Chapitre 7

Estimation des paramètres

Ghislain Léveillé
École d'actuariat, Université Laval

Copyright © 2017

7.1 Contenu du chapitre

Dans le chapitre précédent, nous avons défini des outils qui nous ont aidé à spécifier les modèles qui semblaient le mieux modéliser nos séries chronologiques. Ainsi, après avoir retenu le meilleur modèle ARMA(p,q) (ou ARIMA(p,d,q)) modélisant nos données, nous allons nous attarder dans ce chapitre à estimer les paramètres de ce modèle, essentiellement les paramètres du modèle stationnaire ARMA(p,q), en les obtenant directement ou en opérant d'abord une transformation et/ou une (ou plusieurs) différenciation(s).

Nous considérerons donc trois méthodes d'estimation courantes: soit la méthode des moments, la méthode des moindres carrés et la méthode du maximum de vraisemblance. Les propriétés (asymptotiques) de ces divers estimés seront étudiées et illustrées avec des simulations et des exemples extraits de nos bases de données TSA.

7.2 La méthode des moments (MOM)

La méthode des moments est une des méthodes d'estimation les plus simples, sinon la plus « expéditive », pour estimer les paramètres du modèle retenu pour une série chronologique.

Cette méthode consiste d'abord à égaler les moments échantillonnaux avec les moments théoriques, et par la suite à résoudre le système d'équations ainsi obtenu. Toutes les formules utiles pour faire ces estimations sont très bien implantées dans la plupart des logiciels de statistiques, dont le R.

Nous allons donc d'abord appliquer cette méthode aux processus AR, puis aux processus MA et enfin aux processus ARMA. L'estimation de σ_e^2 ne se fera faite pour chacun des trois processus précédents que lorsque les paramètres γ_0, ϕ_i et θ_i auront été estimés.

7.2.1 Modèles AR

(a) Considérer le processus (stationnaire) AR(1)

$$Y_t = \phi Y_{t-1} + e_t \quad , \quad e_t \sim WN\left(0, \sigma_e^2\right) \quad , \quad e_t \text{ ind. de } Y_{t-1}, Y_{t-2}, \ldots$$

Pour ce modèle, nous avons

$$\rho_1 = \phi \Rightarrow \hat{\phi} = r_1 \quad .$$

Pour l'estimation de σ_e^2, nous aurons besoin des estimés de γ_0 et de ϕ. Ainsi, en notant d'abord que

$$\gamma_0 = \frac{\sigma_e^2}{1 - \phi\rho_1} \Leftrightarrow \sigma_e^2 = \left(1 - \phi\rho_1\right)\gamma_0 \quad ,$$

il nous suffira d'utiliser les estimateurs

$$\hat{\gamma}_0 = S^2 = \frac{1}{n-1}\sum_{t=1}^{n}\left(Y_t - \overline{Y}\right)^2 \quad , \quad \hat{\phi} = r_1 \quad ,$$

pour ainsi obtenir l'estimé de σ_e^2, soit

$$\hat{\sigma}_e^2 = \left(1 - \hat{\phi} r_1\right) S^2 = \left(1 - r_1^2\right) S^2 \quad .$$

(b) Considérer le processus (stationnaire) AR(2)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t \quad , \quad e_t \sim WN\left(0, \sigma_e^2\right) \quad , \quad e_t \text{ indép. de } Y_{t-1}, Y_{t-2}, \ldots$$

Pour ce modèle, nous avons

$$\rho_1 = \phi_1 + \rho_1 \phi_2 \quad , \quad \rho_2 = \rho_1 \phi_1 + \phi_2 \quad .$$

En posant $\rho_1 = r_1$ et $\rho_2 = r_2$, nous obtenons les estimateurs

$$\hat{\phi}_1 = \frac{r_1\left(1-r_2\right)}{1-r_1^2} \ , \ \hat{\phi}_2 = \frac{r_2 - r_1^2}{1-r_1^2} \quad .$$

Pour l'estimation de σ_e^2, nous aurons besoin des estimés de γ_0, de ϕ_1, ϕ_2 et de ρ_1, ρ_2. Ainsi, en notant que

$$\gamma_0 = \frac{\sigma_e^2}{1-\phi_1\rho_1 - \phi_2\rho_2} \ \Leftrightarrow \ \sigma_e^2 = \left(1-\phi_1\rho_1 - \phi_2\rho_2\right)\gamma_0 \quad ,$$

il nous suffira d'utiliser les estimés des quantités précédentes pour obtenir l'estimé de σ_e^2, soit

$$\hat{\sigma}_e^2 = \left(1-\hat{\phi}_1 r_1 - \hat{\phi}_2 r_2\right)S^2 \quad .$$

(c) Considérer le processus (stationnaire) AR(p)

$Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + e_t$, $e_t \sim WN\left(0, \sigma_e^2\right)$, e_t indép. de $Y_{t-1}, Y_{t-2}, ...$

Pour ce modèle, nous avons le système linéaire de Yule-Walker

$$\begin{aligned} \rho_1 &= \phi_1 + \rho_1 \phi_2 + ... + \rho_{p-1} \phi_p \\ &\vdots \\ \rho_p &= \rho_{p-1} \phi_1 + \rho_{p-2} \phi_2 + ... + \phi_p \end{aligned}$$

