Cours de DEUG
Probabilités et Statistiques

Avner Bar-Hen

Université Aix-Marseille III

2002–2003

Table des matières

Chapitre 0

Analyse combinatoire

L'analyse combinatoire fournit des méthodes de dénombrement particulièrement utiles en théorie des probabilités. Une application intéressante est la démonstration du développement du binôme de Newton ; ce développement sera beaucoup utilisé par la suite.

1 Arrangements

Définition 0.1 *Étant donné un ensemble de n objets, on appelle arrangements de p objets toute suite de p de ces objets.*

Cette définition implique que, pour obtenir un arrangement, il faut choisir p objets parmi n et les ordonner, par exemple en leur attribuant une place parmi p ou un numéro de 1 à p. Dans le langage de la théorie des ensemble, on dit :
"Soit un ensemble E à n éléments. On appelle arrangement de p éléments de E, une injection de l'ensemble $\{1, 2, 3, \ldots, p\}$ dans E".
Deux arrangements formés de p objets peuvent donc être distincts soit par la nature des objets, soit par leur ordre. On précise parfois "arrangement sans répétition", parce qu'à un objet ne peut être attribué qu'une place.

Exemple 0.1 *Combien d'arrangements peut-on réaliser en prenant deux objets parmi 4 ?*
Soient les objets $a, , b, c, d$. En choisissant 2 objets et en les ordonnant, par exemple en les alignant, on peut obtenir 12 arrangements :

$$\begin{array}{cccccc} a,b & a,c & a,d & b,c & b,d & c,d \\ b,a & c,a & d,a & c,b & d,b & d,c \end{array}$$

1.1 Dénombrement

On se propose de dénombrer les arrangements possibles de p objets parmi n. On notera A_n^p le nombre de ces arrangements.

Il est aisé de calculer $A_n^1 = n$, puis $A_n^2 = n(n-1)$; il existe en effet n façons de choisir le premier objet et $(n-1)$ façons de choisir le deuxième lorsqu'on a déjà le premier.
Pour déterminer A_n^p, on raisonne par récurrence. On suppose A_n^{p-1} connu. On en déduit :

$$A_n^p = A_n^{p-1}(n-p+1)$$

Il existe en effet A_n^{p-1} façons de choisir les $(p-1)$ premiers objets de la suite et pour chacune de ces possibilité, il reste $n-(p-1) = (n-p+1)$ façons de choisit le dernier.
On a donc successivement :

$$\begin{aligned} A_n^1 &= n \\ A_n^2 &= n(n-1) \\ A_n^3 &= n(n-1)(n-2) \\ \vdots &= \vdots \\ A_n^p &= n(n-1)(n-2)\cdots(n-p+1) \end{aligned}$$

On peut écrire : $A_n^p = n(n-1)(n-2)\cdots(n-p+1)\frac{(n-p)(n-p-1)\cdots(2)(1)}{(n-p)(n-p-1)\cdots(2)(1)}$
Si pour $k \in \mathbb{N}$ et $k \neq 0$, on note $k! = 1 \times 2 \times 3 \times \ldots \times (k-1) \times k$ ($k!$ se lit : factorielle k), on a plus simplement :

$$A_n^p = \frac{n!}{(n-p)!} \tag{1}$$

Exemple 0.2 *En prenant 3 objets parmi 4, on peut réaliser $A_4^3 = \frac{4!}{1!} = 24$ arrangements. Pour vérifier ce résultat, on peut utiliser l'exemple 0.1. Soient les objets a, b, c, d. Les arrangements formés de trois objets parmi ces quatre s'obtiennent en ajoutant un troisième objet aux arrangements formés de deux objets. On obtient :*

$$\begin{array}{cccccc} a,b,c & a,c,b & a,d,b & b,c,a & b,d,a & c,d,a \\ a,b,d & a,c,d & a,d,c & b,c,d & b,d,c & c,d,b \\ b,a,c & c,a,c & d,a,b & c,b,a & d,b,a & d,c,a \\ b,a,d & c,a,d & d,a,c & c,b,d & d,b,c & d,c,b \end{array}$$

On a bien $A_4^3 = A_4^2 \times 2 = 12 \times 2$

Un arrangement de p objets choisis parmi n peut être obtenu en tirant d'abord un objet parmi les n, puis un deuxième parmi les $(n-1)$ restants, etc. Le rang du tirage sert alors à ordonner les objets retenus.
On peut imaginer un type de tirage entièrement différent : on tire d'abord un objet, on le remet parmi les n après avoir noté sa nature, et on répète p fois l'opération. La suite obtenue s'appelle un "arrangement avec répétition " de p objets parmi n.
Il est clair que le nombre d'arrangements avec répétition de p objets parmi n est n^p. On a en effet n possibilités pour chaque place, soit $n \times n \times \cdots \times n = n^p$ possibilités d'arrangement.

2 Permutations

Définition 0.2 *Étant donné un ensemble de n objets, on appelle permutation toute suite de ces n objets.*

Une permutation est donc obtenue en ordonnant l'ensemble de ces n objets. C'est un arrangement particulier où tous les objets sont choisis ; deux permutations de n objets donnés ne peuvent donc différer que par l'ordre de ces objets.

Exemple 0.3 *Avec deux objets x et y, on peut obtenir 2 permutations : x, y et y, x si on les ordonne en les alignant.*

2.1 Dénombrement

Soit P_n le nombre de permutations possibles avec n objets donnés. Le calcul de P_n est simple.
On a en effet : $P_n = A_n^n = n(n-1)(n-2)\cdots 1$
D'où

$$P_n = n! \tag{2}$$

Exemple 0.4 *Avec trois objets a, b, c on obtient $P_3 = 3! = 6$ permutations, à savoir :* $a, b, c \quad b, c, a \quad c, a, b \quad c, b, a \quad b, a, c \quad a, c, b$

3 Combinaisons

Définition 0.3 *Étant donné un ensemble de n objets distincts, on appelle combinaison de p de ces objets tout ensemble de p de ces objets.*

La notion d'ordre a maintenant complètement disparu. Deux combinaisons contenant p objets peuvent donc seulement différer par la nature des objets.

Exemple 0.5 *Soient les nombre 1, 2 , 3, 4. En choisissant 2 nombres, on peut obtenir 6 combinaisons :* $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$, $\{2, 4\}$, $\{3, 4\}$,

3.1 Dénombrement

On se propose de dénombrer les combinaisons possibles de p objets pris parmi n. On note C_n^p le nombre de ces combinaisons. On note aussi parfois $\begin{pmatrix} n \\ p \end{pmatrix}$.
On remarque qu'on peut ordonner une combinaison contenant p objets donnés de $p!$ manières (autant que de permutations). En procédant de même pour chacune des C_n^p combinaisons, on doit obtenir A_n^p arrangements.
On a donc :

$$p! C_n^p = A_n^p$$

D'où :

$$C_n^p = \frac{n(n-1)(n-2)\dots(n-p+1)}{p!}$$

ce que l'on peut écrire

$$C_n^p = \frac{n!}{p!(n-p)!} \tag{3}$$

La definition 0.2 implique : $C_n^1 = n$ et $C_n^n = 1$. Pour généraliser la relation 3 à ce dernier cas, on posera par convention : $0! = 1$.
Ceci permet d'écrire : $1 = \frac{n!}{n!0!} = C_n^0$; si on ne prend aucun objet parmi n, on obtient l'ensemble vide.

3.2 Deux relations importantes

1.
$$C_n^p = C_n^{n-p} \tag{4}$$

   On a en effet, d'après la relation 3 : $\frac{n!}{p!(n-p)!} = \frac{n!}{(n-p)!p!}$
   On peut aussi remarquer qu'à tout sous-ensemble de p objets correspond son complémentaire formé des $(n-p)$ objets restants.

2.
$$C_n^p = C_{n-1}^p + C_{n-1}^{p-1} \tag{5}$$

   En explicitant on a successivement :

$$\begin{aligned}
C_{n-1}^p + C_{n-1}^{p-1} &= \frac{(n-1)!}{(n-1-p)!p!} + \frac{(n-1)!}{(n-1-p+1)!(p-1)!} \\
&= \frac{(n-1)(n-2)\dots(n-p)}{p!} + \frac{(n-1)(n-2)\dots(n-p+1)}{(p-1)!} \\
&= \frac{(n-1)(n-2)\dots(n-p+1)}{p!}(n-p+p) \\
&= \frac{n(n-1)(n-2)\dots(n-p)}{p!} \\
&= C_n^p
\end{aligned}$$

   On peut aussi démontrer la relation 5 en notant que, si on désigne à l'avance un objet parmi n, les combinaisons possibles avec ces n objets pris p à p se décomposent en deux catégories :
   - celles qui contiennent cet objet. Il y en a C_{n-1}^{p-1} ; pour compléter chaque combinaison il suffit en effet de choisir $(p-1)$ objets parmi les $(n-1)$ restants.
   - celles qui ne contiennent pas cet objet. Il y en a C_{n-1}^p ; pour les obtenir il faut en effet choisir p objets parmi les $(n-1)$ qui sont différents de l'objet désigné.

4 Applications

4.1 *Développement du binôme de Newton* [1]

L'analyse combinatoire permet aisément le développement de $(a+x)^n$ où n est un entier strictement positif ($n \in \mathbb{N}$, $n \neq 0$)
Élever $(a+x)$ à la puissance n revient à multiplier n binômes identiques à $(a+x)$. Le résultat est une somme ; chaque élément de cette somme est le produit de n facteurs, du type a ou x, choisis chacun dans un binôme différent.
On obtient donc des termes de la forme $a^p x^{n-p}$ $(0 \leq p \leq n)$. Chacun d'eux sera présent autant de fois qu'il existe de façons de choisir les p facteurs a, c'est-à-dire de choisir p binômes parmi les n (les binômes où l'on choisit les facteurs x sont alors bien déterminés) ; il y en a donc C_n^p. Sommer C_n^p termes identiques à $a^p x^{n-p}$ revient à multiplier $a^p x^{n-p}$ par le coefficient numérique C_n^p. On obtient ainsi, si $n \in \mathbb{N}$ et $n \neq 0$:

$$(a+x)^n = x^n + C_n^1 a x^{n-1} + \cdots + C_n^p a^p x^{n-p} + \cdots + C_n^{n-2} a^{n-2} x^2 + C_n^{n-1} a^{n-1} x + a^n \quad (6)$$

Comme $C_n^{n-p} = C_n^p$ cette expression met en évidence la symétrie des coefficients numériques du développement du binôme de Newton rangé suivant les puissances croissantes de a (ou de x).

Exemple 0.6 *En donnant à n successivement les valeurs 1, 2, 3 on obtient :*

$$\begin{aligned}
(a+x)^1 &= 1a + 1x \\
(a+x)^2 &= 1a^2 + 2ax + 1x^2 \\
(a+x)^3 &= 1a^3 + 3a^2x + 3ax^2 + 1x^3
\end{aligned}$$

4.2 *Le triangle de Pascal* [2]

On considère les développements :

$$\begin{aligned}
(x+a)^{n-1} &= \underbrace{C_{n-1}^0 x^{n-1} + C_{n-1}^1 a x^{n-2}} + \cdots + \underbrace{C_{n-1}^{p-1} a^{p-1} x^{n-p} + C_{n-1}^p a^p x^{n-p-1}} + \cdots \\
(x+a)^n &= C_n^0 x^n + C_n^1 a x^{n-1} + \cdots\cdots\cdots\cdots\cdots\cdots + C_n^p a^p x^{n-p} + \cdots
\end{aligned}$$

On s'aperçoit que le coefficient C_n^p, d'après la relation 5 peut être obtenu par l'addition de C_{n-1}^p et de C_{n-1}^{p-1}. D'où l'idée du triangle de Pascal qui permet d'obtenir, par récurrence, les coefficients numériques du développement du binôme de Newton.

Exemple 0.7 *En utilisant le triangle de Pascal (Table 1), on peut écrire :*

$$(a+x)^6 = x^6 + 6ax^5 + 15a^2x^4 + 20a^3x^3 + 15a^4x^2 + 6a^5x + a^6$$

[1]NEWTON Issac (1642-1727), mathématicien et physicien anglais
[2]PASCAL Blaise (1623-1662), mathématicien, physicien, philosophe et écrivain français

L'analyse combinatoire conduit à utiliser des factorielles.
Quand on sait que $20! \approx 2.432\ 10^{18}$, on se rend compte des difficultés rencontrées dans les calculs. Par la suite il sera surtout intéressant de connaître l'ordre de grandeur d'une probabilité. Il est donc utile d'obtenir une valeur approchée de $n!$ pour les valeurs élevées de n ; ceci est possible par l'intermédiaire de la formule de Stirling[3] :

$$n! = n^n e^{-n}\sqrt{2\pi n}[1 + \epsilon(n)] \tag{7}$$

où $\epsilon(n) \longrightarrow 0$ lorsque $n \longrightarrow +\infty$
Si n est suffisamment grand, on peut donc, dans les calculs, écrire :

$$n! \approx n^n e^{-n}\sqrt{2\pi n} \tag{8}$$

En utilisant les logarithmes, il est alors facile d'obtenir une bonne approximation de $n!$

Exemple 0.8 *On se propose d'estimer* $20!$ *en utilisant l'équation 8.*

$$\begin{aligned}
\log(20!) &\approx 20\log 20 - 20\log e + \frac{1}{2}\log 20 + \log\sqrt{2} + \frac{1}{2}\log\pi \\
&\approx 20.5\log 20 - 20\log e + 0.39908
\end{aligned}$$

On obtient : $\log(20!) \approx 18.38439$*, d'où* $20! \approx 2.423\ 10^{18}$*, soit une erreur relative qui est seulement de l'ordre de* $4.3\ 10^{-3}$

[3]STIRLING James (1698-1770), mathématicien anglais

Tableau 1 – Triangle de Pascal

$n \backslash p$	0	1	2	3	4	5	6	7
1	1	1						
2	1	2	1					
3	1	3	3	1				
4	1	4	6	4	1			
5	1	5	10	10	5	1		
6	1	6	15	20	15	6	1	
7	1	7	21	35	35	21	7	1

On lit : $C_5^2 = 10$. Ce résultat s'obtient en ajoutant C_4^2 (même colonne, ligne du dessus) et C_4^1 (colonne de gauche, ligne du dessus)

5 Exercices

1. Combien de nombres peut-on former avec les chiffres 0,1,2,3,4,5,6,7, chaque chiffre n'étant présent q'une fois, de façon que chaque nombre commence par un 7 et soit divisible par 5,
   - si les nombres sont de 8 chiffres ?
   - si les nombres sont de 6 chiffres ?

   Réponses : 1440 ; 720
2. Cinq hommes et quatre femmes vont au cinéma. Ils disposent d'une rangée de neuf places. De combien de façons différentes peuvent-ils s'asseoir si l'on veut que chaque femme soit entourée de deux hommes ?

   Réponse : 2880
3. On dispose de n boules. On veut former k groupes contenant respectivement $r_1, r_2, \ldots, r_k$ boules, en utilisant toutes les boules ($r_1 + r_2 + \cdots + r_k = n$). De combien de façons peut-on le réaliser ?

   Réponses : $\frac{n!}{r_1!r_2!\cdots r_k!}$
4. On dispose de 5 billes rouges, 2 blanches et 3 bleues. Si les billes de même couleur sont indiscernables, de combien de façons peut-on les aligner ?

   Réponses : 2520
5. Un groupe de 5 mathématiciens et 7 physiciens doit élire un comité représentatif formé de 2 mathématiciens et 2 physiciens. Quel est le nombre de résultats possibles si :
   - les 12 personnes sont éligibles ?
   - un physicien est élu d'office ?
   - 2 mathématiciens ne sont pas éligibles ?

   Réponses : 210 ; 60 ; 63
6. Le jeu de l'écarté se joue avec 32 cartes, chaque joueur en recevant 5.
   - combien de mains différentes peut avoir un joueur ?
   - combien y-a-t-il de mains contenant 2 atouts et 2 seulement ?

   Réponses : 201376 ; 56672
7. Quel est le nombre de groupes de six personnes que l'on peut former avec 4 garçons et 6 filles si l'on veut qu'ils contiennent obligatoirement 2 garçons,
   - donnés ?
   - seulement ?
   - au moins ?

   Réponses : 70 ; 90 ; 185
8. Démontrer les trois relations :

$$\sum_{i=0}^{n} C_n^i = 2^n \tag{9}$$

$$\sum_{i=1, i \text{ impair}}^{n} C_n^i = \sum_{i=1, i \text{ pair}}^{n} C_n^i = 2^{n-1} \tag{10}$$

(11)

9. Développer : $(P+Q)^{10}$
   Réponse : $Q^{10}+10PQ^9+45P^2Q^8+120P^3Q^7+210P^4Q^6+252P^5Q^5+210P^6Q^4+120P^7Q^3+45P^8Q^2+10P^9Q+P^{10}$
10. Estimer c_{100}^{50}
    Réponse : $1.011\ 10^{29}$

Chapitre 1

Premiers éléments de théorie des probabilités

Le but de la théorie des probabilités est de dégager des méthodes permettant de faire des prédictions, qualitatives ou quantifiées, sur le déroulement des phénomènes qui sont régis par le hasard. On va d'abord préciser et isoler l'objet de notre étude :

1 Notion d'expérience aléatoire

Dans chaque cas, le phénomène étudié peut être conçu comme une "expérience" dont le résultat est aléatoire : lorsqu'on reproduit l'expérience (dans des conditions précisées) le résultat de l'expérience varie et semble dépendre du hasard. On dit qu'il s'agit d'une *expérience aléatoire* ; on la représentera par la lettre $\mathcal{E}$.

Exemple 1.1 *on jette deux dés de couleurs différentes : le résultat de l'expérience est exprimé par la donnée des nombres affichés par chacun des dés.*

Exemple 1.2 *On tire "au hasard" successivement et sans remise, deux boules d'une urne contenant 10 boules numérotées de 1 à 10 : le résultat peut être exprimé par la succession des deux numéros des boules tirées.*

Exemple 1.3 *On note la taille d'un individu pris "au hasard" dans une population donnée, le nombre consiste en un nombre réel (l'unité de longueur ayant été choisie)*

Remarquons que très souvent le hasard résulte, ou bien d'un manque d'informations sur les conditions expérimentales, ou bien de l'impossibilité pratique d'exploiter les données expérimentales pour prévoir le résultat.

Exemple 1.4 *jeux de cartes, roulettes, jeu de dés, etc.*

Pour étudier un phénomène aléatoire, il faudra d'abord l'assimiler à une expérience aléatoire (qui est presque toujours une notion idéale ou virtuelle) et associer ensuite à ce phénomène un *modèle mathématique* ; c'est sur ce modèle qu'on pourra le plus commodément raisonner et calculer. Cette modélisation est l'objet des paragraphes suivants.

2 L'ensemble fondamental Ω

On appelle ensemble fondamental associé à $\mathcal{E}$, tout ensemble Ω dont les éléments représentent toutes les issues (ou résultats) possibles de $\mathcal{E}$. Pour chacun des trois exemples mentionnés précédemment, on a :

Exemple 1.1 $\Omega = \{[a,b]; a,b$ entiers compris entre 1 et 6 $\}$

Exemple 1.2 $\Omega = \{[a,b]; a,b$ entiers *distincts* compris entre 1 et 10 $\}$

Exemple 1.3 $\Omega = \mathbb{R}_+ = [0,+\infty[$. On pourra en pratique se restreindre à un intervalle plus petit

Exemple 1.5 *(jeu de pile ou face) si $\mathcal{E}$ consiste à jouer 2 fois à pile ou face*

$$\Omega = \{PP, PF, FP, FF\}$$

On a un modèle presque identique pour l'expérience $\mathcal{E}$ suivante : on choisit au hasard une famille parmi les familles de 2 enfants, et on note le sexe des enfants, celui du plus jeune, puis celui de l'aîné : $\Omega = \{FF, FG, GF, GG\}$.
Si on note seulement le sexe des enfants sans se préoccuper de l'ordre des naissances, on aura un ensemble Ω différent $\Omega = \{FF, FG, GG\}$

Exemple 1.6 *(modèle mendélien de transmission d'un caractère) on considère un gêne admettant deux allèles A et a ; $\mathcal{E}$ consiste en l'observation du descendant direct de deux hétérozygotes ; on prend $\Omega = \{AA, Aa, aa\}$ (ensemble des génotypes possibles).*
Si on considère à la fois deux gênes, admettant chacun deux allèles A, a et B, b, on prendra $\Omega = \{AABB, AABb, AaBB, \ldots\}$ (il y 9 issues possibles).

3 Les événements

Considérons une propriété (ou qualité) E liée au résultat (ou à l'issue) de l'expérience aléatoire $\mathcal{E}$: à chaque mise en œuvre de $\mathcal{E}$, ou bien E est réalisée, ou bien E n'est pas réalisée.

Exemple 1.7 *Dans l'exemple 1.1, on a E="les deux dés affichent des nombres pairs" ; dans l'exemple 1.2, on a E="on a tiré au moins une boule de numéro impair", etc.*

Une telle propriété E permet de partager l'ensemble fondamental Ω en deux parties : d'une part, l'ensemble A_E des points $\omega \in \Omega$ représentant une issue de $\mathcal{E}$ pour laquelle E a lieu, et d'autre part, la partie $\overline{A_E} \subset \Omega$ formée des points $\omega \in \Omega$ associés à une issue ne réalisant pas E.
On obtient ainsi une représentation de la propriété E par une partie A_E de Ω.
On dit que A_E est l'*événement* attaché à la propriété E ; on dit aussi que A_E est l'événement "E est réalisé". On voit immédiatement que $\overline{A_E}$ est aussi un événement, c'est l'événement "non E est réalisé", c'est-à-dire "E n'est pas réalisé".

Exemple 1.8 *Dans l'exemple 1.1, l'événement "la somme des nombres obtenus est 6" est l'ensemble* $A = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$
Dans l'exemple 1.6, considérons que le modèle de transmission d'un caractère lié à un gêne admettant deux allèles A, a *soit une couleur : vert lié à* A*, ou jaune lié à* a *et supposons* A *dominant. L'événement "le descendant présente le caractère vert" est représenté par* $\{AA, Aa\}$.

Événement certain : c'est l'événement lié à une propriété réalisée lorsqu'on effectue $\mathcal{E}$; cet événement est l'ensemble Ω tout entier.

Événement impossible : c'est l'événement associé à une propriété qui n'est jamais réalisée ; cet événement est la partie vide de E, notée $\emptyset$.

Événement complémentaire : si A est un événement, on dit que $\overline{A}$ est l'événement complémentaire de A dans Ω. On retiendra que $\overline{A}$ est l'événement "non A".

Événement élémentaire : on qualifie ainsi les événements liés à $\mathcal{E}$ qui ne sont réalisés que pour une issue de $\mathcal{E}$; ce sont les parties de Ω réduites à un point.

4 Algèbre des événements

Outre l'opération de complémentation, il existe deux autres opérations importantes sur l'ensemble $\mathcal{A}$ des événements liés à une expérience aléatoire donnée $\mathcal{E}$:

La réunion : si A et B sont deux événements (pour $\mathcal{E}$), l'événement "A ou B" est la partie de Ω réunion des parties de A et B, et on a : "A ou B"=$A \cup B$.

L'intersection : l'événement "A et B" est la partie de Ω intersection de A et B, et on a : "A et B"=$A \cap B$.

Exemple 1.9 *sur le jeu de pile ou face (exemple 1.5) avec deux lancers : l'événement élémentaire "*PP*" (2 fois pile) est intersection de l'événement* A*="d'abord pile"=*$\{PP, PF\}$*, avec l'événement* B*="Pile au 2ème lancer"=*$\{PP, FP\}$.
L'événement réunion $A \cup B$ *est l'événement "au moins une fois pile" ; l'événement complémentaire* $\overline{A \cup B}$ *de* $A \cup B$ *est l'événement "jamais pile", c'est-à-dire "deux fois face".*

L'ensemble $\mathcal{A}$ des événements est donc naturellement muni des opération de complémentation, de réunion et d'intersection ; on traduit ce fait en disant que $\mathcal{A}$ est une algèbre de parties de Ω. Une étude approfondie de la théorie des probabilités nécessite aussi la considération de réunion (ou d'intersection) de suites *infinies* d'événements.
Dans ce cours, on se limitera presque toujours à des opérations portant sur un nombre fini d'événements.

Événements incompatibles : on dit que les événements A et B (relatifs à une même expérience aléatoire) sont incompatibles, si ils ne peuvent être réalisés simultanément, autrement dit si $A \cap B = \emptyset$.

Système complet d'événements : on appelle ainsi toute suite $A_1, \dots, A_n$ d'événements deux à deux incompatibles, telle que $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$; les ensembles A_i, $1 \leq i \leq n$, forment alors une partition de Ω.

Exemple 1.10 *$\mathcal{E}$= lancer de deux dés ; on pose, pour $i = 1, 2, \dots, 6$, $A_i =$"le premier dé affiche i".*

Pour achever de décrire le modèle mathématique associé à l'expérience aléatoire, il reste à introduire la notion de probabilité.

5 La probabilité P

On admet qu'à chaque événement $A \in \mathcal{A}$ est associé un nombre $\mathbb{P}(A)$, compris entre 0 et 1 et qui mesure la probabilité de réalisation de A.
Intuitivement, $\mathbb{P}(A)$ est la fréquence de réalisation de A au cours d'un très grand nombre de mises en œuvre de l'expérience aléatoire considérée.
$\mathbb{P}$ est une application, définie sur $\mathcal{A}$, (l'algèbre des événements attachés à $\mathcal{E}$) et à valeurs dans l'intervalles $[0, 1]$:

$$\mathbb{P} : \mathcal{A} \mapsto [0, 1]$$

Il est clair qu'on doit avoir $\mathbb{P}(\emptyset) = 0$ et $\mathbb{P}(\Omega) = 1$.
Considérons maintenant deux événements *incompatibles* A et B ; imaginons qu'on effectue l'expérience $\mathcal{E}$ N fois (N très grand) et qu'on note n_A et n_B le nombre de fois où on a observé A et B (respectivement) au cours de ces N réalisations de $\mathcal{E}$.
Comme A et B sont incompatibles, il est clair que le nombre $n_{A\cup B}$ de réalisations de "A ou B" est exactement $n_{A\cup B} = n_A + n_B$. D'où

$$\frac{n_{A\cup B}}{N} = \frac{n_A}{N} + \frac{n_B}{N}$$

autrement dit, la fréquence de réalisation de $A \cup B$ est la somme des fréquences de réalisation de A et de B. Ce raisonnement conduit à la propriété suivante :
Si A et B sont deux événements incompatibles, on a :

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

Plus généralement, si $A_1, A_2, \dots, A_n$ sont deux à deux incompatibles :

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n)$$

Conséquences :

1. pour tout événement $A \in \mathcal{A}$:

$$\mathbb{P}(A) = 1 - \mathbb{P}(\overline{A})$$

2. si $A \in \mathcal{A}, B \in \mathcal{A}$:

$$\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$$

et en particulier

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$$

On pourra établir ces relations à titre d'exercice.
Soulignons encore un propriété générale des probabilités : Soient $A \in \mathcal{A}, B \in \mathcal{A}$, si $A \subset B$, la réalisation de A entraîne celle de B et on dit que A implique B ; comme $\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap \overline{A})$ (pourquoi ?), on voit que :

$$\text{si } A \subset B, \text{ alors } \mathbb{P}(A) \leq \mathbb{P}(B)$$

5.1 *Notion d'espace de probabilité*

On dit que le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ constitué par la donnée de Ω, de l'algèbre des événements $\mathcal{A}$ et de la probabilité $\mathbb{P}$, est un espace de probabilité. C'est le modèle mathématique qu'on associe à l'expérience aléatoire $\mathcal{E}$.
Dans chaque problème concret, des hypothèses naturelles, ou des faits expérimentaux, permettent de déterminer la probabilité $\mathbb{P}$ ou du moins celle de certains événements qu'on est conduit à considérer.
Considérons un premier exemple :

6 Le cas élémentaire

Supposons l'ensemble fondamental Ω fini, et supposons que toutes les issues ont la même probabilité (on dit qu'elles sont équiprobables). Il est alors facile de déterminer $\mathbb{P}$:
Notons $|A|$ le nombre d'éléments de la partie A de Ω. Comme $\mathbb{P}(\Omega)$ est la somme des probabilités des événements élémentaires $\{\omega\}$, (ω parcourant Ω) et comme les $\mathbb{P}(\{\omega\})$ sont par hypothèses égaux entre eux, on obtient :

$$\forall \omega \in \Omega : \quad \mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}$$

Si A est une partie de Ω, $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$; d'où :

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

c'est-à-dire le rapport du nombre des issues "favorables" au nombre de toutes les issues possibles.
Dans ce cas, le calcul de la probabilité d'un événement peut donc être ramené à un dénombrement.

Exemple 1.11 *Prenons $\mathcal{E}$= on joue deux fois avec un dé.*

$$\Omega = \{[a,b]; a,b \textit{ entiers }, 1 \leq a,b \leq 6\}$$

Si le dé est équilibré, il est naturel de supposer les événements élémentaires équiprobables, d'où, $\mathbb{P}(\{\omega\}) = 1/36$ pour tout $\omega \in \Omega$.
La probabilité que la somme des nombres obtenus soit 6 est $5/36$.

7 Rappel de combinatoire : applications

7.1 Arrangements avec répétition

Soient $X_1, X_2, \ldots, X_n$ n ensembles finis et $X = X_1 \times X_2 \times \cdots \times X_n$ le produit cartésien de ces ensembles :

$$X = \{(x_1, x_2, \ldots, x_n); x_1 \in X_1, x_2 \in X_2, \ldots, x_n \in X_n\}$$

(**Rappel :** un n-uple, tel que $(x_1, x_2, \ldots, x_n)$ est une suite de n éléments rangés dans un certain ordre)
Alors, le cardinal (ou nombre d'éléments) de X est le produit des cardinaux de $X_1, X_2, \ldots, X_n$:

$$|X| = |X_1| \times |X_2| \times \cdots \times |X_n|$$

Exemple 1.12 *On jette trois dés (= $\mathcal{E}$).*
Alors $\Omega = \{(a,b,c); a,b,c$ entiers, $1 \leq a,b,c \leq 6\}$ contient $6 \times 6 \times 6 = 216$ éléments.
La probabilité d'obtenir deux 1 et un 3 est $3/216$.
Remarque : *on a modélisé l'expérience en faisant comme si les trois dés étaient distinguables ; c'est ce qui permet de supposer toutes les issues équiprobables.*

Soient X un ensemble à n éléments et p un entier. On appelle p-arrangement (avec répétition) d'éléments de X la donnée de p éléments de X, *non nécessairement distincts*, rangés dans un certain ordre.
Il revient au même de se donner un éléments de $X^p = \underbrace{X \times X \times \ldots \times X}_{p \text{ fois}}$.
Il y a donc n^p p-arrangements (avec répétition) d'éléments de X.

7.2 Arrangements

Considérons un ensemble X à n éléments et p un entier compris entre 0 et n. Un p-arrangement d'éléments de X est la donnée de p éléments *distincts* de X, rangés dans un certain ordre. Le nombre A_n^p de tels arrangements est donné par la formule :

$$A_n^p = n(n-1)\cdots(n-p+1)$$

Il y a en effet n façons de choisir le 1er élément de l'arrangement, puis, ce choix étant fait, $(n-1)$ choix de pour le 2ème, ..., et $(n-p+1)$ choix possibles pour le $p^{\text{ème}}$ éléments de l'arrangement lorsqu'on a choisi les $p-1$ premiers. Ce qui justifie la formule.
En utilisant la notation factorielle ($k! = 1 \times 2 \times \ldots \times k$ si k entier et $k \geq 1$, et par convention $0! = 1$), on peut aussi écrire :

$$A_n^p = \frac{n!}{(n-p)!}$$

Prenons $p = n$: un n-arrangement d'éléments de X est la donnée d'un ordre total sur X ; on dit aussi que c'est une *permutation* de X. Il y a en tout $n! = A_n^n$ permutations de X.

Exemple 1.13 *$\mathcal{E}$="on jette 3 dés". Il y a $3! = 6$ manières d'obtenir un 1, un 2 et un 5. La probabilité de cet événement est donc $^6/_{216}$=$^1/_{36}$.*

Exemple 1.14 *$\mathcal{E}$= "on tire au hasard et sans remise 3 boules dans une urne contenant 10 boules dont 5 rouges, 3 noires et 2 blanches. Cherchons la probabilité d'obtenir successivement une boule rouge, une boule noire et une boule blanche (événement RNB). On prend (en imaginant une numérotation des boules) :*

$$\Omega = \{(a,b,c); a,b,c \text{ entiers 2 à 2 distincts}, 1 \leq a,b,c \leq 10\}$$

Ω contient $A_{10}^3 = 10 \times 9 \times 8 = 720$ éléments qu'il est naturel de supposer équiprobables. Le nombre d'issues "favorables" est $5 \times 3 \times 2 = 30$. D'où $\mathbb{P}(RNB) = {}^{30}/_{720} = {}^1/_{24}$

7.3 Combinaisons

Une p-combinaisons d'éléments de X est la donnée d'une partie de X à p éléments (sans qu'on ait défini un ordre de succession de ces éléments). Le nombre C_n^p des p-combinaisons de X (X ayant n éléments) est donné par la formule :

$$C_n^p = \frac{A_n^p}{p!} = \frac{n!}{p!(n-p)!}$$

Chaque p-combinaison donne naissance à $p!$ p-arrangements, obtenus en choisissant un ordre de succession sur les éléments de la combinaison.
En particulier $C_n^p = C_n^{n-p}$, et on vérifie : $C_{n+1}^{p+1} = C_n^p + C_n^{p+1}$

Exemple 1.15 *Au jeu de Bridge, quelle est la probabilité d'avoir 4 As dans une main de 13 cartes ? (on utilise un jeu de 52 cartes)..*
Se donner une main de 13 cartes, c'est se donner une 13-combinaison extraite de l'ensemble des 52 cartes.
Prenons pour $\mathcal{E}$ l'observation des 13 cartes d'un jeu donné, et donc Ω = l'ensemble des 13-combinaisons pris dans l'ensemble des 52 cartes. Si on admet que toutes les

issues sont équiprobables, chaque issue a la probabilité $1/C_{52}^{13}$ *; comptons alors les 13-combinaisons comprenant les 4 As : il y en a autant que de 9 combinaisons parmi 48 cartes, d'où :*

$$\mathbb{P}(4\,As) = \frac{C_{48}^{9}}{C_{52}^{13}} = \frac{(48)!(13)!}{(52)!(9)!} = \frac{10 \times 11 \times 12 \times 13}{49 \times 50 \times 51 \times 52} \approx 2.64\,10^{-3}$$

Remarque : *on aurait pu prendre pour* $\mathcal{E}$ *l'observation des cartes d'un joueur donné dans l'ordre où elles ont été distribuées ; alors* Ω *est l'ensemble des* A_{52}^{13} *13-arrangements possibles ; le nombre de ces arrangements contenant les 4 As est :*

$$m = A_{13}^{4} \times A_{48}^{9}$$

Il y a A_{13}^{4} *dispositions possibles pour les 4 As, et, cette disposition étant fixée, il y a* A_{48}^{9} *arrangements possibles.*
Comme les différents 13-arrangements sont équiprobables, on trouve :

$$\mathbb{P}(4\,As) = \frac{A_{13}^{4} \times A_{48}^{9}}{A_{52}^{13}} = \frac{10 \times 11 \times 12 \times 13}{49 \times 50 \times 51 \times 52}$$

c'est-à-dire le même résultat que précédemment mais par un raisonnement plus difficile.

Les notions de *probabilité conditionnelle* et d'*indépendance* jouent un rôle fondamental dans toute la suite ; elles sont indispensables dans la conduite des calculs dès qu'on sort du cas élémentaire considéré précédemment. D'ailleurs même dans ce cas, et dès qu'on a acquis un peu l'habitude, elles permettent souvent des calculs plus simples et plus sûrs.

8 Probabilité conditionnelle

8.1 *Probabilité conditionnelle*

Considérons une expérience aléatoire $\mathcal{E}$ d'espace de probabilité associé $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 1.1 *Soient* $A, B \in \mathcal{A}$ *et supposons* $\mathbb{P}(A) \neq 0$. *La* probabilité conditionnelle de B, A étant donné *est le nombre :*

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

On dit aussi que $\mathbb{P}(B|A)$ *est la* probabilité de B sachant A.

Signification : imaginons qu'on effectue l'expérience $\mathcal{E}$ et qu'on donne à un individu l'information que A a été observé lors de cette réalisation de $\mathcal{E}$; on ne donne aucune information (en dehors de la description de $\mathcal{E}$). Pour cet individu, la probabilité de réalisation de B sera en général différente de ce qu'elle était avant de savoir que A a été réalisé.

Exemple 1.16 *On tire 5 cartes parmi 52. Si on sait que la première carte est l'as de trèfle (événement A), la probabilité d'avoir 4 As (événement B) est intuitivement :*

$$p = \frac{48}{C_{51}^4} = \frac{4!48!}{51!}$$

Il y a C_{51}^4 combinaisons possibles pour les quatre dernières cartes, et 48 de ces combinaisons comportent les 3 As manquants.
D'autre part $\mathbb{P}(A) = 1/52$ et $\mathbb{P}(A \cap B) = 1/5 \mathbb{P}(B) = \frac{48}{5C_{52}^5} = \frac{5!48!}{5 \times 52!} = \frac{4!48!}{52!}$
On vérifie que : $p = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$

Exemple 1.17 *On jette un dé ; si on sait que le nombre obtenu est pair, et en l'absence de toute autre information, la probabilité d'avoir 2 est intuitivement 1/3, celle d'avoir 1 est évidemment 0 ; ce qui est en accord avec la définition précédente de $\mathbb{P}(B|A)$.*

La formule de la probabilité conditionnelle peut se justifier intuitivement. Effectuons l'expérience $\mathcal{E}$ N fois, (N très grand) ; on note N_A, $N_{A\cap B}$ le nombre de fois où "A" et "A et B" ont été effectivement réalisés. La fréquence de réalisation de B lorsqu'on ne considère que les réalisation de $\mathcal{E}$ (parmi les N effectués) qui ont conduit à A est :

$$f = \frac{N_{A\cap B}}{N_A} = \frac{N_{A\cap B}}{N} \times \frac{N}{N_A}$$

D'après la signification de la probabilité d'un événement : $\frac{N_{A\cap B}}{N} \approx \mathbb{P}(A \cap B)$ et $\frac{N_A}{N} \approx \mathbb{P}(A)$ (dans la mesure où N est grand).
On voit que $f \approx \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)} = \mathbb{P}(B|A)$
Par convention, si $A, B \in \mathcal{A}$ et si $\mathbb{P}(A) = 0$, on convient que $\mathbb{P}(B|A) = 0$
Les formules suivantes (*formules des probabilités composées*) découlent immédiatement de la définition de la probabilité conditionnelle. Elles sont néanmoins très importantes dans les calculs (pratiques ou théoriques) :

$$\text{Si } A, B \in \mathcal{A} \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$$

$$\text{Si } A, B, C \in \mathcal{A} \quad \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|A \cap B)$$

Et plus généralement, si $A_1, \ldots, A_n \in \mathcal{A}$ alors :

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \ldots \mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1})$$

Exemple 1.18 *On va calculer des probabilités attachées à l'expérience du tirage sans remise de 3 boules d'une urne ; on suppose que l'urne contient initialement 10 boules dont 5 rouges, 3 noires et 2 blanches. Calculons à l'aide des probabilités conditionnelles $\mathbb{P}(RNB)$ (voir exemple 1.14). On a :*

$$RNB = A_1 \cap A_2 \cap A_3$$

où A_1="la $1^{ère}$ boule tirée est rouge", A_2 ="la $2^{ème}$ boule tirée est noire" et A_3=" la $3^{ème}$ boule tirée est blanche".
Clairement $\mathbb{P}(A_1) = 5/10 = 1/2$ (toutes les boules ayant la même probabilité d'être d'abord tirées) ; de plus $\mathbb{P}(A_2|A_1) = 3/9 = 1/3$ et $\mathbb{P}(A_3|A_1 \cap A_2) = 2/8 = 1/4$. D'où d'après la formule des probabilités composées :

$$\mathbb{P}(RNB) = \frac{5}{10\frac{3}{9}\frac{2}{8}} = \frac{1}{2 \times 3 \times 4} = \frac{1}{24}$$

À titre d'exercice on peut calculer $\mathbb{P}(NRB)$ et $\mathbb{P}(NBR)$. Est-ce étonnant de trouver $1/24$?

8.2 Formule de Bayes

La formule des événements complémentaires permet d'écrire que pour tout $A \in \mathcal{A}$, on

$$\mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1$$

Pour tout couple d'événements A, B la formule des probabilités composées devient donc :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(\overline{A})\mathbb{P}(B|\overline{A})}$$

Cette formule s'appelle la *formule de Bayes*. On parle aussi formule des probabilités des causes.

Exemple 1.19 *(efficacité d'un test de dépistage). On a mis au point un test permettant de dépister une maladie ; ce test n'étant pas infaillible, on souhaite connaître la probabilité d'être atteint par la maladie pour un individu présentant un test positif.*
On suppose connues :

1. *la proportion de la population atteinte par la maladie ;*
2. *la probabilité pour un sujet malade de présenter un test positif ;*
3. *la probabilité pour un sujet sain de fournir un test positif.*

L'expérience aléatoire consiste ici à prendre "au hasard" un individu (dans la population considérée) et à lui faire subir le test. On note T, M, S les événements "le test est positif", "l'individu est malade", "l'individu est sain". Le problème est de trouver $\mathbb{P}(M|T)$ connaissant $\mathbb{P}(M)$, $\mathbb{P}(T|M)$, $\mathbb{P}(T|S)$. On a :

$$\begin{aligned}
\mathbb{P}(M|T) &= \frac{\mathbb{P}(M \cap T)}{\mathbb{P}(T)} \\
&= \frac{\mathbb{P}(M)\mathbb{P}(T|M)}{\mathbb{P}(T \cap M) + \mathbb{P}(T \cap S)} \\
&= \frac{\mathbb{P}(M)\mathbb{P}(T|M)}{\mathbb{P}(M)\mathbb{P}(T|M) + \mathbb{P}(S)\mathbb{P}(T|S)}
\end{aligned}$$

Exemple numérique : *si $\mathbb{P}(M) = 1/100$, $\mathbb{P}(T|M) = 0.8$, $\mathbb{P}(T|S) = 0.1$, on trouve un résultant surprenant : $\mathbb{P}(M|T) \approx 0.075$. Dans ce cas, le test est d'une efficacité presque nulle. Avec $\mathbb{P}(M) = 1/1000$, $\mathbb{P}(T|M) = 0.95$, $\mathbb{P}(T|S) = 0.05$, $\mathbb{P}(M|T) \approx 0.013$.*

9 Indépendance

9.1 *Indépendance de deux événements*

Si A et B sont deux événements (associés à une expérience aléatoire $\mathcal{E}$), on dit que A et B sont indépendants si :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Il revient au même de dire que $\mathbb{P}(B|A) = \mathbb{P}(B)$: ainsi, savoir que A a été réalisé lors d'une mise en œuvre particulière de $\mathcal{E}$ ne modifie pas la probabilité de réalisation de B (et donc ne nous apprend rien de nouveau sur la probabilité que B ait lieu).
Attention à ne pas confondre *indépendants* et *incompatibles*. Si A et B sont incompatibles $A \cap B = \emptyset$ et $\mathbb{P}(A \cap B) = 0$; A et B ne pourront donc être indépendants que si $\mathbb{P}(A) = 0$ ou $\mathbb{P}(B) = 0$.

Exemple 1.20 *On tire une carte au hasard dans un jeu de 52 cartes ; les événements "Pique" et "Valet" sont indépendants (à vérifier), mais non incompatibles (pourquoi ?). Les événements "Pique" et "Trèfle" sont incompatibles mais ils ne sont pas indépendants.*

Si A et B sont indépendants, $\overline{A}$ et B, A et $\overline{B}$, $\overline{A}$ et $\overline{B}$ sont également indépendants (deux à deux).
Vérifions le par exemple pour A et $\overline{B}$: comme $A \cap \overline{B}$ et $A \cap B$ sont incompatibles, et comme $(A \cap \overline{B}) \cup (A \cap B) = A$, on a :

$$\mathbb{P}(A \cap \overline{B}) + \mathbb{P}(A \cap B) = \mathbb{P}(A)$$

D'où :

$$\begin{aligned}
\mathbb{P}(A \cap \overline{B}) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\
&= \mathbb{P}(A)(1 - \mathbb{P}(B)) \\
&= \mathbb{P}(A)\mathbb{P}(\overline{B})
\end{aligned}$$

9.2 *Indépendance de plusieurs événements*

Considérons d'abord le cas de 3 événements A, B, C (relatifs à $\mathcal{E}$). On dira qu'ils sont indépendants (dans leur ensemble) si ils sont indépendants 2 à 2 et si de plus :

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

On peut montrer que cette définition signifie que toute information sur la réalisation de deux d'entre eux au cours d'une mise en œuvre de $\mathcal{E}$ ne modifie pas la probabilité de réalisation du troisième.
Calculons par exemple $\mathbb{P}(A|B \cap \overline{C}) = \frac{\mathbb{P}(A \cap B \cap \overline{C})}{\mathbb{P}(B \cap \overline{C})}$
D'après ce qui précède, B et C sont indépendants et $\mathbb{P}(B \cap \overline{C}) = \mathbb{P}(B)\mathbb{P}(\overline{C})$

D'autre part : $\mathbb{P}(A \cap B \cap \overline{C}) = \mathbb{P}(A \cap B) - \mathbb{P}(A \cap B \cap C)$ car $A \cap B \cap C$ et $A \cap B \cap \overline{C}$ sont incompatibles et leur réunion vaut $A \cap B$.
D'où, $\mathbb{P}(A \cap B \cap \overline{C}) = \mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(\overline{C})$ et donc finalement $\mathbb{P}(A|B \cap \overline{C}) = \mathbb{P}(A)$
Si les événements A, B, C sont deux à deux indépendants, on **ne** peut **pas** en déduire en général qu'ils sont indépendants (dans leur ensemble).

Exemple 1.21 *Prenons $\mathcal{E}$="on jette deux fois un dé", A="on a obtenu 6 au 1^{er} jet", B="on a obtenu 6 au $2^{ème}$ jet", C="on a obtenu le même résultat les 2 fois". Il est à peu près intuitif que ces trois événements sont indépendants* deux à deux. *En effet $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/6$ (par dénombrement, en remarquant qu'il y a en tout 36 issues équiprobables), et $A \cap B = B \cap C = A \cap C =$ "on a obtenu deux fois 6" ; d'où $\mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \mathbb{P}(A \cap C) = 1/36$.*
Par contre $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B) = 1/36$, alors que $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = 1/6^3 = 1/216$

De manière générale on dit que les événements $A_1, A_2 \ldots, A_n$ sont indépendants dans leur ensemble si pour chaque choix d'une partie $\{i_1, \ldots, i_k\}$ de $\{1, \ldots, n\}$ (les i_j étant supposés distincts), on a :

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \ldots \mathbb{P}(A_{i_k})$$

On peut montrer que cette définition équivaut à dire que pour chaque $J = \{i_1, \ldots, i_k\}$ de $\{1, \ldots, n\}$ et chaque $i \notin J$, $1 \leq i \leq n$, on a :

$$\mathbb{P}(A_i|A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_i)$$

Cette relation donne une signification intuitive de l'indépendance.
Il est important de noter que :

- si les événements $A_1, A_2, \ldots, A_n$ sont indépendants, il en est de même pour $A_1, A_2, \ldots, A_{n-1}, \overline{A_n}$. Plus généralement on peut remplacer certains (ou tous) A_i par leur complémentaire $\overline{A_i}$ sans perdre la propriété d'indépendance ;
- Si $A_1, A_2, \ldots, A_n$ sont indépendants, toute "sous-suite" $(A_{i_1}, A_{i_2}, \ldots, A_{i_k})$ $\{i_1, \ldots, i_k\} \in \{1, \ldots, n\}$ est encore une suite d'événements indépendants.

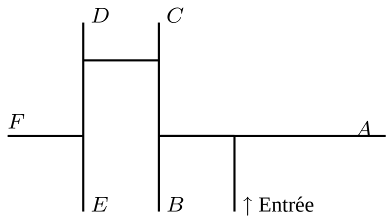
Exemple 1.22 *(schéma de Bernoulli) $\mathcal{E}$ consiste à jouer 3 fois à pile ou face ; on suppose que la probabilité d'obtenir pile au cours d'un lancer est p $(0 \leq p \leq 1)$, (celle d'obtenir face est donc $q = 1 - p$) et on suppose que les lancers sont indépendants.*
p n'est pas forcément égal à $\frac{1}{2}$.
On peut représenter le résultat par un mot : "pile, pile, pile", "pile, face, pile", "face, face, pile", etc. On a, par exemple, :

$$\mathbb{P}(\textit{pile, pile, face}) = \mathbb{P}(\textit{pile})\mathbb{P}(\textit{pile})\mathbb{P}(\textit{face}) = p^2q$$

Ce résultat peut s'énoncer de manière générale : pour n entier fixé, si on joue n fois à pile ou face : la probabilité d'un événement élémentaire A, (c'est-à-dire un événement de la forme "pile, face,..." mot de longueur n formé de pile et de face) est p^iq^j où i est le nombre de pile requis par A et j le nombre de faces ($i + j = n$).

10 Exercices

1. Dans une loterie, 400 billets, dont 4 gagnants, sont mis en vente. Si on achète 10 billets, quelle est la probabilité de gagner au moins un lot ?
2. Soit le labyrinthe ci dessous, dans lequel on fait pénétrer des rat. Quelles sont les probabilités, pour un animal de sortir par A, B, C, D, E ou F ?



3. Trois chasseurs tirent sur un oiseau. Quelle est la probabilité pour que l'oiseau soit touché, sachant que les probabilités de toucher l'oiseau, pour chaque homme, sont $\mathbb{P}(A) = 1/2$, $\mathbb{P}(B) = 2/3$, $\mathbb{P}(C) = 1/4$?
4. Une population est composée de 40% d'hommes et de 60% de femmes ; 50% des hommes et 30% des femmes fument. Quelle est la probabilité pour qu'un fumeur, choisi au hasard soit une femme ?
5. Deux amis se trouvent dans la queue à l'entrée du restaurant universitaire. Sachant que la queue comporte n personnes alignés, quelle est la probabilité pour qu'ils soient séparés par 5 personnes exactement ?
6. Quelle est la probabilité pour que, dans un groupe de 25 personnes, au moins 2 personnes aient leur anniversaire le même jour ?
7. On vous présente un homme. Dans la brève conversation qui suit, vous apprenez que cet homme a deux enfants dont au moins une fille. Quelle est la probabilité pour qu'il ait 2 filles ?
8. On considère des dés à jouer réguliers. Est-il plus probable d'obtenir au moins une fois la face 6 en jetant quatre fois un dé ou d'obtenir au moins une fois le double 6 en jetant vingt quatre fois deux dés ?
9. Soit une espèce diploïde. Un caractère est géré par un couple d'allèles ; tout individu porte donc donc deux de ces allèles qui sont ici de type A ou a. On suppose qu'aux trois génotypes possibles correspondent seulement deux phénotypes : D pour AA et Aa, R pour aa ; de plus, le caractère n'est pas lié au sexe et les accouplements ont lieu au hasard.

   On désigne par u_i, v_i, w_i les probabilités pour qu'un individu, choisi au hasard dans la génération F_i, soit respectivement de génotype AA, Aa, ou aa. En F_0, on avait seulement des individus de type D, avec $u_0 = 1/3$ et $v_0 = 2/3$. on s'intéresse maintenant à la génération F_1 :
   (a) calculer w_1, puis v_1, puis u_1 ;
   (b) si on choisit au hasard un individu de phénotype D, quelle est la probabilité pour qu'il soit de génotype AA ?

10. On a mélangé, par inadvertance, des graines de deux provenances différentes : A et B. On a ainsi un ensemble de graines dont $1/3$ provient de A et $2/3$ de B. La moitié des graines de A et les trois quarts des graines de B sont noires.

    On choisit une graine au hasard : elle est noire. Quelle est la probabilité pour qu'elle provienne de A ?

11. Cinq hommes et quatre femmes vont au cinéma. Ils disposent de 9 places alignées. Ils s'assoient au hasard, mais de manière à ce que chaque femme soit entourée de deux hommes. Quelle est la probabilité pour que monsieur P soit à côté de mademoiselle C ?

12. On dispose de quatre jetons : A, B, C, D tels que A et B ont deux faces blanches, C a une face noire et une face blanche, D a deux faces noires. On tire un jeton au hasard et on ne voit que l'une de ses faces : elle est blanche. Calculer la probabilité pour que l'autre face soit blanche.

13. Soit une urne contenant 10 boules rouges, 5 blanches, 3 noires et 2 jaunes.
    (a) On tire successivement 4 boules (tirage exhaustif) :
        - calculer la probabilité d'obtenir dans l'ordre : 1 rouge, 1 noire, 1 blanche, 1 jaune ;
        - calculer la probabilité d'obtenir seulement 3 boules rouges sur les 4 tirées ;
        - calculer la probabilité pour que la deuxième boule soit rouge.

    (b) On tire successivement 3 boules (tirage non exhaustif) :
        - calculer la probabilité d'obtenir seulement 2 boules rouges sur les 3 tirées ;
        - calculer la probabilité d'obtenir au moins 2 boules rouges sur les 3 tirées ;
        - calculer la probabilité pour que la deuxième boule soit rouge.

14. On admet le modèle suivant :
    - La coloration de l'iris, chez l'homme, est gérée par un couple d'allèles ; elle est indépendantes du sexe ; la coloration bleue est récessive.
    - Madame Dupont n'a pas les yeux bleus, ses parents n'avaient pas les yeux bleus, mais elle a un frère qui a les yeux bleus.
    - Monsieur Dupont n'a pas les yeux bleus, mais il a eu d'un premier mariage, un enfant aux yeux bleus.

    (a) Monsieur et Madame Dupont vont avoir leur premier enfant. Quelle est la probabilité pour qu'il ait les yeux bleus ?

    (b) Monsieur et Madame Dupont ont déjà eu un premier enfant qui n'a pas les yeux bleus, quelle est la probabilité pour que cet enfant puisse transmettre la coloration bleue ?

    (c) si Monsieur et Madame Dupont ont déjà eu un premier enfant aux yeux bleus, quelle est la probabilité pour que leur deuxième enfant ait les yeux bleus ?

15. On considère deux locus homologues d'une plante diploïde et on admet que les gènes correspondants ont deux allèles A et a.

    Partant d'une plante hétérozygote (à la génération 0) on constitue par autofécondation des générations successives.

On désigne par x_n, y_n et z_n les probabilité des génotypes AA, Aa et aa à la $n^{\text{ème}}$ génération ($x_0 = z_0 = 0$, $y_0 = 1$).

(a) Calculer x_n, y_n et z_n en fonction de x_{n-1}, y_{n-1} et z_{n-1} ;

(b) en déduire x_1, y_1, z_1, puis x_2, y_2, z_2, puis x_3, y_3, z_3 ;

(c) calculer x_n, y_n, z_n ;

(d) soit p_n la probabilité pour qu'une plante de la $n^{\text{ème}}$ génération soit homozygote. Quels sont les nombres n_1 et n_2 de générations successives pour que $p_{n_1} \geq 0.95$ et $p_{n_2} \geq 0.999$?

16. Dix couples sont réunis dans une soirée. On admet que, pour danser, chaque homme choisit une femme au hasard.
    - Quelle est la probabilité pour que chacun des 10 hommes danse avec son épouse ?
    - Quelle est la probabilité pour que monsieur Dupond danse avec son épouse ?
    - Quelle est la probabilité pour que monsieur Dupond et monsieur Durand dansent avec leur épouse ?
    - Quelle est la probabilité pour que monsieur Dupond ou monsieur Durand dansent avec son épouse ?
17. Dans une population donnée, un individu peut être atteint d'une affection A avec la probabilité $p_A = 1/100$ et d'une affection B, indépendante de A, avec une probabilité $p_B = 1/20$. Quelle est la probabilité pour qu'un individu choisi au hasard soit atteint d'au moins une des deux maladies ?
18. On appelle "durée de vie" d'un composant électronique le temps pendant lequel ce composant mis sous tension, fonctionne correctement. Un critère définissant la qualité d'une fabrication donnée est : la probabilité pour que la durée de vie d'un composant soit au moins égale à 10000 heures est de 80%. On admet que les pannes de composants différents sont des événements indépendants.

    (a) Avec 3 composants, on réalise un premier montage. Il suffit que l'un des composants soit défectueux pour que le système réalisé ne remplisse plus sa fonction (montage "série"). Calculer la probabilité pour que le système fonctionne correctement au moins 10 000 heures.

    (b) Avec 3 composants, on réalise un deuxième montage. Dans ce cas, pour que le système ne remplisse plus sa fonction, il faut que les trois composants soient défectueux (montage "parallèle").
    - calculer la probabilité pour que le système fonctionne au moins 10 000 heures ;
    - le système ayant été laissé sous tension pendant 10 000 heures, on vient vérifier son fonctionnement : il est correct ; quelle est la probabilité de trouver au moins un composant défectueux ?

11 Solutions

1. On considère l'expérience aléatoire qui consiste à choisir successivement 4 numéros parmi 400 pour désigner les gagnants (schéma de tirage exhaustif).

   Soit un individu qui possède 10 billets. On note A_i ($i = 1, \ldots, 4$) l'événement : "le $i^{\text{ème}}$ numéro tiré n'est pas sur l'un des 10 billets.

   Soit E l'événement : "l'individu gagne au moins un lot". $\overline{E}$ est donc l'événement : "l'individu ne gagne aucun lot".

   On a $\mathbb{P}(E) = 1 - \mathbb{P}(\overline{E})$ (événements contraires).

   $E = A_1 \cap A_2 \cap A_3 \cap A_4$. D'où :

   $$\mathbb{P}(\overline{E}) = \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4)$$

   Les événements A_1, A_2, A_3, A_4 ne sont pas indépendants (tirage exhaustif). Donc :

   $$\begin{aligned} \mathbb{P}(\overline{E}) &= \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_2 \cap A_1)\mathbb{P}(A_4|A_3 \cap A_2 \cap A_1) \\ &= \frac{390}{400}\frac{389}{399}\frac{388}{398}\frac{387}{397} \\ &\approx 90.33\% \end{aligned}$$

   D'où $\mathbb{P}(E) \approx 9.67\%$

2. L'expérience consiste à faire entrer un rat dans le labyrinthe et à observer la sortie choisie. Si on admet que le rat, à chaque embranchement choisit indifféremment l'une des deux voies qui s'offrent à lui : G (pour gauche) et $\overline{G}$ (pour droite), on aura, à chaque embranchement : $\mathbb{P}(G) = \mathbb{P}(\overline{G})$ (absence de latéralisation et indépendance des choix successifs).

   Dans cette hypothèse, on a :

   $$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(\overline{G}) = \frac{1}{2} \\ \mathbb{P}(B) &= \mathbb{P}(G \cap C) = \mathbb{P}(G)\mathbb{P}(G) = \frac{1}{4} \\ \mathbb{P}(C) &= \mathbb{P}(G \cap \overline{G} \cap \overline{G}) = \frac{1}{8} \\ \mathbb{P}(D) &= \mathbb{P}(G \cap \overline{G} \cap G \cap \overline{G}) = \frac{1}{16} \\ \mathbb{P}(E) &= \mathbb{P}(G \cap \overline{G} \cap G \cap G \cap G) = \frac{1}{32} \\ \mathbb{P}(F) &= \mathbb{P}(G \cap \overline{G} \cap G \cap G \cap \overline{G}) = \frac{1}{32} \end{aligned}$$

   On peut vérifier que les événements A, B, C, D, E, F forment, mutuellement incompatibles, forment un système complet :

   $$\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) + \mathbb{P}(D) + \mathbb{P}(E) + \mathbb{P}(F) = 1$$

3. L'expérience est : les chasseurs A, B, C tirent. On note E l'événement "l'oiseau est touché" et $\overline{E}$ l'événement "l'oiseau n'est pas touché".
   On a : $\mathbb{P}(E) = 1 - \mathbb{P}(\overline{E})$ (événements contraires).
   Soit $\overline{A}$ l'événement "l'oiseau n'est pas touché par A".
   On a de même $\mathbb{P}(A) = 1 - \mathbb{P}(\overline{A}) = 1 - \frac{1}{2} = \frac{1}{2}$.
   Avec les mêmes notations, on a également $\mathbb{P}(\overline{B}) = 1 - \mathbb{P}(B) = \frac{1}{3}$ et $\mathbb{P}(\overline{C}) = 1 - \mathbb{P}(C) = \frac{3}{4}$.
   On peut alors écrire $\overline{E} = \overline{A} \cap \overline{B} \cap \overline{C}$ et donc $\mathbb{P}(\overline{E}) = \mathbb{P}(\overline{A})\mathbb{P}(\overline{B})\mathbb{P}(\overline{C}) = \frac{1}{2}\frac{1}{3}\frac{3}{4} = \frac{1}{8}$, en supposant les événements $\overline{A}, \overline{B}, \overline{C}$ indépendants. On en déduit donc que

$$\mathbb{P}(E) = \frac{7}{8}$$

4. L'expérience est le choix d'un individu, au hasard dans la population. On note :

   H l'événement : l'individu choisi est un homme
   F l'événement : l'individu choisi est une femme
   T l'événement : l'individu choisi fume

   On veut calculer la probabilité conditionnelle $\mathbb{P}(F|T)$.
   Par définition $\mathbb{P}(F|T) = \frac{\mathbb{P}(F \cap T)}{\mathbb{P}(T)}$
   - $\mathbb{P}(F \cap T) = \mathbb{P}(F)\mathbb{P}(T|F) = 0.6 \times 0.3 = 0.18$
   - $T = (H \cap T) \cup (F \cap T)$ et donc $\mathbb{P}(T) = \mathbb{P}(H \cap T) + \mathbb{P}(F \cap T)$ puisque les événements sont incompatibles.
   - $\mathbb{P}(T) = \mathbb{P}(H)\mathbb{P}(T|H) + \mathbb{P}(F)\mathbb{P}(T|F) = 0.4 \times 0.5 + 0.6 \times 0.3 = 0.2 + 0.18 = 0.38$

   Et donc finalement $\mathbb{P}(F|T) = \frac{0.18}{0.38} \approx 47.37\%$
5. n étudiants s'alignent à l'entrée du restaurant. Soit E l'événement "les deux amis sont séparés par r personnes". Pour que l'événement soit possible, il faut $n \geq r + 2$.
   Si on ne tient pas compte de l'ordre relatif des deux amis, il y a, a priori, $C_n^2 = \frac{n(n-1)}{2}$ couples de places possibles pour eux, tous équiprobables, donc chacun de probabilité $\frac{2}{n(n-1)}$.
   L'intervalle fermé délimité par les deux amis comprend $r + 2$ places. Il y autant de couples réalisant E que de manières de choisir la place la plus près de l'entrée, c'est-à-dire :

$$n - (r + 2) + 1 = n - r - 1$$

   Les différentes éventualités réalisant E étant mutuellement incompatibles, on a :

$$\mathbb{P}(E) = \frac{2(n - r - 1)}{n(n - 1)}$$

6. En ce qui concerne les jours de naissance, le groupe est considéré comme choisi au hasard dans une population. L'année comprend 365 jours et forme un ensemble A de 365 éléments ; on suppose a priori, hypothèse discutable, que chaque jour est également probable comme jour d'anniversaire d'un personne choisie au hasard.

L'événement F : 2 personnes du groupe, au moins, ont leur anniversaire le même jour", a pour événement contraire $\overline{F}$: "toutes les personnes du groupe ont des jours d'anniversaire différents".

On choisit successivement les 25 personnes. On note :
- A_k le jour d'anniversaire de la $k^{\text{ème}}$ personne, $k \in \{1, 2, \ldots, 25\}$;
- B_k l'ensemble $\{A_1, A_2, \ldots, A_k\}$ et $\overline{B_k}$ le complémentaire de B_k dans A ;
- E_k l'événement "les k premières personnes ont des anniversaires différents".

On note qu'alors B_k contient k éléments et $\overline{B_k}$ contient $(365 - k)$ éléments. On a successivement :

$$\mathbb{P}(E_2) = \mathbb{P}(A_2 \neq A_1)\mathbb{P}(A_2 \in \overline{B_1}) = \frac{364}{365}\left(1 - \frac{1}{365}\right)$$

$$\mathbb{P}(E_{k+1} = \mathbb{P}(E_k \cap A_{k+1} \in \overline{B_k}) = \mathbb{P}(E_k)\mathbb{P}(A_{k+1} \in \overline{B_k}|E_k) = \mathbb{P}(E_k)\left(1 - \frac{k}{365}\right)$$

D'où par récurrence :

$$\mathbb{P}(\overline{F}) = \mathbb{P}(E_{25}) = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\ldots\left(1 - \frac{24}{365}\right)$$

ce que l'on peut écrire :

$$\mathbb{P}(\overline{F}) = \frac{364}{365}\frac{363}{365}\ldots\frac{341}{365} = \frac{365!}{340!}\frac{1}{365^{25}}$$

Et donc

$$\mathbb{P}(F) = 1 - \frac{365!}{340!}\frac{1}{365^{25}}$$

Calcul numérique : $\mathbb{P}(F) \approx 1 - 0.4313 = 56.87\%$

7. On rencontre au par hasard un homme qui a deux enfants. En supposant qu'à chaque naissance la probabilité d'avoir un garçon G, égale à celle d'avoir une fille F est de ½, quatre éventualités sont à envisager : $G \cap G$, $G \cap F$, $F \cap G$, $F \cap F$. On en déduit :

$$\mathbb{P}(F \cap F|\overline{G \cap G}) = \frac{\mathbb{P}((F \cap F) \cap (\overline{G \cap G}))}{\mathbb{P}(\overline{G \cap G})} = \frac{\mathbb{P}(F \cap F)}{1 - \mathbb{P}(G \cap C)} = \frac{1/4}{1 - 1/4} = \frac{1}{3}$$

8. On considère d'abord l'événement le jet de 1 dé. On désigné par 6 l'événement "la face est six". et par $\overline{6}$ l'événement contraire : " la face n'est pas six". On a $\mathbb{P}(6) = \frac{1}{6}$ et $\mathbb{P}(\overline{6}) = \frac{5}{6}$.
   Pour quatre jets de 1 dé, on a :

$$\mathbb{P}(\text{au moins un } 6) = 1 - \mathbb{P}(\text{pas de 6 en 4 jets}) = 1 - \mathbb{P}(\overline{6} \cap \overline{6} \cap \overline{6} \cap \overline{6})$$

   et à cause de l'indépendance des événements successifs !

$$\mathbb{P}(\text{au moins un } 6) = 1 - \left(\frac{5}{4}\right)^6 \approx 51.77\%$$

Pour un jet de deux dés, on a $\mathbb{P}(\text{double } 6) = \mathbb{P}(6 \cap 6) = \frac{1}{36}$ D'où $\mathbb{P}(\overline{\text{double } 6}) = 1 - \frac{1}{36} = \frac{35}{36}$
Pour vingt quatre jets de 2 dés, on obtient, en raisonnant comme dans le cas précédent :

$$\mathbb{P}(\text{au moins un double } 6) 1 - \mathbb{P}(\overline{\text{double } 6}) = 1 - \left(\frac{35}{36}\right)^{24} \approx 49.14\%$$

N.B. : le chevalier de Méré pour qui les deux éventualités étaient égales, ne voulait pas admettre la solution de Pascal.

9. Il faut d'abord dénombrer les croisements possibles dans la génération F_0 et calculer leur probabilité :
Croisement I : ($H(AA) \cap F(AA)$**)**

$$\mathbb{P}(I) = \mathbb{P}(H(AA))\mathbb{P}(F(AA)) = u_0^2 = \frac{1}{9}$$

Croisement II :($(H(AA) \cap F(Aa)) \cup (H(Aa) \cap F(AA)$**)**

$$\mathbb{P}(II) = \mathbb{P}(H(AA))\mathbb{P}(F(Aa)) + \mathbb{P}(H(Aa))\mathbb{P}(F(AA)) = 2u_0v_0 = \frac{4}{9}$$

Croisement III :($H(Aa) \cap F(Aa)$**)**

$$\mathbb{P}(III) = \mathbb{P}(H(Aa))\mathbb{P}(F(Aa)) = v_0^2 = \frac{4}{9}$$

Les trois éventualités forment un système complet d'événements.
Il faut ensuite, pour chaque croisement, calculer la probabilité pour qu'un descendant est un génotype donné. Il est facile de trouver :

$$\mathbb{P}(AA|I) = 1 \;;\; \mathbb{P}(Aa|I) = 0 \;;\; \mathbb{P}(aa|I) = 0$$
$$\mathbb{P}(AA|II) = \frac{1}{2} \;;\; \mathbb{P}(Aa|II) = \frac{1}{2} \;;\; \mathbb{P}(aa|II) = 0$$
$$\mathbb{P}(AA|III) = \frac{1}{4} \;;\; \mathbb{P}(Aa|III) = \frac{1}{2} \;;\; \mathbb{P}(aa|III) = \frac{1}{4}$$

On a alors, en appliquant l'axiome des probabilités totales et la relation des probabilités composés :

$$\begin{aligned}
w_1 &= \mathbb{P}(aa) = \mathbb{P}(III \cap aa) = \mathbb{P}(III)\mathbb{P}(aa|III) = \frac{4}{9}\frac{1}{4} = \frac{1}{9} \\
v_1 &= \mathbb{P}(Aa) = \mathbb{P}(II \cap Aa) + \mathbb{P}(III \cap Aa) = \mathbb{P}(II)\mathbb{P}(Aa|II) + \mathbb{P}(III)\mathbb{P}(Aa|III) \\
&= \frac{4}{9}\frac{1}{2} + \frac{4}{9}\frac{1}{2} = \frac{4}{9} \\
u_1 &= \mathbb{P}(AA) = \mathbb{P}(I \cap AA) + \mathbb{P}(II \cap AA) + \mathbb{P}(III \cap AA) = \frac{1}{9} + \frac{4}{9 \times 2} + \frac{4}{9 \times 4} = \frac{4}{9}
\end{aligned}$$

On vérifie que les 3 éventualités forment un système complet.
Enfin, on a :

$$\mathbb{P}(AA|D) = \frac{\mathbb{P}(AA \cap D)}{\mathbb{P}(D)} = \frac{\mathbb{P}(AA)}{1 - w_1} = \frac{1}{2}$$

10. L'expérience consiste à choisir une graine au hasard dans le mélange. On

$$\mathbb{P}(A|N) = \frac{\mathbb{P}(A \cap N)}{\mathbb{P}(N)} = \frac{1/6}{4/6} = \frac{1}{4}$$

En effet :

$$\begin{aligned} \mathbb{P}(A \cap N) &= \mathbb{P}(A)\mathbb{P}(N|A) = \frac{1}{3}\frac{1}{2} = \frac{1}{6} \\ \mathbb{P}(B \cap N) &= \mathbb{P}(B)\mathbb{P}(N|B) = \frac{2}{3}\frac{3}{4} = \frac{1}{2} \\ \mathbb{P}(N) = &= \mathbb{P}(A \cap N) + \mathbb{P}(B \cap N) = \frac{4}{6} \end{aligned}$$

11. Réponse : $\frac{2}{5}$
12. Réponse : L'expérience consiste à choisir un jeton et à le poser sur une table ; on ne voit ainsi qu'une face, la deuxième est cachée. On a :

$$\mathbb{P}(B_2|B_1) = \frac{\mathbb{P}(B_2 \cap B_1)}{\mathbb{P}(B_1)} = \frac{\mathbb{P}(A \cup B}{B_1} = \frac{4/8}{5/8} = \frac{4}{5}$$

En effet $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = \frac{1}{2}$ (événement incompatibles et $\mathbb{P}(B_1) = \mathbb{P}(A \cup B \cup (C \cup B_1)) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) + \mathbb{P}(B_1|C) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} \times \frac{1}{2} = \frac{5}{8}$

13. (a) $5/1938$; $80/323$; $1/2$
    (b) $3/8$; $1/2$; $1/2$
14. On note a l'allèle correspondant à la coloration bleue et A l'allèle correspondant à la coloration non bleue ; A domine a. On aura besoin d'envisager deux types de croisement

Croisement de type I $H(Aa) \times F(Aa)$

On a 4 événements équiprobables pour le descendant : AA, Aa, aA, aa ;

Croisement de type II $H(Aa) \times F(AA)$

On a 4 événements équiprobables pour le descendant : AA, Aa, aA, AA

Le génotype de Monsieur Dupont est certainement Aa, puisqu'il n'a pas les yeux bleus mais qu'il peut transmettre a. Pour le génotype de Madame Dupont, qui n'a pas les yeux bleus, on a le choix entre deux éventualités : AA ou Aa ; comme elle est issue d'un croisement de type I et qu'elle n'a pas les yeux bleus, on en déduit, pour le croisement des Dupont :

$$\mathbb{P}(II) = \mathbb{P}(F(AA)|F(\text{non aa}))\frac{\mathbb{P}(F(AA \cap F(\text{non } aa)}{\mathbb{P}(F(\text{non } aa))} = \frac{\mathbb{P}(F(AA))}{\mathbb{P}(F(\text{non } aa))} = \frac{1/4}{3/4} = 1/3$$

$$\mathbb{P}(I) = \mathbb{P}(F(Aa)|F(\text{non aa}))\frac{\mathbb{P}(F(Aa \cap F(\text{non } aa)}{\mathbb{P}(F(\text{non } aa))} = \frac{\mathbb{P}(F(Aa))}{\mathbb{P}(F(\text{non } aa))} = \frac{2/4}{3/4} = 2/3$$

En effet si $E_1 \subset E_2 \Rightarrow E_1 \cap E_2 = E_1$

En ce qui concerne l'enfant des Dupont, on peut enfin écrire :

$$\mathbb{P}(AA|I)=\frac{1}{4}\ ;\ \mathbb{P}(Aa|I)=\frac{1}{2}\ ;\ \mathbb{P}(aa|I)=\frac{1}{4}$$
$$\mathbb{P}(AA|II)=\frac{1}{2}\ ;\ \mathbb{P}(Aa|II)=\frac{1}{2}\ ;\ \mathbb{P}(aa|II)=0$$

(a) L'expérience consiste à observer la couleur des yeux du premier enfant. On a : (yeux bleus)=(enfant aa)=($I\cap$ enfant aa)$\cup$($II\cap$ enfant aa) ; d'où, en appliquant l'axiome des probabilités totales et la relation des probabilités composées :

$$\mathbb{P}(\text{enfant } aa)=\mathbb{P}(I)\mathbb{P}(\text{enfant } aa|I)+\mathbb{P}(II)\mathbb{P}(\text{enfant } aa|II)=\frac{2}{3}\frac{1}{4}+\frac{1}{3}0=\frac{1}{6}$$

(b) Il s'agit de la même expérience. On a

$$\mathbb{P}(\text{enfant } Aa|\text{enfant non } aa)=\frac{\mathbb{P}(\text{enfant } Aa)}{\mathbb{P}(\text{enfant non} aa)}=\frac{3/6}{5/6}=\frac{3}{5}$$

(c) L'incertitude de Madame Dupont est levée. On est certain d'être dans le cas I. D'où

$$\mathbb{P}(2^e aa|1^e aa)=\mathbb{P}(2^e aa|I)=\frac{1}{4}$$

15. (a) On désigne par E_n le génotype d'une plante choisie au hasard dans la $^{\text{ème}}$ génération et par E_{n-1} son ascendant immédiat. Comme on passe d'une génération à l'autre par autofécondation, on a :

$$\mathbb{P}(E_n(AA)|E_{n-1}(AA))=1\ ;\ \mathbb{P}(E_n(AA)|E_{n-1}(Aa))=\frac{1}{4}\ ;\ \mathbb{P}(E_n(AA)|E_{n-1}(aa))=0$$
$$\mathbb{P}(E_n(Aa)|E_{n-1}(AA))=0\ ;\ \mathbb{P}(E_n(Aa)|E_{n-1}(Aa))=\frac{1}{2}\ ;\ \mathbb{P}(E_n(Aa)|E_{n-1}(aa))=0$$
$$\mathbb{P}(E_n(aa)|E_{n-1}(AA))=0\ ;\ \mathbb{P}(E_n(aa)|E_{n-1}(Aa))=\frac{1}{4}\ ;\ \mathbb{P}(E_n(aa)|E_{n-1}(aa))=1$$

On peut alors écrire :

$$\mathbb{P}(E_n(AA))=\mathbb{P}((E_{n-1}(AA)\cap E_n(AA))\cup(E_{n-1}(Aa)\cap E_n(AA))\cup(E_{n-1}(aa)\cap E_n(AA)))$$

En appliquant successivement l'axiome des probabilités totales et la relation des probabilités composées, on obtient :

$$\begin{aligned}x_n &= \mathbb{P}(E_{n-1}(AA))\mathbb{P}(E_n(AA)|E_{n-1}(AA))+\mathbb{P}(E_{n-1}(Aa))\mathbb{P}(E_n(AA)|E_{n-1}(Aa)) \\ &\quad +\mathbb{P}(E_{n-1}(aa))\mathbb{P}(E_n(AA)|E_{n-1}(aa))))\end{aligned}$$

En raisonnant de même pour y_n et z_n, on obtient :

$$\begin{aligned}x_n &= 1x_{n-1}+\frac{1}{4}y_{n-1}+0z_{n-1}\\ y_n &= 0x_{n-1}+\frac{1}{2}y_{n-1}+0z_{n-1}\\ z_n &= 0x_{n-1}+\frac{1}{4}y_{n-1}+1z_{n-1}\end{aligned}$$

(b) On part de $x_0 = z_0 = 0$, $y_0 = 0$ En utilisant les résultat précédent, on obtient :

$$x_1 = z_1 = \frac{1}{4} \quad , \quad y_1 = \frac{1}{2}$$
$$x_2 = z_2 = \frac{1}{4} + \frac{1}{8} = \frac{1}{4}\left(1 + \frac{1}{2}\right) \quad , \quad y_2 = \frac{1}{2^2}$$
$$x_3 = z_3 = \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{1}{4}\left(1 + \frac{1}{2} + \frac{1}{4}\right) \quad , \quad y_2 = \frac{1}{2^3}$$

(c) On suppose $x_{n-1} = z_{n-1} = \frac{1}{4}\left(1 + \frac{1}{2} + \ldots + \frac{1}{2^{n-2}}\right)$, $y_{n-1} = \frac{1}{2^{n-1}}$.
On déduit de la question précédente : $x_n = z_n = \frac{1}{4}\left(1 + \frac{1}{2} + \ldots + \frac{1}{2^{n-1}}\right)$, $y_n = \frac{1}{2^n}$. Par récurrence, à partir de $n = 2$, ces égalités sont bien vérifiées.
On reconnaît, entre parenthèses, la somme des n premiers termes d'une suite géométrique, de terme 1 et de raison $1/2$.

$$\frac{1}{4}\left(1 + \frac{1}{2} + \ldots + \frac{1}{2^{n-1}}\right) = \frac{1}{4}\frac{1 - \frac{1}{2^n}}{1 - \frac{1}{2}} = \frac{1}{2}\left(1 - \frac{1}{2^n}\right)$$

Finalement
$$x_n = z_n = \frac{1}{2}\left(1 - \frac{1}{2^n}\right) \text{ , } y_n = \frac{1}{2^n}$$

(d)

$$P_{n_1} = x_{n_1} + z_{n_1} = 1 - \frac{1}{2^{n_1}} \quad ; \quad P_{n_1} \geq 0.95 \Rightarrow \frac{1}{2^{n_1}} \leq 0.05 \Rightarrow n_1 \geq 5$$
$$P_{n_2} = x_{n_2} + z_{n_2} = 1 - \frac{1}{2^{n_2}} \quad ; \quad P_{n_2} \geq 0.999 \Rightarrow \frac{1}{2^{n_2}} \leq 0.05 \Rightarrow n_2 \geq 10$$

On remarque que $x_n + y_n + z_n = 1$, ce qui est normal puisque les événements AA, Aa et aa forment un système complet.

16. Réponse : $1/10!$; $1/10$; $1/90$; $17/90$
17. L'expérience consiste à choisir un individu au hasard dans la population. On a :

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \end{aligned}$$

ou encore $\mathbb{P}(A \cup B) = 1 - \mathbb{P}(\overline{A} \cap \overline{B}) = 1 - \mathbb{P}(\overline{A})\mathbb{P}(\overline{B})$.
Réponse : 5.95%

18. Montage série :
$$\mathbb{P}(E) = (0.8)^3 = 0.512$$

Montage parallèle :
$$\mathbb{P}(E) = 1 - (0.2)^3 = 0.992$$
$$\mathbb{P}(F|E) = \frac{3(0.2)(0.8)^2 + 3(0.2)^2(0.8)}{0.992} \approx 0.484$$

Chapitre 2

Variables aléatoires discrètes

1 Variable aléatoire à valeurs dans $\mathbb{R}$

Définition 2.1 *Soit un espace de probabilité* $\{\Omega, \mathcal{A}, \mathbb{P}\}$ *associé à une expérience aléatoire* $\mathcal{E}$. *On appelle variable variable aléatoire à valeurs dans* $\mathbb{R}$ *une application* X, *de* Ω *dans* $\mathbb{R}$, *possédant la propriété :*

$$\textit{Pour tout intervalle } I \subset \mathbb{R}, \quad \{\omega | X(\omega) \in I\} \in \mathcal{A} \tag{2.1}$$

Cette définition comporte trois points :

1. Tout d'abord une variable aléatoire permet d'associer à tout événement élémentaire ω un nombre réel $X(\omega)$; ceci permet de "traduire" une observation par un nombre réel. L'ensemble $X(\Omega)$ s'appelle l'ensemble des observables
2. La propriété 2.1 signifie que $\{\omega | X(\omega) \in I\}$ fait partie de la tribu $\mathcal{A}$. On désigne par "$X \in I$" cet événement que l'on exprime en langage courant par "la variable aléatoire X prend une valeur dans l'intervalle I".

   On sera amené, par la suite, à distinguer les types d'intervalle :
   - si $I = [a, b[$, "$X \in I$"="$a \leq X < b$" ;
   - si $I = [a, b]$, "$X \in I$"="$a \leq X \leq b$" ;
   - si $I = [a]$, "$X \in I$"="$X = a$", etc.

   Néanmoins, ce qui est essentiel c'est que l'introduction d'une variable aléatoire permet à tout intervalle, éventuellement réduit à un point, de représenter un événement.
3. Puisque l'épreuve $\{\Omega, \mathcal{A}\}$ est probabilisé, on a :

$$\mathbb{P}(\{\omega | X(\omega) \in I\}) = \mathbb{P}(X \in I)$$

   La définition d'une variable aléatoire permet d'attribuer une probabilité à tout intervalle I identifié à l'événement dont il est l'image.

1.1 *Loi de probabilité*

La loi de probabilité d'une variable aléatoire X, on dit aussi la distribution de X, est le mode de répartition de la probabilité de Ω sur l'ensemble des observables.

Il en résulte que connaître la loi de probabilité de X, c'est être capable d'attribuer une probabilité à tout événement "$X \in I$".

1.2 Fonction de répartition

Si on considère, en particulier, l'intervalle $]-\infty, t[$, qui définit l'événement $(X < t) = \{\omega | X(\omega) < t\}$, la définition de la variable aléatoire X implique qu'à tout réel t on peut associer la valeur $\mathbb{P}(X < t)$; il existe donc une fonction F_X, définie sur $\mathbb{R}$ et telle que $F_X(t) = \mathbb{P}(X < t)$:

Définition 2.2 *Étant donnée une variable aléatoire X, à valeurs dans $\mathbb{R}$, on appelle fonction de répartition la fonction :*

$$F_X : \left\{ \begin{array}{lll} \mathbb{R} & \to & [0,1] \\ t & \mapsto & F_X(t) = \mathbb{P}(X < t) \end{array} \right.$$

1.3 Propriétés de la fonction de répartition

1. La fonction de répartition d'une variable aléatoire à valeurs dans $\mathbb{R}$ est une fonction croissante au sens large.
   En effet, si $A = (X < a)$ et $B = (X < b)$, $a < b \Rightarrow A \subset B$. On en déduit $\mathbb{P}(A) \leq \mathbb{P}(B)$ (voir chapitre 1) c'est-à-dire $F_X(a) \leq F_X(b)$
2. $\lim_{t\to+\infty} F_X(t) = 1^-$, $\lim_{t\to-\infty} F_X(t) = 0^+$
   En effet
   $$\lim_{t\to+\infty} F_X(t) = \lim_{t\to+\infty} \mathbb{P}(X < t) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1^-$$
   par valeurs inférieures puisqu'une probabilité est un nombre inférieur ou égal à 1.
   $$\lim_{t\to-\infty} F_X(t) = \lim_{t\to-\infty} \mathbb{P}(X < t) = \mathbb{P}(\emptyset) = 0$$
   par valeurs supérieures puisqu'une probabilité est un nombre supérieur ou égal à 0
3. Si $a < b$, $F_X(b) - F_X(a) = \mathbb{P}(a \leq X < b)$.
   En effet si si $A = (X < a)$ et $B = (X < b)$, $a < b \Rightarrow A \subset B$.
   $A \in \mathcal{A}$ et $B \in \mathcal{A} \Rightarrow B \cap A \in \mathcal{A}$ et $B \cap \overline{A} \in \mathcal{A}$.
   On peut écrire $B = (B \cap A) \cup (B \cap \overline{A})$.
   D'où :
   $$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(B \cap A) + \mathbb{P}(B \cap \overline{A}) \quad \text{événements incompatibles} \\ &= \mathbb{P}(A) + \mathbb{P}(B \cap \overline{A}) \quad (A \subset B \Rightarrow B \cap A = A) \end{aligned}$$
   On en déduit $\mathbb{P}(B \cap \overline{A}) = \mathbb{P}(B) - \mathbb{P}(A)$.
   Or $(B \cap \overline{A}) = (X < b \cap \overline{X < a}) = (X < b \cap X \geq a) = (a \leq X < b)$
   Finalement : $\mathbb{P}(a \leq X < b) = \mathbb{P}(X < b) - \mathbb{P}(X < a) = F_X(b) - F_X(a)$.
   La fonction de répartition d'une variable aléatoire X permet de calculer $\mathbb{P}(X \in I)$ pour tout intervalle I, fermé à droite, ouvert à gauche.

4. La fonction de répartition d'une variable aléatoire à valeurs dans $\mathbb{R}$ est continue à gauche en tout point.
   Il résulte de la propriété 3 que pour tout t et pour tout ϵ positif, on a :
$$\mathbb{P}(t-\epsilon \leq X < t) = F_X(t) - F_X(t-\epsilon)$$
$$\Rightarrow \lim_{\epsilon \to 0^+} \mathbb{P}(t-\epsilon \leq X < t) = F_X(t) - \lim_{\epsilon \to 0^+} F_X(t-\epsilon)$$
   Si on admet $\lim_{\epsilon \to 0^+} \mathbb{P}(t-\epsilon \leq X < t) = \mathbb{P}(\emptyset) = 0$, alors
$$\lim_{\epsilon \to 0^+} F_X(t-\epsilon) = F_X(t)$$

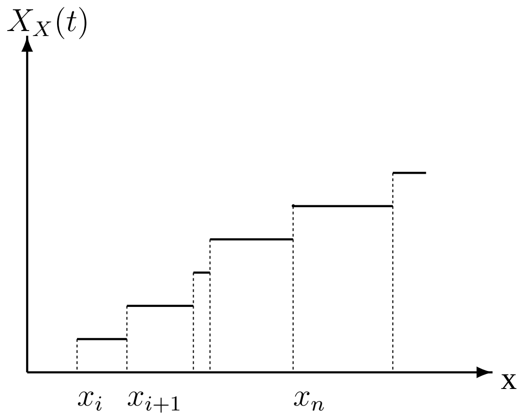
1.4 *Fonction de répartition et loi de probabilité*

Comme on va le montrer par la suite, la loi de probabilité d'une variable aléatoire, à valeurs dans $\mathbb{R}$, est parfaitement définie par sa fonction de répartition, le calcul de $\mathbb{P}(X = t)$ en tout point t combiné avec la propriété 3 permettra en effet d'ouvrir ou de fermer les intervalles à volonté. La fonction de répartition, concept commun à toutes les variables aléatoires, va jouer un rôle essentiel ; il convient donc d'en connaître parfaitement la définition et les propriétés.
C'est à partir de la fonction de répartition que l'on va différencier différents types de distribution ; pour chaque type on montrera qu'il existe une autre manière de définir la loi de probabilité.

2 Variables aléatoires discrètes

2.1 *Fonction de répartition en escalier*

Un premier type de distribution est caractérisé par un fonction de répartition F_X dite "en escalier" ; le graphe de F_X présente une suite croissante (finie ou non) de discontinuités : $\{x_1, x_2, \ldots, x_n, \ldots\}$ et est constitué de paliers successifs d'ordonnées croissantes compte tenu des propriétés de F_X.



En chaque point x_i de discontinuité, F_X est continue à gauche ($\lim_{\epsilon \to 0^+} F_X(x_i - \epsilon) = l_i = F_X(x_i)$) et admet un limite à droite ($\lim_{\epsilon \to 0^+} F_X(x_i + \epsilon) = l_{i+1} = F_X(x_{i+1})$). Partout ailleurs, F_X est continue.
Si on prend un point t et un point $t + \epsilon$ ($\epsilon > 0$) situés dans l'intervalle $]x_i, x_{i+1}[$, on a, puisque les points correspondants du graphe sont sur le même palier :

$$F_X(t + \epsilon) - F_X(t) = l_{i+1} - l_i = 0 = \mathbb{P}(t \leq X < t + \epsilon)$$

On en déduit

$$\mathbb{P}(X = t) = \lim_{\epsilon \to 0^+} \mathbb{P}(t \leq X < t + \epsilon) = 0$$

Toute valeur réelle distincte d'un point de discontinuité et tout intervalle ne comportant pas de point de discontinuité sont affectés d'une probabilité nulle.
Par contre, pour tout point x_i de discontinuité, on peut écrire, si $0 < \epsilon < x_{i+1} - x_i$:

$$F_X(x_i + \epsilon) - F_X(x_i) = l_{i+1} - l_i = \mathbb{P}(x_i \leq X < +\epsilon)$$

On en déduit

$$\mathbb{P}(X = x_i) = \lim_{\epsilon \to 0^+} \mathbb{P}(x_i \leq X < x_i + \epsilon) = l_{i+1} - l_i$$

La conclusion de cette analyse est que l'ensemble dénombrable $\{x_1, x_2, \ldots, x_n, \ldots\}$ regroupe les valeurs de probabilité non nulle et que la probabilité de Ω est répartie uniquement sur ces points. L'ensemble $\{x_1, x_2, \ldots, x_n, \ldots\}$ est donc l'ensemble des observables de X et la probabilité attribuée à chacun de ces points est la hauteur de la marche correspondante.
Réciproquement, si on connaît l'ensemble des observables $\{x_1, x_2, \ldots, x_n, \ldots\}$ et les probabilités attribuées à chacune d'elles, il est facile de retrouver la fonction de répartition et de déterminer la probabilité attribuée à tout intervalle :

- Si $a \in]x_i, x_{i+1}]$, $(X < a) = (X = x_1) \cup (X = x_2) \cup \cdots \cup (X = x_i)$. On en déduit, les événements étant incompatibles :

$$F_X(a) = \mathbb{P}(X < a) = \mathbb{P}(X = x_1) + \mathbb{P}(X = x_2) + \cdots + \mathbb{P}(X = x_i) = \sum_{k=1}^{i} \mathbb{P}(X = x_k)$$

- Si $a \in]x_i, x_{i+1}]$ et $b = x_{i+3}$, $(a \leq X \leq b) = (X = x_{i+1}) \cup (X = x_{i+2}) \cup (X = x_{i+3})$. On en déduit, les événements étant incompatibles :

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(X = x_{i+1}) + \mathbb{P}(X = x_{i+2}) + \mathbb{P}(X = x_{i+3})$$

Définition 2.3 *Une variable aléatoire X, à valeurs dans $\mathbb{R}$, dont la fonction de répartition F_X est en escalier, est dite variable aléatoire* discrète. *Les valeurs observables de X, qui correspondent aux discontinuités de F_X forment un ensemble dénombrable $\{x_1, x_2, \ldots, x_n, \ldots\}$. La loi de probabilité de X est définie par la loi de correspondance :*

$$\mathbb{P} : \left\{ \begin{array}{lcl} \{x_1, x_2, \ldots, x_n, \ldots\} & \to & [0, 1] \\ x_i & \mapsto & \mathbb{P}(X = x_i) = p_i \end{array} \right.$$

Dans la pratique, c'est par la loi $\mathbb{P}$ que l'on définit la loi de probabilité d'une variable aléatoire discrète. Il est clair que les probabilités p_i vérifient la relation $\sum_i p_i = 1$, puisque l'ensemble des événements $(X = x_i)$ forme une partition de Ω

Exemple 2.1 *La fonction de répartition d'une variable aléatoire est définie par :*

$$F : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 1 \\ t \mapsto \frac{2}{6} & , \text{ si } 1 < t \leq 2 \\ t \mapsto \frac{5}{6} & , \text{ si } 2 < t \leq 4 \\ t \mapsto 1 & , \text{ si } t > 4 \end{cases}$$

On en déduit la loi de probabilité de cette variable, définie par :

valeurs observables x_i	*1*	*2*	*4*
probabilités p_i	2⁄6	3⁄6	1⁄6

Exemple 2.2 *On tire une carte d'un jeu de 32 cartes. Soit X la variable aléatoire qui associe à la carte tirée une valeur numérique suivant la règle du jeu de bridge : 4 pour un as, 3 pour un roi, 2 pour une dame, 1 pour un valet et 0 pour une autre carte.*
La loi de probabilité de X est définie par :

valeurs observables x_i	*0*	*1*	*2*	*3*	*4*
probabilités p_i	$\frac{16}{32}$	$\frac{4}{32}$	$\frac{4}{32}$	$\frac{4}{32}$	$\frac{4}{32}$

ou par la fonction de répartition F_X :

t	$t \leq 0$	$0 < t \leq 1$	$1 < t \leq 2$	$2 < t \leq 3$	$3 < t \leq 4$	$t > 4$
$F_X(t)$	0	4⁄8	5⁄8	6⁄8	7⁄8	1

2.2 Représentation graphique : l'histogramme

Considérons un ensemble de n données :

$$(x_1, x_2, \ldots, x_i, \ldots, x_n) = (x_i | 1 \leq i \leq n)$$

Le jeu de données peut constituer un échantillon issu de l'observation d'une distribution aléatoire ou bien un répertoire exhaustif de n valeurs enregistrées lors d'une opération de compilation.
Parrallèlement, soit une partition en m intervalles de I, un intervalle de $\mathbb{R}$ contenant l'ensemble du jeu de données :

$$(I_1 = [b_1, b_2[, \ldots, I_j = [b_j, b_{j+1}], \ldots, I_m = [b_m, b_{m+1}]) = (I_j | 1 \leq j \leq m)$$

$I_1, \ldots, I_m$ étant une partition, nous avons :

$$\bigcup_j I_j = I \text{ et } \forall (j,k) I_j \cap I_k = \emptyset$$

On parle de la classe j pour désigner les valeurs de $(x_i | 1 \leq \ldots \leq n)$ appartenant à I_j.
La classe j est caractérisée par :

- son amplitude $a_j = b_{j+1} - b_j$;
- n_j, le nombre d'observations x_i appartenant à la classe j. On parle aussi de fréquence absolue. $\sum_j n_j = n$;
- $f_j = \frac{n_j}{n}$, la fréquence relative. $\sum_j f_j = 1$.

L'histogramme est une représentation dans un diagramme cartésien qui associe les fréquences f_j aux classes j. De manière pratique, l'histogramme est obtenu en juxtaposant des rectangles dont la base est égale à l'amplitude de la classe et dont la surface est égale à la fréquence de la classe.
Si l'amplitude de toutes les classes est égale, la hauteur des rectangles correspond à la fréquence de chacune des classes et la largeur des rectangles correspond à l'amplitude des classes.
Dans le cas de classes d'amplitudes différentes, la hauteur des rectangles vaut $\frac{f_i}{a_i}$. Cette correction s'explique par la volonté de proposer une représentation en rapport avec l'interprétation visuelle, laquelle est sensible à la surface s_j du rectangle : $s_j = \frac{f_j}{a_j} a_j$ et donc $\sum_j s_j = 1$.
Il existe donc autant d'histogrammes que de partitions de I. Ainsi le choix de la partition conduit à une représentation plus ou moins informative.

3 Paramètres descriptifs d'une distribution discrète

Soit X une variable aléatoire discrète dont la loi de probabilité est définie par

$$\mathbb{P} : \left\{ \begin{array}{lcl} \{x_1, x_2, \ldots, x_n, \ldots\} & \rightarrow & [0, 1] \\ x_i & \mapsto & \mathbb{P}(X = x_i) = p_i \end{array} \right.$$

Dans le cadre de ce cours, nous utiliserons les paramètres descriptifs qui suivent ; l'exposé n'est pas exhaustif.

Mode

On appelle mode de la variable X, ou de la distribution de X, une valeur x pour laquelle $\mathbb{P}(X = x)$ présente un maximum.
Les distributions discrètes classiques présentent en général un seul mode (parfois deux modes successifs équiprobables) et sont dites unimodale, par opposition aux distributions à plusieurs "bosses" dites plurimodales.
Dans l'exemple 2.1, le mode de X est la valeur 2 ; dans l'exemple 2.2, le mode de X est la valeur 0

Espérance

On appelle espérance de X, et on note μ_X ou $\mathbb{E}(X)$, le nombre réel, s'il existe, défini par

$$\mu_X = \mathbb{E}(X) = \sum_i (x_i \mathbb{P}(X = x_i)) = \sum_i (x_i p_i)$$

La réserve ”s'il existe” concerne le cas où l'ensemble des observables contient un nombre infini de valeurs ; l'existence de $\mathbb{E}(X)$ dépend alors de la convergence de la somme.
Analogie mécanique : imaginons une tige sans masse sur laquelle on place des masses ”ponctuelles” : à l'abscisse x_i se trouve la masse m_i. L'abscisse x du centre de masse de ces points est donné par $(\sum_i m_i)x = \sum_i(m_i x_i)$. L'espérance de X définit donc le barycentre de points matériels x_i de masse p_i, de masse totale 1, situés sur un axe.
On peut également donner une autre définition de $\mathbb{E}(X)$: À tout événement élémentaire ω, de probabilité $\mathbb{P}(\omega)$, correspond une image $X(\omega)$. On peut écrire :

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega)$$

En effet, si on regroupe dans cette dernière somme les événements élémentaires de même image, on obtient :

$$\mathbb{E}(X) = \sum_i \left(\sum_{\omega | X(\omega)=x_i} \mathbb{P}(\omega) \right) = \sum_i x_i \mathbb{P}(X = x_i) = \sum_i x_i p_i$$

On dit que le mode et l'espérance, qui représentent une certaine tendance centrale, sont des paramètres de position.

Moments d'ordre k

On appelle moment d'ordre k ($k \in \mathbb{N}$) de X, et on note $\mathbb{E}(X^k)$, le nombre réel, s'il existe, défini par : $\mathbb{E}(X^k) = \sum_i [x_i^k \mathbb{P}(X = x_i)] = \sum_i x_i^k p_i$ On a

$$\mathbb{E}(X^0) = \sum_i p_i = 1$$
$$\mathbb{E}(X^1) = \sum_i x_i p_i = \mathbb{E}(X)$$

On utilisera $\mathbb{E}(X^2) = \sum_i x_i^2 p_i$

Variance

On appelle variance de X, et on note $\mathbb{V}(X)$ ou σ_X^2, le nombre réel, s'il existe, défini par :

$$\mathbb{V}(X) = \sigma_X = \sum_i (x_i - \mu_X)^2 p_i$$

On appelle également la variance de X le moment centré d'ordre 2 de X et on note :

$$\mathbb{E}\left((X - \mu_X)^2\right) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$$

On notera qu'une variance, somme de termes positifs, ne peut avoir qu'une valeur positive. Une valeur élevée de la variance signifie que les valeurs éloignées de l'espérance ont

une forte probabilité ; on dit que la variance est un paramètre de dispersion. Une valeur nulle de la variance signifie qu'il n'existe qu'une seule valeur observable de valeur égale à l'espérance (dispersion minimum).
L'analogie mécanique de la variance est le moment d'inertie de masses ponctuelles par rapport à leur centre de masse.
On peut écrire : $\mathbb{V}(X) = \sum_i (x_i^2 - 2\mu_X x_i + \mu_X^2)p_i$
D'où $\mathbb{V}(X) = \sum_i (x_i^2 p_i) - 2\mu_X \sum_i (x_i p_i) + \mu_X^2 \sum_i p_i$
Or $\sum_i p_i = 1$ et $\sum_i p_i x_i = \mu_x$ donc $\mathbb{V}(X) = \sum_i (x_i^2 p_i) - \mu_X^2$
D'où la relation (théorème de Guldin) :

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

Cette relation est très commode dans le calcul de la variance.

Écart-type

On appelle écart-type de X la racine carré de sa variance. On note σ_X.
Un écart-type est donc un réel positif ; il s'agit d'un indice de dispersion qui présente l'avantage d'avoir la même unité que les observables.

Exemple 2.3 *Reprenons l'exemple 2.1 On a*

$$\begin{aligned}
\mathbb{E}(X) &= \frac{2}{6}1 + \frac{3}{6}2 + \frac{1}{6}4 = 2 \\
\mathbb{E}(X^2) &= \frac{2}{6}1 + \frac{3}{6}4 + \frac{1}{6}16 = 5 \\
V(X) &= 5 - 4 = 1 \\
\sigma_X &= 1
\end{aligned}$$

Exemple 2.4 *Reprenons l'exemple 2.2 On a :*

$$\begin{aligned}
\mathbb{E}(X) &= \frac{16}{32}0 + \frac{4}{32}1 + \frac{4}{32}2 + \frac{4}{32}3 + \frac{4}{32}4 = \frac{40}{32} = \frac{5}{4} \\
\mathbb{E}(X^2) &= \frac{16}{32}0 + \frac{4}{32}1 + \frac{4}{32}4 + \frac{4}{32}9 + \frac{4}{32}16 = \frac{120}{32} = \frac{60}{16} \\
V(X) &= \frac{60}{16} - \frac{25}{16} = \frac{35}{16} \\
\sigma_X &\approx 1.479
\end{aligned}$$

On peut également calculer $\mathbb{V}(X)$ en utilisant la définition de $\mathbb{V}(X)$:

$$\begin{aligned}
\mathbb{V}(X) &= \frac{16}{32}\left(0 - \frac{5}{4}\right)^2 + \frac{4}{32}\left(1 - \frac{5}{4}\right)^2 + \frac{4}{32}\left(2 - \frac{5}{4}\right)^2 + \frac{4}{32}\left(3 - \frac{5}{4}\right)^2 + \frac{4}{32}\left(4 - \frac{5}{4}\right)^2 \\
&= \frac{1}{32}\left(25 + \frac{1}{4} + \frac{9}{4} + \frac{49}{4} + \frac{121}{4}\right) \\
&= \frac{35}{16}
\end{aligned}$$

4 Variable indicatrice

Définition 2.4 *Soit une expérience aléatoire $\mathcal{E}$ pouvant entraîner l'observation d'un élément E ou de son contraire $\overline{E}$. On appelle variable indicatrice de l'événement E une variable aléatoire X qui associe à E la valeur un et à $\overline{E}$ la valeur zéro.*

4.1 Loi de probabilité

On pose $\mathbb{P}(E) = p$ et $\mathbb{P}(\overline{E}) = q = 1 - p$. On a

$$\mathbb{P}(X = 1) = \mathbb{P}(E) = p$$

et

$$\mathbb{P}(X = 0) = \mathbb{P}(\overline{E}) = q = 1 - p$$

La loi de X est donc définie par :

k	0	1
$\mathbb{P}(X = k)$	$1 - p$	p

4.2 Paramètres descriptifs

$$\begin{aligned}
\mathbb{E}(X) &= 0(1-p) + 1p = p \\
\mathbb{E}(X^2) &= 0(1-p) + 1p = p \\
\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) = p - p^2 = p(1-p) = pq
\end{aligned}$$

Le mode est 1 si $p > \frac{1}{2}$, et 0 si $p < \frac{1}{2}$. Il n'y a pas de mode si $p = \frac{1}{2}$

5 Loi discrète uniforme

5.1 Définition

Soit une expérience aléatoire $\mathcal{E}$ à laquelle est associée une variable aléatoire X dont l'ensemble des observables contient un nombre fini de valeurs réelles : $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$. La loi de probabilité de X est dite uniforme si elle est définie par

$$\mathbb{P}(X = x_i) = \frac{1}{n} \quad \forall i \in \{1, 2, \ldots, n\}$$

Exemple 2.5 *Dans le jet d'un dé régulier, si X est la variable aléatoire qui associe à la face obtenue le nombre de points qu'elle représente, la loi de X est définie par :*

k	*1*	*2*	*3*	*4*	*5*	*6*
$\mathbb{P}(X = k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

La loi de X est uniforme sur $\{1, 2, 3, 4, 5, 6\}$

5.2 Paramètres descriptifs

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i) = \frac{1}{n}\sum_{i=1}^{n} x_i \\
\mathbb{E}(X^2) &= \sum_{i=1}^{n} x_i^2 \mathbb{P}(X = x_i) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \\
\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X)
\end{aligned}
$$

Une distribution uniforme ne présente, évidemment, pas de mode.
Un cas particulier intéressant est celui où l'ensemble des observables est constitué des n premiers entiers. On obtient alors :

$$
\begin{aligned}
\mathbb{E}(X) &= \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2} \\
\mathbb{E}(X^2) &= \frac{1}{n}\sum_{i=1}^{n} i^2 = \frac{1}{n}\frac{n(n+½)(n+1)}{3} = \frac{(2n+1)(n+1)}{6} \\
\mathbb{V}(X) &= \frac{(2n+1)(n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}
\end{aligned}
$$

6 La loi hypergéométrique

6.1 Schéma du tirage exhaustif ou sans remise

Une urne contient N boules dont a sont rouges et $N - a$ "non-rouges".
On considère l'expérience $\mathcal{E}$ qui consiste à prélever, sans remise, un échantillon de n boules de l'urne ($n \leq N$). Si X est la variable aléatoire associant à un tel échantillon le nombre de boules rouges qu'il peut contenir, la loi de probabilité de X est définie par : $\mathbb{P}(X = k) = \frac{C_a^k C_{N-a}^{n-k}}{C_N^n}$ Un tel schéma est utilisé dans les sondages.

Définition 2.5 *On dit qu'une variables aléatoire X, à valeurs dans $\mathbb{N}$, suit une loi hypergéométrique si la loi de probabilité est définie par :*

$$\mathbb{P}(X = k) = \frac{C_a^k C_{N-a}^{n-k}}{C_N^n}$$

où k, n, a, N sont des entiers positifs tels que $0 \leq n \leq N$, $0 \leq k \leq n$, $0 \leq k \leq a$, $0 \leq n - k \leq N - a$.

6.2 Paramètres descriptifs

On notera, à titre d'information, les résultats suivants :

$$\mathbb{E}(X) = \frac{a}{N}n\,; \quad \mathbb{V}(X) = n\frac{a}{N}\frac{N-a}{N}\frac{N-n}{N-1}$$

7 La loi binomiale

7.1 Schéma de Bernoulli

Soit une suite finie de n expériences aléatoires $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_n$, obéissant aux conditions suivantes :

1. chaque expérience peut entraîner l'observation d'un événement E ou de son contraire $\overline{E}$;
2. la probabilité de E, notée p, est la même pour chaque expérience ; ceci est alors également vraie pour la probabilité de $\overline{E}$, notée $q = 1 - p$;
3. le résultat d'une expérience est indépendant des résultats des autres expériences.

On note E_k l'événement "E se réalise à la $k^{\text{ème}}$ expérience" et A_k l'événement "E se réalise exactement k fois dans la suite d'expériences".
L'événement A_k peut se réaliser de plusieurs manières mutuellement incompatibles. La probabilité de A_k est donc la somme des probabilités de chacune de ces éventualités. L'une d'elles est, par exemple : "$E_1 \cap \cdots \cap E_k \cap \overline{E_{k+1}} \cap \cdots \cap \overline{E_n}$; sa probabilité est $p^k q^{n-k}$, à cause de l'indépendance des événements. Toute autre éventualité réalisant A_k aura la même probabilité, obtenue par le produit de k termes égaux à p et de $n-k$ termes égaux à q.
Pour obtenir la probabilité de A_k, il suffit donc de dénombrer les éventualités qui réalisent A_k. Il est clair qu'il y en a autant que de manières de choisir, parmi les n expériences, celles qui réalisent E, c'est-à-dire C_n^k et on écrit $\mathbb{P}(A_k) = C_n^k p^k q^{n-k}$.
Ce schéma, dit de Bernoulli, s'applique par exemple, à une suite de jets d'une pièce de monnaie (E=pile) ou un tirage avec remise, ou non exhaustif, de n boules dans une urne à deux catégories (E=boule tirée est rouge).
Si on associe à une suite d'expérience de Bernoulli la variable aléatoire représentant le nombre d'événements E que l'on peut observer, l'événement A_k s'écrit "$X = k$" et on a : $\mathbb{P}(X = k) = C_n^k p^k q^{n-k}$.

Définition 2.6 *On dit qu'une variable aléatoire X, à valeurs dans $\mathbb{N}$, suit une loi binomiale si sa loi de probabilité est donnée par :*

$$\mathbb{P}(X = k) = C_n^k p^k q^{n-k}$$

avec $k \in \{0, 1, 2, \dots, n\}$, où n est un entier donné et où p est un réel tel que $0 < p < 1$.

Le nom de cette loi provient du fait que $\mathbb{P}(X = k)$ est donné par le terme de rang $k+1$ du développement, suivant les puissances croissantes de p, du binôme de Newton : $(p+q)^n$.
On vérifie immédiatement que $\sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^n = 1$.

Une loi binomiale étant parfaitement définie par les paramètres n et p, on écrit symboliquement : X suit une loi $\mathcal{B}(n,p)$.

Exemple 2.6

1. *Si X est la variable aléatoire qui représente le nombre de "piles" que l'on peut obtenir en jetant 10 fois une pièce régulière, X suit une loi $\mathcal{B}\left(10,\frac{1}{2}\right)$. On a :*

k	0	1	2	3	4	5	6	7	8	9	10
$\mathbb{P}(X=k)$	$\frac{1}{1024}$	$\frac{10}{1024}$	$\frac{45}{1024}$	$\frac{120}{1024}$	$\frac{210}{1024}$	$\frac{252}{1024}$	$\frac{210}{1024}$	$\frac{120}{1024}$	$\frac{45}{1024}$	$\frac{10}{1024}$	$\frac{1}{1024}$

2. *Une urne contient des boules rouges en proportion ¼. On tire successivement 25 boules en remettant chaque fois la boule tirée. Si Y est le nombre de boules rouges que l'on peut obtenir, Y suit un loi $\mathcal{B}\left(25,\frac{1}{4}\right)$ (tirage non exhaustif ou avec remise).*
3. *La probabilité pour qu'un individu, issu d'un certain croisement, possède le phénotype R est de ¾. Si Z est la variable aléatoire représentant le nombre d'individus portant le phénotype R, que l'on peut observer sur 50 descendants issus d'un tel croisement, Z suit une loi $\mathcal{B}\left(50,\frac{3}{4}\right)$*

7.2 *Paramètres descriptifs*

On admettra les résultats qui suivent ; il est recommandé d'en connaître une démonstration (voir par exemple le corrigé de l'exercice 11)
Si X est une variable aléatoire de loi $\mathcal{B}(n,p)$,
- L'espérance de X est donné par $\mu_X = \mathbb{E}(X) = np$
- La variance de X est donnée par $\sigma_X^2 = \mathbb{V}(X) = np(1-p) = npq$
- la distribution est de type unimodal et admet pour mode l'entier, en général unique, situé dans l'intervalle $[(n+1)p-1,(n+1)p]$ ou exceptionnellement deux modes successifs, équiprobables, bornes de l'intervalle ci-dessus lorsque $(n+1)p$ a une valeur entière.

