Module 2.3
Estimation of Second Order Properties

S. Lakshmivarahan

School of Computer Science
University of Oklahoma
Norman, OK, 73071
USA

Statement of the problem

- Let $\{x_t\}$ be a real valued second-order stationary time series.
- Assume that we are given a finite segment $x(1:n) = \{x_1, x_2, \cdots, x_n\}$ of observations from the time series.
- Goal is to estimate the unknown second-order statistics-mean, variance and auto covariance function.
- Also need properties of the estimates: unbiasedness, consistency, asymptotic distribution, standard error, etc. so that we can test various hypotheses about the estimates.

Role of Ergodicity

- We are given a finite sample of observation but are asked to estimate the unknown parameters-mean, variance, autocorrelation of the ensemble.
- Fundamental to accomplishing this task is the ergodicity assumption in Module 3.
- From Module 1.3: A second-order stationary process is ergodic in the first two moments if:

$$\sum_{k=0}^{\infty} |\gamma(k)| < \infty \tag{1}$$

- Henceforth, we assume that this condition of absolute summability of $\gamma(k)$ holds.

Estimation of the mean

- Since x_t is stationary, a natural estimate of the unknown μ is the sample average given by:

$$\overline{x_N} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2}$$

- It can be verified that $\overline{x_N}$ is an unbiased, consistent estimate of μ.

Asymptotic properties of $\overline{x}_n$

- Let $e_N = \overline{x}_n - \mu$ denote the error in the estimate.
- Let $\gamma(k)$ denote the auto covariance function of $\{x_t\}$ satisfying condition (1).
- Condition (1) implies, $\gamma(k) \to 0 \;\text{ as }\; k \to \infty$
- CLAIM 1: Under the condition (1), the following properties of the mean squared errors hold:

$$\text{a) } \mathbf{var}(\overline{x}_n) = \mathbf{E}(e_N^2) \to 0 \tag{3}$$

$$\text{b) } N\,\mathbf{var}(\overline{x}_n) = N\,\mathbf{E}(e_N^2) \to V, \tag{4}$$

$$\text{where } \; V = \sum_{k=-\infty}^{\infty} \gamma(k) \tag{5}$$

- Refer to the books by Brockwell and Davis (2013) and Fuller (2009) for detailed proof.

Interpretation: Claim 1(a)

- Claim 1(a) implies that the mean squared error in the estimate $\overline{x}_n$ vanishes asymptotically.
- Since $\overline{x}_n$ is an unbiased estimate, mean squared error is also the variance of $\overline{x_n}$.
- By Chebyshev inequality, (2) implies that $\overline{x}_n$ converges in probability to μ.
- That is, $\overline{x}_n$ is a consistent estimate.

- Claim 1(b) implies,in addition,that the scaled variance given by $NE(\overline{x_n} - \mu)^2$ tends to a finite number which is the total sum of the correlations.
- Thus, for large N, the mean square error tends to zero at the rate $\mathbb{O}(\frac{1}{N})$
- That is, the standard error in $\overline{x}_n$ is $\mathbb{O}(\frac{1}{\sqrt{N}})$ for large N.

A Basic result

- A fundamental result due to Wold is that any stationary tie series x_t admits an MA(∞) representation:

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \varepsilon(t-j), \tag{6}$$

where $\{\varepsilon_t\}$ is the white noise sequence with mean zero and variance σ^2, and

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \tag{7}$$

A refinement of claim 1 (b)

- Claim 2: Under the conditions (4) and (5),

$$\bar{x}_n \to \text{AN}\left(\mu, \frac{V}{N}\right), \tag{8}$$

where

$$V = \sum_{j=-\infty}^{\infty} \gamma_j = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \Psi_j \right)^2 \tag{9}$$

- AN refers to asymptotic normality.
- Refer to Davis and Rockwell (2013) for a proof of this claim.

Interpretation of Claim 2

- If x_t is generated by an ARMA(p, q) process, then it readily admits an $MA(\infty)$ representation as in (4) when the condition (1) holds.
- Asymptotic normality result provide guidance to deriving confidence intervals and enables to test a variety of hypotheses about $\overline{x}_n$.