En posant $\rho_1 = r_1$, ... , $\rho_p = r_p$, nous obtiendrons les estimateurs $\hat{\phi}_1$, ..., $\hat{\phi}_p$ en résolvant le système d'équations

$$\begin{aligned} r_1 &= \phi_1 + r_1 \phi_2 + ... + r_{p-1} \phi_p \\ &\vdots \\ r_p &= r_{p-1} \phi_1 + r_{p-2} \phi_2 + ... + \phi_p \end{aligned}$$

Pour l'estimation de σ_e^2, nous aurons besoin des estimés de γ_0, de $\phi_1,\ldots,\phi_p$ et de $\rho_1,\ldots,\rho_p$. Ainsi, en notant que

$$\gamma_0 = \frac{\sigma_e^2}{1-\phi_1\rho_1 - \ldots - \phi_p\rho_p} \Leftrightarrow \sigma_e^2 = \left(1-\phi\rho_1 - \ldots - \phi_p\rho_p\right)\gamma_0 \quad ,$$

il nous suffira d'utiliser les estimés des quantités précédentes pour obtenir l'estimé de σ_e^2, soit

$$\hat{\sigma}_e^2 = \left(1-\hat{\phi}_1 r_1 - \ldots - \hat{\phi}_p r_p\right)S^2 \quad .$$

Remarque : Il pourrait arriver que les estimés $\hat{\varphi}_1,\ldots,\hat{\phi}_p$ produisent des racines du polynôme (estimé) $\hat{\varphi}\left(x\right) = 1-\hat{\phi}_1 x - \ldots - \hat{\phi}_p x^p$ dont les valeurs absolues (ou normes) n'excéderont pas 1 même si le processus AR est stationnaire. Donc, la prudence est de mise ...

7.2.2 Modèles MA

(a) Considérer le processus MA(1)

$$Y_t = e_t - \theta e_{t-1} \quad , \quad e_t \sim WN\left(0, \sigma_e^2\right) \quad .$$

Pour ce modèle, nous avons

$$\rho_1 = -\frac{\theta}{1+\theta^2} \Leftrightarrow \rho_1 \theta^2 + \theta + \rho_1 = 0 \quad .$$

Ainsi, en posant $\rho_1 = r_1$, $\hat{\theta}$ sera solution de l'équation quadratique précédente, i.e

$$\hat{\theta} = \frac{-1 \pm \sqrt{1 - 4r_1^2}}{2r_1} \quad .$$

- Si $|r_1| > 0.5$, alors il n'y a pas de solution réelle.

- Si $|r_1| = 0.5$, alors $\hat{\theta} = \pm 1$ et cela implique que le modèle MA(1) n'est pas inversible.

- Si $|r_1| < 0.5$, alors $\hat{\theta} = \dfrac{-1 + \sqrt{1 - 4r_1^2}}{2r_1}$ est l'estimateur qui correspond à un modèle MA(1) inversible (i.e. tel $|\hat{\theta}| < 1$)

Pour l'estimation de σ_e^2, nous aurons besoin des estimés de γ_0, de ρ_1 et de θ. Ainsi, en notant que

$$\gamma_0 = \left(1 + \theta^2\right)\sigma_e^2 \iff \sigma_e^2 = \frac{\gamma_0}{1 + \theta^2} \quad ,$$

il nous suffira d'utiliser les estimés des quantités précédentes pour obtenir l'estimé de σ_e^2, soit

$$\hat{\sigma}_e^2 = \frac{S^2}{1+\hat{\theta}^2} \quad .$$

(b) Considérer le processus MA(q) , $q \geq 2$,

$$Y_t = e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q} \quad , \quad e_t \sim WN\left(0, \sigma_e^2\right) \quad .$$

Pour $q \geq 2$, l'estimation des paramètres $\theta_1, \ldots, \theta_q$ est beaucoup plus complexe parce qu'il s'agit alors de résoudre un système non linéaire (en posant $\rho_k = r_k$)

$$r_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \ldots + \theta_q\theta_{q-k}}{1+\theta_1^2 + \ldots + \theta_q^2} \quad , \quad k = 1, 2, \ldots, q-1 \quad ,$$

et

$$r_q = \frac{-\theta_q}{1+\theta_1^2+\ldots+\theta_q^2} \quad .$$

Les estimés de $\theta_1,\ldots,\theta_q$ seront aussi multiples et seulement certains d'entre eux donneront un processus MA(q) inversible.

Pour l'estimation de σ_e^2, nous aurons besoin des estimés de γ_0, de $\rho_1,\ldots,\rho_q$ et de $\theta_1,\ldots,\theta_q$. Ainsi, en notant que

$$\gamma_0 = \left(1+\theta_1^2+\ldots+\theta_q^2\right)\sigma_e^2 \iff \sigma_e^2 = \frac{\gamma_0}{1+\theta_1^2+\ldots+\theta_q^2} \quad ,$$

il nous suffira d'utiliser les estimés des quantités précédentes pour obtenir l'estimé de σ_e^2, soit

$$\hat{\sigma}_e^2 = \frac{S^2}{1+\hat{\theta}_1^2+\ldots+\hat{\theta}_q^2} \quad .$$

Remarque : Il est clair que les estimés obtenus par la méthode des moments ne sont pas recommandés pour le modèle MA. Ils sont difficiles à obtenir et ils ne sont pas nécessairement de très bons estimés.

