PROCESSUS EMPIRIQUE (F2, F3)
(29 / 07 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Soit $X = \{(\Omega, \mathcal{T}, P), (\mathcal{X}, \mathcal{B}), (X_t)_{t \in T}\}$ un **processus** en **temps** discret (avec $T = \mathbf{N}^*$) et $P^{X(t)}$ la **loi de probabilité** de X_t , $\forall t \in T$.

On appelle **loi empirique** associée à X la **loi de probabilité** définie selon :

(1) $\quad P_N = N^{-1} . \Sigma_{n=1}^N \delta (X_n), \quad \forall N \in \mathbf{N}^*.$

Le **processus empirique** associé à X est le processus $Q = (Q_t)_{t \in T}$ défini par :

(2) $\quad Q_N = N^{1/2} . (P_N - \overline{P}_N^X), \quad \forall N \in \mathbf{N}^*,$

où $\overline{P}_N^X = N^{-1} \Sigma_{n=1}^N P^{X(n)}$ désigne la **moyenne** des lois théoriques $P^{X(n)}$.

Autrement dit, Q est un processus en temps discret constitué de va Q_N qui ne sont autres que les lois empiriques (mesures aléatoires) centrées pr à $\overline{P}_N^X$ (cf aussi **mesure stochastique**).

(ii) Si $X = (X_n)_{n \in \mathbf{N}}$ désigne une **suite iid** constituée de **vars**, le **processus empirique** associé est donc défini par :

(2)' $\quad N = (N_t)_{t \in \mathbf{R}} , \qquad$ avec $\; N_t = \Sigma_{n=0}^N \mathbf{1}([X_n \leq t]),$

où $\mathbf{1}(A)$ désigne l'**indicatrice** d'une **partie** $A \subset \mathbf{R}$.

La notion est généralement utilisée lorsque X est indépendant (resp iid) (cf **suite indépendante**, **suite iid**, **théorème fondamental de la Statistique**).

(iii) Si $f : \mathcal{X} \mapsto \mathbf{R}$ est une **fonction** mesurable (cf **application mesurable**) et Q_N-intégrable (pour au moins un $N \in \mathbf{N}^*$), on montre que :

(3) $\quad E_{Q(N)} f = \int f \, dQ_N = N^{-1/2} . \Sigma_{n=1}^N \{f (X_n) - E f (X_n)\},$

où $E_{Q(N)}$ désigne l'opérateur **espérance** calculée avec Q_N .

(iii) En pratique, si $\mathcal{X} = \mathbf{R}^K$ et si l'on note F_N la **fonction de répartition empirique** associée à X (supposé iid avec F pour **fr** théorique commune), on appelle aussi **processus empirique** le processus en temps discret $G = (G_N)_{N \in \mathbf{N}^*}$ défini selon :

(4) $\quad G_N = N^{1/2} (F_N - F), \qquad \forall N \in \mathbf{N}^*.$

Le processus G est ici à valeurs dans un espace fonctionnel (celui des fr).

On peut aussi définir ponctuellement un **processus empirique** $G(x) = (G_N(x))_{N \in \mathbf{N}^*}$ où, $\forall x \in \mathcal{X}$:

(5) $G_N(x) = N^{1/2} . \{F_N(x) - F(x)\})$, $\forall N \in \mathbf{N}^*$.

Si $X = (X_n)_{n \in \mathbf{N}}$ désigne une suite iid constituée de vars, le **processus empirique** associé est aussi défini par :

(6) $N = (N_t)_{t \in \mathbf{R}}$, avec $N_t = \Sigma_{n=0}^{N} \mathbf{1}(X_n \leq t)$.