**ÉTENDUE (C5, F3, F6)**
(07 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'**étendue** désigne une **caractéristique** de **dispersion** simple qui peut concerner :

(a) soit une **va** ou sa **lp** (étendue théorique). La notion peut se rapprocher de la notion de support (cf **support d'une probabilité**, **support d'une fonction**) ;

(b) soit un échantillon engendré par une telle va (étendue empirique), et notamment un **échantillon iid**. Dans ce cas, la notion peut se rapprocher de celle de **valeur extrême**.

(i) Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé**, $\xi : \Omega \mapsto \mathbf{R}$ une **vars** dont la **loi de probabilité** est $P^{\xi}$, f la **densité de probabilité** de $P^{\xi}$ pr à $\lambda_1$ et Supp f = I $\in \mathcal{I}(\mathbf{R})$ (intervalles de **R**) le **support** de f.

On appelle **étendue (théorique)** de $\xi$ (ou de $P^{\xi}$, ou encore de f) le nombre :

(1) $\quad E = \sup I - \inf I,$

qui peut être infini, et qui s'interprète comme un **paramètre** de **dispersion**.

Si I est borné, eg si I = [a, b], avec $a < \infty$ et $b < \infty$, l'étendue est E = b - a, la **semi-étendue** (théorique) étant alors définie par E' = (b - a) / 2.

(ii) Soit $X = (X_1,..., X_N) : \Omega \mapsto \mathbf{R}^N$ un **vecteur aléatoire** (**échantillon**) donné et $X^{(.)} = (X^{(1)},..., X^{(N)})$ sa **statistique d'ordre**, définie par :

(2) $$\begin{cases} X^{(1)} \leq X^{(2)} \leq ... \leq X^{(N-1)} \leq X^{(N)} \\ \{X^{(1)},..., X^{(N)}\} = \{X_1,..., X_N\}. \end{cases}$$

On appelle **étendue**, ou **intervalle de variabilité**, **(empirique)** de X la **statistique** :

(3) $\quad E_N = X^{(N)} - X^{(1)} = \max_{n=1}^{N} X_n - \min_{n=1}^{N} X_n ,$

aussi notée $R_N$ (de l'anglais « range »), $U_N$ ou simplement U.

L'étendue empirique $E_N$ est un estimateur naturel de l'étendue théorique (cf **statistique naturelle**).

On appelle **semi-étendue**, ou parfois **étendue moyenne**, (empirique) la statistique :

(4) $\quad E_N' = E_N / 2 = (X^{(N)} - X^{(1)}) / 2.$

On appelle **étendue réduite** (empirique) la statistique :

(5) $\quad E_N'' = E_N / S_N ,$

où $S_N^2 = N^{-1} \Sigma_n (X_n - \bar{X}_N)^2$ (**variance empirique** naturelle) et $\bar{X}_N = N^{-1} \Sigma_{n=1}^N X_n$ (**moyenne empirique**). Dans un **problème d'estimation**, la variance empirique est parfois remplacée par la variance corrigée $(N-1)^{-1} N S_N^2$.

(iii) Si $X = (X_1 ,..., X_N)$ est un échantillon iid comme la **variable parente** $\xi$ précédente et que Supp f = [a, b], un estimateur simple du paramètre $(a, b) \in \mathbf{R}^2_{\leq}$ est :

(6) $\quad (a_N^{\tilde{}}, b_N^{\tilde{}}) = (X^{(1)}, X^{(N)})$,

et l'étendue théorique E = b - a peut alors être estimée par la statistique :

(7) $\quad E_N = b_N^{\tilde{}} - a_N^{\tilde{}}$.

(iv) Lorsque X contient des **aberrations** à ses deux extrêmités, ie $X^{(1)} ,..., X^{(L)}$ et $X^{(M)} ,..., X^{(N)}$ (avec $1 \leq L \leq N-1$ et $2 \leq M \leq N$), l'estimateur modifié $(a_N^{\#}, b_N^{\#}) = (X^{(L+1)}, X^{(M-1)})$ (avec L < M) définit l'**étendue réelle** (empirique) $E_N = b_N^{\#} - a_N^{\#}$.

(ii) On montre que :

(a) si $\xi : \Omega \mapsto \mathbf{R}$ est la variable parente de loi $P^{\xi}$ qui engendre X (échantillon iid), f la densité de probabilité de $P^{\xi}$ pr à $\lambda_1$ et F la **fr** de $P^{\xi}$. On note $\mathscr{L}(E_N)$ la **loi** de l'étendue $E_N$ . La densité h de $\mathscr{L}(E_N)$ pr à $\lambda_1$ s'explicite alors selon :

(8) $\quad h(u) = N(N-1) . \int_{\mathbf{R}} \{F(z+u) - F(z)\}^{N-2} f(z+u) f(z) dz, \qquad \forall u \in \mathbf{R}$ ;

(b) si $X \sim \mathscr{N}_N (\mu, \sigma^2 . I_N)$ (**loi normale multidimensionnelle** à dispersion « scalaire »), on utilise aussi un autre **estimateur** de $\sigma^2$ que $S_N^2$ associé à (5). Cet estimateur $T_N$ est indépendant de $E_N$ et vérifie :

(9) $\quad \mathscr{L}(\nu T_N / \sigma^2) = \mathscr{X}_\nu^2$ (**loi du chi-deux** à $\nu$ **degrés de liberté**).

La loi de la statistique $E_N''$ définie en (5) est alors appelée **loi de l'étendue réduite**.

(iii) On appelle :

(a) **étendue d'ordre L**, ou **L-étendue**, (empirique) la statistique :

(10) $\quad E_N(L) = X^{(N-L+1)} - X^{(L)}, \qquad \forall L = 1 ,..., N$ ;

(b) **quasi-étendue d'ordre L** (empirique) la statistique :

(11) $\quad E_N'(L) = X^{(N-L)} - X^{(L+1)}, \qquad \forall L = 0, 1 ,..., N-1$.

On établit les relations élémentaires suivantes :

(12) $\quad E_N(1) = E_N'(0) = E_N$ et $E_N(L) = - E_N(N-L+1)$.