

STATISTIQUE MATHÉMATIQUE – PARTIEL

M1 Mathématiques Fondamentales, Université Paris-Saclay
2025-2026

Lundi 16 février 2026
9h – 12h

Une feuille A4 recto manuscrite est autorisée en tant que support. La calculatrice est autorisée.

Avant de commencer :

- *Le sujet comporte une question de cours, un exercice et un problème. Il est demandé de les traiter dans l'ordre sur la copie.*
- *Des résultats pourront être admis d'une question à l'autre, à la condition de l'écrire clairement.*
- *Une attention particulière sera portée aux questions d'interprétation, à la rigueur et à la précision de la rédaction.*
- *Un barème indicatif est donné :*
 - Question de cours : 2 points,*
 - Exercice : 5 points,*
 - Problème : 13 points (et plus).*

Bon courage et amusez-vous !

Une question de cours

On se place dans un modèle paramétrique où $\Theta \subset \mathbb{R}^p$. Soit $T(X)$ un estimateur de θ tel que $\text{Var}_\theta(T(X))$ est finie pour tout $\theta \in \Theta$. Démontrer proprement la décomposition biais-variance du risque quadratique de $T(X)$. Il sera possible (mais non obligatoire) d'accompagner la démonstration d'un dessin.

Exercice – Une Rao-Blackwellisation

On considère le modèle où $X_1, \dots, X_n$ sont i.i.d. de loi¹ $\text{Poi}(\lambda)$, avec $\lambda > 0$ inconnu. 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 ? *Solution.* On a, pour tout $\lambda > 0$, $\mathbb{E}_\lambda[T(X)] = \mathbb{P}_\lambda(X_1 = 0) = \varphi(\lambda)$, $T(X)$ est donc sans biais. On a $\text{Var}_\lambda(T(X)) = \mathbb{E}_\lambda[T(X)^2] - \mathbb{E}_\lambda[T(X)]^2 = \varphi(\lambda)(1 - \varphi(\lambda)) = e^{-\lambda}(1 - e^{-\lambda}) > 0$, qui est d'ordre 1 et ne tend pas vers 0 avec n (logique, on a pris qu'une des n observations pour construire T). Ce n'est donc pas un bon estimateur.
2. Prouver que $S(X) = \sum_{i=1}^n X_i$ est une statistique complète et suffisante pour le modèle. On pourra utiliser sans démonstration que si $Y_1 \sim \text{Poi}(\lambda_1)$ et $Y_2 \sim \text{Poi}(\lambda_2)$ sont indépendants, alors $Y_1 + Y_2 \sim \text{Poi}(\lambda_1 + \lambda_2)$. *Solution.* La densité de X (par rapport à $\mu^{\otimes n}$ où μ est la mesure de comptage sur $\mathbb{N}$) s'écrit, pour tout $k \in \mathbb{N}^n$,

$$p_X(k) = e^{-\lambda n} \prod_{i=1}^n \frac{\lambda^{k_i}}{(k_i)!} = \prod_{i=1}^n \underbrace{\frac{1}{(k_i)!}}_{=:h(k)} \underbrace{e^{-\lambda n + S(k) \log \lambda}}_{=:g_\lambda(S(k))}.$$

$S(X)$ est donc suffisante par le théorème de Neyman-Fisher. Montrons qu'elle est complète. Soit $f : \mathbb{N} \rightarrow \mathbb{R}$ telle que pour tout $\lambda > 0$, $\mathbb{E}_\lambda[f(S(X))] = 0$. D'après l'indication, on a que sous $\mathbb{P}_\lambda$, $S(X) \sim \text{Poi}(\lambda n)$, donc ce qui précède équivaut à

$$\forall \lambda > 0, e^{-\lambda n} \sum_{k \geq 0} f(k) \frac{(\lambda n)^k}{k!} = 0$$

et par unicité des coefficients de la série entière

$$z \mapsto \sum_{k \geq 0} f(k) \frac{z^k}{k!},$$

de rayon infini, on a pour tout $k \geq 0$, $\frac{f(k)}{k!} = 0$, c'est-à-dire $f(S(X)) = 0$, $\mathcal{M}$ -p.s. $S(X)$ est complète.

