Statistique mathématique – recueil d'exercices

Luca Ganassali

Université Paris-Saclay

2025 – 2026

(Last update: February 9, 2026)

1. Outils probabilistes pour le statisticien

Cette première feuille d'exercices est particulièrement longue. Il n'est pas attendu de tout traiter.

Exercice 1.1 (Maximum de variables exponentielles i.i.d.)**.** Soient $(X_n)_{n\geq 1}$ sont i.i.d. et distribuées exponentiellement avec paramètre $\lambda > 0$, c'est-à-dire de densité $x \mapsto \lambda e^{-\lambda x} \mathbb{1}_{x>0}$ par rapport à la mesure de Lebesgue sur $\mathbb{R}$. Pour tout $n \geq 1$, on définit

$$M_n := \max(X_1, \dots, X_n)\,.$$

Equivalent presque sûr de M_n (optionnel).

1. À l'aide du Lemme de Borel-Cantelli, montrer que $\liminf \frac{M_n}{\log n} \geq 1/\lambda$, p.s. *Solution.* *Pour 1., prenons $\varepsilon > 0$. On a*

$$\mathbb{P}\left(\frac{M_n}{\log n} \leq \frac{1}{\lambda} - \varepsilon\right) = \mathbb{P}(M_n \leq (1/\lambda - \varepsilon)\log n) \\ = \mathbb{P}(X_1 \leq (1/\lambda - \varepsilon)\log n)^n \\ = (1 - n^{-1+\lambda\varepsilon})^n \\ = \exp(-n^{\lambda\varepsilon}(1 + o(1))).$$

Le terme ci-dessus est sommable en n, donc d'après Borel–Cantelli, p.s, $\frac{M_n}{\log n} \leq \frac{1}{\lambda} - \varepsilon$ n'arrive qu'un nombre fini de fois. En "rationnalisant" les ε, on obtient, que p.s., pour tout $m \geq 0$, $\frac{M_n}{\log n} \leq \frac{1}{\lambda} - 2^{-m}$ n'arrive qu'un nombre fini de fois, i.e., p.s. pour tout $m \geq 0$, $\liminf \frac{M_n}{\log n} \leq \frac{1}{\lambda} - 2^{-m}$, i.e. p.s. $\liminf \frac{M_n}{\log n} \geq 1/\lambda$, p.s., comme voulu.

2. À l'aide du Lemme de Borel-Cantelli, montrer que $\limsup \frac{X_n}{\log n} = 1/\lambda$ p.s. *Solution.* *Pour la borne supérieure, prenons $\varepsilon > 0$. On a*

$$\mathbb{P}(X_n/\log(n) \geq 1/\lambda + \varepsilon) = \mathbb{P}(X_n \geq (1/\lambda + \varepsilon)\log(n)) = n^{-(1+\lambda\varepsilon)},$$

qui est sommable en n: par Borel–Cantelli avec les mêmes étapes que précédemment, on a que $\limsup \frac{X_n}{\log n} \leq 1/\lambda$ p.s. Pour la borne inférieure, il faut utiliser Borel-Cantelli avec l'indépendance: pour $0 < \varepsilon < 1/\lambda$,

$$\mathbb{P}(X_n/\log(n) \geq 1/\lambda - \varepsilon) = \mathbb{P}(X_n \geq (1/\lambda - \varepsilon)\log(n)) = n^{-(1-\lambda\varepsilon)},$$

qui somme en n à $+\infty$. Par Borel-Cantelli, comme les X_n sont indépendants, on a que p.s., $X_n/\log(n) \geq 1/\lambda - \varepsilon$ une infinité de fois, i.e. p.s., $\limsup X_n/\log(n) \geq 1/\lambda - \varepsilon$. En rationnalisant les ε comme précédemment, on obtient que p.s., $\limsup X_n/\log(n) \geq 1/\lambda$. Cela conclut la question.

3. Montrer finalement que $\frac{M_n}{\log n} \xrightarrow[n\to\infty]{p.s.} 1/\lambda$. *On pourra faire une preuve "à la Cesaro".* *Solution.* *Au vu de 1., il ne nous reste qu'à montrer que $\limsup \frac{M_n}{\log n} \leq 1/\lambda$, p.s. Notons qu'on ne peut pas utiliser B-C ici car les M_n ne sont pas indépendants ! On utilise la question 2., Pour tout $\omega \in \Omega$ sauf sur un ensemble de mesure nulle, pour tout $\varepsilon > 0$, il existe une constante $N_\omega = N_\omega(\varepsilon)$ telle que $\frac{X_k(\omega)}{\log k} - 1/\lambda \leq \varepsilon$ pour tout $k \geq N_\omega$. Fixons un tel ω et prenons $\varepsilon > 0$ quelconque. Ensuite, "à la Cesàro",*

$$\frac{M_n(\omega)}{\log n} \leq \frac{\max_{1\leq k\leq N_\omega} X_k(\omega)}{\log n} + \max_{N_\omega+1\leq k\leq n} \frac{X_k(\omega)}{\log k}\,,$$

le premier terme converge p.s. vers 0 *car* N_ω *est fini et est donc* $\leq \varepsilon$ *pour* n *assez grand, et le second terme est* $\leq \frac{1}{\lambda} + \varepsilon$ *pour* n *assez grand. On a donc* $\frac{M_n(\omega)}{\log n} \leq 1/\lambda + 2\varepsilon$ *pour* n *assez grand. Ceci pour tout* ε*, donc* $\limsup \frac{M_n(\omega)}{\log n} \leq 1/\lambda$*. Finalement,* $\limsup \frac{M_n}{\log n} \leq 1/\lambda$*, p.s. et on a la limite presque sûre voulue.*

Développement limité de M_n puis de $1/M_n$.

4. Montrer que $M_n - \frac{\log n}{\lambda}$ converge en loi vers une loi notée G_λ que vous caractériserez. *On passera par la fonction de répartition.* *Solution. En regardant la limite de la f.d.r., on obtient que*

$$\mathbb{P}(M_n - \log n/\lambda \leq t) = \mathbb{P}(M_n \leq t - \log n/\lambda) = \mathbb{P}(X_1 \leq t - \log n/\lambda)^n$$
$$= \left(1 - e^{-\lambda t}/n\right)^n \to e^{-e^{-\lambda t}}.$$

La limite en loi $M_n - \log n/\lambda$ *est la loi sur* $\mathbb{R}$ *de f.d.r.* $x \mapsto e^{-e^{-\lambda x}}$ *(on parle de loi de Gumbel).*

5. En appliquant la méthode Delta et la question précédente, montrer que

$$\log n \times \left(\frac{\log n}{M_n} - \lambda\right) \xrightarrow[n\to\infty]{(d)} H_\lambda,$$

et préciser la fonction de répartition de H_λ. *Solution. On a établi que*

$$\log n \times (M_n/\log n - 1/\lambda) \xrightarrow[n\to\infty]{(d)} G_\lambda.$$

applique la méthode Delta pour la transformation $g : u \mapsto 1/u$*, dérivable en* $1/\lambda$*. On a alors :*

$$\log n \times ((\log n)/M_n - \lambda) \xrightarrow[n\to\infty]{(d)} g'(1/\lambda) \times G_\lambda = -\lambda^2 G_\lambda,$$

d'où

$$\frac{\log^2 n}{M_n} - \lambda \log n \xrightarrow[n\to\infty]{(d)} H_\lambda,$$

avec $H_\lambda \overset{(d)}{=} -\lambda^2 G_\lambda$ *(égalité en loi). On peut calculer la f.d.r. de la loi* H_λ*: si* $Y \sim G_\lambda$ *et* $t \in \mathbb{R}$*, on a*

$$\mathbb{P}(-\lambda^2 Y \leq t) = \mathbb{P}(Y \geq -t/\lambda^2) = 1 - e^{-e^{\lambda t/\lambda^2}} = 1 - e^{-e^{t/\lambda}}.$$

Application: intervalle de confiance pour λ.

6. Pour $\beta \in]0, 1[$, calculer explicitement le quantile d'ordre β de la loi de H_λ, noté $h_{\lambda,\beta}$. *Solution. On a la fonction de répartition, il suffit de l'inverser :*

$$1 - e^{-e^{t/\lambda}} = \beta \iff -e^{t/\lambda} = \log(1-\beta) \iff t = \lambda \log(-\log(1-\beta)),$$

d'où $h_{\lambda,\beta} = \lambda \log(-\log(1-\beta))$*.*

7. Montrer avec tout ce qui précède que pour tout $\alpha \in]0, 1[$, en notant

$$I_1(X_1, \dots, X_n; \alpha) := \left[\frac{\frac{\log n}{M_n}}{1 + \frac{\log(-\log(\alpha/2))}{\log n}}, \frac{\frac{\log n}{M_n}}{1 + \frac{\log(-\log(1-\alpha/2))}{\log n}}\right],$$

on a

$$\mathbb{P}\left(I_1(X_1,\dots,X_n;\alpha)\ni\lambda\right) \underset{n\to\infty}{\longrightarrow} 1-\alpha\,.$$

$I_1(X_1,\dots,X_n)$ est un intervalle aléatoire appelé intervalle de confiance asymptotique pour λ de probabilité de couverture $1-\alpha$. Solution. *D'après ce qui précède, pour tout* $\alpha\in]0,1[$,

$$\begin{aligned}
&\mathbb{P}\left(I_1(X_1,\dots,X_n;\alpha)\ni\lambda\right)\\
&=\mathbb{P}\left(\lambda\log(-\log(1-\alpha/2))\le\frac{\log^2 n}{M_n}-\lambda\log n\le\lambda\log(-\log(\alpha/2))\right)\\
&\underset{n\to\infty}{\longrightarrow}\mathbb{P}\left(h_{\lambda,\alpha/2}\le H_\lambda\le h_{\lambda,1-\alpha/2}\right)=1-\alpha\,.
\end{aligned}$$

8. Quel est l'équivalent presque sûr de $\mathrm{diam}(I_1(X_1,\dots,X_n;\alpha))$? Solution. *On utilise tous les résultats précédents, et p.s.,*

$$\begin{aligned}
\mathrm{diam}(I_1(X_1,\dots,X_n;\alpha))&=\frac{\log n}{M_n}\left[\frac{1}{1+\frac{\log(-\log(1-\alpha/2))}{\log n}}-\frac{1}{1+\frac{\log(-\log(\alpha/2))}{\log n}}\right]\\
&\sim\lambda\frac{\log(-\log(\alpha/2))-\log(-\log(1-\alpha/2))}{\log n}
\end{aligned}$$

(NB : les limites sont p.s. et constantes ici, pas besoin du Lemme de Slutsky).

