

TD2 : Échantillonnage et estimation



Version web

Ex 1

Propriétés fondamentales des estimateurs

f.JtB

Soit $(X_1, X_2, \dots, X_n)$ un n -échantillon d'une variable aléatoire X d'espérance $\mu = \mathbb{E}(X)$ et de variance $\sigma^2 = \text{Var}(X)$ finies. On définit les estimateurs suivants :

$$T_1 = \frac{1}{n} \sum_{i=1}^n X_i \quad T_2 = \frac{1}{n-1} \sum_{i=1}^n X_i \quad T_3 = X_1$$

1. Rappeler les définitions d'un estimateur sans biais et d'un estimateur convergent.
2. Les trois estimateurs T_1 , T_2 et T_3 sont-ils des estimateurs sans biais de μ ? Justifier.

i

Calculer l'espérance de chaque estimateur en utilisant la linéarité de l'espérance.

3. Calculer la variance de chacun des trois estimateurs.

i

Utiliser le fait que les variables X_i sont indépendantes et identiquement distribuées.

4. Parmi les estimateurs sans biais de μ , lequel est le plus efficace ? Justifier.

i

L'estimateur le plus efficace est celui qui a la plus petite variance.

5. On définit l'erreur quadratique moyenne (EQM) d'un estimateur $\hat{\theta}$ de θ par :

$$\text{EQM}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$$

Montrer que $\text{EQM}(\hat{\theta}) = \text{Var}(\hat{\theta}) + b^2(\hat{\theta})$ où $b(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$ est le biais.

i

Développer $(\hat{\theta} - \theta)^2 = (\hat{\theta} - \mathbb{E}(\hat{\theta}) + \mathbb{E}(\hat{\theta}) - \theta)^2$.

6. Calculer l'erreur quadratique moyenne de chacun des trois estimateurs. Quel est le meilleur estimateur de μ au sens de l'EQM ?

i

Utiliser la formule établie à la question précédente.

Ex 2

Lois statistiques et étude d'estimateurs

3kdw

On considère un échantillon (X_i) de taille $n = 5$ dans une population suivant une loi normale de paramètres μ et σ^2 . On pose

$$T_1 = \frac{1}{5} \sum_{i=1}^5 X_i \quad T_2 = \frac{1}{5}(X_1 + X_2) + \frac{1}{4}(X_3 + X_4 + X_5) \quad T_3 = \frac{1}{10}(2X_1 + 3X_2) + \frac{1}{8}(X_3 + 2X_4 + X_5)$$

$$U = \frac{1}{\sigma^2} \sum_{i=1}^5 (X_i - \mu)^2 \quad V = \frac{1}{\sigma^2} \sum_{i=1}^5 (X_i - T_1)^2$$

1. Quelle est la loi suivie par la variable $X_1 - X_2$? Justifier.
2. On cherche à estimer μ à l'aide des estimateurs T_1, T_2, T_3 . Étudier leur biais et comparer l'efficacité des estimateurs sans biais.
3. Quelle est la loi suivie par la variable U ? la variable V ? justifier.
4. Déterminer $x \in \mathbb{R}$ tel que $P(U > x) = 0,05$.
5. En utilisant T_1 et U , construire une variable Y qui suive une loi de Student dont on précisera le paramètre.

Ex 3

Maximum de vraisemblance pour une loi géométrique

jemd

On rappelle qu'une variable X suit une loi géométrique de paramètre $p \in]0; 1[$ si X est à valeurs dans $\mathbb{N}^* = \{1; 2; 3; \dots\}$ et si pour tout $k \in \mathbb{N}^*$,

$$P(X = k) = p(1 - p)^{k-1}$$

On cherche à estimer ce paramètre p à partir d'un échantillon.

1. On considère un échantillon $(X_1, X_2, X_3, X_4, X_5)$ ayant pour loi mère une loi géométrique de paramètre p et on suppose que la suite $(3; 4; 4; 2; 3)$ constitue une réalisation de cet échantillon.
 - (a) Exprimer la fonction de vraisemblance associée à cet échantillon.
 - (b) En déduire une estimation de p par la méthode du maximum de vraisemblance.
2. Afin de déterminer un estimateur de p , on considère maintenant un n -échantillon $(X_1, \dots, X_n)$ ayant pour loi mère une loi géométrique de paramètre p et n entiers non nuls $(x_1, \dots, x_n)$ constituant une réalisation de cet échantillon.
 - (a) Exprimer la fonction de vraisemblance associée à cet échantillon.
 - (b) En utilisant la méthode du maximum de vraisemblance, déterminer un estimateur du paramètre p .