3. Calculer l'estimateur Rao-Blackwellisé T^* de T par S , et montrer que T^* est UVMB² pour l'estimation de $\varphi(\lambda)$. *Solution.* Calculons l'espérance conditionnelle de T sachant S . Celle-ci vaut $\mathbb{E}_\lambda[T(X) | S(X)] = \mathbb{P}_\lambda(X_1 = 0 | S(X))$. Fixons $s \in \mathbb{N}$ et $\lambda > 0$. On a

$$\begin{aligned} \mathbb{P}_\lambda(X_1 = 0 | S(X) = s) &= \frac{\mathbb{P}_\lambda(X_1 = 0, S(X) = s)}{\mathbb{P}_\lambda(S(X) = s)} \\ &= \frac{\mathbb{P}_\lambda(X_1 = 0, X_2 + \dots + X_n = s)}{\mathbb{P}_\lambda(S(X) = s)} \end{aligned}$$

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

²UVMB : sans biais et uniformément de variance minimale.

$$\begin{aligned}
&= \frac{e^{-\lambda} e^{-\lambda(n-1)} \frac{(\lambda(n-1))^s}{s!}}{e^{-\lambda n} \frac{(\lambda n)^s}{s!}} \\
&= (1 - 1/n)^s.
\end{aligned}$$

On a donc $T^*(X) = (1 - 1/n)^{S(X)}$. T^* étant le Rao-Blackwellisé de T par S qui est sans biais, complète et suffisante, on sait qu'il est UVMB par le théorème de Lehman-Scheffé.

Problème – Estimation dans un modèle de translation

Dans tout le problème, $\|\cdot\|$ désigne la norme euclidienne canonique sur $\mathbb{R}^d$, et ∇ désigne le gradient sur $\mathbb{R}^d$. On considère le modèle où $X = (X_1, \dots, X_n)$ avec X_i vecteurs aléatoires i.i.d. à valeurs dans $\mathbb{R}^d$. Pour $\theta \in \Theta \subseteq \mathbb{R}^d$, sous $\mathbb{P}_\theta$, X_1 possède une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ donnée par

$$x \mapsto f(x - \theta),$$

où $f : \mathbb{R}^d \rightarrow \mathbb{R}_+^*$ est une densité de probabilité, strictement positive, supposée connue. On parle alors de *modèle de translation*. On suppose que le modèle est régulier, et que

(H1) f est de classe C^1 sur $\mathbb{R}^d$ et $\nabla \log f$ vérifie $\sup_{x \in \mathbb{R}^d} \|\nabla \log f(x)\| < +\infty$.

(H2) Θ est un compact de $\mathbb{R}^d$.

Partie I – Propriétés générales du modèle.

1. Montrer que le modèle est identifiable. *Solution.* Pour $\theta, \theta' \in \Theta$, $\mathbb{P}_\theta = \mathbb{P}_{\theta'}$ si et seulement si pour Lebesgue-presque tout $x \in \mathbb{R}^d$, $f(x - \theta) = f(x - \theta')$, c'est-à-dire que f est égale Lebesgue-presque partout à une fonction $\theta - \theta'$ périodique. Or une fonction périodique non nulle ne peut pas être intégrable. Montrons-le ici proprement. Supposons que $T_0 := \theta - \theta' \neq 0$. L'idée est de trouver des boules sur lesquelles l'intégrale de f est la même, et strictement positive. Notons $I = \int f(x) \mathbb{1}_{\|x\| \leq T_0} dx > 0$ et pour tout $k \geq 0$, $B_k = B(2kT_0, T_0)$, boule ouverte de centre $2kT_0$ et de rayon T_0 . Il est aisé de voir que les B_k sont 2 à 2 disjoints (on pourra faire un dessin), et que pour tout $k \geq 0$, $\int f(x) \mathbb{1}_{B_k}(x) dx \stackrel{u=x-2kT_0}{=} \int f(u) \mathbb{1}_{B_0}(u) du = I$. Donc $\int f(x) dx \geq \sum_{k \geq 0} \int f(x) \mathbb{1}_{B_k}(x) dx = \sum_{k \geq 0} I = +\infty$, ce qui contredit l'intégrabilité de f . Donc $T_0 = 0$, et le modèle est identifiable.

