数理解析研究所講究録
第 312 巻 1977 年 144-160

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に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 $\mathcal{F} = \{ f_\theta(x_0 \| x_{-1}, \cdots, x_{-p}),\ \theta \in \Theta,\ x_i \in \mathcal{X} \}$ を考える。ここで $\Theta \subset R^d$ は開集合であり、任意の $x_0, x_{-1}, \cdots, x_{-p} \in \mathcal{X}$ と $\theta \in \Theta$ に対して $f_\theta(x_0 \| x_{-1}, \cdots, x_{-p}) > 0$ a.s. とする。

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

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

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

を満たす推定量 $T_n = T_n(X_0, X_1, \cdots, X_n)$ 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 の正則条件はまったく同様である。

定理 1

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

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

とおけ。T_n が或るパラメータに確率収束することを仮定し

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

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

(N 1) $\psi_\theta(x_0 \| x_{-1},\cdots,x_{-p})$ はθを固定して$\mathcal{B}^p$–可測でθについて可分である。(Huber [4] (A 1) を見よ。)

$\mathfrak{M}_a^b$, $-\infty \leqq a < b \leqq \infty$, は確率過程 $X_a, X_{a+1},\cdots,X_b$ によって生成されるσ–加法族である。ξ, ηをそれぞれ$\mathfrak{M}_{-\infty}^0$, $\mathfrak{M}_k^\infty$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$\mathfrak{M}_{-\infty}^0$と$\mathfrak{M}_k^\infty$の最大相関係数$\varphi(k)$は次のように与えられる。

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

(N 2) 適当な定数$\delta > 0$に対して

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

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

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

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。

(N 3)　$\lambda(\zeta)=0$ をみたす $\zeta\in\Theta$ が存在する。

(N 4)　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)　$E\{u_{\theta,d}(X_0\|X_{-1},\cdots,X_{-p})\}\leqq bd$,　$|\theta-\zeta|+d\leqq d_0$, $d>0$,

(iii)　$E\{u_{\theta,d}(X_0\|X_{-1},\cdots,X_{-p})^2\}\leqq cd$,　$|\theta-\zeta|+d\leqq d_0$, $d>0$。

仮定(N 2)-(ii)から $\{X_t;t\in\mathbb{Z}\}$ は係数が $\varphi(k)$ の strongly mixing (Rosenblattのいみでの)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

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

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

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

最大相関係数の定義からただちに次を得る。

補題1

$\xi(x_1,\cdots,x_p)$, $\eta(x_1,\cdots,x_q)$ 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

$$|E\{\xi_0\eta_k\}-E\xi_0E\eta_0|\leqq 4\varphi(k-p-q)E\{\xi_0^2\}^{1/2}E\{\eta_0^2\}^{1/2},\quad k>p+q,$$

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

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

さて Huber [4] にならって

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

とおくと Huber の Lema 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 を使用するところだけ違っている。

補題 2

仮定(N1)—(N4)のもとに $n\to\infty$ ならば

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

定理 2

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

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

系 1

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

ち $\Lambda(\xi)=E\{\frac{\partial^2}{\partial\theta\partial\theta'}\log f_\xi(X_0\|X_{-1},\cdots,X_{-p})\}$。すると

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

系 2.

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

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

§3. 正規自己回帰(AR)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回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回帰モデル（ARモデル）を考える。 $\{X_t\}$があたかも

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

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考える。

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

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

で与えられる。ここにΘ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によって具体的に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$\{X_t\}$に対する最尤推定量はパラメータ $_p\zeta=(s_p, \zeta_1, \cdots, \zeta_p)$ に a.s に収束する。但し $_p\zeta$ は $\frac{\partial}{\partial\theta}K({}_p\zeta)=0$ でもあり、これから形式的にいわゆるYule-Walkerの方程式に一致し

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

の解として一意に定まる。但し $\rho(i-j)=E X_i X_j$。

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$N(\underline{0}, \Lambda^{-1}\Gamma\Lambda^{-1})$に従うことがわかる。ここで行列$\Lambda, \Gamma$はそれぞれ

$$\Lambda=\Lambda({}_p\zeta)=\{\lambda_{ij}\}_{i,j=0,1,2,\cdots,p}$$

(3.4)
$$=1/s_p^2\left[\begin{array}{c|c} 2 & \underline{0}' \\ \hline \underline{0} & (C_{ij})_{i,j=1,2,\cdots,p}\end{array}\right]$$

ここで $C_{ij}=\rho(i-j)$.

