

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そ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ない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(エントロピーのいみで)近似するときの議論につかえるように拡張し、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する漸近理論 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の情報量規準(AIC)の正当化を試みる。

§2. 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 $\mathcal{F} = \{f_\theta(x_0 \| x_{-1}, \dots, x_{-p}), \theta \in \Theta, x_i \in \mathcal{X}\}$ 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 $f_\theta(x_0 \| x_{-1}, \dots, x_{-p}) > 0$ a.s. とする。

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

としたとき 観測 $\{X_t, t \in \mathbb{Z}\}$ に対し

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

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。 ρ_θ の正則条件はま、たく同様である。

定理 1

(2.3) $K(\theta) = E[-\log \rho_\theta(X_0 \| X_{-1}, \dots, X_{-p})]$, $\theta \in \Theta$, を最小にする $\zeta \in \Theta$ が唯一つ存在するならば (2.2) を満たす任意の統計量 τ_n は ζ に収束する。(a.s. または in prob.)

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

とおけ。 τ_n が或るパラメータ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 τ_n の列が漸近的に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 θ を固定して $\mathcal{P}$ -可測で θ について可分である。(Huber [4] (A1) を見よ。)

$\mathcal{M}_a^b$, $-\infty \leq a < b \leq \infty$, は確率過程 $X_a, X_{a+1}, \dots, X_b$ に $\mathcal{F}$,
で生成される σ -加法族である。 ξ, η をそれぞれ $\mathcal{M}_{-\infty}^0, \mathcal{M}_k^\infty$
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] によって定義された $\mathcal{M}_{-\infty}^0$ と $\mathcal{M}_k^\infty$ の
最大相関係数 $\varphi(k)$ は次のように与えられる。

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

(N2) 適当な定数 $\delta > 0$ に対して

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

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

$$(2.7) \quad \begin{aligned} \lambda(\theta) &= E\{\psi_\theta(X_0 \| X_{-1}, \dots, X_{-p})\} \\ \mathcal{U}_{\theta, d}(X_0 \| X_{-1}, \dots, X_{-p}) &= \sup_{|\tau - \theta| < d} |\psi_\tau(X_0 \| X_{-1}, \dots, X_{-p}) - \psi_\theta(X_0 \| X_{-1}, \dots, X_{-p})| \end{aligned}$$

と置く。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

$$(i) |\lambda(\theta)| > a|\theta - \zeta|, \quad |\theta - \zeta| < d_0,$$

$$(ii) E\{u_{\theta,d}(X_0 \| X_{-1}, \dots, X_{-p})\} \leq bd, \quad |\theta - \zeta| + d \leq d_0, d > 0,$$

$$(iii) E\{u_{\theta,d}(X_0 \| X_{-1}, \dots, X_{-p})^2\} \leq cd, \quad |\theta - \zeta| + d \leq d_0, d > 0.$$

仮定 (N2)-(ii) から $\{X_t; t \in \mathbb{Z}\}$ は係数が $\varphi(k)$ の *strongly mixing* (Rosenblatt の意味での) であることがいえて汎関数の中心極限定理 ([6] 定理 18.6.2) によって次が出る。

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

$\xi(x_1, \dots, x_p), \eta(x_1, \dots, x_q)$ はそれぞれ $\mathcal{B}^p$ -および $\mathcal{B}^q$ -可測な関数で、観測 $\{X_t\}$ の確率のいみで 2 乗可積分ならば

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

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

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

さて Huber [4] にならうと

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

とあくと Huber の Lemma 3, Theorem 3. Corollary に対応する以下が得られる。証明は観測の独立性のかわりに補題 1 を使用するとこが違っている。

補題 2

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

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

定理 2

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

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

系 1

$\lambda(\theta)$ から $\xi \in \Theta$ で正則な微分行列をもつとせよ。すなわ

ち $\Lambda(\zeta) = E \left\{ \frac{\partial^2}{\partial \theta \partial \theta'} \log f_{\zeta}(X_0 \| X_1, \dots, X_p) \right\}$ とする

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

系 2.