9. ($\star$ car moins guidé) Pourrait-on trouver un autre intervalle de confiance dont le diamètre décroit sensiblement plus vite en n que l'intervalle précédent ?

Exercice 1.2 (Empirical risk minimisation)**.** On observe $(X_1,Y_1),\dots,(X_n,Y_n)$ i.i.d., de loi inconnue, avec X_i vecteurs aléatoires de $\mathbb{R}^d$ (*features*), et $Y_i\in\{0,1\}$ (*labels*). On appelle *classifieur* une fonction mesurable $h:\mathbb{R}^d\to\{0,1\}$. Soit $\mathcal{H}$ un ensemble fini de classifieurs. On considère le *classification loss* ℓ défini par

$$\ell(h(X),Y)=\mathbb{1}_{\{h(X)\neq Y\}}\,.$$

Pour $h\in\mathcal{H}$, on définit son *risque* $R(h)$ et son *risque empirique* $\widehat{R}_n(h)$, donnés par :

$$R(h)=\mathbb{E}[\ell(h(X),Y)],\qquad \widehat{R}_n(h)=\frac{1}{n}\sum_{i=1}^n\ell(h(X_i),Y_i)\,.$$

Le but du statisticien est de trouver le meilleur classifieur $h\in\mathcal{H}$ au sens du risque R:

$$h^\star\in\arg\min_{h\in\mathcal{H}}R(h)\,.$$

Problème : R dépend de la loi des données qui est inconnue : on ne peut pas calculer $R(h)$. La seule chose à laquelle nous avons accès, ce sont les données. On définit l'estimateur ERM $\hat{h}_n$ (empirical risk minimizer) par :

$$\hat{h}_n\in\arg\min_{h\in\mathcal{H}}\widehat{R}_n(h)\,.$$

Notons qu'on a par définition $R(h^\star)\le R(\hat{h}_n)$. Le but de l'exercice est d'établir une inégalité dans l'autre sens (avec des termes en plus) afin de montrer que $\hat{h}_n$ n'est pas trop mauvais par rapport à $h^\star$ au sens du risque R.

1. A l'aide de l'inégalité de Hoeffding, montrer que pour tout $h \in \mathcal{H}$ et tout $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\widehat{R}_n(h) - R(h)\right| \geq \varepsilon\right) \leq 2e^{-2n\varepsilon^2}.$$

Solution. *Les variables $\ell(h(X_i), Y_i)$ est bornée entre 0 et 1 et indépendantes. L'inégalité d'Hoeffding donne directement le résultat.*

2. En utilisant une *union bound*, montrer que pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{h\in\mathcal{H}}\left|\widehat{R}_n(h) - R(h)\right| \geq \varepsilon\right) \leq 2|\mathcal{H}|\, e^{-2n\varepsilon^2}.$$

En déduire qu'avec probabilité au moins $1 - \delta$,

$$\sup_{h\in\mathcal{H}}\left|\widehat{R}_n(h) - R(h)\right| \leq \sqrt{\frac{1}{2n}\log\left(\frac{2|\mathcal{H}|}{\delta}\right)}.$$

Solution. *On applique l'union bound :*

$$\mathbb{P}\left(\exists h \in \mathcal{H} : \left|\widehat{R}_n(h) - R(h)\right| \geq \varepsilon\right) \leq \sum_{h\in\mathcal{H}} 2e^{-2n\varepsilon^2} = 2|\mathcal{H}|e^{-2n\varepsilon^2}.$$

On inverse la borne pour écrire l'inégalité avec probabilité $1 - \delta$.

3. Montrer enfin qu'avec probabilité au moins $1 - \delta$,

$$R(\hat{h}_n) \leq R(h^\star) + 2\sqrt{\frac{1}{2n}\log\left(\frac{2|\mathcal{H}|}{\delta}\right)}.$$

Interpréter. *Solution.* *On note que par définition de $\hat{h}_n$, $\widehat{R}_n(\hat{h}_n) \leq \widehat{R}_n(h^\star)$. On ajoute et soustrait des termes de risque et on utilise par deux fois la borne uniforme de la question 2. :*

$$\begin{aligned} R(\hat{h}_n) &\leq \widehat{R}_n(\hat{h}_n) + \sup_{h\in\mathcal{H}}|R(h) - \widehat{R}_n(h)| \leq \widehat{R}_n(h^\star) + \sup_{h\in\mathcal{H}}|R(h) - \widehat{R}_n(h)| \\ &\leq R(h^\star) + 2\sup_{h\in\mathcal{H}}|R(h) - \widehat{R}_n(h)|. \end{aligned}$$

*En remplaçant par la borne de 2., on obtient l'inégalité annoncée. L'erreur $R(h^\star)$ correspond à l'*erreur d'approximation *(limite imposée par $\mathcal{H}$), tandis que le terme supplémentaire correspond à l'*erreur d'estimation du risque*.*

Exercice 1.3 (Loi et espérance conditionnelle)**.** Soit $X \sim \mathrm{Exp}(1)$, et Y une variable aléatoire dont la densité conditionnelle par rapport à la mesure de comptage sur $\mathbb{N}$ est donnée par :

$$p_{Y\,|\,X}(y\,|\,x) = \frac{x^y e^{-x}}{y!}\,.$$

1. Trouver la densité jointe de (X, Y) et identifier la mesure dominante. *Solution.* *La densité jointe est donnée par*

$$p_{X,Y}(x, y) = p_X(x)p_{Y|X}(y, x) = \frac{x^y e^{-2x}}{y!}\,.$$

La mesure dominante est ici $\mathrm{Leb}_{\mathbb{R}_+} \times \mu_{\mathbb{N}}$, où $\mu_{\mathbb{N}}$ désigne la mesure de comptage sur $\mathbb{N}$.

2. Quelle est la densité marginale de Y ? En déduire $\mathbb{E}[Y]$. *Solution. On obtient la densité marginale de Y comme suit :*

$$p_Y(y) = \int_{\mathbb{R}_+} p_{X,Y}(x,y)dx = \int_{\mathbb{R}_+} \frac{x^y e^{-2x}}{y!} dx = \frac{1}{y!}\int_{\mathbb{R}_+} (u/2)^y e^{-u}(du/2) = \frac{\Gamma(y+1)}{2^{y+1}y!} = \frac{1}{2^{y+1}}\,.$$

On a donc $Y \sim \mathcal{G}(1/2)$, la loi géométrique sur $\mathbb{N}$ de paramètre de probabilité de succès $p = 1/2$. Ainsi $\mathbb{E}[Y] = \frac{1-p}{p} = 1$.

3. Reconnaissez-vous la loi de Y conditionnellement à X ? Vérifier que la loi de l'espérance totale donne la même valeur pour $\mathbb{E}[Y]$. *Solution. On voit que $Y|X \sim \mathrm{Poi}(X)$, donc $\mathbb{E}[Y|X] = X$ et $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[X] = 1$, ce qui donne le même résultat.*

4. Trouver la densité conditionnelle de X sachant $Y = y$. *Solution. On a*

$$p_{X|Y}(x,y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{2^{y+1}x^y e^{-2x}}{y!}\,.$$

On reconnaît (peut-être) une loi Gamma : on peut noter $X|Y \sim \Gamma(Y+1, 1/2)$.

Exercice 1.4 (Fonction caractéristique des variables gaussiennes)**.** Nous voulons montrer que si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors pour tout $t \in \mathbb{R}$,

$$\Phi_X(t) := \mathbb{E}[e^{itX}] = \exp\left(i\mu t - \frac{\sigma^2}{2}t^2\right)\,.$$

0. Montrer le résultat lorsque X est dégénérée ($\sigma = 0$). Nous supposons désormais $\sigma > 0$.

1. Montrer que pour tout $t \in \mathbb{R}$, $\Phi_X(t) = e^{it\mu} f_\sigma(t)$, où

$$f_\sigma(t) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\sigma t y - y^2/2} dy\,.$$

Solution. il suffit d'écrire la définition et de faire le changement de variables $y = \frac{x-\mu}{\sigma}$.

2. Montrer que pour tout $t \in \mathbb{R}$, $f_\sigma(t) \in \mathbb{R}$, que f_σ est différentiable et trouver une équation différentielle qu'elle satisfait. *Solution. on prend le conjugué et on fait le changement $y \to -y$. On peut utiliser la convergence dominée : chaque intégrande est dérivable en t et $|-\sigma \sin(\sigma t y) y e^{-y^2/2}|$ est facilement dominé. Par intégration par parties (on intègre $ye^{-y^2/2}$),*

$$\begin{aligned} f'_\sigma(t) &= -\frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} \sin(\sigma t y) y e^{-y^2/2} dy \\ &= 0 + \frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} \sigma t \cos(\sigma t y)(-e^{-y^2/2}) dy \\ &= -\sigma^2 t f_\sigma(t)\,. \end{aligned}$$

3. Conclure. *Solution. nous avons la condition initiale $f_\sigma(0) = 1$. C'est terminé.*

4. En déduire que si $X \sim \mathcal{N}(\mu, \sigma^2)$, et $a, b \in \mathbb{R}$, alors $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

5. En déduire que, si $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ et X_1 et X_2 sont indépendantes, alors

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)\,.$$

Exercice 1.5 (Méthode Delta pour une loi multinomiale)**.** *Cet exercice suppose une connaissance préalables des propriétés des vecteurs gaussiens (que nous reverrons plus tard dans le cours).* Soit

$$(A_n, B_n, C_n) \sim \text{Multinomial}(n; p_1, p_2, 1 - p_1 - p_2)$$

avec $p_1, p_2 \in (0,1)$ et $p_1 + p_2 < 1$. Pour tout $n \geq 1$, on définit

$$U_n := \frac{A_n}{n}, \quad V_n := \frac{B_n}{n}, \quad Z_n := \log\left(\frac{U_n}{V_n}\right).$$

1. Montrer que

$$\sqrt{n}\big((U_n, V_n) - (p_1, p_2)\big) \xrightarrow[n\to\infty]{(d)} \mathcal{N}\big(0, \Sigma\big),$$

où

$$\Sigma = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 \\ -p_1p_2 & p_2(1-p_2) \end{pmatrix}.$$

2. En appliquant la méthode Delta à la fonction

$$g(x,y) = \log(x) - \log(y),$$

montrer que

$$\sqrt{n}\left(Z_n - \log\left(\frac{p_1}{p_2}\right)\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}\Big(0, \frac{1}{p_1} + \frac{1}{p_2}\Big).$$

Interpréter la symétrie dans la variance limite en p_1, p_2.

