Fiche TD avec le logiciel : tdr321

Quelques tests liés aux variables discrètes

D. Chessel, A.B. Dufour & J.R. Lobry

Khi2 de contingence, test exact de Fisher, test de McNemar : quelques exemples

Table des matières

1 Quelques rappels

1.1 La table de contingence observée

Soient A et B, deux variables qualitatives ayant respectivement p et q modalités. Soit n, le nombre d'individus sur lesquels A et B ont été observées. La *table de contingence* observée est un tableau croisé où les colonnes correspondent aux q modalités de la variable B et les lignes aux p modalités de la variable A. On note n_{ij} le nombre d'individus possédant à la fois la modalité i de la variable A et la modalité j de la variable B.

	$B1$	$B2$	$\cdots$	Bj	$\cdots$	Bq	total
$A1$	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1q}	$n_{1\cdot}$
$A2$	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2q}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Ai	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iq}	$n_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Ap	n_{p1}	n_{p2}	$\cdots$	n_{pj}	$\cdots$	n_{pq}	$n_{p\cdot}$
total	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot q}$	$n_{\cdot\cdot}$

Remarques :

- les sommes marginales lignes sont $n_{i\cdot} = \sum_{j=1}^{q} n_{ij}$
- les sommes marginales colonnes sont $n_{\cdot j} = \sum_{i=1}^{p} n_{ij}$
- Les totaux des lignes sont identiques aux fréquences absolues issues de l'étude univariée de A.
- Les totaux des colonnes sont identiques aux fréquences absolues issues de l'étude univariée de B.
- Les sommes marginales sont liées entre elles par $n = n_{\cdot\cdot} = \sum_{j=1}^{q} n_{\cdot j} = \sum_{i=1}^{p} n_{i\cdot}$
- L'ordre d'entrée des variables dans la table de contingence n'a aucune importance. Mais on peut privilégier une des variables en constituant un tableau de profils associés aux lignes (respectivement aux colonnes).
- Le *tableau des profils* lignes (respectivement colonnes) est défini par les fréquences conditionnelles : $\frac{n_{ij}}{n_{i\cdot}}$ (respectivement $\frac{n_{ij}}{n_{\cdot j}}$). La somme de chaque ligne (respectivement colonnes) est alors ramenée à l'unité.

1.2 Le Chi-Deux de Contingence

Afin de mesurer l'intensité de la relation entre deux variables qualitatives, on calcule un paramètre statistique appelé *Chi-deux*, lié à la loi de probabilité notée χ^2. Pour éviter les confusions, on utilisera la notation χ^2_{obs} pour la statistique calculée à partir des observations et χ^2 pour désigner la loi (χ^2_n pour un χ^2 à n degrés de liberté). La statistique χ^2_{obs} permet de comparer les valeurs de la table de contingence observée avec les valeurs d'une table de contingence théorique. Les données de la table de contingence théorique sont définies par :

- les sommes marginales lignes sont identiques à celles de la table observée ;
- les sommes marginales colonnes sont identiques à celles de la table observée ;

- le nombre d'individus possédant à la fois la modalité i de la variable A et la modalité j de la variable B est $\frac{n_{i\cdot}n_{\cdot j}}{n}$

La valeur du Chi-Deux est définie par

$$\chi^2_{obs} = \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{\left(n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot}n_{\cdot j}}{n}}$$

Si $\chi^2_{obs} = 0$, il y a indépendance entre les variables A et B. Si χ^2_{obs} est petit, les effectifs observés sont presque identiques aux effectifs théoriques. Les deux variables sont peu liées entre elles. Si χ^2_{obs} est grand, les effectifs observés sont différents des effectifs théoriques. Les deux variables sont liées entre elles. Afin d'évaluer le degré de relation entre les deux variables qualitatives, divers indices ont été proposés. Une valeur proche de 0 caractérise l'indépendance. Une valeur proche du maximum de l'indice caractérise la liaison fonctionnelle.

1.3 Indices descriptifs

- Le coefficient de contingence de Pearson est $C = \sqrt{\frac{\chi^2_{obs}}{\chi^2_{obs}+n}}$ Le nombre de lignes et de colonnes de la table de contingence détermine la valeur maximale de C. Elle est égale à $\sqrt{\frac{k-1}{k}}$ où $k = \min(p, q)$ et reste toujours inférieure à 1.
- Le coefficient de Tschuprow est $T = \sqrt{\frac{\chi^2_{obs}}{n\sqrt{(p-1)(q-1)}}}$ Il ne peut atteindre 1 que pour les tableaux carrés. Et il n'est comparable que pour des tableaux de même taille.
- Le coefficient de Cramer : $V = \sqrt{\frac{\chi^2_{obs}}{n\ \min(p-1,q-1)}}$ Ce coefficient est le seul qui soit normé (maximum égale à 1) quelle que soit la dimension de la table de contingence.

