System Identification

6.435

SET 9

– Asymptotic distribution of PEM

Munther A. Dahleh

Central Limit Theorem (Generalization)

- <u>Basic Theorem II</u>:

Consider $$X_N = \frac{1}{N}\sum_{t=1}^{N} \Psi(t,\theta)v_o(t)$$

$$EX_N = 0$$

$\Psi(t,\theta), v_o$ are both ARMA processes, possibly correlated, with underlying white noise (bounded 4^{th} moment)

- Then:

1) $$\sqrt{N} X_N \xrightarrow{\text{distribution}} N(0, P)$$

2) $$P = \lim_{N \to \infty} E \quad N X_N X_N^T$$

- Proof:

If $\Psi(t, \theta) v_o(t)$ where independent for different t, then result follows from central limit theorem. It can be shown that the dependence decays for large N.

Application to Prediction Error Methods (Special Case)

- ARX case:

$$D_T(\delta, m) = \{\theta_o\} \neq \Phi$$ (Both identifiable & $\delta \in m$)

Data informative.

$$\widehat{\theta}_N = \left[\frac{1}{N}\sum_{t=1}^{N} \Phi\Phi^T(t)\right]^{-1} \left[\frac{1}{N}\sum_{t=1}^{N} \Phi(t)y(t)\right]$$

$$= \theta_o + \underbrace{\left[\frac{1}{N}\sum_{t=1}^{N} \Phi\Phi^T(t)\right]^{-1}}_{\left(\bar{E}\Phi\Phi^T(t)\right)^{-1}} \underbrace{\left[\frac{1}{\sqrt{N}}\sum_{t=1}^{N} \Phi(t)e_o\right]}_{N(0,P)} \frac{1}{\sqrt{N}}$$

$$P = \lim_{N\to\infty} \frac{1}{N} E \sum_{\substack{t=1 \\ s=1}}^{N} \Phi(t) e_o(t) \Phi^T(s) e_o(s) = \lim_{N\to\infty} \frac{\lambda_o}{N} E \sum_{t=1}^{N} \Phi(t)\Phi^T(t)$$

$$= \lambda_o \bar{E}\left(\Phi(t)\Phi^T(t)\right)$$

$$\sqrt{N}\left(\hat{\theta}_N - \theta_o\right) \sim \text{Asym } N(0, P_\theta)$$

$$P_\theta = \bar{E}\left(\Phi(t)\Phi^T(t)\right)^{-1} \left[\lambda_o \bar{E}\left(\Phi(t)\Phi^T(t)\right)\right] \bar{E}\left(\Phi(t)\Phi^T(t)\right)^{-1}$$

$$= \lambda_o \left[\bar{E}\left(\Phi(t)\Phi^T(t)\right)\right]^{-1}$$

General Case

$$\hat{\theta}_N = \underset{\hat{\theta} \in D_m}{\text{argmin}}\; V_N\left(\theta, Z^N\right) \qquad \begin{cases} \{\theta_o\} = D^T(\delta, m) \\ \text{Data is informative} \end{cases}$$

It follows that $V_N{}'\left(\hat{\theta}_N, Z^N\right) = 0$

Expand $V_N{}'\left(\theta, Z^N\right)$ around θ_o and evaluate at $\hat{\theta}_N$

$$0 = V_N{}'\left(\hat{\theta}_N, Z^N\right) = V_N{}'\left(\theta_o, Z^N\right) + V_N{}''\left(\xi, Z^N\right)\left(\hat{\theta}_N - \theta_o\right)$$

ξ is a vector “between” $\hat{\theta}_N$ & θ_o

As $N \to \infty$: $\begin{cases} \widehat{\theta}_N \to \theta_o \\ {V_N}''\left(\xi, Z^N\right) \to \bar{V}''(\theta_o) \end{cases}$

Assume that $\bar{V}''(\theta_o)$ is nonsingular, then

$$\widehat{\theta}_N - \theta_o \cong -\left[\bar{V}''(\theta_o)\right]^{-1} {V_N}'\left(\theta_o, Z^N\right)$$

$${V_N}'\left(\theta_o, Z^N\right) = \frac{d}{d\theta}\frac{1}{2N}\sum_{t=1}^{N} \varepsilon^2(t,\theta)\bigg|_{\theta=\theta_o}$$

$$= -\frac{1}{N}\sum_{t=1}^{N} \Psi(t,\theta_o)e_o(t)$$

From Basic Theorem II

$$= \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \Psi(t,\theta_o) e_o(t) \xrightarrow{\text{asym}} N(0,P)$$

$$P = \lim_{N\to\infty} EN \frac{1}{N^2} \sum_{t=1}^{N} \Psi(t,\theta_o) e_o(t) \sum_{s=1}^{N} \Psi^T(s,\theta_o) e_o(s)$$

$$= \frac{1}{N} \lim_{N\to\infty} \lambda_o \sum_{t=1}^{N} E\Psi(t,\theta_o)\Psi^T(t,\theta_o)$$

$$= \lambda_o \bar{E} \Psi(t,\theta_o)\Psi^T(t,\theta_o)$$

From this, it follows that

$$\sqrt{N}\left(\hat{\theta}_N - \theta_o\right) \simeq -\sqrt{N}\left[\bar{V}''(\theta_o)\right]^{-1} V_N{}'\left(\theta_o, Z^N\right)$$

