Modélisation des séries temporelles

Séance 2

Les processus ARIMA

Frédéric Sur

https://members.loria.fr/FSur/enseignement/modseries/

Motivation / objectif

Box et Jenkins 1970 (à la suite de Yule et Slutsky 1927) :

de nombreuses phénomènes temporels (ou leur dérivée*), dans de nombreux domaines, peuvent être représentés par des* **processus stationnaires***.*

Rappel :
les processus stationnaires considérés sont de la forme

$$X_t = m + \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}$$

avec (ε_t) **bruit blanc gaussien**

Séance 2

Les processus MA *(=moyenne mobile)*

Définition **: processus MA(q)** (Slutsky 1927)

Ce sont les processus (X_t) du type :

$$\forall t,\ X_t = m + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

où $q \geqslant 1$, $(\theta_i)_{1\leqslant i \leqslant q}$ réels, et (ε_t) est un bruit blanc gaussien

Convention d'écriture

$$X_t - m = \Theta(B)\varepsilon_t$$

où $\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$
et B est l'*opérateur retard* : $B^k \varepsilon_t = \varepsilon_{t-k}$

Stationnarité des processus moyenne-mobile

Propriété

les processus MA(q) sont **stationnaires** (à l'ordre 2)

En effet, quelque soit t :

- $E(X_t) = m$
- $\forall\, h,\ \mathrm{Cov}(X_t, X_{t-h}) = \begin{cases} (1+\theta_1^2+\cdots+\theta_q^2)\,\sigma^2 & \text{si } h=0 \\ (\theta_h+\theta_{h+1}\theta_1+\cdots+\theta_q\theta_{q-h})\,\sigma^2 & \text{si } 1 \leqslant h \leqslant q \\ 0 & \text{si } h > q \end{cases}$

(calcul algébrique simple, cf poly.)

où σ^2 est la variance du b.b.g. ε_t.

$\rightarrow$ les deux premiers moments ne dépendent pas de t.

Caractérisation de l'ordre d'un MA(q)

$$\forall h \geqslant 0,\ \rho(h) = \frac{\mathrm{Cov}(X_t, X_{t+h})}{\sqrt{\mathrm{Var}(X_t)\mathrm{Var}(X_{t+h})}} = \frac{\gamma(h)}{\gamma(0)}.$$

Propriété : fonction d'autocorrélation (ACF) d'un MA(q)

$$\rho(h) = \begin{cases} 1 & \text{si } h=0 \\ \frac{\theta_h+\theta_{h+1}\theta_1+\cdots+\theta_q\theta_{q-h}}{\theta_1^2+\cdots+\theta_q^2} & \text{si } 1 \leqslant h \leqslant q \\ 0 & \text{si } h > q \end{cases}$$

Intérêt : détermination de q.
(cf séance 1)

Les processus AR *(=autorégressif)*

Définition : processus AR(p) (Yule 1927)

Ce sont les processus (X_t) du type :

$$\forall t,\ X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$$

où $p \geqslant 1$, $(\phi_i)_{1\leqslant i\leqslant p}$ réels, et (ε_t) est un bruit blanc gaussien

Convention d'écriture : (constante c vs. moyenne m)

$$\Phi(B)X_t = c + \varepsilon_t \quad \text{ou :} \quad \Phi(B)(X_t - m) = \varepsilon_t$$

avec $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ (B : opérateur retard)
et : $(1 - \phi_1 - \cdots - \phi_p)m = c$

Question : un processus AR(p) est-il stationnaire ?

Exemple : processus AR(1), $m = c = 0$

$$\forall t,\ X_t = \phi X_{t-1} + \varepsilon_t \quad (\text{ ou } (1-\phi B)X_t = \varepsilon_t\ )$$

Exemple-type, car : $\Phi(B)X_t = \varepsilon_t \ \Leftrightarrow\ \prod(1-\phi_i B)\ X_t = \varepsilon_t$

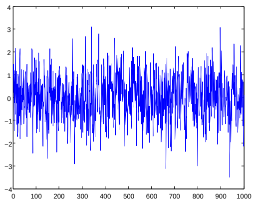
- si $|\phi| < 1$ (« racine » $1/\phi$ à l'extérieur du disque unité) alors
$X_t = \left(1 + \phi B + \phi^2 B^2 + \dots\right)\varepsilon_t$
(car $(1-\phi B)^{-1} = 1 + \phi B + \phi^2 B^2 + \dots$)
$\rightarrow$ (X_t) **stationnaire**, « **équivalent** » **à MA(∞)**.

