
Series temporelles : ARIMA, LSTM?

Jérôme Petit

9 avril 2019

1 INTRODUCTION

2 SÉRIES TEMPORELLES

Une série temporelle ou série chronologique est une suite d'observations d'un phénomène faites au cours du temps. Cette notion s'étend à de nombreux domaines, tout relevé de données, de flux ou d'information est une série temporelle : indice boursier, population française, trafic Internet, etc. Ses données réelles obtenues sont des suites finies indexés par un temps continu ou discret. Les modèles mathématiques utilisées pour modélisées ses séries temporelles sont eux de dimension infinie.

2.1 Stationnarité

Definition 2.1. *Stationnarité stricte ou forte : $(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire au sens strict si : $\forall n \in \mathbb{N}, \forall (t_1, \dots, t_n), \forall h \in \mathbb{Z}$ la loi de $(X_{t_1}, \dots, X_{t_n})$ est identique à la loi de $(X_{t_1+h}, \dots, X_{t_n+h})$*

D'après le théorème de Kolmogorov, on en déduit qu'un processus (X_t) est stationnaire au sens fort si et seulement si la loi de X_t est identique à la loi de X_{t+h} quelque soit h . Cette condition de stationnarité est très contraignante et relativement peu observée en pratique. On remarque pour un processus stationnaire strict X_t admettant des moments d'ordre 1 et 2 finis, on a :

$$\mathbb{E}(X_t) = \mathbb{E}(X_{t+h}), \forall h \in \mathbb{N}$$

on pose : $\mathbb{E}(X_t) = \mathbb{E}(X_0) = \mu$. Pour la covariance, on a :

$$\begin{aligned} \text{Cov}(X_t, X_{t+h}) &= \mathbb{E}((X_t - \mu)(X_{t+h} - \mu)) \\ &= \mathbb{E}((X_0 - \mu)(X_{h_0} - \mu)) \end{aligned}$$

la dernière égalité vient du fait que (X_t, X_{t+h}) a la même distribution que (X_0, X_h) . On remarque donc que les processus stationnaire strict ont des moments d'ordre 1 et 2 indépendant du temps t . C'est cette notion qui va être utilisée pour des définir les processus stationnaires au sens faible.

Definition 2.2. Stationnarité faible : $(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire du second ordre (ou processus faiblement stationnaire) s'il vérifie :

- i) $\forall t \in \mathbb{Z}, \mathbb{E}(X_t) = m$
- ii) $\forall t \in \mathbb{Z}, \mathbb{V}(X_t) = \sigma^2$
- iii) $\forall t, h \in \mathbb{Z}, \text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_0, X_h) = \gamma(h)$

On note $\gamma(h)$ l'autocovariance d'ordre h de X_t .

Dans la suite les processus stationnaires désigneront les processus stationnaire du second ordre. La fonction d'autocovariance est paire et $\gamma(0) = \mathbb{V}(X_t)$.

Definition 2.3. Le coefficient d'autocorrélation d'ordre h d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ est :

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

La fonction $h \mapsto \rho(h)$ est appelée fonction d'autocorrélation. Parmi les processus stationnaires, il y a les processus appelés bruit blanc.

Definition 2.4. Un bruit blanc ϵ_t est une suite de variable aléatoire non corrélées, mais pas nécessairement indépendantes de moyenne nulle et de variance constante σ^2 . On note $\epsilon_t \sim \mathcal{BB}(0, \sigma^2)$.

Un bruit blanc est donc un processus stationnaire d'ordre 2 tel que :

- $\mathbb{E}(X_t) = 0$
- $\gamma(h) = 0$ si $h \neq 0$

Un bruit blanc n'est pas nécessairement fortement stationnaire car les variables aléatoires ne sont pas nécessairement indépendantes.

Definition 2.5. Un bruit blanc gaussien ϵ_t est une suite de variable aléatoire i.i.d de loi normale $\mathcal{N}(0, \sigma^2)$. On note $\epsilon \sim \mathcal{BBN}(0, \sigma^2)$

Un bruit blanc gaussien est un processus stationnaire au sens strict.

