

(14)「統計科学の理論と応用の新展開」に関する研究報告

叶 雄 (総合研究大学院大学), 大西俊郎 (統計数理研究所): von Mises 分布における経験 Bayes 推定	595
大西俊郎, 柳本武美 (統数研・領域): Dual structure in the conjugate analysis of curved exponential families	597
増田弘毅 (九州大学大学院数理学研究院): Notes on Statistical Inference for Discretely Observed OU Processes driven by a Fractional Wiener Process	599
阪本雄二 (広島国際大学): Asymptotic expansion of discriminant functions for ergodic diffusions	601
二宮嘉行 (九州大学大学院数理学研究院): 変化点モデルに対する漸近理論の拡張: 独立系列から相関のある系列へ	603
金川秀也 (武蔵工業大学工学部): 従属確率変数列に対する対称統計量の極限定理について	605
藤澤洋徳 (統計数理研究所): 頑健性と効率性を同時に追求める適応的なパラメータ推定法	607
丸山祐造 (東大・空間情報), W. E. Strawderman (Rutgers 大): MSE を改善する安定したリッジ回帰推定量について	609
藤澤洋徳 (統数研), 磯村 実 (癌研究会), 江口真透 (統数研), 牛嶋 大 (癌研究会), 宮田 敏 (癌研究会), 松浦正明 (癌研究会): 先祖由来モデルと MDL 原理に基づいたハプロタイプブロック同定	611
鈴川晶夫 (北海道大学大学院経済学研究科): 競合危険モデルにおける原因別ハザード関数の推定	613
奥村英則 (中国短大・情報ビジネス), 内藤貫太 (島根大・総合理工): Nonparametric kernel regression for multinomial data	615
種市信裕 (帯広畜産大学・畜産), 関谷祐里 (北海道教育大学・教育): ロジスティック回帰分析におけるロジットの推定量の分布の近似について	617
大和 元 (鹿児島大学理学部), 戸田光一郎 (鹿児島高等予備校): A CONVEX COMBINATION of TWO-SAMPLE U -STATISTICS	619
白石 博 (早稲田大学大学院理工学研究科), 谷口正信 (早稲田大学理工学部): Statistical Estimation of Optimal Portfolios for Dependent Returns of Assets	621
蛭川潤一 (早稲田大学理工学部): Discriminant Analysis for Multivariate Non-Gaussian Locally Stationary Processes.	623
松本浩一 (九州大学大学院経済学研究科): Influence of Liquidity on the Optimal Strategy	625

Sangyeol Lee (ソウル大学), 西山陽一 (統計数理研究所), 吉田朋広 (東京 大学大学院数理科学研究科) : Test for parameter change in diffusion pro- cesses by cusum statistics based on one-step estimators	627
Nakahiro YOSHIDA (Graduate School of Mathematical Sciences, University of Tokyo) : First and higher-order asymptotic properties of estimators for sto- chastic differential equations with jumps	629

von Mises 分布における経験 Bayes 推定

総合研究大学院大学 叶 雄

統計数理研究所 大西 俊郎

1 von Mises 分布の共役性

確率変数 X は密度関数をもつ von Mises 分布に従うとする.

$$p(x; \mu, \tau) = \frac{1}{2\pi I_0(\tau)} \exp\{\tau \cos(x - \mu)\}, \quad 0 < x \leq 2\pi. \quad (1)$$

ただし, $0 < \mu < 2\pi, \tau > 0, I_0(\tau)$ は第一種修正ベッセル関数である. 位置母数 μ に対して, 次の事前密度関数を仮定する.

$$\pi(\mu; m, \delta) = \frac{1}{2\pi I_0(\delta)} \exp\{\delta \cos(\mu - m)\}. \quad (2)$$

ここで, m と δ は超母数である.

命題 1.1. (Mardia and EL-Atoum(1976)) 標本密度関数は式 (1) とし, 事前密度関数は式 (2) とすると, 事後密度は

$$\pi(\mu|x) = \frac{1}{2\pi I_0(R)} \exp\{R \cos(\mu - \hat{\mu}_{MAP})\} \quad (3)$$

に書ける. ここで, $R = \sqrt{\tau^2 + \delta^2 + 2\tau\delta \cos(x - m)}$ であり, $\hat{\mu}_{MAP}$ は事後モードである.

事前密度関数 (2) と事後密度関数 (3) は同じ分布族である. 従って, von Mises 分布 (1) は共役事前分布を持つ.

2 事後モードの最適性

損失関数として, $d(\hat{\mu}, \mu) = 1 - \cos(\hat{\mu} - \mu)$ を採用する. 事後モードの最適性を事後リスクに関するピタゴラス関係を用いて証明する.

命題 2.1. 任意の推定量 $\hat{\mu}$ とすると,

$$\int_0^{2\pi} \{d(\hat{\mu}, \mu) - d(\hat{\mu}_{MAP}, \mu)\} \pi(\mu|x) d\mu = \frac{I'_0(R)}{I_0(R)} d(\hat{\mu}_{MAP}, \hat{\mu}) \quad (4)$$

が成り立つ. 事後モード $\hat{\mu}_{MAP}$ は任意の推定量を優越する.