- A natural estimate $\hat{\gamma}(k)$ for $\gamma(k)$ is:

$$\hat{\gamma}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \overline{x}_N)(x_{t+k} - \overline{x}_N), \quad \text{where} \quad 0 \leq k \leq N \tag{10}$$

- Then $\rho(k)$ is estimated using

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} \tag{11}$$

Properties of $\hat{\gamma}(k)$

- The estimate $\hat{\gamma}(k)$ in (8) is biased but the bias tends to zero and has mean γ_k as $N \to \infty$.
- An unbiased estimate for $\gamma(k)$ is given by:

$$\hat{\gamma}(k) = \frac{1}{N-k} \sum_{t=1}^{N-k} (x_t - \overline{x}_N)(x_{t+k} - \overline{x}_N), \quad \text{where } 0 \leq k \leq N \tag{12}$$

Properties of $\hat{\gamma}(k)$

- A desirable property of $\gamma(k)$ is that the matrix:

$$\hat{\Gamma}_N = [\Gamma_N(|i-j|)] = [\hat{\gamma}(|i-j|)] \in \mathbb{R}^{N\times N} \tag{13}$$

  is non-negative definite.

- To verify this, recall that $\hat{\Gamma}_N$ admits a decomposition by definition as:

$$\hat{\Gamma}_N = \frac{1}{N} DD^T, \tag{14}$$

  where $D \in \mathbb{R}^{N\times 2N}$ is built out of the given data.

Structure of D in (12)

- Assume for example N $= 4$. Then using $\tilde{x}_i = x_i - \bar{x}_N$, D is given by:

$$D = \begin{bmatrix} 0 & 0 & 0 & 0 & \tilde{x}_1 & \tilde{x}_2 & \tilde{x}_3 & \tilde{x}_4 \\ 0 & 0 & 0 & \tilde{x}_1 & \tilde{x}_2 & \tilde{x}_3 & \tilde{x}_4 & 0 \\ 0 & 0 & \tilde{x}_1 & \tilde{x}_2 & \tilde{x}_3 & \tilde{x}_4 & 0 & 0 \\ 0 & \tilde{x}_1 & \tilde{x}_2 & \tilde{x}_3 & \tilde{x}_4 & 0 & 0 & 0 \end{bmatrix} \qquad (15)$$

- Substituting D in (12) and simplifying it can be verified that elements of $\hat{\Gamma}_4$ agree with the definition in (8).
- Hence, the expression for $\hat{\Gamma}_N$ in (12) holds true for any N.

$\hat{\Gamma}_N$ is positive semi-definite

- Let $y \in (y_1, y_2, \cdots, y_N)^T \in \mathbb{R}^N$ be a real vector.
- Then

$$y^T \hat{\Gamma}_N y = \frac{1}{N}[y^T D D^T y] = \frac{1}{N}[D^T y]^T [D^T y] = \frac{1}{N}||D^T y||^2$$

- Hence, $\hat{\Gamma}_N \geq 0$ is positive semi-definite and $\hat{\gamma}_k$ is the auto covariance function.

A guide in computing $\hat{\gamma}_k$

- It can be shown that if we define the elements of $\hat{\Gamma}_N$ using $\hat{\gamma}_4(k)$ in (10), $\hat{\Gamma}_N$ <u>may not</u> be positive definite.
- The number of terms $(N-k)$ on the right hand sum in (8) decreases as k increases.To maintain reasonable accuracy it is suggested that $N \geq 50$ and $k \leq \frac{N}{4}$.

- Define

$$\hat{\rho}(1:k) = (\hat{\rho}(1), \hat{\rho}(2), \cdots, \hat{\rho}(k))^T \in \mathbb{R}^k,$$
$$\rho(1:k) = (\rho(1), \rho(2), \cdots, \rho(k))^T \in \mathbb{R}^k.$$

- Claim 3: Let x_t be a stationary process represented by

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \qquad \varepsilon_t \sim \textbf{iid}(0, \sigma^2) \tag{16}$$

with

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \quad \text{and} \quad \sum_{j=-\infty}^{\infty} {\psi_j}^2 |j| < \infty \tag{17}$$

- Then for each k,

$$\hat{\rho}(k) \ \text{ is AN}(\rho(1:k), N^{-1}\Sigma), \tag{18}$$

where $\Sigma \in \mathbb{R}^{k\times k}$ covariance matrix given below

Elements of ε - Bartlelt's formula

- It can be shown that

$$\Sigma_{ij} = \sum_{k=1}^{\infty} [\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)] \\ [\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)] \quad (19)$$

- Claim 3 is proved in Fuller (2009) and Brockwell and Davis (2013)

Import of Claim 3

- It readily provides an estimate of the standard errors which in turn helps us to identify when $\gamma(k)$ is zero.
- Claim 3 is applicable to all processes modeled by ARMA(p, q) models.

