

## 定常過程の近似としての自己回帰モデル

統計数理研究所 尾形良彦

## §1. はじめに

Huber [4] は最尤推定量の一致性と漸近正規性をこれまでより弱い正則条件のもとで証明した。重要なのはパラメータ分布族  $\{f(x, \theta) ; \theta \in \Theta\}$  が観測を支配する真の分布を含んでいなければならぬという仮定をはずしたことにある。

ここでは Huber の結果を定常過程を観測するときにはパラメータ分布族を有限個の母数をもつマルコフ過程をあてはめてその最大尤度でもとの確率過程を（エントロピーのひみで）近似するときの議論につかえるよう拡張し、そのための仮定と証明を与える。正規自己回帰モデルによって十分良く近似できる観測時系列に対しての漸近理論 議論を試みる。エントロピー最大化原理に基づいて、最良の近似モデルを得るために情報量規準 (AIC) の正当化を試みる。

## 32. マルコフ過程モデルの最尤推定量

$(\mathcal{X}, \beta)$  は可測空間。 $\{X_t, t \in \mathbb{Z}\}$  は  $\mathcal{X}$  に値をもつ定常エルゴード過程。これが観測される時に遷移確率密度分布のパラメータ族  $\mathcal{F} = \{f_\theta(x_0 \mid x_{-1}, \dots, x_{-p}), \theta \in \Theta, x_i \in \mathcal{X}\}$  を考える。ここで  $\Theta \subset \mathbb{R}^d$  は開集合であり、任意の  $x_0, x_1, \dots, x_p \in \mathcal{X}$  と  $\theta \in \Theta$  に対して  $f_\theta(x_0 \mid x_{-1}, \dots, x_{-p}) > 0$  a.s. とする。

$$(2.1) \quad P_\theta(x_0 \mid x_{-1}, \dots, x_{-p}) = -\log f_\theta(x_0 \mid x_{-1}, \dots, x_{-p})$$

Y(いたとき 観測  $\{X_t, t \in \mathbb{Z}\}$  に対する

$$(2.2) \quad \frac{1}{n} \sum_{t=0}^n P_{T_n}(X_t \mid X_{t-1}, \dots, X_{t-p}) - \inf_{\theta \in \Theta} \sum_{t=0}^n P_\theta(X_t \mid X_{t-1}, \dots, X_{t-p}) \rightarrow 0 \quad (\text{a.s. または in prob.})$$

を満たす推定量  $\tau_n = \tau_n(x_0, x_1, \dots, x_n)$  を考える。

エルゴード定理の助けを借りれば Huber [4] が議論した方法と同じようにして次の結果をいうことができる。 $P_\theta$  の正則条件はまたく同様である。

### 定理 1

(2.3)  $L(\theta) = E[-\log P_\theta(x_0 \mid x_{-1}, \dots, x_{-p})]$ ,  $\theta \in \Theta$ , を最小化する  $\hat{\theta} \in \Theta$  が唯一つ存在するならば (2.2) を満たす任意の統計量  $\tau_n$  は  $\hat{\theta}$  に収束する。(a.s. または in prob.)

$$(2.4) \quad \psi_\theta(x_0 \mid x_{-1}, \dots, x_{-p}) = \frac{\partial}{\partial \theta} \log f_\theta(x_0 \mid x_{-1}, \dots, x_{-p})$$

とおけ。 $\tau_n$  が或るパラメータ  $\psi$  に確率収束することを仮定し

$$(2.5) \quad \frac{1}{\sqrt{n}} \sum_{t=0}^n \psi_{\tau_n}(X_t \| X_{t-1}, \dots, X_{t-p}) \rightarrow 0 \quad (\text{in prob.})$$

を満足する  $\tau_n$  の引か漸近的正規分布に従うための充分条件を見出したい。

### 仮定

(N1)  $\psi_\theta(X_0 \| X_{-1}, \dots, X_{-p})$  は  $\theta$  を固定して  $\mathcal{P}$ -可測で  $\mathcal{A}_{t-1}$  にて可分である。(Huber [4] (A1) を見よ。)

