SÉRIE TEMPORELLE (N)
(22 / 03 / 2020, © Monfort, Dicostat2005, 2005-2020)

La notion de **série temporelle** intervient dans tous les **domaines de connaissance**. Elle se réfère notamment aux notions d'**ordre** total et d'**irréversibilité** (eg données « historiques ») qui interviennent dans de nombreux **phénomènes**.

A la différence de la notion de **coupe instantanée**, celle de série temporelle permet :

(a) d'étudier la « dynamique » des **phénomènes** relevant de ces **sciences**, ainsi que leur déroulement dans le temps ;

(b) de préciser la notion de **causalité** et ses rapports avec celle d'**antériorité temporelle** ;

(c) d'interpréter concrètement le concept de « **prévision** », etc.

(i) Une **série temporelle** $x = (x_t)_{t \in T}$ est simplement une « réalisation » d'un **processus stochastique** X, lequel est alors appelé **processus générateur**, ou **processus sous-jacent**, à x (cf aussi **processus générateur de données**). Elle relève donc de la même théorie (même terminologie). Elle peut se concevoir aussi bien :

(a) (**origine des données**) dans le cadre d'une **expérience aléatoire** que dans celui d'un **sondage** (suivis au cours du temps) ;

(b) (**répétitivité**) avec des données répétitives qu'avec des données non répétitives (unicité de l'histoire) (cf **répétition**). Dans ce cas, il n'existe pas (sauf hypothèses supplémentaires) de véritable expérimentation. On ne peut donc procéder à des **expériences** renouvelables dans des condition identiques et l'on ne dispose ainsi que d'une seule « observation » $x = (x_t)_{t \in T}$ de X. Dans ces situations, les **procédures** d'**inférence statistique** font souvent usage de la propriété d'**ergodicité**, qui pallie, sous certaines hypothèses, l'inconvénient lié à cette unicité de l'histoire. Ces hypothèses doivent alors faire l'objet de tests appropriés ;

(c) (**observabilité**) avec des données observables ou non (**facteurs** cachés) (cf aussi **observable**) ;

(d) (**contrôlabilité**) avec des données « contrôlables » (expérience déjà réalisée et en cours d'examen, ou expérience à réaliser séquentiellement et à planifier), partiellement contrôlées, ou totalement non contrôlées.

(ii) Soit $X = \{(\Omega, \mathscr{T}, P), (\mathscr{X}, \mathscr{B}), (X_t)_{t \in T}\}$ un **processus** dont l'ensemble du (des) « **temps** » T est totalement ordonné par une **relation d'ordre** $\leq$. Cet ensemble exprime une « histoire ».

On appelle **série temporelle**, ou **série chronologique**, ou encore **séquence temporelle** ou **chronique temporelle**, toute « **trajectoire** » :

(1) $t \in T \mapsto X_t(\omega) = x_t$

de X.

Cette trajectoire est donc, par définition, associée à une « réalisation » $\omega \in \Omega$ (**événement** élémentaire) de X. On la note donc souvent $x = (x_t)_{t \in T}$.

(a) lorsque t est un **indice** ponctuel (élément de T), on dit que t est un **instant d'observation**. Dans ce cas, x peut représenter la mesure d'un **stock**, ou d'un **en-cours**, observé à un instant t donné ;

(b) lorsque t est un **indice** ensembliste $[t, t + \Delta t]$ (parties, ordonnées entre elles, de T), on dit que t est une **période d'observation**. Dans ce cas, x peut représenter un **flux**, ou une **transaction**, observé(e) pendant une période de temps t donnée.

T étant totalement ordonné, on dit que la valeur de X_t (observable ou non-observable selon x_t) a lieu **avant** l'instant (ou la période) t_0 , ou **après** cet instant (ou cette période), selon que $t < t_0$ ou que $t_0 < t$.

(iii) En pratique, on classe les séries temporelles en fonction de divers **critères** :

(a) selon le **type de temps** T propre à la série :

$(a)_1$ une **série en temps discret** est tq T est un ensemble (au plus) dénombrable, souvent un ensemble fini.

Par exemple, $T = \{0, 1,..., T\} = N_T$ ou $T = \{1,..., T\} = N_T^*$. On peut souvent noter, sans inconvénient, à l'aide du même symbole T, aussi bien l'ensemble des temps (ie l'ensemble des « indices » de X ou de x) que sa borne supérieure $\sup T = \max_{t \in T} t$.

Dans d'autres situations, $T \subset \mathbf{N}$, ou $T \subset \mathbf{N}^*$, ou encore $T \subset \mathbf{Z}$.

On parle aussi de chronique en temps discret (fini si T est fini) : eg (économie) produit intérieur brut (PIB ou GDP) annuel ou trimestriel, ou encore indice des prix à la consommation (IPC) trimestriel ou mensuel, etc.

$(a)_2$ une **série en temps continu** est tq T est un partie de **R**, souvent une partie bornée.

