Estimation for MA(q) processes

The method of moments is not linear, nor efficient for MA(q) processes.

Estimation for MA(q) processes

The method of moments is not linear, nor efficient for MA(q) processes.
Simple example: estimation of MA(1) process:

$$\gamma(0) = (1+\vartheta^2)\sigma^2 \qquad \gamma(1) = \vartheta\sigma^2 \Longrightarrow \rho(1) = \frac{\vartheta}{1+\vartheta^2}.$$

Method of moments: $\gamma(0) = \hat{\gamma}(0), \quad \rho(1) = \hat{\rho}(1)$. Hence

$$\frac{\hat{\vartheta}}{1+\hat{\vartheta}^2} = \hat{\rho}(1) \quad \text{(a 2nd degree equation for } \hat{\vartheta}\text{).}$$

Estimation for MA(q) processes

The method of moments is not linear, nor efficient for MA(q) processes.
Simple example: estimation of MA(1) process:

$$\gamma(0) = (1+\vartheta^2)\sigma^2 \qquad \gamma(1) = \vartheta\sigma^2 \Longrightarrow \rho(1) = \frac{\vartheta}{1+\vartheta^2}.$$

Method of moments: $\gamma(0) = \hat{\gamma}(0), \quad \rho(1) = \hat{\rho}(1).$ Hence

$$\frac{\hat{\vartheta}}{1+\hat{\vartheta}^2} = \hat{\rho}(1) \quad \text{(a 2nd degree equation for } \hat{\vartheta}\text{)}.$$

It may have no solutions (if $|\hat{\rho}(1)| > 1/2$) or 2 solutions (if $|\hat{\rho}(1)| > 1/2$) of whoch we may choose that yielding an **invertible** process.

Estimation for MA(q) processes

The method of moments is not linear, nor efficient for MA(q) processes.
Simple example: estimation of MA(1) process:

$$\gamma(0) = (1+\vartheta^2)\sigma^2 \qquad \gamma(1) = \vartheta\sigma^2 \Longrightarrow \rho(1) = \frac{\vartheta}{1+\vartheta^2}.$$

Method of moments: $\gamma(0) = \hat{\gamma}(0), \quad \rho(1) = \hat{\rho}(1).$ Hence

$$\frac{\hat{\vartheta}}{1+\hat{\vartheta}^2} = \hat{\rho}(1) \quad \text{(a 2nd degree equation for } \hat{\vartheta}\text{)}.$$

It may have no solutions (if $|\hat{\rho}(1)| > 1/2$) or 2 solutions (if $|\hat{\rho}(1)| > 1/2$) of whoch we may choose that yielding an **invertible** process.
Moreover, it can be shown that it is **not efficient** (there are other simple estimators with asymptotically lower variance).

Innovations algorithm for estimation of MA(q) processes

The method consists in applying the innovations algorithm to the sample autocovariances:
Start with $\hat{v}_0 = \hat{\gamma}(0)$. Then for each n,

Innovations algorithm for estimation of MA(q) processes

The method consists in applying the innovations algorithm to the sample autocovariances:
Start with $\hat{v}_0 = \hat{\gamma}(0)$. Then for each n,

$$\hat{\vartheta}_{n,n} = \hat{\gamma}(n)/v_0$$

$$\hat{\vartheta}_{n,j} = [\hat{\gamma}(j) - \sum_{k=1}^{n-j} \hat{\vartheta}_{n-j,k}\hat{\vartheta}_{n,j+k}\hat{v}_{n-j-k}]/\hat{v}_{n-j}, \quad j = n-1, \ldots, 1.$$

Innovations algorithm for estimation of MA(q) processes

The method consists in applying the innovations algorithm to the sample autocovariances:

Start with $\hat{v}_0 = \hat{\gamma}(0)$. Then for each n,

$$\hat{\vartheta}_{n,n} = \hat{\gamma}(n)/v_0$$

$$\hat{\vartheta}_{n,j} = [\hat{\gamma}(j) - \sum_{k=1}^{n-j} \hat{\vartheta}_{n-j,k}\hat{\vartheta}_{n,j+k}\hat{v}_{n-j-k}]/\hat{v}_{n-j}, \quad j = n-1, \ldots, 1.$$

$$\hat{v}_n = \hat{\gamma}(0) - \sum_{j=1}^{n} \hat{\vartheta}^2_{n,j}\hat{v}_{n-j}.$$

Innovations algorithm for estimation of MA(q) processes

The method consists in applying the innovations algorithm to the sample autocovariances:
Start with $\hat{v}_0 = \hat{\gamma}(0)$. Then for each n,

$$\hat{\vartheta}_{n,n} = \hat{\gamma}(n)/v_0$$

$$\hat{\vartheta}_{n,j} = [\hat{\gamma}(j) - \sum_{k=1}^{n-j} \hat{\vartheta}_{n-j,k}\hat{\vartheta}_{n,j+k}\hat{v}_{n-j-k}]/\hat{v}_{n-j}, \quad j = n-1, \dots, 1.$$

$$\hat{v}_n = \hat{\gamma}(0) - \sum_{j=1}^{n} \hat{\vartheta}_{n,j}^2 \hat{v}_{n-j}.$$

$\hat{\vartheta}_{m,1}, \dots, \hat{\vartheta}_{m,q}$ are estimators of $\vartheta_1, \dots, \vartheta_q$ (if m large enough).

Innovations algorithm for estimation of MA(q) processes

The method consists in applying the innovations algorithm to the sample autocovariances:
Start with $\hat{v}_0 = \hat{\gamma}(0)$. Then for each n,

$$\hat{\vartheta}_{n,n} = \hat{\gamma}(n)/v_0$$

$$\hat{\vartheta}_{n,j} = [\hat{\gamma}(j) - \sum_{k=1}^{n-j} \hat{\vartheta}_{n-j,k}\hat{\vartheta}_{n,j+k}\hat{v}_{n-j-k}]/\hat{v}_{n-j}, \quad j = n-1, \ldots, 1.$$

$$\hat{v}_n = \hat{\gamma}(0) - \sum_{j=1}^{n} \hat{\vartheta}_{n,j}^2 \hat{v}_{n-j}.$$

$\hat{\vartheta}_{m,1}, \ldots, \hat{\vartheta}_{m,q}$ are estimators of $\vartheta_1, \ldots, \vartheta_q$ (if m large enough).
$\hat{v}_m$ an estimator of σ^2.