$m_a^\theta, -\infty \leq a < \theta \leq \infty$ , は確率過程  $X_a, X_{a+1}, \dots, X_b$  は  $\mathcal{F}$ , て生成される  $\sigma$ -加法族である。それをそれぞれ  $m_{-\infty}^\theta, m_k^\theta$  可測な 2 次モーメントをもつ任意の確率変数とするとき, Kolmogorov と Rozanov [6] にて定義された  $m_{-\infty}^\theta$  と  $m_k^\theta$  の最大相関係数  $\varphi(k)$  は次のようく与えられる。

$$(2.6) \quad \varphi(k) = \sup_{\xi, \eta} \frac{E\{(\xi - E\xi)(\eta - E\eta)\}}{\{E(\xi - E\xi)^2 E(\eta - E\eta)^2\}^{1/2}}.$$

(N2) 適当な定数  $\delta > 0$  に対して

$$(i) \quad E\{| \psi_\theta(X_0 \| X_{-1}, \dots, X_{-p}) |^{2+\delta}\} < \infty, \theta \in \Theta,$$

$$(ii) \quad \sum_{k=0}^{\infty} \varphi(k)^{\delta/2+\delta} < \infty.$$

$$(2.7) \quad \lambda(\theta) = E\{\psi_\theta(X_0 \| X_{-1}, \dots, X_{-p})\}$$

$$u_{\theta, d}(X_0 \| X_{-1}, \dots, X_{-p}) = \sup_{|\tau - \theta| < d} |\psi_\tau(X_0 \| X_{-1}, \dots, X_{-p}) - \psi_\theta(X_0 \| X_{-1}, \dots, X_{-p})|$$

とおく。ここで  $|\cdot|$  はユークリッドのノルムと同値な任意のノルム。

(N3)  $\lambda(\zeta) = 0$  をみたす  $\zeta \in \Theta$  が存在する。

(N4) 適当な正定数  $a, b, c$  および  $d_0$  に対して

- (i)  $|\lambda(\theta)| > a|\theta - \zeta|$ ,  $|\theta - \zeta| < d_0$ ,
- (ii)  $E\{\psi_{\theta, d}(X_0 \| X_1, \dots, X_p)\} \leq bd$ ,  $|\theta - \zeta| + d \leq d_0$ ,  $d > 0$ ,
- (iii)  $E\{\psi_{\theta, d}(X_0 \| X_1, \dots, X_p)^2\} \leq cd$ ,  $|\theta - \zeta| + d \leq d_0$ ,  $d > 0$ .

仮定(N2)-(ii) から  $\{X_t; t \in \mathbb{Z}\}$  は係数が  $\varphi(k)$  の strongly mixing (Rosenblatt の "mixing" ) であることが示され、て汎関数の中心極限定理([6] 定理 18.6.2) によって次がわかる。

### 命題1

$\frac{1}{\sqrt{n}} \sum_{t=1}^n \psi_\theta(X_t \| X_{t-1}, \dots, X_{t-p})$  は漸近的に平均ベクトル  $\lambda(\theta)$  で分散共分散行列

$$(2.8) \quad \Gamma(\theta) = E\{\bar{\psi}_\theta(0) \bar{\psi}_\theta(0)'\} + 2 \sum_{k=1}^{\infty} E\{\bar{\psi}_\theta(0) \bar{\psi}_\theta(k)'\}$$

の正規分布に従う。ここで  $\bar{\psi}_\theta(k) = \psi_\theta(X_k \| X_{k-1}, \dots, X_{k-p}) - \lambda(\theta)$ 。

最大相関係数の定義からただちに次を得る。

### 補題1

$\xi(x_1, \dots, x_p)$ ,  $\eta(x_1, \dots, x_q)$  はともに  $\mathcal{B}^p$ -および  $\mathcal{B}^q$ -可測な関数で、観測  $\{X_t\}$  の確率の  $\lambda$  について 2乗可積分ならば

$$|E\{\xi_0 \eta_k\} - E\xi_0 E\eta_k| \leq 4\varphi(k-p-q) E\{\xi_0^2\}^{1/2} E\{\eta_0^2\}^{1/2}, \quad k > p+q,$$

$$E\left\{ \left| \sum_{i=1}^n (\xi_i - E\xi_0) \right|^2 \right\} \leq n \left\{ 4p + 16 \sum_{i=1}^n \varphi(i) \right\} E\{\xi_0^2\}$$

たとえ  $\xi_k = \xi(X_{k+1}, X_{k+2}, \dots, X_{k+p})$ ,  $\eta_0 = \eta(X_0, \dots, X_q)$ 。

さて Huber [4] による

$$(2.9) \quad Z_n(\tau, \theta) = \frac{\left| \sum_{i=0}^n \{\psi_\tau(X_i \| X_{i-1}, \dots, X_{i-p}) - \psi_\theta(X_i \| X_{i-1}, \dots, X_{i-p}) - \lambda(\tau) + \lambda(\theta)\} \right|}{\sqrt{n} + n|\lambda(\tau)|}$$

