

## Fiche de TD/TP : Introduction

### Ex.1. Analyse empirique de séries

1. Visualiser les séries suivantes ; quels commentaires pouvez-vous tirer de l'observation de ces séries ?
  - (a) *airpass* : nombre mensuel de passagers aériens de janvier 1949 à décembre 1960 ;
  - (b) *beer* : production mensuelle de bière en Australie ;
  - (c) *deaths* : morts accidentelles aux Etats Unis de janvier 1973 à décembre 1978 ;
  - (d) *lake* : niveau annuel du lac Huron de 1875 à 1972 ;
  - (e) *lynx* : nombre annuel de lynx chassés au Canada de 1821 à 1934 ;
  - (f) *signal* : série simulée :  $X_t = \cos(t) + \varepsilon_t, \varepsilon_t \stackrel{i.i.d.}{\rightsquigarrow} \mathcal{N}(0, 0.5^2), t \in \{0.1, \dots, 20.0\}$  ;
  - (g) *strike* : nombre annuel de grèves aux Etats Unis de 1951 à 1980 ;
  - (h) *sunspot* : nombre de taches solaires de 1770 à 1869 ;
  - (i) *uspop* : population des Etats Unis de 1790 à 1990 (pas de temps : 10 ans) ;
  - (j) *wine* : ventes mensuelles de vin rouge australien de janvier 1980 à octobre 1991.
2. Visualiser à nouveau la série *wine*. Effectuer une transformation logarithmique sur cette série.
3. Représentez les fonctions d'autocorrélations de ces séries à l'aide de la procédure ARIMA, en particulier sur la dernière série transformée.

### Ex.2. 1. Soit la série $(X_t)_{t \in \mathbb{Z}}$ où

$$X_t = 100 + 3t + 50 \cos(2\pi t/3) + \varepsilon_t,$$

où  $(\varepsilon_t)_{t \in \mathbb{Z}}$  est une suite i.i.d. de variables  $\mathcal{N}(0, 4)$ . Créer une table SAS contenant des valeurs simulées de  $X_t$  pour  $1 \leq t \leq 150$ .

2. A partir de la table précédente et de la procédure REG, estimer la tendance de la série.
3. Proposer une méthode d'estimation de la composante saisonnière, et la mettre en oeuvre.
4. Tracer le graphe des résidus estimés.

### Ex.3. Simulation de séries et exemples de filtrage.

1. Créer les séries suivantes dans une table SAS (pour  $t \in \{1, \dots, 200\}$ ) :
  - (a)  $T_t$  : une tendance linéaire :  $T_t = 100 + 3t$
  - (b)  $S_t^1$  : une série périodique de période 3 :  $S_t^1 = 10 \cos\left(\frac{2\pi}{3}t\right)$
  - (c)  $S_t^2$  : une série périodique de période 4 :  $S_t^2 = 10 \sin\left(\frac{\pi}{4} + \frac{\pi}{2}t\right)$
  - (d)  $S_t$  : une série périodique de période 12 :  $S_t = S_t^1 + S_t^2$
  - (e)  $\varepsilon_t$  : une perturbation (bruit blanc fort) suivant une loi normale :  $\varepsilon_t \rightsquigarrow N(0, 4)$  avec  $V(\varepsilon_t) = 4$
  - (f)  $X_t$  :  $X_t = T_t + S_t + \varepsilon_t$
2. Exemples du cours.
  - a. Donner le graphe de l'ACF empirique de  $\varepsilon_t$ .

Comme  $\rho(h) = 0$ , pour  $h > 0$ , on attend donc une ACF empirique proche de 0 pour  $h > 0$ . En fait, on verra que l'ACF empirique  $(\hat{\rho}_n(h), h > 0)$  d'un bruit blanc fort de variance finie est asymptotiquement (i.e. pour  $n$  grand) normale  $\mathcal{N}(0, 1/n)$ . Par conséquent, approximativement 95% de l'ACF empirique devrait se trouver entre les bornes  $\pm 1.96/\sqrt{n}$ . Sur le graphe de l'ACF, combien de valeurs de  $\hat{\rho}_n(h)$  peuvent tomber hors de ces bornes, sans que ça ne remette en question la nature de la série ?

b. On considère la marche aléatoire définie par :

$$S_0 = 0 \text{ et } S_t = \sum_{j=1}^t \varepsilon_j \text{ pour } t > 0$$

Simuler la série associée.

