Sorbonne Université, Master 1 | **MU4MA015, Statistique, 2023-2024**
Cours : A. Ben-Hamou | TD : A. Godichon, A. Guyader et M. Sangnier

Correction du TD 10

Exercice 1 (Borne de Le Cam)

1. (a) L'EMV est évidemment $\overline{X}_n$ et son risque quadratique vaut, pour tout $\theta \in \mathbb{R}$:

$$\mathbf{R}(\theta, \overline{X}_n) = \mathbb{E}[(\overline{X}_n - \theta)^2] = \mathrm{Var}(\overline{X}_n) = \frac{1}{n}.$$

(b) Soient $P_\theta \in \mathcal{P}$ et $P_{\theta'} \in \mathcal{P}$. On note p_θ et $p_{\theta'}$ leurs densités respectives. On rappelle que $h^2(P_\theta, P_{\theta'}) = 2\,(1 - \rho(P_\theta, P_{\theta'}))$, où ρ est l'affinité de Hellinger :

$$\begin{aligned}
\rho(P_\theta, P_{\theta'}) &= \int_{\mathbb{R}} \sqrt{p_\theta(x) p_{\theta'}(x)}\,\mathrm{d}x \\
&= \int_{\mathbb{R}} \left(\frac{e^{-\frac{1}{2}(x-\theta)^2}}{\sqrt{2\pi}} \frac{e^{-\frac{1}{2}(x-\theta')^2}}{\sqrt{2\pi}} \right)^{1/2} \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}x^2 + \frac{\theta+\theta'}{2}x - \frac{\theta^2+\theta'^2}{4}}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta^2+\theta'^2}{4}} \int_{\mathbb{R}} e^{-\frac{1}{2}\left(x - \frac{\theta+\theta'}{2}\right)^2 + \frac{(\theta+\theta')^2}{8}}\,\mathrm{d}x \\
&= \mathrm{e}^{-\frac{(\theta-\theta')^2}{8}}.
\end{aligned}$$

Ainsi

$$h(P_\theta, P_{\theta'}) = \sqrt{2 - 2\,\mathrm{e}^{-\frac{(\theta-\theta')^2}{8}}}.$$

Lorsque $\theta' = \theta + \varepsilon$ et $\varepsilon \to 0$,

$$h^2(P_\theta, P_{\theta+\varepsilon}) = 2 - 2\,\mathrm{e}^{-\frac{\varepsilon^2}{8}} = \frac{\varepsilon^2}{4}(1 + o(1)).$$

On a alors :

$$h(P_\theta, P_{\theta+\varepsilon}) \sim \frac{|\varepsilon|}{2}.$$

Remarque : ce calcul est cohérent avec le résultat $h(P_\theta, P_{\theta+\varepsilon}) \sim \frac{\sqrt{I(\theta)}}{2}\varepsilon$ cité en cours, car ici $I(\theta) = 1$ pour tout $\theta \in \mathbb{R}$.

(c) Pour $\theta \in \mathbb{R}$, le risque minimax local en θ est défini par

$$\mathbf{R}_M^\theta = \sup_{\theta' \in \mathbb{R}} \inf_T \max\left\{ \mathbf{R}(\theta, T), \mathbf{R}(\theta', T) \right\}.$$

Par ailleurs, dans le cas de la perte quadratique, le théorème de Le Cam assure que pour tout $\theta \in \mathbb{R}$:

$$\mathbf{R}_M^\theta \geq \sup_{\theta' \in \mathbb{R}} \frac{(\theta - \theta')^2}{4} \left(1 - \sqrt{n} h(P_\theta, P_{\theta'})\right) = \sup_{\varepsilon \in \mathbb{R}} \frac{\varepsilon^2}{4} \left(1 - \sqrt{n} h(P_\theta, P_{\theta+\varepsilon})\right).$$

Il apparaît que le terme à maximiser peut prendre des valeurs positives et négatives (cela arrive lorsque $h(P_\theta, P_{\theta+\varepsilon})$ est trop grand). Le supremum est donc forcément positif. Or, pour tout $\varepsilon \in \mathbb{R}$:

$$\begin{aligned}\frac{\varepsilon^2}{4}\left(1-\sqrt{n}h(P_\theta,P_{\theta+\varepsilon})\right)\geq 0 &\iff h^2(P_\theta,P_{\theta+\varepsilon})\leq \frac{1}{n}\\ &\iff \mathrm{e}^{-\frac{\varepsilon^2}{8}}\geq 1-\frac{1}{2n}\\ &\iff \varepsilon^2\leq -8\log\left(1-\frac{1}{2n}\right)\\ &\iff \varepsilon^2\leq 8\log\left(1+\frac{1}{2n-1}\right).\end{aligned}$$

Ainsi, la fonction à maximiser est positive uniquement pour $|\varepsilon| \leq 2\sqrt{2}\{\log(1+\frac{1}{2n-1})\}^{\frac{1}{2}}$, et atteint donc son maximum pour ε proche de 0 puisque $2\sqrt{2}\{\log(1+\frac{1}{2n-1})\}^{\frac{1}{2}} \xrightarrow[n\to\infty]{} 0$.