とおくと Huber の Lemma 3, Theorem 3, Corollary は成り立つ  
る以下が得られる。証明は観測の独立性のかわりに補題 1 を  
使用するとこだりが違っている。

### 補題 2

仮定 (N1) ~ (N4)のもとで  $n \rightarrow \infty$  ならば

$$\sup_{|\tau - \zeta| \leq d_0} Z_n(\tau, \zeta) \rightarrow 0, \text{ in prob.}$$

### 定理 2

(N1) ~ (N4)を仮定し  $\tau_n$  は (2.5) を満たす。 $n \rightarrow \infty$  に対して  
 $P\{| \tau_n - \zeta | \leq d_0\} \rightarrow 1$  ならば

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\zeta(X_i \| X_{i-1}, \dots, X_{i-p}) + \sqrt{n}\lambda(\tau_n) \rightarrow 0, \text{ in prob.}$$

### 系 1

$\lambda(\theta)$  かつ  $\zeta \in \Theta$  で 正則な微分行列をもつとせよ。すなわち

$\zeta \Lambda(\zeta) = E\left\{\frac{\partial^2}{\partial \theta^2} \log f_\zeta(x_0 || x_1, \dots, x_p)\right\}$ 。すなは

$$(2.10) \frac{1}{\sqrt{n}} \sum_{i=0}^n \frac{\partial}{\partial \theta} \log f_\zeta(x_i || x_{i-1}, \dots, x_{i-p}) - \sqrt{n}(\tau_n - \zeta) \Lambda(\zeta) \rightarrow 0 \text{ in prob.}$$

系 2.

$\sqrt{n}(\tau_n - \zeta)$  は漸近的に平均ベクトル  $\zeta$ , 共分散行列  $\Lambda^{-1}(\zeta) \Gamma(\zeta) \Lambda^{-1}(\zeta)$  の正規分布に従う。

§3. 正規自己回帰モデルによる近似と漸近理論。  
(AR)

前節の結果は定常過程に対して非線形な自己回帰モデルをあてはめた議論を含んでいる。この節では普通の線形正規自己回帰モデル(ARモデル)を考える。ただし  $\zeta$  が  $\theta$  である。

$$(3.1) Y_t + \theta_1 Y_{t-1} + \dots + \theta_p Y_{t-p} = \varepsilon_t, p=1, 2, \dots$$

によって生成されるとして考える。ここに  $\{\varepsilon_t\}$  は  $N(0, s^2)$  に従い独立性をもつているものとする。

ARモデルの遷移確率密度のパラメータ族  $\{f_\theta(y_0 || y_{-1}, \dots, y_{-p})$ ,  $\theta = (s^2, \theta_1, \dots, \theta_p)'$  が  $\textcircled{1}$  } はこうして

$$(3.2) f_\theta(y_0 || y_{-1}, \dots, y_{-p}) = \frac{1}{\sqrt{2\pi}s} e^{-\frac{1}{2s^2}(y_0 + \theta_1 y_{-1} + \dots + \theta_p y_{-p})^2}$$

で与えられる。ここに  $\textcircled{1}$  は  $R^{p+1}$  の開集合で AR 過程 (3.1) を安定にする領域とする。定理 1 に  $f$ ,  $\gamma$  具体的計算することができ次を得る。

定理 3

次数  $p$  を固定したモデル (3.1) の観測  $\{X_t\}$  に対する最も推定量はパラメータ  $p\zeta = (\zeta_0, \zeta_1, \dots, \zeta_p)$  は a.s で収束する。但し  $p\zeta$  は  $\frac{\partial}{\partial \theta} K(p\zeta) = 0$  であります。これから形式的に、  
いわゆる Yule-Walker の方程式に一致し

$$(3.3) \quad \begin{aligned} \sum_{j=0}^p p(i-j)\zeta_j &= 0, \quad \zeta_0 = 1, \quad i=1, 2, \dots, p \\ S_p^2 &= \sum_{i,j=0}^p p(i-j)\zeta_i\zeta_j \end{aligned}$$