- si $|\phi| > 1$ alors
$X_t = -\left(\phi^{-1}B^{-1} + \phi^{-2}B^{-2} + \dots\right)\varepsilon_t$
(car $(1-\phi B)^{-1} = -(1-\phi^{-1}B^{-1})^{-1}\phi^{-1}B^{-1} = -\phi^{-1}B^{-1} - \phi^{-2}B^{-2} + \dots$)
$\rightarrow$ (X_t) **stationnaire** mais **non causal sous cette forme**.
On démontre qu'il existe un b.b.g. ε' t.q. $X_t = \phi^{-1}X_{t-1} + \varepsilon'_t$.

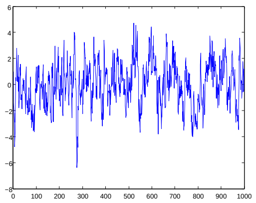
- si $\phi = 1$ (marche aléatoire) : $X_t = X_{t-1} + \varepsilon_t$.
On a bien $\forall t$, $\mathbb{E}(X_t) = \mathbb{E}(X_{t-1})$.
Mais $\mathrm{var}(X_t) = \mathrm{var}(X_{t-1}) + \sigma^2$ dépend de t (non constante).
$\rightarrow$ (X_t) **non-stationnaire**.

Simulation de processus AR(1)

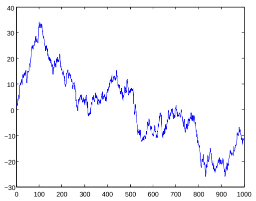
(ε_t) i.i.d. $\sim \mathcal{N}(0,1)$



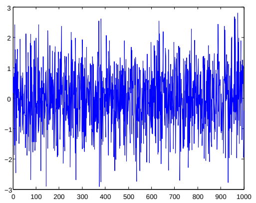
$\phi = 0.2$



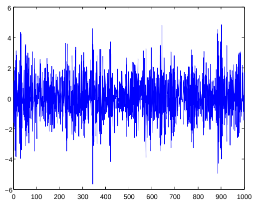
$\phi = 0.8$



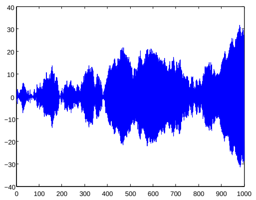
$\phi = 1$



$\phi = -0.2$



$\phi = -0.8$



$\phi = -1$

Et pour identifier l'ordre d'un processus AR ?

Exemple : AR(1) stationnaire, $|\phi| < 1$

$$X_t = \phi X_{t-1} + \varepsilon_t$$

(quitte à retrancher la moyenne, on suppose $E(X_t) = 0$)

On calcule :

$\gamma(1) = \mathrm{Cov}(X_t, X_{t-1}) = \mathrm{Cov}(\phi X_{t-1} + \varepsilon_t, X_{t-1}) = \phi\gamma(0)$

$$\begin{aligned}\gamma(2) &= \mathrm{Cov}(X_t, X_{t-2}) = \mathrm{Cov}(\phi X_{t-1} + \varepsilon_t, X_{t-2}) \\ &= \mathrm{Cov}(\phi^2 X_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t, X_{t-2}) = \phi^2\gamma(0)\end{aligned}$$

...

Preuve : (ε_t) bruit blanc + causalité

→ l'ACF ne permet pas d'identifier l'ordre p d'un processus AR aussi simplement que pour un processus MA

→ ici, X_t est corrélé à X_{t-2} par l'intermédiaire de X_{t-1}.