1.4 Le test du Chi-Deux de Contingence

Le test du Chi-Deux est destiné à décider si la valeur observée est compatible avec la variabilité aléatoire d'un tirage sur deux variables indépendantes. Il est fondé sur la loi multinomiale qui induit la normalité approchée des fréquences observées dans chacune des cases de la table de contingence.

Reprenons encore une fois le raisonnement par simulation, introduit dans :

`http://pbil.univ-lyon1.fr/R/tdr32.pdf`

Supposons que la proportion de campeurs dans l'ensemble des touristes, un jour précis dans une station donnée, soit de $\frac{1}{3}$ et que les touristes soient de trois nationalités, disons Français pour la moitié, Allemands pour $\frac{1}{4}$ et Hollandais pour la même proportion. Si on interroge 100 touristes au hasard on aura en gros une moitié de Français, un quart d'Allemands et un quart de Hollandais. Si le mode de logement est indépendant de la nationalité on aura dans chaque catégorie un tiers en gros de campeurs.

On n'aura jamais exactement 16.67 touristes français campeurs. La probabilité qu'un touriste soit français est de $\frac{1}{2}$. La probabilité qu'un touriste soit campeur est de $\frac{1}{3}$, la probabilité qu'un touriste soit campeur français est de $\frac{1}{6}$.

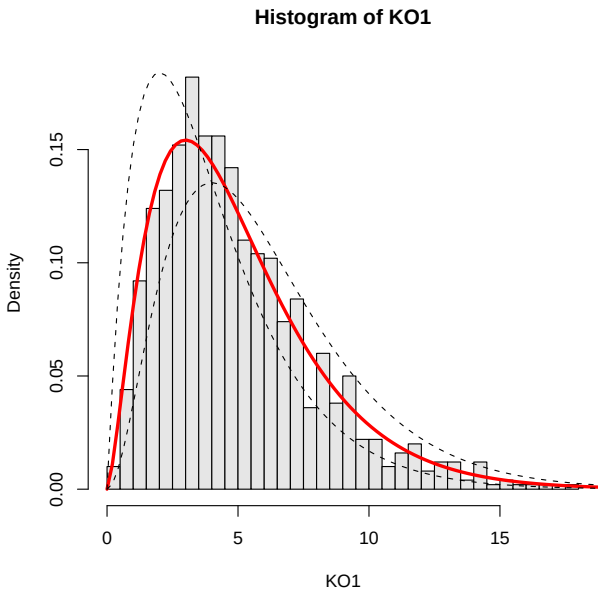
On a une distribution de fréquences multinomiale à 6 catégories FC, FNC, AC, ANC, HC et HNC avec les probabilités $(\frac{1}{6}, \frac{1}{3}, \frac{1}{12}, \frac{1}{6}, \frac{1}{12}, \frac{1}{6})$.

```
table(sample(c("FC", "FNC", "AC", "ANC", "HC", "HNC"), 100, rep = T,
    prob = c(1/6, 1/3, 1/12, 1/6, 1/12, 1/6)))
```

```
AC ANC  FC FNC  HC HNC
 7  24  17  29  10  13
```

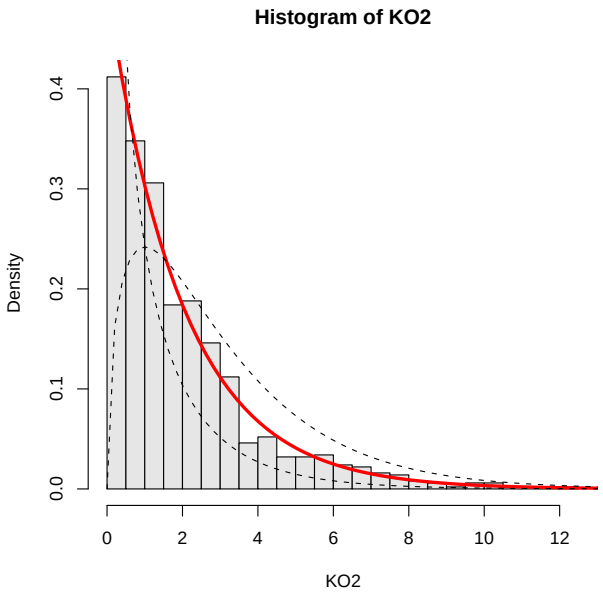
Chacun des effectifs suit une loi binomiale mais ces lois ne sont pas indépendantes car leur somme fait 100 (si une catégorie est bien représentée, une autre l'est forcément moins). La variabilité autour du modèle, inhérente au tirage aléatoire est la *variabilité d'échantillonnage*. L'écart entre l'observation et l'attendu mesuré par le χ^2_{obs} a lui-même une *variabilité d'échantillonnage*.

```
proba <- c(1/6, 1/3, 1/12, 1/6, 1/12, 1/6)
fun1 <- function(k) {
    w <- sample(c("FC", "FNC", "AC", "ANC", "HC", "HNC"), 100, rep = T,
        prob = proba)
    w <- factor(w, levels = c("FC", "FNC", "AC", "ANC", "HC", "HNC"))
    w <- as.numeric(table(w))
}
w <- matrix(sapply(1:1000, fun1), nrow = 6)
KO1 <- as.numeric(apply(w, 2, function(x) sum((x - 100 * proba)^2/100/proba)))
hist(KO1, proba = T, nclass = 30, col = grey(0.9))
x0 <- seq(0, 20, le = 100)
lines(x0, dchisq(x0, df = 5), lwd = 3, col = "red")
lines(x0, dchisq(x0, df = 4), lty = 2)
lines(x0, dchisq(x0, df = 6), lty = 2)
```