の解として一意に定まる。但し  $p(i-j) = E X_i X_j$ 。

同様にして定理 2 の系 2 をこの場合に適用して計算すると  
最も推定量  $\hat{\theta}_n$  に対して  $\sqrt{n}(\hat{\theta}_n - p\zeta)$  は漸近的に  $N(0, \Lambda^{-1}\Gamma\Lambda)$   
に従うことかわかる。ここで行列  $\Lambda, \Gamma$  はそれぞれ

$$(3.4) \quad \begin{aligned} \Lambda &= \Lambda(p\zeta) = \{ \lambda_{ij} \}_{i,j=0,1,2,\dots,p} \\ &= \sqrt{S_p^2} \left[ \begin{array}{cccccc} 2 & & & 0' & & \\ -1 & \ddots & \cdots & - & \cdots & \cdots \\ & \ddots & \ddots & & & \\ 0 & & & (c_{ij})_{i,j=1,2,\dots,p} & & \end{array} \right] \end{aligned}$$

ここで  $c_{ij} = p(i-j)$ .

$$(3.5) \quad \Gamma = \Gamma(p\zeta) = \{ \gamma_{ij} \}_{i,j=0,1,2,\dots,p}$$

たとえ  $p\bar{\zeta}_t = X_t + \zeta_1 X_{t-1} + \dots + \zeta_p X_{t-p}$  とおいたとき

$i, j = 1, 2, \dots; p \quad i = j + l - 2$

$$\gamma_{ij} = S_p^{-4} E \left\{ p \xi_t^2 X_{t-i} X_{t-j} + 2 \sum_{k=1}^{\infty} p \xi_t \cdot p \xi_{t+k} \cdot X_{t+k-i} X_{t-j} \right\}$$

$$\gamma_{i,0} = S_p^{-5} E \left\{ (p \xi_t^2 - S_p^2) p \xi_t X_{t-i} + 2 \sum_{k=1}^{\infty} (p \xi_{t+k}^2 - S_p^2) p \xi_t X_{t-i} \right\}$$

$$\gamma_{0,0} = S_p^{-6} E \left\{ (p \xi_t^2 - S_p^2)^2 + 2 \sum_{k=1}^{\infty} (p \xi_t^2 - S_p^2) (p \xi_{t+k}^2 - S_p^2) \right\}$$

これから の意義論のためには  $\{X_t\}$  のスペクトル密度が存在して、すなはち  $\rho(n) = \int_{-\pi}^{\pi} e^{inx} f(\lambda) d\lambda$  で、 $f(\lambda)$  は  $[-\pi, \pi]$  で 0 点をもつて 解析的な関数  $g(\lambda)$  といふと  $\gamma_{ij}$  で一致するとする。すなはち  $g(\lambda)$  は 単位円上 の関数とみなされ、これが 適当な半径  $R > 1$  の 内盤 の中で 正則な関数  $B(z)$  に 解析接続できる。

$$B(z) = \sqrt{2\pi} \exp \left[ \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{e^{-i\lambda} + z}{e^{i\lambda} - z} \log f(\lambda) d\lambda \right]$$

ここで  $B(z)$  は  $|z| < R$  で 0 点をもたない。 $|z| < R$  で

$$B(z) = \sum_{j=1}^{\infty} b_j z^j, \quad b_0 = 1,$$

$$A(z) = 1/B(z) = \sum_{j=1}^{\infty} a_j z^j, \quad a_0 = 1,$$

と 展開できることから  $X_t$  は 次のようないわば表現を持つ。

$$(3.6) \quad X_t = \xi_t + a_1 \xi_{t-1} + a_2 \xi_{t-2} + \dots$$

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} + \dots = \xi_t$$

$i \in I = \{\xi_t; t \in \mathbb{Z}\}$  は 無相関な確率変数列である。

さて、いま適当な整数  $q > 0$  について  $a_q \neq 0$ かつ  $a_{q+1} = a_{q+2} = \dots = 0$  であるとする。 $\{X_t\}$  は線形過程であるとする。すなはち  $X_t = E\{X_t | X_{t-1}, X_{t-2}, \dots\} + \xi_t$  で  $\xi_t$  は  $X_{t-1}, X_{t-2}, \dots$  と無相関。

