Introduction aux séries temporelles

—

Généralités sur les processus à temps discret

Gilles Faÿ
gilles.fay@ecp.fr

Laboratoire MAS, Ecole Centrale Paris

Cours 1/4 ; OMA, 7/1/2014

Processus stochastique à temps discret

Définition

On appelle **série temporelle** ou **série chronologique** tout processus stochastique à temps discret ($t \in \mathbb{Z}$ par exemple) et à valeur dans $\mathbb{R}^d$ ou $\mathbb{C}^d$:
$(X_t)_{t\in\mathbb{Z}}$ forme une collection de v.a. définies sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

Cadre statistique

Observations : $X_1, \ldots, X_n$
Un processus peut donner lieu à des procédures statistiques différentes d'une suite i.i.d. si il a une structure qu'on peut apprendre de ces seules observations
La propriété la plus importante est celle de stationnarité.

Stationnarité stricte

Définition

$(X_t)_{t\in\mathbb{Z}}$ est dit **strictement stationnaire** si et seulement si

$$\forall n \in \mathbb{N}^*, \forall (t_1, \ldots, t_n) \in \mathbb{Z}^n, \forall h \in \mathbb{Z},$$
$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_n+h})$$

où $\stackrel{d}{=}$ désigne l'égalité en loi.

Exemples

- Suite i.i.d. (**bruit blanc fort** si de plus moyenne nulle) ;
- $(f(X_t))_{t\in\mathbb{Z}}$ avec X_t i.i.d.
- Chaîne de Markov dans son régime stationnaire ;
- A-t-on X_t, Y_t fortement stationnaire $\Longrightarrow X_t + Y_t$ aussi ?

Stationnarité faible

Covariance

Pour tout processus dans $\mathbb{L}^2$, on définit sa **fonction d'autocovariance**

$$\gamma_X(t,s) = \operatorname{cov}(X_t, X_s).$$

Stationnarité au second ordre (ou faible)

$(X_t)_{t\in\mathbb{Z}}$ est dit **stationnaire au second ordre** (ou **au sens faible**) si et seulement si

- $\forall t, X_t \in \mathbb{L}^2$,
- $\exists \mu_X \in \mathbb{R}, \forall t, \mathbb{E}(X_t) = \mu_X$,
- $\exists \gamma_X : \mathbb{Z} \to \mathbb{R}, \forall t \in \mathbb{Z}, \forall h \in \mathbb{Z}, \operatorname{cov}(X_t, X_{t+h}) = \gamma_X(h)$

Type positif

Une fonction d'autocovariance d'un processus réel (resp. complexe) est paire (resp. hermitienne), $\gamma(s, s) \geq 0$ et vérifie

$$\forall n, \forall \boldsymbol{a} := (a_1, \ldots, a_n)' \in \mathbb{C}^n, \forall t_1, \ldots, t_n, \boldsymbol{a}' \Gamma_{t_1,\ldots,t_n} \overline{\boldsymbol{a}} \geq 0$$

où

$$[\Gamma_{t_1,\ldots,t_n}]_{i,j} = [\gamma(t_i, t_j)]_{i,j} \ .$$

Stationnarité faible

Exemples

- Suite de variables décorrélées (on parle de **bruit blanc faible** si de plus de moyenne nulle). Ex. $X_t = Z_t Z_{t-1}$, Z_t i.i.d centré.
- Somme de deux processus stationnaires au second ordre et décorrélés entre eux.

Stationnarité : relations

Dans L^2

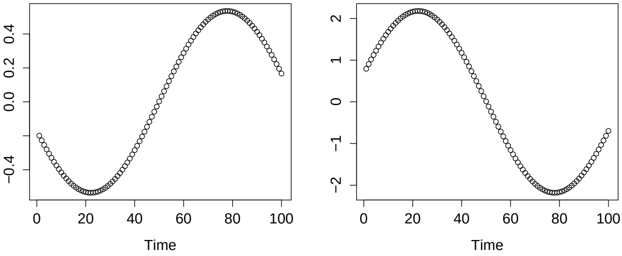
stationnarité forte $\Longrightarrow$ stationnarité faible.

En dehors...

Une suite i.i.d. de variable de Student à 2 degrés de liberté (telle que la moyenne existe mais pas la variance) est fortement stationnaire mais pas faiblement.

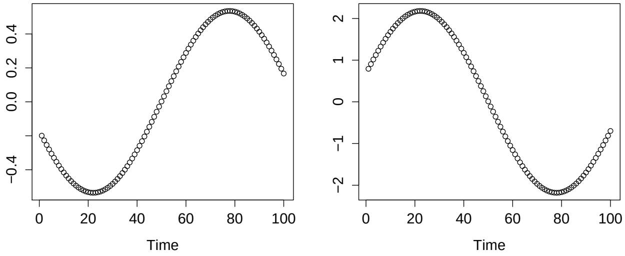
Question difficile/problème mal posé

Stationnaire ou pas ?



