Université de Pau et des Pays de l'Adour — **Semestre printemps 2023–2024**

L3 Math — Introduction à la Statistique

CC 2 - Estimation statistique et Intervalle de confiance Correction

---

Exercice 1 Autour de la loi de Laplace (13pts)

Soit X une variable aléatoire réelle de fonction de densité pour $x \in \mathbb{R}$,

$$f_\theta(x) = C \exp(-2|x|/\theta),$$

avec $\theta > 0$ un paramètre inconnu et C une constante dépendant de θ. On se place dans le cadre de n observations $(X_1, \ldots, X_n)$ i.i.d.

Partie 1 : Construction de nos estimateurs

1. (1pt) Décrire le modèle statistique de cet exercice et pour $\theta > 0$, montrer que $C = 1/\theta$.

La parité de la densité implique que

$$\int_{\mathbb{R}} f_\theta(x)\ \mathrm{d}x = 2\int_0^{+\infty} f_\theta(x)\ \mathrm{d}x.$$

2. (2pts) Calculer l'espérance et la variance de la variable X sous Q_θ.

Indication : on pourra utiliser la symétrie de la densité pour simplifier les calculs.

Comme f_θ est pair alors $x \mapsto x f_\theta(x)$ est impair. Par croissance comparée, cette fonction est intégrable. Donc l'imparité implique que son air est nulle (faire changement de variable pour s'en convaincre). Donc

$$\mathbb{E}_\theta[X] = 0.$$

Pour la variance, la parité implique qu'il faut donc calculer

$$\begin{aligned}
\mathbb{E}_\theta[X^2] = 2\int_0^{+\infty} x^2 f_\theta(x)\ \mathrm{d}x &= \int_0^{+\infty} \frac{2x^2}{\theta}\exp(-2x/\theta)\ \mathrm{d}x \\
&= [-x^2\exp(-2x/\theta)]_{x=0}^{+\infty} + \int_0^{+\infty} 2x\exp(-2x/\theta)\ \mathrm{d}x \\
&= [-\theta x\exp(-2x/\theta)]_{x=0}^{+\infty} + \theta\int_0^{+\infty} \exp(-2x/\theta)\ \mathrm{d}x \\
&= [-\theta^2/2\exp(-2x/\theta)]_{x=0}^{+\infty} = \theta^2/2.
\end{aligned}$$

Chaque IPP est légale car toutes les intégrales existent par croissance comparée. Ainsi la variance vaut $\theta^2/2$.

3. (1pt) Déterminer un estimateur $\hat{\theta}_{MM}$ par la méthode des moments.

Comme l'espérance est nulle, on ne va l'utiliser pour construire un estimateur. On utilise donc les moments d'ordre 2. Ici, on a

$$\mathbb{E}_\theta[X^2] = \mathrm{Var}_\theta(X),$$

on peut donc utiliser $U_2(n)$ ou S_n^2 pour construire un EMM. Utilisons la variance empirique, il faut donc résoudre

$$S_n^2 = \hat{\theta}_{MM}^2/2,$$

donc $\hat{\theta}_{MM} = \sqrt{2\S_n^2}$.

Comme $\mathbb{E}_\theta[X^2] = \mathrm{Var}_\theta(X)$, on peut construire l'estimateur $\hat{\theta}_{MM}$ soit avec la variance empirique S_n^2 soit avec le moment d'ordre 2 empirique $U_2(n)$.

4. (2pts) Montrer que l'estimateur de maximum de vraisemblance est

$$\hat{\theta}_{MV} = \frac{2}{n}\sum_{i=1}^n |X_i|.$$

La log-vraisemblance de ce modèle est pour $x_1, \dots, x_n \in \mathbb{R}$ et $\theta > 0$,

$$\log L_n(x_1, \dots, x_n; \theta) = -n\log\theta - \frac{2}{\theta}\sum_{i=1}^n |x_i|.$$

Donc en dérivant, on obtient

$$\partial_\theta \log L_n(x_1, \dots, x_n; \theta) = -\frac{n}{\theta} + \frac{2}{\theta^2}\sum_{i=1}^n |x_i|.$$

Une étude de signe montre que l'unique point critique est un maximum et le point critique est atteint pour

$$-\frac{n}{\theta} + \frac{2}{\theta^2}\sum_{i=1}^n |x_i| = 0 \Leftrightarrow \theta = \frac{2}{n}\sum_{i=1}^n |x_i|.$$

Partie 2 : Propriétés de l'estimateur de maximum de vraisemblance

5. (1pt) Montrer que l'EMV est sans biais.