7.2.3 Modèles ARMA

Étant donné les difficultés mentionnées en section 7.2.2, nous ne considérerons que le processus (stationnaire) ARMA(1,1)

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1} \quad , \quad e_t \sim WN\left(0, \sigma_e^2\right) \quad , \quad e_t \text{ ind. de } Y_{t-1}, Y_{t-2}, \ldots$$

Pour ce modèle, nous avons

$$\rho_k = \frac{\left(1-\theta\phi\right)\left(\phi-\theta\right)}{1-2\theta\phi+\theta^2}\phi^{k-1} \quad ,$$

et donc,

$$\frac{\rho_2}{\rho_1} = \phi \;\Rightarrow\; \hat{\phi} = \frac{r_2}{r_1} \quad .$$

Pour l'estimation de θ, il faudra résoudre l'équation quadratique

$$r_1 = \frac{\left(1-\theta\hat{\phi}\right)\left(\hat{\phi}-\theta\right)}{1-2\theta\hat{\phi}+\theta^2} \quad .$$

Pour l'estimation de σ_e^2, nous aurons besoin des estimés de γ_0, θ et ϕ. Ainsi, en notant que

$$\gamma_0 = \frac{1-2\phi\theta+\theta^2}{1-\phi^2}\sigma_e^2 \;\Leftrightarrow\; \sigma_e^2 = \frac{1-\phi^2}{1-2\phi\theta+\theta^2}\gamma_0 \quad ,$$

nous obtenons

$$\hat{\sigma}_e^2 = \frac{1-\hat{\phi}^2}{1-2\hat{\phi}\hat{\theta}+\hat{\theta}^2}S^2 \quad .$$

7.2.4 Exemples

(1) Considérer le fichier « hare », étudié à la section 6.7.1, recensant la population annuelle de lièvres au Canada de 1905 à 1935.

Après avoir transformé les données avec une racine carrée, nous avons vu que la fonction PACF semblait suggérer, pour la série chronologique « hare^0.5 », un modèle AR(2) ou même AR(3), et que le critère AIC (corroboré par le test ADF) suggérait plus fortement un modèle AR(3).

Pour fins d'illustration de la méthode des moments, nous estimerons les paramètres du modèle AR(2). Ainsi, nous évaluons d'abord les coefficients r_k (dont r_1 et r_2),

```
> data(hare)
> x=as.vector(acf(hare^0.5))
```

> x

Autocorrelations of series ‘hare^0.5’, by lag

1	2	3	4	5	6	7	8	9	10	11	12	13	14
0.736	0.304	-0.169	-0.497	-0.612	-0.584	-0.357	-0.059	0.261	0.448	0.436	0.279	0.051	-0.176

Nous obtenons

$$\hat{\phi}_1 = \frac{r_1\left(1-r_2\right)}{1-r_1^2} = \frac{0.736\left(1-0.304\right)}{1-\left(0.736\right)^2} = 1.1178 \quad ,$$

et

$$\hat{\phi}_2 = \frac{r_2 - r_1^2}{1-r_1^2} = \frac{0.304-\left(0.736\right)^2}{1-\left(0.736\right)^2} = -0.519 \quad .$$

Nous estimons ensuite $\overline{\sqrt{Y_t}}$ et $Var\left(\sqrt{Y_t}\right)$,

```
> data(hare)
> mean(hare^0.5)

[1] 5.818966

> var(hare^0.5)

[1] 5.877627
```

Nous obtenons ainsi,

$$
\begin{aligned}
\hat{\sigma}_e^2 &= \left(1-\hat{\varphi}_1 r_1 - \hat{\varphi}_2 r_2\right)s^2 \\
&\approx \left[1-\left(1.1178\right)\left(0.736\right)-\left(-0.519\right)\left(0.304\right)\right]\left(5.88\right) \\
&= 1.97
\end{aligned}
$$

Le modèle estimé ainsi obtenu est

$$\sqrt{Y_t} - 5.82 = 1.1178\left(\sqrt{Y_{t-1}} - 5.82\right) - 0.519\left(\sqrt{Y_{t-2}} - 5.82\right) + e_t$$

$$\Updownarrow$$

$$\sqrt{Y_t} = 2.335 + 1.1178\sqrt{Y_{t-1}} - 0.519\sqrt{Y_{t-2}} + e_t$$

Remarque :

Vous auriez pu écrire directement

$$\sqrt{Y_t} = c + 1.1178\sqrt{Y_{t-1}} - 0.519\sqrt{Y_{t-2}} + e_t \quad ,$$

puis évaluer

$$c = \left(1 - 1.1178 + 0.519\right) E\left[\sqrt{Y_t}\right] \approx \left(0.4012\right)\left(5.8777\right) \approx 2.335 \quad .$$

(2) Considérer le fichier « oil.price », étudié à la section 6.7.2, donnant le prix mensuel du baril de pétrole brut de janvier 1986 à janvier 2006.

