PSTAT 174/274 Time Series

Tomoyuki Ichiba
Department of Statistics & Applied Probability,
Center for Financial Mathematics and Actuarial Research,
University of California, Santa Barbara

Spring 2024
Lecture 8: Estimation

Review for the midterm: Lecture 1 - 8

- White noise
- ACF, CCF,
- Difference equation
- ARIMA (p, d, q), ARMA (p, q)
- Prediction, best linear predictor;

- Practice midterm - NO HOMEWORK over this weekend, instead, please work on the practice midterm posted.

Estimation Problem

We shall estimate the coefficients

$$\phi_1, \ldots, \phi_p\,, \quad \theta_1, \ldots, \theta_q$$

for the ARMA (p, q) process $w \sim wn(0, \sigma^2)$

$$x_t = \underbrace{\phi_1}_{\text{unknown}} x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

for $t \in \mathbb{Z}$. How to estimate ϕ and θ.

- Method of Moments estimators - matching the moments
- Maximum Likelihood estimators

YULE-WALKER equation

Assume $\mathbb{E}[x_t]=0$, $\mathbb{E}[x_t \cdot x_{t-h}] = Cov(x_t, x_{t-h}) = \gamma(h)$

- Recall that the autocovariance γ of the AR (p) model

$\mathbb{E}[x_{t-1} \cdot x_{t-h}] = Cov(x_{t-1}, x_{t-h}) = \gamma(h-1)$

$\times x_{t-h}$ both side

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t ; \quad t \in \mathbb{Z}$$

$x_t x_{t-h} = \phi_1 x_{t-1} x_{t-h} + \cdots + \phi_p x_{t-p} x_{t-h} + w_t \cdot x_{t-h}$

satisfies the difference equations (YULE-WALKER equations)

$$\gamma(h) = \phi_1 \gamma(h-1) + \cdots + \phi_p \gamma(h-p) , \quad h = 1, \ldots, p ,$$

last term: $\mathbb{E}[x_{t-h} \cdot w_t] = 0 \cdot 0 = 0$

with $\sigma^2 = \gamma(0) - \phi_1 \gamma(1) - \cdots - \phi_p \gamma(p)$. In the matrix representation

use $Cov(x_t, w_t)$

$x_{t-h} \cdot w_t$ is not corr. $\because$ x_t is causal : x_t can be explained by only previous w's.

$$\Gamma \phi = \gamma ,$$

$\gamma(h) = \phi_1 \gamma(h-1)$

$\ldots, \ldots$

where

$$\Gamma := \{\gamma(i-j)\}_{i,j=1}^{p} , \quad \phi := (\phi_1, \ldots, \phi_p)' , \quad \gamma := (\gamma(1), \ldots, \gamma(p))' .$$

Yule-Walker estimators

$\hat{\gamma}(h)$ = estimator of $\gamma(h)$.

$\gamma(0) = Var(X_t)$ $\hat{\gamma}(0)$ sample variance

$\gamma(1) = Cov(X_t, X_{t-1})$.

Let us denote by $\widehat{\mathbf{\Gamma}}, \widehat{\gamma}$ the sample analogue of $\mathbf{\Gamma}$ and γ. The method-of-moments estimators

$\hat{\gamma}(0) = \frac{1}{n-1}\sum_{t=1}^{n}(X_t - \bar{X})(X - \bar{X})$.

$$\widehat{\phi} := \widehat{\mathbf{\Gamma}}^{-1}\widehat{\gamma}\,, \quad \widehat{\sigma}^2 = \widehat{\gamma}(0) - \widehat{\gamma}'\widehat{\mathbf{\Gamma}}_n^{-1}\widehat{\gamma}$$

$\hat{\gamma}(h) = \frac{1}{n-h}\sum_{t=h+1}^{n}(X_t - \bar{X})(X_{t-h} - \bar{X})$

are called Yule-Walker estimators.

In practice, using the Durbin-Levinson Algorithm with the sample ACF $\widehat{\rho}$, we obtain $\widehat{\phi} = (\widehat{\phi}_{n,1}, \ldots, \widehat{\phi}_{n,n})'$ from

$$\widehat{\phi}_{n,n} = \left(1 - \sum_{k=1}^{n-1}\widehat{\phi}_{n-1,k}\widehat{\rho}(k)\right)^{-1}\left(\widehat{\rho}(n) - \sum_{k=1}^{n-1}\widehat{\phi}_{n-1,k}\widehat{\rho}(n-k)\right),$$

$$\widehat{\phi}_{0,0} = 0\,, \quad \widehat{P}_1^0 = \widehat{\gamma}(0)\,, \quad \widehat{P}_{n+1}^n = \widehat{P}_n^{n-1}(1 - \widehat{\phi}_{n,n}^2)\,,$$

$$\widehat{\phi}_{n,k} = \widehat{\phi}_{n-1,k} - \widehat{\phi}_{n,n}\widehat{\phi}_{n-1,n-k}\,; \quad k = 1, \ldots, n-1\,.$$

Method of Moment Estimators

taking expectations.

sample average.

