Maximum Likelihood Estimation.

Predictor Model

$$Y_K = f_K\left(\theta; Y^{K-1}, u^{K-1}\right) + E_K, \quad K = 0, 1, \cdots$$

$$\hat{Y}_K(\theta) = E\left(Y_K \middle| Y^{K-1}, u^{K-1}\right)$$

$$:= f_K\left(\theta; Y^{K-1}, u^{K-1}\right)$$

$$\varepsilon_K(\theta) = Y_K - \hat{Y}_K(\theta)$$

$\theta \in \mathbb{R}^q$

$Y_K \in \mathbb{R}^r$

Identification Criterion : defined in terms of

$$V_N\left(\theta; Y^N, u^{N-1}\right) = \frac{1}{N}\sum_{K=1}^{N} \varepsilon_K^T(\theta) W_K \varepsilon_K(\theta)$$

$$W_K > 0$$

$$V_N\left(\theta; Y^N, u^{N-1}\right) = \det\left[\frac{1}{N}\sum_{K=1}^{N} \varepsilon_K^T(\theta)\varepsilon_K(\theta)\right]$$

E_K's indep. & indep. of u_K

$Cov(E_K) = \Lambda_K$.

$$\log p\left(Y_N, \ldots, Y_1 \middle| U^{N-1}; \theta\right).$$

$$= -\frac{Nr}{2}\log 2\pi - \frac{1}{2}\sum_{K=1}^{N}\log\det \Lambda_K$$

$$-\frac{1}{2}\sum_{K=1}^{N}\left\| Y_K - \hat{Y}_K(\theta)\right\|^2_{\Lambda_K^{-1}(\theta)}$$

Max. Likelihood $\Longrightarrow$ Minimize

$$L(\theta) = \sum_{K=1}^{N}\left\|\varepsilon_K(\theta)\right\|^2_{\Lambda_K^{-1}(\theta)} + \sum_{K=1}^{N}\log\det\Lambda_K(\theta)$$

<u>Case I</u> Λ_K : known $\Lambda_K > 0$

Cost:

$$V(\theta) = \frac{1}{N}\sum_{K=1}^{N}\left\|\varepsilon_K(\theta)\right\|^2_{\Lambda_K^{-1}}.$$

<u>Case II</u> : $\Lambda_K = \Sigma$, Σ unknown

ψ : vector

$$\psi = (\theta, \Sigma), \quad \theta \in \tilde{D}, \quad \Sigma > 0$$

$$V(\theta) = \frac{1}{N}\sum_{K=1}^{N}\left\|\varepsilon_K(\psi)\right\|^2_{\Sigma^{-1}} + \log\det\Sigma$$

General form for criterion.

$$\psi \longmapsto h\left(\frac{1}{N}\sum_{K=1}^{N} l_K(\psi; \varepsilon_K(\psi))\right)$$

$l_K : \mathbb{R}^q \times \mathbb{R}^n \longrightarrow$ space of matrices $(d \times d)$ $\mathcal{L}(\mathbb{R}^d; \mathbb{R}^d)$.

$h : \mathcal{L}(\mathbb{R}^d; \mathbb{R}^d) \longrightarrow \mathbb{R}$.

For Case I

$$l_K = \varepsilon_K \varepsilon_K^T \Lambda_K^{-1}$$

$$h(\cdot) = \text{Trace}\ (\cdot).$$

Case II

$$h(\cdot) = \det(\cdot).$$

$$l_K = \varepsilon \varepsilon^T.$$

Asymptotic Distribution of Parameter Estimates : $\Lambda_K = \Lambda$ > 0 known

Assume:

$$\overline{V}(\theta) = \lim_{N\to\infty} h\left(\frac{1}{N}\sum_{K=1}^{N} E\, l_K(\theta; \varepsilon_K(\theta))\right).$$

exists for each θ and for N sufficiently large, estimates $\hat{\theta}_N$ are confined to closed ball B s.t.

