Plan du cours de
Séries Chronologiques

1 Introduction

Exemples. Séries chronologiques climatiques, hydroliques, financières,...

Objectifs.
- décrire une série chronologique.
- modéliser une série chronologique.
- comprendre et expliquer des phénomènes attenant à une série.
- prévoir le comportement futur d'une série.

2 Description d'une série chronologique

2.1 Processus aléatoire : définitions et premières propriétés

Définition.
- Processus aléatoires sur T à valeurs dans $\mathbb{R}^k$.
- Trajectoire d'un processus.
- Processus aléatoires du second ordre.
- Fonctions espérance, variance, covariance et corrélation d'un processus à valeurs réelles.
- Série chronologique.

Rappel.
- Variables aléatoires indépendantes.
- Matrice de covariance.
- Indépendance $\Longrightarrow$ non-corrélation, mais réciproque fausse en générale.

Définition.
- Processus aléatoires (séries chronologiques) gaussiens.

Propriété.
- Si X et Y sont des v. a. gaussiennes, alors (Indépendance $\Longleftrightarrow$ non-corrélation).
- Un processus gaussien est entièrement défini par ses fonctions espérance $m(t) = \mathbb{E}X(t)$ et covariance $\gamma(s,t) = \mathbb{E}X_sX_t$. Réciproquement, la connaissance d'une fonction $\gamma(s,t)$ définie positive et d'une fonction $m(t)$, définit un unique processus gaussien.

Définition.
- Série chronologique (strictement) stationnaire. Conséquences sur les fonctions espérance, variance, covariance.

- Série chronologique stationnaire d'ordre 2.
- Série chronologique à accroissements stationnaires.

Propriété.
- (Stationnarité (stricte) $\Longrightarrow$ Stationnarité d'ordre 2), mais la réciproque est fausse.
- Si X est une série chronologique gaussienne, alors (Stationnarité (stricte) $\Longleftrightarrow$ Stationnarité d'ordre 2).

Définition.
- Bruit blanc (fort).
- Bruit blanc faible.
- Bruit blanc gaussien.

2.2 Tendance et composante saisonnière

Définition. Toute série chronologique X peut s'écrire sous la forme

$$X_t = a(t) + S(t) + \varepsilon_t, \quad \text{pour } t \in T,$$

où $t \mapsto a(t)$ et $t \mapsto S(t)$ sont deux fonctions déterministes telle que $\mathbb{E}X_t = a(t) + S(t)$, avec

1. si $t \mapsto S(t)$ est une fonction périodique non nulle de période $r > 0$ telle que $\sum_{i=1}^{r} S(i) = 0$, alors S est la composante saisonnière, saisonnalité de X,
2. si $t \mapsto a(t)$ est non nulle, a est la tendance de X,
3. si ε_t est une série chronologique centrée, ε est le bruit (résidu) de X.

Exemples. Tendances polynomiales, saisonnalité annuelle,..

Définition.
- Série détendancialisée: $X_t - a_t$.
- Série désaisonnalisé (ou série corrigée des variations saisonnières): $X_t - S(t)$.

2.3 Description à l'aide d'une régression

On voudrait connaître la tendance et la saisonnalité de la série chronologique en supposant connue $(X_1, .., X_N)$.
On suppose que la tendance et la saisonnalité s'écrivent sous une forme connue a priori (les f_i, fonctions quelconques, et r sont connues), soit:

$$a(t) = \sum_{i=1}^{k} a_i f_i(t) \quad \text{et} \quad S(t) = \sum_{i=1}^{r-1} s_i(g_i(t) - g_r(t)) \quad \text{pour } t \in T,$$

avec $g_i(t) = \mathbb{I}_{\{t=i,\ [r]\}}$ sont r-périodiques (on a ainsi $\sum_{i=1}^{r} S(i) = 0$).

Notations. • $X = {}^t(X_1, .., X_N)$.

- $F_i = {}^t(f_i(1), .., f_i(N))$ pour $i = 1, .., k$ et $G_i = {}^t(g_i(1) - g_r(1), .., g_i(N) - g_r(N))$ pour $i = 1, .., r-1$.
- $U = (\varepsilon(1), .., \varepsilon(N))$.