Fitted innovations algorithm. Intuitive explanation

$$X_{m+1} = Z_{m+1} + \vartheta_1 Z_m + \ldots + \vartheta_q Z_{m-q+1}.$$

Innovations algorithm on the true ACVF gives

$$\hat{X}_{m+1} = \vartheta_{m,1}(X_m - \hat{X}_m) + \ldots + \vartheta_{m,m}(X_1 - \hat{X}_1).$$

Fitted innovations algorithm. Intuitive explanation

$$X_{m+1} = Z_{m+1} + \vartheta_1 Z_m + \ldots + \vartheta_q Z_{m-q+1}.$$

Innovations algorithm on the true ACVF gives

$$\hat{X}_{m+1} = \vartheta_{m,1}(X_m - \hat{X}_m) + \ldots + \vartheta_{m,m}(X_1 - \hat{X}_1).$$

If the process is **invertible**, Z_m can be predicted from X_m, $X_{m-1}, \ldots$.
Hence for m large $X_m - \hat{X}_m \approx Z_m$.

Fitted innovations algorithm. Intuitive explanation

$$X_{m+1} = Z_{m+1} + \vartheta_1 Z_m + \ldots + \vartheta_q Z_{m-q+1}.$$

Innovations algorithm on the true ACVF gives

$$\hat{X}_{m+1} = \vartheta_{m,1}(X_m - \hat{X}_m) + \ldots + \vartheta_{m,m}(X_1 - \hat{X}_1).$$

If the process is **invertible**, Z_m can be predicted from X_m, $X_{m-1}, \ldots$.
Hence for m large $X_m - \hat{X}_m \approx Z_m$.
Writing

$$X_{m+1} = (X_{m+1} - \hat{X}_{m+1}) + \hat{X}_{m+1}$$
$$= (X_{m+1} - \hat{X}_{m+1}) + \vartheta_{m,1}(X_m - \hat{X}_m) + \vartheta_{m,m}(X_1 - \hat{X}_1)$$

we expect $\vartheta_{m,j} \approx \vartheta_j$ for $j = 1 \ldots q$, $\quad \vartheta_{m,j} \approx 0$ for $j > q$.

Convergence of fitted innovations estimator

This theorem (Brockwell-Davis, 1988) is stated for ARMA(p,q) processes

Theorem

Let $\{X_t\}$ *be the* **causal, invertible** *ARMA process specified by* $\varphi(B)X = \vartheta(B)Z$ *with* $\{Z_t|\} \sim IID(0, \sigma^2)$ *and* $\mathbb{E}(|Z_t|^4 < \infty$, *and let* $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\vartheta(z)}{\varphi(z)}$. *Then for any sequence* m_n *such that* $m_n < n$, $\lim_{n\to\infty} m_n = +\infty$ *and* $\lim_{n\to\infty} n^{-1/3} m_n = 0$ *and for any* $k > 0$

$$n^{1/2}(\vartheta_{m_n,1} - \psi_1, \ldots \vartheta_{m_n,k} - \psi_k) \Longrightarrow N(0, A) \quad \text{as } n \to \infty$$

where $A_{ij} = \sum_{r=1}^{\min(i,j)} \psi_{i-r}\psi_{j-r}$.

Moreover $\hat{v}_m \to \sigma^2$ *in probability.*

Remarks on convergence of fitted innovations

- $(\hat{\vartheta}_{q,1}, \ldots, \hat{\vartheta}_{q,q})$ is not a consistent estimator of $(\vartheta_1, \ldots, \vartheta_q)$. *It is not enough that $n \to \infty$, need also $m \to \infty$ (with appropriate speed).*

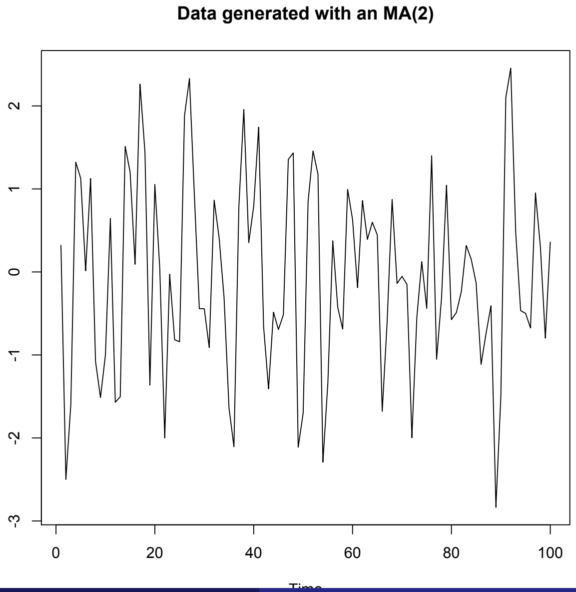
Remarks on convergence of fitted innovations

- $(\hat{\vartheta}_{q,1}, \ldots, \hat{\vartheta}_{q,q})$ is not a consistent estimator of $(\vartheta_1, \ldots, \vartheta_q)$. *It is not enough that $n \to \infty$, need also $m \to \infty$ (with appropriate speed).*
- in practice one has finite n and has to choose a finite m. Increase m until $(\hat{\vartheta}_{m,1}, \ldots, \hat{\vartheta}_{m,q})$ stabilize.

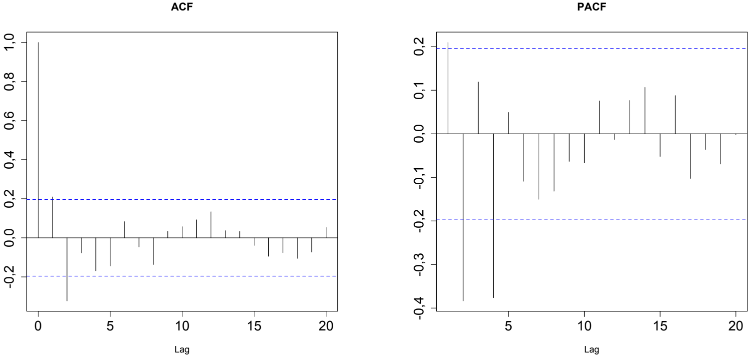
Remarks on convergence of fitted innovations

- $(\hat{\vartheta}_{q,1}, \ldots, \hat{\vartheta}_{q,q})$ is not a consistent estimator of $(\vartheta_1, \ldots, \vartheta_q)$. *It is not enough that $n \to \infty$, need also $m \to \infty$ (with appropriate speed).*
- in practice one has finite n and has to choose a finite m. Increase m until $(\hat{\vartheta}_{m,1}, \ldots, \hat{\vartheta}_{m,q})$ stabilize.
- asymptotic confidence intervals can be found from the elements of the matrix A.

An example on simulated data



ACF and PACF of simulated data



An MA(2) process is suggested by these plots

Using the fitted innovations algorithm

m	$\vartheta_{m,1}$	$\vartheta_{m,2}$	v_m
2	0.290	-0.323	1.028
3	0.336	-0.321	1.015
4	0.380	-0.448	0.977
5	0.399	-0.436	0.978
	...	...	
10	0.355	-0.508	0.971
	...	...	
15	0.360	-0.507	0.971
	...	...	