Large sample distributions

When the sample size is large, the asymptotic behavior of YULE-WALKER estimators $\widehat{\phi}$ for the causal AR (p) process is normal, that is,

$\hat{\phi} \sim \mathcal{N}(\phi, \sigma^2 \Gamma^{-1})$

$$\lim_{n\to\infty} \mathbb{P}(\sqrt{n}(\widehat{\phi} - \phi) \leq \mathrm{u}) = \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_p} f(v_1, \ldots, v_p) \mathrm{d}u_1 \cdots \mathrm{d}u_p$$

for every $\mathrm{u} := (u_1, \ldots, u_p)'$, where $f(v_1, \ldots, v_p)$ is p-dimensional normal density function with mean vector 0 and variance-covariance matrix $\sigma^2\boldsymbol{\Gamma}^{-1}$.

In particular, for PACF, $\sqrt{n}\widehat{\phi}_{h,h}$ is asymptotically standard normal with mean 0 and variance 1 for $h > p$.
This result is used for hypothesis testing.

Appendix B.3 for the details.

yule - walker estimeeter

```
rec.yw = ar.yw(rec, order=2)
rec.yw$x.mean # = 62.26 (mean estimate)
rec.yw$ar # = 1.33, -.44 (coefficient estimates)
sqrt(diag(rec.yw$asy.var.coef)) # = .04, .04 (standard errors)
rec.yw$var.pred # = 94.80 (error variance estimate)
```

prediction variance

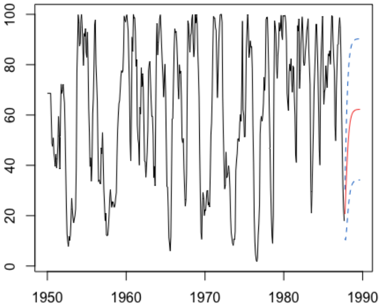
```
rec.pr = predict(rec.yw, n.ahead=24)
ts.plot(rec, rec.pr$pred, col=1:2)
lines(rec.pr$pred + rec.pr$se, col=4, lty=2)
lines(rec.pr$pred - rec.pr$se, col=4, lty=2)
```

24 steps ahead prediction

Prediction (red) and Confidence Interval (blue) of REC

$x_t = \mu + \phi_1 x_{t-1} + \phi_2 t_{t-2} + w_t.$

$x_t = \hat{\mu} + 1.33\, x_{t-1} - 0.44\, x_{t-2} + \hat{w}_t.$



Example: MA (1)

For the MA (1)

$$x_t = w_t + \theta w_{t-1}, \quad \text{or} \quad x_t = \sum_{j=1}^{\infty} (-\theta)^j x_{t-j} + w_t,$$

the method of moments estimator for θ can be computed as follows:

$$\gamma(0) = \sigma^2(1+\theta^2), \quad \gamma(1) = \sigma^2\theta, \quad \rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta}{1+\theta^2},$$

and hence, solving the equation, we obtain

sample analogue

$$\widehat{\rho}(1) = \frac{\widehat{\gamma}(1)}{\widehat{\gamma}(0)} = \frac{\widehat{\theta}}{1+\widehat{\theta}^2}.$$

Recall discussion section problem.

$\hat{\rho}(1+\hat{\theta}^2) = \hat{\theta}$. $\quad \hat{\rho}\hat{\theta}^2 - \hat{\theta} + \hat{\rho} = 0$

$\frac{\theta}{1+\theta^2} \leq \frac{1}{2}$.

In the discussion section, we verified

solve:

$$|\rho(h)| \leq \frac{1}{2}\,; \quad h \in \mathbb{Z}\,,$$

$(\theta-1)^2$.

$\theta^2+1-2\theta \geq 0$. $\theta^2+1 \geq 2\theta$.

however, the sample version $\widehat{\rho}(h)$ does not necessarily satisfy this inequality.
When $|\widehat{\rho}(1)| < 1/2$, then the invertible estimate is

$\hat{\theta} = \frac{1 \pm \sqrt{1-4(\hat{\rho}(1))^2}}{2\hat{\rho}(1)}$.

$\hat{\rho}(1) = 0.05$.

$$\widehat{\theta} = \frac{1 - \sqrt{1-4(\widehat{\rho}(1))^2}}{2\widehat{\rho}(1)}$$

and $\boxed{\text{for large } n}$, its approximated by the normal distribution with mean θ and variance

$$\frac{1+\theta^2+4\theta^4+\theta^6+\theta^8}{n(1-\theta^2)^2}\,.$$

$x_t = \sum_{j=1}^{\infty} (-\theta)^j x_{t-j}$.

See Appendix A.7.

$\theta < 1$ for the infinite series to converge.

Maximum likelihood estimator

For the ARMA (p, q) process, the unknown parameters are

$$\phi := (\phi_1, \ldots, \phi_p)', \quad \boldsymbol{\theta} := (\theta_1, \ldots, \theta_q)', \quad \mu, \sigma^2 .$$

The likelihood function $L(\phi, \boldsymbol{\theta}, \mu, \sigma^2; \mathrm{x})$ is the joint density function of the observation $\mathrm{x} := (x_1, \ldots, x_n)'$.