Le modèle s'écrit alors vectoriellement:

$$X = \sum_{i=1}^{k} a_i f_i + \sum_{i=1}^{r} s_i (g_i - g_r) + U.$$

Proposition. On peut estimer les coefficients (a_i) et (s_i) par une régression par moindres carrés en minimisant une distance dans $\mathbb{R}^n$:

$\| X - (a_1 F_1 + .. + s_{r-1} G_{r-1}) \|^2$, et on notant Z la matrice $Z = (F_1 F_2 ... G_{r-1})$,

1. si U est un bruit blanc, ou un bruit inconnu, on utilise une estimation par moindres carrés ordinaires et:

$$ {}^t(\widehat{a_1}, .., \widehat{a_k}, \widehat{s_1}, .., \widehat{s_{r-1}}) = ({}^t Z.Z)^{-1\,t} Z X,$$

2. si U est un bruit tel que $\mathbb{E} U {}^t U = \Sigma$ est une matrice connue, on utilise une estimation par moindres carrés généralisés, et:

$$ {}^t(\widehat{a_1}, .., \widehat{a_k}, \widehat{s_1}, .., \widehat{s_{r-1}}) = ({}^t Z \Sigma^{-1} Z)^{-1\,t} Z \Sigma^{-1} X.$$

Remarque. • Les estimateurs sont sans biais, et sont (en général) convergents avec N (on peut calculer la matrice de covariance dans les deux cas).

- On peut noter que ces estimateurs sont des combinaisons linéaires des X_i.
- Une telle régression nécessite la connaissance a priori des fonctions f_i, ce qui rend la méthode efficace essentiellement dans le cas polynômial (avant ou après anamorphose).
- Pour répondre à la précédente remarque, on peut utiliser des **régressions localisées** de type Loess ou Lowess, où l'on fixe une taille de fenêtre et on fait une régression linéaire de X dans la fenêtre que l'on fait glisser.

Cas particulier de l'estimation du saisonnier

L'estimation par moindres carrés permet d'estimer le saisonnier d'une série chronologique. On suppose connue $(X_1, .., X_{rN})$, où r est la période (connue) du saisonnier. On écrit $X_t = S(t) + Y_t$, où $S(t) = \sum_{i=1}^{r-1} s_i(g_i(t) - g_r(t))$ pour $t \in T$, avec $g_i(t) = \mathbb{I}_{\{t=i,\ [r]\}}$ (voir plus haut). Dans ce cas, on a :

$$Z = \begin{pmatrix} 1 & 0 & 0 & . & 0 \\ 0 & 1 & 0 & . & 0 \\ 0 & 0 & 1 & . & 0 \\ . & . & . & . & . \\ -1 & -1 & -1 & . & -1 \\ 1 & 0 & 0 & . & 0 \\ 0 & 1 & 0 & . & 0 \\ . & . & . & . & . \\ -1 & -1 & -1 & . & -1 \\ 1 & 0 & 0 & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ -1 & -1 & -1 & . & -1 \end{pmatrix}, \qquad {}^tZ.Z = N \begin{pmatrix} 2 & 1 & . & 1 \\ 1 & 2 & . & 1 \\ . & . & . & . \\ 1 & 1 & . & 2 \end{pmatrix},$$

et

$$({}^tZ.Z)^{-1} = \tfrac{1}{rN} \begin{pmatrix} r-1 & -1 & . & -1 \\ -1 & r-1 & . & -1 \\ . & . & . & . \\ -1 & -1 & . & r-1 \end{pmatrix}.$$

On montre alors que dans le cadre d'une estimation par moindres carrés ordinaires :

$$s_i = \frac{1}{N} \sum_{k=1}^{N} X_{i+r(k-1)} - \frac{1}{rN} \sum_{k=1}^{rN} X_k, \quad \text{pour } i = 1, \cdots, r-1.$$

2.4 Description à l'aide d'un filtre linéaire

Définition. Soit $X = \{X_k, k \in \mathbb{Z}\}$ une série chronologique. On appelle filtre linéaire une famille de réels $a = (a_n)_{-p \le n \le q}$, où $(p, q) \in \mathbb{N}^2$, et la série filtrée est $Y = \{X_k, k \in \mathbb{Z}\}$ telle que $Y_k = \sum_{n=-p}^{q} a_n X_{k+n}$.