Using the fitted innovations algorithm

m	$\vartheta_{m,1}$	$\vartheta_{m,2}$	v_m
2	0.290	-0.323	1.028
3	0.336	-0.321	1.015
4	0.380	-0.448	0.977
5	0.399	-0.436	0.978
	...	...	
10	0.355	-0.508	0.971
	...	...	
15	0.360	-0.507	0.971
	...	...	

Confidence intervals:

$$A_{11} = 1 \quad \hat{A}_{22} = 1 + 0.36^2 \approx 1.1296$$

$$\vartheta_1 = 0.36 \pm 1.96\sqrt{\frac{1}{n}} = 0.36 \pm 0.20$$

$$\vartheta_2 = -0.507 \pm 1.96\sqrt{\frac{1.1296}{n}} = -0.51 \pm 0.21$$

'True' values were
$\vartheta_1 = 0.5$, $\vartheta_2 = -0.4$.

Using innovations for an ARMA(p,q)

Fitted innovations yield estimates $\hat{\vartheta}_{m,1}$, $\hat{\vartheta}_{m,2}, \dots$ for ψ_1, $\psi_2, \dots$.
As in an ARMA(p,q)

$$\psi_j = \vartheta_j + \sum_{i=1}^{\min(j,p)} \varphi_i \psi_{j-i} \quad j = 1, 2, \dots$$

we set

$$\hat{\vartheta}_{m,j} = \hat{\vartheta}_j + \sum_{i=1}^{\min(j,p)} \hat{\varphi}_i \hat{\vartheta}_{m,j-i} \quad j = 1, \dots, p+q.$$

Using innovations for an ARMA(p,q)

Fitted innovations yield estimates $\hat\vartheta_{m,1}$, $\hat\vartheta_{m,2}, \dots$ for ψ_1, $\psi_2, \dots$.
As in an ARMA(p,q)

$$\psi_j = \vartheta_j + \sum_{i=1}^{\min(j,p)} \varphi_i \psi_{j-i} \quad j = 1, 2, \dots$$

we set

$$\hat\vartheta_{m,j} = \hat\vartheta_j + \sum_{i=1}^{\min(j,p)} \hat\varphi_i \hat\vartheta_{m,j-i} \quad j = 1, \dots, p+q.$$

The equations for $j = q+1, \dots, q+p$ do not involve $\hat\vartheta$ and can be solved for $\hat\varphi_1, \dots, \hat\varphi_p$.
Then equations for $j = 1, , \dots, q$ yield $\hat\vartheta_j$ directly.

Using innovations for an ARMA(p,q)

Fitted innovations yield estimates $\hat{\vartheta}_{m,1}$, $\hat{\vartheta}_{m,2}, \dots$ for ψ_1, $\psi_2, \dots$.
As in an ARMA(p,q)

$$\psi_j = \vartheta_j + \sum_{i=1}^{\min(j,p)} \varphi_i \psi_{j-i} \quad j = 1, 2, \dots$$

we set

$$\hat{\vartheta}_{m,j} = \hat{\vartheta}_j + \sum_{i=1}^{\min(j,p)} \hat{\varphi}_i \hat{\vartheta}_{m,j-i} \quad j = 1, \dots, p+q.$$

The equations for $j = q+1, \dots, q+p$ do not involve $\hat{\vartheta}$ and can be solved for $\hat{\varphi}_1, \dots, \hat{\varphi}_p$.
Then equations for $j = 1, , \dots, q$ yield $\hat{\vartheta}_j$ directly.

Note that $\hat{\varphi}_1, \dots, \hat{\varphi}_p$ do not necessarily yield a **causal** process.

Maximum likelihood estimation

A standard method of parameter estimate is maximum likelihood estimation (MLE).

Maximum likelihood estimation

A standard method of parameter estimate is maximum likelihood estimation (MLE).
Likelihood of ARMA processes is computed assuming $\{Z_t\} \sim IIDN(0, \sigma^2)$.
Then $\{X_t\}$ is also normally distributed.

Maximum likelihood estimation

A standard method of parameter estimate is maximum likelihood estimation (MLE).
Likelihood of ARMA processes is computed assuming $\{Z_t\} \sim IIDN(0, \sigma^2)$.
Then $\{X_t\}$ is also normally distributed.

Example of AR(1): $X_t = \mu + \varphi(X_{t-1} - \mu) + Z_t$, $|\varphi| < 1$.
Then $X_t \sim N(\mu, \frac{\sigma^2}{1-\varphi^2})$.

Maximum likelihood estimation

A standard method of parameter estimate is maximum likelihood estimation (MLE).
Likelihood of ARMA processes is computed assuming $\{Z_t\} \sim IIDN(0, \sigma^2)$.
Then $\{X_t\}$ is also normally distributed.

Example of AR(1): $X_t = \mu + \varphi(X_{t-1} - \mu) + Z_t$, $|\varphi| < 1$.
Then $X_t \sim N(\mu, \frac{\sigma^2}{1-\varphi^2})$.
Joint distributions can be obtained by conditional distributions:
conditional on the value of X_{t-1}, $X_t \sim N(\mu + \varphi(X_{t-1} - \mu), \sigma^2)$.

Maximum likelihood estimation

A standard method of parameter estimate is maximum likelihood estimation (MLE).
Likelihood of ARMA processes is computed assuming $\{Z_t\} \sim IIDN(0, \sigma^2)$.
Then $\{X_t\}$ is also normally distributed.

Example of AR(1): $X_t = \mu + \varphi(X_{t-1} - \mu) + Z_t$, $|\varphi| < 1$.
Then $X_t \sim N(\mu, \frac{\sigma^2}{1-\varphi^2})$.
Joint distributions can be obtained by conditional distributions:
conditional on the value of X_{t-1}, $X_t \sim N(\mu + \varphi(X_{t-1} - \mu), \sigma^2)$.
Hence likelihood of data $(x_1, x_2, \ldots, x_n)$ can be obtained as

$$f(x_1)f(x_2|x_1)\cdots f(x_n|x_{n-1})$$

where $f(\cdot)$ is the (unconditional) density of X_t,
while $f(\cdot|x)$ is the density of X_t conditional on $X_{t-1} = x$.