Autocorrélation partielle

Retirons la *dépendance linéaire* de X_{t-1} sur X_t et X_{t-2} :

→ régression de X_t sur X_{t-1} : on cherche α minimisant

$$E(X_t - \alpha X_{t-1})^2 = \gamma(0) - 2\alpha\gamma(1) + \alpha^2\gamma(0)$$

d'où : $\alpha = \gamma(1)/\gamma(0) = \phi$

notation : $\widetilde{X}_t = \phi X_{t-1}$ (« influence » de X_{t-1} sur X_t)

→ régression de X_{t-2} sur X_{t-1} : on cherche β minimisant

$$E(X_{t-2} - \beta X_{t-1})^2 = \gamma(0) - 2\beta\gamma(1) + \beta^2\gamma(0)$$

d'où : $\beta = \phi$

notation : $\widetilde{X}_{t-2} = \phi X_{t-1}$ (« influence » de X_{t-1} sur X_{t-2})

Conséquence :

$\mathrm{Cov}(X_t - \widetilde{X}_t, X_{t-2} - \widetilde{X}_{t-2}) = \mathrm{Cov}(\varepsilon_t, X_{t-2} - \phi X_{t-1}) = 0$

Conclusion : la notion utile pour identifier l'ordre d'un processus AR est l'*autocorrélation partielle* plutôt que l'autocorrélation.

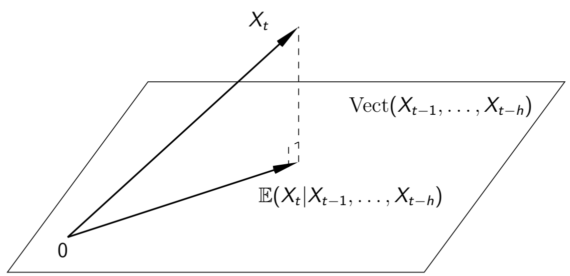
Espérance conditionnelle si (X_t) gaussien

Définition **("rappel") : espérance conditionnelle, cas gaussien**

$$\mathbb{E}(X_t | X_{t-1}, \ldots, X_{t-h}) = a_1(h)X_{t-1} + \cdots + a_h(h)X_{t-h}$$

est la « meilleure » approximation de X_t par une combinaison linéaire de $(X_{t-1}, \ldots, X_{t-h})$ dans L^2 muni du produit scalaire $< X, Y >= \mathbb{E}(XY)$.

(cf régression & cours « apprentissage automatique »)



Fonction d'autocorrélation partielle (PACF)

(X_t) stationnaire donc dans $(L^2, \mathbb{E}(XY))$.

$$\mathbb{E}(X_t|X_{t-1},\dots,X_{t-h}) = a_1(h)X_{t-1} + \cdots + a_h(h)X_{t-h}$$

Définition : PACF de (X_t)

Le terme $r(h) = a_h(h)$ est la *corrélation partielle* d'ordre h.

Propriété (cf poly : géométrie dans un espace de Hilbert)

$$\forall h \geqslant 1,\ a_h(h) = \operatorname{corr}\left(X_t - \widetilde{X}_t, X_{t-h} - \widetilde{X}_{t-h}\right)$$

où : $\widetilde{X}_t = \mathbb{E}(X_t|X_{t-1},\dots,X_{t-h+1})$

et : $\widetilde{X}_{t-h} = \mathbb{E}(X_{t-h}|X_{t-1},\dots,X_{t-h+1})$

Interprétation : corrélation entre X_t et X_{t-h} lorsque l'influence des variables $X_{t-1},\dots,X_{t-h+1}$ a été retirée.

$\rightarrow$ on vient de voir $a_1(1) = \gamma(1)/\gamma(0) = \operatorname{corr}(X_t, X_{t-1})$.

Retour aux processus MA(q)

L'expression générale du PACF d'un processus MA(q) est compliquée

Cas particulier : MA(1)

$$\forall t,\ X_t = m + \varepsilon_t + \theta\varepsilon_{t-1}.$$

alors :

$$r(h) = a_h(h) = \frac{(-\theta)^h(1-\theta^2)}{1-\theta^{2h+2}}.$$

$\rightarrow$ $(|r(h)|)_{h\geqslant 1}$ décroît « exponentiellement vite » (si $\theta \neq 1$)

Résumé : les processus MA *(moyennes mobiles)*

Définition : processus MA(q)

$$X_t = m + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

$$X_t - m = \Theta(B)\varepsilon_t$$

avec $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$

Propriété

Un processus MA(q) est stationnaire.