Ex 4

Estimation du paramètre d'une loi

dXNn

On fixe un réel $a > 0$ et on définit une variable aléatoire X sur $\mathbb{N}$ dont la loi de probabilité est

définie par

$$P(X = k) = \frac{a^k}{(1+a)^{k+1}}$$

pour tout $k \in \mathbb{N}$.

A l'aide de la méthode du maximum de vraisemblance, définir un estimateur de a et déterminer son biais.

Ex 5

Estimation des paramètres d'une loi normale

PwkL

★ ★

On considère un échantillon $X_1, \dots, X_n$ suivant une loi normale $\mathcal{N}(\mu, \sigma)$. On cherche un estimateur de μ et de σ par la méthode du maximum de vraisemblance. On note $(x_1, \dots, x_n)$ une réalisation de cet échantillon. On rappelle que la densité d'une loi normale est

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

1. Exprimer la fonction de vraisemblance $L(x_1, \dots, x_n, \mu, \sigma)$, puis son logarithme.
2. Dériver $\ln L(x_1, \dots, x_n, \mu, \sigma)$ par rapport à μ .
3. En déduire un estimateur de μ .
4. Déterminer un estimateur de σ avec une démarche analogue.

Ex 6

Estimateurs et intervalles de confiance

0e0x

Un centre médical mène des études statistiques sur le taux de cholestérol de son personnel, mesuré en centigramme par litre de sang. Les observations sur 20 individus supposés sains, tirés au sort, sont les suivantes :

Taux de cholestérol en cg	120	160	200	240	280	320
Nombre de sujets observés	2	4	5	4	3	2

Le taux de cholestérol chez un individu sain (noté X) est supposé suivre une loi normale.

1. Donner une estimation $\bar{x}$ de l'espérance mathématique de X et une estimation s^2 de la variance de X . On précisera les estimateurs choisis et on en donnera les propriétés.
2. Donner l'intervalle de confiance de l'espérance de X , centré sur $\bar{x}$, au risque $\alpha = 0.01$.

Ex 7

Intervalles de confiance

gk1A

Une association de consommateurs a effectué une enquête sur le prix d'un produit dans le supermarchés. Pour ce produit, les prix suivants (en euros) ont été relevés dans 7 supermarchés différents :

52 52 43 51 69 55 49

1. En supposant que le prix X de ce produit est distribué selon une loi normale $\mathcal{N}(\mu, \sigma)$, calculer une estimation $\bar{x}$ de l'espérance $\mathbb{E}(X)$ et une estimation s de l'écart-type $\sigma(X)$. On précisera les estimateurs choisis et on en donnera les propriétés.
2. Déterminer les intervalles de confiance symétriques au seuil de 90% et 99% centrés en $\mu = \mathbb{E}(X)$.

Ex 8

Intervalles de confiance

5cbY

Pour étudier le degré de pollution des eaux d'une rivière par les déchets d'une usine, on mesure la teneur en milligrammes d'un certain polluant par litre d'eau.

1. On effectue 17 mesures pour lesquelles les valeurs observées des teneurs en polluant ont une moyenne de 5.2 mg/L et un écart-type de 1.2 mg/L.
 - (a) Donner une estimation sans biais de la moyenne et de l'écart type de la teneur en polluant dans cette rivière.
 - (b) Donner un intervalle de confiance au seuil de 5% pour la teneur moyenne en polluant de cette rivière.
 - (c) A partir de ces résultats, peut-on considérer que l'usine respecte la législation en vigueur selon laquelle cette teneur moyenne en polluant ne doit pas dépasser 7 mg/L ?
2. On suppose ici l'écart type de la population connu et égal à 1.2 mg/L. Combien de mesures devrait-on faire pour estimer la teneur moyenne en polluant avec une précision de 0.1 mg/L au niveau de confiance 95% ?

i

Attention, l'écart-type de la population σ est ici supposé connu. Quelle loi statistique doit-on utiliser dans ce cas, quelle que soit la taille de l'échantillon ? De plus, la "précision" correspond à la marge d'erreur, c'est-à-dire au demi-intervalle.

Ex 9

Intervalle de confiance pour une proportion

iboj

On effectue un contrôle de fabrication sur des pièces dont la proportion p est défectueuse.

