**ESTIMATEUR DE LA DENSITÉ (C5, H)**
(27 / 08 / 2020, © Monfort, Dicostat2005, 2005-2020)

Une **densité de probabilité** est un **paramètre** particulier, à caractère « fonctionnel ». Son **estimation** peut s'effectuer dans un **contexte** paramétrique ou non.

On considère un **échantillon** $X = (X_1 ,..., X_N)$ constitué de **copies** issues d'une **va** réelle (scalaire ou vectorielle) $\xi$, dont la **loi** $P^{\xi}$ admet une **densité** f pr à une **mesure** donnée $\mu$. L'objectif est l'estimation de f.

(i) Divers **estimateurs** de f ont été définis, soit à l'aide d'une méthode **paramétrique**, soit par une **méthode non paramétrique**. Les plus classiques sont les suivants :

(a) estimateur par le **noyau** : méthode de PARZEN, méthode de ROSENBLATT (cf **méthode du noyau**) ;

(b) **estimateur du maximum de vraisemblance** (avec **pénalisation** éventuelle), notamment lorsque f dépend d'un **paramètre** $\theta$ ;

(c) **estimateur par tamisage** (cf **méthode des tamis**) ;

(d) **estimateur par les fonctions orthogonales**, basé sur un développement de la densité pr à une base de **fonctions orthogonales** (ou de fonctions orthonormales) (cf **méthode des fonctions orthogonales**) ;

(e) estimateur par **fonction spline** : **estimateur spline de la densité**, ou « **histo-spline** » ;

(f) estimateur par les **moments** (cf **méthode des moments**) ;

(g) estimateur par développement en série (tronqué) : cf eg **série de CHARLIER**, **développement de CHARLIER-GRAM** ;

(h) estimateur par une **méthode des moindres carrés** ou, plus généralement, par moindre **distance** ou par une **méthode de moindre norme** (cf **estimateur à distance minimale**, **estimateur à distance minimum**).

Ainsi, lorsque f dépend d'un paramètre $\theta \in \mathbf{R}^Q$, on en construit un **histogramme** h en K classes, puis l'on estime $\theta$ par **régression** des **fréquences relatives** (empiriques) $f_k$ (avec $k = 1 ,..., K$) sur des points $x_k \in I_k$ (eg les centres des classes $I_k$) (cf **fréquence empirique**).

Par suite (en supposant que $Q < K$), le **modèle de régression** (non linéaire) :

(1) $$f_k = f(x_k , \theta) + u_k , \quad \forall k = 1 , ..., K,$$

conduit à un estimateur non strict (cf **estimateur strict**) ;

(i) estimateur par **dérivation** : on estime la **fonction de répartition** F de $\xi$ à l'aide d'une fonction $F_N^{\sim}$, supposée dérivable, puis on estime f par la dérivée $(F_N^{\sim})'$ de $F_N^{\sim}$.

(ii) Lorsque la densité f est paramétrable par $\theta \in \Theta$, avec $\Theta \subset \mathbf{R}^Q$, les méthodes généralement utilisées sont la **méthode du mv** et la **méthode des moments**.

On peut parfois estimer une **caractéristique** donnée de $P^{\xi}$ (ou de f) (eg le **mode**, un p-**quantile** ou un **moment**) de façon indirecte : on définit l'estimateur à partir de la caractéristique analogue de l'estimateur $f_N^{\sim}$ de f (ie celle de l'estimateur de $P^{\xi}$ qui s'en déduit implicitement).