UNIVERSITÉ DE NANTES

Master 2 Ingénierie Mathématique

Séries temporelles
Exercices

Ex 1. Soit $(\epsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc Gaussien standard (de variance 1), montrer que le processus

$$X_t = \begin{cases} \epsilon_t & \text{si } t \text{ est pair,} \\ (\epsilon_{t-1}^2 - 1)/\sqrt{2} & \text{si } t \text{ est impair,} \end{cases}, \quad t \in \mathbb{Z},$$

n'est pas une suite de variables aléatoires indépendantes mais que c'est un bruit blanc faible.

Ex 2. Soient $(X_t)_{t\in\mathbb{Z}}$ et $(Y_t)_{t\in\mathbb{Z}}$ deux processus stationnaires et mutuellement décorrélés, d'autocovariances respectives γ_X et γ_Y.

1) Montrer que le processus $(X_t+Y_t)_{t\in\mathbb{Z}}$ est stationnaire et exprimer son autocovariance γ_{X+Y} en fonction de γ_X et de γ_Y.

2) Proposer un contre-exemple montrant que, si l'on enlève l'hypothèse de non-corrélation entre $(X_t)_{t\in\mathbb{Z}}$ et $(Y_t)_{t\in\mathbb{Z}}$, alors $(X_t + Y_t)_{t\in\mathbb{Z}}$ n'est pas nécessairement stationnaire.

Ex 3. Soit $(\epsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc fort.

1. Ecrire une fonction de (n, ω, a, b) qui génère une trajectoire de taille n de la série

$$X_t = a\cos(\omega t) + bt + \epsilon_t, t = 1, ..., n.$$

2. Pour $a = 0$ et $b = 1$ et $\omega = 0$, simuler à l'aide de la fonction `diff()` une trajectoire de taille $n = 200$ de la série différenciée $Y_t = X_t - X_{t-1}$.

3. Calculer les autocorrélations empiriques de $(X_t)_{t=1,...,n}$ et de $(Y_t)_{t=1,...,n}$ (commande `acf()`).

4. Pour $a = 1$, $b = 0$ et $\omega = \pi/6$, simuler une trajectoire de taille $n = 200$ de la série différenciée à l'ordre 12, $Z_t = X_t - X_{t-12}$.

5. Représenter les acf des séries X_t, Y_t et Z_t.

Ex 4. Soit $(\epsilon_t)_{t\in\mathbb{N}}$ une suite de variables iid de loi $\mathcal{N}(0, 1)$ (un bruit blanc Gaussien).

1. A l'aide de la fonction `ts()`, simuler une trajectoire de taille 100 des séries temporelles suivantes:

a) $X_t = \epsilon_t$	*b)* $X_t = t + 3 + 10\epsilon_t$	*c)* $X_t = t^2 + 1000\epsilon_t$
d) $X_t = \epsilon_t + 0.5\epsilon_{t-1}$	*e)* $X_t = \epsilon_t - 0.5\epsilon_{t-1}$	*f)* $X_t = 0.5X_{t-1} + \epsilon_t$
g) $X_t = X_{t-1} + \epsilon_t$	*h)* $X_t = -0.9X_{t-1} + \epsilon_t$	*i)* $X_t = \sin(t\pi/6) + \epsilon_t$

2. Représenter les séries graphiquement ainsi que les acf et pacf empiriques.

Ex 5. On considère le processus (X_t) défini, pour tout $t \in \mathbb{Z}$, par

$$X_t = \frac{5}{6}\,\epsilon_{t-1} - \frac{1}{6}\,\epsilon_{t-2} + \epsilon_t$$

où (ϵ_t) est un bruit blanc de variance σ^2.