Notice that

$$\bar{V}''(\theta_o) = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{t=1}^{N}\Psi(t,\theta_o)\Psi^T(t,\theta_o) - \underbrace{\frac{1}{N}\sum_{t=1}^{N}\frac{\partial\Psi(t,\theta_o)}{\partial\theta}e_o}_{\longrightarrow\, 0}\right)$$

$$= \bar{E}\Psi(t,\theta_o)\Psi^T(t,\theta_o)$$

$\Rightarrow$

$$\sqrt{N}\left(\hat{\theta}_N - \theta_o\right) \longrightarrow N(0, P_\theta) \qquad P_\theta = \lambda_o\left(\bar{E}\Psi(t,\theta_o)\Psi^T(t,\theta_o)\right)$$

Efficiency

<u>Recall</u>: Cramer-Rao Bound for normally distributed noise:

$$\text{Cov}\left(\sqrt{N}\hat{\theta}_N\right) \geq \lambda_o \left(\bar{E}\Psi(t,\theta_o)\Psi^T(t,\theta_o)\right)$$

$\Rightarrow$ Predication error estimates are asymptotically efficient if $e_o(t)$ is normally distributed.

Estimates for accuracy

$$\hat{P}_N = \hat{\lambda}_N \left[\frac{1}{N}\sum_{t=1}^{N} \Psi\left(t,\hat{\theta}_N\right)\Psi^T\left(t,\hat{\theta}_N\right)\right]^{-1}$$

$$\hat{\lambda}_N = \frac{1}{N}\sum_{t=1}^{N} \varepsilon^2\left(t,\hat{\theta}_N\right)$$

Examples

ARX:

$$y(t) + a_o y(t-1) = u(t-1) + e_o(t) \qquad \begin{cases} u \sim WN \text{ with var} = \mu \\ e_o \sim WN \text{ with var} = \lambda_o \end{cases}$$

$$\hat{y} = -ay(t-1) + u(t-1)$$

$$\Psi(t,\theta) = \frac{d}{d\theta}\hat{y} = -y(t-1)$$

$$\bar{E}\left(\Psi(t,\theta_o)\Psi^T(t,\theta_o)\right) = \bar{E}y^2(t-1) = \frac{\mu+\lambda_o}{1-a_o^2}$$

$$P_\theta = \lambda_o \frac{1-a^2}{\mu+\lambda_o}$$

$$\text{Cov } \hat{a}_N \sim \frac{1}{N}\lambda_o \frac{1-a^2}{\mu+\lambda_o}$$

MA: $$y(t) = \left(1 + c_o q^{-1}\right) e_o(t)$$

$$\hat{y} = \left(1 - \frac{1}{cq^{-1} + 1}\right) e_o$$

or $\hat{y}(t) + c\hat{y}(t-1) = cy(t-1)$

$$\Psi(t,\theta) = \frac{d}{d\theta}\hat{y}(t)$$

$$\Rightarrow \quad \Psi(t,\theta) + c\Psi(t-1,\theta) = \overbrace{y(t-1) - \hat{y}(t-1)}^{= e_o(t-1) \text{ at } c = c_o}$$

at c_o: $$\Psi(t,\theta) = \frac{1}{1 + c_o q^{-1}} e_o(t-1)$$

$$E\left(\Psi(t,\theta_o)\Psi^T(t,\theta_o)\right) = \frac{1}{1 - c_o^2}\lambda_o$$

$$\text{Cov } \hat{c}_N \sim \frac{1}{N}\lambda_o \frac{1 - c_o^2}{\lambda_o} = \frac{1 - c_o^2}{N}$$

<u>ARMA</u>:

$$y(t) + ay(t-1) = e(t) + ce(t-1)$$

$$\hat{y}(t,\theta) = \left(1 - \frac{1+aq^{-1}}{1+cq^{-1}}\right) y(t)$$

$$\Psi(t,\theta) = -\frac{d}{d\theta}\varepsilon(t,\theta) = -\frac{d}{d\theta}\left(\frac{1+aq^{-1}}{1+cq^{-1}}\right) y(t)$$

$$\Psi_1(t,\theta) = -\frac{d}{da}\varepsilon(t,\theta) = \frac{-q^{-1}}{1+cq^{-1}} y(t) \triangleq y_F(t,\theta)$$

$$\Psi_2(t,\theta) = -\frac{d}{dc}\varepsilon(t,\theta) = +\frac{q^{-1}\left(1+aq^{-1}\right)}{\left(1+cq^{-1}\right)^2} y(t)$$

$$= \frac{q^{-1}}{1+cq^{-1}}\varepsilon(t,\theta) = \varepsilon_F(t,\theta)$$

$$\dot{E}\left(\Psi\Psi^T\right) = \left.\begin{pmatrix} E{y_F}^2 & -Ey_F\varepsilon_F \\ -Ey_F\varepsilon_F & E{\varepsilon_F}^2 \end{pmatrix}\right|_{\theta=\theta_o}$$

$$P_\theta = \lambda_o \begin{bmatrix} \frac{\lambda_o}{1-a_o^2} & \frac{-\lambda_o}{1-a_oc_o} \\ \frac{-\lambda_o}{1-a_oc_o} & \frac{\lambda_o}{1-c_o^2} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{1}{1-a_o^2} & \frac{-1}{1-a_oc_o} \\ \frac{-1}{1-a_oc_o} & \frac{1}{1-c_o^2} \end{bmatrix}$$

<u>Comments</u>: As $a \rightarrow c$, then $P_\theta \rightarrow \infty$. If $a \rightarrow c$, then the model structure is over parametrized so $D_T \supset \{\theta_o\}$ ($\frac{C}{A}$ has pole/zero cancellation) $\widehat{\theta}_N \rightarrow$ set, not just a point.