Exemple 2.7

1. *Si X suit une loi $\mathcal{B}\left(10,\frac{1}{2}\right)$, on a*

$$\mathbb{E}(X) = 10\frac{1}{2} = 5$$

Vérifions :

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{10} kC_n^k p^k q^{n-k} \\
&= \frac{1}{1024}(0+10+90+360+840+1260+1260+840+360+90+10) \\
&= \frac{5120}{1024} \\
&= 5
\end{aligned}$$

$$\mathbb{V}(X) = 10\frac{1}{2}\frac{1}{2} = 2.5$$

Vérifions :

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=0}^{10} k^2 C_n^k p^k q^{n-k} \\
&= \frac{1}{1024}(0 + 10 + 180 + 1080 + 3360 + 6300 + 7560 + 5880 + 2880 + 910 + 100) \\
&= \frac{28160}{1024} \\
&= 27.5
\end{aligned}$$

$$\Rightarrow \mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = 2.5$$

Le mode est 5, entier de l'intervalle $[4.5, 5.5]$

2. *Si Y suit une loi $\mathcal{B}\left(25, \frac{1}{4}\right)$, on a*

$$\mathbb{E}(X) = \frac{25}{4}, \quad \mathbb{V}(X) = \frac{75}{16}, \quad \textit{le mode est } 6 \in [5.5, 6.5]$$

3. *Si Z suit une loi $\mathcal{B}\left(50, \frac{3}{4}\right)$, on a*

$$\mathbb{E}(X) = \frac{150}{4}, \quad \mathbb{V}(X) = \frac{150}{16}, \quad \textit{le mode est } 38 \in [37.25, 38.25]$$

7.3 *Formes de la distribution*

La forme d'une distribution binomiale se déduit aisément de l'étude du mode :

1. si $p = \frac{1}{2} = q$, forme en cloche symétrique.
   En effet $\mathbb{P}(X = k) = C_n^k \left(\frac{1}{2}\right)^k = \mathbb{P}(X = n - k)$. Car $C_n^k = C_n^{n-k}$. Deux modes successifs équiprobables pour n impair
2. Si $\frac{1}{n+1} < p < \frac{1}{2}$, forme en cloche dissymétrique, le mode étant déplacé vers la gauche. Éventuellement, deux modes successifs équiprobables.
3. Si $p \leq \frac{1}{n+1}$, forme en L ; mode=0, éventuellement deux modes 0 et 1.
4. Si $\frac{1}{2} < p < \frac{n}{n+1}$, forme en cloche dissymétrique, le mode étant déplacé vers la droite. Éventuellement, deux modes successifs équiprobables.
5. Si $p \geq \frac{n}{n+1}$, forme en J ; Éventuellement, deux modes en $n - 1$ et n.

Comme on le verra plus loin, il est très important, en statistique, de savoir déterminer la forme d'une distribution.

Exemple 2.8

1. *Si X suit une loi $\mathcal{B}\left(10, \frac{1}{2}\right)$,la distribution est en cloche symétrique par rapport au mode 5, valeur égale aussi à l'espérance de X.*

2. *Si Y suit une loi $\mathcal{B}\left(25, \frac{1}{4}\right)$, la distribution est en cloche dissymétrique ; le mode 6 est plus proche de 0 que de 25.*
3. *Si X suit une loi $\mathcal{B}\left(50, \frac{3}{4}\right)$,la distribution est en cloche dissymétrique ; le mode 38 est plus proche de 50 que de 0.*

7.4 *Approximation d'une loi hypergéométrique*

Soit X une variable aléatoire de loi hypergéométrique. On a :

$$\begin{aligned}
\mathbb{P}(X=k) &= \frac{C_a^k C_{N-a}^{n-k}}{C_N^n} \\
&= \frac{a!}{k!(a-k)!}\frac{(N-a)!}{(n-k)!(N-a-n+k)!}\frac{n!(N-n)!}{N!} \\
&= \frac{n!}{k!(a-k)!}\frac{a(a-1)\dots(a-k+1)}{N(N-1)\dots(N-k+1)}\frac{(N-a)(N-a-1)\dots(N-a-n+k+1)}{(N-k)(N-k-1)\dots(N-n+1)} \\
&= C_n^k \frac{a}{N}\frac{a-1}{N-1}\dots\frac{a-k+1}{N-k+1}\frac{N-a}{N-k}\frac{N-a-1}{N-k-1}\dots\frac{N-a-n+k+1}{N-n+1}
\end{aligned}$$

Si on suppose maintenant que $N \longrightarrow +\infty$ et que $\frac{a}{N} \longrightarrow p$, on a :

- $\frac{a}{N} \longrightarrow p$, $\frac{a-1}{N-1} = \frac{\frac{a}{N}-\frac{1}{N}}{1-\frac{1}{N}} \longrightarrow p, \dots, \frac{a-k+1}{N-k+1} = \frac{\frac{a}{N}-\frac{k-1}{N}}{1-\frac{k-1}{N}} \longrightarrow p$; on dénombre k fractions de ce type.
- $\frac{N-a}{N-k} = \frac{1-\frac{a}{N}}{1-\frac{k}{N}} \longrightarrow 1-p, \dots, \frac{N-a-n+k+1}{N-n+1} = \frac{1-\frac{a}{N}-\frac{n-k-1}{N}}{1-\frac{n-1}{N}} \longrightarrow 1-p$; on dénombre $n-k$ fractions de ce type.

En conclusion : si $N \longrightarrow +\infty$ et si $\frac{a}{N} \longrightarrow p$, on a :

$$\frac{C_a^k C_{N-a}^{n-k}}{C_N^n} \longrightarrow C_n^k p^k (1-p)^{n-k}$$

Ce résultat mathématique peut être utilisé de la manière suivante :
La variable X représente le nombre d'individus d'un type donné (de couleur rouge pour une boule) que l'on peut trouver dans un échantillon de taille n extrait d'une population de taille N (l'urne) contenant a individus du type considéré. L'échantillon peut être obtenu par n tirages successifs sans remise. Si la taille de la population est très grande devant celle de l'individu ($n \ll N$), de manière à ce que l'on puisse considérer qu'à chaque tirage la probabilité d'obtenir un individu du type considéré demeure pratiquement égale à celle du premier tirage ($\frac{a}{N}$), la loi de X peut être assimilée à une loi binomiale $\mathcal{B}\left(n, p=\frac{a}{N}\right)$. On remarquera, en particulier, que cette approximation conserve l'espérance :

$$\mathbb{E}(X) = n\frac{a}{N} = np$$

et, pratiquement, la variance :

$$\mathbb{V}(X) = n\frac{a}{N}\left(1-\frac{a}{N}\right)\left(\frac{1-\frac{n}{N}}{1-\frac{1}{N}}\right) \approx np(1-p)$$

car $\left(\frac{1-\frac{n}{N}}{1-\frac{1}{N}}\right) \approx 1$

8 La loi de Poisson

Définition 2.7 *On dit qu'une variable aléatoire X, à valeurs dans $\mathbb{N}$, suit une loi de Poisson de paramètre λ si, λ étant un réel donné strictement positif, la loi de X est définie par :*

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{où } k \in \mathbb{N}$$

Une loi de Poisson étant parfaitement définie par le paramètre λ, on écrit alors : X suit une loi $\mathcal{P}(\lambda)$. On note que :

$$\sum_{k=0}^{+\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1$$

8.1 *Calcul pratique des probabilités*

Les différentes probabilités se calculent aisément grâce à la formule de récurrence :

$$\mathbb{P}(X = k+1) = \mathbb{P}(X = k)\frac{\lambda}{k+1}$$

Si, par exemple, on suppose que X suit une loi $\mathcal{P}(2,2)$, on a :

$$\begin{aligned}
\mathbb{P}(X = 0) &= e^{-2.2} \approx 0.110803 \\
\mathbb{P}(X = 1) &= \mathbb{P}(X = 0)\frac{2.2}{1} \approx 0.243767 \\
\mathbb{P}(X = 2) &= \mathbb{P}(X = 1)\frac{2.2}{2} \approx 0.268144 \\
\mathbb{P}(X = 3) &= \mathbb{P}(X = 2)\frac{2.2}{3} \approx 0.196639 \\
& \text{etc.}
\end{aligned}$$

On trouvera dans la table 11.4 les valeurs de $\mathbb{P}(X = k)$, avec quatre chiffres significatifs, pour différentes valeurs de λ. On peut, pour une valeur de λ non tabulée, procéder par interpolation linéaire.
Si, par exemple, X suit une loi $\mathcal{P}(2,2)$, on obtient :

$$\left.\begin{array}{l}\text{Pour } \lambda = 2 \quad \mathbb{P}(X = 1) \approx 0.2707 \\ \text{Pour } \lambda = 2.5\, \mathbb{P}(X = 1) \approx 0.2052\end{array}\right\} \text{Pour } \lambda = 2.2\, \mathbb{P}(X = 1) \approx 0.2702 - \frac{2}{5}(0.2707 - 0.2052)$$

D'où $\mathbb{P}(X = 1) \approx 0.2445$; la comparaison avec la valeur trouvée plus haut donné une idée de la précision du procédé.
On notera que la somme des probabilités des observables concentrées autour du mode est pratiquement égale à un. Ceci explique que les probabilités affectées aux observables extérieures à cette zone ne sont pas tabulées ; il convient cependant de préciser que non seulement chacune de ces probabilités est négligeable mais qu'il en est également de même de la somme de ces probabilité (attention, cette dernière remarque n'est pas triviale : cette somme comprend un nombre infini de termes !).

8.2 Paramètres descriptifs

On admettra les résultats qui suivent ; il est recommandé d'en connaître une démonstration (voir par exemple le corrigé de l'exercice 10)
Si X est une variable aléatoire de loi $\mathcal{P}(\lambda)$,
- L'espérance de X est donné par $\mu_X = \mathbb{E}(X) = \lambda$
- La variance de X est donnée par $\sigma_X^2 = \mathbb{V}(X) = \lambda$
- la distribution est de type unimodal et admet pour mode l'entier, en général unique, situé dans l'intervalle $[\lambda - 1, \lambda]$ ou exceptionnellement deux modes successifs, équiprobables, bornes de l'intervalle ci-dessus lorsque λ a une valeur entière. La position du mode se déduit aisément de la formule de récurrence donnée dans le paragraphe précédent.

8.3 Forme de la distribution

La forme d'une distribution de Poisson se déduit de la valeur du mode :
- Forme en L si $\lambda \leq 1$;
- forme en cloche dissymétrique pour $\lambda > 1$. Le mode se déplace vers la droite lorsque λ augmente.

8.4 Processus de Poisson

L'étude de certains phénomènes aléatoires, qui sont dits suivre un processus de Poisson (voir exercice 28) permet d'introduire directement la loi de Poisson. C'est ainsi que le nombre d'impulsions que l'on peut enregistrer par unité de temps, lors de la mesure d'un rayonnement par un compteur, est une variable aléatoire qui suit une loi de Poisson ; l'espérance de cette loi, qui caractérise le rayonnement, est le paramètre qu'il convient de connaître.

Exemple 2.9 *Le mouvement propre d'un compteur, mesuré pendant 30 secondes, est une variable aléatoire X qui suit une loi $\mathcal{P}(10)$. Quelle est la probabilité d'observer un nombre d'impulsions compris dans l'intervalle $[8; 12]$ lors d'un comptage de 30 secondes ?*
On a :

$$\begin{aligned}
\mathbb{P}(8 \leq X \leq 12) &= \mathbb{P}(X = 8) + \mathbb{P}(X = 9) + \mathbb{P}(X = 10) + \mathbb{P}(X = 11) + \mathbb{P}(X = 12) \\
&= 0.1126 + 0.1251 + 0.1251 + 0.1137 + 0.0948 \\
&\approx 57\%
\end{aligned}$$

8.5 Approximation d'une loi binomiale

On utilisera, le plus souvent, la loi de Poisson comme approximation d'une loi binomiale. Soit X une variable aléatoire de loi $\mathcal{B}(n, p)$. On cherche la limite de $\mathbb{P}(X = k)$ lorsque : n tend vers l'infini, p tend vers zéro, le produit np tend vers une valeur finie λ. On a :

$$\mathbb{P}(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \frac{n!}{(n-k)! n^k} \frac{(np)^k}{k!} \frac{(1-p)^n}{(1-p)^k}$$

- $\lim_{n\to+\infty} \frac{n!}{(n-k)!n^k} = 1$ car :

$$\begin{aligned} \frac{n!}{(n-k)!n^k} &= \frac{n}{n}\frac{n-1}{n}\frac{n-2}{n}\dots\frac{n-k+1}{n} \\ &= 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{k-1}{n}\right) \end{aligned}$$

- $\lim_{p\to 0}(1-p)^k = 1$
- $\log(1-p)^n = n\log(1-p) \approx -np$ lorsque $n \to +\infty$ et $p \to 0$.

On en déduit que :

$$\lim_{n\to+\infty, p\to 0, np\to\lambda} (1-p)^n = e^{-\lambda}$$

En conclusion, lorsque n tend vers l'infini, p tend vers zéro, le produit np tend vers une valeur finie λ, la loi binomiale $\mathcal{B}(n,p)$ converge vers une loi de poisson $\mathcal{P}(\lambda)$ puisque :

$$C_n^k p^k (1-p)^{n-k} \to e^{-\lambda}\frac{\lambda^k}{k!}$$

Dans la pratique nous utiliserons ce résultat mathématique de la manière suivante : lorsque n a une valeur élevée ($n \geq 100$) et une valeur p très faible ($p \leq$ quelques pour cent), une loi binomiale $\mathcal{B}(n,p)$ peut être approchée par une loi de Poisson $\mathcal{P}(\lambda = np)$

Exemple 2.10 *La probabilité d'observer une mutation sur un individu étant de* 10^{-4}*, combien d'individus faut-il s'attendre à examiner pour être "pratiquement sûr" d'observer au moins un mutant ?*
Soit n *ce nombre et soit* X *la variable aléatoire qui associe à ces* n *individus le nombre de mutants que l'on peut observer parmi eux.*
X *suit une loi* $\mathcal{B}(n, 10^{-4})$*. Comme* 10^{-4} *est une valeur très faible et que* n *a certainement une valeur élevée, on peut dire que* X *suit pratiquement une loi* $\mathcal{P}(\lambda = n10^{-4})$*.*
Si on considère comme pratiquement sûr un événement de probabilité au moins égale à 0.95, on écrit :

$$\mathbb{P}(X \geq 1) \geq 0.95 \Leftrightarrow \mathbb{P}(X < 1) \leq 0.05$$

Or $\mathbb{P}(X < 1) = \mathbb{P}(X = 0)$ *et* $\mathbb{P}(X = 0) = e^{-\lambda}$ *pour une loi de Poisson* $\mathcal{P}(\lambda)$*.*
On veut donc $e^{-n10^{-4}} \leq 0.05$*. D'où :*

$$\begin{aligned} -n10^{-4} &\leq \log 0.05 \\ n10^{-4} &\geq -\log 0.05 \\ n &\geq 2.99573\,10^{+4} \end{aligned}$$

$$\Rightarrow n \geq 29958$$

Exemple 2.11 *Le tableau ci-dessous donne une idée de l'approximation*

	$\mathcal{B}\,(300, 10^{-2})$	$\mathcal{P}(3)$
$\mathbb{P}(X=0)$	*0.0490*	*0.0498*
$\mathbb{P}(X=1)$	*0.1486*	*0.1493*
$\mathbb{P}(X=2)$	*0.2244*	*0.2240*
$\mathbb{P}(X=3)$	*0.2252*	*0.2240*
$\mathbb{P}(X=4)$	*0.1689*	*0.1680*
$\mathbb{P}(X=5)$	*0.1010*	*0.1008*
$\mathbb{P}(X=6)$	*0.0501*	*0.0504*
$\mathbb{P}(X=7)$	*0.0212*	*0.0216*
$\mathbb{P}(X=8)$	*0.0079*	*0.0081*

9 La loi géométrique

Soit une suite d'expériences aléatoires $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n, \ldots$ obéissant aux conditions du schéma de Bernoulli :

1. chaque expérience peut entraîner l'observation d'un événement E ou de son contraire $\overline{E}$;
2. la probabilité de E, notée p, est la même pour chaque expérience ;
3. le résultat d'une expérience est indépendant des résultats des autres expériences.

On note E_k l'événement "E se réalise à la $k^{\text{ème}}$ expérience".
Si on considère maintenant la variable aléatoire X qui associe à toute suite de résultats le rang du premier événements E, on a :

$$\mathbb{P}(X=k) = \mathbb{P}(\overline{E_1} \cap \overline{E_2} \cap \cdots \cap \overline{E_{k-1}} \cap \mathbb{E}_k)$$

L'indépendance des événement permet d'écrire :

$$\mathbb{P}(X=k) = \mathbb{P}(\overline{E_1})\mathbb{P}(\overline{E_2})\cdots\mathbb{P}(\overline{E_{k-1}})\mathbb{P}(\mathbb{E}_k) = (1-p)^{k-1}p$$

ou, en posant $q = 1 - p = \mathbb{P}(\overline{E})$:

$$P(X=k) = q^{k-1}p$$

on remarque le fait que $k \in \mathbb{N}^*$.

Définition 2.8 *On dit qu'une variable aléatoire X, à valeurs dans $\mathbb{N}^*$, suit une loi géométrique, ou de Pascal, si sa loi de probabilité est définie par $\mathbb{P}(X=k) = (1-p)^{k-1}p$ où $k \in \mathbb{N}^*$ et où p est un réel tel que $0 < p < 1$.*

9.1 Paramètres descriptifs

On démontre (voir exercice 18) les résultats suivants :

$$\mathbb{E}(X) = \frac{1}{p} \quad , \quad \mathbb{V}(X) = \frac{1-p}{p^2} = \frac{q}{p^2}$$

La formule de récurrence $\mathbb{P}(X=k+1) = \mathbb{P}(X=k)(1-p)$ montre, puisque $0 < 1 - p < 1$, que les probabilités successives décroissent constamment ; le mode a donc pour valeur 1.

10 Exercices

1. Paul a dans sa poche deux boîtes d'allumettes indiscernables ; l'une contient 5 allumettes, l'autre 2. Il choisit au hasard une des boîtes, allume sa cigarette avec une seule allumette, puis remet la boîte dans sa poche si elle n'est pas vide, ou la jette lorsqu'elle est vide. Soit X la variable aléatoire représentant le nombre de cigarettes allumées avant de jeter une des deux boîtes.

   Déterminer la loi de X. Calculer l'espérance et la variance de X.

2. Un sac contient trois tickets numérotés 1,2,3
   (a) Soit X le numéro que l'on peut obtenir en tirant un ticket. Quelle est sa loi de probabilité ? Calculer son espérance et son écart-type.
   (b) Soit $\overline{X_2}$ la variable représentant la moyenne arithmétique des deux numéros que l'on peut obtenir en tirant deux tickets. Déterminer sa loi de probabilité et calculer son espérance et son écart-type selon les deux modes de tirages : exhaustif et non exhaustif.
   (c) Même question pour $\overline{X_3}$, moyenne arithmétique des trois numéros que l'on peut obtenir en tirant trois tickets.

3. Un gardien de nuit a 10 clés, dont une seule marche, pour ouvrir une porte. Il emploie deux méthodes :

   A : méthode rationnelle ; à jeun, il retire les clés déjà essayées.

   B : ivre, chaque clé peut être essayée plusieurs fois.

   Soit X_A le nombre de clés déjà essayées avant d'ouvrir, y compris la bonne, dans le cas A, X_B dans le cas B.
   (a) Déterminer la loi de probabilité et la fonction de répartition de X_A et X_B ;
   (b) Calculer les espérance de X_A et X_B ;
   (c) on sait que le gardien est ivre un jour sur trois. Un jour, après avoir essayé 8 clés, le gardien n'a toujours pas ouvert la porte. Calculer la probabilité pour qu'il soit ivre.

4. Chez une espèce végétale, la longueur des entre-nœuds séparant deux feuilles successives ayant terminé leur croissance est gouverné par un couple d'allèles Aa. Les longueurs des entre-nœuds des génotypes AA, Aa, aa sont respectivement égales à 3cm, 5cm et 2cm.

   On considère une population composée des trois types d'individus avec les fréquences $(1-q)^2$ pour le génotype AA, $2q(1-q)$ pour le génotype Aa et q^2 pour le génotype aa.
   (a) i. Un individu est tiré au hasard dans la population. Quelles sont les probabilités :
      - de tirer un individu de génotype AA ;
      - de tirer un individu de génotype Aa ;
      - de tirer un individu de génotype aa ?

ii. Calculer, en fonction de q, l'espérance mathématique μ_1 de la variable aléatoire X_1 égale à la longueur des entre-nœuds de l'individu tiré au hasard dans la population et étudier les variations de μ_1, lorsque q varie de 0 à 1 (bornes incluses).

iii. Pour quelle valeur de q, l'espérance mathématique de la longueur des entre-nœuds est-elle maximum ? Quelles sont alors les probabilités de tirer :
- de tirer un individu de génotype AA ;
- de tirer un individu de génotype Aa ;
- de tirer un individu de génotype aa ?

iv. Calculer, en fonction de q, les probabilités pour qu'un gamète issu d'un individu tiré dans la population :
- porte l'allèle A ;
- porte l'allèle a.

(b) Un sélectionneur décide d'éliminer de la reproduction tous les individus dont la longueur des entre-nœuds est égale à 2 cm.

i. Calculer, en fonction de q, pour un individu tiré au hasard dans la population, la probabilité :
- d'être de génotype AA ;
- d'être de génotype Aa ?

ii. Calculer, en fonction de q, l'espérance mathématique de la variable aléatoire X_2 égale à la longueur des entre-nœuds d'un individu tiré dans la population conservée pour la reproduction et étudier les variations de μ_2, lorsque $0 \leq q < 1$.

iii. Calculer, en fonction de q, les probabilités pour qu'un gamète issu d'un individu tiré dans la population conservée pour la reproduction :
- porte l'allèle A ;
- porte l'allèle a.

Comparer ces deux probabilités à celles de la question précédente.

N.B. 1 On supposera toujours que tous les individus de la population ont la même probabilité d'être tirés lors du tirage au hasard.

N.B. 2 On rappelle qu'un individu de génotype AA produit des gamètes portant l'allèle A avec la probabilité 1, qu'un individu de génotype aa produit des gamètes portant l'allèle a avec la probabilité 1 et qu'un individu de génotype Aa produit des gamètes portant l'allèle A avec la probabilité ½, et portant l'allèle a avec la probabilité ½.

5. Un groupe de 10 personnes est composé de 4 hommes et de 6 femmes. On choisit dans ce groupe un échantillon de 4 personnes. Déterminer la loi de probabilité du nombre de femmes que l'on peut observer dans un tel échantillon ; calculer son espérance et sa variance.

6. Déterminer la loi de probabilité, le mode, l' espérance et l'écart-type de variable aléatoire donnée par la fonction de répartition :

$$\begin{aligned} F(t) &= 1 \quad , \quad \text{si } t > 3 \\ F(t) &= \frac{4}{5} \quad , \quad \text{si } 2 < t \leq 3 \\ F(t) &= \frac{1}{5} \quad , \quad \text{si } 1 < t \leq 2 \\ F(t) &= 0 \quad , \quad \text{si } t \leq 1 \end{aligned}$$

7. Une urne contient 10 boules numérotées : 3 numéros 1, 2 numéros 2, 5 numéros 3. On tire deux boules et on considère la variable aléatoire représentant le total des nombres marqués sur les deux boules :
   - déterminer sa loi de probabilité ;
   - déterminer sa fonction de répartition ;
   - calculer son espérance et sa variance ;
   - quelle est la probabilité pour que ce total prenne une valeur au moins égal à 6 ? strictement compris entre 2 et 6 ?
8. Combien de fois faut-il envisager de lancer une pièce de monnaie pour être "pratiquement sûr" d'observer au moins une fois l'événement "face" ?
9. Soit un restaurant de 50 places. La probabilité, pour qu'une personne, ayant réservé une table, ne vienne pas, est de 20%. Un jour, le patron a pris 52 réservations. Quelle est la probabilité pour qu'il se trouve dans une situation embarrassante ?
10. On considère deux races de plantes pures d'une même espèce diploïdes, l'une à fleurs rouges, l'autre à fleurs blanches ; la coloration de la fleur est déterminée par un couple d'allèles. On effectue un croisement pour obtenir la génération $F1$, la génération $F2$ étant obtenue par autofécondation.
    - Les plantes observés en $F1$ sont rouges. Déterminer la loi de probabilité de X, nombre de plantes à fleurs rouges que l'on peut observer en $F2$ sur cinq plantes.
    - Les plantes observés en $F1$ sont roses. Déterminer la loi de Y, nombre de plantes à fleurs roses, que l'on peut observer en $F2$, toujours sur 5 plantes.
    - Comparer la forme de ces trois distributions.
11. Soit la variable X, de loi binomiale $\mathbb{P}(X = k) = C_n^k p^k q^{n-k}$ avec $k \in \{0, 1, 2, \ldots, n\}$, n un entier donné et p un réel tel que $0 < p < 1$.
    - Calculer l'espérance et la variance de X ;
    - déterminer le mode de Xet en déduire la forme des distributions de l'exercice précédent.
12. Soit la variable aléatoire représentant le nombre de filles que l'on peut observer dans une famille de 6 enfants. Déterminer sa loi de probabilité, son mode, son espérance, sa variance.
13. Un amplificateur est fabriqué à l'aide de 100 composants électroniques. Chaque composant a une probabilité de 10^{-3} de se révéler défectueux à la mise en service. On admet qu'il suffit que l'un des composants soit défectueux pour que l'amplificateur ne fonctionne pas et qu'il n'y a pas d'autres causes de panne.

- Soit P la probabilité pour qu'un amplificateur ne fonctionne pas. On posera pour la suite $P = \frac{1}{10}$. Justifier ce résultat.
- Soit le nombre d'amplificateurs défectueux que l'on peut obtenir dans une série de 4. Déterminer sa loi de probabilité.

14. On considère deux races d'une même espèce. La première est caractérisée par ses fleurs rouges et ses feuilles découpées, la deuxième par ses fleurs blanches et ses feuilles entières.

    On observe en $F1$, après croisement, des plantes à fleurs rouges et à feuilles découpées. On admet les hypothèses suivantes : chaque caractère est déterminé par un couple d'allèles, les populations parentales sont pures, les deux couples d'allèles sont indépendants.

    Soit X la variable aléatoire représentant le nombre de plantes à fleurs blanches et à feuilles entières qu'on peut observer, sur 100 plantes obtenues en $F2$ par autofécondation. Quelle est la loi de X ? Déterminer le mode de X et donner la forme de sa distribution. Calculer l'espérance et la variance de X.

15. Deux personnes lancent, 5 fois chacune, une pièce de monnaie. Calculer la probabilité pour qu'elles obtiennent le même nombre de faces.

16. Dans un stock de boîtes de lait de longue conservation, il peut exister des boîtes non stériles et donc défectueuses ; on notera N le nombre de boîtes de lait du stock, D le nombre de celles qui sont défectueuses et $q = \frac{D}{N}$

    Pour s'assurer que la proportion de boîtes défectueuses est suffisamment faible, on contrôle un échantillon de n boîtes prélevé dans le stock sous les conditions :

    Chaque boîte a la même probabilité $1/N$ d'être prélevée. La taille n de l'échantillon est petite par rapport à la taille N du stock et on peut faire l'approximation que les prélèvements des n boîtes sont des événements indépendants.

    (a) Quelle est, sous ces conditions, la loi de la variable aléatoire X égale au nombre de boîtes non stériles dans l'échantillon ?

    Quelles sont les probabilités α et β des deux événements :
    - les boîtes de l'échantillon sont toutes stériles ;
    - une boîte au moins de l'échantillon est non stérile.

    (b) On veut que la probabilité β soit supérieure ou égale à 0.95 si q dépasse un seuil fixé q_0 (q_0 est la valeur à partir de laquelle on estime que la qualité du produit est inacceptable).

    Quel nombre n de boîtes faut-il prélever pour que cette condition soit satisfaite ?

    Donner la valeur minimale de n si $q_0 = 0.01$

17. On jette simultanément deux dés réguliers. On s'intéresse à l'événement "sortie d'une paire" ; quel est sa probabilité ?
    - Soit X le nombre de paires que l'on peut obtenir en jetant les deux dés quatre fois de suite. Déterminer sa loi de probabilité. Calculer son espérance et son écart-type.

- Laurent propose à Paul le jeu suivant : tous deux misent la même somme puis Paul lance quatre fois les deux dés ; s'il obtient au moins une paire il emporte la mise, si non c'est Laurent. Le jeu est-il équitable ?

18. On considère une suite chronologique d'expériences aléatoires : $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n, \ldots$. Chaque expérience n'admet que deux résultats possibles : l'événement A, de probabilité P ou l'événement B de probabilité $1 - P$; ces deux événements sont incompatibles et les réalisations successives sont indépendantes. Soit U la variable aléatoire associant à la suite d'expériences le rang du premier événement A :
    - déterminer la loi de probabilité de U ;
    - calculer l'espérance et la variance de U.
19. On considère une variable aléatoire X dont la loi de probabilité est une loi de Poisson : $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, pour $k \in \mathbb{N}$, λ étant un réel donné strictement positif. Montrer que $\mathbb{E}(X) = \mathbb{V}(X) = \lambda$. Déterminer le mode de X et en déduire la forme de la distribution.
20. Soit une variable aléatoire X, dont la loi est une loi de Poisson d'espérance $\lambda = 0.43$. Calculer, à 10^{-4} près, la probabilité de chacune des premières valeurs de X ; justifier l'arrêt des calculs.
21. On admet que la probabilité d'apparition d'une mutation sur un individu est 10^{-6}. Combien d'individus faut-il s'attendre à examiner pour être pratiquement sûr d'observer au moins un mutant ?
22. On choisit au hasard une famille dans une population donnée. Soit X le nombre d'enfants qu'elle peut contenir ; on admet que X est une variable de Poisson d'espérance μ. On sait, d'autre part, qu'à chaque naissance la probabilité d'avoir un garçon est ½.
    - Comparer les histogrammes de la loi de X dans les cas $\mu = 2$ et $\mu = 3.5$;
    - exprimer la probabilité d'observer k garçons dans une famille dont on sait qu'elle contient n enfants,
    - déterminer la loi de probabilité du nombre de garçons que l'on peut observer dans une famille extraite de la population.
23. Le bureau d'étude d'un centre spatial doit donner son avis sur un projet de fusée. Le moteur et l'équipement électronique se composent de 1500000 pièces distinctes. Chaque pièce a une chance sur 10 millions de se révéler défectueuse ; la défectuosité d'une seule pièce suffit à empêcher le bon fonctionnement de la fusée, les causes de défectuosité des diverses pièces étant indépendantes.

    Que penser du projet ?
24. La probabilité pour que l'injection d'un certain sérum provoque une réaction allergique, sur un individu choisi au hasard, est de 0.1%. Quelle est la probabilité pour que, sur 900 individus traités on observe l'allergie dans :
    - exactement 3 cas ;
    - au plus 3 cas ;
    - au moins 3 cas ?
25. Dans une espèce donnée, la proportion des albinos est de 0.2%. Calculer la probabilité d'observer au moins 2 albinos dans un échantillon de 230 individus.

26. Dans une population humaine, la proportion de gauchers est de 1%. Quelle est la probabilité d'observer au moins 4 gauchers dans un échantillon de 230 personnes ?
27. Une usine fabrique des petits gâteaux aux raisins. La pâte et les raisins sont mélangés avant le découpage des gâteaux. Quel nombre moyen de raisins faut-il prévoir par gâteau si l'on veut que la probabilité pour qu'un gâteau contienne au moins un raisin soit de 99% ?
28. (Processus poissonien). On considère une suite d'événements identiques qui se succèdent dans le temps. Le nombre d'événements qui se produisent dans une intervalle de temps est aléatoire. Soit $w_k(t)$ la probabilité d'avoir k événements dans un intervalle de temps de longueur t. Calculer $w_k(t)$ en supposant :
    (a) $w_k(t)$ ne dépend que de la longueur t de l'intervalle de temps et pas de l'instant initial ;
    (b) la probabilité d'observer un événement dans un intervalle de temps dt est un infiniment petit du même ordre que dt, donc équivalent à λdt ($\lambda =$ constante positive) ;
    (c) la probabilité d'observer plus d'un événement dans un intervalle de temps dt est un infiniment petit d'ordre supérieur à dt ;
    (d) si t et t' sont des intervalles de temps sans partie commune, le nombre d'événements se produisant dans t est indépendant du nombre d'événements se produisant dans t'.

11 Solutions

1. L'expérience $\mathcal{E}$ consiste à allumer successivement 7 cigarettes. On note X la variable qui associe à $\mathcal{E}$ le nombre de cigarettes allumées avant de jeter une des deux boîtes. X est une variable discrète. On se propose de calculer $\mathbb{P}(X = k)$ avec $k \in \{2, 3, 4, 5, 6\}$

   Soient les événements :

   A **:** pour une cigarette on tire la première boîte (de 5 allumettes) ;

   B **:** pour une cigarette on tire la deuxième boîte (de 2 allumettes) ;

   (iA, jB) **:** les $(i + j)$ premiers choix ont donné i fois A et j fois B ;

   (dernier A**)** le dernier choix est A ;

   (dernier B**)** le dernier choix est B

   On a : $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}$ tant qu'il y a deux boîtes et $\mathbb{P}((iA, jB)) = C^i_{i+j} \left(\frac{1}{2}\right)^{i+j}$, si $(iA, jB) \neq 0$.

   On écrit successivement

$$
\begin{aligned}
\mathbb{P}(X=2) &= \mathbb{P}(1B \cap \text{dernier } B) = \mathbb{P}(B)\mathbb{P}(B) = \frac{1}{4} \\
\mathbb{P}(X=3) &= \mathbb{P}((1A,1B) \cap \text{dernier } B) = \mathbb{P}((1A,1B))\mathbb{P}(B) = C^1_2 \left(\frac{1}{2}\right)^2 \frac{1}{2} = \frac{2}{8} \\
\mathbb{P}(X=4) &= \mathbb{P}((2A,1B) \cap \text{dernier } B) = \mathbb{P}((2A,1B))\mathbb{P}(B) = C^1_3 \left(\frac{1}{2}\right)^3 \frac{1}{2} = \frac{3}{16} \\
\mathbb{P}(X=5) &= \mathbb{P}((3A,1B) \cap \text{dernier } B) + \mathbb{P}(4A \cap \text{dernier } A) \text{ (événements incompatibles)} \\
&= \mathbb{P}((3A,1B))\mathbb{P}(B) + \mathbb{P}((4A)\mathbb{P}(B) = C^3_4 \left(\frac{1}{2}\right)^4 \frac{1}{2} + \left(\frac{1}{2}\right)^4 \frac{1}{2} = \frac{5}{32} \\
\mathbb{P}(X=26) &= \mathbb{P}((4A,1B) \cap \text{dernier } A \cup B) = \mathbb{P}((4A,1B)) = C^1_5 \left(\frac{1}{2}\right)^5 = \frac{5}{32}
\end{aligned}
$$

   D'où la loi de probabilité :

x_i	2	3	4	5	6
p_i	$\frac{8}{32}$	$\frac{8}{32}$	$\frac{6}{32}$	$\frac{5}{32}$	$\frac{5}{32}$

   On vérifie que $\sum_i p_i = 1$.

   On en déduit :

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{i=1}^{n} x_i p_i = \frac{1}{32}(16 + 24 + 24 + 25 + 30) = \frac{119}{32} \approx 3.719 \\
\mathbb{E}(X^2) &= \sum_{i=1}^{n} x_i^2 p_i = \frac{1}{32}(32 + 72 + 96 + 125 + 180) = \frac{505}{32} \approx 15.78 \\
\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{1999}{1024} \approx 1.95
\end{aligned}
$$

2. (a) X est une variable aléatoire discrète de loi uniforme sur $\{1, 2, 3\}$

$$\mathbb{P}(X=1)=\mathbb{P}(X=2)=\mathbb{P}(X=3)=\frac{1}{3}$$

$$\begin{aligned}\mathbb{E}(X) &= \mu=\frac{1}{3}(1+2+3)=2\\ \mathbb{V}(X) &= \mathbb{E}\left((X-\mu)^2\right)=\frac{1}{3}(1+0+1)=\frac{2}{3}\end{aligned}$$

$$\sigma(X)=\frac{\sqrt{2}}{\sqrt{3}}$$

On note, pour la suite X, Y et Z les numéros tirés en premier, deuxième et troisième rang.

(b) **Tirage exhaustif** On a pour résultat possibles :

$$\begin{aligned}\left(\overline{X_2}=\frac{3}{2}\right) &= (X=1\cap Y=2)\cup(X=2\cap Y=1)\\ \left(\overline{X_2}=2\right) &= (X=1\cap Y=3)\cup(X=3\cap Y=1)\\ \left(\overline{X_2}=\frac{5}{2}\right) &= (X=2\cap Y=3)\cup(X=3\cap Y=2)\end{aligned}$$

En appliquant l'axiome des probabilités totales (événements incompatibles) puis la relation des probabilités composées (événements *dépendants*, on a, par exemple, :

$$\begin{aligned}\mathbb{P}\left(\overline{X_2}=\frac{3}{2}\right) &= \mathbb{P}(X=1)\mathbb{P}\left(Y=2|X=1\right)+\mathbb{P}(X=2)\mathbb{P}\left(Y=1|X=2\right)\\ &= \frac{1}{3}\frac{1}{2}+\frac{1}{3}\frac{1}{2}\\ &= \frac{1}{3}\end{aligned}$$

On a la loi :

x_i	$\frac{3}{2}$	2	$\frac{5}{2}$
$\mathbb{P}(\overline{X_2}=x_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

On en déduit :

$$\begin{aligned}\mathbb{E}(\overline{X_2}) &= \mu=\frac{1}{3}(3/2+0+5/2)=2=\mathbb{E}(X)\\ \mathbb{V}(\overline{X_2}) &= \mathbb{E}\left((\overline{X_2}-\mu)^2\right)=\frac{1}{3}(1/4+1/4)=\frac{1}{6}\\ \sigma(\overline{X_2}) &= \frac{1}{\sqrt{6}}\end{aligned}$$

Tirage non exhaustif On a pour résultat possibles :

$$\begin{aligned}
\left(\overline{X_2} = 1\right) &= (X = 1 \cap Y = 1) \\
\left(\overline{X_2} = \frac{3}{2}\right) &= (X = 1 \cap Y = 2) \cup (X = 2 \cap Y = 1) \\
\left(\overline{X_2} = 2\right) &= (X = 1 \cap Y = 3) \cup (X = 2 \cap Y = 2) \cup (X = 3 \cap Y = 1) \\
\left(\overline{X_2} = \frac{5}{2}\right) &= (X = 2 \cap Y = 3) \cup (X = 3 \cap Y = 2) \\
\left(\overline{X_2} = 3\right) &= (X = 3 \cap Y = 3)
\end{aligned}$$

En appliquant l'axiome des probabilités totales (événements incompatibles) puis la relation des probabilités composées (événements *indépendants*, on a, par exemple, :

$$\begin{aligned}
\mathbb{P}\left(\overline{X_2} = 2\right) &= \mathbb{P}(X = 1)\mathbb{P}(Y = 3) + \mathbb{P}(X = 2)\mathbb{P}(Y = 2) + \mathbb{P}(X = 3)\mathbb{P}(Y = 1) \\
&= 3\frac{1}{3}\frac{1}{3} \\
&= \frac{3}{9}
\end{aligned}$$

On a la loi :

x_i	1	$\frac{3}{2}$	2	$\frac{5}{2}$	3
$\mathbb{P}(\overline{X_2} = x_i)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

On en déduit :

$$\begin{aligned}
\mathbb{E}(\overline{X_2}) &= \mu = \frac{1}{9}(1 + 3 + 6 + 5 + 3) = 2 = \mathbb{E}(X) \\
\mathbb{V}(\overline{X_2}) &= \mathbb{E}\left((\overline{X_2} - \mu)^2\right) = \frac{1}{9}(1 + 2\frac{1}{4} + 0 + 2\frac{1}{4} + 1) = \frac{1}{3} \\
\sigma(\overline{X_2}) &= \frac{1}{\sqrt{3}} = \frac{\sigma(X)}{\sqrt{2}}
\end{aligned}$$

(c) **tirage exhaustif** Une seule valeur observable $\overline{X_3}$: 2 Donc $\mathbb{P}(\overline{X_3} = 2) = 1$ et $\mathbb{E}(\overline{X_3}) = 1 \times 2 = \mathbb{E}(X)$ et $\mathbb{V}(\overline{X_3}) = 0 = \sigma(\overline{X_3})$

tirage non exhaustif En raisonnant comme dans la question précédente, on obtient :

x_i	1	$\frac{4}{3}$	$\frac{5}{3}$	2	$\frac{7}{3}$	$\frac{8}{3}$	3
$\mathbb{P}(\overline{X_2} = x_i)$	$\frac{1}{27}$	$\frac{3}{27}$	$\frac{6}{27}$	$\frac{7}{27}$	$\frac{6}{27}$	$\frac{3}{27}$	$\frac{1}{27}$

On en déduit :

$$\begin{aligned}
\mathbb{E}(\overline{X_2}) &= \mu = \frac{1}{27}(1 + 4 + 10 + 14 + 14 + 8 + 3) = 2 = \mathbb{E}(X) \\
\mathbb{V}(\overline{X_2}) &= \mathbb{E}\left((\overline{X_2} - \mu)^2\right) = \frac{1}{27}\left(1 + 3\frac{4}{9} + 6\frac{1}{9} + 0 + 6\frac{1}{9} + 3\frac{4}{9} + 1\right) = \frac{2}{9} \\
\sigma(\overline{X_2}) &= \frac{\sqrt{2}}{3} = \frac{\sigma(X)}{\sqrt{3}}
\end{aligned}$$

3. On note E_i l'événement "la bonne clef est tirée la $i^{\text{ème}}$ fois et $\overline{E_i}$ l'événement "la bonne clef n'est pas tirée la $i^{\text{ème}}$ fois.

   Si X est le nombre d'essais, on a :$(X = k) = \overline{E_1} \cap \overline{E_2} \cap \ldots \overline{E_{k-1}} \cap E_k$, $k \in \{1, 2, \ldots, 10\}$

   Méthode A (tirage exhaustif)

$$\begin{aligned}
\mathbb{P}(X_A = 1) &= \mathbb{P}(E_1) = \frac{1}{10} \\
\mathbb{P}(X_A = 2) &= \mathbb{P}(\overline{E_1} \cap E_2) = \mathbb{P}(\overline{E_1})\mathbb{P}(E_2|\overline{E_1}) = \frac{9}{10}\frac{1}{9} = \frac{1}{10} \\
\mathbb{P}(X_A = 3) &= \mathbb{P}(\overline{E_1} \cap \overline{E_2} E_2) = \frac{9}{10}\frac{8}{9}\frac{1}{8} = \frac{1}{10} \\
\vdots &= \vdots \\
\mathbb{P}(X_A = 10) &= \mathbb{P}(\overline{E_1} \cap \ldots \cap \overline{E_9} \cap E_{10}) = \frac{9}{10}\frac{8}{9}\frac{7}{8}\cdots\frac{3}{2}\frac{1}{2}1 = \frac{1}{10}
\end{aligned}$$

   On peut également dire : tirer 10 clefs en tirage exhaustif revient à ordonner les 10 clefs. La bonne clef a une probabilité de $\frac{9!}{10!} = \frac{1}{10}$ d'être à une place donnée. D'où la loi de X_A :

$$\mathbb{P}(X_A = k) = \frac{1}{10} \quad , \quad \forall k \in \{1, 2, 3, \ldots, 10\}$$
$$\mathbb{P}(X_A = k) = 0 \quad \text{si} \quad k \notin \{1, 2, 3, \ldots, 10\}$$

   Il s'agit d'une loi uniforme sur $\{1, 2, 3, \ldots, 10\}$.
   On en déduit, en posant $F(t) = \mathbb{P}(X_A < t)$:

$$\begin{aligned}
F(t) = 0 \quad &, \quad \text{si } t \leq 1 \\
F(t) = \frac{k}{10} \quad &, \quad \text{où } k \text{ est l'entier immédiatement inférieur à } t\text{, si } 0 < t \leq 10 \\
F(t) = 1 \quad &, \quad \text{si } t \geq 10
\end{aligned}$$

   Méthode B (tirage non exhaustif) $k \in \mathbb{N}^*$

$$\mathbb{P}(X = k) = \mathbb{P}(\overline{E_1} \cap \overline{E_2} \cap \ldots \overline{E_{k-1}} \cap E_k) = \left(\frac{9}{10}\right)^{k-1} \frac{1}{10}$$

   D'où la loi géométrique : $\mathbb{P}(X_B = k) = \left(\frac{9}{10}\right)^{k-1} \frac{1}{10}$ pour $k \in \mathbb{N}^*$ et $\mathbb{P}(X_B = k) = 0$ pour $k \notin \mathbb{N}^*$

$$\begin{aligned}
G(t) = 0 \quad &, \quad \text{si } t \leq 1 \\
G(t) = 1 - \left(\frac{9}{10}\right)^{k-1} \quad &, \quad \text{où } k \text{ est l'entier immédiatement inférieur à } t\text{, si } t > 1
\end{aligned}$$

(somme des termes d'une suite géométrique : $\frac{1}{10}(1 + {}^9\!/\!_{10} + \cdots + {}^{9^{k-1}}\!/\!_{10}) = \frac{1}{10}\frac{1-\left(\frac{9}{10}\right)^k}{1-{}^9\!/\!_{10}}$
Pour $|x| < 1$, $1 + x + x^2 + \cdots + x^n + \cdots = \frac{1}{1-x}$ donc $1 + 2x + 3x^2 + \cdots + nx^{n-1} + \cdots = \frac{1}{(1-x)^2}$. On en déduit :

$$\mathbb{E}(X_B) = \frac{1}{10}\left(1 + 2\frac{9}{10} + 3\left(\frac{3}{10}\right)^2 + \cdots\right) = \frac{1}{10}\frac{1}{({}^1\!/\!_{10})^2} = 10$$

Pour la dernière question on note que

$$\begin{aligned}
\mathbb{P}(X_A > 8) &= \mathbb{P}(X_A = 9) + \mathbb{P}(X_A = 10) = \frac{2}{10} = 0.2 \\
\mathbb{P}(X_B > 8) &= 1 - \mathbb{P}(X_B \leq 8) = 1 - G(9) = \left(\frac{9}{10}\right)^8 \approx 0.43 \\
\mathbb{P}(X > 8) &= \mathbb{P}(X = X_A)\mathbb{P}(X > 8|X = X_A) + \mathbb{P}(X = X_B)\mathbb{P}(X > 8|X = X_B) \\
&= \frac{2}{3}\mathbb{P}(X_A > 8) + \frac{1}{3}\mathbb{P}(X_B > 8) \\
&\approx \frac{2}{3}0.2 + \frac{1}{3}0.43 \approx \frac{1}{3}0.83
\end{aligned}$$

On peut alors écrire :

$$\mathbb{P}(\text{gardien ivre}|X > 8) = \mathbb{P}(X = X_B|X > 8) = \frac{\mathbb{P}(X = X_B \cap X > 8)}{\mathbb{P}(X > 8)} \approx \frac{\frac{1}{3}0.43}{\frac{1}{3}0.83} \approx 0.518$$

4. Problème très simple
   (a) i. $\mathbb{P}(AA) = (1-q)^2$; $\mathbb{P}(Aa) = 2q(1-q)$; $\mathbb{P}(aa) = q^2$
   ii. Loi de X_1 : $\mathbb{P}(X_1 = 2) = q^2$; $\mathbb{P}(X_1 = 3) = (1-q)^2$; $\mathbb{P}(X_1 = 5) = 2q(1-q)$. Donc $\mu_1 = \mathbb{E}(X_1) = 2q^2 + 3(1-q)^2 + 10q(1-q) = -5q^2 + 4q + 3$
   iii. μ_1 est maximum pour $q = {}^2\!/\!_5$. Dans ce cas on a $\mathbb{P}(AA) = 36\%$; $\mathbb{P}(Aa) = 48\%$; $\mathbb{P}(aa) = 16\%$
   iv. $\mathbb{P}(A) = \mathbb{P}(\text{donneur } AA \cap A) + \mathbb{P}(\mathbb{P}(\text{donneur } Aa \cap A) = (1-q)^2 \times 1 + 2q(1-q) \times \frac{1}{2} = 1 - q$
   $\mathbb{P}(a) = \mathbb{P}(\text{donneur } Aa \cap a) + \mathbb{P}(\mathbb{P}(\text{donneur } aa \cap a) = 2q(1-q) \times \frac{1}{2} + q^2 \times 1 = q$
   (b) i. $\mathbb{P}(AA|\text{non } aa) = \frac{\mathbb{P}(AA \cap \text{non } aa)}{\mathbb{P}(\text{non } aa)} = \frac{\mathbb{P}(AA)}{\text{non } aa} = \frac{(1-q)^2}{1-q^2} = \frac{1-q}{1+q}$
   $\mathbb{P}(Aa|\text{non } aa) = \frac{\mathbb{P}(Aa \cap \text{non } aa)}{\mathbb{P}(\text{non } aa)} = \frac{\mathbb{P}(AAa)}{\text{non } aa} = \frac{2q(1-q)}{1-q^2} = \frac{2q}{1+q}$
   ii. Loi de X_2 : $\mathbb{P}(X_2 = 3) = \frac{1-q}{1+q}$; $\mathbb{P}(X_2 = 5) = \frac{2q}{1+q}$ Donc $\mu_2 = \mathbb{E}(X_2) = \frac{1}{1+q}(3 - 3q + 10q) = \frac{3+7q}{1+q}$

iii. $\mathbb{P}(A) = \mathbb{P}(\text{donneur } AA \cap A) + \mathbb{P}(\text{donneur } Aa \cap A) = \frac{1-q}{1+q} \times 1 + \frac{2q}{1+q} \times \frac{1}{2} = \frac{1}{1+q}$ ($\geq 1-q$ car $1 \geq 1 - q^2$)
$\mathbb{P}(a) = \mathbb{P}(\text{donneur } Aa \cap a) + \mathbb{P}(\mathbb{P}(\text{donneur } aa \cap a) = \frac{2q}{1+q} \times \frac{1}{2} = \frac{q}{1+q}$ ($\leq q$ car $q \leq q + q^2$)

5. Il s'agit d'une loi hypergéométrique $\mathbb{E}(X) = 2.4$; $\mathbb{V}(X) = 0.64$

x_i	0	1	2	3	4
p_i	$\frac{1}{210}$	$\frac{24}{210}$	$\frac{90}{210}$	$\frac{80}{210}$	$\frac{15}{210}$

6. On trouve aisément la loi de T :

t_i	1	2	3
p_i	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$

puis $\mathbb{E}(T) = 2$, $\sigma(T) = \sqrt{\frac{2}{5}}$, mode=2.

7. L'espace des événements Ω associé à l'expérience $\mathcal{E}$ qui consiste à choisir 2 boules contient $C_{10}^2 = 45$ événements élémentaires également probables. On peut alors distinguer les événements suivants :

(1,1) de probabilité $\frac{1}{45}C_3^2 = \frac{3}{45}$
(1,2) de probabilité $\frac{1}{45}3 \times 2 = \frac{6}{45}$
(1,3) de probabilité $\frac{1}{45}3 \times 5 = \frac{15}{45}$
(2,2) de probabilité $\frac{1}{45}C_2^2 = \frac{1}{45}$
(2,3) de probabilité $\frac{1}{45}2 \times 5 = \frac{10}{45}$
(3,3) de probabilité $\frac{1}{45}C_5^2 = \frac{10}{45}$
On en déduit aisément la loi de probabilité :

x_i	2	3	4	5	6
p_i	$\frac{3}{45}$	$\frac{6}{45}$	$\frac{16}{45}$	$\frac{10}{45}$	$\frac{10}{45}$

puis $\mathbb{E}(X) = 4.4$ et $\mathbb{V}(X) \approx 1.35$
$\mathbb{P}(X \geq 6) = \mathbb{P}(X = 6) = \frac{10}{45}$; $\mathbb{P}(2 < X < 6) = 1 - \frac{13}{45} = \frac{32}{45}$

8. Chaque jet peut entraîner l'observation de l'événement "Face" avec la probabilité ½, ou "Pile" avec la probabilité ½, dans l'hypothèse d'un jet régulier. Les différents jets sont indépendants. Si X est la variable aléatoire "Nombre de faces que l'on peut obtenir sur n jets", on a $X \sim \mathcal{B}(n, P)$. Si l'on qualifie de "pratiquement sûr" un événement de probabilité supérieur à 0.99, il faut déterminer n pour que :

$$\mathbb{P}(X \geq 1) \geq 0.99 \Leftrightarrow \mathbb{P}(X = 0) \leq 0.01$$

Or $\mathbb{P}(X = 0) = \frac{1}{2^n}$
D'où l'inégalité $\frac{1}{2^n} \leq 0.01 \Rightarrow 2^n \geq 100$ ce qui est réalisé pour $n \geq 7$ ($2^7 = 128$).
Si on veut $\mathbb{P}(X \geq 1) \geq 0.999$ il faut choisir $n \geq 10$.

9. Pour chaque client qui a réservé, deux résultats possibles : "il vient" de probabilité 0.8, ou "il ne vient pas" de probabilité 0.2. On considère que les choix des 52

clients qui ont réservé sont indépendants. Si X est la variable aléatoire ”nombre de clients qui se présentent le soir, parmi les 52”, on a $X \sim \mathcal{B}(52; 0.8)$

La situation embarrassante correspond à l'événement ”$X > 50$.

”$X > 50$”=”$X = 51$”∪”$X = 52$” ; les deux événement étant incompatibles :

$$\mathbb{P}(X > 50) = \mathbb{P}(X = 51) + \mathbb{P}(X = 52) = 52 \left(\frac{8}{10}\right)^{51} \frac{2}{10} + \left(\frac{8}{10}\right)^{52} \approx 1.28\,10^{-4}$$

10. Exemple classique d'utilisation de la loi binomiale : observer 5 descendants dont les parents ont des génotypes donnés revient à effectuer 5 expériences identiques et indépendantes ; pour chaque descendant il a possibilité d'apparition d'un phénotype donné (probabilité P) ou non (probabilité $1 - P$).
   - $X \sim \mathcal{B}\left(5, \frac{3}{4}\right)$. Distribution en cloche dissymétrique avec mode à droite (pratiquement en J).

x_i	0	1	2	3	4	5
p_i	$\frac{1}{1024}$	$\frac{15}{1024}$	$\frac{90}{1024}$	$\frac{270}{1024}$	$\frac{405}{1024}$	$\frac{243}{1024}$

   - $X \sim \mathcal{B}\left(5, \frac{1}{4}\right)$. Distribution en cloche dissymétrique avec mode à gauche (pratiquement en L).

y_i	0	1	2	3	4	5
p_i	$\frac{243}{1024}$	$\frac{405}{1024}$	$\frac{270}{1024}$	$\frac{90}{1024}$	$\frac{15}{1024}$	$\frac{1}{1024}$

   - $X \sim \mathcal{B}\left(5, \frac{1}{2}\right)$. Distribution en cloche symétrique.

z_i	0	1	2	3	4	5
p_i	$\frac{1}{32}$	$\frac{15}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

11.

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{n} k\mathbb{P}(X = k) \\
&= \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=0}^{n} k \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}
\end{aligned}$$

On pose $i = k - 1 \Rightarrow k = i + 1 \Rightarrow n - k = n - i - 1$. On a

$$\mathbb{E}(X) = np \sum_{i=0}^{n-1} k \frac{(n-1)!}{i!(n-i-1)!} p^i (1-p)^{n-i-1} = np(p + 1 - p)^{n-1} = np$$

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=0}^{n} k^2 \mathbb{P}(X=k) \\
&= \sum_{k=0}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} (k-1+1) \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} (k-1) \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} + \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} + \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} (1-p)^{n-k} + np
\end{aligned}$$

On pose, comme plus haut $i = k-2 \Rightarrow k = i+12 \Rightarrow n-k = n-i-2$. On a :

$$\mathbb{E}(X^2) = n(n-1)p^2 \sum_{i=0}^{n-1} \frac{(n-2)!}{i!(n-k)!} p^i (1-p)^{n-2-i} + np = n(n-1)p^2 + np = n^2p^2 + np(1-p)$$

et donc

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = np(1-p)$$

Pour déterminer le mode on considère le support des probabilités de deux valeurs observables successives. On a :

$$\frac{\mathbb{P}(X=k+1)}{\mathbb{P}(X=k)} = \frac{C_n^{k+1} p^{k+1} (1-p)^{n-k-1}}{C_n^k p^k (1-p)^{n-k}} = \frac{n-k}{k+1} \frac{p}{1-p}$$

On en déduit :

$$\begin{cases}
\mathbb{P}(X=k+1) > \mathbb{P}(X=k) & \text{si } k < (n+1)P - 1 \\
\mathbb{P}(X=k+1) < \mathbb{P}(X=k) & \text{si } k > (n+1)P - 1 \\
\mathbb{P}(X=k+1) = \mathbb{P}(X=k) = k(n+1)P - 1 & \text{si } k \in [0,n] \text{ et } (n+1)P - 1 \text{ a une valeur entière}
\end{cases}$$

On résulte ces informations en disant que le mode est un entier $l \in [(n+1)P - 1, (n+1)P]$

On peut noter que $(n+1)P > 0$ car $P > 0$ et $(n+1)P - 1 < n$, car $P < \frac{n+1}{n+1} = 1$; l est donc bien une observable $\in \{0, 1, 2, \ldots, n\}$. L'intervalle a pour amplitude 1 et en général ill est unique, si $(n+1)P - 1$ a une valeur entière, il peut y avoir deux modes successifs. Il est aisé de retrouver les résultats de l'exercice 10

12. Réponses : $\mathcal{B}\left(6, \frac{1}{2}\right)$; espérance=mode=3 ; variance=1.5.
13. Réponses : $P = 1 - (1 - 10^{-3})^{100} \approx 10^{-1}$ car $(1-x)^n \approx 1 - nx$; $\mathcal{B}\left(4, \frac{1}{10}\right)$
14. Réponses : $\mathcal{B}\left(100, \frac{1}{16}\right)$.
    Distribution pratiquement en L : forme en cloche très dissymétrique, de mode 6. $\mathbb{E}(X) = 6.25$; $\mathbb{V}(X) \approx 5.86$
15. Réponse : X et Y suivent une loi $\mathcal{B}\left(5, \frac{1}{2}\right)$

$$\begin{aligned} \mathbb{P}(X = Y) &= \sum_{i=0}^{5} \mathbb{P}(X = i)\mathbb{P}(Y = i) \\ &= \sum_{i=0}^{5} \left(C_5^i\right)^2 \frac{1}{2^{10}} \\ &= \frac{63}{256} \end{aligned}$$

16. Réponses :
    - X suit une loi hypergéométrique que l'on peut approcher par $\mathcal{B}(n, q)$;
    - $\alpha = (1-q)^n$; $\beta = 1 - (1-q)^n$
    - $n \geq \frac{\log 0.05}{\log(1-q_0)}$; $\log 0.05 \approx -3$, $\log(1 - 10^{-2}) \approx -10^{-2} \Rightarrow n \geq 300$
17. Réponses : $X \sim \mathcal{B}\left(4, \frac{1}{6}\right)$; $\mathbb{E}(X) = \frac{2}{3}$; $\sigma(X) = \frac{\sqrt{5}}{3}$
    $\frac{671}{1296} > \frac{625}{1296}$; le jeu n'est pas équitable.
18. On note B_i l'événement "B se produit à la $i^{\text{ème}}$ expérience" et A_j l'événement "A se produit à la $i^{\text{ème}}$ expérience".
    On a : $(U = k) = B_1 \cap B_2 \cap \ldots \cap B_{k-1} \cap A_k$, $k \in \mathbb{N}^*$
    On en déduit, à cause de l'indépendance des réalisation successives : $\mathbb{P}(U = k) = \mathbb{P}(B_1)\mathbb{P}(B_2)\ldots\mathbb{P}(B_{k-1})\mathbb{P}(A_k) = (1-P)^{k-1}P$
    U suit une loi géométrique :

$$\mathbb{P}(U = k) = (1-P)^{k-1}P \text{ pour } k \in \mathbb{N}^*$$

On sait, que pour $|x| < 1$, on a $\sum_{k=0}^{+\infty} x^k = 1 + x + x^2 + \cdots + x^n + \cdots = \frac{1}{1-x}$
On pose $f(x) = \sum_{k=0}^{+\infty} x^k = \frac{1}{1-x}$.
On en déduit :

$$\begin{aligned} f'(x) = \sum_{k=1}^{+\infty} kx^{k-1} &= \frac{1}{(1-x)^2} \\ xf'(x) = \sum_{k=1}^{+\infty} kx^k &= \frac{x}{(1-x)^2} \\ \phi(x) = (xf'(x))' = \sum_{k=1}^{+\infty} k^2x^{k-1} &= \frac{1}{(1-x)^2} + \frac{2x}{(1-x)^3} = \frac{1+x}{(1-x)^3} \end{aligned}$$

On peut alors écrire :

$$\begin{aligned}
\sum_{k=1}^{+\infty} \mathbb{P}(U=k) &= P\sum_{k=1}^{+\infty}(1-P)^{k-1} = Pf(1-P) = \frac{P}{P} = 1 \\
\mathbb{E}(X) &= \sum_{k=1}^{+\infty} k(1-P)^{k-1}P = Pf'(1-P) = \frac{P}{P^2} = \frac{1}{P} \\
\mathbb{E}(X^2) &= \sum_{k=1}^{+\infty} k^2(1-P)^{k-1}P = P\phi(1-P) = P\frac{2-P}{P^3} = \frac{2-P}{P^2} \\
\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{2-P}{P^2} - \frac{1}{P^2} = \frac{1-P}{P^2}
\end{aligned}$$

Comme $\frac{\mathbb{P}(U=k+1)}{\mathbb{P}(U=k)} = 1-P < 1$, il est clair que la variable U admet pour mode 1 ; les probabilités décroissent quand k augment.

19.

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{+\infty} k\mathbb{P}(X=k) = \sum_{k=0}^{+\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = \lambda\sum_{k=1}^{+\infty} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\sum_{i=0}^{+\infty}\frac{\lambda^i}{i!} = \lambda e^{-\lambda}e^{\lambda} \\
\mathbb{E}(X^2) &= \sum_{k=0}^{+\infty} k^2\mathbb{P}(X=k) = \sum_{k=0}^{+\infty} k^2e^{-\lambda}\frac{\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k=1}^{+\infty} k\frac{\lambda^{k-1}}{(k-1)!}
\end{aligned}$$

Or

$$\begin{aligned}
\sum_{k=1}^{+\infty} k\frac{\lambda^{k-1}}{(k-1)!} &= \sum_{k=1}^{+\infty}(k-1+1)\frac{\lambda^{k-1}}{(k-1)!} \\
&= \sum_{k=2}^{+\infty}\frac{\lambda^{k-1}}{(k-2)!} + \sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!}
\end{aligned}$$

$$\lambda\sum_{i=0}^{+\infty}\frac{\lambda^i}{i!} + \sum_{i=0}^{+\infty}\frac{\lambda^i}{i!}$$

Et donc

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \lambda$$

Détermination du mode On a la relation $\frac{\mathbb{P}(X=k+1)}{\mathbb{P}(X=k)} = \frac{\lambda}{k+1}$

On en déduit :

$$\begin{cases}
\mathbb{P}(X=k+1) > \mathbb{P}(X=k) & \text{si } k < \lambda - 1 \\
\mathbb{P}(X=k+1) < \mathbb{P}(X=k) & \text{si } k > \lambda - 1 \\
\mathbb{P}(X=k+1) = \mathbb{P}(X=k) = k\lambda - 1 & \text{si } k \in \mathbb{N} \text{ et } \lambda \text{ a une valeur entière} \geq 0
\end{cases}$$

On résulte ces informations en disant que le mode est un entier $l \in [\lambda-1, \lambda]$ L'intervalle a pour amplitude 1, et en général, l est unique ; si λ a une valeur entière ≥ 1, on aura deux modes successifs.

Forme de la distribution En L si $\lambda \leq 1$
En cloche, si $\lambda > 1$

20. On obtient, à 10^{-4} près, pour $\mathbb{E}(X) = 0.43$

	Calcul précis	Interpolation linéaire à partir des tables
$\mathbb{P}(X=0) = e^{-0.43}$	≈0.6505	≈0.6511
$\mathbb{P}(X=1) = \mathbb{P}(X=0)\frac{0.43}{1}$	≈0.2797	≈0.2787
$\mathbb{P}(X=2) = \mathbb{P}(X=1)\frac{0.43}{2}$	≈0.0601	≈0.0602
$\mathbb{P}(X=3) = \mathbb{P}(X=2)\frac{0.43}{3}$	≈0.0086	≈0.0088
$\mathbb{P}(X=4) = \mathbb{P}(X=3)\frac{0.43}{4}$	≈0.0009	≈0.0010
$\mathbb{P}(X=5) = \mathbb{P}(X=4)\frac{0.43}{5}$	≈0.0001	≈0.0001
$\mathbb{P}(X \leq 5)$	≈0.9999	≈0.9999

Chaque valeur entière ≥ 6 a une probabilité inférieure à 10^{-4}. On remarque, de plus, que l'ensemble des entiers ≥ 6 constitue un événement presque impossible $\mathbb{P}(X \geq 6) = 1 - \mathbb{P}(X \leq 5) \approx 0.0001$. On peut donc décider d'arrêter la tabulation à $X = 5$.

21. Soit $P = 10^{-6}$ la probabilité, pour un individu, d'être mutant. Soit X le nombre de mutants que l'on peut observer sur n individus examinés. On a $X \sim \mathcal{B}\left(n, 10^{-6}\right)$.
Comme la probabilité P a une valeur très faible et n une valeur très élevée, on peut écrire $X \sim \mathcal{P}\left(\lambda = n10^{-6}\right)$.
Si on considère comme pratiquement sûr un événement de probabilité ≥ 0.99, on a :

$$\mathbb{P}(X \geq 1) \geq 0.99 \Rightarrow \mathbb{P}(X = 0) \leq 0.01 \Leftrightarrow e^{-\lambda} \leq 0.01 \Rightarrow \lambda \geq 4.61$$

D'où $n10^{-6} \geq 4.61$ et $n \geq 4.61\ 10^6$
Pour une probabilité ≥ 0.999 on aurait $n \geq 6.91\ 10^6$.

22. Soient X le nombre d'enfants et Y le nombre de garçons qu'une famille choisie au hasard peut contenir. On sait que $X \sim \mathcal{P}(\mu)$.
Si $\mu = 2$, X a une distribution en cloche avec 2 modes successifs : 1 et 2. Si $\mu = 3.5$, X a une distribution en cloche, de mode 3, donc déplacé sur la droite (voir 19)
Si on sait que la famille choisie contient n enfants, il est clair que le nombre de garçons suit une loi $\mathcal{B}\left(n, \frac{1}{2}\right)$. On écrit :

$$\mathbb{P}(Y = k|X = n) = C_n^k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = C_n^k \left(\frac{1}{2}\right)^n$$

avec $k \in \{0, 1, \ldots, n\}$
On a $(Y = k) = (Y = k \cap X = k) \cup (Y = k \cap X = k + 1) \cup \cdots$
Les différentes éventualités étant deux à deux incompatibles, on a :

$$\mathbb{P}(P(Y = k) \quad = \quad \sum_{i=0}^{+\infty} \mathbb{P}(X = k + i \cap Y = k)$$

$$
\begin{aligned}
&= \sum_{i=0}^{+\infty} \mathbb{P}(X=k+i)\mathbb{P}(Y=k|X=k+i) \\
&= \sum_{i=0}^{+\infty} e^{-\mu}\frac{\mu^{k+i}}{(k+i)!}C_{k+i}^{k}\left(\frac{1}{2}\right)^{k+i} \\
&= \sum_{i=0}^{+\infty} e^{-\mu}\frac{\mu^{k+i}}{(k+i)!}\frac{(k+i)!}{k!i!}\frac{1}{2^{k+i}} \\
&= e^{-\mu}\frac{\mu^k}{2^k}\frac{1}{k!}\sum_{i=0}^{+\infty}\frac{\mu^i}{2^i}\frac{1}{i!} \\
&= e^{-\mu}\frac{\left(\mu/2\right)^k}{k!}e^{\mu/2} \\
&= e^{-\mu/2}\frac{\left(\mu/2\right)^k}{k!}
\end{aligned}
$$

On en déduit

$$Y \sim \mathcal{P}\left(\frac{\mu}{2}\right)$$

23. Soit X le nombre de pièces défectueuses que l'on peut observer sur les $1\ 500\ 000$ pièces. Étant données les hypothèse, on a

$$X \sim \mathcal{B}\left(1.5\ 10^6, 10^{-7}\right)$$

Comme la probabilité, pour une pièce, d'être défectueuse est très faible et comme le nombre de pièces est très élevé, on peut écrire

$$X \sim \mathcal{P}(0.15)$$

On a

$$\mathbb{P}(X=0) = e^{-0.15} \approx 0.86$$

La fusée a donc 14 chances sur 100 de ne pas partir. Le projet n'est pas fameux, compte-tenu de prix d'une telle opération !
À partir de la loi binomiale, on obtient : $\mathbb{P}(X=0) = (1-10^{-7})^{1.5\ 10^6}$. D'où, en utilisant l'approximation $(1-x)^n \approx 1-nx$

$$\mathbb{P}(X=0) \approx 1-0.15 = 0.85$$

24. Réponses : 4.94% ; 98.66% ; 6.28%
25. Réponse : 94.37%
26. Réponse : 20.26%
27. Réponse : $\lambda = \log 0.01 \approx 4.6$

28. On va déterminer successivement $w_o(t), w_1(t), w_(2(t)$ puis, par récurrence $w_k(t)$. On note X_t la variable aléatoire ”nombre d'événement pendant t.

$$w_0(t) = \mathbb{P}(X_t = 0)$$

•

$$\begin{aligned}
w_0(t+\mathrm{d}t) &= \mathbb{P}(X_{t+\mathrm{d}t} = 0) = \mathbb{P}(X_t = 0 \cap X_{\mathrm{d}t} = 0) \\
&= \mathbb{P}(X_t = 0)\mathbb{P}(X_{\mathrm{d}t} = 0) \text{ 4}^{\text{ème}} \text{ hypothèse} \\
&= w_0(t)[1 - \mathbb{P}(X_{\mathrm{d}t} \geq 1)] = w_0(t)[1 - \mathbb{P}(X_{\mathrm{d}t} = 1) - \mathbb{P}(X_{\mathrm{d}t} > 1)] \\
&= w_0(t)(1 - \lambda\mathrm{d}t - \epsilon\mathrm{d}t)
\end{aligned}$$

où $\epsilon \longrightarrow 0$ si d$t \longrightarrow 0$ (2$^{\text{ème}}$ et 23$^{\text{ème}}$ hypothèses).
On en déduit :

$$\frac{w_0(t+\mathrm{d}t) - w_0(t)}{\mathrm{d}t} = \lambda w_0(t) - w_0(t)\epsilon$$

Et, faisant tendre dt vers zéro :

$$w_0'(t) = -\lambda w_0(t) \tag{2.2}$$

L'équation différentielle 2.2 a pour solution générale $w_0(t) = Ce^{-\lambda t}$
Par les 2$^{\text{ème}}$ et 23$^{\text{ème}}$ hypothèses $w_0(0) = 1$ donc $w_0(t) = e^{-\lambda t}$

•

$$\begin{aligned}
w_1(t+\mathrm{d}t) &= \mathbb{P}(X_{t+\mathrm{d}t} = 1) \\
&= \mathbb{P}(X_t = 1 \cap X_{\mathrm{d}t} = 0) + \mathbb{P}(X_t = 0 \cap X_{\mathrm{d}t} = 1) \text{ , événements incompatibles} \\
&= w_1(t)(1 - \lambda\mathrm{d}t - \epsilon\mathrm{d}t) + w_0(t)\lambda\mathrm{d}t
\end{aligned}$$

où $\epsilon \longrightarrow 0$ si d$t \longrightarrow 0$
On en déduit :

$$\frac{w_1(t+\mathrm{d}t) - w_1(t)}{\mathrm{d}t} = \lambda w_1(t) - w_01(t)\epsilon + \lambda w_0(t)$$

Et, faisant tendre dt vers zéro :

$$w_1'(t) = -\lambda w_1(t) + \lambda e^{-\lambda t} \tag{2.3}$$

L'équation différentielle 2.3 a pour solution générale $w_0(t) = (C + \lambda t)e^{-\lambda t}$
$w_1(0) = 0$ donc $w_1(t) = \lambda t e^{-\lambda t}$

• On a de même :

$$\begin{aligned}
w_2(t+\mathrm{d}t) &= \mathbb{P}(X_{t+\mathrm{d}t} = 2) \\
&= \mathbb{P}(X_t = 2 \cap X_{\mathrm{d}t} = 0) + \mathbb{P}(X_t = 0 \cap X_{\mathrm{d}t} = 2) + \mathbb{P}(X_t = 1 \cap X_{\mathrm{d}t} = 1) \\
&= w_2(t)(1 - \lambda\mathrm{d}t - \epsilon\mathrm{d}t) + w_1(t)\lambda\mathrm{d}t + w_0(t)\epsilon'\mathrm{d}t
\end{aligned}$$

où $\epsilon, \epsilon' \rightarrow 0$ si d$t \rightarrow 0$
On en déduit :

$$w_2'(t) = -\lambda w_2(t) + \lambda^2 t e^{-\lambda t} \tag{2.4}$$

L'équation différentielle 2.4 a pour solution générale $w_2(t) = \left(C + \frac{(\lambda t)^2}{2}\right) e^{-\lambda t}$
$w_2(0) = 0$ donc $w_2(t) = \frac{(\lambda t)^2}{2} e^{-\lambda t}$
• Plus généralement, si $w_{k-1}(t) = \frac{(\lambda t)^{k-1}}{k-1} e^{-\lambda t}$, on écrira :

$$w_k(t + \mathrm{d}t) \quad = \quad w_k(t)(1 - \lambda \mathrm{d}t - \epsilon \mathrm{d}t) + w_{k-1}(t)\lambda \mathrm{d}t + w_{k-2}(t)\epsilon' \mathrm{d}t$$

On en déduit l'équation différentielle

$$w_k'(t) = -\lambda w_k(t) + \lambda e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!}$$

de solution générale $w_k(t) = \left(C + \frac{(\lambda t)^k}{k!}\right) e^{-\lambda t}$
$w_k(0) = 0$ donc $w_k(t) = \frac{(\lambda t)^k}{k} e^{-\lambda t}$
La variable X_t suit une loi de Poisson de paramètre λt

Chapitre 3

Variables aléatoires à densité

1 Densité de probabilité

1.1 Variable aléatoire de distribution continue

Un deuxième type de distribution, pour une variable aléatoire à valeur dans $\mathbb{R}$, est caractérisé par sa fonction de répartition continue sur $\mathbb{R}$. On dit alors que la variable aléatoire est de distribution continue.
Soit X une telle variable, de fonction de répartition F_X.
Comme pour toute variable aléatoire, si t et $t+\epsilon$ ($\epsilon > 0$) sont deux réels quelconques, on a

$$\mathbb{P}(t \leq X < t+\epsilon) = F_X(t+\epsilon) - F_X(t)$$

On en déduit

$$\mathbb{P}(X = t) = \lim_{\epsilon \to 0^+} \mathbb{P}(t \leq X < t+\epsilon) = \lim_{\epsilon \to 0^+} F_X(t+\epsilon) - F_X(t)$$

La continuité de F_X entraîne ici

$$\mathbb{P}(X = t) = F_X(t) - F_X(t) = 0$$

Pour une variable aléatoire de distribution continue, la probabilité d'observer une valeur réelle t donnée est donc nulle quelle que soit t. Les événements de probabilité non nulle seront donc obligatoirement représentés par des intervalles.
Pour mieux comprendre, on peut reprendre l'analogie mécanique de la masse d'un solide. Si on découpe le solide en morceaux de plus en plus petits, on peut encore parler de la masse de chacun d'eux. À la limite, la masse d'un "point matériel", notion tout à fait abstraite, est évidemment nulle alors que la masse d'un morceau ne l'est pas.
Une telle modélisation probabiliste s'adapte, d'ailleurs très bien à la réalité ; lorsqu'on dit que la taille d'un individu, choisi au hasard dans une population donnée, est de 178 cm, il s'agit d'un abus de langage : la seule affirmation correcte est que l'observation se situe entre 177.5cm et 178.5cm.
Il résulte de ce qui précède que, si X est de distribution continue, on a

$$F_X(b) - F_X(a) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a < X < b)$$

On peut alors noter

$$\mathbb{P}\left(X \in [a, b]\right) = F_X(b) - F_X(a) \quad (a < b) \tag{3.1}$$

L'ensemble des observable $X(\Omega)$ est alors constitué d'un intervalle ou d'une réunion d'intervalles disjoints ; c'est un ensemble non dénombrable.

1.2 *Densité de probabilité*

Définition 3.1 *Soit X une variable aléatoire à valeurs dans $\mathbb{R}$, de fonction de répartition F_X. S'il existe une fonction f, fonction numérique d'une variable réelle, telle que :*

$$\forall t \in \mathbb{R} \quad , \quad F_X(t) = \int_{-\infty}^{t} f(x)dx \tag{3.2}$$

on dit que f est la densité de probabilité de X et on la note f_X. On dit également que X est une variable aléatoire à densité, ou de distribution absolument continue.

La théorie de l'intégration permet de faire quelques remarques. Si la densité f existe, alors :

1. la fonction de répartition, définie, sur $\mathbb{R}$, par une intégrale fonction de sa borne supérieure, est continue sur $\mathbb{R}$;
2. en tout point t où f admet une limite à droite $f(t+\epsilon)$ et une limite à gauche $f(t-\epsilon)$, la fonction de répartition admet une dérivée à droite égale à $f(t+\epsilon)$ et une dérivée à gauche égale à $f(t-\epsilon)$;
3. en tout point t où f est continue $(f(t-\epsilon) = f(t+\epsilon) = f(t))$, la fonction de répartition admet pour dérivée $f(t)$.

Il résulte du point 1 que toute distribution absolument continue est continue. La réciproque n'est pas vraie. Cependant, toutes les distributions continues usuelles sont absolument continues. La densité, continue presque partout, est alors obtenue par dérivation de la fonction de répartition (point 3), sauf en un nombre fini de points où f n'est pas forcément définie mais admet éventuellement des limites à gauche et à droite (point 2).

1.3 *Densité de probabilité et loi de probabilité*

Une densité de probabilité définit parfaitement la distribution d'une variable aléatoire X à valeurs dans $\mathbb{R}$. Il suffit de remarquer que, connaissant la densité f, la relation 3.2 permet de déterminer la fonction de répartition F_X. Sans chercher à déterminer F_X, on peut également écrire :

$$\begin{aligned}
\mathbb{P}(X \in [a, b]) &= F_X(b) - F_X(a) \quad \text{voir la relation 3.1} \\
&= \int_{-\infty}^{b} f(x)\mathrm{d}x - \int_{-\infty}^{a} f(x)\mathrm{d}x \quad \text{voir la relation 3.2}
\end{aligned}$$

D'où pour tout intervalle $[a,b]$ avec $a < b$:

$$\mathbb{P}(X \in [a,b]) = \int_a^b f(x)\mathrm{d}x \tag{3.3}$$

Pour une fonction de répartition on utilise un segment AB, pour une densité de probabilité, on utilise une aire de surface.
On soulignera le fait que le passage de F_X à f s'obtient par dérivation, $f(x)$ n'est pas une probabilité, mais une probabilité par unité de mesure, d'où la notion de densité de probabilité à rapprocher de la notion de la masse spécifique d'un solide.
Par contre la quantité $f(t)\mathrm{d}t$, parfois appelée probabilité élémentaire, est, lorsque dt est un infiniment petit, équivalente à : $\mathbb{P}(t \leq X < t + \mathrm{d}t) = F_X(t + \mathrm{d}t) - F_X(t)$

1.4 Propriétés d'une densité de probabilité

La densité de probabilité f d'une variable aléatoire X possède les propriétés suivantes :

1. $f(x) \geq 0$, sur $\mathbb{R}$, sauf éventuellement en un nombre fini de points. En effet, si $f(x)$ pouvait avoir une valeur négative sur un intervalle $[\alpha, \beta]$, on aurait $\mathbb{P}(X \in [\alpha,\beta]) = \int_\alpha^\beta f(x)\mathrm{d}x < 0$, $(\alpha < \beta)$, ce qui n'est pas admissible.
2. $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$. En effet :

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \lim_{b\to+\infty} \int_{-\infty}^{b} f(x)\mathrm{d}x = \lim_{b\to+\infty} F_X(b) = 1$$

   Cette deuxième propriété implique que f est intégrable sur $\mathbb{R}$, donc sur tout intervalle.

Exemple 3.1 *Soit X une variable aléatoire, à valeurs dans $\mathbb{R}$, de fonction de répartition F_X définie par :*

$$F_X : \begin{cases} t \mapsto 0 & , \textit{ si } t \leq 0 \\ t \mapsto \frac{1}{2}t & , \textit{ si } 0 < t \leq 2 \\ t \mapsto 1 & , \textit{ si } t > 2 \end{cases}$$

X est de distribution continue (F_X est continue sur $\mathbb{R}$). L'ensemble des observables constitue l'intervalle $]0,2]$ puisque la probabilité de tout intervalle extérieur à $]0,2]$ est nulle.
X admet pour densité :

$$f_X : \begin{cases} x \mapsto 0 & , \textit{ si } x \leq 0 \textit{ ou } t > 2 \\ x \mapsto \frac{1}{2} & , \textit{ si } 0 < x \leq 2 \end{cases}$$

La densité f_X s'obtient par dérivation presque partout, sauf :
- *en 0, où f_x admet pour limite à gauche 0 (dérivée à gauche de F_X) et pour limite à droite $\frac{1}{2}$ (dérivée à droite de F_X) ;*
- *en 2, où f_x admet pour limite à gauche $\frac{1}{2}$ (dérivée à gauche de F_X) et pour limite à droite 0 (dérivée à droite de F_X)*

En ces deux points, donner ou non une valeur à f_X ne modifie pas l'expression de F_X, ce qu'il est aisé de vérifier en utilisant la relation 3.2.

Exemple 3.2 *Soit Y une variable aléatoire, à valeurs dans $\mathbb{R}$, de densité de probabilité f_Y définie par :*

$$f_Y : \begin{cases} y \mapsto 0 & , \text{ si } y < 0 \\ y \mapsto \frac{1}{2}e^{-y/2} & , \text{ si } y \geq 0 \end{cases}$$

L'ensemble des observables est $[0, +\infty[$: tout intervalle à gauche de zéro a une probabilité nulle (intégrale définie de la fonction nulle). La fonction de répartition F_Y est donnée par :

$$F_Y(t) = \mathbb{P}(Y < t) = \int_{-\infty}^{t} f(y)dy$$

On obtient :

$$\text{si } t \leq 0 \; F_Y(t) = \lim_{a \to -\infty} \int_a^t 0dy = 0$$

$$\text{si } t \geq 0 \; F_Y(t) = \int_{-\infty}^{t} 0dy + \int_0^t \frac{1}{2}e^{-y/2}dy = 0 + \left[-e^{-y/2}\right]_0^t = 1 - e^{-t/2}$$

On obtient finalement :

$$F_Y : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto 1 - e^{-t/2} & , \text{ si } t \geq 0 \end{cases}$$

On remarque qu'en zéro, F_Y admet pour dérivée à droite $\frac{1}{2}$, égale à la limite à droite de f_Y, et pour dérivée à gauche 0, égale à la limite à gauche de f_Y. Le choix de $f_Y(0) = \frac{1}{2}$ est ici arbitraire et sans importance.
On peut alors calculer $\mathbb{P}(1 \leq Y < 2)$ de deux manières :

$$\begin{aligned} \mathbb{P}(1 \leq Y < 2) &= \int_1^2 \frac{1}{2}e^{-y/2}\mathrm{d}y \\ &= \left[-e^{-y/2}\right]_1^2 = e^{-1/2} - e^{-1} \end{aligned}$$

$$\begin{aligned} \mathbb{P}(1 \leq Y < 2) &= F_Y(2) - F_Y(1) \\ &= (1 - e^{-1}) - (1 - e^{-1/2}) \\ &\approx 23.86\% \end{aligned}$$

1.5 Représentation graphique

Une distribution à densité est en général représentée par le graphe de la densité de probabilité. La forme de la courbe obtenue s'appelle la forme de la distribution : uniforme (voir 3.1), en L (voir exemple 3.2), en cloche, etc.

2 Paramètres descriptifs d'une distribution à densité

Soit X une variable aléatoire, à valeurs dans $\mathbb{R}$, de densité de probabilité f. On peut associer à la distribution de X un certain nombre de paramètres descriptifs. Ces paramètres ont la même dénomination que dans le cas d'une distribution discrète parce qu'ils recouvrent les mêmes concepts.

2.1 *Mode*

On appelle mode de la variable aléatoire X, ou de la distribution de X, une valeur réelle x pour laquelle la densité de probabilité de X présente un maximum.
Les distribution à densité usuelles sont, en général, unimodales.
La distribution de l'exemple 3.1 n'a pas de mode, celle de l'exemple 3.2 a pour mode 0.

2.2 *Espérance*

On appelle espérance de la variable aléatoire X, et on note $\mathbb{E}(X)$ ou μ_X le nombre réel, s'il existe, défini par :

$$\mathbb{E}(X) = \mu_X = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$$

L'existence de $\mathbb{E}(X)$ est lié à la convergence de l'intégrale.
L'espérance de X définit la position du barycentre de l'ensemble continu des points d'une droite qui représentent l'ensemble des observables. Pour aboutir à la définition ci-dessus, on découpe cette droite en n intervalles ; au $i^{\text{ème}}$ intervalle on associe son barycentre partiel ξ_i pondéré par la probabilité de l'intervalle qui, si Δx_i est la largeur de l'intervalle, est peu différente de la probabilité élémentaire $f(\xi_i)\Delta x_i$. $\mathbb{E}(X)$ est alors peu différente de $\sum_{i=1}^{n} \xi_i f(\xi_i)\Delta x_i$; à la limite ($n \to +\infty$, $\Delta x_i \to 0$), on obtient : $\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$
On rapprochera cette définition de celle du centre de masse d'un solide à une dimension et de masse totale égale à 1, chaque point x du solide étant affecté d'une masse spécifique $f(x)$.

Exemple 3.3 *(suite de l'exemple 3.1)*

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{+\infty} xf_X(x)dx \\
&= \int_{-\infty}^{0} xf_X(x)dx + \int_{0}^{2} xf_X(x)dx + \int_{2}^{+\infty} xf_X(x)dx \\
&= 0 + \int_0^2 \frac{1}{2}xdx + 0 \\
&= \left[\frac{x^2}{4}\right]_0^2 \\
&= 1
\end{aligned}$$

Exemple 3.4 *(suite de l'exemple 3.2)*

$$\begin{aligned}
\mathbb{E}(Y) &= \int_{-\infty}^{+\infty} y f_Y(y)dy \\
&= \int_{-\infty}^{0} y f_Y(y)dy + \int_{0}^{+\infty} y f_Y(y)dy \\
&= 0 + \int_{0}^{+\infty} y \frac{1}{2} e^{-y/2} dy \\
&= \left[-ye^{-y/2}\right]_0^{+\infty} + \int_0^{+\infty} e^{-y/2}dy \quad \text{, en intégrant par parties} \\
&= \int_0^{+\infty} e^{-y/2}dy \quad (\text{car } \lim_{y\to+\infty} ye^{-y/2} = 0 \text{ et } ye^{-y/2} = 0 \text{ pour } y = 0) \\
&= \left[-2e^{-y/2}\right]_0^{+\infty} \\
&= 2
\end{aligned}$$

Exemple 3.5 *Soit Z une variable aléatoire de densité de probabilité :*

$$f_Z : t \mapsto \frac{1}{\pi}\frac{1}{1+t^2} \quad , \quad \forall t \in \mathbb{R}$$

On dit que Z suit une loi de Cauchy. Z n'a pas d'espérance ; en effet :

$$\begin{aligned}
\mathbb{E}(Z) = \int_{-\infty}^{+\infty} \frac{t}{\pi}\frac{1}{1+t^2}dt &= \int_{-\infty}^{0} \frac{t}{\pi}\frac{1}{1+t^2}dt + \int_{0}^{+\infty} \frac{t}{\pi}\frac{1}{1+t^2}dt \\
&= \frac{1}{2\pi}\left[\log(1+t^2)\right]_{-\infty}^{0} + \frac{1}{2\pi}\left[\log(1+t^2)\right]_{0}^{+\infty}
\end{aligned}$$

L'intégrale diverge puisque

$$\lim_{t\to-\infty} \log(1+t^2) = \lim_{t\to+\infty} \log(1+t^2) = +\infty$$

2.3 Moment d'ordre k

On appelle moment d'ordre k ($k \in \mathbb{N}$) de la variable aléatoire X, et on note $\mathbb{E}(X^k)$, le nombre réel s'il existe, défini par :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x^k f_X(x)\mathrm{d}x$$

On a :

$$\begin{aligned}
\mathbb{E}(X^0) &= \int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = 1 \\
\mathbb{E}(X) &= \int_{-\infty}^{+\infty} x f_X(x)\mathrm{d}x = \mathbb{E}(X)
\end{aligned}$$

2.4 Variance

On appelle variance de la variable aléatoire X, et on note $\mathbb{V}(X)$, le nombre réel s'il existe, défini par :

$$\mathbb{V}(X) = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f_X(x)\mathrm{d}x$$

La variance de X s'appelle également moment centré d'ordre 2 de X et se note :

$$\mathbb{E}\left[(x-\mu_X)^2\right] = \mathbb{E}\left[(x-\mathbb{E}(X))^2\right]$$

On note que la variance d'une variable à densité, définie comme l'intégrale d'une fonction positive presque partout et même strictement positive sur certains intervalles ne peut avoir qu'une valeur strictement positive.
Une valeur élevée de la variance signifie que les valeurs éloignées de l'espérance ont une forte densité de probabilité. Une faible valeur de la variance signifie que les valeurs de forte densité de probabilité sont situées près de l'espérance. La variance est donc un paramètre de dispersion autour de l'espérance. L'analogie mécanique de la variance est le moment d'inertie d'un solide à une dimension par rapport à son centre de masse.
On peut écrire

$$\begin{aligned}
\mathbb{V}(X) &= \int_{-\infty}^{+\infty} (x^2 - 2x\mu_X + \mu_X^2) f_X(x)\mathrm{d}x \\
&= \int_{-\infty}^{+\infty} x^2 f_X(x)\mathrm{d}x - 2\mu_X \int_{-\infty}^{+\infty} x f_X(x)\mathrm{d}x + \mu_X^2 \int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x \\
&= \mathbb{E}(X^2) - 2\mu_X^2 + \mu_X^2 \\
&= \mathbb{E}(X^2) + \mu_X^2
\end{aligned}$$

On retrouve la relation très commode dans le calcul de la variance (théorème de Guldin) :

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

Écart-type

On appelle écart-type de X la racine carrée de sa variance. On le note σ_X.
Un écart-type est donc un réel positif ; il s'agit d'un indice de dispersion qui présente l'avantage d'avoir la même unité que les observables.

Exemple 3.6 *(suite de l'exemple 3.1)*

$$\begin{aligned}
\mathbb{V}(X) &= \int_{-\infty}^{+\infty} (t-1)^2 f_X(t)dt \\
&= \int_0^2 (t^2 - 2t + 1)\frac{1}{2}dt \\
&= \frac{1}{2}\left[\frac{t^3}{3} - t^2 + t\right]_0^2
\end{aligned}$$

$$= \frac{1}{3}$$

et donc

$$\sigma_X = \frac{\sqrt{3}}{3}$$

Exemple 3.7 *(suite de l'exemple 3.2)*

$$\mathbb{E}(Y^2) = \int_{-\infty}^{+\infty} u^2 f_Y(u) du = \int_0^{+\infty} u^2 \frac{1}{2} e^{-u/2} du$$

On intègre par parties :

$$\begin{aligned} E(Y^2) &= \left[-u^2 e^{-u/2}\right]_0^{+\infty} + 2\int_0^{+\infty} u e^{-u/2} du \\ &= 0 + 4\int_0^{+\infty} \frac{1}{2} u e^{-u/2} du \\ &= 4\mathbb{E}(Y) \\ &= 8 \end{aligned}$$

En effet $\lim_{u\to+\infty} -u^2 e^{-u/2} = 0$ *et* $u^2 e^{-u/2} = 0$ *pour* $u = 0$
Finalement :

$$\mathbb{V}(Y) = \mathbb{E}(Y^2) - \mathbb{E}^2(Y) = 8 - 4 = 4 \text{ , } \sigma_Y = 2$$

Exemple 3.8 *(suite de l'exemple 3.5) Z, n'ayant pas d'espérance, n'a pas de variance. On peut d'ailleurs remarquer que, l'intégrale qui définit $\mathbb{E}(Z)$ divergeant, l'intégrale qui définit $\mathbb{E}(Z^2)$ diverge a fortiori.*

Mode et espérance sont les paramètres de position d'une distribution qui sont les plus utilisées. Il sera fait appel, ultérieurement, à un troisième paramètre de position qui, lui aussi, décrit une certaine tendance centrale : la médiane. L'idée de départ est de définir un nombre qui sépare l'ensemble des observables, rangés par valeurs croissantes, en deux parties d'égales probabilité ; on rencontre dans cette démarche deux difficultés : le partage idéal en deux zones équiprobables n'est pas toujours possible (variables discrètes) et le nombre cherché n'est pas toujours unique (on peut obtenir un intervalle). On retiendra cependant la définition suivante :

Définition 3.2 *On appelle médiane d'une variable aléatoire X, à valeurs dans $\mathbb{R}$ et de distribution continue un nombre η vérifiant la propriété :*

$$\mathbb{P}(X < \eta) = \mathbb{P}(X > \eta) = \frac{1}{2}$$

Pour les distribution à densité usuelles la médiane est unique.

3 Loi uniforme

Définition 3.3 *On dit qu'une variable aléatoire X, à valeurs réelles, est de loi uniforme sur $[a, b]$, a et b étant deux réels vérifiant : $a < b$, si sa densité de probabilité f est définie par :*

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x \notin [a, b] \\ x \mapsto \frac{1}{b-a} & , \text{ si } x \in [a, b] \end{cases}$$

On dit que la loi est uniforme sur $[a, b]$

3.1 Fonction de répartition

En utilisant la relation 3.2, on a :

$$\mathbb{P}(X < t) = F_X(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x$$

On en déduit :

$$\begin{aligned} \text{si } t \leq a, &\qquad F_X(t) = \int_{-\infty}^{t} 0\mathrm{d}x = 0 \\ \text{si } a \leq t \leq b, &\qquad F_X(t) = \int_{-\infty}^{a} 0\mathrm{d}x + \int_{a}^{t} \frac{1}{b-a}\mathrm{d}x = \frac{t-a}{b-a} \\ \text{si } t \geq b, &\qquad F_X(t) = \int_{-\infty}^{a} 0\mathrm{d}x + \int_{a}^{b} \frac{1}{b-a}\mathrm{d}x + \int_{b}^{t} 0\mathrm{d}x = 1 \end{aligned}$$

3.2 Paramètres descriptifs

Une distribution uniforme n'a pas de mode.
Sa médiane η est définie par $F_X(\eta) = \frac{1}{2} = \frac{\eta-a}{b-a} \Rightarrow \eta = \frac{a+b}{2}$

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x = \int_{a}^{b} \frac{x}{b-a}\mathrm{d}x = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{2}\frac{1}{b-a}(b^2-a^2) = \frac{a+b}{2}$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2f(x)\mathrm{d}x = \int_{a}^{b} \frac{x^2}{b-a}\mathrm{d}x = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{3}\frac{1}{b-a}(b^3-a^3) = \frac{1}{3}(b^2+ab+a^2)$$

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{1}{3}(b^2+ab+a^2) - \frac{1}{4}(b^2+2ab+b^2) = \frac{(b-a)^2}{12}$$

L'exemple 3.1 représente une loi uniforme sur $[0, 2]$.

4 Loi exponentielle

Définition 3.4 *On dit qu'une variable aléatoire X, à valeurs dans $\mathbb{R}$, suit une loi exponentielle de paramètre λ si sa densité de probabilité f est définie par :*

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x < 0 \\ x \mapsto \lambda e^{-\lambda x} & , \text{ si } x \geq 0 \end{cases}$$

où λ est un réel donné, strictement positif.

4.1 *Fonction de répartition*

On a :

$$\mathbb{P}(X < t) = F_X(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x$$

On en déduit :

$$\text{si } t \leq 0, \qquad F_X(t) = \int_{-\infty}^{t} 0\mathrm{d}x = 0$$

$$\text{si } t \geq 0, \qquad F_X(t) = \int_{-\infty}^{0} 0\mathrm{d}x + \int_{0}^{t} \lambda e^{-\lambda x}\mathrm{d}x = 0 + \left[-e^{-\lambda x}\right]_0^t = 1 - e^{-\lambda t}$$

4.2 *Paramètres descriptifs*

Une distribution exponentielle a pour mode zéro.
Sa médiane est définie par

$$F_X(\eta) = \frac{1}{2} = 1 - e^{-\lambda\eta} \Rightarrow e^{-\lambda\eta} = \frac{1}{2} \Rightarrow -\lambda\eta = -\log 2 \Rightarrow \eta = \frac{\log 2}{\lambda}$$

$$\mathbb{E}(X) = \int_0^{+\infty} x\lambda e^{-\lambda x}\mathrm{d}x = \left[-xe^{-\lambda x}\right]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x}\mathrm{d}x = 0 + \frac{1}{\lambda}\int_0^{+\infty} \lambda e^{-\lambda x}\mathrm{d}x = \frac{1}{\lambda}$$

En effet $\lim_{x\to+\infty} -xe^{-\lambda x} = 0$, si $\lambda > 0$ et $xe^{-\lambda x} = 0$ pour $x = 0$.
On remarque de plus que $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \int_{-\infty}^{+\infty} \lambda e^{-\lambda x}\mathrm{d}x = 1$

$$\mathbb{E}(X^2) = \int_0^{+\infty} x^2\lambda e^{-\lambda x}\mathrm{d}x = \left[-x^2e^{-\lambda x}\right]_0^{+\infty} + \int_0^{+\infty} 2xe^{-\lambda x}\mathrm{d}x = 0 + \frac{2}{\lambda}\int_0^{+\infty} \lambda xe^{-\lambda x}\mathrm{d}x = \frac{2}{\lambda^2}$$

En effet $\lim_{x\to+\infty} -x^2e^{-\lambda x} = 0$, si $\lambda > 0$ et $x^2e^{-\lambda x} = 0$ pour $x = 0$. On remarque de plus que la dernière intégrale est égale à $\mathbb{E}(X)$ et donc à $\frac{1}{\lambda}$

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

$\Rightarrow \sigma_X = \frac{1}{\lambda}$
L'exemple 3.2 concernait une distribution exponentielle de paramètre $\frac{1}{2}$.

5 Loi normale

Définition 3.5 *On dit qu'une variable aléatoire X, à valeurs dans $\mathbb{R}$, suit une loi normale, ou de Laplace-Gauss, si sa densité de probabilité f est définie par*

$$f : x \mapsto \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \forall x \in \mathbb{R}, \text{ où } \mu \in \mathbb{R} \text{ et } \sigma \in \mathbb{R}_+^* \text{ sont deux réels donnés}$$

On dit également que X est une variable aléatoire normale.

5.1 *Paramètres descriptifs*

Le mode, la médiane et l'espérance d'une variable aléatoire normale X ont pour valeur commune μ. En effet :

$$f'(x) = \frac{1}{\sigma\sqrt{2\pi}} \left(-\frac{x-\mu}{\sigma}\right) e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$f'(x) > 0$ pour $x < \mu$, $f'(x) = 0$ pour $x = \mu$ et $f'(x) < 0$ pour $x > \mu$. $f(x)$ présente donc un maximum pour μ.
La droite d'équation $x = \mu$ est un axe de symétrie pour le graphe de f ($\forall a > 0$, $f(\mu + a) = f(\mu - a)$) ; cette droite sépare donc l'aire de la surface comprise entre la courbe et l'axe des abcisses en deux parts égales ; d'où : $\mathbb{P}(X < \mu) = \mathbb{P}(X > \mu) = \frac{1}{2}$.

$$\mathbb{E}(xX) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x$$

On effectue le changement de variable $t = \frac{x-\mu}{\sigma}$ $\left(\Rightarrow x = \sigma t + \mu \text{ et } \mathrm{d}t = \frac{\mathrm{d}x}{\sigma}\right)$. On obtient :

$$\begin{aligned}
\mathbb{E}(X) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma t + \mu) e^{-t^2/2} \mathrm{d}t \\
&= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t e^{-t^2/2} \mathrm{d}t + \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \mathrm{d}t \\
&= \frac{\sigma}{\sqrt{2\pi}} \left[-e^{-t^2/2}\right]_{-\infty}^{+\infty} + \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \mathrm{d}t \\
&= 0 + \mu
\end{aligned}$$

En effet $\lim_{t\to\pm\infty} -e^{-t^2/2} = 0$. On admet de plus que $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \mathrm{d}t = 1$; ce résultat que l'on démontrera ultérieurement est évident a priori puisqu'il s'agit de l'intégrale sur $\mathbb{R}$ de la densité de probabilité d'une variable aléatoire normale avec $\mu = 0$ et $\sigma = 1$.

La variance de la variable aléatoire normale X est égale à σ^2. En effet :

$$\mathbb{V}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x$$

On effectue le changement de variable $t = \frac{x-\mu}{\sigma}$ $\left(\Rightarrow x = \sigma t + \mu \text{ et } \mathrm{d}t = \frac{\mathrm{d}x}{\sigma}\right)$. On obtient :

$$\mathbb{V}(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma^2 t^2 e^{-t^2/2} \mathrm{d}t$$

On intègre par parties $\left(\int u\mathrm{d}v = uv - \int v\mathrm{d}u\right)$ en posant $u = t$, $\mathrm{d}v = te^{-t^2/2}\mathrm{d}t$:

$$\mathbb{V}(X) = \frac{\sigma^2}{\sqrt{2\pi}} \left[-te^{-t^2/2}\right]_{-\infty}^{+\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-t^2/2}\mathrm{d}t$$

En notant que $\lim_{t\to\pm\infty} te^{-t^2/2} = 0$ et $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\mathrm{d}t = 1$, on obtient

$$\mathbb{V}(X) = \sigma^2$$

5.2 Notation

On comprend maintenant l'utilisation des symboles "réservés" μ et σ pour les deux paramètres présents dans l'expression de la densité. Une loi normale est complètement définie par ces deux paramètres ; la variance intervenant plus souvent que l'écart-type dans les modèles probabilistes, on utilisera la notation : X suit une loi $\mathcal{N}\left(\mu, \sigma^2\right)$

5.3 Forme de la distribution

L'étude des variations de f montre que la distribution est en cloche, symétrique par rapport à $x = \mu$, dont la dispersion augmente avec σ

Définition 3.6 *Si X^* est une variable aléatoire de loi $\mathcal{N}(0, 1)$, on dit que X^* est une variable aléatoire normale, réduite (ou normale, centrée, réduite voir définition 4.2). Sa densité est alors définie par*

$$f^*(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad , \quad \forall x \in \mathbb{R}$$

5.4 Calcul des probabilités

Si a et b sont deux réels quelconques $(a < b)$, comment calculer $\mathbb{P}\left(X \in [a, b]\right)$ lorsque X suit une loi $\mathcal{N}\left(\mu, \sigma^2\right)$? L'utilisation de la relation 3.1 : $\mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a)$ ou de la relation 3.3 : $\mathbb{P}(X \in [a, b]) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x$ nécessiterait de connaître l'expression analytique de $F_X(t)$ ou de disposer de tables à triple entrée (il faut se donner μ, σ, t)
Le changement de variable $t = \frac{x-\mu}{\sigma}$, déjà introduit précédemment va permettre une grande simplification. On obtient en effet :

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{t^2}{2}} \mathrm{d}t$$

On reconnaît, sous le signe intégral, la densité de probabilité d'une variable aléatoire X^* de loi $\mathcal{N}(0,1)$. On a donc :

$$\mathbb{P}\left(X \in [a,b]\right) = \mathbb{P}\left(X^* \in \left[\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}\right]\right) \tag{3.4}$$

La relation 3.4 permet de ramener tout calcul de probabilité concernant une loi $\mathcal{N}\left(\mu,\sigma^2\right)$ à un calcul de probabilité concernant une loi $\mathcal{N}(0,1)$. Certes, on ne connaît pas d'expression analytique pour les primitives de $f^* \mapsto \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$; les intégrales définies de cette fonction continue sont cependant calculables par des méthodes numériques. Il suffit de disposer de tables de valeurs numériques (tables 11.6, 11.7, 11.8) pour résoudre le problème posé ; ces tables, à une seule entrée, sont aisément utilisables.

5.5 *Utilisation de la table 11.6*

On pose, pour x positif : $\int_0^x \frac{1}{\sqrt{2\pi}}^{-\frac{t^2}{2}}\mathrm{d}t = \frac{1}{2}\theta(x)$
Soit maintenant X une variable aléatoire de loi $\mathcal{N}(2,4)$:

$$\begin{aligned}
\mathbb{P}(3 \leq X < 4) &= \mathbb{P}\left(\frac{1}{2} \leq X^* < 1\right)\\
&= \frac{1}{2}\theta(1) - \frac{1}{2}\theta\left(\frac{1}{2}\right)\\
&\approx 0.3413 - 0.1915\\
&\approx 0.1498
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(0 \leq X \leq 1) &= \mathbb{P}\left(-1 \leq X^* \leq -\frac{1}{2}\right)\\
&= \frac{1}{2}\theta(1) - \frac{1}{2}\theta\left(\frac{1}{2}\right)\\
&\approx 0.1498
\end{aligned}$$

à cause de la symétrie par rapport à zéro de la distribution $\mathcal{N}(0,1)$

$$\begin{aligned}
\mathbb{P}(0 < X \leq 3) &= \mathbb{P}\left(-1 < X^* \leq \frac{1}{2}\right)\\
&= \frac{1}{2}\theta\left(\frac{1}{2}\right) + \frac{1}{2}\theta(1)\\
&\approx 0.1915 + 0.3413\\
&\approx 0.5328
\end{aligned}$$

la symétrie de la distribution impose en effet $\int_{-1}^0 f^*(t)\mathrm{d}t = \int_0^1 f^*(t)\mathrm{d}t$

$$\begin{aligned}
\mathbb{P}(-0.4 < X < 4.4) &= \mathbb{P}\left(-1.2 < X^* < 1.2\right) \\
&= 2\frac{1}{2}\theta\left(1.2\right) \\
&\approx 0.7698
\end{aligned}$$

5.6 *Utilisation de la table 11.7*

Il s'agit là d'une table de la fonction de répartition de la variable X^*. Elle donne, pour μ positif : $F_{X^*}(u) = \mathbb{P}(X^* < u)$.
La symétrie de la distribution de X^* par rapport à zéro permet d'écrire, pour u négatif :

$$F_{X^*}(u) = \mathbb{P}(X^* < u) = \mathbb{P}(X^* > -u) = 1 - F_{X^*}(-u)$$

Si X suit une loi $\mathcal{N}(2, 4)$:

$$\begin{aligned}
\mathbb{P}(3 \leq X \leq 4) &= \mathbb{P}\left(\frac{1}{2} \leq X^* < 1\right) \\
&= F_{X^*}(1) - F_{X^*}\left(\frac{1}{2}\right) \\
&\approx 0.8413 - 0.6915 \\
&\approx 0.1498
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(0 \leq X \leq 1) &= \mathbb{P}\left(-1 \leq X^* \leq -\frac{1}{2}\right) \\
&= F_{X^*}\left(-\frac{1}{2}\right) - F_{X^*}\left(-1\right) \\
&= 1 - F_{X^*}\left(\frac{1}{2}\right) - \left(1 - F_{X^*}\left(-1\right)\right) \\
&= F(1) - F\left(\frac{1}{2}\right) \\
&\approx 0.1498
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(0 < X \leq 3) &= \mathbb{P}\left(-1 < X^* \leq \frac{1}{2}\right) \\
&= F_{X^*}\left(\frac{1}{2}\right) - F_{X^*}(-1) \\
&= F_{X^*}\left(\frac{1}{2}\right) - \left(1 - F_{X^*}(1)\right)
\end{aligned}$$

$$\begin{aligned} &\approx 0.6915 + 0.8413 - 1 \\ &\approx 0.5328 \end{aligned}$$

$$\begin{aligned} \mathbb{P}(-0.4 < X < 4.4) &= \mathbb{P}\left(-1.2 < X^* < 1.2\right) \\ &= F_{X^*}(1.2) - F_{X^*}(-1.2) \\ &= F_{X^*}(1.2) - (1 - F_{X^*}(1.2)) \\ &= 2F_X(1.2) - 1 \\ &= 2 \times 0.8849 - 1 \\ &\approx 0.7698 \end{aligned}$$

5.7 Utilisation de la table 11.8

Cette table est d'une grande utilité en statistique. Si u est un réel positif et si on pose $\alpha = \mathbb{P}(|X^*| \geq u)$, la table donne u en fonction de α. On note que :

$$\begin{aligned} \alpha &= \mathbb{P}(|X^*| \geq u) \\ &= \mathbb{P}\left[(X^* \geq u) \cup (X^* \leq -u)\right] \\ &= \mathbb{P}\left(X^* \geq u\right) + \mathbb{P}\left(X^* \leq -u\right) \quad \text{événements incompatibles} \\ &= 2\mathbb{P}\left(X^* \geq u\right) = 2\mathbb{P}\left(X^* \leq -u\right) \quad \text{symétrie de la distribution} \\ &= 1 - \mathbb{P}(-u < X < u) \quad \text{événement contraire} \end{aligned}$$

X suivant une loi $\mathcal{N}\left(\mu, \sigma^2\right)$, on retiendra les résultats suivants :

$$\begin{aligned} \mathbb{P}(|X^*| \geq 1.645) = 10\% &= \mathbb{P}\left[(X^* \geq 1.645) \cup (X^* \leq -1.645)\right] \\ &= \mathbb{P}\left[(X \geq \mu + 1.645\sigma) \cup (X \leq \mu - 1.645\sigma)\right] \end{aligned}$$

De la symétrie de distribution, on déduit alors :

$$\begin{aligned} \mathbb{P}(X^* \geq 1.645) &= \mathbb{P}\left(X \geq \mu + 1.645\sigma\right) = 5\% \\ \mathbb{P}(X^* \leq -1.645) &= \mathbb{P}(X \leq \mu - 1.645\sigma) = 5\% \end{aligned}$$

De même

$$\begin{aligned} \mathbb{P}(|X^*| \geq 1.960) = 5\% &= \mathbb{P}\left[(X^* \geq 1.960) \cup (X^* \leq -1.960)\right] \\ &= \mathbb{P}\left[(X \geq \mu + 1.960\sigma) \cup (X \leq \mu - 1.960\sigma)\right] \end{aligned}$$

D'où on peut déduire :

$$\begin{aligned} \mathbb{P}(X^* \geq 1.960) &= \mathbb{P}\left(X \geq \mu + 1.960\sigma\right) = 2.5\% \\ \mathbb{P}(X^* \leq -1.960) &= \mathbb{P}(X \leq \mu - 1.960\sigma) = 2.5\% \end{aligned}$$

5.8 Importance de la loi normale

La loi normale est certainement la loi de probabilité la plus utilisée. Comme on le verra ultérieurement, elle intervient souvent comme loi limite vers laquelle convergent certains modèles : on peut citer la moyenne de n mesures, indépendantes et de même loi, dont la distribution, lorsque n augmente indéfiniment, se rapproche de plus en plus d'une distribution normale.
Introduite comme "loi normale des erreurs", d'où son nom, la loi normale semble décrire assez bien la distribution de certains caractères biométriques, par exemple la taille d'un individu choisi au hasard dans une population donnée.
La loi normale est aussi à l'origine du développement de modèles probabilistes, les modèles gaussiens, dont on abordera l'étude et l'utilisation plus loin : lois du chi-deux, lois de Student, lois de Fisher-Snedecor,etc.
Dans un premier temps, on l'utilisera essentiellement comme approximation des lois, discrètes, binomiale et de Poisson.