1. Montrer que (X_t) est un processus stationnaire et calculer sa fonction d'autocorrélation ρ.
2. Générer une trajectoire $(x_n = X_n(\omega))$ de votre modèle sur l'intervalle $\{1, \ldots, n\}$ pour $n = 300$.
3. Proposer un estimateur naturel $\widehat{\rho}$ de sa fonction d'autocorrélation ρ, et calculer $\widehat{\rho}(0), \ldots, \widehat{\rho}(5)$. Représenter les valeurs sous forme de peigne, et les comparer avec les sorties de la fonction `acf`.
4. Ces observations sont-elles cohérentes avec la théorie ?
5. Tester la stationnarité de votre trajectoire en combinant l'utilisation des fonctions `adf.test` et `kpss.test` de la librairie `tseries`.
6. Estimer vos paramètres en utilisant la fonction `arima`.

Ex 6. On considère le processus (Z_t) défini par

$$Z_t = \frac{3}{2} Z_{t-1} - \frac{1}{2} Z_{t-2} + \epsilon_t - \frac{1}{2} \epsilon_{t-1} \ , \ t \in \mathbb{Z}$$

où $(\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance $\sigma^2 > 0$.

1. Générer une trajectoire $z_t = Z_t(\omega)$ pour $t = 1, ..., 300$ et la représenter graphiquement. Que constate-t-on ?
2. Montrer que le processus se comporte comme une marche aléatoire. Est-il stationnaire ?
3. Les sorties de `adf.test` et `kpss.test` sont-elles conformes à la théorie ?
4. Proposer une transformation de données Z'_t de Z_t qui rende le processus $(Z'_t)_{t \in \mathbb{Z}}$ stationnaire.
5. Calculer la trajectoire $(z'_t = Z'_t(\omega))$ associée, et vérifier qu'elle est bien stationnaire par l'intermédiaire de `adf.test` et de `kpss.test`.
6. Observer et commenter le comportement de $\widehat{\rho}$ et de $\widehat{\alpha}$ à l'aide des fonctions `acf` et `pacf`.

Ex 7. La série "champagne.txt" représente les ventes v_t mensuelles de champagne.

1. Tracer la série et ses autocorrélations empiriques.
2. Peut-on détecter une tendance et une composante périodique dans cette série?
3. Quels filtres peut-on appliquer pour éliminer la tendance et la composante saisonnière?
4. Tracer la série obtenue après élimination de la tendance et de la saisonnalité, ainsi que la suite de ses autocorrélations empiriques.
5. La série obtenue peut-elle être modélisée par un bruit blanc?
6. Donner des estimations $\hat{m}_t$ et $\hat{v}_t$ de la tendance et de la composante saisonnière s_t.
7. Tracer la série des résidus $\hat{\epsilon}_t = v_t - \hat{m}_t - \hat{s}_t$ et ses autocorrélations empiriques. Est-ce un bruit blanc?
8. Proposer une prévision à l'ordre 12 des ventes de champagne.

Ex 8. À l'aide de l'algorithme de Durbin-Levinson, calculer les 4 premières valeurs de l'autocorrélation partielle $\alpha(0), \ldots, \alpha(3)$ d'un processus stationnaire en fonction de son autocovariance γ.

Ex 9. La série "varve.txt" représente l'épaisseur de 634 varves de la période glacière. Les varves sont des dépôts sédimentaires issus de la fonte des glaciers, dont l'épaisseur est corrélée positivement avec la température moyenne de l'année.

1. En utilisant la fonction `rollapply()` de la librairie `zoo`, calculer les moyennes et variances glissantes de la série varve. La série semble-t-elle stationnaire?

2. Pour stabiliser la variance, on considère le logarithme de la série. La série Y_t obtenue semble-t-elle stationnaire?

3. Représenter la série différenciée $Z_t = Y_t - Y_{t-1}$ ainsi que ses autocorrélations empiriques. Que vaut l'estimation de la corrélation entre Z_t et Z_{t-1}? entre Z_t et Z_{t-2}? La série est-elle un bruit blanc?