(d) La fonction à maximiser dans la borne de Le Cam est :

$$\frac{\varepsilon^2}{4}\left(1-\sqrt{n}h(P_\theta,P_{\theta+\varepsilon})\right)\sim\frac{\varepsilon^2}{4}\left(1-\sqrt{n}\frac{|\varepsilon|}{2}\right),$$

lorsque $\varepsilon \to 0$. Le terme de droite étant maximal pour $\varepsilon_0 = \pm\frac{4}{3\sqrt{n}}$, on obtient une borne de Le Cam de l'ordre de :

$$\mathbf{R}_M^\theta \geq \sup_{\varepsilon\in\mathbb{R}}\frac{\varepsilon^2}{4}\left(1-\sqrt{n}h(P_\theta,P_{\theta+\varepsilon})\right)\geq\frac{\varepsilon_0^2}{4}\left(1-\sqrt{n}h(P_\theta,P_{\theta+\varepsilon_0})\right)\sim\frac{\varepsilon_0^2}{4}\left(1-\sqrt{n}\frac{\varepsilon_0}{2}\right)=\frac{4}{27n}\geq\frac{1}{8n},$$

la dernière minoration étant obtenue pour $\varepsilon = \frac{1}{\sqrt{n}}$. Par ailleurs, on sait que le risque minimax $\mathbf{R}_M$ est minoré par le risque minimax local, donc, en prenant comme référence l'EMV, pour tout $\theta \in \mathbb{R}$:

$$\frac{1}{n}=\mathbf{R}_{\max}(\overline{X}_n)\geq\mathbf{R}_M\geq\sup_{\theta\in\mathbb{R}}\mathbf{R}_M^\theta\geq\frac{\varepsilon_0^2}{4}\left(1-\sqrt{n}h(P_\theta,P_{\theta+\varepsilon_0})\right)\sim\frac{4}{27n}\geq\frac{1}{8n}.$$

Le risque minimax $\mathbf{R}_M$ est donc encadré par deux suites qui évoluent en $\frac{1}{n}$. On en conclut que l'ordre du risque minimax $\mathbf{R}_M$ est $\frac{1}{n}$ et que la borne de Le Cam est fine, puisqu'elle atteint l'ordre de grandeur de la borne supérieure, et enfin que l'EMV a la vitesse optimale en terme de risque minimax puisqu'elle est en $\frac{1}{n^2}$.

2. (a) Pour le modèle des lois exponentielles translatées, l'EMV est $X_{(1)}$ et nous avons vu lors d'un précédent TD que $n(X_{(1)}-\theta) \sim \mathcal{E}(1)$, donc pour tout $\theta \in \mathbb{R}$:

$$\mathbf{R}(\theta, X_{(1)}) = \mathbb{E}_\theta\left[(X_{(1)}-\theta)^2\right]=\frac{1}{n^2}\mathbb{E}[\mathcal{E}(1)^2]=\frac{2}{n^2}.$$

(b) Pour qu'un modèle $\mathcal{P} = \{P_\theta,\, \theta \in \Theta\}$ avec $\mathrm{d}P_\theta = p_\theta\,\mathrm{d}\mu$ soit régulier, il faut au moins que pour μ-presque tout $x \in E$, la fonction $\theta \mapsto p_\theta(x)$ soit continue sur Θ. Or ici, pour tout $x \in \mathbb{R}$, la fonction $\theta \mapsto \mathrm{e}^{-(x-\theta)}\mathbb{1}_{x\geq\theta}$ présente un saut en $\theta = x$. Ce modèle n'est donc pas régulier. Il se trouve que cette irrégularité se traduit dans le comportement de la distance de Hellinger. Supposons que $\theta \leq \theta'$ (par symétrie de la distance de Hellinger, l'autre cas s'obtient en échangeant les rôles de θ et θ'). On a

$$\rho(P_\theta,P_{\theta'})=\int_{\mathbb{R}}\left(\mathrm{e}^{\theta-x}\,\mathrm{e}^{\theta'-x}\right)^{1/2}\mathbb{1}_{[\theta',\infty)}(x)\,\mathrm{d}x=\mathrm{e}^{\frac{\theta+\theta'}{2}}\int_{\theta'}^{\infty}\mathrm{e}^{-x}\,\mathrm{d}x=\mathrm{e}^{-\frac{\theta'-\theta}{2}}.$$

On en déduit que

$$h(P_\theta, P_{\theta'}) = \sqrt{2 - 2\,\mathrm{e}^{-\frac{|\theta-\theta'|}{2}}}.$$

Lorsque $\theta' = \theta + \varepsilon$ et $\varepsilon \to 0$,

$$h^2(P_\theta, P_{\theta+\varepsilon}) = 2 - 2\,\mathrm{e}^{-\frac{|\varepsilon|}{2}} = |\varepsilon|(1 + o(1)).$$

On a alors :

$$h(P_\theta, P_{\theta+\varepsilon}) \sim \sqrt{|\varepsilon|}.$$

(c) De manière identique à la question précédente, pour tout $\theta \in \mathbb{R}$:

$$\mathbf{R}_M^\theta \geq \sup_{\varepsilon\in\mathbb{R}} \frac{\varepsilon^2}{4}\left(1 - \sqrt{n}h(P_\theta, P_{\theta+\varepsilon})\right)$$

et pour tout $\varepsilon \in \mathbb{R}$:

$$\begin{aligned}
\frac{\varepsilon^2}{4}\left(1 - \sqrt{n}h(P_\theta, P_{\theta+\varepsilon})\right) \geq 0 &\iff h^2(P_\theta, P_{\theta+\varepsilon}) \leq \frac{1}{n} \\
&\iff \mathrm{e}^{-\frac{|\varepsilon|}{2}} \geq 1 - \frac{1}{2n} \\
&\iff |\varepsilon| \leq 2\log\left(1 + \frac{1}{2n-1}\right).
\end{aligned}$$

Ainsi, la fonction à maximiser est positive uniquement pour $|\varepsilon| \leq 2\log\left(1+\frac{1}{2n-1}\right)$, et atteint donc son maximum proche de 0 au sens où $2\log\left(1+\frac{1}{2n-1}\right) \xrightarrow[n\to\infty]{} 0$.