Solution. *Pour 1., on utilise le théorème central limite multidimensionnel pour les composantes d'une loi multinomiale. En effet, la covariance entre A_n et B_n est $-np_1p_2$, et les variances sont $np_1(1-p_1)$ et $np_2(1-p_2)$, d'où le résultat après normalisation par n. Pour 2., la fonction g est différentiable en (p_1, p_2) avec gradient*

$$\nabla g(p_1, p_2) = \left(\frac{1}{p_1}, -\frac{1}{p_2}\right).$$

Par la méthode Delta multidimensionnelle,

$$\sqrt{n}\big(g(U_n, V_n) - g(p_1, p_2)\big) \xrightarrow[n\to\infty]{(d)} \mathcal{N}\Big(0, \nabla g(p_1,p_2)^\top \Sigma \nabla g(p_1, p_2)\Big).$$

On calcule

$$\nabla g(p_1,p_2)^\top \Sigma \nabla g(p_1,p_2) = \frac{1}{p_1^2} p_1(1-p_1) + \frac{1}{p_2^2} p_2(1-p_2) + 2\frac{-1}{p_1p_2}(-p_1p_2) = \frac{1}{p_1} + \frac{1}{p_2}.$$

Ainsi,

$$\sqrt{n}\left(Z_n - \log\left(\frac{p_1}{p_2}\right)\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}\Big(0, \frac{1}{p_1} + \frac{1}{p_2}\Big),$$

comme souhaité.

Exercice 1.6 (Plus rapide que $1/\sqrt{n}$?)**.** Supposons que $X_1, \ldots, X_n$ sont i.i.d., centrées, de variance finie $\sigma^2 > 0$ inconnue. On note classiquement $\bar{X}_n$ la moyenne empirique des X_i et

$$\hat{\sigma}^2 := \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2.$$

1. Étudier la convergence presque sûre de $\bar{X}_n$ et de $\hat{\sigma}^2$.

2. Montrer que
$$\frac{\sqrt{n}\bar{X}_n}{\sqrt{\hat{\sigma}^2}} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1)\,.$$

Dans toute la suite, on suppose que $\sigma^2 = 1$ et qu'il est connu.

3. Quelle est la limite en loi de $\sqrt{n}(\cos(\bar{X}_n) - 1)$?

4. Trouver une suite $(a_n)_{n\geq 1}$ telle que $a_n(\cos(\bar{X}_n) - 1)$ converge en loi vers une v.a. Z qui n'est pas presque sûrement constante. Caractériser la loi de Z. *Solution. On revient à la preuve du théorème de la méthode Delta, mais on va jusqu'à l'ordre 2 ! On a* $\cos(x) - 1 = -x^2/2 + x^2\varepsilon(x)$, *avec ε continue sur $\mathbb{R}\setminus\{0\}$, et $\varepsilon(x) \xrightarrow[x\to 0]{} 0$. On peut alors prolonger ε par continuité en 0. Puisque $\bar{X}_n \to 0$ p.s., alors par continuité, $\varepsilon(\bar{X}_n) \to \varepsilon(0) = 0$ p.s., donc aussi en probabilité. Ainsi, pour tout n,*
$$\cos(\bar{X}_n) = 1 + 0 - \frac{(\bar{X}_n)^2}{2}(1 + \varepsilon(\bar{X}_n))\,,$$
avec $1 + \varepsilon(\bar{X}_n) \xrightarrow[n\to\infty]{\mathbb{P}} 1$. On applique de nouveau le lemme de Slutsky pour conclure :
$$n(\cos(\bar{X}_n) - 1) = -(1 + \varepsilon(\bar{X}_n)) \times \frac{(n\bar{X}_n)^2}{2} \xrightarrow[n\to\infty]{(d)} -Z^2/2,$$
avec $Z \sim \mathcal{N}(0,1)$. Remarquons pour les plus sagaces qu'en appliquant une nouvelle fois la méthode Delta avec arccos, *on pourrait obtenir $n\bar{X}_n \to Z'$, ce qui contredirait le résultat de 3. Mais, bien sûr, on ne peut pas faire cela (voyez-vous pourquoi ?).*

Exercice 1.7 (Une inégalité de type Bennett)**.** On note h la fonction définie sur $\mathbb{R}_+$ par $h(x) = (1+x)\log(1+x) - x$, et ϕ définie sur $\mathbb{R}$ par $\phi(x) = \exp(x) - x - 1$. Soit $X \sim \mathrm{Poi}(\theta)$ avec $\theta > 0$.

1. Pour tout $\lambda \in \mathbb{R}$, calculer $\mathbb{E}[\exp(\lambda(X-\theta))]$ en fonction de ϕ. *Solution. On obtient directement*
$$\mathbb{E}[\exp(\lambda(X-\theta))] = e^{-\theta}\sum_{k\geq 0}\frac{\theta^k e^{\lambda(k-\theta)}}{k!} = e^{-\theta-\lambda\theta+\theta e^{\lambda}} = e^{\theta\phi(\lambda)}$$

2. Montrer que pour tout $x > 0$,
$$\mathbb{P}(X - \theta \geq x) \leq \exp(-\theta h(x/\theta))\,.$$
Solution. Par la méthode de Chernoff,
$$\mathbb{P}(X - \theta \geq x) \leq \exp(-\theta \sup_{\lambda\geq 0}(\lambda x/\theta - \phi(\lambda)))$$
et ce sup est atteint en étudiant $f(\lambda) = \lambda y - \phi(\lambda)$, $f'(\lambda) = y + 1 - e^{\lambda} > 0 \iff \lambda <$ $\log(y+1)$. Il y a donc un maximum global en $\lambda = \log(y+1) > 0$ et ce maximum vaut
$$y\log(y+1) - \phi(\log(y+1)) = y\log(y+1) - (y+1) + \log(y+1) + 1 = h(y)\,.$$

3. On donne $e^{-2h(4)} \sim 0.003$. Comment cette inégalité se compare-t-elle à Bienaymé–Tchebychev

pour $\mathbb{P}(X \geq 10)$ lorsque $\theta = 2$? *Solution. Ici, $\theta = 2$ et $x = 10 - 2 = 8$. Bennett donne*

$$\mathbb{P}(X \geq 10) \leq e^{-2h(4)} \sim 0.003$$

tandis que Bienaymé–Tchebychev donne

$$\mathbb{P}(X \geq 10) \leq \mathbb{P}(|X - \theta| \geq 8) \leq \frac{\theta}{8^2} = \frac{1}{32},$$

ce qui est (nettement) moins bon.

Exercice 1.8 (Canaux gaussiens)**.** Supposons que $Y \sim \mathcal{N}(t, \tau^2)$ et que, conditionnellement à $Y = y$, $X_1, \ldots, X_n$ soient i.i.d. de loi $\mathcal{N}(y, \sigma^2)$. Montrer que

$$Y \,|\, (X_1, \ldots, X_n) \sim \mathcal{N}\left(\frac{t/\tau^2 + n\bar{X}_n/\sigma^2}{1/\tau^2 + n/\sigma^2}, \frac{1}{1/\tau^2 + n/\sigma^2}\right).$$

Pour une variable gaussienne, on appelle *précision* l'inverse de sa variance. Quelle propriété de la précision a-t-on mise en évidence ici ? *Solution. La densité de $X = (X_1, \ldots, X_n)$ conditionnellement à $Y = y$ est donnée par*

$$\begin{aligned} p_{X|Y}(x_1, \ldots, x_n, y) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - y)^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - y)^2\right), \end{aligned}$$

d'où la densité jointe

$$\begin{aligned} p_{X,Y}(x_1, \ldots, x_n, y) &= \frac{1}{\sqrt{2\pi\tau^2}(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\tau^2}(y - t)^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - y)^2\right) \\ &\propto_y \exp\left(-\frac{1}{2}\left[(1/\tau^2 + n/\sigma^2)y^2 - (2/\tau^2 + 2n\bar{X}_n/\sigma^2)y\right]\right) \\ &\propto_y \exp\left(-\frac{1}{2(1/\tau^2 + n/\sigma^2)^{-1}}\left[y - \frac{1/\tau^2 + n\bar{X}_n/\sigma^2}{1/\tau^2 + n/\sigma^2}\right]^2\right), \end{aligned}$$

d'où

$$p_{Y|X}(y, x_1, \ldots, x_n) = \frac{p_{X,Y}(x_1, \ldots, x_n, y)}{p_X(x_1, \ldots, x_n)} \propto_y p_{X,Y}(x_1, \ldots, x_n, y),$$

ce qui donne le résultat voulu. Nous venons de montrer que la précision de la loi conditionnelle est la somme des précisions sur X_i et Y (et la moyenne est une moyenne pondérée par les précisions).

Exercice 1.9 (Loi bêta)**.** Supposons que X et Y sont des v.a. réelles positives et indépendantes.

