COEFFICIENT DE VARIATION (C5, F3)
(17 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

La notion courante (eg en physique) d'**erreur relative** (ie de **pourcentage d'erreur**) portant sur une variable ξ s'exprime généralement par une grandeur sans dimension tq Δx / x. Le concept, probabiliste et statistique, de **coefficient de variation** généralise cette notion.

(i) Soit $\xi \in \mathscr{L}_{\mathscr{X}}^2 (\Omega, \mathscr{T}, P)$ une **va** donnée, dont l'**espérance** est $E\ \xi \neq 0$ et la **variance** est $(\sigma(\xi))^2 = V\ \xi$. On suppose que l'ensemble d'observation $\mathscr{X} \in \{N_n , \mathbf{N} , \mathbf{R}_+ , \mathbf{Z} , \mathbf{R}\}$.

On appelle **coefficient de variation (théorique)** de ξ le rapport (nombre ou scalaire sans dimension) :

(1) $\gamma(\xi)$ ou $\gamma_\xi = \sigma(\xi) / |E\ \xi|$.

Ce rapport est aussi noté $CV(\xi)$ ou CV_ξ , etc.

On l'appelle aussi **coefficient de K. PEARSON**, ou **rapport du bruit au signal**, ou **écart-type relatif**, ou encore (après multiplication par 100) **pourcentage d'erreur**, etc.

Le nombre γ_ξ^2 est appelé **variance relative** de ξ, et le nombre γ_ξ^{-1} **rapport du signal au bruit**.

Le plus souvent, on suppose que ξ est une variable positive. L'**égalité de KOENIG** s'écrit sous la forme (cf **formule de KOENIG-HUYGENS**) :

(2) $E\ \xi^2 = (\gamma_\xi^2 - 1) . E^2\ \xi$.

La forme élémentaire de l'**inégalité de BIENAYMÉ-CHEBICHEV** :

(3) $P\ \{|(\xi - E\ \xi) / E\ \xi| \leq \lambda\} \geq 1 - \{\sigma(\xi) / (\lambda\ E\ \xi)\}^2, \quad \forall\ \lambda > 0$,

peut se réécrire sous la forme :

(3)' $P\ ([|\varepsilon| \leq \lambda]) \geq 1 - (\gamma_\xi / \lambda)^2$,

avec $\varepsilon = (\xi - E\ \xi) / E\ \xi$ (**variable normée** ou normalisée).

(ii) Si $X = (X_1 ,..., X_N)'$ est un N-**échantillon aléatoire** constitué de N **copies** de ξ indépendantes entre elles, alors, en posant :

$$T_N = e_N'\ X = \Sigma_{n=1}^N X_n \quad \text{(total empirique)},$$

(4)

$$\overline{X}_N = e_N'\ X / e_N'\ e_N \quad \text{(\textbf{moyenne empirique})},$$

on montre que :

(5) $\gamma (T_N) = \gamma (\bar{X}_N) = N^{-1/2} . \gamma (\xi)$.

Dans le même cadre que le précédent, si l'on pose :

(6) $S_N^2 = N^{-1} \Sigma_{n=1}^N (X_n - \bar{X}_N)^2 = X' P X / e_N' e_N$ (**variance empirique**),

on appelle **coefficient de variation empirique** la **statistique** (sans dimension) :

(7) $c_N (\xi)$ ou $c_N (X) = S_N / | \bar{X}_N |$.

Si $N \geq 2$ et si $X_n \geq 0$, $\forall n \in N_N^*$, ce nombre vérifie :

(8) $c_N (X) \leq (N - 1)^{1/2}$.

(iii) Il existe diverses généralisations des notions précédentes :

(a) par changement des concepts de **centralité** ou de **dispersion** (cf **tendance centrale**, **valeur centrale**, **paramètre de position**, **paramètre d'échelle**).

Si ξ admet pour paramètre de centralité (resp de position) $C \xi$ (resp α) et pour paramètre de dispersion (resp d'échelle) $D \xi \geq 0$ (resp $\beta \geq 0$), on appelle alors **coefficient de variation généralisé (théorique)** le nombre (sans dimension) :

(9) $\gamma' (\xi) = D \xi / |C \xi|$ (si $C \xi \neq 0$) (resp $\beta / |\alpha|$ (si $\alpha \neq 0$)) ;

(b) par changement de dimensionnalité. On suppose que ξ est un **vecteur aléatoire** à valeurs dans l'**espace vectoriel mesurable** $\mathscr{X}^K$, ou une **va** à valeurs dans un espace produit $\mathscr{X}^K$. On appelle alors **coefficient de variation généralisé (théorique)** de ξ :

(b_1) soit le vecteur (sans dimension) :

(10) $\gamma (\xi) = (\gamma_1 (\xi_1) ,..., \gamma_K (\xi_K))'$,

ou la **matrice diagonale** associée (cf **matricialisation**), avec $\gamma_k (\xi_k) = \sigma (\xi_k) / |E \xi_k|$, $\forall k \in N_K^*$;

(b_2) soit le scalaire réel positif sans dimension γ_ξ tq :

(11) $\gamma_\xi^2 = \text{Dét} (V \xi) / ||E \xi||^2$;

(b_3) soit même la matrice (sans dimension) :

(12) $\Gamma (\xi) = ((|E \xi|)^\mu)^{-1} . V \xi$,

expression dans laquelle $V\,\xi$ est supposée être une **matrice définie positive**, $|E\,\xi| = (|E\,\xi_1|, ..., |E\,\xi_K|)'$ et $(v)^\mu$ désigne le matricialisé du vecteur $v \in \mathbf{R}^K$ (cf **matricialisation**).

(c) en particulier, si $\mathscr{X} = \mathbf{R}$ (ie $\xi : \Omega \mapsto \mathbf{R}^K$) et si $\xi \in L_{\mathbf{R}^K}^2$, on définit le coefficient :

(13) $\gamma_\xi^2 = (\mu' \Sigma^{-1} \mu)^{-1}$, où $\mu = E\,\xi$ et $\Sigma = V\,\xi$,

aussi appelé **coefficient de variation généralisé**.