$$\frac{\partial^2}{\partial \theta^2} \overline{V}(\theta) > \delta . I \quad \forall \theta \in B . \quad \text{(Convexity)}$$

Then $\hat{\theta}_N \longrightarrow \theta^*$ where θ^* minimizes $\overline{V}(\theta)$ over B:

<u>Interpretation:</u>

θ^* = parameter value associated with a model which best approximates the system measured by average value of the I.D. criterion in the limit as $N \longrightarrow \infty$

Suppose θ^* provides a true description of the system in the sense:

$\left(\varepsilon_K(\theta^*)\right)$ = sequence of zero mean, indep. random variables with common covariance Σ_0

Proposition: (Max Likelihood case)

Let

$$\left(g_K(\theta)\right)_{i,j} = \frac{\partial}{\partial \theta_j}\left(\hat{Y}_K(\theta)\right)$$

and assume $g_K(\theta^*)$ is a stationary process.

Then, let $P = \left[E\, g_K^T(\theta^*)\Sigma_0^{-1} g_K(\theta^*)\right]^{-1}$.

Then $N^{1/2}\left(\hat{\theta}_N - \theta\right)$ converges to $N(0; P)$

(normal distribution with covariance P).

Use for estimating Confidence Region.

Characterization of Set into which θ_N converges.

$$\theta_N \text{ minimizer of } V_N\left(\theta; y^{N}, u^{N-1}\right)$$

If $\lim_{N\to\infty} E\, V_n\left(\theta; Y^N, u^{N-1}\right)$ exists

$\forall \theta \in D$, we expect that θ_N converges to

$$\left\{\theta \,\middle|\, \lim_{N\to\infty} EV_n\left(\theta; Y^N, u^{N-1}\right) = \min_{\psi\in D} \lim_{N\to\infty} EV_N(\psi; \cdots\right.$$

$\Rightarrow$ Identification criterion gives a parameter value which minimizes the expected value of limit

In terms of the discussion previously, we can write $E V_N(\theta; Y^N, u^{N-1})$

$$= E h\left(Q_N(\theta; Y^N, u^{N-1})\right)$$

where $Q_N = \frac{1}{N}\sum_{K=1}^{N} l_K(\theta; Y^N, u^{N-1})$

Replace this by:

$$h(E Q_N)$$

Non-existence of limit allowed: replace limit by

$$\liminf_{N\to\infty} h(E(Q_N))$$

Th. 1 Suppose that the system which generates the data is stable, the prediction associated with the model set are uniformly stable. Then

$$\theta_N \longrightarrow D_I \quad , \text{a.s. as } N\to\infty$$

where

$$D_I = \left\{ \theta \,\Big|\, \underset{\psi \in D}{\mathrm{Max}} \lim_{N \to \infty} \inf \left[h\left(E\,Q_N(\theta)\right) - h\left(E\,Q_N(\psi)\right) \right] = 0 \right\}. \quad (*)$$

(Convergence : a.s. , $\forall\, \varepsilon > 0$, $\exists\, \overline{N}$ s.t. $\forall\, N > \overline{N}$, the set $\left\{ \theta \,\big|\, |\theta - \theta_N| \le \varepsilon \right\} \cap D_I$ is non-empty.)

<u>Asymptotic Indistinguishability</u>

Data generated by the system is asymptotically indistinguishable from one of the models of the set

$$D_T = \left\{ \theta \in D \,\Big|\, \lim_{N \to \infty} \frac{1}{N} \sum_{K=1}^{N} \|\hat{Y}_K - \hat{Y}_K(\theta)\|^2 = 0 \right\} \quad (**)$$

is non-empty. Here $\hat{Y}_K$ is the true conditional expectation.