Après avoir transformé les données avec un logarithme, nous avons vu que les fonctions ACF et EACF de la différence du logarithme de nos données semblaient suggérer un modèle MA(1), et qu'au moins 3 autres modèles pouvaient être considérés. La moyenne de la différence du logarithme des données étant presque négligeable (~0.004), nous la considérerons donc nulle.

Pour fins d'illustration de la méthode des moments, nous estimerons les paramètres du modèle MA(1). Ainsi, nous évaluons d'abord les coefficients r_k (dont r_1),

```
> data(oil.price)
> x=acf(diff(as.vector(log(oil.price))))
```

```
> x

Autocorrelations of series 'diff(as.vector(log(oil.price)))', by lag

     1      2      3      4      5      6      7      8      9     10     11     12     13     14
 0.212 -0.087 -0.046 -0.076 -0.054 -0.113 -0.021  0.060  0.034  0.099  0.089 -0.004 -0.125 -0.072
```

Nous obtenons

$$\hat{\theta} = \frac{-1 + \sqrt{1 - 4(0.212)^2}}{2(0.212)} \approx -0.222 \quad .$$

Nous évaluons ensuite γ_0,

```
> data(oil.price)
> var(diff(as.vector(log(oil.price))))

[1] 0.007162272
```

Le modèle estimé ainsi obtenu est

$$\log Y_t = \log Y_{t-1} + e_t + 0.222 e_{t-1} \quad .$$

Remarque : En posant $X_t = \nabla \log Y_t$ et en utilisant la formule de la variance de $\bar{X}$ obtenue en section 3.2.1, soit

$$Var\left(\bar{X}\right) = \frac{Var\left(X_t\right)}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k\right] \approx \frac{\hat{\gamma}_0}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)r_k\right] \quad ,$$

nous pouvons vérifier que $\sqrt{Var\left(\bar{X}\right)} = \sqrt{Var\left(\overline{\nabla \log Y_t}\right)} \approx 0.006$, ce qui nous permets de conclure plus rigoureusement que la moyenne de $\nabla \log Y_t$ (~0.004) est réellement négligeable.

7.3 La méthode des moindres carrés conditionnels (CLS)

Nous avons vu à la section 7.2 que la méthode d'estimation par les moments n'était pas recommandée pour des séries chronologiques si elles sont modélisées par des processus MA ou des processus ARMA (avec q > 1). Nous allons donc considérer un autre type d'estimateur, basée sur la méthode des moindres carrés (conditionnels).

7.3.1 Modèles AR

(a) Considérer le processus stationnaire AR(1), avec moyenne non nulle μ, soit

$$Y_t - \mu = \phi\left(Y_{t-1} - \mu\right) + e_t \ , \ e_t \sim WN\left(0, \sigma_e^2\right) \text{, } e_t \text{ ind. de } Y_{t-1}, Y_{t-2}, \ldots$$

À partir des n observations de notre série chronologique, et en considérant ce modèle comme une régression de Y_t sur Y_{t-1}, nous définissons la somme des carrés conditionnels

$$S_c(\phi,\mu) = \sum_{t=2}^{n} \left[(Y_t - \mu) - \phi (Y_{t-1} - \mu) \right]^2 \quad .$$

La méthode des moindres carrés conditionnels consiste à choisir les paramètres ϕ et μ qui minimiseront la somme précédente. Pour notre modèle AR(1), nous avons

- $$\frac{\delta S_c(\phi,\mu)}{\delta\phi} = 0 \ \Rightarrow \ \sum_{t=2}^{n} \left[(Y_t - \mu) - \phi (Y_{t-1} - \mu) \right] (Y_{t-1} - \mu) = 0$$

- $$\frac{\delta S_c(\phi,\mu)}{\delta\mu} = 0 \ \Rightarrow \ \sum_{t=2}^{n} \left[(Y_t - \mu) - \phi (Y_{t-1} - \mu) \right] (\phi - 1) = 0$$

D'où, pour n « grand »,

$$\hat{\mu} = \frac{\sum_{t=2}^{n} Y_t - \hat{\phi} \sum_{t=2}^{n} Y_{t-1}}{(n-1)(1-\hat{\phi})} \approx \bar{Y} \quad , \quad \hat{\phi} = \frac{\sum_{t=2}^{n} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=2}^{n} (Y_t - \bar{Y})^2}$$

Remarques : (1) Pour n grand, $\hat{\phi} \approx r_1$, puisque seul le terme $t = 1$ manque au dénominateur de l'expression pour $\hat{\phi}$ (voir Chapitre 3, page 67).

(2) Comme dans la méthode des moments, nous aurons

$$\hat{\sigma}_e^2 = (1 - \hat{\phi} r_1) S^2 \quad .$$

(b) Considérer le processus stationnaire AR(p), avec moyenne non nulle μ, soit

$$Y_t - \mu = \phi_1\left(Y_{t-1} - \mu\right) + \ldots + \phi_p\left(Y_{t-p} - \mu\right) + e_t \quad ,$$

où $e_t \sim WN\left(0, \sigma_e^2\right)$, e_t ind. de $Y_{t-1}, Y_{t-2}, \ldots$

