

TD16-Estimation

1 Estimation ponctuelle

Exercice 1

Soit X une variable aléatoire suivant la loi de Poisson de paramètre $\theta \in \mathbb{R}_+^*$ inconnu. Soit (X_1, \dots, X_n) un n -échantillon de la loi de X .

- (a) Montrer que $e^{-\bar{X}_n}$ est un estimateur asymptotiquement sans biais de $e^{-\theta}$.
- (b) Déterminer son risque quadratique.
- (c) Est-ce un estimateur convergent de θ ?

Exercice 2

Soit X une variable aléatoire suivant la loi uniforme sur $[0, \theta]$ où $\theta \in \mathbb{R}_+^*$ est un paramètre inconnu. Soit (X_1, \dots, X_n) un n -échantillon de la loi de X . On considère les estimateurs suivants :

$$T_n = 2\bar{X}_n \quad ; \quad T'_n = \max(X_1, \dots, X_n) \quad ; \quad T''_n = \frac{n+1}{n} T'_n.$$

- (a) Montrer que T_n est un estimateur sans biais de θ .
- (b) Déterminer son risque quadratique.
- (c) L'estimateur T_n est-il un estimateur convergent de θ ?
- (a) Déterminer le biais et le risque quadratique de T'_n .
- (b) Donner un équivalent simple de $r_\theta(T'_n)$ quand n tend vers $+\infty$.
- Mêmes questions pour T''_n .
- Quel est le meilleur estimateur de θ ?
- Ecrire un programme Scilab simulant ces trois estimateurs.

Exercice 3

La durée de vie d'une lampe est modélisée par une variable aléatoire réelle suivant une loi exponentielle de paramètre λ .

On cherche à estimer la durée de vie moyenne $\frac{1}{\lambda}$ de la lampe.

On prélève un échantillon de n lampes et on note X_1, \dots, X_n leurs durées de vie.

- Montrer que \bar{X}_n est un estimateur sans biais de $\frac{1}{\lambda}$ et déterminer son risque quadra-

tique.

- On pose $Y_n = \min(X_1, \dots, X_n)$.
 - Déterminer la loi de Y_n .
 - En déduire que nY_n est un estimateur sans biais de $\frac{1}{\lambda}$.
 - Calculer son risque quadratique.
- Comparer les deux estimateurs.

Exercice 4 (Estimateur sans biais du paramètre d'une loi géométrique)

Soit n un entier naturel supérieur ou égal à 2 et soit $p \in]0, 1[$. On pose $q = 1 - p$. On considère T_1, \dots, T_n des variables aléatoires mutuellement indépendantes de loi géométrique de paramètre p . On pose $S_n = T_1 + \dots + T_n$.

- Déterminer $S_n(\Omega)$ puis montrer que : $\forall k \in S_n(\Omega), P(S_n = k) = \binom{k-1}{n-1} p^n q^{k-n}$.
- (a) Rappeler la valeur de $P(T_n = 1)$ et calculer $P_{[S_n=k]}(T_n = 1)$ pour tout $k \geq n$.
- (b) En déduire : $\forall N \in \mathbb{N}, \sum_{i=N}^{+\infty} \binom{i}{N} q^{i-N} = \frac{1}{(1-q)^{N+1}}$.
- Montrer que $U_n = \frac{n-1}{S_n-1}$ est un estimateur sans biais de p .
- On suppose dans cette question que n est supérieur ou égal à 3.
 - Montrer que la variable aléatoire $V_n = \frac{(n-1)^2}{(S_n-1)(S_n-2)}$ possède une espérance et la calculer.
 - En déduire que U_n admet un moment d'ordre 2 puis que le risque quadratique de U_n est majoré par $\frac{p^2}{n-2}$.

