TD 4 : Maximum de vraisemblance

Exercice 1 (Modèles de translation) Soit $X_1, \dots, X_n$ i.i.d. de densité commune $f_\theta(x) = f(x-\theta)$, $\theta \in \mathbb{R}$. Déterminer l'estimateur du maximum de vraisemblance dans les cas suivants :

1. f est la densité de la loi $\mathcal{N}(0, \sigma^2)$ (σ connue) ;
2. $f(x) = \frac{1}{2}e^{-|x|}$ (loi de Laplace) ;
3. $f(x) = \frac{3}{4}(1-x^2)$ sur $[-1, 1]$.

Exercice 2 (Loi discrète) Soit $\theta \in]0,1[$ un paramètre inconnu, on note X une variable aléatoire de loi définie par

$$\mathbb{P}_\theta(X = k) = (k+1)(1-\theta)^2\theta^k, \text{ pour tout } k \in \mathbb{N}.$$

On donne

$$\mathbb{E}_\theta[X] = \frac{2\theta}{1-\theta} \quad \text{et} \quad \mathrm{Var}_\theta[X] = \frac{2\theta}{(1-\theta)^2}.$$

On souhaite estimer θ à partir d'un échantillon $X_1, \dots, X_n$ de même loi que X.

1. Donner un estimateur $\widehat{\theta}_n$ de θ par la méthode des moments.
2. L'estimateur du maximum de vraisemblance $\widetilde{\theta}_n$ de θ est-il bien défini ?
3. Étudier la consistance de $\widehat{\theta}_n$ et déterminer sa loi limite.

Exercice 3 (Loi exponentielle translatée) On observe un échantillon $X_1, \dots, X_n$ dont la loi admet la densité

$$f_\theta(x) = \exp(-(x-\theta))\mathbb{1}_{[\theta,+\infty[}(x),$$

où θ est un paramètre réel inconnu.

1. Quels estimateurs de θ pouvez-vous proposer en utilisant les méthodes usuelles ?
2. Déterminer la loi de $n(\widehat{\theta}_n - \theta)$, où $\widehat{\theta}_n$ est l'estimateur du maximum de vraisemblance. En déduire un intervalle de confiance pour θ de niveau $(1-\alpha)$, pour $\alpha \in]0,1[$.
3. On souhaite tester au niveau α

$$H_0 : \theta \geq 0 \text{ contre } H_1 : \theta < 0.$$

   (a) Construire un test à partir de l'intervalle de confiance de la question 2, calculer sa puissance et donner son allure (pour n et α fixés). Quelle est sa taille $\alpha^\star$?
   (b) Proposer un autre test qui soit, lui, de taille α.
   (c) Calculer la fonction puissance du test. La représenter en fonction de θ pour n et α fixés.
   (d) Comment varie la puissance en fonction de α ? en fonction de n ?

4. Proposer un test de niveau α de H_0 : « La loi de X_1 est une loi exponentielle » contre H_1 : « La loi de X_1 n'est pas une loi exponentielle ». Calculer la puissance de ce test.

Exercice 4 (Densité en escalier) Pour tout paramètre réel θ, on définit sur $\mathbb{R}$ la fonction

$$f_\theta(x) = \begin{cases} 1-\theta & \text{si } -1/2 < x \leq 0, \\ 1+\theta & \text{si } 0 < x \leq 1/2, \\ 0 & \text{sinon.} \end{cases}$$

1. Quelles conditions doit vérifier θ pour que f_θ soit une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$? On suppose ceci vérifié dans la suite.
2. Soit $(X_1, \ldots, X_n)$ un échantillon de densité f_θ. L'estimateur du maximum de vraisemblance $\widehat{\theta}_n$ de θ est-il bien défini ?
3. L'estimateur $\widehat{\theta}_n$ est-il sans biais ? consistant ? asymptotiquement normal ?

Exercice 5 (Données censurées) Soient $\theta^\star > 0$ et $X_1, \ldots, X_n$ un échantillon de variables aléatoires i.i.d. distribuées selon $\mathcal{E}(\theta^\star)$. On note, pour tout $i \in \{1, \ldots, n\}$, $Y_i := \min(X_i, 1)$.