1. Montrer que pour tout $0 < x < 1$,

$$\mathbb{P}\left(\frac{X}{X + Y} \leq x\right) = \mathbb{E}\left[F_X\left(\frac{xY}{1 - x}\right)\right].$$

Solution. C'est simplement la loi de l'espérance totale, et comme X et Y sont indépen-

dantes,

$$\mathbb{P}\left(\frac{X}{X+Y}\leq x\,|\,Y=y\right)=\mathbb{E}\left[\mathbb{1}_{X\leq\frac{xY}{1-x}}\,|\,Y=y\right]=\mathbb{E}\left[\mathbb{1}_{X\leq\frac{xy}{1-x}}\right]=F_X\left(\frac{xy}{1-x}\right).$$

2. Supposons que $\mathbb{E}[Y]<\infty$ et que X admette une densité p_X par rapport à μ. Montrer que $V=X/(X+Y)$ admet aussi une densité p_V par rapport à μ, donnée par

$$p_V(x)=\mathbb{E}\left[\frac{Y}{(1-x)^2}p_X\left(\frac{xY}{1-x}\right)\right].$$

Solution. C'est parce qu'on peut dériver sous le signe intégral à la question 1. C'est une conséquence du théorème de convergence dominée (version différentiable), puisque pour tout $0<x<b<1$,

$$\left|\frac{Y}{(1-x)^2}p_X\left(\frac{xY}{1-x}\right)\right|\leq\frac{Y}{(1-b)^2},$$

ce qui est intégrable.

Pour $\alpha,\beta>0$, la loi Gamma $\Gamma(\alpha,\beta)$ est la loi de densité

$$x\mapsto\frac{1}{\Gamma(\alpha)}\beta^\alpha x^{\alpha-1}e^{-\beta x}$$

par rapport à $\mathrm{Leb}_{\mathbb{R}_+}$, et la loi bêta $\mathrm{Beta}(\alpha,\beta)$ est la loi de densité

$$x\mapsto\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

par rapport à $\mathrm{Leb}_{[0,1]}$.

3. Montrer que si $X\sim\Gamma(\alpha,1)$ et $Y\sim\Gamma(\alpha',1)$ sont indépendantes, alors $\frac{X}{X+Y}\sim\mathrm{Beta}(\alpha,\alpha')$. *Solution. Par le changement de variables* $u=\frac{y}{1-x}$ *dans la densité précédente, on obtient*

$$\begin{aligned}p_V(x)&=\frac{1}{\Gamma(\alpha)\Gamma(\alpha')}\int_{\mathbb{R}_+}\frac{y}{(1-x)^2}\left(\frac{xy}{1-x}\right)^{\alpha-1}e^{-x}y^{\alpha'-1}e^{-y}dy\\&=\frac{x^{\alpha-1}(1-x)^{\alpha'-1}}{\Gamma(\alpha)\Gamma(\alpha')}\int_{\mathbb{R}_+}u^{\alpha+\alpha'-1}e^{-u}du\\&=\frac{\Gamma(\alpha+\alpha')}{\Gamma(\alpha)\Gamma(\alpha')}x^{\alpha-1}(1-x)^{\alpha'-1},\end{aligned}$$

qui est bien la densité voulue.

Exercice 1.10 (Absence de géométrie dans les graphes d'Erdős-Rényi)**.** Pour $n\geq 1$ on note $[n]=\{1,2,\dots,n\}$. Soit $G\sim G(n,p)$ un graphe aléatoire Erdős-Rényi d'ensemble de sommets $[n]$, où chaque arête $e=\{u,v\}\in\binom{[n]}{2}$ est présente dans G indépendamment avec probabilité $p\in[0,1]$. On note $e(G)$ le nombre d'arêtes de G.

1. Existe-t-il une valeur de p telle que $G\sim G(n,p)$ soit uniforme sur l'ensemble des graphes d'ensemble de sommets $[n]$? *Solution. Pour tout graphe g de sommet $[n]$,*

$$\mathbb{P}(G=g)=p^{e(g)}(1-p)^{\binom{n}{2}-e(g)}.$$

Si $p = 1/2$, $\mathbb{P}(G = g)$ est constante égale à $1/2^{\binom{n}{2}}$, c'est-à-dire uniforme sur l'ensemble des graphes sur $[n]$, qui a taille $1/2^{\binom{n}{2}}$.

2. Pour $G \sim G(n,p)$, quelle est la loi de $e(G)$? *Solution. Chaque arête suit une loi $\mathrm{Ber}(p)$, les arêtes sont indépendantes, donc $e(G) \sim \mathrm{Bin}(\binom{n}{2}, p)$.*

3. Pour $G \sim G(n,p)$, quelle est la loi conditionnelle de G sachant $e(G)$? *Solution. Elle est donnée par*

$$\begin{aligned}\mathbb{P}(G = g \,|\, e(G) = m) &= \frac{\mathbb{P}(G = g, e(G) = m)}{\mathbb{P}(e(G) = m)} \\ &= \mathbb{1}_{e(g)=m} \frac{p^m (1-p)^{\binom{n}{2}-m}}{\binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}} \\ &= \mathbb{1}_{e(g)=m} \frac{1}{\binom{\binom{n}{2}}{m}},\end{aligned}$$

ce qui est la loi uniforme sur l'ensemble des graphes de sommet $[n]$ et comportant $m = e(G)$ arêtes.

4. Expliquer le titre de l'exercice. *Solution. Une fois le nombre d'arêtes fixé, le graphe est uniforme : aucune contrainte géométrique particulière n'est imposée. Cela vient du fait que les arêtes sont i.i.d.*

2. Modélisation statistique, suffisance et complétude

Exercice 2.1. On considère le modèle où $(X_1, \dots, X_n)$ sont i.i.d. avec $X_1 \sim \mathrm{Unif}([\theta, \theta+1])$, avec $\theta > 0$ inconnu.

1. Écrire le modèle sous la forme canonique du cours. Montrer qu'il est identifiable.

2. Montrer que $S(X) = (\min_i X_i, \max_i X_i)$ est suffisante. *Solution. On utilise la factorisation de Neyman–Fisher, la densité jointe s'écrit* $\prod_{i=1}^n \mathbb{1}_{[\theta,\theta+1]}(x_i) = \mathbb{1}_{\min x_i \geq \theta} \mathbb{1}_{\max x_i \leq \theta+1}$.

3. Montrer que $A(X) = \max_i X_i - \min_i X_i$ est auxiliaire. *On pourra chercher la loi de $X_1 - \theta$ sous $\mathbb{P}_\theta$.Solution. Posons $Y_1 = X_1 - \theta$. Notons que $Y_1 \in [0,1]$, $\mathbb{P}_\theta$−p.s. Pour tout $t \in [0,1]$, $\mathbb{P}_\theta(Y_1 \leq t) = \mathbb{P}_\theta(X_1 \leq t + \theta) = t$, donc $X_1 \sim \mathrm{Unif}([0,1])$ (remarque : ce n'est pas une statistique, car l'opération appliquée à X_1 pour obtenir Y_1 dépend de θ). En notant $Y = (Y_1, \dots, Y_n) = (X_1 - \theta, \dots, X_n - \theta)$, on a que sous $\mathbb{P}_\theta$, $Y \sim \mathrm{Unif}([0,1])^{\otimes n}$. Or*
$$A(X) = \max_i (X_i - \theta) - \min_i (X_i - \theta) = \max_i Y_i - \min_i Y_i,$$
variables aléatoire dont la loi ne dépend pas de θ : $A(X)$ est donc auxiliaire.

4. En déduire que $S(X) = (\min_i X_i, \max_i X_i)$ n'est pas complète. *Solution. Supposons que $S(X)$ est complète. D'après la question précédente, $\mathbb{E}_\theta[f(S(X)] = 0$, où $f(s_1, s_2) = s_2 - s_1 - \alpha$ avec $\alpha = \mathbb{E}_\theta[\max_i Y_i - \min_i Y_i]$ (qui ne dépend pas de θ). Donc, comme $S(X)$ est complète, $\mathcal{M}$-p.s., $f(S(X)) = 0$ i.e. $\max_i X_i - \min_i X_i = \alpha$. Montrons que cela est absurde. Si $\alpha = 0$, alors $\max_i X_i = \min_i X_i$, tous les X_i sont égaux p.s., ce qui est absurde car les X_i sont à densité. Si $\alpha > 0$,*
$$\begin{aligned}\mathbb{P}_\theta(\min_i X_i - \max_i X_i \leq \alpha/2) &\geq \mathbb{P}_\theta(\theta \leq X_1 \leq \theta + \alpha/2, \dots, \theta \leq X_n \leq \theta + \alpha/2) \\ &= (\alpha/2)^n > 0,\end{aligned}$$
et donc $\min_i X_i - \max_i X_i$ n'est pas constant à α p.s. et f n'est pas nulle, contradiction.

Exercice 2.2. On considère un modèle dans lequel l'observation X est de la forme $X = a(Y - Z)$ avec $Y \sim \mathrm{Exp}(\lambda)$ et $Z \sim \mathrm{Exp}(\mu)$ indépendantes, $\lambda, \mu > 0$ inconnus.

On suppose d'abord a connu et $a = 1$.

1. Quels sont ici les paramètres du modèle ? *Solution. On a ici $\theta = (\lambda, \mu)$ et $\Theta = (\mathbb{R}_+^*)^2$.*

2. En posant $b = 1/\lambda$ et $c = 1/\mu$, trouver deux équations satisfaites par b et c. En raisonnant sur ces équations, montrer que le modèle est identifiable. *Solution. Les équations sont $b - c = \mathbb{E}[X] =: A$ et $b^2 + c^2 = \mathrm{Var}(X) =: B$. Il s'agit de trouver un point (b, c) sur le cercle centré d'équation $b^2 + c^2 = B$ à l'intersection avec la droite $c = b - A$. On voit facilement que si ces lieux géométriques s'intersectent, ils le font au plus une fois dans le quadrant positif ; comme $b = 1/\lambda$ et $c = 1/\mu$ doivent être des solutions, il y a une solution unique. Le modèle est identifiable.*

Supposons à présent que a est inconnu dans $\mathbb{R}$.