系 2.1. 等式 (4) の右辺は二つの分布 $\pi(\mu; \hat{\mu}_{MAP}, R)$ および $\pi(\mu; \hat{\mu}, R)$ の Kullback-Leibler 分離度に比例する. Kullback-Leibler 分離度が二つの推定量 $\mu, \hat{\mu}_{MAP}$ の事後リスクの差の R 倍である.

$$KL(\pi(\mu; \hat{\mu}_{MAP}, R), \pi(\mu; \hat{\mu}, R)) = R \times \frac{I'_0(R)}{I_0(R)} d(\hat{\mu}_{MAP}, \hat{\mu}).$$

3 経験ベイズ推定

超母数 δ を推定するときに重要な役割を果たすピタゴラス関係を示す.

命題 3.1. 標本密度関数 (1) の下で, 次式が成立する.

$$\int_0^{2\pi} \left[d(x, m) - d(x, \mu) - \frac{I'_0(\tau)}{I_0(\tau)} d(\mu, m) \right] p(x; \mu, \tau) = 0. \quad (5)$$

標本 x_1, \dots, x_p は密度関数 $p(x_i; \mu_i, \tau)$ に従う分布から抽出されたものとし, 事前密度関数として, $\pi(\mu_i; m_i, \delta)$ を仮定する. ただし, τ と m_i は既知とする.

命題 3.2. 式 (5) によって, 次の等式が成立する.

$$E_\pi E_p \left[\sum_{i=1}^p \{1 - \cos(x_i - m_i)\} \right] = p \left\{ 1 - \frac{I'_0(\tau) I'_0(\delta)}{I_0(\tau) I_0(\delta)} \right\}.$$

命題 3.2. を用いてモーメント法により δ を推定する. すなわち, 推定方程式は次のようになる.

$$\frac{1}{p} \sum_{i=1}^p \{\cos(x_i - m_i)\} = \frac{I'_0(\tau) I'_0(\hat{\delta})}{I_0(\tau) I_0(\hat{\delta})}.$$

ただし,

$$\frac{1}{p} \sum_{i=1}^p \{\cos(x_i - m_i)\} \geq \frac{I'_0(\tau)}{I_0(\tau)}$$

のとき, $\hat{\delta} = \infty$ とする.

$$\frac{1}{p} \sum_{i=1}^p \{\cos(x_i - m_i)\} \leq -\frac{I'_0(\tau)}{I_0(\tau)}$$

のとき, $\hat{\delta} = -\infty$ とする.

$\hat{\delta}$ で経験 Bayes 推定量 $\hat{\mu}_{EB}$ を計算する, $\hat{\mu}_{EB} = \hat{\mu}_{MAP}(m_1, \dots, m_p, \hat{\delta})$.

4 結果

1. 事後モード $\hat{\mu}_{MAP}$ は任意の推定量を優越する. 2. 真の値 (μ_1, \dots, μ_p) と (m_1, \dots, m_p) が近いなら, 又は極端に離れる時, 経験 Bayes 推定量のリスクが小さく, 経験 Bayes 推定量は最尤推定量を圧倒的に優越する. 真の値が 2 つのグループに分かれている場合には, m が真値の平均値 (又は $\mu + \pi$) の近くにある時, 経験 Bayes 推定量のリスクは最も小さい. 経験 Bayes 推定量が最尤推定量より劣る時には, リスクの差はそれほど大きくない. 3. 母数 τ が大きくなると, 経験 Bayes 推定量が最尤推定量に近付いていく.

参考文献

K.V. Mardia, S.A.M. EL-Atoum, Bayesian Inference for the Von Mises-Fisher Distribution, *Biometrika*, **63**, (1976) 203-206.

1. 問題設定

曲指数型分布族に属する標本密度 $p(x; \mu) = \exp\{-d(x, \mu)\} a(x)$ における共役解析を議論する。ここで、 x および μ は p 次元であり、非負関数 $d(a, t)$ は

$$d(a, t) = \sum_{j=1}^{p+1} h_j(a) \{f_j(t) - f_j(a)\} \quad (1)$$

で与えられる。関数 $f_1(a), \dots, f_{p+1}(a)$ および $h_1(a), \dots, h_{p+1}(a)$ は適当な条件を満たすものとする。指数型分布族および von Mises 分布が特殊ケースとして含まれる。

本発表の目的は、上記の標本密度における共役解析がもつ双対構造を明らかにすることである。互いに双対な性質をもつ2種類の事前分布の下でそれぞれ事後リスクに関するピタゴラス関係を導く。この等式は Bayes 推定量の最適性を示す。なお、それぞれの事前分布を仮定したときに採用する損失関数も互いに双対である。

2. 相対的な正準パラメータおよび平均パラメータ

指数型分布族の共役解析の場合と平行に議論を進めるために互いに双対な2つの新しいパラメータを導入する。関数 $h_{p+1}(\mu)$ は恒等的に正であり、行列 $\{\partial f_i(\mu)/\partial \mu_j\}_{1 \leq i, j \leq p}$ は正則であるという仮定のもとで、相対的な正準パラメータ $\eta = (\eta_1, \dots, \eta_p)^T$ を

$$\eta_j = \eta_j(\mu) = -f_j(\mu)$$

のように導入する。パラメータ μ および η は1対1対応であり、 $f_{p+1}(\mu)$ を η の関数と見て

$$\psi(\eta) = f_{p+1}(\mu)$$

とおく。これは指数型分布族ではキュムラント関数になる。関数 $\psi(\eta)$ の凸性が示されるので、ルジャンドル変換により共役なパラメータ $\theta = (\theta_1, \dots, \theta_p)^T$ および凸関数 $\phi(\theta)$ を導入する：

$$\theta_j = \theta_j(\mu) = \frac{h_j(\mu)}{h_{p+1}(\mu)} \quad \text{および} \quad \phi(\theta) = \theta^T \eta(\theta) - \psi(\eta(\theta)).$$

