LSTAT 2040 - TP 2 : Solutions

Estimation ponctuelle

Exercice 1

(a) Comme $\mathrm{E}[W] = (a+b)\mu$, W est sans biais pour μ dès que $a+b=1$.

(b) On cherche les valeurs de a et de b telles que $a+b=1$ et qui minimisent le risque quadratique de W donné par

$$\mathrm{MSE}(W) = \mathrm{Bias}(W)^2 + \mathrm{Var}(W) = \mathrm{Var}(W) = a^2\sigma^2 + 4b^2\sigma^2 = (a^2+4b^2)\sigma^2.$$

Il faut donc minimiser la fonction $(a,b) \in \mathbb{R}^2 \mapsto a^2 + 4b^2$ sous la contrainte $b = 1-a$. En substituant la contrainte dans le problème, on trouve que l'on doit minimiser la fonction $a \in \mathbb{R} \mapsto a^2 + 4(1-a)^2 = 5a^2 - 8a + 4$, et par dérivation, on voit que son minimum est atteint en $a = 4/5$. On déduit la valeur de $b = 1/5$ de la contrainte. Le choix $(a,b) = (4/5, 1/5)$ est donc la meilleur choix à effectuer si l'on souhaite un estimateur sans bias pour μ de variance minimale.

Exercice 2

Il faut montrer que

$$\mathrm{E}[X_{(1)}] \to \theta, \qquad n \to \infty.$$

Afin de calculer l'espérance, on commence par trouver la distribution de $X_{(1)}$. Pour cela, on observe que pour tout $x \geq \theta$,

$$\Pr(X_{(1)} \geq x) = \Pr(X_1 \geq x)^n,$$

car l'échantillon est iid. Dès lors, pour $x \geq \theta$,

$$f_{X_{(1)}}(x) = \frac{d}{dx}\Pr(X_{(1)} \leq x) = n\Pr(X_1 \geq x)^{n-1} f(x,\theta).$$

Par intégration, on trouve $\Pr(X_1 \geq x) = \exp(-(x-\theta))$ pour $x \geq \theta$. Comme $f_{X_{(1)}} \equiv 0$ sur $(-\infty, \theta)$, on déduit

$$f_{X_{(1)}}(x) = n\exp(-n(x-\theta))\mathrm{I}(x \geq \theta).$$

De là, on calcule

$$\mathrm{E}[X_{(1)}] = \int_{\mathbb{R}} x f_{X_{(1)}}(x)\,dx = \theta + \frac{1}{n},$$

et la conclusion suit.

Exercice 3

Pour $\widehat{\theta}_1$, on calcule facilement que

$$\mathrm{E}[\widehat{\theta}_1] = \frac{3}{2}\,\mathrm{E}[\overline{X}_n] = \frac{3}{2}\,\mathrm{E}[X_1] = \theta.$$

Pour $\widehat{\theta}_2$, on a

$$\mathrm{E}[\widehat{\theta}_2] = \frac{2n+1}{2n}\,\mathrm{E}[X_{(n)}],$$

et nous avons besoin de $f_{X_{(n)}}$ pour calculer l'espérance. On a pour tout $0 < x < \theta$,

$$f_{X_{(n)}}(x) = \frac{d}{dx}\Pr(X_{(n)} \leq x) = \frac{d}{dx}\Pr(X_1 \leq x)^n = n\Pr(X_1 \leq x)^{n-1}f(x,\theta) = \frac{2n}{\theta^{2n}}x^{2n-1},$$

et $f_{X_{(n)}} \equiv 0$ ailleurs. On déduit

$$\mathrm{E}[X_{(n)}] = \int_{\mathbb{R}} x f_{X_{(n)}}(x)\,dx = \frac{2n}{2n+1}\theta,$$

et donc $\mathrm{E}[\widehat{\theta}_2] = \theta$.

Pour les variances, on a

$$\mathrm{Var}[\widehat{\theta}_1] = \frac{9}{4}\mathrm{Var}[\overline{X}_n] = \frac{9}{4n}\mathrm{Var}[X_1] = \frac{\theta^2}{8n}.$$

Pour $\widehat{\theta}_2$, on a

$$\mathrm{Var}[\widehat{\theta}_2] = \frac{(2n+1)^2}{(2n)^2}\mathrm{Var}[X_{(n)}]$$

où

$$\mathrm{Var}[X_{(n)}] = \int_{\mathbb{R}} x^2 f_{X_{(n)}}(x)\,dx - \left(\int_{\mathbb{R}} x f_{X_{(n)}}(x)\,dx\right)^2.$$

Après calculs, on trouve $\mathrm{Var}[\widehat{\theta}_2] = \frac{\theta^2}{4n(n+1)}$. Dès lors, dès que $n > 1$, on aura $\mathrm{Var}[\widehat{\theta}_2] < \mathrm{Var}[\widehat{\theta}_1]$.

Exercice 4

(a) Comme $\mathrm{E}[X_1] = 1/\lambda$, on propose l'estimateur $\widehat{\lambda} = 1/\overline{X}_n$.

