

Université d'Orléans - Licence Economie et Gestion

Statistique Mathématique

C. Hurlin. Examen Mai 2007

Pour chaque question répondez par VRAI ou FAUX et justifiez précisément votre réponse, en développant si nécessaire les calculs. **Toute réponse non justifiée ou mal justifiée sera comptée comme fausse.**

Question 1 (3 points) Un institut d'études affirme, avec un seuil de risque de 5%, que la crise de la presse écrite française est liée à la structure par âge de la population. Pour cela, il vous communique les résultats suivants d'un sondage sur 500 personnes portant sur la lecture de la presse (variable X) et l'âge des individus (variable Y).

X \ Y	- de 25 ans	25-50 ans	50 ans et +	Total
rare	30	50	20	100
occasionnelle	75	115	50	240
fréquente	50	50	60	160
Total	155	215	130	500

Question 2 (3 points) On considère un échantillon de 100 notes d'examen comprises entre 1 et 10 dont la répartition est donnée dans le tableau suivant. On vous affirme que pour un risque de première espèce de 5%, on est sûr que ces notes ont été distribuées au hasard.

Valeur de X	1	2	3	4	5	6	7	8	9	10
Effectif Empirique	10	9	8	8	14	10	11	9	9	12

Question 3 (2 points) On cherche à tester une proportion p en utilisant un test d'hypothèses simples :

$$H_0 : p = 0.10 \quad H_1 : p = 0.30 \tag{1}$$

On considère un échantillon Z de taille N et l'on suppose que la fréquence d'échantillon f_N est une réalisation de la variable aléatoire F_N telle que:

$$F_N \sim N\left(p, \frac{p(1-p)}{N}\right) \tag{2}$$

Dans ce cas, la région critique associée obtenue par la méthode de Neyman-Pearson est de la forme :

$$W = \{Z | F_n > A\} \tag{3}$$

où A est le seuil critique pour un niveau de risque de première espèce de $\alpha\%$.

Question 4 (2 points) Soit X une v.a.r. qui suit une loi normale de moyenne m inconnue et de variance σ^2 elle aussi inconnue. On note S_N^2 la variance empirique corrigée associée à cet échantillon telle que :

$$S_N^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X}_N)^2 \tag{4}$$

où $\bar{X}_N = (1/N) \sum_{i=1}^N X_i$ désigne la moyenne empirique. On cherche à tester la moyenne par un test unilatéral de la forme :

$$H_0 : m = 0 \quad (5)$$

$$H_1 : m > 0 \quad (6)$$

La région critique de ce test est de la forme :

$$W = \left\{ X_1, \dots, X_N \mid \bar{X}_N > F^{-1}(1 - \alpha) \frac{S_N}{\sqrt{N}} \right\} \quad (7)$$

où K est une constante déterminée par le niveau de risque de première espèce égal à $\alpha\%$ et où $F(\cdot)$ désigne la fonction de répartition de la loi de Student à $N - 1$ degrés de liberté.

Question 5 (2 points) Soit X une v.a.r. qui suit une loi normale de moyenne m inconnue et de variance σ^2 elle aussi inconnue. On note S_N^2 la variance empirique corrigée associée à cet échantillon telle que :

$$S_N^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X}_N)^2 \quad (8)$$

où $\bar{X}_N = (1/N) \sum_{i=1}^N X_i$ désigne la moyenne empirique. On considère un échantillon de taille $N = 16$ tel que $S_N^2 = 4$ et $\bar{X}_N = 1.5$. L'intervalle de confiance sur la vraie valeur de m pour un niveau de risque de 5% est défini par les bornes :

$$\Pr [0.4343 < m < 2.5657] = 0.95 \quad (9)$$

Question 6 (2 points) On considère un échantillon de N réalisations $\{y_1, \dots, y_N\}$ d'une variable aléatoire normale $N(m, \sigma^2)$. L'estimateur du Maximum de Vraisemblance de m correspond à la moyenne empirique :

$$\hat{m} = \frac{1}{N} \sum_{i=1}^N y_i \quad (10)$$

Question 7 (2 points) On désire estimer la variance σ^2 d'une variable aléatoire normale X à partir d'un N échantillon de tirages indépendants $\{X_1, \dots, X_N\}$. On suppose que la moyenne $m = 50$ est connue et que $N = 101$. On suppose enfin que :

$$\sum_{i=1}^{20} (X_i - \bar{X})^2 = 900 \quad (11)$$

Un intervalle de confiance à 95% sur la variance de la variable X est égal à :

$$IC = [6.94; 12, 12] \quad (12)$$

Question 8 (2 points) On considère un test sur la moyenne m d'une variable aléatoire normale X de variance connue, avec $\sigma^2 = 36$. Plus précisément, à partir d'un échantillon X de taille $N = 100$, on cherche à tester l'hypothèse :

$$H_0 : m = 100 \quad H_1 : m > 100 \quad (13)$$

La puissance du test UPP associé à un niveau de risque de première espèce de 5%, notée $P(m)$, vaut en fonction des valeurs de m :

$$P(100) = 0.05 \quad (14)$$

$$P(120) = 0.29 \quad (15)$$

$$P(140) = 0.70 \quad (16)$$

Question 9 (3 points) Soit une variable aléatoire suivant une loi continue de densité $f(t)$ telle que :

$$f(t) = \frac{1}{a} e^{-\frac{t}{a}} \quad \forall t > 0, a > 0 \quad (17)$$

où a est un paramètre inconnu que l'on veut estimer à l'aide d'observations indépendantes t_1, t_2, \dots, t_n . On admet que :

$$E(T) = a \quad V(T) = a^2 \quad (18)$$

La moyenne empirique $\hat{a} = \frac{1}{n} \sum_{i=1}^n t_i$ est un estimateur efficace de a .