Statistiques
Master Statistique et econométrie
TD sur les tests - Feuille n^o 1

V. Monbet

Master 1 - 2013-2014

Exercice 1

Dans le contexte de l'évaluation du nombre X d'accidents cérébraux consécutifs au vaccin anticoqueluche (sur 500000 vaccins), on veut déterminer si ce nombre est inférieur ou égal à 5 ou supérieur à 10. On modélise la distribution de X par une loi de Poisson de paramètre λ de sorte qu'on peut poser les hypothèses de test suivantes :

$$H_0 : \lambda \leq 5 \text{ contre } H_1 : \lambda \geq 10$$

1. Calculer le risque α et le manque de puissance β associés à la règle de décision suivante :
   on choisit H_0 si $x \leq 7$ et H_1 si $x \geq 8$ avec x le nombre d'accidents observé.

2. Chaque année 500000 enfants de moins de un an sont vaccinés en France. Une année, on observe 6 accidents. Que concluez-vous? Donner le degré de signification réel du test.

3. Même question pour 12 accidents.

Exercice 2

On se place dans un modèle gaussien où $P_\theta = \mathcal{N}(\mu, \sigma^2)$ et où la variance σ^2 est supposée connue. On veut tester

$$H_0 : \ \mu = \mu_0 \text{ contre } H_1 : \mu = \mu_1$$

avec $\mu_0 \neq \mu_1$. On suppose pour cela que l'on dispose d'un échantillon de taille n de la loi $\mathcal{N}(\mu, \sigma^2)$.

1. Quel est le meilleur test de niveau α pour tester H_0 contre H_1? En quel sens ce test est-il le meilleur? Donner la forme de la région critique.

2. En prenant $\mu_0 = 0$, $\mu_1 = 1$, $\sigma^2 = 1$ et $\alpha = 0.05$, quel doit être la taille de l'échantillon pour que la puissance du test soit égale à 0.99? Quel est l'erreur de type II correspondante?

Exercice 3

Détection d'un signal dans un canal bruité
Une antenne de communication cherche à détecter l'émission d'un signal. A cause des perturbations atmosphériques, même en l'absence de signal, l'antenne perçoit du "bruit". On procède à n mesures. On modélise le bruit par des variables indépendantes identiquement distribuées (vaiid) de loi de Gauss de moyenne 0 et de variance σ^2. Le signal est modélisé par des vaiid de loi de Gauss de moyenne 0 et de variance ξ^2. Le signal et le bruit sont considérés indépendants. On veut tester l'hypothèse H_0 *"pas de signal"* contre l'alternative H_1 *"signal"*.

1. Expliquer pourquoi le test se ramène à la situation suivante. On observe un échantillon de loi $\mathcal{N}(0, s^2)$ avec s inconnu et on teste

$$H_0 : \; s^2 = \sigma^2 \text{ contre } H_1 : \; s = \sigma^2 + \xi^2$$

2. Déterminer la statistique de test du rapport de vraisemblance. En déduire que le test uniformément plus puissant est basé sur une statistique de loi du χ^2. Quel est son nombre de degrés de liberté?

Exercice 4

Loi de Poisson (suite de l'exercice 1)
On considère un échantillon $(X_1, \cdots, X_n)$ tel que X_i suit une loi de Poisson de paramètre λ. X_i modélise par exemple le nombre de complications suite à une vaccination. On souhaite tester l'hypothèse

$$H_0 : \lambda = 7 \text{ contre } H_1 : \lambda \geq 8$$

1. Le test du rapport de vraisemblance est-il uniformément plus puissant?
2. Montrer qu'il est basé sur la statistique de test $S = X_1 + \cdots + X_n$.
3. Pour n petit, expliquer comment se met en place le test (quand rejeter H_0)?
4. Pour n grand, on peut mettre en place un test plus simple grâce au théorème limit central. Comment estimer la puissance de ce test? Comparer cette puissance à celle du test de la question 1 de l'exercice 1.