5.9 Approximation d'une loi binomiale

L'approximation a pour origine le théorème de De Moivre-Laplace :

$$\lim_{n\to+\infty} \frac{C_n^k p^k (1-p)^{n-k}}{\frac{1}{\sqrt{2\pi}\sqrt{np(1-p)}} e^{-\frac{1}{2}\left(\frac{k-np}{\sqrt{np(1-p)}}\right)^2}} = 1$$

si $k \to +\infty$ avec n de telle manière que $\frac{|k-np|}{\sqrt{np(1-p)}}$ demeure borné.
Si X est une variable aléatoire de loi $\mathcal{B}(n,p)$ et si on désigne par f la densité de probabilité d'une loi $\mathcal{N}(np, np(1-p))$, le théorème ci-dessus peut être interprété de la manière suivante : lorsque n est suffisamment grand et pour les valeurs entières k situées dans la zone proche de l'espérance np, de X, où se trouve concentrée toute la probabilité, on a :

$$\mathbb{P}(X = k) \approx f(k) \tag{3.5}$$

Deux précisions sont à ajouter :
- si $p = \frac{1}{2}$ (distribution binomiale de symétrie parfaite), la "convergence" est très rapide ;
- si $p \neq \frac{1}{2}$ (distribution binomiale dissymétrique), la "convergence" est d'autant plus lente que p est loin de $\frac{1}{2}$.

Remplacer une expression analytique par une autre n'est pas forcément d'un grand intérêt. Une deuxième étape va permettre de simplifier au maximum les calculs et conduire, de fait, à une meilleure approximation.
En statistique, c'est le calcul de probabilités affectées à des intervalles qui présentent un intérêt. Si X suit une $\mathcal{B}(n,p)$, on se propose de calculer $\mathbb{P}(a \le X \le b)$. On a

$$\mathbb{P}(a \le X \le b) = \sum_{(a\le k\le b)} \mathbb{P}(X = k) \quad , \quad a, b, k \in \{0, 1, 2, \ldots, n\}$$

Si on considère l'histogramme de la loi de X, cette somme est aussi donnée par la somme des aires des rectangles de largeur 1. Il est alors clair que :

$$\mathbb{P}(a \leq X \leq b) \approx \int_{a-0.5}^{b+0.5} f(x)\mathrm{d}x$$

cette dernière intégrale, qui porte sur la densité de probabilité d'une loi $\mathcal{N}(np, np(1-p))$ se calcule en utilisant une variable normale réduite X^* :

$$\mathbb{P}(a \leq X \leq b) \approx \mathbb{P}\left(\frac{a-0.5-np}{\sqrt{np(1-p)}} \leq X^* \leq \frac{b+0.5-np}{\sqrt{np(1-p)}}\right) \tag{3.6}$$

Dans la pratique, si les deux moitiés de rectangle situées à droite et à gauche de l'intervalle $[a, b]$ ont une aire négligeable par rapport à l'ensemble, on se permettra d'écrire :

$$\mathbb{P}(a \leq X \leq b) \approx \int_a^b f(x)\mathrm{d} = \mathbb{P}\left(\frac{a-np}{\sqrt{np(1-p)}} \leq X^* \leq \frac{b-np}{\sqrt{np(1-p)}}\right) \tag{3.7}$$

En conclusion : si n est suffisamment grand et si la distribution n'est pas trop dissymétrique, une variable aléatoire X de loi $\mathcal{B}(n, p)$ peut être considérée comme une variable aléatoire de loi $\mathcal{N}(np, np(1-p))$.
Pour le calcul des probabilités, on peut utiliser 3.7 ou 3.6. Lorsqu'on utilise 3.6, ce qui améliore la précision de l'approximation, on dit qu'on effectue une "correction de continuité". Le passage de 3.7 à 3.6 s'effectue très aisément si on remarque que, puisque X est à valeurs entières, on a toujours, pour a et b entiers :

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a - 0.5 \leq X \leq b + 0.5)$$

c'est dans le calcul approché, où X est considéré comme une variable à densité, que les deux membres de l'égalité conduisent à des résultats différents.
Quand peut-on décider de la validité de l'approximation ? On utilise très souvent le critère suivant, d'origine empirique :

$$np \geq 10 \text{ et } n(1-p) \geq 10$$

qui permet de tenir compte de la valeur de n et de la dissymétrie. On obtient par exemple : si $p = \frac{1}{2}$, $n \geq 20$; si $p = \frac{1}{4}$, $n \geq 40$; si $p = 0.9$, $n \geq 100$, etc. bien entendu, pour p donné, plus n est grand, meilleure est l'approximation.
Quand est-il nécessaire d'effectuer la correction de continuité ? En réalité, puisque cette correction donne de meilleurs résultats, il faut se poser la question autrement : quand peut-on abandonner la correction de continuité ? La réponse est simple en théorie : dès que la modification qu'elle apporte est négligeable. C'est le cas

- lorsque n est très grand ; la correction de continuité déplace les bornes concernant X^* de $\pm\frac{0.5}{\sqrt{np(1-p)}}$. Un déplacement de 0.01 (qui correspond pour $p = \frac{1}{2}$ à $n = 10000$) a peu d'importance.

- lorsque ces bornes sont situées dans une zone de faible probabilité (consulter les tables).
- lorsque l'intervalle $[a, b]$ contient beaucoup d'entiers (faible erreur relative). On remarquera à ce propos que pour $a = b$, la formule 3.6 donne :

$$\mathbb{P}(X = a) \approx \mathbb{P}\left(\frac{a - 0.5}{\sqrt{np(1-p)}} \leq X \leq \frac{a + 0.5}{\sqrt{np(1-p)}}\right) \tag{3.8}$$

Il est évident qu'alors, quel que soit n, il n'est pas question de supprimer ± 0.5.

Pour terminer, on attirera l'attention sur deux points

- La variable aléatoire X, de loi $\mathcal{B}(n, p)$, est une variable discrète. Il est donc important de préciser si les bornes de l'intervalle sont ou non inclues dans l'intervalle ; ceci peut modifier en particulier le signe du terme correctif 0.5.
- Il n'est pas choquant d'approcher une distribution sur un ensemble fini de valeurs par une distribution sur $\mathbb{R}$ puisqu'on sait que pour les lois $\mathcal{B}(n, p)$ et $\mathcal{N}(np, np(1-p))$ pratiquement toute la probabilité est concentrée autour de l'espérance np.

Exemple 3.9 *Soit X de loi $\mathcal{B}\left(20, \frac{1}{2}\right)$. Le tableau ci-dessous donne une idée de la validité de l'approximation par une loi normale.*

	$\mathbb{P}(X = k)$	
k	*Loi binomiale*	*Loi normale (eq. 3.8)*
10	*0.176197*	*0.176929*
9 ou 11	*0.160179*	*0.160367*
8 ou 12	*0.120134*	*0.119390*
7 ou 13	*0.073929*	*0.073013*
6 ou 14	*0.036964*	*0.036678*
etc.		

Exemple 3.10 *On joue 10000 fois à pile ou face. Calculer la probabilité pour que le nombre de piles soir dans l'intervalle* $[4900, 5100]$.
Soit X le nombre de piles. X suit une loi $\mathcal{B}\left(10000, \frac{1}{2}\right)$ que l'on peut approcher par une loi $\mathcal{N}(5000, 2500)$. On obtient, en désignant par X^* une variable $\mathcal{N}(0, 1)$:

- *sans correction de continuité :*

$$\mathbb{P}(4900 \leq X \leq 5100) = \mathbb{P}(-2 \leq X^* \leq 2) \approx 95.45\%$$

- *avec correction de continuité :*

$$\mathbb{P}(4900 \leq X \leq 5100) = \mathbb{P}(4899.5 \leq X \leq 5100.5) = \mathbb{P}(-2.01 \leq X^* \leq 2.01) \approx 95.55\%$$

Exemple 3.11 *Soit X de loi $\mathcal{B}(100, 0.3)$. Estimer $\mathbb{P}(24 \leq X \leq 29)$.*
On considère que X suit une loi $\mathcal{N}(30, 21)$ (puisque $n(1-p) > 10$ et $np > 10$)
On obtient :

- *sans correction de continuité :*

$$\mathbb{P}(24 \leq X \leq 29) = \mathbb{P}(-1.3093 \leq X^* \leq -0.2182) \approx 0.3145$$

- *avec correction de continuité :*

$$\mathbb{P}(23.5 \leq X \leq 29.5) = \mathbb{P}(-1.4184 \leq X^* \leq -0.1091) \approx 0.3785$$

Le résultat exact est ≈ 0.3868.

5.10 Approximation d'une loi de Poisson

On montre que si X est une variable aléatoire de loi $\mathcal{P}(\lambda)$, on peut considérer que X est de loi $\mathcal{N}(\lambda, \lambda)$ lorsque λ a une valeur suffisamment élevée.
Pour décider de la validité de l'approximation, le critère, d'origine empirique, généralement utilisé est : $\lambda \geq 20$.
Comme dans le cas précédent, l'approximation revient à remplacer des sommes de probabilités par des intégrales ; il est donc également nécessaire d'introduire une correction de continuité.

Exemple 3.12 *Soit X de loi $\mathcal{P}(100)$. Évaluer $\mathbb{P}(90 \leq X \leq 100)$.*
X suit pratiquement une loi $\mathcal{N}(100)$ et on peut écrire si X^* est de loi $\mathcal{N}(0,1)$, et en effectuant la correction de continuité :

$$\mathbb{P}(89.5 \leq X \leq 100.5) = \mathbb{P}(-1.05 \leq X^* \leq 1.05) \approx 0.7063$$

En conclusion on retiendra le schéma ci-dessous qui résume les principales approximations :

$$\begin{tikzcd}
 & X \sim \mathcal{B}(n,p) \arrow[dl, "\begin{array}{l} n \geq 100 \\ p \leq 0.1 \end{array}"'] \arrow[dr, "\begin{array}{l} np \geq 10 \\ np(1-p) \geq 10 \end{array}"] & \\
X \sim \mathcal{P}(\lambda = np) & & X \sim \mathcal{N}(np, np(1-p)) \\
X \sim \mathcal{P}(\lambda) \arrow[rr, "\lambda \geq 20"'] & & X \sim \mathcal{N}(\lambda, \lambda)
\end{tikzcd}$$

6 Autres lois

Par la suite, on aura l'occasion de définir des variables aléatoires et de déterminer leur densité de probabilité. Deux distributions sont cependant intéressantes à connaître, la première parce qu'elle intervient très souvent dans les livres de probabilité et de statistique, la deuxième à cause de ses applications.

6.1 *La loi gamma*

On dit qu'une variable aléatoire X, à valeurs réelles, suit une loi gamma de paramètres n et λ, si sa densité de probabilité f est définie par :

$$f : \left\{ \begin{array}{ll} x \mapsto 0 & , \text{ si } x \leq 0 \\ x \mapsto \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} & , \text{ si } x > 0 \end{array} \right.$$

où n et λ sont deux réels donnés, strictement positif, et où $\Gamma(n)$ est une constante réelle strictement positive qui ne dépend que de n.
La constante $\Gamma(n)$ est aisément calculable. En effet, on doit avoir :

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1 \Rightarrow \int_0^{+\infty} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}\mathrm{d}x = 1 \tag{3.9}$$

D'où $\Gamma(n) = \int_0^{+\infty} \lambda^n x^{n-1} e^{-\lambda x}\mathrm{d}x = \int_0^{+\infty} t^{n-1} e^{-t}\mathrm{d}t$, en effectuant le changement de variable $t = \lambda x$. Lorsque n est un entier positif, on obtient, en intégrant par parties :

$$\Gamma(n) = (n-1)! \quad , \quad n \in \mathbb{N}^*$$

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = \int_0^{+\infty} \frac{\lambda^n}{\Gamma(n)} x^n e^{-\lambda x}\mathrm{d}x = \frac{1}{\Gamma(n)} \int_0^{+\infty} \lambda^{n-1} x^n \lambda e^{-\lambda x}\mathrm{d}x$$

On intègre par parties :

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{\Gamma(n)} \left[-\lambda^{n-1} x^n e^{-\lambda x}\right]_0^{+\infty} + \frac{n}{\Gamma(n)} \int_0^{+\infty} \lambda^{n-1} x^{n-1} e^{-\lambda x}\mathrm{d}x \\ &= \frac{1}{\Gamma(n)} \left[-\lambda^{n-1} x^n e^{-\lambda x}\right]_0^{+\infty} + \frac{n}{\lambda} \int_0^{+\infty} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}\mathrm{d}x \\ &= \frac{n}{\lambda} \end{aligned}$$

En effet $\lim_{x\to+\infty} x^n e^{-\lambda x} = 0$, si $\lambda > 0$ et $x^n e^{-\lambda x} = 0$ pour $x = 0$. On remarque de plus que la dernière intégrale vaut 1 (voir la relation 3.9).

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \int_0^{+\infty} \frac{\lambda^n}{\Gamma(n)} x^{n+1} e^{-\lambda x}\mathrm{d}x = \frac{1}{\Gamma(n)} \int_0^{+\infty} \lambda^{n-1} x^{n+1} \lambda e^{-\lambda x}\mathrm{d}x$$

On intègre par parties :

$$\begin{aligned} \mathbb{E}(X^2) &= \frac{1}{\Gamma(n)} \left[-\lambda^{n-1} x^{n+1} e^{-\lambda x}\right]_0^{+\infty} + \frac{n+1}{\Gamma(n)} \int_0^{+\infty} \lambda^{n-1} x^n e^{-\lambda x}\mathrm{d}x \\ &= \frac{1}{\Gamma(n)} \left[-\lambda^{n-1} x^{n+1} e^{-\lambda x}\right]_0^{+\infty} + \frac{n+1}{\lambda} \int_0^{+\infty} \frac{\lambda^n}{\Gamma(n)} x^n e^{-\lambda x}\mathrm{d}x \end{aligned}$$

D'où finalement

$$\mathbb{E}(X^2) = 0 + \frac{n+1}{\lambda}\mathbb{E}(X) = \frac{n(n+1)}{\lambda^2}$$

On en déduit

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{n(n+1)}{\lambda^2} - \frac{n^2}{\lambda^2} = \frac{n}{\lambda^2}$$

6.2 La loi log-normale

On dit qu'une variable aléatoire X à valeurs réelles, suit une loi log-normale si sa densité de probabilité f est définie par :

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x \leq 0 \\ x \mapsto \frac{1}{ax\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x - b}{a}\right)^2} & , \text{ si } x > 0 \end{cases}$$

où $a \in \mathbb{R}_+^*$ et $b \in \mathbb{R}$ sont deux réels donnés.
La loi log-normale est utilisée en granulométrie et en biométrie pour décrire, par exemple, la distribution de la taille des grains de sédiment provenant d'une couche géologique donnée ou des cellules dans une culture donnée.
À titre d'exercice, calculons $\mathbb{E}(X)$:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} x e^{-\frac{1}{2}\left(\frac{\log x - b}{a}\right)^2} \frac{\mathrm{d}x}{ax}$$

On pose $t = \frac{\log x - b}{a} \Rightarrow \mathrm{d}t = \frac{\mathrm{d}x}{ax}$ et $x = e^{at+b}$

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2 + at + b} \mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(t^2 - 2at) + b} \mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(t-a)^2 + \frac{a^2}{2} + b} \mathrm{d}t \\ &= e^{\frac{a^2}{2} + b} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(t-a)^2} \mathrm{d}t \end{aligned}$$

On reconnaît l'intégrale sur $\mathbb{R}$ de la densité d'une $\mathcal{N}(a, 1)$ d'où finalement :

$$\mathbb{E}(X) = e^{\frac{a^2}{2} + b}$$

7 Mélanges de distributions

7.1 Introduction

On considère deux variables aléatoires X et Y, à valeurs dans $\mathbb{R}$, de fonctions de répartition respectives F_X et F_Y. Soit maintenant une expérience aléatoire à laquelle on associe la variable aléatoire Z définie ainsi : avec la probabilité p, Z a la loi de X ; avec la probabilité $(1 - p)$, Z a la loi de Y.
On note $p = \mathbb{P}(Z = X)$ et $1 - p = \mathbb{P}(Z = Y)$.

La distribution de Z peut être définie par sa fonction de répartition F_Z. On peut écrire : $(Z < t) = (Z = X \cap Z < t) \cup (Z = Y \cap Z < t)$. Les deux éventualités étant incompatibles, on obtient :

$$\begin{aligned} F_Z(t) = \mathbb{P}(Z < t) &= \mathbb{P}(Z = X \cap Z < t) + \mathbb{P}(Z = Y \cap Z < t) \\ &= \mathbb{P}(Z = X)\mathbb{P}(Z < t | Z = X) + \mathbb{P}(Z = Y)\mathbb{P}(Z < t | Z = Y) \end{aligned}$$

D'où

$$F_Z(t) = pF_X(t) + (1-p)F_Y(t) \tag{3.10}$$

On dit que la distribution de Z est un mélange de la distribution de X et de la distribution de Y. On généralise aisément à un mélange de plus de deux distributions.

7.2 *Mélanges de distribution discrètes*

Si F_X et F_Y sont en escalier, F_Z est forcément en escalier. Z est donc une variable discrète et la probabilité est répartie sur les points de discontinuité dont l'ensemble, ensemble des observables de Z, est la réunion des deux ensembles d'observables de X et de Y.
En chaque point t_i de discontinuité, on peut écrire :

$$\begin{aligned} \mathbb{P}(t_i \leq Z \leq t_i + \epsilon) &= F_Z(t_i + \epsilon) - F_Z(t_i) \\ &= p\left(F_X(t_i + \epsilon) - F_X(t_i)\right) + (1-p)\left(F_Y(t_i + \epsilon) - F_Y(t_i)\right) \\ &= p\mathbb{P}(t_i \leq X < t_i + \epsilon) + (1-p)\mathbb{P}(t_i \leq Y < t_i + \epsilon) \end{aligned}$$

si $\epsilon \longrightarrow 0^+$, on obtient :

$$\mathbb{P}(Z = t_i) = p\mathbb{P}(X = t_i) + (1-p)\mathbb{P}(Y = t_i)$$

Relation que l'on peut obtenir directement à partir de l'événement $(Z = t_i) = (Z = X \cap Z = t_i) \cup (Z = Y \cap Z = t_i)$. On déduit en particulier :

$$\mathbb{E}(Z) = \sum_i t_i \mathbb{P}(Z = t_i) = p\sum_i t_i \mathbb{P}(X = t_i) + (1-p)\sum_i t_i \mathbb{P}(Y = t_i)$$

Donc

$$\mathbb{E}(Z) = p\mathbb{E}(X) + (1-p)\mathbb{E}(Y)$$

En application, se rapporter à l'exercice 3

7.3 *Mélange de distributions à densité*

Par dérivation de la relation 3.10, on obtient dans ce cas :

$$f_Z(t) = pf_X(t) + (1-p)f_Y(t)$$

où f_Z, f_X, f_Y sont les densités respectives de Z, X et Y. On déduit en particulier :

$$\mathbb{E}(Z) = \int_{-\infty}^{+\infty} tf_Z(t)\mathrm{d}t = p\int_{-\infty}^{+\infty} tf_X(t)\mathrm{d}t + (1-p)\int_{-\infty}^{+\infty} tf_Y(t)\mathrm{d}t$$

et on retrouve donc :

$$\mathbb{E}(Z) = p\mathbb{E}(X) + (1 - p)\mathbb{E}(Y)$$

En application, se reporter aux exercices 9 et 19.

7.4 Cas intermédiaire

On suppose maintenant que X a une distribution discrète et que Y a une distribution à densité. La distribution de Z peut être s'interpréter de la manière suivante :

- une partie, p, de la probabilité de Ω est répartie sur les points de discontinuité qui constituent l'ensemble des observables de X ;
- l'autre partie, $1 - p$, de la probabilité de Ω est répartie de manière continue sur l'ensemble non dénombrable des observables de Y.

Il est aisé de montrer que, sous réserve d'existence, on a toujours :

$$\mathbb{E}(Z) = p\mathbb{E}(X) + (1 - p)\mathbb{E}(Y)$$

puisqu'on peut remplacer les points de la distribution de X par leur barycentre partiel $\mathbb{E}(X)$ affecté de la pondération globale p et procéder de même pour les points de la distribution de Y.

8 Exercices

1. Soit une variable aléatoire X, à valeurs réelles, de fonction de répartition :

$$F : \begin{cases} x \mapsto 0 & , \text{ si } x < 0 \\ x \mapsto 1 - e^{-x/2}\left(1 + \frac{x}{2}\right) & , \text{ si } x \geq 0 \end{cases}$$

   (a) F est-elle continue sur $\mathbb{R}$?
   (b) Déterminer $\lim_{x \to +\infty} F(x)$. Interpréter.
   (c) Déterminer la densité de probabilité de X.
   (d) Déterminer le mode de X.
   (e) Calculer l'espérance et l'écart-type de X.
   (f) Déduire de ce qui précède les variations et le graphe de F.
   (g) Calculer $\mathbb{P}(1 \leq X < 2)$

2. Soit une variable aléatoire T, à valeurs réelles, de densité de probabilité :

$$f : \begin{cases} t \mapsto 0 & , \text{ si } t \notin [-1, 1] \\ t \mapsto \lambda(1 - t^2) & , \text{ si } t \in [-1, 1] \end{cases}$$

   (a) calculer λ. Construire le "graphe" de f.
   (b) Déterminer la fonction de répartition de T et construire son graphe.
   (c) Calculer la probabilité de l'événement $|T| \geq \frac{1}{2}$. Représenter cette probabilité sur les deux graphes précédents.
   (d) Calculer l'espérance et la variance de T. Quelles est sa médiane ?

3. Soit une variable aléatoire X, à valeurs réelles, de densité de probabilité, définie pour $n > 1$:

$$f : \begin{cases} x \mapsto ax^{n-1} & , \text{ si } 0 \leq x < 1, a \in \mathbb{R} \\ x \mapsto 0 & , \text{ si } x \notin [0, 1[\end{cases}$$

   (a) calculer a.
   (b) Calculer l'espérance et la variance de X.
   (c) Déterminer la fonction de répartition de X.

4. Soit une variable aléatoire Y, à valeurs réelles, de densité de probabilité :

$$f : \begin{cases} y \mapsto 0 & , \text{ si } y \notin \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \\ y \mapsto \frac{2}{\pi}\cos^2 y & , \text{ si } y \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \end{cases}$$

   (a) Calculer l'espérance, l'écart-type et le mode de Y
   (b) Déterminer la fonction de répartition de X. Quelle est la médiane de Y.
   (c) Calculer la probabilité de l'événement $Y > \frac{4\pi}{10}$.

5. Soit une variable aléatoire X, à valeurs réelles, de densité de probabilité :

$$f : \begin{cases} x \mapsto a\sqrt{4 - x^2} & , \text{ si } x \in [0, 2], a \in \mathbb{R} \\ x \mapsto 0 & , \text{ si } x \notin [0, 2] \end{cases}$$

   (a) calculer a.
   (b) Calculer l'espérance, la variance et le mode de X.
   (c) Déterminer la fonction de répartition de X et tracer son graphe.
6. Soit une variable aléatoire X, à valeurs réelles, de densité de probabilité :

$$f : x \mapsto \frac{2}{\pi} \frac{1}{(1+x^2)^2} \quad , \quad x \in \mathbb{R}$$

   (a) Vérifier que f possède les propriétés d'une densité de probabilité.
   (b) Calculer l'espérance et l'écart-type de X.
   (c) Déterminer la fonction de répartition de X. vérifier ses propriétés.
   (d) Calculer $\mathbb{P}(X \geq 1)$
7. Soit l'équation différentielle

$$(1+x^2)\frac{\mathrm{d}y}{\mathrm{d}x} + 4xy = 0 \tag{3.11}$$

   Soit X une variable aléatoire, à valeurs réelles, dont la densité de probabilité f est solution de 3.11 sur $]-1,1[$ et nulle ailleurs.
   (a) Déterminer f et donner la forme de son graphe
   (b) Déterminer la fonction de répartition de X et donner la forme de son graphe.
   (c) Calculer l'espérance et la variance de X.
   (d) Calculer la probabilité des événements : $\left(X \geq \frac{1}{2}\right)$ et $\left(-\frac{1}{2} \leq X \leq \frac{1}{2}\right)$.
8. Soit l'équation différentielle

$$y'' + 2y' = 0 \tag{3.12}$$

   Soit X une variable aléatoire, à valeurs réelles, dont la fonction de répartition F possède les propriétés : (i) F est continue sur $\mathbb{R}$; F est nulle pour $x \leq 0$; F est solution de 3.12 pour $x > 0$
   (a) Déterminer F et tracer son graphe. Quelle est la médiane de X ?
   (b) Déterminer la densité de X et tracer son graphe. Quelle est le mode de X ?
   (c) Calculer l'espérance et la variance de X.
9. Au cours d'une récolte de fruits, les variétés A et B ont été mélangées. Le poids d'un fruit de la variété A est une variable aléatoire X, de fonction densité f_1, de fonction de répartition F_1, d'espérance μ_1 et de variance σ_1^2. Pour le poids de Y d'un fruit de la variété B, on a la densité f_2, la fonction de répartition F_2, l'espérance μ_2 et la variance σ_2^2.
   Les fruits sont mélangés dans les proportions p pour la variété A et q pour la variété B ($p + q = 1$). On appelle Z la variable aléatoire égale au poids d'un fruit quelconque de la récolte.
   - Déterminer la fonction de répartition F et la fonction densité f de Z ;
   - Calculer l'espérance μ et la variance σ^2 de Z.

10. Calculer l'espérance et la variance de la variance X, à valeurs réelles, de loi uniforme sur $[a, b]$. Déterminer sa fonction de répartition.
11. Soit une variable aléatoire X, à valeurs réelles, de densité de probabilité :

$$f : \begin{cases} x \mapsto 0 & , \text{ si } 0 < x \\ x \mapsto \lambda e^{-\lambda x} & , \text{ si } x \geq 0 \text{ avec } \lambda \in \mathbb{R}_+^* \end{cases}$$

   (a) Calculer l'espérance et la variance de X.
   (b) Déterminer la fonction de répartition de X.
   (c) Déterminer le mode et la médiane de X.
   (d) Pour $\lambda = \frac{1}{2}$, calculer la probabilité des événements : $(2 \leq X < 3)$, $(X \geq 5)$, $(X < 7)$.
12. On considère la variable aléatoire X, à valeurs réelles, de loi $\mathcal{N}(\mu, \sigma^2)$.
   (a) Tracer le graphe de la densité de probabilité de X ; déterminer sa largeur à mi-hauteur.
   (b) Calculer l'espérance et la variance de X.
   (c) Déterminer les nombres réels positifs, a et b, tels que :
   $\mathbb{P}(\mu - a < X < \mu + a) = 0.95$; $\mathbb{P}(\mu - b < X < \mu + b) = 0.99$
   (d) Déterminer les nombre réels c et d tels que : $\mathbb{P}(X \leq c) = \mathbb{P}(X \geq d) = 0.05$.
13. Soit la variable aléatoire normale X, d'espérance 4 et de variance 4.
   (a) Calculer la probabilité de l'événement : $(X > 6)$
   (b) Calculer la probabilité de l'événement : $(|X| > 6)$
   (c) Calculer la probabilité de l'événement : $(2 \leq X \leq 6)$
   (d) Calculer la probabilité de l'événement : $(1.7 \leq X \leq 2.82)$
14. On veut transporter rapidement entre deux villes A et B, dans des conditions de confort acceptables, 1600 voyageurs se présentant, pratiquement en même temps, à la gare A. On met à leur disposition deux trains identiques. On suppose que chaque individu choisit au hasard l'une ou l'autre rame et qu'il n'a pas le temps d'en changer.

   Combien faut-il prévoir de places assises dans chaque rame si l'on veut que la probabilité, pour que des voyageurs soient obligés de rester debout, soit inférieure à $3\ 10^{-3}$?
15. On place un compteur Geiger devant une source de rayonnement. Le nombre d'impulsions, que l'on peut enregistrer pendant une minute, est une variable aléatoire de Poisson, dont l'espérance λ mesure l'intensité du rayonnement. Si $\lambda = 10000$ coups/minute, calculer la probabilité d'observer, en une minute, un nombre d'impulsions situé dans l'intervalle : $[9800, 10200]$.
16. On admet que, dans la population française adulte, la taille d'un individu mâle, choisi au hasard, est une variable aléatoire normale d'espérance 170 cm et d'écart-type 6.2 cm.

Si on choisit au hasard 200 individus, calculer la probabilité d'obtenir au moins 5 individus de taille supérieure ou égale à 186 cm, condition pour former une équipe de basket valable.

17. On admettra, ce que l'on peut discuter, que la durée de fonctionnement correct d'un appareil donné est une variable aléatoire normale, d'espérance 1500 heures, et d'écart-type 300 heures.

    Une révision périodique le remet complètement à neuf. Au bout de combien de temps faut-il effectuer la révision pour que la probabilité de trouver l'appareil en panne soit inférieure ou égale à 0.5% ?

18. Un garage contient 900 places. Si on admet que, chaque jour, le taux d'absentéisme est de 10%, combien peut-on avoir d'abonnés pour être pratiquement sûr de ne pas laisser de voitures dehors, un jour donné ?

19. On s'intéresse à la vitesse de déplacement de piétons, sur un parcours déterminé. On admet le modèle suivant : la vitesse d'un individu isolé, choisi au hasard, est une variable aléatoire normale
    - d'espérance 1.4 m/s et d'écart-type 0.4 m/s pour une femme ;
    - d'espérance 1.6 m/s et d'écart-type 0.4 m/s pour une homme.

    La population est composée d'un tiers d'hommes et de deux tiers de femmes.

    Soit V la vitesse d'un individu isolé, choisi au hasard, dont on ignore le sexe.

    (a) Calculer la probabilité de l'événement : (V <1.2 m/S)

    (b) la vitesse d'un individu a été trouvé inférieure à 1.2m/s. Quelle est la probabilité pour qu'il s'agisse d'une femme ?

    (c) Déterminer l'expression explicite de la densité de V. Calculer l'espérance et l'écart-type de V.

20. Une entreprise comprend 900 employés. Elle dispose d'une cantine de 500 places qui assure deux services. Si chaque employé choisit indifféremment l'un ou l'autre service, quelle est la probabilité pour qu'un jour donné on refuse du monde à la cantine ?

21. On désigne par P la probabilité d'observer un phénotype donné sur un individu issu d'un certain croisement. Calculer la probabilité d'observer moins de 250 individus possédant le phénotype, sur un échantillon de 384 descendants, sous les hypothèses : $P = \frac{3}{4}$, puis $P = \frac{9}{16}$.

22. Soit X une variable aléatoire, à valeurs réelles, de densité de probabilité

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x < 0 \\ x \mapsto xe^{-x^2/2} & , \text{ si } x \geq 0 \end{cases}$$

- Déterminer le mode de X et tracer le graphe de f.
- Calculer l'espérance et la variance de X.
- Déterminer la fonction de répartition de X et calculer la probabilité de l'événement $(0 \leq X < 2)$

23. (a) On pose $I_n = \int_0^{+\infty} x^n e^{-x^2} \mathrm{d}x$, où $n \in \mathbb{N}$. On admet $I_0 = \frac{\sqrt{\pi}}{2}$. Calculer I_1. Par intégration par parties, trouver la relation entre I_{n+2} et I_n. En déduire les valeurs de I_2, I_3, I_4, I_5, I_6.

    (b) Dans la théorie de Maxwell, la vitesse d'une molécule d'un gaz donné, dans des conditions données, est une variable aléatoire V, dont la densité de probabilité f est donnée par : $f(v) = 0$ pour $v \leq 0$; $f(v) = Av^2 e^{-v^2/a^2}$ pour $v \geq 0$, où a est un paramètre positif et A un facteur de proportionnalité.
    - Calculer A en fonction de a. Déterminer le mode de V.
    - Exprimer, en fonction de a et des nombres I_n, le moment d'ordre k de V.
    - En déduire, en fonction de a, l'espérance et la variance de V.

24. La durée de vie d'un composant électronique, choisi au hasard dans un fabrication donnée, est une variable aléatoire T de densité :

$$f : \begin{cases} t \mapsto 0 & , \text{ si } t < 0 \\ t \mapsto \lambda^2 t e^{-\lambda t} & , \text{ si } t \geq 0 \end{cases}$$

où λ est un réel positif.
- Calculer le mode, l'espérance et la variance de T.
- Déterminer la fonction de répartition de la variable T.
- on pose $\lambda = 2\,10^{-4}$, t étant exprimé en heures. Calculer les probabilités P_1 de l'événement $(T < 500 \text{ h})$ et P_2 de l'événement $(T \geq 10000 \text{ h})$

(a) En posant, pour simplifier $P_1 = 0.5\%$ et $P_2 = 40\%$, calculer la probabilité de trouver, dans un lot de 600 composants choisis indépendamment :
- plus de 6 composants de durée de vie inférieur à 500h ;
- plus de 220 composants de durée de vie au moins égale à 10000h.

(b) On a mélangé 10 composants de fabrication caractérisée par $\lambda = a$ et 5 composants d'une fabrication caractérisée par $\lambda = b$. On choisit au hasard un élément de ce mélange ; soit X sa durée de vie.
- Déterminer la densité de X.
- Calculer l'espérance et la variance de X

9 Solutions

1. (a) F est définie sur $\mathbb{R}$. F est continue pour $x < 0$ et pour $x > 0$. Le problème de la continuité se pose en 0. On a $F(0) = 1 - 1 = 0$. De plus :

$$\lim_{x\to 0^-} = 0 = F(0); \lim_{x\to 0^+} = 1 - \lim_{x\to 0^+} e^{-x/2}\left(1 + \frac{x}{2}\right) = 1 - 1 = 0 = F(0)$$

F est continue en zéro et donc sur $\mathbb{R}$

(b) $\lim_{x\to+\infty} F(x) = 1 - \lim_{x\to+\infty} e^{-x/2} - \lim_{x\to+\infty} e^{-x/2}\frac{X}{2} = 0$

En effet $\lim_{x\to+\infty} e^{-x/2} = 0^+$ et $\lim_{x\to+\infty} e^{-x/2}\frac{X}{2} = 0^+$ (l'exponentielle d'une fonction puissance l'emporte sur toute fonction puissance)

Toute fonction de répartition possède cette propriété ($\mathbb{P}(\Omega) = 1$)

(c) Par dérivation, on obtient la densité f :

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x < 0 \\ x \mapsto e^{-x/2}\frac{x}{4} & , \text{ si } x \geq 0 \end{cases}$$

X suit une loi gamma d'ordre 1 et de paramètre ½.

(d) Pour $x > 0$ $f'(x) = \frac{1}{8}e^{-x/2}(2 - x)$.

$f'(x) > 0$ pour $x < 2$; $f'(x) = 0$ pour $x = 2$; $f'(x) < 0$ pour $x > 2$.

f admet donc un maximum pour $x = 2$. La variable aléatoire X a pour mode 2.

(e)

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x = \int_0^{+\infty} e^{-x/2}\frac{x^2}{4}\mathrm{d}x = \left[-e^{-x/2}\frac{x^2}{2}\right]_0^{+\infty} + 4\int_0^{+\infty} e^{-x/2}\frac{x}{4}\mathrm{d}x = 4$$

$$\mathbb{E}(X^2) = \int_0^{+\infty} e^{-x/2}\frac{x^3}{4}\mathrm{d}x = \left[-e^{-x/2}\frac{x^3}{2}\right]_0^{+\infty} + 6\int_0^{+\infty} e^{-x/2}\frac{x^2}{4}\mathrm{d}x = 6\times 4 = 24$$

On en déduit $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = 24 - 16 = 8$ et donc $\sigma(X) = 2\sqrt{2}$

(f) tableau récapitulatif ($f(2) \approx 0.184$, $F(2) \approx 0.264$

(g) $\mathbb{P}(1 \leq X < 2) = F(2) - F(1) = 1 - 2e^{-1} - 1 + \frac{3}{2}e^{-1/2} = 2e^{-1} + \frac{3}{2}e^{-1/2} = \approx 17.40\%$

On a aussi $\mathbb{P}(1 \leq X < 2) = \int_1^2 \frac{x}{4}e^{-x/2}\mathrm{d}x = \left[-\frac{x}{2}e^{-x/2} - e^{-x/2}\right]_1^2 = 2e^{-1} + \frac{3}{2}e^{-1/2} = \approx 17.40\%$

2. (a) Pour que f soit une densité il faut que λ soit un réel positif ($f(t) \geq 0$ et $\int_{-\infty}^{+\infty} f(t)\mathrm{d}t = 1$.

D'où $\int_{-1}^{1} \lambda(1 - t^2)\mathrm{d}t = \lambda\left[t - \frac{t^3}{3}\right]_{-1}^{1} = \lambda\frac{4}{3} = 1$ et donc

$$\lambda = \frac{3}{4}$$

Il est aisé d'étudier les variations et de construire le graphe de f.

(b) Soit F la fonction de répartition de T. On a $F(t) = \mathbb{P}(T < t) = \int_{-\infty}^{t} f(x)\mathrm{d}x$
On en déduit aisément :

$$F : \begin{cases} t \mapsto 0 & , \text{ si } t \leq -1 \\ t \mapsto \frac{1}{2} + \frac{3}{4}\left(t - \frac{t^3}{3}\right) & , \text{ si } -1 < t \leq 1 \\ t \mapsto 1 & , \text{ si } t > 1 \end{cases}$$

(c)

$$\begin{aligned} \mathbb{P}\left(|T| \geq \frac{1}{2}\right) &= \int_{-\infty}^{-1/2} f(t)\mathrm{d}t + \int_{1/2}^{+\infty} f(t)\mathrm{d}t \\ &= \int_{-1}^{-1/2} \frac{3}{4}(1-t^2)\mathrm{d}t + \int_{1/2}^{1} \frac{3}{4}(1-t^2)\mathrm{d}t \\ &= 2\frac{3}{4}\left[t - \frac{t^3}{3}\right]_{1/2}^{1} \text{ à cause de la symétrie de la distribution} \\ &= \frac{3}{2}\left[\frac{2}{3} - \frac{1}{2} + \frac{1}{24}\right] \\ &= \frac{5}{16} \end{aligned}$$

On pourrait également écrire :

$$\begin{aligned} \mathbb{P}\left(|T| \geq \frac{1}{2}\right) &= \mathbb{P}\left(T \geq \frac{1}{2}\right) + \mathbb{P}\left(T \leq frac12\right) \\ &= 1 - F\left(\frac{1}{2}\right) + F\left(-\frac{1}{2}\right) \\ &= 2F\left(-\frac{1}{2}\right) \\ &= 2\left(1 - F\left(\frac{1}{2}\right)\right) \text{ toujours à cause de la symétrie} \end{aligned}$$

(d) Il est clair que T a pour mode 0 et pour médiane 0
$\mathbb{E}(T) = \int_{-\infty}^{+\infty} tf(t)\mathrm{d}t = \int_{-1}^{1} \frac{3}{4}(t - t^3)\mathrm{d}t = 0$ (fonction impaire)
On en déduit

$$\begin{aligned} \mathbb{V}(T) &= \mathbb{E}(T^2) - \mathbb{E}^2(T) = \mathbb{E}(T^2) \\ &= \int_{-\infty}^{+\infty} t^2 f(t)\mathrm{d}t \\ &= \int_{-1}^{1} \frac{3}{4}(t^2 - t^4)\mathrm{d}t \\ &= 2\int_{0}^{1} \frac{3}{4}(t^2 - t^4)\mathrm{d}t \text{ (fonction paire)} \end{aligned}$$

$$\begin{aligned} &= \frac{3}{2}\left[\frac{t^3}{3}-\frac{t^5}{5}\right]_0^1 \\ &= \frac{1}{5} \end{aligned}$$

3. Réponses : $a = n$; $\mathbb{E}(X) = \frac{n}{n+1}$; $\mathbb{V}(X) = \frac{n}{(n+2)(n+1)^2}$

$$F : \begin{cases} x \mapsto 0 & , \text{ si } x \leq 0 \\ x \mapsto x^n & , \text{ si } 0 < x \leq 1 \\ x \mapsto 1 & , \text{ si } x > 1 \end{cases}$$

4. Réponses : $\mathbb{E}(Y) = 1$ mode de Y= médiane de Y=0 $\sigma(Y) = \frac{\pi}{2\sqrt{3}}$

$$F : \begin{cases} y \mapsto 0 & , \text{ si } y \leq -\frac{\pi}{2} \\ y \mapsto \frac{1}{2} + \frac{1}{\pi}\left(x + \frac{\sin 2x}{2}\right) & , \text{ si } -\frac{\pi}{2} < y \leq \frac{\pi}{2} \\ y \mapsto 1 & , \text{ si } y > \frac{\pi}{2} \end{cases}$$

$$\mathbb{P}\left(Y > \frac{4\pi}{10}\right) = 1 - \left(0.9 + \frac{1}{2\pi}\sin\frac{4\pi}{5}\right) \approx 0.645\%$$

Indications : pour le calcul des intégrales, on pose $\cos^2 x = \frac{1+\cos 2x}{2}$; la fonction $x^2 \cos 2x$ s'intègre par parties. La distribution est en cloche, symétrique par rapport à la valeur zéro.

5. Réponses : $a = \frac{1}{\pi}$; $\mathbb{E}(X) = \frac{8}{3\pi} \approx 0.8488$; $\mathbb{V}(X) = 1 - \left(\frac{8}{3\pi}\right)^2 \approx 0.2795$; mode de X=0

$$F : \begin{cases} x \mapsto 0 & , \text{ si } x \leq 0 \\ x \mapsto \frac{2}{\pi}\arcsin\frac{x}{2} + \frac{1}{\pi}x\sqrt{1-\frac{x^2}{4}} & , \text{ si } 0 < x \leq 12 \\ x \mapsto 1 & , \text{ si } x > 2 \end{cases}$$

Indications : on note que le graphe de f est un quart de cercle centré à l'origine et de rayon 2. On remarque que f est discontinue en zéro ; ceci entraîne, pour F, l'existence, en zéro, d'une dérivée à gauche et d'une dérivée à droite respectivement égales à la limite à gauche et à la limite à droite de f.
La seule difficulté de l'exercice réside dans le calcul des intégrale. On écrit $\sqrt{4-x^2}\mathrm{d}x = 2\sqrt{1-\frac{x^2}{4}}\mathrm{d}x$ et on effectue le changement de variable :

$$\begin{aligned} u = \arcsin\frac{x}{2} \quad &\Leftrightarrow \quad \frac{x}{2} = \sin u \text{ et } -\frac{\pi}{2} \leq u \leq \frac{\pi}{2} \\ &\Rightarrow \quad \mathrm{d}x = 2\cos u \mathrm{d}u \\ &\qquad \sqrt{1-\frac{x^2}{4}} = |\cos u| = \cos u \end{aligned}$$

On a ainsi $I_0 = \int_0^2 \sqrt{4-x^2}\mathrm{d}x = 4\int_0^{\frac{\pi}{2}} \cos^2 u\mathrm{d}u = 4\int_0^{\frac{\pi}{2}} \frac{1+\cos 2u}{2}\mathrm{d}u$
$I_1 = \int_0^2 x\sqrt{4-x^2}\mathrm{d}x = 8\int_0^{\frac{\pi}{2}} \cos^2 u \sin u\mathrm{d}u = -\frac{8}{3}\left[\cos^3 u\right]_0^{\frac{\pi}{2}}$
$I_2 = \int_0^2 x^2\sqrt{4-x^2}\mathrm{d}x = 16\int_0^{\frac{\pi}{2}} \cos^2 u \sin^2 u\mathrm{d}u = 4\int_0^{\frac{\pi}{2}} \sin^2 2u\mathrm{d}u = 4\int_0^{\frac{\pi}{2}} \frac{1-\cos 4u}{2}\mathrm{d}u$

6. (a) On doit avoir $f(x) \geq 0$ et $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$. Or $\frac{2}{\pi}\frac{1}{(1+x^2)^2} > 0$ sur $\mathbb{R}$; il reste à calculer $\int_{-\infty}^{+\infty} \frac{2}{\pi}\frac{1}{(1+x^2)^2}\mathrm{d}x$.
On fait le changement de variable :

$$\begin{aligned} u = \arctan x \quad &\Leftrightarrow \quad x = \tan u \text{ et } -\frac{\pi}{2} < u < \frac{\pi}{2} \\ &\Rightarrow \quad \mathrm{d}x = (1 + tan^2 u)\mathrm{d}u \end{aligned}$$

On obtient $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} \frac{1}{1+\tan^2 u}\mathrm{d}u = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} \cos^2 u\mathrm{d}u = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} \frac{1+\cos 2u}{2}\mathrm{d}u = \frac{1}{\pi}\left[u + \frac{\sin 2u}{2}\right]_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} = 1$

(b) Sous réserve de la convergence des intégrales, on a : $\mathbb{E}(X) = \mu = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$, $\mathbb{V}(X) = \sigma^2(X) = \mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{+\infty}(X-\mu)^2 f(x)\mathrm{d}x$
$\mathbb{E}(X) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{1}{(1+x^2)^2}2x\mathrm{d}x = \frac{1}{\pi}\left[-\frac{1}{1+x^2}\right]_{-\infty}^{+\infty} = 0.$
Comme $\mathbb{E}(X) = \mu = 0$, on a, dans cas :
$\mathbb{V}(X) = \mathbb{E}(X^2) = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} \frac{\tan^2 u}{1+\tan^2 u}\mathrm{d}u = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} \sin^2 u\mathrm{d}u = \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} \frac{1+\cos 2u}{2}\mathrm{d}u = \frac{1}{\pi}\left[u + \frac{\sin 2u}{2}\right]_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} = 1$
Finalement :

$$\mathbb{E}(X) = 0, \sigma(X) = \sqrt{1} = 1$$

(c) Par définition, la fonction de répartition est une fonction F, définie sur $\mathbb{R}$, et telle que : $F(x) = \mathbb{P}(X < x)$. On a donc :

$$F(x) = \mathbb{P}(X < x) = \int_{-\infty}^{x} f(t)\mathrm{d}t = \frac{2}{\pi}\int_{-\infty}^{x} \frac{\mathrm{d}t}{(1+t^2)^2}, \forall t \in \mathbb{R}$$

En posant $u = \arctan t$, on obtient : $F(x) = \frac{1}{\pi}\left[u + \frac{\sin 2u}{2}\right]_{-\frac{\pi}{2}}^{\arctan x}$. En remarquant que $\sin 2u = \frac{2\tan u}{1+\tan^2 u}$:
$F(x) = \frac{1}{\pi}\left[u + \frac{\tan u}{1+\tan^2 u}\right]_{-\frac{\pi}{2}}^{\arctan x} = \frac{1}{\pi}\left[\arctan x + \frac{x}{1+x^2} + \frac{\pi}{2}\right]$. D'où :

$$F : x \mapsto \frac{1}{2} + \frac{1}{\pi}\left[\arctan x + \frac{x}{1+x^2}\right], \forall x \in \mathbb{R}$$

On pet vérifier que F est strictement croissante (dérivée positive), que $\lim_{x\to-\infty} F(x) = \frac{1}{2} + \frac{1}{\pi}\left(-\frac{\pi}{2}\right) = 0$ et que $\lim_{x\to 1\infty} F(x) = \frac{1}{2} + \frac{1}{\pi}\left(\frac{\pi}{2}\right) = 1$

(d) $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X < 1) = 1 - F(1) = 1 - \frac{1}{2} - \frac{1}{\pi}\left[\frac{\pi}{4} + \frac{1}{2}\right] = \frac{1}{4} - \frac{1}{2\pi} \approx 9.08\%$

7. L'équation différentielle 3.11 est du premier ordre, linéaire et homogène. Elle se résout par séparation des variables et admet sur $]-1, 1[$, la solution générale ϕ : $x \mapsto \lambda(1-x^2)^2, \forall\lambda \in \mathbb{R}$. Sur $]-1, 1[$, f est solution particulière de 3.11, déterminée par la condition $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$.

(a) On trouve facilement :

$$F : \begin{cases} x \mapsto 0 & , \text{ si } x \notin]-1,1[\\ x \mapsto \frac{15}{16}(x^4 - 2x^2 + 1) & , \text{ si } x \in]-1,1[\end{cases}$$

ainsi que :

$$F : \begin{cases} x \mapsto 0 & , \text{ si } x \leq -1 \\ x \mapsto \frac{1}{2} + \frac{15}{16}\left(\frac{x^5}{5} - 2\frac{x^3}{3} + x\right) & , \text{ si } -1 < x \leq 1 \\ x \mapsto 1 & , \text{ si } x > 1 \end{cases}$$

(b) $\mathbb{E}(X)$= mode de X= médiane de X=0 - $\mathbb{V}(X) = 1/7$

(c) $\mathbb{P}(X \geq 1/2) = 1 - F(1/2) = \frac{1}{2} - \frac{15}{16}\left(\frac{1}{160} - \frac{1}{12} + \frac{1}{2}\right) \approx 10.35\%$ et à cause de la symétrie de la distribution
$\mathbb{P}(-1/2 \leq X \leq 1/2) = 1 - 2\mathbb{P}(X \geq 1/2) \approx 79.30\%$

8. L'équation différentielle 3.12 est du premier ordre, linéaire et homogène. De l'équation caractéristique $r(r+2) = 0$, on déduit la solution générale $\phi : x \mapsto A + Be^{-2x}$, sur $\mathbb{R}$, $\forall A$, $\forall B$. On aura donc $F(x) = 0$ pour $x \leq 0$ et $F(x) = A + Be^{-2x}$ pour $x > 0$.

(a) Le problème de la continuité de F se pose en zéro. Or $F(0) = 0 = \lim_{x \to 0^-} F(x)$. Il faut donc $F(0) = 0 = \lim_{x \to 0^+} F(x) = A + B$. Donc :

$$B = -A$$

(b) F, fonction de répartition doit être non décroissante. Or pour $x > 0$, $F'(x) = 2Ae^{-2x}$. On doit avoir $A > 0$.

(c) De plus $\lim_{x \to +\infty} F(x) = 1 - 0 \Rightarrow \lim_{x \to +\infty} A - Ae^{-2x} = A = 1$

(d) On obtient finalement :

$$F : \begin{cases} x \mapsto 0 & , \text{ si } x \leq 0 \\ x \mapsto 1 - e^{-2x} & , \text{ si } x > 0 \end{cases}$$

puis la densité

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x \leq 0 \\ x \mapsto 2e^{-2x} & , \text{ si } x > 0 \end{cases}$$

F admet en zéro une dérivée à gauche et une dérivée à droite. On peut choisir arbitrairement.

(e) Mode de $X = 0$; $\mathbb{E}(X) = \frac{1}{2}$; $\mathbb{V}(X) = \frac{1}{4}$

(f) $F(x) = \frac{1}{2}$ pour $e^{-2x} = \frac{1}{2} \Rightarrow 2x = \log 2$. Médiane de $X = \frac{1}{2}\log 2 \approx 0.3465$

9. On a $F(t) = \mathbb{P}(Z < t), \forall t \in \mathbb{R}$
Or l'événement $(Z < t)$ peut s'écrire $(Z < t) = (Z = X \cap Z < t) \cup (Z = Y \cap Z < t)$.

Les deux éventualités étant incompatibles, on a :

$$\begin{aligned}\mathbb{P}(Z<t) &= \mathbb{P}(Z=X\cap Z<t)+\mathbb{P}(Z=Y\cap Z<t)\\ &= \mathbb{P}(Z=X)\mathbb{P}(Z<t|Z=X)+\mathbb{P}(Z=Y)\mathbb{P}(Z<t|Z=Y)\\ &= \mathbb{P}(Z=X)\mathbb{P}(Z=X)+\mathbb{P}(Z=Y)\mathbb{P}(Z=Y)\end{aligned}$$

Finalement $F(t)=pF_1(t)+qF_2(t)$ et par dérivation $f(t)=pf_1(t)+qf_2(t)$. On en déduit :

$$\begin{aligned}\mathbb{E}(Z^\alpha) &= \int_{-\infty}^{+\infty} t^\alpha f(t)\mathrm{d}t\\ &= \int_{-\infty}^{+\infty} t^\alpha (pf_1(t)+qf_2(t))\mathrm{d}t\\ &= p\int_{-\infty}^{+\infty} t^\alpha f_1(t)\mathrm{d}t+q\int_{-\infty}^{+\infty} t^\alpha f_2(t)\mathrm{d}t\end{aligned}$$

$$p\mathbb{E}(X^\alpha)+q\mathbb{E}(Y^\alpha)$$

En particulier

$$\begin{aligned}\mathbb{E}(Z)=p\mathbb{E}(X)+q\mathbb{E}(Y) &\Leftrightarrow \mu=p\mu_1+q\mu_2\\ \mathbb{E}(Z^2)=p\mathbb{E}(X^2)+q\mathbb{E}(Y^2) &\end{aligned}$$

On sait, d'autre part, que pour toute variable T

$$\mathbb{V}(T) = \mathbb{E}(T^2)-\mathbb{E}^2(T) \qquad (3.13)$$

$$\mathbb{E}(T^2) = \mathbb{V}(T)+\mathbb{E}^2(T) \qquad (3.14)$$

Donc, d'après 3.14 : $\mathbb{E}(Z^2)=p(\mu_1^2+\sigma_1^2)+q(\mu_2^2+\sigma_2^2)$
et, d'après 3.13 : $\mathbb{V}(Z)=p(\mu_1^2+\sigma_1^2)+q(\mu_2^2+\sigma_2^2)-(p\mu_1+q\mu_2)^2$
D'où :

$$\begin{aligned}\mathbb{V}(Z) &= p\mu_1^2+q\mu_2^2-p^2\mu_1^2-q^2\mu_2^2-2pq\mu_1\mu_2+p\sigma_1^2+q\sigma_2^2\\ &= p\mu_1^2(1-p)+q\mu_2^2(1-q)-2pq\mu_1\mu_2+p\sigma_1^2+q\sigma_2^2\\ &= pq(\mu_1^2+\mu_2^2-2\mu_1\mu_2)+p\sigma_1^2+q\sigma_2^2\\ &= pq(\mu_1-\mu_2)^2+p\sigma_1^2+q\sigma_2^2\end{aligned}$$

10. La densité de X est :

$$f:\left\{\begin{array}{lll} x\mapsto 0 & , & \text{si } x\notin[a,b]\\ x\mapsto \frac{1}{b-a} & , & \text{si } x\in[a,b]\end{array}\right.$$

$$\mathbb{E}(X)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x=\int_a^b \frac{x}{b-a}\mathrm{d}x=\frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b=\frac{1}{2}\frac{b^2-a^2}{b-a}=\frac{b+a}{2}$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \int_a^b \frac{x^2}{b-a}\mathrm{d}x = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{3}\frac{b^3-a^3}{b-a} = \frac{1}{3}(b^2+ab+a^2)$$

D'où, d'après 3.13 :

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{1}{3}(b^2+ab+a^2) - \frac{(b+a)^2}{4} = \frac{1}{12}(b-a)^2$$

On a $F(t) = \mathbb{P}(T < t) = \int_{-\infty}^{t} f(x)\mathrm{d}x\ \forall t \in \mathbb{R}$. On en déduit : $F(t) = 0$, si $t \le a$
$F(t) = \int_{-\infty}^{a} f(x)\mathrm{d}x + \int_a^t f(x)\mathrm{d}x = 0 + \int_a^t \frac{1}{b-a}\mathrm{d}x = \frac{t-a}{b-a}$, si $a < t \le b$
$F(t) = \int_{-\infty}^{a} f(x)\mathrm{d}x + \int_a^b f(x)\mathrm{d}x + + \int_b^t f(x)\mathrm{d}x = 0 + \frac{b-a}{b-a} + 0 = 1$, si $t > b$

11. La variable X suit une loi exponentielle, de paramètre λ
   (a) On remarque que f possède bien les propriétés d'une densité : $f(x) \ge 0$ sur $\mathbb{R}$
   $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \int_0^{+\infty} \lambda e^{-\lambda x}\mathrm{d}x = \left[-e^{-\lambda x}\right]_0^{+\infty} = 1$
   (b)

$$\mathbb{E}(X) = int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = \int_0^{+\infty} \lambda x e^{-\lambda x}\mathrm{d}x = \left[-xe^{-\lambda x}\right]_0^{+\infty} + \frac{1}{\lambda}\int_0^{+\infty} \lambda e^{-\lambda x}\mathrm{d}x = \frac{1}{\lambda}$$

$$\mathbb{E}(X^2) = int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \int_0^{+\infty} \lambda x^2 e^{-\lambda x}\mathrm{d}x = \left[-x^2 e^{-\lambda x}\right]_0^{+\infty} + \frac{2}{\lambda}\int_0^{+\infty} \lambda x e^{-\lambda x}\mathrm{d}x = \frac{2}{\lambda}$$

$$V(X) = \frac{2}{\lambda} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

   (c) Soit F la fonction de répartition. On a

$$\begin{aligned} F(t) = \mathbb{P}(X < t) &= \int_{-\infty}^{t} f(x)\mathrm{d}x = 0 \text{ si } t \le 0 \\ &= \int_{-\infty}^{0} f(x)\mathrm{d}x + \int_0^t f(x)\mathrm{d}x = \left[-e^{-\lambda x}\right]_0^t = 1 - e^{-\lambda t} \text{ si } t > 0 \end{aligned}$$

   (d) La densité est nulle à gauche de zéro et décroissante à droite de zéro. ($f'(x) < 0$). Le mode de X est donc 0.
   $F(t) = 0.5$ si $e^{-\lambda t} = 0.5$. X a donc pour médiane $t = \frac{1}{\lambda}\log 2$ (voir exercice 7 pour $\lambda = 2$)
   (e) Si $\lambda = \frac{1}{2}$:

$$\mathbb{P}(2 \le X \le 3) = F(3) - F(2) = e^{-1} - e^{-3/2} \approx 14.47\%$$

$$\mathbb{P}(X \ge 5) = 1 - F(5) = e^{-5/2} \approx 8.21\%$$

$$\mathbb{P}(X < 7) = F(7) = 1 - e^{-7/2} \approx 96.98\%$$

12. X a pour densité $f : x \mapsto \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $\forall x \in \mathbb{R}$
Le graphe de f est symétrique par rapport à $x = \mu$.
Point d'inflexion pour $x = \mu \pm \sigma$.
Donner les valeurs de $\mu + i\sigma$ pour $i = 1, 2, 3$ dans le tableau de variation.
On a $f(x) = \frac{1}{2}\frac{1}{\sigma\sqrt{2\pi}}$ pour $e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{2} \Rightarrow \left(\frac{x-\mu}{\sigma}\right) = 2\log 2 \Rightarrow x = \mu \pm \sigma\sqrt{2\log 2}$
D'où la largeur à mi-hauteur :
$$2\sigma\sqrt{2\log 2} = 2.3548\sigma$$
En effectuant le changement de variable $z = \frac{x-\mu}{\sigma}$ et en admettant $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \mathrm{d}x = 1$ on trouve facilement $\mathbb{E}(X) = \mu$ et $\mathbb{V}(X) = \sigma^2$.
En tenant compte de la symétrie de la distribution, en désignant par Z une variable normale centrée réduite, on aura (voir table) :
$$\mathbb{P}(\mu - a < X < \mu + a) = 0.95 \Leftrightarrow \mathbb{P}(\mu - a < X) = 0.975 \Leftrightarrow \mathbb{P}(Z < a/\sigma) = 0.975 \Leftrightarrow a \approx 1.96\sigma$$
$$\mathbb{P}(\mu - b < X < \mu + b) = 0.99 \Leftrightarrow b \approx 2.576\sigma$$
$$\mathbb{P}(X \leq c) = 0.05 \Leftrightarrow c \approx \mu - 1.645\sigma$$
$$\mathbb{P}(X \geq d) = 0.05 \Leftrightarrow d \approx \mu + 1.645\sigma$$

13. Soit Z une variable normale centrée réduite, de fonction de répartition F. Si $X \sim \mathcal{N}(4, 2^2)$, on a :
   (a) $\mathbb{P}((X > 6) = \mathbb{P}(Z > \frac{6-4}{2}) = 1 - \mathbb{P}(Z \leq 1) = 1 - F(1)$.
      - Dans la table 11.6, on lit $F(1) \approx 0.8413$
      - Dans la table 11.7, on lit $F(1) = \frac{1}{2} + \theta(1) \approx 0.5 + .3413$
      - Dans la table 11.8, on estime $F(1) = 1 - \frac{1}{2}\mathbb{P}(|Z| \geq 1) \approx 1 - \frac{1}{2}(31 + \frac{15}{21})\frac{1}{100} \approx 0.8414$

      Finalement $\mathbb{P}(X > 6) = 1 - 0.8413 \approx 15.87\%$
   (b) $\mathbb{P}((|X| > 6) = \mathbb{P}(-6 < X < 6) = 1 - \mathbb{P}(-5 < Z < 1) = 1 - (F(1) - F(-5)$
      Dans la table 11.6, on lit $F(1) \approx 0.8413$
      Dans la table 11.8, on tire, $F(-5) = \frac{1}{2}\mathbb{P}(|X| > 5) < \frac{1}{2}10^{-6}$ donc négligeable par rapport à $F(1)$.
      D'où $\mathbb{P}(X > 6) = 1 - 0.8413 \approx 15.87\%$
   (c) $\mathbb{P}((2 \leq X \leq 6) = \mathbb{P}(-1 < Z < 1) = 68.26\%$
   (d) $\mathbb{P}((1.7 \leq X \leq 2.82) = \mathbb{P}(-1.15 < Z < -0.59) = \mathbb{P}(0.59 < Z < 1.15)$ (par symétrie) $= F(1.15) - F(0.59) = 15.25\%$

14. Soit l'expérience : un voyageur se présente ; deux résultats sont possibles. Il prend la rame A ou il prend la rame B, également probables. On pose $p = \mathbb{P}(\text{Il prend la rame } A) = \frac{1}{2}$.
Soit l'expérience $\mathcal{E}$: 1600 voyageurs se présentent. On lui associe la variable X : nombre de voyageurs susceptibles de prendre le train A. En supposant l'indépendance des choix, on a : $X \sim \mathcal{B}\left(1600, \frac{1}{2}\right)$. Les conditions sont idéales pour utiliser l'approximation $X \sim \mathcal{N}(800, 20^2)$.
Soit n le nombre de places assises dans une rame. L'événement à éviter est "$X > n$" mais aussi "$X < 1600 - n$" ; ce dernier résultat entraînerait "$1600 - X > n$", donc une saturation de la rame B.
On veut : $\mathbb{P}(X > n) + \mathbb{P}(X < 1600 - n) < 3\,10^{-3}$ (probabilités totales).
Si $Z \sim \mathcal{N}(0,1)$, cette inégalité équivaut à : $\mathbb{P}\left(Z > \frac{n-800}{20}\right) + \mathbb{P}\left(Z < \frac{1600-n-800}{20}\right) < 3\,10^{-3}$. Ce que l'on peut écrire, puisque n est sûrement supérieur à 800 : $\mathbb{P}\left(|Z| > \frac{n-800}{20}\right) < 3\,10^{-3}$. Cette inégalité est vérifiée pour $\frac{n-800}{20} > 2.97$ (voir table 11.7). On en déduit $n \geq 860$
15. Soit $X \sim \mathcal{P}(\lambda)$. Comme $\lambda = 10^{-4}$, on peut écrire : $X \sim \mathcal{N}(1000, 100^2)$. On a donc, si Z est une variable normale centrée réduite : $\mathbb{P}(9800 \leq X \leq 10200) \approx \mathbb{P}(-2 \leq Z \leq 2) = 2\frac{1}{2}\theta(2)$ (voir table 11.6 ou 11.7) . Il est inutile d'effectuer la correction de continuité $= 2F(2) - 1 \approx 95.44\%$
16. Si on choisit au hasard un individu mâle, 2 résultats sont possibles : sa taille est $\geq 186cm$ ou sa taille est $< 186cm$, éventualités disjointes. Si $Z \sim \mathcal{N}(0,1)$, on a :
$$P = \mathbb{P}(X \geq 186\text{cm}) = 1 - \mathbb{P}(X < 186) = 1 - \mathbb{P}\left(Z < \frac{186-170}{6.2}\right) = 1 - \mathbb{P}(Z < 2.58) \approx 0.5\%$$
(voir table 11.7
Soit l'expérience $\mathcal{E}$: on choisit 200 individus mâles ; on peut lui associer la variable Y : nombre d'individus de taille $\geq$ 186cm. On a $Y \sim \mathcal{B}(200, 5\,10^{-3}) \sim \mathcal{P}(1)$ (nombre d'individus élevé, et probabilité P très faible).
On déduit de la table 11.4 :
$$\mathbb{P}(Y \geq 5) = 1 - \mathbb{P}(Y < 5) \approx 1 - (0.3679 + 0.3679 + 0.1839 + 0.0613 + 0.0153) \approx 37\,10^{-4}$$
17. Soit T la durée de fonctionnement correct à partir d'un révision. Soit a le temps séparant deux révisions. On veut déterminer a pour que : $\mathbb{P}(T < a) \leq 0.05$. En désignant par Z une variable normale centrée réduite, on a : $\mathbb{P}(T < a) = \mathbb{P}\left(Z < \frac{a-1500}{300}\right)$. La condition $\mathbb{P}\left(Z < \frac{a-1500}{300}\right) < 0.005$ est réalisée pour $\frac{a-1500}{300} < -2.576 \Rightarrow a \leq 727$ heures
18. Soit X la variable qui associe à un jour le nombre de clients qui se présentent sur les n abonnés. On admet que chaque abonné a une probabilité de $9/10$ de se présenter et que les décisions sont indépendantes. Dans ces conditions $X \sim \mathcal{B}(n, 9/10)$. Donc $\mathbb{E}(X) = 0.9n$ et $\mathbb{V}(X) = 0.09n$.
L'objectif étant d'avoir plus de 900 abonnés, on a certainement : $n > 900$ et $n\,9/10 > n\,1/10 > 90$. On peut donc utiliser l'approximation normale $X \sim \mathcal{N}(0.9n, 0.09n)$.

Il s'agit de choisir n pour que $\mathbb{P}(X > 900) \approx 0$. On a évidemment zéro pour $n \leq 900$. Pour envisager un bénéfice supplémentaire, il est clair qu'il faut courir un risque. En choisissant $\mathbb{P}(X > 900) \leq 10^{-3}$, et avec $Z \sim \mathcal{N}(0, 1)$, on obtient :

$$\mathbb{P}(X > 900) = \mathbb{P}\left(Z > \frac{900 - 0.9n}{0.3\sqrt{n}}\right) \leq 10^{-3}$$

En utilisant la table 11.8, on : $\frac{900-0.9n}{0.3\sqrt{n}} \geq 3.090 \Rightarrow 0.9n + 0.927\sqrt{n} - 900 \leq 0 \Rightarrow (\sqrt{n})^2 + 1.03\sqrt{n} - 1000 \leq 0$.

Ce trinôme a deux racines : -32.142 et 31.112 ; en tenant compte du fait que $\sqrt{n}$ est un nombre positif, le trinôme est négatif ou nul pour $\leq \sqrt{n} \leq 31.112$, c'est-à-dire $n \leq 967$ pour un risque de 10^{-3}.

On note qu'effectuer la correction de continuité consiste à écrire ici : $\mathbb{P}(X > 900.5)$. On obtient alors $\sqrt{n} \leq 31.120$ et $n \leq 968$.

19. On note
- V_1 la vitesse d'un homme $V_1 \sim \mathcal{N}(1.6, 0.4^2)$;
- V_2 la vitesse d'un homme $V_1 \sim \mathcal{N}(1.4, 0.4^2)$;
- H l'événement : "l'individu observé est un homme ;
- F l'événement : "l'individu observé est une femme
- Z une variable normale centrée réduite.

(a)

$$\begin{aligned}
\mathbb{P}(V < 1.2) &= \mathbb{P}(H \cap V < 1.2) + \mathbb{P}(F \cap V < 1.2) \text{ événements incompatibles} \\
&= \mathbb{P}(H)\mathbb{P}(V < 1.2|H) + \mathbb{P}(F)\mathbb{P}(V < 1.2|F) \\
&= \frac{1}{3}\mathbb{P}(V_1 < 1.2) + \frac{2}{3}\mathbb{P}(V_2 < 1.2) \\
&= \frac{1}{3}\mathbb{P}\left(Z < \frac{1.2 - 1.6}{0.4}\right) + \frac{2}{3}\mathbb{P}\left(Z < \frac{1.2 - 1.6}{0.4}\right) \\
&= \frac{1}{3}\mathbb{P}(Z < -1) + \frac{2}{3}\mathbb{P}(Z < -0.5) \\
&= \frac{1}{3}0.1587 + \frac{2}{3}0.3085 \\
&\approx 25.85\%
\end{aligned}$$

(b) $\mathbb{P}(F|V < 1.2) = \frac{\mathbb{P}(F \cap V < 1.2)}{\mathbb{P}(V < 1.2)} \approx \frac{0.2056}{0.2585} \approx 79.53\%$

(c) On note F et f la fonction de répartition et la densité de V ; on utilisera les fonctions F_1 et f_1 pour V_1, F_2 et f_2 pour V_2.
$F(x) = \mathbb{P}(V < x) = \mathbb{P}(H \cap V < x) + \mathbb{P}(F \cap V < x) = \frac{1}{3}F_1(x) + \frac{2}{3}F_2(x)$
Par dérivation $f(x) = \frac{1}{3}f_1(x) + \frac{2}{3}f_2(x)$
D'où $f(x) = \frac{1}{3}\frac{1}{0.4\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-1.6}{0.4}\right)^2} + \frac{2}{3}\frac{1}{0.4\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-1.4}{0.4}\right)^2} \forall x \in \mathbb{R}$

(d) $\mathbb{E}(V) = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x = \int_{-\infty}^{+\infty} x(\frac{1}{3}f_1(x) + \frac{2}{3}f_2(x))\mathrm{d}x = \frac{1}{3}\mathbb{E}(V_1) + \frac{2}{3}\mathbb{E}(V_2) \approx 1.4667$ m/s

$\mathbb{E}(V^2) = \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \frac{1}{3}\mathbb{E}(V_1^2) + \frac{2}{3}\mathbb{E}(V_2^2) = \frac{1}{3}(\mathbb{V}(V_1) + \mathbb{E}^2(V_1)) + \frac{2}{3}(\mathbb{V}(V_2) + \mathbb{E}^2(V_2))$
$\mathbb{E}(V^2) \approx \frac{1}{3}(0.16 + 2.56) + \frac{2}{3}(0.16 + 1.96) = 2.32\,{}^{m^2}\!/_{s^2}$
D'où la variance $\mathbb{V}(V) = \mathbb{E}(V^2) - \mathbb{E}^2(V) \approx 0.1688$ et l'écart-type $\sigma(V) \approx 0.41\,{}^{\mathrm{m}}\!/_{\mathrm{s}}$

20. Soit X la variable qui associe à un jour le nombre de clients du premier service. $X \sim \mathcal{B}\left(900, \frac{1}{2}\right)$. L'événement "on refuse du monde à l'un des deux services" s'écrit : $E = (X > 500 \cup X < 400)$. En utilisant l'approximation normale on obtient $\mathbb{P}(E) \approx 8$ ou $9\,10^{-4}$. La correction de continuité est ici négligeable.
21. Réponses :
   - si $P = \frac{3}{4}$ on obtient une probabilité de l'ordre de $3\,10^{-6}$;
   - si $P = \frac{9}{16}$ on obtient une probabilité de l'ordre de 99.98%
22. Réponses : Mode=1 ; $\mathbb{E}(X) = \sqrt{{}^{\pi}\!/_{2}}$; $\mathbb{V}(X) = 2 - {}^{\pi}\!/_{2}$.

$$F : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto 1 - e^{-t^2/2} & , \text{ si } t > 0 \end{cases}$$

   $\mathbb{P}(0 \leq X < 2) \approx 86.466\%$
23. (a)

$$I_1 = \int_0^{+\infty} x e^{-x^2}\mathrm{d}x = \left[-\frac{1}{2}e^{-x^2}\right]_0^{+\infty} = \frac{1}{2}$$

$$\begin{aligned} I_{n+2} &= \int_0^{+\infty} x_{n+2} e^{-x^2}\mathrm{d}x = \int_0^{+\infty} x_{n+1} \times x e^{-x^2}\mathrm{d}x \\ &= \left[-\frac{1}{2}x_{n+1}e^{-x^2}\right]_0^{+\infty} + \frac{n+1}{2}\int_0^{+\infty} x_n e^{-x^2}\mathrm{d}x \\ &= \frac{n+1}{2} I_n \end{aligned}$$

   On en déduit $I_0 = \frac{\sqrt{\pi}}{2}$, $I_2 = \frac{\sqrt{\pi}}{4}$, $I_4 = \frac{3\sqrt{\pi}}{8}$, $I_6 = \frac{15\sqrt{\pi}}{6}$
   et $I_1 = \frac{1}{2} = I_3$, $I_5 = 1$.

   (b) La condition $\int_{-\infty}^{+\infty} f(v)\mathrm{d}v = 1$ permet de calculer A. On obtient : $A\int_{-\infty}^{+\infty} v^2 e^{-v^2/a^2}\mathrm{d}v = 1$. On intègre par changement de variable $x = \frac{v}{a}$, $\mathrm{d}v = a\mathrm{d}x$

$$Aa^3\int_0^{+\infty} x^2 e^{-x^2}\mathrm{d}x = Aa^3 I_2 = 1 \Rightarrow A = \frac{4}{a^3\sqrt{\pi}}$$

   $f'(v) = 0$ si $v \leq 0$; $f'(v) = 2Ave^{-v^2/a^2}\,(1 - {}^{v^2}\!/_{a^2})$ si $v > 0$.
   $f'(v) > 0$ si $0 < v < a$, $f'(v) = 0$ si $v = a$ et $f'(v) < 0$ si $v > a \Rightarrow$ Mode=a.

$\mathbb{E}(V^k) = \int_{-\infty}^{+\infty} v^k f(v)\mathrm{d}v = A\int_{-\infty}^{+\infty} v^{k+2} e^{-v^2/a^2}\mathrm{d}v = Aa^{k+3}\int_0^{+\infty} x^{k+2}e^{-x^2}\mathrm{d}x = Aa^{k+3}I_{k+2}$

$$\mathbb{E}(V^k) = \frac{4}{\sqrt{\pi}}a^k I_{k+2}$$

On en déduit

$$\begin{aligned}
\mathbb{E}(V) &= \frac{4}{\sqrt{\pi}}aI_3 = \frac{2a}{\sqrt{\pi}} \\
\mathbb{E}(V^2) &= \frac{4}{\sqrt{\pi}}a^2 I_4 = \frac{3}{2}a^2 \\
\mathbb{V}(V) &= \mathbb{E}(V^2) - \mathbb{E}^2(V) = a^2\left(\frac{3}{2} - \frac{4}{\pi}\right)
\end{aligned}$$

24. (a) f est continue sur $\mathbb{R}$. Pour $t > 0$, $f'(t) = \lambda^2 te^{-\lambda t}(1 - \lambda t)$. On obtient donc le tableau de variation et le mode vaut $\frac{1}{\lambda}$

$$\begin{aligned}
\mathbb{E}(T) &= \int_{-\infty}^{+\infty} tf(t)\mathrm{d}t = \int_0^{+\infty} \lambda^2 t^2 e^{-\lambda t}\mathrm{d}t = \left[-\lambda t^2 e^{-\lambda t}\right]_0^{+\infty} + \frac{2}{\lambda}\int_0^{+\infty} \lambda^2 te^{-\lambda t}\mathrm{d}t = \frac{2}{\lambda} \\
\mathbb{E}(T^2) &= \int_{-\infty}^{+\infty} t^2 f(t)\mathrm{d}t = \int_0^{+\infty} \lambda^2 t^3 e^{-\lambda t}\mathrm{d}t \\
&= \left[-\lambda t^3 e^{-\lambda t}\right]_0^{+\infty} + \frac{3}{\lambda}\int_0^{+\infty} \lambda^2 t^2 e^{-\lambda t}\mathrm{d}t = \frac{6}{\lambda^2} \\
\mathbb{V}(T) &= \frac{2}{\lambda^2}
\end{aligned}$$

$$\begin{aligned}
F(t) &= \mathbb{P}(T < t) = 0 \text{ si } t \leq 0 \\
&= \int_0^t \lambda^2 xe^{-\lambda x}\mathrm{d}x = \left[-\lambda xe^{-\lambda x}\right]_0^t + \int_0^t \lambda xe^{-\lambda x}\mathrm{d}x \\
&= \left[-\lambda xe^{-\lambda x} - e^{-\lambda x}\right]_0^t \\
&= 1 - (1 + \lambda t)e^{-\lambda t} \text{ si } t > 0
\end{aligned}$$

$P_1 = \mathbb{P}(T < 500) = F(500) = 1 - (1 + 0.1)e^{-0.1} \approx 4.68\,10^{-3}$
$P_2 = \mathbb{P}(T \geq 10^{-4}) = 1 - F(10^4) = (1 + 2)e^{-2} \approx 40.60\%$

(b) Si X_1 est la variable associant à un tel échantillon le nombre de composants de durée de vie inférieure à 500h, on a $X_1 \sim \mathcal{B}(600, 5\,10^{-3}) \sim \mathcal{P}(3)$

$$P(X_1 > 6) = 1 - \mathbb{P}(X_1 \leq 6) \approx 1 - 0.9664 \approx 3.36\% \text{ (voir table 11.4)}$$

Si X_2 est la variable associant à un tel échantillon le nombre de composants de durée de vie supérieure à 10000h, on a $X_1 \sim \mathcal{B}(600, 0.4) \sim \mathcal{N}(240, 12^2)$

Si $Z \sim \mathcal{N}(0,1)$, on a : $P(X_2 > 220) = \mathbb{P}(Z > -\frac{20}{12}) = \mathbb{P}(Z > -1.666)$.
Sans correction de continuité, on obtient :

$$\mathbb{P}(Z > -\frac{20}{12}) = \mathbb{P}(Z > -1.666) = \mathbb{P}(Z < 1.666) \approx 0.9522$$

Avec correction de continuité, on obtient :

$$\mathbb{P}(Z > -\frac{19.5}{12}) = \mathbb{P}(Z > -1.625) = \mathbb{P}(Z < 1.625) \approx 0.9479$$

(c) On détermine G, fonction de répartition de X. On obtient : $G(t) = \mathbb{P}(X < t) = \mathbb{P}(\lambda = a \cap X < t) + \mathbb{P}(\lambda = b \cap X < t) = \mathbb{P}(\lambda = a)\mathbb{P}(X < t|\lambda = a) + \mathbb{P}(\lambda = b)\mathbb{P}(X < t|\lambda = b) = \frac{2}{3}F_a(t) + \frac{1}{3}F_b(t)$
En dérivant on obtient la densité : $g(t) = \frac{2}{3}f_a(t) + \frac{1}{3}f_b(t)$. Donc :

$$g : \begin{cases} t \mapsto 0 & , \ \text{si } t < 0 \\ t \mapsto \frac{2}{3}a^2te^{-at} + \frac{1}{3}b^2te^{-bt} & , \ \text{si } t \geq 0 \end{cases}$$

On en déduit : $\mathbb{E}(X) = \frac{2}{3}\int_{-\infty}^{+\infty} tf_a(t)\mathrm{d}t + \frac{1}{3}\int_{-\infty}^{+\infty} tf_b(t)\mathrm{d}t = \frac{2}{3}\left(\frac{2}{a} + \frac{1}{b}\right)$. De même $\mathbb{E}(X^2) = \frac{4}{a^2} + \frac{2}{b^2}$.

Chapitre 4

Fonction d'une variable aléatoire

1 Généralités

Définition 4.1 *Soit X une variable aléatoire à valeurs réelles, associée à un espace de probabilité $\{\Omega, \mathcal{A}, \mathbb{P}\}$. Soit ϕ une fonction numérique d'une variable réelle définie sur $X(\Omega)$. L'application $Y = \phi \circ X$, définie sur Ω, est une variable aléatoire à valeurs réelles si elle possède la propriété :*

$$\text{Pour tout intervalle } I \subset \mathbb{R}, \quad \{\omega | Y(\omega) = \phi(X(\omega)) \in I\} \in \mathcal{A} \tag{4.1}$$

On dit alors que Y est une variable aléatoire fonction de la variable aléatoire X et on note $Y = \phi(X)$

Remarques :

1. Lorsqu'une expérience aléatoire est associée à une variable aléatoire X, toutes observation est traduite par un nombre réel x. La fonction ϕ permet de transformer ce réel en $y = \phi(x)$ que l'on considère alors comme la valeur observée d'une nouvelle variable aléatoire Y.
2. La propriété 4.1 nécessaire à la définition de toute variable aléatoire (voir chapitre 2) implique que l'on peut définir la loi de probabilité de Y, ce qui, a priori, n'est sans doute pas possible pour n'importe quelle fonction numérique.

Le but de ce chapitre étant essentiellement de montrer comment, à partir de la loi de X, on peut déduire la loi de Y, il sera implicite, dans ce qui suit, que les fonction ϕ considérées sont de "bonnes" fonctions.

1.1 Fonction d'une variable discrète

Lorsque la variable aléatoire X est une variable discrète, la variable $Y = \phi(X)$ est, elle-même discrète et la détermination de sa loi, à partir de celle de X, ne présente en général pas de difficultés.

Exemple 4.1 *Supposons que la loi de X soit donnée par le tableau suivant :*

x_i	-2	-1	0	1	2	3
p_i	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{2}{16}$

1. *Soit la variable $Y = 2X-3$. L'ensemble des observables de Y est $\{-7, -5, -3, -1, +1, +3\}$. La fonction $\phi : x \mapsto y = 2x-3$ est bijective et on a : $p_j = \mathbb{P}(Y = y_j) = \mathbb{P}\left(X = \frac{y_j+3}{2}\right)$.*
   On en déduit la loi de Y :

y_j	-7	-5	-3	-1	1	3
p_j	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{2}{16}$

2. *Soit la variable $Z = X^2$. L'ensemble des observables de Z est $\{0, 1, 4, 9\}$*
   On a, pour toute observable z_k strictement positive :

$$\begin{aligned} p_k = \mathbb{P}(Z = z_k) &= \mathbb{P}\left[(X = -\sqrt{z_k}) \cup (X = \sqrt{z_k})\right] \\ &= \mathbb{P}\left(X = -\sqrt{z_k}\right) + \mathbb{P}\left(X = \sqrt{z_k}\right) \end{aligned}$$

   car les événements sont incompatibles et on a $\mathbb{P}(Z = 0) = \mathbb{P}(X = 0)$. On en déduit la loi de Z :

z_k	0	1	4	9
p_k	$\frac{3}{16}$	$\frac{6}{16}$	$\frac{5}{16}$	$\frac{2}{16}$

3. *Peut-on définir la variable $\frac{1}{X}$?*

1.2 Fonction d'une variable à densité

Si X est une variable aléatoire de densité f, et de fonction de répartition F, on cherchera à déterminer la fonction de répartition G de la variable $Y = \phi(X)$; cette détermination sera plus ou moins compliqué suivant la nature de la fonction ϕ.

Exemple 4.2 *On suppose que la densité de X est :*

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x \leq 0 \\ x \mapsto \lambda^2 x e^{-\lambda x} & , \text{ si } x > 0 \end{cases}$$

où λ est un réel strictement positif. On cherche à déterminer la loi de $Y = \frac{1}{X}$.
On a $G(u) = \mathbb{P}(Y < u) = \mathbb{P}\left(\frac{1}{X} < u\right)$.
Pour $u \leq 0$, il est clair que l'événement "$\frac{1}{X} < u$" est un événement impossible puisque X, et donc $\frac{1}{X}$, ne peut prendre ici de valeurs négatives. On en déduit :

$$G(u) = \text{ , pour } u \leq 0$$

Pour $u > 0$, on a :

$$\begin{aligned} \mathbb{P}\left(\frac{1}{X} < u\right) &= \mathbb{P}\left(X > \frac{1}{u}\right) \\ &= 1 - \mathbb{P}\left(X \leq \frac{1}{u}\right) \\ &= 1 - \mathbb{P}\left(X < \frac{1}{u}\right) \text{ distribution continue} \end{aligned}$$

On en déduit :

$$G(u) = 1 - F\left(\frac{1}{u}\right) \text{ , pour } u > 0$$

Par dérivation, on obtient pour $u \leq 0$: $g(u) = 0$ et pour $u > 0$:
$g(u) = -\left(-\frac{1}{u^2}\right) f\left(\frac{1}{u}\right) = \frac{\lambda^2}{u^2}\frac{1}{u}e^{-\lambda/u}$.
D'où la densité de Y :

$$g : \begin{cases} u \mapsto 0 & , \text{ si } u \leq 0 \\ u \mapsto \frac{\lambda^2}{u^3}e^{-\lambda/u} & , \text{ si } u > 0 \end{cases}$$

Connaissant la fonction de répartition de X :

$$F : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto 1 - e^{-\lambda t}\left(1 + \lambda t\right) & , \text{ si } t > 0 \end{cases}$$

on peut évidemment expliciter :

$$G : \begin{cases} u \mapsto 0 & , \text{ si } u \leq 0 \\ u \mapsto 1 - e^{-\lambda u}\left(1 + \frac{\lambda}{u}\right) & , \text{ si } u > 0 \end{cases}$$

Exemple 4.3 *Soit X une variable aléatoire de loi $\mathcal{N}(0, 1)$. On se propose de déterminer la loi de $Y = X^2$.*
On écrit $G(t) = \mathbb{P}(Y < t) = \mathbb{P}(X^2 < t)$
Il est clair que, pour $t \leq 0$, $G(t) = 0$, puisque l'événement "$X^2 < t$" est impossible : un carré ne peut pas prendre une valeur négative ; par dérivation : $g(t) = 0$.
Pour $t > 0$, on a $\mathbb{P}(X^2 < t) = \mathbb{P}(-\sqrt{t} < X\sqrt{t})$ et, X étant de distribution continue : $G(t) = F(\sqrt{t}) - F(-\sqrt{t})$; par dérivation

$$\begin{aligned} g(t) &= \frac{1}{2\sqrt{t}}f(\sqrt{t}) - \left(-\frac{1}{2\sqrt{t}}\right)f(-\sqrt{t}) \\ &= \frac{1}{2\sqrt{2\pi t}}e^{-t/2} + \frac{1}{2\sqrt{2\pi t}}e^{-t/2} \end{aligned}$$

(on rappelle que f est une fonction paire définie sur $\mathbb{R}$ par $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$)
Finalement la densité de Y est :

$$g : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto \frac{1}{\sqrt{2\pi t}}e^{-t/2} & , \text{ si } t > 0 \end{cases}$$

Contrairement à l'exemple précédent, F n'ayant pas d'expression analytique, il est impossible, ici, d'expliciter G. On notera, à propos de cet exemple, que la loi d'une variable aléatoire définie comme le carré d'une variable normale réduite est une loi classique dite "loi du chi-deux à un degré de liberté". Les lois du chi-deux font partie des modèles gaussiens qui seront décrits dans les derniers chapitres de ce cours et qui sont très utilisées en statistique.

2 Un cas particulier : $Y = aX + b$

La modélisation de ce cas particulier est très aisé et se prête à des développement intéressants. Les paramètres a et b étant réels, on suppose, bien entendu, que a est différent de zéro, sinon l'événement "$Y = b$" est un événement certain.

2.1 *X est une variable discrète*

Les observables de X, rangées par valeurs croissantes, sont numérotées par un indice i représentant leur rang.
La fonction $\phi\ : x \mapsto y = ax + b$, $a \neq 0$, est bijective ; les valeurs possibles de X et Y sont donc associées par paires et on peut poser $y_i = ax_i + b$, ce qui revient à classer les observables de Y par valeurs croissantes si a est positif, ou décroissantes si a est négatif et on aura :

$$\mathbb{P}(Y = y_i) = \mathbb{P}\left(X = \frac{y_i - b}{a}\right) = \mathbb{P}(X = x_i)$$

La loi de Y se déduit donc aisément de celle de X (voir exemple 4.1). Intéressons-nous maintenant au calcul de $\mathbb{E}(Y)$ et de $\mathbb{V}(Y)$
On a : $\mathbb{E}(Y) = \sum_i y_i \mathbb{P}(Y = y_i)$, par définition.
Or $y_i = ax_i + b$ et $\mathbb{P}(Y = y_i) = \mathbb{P}(X = x_i)$.Donc :

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_i (ax_i + b)\mathbb{P}(X = x_i) \\
&= \sum_i ax_i \mathbb{P}(X = x_i) + \sum_i b\mathbb{P}(X = x_i) \\
&= a\sum_i x_i \mathbb{P}(X = x_i) + b\sum_i \mathbb{P}(X = x_i)
\end{aligned}$$

comme $\sum_i \mathbb{P}(X = x_i) = 1$, on a :

$$\mathbb{E}(Y) = a\mathbb{E}(X) + b \tag{4.2}$$

D'autre part :

$$\begin{aligned}
\mathbb{V}(Y) &= \sum_i (y_i - \mathbb{E}(y_i))^2\, \mathbb{P}(Y = y_i) \\
&= \sum_i (ax_i + b - a\mathbb{E}(X) - b)^2\, \mathbb{P}(X = x_i) \\
&= a^2 \sum_i (x_i - \mathbb{E}(x_i))^2\, \mathbb{P}(X = x_i)
\end{aligned}$$

D'où :

$$\mathbb{V}(Y) = a^2 \mathbb{V}(X) \tag{4.3}$$

2.2 X est une variable à densité

On désigne toujours par f et F la densité et la fonction de répartition de X et par g et G la densité et la fonction de répartition de Y. Deux cas sont à envisager :

a **est positif :** $\phi : x \mapsto t = ax + b$ est alors une fonction continue strictement croissante, on a : $G(t) = \mathbb{P}(Y < t) = \mathbb{P}(aX + b < t) = \mathbb{P}(aX < t - b) = \mathbb{P}\left(X < \frac{t-b}{a}\right)$
On en déduit : $G(t) = F\left(\frac{t-b}{a}\right)$ d'où

$$g(t) = \frac{1}{a} f\left(\frac{t-b}{a}\right)$$

$\mathbb{E}(Y) = \int_{-\infty}^{+\infty} tg(t)\mathrm{d}t = \int_{-\infty}^{+\infty} tf\left(\frac{t-b}{a}\right)\frac{\mathrm{d}t}{a}$
On pose $x = \frac{t-b}{a} \Leftrightarrow t = ax + b \Rightarrow \mathrm{d}t = a\mathrm{d}x$.

$$\begin{aligned}\mathbb{E}(Y) &= \int_{-\infty}^{+\infty} (ax+b) f(x)\mathrm{d}x \\ &= a\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x + b\int_{-\infty}^{+\infty} f(x)\mathrm{d}x \\ &= a\mathbb{E}(X) + b\end{aligned}$$

on retrouve bien la relation 4.2
$\mathbb{V}(Y) = \int_{-\infty}^{+\infty} (t - \mathbb{E}(Y))^2 g(t)\mathrm{d}t = \int_{-\infty}^{+\infty} (t - a\mathbb{E}(X) - b)^2 f\left(\frac{t-b}{a}\right)\frac{\mathrm{d}t}{a}$.
Le même changement de variable que précédemment donne :

$$\begin{aligned}\mathbb{V}(X) &= \int_{-\infty}^{+\infty} (ax - a\mathbb{E}(X))^2 f(x)\mathrm{d}x \\ &= a^2\int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f(x)\mathrm{d}x \\ &= a^2\mathbb{V}(X)\end{aligned}$$

on retrouve bien la relation 4.3

a **est négatif :** $\phi : x \mapsto t = ax + b$ est une fonction continue strictement décroissante, on a : $G(t) = \mathbb{P}(Y < t) = \mathbb{P}(aX + b < t) = \mathbb{P}(aX < t - b) = \mathbb{P}\left(X > \frac{t-b}{a}\right) = 1 - \mathbb{P}\left(X \leq \frac{t-b}{a}\right) = 1 - F\left(\frac{t-b}{a}\right)$ (distribution continue).
On en déduit :

$$g(t) = -\frac{1}{a} f\left(\frac{t-b}{a}\right)\ , (a < 0)$$

$\mathbb{E}(Y) = \int_{-\infty}^{+\infty} tg(t)\mathrm{d}t = -\int_{-\infty}^{+\infty} tf\left(\frac{t-b}{a}\right)\frac{\mathrm{d}t}{a}$
Le changement de variable $(x = \frac{t-b}{a} \Leftrightarrow t = ax + b) \Rightarrow \mathrm{d}t = a\mathrm{d}x$ donne ici :

$$\begin{aligned}\mathbb{E}(Y) &= -\int_{-\infty}^{+\infty} (ax+b) f(x)\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} (ax+b) f(x)\mathrm{d}x \\ &= a\mathbb{E}(X) + b\end{aligned}$$

on retrouve bien la relation 4.2

$$\begin{aligned}
\mathbb{V}(X) &= \int_{-\infty}^{+\infty} (t - \mathbb{E}(Y))^2 g(t)\mathrm{d}t \\
&= -\int_{-\infty}^{+\infty} (t - a\mathbb{E}(X) - b)^2 f\left(\frac{t-b}{a}\right)\frac{\mathrm{d}t}{a} \\
&= -\int_{-\infty}^{+\infty} (ax - a\mathbb{E}(X))^2 f(x)\mathrm{d}x \\
&= a^2\int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f(x)\mathrm{d}x \\
&= a^2\mathbb{V}(X)
\end{aligned}$$

on retrouve bien la relation 4.3

Il résulte de ces deux cas que :

- sous la réserve évidente de l'existence de $\mathbb{E}(X)$ et de $\mathbb{V}(X)$, les relations 4.2 et 4.3 sont absolument générales ;
- si X est une variable aléatoire de densité f, la densité g de la variable aléatoire $Y = aX + b$ ($a \neq 0$), est donnée par :

$$g(t) = \frac{1}{|a|} f\left(\frac{t-b}{a}\right) \tag{4.4}$$

Cette expression rend compte des deux cas $a > 0$ et $a < 0$.

Ces deux remarques permettent d'énoncer les théorèmes qui suivent.

Théorème 4.1 *Soit une variable aléatoire X dont l'espérance et la variance existent. Soit la variable aléatoire $Y = aX + b$ (a et b réels, $a \neq 0$). On a :*

$$\mathbb{E}(Y) = \mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

et

$$\mathbb{V}(Y) = \mathbb{V}(aX + b) = a^2\mathbb{V}(X)$$

La démonstration de ce théorème a été faite lorsque X est une variable discrète et lorsque X est une variable à densité.

Théorème 4.2 *Soit X une variable aléatoire de loi $\mathcal{N}(\mu, \sigma^2)$.*
La variable aléatoire $Y = aX + b$ (a et b réels, $a \neq 0$) est de loi $\mathcal{N}(a\mu + b, a^2\sigma^2)$.

Pour démontrer ce théorème, il suffit d'appliquer la relation 4.4.

Si $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, on a en effet :

$$\begin{aligned}
g(t) &= \frac{1}{|a|\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\frac{t-b}{a}-\mu}{\sigma}\right)^2} \\
&= \frac{1}{|a|\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-(a\mu+b)}{a\sigma}\right)^2}
\end{aligned}$$

Exemple 4.4 *Soit X une variable aléatoire de loi $\mathcal{B}(n,p)$. Cette variable a été définie comme le nombre d'événements favorables que l'on peut observer sur n expériences identiques et indépendantes. On sait que $\mathbb{E}(X) = np$ et $\mathbb{V}(X) = np(1-p)$.*
On peut également définir la variable aléatoire $Y = \frac{X}{n}$ qui représente alors la fréquence des événements favorables sur les n expériences. On déduit du théorème 4.1 :

$$\begin{aligned}
\mathbb{E}(Y) = \mathbb{E}\left(\frac{X}{n}\right) = \frac{1}{n}\mathbb{E}(X) &= \frac{1}{n}np = p \\
\mathbb{V}(Y) = \mathbb{V}\left(\frac{X}{n}\right) = \frac{1}{n^2}\mathbb{V}(X) &= \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}
\end{aligned}$$

Si, de plus, les conditions sont satisfaites pour que l'on puisse assimiler X à une variable normale, alors on peut considérer, d'après le théorème 4.2, que la variable Y suit une loi $\mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$.
L'utilisation de la variable Y représente un intérêt certain parce que la tendance centrale de la loi de Y est p, probabilité de l'événement favorable sur une expérience, et que la dispersion de Y autour de p tend vers zéro lorsque n tend vers l'infini.

2.3 Variable centrée–variable réduite

Définition 4.2 *une variable aléatoires à valeurs dans $\mathbb{R}$, est dite "centrée" si son espérance est égale à zéro. Une variable aléatoire, à valeurs dans $\mathbb{R}$, est dite "réduite" (on dit aussi centrée et réduite) si son espérance est égale à zéro et sa variance égale à un.*

À toute variable aléatoire X, d'espérance μ et d'écart-type σ, on peut associer :
- une variable aléatoire $X' = X - \mu$
- une variable aléatoire $X^* = \frac{X-\mu}{\sigma}$

La variable aléatoire X' est alors centrée et de variance σ^2. D'après le théorème 4.1, on a en effet :

$$\begin{aligned}
\mathbb{E}(X') = \mathbb{E}(X-\mu) &= \mathbb{E}(X) - \mu = \mu - \mu = 0 \\
\mathbb{V}(X') = \mathbb{V}(X-\mu) &= \mathbb{V}(X) = \sigma^2
\end{aligned}$$

La variable aléatoire X^* est alors réduite puisque :

$$\begin{aligned}
\mathbb{E}(X^*) = \mathbb{E}\left(\frac{X-\mu}{\sigma}\right) &= \frac{1}{\sigma}\mathbb{E}(X) - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0 \\
\mathbb{V}(X^*) = \mathbb{V}\left(\frac{X-\mu}{\sigma}\right) &= \frac{1}{\sigma^2}\mathbb{V}(X) = \frac{\sigma^2}{\sigma^2} = 1
\end{aligned}$$

Si, de plus, X est de loi $\mathcal{N}(\mu,\sigma^2)$, il résulte du théorème 4.2 que X' est de loi $\mathcal{N}(0,\sigma^2)$ et que X^* est de loi $\mathcal{N}(0,1)$.
Le passage de la variable normale X à la variable normale réduite, introduit au chapitre 3, peut en fait s'interpréter comme le choix d'une variable X^*, plus commode à manipuler, définie à partir de X par la relation $X^* = \frac{X-\mu}{\sigma}$; au lieu de raisonner sur la valeur observée x de X, on raisonne sur la valeur observée $\frac{x-\mu}{\sigma}$ de X^*.
Réciproquement, si X' est de loi $\mathcal{N}(0,\sigma^2)$, la variable $X = X' + \mu$ est de loi $\mathcal{N}(\mu,\sigma^2)$ et si X^* est de loi $\mathcal{N}(0,1)$, la variable $X = \sigma X^* + \mu$ est de loi $\mathcal{N}(\mu,\sigma^2)$.

3 Espérance de $Y = \phi(X)$

Soit $Y = \phi(X)$ une variable aléatoire fonction de la variable aléatoire X.
On pose, sous réserve de la convergence de la somme ou de l'intégrale :

$$\mathbb{E}[\phi(X)] = \sum_i \phi(x_i)\mathbb{P}(X = x_i)$$

si X est une variable discrète et

$$\mathbb{E}[\phi(X)] = \int_{-\infty}^{+\infty} \phi(x) f_X(x)\mathrm{d}x$$

si X est une variable de densité f_X.
Si $\mathbb{E}(Y)$ existe, alors on :

$$\mathbb{E}(Y) = \mathbb{E}[\phi(X)]$$

Cette propriété signifie que pour calculer l'espérance de Y, il n'est pas nécessaire de connaître la loi de Y : celle de X suffit.
Nous avons démontré cette propriété pour la variable $Y = aX + b$, ce qui correspond à l'égalité : $\mathbb{E}(Y) = a\mathbb{E}(X) + b$ dans l'énoncé du théorème 4.1.
Il est relativement facile de démontrer cette propriété, de manière générale, lorsque X est une variable discrète. Il suffit d'utiliser comme définition de l'espérance $\mathbb{E}(Y) = \sum_{\omega\in\Omega} Y(\omega)\mathbb{P}(\omega)$ (voir chapitre 2)
En regroupant tous les termes qui conduisent à la même valeur y_j, on retrouve la définition classique :

$$\mathbb{E}(Y) = \sum_j y_j \sum_{\omega|Y(\omega)=y_j} \mathbb{P}(\omega) = \sum_j y_j\mathbb{P}(Y = y_j)$$

Mais on peut également écrire $\mathbb{E}(Y) = \sum_{\omega\in\Omega} (\phi(X(\omega))\mathbb{P}(\omega))$, expression qui donne, en regroupant tous les termes qui conduisent à la même valeur x_i :

$$\mathbb{E}(Y) = \sum_i \phi(x_i) \sum_{\omega|X(\omega)=x_i} \mathbb{P}(\omega) = \sum_i \phi(x_i)\mathbb{P}(X = x_i)$$

On admettra cette propriété, qui sera vérifiée sur des exemple, lorsque X est une variable à densité.
La notation $\mathbb{E}[\phi(X)]$ a déjà été introduites dans les chapitres précédents pour définir les paramètres descriptifs d'une distribution ; on peut maintenant en donner une autre interprétation :

- le moment d'ordre k de la variable X est aussi l'espérance de la variable $Y = X^k$;
- la variance de la variable X est aussi l'espérance de la variable $Z = (X - \mu)$

Exemple 4.5 *(suite de l'exemple 4.1)*
On a déterminé la loi de probabilité de la variable $Z = X^2$. On a :

$$\mathbb{E}(Z) = \frac{3}{16}0 + \frac{6}{16}1 + \frac{5}{16}4 + \frac{2}{16}9 = \frac{44}{16}$$

et

$$\mathbb{E}(X^2) = \frac{3}{16}(-2)^2 + \frac{2}{16}(-1)^2 + \frac{3}{16}(0)^2 + \frac{4}{16}(1)^2 + \frac{4}{16}(2)^2 + \frac{2}{16}(3)^2 = \frac{44}{16}$$

On a bien $\mathbb{E}(Z) = \mathbb{E}(X^2)$, ce qui n'est pas un hasard puisque

$$\mathbb{E}(X^2) = \frac{3}{16}0 + \left(\frac{2}{16} + \frac{4}{16}\right)1 + \left(\frac{4}{16} + \frac{1}{16}\right)4 + \frac{2}{16}9 = \sum_k z_k \mathbb{P}(Z = z_k)$$

Exemple 4.6 *(suite de l'exemple 4.3)*
On se propose de calculer l'espérance de la variable $Y = X^2$ égale au carré d'une variable normale réduite.
On peut utiliser la densité de Y que l'on a déjà déterminé :

$$\mathbb{E}(Y) = \int_{-\infty}^{+\infty} tg(t)dt = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi t}} \frac{t}{\sqrt{t}} e^{-t/2} dt$$

Si on effectue le changement de variable $x = \sqrt{t}$ (d'où $dx = \frac{dt}{2\sqrt{t}}$), on obtient :

$$\mathbb{E}(Y) = \int_0^{+\infty} \frac{1}{\sqrt{2\pi t}} 2x^2 e^{-x^2/2} dx$$

On reconnaît, dans cette expression, $\mathbb{E}(X^2)$, puisque X est de loi $\mathcal{N}(0,1)$. On vient de démontrer, sur cet exemple $\mathbb{E}(Y) = \mathbb{E}(X^2)$, il aurait été plus simple d'écrire ceci directement. Le calcul peut se poursuivre en intégrant par parties. Il est plus rapide de remarquer que, puisque $\mathbb{E}(X) = 0$ et $\mathbb{V}(X) = 1$, la relation générale $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$ permet d'écrire $\mathbb{E}(X^2) = 1$.
On a donc : $\mathbb{E}(Y) = \mathbb{E}(X^2) = 1$
Calculons maintenant le moment d'ordre 2 de Y. On a :

$$\mathbb{E}(Y^2) = \mathbb{E}(X^4) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^4 e^{-x^2/2} dx$$

En intégrant par parties, on obtient $\mathbb{E}(Y^2) = \left[\frac{-x^3}{\sqrt{2\pi}} e^{-x^2/2}\right]_{-\infty}^{+\infty} + 3\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-x^2/2} dx$
Puisque $\lim_{x\to\pm\infty} x^3 e^{-x^2/2} = 0$: $\mathbb{E}(Y^2) = 3\mathbb{E}(X^2) = 3$.
D'où

$$\mathbb{V}(Y) = \mathbb{E}(Y^2) - \mathbb{E}^2(Y) = 3 - 1 = 2$$

On vient de démontrer, et ceci est un résultat qui sera utilisé ultérieurement, que pour une variable Y qui suit une loi du chi-deux à un degré de liberté,

$$\mathbb{E}(Y) = 1 \quad \textit{et} \quad \mathbb{V}(Y) = 2$$

4 Exercices

1. Soit X une variable aléatoire de densité de probabilité :

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x \leq 0 \\ x \mapsto \frac{1}{ax\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x - b}{a}\right)^2} & , \text{ si } x > 0 \end{cases}$$

   où $a \in \mathbb{R}_+^*$ et $b \in \mathbb{R}$ sont deux réels donnés (loi log-normale).
   Déterminer la densité de probabilité de la variable $Y = \log X$

2. Soit X une variable aléatoire de densité de probabilité :

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x \leq 0 \\ x \mapsto e^{-x} & , \text{ si } x > 0 \end{cases}$$

   Déterminer la densité de probabilité de la variable $Y = 2X+3$. Calculer l'espérance et la variance de Y.

3. Soit X une variable aléatoire de densité de probabilité :

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x \leq 0 \\ x \mapsto \lambda^2 x e^{-\lambda x} & , \text{ si } x > 0 \end{cases}$$

   où $\lambda \in \mathbb{R}_+^*$.
   Calculer l'espérance et la variance de $Y = \lambda X$.
   Déterminer la densité de probabilité de la variable $Y = \lambda X$ et $Z = X^2$.

4. Soit V une variable aléatoire de densité de probabilité :

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x \leq 0 \\ x \mapsto \lambda e^{-\lambda x} & , \text{ si } x > 0 \end{cases}$$

   où $\lambda \in \mathbb{R}_+^*$.

   (a) Calculer l'espérance et la variance de V.
   (b) Déterminer la densité de probabilité de la variable $U = 5V$. Calculer son espérance et sa variance.
   (c) Déterminer la densité de probabilité de la variable $W = V^2$.
   (d) Déterminer la densité de probabilité de la variable $Z = \frac{1}{V}$.

5. Soit X une variable aléatoire à valeurs réelles, de fonction de répartition F, continue et strictement croissante. Soit la variable aléatoire $Y = F(X)$. Montrer que Y est uniformément distribuée sur $[0, 1]$.

6. Soit X une variable aléatoire de densité de probabilité :

$$f : \begin{cases} x \mapsto \frac{2}{15} x & , \text{ si } x \in [1, 4] \\ x \mapsto 0 & , \text{ si } x \notin [1, 4] \end{cases}$$

   Déterminer la densité de probabilité de la variable $Y = (X-2)^2$.

7. Des résistances sont fabriquées en série. La résistance d'un élément choisi au hasard dans la fabrication est une variable aléatoire de distribution uniforme entre 9 et 11 ohms. Déterminer la densité de probabilité de la conductance $C = \frac{1}{R}$.
8. Un générateur produit une tension de la forme $V = \sin \omega t$. Pour quelqu'un qui vient mesurer la tension V à un instant quelconque, la mesure est une variable aléatoire ; déterminer sa densité de probabilité.
9. On observe une population de bactéries en croissance asynchrone. Soit T l'âge d'une bactérie choisie au hasard. On admet que la densité de probabilité de T est :

$$f : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto k2^{1-t/\theta} & , \text{ si } 0 < t \leq \theta \\ t \mapsto 0 & , \text{ si } t \geq \theta \end{cases}$$

   (a) Calculer k.
   (b) Calculer la probabilité pour que l'âge d'une bactérie choisie au hasard soit compris entre 0 et $\theta/2$.
   (c) On admet que la masse d'une bactérie est fonction linéaire de son âge : $m = m_0 \left(1 + \frac{t}{\theta}\right)$. Calculer l'espérance de la masse d'une bactérie choisie au hasard.
10. Soit X une variable aléatoire de densité de probabilité :

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x \notin [0,1] \\ x \mapsto 2x & , \text{ si } x \in [0,1] \end{cases}$$

   Soit la variable aléatoire $Y = \arctan Y$
   (a) Calculer l'espérance et la variance de Y.
   (b) Déterminer la fonction de répartition, puis la densité de probabilité de Yet vérifier les résultats précédents.
11. Soit X une variable aléatoire normale réduite. Écrire la densité de probabilité de la variable $Y = \frac{X}{2} - 1$. Déterminer la densité de probabilité de la variable $Z = X^3$.
12. La probabilité pour qu'un individu, issu d'un certain croisement présente un phénotype donné est égale à $3/4$. Soit X la variable aléatoire associant à un échantillon de n individus,, issus du même croisement, la proportion des individus portant ce phénotype. Calculer n pour que l'on ait

$$\mathbb{P}(0.70 < X < 0.80) \geq 0.95$$

13. Soit X la variable aléatoire à valeurs réelles de densité de probabilité :

$$f : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto \frac{1}{2} e^{-t/2} & , \text{ si } t > 0 \end{cases}$$

   Déterminer la densité de probabilité de la variable $T = \sqrt{X}$.

5 Solutions

1. Soit F la fonction de répartition de X. On a : $F(u) = \mathbb{P}(X < u) = \int_{-\infty}^{u} f(x)\mathrm{d}x$
   Si $u \leq 0$, $F(u) = 0$.
   Si $u > 0$, $F(u) = \int_{-\infty}^{0} f(x)\mathrm{d}x + \int_{0}^{u} f(x)\mathrm{d}x = int_0^u f(x)\mathrm{d}x$.
   Le fonction bijective
   $$\phi : \left\{ \begin{array}{ccc} \mathbb{R}_+^* & \rightarrow & \mathbb{R} \\ u & \mapsto & \log u \end{array} \right.$$
   n'est pas définie pour $u \leq 0$; cela n'a pas d'importance ici puisque la variable X ne peut pas prendre de valeur dans $\mathbb{R}^-$.
   Soit G et g la fonction de répartition et la densité de la variable aléatoire $Y = \log X$. On a, puisque ϕ est strictement croissante et admet, pour fonction inverse ϕ^{-1}, la fonction exponentielle :
   $$G(t) = \mathbb{P}(Y < t) = \mathbb{P}(\log t) = \mathbb{P}(0 < X < e^t) = \int_0^{e^t} f(x)\mathrm{d}x = F(e^t)\ , \forall t$$
   On notera que $e^t > 0\ , \forall t$.

