*Université Paris I, Panthéon - Sorbonne*

**Première Année Master M.A.E.F.** 2011 – 2012

# Statistiques II

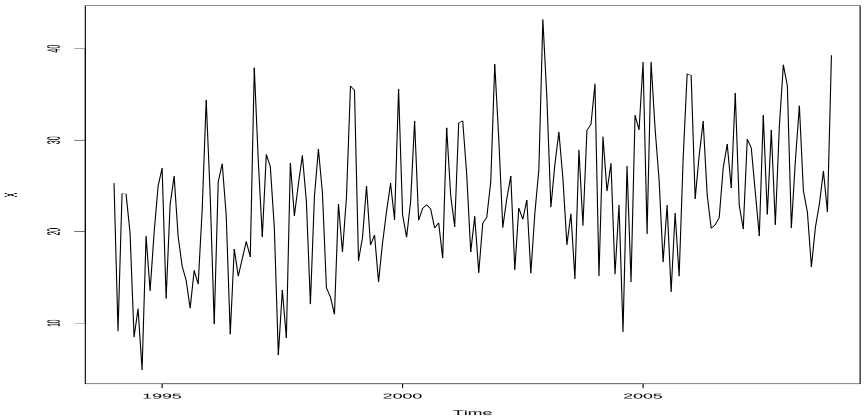
Examen final, mai 2012

*Examen de 3h00. Tout document ou calculatrice est interdit.*

1. **(18 points)** Soit $(\varepsilon_n)_{n\in\mathbf{Z}}$ un bruit blanc gaussien de variance $\sigma_\varepsilon^2$, où $\sigma_\varepsilon^2 > 0$ est inconnu et soit $X = (X_n)_{n\in\mathbf{Z}}$ le processus défini par
$$X_{n+1} - \alpha X_n = \varepsilon_{n+1} - 2\alpha\varepsilon_n \quad \text{pour} \quad n \in \mathbf{Z},$$
avec $|\alpha| < 1$ un réel inconnu.
   (a) Quel processus est $X$? Est-il centré? stationnaire? gaussien? (justifier)
   (b) Montrer que pour tout $n \in \mathbf{Z}$,
$$X_n = \varepsilon_n - \sum_{i=1}^{\infty} \alpha^i\,\varepsilon_{n-i}.$$
   (c) Déduire l'expression de l'autocovariance $r_X(k)$ de $X$ pour tout $k \in \mathbf{Z}$ (on traitera à part le cas $k = 0$). Montrer que pour tout $p \in \mathbf{N}$, $\lim_{n\to\infty} n^p\, r(n) = 0$. Que se passe-t-il lorsque $\alpha = \pm\frac{1}{\sqrt{2}}$?
   (d) Déterminer l'expression de la densité spectrale $f$ de $X$. Determiner la loi asymptotique de $\sqrt{n}\,\overline{X}_n$ lorsque $n \to \infty$, où $\overline{X}_n := \frac{1}{n}\big(X_1 + \cdots + X_n\big)$.
   (e) Soit $(X_1, \ldots, X_N)$ une trajectoire observée de $X$. Déduire de ce qui précède la limite en probabilité de $N\left(\overline{X}_N\right)^2$ quand $N \to \infty$.
   (f) Soit $\widehat{\sigma}_N^2 := \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} X_i^2$ et $\widehat{r}_N(1) := \dfrac{1}{N}\displaystyle\sum_{i=1}^{N-1} X_i X_{i+1}$. Montrer que le vecteur $\big(\widehat{\sigma}_N^2, \widehat{r}_N(1)\big)$ suit un théorème de la limite centrale (on écrira la matrice de covariance asymptotique à l'aide de $f$ et sans calculer les intégrales).
   (g) Déduire de ce qui précède un estimateur $(\widehat{\alpha}, \widehat{\sigma}_\varepsilon^2)$ de $(\alpha, \sigma_\varepsilon^2)$ et expliquer pourquoi cet estimateur est convergent.
   (h) On veut obtenir une prédiction $\widehat{X}_{N+1}$ de $X_{N+1}$, la trajectoire $(X_1, \ldots, X_N)$ étant connue. Sous une condition portant sur $\alpha$ que l'on précisera, donner l'expression de $\varepsilon_N$ en fonction de $(X_{N-k})_{k\in\mathbf{N}}$. En déduire une expression d'un prédicteur $\widehat{X}_{N+1}$ obtenu uniquement à partir de $(X_1, \ldots, X_N)$.