Question difficile/problème mal posé

Stationnaire ou pas ?



Exercice

Soit $X_t = Z_1 \cos(ct) + Z_2 \sin(ct)$ avec Z_1 et Z_2 décorrélées, centrées et de variance 1.

1. X_t est-il stationnaire au second ordre ?
2. strictement stationnaire ?

$AR(1)$

Dans quels cas existe-t-il une soltion stationnaire à l'équation de récurrence

$$X_t = aX_{t-1} + Z_t$$

où Z_t est i.i.d. ?
Cette solution est-elle unique ?

Exemple de processus non-stationnaires

Marche aléatoire

- $X_0 = 0$;
- U_t i.i.d ;
- $X_t = X_{t-1} + U_t$.

Pourquoi ?

Exemple de processus non-stationnaires

Marche aléatoire

- $X_0 = 0$;
- U_t i.i.d ;
- $X_t = X_{t-1} + U_t$.

Processus de Poisson

- $T_i = \sum_{j=1}^{i} E_j$, $E_j \sim \mathcal{E}(\lambda)$ i.i.d
- $X_t = \sum_{j \in \mathbb{N}^*} \mathbf{1}_{T_i \leq t}$.

Pourquoi ?

Exemple de processus non-stationnaires

Marche aléatoire

- $X_0 = 0$;
- U_t i.i.d ;
- $X_t = X_{t-1} + U_t$.

Processus de Poisson

- $T_i = \sum_{j=1}^{i} E_j$, $E_j \sim \mathcal{E}(\lambda)$ i.i.d
- $X_t = \sum_{j\in\mathbb{N}^*} \mathbf{1}_{T_i \leq t}$.

Processus de Poisson recentré

- $\tilde{X}_t = X_t - \mathbb{E}(X_t)$.

Pourquoi ?

Se ramener au cas stationnaire

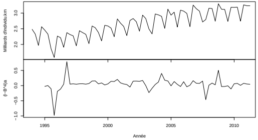
Box-Jenkins (1970)

On modélise les séries au moyen de trois composantes :

- une **tendance** déterministe M_t ;
- une **composante saisonnière** S_t ;
- une partie stationnaire aléatoire Z_t ;

Modèles

- Additif : $X_t = M_t + S_t + Z_t$;
- Multiplicatif : $M_t S_t + Z_t$;
- etc.



Trafic sur le réseau ferré de la RATP

(Source :INSEE,RATP.)

Se ramener au cas stationnaire (2)

Estimer et soustraire

- Poser un modèle simple et identifiable sur M_t et S_t ;
- L'estimer par moindres carrés (pondérés...)
- Soustraire pour obtenir $\hat{Y}_t = X_t - \hat{M}_t - \hat{S}_t$.

Filtrer

- Trouver un filtre Φ qui élimine M_t et S_t ;
- Continuer l'estimation sur ϕX_t.
- Exemple

Processus gaussiens

Définition

On dit que $(X_t)_{t\in\mathbb{Z}}$ est un **processus gaussien** si toutes les distributions fini-dimensionnelles de X_t sont des vecteurs gaussiens.

Propriété

Pour toute suite $\mu : \mathbb{Z} \to \mathbb{R}$ et toute fonction $\Gamma : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ de type positif, il existe un processus gaussien indexé par $\mathbb{Z}$ admettant μ et Γ comme moyenne et fonction d'autocovariance.

Corollaire

Toute fonction de type positif est la fonction d'autocovariance d'un processus, qui peut être choisi gaussien.

Exercice

On considère $(Z_t)_{t\in\mathbb{Z}}$ une suite i.i.d. de variables gaussiennes centrées de variance σ^2. Parmi les processus suivants, déterminer lesquels sont gaussiens et lesquelles sont stationnaires au second ordre. Pour ces derniers, donner la structure au second ordre.

(a) $X_t = a + bZ_t + cZ_{t-1}$;

(b) $X_t = a + bZ_0$;

(c) $X_t = Z_1 \cos(ct) + Z_2 \sin(ct)$;

(d) $X_t = Z_0 \cos ct$;

(e) $X_t = Z_t \cos(ct) + Z_{t-1} \sin(ct)$;

(f) $X_t = Z_t Z_{t-1}$.