(d) La fonction à maximiser dans la borne de Le Cam est :

$$\frac{\varepsilon^2}{4}\left(1 - \sqrt{n}h(P_\theta, P_{\theta+\varepsilon})\right) \sim \frac{\varepsilon^2}{4}\left(1 - \sqrt{n|\varepsilon|}\right),$$

lorsque $\varepsilon \to 0$. Le terme de droite étant maximal pour $\varepsilon_0 = \pm\frac{16}{25n}$, on obtient une borne de Le Cam de l'ordre de :

$$\mathbf{R}_M^\theta \geq \sup_{\varepsilon\in\mathbb{R}} \frac{\varepsilon^2}{4}\left(1 - \sqrt{n}h(P_\theta, P_{\theta+\varepsilon})\right) \geq \frac{\varepsilon_0^2}{4}\left(1 - \sqrt{n}h(P_\theta, P_{\theta+\varepsilon_0})\right) \sim \frac{\varepsilon_0^2}{4}\left(1 - \sqrt{n|\varepsilon_0|}\right),$$

avec

$$\frac{\varepsilon_0^2}{4}\left(1 - \sqrt{n|\varepsilon_0|}\right) = \frac{64}{3125n^2} \geq \frac{1}{128n^2},$$

la dernière minoration étant obtenue pour $\varepsilon = \frac{1}{4n}$. Par ailleurs, pour tout $\theta \in \mathbb{R}$:

$$\frac{2}{n^2} = \mathbf{R}_{\max}(X_{(1)}) \geq \mathbf{R}_M \geq \sup_{\theta\in\mathbb{R}} \mathbf{R}_M^\theta \geq \frac{\varepsilon_0^2}{4}\left(1 - \sqrt{n}h(P_\theta, P_{\theta+\varepsilon_0})\right) \sim \frac{64}{3125n^2} \geq \frac{1}{128n^2}.$$

Le risque minimax $\mathbf{R}_M$ est donc encadré par deux suites qui évoluent en $\frac{1}{n^2}$. On en conclut que l'ordre du risque minimax $\mathbf{R}_M$ est $\frac{1}{n^2}$, que la borne de Le Cam est fine, puisqu'elle atteint l'ordre de grandeur de la borne supérieure, et enfin que l'EMV a la vitesse optimale en terme de risque minimax puisqu'elle est en $\frac{1}{n^2}$.

Remarque : sur cet exemple, si au lieu de l'EMV $X_{(1)}$, on avait considéré $\widehat{\theta}_n := \overline{X}_n - 1$ issu de la méthode des moments (puisque $\mathbb{E}[X_1] = 1 + \theta$), tout ce qui précède s'appliquerait encore et la suite d'inégalités ci-dessus deviendrait

$$\frac{1}{n} = \mathbf{R}_{\max}(\widehat{\theta}_n) \geq \mathbf{R}_M \geq \sup_{\theta\in\mathbb{R}} \mathbf{R}_M^\theta \geq \frac{\varepsilon_0^2}{4}\left(1 - \sqrt{n}h(P_\theta, P_{\theta+\varepsilon_0})\right) \sim \frac{64}{3125n^2} \geq \frac{1}{128n^2}.$$

On en déduirait que le risque minimax a une vitesse entre $\frac{1}{n^2}$ et $\frac{1}{n}$ mais rien de plus ! On ne saurait conclure ni sur la précision de la borne de Le Cam, ni sur l'optimalité ou non de $\widehat{\theta}_n$ en terme de risque minimax. Bref, ce serait nettement moins pertinent.

Exercice 2 (Test bayésien I)
Notons qu'une mesure dominante pour la loi a priori est $\nu = \delta_{1/2} + \lambda_{[0,1]}$, où $\lambda_{[0,1]}$ désigne la mesure de Lebesgue sur $[0,1]$. La densité de la loi a priori Π par rapport à cette mesure dominante est

$$\pi(\theta) = \alpha \mathbb{1}_{\theta=1/2} + (1-\alpha)\mathbb{1}_{[0,1/2[\cup]1/2,1]}(\theta).$$

Rappelons que ceci signifie que pour toute fonction borélienne bornée $\varphi : \mathbb{R} \to \mathbb{R}$, l'intégrale de φ par rapport à Π s'écrit

$$\int_{\mathbb{R}} \varphi(\theta)\Pi(d\theta) = \int_{\mathbb{R}} \varphi(\theta)\pi(\theta)\nu(d\theta) = \alpha\varphi(1/2) + (1-\alpha)\int \varphi(\theta)\mathbb{1}_{[0,1]}(\theta)(\theta)\,\mathrm{d}\theta.$$