Exemples. Filtres des accroissements, moyenne empirique sur des fenêtres mobiles,$\cdots$

Propriété.

- Si X stationnaire alors Y est stationnaire;
- Si X est gaussien alors Y est gaussien;
- Si X est centré alors Y est centré;
- La composée de 2 filtres linéaires est un filtre linéaire.

Propriété. Si X est une suite numérique réelle définie par récurrence par $\sum_{n=-p}^{q} a_n X_{k+n} = 0$, alors on résout l'équation caractéristique associée: $\sum_{n=0}^{p+q} a_{n-p} x^n = 0$, ayant pour solutions $(\lambda_i)_{1\leq i\leq m} \in \mathbb{C}^m$ et:

$$X_n = \sum_{k=0}^{m} |\lambda_k|^n \left(P_k(n)\cos(\arg(\lambda_k)n) + Q_k(n)\sin(\arg(\lambda_k)n)\right),$$

où les P_k et Q_k sont des polynômes de degré la multiplicité de λ_k moins un.

Propriété. On considère les filtres linéaires particuliers suivants:

1. $a_i = 1/k$ pour $i = 1, \cdots, k$ (à une translation près). Alors si X a une composante saisonnière de période r, elle est annulée dès que k est un multiple de r. De plus, si la partie bruit de X est un bruit blanc, alors la série Y a un bruit de variance plus petite que celle de X. Enfin, seuls les polynômes de degré 0 et 1 sont invariants par ce filtre.

2. $a_i = (-1)^i C_i^k$ pour $i = 0, 1, \cdots, k$ (à une translation près). Alors si la tendance de X est une tendance polynômiale de degré $k-1$, elle est annulée.

3. $a_0 = -1$, $a_r = 1$ et $a_i = 0$ pour $i = 1, \cdots, r-1$. Alors si X a une composante saisonnière de période r, elle est annulée.

Propriété. Quelque soit le filtre linéaire utilisé, si Y est la série filtrée, on peut reconstruire la série X à partir de Y (+ les premières valeurs de X).

3 Prédiction

3.1 Propriétés générales

Rappel. Si X et Y sont deux variables aléatoires continues de densité jointe $f_{(X,Y)}(x,y)$ alors

$$\mathbb{E}(X \mid Y = y) = \frac{\int x f_{(X,Y)}(x,y)dx}{\int f_{(X,Y)}(x,y)dx}.$$

Définition. On suppose que $(X_1, \cdots, X_N)$ est connue. On appelle prédiction $\widehat{X}_{N+p}$ à l'horizon p $\widehat{X}_{N+p} = f_p(X_1, \cdots, X_N)$ où f_p est une fonction mesurable et $\mathbb{L}^2$.

Définition. Soit X une série chronologique de carré intégrable. On appelle erreur moyenne quadratique (MSE) de a prédiction $\widehat{X}_{N+p}$ le réel

$$\mathbb{E}(\widehat{X}_{N+p} - X_{N+p})^2.$$

Proposition. On suppose que X une série chronologique de carré intégrable. La prédiction $\widehat{X}_{N+p}$ optimale au sens de la moyenne quadratique (ou des moindres carrés) est définie par :

$$\widehat{X}_{N+p} = \widehat{f}(X_1, \cdots, X_N) \text{ avec } \widehat{f} = \text{Argmin}_{f \in \mathbb{L}^2} \left(\mathbb{E}(X_{N+p} - f(X_1, \cdots, X_N))^2 \right).$$

Alors,

$$\widehat{X}_{N+p} = \mathbb{E}(X_{N+p} \mid (X_1, \cdots, X_N)).$$

$\widehat{X}_{N+p}$ est donc une variable aléatoire dépendant de $(X_1, \cdots, X_N)$ telle que

$$\mathbb{E}(X_{N+p} \mid (X_1, \cdots, X_N)) = \frac{\int x_{N+p} f_{(X_1, \cdots, X_N, X_{N+p}))}(X_1, \cdots, X_N, x_{N+p}))dx_{N+p}}{\int f_{(X_1, \cdots, X_N, X_{N+p}))}(X_1, \cdots, X_N, x_{N+p}))dx_{N+p}}.$$

Proposition. On suppose que X une série chronologique de carré intégrable. Alors le prédiction **linéaire** $\widehat{X}_{N+1}$ optimale au sens de la moyenne quadratique (ou des moindres carrés) est définie par :

$$\widehat{X}_{N+p} = \widehat{a_1} X_1 + \cdots + \widehat{a_n} X_N + \widehat{b} \text{ avec}$$
$$(\widehat{a_i}) = \text{Argmin}_{(a_i) \in \mathbb{R}^n} \left\{ \mathbb{E} \left(X_{N+p} - (a_1 X_1 + \cdots + a_N X_N + b) \right)^2 \right\}.$$

Les coefficients estimés sont obtenus par régression.

