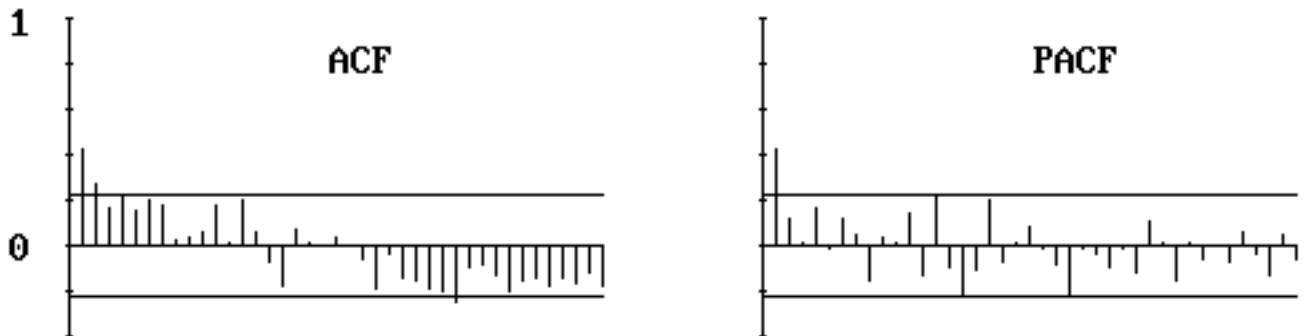


**Examen Mai 2001**

*Aucun document n'est autorisé.*

- Problème 1.**
- 1) Soit le processus  $Y = (Y_t, t \in \mathbb{Z})$ , défini par  $Y_t(\omega) = f(t) + X_t(\omega)$ , avec  $X$  processus stationnaire centré, et  $f$  fonction déterministe définie par  $f(t) = P_k(t) + s(t)$ ,  $P_k$  étant un polynôme de degré  $k$  fixé et  $s$  une fonction périodique de période 4 non nulle.
    - a)  $Y$  est-il un processus stationnaire?
    - b) Rappeler la définition d'un processus  $ARMA(p, q)$ .  
Peut-on modéliser  $Y$  par un  $ARMA(p, q)$ ?
    - c) On décide d'appliquer l'opérateur linéaire  $(I - B)^k(I - B^4)$  à  $Y$ . Justifier ce choix de différenciation, en particulier par le calcul, en considérant le cas où  $k = 1$ .  
Peut-on alors modéliser  $(I - B)^k(I - B^4)Y$  par un  $ARMA(p, q)$ ? S'agit-il d'un processus centré?
    - d) Quelle procédure alternative adopter, si l'on souhaite modéliser  $Y$  sans transformation préalable de  $X$ ?
  - 2) On choisit la méthode de différenciation et à partir des 77 observations de la série transformée  $Z$  définie par  $Z_t := (I - B)^k(I - B^4)Y_t, t \in \mathbb{Z}$ , on trace les graphes de la fonction d'autocorrélation empirique et de la fonction d'autocorrélation partielle empirique.



- a) Rappeler la définition des fonction d'autocovariance, d'autocorrélation et d'autocorrélation partielle d'un processus stationnaire, puis donner un exemple d'estimateur empirique de la fonction d'autocovariance.
- b) A quoi correspondent les bornes  $\pm \frac{1.96}{\sqrt{77}}$  sur les graphes donnés ci-dessus? (comment sont-elles calculées?)
- c) En vous basant sur ces graphes, quels degrés  $p$  et  $q$  choisiriez-vous pour la modélisation de la série  $(Z_t)$  par un processus  $ARMA(p, q)$ ? On aura soin de justifier sa réponse, en rappelant les propriétés utilisées.

3)  $p$  et  $q$  étant supposés connus, on cherche à identifier le modèle  $ARMA(p, q)$ .

- a) Quels sont les paramètres à estimer? Après avoir effectué une estimation préliminaire de ces paramètres, quelle méthode d'estimation utilise-t-on? A quoi sert l'estimation préliminaire?
- b) On s'intéresse à l'estimation préliminaire des paramètres, en particulier dans le cas d'un processus centré  $AR(p)$  noté  $U = (U_t, t \in \mathbb{Z})$ , et défini par

$$U_t + \sum_{j=1}^p \phi_j U_{t-j} = \varepsilon_t, \quad \text{avec } \varepsilon \text{ bruit blanc de variance } \sigma_u^2.$$

Démontrer que la fonction d'autocovariance  $\gamma_u$  de  $U$  satisfait les équations, dites de Yule-Walker, données par

$$\begin{cases} -\sum_{j=1}^p \phi_j \gamma_u(k-j) = \gamma_u(k) & \forall k > 0 \\ -\sum_{j=1}^p \phi_j \gamma_u(j) + \sigma_u^2 = \gamma_u(0) \end{cases}$$

En quoi ce système peut-il être utile pour l'estimation préliminaire des paramètres?

