DISCRÉTISATION (C2, C5, C10)
(29 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

La **discrétisation** d'une **loi de probabilité** (resp de sa **densité**) consiste à la transformer en une **loi discrète** (resp en une densité discrète). La loi initiale est souvent de type « continu » (cf **loi absolument continue**, **variable continue**). Il existe deux définitions usuelles.

(i) **Première définition**. Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé** et $\xi : \Omega \mapsto \mathbf{R}$ une **vars** dont la loi P^{ξ} est supposée absolument continue pr à la **mesure de LEBESGUE** λ et dont la densité (**dérivée de NIKODYM-RADON**) est f :

(1) $f = d P^{\xi} / d\lambda$.

Soit $D \subset \mathbf{R}$ une **partie discrète** (ie au plus dénombrable) de $\mathbf{R}$ (en général, $D \subseteq \mathbf{N}$ ou $D \subseteq \mathbf{Z}$, inclusions au sens large). On pose (**restriction** de f à D) :

(2) $f^* = f_{/D}$.

On appelle alors :

(a) **(densité) discrétisée** de f la fonction $f_d : D \mapsto \mathbf{R}_+$ définie par :

(3) $f_d = f^* / \Sigma_{x \in D} f^* (x)$.

L'application $f \mapsto f_d$ est appelée **discrétisation** de f. On dit aussi que f* est la **discrétisée** de f et f_d la **discrétisée « normée »** de f.

La discrétisée f_d de f est donc une répartition de masses probabiliste qui vérifie les propriétés usuelles :

(4)
$f_d \geq 0$ (positivité),
$\Sigma_{x \in D} f_d (x) = 1$ (normalisation).

C'est donc une densité pr à la **mesure de comptage** $\nu : \mathcal{P}(D) \mapsto \mathbf{N}$ définie par :

(5) $\nu (B) = \Sigma_{x \in D} \delta_x (B), \qquad \forall B \in \mathcal{P} (D)$.

où δ_x désigne la **loi de DIRAC** placée au point x.

(b) **(loi de probabilité) discrétisée** de P^{ξ} la loi P^{ξ}_d définie par :

(6) $P^{\xi}_d = \Sigma_{x \in D} f_d (x) \delta_x$.

L'application $P^{\xi} \mapsto P^{\xi}_d$ est appelée **discrétisation** de P^{ξ}, ou encore discrétisation sur $\mathscr{P}^{\xi}$ si l'on considère une famille $\mathscr{P}^{\xi}$ de lois tq P^{ξ}. La va $\xi_d : \Omega \mapsto D$, dont la loi est P^{ξ}, est dite **(variable) discrétisée** de ξ.

On peut écrire :

(7) $\xi_d = \mathbf{1}_A . \xi$, où $A = \xi^{-1} (D) \in \mathscr{T}$.

(ii) La définition précédente s'étend directement à un **vecteur aléatoire** (ie à une va multidimensionnelle) réelle $\xi : \Omega \mapsto \mathbf{R}^K$.

Dans le cas général, si $(E, \mathscr{A})$ et $(F, \mathscr{B})$ sont des **espaces mesurables**, μ une **mesure abstraite** positive finie sur $\mathscr{A}$, $\xi : E \mapsto F$ une application $(\mathscr{A}, \mathscr{B})$-mesurable, μ^{ξ} l'image de μ par ξ, supposée absolument continue pr à une **mesure positive** ν définie sur $\mathscr{B}$, et D une partie (au plus) dénombrable de F, la discrétisation de $f = d\mu^{\xi} / d\nu$, ξ et μ^{ξ} peut s'effectuer en quatre étapes selon le même procédé :

$$f^* = f_{/D} ,$$

(8)

$$f_d = f^* / \Sigma_{x \in D} f^* (x) \qquad \text{(discrétisée normée de f)},$$

$$\mu^{\xi}_d = \Sigma_{x \in D} f_d (x) \delta_x \qquad \text{(discrétisée de } \mu^{\xi}\text{)},$$

$$\xi_d = \mathbf{1}_A \xi : \Omega \mapsto D \qquad \text{(discrétisée de } \xi, \text{ avec } A = \xi^{-1} (D) \in \mathscr{A}\text{)}.$$

(iii) **Seconde définition**. On appelle ici **discrétisation** une transformation définie comme suit. Soit $(\Omega, \mathscr{T}, P)$ un **espace probabilisé**, $\xi : \Omega \mapsto \mathbf{R}$ une **vars** (eg discrète ou continue) de loi P^{ξ} et $\Pi_{\mathbf{R}} = (I_k)_{k=1,\dots,K}$ une **partition** finie de $\mathbf{R}$, avec (cf aussi **histogramme**) :

$$I_1 =]- \infty, a_1[,$$

(12)

$$I_k = [a_{k-1} , a_k[, \qquad \forall k = 2,\dots, K-1,$$

$$I_K = [a_K, + \infty[.$$

On appelle alors **discrétisée** de ξ la **variable qualitative** $\eta : \Omega \mapsto \Pi_{\mathbf{R}}$ associée à ξ selon :

(13) $\eta (\omega) = I_k , \ \forall \omega \in \Omega \text{ tq } \xi (\omega) \in I_k , \forall k \in N_K^*$.

Autrement dit, η prend la « valeur » (ie la **modalité**) I_k lorsque ξ prend l'une des valeurs $x_k \in I_k$.

On écrit encore :

(14) $[\eta = I_k] \Rightarrow [\xi \in I_k]$, $\forall\ k \in N_K^*$.

La lp P^η de η, définie selon :

(15) $P^\eta\,(I_k) = P\,([\xi \in I_k]) = \int \mathbf{1}(I_k)\, dP$, $\forall\ k \in N_K^*$,

est aussi dite **(loi) discrétisée** de P^ξ. C'est donc une **loi qualitative**.

(iv) La définition précédente s'étend :

(a) à des partitions dénombrables de **R** ;

(b) à des partitions de $\mathbf{R}^K$ (avec $K > 1$).

Un tel procédé permet de définir une **variable qualitative** à partir d'une **variable quantitative** (cf aussi **classification**).