Example 1: Pure White Noise

- Let $x_t = \varepsilon_t$ be the white noise sequence.
- Clearly
$$\rho(0) = 1 \text{ and } \rho(k) = 0 \text{ for } k > 0. \tag{20}$$
- Substituting (20) in (19) and simplifying:

$$\begin{aligned}\Sigma_{ij} &= \sum_{k=1}^{\infty}[\rho(k+i)+\rho(k-i)][\rho(k+j)+\rho(k-j)] \\ &= \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}\end{aligned} \tag{21}$$

Example 1

- By Claim 3,the elements of the vector $\hat{\rho}(1:k)$ are uncorrelated and independent for large N
- Hence,

$$\mathbf{var}(\hat{\rho}(i)) = \frac{1}{N} \text{ for } \quad 1 \leq i \leq N \tag{22}$$

- Consequently, 95 % of the time, the estimate $\hat{\rho}(i)$ must lie in the confidence interval $\pm\frac{1.96}{\sqrt{N}}$.
- This property is used to confirm that the given sequence is indeed a white noise sequence.

Example 2: MA(1) model

- Consider MA(1):

$$x_t = \varepsilon_t + \theta\varepsilon_{t-1}, \text{ with } \theta < 1 \tag{23}$$

- It is well known that (refer to Module 4.3)

$$\begin{cases} \rho(0) = 1 \\ \rho(1) = \left(\frac{\theta}{1+\theta^2}\right) \neq 0 \\ \rho(i) = 0 \quad \text{for } i > 1 \end{cases} \tag{24}$$

- From (19):

$$\Sigma_{ii} = \sum_{k=1}^{\infty} [\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)]^2 \tag{25}$$

Example 2

- Combining (24) and (25): for $i > 1, \rho(i) = 0$ and

$$\Sigma_{ii} = \sum_{k=1}^{\infty} [\rho(k+i) + \rho(k-i)]^2 \tag{26}$$

- Set $i = 2$.Then

$$\begin{aligned}
\Sigma_{22} &= \sum_{k=1}^{\infty} [\rho(k+2) + \rho(k-2)]^2 \\
&= \sum_{k=1}^{\infty} \rho^2(k-2) \qquad (27) \\
&= \rho^2(-1) + \rho^2(0) + \rho^2(1) \\
&= 1 + 2\rho^2(1)
\end{aligned}$$

Example 2

- Setting $i = 3$

$$\begin{aligned}
\Sigma_{33} &= \sum_{k=1}^{\infty} [\rho(k+3) + \rho(k-3)]^2 \\
&= \sum_{k=1}^{\infty} \rho^2(k-3) \\
&= \rho^2(-1) + \rho^2(0) + \rho^2(1) \\
&= 1 + 2\rho^2(1)
\end{aligned}$$

- Verify

$$\Sigma_{ii} = 1 + 2\rho^2(1) \quad \text{for all } i > 1 \tag{28}$$

Example 2

- Let $\theta = \frac{1}{2}, \rho(1) = \frac{\theta}{1+\theta^2} = 0.4$
- $\sum_{ii} = 1 + 2 \times 0.4^2 = 1.32$
- Hence, 95% of the time, the estimate $\hat{\rho}(i)$ for i $>$ 1 will lie in the error band.
  $\pm 1.96 \frac{1+2\rho^2(1)}{\sqrt{N}} = \pm \frac{1.96x1.32}{\sqrt{N}} = \pm \frac{2.59}{\sqrt{N}}$.