Par exemple, $T = |a, b|$ ou $T = |0, T| = R_T$ (intervalles de **R**). On peut, ici aussi, noter à l'aide du même symbole T, aussi bien l'ensemble des temps (ie l'ensemble des « indices » de X ou de x) que sa borne supérieure $\sup T = \sup_{t \in T} t$.

Dans d'autres situations, $T \subset \mathbf{R}_+$, ou $T \subset \mathbf{R}_+^*$, ou encore $T \subset \mathbf{R}$.

On parle alors de chronique en temps continu (borné si T est borné) : eg (physique : données météorologiques), ou encore (biologie) électro-cardiogramme (ECG) ou électro-encéphalogramme (EEG), etc.

(b) selon le **type d'espace des états** $\mathcal{X}$ (états **observables** ou non) :

$(b)_1$ une **série à espace d'état scalaire** (ou « simple », ou encore « univarié ») est une série pour laquelle généralement $\mathcal{X} \subset \mathbf{N}$, $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$ ou $\mathbf{C}$ (X et x sont donc à valeurs scalaires). Autrement dit, $\mathcal{X}$ peut être un **espace vectoriel** de dimension 1 (ou un corps) et l'on peut alors parler de **série scalaire** ;

$(b)_2$ une **série à espace d'état vectoriel** est une série pour laquelle généralement $\mathcal{X} \subset \mathbf{N}^p$, $\mathbf{Z}^p$, $\mathbf{R}^p$ ou $\mathbf{C}^p$ (X et x sont donc à valeurs vectorielles). Autrement dit, $\mathcal{X}$ peut être un **espace vectoriel** de dimension > 1. On peut parler de **série vectorielle** ;

$(b)_3$ en suivant la même démarche que dans les cas $(b)_1$ et $(b)_2$ précédents, X (resp x) peut, dans certaines situations, prendre ses valeurs dans un ensemble numérique $\mathcal{X}$: on parle de **processus numérique** ou de **série numérique** (cf **variable quantitative**). Dans d'autres situations, X (resp x) peut prendre ses valeurs dans un ensemble non numérique $\mathcal{X}$, et l'on parle de **processus qualitatif** (resp de **série qualitative**) (cf aussi **variable qualitative**). Un processus qualitatif (resp une série qualitative) relate ainsi l'évolution de **variables qualitatives** (eg couleurs, formes ou opinions successives, etc).

Il existe aussi des séries mixtes, ie combinant des composantes numériques et des composantes qualitatives.

(iv) Le **mode d'observation** d'un processus (resp d'une série) est important. En général, eg pour des raisons budgétaires ou matérielles, le dispositif d'observation (**système statistique**, **dispositif expérimental**) possède des limites ou des contraintes.

(a) les dates ou périodes d'observation des variables décrivant un phénomène ne sont pas nécessairement régulières ou périodiques (eg analyses médicales individuelles d'un type donné, pendant une durée de vie) ;

$$$ Les instants ou les périodes de temps $t \in S$ n'ont pas nécessairement des intervalles (ou des durées) de temps réguliers (régulières) ;

(b) souvent, X est partiellement observé, ie observé seulement sur une « portion » $S \subset T$ de T : on n'observe donc une « portion » $x_S = (x_s)_{s \in S}$, avec $S \neq \emptyset$ et $S \neq T$, de la trajectoire « complète » $x = (x_t)_{t \in T}$. On parle parfois de « **réalisation partielle** » (ou « réalisation inachevée ») du processus X.

(c) de plus, le phénomène décrit par le processus X peut être indexé par un **temps « théorique »** T différent du **temps « observé »** S pour la série x : eg observation discontinue d'un phénomène continu ou inversement. En effet, le **système d'observation** statistique (cf **production statistique**) ne prend pas nécessairement en compte ce phénomène avec la même notion de « temps » (**non superposition**). Ainsi :

$(c)_1$ un processus en temps continu ($T \subset \mathbf{R}$) peut être « observé » de façon continue ou de façon « discontinue » ou « intermittente » (ie en temps discret : eg $S \subset \mathbf{Z}$) ;

$(c)_2$ un processus en temps discret (eg $T \subset \mathbf{Z}$) peut être observé de façon discontinue ou observé de façon continue (graphe en escalier). Dans le premier cas, les instants ou périodes ne sont pas nécessairement ceux (celles) du processus lui-même (ceci dépend du mode d'observation et des conditions d'observation). L'observation en des instants différents des instants de manifestation du phénomène est donc impossible : eg observation aux instants $T = [1, 3, 5, ...] = 2\,\mathbf{N} + 1$ et réalisations aux instants $[0, 2, 4, 6, ...] = 2\,\mathbf{N} + 2$.

(v) On peut alors étudier diverses particularités de x (cf **composante d'une série temporelle**) : décompositions diverses, comportant des périodicités d'intérêt (**saisonnalité**, distribution des **fréquences**, etc), **valeurs extrêmes**, valeurs passées, valeurs futures possibles, etc.