(b) Comme suggéré dans le premier indice, il faut remarquer que

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^n X_i}, \qquad \sum_{i=1}^n X_i \sim \mathrm{Gamma}(n,\lambda).$$

Dès lors, pour tout $x \geq 0$,

$$\begin{aligned} f_{\widehat{\lambda}}(x) &= \frac{d}{dx}\Pr\left(\frac{n}{\sum_{i=1}^n X_i} \leq x\right) = \frac{d}{dx}\Pr\left(\sum_{i=1}^n X_i \geq \frac{n}{x}\right) \\ &= \frac{n}{x^2}f_{\sum_{i=1}^n X_i}(n/x) = \frac{\lambda^n}{(n-1)!}\left(\frac{n}{x}\right)^n\frac{1}{x}\exp(-\lambda n/x). \end{aligned}$$

On peut alors calculer le risque quadratique de l'estimateur

$$\mathrm{MSE}[\widehat{\lambda}] = \mathrm{Bias}[\widehat{\lambda}]^2 + \mathrm{Var}[\widehat{\lambda}].$$

Pour le calcul du biais, on doit obtenir $\mathrm{E}[\widehat{\lambda}]$. On a,

$$\mathrm{E}[\widehat{\lambda}] = \frac{\lambda^n}{(n-1)!}n^n\int_0^\infty x^{-n}\exp(-\lambda n/x)\,dx = \frac{\lambda n}{(n-1)!}\int_0^\infty u^{n-2}\exp(-u)\,du,$$

où la seconde égalité est obtenue par le changement de variable $u = \lambda n/x$. Par le second indice, on remarque que $\int_0^\infty u^{n-2}\exp(-u)\,du = \Gamma(n-1) = (n-2)!$. On déduit

$$\mathrm{E}[\widehat{\lambda}] = \lambda\frac{n}{n-1}.$$

Pour la variance, un calcul similaire donne

$$\mathrm{Var}[\widehat{\lambda}] = \frac{n^2\lambda^2}{(n-1)^2(n-2)}.$$

On déduit que le risque quadratique est donné par

$$\mathrm{MSE}[\widehat{\lambda}] = \frac{(n+2)\lambda^2}{(n-1)(n-2)}.$$

(c) On voit directement que, comme $\mathrm{E}[\widehat{\lambda}] = \lambda n/(n-1)$, l'estimateur

$$\tilde{\lambda} = \frac{n-1}{n}\widehat{\lambda},$$

sera sans biais pour λ.

Exercice 5

(a) Intuitivement, on a envie de proposer

$$\widehat{a} = X_{(1)} \quad \text{et} \quad \widehat{b} = X_{(n)}.$$

(b) Par des raisonnements similaires à ce que l'on a déjà fait dans les exercices précédents, on trouve

$$f_{\widehat{a}}(x) = \frac{n}{(b-a)^n}(b-x)^{n-1}\mathrm{I}(a \leq x \leq b)$$

et

$$f_{\widehat{b}}(x) = \frac{n}{(b-a)^n}(x-a)^{n-1}\mathrm{I}(a \leq x \leq b).$$

(c) En utilisant les densités trouvées au point précédent et l'indice pour calculer les intégrales, on trouve

$$\mathrm{Bias}[\widehat{a}] = \frac{b-a}{n+1} \qquad \mathrm{Var}[\widehat{a}] = \frac{n(b-a)^2}{(n+1)^2(n+2)}$$

et

$$\mathrm{Bias}[\widehat{b}] = -\frac{b-a}{n+1} \qquad \mathrm{Var}[\widehat{b}] = \frac{n(b-a)^2}{(n+1)^2(n+2)}.$$

Le signe des biais peut s'expliquer comme suit : puisque toutes les données X_i de l'échantillon sont strictement supérieures à a et inférieures à b avec probabilité 1, le minimum de celles-ci $X_{(1)} = \widehat{a}$ aura tendance à sur-estimer a, le biais est donc positif, tandis que $X_{(n)} = \widehat{b}$ aura tendance à sous-estimer b, le biais est donc négatif.

(d) Dans le cas où $b = 2a$, on a, par nos calculs précédents,

$$\mathrm{E}[\alpha X_{(1)} + \beta X_{(n)}] = a\left(\alpha\Big(2 - \frac{n}{n+1}\Big) + \beta\Big(2 - \frac{1}{n+1}\Big)\right).$$

Si on veut que l'estimateur soit sans biais pour a, il faut donc choisir (α, β) tel que

$$\alpha\Big(2 - \frac{n}{n+1}\Big) + \beta\Big(2 - \frac{1}{n+1}\Big) = 1,$$

autrement dit, pour une valeur de α, il faut prendre

$$\beta = \frac{n+1-\alpha(n+2)}{2n+1}.$$

Exercice 6

(a) On calcule

$$\mathrm{MSE}[\widehat{p}_1] = \mathrm{Bias}[\widehat{p}_1]^2 + \mathrm{Var}[\widehat{p}_1] = 0^2 + \frac{p(1-p)}{n} = \frac{p(1-p)}{n}$$

et

$$\mathrm{MSE}[\widehat{p}_2] = \mathrm{Bias}[\widehat{p}_2]^2 + \mathrm{Var}[\widehat{p}_2] = \left(\frac{1-2p}{n+2}\right)^2 + \frac{np(1-p)}{(n+2)^2} = \frac{(1-2p)^2 + np(1-p)}{(n+2)^2}.$$

