# Séries temporelles – TD/TP 6

Estimer, comparer et diagnostiquer des modèles ARMA

M1 Actuariat – Université Gustave Eiffel

Durée indicative : 2 heures

**Objectifs**

Le TD propose deux situations concrètes : interpréter un AR(1) estimé et choisir entre plusieurs modèles en confrontant les critères et les résidus. Le TP applique ensuite cette démarche à une série réelle sous R. Les questions demandent de justifier les réponses, sans calculs techniques excessifs.

## Exercice 1 – Comprendre un AR(1) estimé

Une caisse d'assistance analyse une série hebdomadaire $(X_t)$ de demandes reçues. Sur $n = 150$ semaines, elle ajuste le modèle

$$X_t = c + \phi X_{t-1} + Z_t,$$

avec des innovations centrées. Le logiciel fournit les estimations

$$\widehat{c} = 21, \qquad \widehat{\phi} = 0{,}65, \qquad \mathrm{SE}(\widehat{\phi}) = 0{,}09, \qquad \widehat{\sigma}^2 = 25.$$

On suppose que l'AR(1) ajusté est stationnaire. Un collègue déclare : « Le nombre moyen de demandes est de 21 par semaine, car la constante vaut 21. »

1) **Interpréter les coefficients.** Le collègue a-t-il raison ? En prenant l'espérance dans l'équation du modèle, retrouver la moyenne $\mu$ en fonction de $c$ et $\phi$, puis calculer la moyenne estimée. Réécrire le modèle ajusté sous la forme

$$X_t - \widehat{\mu} = \widehat{\phi}(X_{t-1} - \widehat{\mu}) + \widehat{Z}_t.$$

2) **Examiner trois observations.** Voici trois semaines consécutives de la série :

| Semaine $t$ | 31 | 32 | 33 |
|---|---|---|---|
| Observation $x_t$ | 63 | 70 | 56 |

On sait aussi que $x_{30} = 65$. Pour chacune des semaines 31, 32 et 33, calculer la valeur ajustée $\widehat{x}_t = \widehat{c} + \widehat{\phi} x_{t-1}$, puis le résidu $\widehat{z}_t = x_t - \widehat{x}_t$. Expliquer, avec les mots du problème, ce que signifie un résidu négatif.

3) **Ne pas surinterpréter un résidu.** Le résidu de la semaine 33 est assez grand en valeur absolue. Peut-on en déduire à lui seul que le modèle est mauvais ? Donner une raison.

4) **Évaluer la précision et le diagnostic.** Que mesure l'erreur standard 0,09 ? Pourquoi est-elle différente de l'écart-type estimé des innovations ? Après ajustement, l'ACF des résidus présente une autocorrélation de $-0{,}23$ au retard 3, les autres premières autocorrélations étant proches de zéro. Le test de Ljung–Box à 12 retards donne une $p$-valeur de 0,018. Que conclure au seuil de 5 % ? Quelle prochaine étape raisonnable proposeriez-vous pour améliorer l'analyse ?

**Idée à retenir**

Un coefficient AR décrit une dépendance temporelle, il ne suffit pas à déterminer le niveau moyen. Et des coefficients bien estimés ne garantissent pas que le modèle ait expliqué toute l'autocorrélation de la série.

## Exercice 2 – Faut-il suivre AIC, BIC ou les résidus ?

On étudie une série de $n = 180$ observations mensuelles du nombre de dossiers ouverts par un assureur. Plusieurs modèles *stationnaires* sont estimés par maximum de vraisemblance sur les *mêmes observations*, selon les mêmes conventions. On obtient les résultats fictifs suivants :

| | Modèle | AIC | BIC |
|---|---|---|---|
| $A$ | AR(1) | 410,0 | 419,6 |
| $B$ | ARMA(1,1) | 408,0 | 420,8 |
| $C$ | SARIMA$(1,0,0)(1,0,0)_{12}$ | 408,4 | 421,2 |

Le diagnostic des résidus donne :

| Modèle | Autocorrélation au retard 12 | Ljung–Box ($m = 24$) : $p$-valeur | Chronogramme |
|---|---|---|---|
| $A$ | 0,30 | 0,002 | Dispersion globalement stable |
| $B$ | 0,25 | 0,013 | Dispersion globalement stable |
| $C$ | 0,04 | 0,39 | Dispersion globalement stable |

1) **Premier choix.** Si on ne comparait que $A$ et $B$, lequel retiendrait-on selon AIC ? Selon BIC ? Donner une explication statistique à ce désaccord, *sans recalculer les formules des critères.*

2) **Lire les diagnostics.** Que révèlent l'ACF au retard 12 et le test de Ljung–Box pour $A$ et $B$ ? Pourquoi une petite valeur de AIC ou de BIC ne suffit-elle pas ici pour conclure ?

3) **Comprendre le troisième modèle.** Quel élément du modèle $C$ a été ajouté pour prendre en compte une éventuelle dépendance annuelle ? Pourquoi ce choix est-il raisonnable pour une série *mensuelle* ?

4) **Prendre une décision.** Quel modèle conseilleriez-vous parmi $A$, $B$ et $C$, en tenant compte *à la fois* des critères et des diagnostics ? Justifier en quelques phrases. Peut-on affirmer que le modèle retenu est entièrement validé ?

**Idée à retenir**

AIC et BIC aident à comparer les modèles, mais ils ne remplacent pas l'examen des résidus. Un modèle légèrement moins bien classé peut être préférable si le modèle classé premier laisse une dépendance temporelle nette dans ses résidus.