Computing likelihood for AR(1)

$$f(x) = (2\pi\frac{\sigma^2}{1-\varphi^2})^{-1/2}\exp\{-\frac{(x-\mu)^2}{2\sigma^2/(1-\varphi^2)}\}$$
$$= \left(\frac{1-\varphi^2}{2\pi\sigma^2}\right)^{1/2}\exp\{-\frac{(1-\varphi^2)(x-\mu)^2}{2\sigma^2}\}$$
$$f(y,x) = (2\pi\sigma^2)^{-1/2}\exp\{-\frac{(y-\mu-\varphi(x-\mu))^2}{2\sigma^2}\}.$$

Computing likelihood for AR(1)

$$f(x) = (2\pi \frac{\sigma^2}{1-\varphi^2})^{-1/2} \exp\{-\frac{(x-\mu)^2}{2\sigma^2/(1-\varphi^2)}\}$$

$$= \left(\frac{1-\varphi^2}{2\pi\sigma^2}\right)^{1/2} \exp\{-\frac{(1-\varphi^2)(x-\mu)^2}{2\sigma^2}\}$$

$$f(y,x) = (2\pi\sigma^2)^{-1/2} \exp\{-\frac{(y-\mu-\varphi(x-\mu))^2}{2\sigma^2}\}.$$

Parameters of the model are μ, φ, σ^2:

$$L(\mu,\varphi,\sigma^2) = (1-\varphi^2)^{1/2}(2\pi\sigma^2)^{-n/2} \exp\{-\frac{(1-\varphi^2)(x_1-\mu)^2}{2\sigma^2}\}$$

$$\times \prod_{j=2}^{n} \exp\{-\frac{(x_j-\mu-\varphi(x_{j-1}-\mu))^2}{2\sigma^2}\}$$

Computing likelihood for AR(1)

$$f(x) = (2\pi \frac{\sigma^2}{1-\varphi^2})^{-1/2} \exp\{-\frac{(x-\mu)^2}{2\sigma^2/(1-\varphi^2)}\}$$

$$= \left(\frac{1-\varphi^2}{2\pi\sigma^2}\right)^{1/2} \exp\{-\frac{(1-\varphi^2)(x-\mu)^2}{2\sigma^2}\}$$

$$f(y,x) = (2\pi\sigma^2)^{-1/2} \exp\{-\frac{(y-\mu-\varphi(x-\mu))^2}{2\sigma^2}\}.$$

Parameters of the model are μ, φ, σ^2:

$$L(\mu,\varphi,\sigma^2) = (1-\varphi^2)^{1/2}(2\pi\sigma^2)^{-n/2} \exp\{-\frac{(1-\varphi^2)(x_1-\mu)^2}{2\sigma^2}\}$$

$$\times \prod_{j=2}^{n} \exp\{-\frac{(x_j-\mu-\varphi(x_{j-1}-\mu))^2}{2\sigma^2}\}$$

$$= \frac{(1-\varphi^2)^{1/2}}{(2\pi\sigma^2)^{n/2}} \exp\{-\frac{1}{2\sigma^2}\left[(1-\varphi^2)(x_1-\mu)^2 + \sum_{t=2}^{n}(x_t-\mu-\varphi(x_{t-1}-\mu))^2\right]\}$$

Maximum likelihood estimation: example of AR(1)

Assume $X_t = \mu + \varphi(X_{t-1} - \mu) + Z_t$ with $\{Z_t\} \sim IID\ N(0, \sigma^2)$. Then the likelihood given data $(x_1, x_2, \ldots, x_n)$ is

$$L(\mu, \varphi, \sigma^2) = f(x_1) f(x_2|x_1) \cdots f(x_n|x_{n-1})$$
$$= (2\pi\sigma^2)^{-n/2}(1-\varphi^2)^{1/2} \cdot \exp\{-\frac{S(\mu,\varphi)}{2\sigma^2}\}$$

with $S(\mu, \varphi) = (1-\varphi^2)(x_1 - \mu)^2 + \sum_{t=2}^{n}(x_t - \mu - \varphi(x_{t-1} - \mu))^2$

Maximum likelihood estimation: example of AR(1)

Assume $X_t = \mu + \varphi(X_{t-1} - \mu) + Z_t$ with $\{Z_t\} \sim IID\ N(0, \sigma^2)$. Then the likelihood given data $(x_1, x_2, \ldots, x_n)$ is

$$L(\mu, \varphi, \sigma^2) = f(x_1) f(x_2|x_1) \cdots f(x_n|x_{n-1})$$
$$= (2\pi\sigma^2)^{-n/2}(1-\varphi^2)^{1/2} \cdot \exp\{-\frac{S(\mu,\varphi)}{2\sigma^2}\}$$

$$\text{with } S(\mu,\varphi) = (1-\varphi^2)(x_1-\mu)^2 + \sum_{t=2}^{n}(x_t - \mu - \varphi(x_{t-1}-\mu))^2$$

Maximizing with respect to σ^2 (easier going to logarithms) yields

$$\hat{\sigma^2} = n^{-1} S(\hat{\mu}, \hat{\varphi})$$

where $\hat{\mu}$ and $\hat{\varphi}$ are MLE.

Maximum likelihood estimation: example of AR(1)

Assume $X_t = \mu + \varphi(X_{t-1} - \mu) + Z_t$ with $\{Z_t\} \sim IID\ N(0, \sigma^2)$. Then the likelihood given data $(x_1, x_2, \ldots, x_n)$ is

$$L(\mu, \varphi, \sigma^2) = f(x_1) f(x_2|x_1) \cdots f(x_n|x_{n-1})$$
$$= (2\pi\sigma^2)^{-n/2}(1-\varphi^2)^{1/2} \cdot \exp\{-\frac{S(\mu,\varphi)}{2\sigma^2}\}$$
$$\text{with } S(\mu,\varphi) = (1-\varphi^2)(x_1-\mu)^2 + \sum_{t=2}^{n}(x_t - \mu - \varphi(x_{t-1}-\mu))^2$$

Maximizing with respect to σ^2 (easier going to logarithms) yields

$$\hat{\sigma^2} = n^{-1} S(\hat{\mu}, \hat{\varphi})$$

where $\hat{\mu}$ and $\hat{\varphi}$ are MLE. Now substitute $\hat{\sigma^2}$ and maximize w.r. to (μ, φ).

MLE and least square estimation for AR(1)

Taking logarithm of L, substituting $\hat{\sigma^2}$ and ignoring additive constants yields $-\frac{n}{2}\log(n^{-1}S(\mu,\varphi)) + \frac{1}{2}\log(1-\varphi^2)$.

MLE and least square estimation for AR(1)

Taking logarithm of L, substituting $\hat{\sigma^2}$ and ignoring additive constants yields $-\frac{n}{2}\log(n^{-1}S(\mu,\varphi)) + \frac{1}{2}\log(1-\varphi^2)$. Hence (μ,φ) that maximize L minimize

$$\mathcal{L}(\mu,\varphi) = \log(n^{-1}S(\mu,\varphi)) - n^{-1}\log(1-\varphi^2).$$

As it is, the problem is highly nonlinear and requires numerical solutions.