パラメータ θ_j は $h_j(x)$ と $h_{p+1}(x)$ の期待値の比であり、この意味で相対的な平均値パラメータというべき量である。パラメータ η と θ の1対1対応も凸関数の性質から保証される。

本発表では損失関数として $L(\hat{\mu}, \mu)$ または $L(\mu, \hat{\mu})$ を採用する。ここで、

$$L(\mu_1, \mu_2) = \phi(\theta_1) + \psi(\eta_2) - \theta_1^T \eta_2$$

であり、関数(1)を用いて $L(\mu_1, \mu_2) = d(\mu_1, \mu_2)/h_{p+1}(\mu_1)$ と書ける。

3. 事前分布の共役性および最小情報量性

相対的な正準パラメータ η に対して仮定する事前分布は次の形の密度をもつ。

$$\pi(\eta; m, \delta) \propto \exp\{-\delta d(m, \mu)\} b(\eta). \quad (2)$$

ここで、 $b(\eta)$ は(2)がプロパーになる任意の非負関数である。まず、事前密度(2)の共役性を示す。命題に現れる推定量 $\hat{\mu}_{smap}$ および $\hat{\theta}_{smap}$ は Yanagimoto and Ohnishi (2005b) が提案した標準化事後モードである。

命題 3.1 事前密度(2) は共役であり, 事後密度は $\pi(\eta; \hat{\mu}_{smap}, \rho(x, m, \delta))$ と書ける. ここで,

$$\hat{\theta}_{smap} = \frac{h_{p+1}(x)\theta(x) + \delta h_{p+1}(m)\theta(m)}{h_{p+1}(x) + \delta h_{p+1}(m)} \quad \text{かつ} \quad \hat{\theta}_{smap} = \theta(\hat{\mu}_{smap})$$

であり, $\rho(x, m, \delta) = \{h_{p+1}(x) + \delta h_{p+1}(m)\}/h_{p+1}(\hat{\mu}_{smap})$ である.

次に, 事前分布の共役性が最小情報量性を含意することを示す. 一般の事前密度を $\pi(\eta)$, それに対応する事後密度を $\pi(\eta|x)$ とする. データ x が与えられているときに,

$$E[f_j(\mu) \mid \pi(\eta|x)] = E[f_j(\mu) \mid \pi(\eta; \hat{\mu}_{smap}, \rho(x, m, \delta))] \quad \text{for } j = 1, \dots, p+1$$

を満たす事前密度 $\pi(\eta)$ の集合を \mathcal{P}_x と書く. 適当な解釈の下で, 次の命題は共役事前密度 $\pi(\eta; m, \delta)$ の最小情報量性を意味する.

命題 3.2 事前密度の集合 \mathcal{P}_x の中で汎関数 $G[\pi(\eta)] = \text{KL}(\pi(\eta|x), \pi(\eta; x, 1))$ を最小化するのは, 共役事前密度 $\pi(\eta; m, \delta)$ である.

4. 互いに双対な共役事前分布およびピタゴラス関係

2種類の共役事前分布を仮定し, それぞれの下で Bayes 推定量の最適性を示すピタゴラス関係を導く. 仮定する事前分布は, それぞれ Yanagimoto and Ohnishi (2005a) が定義した m -preserving 共役事前分布および c -preserving 共役事前分布の拡張になっている.

まず, 共役事前分布(2) において $b(\eta) \equiv 1$ の場合を考える.

$$\pi_m(\eta; m, \delta) \propto \exp\{-\delta d(m, \mu)\}. \quad (3)$$

命題 4.1 (i) 事前分布(3) は次の等式を満たす.

$$E[\theta - \theta(m) \mid \pi_m(\eta; m, \delta)] = 0.$$

(ii) 損失関数 $L(\mu, \hat{\mu})$ に対する事後リスクについて次のピタゴラス関係が成立する.

$$E[L(\mu, \hat{\mu}) - L(\mu, \hat{\mu}_{smap}) - L(\hat{\mu}_{smap}, \hat{\mu}) \mid \pi_m(\eta; \hat{\mu}_{smap}, \rho(x, m, \delta))] = 0.$$

次に, 積分 $\int \exp[-\delta_1\{\phi(\theta_0) + \psi(\eta) - \theta_0^T \eta\}] b_c(\eta) d\eta$ が δ_1 にのみ依存するような非負関数 $b_c(\eta)$ が存在する場合を考え, 次の事前分布を仮定する.

$$\pi_c(\eta; m, \delta) \propto \exp[-\delta d(m, \mu)] b_c(\eta). \quad (4)$$

命題 4.2 (i) 事前分布(4) は次の等式を満たす.

$$E[\eta - \eta(m) \mid \pi_c(\eta; m, \delta)] = 0.$$

(ii) 損失関数 $L(\hat{\mu}, \mu)$ に対する事後リスクについて次のピタゴラス関係が成り立つ.

$$E[L(\hat{\mu}, \mu) - L(\hat{\mu}_{smap}, \mu) - L(\hat{\mu}, \hat{\mu}_{smap}) \mid \pi_c(\eta; \hat{\mu}_{smap}, \rho(x, m, \delta))] = 0.$$

参考文献

- YANAGIMOTO, T. AND OHNISHI, T. (2005a). Extensions of a conjugate prior through the Kullback-Leibler separators, *J. Multivariate Anal.*, **92**, 116–133.
- YANAGIMOTO, T. AND OHNISHI, T. (2005b). Standardized posterior mode for the flexible use of a conjugate prior, *J. Statist. Plann. Inference*, In press.