Quand tous les paramètres sont connus, l'écart suit une loi χ^2_5. Mais quand on les ignore, si on utilise les marges comme estimation des probabilités, il n'en est rien.

```
KO2 <- as.numeric(apply(w, 2, function(x) chisq.test(matrix(x, nrow = 2))$statistic))
hist(KO2, proba = T, nclass = 30, col = grey(0.9))
lines(x0, dchisq(x0, df = 2), lwd = 3, col = "red")
lines(x0, dchisq(x0, df = 1), lty = 2)
lines(x0, dchisq(x0, df = 3), lty = 2)
```



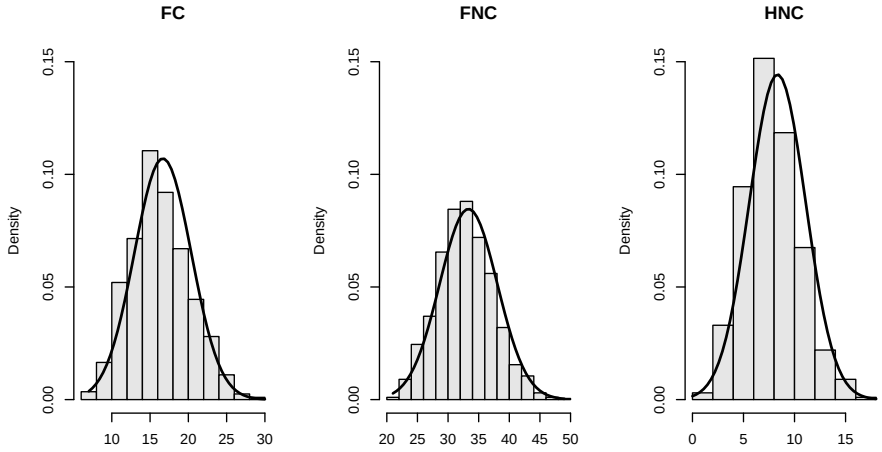
Le test consiste à replacer l'obsevation par rapport à l'ensemble des résultats aléatoires donc à préciser le caractère anormal de l'observation par la *probabilité critique* :

$$P(\chi^2_{(p-1)(q-1))} > \chi^2_{obs})$$

Si celle-ci est trop petite, l'écart entre les données et le modèle est trop grand et l'hypothèse d'indépendance est rejetée. Une littérature ancienne et abondante a longuement discuté des conditions d'utilisation de la loi, sachant qu'on utilise un théorème d'approximation.

Il est logique d'admettre que l'approximation tient si chacune des variables qui permettent de la calculer, donc les effectifs par case, supporte l'approximation normale. Par exemple, ici, c'est très vrai (*justifier les calculs*) :

```
par(mfrow = c(1, 3))
hist(w[1, ], main = "FC", proba = T, col = grey(0.9), ylim = c(0,
    0.15), nclass = 12, xlab = "")
lines(x0 <- seq(min(w[1, ]), max(w[1, ]), le = 50), dnorm(x0, 100/6,
    sqrt(500/36)), lwd = 2)
hist(w[2, ], main = "FNC", proba = T, col = grey(0.9), ylim = c(0,
    0.15), nclass = 12, xlab = "")
lines(x0 <- seq(min(w[2, ]), max(w[2, ]), le = 50), dnorm(x0, 100/3,
    sqrt(200/9)), lwd = 2)
hist(w[5, ], main = "HNC", proba = T, col = grey(0.9), ylim = c(0,
    0.15), nclass = 12, xlab = "")
lines(x0 <- seq(min(w[5, ]), max(w[5, ]), le = 50), dnorm(x0, 50/6,
    sqrt(1100/144)), lwd = 2)
```