1. On contrôle un lot de 200 pièces et on trouve $n = 24$ pièces défectueuses. Donner les intervalles de confiance pour l'estimation de p , au risque $\alpha = 0.05$.

i

La première étape consiste à calculer la fréquence observée f de pièces défectueuses dans l'échantillon. Ensuite, il faut déterminer la valeur critique $z_{\alpha/2}$ correspondant à un risque $\alpha = 0.05$ pour une loi normale centrée réduite. L'intervalle de confiance pour une proportion p est donné par la formule : $I_C = \left[f - z_{\alpha/2} \sqrt{\frac{f(1-f)}{N}}; f + z_{\alpha/2} \sqrt{\frac{f(1-f)}{N}} \right]$, où N est la taille du lot.

2. Combien faudrait-il contrôler de pièces pour que, au risque $\alpha = 0.05$, l'intervalle de confiance centré soit d'amplitude 0.04 ?

i L'amplitude A de l'intervalle de confiance est la différence entre la borne supérieure et la borne inférieure. Elle est égale à $A = 2 \times z_{\alpha/2} \sqrt{\frac{f(1-f)}{n}}$. Il faut isoler n dans cette équation. Comme la proportion p (et donc la fréquence f) de la future population est inconnue, on peut utiliser l'estimation $f = 0.12$ de la question précédente, ou prendre le cas le plus défavorable qui maximise le produit $f(1-f)$, c'est-à-dire $f = 0.5$.

Ex 10**Étude complète d'un processus industriel**

p7QA

★★★★

Une entreprise fabrique des câbles électriques dont la résistance R (en ohms) est modélisée par une variable aléatoire suivant une loi exponentielle de paramètre $\lambda > 0$. La densité de cette loi est donnée par :

$$f(x) = \lambda e^{-\lambda x} \quad \text{pour } x \geq 0$$

On rappelle que pour $X \sim \mathcal{E}(\lambda) : \mathbb{E}(X) = \frac{1}{\lambda}$ et $\text{Var}(X) = \frac{1}{\lambda^2}$.

Pour contrôler la qualité de production, on mesure la résistance de $n = 10$ câbles tirés au hasard. On obtient les valeurs suivantes (en ohms) :

2.3 1.8 3.1 2.7 1.5 2.9 2.1 3.4 2.5 1.9

Les normes de l'industrie exigent que la résistance moyenne des câbles soit au moins égale à 2.5 ohms, c'est-à-dire $\mathbb{E}(R) \geq 2.5$.

1. Exprimer la fonction de vraisemblance $L(r_1, \dots, r_{10}, \lambda)$ associée à cet échantillon, où $r_1, \dots, r_{10}$ désignent les valeurs observées.

i Utiliser le fait que les observations sont indépendantes et identiquement distribuées.

2. En utilisant la méthode du maximum de vraisemblance, déterminer un estimateur $\hat{\Lambda}_n$ du paramètre λ à partir d'un n -échantillon $(X_1, \dots, X_n)$.

i Calculer le logarithme de la vraisemblance et annuler sa dérivée par rapport à λ .

3. L'estimateur $\hat{\Lambda}_n$ est-il sans biais? Justifier.

i Calculer $\mathbb{E}(\hat{\Lambda}_n)$ et comparer à λ . Attention, $\mathbb{E}(1/\bar{X}_n) \neq 1/\mathbb{E}(\bar{X}_n)$ en général.

4. Calculer une estimation $\hat{\lambda}$ du paramètre λ à partir des données observées.

i Calculer la moyenne empirique des observations.

5. En déduire une estimation ponctuelle de l'espérance $\mathbb{E}(R)$ de la résistance.

i Utiliser la relation $\mathbb{E}(R) = 1/\lambda$.

6. On admet que pour un n -échantillon de loi exponentielle $\mathcal{E}(\lambda)$, la variable $2\lambda n\overline{X}_n$ suit une loi du chi-deux à $2n$ degrés de liberté. Déterminer un intervalle de confiance à 95% pour le paramètre λ .

i

Utiliser le fait que $\mathbb{P}(\chi_{0.025, 2n}^2 \leq 2\lambda n\overline{X}_n \leq \chi_{0.975, 2n}^2) = 0.95$ où $\chi_{\alpha, k}^2$ désigne le quantile d'ordre α de la loi $\chi^2(k)$.

7. En déduire un intervalle de confiance à 95% pour l'espérance $\mathbb{E}(R)$.

i

Utiliser la relation $\mathbb{E}(R) = 1/\lambda$ et inverser les bornes de l'intervalle.

8. Au vu de ces résultats, peut-on affirmer avec 95% de confiance que l'entreprise respecte les normes de l'industrie ($\mathbb{E}(R) \geq 2.5$ ohms) ? Justifier votre réponse.

i

Comparer la borne inférieure de l'intervalle de confiance avec la norme exigée.