統計数理（1999）
第47巻 第1号 157-174

特集「統計基礎理論」
［研究詳解］

マルチンゲール確率場に対する エントロピー法とその統計的応用

統計数理研究所　西　山　陽　一

（受付 1998年9月24日；改訂 1999年1月18日）

要　旨

エントロピー法は，集合や関数によって添字づけられたI.I.D.データの経験分布過程に対する大数の法則や中心極限定理を確立するために，1980年代に研究された．さらに，最近のいくつかの論文はエントロピー法が他の統計的問題にも有用であることを示してきている．しかしながら，この方法のある部分はI.I.D.でないデータにも適用出来る可能性を秘めているにも拘わらず，マルチンゲールの枠組みでの体系的な研究はまだなされていない．マルチンゲールは重要な概念であることを考慮し，我々は文献のこのギャップを埋めるためのステップをつくることを意図する．

1節ではOssianderの中心極限定理の一般化に関する直感的な説明を通じて，エントロピー法の一般化のポイントがどこにあるかを解説する．簡明のため，本稿の残りの部分は連続マルチンゲールに限定して話を進める．2節では，本研究の鍵となるquadratic modulusという量を導入し，その言葉で最大不等式を与える．3節では弱収束定理を与える．それらを用いてカーネル推定量の局所確率場の漸近挙動，いくつかのM-推定量の収束率，積分型推定量の漸近正規性などを導出する．

キーワード：マルチンゲール，最大不等式，中心極限定理，カーネル推定量，変化点，最尤推定量．

1. 序 —— I.I.D.からマルチンゲールへ ——

本研究の目的は，これまで主としてI.I.D.データの経験分布過程に対して発展してきたエントロピー法の核心部分をマルチンゲールの枠組みに一般化し，確率過程の統計的推測への応用を可能ならしめる事である．ここで経験分布過程と呼んでいるのは，Dudleyらによって研究されてきた集合や関数を添字にもつ確率場の事である．直感的な説明をかねて，まずOssiander (1987) によるI.I.D.の場合に対する中心極限定理の紹介をし，その後にそれをdependentデータの最も原始的な場合に一般化する．その説明の中で，どこが核心部分であるかを解説する．

$(E, \mathcal{E})$は可測空間であるとする．$\{Z_i\}_{i\in\mathbb{N}}$はE-値のI.I.D.確率変数であるとし，その共通の分布をPと表す．Ψは$\mathcal{L}^2(P)=\mathcal{L}^2(E,\mathcal{E},P)$の部分集合であって，ある$\varphi\in\mathcal{L}^2(P)$に対し全ての$\psi\in\Psi$が$|\psi|\le\varphi$を満たすものとする（$\varphi$は$\Psi$に含まれていなくてもよい）．いま

$$X^n(\psi)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\psi(Z_i)-\int_E\psi(z)\,P(dz)\right\}$$

によって定義される確率過程 $\psi \rightsquigarrow X^n(\psi)$ を考える．Ossiander (1987) による L^2-bracketing とは次のようなものである；各 $\varepsilon \in (0,1]$ に対し，$N(\varepsilon)$ 個の $\mathcal{L}^2(P)$ の要素のペア $[l^{\varepsilon,k}, u^{\varepsilon,k}]$, $k = 1,\ldots, N(\varepsilon)$, であって，任意の $\psi \in \Psi$ に対して関係 $l^{\varepsilon,k} \le \psi \le u^{\varepsilon,k}$ がいずれかの k に対して成り立ち，しかも

$$\sqrt{\int_E |u^{\varepsilon,k}(z) - l^{\varepsilon,k}(z)|^2 P(dz)} \le \varepsilon \tag{1.1}$$

が成り立つ．Ossiander の定理の主張は「もしもこの bracketing procedure がエントロピー条件

$$\int_0^1 \sqrt{\log N(\varepsilon)}\, d\varepsilon < \infty, \tag{1.2}$$

を満たすように達成されるならば，確率過程の列 $\psi \rightsquigarrow X^n(\psi)$ は Ψ によって添字付けられたブラウン橋，すなわち，平均ゼロの正規確率過程 $\psi \rightsquigarrow G(\psi)$ であって $EG(\psi)\, G(\phi) = \int_E \psi(z)\, \phi(z)\, P(dz) - \int_E \psi(z)\, P(dz) \int_E \phi(z)\, P(dz)$ を満たすものに $\ell^\infty(\Psi)$ の中で弱収束する」という事である．ただし $\ell^\infty(\Psi)$ は Ψ 上の有界関数全体の空間とし，これに supremum norm を添付して完備距離空間とする．

さて，$\{Z_i\}_{i\in\mathbb{N}}$ は任意の E-値確率変数の列とし，P_i は $\mathcal{F}_{i-1} = \sigma\{Z_1,\ldots, Z_{i-1}\}$ (ただし $\mathcal{F}_0$ は null σ-field) を与えたもとでの Z_i の条件付き分布を表すものとする．いま

$$X^n(\psi) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \psi(Z_i) - \int_E \psi(z)\, P_i(dz) \right\}$$

によって与えられる確率過程 $\psi \rightsquigarrow X^n(\psi)$ を考える．上の bracketing procedure のうち，(1.1) を

$$\sqrt{\int_E |u^{\varepsilon,k}(z) - l^{\varepsilon,k}(z)|^2 P_i(dz)} \le K_i \varepsilon \quad \text{almost surely}$$

に取り替えたものを考えよう．ただし K_i は ε や k に依存しない確率変数であって $\mathcal{F}_{i-1}$-可測なるものである．いまの場合，左辺がランダムであるから右辺にもランダムな係数 K_i を許したのである．このとき，もしもエントロピー条件 (1.2) が満たされ，しかも

$$\bar{K}^n = \sqrt{\frac{1}{n} \sum_{i=1}^n |K_i|^2}$$

によって定義される確率変数が確率有界であるならば，確率過程 $\psi \rightsquigarrow X^n(\psi)$ の緊密性は有限次元分布の収束と，ある種の Lindeberg 条件から従う．Ossiander の定理は $P_i \equiv P$ かつ $K_i \equiv 1$ の場合であると考えられる．本研究の鍵となるランダム量 "quadratic modulus" は，この確率変数 $\bar{K}^n$ と類似したものである．より近い表現は

$$\text{"quadratic modulus"} = \sup_{\varepsilon \in (0,1]} \max_{1 \le k \le N(\varepsilon)} \frac{\sqrt{n^{-1} \sum_{i=1}^n \int_E |u^{\varepsilon,k}(z) - l^{\varepsilon,k}(z)|^2 P_i(dz)}}{\varepsilon}$$

である．このランダム量は bracket の選択に見苦しい形で依存しており，また我々は重み関数 ψ そのものがランダムである場合も考えたいので，実際には "quadratic modulus" は少し異なる形で定義される．それ以前に，本研究では「離散時間のマルチンゲール」,「点過程」,「連続マルチンゲール」の三つの場合に分けてアプローチをしているが，上述のものはその最初のものにすぎない．後 2 者のうちの特に「連続マルチンゲール」は異なる発想による "quadratic modulus"を定義する必要がある．本稿では，紙数の関係と説明の簡明のため（これまでのイントロとは別のものとなってしまうが，敢えて）「連続マルチンゲール」の場合における結果のみ報告する．しかし，他の2者についても同様のアプローチにより類似した結果が得られる事は強調しておきたい．

エントロピー法の統計学における有用性は，はじめは主としてより一般の一様大数の法則や中心極限定理（特に多次元のDonskerの定理）を求める努力を通じて認識されてきた．しかしながら，その核心部分，すなわちエントロピーによって制御されたchainingとbracketingの技術は，それらの極限定理に直接関係のない統計的問題にも適用できる事が，最近のいくつかの論文により示されてきている．例えば本稿でも扱うM-推定がそのよい例である．我々にとって，エントロピー法の利点の一つは，モデルがI.I.D.でなくても$\{\bar{K}^n \leq L\}$という形の集合上ではその技術がそのまま適用できるという事である．それ故，確率過程の統計的推測に関する問題のいくつかは，上述のようなtruncationを導入し，その補集合$\{\bar{K}^n > L\}$が大きな定数$L > 0$に対して無視できる事を示す事によって解決できる．これが本研究で繰り返し用いるアプローチである．