2. Montrer que l'information de Fisher s'écrit

$$\forall \theta \in \Theta, I(\theta) = n \int_{\mathbb{R}^d} \frac{(\nabla f(x))(\nabla f(x))^\top}{f(x)} dx.$$

Solution. Le modèle est régulier et produit. La log-vraisemblance du modèle à un échantillon s'écrit ici

$$\ell(\theta, x_1) = \log f(x_1 - \theta),$$

et le score vaut

$$\dot{\ell}(\theta, x_1) = \frac{\nabla f(x_1 - \theta)}{f(x_1 - \theta)},$$

il est centré et de matrice de covariance

$$\begin{aligned}
I(\theta) &= \text{Var}_\theta(\dot{\ell}(\theta, X_1)) \\
&= \mathbb{E}_\theta[\dot{\ell}(\theta, X_1) \dot{\ell}(\theta, X_1)^\top]
\end{aligned}$$

$$\begin{aligned}
&= \int \frac{\nabla f(x - \theta)}{f(x - \theta)} \left(\frac{\nabla f(x - \theta)}{f(x - \theta)} \right)^\top f(x - \theta) dx \\
&= \int \frac{(\nabla f(u))(\nabla f(u))^\top}{f(u)} du,
\end{aligned}$$

où la dernière ligne est obtenue par le changement de variable $u = x - \theta$. On conclut par additivité de l'information de Fisher.

3. Comment interprétez-vous le fait que $I(\theta)$ ne dépende pas de θ ? *Solution.* La difficulté d'estimation de θ est la même partout, quelque soit θ . Cela vient du fait que le modèle est un modèle de translation, i.e., si $X \sim \mathbb{P}_\theta$, alors $X - \theta \sim \mathbb{P}_0$ (on peut le voir avec la méthode de la fonction muette).
4. Montrer qu'il existe un estimateur du maximum de vraisemblance (EMV, pas forcément unique), et que tout EMV $\hat{\theta}_n(X)$ vérifie

$$\hat{\theta}_n(X) \in \arg \max_{t \in \Theta} M_n(X, t),$$

avec

$$M_n(X, t) = \frac{1}{n} \sum_{i=1}^n \log f(X_i - t).$$

Solution. La log-vraisemblance s'écrit $\ell(\theta, x) = \sum_{i=1}^n \log f(x_i - \theta) = nM_n(x, \theta)$. pour tout vecteur $x \in \mathbb{R}^n$, $t \rightarrow M_n(x, t)$ est continue sur un compact : elle est bornée et atteint ses bornes. L'argmax est donc bien défini (mais pas forcément unique, d'où le fait qu'il existe au moins un EMV).

Pour $t \in \Theta$, on note $M(t) = \mathbb{E}_\theta[\log f(X_1 - t)]$.

5. Montrer que pour tout $t \in \Theta$, $M_n(X, t) \xrightarrow[n \rightarrow \infty]{p.s.} M(t)$. *Solution.* Les X_i sont i.i.d., et c'est la loi des grands nombres.
6. Montrer que M admet un maximum global atteint uniquement en $t = \theta$. On pourra montrer que pour tout $t \in \Theta$,

$$\int_{\mathbb{R}^d} f(x - \theta) \log \left(\frac{f(x - \theta)}{f(x - t)} \right) dx \geq 0,$$

et étudier le cas d'égalité. *Solution.* On note

$$\Delta(t) = \int f(x - \theta) \log \left(\frac{f(x - \theta)}{f(x - t)} \right) dx.$$

On a déjà fait des choses similaires en TD. Notons que cette intégrale a toujours un sens dans $\mathbb{R} \cup \{+\infty\}$: en effet, sa partie négative est toujours intégrable, car $p \log p/q$ est toujours minorée par $-\frac{1}{e}q$, comme $u \log u \geq -1/e$ sur $\mathbb{R}_+^*$. Ce minorant est intégrable, donc $\Delta(t)$ a bien une valeur dans $\mathbb{R} \cup \{+\infty\}$. Ensuite, on remarque que $\Delta(t) = \mathbb{E}_t[Z \log Z]$, avec $Z = \frac{f(X - \theta)}{f(X - t)}$. La fonction $u \rightarrow u \log u$ étant strictement convexe sur $\mathbb{R}_+^*$, d'après l'inégalité de Jensen, on a

$$\Delta(t) \geq \mathbb{E}_t[Z] \log \mathbb{E}_t[Z] = 0,$$

avec égalité si et seulement si Z est constante p.s., i.e. $f(x - \theta) = C f(x - t)$ Lebesgue presque partout, ce qui implique $C = 1$. Puis, par identifiabilité du modèle, cela équivaut

à $t = \theta$. M vérifie $M(t) = -\Delta(t) + \int_{\mathbb{R}^d} f(x-\theta) \log(f(x-\theta)) dx$ admet donc un unique maximum global atteint uniquement en $t = \theta$.