Exercice 5 (ESC 2006)

Dans cet exercice R désigne un réel fixé strictement positif et on considère la fonction f définie sur \mathbb{R} par :

$$\begin{cases} f(t) = 0 & \text{si } t \notin [0; R] \\ f(t) = \frac{2t}{R^2} & \text{si } t \in [0; R] \end{cases}$$

1. Montrer que f est une densité de probabilité.
2. On note dans toute la suite X une variable aléatoire réelle de densité f . Déterminer sa fonction de répartition F_X .
3. (a) Montrer que X admet une espérance et que $E(X) = \frac{2R}{3}$.
(b) Montrer que X admet une variance et que $V(X) = \frac{R^2}{18}$.

Dans toute la suite n désigne un entier naturel non nul et X_1, X_2, \dots, X_n des variables aléatoires indépendantes et de même loi que X . On cherche à estimer le réel R à l'aide de X_1, X_2, \dots, X_n .

4. On note $T_n = \frac{3}{2n} \sum_{k=1}^n X_k$ et on cherche à estimer R avec T_n .

Montrer que T_n est un estimateur sans biais de R et calculer son risque quadratique.

1. On note $M_n = \max(X_1, \dots, X_n)$.
(a) Montrer que pour tout réel x , $P(M_n \leq x) = (F_X(x))^n$. En déduire la fonction de répartition de M_n , puis montrer que M_n est une variable aléatoire à densité.
(b) Montrer qu'une densité possible de M_n est la fonction g_n définie sur \mathbb{R} par :

$$\begin{cases} g_n(t) = 2n \frac{t^{2n-1}}{R^{2n}} & \text{si } t \in [0; R] \\ g_n(t) = 0 & \text{si } t \notin [0; R] \end{cases}.$$

- (c) Montrer que M_n admet une espérance et une variance, et que :

$$E(M_n) = \frac{2n}{2n+1}R \quad \text{et} \quad V(M_n) = \frac{n}{(n+1)(2n+1)^2}R^2.$$

- (d) On cherche à estimer R avec M_n :

Calculer le biais de M_n et son risque quadratique.

2. (a) Déterminer un équivalent simple lorsque n tend vers $+\infty$ de $b_R(M_n)$ et $r_R(M_n)$.
(b) Quels sont les avantages et les inconvénients des estimateurs T_n et M_n ?

2 Intervalle de confiance

Exercice 6

Lors d'une élection, les candidats A et B s'opposent. On souhaite réaliser un sondage afin de connaître, avec un niveau de confiance de 0.95 (95%), le futur vainqueur. Un précédent sondage ayant donné les deux candidats au coude à coude, on souhaite une

marge d'erreur de 0.01 (1%).

1. Combien de personnes faut-il interroger au minimum si on se fie à l'inégalité de Bienaymé-Tchebychev ?
2. Combien de personnes faut-il interroger au minimum si on se fie au théorème central limite ?

Exercice 7 (EDHEC 2000 voie S)

Un sondage consiste à proposer l'affirmation « A » à certaines personnes d'une population donnée. Le sujet abordé étant délicat, le stratagème suivant est mis en place afin de mettre en confiance les personnes sondées pour qu'elles ne mentent pas...

L'enquêteur dispose d'un paquet de 20 cartes, numérotées de 1 à 20, qu'il remet à la personne sondée. Celle-ci tire une carte au hasard et ne la montre pas à l'enquêteur. La règle est alors la suivante :

- si la carte porte le numéro 1, la personne sondée répond "vrai" si elle est d'accord avec l'affirmation A et "faux" sinon ;
- si la carte porte un autre numéro, la personne sondée répond "vrai" si elle n'est pas d'accord avec l'affirmation A et "faux" sinon.

Le but de l'enquête est d'évaluer la proportion p de gens de cette population qui sont réellement d'accord avec l'affirmation « A ».