本研究の対象としてマルチンゲールを選ぶ理由は3つある．理由の第1は，I.I.D.の場合において基本的な道具であったBernsteinの不等式が，マルチンゲールの枠組みでも，predictable quadratic variationに基づくtruncationが必要であるという修正のもとで，すでに与えられている事である．第2に，我々の状況における有限次元分布の収束を示すために，よく発達したマルチンゲール中心極限定理を利用する事ができる．第3に，マルチンゲールという概念により，統計的モデルの豊富なクラスを扱う事ができる（例えば，生存解析におけるmultiplicative intensityモデル，マルコフ連鎖，正規白色ノイズモデル，拡散過程など）．

本稿の構成は次の通り．先にもふれたが，2節以降は敢えて「連続マルチンゲール」とその応用のみ扱う．まず2節において，基本となる最大不等式を述べる．3節では，それを用いて連続マルチンゲールから生成される確率場に対する中心極限定理を与える．4節において，中心極限定理を応用して，カーネル推定量の局所確率場の弱収束を導出する．5節と6節ではM-推定への応用を述べる．まず5.1節においてM-推定量の収束率を判定するための一般的な定理を紹介する．それは5.2, 5.3および6節において有用となる．5.2節では，4節の結果も用いて，正規白色ノイズモデルの回帰関数の最大点の位置の推定問題を考察し，5.3節では回帰関数のジャンプ点の位置の推定を行う．6節では，正規白色ノイズモデルの回帰関数のノンパラメトリック最尤推定量の収束率を導出する．特に，回帰関数が単調である場合と滑らかである場合についての例を検討する．7節では，正規白色ノイズモデルが部分的に観測されている（すなわちcensoringがある）場合について，回帰関数の積分値に対するNelson-Aalen型の推定量の漸近正規性を導出する．

若干の文献紹介をしてこの節を終える．集合や関数を添字にもつ経験分布過程の理論はI.I.D.の場合に対してDudley (1978)によって創始され，1980年代に列独立な三角列の場合まで研究された．van der Vaart and Wellner (1996)の本がその良い解説ならびに文献紹介を含んでいるので，ここでは敢えて述べない．1990年代に入り，独立性をはずす努力が始められた．定常過程に対するmixing conditionに関してはDoukhan et al. (1995)があり，マルチンゲールに関してはBae and Levental (1995a, 1995b)，van de Geer (1995)，Nishiyama (1997a)がある．本稿はNishiyama (1997a)の発展版であるNishiyama (1998)の一部を紹介したものであるが，話題の中心を連続マルチンゲールとその正規白色ノイズモデルへの応用に据えたため，大部分はNishiyama (1997b)で扱ったものとなった．

Bae and Levental (1995a)は，マルコフ連鎖の経験分布過程の中心極限定理を導出するために，離散時間のマルチンゲールに対する本稿の1節の第三段落と類似した主張を得た．また

Bae and Levental (1995b) は，パラメータ ψ に関する連続性も仮定した上で，本稿の定理 3.2 と同じ結論を得た．彼らの議論の中に，quadratic modulus という量の起源は見いだす事はできるが，その量の導入のためには van der Vaart and Wellner (1996) による新しい緊密性判定定理も大きな動機になっていることを付記しておきたい．一方 van de Geer (1995) は Bernstein の不等式のマルチンゲール版のジャンプの大きさに関する仮定をゆるめると同時に，Ossiander (1987) による bracketing の技術を用いて点過程のノンパラメトリック最尤推定量の収束率を導出した．Nishiyama (1997a) は同様の技術を用いて，弱収束定理を証明し，Nelson-Aalen 推定量への応用を得た．以上のように，本研究は，Dudley ら (特に Ossiander) による経験分布過程の研究をマルチンゲールに一般化するにあたり，特別な場合や少し異なる視点から考えた Bae and Levental や van de Geer によるアイデアと，van der Vaart and Wellner の本のタイムリーな出版にも負って得られた成果の集大成であると言える．

2. 連続マルチンゲールの確率場に対する最大不等式

$\boldsymbol{B}=(\Omega,\ \mathcal{F},\ \boldsymbol{F}=(\mathcal{F}_t)_{t\in\mathbb{R}_+},\ P)$ は filtration が添付された確率空間であるとし，$(\Psi,\ \rho)$ は距離空間であるとする．各 $\psi\in\Psi$ に対し $\boldsymbol{B}$ 上の連続局所マルチンゲール X^ψ であって $X_0^\psi=0$ なるものが与えられたとする．本節の目標は，その族 $X=(X^\psi|\psi\in\Psi)$ に対する最大不等式，すなわち $E\sup_{\psi,\phi}|X_t^\psi-X_t^\phi|$ の形の値を押さえる不等式を導出することである．第1節でも述べたが，それは以後の全ての結果の道具として有用である．

まず連続局所マルチンゲールの quadratic modulus を定義する事から始めよう．

定義 2.1. 与えられた部分集合 $S\subset\Psi$ に対し，Xの quadratic $(S,\ \rho)$-modulus を

$$\|X\|_{(S,\rho),t}=\sup_{\substack{\psi,\phi\in S\\ \psi\neq\phi}}\frac{\sqrt{\langle X^\psi-X^\phi,\ X^\psi-X^\phi\rangle_t}}{\rho(\psi,\ \phi)}\qquad \forall t\in\mathbb{R}_+$$

と定義する．

S が可算であれば $[0,\ \infty]$-値確率変数 $\|X\|_{(S,\rho),t}$ は $\mathcal{F}_t$-可測となるが，S が非可算であれば $\mathcal{F}$-可測性すら保障されないことに注意されたい．以下で，距離空間 (Ψ,ρ) が全有界であるとは任意の $\varepsilon>0$ に対して Ψ が有限個の半径 ε の閉球によって覆われることを意味するものとし，また $N(\Psi,\rho;\varepsilon)$ によってそういう閉球の最小個数を表すものとする．次が本節の主定理である．

定理 2.2. (Ψ,ρ) は全有界距離空間であるとする．τ は $\boldsymbol{B}$ 上の有限停止時刻であるとする．このとき，任意の δ，$K>0$ に対し

$$\sup_{\substack{\Psi^*\subset\Psi\\ \text{countable}}}E\sup_{t\in[0,\ \tau]}\sup_{\substack{\psi,\phi\in\Psi^*\\ \rho(\psi,\phi)\leq\delta}}|X_t^\psi-X_t^\phi|1_{\{\|X\|_{(\Psi^*,\rho),\tau}\leq K\}}\leq CK\int_0^\delta\sqrt{\log(1+N(\Psi,\ \rho\,;\ \varepsilon))}\,d\varepsilon$$

が成り立つ．ただし $C>0$ は universal constant で，最初の supremum は Ψ の全ての可算部分集合でとられる．

ところで，我々はパス $\psi\rightsquigarrow X_t^\psi$ や $(t,\ \psi)\rightsquigarrow X_t^\psi$ が連続あるいは有界であるかどうかという疑問にしばしば出くわす．上の結果を適用して，この問題に対する2種類の解答を提示する事ができる．1つ目のものは Ψ が可算集合である場合に関するものである．

定理 2.3.　定理 2.2 の状況を考える．もしも Ψ が可算でしかも

$$P(\|X\|_{(\Psi,\rho),\tau} < \infty) = 1 \quad \text{かつ} \quad \int_0^1 \sqrt{\log N(\Psi, \rho;\, \varepsilon)}\, d\varepsilon < \infty$$

であるならば，$\psi \rightsquigarrow X_\tau^\psi$ のほとんど全てのパスは Ψ 上で ρ に関して一様連続である；特に，それらは $\ell^\infty(\Psi)$ に属している．さらに $\tau > 0$ が定数であるならば，$(t, \psi) \rightsquigarrow X_t^\psi$ のほとんど全てのパスは $\tilde{\rho}((t, \psi), (s, \phi)) = |t - s| \vee \rho(\psi, \phi)$ によって定義される $[0, \tau] \times \Psi$ 上の擬距離 $\tilde{\rho}$ に関して一様連続である；特に，それらは $\ell^\infty([0, \tau] \times \Psi)$ に属している．

Ψ が非可算である場合，次の定理が $\psi \rightsquigarrow X_\tau^\psi$ の連続ヴァージョンの存在のための十分条件を与える．

定理 2.4.　定理 2.2 の状況を考える．もしも $\|X\|_{(\Psi,\rho),\tau} \leq Y$ (identically) なる $\mathcal{F}_\tau$-可測，$[0, \infty]$-値確率変数 Y が存在し，しかも

$$P(Y < \infty) = 1 \quad \text{かつ} \quad \int_0^1 \sqrt{\log N(\Psi, \rho;\, \varepsilon)}\, d\varepsilon < \infty$$