$\sqrt{n}(\tau_n - \zeta)$ は漸近的に平均ベクトル 0, 共分散行列

$\Lambda^{-1}(\zeta) \Gamma(\zeta) \Lambda^{-1}(\zeta)$ の正規分布に従う。

§3. 正規自己帰帰モデルによる近似と漸近理論。

前節の結果は定常過程に対して非線形な自己帰帰モデルをあてはめる議論を含んでいる。この節では普通の線形正規自己帰帰モデル (AR モデル) を考える。 $\{X_t\}$ が満たす

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

によって生成されていると考えるのである。ここに $\{\varepsilon_t\}$ は $N(0, s^2)$ に従い独立性をもっているものとする。

AR モデルの遷移確率密度のパラメータ族 $\{f_{\theta}(y_0 \| y_1, \dots, y_p)\}$, $\theta = (s^2, \theta_1, \dots, \theta_p)' \in \mathcal{U}\}$ はこうして

$$(3.2) \quad f_{\theta}(y_0 \| y_1, \dots, y_p) = \frac{1}{\sqrt{2\pi}s} e^{-\frac{1}{2s^2}(y_0 + \theta_1 y_1 + \dots + \theta_p y_p)^2}$$

で与えられる。ここに $\mathcal{U}$ は $\mathbb{R}^{p+1}$ の開集合で AR 過程 (3.1) を安定にする領域とする。定理 1 によって具体的に計算することから次を得る。

定理 3

次数 p を固定したモデル (3.1) の観測 $\{X_t\}$ に対する最尤推定量はパラメータ $p\zeta = (\zeta_p, \zeta_1, \dots, \zeta_p)$ に a.s. に収束する。

但し $p\zeta$ は $\frac{\partial}{\partial \theta} K(p\zeta) = 0$ であり、これから形式的に
いの中々の Yule-Walker の方程式に一致し

$$(3.3) \quad \sum_{j=0}^p p(i-j) \zeta_j = 0, \quad \zeta_0 = 1, \quad i=1, 2, \dots, p$$

$$s_p^2 = \sum_{i,j=0}^p p(i-j) \zeta_i \zeta_j$$

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尤推定量 $\hat{\theta}_n$ に対して $\sqrt{n}(\hat{\theta}_n - p\zeta)$ は漸近的に $N(0, \Lambda^{-1}\Gamma\Lambda^{-1})$ に従うことがわかる。ここで行列 Λ, Γ はそれぞれ

$$(3.4) \quad \Lambda = \Lambda(p\zeta) = \{ \chi_{ij} \}_{i,j=0,1,2,\dots,p}$$

$$= \frac{1}{s_p^2} \begin{bmatrix} -2 & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & 0 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & (c_{ij})_{i,j=1,2,\dots,p} & \vdots & \vdots & \vdots \end{bmatrix}$$

ここで $c_{ij} = p(i-j)$ 。

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

ただし $p\zeta_t = X_t + \zeta_1 X_{t-1} + \dots + \zeta_p X_{t-p}$ とおいたとき

$$i, j = 1, 2, \dots, p \quad i \geq j+1$$

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

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

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

こゝからの議論のために $\{X_t\}$ のスペクトル密度が存在して、すなわち $\rho(n) = \int_{-\pi}^{\pi} e^{in\lambda} f(\lambda) d\lambda$ で、 $f(\lambda)$ は $[-\pi, \pi]$ で 0 点をもたない解析的関数 $g(\lambda)$ といえるところと一致するとする。すると $g(\lambda)$ は $\mathbb{C}$ 上の関数とみるとき、こゝが適当な半径 $R > 1$ の円盤の中で正則な関数 $B(z)$ に解析接続でき

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

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

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

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

と展開できるから X_t は次のような表現を持つ。

$$(3.6) \quad X_t = \xi_t + b_1 \xi_{t-1} + b_2 \xi_{t-2} + \dots$$

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

こゝに $\{\xi_t; t \in \mathbb{Z}\}$ は無相関な確率変数列である。

さて、いま適当な整数 $q > 0$ で $a_q \neq 0$ かつ $a_{q+1} = a_{q+2} = \dots = 0$ であるとしよう。さらに $\{X_t\}$ は線形過程であるとす。すなわち $X_t = E\{X_t | X_{t-1}, X_{t-2}, \dots\} + \xi_t$ で ξ_t は $X_{t-1}, X_{t-2}, \dots$ と無相関。

$$E(\xi_t | X_{t-1}, X_{t-2}, \dots) = 0, \quad E(\xi_t^2 - E\xi_t^2 | X_{t-1}, X_{t-2}, \dots) = 0$$

