

時系列推定における高次の漸近有効性について

竹内 啓

§ 1

時系列における母数推定の理論は、一面から見れば過去の数十年間に大いに進歩したが、しかしなお最適推定論 optimum estimation の観点からは、極めて不十分なところがあるといわざるを得ない。

いま時系列データ $X_1(\omega), X_2(\omega), \dots, X_t(\omega), \dots$ が与えられたものとしよう。このとき推定論の問題を明確に定式化するためには、次のものを明確に定義しなければならぬ。

1. $\Omega = \{\omega\}$ 上で定義された確率測度の族 $\{P_\omega^0\}$ $\theta \in \Theta$
2. 推定すべき母数 (実、または実ベクトル) $Y = g(\theta)$
3. 考慮される推定量のクラスと、その「よさ」の基準

時系列における推定論の固有の問題は、上記の 1~3 が明確に定義されることと前提にした上で、最適な推定量を求めること、或いは特定の推定量の相対的な効率を求めることである。従って時に問題にされる推定量の計算手続さの問題は推定論の固有の問題ではない。勿論、いわゆる「簡便法」の相対効率の問題となつて得るけれども。

時系列解析の書物において、上記の 1~3 の点、とくに 1 と 3 に関して明確の規定が欠けていることが、問題を不明確

にしていることが少なくない。例えばいわゆる「スペクトル密度の推定」において、考慮の対象になっている分布の集合がどのようなものか、より具体的に言えば、「真の」密度関数からどのような集合に属するものであるかを明らかにすることが、いろいろの推定法の比較を直観的な議論の水準にとどめている。

実際時系列解析において、いわゆる「Non-parametric」すなわち具体的な分布形を仮定しないモデルについて、厳密な推定論を展開することは、抽象的な測度論のレベルでも、また推定法の比較という観点から、著しく困難であるように思われる。推定論の観点からは「parametric」な、より具体的に言えばカウス過程を中心としたモデルだけが manageable であるといわなければならない。しかし現実のデータがこのような特定の分布形を持つ分布で厳密に従うということは保証されないかもしれない。このような場合を処理するためには、モデルからの乖離によって影響を受けることの少ないような手法を求める robustness の観点を導入しなければならない。

しかしここでは robustness の問題はとらあけたい。ことにしたい。すなわち「正しい」parametric model が与えられているものとする。更に分布を規定する母教団自体が

有限次元の実ベクトルで表せるものに限定する。このときには推定すべき母数をその自体に与えてよい。

ところで時系列モデルにおいては、厳密な小標本理論を展開することは不可能に近い。例えば最小分散不偏推定量が存在するような場合はいくつもない。そもそも不偏推定量の存在自体が疑問である場合が多い。もっともその不存在的証明は一つの興味ある問題であるが。

従って理論的に期待し得る結果は、ほとんど漸近的なものに限られるであろう。実際これまでに得られている結果も、ほとんど漸近理論の範囲内のものである。しかしこの分野にもなおいくつかの問題がある。

1 推定量の漸近的最適性 asymptotic optimality の定義と厳密な定式化。

2, 漸近分布への収束の速さと高次の漸近最適性 asymptotic optimality of higher order の問題

1 に関して、しばしば最尤推定量の漸近最適性について厳密な裏付けが示されていることが多い。或いは独立同一分布の観測値からなる標本についての結果が、そのまま援用されていることも少なくない。いうまでもなくこのことは理論的に妥当でないが、しかしながら時系列解析のいろいろなモデルに関する限り、最尤推定量の漸近正規性、漸近有効性等

について、結果的には得られた結論に等しくなることはわかって
いる。標準的な漸近理論の元（竹内啓「統計的推定
の漸近理論」参照）が、中心極限定理の適用などにおいてし
かるべき注意を怠らなければ、ここでも適用できる。

しかしながら時系列モデルにおいては、極限分布への収束
はもう少し速くたいてい通例である。また同じ理由によること
であるが、漸近的には同等になる二つの推定量が、現実には有
意に異なる値を示すこともある。従って高次の漸近分布につい
ての立ち入った議論が必要である。

しかしこの点に関しては、独立同一分布の場合についても
まだ理論は完成していない。またこれまでに得られた理論を
時系列モデルに適用するについては、独立でない確率変数の
和の分布の Edgeworth 展開を理論的に基礎づけたわけではないが、
いれどもより容易ではない。

以下においては、最近になって明らかになった高次の漸
近有効性に関する「拡張された正則推定量」の性質について
のべ、その時系列モデルへの適用について論じたい。いかに
よって問題を考察するための手がかりを得るこれが、この文
の主要な目的である。