Observe-t-on la non-stationnarité démontrée théoriquement en cours ?

3. A l'aide de la procédure EXPAND, appliquer les moyennes mobiles suivantes centrées d'ordre 3, 4 et 12 sur les six séries de la question 1).

(a)  $M_3 = \frac{1}{3}F(I + B + B^2 + B^3)$

(b)  $M_4 = \frac{1}{4}F^2(\frac{1}{2}I + B + B^2 + B^3 + \frac{1}{2}B^4)$

(c)  $M_{12} = \frac{1}{12}F^6(\frac{1}{2}I + B + B^2 + B^3 + B^4 + B^5 + B^6 + B^7 + B^8 + B^9 + B^{10} + B^{11} + \frac{1}{2}B^{12})$

Quelles conclusions pouvez vous tirer ?

**Ex.4.** On définit les moyennes mobiles suivantes,

$$M_1 = \frac{1}{2}B(I + F),$$

$$M_2 = \frac{1}{4}B(I + F + F^2 + F^3).$$

1. Quelles sont les séries absorbées par  $M_1$  ? Invariantes ? Calculer le pouvoir de réduction de la variance de  $M_1$ .
2. Même question pour  $M_2$ .
3. Montrer que  $M_1 M_2$  laisse invariant les tendances linéaires.
4. Quelles sont les séries temporelles absorbées par  $M_2 M_2$  ?

**Ex.5. Etude théorique de la stationnarité de processus.**

Soit  $\{\varepsilon_t\}_{t \in \mathbb{N}}$  une suite de v.a.r. i.i.d. de loi  $N(0, \sigma^2)$ .

On pose  $X_t = \varepsilon_t$  et  $Y_t = (-1)^t \varepsilon_t, t \in \mathbb{N}$ .

1. a. Les processus  $X$  et  $Y$  sont-ils stationnaires à l'ordre 2 ?  
b. Sont-ils strictement stationnaires ?
2. Qu'en est-il du processus  $(X_t + Y_t, t \in \mathbb{N})$  ?
3. Même question pour le processus bivariable  $((X_t, Y_t), t \in \mathbb{N})$ .

**Ex.6.** Soit  $(\varepsilon_t)_{t \in \mathbb{Z}}$  un bruit blanc de variance  $\sigma^2$ . Etudier la stationnarité des processus  $(X_t)_{t \in \mathbb{Z}}$  et  $(Y_t)_{t \in \mathbb{Z}}$  suivants.

1.  $X_t = a + b\varepsilon_t + c\varepsilon_{t-1}$ , où  $(a, b, c) \in \mathbb{R}^3$ .
2.  $Y_t = \varepsilon_t \cos(\omega t) + \varepsilon_{t-1} \sin(\omega t)$ , où  $\omega \in [0, 2\pi[$ .

**Ex.7.** Soit  $(\varepsilon_t)_{t \in \mathbb{Z}}$  un bruit blanc de variance  $\sigma^2$ . Pour  $t \in \mathbb{Z}$ , on définit, pour  $(a, b) \in \mathbb{R}^2$ ,

$$X_t = at + b + \varepsilon_t.$$

1. A quelle condition  $(X_t)_{t \in \mathbb{Z}}$  est-il stationnaire ?
2. Etudier la stationnarité du processus "différence première"  $(Y_t)_{t \in \mathbb{Z}}$  :

$$Y_t = X_t - X_{t-1}.$$

**Ex.8.** Soit  $(\varepsilon_t)_{t \in \mathbb{Z}}$  un bruit blanc de variance  $\sigma^2$ . On définit, pour  $t \in \mathbb{Z}$ ,

$$X_t = \sum_{i=0}^N \theta_i \varepsilon_{t-i},$$

où  $N \in \mathbb{N}^*$  et  $(\theta_0, \theta_1, \dots, \theta_n) \in \mathbb{R}^{N+1}$ .

