

TP5 : ESTIMATEURS CLASSIQUES

1. MAXIMUM DE VRAISEMBLANCE

Exercice 1 [Etude d'une famille exponentielle]. Soit $X_1, \dots, X_n$ un échantillon de loi $\mathcal{N}(\theta, \theta^2)$, où θ est un réel strictement positif. La densité correspondante est notée $p_\theta(x)$.

- (1) A quels estimateurs pouvez-vous penser pour estimer le paramètre θ ? Quelle est leur variance?
- (2) Calculez la log-vraisemblance associée aux observations pour le paramètre θ et en déduire que le maximum de vraisemblance $\hat{\theta}_n$ est défini par :

$$\hat{\theta}_n = -\frac{\overline{X_n}}{2} + \sqrt{\overline{X_n^2} + \frac{1}{4}\overline{X_n}^2},$$

où $\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$ et $\overline{X_n^2} = \frac{1}{n} \sum_{i=1}^n X_i^2$ désignent respectivement la moyenne empirique et le moment empirique d'ordre 2.

- (3) Ecrivez un programme simulant N réalisations de l'estimateur $\hat{\theta}_n$ et illustrez sa consistance, que vous vérifierez également par le calcul. Comparez les convergences vers θ de $\hat{\theta}_n$ et des estimateurs proposés en (1).

Notons $\dot{l}_\theta(x)$ la dérivée partielle en θ de $(\theta, x) \rightarrow \ln p_\theta(x)$. Cette quantité s'appelle *fonction score* pour l'estimation de θ . On note aussi $I_\theta = \mathbb{E}(\dot{l}_\theta(X)^2)$, où X suit la loi $\mathcal{N}(\theta, \theta^2)$. Cette quantité est l'*information de Fisher* associée au modèle au point θ . On peut alors montrer le résultat suivant :

$$\sqrt{n}(\hat{\theta}_n - \theta) = I_\theta^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_\theta(X_i) + o_P(1),$$

où la notation $o_P(1)$ désigne une fonction qui tend en probabilité vers 0.

- (4) Illustrez le résultat précédent. Pour cela, montrez que I_θ vaut $3/\theta^2$ puis voyez si le N-échantillon construit en (2) a une variance proche de $(nI_\theta)^{-1}$. On peut en fait démontrer qu'il s'agit de la variance asymptotique optimale.

Exercice 2. Dans cet exercice, on part d'un échantillon $X_1, \dots, X_n$ de loi uniforme sur l'intervalle $[0, \theta]$, où θ est un réel strictement positif.

- (1) Montrez que le maximum de vraisemblance $\hat{\theta}_n$ associé à l'estimation de θ est le maximum des X_i ($i = 1, \dots, n$).

- (2) Simulez un N-échantillon de la variable $\hat{\theta}_n$. L'estimateur est-il consistant ? Que pouvez-vous dire de la quantité $\sqrt{n}(\hat{\theta}_n - \theta)$? Illustrez puis vérifiez par le calcul qu'il y a convergence en loi (ou en probabilité, c'est ici la même chose ; au fait, pourquoi ?) vers 0.
- (3) Cherchons à obtenir une convergence plus intéressante. Pour cela, on renormalise par n la différence $(\hat{\theta}_n - \theta)$. A partir du N-échantillon précédent, illustrez le fait que $-n(\hat{\theta}_n - \theta)$ converge en loi. Reconnaissez-vous la loi limite ? Etablir cette convergence par le calcul.
- (4) Comparez les propriétés de convergence de l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ et de l'estimateur Y_n défini par $Y_n = (1 + \frac{1}{n})\hat{\theta}_n$. Lequel préférez-vous et pourquoi ? L'estimateur du maximum de vraisemblance a-t-il toujours une variance asymptotique optimale ?

Exercice 3 [Modèle linéaire gaussien]. On considère le modèle de régression linéaire suivant :

$$Y = \beta_1 f(x) + \beta_2 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2),$$

où β_1 , β_2 et σ^2 sont des paramètres inconnus à estimer.

- (1) Donner l'expression des estimateurs des paramètres à estimer pour un échantillon $(x_1, Y_1), \dots, (x_n, Y_n)$. Ecrivez une fonction qui simule le nuage de points correspondant à l'échantillon précédent pour des points $x_1, \dots, x_n$ répartis uniformément sur $[0, 1]$ et la fonction f définie par $f(x) = (1 + x)^2$.
- (2) Ecrivez une fonction qui trace la courbe de régression et qui donne des intervalles de confiance de niveau $1 - \alpha$ pour les paramètres à estimer. Trouvez la fonction de Stixbox qui permet de traiter directement cette question et comparez le programme correspondant (accessible avec la commande `open`) au vôtre.
- (3) Tracer deux régions de confiance de niveau 95% pour le couple (β_1, β_2) : l'une utilisant les intervalles de confiance précédents et l'autre plus globale.