1. On interroge une personne selon ce procédé et on considère l'événement suivant, notée V : « la personne répond "vrai" ». On note $\theta = P(V)$.
En utilisant la formule des probabilités totales, exprimer θ en fonction de p , puis en déduire p en fonction de θ .
2. On considère un échantillon aléatoire, de taille n , extrait de la population considérée et on note S_n , le nombre de réponses "vrai" obtenues. On suppose n assez grand pour pouvoir considérer que cet échantillonnage est assimilable à un tirage avec remise.
(a) Donner la loi de S_n ainsi que son espérance et sa variance.
(b) Montrer que $\frac{S_n}{n}$ est un estimateur sans biais et convergent de θ .
3. Dans cette question, on suppose que l'on a réalisé un échantillon de 100 personnes et on constate que 23 personnes ont répondu "vrai".
(a) Donner une estimation ponctuelle de θ et de p .
(b) Donner un intervalle de confiance à 95% de θ puis de p .
On rappelle que, si Φ désigne la fonction de répartition d'une variable X suivant la loi normale $\mathcal{N}(0, 1)$, alors $\Phi(1,96) = 0,975$.

Exercice 8 (ESSEC 2005 maths II)

Une entreprise souhaite acquérir une machine qui fabrique un certain type d'objets et

qui, en fonctionnement normal, produit une proportion $p \in]0, 1[$ d'objets défectueux. Le directeur veut connaître la valeur de p . Pour cela il teste la machine et prélève un échantillon de $n \in \mathbb{N}^*$ objets qu'il analyse. Pour tout $i \in \llbracket 1, n \rrbracket$, soit X_i la variable aléatoire de Bernoulli définie par

$$X_i = \begin{cases} 1 & \text{si le } i\text{-ème objet est défectueux} \\ 0 & \text{sinon} \end{cases}.$$

On suppose que dans les conditions de prélèvement, les variables aléatoires X_1, \dots, X_n sont indépendantes.

- Montrer que $F_n = \frac{S_n}{n}$ est un estimateur sans biais de p .
 - Calculer le risque quadratique $r_n = E((F_n - p)^2)$. Déterminer $\lim_{n \rightarrow +\infty} r_n$.
- Soit α un réel de $]0, 1[$. On souhaite déterminer dans cette question un intervalle de confiance du paramètre p inconnu, au niveau de confiance $1 - \alpha$, à partir de l'échantillon (X_1, \dots, X_n) .
 - Quelle est la limite en loi de la suite $\left(\sqrt{n} \frac{F_n - p}{\sqrt{p(1-p)}} \right)_{n \in \mathbb{N}^*}$?
 - Soit f_n la réalisation de F_n sur l'échantillon considéré. Soit t_α le réel défini par $\Phi(t_\alpha) = 1 - \frac{\alpha}{2}$, où Φ désigne la fonction de répartition de la loi normale centrée, réduite. Montrer qu'un intervalle de confiance de p au niveau $1 - \alpha$ est donné par $[U_n, V_n]$ tel que

$$P(U_n \leq p \leq V_n) \geq 1 - \alpha$$

avec

$$U_n = f_n - \frac{t_\alpha}{2\sqrt{n}}, \quad V_n = f_n + \frac{t_\alpha}{2\sqrt{n}}.$$

3 Compléments

Exercice 9 (ESSEC 2009)

Dans tout l'exercice (Ω, \mathcal{F}, P) désignera un espace probabilisé et les variables aléatoires utilisées seront toutes définies sur cet espace probabilisé.

I. Premier problème d'estimation

Dans ce premier problème d'estimation, on dispose d'une seule observation notée X .

On suppose que X admet pour densité f_θ définie sur \mathbb{R} par

$$\forall x \in \mathbb{R}, \quad f_\theta(x) = \begin{cases} \frac{k+1}{\theta^{k+1}} x^k & \text{si } x \in [0, \theta] \\ 0 & \text{sinon} \end{cases}$$

où k est un entier naturel non nul et θ un paramètre réel inconnu strictement positif que l'on souhaite estimer.