Les processus non stationnaires sont des processus dont les moments d'ordre 1 ou 2 ne sont pas un constant. La marche aléatoire est un exemple de processus non stationnaire car la variance est fonction du temps.

Dans les cas pratiques, on aura une série observée pour $t = 1, \dots, T$, dans ce cas là on devra calculer l'autocavariance (resp. autocorrélation) empirique :

$$\begin{aligned} \bar{\gamma}(h) &= \frac{1}{T} \sum_{t=h+1}^T (y_t - \bar{y})(y_{t-h} - \bar{y}), 0 \leq h \leq T-1 \\ \bar{\rho}(h) &= \frac{\sum_{t=h+1}^T (y_t - \bar{y})(y_{t-h} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}, 0 \leq h \leq T-1, \bar{y} = \frac{1}{T} \sum_{t=1}^T y_t \end{aligned}$$

Dans le cas ou le processus y_t est stationnaire l

2.2 Outil pour l'étude des processus stationnaires

Théorème 2.6. Soit X_t un processus stationnaire alors il existe une bruit blanc au sens faible ϵ_t et des coefficients réels (Ψ_t) tel que :

$$X_t = m + \sum_{i \geq 0} \Psi_i \epsilon_i$$

avec le moment d'ordre 1 de X_t

Cette décomposition est appelé décomposition de Wold. ou transformé par moyenne mobile. Afin d'étudier les processus stationnaires, on utilisera les coefficients d'auto-corrélation et les coefficients d'auto corrélation partielles.

Definition 2.7. On appelle auto-corrélation partielle d'ordre p :

$$r(p) = \frac{\text{Cov}(X_t - \mathbb{E}(X_t | X_{t-1}, \dots, X_{t-p}), X_{t-p} - \mathbb{E}(X_{t-p} | X_{t-1}, \dots, X_{t-p+1}))}{\mathbb{V}(X_t | X_{t-1}, \dots, X_{t-p}) \mathbb{V}(X_{t-p} | X_{t-1}, \dots, X_{t-p+1})}$$

3 PROCESSUS

3.1 processus MA

Definition 3.1. Le processus $(X_t)_{t \in \mathbb{Z}}$ est une moyenne mobile d'ordre q s'il existe un bruit blanc $\epsilon_t \sim BB(0, \sigma^2)$ et des réels $\theta_1, \dots, \theta_q$ tels que :

$$X_t = m + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

On note $MA(q)$ une moyenne mobile d'ordre q .

Un processus $MA(q)$ est stationnaire, la fonction d'auto-covariance est :

$$\gamma(h) = \begin{cases} 0 & \text{si } |h| > q \\ \sigma^2(1 + \sum_{i=1}^q \theta_i^2) & h = 0 \\ \sigma^2 \theta_q^2 & h = q \\ \sigma^2(\theta_h + \sum_{i=h+1}^q \theta_i \theta_{i-h}) & \text{si } 0 < h < q \end{cases}$$

Il n'y a pas de résultats particulier pour les auto-corrélations partielles.

3.2 processus AR

Definition 3.2. Un processus $(X_t)_{t \in \mathbb{Z}}$ est un processus auto-régressif d'ordre p noté $AR(p)$ si :

- i) $(X_t)_{t \in \mathbb{Z}}$ est stationnaire
- ii) $(X_t)_{t \in \mathbb{Z}}$ vérifie :

$$X_t = \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t$$

avec $\phi_p \neq 0$ et $\epsilon_t \sim BB(0, \sigma^2)$

On pose $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$, un processus $AR(p)$ peut s'écrire à l'aide de polynôme de retard :

$$\Phi(L)X_t = \mu + \epsilon_t$$

A l'aide de cette formulation, on en déduit les racines du polynôme Φ sont toutes de module strictement supérieur à 1.

