LSTAT 2040 - TP 2

Estimation ponctuelle

Exercice 1

Soit $X \sim \mathrm{N}(\mu, \sigma^2)$ et $Y \sim \mathrm{N}(\mu, 4\sigma^2)$, avec μ et σ^2 inconnus et $X \perp\!\!\!\perp Y$. Soit aussi $W = aX + bY$.

(a) Quelles valeurs peut-on donner à a et b pour que W soit un estimateur sans biais de μ ?

(b) Parmi toutes les valeurs possibles de a et b pour que W soit un estimateur sans biais de μ, lesquelles préfèrent-on ? Pourquoi ?

Exercice 2

Soit $X_1, \ldots, X_n$ un échantillon aléatoire iid issu de la fonction de densité suivante :

$$f(x, \theta) = \exp\big(-(x - \theta)\big)\mathrm{I}(x \geq \theta),$$

où $\theta > 0$. Montrer que $\min(X_1, \ldots, X_n)$ est un estimateur de θ dont le biais tend vers 0.

Exercice 3

Soit $X_1, \ldots, X_n$ un échantillon aléatoire iid issu de la fonction de densité suivante :

$$f(x, \theta) = \frac{2x}{\theta^2}\mathrm{I}(0 < x < \theta),$$

où $\theta > 0$. Soit $\overline{X}_n$ la moyenne empirique et $X_{(n)} = \max(X_1, \ldots, X_n)$. Montrer que $\widehat{\theta}_1 = \frac{3\overline{X}_n}{2}$ et $\widehat{\theta}_2 = \frac{2n+1}{2n}X_{(n)}$ sont deux estimateurs non biaisés de θ et comparer leurs variances.

Exercice 4

Soit un échantillon iid $X_1, \ldots, X_n$ de distribution commune $\mathrm{Exp}(\lambda)$.

(a) Proposer un estimateur du paramètre λ.

(b) Calculer la densité et le risque quadratique (MSE) de cet estimateur.
Indice : voir le formulaire pour la distribution d'une somme de variables aléatoires iid distribuées selon une loi exponentielle.
Indice : effectuer une intégration par substitution et utiliser la définition (et les propriétés) de la fonction Γ (voir le formulaire).

(c) Sur base de ces résultats, donner un estimateur sans biais de λ.

Exercice 5

Soit $X_1, \ldots, X_n$ un échantillon iid de distribution commune $\mathrm{Unif}[a, b]$.

(a) Proposer un estimateur pour a et un estimateur pour b.

(b) Calculer la distribution de ces deux estimateurs.

(c) Calculer le biais et la variance de ces deux estimateurs. Expliquer pourquoi le biais qu'on obtient est logique.
Indice : dans les différentes intégrales à calculer, si x désigne la variable d'intégration, effectuer des substitutions comme $u = x - a$ ou $v = b - x$.

(d) Dans le cas particulier où $b = 2a$, considérer un estimateur pour a de la forme $\alpha X_{(1)} + \beta X_{(n)}$. Quelles valeurs peut-on donner à α et à β pour que cet estimateur soit sans biais ?

Exercice 6

Soit $Y \sim \text{Bin}(n, p)$. Définissons $\widehat{p}_1 = \frac{Y}{n}$ et $\widehat{p}_2 = \frac{Y+1}{n+2}$ deux estimateurs de p.

(a) Calculer $\text{MSE}(\widehat{p}_1)$ et $\text{MSE}(\widehat{p}_2)$.
(b) Pour quelles valeurs de p a-t-on $\text{MSE}(\widehat{p}_1) < \text{MSE}(\widehat{p}_2)$?

Exercice 7

Soit $X_1, \ldots, X_n$ un échantillon iid dont la fonction de répartition commune est notée F. Sans hypothèse paramétrique sur la distribution des données, si nous souhaitons estimer $F(x)$, une manière raisonnable d'effectuer l'estimation est de considérer l'estimateur

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \text{I}(X_i \leq x),$$

i.e., la fonction de répartition empirique évaluée en x.

(a) L'estimateur $\widehat{F}_n(x)$ est-il sans biais ? Calculez sa MSE.
(b) Montrez que la MSE de $\widehat{F}_n(x)$ est maximale quand x est la médiane de F.

Exercice 8

Soit $X \sim \text{N}(\theta, 1)$ et soit la fonction de perte (loss function) résultant de l'estimation de θ par $\widehat{\theta}(X)$ définie par

$$L(\widehat{\theta}(X), \theta) = \begin{cases} a(\widehat{\theta}(X) - \theta) & \text{si } \widehat{\theta}(X) \geq \theta \\ b(\theta - \widehat{\theta}(X)) & \text{si } \widehat{\theta}(X) \leq \theta, \end{cases}$$

avec $a > 0$ et $b > 0$.

(a) Montrer que le risque associé à l'estimateur $\widehat{\theta}_k(X) = X - k$ peut s'écrire sous la forme :

$$(a+b)[\phi(k) + k\Phi(k)] - ka,$$

où ϕ et Φ sont respectivement la fonction de densité et la fonction de répartition d'une variable aléatoire $\text{N}(0, 1)$.

(b) Montrer que dans la classe d'estimateurs $\{\widehat{\theta}_k : k \in \mathbb{R}\}$, il existe un estimateur qui minimise le risque uniformément en θ, et que le risque minimal est donné par :

$$(a+b)\phi\left[\Phi^{-1}\left(\frac{a}{a+b}\right)\right].$$