Notes on Statistical Inference for Discretely Observed OU Processes driven by a Fractional Wiener Process

九州大学大学院数理学研究院 増田弘毅

1. 概要

フラクショナルウィーナー過程で駆動されるオルンシュタイン-ウーレンベック過程について、離散観測に基づいて未知母数の推定を考える。ガウス性により、データに対応する尤度を正確に書き下すことは可能であるが、その表示はかなり複雑なものとなる。ここではエルゴード的な場合を対象とし、拡散係数(定数)と Hurst 指数 $H \in (1/2, 1)$ は既知であるとして、ドリフト項に入る未知母数に関する簡単な積率型推定量を考える。強一致性が導出される。

2. マルコフ型ガウス OU 過程

本講で対象とするのは、“一次元フラクショナルウィーナー過程で駆動されるオルンシュタイン-ウーレンベック過程”(略して FOU 過程)である。その前にマルコフ型ガウス OU 過程について簡単に触れる。

$X^{1/2} = (X_t^{1/2})_{t \in \mathbf{R}_+}$ で一次元ランジュバン積分方程式

$$X_t^{1/2} = X_0^{1/2} + \int_0^t \lambda(\gamma - X_s^{1/2})ds + \sqrt{\sigma}w_t^{1/2} \quad (1)$$

の解を表す。ここで w は一次元標準ウィーナー過程、 $\gamma \in \mathbf{R}$ および $\lambda, \sigma > 0$ は未知母数である。これは一次元拡散過程の中で極めて特殊なものであるといえる。この確率過程は連続時間のガウス型自己回帰過程であり、解は

$$X_t^{1/2} = e^{-\lambda t} X_0^{1/2} + \gamma(1 - e^{-\lambda t}) + \sqrt{\sigma} \int_0^t e^{-\lambda(t-s)} dw_s^{1/2}, \quad t \in \mathbf{R}_+, \quad (2)$$

と陽に与えられる。 $X^{1/2}$ は不変分布として $N_1(\gamma, \sigma/(2\lambda))$ を持つ。

一方で、これは例えば平均回帰平方根拡散過程など、応用上よく現れる大半のエルゴード的な拡散過程についていえることであるが、マルコフ型であるが故、その自己相関構造が非常に制限されたものになってしまう(減衰が早い、即ち弱相関過程)という欠点がある: 実際、 $X^{1/2}$ の自己相関関数は $r^{1/2}(t) = \exp(-\lambda t)$ で与えられる。これは(2)において $w^{1/2}$ をより一般のレヴィ過程に取り替えたバージョンについても成り立つ事実である。ここで、独立なマルコフ型 OU 過程の重ね合わせによって、より減衰の遅い自己相関構造を入れることも可能であることに注意しておく。

3. フラクショナル OU 過程

まず、長期記憶型の一次元フラクショナルウィーナー過程の定義を思い出しておく。 $H \in (1/2, 1)$ に対し、ある確率過程 $w^H = (w_t^H)_{t \in \mathbf{R}}$ がフラクショナルウィーナー過程(FWP)であるとは、それが平均零のガウス過程であり、共分散構造が

$$\text{Cov}[w_s^H, w_t^H] = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}), \quad s, t \in \mathbf{R}, \quad (3)$$

で与えられることであった。この場合 w^H はウィーナー過程と違って独立増分性を持たず、セミマルチンゲールの枠組みから外れる(但し、“fundamental martingale”と呼ばれるガウス型マルチンゲールを対応付けることが可能である)。FWP については、ウィーナー積分による様々な表現が知られている。

FOU 過程は, (2) において $w^{1/2}$ を w^H に取り替えたものとして定義される:

$$X_t^H = e^{-\lambda t} X_0^H + \gamma(1 - e^{-\lambda t}) + \sqrt{\sigma} \int_0^t e^{-\lambda(t-s)} dw_s^H, \quad t \in \mathbf{R}_+. \quad (4)$$

フラクショナルウィーナー積分の共分散公式によって X^H の不変分布を実際に求めることができ, それは $N_1(\gamma, \sigma H \Gamma(2H) \lambda^{-2H})$ で与えられる; 形式的に $H = 1/2$ とすればマルコフ型の場合になることに注意. また, 自己相関関数 $r^H(t)$ は, 任意の $N \in \mathbf{N}$ に対して $t \rightarrow \infty$

$$r^H(t) = \frac{\sigma}{2} \sum_{k=1}^N \lambda^{-2k} \left(\prod_{l=0}^{2k-1} (2H - l) \right) t^{2H-2k} + O(t^{2H-2N-2}) \quad (5)$$

で与えられる. 従って X^H は長期記憶性を持ち, 更に X^H はエルゴード性を持つ.

4. FOUに関する統計推測について: 積率推定量の強一致性

(4) で与えられる X^H をモデルとして想定し, 離散的な観測 $(X_{t_i^n})_{i=1}^n$ に基づいて $\theta = (\gamma, \lambda) \in \mathbf{R} \times (0, \infty)$ の簡単な積率型推定を行う. ここで $(t_i^n)_{i=0}^n$ は $0 = t_0^n < t_1^n < \dots < t_n^n$ で $n \rightarrow \infty$ のとき $t_i^n - t_{i-1}^n = h_n \rightarrow 0$ かつ $t_n^n = nh_n \rightarrow \infty$ となる時点であり, また $(\sigma, H) \in (0, \infty) \times (1/2, 1)$ は既知とする.