4. Proposer une modélisation de la série Z_t. Comment estimer les paramètres du modèle à partir des acf?

5. Proposer une prédiction de la série à l'année suivante en utilisant que si X_t est un MA(1): $X_t = \epsilon_t + a\epsilon_{t-1}$ avec $a \in]-1, 1[$ et (ϵ_t) un bruit blanc faible, alors la meilleure prévision linéaire de X_{t+1} sachant le passé est
$$\hat{X}_{t+1} = a \sum_{k \geq 0} (-a)^k X_{t-k}.$$

Ex 10. On sait que si $(X_t)_{t \in \mathbb{Z}}$ est une série chronologique stationnaire centrée, alors son innovation définie pour $t \in \mathbb{Z}$ par
$$\eta_t = X_t - \Pi_{H_{t-1}}(X_t)$$
est un bruit blanc, où l'on a noté $\Pi_{H_{t-1}}$ l'opérateur de projection dans $\mathbb{L}^2$ sur $H_{t-1} = \text{vect}\{X_s, s \leq t-1\}$.

1. Vérifier que le processus défini, pour $t \in \mathbb{Z}$, par
$$X_t = \epsilon_t - \frac{5}{6}\,\epsilon_{t-1} - \frac{1}{18}\,\epsilon_{t-2} + \frac{1}{9}\,\epsilon_{t-3}$$
où $(\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2, est un processus stationnaire centré admettant une écriture causale. En déduire que $(\epsilon_t)_{t \in \mathbb{Z}}$ est le bruit blanc d'innovation de $(X_t)_{t \in \mathbb{Z}}$.

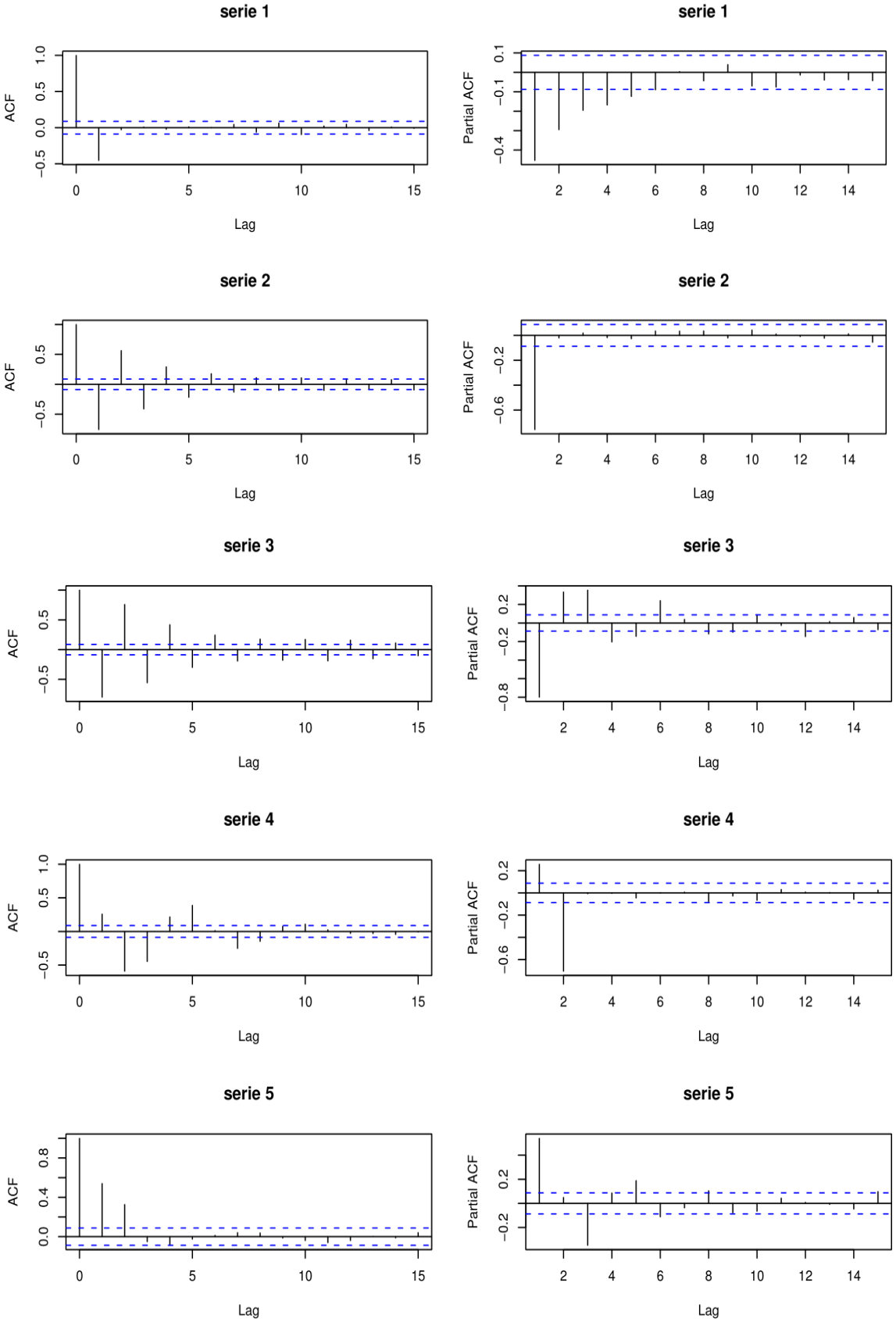
2. Simuler une trajectoire $(x_n = X_n(\omega))$ sur l'intervalle $\{1, \ldots, n\}$ pour n assez grand, en prenant soin d'estomper l'effet des valeurs initiales (par exemple en le simulant sur l'intervalle $n + d$ avant de supprimer les d premières valeurs).