Proposition. Si X est une série gaussienne, alors la prédiction linéaire par moindres carrés est aussi la prédiction par moindres carrés (donc l'espérance conditionnelle).

3.2 Prédiction par filtrage exponentiel

L'étude théorique précédente nécessite la connaissance de la loi du processus. Lorsque celle-ci est inconnue, mais que la tendance et la saisonnalité sont supposées composées de certaines fonctions, le lissage exponentiel est une démarche "naturelle" pour prédire X_{N+p} lorsque $(X_1, \cdots, X_N)$ est connu.

On suppose que la tendance et la saisonnalité s'écrivent sous une forme connue a priori (les f_i, fonctions non constantes, et r sont connues), soit:

$$a(t) = \sum_{i=1}^{k} a_i f_i(t) \quad \text{et} \quad S(t) = \sum_{i=1}^{r} s_i g_i(t) \quad \text{pour } t \in T,$$

avec $g_i(t) = \mathbb{I}_{\{t=i,\ [r]\}}$ sont r-périodiques (on ne suppose pas ici que $\sum_{i=1}^{r} S(i) = 0$).

Notations. • $X = (X_1, .., X_N)$.

- $F_i = (f_i(1), .., f_i(N))$ pour $i = 1, .., k$ et $G_i = (g_i(1), .., g_i(N))$ pour $i = 1, .., r$.
- $U = (\varepsilon(1), .., \varepsilon(N))$.

Le modèle s'écrit alors vectoriellement:

$$X = \sum_{i=1}^{k} a_i f_i + \sum_{i=1}^{r} s_i g_i + U.$$

Proposition. On peut estimer les coefficients (a_i) et (s_i) par une régression linéaire par moindres carrés pondérés qui accorde un poids exponentiellement plus important à X_N qu'à X_M si $M < N$. Pour cela, on minimise la distance dans $\mathbb{R}^N$ dépendant d'un paramètre β:

$$\| X - (a_1F_1 + .. + s_rG_r) \|^2_\beta = {}^t(X - (a_1F_1 + .. + s_rG_r)\Omega^{-1}(\beta)(X - (a_1F_1 + .. + s_rG_r),$$

en notant $\Omega(\beta) = \begin{pmatrix} \beta^N & 0 & 0 & .. & 0 \\ 0 & \beta^{N-1} & 0 & .. & 0 \\ : & : & : & : & : \\ 0 & 0 & 0 & .. & 1 \end{pmatrix}$ et avec Z la matrice $Z = (F_1F_2...G_r)$,

Alors $\widehat{X_{N+p}} = (f_1(N+p), \cdots, g_r(N+p))({}^tZ\Omega(\beta)Z)^{-1}{}^tZ\Omega(\beta)X.$

Remarque. En pratique, on choisit souvent β entre 0.5 et 0.8.

4 Analyse spectrale d'une série chronologique

On s'intéresse ici à l'identification de la partie bruit d'une série chronologique.

4.1 Rappel : tests d'adéquation

Rappel. On suppose connu $(X_1, \cdots, X_N)$ suite de v.a.i.i.d. et on veut connaître la loi de X:

- Test du Chi-Deux.
- Test de Kolmogorov-Smirnov

Mais que se passe-t-il lorsque les variables ne sont plus indépendantes? Comment estimer la structure de covariance du bruit?

4.2 Définition - Propriétés

Dans toute la suite, on suppose que $X = \{X_k, k \in \mathbb{Z}\}$ est une série chronologique centrée **stationnaire** (pas de tendance ni de saisonnalité) à valeurs réelles.