Example 3: MA(q) Model

- Consider

$$x_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{29}$$

- From Module 4.3:

$$\left.\begin{aligned} &\rho(0) = 1 \\ &\rho(i) \neq 0 \quad \text{for} \quad 1 \leq i \leq q \\ &\rho(i) = 0 \quad \text{for} \quad i > q \end{aligned}\right\} \tag{30}$$

- It can be verified that, for $i > q$

$$\Sigma_{ii} = [1 + 2\rho^2(1) + 2\rho^2(2) + \cdots + 2\rho^2(q)] \tag{31}$$

- From the known values of $\rho(i)$ for $1 \leq i \leq q$, we can likewise compute the 95% error band for $\hat{\rho}(i)$ for$i > q$

Example 4: AR(1) model

- $$x_t = c + \phi x_{t-1} + \varepsilon_t \tag{32}$$
- It can be shown Module 4.4 that
$$\rho(i) = \phi^{|i|} for - \infty < i < \infty. \tag{33}$$
- Hence, from (33) and (19):
$$\begin{aligned}\Sigma_{ii} &= \sum_{k=1}^{\infty}[\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)]^2 \\ &= \sum_{k=1}^{\infty}[\phi^{|k-i|} - \phi^{i+k}]^2\end{aligned} \tag{34}$$

- Splitting the sum in (34):

$$\Sigma_{ii} = \underbrace{\sum_{k=1}^{i} \phi^{2i}(\phi^{-k} - \phi^{k})^2}_{I} + \underbrace{\sum_{k=i+1}^{\infty} \phi^{2k}(\phi^{-i} - \phi^{i})^2}_{II}. \qquad (35)$$

Example 4

- $$\begin{aligned}\text{Term I} &= \phi^{2i}\sum_{k=1}^{i}\frac{1-2\phi^{2k}+\phi^{4k}}{\phi^{2k}} \\ &= \phi^{2i}\left[\sum_{k=1}^{i}\frac{1}{\phi^{2k}} - 2i + \sum_{k=1}^{i}\phi^{2k}\right]\end{aligned} \tag{36}$$

- But

$$\phi^{2i}\sum_{k=1}^{i}\frac{1}{\phi^{2k}} = \frac{1-\phi^{2i}}{1-\phi^{2}} \tag{37}$$

$$\phi^{2i}\sum_{k=1}^{i}\phi^{2k} = \phi^{2i+2}\frac{1-\phi^{2i}}{1-\phi^{2}} \tag{38}$$

- Substituting (37), (38) in (36):

$$\text{Term I} = \frac{1-\phi^{2i}}{1-\phi^2} - 2i\phi^{2i} + \phi^{2i+2}\frac{1-\phi^{2i}}{1-\phi^2} \qquad (39)$$

Example 4

- $$\text{Term II} = (\phi^{-i} - \phi^{i})^2 \sum_{k=i+1}^{\infty} \phi^{2k} \tag{40}$$

- $$But \sum_{k=i+1}^{\infty} \phi^{2k} = \frac{\phi^{2i+2}}{1-\phi^2} \tag{41}$$

- Substituting (41) in (40) and simplifying

$$\text{Term II} = \frac{\phi^2}{1-\phi^2} - 2\frac{\phi^{2i+2}}{1-\phi^2} + \frac{\phi^{4i+2}}{1-\phi^2} \tag{42}$$

Example 4

- Adding (39) and (42) and simplifying,from (35):

$$\Sigma_{ii} = \frac{1+\phi^2}{1-\phi^2}(1-\phi^{2i}) - 2i\phi^{2i} \tag{43}$$

- Since $\mid \phi \mid < 1$, for large i

$$\Sigma_{ii} \approx \frac{1+\phi^2}{1-\phi^2} \tag{44}$$

Example 4

- For $\phi = \frac{1}{2}, \Sigma_{ii} = \frac{1.25}{0.75} = 1.667$
- Hence, for large i,the error band for $\hat{\rho}(i)$ is $\pm\frac{1.667}{\sqrt{N}}$
- However, Partial Auto Correlation(PACF) for AR(p) models are better indicators since PACF(k) $= 0$ for $k > p$ for AR(p) model.
- This is similar to ACF $(k) = 0$ for $k > q$ for MA(q) models.
- In practice,we need to combine the information from ACF and PACF to decide on the order of the model.