Exercice 5

Neyman Pearson
Avant le second tour de l'élection présidentielle, un candidat commande un sondage à une société spécialisée pour connaître ses chances d'être élu. Si π est la proportion d'électeurs qui lui sont favorables dans la population, il souhaite résoudre le problème suivant : $\pi = 0.48$ ou bien $\pi = 0.52$?

1. Comment choisir H_0 et H_1 pour prémunir le candidat contre de faux espoirs ?
2. Déterminer la région critique du test UPP de niveau 0.1 lorsque le sondage est réalisé sur 100 personnes. Que penser du résultat ?
   Indication : on utilisera l?approximation normale de la binomiale pour $n = 100$.
3. Indiquer comment varient la région critique et la puissance du test en fonction de n. Calculer la puissance pour $n = 100$ et $n = 500$.
4. Calculer la taille de l'échantillon pour que la puissance soit supérieure à 0.90. Quelle est alors la région critique ?
5. Le candidat veut désormais savoir si $\pi \leq 0.5$ ou bien si $\pi > 0.5$. Proposer un test UPP, toujours en prémunissant le candidat contre de faux espoirs.

Exercice 6

On considère un échantillon i.i.d. de taille n, noté $X_1, \cdots, X_n$, de loi $\mathcal{N}(\mu, \sigma^2)$. On cherche à tester $H_0 : \mu \leq \mu_0$ contre $H_1 : \mu > \mu_0$.

1. On suppose que la variance σ^2 est connue. Existe-t-il un test UPP de niveau α ? Donner la forme de ce test.
   Application - Pour un plein de carburant, une voiture en version économique a une autonomie de 900 km en moyenne avec un écart-type de 50 km. Le constructeur prétend qu'un aileron dynamique peut augmenter cette autonomie. Une association de consommateurs veut vérifier cette affirmation et teste l?efficacité de l?aileron sur un échantillon de 20 véhicules. Elle observe une autonomie moyenne de 925 km (par véhicule). Conclure.
2. On suppose que σ^2 est inconnue. Rappeler la loi de la statistique
   $$T(X_1, \cdots, X_n) = \frac{\sqrt{n}}{\bar{X}_n - \mu} S_n$$
   avec
   $$S_n{}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$
   En s'inspirant de la première question, proposer un test de H_0 contre H_1. *Application* - Il est prouvé qu'une exposition à un volume sonore supérieur à 90 décibels peut provoquer des lésions auditives irréversibles. Un gérant de boite de nuit veut savoir si son établissement est dangereux de ce point de vue. Il fait procéder à 65 mesures, et constate que le volume sonore moyen mesuré est de 86 décibels avec un écart-type de 3 décibels. On propose un test de niveau 0.005 permettant à cette personne de décider si elle doit modifier le volume sonore de son établissement (on suppose la normalité de ce volume sonore). Pourquoi ce niveau est-il inhabituel, et pourquoi est-ce justifié ici? Quel est le résultat du test ?

Exercice 7

(Contrôle continu 2012-13)

Un sondage effectué sur 4000 personnes donne, pour un référendum, 1800 intentions de vote positif (oui). Les personnes interrogées devaient répondre oui ou non. Peut-on en conclure que le oui l'"emportera". Soit π le taux de "oui". On voudrait tester $H_0 : \pi \leq 1/2$ contre $H_1 : \pi > 1/2$.

1. La variable d'intérêt X suit une loi de Bernouilli de paramètre π . Justifier cette assertion.

2. On dispose de $n = 1800$ observations $x_1, \cdots, x_n$ de X. Montrer que l'estimateur du maximum de vraisemblance de π est la moyenne empirique.

3. Proposer un test du rapport de vraisemblance pour tester $H_0 : \pi \leq 1/2$ contre $H_1 : \pi > 1/2$ au niveau $\alpha = 5\%$.

4. Quelles sont les propriétés du test de la question précédente? Justifier la réponse.

5. Réaliser le test et conclure.

Exercice 8

(Contrôle continu 2012-13) Pour avoir la certification "bio", un fabriquant de produits "bios" doit garantir pour chaque lot un pourcentage d'OGM inférieur à 1%. Il prélève donc n=25 produits par lot et teste si le pourcentage d'OGM est inférieur à 1%. On note X_i le logarithme du pourcentage d'OGM du paquet numéro i.