3. En raisonnant sur la transformation $(Y, Z) \mapsto (Z, Y)$, montrer que le modèle n'est pas identifiable. *Solution. Quand $\alpha \in \mathbb{R}$, les lois $\mathbb{P}_{(\alpha,\lambda,\mu)}$ et $\mathbb{P}_{(-\alpha,\mu,\lambda)}$ sont les mêmes, donc le modèle n'est pas identifiable.*

Supposons finalement que a est inconnu dans $\mathbb{R}_+$. L'argument précédent ne tient plus.

4. Montrer que si $Y \sim \mathrm{Exp}(\lambda)$ et $\alpha > 0$, $Y \sim \mathrm{Exp}(\lambda/\alpha)$. *Solution. via la fonction de répartition.*

5. En déduire que le modèle n'est pas identifiable. *Solution. La question 4. montre qu'i y a encore une symétrie $(\alpha, \lambda, \mu) \mapsto (1, \lambda/\alpha, \mu/\alpha)$ lorsque $\alpha > 0$. Le modèle n'est pas identifiable.*

Exercice 2.3. On considère le modèle où $(X_1, \ldots, X_n)$ sont i.i.d. avec densité $x \mapsto C(\theta)\, x \mathbb{1}_{[0,\theta]}(x)$ pour un certain paramètre $\theta > 0$.

1. Trouver la valeur de $C(\theta)$ pour tout $\theta > 0$. *Solution. Une intégration simple montre que $C(\theta) = \frac{2}{\theta^2}$.*

2. Trouver la densité de X_1/θ. *Solution. Avec la fonction de répartition. Pour tout $x \in [0, 1]$,*
$$\mathbb{P}_\theta(X_1/\theta \le x) = \mathbb{P}_\theta(X_1 \le x\theta) = \frac{2}{\theta^2} \int_0^{x\theta} u\, du = \frac{2(x\theta)^2}{2\theta^2} = x^2 .$$
Donc, en dérivant, X_1/θ a pour densité $x \mapsto 2x \mathbb{1}_{[0,1]}(x)$.

3. On trie les X_i par ordre croissant : $X_{(1)} \le \ldots \le X_{(n)}$ sont appelées les statistiques d'ordre de l'échantillon. Montrer que $X_{(n)}$ et $X_{(1)}/X_{(n)}$ sont indépendantes. *On utilisera le théorème de Basu, après avoir montré que $X_{(1)}/X_{(n)}$ est auxiliaire à l'aide de la question précédente. Solution. Montrons que $X_{(1)}/X_{(n)}$ est auxiliaire. Comme diviser par $\theta > 0$ ne change pas l'ordre des variables, on peut poser $Y_i = X_i/\theta$. Dans ce modèle, les Y_i sont i.i.d et la loi de Y ne dépend plus de θ. De plus, $Y_{(1)}/Y_{(n)} = X_{(1)}/X_{(n)}$ en loi. La loi du membre de gauche ne dépend pas de θ, donc $X_{(1)}/X_{(n)}$ est auxiliaire. Il reste à montrer que $X_{(n)} = \max_i X_i$ est suffisante et complète. On peut le démontrer directement en écrivant d'abord la densité jointe de X sous $\mathbb{P}_\theta$, qui s'écrit*
$$\frac{2^n}{\theta^{2n}} \prod_{i=1}^n x_i \mathbb{1}_{[0,\theta]}(x_i) = 2^n \Big(\prod_{i=1}^n x_i\Big) \mathbb{1}_{\min_i x_i \ge 0} \frac{1}{\theta^{2n}} \mathbb{1}_{\max_i x_i \le \theta}.$$
Puis, via la fonction de répartition, comme dans l'exemple du cours, on obtient que $X_{(n)}$ a pour densité $u \mapsto \frac{2nu^{2n-1}}{\theta^{2n}} \mathbb{1}_{[0,\theta]}(u)$, et l'argument est similaire à celui du cours : si f est telle que $\int_0^\theta f(u) u^{2n-1}\, \mathrm{d}u = 0$ pour tout θ, alors la fonction $h : s \mapsto f(s)s^{2n-1}$ a intégrale nulle sur tout intervalle de $\mathbb{R}_+$, donc sur tout borélien par le lemme de la classe monotone, en particulier sur le borélien $B = \{s : f(s) > 0\}$. Comme $\int \mathbb{1}_B(s) f(s) s^{2n-1} ds = 0$ implique que $f = 0$ presque sûrement au sens de Lebesgue.

Exercice 2.4 (Identifier une direction causale, $\star$)**.** On considère un modèle avec deux variables aléatoires X_1, X_2. On dit que X_1 *cause* X_2 si la loi du modèle est de la forme suivante :
$$X_1 = U_1 \quad X_2 = f(X_1, U_2),$$
où U_1 et U_2 sont des bruits indépendants et f est une fonction mesurable.

Une question fondamentale est la suivante : *en général, peut-on inférer quoi que ce soit sur la direction du lien causal entre X_1 et X_2 en observant quelques réalisations ?*

L'objectif de cet exercice est de répondre à la négative via un contre-exemple à ce fait. On considère le modèle
$$X_1 = U_1 \quad X_2 = X_1 + U_2,$$
où U_1 et U_2 sont des variables i.i.d. de loi $\mathrm{Exp}(1)$. Montrer qu'il existe un modèle causal inversé de la forme
$$X_2 = U_2' \quad X_1 = h(X_2, U_1'),$$
où U_1' et U_2' sont indépendants, et qui préserve la loi jointe de (X_1, X_2). *Solution. On peut savoir que $X_2 \sim \Gamma(2, 1)$, donc la loi de U_2' est fixée. Si l'on ne le sait pas, on peut la retrouver*

à partir de la densité jointe que l'on calcule. Pour toute fonction continue bornée f,

$$\begin{aligned}\mathbb{E}[f(X_1, X_2)] &= \mathbb{E}[f(U_1, U_1 + U_2)] \\ &= \int_0^{+\infty} \int_0^{+\infty} f(u_1, u_1 + u_2) e^{-u_1 - u_2} \, du_1 \, du_2 \\ &= \int_0^{+\infty} \int_{x_1}^{+\infty} f(x_1, x_2) e^{-x_2} \mathbb{1}_{x_2 \geq x_1} \, dx_1 \, dx_2.\end{aligned}$$

On en déduit que la densité jointe est $p(x_1, x_2) = e^{-x_2} \mathbb{1}_{x_2 \geq x_1 \geq 0}$ et que la seconde marginale est $p(x_2) = x_2 e^{-x_2} \mathbb{1}_{x_2 \geq 0}$. À présent, la densité conditionnelle de X_1 sachant X_2 est donnée par

$$p(x_1 \mid x_2) = \frac{p(x_1, x_2)}{p(x_2)} = \frac{e^{-x_2} \mathbb{1}_{x_2 \geq x_1 \geq 0}}{x_2 e^{-x_2} \mathbb{1}_{x_2 \geq 0}} = \frac{1}{x_2} \mathbb{1}_{x_1 \in [0, x_2]} \, .$$

Il s'agit d'une loi uniforme sur $[0, x_2]$; on peut donc définir $X_2 = U_2' \sim \Gamma(2, 1)$ et $X_1 = X_2 U_1'$, où $U_1' \sim \mathrm{Unif}([0, 1])$ est indépendante de X_2.

Exercice 2.5. On observe une réalisation de $G \sim G(n, p)$, c'est-à-dire un graphe aléatoire à n sommets (n est connu) dont chaque arête est présente indépendamment avec probabilité $p \in [0, 1]$ inconnu. Trouver une statistique suffisante et complète. Interpréter.

Exercice 2.6. On considère le modèle $(\mathbb{R}, \mathcal{B}(\mathbb{R}), (\mathrm{Unif}([0, \theta]))_{\theta \geq 0})$. Trouver une mesure dominante minimale pour ce modèle. *Solution. Nous allons montrer que $\lambda = \mathrm{Leb}_{\mathbb{R}_+}$, la mesure de Lebesgue sur $\mathbb{R}_+$, est une mesure dominante minimale. Tout d'abord, il est clair que λ domine le modèle. Ensuite, soit une mesure μ telle que $\mathcal{M} \ll \mu$ mais $\lambda \not\ll \mu$. Il existe donc un borélien $A \subseteq \mathbb{R}_+$ tel que $\mu(A) = 0$ et $\lambda(A) > 0$. Comme $\lambda(A) > 0$, il existe $t > 0$ tel que $\lambda(A \cap [0, t]) > 0$, et donc $P_t(A) > 0$, où P_t désigne la loi uniforme sur $[0, t]$. Mais dans le même temps, $P_t(A) = \int_A \frac{dP_t}{d\mu} \, d\mu = 0$ puisque $\mu(A) = 0$.*

Exercice 2.7 ($\star$). Reprendre l'exemple du modèle uniforme i.i.d. du cours. Montrer que $S(X) = \max_i(X_i)$ est suffisante sans utiliser le théorème de Neyman-Fisher, en donnant la distribution de X sachant $S(X)$. Interpréter la forme de cette distribution.

3. Estimation paramétrique

Dans ces exercices, l'abbréviation UVMB désigne "uniformément de variance minimale et sans biais".

Exercice 3.1 (Deux estimateurs dans le modèle i.i.d. uniforme)**.** On considère le modèle où $X_1, \ldots, X_n$ sont i.i.d. de loi $\mathrm{Unif}([0, \theta])$, avec $\theta > 0$ inconnu.