§ 2

最初に筆者によって最近得られた結果を紹介する。

$X_1, X_2, \dots, X_n, \dots$ が互いに独立に次のような密度を持つ指数型分布に従うとする。

$$f(x, \theta) = c(\theta) h(x) \exp \sum_{i=1}^p \lambda_i(\theta) t_i(x)$$

ここに t_i は実数値をとる関数で、 t_i は互いに線形独立とする。また $\theta \in \Theta \subset \mathbb{R}^p$ は p 次元実母数で $p < \infty$, $\theta \in \Theta$ に対し、 $(\lambda_1(\theta), \dots, \lambda_p(\theta))$ は $(\lambda_1, \dots, \lambda_p)$ のとり得る値の自然な範囲の内点になつてゐるとする。また λ_i は θ に関し適当な次数まで連続微分可能であると仮定する。

上記の分布に対して十分統計量は

$$T = (T_1, \dots, T_p) \quad T_i = \frac{1}{n} \sum_{j=1}^n t_i(X_j)$$

で與えられる。 T の分布は $n \geq p$ のとき Lebesgue 測度に関し絶対連続であると仮定する。このとき T はすべての次数のモーメントを持ち、かつこの分布の密度関数を Edgeworth 展開することができる。

θ の「拡張された正則推定量」 extended regular estimator とは次のような形に表される推定量をいう。

$$\hat{\theta} = g(T_1, \dots, T_p) + \frac{1}{n} h_1(T_1, \dots, T_p) + \dots + \frac{1}{n^k} h_k(T_1, \dots, T_p)$$

ここに $g, h_1, \dots, h_k$ はすべて n とおくまでもない関数で g は $2k-1$ 回、 h_j は $k-j$ 回それぞれ連続微分可能と仮定する。

このとき $\hat{\theta}$ の Taylor 展開が可能であつて、これにより $\hat{\theta}$ の漸近分布の密度関数の Edgeworth 展開と與えることが

ができる。

ここで $k=1$ とすると、 $\hat{\theta}$ の密度の $1/n$ の order までの漸近展開が得られる。いま $\sqrt{n}(\hat{\theta} - \theta)$ の漸近 cumulant を

$$E(\sqrt{n}(\hat{\theta}_\alpha - \theta_\alpha)) = \mu_\alpha/\sqrt{n} + o(1/\sqrt{n})$$

$$E(\sqrt{n}(\hat{\theta}_\alpha - \theta_\alpha)(\hat{\theta}_\beta - \theta_\beta)) = \sigma_{\alpha\beta} + Q_{\alpha\beta}/\sqrt{n} + o(1/\sqrt{n})$$

$$E(\sqrt{n}(\hat{\theta}_\alpha - \theta_\alpha)(\hat{\theta}_\beta - \theta_\beta)(\hat{\theta}_\gamma - \theta_\gamma)) = \beta_{\alpha\beta\gamma}/\sqrt{n} + o(1/\sqrt{n})$$

$$E(\sqrt{n}(\hat{\theta}_\alpha - \theta_\alpha)(\hat{\theta}_\beta - \theta_\beta)(\hat{\theta}_\gamma - \theta_\gamma)(\hat{\theta}_\delta - \theta_\delta)) = \sigma_{\alpha\beta}\sigma_{\gamma\delta} + \sigma_{\alpha\gamma}\sigma_{\beta\delta} \\ - \sigma_{\alpha\delta}\sigma_{\beta\gamma} = \gamma_{\alpha\beta\gamma\delta}/\sqrt{n} + o(1/\sqrt{n}),$$

という形で得られる。

ここで $\Sigma = \{\sigma_{\alpha\beta}\} \geq I^{-1}$ (I は情報行列) という関係が成り立ち、かつ等号は $\hat{\theta}$ が最尤推定量 あるいはそれと漸近的に同等な BAN 推定量の場合に限って成り立つ。

更にはすべての BAN 推定量について、 $\beta_{\alpha\beta\gamma}$ および $\gamma_{\alpha\beta\gamma\delta}$ は等しいことが示される。従って漸近分布の $1/n$ の order までの展開において、拡張された BAN 推定量のクラスの中で一定で正しいのは μ_α および $Q_{\alpha\beta}$ のみである。ところで μ_α は $\hat{\theta}$ について高次の漸近不偏性、高次の漸近中央値不偏性、あるいは高次のモード不偏性のいづれかを仮定するならば、 $\beta_{\alpha\beta\gamma}$ も定められしむ。従ってこのように何らかの意味での高次の正則性をみたすクラスの中では、 $Q_{\alpha\beta}$ のみが問題となる。ここで次のことが成り立つことが証明される。

$$\sum_{\alpha} \sum_{\beta} Q_{\alpha\beta} t_{\alpha} t_{\beta} \geq \sum_{\alpha} \sum_{\beta} Q_{\alpha\beta}^* t_{\alpha} t_{\beta} \quad \forall t_{\alpha}$$

ただし $Q_{\alpha\beta}^*$ は (拡張された) 最尤推定量に対応する値である。このことから次のことがいえる。 $\hat{\theta}^*$ と最尤推定量を高次の漸近正則性をみたすように補正した推定量 $\hat{\theta}$ と同じ漸近正則性をみたす (拡張された) 正則推定量とすると、原点を中心とする任意の凸集合 C に対して

$$\lim_{n \rightarrow \infty} n [\Pr\{\sqrt{n}(\hat{\theta}^* - \theta) \in C\} - \Pr\{\sqrt{n}(\hat{\theta} - \theta) \in C\}] \geq 0$$