De la même façon que précédemment, nous faisons une régression de Y_t sur $Y_{t-1}, \ldots, Y_{t-p}$ et nous minimisons la fonction suivante par rapport aux paramètres $\phi_1, \ldots, \phi_p$ et μ, soit

$$S_c\left(\phi, \mu\right) = \sum_{t=p+1}^{n} \left[\left(Y_t - \mu\right) - \phi_1\left(Y_{t-1} - \mu\right) - \ldots - \phi_p\left(Y_{t-p} - \mu\right)\right]^2 \quad .$$

Les expressions obtenues sont généralement plus complexes mais, pour n grand, $\hat{\mu} \approx \bar{Y}$ et les paramètres $\phi_1, \ldots, \phi_p$ pourront être estimés par les équations de Yule-Walker

$$\begin{aligned}
r_1 &= \phi_1 + \phi_2 r_1 + \phi_3 r_2 + \ldots + \phi_p r_{p-1} \\
r_2 &= \phi_1 r_1 + \phi_2 + \phi_3 r_1 + \ldots + \phi_p r_{p-2} \\
&\vdots \\
r_p &= \phi_1 r_{p-1} + \phi_2 r_{p-2} + \phi_3 r_{p-3} + \ldots + \phi_p
\end{aligned}$$

Remarques : (1) Donc, pour les modèles stationnaires AR(p), les estimateurs des moments et les estimateurs de moindres carrés conditionnels sont approximativement égaux.

(2) Tout comme dans la méthode des moments, nous aurons

$$\hat{\sigma}_e^2 = \left(1 - \hat{\phi}_1 r_1 - \ldots - \hat{\phi}_p r_p\right) S^2 \quad .$$

7.3.2 Modèles MA

(a) Considérer le processus MA(1), soit

$$Y_t = e_t - \theta e_{t-1} \ , \ e_t \sim WN\left(0,\sigma_e^2\right) \ .$$

À partir des n observations de notre série chronologique, et en considérant les n « erreurs » e_t réécrites comme suit (ainsi que le fait d'imposer la condition $e_0 = 0$)

$$e_1 = Y_1 \, , e_2 = Y_2 + \theta e_1 \, , \ldots , e_n = Y_n + \theta e_{n-1} \quad ,$$

nous aurons alors à minimiser par rapport à θ la fonction

$$S_c\left(\theta\right) = \sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n} \left(Y_t + \theta e_{t-1}\right)^2 \quad .$$

Remarques : (1) Cette valeur de θ sera obtenue numériquement et devra être comprise entre -1 et 1, pour l'inversibilité du processus.

(2) Comme dans la méthode des moments, nous aurons

$$\hat{\sigma}_e^2 = \frac{S^2}{1+\hat{\theta}^2} \quad .$$

(b) Considérer le processus MA(q), soit

$$Y_t = e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q} \ \ , \ e_t \sim WN\left(0, \sigma_e^2\right) \ .$$

Comme précédemment, en posant $e_0 = e_{-1} = \ldots = e_{1-q} = 0$, nous aurons à minimiser par rapport aux paramètres $\theta_1, \ldots, \theta_q$ la fonction

$$S_c\left(\theta\right) = \sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n} \left(Y_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q}\right)^2 \quad .$$

Remarques : (1) Ces valeurs de $\theta_1,\ldots,\theta_q$ seront obtenues numériquement et les racines du polynôme caractéristique $\theta(x)$ devront avoir leur "norme" supérieure à 1, pour l'inversibilité du processus.

(2) Comme dans la méthode des moments, nous aurons

$$\hat{\sigma}_e^2 = \frac{S^2}{1+\hat{\theta}_1^2+\ldots+\hat{\theta}_q^2} \quad .$$

7.3.3 Modèles ARMA

Nous ne considérerons ici que le modèle stationnaire ARMA(1,1), soit

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1} \quad , \quad e_t \sim WN\left(0,\sigma_e^2\right) \quad , \quad e_t \text{ ind. de } Y_{t-1}, Y_{t-2}, \ldots$$

Afin d'utiliser au mieux la série chronologique, nous réécrivons d'abord le modèle comme suit

$$e_t = Y_t - \phi Y_{t-1} + \theta e_{t-1} \quad ,$$

et posons $e_1 = 0$. Nous avons ainsi

$$e_2 = Y_2 - \phi Y_1 \, , e_3 = Y_3 - \phi Y_2 + \theta e_2 \, , \ldots , e_n = Y_n - \phi Y_{n-1} + \theta e_{n-1} \quad .$$

La fonction à minimiser par rapport aux paramètres ϕ et θ sera alors

$$S_c(\phi,\theta) = \sum_{t=2}^{n} e_t^2 = \sum_{t=2}^{n} \left(Y_t - \phi Y_{t-1} + \theta e_{t-1} \right)^2 \quad .$$

Nous obtenons les estimés suivants:

$$\hat{\phi} = \frac{\sum_{t=2}^{n} Y_t Y_{t-1} \sum_{t=2}^{n} e_{t-1}^2 - \sum_{t=2}^{n} Y_t e_{t-1} \sum_{t=2}^{n} Y_{t-1} e_{t-1}}{\sum_{t=2}^{n} Y_{t-1}^2 \sum_{t=2}^{n} e_{t-1}^2 - \sum_{t=2}^{n} Y_{t-1} e_{t-1} \sum_{t=2}^{n} Y_{t-1} e_{t-1}} \quad ,$$

et

$$\hat{\theta} = \frac{\sum_{t=2}^{n} Y_t Y_{t-1} \sum_{t=2}^{n} Y_{t-1} e_{t-1} - \sum_{t=2}^{n} Y_t e_{t-1} \sum_{t=2}^{n} Y_{t-1}^2}{-\sum_{t=2}^{n} Y_{t-1} e_{t-1} \sum_{t=2}^{n} Y_{t-1} e_{t-1} + \sum_{t=2}^{n} e_{t-1}^2 \sum_{t=2}^{n} Y_{t-1}^2} \quad .$$