Une mesure dominante pour les lois $(P_\theta)_{0\leq\theta\leq 1}$ est $\mu = (\delta_0 + \delta_1)^{\otimes n}$ et, sachant $\boldsymbol{\theta} = \theta$, la densité associée à une observation $\mathbf{X} = (X_1, \ldots, X_n) \in \{0,1\}^n$ (ou vraisemblance) est

$$p_\theta(\mathbf{X}) = \theta^{n\overline{X}_n}(1-\theta)^{n-n\overline{X}_n}.$$

Le test de Bayes s'écrit

$$\varphi(\mathbf{X}) = \mathbb{1}_{\{\Pi(\Theta_0|\mathbf{X})\leq\Pi(\Theta_1|\mathbf{X})\}} = \mathbb{1}_{\{\Pi(\Theta_0|\mathbf{X})\leq 1/2\}}\,.$$

On a

$$\Pi(\Theta_0|\mathbf{X}) = \Pi(\{1/2\}|\mathbf{X}) = \frac{\pi(1/2)p_{1/2}(\mathbf{X})}{f(\mathbf{X})}$$

avec

$$f(\mathbf{X}) = \int_{[0,1]} \pi(\theta)p_\theta(\mathbf{X})\,\mathrm{d}\nu(\theta) = \alpha p_{1/2}(\mathbf{X}) + (1-\alpha)\int_{[0,1]\setminus\{1/2\}} \pi(\theta)p_\theta(\mathbf{X})\,\mathrm{d}\theta,$$

c'est-à-dire

$$f(\mathbf{X}) = \alpha 2^{-n} + (1-\alpha)\int_0^1 \theta^{n\overline{X}_n}(1-\theta)^{n-n\overline{X}_n}\,\mathrm{d}\theta = \alpha 2^{-n} + (1-\alpha)B(n\overline{X}_n+1, n-n\overline{X}_n+1)\,.$$

Ainsi

$$\Pi(\Theta_0|\mathbf{X}) = \frac{\alpha 2^{-n}}{\alpha 2^{-n} + (1-\alpha)B(n\overline{X}_n+1, n-n\overline{X}_n+1)}\,.$$

En utilisant le fait que

$$B(n\overline{X}_n+1, n-n\overline{X}_n+1) = \frac{\Gamma(n\overline{X}_n+1)\Gamma(n-n\overline{X}_n+1)}{\Gamma(n+2)} = \frac{(n\overline{X}_n)!(n-n\overline{X}_n)!}{(n+1)!} = \frac{1}{(n+1)\binom{n}{n\overline{X}_n}}\,,$$

on obtient

$$\varphi(\mathbf{X}) = \mathbb{1}_{\left\{\alpha 2^{-n}\leq\frac{1-\alpha}{(n+1)\binom{n}{n\overline{X}_n}}\right\}} = \mathbb{1}_{\left\{\binom{n}{n\overline{X}_n}\leq\frac{(1-\alpha)2^n}{\alpha(n+1)}\right\}}.$$

Remarque : Ce résultat peut s'interpréter comme suit : si H_0 est vraie, alors $\overline{X}_n$ tend presque sûrement vers $\theta = 1/2$, or c'est lorsque $n\overline{X}_n$ est de l'ordre de $n/2$ que le coefficient binomial est le plus grand, donc qu'on a en effet le plus de chances d'accepter H_0 dans le test ci-dessus. En particulier, pour forcer le trait, si on avait exactement $n\overline{X}_n = n/2$, alors de la formule de Stirling [1] on déduirait que

$$\binom{n}{n\overline{X}_n} = \binom{n}{n/2} \sim \sqrt{\frac{2}{\pi n}}\, 2^n \gg \frac{(1-\alpha)2^n}{\alpha(n+1)}.$$

Bien entendu, sous H_0, on n'a pas exactement $n\overline{X}_n = n/2$, mais seulement $n\overline{X}_n$ de l'ordre de $n/2$ avec des fluctuations d'ordre $\sqrt{n}$ en raison du théorème central limite.

Exercice 3 (Test bayésien II)

1. Dans le cas des hypothèses H_0 et H_1, l'hypothèse nulle est ponctuelle, donc il est naturel de choisir une loi a priori Π qui met une masse strictement positive sur $\{0\}$, de la forme

$$\Pi = \alpha\delta_0 + (1-\alpha)\mathcal{N}(0,\sigma^2),$$

avec $0 < \alpha < 1$. Il s'agit d'une loi de mélange $\alpha\delta_0 + (1-\alpha)G$ avec $\delta_0(\Theta_0) = 1$ et $G(\Theta_1) = G(\mathbb{R}) = 1$. Une mesure dominante pour la loi a priori est $\nu = \delta_0 + \lambda_{\mathbb{R}}$, où $\lambda_{\mathbb{R}}$ désigne la mesure de Lebesgue sur $\mathbb{R}^*$. La densité de la loi a priori Π par rapport à cette mesure dominante est

$$\pi(\theta) = \alpha \mathbb{1}_{\theta=0} + \frac{1-\alpha}{\sqrt{2\pi}\sigma}\,\mathrm{e}^{-\frac{\theta^2}{2\sigma^2}} \mathbb{1}_{\theta\neq 0}.$$

Une mesure dominante pour les lois $(P_\theta)_{0\leq\theta\leq 1}$ est $\mu = \lambda^{\otimes n}$ et, sachant $\boldsymbol{\theta} = \theta$, la densité associée à une observation $\mathbf{X} = (X_1, \ldots, X_n) \in \mathbb{R}^n$ (ou vraisemblance) est

$$p_\theta(\mathbf{X}) = \frac{1}{(\sqrt{2\pi})^n}\,\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^n (X_i-\theta)^2}.$$