Corrélogrammes d'un MA(q) :

- ACF : $\rho(h) = 0$ si $h > q$,
- PACF : $|r(h)|$ tend vers 0 ($q = 1$ « vitesse exponentielle »)

$\longrightarrow$ permet l'identification d'un MA(q).

Caractérisation de l'ordre d'un AR(p) stationnaire

Propriété : PACF d'un processus AR(p)

Si $r(h)$ est la fonction d'autocorrélation partielle :

$$\forall h > p,\ r(h) = 0$$

Preuve : $X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=p+1}^{h} 0 \cdot X_{t-i} + \varepsilon_t$, puis projection sur $(X_{t-1},\dots,X_{t-h})$ sachant ε_t décorrélé de $X_{t-1},\dots,X_{t-h}$.

Intérêt : détermination de p.

Propriété : ACF d'un processus AR(p)

$(|\rho(h)|)_{h\geqslant 1}$ décroît vers 0 « exponentiellement vite ».

Preuve : cf équations de Yule-Walker, qui permettent de retrouver les ϕ_i à partir des $\rho(h)$. (poly)

$\rightarrow$ cf calcul précédent pour AR(1) : $\rho(h) = \phi^h$

Résumé : les processus AR *(autorégressifs)*

Définition : processus AR(p)

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$$

$$\Phi(B)(X_t - m) = \varepsilon_t \qquad \Phi(B)(X_t) = c + \varepsilon_t$$

avec $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$

Propriété

Un processus AR(p) est stationnaire si Φ n'a pas de racine unité.

Corrélogrammes d'un AR(p) stationnaire :

- ACF : $|\rho(h)|$ décroît de « manière exponentielle » vers 0
- PACF : $\forall h > p,\ r(h) = 0$

$\longrightarrow$ permet l'identification d'un AR(p).

Les processus ARMA *(=autorégressif moyenne mobile)*

Définition **: processus ARMA(p,q)**

Ce sont les processus (X_t) du type :

$$\forall t,\ X_t = \theta_0 + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

où $p \geqslant 1$, $(\phi_i)_{1 \leqslant i \leqslant p}$ et $(\theta_j)_{0 \leqslant j \leqslant q}$ réels, et (ε_t) b.b. gaussien.

Notation symbolique : $\Phi(B)X_t = \theta_0 + \Theta(B)\varepsilon_t$.
ou : $\Phi(B)(X_t - \mu) = \Theta(B)\varepsilon_t$
avec : $\theta_0 = \Phi(B)\mu$ (θ_0 joue le rôle de c dans un proc. AR)

Remarque : si Φ a ses racines à l'extérieur du cercle unité, alors on peut écrire :

$$X_t = \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j} \quad \text{(sous forme MA}(+\infty)\text{) + relations entre } a_j\text{)}$$

$\rightarrow$ *cf. principe de parcimonie : MA($+\infty$) représenté avec p+q+1 param.*

Séance 2

Les processus non stationnaires

Rappel : forme générale d'une chronique (modèle linéaire)

$$X_t = T_t + S_t + u_t$$

avec T_t tendance, S_t saisonnalité, u_t perturbation aléatoire

$\rightarrow$ (X_t) n'est pas stationnaire

$\rightarrow$ dans la suite du cours, les chroniques non stationnaires seront modélisés comme des processus ARIMA ou SARIMA

Les processus ARIMA *(=ARMA intégré)*

Opérateur de différenciation : $\nabla X_t = X_t - X_{t-1} = (1-B)X_t$.

Définition : processus ARIMA(p,d,q)

Ce sont les processus (X_t) du type :

$$\Phi(B)(1-B)^d X_t = \theta_0 + \Theta(B)\varepsilon_t$$

où Φ sans racine unitaire, $p, d, q \geqslant 0$,
$\deg(\Phi) = p$, $\deg(\Theta) = q$,
et (ε_t) est un bruit blanc gaussien (écart-type σ).

