Partiel du module RFIDEC

C. Gonzales

Durée : 2 heures

Seuls documents autorisés :

Seuls documents autorisés: les transparents de cours. Calculatrices autorisées.

Exercice 1 (2 pts)

Soit un échantillon de taille 10 :

5	1	9	2	4	2	8	7	3	9

.

Q 1.1 Déterminez la médiane de cet échantillon.

On commence par retrier l'échantillon par ordre croissant :

1	2	2	3	4	5	7	8	9	9

.
Il y a donc ici deux médianes : 4 et 5.

Q 1.2 Estimez, sans biais, la variance de la population d'où est issu l'échantillon.

Pour estimer la variance de la population, il faut déterminer la moyenne de l'échantillon : $\mu = 50/10 = 5$. Ensuite, on calcule la variance corrigée : $\sigma^2 = \frac{n}{n-1}S^2$, où S^2 est la variance de l'échantillon. $S^2 = 84/10$. On en déduit donc l'estimation de σ^2 : $84/9 \approx 9,33$.

Exercice 2 (4 pts)

Soit une urne contenant des boules de 5 couleurs différentes : (R)ouges, (B)leues, (V)ertes, (J)aunes, (N)oires. On suspecte que la distribution de probabilité sur les couleurs des boules de l'urne est la suivante :

$$P(R) = 0,2 \quad P(B) = 0,4 \quad P(V) = 0,1 \quad P(J) = 0,2 \quad P(N) = 0,1.$$

Par ailleurs, on a tiré un échantillon i.i.d. de 20 boules et on a noté le nombre de boules de chaque couleur :

Couleur	R	B	V	J	N
Nb boules	2	9	4	5	0

Faites un test d'ajustement avec un niveau de confiance $1 - \alpha = 90\%$ pour déterminer si, oui ou non, la distribution de probabilité sur les couleurs des boules est celle indiquée ci-dessus.

On calcule le nombre de boules que l'on aurait du tirer théoriquement dans l'urne :

Couleur	R	B	V	J	N
Proba	$0,2$	$0,4$	$0,1$	$0,2$	$0,1$
Nb boules	4	8	2	4	2

Il suffit maintenant de calculer la statistique d'ajustement :

$$D^2 = \frac{(2-4)^2}{4} + \frac{(9-8)^2}{8} + \frac{(4-2)^2}{2} + \frac{(5-4)^2}{4} + \frac{(0-2)^2}{2} = \frac{43}{8} = 5.375.$$

Or nous savons que $D^2 \sim \chi^2_4$. Suivant cette loi, et pour $\alpha = 10\%$, d_4^α, le seuil au delà duquel on peut considérer qu'il y a divergence entre la théorie et l'observation, est égal à $7,78$. Puisque $D^2 < d_4^\alpha$. Donc la loi de probabilité de l'énoncé semble être celle dont est tiré l'échantillon.

Exercice 3 (8 points)

On sait, par expérience, que les notes de partiel de RFIDEC suivent une loi normale $\mathcal{N}(\mu; 6^2)$. On considère l'échantillon de notes i.i.d. suivant :

10	8	13	20	12	14	9	7	15

.

Q 3.1 Estimez la moyenne de la population des notes de partiel de RFIDEC grâce à un intervalle de confiance de niveau de confiance $1 - \alpha = 95\%$.

On sait, grâce au théorème 1 du cours 2 que $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0;1)$. On cherche ici le seuil c tel que $P(-c \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq c) = 1 - \alpha$. Par symétrie de la loi normale centrée réduite, cela revient à chercher c tel que $P(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \geq c) = \alpha/2 = 0,025$. Or, d'après la loi normale, $c = 1,96$. Par conséquent, $P(-1,96 \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq 1,96) = 0,95$, ou encore $P(\overline{X} - 1,96 \times 6/3 \leq \mu \leq \overline{X} - 1,96 \times 6/3)$. L'intervalle de confiance est donc $[\overline{X} - 3,92; \overline{X} + 3,92]$. La moyenne de l'échantillon, $\overline{x}$, est égale à 12. Par conséquent l'estimation de μ par intervalle de confiance est égale à $[8,08; 15,92]$.