であるならば，ある $\mathcal{F}_\tau$-可測確率変数の族 $\{\tilde{X}(\psi) : \psi \in \Psi\}$ であって各 $\psi \in \Psi$ に対し確率 1 で $\tilde{X}(\psi) = X_\tau^\psi$ であり，しかも $\psi \rightsquigarrow \tilde{X}(\psi)$ のほとんど全てのパスが Ψ 上で ρ に関して一様連続であるものが存在する；特に，それらは $\ell^\infty(\Psi)$ に属している．(そのような確率過程 $\psi \rightsquigarrow \tilde{X}(\psi)$ は $\psi \rightsquigarrow X_\tau^\psi$ の ρ-連続ヴァージョンと呼ばれる．)

構成された $\tilde{X}(\psi)$ はもはや連続局所マルチンゲールの終点値ではない事に注意されたい．

定理 2.2 において，ρ が Ψ 上の (擬距離ではなく) 距離であるという仮定は応用上強すぎる事もある．次の定理は，より一般に，ランダムな擬距離への修正に関するものである．ただし，エントロピーは依然としてある距離に関して計算しなければならない．

定理 2.5.　定理 2.2 の状況を考える．もしも与えられた Ψ 上のランダム擬距離 d が

$$\sqrt{\langle X^\psi - X^\phi, X^\psi - X^\phi \rangle_\tau} \leq d(\psi, \phi) \qquad \forall \psi, \phi \in \Psi \quad \text{identically}$$

を満たすならば，任意の $\delta, K > 0$ に対し

$$\sup_{\substack{\Psi^* \subset \Psi \\ \text{countable}}} E^* \sup_{t \in [0,\tau]} \sup_{\substack{\psi,\phi \in \Psi^* \\ d(\psi,\phi) \leq K\delta}} |X_t^\psi - X_t^\phi| 1_{\{\|d\|_\rho \leq K\}} \leq CK \int_0^\delta \sqrt{\log(1 + N(\Psi, \rho;\, \varepsilon))}\, d\varepsilon,$$

ただし

$$\|d\|_\rho = \sup_{\substack{\psi,\phi \in \Psi \\ \psi \neq \phi}} \frac{d(\psi, \phi)}{\rho(\psi, \phi)},$$

が成り立つ．ここに $C > 0$ は universal constant で，最初の supremum は Ψ の全ての可算部分集合 Ψ^* でとられる．

3.　連続マルチンゲールの確率場に対する中心極限定理

まず 1 つの定義を用意するところから始めよう．

定義 3.1. 距離空間(Ψ, ρ)によって添字付けられた連続局所マルチンゲールの族$X=(X^\psi|\psi\in\Psi)$がρ-可分であるとは，ある可算部分集合$\Psi^*\subset\Psi$とP-ゼロ集合$N\in\mathcal{F}$が存在して，任意の$\varepsilon>0$と$\omega\in\Omega\backslash N$に対し

$$X_t^\psi(\omega)\in\overline{\{X_t^\phi(\omega): \phi\in\Psi^*,\ \rho(\psi,\phi)<\varepsilon\}}\qquad \forall t\in\mathbb{R}_+,\ \forall\psi\in\Psi,$$

が成り立つときにいう．ただし閉包は$\mathbb{R}\cup\{-\infty, +\infty\}$においてとられる．

もしも(Ψ,ρ)が可分ならば，このρ-可分性のための十分条件は，ほとんど全てのパス$\psi \rightsquigarrow X_t^\psi$が$\rho$-連続である事である．しかし，一般にはパスの連続性をチェックするのは容易ではない．一方，添字集合Ψが可算である場合には，どのような族もΨ上の任意の距離ρに関してρ-可分である事は明らかである．

では弱収束の文脈に移ろう．(Ψ,ρ)は全有界距離空間であるとする．各$n\in\mathbb{N}$に対し，filtrationが添付された確率空間$\boldsymbol{B}^n=(\Omega^n, \mathcal{F}^n, \boldsymbol{F}^n=(\mathcal{F}_t^n)_{t\in\mathbb{R}_+}, P^n)$が与えられたとする．$X^n=(X^{n,\psi}|\psi\in\Psi)$は$\boldsymbol{B}^n$上の（必ずしも$\rho$-可分ではない）連続局所マルチンゲール$X^{n,\psi}$であって$X_0^{n,\psi}=0$であるものの族であるとする．$\tau^n$は同じ確率空間上で定義された有限停止時刻であるとする．ここでMetric Entropy conditionを導入しよう．

［**ME**］ 与えられた$\boldsymbol{B}^n$上の有限停止時刻τ^nに対し

$$\|X^n\|_{(\Psi^*,\rho),\tau^n}=O_{P^n}(1)\quad かつ\quad \int_0^1\sqrt{\log N(\Psi,\rho;\varepsilon)}\,d\varepsilon<\infty.$$

定理 3.2. 上の状況において族$X^n=(X^{n,\psi}|\psi\in\Psi)$は$\rho$-可分であるとする．$\psi\rightsquigarrow X_{\tau^n}^{n,\psi}$は確率1で$\ell^\infty(\Psi)$に値をとり，さらにその任意の有限次元マージナルはあるBorel可測分布に弱収束するとする．このとき，もしも条件[**ME**]が満たされるならば，$X_{\tau^n}^n$は$\ell^\infty(\Psi)$の中である緊密Borel可測分布に弱収束する．

証明のポイントのみ述べる．一般に，全有界擬距離空間(T,ρ)によって添字付けられた，$\ell^\infty(T)$に値をとる確率場の列$t\rightsquigarrow X^n(t), n=1,2,\ldots$が緊密であるための十分条件は，任意の有限次元マージナルが弱収束する事と任意の$\varepsilon, \eta>0$に対しある$\delta>0$が存在して

$$\limsup_{n\to\infty}P^{n*}\left(\sup_{\rho(t,s)\le\delta}|X^n(t)-X^n(s)|>\varepsilon\right)\le\eta$$

である事である．定理3.2の枠組みの場合，これは2節で与えた最大不等式（定理2.2）とMarkovの不等式から直ちに従う．

さて，定理2.3によれば，もしもΨが可算で［**ME**］が満たされるならば，$\psi\rightsquigarrow X_{\tau^n}^{n,\psi}$が確率1で$\ell^\infty(\Psi)$に値をとるための十分条件は$P^n(\|X^n\|_{(\Psi,\rho),\tau^n}<\infty)=1$が成り立つ事である．$\Psi$が非可算である場合には，$\psi\rightsquigarrow X_{\tau^n}^{n,\psi}$そのものではなくその$\rho$-連続ヴァージョンを扱った次の系が有用となる．$\rho$-連続ヴァージョンの存在自体は定理2.4によって保証されている事を思い出そう．

系 3.3. 上の状況において，$\tau^n\equiv\tau$は定数であるとし，ある定数の族$\{C(\psi,\phi);\psi,\phi\in\Psi\}$に対し

$$\langle X^{n,\psi}, X^{n,\phi}\rangle_\tau\xrightarrow{P^n}C(\psi,\phi)\qquad\forall\psi,\phi\in\Psi$$

である事を仮定する．ある $\mathcal{F}_\tau^n$-可測，$[0, \infty]$-値確率変数 Y^n であって $\|X^n\|_{(\Psi,\rho),\tau} \leq Y^n$ identically と $P^n(Y^n < \infty) = 1$ を満たすものが存在する事を仮定する．このとき，もしも [**ME**] が満たされるならば，$\psi \rightsquigarrow X_\tau^{n,\psi}$ の ρ-連続ヴァージョン $\psi \rightsquigarrow \widetilde{X}^n(\psi)$ は，平均ゼロの正規過程 $\psi \rightsquigarrow G(\psi)$ であって $EG(\psi)\,G(\phi) = C(\psi, \phi)$ なるものに $\ell^\infty(\Psi)$ の中で弱収束する．

4. 応用１：カーネル推定量の局所確率場の弱収束

I.I.D.データの密度関数の推定問題において，カーネル推定量が各点ごとの漸近正規性をもつことはよく知られている．本節では，それを局所パラメータに関する汎関数の意味での漸近正規性に拡張する問題を考える．このことは，I.I.D.データをはじめとする様々な状況で可能であるが，ここでは次の式で与えられる正規白色ノイズモデルを考察する：

$$dX_t^n = f(t)\,dt + n^{-1/2}dB_t \qquad X_0^n = x_0 \in \mathbb{R}\,. \tag{4.1}$$