Mais évidemment ces variables ne sont pas indépendantes :

```
signif(cor(t(w)), 3)
```

```
       [,1]   [,2]   [,3]   [,4]    [,5]    [,6]
[1,]  1.000 -0.349 -0.105 -0.201 -0.1450 -0.2320
[2,] -0.349  1.000 -0.186 -0.301 -0.1770 -0.3000
[3,] -0.105 -0.186  1.000 -0.120 -0.1600 -0.1700
[4,] -0.201 -0.301 -0.120  1.000 -0.1510 -0.2020
[5,] -0.145 -0.177 -0.160 -0.151  1.0000 -0.0937
[6,] -0.232 -0.300 -0.170 -0.202 -0.0937  1.0000
```

Démontrer que cette matrice de corrélation a une valeur propre nulle.
Toutes les corrélations sont négatives car la somme des variables est constante. C'est encore beaucoup plus sensible si on fixe les marges de la table de contingence. Dans ce cas, les variables dans chaque case sont hypergéométriques et les corrélations sont amplifiées (c'est pourquoi on perd les degrés de liberté) :

```
a <- w[, 1]
campi <- rep(c("C", "NC", "C", "NC", "C", "NC"), a)
natio <- rep(c("F", "F", "A", "A", "H", "H"), a)
table(campi, natio)
```

```
     natio
campi  A  F  H
   C  13 15  5
   NC 17 36 14
```

```
fun2 <- function(k) {
    return(as.numeric(table(campi, sample(natio, 100))))
}
w <- matrix(sapply(1:1000, fun2), nrow = 6)
signif(cor(t(w)), 3)
```

```
       [,1]   [,2]   [,3]   [,4]   [,5]   [,6]
[1,]  1.000 -1.000 -0.664  0.664 -0.351  0.351
[2,] -1.000  1.000  0.664 -0.664  0.351 -0.351
[3,] -0.664  0.664  1.000 -1.000 -0.467  0.467
[4,]  0.664 -0.664 -1.000  1.000  0.467 -0.467
[5,] -0.351  0.351 -0.467  0.467  1.000 -1.000
[6,]  0.351 -0.351  0.467 -0.467 -1.000  1.000
```

Démontrer que cette matrice de corrélation a 4 valeurs propres nulles.
On a dit que pour appliquer un test du Chi-Deux, il faut que l'effectif total n soit grand et que les effectifs **théoriques** soient tous supérieurs à 5. On a dit ensuite qu'il faut que moins de 20% des effectifs théoriques soient inférieurs à 5 mais supérieurs à 1. Les conseils les plus célèbres sont ceux de W.G. Cochran[2].
- si $n < 20$, il faut appliquer le test exact de Fisher.
- si $20 \leq n < 40$ et si les effectifs théoriques sont supérieurs ou égaux à 5, on peut réaliser le test du Chi-Deux avec la correction de continuité de Yates

$$\chi^2_c = \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{\left(\left|n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n}\right| - 0,5\right)^2}{\frac{n_{i\cdot}n_{\cdot j}}{n}}$$

En effet, les effectifs observés varient par saut d'une unité et leur approximation à la loi Normale qui est continue, engendre une sur-évaluation systématique du Chi-Deux.
- si $n \geq 40$ et si les proportions ne sont voisines ni de 0, ni de 1, alors on peut réaliser le test classique du Chi-Deux. Ces conseils n'ont plus qu'un intérêt historique : en cas de doute, on fait 100000 simulations à marges fixées et on aura une idée précise du caractère non aléatoire de l'observation.

2 Exemples

2.1 Enquête sociologique

Lors d'une enquête[1] sur le tourisme et les pratiques de loisir en Ardèche, 2953 personnes ont été interrogées à l'aide d'un questionnaire bilingue français / anglais. 592 d'entre eux proviennent de la zone du moyen Vivarais. Peut-on mettre en évidence une différence d'âge entre campeurs et non campeurs ?

