Université du Littoral Côte d'Opale
Probabilités et statistiques
Master 1 TSI
Année 2022-2023

Fiche n°3. Estimation.

1. Estimation ponctuelle

Exercice 1. Le staff médical d'une grande entreprise fait ses petites statistiques sur le taux de cholestérol de ses employés. Les observations faites sur 100 employés tirés au sort sont les suivantes :

Taux de cholestérol (en cg)	Effectif d'employés
120	9
160	22
200	25
240	21
280	16
320	7

Donner des estimations ponctuelles de la moyenne et de l'écart-type pour le taux de cholestérol dans toute l'entreprise, en précisant quelles hypothèses sont faites.

Exercice 2. Soit $X_1, \ldots, X_n$ un n-échantillon de loi P_θ, avec $\theta \in \Theta$. Estimer, par maximum de vraisemblance, le paramètre θ dans les quatre cas suivants :

(1) P_θ est une loi de Poisson de paramètre $\theta \in]0, \infty[$;
(2) P_θ est une loi géométrique de paramètre $\theta \in]0, 1[$;
(3) P_θ est une loi exponentielle de paramètre $\theta \in]0, \infty[$;
(4) P_θ est la loi uniforme sur $[0, \theta]$, avec $\theta \in]0, \infty[$.

Exercice 3. Soit $X_1, \ldots, X_n$ un n-échantillon de loi normale de moyenne $\mu \in \mathbf{R}$ et $\sigma^2 > 0$.

(1) On suppose σ^2 connue. Donner l'estimateur du maximum de vraisemblance de μ.
(2) On suppose que μ et σ^2 sont inconnues. Donner l'estimateur du maximum de vraisemblance du couple (μ, σ^2).

Exercice 4. Soit $X_1, \ldots, X_n$ un n-échantillon de loi de Poisson de paramètre $\theta > 0$. Posons $\hat{\theta}_n = \frac{X_1 + \cdots + X_n}{n}$.

(1) Montrer que $\hat{\theta}_n$ est un estimateur consistant de θ.
(2) Calculer le biais de $\hat{\theta}_n$.
(3) Calculer le risque quadratique de $\hat{\theta}_n$.

Exercice 5. Une variable aléatoire X suit une loi Gamma de paramètres (a, b), avec $a > 0$ et $b > 0$, si sa loi admet pour densité :

$$f_{a,b}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \mathbf{1}_{]0,\infty[(x)},$$

avec $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} \mathrm{d}x$. On considère un n-échantillon $X_1, \ldots, X_n$ de loi Gamma de paramètres inconnus (a, b). Le but de cet exercice est d'estimer a et b par la méthode des moments.

(1) Posons $m_1 = \mathbb{E}[X]$ et $m_2 = \mathbb{E}\left[X^2\right]$. Montrer que $m_1 = \frac{a}{b}$ et $m_2 = \frac{a(a+1)}{b^2}$.
(2) Exprimer a et b en fonction de m_1 et m_2.
(3) Proposer des estimateurs pour les paramètres a et b. Ces estimateurs sont-ils fortement consistants ?

2. Intervalles de confiance

Exercice 6. Soit $X_1, \dots, X_n$ un n-échantillon de moyenne inconnue μ et de variance connue σ^2. Notons $\hat{\mu} = \overline{X_n} = \frac{X_1+\cdots+X_n}{n}$ la moyenne empirique.

(1) Montrer que pour tout $\alpha \in [0,1[$, on a
$$\mathbb{P}\left(\mu \in \left[\hat{\mu} - \sqrt{\frac{\sigma^2}{n}}u_{1-\alpha/2},\ \hat{\mu} + \sqrt{\frac{\sigma^2}{n}}u_{1-\alpha/2}\right]\right) \underset{n\to\infty}{\longrightarrow} 1-\alpha,$$
où $u_{1-\alpha/2}$ désigne le quantile de la loi normale centrée réduite.
(2) Sur les cinquante dernières années, on a relevé en France une pluviométrie annuelle moyenne de 800mm pour un écart-type connu de 100mm. Déterminer un intervalle de confiance (asymptotique) de la moyenne au niveau de confiance de 90% en précisant quelles hypothèses sont faites.

Exercice 7. Soit $X_1, \dots, X_n$ un n-échantillon de loi de Bernoulli de paramètre $p \in]0,1[$. Notons $\hat{p}$ la proportion empirique.

(1) Montrer que pour tout $\alpha \in [0,1[$, on a
$$\mathbb{P}\left(p \in \left[\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}u_{1-\alpha/2},\ \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}u_{1-\alpha/2}\right]\right) \underset{n\to\infty}{\longrightarrow} 1-\alpha,$$
où $u_{1-\alpha/2}$ désigne le quantile de la loi normale centrée réduite.
(2) On veut étudier la proportion p de gens qui vont au cinéma chaque mois. On prélève un n-échantillon de taille $n = 100$. On observe une proportion $\hat{p}(\omega) = 0.1$ de gens qui vont chaque mois au cinéma. Donner un intervalle de confiance pour p de niveau de confiance 90% issu de l'observation.

Exercice 8. On considère un n-échantillon $X_1, \dots, X_n$ de loi de Poisson de paramètre inconnu $\lambda > 0$.