   Par dérivation, on obtient : $g(t) = e^t f(e^t) = e^t \frac{1}{a\sqrt{2\pi}e^t} e^{-\frac{1}{2}\left(\frac{\log e^t - b}{a}\right)^2}$
   D'où la densité de Y :
   $$g : t \mapsto \frac{1}{a\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-b}{a}\right)^2}\ , \forall t, a \in \mathbb{R}_+^*, b \in \mathbb{R}$$
   On en conclut que Y suit une loi $\mathcal{N}(b, a^2)$
2. la fonction :
   $$\phi : \left\{ \begin{array}{ccc} \mathbb{R} & \rightarrow & \mathbb{R} \\ u & \mapsto & 2u+3 \end{array} \right.$$
   est une bijection, strictement croissante. On a, d'autre part : $F_X(u) = \mathbb{P}(X < u)$, d'où :
   $F_X(u) = 0$, si $u \leq 0$
   $F_X(u) = \int_0^u e^{-x}\mathrm{d}x = 1 - e^{-u}$, si $u > 0$ (loi exponentielle).
   On a $F_Y(t) = \mathbb{P}(Y < t) = \mathbb{P}(2X + 3 < t) = \mathbb{P}\left(X < \frac{t-3}{2}\right) = F_X\left(\frac{t-3}{2}\right)$
   Par dérivation, on obtient : $f_Y(t) = \frac{1}{2} f_X\left(\frac{t-3}{2}\right)$.
   D'où la densité (voir relation 4.4) :
   $$f_Y : \left\{ \begin{array}{ll} t \mapsto 0 & , \text{ si } t \leq 3 \\ x \mapsto e^{-\frac{t-3}{2}} & , \text{ si } t > 3 \end{array} \right.$$
   On a d'autre part (théorème 4.1) : $\mathbb{E}(Y) = \mathbb{E}(2X + 3) = 2\mathbb{E}(X) + 3 = 5$ et $\mathbb{V}(Y) = \mathbb{V}(2X) = 4\mathbb{V}(X) = 4$

3. On a

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x = \int_{0}^{+\infty} \lambda^2 x^2 e^{-\lambda x}\mathrm{d}x = \left[-\lambda x^2 e^{-\lambda x}\right]_0^{+\infty} + \frac{2}{\lambda}\int_0^{+\infty} \lambda^2 x e^{-\lambda x}\mathrm{d}x = \frac{2}{\lambda}$$

En effet $\lim_{x\to+\infty} \lambda x^2 e^{-\lambda x} = 0$ et $\lambda x^2 e^{-\lambda x} = 0$ pour $x = 0$, $int_0^{+\infty}\lambda^2 x e^{-\lambda x}\mathrm{d}x = 1$.
De même :

$$\mathbb{E}(X^2) = \int_{0}^{+\infty} \lambda^2 x^3 e^{-\lambda x}\mathrm{d}x = \left[-\lambda x^3 e^{-\lambda x}\right]_0^{+\infty} + \frac{3}{\lambda}\int_0^{+\infty} \lambda^2 x^2 e^{-\lambda x}\mathrm{d}x = \frac{3}{\lambda}\mathbb{E}(X) = \frac{6}{\lambda^2}$$

et donc $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{2}{\lambda^2}$
On en déduit (voir théorème 4.1) :

$$\begin{aligned} \mathbb{E}(Y) &= \mathbb{E}(\lambda X) = \lambda\mathbb{E}(X) = 2 \\ \mathbb{V}(Y) &= \mathbb{V}(\lambda X) = \lambda^2\mathbb{V}(X) = 2 \end{aligned}$$

On désignera, dans l'exercice par F_X la fonction de répartition de X. On n'aura pas besoin de son expression analytique ; il est cependant facile de montrer que :

$$F_X : \begin{cases} u \mapsto 0 & , \text{ si } u \leq 0 \\ u \mapsto 1 - (1+\lambda u)e^{-\lambda u} & , \text{ si } u > 0 \end{cases}$$

On a, $\forall t \in \mathbb{R}$, $F_Y(t) = \mathbb{P}(Y < t) = \mathbb{P}(\lambda X < t) = \mathbb{P}(X < t/\lambda) = F_X\left(\frac{t}{\lambda}\right)$
On en déduit la densité de Y : $f_Y(t) = \frac{1}{\lambda} f_X\left(\frac{t}{\lambda}\right)$
On obtient finalement :

$$f_Y : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto te^{-t} & , \text{ si } t > 0 \end{cases}$$

On constate que Y suit une loi du même type que celle de X avec $\lambda = 1$.
Contrairement aux exemples précédents, la fonction $\phi : x \mapsto x^2$ n'est pas bijective.
On aura cependant, $\forall t \in \mathbb{R}$, $F_Z(t) = \mathbb{P}(Z < t) = \mathbb{P}(X^2 < t)$
Pour $t \leq 0$, il est clair que $(X^2 < t) = \emptyset \Rightarrow F_Z(t) = 0 \Rightarrow f_Z(t) = 0$
Pour $t > 0$: $\mathbb{P}(X^2 < t) = \mathbb{P}(-\sqrt{t} < X < \sqrt{t}) = F_X(\sqrt{t}) - F_X(-\sqrt{t}) \Rightarrow f_Z(t) = \frac{1}{2\sqrt{t}} f_X(\sqrt{t}) + \frac{1}{2\sqrt{t}} f_X(-\sqrt{t}) = \frac{\lambda^2}{2\sqrt{t}}\sqrt{t}e^{-\lambda\sqrt{t}} + 0$ ($-\sqrt{t} < 0$).
On peut éventuellement vérifier que $\mathbb{E}(Z) = \mathbb{E}(X^2) = \frac{6}{\lambda^2}$

4. Réponses :
   (a) $\mathbb{E}(V) = \frac{1}{\lambda}$; $\mathbb{V}(V) = \frac{1}{\lambda^2}$
   (b) $\mathbb{E}(U) = 5\mathbb{E}(V) = \frac{5}{\lambda}$; $\mathbb{V}(U) = 25\mathbb{V}(V) = \frac{25}{\lambda^2}$

$$f_U : \begin{cases} t \mapsto 0 & , \text{ si } xt \leq 0 \\ t \mapsto \frac{\lambda}{5}e^{-\lambda t/5} & , \text{ si } t > 0 \end{cases}$$

(c)

$$f_W : \begin{cases} t \mapsto 0 & , \quad \text{si } xt \leq 0 \\ t \mapsto \frac{\lambda}{2\sqrt{t}} e^{-\lambda\sqrt{t}} & , \quad \text{si } t > 0 \end{cases}$$

(d)

$$f_Z : \begin{cases} t \mapsto 0 & , \quad \text{si } t \leq 0 \\ t \mapsto \frac{\lambda}{t^2} e^{-\lambda/t} & , \quad \text{si } t > 0 \end{cases}$$

5. On désigne par G et g la fonction de répartition et la densité de probabilité de la variable aléatoire $Y = F(X)$. On a $\forall t : G(t) = \mathbb{P}(Y < t) = \mathbb{P}(F(X) < t)$.

   Il semble, a priori, difficile d'aller plus loin puisqu'on ne connaît pas la loi de X. Pourtant nos informations ne sont pas nulles.

   Considérons :

$$F : \begin{cases} \mathbb{R} & \rightarrow & [0,1] \\ x & \mapsto & F(x) = \mathbb{P}(X < x) \end{cases}$$

   Comme toutes les fonction de répartition, $F(X)$ croît de 0 à 1. On peut en déduire :

$$\begin{cases} \mathbb{P}(F(X) < t) & = & 0 \text{ si } t \leq 0 \\ \mathbb{P}(F(X) < t) & = & 1 \text{ si } t > 1 \end{cases}$$

   Il est, d'autre part, précisé dans l'énoncé que F est continue et *strictement croissante* ; F admet donc une fonction inverse F^{-1} définie sur $]0,1]$. On a donc, pour $0 < t \leq 1 : G(t) = \mathbb{P}(F(X) < t) = \mathbb{P}(X < F^{-1}(t)) = F\left(F^{-1}(t)\right) = t$.

   Finalement :

$$G : \begin{cases} t \mapsto 0 & , \quad \text{si } t \leq 0 \\ t \mapsto t & , \quad \text{si } 0 < t \leq 1 \\ t \mapsto 1 & , \quad \text{si } t > 1 \end{cases} \tag{4.5}$$

   et donc

$$g : \begin{cases} t \mapsto 0 & , \quad \text{si } t \leq 0 \text{ ou } t > 1 \\ t \mapsto 1 & , \quad \text{si } 0 < t \leq 1 \end{cases}$$

   La variable aléatoire $Y = F(X)$ est uniformément distribuée sur $[0,1]$

6. Soit F la fonction de répartition de la variable aléatoire X. On remarque que :

$$F(u) = \int_{-\infty}^{u} f(x)\mathrm{d}x : \begin{cases} = 0 & , \quad \text{si } u \leq 1 \\ = \frac{1}{15}(u^2 - 1) & , \quad \text{si } 1 < t \leq 4 \\ = 1 & , \quad \text{si } u > 4 \end{cases}$$

   On désigne par G et g la fonction de répartition et la densité de probabilité de la variable aléatoire $Y = (X-2)^2$. La fonction :

$$F_X : \begin{cases} \mathbb{R} & \rightarrow & \mathbb{R}_+ \\ x & \mapsto & (x-2)^2 \end{cases}$$

   n'est pas bijective. On a cependant :

- Si $t \leq 0$ $G(t) = \mathbb{P}(Y < t) = 0$
- Si $t > 0$ $G(t) = \mathbb{P}(Y < t) = \mathbb{P}(X-2)^2 < t) = \mathbb{P}(-\sqrt{t} < X - 2 < \sqrt{t}) = \mathbb{P}(2 - \sqrt{t} < X < 2 + \sqrt{t}) = F_X(2+\sqrt{t}) - F_X(2-\sqrt{t})$

Il faut discuter, en fonction de t, les expressions de $F_X(2+\sqrt{t})$ et de $F_X(2-\sqrt{t})$, qu'il faut choisir selon les équations 4.5

$$F_X(2+\sqrt{t}) \begin{cases} 2+\sqrt{t} \leq 1 & , \text{ impossible} \\ 1 < 2+\sqrt{t} \leq 4 & , \text{ si } 0 < t \leq 4 \\ 2+\sqrt{t} > 4 & , \text{ si } t > 4 \end{cases}$$

$$F_X(2-\sqrt{t}) \begin{cases} 2-\sqrt{t} \leq 1 & , \text{ si } t \geq 1 \\ 1 < 2-\sqrt{t} \leq 4 & , \text{ si } 0 < t \leq 1 \\ 2-\sqrt{t} > 4 & , \text{ impossible} \end{cases}$$

On en déduit, pour $t > 0$:

$$\begin{aligned} \text{Si } 0 < t < 1 \quad &, \quad G(t) = \frac{1}{15}[(2+\sqrt{t})^2 - 1] - \frac{1}{15}[(2-\sqrt{t})^2 - 1] = \frac{8}{15}\sqrt{t} \\ \text{Si } 1 \leq t \leq 4 \quad &, \quad G(t) = \frac{1}{15}[(2+\sqrt{t})^2 - 1] - 0 \\ \text{Si } t > \quad &, \quad G(t) = 1 - 0 \end{aligned}$$

D'où la densité :

$$g : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto \frac{4}{15}\frac{1}{\sqrt{t}} & , \text{ si } 0 < t \leq 1 \\ t \mapsto \frac{1}{15}\left(\frac{2}{\sqrt{t}} + 1\right) & , \text{ si } 1 < t \leq 4 \\ t \mapsto 1 & , \text{ si } t > 4 \end{cases}$$

7. Soient f et F la densité et la fonction de répartition de R, g et G celles de C. On a :

$$f : \begin{cases} t \mapsto 0 & , \text{ si } t \notin [9, 11] \\ t \mapsto \frac{1}{2} & , \text{ si } t \in [9, 11] \end{cases}$$

$$F : \begin{cases} u \mapsto 0 & , \text{ si } u < 9 \\ u \mapsto \frac{1}{2}(u-9) & , \text{ si } 9 \leq u \leq 11 \\ u \mapsto 1 & , \text{ si } u > 11 \end{cases}$$

La fonction

$$\phi : \begin{cases} [9, 11] & \rightarrow & \left[\frac{1}{11}, \frac{1}{9}\right] \\ x & \mapsto & \frac{1}{x} \end{cases}$$

qui permet de définir C à partir de R est ici continue et strictement décroissante ; il s'agit encore d'une bijection.

soit $G(t) = \mathbb{P}(C < t)$. Il est clair que $\mathbb{P}(C < t) = 0$ si $t \leq 1/11$ et que $\mathbb{P}(C < t) = 1$ si $t > 1/9$. Pour $1/11 < t \leq 1/9$, on a : $\mathbb{P}(C < t) = \mathbb{P}\left(\frac{1}{R} < t\right) = \mathbb{P}\left(R > \frac{1}{t}\right) = 1 - F\left(\frac{1}{t}\right)$.

Par dérivation, on obtient la densité :

$$g : \begin{cases} t \mapsto 0 & , \text{ si } t \notin [9, 11] \\ t \mapsto \frac{1}{2t^2} & , \text{ si } t \in [9, 11] \end{cases}$$

8. Exercice difficile. La variable T "instant où on fait la mesure" sera considérée comme uniformément distribuée sur $\left[-\frac{\pi}{\omega}, \frac{\pi}{\omega}\right]$. On tracera le graphe de $\sin \omega t$ en fonction de ωt pour retrouver les événements équivalents à "$V < u$".
   Réponses : $f_V(u) = \frac{1}{\pi\sqrt{1-u^2}}$, si $-1 < u < 1$; $f_V(u) = 0$ sinon.
9. On notera que $2^{1-t/\theta} = 2 \times 2^{-t/\theta} = 2e^{-\frac{t}{\theta}\ln 2}$
   Réponses :
   (a) $k = \frac{\ln 2}{\theta}$;
   (b) $2\frac{\ln 2}{\theta}\int_0^{\theta/2} e^{-\frac{t}{\theta}\ln 2}\mathrm{d}t = 2\left(1 - \frac{\sqrt{2}}{2}\right) \approx 58.6\%$
   (c) $\mathbb{E}(M) = m_0 + \frac{m_0}{\theta}\mathbb{E}(T) = \frac{m_0}{\ln 2}$
10. (a) On peut écrire

$$\begin{aligned}
\mathbb{E}(Y) &= \mathbb{E}(\arctan X) = \int_{-\infty}^{+\infty} \arctan x f(x)\mathrm{d}x = \int_0^1 2x\arctan x\mathrm{d}x \\
&= \left[x^2\arctan x\right]_0^1 - \int_0^1 \frac{x^2}{1+x^2}\mathrm{d}x = \frac{\pi}{4} - \int_0^1 \frac{x^2+1-1}{1+x^2}\mathrm{d}x \\
&= \frac{\pi}{4} - [x - \arctan x]_0^1 \\
&= \frac{\pi}{2} - 1
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(Y^2) &= \mathbb{E}((\arctan X)^2) = \int_{-\infty}^{+\infty} (\arctan x)^2 f(x)\mathrm{d}x = \int_0^1 2x(\arctan x)^2\mathrm{d}x \\
&= \left[x^2(\arctan x)^2\right]_0^1 - \int_0^1 2x^2\frac{\arctan x}{1+x^2}\mathrm{d}x = \frac{\pi^2}{16} - 2\int_0^1 \frac{x^2+1-1}{1+x^2}\arctan x\mathrm{d}x \\
&= \frac{\pi^2}{16} - 2\int_0^1 \arctan x\mathrm{d}x + \int_0^1 \frac{2\arctan x}{1+x^2}\mathrm{d}x\,[x - \arctan x]_0^1 \\
&= \frac{\pi^2}{16} - 2[x\arctan x]_0^1 + 2\int_0^1 \frac{x}{1+x^2}\mathrm{d}x + [(\arctan x)^2]_0^1 \\
&= \frac{\pi^2}{16} - \frac{\pi}{2} + [\log(1+x^2)]_0^1 + \frac{\pi^2}{16} \\
&= \frac{\pi^2}{8} - \frac{\pi}{2} + \log 2
\end{aligned}$$

$$V(Y) = \mathbb{E}(Y^2) - \mathbb{E}^2(Y) = \frac{\pi^2}{8} - \frac{\pi}{2} + \log 2 - \frac{\pi^2}{4} + \pi - 1 = \frac{\pi}{2} + \log 2 - \frac{\pi^2}{8} - 1$$

   (b) On désigne par f et F la densité et la fonction de répartition de X et par g et G celles de Y. On a :

$$F : \begin{cases} x \mapsto 0 & , \text{ si } x \leq 0 \\ x \mapsto x^2 & , \text{ si } 0 < x \leq 1 \\ x \mapsto 1 & , \text{ si } x > 1 \end{cases}$$

la fonction

$$\phi : \begin{cases} \mathbb{R} & \to & \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[\\ x & \mapsto & \arctan x \end{cases}$$

est une fonction bijective, continue et strictement croissante. On peut donc écrire :

$$G(t) = \mathbb{P}(Y < t) = \mathbb{P}(\arctan X < t) = \mathbb{P}(X < \tan t)$$

d'où $G(t) = F(\tan t)$, $\forall t \in \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$

$$G : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \text{ en fait } -\frac{\pi}{2} < t \leq 0 \text{ mais } \arctan X \leq -\frac{\pi}{2} \text{ impossible} \\ t \mapsto \tan^2 t & , \text{ si } 0 < t \leq \frac{\pi}{4}\ X(\Omega =]0,1] \Rightarrow Y(\Omega) = \left]0, \frac{\pi}{4}\right] \\ t \mapsto 1 & , \text{ si } t > \frac{\pi}{4} \text{ en fait } \frac{\pi}{4} < t < \frac{\pi}{2} \text{ mais } \arctan X < \frac{\pi}{2} \text{ est certain} \end{cases}$$

$$g : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \text{ ou } t > \frac{\pi}{4} \\ t \mapsto 2\tan t(1+\tan^2 t) & , \text{ si } 0 < t \leq \frac{\pi}{4} \end{cases}$$

(c) Il est facile de vérifier
$\mathbb{E}(Y) = \int_0^{\pi/4} t 2\tan t(1+\tan^2 t)\mathrm{d}t = \int_0^{\pi/4} t\mathrm{d}(\tan^2 t) = [t\tan^2 t]_0^{\pi/4} - \int_0^{\pi/4} \tan^2 t\mathrm{d}t = \frac{\pi}{4} - \int_0^{\pi/4}(\tan^2 t + 1 - 1)\mathrm{d}t = \frac{\pi}{4} - [\tan t - t]_0^{\pi/4} = \frac{\pi}{2} - 1$

11. Réponses : $Y \sim \mathcal{N}(-1, 1/4)$; $f_Z(t) = \frac{1}{3\sqrt{2\pi}\sqrt[3]{t^2}} e^{-\frac{1}{2}\sqrt[3]{t^2}}$, $\forall t \neq 0$.
12. Réponse : $n \geq 289$
13. Réponse :

$$f_T : \begin{cases} u \mapsto 0 & , \text{ si } u \leq 0 \\ u \mapsto ue^{-u^2/2} & , \text{ si } u > 0 \end{cases}$$

Chapitre 5

Tests d'hypothèses

1 Problématique

Si un expérimentateur place des animaux ou des plantes dans certaines conditions particulières, c'est qu'il a certainement fait l'hypothèse préalable que ces conditions doivent avoir une influence sur la taille ou la vitesse de développement, ou la forme, de ces animaux ou de ces plantes. De même si un écologue observe dans une région donnée un micro climat et/ou une zone dont la pédologie ou l'hydrologie diffère sensiblement du reste de cette région, il s'attend à ce que cet environnement particulier ait une influence sur la taille, la croissance, etc.. des êtres vivants qui y habitent.
Pour démontrer l'influence de l'environnement ou des conditions créés par l'expérimentateur, on effectue généralement des mesures sur les êtres vivants étudiés. En général on ne peut mesurer qu'un nombre limité d'unités expérimentales.
Le principe du test statistique consiste à supposer que ces conditions n'ont pas d'influence, ce sera l'hypothèse privilégiée (appelée H_0), et à voir si les mesures faites auraient pu arriver par le simple jeu du "hasard". Si la probabilité d'observer de telles mesures est trop faible on rejettera H_0 au profit d'une hypothèse alternative (appelée H_1) qui formulera en termes précis l'influence étudiée.

Exemple 5.1 *La médiane des tailles des français (âgés de plus de 18 ans) est de 170 cm. On aimerait savoir si c'est aussi le cas parmi les étudiants de la faculté Saint Jérôme. Pour cela cela on mesure la taille d'un échantillon de n étudiants.*

1.1 Hypothèse

Construire un test de l'hypothèse nulle H_0 contre l'hypothèse alternative H_1, c'est établir un critère de décision permettant de choisir entre l'hypothèse H_0 et H_1.

H_0 hypothèse privilégiée

H_0 est l'hypothèse privilégiée : c'est celle que l'on garde si le résultat de l'expérience n'est par clair. On dit que le test d'hypothèse est *conservatif* car il conserve H_0 sauf si

les données conduisent à la rejeter. En ce sens il y a une analogie entre un test d'hypothèse et un procès : tout suspect est présumé innocent et l'accusation doit apporter la preuve de sa culpabilité avant que la justice décide de le condamner. Cette preuve doit de plus s'appuyer sur des éléments matériels, comme le test qui utilise les données pour rejeter l'hypothèse. Quand on accepte H_0, on ne prouve pas qu'elle est vraie, on accepte de conserver H_0 parce qu'on a pas pu accumuler suffisamment de preuves contre elle. Accepter H_0 c'est *acquitter faute de preuve*. Dans l'exemple 5.1, on a H_0 : la médiane des tailles des étudiants est de 170 cm.
On appelle hypothèse alternative, que l'on note H_1, une hypothèse différente de H_0. On dit que l'on test H_0 contre H_1.
Dans ce qui suit θ est un paramètre relatif à une population, donc inconnu. Dans l'exemple 5.1, θ est la médiane.
Différents cas sont possibles pour le couple (H_0, H_1) :

- Hypothèses simples :
$$H_0 \ : \theta = \theta_0 \ \text{ et } \ H_1 \ : \theta = \theta_1$$
Par exemple $H_1 : \theta = 180$cm pour l'exemple 5.1
- H_0 simple et test bilatéral
$$H_0 \ : \theta = \theta_0 \ \text{ et } \ H_1 \ : \theta \neq \theta_0$$
Par exemple $H_1 : \theta \neq 170$cm pour l'exemple 5.1
- H_0 composite et test unilatéral :
$$H_0 \ : \theta \leq \theta_0 \ \text{ et } \ H_1 \ : \theta > \theta_0$$
- H_0 composite et test bilatéral :
$$H_0 \ : \theta_1 \leq \theta \leq \theta_2 \ \text{ et } \ H_1 \ : \{\theta > \theta_2\} \cup \{\theta < \theta_1\}$$

2 Niveau et puissance

On détermine la loi de la variable aléatoire d'étude sous H_0 et sous H_1. L'ensemble des valeurs observées pour lesquelles l'hypothèse nulle H_0 est admissible forme la région d'acceptation A ou de non-rejet de H_0 et les autres valeurs constituent la région de rejet ou domaine de rejet R ou région critique. Évidemment le type de test (unilatéral ou bilatéral) affecte la forme de la région d'acceptation A et de la zone de rejet R.
Le hasard de l'échantillonnage peut fausser les conclusions. Quatre situations doivent être envisagées :

- l'acceptation de l'hypothèse nulle alors qu'elle est vraie ;
- le rejet de l'hypothèse nulle alors qu'elle est vraie ;
- l'acceptation de l'hypothèse nulle alors qu'elle est fausse ;
- le rejet de l'hypothèse nulle alors qu'elle est fausse.

Dans le premier et le dernier cas, la conclusion obtenue est correcte, mais non dans les deux cas intermédiaires.

L'erreur qui consiste à rejeter une hypothèse vraie est appelée erreur de première espèce. Si H_0 est simple, la probabilité de rejeter à tort H_0 vaut

$$\alpha = \mathbb{P}_{H_0}(Z \in R)$$

L'erreur commise en acceptant une hypothèse fausse est l'erreur de seconde espèce. Si H_1 est simple, la probabilité de rejeter à tort H_1 vaut

$$\beta = \mathbb{P}_{H_1}(Z \in A)$$

En résumé, nous avons le tableau suivant :

		Réalité	
		H_0	H_1
Décision	H_0	OK	risque de 2ème espèce
	H_1	risque de 1ère espèce	OK

Pratiquement, on se donne une limite supérieure du risque de première espèce, le plus souvent 5% (significatif), 1% (très significatif) ou l pour mille (hautement significatif). Cette limite constitue aussi le niveau de signification du test et permet de définir la condition de rejet de l'hypothèse nulle. Si l'hypothèse est composite, on choisit le domaine de rejet de H_0 de manière à ce que $\alpha = \max\{P_{H_0}(Z \in R)\}$. On rejette alors l'hypothèse nulle au niveau de signification nominal choisi (par exemple 0.05) si (et seulement si) le niveau de signification réel est inférieur ou égal au niveau de signification nominal ($p = 0.003 < 0.05$, rejet d'H_0). Cette attitude est bien sûr conservatrice.
Le risque de première espèce étant donné, on peut s'efforcer de calculer le risque de deuxième espèce, grâce à la notion de puissance de test ($P = 1 - \beta$). La puissance d'un test mesure dans un certain sens la capacité du test à différencier la valeur d'échantillonnage de celle de la population. Plus l'échantillon est grand, plus les estimateurs des paramètres de la population à étudier sont précis et plus le test d'hypothèses fondé sur ces estimateurs devient discriminatoire. En effet, plus l'échantillon est grand, plus il devient improbable qu'une différence observée entre l'estimateur et la valeur hypothétique soit uniquement attribuable au hasard de l'échantillonnage. On peut, au contraire penser à juste raison qu'il existe une différence réelle et donc rejeter l'hypothèse de départ. La performance d'un test est meilleure si la taille de l'échantillon est grande.
Un bon test possède des risques de première et de deuxième espèce faibles. Attention ces risques ne sont pas indépendants. C'est le principe du vendeur d'oranges : imaginons un vendeur qui vend des oranges avec pépins (bon marché) et des oranges sans pépins (plus cher). Son risque est de vendre peu cher des oranges sans pépins ; alors que le risque du consommateur est d'acheter cher des oranges avec pépins.

3 Stratégie d'un test d'hypothèse

Construire un test de l'hypothèse nulle H_0 contre l'hypothèse alternative H_1, c'est établir un critère de décision permettant de choisir entre l'hypothèse H_0 et H_1. Pour cela, il faut :

- Préciser les deux hypothèses H_0 et H_1. Quelle est l'hypothèse nulle ? Quelle est l'hypothèse alternative ?
- Établir un protocole expérimental et lui associer une variable aléatoire Z dont la loi évolue entre H_0 et H_1 mais dont la valeur observée z_0 sera parfaitement connue après réalisation de l'expérience.
- Déterminer la loi de Z sous H_0 et la loi de Z sous H_1. En déduire le type de test (unilatéral, bilatéral) c'est-à-dire la position du domaine de rejet R et du domaine d'acceptation A de H_0.
- Choisir le niveau α du test. En déduire précisément les domaines d'acceptation et de rejet de H_0.

Tout cela étant fait, la décision ne dépend plus que de la valeur observée z_0 de Z.

4 Tests non paramétriques

Un test d'hypothèse est dit paramétrique lorsque :
- la question que l'on se pose au départ concerne une variable aléatoire X (éventuellement l'indicatrice d'un événement).
- l'hypothèse H_0 est : $X \sim \mathcal{L}_0(\theta_0 \in I)$ où $\mathcal{L}_0$ est une loi de type donné et θ_0 un paramètre dont la valeur appartient à I.
- l'hypothèse H_1 est : $X \sim \mathcal{L}_1(\theta_1 \in J)$ où $\mathcal{L}_1$ est une loi de type donné et θ_1 un paramètre dont la valeur appartient à J.

Cette définition assez générale signifie que l'on connaît les lois de la population à un paramètre près (que l'on veut tester). Le cas gaussien qui est central sera traité au chapitre 9.
Dans tous les autres cas (l'une au moins des trois conditions n'est pas réalisée), le test est dit non-paramétrique. Un test paramétrique requiert un modèle à fortes contraintes (par exemple normalité des distributions, égalité des variances) pour lequel les mesures doivent avoir été réalisées dans une échelle au moins d'intervalle. Ces hypothèses sont d'autant plus difficiles à vérifier que les effectifs étudiés sont plus réduits.
Un test non paramétrique est un test dont le modèle ne précise pas les conditions que doivent remplir les paramètres de la population dont a été extrait l'échantillon. Cependant certaines conditions d'application doivent être vérifiées. Les échantillons considérés doivent être aléatoires (lorsque tous les individus ont la même probabilité de faire partie de l'échantillon) et simples (tous les individus qui doivent former l'échantillon sont prélevés indépendamment les uns des autres), et éventuellement indépendants les uns des autres. Les variables aléatoires prises en considération sont généralement supposées continues.

Avantages des tests non paramétriques

Leur emploi se justifie lorsque les conditions d'applications des autres méthodes ne sont pas satisfaites, même après d'éventuelles transformation de variables.
Les probabilités des résultats de la plupart des tests non paramétriques sont des probabilités exactes quelle que soit la forme de la distribution de la population dont est tiré l'échantillon.

Pour des échantillons de taille très faible jusqu'à $N = 6$, la seule possibilité est l'utilisation d'un test non paramétrique, sauf si la nature exacte de la distribution de la population est précisément connue. Ceci permet une diminution du coût ou du temps nécessaire à la collecte des informations.
Il existe des tests non paramétriques permettant de traiter des échantillons composés à partir d'observations provenant de populations différentes. De telles données ne peuvent être traitées par les tests paramétriques sans faire des hypothèses irréalistes.
Seuls des tests non paramétriques existent qui permettent le traitement de données qualitatives : soit exprimées en rangs ou en plus ou moins (échelle ordinale), soit nominales.
Les tests non paramétriques sont plus facile à apprendre et à appliquer que les tests paramétriques. Leur relative simplicité résulte souvent du remplacement des valeurs observées soit par des variables alternatives, indiquant l'appartenance à l'une ou à l'autre classe d'observation, soit par les rangs, c'est-à-dire les numéros d'ordre des valeurs observées rangées par ordre croissant. C'est ainsi que la médiane est généralement préférée à la moyenne, comme paramètre de position.

Désavantages des tests non paramétriques

Les tests paramétriques, quand leurs conditions sont remplies, sont plus puissants que les tests non paramétriques.
Un second inconvénient réside dans la difficulté a trouver la description des tests et de leurs tables de valeurs significatives, surtout en langue française. Heureusement, les niveaux de significativité sont donnés directement par les logiciels statistiques courants.
On choisira les tests appropriés en fonction du type de mesure, de la forme de la distribution de fréquences et du nombre d'observations dont on dispose.

5 Quelques applications pratiques des méthodes de statistique non paramétrique

5.1 *Cas d'un échantillon isolé*

Des tests permettent de vérifier si un échantillon observé peut être considéré comme extrait d'une population donnée (Test d'ajustement). Ces tests peuvent permettre de répondre aux questions suivantes :
- Y a t-il une différence significative de localisation (tendance centrale) entre l'échantillon et la population ?
- Y a t-il une différence significative entre les fréquences observées et les fréquences attendues sur la base d'un principe ?
- Y a t-il une différence significative entre des proportions observées et des proportions espérées ?
- Est-il raisonnable de penser que cet échantillon a été tiré d'une population d'une forme particulière ?
- Est-il raisonnable de penser que cet échantillon est un échantillon d'une certaine population connue ?

5.2 Un test d'ajustement : le test binomial

Il y a des populations où seulement deux classes sont distinguées : mâle et femelle ; lettré et illettré... Dans un tel cas, toutes les observations de cette population tomberont dans l'une ou l'autre classe. Pour toutes ces populations, si nous connaissons la proportion des cas d'une classe (P), nous connaissons celle de l'autre classe ($1 - P = Q$). Ces proportions sont fixées pour une population donnée. Cependant, ces proportions exactes ne se retrouverons pas dans un échantillon prélevé au hasard dans cette population. De telles différences entre les valeurs observées et celles de la population sont dues au processus d'échantillonnage. Bien entendu, de faibles différences sont plus probables que de fortes différences.
Le test binomial nous permet de dire si il est raisonnable de penser que les proportions (ou fréquences) observées dans notre échantillon proviennent d'une population ayant une valeur donnée de P.

Méthode

La loi binomiale ne dépend que d'un paramètre, la probabilité p de "l'événement favorable". La probabilité d'obtenir x objets dans une catégorie et $N - x$ dans l'autre est donnée par la formule :

$$\mathbb{P}(X = x) = \frac{N!}{x!(N-x)!} P^x Q^{N-x}$$

où N est le nombre d'observations ; P la proportion de cas attendus dans une catégorie et $Q = 1 - P$ la proportion de cas attendus dans l'autre catégorie.
Nous pouvons alors répondre à la question suivante : quelle est la probabilité exacte d'obtenir les valeurs observées. Mais le plus souvent nous posons la question : Quelle est la probabilité d'obtenir les valeurs observées ou des valeurs encore plus extrêmes ? La distribution d'échantillonnage est alors

$$\mathbb{P}(X \leq x) = \sum_{j=0}^{x} \frac{N!}{j!(N-j)!} P^j Q^{N-j} \tag{5.1}$$

Exemple 5.2 *Quelle est la probabilité d'obtenir deux six ou moins de deux six après cinq jets d'un dé non pipé ?*
On a $N = 5$ *(le nombre de jets)* $x = 2$ *(le nombre de six)* $P = 1/6$ *(proportion de six attendue)* $Q = 5/6$
Soit $p(i)$ *la probabilité d'obtenir* i *six ; On a*
- $\mathbb{P}(X = 0) = \frac{5!}{0!5!}(1/6)^0(5/6)^5 = 1 \times 1 \times 0.40 = 0.40$
- $\mathbb{P}(X = 1) = \frac{5!}{1!5!}(1/6)^1(5/6)^4 = 5 \times 0.1666 \times 0.4822 = 0.40$
- $\mathbb{P}(X = 2) = \frac{5!}{2!5!}(1/6)^2(5/6)^3 = 10 \times 0.0277 \times 0.578 = 0.16$

Et donc :

$$\mathbb{P}(X \leq 2) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = 0.40 + 0.40 + 0.16 = 0.96$$

La probabilité d'obtenir sous H_0 *deux six ou moins lorsqu'un dé non pipé est lancé cinq fois est* $p = 0.96$.

Petits échantillons

Dans le cas d'un échantillon à deux classes, une situation commune est celle où $P = {}^{1}\!/_{2}$. Lorsque l'effectif est inférieur à 10, la table 11.3 donne les probabilités associées à diverses valeurs de x (la plus petite des fréquences observées) pour différents effectifs N.

Exemple 5.3 *Dans une étude des effets du stress, on enseigne à 18 étudiants deux méthodes différentes de faire un noeud. La moitié des sujets (choisie au hasard dans le groupe de 18) apprend d'abord la méthode A, puis la méthode B. L'autre moitié apprend en premier la méthode B, puis la méthode A. Après avoir subi un examen de quatre heures, on demande à chaque sujet de faire le noeud. L'hypothèse est que le stress induit une régression, c'est-à-dire, que les sujets utiliseront la première méthode apprise. Chaque sujet est catégorisé suivant qu'il utilise la première méthode apprise ou la seconde après le stress.*
Hypothèse nulle $H_0\ : p_1 = p_2 = {}^{1}\!/_{2}$
Il n'y a pas de différence entre la probabilité d'utiliser la première méthode apprise (p_1*) et celle d'utiliser la seconde méthode apprise (*p_2*), après le stress.*
$H_1\ : p_1 > p_2$ *le test est donc unilatéral*
Le test binomial est choisi car les données rentrent dans deux catégories discrètes et l'échantillon est unique. L'apprentissage en premier ou second des deux méthodes A et B étant reparti au hasard, il n'y a pas de raison de penser que la première méthode apprise soit préférée à la seconde, compte tenu de H_0*, et de* $P = Q = {}^{1}\!/_{2}$*.*
On choisit le coefficient de sécurité $\alpha = 0.01$
La région de rejet comprend toutes les valeurs de x *(nombre de sujets qui ont utilisé, après le stress, la seconde méthode apprise) qui sont si faibles que leur probabilité associée sous* H_0 *est égale ou inférieure à* $\alpha = 0.01$*. Les petites valeurs favorisent l'hypothèse* H_1*. le test est donc unilatéral avec region de rejet à gauche.*
Dans cette expérience les résultats obtenus après le stress sont les suivants :

Méthode choisie		
Première apprise	*Deuxième apprise*	*Total*
16	*2*	*18*

La probabilité associée avec $x \leq 2$ *est* 0.0011*. Attendu que cette probabilité est inférieure à* $\alpha = 0.01$*, nous pouvons rejeter* H_0 *en faveur de* H_1*. Nous concluons que* $p_1 > p_2$*, c'est-à-dire, les personnes soumises à un stress utilisent la première des deux méthodes apprises.*
Lorsque N *s'accroît, la distribution binomiale tend vers la distribution normale. Cette tendance rapide lorsque* P *est proche de* ${}^{1}\!/_{2}$ *, se ralentit lorsque* P *est voisin de 0 ou de 1. Donc, plus la disparité entre* P *et* Q *est importante, plus l'échantillon devra être important pour que l'approximation soit utile.*

5.3 *Test des signes*

Soit une variable aléatoire T, de distribution continue. On veut tester H_0 : "la médiane de T vaut a" contre H_1 : "la médiane de T est différente de a" (supérieure ou inférieure)..
Si on considère la variable $T - a$, les hypothèses deviennent :

$$H_0 \ : \mathbb{P}(T - a \geq 0) = p = \frac{1}{2} \ , \ H_1 \ : \mathbb{P}(T - a \geq 0) = p \neq \frac{1}{2}$$

Le test du signe est aussi utilisé dans le cas des observations (X, Y) appariées. On considère la variable $T = X - Y$ et on teste H_0 : "la médiane de T vaut 0" contre H_1 : "la médiane de T est différente de 0" (positive ou négative).
L'hypothèse nulle peut s'écrire

$$H_0 \ : \mathbb{P}(T \geq 0) = \mathbb{P}(T \leq 0) = 1/2$$

Lorsque l'hypothèse nulle est vraie et pour N paires d'observations, le nombre de différences positives (ou négatives) est une variable binomiale de paramètres $P = Q = {}^1\!/_2$ et N. Le test permet de comparer, grâce à cette distribution, le nombre observé de signes plus (ou moins) et le nombre attendu ${}^N\!/_2$. Quand certaines différences sont nulles, les paires d'observations correspondantes sont écartées de l'analyse et la valeur de N est réduite en conséquence.
Lorsque $N < 100$, la table 11.2 donne en fonction de N et du niveau α, le plus grand entier k, tel que $\mathbb{P}(T \leq k) \leq \alpha$ pour un test unilatéral ou tel que $2\mathbb{P}(T \leq k) \leq \alpha$ pour un test bilatéral.
Le test des signes peut être unilatéral lorsque l'on prédit quel signe $+$ ou $-$ sera le plus fréquent ou bilatéral lorsque les fréquences des deux signes seront simplement différentes.

Exemple 5.4

1. *Vingt paires sont observées ; 16 présentent une différence positive et les 4 autres une différences négative. Donc $N = 20$ et $k = 16$.*

   Si H_1 prédit que les signes $+$ sont les plus fréquents, on observera moins de valeurs négatives sous H_1 : le test est unilatéral et la région de rejet de H_0 est à droite. La Table 11.2 révèle que le domaine de rejet de H_0, au seuil 1% est $k = \{15, 16, 17, 18, 19, 20\}$; on peut aussi dire que le domaine d'acceptation est $k = \{0, 1, \ldots, 13, 14\}$.

   Si H_1 prédit simplement que la différence entre les fréquences des deux signes est différente (test bilatéral), la Table 11.2 révèle que le domaine de rejet de H_0, au seuil 1% est $k = \{0, 1, 2, 3, 17, 18, 19, 20\}$

2. *Douze arbres sont mesurés alors qu'ils sont debout, par une mesure trigonométrique. Puis les mêmes arbres sont mesurés au sol, après abattage. La première méthode donne-t-elle des résultats significativement trop faibles ou trop élevés ?*

   H_0 : *Il n'y a pas de différences entre les mesures obtenues par la première et la seconde méthode ;*

H_1 **:** *il y a une différence significative, au seuil* $\alpha = 0.05$.

Les hauteurs obtenues (en mètres) sont les suivantes :

Arbres plantés	*Arbres abattus*	*Différences*
20.4	*21.7*	*-1.3*
25.4	*26.3*	*-0.9*
25.6	*26.8*	*-1.2*
25.6	*28.1*	*-2.5*
26.6	*26.2*	*0.4*
28.6	*27.3*	*1.3*
28.7	*29.5*	*-0.8*
29.0	*32.0*	*-3.0*
29.8	*30.9*	*-1.1*
30.5	*32.3*	*-1.8*
30.9	*32.3*	*-1.4*
31.1	*31.7*	*-0.6*

$N = 12$ *; nombre de différences positives :* $k = 2$.
Le but étant de comparer les deux méthodes de mesure, il y a autant de chances d'avoir de grandes valeurs que de petites valeurs sous H_1. *Le test est donc bilatéral. La table 11.2 révèle que l'hypothèse d'identité des résultats obtenus par les deux méthodes de mesure doit être rejetée au seuil de signification 0.05.*

Lorsque $N > 90$, on peut utiliser l'approximation normale de la loi binomiale en faisant intervenir une correction de continuité. Il suffit de calculer la valeur

$$z = \frac{(x \pm 0.5) - {}^1\!/\!_2 N}{\frac{1}{2}\sqrt{N}} \tag{5.2}$$

où $x + 0.5$ est utilisé lorsque $x < {}^1\!/\!_2 N$ et $x - 0.5$ lorsque $x > {}^1\!/\!_2 N$. La signification d'un tel z peut être déterminée par référence à la table de la loi normale.

Exemple 5.5

1. *Si l'on reprend l'exemple de comparaison des mesures des arbres, l'approximation normale donnerait :*
$$z = \frac{(2 + 0.5) - 6}{0.5\sqrt{12}} = \frac{3.5}{1.7320508} = 2.02$$
La table 1 révèle que pour z = 2.02, la probabilité bilatérale associée est $(1 - 0.97831) \times 2 = 0.04438$. *Cette valeur conduirait à rejeter l'hypothèse nulle au seuil 0.05. Bien que les échantillons ne contiennent chacun que douze individus, l'approximation est déjà très satisfaisante car la valeur exacte est* $p = 0.038$.
2. *Supposons qu'un chercheur veuille déterminer si la vision d'un film sur la délinquance juvénile change les opinions des membres d'une communauté sur la sévérité des sanctions à donner à des délinquants juvéniles. Il extrait un échantillon aléatoire de 100*

adultes de la communauté. Chaque sujet sera son propre contrôle. Il leur demande de prendre position sur la sévérité plus ou moins grande des punitions à infliger aux délinquants juvéniles. Il leur présente ensuite le film et réitère sa question après.

H_0 **:** *le film n'a pas d'effet sur l'opinion des sujets ;*

H_1 **:** *le film a un effet systématique.*

Le test des signes est choisi pour cette étude portant sur deux groupes appariés et dont les mesures sont réalisées dans l'échelle ordinale. Les différences pourront être représentées par des plus ou des moins.
Soit $\alpha = 0.01$; $N =$ le nombre de sujets qui changent d'opinion, quel qu'en soit le sens.
$N > 90$ donc z est calculé avec la formule 5.2 et les tables de la loi normale donnent la probabilité associée aux valeurs aussi extrêmes que le z obtenu.
Comme H_1 ne prédit pas la direction des différences, la région de rejet est bilatérale.
Les résultats de cette étude fictive sont les suivants :

		Opinion avant le film	
		Moins	*Plus*
Opinion après le film	*Plus*	*59*	*7*
	Moins	*8*	*26*

Ces données montrent que 15 adultes ($8 + 7$) n'ont pas été affectés par la vision du film et 85 l'ont été. Si le film n'a pas d'effet systématique, nous nous attendrions à ce que à peu près la moitié de ceux qui ont modifié leur jugement entre avant et après a changé de plus à moins et à peu près la moitié a changé de moins à plus. Soit 42.5 sujets auraient modifié leur jugement dans un sens ou dans l'autre.

$$x = 26\text{ ; } N = 85 \text{ donc } x < {}^1\!/\!_2 N$$

$$z = \frac{(26 + 0.5) - 42.5}{0.5\sqrt{85}} = \frac{16}{4.609772} = 3.47$$

Au seuil 1%, nous pouvons rejeter l'hypothèse nulle. Nous pouvons conclure, dans cette étude fictive, que la vision du film a eut un effet significatif sur l'opinion des adultes concernant la sévérité des peines à infliger aux délinquants juvéniles.

5.4 Test des rangs appliqué au cas d'échantillons appariés (Wilcoxon)

Le test précédent n'utilise que l'information sur la direction des différences entre paires. Si nous pouvons prendre en compte en plus la grandeur des différences, un test plus puissant peut être utilisé. Le test de Wilcoxon donne plus de poids à une paire qui montre une large différence entre les deux conditions qu'à une paire ayant une faible différence. Cela implique que l'on puisse dire quel membre d'une paire est plus grand que l'autre (donner le signe de la différence), mais aussi que l'on puisse ranger les différences en ordre croissant.

Méthode

- d_i = différence entre chaque paire, représentant la différence entre les scores appariés obtenus lors des deux traitements. Chaque paire a un d_i.
- Ranger tous les d_i sans tenir compte de leur signe. Dans ce cas, lorsque l'on range les d_i, un d_i de -1 est affecté d'un rang inférieur à celui d'un d_i de -2 ou $+2$. Puis réaffecter à chaque rang le signe de la différence.
- Si les traitements A et B sont équivalents, donc si H_0 est vraie, la somme des rangs ayant un signe positif et celle des rangs ayant un signe négatif devraient être à peu près égales. Mais si la somme des rangs de signes positifs est très différente de celle des rangs de signes négatifs, nous en déduirons que le traitement A diffère du traitement B, et rejetterons l'hypothèse nulle. Donc, il y a rejet d'H_0 que la somme des rangs de signe négatif ou que celle des rangs de signe positif soit faible.

Il est possible que les deux scores d'une quelconque paire soient égaux. Il n'y a pas de différence observée entre les deux traitements pour cette paire ($d = 0$). De telles paires sont abandonnées. N est alors égal au nombre de paires dont la différence entre les traitements n'est pas nulle. Mais deux ou plus des différences observées entre paires peuvent être égales entre elles. On donne alors le même rang à ces valeurs liées. Le rang affecté est la moyenne des rangs qu'auraient eu les diverses valeurs si elles avaient différé. Ainsi, trois des paires observées présentent les différences suivantes : -1, -1 et $+1$. Chaque paire aura le rang 2, car $\frac{1+2+3}{3} = 2$. La différence suivante aura alors le rang 4, puisque les rangs 1, 2, et 3 ont déjà été utilisés.

Petits échantillons

T = la somme des rangs du signe observé le moins fréquent. La table 11.16 donne les valeurs critiques de T et leurs niveaux de signification associés pour $N < 25$. Si le T observé est égal ou inférieur à la valeur donnée dans la table pour un niveau de signification et pour le nombre de différences non nulles N, l'hypothèse nulle peut être rejetée à ce niveau de signification.

Exemple 5.6 *Un psychologue de l'enfance veut tester l'effet de l'assistance à l'école maternelle sur la compréhension sociale des enfants. Il estime cette compréhension à partir des réponses que les enfants donnent à une série de questions portant sur des images représentant diverses situations sociales. Chaque enfant obtient ainsi un score compris entre 0 et 100. Le psychologue ne peut pas affirmer que les différences observées entre scores sont numériquement exactes (il ne peut pas dire qu'un score de 60 est le double d'un score de 30, ni que la différence entre 60 et 40 est exactement le double de la différence entre 40 et 30). Cependant, il pense que les scores sont suffisamment précis pour qu'il puisse les ranger selon leur valeur absolue. Pour tester l'effet de l'assistance à l'école maternelle sur la compréhension sociale des enfants, il utilise 8 paires de jumeaux. L'un des jumeaux est envoyé à l'école, alors que l'autre reste à la maison pendant un trimestre. L'affectation se faisant au hasard. A la fin du trimestre, il estime la compréhension sociale de chacun des enfants.*
L'hypothèse nulle : il n'y pas de différence entre la compréhension sociale des enfants restés à la maison et celle des enfants ayant suivi l'école.

Les résultats sont donnés dans le tableau ci-dessous.

Paires	*Score enfants scolarisés*	*Score enfants Non scolarisé*	d	*Rang de* d	*Rang avec le signe le − fréquent*
a	*82*	*63*	*19*	*7*	
b	*69*	*42*	*27*	*8*	
c	*73*	*74*	-1	-1	*1*
d	*43*	*37*	*6*	*4*	
e	*58*	*51*	*7*	*5*	
f	*56*	*43*	*13*	*6*	
g	*76*	*80*	-4	-3	*3*
h	*65*	*62*	*3*	*2*	
					$T = 4$

La table 11.16 montre que pour $N = 8$, un $T = 4$ nous permet de rejeter l'hypothèse nulle au seuil 0.05 pour un test bilatéral. Par conséquent, nous conclurions, dans cette étude fictive, que l'expérience de l'école affecte la compréhension sociale des enfants.
Ces données sont aussi traitables par le test des signes. Dans ce cas, $x = 2$ et $N = 8$, la table 11.2 montre que nous ne pourrions pas rejeter H_0 au seuil 0.05.

Grands échantillons

Lorsque N est supérieur à 25, il peut être démontré que la somme des rangs T est pratiquement normale et que l'on peut calculer :

$$z = \frac{T - {}^{\mathrm{N\,(N+1)}}/_{4}}{\sqrt{{}^{\mathrm{N(N+1)(2N+1)}}/_{24}}} \tag{5.3}$$

et se référer à aux tables de la loi normale
Pour montrer la précision de l'approximation, nous pouvons traiter les données précédentes $N = 8$, $T = 4$, dans ce cas $z = -1.96$
Pour $z = -1.96$, $p = 0.05$, c'est-à-dire la même probabilité qu'en utilisant la table des valeurs critiques de T.

5.5 *Test des rangs appliqué au cas des échantillons indépendants (Mann-Whitney)*

On considère un premier ensemble $(X_1, X_2, \dots, X_n)$ de n variables aléatoires indépendantes et de même loi $\mathcal{L}_X$ (on dit qu'il s'agit d'un n-échantillon). On considère un deuxième ensemble $(Y_1, Y_2, \dots, Y_p)$, indépendant du précédent, de p variables aléatoires indépendantes et de même loi $\mathcal{L}_Y$.
On choisit les notations pour que $n \leq p$.
On veut tester :

H_0 **:** les variables X et Y ont même loi ($\mathcal{L}_X = \mathcal{L}_Y$)

contre

H_1 **:** la loi $\mathcal{L}_X$ est plus étalée à droite ou à gauche ou des deux côtés par rapport à la loi de $\mathcal{L}_Y$.

On envisage le classement des $n+p$ observations par valeurs croissantes : si l'on dispose d'un groupe expérimental de 3 cas (p) et d'un groupe contrôle de 4 cas (n) et que les observations sont les suivantes :

Observations E	9	11	15	
Observations C	6	8	10	13

Nous rangeons les observations en ordre croissant en conservant l'identité de chacune d'entre-elles.

6	8	9	10	11	13	15
C	C	E	C	E	C	E

Maintenant considérons le groupe E et calculons le nombre d'observations C qui précèdent chacune de celle du groupe E. Pour l'observation 9 de E, deux observations de C précèdent ; pour l'observation 11 de E, trois C précèdent ; pour l'observation 15 de E, quatre C précèdent. Donc $U = 2 + 3 + 4 = 9$.
Le principe du test consiste à rejeter l'hypothèse d'identité des deux distributions lorsque la valeur observée U, s'écarte trop de la valeur attendue correspondante. Pour des échantillons très petits ($3 < n < 8$), on dispose de tables qui donnent la probabilité exacte d'obtenir tout U aussi extrême que celui qui est observé . Il suffit alors de connaître p (la taille du plus petit échantillon), n et U et de se reporter à la table 11.15 pour la valeur de l'échantillon n. Les probabilités données dans ces tables sont unilatérales. Pour un test bilatéral, il faut doubler la valeur de la table.
Dans notre exemple : $p = 3, n = 4, U = 9$, nous consultons la table 11.15 pour $n = 4$, mais la valeur observée de U n'apparaît pas dans la table. Par contre, si nous avions calculé le nombre d'observations E qui précèdent celle du groupe C, le U obtenu serait égal à 0 + 0 + 1 + 2 = 3. Cette valeur se trouve dans la table. Il est toujours possible de rechercher le plus petit U observé par la formule

$$U = np - U' \tag{5.4}$$

La probabilité unilatérale d'obtenir un $U \leq 3$ est $p = 0.200$. Lorsque la taille de n et p augmentent la méthode de comptage décrite devient rapidement inutilisable et une méthode alternative rend ce calcul plus aisé.

$$U = np + \frac{n(n+1)}{2} - R_1 \tag{5.5}$$

ou de façon équivalente

$$U = np + \frac{p(p+1)}{2} - R_2 \tag{5.6}$$

où R_1 = somme des rangs assignés à l'échantillon le plus petit (n) et R_2 = somme des rangs assignés à l'autre échantillon.

Exemple 5.7 *Cinq rats sont entraînés à imiter un rat leader dans un labyrinthe en T, pour atteindre une source de nourriture. Puis ces rats sont ensuite transférés dans une situation où par imitation d'un rat leader, ils apprennent à éviter un choc électrique. Leur comportement dans cette situation est comparé à celui de rats n'ayant pas été entraînés à suivre un leader. La comparaison se fait en terme de nombre d'essais nécessaire à chaque rat pour obtenir 10 réponses d'évitement lors de 10 essais. On fait l'hypothèse que les 5 rats préalablement conditionnés à imiter un congénère réussiront plus rapidement que les autres à éviter les chocs.*
Soit $\alpha = 0.05$ *;* $n = 4$ *rats témoins et* $p = 5$ *rats expérimentaux. Les résultats sont les suivants :*

Exp	*Rang*	*Témoins*	*Rang*
78	*7*	*110*	*9*
64	*4*	*70*	*5*
75	*6*	*53*	*3*
45	*1*	*51*	*2*
82	*8*		
	$R_2 = 26$		$R_1 = 19$

donc en appliquant la formule 5.6, nous avons $U = 4 \times 5 + \frac{5(5+1)}{2} - 26 = 9$
La probabilité d'obtenir un $U \leq 9$ *dans ces conditions est* $p = 0.452$ *(Table 11.15,* $p = 5$*).*
Les données ne supportent pas l'hypothèse selon laquelle un entraînement à l'imitation préalable est généralisé à d'autres situations.

Échantillons dont n2 est compris entre 9 et 20

La table 11.15 n'est plus utilisable lorsque p devient supérieur à 8. Mais on peut alors faire usage de la table 11.13pour les échantillons dont la taille p est comprise entre 9 et 20 et $n \geq 20$. Cette table 11.13 donne les valeurs critiques de U à différents niveaux de signification. Ainsi, lorsque la valeur du U observé est inférieur ou égale à celle de la table, H_0 peut être rejetée au niveau de signification correspondant.

Grands échantillons ($p > 20$)

Quand la taille de n et de p augmente, la distribution de U s'approche de la distribution normale. L'approximation normale se calcule de la façon suivante :

$$z = \frac{U - \frac{np}{2}}{\sqrt{\frac{np(n+p-1)}{12}}} \tag{5.7}$$

Exemple 5.8 *Dans des sociétés humaines à culture non écrite, les ethnologues peuvent classer ces sociétés en fonction du degré d'anxiété présenté par les enfants à la suite de la socialisation (ce classement va de 6 à 17). Il est aussi possible de distinguer deux groupes suivant que ces sociétés disposent d'explications orales de la maladie ou non.*

Explic. absente	*Estim. anxiété*	*Rang*	*Explic. présente*	*Estim. anxiété*	*Rang*
Lapp	*13*	*29.5*	*Marquesan*	*17*	*39*
Chamorro	*12*	*24.5*	*Dobuan*	*16*	*38*
Samoan	*12*	*24.5*	*Baiga*	*15*	*36*
Arapesh	*10*	*16*	*Kwoma*	*15*	*36*
Balinese	*10*	*16*	*Thonga*	*15*	*36*
Hopi	*10*	*16*	*Alorese*	*14*	*33*
Tanala	*10*	*16*	*Chagga*	*14*	*33*
Paiute	*9*	*12*	*Navaho*	*14*	*33*
Chenchu	*8*	*9.5*	*Dahomean*	*13*	*29.5*
Teton	*8*	*9.5*	*Lesu*	*13*	*29.5*
Flathead	*7*	*5*	*Masai*	*13*	*29.5*
Papago	*7*	*5*	*Lepcha*	*12*	*24.5*
Venda	*7*	*5*	*Maori*	*12*	*23.5*
Warrau	*7*	*5*	*Pukapukan*	*12*	*24.5*
Wogeo	*7*	*5*	*Trobriander*	*12*	*24.5*
Ontong-Javanese	*6*	*1.5*	*Kwakiutl*	*11*	*20.5*
		$R_1 = 200$	*Manus*	*11*	*20.5*
			Chiricahua	*10*	*16*
			Comanche	*10*	*16*
			Siriono	*10*	*16*
			Bena	*8*	*9.5*
			Slave	*8*	*9.5*
			Kurtachi	*6*	*1.5*
					$R_2 = 580$

Noter que les valeurs d'anxiété qui sont ex-aequo, où qu'elles se présentent, sont affectées d'un rang égal à la moyenne des rangs revenant normalement à ces différentes valeurs.
Nous calculons la valeur de U *par la formule 5.5 :* $U = 16 \times 23 + \frac{16(16+1)}{2} - 200 = 304.$
Nous substituons la valeur de U *dans la formule 5.7 pour calculer*

$$z = \frac{304 - \frac{16\times23}{2}}{\sqrt{\frac{16\times23\times(16+23+1)}{12}}} = 3.43$$

La référence à la table de la loi normale révèle que z *égal ou supérieur à 3,.43 a une probabilité unilatérale de* $p < 0.0003$*. Comme ce* p *est inférieur à* $\alpha = 0.01$*, nous pouvons rejeter* H_0*. Nous concluons que les sociétés à explications orales de la maladie ont une socialisation de l'anxiété supérieure aux autres sociétés.*

Problème des ex-aequo

Lorsque deux ou plus d'observations du même groupe ont des valeurs égales, la valeur de U n'est pas affectée. Par contre, si des valeurs identiques se présentent dans les deux échantillons, la valeur du U est affectée. Bien que cet effet soit souvent négligeable, une correction pour les ex-aequo existe. La formule corrigée pour les ex-aequo est la suivante :

$$z = \frac{N - \frac{np}{2}}{\sqrt{\frac{np}{N(N-1)}\left(\frac{N^3-N}{12} - T\right)}} \tag{5.8}$$

où $N = n + p$ et $T = \frac{t^3-t}{12}$; t est le nombre d'ex-aequo pour un rang donné.
Pour les données précédentes $n + p = 16 + 23 = 39 = N$, nous observons les ex-aequo suivants :
2 pour 6 ; 5 pour 7 ; 4 pour 8 ; 7 pour 10 ; 2 pour 11 ; 6 pour 12 ; 4 pour 13 ; 3 pour 14 ; 3 pour 15.

$$T = 0.5 + 10.0 + 5.0 + 28.0 + 0.5 + 17.5 + 5.0 + 2.0 + 2.0 = 70.5$$

L'utilisation de ces valeurs dans la formule 5.8 donne :

$$z = \frac{304 - \frac{16\times 23}{2}}{\sqrt{\frac{16\times 23}{39(39-1)}\left(\frac{39^3-39}{12} - 70.5\right)}} = 3.45$$

Cet exemple confirme que les ex-aequo ont un effet négligeable sur la valeur du z. Aussi la correction peut n'être faite que lorsque le nombre d'ex-aequo est très important où lorsque la valeur du z obtenue sans correction est voisine de la signification au seuil choisi.

6 Exercices

1. Un directeur de laboratoire pharmaceutique refuse la mise en fabrication d'un nouveau vaccin proposé par un des chercheurs du laboratoire. Il invoque pour cela les résultats statistiques peu concluants obtenus suite aux tests : le vaccin proposé n'est pas significativement plus efficace que celui utilisé actuellement. Les frais supplémentaires entraînés pour le produire ne sont donc pas justifiés.
   (a) Soient H_0 et H_1 les hypothèses possible :
   H_0 : le vaccin proposé n'est pas plus efficace que celui déjà en production ;
   H_1 : le vaccin proposé est plus efficace que celui déjà en production.
   Quels sont les deux types d'erreurs que le directeur pourrait commettre relativement à ces deux hypothèses ?
   (b) En prenant la décision de ne pas mettre en fabrication le vaccin proposé, lequel des deux types d'erreur le directeur a-t-il tenté de contrôler ?
2. Soit une expérience $\mathcal{E}$ à laquelle est associée une variable aléatoire Z de fonction de répartition continue et strictement croissante. On réalise quatorze expériences indépendantes et identiques à $\mathcal{E}$ qui donnent les résultats suivant :

4.65	4.86	4.40	3.20	5.17	4.60	4.18
4.85	5.28	5.75	5.35	6.33	2.69	3.95

   Peut-on accepter l'hypothèse : "la médiane est égale à 5" ?
3. À une expérience $\mathcal{E}$ est associée une variable aléatoire X de fonction de répartition continue et strictement croissante. Construire, au niveau 5%, un test de l'hypothèse H_0 : "la médiane de X est nulle". contre l'hypothèse H_1 : "la médiane de X est négative, à partir de n expériences indépendantes et identiques à $\mathcal{E}$,
   (a) lorsque $n = 14$;
   (b) lorsque $n = 196$.
4. Soit T le résultat possible d'une expérience aléatoire $\mathcal{E}$. T est une variable aléatoire à densité.
   Soit $\{X_1, X_2, \ldots, X_n\}$ un ensemble de n variables aléatoires indépendantes et de même loi que T. Cet ensemble représente les résultats possibles d'une suite finie de n expériences indépendantes et identiques à $\mathcal{E}$.
   Les quatorze premières expériences ont donné les résultats suivants :

-0.35	-0.15	-0.14	+0.28	-0.60	+0.75	-1.80
+0.35	+0.17	+1.33	-0.40	-2.31	-0.82	-1.05

   Peut-on accepter l'hypothèse : "la densité de T est symétrique par rapport à zéro" ?
5. Pour une maladie donnée, les traitements habituels donnent au malade une chance sur deux de guérison. Un nouveau traitement est appliqué à 11 malades. Construire un test permettant de décider si le nouveau traitement est plus efficace que les anciens. À votre avis, quelle critique peut-on faire à ce test ?

6. On désigne par p la probabilité d'observer un phénotype donné sur un individu issu d'un certain croisement.

   Pour tester l'hypothèse $p = \frac{9}{16}$ contre l'hypothèse $p = \frac{9}{15}$, on procède ainsi : on observe 2400 individus issus du croisement en question ; si le nombre d'individus présentant le phénotype est inférieur ou égal à 1395, on choisit $9/16$, dans le cas contraire, on choisit $9/15$. Justifier le principe de ce test ; calculer son niveau et sa puissance.

7. On soupçonne que les conditions de travail, dans une entreprise donnée, entraînent une élévation progressive du taux de glucose dans le sang. Le tableau ci-dessous donne, pour un groupe de quatorze employés, les valeurs de ce taux (en g/l) observées à l'embauche (X) et après six mois de travail (Y).

Employé	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Taux X	0.88	0.85	0.94	0.96	1.19	0.96	0.92	0.90	0.76	0.94	1.05	1.01	0.96	0.94
Taux Y	1.10	0.84	1.02	1.06	1.07	1.16	0.88	1.18	0.89	1.10	0.96	1.16	1.19	0.89

   Que peut-on conclure ?

   On présentera deux interprétations différentes en précisant dans les deux cas :
   - les hypothèses testées et la variable choisie ;
   - la construction du test ;
   - le niveau exact du test.

8. On a mesuré sur *Dunaliella Marina*, la quantité d'azote protéique par cellule, à la même date et dans des conditions expérimentales identiques, sur une culture témoin et sur une culture préalablement irradiée. On pense que l'irradiation favorise un développement anormal des cellules. Interpréter les résultats ($y\,10^{-4}$)

   Culture témoin : 1.65 ; 2.00 ; 1.69 ; 2.20 ; 2.13 ; 1.66 ; 2.30 ; 1.87 ; 1.74 ; 1.97

   Culture irradiée : 2.29 ; 2.57 ; 2.66 ; 2.45 ; 2.97 ; 2.27 ; 1.76 ; 2.74 ; 2.36

9. Les apiculteurs du Texas s'inquiètent de la progression des abeilles africaines (*killer bees*), plus agressives mais moins productives que les abeilles domestiques. Les pouvoirs publics sont prêts à donner des fonds pour combattre ce phénomène si on peut démontrer que la proportion d'abeilles africaines a augmenté de manière significative ces dernières années.

   Les données sont récoltées à l'aide de pièges répartis sur le territoire texan. Des spécialistes identifient les abeilles capturées et les dénombrent, ce qui permet d'associer à chaque piège la proportion d'abeilles africaines que l'on a observé. Deux séries de données ont ainsi été obtenues ; l'une en 1980, l'autre en 1990.

Piège	1	2	3	4	5	6	7	8	9	10
% en 1980	0.330	0.146	0.518	0.339	0.693	0.249	0.438	0.695	0.135	0.388
% en 1990	0.360	0.177	0.524	0.447	0.140	0.392	0.534	0.263	0.157	0.566

   - En l'absence de toute information sur la manière dont on a réparti les pièges en 1980 et en 1990, que peut-on conclure, au niveau 5%, avec un test sur la somme des rangs ? Pensez-vous que ce test soit applicable ?
   - En fait, les pièges ont été localisés aux mêmes endroits en 1980 et 1990 ? Que peut-on conclure avec un test du signe, au niveau 5% puis au niveau 5.5% ?

10. On désire savoir si une certaine variété de blé a, dans une région A, un rendement supérieur à celui qu'elle a dans une région B. Pour cela, on dispose de résultats, exprimés en quintaux par hectare, obtenus sur seize parcelles différentes et ainsi répartis :

Région A	48.0	48.2	50.3	53.5	54.6	56.4	57.8	58.5	60.5
Région B	44.2	46.3	48.3	48.5	50.5	51.2	55.4		

La réponse s'appuiera sur un test construit au niveau 2.5%, en l'absence de toute hypothèse sur le type de loi suivie par le rendement d'une parcelle.

7 Solutions

1.
2. On fait un test du signe.
   On note $Z_1, \ldots, Z_{14}$ les résultats des 14 expériences. Les Z_i sont de même loi que Z.
   On teste :
   H_0 : la médiane de Z est 5 : $\mathbb{P}(Z \geq 5) = 1/2$
   contre H_1 : la médiane de Z est différente de 5 : $\mathbb{P}(Z \geq 5) \neq 1/2$
   On utilise la variable V : nombre de fois où $Z_i \geq 5$ parmi les 14 valeurs.
   sous H_0, $Z \sim \mathcal{B}(14, \frac{1}{2})$.
   Sous H_1, $Z \sim \mathcal{B}(14, p)$.
   Donc, sous H_1, V prend des valeurs plus grandes ou plus petites que sous H_0. Dans la table 11.2, au niveau 5%, on rejette H_0 si $V \leq 2$ ou $V \geq 12$, la region de rejet est donc $R_{H_0} = \{0, 1, 2\} \cup \{12, 13, 14\}$.
   Parmi les 14 valeurs observées ici, 5 sont au dessus de 5 : $v_0 = 5 \notin R_{H_0}$. Donc on ne rejette pas H_0 : la médiane de Z est 5.
3. On va traduire H_0 et H_1 sous les formes équivalentes :
   H_0 : $\mathbb{P}(X \geq 0) = \frac{1}{2}$
   H_1 : $\mathbb{P}(X \geq 0) < \frac{1}{2}$
   Aux n expériences, on associe la variable Z = le nombre de valeurs positives de X que l'on peut observer. Il est clair que :
   sous H_0, $Z \sim \mathcal{B}(n, \frac{1}{2})$.
   Sous H_1, $Z \sim \mathcal{B}(n, p < \frac{1}{2})$.
   On construira donc, à partir de Z un test unilatéral avec domaine de rejet de H_0 à gauche. Il s'agit d'un test du signe.
   - Si $n = 14$, au niveau $\alpha = 5\%$, on peut séparer l'ensemble des observables (table du signe) :
     Domaine de rejet de H_0 : $\{0, 1, 2, 3\}$, c'est-à-dire domaine d'acceptation de H_0 : $\{4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}$
     Donc si on observe $z \in \{0, 1, 2, 3\}$ on choisit H_0, et,
     si on observe $z \in \{4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}$ on choisit H_1
   - Si $n = 196$, on peut considérer que, sous H_0, $Z \sim \mathcal{N}(98, 49)$. (On note que $98 > 10$). Pour la variable Z^*, de loi $\mathcal{N}(0, 1)$, c'est la valeur -1.645 qui sépare le domaine de rejet du domaine d'acceptation de H_0.
     Pour la variable Z, ce sera la valeur $98 - 1.645 \times \sqrt{49} = 86.485$
     D'où un test, au niveau de 5% : si on observe une valeur $z \leq 86$, on rejette H_0 et si on observe $z \geq 87$, on conserve H_0.
     On peut vérifier que le niveau de ce test est très proche de 5% car il correspond, avec la correction de continuité à la valeur $z^* = \frac{86.5-98}{7} \approx -1.643$
4. On va tester l'hypothèse H_0 : "la loi de T est symétrique par rapport à zéro" contre l'hypothèse H_1 : la loi de T n'est pas symétrique par rapport à zéro".

Pour cela on classe les 14 valeurs X_i par valeur absolue croissante et on considère la variable aléatoire U_1 représentant la somme des rangs des valeurs positives que l'on est susceptible d'observer.

Sous H_0, la loi de U_1 est en cloche symétrique sur $\{0, 1, 2, \dots, 104, 105\}$.

Sous H_1, la loi de U_1 se déforme vers la droite ou la gauche de $\{0, 1, 2, \dots, 104, 105\}$.

On construit un test bilatéral.

An niveau 5%, le domaine d'acceptation de H_0 est l'ensemble des entiers : $\{22, 23, \dots, 82, 83\}$ (test signe et rang).

Les observations rangées par valeur absolue croissante :

Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14
valeurs	-0.14	-0.15	$+0.17$	$+0.28$	-0.35	$+0.35$	-0.40	-0.60	$+0.75$	-0.82	-1.05	$+1.33$	-1.80	-2.31

conduisent à la valeur $u_1 = 33.5$ (deux ex-aequo)

On accepte donc H_0 (cela ne change rien si on choisit $u_1 = 33$ ou $u_1 = 34$).

5. Soit p la probabilité de guérison d'un malade avec le nouveau traitement.

   On veut tester :

   H_0 **:** $p = \frac{1}{2}$ (pas de changement)

   contre

   H_1 **:** $p > \frac{1}{2}$ (amélioration)

   Pour cela on choisit la statistique $X =$ nombre de guérisons que l'on peut observer sur 11 malades traités.

   Sous H_0 : $X \sim \mathcal{B}(11, {}^1\!/_2)$ (loi en cloche symétrique. Sous H_1, le mode est déplacé à droite. On a donc un test unilatéral avec domaine de rejet de H_0 à droite.

   Au niveau 5% (voir table 11.2)

   - Domaine de rejet de H_0 : $\{9, 10, 11\}$
   - Domaine d'acceptation de H_0 : $\{0, 1, 2, \dots, 7, 8\}$

   Le calcul exact de α donne $\alpha = \frac{1+11+55}{2048} \approx 3.27\%$.

   Critique : compte tenu du faible nombre d'essais (11), la courbe de puissance du test, représentative de $\pi(p) = \mathbb{P}(X \geq 9)$, ne va pas croître très vite ($p > \frac{1}{2}$). Seules les améliorations ($p \approx 0.9$) seront détectées à coup sûr par u tel test.

6. Soit X la variable aléatoire qui représente le nombre de porteurs de phénotype que l'on peut observer sur 2400 individus.

   Sous H_0 **:** $X \sim \mathcal{B}\left(2400, \frac{9}{16}\right)$

   Sous H_1 **:** $X \sim \mathcal{B}\left(2400, \frac{9}{15}\right)$

   En fait, compte tenu de la taille de l'échantillon, on peut considérer que l'approximation de la loi binomiale par une loi normale est excellente ($2400 \times {}^9\!/_{16} > 1000$) donc :

   sous H_0 **:** $X \sim \mathcal{N}\left(1350, 24.30^2\right)$

   sous H_1 **:** $X \sim \mathcal{N}\left(1440, 24^2\right)$

Les deux distributions sont bien séparées.

Il est clair que l'on est amené à construire un test unilatéral avec domaine de rejet de H_0 à droite. La valeur 1395, située entre les deux modes peut parfaitement être choisie pour séparer le domaine d'acceptation de H_0 du domaine de rejet de H_0.

Est-ce un bon test ? Calculons son niveau α et sa puissance π.

$$\begin{aligned}
\alpha = \mathbb{P}_{H_0}(X \geq 1395) &= \mathbb{P}\left(X^* \geq \frac{1395 - 1350}{24.3}\right) = \mathbb{P}(X^* \geq 1.852) \approx 3.2\% \\
\pi = \mathbb{P}_{H_1}(X \geq 1395) &= \mathbb{P}\left(X^* \geq \frac{1395 - 1440}{24}\right) = \mathbb{P}(X^* \geq -1.875) \approx 96.96\%
\end{aligned}$$

Il s'agit d'un très bon test dont les risques de première et de deuxième espèce sont pratiquement égaux et inférieurs à 5%

7.

(a) On va tester :

H_0 **:** la médiane de $Y - X$ vaut zéro $\Leftrightarrow \mathbb{P}(Y - X \geq 0) = \frac{1}{2}$

contre

H_1 **:** la médiane de $Y - X$ est positive $\Leftrightarrow \mathbb{P}(Y - X \geq 0) > \frac{1}{2}$

Pour cela, on utilise le test du signe.

Soit V la variable aléatoire qui représente le nombre d'écarts $(Y - X)$ positifs que l'on peut observer sur 14 individus.

$$V(\Omega) = \{0, 1, 2, 3, \ldots, 8, 9, 10, 11, 12, 13, 14\}$$

Sous H_0 **:** $V \sim \mathcal{B}\left(14, \frac{1}{2}\right)$, distribution en cloche symétrique.

Sous H_1 **:** $V \sim \mathcal{B}\left(14, p > \frac{1}{2}\right)$, distribution déplacée, par rapport à la précédente, vers la droite de $V(\Omega)$.

D'où un test unilatéral avec domaine de rejet de H_0 à droite.

Au niveau 5% (voir table 11.2), on a :

- Domaine de rejet de H_0 : $\{11, 12, 13, 14\}$
- Domaine d'acceptation de H_0 : $\{0, 1, 2, \ldots, 8, 9, 10\}$

On peut préciser le niveau réel de ce test.

$$\alpha = \mathbb{P}_{H_0}(V \in \{11, 12, 13, 14\}) = \left(\frac{1}{2}\right)^{14} \left(C_{14}^{11} + C_{14}^{12} + C_{14}^{13} + C_{14}^{14}\right) \approx 2.87\%$$

Notons que si l'on choisissait comme domaine de rejet $\{10, 11, 12, 13, 14\}$ le niveau de test deviendrait $\alpha \approx 8.98\%$ que l'on peut considérer comme un risque plus élevé.

Dans l'échantillon, on a observé 9 valeurs positives, soit $v_0 = 9$. L'hypothèse H_0 est donc choisie : on peut considérer que la médiane de $Y - X$ est nulle.

(b) On peut également construire un test de

H_0 **:** la loi de $Y - X$ est symétrique par rapport à zéro

contre

H_1 **:** la loi de $Y - X$ est plus étalée vers les valeurs positives.

Pour cela, on va classer les 14 observations de $Y - X$ par valeurs absolues croissantes et on utilise la variable T_+ = somme des rangs des observations positives et on construit un test de Wilcoxon (section 5.4).

$$T_+(\Omega) = \{0, 1, 2, 3, \ldots, 102, 103, 104, 105\}$$

Sous H_0, T_+ a une distribution en cloche et symétrique sur l'ensemble des observables.

Sous H_1, on ne sait rien sur la loi de T_+ mais il est clair que les observations positives ayant tendance à prendre des valeurs (absolues) élevées donc des rangs élevés, on observe une valeur de T_+ sur la droite de $T_+(\Omega)$.

On construit donc un test unilatéral avec domaine de rejet de H_0 à droite.

Au niveau 5% (voir table 11.14), on obtient :

- Domaine de rejet de H_0 : $\{79, 80, 81, \ldots, 104, 105\}$
- Domaine d'acceptation de H_0 : $\{0, 1, 2, \ldots, 76, 77, 78\}$

Effectuons le classement des 14 observations :

Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$Y - X$	-0.01	-0.04	-0.05	0.08	-0.09	$+0.10$	-0.12	$+0.13$	$+0.15$	0.16	$+0.20$	$+0.22$	$+.23$	$+0.28$
Patient	B	G	N	C	K	D	E	I	L	J	F	A	M	H

On a donc observé la valeur $T_+ = 87$.

L'hypothèse H_0 est donc rejetée. Le test signe et rang montre que, dans l'entreprise, le taux de glucose sanguin a tendance à augmenter.

8. Soit X le résultat d'un dosage effectué sur une culture témoin. Soit Y le résultat d'un dosage effectué sur une culture irradiée.

   On teste l'hypothèse H_0 : X et Y ont même loi. contre l'hypothèse H_1 : Y a tendance à prendre des valeurs plus élevées que celles de X.

   On va classer, par valeurs croissantes, l'ensemble des observations et on va choisir la variable W définie comme la somme des rangs des observations faites sur les cultures irradiées. $W(\Omega) = \{45, 46, \ldots, 134, 135\}$. Sous H_0, W a une distribution en cloche symétrique. Sous H_1, W prendra de préférence, des valeurs plus proches de 135. On construit donc un test unilatéral avec un domaine de rejet de H_0 à droite.

   Au niveau 5% (4.7%) (table X, test de la somme des rangs) :
   Domaine de rejet de H_0 : $\{111, 112, \ldots, 134, 135\}$
   Domaine d'acceptation de H_0 : $\{45, 46, , \ldots, 109, 110\}$

   On a observé $W_0 = 127$, on choisit donc H_1 : la quantité d'azote protéique par cellule a tendance à être plus élevée sur les cultures irradiées.

9. •

Piège	1	2	3	4	5	6	7	8	9	10
1980	0.330 (8)	0.146 (3)	0.518 (15)	0.339 (9)	0.693 (19)	0.249 (6)	0.438 (13)	0.695 (20)	0.135 (1)	0.388 (11)
1990	0.360 (10)	0.177 (5)	0.524 (16)	0.447 (14)	0.140 (2)	0.392 (12)	0.534 (17)	0.263 (7)	0.157 (4)	0.566 (18)

On considère un ensemble $(X_1, X_2, \ldots, X_{10})$ de variables aléatoires représentant les proportions d'abeilles africaines observables sur 10 pièges en 1980. Ces variables sont supposées indépendantes et de même loi $\mathcal{L}_X$. Un deuxième ensemble $(Y_1, Y_2, \ldots, Y_{10})$ représente la proportion d'abeilles africaines indépendantes et de même loi $\mathcal{L}_Y$. Les deux ensembles sont indépendants.
On veut tester H_0 : les variables X_i et Y_j ont même loi ($\mathcal{L}_Y = \mathcal{L}_X$) contre H_1 : les variables Y_j ont tendance à prendre des valeurs plus élevées que celles des X_i.
On choisit la variable W associant aux 20 résultats possibles, rangés par valeurs croissantes, la somme des rangs des résultats obtenus en 1990. $W(\Omega) = \{55, 56, \ldots, 153, 154, 155\}$.
Sous H_0, W a une distribution symétrique, en cloche. Sous H_1, W aura tendance à prendre des valeurs plus élevées, c'est-à-dire vers la droite de $W(\Omega)$. On construit donc un test unilatéral avec domaine de rejet de H_0 à droite. Au niveau 5%, on obtient (table X) :
Domaine de rejet de H_0 : $\{127, 128, \ldots, 154, 155\}$
Domaine d'acceptation de H_0 : $\{55, 56, , \ldots, 125, 126\}$
(somme des rangs entre parenthèses dans le tableau)
On a observé $W_0 = 105$, on choisit donc H_0 et on conclut que la proportion d'abeilles africaines n'est pas en augmentation.
La dispersion des résultats favorisent leur interclassement. Il n'est donc pas étonnant de trouver un résultat non significatif. On peut cependant faire une critique de fond : la variable proportion observable dans un piège a une loi qui dépend du nombre n d'insectes capturés dans le piège. Il est évident que n varie d'un piège à l'autre et donc que les variables X_i ne peuvent pas avoir la même loi $\mathcal{L}_X$. Il en est de même pour les Y_j.

•

Piège	1	2	3	4	5	6	7	8	9	10
% en 1980	0.330	0.146	0.518	0.339	0.693	0.249	0.438	0.695	0.135	0.388
% en 1990	0.360	0.177	0.524	0.447	0.140	0.392	0.534	0.263	0.157	0.566
Différence	+	+	+	+	-	+	+	-	+	+

Les observations sont maintenant appariées. On peut alors porter son attention, pour un lieu i, sur la variable $Z_i = Y_i - X_i$ qui représente la variation de la proportion des abeilles africaines. Est-elle stable ? Est-elle en augmentation ? Le test du signe va permettre de tester : H_0 : la médiane des Z_i est nulle $\Leftrightarrow p = \mathrm{P}(Z_i \geq 0) = 1/2$ contre H_1 : la médiane des Z_i est positive $\Leftrightarrow p = \mathrm{P}(Z_i \geq 0) > 1/2$
On choisit la variable V représentant le nombre d'écarts positifs que l'on peut observer sur 10 stations différents. Il est clair que $V(\Omega) = \{0, 1, 2, \ldots, 9, 10\}$ et que $V \sim \mathcal{B}(10, p)$.
Sous H_0 : $V \sim \mathcal{B}(10, 1/2)$ distribution en cloche symétrique.

Sous $H_1 : V \sim \mathcal{B}(10, p > 1/2)$ distribution en cloche ou en J mais dissymétrique avec le mode à droite.
On construit donc un test unilatéral avec domaine de rejet à droite. Sachant que, sous H_0 : P$(V = 10) = 1/1024$, P$(V = 9) = 10/1024$, P$(V = 8) = 45/1024$, on a
Au niveau 5% (en fait $11/1024 \approx 1.07$%) le domaine de rejet $\{9, 10\}$.
Au niveau 5.5% (en fait $56/1024 \approx 5.47$%) le domaine de rejet $\{8, 9, 10\}$.
Comme $z = 8$ et compte-tenu des niveaux réels, on peut conclure à la progression des abeilles africaines.

10. On peut tester l'hypothèse H_0 : "les variables X_i et Y_j ont même loi" contre l'hypothèse H_1 : "les variables X_i ont tendance à prendre des valeurs plus grandes que celles des Y_j".

    Soit W la variable aléatoire associant aux 16 résultats possibles, rangés par valeurs croissantes, la somme des rangs des 7 résultats de B. W est à valeurs entières sur {28, 29,...,90,91}. Sous H_0, W a une distribution en cloche, symétrique. Sous H_1, W aura tendance à prendre des valeurs situées à gauche de la distribution. D'où un test unilatéral qui admet, au niveau 2.5%, pour domaine de rejet H_0 : $\{28, 29, \ldots, 40, 41\}$

    les 16 observations sont rangés par valeurs croissantes :

    $y_1, y_2, x_1, x_2, y_3, y_4, x_3, y_5, y_6, x_4, x_5, y_7, x_6, x_7, x_8, x_9$
    On en déduit $W = 43 \Rightarrow$. Les rendements sont les mêmes.

Chapitre 6

Variables aléatoires à valeurs dans $\mathbb{R}^2$

1 Introduction

Définition 6.1 *Soit un espace de probabilité* $\{\Omega, \mathcal{A}, \mathbb{P}\}$ *associé à une expérience aléatoire* $\mathcal{E}$*. On appelle variable aléatoire à valeurs dans* $\mathbb{R}^2$ *une application, notée* (X, Y)*, de* Ω *dans* $\mathbb{R}^2$*, qui permet d'associer à tout événement élémentaire* ω *un élément* $(X(\omega), Y(\omega))$ *de* $\mathbb{R}^2$ *et qui possède la propriété :*

$$\textit{pour tout couple d'intervalles } I \textit{ et } J \textit{ de } \mathbb{R}, \{\omega \textit{ tel que } [X(\omega) \in I \cap Y(\omega) \in J]\} \in \mathcal{A} \tag{6.1}$$

Cette définition conduit à plusieurs commentaires :

1. une variable aléatoires à valeurs dans $\mathbb{R}^2$ permet de "traduire" toute observation par un doublet ordonné de réels. Si par exemple, à l'expérience qui consiste à choisir un individu dans une population, on associe la taille et le poids de cet individu, il est clair que le choix d'un individu donné implique l'observation de deux réels donnés.
2. La propriété 6.1 signifie que l'événement contenant tous les événement élémentaires qui ont une image dans le domaine $I \times J$ (I et J éventuellement réduits à un point) fait partie de la tribu $\mathcal{A}$ et, puisque l'espace $\{\Omega, \mathcal{A}\}$ est probabilisé, possède une probabilité. On notera désormais cet événement $(X \in I \cap Y \in J)$ ou $[(X, Y) \in I \times J]$ et on écrira :
$$\mathbb{P}(\{\omega | [X(\omega) \in I \cap Y(\omega) \in J]\}) = \mathbb{P}(X \in I \cap Y \in J)$$
La définition d'une variable aléatoire à valeurs dans $\mathbb{R}^2$, implique donc l'attribution d'une probabilité à tout "pavé" $I \times J$, en identifiant ce domaine à l'événement dont il est l'image.
3. Soit (X, Y) une variable aléatoire à valeurs dans $\mathbb{R}^2$. Il est clair que l'application de Ω dans $\mathbb{R}$ qui, à tout événement élémentaire ω, associe le premier réel $X(\omega)$ définit une variable aléatoire X puisque, pour tout intervalle I, on peut écrire :
$$\mathbb{P}(X \in I) = \mathbb{P}(X \in I \cap Y \in \mathbb{R}) \tag{6.2}$$

On définit, de même, la variable aléatoire Y. C'est la raison pour laquelle, le plus souvent, on dit que (X, Y) est un *couple de variables aléatoires à valeurs dans* $\mathbb{R}$. Certains auteurs qualifient ces variables de simultanées, ce qui signifie qu'une même expérience permet d'observer une valeur pour chacune des variables

4. On peut signaler, pour terminer, que l'on parle également de *vecteur aléatoire* $\begin{pmatrix} X \\ Y \end{pmatrix}$ au lieu de *variable aléatoire* (X, Y).

1.1 Loi de probabilité

La loi de probabilité d'un couple (X, Y) de variables aléatoires à valeurs dans $\mathbb{R}$ est l'application qui permet d'associer à tout couple d'intervalles I et J, de $\mathbb{R}$, la probabilité de l'événement $(X \in I \cap Y \in J)$.

1.2 Fonction de répartition

Si (X, Y) est un couple de variables aléatoires à valeurs dans $\mathbb{R}$, il résulte de la définition d'un tel couple, qu'il existe, en particulier, une fonction F, telle que :

$$F : \begin{cases} \mathbb{R}^2 & \rightarrow & [0, 1] \\ (x, y) & \mapsto & F(x, y) = \mathbb{P}(X < x \cap Y < y) \end{cases}$$

Cette fonction s'appelle la fonction de répartition du couple (X, Y).
On admettra que la fonction de répartition possède les propriétés suivantes, que l'on peut aisément justifier :

1. F est une fonction croissante au sens large, relativement à chacune des variables x et y :

$$\begin{array}{lcl} F(a, y) \leq F(b, y) & , & \text{si } a < b, \quad \forall y \in \mathbb{R} \\ F(x, c) \leq F(x, d) & , & \text{si } c < d, \quad \forall x \in \mathbb{R} \end{array}$$

2.

$$\begin{aligned} \lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) &= \mathbb{P}(\emptyset) = 0^+ \\ \lim_{x, y \to +\infty} F(x, y) &= \mathbb{P}(\Omega) = 1^- \end{aligned}$$

3. Pour $a < b$ et $c < d$

$$\mathbb{P}(X \in [a, b[\cap Y \in [c, d[) = F(b, d) - F(a, d) - F(b, c) + F(a, c) \qquad (6.3)$$

On a en effet :

$$\begin{aligned} \mathbb{P}(a \leq X < b \cap c \leq Y < d) &= \mathbb{P}(X < b \cap Y < d) - \mathbb{P}(X < a \cap Y < d) \\ &\quad - \mathbb{P}(X < b \cap Y < c) + \mathbb{P}(X < a \cap Y < c) \end{aligned}$$

4. En tout point (x, y) de $\mathbb{R}^2$, F est continue à gauche relativement à chacune des variables x et y :

$$\lim_{\epsilon \to 0^+} F(x-\epsilon, y) = \lim_{\epsilon \to 0^+} F(x, y-\epsilon) = F(x, y)$$

   On a, par exemple :

$$\begin{aligned} \lim_{\epsilon \to 0^+} [F(x,y) - F(x-\epsilon, y)] &= \lim_{\epsilon \to 0^+} \mathbb{P}(x - \epsilon \leq X < x \cap Y < y) \\ &= \mathbb{P}(\emptyset) = 0^+ \end{aligned}$$

La fonction de répartition F définit complètement la loi de probabilité du couple (X, Y). Elle permet en effet d'attribuer une probabilité à tout pavé semi-ouvert (voir equation 6.3) ; comme dans le cas des variables aléatoires à valeurs dans $\mathbb{R}$, l'étude des discontinuités de F permet d'ouvrir, ou de fermer, ces pavés à volonté.
Dans la pratique, on ne rencontrera que deux types de variables aléatoires à valeurs dans $\mathbb{R}^2$. Les lois de probabilités correspondantes seront, heureusement, définies plus simplement.

1.3 Les couples de variables discrètes

Lorsque la probabilité de Ω, égale à 1, est entièrement répartie sur un ensemble dénombrable de points de $\mathbb{R}^2$, on dit que le couple (X, Y) est un couple de variables aléatoires discrètes ; il est, en effet, clair que l'ensemble $E_1 = \{x_1, x_2, \ldots, x_n, \ldots,\}$ des observables de X est dénombrable et qu'il en est de même pour $E_2 = \{y_1, y_2, \ldots, y_p, \ldots\}$ des observables de Y.
La loi de probabilité du couple (X, Y) est alors parfaitement définie par l'application :

$$\mathbb{P} : \begin{cases} E_1 \times E_2 & \rightarrow \quad [0,1] \\ (x_i, y_j) & \mapsto \quad \mathbb{P}(X = x_i \cap Y = y_j) = P_{ij} \end{cases}$$

puisque la probabilité d'un événement $(X \in I \cap Y \in J)$ est égale à la somme des probabilités des points de $E_1 \times E_2$ situés à l'intérieur du domaine $I \times J$, seules éventualités mutuellement incompatibles réalisant l'événement. On peut, d'ailleurs, procéder de même pour tout domaine, quelle que soit sa forme.
Tout point n'appartenant pas à $E_1 \times E_2$ et tout domaine ne contenant aucun point de $E_1 \times E_2$ ont, évidemment, une probabilité nulle. Les probabilités P_{ij} définissant la loi du couple vérifient, bien entendu, la relation :

$$\sum_i \sum_j P_{ij} = \mathbb{P}(X \in \mathbb{R} \cap Y \in \mathbb{R}) = \mathbb{P}(\Omega) = 1 \tag{6.4}$$

Exemple 6.1 *Une urne contient trois boules numérotés 1,2,3. On tire successivement et sans remise, deux boules (tirage exhaustif). On peut associer à cette expérience un couple*

(X, Y) de variables aléatoires discrètes, X représentant le numéro que l'on peut obtenir en premier et Y représentant le numéro que l'on peut obtenir en second.
La loi du couple (X, Y) est donnée par le tableau à double entrée :

$Y \backslash X$	1	2	3
1	O	1/6	1/6
2	1/6	0	1/6
3	1/6	1/6	0

Par exemple : $\mathbb{P}(X = 1 \cap Y = 2) = \mathbb{P}(X = 1)\mathbb{P}(Y = 2|X = 1) = \frac{1}{3}\frac{1}{2} = \frac{1}{6}$.
On peut noter, pour tout $i, j \in \{1, 2, 3\}$:

$$\begin{cases} P_{ij} = \frac{1}{6} \text{ , si } i \neq j \\ P_{ij} = 0 \text{ , si } i = j \end{cases}$$

Exemple 6.2 *Si, avec la même urne, on tire, successivement et avec remise, deux boules (tirage non exhaustif), la loi du couple (X, Y) devient :*

$Y \backslash X$	1	2	3
1	1/9	1/9	1/9
2	1/9	1/9	1/9
3	1/9	1/9	1/9

Par exemple : $\mathbb{P}(X = 1 \cap Y = 2) = \mathbb{P}(X = 1)\mathbb{P}(Y = 2|X = 1) = \frac{1}{3}\frac{1}{3} = \frac{1}{9}$.
On peut noter, pour tout $i, j \in \{1, 2, 3\}$:

$$P_{ij} = \frac{1}{9}$$

1.4 Les couples à densité

On dit qu'un couple (X, Y) de variables aléatoires à valeurs réelles, est un couple à densité, s'il existe une fonction numérique de deux variables réelles, g, possédant la propriété : quel que soit le domaine D, de $\mathbb{R}^2$, on a

$$\mathbb{P}[(X, Y) \in D] = \int\int_D g(x, y)\mathrm{d}x\mathrm{d}y \tag{6.5}$$

On dit, encore, que le couple (X, Y) est de distribution absolument continue.
Il est clair que la loi de probabilité du couple (X, Y) est complètement définie par sa densité g.
En particulier, si F est la fonction de répartition du couple (X, Y), on a :

$$F(u, v) = \mathbb{P}(X < u \cap Y < v) = \int_{-\infty}^{u}\int_{-\infty}^{v} g(x, y)\mathrm{d}x\mathrm{d}y$$

Réciproquement, on peut montrer que, en tout point (x, y) où g est continue, on :

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = g(x,y)$$

La fonction densité g d'un couple de variables aléatoires à valeurs réelles est, par définition, une fonction intégrable sur $\mathbb{R}^2$. Elle possède les propriétés suivantes, que l'on rapprochera de celles de la densité d'une variable aléatoire à valeurs dans $\mathbb{R}$:

1. $g(x, y) \geq 0$, sur $\mathbb{R}^2$, sauf éventuellement en "certains points". Cette condition est indispensable pour que, pour tout domaine D, on ait $\mathbb{P}[(X, Y) \in D] \geq 0$
2. $\int\int_{\mathbb{R}^2} g(x, y)\mathrm{d}x\mathrm{d}y = \mathbb{P}[(X, Y) \in \mathbb{R}^2] = \mathbb{P}(\Omega) = 1$

On note, enfin, que si dx et dy sont des infiniment petits, on a :

$$\mathbb{P}(x \leq X < x + \mathrm{d}x \cap y \leq Y < y + \mathrm{d}y) \sim g(x,y)\mathrm{d}x\mathrm{d}y$$

quantité dite *probabilité élémentaire*

Exemple 6.3 *Soit (X, Y) un couple de variables aléatoires, à valeurs réelles, de densité de probabilité :*

$$g : \begin{cases} (x,y) \mapsto e^{-x}, si (x,y) \in D \\ (x,y) \mapsto 0 \ \ , si (x,y) \notin D \end{cases}$$

où D est défini par $0 \leq y \leq x$
On peut vérifier dans un premier temps, que la fonction g possède les propriétés d'une densité de probabilité. D'une part : $g(x, y) \geq 0$ sur $\mathbb{R}^2$. D'autre part :

$$\begin{aligned} \int\int_{\mathbb{R}^2} g(x,y)dxdy &= \int\int_D e^{-x}dxdy \\ &= \int_o^{+\infty} e^{-x}dx \int_0^x dy \\ &= \int_0^{+\infty} xe^{-x}dx \\ &= \left[-xe^{-x} - e^{-x}\right]_0^{+\infty} \\ &= 1 \end{aligned}$$

Calculons maintenant la probabilité de l'événement $[(X, Y) \in \Delta]$, où Δ est défini par : $0 \leq x + y \leq t$:

$$\begin{aligned} \mathbb{P}[(X,Y) \in \Delta] &= \int\int_\Delta g(x,y)dxdy \\ &= \int\int_{\Delta \cap D} e^{-x}dxdy \quad g(x,y) = 0 \textit{ pour } (x,y) \notin \Delta \cap D \\ &= \int_0^{t/2} dy \int_y^{t-y} e^{-x}dx \end{aligned}$$

$$\begin{aligned}
&= \int_0^{t/2} \left(-e^{-t+y} + e^{-y}\right) dy \\
&= \left[-e^{-t+y} - e^{-y}\right]_0^{t/2} \\
&= 1 - 2e^{-t/2} + e^{-t}
\end{aligned}$$

La fonction de répartition F du couple (X,Y) est définie par : $F(u,v) = \int_{-\infty}^{u}\int_{-\infty}^{v} g(x,y)dxdy$, on obtient (à vérifier) :

$$F : \begin{cases} F(u,v) = 0 & , si \quad u \leq 0 \text{ ou } v \leq 0 \\ F(u,v) = 1 - e^{-u} - ue^{-u}, si & 0 \leq u \leq v \\ F(u,v) = 1 - e^{-v} - ve^{-u}, si & 0 \leq v \leq u \end{cases}$$

Exemple 6.4 *Soit (X,Y) un couple de variables aléatoires, à valeurs réelles, de densité de probabilité*

$$g : (x,y) \mapsto \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)},\ \forall (x,y) \in \mathbb{R}^2$$

La fonction g possède les propriétés d'une densité : $g(x,y) > 0$ sur $\mathbb{R}^2$ et

$$\int\int_{\mathbb{R}^2} g(x,y)dxdy = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = 1 \times 1 = 1$$

Si Δ est défini par : $x \geq 0$, $y \geq 0$, $x^2 + y^2 \leq 2$ (quart de cercle), on a :

$$\begin{aligned}
\mathbb{P}[(X,Y) \in \Delta] &= \int\int_{\Delta} g(x,y)dxdy \\
&= \frac{1}{2\pi} \int\int_{\Delta} e^{-\frac{1}{2}(x^2+y^2)} dxdy
\end{aligned}$$

On calcule cette intégrale en passant en coordonnées polaires :

$$\begin{aligned}
\mathbb{P}[(X,Y) \in \Delta] &= \frac{1}{2\pi} \int_0^{\pi/2} d\theta \int_0^{\sqrt{2}} e^{-r^2/2} r dr \\
&= \frac{1}{2\pi} \frac{\pi}{2} \left[-e^{-r^2/2}\right]_0^{\sqrt{2}} \\
&= \frac{1}{4}[1 - e^{-1}] \\
&\approx 15.80\%
\end{aligned}$$

Si G est la fonction de répartition du couple (X,Y), on obtient ici :

$$\begin{aligned}
G(u,v) &= \int_{-\infty}^{u} \int_{-\infty}^{v} g(x,y)dxdy \\
&= \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \int_{-\infty}^{v} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\
&= F_{X^*}(u) F_{X^*}(v)
\end{aligned}$$

où X^* est une variable normale réduite et F_{X^*} est sa fonction de répartition.

2 Lois marginales

Comme il a déjà été mentionné (voir équation 6.2), de la loi d'un couple (X, Y), on doit toujours pouvoir déduire la loi de chacune des deux variables aléatoires X et Y. On dit que ces deux lois de probabilité sont des *lois marginales*, ce qui signifie qu'elles ne sont qu'une conséquence de la loi du couple.
On va montrer comment procéder pour déterminer, dans les cas usuels, les lois marginales.

2.1 *Couples de variables discrètes*

Soit l'application

$$P : \left\{ \begin{array}{lcl} E_1 \times E_2 & \rightarrow & [0,1] \\ (x_i, y_j) & \mapsto & P_{ij} = \mathbb{P}(X = x_i \cap Y = y_j) \end{array} \right.$$

qui définit la loi du couple (X, Y) de variables discrètes.
Il est clair que E_1 est l'ensemble, dénombrable, des observables de X. Pour chaque valeur $x_i \in E_1$, on peut alors écrire :

$$(X = x_i) = (X = x_i \cap Y = y_1) \cup (X = x_i \cap Y = y_2) \cup (X = x_i \cap Y = y_3) \cup \cdots$$

Les différentes éventualités étant incompatibles, on en déduit :

$$\mathbb{P}(X = x_i) = \sum_j \mathbb{P}(X = x_i \cap Y = y_j)$$

ou, en abrégé :

$$P_{i.} = \sum_j P_{ij} \tag{6.6}$$

On aura de même pour tout $y_j \in E_2$, ,$\mathbb{P}(Y = y_j) = \sum_i \mathbb{P}(X = x_i \cap Y = y_j)$, c'est-à-dire :

$$P_{.j} = \sum_i P_{ij} \tag{6.7}$$

Exemple 6.5 *(suite de l'exemple 6.1) À partir de la loi du couple donné par le tableau, on obtient la loi de X en additionnant les probabilités par colonne et la loi de Y en additionnant les probabilités par ligne*

Y \ X	1	2	3	
1	0	1/6	1/6	1/3
2	1/6	0	1/6	1/3
3	1/6	1/6	0	1/3
	1/3	1/3	1/3	

Et donc

x_i	*1*	*2*	*3*
$P_{i.}$	⅓	⅓	⅓

y_i	*1*	*2*	*3*
$P_{.j}$	⅓	⅓	⅓

Par exemple : $\mathbb{P}(Y=2) = \mathbb{P}(X=1 \cap Y=2) + \mathbb{P}(X=2 \cap Y=2) + \mathbb{P}(X=3 \cap Y=2) = 1/6 + 0 + 1/6 = 1/3$.

Exemple 6.6 *(suite de l'exemple 6.2) On obtient de la même manière :*

$Y \backslash X$	*1*	*2*	*3*	
1	1/9	1/9	1/9	1/3
2	1/9	1/9	1/9	1/3
3	1/9	1/9	1/9	1/3
	1/3	1/3	1/3	

Et donc

x_i	*1*	*2*	*3*
$P_{i.}$	⅓	⅓	⅓

y_i	*1*	*2*	*3*
$P_{.j}$	⅓	⅓	⅓

Par exemple : $\mathbb{P}(Y=2) = \mathbb{P}(X=1 \cap Y=2) + \mathbb{P}(X=2 \cap Y=2) + \mathbb{P}(X=3 \cap Y=2) = 1/9 + 1/9 + 1/9 = 1/3$.

On remarque que dans ces deux exemples, X et Y ont toujours les mêmes lois ; ceci n'est pas surprenant car, quelque soir le schéma, chaque boule a autant de chance qu'une autre de sortir au premier ou au deuxième tirage, c'est-à-dire une chance sur trois. Ce qui différencie les deux schémas, c'est la manière dont les observables sont appariées. Il semble, en, particulier, relativement facile dans le deuxième exemple, d'établir une règle pour reconstruire la loi du couple (X,Y) à partir de la loi de X et de la loi de Y ; ce n'est pas le cas dans le premier exemple.

2.2 *Couples à densité*

Soit g la densité de probabilité d'un couple (X,Y) de variables aléatoires à valeurs réelles.
La loi de probabilité de X peut être définie par sa fonction de répartition F_X. On a, en effet :

$$F_X(t) = \mathbb{P}(X<t) = \mathbb{P}(X<t \cap Y \in \mathbb{R}) = \int\int_D g(x,y)\mathrm{d}x\mathrm{d}y$$

où D est défini par $x<t$ et $y \in \mathbb{R}$. On a ;

$$F_X(t) = \int_{-\infty}^{t} \mathrm{d}x \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}y$$

Si on pose $f_X(x) = \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}y$, on en déduit que X est une variable à densité, de densité :

$$f_X : x \mapsto \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}y, \quad \forall x \in \mathbb{R} \tag{6.8}$$

De même on a

$$f_Y : y \mapsto \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}x, \quad \forall y \in \mathbb{R} \tag{6.9}$$

Exemple 6.7 *(suite de l'exemple 6.4) Ici la densité f_X de la variables aléatoires est définie, pour tout $x \in \mathbb{R}$:*

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} g(x,y)dy \\ &= \int_{-\infty}^{+\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \end{aligned}$$

De même, la densité f_Y de la variable aléatoire Y, est définie par

$$f_Y(y) = \int_{-\infty}^{+\infty} g(x,y)dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \forall y \in \mathbb{R}$$

On dit que le couple (X,Y) est un couple de variables aléatoires normales réduites. On note qu'ici, contrairement à l'exemple 6.3, la densité du couple est égale au produit des densités de X et de Y.

3 Variables indépendantes

Définition 6.2 *Soit (X,Y) un couple de variables aléatoires à valeurs dans $\mathbb{R}$. On dit que les deux variables X et Y sont indépendantes si, pour tout couple I et J d'intervalles de $\mathbb{R}$, on a la propriété :*

$$\mathbb{P}(X \in I \cap Y \in J) = \mathbb{P}(X \in I)\mathbb{P}(Y \in J) \tag{6.10}$$

Dans tous les autres cas, on dit que les variables X et Y sont dépendantes, ou liées, ou corrélées.
La définition 6.10 est remplacée, pour les couple discrets et pour les couples à densité, par une définition équivalente et plus commode à manipuler.

3.1 Couples de variables discrètes

Soit

$$\mathbb{P} : \left\{ \begin{array}{lcl} E_1 \times E_2 & \rightarrow & [0,1] \\ (x_i, y_j) & \mapsto & \mathbb{P}(X = x_i \cap Y = y_j) = P_{ij} \end{array} \right.$$

l'application qui définit la loi du couple (X, Y). On dit que les variables aléatoires X et Y sont indépendantes si et seulement si :

$$\mathbb{P}(X = x_i \cap Y = Y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) \quad \forall (x_i, y_j) \in E_1 \times E_2 \tag{6.11}$$

Il est clair que l'équation 6.10 implique l'équation 6.11 : il suffit de réduire le domaine de $I \times J$ au point x_i, y_j.
Réciproquement l'équation 6.11 implique l'équation 6.10. Si on considère un intervalle I contenant les observables $x_n, x_{n+1}, \ldots, x_{n+k}$ de X et un intervalle J contenant les observables $y_p, y_{p+1}, \ldots, y_{p+l}$ de Y, on a, en effet :

$$\mathbb{P}(X \in I \cap Y \in J) = \sum_{i=n}^{n+k} \sum_{j=p}^{p+l} \mathbb{P}(X = x_i \cap Y = y_j)$$

$$\begin{aligned}
(6.11) \Rightarrow \mathbb{P}(X \in I \cap Y \in J) &= \sum_{i=n}^{n+k} \sum_{j=p}^{p+l} \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) \\
&= \sum_{i=n}^{n+k} \mathbb{P}(X = x_i) \sum_{j=p}^{p+l} \mathbb{P}(X = x_i) \\
&= \mathbb{P}(X \in I)\mathbb{P}(Y \in J)
\end{aligned}$$

Si on remarque que la relation 6.10 est triviale si l'un au moins des deux intervalles ne contient aucun observable.
En abrégé, la relation 6.11 peut s'écrire :

$$P_{ij} = P_i P_j \quad \forall i \text{ et } \forall j \tag{6.12}$$

Exemple 6.8 *(suite des exemples 6.1 et 6.2) La relation 6.12 étant vérifiée pour tout $i, j = 1, 2, 3$, les variables X et Y de l'exemple 6.2 sont indépendantes. Par contre les variables X et Y de l'exemple 6.1 ne sont pas indépendantes.*

3.2 Couples à densité

Soit (X, Y) un couple de variables aléatoires à valeurs réelles et de densité g. On désigne par f_X la densité de X et par f_Y la densité de Y (voir équation 6.8 et 6.9)
Les variables aléatoires Xet Y sont indépendantes si et seulement si :

$$g(x, y) = f_X(x) f_Y(y) \tag{6.13}$$

L'indépendance de X et de Y implique, en particulier, $\forall (u, v) \in \mathbb{R}^2$:

$$\mathbb{P}(X < u \cap Y < v) = \mathbb{P}(X < u)\mathbb{P}(Y < v)$$

et donc :

$$\int_{-\infty}^{u}\int_{-\infty}^{v} g(x,y)\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{u} f_X(x)\mathrm{d}x \int_{-\infty}^{v} f_Y(y)\mathrm{d}y$$

On en déduit : $g(x,y) = f_X(x)f_Y(y)$
Réciproquement, si $g(x,y) = f_X(x)f_Y(y)$ on aura pour tout domaine $I \times J$:

$$\begin{aligned}
\mathbb{P}(X \in I \cap Y \in J) &= \int\int_{I\times J} g(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int\int_{I\times J} f_X(x)f_Y(y)\mathrm{d}x\mathrm{d}y \\
&= \int_I f_X(x)\mathrm{d}x \int_J f_Y(y)\mathrm{d}y \\
&= \mathbb{P}(X \in I)\mathbb{P}(Y \in J)
\end{aligned}$$

La relation 6.13 est donc bien équivalente à la relation 6.10

Exemple 6.9 *(suite de l'exemple 6.3 et 6.4) Les variables X et Y de l'exemple 6.4 sont indépendantes puisque, sur $\mathbb{R}^2$, la relation 6.13 est vérifiée.*
Par contre, les variables X et Y de l'exemple 6.3 ne sont pas indépendantes.

4 Lois conditionnelles

Comme nous l'avons vu, de la loi d'un couple de variables aléatoires, on peut toujours déduire la loi de chacune des deux variables. Il convient maintenant de s'intéresser à la liaison entre ces deux variables, c'est-à-dire à la manière dont les observables sont appariées dans la loi du couple. L'indépendance, ou absence de liaison, décrit une situation particulière ; les lois conditionnelles vont permettre une description plus générale de la liaison entre deux variables.
Une loi conditionnelle donne, pour une valeur fixée de l'une des variables, la loi de probabilité de l'autre. On conçoit l'intérêt des lois conditionnelles. Si par exemple, on s'intéresse à la taille et au poids d'un individu choisi au hasard dans une population donnée, il est important de pouvoir décrire la distribution de la taille des individus de même poids et de suivre l'évolution de cette distribution en fonction du poids.

4.1 Couples de variables discrètes

La loi du couple (X,Y) est, dans ce cas, définie par une application

$$\mathbb{P} : \left\{ \begin{array}{lcl} E_1 \times E_2 & \longrightarrow & [0,1] \\ (x_i, y_j) & \longmapsto & \mathbb{P}(X = x_i \cap Y = y_j) = P_{ij} \end{array} \right.$$

La relation des probabilités composées permet d'écrire, dans tous les cas :

$$\mathbb{P}(X = x_i \cap Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j | X = x_i)$$

Pour toute observable $x_i \in E_1$, de probabilité non nulle, il existe donc une application :

$$\begin{cases} E_2 & \to & [0,1] \\ y_j & \mapsto & \mathbb{P}(Y=y_j|X=x_i) = \frac{P_{ij}}{P_{i.}} \end{cases}$$

Cette application définit ce que l'on appelle la loi conditionnelle de Y sachant que X à la valeur x_i. Il existe autant de lois conditionnelles de Y que de valeurs de X de probabilité non nulle.
Si on écrit : $\mathbb{P}(X = x_i \cap Y = y_j) = \mathbb{P}(X = x_i|Y = y_j)\mathbb{P}(Y = y_j)$, on peut définir de même, pout toute observable $y_j \in E_2$, de probabilité non nulle, la loi conditionnelle de X sachant que Y à la valeur y_j par l'application

$$\begin{cases} E_1 & \to & [0,1] \\ x_i & \mapsto & \mathbb{P}(X=x_i|Y=y_j) = \frac{P_{ij}}{P_{.j}} \end{cases}$$

On remarque que ces applications définissent bien des lois de probabilité puisque, par exemple :

$$0 \leq P_{ij} \leq P_{.j} \Rightarrow 0 \leq \frac{P_{ij}}{P_{ij}} \leq 1, (P_{ij} \neq 0)$$

et

$$\sum_i \frac{P_{ij}}{P_{i.j}} = \frac{1}{P_{.j}} \sum_i P_{ij} = \frac{P_{.j}}{P_{.j}} = 1$$

Exemple 6.10 *(suite de l'exemple 6.1)*

Lois conditionnelles de X

x_i	1	2	3
$\mathbb{P}(X=x_i\|Y=1)$	0	½	½

x_i	1	2	3
$\mathbb{P}(X=x_i\|Y=2)$	½	0	½

x_i	1	2	3
$\mathbb{P}(X=x_i\|Y=3)$	½	½	0

Lois conditionnelles de Y

y_i	1	2	3
$\mathbb{P}(Y=y_j\|X=1)$	0	½	½

y_i	1	2	3
$\mathbb{P}(Y=y_j\|X=2)$	½	0	½

y_i	1	2	3
$\mathbb{P}(Y=y_j\|X=3)$	½	½	0

Par exemple : $\mathbb{P}(Y=2|X=1) = \frac{\mathbb{P}(X=1\cap Y=2)}{\mathbb{P}(X=1)} = \frac{1/6}{1/3} = \frac{1}{2}$

Exemple 6.11 *(suite de l'exemple 6.2) Les lois conditionnelles de X ou Y peuvent être déterminées comme précédemment. On note que $\mathbb{P}(X = i|Y = j) = \mathbb{P}(X = i)$ a par exemple :* $\mathbb{P}(X=2|Y=1) = \frac{\mathbb{P}(X=2\cap Y=1)}{\mathbb{P}(Y=1)} = \frac{1/9}{1/3} = \frac{1}{3} = \mathbb{P}(X=2)$
Ce résultat est général :

$$\mathbb{P}(X=i|Y=j) = \frac{P_{ij}}{P_{.j}} = \frac{P_{i.}P_{.j}}{P_{.j}} = P_{i.} = \mathbb{P}(X=x_i)\,, \forall y_j \in E_2$$

$$\mathbb{P}(Y=j|X=i) = \frac{P_{ij}}{P_{i.}} = \frac{P_{i.}P_{.j}}{P_{i.}} = P_{.j} = \mathbb{P}(Y=y_j)\,, \forall x_i \in E_1$$

Dans l'exemple 6.2, les variables sont indépendantes et les lois conditionnelles de chacune des variables n'évoluent pas en fonction des valeurs prises par l'autre ; dans l'exemple 6.1 ces lois évoluent et les deux variables sont liées.
On notera que le passage de $\mathbb{P}(X = x_i \cap Y = y_j)$ à $\mathbb{P}(X = i|Y = j)$ s'effectue par l'intermédiaire de la division par $\mathbb{P}(Y = y_j)$; ceci correspond au changement d'ensemble fondamental qui intervient dans toute probabilité conditionnelle et se traduit, dans le schéma, par un simple changement d'échelle.