$E(\xi_t | X_{t-1}, X_{t-2}, \dots) = 0$ ,  $E(\xi_t^2 - E\xi_t^2 | X_{t-1}, X_{t-2}, \dots) = 0$  であるから  $q \leq p$  の  $p$  に対する (3.5) はより簡単な  $\gamma_{ij} = P(i-j)/\sigma^2$ ,  $\gamma_{i,0} = E\{\xi_t^3 X_{t-i}\}/\sigma^5$ ,  $\gamma_{0,0} = E\{\xi_t^4\}/\sigma^6$  という値となる。ここで  $\{X_t\}$  が移動平均過程ならば、すなはち  $\{\xi_t\}$  が i.i.d. のとき,  $q \leq p$  に対して

$$\Gamma = \Gamma(p\xi) = \{\gamma_{ij}\}_{i,j=0,1,2,\dots,p}$$

$$= \sigma^{-2} \begin{bmatrix} \frac{\mu_4 - \sigma^4}{\sigma^4} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 0 \\ 0 & \ddots & & \{P(i-j)\}_{i,j=1,2,\dots,p} \end{bmatrix}$$

$$\text{と } \mu_4 = E\xi_t^4.$$

次に適当な整数  $q > 0$  に対して  $a_q \neq 0$  であって  $q$  以後の項は極めて小さく  $|a_{q+1}| + |a_{q+2}| + \dots \rightarrow 0$  となる、  
とすると。 $q \leq p$  のとき  $|p\xi_t - \xi_t| \rightarrow 0$  in prob.  
である。一方  $\xi_t \leq |X_t| + |a_1 X_{t-1}| + \dots$  であるから  $X_t$  の  
4次のモーメントがみれば補題1は  $\Rightarrow$  (3.5) の成分

たちは  $(\sigma, a_1, \dots, a_p, 0, 0, \dots)$  の近傍で一様に有界である。

ルベーグの収束定理がつかえて次の定理を得る。

#### 定理4

移動平均過程(3.6)が有限次元のAR過程で十分に近似されるとする。すなはち或る  $q > 0$  に対して  $|a_q| \neq 0$ かつ  $|a_{q+1}| + |a_{q+2}| + \dots \rightarrow 0$  とするとき  $p \geq q$  を 3p 次の自己回帰モデル(3.1)の最大推定量  $p\hat{\theta}_n$  は  $\sqrt{n}(\hat{\theta}_n - \theta)$  は漸近的  $N(\underline{0}, \Sigma)$  に従う。ただし

$$\Sigma = \sigma^2 \begin{bmatrix} \frac{a_4 - a_2^2}{4a_2^4} & \underline{0}' \\ \underline{0} & \left\{ p(i-j) \right\}_{i,j=1,\dots,p}^{-1} \end{bmatrix}.$$

いま AR モデル(3.1)の近似のよさをエントロピー(2.3)ではかるとする。

$$K(\theta) = \log \sqrt{2\pi} s + \frac{1}{2s^2} \int_{-\pi}^{\pi} |1 + \theta_1 z + \dots + \theta_p z^p|^2 f(\lambda) d\lambda,$$

$$z = e^{i\lambda}, \quad \lambda \mapsto \pi \text{ と } -\pi$$

$$K(p\zeta) = \inf_{s, \theta_1, \dots, \theta_p} K(\theta) = \log \sqrt{2\pi} s_p + \frac{1}{2}$$

ただし

$$\begin{aligned} s_p^2 &= \int_{-\pi}^{\pi} |1 + \zeta_1 z + \dots + \zeta_p z^p|^2 f(\lambda) d\lambda \\ &= \inf_{\theta} \int_{-\pi}^{\pi} |1 + \theta_1 z + \dots + \theta_p z^p|^2 f(\lambda) d\lambda. \end{aligned}$$

$s_p^2$  は  $p \rightarrow \infty$  とともに単調に減少して  $\sigma^2 = E \xi_t^2$  に収束するることはすぐわかる。ここで以下は Grenander & Rosenblatt の結果を擧げる。

### Theorem [3]

$p \rightarrow \infty$  のとき  $s_p^2 = s_p^2 - \sigma^2$  が遙くよりも指數的に 0 に収束する: すなはち  $f(\lambda)$  が  $[-\pi, \pi]$  で 0 点をもたない角折閏数  $g(\lambda)$  といたるとき一致することは同値である。