1. Trouver un estimateur $\hat{\theta}_1$ de θ en utilisant la méthode des moments appliquée au premier moment. *Solution. Puisque $\mathbb{E}_\theta[X_1] = \theta/2$, la méthode des moments donne l'estimateur $\hat{\theta}_1 = 2\bar{X}_n$.*

2. Déterminer la vraisemblance dans ce modèle et montrer qu'il existe un estimateur du maximum de vraisemblance unique $\hat{\theta}_2$. Trouver son expression. *Solution. La vraisemblance s'écrit*

$$L(\theta, x) = \prod_{i=1}^n \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x_i) = \frac{1}{\theta^n} \mathbb{1}_{\{0 \le \min x_i\}} \mathbb{1}_{\{\max x_i \le \theta\}} \overset{\mathbb{P}_\theta - p.s.}{=} \begin{cases} 0 & \text{si } \theta < \max x_i, \\ \frac{1}{\theta^n} & \text{si } \theta \ge \max x_i, \end{cases}$$

ce qui possède un maximum global unique pour $\theta = \max_i x_i$. Ainsi, l'estimateur du maximum de vraisemblance est unique et s'écrit

$$\hat{\theta}_2 = \max_i X_i.$$

3. Calculer le biais, la variance et le risque quadratique de $\hat{\theta}_1$. *Solution. Pour $\hat{\theta}_1$, on a $\mathbb{E}_\theta[\hat{\theta}_1] = \theta$ pour tout $\theta > 0$, donc il est non biaisé. Sa variance est*

$$\mathrm{Var}_\theta(\hat{\theta}_1) = \frac{4}{n} \mathrm{Var}_\theta(X_1) = \frac{\theta^2}{3n},$$

d'où

$$R_\theta(\hat{\theta}_1) = 0 + \frac{\theta^2}{3n} = \frac{\theta^2}{3n}.$$

4. Calculer la densité de $\hat{\theta}_2$. En déduire son biais, sa variance et son risque quadratique. *Solution. Pour $\hat{\theta}_2$, la densité s'obtient via la fonction de répartition :*

$$g_\theta(x) = \frac{n x^{n-1}}{\theta^n} \mathbb{1}_{[0,\theta]}(x).$$

Ensuite,

$$\mathbb{E}_\theta[\hat{\theta}_2] = \int_0^\theta x\, g_\theta(x) dx = \frac{n}{n+1}\theta, \quad \mathbb{E}_\theta[\hat{\theta}_2^2] = \int_0^\theta x^2\, g_\theta(x) dx = \frac{n}{n+2}\theta^2,$$

ce qui donne

$$b_\theta(\hat{\theta}_2) = -\frac{\theta}{n+1}, \quad \mathrm{Var}_\theta(\hat{\theta}_2) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2,$$

et finalement

$$R_\theta(\hat{\theta}_2) = \left(\frac{\theta}{n+1}\right)^2 + \frac{n}{(n+1)^2(n+2)}\theta^2 = \frac{2}{(n+1)(n+2)}\theta^2.$$

5. Quel est le meilleur estimateur pour n assez grand ? Y a-t-il une grande différence ? *Solution.* *On remarque facilement que $R_\theta(\hat{\theta}_2) \leq R_\theta(\hat{\theta}_1)$ dès que $n \geq 2$, donc $\hat{\theta}_2$ est meilleur que $\hat{\theta}_1$. En fait, il est encore bien meilleur puisque lorsque $n \to \infty$, $R_\theta(\hat{\theta}_2) \sim \frac{2\theta^2}{n^2}$ tandis que $R_\theta(\hat{\theta}_1) \sim \frac{\theta^2}{3n}$, beaucoup plus grand.*

Exercice 3.2 (Information de Fisher dans un modèle linéaire à deux paramètres)**.** On considère le modèle statistique où $X_1, \ldots, X_n$ sont des variables réelles indépendantes, avec pour tout $1 \leq i \leq n$, $X_i \sim \mathcal{N}(\alpha + \beta t_i, 1)$, où les $(t_i)_{1 \leq i \leq n}$ sont des réels *connus* et $\alpha, \beta \in \mathbb{R}$ sont des paramètres *inconnus*. On admettra que ce modèle est régulier.

1. Déterminer une condition nécessaire et suffisante sur les $(t_i)_{1 \leq i \leq n}$ pour que ce modèle soit identifiable. *Solution.* *On peut commencer par regarder ce que donnent les moments d'ordre 1. L'application*
$$(\alpha, \beta) \mapsto \begin{pmatrix} \mathbb{E}_{(\alpha,\beta)}[X_1] \\ \vdots \\ \mathbb{E}_{(\alpha,\beta)}[X_n] \end{pmatrix} = \begin{pmatrix} \alpha + \beta t_1 \\ \vdots \\ \alpha + \beta t_n \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}}_{=:X} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$
est injective si et seulement si la matrice X a un noyau réduit à $\{0_2\}$, ce qui arrive si et seulement si elle est de rang 2, ssi les t_i ne sont pas constants. Réciproquement, si les t_i sont constants, l'identifiabilité est alors mise en défaut : toutes les variables sont i.i.d. $\mathcal{N}(\alpha + \beta t, 1)$ et on a une symétrie $(\alpha, \beta) \mapsto (0, \alpha + \beta t)$.

2. Montrer que la matrice d'information de Fisher $I(\alpha, \beta)$ s'écrit
$$I(\alpha, \beta) = \begin{pmatrix} n & \sum_{i=1}^n t_i \\ \sum_{i=1}^n t_i & \sum_{i=1}^n t_i^2 \end{pmatrix}.$$
Solution. *On écrit la log-vraisemblance qui est bien définie et vaut $\ell(\alpha, \beta, x) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n (x_i - \alpha - t_i\beta)^2$, différentiable sur $\mathbb{R}^2$ et de gradient donné par $\nabla_{\alpha,\beta}\ell(\alpha, \beta, x) = \begin{pmatrix} \sum_{i=1}^n (x_i - \alpha - t_i\beta) \\ \sum_{i=1}^n t_i(x_i - \alpha - t_i\beta). \end{pmatrix}$ La matrice de covariance de $\nabla_{\alpha,\beta}\ell(\alpha, \beta, X)$ existe et vaut*
$$I(\alpha, \beta) = \begin{pmatrix} n & \sum_{i=1}^n t_i \\ \sum_{i=1}^n t_i & \sum_{i=1}^n t_i^2 \end{pmatrix}.$$
On pouvait calculer la matrice d'information de Fisher pour un seul X_1 et utiliser l'additivité de I.

3. Déterminer une condition nécessaire et suffisante sur les $(t_i)_{1 \leq i \leq n}$ pour que $I(\alpha, \beta)$ soit inversible. Que remarque-t-on ? *Solution.* *A noter que l'information de Fisher est ici constante. Cela est du au fait que pour passer toutes les lois considérées sont reliées par translation : $X_i = \alpha + \beta t_i + Z_i$ avec l'aléa seulement dans Z_i, ne dépendant pas de α ni de β. La matrice $I(\alpha, \beta)$ est (symétrique) et inversible si et seulement si $\det I(\alpha, \beta) \neq 0$ i.e.*
$$n \sum_{i=1}^n t_i^2 \neq \left(\sum_{i=1}^n t_i \right)^2.$$
Mais par Cauchy-Schwarz, $n \sum_{i=1}^n t_i^2 \geq (\sum_{i=1}^n t_i)^2$ avec égalité si et seulement si $(t_1, \ldots, t_n)$ et $(1, \ldots, 1)$ sont colinéaires c'est-à-dire si les t_i sont constants. $I(\alpha, \beta)$ est inversible

sous les mêmes conditions que la question 1. On pourra creuser ce lien entre identifiabilité et information de Fisher dans l' exo:fisher$_i$dentifiabilite.

Exercice 3.3 (Un modèle autorégressif)**.** Nous considérons un modèle avec $(X_1, \dots, X_n)$ tels que $X_1 = \theta + Z_1$, et récursivement, $X_{j+1} = \rho(X_j - \theta) + \theta + Z_{j+1}$ avec les Z_j i.i.d. de loi $\mathcal{N}(0, \sigma^2)$. Nous supposons que $\rho \in [0, 1]$ est un paramètre connu, et que $\theta \in \mathbb{R}$ et $\sigma^2 > 0$ sont inconnus. On admettra la régularité de ce modèle.

1. Calculer la matrice d'information de Fisher $I(\theta, \sigma^2)$ en fonction de θ, σ^2, ρ et n. Solution. *Le modèle est markovien, et log-vraisemblance s'écrit :*

$$\ell(\theta, \sigma^2) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{(X_1 - \theta)^2}{2\sigma^2} - \frac{1}{2\sigma^2}\sum_{j=1}^{n-1}(X_{j+1} - \theta - \rho(X_j - \theta))^2 .$$

Son gradient vaut

$$\nabla_{\theta,\sigma^2}\ell(\theta, \sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2}\left[(X_1 - \theta) + (\rho + 1)\sum_{j=1}^{n-1}(X_{j+1} - \theta - \rho(X_j - \theta))\right] \\ \frac{1}{2\sigma^2}\left[-n + \frac{(X_1-\theta)^2}{\sigma^2} + \frac{1}{\sigma^2}\sum_{j=1}^{n-1}(X_{j+1} - \theta - \rho(X_j - \theta))^2\right] \end{pmatrix} .$$

Pour calculer la matrice de covariance associée, il peut être bon de le reformuler en fonction des Z_j :

$$\nabla_{\theta,\sigma^2}\ell(\theta, \sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2}\left[Z_1 + (1 - \rho)\sum_{j=2}^{n} Z_j\right] \\ \frac{1}{2\sigma^2}\left[-n + \frac{1}{\sigma^2}\sum_{j=1}^{n} Z_j^2\right] \end{pmatrix} .$$

En utilisant que les Z_j sont i.i.d., que moments impairs d'une gaussienne centrée sont nuls, et $\mathrm{Var}(Z_1^2) = \sigma^4(3 - 1^2) = 2\sigma^4$, on a :

$$I(\theta, \sigma^2) = \begin{pmatrix} \frac{1+(n-1)(1-\rho)^2}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix} .$$

Notons qu'elle ne dépend pas de θ : l'estimation de θ est de la même difficulté partout.