連続観測の場合, 即ち観測値が $(X_t^H)_{0 \leq t \leq T}$ である場合の推測についてはいくつかの既存の結果がある. それらの殆どが最近 (2000 年以降) のものである.

定常分布は $N_1(\gamma, \sigma H \Gamma(2H) / \lambda^{2H})$ であったことを思い出す. ここで扱うのは最も簡単な積率推定量である. 正確な尤度を書き下すのは可能であるが, 非マルコフ性に乗じてその表現は複雑である. MLE の考察に先駆け, 計算し易い推定量を確保することは重要と言える.

$q \in \mathbf{N}$ と多項式 f について成り立つ評価

$$\left\| \frac{1}{n} \sum_{i=1}^n f(X_{t_{i-1}^n}) - \frac{1}{nh_n} \int_0^{nh_n} f(X_s) ds \right\|_{L^q(P_0)}^q \lesssim (1 - e^{-\lambda h_n})^q + O(h_n^{Hq}) \quad (6)$$

とエルゴード定理を介して, 以下が成り立つ:

次で与えられる積率推定量 $\hat{\theta}_n = (\hat{\gamma}_n, \hat{\lambda}_n)^\top$ は強一致性を持つ.

$$\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^n X_{t_{i-1}^n}, \quad (7)$$

$$\hat{\lambda}_n = \frac{[\sigma H \Gamma(2H)]^{1/(2H)}}{\left\{ \frac{1}{n} \sum_{i=1}^n X_{t_{i-1}^n}^2 - \left(\frac{1}{n} \sum_{i=1}^n X_{t_{i-1}^n} \right)^2 \right\}^{1/(2H)}} \quad (8)$$

もし $(\gamma, \lambda, \sigma, H)$ の全てが未知母数であるとする, ガウス性から, 積率推定では平均と分散の推定しかできないという不都合が生じてしまう点に注意する. また, 推定量の極限分布に関する結果はまだ出せていないが, これが正規分布の場合, 時系列の場合とどう異なるのか非常に興味深い.

Asymptotic expansion of discriminant functions for ergodic diffusions

広島国際大学 阪本 雄二

Given $\Theta \subset \mathbb{R}^p$, let θ_1 and θ_2 be two p -dimensional vectors in Θ . For each $k = 1, 2$, let Π_k denote the total of sample paths of the d -dimensional stationary diffusion process $X^{(k)} = (X_t^{(k)})_{t \in [0, T_k]}$ satisfying $dX_t^{(k)} = V_0(X_t^{(k)}, \theta_k)dt + V(X_t^{(k)})dw_t^{(k)}$, $X_0^{(k)} \sim \nu_{\theta_k}$, where T_k is a positive constant, ν_{θ_k} is the stationary distribution with a positive density $\partial \nu_{\theta_k} / dx$, V_0 is a function from $\mathbb{R}^d \times \mathbb{R}^p$ to \mathbb{R}^d , V is a function from \mathbb{R}^d to $\mathbb{R}^d \otimes \mathbb{R}^r$, $w^{(k)}$, $k = 1, 2$, are r -dimensional standard Wiener processes, and they are independent of each other. In this article, we will consider the discrimination problem of how to decide from which of populations Π_1 and Π_2 new observations $X^{(0)} = (X_t^{(0)})_{t \in [0, T_0]}$ is drawn.

In the case where θ_1 and θ_2 are known, we may use the Bayes rule: $X^{(0)} = (X_t^{(0)})_{t \in [0, T_0]}$ is classified to Π_1 if $d_B(\theta_1, \theta_2) > K$ and otherwise to Π_2 , where d_B is the Bayes discriminant function defined by $d_B(\theta_1, \theta_2) = \ell(\theta_1; (X_t^{(0)})_{t \in [0, T_0]}) - \ell(\theta_2; (X_t^{(0)})_{t \in [0, T_0]})$ and ℓ is the log-likelihood function. Here K is a constant dependent of two kinds of misdiscrimination costs $c(1|2)$ and $c(2|1)$, and the prior probabilities π_k , $k = 1, 2$ such that new observations $X^{(0)}$ is drawn from Π_k with probability π_k .

In the case where θ_1 and θ_2 are unknown, we need training samples $X^{(1)} = (X_t^{(1)})_{t \in [0, T_1]}$ and $X^{(2)} = (X_t^{(2)})_{t \in [0, T_2]}$ drawn from Π_1 and Π_2 , respectively. A natural extension of the Bayes rule is the so-called W -rule such that $X^{(0)}$ is classified to Π_1 if $d_B(\hat{\theta}_1^{(1)}, \hat{\theta}_2^{(2)}) > K$, and otherwise to Π_2 . Here, $\hat{\theta}_k^{(k)}$ is the maximum likelihood estimator defined as a solution of $(\delta_a \ell(\hat{\theta}_k^{(k)}; (X_t^{(k)})_{t \in [0, T_k]}))_{a=1}^p = 0$, δ_a is the differential operator defined by $\delta_a = \partial / \partial \theta^a$, and θ^a is the a -th coordinate of θ . We abbreviate a plug-in version $d_B(\hat{\theta}_1^{(1)}, \hat{\theta}_2^{(2)})$ to d_W . Unlike the Bayes rule in the known parameter case, the optimality of the W -rule is not clear.

