Biostatistiques MIV
TD 2 : Tests

M. Bailly-Bechet

18 mai 2010

1 Autour du maximum – Examen Biostatistiques MIV 2009

1.1 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 :
$$P\left(\max\left(x_1, x_2\right) \leq k\right) = \left(F(k)\right)^2, \tag{1}$$
avec $F(X)$ la fonction de répartition de X.
4. Généralisez cette formule au cas de n tirages, pour calculer $P\left(\max\left(x_i, i = 1..n\right) \leq k\right)$.

1.2 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 :

$$H_0 : \quad \text{X suit la loi de densité } p_0 \tag{2}$$

$$H_1 : \quad \text{X ne suit pas la loi de densité } p_0 \tag{3}$$

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 :

$$(F_0(z_\alpha))^n = 1 - \alpha, \tag{4}$$

   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 :

$$z_\alpha = \frac{-1}{a} \ln \frac{\alpha}{n}, \tag{5}$$

   en employant l'approximation $(1 - \alpha)^{\frac{1}{n}} = 1 - \dfrac{\alpha}{n}$, 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 ?