Définition.
- On note $r(k) = \mathbb{E}X_0X_k = \mathbb{E}X_iX_{i+k}$ pour $k \in \mathbb{Z}$, la covariance de X. $r(0)$ est la variance de la série.
- On note $\rho(k) = r(k)/r(0)$ la corrélation de X. On a $-1 \leq \rho(k) \leq 1$ pour $k \in \mathbb{Z}$.

Définition. S'il existe une fonction $f : [-\pi, \pi[\to \mathbb{C}$ telle que $\forall k \in \mathbb{Z}$, $r(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda$, alors on dit que X admet une densité spectrale f.

Propriété. Si f existe alors f est paire et $f(\lambda) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} r(k) e^{-ik\lambda}$ pour tout $\lambda \in [-\pi, \pi[$.

Propriété.
1. (Les covariances $(r(k))$ sont telles que $\sum |r(k)|^2 < \infty$) $\iff$ (f existe et f est de carré intégrable).
2. (Les covariances $(r(k))$ sont telles que $\sum |r(k)| < \infty$) $\Longrightarrow$ (f existe et f continue sur $[-\pi, \pi[$).

Exemples. Bruits blancs et (X_n) telle que $X_n = \varepsilon_n - a\varepsilon_{n-1}$.

Définition. On appelle covriance empirique de $(X_1, \cdots, X_n)$ pour $p = 0, 1, \cdots, n-1$,

$$c_n(p) = \frac{1}{n} \sum_{k=1}^{n-p} (X_{k+p} - \bar{X}_n)(X_k - \bar{X}_n) \quad \text{et } c_n(-p) = c_n(p).$$

Lorsque l'on sait que le processus est centré alors pour $p = 0, 1, \cdots, n-1$,

$$c_n(p) = \frac{1}{n} \sum_{k=1}^{n-p} X_{k+p} X_k \quad \text{et } c_n(-p) = c_n(p).$$

On appelle corrélation empirique de $(X_1, \cdots, X_n)$ pour $p = 0, 1, \cdots, n-1$,

$$\rho_n(p) = \frac{c_n(p)}{c_n(0)}.$$

Définition. On appelle périodogramme de $(X_1, \cdots, X_n)$ la fonction $I_n(\lambda)$: $[-\pi, \pi[\to \mathbb{R}_+$ telle que lorsque X est centrée :

$$I_n(\lambda) = \frac{1}{2\pi} \sum_{k=1-n}^{n-1} c_n(k) e^{-ik\lambda} = \frac{1}{2\pi n} \left| \sum_{k=1}^{n} X_k e^{-ik\lambda} \right|^2.$$

Propriété. Si les covariances $(r(k))$ sont telles que $\sum |r(k)| < \infty$)

1. $\lim_{n \to \infty} \mathbb{E}(I_n(\lambda)) = f(\lambda)$ pour tout $\lambda \in [-\pi, \pi[$.
2. $\lim\limits_{n \to \infty} Var(I_n(\lambda)) = \begin{cases} f^2(\lambda) & \text{si } \lambda \in]-\pi, 0[\cup]0, \pi[\\ 2f^2(\lambda) & \text{si } \lambda = -\pi, 0. \end{cases}$
3. $\lim\limits_{n \to \infty} \operatorname{cov}(I_n(\lambda), I_n(\lambda')) = 0$ si $\lambda \neq \lambda'$.

4.3 Statistique asymptotique des séries chronologiques stationnaires

Proposition. Soit $(X_1, \cdots, X_n)$ une série chronologique stationnaire d'espérance m et de densité spectrale f continue en 0. Alors :

$$\begin{array}{rcl} \bar{X}_n & \xrightarrow[n\to+\infty]{p.s.} & m \quad \text{et} \\ \lim\limits_{n\to\infty} n\mathbb{E}(\bar{X}_n - m)^2 & = & m. \end{array}$$

Proposition. Soit $(X_1, \cdots, X_n)$ une série chronologique stationnaire gaussienne d'espérance m et de densité spectrale f. Alors :

1. Si $\lim_{k\to\infty} r(k) = 0$ alors $\bar{X}_n \xrightarrow[n\to+\infty]{p.s.} m$.
2. Si f est continue en 0, alors $\sqrt{n}(\bar{X}_n - m) \xrightarrow[n\to+\infty]{\mathcal{L}.} \mathcal{N}(0, 2\pi f(0))$.
3. Si f est continue, alors pour p fixé, $c_n(p) \xrightarrow[n\to+\infty]{p.s.} r(p)$.

