

Contents

Choix de la fenetre :

1

output: pdf_document: toc: yes html_document: default —

Choix de la fenetre :

Déjà on se ramene a la question du choix de h (choix de la fenetre) parceque a la base on cherchais a minimiser le risque R de l'estimateur \hat{f} pour l'estimation f .

On veut que le risque \mathcal{R} tend vers zero quand le nombre des observation tend vers l'infini

On a pour on introduit deux methodes d'estimation :

Pour la methode a noyau :

Ce qui rend cette méthode interessante est le fait que la convolution $K_h * f$ tend vers f sur L^2 quand h tend vers 0

On peut donc approximer la vrai densité f par le produit de convolution $K_h * f$ qui satisfait $K_h * f(x) = E_f[k_h(x - X_1)]$.

Ce qui nous ramène a l'expression suivante de l'estimateur a noyau de f pour un paramètre $h > 0$ fixé :

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i)$$

Ensuite ; On a que : $E_f[\hat{f}_h] = K_h * f$

Et :

$$\mathcal{R}(\hat{f}_h, f) = E_f[||\hat{f}_h - f||^2]$$

De plus

$$E_f[||\hat{f} - f||^2] = E_f[||\hat{f} + E_f(\hat{f}) - (E_f(\hat{f}) - f)||^2]$$

$$E_f[||\hat{f} - f||^2] = E_f[||\hat{f} + E_f(\hat{f})||^2] + E_f[||E_f(\hat{f}) - f||^2] - 2E_f(<\hat{f} - E_f(\hat{f}); E_f(\hat{f}) - f>)$$

Comme \hat{f} est déterministe :

$$2E_f(<\hat{f} - E_f(\hat{f}); E_f(\hat{f}) - f>) = 2 <0, E_f(\hat{f}) - f> = 0 \text{ Ainsi } ||E_f(\hat{f}) - f|| \text{ est déterministe :}$$

On obtient :

$$R(\hat{f}, f) = E_f[||\hat{f} - f||^2] = ||f - K_h * f||^2 + E_f[||\hat{f} - K_h * f||^2]$$

$$E_f[||\hat{f}_h - K_h * f||^2] = \int Var_f(\hat{f}_h(x))dx < \frac{1}{n} \int E_f[K_h^2(x - X_1)]dx$$

Et comme $||K_h||^2 = ||K||^2/h$

$$E_f[||\hat{f}_h - K_h * f||^2] < ||K||^2/nh$$

Ca nous ramène donc a :

$$\mathcal{R}(\hat{f}_h, f) < ||f - K_h * f||^2 = ||K||^2/nh$$

On voit donc bien que pour minimiser le risque le biais tend vers 0 quand h tend vers 0 mais plus h est petit plus la variance devient grande .

Il faut donc chercher le meilleur compromis entre ces deux

Remarque : le choix de h est beaucoup plus crucial que celui de la noyau

Pour répondre a ce problème on va introduire la méthode de Goldenshluger-Lepski :

L'idée est d'estimer le biai du risque avec des fenetres fixées

Soit $(\hat{f}_h)_{h \in \mathcal{H}_\setminus}$

On va choisir

$$\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}_\setminus} (A(h) + V(h))$$

Avec V est le terme de pénalisation qui estime l'erreur stochastique .

$$V(h) = ||\frac{||K||_1^2 ||K||^2}{nh}, \text{ et } A(h) = \max_{h' \in \mathcal{H}_\setminus} (||\hat{f}_{h'} - \hat{f}_{h,h'}||^2 - V(h'))_+$$