ただし $B = (B_t)_{t\in[0,1]}$ は標準ブラウン運動を表す．いま，ある一点 $t_0 \in (0,1)$ を固定する．ゼロに収束する正定数の列 b_n をとる．関数 $u \rightsquigarrow f(t_0 + b_n u)$ を推定することを我々の目標とする．ただし u は $\mathbb{R}$ の有界部分集合 U の中を動くものとする．自然な推定量は

$$\hat{f}_n(t_0 + b_n u) = \frac{1}{b_n}\int_0^1 K\left(\frac{t - t_0}{b_n} - u\right) dX_t^n \qquad \forall u \in U$$

である．これに関する 2 種類の残差確率場を考えよう：

$$Z_n(u) = \sqrt{nb_n}\,\{\hat{f}_n(t_0 + b_n u) - \tilde{f}_n(t_0 + b_n u)\} \qquad \forall u \in U;$$
$$R_n(u) = \sqrt{nb_n}\,\{\hat{f}_n(t_0 + b_n u) - f(t_0 + b_n u)\} \qquad \forall u \in U\,.$$

ただし

$$\tilde{f}_n(t_0 + b_n u) = \frac{1}{b_n}\int_0^1 K\left(\frac{t - t_0}{b_n} - u\right) f(t)\,dt \qquad \forall u \in U$$

とする．カーネル関数 K には次の仮定をおく．

条件 4.1. 関数 $K : \mathbb{R} \to [0, \infty)$ は，原点対称，$\int_{\mathbb{R}} K(x)\,dx = 1$，かつコンパクトな台をもち，しかも次の(i)あるいは(ii)のいずれかを満たす：

(i) $x \rightsquigarrow K(x)$ は Lipschitz 連続；
(ii) $x \rightsquigarrow K(x)$ は $[0, \infty)$ 上で単調減少．

定理 4.2. カーネル関数 K は条件 4.1 を満たすものとする．定数列 b_n を $b_n \downarrow 0$ かつ $nb_n \uparrow \infty$ となるようにとる．

(i) もし $t \rightsquigarrow f(t)$ が点 $t_0 \in (0, 1)$ において連続であるならば，$Z^n \xRightarrow{P} Z$ in $\ell^\infty(U)$．ただし $u \rightsquigarrow Z(u)$ は平均ゼロの正規過程で

$$EZ(u_1)\,Z(u_2) = \int_{\mathbb{R}} K(x - u_1)\,K(x - u_2)\,dx \qquad u_1, u_2 \in U$$

を満たすものである．
(ii) もし $t \rightsquigarrow f(t)$ が点 $t_0 \in (0, 1)$ の近傍において 2 回連続的微分可能で，しかも

$$\lim_{n\to\infty} nb_n^5 = h < \infty$$

であるならば，$R^n \overset{P}{\Longrightarrow} z_0 + Z$ in $\ell^\infty(U)$．ただし

$$z_0 = \frac{\sqrt{h}}{2}\cdot\int_{\mathbb{R}} x^2 K(x)\,dx\cdot \left.\frac{d^2 f(t)}{dt^2}\right|_{t=t_0}.$$

上記の主張に現れる仮定は，各点ごとの漸近正規性の文脈におけるものと同一である．従って，我々の結論は「f の局所的な滑らかさは，各点ごとでの漸近正規性のみならず，局所的汎関数の意味での分布収束をも保証する」ということである．

なお，定理4.2の場合には $u \rightsquigarrow Z^n(u)$ は連続となるので，空間 C に値をとる確率変数に対する古典的な弱収束理論による扱いも可能である．しかしながら，正規白色ノイズモデル以外の例では $u \rightsquigarrow Z^n(u)$ は必ずしも連続とはならない．それ故，特に多次元の場合，空間 $\ell^\infty(U)$ に値をとる確率変数に対する新理論が自然なアプローチとなる．

5. 応用2：M-推定量の漸近挙動

5.1 一般的な M-推定量の収束率判定定理

まず M-推定量の収束率を導出するための一般的な判定定理を紹介する．(Θ, d) を距離空間とする．$\theta \rightsquigarrow \gamma(\theta)$ は Θ 上で定義された deterministic な関数であるとする．我々は，その最大点の位置 $\theta_0 = \mathrm{argmax}_\theta\, \gamma(\theta)$ を推定したいものとする．そのために，パラメータ空間が Θ であるような確率場 $\theta \rightsquigarrow \Gamma^n(\theta)$ が与えられたとする．後に $\theta \rightsquigarrow \Gamma^n(\theta)$ が $\theta \rightsquigarrow \gamma(\theta)$ の良い近似になっているという仮定をおくが，そのような場合には，θ_0 の自然な推定量は $\hat{\theta}_n = \mathrm{argmax}_\theta\, \Gamma^n(\theta)$ であろう．以下では $\theta \rightsquigarrow \Gamma^n(\theta)$ を criterion process，$\theta \rightsquigarrow \gamma(\theta)$ を contrast function と呼ぶ．

この枠組みの中で，例えばパラメトリック I.I.D.モデルの最尤推定量 (MLE) は次のように捉えられる．$\{f_\theta : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}$ は密度関数の族であるとし，$X_1,\ldots, X_n$ は密度 f_{θ_0} をもつ I.I.D. のデータであるとする．いま

$$\gamma(\theta) = E_{\theta_0} \log f_\theta(X_1) \qquad \forall \in \Theta$$

とおく．適当な正則条件のもとで，contrast function $\theta \rightsquigarrow \gamma(\theta)$ は，θ_0 の周りでの2項 Taylor 展開により，

$$\gamma(\theta) - \gamma(\theta_0) \approx -I(\theta_0)\,|\theta - \theta_0|^2$$

と近似的に表される．ただし $I(\theta_0)$ は Fisher 情報量を表す．従って，確かに contrast function $\theta \rightsquigarrow \gamma(\theta)$ の最大点は θ_0 となる．一方，criterion process $\theta \rightsquigarrow \Gamma^n(\theta)$ は自然に

$$\Gamma^n(\theta) = \frac{1}{n}\sum_{i=1}^n \log f_\theta(X_i) \qquad \forall\, \theta \in \Theta$$

と定義されるが，その最大点は MLE に他ならない．正則なモデルでの MLE の収束率は $n^{-1/2}$ であること，すなわち $n^{1/2}|\hat{\theta}_n - \theta_0| = O_{P_{\theta_0}}(1)$ であることはよく知られている．しかし，正則性の仮定が崩れた場合，あるいは，未知パラメータ θ_0 を定義する contrast function $\theta \rightsquigarrow \gamma(\theta)$ そのものが全く別のものである場合，収束率はどのように変化するのであろうか？ この疑問に1つの解答を与えるのが次の一般的な判定定理である．

以下，枠組みは I.I.D. を離れて，本節の第一段落で与えた一般的なものに戻る．$B_d(\theta_0, \delta)$ は θ_0

を中心とした距離 d に関する半径 δ の閉球，すなわち，$B_d(\theta_0, \delta) = \{\theta \in \Theta : d(\theta, \theta_0) \leq \delta\}$ を表すものとする．

定理 5.1. (van der Vaart and Wellner (1996)) 次の条件 "M-CRITERION" が，ある $p > 0$, $a \in (0, p)$ および $\delta_0 \in (0, \infty]$ (いずれも n に依存しない) と，ある関数 $\phi^n : (0, \delta_0) \to (0, \infty)$ であって $\delta \rightsquigarrow \delta^{-a}\phi^n(\delta)$ が単調減少なるものに対して満たされたと仮定する．

M-CRITERION. 任意の $\varepsilon > 0$ に対しある定数 c_ε, $C_\varepsilon > 0$ とある $n_\varepsilon \in \mathbb{N}$ が存在して：各 $n \geq n_\varepsilon$ に対し集合 $\tilde{\Omega}^n(\varepsilon) \subset \Omega^n$ が存在して

$$\gamma(\theta) - \gamma(\theta_0) \leq -c_\varepsilon d(\theta, \theta_0)^p \qquad \forall \theta \in B_d(\theta_0, \delta_0)$$

と

$$E^{n*} \sup_{\theta \in B_d(\theta_0, \delta)} |(\Gamma^n - \gamma)(\theta) - (\Gamma^n - \gamma)(\theta_0)| 1_{\tilde{\Omega}^n(\varepsilon)} \leq C_\varepsilon \phi^n(\delta) \qquad \forall \delta \in (0, \delta_0)$$

が成り立ち，しかも $P^{n*}(\Omega^n \backslash \tilde{\Omega}^n(\varepsilon)) \leq \varepsilon$ である．