4.2 Couple à densité

On rappelle les notations : le couple (X, Y) a pour densité g, f_X et f_Y sont les densités respectives de X et de Y. On a :

$$\mathbb{P}(x \leq X < x+\mathrm{d}x \cap y \leq Y < y+\mathrm{d}y) = \mathbb{P}(x \leq X < x+\mathrm{d}x)\mathbb{P}(y \leq Y < y+\mathrm{d}y|x \leq X < x+\mathrm{d}x)$$

Si dx et dy tendent vers zéro, on peut écrire :

$$\begin{aligned} \mathbb{P}(x \leq X < x + \mathrm{d}x \cap y \leq Y < y + \mathrm{d}y) &\sim g(x,y)\mathrm{d}x\mathrm{d}y \\ \mathbb{P}(x \leq X < x + \mathrm{d}x) &\sim f_X(x)\mathrm{d}x \end{aligned}$$

On en déduit :

$$\mathbb{P}(y \leq Y < y + \mathrm{d}y|X = x) \sim \frac{g(x,y)}{f_X(x)} \text{, si } f_X(x) \neq 0$$

Pour tout réel x tel que $f_X(x) \neq 0$, on peut définir une fonction $f_{Y|X=x}$:

$$f_{Y|X=x} : y \mapsto f_{Y|X=x}(y) = \frac{g(x,y)}{f_X(x)}$$

Cette fonction s'appelle la densité conditionnelle de Y sachant que X a la valeur x.
De même, pour tout réel y tel que $f_Y(y) \neq 0$, on peut définir une fonction $f_{X|Y=y}$:

$$f_{X|Y=y} : x \mapsto f_{X|Y=y}(x) = \frac{g(x,y)}{f_Y(y)}$$

appelée densité conditionnelle de X sachant que Y a la valeur y.
On notera que ces fonctions ont bien les propriétés d'une densité puisque, par exemple :

$$\begin{aligned} f_{X|Y=y}(x) &= \frac{g(x,y)}{f_Y(y)} \geq 0 \\ \int_{-\infty}^{+\infty} f_{X|Y=y}(x)\mathrm{d}x &= \frac{1}{f_Y(y)} \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}x \\ &= \frac{f_Y(y)}{f_Y(y)} = 1 \end{aligned}$$

On remarque au passage que cette dernière intégrale a la valeur 1 à cause de la division par $f_Y(y)$ ce qui correspond au changement d'ensemble fondamental déjà signalé dans le cas des variables discrètes.
Les lois conditionnelles de X sont ici définies par une densité où la valeur y prise par Y intervient comme un paramètre. Ces lois vont se déformer lorsque y varie ; cette évolution est un des effets de la liaison entre les variables. Il en est de même des lois conditionnelles de Y qui dépendront de la valeur x prise par X.
Il est clair que si X et Y sont des variables indépendantes les lois conditionnelles de l'une n'évoluent pas en fonction des valeurs prises par l'autre et que : $f_{Y|X=x} = f_Y$; $f_{X|Y=y} = f_X$ à cause de la relation 6.13

Exemple 6.12 *(suite de l'exemple 6.4) Les variables sont indépendantes et on a, ici :*

$$\forall y \in \mathbb{R}, \quad f_{X|Y=y}(x) = f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, x \in \mathbb{R}$$

$$\forall y \in \mathbb{R}, \quad f_{Y|X=x}(y) = f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, x \in \mathbb{R}$$

Exemple 6.13 *(suite de l'exemple 6.3) On peut définir la densité conditionnelle de X sachant $Y = y$ seulement si $y \geq 0$ ($f_Y(y) \neq 0$). On a alors :*

$$f_{X|Y=y} = \frac{g(x,y)}{f_Y(y)} = \begin{cases} e^{-x+y} & \text{, si } x \geq y \\ 0 & \text{, si } x < y \end{cases}$$

On peut aussi définir la loi conditionnelle de Y sachant $X = x$ seulement si $x > 0$ ($f_X(x) \neq 0$). On a alors :

$$f_{Y|X=x} = \frac{g(x,y)}{f_X(x)} = \begin{cases} 1/x & \text{, si } 0 \leq y \leq x \\ 0 & \text{, si } x < y \text{ ou } y < 0 \end{cases}$$

5 Espérances conditionnelles et courbes de régression

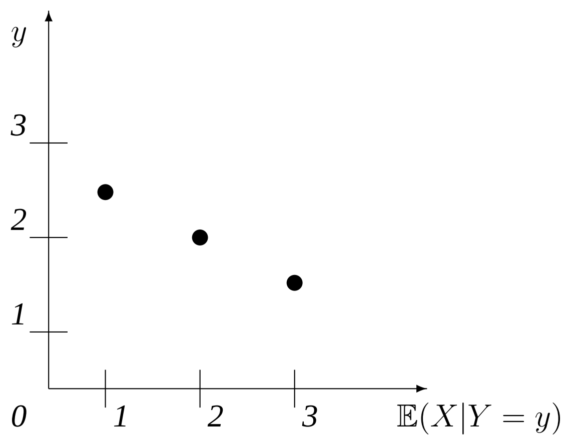
On peut associer à la distribution conditionnelle de Y sachant que X a la valeur x, tous les paramètres descriptifs usuels des distributions d'une variable aléatoire dans $\mathbb{R}$.
En particulier, sous réserve de la convergence de la somme, ou de l'intégrale correspondante, on peut lui associer une espérance. Cette espérance s'appelle l'espérance conditionnelle de Y sachant que X à la valeur x et on la note $\mathbb{E}(Y|X = x)$.
Le point $(x, \mathbb{E}(Y|X = x)$ est le barycentre de la distribution de Y sachant $X = x$. L'ensemble de ces points s'appelle la courbe de régression de l'espérance conditionnelle $\mathbb{E}(Y|X = x)$; cette courbe de régression représente donc l'évolution de $\mathbb{E}(Y|X = x)$ en fonction de x.
On définit de même $\mathbb{E}(X|Y = y)$, espérance conditionnelle de X lorsque Y a pris la valeur y, et la courbe de régression correspondante, ensemble des barycentres $(\mathbb{E}(X|Y = y), y)$ des distributions conditionnelles de X.

Les courbes de régression $\mathbb{E}(Y|X = x)$ et de $\mathbb{E}(X|Y = y)$ sont, a priori, différentes : si l'espérance de la taille d'un individu de 80 kilos est de 180 centimètres, il n'y a pas de raison que l'espérance du poids d'un individu de 80 kilos soit de 180 cm.
Les courbes de régression représentent, d'une manière synthétique, un des aspects de la liaison entre deux variables mais donnent, bien entendu, moins d'informations sur cette liaison que l'ensemble des lois conditionnelles ; toutes les informations sont, par ailleurs, contenues dans la loi du couple.

Exemple 6.14 *(suite de l'exemple 6.1) On a :*

$$\begin{array}{rcl}\mathbb{E}(X|Y=1) & = & 2.5\\ \mathbb{E}(X|Y=2) & = & 2\\ \mathbb{E}(X|Y=3) & = & 1.5\end{array}$$

D'où le dessin :



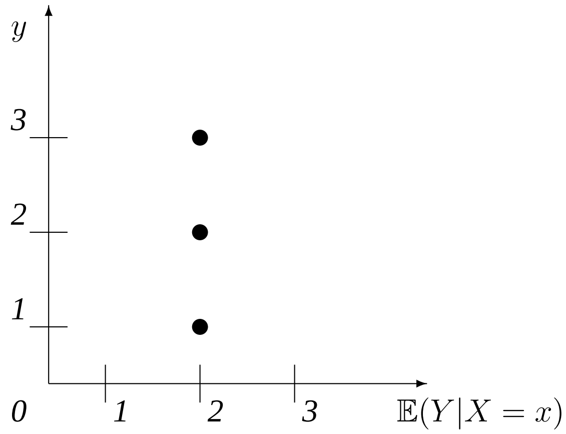
De même
On a :

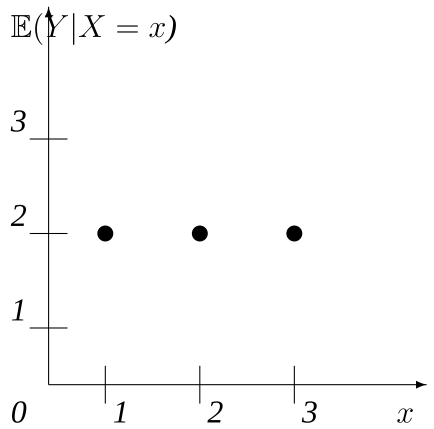
$$\begin{array}{rcl}\mathbb{E}(Y|X=1) & = & 2.5\\ \mathbb{E}(Y|X=2) & = & 2\\ \mathbb{E}(Y|X=3) & = & 1.5\end{array}$$

Exemple 6.15 *(suite de l'exemple 6.2)*
Dans ce cas, pour tout $x_i, y_j = 1, 2, 3$, on a

$$\mathbb{E}(X|Y = y_j) = \mathbb{E}(Y|X = x_i) = \frac{1}{3}1 + \frac{1}{3}2 + \frac{1}{3}3 = 2$$

D'où le dessin :





Les variables étant indépendantes, les lois conditionnelles n'évoluent pas et les espérances conditionnelles restent constantes. Les points représentatifs sont donc situés sur une parallèle à l'axe des x ou sur une parallèle à l'axe des y.
Il en sera de même dans l'exemple 6.4 où $\mathbb{E}(X|Y=y)=0\ \forall y\in\mathbb{R}$ la courbe de régression est donc parallèle à l'axe $y'Oy$.
$\mathbb{E}(Y|X=x)=0\ \forall x\in\mathbb{R}$ la courbe de régression est donc parallèle à l'axe $x'Ox$.

Il convient de souligner que les lois conditionnelles peuvent évoluer, c'est-à-dire que les variables peuvent être liées, tout en conservant des espérances conditionnelles constantes ; la "verticalité" et l'"horizontalité" des courbes de régression ne signifient donc pas forcément l'indépendance des variables.

Exemple 6.16 *(suite de l'exemple 6.3)*
Pour tout $y\geq 0$, on a

$$\begin{aligned}\mathbb{E}(X|Y=y) &= \int_{-\infty}^{+\infty} x f_{X|Y=y}(x)dx \\ &= \int_{-\infty}^{y} x f_{X|Y=y}(x)dx + \int_{y}^{+\infty} x f_{X|Y=y}(x)dx\end{aligned}$$

$$\begin{aligned} &= 0 + \int_y^{+\infty} xe^{-x+y} dx \\ &= e^y \left[-xe^{-x} - e^{-x}\right]_y^{+\infty} \\ &= 1 + y \end{aligned}$$

Et donc la courbe de régression correspondante est une droite.
De même, pour $x > 0$, on a :

$$\begin{aligned} \mathbb{E}(Y|X = x) &= \int_{-\infty}^{+\infty} y f_{Y|X=x}(y) dy \\ &= \int_{-\infty}^{0} y f_{Y|X=x}(y) dy + \int_0^x y f_{Y|X=x}(y) dy + \int_x^{+\infty} y f_{Y|X=x}(y) dy \\ &= 0 + \int_0^x y \frac{1}{x} dy + 0 \\ &= \frac{1}{x} \left[\frac{y^2}{2}\right]_0^x \\ &= \frac{x}{2} \end{aligned}$$

La courbe de régression correspondante est encore une droite qui, cette fois passe par l'origine.

6 Covariance et coefficient de corrélation

Soit (X, Y) un couple de variables aléatoires à valeurs réelles. On suppose que chacune des deux variables possède une espérance et une variance, calculées à partir des lois marginales, par exemple, et on pose :
$\mathbb{E}(X) = \mu_X$; $\mathbb{E}(Y) = \mu_Y$; $\mathbb{V}(Y) = \sigma_X^2$; $\mathbb{V}(Y) = \sigma_Y^2$

6.1 *Définition*

On appelle covariance des variables X et Y et on note Cov(X, Y), σ_{XY} ou plus généralement $\mathbb{E}[(X - \mu_x)(Y - \mu_Y)]$, le nombre, s'il existe, défini par :

$$\sigma_{XY} = \sum_i \sum_j (x_i - \mu_X)(y_j - \mu_Y) P_{ij}, \text{ si les variables sont discrètes}$$

$$\sigma_{XY} = \int\int_{\mathbb{R}^2} (x - \mu_X)(y - \mu_Y) g(x, y) \mathrm{d}x\mathrm{d}y, \text{ pour un couple à densité}$$

On appelle coefficient de corrélation des variables X et Y, et on note ρ_{XY}, le nombre, s'il existe, défini par :

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

6.2 *Covariance*

Ce premier coefficient, qui s'introduit naturellement dans certains calculs probabilistes (voir chapitre 7), possède la propriété d'être nul lorsque les variables sont indépendantes. On a, en effet, dans ce cas :

$$\sigma_{XY} = \sum_i \sum_j (x_i - \mu_X)(y_j - \mu_Y) P_{i.} P_{.j} = \sum_j (y_j - \mu_Y) P_{.j} \sum_i (x_i - \mu_X) P_{i.} = 0 \times 0 = 0$$

si les variables sont discrètes, et si les variables ont une densité :

$$\begin{aligned} \sigma_{XY} &= \int\int (x-\mu_X)(y-\mu_Y) f_X(x) f_Y(y) \mathrm{d}x \mathrm{d}y \\ &= \int_{\mathbb{R}} (x-\mu_X) f_X(x) \mathrm{d}x \int_{\mathbb{R}} (y-\mu_Y) f_Y(y) \mathrm{d}y \\ &= 0 \end{aligned}$$

puisque, pour toute variable Z d'espérance μ_Z : $\mathbb{E}(Z - \mu_Z) = \mathbb{E}(Z) - \mu_Z = 0$.
Réciproquement, une covariance nulle n'implique pas l'indépendance entre les variables sauf, éventuellement, à l'intérieur de certains modèles, par exemple les couples gaussiens. Malgré cette remarque, la covariance donne les indications suivantes sur la liaison entre deux variables :

- Lorsque $\sigma_{XY} \neq 0$, les variables X et Y sont liées ; ceci résulte de la propriété précédente ;
- lorsque $\sigma_{XY} < 0$, les écarts $(X - \mu_X)$ et $(Y - \mu_Y)$ ont tendance à être de signe contraire et X et Y ont tendance à évoluer en sens contraire ;
- lorsque $\sigma_{XY} > 0$, les écarts $(X - \mu_X)$ et $(Y - \mu_Y)$ ont tendance à être de même signe et X et Y ont tendance à évoluer en même sens.

6.3 *Coefficient de corrélation*

Le coefficient de corrélation est utilisé pour mesurer le degré de liaison entre deux variables. Ceci est justifié par ses propriétés :

1. $-1 \leq \rho_{XY} \leq 1$ (voir chapitre 7).
   On dispose donc d'une échelle commune à tous les couples
2. $|\rho_{XY}| = 1 \Leftrightarrow Y = aX + b$. De plus $\rho = 1 \Rightarrow a > 0$ et $\rho = -1 \Rightarrow a < 0$ (voir chapitre 7).
   On dit qu'il existe entre X et Y une liaison fonctionnelle : quand X a pris une valeur, celle de Y est imposée ; la liaison linéaire est, dans l'échelle choisie, considérée comme la plus forte.
3. L'indépendance de X et Y implique $\rho_{XY} = 0$ (voir la définition de σ_{XY}).
   C'est la liaison la plus faible. Malheureusement la réciproque n'est pas vraie sauf pour des cas particuliers comme les couples gaussiens.
4. $\rho_{XY} > 0 \Rightarrow X$ et Y ont tendance à évoluer dans le même sens.
   $\rho_{XY} < 0 \Rightarrow X$ et Y ont tendance à évoluer en sens contraire.

Exemple 6.17 *(suite de l'exemple 6.1) On déduit des lois marginales :*
$\mathbb{E}(X) = \mathbb{E}(Y) = \frac{1}{3}(1+2+3) = 2$
$\mathbb{V}(X) = \mathbb{V}(Y) = \frac{1}{3}(-1)^2 + \frac{1}{3}(0)^2 + \frac{1}{3}(1)^2 = \frac{2}{3}$
Le calcul de la covariance donne, sans tenir compte des termes nuls :

$$\sigma_{XY} = (3-2)(1-2)\frac{1}{6} + (1-2)(3-2)\frac{1}{6} = -\frac{2}{6}$$

On en déduit le coefficient de corrélation

$$\rho_{XY} = \frac{-2/6}{\sqrt{2/3}\sqrt{2/3}} = -\frac{1}{2}$$

Les variables sont liées. Les valeurs négatives obtenues pour σ_{XY} et ρ_{XY} confirment l'évolution des distributions conditionnelles.

Exemple 6.18 *(suite de l'exemple 6.3) On déduit des lois marginales :*
$\mathbb{E}(X) = \int_0^{+\infty} x^2 e^{-x} dx = [-x^2 e^{-x}]_0^{+\infty} + 2\int_0^{+\infty} x e^{-x} dx = 0 + 2 = 2$
$\mathbb{V}(X) = \int_0^{+\infty} x^3 e^{-x} dx = [-x^3 e^{-x}]_0^{+\infty} + 3\int_0^{+\infty} x^2 e^{-x} dx = 0 + 6 = 6$

$\mathbb{E}(Y) = \int_0^{+\infty} y e^{-y} dy = 1$
$\mathbb{E}(Y^2) = \int_0^{+\infty} y^2 e^{-y} dy = 2$
$\mathbb{V}(Y) = 2 - 1 = 1$
Le calcul de la covariance donne :

$$\begin{aligned}
Cov(X,Y) &= \int\int_{\mathbb{R}^2} (x-2)(y-1)g(x,y)dxdy \\
&= \int\int_{\mathbb{R}^2} (x-2)(y-1)e^{-x}dxdy \\
&= \int_0^{+\infty} (x-2)e^{-x}dx \int_0^x (y-1)dy \\
&= \int_0^{+\infty} \left(\frac{x^3}{2} - 2x^2 + 2x\right) e^{-x}dx \\
&= 3 - 4 + 2 \\
&= 1
\end{aligned}$$

D'où le coefficient de corrélation

$$\rho_{XY} = \frac{1}{\sqrt{1}\sqrt{2}} = \frac{\sqrt{2}}{2}$$

À votre avis, les valeurs positives obtenues pour σ_{XY} et ρ_{XY} confirment-elles l'évolution des distributions conditionnelles ?

7 Variables aléatoires dans $\mathbb{R}^n$

Ce qui a été dit précédemment pour les variables aléatoires à valeurs dans $\mathbb{R}^2$ peut aisément être généralisé à $\mathbb{R}^n$:

Définition 6.3 *Soit un espace de probabilité $\{\Omega, \mathcal{A}, \mathbb{P}\}$ associé à une expérience aléatoire $\mathcal{E}$.*
On appelle variable aléatoire à valeurs dans $\mathbb{R}^n$ une application, notée $(X_1, X_2, \ldots, X_n)$, de Ω dans $\mathbb{R}^n$, qui permet d'associer à tout événement élémentaire ω un élément $(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$ de $\mathbb{R}^n$ et qui possède la propriété :
pour tout ensemble d'intervalles $I_1, I_2, \ldots, I_n$ de $\mathbb{R}$,

$$\{\omega \text{ tel que } [X_1(\omega) \in I_1 \cap X_2(\omega) \in I_2 \cdots \cap X_n(\omega) \in I_n]\} \in \mathcal{A}$$

Une variable aléatoire à valeur dans $\mathbb{R}^n$ ou, ce qui revient au même, un ensemble de n variables aléatoires à valeurs réelles, permet de traduire toute observation par un n-uplet de valeurs numériques. La définition ci-dessus implique que l'on puisse attribuer une probabilité à tout pavé $I_1 \times I_2 \times \cdots \times I_n$. On distinguera les ensembles de variables discrètes et les ensembles de variables à densité.

Exemple 6.19 *Si (X, Y, Z) est un triplet de variables discrètes, la loi de probabilité du triplet peut être définie par :*

$$\mathbb{P} : \left\{ \begin{array}{lcl} E_1 \times E_2 \times E_3 & \rightarrow & [0, 1] \\ (x_i, y_j, z_k) & \mapsto & \mathbb{P}(X = x_i \cap Y = y_j \cap Z = z_k) = P_{ijk} \end{array} \right.$$

Avec la relation $\sum_i \sum_j \sum_k P_{ijk} = 1$
On déduit de la loi d'un triplet différentes lois marginales. Par exemple la loi de X est

$$\mathbb{P}(X = x_i) = \sum_j \sum_k P_{ijk} = P_{i..}$$

$$\mathbb{P}(X = x_i \cap Y = y_j) = \sum_k P_{ijk} = P_{ij.}$$

Exemple 6.20 *On dit qu'un triplet (X, Y, Z) de variables aléatoires est un triplet à densité s'il existe une fonction numérique de trois variables réelles, h, possédant la propriété : quelque soit le domaine D de $\mathbb{R}^3$, on a :*

$$\mathbb{P}[(X, Y, Z) \in \mathbb{R}^3] = \int\int\int_D h(x, y, z) dx dy dz$$

h, densité du triplet, vérifie :

1. *$h(x, y, z) \geq 0$*
2. *$\int\int\int_{\mathbb{R}^3} h(x, y, z) dx dy dz = 1$*

On peut, à partir de h obtenir différentes densités marginales. Par exemple la loi de X est

$$f_X(x) = \int\int_{\mathbb{R}^2} h(x, y, z) dy dz$$

Pour la loi du couple (X, Y), on a :

$$g(x, y) = \int_{\mathbb{R}} h(x, y, z) dz$$

7.1 *Indépendance*

On dit que les variables $(X_1, X_2, \ldots, X_n)$ sont indépendantes entre elles si, pour tout ensemble $I_1, I_2, \ldots, I_n$ de $\mathbb{R}$, on a :

$$\mathbb{P}(X_1 \in I_1 \cap X_2 \in I_2 \cap \cdots \cap X_n \in I_n) = \mathbb{P}(X_1 \in I_1)\mathbb{P}(X_2 \in I_2) \ldots \mathbb{P}(X_n \in I_n)$$

On notera que l'indépendance des variables entre elles entraîne leur indépendance deux à deux (la réciproque n'est pas vraie) ainsi que l'indépendance entre elles des variables $\phi_1(X_1), \phi_2(X_2), \ldots, \phi_n(X_n)$, où ϕ est une "bonne" fonction.
Pour le triplet de l'exemple 6.19, l'indépendance des trois variables s'écrit :

$$P_{ijk} = P_{i..}P_{.j.}P_{..k} \;\; \forall(i, j, k)$$

Pour le triplet de l'exemple 6.20, l'indépendance des trois variables s'écrit :

$$h(x, y, z) = f_X(x)f_Y(y)f_Z(z) \;\; \forall(x, y, z)$$

8 Exercices

1. Soit (X, Y) un couple de variables aléatoires à valeurs réelles, de densité de probabilité g définie par :

$$g : \begin{cases} (x, y) \mapsto kxy & , \text{ si } (x, y) \in D \\ (x, y) \mapsto 0 & , \text{ si } (x, y) \notin D \end{cases}$$

   où $D = \{(x, y) \in \mathbb{R}^2 \text{ tel que } x \geq 0,\, y \geq 0, x^2 + y^2 \leq a^2 \ (a \in \mathbb{R}_+^*)\}$.
   (a) Calculer k.
   (b) calculer la probabilité de l'événement "$X + Y < t$" où $0 \leq t \leq a$.
   (c) déterminer les densités de probabilité de X et de Y. Ces variables sont-elles indépendantes ?
   (d) déterminer la densité conditionnelle de Y, lorsque $X = x$, et tracer la courbe de régression de l'espérance conditionnelle correspondante.
2. Soit un couple de variables aléatoires à valeurs réelles de densité de probabilité :

$$g : \begin{cases} (x, y) \mapsto \frac{1}{4}[1 + xy(x^2 - y^2)] & , \text{ si } (x, y) \in D \\ (x, y) \mapsto 0 & , \text{ si } (x, y) \notin D \end{cases}$$

   où $D = \{(x, y) \in \mathbb{R}^2 \text{ tel que } -1 \leq x \leq 1, -1 \leq y \leq 1\}$.
   - Vérifier que g possède les propriétés d'une densité de probabilité.
   - Calculer la probabilité de l'événement :"$(0 \leq x \leq 1/2) \bigcap (0 \leq y \leq 1)$"
   - Déterminer la densité de probabilité de chacune des deux variables aléatoires ; sont-elles indépendantes ?
   - Calculer les espérances conditionnelles E($Y|X = x$) et E($X|Y = y$) ; tracer les courbes de régression correspondantes.
   - Calculer la covariance et le coefficient de corrélation des deux variables.
3. Une expérience aléatoire ϵ peut avoir trois résultats possibles formant un système complet d'événements : E_1 de probabilité p_1, E_2 de probabilité p_2, E_3 de probabilité p_3. On veut effectuer n expériences, indépendantes et identiques à $\mathcal{E}$ et on considère les variables X_1, X_2 et X_3 représentant respectivement le nombre d'événements E_1, E_2 et E_3 que l'on peut observer.
   - Montrer que la loi de probabilité du triplet est donnée par :

$$\mathbb{P}(X_1 = k, X_2 = l, X_3 = m) = \frac{n!}{k!l!m!} p_1^k p_2^l p_3^m$$

     avec $k, l, m \in \{0, 1, 2, \ldots, n\}$ et $k + l + m = n$
   - Généraliser à plus de trois variables (loi multinomiale).
   - En déduire les lois de probabilité de X_1, X_2 et X_3. Ces variables sont-elles indépendantes ?
   - Appliquer à l'exemple suivant : on croise des mufliers ivoires et des mufliers rouges. On obtient en F1 des mufliers "pâles". On admet que la coloration de la fleur est gérée par un couple d'allèles. On observe en F2, après autofécondation, 100 descendants.

4. On considère deux variables aléatoires réelles X et Y indépendantes et de densités de probabilité respectives :

$$f_X : \begin{cases} x \mapsto 0 & , \text{ si } x < 0 \\ x \mapsto \lambda e^{-\lambda x} & , \text{ si } x \geq 0 \end{cases}$$

et

$$f_Y : \begin{cases} y \mapsto 0 & , \text{ si } y < 0 \\ y \mapsto \mu e^{-\mu y} & , \text{ si } y \geq 0 \end{cases}$$

où λ et μ sont deux paramètres réels strictement positifs.

Le couple (X, Y) est un couple d'observations que l'on peut effectuer sur un individu choisi au hasard dans une population donnée. Les paramètres λ et μ varient d'un individu à l'autre, mais on sait que leur rapport reste constant. On désire tester l'hypothèse $H_0 : \mu = \lambda$ contre l'hypothèse $H_1 : \mu = 9\lambda$.

- Calculer la probabilité d'observer sur un individu un couple (x, y) tel que $y \leq x$ sous H_0 et sous H_1.
- Utiliser ces résultats pour construire, au niveau 6%, un test de H_0 contre H_1, à partir d'un échantillon de 10 couples indépendants. Calculer la puissance de ce test.

5. Paul et Valérie ont rendez-vous chez Robert, entre 12h et 14h. Par hypothèse, les instants d'arrivée de Paul et Valérie sont des variables aléatoires X et Y indépendantes, de distribution uniforme sur $[0, 2]$, l'instant zéro correspondant à midi, l'unité de temps étant l'heure.

   (a) Soit U la variable aléatoire représentant le temps d'attente de Robert jusqu'à la première arrivée. Déterminer la densité de probabilité de U.

   (b) Soit V la variable aléatoire représentant le temps d'attente de Robert jusqu'à ce que ses deux amis soient arrivés. Déterminer la densité de probabilité de V.

   (c) Soit la variable $Z = X - Y$. Calculer l'espérance et la variance de Z. Déterminer la densité de probabilité de Z.

   (d) Soit W la variable aléatoire représentant le temps d'attente de Robert entre les deux arrivées. Déterminer la densité de probabilité de W.

9 Solutions

1. • g doit satisfaire aux deux conditions :
   (a) $g(x,y) \geq 0, \forall (x,y) \in \mathbb{R}^2$.
   Or $g(x,y) = 0$ si $(x,y) \notin D$ et $g(x,y) \geq 0$ si $(x,y) \in D$
   $\Rightarrow k \geq 0$
   (b) $\int\int_{\mathbb{R}^2} g(x,y)\mathrm{d}x\mathrm{d}y = 1$.
   $\Rightarrow I = \int\int_D kxy\mathrm{d}x\mathrm{d}y = 1$
   On passe en coordonnées polaires pour calculer I
   $I = k\int\int_\Delta r^2 \cos\theta \sin\theta r\mathrm{d}r\mathrm{d}\theta$ avec Δ défini par : $0 \leq \theta \leq \frac{\pi}{2}$ et $0 \leq r \leq R$
   D'où $I = k\int_0^{\frac{\pi}{2}} \cos\theta\sin\theta\mathrm{d}\theta \int_0^R r^3\mathrm{d}r = k\left[\frac{sin^2\theta}{2}\right]_0^{\frac{\pi}{2}} \frac{R^4}{4} = k\frac{R^4}{8}$

   $I = 1 \Rightarrow k = \frac{8}{R^4}$

   • On a $\mathbb{P}(X+Y<t) = \int\int_{x+y<t} g(x,y)\mathrm{d}x\mathrm{d}y = 0$ si $t \leq 0$

   si $t > 0$ $\mathbb{P}(X+Y<t) = \int\int_{D_1} \frac{8}{R^4}xy\mathrm{d}x\mathrm{d}y$
   D_1 étant défini par $x \geq 0$, $y \geq 0$, $0 \leq x+y < t \leq R$
   On note que la condition $t \leq R$ entraîne que D_1 est un triangle.
   On a :

$$
\begin{aligned}
\frac{8}{R^4}\int\int_{D_1} xy\mathrm{d}x\mathrm{d}y &= \frac{8}{R^4}\int_0^t x\mathrm{d}x \int_0^{t-x} y\mathrm{d}y \\
&= \frac{8}{R^4}\int_0^t x\frac{(t-x)^2}{2}\mathrm{d}x \\
&= \frac{4}{R^4}\int_0^t (t^2x - 2tx^2 + x^3)\mathrm{d}x \\
&= \frac{4}{R^4}\left[t^2x^2 - 2t\frac{x^3}{3} + \frac{x^4}{4}\right]_0 t \\
&= \frac{4t^4}{R^4}\frac{1}{12} \\
&= \frac{t^4}{3R^4}
\end{aligned}
$$

   Finalement $\mathbb{P}(X+Y<t) = 0$ si $t \leq 0$; $\mathbb{P}(X+Y<t) = \frac{t^4}{3R^4}$ si $t > 0$

   • Soient f_1 la densité de X et f_2 la densité de Y. On a :

$$
\begin{aligned}
f_1(x) &= \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}x\mathrm{d}y \\
&= 0 \text{ si } x < 0 \text{ ou si } x > R \\
\\
&= \frac{8}{R^4}\int_0^{\sqrt{R^2-x^2}} xy\mathrm{d}y
\end{aligned}
$$

$$\begin{aligned} &= \frac{8x}{R^4}\left[\frac{y^2}{2}\right]_0^{\sqrt{R^2-x^2}} \\ &= \frac{4x}{R^4}(R^2-x^2) \text{ si } 0 \le x \le R \end{aligned}$$

Finalement :

$$f_1 : \begin{cases} x \mapsto 0 & , \text{ si } x < 0 \text{ ou si } x > R \\ x \mapsto \frac{4x}{R^4}(R^2-x^2) & , \text{ si } 0 \le x \le R \end{cases}$$

De même, X et Y jouant un rôle symétrique :

$$f_2 : \begin{cases} y \mapsto 0 & , \text{ si } y < 0 \text{ ou si } y > R \\ y \mapsto \frac{4y}{R^4}(R^2-y^2) & , \text{ si } 0 \le y \le R \end{cases}$$

Les deux variables ne sont pas indépendantes car on n'a pas l'égalité :

$$g(x,y) = f_1(x)f_2(y)\forall(x,y) \in \mathbb{R}^2$$

2. • g doit satisfaire aux deux conditions :
   (a) $g(x,y) \ge 0$. On a $g(x,y) = 0$ pour tout $(x,y) =\notin D$.
   Il reste à vérifier que sur D, $g(x,y) \ge 0$.
   En passant en coordonnées polaires, on a :

$$\begin{aligned} xy(x^2-y^2) &= r^4\cos\theta\sin\theta(\cos^2\theta - \sin^2\theta) \\ &= \frac{r^4}{2}\sin(2\theta)\cos(2\theta) \\ &= \frac{r^4}{4}\sin(4\theta) \end{aligned}$$

   $|\sin(4\theta)| \le 1 \Rightarrow |xy(x^2-y^2)| \le \frac{r^4}{4}$.
   Or $r^2 \le 2$. Donc $|xy(x^2-y^2)| \le 1$ et par conséquent $1 + xy(x^2-y^2) \ge 0$
   (b) $\int\int_{\mathbb{R}^2} g(x,y)\mathrm{d}x\mathrm{d}y = 1$.
   Or

$$\begin{aligned} \int\int_{\mathbb{R}^2} g(x,y)\mathrm{d}x\mathrm{d}y &= \int\int_D \frac{1}{4}(1+x^3y-xy^3)\mathrm{d}x\mathrm{d}y \\ &= \frac{1}{4}\int_{-1}^{1}\mathrm{d}x\int_{-1}^{1}(1+x^3y-xy^3)\mathrm{d}y \\ &= \frac{1}{4}\int_{-1}^{1}\mathrm{d}x\left[y+x^3\frac{y^2}{2}-x\frac{y^4}{4}\right]_{-1}^{1} \\ &= \frac{1}{4}\int_{-1}^{1}2\mathrm{d}x \\ &= \frac{1}{2}[x]_{-1}^{1} \\ &= 1 \end{aligned}$$

- Soit Δ le domaine. On a

$$\begin{aligned}
\mathrm{P}((X,Y)\in\Delta) &= \int\int_{\Delta} g(x,y)\mathrm{d}x\mathrm{d}y \\
&= \frac{1}{4}\int_0^{1/2}\mathrm{d}x\int_0^1 (1+x^3y-xy^3)\mathrm{d}y \\
&= \frac{1}{4}\int_{-1}^{1/2}(1+{}^{x^3}\!/_3-{}^{x}\!/_4)\mathrm{d}x \\
&= \frac{1}{4}\left[x+\frac{x^4}{8}-\frac{x^2}{8}\right]_0^{1/2} \\
&= \frac{1}{4}\left[\frac{1}{2}+\frac{1}{128}-\frac{1}{32}\right] \\
&= \frac{61}{512} \\
&\approx 11.91\%
\end{aligned}$$

- Soient f_1 la densité de X et f_2 la densité de Y. On a :

$$\begin{aligned}
f_1(x) &= \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}y \\
&= 0 \text{ si } x<-1 \text{ ou si } x>1
\end{aligned}$$

$$\begin{aligned}
f_1(x) &= \frac{1}{4}\int_{-1}^{1}(1+x^3y-xy^3)\mathrm{d}y \\
&= \frac{1}{4}\left[y+x^3\frac{y^2}{2}-x\frac{y^4}{4}\right]_{-1}^{1} \\
&= \frac{1}{2} \text{ si } -1\le x\le 1
\end{aligned}$$

$$\begin{aligned}
f_2(x) &= \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}x \\
&= 0 \text{ si } y<-1 \text{ ou si } y>1
\end{aligned}$$

$$\begin{aligned}
f_2(x) &= \frac{1}{4}\int_{-1}^{1}(1+x^3y-xy^3)\mathrm{d}x \\
&= \frac{1}{4}\left[x+y\frac{x^4}{4}-y^3\frac{x^2}{2}\right]_{-1}^{1} \\
&= \frac{1}{2} \text{ si } -1\le y\le 1
\end{aligned}$$

Les deux variables X et Y ont donc même loi : il s'agit de la loi uniforme sur $[-1;1]$. Il est clair que les deux variables ne sont pas indépendantes puisqu'on n'a pas l'égalité $g(x,y)=f_1(x)f_2(y)\ \forall(x,y)\in\mathbb{R}^2$

- Soit h_X la densité conditionnelle de Y lorsque $X = x$. On a : $h_x(y) = \frac{g(x,y)}{f_1(x)}$. Cette expression n'a de sens que si $f_1(x) \neq 0$, c'est-à-dire si $-1 \leq x \leq 1$. D'autre part, pour x donné entre -1 et 1, la densité g est nulle si $y < -1$ ou si $y > 1$. D'où, pour $-1 \leq x \leq 1$, h_x :

$$h_x : \begin{cases} y \mapsto 0 & , \text{ si } y < -1 \text{ ou } y > 1 \\ y \mapsto \frac{1}{2}(1 + x^3 y - x y^3) & , \text{ si } -1 \leq y \leq 1 \end{cases}$$

On en déduit l'espérance conditionnelle de Y :

$$\begin{aligned} \mathrm{E}\left(Y|X = x\right) &= \int_{-\infty}^{+\infty} y h_x(y) \mathrm{d}y \\ &= \frac{1}{2} \int_{-1}^{1} (y + x^3 y^2 - x y^4) \mathrm{d}y \\ &= \frac{1}{3} x^3 - \frac{1}{5} x \text{ pour } -1 \leq x \leq 1 \end{aligned}$$

La courbe de régression correspondante est la courbe représentative de la fonction :

$$\phi : \begin{cases} [-1;1] & \mapsto \mathbb{R} \\ x & \mapsto \frac{1}{3}x^3 - \frac{1}{5}x \end{cases}$$

Pour h_y, densité conditionnelle de X lorsque $Y = y$, on obtient, pour $-1 \leq y \leq 1$:

$$h_y : \begin{cases} x \mapsto 0 & , \text{ si } x < -1 \text{ ou } x > 1 \\ x \mapsto \frac{1}{2}(1 + x^3 y - x y^3) & , \text{ si } -1 \leq x \leq 1 \end{cases}$$

On en déduit l'espérance conditionnelle :

$$\begin{aligned} \mathrm{E}\left(X|Y = y\right) &= \int_{-\infty}^{+\infty} x h_y(x) \mathrm{d}x \\ &= \frac{1}{2} \int_{-1}^{1} (x + x^4 y - x^2 y^3) \mathrm{d}x \\ &= \frac{1}{5} y - \frac{1}{3} y^3 \text{ pour } -1 \leq y \leq 1 \end{aligned}$$

La courbe de régression correspondante est la courbe représentative de la fonction :

$$\psi : \begin{cases} [-1;1] & \mapsto \mathbb{R} \\ y & \mapsto \frac{1}{5}y - \frac{1}{3}y^3 \end{cases}$$

- Dans cet exemple, X et Y ont même loi. On a donc :
E(X)=E(Y)=$\int_{-\infty}^{+\infty} x f_1(x) \mathrm{d}x = \int_{-1}^{1} \frac{x}{2} \mathrm{d}x = \left[\frac{x^2}{4}\right]_{-1}^{1} = 0$

  V(X)=V(Y)=E(X^2) puisque E(X)=0
et $\int_{-\infty}^{+\infty} x^2 f_1(x) \mathrm{d}x = \int_{-1}^{1} \frac{x^2}{2} \mathrm{d}x = \left[\frac{x^3}{6}\right]_{-1}^{1} = \frac{1}{3}$

Il en résulte que

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \int\int_{\mathbb{R}^2}(x-\mu_x)(y-\mu_y)g(x,y)\mathrm{d}x\mathrm{d}y \\
&= \frac{1}{4}\int\int_D xy(1+x^3y-xy^3)\mathrm{d}x\mathrm{d}y \\
&= \frac{1}{4}\int_{-1}^{1}\mathrm{d}x\left[x\frac{y^2}{2}+x^4\frac{y^3}{3}-x^2\frac{y^5}{5}\right]_{-1}^{1} \\
&= \frac{1}{4}\int_{-1}^{1}\left[\frac{2}{3}x^4-\frac{2}{5}x^2\right]\mathrm{d}x \\
&= \frac{1}{2}\left[\frac{x^5}{15}-\frac{x^3}{15}\right]_{-1}^{1} \\
&= 0
\end{aligned}$$

La covariance de X et Y étant nulle, le coefficient de corrélation est aussi nul. C'est une illustration du fait que l'annulation de ces coefficients ne signifie pas forcément l'indépendance des variables.

3. (a) On peut numéroter de 1 à n les n expériences et ainsi représenter tout résultat possible de cet en ensemble d'expériences par une suite ordonnée de n événements. L'événement $A =$ "$X_1 = k \bigcap X_2 = l \bigcap X_3 = m$" est réalisé chaque fois que l'on observe une suite d'événements comportant k fois E_1, l fois E_2 et m fois E_3 avec $k+l+m=n$

   Or, les expériences étant identiques et indépendantes, la probabilité d'une telle suite est obtenue par le produit de k termes égaux à P_1, l termes égaux à P_2, m termes égaux à P_3, c'est-à-dire $p_1^k p_2^l p_3^m$.

   De plus il a autant de suites de ce type que de manière d'ordonner les k événements de E_1, les l événements de E_2 et les $m = n-k-l$ événements de E_3. On peut choisir la place des k événements de E_1 de C_n^k façons différentes ; pour chacune d'elles on a C_{n-k}^l façons différentes de choisir la place des l événements E_2, la place des $m = n-k-l$ événements E_3 est alors imposée. Il y a donc :
   $C_n^k C_{n-k}^l = \frac{n!}{k!(n-k)!}\frac{(n-k)!}{l!(n-k-l)!} = \frac{n!}{k!l!m!}$ manières, mutuellement incompatibles, de réaliser A.

   L'axiome des probabilités totales permet d'écrire : $\mathbb{P}(X_1 = k, X_2 = l, X_3 = m) = \frac{n!}{k!l!m!}p_1^k p_2^l p_3^m$ avec $k,l,m \in \{0,1,2,\ldots,n\}$ et $k+l+m=n$ et où $p_1+p_2+p_3=1$

   La loi du triplet de variables discrètes (X_1, X_2, X_3) est alors dite "loi trinomiale". La somme des probabilités de tous les triplets "k,l,m" observables est égale au développement du trinôme $(p_1+p_2+p_3)^n = 1$

   (b) On généralise aisément au cas où les résultats possibles d'une expérience sont en nombre i, à savoir $E_1, E_2, \ldots, E_i$ avec les probabilités $p_1, p_2, \ldots, p_i$ $(p_1+p_2+\cdots+p_i=1)$

On désigne par $X_1, X_2, \ldots, X_i$ le nombre d'événements $E_1, E_2, \ldots, E_i$ que l'on peut observer en effectuant n expériences identiques et indépendantes

L'événement $A =$ "$X_1 = k \bigcap X_2 = l \bigcap \cdots \bigcap X_i = r$" est réalisé chaque fois que l'on observe une suite d'événements comportant k fois E_1, l fois E_2, ... , r fois E_i à savoir :

$C_n^k C_{n-k}^l \cdots C_r^r = \frac{n!}{k!(n-k)!} \frac{(n-k)!}{l!(n-k-l)!} \cdots \frac{r!}{r!} = \frac{n!}{k!l!\ldots r!}$

Chaque éventualité à la probabilité $p_1^k p_2^l \ldots p_i^r$.

Et finalement :

$\mathbb{P}(X_1 = k \bigcap X_2 = l \bigcap \cdots \bigcap X_i = r) = \frac{n!}{k!l!\ldots r!} p_1^k p_2^l \ldots p_i^r$ avec $k, l, \ldots, r \in \{0, 1, 2, \ldots, n\}$ et $k + l + \ldots = r = n$ et où $p_1 + p_2 + \cdots + p_i = 1$

Une telle loi est dite loi multinomiale ; la loi binomiale ($i = 2$) et trinomiale ($i = 3$) en sont des cas particuliers.

(c) L'événement "$X_1 = l$" est réalisé chaque fois que l'on a : "$X_1 = k \bigcap X_2 = l \bigcap X_3 = m$" avec $l + m = n - k$ ce que l'on peut écrire "$X_1 = k \bigcap X_2 = l \bigcap X_3 = n - k - l$" avec $k \in \{0, 1, 2, \ldots, n\}$. L'axiome des probabilités totales permet d'écrire :

$$\begin{aligned}
\mathrm{P}(X_1 = k) &= \sum_{l=0}^{n-k} \frac{n!}{k!l!m!(n-k-l)!} p_1^k p_2^l p_3^{n-k-l} \\
&= \frac{n!}{k!(n-k)!} p_1^k \sum_{l=0}^{n-k} \frac{(n-k)!}{l!(n-k-l)!} p_2^l p_3^{n-k-l} \\
&= \frac{n!}{k!(n-k)!} p_1^k (p_2 + p_3)^{n-k}
\end{aligned}$$

Comme $p_2 + p_3 = 1 - p_1$, on en déduit que X_1 est une variable binomiale $\mathcal{B}(n, p_1)$. De même $X_2 \sim \mathcal{B}(n, p_2)$ et $X_3 \sim \mathcal{B}(n, p_3)$. Les trois variables ne sont pas indépendantes puisque :

$$\mathrm{P}(X_1 = k \bigcap X_2 = l \bigcap X_3 = m) \neq \mathrm{P}(X_1 = k)\mathrm{P}(X_2 = l)\mathrm{P}(X_3 = m)$$

(d) L'expérience ϵ est ici l'observation d'un individu de la F2. Soient p_1, p_2, p_3 les probabilités d'apparition des phénotypes "ivoire", "pâle" ou "rouge" sur un individu de la F2. On a, ici, $p_1 = 1/4$, $p_2 = 1/2$, et $p_3 = 1/4$.

Soient X_1 le nombre de plantes à fleurs ivoires, X_2 le nombre de plantes à fleurs pâles et X_3 le nombre de plantes à fleurs rouges que l'on peut observer sur 100 plantes de la F2. On a :

$$\mathrm{P}(X_1 = k \bigcap X_2 = l \bigcap X_3 = m) = \frac{100!}{k!l!m!} \left(\frac{1}{4}\right)^k \left(\frac{1}{2}\right)^l \left(\frac{1}{4}\right)^m$$

où $k, l, m \in \{0, 1, 2, \ldots, 100\}$ et $k + l + m = 100$

Si on s'intéresse au nombre X_3 de plantes à fleurs rouges, on a $X_3 \sim \mathcal{B}\left(100, \frac{1}{4}\right)$

4. (a) Le couple (X, Y) a pour densité

$$\begin{cases} g(x,y) = \lambda\mu e^{-\lambda x} e^{-\mu y} & \text{, si} \quad x \geq 0 \text{ et } y \geq 0 \\ g(x,y) = 0 & \text{, si} \quad x < 0 \text{ ou } y < 0 \end{cases}$$

On a $\mathbb{P}(Y \leq X) = \int\int_\Delta \lambda\mu e^{-\lambda x} e^{-\mu y} \mathrm{d}x\mathrm{d}y$
où Δ est défini par $x \geq 0$, $y \geq 0$, $y \leq x$
On en déduit

$$\begin{aligned} \mathrm{P}(Y \leq X) &= \int_0^{+\infty} \lambda e^{-\lambda x} \mathrm{d}x \int_0^x \mu e^{-\mu y} \mathrm{d}y \\ &= \int_0^{+\infty} \lambda e^{-\lambda x} \mathrm{d}x \left[-e^{-\mu y}\right]_0^x \\ &= \int_0^{+\infty} \lambda e^{-\lambda x}(1 - e^{-\mu x}) \mathrm{d}x \\ &= \int_0^{+\infty} \lambda e^{-\lambda x} - \int_0^{+\infty} \lambda e^{-(\lambda+\mu)x} \\ &= \left[-e^{-\lambda x} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)x}\right]_0^{+\infty} \\ &= 1 - \frac{\lambda}{\lambda+\mu} \end{aligned}$$

Sous H_0 $\mu = \lambda$: $\mathbb{P}(Y \leq X) = p_0 = 1 - \frac{\lambda}{2\lambda} = \frac{1}{2}$
Sous H_1 $\mu = 9\lambda$: $\mathbb{P}(Y \leq X) = p_1 = 1 - \frac{\lambda}{10\lambda} = \frac{9}{10}$

(b) Pour chaque couple d'observations, deux résultats sont possibles : $(Y \leq X)$, de probabilité P ou $(Y > X)$ de probabilité $1 - P$. Soit Z la variable associant à 10 couples d'observations indépendants le nombre de résultats $(Y \leq X)$.
On a $X \sim \mathcal{B}(10, P)$
Sous H_0, on a $Z \sim \mathcal{B}(10, 1/2)$. Cette distribution est en cloche symétrique par rapport à la valeur 5.
Sous H_1, on a $Z \sim \mathcal{B}(10, 9/10)$. Cette distribution est très dissymétrique et a pour module la valeur 9.
Il est clair que c'est l'observation d'une valeur élevée de Z qui conduit à choisir H_1.
On construit donc un test de H_0 contre H_1, unilatéral, avec domaine de rejet à droite.
On a :
$\mathbb{P}_{H_0}(Z = 10) = \left(\frac{1}{2}\right)^{10} = \frac{1}{1024}$
$\mathbb{P}_{H_0}(Z = 9) = C_{10}^9 \left(\frac{1}{2}\right)^{10} = \frac{10}{1024}$
$\mathbb{P}_{H_0}(Z = 8) = C_{10}^8 \left(\frac{1}{2}\right)^{10} = \frac{45}{1024}$
$\mathbb{P}_{H_0}(Z = 7) = C_{10}^7 \left(\frac{1}{2}\right)^{10} = \frac{120}{1024}$

On en déduit :
$\mathbb{P}_{H_0}(Z \geq 8) = \frac{56}{1024} \approx 5.47\% < 6\%$
$\mathbb{P}_{H_0}(Z \geq 7) = \frac{176}{1024} \approx 17.19\% > 6\%$
D'où le test : au niveau 6%, l'observation d'une valeur de $\{0, 1, 2, 3, 4, 5, 6, 7\}$ conduit à rejeter H_0 pour choisir H_1
La puissance du test correspond à la probabilité de choisir H_1 avec raison. C'est donc :

$$\begin{aligned}
\mathrm{P}_{H_1}(Z \geq 8) &= \mathrm{P}_{H_1}(Z = 8) + \mathrm{P}_{H_1}(Z = 9) + \mathrm{P}_{H_1}(Z = 10) \\
&= C_{10}^8 \left(\frac{9}{10}\right)^8 \left(\frac{1}{10}\right)^2 + C_{10}^9 \left(\frac{9}{10}\right)^8 \left(\frac{1}{10}\right) + \left(\frac{9}{10}\right)^{10} \\
&= \left(\frac{9}{10}\right)^8 \left[\frac{45}{100} + \frac{890}{100} + \frac{81}{100}\right] \\
&\approx 0.4305 \times 2.16 \\
&\approx 93\%
\end{aligned}$$

(a) Le couple (X, Y) a pour densité

$$\begin{cases} g(x, y) = \lambda\mu e^{-\lambda x} e^{-\mu y} & \text{, si } \ x \geq 0 \text{ et } y \geq 0 \\ g(x, y) = 0 & \text{, si } \ x < 0 \text{ ou } y < 0 \end{cases}$$

On a $\mathbb{P}(Y \leq X) = \int\int_\Delta \lambda\mu e^{-\lambda x} e^{-\mu y} \mathrm{d}x\mathrm{d}y$
où Δ est défini par $x \geq 0$, $y \geq 0$, $y \leq x$
On en déduit

$$\begin{aligned}
\mathrm{P}(Y \leq X) &= \int_0^{+\infty} \lambda e^{-\lambda x} \mathrm{d}x \int_0^x \mu e^{-\mu y} \mathrm{d}y \\
&= \int_0^{+\infty} \lambda e^{-\lambda x} \mathrm{d}x \left[-e^{-\mu y}\right]_0^x \\
&= \int_0^{+\infty} \lambda e^{-\lambda x} (1 - e^{-\mu x}) \mathrm{d}x \\
&= \int_0^{+\infty} \lambda e^{-\lambda x} - \int_0^{+\infty} \lambda e^{-(\lambda+\mu)x} \\
&= \left[-e^{-\lambda x} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)x}\right]_0^{+\infty} \\
&= 1 - \frac{\lambda}{\lambda + \mu}
\end{aligned}$$

Sous H_0 $\mu = \lambda : \mathbb{P}(Y \leq X) = p_0 = 1 - \frac{\lambda}{2\lambda} = \frac{1}{2}$
Sous H_1 $\mu = 9\lambda : \mathbb{P}(Y \leq X) = p_1 = 1 - \frac{\lambda}{10\lambda} = \frac{9}{10}$

(b) Pour chaque couple d'observations, deux résultats sont possibles : $(Y \leq X)$, de probabilité P ou $(Y > X)$ de probabilité $1 - P$. Soit Z la variable associant à 10 couples d'observations indépendants le nombre de résultats $(Y \leq X)$.

On a $X \sim \mathcal{B}(10, P)$

Sous H_0, on a $Z \sim \mathcal{B}(10, 1/2)$. Cette distribution est en cloche symétrique par rapport à la valeur 5.

Sous H_1, on a $Z \sim \mathcal{B}(10, 9/10)$. Cette distribution est très dissymétrique et a pour module la valeur 9.

Il est clair que c'est l'observation d'une valeur élevée de Z qui conduit à choisir H_1.

On construit donc un test de H_0 contre H_1, unilatéral, avec domaine de rejet à droite.

On a :

$\mathbb{P}_{H_0}(Z = 10) = \left(\frac{1}{2}\right)^{10} = \frac{1}{1024}$

$\mathbb{P}_{H_0}(Z = 9) = C_{10}^{9}\left(\frac{1}{2}\right)^{10} = \frac{10}{1024}$

$\mathbb{P}_{H_0}(Z = 8) = C_{10}^{8}\left(\frac{1}{2}\right)^{10} = \frac{45}{1024}$

$\mathbb{P}_{H_0}(Z = 7) = C_{10}^{7}\left(\frac{1}{2}\right)^{10} = \frac{120}{1024}$

On en déduit :

$\mathbb{P}_{H_0}(Z \geq 8) = \frac{56}{1024} \approx 5.47\% < 6\%$

$\mathbb{P}_{H_0}(Z \geq 7) = \frac{176}{1024} \approx 17.19\% > 6\%$

D'où le test : au niveau 6%, l'observation d'une valeur de $\{0, 1, 2, 3, 4, 5, 6, 7\}$ conduit à rejeter H_0 pour choisir H_1

La puissance du test correspond à la probabilité de choisir H_1 avec raison. C'est donc :

$$\begin{aligned}
\mathrm{P}_{H_1}(Z \geq 8) &= \mathrm{P}_{H_1}(Z = 8) + \mathrm{P}_{H_1}(Z = 9) + \mathrm{P}_{H_1}(Z = 10) \\
&= C_{10}^{8}\left(\frac{9}{10}\right)^{8}\left(\frac{1}{10}\right)^{2} + C_{10}^{9}\left(\frac{9}{10}\right)^{8}\left(\frac{1}{10}\right) + \left(\frac{9}{10}\right)^{10} \\
&= \left(\frac{9}{10}\right)^{8}\left[\frac{45}{100} + \frac{890}{100} + \frac{81}{100}\right] \\
&\approx 0.4305 \times 2.16 \\
&\approx 93\%
\end{aligned}$$

5. Si T est une variable de loi uniforme sur $[0, 2]$, elle admet pour densité :

$$f_T : \begin{cases} t \mapsto 0 \;, & \text{si } t \notin [0, 2] \\ t \mapsto \frac{1}{2} \;, & \text{si } t \in [0, 2] \end{cases}$$

et pour fonction de répartition

$$F_T : \begin{cases} t \mapsto 0 \;, & \text{si } t < 0 \\ t \mapsto \frac{t}{2} \;, & \text{si } t \in [0, 2] \\ t \mapsto 1 \;, & \text{si } t > 2 \end{cases}$$

(a) Soit F_u la fonction de répartition de U. On peut écrire :

$$\begin{aligned} F_u(t) = \mathbb{P}(U < t) &= 1 - \mathbb{P}(U \geq t) \\ &= 1 - \mathbb{P}(X \geq t \cap Y \geq t) \\ &= 1 - \mathbb{P}(X \geq t)\mathbb{P}(Y \geq t) \text{ (variables indépendantes)} \\ &= 1 - (1 - \mathbb{P}(X < t))(1 - \mathbb{P}(Y < t)) \\ &= 1 - (1 - F_X(t))(1 - F_Y(t)) \end{aligned}$$

X et Y ont même loi que T, on en déduit F_u et f_u :

$$F_U : \begin{cases} t \mapsto 0 & , \text{ si } t < 0 \\ t \mapsto 1 - \left(1 - \frac{t}{2}\right)^2 & , \text{ si } t \in [0, 2] \\ t \mapsto 1 & , \text{ si } t > 2 \end{cases}$$

et donc

$$f_U : \begin{cases} t \mapsto 0 & , \text{ si } t \notin [0, 2] \\ t \mapsto 1 - \frac{t}{2} & , \text{ si } t \in [0, 2] \end{cases}$$

(b) Soit F_V la fonction de répartition de V. On peut écrire :
$F_V(t) = \mathbb{P}(V < t) = \mathbb{P}(X < t \cap Y < t) = \mathbb{P}(X < t)\mathbb{P}(Y < t) = F_X(t)F_Y(t) = F_T^2(t)$.
On en déduit F_V puis la densité f_V :

$$F_V : \begin{cases} t \mapsto 0 & , \text{ si } t < 0 \\ t \mapsto \frac{t^2}{4} & , \text{ si } t \in [0, 2] \\ t \mapsto 1 & , \text{ si } t > 2 \end{cases}$$

et donc

$$f_V : \begin{cases} t \mapsto 0 & , \text{ si } t \notin [0, 2] \\ t \mapsto \frac{t}{2} & , \text{ si } t \in [0, 2] \end{cases}$$

(c) Les variables X et Y ayant même loi que T, on a :

$$\begin{aligned} \mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{E}(T) &= \int_0^2 \frac{t}{2} \mathrm{d}t = \left[\frac{t^2}{4}\right]_0^2 = 1 \\ \mathbb{E}(X^2) = \mathbb{E}(Y^2) = \mathbb{E}(T^2) &= \int_0^2 \frac{t^2}{2} \mathrm{d}t = \left[\frac{t^3}{6}\right]_0^2 = \frac{4}{3} \\ \mathbb{V}(X) = \mathbb{V}(Y) = \mathbb{V}(T) &= \mathbb{E}(T^2) - \mathbb{E}^2(T) = \frac{4}{3} - 1 = \frac{1}{3} \end{aligned}$$

On en déduit

$$\begin{aligned} \mathbb{E}(Z) &= \mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = 0 \text{ (linéarité de l'espérance)} \\ \mathbb{V}(Z) &= \mathbb{V}(X - Y) = \mathbb{V}(X) + \mathbb{V}(Y) = \frac{2}{3} \text{ (indépendance des variables)} \end{aligned}$$

Les variables X et Y étant indépendantes, le couple (X, Y) a pour densité g telle que $g(x, y) = f_X(x) f_Y(y)$. D'où :

$$g : \begin{cases} (x, y) \mapsto \frac{1}{4} & , \text{ si } (x, y) \in [0, 2]^2 \\ (x, y) \mapsto 0 & , \text{ si } (x, y) \notin [0, 2] \times [0, 2] \end{cases}$$

Si F_Z est la fonction de répartition de Z, on a :

$$F_Z(t) = \mathbb{P}(X - Y < t) = \int \int_{y > x - t} g(x, y) \mathrm{d}x \mathrm{d}y$$

Quatre cas sont à envisager :

$t \leq -2$ $F_Z(t) = 0$ car $g(x, y) = 0$ sur le domaine d'intégration

$-2 \leq t \leq 0$ $F_Z(t) = \frac{1}{4} \times \frac{1}{2}(2 + t)^2 = \frac{1}{2} + \frac{t}{2} + \frac{t^2}{8}$

$0 \leq t \leq 2$ $F_Z(t) = \frac{1}{4}\left(4 - \frac{1}{2}(2 - t)^2\right) = \frac{1}{2} + \frac{t}{2} - \frac{t^2}{8}$

$t \geq 2$ $F_Z(t) = 1$

On obtient la densité f_Z par dérivation :

$$f_Z : \begin{cases} t \mapsto 0 & , \text{ si } t < -2 \text{ ou } t > 2 \\ t \mapsto \frac{1}{2} + \frac{t}{4} & , \text{ si } t \in [-2, 0] \\ t \mapsto \frac{1}{2} - \frac{t}{4} & , \text{ si } t \in [0, 2] \end{cases}$$

(d) On a clairement $W = |X - Y|$. Si F_W est la fonction de répartition de W, on a :

$$F_W(t) = \mathbb{P}(W < t) = \mathbb{P}(|Z| < t)$$

- Si $t \leq 0$ $F_W(t) = \mathbb{P}(\emptyset) = 0$
- Si $t > 0$, $F_W(t) = \mathbb{P}(-t < Z < t) = \mathbb{P}(-t \leq Z < t) = F_Z(t) - F_Z(-t)$
- Si $0 \leq t \leq 2$ $F_W(t) = \frac{1}{2} + \frac{t}{2} - \frac{t^2}{8} - \left(\frac{1}{2} - \frac{t}{2} + \frac{t^2}{8}\right) = t - \frac{t^2}{4}$
- Si $t \geq 2$ $F_W(t) = 1$

Par dérivation, on obtient la densité f_W :

$$f_W : \begin{cases} t \mapsto 0 & , \text{ si } t \notin [0, 2] \text{ ou } t > 2 \\ t \mapsto 1 - \frac{t}{2} & , \text{ si } t \in [0, 2] \end{cases}$$

Chapitre 7

Fonctions de plusieurs variables aléatoires

On considère une expérience aléatoire $\mathcal{E}$ à laquelle est associée une variable aléatoire à valeurs dans $\mathbb{R}^n$, c'est-à-dire un ensemble de variables aléatoires à valeurs réelles : $(X_1, X_2, \ldots, X_n)$.
Si, par une fonction numérique ϕ, définie sur l'ensemble des observables, on fait correspondre à une réalisation $(x_1, x_2, \ldots x_n)$ son image $z = \phi(x_1, x_2, \ldots x_n)$, on peut dire que l'expérience a entraîné l'observation de z et considérer cette valeur z comme une valeur particulière d'une variable aléatoire Z. Cette nouvelle variable est dite *fonction des variables aléatoires* $(X_1, X_2, \ldots, X_n)$ et on la représente par $Z = \phi(X_1, X_2, \ldots, X_n)$.

Exemple 7.1 *À l'expérience $\mathcal{E}$ est associé le couple (X, Y). La réalisation de $\mathcal{E}$ entraîne l'observation $(X = 3 \cap Y = 5)$*

	La variable	*a pris la valeur*
somme :	$Z = X + Y$	$z = 8$
moyenne :	$Z = \frac{1}{2}(X + Y)$	$z = 4$
différence :	$Z = X - Y$	$z = -2$
quotient :	$Z = \frac{X}{Y}$	$\frac{3}{5}$

La connaissance de la loi de probabilité de chacune des variables $(x_1, x_2, \ldots x_n)$ ne suffit pas, sauf dans le cas de l'indépendance, pour déterminer la loi de Z. Il est clair en effet que l'on doit tenir compte de la pondération de toutes les associations de valeurs possibles conduisant au même résultat z. C'est donc seulement à partir de la loi de la variable, à valeurs dans $\mathbb{R}^n$, $(x_1, x_2, \ldots x_n)$, que l'on peut déterminer celle de $Z = \phi(X_1, X_2, \ldots, X_n)$.
Le but de ce chapitre est de montrer comment déterminer cette loi, dans les cas les plus usuels, et de donner un certain nombre de résultats généraux.

1 Espérance

Pour faciliter les démonstrations ultérieures, on notera dès maintenant que, comme pour les fonctions d'une variable aléatoire (*cf. chapitre 4*), il n'est pas nécessaire de connaître

la loi de la variable $\phi(X_1, X_2, \ldots, X_n)$, pour calculer son espérance : la loi de la variable $(X_1, X_2, \ldots, X_n)$ à valeurs dans $\mathbb{R}^n$ suffit.
À condition que l'espérance de la variable $Z = \phi(X_1, X_2, \ldots, X_n)$ existe (problèmes possible de convergence) :

- Si les variables $(X_1, X_2, \ldots, X_n)$ sont discrètes, il est facile de montrer l'égalité :

$$\mathbb{E}(Z) = \sum_{i,j,\ldots,l} \phi(x_{1i}, x_{2j}, \ldots, x_{nl}) \mathbb{P}(X_1 = x_{1i} \cap X_2 = x_{2j} \cap \cdots \cap X_n = x_{nl}) \quad (7.1)$$

- Si la fonction $(X_1, X_2, \ldots, X_n)$ admet une densité g, on a, pour les fonctions "raisonnables" ϕ utilisées dans la pratique (fonctions mesurables) :

$$\mathbb{E}(Z) = \int\int \cdots \int_{\mathbb{R}^n} \phi(x_1, x_2, \ldots, x_n) g(x_1, x_2, \ldots, x_n) \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n \quad (7.2)$$

On notera la généralité de ce formalisme. En particulier, pour un couple (X, Y), de densité g, et des densités marginales f_1 (pour X) et f_2 (pour Y), on peut utiliser la fonction $\phi : (x, y) \mapsto x$ qui définit la variable $X = \phi(X, Y)$. On obtient :

$$\begin{aligned}
\mathbb{E}(X) &= \int\int_{\mathbb{R}^2} \phi(x, y) g(x, y) \mathrm{d}x \mathrm{d}y \\
&= \int\int_{\mathbb{R}^2} x g(x, y) \mathrm{d}x \mathrm{d}y \\
&= \int_{-\infty}^{+\infty} x \mathrm{d}x \int_{-\infty}^{+\infty} g(x, y) \mathrm{d}y \\
&= \int_{-\infty}^{+\infty} x f_1(x) \mathrm{d}x
\end{aligned}$$

2 Variable somme

Soit la variable $Z = X + Y$. On se propose de déterminer sa loi de probabilité à partir de celle du couple (X, Y)

2.1 couple de variables discrètes

Dans ce cas, la détermination de la loi de la variable somme est, en général, aisée. Elle s'appuie sur l'axiome des probabilité totales. On peut l'illustrer sur deux exemples :

Exemple 7.2 *(voir exemple 6.1) Soit un couple (X, Y) dont la loi de probabilité est donné par le tableau ci-dessous :*

Y \ X	1	2	3
1	0	1/6	1/6
2	1/6	0	1/6
3	1/6	1/6	0

On a $\mathbb{P}(Z = 2) = \mathbb{P}(Z = 6) = 0$

$$\mathbb{P}(Z = 3) = \mathbb{P}(X = 1 \cap Y = 2) + \mathbb{P}(X = 2 \cap Y = 1) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

$$\mathbb{P}(Z = 4) = \mathbb{P}(X = 1 \cap Y = 3) + \mathbb{P}(X = 3 \cap Y = 1) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

$$\mathbb{P}(Z = 5) = \mathbb{P}(X = 2 \cap Y = 3) + \mathbb{P}(X = 3 \cap Y = 2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

D'où la loi de Z

z_i	4	5	6
$\mathbb{P}(Z = z_i)$	1/3	1/3	1/3

Exemple 7.3 *(voir exemple 6.2) Soit un couple* (X, Y) *dont la loi de probabilités est donné par le tableau ci-dessous :*

Y \ X	1	2	3
1	1/9	1/9	1/9
2	1/9	1/9	1/9
3	1/9	1/9	1/9

Par le même raisonnement que dans l'exemple précédent, on obtient :

z_i	2	3	4	5	6
$\mathbb{P}(Z = z_i)$	1/9	2/9	3/9	2/9	1/9

On remarque que, bien que les variables X et Y aient même loi dans les deux cas (voir chapitre 6), dans un cas il y a indépendance et dans l'autre il y a dépendance, d'où la forme différente de la loi de $X + Y$.

2.2 Couples de variables à densité

Soit g la densité d'un couple (X, Y). Pour déterminer la loi de la variable $Z = X + Y$, la méthode générale est de déterminer sa fonction de répartition sur Z. On a

$$\begin{aligned} F(t) &= \mathbb{P}(Z < t) \\ &= \int\int_D g(x, y)\mathrm{d}x\mathrm{d}y \end{aligned}$$

où D est le domaine défini par $x + y < t$.
Dans la pratique le calcul d'un telle intégrale double est relativement simple.
Par dérivation de F, on obtient la densité f de Z.

Exemple 7.4 *On se propose de déterminer la densité de probabilité de la somme de deux variables aléatoires indépendantes, de même loi exponentielle de paramètre λ ($\lambda \in \mathbb{R}_+^*$) La première variable X a donc pour densité :*

$$f_1 : \begin{cases} x \mapsto 0 & , \textit{ si } x < 0 \\ x \mapsto \lambda e^{-\lambda x} & , \textit{ si } x \geq 0 \end{cases}$$

La deuxième variable Y a pour densité

$$f_2 : \begin{cases} y \mapsto 0 & , \textit{ si } y < 0 \\ y \mapsto \lambda e^{-\lambda y} & , \textit{ si } y \geq 0 \end{cases}$$

À cause de l'indépendance des deux variables, le couple (X, Y) a pour densité :

$$g : \begin{cases} (x, y) \mapsto 0 & , \textit{ si } x < 0 \textit{ ou } y < 0 \\ (x, y) \mapsto \lambda^2 e^{-\lambda(x+y)} & , \textit{ si } x \geq 0 \textit{ et } y \geq 0 \end{cases}$$

Soit $Z = X + Y$, de fonction de répartition F et de densité f. On a :

$$\begin{aligned} F(t) &= \mathbb{P}(Z < t) \\ &= \int\int_D g(x, y) dx dy = 0, \textit{ si } t \leq 0 \\ &= \int\int_\Delta \lambda^2 e^{-\lambda(x+y)} dx dy, \textit{ si } t > 0 \textit{ et } \Delta : \{x \geq 0, y \geq 0, x + y < t\} \\ &= \int_0^t \lambda e^{-\lambda x} dx \int_0^{t-x} \lambda e^{-\lambda y} dy \\ &= \int_0^t \lambda e^{-\lambda x} dx \left[-e^{-\lambda y}\right]_0^{t-x} \\ &= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} \end{aligned}$$

D'où la fonction de répartition de Z :

$$F : \begin{cases} t \mapsto 0 & , \textit{ si } t < 0 \\ t \mapsto 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} & , \textit{ si } t \geq 0 \end{cases}$$

On en déduit la densité de Z :

$$f : \begin{cases} t \mapsto 0 & , \textit{ si } t < 0 \\ t \mapsto \lambda^2 t e^{-\lambda t} & , \textit{ si } t \geq 0 \end{cases}$$

On peut noter qu'il est *possible* de déterminer plus directement la densité f de la variable $X + Y$, à partir de la densité du couple (X, Y).
Dans l'intégrale $F(t) = \int\int_{x+y<t} g(x, y) \mathrm{d}x\mathrm{d}y$, on peut faire le changement de variables :

$$\begin{cases} u = x \\ v = x + y \end{cases} \Leftrightarrow \begin{cases} x = u \\ y = v - u \end{cases} \text{ de Jacobien } \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$

Et donc

$$\begin{aligned} F(t) &= \int\int_{v<t} g(u, v-u)\mathrm{d}u\mathrm{d}v \\ &= \int_{-\infty}^{t} \mathrm{d}v \int_{-\infty}^{+\infty} g(u, v-u)\mathrm{d}u \end{aligned}$$

Or $F(t) = \int_{-\infty}^{t} f(v)\mathrm{d}v$. On en déduit $f(v) = \int_{-\infty}^{+\infty} g(u, v-u)\mathrm{d}u$.
En remarquant que le changement de variables

$$\begin{cases} u = x + y \\ v = y \end{cases}$$

aurait conduit à $f(u) = \int_{-\infty}^{+\infty} g(u-v, v)\mathrm{d}v$, on peut conclure que la densité f de la variable $X + Y$ est donnée par les relations

$$f(t) = \int_{-\infty}^{+\infty} g(x, t-x)\mathrm{d}x = \int_{-\infty}^{+\infty} g(t-y, y)\mathrm{d}y \tag{7.3}$$

Exemple 7.5 *(suite de l'exemple 7.4) On aurait pu écrire directement :*

$$f(t) = \int_{-\infty}^{+\infty} g(x, t-x)dx$$

Or $g(x, t-x) \neq 0$ seulement si $x \geq 0$ et $t - x \geq 0 \Rightarrow 0 \leq x \leq t$. Donc

$$\begin{aligned} &\textit{Pour } t < 0 \quad , \quad f(t) = 0 \\ &\textit{Pour } t \geq 0 \quad , \quad f(t) = \int_0^t \lambda^2 e^{-\lambda(x+t-x)} dx = \lambda^2 e^{-\lambda t} \int_0^t dx = \lambda^2 t e^{-\lambda t} \end{aligned}$$

2.3 *Somme de n variables*

La détermination de la loi de probabilité d'une variable somme de n variables aléatoires ($n > 2$) se fait en suivant les mêmes principes que ceux qui viennent d'être exposés pour un couple. Seul le formalisme devient plus compliqué. On obtient donc le résultat suivant :

Théorème 7.1 *Si les variables aléatoires $X_1, X_2, \ldots, X_n$ ont des espérances, la somme de ces variables a pour espérance la somme des espérances de chacune d'elles.*

On va démontrer ce théorème dans le cas d'un couple (X, Y) de densité g, de densités marginales f_1 et f_2. La démonstration est similaire pour un couple discret.
On peut les démontrer pour n variables, directement ou, plus simplement, par récurrence.

On a, d'après, l'équation 7.2 :

$$\begin{aligned}
\mathbb{E}(X+Y) &= \int\int_{\mathbb{R}^2} (x+y)g(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int\int_{\mathbb{R}^2} xg(x,y)\mathrm{d}x\mathrm{d}y + \int\int_{\mathbb{R}^2} yg(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} x\mathrm{d}x \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}y + \int_{-\infty}^{+\infty} y\mathrm{d}y \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}x \\
&= \int_{-\infty}^{+\infty} xf_1(x)\mathrm{d}x + \int_{-\infty}^{+\infty} yf_2(y)\mathrm{d}y \qquad \text{voir chapitre 6} \\
&= \mathbb{E}(X) + \mathbb{E}(Y)
\end{aligned}$$

On note que ce résultat suppose seulement l'existence des espérances.

Théorème 7.2 *Si les variables aléatoires $X_1, X_2, \dots, X_n$ ont des variances et sont* indépendantes, *la somme de ces variables a pour variance la somme des variances de chacune d'elles.*

On va démontrer ce théorème dans le cas d'une couple (X, Y) de densité g, de densités marginales f_1 et f_2. La démonstration est similaire pour un couple discret.
La généralisation à n variables est facile.
On pose $\mathbb{E}(X) = \mu_1$, $\mathbb{E}(Y) = \mu_2$. On remarque que les espérances existent puisque les variances existent. On a $\mathbb{E}(Z) = \mu_1 + \mu_2$, d'après le théorème 7.1.
D'après l'équation 7.2, on peut écrire :

$$\begin{aligned}
\mathbb{V}(X+Y) &= \mathbb{E}\left((X+Y-\mu_1-\mu_2)^2\right) \\
&= \int\int_{\mathbb{R}^2} (x+y-\mu_1-\mu_2)^2 g(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int\int_{\mathbb{R}^2} (x-\mu_1)^2 g(x,y)\mathrm{d}x\mathrm{d}y + \int\int_{\mathbb{R}^2} (y-\mu_2)^2 g(x,y)\mathrm{d}x\mathrm{d}y \\
&\quad +2\int\int_{\mathbb{R}^2} (x-\mu_1)(y-\mu_2)g(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} (x-\mu_1)^2\mathrm{d}x \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}y + \int_{-\infty}^{+\infty} (y-\mu_2)^2\mathrm{d}y \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}x \\
&\qquad + 2\mathrm{Cov}(X,Y) \\
&= \int_{-\infty}^{+\infty} (x-\mu_1)^2 f_1(x)\mathrm{d}x + \int_{-\infty}^{+\infty} (y-\mu_2)^2 f_2(y)\mathrm{d}y + 2\mathrm{Cov}(X,Y)
\end{aligned}$$

D'où une première relation générale :

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + V(Y) + 2\mathrm{Cov}(X,Y) \tag{7.4}$$

Si les variables sont indépendantes, alors $\mathrm{Cov}(X, Y) = 0$ (voir chapitre 6) et le théorème est démontré : $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.
Par rapport au théorème 7.1, on remarque la nécessité de l'indépendance (ou au moins de l'annulation de la covariance).

Exemple 7.6 *(suite des exemples 7.2 et 7.3) Dans ces deux exemples, les variables X et Y ont même loi (voir chapitre 6 :* $\mathbb{P}(X = i) = \mathbb{P}(Y = j) = 1/3$ *pour* $i, j = 1, 2, 3$*). On en déduit :*

$$\begin{aligned}
\mathbb{E}(X) = \mathbb{E}(Y) &= \frac{1}{3}(1+2+3) = 2 \\
\mathbb{V}(X) = \mathbb{V}(Y) &= \frac{1}{3}(1+0+1) = \frac{2}{3}
\end{aligned}$$

Dans l'exemple 7.2, on a :

$$\begin{aligned}
\mathbb{E}(X+Y) &= \frac{1}{3}(3+4+5) = 4 = \mathbb{E}(X) + \mathbb{E}(Y) \\
\mathbb{V}(X+Y) &= \frac{1}{3}(0+1+0) = \frac{2}{3} \neq \mathbb{V}(X) + \mathbb{V}(Y)
\end{aligned}$$

Le théorème 7.1 est vérifié, mais pas le théorème 7.2, à cause de la dépendance entre les variables.
Dans l'exemple 7.3, on a :

$$\begin{aligned}
\mathbb{E}(X+Y) &= \frac{1}{9}(2+6+12+10+6) = 4 = \mathbb{E}(X) + \mathbb{E}(Y) \\
\mathbb{V}(X+Y) &= = \frac{1}{9}(4+2+0+2+4) = \frac{4}{3} = \mathbb{V}(X) + \mathbb{V}(Y)
\end{aligned}$$

Le théorème 7.1 est toujours vérifié, ainsi que le théorème 7.2 à cause de l'indépendance des deux variables

Exemple 7.7 *(suite de l'exemple 7.4) Les variables X et Y ayant même loi, on a :*

$$\begin{aligned}
\mathbb{E}(X) = \mathbb{E}(Y) &= \int_0^{+\infty} \lambda x e^{-\lambda x} dx = \left[-x e^{-\lambda x}\right]_0^{+\infty} + \frac{1}{\lambda}\int_0^{+\infty} \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \\
\mathbb{E}(X^2) = \mathbb{E}(Y^2) &= \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx = \left[-x^2 e^{-\lambda x}\right]_0^{+\infty} + \frac{2}{\lambda}\int_0^{+\infty} \lambda x e^{-\lambda x} dx = \frac{2}{\lambda^2} \\
\mathbb{V}(X) = \mathbb{V}(Y) &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}
\end{aligned}$$

On en déduit, ce qu'il est facile de vérifier :

$$\begin{aligned}
\mathbb{E}(X+Y) &= \mathbb{E}(X) + \mathbb{E}(Y) = \frac{2}{\lambda} \qquad \textit{(théorème 7.1)} \\
\mathbb{V}(X+Y) &= \mathbb{V}(X) + \mathbb{V}(Y) = \frac{2}{\lambda^2} \qquad \textit{(théorème 7.2, à cause de l'indépendance)}
\end{aligned}$$

Théorème 7.3 *Si les variables $X_1, X_2, \ldots, X_n$ sont des variables* normales *et* indépendantes*, la somme de ces variables est une variable normale, ayant pour espérance la somme des espérances de chacune d'elles et pour variance la somme des variances de chacune d'elles.*

La dernière partie de ce théorème est déjà donnée par les théorèmes 7.1 et 7.2. L'élément nouveau concerne la forme de la loi. Ce théorème peut se démontrer pour deux variables puis, par récurrence, pour n variables.
Pour simplifier l'écriture, on va démontrer ce théorème dans le cas de deux variables normales centrées et réduites.
Si $X \sim \mathcal{N}(0,1)$ et $Y \sim \mathcal{N}(0,1)$, la loi du couple est donné par la densité

$$g : (x,y) \mapsto \frac{1}{2\pi} e^{\left(-\frac{1}{2}(x^2+y^2)\right)}$$

Si f est la densité de $X+Y$, elle est donnée par la relation 7.3 :

$$f(t) = \int_{-\infty}^{+\infty} \frac{1}{2\pi} e^{\left(-\frac{1}{2}(x^2+(t-x)^2)\right)} \mathrm{d}x$$

D'où

$$\begin{aligned}
f(t) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{\left(-\frac{1}{2}(2x^2-2tx+t^2)\right)} \mathrm{d}x \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{\left(-\frac{1}{2}(x\sqrt{2}-\frac{t}{\sqrt{2}})^2+\frac{t^2}{2}\right)} \mathrm{d}x \\
&= \frac{1}{2\pi} e^{\left(-\frac{t^2}{4}\right)} \int_{-\infty}^{+\infty} e^{\left(-\frac{1}{2}(x\sqrt{2}-\frac{t}{\sqrt{2}})^2\right)} \mathrm{d}x
\end{aligned}$$

On change de variable : $x\sqrt{2} - \frac{t}{\sqrt{2}} = u \Rightarrow \mathrm{d}x = \frac{\mathrm{d}u}{\sqrt{2}}$. On obtient :

$$\begin{aligned}
f(t) &= \frac{1}{2\pi} e^{\left(-\frac{t^2}{4}\right)} \int_{-\infty}^{+\infty} e^{\left(-\frac{u^2}{2}\right)} \frac{\mathrm{d}u}{\sqrt{2}} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{\left(-\frac{1}{2}(t/\sqrt{2})^2\right)}
\end{aligned}$$

Car $\int_{-\infty}^{+\infty} \frac{1}{2\pi} e^{\left(-\frac{u^2}{2}\right)} \mathrm{d}x$ On a donc

$$X+Y \sim \mathcal{N}(0,2)$$

Théorème 7.4 *(théorème de la limite centrée)*
Soit une suite $X_1, X_2, \ldots, X_n, \ldots$ de variables aléatoires indépendantes et de même loi. *On désigne par μ leur espérance commune et σ^2 leur variance commune. La loi de probabilité de la variable $S_n = \sum_{i=1}^n X_i$ tend vers une loi normale $\mathcal{N}(n\mu, n\sigma^2)$ lorsque n augmente indéfiniment.*

Il faut noter que $\mathbb{E}(S_n) = n\mu$ et $\mathbb{V}(S_n) = n\sigma^2$ sont des conséquences des théorèmes 7.1 et 7.2.

On admettra ce théorème que l'on utilisera plus loin. Il met en évidence l'importance de la loi normale qui apparaît ainsi comme loi asymptotique dans un certain nombre de modèles probabilistes.
Cette convergence en loi permet, si l'on dispose d'un nombre de variables suffisamment grand, d'utiliser la loi normale comme loi de la variable S_n même si on ignore la loi commune des variables X_i. Le seul problème est de savoir à partir de quand on peut utiliser cette approximation. Un usage est de choisir $n \geq 30$. Il est cependant clair que si la loi des variables X_i s'écarte beaucoup du modèle gaussien, la convergence sera assez lente : l'exemple de la loi binomiale incite à la prudence.

2.4 *Application à la loi binomiale*

Soit une expérience aléatoire $\mathcal{E}$ constitué de n expériences $e_1, e_2, \ldots, e_n$ identiques et indépendantes.
Au cours de chaque expérience e_i on peut observer un événement A de probabilité P, ou son contraire $\overline{A}$ de probabilité $Q = 1 - P$.
Si Z est la variable qui associe à $\mathcal{E}$ le nombre d'événements A que l'on peut observer, on sait que Z suit une loi binomiale $\mathcal{B}(n, P)$.
Supposons maintenant qu'à chaque expérience e_i, on associe une variable X_i qui prend la valeur 1 si A se réalise ou la valeur 0 si A ne se réalise pas. Une telle variable s'appelle l'indicatrice de l'événement A.
On a :

$$\begin{aligned} \mathbb{P}(X_i = 1) = \mathbb{P}(A) &= P \\ \mathbb{P}(X_i = 0) = \mathbb{P}(\overline{A}) = Q = 1 - P \end{aligned}$$

On en déduit :

$$\begin{aligned} \mathbb{E}(X_i) &= 1P + 0Q = P \\ \mathbb{V}(X_i) &= (1-P)^2P + (0-P)^2Q = QP(P+Q) = PQ \end{aligned}$$

La variable Z apparaît alors comme une variable somme de n variables indépendantes et de même loi : $Z = \sum_{i=1}^{n} X_i$
On déduit du théorème 7.1 :

$$\mathbb{E}(Z) = \sum_{i=1}^{n} \mathbb{E}(X_i) = nP$$

On déduit du théorème 7.2 :

$$\mathbb{V}(Z) = \sum_{i=1}^{n} \mathbb{V}(X_i) = nPQ$$

Enfin le théorème de la limite centrée permet d'écrire, pour n suffisamment grand :

$$Z \sim \mathcal{N}(nP, nPQ)$$

2.5 *Généralisation : combinaisons linéaire de variables*

À un ensemble de variables aléatoires $(X_1, X_2, \ldots, X_n)$, à valeurs réelles, on peut plus généralement associer un variable :

$$Z = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n \quad \alpha_i \in \mathbb{R}$$

On supposera, par la suite, l'existence de l'espérance et de la variance de chacune des variables.
Le théorème 7.1 permet d'écrire :

$$\mathbb{E}(\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n = \mathbb{E}(\alpha_1 X_1) + \mathbb{E}(\alpha_2 X_2) + \cdots + \mathbb{E}(\alpha_n X_n)$$

et d'après le théorème 4.2 :

$$\mathbb{E}(\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n = \alpha_1 \mathbb{E}(X_1) + \alpha_2 \mathbb{E}(X_2) + \cdots + \alpha_n \mathbb{E}(X_n) \tag{7.5}$$

Si les variables sont indépendantes, le théorème 7.2 permet d'écrire :

$$\mathbb{V}(\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n = \mathbb{V}(\alpha_1 X_1) + \mathbb{V}(\alpha_2 X_2) + \cdots + \mathbb{V}(\alpha_n X_n)$$

et d'après le théorème 4.2 :

$$\mathbb{V}(\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n = \alpha_1^2 \mathbb{V}(X_1) + \alpha_2^2 \mathbb{V}(X_2) + \cdots + \alpha_n^2 \mathbb{V}(X_n) \tag{7.6}$$

Si enfin les variables $X_1, X_2, \ldots, X_n$ sont des variables normales et indépendantes, les variables $\alpha_1 X_1, \alpha_2 X_2, \ldots, \alpha_n X_n$ sont indépendantes et normales d'après le théorème 4.1.
Le théorème 7.3 permet d'affirmer que la variables $Z = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n$ est encore une variable normale dont l'espérance et la variance sont données par les relation 7.5 et 7.6
Ces quelques remarques ne sont pas exhaustives. On ne doit pas oublier que, quelle que soit la loi d'une variable X_i, on sait déterminer la loi de la variable $\alpha_i X_i$ (multiplication par une constante, voir chapitre 4) ; on peut donc envisager d'autres modèles.

3 Variable différence

Soit un couple (X_1, X_2) et soir la variable $Z = X_1 - X_2$. Une telle variable est utilisée dans certains tests. On peut considérer Z comme une combinaison linéaire de X_1 et X_2 avec les coefficients $+1$ et -1.
On peut donc écrire, en posant $\mathbb{E}(X_1) = \mu_1$, $\mathbb{E}(X_2) = \mu_2$, $\mathbb{V}(X_1) = \sigma_1^2$, $\mathbb{V}(X_2) = \sigma_2^2$ et en appliquant les résultats précédents :

$$\mathbb{E}(X_1 - X_2) = \mathbb{E}(X_1) - \mathbb{E}(X_2) = \mu_1 - \mu_2$$

Dans le cas de l'indépendance :

$$\mathbb{V}(X_1 - X_2) = \mathbb{V}(X_1) + V(X_2) = \sigma_1^2 + \sigma_2^2$$

Dans le cas de variables normales et indépendantes :

$$X_1 - X_2 \sim \mathcal{N}\left(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2\right)$$

4 Variable moyenne

À un ensemble $(X_1, X_2, \ldots, X_n)$ de n variables à valeurs réelles, on peut associer la variable $m = \frac{1}{n}\sum_{i=1}^n X_i$ appelée variable moyenne. En fait cette variable est obtenue en multipliant la variable somme par la constante $\frac{1}{N}$

4.1 Cas d'un n-échantillon

Un cas particulier très important dans la pratique est celui où les n variables sont *indépendantes et de même loi*. On parle alors d'un n-échantillon et on désignera la moyenne par le symbole $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$.
On appellera moyenne empirique la valeur prise par $\overline{X}$ lorsque l'expérience est réalisée ; il s'agit de la moyenne arithmétique des observations $\overline{x} = \frac{1}{n}\sum_{i=1}^n Xx_i$.
La variable $\overline{X}$ peut être considérée comme une combinaison linéaire des observation X_i avec des coefficients tous égaux à $\frac{1}{n}$. Comme les variables X_i ont même loi, on pose pour tout i : $\mathbb{E}(X_i) = \mu_i$ et $\mathbb{V}(X_i) = \sigma_i^2$, et on déduit des formules 7.5 et 7.6 :

$$\begin{cases} \mathbb{E}(\overline{X}) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n}n\mu = \mu \\ \mathbb{V}(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n} \end{cases} \tag{7.7}$$

La variable $\overline{X}$ a donc pour espérance l'espérance d'une variable mais elle est beaucoup moins dispersée.
Il résulte du théorème 7.3 que si les variables X_i suivent une loi $\mathcal{N}(\mu, \sigma^2)$, alors $\overline{X}$ suit une loi $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.
Enfin, même si on ignore la loi suivie par les variables X_i, le théorème central limite (combiné avec le théorème 4.1) permet d'affirmer que la loi asymptotique de la variable $\overline{X}$ est une loi $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$; en fait cette loi est utilisée pratiquement pour $n \geq 30$.
On comprend ainsi l'intérêt de la variable $\overline{X}$ pour construire des tests d'hypothèses concernant le paramètre μ.

4.2 Application a la loi binomiale

On a parlé précédemment de la variable Z, obéissant à une loi $\mathcal{B}(n, P)$, associant à l'expérience $\mathcal{E}$ le nombre d'événements A que l'on peut observer. On a vu que cette variable pouvait être considérée comme la somme de n variables indicatrices X_i.
Soit maintenant la variable Z/n associant à $\mathcal{E}$ la proportion d'événements A que l'on peut observer sur les n résultats. On a :

$$\frac{Z}{n} = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}$$

Cette variable est la moyenne de n variables indépendantes de de même loi. On en déduit

$$\mathbb{E}(\overline{X}) = \mathbb{E}(X_i) = P\,;\;\; \mathbb{V}(\overline{X}) = \frac{\mathbb{V}(X_i)}{n} = \frac{PQ}{n}$$

Enfin, pour n suffisamment grand, on aura

$$\overline{X} \sim \mathcal{N}\left(P, \frac{PQ}{n}\right)$$

Exemple 7.8 *Soit X une variable normale d'espérance μ et de variance μ. On se propose de construire au niveau de 5%, à partir de la moyenne de 10 variables, indépendantes et de même loi que X, un test de l'hypothèse $\mu = 100$ contre l'hypothèse $\mu = 90$ et de calculer la puissance de ce test.*
Soit $\overline{X}$ la moyenne de 10 variables indépendantes de loi $\mathcal{N}(\mu, \mu)$.
On en déduit $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\mu}{10}\right)$
Sous H_0 : $\overline{X} \sim \mathcal{N}(100, 10)$
Sous H_1 : $\overline{X} \sim \mathcal{N}(90, 9)$
Les faibles valeurs de $\overline{X}$ étant plus probables sous H_1 que sous H_0, on construit un test unilatéral avec domaine de rejet de H_0 à gauche. Au niveau 5% les valeurs qui conduisent à rejeter H_0 sont les valeurs de $\overline{X}$ inférieures ou égales à : $100 - 1.645\sqrt{10} \approx 94.8$. D'où le test : si on observe une valeur de $\overline{X}$ supérieure à 94.8 on choisit H_0 ; si on observe une valeur de $\overline{X}$ inférieure ou égales à 94.8, on choisit H_1.
La puissance du test est la probabilité de choisir H_1 avec raison. Soit $Z \sim \mathcal{N}(0, 1)$. On a :

$$\pi = \mathbb{P}_{H_1}(\overline{X} \leq 94.8) = \mathbb{P}\left(Z \leq \frac{94.8 - 90}{\sqrt{9}}\right) = \mathbb{P}(Z \leq 1.60) \approx 94.52\%$$

5 Variable produit

Soit (X, Y) un couple de variables aléatoires à valeurs réelles. La variables $Z = XY$ est la variable produit de X et Y. La détermination de la loi de Z, dans le cas d'un couple discret, ne présente pas plus de difficultés que dans le cas d'un couple de densité g. Si F et f sont respectivement la fonction de répartition et la densité de Z, on a :

$$F(t) = \mathbb{P}(Z < t) = \int\int_D g(x, y)\mathrm{d}x\mathrm{d}y$$

où D est défini par $xy < t$.
On effectue le changement de variables :

$$\left\{ \begin{array}{c} u = x \\ v = xy \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} x = u \\ y = \frac{v}{u} \end{array} \right. \text{ de Jacobien } \left| \begin{array}{cc} 1 & 0 \\ -\frac{v}{u^2} & \frac{1}{u} \end{array} \right| = \frac{1}{u}$$

On en déduit :

$$\begin{aligned} F(t) &= \int\int_{v<t} g\left(u, \frac{v}{u}\right) \left|\frac{1}{u}\right| \mathrm{d}u\mathrm{d}v \\ &= \int_{-\infty}^{t} \mathrm{d}v \int_{-\infty}^{+\infty} \frac{1}{|u|} g(u, \frac{v}{u})\mathrm{d}u \end{aligned}$$

D'où $f(t) = \int_{-\infty}^{+\infty} \frac{1}{|u|} g\left(u, \frac{t}{u}\right) \mathrm{d}u$, que l'on écrit :

$$f(t) = \int_{-\infty}^{+\infty} \frac{1}{|x|} g\left(x, \frac{t}{x}\right) \mathrm{d}x = \int_{-\infty}^{+\infty} \frac{1}{|y|} g\left(y, \frac{t}{y}\right) \mathrm{d}y \tag{7.8}$$

La deuxième égalité provenant du changement de variables : $\left\{ \begin{array}{l} u = xy \\ v = y \end{array} \right.$

Théorème 7.5 *Si les variables aléatoires $X_1, X_2, \ldots, X_n$ admettent des espérance et sont indépendantes, le produit de ces variables a pour espérance le produit des espérances de chacune d'elles.*

On va démontrer ce théorème dans le cas d'un couple (X, Y) de densité g, de densités marginales f_1 et f_2. La démonstration pour un couple discret et la généralisation à n variables est aisée.
On a, d'après 7.2 :

$$\begin{aligned}
\mathbb{E}(XY) &= \int\int_{\mathbb{R}^2} xyg(x,y)\mathrm{d}x\mathrm{d}y \\
&= = \int\int_{\mathbb{R}^2} xyf_1(x)f_2(y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} xf_1(x)\mathrm{d}x \int_{-\infty}^{+\infty} yf_2(y)\mathrm{d}y \\
&= \mathbb{E}(X)\mathbb{E}(Y)
\end{aligned}$$

6 Variable quotient

Soit (X, Y) un couple de variables aléatoires à valeurs réelles. On peut lui associer la variable quotient $Z = \frac{X}{Y}$.
Les lois de Student et de Fisher-Snedecor sont définies à partir de variables quotient.
Comme plus haut, à titre d'entraînement, on va déterminer la densité Z dans le cas d'un couple de densité g. Si F et f sont respectivement la fonction de répartition et la densité de Z, on a :

$$F(t) = \mathbb{P}\left(\frac{X}{Y} < t\right) = \int\int_D g(x,y)\mathrm{d}x\mathrm{d}y$$

où D est défini par $\frac{x}{y} < t$.
On effectue le changement de variables :

$$\left\{ \begin{array}{l} u = \frac{x}{y} \\ v = y \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} x = uv \\ y = v \end{array} \right. \text{ de Jacobien } \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$$

On en déduit :

$$\begin{aligned}
F(t) &= \int\int_{vu<t} g(uv, v)\,|v|\,\mathrm{d}u\mathrm{d}v \\
&= \int_{-\infty}^{t} \mathrm{d}u \int_{-\infty}^{+\infty} |v| g(uv, v)\mathrm{d}v
\end{aligned}$$

D'où $f(t) = \int_{-\infty}^{+\infty} |v| g(tv, v)\mathrm{d}v$, que l'on écrit :

$$f(t) = \int_{-\infty}^{+\infty} |y| g(ty, y)\mathrm{d}y \tag{7.9}$$

7 Variable variance

Soit un ensemble $(X_1, X_2, \ldots, X_n)$ de variables aléatoires, à valeurs dans $\mathbb{R}$, indépendantes et de même loi.
À un tel n-échantillon on a déjà associé la variable $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, la réalisation d'une expérience entraîne l'observation dune valeur $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, moyenne empirique des mesures.
On peut également associer à ce n-échantillon la variable variance définie par :

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

La réalisation d'une expérience (n mesures) entraîne l'observation d'une valeur de S^2 : $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ appelée variance empirique des mesures.
La variable S^2, ainsi d'ailleurs que la variable $\overline{X}$, sera souvent utilisée par la suite. On peut remarquer qu'elle peut s'écrire :

$$\begin{aligned} S^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + \overline{X}^2) \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{2\overline{X}}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}n\overline{X}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}^2 + \overline{X}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 \end{aligned}$$

La détermination de la loi de S^2 sera développée ultérieurement. Cependant on a assez d'éléments pour déterminer son espérance :

$$\mathbb{E}(S^2) = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}\right]$$

Or en posant $\mathbb{E}(X_i) = \mu$ et $\mathbb{V}(X_i) = \sigma^2$, on a $\mathbb{E}(X_i^2) = \mu^2 + \sigma^2$
Et $\mathbb{E}(\overline{X}) = \mu$ et $\mathbb{V}(\overline{X}) = \sigma^2/n$, on a $\mathbb{E}(\overline{X}^2) = \mu^2 + \sigma^2/n$
(on utilise le fait que pour toute variable Y : $\mathbb{V}(Y) = \mathbb{E}(Y^2) - \mathbb{E}^2(Y)$)

Il en résulte

$$\begin{aligned}
\mathbb{E}(S^2) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i^2) - \mathbb{E}(\overline{X}^2) \\
&= \frac{1}{n}n(\mu^2+\sigma^2) - \left(\mu^2 + \frac{\sigma^2}{n}\right) \\
&= \sigma^2\left(1-\frac{1}{n}\right) \\
&= \frac{n-1}{n}\sigma^2 \qquad (7.10)
\end{aligned}$$

8 Inégalité de Schwarz

Soit (X, Y) un couple de variables aléatoires à valeurs réelle. On définit la variable $Z = X + \lambda Y$, avec $\lambda \in \mathbb{R}$. On a :

$$\mathbb{E}\left[(X+\lambda Y)^2\right] = \mathbb{E}(X^2 + \lambda^2 Y^2 + 2\lambda XY) = \lambda^2\mathbb{E}(Y^2) + 2\lambda\mathbb{E}(XY) + \mathbb{E}(X^2)$$

Il est clair que $\mathbb{E}\left[(X+\lambda Y)^2\right] \geq 0$, $\forall\lambda \in \mathbb{R}$ ce qui implique que le discriminant du trinôme du second degré en λ est forcément négatif ou nul. D'où $\Delta' = \mathbb{E}^2(XY) - \mathbb{E}(X^2)\mathbb{E}(Y^2) \leq 0$; et donc :

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)} \qquad (7.11)$$

8.1 *Applications*

On peut appliquer l'inégalité de Schwarz au couple (X', Y') où X' et Y' sont les variables centrées $X' = X - \mu_X$, $Y' = Y - \mu_Y$:

$$|\mathbb{E}(X'Y')| \leq \sqrt{\mathbb{E}(X'^2)\mathbb{E}(Y'^2)}$$

Si on remarque que $\mathbb{E}(X'Y') = \text{Cov}(X, Y)$, $\mathbb{E}(X'^2) = \sigma_X^2$ et $\mathbb{E}(Y'^2) = \sigma_Y^2$, on obtient :

$$\left|\frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y}\right| \leq 1$$

D'où la propriété du coefficient de corrélation de X et Y (voir chapitre 4) :

$$-1 \leq \rho_{XY} \leq +1$$

On note que $|\rho_{XY}| = 1$ correspond au cas $\Delta' = 0$; il existe alors *une* valeur λ_0 (racine double du trinôme du second degré en λ) telle que $\mathbb{E}([(X'+\lambda Y')^2] = 0$, ce qui signifie que la seule observable de la variable $X' + \lambda_0 Y'$ est la valeur 0. On en déduit $X' = \lambda Y'$, c'est-à-dire

$$X = \mu_X - \lambda_0(Y - \mu_Y)$$

X et Y sont liées par une relation fonctionnelle linéaire (voir chapitre 4).

9 Exercices

1. On désigne par X et Y deux variables aléatoires indépendantes, ayant chacune une loi de Poisson de paramètres respectifs λ et μ.
   (a) Calculer la loi de la variable $Z = X + Y$
   (b) Déterminer la loi conditionnelle de X sachant que $Z = n$.
2. On considère deux variables aléatoires X et Y, indépendantes, à valeurs dans $\mathbb{R}$, et de densités de probabilité respectives :

$$f_X : \begin{cases} x \mapsto 0 & , \text{ si } x < 0 \\ x \mapsto \lambda e^{-\lambda x} & , \text{ si } x \geq 0 \end{cases}$$

$$f_Y : \begin{cases} y \mapsto 0 & , \text{ si } y < 0 \\ y \mapsto \mu e^{-\mu y} & , \text{ si } y \geq 0 \end{cases}$$

   où λ et μ sont deux paramètres réels strictement positifs.
   (a) Soit la variable aléatoire $U = Y + X$. Calculer l'espérance et la variance de U, puis déterminer sa densité de probabilité ;
   (b) Soit la variable aléatoire $V = Y - X$. Calculer l'espérance et la variance de V, puis déterminer sa densité de probabilité.
3. Soit (X, Y) un couple de variables aléatoires à valeurs réelles. Démontrer la relation : $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
4. On considère deux variables aléatoires à valeurs réelles, X et Y, indépendantes et de même loi ; cette loi est définie par la densité de probabilité :

$$f : \begin{cases} t \mapsto 0 & , \text{ si } t < 0 \\ t \mapsto e^{-t} & , \text{ si } t \geq 0 \end{cases}$$

   On définit les variables aléatoires $U = \frac{1}{2}(X + Y)$ et $V = \frac{1}{2}(X - Y)$
   (a) Calculer l'espérance et la variance de U et de V.
   (b) Déterminer la densité de probabilité de U.
   (c) Déterminer la densité de probabilité de V.
5. Soit (X, Y) un couple de variables aléatoires à valeurs réelles, de densité de probabilité :

$$g : \begin{cases} (x, y) \mapsto \frac{2}{\pi}(1 - x^2 - y^2) & , \text{ si } (x, y) \in D \\ (x, y) \mapsto 0 & , \text{ si } (x, y) \notin D \end{cases}$$

   où D est défini par : $x^2 + y^2 \leq 1$
   (a) Calculer, à partir de $g(x, y)$, l'espérance et de la variance de X. En déduire l'espérance et la variance de Y.
   (b) Déterminer les densités de probabilités de X et de Y.
   (c) Calculer la covariance de X et Y. Ces variables sont-elles indépendantes ?

(d) Soit la variable aléatoire $Z = X^2 + Y^2$. Calculer son espérance et déterminer sa densité de probabilité.

6. Vous avez à décider si la concentration d'un certain produit est plus élevée dans une solution A que dans une solution B. Des dosages, effectués à cet effet, ont donné les résultats suivants :

   Solution A : 12.8 ; 13.1 ; 13.3 ; 13.6 ; 13.8 ; 14.1 mg/litre

   Solution B : 11.8 ; 12.2 ; 13.2 ; 13.4 mg/litre

   On admet que le résultat d'un dosage est la valeur d'une variable aléatoire normale dont l'espérance est la concentration du produit dans la solution choisie et dont l'écart-type, pour la technique expérimentale utilisée, est égal à 0.6 mg/litre. On admet également que tous les dosages sont effectués de manière indépendante.

   Donnez votre conclusion à partir d'un test, soigneusement justifié, utilisant les moyennes des deux séries de résultats.

   À quelle conclusion conduirait un test de la somme des rangs ?

7. Soit X une variable aléatoire d'espérance μ et de variance $\sigma^2 = \mu$. En utilisant la moyenne de n variables indépendantes et de même loi que X, on veut construire un test de l'hypothèse $H_0 \ : \mu = 64$ contre l'hypothèse $H_1 \ : \mu = 49$ avec la condition suivante : les risque de première espèce et de deuxième espèce sont égaux et valent au plus 1%.

   Quelle est la plus petite valeur qu'il faut donner à n ? Préciser, pour cette valeur, le niveau et la puissance du test ainsi que les intervalles d'acceptation et de rejet de H_0.

8. À une expérience $\mathcal{E}$ sont associées deux variables aléatoires X et Y, à valeurs réelles, indépendantes.

   X est de loi $\mathcal{N}\left(a, (0.4)^2\right)$ et Y de loi $\mathcal{N}\left(b, (0.3)^2\right)$. Construire, au niveau 5%, un test de l'hypothèse "$H_0 \ : a = b$" contre l'hypothèse "$H_1 \ : a < b$", à partir d'un ensemble de 100 expériences indépendantes et identiques à $\mathcal{E}$. Calculer la puissance de ce test dans le cas où l'hypothèse alternative s'écrit : "$H_1 \ : a = 0.15$"

9. Soit X une variable aléatoire, à valeurs réelles, qui suit une loi exponentielle de paramètre λ ; on rappelle que X admet pour espérance $1/\lambda$ et pour variance $1/\lambda^2$.

   En utilisant la moyenne de 400 variables indépendantes et de même loi que X, construire, au niveau 1%, un test de l'hypothèse $H_0 \ : \lambda = 2$ contre l'hypothèse $H_1 \ : \lambda = 2.5$.

   Calculer la puissance de ce test.

10. On désigne par Z_n une variable aléatoire définie comme la somme des carrés de n variables normales, centrées, réduites et indépendantes.

    (a) Démontrer que Z_n a pour espérance n et pour variance $2n$.

    (b) Déterminer la densité de probabilité de la variable Z_2.

    (c) Déterminer la densité de probabilité de la variable Z_4.

10 Solutions

1. Soit $Z = X + Y$. On a clairement $Z(\Omega) = \mathbb{N}$
   (a) $(Z = n) = (X + Y = n) = (X = 0 \cap Y = n) \cup (X = 1 \cap Y = n - 1) \cup \cdots \cup (X = n \cap Y = 0) = \bigcup_{k=0}^{n} [(X = k) \cap (Y = n - k)]$
   Les différentes éventualités étant incompatibles, on peut écrire :
$$\mathbb{P}(Z = n) = \sum_{k=0}^{n} \mathbb{P}[(X = k) \cap (Y = n - k)]$$
   L'indépendance de X et Y implique :
$$\begin{aligned}
\mathbb{P}(Z = n) &= sum_{k=0}^{n} \mathbb{P}(X = k)\mathbb{P}(Y = n - k) \\
&= \sum_{k=0}^{n} e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^{n-k}}{(n-k)!} \\
&= e^{-(\lambda+\mu)} \sum_{k=0}^{n} \frac{1}{k!(n_k)!} \lambda^k \mu^{n-k} \\
&= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n_k)!} \lambda^k \mu^{n-k} \\
&= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^n}{n!}
\end{aligned}$$
   Z suit une loi de Poisson de paramètre $\lambda + \mu$
   (b) $\forall n \in \mathbb{N}\ : \mathbb{P}(X = k | Z = n) = \frac{\mathbb{P}[(X=k) \cap (Z=n)]}{\mathbb{P}(Z=n)}$
   Il convient de remarquer que $\mathbb{P}[(X = k) \cap (Z = n)] = \mathbb{P}[(X = k) \cap (Y = n - k)]$. L'indépendance des variables X et Y permet alors d'écrire :
$$\mathbb{P}[(X = k) \cap (Z = n)] = \mathbb{P}(X = k)\mathbb{P}(Y = n - k) = e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^{n-k}}{(n-k)!}$$
   Finalement :
$$\begin{aligned}
\mathbb{P}(X = k | Z = n) &= \frac{n!}{e^{-(\lambda+\mu)}(\lambda+\mu)^n} \frac{e^{-\lambda}\lambda^k e^{-\mu}\mu^{n-k}}{k!(n-k)!} \\
&= C_n^k \frac{\lambda^k \mu^{n-k}}{(\lambda+\mu)^n} \\
&= C_n^k \left(\frac{\lambda}{(\lambda+\mu)} \right)^k \left(\frac{\mu}{(\lambda+\mu)} \right)^{n-k}
\end{aligned}$$
   La loi conditionnelle de $(X|Z = n)$ est donc une loi $\mathcal{B}\left(n, \frac{\lambda}{(\lambda+\mu)}\right)$

2. On reconnaît deux lois exponentielles de paramètres respectifs λ et μ.

$$\begin{array}{rcl} \mathbb{E}(U) & = & \mathbb{E}(X)+\mathbb{E}(Y)=\dfrac{1}{\lambda}+\dfrac{1}{\mu} \\ \mathbb{V}(U) & = & \mathbb{V}(X)+\mathbb{V}(Y)=\dfrac{1}{\lambda^2}+\dfrac{1}{\mu^2} \text{ car variables indépendantes} \\ \mathbb{E}(V) & = & \mathbb{E}(X)-\mathbb{E}(Y)=\dfrac{1}{\lambda}-\dfrac{1}{\mu} \\ \mathbb{V}(V) & = & \mathbb{V}(X)+\mathbb{V}(Y)=\dfrac{1}{\lambda^2}+\dfrac{1}{\mu^2} \text{ car variables indépendantes} \end{array}$$

(a) On détermine la fonction de répartition F_u

$$F_u(t)=\mathbb{P}(U<t)=\mathbb{P}(X+Y<t)=\int\int_{y<t-x} g(x,y)\mathrm{d}x\mathrm{d}y$$

Les variables étant indépendantes $g(x,y)=f_x(x)f_y(y)$ donc :

$$\begin{cases} g(x,y)=\lambda\mu e^{-\lambda x}e^{-\mu y} & , \text{ si } x>0 \text{ et } y>0 \\ g(x,y)=0 & , \text{ si } x\leq 0 \text{ ou } y\leq 0 \end{cases}$$

On en déduit :

$$F_U(t)=0 \text{ si } t\leq 0 \text{ car } g(x,y)=0 \text{ sur le domaine d'intégration}$$

$$\begin{array}{rcl} F_U(t) & = & \displaystyle\int\int_\delta \lambda\mu e^{-\lambda x}e^{-\mu y}\mathrm{d}x\mathrm{d}y \text{ pour } t>0 \\ & = & \displaystyle\int_0^t \lambda e^{-\lambda x}\mathrm{d}x\int_0^{t-x}\mu e^{-\mu y}\mathrm{d}y \\ & = & \displaystyle\int_0^t \lambda e^{-\lambda x}\left[1-e^{-\mu t+\mu x}\right]\mathrm{d}x \\ & = & \displaystyle\int_0^t \lambda e^{-\lambda x}\mathrm{d}x-\lambda e^{-\mu t}\int_0^t e^{x(\mu-\lambda)}\mathrm{d}x \end{array}$$

Premier cas : $\lambda=\mu$

$$F_u(t)=1-e^{-\lambda t}-\lambda te^{-\lambda t}$$

D'où la densité :

$$f_U:\begin{cases} t\mapsto 0 & , \text{ si } t\leq 0 \\ t\mapsto \lambda^2te^{-\lambda t} & , \text{ si } t>0 \end{cases}$$

Deuxième cas : $\lambda\neq\mu$

$$\begin{array}{rcl} F_u(t) & = & 1-e^{-\lambda t}-\dfrac{\lambda}{\mu-\lambda}e^{-\mu t}\left[e^{x(\mu-\lambda)}\right]_0^t \\ & = & 1-e^{-\lambda t}-\dfrac{\lambda}{\mu-\lambda}e^{-\lambda t}+\dfrac{\lambda}{\mu-\lambda}e^{-\mu t} \\ & = & 1+\dfrac{\lambda e^{-\mu t}-\mu e^{-\lambda t}}{\mu-\lambda} \end{array}$$

D'où la densité :

$$f_U : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto \frac{\lambda\mu}{\mu-\lambda}\left(e^{-\lambda t} - e^{-\mu t}\right) & , \text{ si } t > 0 \end{cases}$$

(b) On cherche à déterminer la fonction de répartition F_V

$$F_V(t) = \mathbb{P}(V < t) = \mathbb{P}(X - Y < t) = \int\int_{y > x-t} g(x,y)\mathrm{d}x\mathrm{d}y$$

On distinguera deux cas :

$t \geq 0$

$$\begin{aligned} F_V(t) &= \int\int_D \lambda\mu e^{-\lambda x} e^{-\mu y}\mathrm{d}x\mathrm{d}y \\ &= \int_0^{+\infty} \mu e^{-\mu y}\mathrm{d}y \int_0^{y+t} \lambda e^{-\lambda x}\mathrm{d}x \\ &= \int_0^{+\infty} \mu e^{-\mu y}\left[1 - e^{-\lambda(y+t)}\right]\mathrm{d}y \\ &= \int_0^{+\infty} \mu e^{-\mu y}\mathrm{d}y - \mu e^{-\lambda t}\int_0^{+\infty} e^{-y(\mu+\lambda)}\mathrm{d}y \\ &= 1 + \mu e^{-\lambda t}\left[\frac{e^{-y(\mu+\lambda)}}{\mu+\lambda}\right]_0^{+\infty} \\ &= 1 - \frac{\mu}{\mu+\lambda}e^{-\lambda t} \end{aligned}$$

$t \leq 0$

$$\begin{aligned} F_V(t) &= \int_0^{+\infty} \mu e^{-\mu y}\mathrm{d}y \int_0^{y+t} \lambda e^{-\lambda x}\mathrm{d}x \\ &= \int_0^{+\infty} \mu e^{-\mu y}\left[1 - e^{-\lambda(y+t)}\right]\mathrm{d}y \\ &= \int_0^{+\infty} \mu e^{-\mu y}\mathrm{d}y - \mu e^{-\lambda t}\int_0^{+\infty} e^{-y(\mu+\lambda)}\mathrm{d}y \\ &= \left[-e^{-\mu y}\right]_{-t}^{+\infty} + \mu e^{-\lambda t}\left[\frac{e^{-y(\mu+\lambda)}}{\mu+\lambda}\right]_{-t}^{+\infty} \\ &= e^{\mu t} - \frac{\mu}{\mu+\lambda}e^{-\lambda t + (\mu+\lambda)t} \\ &= e^{\mu t}\left(1 - \frac{\mu}{\mu+\lambda}\right) \\ &= \frac{\lambda}{\mu+\lambda}e^{\mu t} \end{aligned}$$

D'où la densité :

$$f_V : \begin{cases} t \mapsto \frac{\lambda\mu}{\mu+\lambda} e^{\mu t} & , \quad \text{si } t < 0 \\ t \mapsto \frac{\lambda\mu}{\mu+\lambda} e^{-\lambda t} & , \quad \text{si } t \geq 0 \end{cases}$$

3.