MLE and least square estimation for AR(1)

Taking logarithm of L, substituting $\hat{\sigma}^2$ and ignoring additive constants yields $-\frac{n}{2}\log(n^{-1}S(\mu,\varphi)) + \frac{1}{2}\log(1-\varphi^2)$. Hence (μ,φ) that maximize L minimize

$$\mathcal{L}(\mu,\varphi) = \log(n^{-1}S(\mu,\varphi)) - n^{-1}\log(1-\varphi^2).$$

As it is, the problem is highly nonlinear and requires numerical solutions. Dropping the last term, one may proceed by *unconditional* least squares, i.e. minimize

$$S(\mu,\varphi) = (1-\varphi^2)(x_1-\mu)^2 + \sum_{t=2}^{n}(x_t-\mu-\varphi(x_{t-1}-\mu))^2,$$

MLE and least square estimation for AR(1)

Taking logarithm of L, substituting $\hat{\sigma^2}$ and ignoring additive constants yields $-\frac{n}{2}\log(n^{-1}S(\mu,\varphi)) + \frac{1}{2}\log(1-\varphi^2)$. Hence (μ,φ) that maximize L minimize

$$\mathcal{L}(\mu,\varphi) = \log(n^{-1}S(\mu,\varphi)) - n^{-1}\log(1-\varphi^2).$$

As it is, the problem is highly nonlinear and requires numerical solutions. Dropping the last term, one may proceed by *unconditional* least squares, i.e. minimize

$$S(\mu,\varphi) = (1-\varphi^2)(x_1-\mu)^2 + \sum_{t=2}^{n}(x_t-\mu-\varphi(x_{t-1}-\mu))^2,$$

or with a further simplification by *conditional* least squares, i.e. minimize

$$S_c(\mu,\varphi) = \sum_{t=2}^{n}(x_t-\mu-\varphi(x_{t-1}-\mu))^2.$$

Conditional least squares

$$S_c(\mu, \varphi) = \sum_{t=2}^{n} (x_t - \mu - \varphi(x_{t-1} - \mu))^2 = \sum_{t=2}^{n} (x_t - (\alpha + \varphi x_{t-1}))^2$$

with $\alpha = \mu(1 - \varphi)$.

Conditional least squares

$$S_c(\mu,\varphi) = \sum_{t=2}^{n}(x_t - \mu - \varphi(x_{t-1} - \mu))^2 = \sum_{t=2}^{n}(x_t - (\alpha + \varphi x_{t-1}))^2$$

with $\alpha = \mu(1-\varphi)$.
This is a linear regression and yields

Conditional least squares

$$S_c(\mu,\varphi) = \sum_{t=2}^{n} (x_t - \mu - \varphi(x_{t-1} - \mu))^2 = \sum_{t=2}^{n} (x_t - (\alpha + \varphi x_{t-1}))^2$$

with $\alpha = \mu(1-\varphi)$.
This is a linear regression and yields

$$\hat{\alpha}_c = \bar{x}_{(2)} - \hat{\varphi}_c \bar{x}_{(1)} \quad \text{with } \bar{x}_{(1)} = \frac{\sum_{t=1}^{n-1} x_t}{n-1} \text{ and } \bar{x}_{(2)} = \frac{\sum_{t=2}^{n} x_t}{n-1}.$$

Conditional least squares

$$S_c(\mu,\varphi) = \sum_{t=2}^{n}(x_t - \mu - \varphi(x_{t-1} - \mu))^2 = \sum_{t=2}^{n}(x_t - (\alpha + \varphi x_{t-1}))^2$$

with $\alpha = \mu(1-\varphi)$.

This is a linear regression and yields

$$\hat{\alpha}_c = \bar{x}_{(2)} - \hat{\varphi}_c \bar{x}_{(1)} \quad \text{with } \bar{x}_{(1)} = \frac{\sum_{t=1}^{n-1} x_t}{n-1} \text{ and } \bar{x}_{(2)} = \frac{\sum_{t=2}^{n} x_t}{n-1}.$$

$$i.e.\ \hat{\mu}_c = \frac{\bar{x}_{(2)} - \hat{\varphi}_c \bar{x}_{(1)}}{1 - \hat{\varphi}_c}$$

Conditional least squares

$$S_c(\mu,\varphi) = \sum_{t=2}^{n}(x_t - \mu - \varphi(x_{t-1}-\mu))^2 = \sum_{t=2}^{n}(x_t - (\alpha + \varphi x_{t-1}))^2$$

with $\alpha = \mu(1-\varphi)$.

This is a linear regression and yields

$$\hat{\alpha}_c = \bar{x}_{(2)} - \hat{\varphi}_c \bar{x}_{(1)} \quad \text{with } \bar{x}_{(1)} = \frac{\sum_{t=1}^{n-1} x_t}{n-1} \text{ and } \bar{x}_{(2)} = \frac{\sum_{t=2}^{n} x_t}{n-1}.$$

$$i.e.\ \hat{\mu}_c = \frac{\bar{x}_{(2)} - \hat{\varphi}_c \bar{x}_{(1)}}{1-\hat{\varphi}_c} \approx \bar{x}$$

Conditional least squares

$$S_c(\mu,\varphi) = \sum_{t=2}^{n} (x_t - \mu - \varphi(x_{t-1} - \mu))^2 = \sum_{t=2}^{n} (x_t - (\alpha + \varphi x_{t-1}))^2$$

with $\alpha = \mu(1-\varphi)$.
This is a linear regression and yields

$$\hat{\alpha}_c = \bar{x}_{(2)} - \hat{\varphi}_c \bar{x}_{(1)} \quad \text{with } \bar{x}_{(1)} = \frac{\sum_{t=1}^{n-1} x_t}{n-1} \text{ and } \bar{x}_{(2)} = \frac{\sum_{t=2}^{n} x_t}{n-1}.$$

$$i.e.\ \hat{\mu}_c = \frac{\bar{x}_{(2)} - \hat{\varphi}_c \bar{x}_{(1)}}{1 - \hat{\varphi}_c} \approx \bar{x}$$

$$\hat{\varphi}_c = \frac{\sum_{t=2}^{n} (x_t - \bar{x}_{(2)})(x_{t-1} - \bar{x}_{(1)})}{\sum_{t=2}^{n} (x_t - \bar{x}_{(2)})^2}$$

Conditional least squares

$$S_c(\mu,\varphi) = \sum_{t=2}^{n}(x_t - \mu - \varphi(x_{t-1} - \mu))^2 = \sum_{t=2}^{n}(x_t - (\alpha + \varphi x_{t-1}))^2$$

with $\alpha = \mu(1-\varphi)$.
This is a linear regression and yields

$$\hat{\alpha}_c = \bar{x}_{(2)} - \hat{\varphi}_c \bar{x}_{(1)} \quad \text{with } \bar{x}_{(1)} = \frac{\sum_{t=1}^{n-1} x_t}{n-1} \text{ and } \bar{x}_{(2)} = \frac{\sum_{t=2}^{n} x_t}{n-1}.$$

$$i.e.\ \hat{\mu}_c = \frac{\bar{x}_{(2)} - \hat{\varphi}_c \bar{x}_{(1)}}{1 - \hat{\varphi}_c} \approx \bar{x}$$

$$\hat{\varphi}_c = \frac{\sum_{t=2}^{n}(x_t - \bar{x}_{(2)})(x_{t-1} - \bar{x}_{(1)})}{\sum_{t=2}^{n}(x_t - \bar{x}_{(2)})^2} \approx \hat{\rho}(1)$$