	Campeurs	*Non Campeurs*
16-19 ans	7	4
20-24 ans	43	21
25-29 ans	37	37
30-39 ans	99	78
40-49 ans	79	62
50-54 ans	19	25
55-59 ans	15	15
60-65 ans	9	15
Plus de 65 ans	2	24

Un individu manquant donne $n = 591$ pour une table de contingence avec $p = 2$ et $q = 9$. Saisir les données et faire des calculs élémentaires :

```
noncampeur <- c(4, 21, 37, 78, 62, 25, 15, 15, 24)
campeur <- c(7, 43, 37, 99, 79, 19, 15, 9, 2)
client <- matrix(c(campeur, noncampeur), nrow = 9)
nomligne = c("16-19", "20-24", "25-29", "30-39", "40-49", "50-54",
    "55-59", "60-65", "+65")
nomcol = c("camping", "non_camping")
rownames(client) <- nomligne
colnames(client) <- nomcol
totligne <- apply(client, 1, sum)
totcol <- apply(client, 2, sum)
```

ou encore :

```
rowSums(client)
```

```
16-19 20-24 25-29 30-39 40-49 50-54 55-59 60-65   +65
   11    64    74   177   141    44    30    24    26
```

```
colSums(client)
```

```
    camping non_camping
        310         281
```

ou encore :

```
margin.table(client, 1)
```

```
16-19 20-24 25-29 30-39 40-49 50-54 55-59 60-65   +65
   11    64    74   177   141    44    30    24    26
```

```
margin.table(client, 2)
```

```
    camping non_camping
        310         281
```

Calculer les statistiques descriptives χ^2_{obs} (32.51), C (0.2283), T(0.1395) et V (0.2345). Exécuter le test du Khi2. Éditer les profils lignes et colonnes. Éditer les profils théoriques. Caractériser l'écart entre les données et le modèle. Quelques questions simples mais de fond quand on pratique R :

1. Quel est la nature de l'objet `ttt` ?
2. Quel est le nombre de ses composantes ?
3. Quels sont les noms des composantes ?
4. Que signifie chacune des composantes ?
5. Comment est calculée la composante `residuals` ?
6. Donner la part de chaque classe d'âge dans la constitution du χ^2_{obs}.
7. A partir de quel âge le camping est-il vraiment délaissé ?
8. Représenter le taux de campeurs en fonction de l'âge.
9. Représenter une des colonnes de résidus en fonction de l'âge.
10. Résumer le lien mis en évidence entre âge et mode de résidence des touristes en Ardèche.

Entre faire un test et analyser les données il y a un écart ! Pour approfondir cet aspect, lire un article de fond[5].

Plus difficile :

```
chisq.test(client[-9, ])
```

```
        Pearson's Chi-squared test

data:  client[-9, ]
X-squared = 10.7239, df = 7, p-value = 0.1511
```

Commenter. Pour lever la difficulté, on pourra recourir au modèle linéaire généralisé :

```
x = 1:8
glm0 <- glm(client[-9, ] ~ 1 + x, family = binomial)
summary(glm0)
anova(glm0, test = "Chisq")
```

2.2 Latéralité manuelle

On connaît pour 119 étudiants en Activités Physiques et Sportives (APS) et 88 étudiants en Biologie la main préférentielle d'écriture :

	droite	gauche	total
APS	101	18	119
Biologie	81	7	88
total	182	25	207

Éditer le tableau des profils lignes qui permet de répondre à la question : quelle est, pour chaque filière, la proportion de droitiers et de gauchers ? Éditer le tableau des profils colonnes qui permet de répondre à la question : quelle est la répartition des droitiers (resp. des gauchers) dans les deux filières ? Calculer les statistiques descriptives χ^2_{obs} (1.821), C (0.05543), T(0.03301) et V (0.05552). Lorsqu'on a une table de contingence 2x2, le coefficient de Tschuprow est identique au coefficient de Cramer. Du point de vue descriptif, les coefficients sont proches de 0 et il n'existe pas de lien entre la main d'écriture et la filière.

Au plan inférentiel, le test du Chi2 est ici un test de comparaison de proportion :

```
chisq.test(matrix(c(101, 81, 18, 7), ncol = 2))
```

```
        Pearson's Chi-squared test with Yates' continuity correction

data:  matrix(c(101, 81, 18, 7), ncol = 2)
X-squared = 1.8214, df = 1, p-value = 0.1771
```

```
prop.test(matrix(c(101, 81, 18, 7), ncol = 2))
```

```
        2-sample test for equality of proportions with continuity correction

data:  matrix(c(101, 81, 18, 7), ncol = 2)
X-squared = 1.8214, df = 1, p-value = 0.1771
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.16727504  0.02384494
sample estimates:
   prop 1    prop 2
0.8487395 0.9204545
```

Le test de comparaison de deux proportions se retrouve facilement dans le cas d'une table de contingence 2x2. Dans l'exemple ci-dessus, supposons que l'on désire comparer la proportion de gauchers dans chacune des deux populations d'étudiants. L'hypothèse nulle est l'identité des proportions. La valeur de la statistique du test de comparaison de deux proportions est :

$$z = \frac{f_1 - f_2}{\sqrt{\hat{p}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \text{ avec } \hat{p} = \frac{k_1 + k_2}{n_1 + n_2}$$

avec n_j, k_j, f_j $(j = 1, 2)$ l'effectif total, les fréquences absolue et relative du groupe j.