## Exercice 3 – La série `Nile` : quel modèle conserve-t-on ?

La série annuelle `Nile`, disponible directement dans R, mesure le débit annuel du Nil à Assouan (1871–1970). Son intérêt est que la dynamique n'est *pas* donnée à l'avance : il faudra comparer des modèles et interpréter leurs diagnostics. Pour l'exercice, on travaillera sur les observations dans leur unité initiale, sans transformation logarithmique.

## A. Explorer la série et proposer des candidats

```
data(Nile)
x <- Nile
length(x)
start(x)
end(x)
frequency(x)
plot.ts(x, ylab = "Debit annuel", main = "Serie Nile")
acf(x, main = "ACF de Nile")
```

**Décodage du code R**

`data(Nile)` charge une série déjà fournie par R. `length(x)` donne sa taille. `frequency(x)` indique sa fréquence de mesure : ici annuelle. `acf(x)` trace les corrélations empiriques de $x_t$ et de ses valeurs passées à différents retards. Les barres horizontales du graphe sont des repères approximatifs de significativité, et non un test global.

1) Combien y a-t-il d'observations ? Décrire l'évolution du niveau et la dépendance temporelle visibles sur les deux graphiques. Est-ce qu'une saisonnalité trimestrielle ou mensuelle a du sens ici ?

2) Pour commencer, comparer les familles : bruit blanc autour d'une moyenne, AR(1), AR(2), ARMA(1,1). Quelle différence de structure y a-t-il entre elles ? Ne pas décider du meilleur modèle avant l'ajustement.

## B. Estimer et comparer par AIC/BIC

```
f0 <- arima(x, order = c(0, 0, 0), method = "ML")
f1 <- arima(x, order = c(1, 0, 0), method = "ML")
f2 <- arima(x, order = c(2, 0, 0), method = "ML")
f3 <- arima(x, order = c(1, 0, 1), method = "ML")

f1                # coefficients et erreurs standards
sqrt(diag(f1$var.coef))
AIC(f0, f1, f2, f3)
BIC(f0, f1, f2, f3)
```

**Décodage du code R**

- `order = c(p, d, q)` spécifie les ordres ordinaires du modèle ; ici $d = 0$ pour tous les candidats. Comme il s'agit de modèles non différenciés, R estime par défaut une moyenne (`include.mean = TRUE`).
- `method = "ML"` demande l'estimation par maximum de vraisemblance. `f1$var.coef` contient une matrice estimée de covariance des coefficients ; sa diagonale fournit leurs variances estimées.
- `AIC(...)` et `BIC(...)` comparent les ajustements sur la *même série* avec la *même méthode* d'estimation. Dans les deux cas, le plus petit critère est préféré.

3) Pour `f1`, identifier le coefficient AR estimé, l'estimation de la moyenne, la variance d'innovation estimée et l'erreur standard du coefficient AR. Expliquer la différence entre *écart-type des innovations* et *erreur standard d'un estimateur*.

4) À partir des tableaux AIC/BIC, relever le meilleur modèle selon chaque critère. Les critères sont-ils d'accord ? Donner une raison pour laquelle le modèle de meilleure log-vraisemblance n'est pas nécessairement le meilleur selon BIC.

## C. Vérifier ce que les modèles n'ont pas expliqué

```
# Les deux ajustements a examiner en priorite
r1 <- residuals(f1)
r3 <- residuals(f3)

par(mfrow = c(2, 2))
plot.ts(r1, main = "Residus AR(1)")
acf(r1, main = "ACF residus AR(1)")
plot.ts(r3, main = "Residus ARMA(1,1)")
acf(r3, main = "ACF residus ARMA(1,1)")
par(mfrow = c(1, 1))

Box.test(r1, lag = 12, type = "Ljung-Box", fitdf = 1)
Box.test(r3, lag = 12, type = "Ljung-Box", fitdf = 2)
```

**Décodage du code R**

- `residuals(f1)` extrait les résidus calculés après ajustement. `par(mfrow = c(2, 2))` organise quatre graphiques dans la même fenêtre ; la dernière commande rétablit un seul graphique.
- `lag = 12` examine simultanément les autocorrélations aux retards $1, \ldots, 12$. `fitdf` retranche les ordres AR et MA estimés : ici 1 pour AR(1) et 2 pour ARMA(1,1).
- Une *p-valeur* supérieure à 0,05 signifie que le test ne rejette pas l'absence d'autocorrélation aux retards étudiés ; cela ne démontre ni l'indépendance ni la validité de toutes les hypothèses du modèle.

5) Pour chaque modèle examiné, commenter le chronogramme des résidus et leur ACF. Des retards ressortent-ils clairement ? Les résidus paraissent-ils de dispersion stable ?

6) Lire les deux p-valeurs de Ljung–Box. Formuler *la décision du test* au seuil de 5 %, sans écrire que l'hypothèse nulle a été « prouvée ».

7) Donner votre conclusion : quel modèle retiendriez-vous parmi les candidats étudiés ? Appuyez-vous sur *trois arguments distincts* (critères, ACF, test). Citer enfin une limite que ces diagnostics ne permettent pas d'écarter.

## Tableau de synthèse à compléter après les calculs

Remplissez le tableau à partir de la sortie de `R` ; puis rédigez une conclusion argumentée de cinq à huit lignes.

| Modèle | AIC | BIC | ACF des résidus | Ljung–Box à 12 retards |
|---|---|---|---|---|
| Bruit blanc | | | | — |
| AR(1) | | | | |
| AR(2) | | | | — |
| ARMA(1,1) | | | | |

*Dans la dernière colonne, indiquez « rejet » ou « non-rejet » lorsque vous avez effectivement réalisé le test. Pour les modèles non testés, laissez un tiret : il ne faut pas inventer de diagnostic.*