Définition. Pour g une fonction continue sur $[-\pi, \pi[$, on considère :

$$J_n(g) = \int_{-\pi}^{\pi} g(\lambda) I_n(\lambda) d\lambda.$$

Proposition. Soit $(X_1, \cdots, X_n)$ une série chronologique stationnaire gaussienne centrée et de densité spectrale f continue. Alors :

1. Pour g une fonction continue sur $[-\pi, \pi[$, $J_n(g) \xrightarrow[n\to+\infty]{p.s.} \int_{-\pi}^{\pi} f(\lambda)g(\lambda)d\lambda$.
2. Pour $(g_i)_{1\leq i_l eq m}$ des fonctions continues sur $[-\pi, \pi[$, et si f est strictement positive et de classe $\mathcal{C}^1$ sur $[-\pi, \pi[$, alors :

$$\sqrt{n}\left(J_n(g_i) - \int_{-\pi}^{\pi} f(\lambda)g(\lambda)d\lambda\right)_{1\leq i\leq m} \xrightarrow[n\to+\infty]{\mathcal{L}.} \mathcal{N}_m(0, \Sigma)$$

avec $\sigma_{ij} = 4\pi \int_{-\pi}^{\pi} f^2(\lambda) g_i(\lambda) g_j(\lambda) d\lambda$.

Conséquence.

- Comme $c_n(p) = \int_{-\pi}^{\pi} \cos(p\lambda) I_n(\lambda) d\lambda$, on peut appliquer le théorème de limite centrale précédent.
- En utilisant un noyau comme fonction g que l'on fait se rapprocher d'une impulsion de Dirac en λ_0, on obtient un estimateur de la densité $f(\lambda_0)$.

Notation : On suppose que l'on a un modèle paramétrique de paramètre $\theta_0 = (\theta_1, \cdots, \theta_p)$ tel que la forme générale f_θ soit connue, mais le paramètre θ_0 est inconnu. On définit $L_n(\theta, \omega)$ une vraisemblance pour $\theta \in \Theta$ (compact). Soit

$$\widehat{\theta_n} = \text{Argmax}_{\theta\in\Theta} L_n(\theta) = \text{Argmax}_{\theta\in\Theta} \log L_n(\theta).$$

Proposition. Soit $(X_1, \cdots, X_n)$ une série chronologique stationnaire gaussienne centrée et de densité spectrale f_θ bijective telle que $\sup(f_\theta)$, $\sup(1/f_\theta)$ et $\sup(\text{grad} f_\theta)$ sont bornées pour $\theta \in \Theta$. Alors : $\widehat{\theta_n} \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} \theta_0$.

Proposition. Soit $(X_1, \cdots, X_n)$ une série chronologique stationnaire gaussienne centrée et de densité spectrale f_θ bijective telle que $(\theta, \lambda) \mapsto f_\theta(\lambda)$ soit strictement positive et de classe $\mathcal{C}^3$ sur $\Theta \times [-\pi, \pi[$, alors :

$$\sqrt{n}(\widehat{\theta_n} - \theta_0) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}_m(0, I(\theta_0)^{-1}),$$

où $I_{ij}(\theta) = \dfrac{1}{4\pi}\displaystyle\int_{-\pi}^{\pi} \frac{\partial \log f_\theta(\lambda)}{\partial\theta_i}.\frac{\partial \log f_\theta(\lambda)}{\partial\theta_j} d\lambda.$

Remarque. La dernière proposition est encore valable dans le cas non gaussien en rajoutant des conditions de régularité sur $\dfrac{\partial^2 \log L(\theta)}{\partial\theta^2}$.

