

TD 10 : TESTS BAYÉSIENS

EXERCICE 1 (Borne de Le Cam)

1. On considère le modèle $\mathcal{P} = \{P_\theta^{\otimes n}, \theta \in \mathbb{R}\}$ où $P_\theta = \mathcal{N}(\theta, 1)$.
 - (a) Rappeler l'EMV dans \mathcal{P} et déterminer son risque quadratique.
 - (b) Exprimer, pour $P_\theta \in \mathcal{P}$ et $P_{\theta'} \in \mathcal{P}$, la distance de Hellinger entre P_θ et $P_{\theta'}$, notée $h(P_\theta, P_{\theta'})$ et en déduire, pour tous $\theta \in \mathbb{R}$ et $\varepsilon \in \mathbb{R}$, un équivalent de $h(P_\theta, P_{\theta+\varepsilon})$ lorsque $\varepsilon \rightarrow 0$.
 - (c) Exprimer la borne de Le Cam pour le risque quadratique en fonction de ε et justifier que le supremum est atteint pour ε proche de 0.
 - (d) En utilisant l'équivalent de $h(P_\theta, P_{\theta+\varepsilon})$ lorsque $\varepsilon \rightarrow 0$, donner l'ordre de la borne de Le Cam en fonction de n et le comparer au risque maximal de l'EMV. Conclure sur l'ordre du risque minimax.
2. Reprendre la question 1 avec $\mathcal{P} = \{P_\theta^{\otimes n}, \theta \in \mathbb{R}\}$ où $P_\theta = \mathcal{E}(1) + \theta$.

EXERCICE 2 (Test bayésien I)

Soit $\mathbf{X} = (X_1, \dots, X_n) \mid \boldsymbol{\theta} \sim \mathcal{B}(\boldsymbol{\theta})^{\otimes n}$, avec $\boldsymbol{\theta} \in]0, 1[$. On souhaite tester $H_0 : \boldsymbol{\theta} = 1/2$ contre $H_1 : \boldsymbol{\theta} \neq 1/2$. On considère une loi a priori de la forme

$$\Pi = \alpha \delta_{1/2} + (1 - \alpha) \text{Unif}[0, 1], \quad \alpha \in]0, 1[.$$

Montrer que le test de Bayes pour H_0 contre H_1 pour cette loi a priori et la fonction de perte équilibrée peut s'écrire

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

Rappel : pour tous $a > 0$ et $b > 0$, $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

EXERCICE 3 (Test bayésien II)

Soit $\mathbf{X} = (X_1, \dots, X_n) \mid \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}, 1)^{\otimes n}$.

1. On considère le problème de test $H_0 : \boldsymbol{\theta} = 0$ versus $H_1 : \boldsymbol{\theta} \neq 0$, et on choisit une loi a priori Π de la forme $\Pi = \alpha \delta_0 + (1 - \alpha) \mathcal{N}(0, \sigma^2)$ avec $\alpha \in]0, 1[$ et $\sigma^2 > 0$. Montrer que le test de Bayes pour H_0 contre H_1 pour cette loi a priori et la fonction de perte équilibrée peut s'écrire

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

2. Pour $\varepsilon > 0$, on considère $H'_0 : |\boldsymbol{\theta}| \leq \varepsilon$ versus $H'_1 : |\boldsymbol{\theta}| > \varepsilon$. On choisit cette fois la loi a priori $\Pi' = \mathcal{N}(0, \sigma^2)$. Déterminer le test de Bayes pour Π' et la perte équilibrée.

EXERCICE 4 (Le phénomène de Hodges)

On considère, pour X_1, \dots, X_n i.i.d. de loi $\mathcal{N}(\theta, 1)$ avec $\theta \in \mathbb{R}$, l'estimateur $T_n = \bar{X}_n$ ainsi que l'estimateur de Hodges $S_n = \bar{X}_n \mathbf{1}_{|\bar{X}_n| > n^{-1/4}}$.

1. En notant \mathbf{R} le risque quadratique, rappeler ce que vaut $\mathbf{R}(\theta, T_n)$ et montrer que

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

2. Prouver que

$$\lim_{n \rightarrow +\infty} \frac{\mathbf{R}(\theta, S_n)}{\mathbf{R}(\theta, T_n)} = \begin{cases} 0 & \text{si } \theta = 0, \\ 1 & \text{si } \theta \neq 0. \end{cases}$$

Indication : pour majorer $\mathbb{P}_\theta(|\bar{X}_n| \leq n^{-1/4})$ lorsque $\theta \neq 0$, on pourra utiliser l'inégalité de Markov ou la majoration classique pour les queues gaussiennes : pour tout $x > 0$,

$$\mathbb{P}(\mathcal{N}(0, 1) > x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-t^2/2} dt \leq \frac{1}{x\sqrt{2\pi}} \int_x^{+\infty} te^{-t^2/2} dt = \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}.$$

3. Montrer que pour tout $\theta \in \mathbb{R}$,

$$\mathbb{E}_\theta[(S_n - \theta)^2] \geq \theta^2 \mathbb{P} \left(0 \leq \mathcal{N} \left(\theta, \frac{1}{n} \right) \leq \frac{1}{n^{1/4}} \right).$$

4. En déduire que pour $\theta_\star = n^{-1/4}$, on a

$$\mathbb{E}_{\theta_\star}[(S_n - \theta_\star)^2] \geq \frac{1}{\sqrt{n}} \left(\frac{1}{2} - o(1) \right).$$

5. Comparer le risque maximal de T_n à celui de S_n . Ce dernier est-il "uniformément bon" ?

EXERCICE 5 (★) (Distance en variation totale et couplage)

Soient P et Q deux lois de probabilités sur \mathbb{N} . Pour tout $k \in \mathbb{N}$, on note $p_k = P(\{k\})$ et $q_k = Q(\{k\})$. On rappelle que la distance en variation totale entre P et Q est définie par $d_{\text{VT}}(P, Q) = \sup_{A \subset \mathbb{N}} |P(A) - Q(A)|$.

1. Montrer que le supremum dans la définition de $d_{\text{VT}}(P, Q)$ est atteint en une partie de \mathbb{N} que l'on déterminera, et en déduire que $d_{\text{VT}}(P, Q) = \sum_{k \in \mathbb{N}} (p_k - q_k)_+$ où pour $z \in \mathbb{R}$, $z_+ = \max(z, 0)$.
2. Montrer que $d_{\text{VT}}(P, Q) = 1 - \sum_{k \in \mathbb{N}} \min(p_k, q_k)$.
3. On appelle *couplage* de P et Q un couple de variables aléatoires (X, Y) tel que $X \sim P$ et $Y \sim Q$. Montrer que pour tout couplage (X, Y) de P et Q , on a $d_{\text{VT}}(P, Q) \leq \mathbb{P}(X \neq Y)$ et vérifier que le couple (X, Y) de loi donnée par : pour $(x, y) \in \mathbb{N}^2$,

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

est un couplage de P et Q tel que $\mathbb{P}(X \neq Y) = d_{\text{VT}}(P, Q)$.

4. Montrer que $d_{\text{VT}}(P^{\otimes n}, Q^{\otimes n}) \leq n d_{\text{VT}}(P, Q)$.
5. Soit $\lambda > 0$. On prend $P = \mathcal{B}(n, \frac{\lambda}{n})$ et $Q = \mathcal{P}(\lambda)$. Montrer que $d_{\text{VT}}(P, Q) \leq \lambda \left(1 - e^{-\frac{\lambda}{n}} \right)$.