2. **(8.5 points)** A partir de la liste de commandes et de résultats qui suivent:
   (a) Commenter chacune des commandes écrites, en expliquant ce que ces commandes sont supposées faire.
   (b) Donner la légende des différentes figures et commenter ces figures.
   (c) Pour le résultat de la commande `summary(X.BIC)`, expliquer ce que signifie mathématiquement `Pr(>|t|)` et ce que l'on conclue contrétement.
   (d) Que peut-on conclure de la figure issue de la commande `acf(X.BIC$res)` puis de la commande `Box.test(X.BIC$res, lag = 5, type="Ljung")` (test portemanteau)?
   (e) Sans faire le calcul, mais en précisant explicitement les constantes, donner exactement le modèle obtenu pour $X_t$ lorsque $t = 2000 + 2/12$.
   (f) Que représente le résultat 23.39663 dans le résultat issue de la toute dernière commande `Xpred`?

Figure 1: Chiffre d'affaire de l'entreprise




On s'intéresse au chiffre d'affaire mensuel d'une entreprise chocolatière entre les mois de janvier 1994 et décembre 2008. Le tracé de ce chiffre d'affaire noté $X = (X_t)_t$ est donné sur la Figure 1. On tape alors les commandes suivantes en R:

```
monthplot(X)
s=stl(X,s.win="perio")
sais=s$time.series[,1]
sais
```

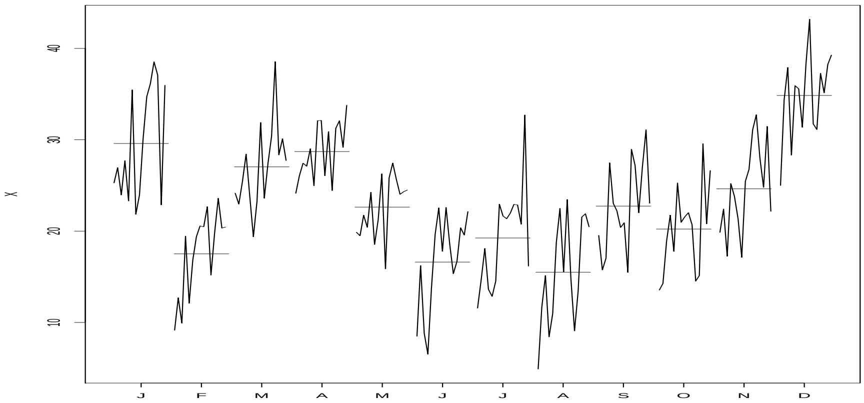
On obtient ainsi les résultats numériques et la Figure 2:

```
  sais[,1]
           Jan        Feb        Mar        Apr        May        Jun
1994  6.5508133 -5.5812967  3.8918706  5.5201549 -0.6064964 -6.6615221
1995  6.5508133 -5.5812967  3.8918706  5.5201549 -0.6064964 -6.6615221
1996  6.5508133 -5.5812967  3.8918706  5.5201549 -0.6064964 -6.6615221
1997  6.5508133 -5.5812967  3.8918706  5.5201549 -0.6064964 -6.6615221
1998  6.5508133 -5.5812967  3.8918706  5.5201549 -0.6064964 -6.6615221
1999  6.5508133 -5.5812967  3.8918706  5.5201549 -0.6064964 -6.6615221
2000  6.5508133 -5.5812967  3.8918706  5.5201549 -0.6064964 -6.6615221
2001  6.5508133 -5.5812967  3.8918706  5.5201549 -0.6064964 -6.6615221
2002  6.5508133 -5.5812967  3.8918706  5.5201549 -0.6064964 -6.6615221
2003  6.5508133 -5.5812967  3.8918706  5.5201549 -0.6064964 -6.6615221
2004  6.5508133 -5.5812967  3.8918706  5.5201549 -0.6064964 -6.6615221
......................................................................
......................................................................
```

On tape ensuite les commandes suivantes:

```
XX=X-sais
ts.plot(XX)
library(MASS)
kmax=10
Z=matrix(nrow=180,ncol=kmax)
for (i in 1:180)
  for (j in 1:(kmax))
    Z[i,j]=i^j
ZZ=as.data.frame(Z)
X.lm=lm(XX~.,data=ZZ)
X.BIC=stepAIC(X.lm,k=log(180))
summary(X.BIC)
```

Figure 2:




