

Time Series Analysis : TD4.

Exercice 1. This exercise is using the properties of the projection in order to get an efficient algorithm for determining the best linear prediction $\Pi_t(X_{t+1})$ and the associated risk R_t^L . Consider a $WN(\sigma^2)$ (Z_t) and the MA(1) (X_t) defined as

$$X_t = Z_t + \theta Z_{t-1}, \quad t \in \mathbb{Z},$$

with $|\theta| < 1$.

1. Express the coefficients (φ_j) of the causal solution $X_t = \sum_{j=1}^{\infty} \varphi_j X_{t-j} + Z_t$ of the MA(1) model in term of θ .
2. Deduce $\Pi_{\infty}(X_{n+1})$ and the associated risk R_{∞}^L .
3. Show that $\Pi_n(X_{n+2}) = 0$ and $\mathbb{E}[X_{n+1}\Pi_{n-1}(X_n)] = 0$.
4. Deduce from the projection decomposition the recursive formula called

$$\Pi_n(X_{n+1}) = \frac{\sigma^2\theta}{R_n^L}(X_n - \Pi_{n-1}(X_n)), \quad n \geq 1.$$

5. Deduce the recursive formula $R_{n+1}^L = \sigma^2(1+\theta^2) - \sigma^4\theta^2/R_n^L$ for $n \geq 1$ and the innovation algorithm that update $(\Pi_n(X_{n+1}), R_n^L)$ recursively.

Exercice 2.¹

Un statisticien étudie un jeux de données composé de 4 séries temporelles X^1, X^2, X^3 et X^4 représentées en fin de TD.

1. Proposer une démarche de modélisation pour chacune de ces séries, justifier.

Notre statisticien s'intéresse ensuite à la série X^5 . Il propose et cherche ensuite à valider un modèle. Il obtient les sorties suivantes :

```
##
## Call:
## arima(x = X5, order = c(4, 1, 5), include.mean = T, method = c("ML"))
##
## Coefficients:
##      ar1 ar2 ar3 ar4 ma1 ma2 ma3 ma4
##      0.7730 0.0375 -0.6437 0.0846 -0.6368 -0.0851 -0.3488 0.5998
## s.e. 0.2432 0.1432 0.1075 0.1694 0.2359 0.1121 0.1006 0.1147
##      ma5
##      -0.3280
## s.e. 0.1131
##
```

1. <https://www.imo.universite-paris-saclay.fr/yannig.goude/teaching.html>

```
## sigma^2 estimated as 3.876: log likelihood = -629.03, aic = 1278.05
## pvalue student-test:
## ar1 ar2 ar3 ar4 ma1 ma2 ma3 ma4 ma5
## 0.00 0.79 0.00 0.62 0.01 0.45 0.00 0.00 0.00
```

2. Préciser et commenter le modèle choisi par notre statisticien.
3. A quoi correspond “sigma^2” ? Que signifie l’option “method = c(“ML”)” ? A quoi correspondent les p-values affichées ?
4. Il décide de tester un modèle ARIMA(5,1,6) et obtient pour ce modèle une log-vraisemblance de 627.4, cela vous paraît-il logique ? Quel est l’AIC de ce modèle ?
5. Au vu de ces résultats que doit faire notre statisticien ?
6. Il souhaite ensuite valider son modèle, quels autres outils de diagnostic lui proposez vous ?