であるから $q \leq p$ なる p に対して (3.5) はより簡単な

$\gamma_{ij} = \rho(i-j)/\sigma^2$, $\gamma_{i,0} = E\{\xi_t^3 X_{t-i}\}/\sigma^5$, $\gamma_{0,0} = E\{\xi_t^4 - \sigma^4\}/\sigma^6$ という値をとる。とくに $\{X_t\}$ が移動平均過程ならば、すなわち $\{\xi_t\}$ が i.i.d. のとき、 $q \leq p$ に対して

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

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

となる。ただし $\mu_4 = E\xi_t^4$ 。

次に適当な整数 $q > 0$ に対して $a_q \neq 0$ であってそれ以後の項は極めて小さく $|a_{q+1}| + |a_{q+2}| + \dots \rightarrow 0$ となっているとする。 $q \leq p$ に対してこのとき $|p\xi_t - \xi_t| \rightarrow 0$ 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 次のモーメントが有限ならば補題 1 によって (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。

ルベグ-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。すなわち或る $q > 0$ に対して $|a_q| \neq 0$ かつ $|a_{q+1}| + |a_{q+2}| + \dots \rightarrow 0$ とするとき $p \geq q$ なる p -次の自己回帰モデル (3.1) の最尤推定量 $p\hat{\theta}_n$ について $\sqrt{n}(p\hat{\theta}_n - p\zeta)$ は漸近的に $N(0, \Sigma)$ に従う。ただし

$$\Sigma = \sigma^2 \begin{bmatrix} \frac{114 - \sigma^4}{4\sigma^4} & \underline{0}' \\ \underline{0} & \{r(i-j)\}_{i,j=1,\dots,p}^{-1} \end{bmatrix}.$$

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。

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

$z = e^{i\lambda}$, となつてゐることから

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

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 $\sigma^2 = E \xi_t^2$ に収束することはすぐわかる。ここで以下に Grenander と Rosenblatt の結果を掲げる。

Theorem [3]

$p \rightarrow \infty$ とともに $\delta_p = s_p^2 - \sigma^2$ が速くとも指数的に 0 に収束することと $f(\lambda)$ が $[-\pi, \pi]$ で 0 点をもたない解析関数 $g(\lambda)$ といえるところと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 $K(p)$ は $K(\infty)$ に速くとも指数的に収束することからわかった。観測 $\{X_t\}$ が正規過程ならば $K(\infty) = 0$ である。以後 $\{X_t\}$ は正規としよう。そうすると最大相関係数 (2.6) が 0 に収束することと strongly mixing は同値である (Kolmogorov-Rozanov [7])。さらに $f(\lambda)$ が解析関数 $g(\lambda)$ といえるところと一致することと $\rho(k) \rightarrow 0$ が速くとも指数的であることは同値である (Ibragimov [5])。故に (N1) ~ (N4) を満たすから $i, j = 1, 2, \dots, p$ に對して $\lambda_{ij} = E \{ \xi_t^2 X_{t-i} X_{t-j} \} / s_p^2 \sigma^2$ となることに注意すると補題 1 によって

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

$E_p \xi_t = E \xi_t = 0$, $S_p^2 - \sigma^2 \rightarrow 0$ (遅くとも指数的) である正規性から $E\{(p\xi_t - \xi_t)^4\} \rightarrow 0$, $E\{(p\xi_t^2 - \xi_t^2)^2\} \rightarrow 0$ (いずれも遅くとも指数的) が出る。すなわち適当な $C > 0$, $0 < p < 1$ に對して