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \mathbb{E}(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) \\ &= \mathbb{E}(XY) - \mu_X \mathbb{E}(Y) - \mu_Y \mathbb{E}(X) + \mu_X \mu_Y \\ &= \mathbb{E}(XY) - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \end{aligned}$$

Si les variables sont indépendantes $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ et donc $\text{Cov}(X, Y) = 0$

4. (a)

$$\begin{aligned} \mathbb{E}(X) = \mathbb{E}(Y) &= \int_{-\infty}^{+\infty} t f(t) \text{d}t = \int_0^{+\infty} t e^{-t} \text{d}t = \left[-te^{-t} - e^{-t}\right]_0^{+\infty} = 1 \\ \mathbb{E}(X^2) = \mathbb{E}(Y^2) &= \int_{-\infty}^{+\infty} t^2 f(t) \text{d}t = \int_0^{+\infty} t^2 e^{-t} \text{d}t = \left[-t^2 e^{-t}\right]_0^{+\infty} + \int_0^{+\infty} t f(t) \text{d}t = 2 \\ \mathbb{V}(X) = \mathbb{V}(Y) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) = 1 \end{aligned}$$

On en déduit :

$$\begin{aligned} \mathbb{E}(U) = \mathbb{E}\left[\frac{1}{2}(X + Y)\right] &= \frac{1}{2}\mathbb{E}(X + Y) = \frac{1}{2}(\mathbb{E}(X) + \mathbb{E}(Y)) = 1 \\ \mathbb{E}(V) = \mathbb{E}\left[\frac{1}{2}(X - Y)\right] &= \frac{1}{2}\mathbb{E}(X - Y) = \frac{1}{2}(\mathbb{E}(X) - \mathbb{E}(Y)) = 0 \end{aligned}$$

Et puisque les variables sont indépendantes :

$$\begin{aligned} \mathbb{V}(U) = \mathbb{V}\left[\frac{1}{2}(X + Y)\right] &= \frac{1}{4}\mathbb{V}(X + Y) = \frac{1}{4}(\mathbb{V}(X) + \mathbb{V}(Y)) = \frac{1}{2} \\ \mathbb{V}(V) = \mathbb{V}\left[\frac{1}{2}(X - Y)\right] &= \frac{1}{4}\mathbb{V}(X - Y) = \frac{1}{4}(\mathbb{V}(X + \mathbb{V}(Y)) = \frac{1}{2} \end{aligned}$$

(b) Soit $g(x, y)$ la densité du couple (X, Y). On aura, à cause de l'indépendance des variables :

$$g(x, y) = \begin{cases} e^{-x-y} & , \quad \text{si } x > 0 \text{ et } y > 0 \\ 0 & , \quad \text{si } x \leq 0 \text{ ou } y \leq 0 \end{cases}$$

Soit F_U la fonction de répartition de U.
$F_U(t) = \mathbb{P}(U < t) = \mathbb{P}(X + Y < 2t) = \int\int_{x+y<2t} g(x, y) \text{d}x \text{d}y$.
On en déduit : $F_u(t) = 0$ si $t \leq 0$ et, pour $t > 0$

$F_U(t) = \int_0^{2t} e^{-x}\mathrm{d}x \int_0^{2t-x} e^{-y}\mathrm{d}y = \int_0^{2t} e^{-x}\mathrm{d}x\,[1-e^{-2t+x}] = \int_0^{2t} e^{-x}\mathrm{d}x - \int_0^{2t} e^{-2t}\mathrm{d}x = 1-e^{-2t}-2te^{-2t}$.

D'où la densité :

$$f_U : \begin{cases} t \mapsto 0 & , \text{ si } t \le 0 \\ t \mapsto 4te^{-2t} & , \text{ si } t > 0 \end{cases}$$

(c) Soit F_V la fonction de répartition de V. On a :

$F_V(t) = \mathbb{P}(V < t) = \mathbb{P}(X - Y < 2t) = \int\int_{y>x-2t} g(x,y)\mathrm{d}x\mathrm{d}y$.

Pour $t \le 0$, on a :

$F_UV(t) = \int_{-2t}^{+\infty} e^{-y}\mathrm{d}y \int_0^{y+2t} e^{-x}\mathrm{d}x = \int_{-2t}^{+\infty} e^{-y}\,[1-e^{-y-2t}]\,\mathrm{d}y = \int_{-2t}^{+\infty} e^{-y}\mathrm{d}y + e^{-2t}\int_{-2t}^{+\infty} -e^{-2y}\mathrm{d}y = [-e^{-y}]_{-2t}^{+\infty} + e^{-2t}\left[\frac{e^{-y}}{2}\right]_{-2t}^{+\infty} = e^{2t} - \frac{1}{2}e^{2t} = \frac{1}{2}e^{2t}$

Si$t > 0$, on a :

$F_V(t) = \int_0^{+\infty} e^{-y}\mathrm{d}y \int_0^{y+2t} e^{-x}\mathrm{d}x = \int_0^{+\infty} e^{-y}\mathrm{d}y + e^{-2t}\int_0^{+\infty} -e^{-2y}\mathrm{d}y = 1 + e^{-2t}\left[\frac{e^{2y}}{2}\right]_0^{+\infty} = 1 - \frac{1}{2}e^{-2t}$

D'où la densité :

$$f_V : \begin{cases} t \mapsto e^{2t} & , \text{ si } t \le 0 \\ t \mapsto e^{-2t} & , \text{ si } t > 0 \end{cases}$$

5. (a) On a : $\mathbb{E}(X) = \int\int_{\mathbb{R}^2} xg(x,y)\mathrm{d}x\mathrm{d}y = \int\int_D \frac{2}{\pi}x(1-x^2-y^2)\mathrm{d}x\mathrm{d}y = \frac{2}{\pi}\int_0^{2\pi}\cos(\theta)\mathrm{d}\theta \int_0^1 r^2(1-r^2)\mathrm{d}r = \frac{2}{\pi}[\sin(\theta)]_0^{2\pi}\left[\frac{r^3}{3}-\frac{r^5}{5}\right]_0^1 = 0$

$\mathbb{E}(X^2) = \int\int_{\mathbb{R}^2} x^2g(x,y)\mathrm{d}x\mathrm{d}y = \int\int_D \frac{2}{\pi}x^2(1-x^2-y^2)\mathrm{d}x\mathrm{d}y = \frac{2}{\pi}\int_0^{2\pi}\cos^2(\theta)\mathrm{d}\theta \int_0^1 r^3(1-r^2)\mathrm{d}r = \frac{2}{\pi}\int_0^{2\pi}\frac{1+\cos(2\theta)}{2}\mathrm{d}\theta\left[\frac{r^4}{4}-\frac{r^6}{6}\right]_0^1 = \frac{1}{\pi}\left[\theta+\frac{\sin(2\theta)}{2}\right]_0^{2\pi}\times\frac{1}{12} = \frac{2\pi}{\pi}\times\frac{1}{12} = \frac{1}{6}$

$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{1}{6}$

De toute évidence $\mathbb{E}(Y) = \mathbb{E}(X) = 0$ et $\mathbb{E}(Y^2) = \mathbb{E}(X^2) = \mathbb{V}(X) = \mathbb{V}(Y) = \frac{1}{6}$.

Le changement de x en y dans le calcul de $\mathbb{E}(X)$ et $\mathbb{E}(X^2)$ ne doit pas changer le résultat à cause de la symétrie du domaine D et de $g(x,y)$ par rapport à la première bissectrice.

(b) On a $f_X(x) = \int_{-\infty}^{+\infty} g(x,y)\mathrm{d}y = 0$. D'où :

$f_X(x) = 0$ si $x \le -1$ ou $x \ge 1$

et $f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{2}{\pi}(1-x^2-y^2)\mathrm{d}y$ si $-1 \le x \le 1$.

On obtient dans ce cas : $f_X(x) = \frac{2}{\pi}\left[(1-x^2)y - \frac{y^3}{3}\right]_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} = \frac{8}{3\pi}(1-x^2)^{3/2}$

Par symétrie, on obtient la même densité pour Y. X et Y ont donc même loi définie par :

$$f : \begin{cases} t \mapsto 0 & , \text{ si } t \le -1 \text{ ou } t \ge 1 \\ t \mapsto \frac{8}{3\pi}(1-x^2)^{3/2} & , \text{ si } -1 \le t \le 1 \end{cases}$$

(c) Par définition $\text{Cov}(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$. On a ici $\text{Cov}(X,Y) = \mathbb{E}(XY)$. D'où :

$$\begin{aligned}
\text{Cov}(X,Y) &= \int\int_D xyg(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int\int_D \frac{2}{\pi}xy(1-x^2-y^2)\mathrm{d}x\mathrm{d}y \\
&= \frac{2}{\pi}\int_0^{2\pi}\cos\theta\sin\theta\mathrm{d}\theta\int_0^1 r^3(1-r^2)\mathrm{d}r \\
&= \frac{2}{\pi}\left[\frac{\sin^2\theta}{2}\right]_0^{2\pi}\left[\frac{r^4}{4}-\frac{r^6}{6}\right]_0^1 = 0
\end{aligned}$$

Les variables ne sont cependant pas indépendantes puisque sur $\mathbb{R}^2$, $g(x,y) \neq f_X(x)f_Y(y)$

(d) On a $\mathbb{E}(Z) = \mathbb{E}(X^2+Y^2) = \mathbb{E}(X^2)+\mathbb{E}(Y^2) = \frac{1}{6}+\frac{1}{6} = \frac{1}{3}$

Soit H la fonction de répartition de Z. On a $H(t) = \mathbb{P}(Z < t) = \mathbb{P}(X^2 + Y^2 < t)$.

Il est clair que pour $t \leq 0$, on aura $H(t) = 0$ et pour $t > 1$, $H(t) = 1$.

Pour $0 < t < 1$, on obtient :

$H(t) = \int\int_{x^2+y^2<t} g(x,y)\mathrm{d}x\mathrm{d}y = \frac{2}{\pi}\int_0^{2\pi}\mathrm{d}\theta\int_0^1(1-r^2)r\mathrm{d}r = 4\left[\frac{r^2}{2}-\frac{r^4}{4}\right]_0^{\sqrt{t}} = 2t-t^2$.

Finalement :

$$H : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto 2t-t^2 & , \text{ si } 0 \leq t \leq 1 \\ t \mapsto 1 & , \text{ si } t \geq 1 \end{cases}$$

et par dérivation :

$$h : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \text{ ou } t \geq 1 \\ t \mapsto 2-2t & , \text{ si } 0 \leq t \leq 1 \end{cases}$$

6. Soit $X \sim \mathcal{N}(\mu_A; 0.36)$ la variable aléatoire associée à un dosage de A et $\overline{X}$ la moyenne de 6 de ces variables indépendantes.

   On a $\overline{X} \sim \mathcal{N}\left(\mu_A, \frac{0.36}{6}\right)$

   Soit $Y \sim \mathcal{N}(\mu_B; 0.36)$ la variable aléatoire associée à un dosage de B et $\overline{Y}$ la moyenne de 4 de ces variables indépendantes.

   On a $\overline{Y} \sim \mathcal{N}\left(\mu_B, \frac{0.36}{4}\right)$

   On en déduit : $\overline{X}-\overline{Y} \sim \mathcal{N}\left(\mu_A-\mu_B, \frac{0.36}{6}+\frac{0.36}{4} = 0.15\right)$

   On teste $H_0 : \mu_A = \mu_B$ contre $H_1 : \mu_A > \mu_B$.

   Sous $H_0 : \overline{X}-\overline{Y} \sim \mathcal{N}(0, 0.15)$

   Sous $H_1 : \overline{X}-\overline{Y} \sim \mathcal{N}(\mu > 0, 0.15)$.

   On construit un test unilateral avec domaine de rejet de H_0 à droite :

Soit a la borne inférieure du domaine de rejet de H_0. Au niveau 5%, $a = 1.645\sqrt{0.15} = 0.6371$

Donc si $\overline{x} - \overline{y} < 0.6371$ on ne rejette pas H_0 et si $\overline{x} - \overline{y} > 0.6371$ on rejette H_0.

Ici $\overline{x} = 13.45$ et $\overline{y} = 12.65$ donc $\overline{x} - \overline{y} = 0.8$ donc on choisit H_1. Il y a une concentration plus élevée dans A.

Test de la somme des rangs

Soit H_0 : les variables X et Y ont la même loi contre H_1 : les variables Y ont une loi qui favorise les valeurs plus faibles que celles de X.

On va classer les observations par valeurs croissantes et on choisit la variable W, somme des rangs des dosages provenant de B.

Sous H_0, W a une loi en cloche symétrique sur $\{10, 11, 12, \ldots, 32, 33, 34\}$.

Sous H_1, W a tendance a prendre des valeurs plus proches de 10. On construit donc un test unilatéral avec domaine de rejet H_0 à droite. Au niveau 5%, ce domaine de rejet est $\{10, 11, 12, 13, 14\}$.

On a observé : $BBAABABAAA$ donc $w_0 = 1 + 2 + 5 + 7 = 15$. Les dosages ne permettent pas de rejeter H_0.

7. Soit $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

   Sous H_0 : $\mu = \sigma^2 = 64$, $\overline{X} \sim \mathcal{N}\left(64, \frac{64}{n}\right)$

   Sous H_01 : $\mu = \sigma^2 = 49$, $\overline{X} \sim \mathcal{N}\left(49, \frac{49}{n}\right)$

   Pour tester H_0 contre H_1, on est amené à construire un test bilatéral avec domaine de rejet H_0 à gauche. Soit a la borne supérieure du domaine de rejet de H_0. On peut déterminer a de manière à ce que :

   $\alpha = \mathbb{P}_{H_0}(\overline{X} \leq a) = \beta = \mathbb{P}_{H_1}(\overline{X} \geq a)$ avec α et β plus petit que 1%.

   On aura donc $a = 64 - \lambda\frac{8}{\sqrt{n}} = 49 + \lambda\frac{8}{7}\sqrt{n}$ avec $\lambda \geq 2.326$ (loi normale centrée réduite au seuil 1%) .

   Donc $\lambda\frac{15}{\sqrt{n}} = 15 \Rightarrow \lambda = \sqrt{n}$

   $\lambda \geq 2.326 \Rightarrow n \geq 5.41$. On choisit donc $n = 6$.

   Pour $n = 6$, $\lambda = \sqrt{6} \approx 2.4495 \Rightarrow \alpha = \beta \approx 0.71\%$ et donc la puissance $\pi = 1 - \beta \approx 99.29\%$.

   $a = 64 - 8 = 49 + 7 = 56$. D'où le test :

   Si $\overline{x} \leq 56$, on choisit H_1.

   Si $\overline{x} > 56$, on choisit H_0.

8. On considère la variable aléatoire $Z = Y - X$. $Z \sim \mathcal{N}(b - a, 0.5^2)$ puisque X et Y sont indépendantes.

   Soit $\overline{Z} = \frac{1}{100}\sum_{i=1}^{100} Z_i$. Les Z_i étant indépendantes et de même loi on a $\overline{Z} \sim \mathcal{N}\left(b - a, \frac{0.5^2}{100}\right)$.

   On teste H_0 : $a = b$, $\overline{Z} \sim \mathcal{N}\left(0, 0.05^2\right)$ contre H_1 : $a < b$, $\overline{Z} \sim \mathcal{N}\left(\mu > 0, 0.05^2\right)$.

   C'est donc un test unilatéral avec domaine de rejet H_0 à droite.

   Si $\overline{Z} \geq u$, on rejette H_0 ; si $\overline{Z} < u$, on accepte H_0. Au niveau 5% on a $u = 1.645 \times 0.05 = 0.0825$.

Si, sous H_1, $b - a = 0.15$, on a

$$\pi = \mathbb{P}_{H_1}(\overline{Z} \geq 0.0825) = \mathbb{P}\left(\overline{Z^*} \geq \frac{0.0825 - 0.15}{0.05}\right) = \mathbb{P}\left(\overline{Z^*} \geq -1.355\right) \approx 91.23\%$$

9. Soit $\overline{X}$ la moyenne de 400 variables indépendantes et de même loi que X. Deux résultats fondamentaux vont être utilisés :

$$\begin{aligned}\mathbb{E}(\overline{X}) &= \mathbb{E}(X) = {}^1\!/\!_\lambda \\ \mathbb{V}(\overline{X}) &= \frac{1}{n}\mathbb{V}(X) = \frac{1}{n\lambda^2}\end{aligned}$$

On peut enfin, d'après le théorème de la limite centrale et vu la taille de l'échantillon, considérer que $\overline{X}$ suit approximativement un loi normale.
Sous H_0, $\overline{X} \sim \mathcal{N}\left(0.5, \frac{1}{1600}\right)$
Sous H_1, $\overline{X} \sim \mathcal{N}\left(0.4, \frac{1}{2500}\right)$.
On est donc amené à construire un test unilatéral avec domaine de rejet de H_0 à gauche. La borne supérieure a du domaine de rejet est donnée, au niveau 1%, par $a = 0.5 - \frac{2.326}{40}$.
D'où le test : si $\overline{x} \leq 0.44185$, on rejette H_0 si $\overline{x} > 0.44185$, on accepte H_0.
La puissance π de ce test est donné par $\pi = \mathbb{P}_{H_1}(\overline{X} \leq 0.44185)$
Si T est la variable normale réduite associée à $\overline{X}$, on obtient
$\pi = \mathbb{P}\left(T \leq \frac{0.44185-0.40}{1/50}\right) = \mathbb{P}(T \leq 2.0925) \approx 98.18\%$

10. **Définition 7.1** *On appelle loi du chi-deux à n degrés de liberté $(n \in \mathbb{N}^*)$, la loi suivie par une variable aléatoire Z_n définie comme la somme des carrés de n variables normales centrées réduites et indépendantes.*
Si $(X_1, X_2, \ldots, X_n)$ est un n-échantillon de variables $\mathcal{N}(0,1)$, on a donc :$Z_n = X_1^2 + X_2^2 + \cdots + X_n^2$.
On note : $Z_n \sim \chi_n^2$
Les lois du chi-deux sont à la base des modèles gaussiens que nous étudierons dans le chapitre 9.
Il est clair que $Z_n(\Omega) = \mathbb{R}^+$. Si F_n et f_n sont respectivement la fonction de répartition et la densité de Z_n, on a :

$$\forall t \leq 0, F_n(t) = \mathbb{P}(Z_n < t) = \mathbb{P}(\emptyset) = 0 \Rightarrow f_n(t) = 0$$

On rappelle que si G et g représente la fonction de répartition et la densité d'une variable $\mathcal{N}(0,1)$.
Pour $t > 0$, $F_1(t) = \mathbb{P}(Z_1 < t) = \mathbb{P}(X_1^2 < t) = \mathbb{P}(-\sqrt{t} < X_1 < \sqrt{t}) = G(\sqrt{t}) - G(-\sqrt{t})$.
Et donc $f_1(t) = \frac{1}{2\sqrt{t}}g(\sqrt{t}) + \frac{1}{2\sqrt{t}}g(-\sqrt{t}) = \frac{1}{\sqrt{t}}g(\sqrt{t})$, puisque g est paire.
La loi du chi-deux à un degré de liberté est définie par la densité :

$$f_1 : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto \frac{1}{\sqrt{2\pi t}}e^{-t/2} & , \text{ si } t > 0 \end{cases}$$

(a) Si $X \sim \mathcal{N}(0,1)$, $\mathbb{E}(X)=0$, $\mathbb{V}(X)=1$ et $\mathbb{E}(X^2)=\mathbb{V}(X)+\mathbb{E}^2(X)=1$.

$$E(X^4)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t^4e^{-t^2/2}\mathrm{d}t=\frac{1}{\sqrt{2\pi}}\left[-t^3e^{-t^2/2}\right]_{-\infty}^{+\infty}+\frac{3}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t^2e^{-t^2/2}\mathrm{d}t=0+3\mathbb{E}(X^2)$$

Soit $Z_1=X_1^2$, on a $\mathbb{E}(Z_1)=\mathbb{E}(X_1^2)=1$.
$\mathbb{V}(Z_1)=\mathbb{V}(X_1^2)=\mathbb{E}(Z_1^2)-\mathbb{E}^2(Z_1)=\mathbb{E}(X_1^4)-\mathbb{E}^2(X_1^2)=3-1=2$
Plus généralement $\mathbb{E}(Z_n)=\mathbb{E}\left(\sum_{i=1}^n X_i^2\right)=\sum_{i=1}^n\mathbb{E}\left(X_i^2\right)=n\mathbb{E}(X_1^2)=n$
$\mathbb{V}(Z_n)=\mathbb{V}\left(\sum_{i=1}^n X_i^2\right)=\sum_{i=1}^n\mathbb{V}\left(X_i^2\right)=n\mathbb{V}(X_1^2)=2n$

(b) Le couple (X_1,X_2) de variables indépendantes a pour densité h telle que $h(x_1,x_2)=\frac{1}{2\pi}e^{-\frac{1}{2}(x_1^2+x_2^2)}\mathrm{d}x_1\mathrm{d}x_2$.
Le domaine C est un cercle de centre O et de rayon $\sqrt{t}$.
Le passage en coordonnées polaires donne : $F_2(t)=\frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\theta\int_0^{\sqrt{t}}e^{-r^2/2}r\mathrm{d}r=\left[e^{-r^2/2}\right]_0^{\sqrt{t}}=1-e^{-t/2}$.
La loi du chi-deux à deux degrés de liberté est définie par la densité :

$$f_2:\begin{cases} t\mapsto 0 & ,\ \text{si } t\leq 0\\ t\mapsto \frac{1}{2}e^{-t/2} & ,\ \text{si } t>0\end{cases}$$

C'est aussi une loi exponentielle de paramètre ½

(c) $Z_4=X_1^2+X_2^2+X_3^2+X_4^2=U+V$ où U et V sont deux variables indépendantes et de densité f_2. Le couple (U,V) a donc pour densité :

$$l:\begin{cases} (u,v)\mapsto 0 & ,\ \text{si } u\leq 0 \text{ ou } v\leq 0\\ (u,v)\mapsto \frac{1}{4}e^{-\frac{1}{2}(u+v)} & ,\ \text{si } u>0 \text{ et } v>0\end{cases}$$

Pour $t>0$, on a :

$$F_4(t)=\mathbb{P}(Z_4<t)=\mathbb{P}(U+V<t)=\int\int_D l(u,v)\mathrm{d}u\mathrm{d}v=\int\int_\Delta \frac{1}{4}e^{-\frac{1}{2}(u+v)}\mathrm{d}u\mathrm{d}v$$

où $D=\{(u,v)\in\mathbb{R}^2|u+v<t\}$ et $\Delta=\{(u,v)\in\mathbb{R}^2|u+v<t,u>0,v>0\}$.
$F_4(t)=\int_0^t\frac{1}{2}e^{-u/2}\mathrm{d}u\int_0^{t-u}\frac{1}{2}e^{-v/2}\mathrm{d}v=1-\left(1+\frac{t}{2}\right)e^{-t/2}$.
La loi du chi-deux à quatre degrés de liberté est définie par la densité :

$$f_4:\begin{cases} t\mapsto 0 & ,\ \text{si } t\leq 0\\ t\mapsto \frac{t}{4}e^{-t/2} & ,\ \text{si } t>0\end{cases}$$

Chapitre 8

L'estimation

1 Préliminaires

1.1 Une inégalité importante

Soit Y une variable aléatoire à valeurs dans $\mathbb{R}$ admettant un moment d'ordre deux :

$$\forall \epsilon > 0, \mathbb{P}(|Y| \geq \epsilon) \leq \frac{\mathbb{E}(Y^2)}{\epsilon^2} \tag{8.1}$$

Démonstration : soit Z la variable aléatoire que prend la valeur ϵ^2 lorsque $|Y| \geq \epsilon$, ou la valeur 0 lorsque $|Y| < \epsilon$.

$$Z \leq Y^2 \Rightarrow \mathbb{E}(Z) \leq \mathbb{E}(Y^2) \ ; \ \mathbb{E}(Z) = \epsilon^2\mathbb{P}(|Y| \geq \epsilon) + 0\mathbb{P}(|Y| < \epsilon) \Rightarrow (8.1)$$

1.2 Inégalité de Bienaymé-Tchebichev

Soit X une variable aléatoire à valeurs dans $\mathbb{R}$, d'espérance $\mathbb{E}(X) = \mu_X$ et de variance $\mathbb{V}(X) = \sigma_X^2$. En appliquant la relation 8.1 à la variable $(X - \mu_X)$, on obtient :

$$\forall \epsilon > 0 \ , \ \mathbb{P}(|X - \mu_X| \geq \epsilon) \leq \frac{\sigma_X^2}{\epsilon^2} \tag{8.2}$$

Cette inégalité donne une idée de l'écart entre X et μ_x.

2 Estimation ponctuelle

2.1 Estimateur

Un estimateur T d'un paramètre θ est une variable aléatoire dont la valeur $\hat{\theta}$, prise à l'issue d'une expérience, constitue l'estimation de θ.

2.2 *Écart quadratique moyen*

Un bon estimateur T d'un paramètre θ doit avoir une distribution très resserrée autour de θ. Or, en appliquant la relation (8.1) à la variable $(T-\theta)$, on obtient :

$$\forall \epsilon > 0 \ , \ \mathbb{P}(|T-\theta| \geq \epsilon) \leq \frac{\mathbb{E}[(T-\theta)^2]}{\epsilon^2} \tag{8.3}$$

On appelle écart quadratique moyen de T relativement à θ la quantité

$$\mathbb{E}[(T-\theta)^2] = \text{EQM}_\theta(T)$$

T sera un bon estimateur si $\mathbb{E}[(T-\theta)^2]$ est faible.

2.3 *Écart quadratique moyen et variance*

On a $\mathbb{E}[(T-\theta)^2] = \mathbb{E}[(T-\mu_T+\mu_T-\theta)^2] = \mathbb{E}[(T-\mu_T)^2] - 2(\mu_T-\theta)\mathbb{E}[(T-\mu_T)] + (\mu_T-\theta)^2$

$$\mathbb{E}[(T-\theta)^2] = \mathbb{V}(T) + (\mu_T-\theta)^2 = \mathbb{V}(T) + \mathbb{E}^2(T-\theta) \tag{8.4}$$

2.4 *Biais d'un estimateur*

On appelle biais de l'estimateur T de paramètre θ la quantité

$$\mathbb{E}(T-\theta) = \mathbb{E}(T) - \theta = \mu_T - \theta$$

- Si $\mathbb{E}(T-\theta) = 0 \Leftrightarrow \mathbb{E}(T) = \mu_T = \theta$, T est dit *sans biais*.
- Si $\mathbb{E}(T-\theta) \neq 0 \Leftrightarrow \mathbb{E}(T) \neq \theta$, T est dit *biaisé*.

Il résulte de la relation 8.4 que si deux estimateurs de θ ont la même variance, le meilleur est celui de plus faible biais.
Réciproquement, pour deux estimateurs sans biais, le meilleur est celui qui a la variance minimale.

2.5 *Suite consistante d'estimateurs*

Soit $X_1, X_2, \ldots, X_n, \ldots$ une suite de variables aléatoires. Soit $T_n = \phi(X_1, X_2, \ldots, X_n)$ un estimateur du paramètre θ. On dit que la suite T_n est une suite d'estimateurs consistante (ou convergente) pour θ si :

$$\forall \epsilon > 0 \ , \ \mathbb{P}(|T_n-\theta| < \epsilon) \to 1 \text{ lorsque } n \to +\infty$$

Or, en passant par l'événement contraire, il résulte de la relation 8.3 :

$$\forall \epsilon > 0 \ , \ 1 \geq \mathbb{P}(|T_n-\theta| < \epsilon) \geq 1 - \frac{\mathbb{E}[(T_n-\theta)^2]}{\epsilon^2} \tag{8.5}$$

Une suite d'estimateurs est donc consistante quand leur écart quadratique moyen tend vers zéro lorsque n tend vers l'infini.
Par abus de langage, on dit parfois qu'un estimateur T_n, pour n donné, élément d'un suite consistante, est un estimateur consistant.
Un estimateur consistant est un estimateur particulièrement intéressant puisque, en jouant sur n, il permet d'obtenir, presque sûrement, une estimation aussi proche que l'on veut de θ.
On notera, enfin, que si les estimateurs T_n sont sans biais, la relation 8.5 s'écrit :

$$\forall \epsilon > 0 \ , \ 1 \geq \mathbb{P}(|T_n - \theta| < \epsilon) \geq 1 - \frac{\mathbb{V}(T_n)}{\epsilon^2}$$

Une suite T_n d'estimateurs sans biais dont la variance tend vers zéro lorsque n tend vers l'infini est donc consistante.

2.6 *Exemples*

Probabilité d'un événement

Dans le schéma de Bernoulli, la fréquence Y_n d'un événement associée à un ensemble de n expériences peut être choisie comme estimateur de la probabilité p de l'événement considéré. C'est un estimateur sans biais, puisque $\mathbb{E}(Y_n) = p$.
C'est un estimateur consistant puisque c'est un estimateur sans biais et tel que

$$\mathbb{V}(Y_n) = \frac{p(1-p)}{n} \to 0 \text{ lorsque } n \to +\infty$$

Si y_n est la fréquence observée, on écrira : $\hat{p} = y_n$.
On notera que pour avoir : $1 \geq \mathbb{P}(|Y_n - p| < 10^{-2}) \geq 0.99$, il faut :

$$\frac{\mathbb{V}(Y_n)}{10^{-4}} \leq 10^{-2} \Rightarrow \frac{p(1-p)}{n} \leq \frac{1}{4n} \leq 10^{-6} \Rightarrow n \geq 250\,000$$

Il s'agit là d'une approche grossière ; ce nombre pourra être très fortement diminué si on fait intervenir l'information de la loi de Y_n.

Espérance d'une variable aléatoire

On se place dans le cas d'un n-échantillon, c'est-à-dire d'un ensemble $(X_1, X_2, \ldots, X_n)$ de n variables aléatoires, indépendantes et de même loi.
On pose : $\mathbb{E}(X_i) = \mu_X$, $\mathbb{V}(X_i) = \sigma_X^2$, $\forall i \in \{1, 2, \ldots, n\}$.
La variable $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ peut être choisie comme estimateur de μ_X. $\overline{X_n}$ est alors un estimateur sans biais, puisque $\mathbb{E}(\overline{X_n}) = \mu_X$.
$\overline{X_n}$ est un estimateur consistant puisque sans biais et tel que $\mathbb{V}(\overline{X_n}) = \frac{\sigma_X^2}{n} \to 0$ lorsque $n \to +\infty$.
Dans ce cas la relation 8.5, s'écrit :

$$\forall \epsilon > 0 \ , \ 1 \geq \mathbb{P}(|\overline{X_n} - \mu_X| < \epsilon) \geq 1 - \frac{\sigma_X^2}{n\epsilon^2}$$

ce qui constitue la *loi faible des grands nombre*.
On notera enfin que Y_n pouvant représenter la moyenne de n variables indicatrices, l'exemple précédent (probabilité d'un événement) n'est qu'un cas particulier de celui-ci.

Variance d'une variable aléatoire

On reste dans le cas d'un n-échantillon et on considère la variable

$$\widehat{S_n^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

peut être choisie comme estimateur de σ_X^2. Il s'agit, bien sûr d'un estimateur sans biais, puisque $\mathbb{E}(\widehat{S_n^2}) = \sigma_X^2$.
Dans ce cas la relation 8.5, s'écrit :

$$\forall \epsilon > 0 \ , \ 1 \geq \mathbb{P}(|\widehat{S_n^2} - \sigma_X^2| < \epsilon) \geq 1 - \frac{\mathbb{V}(\widehat{S_n^2})}{\epsilon^2}$$

Pour que $\widehat{S_n^2}$ soit un estimateur consistant de σ_X^2, il faut que $\mathbb{V}(\widehat{S_n^2}) \to 0$ quand $n \to +\infty$. Comme on le verra plus tard, c'est bien le cas dans les modèles gaussiens.
On peut également choisir comme estimateur de σ_X^2 la variable $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})$. Il s'agit alors d'un estimateur biaisé car $\mathbb{E}(S_n^2) = \frac{n-1}{n}\sigma_X^2 = \sigma_X^2 - \frac{\sigma_X^2}{n}$ (voir relation 7.10) ; le biais de S_n^2 est égal à ${}^{-\sigma_X^2}/_{n}$. En appliquant la relation 8.4, la relation 8.5 s'écrit alors ;

$$\forall \epsilon > 0, 1 \geq \mathbb{P}(|S_n^2 - \sigma_X^2| < \epsilon) \geq 1 - \frac{\mathbb{V}(\widehat{S_n^2})}{\epsilon^2} - \frac{\sigma_X^4}{n^2\epsilon^2}$$

Si on démontre que $\mathbb{V}(S_n^2) \to 0$ quand $n \to +\infty$, et ce sera le cas dans les modèles gaussiens, S_n^2, bien que biaisé, est un estimateur consistant de σ_X^2.
$\widehat{S_n^2}$ est-il un meilleur estimateur que S_n^2 ? Seule la comparaison de leur écart quadratique moyen peut permettre de trancher ; ce calcul sera fait ultérieurement dans le cadre des modèles gaussiens.

3 Estimation par intervalle

3.1 Intervalle de confiance

Les expérimentateurs préfèrent donner, au lieu d'une estimation ponctuelle $\widehat{\theta}$ d'un paramètre θ, un intervalle I dans lequel ils ont la quasi-certitude de cerner la valeur exacte θ. I s'appelle un intervalle de confiance de θ ; la "quasi-certitude", dont dépend la largeur de I, est "mesurée" par une probabilité appelée niveau de confiance ou coefficient de sécurité de I.
On ne donnera jamais un intervalle de confiance sans l'accompagner du coefficient de sécurité choisi.

3.2 Principe

Pour déterminer un intervalle de confiance I, on choisit une *statistique* X, c'est-à-dire une variable aléatoire, dont la loi dépend du paramètre inconnu θ. Cette statistique n'est pas forcément un estimateur de θ.
Si θ est connu, la loi de X est complètement précisée. On peut alors parier, avant toute expérience, que l'observation future x_0 va se trouver, avec une forte probabilité $(1-\alpha)$ dans un intervalle J ; cet intervalle J est parfaitement déterminé si, par exemple, on affecte une probabilité $\alpha/2$ à sa gauche et de $\alpha/2$ à sa droite.
Réciproquement, lorsqu'on ignore θ donc la loi précise de X, on peut parier que l'événement $x_0 \in J$ a bien eu lieu. On peut alors en déduire l'ensemble des intervalles J qui contiennent l'observation x_0, donc un ensemble de lois possibles et donc un ensemble I de valeurs possibles pour θ.
L'ensemble I est l'intervalle de confiance de θ. La probabilité $1-\alpha$ choisie au départ s'appelle le niveau de confiance ou le coefficient de sécurité associé à I.
On remarquera que cette démarche, lorsque le type de la loi de X est connu, revient à faire un test bilatéral à l'envers. L'intervalle de confiance de θ contient l'ensemble des valeurs de θ qui seraient acceptées dans un tel test, au niveau α, compte tenu de l'observation x_0.
Dans le cadre de ce cours, on adoptera systématiquement la méthode qui vient d'être décrite. On conçoit cependant que, dans le cas de lois dissymétriques, on n'a pas forcément avantage à s'appuyer sur un test bilatéral avec partage de risque.

3.3 Utilisation de la loi binomiale

Soit une variable aléatoire X de loi $\mathcal{B}(n,p)$. On observe la valeur x_0 de X. Peut-on en déduire un intervalle de confiance de p, pour un coefficient de sécurité $(1-\alpha)$?
La loi $\mathcal{B}(n,p)$, pour n donné, se déplaçant de la gauche vers la droite lorsque p augmente, on doit avoir

$$\mathbb{P}_i(X < x_0) = \sum_{k=0}^{x_0-1} C_n^k (p_{\text{inf}})^k (1-p_{\text{inf}})^{n-k} = 1 - \frac{\alpha}{2}$$

$$\mathbb{P}_s(X \leq x_0) = \sum_{k=0}^{x_0} C_n^k (p_{\text{sup}})^k (1-p_{\text{sup}})^{n-k} = \frac{\alpha}{2}$$

Au niveau $(1-\alpha)$, $p_{\text{inf}} < p < p_{\text{sup}}$
Les deux équations ci-dessus peuvent être résolues par approximations successives. On peut trouver des abaques, ou des tables (voir table 11.3 pour $1-\alpha = 95\%$) en donnant les solutions

Exemple 8.1 *On veut estimer la proportion p des mâles à la naissance dans une espèce donnée. On observe x_0 mâles sur n naissances. Qu'en conclure ?*
Pour un coefficient de sécurité de 95% :

- $n = 30$, $x_0 = 12 \Rightarrow 0.227 < p < 0.594$
- $n = 34$, $x_0 = 12 \Rightarrow 0.202 < p < 0.543$.
  En effet $p_{\text{sup}} = 0.594 - \frac{4}{10}(0.594 - 0.465)$ *et* $p_{\text{inf}} = 0.227 - \frac{4}{10}(0.227 - 0.166)$

- $n = 30$, $x_0 = 24 \Rightarrow 0.614 < p < 0.923$.
  24 mâles $\Leftrightarrow$ *6 femelles et* $0.077 < 1 - p < 0.386$

3.4 *Utilisation de la loi de Poisson*

Soit une variable aléatoire X de loi $\mathcal{P}(\mu)$. On observe la valeur x_0 de X. Peut-on en déduire un intervalle de confiance de μ, pour un coefficient de sécurité $(1-\alpha)$?
La méthode suivie est la même que précédemment. Les équation à résoudre sont cette fois :

$$\left.\begin{array}{l} \mathbb{P}_i(X < x_0) = \sum_{k=0}^{x_0-1} e^{-\mu_{\text{inf}}} \frac{\mu_{\text{inf}}^k}{k!} = 1 - \frac{\alpha}{2} \\ \mathbb{P}_i(X \leq x_0) = \sum_{k=0}^{x_0} e^{-\mu_{\text{sup}}} \frac{\mu_{\text{sup}}^k}{k!} = \frac{\alpha}{2} \end{array}\right\} \text{Au niveau de confiance } (1-\alpha) : \mu_{\text{inf}} < \mu < \mu_{\text{sup}}$$

La table 11.5 donne les intervalles de confiance de μ pour un coefficient de sécurité de 95% et ceci pour $x_0 \leq 100$.

Exemples

- Le nombre d'impulsions enregistrées pendant une minute par un compteur est une variable aléatoire X de loi $\mathcal{P}(\mu)$. L'espérance μ caractérise l'intensité de la source de rayonnement. On mesure 59 coups durant une minute. On en déduit, avec un coefficient de sécurité de 95% :
$$44.9 < \mu < 76.1$$
- Sur 12000 individus d'une espèce, on a dénombré 13 albinos. Que peut-on déduire sur la proportion p des albinos dans l'espèce ?
  La variable $X =$ nombre d'albinos que l'on peut observer sur 12000 individus suit une loi $\mathcal{B}(12000, p)$ si on peut considérer le choix de l'échantillon comme non exhaustif. Il est clair que, dans cet exemple n est grand et p faible ; on peut donc dire que la loi de X est une loi $\mathcal{P}(\mu = np)$.
  Avec un coefficient de sécurité de 95% :
  $6.9 < 12000p < 22.3 \Rightarrow 5.75\,10^{-4} < p < 18.58\,10^{-4}$

Grand échantillon

Pour $x_0 > 100$, il est clair que l'espérance μ aura une valeur élevée et que l'on peut considérer que la loi de X peut être assimilée à une loi $\mathcal{N}(\mu, \mu)$. On peut alors parier que la valeur observée x_0 vérifie l'inégalité :

$$|x_0 - \mu| < \lambda\sqrt{\mu} \tag{8.6}$$

où λ est une valeur numérique qui dépend du coefficient de sécurité choisi ($\lambda = 1.645$ pour 90% ; $\lambda = 2.576$ pour 99%).
Les solution de (8.6) constituent l'intervalle $]\mu_{inf}, \mu_{\text{sup}}[$ décrit plus haut, lorsque la distribution de X est du type $\mathcal{N}(\mu, \mu)$.

L'inégalité 8.6 peut être résolue rigoureusement en élevant au carré les deux membres, ce qui conduit à :

$$\mu^2 - 2\mu\left(x_0 + \frac{\lambda^2}{2}\right) + x_0^2 < 0$$

$$\Rightarrow \mu_{\text{inf}} = x_0 + \frac{\lambda^2}{2} - \lambda\sqrt{x_0 + \frac{\lambda^2}{4}} \quad , \quad \mu_{\text{sup}} = x_0 + \frac{\lambda^2}{2} + \lambda\sqrt{x_0 + \frac{\lambda^2}{4}}$$

Comme $x_0 > 100$, l'inégalité 8.6 peut être résolue de manière approchée en écrivant :

$$|x_0 - \mu| < \lambda\sqrt{x_0} \tag{8.7}$$

$\Rightarrow x_0 - \lambda\sqrt{x_0} < \mu < x_0 + \lambda\sqrt{x_0}$, ce qui donne un intervalle de confiance un peu différent du précédent.

Exemple 8.2 *La mesure du rayonnement d'une source radioactive est de 100 coups/mn. On en déduit, pour l'intensité μ de la source, et avec un coefficient de sécurité de 95% :*

- $81.4 < \mu < 121.6$ *(table 11.4 ; résultat rigoureux)*
- $81.22 < \mu < 121.62$ *(inégalité 8.6)*
- $80.40 < \mu < 119.60$ *(inégalité 8.7)*

3.5 Utilisation de la loi normale

Soit une variable aléatoire X de loi $\mathcal{N}(\mu_X, \sigma_X^2)$. On observe la valeur x_0 de X. Peut-on déduire un intervalle de confiance de μ_X, pour un coefficient de sécurité $(1-\alpha)\%$?
Comme précédemment, on va parier sur l'inégalité :

$$|x_0 - \mu_X| < \lambda\sigma_X \tag{8.8}$$

$$\Rightarrow x_0 - \lambda\sigma_X < \mu < x_0 + \lambda\sigma_X$$

On ne saura évidemment donner un intervalle de confiance que si on connaît σ_X. Le développement ultérieur des modèles gaussiens permettra de surmonter cette difficulté. En attendant les seuls cas qui sont accessibles sont ceux où effectivement σ_X est connu ou encore ceux où μ_X et σ_X sont liés par une relation (voir relation 8.6 et 8.9).
L'inégalité 8.8 peut être appliquée à la variable $\overline{X}$; on rappelle que, dans ce cas $\mathbb{E}(\overline{X}) = \mu_X$ et $\sigma_{\overline{X}} = \sigma_X/\sqrt{n}$, lorsque la loi de $\overline{X}$ est normale ou peut être considérée comme telle (théorème limite centrale).

Exemple 8.3 *Reprenons le schéma de Bernoulli et considérons la fréquence Y_n d'un événement au cours d'un ensemble de n expériences. En supposant que l'on puisse considérer que $Y_n \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$, on peut parier, pour un niveau de confiance de $(1-\alpha)$ sur l'inégalité :*

$$|y_0 - p| < \lambda\sqrt{\frac{p(1-p)}{n}} \tag{8.9}$$

($\lambda = 1.96$ pour $1 - \alpha = 95\%$).

L'inégalité 8.9 peut être résolue rigoureusement en élevant au carré les deux membres. On obtient :
$p^2\left(1+\frac{\lambda^2}{n}\right) - 2p\left(y_0+\frac{\lambda^2}{2n}\right) + p_0^2 < 0 \Rightarrow p_{\text{inf}} < p < p_{\text{sup}}$ *avec :*

$$p_{\text{inf}} = \frac{y_0 + \frac{\lambda^2}{2n} - \lambda\sqrt{\frac{y_0(1-y_0)}{n} + \frac{\lambda^2}{4n^2}}}{1+\frac{\lambda^2}{n}} \ , \ p_{\text{sup}} = \frac{y_0 + \frac{\lambda^2}{2n} + \lambda\sqrt{\frac{y_0(1-y_0)}{n} + \frac{\lambda^2}{4n^2}}}{1+\frac{\lambda^2}{n}}$$

L'inégalité 8.9 peut être résolue de manière approchée en la remplaçant par :

$$|y_0 - p| < \lambda\sqrt{\frac{y_0(1-y_0)}{n}} \tag{8.10}$$

$\Rightarrow y_0 - \lambda\sqrt{\frac{y_0(1-y_0)}{n}} < p < y_0 + \lambda\sqrt{\frac{y_0(1-y_0)}{n}}$*, ce qui donne un intervalle de confiance un peu différent du précédent.*
On n'oubliera pas de contrôler a posteriori la validité de l'approximation normale en vérifiant que, sur l'intervalle de confiance, on a toujours np *et* $n(1-p) > 100$.

3.6 *Absence de loi*

L'inégalité de Bienaymé-Tchebichev permet d'écrire successivement :

$$\forall k > 0 \ , \ \mathbb{P}(|X-\mu_X| \geq k\sigma_X) \leq \frac{1}{k^2}$$

$$\Leftrightarrow 1 \geq \mathbb{P}(|X-\mu_X| < k\sigma_X) \geq 1 - \frac{1}{k^2} \tag{8.11}$$

On peut donc, en particulier et ceci quelle que soit la loi de X parier, avec un coefficient de sécurité qui est au moins de 95% sur l'inégalité :

$$|x_0 - \mu_X| < \sqrt{20}\sigma_X \qquad (1 - \frac{1}{20} = 0.95)$$

D'où, si l'on connaît σ_X un intervalle de confiance de μ_X :

$$x_0 - 4.472\sigma_X < \mu_X < x_0 + 4.472\sigma_X$$

4 Exercices

1. Parmi 900 poissons pêchés dans un lac, on a observé 180 porteurs de parasites. Entre quelles limites situez-vous la proportion des individus parasités dans la population des poissons du lac.
2. On photographie au microscope une boîte de Petri. La photographie est divisée en carrés de surfaces égales et on dénombre dans chaque carré des colonies de bactéries. Les résultats expérimentaux sont les suivants : 10 carrés ne contiennent pas de colonie, 24 carrés contiennent 1 colonie, 34 en contiennent 2, 23 en contiennent 3, 6 en contiennent 4, 3 en contiennent 5.

   On désigne par X la variable aléatoire qui dans une telle expérience, associe à un carré le nombre de colonies qu'il contient et par H_0 l'hypothèse : "X suit une loi de Poisson d'espérance μ".

   En admettant H_0 et à partir des résultats expérimentaux, donner, en les justifiant, une estimation de μ, puis un intervalle de confiance de μ pour un coefficient de sécurité de 95%.
3. On sait que, à chaque naissance, la probabilité p d'observer un garçon est très proche de $1/2$. Pour estimer précisément cette probabilité, on recherche son intervalle de confiance pour un coefficient de sécurité de 99.99% à partir de la proportion de garçons observée sur n naissances. Quelle valeur donner à n pour avoir une estimation à 0.001 près ?
4. Pour déterminer la concentration d'un certain produit dans une solution, on effectue des dosages à l'aide d'une technique expérimentale donnée. On admet que le résultat de chaque dosage est une variable aléatoire normale,
   - dont l'espérance μ est la valeur que l'on cherchera à déterminer, ceci en l'absence d'erreur systématique, c'est-à-dire si le protocole expérimental est scrupuleusement suivi.
   - dont l'écart-type est de 0.05 mg/litre pour un expérimentateur convenablement entraîné.

   (a) Vous effectuez cinq dosages indépendants d'une solution de concentration connue égale à 4.00 mg/litre. Vous obtenez les résultats suivants : 4.04 ; 4.01 ; 4.08 ; 3.95 ; 4.02 mg/litre.

   Suivez-vous scrupuleusement le protocole expérimental ?

   (b) Vous effectuez six dosages indépendants d'une solution de concentration inconnue. Vous obtenez les résultats suivants : 2.97 ; 3.01 ; 2.98 ; 2.94 ; 3.03 ; 2.95 mg/litre.

   En déduire un intervalle de confiance de la concentration cherchée, pour un coefficient de sécurité de 95%.

   Bien entendu, vous vous considérez comme normalement entraîné.
5. Soit X une variable aléatoire, associée à une expérience $\mathcal{E}$ de densité de probabilité :

$$f : \begin{cases} x \mapsto 0 & , \text{ si } x < 0 \\ x \mapsto \lambda^2 x e^{-\lambda x} & , \text{ si } x \geq 0 \end{cases}$$

où λ est un réel strictement positif.
- Calculer l'espérance et la variance de X ;
- on a répété 400 fois l'expérience $\mathcal{E}$. La moyenne des 400 valeurs observées pour X est égale à 2. Peut-on en déduire un intervalle de confiance de λ pour le coefficient de sécurité de 90% ?

6. Une boisson, de consommation courante, est vendue en bouteille d'un litre. L'un des éléments de sa composition est un certain produit A qui fait l'objet d'une réglementation particulière.

   On désigne par X la variable aléatoire représentant la quantité A, exprimée en milligrammes, que l'on peut trouver dans une bouteille. On admet que X suit une loi normale d'espérance μ et d'écart-type égal à 5 milligrammes. Pour que a boisson soit conforme à la réglementation en vigueur, la valeur de μ ne doit pas dépasser 50 milligrammes.

   Une association de consommateurs a fait effectuer des dosages de A sur 9 bouteilles choisies au hasard mais d'une marque donnée. Les résultats, en milligrammes, sont les suivants :

   $$48.0; 48.2; 49.3; 53.5; 54.7; 56.4; 57.8; 58.5; 60.5$$

   À l'aide d'un test paramétrique construit au niveau 5%, vérifier si la réglementation est respectée par le fabriquant.

7. On place un compteur devant une source de rayonnement. Le nombre d'impulsions que l'on peut enregistrer pendant une minute est une variable de Poisson d'espérance μ, constante pendant la durée de l'expérience et dont la valeur mesure l'intensité de la source.

   (a) On effectue un comptage d'une minute et on observe 6400 impulsions par minute. En déduire un intervalle de confiance de μ pour un coefficient de sécurité de 90%.

   (b) Pour améliorer la précision de cette estimation, on répète n fois la mesure. Sachant que la moyenne de ces n mesures reste de l'ordre de 6400, quelle valeur faut-il donner à n pour estimer μ à 10 unités près, toujours au niveau de 90% ?

   (c) Sachant que la somme de n variables de Poisson indépendantes est une variable de Poisson, améliorait-on l'estimation avec un comptage de n minutes ?

8. (Méthode de "capture-recapture"). On désire évaluer le nombre N d'individus d'une espèce animale vivant sur une île. Pour cela, on capture 800 individus ; ces individus sont marqués, puis relâchés. En essayant de respecter le schéma du tirage exhaustif, on recapture ultérieurement 1000 animaux parmi lesquels on dénombre 250 animaux marqués. En déduire un intervalle de confiance de N pour un coefficient de sécurité de 90%.

5 Solutions

1. Si p est la proportion des poissons parasités dans le lac, lorsqu'on pêche un poisson, la probabilité pour qu'il soit porteurs de parasites est p.
   Si le nombre de poissons dans le lac est très élevé, pécher 900 poissons revient à répéter 900 fois la même expérience. Dans ces conditions, le nombre de porteurs de parasites que l'on peut observer est une variable aléatoire de loi $\mathcal{B}(900, p)$ que l'on peut approcher par une loi $\mathcal{N}(900p, 900p(1-p))$, p étant de l'ordre de 20%.
   La variable X, proportion des individus parasités sur 900, obéit donc à peu près à une loi $\mathcal{N}\left(p, \frac{p(1-p)}{900}\right)$. On a observé la valeur $x = {}^{180}\!/_{900} = 0.20$. Pour un niveau de confiance de 95%, on peut parier que l'inégalité :
   $$|0.20 - p| < 1.960\frac{\sqrt{p(1-p)}}{30} \tag{8.12}$$
   soit, pratiquement : $|0.20 - p| < 1.960\frac{\sqrt{0.16}}{30}$.
   D'où l'intervalle de confiance : $0.174 < p < 0.226$.
   L'approximation normale était bien justifiée puisque sur tout l'intervalle trouvé : $900(1-p) > 900p > 900(0.174) > 10$
   On peut noter que résoudre rigoureusement l'équation 8.12 aurait conduit à l'inégalité :
   $$1.00426844p^2 - 0.40426844p + 0.04 < 0$$
   dont les solutions sont : $0.175 < p < 0.227$
2. Soit $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ la moyenne de n variables indépendantes et de même loi que X. $\overline{X}$ est un bon estimateur de μ. En effet :
   $$\mathbb{E}(\overline{X}) = \mu \ , \ \mathbb{V}(\overline{X}) = \frac{\mu}{n}$$
   $\overline{X_n}$ est un estimateur consistant de μ puisqu'il est sans biais et que sa variance tend vers zéro lorsque n tend vers l'infini.
   Ici $n = 100$, $\overline{x} = \frac{1}{100}(10 \times 0 + 24 \times 1 + 34 \times 2 + 23 \times 3 + 6 \times 4 + 3 \times 5) = 2$
   D'où l'estimation $\widehat{\mu} = 2$
   Le nombre d'expériences étant ici assez élevé, on peut, en utilisant le théorème limite centrale, considérer que la loi de $\overline{X}$ est pratiquement : $\mathcal{N}(\mu, {}^{\mu}\!/_{100})$
   Avec un coefficient de sécurité de 95%, on peut parier sur l'inégalité
   $$|2 - \mu| < 1.960\frac{\sqrt{\mu}}{10} \tag{8.13}$$
   soit, pratiquement : $|2 - \mu| < 1.960\frac{\sqrt{2}}{10}$.
   D'où l'intervalle de confiance : $1.722 < \mu < 2.278$.
   On peut noter que résoudre rigoureusement l'équation 8.13 conduit à l'inégalité :
   $$\mu^2 - 4.038416\mu + 4 < 0$$
   dont les solutions sont : $1.741 < \mu < 2.297$

3. Soit Y_n la proportion des garçons que l'on peut observer sur n naissances. Comme $p \approx 1/2$ et n certainement très grand, on peut considérer que $Y_n \sim \mathcal{N}\left(p, \frac{p(1-p)}{n},\right)$.
   Si y_0 est la proportion observée et pour un niveau de confiance de 0.9999, on écrira : $|y_0 - p| < 3.891\sqrt{\frac{p(1-p)}{n}}$.
   Soit pratiquement ($p \approx 1/2$) : $|y_0 - p| < 3.891\sqrt{\frac{1}{4n}}$.
   On pourra en déduire p à 0.001 près si : $\frac{3.891}{2\sqrt{n}} < 0.001$. D'où $\sqrt{n} > 3.891 \times 500$, c'est-à-dire $n > 3.785\,10^6$.
4. Soit X la variable aléatoire représentant le résultat d'un dosage. On admet, dans tout le problème que X suit une loi normale $\mathcal{N}(\mu, (0.05)^2)$
   (a) On veut tester $H_0 \ : \mu = 4.00$ contre $H_1 \ : \mu \neq 4.00$
   Pour cela on utilise $\overline{X_5}$, moyenne de 5 variables indépendantes et de même loi que X.
   Sous H_0, $\overline{X_5}$ suit une loi $\mathcal{N}\left(4, \left(\frac{0.05}{\sqrt{5}}\right)^2\right)$.
   Sous H_1 la loi de $\overline{X_5}$ est déplacée soit vers la droite, soit la gauche. On construit un test bilatéral.
   Au niveau 5% le domaine d'acceptation de H_0 est de la forme $]a; b[$ où $a = 4 - 1.960\frac{0.05}{\sqrt{5}}$ et $b = 4 + 1.960\frac{0.05}{\sqrt{5}}$. Donc de la forme $]3.956; 4,044[$. On a observé $\overline{x_5} = 4.02$. On peut considérer que le protocole expérimental est scrupuleusement suivi
   (b) On considère la variable $\overline{X_6}$, moyenne de 6 variables indépendantes et de même loi que X. En l'absence d'erreur systématique, $\overline{X_6}$ suit une loi $\mathcal{N}\left(\mu, \left(\frac{0.05}{\sqrt{6}}\right)^2\right)$.
   On a observé $\overline{x_6} = 2.98$.
   Pour un coefficient de sécurité de 95%, on peut parier sur l'inégalité :
   $$|2.98 - \mu| < 1.96\frac{0.05}{\sqrt{6}}$$
   D'où l'intervalle de confiance : $2.94 < \mu < 3.02$ mg/litre.
5. On a successivement :
   $$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \int_0^{+\infty} \lambda^2 x e^{-\lambda x}\mathrm{d}x = \left[-\lambda x e^{-\lambda x} - e^{-\lambda x}\right]_0^{+\infty} = 1$$
   $$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = \int_0^{+\infty} \lambda^2 x^2 e^{-\lambda x}\mathrm{d}x = \left[-\lambda x^2 e^{-\lambda x}\right]_0^{+\infty} + \frac{2}{\lambda}\int_0^{+\infty} \lambda^2 x e^{-\lambda x}\mathrm{d}x = \frac{2}{\lambda}$$
   $$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \int_0^{+\infty} \lambda^2 x^3 e^{-\lambda x}\mathrm{d}x = \left[-\lambda x^3 e^{-\lambda x}\right]_0^{+\infty} + \frac{3}{\lambda}\int_0^{+\infty} \lambda^2 x^2 e^{-\lambda x}\mathrm{d}x = \frac{6}{\lambda^2}$$

On en déduit :
$$\mathbb{E}(X) = \frac{2}{\lambda}\ ,\ \mathbb{V}(X) = \frac{6}{\lambda^2} - \frac{4}{\lambda^2} = \frac{2}{\lambda^2}$$

Soit $\overline{X}$ la moyenne de 400 variables indépendantes et de même loi que X. On peut écrire (théorème limite centrale) :

$$\overline{X} \sim \mathcal{N}\left(\frac{2}{\lambda}, \frac{2}{400\lambda^2}\right)$$

On a observé la valeur $\overline{x} = 2$. Au niveau de confiance de 90%, on peut accepter les valeur de λ telles que :

$$\left|2 - \frac{2}{\lambda}\right| < 1.645\frac{\sqrt{2}}{20\lambda} \Rightarrow \begin{cases} 2 - \frac{2}{\lambda} < 0.116\frac{1}{\lambda} \Rightarrow \lambda < 1.058 \\ \frac{2}{\lambda} - 2 < 0.116\frac{1}{\lambda} \Rightarrow \lambda > 0.942 \end{cases}$$

D'où l'intervalle de confiance $0.942 < \lambda < 1.058$.

6. Au 9-échantillon $(X_1, X_2, \ldots, X_9)$ de variables indépendantes et de même loi $\mathcal{N}(\mu, 25)$, on associe la variable $\overline{X} = \frac{1}{9}\sum_{i=1}^{9} X_i$ de loi $\mathcal{N}\left(\mu, \frac{25}{9}\right)$.
   On a observé la valeur $\overline{x} = 54.1$ mg.
   On va tester l'hypothèse $H_0\ : \mu = 50$ (ou moins) contre $H_1\ : \mu > 50$.
   Sous $H_0\ : \overline{X} \sim \mathcal{N}\left(50, \frac{25}{9}\right)$.
   Sous $H_1\ : \overline{X} \sim \mathcal{N}\left(\mu > 50, \frac{25}{9}\right)$.
   On construit donc un test unilatéral avec domaine de rejet de H_0 à droite, puisque H_1 favorise les valeurs plus élevées de $\overline{X}$.
   Le domaine d'acceptation de H_0 est $[0; 52.74[$.
   Le domaine de rejet de H_0 est $[52.74, +\infty[$.
   Au niveau 5%, on a en effet : $50 + 1.645\frac{5}{3} \approx 52.7417$.
   Il est clair, compte tenu de $\overline{x}$ que le fabriquant ne respecte pas la réglementation puisque l'on est amené à rejeter H_0.
7. (a) Si X est le nombre d'impulsions qu'on peut observer pendant 1 minute, on a : $X \sim \mathcal{P}(\mu) \approx \mathcal{N}(\mu, \mu)$ car μ a visiblement une valeur très élevée.
   Au niveau de confiance de 90%, on écrira $|6400 - \mu| < 1.645\sqrt{\mu}$, c'est-à-dire, approximativement $|6400 - \mu| < 1.645\sqrt{6400}$. D'où :
   $$6268.4 < \mu < 6531.6$$
   (b) Si $\overline{X}$ est la moyenne de n mesures indépendantes, on a, approximativement : $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\mu}{n}\right)$.
   Si $\overline{x_0}$ est la valeur observé de $\overline{X}$, on a, au niveau de 90% : $|\overline{x_0} - \mu| < 1.645\sqrt{\frac{\mu}{n}}$, soit pratiquement $|\overline{x_0} - \mu| < 1.645\sqrt{\frac{\overline{x_0}}{n}}$.
   On veut $1.645\sqrt{\frac{\overline{x_0}}{n}} < 10$. Pour $\overline{x_0} \approx 6400$, on en déduit $\sqrt{n} > 13.16 \Rightarrow n >$ 174 comptages.

(c) Si S_n est la somme de n variables de même loi que X et indépendantes, $\S_n \sim \mathcal{P}(n\mu) \approx \mathcal{N}(n\mu, n\mu)$.

S_n est aussi le résultat d'un comptage de n minutes ; on a donc $S_n = \sum_{i=1}^{n} X_i = n\overline{X}$.

Il est équivalent d'utiliser $n\overline{X} \sim \mathcal{N}(n\mu, n\mu)$ ou $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\mu}{n}\right)$.

8. La première capture a permis de créer une population à deux catégories d'individus. Toute capture ultérieur d'un individu aura deux résultats possibles : il est marqué, et cette éventualité a la probabilité $p = \frac{800}{N}$, ou il nest pas marqué, et cette éventualité a la probabilité $q = 1 - p = 1 - \frac{800}{N}$.

Si X est le nombre d'individus marqués que l'on peut observer sur 1000 individus, et si le schéma de tirage non exhaustif est respecté, on a : $X \sim \mathcal{B}(1000, p)$. Vu le nombre d'observations, on utilise l'approximation normale. Comme il est plus commode, ici d'utiliser la proportion $Y = \frac{X}{1000}$ des individus marqués, on écrira :

$$Y \sim \mathcal{N}\left(p \frac{p(1-p)}{1000}\right)$$

On a observé effectivement la proportion $y_0 = \frac{250}{1000} = \frac{1}{4}$. Au niveau de 90%, on acceptera les valeurs de p telles que

$$\left|\frac{1}{4} - p\right| < 1.645\sqrt{\frac{p(1-p)}{1000}}$$

soit, approximativement, telles que : $\left|\frac{1}{4} - p\right| < 1.645\sqrt{\frac{1}{4}\frac{1}{3}\frac{1}{1000}}$

D'où l'intervalle de confiance de p pour un coefficient de sécurité : $0.227 < p < 0.273$.

On remarque que l'approximation normale se justifie puisque pour toutes les valeurs p de cet intervalle, on a : $1000q > 1000p > 227 > 10$.

Puisque l'on sait que $p = \frac{800}{N}$, on a :

$$0.227 < \frac{800}{N} < 0.273$$

D'où l'intervalle de confiance de N : $2930 < N < 3524$

Chapitre 9

Les modèles gaussiens et leur utilisation à l'étude des moyennes et des variances

1 Les lois du chi-deux

Définition 9.1 *On appelle variable chi-deux à n degrés de liberté une variable aléatoire somme des carrés de n variables normales centrées réduites indépendantes. On note χ_n^2 une telle variable.*
On a : $\chi_n^2 = \sum_{i=1}^{n} X_i^2$, où $X_1, X_2, \ldots, X_n$ sont des variables indépendantes, telles que $X_i \sim \mathcal{N}(0,1)\ \forall i \in \{1, 2, \ldots, n\}$

1.1 Densité de probabilité

On montre que la densité de probabilité f_n d'une variable $Y = \chi_n^2$ a pour expression

$$f_n : \begin{cases} y \mapsto 0 & , \text{ si } y \leq 0 \\ y \mapsto c_n y^{\frac{n-2}{2}} e^{-y/2} & , \text{ si } y > 0 \end{cases}$$

c_n étant une constante positive dépendant de n.

Exemple 9.1 *On a montré (voir exemple 4.3) que la densité de probabilité de la variable $Y = \chi_1^2$ est :*

$$f_1 : \begin{cases} y \mapsto 0 & , \text{ si } y \leq 0 \\ y \mapsto \frac{1}{\sqrt{2\pi y}} e^{-y/2} & , \text{ si } y > 0 \end{cases}$$

Exemple 9.2 *On cherche à déterminer F_2, puis f_2, fonction de répartition et fonction densité de la variable $Y = \chi_2^2$. On a : $Y = X_1^2 + X_2^2$, X_1 et X_2 étant deux variables normales réduites indépendantes. Le couple (X_1, X_2) a pour densité :*

$$g : (x_1, x_2) \mapsto \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2 + x_2^2)} , (x_1, x_2) \in \mathbb{R}^2$$

car produit des fonctions densité de X_1 et X_2.

On a

$$\begin{array}{rcl} F_2(y) = \mathbb{P}(Y < y) & = & \mathbb{P}(X_1^2 + X_2^2 < y) = 0 \textit{ si } y \leq 0 \\ & = & \displaystyle\int\int_C \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2+x_2^2)} dx_1 dx_2 \textit{ si } y > 0 \end{array}$$

Le domaine C étant défini par $x_1^2 + x_2^2 < y$ et donc représenté par l'ensemble des points à l'intérieur d'un cercle de rayon $\sqrt{y}$.
En passant en coordonnées polaires, on obtient pour $y > 0$:

$$F_2(y) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_0^{\sqrt{y}} e^{-\frac{1}{2}r^2} r dr = 1 - e^{-\frac{1}{2}y}$$

D'où la densité :

$$f_2 : \left\{ \begin{array}{lcl} y \mapsto 0 & , & \textit{si } y \leq 0 \\ y \mapsto \frac{1}{2} e^{-y/2} & , & \textit{si } y > 0 \end{array} \right.$$

Exemple 9.3 *La variable aléatoire de densité de probabilité :*

$$f_4 : \left\{ \begin{array}{lcl} y \mapsto 0 & , & \textit{si } y \leq 0 \\ y \mapsto \frac{y}{4} e^{-y/2} & , & \textit{si } y > 0 \end{array} \right.$$

est une variable $Y = \chi_4^2$.

Exemple 9.4 *La variable aléatoire de densité de probabilité :*

$$f_6 : \left\{ \begin{array}{lcl} z \mapsto 0 & , & \textit{si } z \leq 0 \\ z \mapsto \frac{z^2}{16} e^{-z/2} & , & \textit{si } z > 0 \end{array} \right.$$

est une variable χ_6^2.

1.2 *Propriété d'additivité*

Soient deux variables χ_n^2 et χ_p^2 indépendantes. On a $Z = \chi_n^2 + \chi_p^2 = \chi_{n+p}^2$.
Il est facile de démontrer cette propriété à partir de la forme générale de la densité d'un χ^2. On la vérifiera sur des cas particuliers. Il résulte de cette propriété que toute variable suivant un loi de χ_n^2 peut être considérée comme la somme de n variables χ_1^2 indépendantes.

1.3 *Espérance et variance*

Soit une variable $X \sim \mathcal{N}(0,1)$ et soit une variable $Y = X^2$. On peut écrire directement (voir exemple 4.3) :
$\mathbb{E}(Y) = \mathbb{E}(X^2) = \mathbb{V}(X) = 1$ (puisque $\mathbb{E}(X) = 0 \Rightarrow \mathbb{E}(X^2) = \mathbb{V}(X)$)
$\mathbb{E}(Y^2) = \mathbb{E}(X^4) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^4 e^{-x^2/2} \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \left[-x^3 e^{-x^2/2} \right]_{-\infty}^{+\infty} + \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-x^2/2} \mathrm{d}x = 0 + 3\mathbb{E}(X^2) = 3\mathbb{V}(X) = 3$

On en déduit $\mathbb{V}(Y) = \mathbb{E}(Y^2) - \mathbb{E}^2(Y) = 3 - 1 = 2$

$$\mathbb{E}(\chi_1^2) = 1\,,\,\mathbb{V}(\chi_1^2) = 2$$

On aboutira au même résultat à partir de la densité f_1 de la variable Y (voir exemple 9.1) en effectuant le changement de variable $x = \sqrt{y}$.
Soient maintenant n variables indépendantes Y_i ($i \in \{1, 2, \ldots, n\}$) ayant une loi de χ_1^2.
On déduit de la propriété d'additivité (théorème 7.1) :

$$\mathbb{E}(\chi_n^2) = \mathbb{E}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \mathbb{E}\left(Y_i\right) = n\mathbb{E}(\chi_n^2) = n \times 1 = n$$

À cause de l'indépendance des variables (théorème 7.2) :

$$\mathbb{V}(\chi_n^2) = \mathbb{V}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \mathbb{V}\left(Y_i\right) = n\mathbb{V}(\chi_1^2) = n \times 2 = 2n$$

D'où le résultat général :

$$\mathbb{E}(\chi_n^2) = n\,,\,\mathbb{V}(\chi_n^2) = 2n$$

1.4 Formes de la distribution

La courbe représentative de la densité d'un χ_n^2 a une forme en L pour $n = 1$ et pour $n = 2$. Il suffit d'étudier les variations des fonctions f_1 et f_2 (voir exemples 9.1 et 9.2) pour s'en rendre compte.
Pour $n > 2$, on a :

$$(f_n(y))' = \frac{1}{2} c_n e^{-y/2} y^{\frac{n-4}{2}} [(n-2) - y]$$

On en déduit que la courbe représentative de f_n a une forme en cloche, dissymétrique et qu'une variable χ_n^2 a pour mode $n - 2$ ($n > 2$).
On note, en particulier que plus la valeur de n est élevée, plus le mode se déplace vers la droite.

1.5 Convergence en loi

Puisqu'une variable χ_n^2 peut être considérée comme la somme de n variables χ_1^2 indépendantes, il résulte du théorème central limite que la loi d'un χ_n^2 tend vers une loi normale lorsque n augmente indéfiniment.
Cependant, dans le cas d'une variable χ_n^2, cette convergence est assez lente. Par contre on peut noter que la variable $\sqrt{2\chi_n^2}$ a une loi qui converge assez rapidement vers une loi $\mathcal{N}(\sqrt{2n-1}, 1)$.
Dans la pratique, pour $n > 30$, on écrit :

$$\sqrt{2\chi_n^2} \sim \mathcal{N}(\sqrt{2n-1}, 1)$$

1.6 *Calcul de probabilités*

Comme pour toutes les variables d'usage courant, les fonctions de répartitions des variables χ^2_n sont tabulées.
Soit la relation $F(u) = \mathbb{P}(\chi^2_n < u) = p$. La table 11.10 donne, pour un certain nombre de valeurs p, u en fonction de n.

Exemple 9.5 • *Soit une variable $Y = \chi^2_4$. On a*

$$\mathbb{P}(0.484 < Y < 11.143) = F(11.143) - F(0.484) = 0.975 - 0.025 = 95\%$$

• *Soit une variable $Z = \chi^2_{10}$. On a*

$$\mathbb{P}(Z \geq 18.307) = 1 - \mathbb{P}(Z < 18.307) = 1 - F(18.307) = 1 - 0.95 = 5\%$$

On remarquera que, la densité d'une variable χ^2_n étant nulle en zéro et à gauche de zéro, on a :

$$\begin{aligned} F(u) = 0 \quad &, \quad \text{si } u \leq 0 \\ F(u) = \mathbb{P}(\chi^2_n < u) = \mathbb{P}(0 < \chi^2_n < u) \quad &, \quad \text{si } u > 0 \end{aligned}$$

On remarquera également que si $X \sim \mathcal{N}(0,1)$ et si $Y = X^2 \sim \chi^2_1$, on a :

$$\mathbb{P}(Y < u) = \mathbb{P}(X^2 < u) = \mathbb{P}(-\sqrt{u} < X < \sqrt{u})\text{ , } u > 0$$

En particulier on pourra vérifier :

$$\begin{aligned} \mathbb{P}(-1.960 < X < 1.960) = \mathbb{P}(Y < (1.960)^2) &= \mathbb{P}(Y < 3.841) = 95\% \\ \mathbb{P}(-2.576 < X < 2.576) = \mathbb{P}(Y < (2.576)^2) &= \mathbb{P}(Y < 6.635) = 99\% \end{aligned}$$

enfin, on notera que dans la table 11.10, le nombre de degrés de liberté d'une variable χ^2 ne dépasse pas 30. Pour $n > 30$, on utilisera l'approximation $\sqrt{2\chi^2_n} \sim \mathcal{N}(\sqrt{2n-1}, 1)$
Pour évaluer la précision du procédé, on va comparer, pour $n = 30$, les probabilités vraies, lues dans la table, avec celles obtenues par l'approximation. Pour $Y = \chi^2_{30}$, on a :

- $\mathbb{P}(Y < 46.979) = 0.975$
  Or $\mathbb{P}(Y < 46.979) = \mathbb{P}(2Y < 93.958) = \mathbb{P}(\sqrt{2Y} < 9.693)$
  $\sqrt{2Y} \sim \mathcal{N}(\sqrt{59}, 1) \Rightarrow \sqrt{2Y} - \sqrt{59} \sim \mathcal{N}(0,1)$
  $\Rightarrow \mathbb{P}(\sqrt{2Y} < 9.693) = \mathbb{P}\left(\frac{\sqrt{2y}-\sqrt{59}}{1} < 2.01\right) \approx 0.9778$
- $\mathbb{P}(Y < 29.336) = 0.50$
  Or $\mathbb{P}(Y < 29.336) = \mathbb{P}(2Y < 58.672) = \mathbb{P}(\sqrt{2Y} < 7.660))$
  $= \mathbb{P}\left(\frac{\sqrt{2y}-\sqrt{59}}{1} < -0.0213\right) \approx 1 - 0.5085 \approx 0.4915$
- $\mathbb{P}(Y < 16.791) = 0.025$
  Or $\mathbb{P}(Y < 16.791) = \mathbb{P}(2Y < 33.582) = \mathbb{P}(\sqrt{2Y} < 5.795)$
  $= \mathbb{P}\left(\frac{\sqrt{2y}-\sqrt{59}}{1} < -1.886\right) \approx 1 - 0.9703 \approx 0.0297$

Les écarts sont tous inférieurs à 1%.

2 Les lois de Fisher-Snedecor

Définition 9.2 *Soient deux variables aléatoires indépendantes : $X = \chi^2_n$ et $Y = \chi^2_p$. La variable $Z = \frac{X/n}{Y/p} = \frac{p}{n}\frac{X}{Y}$ est dite variable de Fisher-Snedecor à n et p degrés de liberté. On note : $Z = F(n,p) = \frac{\chi^2_n/n}{\chi^2_p/p} = \frac{p}{n}\frac{\chi^2_n}{\chi^2_p}$.*

2.1 *Densité de probabilité*

On montre que la densité de probabilité $g_{n,p}$ d'une variable $Z = F(n,p)$ a pour expression

$$g_{n,p} : \begin{cases} z \mapsto 0 & , \text{ si } z \leq 0 \\ z \mapsto c_{n,p} z^{\frac{n-2}{2}} (p+nz)^{-\frac{n+p}{2}} & , \text{ si } z > 0 \end{cases}$$

$c_{n,p}$ étant une constante positive dépendant de n et de p.
il est normal que la densité $g_{n,p}$ soit nulle en zéro et à gauche de zéro ; l'événement "$F(n,p) < u$" est en effet impossible si $u \leq 0$, puisque dans ce cas, pour tout χ^2, on a : $\mathbb{P}(\chi^2 < u) = 0$

2.2 *Espérance et variance*

On montre également qu'une variable $Z = F(n,p)$ a pour espérance :

$$\mathbb{E}[F(n,p)] = \frac{p}{p-2} \text{ , définie pour } p > 2$$

et pour variance :

$$\mathbb{V}[F(n,p)] = \left(\frac{p}{p-2}\right)^2 \frac{2(n+p-2)}{n(p-4)} \text{ , définie pour } p > 4$$

2.3 *Forme de la distribution*

L'étude des variations de la fonction densité d'une variable $F(n,p)$ montre que la forme de sa courbe représentative est :

- en L pour $n = 1$ et pour $n = 2$;
- en cloche, dissymétrique pour $n > 2$. Dans ce cas le mode est égal à : $\frac{p(n-2)}{n(p+2)}$ pour $n > 2$.

2.4 *Calcul de probabilités*

Soit la relation :

$$G(u) = \mathbb{P}[F(n,p) < u] = p \tag{9.1}$$

Il est clair que, donner une table de valeurs numériques concernant les fonctions de répartition G des variables $F(n,p)$, conduit à envisager un certain nombre de couples (n,p) et donc à donner un véritable fascicule uniquement pour ce type de variables.
En statistique, certaines valeurs de p sont fondamentales. Pour cette raison, les tables 11.11 et 11.12 correspondent chacune à une valeur de p et indiquent pour divers couples (n,p) (donc tables à doubles entrées) les valeurs u (forcément positives) vérifiant 9.1.

Exemple 9.6 • *Dans la table 11.11, on lit, puisque $p = 0.95$:*

$$\mathbb{P}[F(9,7) < 3.6767] = 0.95\,;\ \mathbb{P}[F(15,14) < 2.4630] = 0.95$$

• *Dans la table 11.12, on lit, puisque $p = 0.99$:*

$$\mathbb{P}[F(9,7) < 6.7188] = 0.99\,;\ \mathbb{P}[F(15,14) < 3.6567] = 0.99$$

• *On note que :*

$$\mathbb{P}[F(n,p) < u] = \mathbb{P}\left(\frac{\chi^2_n/n}{\chi^2_p/p} < u\right) = \mathbb{P}\left(\frac{n\chi^2_p}{p\chi^2_n} > \frac{1}{u}\right) = 1-\mathbb{P}\left(\frac{n\chi^2_p}{p\chi^2_n} \leq \frac{1}{u}\right) = 1-\mathbb{P}\left(F(p,n) \leq \frac{1}{u}\right)$$

On a donc :

- $\mathbb{P}(F(n,p) < u) = 0.01 \Leftrightarrow \mathbb{P}\left(F(p,n) \leq \frac{1}{u}\right) = 0.99$
- $\mathbb{P}(F(n,p) < u) = 0.05 \Leftrightarrow \mathbb{P}\left(F(p,n) \leq \frac{1}{u}\right) = 0.95$
- *Par exemple :* $\mathbb{P}[F(9,7) < 3.6767] = 0.95 \Leftrightarrow \mathbb{P}\left[F(7,9) < \frac{1}{3.6767}\right] = 0.05$

3 Application à l'étude des variances

3.1 *Problèmes à un échantillon*

On considère un ensemble $(X_1, X_2, \ldots, X_n)$, de n variables indépendantes et de même loi ; on notera μ et σ^2 leur espérance et leur variance. La réalisation de l'expérience $\mathcal{E}$ à laquelle est associé cet ensemble entraîne les observations : $x_1, x_2, \ldots, x_n$.
On rappelle que sur un tel n-échantillon, on peut définir les variables $\overline{X}$ (moyenne), S^2 (variance), $\widehat{S^2}$ (estimateur de σ^2). On a :

$$\left\{\begin{array}{lll} \overline{X} = \frac{1}{n}\sum_{i=1}^n X_i & \text{, dont la valeur observée est la moyenne empirique} & \overline{x_0} = \frac{1}{n}\sum_{i=1}^n x_i \\ S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 & \text{, dont la valeur observée est la variance empirique} & s_0^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x_0})^2 \\ \widehat{S^2} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 & \text{, dont la valeur observée est l'estimation de } \sigma^2 & \widehat{s_0^2} = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x_0})^2 \end{array}\right.$$

On suppose maintenant que les variables $X_1, X_2, \ldots, X_n$ sont toutes des variables normales $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $\forall i \in \{1, 2, \ldots, n\}$

Premier résultat

Dans le cadre de ce modèle, il est alors clair que la variable $\sum_{i=1}^n \left(\frac{X_i-\mu}{\sigma}\right)^2$ est la somme des carrés de n variables normales centrées réduites, indépendantes. On peut donc écrire :

$$\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2_n$$

Ce résultat n'est guère utilisable pratiquement car le modèle obtenu dépend des deux paramètres μ et σ^2, difficulté que l'on a déjà rencontrée dans l'utilisation de la variable $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

Deuxième résultat

On peut écrire : $X_i - \mu = (X_i - \overline{X}) + (\overline{X} - \mu)$
D'où :

$$\begin{aligned} (X_i - \mu)^2 &= (X_i - \overline{X})^2 + (\overline{X} - \mu)^2 + 2(X_i - \overline{X})(\overline{X} - \mu) \\ \sum_{i=1}^{n}(X_i - \mu)^2 &= \sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X} - \mu)^2 + 2(\overline{X} - \mu)\sum_{i=1}^{n}(X_i - \overline{X}) \end{aligned}$$

Or il est clair que $\sum_{i=1}^{n}(X_i - \overline{X}) = \sum_{i=1}^{n} X_i - n\overline{X} = n\overline{X} - n\overline{X} = 0$
Et donc :

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 + \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)^2$$

Or on vient de voir que $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2_n$.
De plus $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2_1$
En posant $Z = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2$ on a donc $\chi^2_n = Z + \chi^2_1$
Un théorème, dit théorème de Cochran, (qui sort du cadre du cours) a pour conséquence importante de démontrer que les variables $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$ et $\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$ sont indépendantes.
La propriété d'additivité des χ^2 implique alors naturellement que $Z \sim \chi^2_{n-1}$.
On obtient ainsi le résultat, important car on dispose maintenant d'un modèle qui ne dépend que du paramètre σ^2 :
La variable $Z = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{nS^2}{\sigma^2} = \frac{(n-1)\widehat{S^2}}{\sigma^2}$ est une variable χ^2_{n-1}.
Il devient alors possible de construire des interprétations statistiques concernant le paramètre σ^2. On remarquera que la variable $nS^2 = (n-1)\widehat{S^2}$ est simplement la somme des carrés des écarts à $\overline{X}$.

4 Test d'hypothèse sur σ^2

On se propose de tester l'hypothèse $H_0 \ : \sigma^2 = a^2$ contre l'une des trois hypothèses alternatives suivantes :

$$H_1 \ : \sigma^2 > a^2 \ \ ; \ \ H_2 \ : \sigma^2 < a^2 \ \ ; \ \ H_3 \ : \sigma^2 \neq a^2$$

Sous H_0, la variable $Z = \frac{nS^2}{a^2}$ suit une loi de χ^2_{n-1}.

Sous $H_1 \ : \sigma^2 > a^2$ La variance S^2 aura tendance à prendre des valeurs plus élevées que sous H_0. Ce sont donc les valeurs élevées de $\frac{nS^2}{a^2}$ qui poussent à rejeter H_0 pour choisir H_1. On construit donc un test unilatéral avec un domaine de rejet à droite.

La borne supérieure d du domaine d'acceptation est telle que : $\mathbb{P}(\chi^2_{n-1} < d) = 1 - \alpha$, si α est le niveau choisi.

La valeur d est lue dans la table 11.10 et, pour effectuer le test, il suffit de regarder si la valeur observée $\frac{ns_0^2}{a^2}$, de la variable $\frac{nS^2}{a^2} \sim \chi^2_{n-1}$, se trouve dans le domaine d'acceptation ou dans le domaine de rejet de H_0.

Sous $H_2 \ : \sigma^2 < a^2$ La variance S^2 aura tendance à prendre des valeurs plus faibles que sous H_0. Ce sont donc les valeurs faibles de $\frac{nS^2}{a^2}$ qui poussent à rejeter H_0 pour choisir H_1. On construit donc un test unilatéral avec un domaine de rejet à gauche.

La borne inférieure c est lue dans la table 11.10, table des valeurs numériques de la fonction de répartition de la variable $\frac{nS^2}{a^2} \sim \chi^2_{n-1}$. On a $\mathbb{P}(\chi^2_{n-1} < c) = \alpha$, si α est le niveau choisi.

Sous $H_3 \ : \sigma^2 \neq a^2$ Il est alors clair que l'on est amené à construire un test bilatéral, le domaine de rejet étant constitué d'une partie à droite et d'une partie à gauche de la distribution.

si α est le niveau choisi, les bornes c et d de l'intervalle d'acceptation sont telles que :

$$\mathbb{P}(\chi^2_{n-1} < c) = \frac{\alpha}{2} \quad , \quad \mathbb{P}(\chi^2_{n-1} < d) = 1 - \frac{\alpha}{2}$$

Ces valeurs sont lues dans la table 11.10, table des valeurs numériques de la fonction de répartition de la variable $\frac{nS^2}{a^2} = \chi^2_{n-1}$.

Exemple 9.7 *Un laboratoire pharmaceutique fabrique un médicament dont la teneur, en un certain composant par unité de fabrication, est assurée avec un écart-type de 8 milligrammes.*
Le service de recherche a mis au point un nouveau procédé de fabrication qui sera adopté s'il assure une réduction substantielle de la dispersion. On a fait 10 mesures de teneur sur des unités fabriquées par la nouvelle méthode et obtenus les résultats suivants, exprimés en milligrammes : 725 ; 722 ; 727 ; 718 ; 723 ; 731 ; 719 ; 724 ; 726 ; 725. On suppose que chaque mesure est une réalisation d'une variable normale. Peut-on adopter la nouvelle méthode ?
Soit X_i la variable représentant une mesure. On a $X_i \sim \mathcal{N}(\mu, \sigma^2)$. À un ensemble de 10 mesures, on peut associer les variables $\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$ et $S^2 = \frac{1}{10}\sum_{i=1}^{10}(X_i - \overline{X})^2$. On sait alors que la variable $\frac{10S^2}{\sigma^2}$ obéit à une loi de χ^2_9.
On veut tester l'hypothèse $H_0 \ : \sigma^2 = 64$ mg^2 contre l'hypothèse $H_1 \ : \sigma^2 < 64$ mg^2. Sous H_0, la variable $Z = \frac{10S^2}{64}$ obéit à une loi de χ^2_9. Si $\sigma^2 < 64$, la variable S^2, et donc la variable $Z = \frac{10S^2}{64}$, aura tendance à prendre des valeurs moins élevées que prévu, puisque les variables X_i sont moins dispersées. On est donc amené à construire un test unilatéral avec domaine de rejet à gauche ; on choisira un niveau de 5%. La borne supérieure c du domaine de rejet de H_0 doit vérifier $\mathbb{P}(Z < c) = 0.05$; on en déduit, puisque $Z \sim \chi^2_9$ $c = 3.325$ (table 11.10).
Or on a observé les valeurs :

- $\overline{x_0} = \frac{1}{10}\sum_{i=1}^{10} x_i = 724$ *mg pour* $\overline{X}$
- $s_0^2 = \frac{1}{10}\sum_{i=1}^{10}(x_i - \overline{x_0})^2 = \frac{1}{10}130$ *mg*^{2} *pour* S^2
- *et, sous* H_0*, la valeur* $z_0 = \frac{130}{64} \approx 2.03$ *de la variable* Z*.*

Au niveau 5%, on rejette H_0*. On adoptera donc la nouvelle méthode de fabrication, reconnue plus précise que l'ancienne.*

4.1 Intervalle de confiance d'une variance

Il est facile de déduire du modèle obtenu une estimation par intervalle de σ^2, à partir de l'observation s_0^2. L'intervalle de confiance sera constitué des valeurs de σ^2 qui seraient acceptées dans un test d'hypothèse bilatéral. D'où la méthode :

- on sait que la variable $\frac{nS^2}{\sigma^2}$ obéit à une loi χ^2_{n-1} ;
- on détermine les valeurs c et d telles que $\mathbb{P}(\chi^2_{n-1} < c) = \frac{\alpha}{2}$ et $\mathbb{P}(\chi^2_{n-1} < d) = 1 - \frac{\alpha}{2}$;
- On a observé la valeur s_0^2 de S^2. On parie qu'elle vérifie l'inégalité $c < \frac{ns_0^2}{\sigma^2} < d$, d'où on tire l'intervalle de confiance de σ^2 pour le coefficient de sécurité $1 - \alpha$.

Exemple 9.8 *(suite de l'exemple 9.7). On se propose maintenant de procéder à une estimation par intervalle de la variance obtenue par la nouvelle méthode, afin de pouvoir évaluer sa précision.*
La variable $\frac{10S^2}{\sigma^2}$ *obéit à une loi de* χ^2_9 *; on écrit que la valeur observée* $\frac{130}{\sigma^2}$ *vérifie la double inégalité :* $2.700 < \frac{130}{\sigma^2} < 19.023$*, au niveau de confiance de 95%.*
D'où l'intervalle de confiance : $6.83 < \sigma^2 < 48.15$

4.2 Problèmes à deux échantillons

Soit un ensemble $(X_1, X_2, \ldots, X_n)$ de n variables indépendantes, de même loi normale $\mathcal{N}(\mu_X, \sigma_X^2)$. Soit un deuxième ensemble $(Y_1, Y_2, \ldots, Y_p)$ de p variables indépendantes, de même loi normale $\mathcal{N}(\mu_Y, \sigma_Y^2)$. Les deux ensembles sont supposés indépendants entre eux. On note :

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad S_Y^2 = \frac{1}{p}\sum_{i=1}^{p}(Y_i - \overline{Y})^2$$

$$\widehat{S_X^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad \widehat{S_Y^2} = \frac{1}{p-1}\sum_{i=1}^{p}(Y_i - \overline{Y})^2$$

On sait que, dans le cadre des hypothèses ci-dessus :

$$\frac{nS_X^2}{\sigma_X^2} = \frac{(n-1)\widehat{S_X^2}}{\sigma_X^2} \sim \chi^2_{n-1} \quad , \quad \frac{pS_Y^2}{\sigma_Y^2} = \frac{(p-1)\widehat{S_Y^2}}{\sigma_Y^2} \sim \chi^2_{p-1}$$

On en déduit :

$$F(n-1, p-1) = \frac{p-1}{n-1}\frac{\chi^2_{n-1}}{\chi^2_{p-1}} = \frac{p-1}{n-1}\frac{nS_X^2}{pS_Y^2}\frac{\sigma_Y^2}{\sigma_X^2} = \frac{\sigma_Y^2}{\sigma_X^2}\frac{\widehat{S_X^2}}{\widehat{S_Y^2}}$$

D'où le résultat important :
La variable $\frac{\sigma_Y^2}{\sigma_X^2}\frac{\widehat{S_X^2}}{\widehat{S_Y^2}}$ obéit à une loi $F(n-1,p-1)$.
Il devient alors possible de construire des interprétations statistiques concernant le rapport $\frac{\sigma_Y^2}{\sigma_X^2}$

4.3 *Comparaison de deux variances. Test F*

On se propose de tester l'hypothèse $H_0\ :\ \sigma_X^2 = \sigma_Y^2$ contre l'une des trois hypothèses alternatives suivantes :

$$H_1\ :\sigma_X^2 > \sigma_Y^2\ \ ;\ \ H_2\ :\sigma_X^2 < \sigma_Y^2\ \ ;\ \ H_3\ :\sigma_X^2 \neq \sigma_Y^2$$

Dans le cadre du modèle qui vient d'être décrit, $\frac{\sigma_Y^2}{\sigma_X^2}\frac{\widehat{S_X^2}}{\widehat{S_Y^2}}$ suit une loi $F(n-1,p-1)$, et par conséquent :

Sous H_0, $Z = \frac{\widehat{S_X^2}}{\widehat{S_Y^2}}$ suit une loi de $F(n-1,p-1)$.

Sous $H_1\ :\sigma_X^2 > \sigma_Y^2$ Les variables X_i étant plus dispersées, le rapport $\frac{\widehat{S_X^2}}{\widehat{S_Y^2}}$ aura tendance à prendre de préférence des valeurs plus élevées que sous H_0. Ce sont donc les valeurs élevées de Z qui incitent à choisir H_1 et à rejeter H_0. On construit donc un test unilatéral avec domaine de rejet à droite. L'intervalle d'acceptation de H_0 est donc du type $[0,b[$, b étant déterminée, au niveau α, par :

$$\mathbb{P}[F(n-1,p-1) < b] = 1-\alpha$$

Pour effectuer le test, il suffit de regarder où est située la valeur observée $z_0 = \frac{\widehat{S_X^2}}{\widehat{S_Y^2}}$.

Sous $H_2\ :\sigma_X^2 < \sigma_Y^2$ le rapport $\frac{\widehat{S_X^2}}{\widehat{S_Y^2}}$ aura tendance à prendre de préférence des valeurs plus faibles que sous H_0. On construit donc un test unilatéral avec domaine de rejet à gauche. L'intervalle d'acceptation de H_0 est donc du type $[a,+\infty[$. En fait il est bien plus simple de se ramener au cas précédent en inversant le rapport des estimateurs.
La variable $V = \frac{\widehat{S_Y^2}}{\widehat{S_X^2}}$ obéit de toute évidence à une loi $F(p-1,n-1)$.
On a "$Z > a$"="$\frac{1}{Z} < \frac{1}{a}$"="$V < b = \frac{1}{a}$" et $\mathbb{P}(Z > a) = 1-\alpha = \mathbb{P}(V < b = {}^1\!/_a)$
On peut donc énoncer la règle : dans un test F unilatéral, on met au numérateur l'estimateur qui correspond à la variance la plus forte dans l'hypothèse alternative. L'intervalle d'acceptation est alors du type $[0,b[$ avec $\mathbb{P}[F(\nu_1,\nu_2) < b] = 1-\alpha$, ν_1 étant le nombre de degrés de liberté associé au numérateur.

Sous $H_3\ :\sigma_X^2 \neq \sigma_Y^2$ le rapport $\frac{\widehat{S_X^2}}{\widehat{S_Y^2}}$ aura tendance à prendre soit des valeurs plus faibles, soit des valeurs plus fortes que sous H_0. On construit donc un test bilatéral. L'intervalle d'acceptation de H_0 est donc du type $]a,b[$ avec, au niveau α :

$$\mathbb{P}[F(n-1,p-1) < b] = 1-\frac{\alpha}{2} \ \text{ et }\ \mathbb{P}[F(n-1,p-1) < a] = \frac{\alpha}{2}$$

Pour effectuer le test, il suffit de regarder où est située la valeur observée $z_0 = \frac{\widehat{s_X^2}}{\widehat{s_Y^2}}$.
La valeur b est lue dans la table $1-\frac{\alpha}{2}$ pour $F(n-1, p-1)$. On obtient $\frac{1}{a}$ dans la table $1-\frac{\alpha}{2}$ pour $F(n-1, p-1)$.
En réalité a est toujours inférieur à 1 et b toujours supérieur à 1 (voir tables 11.11 et 11.12). Donc si le rapport des estimations $z_0 = \frac{\widehat{s_X^2}}{\widehat{s_Y^2}}$ est supérieur à 1, on sait que $z_0 > a$ et il suffit de comparer z_0 à b.
D'où la règle : dans un test bilatéral, on met au numérateur l'estimation la plus forte, de manière à obtenir un quotient observé z_0 supérieur à 1. L'intervalle d'acceptation de H_0 étant du type $]a, b[$, on sait alors que $a < z_0$ et il suffit de comparer z_0 à b ; on a : $\mathbb{P}[F(\nu_1, \nu_2) < b] = 1-\frac{\alpha}{2}$, ν_1 étant le nombre de degrés de liberté associé au numérateur.

Exemple 9.9 *Des dosages de calcium sur deux échantillons de végétaux ont donné les résultats suivants, exprimés en mg. de calcium :*

	Effectif	*Moyenne empirique*	*Variance empirique*
Échantillon A	*11*	*3.92*	*0.3130*
Échantillon B	*9*	*4.18*	*0.4231*

On admet que chaque mesure est réalisation d'une variable aléatoire $\mathcal{N}(\mu_X, \sigma_X^2)$ pour A et $\mathcal{N}(\mu_Y, \sigma_Y^2)$ pour B. Peut-on conclure à une différence entre les variances σ_X^2 et σ_Y^2 ?
On va tester l'hypothèse $H_0 : \sigma_X^2 = \sigma_Y^2$ contre $H_1 : \sigma_X^2 \neq \sigma_Y^2$ (test bilatéral).
Soit $\widehat{S_X^2}$ et $\widehat{S_Y^2}$ les estimateurs de σ_X^2 et σ_Y^2. On a $Z = \frac{\widehat{S_X^2}}{\widehat{S_Y^2}} \sim F(8, 10)$.
Or on a observé les valeurs $\widehat{s_X^2} = \frac{11}{10}0.3130 \approx 0.3443$ et $\widehat{s_Y^2} = \frac{9}{8}0.4231 \approx 0.4760$, donc $z_0 = \frac{0.4760}{0.3443} \approx 1.38$.
$\mathbb{P}[F(8, 10) < 3.8549] = 0.975 \Rightarrow 1.38 < 3.8549$
Au niveau 5%, l'hypothèse de l'égalité des variances est acceptable.

5 Les lois de Student

Définition 9.3 *Soit deux variables aléatoires indépendantes : une variable normale centrée réduite, $U \sim \mathcal{N}(0, 1)$ et une variable chi-deux à n degrés de liberté, $V \sim \chi_n^2$. On appelle variable de Student à n degrés de liberté, la variable :*

$$T_n = \frac{U}{\sqrt{\frac{1}{n}V}} = \frac{U\sqrt{n}}{\sqrt{V}}$$

5.1 *Densité de probabilité*

On montre que la densité f_n, d'une variable T_n a pour expression

$$f_n : t \mapsto \frac{c_n}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}, t \in \mathbb{R}$$

où c_n est une constante positive dépendant de n.

Exemple 9.10 *La loi de Student à un degré de liberté, appelée aussi loi de Cauchy est la loi de la variable T_1 dont la densité est*

$$f_1 \ : t \mapsto \frac{1}{\pi}\frac{1}{(1+t^2)}\ , t \in \mathbb{R}$$

Cette variable, qui n'admet pas d'espérance ni de moments d'ordre supérieur à 1 (les intégrales correspondantes divergent), est souvent citée comme exemple.

5.2 *Espérance et variance*

Le calcul de l'espérance et de la variance, lorsqu'elles existent, conduit aux résultats suivants :
$\mathbb{E}(T_n) = 0$, définie pour $n > 1$; $\mathbb{V}(T_n) = \frac{n}{n-2}$, définie pour $n > 2$.

5.3 *Forme de la distribution*

L'étude des variations de la densité d'une variable T_n, montre que la distribution est en cloche, symétrique par rapport à l'axe des ordonnées, et admet pour mode : 0.
Plus le nombre de degrés de liberté augmente, et plus la distribution d'une variable T_n est resserrée autour de l'origine.

5.4 *Convergence en loi*

La distribution d'une variable T_n, de Student, est toujours plus dispersée que celle d'une variable normale centrée réduite. Cependant on montre que la densité f_n d'une variable T_n est telle que :

$$\text{Si } n \to +\infty \quad , \quad f_n(t) \to \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

Les lois de Student convergent donc vers la loi normale centré réduite. On constate dans les tables de valeurs numérique (voir table 11.9) que la différence entre une loi de Student et une loi normale centrée réduite est, en ce qui concerne le calcul des probabilités, peu sensible lorsque $n = 30$, presque négligeable lorsque $n = 120$.

5.5 *Calcul des probabilités*

Soit $F_n(u) = \mathbb{P}(T_n < u) = p$
La table 11.9 permet d'obtenir u, pour certaines valeurs de p, selon le nombre de degrés de liberté de la variable de Student.
À cause de la symétrie par rapport à l'origine, la table n'est construite que pour des valeurs u positives.

Pour $u < 0$, on a, comme pour la loi normale centrée réduite :

$$\mathbb{P}(T_n < u) = \mathbb{P}(T_n > -u) = 1 - \mathbb{P}(T_n \leq -u) = 1 - F(-u)$$

Exemple 9.11 • *Pour* $n = 11$, *on a :* $\mathbb{P}(T_n < 0.540) = 70\%$; $\mathbb{P}(T_n < 1.796) = 95\%$; $\mathbb{P}(T_n < 2.201) = 97.5\%$
- *Pour* $n = 16$, *on a :* $\mathbb{P}(T_n < 0.691) = 75\%$; $\mathbb{P}(T_n < -0.691) = \mathbb{P}(T_n > 0.691) = 25\%$
- *Pour* $n = 25$, *on a :*

$$\begin{aligned}\mathbb{P}(|T_n| < 2.060) &= \mathbb{P}(-2.060 < T_n < 2.060)\\ &= F(2.060) - F(-2.060)\\ &= F(2.060) - [1 - F(2.060)]\\ &= 2F(2.060) - 1\\ &= 2 \times 0.975 - 1\\ &= 95\%\end{aligned}$$

6 Application à l'étude des moyennes

6.1 Problèmes à un échantillon

Comme dans l'étude des variances, on considère un ensemble $(X_1, X_2, \ldots, X_n)$ de n variables indépendantes et de même loi normale $\mathcal{N}(\mu, \sigma^2)$. On rappelle que :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad , \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad , \quad \widehat{S^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

et la relation $nS^2 = (n-1)\widehat{S^2}$
On considère maintenant la variable :

$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{S^2}{n-1}}} = \frac{\overline{X} - \mu}{\sigma}\frac{\sqrt{n-1}}{\sqrt{\frac{1}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\frac{\sqrt{n-1}}{\sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2}}}$$

On sait que $X_i \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow \overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. Donc la première fraction est une variable normale centrée réduite U.
On reconnaît de plus, sous le radical du deuxième dénominateur la variable $V = \frac{nS^2}{\sigma^2}$ qui est une variable χ^2_{n-1} (voir section 3.1).
Le théorème de Cochran (voir section 3.1) permettant d'affirmer l'indépendance entre U et V, il est clair que la variable Z obéit à la définition d'une variable de Student à $(n-1)$ degrés de liberté.

On obtient ainsi un autre résultat important car on dispose maintenant d'un modèle qui ne dépend plus que du paramètre μ :
La variable $Z = \frac{\overline{X}-\mu}{\sqrt{\frac{S^2}{n-1}}} = \frac{\overline{X}-\mu}{\sqrt{\frac{\widehat{S^2}}{n}}}$ est une variable de Student T_{n-1}
Il devient alors possible de construire des interprétations statistiques concernant le paramètre μ.

6.2 *Test d'hypothèse sur* μ

On se propose de tester l'hypothèse $H_0 \ : \mu = a$ contre l'une des trois hypothèses alternatives suivantes :

$$H_1 \ : \mu > a \ \ ; \ \ H_2 \ : \mu < a \ \ ; \ \ H_3 \ : \mu \neq a$$

Dans le cadre du modèle qui vient d'être décrit, la variable $T = \frac{\overline{X}-a}{\sqrt{\frac{S^2}{n-1}}}$ est une variable de Student à $(n-1)$ degrés de liberté.

Sous $H_1 \ : \mu > a$ Les écarts $\overline{X} - a$ auront tendance à être plus élevées que sous H_0, puisque la distribution de $\overline{X}$ est centrée autour de μ.. Ce sont donc les valeurs élevées de T qui incitent à choisir H_1 et à rejeter H_0. On construit donc un test unilatéral avec domaine de rejet à droite. L'intervalle d'acceptation de H_0 est donc du type $]-\infty, b[$, avec au niveau α : $\mathbb{P}[T < b] = 1 - \alpha$.
Pour effectuer le test, il suffit de regarder où est située la valeur observée $t_0 = \frac{\overline{x_0}-a}{\sqrt{\frac{s_0^2}{n-1}}}$

Sous $H_2 \ : \mu < a$ Les écarts faibles de $\overline{X} - a$, c'est-à-dire les écarts négatifs, seront plus probables que sous H_0. On construit donc un test unilatéral avec domaine de rejet à gauche. L'intervalle d'acceptation de H_0 est donc du type $[-b, +\infty[$ avec $\mathbb{P}[T < -b] = \alpha$. En fait on détermine b par la relation, qui découle de la symétrie de la distribution : $\mathbb{P}[T < b] = 1 - \alpha$.

Sous $H_3 \ : \mu \neq a$ Il est alors clair que l'on construit un test bilatéral, l'intervalle d'acceptation de H_0 étant du type $[-b, b]$ avec : $\mathbb{P}[T < b] = 1 - \frac{\alpha}{2}$.

Exemple 9.12 *Sur six étudiants de sexe masculin, choisis dans l'amphi, on a mesuré les tailles suivantes : 172 ; 174 ; 178 ; 180 ; 175 ; 183 cm. Interpréter ces résultats sachant que l'espérance de la taille d'un français est 170cm.*
Soit $X_i \sim \mathcal{N}(\mu, \sigma^2)$ *la taille d'un étudiant choisi au hasard. On va tester l'hypothèse* $H_0 \ : \mu = 170$ *contre l'hypothèse* $H_1 \ : \mu > 170$
Soit $\overline{X} = \frac{1}{6}\sum_{i=1}^{6} X_i$ *et soit* $s^2 = \frac{1}{6}\sum_{i=1}^{6}(X_i - \overline{X})^2$*. Les différentes mesures étant indépendantes, la variable* $T_5 = \frac{\overline{X}-170}{\sqrt{\frac{s^2}{5}}}$ *obéit à une loi de Student à 5 degrés de liberté, sous* H_0.
Ce sont les valeurs élevées de T *qui sont favorables à* H_1*. On construit donc un test unilatéral avec domaine de rejet à droite.*
Au niveau 5%, l'intervalle d'acceptation de H_0 *est* $]-\infty, 2.015]$.

On a observé les valeurs :

$$\begin{aligned}
\overline{x_0} &= 170 + \frac{1}{6}42 = 177cm \textit{ pour } \overline{X} \\
s_0^2 &= \frac{1}{6}(25+9+1+9+4+36) = 14 \textit{ pour } S^2 \\
t_0 &= \frac{177-170}{\sqrt{\frac{14}{5}}} \approx 4.18 \textit{ pour } T_5
\end{aligned}$$

Au niveau 5%, on rejette donc H_0. L'espérance de la taille d'un étudiant est supérieure à l'espérance de la taille d'un français. La moyenne observée est assez significativement différente de 170cm ; même au niveau 0.5% on rejetterait H_0.

6.3 Intervalle de confiance d'une espérance

Il est facile de déduire du modèle obtenu une estimation par intervalle de μ, à partir de l'observation $\overline{x_0}$ et $\widehat{s_0^2}$. L'intervalle de confiance, pour un coefficient de sécurité $1-\alpha$ sera constitué des valeurs de μ qui seraient acceptées dans un test d'hypothèse bilatéral au niveau α. D'où la méthode :

- on sait que la variable $\frac{\overline{X}-\mu}{\sqrt{\frac{S^2}{n-1}}}$ obéit, sous des hypothèses de normalité et d'indépendance, à une loi de Student T_{n-1} ;
- on détermine la valeur b telle que $\mathbb{P}(-b < T_{n-1} < b) = 1-\alpha$;
- On a observé les valeurs $\overline{x_0}$ et $\widehat{s_0^2}$ de la moyenne et de la variance empiriques. On parie qu'elle vérifie la double inégalité :

$$-b < \frac{\overline{x_0}-\mu}{\sqrt{\frac{s_0^2}{n-1}}} < b$$

D'où l'intervalle de confiance :

$$\overline{x_0} - b\sqrt{\frac{s_0^2}{n-1}} < \mu < \overline{x_0} + b\sqrt{\frac{s_0^2}{n-1}}$$

Exemple 9.13 *Sept dosages du glucose sanguin, effectués indépendamment, à partir du sang d'une même provenance ont donné les résultats suivants : 1.17 ; 1.16 ; 1.16 ; 1.19 ;1.19 ; 1.21 ; 1.18 grammes par litre.*
On admet que chaque moyenne est une variable aléatoire normale dont l'espérance μ est la valeur que l'on essaie de mesurer.
Déterminer, pour le coefficient de sécurité de 95%, l'intervalle de confiance de μ.
Soient $\overline{X}$ la moyenne et S^2 la variance des 7 mesures. On sait que la variable $T_6 = \frac{\overline{X}-\mu}{\sqrt{\frac{S^2}{6}}}$ est une variable de Student à 6 degrés de liberté, compte tenu de la normalité et de l'indépendance des mesures.
On a observé les valeurs $\overline{x_0} = 1.18$ et $s_0^2 = \frac{20}{7}10^{-4}$ de $\overline{X}$ et S^2.