Dans le cas d'une fonction de perte équilibrée, le test bayésien pour Π s'écrit

$$\varphi(\mathbf{X}) = \mathbb{1}_{\Pi[\Theta_0|\mathbf{X}]\leq\Pi[\Theta_1|\mathbf{X}]} = \mathbb{1}_{\Pi[\Theta_0|\mathbf{X}]\leq 1/2}.$$

Il faut donc calculer $\Pi[\Theta_0|\mathbf{X}] = \Pi[\{0\}|\mathbf{X}]$. En notant $g(\theta)$ la densité de la loi $\mathcal{N}(0,\sigma^2)$, on a

$$\Pi[\{0\}|\mathbf{X}] = \frac{\alpha p_0(\mathbf{X})}{\alpha p_0(\mathbf{X}) + (1-\alpha)\int_{\mathbb{R}} p_\theta(\mathbf{X}) g(\theta)\,\mathrm{d}\theta}.$$

On calcule ensuite

$$\begin{aligned}\int_{\mathbb{R}} p_\theta(\mathbf{X}) g(\theta)\,\mathrm{d}\theta &= \frac{1}{(\sqrt{2\pi})^n}\frac{1}{\sqrt{2\pi\sigma^2}}\int \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^n (X_i-\theta)^2}\,\mathrm{e}^{-\frac{\theta^2}{2\sigma^2}}\,\mathrm{d}\theta \\ &= \frac{\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^n X_i^2}}{(\sqrt{2\pi})^n}\frac{1}{\sqrt{2\pi\sigma^2}}\int \mathrm{e}^{-\frac{n+\sigma^{-2}}{2}\left(\theta-\frac{n\overline{X}_n}{n+\sigma^{-2}}\right)^2+\frac{(n\overline{X}_n)^2}{2(n+\sigma^{-2})}}\,\mathrm{d}\theta \\ &= \frac{p_0(\mathbf{X})}{\sqrt{\sigma^2(n+\sigma^{-2})}}\,\mathrm{e}^{\frac{(n\overline{X}_n)^2}{2(n+\sigma^{-2})}}.\end{aligned}$$

On en déduit que

$$\varphi(\mathbf{X}) = \mathbb{1}_{\left\{\alpha\sqrt{\sigma^2 n+1}\leq(1-\alpha)\,\mathrm{e}^{\frac{(n\overline{X}_n)^2}{2(n+\sigma^{-2})}}\right\}}.$$

1. $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$.

Remarque : On peut faire le même genre d'interprétation qu'en exercice précédent : si H_0 est vraie, alors $\sqrt{n}\overline{X}_n \sim \mathcal{N}(0,1)$ donc $(\sqrt{n}\overline{X}_n)^2 \sim \chi_1^2$ et par le Lemme de Slutsky, si on note $Y \sim \chi_1^2$, on a donc

$$(1-\alpha)\,\mathrm{e}^{\frac{(n\overline{X}_n)^2}{2(n+\sigma^{-2})}} \xrightarrow[n\to\infty]{d} (1-\alpha)\,\mathrm{e}^{\frac{Y}{2}}$$

tandis que $\alpha\sqrt{\sigma^2 n+1}$ tend vers l'infini.

2. Dans le cas des hypothèses H_0' et H_1', une loi a priori diffuse suffit pour mettre une masse positive sur chacune des deux hypothèses. On peut donc choisir $\Pi' = \mathcal{N}(0,\sigma^2)$. Une mesure dominante pour la loi a priori est $\nu = \lambda_{\mathbb{R}}$, où $\lambda_{\mathbb{R}}$ désigne la mesure de Lebesgue sur $\mathbb{R}$. La densité de la loi a priori Π' par rapport à cette mesure dominante est

$$\pi'(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{\theta^2}{2\sigma^2}}.$$

La loi a posteriori pour Π' est, d'après le calcul standard dans le cas gaussien conjugué,

$$\Pi'[\cdot|\mathbf{X}] = \mathcal{N}\left(\frac{n\overline{X}_n}{\sigma^{-2}+n}, \frac{1}{\sigma^{-2}+n}\right).$$

Ainsi, le test de Bayes pour Π' et la perte équilibrée est donné par

$$\varphi'(\mathbf{X}) = \mathbb{1}_{\left\{\mathbb{P}\left(\left|\mathcal{N}\left(\frac{n\overline{X}_n}{\sigma^{-2}+n}, \frac{1}{\sigma^{-2}+n}\right)\right| \le \varepsilon \,\middle|\, \mathbf{X}\right) \le 1/2\right\}}.$$

En notant comme d'habitude Φ la fonction de répartition de la loi gaussienne standard, ceci revient à dire que $\varphi'(\mathbf{X}) = 1$ si et seulement si

$$\Phi\left(\sqrt{\sigma^{-2}+n}\left(\varepsilon - \frac{n\overline{X}_n}{\sigma^{-2}+n}\right)\right) - \Phi\left(\sqrt{\sigma^{-2}+n}\left(-\varepsilon - \frac{n\overline{X}_n}{\sigma^{-2}+n}\right)\right) \le \frac{1}{2}.$$

Exercice 4 (Le phénomène de Hodges)

