

Contrôle continu : Statistique

Patrick Tardivel
Université de Bourgogne

14/03/2023

Calculatrice autorisée, note de cours interdite, durée : 2h. Le formulaire est au verso.

Exercice 1 (Questions de cours).

1. Soit (X_1, \dots, X_n) un n -échantillon d'une loi P paramétrée par $\theta \in \Theta \subset \mathbb{R}$ et $T = g(X_1, \dots, X_n)$ une statistique. On rappelle que l'erreur quadratique moyenne de la statistique T pour θ vaut

$$\text{EQM}_\theta(T) = \mathbb{E}((T - \theta)^2).$$

Rappeler l'expression du biais de la statistique T pour θ noté $B_\theta(T)$ puis montrer que

$$\text{EQM}_\theta(T) = \text{var}(T) + B_\theta(T)^2.$$

2. Soit $X = \{(x_1, p_1), \dots, (x_n, p_n)\}$ une série statistique pondérée. Rappeler la formule de la moyenne pondérée \bar{X} . La variance pondérée de la série X vaut $v(X) = \sum_{i=1}^n p_i(x_i - \bar{X})^2$. Montrer que $v(X) = \overline{X^2} - \bar{X}^2$.

Exercice 2. Neuf étudiants ont passé l'agrégation de mathématiques et ont obtenu les notes suivantes, sur 20, aux épreuves écrites :

X =Note en mathématiques générales	8	14	11	6	14	17	7	4	6
Y =Note en analyse et probabilités	5	11	13	10	10	14	9	9	8

1. Déterminer la droite des moindres carrés de Y en fonction de X .
2. Calculer le coefficient de corrélation linéaire $r(X, Y)$.
3. Prédire la note moyenne d'un étudiant ayant obtenu 12/20 en mathématiques générales.

Exercice 3. On considère X une variable aléatoire suivant une loi de Pareto de paramètre $\theta > 0$ ayant pour densité :

$$\forall x \in \mathbb{R} \quad f_\theta(x) = \frac{\theta}{x^{\theta+1}} \mathbf{1}_{[1, +\infty[}(x)$$

1. Pour $t \geq 1$ calculer $\int_1^t f_\theta(x) dx$.
 - (a) En déduire que f_θ est une densité.
 - (b) Donner l'expression explicite de la médiane (c'est-à-dire du quantile d'ordre 0,5) de la variable aléatoire X en fonction de $\theta > 0$.
2. Calculer l'estimateur du maximum de vraisemblance $\hat{\theta}^{\text{mv}}$ de θ .
3. On considère l'échantillon expérimental suivant :

1,244 1,073 1,001 1,366 2,225 1,624.

Calculer la valeur expérimentale de l'estimateur du maximum de vraisemblance de θ .

4. En utilisant la valeur expérimentale de l'estimateur du maximum de vraisemblance de θ donner :
 - (a) une estimation de la probabilité que X soit inférieure à 1,2.
 - (b) une estimation de la médiane de X .

Formulaire

Droite des moindres carrés et coefficient corrélation linéaire

Soit $(X, Y) = \{(x_1, y_1), p_1), \dots, ((x_n, y_n), p_n)\}$ une série statistique double pondérée. La droite des moindres carrés pondérés $y = a^*x + b^*$ a pour coefficients

$$a^* = \frac{\text{cov}(X, Y)}{v(X)} \text{ et } b^* = \bar{Y} - a^*\bar{X}.$$

Le coefficient de corrélation linéaire vaut

$$r(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{v(X)}\sqrt{v(Y)}}.$$