Proposition. Soit $(X_1, \cdots, X_n)$ une série chronologique stationnaire gaussienne centrée et de densité spectrale f_θ bijective telle que $(\theta, \lambda) \mapsto f_\theta(\lambda)$ soit strictement positive et de classe $\mathcal{C}^3$ sur $\Theta \times [-\pi, \pi[$. On considère le contrastedit contraste de Whittle :

$$U_n(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \log(f_\theta(\lambda)) + \frac{I_n(\lambda)}{f_\theta(\lambda)} d\lambda.$$

Soit $\widetilde{\theta_n} = \text{Argmin}_{\theta\in\Theta} U_n(\theta)$, appelé estimateur de θ par minimum de contraste (ou par maximum de vraisemblance approché de Whittle). Alors :

$$\sqrt{n}(\widetilde{\theta_n} - \theta_0) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}_m(0, I(\theta_0)^{-1}),$$

où $I_{ij}(\theta) = \dfrac{1}{4\pi}\displaystyle\int_{-\pi}^{\pi} \frac{\partial \log f_\theta(\lambda)}{\partial\theta_i}.\frac{\partial \log f_\theta(\lambda)}{\partial\theta_j} d\lambda.$

5 Processus ARMA et ARIMA

Les processus ARMA et ARIMA sont des exemples paradigmatiques de séries chronologiques que nous allons rapidement étudier.

5.1 Préambules

Définition.
- Opérateur retard B tel que $Y = (Y_n)_{n\in\mathbb{Z}} = B.X$ où $X = (X_n)_{n\in\mathbb{Z}}$ vérifie $Y_n = X_{n-1}$ pour $n \in \mathbb{Z}$ (par abus de notation $B.X_n = X_{n-1}$).
- Puissance B^k tel que $B^k.X_n = X_{n-k}$.
- Polynôme de B.
- Si $\sum_{i=-\infty}^{\infty} |a_i| < \infty$, alors $f(B) = \sum_{i=-\infty}^{\infty} a_i B^i$ transforme X en $Y = (Y_n = \sum_{i=-\infty}^{\infty} a_i X_{n-i})$.

Propriété.
- Additivité des fonctions de B.
- Multiplication d'une fonction de B par une constante.
- Commutativité du produit de fonctions de B.

Proposition. Soit $X = (X_n)_{n\in\mathbb{Z}}$ et $Y = (Y_n)_{n\in\mathbb{Z}} = (1 - \lambda B).X$. Alors,

- si $|\lambda| < 1$, $X_n = \sum_{i=0}^{\infty} \lambda^i Y_{n-i}$;
- si $|\lambda| > 1$, $X_n = -\sum_{i=1}^{\infty} \lambda^{-i} Y_{n+i}$;
- si $|\lambda| = 1$, $1 - \lambda B$ n'est pas inversible.

Conséquence. Si on suppose que $P(B)$ est polynôme dont les racines ne sont pas situées sur le cercle trigonométrique, alors il existe une unique série de la forme $(P(B))^{-1} = \sum_{i=-\infty}^{\infty} a_i B^i$ avec $\sum_{i=-\infty}^{\infty} |a_i| < \infty$

5.2 Processus ARMA

Propriété. Soit $(\varepsilon_n)_{n\in\mathbb{Z}}$ un bruit blanc et soit $(a_i)_{i\in\mathbb{Z}}$ une famille de réels tels que $\sum_{i=-\infty}^{\infty} |a_i| < \infty$. Alors la série chronologique $X = (X_n)_{n\in\mathbb{Z}}$ telle que $X_n = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{n-i}$ pour $n \in \mathbb{Z}$ est stationnaire et appelée processus linéaire (ou processus à moyennes mobiles infinies).

Propriété. Soit la série chronologique $X = (X_n)_{n\in\mathbb{Z}}$ telle que $X_n = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{n-i}$ pour $n \in \mathbb{Z}$, avec $(\varepsilon_n)_{n\in\mathbb{Z}}$ un bruit blanc et $\sum_{i=-\infty}^{\infty} |a_i| < \infty$. Alors X admet une densité spectrale f telle que

$$f(\lambda) = \frac{1}{2\pi} \left| \sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda} \right|^2 , \quad \text{pour } \lambda \in [-\pi, \pi[.$$

Propriété. Soit la série chronologique $Y = (Y_n)_{n\in\mathbb{Z}}$ telle que $Y_n = \sum_{i=-\infty}^{\infty} a_i X_{n-i}$ pour $n \in \mathbb{Z}$, avec $X = (X_n)_{n\in\mathbb{Z}}$ une série chronologique stationnaire de densité spectrale f_X et $\sum_{i=-\infty}^{\infty} |a_i| < \infty$. Alors Y est une série stationnaire de densité :

$$f_Y(\lambda) = f_X(\lambda) \left| \sum_{j=-\infty}^{\infty} a_j e^{-ij\lambda} \right|^2 , \quad \text{pour } \lambda \in [-\pi, \pi[.$$

Définition.