2. ESTIMATEURS BAYÉSIENS

Commençons par quelques rappels sur le principe de l'estimation bayésienne. On part d'un échantillon $X_1, \dots, X_n$ de loi commune P_{θ_0} de densité p_{θ_0} , où $\theta_0 \in \Theta$ est un paramètre à estimer. Le point de vue bayésien consiste à se donner une mesure de probabilité ν sur Θ dite loi *a priori* et de « réactualiser » cette loi à l'aide des observations. On introduit artificiellement de l'aléa sur θ qui était initialement un paramètre fixé : on en fait une variable aléatoire de loi ν . La loi P_θ de X devient alors la loi conditionnelle de X sachant θ , la loi du couple (θ, X) est $p_\theta(x)dx d\nu(\theta)$ et la loi de X est maintenant $f_X(x)dx = \int_\Theta p_\theta(x)dx d\nu(\theta)$. On utilise l'information que nous donnent les observations en considérant la mesure *a posteriori* de θ sachant $X_1, \dots, X_n$ définie par

$$\Pi(d\theta | X_1, \dots, X_n) = \frac{p_\theta^{\otimes n}(X_1, \dots, X_n) d\nu(\theta)}{f_X^{\otimes n}(X_1, \dots, X_n)}.$$

Autrement dit, si ν admet pour densité f par rapport à la mesure de Lebesgue sur Θ , alors la mesure a posteriori a pour densité $C_n(X_1, \dots, X_n)^{-1} \prod_{i=1}^n p_\theta(X_i) f(\theta)$, avec $C_n(X_1, \dots, X_n) = \int_{\Theta} \prod_{i=1}^n p_\theta(X_i) f(\theta) d\theta$ constante de renormalisation.

Sous de très faibles hypothèses sur le modèle, la suite de mesures a posteriori $\Pi(\cdot | X_1, \dots, X_n)$ converge en loi vers la masse de Dirac δ_{θ_0} au point θ_0 et cela presque sûrement, sauf éventuellement sur un ensemble de mesure nulle (théorème de Doob).

Il faut noter en particulier que l'estimateur n'est cette fois pas une variable, comme dans le cas du maximum de vraisemblance, mais une *mesure*, qui dépend des observations $X_1, \dots, X_n$ et est donc aléatoire. Cependant, on peut se ramener à un estimateur au sens classique du terme en considérant par exemple l'espérance de la mesure a posteriori, ou, si cette dernière admet une densité, le point où cette densité est maximale.

Exercice 4. Nous allons essayer d'illustrer la convergence des mesures a posteriori vers δ_θ pour différents choix d'a priori sur deux exemples simples.

- (1) Soit à estimer le paramètre $\theta \in [0, 1]$ d'une loi de Bernoulli $\mathcal{B}(\theta)$. Illustrez la convergence de la mesure a posteriori pour plusieurs n -échantillons simulés, et cela pour une loi a priori ν à choisir. Vous pourrez par exemple considérer pour loi a priori une loi Beta de paramètres α et β , la densité $f_{\alpha, \beta}$ d'une telle loi sur $[0, 1]$ est donnée par :

$$f_{\alpha, \beta}(x) = C_{\alpha, \beta}^{-1} x^{\alpha-1} (1-x)^{\beta-1}, \quad \text{où} \quad C_{\alpha, \beta} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

- (2) On dispose maintenant d'observations suivant une loi normale $\mathcal{N}(\theta, 1)$, où la moyenne θ est à estimer. Illustrez de même la convergence de la mesure a posteriori pour un a priori que vous choisirez.

Exercice 5 [Théorème de Bernstein-Von Mises]. Sous des conditions assez faibles sur le modèle considéré, on peut montrer que la mesure a posteriori pour n observations est proche pour n grand d'une loi normale centrée en l'estimateur du maximum de vraisemblance pour le modèle considéré. Plus précisément, quand n tend vers $+\infty$, la convergence suivante est vérifiée :

$$\left\| \Pi(\cdot | X_1, \dots, X_n) - \mathcal{N}\left(\hat{\theta}_n, \frac{1}{n} I_\theta^{-1}\right) \right\| \rightarrow 0 \quad \text{sous } P_\theta^n,$$

où $\hat{\theta}_n$ désigne l'estimateur du maximum de vraisemblance, I_θ l'information de Fisher associée au modèle et $\|P - Q\|$ désigne pour deux mesures P et Q leur distance en variation totale définie de la façon suivante. Si P et Q sont deux mesures de proba sur un espace $\mathcal{E}$ de densités p et q par rapport à μ , $\|P - Q\| = \sup_{B \in \mathcal{B}(\mathcal{E})} |P(B) - Q(B)| = \int |p - q| d\mu$, où $\mathcal{B}(\mathcal{E})$ désigne la tribu des boréliens sur $\mathcal{E}$.

Proposez une illustration de ce théorème dans le cadre de la question (2) de l'exercice 4.