Analyser les composantes des deux objets. Les deux tests donnent exactement le même résultat. Mais le second est préférable au premier ! POURQUOI ?

2.3 Tennis et badminton : le test exact de Fisher[3]

On considère une table de contingence 2x2. On note A et B les deux variables qualitatives observées sur n individus. Lorsque les effectifs sont trop petits, on transforme l'inconvénient des échantillons de petite taille en bénéfice en énumérant l'ensemble des arrangements possibles des observations puis en calculant les probabilités exactes de chaque arrangement.

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	n

Probabilité du test : La probabilité associée à la table de contingence observée est définie par : $p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$ On calcule ensuite les probabilités des tables de contingence présentant des situations aussi extrêmes que celle observée : p_j avec $j = 1, m$. La probabilité critique finale est : $P = \sum_{j=1}^{m} p_j$

Exemple : on fait passer un test de réactivité visuelle à un groupe de 36 sportifs spécialistes soit de tennis, soit de badminton. Le test se présente de la manière suivante. Deux lampes sont placées à droite et à gauche du sujet. Chaque lampe s'allume de façon aléatoire, avec un temps d'attente variable (entre 0,2s et 0,5s). Le sujet est assis, les mains sur les genoux. Dès qu'une lampe s'allume, il doit frapper une plaque située en dessous de la lampe correspondante. On considère qu'un sujet a réussi le test lorsqu'il a réalisé la bonne association (lumière, frappe) au moins 7 fois sur 10.

Les résultats sont consignés dans la table de contingence ci-dessous.

	Tennis	Badminton	*Total*
Echec	4	2	6
Succès	16	14	30
Total	20	16	36

```
vision <- matrix(c(4, 16, 2, 14), 2, 2)
dimnames(vision) <- list(c("echec", "succes"), c("tennis", "badminton"))
res3 <- chisq.test(vision)
res3$expected
```

```
         tennis badminton
echec   3.333333  2.666667
succes 16.666667 13.333333
```

Remarque : On ne se sert ici du Chi-Deux que pour avoir la table des effectifs théoriques. $n = 36$ et certains effectifs théoriques sont inférieurs à 5. On ne peut pas augmenter la taille de l'échantillon. On applique le test exact de Fisher. Pour ce faire, on construit les tables de contingence des situations encore plus extrêmes que celle observée. On repère l'effectif le plus petit de la table de contingence (échec / badminton : 2) et on fait varier ce dernier jusqu'à 0. On obtient deux possibilités :

	Tennis	Badmington	*Total*
Echec	5	1	6
Succès	15	15	30
Total	20	16	36

	Tennis	Badmington	*Total*
Echec	6	0	6
Succès	14	16	30
Total	20	16	36

```
Hypergeometric              package:stats              R Documentation

The Hypergeometric Distribution

Description:

     Density, distribution function, quantile function and random
     generation for the hypergeometric distribution.

Usage:

     dhyper(x, m, n, k, log = FALSE)
     phyper(q, m, n, k, lower.tail = TRUE, log.p = FALSE)
     qhyper(p, m, n, k, lower.tail = TRUE, log.p = FALSE)
     rhyper(nn, m, n, k)

Arguments:

    x, q: vector of quantiles representing the number of white balls
          drawn without replacement from an urn which contains both
          black and white balls.

       m: the number of white balls in the urn.

       n: the number of black balls in the urn.

       k: the number of balls drawn from the urn.

       p: probability, it must be between 0 and 1.

      nn: number of observations. If 'length(nn) > 1', the length is
          taken to be the number required.
```

Il y a $k = 6$ boules tirées (échec) dans une urne avec $m = 16$ blanches (badminton) et $n = 20$ noires (tennis). il y a $x = 0, 1, 2, 3, 4, 5, 6$ blanches tirées (échec et tennis) et donc la loi complète est :

```
dhyper(0:6, 16, 20, 6)
```

```
[1] 0.019899455 0.127356514 0.298491831 0.327755736 0.177534357 0.044850785
[7] 0.004111322
```

```
sum(dhyper(0:6, 16, 20, 6))
```

```
[1] 1
```

Vérifier en utilisant le formule explicite $p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$. Donner la probabilité critique du test qui est par définition :

```
[1] 0.4457478
```