- Processus AR(p).
- Processus MA(q).
- Processus ARMA(p,q).

Propriété. Si les racines de P sont à l'exterieur du disque trigonométrique, alors un processus ARMA peut s'écrire sous la forme d'un processus linéaire.

Propriété. Lorsque les racines de P sont à l'extérieur du disque trigonométrique, on peut établir des relations de récurrence permettant de calculer la fonction covariance d'un processus ARMA(p,q) tel que

$$X_{n+p} + a_1 X_{n+p-1} + \cdots + a_p X_n = Q(B).\varepsilon_n.$$

On obtient ainsi les équations de Yule-Walker, vérifiant le système :

$$\begin{pmatrix} r(q) & r(q-1) & . & r(q-p+1) \\ r(q+1) & r(q) & . & r(q-p+2) \\ . & . & . & . \\ r(q+p-1) & r(q+p-2) & . & r(q) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ . \\ a_p \end{pmatrix} = - \begin{pmatrix} r(q+1) \\ r(q+2) \\ . \\ r(q+p) \end{pmatrix}.$$

Conséquence. La covariance d'un processus ARMA décroît asymptotiquement exponentiellement vite.

Définition. On appelle fonction d'autocorrélation partielle d'ordre s, la fonction $\phi(s)$ qui est le coefficient α_s de X_{n-s} dans la régression linéaire de X_n sur $(X_{n-1}, \cdots, X_{n-s})$.

Propriété. ($X = (X_n)_{n\in\mathbb{Z}}$ est un processus AR(p)) $\Longleftrightarrow$ (la fonction d'autocorrélation partielle $\phi(.)$ est telle que $\phi(p) \neq 0$ et $\phi(p) = 0$ pour $s \geq p+1$).

Propriété. Les coefficients d'autocorrélation partielle $\phi(.)$ s'obtiennent en inversant les équations de Yule-Walker pour écrire les coefficients en fonction des covariances de $X = (X_n)_{n\in\mathbb{Z}}$.

Propriété. Lorsque les racines de P sont à l'extérieur du disque trigonométrique, un processus ARMA(p,q) (tel que $P(B).X = Q(B).\varepsilon$ avec (ε_n) bruit blanc de variance σ^2) a pour densité spectrale :

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|\frac{Q(e^{i\lambda})}{P(e^{i\lambda})}\right|^2, \quad \text{pour } \lambda \in [-\pi, \pi[.$$

5.3 Estimation et prédiction pour un processus ARMA

■ Pour estimer les coefficients d'un ARMA(p,q) pour lequel $(X_1, \cdots, X_N)$ est connu ainsi que p et q, on peut utiliser :

- L'estimateur par maximum de vraisemblance dans le cas gaussien.
- L'estimateur par minimum de contraste à partir de l'écriture de la densité spectrale (plus intéressant).
- L'estimation des covariances pour inverser les équations de Yule-Walker et estimer les coefficients.

■ Pour estimer l'ordre (p, q) d'un ARMA pour lequel $(X_1, \cdots, X_N)$ est connu, on peut utiliser :

- La convergence de l'estimateur par minimum de contraste.
- Le corrélogramme pour l'estimation de l'ordre q de la partie MA : en effet, théoriquement, si $\rho(k) \neq 0$ pour $k = 0, 1, \cdots, M$ et $\rho(M+1) = 0$ alors un estimateur de q est M.
- Le graphe des corrélations partielles pour l'estimation de l'ordre p de la partie AR : en effet, théoriquement, si $\phi(k) \neq 0$ pour $k = 0, 1, \cdots, M$ et $\phi(M+1) = 0$ alors un estimateur de q est M.

Propriété. Soit $\widehat{X}_{N+1}$ le prédicteur optimal au sens des moindres carrés, lorsque $(X_1, \cdots, X_N)$ est connu et lorsque X est un ARMA(p,q) connu (tous

les coefficients sont connus). Alors $\widehat{X}_{N+1} = -\sum_{k\geq 0} c_{k+1} X_{N-k}$ où les (c_k) sont définis par le développement en série entière de $\frac{P}{Q}(z) = \sum_{k\geq 0} c_k z^k$.