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

ただし ${}_p\xi_t=X_t+\zeta_1X_{t-1}+\cdots+\zeta_pX_{t-p}$ とおいたとき

$i, j = 1, 2, \cdots, p$ に対して

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

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

$$\gamma_{0.0} = S_p^{-6} E\{ ({}_p\xi_t^2 - S_p^2)^2 + 2\sum_{k=1}^{\infty} ({}_p\xi_t^2 - S_p^2)({}_p\xi_{t+k} - S_p^2) \}$$

これ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な関数 $g(\lambda)$ といたるところで一致するとする。すると $g(\lambda)$ を円上の関数とみたとき、これが適当な半径 $R > 1$ の円盤の中で正則な関数 $B(z)$ に解析接続できて

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

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。

$$(3.6) \quad \begin{aligned} X_t &= \xi_t + b_1 \xi_{t-1} + b_2 \xi_{t-2} + \cdots \\ X_t + a_1 X_{t-1} + a_2 X_{t-2} + \cdots &= \xi_t \end{aligned}$$

ここに $\{\xi_t ; t \in \mathbb{Z}\}$ 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 $a_{q+1}=a_{q+2}=\cdots=0$ であるとしよう。さらに $\{X_t\}$ は線形過程であるとする。すなわち $X_t=E\{X_t\|X_{t-1},X_{t-2},\cdots\}+\xi_t$ で ξ_t は $X_{t-1},X_{t-2},\cdots$ と無相関。

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

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

$\gamma_{ij}=\rho(i-j)/\sigma^2,\ \gamma_{i,0}=E\{\xi_t^3X_{t-i}\}/\sigma^5,\ \gamma_{0,0}=E\{\xi_t^4-\sigma^4\}/\sigma^6$

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，$q\leqq p$ に対して

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

$$=\sigma^{-2}\left[\begin{array}{c:c}\dfrac{\mu_4-\sigma^2}{\sigma^4} & \underline{0}' \\ \hdashline \underline{0} & \{\rho(i-j)\}_{i,j=1,2,\cdots p}\end{array}\right]$$

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 $|a_{q+1}|+|a_{q+2}|+\cdots\cdots\longrightarrow 0$ となっているとする。$q\leqq p$ に対してこのとき $|{}_p\xi_t-\xi_t|\longrightarrow 0$ in prob. である。一方 $\xi_t\leqq|X_t|+|a_1X_{t-1}|+\cdots$ であるから X_t の4次のモーメントがあれば補題1によって (3.5) の成分

たちは $(\sigma, a_1, \cdots, a_p, 0, 0, \cdots)$ 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。すなわち或る $q>0$ に対して $|a_q| \neq 0$ かつ $|a_{q+1}|+|a_{q+2}|+\cdots\cdots \to 0$ 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 $N(\underline{0}, \Sigma)$ に従う。ただし

$$\Sigma = \sigma^2 \left[\begin{array}{c:c} \dfrac{\mu_4 - \sigma^4}{4\sigma^4} & \underline{0}' \\ \hdashline \underline{0} & \{\rho(i-j)\}^{-1}_{i,j=1,\cdots,p} \end{array} \right]。$$

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。

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

$z = e^{i\lambda}$, となっていることから

$$K({}_p\zeta) = \inf_{s^2, \theta_1, \cdots, \theta_p} K(\theta) = \log \sqrt{2\pi}\, S_p + \frac{1}{2}$$

ただし

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

S_p^2 は $p\to\infty$ 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]

$p\to\infty$ とともに $\delta_p=S_p^2-\sigma^2$ が遅くとも指数的に 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 $g(\lambda)$ といたるところで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 $K(p\xi)$ は $K(\infty\xi)$ に遅くとも指数的に収束することがわかった。観測 $\{X_t\}$ が正規過程ならば $K(\infty\xi)=0$ 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 $g(\lambda)$ といたるところで一致することと $\varphi(k)\to 0$ が遅くとも指数的であることは同値である。(Ibragimov [5])。故に (N1)~(N4) を満すから $i,j=1,2,\dots,p$ に対して $\lambda_{ij}=E\{\xi_t^2X_{t-i}X_{t-j}\}/S_p^2\sigma^2$ となることに注意すると補題1によって