L'analyse d'une série, comme celle d'un processus, peut s'effectuer :

(a) dans l'**ensemble des états**, parfois dit **ensemble des temps**. Dans ces méthodes temporelles, ce sont les valeurs (**états**) du processus ou de la série qui sont étudiées et mises en relations entre elles (voire même avec le temps). Ces relations (ou équations) sont donc exprimées dans l'ensemble des états $\mathscr{X}$ (cf **espace d'état**). Elles peuvent être statiques (aucun **opérateur avance** ou **opérateur retard** n'intervient) ou dynamiques (avec présence de ces opérateurs ou de leurs itérés).

Diverses méthodes peuvent alors être mises en oeuvre :

$(a)_1$ ainsi, certaines méthodes, parfois qualifiées de « méthodes empiriques » conduisent à décomposer une série x selon (cas d'une forme additive) (cf **schéma**) :

(2) $\quad x = m + c + s + u,$

où $m = (m_t)_{t \in T}$ représente une composante tendantielle (ou **tendance**), $c = (c_t)_{t \in T}$ une composante cyclique (ou **cycle**), $s = (s_t)_{t \in T}$ une composante saisonnière (ou **saisonnalité**) et $u = (u_t)_{t \in T}$ un terme aléatoire (erreur sur l'équation), laquelle est souvent traitée comme une **perturbation aléatoire** (eg dans un **modèle de régression** sur données temporelles) ;

$(a)_2$ les méthodes de type inférentiel (eg **méthodes de BOX-JENKINS**) étudient des modèles reliant des processus entre eux ou des séries entre elles. Elles concernent généralement des **modèles dynamiques** : eg **modèle à retards échelonnés**, **modèle autorégressif**, **modèle autorégressif de moyenne mobile**, **modèle avec autocorrélation temporelle**, **modèle d'interdépendance dynamique**. La nature même d'un processus suggère le type de **modélisation** : eg **processus à retards échelonnés**, **processus autorégressif**, **processus autorégressif de moyenne mobile**, **processus stationnaire** ou non stationnaire, etc ;

(b) dans l'**ensemble des fréquences**, ou **domaine des fréquences**. Ces **méthodes fréquentistes** consistent à décomposer les séries (trajectoires) selon diverses composantes périodiques, dont on étudie les **périodes** ou les **phases** : l'analyse de FOURIER (**série de FOURIER**, **transformation de FOURIER**) est à la base de l'**analyse harmonique** et de l'**analyse spectrale** (**analyse cospectrale** pour un **processus vectoriel**).

Ces méthodes visent donc à décomposer la fonction (trajectoire de X) $t \mapsto x_t$ en termes de **fréquences** (au sens mathématique) et de **périodes**. Ainsi, un **processus stationnaire au sens large** peut s'analyser à l'aide d'une décomposition de la forme :

(3) $X_t = \Sigma_{k=1}^{K} A_k \sin(\omega_k t + \phi_k) + u_t, \qquad \forall t \in T.$

L'inférence statistique est fondée sur la **spécification** temporelle de la **loi** du processus X sous-jacent à x. Cette loi est souvent, pour des raisons de commodité, supposée tq X est un **processus stationnaire**, ou **processus homogène** ou encore un **processus ergodique**.

(vi) Dans la pratique, les notations ci-dessus ne se rencontrent pas toujours : majuscules ou minuscules pour les processus ou leurs trajectoires, lettres latines ou grecques pour les paramètres ou leurs estimateurs, etc.

Cependant le symbolisme général relatif aux processus et séries permet de représenter des situations très variées.

Ainsi, lorsque $T = N_T^*$, avec $T = H . I$, et que x est une série purement saisonnière (de période I) (saison « administrative », ou « calendaire »), on peut écrire :

(4) $x_t = x_{I(h-1)+i} = y_{hi}$,

avec la relation élémentaire entre indices :

(5) $t = I(h-1) + i, \qquad \forall (h, i) \in N_H^* \times N_I^*.$

Autrement dit, pour tout t, il existe un couple (h, i) tq (4) soit vérifiée, et inversement.

La (H,I)-matrice Y d'éléments y_{hi} est appelée **table de C.H.D. BUYS-BALLOTT**.

Ainsi (sociologie), I = 52 représente le nombre annuel de semaines, I = 12 le nombre annuel de mois, I = 4 le nombre annuel de trimestres. Le formalisme {(3),(4)} est ainsi adapté à l'étude de divers mouvements saisonniers.

D'autres formalismes dépendent du problème à traiter (autres domaines de connaissance).

Une décomposition tq (2) est parfois adéquate pour représenter des séries temporelles de nature météorologique, démographique, économique ou sociale. Dans le cas de données relatives à la physique (eg astronomie), les périodicités sont beaucoup plus longues que ci-dessus (eg de l'ordre de l'année ou de l'annéelumière). A l’inverse, dans le cas de données relatives eg à la biologie (ECG, EEG, etc), les périodicités peuvent être beaucoup plus courtes (eg de l’ordre de la seconde ou de la milli-seconde).