Au niveau de confiance de 95%, la valeur $t_0 = \frac{\overline{x_0}-\mu}{\sqrt{\frac{s_0^2}{6}}}$ *vérifie l'inégalité :* $-2.447 < t_0 < 2.447$

$$\Rightarrow -2.447 < \frac{1.18-\mu}{\sqrt{\frac{20}{7\times 6}10^{-4}}} < 2.447$$

D'où l'intervalle de confiance : $1.163 < \mu < 1.197$ *gramme par litre.*

6.4 *Problèmes à deux échantillons*

Soit un ensemble $(X_1, X_2, \dots, X_n)$ de n variables indépendantes, de même loi normale $\mathcal{N}(\mu_X, \sigma^2)$. Soit un deuxième ensemble $(Y_1, Y_2, \dots, Y_p)$ de p variables indépendantes, de même loi normale $\mathcal{N}(\mu_Y, \sigma^2)$. Les variables X et Y sont supposées avoir la même variance ($\sigma_X^2 = \sigma_Y^2 = \sigma^2$). Les deux ensembles sont supposés indépendants entre eux.
On a : $\overline{X} \sim \mathcal{N}\left(\mu_X, \frac{\sigma^2}{n}\right)$ et $\overline{Y} \sim \mathcal{N}\left(\mu_Y, \frac{\sigma^2}{p}\right)$. On en déduit : $\overline{X}-\overline{Y} \sim \mathcal{N}\left(\mu_X - \mu_Y, \sigma^2\left(\frac{1}{n}+\frac{1}{p}\right)\right)$ et, par conséquent :

$$U = \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{p}}} \sim \mathcal{N}(0,1)$$

Soient maintenant les variables S_X^2 et S_Y^2, variances associées aux X et aux Y. On sait que :

$$\frac{nS_X^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{et} \quad \frac{pS_Y^2}{\sigma^2} \sim \chi^2_{p-1}$$

On en déduit :

$$V = \frac{nS_X^2}{\sigma^2} + \frac{pS_Y^2}{\sigma^2} \sim \chi^2_{n+p-2}$$

(somme de deux chi-deux indépendants).
Il en résulte que, compte tenu du théorème de Cochran (voir section 3.1), la variable $Z = \frac{U}{\sqrt{V}}\sqrt{n+p-2}$ obéit à la définition d'une variable de Student à $n+p-2$ degrés de liberté.
Cette variable Z s'écrit :

$$Z = \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{p}}} \frac{\sqrt{n+p-2}}{\sqrt{(nS_X^2+pS_Y^2)\frac{1}{\sigma^2}}} = \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{nS_X^2+pS_Y^2}{n+p-2}}\sqrt{\frac{1}{n}+\frac{1}{p}}}$$

On désigne par

$$\widehat{S^2} = \frac{nS_X^2+pS_Y^2}{n+p-2}$$

l'estimateur sans biais de la variance commune aux X et aux Y (voir exercice 5 du chapitre 8) , ce qui permet d'écrire :

$$Z = \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\widehat{S^2}\frac{1}{n}+\frac{1}{p}}} \sim T_{n+p-2}$$

L'utilisation de ce modèle permet de procéder à la comparaison de deux espérances. Cependant, il est bon de souligner que ce modèle n'est valable que sous des conditions assez restrictives :

- normalité et indépendance des variables X ;
- normalité et indépendance des variables Y ;
- indépendance relative des X et des Y ;
- égalité des variances des X et des Y.

6.5 Comparaison de deux espérances

On désire tester l'hypothèse $H_0 \ : \mu_X = \mu_Y$ contre l'une des trois contre l'une des trois hypothèses alternatives suivantes :

$$H_1 \ : \mu_X > \mu_Y \ \ ; \ \ H_2 \ : \mu_X < \mu_Y \ \ ; \ \ H_3 \ : \mu_X \neq \mu_Y$$

ce que l'on peut écrire : $H_0 \ : \mu_X - \mu_Y = 0$

$$H_1 \ : \mu_X - \mu_Y > 0 \ \ ; \ \ H_2 \ : \mu_X - \mu_Y < 0 \ \ ; \ \ H_3 \ : \mu_X - \mu_Y \neq 0$$

Sous H_0, il résulte de ce qui précède que la variable $T = \frac{\overline{X}-\overline{Y}}{\sqrt{\widehat{S^2}\left(\frac{1}{n}+\frac{1}{p}\right)}}$ est une variable de Student T_{n+p-2}.

Les tests se construisent alors comme dans les problèmes à un échantillon :

contre H_1 : test unilatéral avec domaine de rejet à droite, car, sous H_1, l'écart $\overline{X} - \overline{Y}$ aura tendance à avoir des valeurs plus élevées que sous H_0 ;

contre H_2 : test unilatéral avec domaine de rejet à gauche ;

contre H_3 : test bilatéral.

Pour effectuer le test on calcule, à partir des échantillons : $\overline{x_0}$, $\overline{y_0}$, s_X^2, s_Y^2, $s_0^2 = \frac{ns_X^2+ps_Y^2}{n+p-2}$, et enfin $t_0 = \frac{\overline{x_0}-\overline{y_0}}{\sqrt{\widehat{s_0^2}\left(\frac{1}{n}+\frac{1}{p}\right)}}$.

Il suffit de regarder si t_0 est dans le domaine d'acceptation ou de rejet de H_0.

Exemple 9.14 *(suite de l'exemple 9.9). Peut-on conclure à une différence entre les espérances μ_X et μ_Y ?*

On a montré précédemment que l'hypothèse $\sigma_X^2 = \sigma_Y^2$ est acceptable. Dans ces conditions, la variance commune aux deux séries de mesures peut être estimée par :

$$s_0^2 = \frac{11 \times 0.3130 + 9 \times 0.4231}{18} \approx 0.4028$$

On veut tester l'hypothèse $H_0 \ : \mu_X = \mu_Y$ contre $H_1 \ : \mu_X \neq \mu_Y$.

Soient $\overline{X}$ la moyenne de 11 mesures de type A et $\overline{Y}$ la moyenne de 9 mesures de type B, $\widehat{S^2}$ étant l'estimateur de la variance commune aux deux types de mesure.

Sous H_0, la variable $T_{18} = \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\widehat{S^2}\left(\frac{1}{11}+\frac{1}{9}\right)}}$ obéit à une loi de Student à 18 degrés de liberté.

On effectue ici un test bilatéral. On a observé la valeur

$$t_0 = \frac{3.92 - 4.18}{\sqrt{0.4028\left(\frac{1}{11} + \frac{1}{9}\right)}} = -0.911 \text{ de la variable } T_{18}$$

Cette valeur est comprise dans l'intervalle d'acceptation de H_0 au niveau 5% : $[-2.101; 2.101]$. *On peut donc admettre que les deux types de végétaux ont la même teneur en calcium, en moyenne.*

Ce test de comparaison de deux espérances à partir de deux moyennes empiriques est dit "test d'homogénéité de Student". La condition restrictive d'égalité des variances nécessite, la plupart du temps, une vérification préliminaire à l'aide du test F.

7 Exercices

1. Une nouvelle technique de dosage du glucose sanguin vient d'être mise au point. Sept dosages, effectués à l'aide de cette nouvelle technique, à partir du sang d'une même provenance, ont donné les résultats suivants :

   $$1.17 - 1.16 - 1.16 - 1.19 - 1.19 - 1.21 - 1.18 \text{ gramme/litre}$$

   On admet que chaque mesure est une variable aléatoire normale dont l'espérance μ est la valeur que l'on cherche à déterminer et dont l'écart-type caractérise la précision du procédé. On considère que les n résultats possibles d'un ensemble de n dosages, effectués dans les mêmes conditions sont des variables aléatoires indépendantes et de même loi.

   - La technique utilisée jusqu'à ce jour était caractérisée par un écart-type de 0.05 mg/litre. Peut-on dire que la nouvelle technique est plus précise que l'ancienne ?
   - Déterminer, à partir de l'échantillon, un intervalle de confiance de μ.

2. Une boisson, de consommation courante, est vendue en bouteilles d'un litre. L'un des éléments de sa composition est un certain produit A qui fait l'objet d'une réglementation particulière.

   On désigne par X la variable aléatoire représentant la quantité A, exprimée en milligrammes, que l'on peut trouver dans une bouteille. On admet que X suit une loi normale d'espérance μ. Pour que la boisson soit conforme à la réglementation en vigueur, la valeur de μ ne doit pas dépasser 50 milligrammes.

   Une association de consommateurs a fait effectuer des dosages de A sur 9 bouteilles choisies au hasard mais d'une marque donnée. Les résultats, en milligrammes, sont les suivants :

   60.5 - 58.5 - 57.8 - 56.4 - 54.7 - 53.5 - 49.3 - 48.2 - 48.0

   - Vérifier si la réglementation est respectée par le fabricant.
   - Donner, pour la marque en question, un intervalle de confiance de μ.

3. Les résultats du dosage du glucose sanguin, effectué sur un lot de 9 lapins, sont les suivants :

   1.17 ; 1.16 ; 1.15 ; 1.18 ; 1.19 ; 1.20 ; 1.20 ; 1.16 ; 1.21

   grammes par litre. On admet que chaque mesure est une variable normale d'espérance μ, taux moyen de glycémie dans la population d'où provient le lot. Entre quelles limites peut-on situer μ, au niveau de confiance de 95% ?

4. Pour comparer l'influence de deux régimes alimentaires A et B sur le développement des doryphores, un entomologiste a mesuré, lors de la mue imaginale, le poids d'insectes élevés dans les conditions A pour les uns, dans les conditions B pour les autres. il a obtenu les résultats suivants :

   Régime A (9 insectes mâles) : 100, 94, 119, 111, 113, 84, 102, 107, 99 mg

   Régime B (8 insectes mâles) : 107, 115, 99, 111, 114, 127, 145, 140 mg

Le poids d'un insecte choisi au hasard dans un élevage est une variable aléatoire que l'on désignera par X dans le cas A et par Y dans le cas B. On admet que X et Y sont de loi normale.

- Montrer qu'il n'y a aucune raison de penser que X et Y ont des variances différentes. Pour la suite, on notera σ^2, la valeur commune de ces deux variances.
- En utilisant l'ensemble des résultats, donner une estimation de σ^2 ; on désigne par $\hat{S}^2$ l'estimateur utilisé. Calculer l'espérance et la variance de $\hat{S}^2$; en déduire les qualités de cet estimateur.
- En utilisant l'ensemble des résultats et avec un minimum de justifications, donner un intervalle de confiance de σ^2
- Montrer que le régime B est plus favorable au développement des insectes que le régime A.

5. Un laboratoire produit un vaccin destiné à une population de 100000 personnes. On considère que la proportion p de personnes achetant le vaccin suit une loi gaussienne de moyenne 0.24 et d'écart-type 0.03.
   - Combien de doses doit-il produire pour que le risque de rupture de stock soit de l'ordre de 2% ?
   - Quelle doit être la marge bénéficiaire réalisée sur chaque doses vendue pour que le risque de perte soit de l'ordre de 3% ?
6. On désire savoir si une certaine variété de blé a, dans une région A, un rendement supérieur à celui qu'elle a dans une région B. Pour cela, on dispose de résultats, exprimés en quintaux par hectare, obtenus sur seize parcelles différentes et ainsi répartis :

Région A	48.0	48.2	50.3	53.5	54.6	56.4	57.8	58.5	60.5
Région B	44.2	46.3	48.3	48.5	50.5	51.2	55.4		

   On suppose que le rendement observable sur une parcelle, choisie au hasard dans une région donnée, est une variable aléatoire désignée par X pour la région A et par Y pour la région B.

   (a) Dans un premier temps, on admet que X et Y sont des variables normales dont l'espérance et la variance sont représentées par : μ_A et σ_A^2 pour X, μ_B et σ_B^2 pour Y

      i. Montrer que l'hypothèse "$\sigma_A^2 = \sigma_B^2$" est acceptable. Pour la suite, on désignera par σ^2 la valeur commune de ces deux variances.

      ii. On peut alors disposer de trois estimateurs sans biais de σ^2. Le premier est associé à un ensemble de 9 rendements possibles provenant de la région A. Le deuxième est associé à un ensemble de 7 rendements possibles provenant de la région B. Le troisième associé à l'ensemble des 16 rendements précédents. A partir de leur écart quadratique moyen relativement à σ^2, choisir le meilleur de ces trois estimateurs. En déduite une estimation de σ^2.

      iii. A partir de l'ensemble des 16 observations, donner un intervalle de confiance de σ^2.

      iv. Montrer que l'on peut accepter l'hypothèse "$\mu_A > \mu_B$"

v. Donner un estimateur de $\mu_A - \mu_B$. Calculer son espérance et sa variance ; qu'en conclure ? En déduire une estimation de $\mu_A - \mu_B$

vi. Déterminer un intervalle de confiance de $\mu_A - \mu_B$

(b) En l'absence de toute hypothèse sur le type de loi suivi par X et par Y, peut-on cependant affirmer que le rendement de la variété de blé a tendance à être plus élevé dans la région A que dans la région B ?

7. (a) On considère un ensemble $(X_1, X_2, \ldots, X_n)$, de n variables indépendantes et de même loi $\mathcal{N}(\mu, \sigma^2)$. Soient les variables :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad , \quad S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad , \quad \widehat{S_X^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

À partir des résultats connus sur les lois du χ^2, calculer l'espérance et la variance des variables S_X^2 et $\widehat{S_X^2}$. Quelles sont les qualités des variables S_X^2 et $\widehat{S_X^2}$ considérées comme des estimateurs du paramètre σ^2 ? Quel est le meilleur estimateur de σ^2 ?

(b) On considère un deuxième ensemble $\{Y_1, Y_2, \ldots, Y_p\}$ indépendant du précédent, de variables aléatoires indépendantes et de même loi $\mathcal{N}(\mu_2, \sigma^2)$. Soient les variables :

$$\overline{Y} = \frac{1}{p}\sum_{i=1}^{p} Y_i \quad , \quad S_Y^2 = \frac{1}{p}\sum_{i=1}^{p}(Y_i - \overline{Y})^2 \quad , \quad \widehat{S^2} = \frac{(n-1)\widehat{S_X^2} + (p-1)\widehat{S_Y^2}}{n+p-2}$$

Calculer l'espérance et la variance de $\widehat{S^2}$. Quelles sont les qualités de la variable $\widehat{S^2}$ considérées comme des estimateurs du paramètre σ^2 ?

(c) À votre avis, quel est le meilleur estimateur de σ^2 ?

8. On considère deux variables aléatoires réelles R et S de densités de probabilité respectives :

$$f_R : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto ae^{-at} & , \text{ si } t > 0 \end{cases}$$

et

$$f_S : \begin{cases} t \mapsto 0 & , \text{ si } t \leq 0 \\ t \mapsto be^{-bt} & , \text{ si } t > 0 \end{cases}$$

où a et b sont deux paramètres réels strictement positifs.

On se propose, en se ramenant à des modèles connus, de construire des tests concernant les paramètres de ces distributions exponentielles, et d'imaginer comment procéder à leur estimation.

On rappelle que les densités de probabilité d'une loi du khi-deux à deux degrés de liberté et d'une loi exponentielle de paramètre $\frac{1}{2}$ sont les mêmes.

Construction des modèles • Montrer que la variable aléatoire $T = 2aR$ suit une loi du khi-deux à deux degrés de liberté. en déduire l'espérance et la variance de R.

- Soit un ensemble de n variables aléatoires $(R_1, R_2, \dots, R_n)$ indépendantes et de même loi que R. On associe à ce n-échantillon la variable $Z_n = 2a(R_1 + R_2 + \cdots + R_n)$. Quelle est la loi de Z_n ?
- Soit un deuxième ensemble $(S_1, S_2, \dots, S_p)$ de variables indépendantes entre elles, indépendantes des n précédentes, et de même loi que S. On considère la variable $U(n,p) = \frac{p}{n}\frac{a}{b}\frac{(R_1+R_2+\cdots+R_n)}{(S_1+S_2+\cdots+S_p)}$. Quelle est la loi de $U(n,p)$?

Premier test On suppose $n = 12$. Les valeurs observées des 12 variables considérées sont précisément :

0.100 - 1.197 - 0.152 - 0.182 - 0.418 - 0.192 - 0.029 - 0.885 - 0.161 - 0.633 - 0.278 - 0.008

Construire un test de l'hypothèse $H_0 : a = 3$ contre l'hypothèse $H_1 : a > 3$. Quel est le résultat du test ?

Deuxième test On suppose $p = 10$. Les valeurs observées des 10 variables considérées sont précisément :

0.361 - 0.085 - 0.293 - 0.080 - 0.077 - 0.020 - 0.036 - 0.095 - 0.197 - 0.206

Construire un test de l'hypothèse $H_0 : a = b$ contre l'hypothèse $H_1 : a \neq b$. Quel est le résultat du test ?

Estimation En cas de rejet de l'hypothèse $a = b$, déterminer un intervalle de confiance de b et donner, avec un minimum de justifications, une estimation de b.

8 Solutions

1. On considère un 7-échantillon $(Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7)$ où les variables Y_i sont indépendantes et de loi $\mathcal{N}(\mu, \sigma^2)$.

   On pose : $\begin{cases} \overline{Y} = \frac{1}{7}\sum_{i=1}^{7} Y_i \\ \hat{S}^2 = \frac{1}{6}\sum_{i=1}^{7}\left(Y_i - \overline{Y}\right)^2 \end{cases}$

   On a observé : $\begin{cases} \overline{y} = 1.18 \\ \hat{s}^2 = 3.333.10^{-4} = \frac{20}{6}.10^{-4} \end{cases}$

   On veut tester l'hypothèse $H_0 : \sigma^2 = 25.10^{-4}$ contre l'hypothèse $H_1 : \sigma^2 < 25.10^{-4}$

   Soit la variable $Z = \frac{6\hat{S}^2}{25.10^{-4}}$.

   Sous H_0, Z suit une loi du chi-deux à 6 degrés de liberté.

   Sous H_1, Z aura tendance à prendre des valeurs plus faibles que sous H_0 ; on construit donc un test unilatéral avec un domaine de rejet à gauche.

   On a observé la valeur $z = \frac{20.10^{-4}}{25.10^{-4}} = 0.8 < 1.635$

   Au niveau 5%, on rejette donc H_0. La nouvelle méthode est plus précise que l'ancienne.

   Soit la variable $T = \frac{\overline{Y}-\mu}{\sqrt{\hat{S}^2/7}}$, dont on a observé la valeur $t = \frac{1.18-\mu}{\sqrt{\frac{20}{42}10^{-4}}}$. Cette variable suit une loi de Student à 6 degrés de liberté. Au niveau de confiance de 95%, on peut parier sur l'inégalité $-2.447 < \frac{1.18-\mu}{\sqrt{\frac{20}{42}10^{-4}}} < 2.447$

   D'où l'intervalle de confiance $1.163 < \mu < 1.197$

2. On considère un 9-échantillon $(X_1, X_2, \ldots, X_9)$ où X_i représente la quantité A que l'on peut trouver dans la i$^{\text{ème}}$ bouteille. Les 9 variables aléatoires sont indépendantes et de même loi $\mathcal{N}(\mu, \sigma^2)$. On pose $\overline{X} = \frac{1}{9}\sum_{i=1}^{9} X_i$; $S^2 = \frac{1}{9}\sum_{i=1}^{9}\left(X_i - \overline{X}\right)^2$ ou $\hat{S}^2 = \frac{1}{8}\sum_{i=1}^{9}\left(X_i - \overline{X}\right)^2$

   - On va tester $H_0 : \mu \leq 50$ contre $H_1 : \mu > 50$
     Pour cela, on choisit la variable

     $$T = \frac{\overline{X} - 50}{\sqrt{S^2/8}} = \frac{\overline{X} - 50}{\sqrt{\hat{S}^2/9}}$$

     Sous H_0, si 50=μ, T suit une loi de Student à 8 degrés de liberté. Si $\mu < 50$, T aura tendance à prendre des valeurs négatives.

     Sous H_1, $\overline{X}$ a tendance à prendre des valeurs supérieures à 50 et donc T a tendance à prendre des valeurs positives.

     On construit donc un intervalle unilatéral avec domaine de rejet de H_0 à droite.

     Au niveau $\alpha = 5\%$, le domaine de rejet de H_0 est $[1.860\,; +\infty[$. De manière équivalente, le domaine d'acceptation de H_0 est $]-\infty\,; 1.860[$

     On a observé $\overline{x} = 54.1$, $8\hat{s^2} = 9s^2 = 175.08$ Donc $t = 2.629$ En ce qui concerne μ la réglementation n'est pas respectée.

- On choisit la variable
$$U = \frac{\overline{X} - \mu}{\sqrt{S^2/8}} = \frac{\overline{X} - \mu}{\sqrt{\hat{S}^2/9}}$$
Cette variable suit une loi de Student à 8 degrés de liberté. Pour un coefficient de sécurité de 95%, on peut parier que la valeur observée u_0 vérifie les inégalités :
$$-2.306 < u_0 < 2.306$$
$$\Rightarrow -2.306 < \frac{54.1 - \mu}{\sqrt{175.08/72}} < 2.306$$
D'où l'intervalle de confiance :
$$50.5 < \mu < 57.7$$

3. On considère un 9-échantillon $(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9)$ où X_i représente le résultat possible du $i^{\text{ème}}$ dosage. Les variables X_i sont indépendantes et de même loi $\mathcal{N}(\mu, \sigma^2)$.
On pose : $\overline{X} = \frac{1}{9}\sum_{i=1}^{9} X_i$. On a observé $\overline{x_0} = 1.18$ mg/litre.
On pose $S^2 == \frac{1}{9}\sum_{i=1}^{9}(X_i - \overline{X})^2$. On a observé $s_0^2 = 4\,10^{-4}$ (mg/litre.)^2.
   - Soit $Z = \frac{\overline{X}-\mu}{\sqrt{s^2/8}}$. Cette variable suit une loi de Student à 8 degrés de liberté.
   Au niveau 90%, on peut écrire que la valeur observée z_0 vérifie l'inégalité :
   $$-1.86 < \frac{1.18 - \mu}{\sqrt{\frac{4}{8}10^{-4}}} < 1.86$$
   D'où l'intervalle de confiance : $1.167 < \mu < 1.193$ mg/l
   - Soit $Y = \frac{9S^2}{\sigma^2}$. Cette variable suit une loi du khi-deux à 8 degrés de liberté.
   Pour un niveau de confiance de 90%, on peut parier sur l'inégalité :
   $$2.733 < \frac{36\,10^{-4}}{\sigma^2} < 15.507$$
   D'où l'intervalle de confiance : $2.32\,10^{-4} < \sigma^2 < 13.17\,10^{-4}$.
4. Soit un premier ensemble $(X_1, X_2, \ldots, X_8, X_9)$ de variables indépendantes et de même loi, représentant les mesures possibles sur 9 insectes élevés dans les conditions A.
Soit un deuxième ensemble $(Y_1, Y_2, \ldots, Y_7, Y_8)$ de variables indépendantes et de même loi, représentant les mesures possibles sur 8 insectes élevés dans les conditions B.
On suppose $X_i \sim \mathcal{N}(\mu_A, \sigma_A^2)$ et $Y_i \sim \mathcal{N}(\mu_B, \sigma_B^2)$
On pose : $\begin{cases} \overline{X} = \frac{1}{9}\sum_{i=1}^{9} X_i \\ \hat{S_X^2} = \frac{1}{8}\sum_{i=1}^{9}\left(X_i - \overline{X}\right)^2 \end{cases}$

On a observé : $\begin{cases} \overline{x} \approx 103.222 \\ \hat{s_X^2} \approx 112.944 \\ s_X^2 \approx 100.395 \end{cases}$

On pose : $\begin{cases} \overline{Y} = \frac{1}{8}\sum_{i=1}^{8} Y_i \\ \hat{S^2} = \frac{1}{7}\sum_{i=1}^{8}\left(Y_i - \overline{Y}\right)^2 \end{cases}$

On a observé : $\begin{cases} \overline{y} \approx 119.750 \\ \hat{s_Y^2} \approx 260.786 \\ s_Y^2 \approx 228.188 \end{cases}$

- On veut tester $H_0 : \sigma_A^2 = \sigma_B^2$ contre $H_1 : \sigma_A^2 \neq \sigma_B^2$
  On choisit la variable $Z = \frac{\hat{s_Y^2}}{\hat{s_X^2}}$
  Sous H_0, $Z = \frac{\sigma_A^2}{\sigma_B^2}\frac{\hat{s_Y^2}}{\hat{s_X^2}}$, puisque $\sigma_A^2 = \sigma_B^2$. Donc $Z \sim \mathcal{F}(7, 8)$
  Sous H_1, $Z = \frac{\sigma_B^2}{\sigma_A^2}\left(\frac{\sigma_A^2}{\sigma_B^2}\frac{\hat{s_Y^2}}{\hat{s_X^2}}\right) = \frac{\sigma_B^2}{\sigma_A^2}Z'$ où $Z' \sim \mathcal{F}(7, 8)$. Donc Z aura tendance à prendre des valeurs plus fortes ou plus faibles que sous H_0 selon que $\sigma_A^2 < \sigma_B^2$ ou $\sigma_A^2 > \sigma_B^2$. D'où un test bilatéral. A un niveau 5%, le domaine d'acceptation de H_0 est : $]1/4.8994\,; 5.5286[$
  On a observé $z_0 \approx 2.309$ d'où l'acceptation de H_0
- On prend comme estimateur de la variable σ la variable aléatoire :

$$\hat{S^2} = \frac{8\hat{S_X^2} + 7\hat{S_Y^2}}{15}$$

  ce qui conduit à l'estimation $\hat{\sigma}^2 = \frac{8\times112.944+7\times260.786}{15} \approx 181.937$
  On peut écrire $\frac{15\hat{S^2}}{\sigma^2} = \frac{8\hat{S_X^2}}{\sigma^2} + \frac{7\hat{S_Y^2}}{\sigma^2}$
  Or $\frac{8\hat{S_X^2}}{\sigma^2} \sim \chi_8^2$ et $\frac{7\hat{S_Y^2}}{\sigma^2} \sim \chi_7^2$. Ces deux variables étant indépendantes, $\frac{15\hat{S^2}}{\sigma^2} \sim \chi_{15}^2$.
  On en déduit $\mathrm{E}\left(\frac{15\hat{S^2}}{\sigma^2}\right) = 15 \Rightarrow \frac{15}{\sigma^2}\mathrm{E}\left(\hat{S^2}\right) = 15 \Rightarrow \mathrm{E}\left(\hat{S^2}\right) = \sigma^2$
  $\hat{S^2}$ est donc un estimateur sans biais de σ^2
  On a de même $\mathrm{V}\left(\frac{15\hat{S^2}}{\sigma^2}\right) = 30 \Rightarrow \frac{15^2}{\sigma^4}\mathrm{V}\left(\hat{S^2}\right) = 30 \Rightarrow \mathrm{V}\left(\hat{S^2}\right) = \frac{2\sigma^4}{15}$
  Puisque le biais de $\hat{S^2}$ est nul, la variance est égale à l'écart quadratique moyen de $\hat{S^2}$ relativement à σ^2. Si on estime que cette grandeur est trop élevée, on peut la diminuer en jouant sur la taille des échantillons. $\hat{S^2}$ fait partie d'une suite consistante d'estimateurs.
- Puisque la variable $\frac{15\hat{S^2}}{\sigma^2}$ suit une loi du khi-deux à 15 degrés de liberté, on peut parier que la valeur "observée" de cette variable vérifie la double inégalité, pour un niveau de confiance de 95% :

$$6.262 < \frac{15 \times 181.937}{\sigma^2} < 27.488$$

  d'où l'intervalle de confiance : $99.28 < \sigma^2 < 435.81$

- On va tester $H_0 : \mu_B = \mu_A$ contre $H_1 : \mu_B > \mu_A$ Les variables X_i et Y_j étant considérées comme ayant la même variance σ^2, on choisit la variable :

$$V = \frac{\overline{Y} - \overline{X}}{\sqrt{\hat{S}^2(1/8 + 1/9)}}$$

Sous $H_0 : \mu_B - \mu_A = 0$; V suit une loi de Student à 15 degrés de liberté. Sous H_1, V aura tendance à prendre des valeurs positives préférentiellement. En effet $\mu_B > \mu_A \Rightarrow \overline{Y}$ a tendance à prendre des valeurs supérieures à $\overline{X}$. On construit donc un test unilatéral avec domaine de rejet H_0 à droite.
Au niveau 5%, le domaine d'acceptation de H_0 est $]-\infty\,; 1.753[$. Le domaine de rejet de H_0 est $[1.753\,; +\infty[$
On a observé $v_0 \approx 2.521 \Rightarrow$ on rejette H_0 le régime B favorise le développement des doryphores.

5. (a) D'après la table P$\left(\frac{p-0.24}{0.03} \geq 2.05\right) \approx 0.02$
Donc P($p > 0.3015) \approx 0.02$
Il doit donc produire 30150 doses

   (b) Il y a perte si $p \times 10^5 \times$prix$< 30150\times$coût
C'est-à-dire si $p < 0.3015\times$coût/prix
Or P$\left(\frac{p-0.24}{0.03} < -1.88\right) \approx 0.03$
La marge bénéficiaire doit donc être $\frac{0.3015}{0.1836} - 1 \approx 64\%$

6. (a) Au 9-échantillon $(X_1, X_2, \ldots, X_9)$ des rendements possibles des parcelles de A, on associe les variables aléatoires $\overline{X} = \frac{1}{9}\sum_{i=1}^{9} X_i$; $S^2 = \frac{1}{9}\sum_{i=1}^{9}\left(X_i - \overline{X}\right)^2$ ou $\hat{S}^2 = \frac{1}{8}\sum_{i=1}^{9}\left(X_i - \overline{X}\right)^2$.
Au 7-échantillon $(Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7)$ des rendements possibles des parcelles de B, on associe les variables aléatoires $\begin{cases} \overline{Y} = \frac{1}{7}\sum_{i=1}^{7} Y_i \\ \hat{S}^2 = \frac{1}{6}\sum_{i=1}^{7}\left(Y_i - \overline{Y}\right)^2 \end{cases}$
On suppose $X_i \sim \mathcal{N}\left(\mu_A, \sigma_A^2\right)$ et $Y_i \sim \mathcal{N}\left(\mu_B, \sigma_B^2\right)$

   i. On veut tester $H_0 : \sigma_A^2 = \sigma_B^2$ contre $H_1 : \sigma_A^2 \neq \sigma_B^2$
On choisit la variable $U = \frac{\hat{s_X^2}}{\hat{s_Y^2}}$
Sous H_0, $U = \frac{\sigma_B^2}{\sigma_A^2}\frac{\hat{s_X^2}}{\hat{s_Y^2}}$, puisque $\sigma_A^2 = \sigma_B^2$. Donc $Z \sim \mathcal{F}(8, 6)$
Sous H_1, $U = \frac{\sigma_A^2}{\sigma_B^2}\left(\frac{\sigma_B^2}{\sigma_A^2}\frac{\hat{s_X^2}}{\hat{s_Y^2}}\right) = \frac{\sigma_A^2}{\sigma_B^2}Z'$ où $Z' \sim \mathcal{F}(8, 6)$. Donc U aura tendance à prendre des valeurs plus fortes ou plus faibles que sous H_0 selon que $\sigma_A^2 > \sigma_B^2$ ou $\sigma_A^2 < \sigma_B^2$. D'où un test bilatéral. A un niveau 5%, le domaine d'acceptation de H_0 est : $]1/4.6517\,; 5.5596[$
On a observé $\hat{s_X^2} = 20.785$, $\hat{s_Y^2} = 13.140 \Rightarrow u \approx 1.582$ d'où l'acceptation de H_0

ii. Les trois estimateurs possibles de σ^2 sont $\hat{s_X^2}$, $\hat{s_Y^2}$ et $\hat{S^2} = \frac{8\hat{S_X^2}+6\hat{S_Y^2}}{14}$
On sait $\frac{8\hat{S_X^2}}{\sigma^2} \sim \chi^2_8$ et $\frac{6\hat{S_Y^2}}{\sigma^2} \sim \chi^2_6$.
Donc E($\frac{8\hat{S_X^2}}{\sigma^2}$) =E($\frac{6\hat{S_Y^2}}{\sigma^2}$)=E($\frac{8\hat{S_X^2}+6\hat{S_Y^2}}{14}$) $= \sigma$
et V($\frac{8\hat{S_X^2}}{\sigma^2}$) $= 16$ et donc V($\hat{S_X^2}$) $= \frac{\sigma^4}{4}$; V($\frac{6\hat{S_Y^2}}{\sigma^2}$) $= 12$ et donc V($\hat{S_Y^2}$) $=$ $\frac{\sigma^4}{3}$; V($\frac{14\hat{S^2}}{\sigma^2}$) $= 28$ et donc V($\hat{S^2}$) $= \frac{\sigma^4}{7}$;
les trois estimateurs sont des estimateurs sans biais de σ^2. Dans ces conditions leur écart quadratique moyen est égale à leur variance (EQM=V()+(biais)^2).
Le meilleur estimateur de σ^2 est celui de plus faible écart quadratique moyen et c'est donc $\hat{S^2}$. On en déduit $\hat{\sigma^2} = \frac{8\hat{S_X^2}+6\hat{S_Y^2}}{14} \approx 17.5086$

iii. Pour déterminer un intervalle de confiance de σ^2, on utilise la variable aléatoire $14\hat{S^2}/\sigma^2 \sim \chi^2_{14}$. Pour un niveau de confiance de 95%, on peut parier sur l'inégalité : $5.629 < \frac{245.12}{\sigma^2} < 43.546$ D'où l'intervalle de confiance $9.384 < \sigma^2 < 43.546$

iv. On va tester, au niveau 5%, l'hypothèse $H_0 : \mu_B = \mu_A$ contre $H_1 : \mu_B > \mu_A$ Les variables X_i et Y_j étant considérées comme ayant la même variance σ^2, on choisit la variable :

$$V = \frac{\overline{Y} - \overline{X}}{\sqrt{\hat{S^2}(1/7 + 1/9)}}$$

Sous $H_0 : \mu_B - \mu_A = 0$; V suit une loi de Student à 14 degrés de liberté. Sous H_1, V aura tendance à prendre des valeurs positives préférentiellement. En effet $\mu_B > \mu_A \Rightarrow \overline{Y}$ a tendance à prendre des valeurs supérieures à $\overline{X}$. On construit donc un test unilatéral avec domaine de rejet H_0 à droite.
Au niveau 5%, le domaine d'acceptation de H_0 est $]-\infty\,; 1.761[$. Le domaine de rejet de H_0 est $[1.761\,; +\infty[$
On a observé $v_0 = \frac{54.2-49.2}{\sqrt{17.5086(1/9+1/7)}} \approx 2.3711 \Rightarrow$ on accepte $H_1 : \mu_A > \mu_B$

v. Un estimateur de la variable $\mu_A - \mu_B$ est la variable $Z = \overline{X} - \overline{Y}$
E(Z)=E($\overline{X} - \overline{Y}$)=E($\overline{X}$)$-$E($\overline{Y}$)= $\mu_A - \mu_B$. C'est donc un estimateur sans biais
V(Z)=V($\overline{X} - \overline{Y}$)=V($\overline{X}$)+V($\overline{Y}$) (car indépendance de X et Y) donc V(Z)=$\frac{\sigma}{9} + \frac{\sigma}{7}$. C'est donc un estimateur consistant.
On en déduit : $\mu_A \hat{-} \mu_B = \overline{X} - \overline{Y} = 54.2 - 49.2 = 5$

vi. Pour déterminer un intervalle de confiance de $\mu_A - \mu_B$, on choisit la variable aléatoire

$$T = \frac{\overline{Y} - \overline{X} - (\mu_A - \mu_B)}{\sqrt{\hat{S^2}(1/9 + 1/7)}}$$

Cette variable suit une loi de Student à 14 degrés de liberté.
On a observé $t = \frac{5-(\mu_A-\mu_B)}{2.1087}$. Pour un niveau de confiance de 95%, on peut parier sur l'inégalité $-2.145 < \frac{5-(\mu_A-\mu_B)}{2.1087} < 2.145$. D'où l'intervalle de confiance $0.4768 < \mu_A - \mu_B < 9.5232$

(b) On peut tester l'hypothèse H_0 : "les variables X_i et Y_j ont même loi" contre l'hypothèse H_1 : "les variables X_i ont tendance à prendre des valeurs plus grandes que celles des Y_j".
Soit W la variable aléatoire associant aux 16 résultats possibles, rangés par valeurs croissantes, la somme des rangs des 7 résultats de B. W est à valeurs entières sur $\{28, 29, \ldots, 90, 91\}$. Sous H_0, W a une distribution en cloche, symétrique. Sous H_1, W aura tendance à prendre des valeurs situées à gauche de la distribution. D'où un test unilatéral qui admet, au niveau 5%, pour domaine de rejet $H_0 : \{28, 29, \ldots, 42, 43\}$
les 16 observations sont rangés par valeurs croissantes :
$y_1, y_2, x_1, x_2, y_3, y_4, x_3, y_5, y_6, x_4, x_5, y_7, x_6, x_7, x_8, x_9$
On en déduit $W = 43 \Rightarrow$ Rejet de H_0
Le test non paramétrique de la somme des rangs conduit donc également à la conclusion que la variété a un meilleur rendement dans la région A.

7. (a) On sait que $\frac{nS_X^2}{\sigma^2} = \frac{(n-1)\widehat{S_X^2}}{\sigma^2} \sim \chi^2_{n-1}$. On en déduit donc (pour $n \geq 2$) :

$$\mathbb{E}\left(\frac{nS_X^2}{\sigma^2}\right) = n-1 \Rightarrow \mathbb{E}(S_X^2) = \frac{n-1}{n}\sigma^2$$

S_X^2 est un estimateur biaisé de σ^2, de biais $\left(-\frac{\sigma^2}{n}\right)$.

$$\mathbb{E}\left(\frac{(n-1)\widehat{S_X^2}}{\sigma^2}\right) = n-1 \Rightarrow \mathbb{E}(\widehat{S_X^2}) = \sigma^2$$

S_X^2 est un estimateur sans biais de σ^2

$$\mathbb{V}\left(\frac{nS_X^2}{\sigma^2}\right) = 2(n-1) \quad \Rightarrow \quad \mathbb{V}(S_X^2) = \frac{2(n-1)}{n^2}\sigma^4$$
$$\mathbb{V}\left(\frac{(n-1)\widehat{S_X^2}}{\sigma^2}\right) = 2(n-1) \quad \Rightarrow \quad \mathbb{V}(\widehat{S_X^2}) = \frac{2\sigma^4}{n-1}$$

et donc :

$$\mathbb{E}\left((S_X^2-\sigma^2)^2\right) = \left(\frac{2(n-1)}{n^2}+\frac{1}{n^2}\right)\sigma^4 \ , \ \mathbb{E}\left((\widehat{S_X^2}-\sigma^2)^2\right) = \mathbb{V}(\widehat{S_X^2}) = \frac{2\sigma^4}{(n-1)}$$

En utilisant la relation 8.5 on déduit que $\widehat{S_X^2}$, estimateur sans biais et dont la variance tend vers zéro lorsque n tend vers l'infini est un estimateur consistant de σ^2. En effet, la relation 8.5 s'écrit :

$$\forall \epsilon > 0 \ , \ 1 - \left(\frac{2}{n-1}\right)\frac{\sigma^4}{\epsilon^2} \leq \mathbb{P}(|\widehat{S_X^2} - \sigma^2| < \epsilon) \leq 1$$

Il est clair que $\forall\epsilon > 0$, $\mathbb{P}(|\widehat{S^2_X} - \sigma^2| < \epsilon) \to 1$ quand $n \to +\infty$.
En ce qui concerne S^2_X, on a :

$$\forall\epsilon > 0 \ , \ 1 - \left(\frac{2(n-1)}{n^2}\right)\frac{\sigma^4}{\epsilon^2} \leq \mathbb{P}(|S^2_X - \sigma^2| < \epsilon) \leq 1$$

Il est également clair que S^2_X est un estimateur consistant de σ^2, bien que biaisé : $\forall\epsilon > 0$, $\mathbb{P}(|S^2_X - \sigma^2| < \epsilon) \to 1$ quand $n \to +\infty$.
Le meilleur estimateur est celui qui a le plus faible écart quadratique moyen relativement à σ^2. Contrairement à ce que l'on pourrait penser, dans les modèles gaussiens, il s'agit de S^2_X. On a, en effet :

$$\begin{aligned}
\mathbb{E}\left((S^2_X - \sigma^2)^2\right) < \mathbb{E}\left((\widehat{S^2_X} - \sigma^2)^2\right) \Leftrightarrow \frac{2(n-1)}{n^2}\sigma^4 < \frac{2}{n-1}\sigma^4 & \\
\Leftrightarrow \ & 2n^2 - 3n + 1 < 2n^2 \\
\Leftrightarrow \ & 1 - 3n < 0
\end{aligned}$$

(b) Pour $n \geq 2$ et $p \geq 2$:

$$\begin{aligned}
\mathbb{E}(\widehat{S^2}) &= \frac{(n_1)\mathbb{E}(S^2_X) + (p-1)\mathbb{E}(S^2_Y)}{n+p-2} \\
&= \frac{(n_1)\sigma^2 + (p-1)\sigma^2}{n+p-2} \\
&= \sigma^2
\end{aligned}$$

À cause de l'indépendance des variables $\widehat{S^2_X}$ et $\widehat{S^2_Y}$, on a :

$$\begin{aligned}
\mathbb{V}(\widehat{S^2}) &= \frac{(n_1)^2}{(n+p-2)^2}\mathbb{V}(S^2_X) + \frac{(p-1)^2}{(n+p-2)^2}\mathbb{V}(S^2_Y) \\
&= \frac{(n_1)^2}{(n+p-2)^2}\frac{2\sigma^4}{n-1} + \frac{(p-1)^2}{(n+p-2)^2}\frac{2\sigma^4}{p-1} \\
&= \frac{2\sigma^4}{n+p-2}
\end{aligned}$$

$\widehat{S^2}$, estimateur de la variance σ^2, commune aux deux échantillons, est un estimateur consistant de σ^2 puisque sans biais et de variance qui tend vers zéro lorsque $n + p \to +\infty$.

(c) $\widehat{S^2}$ est le meilleur estimateur de la variance σ^2quand n et $p \geq 3$.
La fonction f définie par $f(x) = \frac{2x-1}{x^2} = \frac{2}{x} - \frac{1}{x^2}$, $x \geq 2$ est strictement décroissante ($f'(x) = -\frac{2}{x^2} + \frac{2}{x^3} = \frac{2(1-x)}{x^3} < 0$).
Donc si $n \geq p \geq 2$, on a : $\frac{2n-1}{n^2} \leq \frac{2p-1}{n^2}$ et S^2_X est le meilleur estimateur de σ^2 de σ^2 par rapport à S^2_Y.
Comparons maintenant les écarts quadratiques moyens de S^2_X. et $\widehat{S^2}$.

$$\begin{aligned}
\mathbb{E}\left(\widehat{S^2}-\sigma^2)^2\right) \leq \mathbb{E}\left(S_X^2-\sigma^2)^2\right) \quad & \Leftrightarrow \quad \frac{2\sigma^4}{n+p-2} \leq \frac{2n-1}{n^2}\sigma^4 \\
& \Leftrightarrow \quad \frac{2}{p+(n-2)} \leq \frac{2n-1}{n^2} \\
& \Leftrightarrow \quad 2n^2 \leq p(2n-1)+(2n-1)(n-2) \\
& \Leftrightarrow \quad 2n^2 \leq 2n^2-5n+2+p(2n-1) \\
& \Leftrightarrow \quad n(2p-5) \geq p-2 \qquad (9.2)
\end{aligned}$$

la relation 9.2 n'est pas vérifiée pour $p = 2$. Par contre, pour $p \geq 3$, la relation 9.2 est vérifiée pour $n \geq \frac{p-2}{2p-5}$. Donc pour $n \geq p \geq 3$. (Le cas $p \geq n$ se traite de la même façon en permutant X et Y, n et p). En effet : $n \geq p \Rightarrow n > p-2 \Rightarrow n > \frac{p-2}{2p-5}$, puisque $2p \geq 6 \Rightarrow 2p-5 \geq 1$.

8. **Construction des modèles**

- Soit F_T la fonction de répartition et f_T la densité de la variable T. Par définition :

$$F_T(t) = \mathrm{P}(T < t) = \mathrm{P}(2aR < t) = \mathrm{P}\left(R < \frac{t}{2a}\right) \qquad \text{car } a > 0$$

Donc :$F_T(t) = F_R\left(\frac{t}{2a}\right)$
Par dérivation :$f_T(t) = \frac{1}{2a} f_R\left(\frac{t}{2a}\right)$
D'où la densité :

$$f_T : \begin{cases} t \mapsto 0 & , \quad \text{si } t \leq 0 \\ t \mapsto \frac{1}{2}e^{-t/2} & , \quad \text{si } t > 0 \end{cases}$$

On reconnaît une loi exponentielle de paramètre 1/2 ou, ce qui revient au même, une loi du khi-deux à deux degrés de liberté. On sait alors que $\mathrm{E}(T)$=2 et $\mathrm{V}(T)$=4. On en déduit :

$$\begin{array}{l} \mathrm{E}(T) = \mathrm{E}(2aR) = 2a\mathrm{E}(R) \;\; \Rightarrow \mathrm{E}(R) = \frac{1}{a} \\ \mathrm{V}(T) = \mathrm{V}(2aR) = 4a^2\mathrm{V}(R) \Rightarrow \mathrm{V}(R) = \frac{1}{a^2} \end{array}$$

On retrouve les résultats connus pour une loi exponentielle de paramètre a

- On a

$$\begin{aligned}
Z_n &= 2a(R_1+R_2+\cdots+R_n) \\
&= 2aR_1+2aR_2+\cdots+2aR_n \\
&= T_1+T_2+\cdots+T_n
\end{aligned}$$

Z_n est la somme de n variables indépendantes et de même loi du khi-deux à deux degrés de liberté. Donc Z_n suit une loi du khi-deux à $2n$ degrés de liberté (propriété d'additivité des χ^2).

- De même $W_p = 2b(S_1 + S_2 + \cdots + S_p)$ suit une loi du khi-deux à $2p$ degrés de liberté et

$$U(n,p) = \frac{p}{n}\frac{a}{b}\frac{(R_1 + R_2 + \cdots + R_n)}{(S_1 + S_2 + \cdots + S_p)} = \frac{Z_n}{2n}\frac{2p}{W_p}$$

  Par définition $U(n,p)$ suit une loi de Fisher-Snédécor à $2n$ et $2p$ degrés de liberté, que l'on note $\mathcal{F}(2n, 2p)$.

Premier test On dispose d'un 12-échantillon $(R_1, R_2, \ldots, R_{12})$ On veut tester $H_0 : a = 3$ contre $H_1 : a > 3$. Pour cela, on choisit la variable $Z_{12} = 2.3(R_1 + R_2 + \cdots + R_{12})$

Sous H_0, puisque $a = 3$, Z_{12} suit une loi du khi-deux à 24 degrés de liberté.

Sous H_1, Z_{12} ne suit plus la même loi, mais on peut écrire :

$$Z_{12} = \frac{3}{a}2a(R_1 + R_2 + \cdots + R_{12}) = \frac{3}{a}Z^*_{12}$$

où Z^*_{12} suit une loi χ^2_{24}. Puisque, sous H_1, $\frac{3}{a} < 1$, on peut affirmer que la loi de Z_{12} est déplacée vers la gauche par rapport à la loi du χ^2_{24}.

On construit donc un test unilatéral avec domaine de rejet de H_0 à gauche.

Au niveau $\alpha = 5\%$, le domaine d'acceptation de H_0 est $]13.848\,; +\infty[$

On a observé la valeur $z_{12} = 6(0.100+1.197+0.152+\cdots+0.278+0.008) = 25.41$. Au niveau 5% l'hypothèse $H_0 : a = 3$ est acceptable.

Deuxième test On dispose d'un 12-échantillon $(R_1, R_2, \ldots, R_{12})$ et d'un 10-échantillon $(S_1, S_2, \ldots, S_{10})$ indépendant du précédent.

On veut tester $H_0 : a = b$ contre $H_1 : a \neq b$. Pour cela, on choisit la variable $U(12, 10) = \frac{10}{12}\frac{(R_1+R_2+\cdots+R_{12})}{(S_1+S_2+\cdots+S_{10})}$

Sous $H_0 : 1 = \frac{a}{b}$ et $U(12, 10)$ suit une loi $\mathcal{F}(24, 20)$.

Sous H_1, $U(12, 10)$ ne suit plus la même loi, mais on peut écrire :

$$Z_{12} = \frac{b}{a}\frac{10}{12}\frac{a}{b}\frac{(R_1 + R_2 + \cdots + R_{12})}{(S_1 + S_2 + \cdots + S_{10})} = \frac{b}{a}U(12, 10)^*$$

où $U^*(12, 10)$ suit une loi $\mathcal{F}(24, 20)$.

Deux cas sont alors à envisager :

$\frac{b}{a} > 1 \Leftrightarrow b > a$ la loi de $U(12, 10)$ est déplacée vers la droite par rapport à la loi $\mathcal{F}(24, 20)$.

$\frac{b}{a} < 1 \Leftrightarrow b < a$ la loi de $U(12, 10)$ est déplacée vers la gauche par rapport à la loi $\mathcal{F}(24, 20)$.

On construit donc un test bilatéral.

Au niveau $\alpha = 5\%$ (2.5% à gauche, 2.5% à droite), le domaine d'acceptation de H_0 est : $]0.4296\,; 2.4076[$

2.4076 est lu directement dans la table pour $\mathcal{F}(24, 20)$. 0.4296=1/2.3273 où 2.3273 est lu sur la table de $\mathcal{F}(24, 20)$.

On a observé la valeur $u(12, 10) = \frac{10}{12}\frac{4.235}{1.450} \approx 2.434$. Au niveau 5%, on rejette donc H_0.

Estimation Pour déterminer un intervalle de confiance de b, on passe par l'intermédiaire de la variable $W_{20} = 2b(S_1 + S_2 + \cdots + S_{10})$ qui suit une loi du khi-deux à 20 degrés de liberté.

Pour un coefficient de sécurité de 95%, on peut parier que la valeur observée w_{20} de W_{20} vérifie les inégalités : $9.591 < w_{20} < 34.170$

or $w_{20} = 2b(1.450) \Rightarrow 9.591 < 2.90b < 34.170$

D'où l'intervalle de confiance : $3.307 < b < 11.1783$

Estimation ponctuelle de b

On sait qu'un bon estimateur du paramètre E(S) est la variable $\overline{S} = \frac{1}{p}(S_1 + S_2 + \cdots + S_p)$; c'est un estimateur consistant puisque :

- il est sans biais : E(S)=E($\overline{S}$)
- sa variance V($\overline{S}$)=V(S)/p tend vers zéro si p tend vers $+\infty$

On en déduit une estimation de E(S)

$$\mathrm{E}(\hat{S}) = \frac{1}{10}(1.450) = 0.145$$

on sait d'autre part que E(S)=1/b. On peut donc en déduire l'estimation :

$$\hat{b} = \frac{1}{\mathrm{E}(\hat{S})} = 6.896 \approx 6.9$$

Chapitre 10

Le test chi-deux

1 Introduction

Le test du χ^2 est fréquemment utilisé par les biologistes. À la différence des modèles gaussiens, par exemple, ce test ne s'appuie pas sur un modèle probabiliste rigoureux, mais sur une loi asymptotique ; il est donc délicat à utiliser et il est parfois préférable de le remplacer, lorsque c'est possible, par un test non paramétrique plus adapté.
On expliquera d'abord les principes du test sur une loi multinomiale, puis, dans ses applications les plus courantes, la méthode non paramétrique qui en découle.

2 Test sur une loi multinomiale

2.1 *Distribution à deux classes*

Soit une expérience aléatoire $\mathcal{E}$ susceptible d'entraîner la réalisation d'un événement E_1 de probabilité $\mathbb{P}(E_1)$ ou d'un événement E_2 de probabilité $\mathbb{P}(E_2)$, E_1 et E_2 formant un système complet ($\mathbb{P}(E_1) + \mathbb{P}(E_2) = 1$, $\mathbb{P}(E_1 \cap E_2) = 0$)
Soit un ensemble de n expériences identiques à $\mathcal{E}$ et indépendantes . On lui associe les variables X_1 et X_2 représentant respectivement le nombre d'événements E_1 et le nombre d'événements E_2 que l'on peut observer ($X_1 + X_2 = n$). La réalisation effective des n expériences entraîne l'observation des valeurs x_1 de X_1 et x_2 de X_2 ($x_1 + x_2 = n$) ; on dit que les résultats sont répartis en deux classes, les variables X_1 et X_2 prenant le nom d'effectifs de classe.
On désire tester l'hypothèse H_0 : "$\mathbb{P}(E_1) = p_1$ et $\mathbb{P}(E_2) = p_2$" contre l'hypothèse H_1 : "$\mathbb{P}(E_1) \neq p_1$ et $\mathbb{P}(E_2) \neq p_2$"
On sait déjà résoudre le problème. En effet, compte-tenu de la relation $\mathbb{P}(E_1) + \mathbb{P}(E_2) = 1$, il suffit de tester $\mathbb{P}(E_1) = p_1$ contre $\mathbb{P}(E_1) \neq p_1$, ce que l'on peut faire à l'aide de la variable $X_1 \sim \mathcal{B}(n, p_1)$
X_1 admet pour loi asymptotique, lorsque n augmente indéfiniment, la loi $\mathcal{N}(np_1, np_1(1-p_1))$; ce dernier modèle est utilisé dans la pratique pour $n \geq 100$, à condition que les produits np_1 et $n(1-p_1)$ aient une valeur minimum de l'ordre de 5 à 10. On construit

alors un test bilatéral avec la variable $Y = \frac{X_1 - np_1}{\sqrt{np_1(1-p_1)}}$ considérée comme pratiquement normale centrée réduite, sous H_0.
Soit maintenant la variable :

$$Z = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2}$$

On a :

$$\begin{aligned}
Z &= \frac{p_2(X_1 - np_1)^2 + p_1(X_2 - np_2)^2}{np_1p_2} \\
&= \frac{(1-p_1)(X_1 - np_1)^2 + p_1(n - X_1 - n(1-p_1))^2}{np_1(1-p_1)} \\
&= \frac{(1-p_1)(X_1 - np_1^2) + p_1(np_1 - X_1)^2}{np_1(1-p_1)} \\
&= \frac{(X_1 - np_1)^2}{np_1(1-p_1)} \\
&= Y^2
\end{aligned}$$

Étant donné la relation $Z = Y^2$ et étant donné le comportement asymptotique de Y, il est clair que Z admet pour loi asymptotique une loi de χ^2_1, sous H_0.
D'autre part, pour un niveau α, on peut écrire :

$$1 - \alpha = \mathbb{P}(-y_{\alpha/2} < Y < y_{\alpha/2}) = \mathbb{P}(0 \leq Y^2 < y^2_{\alpha/2}) = \mathbb{P}(0 \leq Z < z_\alpha) \text{ avec } z_\alpha = y^2_{\alpha/2}$$

Il revient donc au même d'effectuer un test bilatéral sur Y que d'effectuer un test unilatéral sur Z avec domaine de rejet à droite.
On dispose donc d'une deuxième méthode pour tester H_0 contre H_1. La variable Z, qui peut être assimilée sous les conditions précédentes à une variable χ^2_1 ($n \geq 100$, np_1 et $np_2 \geq$ 5 à 10), rend compte globalement des écarts entre les effectifs de classe et leur espérance (souvent appelées "effectifs théoriques"). Une valeur nulle de Z signifie l'accord parfait avec H_0 ; une valeur élevée de Z conduit au rejet de H_0, la borne supérieure z_α de l'intervalle d'acceptation ($3.841 = (1.96)^2$ au niveau 5% ; $6.635 = (2.576)^2$ au niveau 1%) étant lue dans les tables des lois de χ^2.

2.2 *Distribution à r classes*

Plus généralement, soit une expérience aléatoire $\mathcal{E}$ susceptible d'entraîner la réalisation d'un événement E_1 de probabilité $\mathbb{P}(E_1)$ ou E_2 avec la probabilité $\mathbb{P}(E_2)$, ..., ou E_r avec la probabilité $\mathbb{P}(E_r)$; les événements $E_1, E_2, \ldots, E_r$ formant un système complet ($\sum_{i=1}^r \mathbb{P}(E_i) = 1, \mathbb{P}(E_i \cap E_j) = 0$ pour $i \neq j$)
Les résultats de n expériences identiques à $\mathcal{E}$ et indépendantes sont donc répartis en r classes. A un tel ensemble d'expériences, on associe les variables $X_1, X_2, \ldots, X_r$ représentant les effectifs de classe.

Le système $(X_1, X_2, \ldots, X_r)$ obéit à une loi multinomiale (voir exercice 3 du chapitre 6).
On veut tester l'hypothèse H_0 : "$\mathbb{P}(E_1) = p_1$ et $\mathbb{P}(E_2) = p_2$, ..., et $\mathbb{P}(E_r) = p_r$" contre l'hypothèse H_1 : "$\mathbb{P}(E_1) \neq p_1$ ou $\mathbb{P}(E_2) \neq p_2$, ..., ou $\mathbb{P}(E_r) \neq p_r$"
On considère la variable :

$$Z = \sum_{i=1}^{r} \frac{(X_i - np_i)^2}{np_i}$$

On montre que, sous H_0, cette variable admet comme loi asymptotique une loi de χ^2_{r-1}.
Pratiquement, on utilise ce modèle pour $n \geq 100$, à conditions que les effectifs théoriques np_i aient tous une valeur minimum de l'ordre de 5 à 10.
Il est clair que sous H_1 les valeurs élevées de Z sont plus probables que sous H_0. On construit donc un test unilatéral avec domaine de rejet à droite.
On remarquera que la notion de degrés de liberté est assez compréhensible. En fait il n'y a, parmi les variables $X_1, X_2, \ldots, X_r$, que $r - 1$ variables indépendantes. En effet les variables sont liées par la relation : $X_1 + X_2 + \cdots + X_r = n$, dès que le hasard a attribué une valeur numérique à $r - 1$ variables, la valeur de la dernière est imposée.

Exemple 10.1 *Partant de races pures, un sélectionneur a croisé des mufliers ivoires avec des mufliers rouges. Il a obtenu en F1 des mufliers pâles puis en F2, après autofécondation des plantes de la génération F1 : 22 mufliers rouges, 52 mufliers pâles, 23 mufliers ivoires.*
La couleur des fleurs est-elle gérée par un couple d'allèles ?
Soient p_1, p_2, p_3 les probabilités pour qu'une plante de la F2 ait respectivement des fleurs rouges, pâles ou ivoires. Soient X_1, X_2, X_3 les variables représentant le nombre de plantes à fleurs rouges, pâles et ivoires que l'on peut observer sur 97 plantes.
On est amené à tester, après un raisonnement génétique élémentaire, l'hypothèse H_0 : $p_1 = 1/4$, $p_2 = 1/2$, $p_3 = 1/4$ contre l'hypothèse H_1 : $p_1 \neq 1/4$, ou $p_2 \neq 1/2$, ou $p_3 \neq 1/4$.
D'où le tableau :

Phénotypes	*Rouge*	*Pâle*	*Ivoire*	*Total*
Probabilité	$1/4$	$1/2$	$1/4$	*1*
Effectif théorique	*24.25*	*48.50*	*24.25*	*97*
Effectif observé	*22*	*52*	*23*	*97*

Les conditions d'application du test χ^2 sont satisfaites, à savoir :
- *Les classes constituent un système complet d'événements ;*
- *Les 97 expériences sont identiques et indépendantes ;*
- *leur nombre est assez élevé ;*
- *les effectifs théoriques sont suffisamment élevés.*

Dans ces conditions, sous H_0, la variable $Z = \sum_{i=1}^{3} \frac{(X_i - 97p_i)^2}{97p_i}$ est pratiquement une variable χ^2_2. On effectue un test unilatéral avec domaine de rejet à droite. L'intervalle d'acceptation de H_0 est, au niveau 5% : $[0, 5.991]$.
On a observé la valeur $z_0 = \frac{(2.25)^2}{24.25} + \frac{(3.50)^2}{48.5} + \frac{(1.25)^2}{24.25} \approx 0.52$
On peut accepter l'hypothèse que la couleur des fleurs des mufliers est gérée par un couple d'allèles.

3 Test d”ajustement

3.1 Principe

Soit une variable Y associée à une expérience $\mathcal{E}$. On veut tester l'hypothèse H_0 : ”Y a une certaine loi de probabilité” contre l'hypothèse H_1 : ”Y a une loi différente”. Un tel test s'appelle un test d'ajustement.
Le principe du test du χ^2 d'ajustement est le suivant :

- On effectue une partition $E_1, E_2, \ldots, E_r$ de l'ensemble des valeurs possibles de Y, ce qui revient à constituer un ensemble de r classes auxquelles la loi supposée de Y permet d'affecter les probabilités $p_1, p_2, \ldots, p_r$.
- Chaque fois que l'on effectue une expérience $\mathcal{E}$, on peut dire dans quelle classe la valeur observée de Y se trouve. Donc on peut associer à un ensemble de n expériences indépendantes, identiques à $\mathcal{E}$, les effectifs de classe $X_1, X_2, \ldots, X_r$ et effectuer un test χ^2 comme précédemment.

Il doit être cependant clair qu'en réalité on teste ainsi, non pas H_0, mais l'ensemble $(p_1, p_2, \ldots, p_r)$ des probabilités affectées aux classes. Ce n'est que dans le cas d'une variable discrète, si chaque valeur observable constitue réellement une classe, qu'on teste véritablement la loi de Y. Dans le cas d'une variable continue, en particulier, pour que le test garde une certaine signification, il faut un nombre de classes suffisant ; il serait tout à fait ridicule de vouloir tester une loi normale en constituant deux classes de part et d'autre de l'espérance, chacune de probabilité $1/2$.
Bien entendu les conditions d'application du test χ^2 doivent être toujours respectées. On sera souvent amené, par exemple à regrouper des classes d'effectifs théoriques trop faibles.

Exemple 10.2 *On admet que la coloration de l'iris, chez l'homme, est déterminée par un couple d'allèles. La diversité des gènes complique l'étude de la transmission de ce caractère ; on sait cependant que la coloration bleue est récessive.*
Le père de Monsieur Dupont et le père de Madame Dupont ont les yeux bleus. Monsieur et Madame Dupont n'ont pas les yeux bleus ; étant hétérozygotes, s'ils attendent un enfant, la probabilité pour qu'il ait les yeux bleus est $1/4$. Sur cinq enfants, le nombre d'enfants aux yeux bleus qu'ils peuvent avoir obéit à une loi binomiale $\mathcal{B}(5, 1/4)$
On a classé selon le nombre d'enfants aux yeux bleus qu'elles contiennent 1024 familles de 5 enfants et dont les parents ont le même génotype que Monsieur et Madame Dupont.
Soit Y le nombre d'enfants aux yeux bleus d'une telle famille.
On se propose de tester $H_0 : Y \sim \mathcal{B}(5, 1/4)$ contre $H_1 : Y \not\sim \mathcal{B}(5, 1/4)$.

Nombre d'enfants aux yeux bleus	*0*	*1*	*2*	*3*	*4*	*5*	*Total*
Probabilité (sous H_0)	$\frac{243}{1024}$	$\frac{405}{1024}$	$\frac{270}{1024}$	$\frac{90}{1024}$	$\frac{15}{1024}$	$\frac{1}{1024}$	*1*
Effectif théorique	*243*	*405*	*270*	*90*	$\underbrace{15 \quad 1}_{16}$		*1024*
Nombre observé de familles	*252*	*410*	*265*	*87*	$\underbrace{10 \quad 0}_{10}$		*1024*

Le tableau ci-dessus résume l'ensemble des résultats expérimentaux et théoriques.
Les condition d'application du test χ^2 sont remplies (voir exemple 10.1) :
- *Les classes constituent un système complet d'événements ;*
- *observer les 1024 familles revient à effectuer 1024 expériences identiques et indépendantes ;*
- *le nombre d'observations est suffisant ;*
- *en regroupant les deux dernières classes, les effectifs théoriques sont suffisants.*

Soient X_0, X_1, X_2, X_3, X_4 les effectifs des cinq classes retenues.
Sous H_0, la variable $Z = \frac{(X_0-243)^2}{243} + \frac{(X_1-405)^2}{405} + \frac{(X_2-270)^2}{270} + \frac{(X_3-90)^2}{90} + \frac{(X_4-16)^2}{16}$ est une variable χ^2_4. Au niveau 5%, l'intervalle d'acceptation est $[0; 9.488]$.
On a observé $z_0 = 2.84$. L'hypothèse H_0 est acceptée.

Exemple 10.3 *On se propose de vérifier si le nombre quotidien d'accidents d'automobile, dans une vie donnée, obéit à une loi de Poisson d'espérance 1. Pour cela, on dénombre quotidiennement les accidents qui se sont produits sur une durée de 100 jours. Les résultats expérimentaux et théoriques sont résumés dans le tableau suivant :*

Nombre quotidien d'accidents	*0*	*1*	*2*	*3*	*4, ou plus*	*Total*
Probabilité sous l'hypothèse $\mathcal{P}(1)$	*0.3679*	*0.3679*	*0.1839*	*0.0613*	*0.0190*	*1*
Nombre théorique de jours	*36.79*	*36.79*	*18.39*	$\underbrace{6.13 \quad 1.90}_{8.03}$		*100*
Nombre observé de jours	*35*	*40*	*17*	$\underbrace{6 \quad 2}_{8}$		*100*

Si Y est le nombre d'accidents par jour, on veut tester $H_0 : Y \sim \mathcal{P}(1)$ contre $H_1 : Y \not\sim \mathcal{P}(1)$.
Les conditions d'utilisation du test sont satisfaites (voir exemple 10.1 et 10.2). On notera en particulier la nécessité de prévoir l'événement "4 ou plus" pour obtenir un système complet et le regroupement des deux dernières classes en "3 ou plus" pour obtenir un effectif théorique suffisant (en fait, on a observé 2 fois 4 accidents).
Soient X_0, X_1, X_2, X_3, les effectifs des quatre classes finalement retenues (X_2 : nombre possible de jours, sur 100, avec 2 accidents).
Sous H_0, la variable $Z = \frac{(X_0-36.79)^2}{36.79} + \frac{(X_1-36.79)^2}{36.79} + \frac{(X_2-18.39)^2}{18.39} + \frac{(X_3-8.03)^2}{8.03}$ est une variable χ^2_3.
Sous H_1, les valeurs élevées de Z sont plus probables que sous H_0. Au niveau 5% l'intervalle d'acceptation de H_0 est $[0, 7.815]$.
On a observé la valeur $z_0 \approx 0.47$ de Z. On ne rejette pas l'hypothèse que le nombre quotidien d'accidents de voiture est une variable de Poisson d'espérance 1.

3.2 Estimation de paramètres

Les lois que l'on veut tester, dans un test χ^2 d'ajustement, dépendent en général d'un ou plusieurs paramètres. La loi $\mathcal{B}(n, p)$ dépend de p (en général n est connu), la loi $\mathcal{P}(\mu)$ dépend de μ, la loi $\mathcal{N}(\mu, \sigma^2)$ dépend de μ et de σ^2.

Il peut arriver que l'on désire tester un type de loi en ignorant la valeur exacte de chacun des paramètres qui déterminent complètement la loi. On peut alors procéder à l'estimation ponctuelle de ces paramètres, par l'intermédiaire d'estimateurs, à partir des données expérimentales sur lesquelles on désire effectuer le test.
On sait que, les effectifs de classe étant toujours liés par la relation $X_1 + X_2 + \cdots + X_r = n$, on perd toujours un degré de liberté. On admettre que *toute* estimation d'*un* paramètre introduit une liaison nouvelle entre les effectifs de classe ce qui se traduit par la perte supplémentaire d'un degré de liberté. D'où la règle :

> Dans un test χ^2 d'ajustement, le nombre ν de degrés de liberté est :
>
> $$\nu = r - 1 - p$$
>
> où r est le nombre de classes et où p est le nombre de paramètres estimés à partir des observations.

Bien entendu, si les paramètres sont estimés à partir d'une deuxième série d'observations, indépendante de celle sur laquelle on effectue le test, on est ramené au test d'une loi a priori ($p = 0$)

Exemple 10.4 *À partir des observations précédentes (exemple 10.3), peut-on admettre que le nombre quotidien d'accidents de voiture obéit à une loi de Poisson ?*
Le test précédent devient alors : $H_0 \ : Y \sim \mathcal{P}(\lambda)$ *contre* $H_1 \ : Y \not\sim \mathcal{P}(\lambda)$.
On estime λ*, espérance de* Y*, à partir de* $\overline{Y}$*, moyenne de 100 observations. D'où l'estimation :* $\lambda = \overline{y_0} = \frac{1}{100}(35.0 + 40.1 + 17.2 + 6.3 + 2.4) = 1$.
Il se trouve, par hasard, que l'estimation est égale à l'espérance choisie a priori dans l'exemple 10.3. On aboutira donc au même résultat numérique pour la valeur observée de Z *:* $z_0 \approx 0.47$.
Mais Z *est cette fois une variable* χ^2_2*. Comme* $0.47 < 5.991$*, on ne rejette pas l'hypothèse d'un modèle poissonnien pour le nombre quotidien d'accidents.*

Exemple 10.5 *Soit* Y *le nombre d'impulsions enregistré par un détecteur dans la mesure de l'intensité d'un faisceau de rayons X. On sait que* Y *est une variable de Poisson dont l'espérance* μ *caractérise l'intensité du faisceau. Si* μ *a une valeur assez élevée, on peut écrire* $Y \sim \mathcal{N}(\mu, \mu)$.
On se propose de tester l'hypothèse $H_0 \ : Y \sim \mathcal{N}(\mu, \mu)$ *contre* $H_1 \ : Y \not\sim \mathcal{N}(\mu, \mu)$*, à partir de 300 comptages d'une minute, effectués dans des conditions identiques. Ce test présente un intérêt certain car le modèle probabiliste peut être perturbé par différents phénomènes tels que la stabilité du générateur.*
Les 300 comptages ont une moyenne arithmétique de 10000 coups/mn. Ils ont été répartis en classes d'amplitude de 50 coups/mn..

Nombre de coups y	*Nombre d'observations*		*Probabilité (sous* H_0*)*	*Effectifs théoriques*	
$y < 9775$	3	8	0.0123	3.7	12
$9775 \leq y < 9825$	5		0.0278	8.3	
$9825 \leq y < 9875$	15		0.0655	19.7	
$9875 \leq y < 9925$	45		0.1210	36.3	
$9925 \leq y < 9975$	54		0.1747	52.4	
$9975 \leq y < 10025$	65		0.1974	59.2	
$10025 \leq y < 10075$	48		0.1747	52.4	
$10075 \leq y < 10125$	34		0.1210	36.3	
$10125 \leq y < 10175$	24		0.0655	19.7	
$10175 \leq y < 10225$	5	7	0.0278	8.3	12
$10225 \leq y$	2		0.0123	3.7	
Total	300		1.000	300	

Le test n'est possible qu'après une estimation du paramètre μ. L'espérance μ de Y est estimée à l'aide de la moyenne $\overline{Y}$ de 300 observations.
D'où l'estimation $\mu = \overline{y_0} = 10000$. On en déduit $\sigma = \sqrt{\mu} = 100$.
Pour le calcul des probabilités, il faut passer en variable centrée réduite : $Y' = \frac{Y-10000}{100}$.
On obtient (voir table 11.6) :

$$\mathbb{P}(9975 \leq Y < 10025) = \mathbb{P}(-0.25 \leq Y' < 0.25) = 2\frac{1}{2}\theta(0.25) \approx 0.1974$$

$$\left\{\begin{array}{l} \mathbb{P}(10025 \leq Y < 10075) = \mathbb{P}(0.25 \leq Y' < 0.75) \\ \mathbb{P}(9925 \leq Y < 9975) = \mathbb{P}(-0.75 \leq Y' < -0.25) \end{array}\right. = \frac{1}{2}\theta(0.75) - \frac{1}{2}\theta(0.25) \approx 0.1747$$

$$\left\{\begin{array}{l} \mathbb{P}(10075 \leq Y < 10125) = \mathbb{P}(0.75 \leq Y' < 1.25) \\ \mathbb{P}(9875 \leq Y < 9925) = \mathbb{P}(-1.25 \leq Y' < -0.75) \end{array}\right. = \frac{1}{2}\theta(1.25) - \frac{1}{2}\theta(0.75) \approx 0.1210$$

$$\left\{\begin{array}{l} \mathbb{P}(10125 \leq Y < 10175) = \mathbb{P}(1.25 \leq Y' < 1.75) \\ \mathbb{P}(9825 \leq Y < 9875) = \mathbb{P}(-1.75 \leq Y' < -1.25) \end{array}\right. = \frac{1}{2}\theta(1.75) - \frac{1}{2}\theta(1.25) \approx 0.0655$$

$$\left\{\begin{array}{l} \mathbb{P}(10175 \leq Y < 10225) = \mathbb{P}(1.75 \leq Y' < 2.25) \\ \mathbb{P}(9775 \leq Y < 9825) = \mathbb{P}(-2.25 \leq Y' < -1.75) \end{array}\right. = \frac{1}{2}\theta(2.25) - \frac{1}{2}\theta(1.75) \approx 0.0278$$

$$\left\{\begin{array}{l} \mathbb{P}(Y \geq 10225) = \mathbb{P}(Y' \geq 2.25) \\ \mathbb{P}(Y < 9775) = \mathbb{P}(Y' < -2.25) \end{array}\right. = \frac{1}{2} - \frac{1}{2}\theta(2.25) \approx 0.0123$$

D'où les effectifs théoriques.
Les conditions d'applications du test χ^2 sont satisfaites (voir exemple 10.1 et 10.2). On remarquera en particulier que l'on a veillé à ce que les classes constituent une partition des valeurs possibles de Y ; le regroupement des classes extrêmes permet de remplir la condition sur les effectifs théoriques.
Soient $X_1, X_2, \ldots, X_9$ les effectifs des classes finalement retenues d'effectifs théoriques $c_1, c_2, \ldots, c_9$.
Sous H_0, la variable $Z = \sum_{i=1}^{9} \frac{(X_i - c_i)^2}{c_i}$ est une variable χ^2 à $9 - 1 - 1 = 7$ degrés de liberté (un paramètre estimé).

Sous H_1 les valeurs déviées de Z seront plus probables que sous H_0. D'où le domaine d'acceptation de H_0, au niveau 5% : $[0; 14.067]$.
On a observé la valeur de Z :

$$z_0 = \frac{4^2}{12} + \frac{4.7^2}{19.7} + \frac{8.7^2}{36.3} + \frac{1.6^2}{52.4} + \frac{5.8^2}{59.2} + \frac{4.4^2}{52.4} + \frac{2.3^2}{36.3} + \frac{4.3^2}{19.7} + \frac{5^2}{12} \approx 8.69$$

On en conclut à l'acceptation du modèle probabiliste.

4 Tests d'homogénéité

4.1 Principe

Le test χ^2 est également utilisé pour la comparaison de plusieurs échantillons. Le principe du test va être exposé dans un exemple à deux échantillons. On le généralise sans peine pour plusieurs échantillons.

Exemple 10.6 *On a étudié sur deux échantillons provenant de deux populations différentes la répartition des quatre groupes sanguins : O, A, B, AB. Les résultats obtenus sont reportés dans un tableau dit tableau de contingence, à deux lignes et quatre colonnes :*

Groupe	*O*	*A*	*B*	*AB*	*Total*
1er échantillon	*121*	*120*	*79*	*33*	*353*
2ème échantillon	*118*	*95*	*121*	*30*	*364*
Total	*239*	*215*	*200*	*63*	*717*

On veut tester l'hypothèse H_0 : "les quatre groupes sanguins sont répartis de la même manière sur les deux populations" contre l'hypothèse H_1 : "les répartitions sont différentes". Sous H_0, la probabilité, pour un individu prélevé au hasard, d'être d'un groupe donné est la même dans les deux populations. On ne connaît pas cette probabilité, sinon le problème serait résolu ; on peut cependant l'estimer et, toujours sous H_0, la meilleure estimation que l'on puisse en donner est la proportion des individus de ce groupe observée sur l'ensemble des deux échantillons. C'est ainsi que l'on obtient les estimations :

$$\left.\begin{array}{ll} \textit{Pour le groupe O} & p_1 = {}^{239}\!/_{717} \approx 0.333 \\ \textit{Pour le groupe A} & p_2 = {}^{215}\!/_{717} \approx 0.300 \\ \textit{Pour le groupe B} & p_3 = {}^{200}\!/_{717} \approx 0.279 \\ \textit{Pour le groupe AB} & p_4 = {}^{63}\!/_{717} \approx 0.088 \end{array}\right\} p_1 + p_2 + p_3 + p_4 = 1$$

La relation $p_1 + p_2 + p_3 + p_4 = 1$ montre qu'en fait il suffit de trois paramètres pour déterminer complètement le modèle.
On déduit de l'estimation précédente les effectifs théoriques de chaque classe pour un échantillon de taille 353 d'une part et pour un échantillon de taille 364 d'autre part. D'où le tableau :

Groupe	*O*	*A*	*B*	*AB*	*Total*
1[er] échantillon	*121*	*120*	*79*	*33*	*353*
	(117.7)	*(105.9)*	*(98.5)*	*(31.0)*	
2[ème] échantillon	*118*	*95*	*121*	*30*	*364*
	(121.3)	*(109.1)*	*(101.5)*	*(32.0)*	
Total	*239*	*215*	*200*	*63*	*717*

Les effectifs théoriques sont entre parenthèses. On a, par exemple, $117.7 = 0.333 \times 353$. *Soient maintenant les variables* X_1, X_2, X_3, X_4 *représentant les effectifs de classe du premier échantillon et* Y_1, Y_2, Y_3, Y_4 *représentant les effectifs de classe du deuxième échantillon. On pose :*

$$\begin{aligned} Z &= \frac{(X_1-117.7)^2}{117.7}+\frac{(X_2-105.9)^2}{105.9}+\frac{(X_3-98.5)^2}{98.5}+\frac{(X_4-31.0)^2}{31.0} \\ &\quad +\frac{(Y_1-121.3)^2}{121.3}+\frac{(Y_2-109.1)^2}{109.1}+\frac{(Y_3-101.5)^2}{101.5}+\frac{(Y_4-32.0)^2}{32.0} \end{aligned}$$

Les conditions d'application du test χ^2 *étant satisfaites pour chaque échantillon, sous* H_0 *la variable* Z *peut être considérée comme la somme de deux variables* χ^2*. L'indépendance des deux séries d'observations permet de considérer la variable* Z *comme une variable* χ^2 *(propriété d'additivité des* χ^2*). On est tenté de dire qu'il s'agit d'une variable* χ^2 *à* $2(4-1) = 6$ *degrés de liberté ; cependant l'estimation, à partir des observations, des trois paramètres qui déterminent complètement le modèle probabiliste baisse le nombre de degrés de liberté à* $6-3=3$*. D'où* $Z \sim \chi^2_3$*.*
Les valeurs élevées de Z *étant plus probables sous* H_1 *que sous* H_0 *on construit, comme dans les exemples précédents un test unilatéral avec domaine de rejet à droite ; c'est-à-dire, au niveau 5%, l'intervalle* $[7.815; +\infty[$*.*
On a observé la valeur de Z *:*

$$z_0 = \frac{3.3^2}{117.7}+\frac{14.1^2}{105.9}+\frac{19.5^2}{98.5}+\frac{2^2}{31.0}+\frac{3.3^2}{121.3}+\frac{14.1^2}{109.1}+\frac{19.5^2}{101.5}+\frac{2^2}{32.0} \approx 11.74$$

On peut donc conclure au rejet de H_0 *: les quatre groupes sanguins sont répartis différemment sur les deux populations d'où proviennent les échantillons. La valeur observée est même assez significative puisque, même au niveau 1%, on rejetterait* H_0*.*

4.2 Généralisation

On généralise très facilement au cas de m échantillons répartis en r classes. On est alors obligé d'estimer $r-1$ paramètres.
D'où le nombre de degrés de liberté $\nu = m \times (r-1) - (r-1) = (m-1)(r-1)$.
Les données étant représentées dans un tableau à m lignes et r colonnes, on peut retenir que le nombre de degrés de liberté dans un test χ^2 d'homogénéité est égal au nombre de lignes moins un multiplié par le nombre de colonnes moins un.
Il est bon de résumer les conditions d'utilisation d'un test chi-deux d'homogénéité :
- les classes doivent constituer un système complet d'événements ;

- les observation doivent être indépendantes dans chaque série d'observations et les séries indépendantes entre elles ; chaque série d'expériences doit comporter un grand nombre d'observations ;
- les effectifs théoriques doivent avoir des valeurs suffisamment élevées.

4.3 À propos d'un cas particulier

La comparaison de 2 échantillons dont les observations sont réparties en deux classes peut donc être faite, sous réserve des conditions précédentes, par l'intermédiaire d'un test χ^2_1

Exemple 10.7 *On désire interpréter les résultats suivants : le nombre de guérisons du cancer de la peau a été de 1712 individus sur 2015 patients pour un traitement A (85%) et de 757 individus sur 1010 patients pour un traitement B (75%).*
Le test χ^2 permet de tester l'hypothèse H_0 : "un individu a la même probabilité d'être guéri, estimée à $\frac{757+1712}{2015+1010} \approx 0.8162$, dans les deux traitements" contre H_1 : "les deux traitements sont caractérisés par des probabilités de guérison différentes*".*
On obtient, après calculs :

			guérison	non guérison	
Premier échantillon	*Effectifs*	*observés*	*1712*	*303*	*2015*
		théoriques	*(1644.64)*	*(370.36)*	
Deuxième échantillon	*Effectifs*	*observés*	*757*	*253*	*1010*
		théoriques	*(824.36)*	*(185.64)*	
			2469	*556*	*3025*

Les conditions d'utilisation d'un test χ^2 d'homogénéité sont satisfaites.
En désignant par X_1 et X_2 les effectifs de classe du premier échantillon et par Y_1 et Y_2 ceux du deuxième, la variable :

$$Z = \frac{(X_1 - 1644.64)^2}{1644.64} + \frac{(X_2 - 370.36)^2}{370.36} + \frac{(Y_1 - 824.36)^2}{824.36} + \frac{(Y_2 - 185.64)^2}{185.64}$$

est, sous H_0, assimilée à une variable χ^2_1.
On a observé la valeur de Z, $z_0 = \frac{67.36^2}{1644.64} + \frac{67.36^2}{370.36} + \frac{67.36^2}{824.36} + \frac{67.36^2}{185.64} \approx 44.95$.
Ce résultat est très significatif puisque même à un niveau $<< 10^{-3}$ on rejetterait H_0.
On peut, cependant, construire une deuxième interprétation qui présente l'avantage de permettre de tester H_0 : "un individu a la même probabilité d'être guéri avec les deux traitements" contre H'_1 : "le traitement A donne une probabilité de guérison supérieure *à celle du traitement B". Cette interprétation s'appuie sur la comparaison des deux proportions de guérison.*
On désigne par p_1 la probabilité de guérison avec A et par p_2 la probabilité de guérison avec B, pour un individu quelconque.

En conservant les mêmes notations que plus haut, si X_1 désigne le nombre de guérisons susceptibles d'être obtenues sur 2015 individus traités par A, on a : $X_1 \sim \mathcal{B}(2015, p_1)$. Si Y_1 désigne le nombre de guérisons susceptibles d'être obtenus sur 1010 individus traités par B, on a : $Y_1 \sim \mathcal{B}(1010, p_2)$.
Il est clair que l'on peut écrire, dans cet exemple :

$$\frac{X_1}{2015} \sim \mathcal{N}\left(p_1, \frac{p_1(1-p_1)}{2015}\right) \qquad \textit{et} \qquad \frac{Y_1}{1010} \sim \mathcal{N}\left(p_2, \frac{p_2(1-p_2)}{1010}\right)$$

à cause de la convergence de la loi binomiale vers la loi normale, et, à cause de l'indépendance des deux séries d'observations :

$$\frac{X_1}{2015} - \frac{Y_1}{1010} \sim \mathcal{N}\left(p_1 - p_2, \frac{p_1(1-p_1)}{2015} + \frac{p_2(1-p_2)}{1010}\right)$$

Donc, sous $H_0\ : p_1 = p_2 = p$:

$$\frac{X_1}{2015} - \frac{Y_1}{1010} \sim \mathcal{N}\left(0, p(1-p)\left(\frac{1}{2015} + \frac{1}{1010}\right)\right)$$

Pour pouvoir utiliser le modèle, il faut connaître la variance de cette loi normale. On procède alors à l'estimation de paramètre p. La meilleure estimation en est, sous H_0, bien sûr : $\frac{2469}{3025} \approx 0.8162$. On en déduit, sous H_0 :

$$\frac{X_1}{2015} - \frac{Y_1}{1010} \sim \mathcal{N}(0; 2.23\,10^{-4})$$

Soit maintenant la variable :

$$U = \frac{X_1/2015 - Y_1/1010}{\sqrt{2.23}\,10^{-2}} = 66.96\left(\frac{X_1}{2015} - \frac{Y_1}{1010}\right)$$

Sous H_0, on peut donc assimiler cette variable à une variable $\mathcal{N}(0,1)$. Sous H_1', les valeurs élevées de U sont plus probables que sous H_0. On construit donc un test unilatéral avec domaine de rejet à droite, soit $[1.645; +\infty[$ au niveau 5%.
On a observé la valeur de U :

$$u_0 = 66.96\left(\frac{1712}{2015} - \frac{757}{1010}\right) \approx 6.704$$

Le traitement A est donc meilleur que le traitement B. On remarque que ce résultat est très significatif puisque même à un niveau 10^{-3} on rejetterait H_0.

Si on compare les deux méthodes, on peut d'abord vérifier, relativement aisément que $Z = U^2$ (en particulier : $44.95 = 6.704^2$ aux erreurs d'arrondi près), les conditions d'utilisation étant par ailleurs les mêmes.
Il s'agit donc fondamentalement du même modèle probabiliste mais, dans le deuxième cas, on utilise une variable $\mathcal{N}(0,1)$ et, dans le premier cas, son carré qui est une variable χ_1^2. La deuxième méthode permet les tests unilatéraux.
On a pu remarquer, dans la deuxième méthode, que l'estimation du paramètre p ne joue en fait que sur $\sqrt{p(1-p)}$. Il est évident que l'erreur introduite éventuellement par cette estimation ne doit pas jouer énormément puisque :

- On sait que l'estimation d'une probabilité sur un grand échantillon est assez précise ;
- une sous-estimation de p est compensée, dans une certaine mesure, par une sur-estimation de $1 - p$.

5 Exercices

1. Soit une variable aléatoire discrète associée à une expérience donnée. 640 expérience de ce type, identiques et indépendantes, ont donné les résultats suivants :

valeur de la variable	0	1	2	3	4	5	6
Nombre d'observations	9	48	159	220	138	54	12

   Peut-on accepter l'hypothèse : la variable suit une loi binomiale $\mathcal{B}\left(6, \frac{1}{2}\right)$

2. Vous avez conçu un programme informatique générateur de nombre choisis au hasard dans l'ensemble des 10 premiers entiers. Les mille premiers résultats sont répartis comme suit :

Chiffre	0	1	2	3	4	5	6	7	8	9
Nombre d'observations	120	87	115	103	91	109	92	112	94	77

   Peut-on accepter l'hypothèse que votre programme garantit l'équiprobabilité pour chacun des chiffres.

3. Soit une expérience ϵ à laquelle est associée une variable aléatoire W, à valeurs sur $\{0, 1, 2, 3, 4, 5, 6, 7\}$. Un ensemble de 1280 expériences indépendantes et identiques à ϵ a permis d'observer les résultats suivants :

Valeur de la var.	0	1	2	3	4	5	6	7
Nombre d'obs.	13	64	222	322	369	219	62	9

   Peut-on accepter l'hypothèse : "la variable W suit une loi binomiale" ?

4. Le étudiants préparant le DEUG Sciences de la Nature et de la vie à Saint Jérôme sont répartis en deux sections, mais présentent les mêmes examens. Malgré les précautions prises, se pose chaque année le problème de l'égalité des chances.

   Pour une promotion à l'issue des épreuves de la première session, on vient d'enregistrer les résultats suivants : pour la section A 104 reçus sur 175 candidats, pour la section B 78 reçus sur 144 candidats. Qu'en pensez-vous ?

5. Une enquête a été effectué en milieu hospitalier pour déterminer si l'usage du tabac favorise l'apparition du cancer broncho-pulmonaire. Cette enquête a été menée de la manière suivante :
   - les individus interrogés sont répartis en quatre catégories selon leur consommation journalière en cigarettes : A (non fumeurs), B (de 1 à 9), C (de 10 à 19), D (20 ou plus) ; il s'agit d'une consommation moyenne évaluée sur les deux dernières années précédant l'enquête.
   - un premier échantillon est constitué de cancéreux. Un échantillon témoin a ensuite été choisi parmi les accidentés, c'est-à-dire parmi des patients hospitalisés pour des raisons qui n'ont rien à voir avec le tabac ; de plus, pour éliminer tout autre facteur extérieur, à chaque cancéreux correspond un témoin de même âge, de même sexe et interrogé par le même enquêteur.

À partir des résultats ci-dessous, peut-on conclure à l'influence du tabac ?

Catégorie	A	B	C	D	Total
Cancéreux	25	66	177	334	602
Témoins	130	136	165	171	602
Total	155	202	342	505	1204

6 Solutions

1. On veut tester l'hypothèse H_0 : "la variable U suit une loi $\mathcal{B}\left(6, \frac{1}{2}\right)$" contre l'hypothèse H_1 : "la variable U ne suit pas une loi $\mathcal{B}\left(6, \frac{1}{2}\right)$". Il s'agit d'un test d'ajustement.

   Sous H_0, on peut écrire :

$$\begin{aligned} p_1 = \mathbb{P}(U=0) &= \left(\frac{1}{2}\right)^6 = \frac{1}{64} = \mathbb{P}(U=6) = p_7 \\ p_2 = \mathbb{P}(U=1) &= C_6^1 \left(\frac{1}{2}\right)^6 = \frac{6}{64} = \mathbb{P}(U=5) = p_6 \\ p_3 = \mathbb{P}(U=2) &= C_6^2 \left(\frac{1}{2}\right)^6 = \frac{15}{64} = \mathbb{P}(U=4) = p_5 \\ p_4 = \mathbb{P}(U=3) &= C_6^3 \left(\frac{1}{2}\right)^6 = \frac{20}{64} \end{aligned}$$

   D'où le tableau à 7 classes :

k : valeur de la variable	0	1	2	3	4	5	6
Effectifs théoriques : $640\mathbb{P}(U=k)$	10	60	150	200	150	60	10
Effectifs observés	9	48	159	220	138	54	12

   On considère l'ensemble de variables $(X_1, X_2, X_3, X_4, X_5, X_6, X_7)$ où la variable X_i représente l'effectif de la i$^{\text{ème}}$ classe.

   Soit la variable $Z = \sum_{i=1}^{7} \frac{(X_i - 640p_i)^2}{640p_i}$

   Sous H_0, la variable Z suit pratiquement une loi du chi-deux à 6 degrés de liberté (640 expériences identiques et indépendantes, nombre d'expériences suffisamment élevé, effectifs théoriques suffisamment grand, système complet d'événements).

   Sous H_1, les valeurs élevées de Z sont plus probables ; on construit donc un test unilatéral avec domaine de rejet à droite.

   On a observé $z = \frac{1}{10} + \frac{144}{60} + \frac{81}{150} + \frac{400}{200} + \frac{144}{150} + \frac{36}{60} + \frac{4}{10} = 7$

   Or $\mathbb{P}(Z < 7) \approx 67.91\% < 95\%$. Au niveau 5%, on accepte H_0.

2. On va tester H_0 : la probabilité d'apparition d'un chiffre donné est, à chaque tirage, égale à $1/10$ contre l'hypothèse H_1 : les différents chiffres, à chaque tirage, ne sont pas équiprobables.

   Soient $(N_0, N_1, \ldots, N_9)$ les variables aléatoires représentant le nombre d'apparitions de $0, 1, 2, \ldots, 9$ sur 1000 tirages et soit la variable $Z = \sum_{i=1}^{10} \frac{(N_i - 100)^2}{100}$

   Sous H_0, la variable Z suit approximativement une loi du chi-deux à 9 degrés de liberté (100 expériences identiques et indépendantes, nombre d'expériences suffisamment élevé, effectifs théoriques (100) supérieurs à 10, système complet d'événements).

   Sous H_1, Z aura tendance à prendre des valeurs plus élevés que sous H_0 ; on construit donc un test unilatéral avec domaine de rejet à droite.

Au niveau 5%, le domaine d'acceptation de H_0 est : $[0, 16.919[$; le domaine de rejet de H_0 est $[16.919, +\infty[$.

On a observé $z_0 = 17.36$. Au niveau 5%, on rejette l'hypothèse d'équiprobabilité ; il faut revoir le programme.

3. Soit W la variable aléatoire. On veut tester l'hypothèse H_0 : "la variable W suit une loi binomiale" contre l'hypothèse H_1 : "la variable W ne suit pas une loi binomiale". Il s'agit d'un test non paramétrique d'ajustement.

   Pour pouvoir effectuer le test, il faut calculer la probabilité sous H_0 affectée à chacun des observables. Si $W \sim \mathcal{B}(n,p)$, sachant que $n = 7$, le calcul sera rendu possible par l'estimation de p à partir des observations.

   On estimera dans un premier temps E(W)=np par $\overline{W}$, moyenne empirique des observations :

$$\hat{np} = \overline{W} = \frac{1}{1280}(0 \times 13 + 1 \times 64 + 3 \times 322 + \cdots + 7 \times 9) = 3.5$$

   D'où $\hat{p} = 1/2$

   Sous H_0, on peut alors considérer que P($W = 0$)=$\left(\frac{1}{2}\right)^7 = \frac{1}{128}$= P($W = 7$) ; P($W = 1$)=7 $\left(\frac{1}{2}\right)^7 = \frac{7}{128}$= P($W = 6$) ; P($W = 2$)=21 $\left(\frac{1}{2}\right)^7 = \frac{21}{128}$ = P($W = 5$) ; P($W = 3$)=35 $\left(\frac{1}{2}\right)^7 = \frac{35}{128}$ = P($W = 4$)

   D'où le tableau à huit classes :

Observations	13	64	222	322	369	219	62	9
Théorique	10	70	210	350	350	210	70	10

   On considère l'ensemble des variables ($X_0, X_1, X_3, \ldots, X_6, X_7$), où X_i représente l'effectif possible de la i$^{\text{ème}}$ classe et on définit la variable : $Z = \frac{(X_0-10)^2}{10} + \frac{(X_1-70)^2}{70} + \frac{(X_2-210)^2}{210} + \frac{(X_3-350)^2}{350} + \frac{(X_4-350)^2}{350} + \frac{(X_5-210)^2}{210} + \frac{(X_6-70)^2}{70} + \frac{(X_7-10)^2}{10}$ Sous H_0, Z suit approximativement une loi du chi-deux (système complet d'événements, 1280 expériences identiques et indépendantes, grand nombre d'expériences, tous les effectifs supérieurs à 10) à six degrés de liberté (8-1-1 puisqu'un paramètre est estimé).

   Sous H_1, les valeurs élevées de Z sont plus probables. On construit donc, au niveau 5% un test unilatéral avec domaine de rejet à droite. On a observé $z = 6.77$. Puisque $z < 12.592$, z est dans le domaine d'acceptation de $H_0 \Rightarrow W$ suit une loi binomiale.

4.

	Reçus	Collés	Total
A	104 (99.84)	71 (75.16)	175
B	78 (82.16)	66 (61.84)	144
Total	182	137	319

   On désigne par p_1 la probabilité de réussite d'un étudiant en A et par p_2 la probabilité de réussite d'un étudiant en B.

   On veut tester $H_0 : p_1 = p_2 = p$ contre $H_1 : p_1 \neq p_2$

Sous H_0, la valeur commune p peut être estimé par $\hat{p} = 182/319 \approx 0.5705$ On note X_1 le nombre d'étudiants qui étaient susceptibles d'être reçus en A et X_2 le nombre d'étudiants qui étaient susceptibles d'être collés en A.

De même, on note Y_1 le nombre d'étudiants qui étaient susceptibles d'être reçus en B et Y_2 le nombre d'étudiants qui étaient susceptibles d'être collés en B.

Sous H_0 : E(X_1)=$\approx 175\hat{p} \approx 99.84$ E(X_2)=175$(1-\hat{p}) \approx 75.16$
et E(Y_1)=$\approx 144\hat{p} \approx 82.16$ E(Y_2)=144$(1-\hat{p}) \approx 61.84$

On notera que l'ensemble $\{$"reçu,"collé"$\}$ est un système complet d'événements : l'indépendance des résultats des étudiants et des sections, la taille des deux échantillons supérieurs à 100 et les effectifs "théoriques" largement supérieurs à 10.

Sous H_0 et compte tenu du paragraphe précédent, la variable $Z = \frac{(X_1-99.84)^2}{99.84} + \frac{(X_2-75.16)^2}{75.16} + \frac{(Y_1-82.16)^2}{82.16} + \frac{(Y_2-61.84)^2}{61.84}$ suit une loi χ^2_4. Sous H_1, Z aura tendance à prendre des valeurs plus élevées que sous H_0.

D'où un test unilatéral avec domaine de rejet à droite : $]3.841; +\infty[$ au niveau 5%

On a observé $z \approx 0.894$. On est amené à accepter H_0 : pas de différence significative entre les deux résultats.

5. On veut tester H_0 : le tabac a une influence sur l'apparition du cancer contre l'hypothèse H_1 : le tabac n'a pas d'influence.

   On va appliquer le test khi-deux de Pearson appliqué à la comparaison de deux échantillons. On désigne par :
   - p_A^C et p_A^T, les probabilités pour qu'un individu choisi parmi les cancéreux, ou parmi les accidentés, soit non fumeur.
   - p_B^C et p_B^T, les probabilités pour qu'un individu choisi parmi les cancéreux, ou parmi les accidentés, fume de 1 à 9 cigarettes.
   - p_C^C et p_C^T, les probabilités pour qu'un individu choisi parmi les cancéreux, ou parmi les accidentés,fume de 10 à 19 cigarettes.
   - p_D^C et p_D^T, les probabilités pour qu'un individu choisi parmi les cancéreux, ou parmi les accidentés, fume plus de 20 cigarettes.

   Sous H_0
   - $p_A^C = p_A^T = p_A$ que l'on peut estimer par $\widehat{p_A} = \frac{155}{1204}$. Les deux échantillons étant de même taille, on en déduit les effectifs théoriques, pour A, égaux à $602\widehat{p_A} = 77.5$.
   - $p_B^C = p_B^T = p_B$ que l'on peut estimer par $\widehat{p_B} = \frac{202}{1204}$. D'où des effectifs théoriques, pour B, égaux à $602\widehat{p_B} = 101$.
   - $p_C^C = p_C^T = p_C$ que l'on peut estimer par $\widehat{p_C} = \frac{342}{1204}$. D'où des effectifs théoriques, pour C, égaux à $602\widehat{p_B} = 171$.
   - $p_D^C = p_D^T = p_D$ que l'on peut estimer par $\widehat{p_D} = \frac{505}{1204}$. D'où des effectifs théoriques, pour D, égaux à $602\widehat{p_B} = 252.5$.

   Soient X_A, X_B, X_C, X_D et Y_A, Y_B, Y_C, Y_D les effectifs possibles pour les quatre classes envisagées et les deux échantillons.

   Soit la variable : $Z = \frac{(X_A-77.5)^2}{77.5} + \frac{(X_B-101)^2}{101} + \frac{(X_C-171)^2}{171} + \frac{(X_D-252.5)^2}{252.5} + \frac{(Y_A-77.5)^2}{77.5} + \frac{(Y_B-101)^2}{101} + \frac{(Y_C-171)^2}{171} + \frac{(Y_D-252.5)^2}{252.5}$ On a un système complet de classes, les résultats observables sur différents individus cancéreux ou témoins, sont indépendants, les échantillons sont grands et les effectifs théoriques supérieurs à 10. Dans ces conditions, on peut dire que, sous H_0, Z suit approximativement une loi du khi-deux à 3 degrés de liberté (tableau de contingence à 2 lignes et 4 colonnes)

Sous H_1 les effectifs possibles auront tendance à s'éloigner des effectifs théorique calculés qui représentent leur espérance sous H_0. Donc Z aura tendance à prendre des valeurs plus élevées que sous H_0.

On construit donc un test unilatéral avec domaine de rejet de H_0 à droite.

Au niveau 5%, on a :

- Domaine d'acceptation de H_0 : $[0, 7.815[$
- Domaine de rejet de H_0 : $[7.815, +\infty[$

On a observé $z = 2\frac{(52.5)^2}{77.5} + 2\frac{(35)^2}{101} + 2\frac{(6)^2}{171} + 2\frac{(81.5)^2}{252.5} \approx 148.42$ On rejette donc H_0. Ce résultat est même très significatif.

Le tabac a donc une influence sur l'apparition du cancer broncho-pulmonaire. La comparaison des effectifs théoriques et observés montrent clairement qu'elle est positive.

Chapitre 11

Tables de valeurs numériques

Tableau 11.1 – Factorielles

n	$n!$
3	6
4	24
5	120
6	720
7	5040
8	40320
9	362880
10	3628800
11	39916800
12	479001600
13	6227020800
14	87178291200
15	1307674368000
16	20922789888000
17	355687428096000
18	6402373705728000
19	121645100408832000
20	2432902008176640000

Tableau 11.2 – Test du signe

Soit la variable, à valeurs entières, $X \sim \mathcal{B}\left(n, \frac{1}{2}\right)$. La table donne, en fonction de n et du niveau α, le plus grand entier k, tel que $\mathbb{P}(X \leq k) \leq \alpha$ pour un test unilatéral ou tel que $2\mathbb{P}(X \leq k) \leq \alpha$ pour un test bilatéral.

	α (test bilatéral)					α (test bilatéral)			
	0.01	0.05	0.10	0.25		0.01	0.05	0.10	0.25
	α (test unilatéral)					α (test unilatéral)			
n	0.005	0.025	0.05	0.125	n	0.005	0.025	0.05	0.125
1					46	13	15	16	18
2					47	14	16	17	19
3				0	48	14	16	17	19
4				0	49	15	17	18	19
5			0	0	50	15	17	18	20
6		0	0	1	51	15	18	19	20
7		0	0	1	52	16	18	19	21
8	0	0	1	1	53	16	18	20	21
9	0	1	1	2	54	17	19	20	22
10	0	1	1	2	55	17	19	20	22
11	0	1	2	3	56	17	20	21	23
12	1	2	2	3	57	18	20	21	23
13	1	2	3	3	58	18	21	22	24
14	1	2	3	4	59	19	21	22	24
15	2	3	3	4	60	19	21	23	25
16	2	3	4	5	61	20	22	23	25
17	2	4	4	5	62	20	22	24	25
18	3	4	5	6	63	20	23	24	26
19	3	4	5	6	64	21	23	24	26
20	3	5	5	6	65	21	24	25	27
21	4	5	6	7	66	22	24	25	27
22	4	5	6	7	67	22	25	26	28
23	4	6	7	8	68	22	25	26	28
24	5	6	7	8	69	23	25	27	29
25	5	7	7	9	70	23	26	27	29
26	6	7	8	9	71	24	26	28	30
27	6	7	8	10	72	24	27	28	30
28	6	8	9	10	73	25	27	28	31
29	7	8	9	10	74	25	28	29	31
30	7	9	10	11	75	25	28	29	32
31	7	9	10	11	76	26	28	30	32
32	8	9	10	12	77	26	29	30	32
33	8	10	11	12	78	27	29	31	33
34	9	10	11	13	79	27	30	31	33
35	9	11	12	13	80	28	30	32	34
36	9	11	12	14	81	28	31	32	34
37	10	12	13	14	82	28	31	33	35
38	10	12	13	14	83	29	32	33	35
39	11	12	13	15	84	29	32	33	36
40	11	13	14	15	85	30	32	34	36
41	11	13	14	16	86	30	33	34	37
42	12	14	15	16	87	31	33	35	37
43	12	14	15	17	88	31	34	35	38
44	13	15	16	17	89	31	34	36	38
45	13	15	16	18	90	32	35	36	39

Pour de plus grandes valeurs de n, utiliser l'approximation par une loi normale.

Tableau 11.3 – Intervalle de confiance d'une probabilité (loi binomiale)

Soit la variable, à valeurs entières, $X \sim \mathcal{B}(n, p)$. La table donne ou permet de calculer en fonction de n et de la valeur observée x_0, les bornes, exprimées en pour cent, de l'intervalle de confiance de p pour un coefficient de sécurité de 95%.

Exemple : $n = 10$ et $x_0 = 7 \Rightarrow 34.8\% \leq p \leq 93.3\%$

$n \backslash x_0$	0	1	2	3	4	5	6	7	8	9	10
4	0 60.2	0.6 80.6	6.8 93.2	19.4 99.4	39.8 100						
5	0 52.2	0.5 71.6	5.3 85.3	14.7 94.7	28.4 99.5	47.8 100					
6	0 45.9	0.4 64.1	4.3 77.7	11.8 88.2	22.3 95.7	35.9 99.6	54.1 100				
7	0 41.0	0.4 57.9	3.7 71	9.9 81.6	18.4 90.1	29 96.3	42.1 99.6	59 100			
8	0 36.9	0.3 52.7	3.2 65.2	8.5 75.5	15.7 84.3	24.5 91.5	34.8 96.8	47.3 99.7	63.1 100		
9	0 33.6	0.3 48.3	2.8 60	7.5 70.1	13.7 78.8	21.2 86.3	29.9 92.5	40.0 97.2	51.7 99.7	66.4 100	
10	0 30.8	0.3 44.5	2.5 55.6	6.7 65.2	12.2 73.8	18.7 81.3	26.2 81.8	34.8 93.3	44.4 97.5	55.5 99.7	69.2 100

Tableau 11.4 – Lois de Poisson

Soit la variable, à valeurs entière, $X \sim \mathcal{P}(\lambda)$. La table donne, pour différentes valeurs de λ, les probabilités $\mathbb{P}(X = k)$.
Exemple : si $\lambda = 0.9$, $\mathbb{P}(X = 3) = 4.94\%$ et $\mathbb{P}(X = 5) = 0.20\%$

$k \backslash \lambda$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	0.9048	0.8181	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659	0.3679
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647	0.1839
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494	0.0613
4		0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111	0.0153
5				0.0001	0.0002	0.0004	0.0007	0.0012	0.0020	0.0031
6							0.0001	0.0002	0.0003	0.0001

Tableau 11.5 – Intervalle de confiance pour une espérance (loi de Poisson)
Soit la variable, à valeurs entière, $X \sim \mathcal{P}(\lambda)$. La table donne,à partir de la valeur observée x_0, les bornes de l'intervalle de confiance de λ pour un coefficient de sécurité de 95%.
Exemple : si $x_0 = 34 \Rightarrow 23.5 \leq \lambda \leq 47.5$

x_0	0	1	2	3	4	5	6	7	8	9
0		0.025	0.24	0.62	1.09	1.62	2.20	2.81	3.45	4.12
		5.572	7.22	8.76	10.24	11.67	13.06	14.42	16.76	17.08
10	4.8	5.5	6.2	6.9	7.6	8.3	9.0	9.8	10.6	11.4
	18.4	19.7	21.0	22.3	23.6	24.9	26.1	27.3	28.5	29.7
20	12.2	13.0	13.8	14.6	15.4	16.2	17.0	17.8	18.6	19.4
	30.9	32.1	33.3	34.5	35.7	36.9	38.1	39.3	40.5	41.7
30	20.2	21.0	21.8	22.7	23.5	24.4	25.2	26.1	26.9	27.8
	42.8	44.0	45.2	46.3	47.5	48.6	49.8	50.9	52.1	53.3
40	28.6	29.5	30.3	31.2	32.0	32.9	33.7	34.6	35.4	36.3
	54.4	55.5	56.7	57.8	59.0	60.1	61.3	62.5	63.6	64.7
50	37.1	38.0	38.8	39.7	40.5	41.4	42.3	43.1	44.0	44.9
	65.9	67.0	68.2	69.3	70.5	71.6	72.7	73.9	75.0	76.1
60	45.8	46.6	47.5	48.4	49.3	50.2	51.0	51.9	52.8	53.7
	77.2	78.4	79.5	80.6	81.8	82.9	84.0	85.1	86.2	87.3
70	54.6	55.5	56.3	57.2	58.1	59.0	59.9	60.8	61.7	62.6
	88.4	89.5	90.7	91.8	92.9	94.0	95.1	96.2	97.3	98.4
80	63.4	64.3	65.2	66.1	67.0	67.9	68.8	69.7	70.8	71.5
	99.6	100.7	101.8	102.9	104.0	105.1	106.2	107.3	108.4	109.5
90	72.4	73.3	74.2	75.1	76.0	76.9	77.8	78.7	79.6	80.5
	110.6	111.7	112.8	113.9	115.0	116.1	117.2	118.3	119.4	120.5

Tableau 11.6 – Loi normale centrée réduite

Soit la variable, à valeurs réelles $X \sim \mathcal{N}(0,1)$, de densité ρ. La table donne, pour différentes valeurs de x positives :

$$y = \rho(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ et } \frac{1}{2}\theta(x) = \int_0^x \rho(t)\mathrm{d}t$$

x	y	$\frac{1}{2}\theta(x)$	x	y	$\frac{1}{2}\theta(x)$	x	y	$\frac{1}{2}\theta(x)$	x	y	$\frac{1}{2}\theta(x)$
0	0.3989	0	0.7	0.3123	0.2580	1.45	0.1394	0.4265	2.4	0.0224	0.4918
0.05	0.3984	0.0199	0.75	0.3011	0.2734	1.5	0.1295	0.4332	2.5	0.0175	0.4938
0.1	0.3970	0.0398	0.8	0.2897	0.2881	1.55	0.1200	0.4394	2.6	0.0136	0.4953
0.15	0.3945	0.0596	0.85	0.2780	0.3023	1.6	0.1109	0.4452	2.7	0.0104	0.4965
0.2	0.3910	0.0793	0.9	0.2661	0.3159	1.65	0.1023	0.4505	2.8	0.0079	0.4974
0.25	0.3867	0.0987	0.95	0.2541	0.3289	1.7	0.0940	0.4554	2.9	0.0060	0.4981
0.3	0.3814	0.1179	1	0.2420	0.3413	1.75	0.0863	0.4599	3	0.0044	0.49865
0.35	0.3752	0.1368	1.05	0.2299	0.3531	1.8	0.0790	0.4641	3.2	0.0024	0.49931
0.4	0.3683	0.1554	1.1	0.2179	0.3643	1.85	0.0721	0.4678	3.4	0.0012	0.49966
0.45	0.3605	0.1736	1.15	0.2059	0.3749	1.9	0.0656	0.4713	3.6	0.0006	0.499841
0.5	0.3521	0.1915	1.2	0.1942	0.3849	1.95	0.0596	0.4744	3.8	0.0003	0.499928
0.55	0.3429	0.2088	1.25	0.1826	0.3944	2	0.0540	0.4772	4	0.0001	0.499928
0.6	0.3332	0.2257	1.3	0.1714	0.4032	2.1	0.0440	0.4821	4.5		0.499997
0.65	0.3230	0.2422	1.35	0.1604	0.4115	2.2	0.0355	0.4861			
0.6745		0.2500	1.4	0.1497	0.4192	2.3	0.0283	0.4893			

Tableau 11.7 – Loi normale centrée réduite

Soit la variable, à valeurs réelles, $X \sim \mathcal{N}(0,1)$, de densité ρ. La table donne, pour différentes valeurs de u positives :

$$F(u) = \mathbb{P}(X < u) = \int_{-\infty}^{u} \rho(x)\mathrm{d}x$$

Exemple : $F(1.35) = \mathbb{P}(X < 1.35) = 0.9115 = 0.5 + \frac{1}{2}\theta(1.35)$ (voir table 11.6)

u	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Pour $u < 0$, on a $\mathbb{P}(X < u) = 1 - F(-u)$

Tableau 11.8 – Loi normale centrée réduite

Soit la variable, à valeurs réelles, $X \sim \mathcal{N}(0,1)$, de densité ρ. Si u est un nombre positif, et si on pose :

$$\alpha = \mathbb{P}(|X| \geq u) = 2(1 - F(u))$$

la table donne u pour différentes valeurs de α.
Exemple :

$$\begin{aligned} \alpha = \mathbb{P}(|X| \geq 1.2) &\approx 0.23 \\ &= 2(1 - F(1.2)) \text{ (voir table 11.6)} \\ &= 1 - \theta(1.2) \text{ (voir table 11.7)} \end{aligned}$$

u	$\alpha \times 100$	u	$\alpha \times 100$	u	$\alpha \times 100$
0.00	0.00	1.50	86.64	3.00	99.730
0.05	3.99	1.55	87.89	3.05	99.771
0.10	7.97	1.60	89.04	3.10	99.806
0.15	11.92	1.65	90.11	3.15	99.837
0.20	15.85	1.70	91.09	3.20	99.863
0.25	19.74	1.75	91.99	3.25	99.885
0.30	23.58	1.80	92.81	3.30	99.903
0.35	27.37	1.85	93.57	3.35	99.919
0.40	31.08	1.90	94.26	3.40	99.933
0.45	34.73	1.95	94.88	3.45	99.944
0.50	38.29	2.00	95.45	3.50	99.953
0.55	41.77	2.05	95.96	3.55	99.961
0.60	45.15	2.10	96.43	3.60	99.968
0.65	48.43	2.15	96.84	3.65	99.974
0.70	51.61	2.20	97.22	3.70	99.978
0.75	54.67	2.25	97.56	3.75	99.982
0.80	57.63	2.30	97.86	3.80	99.986
0.85	60.47	2.35	98.12	3.85	99.988
0.90	63.19	2.40	98.36	3.90	99.990
0.95	65.79	2.45	98.57	3.95	99.992
1.00	68.27	2.50	98.76	4.00	99.9937
1.05	70.63	2.55	98.92	4.05	99.9949
1.10	72.87	2.60	99.07	4.10	99.9959
1.15	74.99	2.65	99.20	4.15	99.9967
1.20	76.99	2.70	99.31	4.20	99.9973
1.25	78.87	2.75	99.40	4.25	99.9979
1.30	80.64	2.80	99.49	4.30	99.9983
1.35	82.30	2.85	99.56	4.35	99.9986
1.40	83.85	2.90	99.63	4.40	99.9989
1.45	85.29	2.95	99.68	4.45	99.9991

Tableau 11.9 – Lois de Student

Soit T une variable de Student à ν degrés de liberté, de densité ρ. Si u est un nombre positif et si on pose :

$$F(u) = \mathbb{P}(T < u) = P$$

la table donne u pour différentes valeurs de ν et de P.