3. Nous souhaitons estimer le processus d'innovation $(\eta_t)_{t \in \mathbb{Z}}$ sur la trajectoire (x_n). Pour cela, on va noter
$$\widehat{\eta}_k = X_k - \widehat{p}_1\, X_{k-1} - \widehat{p}_2\, X_{k-2} - \ldots - \widehat{p}_\ell\, X_{k-\ell}$$
pour tout $\ell + 1 \leq k \leq n$, où $\widehat{p}_1, \ldots, \widehat{p}_\ell$ sont les estimateurs des moindres carrés associés à la régression linéaire de X_k sur ses k valeurs passées. Pour différentes valeurs de ℓ, estimer le bruit blanc d'innovation.

4. Etudier la blancheur des trajectoires $\widehat{\eta}_{\ell+1}, ..., \widehat{\eta}_n$ en fonction des valeurs de ℓ.

5. Estimer la variance empirique de l'innovation et la comparer avec σ^2.

Ex 11. Les graphiques suivant représentent les autocorrélations et autocorrélations partielles empiriques de trajectoires de longueur $n = 500$ de processus ARMA.



Indentifier les processus AR, MA et ARMA en précisant les ordres dans chaque cas. On pourra utiliser les fonctions:

- `ARMAacf` pour calculer les acf et pacf théoriques d'un processus ARMA
- `arima.sim` pour simuler des trajectoires d'un processus ARMA
- `arima` pour estimer les paramètres d'un modèle ARMA

Ex 12. La série "globtemp.txt" représente les différences entre les températures annuelles depuis 1856 et la moyenne constatée entre 1961 et 1990. On cherche à mettre en évidence un éventuel réchauffement climatique significatif au 20ème siècle.

1. Importer la série sous R et la mettre au format ts.
2. A l'aide de la fonction `window()`, ne conserver que les données depuis 1900. Effectuer une régression linéaire de la série en fonction du temps. Une tendance positive significative apparaît-elle?
3. Représenter les résidus de la régression précédente. Effectuer un test d'autocorrélation des résidus (par exemple Durbin-Watson ou Breush-Godfrey de la librairie `lmtest`).
4. Représenter le nuage de points des résidus à l'instant t en fonction des résidus à l'instant $t-1$. Représenter les autocorrélations empiriques des résidus. Cela confirme-t-il les résultats de la question précédente?
5. En quoi le résultat du test remet-il en cause la significativité de tendance de la question 2?
6. On admet que les résidus ϵ_t du modèle suivent un modèle autorégressif d'ordre 1, c'est-à-dire une relation de la forme $\epsilon_t = a\epsilon_{t-1} + \xi_t$ où $(\xi_t)_{t\in\mathbb{N}}$ est un bruit blanc et $a \in]-1,1[$. Effectuer une régression par moindres carrés généralisés en supposant une structure d'AR(1) des résidus (utiliser la fonction `gls` de la librairie `nlme` en précisant `correlation=corAR1()`).
7. Une tendance positive apparaît-elle suite à cette nouvelle estimation?
8. Récupérer les résidus normalisés par la commande `residuals(.,type="normalized")`. Refaire le test d'autocorrélation. Le modèle est-il satisfaisant?

