Séries temporelles - Questions de cours

1 - Qu'est-ce qu'une processus stationnaire fort et un processus stationnaire faible ? À partir de données, comment procédez-vous pour savoir si un processus est stationnaire (faible)? Donner plusieurs caractérisations possibles en expliquant l'intérêt de ces caractérisations.

* Processus stationnaire fort $(X_t)_{t\in\mathbb{Z}}$
  → (X_t) invariant par translation i.e $\forall k\in\mathbb{N}, h\in\mathbb{Z}$, $(X_1, \dots, X_k) \overset{\mathcal{L}}{=} (X_{1+h}, \dots, X_{k+h})$

* Processus stationnaire faible $(X_t)_{t\in\mathbb{Z}}$
  → moyenne et autocovariances stables par translation
  ↳ * $t \longmapsto \mu_X(t) = \mu \in \mathbb{R} \quad \forall t\in\mathbb{Z}$
     * $t \longmapsto \gamma_X(t, t+h) = \gamma_X(h) \quad \forall t\in\mathbb{Z}$

* Reconnaître si un processus stationnaire est faible?
  → On trace l'autocorrélogramme (partiel) et on doit observer une décroissance exponentielle (sinon pas stationnaire)

* Caractérisat° possibles et intérêts :
  → moyenne stable, ainsi qu'importe l'info et où elle se trouve, on retrouve une valeur moyenne
  → Les observations gardent des corrélations stables entre elles pour un m̂ décalage.
  → On trace l'autocorrélogramme (partiel) et on doit observer une décroissance exponentielle (sinon pas stationnaire)

2 - En quoi consiste la méthode de désaisonnalisation par moyennes mobiles ? Donner ses limites

Moyennes Mobiles

Techniques :
- AR(p)
- MA(q)
- ARMA(p,q)

On suppose pas de saisonnalité

- arithmétique [2m+1] → annule saisonnalité 2m+1.
- ARIMA(p,d,q)
- SARIMA[(p,d,q)(P,D,Q)]
  ↓
  on travaille avec saisonnalité

⊕ * On peut travailler sur le terme d'erreur ε_t ou le terme aléatoire Y_t.
  * on peut annuler / conserver tendance cte / linéaire / exp / géométrique, les composantes saisonnières.

⊖ * on ne peut pas projeter (≠ par régression), c'est à l'utilisateur de construire le modèle.

3 - On suppose qu'il est possible de décomposer une série temporelle en trois termes :

$$X_t = m_t + S_t + Y_t$$

où m_t est une tendance déterministe, S_t est un saisonnalité déterministe et Y_t est un processus stationnaire.

Expliquer quelles sont les différentes méthodes que vous connaissez pour identifier chacun des termes sur des données. Donner les avantages et les défauts de chaque méthode.

① Méthode par regression

1- Méthode Buys-Ballot généralisée : effectuer une régression sur l'ensemble des données

⊕ * possibilité de projeter m_t et $\tilde{S}_t$ au delà de l'horizon T.
* estimateurs convergents (mais pas forcément optimaux).

⊖ * manque de flexibilité (représentation linéaire et stable dans le temps).
* Si m_t et $\tilde{S}_t$ erronés, alors erreurs donc estimat° des coeffs et des project°.

2- Méthode STL : effectuer des régressions linéaires localement

⊕ * possibilité d'ajuster le modèle plus finement autour du point étudié
* ajustement de la fenêtre d'observat° où on effectue localement la regression
* possibilité de pondérer les observat° (⊕ proche du point étudié, ⊕ pondéré !)
* la saisonnalité peut évoluer au cours du temps.

⊖ * Ne traite pas automatiquement les variations journalières ou récurrentes.
* Seulement utilisable pour décomposit° additive (pas multiplicative).

② Moyennes Mobiles

Techniques : - AR(p) - arithmétique [2m+1] ⟶ annule saisonnalité 2m+1.
- MA(q) - ARIMA(p,d,q)
- ARMA(p,q) - SARIMA[(p,d,q)(P,D,Q)]

On suppose pas de saisonnalité

↓ on travaille avec saisonnalité

⊕ * On peut travailler sur le terme d'erreur ε_t ou le terme aléatoire Y_t.
* on peut annuler / conserver tendance de / linéaire / exp / geométrique, les composantes saisonnières.

⊖ * on ne peut pas projeter (≠ par regression), c'est à l'utilisateur de construire le modèle.

4 - Expliquer ce qu'est la méthode d'estimation de Box et Jenkins. Donner les différentes étapes en expliquant leur intérêt, mais sans rentrer dans les détails mathématiques. Pour un modèle ARMA, expliquer la procédure de Box et Jenkins d'ajustement d'un modèle.

1- Transformations données : - homoscédasticité
- stationnarisation

2- Estimation des paramètres p et q par visualisation des autocorrélogrammes et autocorrélogrammes partiels.