このとき $\phi^n(r_n^{-1}) \leq r_n^{-p}$ を満足する定数 $r_n > 0$ をとれ．すると，もしも写像 $\hat{\theta}_n : \Omega^n \to \Theta$ が

$$\lim_{L\to\infty} \limsup_{n\to\infty} P^{n*}(\Gamma^n(\hat{\theta}_n) < \Gamma^n(\theta_0) - Lr_n^{-p}) = 0 \tag{5.1}$$

と

$$\lim_{n\to\infty} P^{n*}(d(\hat{\theta}_n, \theta_0) > \delta_0/2) = 0 \tag{5.2}$$

を満たすならば，

$$\lim_{L\to\infty} \limsup_{n\to\infty} P^{n*}(r_n d(\hat{\theta}_n, \theta_0) > L) = 0$$

が成り立つ．もしも M-CRITERION が $\delta_0 = \infty$ に対して満たされている場合，一致性の仮定 (5.2) は不必要である．

M-CRITERION の中のディスプレイした 2 つの不等式をそれぞれ「第一不等式」,「第二不等式」と呼ぶことにする．第一不等式は，contrast function $\theta \rightsquigarrow \gamma(\theta)$ が真値 θ_0 を他の値 θ から p 次のオーダーで分離する能力があることを意味し，第二不等式は，criterion process $\theta \rightsquigarrow \Gamma^n(\theta)$ が contrast function $\theta \rightsquigarrow \gamma(\theta)$ をどの程度の誤差で近似する能力があるかを規定している．それらを併せて収束率 r_n が関係式 $\phi^n(r_n^{-1}) \leq r_n^{-p}$ によって導出される．

上記の判定定理の $p = 2$ の場合は，特別な場合を扱った 1990 年代前半の何人かの仕事をもとにして，van der Vaart and Wellner ((1996), Theorem 3.2.5) によってほぼこの形に与えられた．さらに彼らは，γ や θ_0 も n に依存するランダムなものである場合への一般化も与えた (彼らの Theorem 3.4.1)．一方 Nishiyama (1997b) は，同定理が一般の p についても成り立つことを指摘し，さらに真値 θ_0 に関する一様性についても議論した．以下の 5.2 および 5.3 節では，敢えて一般の p に対する例のみを考察し，また簡明のため収束の真値に関する一様性は言及しない．

なお，特に Θ がユークリッド空間の部分集合である場合，収束率のみならず漸近分布をも導出できる事がしばしばある．そのための一般的な手続きは以下の通りである (詳しくは van der Vaart and Wellner (1996) の Theorem 3.2.2 を見よ)．いかなる場合においても，まず

$$\mathbb{M}^n(h) = a_n\{\Gamma^n(\theta_0 + r_n^{-1}h) - \Gamma^n(\theta_0)\}$$

というような形とした rescaled criterion process $h \rightsquigarrow \mathbb{M}^n(h)$ を考える事になる．従って，定

数列 r_n を見いだす事が最初の問題となるが，そのために定理 5.1 は有用である．定数列 a_n は r_n との関係を考慮して決定される．次に，argmax continuous mapping theorem (van der Vaart and Wellner (1996) の Theorem 3.2.2) を適用するために次の事を示す必要がある．

(A1) 局所列 $\hat{h}_n = r_n(\hat{\theta}_n - \theta_0)$ の一様緊密性．

(A2) 確率場 $h \rightsquigarrow \mathbb{M}^n(h)$ が連続確率場 $h \rightsquigarrow \mathbb{M}(h)$ へ $\ell^\infty(K)$ の中で弱収束する事．ただし K は局所パラメータの空間の任意のコンパクト部分集合．

(A3) 極限の連続確率場 $h \rightsquigarrow \mathbb{M}(h)$ のパスの最大点 $\hat{h}$ の一意的存在．

これらが示されたとき，“$r_n(\hat{\theta}_n - \theta_0) \stackrel{P}{\Longrightarrow} \hat{h}$” という形の結果が argmax continuous mapping theorem から従う．

我々がこのアプローチに関して有限次元パラメータの場合のみに甘んずる理由は，局所列 $\hat{h}_n$ の一様緊密性 (上のステップ (A1)) が“$r_n|\hat{\theta}_n - \theta_0| = O_P(1)$”と同値であり，そしてこれがまさしく定理 5.1 の主張の形であるからである．このことは，パラメータ空間 (Θ, d) が一般の場合には必ずしも真ではない．しかし，定理 5.1 は収束率の導出のためにやはり有用であり，その例を 6 節で提示する．なお 3 節の弱収束定理やその応用である 4 節の結果は上のステップ (A2) において用いられる．

5.2 例：最高点の位置の推定

4 節で考えた正規白色ノイズモデル (4.1) を考える．ここでは関数 $t \rightsquigarrow f(t)$ の最高点の位置 t_0 の推定問題を考える．採用する criterion process $\Gamma^n(\theta)$ と contrast function $\gamma(\theta)$ は，それぞれ $\Gamma(\theta) = \hat{f}_n(\theta)$ と $\gamma(\theta) = f(\theta)$ である (厳密にいえば，後者は n に依存して $\gamma^n(\theta) = \hat{f}_n(\theta)$ とすべきであるが，簡明のため細かい議論は省略する)．

最高点の近傍における f の滑らかさを考慮に入れた次の仮定をおく．

条件 5.2. ある偶数 $p \leq 2$ に対し，関数 $t \rightsquigarrow f(t)$ はある点 $t_0 \in (0,1)$ の近傍 N で p-回連続的微分可能であり，その導関数 $f^{(m)}$, $m = 1,\ldots, p$ は次を満たす：

(a) 各 $m = 1,\ldots, p-1$ に対し $f^{(m)}(t_0) = 0$;

(b) $\sup_{t \in N} f^{(p)}(t) < 0$.

推定したい未知母数は $\theta_0 = t_0$ である．この条件のもとで，p に依存して，Taylor 展開により

$$f(\theta) - f(\theta_0) \approx f^{(p)}(\theta_0)\,|\theta - \theta_0|^p$$

という近似が成り立つ．従って M-CRITERION の第一不等式は満たされる．第二不等式をチェックするために，2 節で与えた最大不等式を利用するが，その方法はカーネル関数の種類によって若干異なる．具体的に結果のみ記すと，smooth kernel の場合は $\phi^n(\delta) = n^{-1/2} b_n^{-2} \delta^{3/2}$ となり，monotone kernel の場合は $\phi^n(\delta) = n^{-1/2} b_n^{-1} \delta^{1/2}$ となる．いずれにせよ，もしも bandwidth を $b_n = n^{-1/(2p+1)}$ とおくと，関係 $\phi^n(r_n^{-1}) \leq r_n^{-p}$ は $r_n = n^{1/(2p+1)}$ とおくことによって満たされる．

次に $n^{1/(2p+1)}(\hat{\theta}_n - t_0)$ の漸近分布を導出しよう．計算の便宜上，一様カーネル $K(x) = 1/2 \cdot 1_{[-1,1]}(x)$ を使用する．ステップ (A1) はすでに示されている．ステップ (A2) に関しては，次のような分解ができる rescaled criterion function $h \rightsquigarrow \mathbb{M}^n(h)$ を考える．

$$\begin{aligned}\mathbb{M}^n(h) &= b_n^{-p}\{\hat{f}_n(t_0 + b_n h) - f(t_0)\} \\ &= Y^n(h) + Z^n(h),\end{aligned}$$

ただし

$$Y^n(h) = b_n^{-p}\{f_n(t_0+b_nh) - f(t_0)\};$$
$$Z^n(h) = b_n^{-p}\{\hat{f}_n(t_0+b_nh) - f_n(t_0+b_nh)\}.$$

$\sqrt{nb_n} = b_n^{-p}$ に注意すると，任意のコンパクト集合 $K \subset \mathbb{R}$ に対し $Z^n \xrightarrow{P} Z$ in $\ell^\infty(K)$，ただし $Z(h) = \{\mathbb{B}(h+1) - \mathbb{B}(h-1)\}/2$，となる事が，定理 4.2 から導かれる．ここに $h \rightsquigarrow \mathbb{B}(h)$ は両側ブラウン運動である．他方，簡単な計算から，各 $h \in \mathbb{R}$ に対し

$$\begin{aligned}\lim_{n\to\infty} Y^n(h) &= Y(h)\\ &= \frac{f^{(p)}(t_0)}{2p!}\int_{-1}^{1}(h+y)^p\,dy\\ &= \mathbb{A}(h)/2 \text{（この式で}\mathbb{A}(h)\text{を定義する）}\end{aligned}$$