On a directement ce résultat en utilisant la fonction `fisher.test`. Entrer la matrice 2*2 de manière à ce que la première valeur soit la variable à tester (les 4 tests possibles sont évidemment identiques) et utiliser l'alternative `less` pour savoir si l'observation est trop petite, `greater` pour savoir si elle est trop grande ou le défaut `two side` pour savoir si elle est anormale dans un sens ou dans l'autre. Ici, *a priori* nous n'avons aucune raison de faire une hypothèse alternative précise et cet exercice est purement didactique.

```
fisher.test(matrix(c(2, 14, 4, 16), 2, 2), alt = "less")
```

```

        Fisher's Exact Test for Count Data

data:  matrix(c(2, 14, 4, 16), 2, 2)
p-value = 0.4457
alternative hypothesis: true odds ratio is less than 1
95 percent confidence interval:
 0.000000 3.625182
sample estimates:
odds ratio
 0.5801421
```

```
fisher.test(matrix(c(2, 14, 4, 16), 2, 2), alt = "greater")
```

```

        Fisher's Exact Test for Count Data

data:  matrix(c(2, 14, 4, 16), 2, 2)
p-value = 0.8527
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval:
 0.06840447        Inf
sample estimates:
odds ratio
 0.5801421
```

```
fisher.test(matrix(c(2, 14, 4, 16), 2, 2))
```

```

        Fisher's Exact Test for Count Data

data:  matrix(c(2, 14, 4, 16), 2, 2)
p-value = 0.6722
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.04585961 4.78653951
sample estimates:
odds ratio
 0.5801421
```

Retrouver ces valeurs avec la loi hypergéométrique :

```
sum(dhyper(0:2, 16, 20, 6))
```

```
[1] 0.4457478
```

```
sum(dhyper(2:6, 16, 20, 6))
```

```
[1] 0.852744
```

```
sum(dhyper(0:2, 16, 20, 6)) + sum(dhyper((4:6), 16, 20, 6))
```

```
[1] 0.6722443
```

A retenir : quand on fait un test bilatéral, le cas par défaut, on prend exactement le même nombre de valeurs extrêmes de chaque côté, ici 0-1-2 à gauche et 4-5-6 à droite. Autre exemple :

```
fisher.test(matrix(c(0, 3, 6, 4), 2))
```

```

        Fisher's Exact Test for Count Data

data:  matrix(c(0, 3, 6, 4), 2)
p-value = 0.1923
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.000000 2.577685
sample estimates:
odds ratio
         0
```

```
dhyper(0, 3, 10, 6) + dhyper(3, 3, 10, 6)
```

```
[1] 0.1923077
```

2.4 Avec ou sans bruit : test de McNemar[4]

Soit une variable qualitative binaire observée sur un échantillon de n individus dans deux conditions différentes. On note S et E les deux modalités de la variable qualitative. On obtient une table de contingence 2x2.

	Succès en condition 1	Échec en condition 1
Succès en condition 2	N_{SS}	N_{SE}
Échec en condition 2	N_{ES}	N_{EE}

Sur un tel tableau un χ^2_{obs} n'a aucun sens. Il sera d'autant plus significatif qu'il n'y aura aucune signification expérimentale. Oh le beau cas ! Supposons par exemple qu'un ensemble d'individus soit formé de deux groupes. Le premier réussit l'épreuve imposée quoi qu'il arrive, le second ne connaît que l'échec en toute circonstance. Si on fait l'expérience dans deux conditions on obtient (on suppose en plus qu'il y a eu une erreur de mesure !) :

```
res <- matrix(c(12, 1, 0, 7), 2)
dimnames(res) <- list(c("E", "S"), c("E", "S"))
res
```

```
  E S
E 12 0
S  1 7
```

```
fisher.test(res)$p.value
```

```
[1] 0.0001031992
```

```
chisq.test(res)$p.value
```

```
[1] 0.0003990513
```

```
prop.test(res)$p.value
```

```
[1] 0.0003990513
```

Un test très significatif pour dire que ceux qui sont bons sont bons et que ceux qui sont mauvais sont mauvais ! Bravo. En fait, on est dans un cas extrême de non indépendance des observations, le résultat des deux essais pour le même individu étant fortement corrélé. Mais ce n'est pas la question. On demande si le changement de condition a eu un effet. Le problème consiste à mettre en évidence une différence entre les deux situations. On comprend que l'individu réalisant le même score dans les deux conditions n'apporte rien pour la comparaison elle-même. Ce qui importe, ce sont les effectifs dans SE et ES. L'hypothèse nulle est que les deux conditions sont identiques vis à vis de la variable qualitative étudiée.