Remarque : Comme dans la méthode des moments, nous aurons

$$\hat{\sigma}_e^2 = \left[\frac{1-\hat{\phi}^2}{1-2\hat{\phi}\hat{\theta}+\hat{\theta}^2} \right] S^2 \quad .$$

7.3.4 Exemples

(1) Reprenons l'exemple 7.2.4 (1), et utilisons la méthode des moindres carrés conditionnels pour estimer les paramètres du modèle AR(2) sensé représenter la série chronologique « hare^0.5 ».

```
> data(hare)
> arima(hare^0.5, order=c(2,0,0), method='CSS')

Call:
> arima(x = hare^0.5, order = c(2, 0, 0), method = "CSS")

Coefficients:
         ar1      ar2   intercept
      1.3631  -0.7792     5.7086
s.e.  0.1332   0.1362     0.4783

sigma^2 estimated as 1.225:  part log likelihood = -47.13
```

Nous obtenons ainsi le modèle suivant

$$\sqrt{Y_t} - 5.7086 = 1.3631\left(\sqrt{Y_{t-1}} - 5.7086\right) - 0.7792\left(\sqrt{Y_{t-2}} - 5.7086\right) + e_t$$

$$\Updownarrow$$

$$\sqrt{Y_t} = 2.3753 + 1.3631\sqrt{Y_{t-1}} - 0.7792\sqrt{Y_{t-2}} + e_t$$

Remarques : (1) Le nombre 5.7086 est un estimé des moindres carrés conditionnels de la moyenne de $\sqrt{Y_t}$.

(2) Si nous comparons le modèle précédent avec le modèle obtenu par la méthode des moments, soit

$$\sqrt{Y_t} = 2.335 + 1.1178\sqrt{Y_{t-1}} - 0.519\sqrt{Y_{t-2}} + e_t \quad ,$$

tous les paramètres diffèrent « quelque peu ».

(2) Reprenons l'exemple 7.2.4 (2), et utilisons la méthode des moindre carrés conditionnels pour estimer les paramètres du modèle MA(1) sensé représenter la série chronologique « ∇ log(oil.price) ».

```
> data(oil.price)
> arima(log(oil.price), order=c(0,1,1), method='CSS')

Call:
arima(x = log(oil.price), order = c(0, 1, 1), method = "CSS")

Coefficients:
         ma1
      0.2731
s.e.  0.0681

sigma^2 estimated as 0.006731:  part log likelihood = 259.58
```

Nous obtenons ainsi le modèle suivant

$$\log Y_t = \log Y_{t-1} + e_t + 0.2731 e_{t-1} \quad .$$

Remarque : Si nous comparons le modèle précédent avec le modèle obtenu par la méthode des moments, soit

$$\log Y_t = \log Y_{t-1} + e_t + 0.222 e_{t-1} \quad ,$$

le paramètre diffère légèrement.

Notons tout de même que l'estimé des moments de θ (- 0.222) se trouve dans l'intervalle à un écart-type de - 0.2731 obtenu précédemment, soit

$$\left]-0.0681-0.2731, 0.0681-0.2731\right[\quad = \quad \left]-0.3412, -0.2050\right[\quad .$$

7.4 La méthode du maximum de vraisemblance (MLE)

La méthode du maximum de vraisemblance est une des techniques les plus utilisées, non seulement dans les séries chronologiques mais aussi généralement en statistique, pour estimer les paramètres inconnus d'un modèle de série chronologique.

- Un premier avantage est d'utiliser toutes les valeurs de la série chronologique, sans se préoccuper de valeurs initiales comme dans la méthode précédente.

- Un deuxième avantage est que les estimateurs MLE ont de très belles propriétés asymptotiques (n grand).

- Un désavantage important est qu'il faut spécifier la forme de la distribution conjointe des variables aléatoires $Y_1,...,Y_n$ (sans bien sûr avoir la valeur des paramètres ...)

7.4.1 Illustration de la méthode

Rappelons que la fonction du maximum de vraisemblance L est une fonction qui spécifie la distribution conjointe des variables aléatoires $Y_1,\ldots,Y_n$, mais que celle-ci est plutôt vue comme une fonction des paramètres du modèle. Le but est donc de maximiser la valeur de cette fonction par rapport à ses paramètres.

