ESTIMATEUR « SPLINE » DE LA RÉGRESSION (H3)
(27 / 11 / 2019, © Monfort, Dicostat2005, 2005-2019)

En **Statistique non paramétrique**, lorsqu'une **fonction de régression** ne se présente pas sous forme paramétrée, on utilise généralement une **méthode non paramétrique** (**estimation non paramétrique**).

(i) Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé** et $(\xi, \eta) : \Omega \mapsto \mathbf{R}^2$ un **couple aléatoire** réel intégrable. On considère un **modèle de régression** (non linéaire), dont la forme usuelle s'écrit (dans l'**espace des variables**) :

(1) $\eta = \rho(\xi) + \varepsilon$, avec $E\,\varepsilon = 0$ et $V\,\varepsilon = \sigma^2$,

et dans laquelle ρ est la fonction de régression de η conditionnelle à ξ.

Par ailleurs, on suppose données N **copies** iid $(x_n, y_n)_{n=1,\ldots,N}$ de (ξ, η) et l'on pose :

(2) $y_n = \rho(x_n) + u_n$,

avec $E\,u_n = 0$ et $C(u_\alpha, u_\beta) = \delta_{\alpha\beta} \,.\, \sigma^2, \forall (\alpha, \beta)$.

On appelle alors **estimateur spline de la régression** ρ (de degré 3), avec **« nœuds »** (ie limites d'intervalles) positionnés aux points (x_n, y_n) la solution du **problème d'optimisation** suivant :

(2) $\min_\rho \ \Sigma_{n=1}^N (y_n - \rho(x_n))^2 + \lambda^{-1} \,.\, \int \{\rho''(x)\}^2 \, dx$,

dans lequel :

(a) le premier terme de la somme indique la **qualité de l'ajustement** (au sens de la **méthode des mco**) de la fonction ρ : c'est la somme des carrés des écarts entre les observations de l'**endogène** η et les « valeurs » prises par la fonction ρ en chaque valeur observée de l'**exogène** ξ ;

(b) l'**intégrale** mesure le **degré d'irrégularité** (apprécié par la **courbure** moyenne) de ρ (cf **pénalisation**) ;

(c) le **coefficient de régularisation** $\lambda > 0$ mesure un « compromis » entre qualité de l'ajustement et régularité de l'**estimateur** de ρ. En particulier, si $\lambda \to 0+$, on définit un mode d'**interpolation** de ρ.

La fonction ρ n'est, dans cette méthode, astreinte à aucune hypothèse autre que d'être de classe C^2 et tq ρ'' soit de carré intégrable.

(ii) Si l'on remplace dans (2) ρ'' par $\rho^{(p)}$ (dérivée d'ordre $p \geq 2$ de ρ, supposée de classe C^p et tq $\rho^{(p)}$ soit de carré intégrable), on définit, de façon analogue, la notion d'**estimateur spline de degré p+1** de ρ (les noeuds étant les mêmes que précédemment).