3- Estimation des ≠ts modèles

4- Test d'adéquation des modèles (i.e vérifier les modèles et choisir un modèle).

5- Prévis° des valeurs futures à travers le modèle retenu.

5 - Quel est l'intérêt des processus ARMA par rapport aux processus AR ou MA ? Quelles sont les difficultés statistiques rencontrées pour identifier les ordres des processus ARMA par rapport à l'identification des ordres des processus AR ou MA ?

Intérêt ARMA p/rt aux AR et MA

→ moins de paramètres que les AR ou MA pour atteindre m̂ niveau de précis° en général
→ ARMA peut s'écrire comme AR(∞) et MA(∞) si racines de modules > 1.
→ Flexibilité et adaptabilité
→ Efficacité en termes de prédict°.

Difficultés statistiques pour identifier les ordres d'un ARMA p/rt à MA ou AR.

→ ARMA peut s'écrire comme AR ou MA, donc paramètres très reliés.
→ Les racines de Φ et Θ doivent être de modules > 1 (Et pour écriture canonique il faut Φ et Θ 1ers entre eux).

6 - Dans quel cadre utilise-t-on un modèle SARIMA plutôt qu'ARIMA ? Comment savoir à partir de données s'il vaut mieux utiliser un modèle plutôt que l'autre ?

ARIMA → permet de stationnariser un processus afin d'avoir un processus asympt. ARMA

SARIMA → généralisat° des modèles ARIMA contenant une partie saisonnière de nature aléatoire.

Rappel: X_t est ARIMA(p,d,q) si:
$\forall t, (1-L)^d X_t = Y_t$ avec $Y_t \sim ARMA(p,q)$

X_t est $SARIMA_s(p,d,q)$ si: (X_t) est de périodicité s.
$Y_t = (id-L)^d(id-L^s)X_t$ est un ARMA(p,q)
permet de désaisonnaliser la série.

3 critères: - parcimonie : nb de paramètres minimal
- prédict° : 1. R^2 coeff de détermination
2. $\bar{R}^2$ } max
3. σ^2 variance) min
4. Fisher) max
- informat° : 1- variance résiduelle
2 - AIC et AICC
3- BIC
4- Hannan-Quinn } min.

7 - Donner la définition d'un processus *ARMA(p,q)*.

Un processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ est un ARMA(p,q) si :

$$\Phi(L)X_t = \mu + \Theta(L)\varepsilon_t$$

avec $\Phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p \quad \varphi_p \neq 0$

$\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q \quad \theta_q \neq 0$

8 - Soit $X = (X_t)_{t\in\mathbb{Z}}$ un processus stationnaire.
Soit $X_t^* = EL(X_t|1, \underline{X_{t-1}})$ la projection orthogonale de X_t sur l'espace engendré par 1 et $X_{t-1}, X_{t-2}, \ldots$
Montrer que le processus défini par $\epsilon = X_t - X_t^*$ est un bruit blanc.

9 - Soit $\epsilon = (\epsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc de variance σ_ϵ^2. Montrer que sa densité spectrale vaut : $f_\epsilon(\omega) = \dfrac{\sigma_\epsilon^2}{2\pi}$.
En déduire l'écriture de la densité spectrale d'un processus ARMA(p, q).

$(\varepsilon_t) \sim BB(0, \sigma_\varepsilon^2) \iff \mathbb{E}[\varepsilon_t] = 0$

$\mathbb{V}[\varepsilon_t] = \sigma_\varepsilon^2$

$\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}) = 0 \quad \forall t \neq t'$

$$f_\varepsilon(\omega) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \gamma_\varepsilon(h)\, e^{i\omega h}$$

$$= \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \gamma_\varepsilon(h)\cos(\omega h) = \frac{1}{2\pi}\gamma_\varepsilon(0) \qquad \text{car } \forall h \neq 0,\ \gamma_\varepsilon(h) = 0$$

$$= \frac{1}{2\pi}\sigma_\varepsilon^2$$

$$\Rightarrow (\varepsilon_t) \sim BB(0, \sigma_\varepsilon^2) \longleftrightarrow f_\varepsilon(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}$$

Soit $(X_t)_{t\in\mathbb{Z}}$ un processus ARMA(p, q)

$\Rightarrow \exists \Phi$ et Θ tq $\Phi(L)X_t = \Theta(L)\varepsilon_t$ On suppose Φ inversible (|racines|>1)

$\Rightarrow X_t = \Theta(L)\Phi^{-1}(L)\varepsilon_t$

$$\Rightarrow f_X(\omega) = \frac{1}{2\pi}\left|\Theta(e^{i\omega})\Phi^{-1}(e^{i\omega})\right|^2 f_\varepsilon(\omega)$$

$$= \frac{\left|\Theta(e^{i\omega})\Phi^{-1}(e^{i\omega})\right|^2}{2\pi}\sigma_\varepsilon^2$$