Ex 13. Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stationnaire sur $\mathbb{Z}$ de moyenne m et d'autocovariance γ, dont on extrait le vecteur $(X_1,\ldots,X_n)$. On définit la moyenne empirique du processus par

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

1) Montrer que $\overline{X}_n$ est un estimateur non biaisé de m.

2) Montrer que $\overline{X}_n$ converge en probabilité vers m dès que $\gamma(n) \to 0$ quand $n \to \infty$.

3) Si on suppose de plus que $\sum_{h\in\mathbb{Z}} |\gamma(h)| < \infty$, montrer que

$$\lim_{n\to\infty} n \ \mathrm{var}(\overline{X}_n) = \sum_{h\,\in\,\mathbb{Z}} \gamma(h).$$

4) Proposer un estimateur naturel de l'autocovariance γ, de l'autocorrélation ρ et de l'autocorrélation partielle α d'un processus stationnaire.

Ex 14. Soit le polynôme de degré n défini, pour tout $z \in \mathbb{C}$, par

$$P(z) = 1 - p_1\, z - p_2\, z^2 - \ldots - p_n\, z^n.$$

Montrer par l'absurde que la condition $\sum_{i=1}^{n} |p_i| < 1$ est suffisante pour que P soit causal.

Ex 15. Soient $(\epsilon_t)_{t\in\mathbb{Z}}$ un bruit blanc de variance σ^2 et $(\eta_t)_{t\in\mathbb{Z}}$ un bruit blanc de variance $\delta^2\sigma^2$, pour $0 < |\delta| < 1$. On considère les processus MA(1) sur $\mathbb{Z}$ engendrés par les relations

$$X_t = \epsilon_t - \delta\,\epsilon_{t-1} \quad \text{et} \quad Y_t = \eta_t - \frac{1}{\delta}\,\eta_{t-1}.$$

1) Montrer que $(X_t)_{t\in\mathbb{Z}}$ et $(Y_t)_{t\in\mathbb{Z}}$ ont la même fonction d'autocovariance.

2) Montrer que $(\epsilon_t)_{t\in\mathbb{Z}}$ est le bruit blanc d'innovation de $(X_t)_{t\in\mathbb{Z}}$ mais que $(\eta_t)_{t\in\mathbb{Z}}$ n'est pas celui de $(Y_t)_{t\in\mathbb{Z}}$.

3) Montrer que le processus $\widetilde{\eta}_t = (I - \delta B)^{-1}(I - \delta^{-1}B)\,\eta_t, t \in \mathbb{Z}$ est un bruit blanc de même variance que $(\epsilon_t)_{t\in\mathbb{Z}}$, et qu'il s'agit de l'innovation de $(Y_t)_{t\in\mathbb{Z}}$.

Ex 16. Soit le processus défini par

$$X_t = aX_{t-1} + \epsilon_t \ , \ t \in \mathbb{Z},$$

où $(\epsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc de variance σ^2 et $a \in \mathbb{R}$.

1) Montrer que la condition $|a| \neq 1$ implique la stationnarité de $(X_t)_{t\in\mathbb{Z}}$. Quelle condition supplémentaire implique la causalité de $(X_t)_{t\in\mathbb{Z}}$?

2) On suppose maintenant que le processus est défini seulement pour $t \in \mathbb{N}$ avec comme valeur initiale $X_0 \in \mathbb{L}^2$. Montrer que le processus n'est plus nécessairement stationnaire.

3) Montrer qu'un tel processus est asymptotiquement stationnaire dès que $|a| < 1$.