We estimate the unknown parameters by maximizing the likelihood function L with respect to the parameters, i.e.,

$$\max_{\phi, \boldsymbol{\theta}, \mu, \sigma^2} L(\phi, \boldsymbol{\theta}, \mu, \sigma^2; \mathrm{x}) .$$

$f(x_2|x_1) = \frac{f(x_1 x_2)}{f(x_1)}$

- The maximum likelihood estimator is known to have good properties.

- The likelihood function (joint density)

Joint Density func:

$n=2$

$f(x_1, x_2)$

$= f(x_2|x_1) f(x_1)$.

$$L(\phi, \theta, \mu, \sigma^2; \mathrm{x}) = f(\mathrm{x})$$

can be written as a product from

$$f(\mathrm{x}) = f(x_1)f(x_2|x_1)f(x_3|x_1, x_2)\cdots f(x_n|x_1, \ldots, x_{n-1})$$

$$= \prod_{t=1}^{n} f(x_t|x_1, \ldots, x_{t-1}),$$

where $f(x_t|x_1, \ldots, x_{t-1})$ is the conditional density of x_t given $x_1, \ldots x_{t-1}$ for $i = 2, \ldots, n$.

Assume that $\{w_t\}$ is the Gaussian white noise.
Each $f(x_t|x_1,\ldots,x_{t-1})$ of the conditional density in the joint density

$$f(\mathrm{x}) = f(x_1)f(x_2|x_1)f(x_3|x_1,x_2)\cdots f(x_n|x_1,\ldots,x_{n-1})$$

is normal density function with mean

$$\mathbb{E}[x_t|x_1,\ldots,x_{t-1}] = x_t^{t-1} \text{ (one-step ahead prediction)}$$

and variance

$$\mathrm{Var}(x_t|x_{t-1},\ldots,x_{t-1}) = P_t^{t-1} = \sigma^2 r_t \text{ (prediction error)},$$

where

$$r_t := (\sum_{j=0}^{\infty}\psi_j^2)\prod_{j=1}^{t-1}(1-\phi_{j,j}^2).$$

Thus, the MLE is obtained by maximizing the log likelihood

$$\text{max.}\ \log f(\mathrm{x}) = \log f(x_1) + \sum_{t=2}^{n} \log f(x_t|x_1, \ldots x_{t-1})$$

$$= -\frac{n}{2}(\log(2\pi) + \log \sigma^2) - \frac{1}{2}\sum_{t=1}^{n} \log r_t - \frac{1}{2\sigma^2} S(\beta),$$

where $\beta := (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ and

$$S(\beta) := \underbrace{\sum_{t=1}^{n} \frac{(x_t - x_t^{t-1})^2}{r_t}}_{\text{min}}.$$

Example: the causal AR (1)

$\mathbb{E}[x_t | x_{t-1}]$
$= \mathbb{E}[\mu + \phi(x_{t-1} - \mu) + w_t | x_{t-1}]$

$x_t = \mu + \phi(x_{t-1} - \mu) + w_t$, $t = 2, \ldots, n$ with Gaussian white noise $\{w_t\} \sim$ iid $\mathrm{N}(0, \sigma^2)$ and $x_1 \sim N(\mu, \sigma^2/(1-\phi^2))$.
In this case,

$$f(x_t | x_{t-1}, \ldots, x_1) = f(x_t | x_{t-1})$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{1}{2\sigma^2} (x_t - \mu - \phi(x_{t-1} - \mu))^2 \Big)$$

The likelihood function L is given by

$$L = (2\pi\sigma^2)^{-n/2} (1 - \phi^2)^{-1/2} \exp\Big(- \frac{S}{2\sigma^2} \Big),$$

where S is the unconditional sum of squares

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

max log likelihood

Example: MLE for REC series

```
rec.mle = ar.mle(rec, order=2)
rec.mle$x.mean # 62.26
rec.mle$ar # 1.35, -.46
sqrt(diag(rec.mle$asy.var.coef)) # .04, .04
rec.mle$var.pred # 89.34
```

```
> rec.mle = ar.mle(rec, order=2)
> rec.mle$x.mean # 62.26
[1] 62.26153
> rec.mle$ar # 1.35, -.46
[1]  1.3512809 -0.4612736
> sqrt(diag(rec.mle$asy.var.coef)) # .04, .04
[1] 0.04099159 0.04099159
> rec.mle$var.pred # 89.34
[1] 89.33597
```

More topics in estimation

- Conditional Least Square

- Asymptotic optimality - MLE, conditional least square, unconditional least square lead to optimal estimator

See reference books, e.g., Blockwell & Davis.

https://link.springer.com/book/10.1007/978-1-4419-0320-4

- Newton-Raphson methods to maximize the likelihood

- Gauss-Newton methods

- Overfitting problem – model selection

- Bootstrapping