2. Donner une borne inférieure pour la variance d'un estimateur sans biais et lisse de θ. Commenter et interpréter les valeurs de cette borne lorsque $\rho = 0$, et lorsque $\rho = 1$. Solution. *La borne de Cramér-Rao pour un estimateur sans biais $\hat{\theta}$ de θ est l'inverse de l'information de Fisher en θ, donc :*

$$\mathrm{Var}(\hat{\theta}) \geq \left[I(\theta, \sigma^2)^{-1}\right]_{1,1} = \frac{\sigma^2}{1 + (n-1)(1-\rho)^2} .$$

On retrouve d'ailleurs la borne classique de Cramer-Rao lorsque $\rho = 0$: les variables sont alors i.i.d. $\mathcal{N}(\theta, \sigma^2)$. Lorsque $\rho = 1$, en reprenant l'énoncé on $X_{j+1} = X_j + Z_{j+1}$, le bruit ne fait que s'ajouter, et on ne peut pas estimer σ^2 mieux qu'avec X_1, donc avec erreur σ^2, qui ne décroit pas en n. De façon générale, cette borne croît avec ρ, on estime plus difficilement θ quand ρ augmente. En fait tout se passe comme si on travaillait dans le cadre i.i.d. avec une taille effective d'échantillon

$$n_{\text{eff}} = 1 + (n - 1)(1 - \rho)^2 \leq n .$$

Cette taille effective est inférieure à n car le cas dépendant n'amortit pas autant le bruit que le cas i.i.d., plus favorable.

Exercice 3.4 (Une Rao-Blackwellisation)**.** On considère le modèle où $X_1, \dots, X_n$ sont i.i.d. de loi[1] $\mathrm{Poi}(\lambda)$. On souhaite estimer $\varphi(\lambda) = e^{-\lambda}$.

1. Montrer que $T(X) = \mathbb{1}_{X_1=0}$ est un estimateur sans biais de $\varphi(\lambda)$. Calculer sa variance. Est-ce un bon estimateur ?

2. Prouver que $S(X) = \sum_{i=1}^n X_i$ est une statistique complète et suffisante pour le modèle. *Solution. Suffisant par Neyman-Fisher, complet en écrivant* $e^{-\lambda n} \sum_{k\geq 0} f(k) \frac{(\lambda n)^k}{k!} = 0$ *pour tout* λ *et en identifiant la série entière (ou, en dérivant).*

3. Calculer l'estimateur Rao-Blackwellisé T^* de T par S. *Solution. On trouve* $T^* = (1 - 1/n)^S$.

4. Montrer que T^* est UVMB et qu'il est asymptotiquement efficace. *Solution. on sait qu'il est UVMB par Lehman-Scheffé, mais est-il de variance comparable à la borne de Cramer-Rao ? Calculons la. L'information de Fisher vaut* $n \times I_1(\lambda) = n/\lambda$ *(simple calcul via la vraisemblance). La borne de CR vaut donc* $\frac{\varphi'(\lambda)^2}{I(\lambda)} = \frac{\lambda e^{-2\lambda}}{n}$. *On utilise ensuite la transformée de Laplace d'une Loi de Poisson : si* $X \sim \mathrm{Poi}(\lambda)$, $\mathbb{E}[e^{tX}] = e^{\lambda(e^t-1)}$. *Cela donne* $\mathbb{E}_\lambda[T^*] = e^{-\lambda}$, *on le savait, mais surtout* $\mathbb{E}_\lambda[(T^*)^2] = e^{\lambda n \times (e^{2\log(1-1/n)}-1)} = e^{-2\lambda + \lambda/n}$. *Cela donne*
$$\mathrm{Var}_\lambda(T^\star) = e^{-2\lambda}(e^{\lambda/n} - 1) \sim \frac{\lambda e^{-2\lambda}}{n}\,.$$

5. Montrer que T^* est fortement consistant. *Solution.* $T^* = (1 - 1/n)^{S_n} = \exp(n \log(1 - 1/n) \times (S_n/n))$ *et on applique la loi des grands nombres.*

6. ($\star$, car un peu technique) Montrer que T^* est asymptotiquement normal en précisant les paramètres asymptotiques. *Solution. Le TCL appliqué à* S_n/n, *avec* $S = S_n$, *donne*
$$\sqrt{n}\left(\frac{S_n}{n} - \lambda\right) \to \mathcal{N}(0, \lambda)\,.$$
On aimerait écrire $T^* = g(S_n/n)$, *mais ici* g *dépendrait de* n:
$$T^* = (1 - 1/n)^{S_n} = \exp(n \log(1 - 1/n) \times (S_n/n))\,.$$
On sait appliquer la méthode Delta à $V_n = \exp(-S_n/n)$, *ici* $g(x) = \exp(-x)$ *et elle donne*
$$\sqrt{n}(V_n - e^{-\lambda}) \xrightarrow[n\to\infty]{(d)} -e^{-\lambda}\mathcal{N}(0, \lambda) = \mathcal{N}(0, \lambda e^{-2\lambda})\,.$$
Il ne resterait qu'à montrer que $\sqrt{n}(T^* - V_n) \xrightarrow[n\to\infty]{(d)} 0$ *et appliquer le lemme de Slutsky.*

Exercice 3.5 (Information de Fisher et identifiabilité locale)**.** On dit qu'un modèle $\mathcal{M} = (\mathcal{X}, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ est *localement identifiable* en $\theta \in \Theta$ si il existe un voisinage U de θ dans Θ tel que
$$\forall \theta' \in U, (\mathbb{P}_{\theta'} = \mathbb{P}_\theta) \implies \theta = \theta'\,.$$

On suppose que le modèle est paramétrique régulier avec $\Theta \subseteq \mathbb{R}^p$. On fait une hypothèse supplémentaire, un peu plus forte que la régularité du cours :

(H1) la log-vraisemblance ℓ est deux fois différentiable en θ sur Θ, et ses dérivées secondes sont localement dominées uniformément en θ.

[1] rappel : c'est la loi telle que pour tout $k \in \mathbb{N}$, $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$.

(H2) la Hessienne $\theta \mapsto \nabla^2 \ell(\theta, x)$ est localement Lispchitz, uniformément en x au sens ou pour tout $\theta \in \Theta$, il existe V voisinage de θ, $C_\theta > 0$ tels que pour tous $\theta' \in V$, $v \in \mathbb{R}^p$,

$$\sup_x |v^\top (\nabla^2 \ell(\theta', x) - \nabla^2 \ell(\theta, x)) v| \leq C_\theta \|\theta' - \theta\| \|v\|^2 .$$

Le but de cet exercice est de démontrer que sous (H1), si l'information de Fisher $I(\theta)$ est inversible en θ, alors le modèle est localement identifiable en θ.

1. Montrer que sous (H1), pour tout $\theta \in \Theta$,

$$I(\theta) = -\mathbb{E}_\theta[\nabla^2 \ell(\theta, X)] .$$

Pour $\theta, \theta' \in \Theta$, on définit la divergence de Kullback-Leibler entre $\mathbb{P}_\theta$ et $\mathbb{P}_{\theta'}$ par:

$$\mathrm{KL}(\mathbb{P}_\theta \| \mathbb{P}_{\theta'}) := \mathbb{E}_\theta \left[\ell(\theta, X) - \ell(\theta', X) \right] .$$

Pour $\theta \in \Theta$ et $v \in \mathbb{R}^p$ assez petit, un développement de Taylor de $\ell(\theta + v, x)$ autour de $\ell(\theta, x)$ donne

$$\ell(\theta + v, x) = \ell(\theta, x) + v^T \nabla_\theta \ell(\theta, x) + \frac{1}{2} v^T \nabla_\theta^2 \ell(\theta, x) v + R(v, \theta, x),$$

avec

$$R(v, \theta, x) = \frac{1}{2} \int_0^1 (1-t) v^\top (\nabla_\theta^2 \ell(\theta + tv, x) - \nabla_\theta^2 \ell(\theta, x)) v \, dt.$$

2. Soit $\theta \in \Theta$. Montrer qu'avec ce qui précède, pour v tel que $\|v\|$ est assez petit,

$$\mathrm{KL}(\mathbb{P}_\theta \| \mathbb{P}_{\theta+v}) = v^T I(\theta) v + o(\|v\|^2),$$

   où $o(\|v\|^2)$ est une fonction de la forme $\|v\|^2 \varepsilon(v)$ tel que $\varepsilon(v) \to 0$ quand $v \to 0$. *On justifiera également que le membre de gauche est bien défini pour v assez petit.* *Solution. Le membre de gauche est bien défini pour v assez petit car Θ est un ouvert de $\mathbb{R}^p$. On a*

$$\mathrm{KL}(\mathbb{P}_\theta \| \mathbb{P}_{\theta+v}) = \int (\ell(\theta, x) - \ell(\theta + v, x)) \, L(\theta, x) \, d\nu(x).$$

   En substituant le développement limité de $\ell(\theta + v, x)$:

$$\mathrm{KL}(P_\theta \| P_{\theta+v}) = -\int \left(v^T \nabla_\theta \ell(\theta, x) + \frac{1}{2} v^T \nabla_\theta^2 \ell(\theta, x) v + R(v, \theta, x) \right) L(\theta, x) \, dx.$$

   Le premier terme est nul car le score est centré. Le second terme vaut, d'après la question 1,

$$-\int \frac{1}{2} v^T \nabla_\theta^2 \ell(\theta, x) v L(\theta, x) \, dx = -\frac{1}{2} v^T \mathbb{E}_\theta[\nabla_\theta^2 \ell(\theta, x)] v = \frac{1}{2} v^T I(\theta) v.$$

   Pour le troisième terme, on conclut via (H2) et le TCD.