Pour illustrer le fonctionnement de cette méthode, nous allons considérer le modèle stationnaire AR(1), soit

$$Y_t - \mu = \phi\left(Y_{t-1} - \mu\right) + e_t \quad ,$$

où nous supposons

$$e_t \sim N\left(0,\sigma_e^2\right) \quad , \quad e_t \text{ sont iid} \quad .$$

Comme nous aimerions avoir la distribution conjointe de $Y_1,\ldots,Y_n$, nous allons d'abord observer que

- $f\left(y_1,\ldots,y_n\right)=f\left(y_2,\ldots,y_n \middle| y_1\right)f\left(y_1\right)$

- $Y_1 \sim N\left(\mu,\sigma_e^2 \middle/ \left(1-\phi^2\right)\right)$, puisque que $Y_t=\mu+\sum_{k=0}^{\infty}\phi^k e_{t-k}$

- $g\left(e_2,\ldots,e_n\right)=\prod_{t=2}^{n} g\left(e_t\right)=\left(2\pi\sigma_e^2\right)^{-\frac{n-1}{2}}\exp\left(-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n}e_t^2\right)$

En posant $Y_1=y_1$, nous pouvons effectuer une transformation linéaire des variables $Y_2,\ldots,Y_n$ en fonction des variables $e_2,\ldots,e_n$ en utilisant l'équation

$$Y_i=\mu+\phi\left(Y_{i-1}-\mu\right)+e_i \ , \ i=2,\ldots,n \quad .$$

Cette transformation linéaire de $Y_2,...,Y_n$ en fonction de $e_2,...,e_n$ sera précisément la suivante ...

$$\begin{aligned}
Y_2 &= \mu + \phi\left(y_1 - \mu\right) + e_2 \\
Y_3 &= \mu + \phi^2\left(y_1 - \mu\right) + \phi e_2 + e_3 \\
&\vdots \\
Y_n &= \mu + \phi^{n-1}\left(y_1 - \mu\right) + \phi^{n-2} e_2 + ... + \phi e_{n-1} + e_n
\end{aligned}$$

... dont le Jacobien de la transformation est égal à 1, i.e.

$$Det\begin{bmatrix} \delta y_2/\delta e_2 & \cdots & \delta y_2/\delta e_n \\ \vdots & \ddots & \vdots \\ \delta y_n/\delta e_2 & \cdots & \delta y_n/\delta e_n \end{bmatrix} = 1 \quad .$$

Ainsi,

$$
\begin{aligned}
f\left(y_1,\ldots,y_n\right) &= f\left(y_2,\ldots,y_n \middle| y_1\right) f\left(y_1\right) \\
&= g\left(y_2-\mu-\phi\left(y_1-\mu\right),\ldots,y_n-\mu-\phi\left(y_n-\mu\right)\right) f\left(y_1\right) \\
&= \left(2\pi\sigma_e^2\right)^{-\frac{n}{2}}\left(1-\phi^2\right)^{\frac{1}{2}} \\
&\times \exp\left\{-\frac{1}{2\sigma_e^2}\left[\left(1-\phi^2\right)\left(y_1-\mu\right)^2+\sum_{t=2}^{n}\left[\left(y_t-\mu\right)-\phi\left(y_{t-1}-\mu\right)\right]^2\right]\right\}
\end{aligned}
$$

Nous obtenons ainsi la fonction de vraisemblance

$$
L\left(\phi,\mu,\sigma_e^2\right)=\left(2\pi\sigma_e^2\right)^{-\frac{n}{2}}\left(1-\phi^2\right)^{\frac{1}{2}}\exp\left(-\frac{S\left(\phi,\mu\right)}{2\sigma_e^2}\right) \quad ,
$$

où

$$S(\phi,\mu) = (1-\phi^2)(Y_1-\mu)^2 + \sum_{t=2}^{n}\left[(Y_t-\mu)-\phi(Y_{t-1}-\mu)\right]^2$$
$$= (1-\phi^2)(Y_1-\mu)^2 + S_c(\phi,\mu)$$

est dite la fonction des moindres carrés non conditionnelle (ULS).

Remarques : (1) Il apparait plus souvent approprié de maximiser la fonction de log vraisemblance ($\log L$) plutôt que la fonction de vraisemblance (L), ce qui donne dans ce cas-ci

$$\log L(\phi,\mu,\sigma_e^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma_e^2 + \frac{1}{2}\log(1-\phi^2) - \frac{S(\phi,\mu)}{2\sigma_e^2} \quad ,$$

Cette maximisation se fera en égalant à 0 les dérivées partielles d'ordre 1, dont l'évaluation numérique donnera $\hat{\phi}$, $\hat{\mu}$ et $\hat{\sigma}_e^2$.

Nous obtenons ainsi ces équations

$$\frac{\delta S(\phi,\mu)}{\delta\phi} = -\frac{\phi\sigma_e^2}{1-\phi^2} \quad , \quad \frac{\delta S(\phi,\mu)}{\delta\mu} = 0 \quad , \quad \sigma_e^2 = \frac{S(\phi,\mu)}{n} \quad .$$

(2) Dans la littérature sur l'estimation de paramètres de séries chronologiques, la méthode d'estimation ULS constitue un « compromis » entre la méthode des moindres carrés et la méthode du maximum de vraisemblance. Cette méthode consiste à minimiser la fonction $S(\phi,\mu)$ puis, à partir des estimés $\hat{\phi}$ et $\hat{\mu}$, à poser

$$\hat{\sigma}_e^2 = \frac{S(\hat{\phi},\hat{\mu})}{n} \quad .$$

(3) La méthode du maximum de vraisemblance est généralement recommandée pour un nombre de données relativement modéré.