1. Déterminer la loi de Y_1, notée $P_{\theta^\star}$ La variable aléatoire Y_1 est-elle discrète ? Est-elle à densité ?
2. On se donne le modèle statistique $\mathcal{P} = \{P_\theta, \theta > 0\}$. Donner une mesure dominante μ de ce modèle. Pour tout $\theta > 0$, déterminer une densité (notée f_θ) de P_θ par rapport à μ.
3. Calculer l'EMV $\widehat{\theta}_n$ de $\theta^\star$.
4. Montrer que $\widehat{\theta}_n$ est un estimateur consistant de $\theta^\star$.
5. ($\star$) À l'aide de l'indication ci-dessous, montrer que $\widehat{\theta}_n$ est asymptotiquement normal.
   <u>Indication (TCL et méthode delta multivariés)</u> : Soit $(U_n)_{n\geq 1}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^2$, i.i.d. et de carrés intégrables. En notant $\overline{U}_n = \frac{1}{n}\sum_{i=1}^n U_i$, on a alors
   $$\sqrt{n}\left(\overline{U}_n - \mathbb{E}[U_1]\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \mathrm{Var}(U_1)\right),$$
   où $\mathrm{Var}(U_1)$ est la matrice de variance-covariance de $U_1 = \begin{pmatrix} U_x \\ U_y \end{pmatrix}$, définie par
   $$\mathrm{Var}(U_1) = \begin{pmatrix} \mathrm{Var}(U_x) & \mathrm{Cov}(U_x, U_y) \\ \mathrm{Cov}(U_x, U_y) & \mathrm{Var}(U_x) \end{pmatrix}.$$
   De plus, pour une fonction $\varphi : \mathbb{R}^2 \to \mathbb{R}$ différentiable et telle que $\nabla\varphi(\mathbb{E}[U_1]) \neq 0$, la méthode delta assure que
   $$\sqrt{n}\left(\varphi(\overline{U}_n) - \varphi(\mathbb{E}[U_1])\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \nabla\varphi(\mathbb{E}[U_1])^\top \mathrm{Var}(U_1) \nabla\varphi(\mathbb{E}[U_1])\right).$$
6. ($\star$) Montrer que le modèle est régulier. Que pouvez-vous en déduire ?

Exercice 6 (Méthode de capture-recapture) On souhaite estimer le nombre de poissons, $N \in \mathbb{N}^*$, vivant dans un bassin. Pour ce faire, nous mettons en place le protocole suivant :

— on prélève au chalut un groupe de poissons et l'on note $k \in [\![1, N]\!]$ leur nombre ;
— ces poissons sont marqués puis relâchés dans le bassin ;
— pour un nombre de prises $n \in \mathbb{N}^*$ défini et aussi grand qu'on veut, on pêche au hasard puis on relâche successivement n poissons ;
— de retour à quai, on reporte le nombre $x \in [\![0, n]\!]$ de poissons marqués parmi les n que l'on a pêchés puis relâchés.

On fait ensuite appelle à un statisticien afin de proposer une estimation de N à partir de cette expérience.

1. Quelle est ici la quantité mesurée (et enregistrée) ? En déduire une modélisation probabiliste de l'expérience, puis un modèle statistique paramétrique indicé par un intervalle Θ.
2. L'estimateur du maximum de vraisemblance est-il défini ? Proposer un estimateur $\widehat{N}$ de N.
3. Montrer que $\widehat{N}$ est consistant et asymptotiquement normal lorsque n tend vers l'infini.
4. Au regard de la question précédente, quelle valeur de k est pertinente pour l'expérience ?

Exercice 7 (Maximum de vraisemblance et donnée aberrante) Le but de cet exercice est de montrer sur un exemple très simple que l'EMV n'est pas robuste à la présence de données aberrantes. Afin de simplifier les calculs, on fera les approximations suivantes : $\Phi^{-1}(0.975) \approx 2$, $\Phi(-3) \approx 10^{-3}$, $\Phi(-7) \approx 0$ et $99/10^4 \approx 10^{-2}$.

1. Soit $\theta \in \mathbb{R}$ un paramètre inconnu et $X_1, \ldots, X_{100}$ i.i.d. selon une loi $\mathcal{N}(\theta, 1)$. Si on note $\widehat{I} = [\overline{X}_{100} - 0.2 \ ; \ \overline{X}_{100} + 0.2]$, montrer que $\mathbb{P}(\theta \in \widehat{I}) \approx 0.95$. On suppose désormais $\theta = 0$, ce qui donne donc $\mathbb{P}(0 \in \widehat{I}) \approx 0.95$, i.e. la vraie valeur du paramètre se situe 95% du temps dans l'intervalle (aléatoire) $\widehat{I}$.
2. On suppose maintenant la présence d'une seule donnée aberrante (dysfonctionnement du dispositif d'enregistrement des données, etc.) et on veut voir l'incidence de celle-ci sur le résultat précédent. Plus précisément, on suppose que $X_1, \ldots, X_{99}$ sont i.i.d. selon une loi $\mathcal{N}(0, 1)$ et que $X_{100} = 50$. Donner la loi de $\overline{X}_{100}$ sachant $X_{100} = 50$.
3. On note toujours $\widehat{I} = [\overline{X}_{100} - 0.2 \ ; \ \overline{X}_{100} + 0.2]$. En déduire que $\mathbb{P}(0 \in \widehat{I} \mid X_{100} = 50) \approx 10^{-3}$ et conclure.