Proposition 3.3. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $AR(p)$, alors :

- $\mathbb{E}(X_t) = \frac{\mu}{1 - \phi_1 - \dots - \phi_p}$
- on pose $Y_t = X_t - \mathbb{E}(X_t)$ alors Y_t est un processus $AR(p)$ d'espérance nulle

Les processus $AR(p)$ sont des processus stationnaires, on a vu que les racines du polynômes de retard sont toutes strictement supérieure à 1, cela signifie que l'on peut inverser le polynôme de retard. Ainsi tout processus $AR(p)$ peut s'écrire : $X_t = \Phi(L)^{-1}(\mu + \epsilon_t)$. L'inverse donnant une somme infinie, cela veut dire que tout processus $AR(p)$ peut s'écrire sous la forme $MA(\infty)$.

3.2.1 Propriétés des processus AR(p)

On a vu précédemment que d'un processus AR(p) admettant une espérance non nulle, on pouvait toujours se ramener à un processus AR(p) d'espérance nulle. Dans la suite on considérera le processus AR(p) $(X_t)_{t \in \mathbb{Z}}$ d'espérance nulle. L'auto covariance s'écrit pour tout $h > 0$:

$$\begin{aligned}\gamma(h) &= \text{Cov}(X_t, X_{t-h}) \\ &= \mathbb{E}(X_t X_{t-h}) \\ &= \sum_{i=1}^p \phi_i \mathbb{E}(X_{t-i} X_{t-h}) + \mathbb{E}(\epsilon_t X_{t-h}) \\ &= \sum_{i=1}^p \phi_i \gamma(h-i)\end{aligned}$$

Pour $h=0$, on a :

$$\begin{aligned}\gamma(0) &= \mathbb{E}(X_t^2) \\ &= \sum_{i=1}^p \phi_i \mathbb{E}(X_{t-i} X_t) + \mathbb{E}(\epsilon_t X_t) \\ &= \sum_{i=1}^p \phi_i \gamma(i) + \mathbb{E}(\epsilon_t \epsilon_t) \\ &= \sum_{i=1}^p \phi_i \gamma(i) + \sigma^2\end{aligned}$$

On a de même pour l'auto-corrélation : $\rho(h) = \phi_1 \rho(h-1) + \dots + \phi_p \rho(h-p)$ pour tout $h > 0$. Ces équations sont appelés equation de Yule-Walker. On a une relation de récurrence :

$$r(0) = 1 = \sum_{i=1}^p \phi_i \rho(i) + \frac{\sigma^2}{\gamma(0)}$$

d'où :

$$\gamma(0) = \frac{\sigma^2}{1 - \sum_{i=1}^p \phi_i \rho(i)}$$

Les equations de Yule-Walker peuvent s'écrire :

$$\begin{pmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \dots & \rho(p-2) \\ \rho(2) & \rho(1) & \dots & \rho(p-3) \\ \vdots & \vdots & \dots & \vdots \\ \rho(p-1) & \rho(p-2) & \dots & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \vdots \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \vdots \\ \vdots \\ \rho(p) \end{pmatrix}$$

Les solutions de ce système sont données par les valeurs initiales $\rho(i)$. Ainsi si l'on peut estimer les corrélation sur un échantillon donnée on pourra en déduire les coefficients du processus AR(p).

Proposition 3.4. Soit $(X_t)_{t \in \mathbb{Z}}$ un processus AR(p) alors :

- i) $|\rho(h)|$ et $\gamma(h)$ décroissent exponentiellement avec h
- ii) l'auto corrélation partielle est nulle pour $h > p$

3.3 ARMA modèles

3.4 SARMA modèles

3.5 ARIMA modèles

3.6 SARIMA modèles

3.7 ARCH modèles

3.8 GARCH modèles

4 ANNEXE

4.1 Convergence

Convergence \mathcal{L}^2 :

$$X_n \rightarrow_{\mathcal{L}^2} X \Leftrightarrow \lim_{n \rightarrow \infty} \|X_n - X\|_2 = 0$$