7.4.2 Propriétés asymptotiques du MLE

Les propriétés asymptotiques des estimateurs du maximum de vraisemblance et des moindres carrés (conditionnels ou non conditionnels) sont identiques.

Ainsi, considérer un processus stationnaire ARMA(p,q)

$$\phi(B)Y_t = \theta(B)e_t \quad ,$$

où $e_t \sim WN(0,\sigma_e^2)$ et e_t indépendant de $Y_{t-1}, Y_{t-2}, \ldots$.

Alors, pour n grand,

- $\hat{\phi}_i \sim AN\left(\phi_i, \sigma^2_{\hat{\phi}_i} / n\right)$, $i = 1,2,\ldots,p$

- $\hat{\theta}_j \sim AN\left(\theta_j, \sigma^2_{\hat{\theta}_j} / n\right)$, $j = 1,2,\ldots,q$

Cas particuliers:

- AR(1) :

$$\hat{\phi} \sim AN\left(\phi, \frac{1-\phi^2}{n}\right) \quad .$$

- AR(2) :

$$\hat{\phi}_i \sim AN\left(\phi_i, \frac{1-\phi_i^2}{n}\right) \ , \ i = 1,2 \quad .$$

- MA(1) :

$$\hat{\theta} \sim AN\left(\theta, \frac{1-\theta^2}{n}\right) \quad .$$

- MA(2) :

$$\hat{\theta}_i \sim AN\left(\theta_i, \frac{1-\theta_i^2}{n}\right) \ , \ i = 1,2 \quad .$$

- ARMA(1,1) :

$$\hat{\phi} \sim AN\left(\phi, \frac{(1-\phi\theta)^2(\phi-\theta)^{-2}(1-\phi^2)}{n}\right) \quad ,$$

$$\hat{\theta} \sim AN\left(\theta, \frac{(1-\phi\theta)^2(\phi-\theta)^{-2}(1-\theta^2)}{n}\right) \quad .$$

Remarques : (1) Les estimateurs du maximum de vraisemblance, en plus d'obéir asymptotiquement à une loi normale, sont consistants.

(2) À partir de ces distributions asymptotiques, nous pouvons créer des intervalles de confiance pour les paramètres de notre processus ARMA(p,q).

7.4.3 Exemples

(1) Reprenons l'exemple 7.2.4 (1), et utilisons la méthode du maximum de vraisemblance pour estimer les paramètres du modèle AR(2) sensé représenter la série chronologique « hare^0.5 ».

```
> data(hare)
> arima(hare^0.5, order=c(2,0,0), method='ML')

Call:
arima(x = hare^0.5, order = c(2, 0, 0), method = "ML")

Coefficients:
         ar1      ar2    intercept
      1.3514  -0.7763     5.7133
s.e.  0.1286   0.1242     0.4753

sigma^2 estimated as 1.223: log likelihood = -48.46, aic = 102.91
```

Nous obtenons ainsi le modèle suivant

$$\sqrt{Y_t} - 5.7133 = 1.3514\left(\sqrt{Y_{t-1}} - 5.7133\right) - 0.7763\left(\sqrt{Y_{t-2}} - 5.7133\right) + e_t$$
$$\Updownarrow$$
$$\sqrt{Y_t} = 2.4276 + 1.3514\sqrt{Y_{t-1}} - 0.7763\sqrt{Y_{t-2}} + e_t$$

Remarques : (1) Le nombre 5.7133 est un estimé du maximum de vraisemblance de la moyenne de $\sqrt{Y_t}$.

(2) Si nous comparons le modèle précédent avec le modèle obtenu par la méthode des moindres carrés, soit

$$\sqrt{Y_t} = 2.3753 + 1.3631\sqrt{Y_{t-1}} - 0.7792\sqrt{Y_{t-2}} + e_t \quad ,$$

tous les paramètres diffèrent « quelque peu ».

(2) Reprenons l'exemple 7.2.4 (2), et utilisons la méthode du maximum de vraisemblance pour estimer les paramètres du modèle MA(1) sensé représenter la série chronologique « ∇log(oil.price) ».

```
> data(oil.price)
> arima(log(oil.price), order=c(0,1,1), method='ML')

Call:
arima(x = log(oil.price), order = c(0, 1, 1), method = "ML")

Coefficients:
         ma1
      0.2956
s.e.  0.0693

sigma^2 estimated as 0.006689: log likelihood = 260.29,
aic = -518.58
```

Nous obtenons ainsi le modèle suivant

$$\log Y_t = \log Y_{t-1} + e_t + 0.2956 e_{t-1} \quad .$$

Remarque : Si nous comparons le modèle précédent avec le modèle obtenu par la méthode des moindres carrés conditionnels, soit

$$\log Y_t = \log Y_{t-1} + e_t + 0.2731 e_{t-1} \quad ,$$

le paramètre diffère légèrement.

Notons tout de même que l'estimé des moindres carrés conditionnel de θ (-0.2731) se trouve dans l'intervalle à un écart-type de -0.2956 obtenu précédemment, soit

$$\left]-0.0693 - 0.2956, 0.0693 - 0.2956\right[\; = \; \left]-0.3649, -0.2263\right[\quad .$$