Conditional least squares

$$S_c(\mu,\varphi) = \sum_{t=2}^{n}(x_t - \mu - \varphi(x_{t-1} - \mu))^2 = \sum_{t=2}^{n}(x_t - (\alpha + \varphi x_{t-1}))^2$$

with $\alpha = \mu(1-\varphi)$.
This is a linear regression and yields

$$\hat{\alpha}_c = \bar{x}_{(2)} - \hat{\varphi}_c \bar{x}_{(1)} \quad \text{with } \bar{x}_{(1)} = \frac{\sum_{t=1}^{n-1} x_t}{n-1} \text{ and } \bar{x}_{(2)} = \frac{\sum_{t=2}^{n} x_t}{n-1}.$$

$$i.e.\ \hat{\mu}_c = \frac{\bar{x}_{(2)} - \hat{\varphi}_c \bar{x}_{(1)}}{1-\hat{\varphi}_c} \approx \bar{x}$$

$$\hat{\varphi}_c = \frac{\sum_{t=2}^{n}(x_t - \bar{x}_{(2)})(x_{t-1} - \bar{x}_{(1)})}{\sum_{t=2}^{n}(x_t - \bar{x}_{(2)})^2} \approx \hat{\rho}(1)$$

Estimates very similar to Yule-Walker.

Conditional least squares

$$S_c(\mu,\varphi) = \sum_{t=2}^{n}(x_t - \mu - \varphi(x_{t-1}-\mu))^2 = \sum_{t=2}^{n}(x_t - (\alpha + \varphi x_{t-1}))^2$$

with $\alpha = \mu(1-\varphi)$.
This is a linear regression and yields

$$\hat\alpha_c = \bar x_{(2)} - \hat\varphi_c \bar x_{(1)} \quad \text{with } \bar x_{(1)} = \frac{\sum_{t=1}^{n-1} x_t}{n-1} \text{ and } \bar x_{(2)} = \frac{\sum_{t=2}^{n} x_t}{n-1}.$$

$$i.e.\ \hat\mu_c = \frac{\bar x_{(2)} - \hat\varphi_c \bar x_{(1)}}{1-\hat\varphi_c} \approx \bar x$$

$$\hat\varphi_c = \frac{\sum_{t=2}^{n}(x_t - \bar x_{(2)})(x_{t-1} - \bar x_{(1)})}{\sum_{t=2}^{n}(x_t - \bar x_{(2)})^2} \approx \hat\rho(1)$$

Estimates very similar to Yule-Walker. $\hat\mu \approx \bar x$ often suggests to assume $Y_t = X_t - \bar X$ has mean 0 in more complex models.

Likelihood function

Assume $\{X_t\}$ is a 0-mean ARMA(p,q) process with Gaussian distribution,

$$\mathbb{X}_n = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \text{ the random vector} \quad \text{and} \quad \mathbf{x}_n = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ the observed data.}$$

The likelihood can be written in terms of Γ_n, the covariance matrix of $\mathbb{X}_n$:

$$L(\Gamma_n) = (2\pi)^{-n/2} \left|\Gamma_n\right|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{x}_n^t \Gamma_n^{-1} \mathbf{x}_n\right\}.$$

Γ_n will then be a function of $\phi = (\varphi_1, \ldots, \varphi_p)$, $\vartheta = (\vartheta_1, \ldots, \vartheta_q)$, σ^2.

Likelihood function

Assume $\{X_t\}$ is a 0-mean ARMA(p,q) process with Gaussian distribution,

$$\mathbb{X}_n = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \text{ the random vector} \quad \text{and} \quad \mathbf{x}_n = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ the observed data.}$$

The likelihood can be written in terms of Γ_n, the covariance matrix of $\mathbb{X}_n$:

$$L(\Gamma_n) = (2\pi)^{-n/2} |\Gamma_n|^{-1/2} \exp\left\{ -\frac{1}{2} \mathbf{x}_n^t \Gamma_n^{-1} \mathbf{x}_n \right\}.$$

Γ_n will then be a function of $\phi = (\varphi_1, \ldots, \varphi_p)$, $\vartheta = (\vartheta_1, \ldots, \vartheta_q)$, σ^2.

$|\Gamma_n|$ and $\mathbf{x}_n^t \Gamma_n^{-1} \mathbf{x}_n$ can be computed with the innovations algorithm.

Recursive computation of the likelihood

We showed $\mathbb{X}_n = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = C_n(\mathbb{X}_n - \hat{\mathbb{X}}_n)$, where $C_n = \begin{pmatrix} 1 & 0 & 0 \\ \bullet & \ddots & \vdots \\ \bullet & \bullet & 1 \end{pmatrix}$ and

$$\mathbb{V}(\mathbb{X}_n - \hat{\mathbb{X}}_n) = D_n = \text{diag}(v_0 \cdots v_{n-1}) \text{ with } v_j = \mathbb{E}(X_{j+1} - \hat{X}_{j+1})^2.$$

Recursive computation of the likelihood

We showed $\mathbb{X}_n = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = C_n(\mathbb{X}_n - \hat{\mathbb{X}}_n)$, where $C_n = \begin{pmatrix} 1 & 0 & 0 \\ \bullet & \ddots & \vdots \\ \bullet & \bullet & 1 \end{pmatrix}$ and

$$\mathbb{V}(\mathbb{X}_n - \hat{\mathbb{X}}_n) = D_n = \mathrm{diag}(v_0 \cdots v_{n-1}) \text{ with } v_j = \mathbb{E}(X_{j+1} - \hat{X}_{j+1})^2.$$

Furthermore $\Gamma_n = \mathbb{V}(\mathbb{X}_n) = C_n D_n C_n^t$.

Recursive computation of the likelihood

We showed $\mathbb{X}_n = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = C_n(\mathbb{X}_n - \hat{\mathbb{X}}_n)$, where $C_n = \begin{pmatrix} 1 & 0 & 0 \\ \bullet & \ddots & \vdots \\ \bullet & \bullet & 1 \end{pmatrix}$ and

$$\mathbb{V}(\mathbb{X}_n - \hat{\mathbb{X}}_n) = D_n = \mathrm{diag}(v_0 \cdots v_{n-1}) \text{ with } v_j = \mathbb{E}(X_{j+1} - \hat{X}_{j+1})^2.$$

Furthermore $\Gamma_n = \mathbb{V}(\mathbb{X}_n) = C_n D_n C_n^t$.

