Projet X-MAP-311-2017-CHAFAI-2
Estimateur par fenêtres glissantes

Proposé par Djalil Chafaï

Second semestre 2016-2017

Soient $(X_n)_{n\geq 1}$ des variables aléatoires réelles indépendantes et identiquement distribuées de densité f. Soit $(h_n)_{n\geq 1}$ une suite de réels strictement positifs. Pour tous $x \in \mathbb{R}$ et $n \geq 1$, soit

$$\hat{f}_n(x) = \frac{F_n(x+h_n) - F_n(x-h_n)}{2h_n} \quad \text{où} \quad F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{X_k \leq x\}}.$$

Dans tout ce qui suit, on suppose que

$$\lim_{n\to\infty} h_n = 0 \quad \text{et} \quad \lim_{n\to\infty} nh_n = \infty.$$

1. Démontrer que pour tout $n \geq 1$, $\hat{f}_n$ est une densité de probabilité
2. Démontrer que pour tout $x \in \mathbb{R}$, on a la décomposition biais-variance
$$\mathbb{E}((\hat{f}_n(x) - f(x))^2) = (\mathbb{E}(\hat{f}_n(x)) - f(x))^2 + \mathrm{Var}(\hat{f}_n(x)).$$
3. Pour tout $x \in \mathbb{R}$, en posant
$$p_n(x) = F(x+h_n) - F(x-h_n),$$
montrer que
$$\lim_{n\to\infty} \mathbb{E}(\hat{f}_n(x)) = \lim_{n\to\infty} \frac{p_n(x)}{2h_n} = f(x)$$
4. Démontrer que pour tous $x \in \mathbb{R}$ et $n \geq 1$,
$$2nh_n\hat{f}_n(x) = \sum_{k=1}^{n} \mathbf{1}_{\{X_k \in\,]x-h, x+h]\}}$$
5. En déduire que pour tout $x \in \mathbb{R}$,
$$\lim_{n\to\infty} \mathrm{Var}(\hat{f}_n(x)) = \lim_{n\to\infty} \frac{np_n(x)(1-p_n(x))}{4n^2h_n^2} = 0$$
6. En déduire que pour tout $x \in \mathbb{R}$, en probabilité,
$$\lim_{n\to\infty} \hat{f}_n(x) = f(x)$$
7. Montrer que si pour tout $n \geq 1$, S_n est une variable aléatoire de loi binomiale de taille n et de paramètre p_n, alors
$$\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1).$$

8. Supposons à partir de maintenant que les deux propriétés suivantes sont réalisées:

   - $\lim_{n\to\infty} nh_n^3 = 0$
   - $x \in \mathbb{R}$ vérifie $f(x) > 0$ et f est dérivable sur un voisinage de x avec dérivée bornée.

   Démontrer que
$$\frac{2nh_n\left(\hat{f}_n(x) - \mathbb{E}(\hat{f}_n(x))\right)}{\sqrt{np_n(x)(1-p_n(x))}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1)$$

9. En déduire que
$$\sqrt{2nh_n}\frac{\left(\hat{f}_n(x) - \mathbb{E}(\hat{f}_n(x))\right)}{\sqrt{f(x)}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1)$$

10. En déduire que
$$\sqrt{2nh_n}\left(\frac{\hat{f}_n(x) - f(x)}{\sqrt{\hat{f}_n(x)}}\right) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1)$$

11. Proposer un programme informatique fournissant des simulations illustrées par de jolis graphiques de l'estimation de densité par fenêtres glissantes. On pourra prendre pour f un mélange de deux densité gaussiennes (combinaison convexe de deux densités gaussiennes de moyennes et de variances différentes), et jouer sur la taille de fenêtre h_n.

12. Dans toute la suite, soit $K : \mathbb{R} \to \mathbb{R}_+$ une fonction appelée noyau vérifiant
$$\int_{\mathbb{R}} K\mathrm{d}\lambda = 1 \quad \text{et} \quad S = \int_{\mathbb{R}} K^2\mathrm{d}\lambda < \infty.$$

    Pour tout $h > 0$ on note $K_h = K(\cdot/h)/h$, et on définit l'estimateur $\hat{f}_{n,h}$ de f par noyau K et largeur de bande h comme suit: pour tout $x \in \mathbb{R}$,
$$\hat{f}_{n,h}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

13. Démontrer que pour tout $x \in \mathbb{R}$
$$\mathbb{E}(\hat{f}_{n,h_n}(x)) = f * K_{h_n}.$$

14. Démontrer que
$$\mathbb{E}\left(\left\|\hat{f}_{n,h_n} - f\right\|_1\right) \leq \|f * K_{h_n} - f\|_1 + \mathbb{E}\left[\left\|\hat{f}_{n,h_n} - f * K_{h_n}\right\|_1\right]$$

    Il est possible d'établir (admis) que les deux termes du membre de droite convergent vers 0 quand n tends vers l'infini. Le premier est un terme de biais, tandis que le second est un terme de variance, comme le suggeère la question suivante:

15. Démontrer que
$$\mathbb{E}\left[\left\|\hat{f}_{n,h_n} - f * K_{h_n}\right\|_1\right] \leq \int_{\mathbb{R}} \sqrt{\mathrm{Var}(\hat{f}_{n,h_n}(x))}\mathrm{d}x.$$

16. Proposer un programme informatique fournissant des simulations illustrées par de jolis graphiques de l'estimation de densité en utilisant un noyau gaussien. On pourra prendre pour f un mélange de deux densité gaussiennes (combinaison convexe de deux densités gaussiennes de moyennes et de variances différentes), et jouer sur la taille de fenêtre h_n.