1. Le risque quadratique de $\overline{X}_n$ est constant, égal à $\mathbf{R}(\theta, T_n) = 1/n$. Quant à S_n, son risque quadratique vaut

$$\begin{aligned}\mathbb{E}_\theta[(S_n-\theta)^2] &= \mathbb{E}_\theta\left[(S_n-\theta)^2\mathbb{1}_{\{|\overline{X}_n|\le n^{-1/4}\}}\right] + \mathbb{E}_\theta\left[(S_n-\theta)^2\mathbb{1}_{\{|\overline{X}_n|> n^{-1/4}\}}\right]\\ &= \theta^2\mathbb{P}_\theta\left(|\overline{X}_n| \le n^{-1/4}\right) + \mathbb{E}_\theta\left[(\overline{X}_n-\theta)^2\mathbb{1}_{\{|\overline{X}_n|> n^{-1/4}\}}\right].\end{aligned}$$

2. Nous distinguons bien entendu les cas $\theta = 0$ et $\theta \neq 0$.

Première solution :

— Si $\theta = 0$: dans ce cas $Y_n := \sqrt{n}\overline{X}_n \sim \mathcal{N}(0,1)$ et

$$\mathbf{R}(\theta, S_n) = \frac{1}{n}\mathbb{E}_\theta\left[(\sqrt{n}\overline{X}_n)^2\mathbb{1}_{\{\sqrt{n}|\overline{X}_n|> n^{1/4}\}}\right] = \frac{1}{n}\mathbb{E}_\theta\left[Y_n^2\mathbb{1}_{\{|Y_n|> n^{1/4}\}}\right]$$

donc

$$\frac{\mathbf{R}(\theta, S_n)}{\mathbf{R}(\theta, T_n)} = \mathbb{E}_\theta\left[Y_n^2\mathbb{1}_{\{|Y_n|> n^{1/4}\}}\right] = \int_{\mathbb{R}} \frac{y^2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\mathbb{1}_{\{|y|>n^{1/4}\}}dy \xrightarrow[n\to\infty]{} 0$$

par le théorème de convergence dominée.

— Si $\theta \neq 0$: dans ce cas $Y_n := \sqrt{n}(\overline{X}_n - \theta) \sim \mathcal{N}(0,1)$ et

$$\mathbf{R}(\theta, S_n) = \frac{1}{n}\mathbb{E}_\theta\left[Y_n^2 \mathbb{1}_{\{|\theta + Y_n/\sqrt{n}| > n^{-1/4}\}}\right] + \theta^2 \mathbb{P}_\theta\left(|\overline{X}_n| \leq n^{-1/4}\right).$$

Puisque $|\theta| \leq |\overline{X}_n - \theta| + |\overline{X}_n|$ et $|\theta|\sqrt{n} - n^{1/4} > 0$ pour n assez grand, l'inégalité de Markov donne (en se souvenant que $\mathbb{E}[\mathcal{N}(0,1)^4] = 3$)

$$\begin{aligned}\mathbb{P}_\theta\left(|\overline{X}_n| \leq n^{-1/4}\right) &\leq \mathbb{P}_\theta\left(|\overline{X}_n - \theta| \geq |\theta| - n^{-1/4}\right)\\ &= \mathbb{P}_\theta\left(|Y_n| \geq |\theta|\sqrt{n} - n^{1/4}\right)\\ &\leq \frac{3}{(|\theta|\sqrt{n} - n^{1/4})^4}\end{aligned}$$

de sorte que

$$n\theta^2 \mathbb{P}_\theta\left(|\overline{X}_n| \leq n^{-1/4}\right) \xrightarrow[n\to\infty]{} 0.$$

Par ailleurs, une nouvelle application du théorème de convergence dominée assure que

$$\mathbb{E}_\theta\left[Y_n^2 \mathbb{1}_{\{|\theta + Y_n/\sqrt{n}| > n^{-1/4}\}}\right] = \int_{\mathbb{R}} \frac{y^2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \mathbb{1}_{\{|\theta + y/\sqrt{n}| > n^{-1/4}\}} dy \xrightarrow[n\to\infty]{} \mathbb{E}[\mathcal{N}(0,1)^2] = 1.$$

Seconde solution : Si $\theta \neq 0$, on a, en utilisant que $|\theta| \leq |\overline{X}_n - \theta| + |\overline{X}_n|$,

$$\begin{aligned}\mathbb{P}_\theta\left(|\overline{X}_n| \leq n^{-1/4}\right) &\leq \mathbb{P}_\theta\left(|\overline{X}_n - \theta| \geq |\theta| - n^{-1/4}\right)\\ &= \mathbb{P}\left(|\mathcal{N}(0,1)| \geq |\theta|\sqrt{n} - n^{1/4}\right),\end{aligned}$$

où l'on a utilisé le fait que la loi de $\overline{X}_n - \theta$ sous P_θ est $\mathcal{N}(0, 1/n) = \frac{1}{\sqrt{n}}\mathcal{N}(0,1)$. Pour n assez grand $|\theta|\sqrt{n} - n^{1/4} > 0$ et en utilisant la majoration sur les queues gaussiennes, on obtient

$$\mathbb{P}_\theta\left(|\overline{X}_n| \leq n^{-1/4}\right) \leq \frac{2}{(|\theta|\sqrt{n} - n^{1/4})\sqrt{2\pi}} \exp\left(-\frac{(|\theta|\sqrt{n} - n^{1/4})^2}{2}\right) = o\left(e^{-\frac{\theta^2 n}{2}}\right).$$