1. Calculer  $E[X_t]$ .
2. Calculer la fonction d'autocovariance de  $(X_t)_{t \in \mathbb{Z}}$ .

**Ex.9.** Soit  $(\varepsilon_t)_{t \in \mathbb{Z}}$  un bruit blanc de variance  $\sigma^2$ . Pour  $t \in \mathbb{N}$ , on définit

$$X_t = \sum_{i=0}^t (-1)^i \varepsilon_i.$$

1. Le processus  $(X_t)_{t \in \mathbb{N}}$  est-il du second ordre ?
2. Calculer  $\text{Cov}(X_t, X_{t+h})$  pour  $t \in \mathbb{N}$  et  $t+h \in \mathbb{N}$ .
3. Le processus  $(X_t)_{t \in \mathbb{N}}$  est-il stationnaire ?

**Ex.10.** Soit  $(\varepsilon_t)_{t \in \mathbb{Z}}$  un bruit blanc de variance  $\sigma^2$ . On définit, pour  $t \in \mathbb{Z}$ , et  $\theta \in \mathbb{R}$  tel que  $0 < |\theta| < 1$ ,

$$X_t = \varepsilon_t - \theta \varepsilon_{t-1}.$$

1. Calculez l'autocorrélogramme simple de  $(X_t)_{t \in \mathbb{Z}}$ .
2. Soit  $\Psi$  la fonction définie sur  $\mathbb{Z}$  par  $\Psi(0) = 1$ ,  $\Psi(\pm 1) = \rho$ ,  $\Psi(h) = 0$  pour  $|h| > 1$ . Si  $|\rho| \leq 1/2$ , trouver un processus stationnaire dont  $\Psi$  est l'autocorrélogramme simple. Dans le cas  $|\rho| > 1/2$ , montrer qu'il n'existe pas de processus stationnaire dont  $\Psi$  soit l'autocorrélogramme.

**Ex.11.** Soit  $(\varepsilon_t)_{t \in \mathbb{Z}}$  un bruit blanc de variance  $\sigma^2$ . Etudier la stationnarité des processus suivants.

1.  $X_t = \varepsilon_t \varepsilon_{t-1}$  pour  $t \in \mathbb{Z}$ , lorsque  $(\varepsilon_t)_{t \in \mathbb{Z}}$  est un bruit blanc fort. Que peut-on dire si  $(\varepsilon_t)_{t \in \mathbb{Z}}$  bruit blanc faible ?
2.  $Y_t = \sum_{i=0}^t \lambda^i (\varepsilon_{t-i} - \varepsilon_{t-i-1})$ , pour  $\lambda \in \mathbb{R}$ .
3. Pour  $|\lambda| < 1$ ,  $Z_t = \sum_{i=0}^{+\infty} \lambda^i (\varepsilon_{t-i} - \varepsilon_{t-i-1})$ . Montrer par ailleurs que  $(Z_t - Y_t) \rightarrow_{t \rightarrow \infty} 0$  dans  $\mathbb{L}^2$ .

**Exercice 12.** Soit  $(\varepsilon_t)_{t \in \mathbb{Z}}$  un bruit blanc de variance  $\sigma^2$ . On définit, pour  $t \in \mathbb{Z}$ ,

$$X_t = a + bt + s_t + \varepsilon_t,$$

où  $s_t$  est un processus saisonnier (périodique) de période 4, indépendant de  $\varepsilon_t$ , satisfaisant  $s_0 + s_1 + s_2 + s_3 = 0$ . On définit l'opérateur

$$M_4 : \left( (Z_t)_{t \in \mathbb{Z}} \rightarrow \left( \frac{Z_t + Z_{t-1} + Z_{t-2} + Z_{t-3}}{4} \right)_{t \in \mathbb{Z}} \right).$$

1. Soit  $Y = M_4 X$ . Donner l'expression de  $(Y_t)_{t \in \mathbb{Z}}$  et justifier l'intérêt de la transformation  $M_4$ .
2. Soit  $W_t = Y_t - Y_{t-1}$ . Etudier la stationnarité de  $W$  et calculer sa fonction d'autocovariance.