Remarque : processus ARIMA X_t non stationnaire,
mais ordre d tel que $(1-B)^d X_t$ stationnaire.

Cas de non stationarité

Hypothèse : (U_t) stationnaire d'espérance nulle

Tendance déterministe : (*espérance* non constante)
$X_t = f(t) + U_t$ (par ex. $f(t) = at + b$)
alors on estime les paramètres de f par minimisation des moindres carrés et on cherche un modèle ARMA sur les résidus U_t.

Tendance stochastique :
$X_t = X_{t-1} + U_t$ (marche aléatoire)
$\rightarrow$ *variance* non constante
$X_t = \mu + X_{t-1} + U_t$ (marche aléatoire avec *drift* - dérive)
$\rightarrow$ *tendance* + *variance* non constantes
alors $(1-B)X_t = (\mu +)\, U_t$ est stationnaire
(on dit : la chronique X_t est intégrée d'ordre 1)

Les processus SARIMA

Permettent de traiter le cas de la saisonnalité.

$\rightarrow$ séance 3.

Séance 2

Modélisation des chroniques par (S)ARIMA

Comment modéliser une chronique X_t par un ARIMA(p,d,q) ?

$\rightarrow$ méthode de Box et Jenkins (1970)

1. **Identification** des paramètres p, d, q.
2. **Estimation** des θ_j, ϕ_i, et σ
3. **Validation** du modèle.
4. **Prévision** du futur.

Identification du modèle ARIMA(p,d,q)

0. Transformation de la chronique (log, exp, $\sqrt{\cdot}$, Box-Cox) pour passer au « modèle additif ».

1. Identification de d.

Indices de non-stationnarité :

- « ça se voit » sur le graphique (tendance, variance pas constante)
- l'ACF ne décroît pas assez vite
- tests de Dickey-Fuller augmenté (séance 3)

$\rightarrow$ il faut dériver.

Identification du modèle ARIMA(p,d,q)

2. Identification de p et q.

$\rightarrow$ utilisation des corrélogrammes

Rappel :

	ACF	PACF
AR(p)	$\lvert\rho(h)\rvert$: décroissance exponentielle vers 0	$r(h) = 0$ si $h > p$
MA(q)	$\rho(h) = 0$ si $h > q$	$\lvert r(h)\rvert$: décroissance $\sim$ exponentielle vers 0

Remarque 1 : on dispose des ACF et PACF **empiriques**.
R donne des intervalles de confiance pour les $\widehat{\rho}(h)$ et $\widehat{r}(h)$ sous hypothèse $\rho(h) = 0$ et $r(h) = 0$

Remarque 2 : si « vrai » ARMA(p,q), comme on se limitera à $p, q \leqslant 2$, on teste les différents modèles. . .

Estimation des paramètres θ_j, ϕ_i, σ (à grands traits. . . - 1)

Définition : vraisemblance de la réalisation d'un vecteur aléatoire X

Soit f_θ la densité de X, et x une réalisation de X :

$$\mathcal{L}_x(\theta) = f_\theta(x)$$

(θ : paramètres de la densité d'un vecteur aléatoire)

Méthode du **maximum de vraisemblance** :

estimer θ qui maximise $\mathcal{L}_x(\theta)$

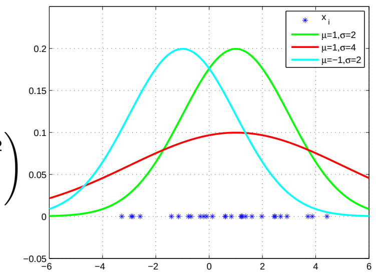
Idée : plus $\mathcal{L}_x(\theta)$ est grand, plus il est plausible que f_θ est la bon modèle

Exemple : $x = (x_1, x_2, \ldots, x_n)$

échantillon i.i.d. gaussien :

$$\mathcal{L}_x(\mu, \sigma) = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2\right)$$

Maximum of likelihood estimation (MLE)



Estimation des paramètres θ_j, ϕ_i, σ (à grands traits... - 2)

$$\forall t,\ X_t = \theta_0 + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