Q 3.2 Par expérience, les années précédentes, la moyenne au partiel de RFIDEC était égale à 14. Dressez un test d'hypothèse de niveau de confiance $1-\alpha = 95\%$ pour confronter les hypothèses $H_0 =$ « la moyenne est égale à 14 » et $H_1 =$ « la moyenne a baissé, i.e., elle est inférieure à 14 ».

On doit déterminer un seuil c tel que si $\overline{x} < c$, alors H_1 est plus probable que H_0 (région de rejet). Sous l'hypothèse H_0, on sait que $\frac{\overline{X}-14}{\sigma/\sqrt{n}} = \frac{\overline{X}-14}{2} \sim \mathcal{N}(0;1)$. On cherche donc c tel que $P(\frac{\overline{X}-14}{2} < \frac{c-14}{2} | \frac{\overline{X}-14}{2} \sim \mathcal{N}(0;1)) = 0,05$. D'après la loi normale, $\frac{c-14}{2} \approx -1,645$, ou encore $c = 10,71$. Donc, la règle de décision du test d'hypothèse est : rejeter H_0 si $\overline{x} < 10,71$. Or, $\overline{x} = 12$. Par conséquent, on ne peut en déduire, à ce niveau de confiance, que la moyenne a baissé.

Q 3.3 Calculez la puissance du test pour une moyenne de 12 (H_1 : la moyenne est égale à 12).

Puissance du test $= 1-\beta(12) = P(\text{rejeter } H-0|H_1) = P(\overline{X} < 10,71 | \frac{\overline{X}-12}{2} \sim \mathcal{N}(0;1))$
$= P(\frac{\overline{X}-12}{2} < -0,645 | \frac{\overline{X}-12}{2} \sim \mathcal{N}(0;1)) \approx 25,95\%$.

Exercice 4 (6 points)

Soit X une variable aléatoire définie sur l'ensemble des nombres entiers positifs. X suit la loi géométrique de paramètre $p \in [0, 1]$ si $P(X = n) = (1 - p)^{n-1}p$. On a observé 5 réalisations (obtenues indépendamment les unes des autres) d'une variable X suivant la loi géométrique : $\boxed{4} \boxed{2} \boxed{6} \boxed{5} \boxed{8}$.

Q 4.1 Estimez par maximum de vraisemblance la valeur du paramètre $\theta = p$ de la loi.

La vraisemblance est égale à $L(\mathbf{x}, \theta) = \prod_{i=1}^{5} P(x_i|\theta)$, où θ est notre estimation de p. En les $P(x_i|\theta)$ par leur valeurs, on obtient que $L(\mathbf{x}, \theta) = (1 - \theta)^{20}\theta^5$. Par conséquent, la log-vraisemblance est égale à $\ln L(\mathbf{x}, \theta) = 20 \ln(1-\theta)+5 \ln \theta$. Donc $\frac{\partial \ln L(\mathbf{x},\theta)}{\partial \theta} = \frac{-20}{1-\theta}+\frac{5}{\theta} = \frac{5-25\theta}{p(1-\theta)}$. La dérivée de la log-vraisemblance est donc positive jusqu'à $\theta = 0,2$ et négative ensuite. Le maximum de vraisemblance est donc atteint pour $p = 0,2$.

Q 4.2 Avant le tirage de l'échantillon, nous avions une connaissance a priori sur le paramètre θ : ce dernier suivait a priori une loi Beta de paramètres 4 et 5, autrement dit $\pi(\theta) \propto \theta^3(1 - \theta)^4$. Estimez la valeur du paramètre $\theta = p$ par maximum a posteriori.

On doit calculer $\text{Argmax}_\theta \pi(\theta|\mathbf{x}) = \text{Argmax}_\theta \pi(\mathbf{x}|\theta)\pi(\theta) = \text{Argmax}_\theta (1-\theta)^{24}\theta^8$. En passant au log, on obtient $\text{Argmax}_\theta 24\ln(1-\theta) + 8\ln\theta$. La dérivée par rapport à θ s'annule lorsque $32\theta = 8$. Autrement dit, l'estimation par maximum a posteriori de p est $0,25$.