が成り立つ。或いは $L(u)$ を

$$L(0) = 0 \quad L(u) > 0 \quad \text{for } u > 0 \quad \text{に対して}$$

集合 $\{u \mid L(u) \leq a\}$ は凸、 $L(u)$ は有界

な関数とすると

$$\lim_{n \rightarrow \infty} n [E\{L(\sqrt{n}(\hat{\theta}^* - \theta))\} - E\{L(\sqrt{n}(\hat{\theta} - \theta))\}] \leq 0$$

が成り立つ。

すなわち (拡張された) 最尤推定量は n^{-1} の order まで、

……を上げると漸近有効性を持つことがいえる。

§ 3

以上の結果は、実は直接に拡張可能であって、時系列モデルにも一部応用できるものである。

すなわち前節の結果を得るための本質的な前提は、有限次元の十分統計量が存在し、かつその密度関数の Edgeworth 展開が可能であるというだけであり、拡張された正則推定

母の漸近モーメントに関する結果は、これから直接に導かれるのである。

従って例えば $X_1, X_2, \dots, X_n, \dots$ がすべて互いに独立で、 X の密度関数 g が

$$f(x_j, \theta) = g_j(\theta) h_j(x_j) \exp \sum_i \eta_i(\theta) t_{ij}(x_j)$$

とこの形に表せるとき、もしある関数 $\bar{c}(\theta)$ が存在して

$$\frac{1}{n} \left(\frac{d}{d\theta} \right)^m \sum \log g_j(\theta) \rightarrow \frac{d^m}{d\theta^m} \bar{c}(\theta)$$

$$m = 1, 2, \dots$$

がなり立てられたら、前節の結果が成立することはいま直ちに示される。すなわちこの場合でも（拡張された）最大推定量は3次の漸近有効性を持っている。また各の漸近モーメントは尤度関数の導関数（高次のものを小さくして）の同時モーメントから定められる。

時系列解析の場合に、有限次元の十分統計量が存在するならば、同じく同じ結果がなり立つ。例えば自己回帰モデル

$$x_{t+k} = \beta_1 x_{t+k-1} + \beta_2 x_{t+k-2} + \dots + \beta_k x_t + \varepsilon_t$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

において $x_1, \dots, x_T$ の同時密度関数は

$$\left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{T-k} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=k+1}^T (x_t - \beta_1 x_{t-1} - \dots - \beta_k x_{t-k})^2 \right\} \\ \times f_0(x_1, \dots, x_k)$$

である。ただし $f_0(x_1, \dots, x_k)$ は "初期値" $x_1, \dots, x_k$ の同時密度関数である。そうすると十分統計量は

$$T_0 = \sum_{t=k+1}^T x_t^2 / (T-k), T_1 = \sum_{t=k+1}^T x_t x_{t-1} / (T-k) \dots T_k = \sum_{t=k+1}^T x_t x_{t-k} / (T-k)$$

$$\text{および } x_1, \dots, x_k; x_T, x_{T-1}, \dots, x_{T-k+1}$$

で与えられることがわかる。(このことは "初期値" の分布についての仮定とは無関係になり立つ)

ここで $T_0, T_1, \dots$ の同時分布については, Edgeworth 展開が可能であるから, これまでの場合と同様に扱うことができるが, $x_1, \dots, x_k, x_T, \dots, x_{T-k+1}$ については その持つ情報量は一定であって, T とともに増加することはない。従ってそれらを推定量の中に入れて入れるについては, 特殊の扱いが必要になる。一つの方法は推定量のクラスとして

$$\hat{\beta} = g(T_0, T_1, \dots, T_k) + \frac{1}{T} h(x_1, \dots, x_k, x_T, \dots, x_{T-k+1}, T_0, \dots, T_k)$$

とこの形のものを考えることである。この場合も問題は残されてゐるが本質的な困難はない。

有限次元の十分統計量が存在しない場合, 例之は 自己回帰・移動平均 (ARMA) モデルの場合などには困難は大きい。一つの方法は, 次元の多くなる統計量

$$T_0 = \sum x_t^2 / T, T_1 = \sum x_t x_{t-1} / (T-1) \dots T_r = \sum x_t x_{t-r} / (T-r)$$

をとって, それらにもとづく (拡張された) 正規推定量を考へ, r を固定して T の漸近分布を考へ, 次に T を無限に

大きくするこゝでろろ。これを固定すれば、"補正された"最
 大推定量の2次の漸近有効性を持つことは、これまでと同様
 の議論によつて示すことができるから、次に上と下になら
 ない通常の速さで増減させればよいでろろ。ただしこのま
 うの点にわたる吟味はまた今後の課題でろろ。