Hypothèse : (ε_t) i.i.d. $\sim \mathcal{N}(0, \sigma^2)$
« donc » $(X_t)_{0 \leqslant t \leqslant T}$ est un vecteur gaussien
(car combinaison linéaire des (ε_t))

Vraisemblance des T observations X_t :

$$\mathcal{L}_X((\phi_i), (\theta_j), \sigma) = \frac{1}{(2\pi)^{T/2}\sqrt{\det \Omega}} \exp\left(-\frac{1}{2}(X-\mu)^T \Omega^{-1} (X-\mu)\right)$$

où μ et Ω dépendent des ϕ_i, θ_j, σ, et $X = (X_t)_{1 \leqslant t \leqslant T}$

→ estimation des paramètres $(\phi_i), (\theta_j), \sigma$ par la méthode (numérique) du *maximum de vraisemblance*.

Remarque : il existe d'autres méthodes d'estimation.
→ *à utiliser lorsque R avertit d'un problème d'optimisation numérique.*

Examen de la validité du modèle

1. Résidus = **bruit blanc gaussien** ?
Estimation e_t des résidus : $X_t - \widehat{X}_{t-1}(1)$ (cf *prévision*, séance 3)
→ on vérifie que les résidus (e_t) sont bien des « réalisations » d'un bruit blanc gaussien (ε_t) :

- vérification graphique : moyenne nulle, variance constante, pas de « structure », QQ-plot (« droite de Henry »).
- ACF **des résidus** ne doit pas présenter de $\rho(h)$ significatif.
- *Portmanteau Test* (Box-Pierce) **sur les résidus** :
  soit $\rho(h)$ l'ACF des résidus (théoriquement, $\rho(h) = 0$ si $h > 0$)
  → $(T - K) \sum_{h=1}^{K} \widehat{\rho}(h)^2$ suit un χ^2 à $K - p - q$ d.d.l.,
  sous hypothèse (ε_t) bruit blanc gaussien.
  (généralement on fait le test pour plusieurs valeurs de K)

2. Significativité des estimations des paramètres
(Z-tests, Student, cf séance 3)
→ si un paramètre n'est pas significatif, on relance la modélisation en imposant sa valeur à 0.

Prévision avec modèle ARIMA(p,d,q)

→ la semaine prochaine.

Syntaxe R

`help(arima)` (on utilisera en fait `Arima` du package `forecast`)

Details

Different definitions of ARMA models have different signs for the AR and/or MA coefficients. The definition used here has

X[t] = a[1]X[t-1] + … + a[p]X[t-p] + e[t] + b[1]e[t-1] + … + b[q]e[t-q]

and so the MA coefficients differ in sign from those of S-PLUS. Further, if `include.mean` is true (the default for an ARMA model), this formula applies to *X - m* rather than *X*. For ARIMA models with differencing, the differenced series follows a zero-mean ARMA model. If am `xreg` term

Avec les notations du cours :

$$\begin{aligned}&(1-B)^d X_t - m \\ &= \mathtt{ar[1]}\left((1-B)^d X_{t-1} - m\right) + \ldots + \mathtt{ar[p]}\left((1-B)^d X_{t-p} - m\right) \\ &\qquad + \varepsilon_t + \mathtt{ma[1]}\varepsilon_{t-1} + \ldots + \mathtt{ma[q]}\varepsilon_{t-q}\end{aligned}$$

D'où :

$$\left(1 - \mathtt{ar[1]}\,B - \cdots - \mathtt{ar[p]}\,B^p\right)\left((1-B)^d X_t - m\right) = \left(1 + \mathtt{ma[1]}\,B + \cdots + \mathtt{ma[q]}\,B^q\right)\varepsilon_t$$

Séance 2

Conclusion

- Processus stationnaire ?
  - AR ≪ pur ≫ (sans racine unité)
  - MA ≪ pur ≫
  - ARMA

  → identification des ordres p / q : ACF et PACF.

- Processus non-stationnaire ?
  - ARIMA (si tendance)
  - SARIMA (si saisonnalité)

  → on ≪ stationnarise ≫ en dérivant

- Il faut aussi valider le modèle identifié
  (résidus b.b.g., significativité des coefficients, etc).

⟶ à pratiquer en TD !