Let $A(x, \theta) = \log \frac{d\nu_\theta}{dx}(x)$, $B(x, \theta) = V_0'(VV')^{-1}V(x, \theta)$ and $C^\vartheta(x, \theta) = B(\theta, x)(B'(\vartheta, x) - \frac{1}{2}B'(\theta, x))$. For any function f of θ , $\Delta f(\theta_1, \theta_2)$ denote the difference $f(\theta_1) - f(\theta_2)$, and put $\nu_\theta(f) = \int_{\mathbb{R}^d} f(x) \nu_\theta(dx)$ if f is integrable w.r.t. ν_θ . Let $\tilde{A}^\vartheta(x, \theta) = A(x, \theta) - \nu_\theta(A(\cdot, \theta))$, Therefore $\tilde{d}_B(\theta_1, \theta_2)$ admits an Edgeworth expansion under some regularity conditions. For any measurable function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, let $\langle f \rangle^\theta$ be a function such that $\mathcal{A}^\theta \langle f \rangle^\theta = f - \nu_\theta(f)$, and $[f]^\theta = -\nabla \langle f \rangle^\theta V$, where ∇ is the gradient operator with ∇f being a row vector, \mathcal{A}^θ is the differential operator defined by

$$\mathcal{A}^\theta = \sum_{i=1}^d V_0^i(x, \theta) \frac{\partial}{\partial x^i} + \frac{1}{2} \sum_{i,j} \sum_{l=1}^r V_l^i(x) V_l^j(x) \frac{\partial^2}{\partial x^i \partial x^j}.$$

Put $\tilde{B}^\vartheta(x, \theta) = B(x, \theta) + [C^\vartheta]^\vartheta(x, \theta)$, $\mathbf{A}^\vartheta(\theta) = \nu_\theta(A(\cdot, \theta))$,

$$\Sigma_B^{\theta_0}(\theta_1, \theta_2) = \nu_{\theta_0} \left(\Delta \tilde{B}^{\theta_0}(\cdot, \theta_1, \theta_2) \cdot \Delta \tilde{B}^{\theta_0}(\cdot, \theta_1, \theta_2) \right),$$

$$\kappa_{B,3}^{\theta_0}(\theta_1, \theta_2) = 3\nu_{\theta_0} \left(\left[\Delta \tilde{B}^{\theta_0}(\cdot, \theta_1, \theta_2) \cdot \Delta \tilde{B}^{\theta_0}(\cdot, \theta_1, \theta_2) \right]^{\theta_0} \cdot \Delta \tilde{B}^{\theta_0}(\cdot, \theta_1, \theta_2) \right).$$

Therefore, the normalized version $\tilde{d}_B(\theta_1, \theta_2)$ of the Bayes discriminant function $d_B(\theta_1, \theta_2)$ is given by $\tilde{d}_B(\theta_1, \theta_2) = \sqrt{T_0} \left(\frac{1}{T_0} d_B(\theta_1, \theta_2) - \Delta \mathbf{C}^{\theta_0}(\theta_1, \theta_2) \right)$ with $\mathbf{C}^\vartheta(\theta) = \nu_\theta(C^\vartheta(\cdot, \theta))$ admits the following asymptotic expansion:

Theorem 0.1. *Under some regularity conditions on V_0 and V , for any measurable function f ,*

$$E[f(\tilde{d}_B(\theta_1, \theta_2))] = \int f(z) \phi(z; \Sigma_B^{\theta_0}(\theta_1, \theta_2)) \left(1 + \frac{1}{\sqrt{T_0}} q_{B,2}^{\theta_0}(z; \theta_1, \theta_2) \right) dz + o(T_0^{-1/2}),$$

where $\phi(z; \sigma)$ is the normal density with mean 0 and covariance σ , Hermite polynomials h_1, h_3 are defined by $h_1(z; \sigma) = \sigma^{-1}z$, $h_3(z; \sigma) = \sigma^{-3}(z^3 - 3\sigma z)$, and

$$q_{B,2}^{\theta_0}(z; \theta_1, \theta_2) = \Delta \mathbf{A}^{\theta_0}(\theta_1, \theta_2) h_1(z; \Sigma_B^{\theta_0}(\theta_1, \theta_2)) + \frac{1}{6} \kappa_{B,3}^{\theta_0}(\theta_1, \theta_2) h_3(z; \Sigma_B^{\theta_0}(\theta_1, \theta_2))$$