- Montrer que f_θ est bien une densité de probabilité.
- Calculer $E(X)$ et $V(X)$.
- Déterminer λ_0 réel dépendant uniquement de k tel que $\lambda_0 X$ estime θ sans biais.
- Montrer le résultat du cours précisant que pour tout T estimateur de θ possédant un moment d'ordre 2 on a :

$$r_\theta(T) = (E(T) - \theta)^2 + V(T).$$

- Donner la valeur de $r_\theta(\lambda_0 X)$.
- Le but de cette question est de déterminer un estimateur de θ ayant un plus petit risque quadratique que celui de $\lambda_0 X$.
 - Montrer que pour tout réel λ

$$r_\theta(\lambda X) = \theta^2 Q(\lambda)$$

où Q est un polynôme de degré 2 dont les coefficients ne dépendent que de k .

- Montrer que la fonction $\lambda \mapsto Q(\lambda)$ atteint son minimum en un unique réel noté λ^* que l'on exprimera en fonction de k .
- Conclure.

II. Second problème d'estimation

Dans ce second problème d'estimation, on dispose de n observations indépendantes ($n \geq 2$) notées X_1, \dots, X_n de même loi de Poisson de paramètre θ inconnu ($\theta \in]0, +\infty[$). On souhaite estimer le paramètre $\exp(-\theta)$. On définit pour tout i élément de $\{1, \dots, n\}$ la variable aléatoire Y_i par

$$\forall \omega \in \Omega, \quad Y_i(\omega) = \begin{cases} 1 & \text{si } X_i(\omega) = 0 \\ 0 & \text{sinon.} \end{cases}$$

- Pour tout i élément de $\{1, \dots, n\}$, donner la loi de Y_i .
- Donner la loi de $\sum_{i=1}^n Y_i$, puis montrer que $E(\overline{Y}_n) = \exp(-\theta)$.
- Calculer $V(\overline{Y}_n)$.

Pour tout k élément de $\{1, \dots, n\}$ on définit $S_k = \sum_{i=1}^k X_i$ et on note pour tout j entier naturel

$$\varphi(j) = P_{\{S_n=j\}}(X_1 = 0).$$

10. Rappeler sans démonstration la loi de S_k pour tout k élément de $\{1, \dots, n\}$.

11. Montrer que pour tout j entier naturel

$$\varphi(j) = \left(1 - \frac{1}{n}\right)^j.$$

On a donc $\varphi(j)$ indépendant du paramètre θ inconnu et on peut définir l'estimateur

$$\varphi(S_n) = \left(1 - \frac{1}{n}\right)^{S_n}.$$

12. Montrer que $\varphi(S_n)$ admet une espérance et que $E(\varphi(S_n)) = \exp(-\theta)$.

13. Montrer que $\varphi(S_n)$ admet une variance vérifiant :

$$V(\varphi(S_n)) = \exp(-2\theta) \left(\exp\left(\frac{\theta}{n}\right) - 1 \right).$$

14. On souhaite comparer les performances de \overline{Y}_n et $\varphi(S_n)$ en tant qu'estimateurs de $\exp(-\theta)$.

(a) En utilisant le théorème des accroissements finis, démontrer que :

$$1 \leq \frac{\exp(\theta) - 1}{\theta} \leq \exp(\theta).$$

(b) Soit la fonction $h : [0, 1] \rightarrow \mathbb{R}$ définie par

$$\forall t \in [0, 1], \quad h(t) = t \exp(\theta) + (1 - t) - \exp(t\theta).$$

Étudier les variations de h .

(c) En déduire :

$$\forall n \in \mathbb{N}^* \quad \exp\left(\frac{\theta}{n}\right) \leq \frac{\exp(\theta)}{n} + \frac{n-1}{n}$$

puis l'inégalité

$$\forall n \in \mathbb{N}^* \quad V(\varphi(S_n)) \leq V(\overline{Y}_n).$$

(d) Comparer les risques quadratiques de \overline{Y}_n et $\varphi(S_n)$.

On reprendra à la fin de la partie IV l'étude de $\varphi(S_n)$.