D'autre part, par l'inégalité de Cauchy–Schwarz, on a

$$\mathbb{E}_\theta\left[(\overline{X}_n - \theta)^2 \mathbb{1}_{\{|\overline{X}_n| \leq n^{-1/4}\}}\right] \leq \sqrt{\mathbb{E}_\theta\left[(\overline{X}_n - \theta)^4\right] \mathbb{P}_\theta\left(|\overline{X}_n| \leq n^{-1/4}\right)} = o\left(e^{-\frac{\theta^2 n}{4}}\right),$$

car $\mathbb{E}_\theta\left[(\overline{X}_n - \theta)^4\right] = \frac{3}{n^2} = o(1)$. Ainsi, pour tout $\theta \neq 0$, on a

$$\mathbb{E}_\theta[(S_n - \theta)^2] = \mathbb{E}_\theta\left[(\overline{X}_n - \theta)^2\right] + o\left(e^{-\frac{\theta^2 n}{4}}\right) \sim \frac{1}{n}\cdot$$

Pour $\theta = 0$, on a, encore par l'inégalité de Cauchy–Schwarz,

$$\mathbb{E}_\theta[(S_n - \theta)^2] = \mathbb{E}_0\left[\overline{X}_n^2 \mathbb{1}_{\{|\overline{X}_n| > n^{-1/4}\}}\right] \leq \sqrt{\mathbb{E}_0\left[\overline{X}_n^4\right] \mathbb{P}_0\left(|\overline{X}_n| > n^{-1/4}\right)},$$

c'est-à-dire

$$\mathbb{E}_\theta[(S_n - \theta)^2] \leq \frac{\sqrt{3}}{n}\sqrt{\mathbb{P}_0\left(|\overline{X}_n| > n^{-1/4}\right)}.$$

Or, en utilisant à nouveau la majoration sur les queues gaussiennes,

$$\mathbb{P}_0\left(|\overline{X}_n| > n^{-1/4}\right) = \mathbb{P}\left(|\mathcal{N}(0,1)| > n^{1/4}\right) \leq \frac{2}{n^{1/4}\sqrt{2\pi}} e^{-\frac{\sqrt{n}}{2}}.$$

Ainsi, pour $\theta = 0$, on a $\mathbb{E}_\theta[(S_n - \theta)^2] = o\left(\frac{1}{n}\right)$.

3. En ne retenant que l'espérance sur l'événement $\{|\overline{X}_n| \leq n^{-1/4}\}$, on a

$$\begin{aligned}\mathbb{E}_\theta[(S_n-\theta)^2] &\geq \theta^2 \mathbb{P}_\theta\left(|\overline{X}_n| \leq n^{-1/4}\right) \\ &= \theta^2 \mathbb{P}_\theta\left(\left|\mathcal{N}\left(\theta, \frac{1}{n}\right)\right| \leq n^{-1/4}\right) \\ &\geq \theta^2 \mathbb{P}\left(0 \leq \mathcal{N}\left(\theta, \frac{1}{n}\right) \leq n^{-1/4}\right).\end{aligned}$$

4. Pour $\theta_\star = n^{-1/4}$, on a

$$\begin{aligned}\mathbb{E}_{\theta_\star}[(S_n-\theta_\star)^2] &\geq \theta_\star^2 \mathbb{P}\left(0 \leq \mathcal{N}\left(\theta_\star, \frac{1}{n}\right) \leq n^{-1/4}\right) \\ &= n^{-1/2}\mathbb{P}\left(0 \leq n^{-1/4} + \mathcal{N}\left(0, \frac{1}{n}\right) \leq n^{-1/4}\right) \\ &= n^{-1/2}\mathbb{P}\left(-n^{-1/4} \leq \mathcal{N}\left(0, \frac{1}{n}\right) \leq 0\right) \\ &= n^{-1/2}\mathbb{P}\left(-n^{1/4} \leq \mathcal{N}(0,1) \leq 0\right).\end{aligned}$$

   Or $\mathbb{P}\left(-n^{1/4} \leq \mathcal{N}(0,1) \leq 0\right) \underset{n\to\infty}{\longrightarrow} \mathbb{P}\left(\mathcal{N}(0,1) \leq 0\right) = \frac{1}{2}$.

5. On en conclut que le risque maximal de S_n est au moins de l'ordre de $\frac{1}{\sqrt{n}}$, alors que celui de T_n vaut $1/n$. Ainsi, même si, au vu de la première question, S_n pouvait paraître meilleur que T_n, l'estimateur S_n n'est pas uniformément bon car son risque maximal est bien plus grand. La figure 1 illustre ce qui se passe.