Let $\delta_a = \partial/\partial\theta^a$, and denote the partial derivatives $\delta_{a_1} \cdots \delta_{a_n} A(\cdot, \theta)$, $\delta_{a_1} \cdots \delta_{a_n} B(\cdot, \theta)$, and $\delta_{a_1} \cdots \delta_{a_n} C(\cdot, \theta)$ by $A_{a_1 \cdots a_n}(\cdot, \theta)$, $B_{a_1 \cdots a_n}(\cdot, \theta)$, and $C_{a_1 \cdots a_n}(\cdot, \theta)$, respectively. Let $\tilde{\nu}_{a_1 \cdots a_n}(\theta; \vartheta) = \nu_{\vartheta}(C_{a_1 \cdots a_n}^{\vartheta}(\cdot, \theta))$, $\tilde{\kappa}_{a_1 \cdots a_n}(\theta) = \tilde{\nu}_{a_1 \cdots a_n}(\theta; \theta)$, $\tilde{A}_{a_1 \cdots a_n}^{\vartheta}(x, \theta) = A_{a_1 \cdots a_n}(x, \theta) - \nu_{\vartheta}(A_{a_1 \cdots a_n}(\cdot, \theta))$, $\mathbf{A}_{a_1 \cdots a_n}^{\vartheta}(\theta) = \nu_{\vartheta}(A_{a_1 \cdots a_n}(\cdot, \theta))$, $\mathbf{A}_a(\theta) = \mathbf{A}_a^{\theta}(\theta)$, $\mathbf{B}_{a_1 \cdots a_n, b_1 \cdots b_m}^{\vartheta}(\theta, \theta') = \nu_{\vartheta}(\tilde{B}_{a_1 \cdots a_n}^{\vartheta}(\cdot, \theta) \cdot \tilde{B}_{b_1 \cdots b_m}^{\vartheta}(\cdot, \theta'))$ and $\mathbf{B}_{a_1 \cdots a_n, b_1 \cdots b_m}^{\theta}(\theta) = \mathbf{B}_{a_1 \cdots a_n, b_1 \cdots b_m}^{\theta}(\theta, \theta)$. Denote by $(\tilde{\kappa}^{ab}(\theta))$ the inverse matrices of $(\tilde{\kappa}_{ab}(\theta))$ if it exists. Put $\tilde{\kappa}_{bc}^{aa'}(\theta) = -\tilde{\kappa}^{aa'}(\theta) \tilde{\kappa}_{a'bc}(\theta)$, $\mathbf{A}^a(\theta) = -\tilde{\kappa}^{aa'}(\theta) \mathbf{A}_{a'}(\theta)$, $\mathbf{B}^{a,b}(\theta) = \tilde{\kappa}^{aa'}(\theta) \tilde{\kappa}^{bb'} \mathbf{B}_{a',b'}(\theta)$, and $\mathbf{B}_{b,c}^{a,b}(\theta) = \tilde{\kappa}^{aa'}(\theta) \tilde{\kappa}^{cc'} \mathbf{B}_{a',b,c'}(\theta)$. Suppose that T_k is proportional to T_0 : $T_k = \tau_k^2 T_0$ for some constant τ_k . Put $\mathbf{c}_a^{\theta_0}(\theta_k) = \tilde{\nu}_a(\theta_0; \theta_k)$, $\mathbf{c}_{ab}^{\theta_0}(\theta_k) = \frac{1}{2} \left(\tilde{\nu}_{ab}(\theta_k; \theta_0) + \tilde{\nu}_{ba}(\theta_k; \theta_0) \right)$, $\mathbf{c}^{\theta_0}(\theta_1, \theta_2) = \Delta \mathbf{A}^{\theta_0}(\theta_1, \theta_2) - \sum_{k=1}^2 \frac{(-1)^k}{\tau_k^2} \tilde{\nu}_a(\theta_k; \theta_0) \mathbf{A}^a(\theta_k)$,

$$\mathbf{B}^{[a,b],c}(\theta_k) = \tilde{\kappa}^{aa'}(\theta_k) \tilde{\kappa}^{bb'}(\theta_k) \tilde{\kappa}^{cc'}(\theta_k) \nu_{\theta_k} \left(\left[\tilde{B}_{a'}^{\theta_k}(\cdot, \theta_k) \cdot \tilde{B}_{b'}^{\theta_k}(\cdot, \theta_k) \right]^{\theta_k} \cdot \tilde{B}_{c'}^{\theta_k}(\cdot, \theta_k) \right),$$

$$\Sigma_W^{\theta_0}(\theta_1, \theta_2) = \Sigma_B^{\theta_0}(\theta_1, \theta_2) + \sum_{k=1}^2 \frac{1}{\tau_k^4} \tilde{\nu}_a(\theta_k; \theta_0) \tilde{\nu}_b(\theta_k; \theta_0) \mathbf{B}^{a,b}(\theta_k),$$

$$\kappa_{W,3}^{\theta_0}(\theta_1, \theta_2) = \kappa_{B,3}^{\theta_0}(\theta_1, \theta_2) - \sum_{k=1}^2 \frac{3(-1)^k}{\tau_k^4} \tilde{\nu}_a(\theta_k; \theta_0) \tilde{\nu}_b(\theta_k; \theta_0) \tilde{\nu}_c(\theta_k; \theta_0) \mathbf{B}^{[a,b],c}(\theta_k).$$

Then $\tilde{d}_W(\theta_1, \theta_2) = \sqrt{T_0} \left(\frac{1}{T_0} d_B(\hat{\theta}_1, \hat{\theta}_2) - \nu_{\theta_0}(\Delta C^{\theta_0}(\cdot, \theta_1, \theta_2)) \right)$ admits the following asymptotic expansion:

Theorem 0.2. Under some regularity conditions on V_0 and V , for any measurable function f ,

$$E[f(\tilde{d}_W(\theta_1, \theta_2))] = \int f(z) \phi(z; \Sigma_W^{\theta_0}(\theta_1, \theta_2)) \left(1 + \frac{1}{\sqrt{T_0}} q_{W,2}^{\theta_0}(z; \theta_1, \theta_2) \right) dz + o\left(\frac{1}{\sqrt{T_0}}\right),$$

where $q_{W,2}^{\theta_0}(z; \theta_1, \theta_2) = Q_{W,1}^{\theta_0}(\theta_1, \theta_2) h_1(z; \Sigma_W^{\theta_0}(\theta_1, \theta_2)) + Q_{W,3}^{\theta_0}(\theta_1, \theta_2) h_3(z; \Sigma_W^{\theta_0}(\theta_1, \theta_2))$,

$$Q_{W,1}^{\theta_0}(\theta_1, \theta_2) = \mathbf{c}^{\theta_0}(\theta_1, \theta_2) - \sum_{k=1}^2 \frac{(-1)^k}{\tau_k^2} \left(\mathbf{c}_a^{\theta_0}(\theta_k) \mathbf{B}_{b,c}^{a,b}(\theta_k) + \mathbf{c}_{ab}^{\theta_0}(\theta_k) \mathbf{B}^{a,b}(\theta_k) \right),$$