III. Information de Fisher

15. **Cas discret.** Dans cette question, on considère I un intervalle de \mathbb{R} , θ un paramètre inconnu de I et X une variable aléatoire à valeurs dans \mathbb{N} ($X(\Omega) \subseteq \mathbb{N}$). On suppose qu'il existe une fonction p définie sur $I \times X(\Omega)$ telle que :

$$\forall k \in X(\Omega), \quad P(X = k) = p(\theta, k) \quad \text{où } \theta \mapsto p(\theta, k) \text{ est dérivable sur } I.$$

On définit sous réserve d'existence l'**information de Fisher** de X par :

$$I_X(\theta) = \sum_{k \in X(\Omega)} \left(\frac{\partial}{\partial \theta} \ln p(\theta, k) \right)^2 p(\theta, k).$$

(a) Dans cette question, on considère X variable aléatoire de loi de Bernoulli de paramètre $\theta \in]0, 1[$. On a alors $X(\Omega) = \{0, 1\}$, $P(X = 1) = p(\theta, 1) = \theta$, $P(X = 0) = p(\theta, 0) = 1 - \theta$.

Montrer que

$$I_X(\theta) = \frac{1}{\theta(1-\theta)}.$$

(b) Dans cette question, on considère X variable aléatoire de loi binomiale de paramètres $N \in \mathbb{N}^*$ et $\theta \in]0, 1[$.

i. Montrer que

$$I_X(\theta) = \frac{1}{(\theta(1-\theta))^2} \sum_{k=0}^N (k - N\theta)^2 \binom{N}{k} \theta^k (1-\theta)^{N-k}.$$

ii. En déduire que

$$I_X(\theta) = \frac{V(X)}{(\theta(1-\theta))^2}.$$

Donner la valeur de $I_X(\theta)$.

(c) Dans cette question, on considère X une variable aléatoire de Poisson de paramètre $\theta \in \mathbb{R}_+^*$. Puisque $X(\Omega) = \mathbb{N}$, on a sous réserve de convergence

$$I_X(\theta) = \sum_{k=0}^{+\infty} \left(\frac{\partial}{\partial \theta} \ln p(\theta, k) \right)^2 p(\theta, k).$$

i. Montrer que la série de terme général $\left(\frac{\partial}{\partial \theta} \ln p(\theta, k) \right)^2 p(\theta, k)$ converge et calculer sa somme $I_X(\theta)$.

ii. Justifier que

$$I_X(\theta) = E \left(\left(\frac{\partial}{\partial \theta} \ln p(\theta, X) \right)^2 \right).$$

16. **Cas d'une variable gaussienne.** Soit X une variable aléatoire qui suit la loi normale de moyenne $\theta \in \mathbb{R}$ et de variance 1 dont la densité est notée $x \mapsto f(\theta, x)$. On définit sous réserve de convergence l'**information de Fisher** de X par

$$I_X(\theta) = \int_{-\infty}^{+\infty} \left(\frac{\partial}{\partial \theta} \ln(f(\theta, x)) \right)^2 f(\theta, x) dx.$$

- (a) Montrer que sous réserve de convergence

$$I_X(\theta) = \int_{-\infty}^{+\infty} (x - \theta)^2 f(\theta, x) dx.$$

- (b) En déduire l'existence et la valeur de $I_X(\theta)$.

- (c) Justifier que

$$I_X(\theta) = E \left(\left(\frac{\partial}{\partial \theta} \ln(f(\theta, X)) \right)^2 \right).$$

IV. Minoration du risque quadratique

17. **Inégalité de Cramer-Rao.** Dans cette question, on considère I un intervalle de \mathbb{R} , θ un paramètre inconnu de I et X une variable aléatoire telle que $X(\Omega) = \{0, \dots, N\}$ ($N \in \mathbb{N}$). On suppose qu'il existe une fonction p définie sur $I \times X(\Omega)$ telle que pour tout $k \in \{0, \dots, N\}$

$$P(X = k) = p(\theta, k)$$

et vérifiant :

- pour tout $k \in \{0, \dots, N\}$, $\theta \mapsto p(\theta, k)$ dérivable sur I ,
- l'information de Fisher de X notée $I_X(\theta)$ définie dans la partie III est non nulle pour tout $\theta \in I$.