このことから我々の場合では  $K(\beta)$  は  $K(\infty)$  に遙くよりも指數的に収束する: とかかってた。範例  $\{X_t\}$  が正規過程ならば  $K(\infty) = 0$  である。以後  $\{X_t\}$  は正規にしよう。そうすると最大相関係数 (2.6) が 0 に収束する: これと strongly mixing は同値である (Kolmogorov-Rozanov [7])。さらに:  $f(\lambda)$  が角折閏数  $g(\lambda)$  といたるとき一致することと  $g(k) \rightarrow 0$  が遙くよりも指數的であることは同値である。(Ibragimov [5])。故に (N1) ~ (N4) を満すが  $i, j = 1, 2, \dots, p$  に対し  $\lambda_{ij} = E\{\xi_t^2 X_{t-i} X_{t-j}\} / s_p^2 \sigma^2$  となる: これは注意する補題 1 によつて

$$\begin{aligned} & | \lambda_{ij} - \lambda_{ij} | \\ & \leq \frac{E\{(p\xi_t^2 - \xi_t^2)^2\}^{1/2} E\{X_t^4\}^{1/2}}{s_p^4} + \frac{(s_p^2 - \sigma^2) \rho(i-j)}{s_p^4} + 2 \sum_{k=1}^{\infty} g(k) E\left\{ \left( p\xi_t - \xi_t \right)^4 \right\}^{1/4} E\left\{ \xi_t^4 \right\}^{1/4} \end{aligned}$$

$E_p \xi_t = E \xi_t = 0$ ,  $S_p^2 - \sigma^2 \rightarrow 0$  (遅くとも指數的) ため  $\xi_t$  正規性  
 やし  $E\{(p\xi_t - \xi_t)^4\} \rightarrow 0$ ,  $E\{(p\xi_t^2 - \xi_t^2)^2\} \rightarrow 0$  (いすゞて  
 遅くとも指數的) が出了。すなはち適當な  $C > 0$ ,  $0 < p < 1$  は  
 ある。

$$|\gamma_{ij} - \lambda_{ij}| \leq C p^p, \quad i, j = 1, 2, \dots, p$$

同様に  $i=1, 2$

$$|\gamma_{i,0}| \leq C p^p, \quad i=1, 2, \dots, p, \quad |\gamma_{00} - z| \leq C p^p$$

である。以上をまとめると次の定理を得る。

### 定理 5

$\{X_t; t \in \mathbb{Z}\}$  は正規定常過程で  $[-\pi, \pi]$  で 0 点をもたない  
 解析複数  $g(\lambda)$  とよどんだる  $\zeta(\lambda)$  一致するスペクトル  
 密度関数  $f(\lambda)$  を持つせよ。すなはち  $g(k)$  および  $K(\rho)$   
 はそのとき  $k \rightarrow \infty$ ,  $\rho \rightarrow \infty$  とともに遅くとも指數的に  $O(1/k^4)$   
 である。さらに適當な  $C > 0$ ,  $0 < p < 1$  が存在する。

$$\|\Gamma(p\beta) - \Lambda(p\beta)\| \leq C p^p, \quad p = 1, 2, \dots,$$

$$\|A\| = \max_{i,j} |a_{ij}|, \quad A = (a_{ij}).$$

### 例

一次の移動平均過程が観測されるとする。

$$X_t = \xi_t - \theta \xi_{t-1}, \quad |\theta| < 1,$$

$\{\xi_t\}$  は i.i.d. で  $E \xi_t = 0$ ,  $E \xi_t^2 = \sigma^2$ ,  $E \xi_t^4 = \mu_4$ 。

これは次のようす表現にも書けた。

$$X_t + \theta X_{t-1} + \theta^2 X_{t-2} + \dots = \xi_t$$

これに対する  $p$  次の AR モデルをみてはめたときの最良近似

は (3.3) 1 =  $\xi_t$

$$X_t + \theta X_{t-1} + \theta^2 X_{t-2} + \dots + \theta^{p-1} X_{t-p+1} + \frac{\theta^p}{1+\theta^2} X_{t-p} = \varepsilon_t$$

とおこらせて  $S_p^2 = E \varepsilon_t^2 = \sigma^2 \left( 1 + \frac{\theta^{2p+2}}{1+\theta^2} \right)$  となる。

#### § 4. Minimum AIC Estimation procedure

$X_0, X_1, X_2, \dots$  は定常過程の観測。 $Y_0, Y_1, Y_2, \dots$  は将来の観測で  $\{X_t\}$  と同法則であるから  $\{X_t\}$  と  $\{Y_t\}$  は互に独立であるとする。マルコフモデルによつて観測の予測を行なうためにとき最も推定量  $\hat{\theta}_n = \hat{\theta}_n(X_0, \dots, X_n)$  をつかって  $f_{\hat{\theta}_n}(y_t | y_{t-1}, \dots, y_{t-p})$  を採用するのは自然なことであろう。次のよきな損失関数および危険関数を考へる。

$$W(\hat{\theta}_n) = 2E_y \left\{ -\log \prod_{t=1}^n f_{\hat{\theta}_n}(Y_t | Y_{t-1}, \dots, Y_{t-p}) \right\} = 2nK(\hat{\theta}_n)$$

$$R(p) = E_x \{ W(\hat{\theta}_n) \} = n E_x \{ K(\hat{\theta}_n) \},$$

ここで  $E_x, E_y$  は  $\{X_t\}, \{Y_t\}$  に対する平均を意味する。

$$\mathcal{F}_p = \{ f_{p\theta}(y_t | y_{t-1}, \dots, y_{t-p}); p\theta = (\theta_0, \theta_1, \dots, \theta_p, 0, \dots) \}$$