Then

$$\mathbb{X}_n^t \Gamma_n^{-1} \mathbb{X}_n = (\mathbb{X}_n - \hat{\mathbb{X}}_n)^t C_n^t \Gamma_n^{-1} C_n (\mathbb{X}_n - \hat{\mathbb{X}}_n)$$

Recursive computation of the likelihood

We showed $\mathbb{X}_n = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = C_n(\mathbb{X}_n - \hat{\mathbb{X}}_n)$, where $C_n = \begin{pmatrix} 1 & 0 & 0 \\ \bullet & \ddots & \vdots \\ \bullet & \bullet & 1 \end{pmatrix}$ and

$$\mathbb{V}(\mathbb{X}_n - \hat{\mathbb{X}}_n) = D_n = \text{diag}(v_0 \cdots v_{n-1}) \text{ with } v_j = \mathbb{E}(X_{j+1} - \hat{X}_{j+1})^2.$$

Furthermore $\Gamma_n = \mathbb{V}(\mathbb{X}_n) = C_n D_n C_n^t$.

Then

$$\mathbb{X}_n^t \Gamma_n^{-1} \mathbb{X}_n = (\mathbb{X}_n - \hat{\mathbb{X}}_n)^t C_n^t \Gamma_n^{-1} C_n (\mathbb{X}_n - \hat{\mathbb{X}}_n)$$

$$= (\mathbb{X}_n - \hat{\mathbb{X}}_n)^t C_n^t (C_n^t)^{-1} D_n^{-1} C_n^{-1} (\mathbb{X}_n - \hat{\mathbb{X}}_n)$$

Recursive computation of the likelihood

We showed $\mathbb{X}_n = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = C_n(\mathbb{X}_n - \hat{\mathbb{X}}_n)$, where $C_n = \begin{pmatrix} 1 & 0 & 0 \\ \bullet & \ddots & \vdots \\ \bullet & \bullet & 1 \end{pmatrix}$ and

$$\mathbb{V}(\mathbb{X}_n - \hat{\mathbb{X}}_n) = D_n = \text{diag}(v_0 \cdots v_{n-1}) \text{ with } v_j = \mathbb{E}(X_{j+1} - \hat{X}_{j+1})^2.$$

Furthermore $\Gamma_n = \mathbb{V}(\mathbb{X}_n) = C_n D_n C_n^t$.

Then

$$\mathbb{X}_n^t \Gamma_n^{-1} \mathbb{X}_n = (\mathbb{X}_n - \hat{\mathbb{X}}_n)^t C_n^t \Gamma_n^{-1} C_n (\mathbb{X}_n - \hat{\mathbb{X}}_n)$$

$$= (\mathbb{X}_n - \hat{\mathbb{X}}_n)^t C_n^t (C_n^t)^{-1} D_n^{-1} C_n^{-1} (\mathbb{X}_n - \hat{\mathbb{X}}_n) = \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{v_{j-1}}.$$

Recursive computation of the likelihood

We showed $\mathbb{X}_n = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = C_n(\mathbb{X}_n - \hat{\mathbb{X}}_n)$, where $C_n = \begin{pmatrix} 1 & 0 & 0 \\ \bullet & \ddots & \vdots \\ \bullet & \bullet & 1 \end{pmatrix}$ and

$$\mathbb{V}(\mathbb{X}_n - \hat{\mathbb{X}}_n) = D_n = \mathrm{diag}(v_0 \cdots v_{n-1}) \text{ with } v_j = \mathbb{E}(X_{j+1} - \hat{X}_{j+1})^2.$$

Furthermore $\Gamma_n = \mathbb{V}(\mathbb{X}_n) = C_n D_n C_n^t$.

Then

$$\mathbb{X}_n^t \Gamma_n^{-1} \mathbb{X}_n = (\mathbb{X}_n - \hat{\mathbb{X}}_n)^t C_n^t \Gamma_n^{-1} C_n (\mathbb{X}_n - \hat{\mathbb{X}}_n)$$

$$= (\mathbb{X}_n - \hat{\mathbb{X}}_n)^t C_n^t (C_n^t)^{-1} D_n^{-1} C_n^{-1} (\mathbb{X}_n - \hat{\mathbb{X}}_n) = \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{v_{j-1}}.$$

Finally $|\Gamma_n| = |C_n|^2 |D_n| = \prod_{i=0}^{n-1} v_i$, as C_n is triangular with all 1's on the diagonal.

Likelihood for an ARMA process: final expressions

Let $(\phi, \vartheta, \sigma^2)$ the parameters of an ARMA model. We found

$$L(\phi, \vartheta, \sigma^2) = (2\pi)^{-n/2}(v_0 \dots v_{n-1})^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n} \frac{(x_j - \hat{x}_j)^2}{v_{j-1}} \right\}$$

with v_j computed through the innovation algorithm.

Likelihood for an ARMA process: final expressions

Let $(\phi, \vartheta, \sigma^2)$ the parameters of an ARMA model. We found

$$L(\phi, \vartheta, \sigma^2) = (2\pi)^{-n/2}(v_0 \dots v_{n-1})^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n} \frac{(x_j - \hat{x}_j)^2}{v_{j-1}} \right\}$$

with v_j computed through the innovation algorithm.

Applying the innovations to $W_t = \begin{cases} \sigma^{-1} X_t & t \leq m = \max\{p, q\} \\ \sigma^{-1}(\Phi(B)X)_t & t > m \end{cases}$

and letting $r_n = \mathbb{E}(W_{n+1} - \hat{W}_{n+1})^2$,

Likelihood for an ARMA process: final expressions

Let $(\phi, \vartheta, \sigma^2)$ the parameters of an ARMA model. We found

$$L(\phi, \vartheta, \sigma^2) = (2\pi)^{-n/2}(v_0 \dots v_{n-1})^{-1/2} \exp\left\{ -\frac{1}{2}\sum_{j=1}^{n} \frac{(x_j - \hat{x}_j)^2}{v_{j-1}} \right\}$$

with v_j computed through the innovation algorithm.

Applying the innovations to $W_t = \begin{cases} \sigma^{-1} X_t & t \leq m = \max\{p, q\} \\ \sigma^{-1}(\Phi(B)X)_t & t > m \end{cases}$

and letting $r_n = \mathbb{E}(W_{n+1} - \hat{W}_{n+1})^2$, one may write

$$L(\phi, \vartheta, \sigma^2) = (2\pi\sigma^2)^{-n/2}(r_0 \dots r_{n-1})^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} S(\phi, \vartheta) \right\}$$

with $S(\phi, \vartheta) = \sum_{j=1}^{n} \frac{(x_j - \hat{x}_j)^2}{r_{j-1}}$.