Si les deux conditions sont identiques, cela signifie que la répartition des individus entre ES et SE est la même. Ceci nous donne un effectif théorique :

$$\frac{N_{ES} + N_{SE}}{2}$$

Si l'effectif théorique est supérieur à 5, on pense immédiatement au test du Chi-Deux :

$$\chi^2_{obs} = \frac{\left(N_{ES} - \left(\frac{N_{ES}+N_{SE}}{2}\right)\right)^2 + \left(N_{SE} - \left(\frac{N_{ES}+N_{SE}}{2}\right)\right)^2}{\left(\frac{N_{ES}+N_{SE}}{2}\right)}$$

soit

$$\chi^2 = \frac{(N_{ES} - N_{SE})^2}{(N_{ES} + N_{SE})}$$

Avec la correction de continuité de Yates, on a :

$$\chi^2 = \frac{(|N_{ES} - N_{SE}| - 1)^2}{(N_{ES} + N_{SE})}$$

Comme on ignore les individus ayant répondu de la même façon dans les conditions, il n'y a qu'un degré de liberté à cette statistique du Chi-Deux. On inclut généralement directement la correction de continuité de Yates. Exemple : on fait passer un test de réactivité visuelle à un groupe de 118 sujets. Chaque sujet passe le test dans deux conditions différentes : avec ou sans bruit dans la salle. Le test se présente comme suit. Deux lampes sont placées à droite et à gauche du sujet. Chaque lampe s'allume de façon aléatoire, avec un temps d'attente variable (entre 0.2s et 0.5s). Le sujet est assis les mains sur les genoux. Dès qu'une lampe s'allume, il doit frapper une plaque située en dessous de la lampe correspondante. On considère qu'un sujet a réussi le test lorsqu'il a réalisé la bonne association " lumière, frappe " au moins 7 fois sur 10.

	avec Succès	*avec* Échecs
sans Succès	62	26
sans Échecs	7	23

Hypothèse H_0 : Qu'il y ait ou non du bruit dans la salle, les sportifs réagissent de la même manière au test de réactivité visuelle.

```
bruit <- matrix(c(62, 7, 26, 23), nrow = 2)
dimnames(bruit) <- list(c("S", "E"), c("S", "E"))
bruit
```

```
   S  E
S 62 26
E  7 23
```

```
mcnemar.test(bruit)
```

```
        McNemar's Chi-squared test with continuity correction

data:  bruit
McNemar's chi-squared = 9.8182, df = 1, p-value = 0.001728
```

Notons que le test exact est bien simple d'accès. Dans le sens succès avec bruit & échec sans bruit, on observe 7 cas. Dans le sens échec avec bruit & succès sans bruit on observe 26 cas. Évidemment le bruit est néfaste au résultat. Mais supposons que nous ayons beaucoup moins d'observations.

```
bruit <- matrix(c(12, 1, 7, 6), nrow = 2)
dimnames(bruit) <- list(c("S", "E"), c("S", "E"))
bruit
```

```
   S E
S 12 7
E  1 6
```

```
mcnemar.test(bruit)
```

```
        McNemar's Chi-squared test with continuity correction

data:  bruit
McNemar's chi-squared = 3.125, df = 1, p-value = 0.0771
```

Le résultat peut être amélioré en faisant directement le test unilatéral. Nous avons 8 répétitions, $X = 1$ pour succès avec bruit & échec sans bruit et $Y = 7$ pour échec avec bruit & succès sans bruit. La probabilité critique du test unilatéral de l'hypothèse nulle le bruit n'a pas d'effet contre l'alternative le bruit a un effet **négatif** est simplement :

```
sum(dbinom(0:1, 8, 0.5))
```

```
[1] 0.03515625
```

On peut encore l'appeler test de McNemar. Mais on peut aussi retenir le raisonnement qu'on emploiera à sa convenance sans qu'on puisse toujours y mettre un nom.

Pour finir, un exercice célèbre :

- Variante 1 : dans une famille bien élevée de 3 enfants, chacun fait la vaisselle à son tour. Au bout d'un mois, il y a 4 assiettes cassées, 3 par le même et la dernière par un autre. Peut-on dire qu'il existe un enfant moins adroit que les autres ?
- Variante 2 : dans une famille bien élevée de 3 enfants, chacun fait la vaisselle à son tour. Au bout d'un mois, il y a 4 assiettes cassées, 3 par le plus jeune et la dernière par un autre. Peut-on dire que le petit n'est pas encore aussi adroit que les autres ?

Références

[1] P. Chazaud, A-B. Dufour, and B. Vignal. Vers une typologie des campeurs. l'exemple de l'ardèche. *Cahiers Espaces*, 36 :61–68, 1994.

[2] W.G. Cochran. Some methods for strenghening the common chi square tests. *Biometrics*, 10 :417–451, 1954.

[3] R.A. Fisher. The logic of inductive inference. *Journal of the Royal Statistical Society Series A*, 98 :39–54, 1935.

[4] I. McNemar. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12 :153–157, 1947.

[5] N. G. Yoccoz. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bulletin of the Ecological Society of America*, 72 :106–111, 1991.