と書くと  $\mathcal{F}_p$  の中の最良近似は  $f_{p\beta}(y_t | y_{t-1}, \dots, y_{t-p})$  と書く

く。すなはち  $K(p\beta) = \inf_{\theta \in \Theta} \{ K(\rho\theta) \}$ 。  $\mathcal{F}_p$  の最大推定量  $\hat{\rho\theta_n}$ ,  $p=1, 2, \dots$ ,  $t_1 < t_2 < t$

$$R(\rho\hat{\theta_n}) = 2nK(p\beta) + 2nE_x \left[ E_y \left\{ \log \frac{f_{p\beta}(Y_0|Y_1, \dots, Y_p)}{f_{\rho\hat{\theta_n}}(Y_0|Y_1, \dots, Y_p)} \right\} \right]$$

とから定理1に依る。

$$\log \left( f_{p\beta} / f_{\rho\hat{\theta_n}} \right) = - \left( \frac{\partial}{\partial \theta} \log f_{p\beta} \right) (\hat{\rho\theta_n} - p\beta)' + \frac{1}{2} (\hat{\rho\theta_n} - p\beta) \left\{ \frac{\partial^2}{\partial \theta \partial \theta} \log f_{p\beta} \right\} (\hat{\rho\theta_n} - p\beta)' + o\left(\frac{1}{n}\right)$$

$E_y \left\{ \frac{\partial}{\partial \theta} \log f_{p\beta} \right\} = 0$  たゞして仮定(2)定理2の系2

をかうと漸近的:

$$2nE_x \left[ E_y \log \left\{ f_{p\beta} / f_{\rho\hat{\theta_n}} \right\} \right] = \frac{1}{2} \text{tr } R(p\beta) \Lambda(p\beta)^{-1}$$

となる。故に漸近的:

$$(4.1) \quad R(p) = 2nK(p\beta) + \text{tr } R(p\beta) \Lambda(p\beta)^{-1}$$

以下、観測  $X_0, X_1, X_2, \dots, X_n$  およびモデル  $\mathcal{F}_p$  の最大推定量  $\hat{\rho\theta_n} = \hat{\rho\theta_n}(X_0, X_1, \dots, X_n)$  をかうて  $R(p)$  の不偏推定

量を見出せん。定理1に依る  $n$  が充分大きくなると

$$-\sum_{t=1}^n \log f_{p\beta} = -\sum_{t=1}^n \log f_{\rho\hat{\theta_n}} + \sum_{t=1}^n \left( \frac{\partial}{\partial \theta} \log f_{p\beta} \right) (\hat{\rho\theta_n} - p\beta)' + \frac{1}{2} (\hat{\rho\theta_n} - p\beta) \left\{ \sum_{t=1}^n \frac{\partial^2}{\partial \theta \partial \theta} \log f_{p\beta} \right\} (\hat{\rho\theta_n} - p\beta)' + o(1)$$

定理1の系2に依る。

$$\sum_{t=1}^n \left( \frac{\partial}{\partial \theta} \log f_{p\beta} \right) (\hat{\rho\theta_n} - p\beta)' - n(\hat{\rho\theta_n} - p\beta) \Lambda(p\beta) (\hat{\rho\theta_n} - p\beta)' \xrightarrow{\text{in prob.}} 0$$

こうして漸近的は

$$-\sum_{t=1}^n \log f_{p\beta} = -\sum_{t=1}^n \log f_{p\hat{\theta}_n} + \frac{1}{2} (\hat{\theta}_n - \beta) \Lambda(\beta) (\hat{\theta}_n - \beta)'$$

と $\beta$ から $nK(\beta)$ を漸近的に平均値とする量

$$-\log \prod_{t=1}^n f_{p\hat{\theta}_n}(x_t | x_{t-1}, \dots, x_{t-p}) + \frac{1}{2} \text{tr } \Gamma(\beta) \Lambda(\beta)^{-1}$$