7. D'après ce qui précède, pourquoi, intuitivement, $\hat{\theta}_n$ a-t-il une chance d'être proche de θ lorsque n est grand ? On attend un argument heuristique (non rigoureux). *Solution.* Quand n est grand, comme $M(X, t)$ est proche de $M(t)$, maximiser $M(X, t)$ revient heuristiquement à maximiser $M(t)$, qui n'atteint son maximum qu'en θ . Donc cela nous amène à penser que $\hat{\theta}_n$ est proche de θ . Reste à montrer que la limite presque sûre et l'argmax commutent, ce qui n'est pas évident à priori.

Partie II – Un résultat d'analyse

Pour rendre rigoureuse l'intuition de la question 7, il nous faut un résultat d'analyse dont cette partie fait l'objet.

Soit Γ un compact de $\mathbb{R}^d$. Soit $(g_n)_{n \geq 0}$ une suite de fonctions continues de Γ dans $\mathbb{R}$, et $g : \Gamma \rightarrow \mathbb{R}$. On suppose que g_n converge uniformément vers g , que g a un maximum global atteint en un unique $x^* \in \Gamma$, et que pour tout n , il existe $x_n \in \Gamma$ tel que $g_n(x_n) = \sup_{x \in \Gamma} g_n(x)$. L'objectif est de démontrer que la suite $(x_n)_{n \geq 0}$ converge vers x^* .

Fixons $\varepsilon > 0$. On considère $\Gamma_\varepsilon := \{x \in \Gamma, \|x - x^*\| \geq \varepsilon\}$, et

$$\delta := g(x^*) - \sup_{x \in \Gamma_\varepsilon} g(x).$$

8. Justifier que $\delta > 0$. *Solution.* Notons que $\delta \geq 0$ par définition. Montrons qu'il est strictement positif. g est limite uniforme de fonctions continues, elle est donc continue. Comme Γ est compact, Γ_ε est encore compact en tant que fermé de Γ (intersection de Γ et d'un fermé). Le supremum est donc un maximum : il existe $y^* \in \Gamma_\varepsilon$ tel que $g(y^*) = \sup_{x \in \Gamma_\varepsilon} g(x)$. Si $\delta = 0$, alors y^* atteint le maximum global et $y^* = x^*$, absurde car $y^* \in \Gamma_\varepsilon$.

9. Montrer qu'il existe $N \geq 0$ tel que pour tout $n \geq N$, pour tout $x \in \Gamma_\varepsilon$,

$$g(x^*) - \frac{1}{3}\delta \leq g_n(x^*) \quad \text{et} \quad g_n(x) \leq g(x^*) - \frac{2}{3}\delta,$$

puis que $g_n(x) < g_n(x^*)$. *Solution.* Par convergence uniforme, il existe $N \geq 0$ tel que pour tout $n \geq N$, $x \in \Gamma$,

$$|g_n(x) - g(x)| \leq \delta/3.$$

En appliquant cette inégalité à x^* , on a exactement la première inégalité voulue. Pour la seconde, notons que pour $n \geq N$ et $x \in \Gamma_\varepsilon$, $g_n(x) \leq g(x) + \delta/3 \leq \sup_{x \in \Gamma_\varepsilon} g(x) + \delta/3 = g(x^*) - \delta + \delta/3 = g(x^*) - 2\delta/3$. Comme $\delta > 0$, on a donc $g_n(x) \leq g(x^*) - 2\delta/3 < g(x^*) - \delta/3 \leq g_n(x^*)$.

10. En déduire que pour $n \geq N$, $x_n \notin \Gamma_\varepsilon$. Conclure. *Solution.* Comme $g_n(x_n) \geq g_n(x^*)$ par définition, d'après la question précédente, pour $n \geq N$, x_n ne peut pas être dans Γ_ε . On a montré que pour $n \geq N$, $\|x_n - x^*\| < \varepsilon$. Cela achève la preuve de la convergence $x_n \rightarrow x^*$.

Partie III – Une loi des grands nombres uniforme

Le but de cette partie est de montrer qu'on a la convergence uniforme presque sûre :

$$\sup_{t \in \Theta} |M_n(X, t) - M(t)| \xrightarrow[n \rightarrow \infty]{p.s.} 0.$$

Notons que ce résultat est plus fort que celui établi en question 5.

Fixons $\varepsilon > 0$.

