APPROXIMATION D'UNE DENSITÉ (A10, C5, C10)
(26 / 08 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'expression **approximation d'une densité** est le nom général donné à toute méthode numérique visant à définir ou à calculer une fonction approchant (un représentant d') une **densité de probabilité**.

Ainsi, $\xi : \Omega \mapsto \mathbf{R}$ étant une **vars** dont la **loi** admet une **densité** pr à la **mesure de LEBESGUE** λ, on considère deux situations classiques.

(i) Le **support** Supp f = [a, b] de f est borné (compact) (cf **partie bornée**, **partie compacte**) et f est continue. On peut alors approximer f par un polynôme $P_n(x) = \Sigma_{i=0}^{n} a_i x^i$. Si l'on note :

(1) $\quad (f, g) \mapsto \int_{[a,b]} f \,.\, g \, d\lambda$

le produit scalaire défini sur l'ensemble $\mathcal{C}_{\mathbf{R}}$ ([a, b]) des fonctions continues à support compact, la **distance** d induite par ce produit scalaire est définie par :

(2) $\quad d^2 (P_n , f) = \int_{[a,b]} (P_n - f)^2 \, d\lambda.$

Cette distance est minimum lorsque le vecteur $a = (a_0 , a_1 ,..., a_n)'$ vérifie les N + 1 **équations normales** (conditions du premier ordre) :

(3) $\quad \Sigma_{i=0}^{n} a_i \int_{[a,b]} x^{i+j} \, dx = \int_a^b x^j f(x) \, dx = E\, \xi^j , \quad \forall j \in N_n .$

En posant :

(4)
$$M = (m_{ij})_{(i,j)} , \qquad \text{avec } m_{ij} = \int_a^b x^{i+j} \, dx,$$
$$b = (b_0 , b_1 ,..., b_n)', \qquad \text{avec } b_j = E\, \xi^j,$$

on obtient le **système linéaire** M a = b, dont la **solution des moindres carrés** est :

(5) $\quad a^{\wedge} = M^{-1} b.$

(ii) Le support Supp f = **R** de f n'est pas borné. On utilise alors généralement une **mesure positive** $\mu \geq 0$ sur $\mathcal{B}_{\mathbf{R}}$, absolument continue pr à λ (ie $\mu << \lambda$) (cf **mesure absolument continue**) et tq le **produit scalaire** :

(6) $\quad (f, g) \mapsto \int f \, g \, d\mu$

ait un sens.

Un raisonnement analogue au précédent conduit à considérer l'expression :

(7) $d^2 (P_n , f) = \int (P_n - f)^2 d\mu$,

dont la minimisation conduit aux équations :

(8) $\Sigma_{i=0}^{n} a_i \int x^{i+j} d\mu (x) = \int x^{j} d\mu (x), \qquad \forall\, i \in N_n$,

Ce système est encore de la forme M a = b, et sa solution (vectorielle) est notée $a^{\wedge}$.

(iii) Les **méthodes d'approximation** précédentes, formellement voisines de la **méthode des moments**, ont l'inconvénient de supposer n connu.

De plus, elles peuvent conduire à des approximations $P_n^{\wedge} (x) = \Sigma_{i=0}^{n} a_i^{\wedge} x^i$ qui peuvent prendre des valeurs négatives (cf **estimateur strict**, **méthode des fonctions orthogonales**).

D'autres méthodes de même type visent à éliminer ces difficultés. Elles font usage de théorèmes (eg **théorème de STONE-WEIERSTRASS**), ou d'approximations fondées sur la décomposition de f pr à une **base** de **fonctions orthonormales** d'un **espace de HILBERT** (cf **estimateur de CHARLIER**, **estimateur de la densité**).

(iv) Une démarche similaire peut s'appliquer à l'approximation de densités multidimensionnelles.