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

同様にして

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

となる。以上をまとめる次の定理を得る。

定理 5

$\{X_t; t \in \mathbb{Z}\}$ は正規定常過程で $[-\pi, \pi]$ で 0 点をもたない解析関数 $g(\lambda)$ とほとんどいたるところで一致するスペクトル密度関数 $f(\lambda)$ を持つとせよ。すると $\varphi(k)$ および $K(p; \beta)$ はそれぞれ $k \rightarrow \infty$, $p \rightarrow \infty$ とともに遅くとも指数的に 0 に収束する。さらに適当な $C > 0$, $0 < p < 1$ が存在して

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

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

例 1

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 $E \xi_t = 0$, $E \xi_t^2 = \sigma^2$, $E \xi_t^4 = \mu_4$ 。

これは次のような表現にも書ける。

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

これに対して p 次の AR モデルをあてはめたときの最良近似は (3.3) によって

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

と与えらる。 $S_p^2 = E \varepsilon_t^2 = \sigma^2 \left(1 + \frac{\theta^{2p+2}}{1+\theta^2} \right)$ 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って観測の予測を行ないたいとき最尤推定量 $\hat{\theta}_n = \hat{\theta}_n(X_0, \dots, X_n)$ 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。次のような損失関数および危険関数を考える。

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

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

ここに E_x, E_y はそれぞれ $\{X_t\}, \{Y_t\}$ に対する平均を意味する。

$\mathcal{F}_p = \{ f_{p\theta}(y_t \| y_{t-1}, \dots, y_{t-p}); p\theta = (\theta_1, \theta_2, \dots, \theta_p, 0, \dots) \}$ と書いて $\mathcal{F}_p$ の中での最良近似を $f_{p\hat{\theta}}(y_t \| y_{t-1}, \dots, y_{t-p})$ と書く

く。すなわち $K(p\zeta) = \inf_{\theta \in \Theta} \{K(p\theta)\}$ 。 f_p の最尤推定量を $p\hat{\theta}_n$, $p=1,2,\dots$, としたとき

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

となるから定理1に よって

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

$E_y \left\{ \frac{\partial}{\partial \theta} \log f_{p\zeta} \right\} = 0$ なることを仮定して定理2の系2

をつかうと漸近的に

$$2nE_x \left[E_y \log \{f_{p\zeta}/f_{p\hat{\theta}_n}\} \right] = \frac{1}{2} \text{tr} \Gamma(p\zeta) \Lambda(p\zeta)^{-1}$$

となる。故に漸近的に

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

以下、観測 $X_0, X_1, X_2, \dots, X_n$ およびモデル f_p の最尤推定量 $p\hat{\theta}_n = p\hat{\theta}_n(X_0, X_1, \dots, X_n)$ をつかって $R(p)$ の不偏推定量を見出したい。定理1に よって n が充分大きいとき

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

定理1の系2に よって

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

こうして漸近的に

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

となるから $nK(p\beta)$ を漸近的に平均値とする次の量

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

を得る。したがって (4.1) によって

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

は平均値(漸近的) $R(p)$ をもつ。後半の量は統計量でないから不可欠の必要がある。 $p\beta$ のかわりに $p\hat{\theta}_n$ を代入することも考えられるかもしれないが (3.5) の形などを見ると絶望的である。

より具体的にいま正規ARモデルを定常過程にあてはめるならば

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

ことに

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

定理4の仮定のいみで定常過程が q 次ARによって充分近似可能とするなら $q \leq p$ に対して

$$2n \Gamma(p\beta) \Lambda(p\beta)^{-1} \doteq 2(p+1) + \gamma_2,$$

ここで $\gamma_2 = (\mu_4/\sigma^4) - 3$ 。 $p \leq q$ なる次数について

は $2n \Gamma(p+1)A(p+1)^{-1}$ は複雑であるが 充分大きな n のもとでは $n \log \hat{S}_p$ の方が大きいから $R(p)$ の最小化によ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るならば 定理5によって、適当な $C > 0$, $0 < p < 1$ が存在してすべての $p = 1, 2, \dots$ に対して

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

となる。

いずれの場合にも、そのモデルの次数に共通な定数を無視することによって $R(p)$ の不偏推定量から定数を引いた情報量規準 [1, 2]

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

が導びかれる。

例

3節と同じ例に対して ε_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

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

のときに最小化される。このことは、観測数、パラメータの値の大きさがモデルのパラメータの次元についての最適次元と関係することを示している。

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.