ν \ P	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995
1	0.1584	0.3249	0.5095	0.7265	1.0000	1.3764	1.9626	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.1421	0.2887	0.4447	0.6172	0.8165	1.0607	1.3862	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.1366	0.2767	0.4242	0.5844	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.1338	0.2707	0.4142	0.5686	0.7407	0.9410	1.1896	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.1322	0.2672	0.4082	0.5594	0.7267	0.9195	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.1311	0.2648	0.4043	0.5534	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.1303	0.2632	0.4015	0.5491	0.7111	0.8960	1.1192	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.1297	0.2619	0.3995	0.5459	0.7064	0.8889	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.1293	0.2610	0.3979	0.5435	0.7027	0.8834	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.1289	0.2602	0.3966	0.5415	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.1286	0.2596	0.3956	0.5399	0.6974	0.8755	1.0877	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.1283	0.2590	0.3947	0.5386	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.1281	0.2586	0.3940	0.5375	0.6938	0.8702	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.1280	0.2582	0.3933	0.5366	0.6924	0.8681	1.0763	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.1278	0.2579	0.3928	0.5357	0.6912	0.8662	1.0735	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.1277	0.2576	0.3923	0.5350	0.6901	0.8647	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.1276	0.2573	0.3919	0.5344	0.6892	0.8633	1.0690	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.1274	0.2571	0.3915	0.5338	0.6884	0.8620	1.0672	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.1274	0.2569	0.3912	0.5333	0.6876	0.8610	1.0655	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.1273	0.2567	0.3909	0.5329	0.6870	0.8600	1.0640	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.1272	0.2566	0.3906	0.5325	0.6864	0.8591	1.0627	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.1271	0.2564	0.3904	0.5321	0.6858	0.8583	1.0614	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.1271	0.2563	0.3902	0.5317	0.6853	0.8575	1.0603	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.1270	0.2562	0.3900	0.5314	0.6848	0.8569	1.0593	1.3178	1.7109	2.0639	2.4922	2.7970
25	0.1269	0.2561	0.3898	0.5312	0.6844	0.8562	1.0584	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.1269	0.2560	0.3896	0.5309	0.6840	0.8557	1.0575	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.1268	0.2559	0.3894	0.5306	0.6837	0.8551	1.0567	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.1268	0.2558	0.3893	0.5304	0.6834	0.8546	1.0560	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.1268	0.2557	0.3892	0.5302	0.6830	0.8542	1.0553	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.1267	0.2556	0.3890	0.5300	0.6828	0.8538	1.0547	1.3104	1.6973	2.0423	2.4573	2.7500
40	0.1265	0.2550	0.3881	0.5286	0.6807	0.8507	1.0500	1.3031	1.6839	2.0211	2.4233	2.7045
50	0.1263	0.2547	0.3875	0.5278	0.6794	0.8489	1.0473	1.2987	1.6759	2.0086	2.4033	2.6778
60	0.1262	0.2545	0.3872	0.5272	0.6786	0.8477	1.0455	1.2958	1.6706	2.0003	2.3901	2.6603
70	0.1261	0.2543	0.3869	0.5268	0.6780	0.8468	1.0442	1.2938	1.6669	1.9944	2.3808	2.6479
80	0.1261	0.2542	0.3867	0.5265	0.6776	0.8461	1.0432	1.2922	1.6641	1.9901	2.3739	2.6387
90	0.1260	0.2541	0.3866	0.5263	0.6772	0.8456	1.0424	1.2910	1.6620	1.9867	2.3685	2.6316
100	0.1260	0.2540	0.3864	0.5261	0.6770	0.8452	1.0418	1.2901	1.6602	1.9840	2.3642	2.6259
200	0.1258	0.2537	0.3859	0.5252	0.6757	0.8434	1.0391	1.2858	1.6525	1.9719	2.3451	2.6006

Tableau 11.10 – Lois du chi-deux

Si ν est le nombre de degrés de liberté d'une variable χ^2 de densité ρ, si u est un nombre positif et si on pose :

$$F(u) = \mathbb{P}(T < u) = P$$

la table donne u pour différentes valeurs de ν et de P.

ν \ P	0.01	0.025	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.99
1	0.00	0.00	0.00	0.01	0.06	0.14	0.27	0.45	0.70	1.07	1.64	2.70	3.84	5.02	6.63
2	0.02	0.05	0.10	0.21	0.44	0.71	1.02	1.38	1.83	2.40	3.21	4.60	5.99	7.37	9.21
3	0.11	0.21	0.35	0.58	1.00	1.42	1.86	2.36	2.94	3.66	4.64	6.25	7.81	9.34	11.34
4	0.29	0.48	0.71	1.06	1.64	2.19	2.75	3.35	4.04	4.87	5.98	7.77	9.48	11.14	13.27
5	0.55	0.83	1.14	1.61	2.34	2.99	3.65	4.35	5.13	6.06	7.28	9.23	11.07	12.83	15.08
6	0.87	1.23	1.63	2.20	3.07	3.82	4.57	5.34	6.21	7.23	8.55	10.64	12.59	14.44	16.81
7	1.23	1.68	2.16	2.83	3.82	4.67	5.49	6.34	7.28	8.38	9.80	12.01	14.06	16.01	18.47
8	1.64	2.17	2.73	3.48	4.59	5.52	6.42	7.34	8.35	9.52	11.03	13.36	15.50	17.53	20.09
9	2.08	2.70	3.32	4.16	5.38	6.39	7.35	8.34	9.41	10.65	12.24	14.68	16.91	19.02	21.66
10	2.55	3.24	3.94	4.86	6.17	7.26	8.29	9.34	10.47	11.78	13.44	15.98	18.30	20.48	23.20
11	3.05	3.81	4.57	5.57	6.98	8.14	9.23	10.34	11.52	12.89	14.63	17.27	19.67	21.92	24.72
12	3.57	4.40	5.22	6.30	7.80	9.03	10.18	11.34	12.58	14.01	15.81	18.54	21.02	23.33	26.21
13	4.10	5.00	5.89	7.04	8.63	9.92	11.12	12.33	13.63	15.11	16.98	19.81	22.36	24.73	27.68
14	4.66	5.62	6.57	7.78	9.46	10.82	12.07	13.33	14.68	16.22	18.15	21.06	23.68	26.11	29.14
15	5.22	6.26	7.26	8.54	10.30	11.72	13.02	14.33	15.73	17.32	19.31	22.30	24.99	27.48	30.57
16	5.81	6.90	7.96	9.31	11.15	12.62	13.98	15.33	16.77	18.41	20.46	23.54	26.29	28.84	31.99
17	6.40	7.56	8.67	10.08	12.00	13.53	14.93	16.33	17.82	19.51	21.61	24.76	27.58	30.19	33.40
18	7.01	8.23	9.39	10.86	12.85	14.43	15.89	17.33	18.86	20.60	22.75	25.98	28.86	31.52	34.80
19	7.63	8.90	10.11	11.65	13.71	15.35	16.85	18.33	19.91	21.68	23.90	27.20	30.14	32.85	36.19
20	8.26	9.59	10.85	12.44	14.57	16.26	17.80	19.33	20.95	22.77	25.03	28.41	31.41	34.16	37.56
21	8.89	10.28	11.59	13.23	15.44	17.18	18.76	20.33	21.99	23.85	26.17	29.61	32.67	35.47	38.93
22	9.54	10.98	12.33	14.04	16.31	18.10	19.72	21.33	23.03	24.93	27.30	30.81	33.92	36.78	40.28
23	10.19	11.68	13.09	14.84	17.18	19.02	20.69	22.33	24.06	26.01	28.42	32.00	35.17	38.07	41.63
24	10.85	12.40	13.84	15.65	18.06	19.94	21.65	23.33	25.10	27.09	29.55	33.19	36.41	39.36	42.97
25	11.52	13.11	14.61	16.47	18.93	20.86	22.61	24.33	26.14	28.17	30.67	34.38	37.65	40.64	44.31
26	12.19	13.84	15.37	17.29	19.82	21.79	23.57	25.33	27.17	29.24	31.79	35.56	38.88	41.92	45.64
27	12.87	14.57	16.15	18.11	20.70	22.71	24.54	26.33	28.21	30.31	32.91	36.74	40.11	43.19	46.96
28	13.56	15.30	16.92	18.93	21.58	23.64	25.50	27.33	29.24	31.39	34.02	37.91	41.33	44.46	48.27
29	14.25	16.04	17.70	19.76	22.47	24.57	26.47	28.33	30.28	32.46	35.13	39.08	42.55	45.72	49.58
30	14.95	16.79	18.49	20.59	23.36	25.50	27.44	29.33	31.31	33.53	36.25	40.25	43.77	46.97	50.89
31	15.65	17.53	19.28	21.43	24.25	26.43	28.40	30.33	32.34	34.59	37.35	41.42	44.98	48.23	52.19
32	16.36	18.29	20.07	22.27	25.14	27.37	29.37	31.33	33.38	35.66	38.46	42.58	46.19	49.48	53.48
33	17.07	19.04	20.86	23.11	26.04	28.30	30.34	32.33	34.41	36.73	39.57	43.74	47.39	50.72	54.77
34	17.78	19.80	21.66	23.95	26.93	29.24	31.31	33.33	35.44	37.79	40.67	44.90	48.60	51.96	56.06
35	18.50	20.56	22.46	24.79	27.83	30.17	32.28	34.33	36.47	38.85	41.77	46.05	49.80	53.20	57.34
36	19.23	21.33	23.26	25.64	28.73	31.11	33.25	35.33	37.50	39.92	42.87	47.21	50.99	54.43	58.61
37	19.96	22.10	24.07	26.49	29.63	32.05	34.22	36.33	38.53	40.98	43.97	48.36	52.19	55.66	59.89
38	20.69	22.87	24.88	27.34	30.53	32.99	35.19	37.33	39.56	42.04	45.07	49.51	53.38	56.89	61.16
39	21.42	23.65	25.69	28.19	31.44	33.93	36.16	38.33	40.59	43.10	46.17	50.65	54.57	58.12	62.42
40	22.16	24.43	26.50	29.05	32.34	34.87	37.13	39.33	41.62	44.16	47.26	51.80	55.75	59.34	63.69
41	22.90	25.21	27.32	29.90	33.25	35.81	38.10	40.33	42.65	45.22	48.36	52.94	56.94	60.56	64.95
42	23.65	25.99	28.14	30.76	34.15	36.75	39.07	41.33	43.67	46.28	49.45	54.09	58.12	61.77	66.20
43	24.39	26.78	28.96	31.62	35.06	37.69	40.04	42.33	44.70	47.33	50.54	55.23	59.30	62.99	67.45
44	25.14	27.57	29.78	32.48	35.97	38.64	41.02	43.33	45.73	48.39	51.63	56.36	60.48	64.20	68.70
45	25.90	28.36	30.61	33.35	36.88	39.58	41.99	44.33	46.76	49.45	52.72	57.50	61.65	65.41	69.95
46	26.65	29.16	31.43	34.21	37.79	40.52	42.96	45.33	47.78	50.50	53.81	58.64	62.82	66.61	71.20
47	27.41	29.95	32.26	35.08	38.70	41.47	43.94	46.33	48.81	51.56	54.90	59.77	64.00	67.82	72.44
48	28.17	30.75	33.09	35.94	39.62	42.42	44.91	47.33	49.84	52.61	55.99	60.90	65.17	69.02	73.68
49	28.94	31.55	33.93	36.81	40.53	43.36	45.88	48.33	50.86	53.66	57.07	62.03	66.33	70.22	74.91
50	29.70	32.35	34.76	37.68	41.44	44.31	46.86	49.33	51.89	54.72	58.16	63.16	67.50	71.42	76.15
60	37.48	40.48	43.18	46.45	50.64	53.80	56.62	59.33	62.13	65.22	68.97	74.39	79.08	83.29	88.37
70	45.44	48.75	51.73	55.32	59.89	63.34	66.39	69.33	72.35	75.68	79.71	85.52	90.53	95.02	100.42
80	53.54	57.15	60.39	64.27	69.20	72.91	76.18	79.33	82.56	86.11	90.40	96.57	101.87	106.62	112.32
90	61.75	65.64	69.12	73.29	78.55	82.51	85.99	89.33	92.76	96.52	101.05	107.56	113.14	118.13	124.11
100	70.06	74.22	77.92	82.35	87.94	92.12	95.80	99.33	102.94	106.90	111.66	118.49	124.34	129.56	135.80
200	156.43	162.72	168.27	174.83	183.00	189.04	194.31	199.33	204.43	209.98	216.60	226.02	233.99	241.05	249.44

Tableau 11.11 – Lois de Fisher-Snedecor

Soit F une variable dont la loi est celle de Fisher Snedecor, à ν_1 et ν_2 degrés de liberté, de densité ρ. La table donne, pour divers couples (ν_1, ν_2) les valeurs u pour lesquelles $\mathbb{P}(F < u) = G(u) = 0.95$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	12	24	$>$25
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.08	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
$>$120	3.84	2.99	2.60	2.37	2.21	2.10	1.94	1.75	1.52	1.00

Tableau 11.12 – Lois de Fisher-Snedecor

Soit F une variable dont la loi de Fisher Snedecor, à ν_1 et ν_2 degrés de liberté, de densité ρ. La table donne, pour divers couples (ν_1, ν_2) les valeurs u pour lesquelles $\mathbb{P}(F < u) = G(u) = 0.99$

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	>25
1	4052.	4999.	5403.	5625.	5764.	5859.	5982.	6106.	6234.	6366.
2	98.50	99.00	99.17	99.25	99.30	99.33	99.37	99.42	99.46	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.29	9.89	9.47	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	3.91
11	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
13	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.59	3.16
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.43	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.08	2.65
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
22	7.94	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
>120	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00

Tableau 11.13 – Valeurs critiques du test de Mann-Whitney

n_1 est l'effectif du plus petit des deux échantillons. Quand la valeur observée de U est égale ou inférieure à la valeur donnée dans la table, H_0 peut être rejeté au niveau de signification choisi.

Seuil de signification à 0.001 pour un test unilatéral
Seuil de signification à 0.002 pour un test bilatéral

n_1 \ n_2	9	10	11	12	13	14	15	16	17	18	19	20
1												
2												
3									0	0	0	0
4		0	0	0	1	1	1	2	2	3	3	3
5	1	1	2	2	3	3	4	5	5	6	7	7
6	2	3	4	4	5	6	7	8	9	10	11	12
7	3	5	6	7	8	9	10	11	13	14	15	16
8	5	6	8	9	11	12	14	15	17	18	20	21
9	7	8	10	12	14	15	17	19	21	23	25	26
10	8	10	12	14	17	19	21	23	25	27	29	32
11	10	12	15	17	20	22	24	27	29	32	34	37
12	12	14	17	20	23	25	28	31	34	37	40	42
13	14	17	20	23	26	29	32	35	38	42	45	48
14	15	19	22	25	29	32	36	39	43	46	50	54
15	17	21	24	28	32	36	40	43	47	51	55	59
16	19	23	27	31	35	39	43	48	52	56	60	65
17	21	25	29	34	38	43	47	52	57	61	66	70
18	23	27	32	37	42	46	51	56	61	66	71	76
19	25	29	34	40	45	50	55	60	66	71	77	82
20	26	32	37	42	48	54	59	65	70	76	82	88

Seuil de signification à 0.01 pour un test unilatéral
Seuil de signification à 0.02 pour un test bilatéral

n_1 \ n_2	9	10	11	12	13	14	15	16	17	18	19	20
1												
2					0	0	0	0	0	0	1	1
3	1	1	1	2	2	2	3	3	4	4	4	5
4	3	3	4	5	5	6	7	7	8	9	9	10
5	5	6	7	8	9	10	11	12	13	14	15	16
6	7	8	9	11	12	13	15	16	18	19	20	22
7	9	11	12	14	16	17	19	21	23	24	26	28
8	11	13	15	17	20	22	24	26	28	30	32	34
9	14	16	18	21	23	26	28	31	33	36	38	40
10	16	19	22	24	27	30	33	36	38	41	44	47
11	18	22	25	28	31	34	37	41	44	47	50	53
12	21	24	28	31	35	38	42	46	49	53	56	60
13	23	27	31	35	39	43	47	51	55	59	63	67
14	26	30	34	38	43	47	51	56	60	65	69	73
15	28	33	37	42	47	51	56	61	66	70	75	80
16	31	36	41	46	51	56	61	66	71	76	82	87
17	33	38	44	49	55	60	66	71	77	82	88	93
18	36	41	47	53	59	65	70	76	82	88	94	100
19	38	44	50	56	63	69	75	82	88	94	101	107
20	40	47	53	60	67	73	80	87	93	100	107	114

Seuil de signification à 0.025 pour un test unilatéral
Seuil de signification à 0.05 pour un test bilatéral

n_1 \ n_2	9	10	11	12	13	14	15	16	17	18	19	20
1												
2	0	0	0	1	1	1	1	1	2	2	2	2
3	2	3	3	4	4	5	5	6	6	7	7	8
4	4	5	6	7	8	9	10	11	11	12	13	13
5	7	8	9	11	12	13	14	15	17	18	19	20
6	10	11	13	14	16	17	19	21	22	24	25	27
7	12	14	16	18	20	22	24	26	28	30	32	34
8	15	17	19	22	24	26	29	31	34	36	38	41
9	17	20	23	26	28	31	34	37	39	42	45	48
10	20	23	26	29	33	36	39	42	45	48	52	55
11	23	26	30	33	37	40	44	47	51	55	58	62
12	26	29	33	37	41	45	49	53	57	61	65	69
13	28	33	37	41	45	50	54	59	63	67	72	76
14	31	36	40	45	50	55	59	64	67	74	78	83
15	34	39	44	49	54	59	64	70	75	80	85	90
16	37	42	47	53	59	64	70	75	81	86	92	98
17	39	45	51	57	63	67	75	81	87	93	99	105
18	42	48	55	61	67	74	80	86	93	99	106	112
19	45	52	58	65	72	78	85	92	99	106	113	119
20	48	55	62	69	76	83	90	98	105	112	119	127

Seuil de signification à 0.05 pour un test unilatéral
Seuil de signification à 0.1 pour un test bilatéral

n_1 \ n_2	9	10	11	12	13	14	15	16	17	18	19	20
1										0	0	
2	1	1	1	2	2	2	3	3	3	4	4	4
3	3	4	5	5	6	7	7	8	9	9	10	11
4	6	7	8	9	10	11	12	14	15	16	17	18
5	9	11	12	13	15	16	18	19	20	22	23	25
6	12	14	16	17	19	21	23	25	26	28	30	32
7	15	17	19	21	24	26	28	30	33	35	37	39
8	18	20	23	26	28	31	33	36	39	41	44	47
9	21	24	27	30	33	36	39	42	45	48	51	54
10	24	27	31	34	37	41	44	48	51	55	58	62
11	27	31	34	38	42	46	50	54	57	61	65	69
12	30	34	38	42	47	51	55	60	64	68	72	77
13	33	37	42	47	51	56	61	65	70	75	80	84
14	36	41	46	51	56	61	66	71	77	82	87	92
15	39	44	50	55	61	66	72	77	83	88	94	100
16	42	48	54	60	65	71	77	83	89	95	101	107
17	45	51	57	64	70	77	83	89	96	102	109	115
18	48	55	61	68	75	82	88	95	102	109	116	123
19	51	58	65	72	80	87	94	101	109	116	123	130
20	54	62	69	77	84	92	100	107	115	123	130	138

Tableau 11.14 – Valeur critique du test du nombre de paires au seuil de 5%

Valeurs inférieures

$n_1 \backslash n_2$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

Valeurs supérieures

$n_1 \backslash n_2$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2																			
3																			
4				9	9														
5			9	10	10	11	11												
6			9	10	11	12	12	13	13	13	13								
7				11	12	13	13	14	14	14	14	15	15	15					
8				11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17
9					13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
10					13	14	15	16	16	17	17	18	18	18	19	19	19	20	20
11					13	14	15	16	17	17	18	19	19	19	20	20	20	21	21
12					13	14	16	16	17	18	19	19	20	20	21	21	21	22	22
13						15	16	17	18	19	19	20	20	21	21	22	22	23	23
14						15	16	17	18	19	20	20	21	22	22	23	23	23	24
15						15	16	18	18	19	20	21	22	22	23	23	24	24	25
16							17	18	19	20	21	21	22	23	23	24	25	25	25
17							17	18	19	20	21	22	23	23	24	25	25	26	26
18							17	18	19	20	21	22	23	24	25	25	26	26	27
19							17	18	20	21	22	23	23	24	25	26	26	27	27
20							17	18	20	21	22	23	24	25	25	26	27	27	28

Tableau 11.15 – Probabilités associées aux valeurs aussi extrêmes que les valeurs de U observées du test de Mann-Whitney

La taille des deux échantillons ne peut être supérieure à 8.

$n_2 = 3$

U \ n_1	1	2	3
0	0.25	0.1	0.05
1	0.50	0.2	0.10
2	0.75	0.4	0.20
3		0.60	0.35
4			0.50
5			0.65

$n_2 = 4$

U \ n_1	1	2	3	4
0	0.200	0.067	0.028	0.014
1	0.400	0.133	0.057	0.029
2	0.600	0.267	0.114	0.057
3		0.400	0.200	0.100
4		0.600	0.314	0.171
5			0.429	0.243
6			0.571	0.343
7				0.443
8				0.557

$n_2 = 5$

U \ n_1	1	2	3	4	5
0	0.167	0.047	0.018	0.008	0.004
1	0.333	0.095	0.036	0.016	0.008
2	0.500	0.190	0.071	0.032	0.016
3	0.667	0.286	0.125	0.056	0.028
4		0.429	0.196	0.095	0.048
5		0.571	0.286	0.143	0.075
6			0.393	0.206	0.111
7			0.500	0.278	0.155
8			0.607	0.365	0.210
9				0.452	0.274
10				0.548	0.345
11					0.421
12					0.500
13					0.579

$n_2 = 6$

U \ n_1	1	2	3	4	5	6
0	0.143	0.036	0.012	0.005	0.002	0.001
1	0.286	0.071	0.024	0.010	0.004	0.002
2	0.428	0.143	0.048	0.019	0.009	0.004
3	0.571	0.214	0.083	0.033	0.015	0.008
4		0.321	0.131	0.057	0.026	0.013
5		0.429	0.190	0.086	0.041	0.021
6		0.571	0.274	0.129	0.063	0.032
7			0.357	0.176	0.089	0.047
8			0.452	0.238	0.123	0.066
9			0.548	0.305	0.165	0.090
10				0.381	0.214	0.120
11				0.457	0.268	0.155
12				0.545	0.331	0.197
13					0.396	0.242
14					0.465	0.294
15					0.535	0.350
16						0.409
17						0.469
18						0.531

$n_2 = 7$

U \ n_1	1	2	3	4	5	6	7
0	0.125	0.028	0.008	0.003	0.001	0.001	0.000
1	0.250	0.056	0.017	0.006	0.003	0.001	0.001
2	0.375	0.111	0.033	0.012	0.005	0.002	0.001
3	0.500	0.167	0.058	0.021	0.009	0.004	0.002
4	0.625	0.250	0.092	0.036	0.015	0.007	0.003
5		0.333	0.133	0.055	0.024	0.011	0.006
6		0.444	0.192	0.082	0.037	0.017	0.009
7		0.556	0.258	0.115	0.053	0.026	0.013
8			0.333	0.158	0.074	0.037	0.019
9			0.417	0.206	0.101	0.051	0.027
10			0.500	0.264	0.134	0.069	0.036
11			0.583	0.324	0.172	0.090	0.049
12				0.394	0.216	0.117	0.064
13				0.464	0.265	0.147	0.082
14				0.538	0.319	0.183	0.104
15					0.378	0.223	0.130
16					0.438	0.267	0.159
17					0.500	0.314	0.191
18					0.562	0.365	0.228
19						0.418	0.267
20						0.473	0.310
21						0.527	0.355
22							0.402
23							0.451
24							0.500
25							0.549

$n_2 = 8$

U \ n_1	1	2	3	4	5	6	7	8
0	0.111	0.022	0.006	0.002	0.001	0.000	0.000	0.000
1	0.222	0.044	0.012	0.004	0.002	0.001	0.000	0.000
2	0.333	0.089	0.024	0.008	0.003	0.001	0.001	0.000
3	0.444	0.133	0.042	0.014	0.005	0.002	0.001	0.001
4	0.556	0.200	0.067	0.024	0.009	0.004	0.002	0.001
5		0.267	0.097	0.036	0.015	0.006	0.003	0.001
6		0.356	0.139	0.055	0.023	0.010	0.005	0.002
7		0.444	0.188	0.077	0.033	0.015	0.007	0.003
8		0.556	0.248	0.107	0.047	0.021	0.010	0.005
9			0.315	0.141	0.064	0.030	0.014	0.007
10			0.387	0.184	0.085	0.041	0.020	0.010
11			0.461	0.230	0.111	0.054	0.027	0.014
12			0.539	0.285	0.142	0.071	0.036	0.019
13				0.341	0.177	0.091	0.047	0.025
14				0.404	0.217	0.114	0.060	0.032
15				0.467	0.262	0.141	0.076	0.041
16				0.533	0.311	0.172	0.095	0.052
17					0.362	0.207	0.116	0.065
18					0.416	0.245	0.140	0.080
19					0.472	0.286	0.168	0.097
20					0.528	0.331	0.198	0.117
21						0.377	0.232	0.139
22						0.426	0.268	0.164
23						0.475	0.306	0.191
24						0.525	0.347	0.221
25							0.389	0.253
26							0.433	0.870
27							0.478	0.323
28							0.522	0.360
29								0.399
30								0.439
31								0.480
32								0.520

Tableau 11.16 – Valeurs critiques du test des rangs pour échantillons appariés, de Wilcoxon

	α (test unilatéral)			
	0.05	0.025	0.01	0.005
	α (test bilatéral)			
n	0.1	0.05	0.02	0.01
6	2	0		
7	3	2	0	
8	5	4	2	0
9	8	6	3	2
10	10	8	5	3
11	13	11	7	5
12	17	14	10	7
13	21	17	13	10
14	25	21	16	13
15	30	25	20	16
16	35	30	24	20
17	41	35	28	23
18	47	40	33	28
19	53	46	38	32
20	60	52	43	38
21	67	59	49	43
22	75	66	56	49
23	83	73	62	55
24	91	81	69	61
25	100	89	77	68

Annexe A

Fonction de plusieurs variables

1 Introduction

Définition A.1 *On appelle fonction numérique de n variables réelles une application f d'une partie D de $\mathbb{R}^n$ dans $\mathbb{R}$. On dit que f est définie sur le domaine D.*

On note :

$$f : \begin{cases} D & \rightarrow \quad \mathbb{R} \\ (x_1, x_2, \ldots, x_n) & \mapsto \quad f(x_1, x_2, \ldots, x_n) \end{cases}$$

ou plus simplement $f : (x_1, x_2, \ldots, x_n) \mapsto f(x_1, x_2, \ldots, x_n)$ s'il n'y a pas d'ambiguïté sur D.
On note également :

$$f : \begin{cases} D & \rightarrow \quad \mathbb{R} \\ H & \mapsto \quad f(H) \end{cases}$$

où $H = (x_1, x_2, \ldots, x_n)$ est un "point" de $\mathbb{R}^n$ appartenant à D.

Exemple A.1

- *la fonction $f : (x, y) \mapsto ax^2 + bxy + cy^2$ où a, b et c sont des réels donnés, est une fonction numérique de deux variables réelles, définie sur $\mathbb{R}^2$.*
- *la fonction $f : (x, y) \mapsto \sqrt{R^2 - (x^2 + y^2)}$ est une fonction numérique de deux variables réelles définie pour tout couple (x, y) tel que $(x^2 + y^2) \leq R^2$, R étant un réel positif donné.*
- *la fonction $f : (x, y) \mapsto \frac{x^2 - y^2}{x^2 + y^2}$ est une fonction numérique de deux variables réelles définie sur $D = \mathbb{R}^2 \setminus \{(0, 0)\}$.*
- *la fonction $f : (x, y) \mapsto \frac{x^2\sqrt{y-2}}{z-3}$ est une fonction numérique de trois variables réelles définie pour tout triplet (x, y, z) tel que : $y \geq 2$, $z \neq 3$.*

1.1 Représentation graphique

Dans le cas particulier d'une fonction de deux variables, on peut utiliser des représentations géométriques.

Si on choisit, par exemple, un trièdre orthonormé direct, un couple (x, y) de D est représenté par un point H du plan xOy, auquel on peut associer les $M(x, y, f(x, y))$.
Si D est représenté par une surface plane, alors l'ensemble des points $M(x, y, f(x, y))$ constitue également une surface.

Exemple A.2 *Soit $f : (x, y) \mapsto \sqrt{R^2 - (x^2 + y^2)}$. Le domaine de définition D défini par $(x^2 + y^2) \leq R^2$ est représenté par l'ensemble des points du disque circulaire de centre O et de rayon R.*
$z = \sqrt{R^2 - (x^2 + y^2)} \Leftrightarrow (x^2+y^2+z^2 = R^2$ et $z \geq 0)$. L'ensemble des points $M(x, y, z)$ forme une demi-sphère de centre O et de rayon R.

Dans le cas particulier d'une fonction de trois variables, il est encore possible de représenter le domaine de définition D par un ensemble de points de l'espace réel.
Pour une fonction de n variables, avec $n > 3$, on ne peut évidemment plus donner de représentation globale.

1.2 Limite

La notion de limite d'une fonction de plusieurs variables s'introduit comme pour les fonctions d'une variable. On a, par exemple, pour une fonction de deux variables (la généralisation est aisée) :

Définition A.2 *On dit que $f(x, y) = f(H)$ tend vers une limite L lorsque (x, y) tend (x_0, y_0) ou lorsque $H(x, y)$ tend $H(x_0, y_0)$, si :*

$$\exists \epsilon > 0, \exists \alpha > 0 \text{ tel que } (|x - x_0| < \alpha \text{ et } |y - y_0| < \alpha \text{ avec } (x, y) \neq (x_0, y_0)) \Rightarrow |f(H) - L| < \epsilon$$

On écrit :

$$\lim_{H \to H_0} f(H) = \lim_{(x,y) \to (x_0,y_0)} f(x, y) = L$$

On notera :
- que la définition implique que la fonction f est définie sur un pavé ouvert entourant H_0 (un voisinage de H_0).
- que la définition entraîne l'unicité de la limite ; lorsqu'elle existe ;
- qu'on peut écrire un définition équivalente :

$$\exists \epsilon > 0, \exists \beta > 0 \text{ tel que } ||HH_0|| < \beta \text{ avec } H \neq H_0 \Rightarrow |f(H) - L| < \epsilon$$

- que les théorèmes, concernant la somme, le produit, et le quotient de deux fonctions admettant une limite lorsque H tend vers H_0, s'énoncent et s'établissent comme dans le cas des fonctions d'une variable ;
- qu'il est aisé de définir une limite infinie ainsi qu'une limite de $f(H)$ quand H s'éloigne indéfiniment.

Exemple A.3 *Soit $f : (x, y) \mapsto (x + y) \sin \frac{1}{x^2+y^2}$*
f n'est pas définie en $(0, 0)$ mais on a $\lim_{(x,y) \to (0,0)} f(x, y) = 0$. En effet $(x + y) \to 0$ et $\sin \frac{1}{x^2+y^2}$ demeure borné.

Exemple A.4 *Soit* $f:(x,y)\mapsto \frac{x^2\sqrt{y-2}}{z-3}$.
On considère $H_0(1,3,3)$ *où* f *n'est pas définie. On a :*

$$\lim_{H\to H_0} f(H) = +\infty \quad \textit{si} \quad z\to z+0$$
$$\lim_{H\to H_0} f(H) = -\infty \quad \textit{si} \quad z\to z-0$$

Il convient de préciser si H *tend vers* H_0 *avec une côte supérieure ou une côte inférieure.*

1.3 Continuité

La notion de continuité découle naturellement de la notion de limite. On aura en particulier pour une fonction de deux variables :

Définition A.3 *On dit qu'une fonction* $f:(x,y)\mapsto f(x,y)$ *est continue au point* (x_0,y_0) *si :*

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = f(x_0,y_0)$$

Une fonction continue en tout point d'un domaine est dite continue sur ce domaine.

Théorème A.1 *Si* f *et* g *sont deux fonctions continues au point* M_0*, il en est de même pour des fonctions* $f+g$*,* αf *(*$\alpha\in\mathbb{R}$*),* $f\times g$ *et si* $g(M_0)\neq 0$*,* f/g*.*

Il s'agit là d'une conséquence des théorèmes sur les limites.

Exemple A.5 *Soit la fonction* $f:(x,y)\mapsto 3x-y+1$.
Cette fonction est définie sur $\mathbb{R}^2$*. Elle est continue sur* $\mathbb{R}^2$*, ce que l'on peut démontrer en s'appuyant sur la définition de la continuité. Quel que soit* (x_0,y_0)*, on a :*

$$\forall\epsilon>0, \exists\alpha=\frac{\epsilon}{4}>0 \textit{ tel que : } (|x-x_0|<\alpha, |y-y_0|<\alpha \textit{ et } (x,y)\neq(x_0,y_0)) \Rightarrow |f(x,y)-f(x_0,y_0)|<\epsilon$$

En effet $|x-x_0|<\frac{\epsilon}{4}\Rightarrow 3|x-x_0|<\frac{3\epsilon}{4}$
Or $|f(x,y)-f(x_0,y_0)| = |3(x-x_0)-(y-y_0)| \leq 3|x-x_0|+|y-y_0|$.
Donc $|x-x_0|<\frac{\epsilon}{4}$ *et* $|y-y_0|<\frac{\epsilon}{4}\Rightarrow |f(x,y)-f(x_0,y_0)|<\epsilon$

2 Fonctions composées

On peut envisager plusieurs types de composition

2.1 Fonction de fonction

Soit u une fonction de variables $x_1,x_2,\ldots,x_n$. Soit f une fonction d'une variable. On peut définir la fonction :

$$\phi:(x_1,x_2,\ldots,x_n)\mapsto f\left(u(x_1,x_2,\ldots,x_n)\right)$$

sous réserve, bien entendu, que $u(x_1,x_2,\ldots,x_n)$ appartienne au domaine de définition de f.

Exemple A.6 *La fonction $\phi : (x,y,z) \mapsto \sqrt{x^2+y^2+z^2}$ est obtenue par composition de la fonction $u : (x,y,z) \mapsto x^2+y^2+z^2$ par la fonction $f : u \mapsto \sqrt{u}$.*

Théorème A.2 *Si u est une fonction, de deux variables, continue au point (x_0,y_0) et si f est une fonction, d'une variable, continue au point (x_0,y_0), la fonction $\phi : (x,y) \mapsto f(u(x,y))$ est continue au point (x_0,y_0).*
On généralise aisément au cas où u est une fonction de n variables.

En effet puisque f est continue au point $u_0 = u(x_0,y_0)$, on peut trouver $\beta > 0$ tel que : $|u-u_0| < \beta \Rightarrow |f(u)-f(u_0)| < \epsilon$
Mais u est continue au point (x_0,y_0) ; il est donc possible de choisir $\alpha > 0$ tel que : $(|x-x_0| < \alpha$ et $|y-y_0| < \alpha) \Rightarrow |u(x,y)-u(x_0,y_0)| < \beta$. Donc il est possible d choisir $\alpha > 0$ tel que :

$$(|x-x_0| < \alpha \text{ et } |y-y_0| < \alpha) \Rightarrow |u(x,y)-u(x_0,y_0)| < \beta \Rightarrow |f(u(x,y))-f(u(x_0,y_0))| < \epsilon$$

Exemple A.7 *La fonction $\phi : (x,y,z) \mapsto \sqrt{x^2+y^2+z^2}$ est continue dans $\mathbb{R}^3$*

2.2 Fonction composée

Soit $\phi : (u,v,w) \mapsto \phi(u,v,w)$ une fonction de 3 variables. Soient $f : x \mapsto u = f(x)$, $g : x \mapsto v = g(x)$, $h : x \mapsto h(x)$, trois fonctions d'une variable. On peut donc définit la fonction :

$$F : x \mapsto \phi(f(x), g(x), h(x))$$

sous réserve que le point $(f(x), g(x), h(x))$ appartienne au domaine de définition de ϕ. Une telle fonction est dite fonction composée de f, g et h par ϕ. On notera que F est fonction d'une variable. On généralise aisément au cas où F est fonction de n variables.

Exemple A.8 *La fonction*

$$F :\mapsto \frac{\cos^2 x \times e^{x^2}}{x^2}$$

est une fonction composée. Elle est obtenue en composant les fonctions : $x \mapsto u = \cos^2 x$, $x \mapsto v = e^{x^2}$, $x \mapsto w = x^2$ par la fonction $(u,v,w) \mapsto \frac{uv}{w}$

Théorème A.3 *Si f, g et h sont des fonctions, d'une variable, continues en x_0 et si ϕ est une fonction, de trois variables, continue au point $u_0 = f(x_0)$, $v_0 = g(x_0)$, $w_0 = h(x_0)$, la fonction $F : x \mapsto \phi(f(x), g(x), h(x))$ est continues en x_0.*
On généralise aisément au cas où ϕ est une fonction de n variables.

Ce théorème se démontre aisément, en utilisant le même schéma que pour la démonstration du théorème A.2.

Exemple A.9 *La fonction*

$$F :\mapsto \frac{\cos^2 x \times e^{x^2}}{x^2}$$

est continue sur $\mathbb{R}^*$.

2.3 Autre cas

Soient les fonctions $u : (x, y) \mapsto u(x, y)$, $v : (x, y) \mapsto v(x, y)$, $w : (x, y) \mapsto w(x, y)$ et soit la fonction $f : (u, v, w) \mapsto f(u, v, w)$. On peut alors définir la fonction :

$$F : (x, y) \mapsto f(u(x, y), v(x, y), w(x, y))$$

sous réserve que le point $(u(x, y), v(x, y), w(x, y))$ appartienne au domaine de définition de f.
On généralise aisément à la composition de n fonctions de p variables par une fonction de n variables.

Exemple A.10 *La fonction*

$$F \mapsto \frac{x^2 - y^2}{x^2 + y^2}$$

est obtenue en composant les fonctions $u(x, y) \mapsto u = x^2 - y^2$ *et* $(x, y) \mapsto v = x^2 + y^2$ *par la fonction* $(u, v) \mapsto \frac{u}{v}$

Théorème A.4 *Si les fonctions, de deux variables,* u, v *et* w *sont continues au point* (x_0, y_0) *et si la fonction de trois variables, est continue au point* $(u(x_0, y_0), v(x_0, y_0), w(x_0, y_0))$, *la fonction*

$$F : (x, y) \mapsto f(u(x, y), v(x, y), w(x, y))$$

est continue au point (x_0, y_0).

Ce théorème se démontre aisément, en utilisant le même schéma que pour la démonstration du théorème A.2.

Exemple A.11 *La fonction*

$$F \mapsto \frac{x^2 - y^2}{x^2 + y^2}$$

est continue sur $\mathbb{R}^2 \setminus \{(0, 0)\}$

3 Dérivées partielles

Définition A.4 *Soit la fonction de deux variables* $f : (x, y) \mapsto f(x, y)$.
On appelle dérivée partielle de f, *par rapport à la variable* x *et au point* (x_0, y_0), *le nombre s'il existe :*

$$f'_x(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

On appelle de même dérivée de f, *par rapport à la variable* y *et au point* (x_0, y_0), *le nombre s'il existe :*

$$f'_y(x_0, y_0) = \frac{\partial f}{\partial y}(x_0, y_0) = \lim_{k \to 0} \frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k}$$

Si en tout point d'un domaine D il existe des dérivées partielles, les fonctions $\frac{\partial f}{\partial x}$: $(x, y) \mapsto \frac{\partial f}{\partial x}(x, y)$ et $\frac{\partial f}{\partial y}$: $(x, y) \mapsto \frac{\partial f}{\partial y}(x, y)$ sont appelées dérivées partielles premières de f, par rapport à x et par rapport à y.

On généralise aisément cette définition à une fonction de n variables.
On remarque que la dérivée partielle par rapport à x, au point (x_0, y_0), n'est autre que la dérivée, au point x_0, de la fonction $\phi : x \mapsto f(x, y_0)$. Pour la dérivée partielle par rapport à y, il s'agit de la dérivée, au point y_0, de la fonction $\psi : y \mapsto f(x_0, y)$.
On en déduit la règle pratique : pour déterminer une dérivée partielle, il suffit de dériver comme on en a l'habitude par rapport à la variable considérée, les autres variables étant considérées comme des constantes.

Exemple A.12 *La fonction $f : (x, y) \mapsto 5x^2 - xy + 2y^2$ est définie et continue sur $\mathbb{R}^2$. Sur $\mathbb{R}^2$, elle admet pour dérivées partielles :*

$$\frac{\partial f}{\partial x} : (x, y) \mapsto 10x - y \textit{ et } \frac{\partial f}{\partial y} : (x, y) \mapsto -x + 4y$$

Exemple A.13 *La fonction $f : (x, y) \mapsto x^2 \sin y$ est définie et continue sur $\mathbb{R}^2$. Sur $\mathbb{R}^2$, elle admet pour dérivées partielles :*

$$\frac{\partial f}{\partial x} : (x, y) \mapsto 2x \sin y \textit{ et } \frac{\partial f}{\partial y} : (x, y) \mapsto x^2 \cos y$$

3.1 *Dérivées partielles secondes*

Les fonctions dérivées partielles premières, d'une fonction f de plusieurs variables, peuvent elles-mêmes admettre des dérivées partielles ; on les appelle dérivées partielles secondes de f.
On peut évidemment envisager également des dérivées partielles d'ordre supérieur.
Pour une fonction de deux variables $f : (x, y) \mapsto f(x, y)$, il faut a priori envisager quatre dérivées partielles secondes :

$$\begin{array}{ccc}\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = f''_{x^2} & , & \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x \partial y} = f''_{xy} \\ \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y \partial x} = f''_{yx} & , & \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2} = f''_{y^2}\end{array}$$

Exemple A.14 *La fonction $f : (x, y) \mapsto 5x^2 - xy + 2y^2$admet, sur $\mathbb{R}^2$, les dérivées partielles secondes :*

$$\begin{array}{ccc}\frac{\partial^2 f}{\partial x^2}(x, y) \mapsto 10 & ; & \frac{\partial^2 f}{\partial x \partial y}(x, y) \mapsto -1 \\ \frac{\partial^2 f}{\partial y \partial x}(x, y) \mapsto -1 & ; & \frac{\partial^2 f}{\partial y^2}(x, y) \mapsto 4\end{array}$$

Exemple A.15 *La fonction* $f : (x, y) \mapsto x^2 \sin y$ *admet, sur* $\mathbb{R}^2$*, les dérivées partielles secondes :*

$$\begin{array}{rcl}\frac{\partial^2 f}{\partial x^2}(x,y) \mapsto 2\sin y & ; & \frac{\partial^2 f}{\partial x \partial y}(x,y) \mapsto 2x\cos y \\ \frac{\partial^2 f}{\partial y \partial x}(x,y) \mapsto 2x\cos y & ; & \frac{\partial^2 f}{\partial y^2}(x,y) \mapsto -x^2\sin y\end{array}$$

On remarque, sur ces deux exemples que l'on a : $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$

Théorème A.5 *Si, dans le voisinage du point* (x_0, y_0)*, la fonction* $f : (x, y) \mapsto f(x, y)$ *admet des dérivées partielles premières continues et si les dérivées partielles secondes* f''_{xy} *et* f''_{yx} *existent et sont continues, on a :*

$$f''_{xy}(x_0, y_0) = f''_{yx}(x_0, y_0)$$

On admettra ce théorème qui se généralise aisément et permet de considérer les dérivations successives par rapport aux variables dans un ordre quelconque.
Sous réserve de la continuité des dérivées partielles, la fonction $f : (x, y) \mapsto f(x, y)$ n'admet que trois dérivées partielles secondes $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial x \partial y}$ et $\frac{\partial^2 f}{\partial y^2}$ et quatre dérivées partielles troisièmes : $\frac{\partial^3 f}{\partial x^3}$, $\frac{\partial^3 f}{\partial x^2 \partial y}$, $\frac{\partial^3 f}{\partial x \partial y^2}$ et $\frac{\partial^3 f}{\partial y^3}$

Exemple A.16 *Reprenons la fonction* $f : (x, y) \mapsto x^2 \sin y$*. On peut facilement vérifier l'égalité des dérivées partielles :*

$$\begin{array}{rcl}\frac{\partial^3 f}{\partial x \partial y \partial y} & : & (x,y) \mapsto -2x\sin y \\ \frac{\partial^3 f}{\partial y \partial x \partial y} & : & (x,y) \mapsto -2x\sin y \\ \frac{\partial^3 f}{\partial y \partial y \partial x} & : & (x,y) \mapsto -2x\sin y\end{array}$$

On note que $\frac{\partial^3 f}{\partial x \partial y^2} : (x, y) \mapsto -2x \sin y$.

Théorème A.6 *Soient les fonctions* $u : x \mapsto u(x)$ *et* $v : x \mapsto v(x)$ *dérivables en* x_0 *; on pose* $u_0 = u(x_0)$ *et* $v_0 = v(x_0)$*. Soit la fonction* $\phi : (u, v) \mapsto \phi(u, v)$ *admettant au voisinage du point* (u_0, v_0) *des dérivées partielles premières continues en* (u_0, v_0)*. La fonction composée* $F : x \mapsto \phi[u(x), v(x)]$ *admet pour dérivée en* x_0 *:*

$$F' : \phi'_u [u(x_0), v(x_0)] \, u'(x_0) + \phi'_v [u(x_0), v(x_0)] \, v'(x_0)$$

On considère les valeurs x_0 et $x_0 + \Delta x$ de la variable x. À l'accroissement Δx de la variable correspondent les accroissements :

$$\begin{array}{rcl}\Delta u & = & u(x_0 + \Delta x) - u(x_0) \\ \Delta v & = & v(x_0 + \Delta x) - v(x_0) \\ \Delta F & = & F(x_0 + \Delta x) - F(x_0)\end{array}$$

On a

$$\begin{aligned} F(x_0+\Delta x)-F(x_0) &= \phi\left[u(x_0+\Delta x), v(x_0+\Delta x)\right]-\phi\left[u(x_0), v(x_0)\right] \\ &= \phi\left(u_0+\Delta u, v_0+\Delta v\right)-\phi(u_0, v_0) \end{aligned}$$

On peut écrire :

$$F(x_0+\Delta x)-F(x_0)=\phi\left(u_0+\Delta u, v_0+\Delta v\right)-\phi(u_0, v_0+\Delta v)+\phi(u_0, v_0+\Delta v)-\phi(u_0, v_0)$$

On applique le théorème des accroissements finis à la fonction : $u \mapsto \phi(u, v_0+\Delta v)$, qui est dérivable :

$$\phi\left(u_0+\Delta u, v_0+\Delta v\right)-\phi(u_0, v_0+\Delta v)=\Delta u \phi'_u\left(u_0+\theta_1 \Delta u, v_0+\Delta v\right) \quad 0<\theta_1<1$$

On applique le théorème des accroissements finis à la fonction : $v \mapsto \phi(u_0+\Delta u, v)$, qui est dérivable :

$$\phi\left(u_0, v_0+\Delta v\right)-\phi(u_0, v_0)=\Delta v \phi'_v\left(u_0, v_0+\theta_2 \Delta v\right) \quad 0<\theta_2<1$$

Il en résulte :

$$F(x_0+\Delta x)-F(x_0)=\Delta u \phi'_u\left(u_0+\theta_1 \Delta u, v_0+\Delta v\right)+\Delta v \phi'_v\left(u_0, v_0+\theta_2 \Delta v\right)$$

Et donc :

$$\frac{F(x_0+\Delta x)-F(x_0)}{\Delta x}=\frac{\Delta u}{\Delta x} \phi'_u\left(u_0+\theta_1 \Delta u, v_0+\Delta v\right)+\frac{\Delta v}{\Delta x} \phi'_v\left(u_0, v_0+\theta_2 \Delta v\right)$$

Lorsqu'on fait tendre Δx vers zéro, étant données la dérivabilité de u et v et la continuité de ϕ'_u et ϕ'_v, on obtient comme limite du second membre : $u'(x_0)\phi'_u(u_0, v_0)+v'(x_0)\phi'_v(u_0, v_0)$. Il en résulte que F est dérivable en x_0 et que :

$$F'(x_0)=u'(x_0)\phi'_u(u_0, v_0)+v'(x_0)\phi'_v(u_0, v_0)$$

Lorsque u et v sont dérivables sur un intervalle $[a, b]$ et ϕ'_u et ϕ'_v sont continues sur le domaine correspondant à $[a, b]$ on obtiendra comme fonction dérivée de F :

$$F'=u'\phi'_u+v'\phi'_v$$

Exemple A.17 *Soit* $F: x \mapsto \frac{u(x)}{v(x)}$.
On obtient : $F'=\frac{1}{v}u'-\frac{u}{v^2}v'=\frac{vu'-uv'}{v^2}$

Exemple A.18 *Soit* $F: x \mapsto \log[u(x)v(x)]$.
On obtient : $F'=\frac{1}{u}u'+\frac{1}{v}v'=\frac{u'}{u}+\frac{v'}{v}$

Exemple A.19 *Soit* $F: x \mapsto u(x)v(x)$.
On obtient : $F'=vu'+uv'$

Quelques remarques :

1. Le théorème A.6 se généralise aisément au cas où ϕ est une fonction de n variables. si, par exemple, on a $F[u(x), v(x), w(x)]$, on obtient :

$$F' = u'\phi'_u + v'\phi'_v + w'\phi'_w$$

   c'est-à-dire que l'on a :

$$F'(x) = u'(x)\phi'_u[u(x), v(x), w(x)] + v'(x)\phi'_v[u(x), v(x), w(x)] + w'(x)\phi'_w[u(x), v(x), w(x)]$$

2. Si on considère la composition de fonctions de plusieurs variables par une autre fonction, le théorème A.6 s'applique également aux dérivées partielles, sous réserve des conditions de validité.
   Soit par exemple $F : (x, y) \mapsto \phi[u(x, y), v(x, y)]$, on aura :

$$F : \begin{cases} \frac{\partial F}{\partial x} & = & \phi'_u \frac{\partial u}{\partial x} + \phi'_v \frac{\partial u}{\partial x} \\ \frac{\partial F}{\partial y} & = & \phi'_u \frac{\partial u}{\partial y} + \phi'_v \frac{\partial u}{\partial y} \end{cases}$$

si les fonctions u et v admettent des dérivées partielles en (x, y) et si ϕ admet des dérivées partielles au voisinage de $(u(x, y), v(x, y))$ continues en ce point.

Exemple A.20 *La fonction $F : x \mapsto \frac{\cos^2 x e^{x^2}}{x^2}$ admet pour dérivées, sur $\mathbb{R}^*$:*

$$F' : x \mapsto \frac{e^{x^2}}{x^2}(-2\cos x \sin x) + \frac{\cos^2 x}{x^2}\left(2xe^{x^2}\right) - \frac{\cos^2 x e^{x^2}}{x^4} 2x$$

Exemple A.21 *La fonction $F : (x, y) \mapsto \frac{x^2 - y^2}{x^2 + y^2}$ admet pour dérivées partielles, sur $\mathbb{R}^*$:*

$$F : \begin{cases} \frac{\partial F}{\partial x}(x, y) \mapsto \frac{1}{x^2+y^2} 2x - \frac{x^2-y^2}{(x^2+y^2)^2} 2x & = & \frac{4xy^2}{(x^2+y^2)^2} \\ \frac{\partial F}{\partial y}(x, y) \mapsto \frac{1}{x^2+y^2}(-2y) - \frac{x^2-y^2}{(x^2+y^2)^2} 2y & = & \frac{-4x^2y}{(x^2+y^2)^2} \end{cases}$$

4 Exercices

1. Déterminer les domaines de définition des fonctions f et g définies par :
   - $f(x,y) = \sqrt{25 - x^2 - y^2}$
   - $g(x,y) = \frac{x}{\sqrt{x^2-y^2}}$ si $x^2 - y^2 \neq 0$, et $g(0,0) = 0$
2. On désigne par :
   - f la fonction à valeurs réelles définie sur $\mathbb{R}^2$ par :

   $$f(x,y) = \begin{cases} \frac{1}{y} & \text{si } 0 \leq x < y \\ 0 & \text{ailleurs} \end{cases}$$

   - g la fonction réelle définie sur $\mathbb{R}$ par :

   $$g(y) = \begin{cases} ye^{-y} & \text{si } y \geq 0 \\ 0 & \text{si } y < 0 \end{cases}$$

   - h la fonction réelle définie sur $\mathbb{R}^2$ par :

   $$h(x,y) = g(y)f(x,y)$$

   (a) Définir D l'ensemble de $\mathbb{R}^2$ sur lequel la fonction h prend des valeurs non nulles et donner une représentation géométrique de cet ensemble.

   (b) Calculer l'intégrale $f(x) = \int_0^{+\infty} h(x,y)\mathrm{d}y$. (on prendra soin d'utiliser la question précédente).

   (c) Donner, en fonction de x et y, l'expression explicite de :

   $$h(x,y) = \begin{cases} \frac{h(x,y)}{f(x)} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

   Pour quelles valeurs des variables x et y cette quantité est-elle non nulle ?

   (d) Calculer l'intégrale suivante dépendant du paramètre a :

   $$Q(a,x) = \int_0^{+\infty} (y-a)^2 h(x,y)\mathrm{d}y$$

   (e) Donner en fonction de x la valeur a_0 du paramètre a qui rend minimum l'intégrale $Q(a,x)$.
3. Déterminer et représenter graphiquement le domaine de définition des fonctions f et g définies par :
   - $f(x,y) = \log[(16x^2 - y^2)(x^2 + y^2 - 4)]$
   - $g(x,y) = \sqrt{6 - (2x + 3y)}$
4. Déterminer si les fonction suivantes ont une limite $(x,y) \mapsto (0,0)$:
   - $f : (x,y) \mapsto \frac{x^2-y^2}{x^2+y^2}$
   - $g : (x,y) \mapsto \frac{|x+y|}{x^2+y^2}$

- $h : (x,y) \mapsto \frac{x^3y^3}{x^2+y^2}$

Ces fonctions, complétées par : $(0,0) \mapsto 0$, sont-elles continues à l'origine ?

5. Peut-on rendre continues à l'origine, en les définissant en ce point, les fonctions définies par :
   - $f(x,y) = \frac{\sin(x+y)}{x+y}$
   - $g(x,y) = \frac{xy}{x^2+y^2}$
6. Déterminer et représenter graphiquement le domaine de définition de la fonction :

$$f : (x,y) \mapsto \frac{\sin x - \sin y}{x - y}$$

   Trouver une fonction g qui prenne la même valeur que f partout où celle-ci est définie et qui soit continue dans tout le plan.
7. Montrer que la fonction $(x,y) \mapsto \frac{x+y}{x^2+y^2}$ admet une limite lorsque le point $M(x,y)$ s'éloigne à l'infini.
8. La fonction f, définie par $f(x,y) = e^{x-y}$ a-t-elle une limite lorsque le point $M(x,y)$ s'éloigne à l'infini.
9. Soit la fonction f :

$$f : \begin{cases} (x,y) \mapsto x^2 + 2y & , \quad \text{si } (x,y) \neq (1,2) \\ (x,y) \mapsto 0 & , \quad \text{si } (x,y) = (1,2) \end{cases}$$

   Cette fonction est-elle continue au point $(1,2)$?
10. Calculer les dérivées partielles des fonction suivantes :
    - $f_1(x,y) \mapsto 2x^2 - 3xy + 4y^2$
    - $f_2(x,y) \mapsto \frac{x^2}{y} + \frac{y^2}{x}$
    - $f_3(x,y) \mapsto \sin(2x - 3y)$
    - $f_4(x,y) \mapsto e^{x^2+xy}$
11. Calculer les dérivées partielles secondes des fonction suivantes :
    - $f_1(x,y) \mapsto x^2 + 3xy + y^2$
    - $f_2(x,y) \mapsto x \cos y - y \cos x$
    - $f_3(x,y) \mapsto xy$
    - $f_4(x,y) \mapsto \log(xy)$
    - $f_5(x,y) \mapsto \arctan \frac{y}{x}$
    - $f_6(x,y) \mapsto \log \frac{1}{\sqrt{x^2+y^2}}$
12. On considère la fonction f :

$$f : \begin{cases} (x,y) \mapsto \frac{2xy}{x^2+y^2} & , \quad \text{si } (x,y) \neq (0,0) \\ (x,y) \mapsto 0 & , \quad \text{si } (x,y) = (0,0) \end{cases}$$

    Montrer que les dérivées partielles $f'_x(0,0)$ et $f'_y(0,0)$ existent. La fonction est-elle continue au point $(0,0)$?

13. Montrer que la fonction définie par $U(x,y,z) = (x^2+y^2+y^2)^{-1/2}$ vérifie l'équation aux dérivées partielles de Laplace :

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = 0$$

14. Soit la fonction $f : (x,y,z) \mapsto e^{-x/y} + e^{-y/z} + e^{-z/x}$. Déterminer le domaine de définition de D de f ; la fonction est-elle continue en D ? Calculer les dérivées partielles, premières et secondes, de f.
15. Soit la fonction $f : (x,y) \mapsto x^2 + y^2$. Calculer, en appliquant le théorème A.6, la dérivée de $F : t \mapsto f(a\cos t, b\sin t)$.
16. Soit la fonction $F : r \mapsto F(r)$. On pose $r = \sqrt{x^2+y^2+z^2}$, ce qui permet de définir la fonction $f : (x,y,z) \mapsto f(\sqrt{x^2+y^2+z^2})$. Calculer $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$.
Déterminer F pour que l'on ait : $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 0$.

Annexe B

Intégrales multiples

1 Intégrale double

1.1 Notion d'intégrale double

Soit une fonction $f : (x, y) \mapsto z = f(x, y)$, définie sur un domaine D fermé et supposée nulle à l'extérieur de D.
L'espace étant muni d'un repère orthormé direct $(O, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$, lorsque le point $H(x, y)$ se déplace dans D, le point $M(x, y, z)$ décrit une surface S.
On désigne par a et b les bornes inférieures et supérieures des abcisses d'un point de contour de D et par c et d les bornes inférieures et supérieures des ordonnées d'un point de ce même contour.
On partage $[a, b]$ par les valeurs $x_0 = a < x_1 < x_2 < \ldots < x_i < \ldots < x_m = b$
et $[c, d]$ par les valeurs $y_0 = c < y_1 < y_2 < \ldots < y_j < \ldots < y_m = d$
ce qui permet de définir un ensemble de mn rectangles recouvrant D.
Soit E_{ij} le rectangle $[x_{i-1}, x_i] \times [y_{j-1}, y_j]$, de dimensions $\Delta x_i = x_i - x_{i-1}$ et $\Delta y_j = y_j - y_{j-1}$.
On choisit arbitrairement dans E_{ij} un point H_{ij}.
On considère la somme $\sum_{i=1}^{n} \sum_{j=1}^{m} f(H_{ij}) \Delta x_i \Delta y_j$.
Si cette somme tend vers une limite I, lorsque m et n tendent vers l'infini de manière à ce que $\sup \Delta x_i$ et $\sup \Delta y_j$ tendent vers zéro, indépendamment de la façon dont on fait les partages et dont on choisit les points H_{ij}, on dit que f est intégrable sur D, que I est l'intégrale de f sur d et on pose :

$$I = \int\int_D f(x, y) \mathrm{d}x \mathrm{d}y$$

On peut démontrer en particulier que si f est continue sur D, l'intégrale existe toujours.

1.2 Calcul d'une intégrale double

Le calcul d'une intégrale double se ramène au calcul de deux intégrales définies.

Limitons-nous, dans un premier temps, au cas où les parallèles aux axes qui coupent le contour de domaine D ont en commun avec D un segment.
On peut choisir des points H_{ij} d'abcisse ξ_i, constante pour tous les rectangles d'une rangée parallèle à Oy et d'ordonnée η_j, constante pour tous les rectangles d'une rangée parallèle à Ox.
On obtient alors la somme double :

$$\sum_{i=1}^{n}\sum_{j=1}^{m} f(\xi_i,\eta_j)\Delta x_i \Delta y_j$$

que l'on peut écrire :
$\sum_{i=1}^{n} \Delta x_i \left(\sum_{j=1}^{m} f(\xi_i,\eta_j)\Delta y_j\right)$
Or lorsque $n \to +\infty$ avec $\sup \Delta y_j \to 0$, il est clair que :

$$\lim \sum_{j=1}^{m} f(\xi_i,\eta_j)\Delta y_j = \int_{\phi_1(\xi_i)}^{\phi_2(\xi_i)} f(\xi_i, y)\mathrm{d}y = F(\xi_i)$$

et lorsque $n \to +\infty$ avec $\sup \Delta x_i \to 0$, on obtient :

$$\lim \sum_{j=1}^{n} F(\xi_i)\Delta x_i = \int_a^b F(x)\mathrm{d}x$$

Finalement :

$$\int\int_D \mathrm{d}x\mathrm{d}y = \int_a^b \mathrm{d}x \int_{\phi_1(\xi_i)}^{\phi_2(\xi_i)} f(x,y)\mathrm{d}y$$

Cette écriture signifie que, x étant fixé, on intègre d'abord par rapport à y ; la fonction ϕ_2 admet pour graphe la partie supérieure du contour de D, la fonction ϕ_1 la partie inférieure.
On obtient alors une fonction de x qu'il reste à intégrer de a à b.
En inversant l'ordre de la mise en facteurs, on obtiendrait également :

$$\int\int_D \mathrm{d}x\mathrm{d}y = \int_c^d \mathrm{d}y \int_{\psi_1(y)}^{\psi_2(y)} f(x,y)\mathrm{d}x$$

Exemple B.1 $\int\int_D (x+y^2)dxdy$ *où* D *est défini par* $\{x \geq 0, y \geq 0, x+y \leq 1\}$. *Il est facile de représenter le domaine (premier quadrant en dessous de la droite* $y = 1-x$*)*
• *Premier calcul :*

$$\begin{aligned} I &= \int_0^1 dx \int_0^{1-x} (x+y^2)dy = \int_0^1 dx \left[xy + \frac{y^3}{3}\right]_0^{1-x} \\ &= \int_0^1 \frac{1}{3}(1-x^3)dx = \frac{1}{3}\left[x - \frac{x^4}{4}\right]_0^1 = \frac{1}{4} \end{aligned}$$

- *Deuxième calcul :*

$$
\begin{aligned}
I &= \int_0^1 dy \int_0^{1-y} (x+y^2)dx = \int_0^1 dy \left[\frac{x^2}{2} + xy^2\right]_0^{1-y} \\
&= \int_0^1 \left(-y^3 + \frac{3}{2}y^2 - y + \frac{1}{2}\right) dy \\
&= \left[-\frac{y^4}{4} + \frac{y^3}{2} - \frac{y^2}{2} + \frac{y}{2}\right]_0^1 = \frac{1}{4}
\end{aligned}
$$

1.3 Propriétés

1. L'intégrale double se calculant par deux intégrales simples successives, il résulte de la linéarité d'une intégrale définie que l'intégrale double est une expression linéaire de la fonction à intégrer, lorsque le domaine d'intégration est fixé :

$$
\begin{aligned}
\int\int_D [g(x,y) + h(x,y)]\mathrm{d}x\mathrm{d}y &= \int\int_D g(x,y)\mathrm{d}x\mathrm{d}y + \int\int_D h(x,y)\mathrm{d}x\mathrm{d}y \\
\int\int_D \lambda f(x,y)\mathrm{d}x\mathrm{d}y &= \lambda \int\int_D f(x,y)\mathrm{d}x\mathrm{d}y \quad , \ \lambda \in \mathbb{R}
\end{aligned}
$$

2. On peut montrer également que si le domaine D est la réunion de deux domaines disjoints D_1 et D_2, on a :

$$
\int\int_D f(x,y)\mathrm{d}x\mathrm{d}y = \int\int_{D_1} f(x,y)\mathrm{d}x\mathrm{d}y + \int\int_{D_2} f(x,y)\mathrm{d}x\mathrm{d}y
$$

Cette propriété permet d'étendre la notion d'intégrale double à des domaines de forme générale.

1.4 Cas particuliers

1. Le domaine d'intégration est un rectangle.
   On a alors :

$$
\int\int_D f(x,y)\mathrm{d}x\mathrm{d}y = \int_a^b \mathrm{d}x \int_c^d f(x,y)\mathrm{d}y
$$

   Les bornes de la première intégrale ne sont pas fonction de x.

2. Le domaine d'intégration est un rectangle et $f(x,y) = g(x)f(y)$.
   Dans ce cas :

$$
\begin{aligned}
\int\int_D f(x,y)\mathrm{d}x\mathrm{d}y &= \int_a^b \mathrm{d}x \int_c^d g(x)f(y)\mathrm{d}y \\
&= \left[\int_a^b g(x)\mathrm{d}x\right]\left[\int_c^d f(y)\mathrm{d}y\right]
\end{aligned}
$$

   L'intégrale double est alors le produit de deux intégrales simples.

Exemple B.2 *$\int\int_D xydxdy$ où D est défini par : $0 \leq x \leq 1$ et $0 \leq y \leq 2$.*
On a : $I = \int_0^1 xdx \int_0^2 ydy = \left[\frac{x^2}{2}\right]_0^1 \left[\frac{y^2}{2}\right]_0^2 = 1$

1.5 Calcul de volumes et de surfaces

Si $f(x, y)$ est positif ou nul en tout point du domaine D, le produit $f(H_{ij})\Delta x_i \Delta y_j$ est égal au volume d'un parallélépipède rectangle de surface de base $\Delta x_i \Delta y_j$ et de hauteur $f(H_{ij})$. En sommant sur i et sur j, on obtient un volume qui, à la limite, est égal au volume V de la portion de cylindre, de génératrices parallèles à Oz, limité par la base D et par la surface S. On posera donc :

$$V = \int\int_D f(x, y)\mathrm{d}x\mathrm{d}y$$

Par contre, si $f(x, y)$ n'est pas forcément positif ou nul, on parlera de volume algébrique. Les portions situées en dessous du plan xOy étant comptées négativement (voir propriété 2).
Si on considère maintenant la fonction $f : (x, y) \mapsto 1$, définie sur un domaine D, il est clair que l'intégrale $\int\int_D \mathrm{d}x\mathrm{d}y$ donne le volume d'un cylindre de base D et de hauteur 1. Le nombre obtenu est donc également l'aire de la surface D. On posera donc :

$$S = \int\int_D \mathrm{d}x\mathrm{d}y$$

Exemple B.3 *(Volume d'une sphère). Une sphère de centre O et de rayon R admet pour équation, par rapport à un repère orthonormé d'origine O : $x^2 + y^2 + z^2 = R^2$*
La demi-sphère S située au-dessus du plan xOy a pour équation :

$$x^2 + y^2 + z^2 = R^2 \text{ et } z \geq 0 \Leftrightarrow z = \sqrt{R^2 - x^2 - y^2}$$

C'est donc le graphe de la fonction $f : (x, y) \mapsto \sqrt{R^2 - x^2 - y^2}$ définie sur le domaine $D : x^2 + y^2 \leq R^2$.
Le volume limité par la demi-sphère S et le plan xOy est donc :

$$\begin{aligned} I &= \int\int_D \sqrt{R^2 - x^2 - y^2}dxdy \\ &= \int_{-R}^{R} dx \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \sqrt{R^2 - x^2 - y^2}dy \text{ sur } D\text{, pour } x \text{ fixé, } -\sqrt{R^2 - x^2} \leq y \leq \sqrt{R^2 - x^2} \\ &= \int_{-R}^{R} dx \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \sqrt{R^2 - x^2}\sqrt{1 - \frac{y^2}{R^2 - x^2}}dy \end{aligned}$$

On pose :

$$\begin{aligned} t = \arcsin \frac{y}{\sqrt{R^2 - x^2}} \quad &\Leftrightarrow \quad \frac{y}{\sqrt{R^2 - x^2}} = \sin t \text{ et } -\frac{\pi}{2} \leq t \leq \frac{\pi}{2} \\ &\Rightarrow \quad dy = \sqrt{R^2 - x^2}\cos t dt \\ &\quad\quad \sqrt{1 - \sin^2 t} = |\cos t| = \cos t \ (\geq 0 \text{ si } -\tfrac{\pi}{2} \leq t \leq \tfrac{\pi}{2}) \end{aligned}$$

Finalement :

$$I = \int_{-R}^{R} (R^2 - x^2)dx \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2 t dt = \int_{-R}^{R} (R^2 - x^2)dx \frac{\pi}{2} = \frac{\pi}{2} \left[R^2 x - \frac{x^3}{3} \right]_{-R}^{R} = \frac{2}{3}\pi R^3$$

D'où le volume de la sphère

$$V = 2I = \frac{4}{3}\pi R^3$$

Exemple B.4 *(Aire du cercle) Soit un cercle de centre O et de rayon R. On a :*

$$\begin{aligned} S = \int\int_D dxdy &= \int_{-R}^{R} dx \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy \\ &= 2\int_{-R}^{R} \sqrt{R^2 - x^2} dx \\ &= 2R \int_{-R}^{R} \sqrt{1 - \frac{x^2}{R^2}} dx \end{aligned}$$

On pose

$$\begin{aligned} u = \arcsin \frac{x}{R} \quad &\Leftrightarrow \quad \frac{x}{R} = \sin u \textit{ et } -\frac{\pi}{2} \leq u \leq \frac{\pi}{2} \\ &\Rightarrow \quad dx = R \cos u du \\ &\quad\quad \sqrt{1 - \sin^2 u} = |\cos u| = \cos u \end{aligned}$$

On obtient

$$S = 2R^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2 u du = 2R^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1 + \cos 2u}{2} du = 2R^2 \left[\frac{u}{2} + \frac{\sin 2u}{4} \right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = 2R^2 \frac{\pi}{2} = \pi R^2$$

1.6 Changement de variables

Soit l'intégrale double $\int\int_D f(x, y) \mathrm{d}x\mathrm{d}y$.
Pour faciliter les calculs, il est souvent commode de changer de variables. Sans aller jusqu'à la justification, relativement compliquée, du procédé, il convient cependant d'en connaître les règles.
Le changement de variables, dans un intégrale double, consiste à passer d'un couple (x, y) à un autre couple (u, v) par l'intermédiaire d'une application $(x, y) \mapsto (u, v)$. Lorsque le point $H(x, y)$ décrit le domaine D, son image $K(u, v)$ va décrire un nouveau domaine Δ.
L'application doit donc être bijective.
Considérons maintenant les fonctions ϕ et ψ définies par : $x = \phi(u, v)$ et $y = \psi(u, v)$, les formules de changement de variables.
On appelle déterminant fonctionnel ou Jacobien des fonctions ϕ et ψ le déterminant :

$$\frac{D(x, y)}{D(u, v)} = \begin{vmatrix} x'_u & x'_v \\ y'_u & y'_v \end{vmatrix} = x'_u y'_v - x'_v y'_u$$

Sous réserve de la continuité des fonctions ϕ et ψ et de leurs dérivées partielles sur Δ, on a :

$$\int\int_D f(x,y)\mathrm{d}x\mathrm{d}y = \int\int_\Delta f\left(\phi(u,v),\psi(u,v)\right)\left|\frac{D(x,y)}{D(u,v)}\right|\mathrm{d}u\mathrm{d}v$$

Application : utilisation des coordonnées polaires

À tout point M du plan muni du repère orthonormé $(O,\overrightarrow{i},\overrightarrow{j},\overrightarrow{k})$, on associe ses coordonnées cartésiennes x et y qui sont telles que : $\overrightarrow{OM} = x\overrightarrow{i} + y\overrightarrow{j}$.
On peut également lui associer ses coordonnées polaires r et θ : si $\overrightarrow{u}$ est le vecteur unitaire parallèle et de même sens que $\overrightarrow{OM}$, on a :

$$\overrightarrow{OM} = r\overrightarrow{u}\ (r \geq 0)\ ,\ \theta = (\widehat{\overrightarrow{i},\overrightarrow{u}})(0 \leq \theta < 2\pi)$$

Dans ces conditions, il est clair que se donner (x,y), c'est se donner (r,θ) et réciproquement ; l'application $(x,y) \mapsto (r,\theta)$ est bijective.
On a : $x = r\cos\theta$ et $y = r\sin\theta$

$$\frac{D(x,y)}{D(r,\theta)} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r$$

$$r \geq 0 \Rightarrow \left|\frac{D(x,y)}{D(r,\theta)}\right| = r$$

et on peut écrire :

$$\int\int_D f(x,y)\mathrm{d}x\mathrm{d}y = \int\int_\Delta f(r\cos\theta, r\sin\theta)r\mathrm{d}r\mathrm{d}\theta$$

Exemple B.5 *On avait à calculer :* $I = \int\int_D \sqrt{R^2-x^2-y^2}dxdy$
On passe en coordonnées polaires. On obtient :
$\int\int_\Delta \sqrt{R^2-r^2}rdrd\theta$ *où* Δ *est défini par :* $0 \leq r \leq R$, $0 \leq \theta < 2\pi$.
D'où

$$\begin{aligned} I &= \int_0^{2\pi} d\theta \int_0^R \sqrt{R^2-r^2}rdr \qquad \textit{cas particulier 2 du paragraphe 1.4} \\ &= 2\pi\left[-\frac{1}{3}(R^2-r^2)^{3/2}\right]_0^R = \frac{2}{3}\pi R^3 \end{aligned}$$

et donc

$$V = \frac{4}{3}\pi R^3$$

Exemple B.6 $\int\int_D x^2+y^2dxdy$ *où* D *est défini par* $x \geq 0$, $y \geq 0$ *et* $x^2+y^2 \leq 2$.
Dans un plan muni d'un repère orthonormé $(O,\overrightarrow{i},\overrightarrow{j},\overrightarrow{k})$, *le domaine* D *est représenté par le quart de cercle, de centre* O *et de rayon* $\sqrt{2}$, *situé dans le premier quadrant. On passe en coordonnées polaires et on obtient :*
$I = \int\int_\Delta r^3drd\theta$ *où* Δ *est défini par :* $0 \leq r \leq \sqrt{2}$, $0 \leq \theta < \frac{\pi}{4}$.
D'où

$$I = \int_0^{\frac{\pi}{4}} d\theta \int_0^{\sqrt{2}} r^3dr = [\theta]_0^{\frac{\pi}{4}}\left[\frac{r^4}{4}\right]_0^{\sqrt{2}} = \frac{\pi}{4}$$

2 Intégrale triple

2.1 Notion d'intégrale triple

Soit une fonction $f : (x, y, z) \mapsto t = f(x, y, z)$, définie sur un domaine D fermé et supposée nulle à l'extérieur de D.
L'espace étant muni d'un repère orthormé direct $(O, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$, D peut être représenté par l'ensemble des points d'un volume limité par une surface fermée S.
On désigne par a et b, c et d, e et f les bornes inférieures et supérieures des abcisses, des ordonnées et des côtes des point de S.
On partage $[a, b]$ par les valeurs $x_0 = a < x_1 < x_2 < \ldots < x_i < \ldots < x_m = b$
$[c, d]$ par les valeurs $y_0 = c < y_1 < y_2 < \ldots < y_j < \ldots < y_m = d$
et $[e, f]$ par les valeurs $z_0 = e < z_1 < z_2 < \ldots < z_k < \ldots < z_p = f$
et on quadrille l'espace par des plans d'abcisses $x_0, x_1, \ldots, x_m$ des plans d'ordonnée $y_0, y_1, \ldots, y_n$ et des plans de côte $z_0, z_1, \ldots, z_p$, ce qui permet de définir un ensemble de mnp parallélépipèdes rectangles recouvrant D.
On désigne par E_{ijk} le parallélépipède rectangle

$$[x_{i-1}, x_i] \times [y_{j-1}, y_j] \times [z_{k-1}, y_k]$$

de dimensions $\Delta x_i = x_i - x_{i-1}$, $\Delta y_j = y_j - y_{j-1}$ et $\Delta z_k = z_k - z_{k-1}$.
On choisit arbitrairement dans chaque parallélépipède E_{ijk} un point M_{ijk}.
Soit maintenant la somme :

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} f(M_{ijk}) \Delta x_i \Delta y_j \Delta z_k$$

Si cette somme tend vers une limite I, lorsque m, n et p tendent vers l'infini de manière à ce que $\sup \Delta x_i$, $\sup \Delta y_j$ et $\sup \Delta z_k$ tendent vers zéro, indépendamment de la façon dont on fait les partages et du point M_{ijk}, on dit que f est intégrable sur D, que I est l'intégrale triple de f sur D et on pose :

$$\int\int\int_D f(x, y, z) \mathrm{d}x \mathrm{d}y \mathrm{d}z$$

On peut démontrer, en particulier, que, si f est continue sur D, l'intégrale existe.

2.2 Calcul d'une intégrale triple

Le calcul d'une intégrale triple se ramène au calcul d'une intégrale double et d'une intégrale simple donc, finalement, au calcul de trois intégrales simples successives.
On peut choisir $M_{ijk} = (\xi_i, \eta_j, \zeta_k)$; ceci revient à choisir même abcisse pour tous les points de premier indice i, même ordonnée pour tous les points de deuxième indice j et même côte pour tous les points de troisième indice k. On peut alors écrire la somme triple :

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} f(\xi_i, \eta_j, \zeta_k) \Delta x_i \Delta y_j \Delta z_k = \sum_{k=1}^{p} \Delta z_k \sum_{i=1}^{m} \sum_{j=1}^{n} f(\xi_i, \eta_j, \zeta_k) \Delta x_i \Delta y_j$$

À la limite, la somme double est l'intégrale double de $f(x, y, \zeta_k)$ par rapport à x et y, ζ_k étant fixé, sur un domaine $D(\zeta_k)$ obtenu par l'intersection de D par le plan de côté $z = \zeta_k$.
Cette intégrale double est évidemment fonction de ζ_k. Il ne reste plus qu'à sommer sur k ce qui, à la limite est une intégrale simple sur z. On écrira donc successivement :

$$\begin{aligned}
\int\int\int_D f(x,y,z)\mathrm{d}x\mathrm{d}y\mathrm{d}z &= \int_e^f \mathrm{d}z \int\int_{D(z)} f(x,y,z)\mathrm{d}x\mathrm{d}y \\
\int\int\int_D f(x,y,z)\mathrm{d}x\mathrm{d}y\mathrm{d}z &= \int_e^f \mathrm{d}z \int_{\phi_1(z)}^{\phi_2(z)} \mathrm{d}y \int_{\psi_1(y,z)}^{\psi_2(y,z)} f(x,y,z)\mathrm{d}x
\end{aligned}$$

Cette notation signifie que x et y étant fixées, on intègre d'abord par rapport à x ; les bornes de cette première intégrale dépendant en général de (y, z). On intègre ensuite par rapport à y, z étant fixé ; les bornes de cette deuxième intégrale dépendant en général de z. Enfin on intègre par rapport à z.
Bien entendu l'ordre d'intégration choisi ici est arbitraire ; on pourrait commencer à intégrer d'abord par rapport à z, x et y étant fixés, puis par rapport à x, y étant fixé, enfin par rapport à y.

Exemple B.7 *$\int\int\int_D xyzdxdydz$ où D est défini par : $x \geq 0$, $y \geq 0$, $z \geq 0$, $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1$ ($a > 0$, $b > 0$, $c > 0$).*
Le domaine D peut ici être représenté par un huitième d'ellipsoïde. On a

$$\begin{aligned}
I &= \int_0^a xdx \int_0^{b\sqrt{1-\frac{x^2}{a^2}}} ydy \int_0^{c\sqrt{1-\frac{x^2}{a^2}-\frac{y^2}{b^2}}} zdz \\
&= \int_0^a xdx \int_0^{b\sqrt{1-\frac{x^2}{a^2}}} ydy \left[\frac{c^2}{2}\left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right)\right] \\
&= \frac{c^2}{2}\int_0^a xdx \left[\frac{y^2}{2} - \frac{x^2}{a^2}\frac{y^2}{2} - \frac{y^4}{4b^2}\right]_0^{b\sqrt{1-\frac{x^2}{a^2}}} \\
&= \frac{c^2}{2}\frac{b^2}{2}\int_0^a \left[\left(1 - \frac{x^2}{a^2}\right)x - \left(1 - \frac{x^2}{a^2}\right)\frac{x^3}{a^2} - \left(1 - \frac{x^2}{a^2}\right)^2 \frac{x}{2}\right] dx \\
&= \frac{b^2c^2}{2}\left(\frac{a^2}{2} - \frac{a^2}{4} - \frac{a^2}{4} + \frac{a^2}{6} - \frac{a^2}{4} + \frac{a^2}{4} - \frac{a^2}{12}\right) \\
&= \frac{a^2b^2c^2}{48}
\end{aligned}$$

Le produit $\Delta x_i \Delta y_j \Delta z_k$ est égal au volume du parallélépipède rectangle E_{ijk}.
Soit maintenant la fonction :

$$g : \begin{cases} (x,y,z) & \mapsto & 1 \text{ sur } D \\ (x,y,z) & \mapsto & 0 \text{ ailleurs} \end{cases}$$

Il est clair qu'en multipliant $\Delta x_i \Delta y_j \Delta z_k$ par $g(M_{ijk})$ et en sommant sur i, j et k, on obtient un volume qui, à la limite, est égal au volume de la portion de l'espace représentant D.

2.3 Changement de variables

Soient les formules de changement de variables :

$$x = \phi(u,v,w)\ ,\ y = \psi(u,v,w) \text{ et } z = \chi(u,v,w)$$

définissant une correspondance bijective entre l'ensemble D des points (x,y,z) et l'ensemble Δ des points (u,v,w). On admettra que, sous réserve de la continuité des fonctions ϕ, ψ et χ et de leurs dérivées partielles :

$$\int\int\int_D f(x,y,z)\mathrm{d}x\mathrm{d}y\mathrm{d}z = \int\int\int_\Delta f\left(\phi(u,v,w),\psi(u,v,w),\chi(u,v,w)\right)\left|\frac{D(x,y,z)}{D(u,v,w)}\right|\mathrm{d}u\mathrm{d}v\mathrm{d}w$$

où $\frac{D(x,y,z)}{D(u,v,w)}$ est le Jacobien des fonctions ϕ, ψ et χ défini par le déterminant :

$$\frac{D(x,y,z)}{D(u,v,w)} = \begin{vmatrix} x'_u & x'_v & x'_w \\ y'_u & y'_v & y'_w \\ z'_u & z'_v & z'_w \end{vmatrix}$$

2.4 Généralisation

On peut envisager des intégrales n-uples de fonctions de n variables. Le support géométrique fera évidemment totalement défaut dès que $n > 3$, ce qui n'est pas nécessairement gênant.
En fait le processus de calcul sera le même que pour les intégrales doubles et triples.

3 Rappels de mécanique

L'espace est muni d'un repère orthonormé $Oxyz$
Pour un ensemble de points isolés $A_i(x_i,y_i,z_i)$ affectés d'une masse m_i, de masse totale $M = \sum_i m_i$, on définit :

le centre de masse $G(X,Y,Z)$ par la relation vectorielle $\sum_i m_i\overrightarrow{GA_i} = \overrightarrow{O}$. On en déduit $M\overrightarrow{OG} = \sum_i m_i\overrightarrow{OA_i}$, d'où les coordonnées de G :

$$X = \frac{1}{M}\sum_i m_i x_i\ ,\ Y = \frac{1}{M}\sum_i m_i y_i\ ,\ X = \frac{1}{M}\sum_i m_i z_i$$

le moment d'inertie par rapport à un point, à un axe, ou à un plan :

$$I = \sum_i m_i d_i^2$$

où d_i est la distance du point A_i au point, à l'axe, ou au plan par rapport auquel le moment d'inertie est calculé.

Pour une tige porté par l'axe Ox, de masse spécifique (linéaire) $\rho(x)$, (a, b) étant l'ensemble des abcisses des points de la tige, on obtient :

$$M = \int_a^b \rho(x)\mathrm{d}x \text{ , } X = \frac{1}{M}\int_a^b x\rho(x)\mathrm{d}x \text{ , } (Y = Z = 0)$$

$$I = \int_a^b d^2(x)\rho(x)\mathrm{d}x$$

où $d(x)$ est la distance du point $A(x)$ de la tige, à l'élément par rapport auquel I est calculé.

Pour un plaque plane situé sur le plan xOy, de masse spécifique (superficielle) $\rho(x, y)$, D étant l'ensemble des couples (x, y) associés aux points de la plaque, on obtient :

$$\begin{aligned} M &= \int\int_D \rho(x, y)\mathrm{d}x\mathrm{d}y \\ X &= \frac{1}{M}\int\int_D x\rho(x, y)\mathrm{d}x\mathrm{d}y \\ Y &= \frac{1}{M}\int\int_D y\rho(x, y)\mathrm{d}x\mathrm{d}y \text{ , } (Z = 0) \end{aligned}$$

$$I = \int\int_D d^2(x, y)\rho(x, y)\mathrm{d}x\mathrm{d}y$$

où $d(x, y)$ est la distance du point $A(x, y)$ de la plaque, à l'élément par rapport auquel I est calculé.

Pour un solide, à 3 dimensions, de masse spécifique $\rho(x, y, z)$, Δ étant l'ensemble des triplets (x, y, z) définissant les points du solide, on a :

$$M = \int\int\int_\Delta \rho(x, y, z)\mathrm{d}x\mathrm{d}y\mathrm{d}z \text{ , } X = \frac{1}{M}\int\int_D x\rho(x, y, z)\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

$$Y = \frac{1}{M}\int\int_D y\rho(x, y, z)\mathrm{d}x\mathrm{d}y\mathrm{d}z \text{ , } Z = \frac{1}{M}\int\int_D z\rho(x, y, z)\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

$$I = \int\int_D d^2(x, y, z)\rho(x, y, z)\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

où $d(x, y, z)$ est la distance du point $A(x, y, z)$ du solide, à l'élément par rapport auquel I est calculé.
Si le solide est homogène, la masse spécifique est constante.

4 Exercices

1. Calculer $\int\int_D xy\mathrm{d}x\mathrm{d}y$, où D est défini par $\{x \geq 0, y \geq 0, y \leq 1-2x\}$
2. Calculer, par deux méthodes différentes : $\int\int_{D_1} xy\mathrm{d}x\mathrm{d}y$, où D_1 est défini par $\{x \geq 0, y \geq 0, x^2+y^2 \leq R^2\}$ Calculer $\int\int_{D_2} xy\mathrm{d}x\mathrm{d}y$, où D_2 est défini par $\{x \geq 0, y \geq 0, x^2+y^2 \leq R^2, x+y \geq R > 0\}$
3. Calculer $\int\int_D x\sin(x+y)\mathrm{d}x\mathrm{d}y$, où D est défini par $\{0 \leq x \leq \pi, 0 \leq y \leq \frac{\pi}{2}\}$
4. Calculer $\int_0^{+\infty}\int_0^{+\infty} \frac{1}{(x^2+y^2+a^2)^2}\mathrm{d}x\mathrm{d}y$
5. Calculer l'aire plane limitée par les courbes d'équations, dans un repère orthonormé $y = x^2$ et $y = 2 - x^2$
6. L'espace est muni d'un repère orthonormé $Oxyz$. Calculer le moment d'inertie, par rapport à l'axe Oz, d'une plaque homogène situé dans le plan xOy et limitée par les courbes d'équations : $y^2 = 4x$, $x = 0$, $y = 2$.
7. Déterminer le centre de masse d'une plaque semi-circulaire homogène, de centre O et de rayon R.
8. Calculer $\int\int_D \frac{1}{(x+y)^2}\mathrm{d}x\mathrm{d}y$ où D est défini par $\{x \geq 1, y \geq 1, x+y \leq 3\}$
9. Calculer $\int\int_D \frac{xy}{x^2+y^2}\mathrm{d}x\mathrm{d}y$ où D est le triangle de sommets : $O(0,0)$, $A(a,a)$, $B(a,0)$ avec $a > 0$, les trois points étant repérés par rapport à un système orthonormé xOy.
10. Calculer $\int\int_D \frac{1}{x+y}\mathrm{d}x\mathrm{d}y$ où D est défini par $\{0 \leq x \leq 1, 0 \leq y \leq 1, x+y \geq 1\}$
11. Calculer $\int\int\int_D z^2\mathrm{d}x\mathrm{d}y\mathrm{d}z$ où D est défini par $\{x^2+y^2+z^2 \leq R^2\}$
12. Calculer $\int\int\int_V z\mathrm{d}x\mathrm{d}y\mathrm{d}z$ où V est défini par $\{x \geq 0, y \geq 0, z \geq 0, x+y+z \leq a\}$
13. Calculer $\int\int\int_V z\mathrm{d}(x+y+z)\mathrm{d}y\mathrm{d}z$ où V est défini par $\{-h \leq z \leq h, x^2+y^2 \leq R^2\}$
14. Calculer le volume, la masse, le moment d'inertie par rapport à son axe de révolution, d'un cône de révolution homogène, de hauteur H et de rayon de base R. Déterminer son centre de masse.
15. Calculer $\int\int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)}\mathrm{d}x\mathrm{d}y$. En déduire $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\mathrm{d}x$.
16. Calculer la surface d'un cercle de rayon R.
17. Calculer le volume d'une sphère de rayon R.
18. Calculer l'aire plane limitée par les courbes d'équation : $y^2 = 4x$ et $y = 2x - 4$, dans un repère orthonormé.
19. Calculer l'aire et déterminer le centre de masse d'une plaque plane homogène limitée par une courbe d'équation polaire :
    (a) $r^2 = a^2\cos 2\theta$ avec $-\frac{\pi}{4} \leq \theta \leq \frac{\pi}{4}$
    (b) $r = a\cos\theta$ avec $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$
20. L'espace est muni d'un repère orthonormé. Calculer le volume et déterminer le centre de masse d'un solide homogène limité par la surface d'équation $x^2+y^2+z = 9$ et le plan $z = 0$.

21. Calculer le volume et déterminer le centre de masse du solide homogène limité par les plans d'équations : $x = 0, y = 0, z = 0, \frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1$, $(a, b, c \in \mathbb{R}_+^*)$
22. Calculer le moment d'inertie par rapport à une de ses arêtes d'une plaque rectangulaire homogène.

5 Solutions

1. On a :

$$\begin{aligned}
I &= \int\int_D xy\mathrm{d}x\mathrm{d}y = \int_0^{1/2} x\mathrm{d}x \int_0^{1-2x} y\mathrm{d}y \\
&= \int_0^{1/2} x\mathrm{d}x \left[\frac{y^2}{2}\right]_0^{1-2x} \\
&= \frac{1}{2}\int_0^{1/2} (x-4x^2+4x^3)\mathrm{d}x \\
&= \frac{1}{2}\left[\frac{x^2}{2} - \frac{4x^3}{3} + x^4\right]_0^{\frac{1}{2}} = \frac{1}{2}\left(\frac{1}{8} - \frac{4}{24} + \frac{1}{16}\right) \\
&= \frac{1}{96}
\end{aligned}$$

2. Soit $I_1 = \int\int_{D_1} xy\mathrm{d}x\mathrm{d}y$.
   (a) on a

$$\begin{aligned}
I_1 &= \int_0^R x\mathrm{d}x \int_0^{\sqrt{R^2-x^2}} y\mathrm{d}y \\
&= \int_0^R x\mathrm{d}x \left[\frac{y^2}{2}\right]_0^{\sqrt{R^2-x^2}} \\
&= \frac{1}{2}\int_0^R (R^2x - x^3)\mathrm{d}x = \frac{1}{2}\left[R^2\frac{x^2}{2} - \frac{x^4}{4}\right]_0^R \\
&= \frac{R^4}{8}
\end{aligned}$$

   (b) On peut également écrire, en passant en coordonnées polaires :

$$\begin{aligned}
I_1 &= \int\int_{\Delta_1} r^2\cos\theta\sin\theta r\mathrm{d}r\mathrm{d}\theta \text{ où } \Delta_1 \text{ est défini par } \{0 \le \theta \le \tfrac{\pi}{2}, 0 \le r \le R\} \\
&= \int_0^{\pi/2} \cos\theta\sin\theta\mathrm{d}\theta \int_0^R r^3\mathrm{d}r \\
&= \left[\frac{\sin^2\theta}{2}\right]_0^{\frac{\pi}{2}} \left[\frac{r^4}{4}\right]_0^R \\
&= \frac{R^4}{8}
\end{aligned}$$

   Soit maintenant $I_2 = \int\int_{D_2} xy\mathrm{d}x\mathrm{d}y$. On a :

$$I_2 = \int_0^R x\mathrm{d}x \int_{R-x}^{\sqrt{R^2-x^2}} y\mathrm{d}y$$

$$\begin{aligned}
&= \int_0^R x\mathrm{d}x \left[\frac{y^2}{2}\right]_{R-x}^{\sqrt{R^2-x^2}} \\
&= \frac{1}{2}\int_0^R (R^2 - x^2 - R^2 - x^2 + 2Rx)x\mathrm{d}x = \frac{1}{2}\left[R^2\frac{x^3}{3} - \frac{2x^4}{4}\right]_0^R \\
&= \frac{R^4}{12}
\end{aligned}$$

3.

$$\begin{aligned}
I &= \int_0^\pi x\mathrm{d}x \int_0^{\pi/2} \sin(x+y)\mathrm{d}y \\
&= \int_0^\pi x\mathrm{d}x \left[-\cos(x+y)\right]_0^{\pi/2} \\
&= \int_0^\pi x\left(-\cos\left(x+\frac{\pi}{2}\right) + \cos x\right)\mathrm{d}x \\
&= \int_0^\pi x(\cos x + \sin x)\mathrm{d}x
\end{aligned}$$

On intègre par parties $\left(\int u\mathrm{d}v = uv - \int v\mathrm{d}u\right)$ en posant $x = u \Rightarrow \mathrm{d}x = \mathrm{d}u$ et $(\cos x + \sin x)\mathrm{d}x = \mathrm{d}v \Rightarrow v = \sin x - \cos x$

$$I = [x(\sin x - \cos x)]_0^\pi - \int_0^\pi (\sin x - \cos x)\mathrm{d}x = \pi - [-\cos x - \sin x]_0^\pi = \pi - 2$$

4. Soit $I = \int_0^{+\infty}\int_0^{+\infty} \frac{1}{(x^2+y^2+a^2)^2}\mathrm{d}x\mathrm{d}y$. Le domaine d'intégration peut être représenté dans un plan, muni d'un repère orthonormé, par le premier quadrant.

   On passe en coordonnées polaires :

$$I = \int_0^{\pi/2} \mathrm{d}\theta \int_0^{+\infty} \frac{r}{(r^2+a^2)^2}\mathrm{d}r = \frac{\pi}{2}\left[-\frac{1}{r^2+a^2}\right]_0^{+\infty} = \frac{\pi}{2a^2}$$

5. Dans un plan muni d'un repère orthonormé, l'aire à calculer est limitée par deux paraboles admettant $y'Oy$ pour axe de symétrie et symétriques entre elles par rapport à la droite d'équation $y = 1$. Ces deux paraboles se coupent en $A(-1,1)$ et $B(1,1)$: $y = 2 - x^2 = x^2 \Rightarrow x^2 = 1 \Rightarrow x = \pm 1$. On a :

$$S \int\int_D \mathrm{d}x\mathrm{d}y = \int_{-1}^1 \mathrm{d}x \int_{x^2}^{2-x^2} \mathrm{d}y = \int_{-1}^1 (2-2x^2)\mathrm{d}x = \left[2x - 2\frac{x^3}{3}\right]_{-1}^1 = \frac{8}{3}$$

6. La plaque plane est la surface entre l'axe Oy, la droite $y = 2$ et la courbe $y^2 = 4x$. On a, si D est l'ensemble des (x,y) définissant la plaque de masse spécifique ρ :

$$I = \int\int_D \rho(x^2+y^2)\mathrm{d}x\mathrm{d}y = \rho\int_0^2 \mathrm{d}y \int_0^{y^2/4} (x^2+y^2)\mathrm{d}x$$

$$\begin{aligned}
&= \rho \int_0^2 \mathrm{d}y \left[\frac{x^3}{3} + y^2 x\right]_0^{y^2/4} = \rho \int_0^2 \left(\frac{y^6}{192} + \frac{y^4}{4}\right) \mathrm{d}y \\
&= \rho \left[\frac{y^7}{1344} + \frac{y^5}{20}\right]_0^2 \\
&= \frac{178}{105}\rho
\end{aligned}$$

Il est d'usage d'exprimer un moment d'inertie en fonction de la masse M du solide. Si S est l'aire de la plaque, on a :

$$M = \int\int_D \rho \mathrm{d}x\mathrm{d}y = \rho \int\int_D \mathrm{d}x\mathrm{d}y = \rho S$$

$$S = \int\int_D \mathrm{d}x\mathrm{d}y = \int_0^2 \mathrm{d}y \int_0^{y^2/4} \mathrm{d}x = \left[\frac{y^3}{12}\right]_0^2 = \frac{2}{3} \Rightarrow M = \frac{2}{3}\rho \Rightarrow I = \frac{3}{2}M\frac{178}{105} = \frac{89}{35}M$$

7. On choisit le repère orthonormé Oxy de manière que Oy soit l'axe de symétrie de la plaque. Soient ρ la masse spécifique (constante) de la plaque, M sa masse, D l'ensemble des (x, y) définissant la plaque. Soit $G(x, y)$ le centre de masse cherché. On a :

$$X = \frac{\rho}{M}\int\int_D x\mathrm{d}x\mathrm{d}y = \frac{\rho}{M}\int\int_\Delta r\cos\theta r\mathrm{d}r\mathrm{d}\theta$$

, en passant en coordonnées polaires. Δ est défini par : $\{0 \leq \theta \leq \pi, 0 \leq r \leq R\}$

$$X = \frac{\rho}{M}\int_0^\pi \cos\theta\mathrm{d}\theta \int_0^R r^2\mathrm{d}r = \frac{\rho}{M}\frac{R^3}{3}[\sin\theta]_0^\pi = 0$$

Ce résultat était prévisible puisque Oy est un axe de symétrie. Le centre de masse est donc situé sur Oy et a pour ordonnée :

$$Y = \frac{\rho}{M}\int\int_D y\mathrm{d}x\mathrm{d}y = \frac{\rho}{M}\int\int_\Delta r\sin\theta r\mathrm{d}r\mathrm{d}\theta = \frac{\rho}{M}\int_0^\pi \sin\theta\mathrm{d}\theta\int_0^R r^2\mathrm{d}r = \frac{\rho}{M}\frac{R^3}{3}[-\cos\theta]_0^\pi$$

c'est-à-dire

$$Y = \frac{2}{3}\frac{\rho}{M}R^3$$

Or $M = \int\int_D \rho\mathrm{d}x\mathrm{d}y = \rho\int\int_D \mathrm{d}x\mathrm{d}y = \frac{1}{2}\pi R^2\rho \Rightarrow \frac{\rho}{M} = \frac{2}{\pi R^2}$. Finalement :

$$Y = \frac{4}{3\pi}R \approx 0.4244R$$

8. Le domaine d'intégration est facile à représenter. On a :

$$\begin{aligned}
\int\int_D \frac{1}{(x+y)^2}\mathrm{d}x\mathrm{d}y &= \int_1^2 \mathrm{d}x\int_1^{3-x}\frac{1}{(x+y)^2}\mathrm{d}y \\
&= \int_1^2 \mathrm{d}x\left[-\frac{1}{(x+y)}\right]_1^{3-x} \\
&= \int_1^2\left(-\frac{1}{3}+\frac{1}{x+1}\right)\mathrm{d}x \\
&= \left[\log(x+1) - \frac{x}{3}\right]_1^2 = \log\frac{3}{2} - \frac{1}{3}
\end{aligned}$$

9. Soit $I = \int\int_D \frac{xy}{x^2+y^2}\mathrm{d}x\mathrm{d}y$. On passe en coordonnées polaires :

$$\begin{aligned}
I &= \int\int_\Delta \frac{r^2\cos\theta\sin\theta}{r^2} r\mathrm{d}r\mathrm{d}\theta \\
&= \int_0^{\pi/4} \cos\theta\sin\theta\mathrm{d}\theta \int_0^{\frac{a}{\cos\theta}} r\mathrm{d}r \\
&= \int_0^{\pi/4} \frac{a^2}{2\cos^2\theta}\cos\theta\sin\theta\mathrm{d}\theta \\
&= \frac{a^2}{2}\int_0^{\pi/4}\frac{\sin\theta}{\cos\theta}\mathrm{d}\theta \\
&= -\frac{a^2}{2}[\log(\cos\theta)]_0^{\pi/4} \\
&= \frac{a^2}{2}\log\sqrt{2} = \frac{a^2}{4}\log 2
\end{aligned}$$

10. Soit $I = \int\int_D \frac{1}{x+y}\mathrm{d}x\mathrm{d}y$.
On a $I = \int_0^1 \mathrm{d}y \int_{1-y}^1 \frac{1}{x+y}\mathrm{d}x = \int_0^1 \mathrm{d}y[\log(x+y)]_{1-y}^1 = \int_0^1 \log(1+y)\mathrm{d}y$.
On intègre par parties : $\left(\int u\mathrm{d}v = uv - \int v\mathrm{d}u\right)$ en posant $\mathrm{d}v = \mathrm{d}y \Rightarrow v = y$ et $\log(1+y) = u \Rightarrow \mathrm{d}u = \frac{\mathrm{d}y}{1+y}$

$$\begin{aligned}
I &= [y\log(1+y)]_0^1 - \int_0^1 \frac{y}{1+y}\mathrm{d}y \\
&= \log 2 - \int_0^1 \frac{1+y}{1+y}\mathrm{d}y + \int_0^1 \frac{1}{1+y}\mathrm{d}y \\
&= \log 2 - [y]_0^1 + [\log(1+y)]_0^1 \\
&= 2\log 2 - 1 \approx 0.3863
\end{aligned}$$

11. Soit $I = \int\int\int_D z^2\mathrm{d}x\mathrm{d}y\mathrm{d}z$.
L'espace étant muni d'un repère orthonormé d'origine O, le domaine d'intégration peut être représenté par l'ensemble des points situés à l'intérieur d'une sphère de centre O et de rayon R. La symétrie de révolution autour de l'axe Oz suggère le passage aux coordonnées cylindres, c'est-à-dire le changement : $x = r\cos\theta$; $y = r\sin\theta$; $z = z$. Le domaine d'intégration Δ étant défini par $0 \leq 2\pi$, $0 \leq r \leq R$, $z^2 + r^2 \leq R^2$.
Le Jacobien de cette transformation est $\frac{D(x,y)}{D(u,v)} = \begin{vmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = r$
D'où $|J| = |r| = r$.
finalement

$$I = \int\int\int_\Delta z^2 r\mathrm{d}r\mathrm{d}\theta\mathrm{d}z = \int_0^{2\pi}\mathrm{d}\theta\int_0^R r\mathrm{d}r\int_{-\sqrt{R^2-z^2}}^{\sqrt{R^2-r^2}}\mathrm{d}z = 2\pi\int_0^R \frac{2}{3}(R^2-r^2)^{3/2} r\mathrm{d}r$$

D'où $I = \frac{2\pi}{3}\left[-\frac{2}{5}(R^2-r^2)^{5/2}\right]_0^R = \frac{4}{15}\pi R^5$

12. Soit $I = \int\int\int_V z\mathrm{d}x\mathrm{d}y\mathrm{d}z$.
On remarque que $x+y+z \geq 0 \Rightarrow a \geq 0$. Les triplets (x,y,z) du domaine V sont tels que :
   - $0 \leq x \leq a-y-z$
   - $0 \leq y \leq a-z$ la condition 12 impliquant $a-y-z \geq 0 \Rightarrow y \leq a-z$
   - $0 \leq z \leq a$ la condition 12 impliquant $a-z \geq 0 \Rightarrow z \leq a$

On en déduit $I = \int_0^a z\mathrm{d}z \int_0^{a-z}\mathrm{d}y \int_0^{a-y-z}\mathrm{d}x = \int_0^a z\mathrm{d}z \int_0^{a-z}(a-y-z)\mathrm{d}y$

$$\int_0^a z\mathrm{d}z\left[(a-z)y-\frac{y^2}{2}\right]_0^{a-z} = \int_0^a \frac{1}{2}(a-z)^2 z\mathrm{d}z = \frac{1}{2}\left[a^2\frac{z^2}{2}-2a\frac{z^3}{3}+\frac{z^4}{4}\right]_0^a = \frac{a^4}{24}$$

L'espace étant muni d'un repère orthonormé, le domaine d'intégration V peut être représenté par une pyramide de sommet O.

13. Soit $I = \int\int\int_V (x+y+z)\mathrm{d}x\mathrm{d}y\mathrm{d}z$.
L'espace étant muni d'un repère orthonormé $Oxyz$, le domaine d'intégration V peut être représenté par l'ensemble des points situés à l'intérieur d'un cylindre de révolution autour de l'axe Oz. On a :

$$\begin{aligned}
I &= \int_{-R}^{R}\mathrm{d}x\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}}\mathrm{d}y\int_{-h}^{h}(x+y+z)\mathrm{d}z\\
&= \int_{-R}^{R}\mathrm{d}x\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}}\left[xz+yz+\frac{z^2}{2}\right]_{-h}^{h}\mathrm{d}y\\
&= \int_{-R}^{R}\mathrm{d}x\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}}2h(x+y)\mathrm{d}y\\
&= \int_{-R}^{R}2h\mathrm{d}x\left[xy+\frac{y^2}{2}\right]_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}}\\
&= \int_{-R}^{R}2h2x\sqrt{R^2-x^2}\mathrm{d}x
\end{aligned}$$

D'où

$$I = 2h\left[-\frac{2}{3}(R^2-x^2)^{3/2}\right]_{-R}^{R} = 0$$

On aurait pu obtenir le même résultat en passant en coordonnées cylindriques (voir exercice 11).

14. On choisit un repère orthonormé $Oxyz$ dont l'origine est le sommet du cône et Oz l'axe de révolution. Soit D l'ensemble des triplets (x,y,z) correspondant aux points situés à l'intérieur du cône. On désignera par : V le volume du cône, ρ sa masse spécifique (constante), M sa masse, I son moment d'inertie par rapport à Oz, et $G(X,Y,Z)$ son centre de masse. On a :

$$V = \int\int\int_D \mathrm{d}x\mathrm{d}y\mathrm{d}z = \int\int\int_\Delta r\mathrm{d}r\mathrm{d}\theta\mathrm{d}z$$

en passant en coordonnées cylindriques.

Le domaine d'intégration Δ est défini par $\{0 \leq \theta \leq 2\pi\,; 0 \leq z \leq H\,; 0 \leq r \leq \frac{R}{H}z\}$.

Donc $V = \int_0^{2\pi} \mathrm{d}\theta \int_0^H \mathrm{d}z \int_0^{\frac{R}{H}z} r\mathrm{d}r = 2\pi\frac{1}{2}\int_0^H \frac{R^2}{H^2}z^2\mathrm{d}z = \pi\frac{R^2}{H^2}\left[\frac{z^3}{3}\right]_0^H = \frac{1}{3}\pi R^2 H$

On a $M = \int\int\int_D \rho\mathrm{d}x\mathrm{d}y\mathrm{d}z = \rho V = \frac{1}{3}\pi R^2 H\rho$

On a $I = \int\int\int_D \rho(x^2+y^2)\mathrm{d}x\mathrm{d}y\mathrm{d}z = \rho\int\int\int_\Delta r\times r^2\mathrm{d}r\mathrm{d}\theta\mathrm{d}z = \rho\int_0^{2\pi}\mathrm{d}\theta\int_0^H\mathrm{d}z\int_0^{\frac{R}{H}z} r^3\mathrm{d}r = 2\pi\rho\frac{1}{4}\int_0^H \frac{R^4}{H^4}z^4\mathrm{d}z = \frac{1}{2}\pi\rho\frac{R^4}{H^4}\frac{H^5}{5} = \frac{1}{10}\pi\rho R^4 H = \frac{3}{10}MR^2$

Le cône étant homogène, il est évident que le centre de masse $G(X, Y, Z)$ est sur l'axe Oz, donc $X = Y = 0$ (à vérifier). On a :

$Z = \frac{1}{M}\int\int\int_D \rho z\mathrm{d}x\mathrm{d}y\mathrm{d}z = \frac{\rho}{M}\int\int\int_\Delta zr\mathrm{d}r\mathrm{d}\theta\mathrm{d}z = \frac{\rho}{M}\int_0^{2\pi}\mathrm{d}\theta\int_0^H z\mathrm{d}z\int_0^{\frac{R}{H}z} r\mathrm{d}r = \frac{\rho}{M}2\pi\frac{1}{2}\int_0^H \frac{R^2}{H^2}z^3\mathrm{d}z = \pi\frac{\rho}{M}\frac{R^2}{H^2}\left[\frac{z^4}{4}\right]_0^H = \frac{1}{4}\pi\frac{\rho}{M}R^2H^2 = \frac{3}{4}H$

15. Il est clair que l'on a : $J = \int\int_{\mathbb{R}\times\mathbb{R}} e^{-\frac{1}{2}(x^2+y^2)}\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\mathrm{d}x\int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2}\mathrm{d}y = I \times I = I^2$ en posant $I = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2}\mathrm{d}u$

    Or si on passe en coordonnées polaires, on : $J = \int\int_\Delta e^{-\frac{1}{2}r^2}r\mathrm{d}r\mathrm{d}\theta$ où Δ est défini par : $\{0 \leq \theta \leq 2\pi, r \geq 0\}$

    donc $J = \int_0^{2\pi}\mathrm{d}\theta\int_0^{+\infty} e^{-\frac{1}{2}r^2}r\mathrm{d}r = [\theta]_0^{2\pi}\left[-e^{-\frac{1}{2}r^2}\right]_0^{+\infty} = 2\pi \times 1 = 2\pi$

    On en déduit $I = \sqrt{2\pi}$ et donc $\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}\mathrm{d}u = 1$

16. On a $S = \int\int_D \mathrm{d}x\mathrm{d}y$ où D est défini par $x^2 + y^2 \leq R^2$. En passant en coordonnées polaires, on obtient : $S = \int\int_\Delta r\mathrm{d}r\mathrm{d}\theta$ où Δ est défini par $\{0 \leq \theta \leq 2\pi, r \leq R$

    donc $S = \int_0^{2\pi}\mathrm{d}\theta\int_0^R r\mathrm{d}r = [\theta]_0^{2\pi}\left[\frac{r^2}{2}\right]_0^R = \pi R^2$

17. On a : $V = \int\int\int_D \mathrm{d}x\mathrm{d}y\mathrm{d}z$ où D est défini par $x^2 + y^2 + z^2 \leq R^2$. En passant en coordonnées polaires, on obtient : $V = \int\int\int_\Delta r\mathrm{d}r\mathrm{d}\theta\mathrm{d}z$ où Δ est défini par $\{0 \leq \theta \leq 2\pi, 0 \leq r \leq R, z^2 + r^2 \leq R^2\}$

    donc $V = \int_0^{2\pi}\mathrm{d}\theta\int_0^R r\mathrm{d}r\int_{-\sqrt{R^2-r^2}}^{\sqrt{R^2-r^2}}\mathrm{d}z = \int_0^{2\pi}\mathrm{d}\theta\int_0^R r2\sqrt{R^2-r^2}\mathrm{d}r = 2\pi\left[-\frac{2}{3}(R^2-r^2)^{3/2}\right]_0^R = \frac{4}{3}\pi R^3$