を得る。(たかく2 (4.1) 1 = 5, 2

$$(4.2) -2 \log \prod_{t=1}^n f_{p\hat{\theta}_n}(x_t | x_{t-1}, \dots, x_{t-p}) + 2 \text{tr } \Gamma(\beta) \Lambda(\beta)^{-1}$$

は平均量(漸近的) $R(p)$ をもつ。後半の量は統計量でないから何とかする必要がある。 $\beta$ のかわりに $p\hat{\theta}_n$ を代入することも考慮されるが、これが(3.5)の形などと合致しないと絶対的である。

より具体的には正規ARモデルを定常過程であればあるだけ

$$-2 \log \prod_{t=1}^n f_{p\hat{\theta}_n}(x_t | x_{t-1}, \dots, x_{t-p}) = n \log \hat{s}_p^2 + (1 + \log 2\pi) n$$

$\approx 2$

$$\hat{s}_p^2 = \inf_{\theta} \frac{1}{n} \sum_{t=1}^n (x_t + \theta_1 x_{t-1} + \dots + \theta_p x_{t-p})^2.$$

定理4の仮定の下で定常過程を $q$ 次AR $\Rightarrow p+2$ 充分近似可能とするなら $q \leq p+1$ で

$$2 \text{tr } \Gamma(\beta) \Lambda(\beta)^{-1} \approx 2(p+1) + \gamma_2,$$

$$\therefore \gamma_2 = (\mu_4/\sigma^4) - 3. \quad p \leq q \text{ すなはち } q \text{ 次数について}$$

は  $\text{Tr } \Gamma(p) A(p)^{-1}$  は複雑であるから充分大きな  $n$  のもとでは  $n \log \widehat{S}_p$  の方が大きいから  $R(p)$  の最小化にとって本質的でないことがわかる。

さて、もし観測が正規定常で定理5の仮定を満たしていれば定理5によると、適当な  $C > 0$ ,  $0 < p < 1$  が存在してすべての  $p = 1, 2, \dots, l = k+1$

$$|\text{Tr } \Gamma(p) A(p)^{-1} - (p+1)| \leq C p^2 p^{-p}$$

である。

いすゞの場合にも、とくにモードルの次数に共通な定数を無視すると  $\text{Tr } \Gamma(p) A(p)^{-1} - R(p)$  の不偏推定量から定数を引いた情報量標準 [1, 2]

$$AIC(p) = n \log \widehat{S}_p^2 + 2(p+1)$$

が導びかれる。

### 例

3節と同じ例に対する  $\widehat{\theta}_t$  が“正規ならば” (4.1) は

$$R(p) = 2n\sigma^2(1+\theta^2)^{-1}\theta^{2(p+1)} + (p+1)$$

となる。これは

$$p = \left\{ \log n + \log \frac{\theta^2 \sigma^2 \log \frac{1}{\theta}}{1+\theta^2} \right\} \left( 2 \log \frac{1}{\theta} \right)^{-1}$$

のときに最小化される。このことは、観測回数、パラメータの値の大きさがモードルのパラメータの次元につれての最適次元を示すことを示している。

## References

- [1] Akaike, H. (1973). Information theory and an extension of maximum likelihood principle, 2nd Int. Symp. Information Theory, B. N. Petrov and F. Csaki, eds., Akademiai Kiado, Budapest, 267-281.
- [2] Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion, System Identification: Advances and Case Studies, R. K. Mehra and D. G. Lainiotis, eds., Academic Press, New York, 27-96.
- [3] Grenander, U. and Rosenblatt, M. (1954). An extension of a theorem of G. Szego and its application to the study of stochastic processes, Transactions of the American Mathematical Society, vol. 76, 112-126.
- [4] Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions, Proc. 5th Berkeley Symp. Math. Statist. Prob., 1, 221-233.
- [5] Ibragimov, I. A. (1970). On the spectrum of stationary Gaussian sequences satisfying the strong mixing condition II. Sufficient conditions. Mixing rate, Theory of prob. and its appl. vol. 15, no. 1
- [6] Ibragimov, I. A. and Linnik, Yu. V. (1971). Independent and Stationary Sequences of Random Variables, Woters-Noordhoff Publishing, Groningen.
- [7] Kolmogorov, A. N. and Rozanov, Yu. A. (1960). On strong mixing conditions for stationary Gaussian processes, Theory of prob. and its appl., vol. 5, no. 2.