Modèle : On suppose que les X_i sont indépendants et suivent une loi gaussienne de moyenne θ et de variance 1.

1. Pour $\theta_1 > \theta_0$, montrez que le test de Neyman-Pearson de niveau α de $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ est de la forme $\bar{X}_n > t_{n,\alpha}$ avec $\bar{X}_n$ la moyenne empirique.

2. Pour le fabricant, le pourcentage d'OGM est inférieur à 1% sauf preuve du contraire. Il veut tester l'hypothèse $H_0 : \theta \leq 0$ contre $H_1 : \theta > 0$ et il souhaite que pour $\theta \leq 0$ le test se trompe avec une probabilité inférieure à 5%. Calculer un seuil $t_{25,5}$ tel que

$$\sup_{\theta \leq 0} P_\theta(\bar{X}_{25} > t_{25,5}) = 5\%.$$

   On pourra utiliser que $P(Z > 1.645) \simeq 5\%$ si Z suit une loi de Gauss standard.

3. Une association "anti-OGM" veut s'assurer qu'il n'y a effectivement pas plus de 1% d'OGM dans les produits labélisés "bio". En particulier, elle s'inquiète de savoir si le test parvient à éliminer les produits pour lesquels le pourcentage d'OGM dépasse de 50% du maximum autorisé. Quelle est la probabilité que le test ne rejette pas H_0 lorsque le pourcentage d'OGM est de 1,5%?

4. Scandalisée par le résultat précédent, l'association milite pour que le test du fabriquant prouve effectivement que le pourcentage d'OGM est inférieur à 1%. Pour elle, le pourcentage d'OGM est supérieur à 1% sauf preuve du contraire, donc H_0 est $\theta > 0$ et H_1 est $\theta < 0$. Proposez un test de H_0 contre H_1 tel que a probabilité que le test rejette à tort H_0 soit inférieure à 5%.

Exercice 9

Dans une population végétale on émet l'hypothèse d?une sous-population (en proportion $1-\lambda$) qui créerait un phénomène de mélange. On modélise ainsi la taille de cette espèce par une loi de densité (par rapport à la mesure de Lebesgue sur $\mathbb{R}$)

$$f(x;\lambda) = \lambda \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right) + (1-\lambda)\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-1)^2}{2\sigma^2}\right)$$

avec σ connu et $\lambda \in [0,1]$ inconnue.

1. Interpréter f.
2. Proposer un estimateur de λ par la méthode des moments. Quelles sont les propriétés de cet estimateur?
3. Proposer un test de niveau 5% pour tester l'absence de la sous-population.

Exercice 10

(Contrôle terminal 2012-13) Une vaste enquête au sein d'un central téléphonique a permis de déterminer que le nombre d'appels reçus durant une seconde suit une loi de Poisson de paramètre 4.
Pour un second central téléphonique, on effectue une enquête de moindre envergure afin de vérifier si le nombre d'appels reçus par seconde suit la même loi. On comptabilise lors de 200 secondes, le nombre d'appels par seconde, ce qui produit les résultats suivants :

Nombre d'appels par seconde	0	1	2	3	4	5	6	7	8	9	10	11
Effectifs observés	6	15	40	42	37	30	10	9	5	3	2	1

On note N_i le nombre d'appels reçus entre les secondes i et $i+1$ et on suppose les N_i indépendantes. Enfin on note $N_i' = \min(N_i, 8)$.

1. On note Q la loi de $\min(N, 8)$ où N suit une loi de Poisson de paramètre 4. Tester l'aqéquation des N_i' observés à cette loi Q.
2. On suppose que les temps entre deux appels consécutifs suivent une loi exponentielle de paramètre λ. Dans ce cas, les N_i sont indépendants et suivent une loi de Poisson de paramètre λ. Proposez un test de niveau asymptotique 5% de $\lambda = 4$ contre $\lambda \neq 4$.
3. Conclure