- c) On reprend l'étude de la série  $Z$ . Sa fonction d'autocovariance empirique prend les valeurs:  $\hat{\gamma}(0) = 0.17992$ ,  $\hat{\gamma}(1) = 0.07590$ ,  $\hat{\gamma}(2) = 0.04885$ , etc...  
En supposant que parmi les modèles possibles pour modéliser  $Z$ , on considère un processus  $AR(1)$ , donner une estimation empirique des paramètres de ce modèle.  
On rappelle le résultat suivant:

Si  $X = (X_t, t \in \mathbb{Z})$  est un processus  $AR(p)$  causal avec  $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$  bruit blanc de var  $\sigma^2$  et si  $\hat{\Phi}$  est l'estimateur de Yule-Walker de  $\Phi$ , alors

$$\sqrt{n}(\hat{\Phi} - \Phi) \rightarrow^d \mathcal{N}(0, \sigma^2 \Gamma_p^{-1}), \quad \text{quand } n \rightarrow \infty,$$

avec  $\Gamma_p = (\gamma(i-j))_{i,j=1,\dots,p}$  (matrice de covariance de  $X$ ),  $\Phi := {}^t(\Phi_1, \dots, \Phi_p)$ , et  $\hat{\Phi} := {}^t(\hat{\Phi}_1, \dots, \hat{\Phi}_p)$ .

Utiliser ce résultat pour construire une région de confiance au niveau 95% pour les paramètres du polynôme autorégressif associé au modèle.

**Problème 2.** Soit  $(\varepsilon_t, t \in \mathbb{Z})$  une suite de v.a.r. i.i.d. centrées et de variance  $\sigma^2$ . On pose

$$X_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}, \quad |\rho| < 1, \quad t \in \mathbb{Z}.$$

- 1) Montrer que la série définissant  $X_t$  est convergente en m.q., i.e. dans  $L^2$  (il suffit de montrer que  $(S_n)_n$ , où  $S_n := \sum_{i=0}^n \rho^i \varepsilon_{t-i}$ , est une suite de Cauchy dans  $L^2$ ).
- 2) Calculer  $E[X_t]$ ,  $Var(X_t)$  et  $Cov(X_s, X_t)$  pour  $s < t$  (on aura soin de justifier les calculs effectués).
- 3) Soit le processus  $X = (X_t, t \in \mathbb{Z})$ .
  - a) Montrer que  $X_t = \rho X_{t-1} + \varepsilon_t$ . De quel processus s'agit-il?
  - b) En déduire que, pour  $h > 0$ ,  $X_t = \rho^h X_{t-h} + \sum_{k=1}^h \rho^{h-k} \varepsilon_{t-(h-k)}$ .
  - c)  $X$  est-il faiblement stationnaire?

4) Montrer que pour tout  $h > 0$ ,  $E[X_t \varepsilon_{t+h}] = 0$ .

5) On suppose désormais que  $(\varepsilon_t, t \in \mathbb{Z})$  est un processus gaussien; on en déduit que  $X$  est un processus gaussien (en utilisant que la limite en m.q. de v.a. gaussienne est encore une v.a. gaussienne).  $X$  est-il strictement stationnaire?

On dispose des observations  $(X_0, \dots, X_T)$ . On cherche à prévoir  $X_{T+1}$ . De façon générale, on note  $\mathcal{M}_k$  l'espace vectoriel engendré par  $(X_0, \dots, X_k)$  et  $\hat{X}_t^k = P_{\mathcal{M}_k}(X_t)$  la projection orthogonale de  $X_t$  sur  $\mathcal{M}_k$ .

a) Montrer que  $\hat{X}_{T+1}^T = \rho X_T$ .

b)  $\rho$  étant inconnu, on veut l'estimer. On considère alors l'estimateur  $\hat{\rho}_T$  qui minimise (par rapport à  $r$ )

$$\sum_{t=1}^T (X_t - r X_{t-1})^2.$$

Montrer que le vecteur  $\hat{\rho}_T$  est la projection orthogonale du vecteur  $Y = (X_T, \dots, X_1)$  sur l'espace vectoriel engendré par  $(X_{T-1}, \dots, X_0)$ . (On en déduit alors l'unicité de la solution).

Calculer  $\hat{\rho}_T$  et en déduire que

$$\hat{\rho}_T - \rho = \frac{\sum_{t=1}^T \varepsilon_t X_{t-1}}{\sum_{t=1}^T X_{t-1}^2}.$$

Par conséquent on a  $\hat{X}_{T+1}^T = \hat{\rho}_T X_T$ .

c) Calculer l'erreur de prédiction  $\hat{X}_{T+1}^T - X_{T+1}$  et montrer qu'elle peut s'écrire comme la somme de l'erreur d'estimation et de l'erreur de prévision.