Estimation

Estimation ponctuelle

Exercice 1 (élémentaire et incontournable)

Soit X une variable suivant une loi de Bernoulli de paramètre p inconnu. Pour n entier naturel non nul, on désigne par $(X_1,\dots,X_n)$ un n-échantillon de variables aléatoires indépendantes de même loi que X. On note $\overline{X_n}=\frac{X_1+\cdots+X_n}{n}$.

1. Montrer que $\overline{X_n}$ est un estimateur de p.
2. Montrer que l'estimateur $\overline{X_n}$ a une espérance égale à p.

Exercice 2 (Pour voir un peu toutes les notions de l'estimation ponctuelle sur un exemple simple)

On considère la variable aléatoire X dont la loi est donnée par : $\mathrm{P}(X=-1)=p \quad \mathrm{P}(X=0)=1-2p \quad \mathrm{P}(X=1)=p$

Pour un certain paramètre $p\in\left]0,\frac{1}{2}\right[$. On dispose d'un n-échantillon $(X_1,\dots,X_n)$ de X, et on cherche à déterminer le paramètre p.

1. Calculer $\mathrm{E}(\overline{X_n})$.
2. Peut-on trouver les réels a et b tels que aX_n+b soit un estimateur sans biais de p (c'est à dire dont l'espérance est égale à p) ?
3. On note $T_n=\frac{1}{2n}\sum\limits_{i=1}^{n}X_i^2$. Montrer que T_n est un estimateur sans biais de de p. Calculer sa variance et montrer qu'il est convergent.

Exercice 3

Soit X une variable aléatoire admettant une variance σ^2. On désigne par $(X_1,\dots,X_n)$ un n-échantillon de variables aléatoires indépendantes de la loi de X et on pose $\overline{X_n}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$ et $T_n=\frac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i-\overline{X_n}\right)^2$.

Montrer que $\mathrm{E}(T_n)=\sigma^2$.

Exercice 4 (Un autre incontournable : Estimateur du maximum de vraisemblance : cas de la loi de Poisson)

Soit $X_1,\dots,X_n$ des variables aléatoires mutuellement indépendantes suivant une loi de Poisson de paramètre $\lambda>0$.

On pose $f_{x_1,\dots,x_n}(\lambda)=\mathrm{P}(X_1=x_1,\dots,X_n=x_n)$

1. Donner une expression de $f_{x_1,\dots,x_n}(\lambda)$.
2. Pour quelle valeur de λ la fonction $f_{x_1,\dots,x_n}$ est-elle maximale ?

Exercice 5 (Estimation de la variance)

Soit $(X_1,\dots,X_n)$ un n-échantillon d'une v.a. X admettant pour espérance m et pour variance σ^2.

1. On suppose que m est connu. Montrer que T_n est un estimateur de σ^2, où $T_n=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-m)^2$ et calculer son espérance.
2. On suppose que m n'est pas connu. On note $\overline{S_n}^2$ la variance empirique, où $\overline{S_n}^2=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X_n})^2$
   - **a.** Montrer que, pour tout $i\in[\![1,n]\!]$, $\mathrm{E}((X_i-\overline{X_n})^2)=\mathrm{V}(X_i-\overline{X_n})$
   - **b.** D'autre part, montrer que, pour tout $i\in[\![1,n]\!]$, $\mathrm{E}((X_i-\overline{X_n})^2)=\mathrm{V}(X_i)-2\mathrm{E}\big((X_i-m)(\overline{X_n}-m)\big)+\mathrm{V}(\overline{X_n})$ (on pourra ajouter et retrancher m dans le carré)
   - **c.** En déduire l'espérance de la variance empirique. En déduire un estimateur sans biais de σ^2.

Exercice 6 (Convergence d'un estimateur)

Soit a un réel strictement positif. On note f la fonction définie sur $\mathbb{R}$ par $f(x)=\begin{cases}0 & \text{si } x<a\\ \frac{3a^3}{x^4} & \text{si } x\geq a\end{cases}$

1. Montrer que f est une densité de probabilité d'une var T et que T admet une espérance et une variance que l'on calculera.
2.
   - **a.** Déterminer la fonction de répartition de T.
   - **b.** Calculer les probabilités $\mathrm{P}(T>2a)$ et $\mathrm{P}_{(T>2a)}(T>6a)$.
3. Montrer que la variable $Z_n=\frac{2}{3n}\sum\limits_{k=1}^{n}T_k$ est un estimateur de a dont on calculera l'espérance. Est-il convergent ?

Estimation par intervalle de confiance

Exercice 7 (Grâce à l'inégalité de BT)

On suppose que le paramètre p, qui exprime la proportion de votants pour un candidat au second tour d'une élection, est inconnu, et on cherche à l'estimer. Pour tout entier n supérieur ou égal à 1, on considère un n-échantillon $(X_1,\dots,X_n)$ de v.a.i.i.d de même loi de Bernoulli p. On pose $\overline{X_n}$ la moyenne empirique. Calculer l'espérance et l'écart-type de $\overline{X_n}$ et à l'aide de l'inégalité de Bienaymé-Tchebychev, montrer que $\left[\overline{X_n}-\sqrt{\frac{5}{n}}\ ,\ \overline{X_n}+\sqrt{\frac{5}{n}}\right]$ est un intervalle de confiance de p au niveau de confiance 0,95.