Le but de la question est de démontrer l'inégalité suivante due à Cramer et Rao :

Théorème 1 (de Cramer-Rao)

Soit $f(X)$ un estimateur sans biais de $g(\theta)$ où g est dérivable sur I . On a alors :

$$V(f(X)) \geq \frac{(g'(\theta))^2}{I_X(\theta)}.$$

- (a) Montrer que pour tout θ élément de I on a : $\sum_{k=0}^N \frac{\partial}{\partial \theta} (p(\theta, k)) = 0$.

- (b) En déduire pour tout θ élément de I : $E \left(\frac{\partial}{\partial \theta} \ln(p(\theta, X)) \right) = 0$.

- (c) En dérivant partiellement par rapport à θ les deux membres de l'égalité de la question précédente, montrer que pour tout θ élément de I

$$E \left(\frac{\partial^2}{\partial \theta^2} \ln(p(\theta, X)) \right) = -E \left(\left(\frac{\partial}{\partial \theta} \ln(p(\theta, X)) \right)^2 \right).$$

- (d) Montrer que pour tout θ élément de I

$$g'(\theta) = \sum_{k=0}^N f(k) \left(\frac{\partial}{\partial \theta} \ln(p(\theta, k)) \right) p(\theta, k)$$

puis que

$$g'(\theta) = E \left((f(X) - g(\theta)) \frac{\partial}{\partial \theta} \ln(p(\theta, X)) \right).$$

- (e) On pose pour tout t réel

$$L(t) = E \left(\left((f(X) - g(\theta)) + t \frac{\partial}{\partial \theta} \ln(p(\theta, X)) \right)^2 \right).$$

- i. Développer le polynôme L suivant les puissances décroissantes de t .
- ii. Calculer le discriminant Δ de L et justifier que $\Delta \leq 0$.
- iii. En déduire l'inégalité de Cramer-Rao.

18. **Extension du théorème de Cramer-Rao.** On reprend dans cette question les notations et hypothèses de la partie II. On admet que, dans ce contexte, le théorème de Cramer-Rao peut se généraliser comme suit :

Théorème 2

Soit $T_n = f(X_1, \dots, X_n)$ un estimateur sans biais de $g(\theta)$ où g est dérivable sur $]0, +\infty[$. On a alors :

$$V(T_n) \geq \frac{(g'(\theta))^2}{n I_{X_1}(\theta)}$$

où $I_{X_1}(\theta)$ est l'information de Fisher d'une variable aléatoire de loi de Poisson de paramètre θ définie et calculée à la partie III.

Il s'agit dans cette section d'exploiter cette nouvelle inégalité de Cramer-Rao.

- (a) Calculer $E(\overline{X}_n)$ et $V(\overline{X}_n)$.
- (b) Déduire de la généralisation de Cramer-Rao, que \overline{X}_n a le plus petit risque quadratique parmi les estimateurs sans biais de θ .

(c) Montrer que pour $g(\theta) = \exp(-\theta)$ où $\theta \in]0, +\infty[$ on a :

$$V(\varphi(S_n)) \underset{n \rightarrow +\infty}{\sim} \frac{(g'(\theta))^2}{nI_{X_1}(\theta)}.$$

(d) Que prouve ce résultat en terme d'optimalité de $\varphi(S_n)$ dans l'estimation de $\exp(-\theta)$?

(e) A la lumière de la partie I, peut-on conclure que lorsque n est grand $\varphi(S_n)$ est le meilleur estimateur de $\exp(-\theta)$ en terme de risque quadratique ?