MLE computations for ARMA

$$L(\phi, \vartheta, \sigma^2) = (2\pi\sigma^2)^{-n/2}(r_0 \ldots r_{n-1})^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} S(\phi, \vartheta)\right\}$$

with $S(\phi, \vartheta) = \sum_{j=1}^{n} \frac{(x_j - \hat{x}_j)^2}{r_{j-1}}$.

Setting equal to 0 the derivative of L (or $\log(L)$) w.r. to σ^2, one obtains

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

MLE computations for ARMA

$$L(\phi, \vartheta, \sigma^2) = (2\pi\sigma^2)^{-n/2}(r_0 \dots r_{n-1})^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} S(\phi, \vartheta)\right\}$$

with $S(\phi, \vartheta) = \sum_{j=1}^{n} \frac{(x_j - \hat{x}_j)^2}{r_{j-1}}$.

Setting equal to 0 the derivative of L (or $\log(L)$) w.r. to σ^2, one obtains

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

Then $(\hat{\phi}, \hat{\vartheta})$ must be the values that minimize

$$LL(\phi, \vartheta) = \log(\frac{S(\phi, \vartheta)}{n}) + \frac{1}{n} \sum_{i=1}^{n} \log(r_{i-1}).$$

MLE computations for ARMA

$$L(\phi, \vartheta, \sigma^2) = (2\pi\sigma^2)^{-n/2}(r_0 \dots r_{n-1})^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} S(\phi, \vartheta)\right\}$$

with $S(\phi, \vartheta) = \sum_{j=1}^{n} \frac{(x_j - \hat{x}_j)^2}{r_{j-1}}$.

Setting equal to 0 the derivative of L (or $\log(L)$) w.r.t. to σ^2, one obtains

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

Then $(\hat{\phi}, \hat{\vartheta})$ must be the values that minimize

$$LL(\phi, \vartheta) = \log(\frac{S(\phi, \vartheta)}{n}) + \frac{1}{n} \sum_{i=1}^{n} \log(r_{i-1}).$$

The *least square estimate* instead minimizes $S(\phi, \vartheta)$.

Remarks on MLE

- Minimization has to be accomplished numerically with an iterative procedure.

Remarks on MLE

- Minimization has to be accomplished numerically with an iterative procedure.
  - An initial estimate (ϕ^0, ϑ^0) has to be supplied; as minimization in a high-dimensional parameter space may be hard, it is useful that it is close to the minimum.

Remarks on MLE

- Minimization has to be accomplished numerically with an iterative procedure.
  - An initial estimate (ϕ^0, ϑ^0) has to be supplied; as minimization in a high-dimensional parameter space may be hard, it is useful that it is close to the minimum.
  - Simpler methods (Yule-Walker, Burg, innovations, Hannan-Rissanen) may be used to provide (ϕ^0, ϑ^0).

Remarks on MLE

- Minimization has to be accomplished numerically with an iterative procedure.
  - An initial estimate (ϕ^0, ϑ^0) has to be supplied; as minimization in a high-dimensional parameter space may be hard, it is useful that it is close to the minimum.
  - Simpler methods (Yule-Walker, Burg, innovations, Hannan-Rissanen) may be used to provide (ϕ^0, ϑ^0).
  - MLE estimate is searched among those causal. Hence the initial estimate has to be causal (*Yule-Walker and Burg ensure this*).

Remarks on MLE

- Minimization has to be accomplished numerically with an iterative procedure.
  - An initial estimate (ϕ^0, ϑ^0) has to be supplied; as minimization in a high-dimensional parameter space may be hard, it is useful that it is close to the minimum.
  - Simpler methods (Yule-Walker, Burg, innovations, Hannan-Rissanen) may be used to provide (ϕ^0, ϑ^0).
  - MLE estimate is searched among those causal. Hence the initial estimate has to be causal (*Yule-Walker and Burg ensure this*).
- $r_0, \ldots, r_{n-1}$ depend only on parameters (ϕ, ϑ) and not on observed data; the same is true for $\vartheta_{n,j}$ of innovations algorithm. Data enter likelihood only through the terms $(x_j - \hat{x}_j)^2$ in $S(\phi, \vartheta)$.

Asymptotic distribution of MLE

Theorem

Let $\{X_t\}$ *be a causal and invertible ARMA(p,q) process satisfying*

$$\Phi(B)X = \Theta(B)Z, \qquad \{Z_t\} \sim IID(0, \sigma^2).$$

Let $(\hat{\phi}, \hat{\vartheta})$ *the values that minimize* $LL_n(\phi, \vartheta)$ *among those yielding a causal and invertible ARMA process*

Asymptotic distribution of MLE

Theorem

Let $\{X_t\}$ be a causal and invertible ARMA(p,q) process satisfying

$$\Phi(B)X = \Theta(B)Z, \qquad \{Z_t\} \sim IID(0, \sigma^2).$$

Let $(\hat{\phi}, \hat{\vartheta})$ the values that minimize $LL_n(\phi, \vartheta)$ among those yielding a causal and invertible ARMA process , and let $\hat{\sigma^2} = \dfrac{S(\hat{\phi}, \hat{\vartheta})}{n}$. Then

$$n^{1/2}((\hat{\phi}, \hat{\vartheta}) - (\phi, \vartheta)) \Longrightarrow N(0, W) \qquad \text{and } \hat{\sigma^2} \xrightarrow{a.s.} \sigma^2$$

Asymptotic distribution of MLE

Theorem

Let $\{X_t\}$ be a causal and invertible ARMA(p,q) process satisfying

$$\Phi(B)X = \Theta(B)Z, \qquad \{Z_t\} \sim IID(0, \sigma^2).$$

Let $(\hat{\phi}, \hat{\vartheta})$ the values that minimize $LL_n(\phi, \vartheta)$ among those yielding a causal and invertible ARMA process , and let $\hat{\sigma^2} = \dfrac{S(\hat{\phi}, \hat{\vartheta})}{n}$. Then

$$n^{1/2}((\hat{\phi}, \hat{\vartheta}) - (\phi, \vartheta)) \Longrightarrow N(0, W) \qquad \text{and } \hat{\sigma^2} \xrightarrow{a.s.} \sigma^2$$

$$W = \sigma^2 \begin{pmatrix} \mathbb{E}(\mathbb{U}_t \mathbb{U}_t^t) & \mathbb{E}(\mathbb{U}_t \mathbb{V}_t^t) \\ \mathbb{E}(\mathbb{V}_t \mathbb{U}_t^t) & \mathbb{E}(\mathbb{V}_t \mathbb{V}_t^t) \end{pmatrix}^{-1} \text{with } \mathbb{U}_t = \begin{pmatrix} U_t \\ \vdots \\ U_{t-p+1} \end{pmatrix} \ \mathbb{V}_t = \begin{pmatrix} V_t \\ \vdots \\ V_{t-q+1} \end{pmatrix}$$

$$(\Phi(B)U)_t = Z_t \qquad (\Theta(B)V)_t = Z_t.$$