Pour $\theta > 0$, calculons en utilisant la parité

$$\mathbb{E}_\theta[|X_1|] = \int_0^{+\infty} \frac{2x}{\theta}\exp(-2x/\theta)\,\mathrm{d}x = \dots = \theta/2.$$

Comme $\hat{\theta}_{MV} = \frac{2}{n}\sum_{i=1}^n |X_i|$, on obtient bien que

$$\mathbb{E}_\theta[\hat{\theta}_{MV}] = \frac{2\sum_{i=1}^n \mathbb{E}_\theta[|X_i|]}{n} = \theta.$$

6. (1pt) Montrer que l'EMV est fortement consistant.

La Loi des Grands Nombres nous dit que presque sûrement

$$\frac{1}{n}\sum_{i=1}^n |X_i| \to \mathbb{E}_\theta[|X_1|],$$

car les observations sont indépendantes et de même loi possédant un moment d'ordre 1. On a vu à la question précédente que $\mathbb{E}_\theta[|X_1|] = \theta/2$, d'où en multipliant par 2, presque sûrement

$$\hat{\theta}_{MV} \to \theta.$$

7. (1pt) Déterminer le risque quadratique de $\hat{\theta}_{MV}$.

La décomposition biais variance dit que

$$\mathcal{R}(\hat{\theta}_{MV}; \theta) = b_n(\theta)^2 + \mathrm{Var}_\theta(\hat{\theta}_{MV}).$$

Nous avons vu que l'estimateur n'avait pas de biais, il reste donc à calculer sa variance. On a

$$\mathrm{Var}_\theta(\hat{\theta}_{MV}) = \frac{4}{n}\mathrm{Var}_\theta(|X_1|) = \frac{4}{n}\left(\mathbb{E}_\theta[X_1^2] - \mathbb{E}_\theta[|X_1|]^2\right) = \frac{\theta^2}{n}.$$

Indication : on pourra remarquer que $|X|^2 = X^2$.

8. (2pts) En calculant l'information de Fisher du modèle, déterminer si l'EMV est efficace.

On rappelle que l'information de Fisher est dans le cas de n observations i.i.d.,

$$I_n(\theta) = n\mathrm{Var}_\theta(\partial_\theta \log L_1(X_1;\theta)) = n\mathrm{Var}_\theta\left(-\frac{1}{\theta} + \frac{2|X_1|}{\theta^2}\right).$$

La variance de $|X_1|$ a déjà été calculée et elle vaut $\theta^2/4$, donc

$$I_n(\theta) = n/\theta^2.$$

Ainsi la borne de Cramer-Rao est θ^2/n qui est le risque de l'EMV, donc l'EMV est efficace.

Partie 3 : Une application de cette distribution

La distribution de Laplace est couramment utilisée pour modéliser le rendement d'un actif financier. Supposons que notre rendement en pourcentage soit représenté par la variable aléatoire Y, définie comme $Y = 3 + X$, où la densité de X est f_θ.

9. (1pt) Voici l'historique des rendements de notre investissement :

2.5	2.9	3.2	2.8	3.5	3.1	2.8	4

Estimer le paramètre θ en utilisant l'estimateur $\hat{\theta}_{MV}$. Nous ne procéderons pas aux calculs, mais nous donnerons une expression explicite de l'estimation.

Il faut considérer $X = Y - 3$, et donc

$$\hat{\theta}_{MV} = \frac{|2.5-3| + |2.9-3| + |3.2-3| + \ldots + |4-3|}{8}.$$

10. (1pt) Pour $\theta > 0$, calculer la probabilité $Q_\theta(Y < 2)$. En déduire une expression explicite de la probabilité que notre rendement soit inférieur à 2%.