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

$E_p\xi_t = E\xi_t = 0$, $s_p^2 - \sigma^2 \to 0$（遅くとも指数的）だから正規性から $E\{({}_p\xi_t - \xi_t)^4\} \to 0$, $E\{({}_p\xi_t^2 - \xi_t^2)^2\} \to 0$（いずれも遅くとも指数的）が出る。すなわち適当な $C>0$, $0<\rho<1$ に対して

$$|\gamma_{ij} - \lambda_{ij}| \leqq C\rho^p \ , \quad i,j=1,2,\cdots,p$$

同様にして

$$|\gamma_{i,0}| \leqq C\rho^p, \ i=1,2,\cdots,p, \quad |\gamma_{00} - 2| \leqq C\rho^p$$

となる。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 $K({}_p\xi)$ はそれぞれ $k\to\infty$, $p\to\infty$ とともに遅くとも指数的に0に収束する。さらに適当な $C>0$, $0<\rho<1$ が存在して

$$\|\Gamma({}_p\xi) - \Lambda({}_p\xi)\| \leqq C\rho^p \ , \quad p=1,2,\cdots,$$

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

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

これは次のような表現にも書ける。

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

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

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

と与えられて、$S_p^2 = E\varepsilon_t^2 = \sigma^2\left(1 + \frac{\theta^{2p+2}}{1+\theta^2}\right)$となる。

§4 Minimum AIC Estimation procedure

$X_0, X_1, X_2, \cdots$は定常過程の観測。$Y_0, Y_1, Y_2, \cdots$は将来の観測で$\{X_t\}$と同法則であるが$\{X_t\}$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${}_p\hat{\theta}_n = {}_p\hat{\theta}_n(x_0, \cdots, x_n)$をつかって$f_{{}_p\hat{\theta}_n}(y_t \| y_{t-1}, \cdots, y_{t-p})$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。

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

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

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。

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

と書いて$\mathcal{F}_p$の中での最良近似を$f_{p\xi}(y_t \| y_{t-1}, \cdots, y_{t-p})$と書

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

$$R({}_p\hat{\theta}_n)=2nK({}_p\zeta)+2nE_x\left[E_y\left\{\log\frac{f_{{}_p\zeta}(Y_0\|Y_{-1},\cdots,Y_{-p})}{f_{{}_p\hat{\theta}_n}(Y_0\|Y_{-1},\cdots,Y_{-p})}\right\}\right]$$

となるから定理1によって

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

$E_y\left\{\frac{\partial}{\partial\theta}\log f_{{}_p\zeta}\right\}=0$ なることを仮定して定理2の系2をつかうと漸近的に

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

となる。故に漸近的に

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

以下観測 $X_0, X_1, X_2, \cdots, X_n$ およびモデル $\mathcal{F}_p$ の最尤推定量 ${}_p\hat{\theta}_n={}_p\hat{\theta}_n(X_0,X_1,\cdots,X_n)$ 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

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

定理1の系2によって

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

こうして漸近的に

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

となるから $nK({}_p\xi)$ を漸近的に平均値とする次の量

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

を得る。したがって (4.1) によって

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

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何とかする必要がある。${}_p\xi$ 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) の形などをみると絶望的である。

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

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

ここに

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

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 $q \leqq p$ に対して

$$2\operatorname{tr}\Gamma({}_p\xi)\Lambda({}_p\xi)^{-1} \fallingdotseq 2(p+1) + \gamma_2,$$

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

は $\mathrm{tr}\,\Gamma(p,\xi)\Lambda(p,\xi)^{-1}$ 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るならば定理5によって、適当な $C>0$, $0<\rho<1$ が存在してすべての $p=1,2,\cdots$ に対して

$$\left|\mathrm{tr}\,\Gamma(p,\xi)\Lambda(p,\xi)^{-1}-(p+1)\right|\leqq C p^2\rho^{-p}$$

となる。

いずれの場合にも、それぞれ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。

<u>例</u>

3節と同じ例に対して ξ_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

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

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 muximum 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., Academie 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.