Exercice

Soit Z_t un bruit blanc faible de variance σ^2 et $X_t := Z_t + \theta Z_{t-1}$. On dit que X est un processus MA(1). On vérifie que X_t ainsi défini est stationnaire au second ordre et que sa fonction d'autocovariance est donnée par

$$\gamma_X(h) = \sigma^2 \times \begin{cases} 1 + \theta^2 & \text{si } h = 0; \\ \theta & \text{si } |h| = 1; \\ 0 & \text{si } |h| > 1. \end{cases}$$

Réciproquement, soit R la fonction donnée par

$$R(h) = \begin{cases} 1 & \text{si } h = 0; \\ \rho & \text{si } |h| = 1; \\ 0 & \text{si } |h| > 1. \end{cases}$$

Montrer que c'est la fonction d'un processus stationnaire au second ordre si et seulement si $|\rho| \leq 1/2$.

Mesure spectrale

Théorème d'Herglotz

Une suite $\{\gamma(h)\}_{h\in\mathbb{Z}}$ est de type positif si et seulement s'il existe une mesure positive sur $\{\Lambda, \mathcal{B}(\Lambda)\}$ telle que :

$$\gamma(h) = \int_\Lambda e^{ih\lambda}\nu(d\lambda) \tag{1}$$

Si la suite $\{\gamma(h)\}_{h\in\mathbb{Z}}$ est de type positif et $\sum_h |\gamma(h)| < \infty$, ν possède une densité f (fonction positive) par rapport à Leb$\{\Lambda, \mathcal{B}(\Lambda)\}$, donnée par la série :

$$f(\lambda) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \gamma(h)e^{-ih\lambda} \geq 0 \tag{2}$$

Lorsque $\gamma = \gamma_X$ pour X un processus stationnaire au second ordre, ν est appelée **mesure spectrale** et la fonction f, lorsque qu'elle existe, est dite **densité spectrale (de puissance)**.

Exercice

Corollaire du théorème d'Herglotz

Une suite $R(h)$ à valeurs complexes absolument sommable est de type positif si et seulement si la fonction :

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} R(h) e^{-ih\lambda}$$

est positive pour tout $\lambda \in \Lambda$.

Autocovariance type MA(1)

$R(h) = \delta_0(h) + \rho\delta_{-1}(h) + \rho\delta_1(h)$ est de module sommable et :

$$f(\lambda) = \frac{1}{2\pi} \sum_k R(h) e^{-ih\lambda} = \frac{1}{2\pi}(1 + 2\rho\cos(\pi\lambda))$$

donc que R est une fonction d'autocovariance ssi $|\rho| \leq 1/2$.

Structure de second ordre

Définition

On appelle structure de second ordre d'un processus stationnaire au second ordre X_t sa moyenne μ_X et sa fonction d'autocovariance γ_X ou de façon équivalente, sa moyenne μ_X et sa mesure spectrale ν_X.

Propriété

Si le processus X_t est gaussien, la connaissance de sa structure de second ordre suffit à déterminer toutes les lois fini-dimentionnelles du processus X_t.

Estimateurs de la structure de second ordre (domaine temporel)

La **moyenne empirique** du processus est

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \ .$$

L'**autocovariance empirique** d'ordre h est définie par

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{i=1}^{n-|h|}(X_i - \bar{X}_n)(X_{i+|h|} - \bar{X}_n)$$

pour $|h| < n$ et 0 pour $|h| \geq n$. De même, l'**autocorrélation empirique** vaut

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

pour $|h| < n$ et 0 pour $|h| \geq n$.

Remarque : Avec normalisation en $(n - |h|)$ la suite $\hat{\gamma}$ ne serait pas de type positif

Estimateurs de la structure de second ordre (domaine fréquentiel)

Le **périodogramme** des observations à la fréquence $\lambda \in \Lambda$ est défini par

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{k=1}^{n} (X_k - \bar{X}_n) \mathrm{e}^{\mathrm{i}\lambda k} \right|^2$$

Exercice

Montrer

$$I_n(\lambda) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \hat{\gamma}(h) e^{-ih\lambda} \; .$$

et en déduire que $\hat{\gamma}$ est de type positif.

Propriétés de base

Moyenne empirique

Soit X un P.S.S.O.

- Si $\gamma_X(h) \to 0$ lorsque $h \to \infty$, $\lim_{n\to\infty} \mathrm{var}(\bar{X}_n) = 0$
- Si de plus $\sum_{h\in\mathbb{Z}} |\gamma(h)| < \infty$ alors $\lim_{n\to\infty} n\mathrm{var}(\bar{X}_n) = \sum_{h\in\mathbb{Z}} \gamma(h)$
- Incidence pour l'estimation ?

Variance empirique

Soit X un bruit blanc fort

- $\forall h, \sqrt{n}\hat{\gamma}(h) \to \mathcal{N}(0,1)$
- Incidence pour un « test de blancheur » ?

Régression dans du bruit corrélé

Solution itérative. Les intervalles sont modifiés.

Prochaine séance

- Notion de filtrage ;
- Théorèmes de filtrage ;
- Classe des processus ARMA.
- Propriétés des ARMA.