が成り立つ事がわかる．$h \rightsquigarrow Y^n(h)$ と $h \rightsquigarrow Y(h)$ はともに連続であるからこの収束は任意のコンパクト集合K上で連続である．それ故，確率場 $h \rightsquigarrow \mathbb{M}^n(h)$ は $h \rightsquigarrow \{Y(h)+Z(h)\}$ へ $\ell^\infty(K)$ の中で弱収束する．これでステップ (A2) が示された．ステップ (A3) はいまの場合自明である．よって，$\mathrm{argmax}_h\,\mathbb{M}^n(h)$ は $\mathrm{argmax}_h\{Y(h)+Z(h)\}$ へ弱収束する．すなわち，$n^{1/(2p+1)}(\hat{\theta}_n - t_0) \xrightarrow{P} \mathrm{argmax}_h\{\mathbb{A}(h) + \mathbb{B}(h+1) - \mathbb{B}(h-1)\}$ が得られた．

5.3 例：ジャンプ点の位置の推定

再び 4 節で考えた正規白色ノイズモデル (4.1) を考える．ここでは関数 $t \rightsquigarrow f(t)$ のジャンプの位置 t_0 の推定問題を考える．t_0 の近傍における f の形状に関して次の仮定をおく．

条件 5.3. ある $t_0 \in (0,1)$ に対し，ある定数 $a \in (0, 1/2)$ が存在して関数 $t \rightsquigarrow f(t)$ は区間 $[t_0 - a, t_0 + a]$ 上で右連続かつ左極限をもち，しかも次の条件を満たす：

$$D = (R_* - L^*) - (L^* - L_*) \vee (R^* - R_*) > 0$$

ただし

$$L^* = \sup\nolimits_{t\in[t_0-a,t_0)} f(t), \qquad R^* = \sup\nolimits_{t\in[t_0,t_0+a]} f(t),$$
$$L_* = \inf\nolimits_{t\in[t_0-a,t_0)} f(t), \qquad R_* = \inf\nolimits_{t\in[t_0,t_0+a]} f(t).$$

上で現れた定数 $a > 0$ は以下で定義する推定量を構成するために既知でなければならない．しかし，我々は関数 $t \rightsquigarrow f(t)$ の具体的な形は何も規定しない．定数 $D > 0$ の値は既知である必要はない．条件 5.3 は「関数 $t \rightsquigarrow f(t)$ は t_0 において正のジャンプ，すなわち $f(t_0) - f(t_0-) \geq R_* - L^*$ をもち，そしてそれが区間 $[t_0 - a, t_0 + a]$ の中の最大のジャンプである」という意味である．この解釈により，この仮定がいかに自然であるかがわかるであろう．

パラメータ空間を $\Theta = [a, 1-a]$ とおく．アイデアは「$t \rightsquigarrow F(t) = \int_0^t f(s)\,ds$ の右側導関数と左側導関数の差が t_0 において最大になるという事実を，それら導関数のカーネル型推定量を用いて表現する」という事である．そのために，定数 $b \in (0, a)$ を固定し，

(5.3) $$w_\theta(t) = k_b(t-\theta) \qquad \forall\,\theta \in [a, 1-a],$$

ただし

$$k_b(x) = \begin{cases} -x-b, & x \in [-b, 0), \\ -x+b, & x \in [0, b], \\ 0, & \text{otherwise} \end{cases}$$

と定義する．採用する criterion process $\theta \rightsquigarrow \Gamma^n(\theta)$ と contrast function $\theta \rightsquigarrow \gamma(\theta)$ はそれぞれ

$$\Gamma^n(\theta) = \int_0^1 w_\theta(t)\, dX_t^n,$$
$$\gamma(\theta) = \int_0^1 w_\theta(t) f(t)\, dt$$

である．結論からいえば，我々は $n(\hat{\theta}^n - \theta_0) \overset{P}{\Longrightarrow} \mathrm{argmax}_{h \in \mathbb{R}} \{\mathbb{A}(h) + \mathbb{B}(h)\}$ in $\mathbb{R}$ を得ることができる．ただし $h \rightsquigarrow \mathbb{A}(h)$ は

$$\mathbb{A}(h) = \begin{cases} h\{(2b)^{-1}\int_{t_0-b}^{t_0+b} f(t)\, dt - f(t_0)\}, & \forall h \geq 0, \\ h\{(2b)^{-1}\int_{t_0-b}^{t_0+b} f(t)\, dt - f(t_0-)\}, & \forall h < 0, \end{cases}$$

によって与えられる deterministic process で，$h \rightsquigarrow \mathbb{B}(h)$ は両側ブラウン運動である．収束率の導出のポイントは，やはり次の2点である．まず，任意の $\theta \in [t_0, t_0 + a - b]$ に対し

$$\gamma(\theta) - \gamma(\theta_0) \leq -|\theta - \theta_0|\{bD - |\theta - \theta_0| R_*\}$$

であり，任意の $\theta \in [t_0 - a + b, t_0)$ に対し

$$\gamma(\theta) - \gamma(\theta_0) \leq -|\theta - \theta_0|\{bD + |\theta - \theta_0| L^*\}$$

である事から，M-CRITERION の第一不等式が $p = 1$ について示される．一方2節の最大不等式を用いて，第二不等式が $\phi^n(\delta) = n^{-1/2}\delta^{1/2}$ に対して示される．これらより収束率 $r_n = n$ が導かれる．証明の残る部分は，5.2節と同様，criterion process $\Gamma^n(\theta)$ の局所確率場の弱収束を示す事が鍵となる．詳細は省略する．

6. 応用3：ノンパラメトリック MLE の収束率

この節では，簡単のため正規白色ノイズモデルのみを考えるが，そのアプローチは，非線形な確率微分方程式のモデルや点過程のモデルにも応用が可能である．

6.1 正規白色ノイズモデルにおける一般論

Θ は $\mathcal{L}^2[0,1]$ の部分集合であるとする．各 $n \in \mathbb{N}$ に対し $t \rightsquigarrow X_t^n$ はある filtered measurable space $(\Omega^n, \mathcal{F}^n, \boldsymbol{F}^n = (\mathcal{F}_t^n)_{t \in [0,1]})$ 上で定義された連続適合過程であるとする．$\boldsymbol{P}^n = \{P_\theta^n : \theta \in \Theta\}$ は Θ によって添字づけられた $(\Omega^n, \mathcal{F}^n)$ 上の確率測度の族であるとする．X^n は P_θ^n のもとで

$$dX_t^n = \theta(t)\, dt + n^{-1/2} dB_t^{n,\theta}, \qquad X_0^n = x_0 \in \mathbb{R}$$

と表されるものとする．ただし $t \rightsquigarrow B^{n,\theta}$ は $(\Omega^n, \mathcal{F}^n, \boldsymbol{F}^n, P_\theta^n)$ 上の標準ブラウン運動とする．ある条件の下で，対数尤度比は

$$(6.1) \quad \log \frac{P_\theta^n | \mathcal{F}_1^n}{P_\vartheta^n | \mathcal{F}_1^n} = \int_0^1 \{\theta(t) - \vartheta(t)\}\, dX_t^n - \frac{1}{2}\{\|\theta\|^2_{\mathcal{L}^2[0,1]} - \|\vartheta\|^2_{\mathcal{L}^2[0,1]}\} \qquad \forall \theta, \vartheta \in \Theta$$

によって与えられる事はよく知られている (例えば Jacod and Shiryaev (1987) の Theorem III. 5.34). よって最尤推定量 (MLE) は

$$\Gamma^n(\theta) = \int_0^1 \theta(t)\, dX_t^n - \frac{1}{2}\|\theta\|^2_{\mathcal{L}^2[0,1]} \qquad \forall \theta \in \mathcal{L}^2[0, 1] \tag{6.2}$$

によって与えられる criterion process $\theta \rightsquigarrow \Gamma^n(\theta)$ の最大点として定義される ("最大点" の存在自体が明らかではないので, 少し異なる正確な定義を後述の定理の中で与える). さて, データが与えられた真値 $\theta_0 \in \Theta$ から生成されたものであったとしよう. このとき, 対応する ($P^n_{\theta_0}$ のもとでの) contrast function $\theta \rightsquigarrow \gamma(\theta)$ は

$$\begin{aligned} \gamma(\theta) &= \int_0^1 \theta(t)\, \theta_0(t)\, dt - \frac{1}{2}\|\theta\|^2_{\mathcal{L}^2[0,1]} \qquad \forall \theta \in \mathcal{L}^2[0, 1] \\ &= -\frac{1}{2}\|\theta - \theta_0\|^2_{\mathcal{L}^2[0,1]} + \frac{1}{2}\|\theta_0\|^2_{\mathcal{L}^2[0,1]} \end{aligned} \tag{6.3}$$

と定義するのが自然である. なぜなら, 残差 $\Gamma^n(\theta) - \gamma(\theta) = n^{-1/2}\int_0^1 \theta(t)\, dB_t^{n,\theta_0}$ が $P^n_{\theta_0}$ のもとでマルチンゲールだからである. 我々は (6.1) が対数尤度比を与えるという事実を使わない. 従って, $\Gamma^n(\theta)$ は単に (6.2) の右辺によって定義された確率変数であると考えられる. さらに, 我々は (6.2) において $\Gamma^n(\theta)$ を Θ の上だけではなく $\mathcal{L}^2[0, 1]$ の全ての要素について定義した. このことは, "sieved" MLE を考える事を可能にする. (6.3) と M-CRITERION の第一不等式の関係を考慮すると, L^2-semimetric $\|\cdot\|_{\mathcal{L}^2[0,1]}$ を Θ 上 (より広く, $\mathcal{L}^2[0,1]$ 上) の擬距離として採用するのが自然である.