(b) Il faut résoudre l'inégalité $\mathrm{MSE}[\widehat{p}_1] < \mathrm{MSE}[\widehat{p}_2]$ pour p. On trouve que cette inégalité est satisfaite pour

$$p < \frac{1}{2} - \sqrt{\frac{n+1}{8n+4}} \quad \text{ou} \quad p > \frac{1}{2} + \sqrt{\frac{n+1}{8n+4}}.$$

Exercice 7

(a) Notons que les variables aléatoires $\mathrm{I}(X_i \leq x)$ sont iid et distribuées selon une distribution $\mathrm{Be}(F(x))$ puisqu'elle ne prennent que les valeurs 0 ou 1 et la probabilité de “succès” est

$$\mathrm{E}[\mathrm{I}(X_i \leq x)] = \Pr(X_i \leq x) = F(x).$$

Dès lors, on trouve

$$\mathrm{E}[\widehat{F}_n(x)] = \mathrm{E}[\mathrm{I}(X_i \leq x)] = F(x)$$

et l'estimateur est sans biais. Pour calculer sa MSE, on calcule

$$\mathrm{MSE}[\widehat{F}_n(x)] = \mathrm{Var}[\widehat{F}_n(x)] = \frac{\mathrm{Var}[\mathrm{I}(X_1 \leq x)]}{n} = \frac{F(x)(1-F(x))}{n}.$$

(b) Le point $x \in \mathbb{R}$ où la MSE de $\widehat{F}_n(x)$ est maximale est le point x^* solution du problème

$$x^* = \operatorname*{argmax}_{x\in\mathbb{R}} \frac{F(x)(1-F(x))}{n} = \operatorname*{argmax}_{x\in\mathbb{R}} F(x)(1-F(x)).$$

Si on note $p^* = F(x^*) \in [0,1]$, alors p^* satisfait

$$p^* = \operatorname*{argmax}_{p\in[0,1]} p(1-p) = 1/2,$$

où le problème peut être résolu par dérivation. On trouve alors que x^* est tel que $F(x^*) = 1/2$, c'est-à-dire, x^* est la médiane de F.

Exercice 8

(a) Par défintion, le risque associé à $\widehat{\theta}_k = X - k$ est

$$\mathrm{E}[L(\widehat{\theta}_k, \theta)] = \int_{\mathbb{R}} L(z,\theta) f_{\widehat{\theta}_k}(z)\, dz.$$

Puisque $X \sim \mathrm{N}(\theta, 1)$, on a $\widehat{\theta}_k = X - k \sim \mathrm{N}(\theta - k, 1)$. Donc, l'intégrale vaut

$$\begin{aligned}
&\int_{\mathbb{R}} L(z,\theta) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(z-(\theta-k))^2\right) dz \\
&= \int_{-\infty}^{\theta} b(\theta - z) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(z-(\theta-k))^2\right) dz + \int_{\theta}^{\infty} a(z-\theta) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(z-(\theta-k))^2\right) dz.
\end{aligned}$$

On considère le même changement de variable $x = z - (\theta - k)$. On trouve alors,

$$\begin{aligned}
&\int_{-\infty}^{k} b(k-x)\phi(x)\, dx + \int_{k}^{\infty} a(x-k)\phi(x)\, dx \\
&= b\left(k\Phi(k) + \int_{-\infty}^{k} -x\phi(x)\, dx\right) + a\left(-\int_{k}^{\infty} -x\phi(x)\, dx - k(1-\Phi(k))\right) \\
&= b\left(k\Phi(k) + \int_{-\infty}^{k} \phi'(x)\, dx\right) + a\left(-\int_{k}^{\infty} \phi'(x)\, dx - k(1-\Phi(k))\right) \\
&= b\left(k\Phi(k) + \phi(k)\right) + a\left(\phi(k) - k(1-\Phi(k))\right) \\
&= (a+b)[\phi(k) + k\Phi(k)] - ka.
\end{aligned}$$

(b) Le valeur de k^* minimisant le risque satisfait

$$\frac{d(a+b)[\phi(k) + k\Phi(k)] - ka}{dk}(k^*) = (a+b)\Phi(k^*) - a = 0.$$

En résolvant l'équation, on trouve donc

$$k^* = \Phi^{-1}\left(\frac{a}{a+b}\right).$$

Notons qu'en particulier, si on avait $a = b$ et que la fonction de perte était symétrique par rapport aux sous-estimations et aux sur-estimations de θ, on aurait $k^* = \Phi^{-1}(0.5) = 0$. Le risque minimal est donné par

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