3. Conclure. *Solution. Nullité de la KL ssi égalité dans Jensen $x \log x \geq 0$ avec égalité ssi $x = 1$.*

4. ($\star$) On se place dans le modèle ou l'on observe X qui a pour loi $\mathcal{N}(a, 1)$ avec probabilité $1/2$, et $\mathcal{N}(b, 1)$ avec probabilité $1/2$. On se place dans l'espace des paramètres $\Theta = \mathbb{R}^2 \backslash \{(a, a), a \in \mathbb{R}\}$. On admet que ce modèle est régulier. Montrer que $I(\theta)$ est inversible en tout point de Θ, mais que le modèle n'est pas identifiable globalement (c'est-à-dire au sens du cours). *Solution. globalement : symétrie $(a, b) \to (b, a)$ mais localement, identifiable dans des intervalles de rayon $< |a - b|/2$ autour de a et de b.*

Exercice 3.6 (Lois de puissance)**.** La loi de puissance de paramètre $\alpha > 1$, notée $\mathrm{Pow}(\alpha)$ ci-après, a pour densité de Lebesgue $f_\alpha : x \mapsto C_\alpha x^{-\alpha}\mathbb{1}_{[1,\infty)}(x)$, où C_α est une constante. On observe $X_1, \ldots, X_n$ i.i.d. $\sim \mathrm{Pow}(\alpha)$ où $\alpha > 1$ est inconnu.

1. Calculer C_α pour tout $\alpha > 1$. *Solution. L'intégration montre que* $C_\alpha = \alpha - 1$.

2. Pour $X \sim \mathrm{Pow}(\alpha)$, X a-t-il une espérance finie ? Une variance finie ? Si oui, les donner. *Solution. Espérance finie seulement si* $\alpha > 2$*, dans ce cas* $\mathbb{E}_\alpha[X] = \frac{\alpha-1}{\alpha-2}$*, le second moment est fini si et seulement si* $\alpha > 3$ *et* $\mathbb{E}_\alpha[X^2] = \frac{\alpha-1}{\alpha-3}$*, et la variance est* $\mathrm{Var}_\alpha(X) = \frac{(\alpha-1)[(\alpha-2)^2-(\alpha-1)(\alpha-3)]}{(\alpha-3)(\alpha-2)^2} = \frac{\alpha-1}{(\alpha-3)(\alpha-2)^2}$.

3. Si $\alpha > 2$, en utilisant le premier moment, fournir un estimateur $\widehat{\alpha}_1$ de α. Montrer que $\widehat{\alpha}_1$ est fortement consistant. *Solution.* $\widehat{\alpha}_1 = \frac{1}{\bar{X}_n - 1} + 2$*. On applique la loi forte des grands nombres pour prouver la forte consistance.*

4. Montrer que l'estimateur du maximum de vraisemblance $\widehat{\alpha}_2$ est donné par :

$$\widehat{\alpha}_2 = 1 + \left(\frac{1}{n}\sum_{i=1}^{n} \log X_i\right)^{-1}.$$

Montrer que $\widehat{\alpha}_2$ est également consistant. *Indice : on peut étudier la distribution de* $\log X_1$ *quand* $X_1 \sim Pow(\alpha)$.

Solution. La log-vraisemblance s'écrit

$$\ell(\alpha, x) = n\log(\alpha - 1) - \alpha \sum \log x_i,$$

qui est concave en α*. Résoudre les conditions du premier ordre donne le résultat désiré. On peut calculer l'espérance de* $\log X_1$ *pour montrer la forte consistance par la loi forte des grands nombres. Mais suivant l'indice, on peut trouver la distribution de* $\log X_1$ *avec une fonction test :*

$$\begin{aligned}\mathbb{E}_\alpha[\varphi(\log X_1))] &= (\alpha-1)\int_1^{+\infty} \varphi(\log x) x^{-\alpha} dx \\ &\underset{u=\log x}{=} (\alpha-1)\int_0^{+\infty} \varphi(u) e^{-u\alpha} e^u du \\ &= \int_0^{+\infty} \varphi(u)(\alpha-1)e^{-(\alpha-1)u} du,\end{aligned}$$

ce qui montre que $\log X_1 \sim \mathrm{Exp}(\alpha - 1)$.

5. Montrer que $\widehat{\alpha}_2$ est asymptotiquement normal, et donner les caractéristiques de la loi limite. Faire de même pour $\widehat{\alpha}_2$ sous une hypothèse appropriée sur α.

6. Sur la base des résultats précédents, quel estimateur préféreriez-vous pour n grand ?

7. Déterminer l'information de Fisher $I(\alpha)$. Comparer avec les résultats précédents. Que pouvez-vous en conclure ? *Solution.* $\ell(\alpha, x) = n\log(\alpha-1) - \alpha\sum_i^n \log x_i$ *et* $\dot{\ell}(\alpha, x) = \frac{n}{\alpha-1} - \sum_i^n \log x_i$*, et* $I(\alpha) = \mathrm{Var}_\alpha(\dot{\ell}(\alpha, X)) = \frac{n}{(\alpha-1)^2}$*, sachant que* $\log X_1 \sim \mathrm{Exp}(\alpha - 1)$*. On voit que l'estimateur MV atteint la borne de Cramér-Rao asymptotiquement.*

Exercice 3.7 (Non-existence d'un UVMB)**.** L'objet de cet exercice est l'étude d'un exemple où des estimateurs sans biais existent, mais où il n'existe pas d'estimateur sans biais de variance uniformément minimale (UVMB). Soit X prenant ses valeurs dans $\{-1, 0, 1, 2, \ldots\}$

avec la distribution suivante de paramètre $p \in (0,1)$: $\mathbb{P}_p(X=-1)=p, \quad \mathbb{P}_p(X=k)=(1-p)^2p^k, \quad k=0,1,2,\dots$ On cherche à estimer p.

1. Montrer que $\mathbb{1}_{\{X=-1\}}$ est un estimateur sans biais de p. *Solution. Oui, $\mathbb{1}_{\{X=-1\}}$ est sans biais puisque $\mathbb{E}_p[\mathbb{1}_{\{X=-1\}}]=\mathbb{P}_p(X=-1)=p$.*

2. Montrer que la classe de tous les estimateurs sans biais de p est un espace affine de dimension 1 que vous décrirez. *Solution. On commence par caractériser l'ensemble des estimateurs sans biais de zéro, $\mathcal{U}_0$, et on montre que $U(k)=-kU(-1)$ pour tout $k\geq 0$. Maintenant, comme tout estimateur sans biais peut s'écrire comme $\mathbb{1}_{\{X=-1\}}+U$ avec $U\in\mathcal{U}_0$, en posant $U(-1)=a$, on obtient la forme:*
$$T_a(X)=(1+a)\mathbb{1}_{\{X=-1\}}-aX\mathbb{1}_{\{X\geq 0\}},$$
ou
$$T_a(X)=\mathbb{1}_{\{X=-1\}}-aX\,.$$
La synthèse découle de ce que $\mathbb{E}_p[X]$ est toujours nul.

3. Pour $p\in(0,1)$, trouver T estimateur non biaisé de p qui minimise $\mathrm{Var}_p(T)$. *On donne, pour $|p|<1$,*
$$\sum_{k=0}^{+\infty}k^2p^k=\frac{p(1+p)}{(1-p)^3}\,.$$
Solution. On minimise : $\mathrm{Var}_p(T_a)=\mathbb{E}_p(T_a^2)=(1+a)^2p+(1-p)^2a^2\sum_{k=0}^{\infty}k^2p^k=(1+a)^2p+a^2\frac{p(p+1)}{1-p}$. En dérivant par rapport à a et en annulant la dérivée, on obtient : $a=-\frac{p}{p+\frac{p(p+1)}{1-p}}=-\frac{1-p}{2}$. Ainsi, l'optimal a dépend de p.

4. Conclure qu'il n'existe pas d'UVMB pour p.

Exercice 3.8 (Fonctions estimables sans biais)**.** On travaille dans un modèle paramétrique et l'on note $\mathcal{U}$ l'ensemble des fonctions de θ à valeurs dans $\mathbb{R}^\ell$ qui sont estimables sans biais, c'est-à-dire
$$\mathcal{U}:=\left\{\varphi:\Theta\to\mathbb{R}^\ell,\exists T:\mathcal{X}\to\mathbb{R}^\ell \text{ mesurable, intégrable, t.q.,}\forall\theta\in\Theta,\mathbb{E}_\theta[T(X)]=\varphi(\theta)\right\}.$$

1. Montrer que $\mathcal{U}$ est un sous-espace vectoriel du $\mathbb{R}-$espace vectoriel des fonctions de Θ dans $\mathbb{R}^\ell$.

2. Montrer que si $\mathcal{X}$ est fini, alors $\dim\mathcal{U}\leq|\mathcal{X}|$.

Dans toute la fin de l'exercice, on se place dans le modèle à une seule observation $X\sim\mathrm{Unif}([0,\theta])$, avec $\theta>0$ inconnu. On considère le cas $\ell=1$.

3. $(\star)$ Montrer que
$$\mathcal{U}=\left\{\varphi:\mathbb{R}_+^*\to\mathbb{R},\varphi \text{ absolument continue}\right\}.$$
on pourra utiliser le théorème fondamental du calcul intégral (de Lebesgue): pour $f:[a,b]\to\mathbb{R}$ Lebesgue-intégrable sur $[a,b]$ et $F:x\mapsto\int_a^x f(t)\,dt$, F est absolument continue donc dérivable presque partout sur $[a,b]$ et, pour presque tout $x\in[a,b]$, $F'(x)=f(x)$. Solution. Si $\varphi\in\mathcal{U}$ alors pour tout $\theta>0$, $\varphi(\theta)=\frac{1}{\theta}\int_0^\theta T(x)dx$. Par le théorème de Lebesgue, φ est absolument continue. Réciproquement, si φ est absolument continue, alors $x\mapsto x\varphi(x)$ l'est aussi, et est dérivable presque partout. En posant $T(x)=[x\varphi(x)]'=\varphi(x)+x\varphi'(x)$ quand cela existe, et 0 ailleurs, T estime $\varphi(\theta)$ sans biais. Notons qu'une fonction discontinue avec un saut, comme $\mathbb{1}_{\theta>1}$ n'est pas dans $\mathcal{U}$, car pas absolument continue. Si elle l'était, elle serait égale à l'intégrale de sa dérivée, c'est à dire 0, absurde.