11. En utilisant (H1), montrer qu'il existe $\eta > 0$ tel que pour tout $\theta, \theta' \in \Theta$,

$$\|\theta - \theta'\| \leq \eta \implies \sup_{x \in \mathbb{R}^d} |\log f(x - \theta) - \log f(x - \theta')| \leq \varepsilon.$$

Solution. Notons $C := \sup_{x \in \mathbb{R}^d} \|\nabla \log f(x)\|$ dans (H1). Si $C = 0$, f est constante sur $\mathbb{R}^d$, strictement positive et intégrable, absurde. Donc $C > 0$. D'après l'inégalité des accroissements finis (valable bien sûr aussi dans $\mathbb{R}^d$), on a que pour tout $x \in \mathbb{R}^d$,

$$|\log f(x - \theta) - \log f(x - \theta')| \leq C \times \|x - \theta - (x - \theta')\| = C \times \|\theta - \theta'\|,$$

et $\eta = \varepsilon/C$ convient.

12. Montrer qu'il existe $m \geq 1$, $t_1, \dots, t_m \in \Theta$ tels que pour tout $t \in \Theta$, il existe $1 \leq j \leq m$ tel que $\|t - t_j\| \leq \eta$. *Solution.* Notons que

$$\Theta \subseteq \bigcup_{t \in \Gamma} B(t, \eta),$$

et que comme Θ est compact, on peut extraire de ce recouvrement un recouvrement fini. Il existe donc $t_1, \dots, t_m \in \Theta$ tels que

$$\Theta \subseteq \bigcup_{j=1}^m B(t_j, \eta),$$

ce qui est exactement ce que l'on veut.

13. Soit $t \in \Theta$. Montrer qu'il existe un entier $1 \leq j \leq m$ tel que :

$$|M_n(X, t) - M(t)| \leq |M_n(X, t_j) - M(t_j)| + 2\varepsilon.$$

Solution. On fait apparaître les quantités qu'on veut. Il existe $1 \leq j \leq m$ tel que $\|t_j - t\| \leq \eta$. D'après 11, cela implique que presque sûrement (et même, partout), pour tout $1 \leq i \leq d$, $|\log f(X_i - t) - \log f(X_i - t_j)| \leq \varepsilon$ et partant $|M_n(X, t) - M_n(X, t_j)| \leq \varepsilon$. Il vient alors,

$$\begin{aligned} |M_n(X, t) - M(t)| &\leq |M_n(X, t) - M_n(X, t_j)| + |M_n(X, t_j) - M(t_j)| + |M(t_j) - M(t)| \\ &\leq \varepsilon + |M_n(X, t_j) - M(t_j)| + \mathbb{E}_\theta[|M_n(X, t) - M_n(X, t_j)|] \\ &\leq |M_n(X, t_j) - M(t_j)| + 2\varepsilon. \end{aligned}$$

14. Conclure. *Solution.* La loi des grands nombres donne que pour tout $1 \leq j \leq m$, $\mathbb{P}_\theta$ -p.s., $|M_n(X, t_j) - M(t_j)| \rightarrow 0$. Mais les t_j sont en nombre fini, et comme réunion finie d'ensembles négligeable est négligeable, on a que $\mathbb{P}_\theta$ -p.s.,

$$\sup_{1 \leq j \leq m} |M_n(X, t_j) - M(t_j)| \rightarrow 0.$$

En prenant le supremum dans l'inégalité de la question précédente, nous avons donc que pour tout $\varepsilon > 0$, $\mathbb{P}_\theta$ -p.s.,

$$\limsup_n \sup_{t \in \Theta} |M_n(X, t) - M(t)| \leq 2\varepsilon.$$

Enfin, on séquentialise les $\varepsilon > 0$ par exemple en posant $\varepsilon_m = 2^{-m}$ pour obtenir que $\mathbb{P}_\theta$ -p.s.,

$$\sup_{t \in \Theta} |M_n(X, t) - M(t)| \rightarrow 0.$$

Partie IV – Conclusion : consistance de tout estimateur du maximum de vraisemblance

15. En utilisant les résultats des parties II et III, montrer que tout EMV $\hat{\theta}_n$ est fortement consistant. *Solution.* D'après la partie III, $\mathbb{P}_\theta$ -p.s., $M_n(X, \cdot)$ converge uniformément vers $M(\cdot)$, qui possède un maximum global atteint uniquement en θ , d'après 6. D'après 4, $\hat{\theta}_n$ vérifie $M_n(X, \hat{\theta}_n) = \sup_{t \in \Theta} M_n(X, t)$. Θ est compact, et les $M_n(X, \cdot)$ sont continues par régularité du modèle : on est exactement dans les hypothèses de la partie II. On conclut donc que $\mathbb{P}_\theta$ -p.s., $\hat{\theta}_n$ converge vers θ . C'est exactement la consistance forte de l'estimateur du maximum de vraisemblance. La limite presque sûre et l'argmax commutent !

Fin du sujet.