

Examen de Statistique du premier février 2010

DUREE 2 HEURES

Documents et calculatrices interdits.

Indications importantes concernant la rédaction Le sujet doit être lu attentivement et en entier. Les réponses (notamment aux questions de cours) devront être correctement rédigées, précises, concises, pertinentes et **lisibles**.

Questions de cours (chaque réponse doit être rédigée en deux lignes au maximum)

Q1 Soit T une statistique et θ un paramètre inconnu. Que signifie que T est un estimateur sans biais de θ ?

Q2 Soit $X_1, \dots, X_n$, n variables indépendantes de loi $\mathbf{N}(m, \sigma^2)$. On pose

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}, \quad \hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Quelle sont les lois des statistiques suivantes : $\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$, $\frac{\sqrt{n}(\bar{X}_n - m)}{\hat{\sigma}_n}$.

Q3 Soit θ un paramètre inconnu et U_n et V_n deux statistiques. Que signifie que $[U_n, V_n]$ est un intervalle de confiance pour θ de niveau 95% ?

Q4 Soit T_n une statistique à valeur dans un ensemble E et soit $\mathcal{R}$ une partie de E . Que signifie que $\mathcal{R}$ est une région de rejet de l'hypothèse H_0 de seuil 5% ?

Q5 Soit t la valeur observée de la statistique de test d'une hypothèse H_0 . Comment s'appelle le plus grand seuil de risque auquel on peut accepter H_0 ?

Exercice 1. On observe un échantillon de 290 pingouins et 285 manchots ayant fait les mêmes études et classés par tranches de salaire. On veut savoir si à qualifications égales, le salaire dépend de l'espèce.

	1000-2000	2000-3000	3000-4000	4000-5000
Pingouins	50	70	110	60
Manchots	60	75	100	50

On utilisera les notations suivantes : soit $n_{1,j}$ et $n_{2,j}$ les nombre respectifs de pingouins et de manchots de la tranche de salaire j , $1 \leq j \leq 4$, n_1 le nombre total de pingouins, n_2 le nombre total de manchots et $n = n_1 + n_2$.

- Donner le nom du test à appliquer.
- Définir l'hypothèse nulle H_0 qui convient.
- Pour $j = 1, \dots, 4$, calculer n_{+j} , le nombre total d'individus de la catégorie de salaire j . Calculer $\sum_{j=1}^4 n_{+j}$.
- Donner la définition de la statistique de test T_n . (Utiliser les notations indiquées)
- Quelle est sa loi approximative sous H_0 ? Quelle est la région de rejet de H_0 au seuil de risque 5% ?
- La valeur observée de T_n est $t = 2.37$. Peut-on accepter au seuil 5% l'hypothèse que le salaire est indépendant de l'espèce ?
- Quel est le plus grand seuil auquel on peut accepter cette hypothèse ?

Exercice 2. On considère un échantillon $(X_1, X_2, \dots, X_n)$ de loi normale $\mathbf{N}(m, 1)$ avec m inconnue. On rappelle que la densité ϕ_m de la loi $\mathbf{N}(m, 1)$ est donnée par :

$$\phi_m(x) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -(x - m)^2 / 2 \right\} .$$

- (i) Déterminer la fonction de vraisemblance V de l'échantillon. (Préciser sur quel ensemble elle est définie).
- (ii) Calculer la log-vraisemblance L et ses dérivées première et seconde par rapport à m .
- (iii) Calculer l'information de Fisher $I_n(m)$.
- (iv) Déterminer l'estimateur du maximum de vraisemblance de m .
- (v) Cet estimateur est-il sans biais ?
- (vi) Cet estimateur est-il convergent ?
- (vii) Cet estimateur est-il efficace ?
- (viii) Quelle est la loi de la moyenne empirique $\bar{X}_n$?
- (ix) Donner un intervalle de confiance pour m de niveau 95%, pour $n = 16$ et la valeur observée $\bar{x}_n = 5$ de $\bar{X}_n$.

Exercice 3. Les notes d'examen de deux groupes d'étudiants sont modélisées par des variables gaussiennes indépendantes d'espérances et de variances respectives m_1, m_2, σ_1^2 et σ_2^2 . Le groupe 1 a $n_1 = 21$ étudiants et le groupe 2 a $n_2 = 31$ étudiants. Pour chaque groupe on mesure la moyenne empirique et la variance empirique (sans biais) des notes.

	$\bar{x}_{n,i}$	$\hat{\sigma}_{n,i}^2$
groupe 1	14.12	3.52
groupe 2	13.25	3.43

On veut tester l'hypothèse que les deux groupes ont le même niveau moyen.

- (i) On teste tout d'abord l'égalité des variances. Quelle est la statistique de test et quelle est sa loi sous l'hypothèse nulle ? Quelle est la région de rejet de H_0 au seuil de risque 5% ?
- (ii) La valeur de la statistique de test est 1.05. L'hypothèse d'égalité des variances peut-elle être acceptée au seuil 5% ?
- (iii) Sous l'hypothèse d'égalité des variances, donner en fonction de $\hat{\sigma}_{n,1}^2$ et $\hat{\sigma}_{n,2}^2$ un nouvel estimateur $\tilde{\sigma}_n^2$ sans biais de la variance commune. (Justifier la réponse)
- (iv) On teste maintenant l'égalité des moyennes. Donner l'expression de la statistique de test T_n à utiliser. Quelle est sa loi sous H_0 ? Quelle approximation peut-on utiliser ? Quelle est la région de rejet de H_0 au seuil de risque 5% ?
- (v) La valeur observée de la statistique de test est $t = 1.65$. L'hypothèse d'égalité des moyennes peut-elle être acceptée au seuil 5% ?
- (vi) Quel est le plus grand seuil auquel on accepte l'hypothèse d'égalité des moyennes ?

Valeurs numériques

- Soit Φ la fonction de répartition de la loi normale centrée réduite.

$$\Phi(1.65) = 95\% , \quad \Phi(1.96) = 97.5\% .$$

- Soit Q_ν la fonction de répartition de la loi du χ^2 à ν degrés de liberté.

$$Q_3(2.37) = 50\% , \quad Q_3(7.81) = 95\% .$$

- Soit $F_{n,m}$ la fonction de répartition de la loi de Fisher à n et m degrés de liberté.

$$F_{20,30}(0.43) = 2.5\% , \quad F_{20,30}(2.19) = 97.5\% .$$