On a pour $\theta > 0$,

$$Q_\theta(Y < 2) = Q_\theta(X < -1) = \int_{-\infty}^{-1} f_\theta(x)\,\mathrm{d}x = \ldots$$

Exercice 2 À la recherche du boson de Higgs (7pts)

Selon Wikipédia, "le boson de Higgs est une particule élémentaire dont l'existence permet d'expliquer la brisure de l'interaction unifiée électrofaible en deux interactions par l'intermédiaire du mécanisme de Brout-Englert-Higgs-Hagen-Guralnik-Kibble et d'expliquer ainsi pourquoi certaines particules ont une masse et d'autres n'en ont pas". Afin d'identifier ce boson, les physiciens provoquent la collision de plusieurs particules au sein du Grand Collisionneur de Hadrons (LHC). Cette collision génère une énergie théorique de μ GeV, une énergie théorique qui diffère selon que l'on suppose l'existence ou non du boson de Higgs. Nous proposons ici une procédure statistique visant à déterminer si l'absence de ce boson est conforme à nos observations.

1. (2pts) On sait que les appareils de mesure du LHC suivent une loi $\mathcal{N}(\mu, \sigma^2)$ avec $\sigma^2 = 1/10000$ et μ l'énergie théorique de la collision. Déterminer l'estimateur du maximum de vraisemblance dans le cadre des observations i.i.d. ainsi que sa loi exacte.

Dans ce cadre, l'EMV est la moyenne empirique et de plus, comme nos observations suivent la distribution $\mathcal{N}(\mu, \sigma^2)$ de manière indépendante, on a

$$\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/2).$$

2. (2pts) Construire un intervalle de confiance bilatéral du paramètre μ de niveau de confiance $1-\alpha$.

En normalisant la moyenne empirique, on obtient que

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0,1).$$

Par la défintion des quantiles d'ordre et la parité de la densité gaussienne, on a

$$Q_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \in [-q_{1-\alpha/2}, +q_{1-\alpha/2}]\right) = 1-\alpha.$$

Il reste donc à "inverser" l'intervalle, et obtenir

$$Q_\mu\left(\mu \in [\bar{X}_n \pm \sigma q_{1-\alpha/2}/\sqrt{n}]\right) = 1-\alpha.$$

Ainsi $[\bar{X}_n \pm \sigma q_{1-\alpha/2}/\sqrt{n}]$ est un IC de niveau de confiance $1-\alpha$.

3. (1pt) D'après la théorie, si le boson de Higgs n'existe pas, l'énergie mesurée par cette expérience doit valoir $\mu_0 = 125$ GeV. Énoncer les hypothèses de cette expérience et proposer un test statistique.

Nous allons tester les hypothèses suivantes

$$(H_0) : "\mu = \mu_0" \text{ Vs } (H_1) : "\mu \neq \mu_0".$$

Nous savons que sous (H_0), la probabilité que l'intervalle $[\bar{X}_n \pm \sigma q_{1-\alpha/2}/\sqrt{n}]$ contient μ_0 est $1-\alpha$. La procédure devient donc

- on rejette si l'intervalle de confiance ne contient $\mu_0 = 125$, i.e.

$$\bar{X}_n \notin [125 \pm \sigma q_{1-\alpha/2}/\sqrt{n}].$$

- on ne rejette pas dans le cas contraire.

4. (1pt) Après 121 mesures, on observe une moyenne empirique de l'énergie $\overline{x} = 125.03$ GeV. Pour être sûr de ne pas rejeter à tort l'hypothèse nulle, les physiciens ont pris un $\alpha = 0.01$. Remplacer toutes les variables par leurs valeurs numériques. On ne fera pas le calcul pour déterminer l'intervalle de confiance évalué.

On doit vérifier si 125.03 est dans l'intervalle

$$[125 \pm 0.01 \times 2.576/11]$$

On admet que μ_0 n'appartient pas à l'intervalle de confiance évalué. Que conclure ?

On rejette !

5. (1pt) Commenter le titre suivant d'un article paru dans *Le Monde* :

Le boson de Higgs découvert avec 99 % de certitude !!

Le journaliste, comme beaucoup de personne, a sur-interprété la conclusion du test. Ce que l'on peut dire, c'est que si le boson de Higgs n'existe pas (ou plus exactement le modèle des particules classique sans le boson de Higgs est vraie) alors nous venons d'observer un évènement qui avait moins de 1% de chance de se produire, donc un évènement peu probable dans le modèle classique. Nous n'avons en aucun cas rejeter l'hypothèse nulle avec 99% de certitude !