**Ex.13. Processus gaussiens.**

Soit  $X$  une v.a. suivant une loi normale centrée réduite. Soit  $c > 0$ . On pose :

$$X_c = X(1_{(|X| > c)} - 1_{(|X| \leq c)}).$$

1. Déterminer la loi de  $X_c$ .

2. Calculer  $cov(X, X_c)$  et montrer que  $\exists c_0 > 0 : cov(X, X_{c_0}) = 0$ .
3. Le vecteur  $(X, X_{c_0})$  est-il gaussien ? Montrer que  $X$  et  $X_{c_0}$  ne sont pas indépendants.

**Ex.14.**

Soient  $A$  et  $B$  deux v.a.r. i.i.d., de loi normale centrée réduite. On considère le processus  $X = (X_t, t \in R)$ , où  $X_t(\omega) = \cos(2\pi t)A(\omega) + \sin(2\pi t)B(\omega)$ ,  $\omega \in \Omega$ .

1. a. Calculer la moyenne puis la fonction d'autocovariance du processus  $X$ .  
b. Ce processus est-il faiblement stationnaire ?
2. a. Calculer la loi de la v.a.r  $X_t$ .  
b. Le processus  $X$  est-il gaussien ? Est-il fortement stationnaire ?
3. Simuler 100 observations de la série  $X$ .  
a. Tracer le graphe de la série et celui de l'ACF associée. Retrouve-t-on la propriété 1b ?  
b. Quel test proposez-vous pour retrouver la propriété 2b ?
4. On pose  $A = R \cos(2\pi\theta)$  et  $B = R \sin(2\pi\theta)$ , avec  $0 \leq \theta \leq 1$  et  $R > 0$ .  
Quelle est la loi de  $R$ ?, de  $\theta$  ?

**Ex.15. Décomposition saisonnière.**

Effectuer la décomposition saisonnière de la série *airpass* (on utilisera la transformation logarithmique). Utiliser la méthode de Shiskin pour effectuer une nouvelle décomposition saisonnière. Comparer les résultats obtenus.

Principe de la décomposition saisonnière

Soit une série pouvant s'écrire sous la forme suivante (modèle dit " additif " ) :

$$X_t = T_t + S_t + \varepsilon_t$$

où  $T_t$  désigne la tendance,  $S_t$  la saisonnalité et  $\varepsilon_t$  la perturbation.

Rq : dans la suite, on considérera une série de saisonnalité 12.

1. Estimation de la tendance

La tendance est estimée par :

$$\widehat{T}_t = M_{12} X_t$$

La moyenne mobile  $M_{12}$  annule en effet les séries périodiques d'ordre 12 et conserve les tendances linéaires.

2. Estimation de la saisonnalité

La saisonnalité est estimée (dans une première étape) par :

$$\widetilde{S}_t = X_t - \widehat{T}_t$$

3. Calcul des coefficients saisonniers

Douze coefficients sont estimés par les moyennes suivantes :

$$\forall j \in \{1, \dots, 12\} : s(j) = \overline{\widetilde{S}_t} \text{ pour } t \equiv 0 [j]$$

4. Normalisation des coefficients saisonniers

$$\widehat{s}(j) = s(j) - \bar{s} \text{ avec } \bar{s} = \frac{1}{12} \sum_{j=1}^{12} s(j)$$

A partir de ces coefficients saisonniers normalisés, il est possible d'obtenir le cycle des valeurs saisonnières :

$$\widehat{S}_t = \widehat{s}(j) \text{ pour } t \equiv 0 [j]$$

On peut ainsi calculer la série corrigée des variations saisonnières :

$$X_t^{CVS} = X_t - \widehat{S}_t$$

5. Estimation des perturbations

$$u_t = X_t - \widehat{T}_t - \widehat{S}_t = X_t^{CVS} - \widehat{T}_t$$

### Méthode de Shiskin

Cette méthode consiste à réestimer la tendance (de l'étape 1 précédente) à l'aide d'une moyenne mobile appliquée à la série corrigée des variations saisonnières (calculées lors de l'étape 5 précédente). Les étapes 2 à 6 sont ensuite réitérées à partir de cette nouvelle estimation de la tendance.

$$\widehat{T}_t = M_5 X_t \text{ avec } M_5 = \frac{1}{5} F^2 (I + B + B^2 + B^3 + B^4)$$

Rq : cette méthode permet de diminuer la variance des perturbations grâce au pouvoir réducteur de la moyenne mobile  $M_5$ .