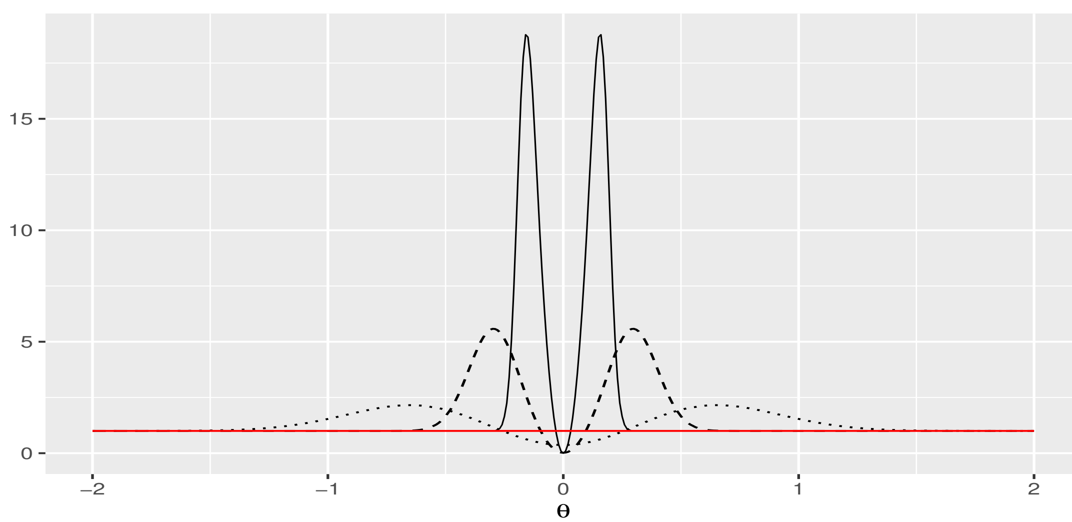

FIGURE 1 – Phénomène de Hodges : $\frac{\mathbf{R}(\theta,S_n)}{\mathbf{R}(\theta,T_n)}$ pour $n = 10$ $(\cdots)$, $n = 100$ (- - -) et $n = 1000$ (trait plein noir).

EXERCICE 5 (Distance en variation totale)

1. Posons $\Lambda = \{k \in \mathbb{N},\, p_k > q_k\}$ et soit A une partie quelconque de $\mathbb{N}$. On a

$$P(A) - Q(A) \leq P(A \cap \Lambda) - Q(A \cap \Lambda) \leq P(\Lambda) - Q(\Lambda)\,,$$

et

$$Q(A) - P(A) \leq Q(\Lambda^c) - P(\Lambda^c) = P(\Lambda) - Q(\Lambda).$$

Ainsi le supremum est atteint en Λ et

$$d_{\mathrm{VT}}(P, Q) = P(\Lambda) - Q(\Lambda) = \sum_{k \in \mathbb{N}} (p_k - q_k)_+\,.$$

2. Il suffit d'écrire $(p_k - q_k)_+ = p_k - \min(p_k, q_k)$.
3. Soit (X, Y) un couplage de P et Q. Pour tout $A \subset \mathbb{N}$, on a

$$\begin{aligned} P(A) - Q(A) &= \mathbb{P}(X \in A) - \mathbb{P}(Y \in A) \\ &\leq \mathbb{P}(X \in A) - \mathbb{P}(Y \in A, X \in A) \\ &= \mathbb{P}(X \in A, Y \notin A) \leq \mathbb{P}(X \neq Y)\,. \end{aligned}$$

D'autre part, soit (X, Y) un couple de variables aléatoires dont la loi est donnée par : pour $(x, y) \in \mathbb{N}^2$,

$$\mathbb{P}(X = x, Y = y) = \min(p_x, q_x)\mathbb{1}_{x=y} + \frac{(p_x - q_x)_+(q_y - p_y)_+}{d(P, Q)}\,.$$

On vérifie que X est de loi P, Y de loi Q, et que $\mathbb{P}(X \neq Y) = d_{\mathrm{VT}}(P, Q)$.
4. Soient $(X_1, Y_1), \ldots, (X_n, Y_n)$ des couples i.i.d. selon le couplage optimal de P et Q (celui de la question précédente). Alors on a

$$\begin{aligned} d_{\mathrm{VT}}(P^{\otimes n}, Q^{\otimes n}) &\leq \mathbb{P}\left((X_1, \ldots, X_n) \neq (Y_1, \ldots, Y_n)\right) \\ &\leq \sum_{i=1}^{n} \mathbb{P}(X_i \neq Y_i) \\ &= n d_{\mathrm{VT}}(P, Q)\,. \end{aligned}$$

5. Soient $(X_1, Y_1), \ldots, (X_n, Y_n)$ des couples i.i.d. avec $X_i \sim \mathcal{B}\left(\frac{\lambda}{n}\right)$ et $Y_i \sim \mathcal{P}\left(\frac{\lambda}{n}\right)$, (X_i, Y_i) distribué selon le couplage optimal entre $\mathcal{B}\left(\frac{\lambda}{n}\right)$ et $\mathcal{P}\left(\frac{\lambda}{n}\right)$. Alors $\sum_{i=1}^n X_i$ est de loi P et $\sum_{i=1}^n Y_i$ est de loi Q. On a

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \neq \sum_{i=1}^{n} Y_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(X_i \neq Y_i) = n d_{\mathrm{VT}}\left(\mathcal{B}\left(\frac{\lambda}{n}\right), \mathcal{P}\left(\frac{\lambda}{n}\right)\right)\,.$$

Or

$$d_{\mathrm{VT}}\left(\mathcal{B}\left(\frac{\lambda}{n}\right), \mathcal{P}\left(\frac{\lambda}{n}\right)\right) = \sum_{k \in \mathbb{N}} \left(\mathbb{P}(X_i = k) - \mathbb{P}(Y_i = k)\right)_+\,.$$

Tous les termes de la somme ci-dessus sont nuls sauf pour $k = 1$. Ainsi $\mathbb{P}(X_i \neq Y_i) = \frac{\lambda}{n}\left(1 - e^{-\lambda/n}\right)$.