Exercice 8 ($\star$) (Gauss et la loi normale [1]) En statistique, la loi normale a été introduite par Gauss pour un problème d'estimation de paramètre en astronomie. Dans le langage moderne, voici le problème : soit $X_1, \ldots, X_n$ i.i.d. de densité commune $f_\theta(x) = f(x - \theta)$, $\theta \in \mathbb{R}$. On a donc affaire à un modèle de translation dans lequel f correspond à la densité de la loi des erreurs et sur laquelle on fait les hypothèses "naturelles" suivantes : f est paire, de classe C^1 et strictement positive sur $\mathbb{R}$. La question à laquelle Gauss souhaitait répondre est la suivante : quelles sont les densités f pour lesquelles l'estimateur des moindres carrés correspond à celui du maximum de vraisemblance ? Comme nous allons le voir, seules les lois normales centrées satisfont cette propriété.

1. Montrer que $\overline{X}_n$ est l'estimateur des moindres carrés.
2. On note $g = \ln f$. Montrer que si $\overline{X}_n$ est l'EMV alors pour tout n-uplet $(x_1, \ldots, x_n)$ de réels, la fonction g doit vérifier $\sum_{i=1}^n g'(x_i - \overline{x}_n) = 0$.
3. En prenant $x_2 = \cdots = x_n = x_1 - nt$, en déduire que, $\forall t \in \mathbb{R}^\star$, $\forall k \in \mathbb{N}^\star$, on a $\frac{g'(kt)}{kt} = \frac{g'(t)}{t}$.
4. En déduire qu'il existe une constante c telle que, pour tout $t \in \mathbb{R}^\star$, $\frac{g'(t)}{t} = c$.
5. Conclure.

Exercice 9 (Loi uniforme translatée) Soient $X_1, \ldots, X_n$ des observations i.i.d. de densité f_θ, θ réel, donnée par $f_\theta(x) = \mathbb{1}_{[\theta, \theta+1]}(x)$.

1. Comment peut-on définir l'estimateur du maximum de vraisemblance de θ ?
2. Déterminer les lois exactes et asymptotiques de $n(X_{(1)} - \theta)$ et de $n(\theta + 1 - X_{(n)})$.

1. Cet exercice est inspiré de l'article *Plaidoyer pour la loi normale*, de Aimé Fuchs.

3. Montrer que pour toute variable aléatoire Y à valeurs dans $\mathbb{R}_+$ et tout $k \in \mathbb{N}^\star$ (**on pourra retenir cette formule, ainsi que sa preuve**)

$$\mathbb{E}[Y^k] = k \int_0^{+\infty} t^{k-1} \mathbb{P}(Y > t)\, \mathrm{d}t.$$

   En déduire les deux premiers moments de $n(X_{(1)} - \theta)$.
4. Étudier la consistance des estimateurs du maximum de vraisemblance de θ.
5. Majorer $\sup_{n \geq 1} n^2 R(\widehat{\theta}_n, \theta)$ pour tout estimateur du maximum de vraisemblance $\widehat{\theta}_n$ de θ.
6. Proposer un intervalle de confiance pour θ au niveau 95%. Donner l'application numérique pour $\alpha = 0.05$ et $\widehat{\theta}(\omega) = 3$, avec $n = 10$ et $n = 100$.

Exercice 10 (Mélange de lois uniformes) Soit $X_1, \ldots, X_n$ i.i.d. dont la loi admet la densité

$$f_{\theta,\lambda}(x) = \lambda \mathbb{1}_{[0,1]}(x) + \frac{1-\lambda}{\theta} \mathbb{1}_{[0,\theta]}(x),$$

où $\theta \geq 1$ et $\lambda \in [0, 1[$ sont deux paramètres inconnus.

