MOYENNE EMPIRIQUE (F3)
(30 / 04 / 2018)

Une **moyenne empirique** est une **statistique naturelle** mise en correspondance avec celle d'**espérance mathématique**. En effet, cette expression concerne, le plus souvent, implicitement, la **moyenne arithmétique**.

(i) Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé** et $\xi : \Omega \mapsto \mathbf{R}$ une **vars** de **loi** P^{ξ}. Soit $X = (X_1, ..., X_N)$ un N-**échantillon** issu de la **variable parente** ξ.

On appelle **moyenne empirique** (**simple**) la **va** (ou **statistique**) suivante :

(1) $\quad \overline{X}_N = N^{-1} \Sigma_{n=1}^{N} X_n = e_N' X / e_N' e_N \quad$ (en notant X en vecteur colonne).

Autrement dit, $\overline{X}_N$ s'interprète comme l'**espérance** de ξ calculée à l'aide de la **loi empirique** P_N associée à X, ie :

(2) $\quad \overline{X}_N = E_N \xi = \int \xi \, dP_N$.

(ii) Si $\lambda \in S_N$ (**simplexe** de $\mathbf{R}^N$) (cf **fonction de poids**), on appelle **moyenne empirique pondérée** (par λ) la va ou statistique suivante, définie comme **combinaison linéaire convexe** des coordonnées de X :

(3) $\quad \overline{X}_N(\lambda) = \Sigma_{n=1}^{N} \lambda_n X_n = \lambda' X$.

(iii) On montre que :

(a) si $\lambda = e_N / e_N' e_N = (1 / N, ..., 1 / N)'$, alors $\overline{X}_N(\lambda) = \overline{X}_N$ (moyenne empirique « ordinaire ») ;

(b) si $\xi \in L_{\mathbf{R}}^1(\Omega, \mathcal{T}, P)$, on a $E \overline{X}_N(\lambda) = \lambda' E X = \lambda' (E \xi) e_N = (E \xi) \lambda' e_N = E \xi = \mu_1$ (**moment** théorique simple d'ordre 1, ie espérance, de ξ) ;

(c) si X est un **échantillon iid** selon P^{ξ} et de carré intégrable, si μ désigne l'espérance de ξ et σ^2 sa **variance**, on a $V \overline{X}_N = N^{-1} \sigma^2$ et $\overline{X}_N$ est asymptotiquement gaussienne (cf **normalité asymptotique**, **théorème de la limite centrale**), ie :

(4) $\quad \mathcal{L}\{N^{1/2} . \sigma^{-1} . (\overline{X}_N - \mu)\} \to_{N \to +\infty} \mathcal{N}(0, 1)$ (**loi normale réduite**).

Si, de plus, $\xi \sim \mathcal{N}(\mu, \sigma^2)$ (**loi normale**), alors (cas où σ^2 est « inconnue ») :

(5) $\quad \mathcal{L}\{N^{1/2} . (S_N^*)^{-1} . (\overline{X}_N - \mu)\} = \mathcal{S}_{N-1} \quad$ (**loi de STUDENT** à N - 1 **degrés de liberté**),

où S_N^* désigne l'écart-type empirique associé à la **variance empirique** « corrigée » $(S_N^*)^2 = (N - 1) \Sigma_{n=1}^N (X_n - \bar{X}_N)^2$;

(d) si $X^i = (X_{i1}, \dots, X_{i,N(i)})$ est un échantillon iid comme la vars ξ_i (avec $i \in N_2^*$), si X_1 est indépendant de X_2 et si $\xi_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, alors la statistique (**écart**) :

(6) $\quad D_{N(1),N(2)} = \bar{X}^1 - \bar{X}^2$

vérifie :

(7) $\quad (D_{N(1),N(2)} - \delta_{12}) / \sigma_{12} \sim \mathcal{N}(0, 1)$,

où $\bar{X}_i$ est la moyenne empirique associée à X^i , S_i^2 sa variance empirique brute (ie non corrigée), $\delta_{12} = \mu_1 - \mu_2$, et $\sigma_{12}^2 = (\sigma_1^2 / N_1) + (\sigma_2^2 / N_2)$ (variance « combinée »). Ce résultat sert à divers **tests** portant sur les différences entre moyennes δ_{12} .

(iii) Si $\xi : \Omega \mapsto \mathbf{R}^K$ est un **vecteur aléatoire** et X un échantillon iid selon ξ, la moyenne empirique (vectorielle) de ξ (ou de X) est définie par le vecteur des moyennes empiriques (scalaires) des coordonnées de ξ, ie :

(8) $\quad \bar{X}_N = N^{-1} \Sigma_{n=1}^N X_n = e_N' X / e_N' e_N$,

X désignant ici la (N,K)-**matrice** $(X_1' /// X_N')$, dont les lignes sont $X_n' = (X_{n1}, \dots, X_{nK})$, $\forall n \in N_N^*$, et par $e_N = (1, \dots, 1)'$ le premier vecteur bissecteur de $\mathbf{R}^N$. On note souvent $\bar{X}_N$ sous la forme :

(9) $\quad \bar{X}_N = (\bar{X}_1, \dots, \bar{X}_k)$, $\quad$ avec $\bar{X}_k = N^{-1} \Sigma_{n=1}^N X_{nk}$, $\quad \forall k \in N_K^*$.

(iv) On montre que :

(a) si $\xi \in L_{\mathbf{R}K}^1 (\Omega, \mathcal{T}, P)$, $\bar{X}_N$ est un **estimateur sans biais** de $E\,\xi$ (moyenne théorique de ξ), ie :

(10) $\quad E\,\bar{X}_N = E\,\xi$;

(b) si $\xi \in L_{\mathbf{R}K}^2 (\Omega, \mathcal{T}, P)$ et si X est un échantillon iid selon P^ξ, alors (en notant $V\,\xi$ la **matrice de covariance** théorique de ξ) :

(11) $\quad V\,\bar{X}_N = N^{-1} . V\,\xi$;

(c) si $\xi \sim \mathcal{N}_K(\mu, \Sigma)$ (**loi normale multidimensionnelle**) et si X est iid selon P^ξ, on a :

$$\bar{X}_N \sim \mathcal{N}_K(\mu, N^{-1}\,\Sigma),$$

(12)

$$N\,.\,(\bar{X}_N - \mu)'\,\Sigma^{-1}\,(\bar{X}_N - \mu) \sim \mathcal{X}_K^2$$ (**loi du chi-deux** à K **dl**).