### Ex.16. Méthode de différenciation.

Soit  $Y = (Y_t, t \in \mathbb{Z})$  processus stationnaire centré.

Soient  $m_t := a + bt$ , avec  $a$  et  $b$  constantes, et  $s_t$  saisonnalité de période 12.

1. a. Montrer que le processus  $Z = (Z_t, t \in \mathbb{Z})$  défini par  $Z_t := \nabla \nabla_{12} (m_t + s_t + Y_t)$  est stationnaire et donner sa fonction d'autocovariance en fonction de celle de  $Y$ .  
b. Application SAS sur la série *death*.
2. Même question pour le processus  $W$  défini par  $W_t := \nabla_{12} \nabla_{12} (m_t s_t + Y_t)$ .

### Ex.17. L'ACF comme indicateur de non-stationnarité.

Soit  $(x_1, \dots, x_n)$  les valeurs observées d'une série temporelle  $X$  aux temps  $1, \dots, n$ .

1. Supposons  $x_t = a + bt$ , avec  $a$  et  $b$  constantes,  $b \neq 0$ .  
Montrer que à  $h$  fixé ( $h \geq 1$ ),  $\lim_{n \rightarrow \infty} \hat{\rho}_X^{(n)}(h) = 1$ .
2. Supposons  $x_t = \cos(\lambda t)$ , avec  $\lambda$  constante telle que  $\lambda \in (-\pi, \pi]$ .  
Montrer que à  $h$  fixé ( $h \geq 1$ ),  $\lim_{n \rightarrow \infty} \hat{\rho}_X^{(n)}(h) = \cos(\lambda h)$ .
3. Application : série *wine* (Cf. Ex.1, question 1j).

### Ex.18. Tests sur les résidus : exemple simple.

On considère  $n = 200$  observations de la série *signal* dont on veut savoir si elles proviennent d'une suite i.i.d. Pour cela, on veut effectuer un certain nombre de tests.

1. Construire sur SAS le test de l'ACF.
2. Construire d'autres tests en fonction des valeurs données ci-dessous.

$$Q_n = n \sum_{h=1}^{20} (\hat{\rho}^{(n)}(h))^2 = 50,40; T_n = 138; S_n = 101; P_n = 10310.$$

Quelle est votre conclusion ?

### Ex.19. Simulation d'échantillons aléatoires.

1. Démontrer le théorème suivant, appelé théorème d'inversion :  
Soit  $F$  une fonction de répartition continue sur la droite réelle. Pour  $u$  compris entre 0 et 1, on définit  $F^{-1}(u) = \inf\{x | F(x) = u\}$ . Si  $U$  suit une loi  $\mathcal{U}[0,1]$ , alors  $F^{-1}(U)$  a pour fonction de répartition  $F$ . Si  $X$  a pour fonction de répartition  $F$ , alors  $F(X)$  suit une loi  $\mathcal{U}[0,1]$ .
2. Générer un échantillon de taille 100 de la loi uniforme sur  $[0,1]$  et l'imprimer.
3. Simuler une loi exponentielle de paramètre 1 ; en donner une représentation sous forme d'histogramme sous SAS ainsi que ses caractéristiques élémentaires.
4. Démontrer que si  $f(x) = \lambda e^{-\lambda x} \mathbf{1}_{x \geq 0}$ , alors  $F(x) = 1 - e^{-\lambda x}$ ,  $X = -\ln(1 - U)/\lambda$ .  
Rq :  $1 - U$  a même fonction de répartition que  $U$ .
5. Soit  $X$  une variable aléatoire de fonction de répartition  $F$  cadlag. Quelle est la loi de  $F(X)$  ?  
En déduire une simulation de la loi exponentielle à partir de la loi uniforme ; en donner une représentation.