1. Montrer que, si on exclut une valeur de θ à préciser, le modèle est identifiable. Sauf indication contraire, on supposera cette condition vérifiée dans la suite.
2. Montrer que la densité $f_{\theta,\lambda}$ peut s'écrire, pour tout $x \in \mathbb{R}$,

$$f_{\theta,\lambda}(x) = \left(\lambda + \frac{1-\lambda}{\theta}\right)^{\mathbb{1}_{[0,1](x)}} \left(\frac{1-\lambda}{\theta}\right)^{\mathbb{1}_{[1,\theta]}(x)} \mathbb{1}_{[0,\theta]}(x).$$

3. Déterminer l'EMV de λ lorsque θ est connu, l'EMV de θ lorsque λ est connu, puis l'EMV de (θ, λ) lorsque les 2 paramètres sont inconnus.
4. Étudier la consistance et la loi limite de l'EMV de θ, que λ soit ou non connu.
5. Étudier la consistance de l'EMV de λ lorsque θ est connu, puis lorsque θ est inconnu. Que se passe-t-il si $\theta = 1$?
6. Construire un test de $H_0 : \theta = 1$ contre $H_1 : \theta > 1$ au niveau α, avec $\alpha \in]0, 1[$. Pour $\lambda \in [0, 1[$ et $\theta > 1$ fixés, étudier la limite de la puissance lorsque n tend vers l'infini.

Exercice 11 ($\star$) (Non-robustesse de l'EMV) Sur l'ensemble des densités de probabilité sur $\mathbb{R}$ (ensemble des classes d'équivalence de fonctions de L_1 positives et d'intégrale 1), on définit le carré de la distance de Hellinger h entre f et g par

$$h^2(f, g) = \frac{1}{2} \int_{\mathbb{R}} \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx.$$

1. Montrer que

$$h^2(f, g) = 1 - \int_{\mathbb{R}} \sqrt{f(x)g(x)}\, dx.$$

   Vérifier que h est une distance sur l'ensemble des densités sur $\mathbb{R}$ (i.e. positivité avec nullité si et seulement si égalité, symétrie et inégalité triangulaire), avec de plus $0 \leq h(f, g) \leq 1$.
2. Soit $\theta > 0$ et f_θ la densité de la loi uniforme sur $[0, \theta]$. Montrer que $h^2(f_\theta, f_{\theta'}) = 1 - \sqrt{\frac{\theta}{\theta'}}$ si $\theta \leq \theta'$.

3. Soit $X_1, \dots, X_n$ i.i.d. de loi uniforme sur $[0, \theta]$ pour un $\theta > 0$ inconnu. Donner l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$, calculer $\mathbb{E}_\theta\left[\sqrt{\widehat{\theta}_n}\right]$ et en déduire que $\mathbb{E}_\theta\left[h^2(f_\theta, f_{\widehat{\theta}_n})\right] = \frac{1}{2n+1}$.
4. Soit la densité
$$f_n^*(x) = 10\left(1 - \frac{1}{n}\right)\mathbf{1}_{0 \le x \le 1/10} + \frac{10}{n}\,\mathbf{1}_{9/10 \le x \le 1}.$$
   (a) Montrer que si, pour tout n, Y_n a pour densité f_n^*, alors la suite (Y_n) converge en loi et identifier la limite.
   (b) Calculer $h^2(f_n^*, f_{1/10})$.
5. Soit $n > 1$. Les variables $X_1, \dots, X_n$ sont désormais i.i.d. de densité f_n^*, mais on les croit toujours i.i.d. suivant une densité uniforme $(f_\theta)_{\theta > 0}$, avec θ inconnu. En particulier, l'estimateur $\widehat{\theta}_n$ est le même que celui défini ci-dessus.
   (a) Montrer qu'avec probabilité $1 - (1 - 1/n)^n$, on a : $9/10 \le \widehat{\theta}_n \le 1$.
   (b) Montrer que, sur l'événement $\{9/10 \le \widehat{\theta}_n \le 1\}$, on a :
$$h^2(f_n^*, f_{\widehat{\theta}_n}) \ge 1 - \frac{1}{3}\sqrt{1 - \frac{1}{n}} - \frac{1}{3\sqrt{n}}.$$
   (c) En notant $\mathbb{E}_n^*[\cdot]$ l'espérance par rapport au modèle où les données $X_1, \dots, X_n$ sont i.i.d. de densité f_n^*, en déduire que $\mathbb{E}_n^*[h^2(f_n^*, f_{\widehat{\theta}_n})] \ge u_n$, avec $\lim_{n\to\infty} u_n = 2/3 \times (1 - 1/e)$.
   (d) Conclure quant à la robustesse de l'estimateur du maximum de vraisemblance à une mauvaise spécification du modèle.