On dit que X_t convergent en loi vers X si et seulement si pour toute fonction bornée ϕ , on a :

$$\lim_n \mathcal{E}(\phi(X_n)) = \mathcal{E}(\phi(X))$$

4.2 Densité spectrale

Proposition 4.1. Soit (X_t) un processus stationnaire de la forme :

$$X_t = m + \sum_{j \geq 0} a_j \epsilon_{t-j}$$

avec ϵ_t un bruit blanc $BB(0, \sigma^2)$ et $\sum_{j \geq 0} |a_j| < +\infty$. Alors :

$$— \sum_{h \in \mathbb{Z}} |\gamma(h)| < +\infty$$

$$— \forall \omega \in [-\pi, \pi], f_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma(h) e^{i\omega h} \text{ est la densité spectrale de } X_t.$$

Sous les hypothèses précédentes, on a :

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma(h) \cos(\omega h)$$

4.3 Estimateur empirique des moments

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire, on cherche à estimer l'espérance, les fonction d'auto-covariance, d'auto-corrélation et d'auto-corrélation partielle. On a un échantillon de taille T . On prend comme estimateur :

$$— \text{moyenne empirique : } \hat{m} = \frac{1}{T} \sum_{t=1}^T X_t = \bar{X}_T$$

$$— \text{auto-covariance empirique d'ordre } h : \hat{\gamma}(h) = \frac{1}{T-h} \sum_{t=h+1}^T (X_t - \bar{X}_T)(X_{t-h} - \bar{X}_T)$$

$$— \text{auto-corrélation empirique d'ordre } h : \hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

— l'auto-corrélation partielle est une régression empirique sur $1, X_1, \dots, X_{t-h}$.

Pour estimer la fonction de densité spectrale on utilise la auto-covariance empirique et un terme d'ajustement (coefficient de Newey-West) afin diminuer le poids de $\hat{\gamma}(h)$ pour des h grand.

Si X_t est un processus stationnaire alors d'après la loi des grands nombres tous les estimateurs empiriques présentés ci-dessus convergent.

Proposition 4.2. Si $X_t = m + \sum_{j \geq 0} a_j \epsilon_{t-j}$ avec ϵ_t un bruit blanc de moment d'ordre 4 fini, alors tous ses estimateurs ont des lois jointes asymptotiquement gaussiennes :

$$\sqrt{T}(\hat{m} - m) \rightarrow \mathcal{N}(0, \sum_{h \in \mathcal{Z}} \gamma(h))$$

convergence en loi

4.4 polynôme de retard

Définition 4.3. L'opérateur de retard est défini sur la classe des processus stationnaires comme étant :

$$L(X_t) = X_{t-1}$$

Définition 4.4. Soit P un polynôme, $P(z) = \sum_{k=0}^p a_k z^k$ avec $a_k \in \mathbb{R}$, on lui associe le polynôme de retard $P(L)$ défini comme suit :

$$P(L) = \sum_{k=0}^p a_k L^k$$

et

$$P(L)X_t = \sum_{k=0}^p a_k X_{t-k}$$

Un polynôme de retard $P(L)$ est inversible s'il existe un polynôme de retard $Q(L)$ tel que $P(L)Q(L) = Id$. Dans \mathbb{C} , tout polynôme réel P se décompose :

$$P(z) = \prod_{i=1}^p (z - z_i) = \prod_{i=1}^p (-z_i) \prod_{i=1}^p (1 - \frac{z}{z_i}) = \alpha \prod_{i=1}^p (1 - \lambda_i z)$$

avec $\lambda_i = \frac{1}{z_i}$. Le polynôme $(1 - \lambda z)$ est inversible si seulement si ses racines sont de modules strictement inférieur à 1. Ceci équivaut à $|\lambda| < 1$. On peut généraliser ce résultat à tout polynôme de degré p .

Proposition 4.5. Soit $P(L) = \alpha \prod_{i=1}^p (1 - \lambda_i L)$ un polynôme de retard, si pour tout i $|\lambda_i| < 1$ alors $P(L)$ est inversible et :

$$P^{-1}(L) = \sum_{k \geq 0} b_k L^k$$