いま, "sieve" Θ^n とは, $\mathcal{L}^2[0,1]$ の部分集合の列 (Θ の部分集合とは限らない) であって, Θ を "良く" 近似するようなものとする. 一般に, 大きな Θ^n をとれば近似は "良く" できるが, 逆に確率的な制御は困難となる. そこで,「sieve が大きすぎない」という意味のエントロピー条件と「sieve が粗すぎない」という意味の近似条件を次のように導入する.

条件 6.1. (エントロピー条件) ある関数 $\varphi: (0, \infty) \to (0, \infty)$ であって $\delta \rightsquigarrow \delta^{-1}\varphi(\delta)$ が非増加なるものが存在し, 十分大きな全ての $n \in \mathbb{N}$ に対し

$$\int_0^\delta \sqrt{\log(1 + N(\Theta^n, \|\cdot\|_{\mathcal{L}^2[0,1]}; \varepsilon))}\, d\varepsilon \le \varphi(\delta) \qquad \forall \delta \in (0, \infty)$$

が成り立つ.

これが成り立つとき, 定数 $r_n > 0$ を $n^{-1/2}\varphi(r_n^{-1}) \le r_n^{-2}$ を満たすようにとる. 近似条件はこれに対して与えられる.

条件 6.2. (近似条件) ある定数 $M > 0$ が存在して, 十分大きな全ての $n \in \mathbb{N}$ に対し

$$\Theta \subset \bigcup_{\theta \in \Theta^n} B(\theta, M r_n^{-1})$$

が成り立つ. ただし $B(\theta, \varepsilon)$ は, θ を中心とする $\|\cdot\|_{\mathcal{L}^2[0,1]}$ に関する半径 ε の閉球を表す.

定理 6.3. 条件 6.1 および 6.2 のもとで, 写像 $\widehat{\theta}_n: \Omega^n \to \Theta^n$ が

$$\Gamma^n(\widehat{\theta}_n) \ge \sup_{\theta \in \Theta^n} \Gamma^n(\theta) - r_n^{-2} \tag{6.4}$$

を満たすならば,

$$\lim_{L\to\infty}\limsup_{n\to\infty} P_{\theta_0}^{n*}(r_n\|\hat{\theta}_n-\theta_0\|_{\mathcal{L}^2[0,1]}>L)=0$$

が成り立つ.

厳密には,この定理を証明するためには,定理 5.1 のパラメータ集合 Θ を Θ^n に取り替えたものを用意する必要がある.しかし,大まかな流れは同じであるから,詳細は省略する.要約すると,第一不等式は contrast function の定義 (6.3) から従い,また第二不等式は 2 節の最大不等式がそのまま適用できる形になっている,という事である.

6.2 例 1:単調関数の族

Θ は $[0, 1]$ から $[0, 1]$ への単調関数の全体を表すものとする.すると,van der Vaart and Wellner (1996) の Theorem 2.7.5 より

$$\int_0^{\delta}\sqrt{\log N(\Theta, \|\cdot\|_{\mathcal{L}^2[0,1]}; \varepsilon)}d\varepsilon \le \text{const.}\delta^{1/2} \qquad \forall\delta>0$$

が成り立つ.このことは,sieve Θ^n を適切にとることにより,条件 6.1 は $\varphi(\delta)=\text{const.}(\delta^{1/2}\vee\delta)$ について成り立つようにするべきである,ということを示唆している.この場合,収束率は $r_n=n^{1/3}$ となり,この率に対して条件 6.2 が満たされるようにするべきである.実際,簡単な計算により,次の命題が成り立つ.

命題 6.4. 分点 $0=t_0^n<t_1^n<\cdots<t_{k_n}^n=1$ を $t_i^n-t_{i-1}^n\le n^{-2/3}$ となるように選び,また $V^n=\{j\cdot n^{-2/3}: j\in\mathbb{Z}\}\cap[0,1]$ とおく.Θ^n を $[0,1]$ から V^n への単調関数であって各区間 $[t_{i-1}^n, t_i^n)$ 上で一定値をとるものであるとする.このとき $\Theta\subset\cup_{\theta\in\Theta^n}B(\theta; \sqrt{2}n^{-1/3})$ が成り立つ.

この Θ^n の構成の場合,$\Theta^n\subset\Theta$ であるから,条件 6.1 も実際満たされている.結局,(6.4) をこのように構成した Θ^n について満たす $\hat{\theta}_n$ であれば,

$$\lim_{L\to\infty}\limsup_{n\to\infty} P_{\theta_0}^{n*}(n^{1/3}\|\hat{\theta}_n-\theta_0\|_{\mathcal{L}^2[0,1]}>L)=0$$

が成り立つ.この収束率 $n^{1/3}$ は,I.I.D.のノイズをもつ回帰関数の推定の文脈における"optimal rate"と一致している.注意すべき点は,$n^{-2/3}$ というオーダーの分点がこの収束率を得るためには十分であり,従って確率過程 $t\rightsquigarrow X_t^n$ をこれらの分点上で離散的に観測できれば,その推定量を計算する事ができる,という事である.

6.3 例 2:滑らかな関数の族

定数 $\alpha>1/2$ と $H>0$ が与えられたとする.$f:[0,1]\to\mathbb{R}$ に対し $\|f\|_\alpha$ を

$$\|f\|_\alpha=\max_{k\le\underline{\alpha}}\sup_{t\in(0,1)}|D^kf(t)|+\sup_{\substack{t,s\in(0,1)\\ t\neq s}}\frac{|D^kf(t)-D^kf(s)|}{|t-s|^{\alpha-\underline{\alpha}}}$$

と定義する.ただし $\underline{\alpha}$ は α よりも真に小さい最大の整数であり,また

$$D^kf(t)=\frac{d^kf(t)}{dt^k}$$

である.関数の族 $C_H^\alpha([0, 1])$ を $\{f:[0, 1]\to\mathbb{R}: \|f\|_\alpha\le H\}$ によって定義し,真値の族 $\Theta=$

$C_H^\alpha([0,1])$ を考える．よく知られているように，

$$\int_0^\delta \sqrt{\log N(C_H^\alpha([0,1]), \|\cdot\|_\infty; \varepsilon)}\, d\varepsilon \le \text{const.}\delta^{1-(1/2\alpha)}$$

が成り立つので，条件 6.1 は $\varphi(\delta) = \text{const.}(\delta^{1-(1/2\alpha)} \vee \delta)$ に対して成り立つべきである．この場合，収束率は $r_n = n^{\alpha/(2\alpha+1)}$ となる．

分点 $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = 1$ を，$t_i^n - t_{i-1}^n \le n^{-\alpha/(2\alpha+1)}$ を満たすように選び，写像 $\pi^n : \Theta \to \ell^\infty[0,1]$ を

$$\pi^n\theta(t) = \sum_{i=1}^{k_n} \frac{1_{[t_{i-1}^n, t_i^n)}(t)}{t_i^n - t_{i-1}^n} \int_{t_{i-1}^n}^{t_i^n} \theta(s)\, ds \qquad \forall t \in [0,1]$$

によって定義する．このとき次の 2 つの事実が成り立つ：

$$\|\pi^n\theta - \pi^n\vartheta\|_{\mathcal{L}^2[0,1]} \le \|\theta - \vartheta\|_{\mathcal{L}^2[0,1]} \le \|\theta - \vartheta\|_\infty \qquad \forall \theta, \vartheta \in \Theta;$$
$$\|\theta - \pi^n\theta\|_{\mathcal{L}^2[0,1]} \le \|\theta - \pi^n\theta\|_\infty \le Hn^{-\alpha/(2\alpha+1)} \qquad \forall \theta \in \Theta.$$

