

AUTOCOVARIANCE (C5, F3, N)

(21 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

La notion d'**autocovariance** est définie comme la **covariance** entre **va** constituant un **processus stochastique** ou une **série temporelle**. On en déduit celle d'**autocorrélation** du processus ou de la série.

Lorsque plusieurs processus (ou séries) sont en relation, on étend ce concept à celui de **covariance croisée retardée**.

(i) Soit $\{(\Omega, \mathcal{F}, P), (\mathcal{X}, \mathcal{B}), (X_t)_{t \in T}\}$ un processus stochastique tq :

(a) $(\mathcal{X}, \mathcal{B}) = (\mathbf{R}, \mathcal{B}_{\mathbf{R}})$ (processus réel scalaire) ;

(b) $(T, +)$ est un **groupe** additif abélien (eg $T = \mathbf{Z}$, $T = \mathbf{Q}$, ou $T = \mathbf{R}$) ;

(c) X est de carré intégrable, ie $X_t \in L^2_{\mathbf{R}}(\Omega, \mathcal{F}, P), \forall t \in T$.

On appelle alors (**coefficient d' autocovariance** (théorique) de type (s, t) de X le nombre réel suivant :

$$(1) \quad \gamma_{st} = C(X_s, X_t) = E(X_s - \mu_s)(X_t - \mu_t), \quad \forall (s, t) \in T^2,$$

où $\mu_t = E X_t$ (**espérance** de X_t), $\forall t \in T$.

La fonction réelle $\gamma : T^2 \mapsto \mathbf{R}$ définie par :

$$(2) \quad \gamma(s, t) = \gamma_{st}, \forall (s, t) \in T^2,$$

est appelée (**fonction d' autocovariance** (théorique) de X . On note aussi $\sigma_t^2 = \gamma_{tt}$ la variance (« instantanée ») de X_t , $\forall t \in T$.

Les coefficients et fonction d'autocovariance précédents sont théoriques, ie définis à partir du processus X et de sa loi $P^X = X(P)$ (image de P par X). On calcule, en effet, l'espérance μ_t de X_t à l'aide de la **loi marginale** de X_t :

$$(3) \quad \mu_t = \int X_t dP = \int x dP^{X_t}(x), \quad \forall t \in T,$$

et chaque coefficient γ_{st} à l'aide de la loi marginale du **couple aléatoire** (X_s, X_t) :

$$(4) \quad \gamma_{st} = \int (X_s - \mu_s)(X_t - \mu_t) dP = \int (x - \mu_s)(y - \mu_t) dP^{(X_s, X_t)}(x, y).$$

(ii) Pour estimer les fonctions $t \mapsto \mu_t$ (**moyenne** « temporelle » théorique) et $(s, t) \mapsto \gamma_{st}$ (autocovariance « temporelle » théorique), on doit distinguer deux situations :

(a) soit il est possible d'observer plusieurs **séries temporelles** issues de X . Ceci est le cas eg d'**expériences** séquentiellement répétables, ou du suivi temporel de plusieurs **unités statistiques**, etc.

Si l'on note $x_i = (x_{it})_{t=1, \dots, T}$ ces séries, en nombre I (ie $i \in N_I^*$), alors :

(a)₁ la moyenne temporelle théorique $t \mapsto \mu_t$ peut être estimée (aux instants $t = 1, \dots, T$) à l'aide de la suite :

$$(5) \quad \bar{x}(t) = I^{-1} \sum_{i=1}^I x_{it}, \quad \forall t \in N_T^* ;$$

(a)₂ l'autocovariance temporelle théorique $(s, t) \mapsto \gamma_{st}$ peut être estimée à l'aide de la suite double :

$$(6) \quad c(s, t) = I^{-1} \sum_{i=1}^I (x_{is} - \bar{x}(s)) (x_{it} - \bar{x}(t)), \quad \forall (s, t) \in (N_T^*)^2.$$

(b) lorsqu'on ne dispose que d'une seule série temporelle (**trajectoire** partielle du processus) (eg expérience non répétable, données temporelles à caractère « historique » ou « non contrôlable »), la situation est plus complexe :

(b)₁ si le processus générateur de la série considérée est **stationnaire** (ou **ergodique**) (ce qui est une hypothèse à tester au préalable), il est encore possible d'effectuer les estimations précédentes ;

(b)₂ à défaut, on peut (1) soit transformer le processus (dans l'**espace des états**) pour obtenir un nouveau processus (quasi-)stationnaire, (2) soit, parfois, formuler une hypothèse bayésienne pour pallier l'unicité de la série (cf **principe bayésien**).