On obtient ainsi les résultats suivants:

```
> X.BIC=stepAIC(X.lm,k=log(180))

Step:  AIC=498.81
XX ~ V2 + V3
       Df Sum of Sq    RSS    AIC
<none>              2637.5 498.81
- V3    1    236.80 2874.3 509.10
- V2    1    426.16 3063.7 520.58

> summary(X.BIC)

Call:
lm(formula = XX ~ V2 + V3, data = ZZ)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.900e+01  5.779e-01  32.876  < 2e-16 ***
V2           9.521e-04  1.780e-04   5.348 2.72e-07 ***
V3          -4.133e-06  1.037e-06  -3.986 9.80e-05 ***
---
Signif. codes:  0 '***' 0.001 '*'0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.86 on 177 degrees of freedom
Multiple R-squared: 0.333,Adjusted R-squared: 0.3255
F-statistic: 44.19 on 2 and 177 DF,  p-value: 2.707e-16
```

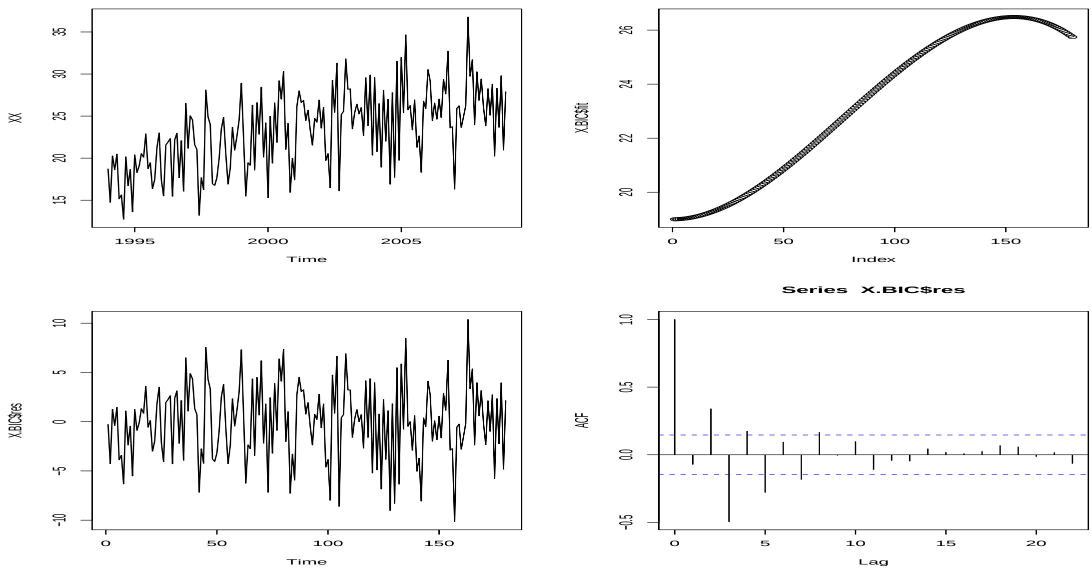
On travaille alors sur la série `X.BIC$res=XX-X.BIC` pour essayer de modéliser cette série:

```
plot(X.BIC$fit)
ts.plot(X.BIC$res)
acf(X.BIC$res)
Box.test(X.BIC$res, lag = 5, type="Ljung")
library(fArma)
X.ARMA=armaFit(~ arma(1,1), data=X.BIC$res, method = c("mle"))
summary(X.ARMA)
```

Voici les résultats, dont la Figure 3:

```
> Box.test(X.BIC$res, lag = 5, type="Ljung")
```

Figure 3:




```
  Box-Ljung test

data:  X.BIC$res
X-squared = 87.5669, df = 5, p-value < 2.2e-16

> summary(X.ARMA)

Call:
 armaFit(formula = ~arma(1, 1), data = X.BIC$res, method = c("mle"))

Model:
 ARIMA(1,0,1) with method: CSS-ML

Coefficient(s):
      ar1        ma1  intercept
-0.923728   0.765415  -0.008483

Coefficient(s):
           Estimate  Std. Error  t value Pr(>|t|)
ar1       -0.923728    0.040440  -22.842   <2e-16 ***
ma1        0.765415    0.057322   13.353   <2e-16 ***
intercept -0.008483    0.242192   -0.035    0.972


preRes=predict(X.ARMA,n.ahead=12)
new=c(181:192)
Xpred=preRes$pred+sais[1:12]+1.900e+01+9.521e-04*new^2-4.133e-06*new^3
Xpred
```

On obtient le résultat suivant:

```
> Xpred

 [1] 30.29616 21.81458 27.79215 32.52088 23.39663 19.97180 19.99761
 [8] 18.45459 23.44601 22.78869 25.25547 37.02333
```