

DÉSAISONNALISATION (N9)

(04 / 09 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'étude d'un **phénomène** temporel peut nécessiter la décomposition d'un **processus stochastique** qui le décrit, ou d'une **série temporelle** $x = (x_t)_{t \in T}$ associée à ce processus, en diverses composantes (cf **composante d'une série temporelle**).

Dans certains **domaines de connaissance** (eg écologie ; sociologie : économie), la décomposition certaines composantes parmi les suivantes :

(a) une **composante tendancielle** $m = (m_t)_{t \in T}$ (cf **tendance**) ;

(b) une **composante cyclique** $c = (c_t)_{t \in T}$ (cf **cycle**) ;

(c) une **composante saisonnière** $s = (s_t)_{t \in T}$ (cf **saisonnalité**) ;

(d) une **composante résiduelle** aléatoire $u = (u_t)_{t \in T}$ (cf **perturbation aléatoire**).

(i) Selon la nature du phénomène, ou la longueur de la **période d'observation**, la signification concrète des composantes peut varier. Dans certains cas (eg physique : météorologie, sociologie), la composante saisonnière constitue un élément important de la décomposition de x .

La décomposition suit, de façon générale, une **modélisation** de la forme :

$$(1) \quad x_t = \psi_t(m_t, c_t, s_t, u_t), \quad \forall t \in T.$$

Ainsi :

(a) dans la **décomposition additive**, on a :

$$(1)_a \quad x_t = m_t + c_t + s_t + u_t,$$

et l'**amplitude** de la **composante saisonnière** s_t ne dépend pas du niveau $m_t + c_t$;

(b) dans la **décomposition multiplicative**, on pose généralement :

$$(1)_b \quad x_t = (m_t + c_t) \cdot s_t \cdot u_t$$

ou encore :

$$(1)_b \quad x_t = (m_t + c_t) \cdot s_t \cdot (1 + u_t).$$

L'amplitude de s_t est alors proportionnelle au niveau $m_t + c_t$;

(c) il existe aussi des décompositions de type puissance :

$$(1)_c \quad x_t = m_t^\alpha \cdot c_t^\beta \cdot s_t^\gamma \cdot (1 + u_t)$$

ou de type exponentiel :

$$(1)_c \quad x_t = \alpha^{m(t)} \cdot \beta^{c(t)} \cdot \gamma^{s(t)} \cdot (1 + u_t),$$

où $m(t)$, $c(t)$ et $s(t)$ désignent resp, pour simplifier, m_t , c_t et s_t .

(ii) On appelle **méthode de désaisonnalisation** une méthode consistant :

(a) soit à estimer s ;

(b) soit, alternativement, à estimer x après en avoir « extrait » s .

Si x est observée en **temps** discret (eg si $T = N_T^* = \{1, \dots, T\}$), et si T comprend H **périodes** saisonnières (eg H mois ou H trimestres) et I parties (eg I années) de longueur H , avec $T = H \cdot I$, on pose $x_t = x_{hi}$, avec $t = (i - 1) \cdot H + h$, et s est supposée strictement périodique de période H , ie :

$$(2) \quad s_t = s_{t+H}, \quad \text{avec } t = (i - 1) \cdot H + h, \quad \forall t \in T,$$

et indépendante des composantes m et c .

Pour tout t , on définit le **coefficient saisonnier** relatif à l'instant t comme le paramètre :

$$(3) \quad s_t = \sigma_h,$$

où σ_h est indépendant de $i \in N_I^*$.

On appelle parfois **table de C.H.D. BUYS-BALLOT** le **tableau statistique**, basé sur la (H, I) -matrice $X = (x_{hi})_{(h,i)}$, qui permet le calcul des coefficients saisonniers.

(iii) Il existe deux principales familles de méthodes de désaisonnalisation :

(a) les méthodes parfois dites « empiriques » (ou descriptives), ie sans **modèle** « statistique » : **méthode des moyennes périodiques**, **méthode des chaînes de rapports**, **méthode des rapports à la tendance** ;

(b) les méthodes proprement « statistiques », basées sur deux principes :

(b)₁ l'un est la connaissance a priori de la **périodicité** de la composante saisonnière ;

(b)₂ l'autre (**modélisation**) est le choix a priori (**spécification**) d'un **modèle** formalisant la **saisonnalité** s , ou même formalisant l'ensemble des termes de la décomposition de x (y compris la saisonnalité s). Ainsi, une décomposition tq :

$$m_t = \alpha \cdot e^{\beta t} \quad (\text{tendance exponentielle}),$$

$$(4) \quad c_t = \sum_k a_k \cdot \sin(\omega_k t + \varphi_k) \quad (\text{cycle sinusoidal}),$$

$$s_t = \sigma_h,$$

conduit à estimer le **modèle de régression** (non linéaire) :

$$(5) \quad x_t = \alpha \cdot e^{\beta t} + \sum_k a_k \cdot \sin(\omega_k t + \varphi_k) + \sigma_h + u_t,$$

qui comporte $2 + 3K + H$ paramètres, avec $E u = 0$, $V u = \sigma^2 \cdot I_T$, sous les contraintes $\sum_h \sigma_h = 0$. En pratique, un **estimation** correcte suppose que $2 + 3K + H \ll T$.

(iv) Le nombre H de « saisons » d'un **processus** ou d'une série constitue généralement une **donnée** dans un problème de désaisonnalisation. La composante saisonnière est souvent une composante périodique importante de ce processus ou de cette série.

Dans la décomposition d'un processus ou d'une série, d'autres composantes, non apparentes, peuvent aussi intervenir : elles jouent alors le rôle d'**inconnues** (cf **analyse harmonique**, **analyse spectrale**).