Th: 2 For a stable system and uniformly stable predictor and assuming that $D_T \neq \phi$, suppose $\exists\ \delta > 0$ s.t

(H1) $$E\left((Y_k-\hat{Y}_k)(Y_k-\hat{Y}_k)^T\right) > \delta I \quad k=1,2\ldots$$

Then

$$\theta_N \longrightarrow \left\{\theta \;\middle|\; \liminf_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} E\left\|\hat{Y}_k-\hat{Y}_k(\theta)\right\|^2 = 0\right\}$$

a.s.

Proof requires the following Lemma:

A, B = symmetric $n \times n$ matrices. Let $A > 0$ and $B \geq 0$. Then

$$\det(A+B) \geq \det(A) + \frac{\det(A)}{n\lambda_{max}(A)}\operatorname{trace}(B)$$

Proof of Th:

Let $\nu_k = Y_k - \hat{Y}_k$

$$E\left(\varepsilon_k(\psi)\varepsilon_k^T(\psi)\right) = E\left((\hat{Y}_k-\hat{Y}_k(\psi)+\nu_k)\left(\quad\right)^T\right)$$

$$= E\left[(\hat{Y}_k-\hat{Y}_k(\psi))(\hat{Y}_k-\hat{Y}_k(\psi))^T\right] + E\left(\nu_k\nu_k^T\right). \quad (***)$$

Take $\psi \in D$. From hypothesis $\theta \in D_I$

$$0 \geq \liminf_{N\to\infty} \left\{ \det E \frac{1}{N} \sum_K \varepsilon_K(\theta)\varepsilon_K^T(\theta) - \det E \frac{1}{N} \sum_K \varepsilon_K(\psi)\varepsilon_K^T(\psi) \right\}$$

$$= \liminf_{N\to\infty} \left\{ \det\left(S_N(\theta)+P_N\right) - \det\left(S_N(\psi)+P_N\right) \right\}$$

where $S_N(\eta) = \frac{1}{N} \sum_K E\left(\hat{Y}_K - \hat{Y}_K(\eta)\right)\left(Y_K - \hat{Y}_K(\eta)\right)^T$.

$$P_N = \frac{1}{N} \sum_K E\, \nu_K \nu_K^T$$

Above inequality can be written as:

$$\liminf_{N\to\infty} \left\{ \det\left(S_N(\theta)+P_N\right) - \det(P_N) - \Delta_N(\psi) \right\} \leq 0 \quad (****)$$

where $\Delta_N(\psi) = \det\left(S_N(\psi) + P_N\right) - \det P_N$.

Now choose $\psi \in D_T$ (indistinguishable set).

$\Rightarrow S_N(\psi) \to 0$ as $N \to \infty$

Now P_N is a bounded sequence

(ν_K has uniformly bdd. second moments) and the fn. $\det(\cdot)$ is continuous

and $L \mapsto \det(L + P_N)$ is continuous at zero,

uniformly in N, $\Rightarrow \lim_{N\to\infty} \Delta_N(\psi)$ exists and is zero.

Hence $(****) \Longrightarrow$.

$$\liminf_{N\to\infty} \left\{ \det(S_N(\theta) + P_N) - \det(P_N) \right\} \leq 0$$

$$(*****)$$

Now using (H1), $\exists\ \alpha, \bar{\alpha} > 0$ s.t

$\det(P_k) < \alpha$ and $\lambda_{max}(P_k) > \bar{\alpha}\ \ \forall k.$

and hence from the Lemma: $\exists\ c > 0$

$$\det(S_N(\theta) + P_N) - \det P_N \geq c \operatorname{Trace}(S_N(\theta))$$

$$= c\left[\frac{1}{N}\sum_k \|\hat{Y}_k - \hat{Y}_k(\theta)\|^2\right] \quad N = 1, 2 \ldots$$

Hence from this inequality and $(*****)$

$$\liminf_{N\to\infty} \left\{ \frac{1}{N}\sum_k \|\hat{Y}_k - \hat{Y}_k(\theta)\|^2 \right\} \leq 0$$

which proves the result.