Exercice 8

On suppose que la charge maximale supportée par un câble, exprimée en tonnes, est une v.a. qui suit une loi normale d'écart-type 0,5 et de moyenne inconnue. Une étude portant sur 50 câbles a donné une moyenne des charges maximales supportées de 12,2 tonnes.

1. Déterminer un intervalle de confiance à 99% de la charge maximale de tous les câbles fabriqués par l'usine.
2. Déterminer la taille minimale de l'échantillon étudié pour que la longueur de l'intervalle de confiance à 99% soit inférieure ou égale à 0,2.

Exercice 9

Lors d'un sondage pré-électoral, 1000 électeurs choisis au hasard ont été interrogés. 520 d'entre eux se sont déclarés favorables à Mme Durand.

1. Déterminer un intervalle de confiance à 95% de la proportion p d'électeurs favorables à Mme Durand dans la population.
2. Quel doit être le nombre minimal d'électeurs interrogés pour que l'intervalle de confiance que l'on puisse en déduire soit de longueur inférieure ou égale à 0,02 ?

Problèmes récapitulatifs et annales

Exercice 10 (Un bon exercice récapitulatif)

Soient θ un réel strictement positif et $(X_1,\dots,X_n)$ un $n-$échantillon d'une variable $X \hookrightarrow \mathscr{U}([0,\theta])$.

1. Montrer que $\overline{X_n}$ est un estimateur de θ et calculer son espérance.
2. Proposer un estimateur V_n de θ dont l'espérance est θ.
3. On considère maintenant l'estimateur $M_n = \max(X_1,\dots,X_n)$.
   - **a.** Déterminer la fonction de répartition F de M_n et en déduire une densité f de M_n.
   - **b.** Montrer alors que $E(M_n) = \frac{n\theta}{n+1}$
   - **c.** En déduire un estimateur Z_n de θ dont l'espérance est θ à partir de M_n.
4.
   - **a.** Établir que $\mathrm{V}(V_n) = \frac{\theta^2}{3n}$
   - **b.** Montrer que $\mathrm{E}\left(Z_n^2\right) = \frac{(n+1)^2}{n^2}\displaystyle\int_0^\theta \frac{nt^{n+1}}{\theta^n}\,\mathrm{d}t$
   - **c.** En déduire que $\mathrm{V}(Z_n) = \frac{\theta^2}{n(n+2)}$
   - **d.** Quel estimateur aura-t-on tendance à préférer en pratique pour des grandes valeurs de n ?
5. Soit $(x_1,\dots,x_n) \in \left(\mathbb{R}_+^*\right)^n$ fixé. On note f_θ la densité de X.

   On introduit la fonction de vraisemblance, définie sur $\mathbb{R}_*^+$ par $L_n(\theta) = \displaystyle\prod_{i=1}^n f_\theta(x_i)$.
   - **a.** Montrer que, pour tout $\theta \geq 0$, on a $L_n(\theta) = \begin{cases} \theta^{-n} & \text{si } \theta \geq \max(x_1,\dots,x_n) \\ 0 & \text{sinon} \end{cases}$
   - **b.** En déduire que l'estimateur du maximum de vraisemblance pour la loi $\mathscr{U}\left([0,\theta]\right)$ est donné par $\hat{\theta}_n = \max(X_1,\dots,X_n)$
6.
   - **a.** Montrer à l'aide de l'inégalité de Bienaymé-Tchebychev que $\mathrm{P}\left(|V_n - \theta| > \varepsilon\right) \leq \frac{\theta^2}{3n\varepsilon^2}$ (V_n est donc un estimateur convergent de θ)
   - **b.** Montrer que $\theta \in \left[V_n - \sqrt{\frac{\theta^2}{3n\alpha}} ; V_n + \sqrt{\frac{\theta^2}{3n\alpha}}\right] \quad \Leftrightarrow \quad \dfrac{V_n}{1+\frac{1}{\sqrt{3n\alpha}}} \leq \theta \leq \dfrac{V_n}{1-\frac{1}{\sqrt{3n\alpha}}}$
   - **c.** En déduire un intervalle de confiance au risque α pour θ.

Exercice 11 (EDHEC 2012, avec estimateur du max de vraisemblance, cas variable à densité entres autres)

On pourra aussi traiter le problème de l'année 2012 de l'EDHEC en remplaçant les questions portant sur Scilab par des questions portant sur Python :

5.
   - **b.** En déduire une fonction python dont l'entête est `def vax(lambda):` qui simule la loi de $|X|$.
   - **c.** Vérifier que la probabilité que X prenne des valeurs positives est égale à la probabilité que X prenne des valeurs positives est égale à la probabilité que X prenne des valeurs négatives.

     En déduire une fonction Python dont l'entête est `def Simulx(lambda):` qui simule la loi de X.