(1) Vers quoi converge $\overline{X_n}$?
(2) Justifier que $\sqrt{\frac{n}{\overline{X_n}}}(\overline{X_n} - \lambda)$ converge en loi vers la loi normale centrée réduite.
(3) En déduire un intervalle de confiance au niveau 95% pour λ.
(4) Le tableau ci-dessous donne les fréquences de valeurs issues d'un échantillon observé de taille 100.

Valeur	0	1	2	3	4	5	6
Fréquence	0.07	0.20	0.33	0.20	0.10	0.08	0.02

(a) Calculer la moyenne empirique de l'échantillon.
(b) Donner un intervalle de confiance calculé à partir de cet échantillon au niveau de 95%.

Exercice 9. On suppose que le chiffre d'affaires mensuel d'une entreprise suive une loi normale de moyenne μ et de variance σ^2 inconnues. Sur les 12 derniers mois, on a observé les chiffres d'affaires (en kilo euros) suivants : $9;\ 9,8;\ 10,3;\ 7;\ 7,2;\ 9,5;\ 10,6;\ 10,1;\ 10;\ 8,9;\ 8,7;\ 11,1.$

(1) Donner des estimations ponctuelles de la moyenne et de la variance.

(2) On rappelle le résultat suivant : si $(X_1, \dots, X_n)$ est un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$ alors les estimateurs $\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$ et $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2$ sont des statistiques indépendantes, de lois respectives données par :

$$\overline{X_n} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{et} \quad \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1).$$

En particulier,

$$\frac{\overline{X_n} - \mu}{\sqrt{S_n^2/n}} \sim \tau(n-1),$$

où $\tau(n-1)$ désigne la loi de Student à $n-1$ degrés de liberté, i.e. la loi dont la densité est donnée par :

$$x \mapsto \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{(n-1)\pi}\,\Gamma\left(\frac{n-1}{2}\right)}\left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}.$$

(a) Donner un intervalle de confiance pour la moyenne de niveau de confiance 90%.

(b) Donner un intervalle de confiance pour la variance de niveau de confiance 90%.

Exercice 10. On considère un n-échantillon $X_1, \dots, X_n$ de loi exponentielle de paramètre $\lambda > 0$. Soit $\alpha \in]0, 1[$.

(1) Montrer que la loi de la variable aléatoire $\lambda \max_{1 \le i \le n} X_i$ ne dépend pas de λ. En déduire un intervalle de confiance pour le paramètre λ au niveau de confiance $1 - \alpha$.

(2) Construire un intervalle de confiance pour λ au niveau $1 - \alpha$ en utilisant la statistique $\overline{X_n}$.

Exercice 11. On considère un n-échantillon $X_1, \dots, X_n$ de loi P_θ, $\theta > 0$, où P_θ a pour fonction de répartition :

$$F_\theta(t) = \begin{cases} 0 & \text{si} \quad t < 0, \\ t^\theta & \text{si} \quad 0 \le t < 1, \\ 1 & \text{si} \quad t \ge 1 \end{cases}$$

(1) Calculer $\mathbb{E}_\theta[X_1]$. En déduire que la variable aléatoire S_n, définie par :

$$S_n = \frac{\overline{X_n}}{1 - \overline{X_n}}, \quad \text{où} \quad \overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i,$$

est un estimateur fortement consistant de θ.

(2) On pose $Y_i = -\log X_i$ pour $1 \le i \le n$. Montrer que les variables aléatoires $(Y_i)_{1 \le i \le n}$ sont indépendantes et suivent une loi exponentielle de paramètre θ sous P_θ.

(3) Calculer $\mathbb{E}_\theta[Y_1]$ et $\mathbb{V}_\theta[Y_1]$.

(4) En déduire que la variable aléatoire T_n, définie par :

$$T_n = \frac{1}{\overline{Y_n}}, \quad \text{où} \quad \overline{Y_n} = \frac{1}{n}\sum_{i=1}^n Y_i,$$

est un estimateur fortement consistant de θ. Est-il sans biais (étudier seulement le cas $n = 1$) ?

(5) Justifier que $\sqrt{n}\left(\frac{\theta}{T_n} - 1\right)$ converge en loi vers la loi normale centrée réduite.

(6) En déduire un intervalle de confiance pour θ au niveau de confiance 0,95 en négligeant l'erreur due à l'approximation gaussienne.

Exercice 12. On considère un n-échantillon $X_1, \ldots, X_n$ de loi Q_θ, $\theta > 0$, où Q_θ a pour densité

$$f_\theta(x) = x\theta^{-x^2/2}\mathbf{1}_{\mathbf{R}_+}(x).$$

On note $\hat{\theta}$ l'estimateur de θ suivant :

$$\hat{\theta} = \exp\left(\frac{2n}{\sum_{i=1}^n X_i^2}\right).$$

(1) Calculer les moments d'ordres 2 et 4 de la loi Q_θ.

(2) Montrer que $\hat{\theta}$ est biaisé et consistant.

(3) Déterminer la loi limite de $\hat{\theta}$.

(4) Soit $\alpha \in]0,1[$. Construire un intervalle de confiance asymptotique pour θ au niveau de confiance $1-\alpha$.