Exercice 12 ($\star$) (Un exemple d'EMV non consistant [2]) On considère l'espace de paramètres $\Theta = [0, 1]$ et la famille suivante de densités sur $\mathbb{R}$ par rapport à la mesure de Lebesgue : $f_1(x) = \frac{1}{2}\mathbf{1}_{[-1,1]}(x)$ si $\theta = 1$ et, pour tout $0 \le \theta < 1$,
$$f_\theta(x) = \theta f_1(x) + (1 - \theta)\frac{1}{c(\theta)}\left(1 - \frac{|x - \theta|}{c(\theta)}\right)\mathbf{1}_{[\theta - c(\theta), \theta + c(\theta)]}(x),$$
où $c : [0, 1[\to]0, 1]$ est continue décroissante telle que $c(0) = 1$ et $0 < c(\theta) \le 1 - \theta$ pour tout $\theta \in]0, 1[$, fonction que nous spécifierons ultérieurement. Autrement dit, P_θ est un mélange pondéré par θ entre une loi uniforme et une loi triangulaire (faire un dessin). Dans la suite, $\theta^\star \in [0, 1]$ désigne la vraie valeur du paramètre et $(X_1, \dots, X_n)$ un échantillon selon $P_{\theta^\star}$.

1. Montrer que, dans le modèle $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, il existe toujours un EMV $\widehat{\theta}_n \in [0, 1]$.
2. En notant comme d'habitude $\ell_n(\theta) = \sum_{i=1}^n \log f_\theta(X_i)$ la log-vraisemblance, montrer que pour tout $\alpha \in [0, 1[$
$$\sup_{0 \le \theta \le \alpha} \frac{1}{n}\ell_n(\theta) \le \log\left(\frac{1}{2} + \frac{1}{c(\alpha)}\right) < \infty.$$
3. Soit $M_n := \max_{1 \le i \le n} X_i$, montrer que (M_n) tend presque sûrement vers 1.
4. Prouver que, si $0 \le M_n \le 1$, alors
$$\frac{1}{n}\ell_n(\widehat{\theta}_n) \ge \frac{n-1}{n}\log\frac{M_n}{2} + \frac{1}{n}\log\frac{1 - M_n}{c(M_n)}$$
et en déduire que, presque sûrement,
$$\liminf_{n\to\infty} \frac{1}{n}\ell_n(\widehat{\theta}_n) \ge \log\frac{1}{2} + \liminf_{n\to\infty} \frac{1}{n}\log\frac{1 - M_n}{c(M_n)}.$$

2. D'après l'article de Thomas S. Ferguson, *An Inconsistent Maximum Likelihood Estimate*, Journal of the American Statistical Association, 1982.

5. Montrer que pour tout $\theta \in [0,1]$, il existe $x_\theta \in [0,1[$ tel que pour tout $x \in [x_\theta, 1]$ on ait $\mathbb{P}_\theta(X > x) \geq \mathbb{P}_0(X > x)$.
6. Soient $0 \leq \theta \leq 1$ et $0 < \varepsilon < 1$ fixés. En déduire que pour n assez grand

$$\mathbb{P}_\theta\left(n^{1/4}\left(1 - M_n\right) \geq \varepsilon\right) \leq \mathbb{P}_0\left(n^{1/4}\left(1 - M_n\right) \geq \varepsilon\right).$$

7. Montrer que pour tout $\varepsilon \in]0,1[$

$$\sum_{n=1}^{\infty} \mathbb{P}_0\left(n^{1/4}\left(1 - M_n\right) \geq \varepsilon\right) \leq \sum_{n=1}^{\infty} e^{-\frac{1}{2}\varepsilon^2\sqrt{n}}.$$

   Qu'en déduisez-vous quant à la convergence de $\left(n^{1/4}\left(1 - M_n\right)\right)$ sous P_0 ? Et sous $P_{\theta^\star}$?
8. Considérons désormais $c(\theta) = (1-\theta)e^{1-(1-\theta)^{-4}}$. Montrer que pour tout $\theta^\star \in [0,1]$, sous $P_{\theta^\star}$,

$$\frac{1}{n}\log\frac{1 - M_n}{c(M_n)} \xrightarrow[n\to\infty]{p.s.} \infty.$$

9. Conclure et interpréter.