従って，もしも $\pi^n\Theta$ の可算部分集合の列 Θ^n であって $\pi^n\Theta \subset \bigcup_{\theta\in\Theta^n} B(\theta; Mn^{-\alpha/(2\alpha+1)})$ が適当な定数 $M > 0$ (n に依存しない) に対して成り立つものをとれば，条件 6.1 と 6.2 は満たされ収束率 $r_n = n^{\alpha/(2\alpha+1)}$ を得る．前の例と同様，この率は回帰分析の文脈における “optimal rate” に一致する．また，オーダー $n^{-\alpha/(2\alpha+1)}$ の分点における離散的観測が得られれば，この収束率を達成するためには十分であることに注意すべきである．

7. 応用 4：積分型推定量の漸近正規性

ここでは正規白色ノイズモデルの部分観測問題を考える．すなわち，各 $i \in \mathbb{N}$ に対し確率過程

$$dX_t^i = f(t)\, Y_t^i dt + Y_t^i dB_t^i, \qquad X_0^i = x_0^i \in \mathbb{R}$$

が与えられたとする．ここに f は deterministic 関数で，$t \rightsquigarrow Y_t^i$ は $\{0,1\}$-値の予測可能過程，$t \rightsquigarrow B_t^i$ は直交ブラウン運動であるとする．問題は，X^i と Y^i の観測から未知関数 f の汎関数 $\psi \rightsquigarrow F(\psi) = \int_0^1 \psi(t) f(t)\, dt$ を推定する事である．ただし ψ は適当な族 $\Psi \subset L^1([0,1], f(t)\, dt)$ の要素である．ここで，

$$\widetilde{Y}_t^{n,i} = \frac{Y_t^i}{\overline{Y}_t^n} \qquad \text{ただし } 0/0 = 0,$$
$$\overline{Y}_t^n = \sum_{i=1}^n Y_t^i,$$

とおき，

$$\widehat{F}^n(\psi) = \sum_{i=1}^n \int_0^1 \psi(t)\, \widetilde{Y}_t^{n,i} dX_t^i$$

によって定義される推定量 $\psi \rightsquigarrow \widehat{F}^n(\psi)$ を提案する．実際，ノイズの部分を無視すれば

$$\widehat{F}^n(\psi) \approx \sum_{i=1}^n \int_0^1 \psi(t)\, \widetilde{Y}_t^{n,i} f(t)\, Y_t^i dt = \int_0^1 \psi(t) f(t)\, 1_{\{\overline{Y}_t^n > 0\}} dt \tag{7.1}$$

が成り立つので，この推定量は自然である．いま，次のような条件を用意する．

条件 7.1. ある定数 $c_1, c_2 > 0$ と関数 $c_1 \le y(t) \le c_2$ が存在して

$$\sup_{t\in[0,1]}\left|\frac{\bar{Y}_t^n}{n} - y(t)\right| = o_P(1).$$

この条件のもとで

$$\sup_{t\in[0,1]} |1_{\{\bar{Y}_t^n>0\}} - 1| = o_P(1)$$

が成り立つので，(7.1) を思い出せば，$\hat{F}^n(\phi) \approx F(\phi)$ という近似が得られる．その残差 $n^{1/2}\cdot\{\hat{F}^n(\phi) - F(\phi)\}$ を計算するために，それと連続マルチンゲール

$$M_t^{n,\phi} = n^{1/2}\sum_{i=1}^n \int_0^t \phi(s)\ \tilde{Y}_s^{n,i}\, Y_s^i dB_s^i$$

の終点値とが集合 $\bigcap_{t\in[0,1]}\{\bar{Y}_t^n > 0\}$ 上では一致する事に注意しよう．このマルチンゲールの quadratic covariation は

$$\begin{aligned}\langle M^{n,\psi}, M^{n,\phi}\rangle_t &= n\sum_{i=1}^n\int_0^t \psi(s)\ \phi(s)\,|\tilde{Y}_s^{n,i}|^2 ds\\ &= n\int_0^t \psi(s)\ \phi(s)\frac{1_{\{\bar{Y}_s^n>0\}}}{\bar{Y}_s^n}\,ds\end{aligned}$$

と与えられるので，条件 7.1 より

$$\langle M^{n,\psi}, M^{n,\phi}\rangle_1 \xrightarrow{P} C(\psi, \phi) = \int_0^1 \frac{\psi(t)\ \phi(t)}{y(t)}\,dt \qquad \forall \psi, \phi \in \Psi$$

が容易に得られる．一方，Ψ に擬距離

$$\rho(\psi, \theta) = \sqrt{\int_0^1 |\psi(t) - \phi(t)|^2 dt}$$

を導入し，これに関し

$$\int_0^1 \sqrt{\log N(\Psi, \rho; \varepsilon)}\, d\varepsilon < \infty$$

を仮定すると，[ME] も満たされている事がわかる．従って，残差確率場 $\phi \rightsquigarrow n^{1/2}\{\hat{F}^n(\phi) - F(\phi)\}$ は平均ゼロで $EG(\psi)\ G(\phi) = C(\psi, \phi)$ なる正規確率場 $\phi \rightsquigarrow G(\phi)$ に $\ell^\infty(\Psi)$ の中で弱収束する．

さらにこれが convolution theorem (例えば van der Vaart and Wellner (1996) の Theorem 3.11.2) や asymptotic minimax theorem (同じく Theorem 3.11.5) の意味において漸近有効であることも示される．詳しい議論は Nishiyama (1998) の 4 章を参照されたい．

参 考 文 献

Bae, J. and Levental, S. (1995a). Uniform CLT for Markov chains and its invariance principle: a martingale approach, *J. Theoret. Probab.*, 8, 549-570.

Bae, J. and Levental, S. (1995b). A uniform CLT for continuous martingales, *J. Korean Statist. Soc.*, **24**, 225-231.

Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes, *Ann. Inst. H. Poincaré Probab. Statist.*, **31**, 393-427.

Dudley, R. M. (1978). Central limit theorems for empirical measures, *Ann. Probab.*, **6**, 899-929 (Correction: *ibid*. (1979). **7**, 909-911).

Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes*, Springer, Berlin, Heidelberg.

Nishiyama, Y. (1997a). Some central limit theorems for ℓ^∞-valued semimartingales and their applications, *Probab. Theory Related Fields*, **108**, 459-494.

Nishiyama, Y. (1997b). A maximal inequality for continuous martingales and M-estimation in a Gaussian white noise model, *Ann. Statist*. (to appear).

Nishiyama, Y. (1998). Entropy methods for martingales, Ph. D. Thesis, Department of Mathematics, University of Utrecht.

Ossiander, M. (1987). A central limit theorem under metric entropy with L_2 bracketing, *Ann. Probab.*, **15**, 897-919.

van de Geer, S. (1995). Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes, *Ann. Statist.*, **23**, 1779-1801.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer, New York.

Entropy Methods for Random Fields Generated by Martingales and Their Statistical Applications

Yoichi Nishiyama
(The Institute of Statistical Mathematics)

The purpose of this study is to develop entropy methods, which were first introduced for empirical processes of I.I.D. data, in order to handle some martingales with applications to statistical inference for stochastic processes.

The motivation is as follows. Since the prominent work of Dudley in 1978, the entropy methods were studied to establish laws of large numbers and central limit theorems for empirical processes indexed by classes of sets or functions in the 80's. Furthermore, some recent works have shown that the methods are useful not only for those limit theorems but also for other problems in statistics. The book by van der Vaart and Wellner in 1996 gives a nice exposition of the methods as well as a lot of applications, with emphasis on I.I.D. data. However, although some parts of the methods have a good potential to be applied also for non-I.I.D. data, no systematic study has been done in the framework of martingales, which are known to be important for analyzing a rich class of statistical models. We intend to make a step to fill this gap in the literature.

Section 1 contains an intuitive explanation about generalization of Ossiander's central limit theorem. For simplicity, the rest part of the paper is devoted only to continuous local martingales and applications to the Gaussian white noise model. Based on maximal inequalities derived in Section 2, a highlight is Section 3 that gives a weak convergence theorem. By using them, we derive the asymptotic behavior of local random fields of kernel estimators, the rate of convergence of some parametric and non-parametric M-estimators, and the asymptotic normality of integral type estimators.

Key words: Martingale, maximal inequality, central limit theorem, kernel estimator, change point, maximum likelihood estimator.