4) Quelles conditions faut-il ajouter pour que le processus défini sur $\mathbb{N}$ soit stationnaire?

Ex 17. Modéliser par des processus ARIMA les séries "simul1", "simul2", "simul3" et "simul4" contenues dans les fichiers du même nom. Représenter la matrice de la méthode du coin pour visualiser les ordres p, q dans chaque cas.

Ex 18. Importer la série "voiture.txt" et ne conserver que les données avant la rupture. Mettre en évidence la présence d'une tendance et d'une composante saisonnière.

1. A l'aide de la commande `stl(,s.window="periodic")`, estimer la tendance et la composante saisonnière de cette série. La série des résidus peut-elle être modélisée par un bruit blanc?

2. Eliminer la tendance et la composante saisonnière en appliquant un ou plusieurs filtres linéaires de la forme $(I - L^s)^d$ (on pourra utiliser la fonction `filter()`). La série obtenue peut-elle être modélisée par un bruit blanc?

Ex 19. Extraire la série `DAX` des données `EuStockMarkets`, que l'on note X_t.

1. Proposer des transformations de la série pour tenter de la rendre stationnaire.

2. Ajuster un modèle GARCH(1, 1) sur les données $Y_t = \ln(X_t) - \ln(X_{t-1})$ (utiliser la fonction `garch`).

3. Extraire les résidus du modèle précédent. Est-ce un bruit blanc?

4. Déduire un bon modèle sur les données X_t.

Ex 20. Modéliser par un processus SARIMA la production mensuelle australienne d'électricité (en millions de kWh) de 1956 à avril 1990 contenue dans le fichier "elec_austr.txt".

Ex 21. On considère le jeu de données `sanfran.dat` contenant les précipitations mesurées mensuellement dans la région de San-Francisco, entre 1932 et 1966.

1. Observer les données, mener une étude de stationnarité sur les données éventuellement différenciées.

2. Proposer un ensemble de modèles ARIMA saisonniers (SARIMA) permettant de modéliser la série chronologique, pour différentes valeurs de la périodicité (par ex: 2 ans, 6 mois ...).

3. Superposer les données réelles et les données modélisées de façon lisible (couleurs différentes, zoom sur un créneau de la série).

4. Supprimer les 3 dernières années de données et les prédire à l'aide de la commande `predict`. Superposer les données réelles et les données prédites.

5. Reprendre la question précédente, mais en supposant que les données intermédiaires sont séquentiellement observées (ce qui revient à faire une série de prédictions à horizon 1 en réestimant le modèle autant de fois que nécessaire).

6. Remarquer visuellement l'existence de changements dans le comportement global de la série, puis expliquer la raison pour laquelle une fenêtre glissante comme historique de modélisation est susceptible d'améliorer les résultats. Mettre en pratique cette méthode pour prédire à nouveau séquentiellement les 3 derniers mois de précipitations.

Ex 22. On veut comparer les différents lissages exponentiels.

1. Simuler une trajectoire de taille $n = 100$ des séries $X_t = \epsilon_t + 1, Y_t = 2t + \epsilon_t + 2\epsilon_{t-1} - \epsilon_{t-3}$ et $Z_t = t + 2\cos(\pi t) + \epsilon_t - \epsilon_{t-1}$ où $(\epsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc Gaussien.

2. Pour chaque série, effectuer la prédiction des 10 dernières valeurs à partir des 90 premières par lissage exponentiel simple, double, puis de Holt-Winters. Comparer les valeurs prédites avec les vraies valeurs mises de côté. Quel lissage semble le mieux adapté à chaque série?

Ex 23. Le fichier `co2` contient les concentrations mensuelles en CO2 à proximité du volcan Mauna Loa (Hawaï) de 1959 à 1997.

1. Ajuster un modèle sur les données.

2. Tester la prédiction obtenue par le modèle sur les 12 dernières valeurs de la série, construite à partir des valeurs passées.

3. Comparer avec la prédiction par un lissage de Holt-Winters.

Ex 24. Proposer une modélisation des données `AirPassengers`.