Examen de biostatistiques
L3 MIV

M. Bailly-Bechet & H. Haned

5 juin 2009

Documents de cours et TD uniquement autorisés. Échanges interdits. Calculatrices inutiles mais autorisées. Durée de l'épreuve : 2h.

Cette épreuve est divisée en deux parties. Il est fortement recommandé de traiter le plus complètement possible une des deux parties, plutôt que de chercher à répondre aux questions de manière éparpillée. Il n'est théoriquement pas possible pour un étudiant de niveau L3 de terminer cet examen en 2h, choisissez-donc une partie et concentrez-vous dessus. Les résultats proposés peuvent être employés même s'ils n'ont pas été démontrés.

1 Quelques développements sur le χ^2 et l'indépendance

1.1 Loi du χ^2

1. Donnez la définition d'une variable du χ^2 à n degrés de liberté.
2. Soit X une variable aléatoire du χ^2 à n degrés de liberté et Y une variable aléatoire du χ^2 à p degrés de liberté. Ces deux variables sont indépendantes. Quelle loi suit la variable aléatoire $X+Y$? Démontrez votre réponse.
3. Rappelez la formule de l'espérance mathématique d'une variable aléatoire X suivant une loi de probabilié $p(X=x)=p(x)$.
4. Démontrez que l'espérance d'une v.a. du χ^2 à 1 degré de liberté vaut 1 (on rappelle la formule de l'intégration par parties : $\int_a^b u'v = [uv]_a^b - \int_a^b uv'$).

5. Que pouvez-vous en déduire sur l'espérance d'une v.a. du χ^2 à n degrés de liberté ? Il n'est pas nécéssaire de démontrer ce résultat, une justification suffira.

1.2 Indépendance

Soient deux variables aléatoires discrètes *non indépendantes*, X et Y. La valeur de Y est dépendante de celle de X ; en effet on a :

$$X = 0 \qquad p(X = 0) = \tfrac{1}{2} \tag{1}$$

$$X = 1 \qquad p(X = 1) = \tfrac{1}{2} \tag{2}$$

$$Y = -1 \quad \begin{cases} p(Y = -1|X = 0) = \frac{2}{3} \\ p(Y = -1|X = 1) = \frac{1}{3} \end{cases} \tag{3}$$

$$Y = 1 \quad \begin{cases} p(Y = 1|X = 0) = \frac{1}{3} \\ p(Y = 1|X = 1) = \frac{2}{3} \end{cases} \tag{4}$$

On rappelle la formule de l'espérance d'une variable aléatoire Y conditionnée à une autre v.a. X :

$$E(Y) = \sum_{X=x} p(x) \left(\sum_{Y=y} y p(y|x) \right) \tag{5}$$

1. Démontrer que $E(X) = \frac{1}{2}$ et $E(Y) = 0$.
2. Écrivez la table des valeurs de la v.a. $X + Y$, c'est à dire l'ensemble de toutes les valeurs possibles de $X + Y$ et la probabilié associée.
3. Calculer $E(X+Y)$. À-t-on l'égalité $E(X+Y) = E(X)+E(Y)$? Celle-ci serait-elle vraie si les variables X et Y étaient indépendantes ?
4. De la même façon, calculer $E(XY)$. À-t-on l'égalité $E(XY) = E(X)E(Y)$? Celle-ci serait-elle vraie si les variables X et Y étaient indépendantes ?

2 Fonctions de répartition et test du maximum

2.1 Statistique du maximum

On va définir une statistique, celle dite du maximum. Celle-ci consiste à regarder la valeur la plus élevée dans un échantillon, et à déterminer la probabilité qu'une telle valeur soit obtenue en tirant n variables avec une loi donnée. Soit X une v.a. continue définie pour toutes les valeurs $x > 0$, de densité de probabilité $p(x)$.

1. Quelle est la probabilité d'obtenir, sur un tirage de la v.a. X, une valeur $x \leq k$?
2. Comment appelle-t-on également cette quantité ? En déduire la probabilité d'obtenir une valeur $x > k$. Ces calculs se font en fonction de $p(x)$, inconnue pour le moment.
3. Si l'on effectue maintenant 2 tirages indépendants de la v.a. X et que l'on ne garde que le maximum, quelle est la probabilité que les deux valeurs tirées x_1 et x_2 soient inférieures à une valeur k donnée ? En déduire que :
$$P\left(max\left(x_1, x_2\right) \leq k\right) = \left(F(k)\right)^2, \tag{6}$$
avec $F(X)$ la fonction de répartition de X.
4. Généralisez cette formule au cas de n tirages, pour calculer $P\left(max\left(x_i, i = 1..n\right) \leq k\right)$.

2.2 Test du maximum

À l'aide de ce qui a été fait dans la partie précédente, on va chercher à caractériser les propriétés d'un test, le test du maximum. Le but de ce test est de comparer la loi de probabilité suivie par la v.a.X à une loi de référence $p_0(x)$ (on pourrait de façon identique comparer leurs fonctions de répartition). Pour cela on procède à un test unilatéral. On tire n valeurs indépendantes d'une variable aléatoire X, et on n'en garde que la valeur maximale x_{max}. On compare la valeur x_{max} à la valeur maximale attendue (suivant la loi de probabilité de référence p_0), avec un risque α, notée z_α, pour tester l'hypothèse :

$$H_0 : \quad \text{X suit la loi de densité } p_0 \tag{7}$$
$$H_1 : \quad \text{X ne suit pas la loi de densité } p_0 \tag{8}$$

Si $x_{max} \leq z_\alpha$, on acceptera H_0 ; sinon on rejettera H_0 au profit de H_1. L'idée intuitive derrière ce test est que si l'on connaît la loi p_0 à laquelle on veut comparer notre échantillon, on peut facilement calculer la probabilité qu'une valeur donnée soit le maximum d'un échantillon, et ainsi réaliser le test. *Par souci de simplicité, on néglige ici la possibilité que x_{max} soit trop petit par rapport à la valeur attendue et que l'on doive rejetter H_0 pour cela.*

1. Expliquez pourquoi z_α est solution de l'équation :
$$(F_0(z_\alpha))^n = 1 - \alpha, \tag{9}$$
avec $F_0(x)$ la fonction de réparition correspondant à la densité de probabilité $p_0(x)$.

On va supposer pour la suite que la loi p_0 est une loi exponentielle de paramètre a. Sa densité de probabilité est $p_0(Z = z) = ae^{-az}$, avec $a > 0$.

2. Si Z suit une loi exponentielle de paramètre a, montrer que l'on a :
$$z_\alpha = \frac{-1}{a} \ln \frac{\alpha}{n}, \tag{10}$$
en employant l'approximation $(1-\alpha)^{\frac{1}{n}} = 1 - \frac{\alpha}{n}$, approximation qui n'est valable que pour α petit.
3. Quelle sera alors la valeur seuil du test comparant un échantillon de 5 valeurs à la loi de probabilité exponentielle de paramètre $a = 2$, au risque 5% ?
4. En conclusion, quels sont, selon vous, les défauts de ce test du maximum ? Quels en sont les avantages pratiques ?