(iii) Les notions précédentes sont souvent utilisées lorsque X est un **processus stationnaire en covariance** (**stationnarité** au second ordre). Dans ce cas, on a :

$$(7) \quad \begin{aligned} \mu_t &= \mu \text{ (moyenne indépendante de } t), \quad \forall t \in T, \\ \gamma_{st} &= C(X_s, X_t) = C(X_s, X_{s+(t-s)}) = \gamma(t-s), \quad \forall (s, t) \in T^2. \end{aligned}$$

Autrement dit, la covariance γ_{st} ne dépend que de la différence $t - s$, et l'on note encore γ la fonction qui associe à $t - s$ la valeur γ_{st} .

Si l'on pose $\theta = t - s \in T$, les relations de définition (5) deviennent :

$$(8) \quad \gamma_\theta = C(X_s, X_{s+\theta}) = E(X_s - \mu)(X_{s+\theta} - \mu) = \gamma(\theta), \quad \forall s \in T.$$

Les coefficients d'autocovariance $\gamma(\theta)$ ($\theta \in T$) définissent ainsi la **fonction d'autocovariance** $\gamma : T \mapsto \mathbf{R}$ d'un processus stationnaire au sens large X .

Par suite, on peut définir les **notions analogues « empiriques »** à partir d'une série temporelle $x = (x_t)_{t=1, \dots, T}$ issue de X et observée en **temps** discret. On pose alors :

$$(9) \quad \bar{x}_T = T^{-1} \sum_{t=1}^T x_t \quad (\text{moyenne empirique temporelle}),$$

$$r_\theta = (T - \theta)^{-1} \sum_{t=1}^{T-\theta} (x_{t+\theta} - \bar{x}_T) (x_t - \bar{x}_T), \quad \forall \theta \in N_{T-1}.$$

La suite des coefficients d'autocovariance empiriques $(r_\theta)_\theta$ ainsi calculée définit la **(fonction d') autocovariance empirique** $r : N_{T-1} \mapsto \mathbf{R}$ tq $r(\theta) = r_\theta$ ($\forall \theta \in N_{T-1}$).

La suite $(r_\theta)_\theta$ permet de définir les coefficients d'autocorrélation et la fonction d'autocorrélation empirique, de la même façon que la famille $(\gamma_\theta)_\theta$ permet de définir les coefficients d'autocorrélation et la fonction d'autocorrélation théoriques (cf **autocorrélation**). Les notions empiriques constituent donc des estimateurs « naturels » de ces dernières (cf **statistique naturelle**). On en dérive aussi la notion de **spectre** : en effet, par définition, la **densité spectrale** de X n'est autre que la **transformée de FOURIER** de sa fonction d'autocovariance.

(iv) Les concepts de coefficients et de fonction d'autocovariance peuvent être étendus à des processus plus généraux que les processus scalaires réels précédents. Ainsi :

(a) si X est scalaire complexe, ie si $(\mathcal{X}, \mathcal{B}) = (\mathbf{C}, \mathcal{B}_{\mathbf{C}})$, on remplace (1) par :

$$(1)' \quad \gamma_{st} = C(X_s, \bar{X}_t) = E(X_s - \mu_s) (\bar{X}_t - \bar{\mu}_t), \quad \forall (s, t) \in T^2,$$

où \bar{z} désigne le nombre complexe conjugué de $z \in \mathbf{C}$ et $\bar{\mu}_t = E \bar{X}_t, \forall t$;

(b) si X est vectoriel réel, ie si $(\mathcal{X}, \mathcal{B}) = (\mathbf{R}^K, \mathcal{B}(\mathbf{R}^K))$, on remplace (1) par :

$$(1)'' \quad \gamma_{st} = C(X_s, X_t) = E(X_s - \mu_s) (X_t - \mu_t)', \quad \forall (s, t) \in T^2,$$

ce qui définit des « coefficients » **d'autocorrélation matriciels** ainsi que la **(fonction d') autocovariance matricielle** associée ;

(c) si X est vectoriel complexe, ie si $(\mathcal{X}, \mathcal{B}) = (\mathbf{C}^K, \mathcal{B}(\mathbf{C}^K))$, on combine les deux définitions précédentes. De telles généralisations peuvent être définies dès lors que la notion de covariance (donc d'espérance mathématique) l'est, eg pour des processus à valeurs dans un **espace de BANACH** (cf **opérateur de covariance**).