$$\begin{aligned} Q_{W,3}^{\theta_0}(\theta_1, \theta_2) = & \frac{1}{6} \kappa_{W,3}^{[0],\theta_0}(\theta_1, \theta_2) + \sum_{k=1}^2 \frac{1}{\tau_k^2} \tilde{\nu}_b(\theta_k; \theta_0) \left(\mathbf{B}_{\phi,a}^{\theta_0}(\theta_1, \theta_k) - \mathbf{B}_{\phi,a}^{\theta_0}(\theta_2, \theta_k) \right) \mathbf{B}^{a,b}(\theta_k) \\ & - \sum_{k=1}^2 \frac{(-1)^k}{\tau_k^4} \left(\mathbf{c}_a^{\theta_0}(\theta_k) \tilde{\nu}_c(\theta_k; \theta_0) \mathbf{B}_{b,c}^{a,b}(\theta_k) \tilde{\nu}_{b'}(\theta_k; \theta_0) \mathbf{B}^{b,b'}(\theta_k) \right. \\ & \left. + \mathbf{c}_{ab}^{\theta_0}(\theta_k) \tilde{\nu}_{a'}(\theta_k; \theta_0) \mathbf{B}^{a,a'}(\theta_k) \tilde{\nu}_{b'}(\theta_k; \theta_0) \mathbf{B}^{b,b}(\theta_k) \right). \end{aligned}$$

References

- [1] Kusuoka, S. and Yoshida, N.(2000). *Probability Theory and Related Fields*, **116**, 457-484.
- [2] Kutoyants, Yu. A.: London: Springer 2004
- [3] Sakamoto, Y. and Yoshida, N. (2004). *Annals of the Institute of Statistical Mathematics*, **56-3**, 545-597
- [4] Uchida, M. and Yoshida, N.(2001). *Statistical Inference for Stochastic Processes*, **4**, 73-98.
- [5] Uchida, M. and Yoshida, N.(2004). *Statistical Inference for Stochastic Processes*, **7**, 35-67.

変化点モデルに対する漸近理論の拡張：独立系列から相関のある系列へ

九州大学大学院数理学研究院 二宮 嘉行

1. 序

変化点解析のための統計理論は、1950年代から現在に至るまで、長い間研究されている。その理由は、この解析が応用上重要であることだけでなく、変化点パラメータがある種の非正則性をもつという理論的興味にもある。例えば、一般に変化点モデルの尤度関数は変化点パラメータで微分できないため、変化点モデル固有の漸近理論が必要となり、研究が進んでいる。ここでは、変化点解析の需要が高い計量経済への応用を考慮し、自己回帰過程における複数変化点モデルを扱う。

2. 変化点推定量の漸近的性質

最も単純なモデルを考える。 ϵ_t は $N(0, \sigma^2)$ に従う白色雑音とし、 X_1, \dots, X_n は

$$\begin{aligned} X_t &= a^{(1)}X_{t-1} + \epsilon_t, & (2 \leq t \leq k^*) \\ &= a^{(2)}X_{t-1} + \epsilon_t, & (k^* + 1 \leq t \leq n) \end{aligned} \quad (1)$$

なるモデルに従うとする (X_1 は適当)。ここで、 $a^{(1)}$, $a^{(2)}$, σ^2 , k^* は未知であり、漸近論 ($n \rightarrow \infty$) を考える際には、 λ を $0 < \lambda < 1$ なる定数として $k^* = [n\lambda]$ を仮定する。変化点を k としたときの (条件付) 最大対数尤度を $T(k)$, 両側に伸びる負のドリフト付ランダムウォークを

$$\begin{aligned} Q_k \equiv & I_{k < k^*} \sum_{t=k+1}^{k^*} \{2X_{t-1}(a^{(2)} - a^{(1)})\epsilon_t - (a^{(2)} - a^{(1)})^2 X_{t-1}^2\} / \sigma^2 \\ & + I_{k > k^*} \sum_{t=k^*+1}^k \{2X_{t-1}(a^{(1)} - a^{(2)})\epsilon_t - (a^{(2)} - a^{(1)})^2 X_{t-1}^2\} / \sigma^2 \end{aligned} \quad (2)$$

と定義すれば、 $k - k^* = O(1)$ ならば $2\{T(k) - T(k^*)\} - Q_k = o_P(1)$ がいえる。 $|k - k^*| \rightarrow \infty$ ならば $Q_k \rightarrow -\infty$ となることから想像できるように、 $|k - k^*| \rightarrow \infty$ ならば $2\{T(k) - T(k^*)\} \rightarrow -\infty$ となる。以上より、最尤推定量 \hat{k} に関して

$$\hat{k} - k^* = O_P(1) \quad \text{かつ} \quad \hat{k} - \arg\max_k Q_k = o_P(1) \quad (3)$$

が成り立つ。これより \hat{k} は一貫性をもつといわれる。また、漸近正規性は満たされないこともわかる。さらに

$$\max_k 2\{T(k) - T(k^*)\} - \max_k Q_k = o_P(1) \quad (4)$$

がいえていることもわかる。

3. 尤度比検定統計量の漸近的性質

この節では変化点数 1 対 2 の尤度比検定の漸近論を扱う。そのために変化点数 2 のモデルを

$$\begin{aligned} X_t &= a^{(1)}X_{t-1} + \epsilon_t, & (2 \leq t \leq k_1^*) \\ &= a^{(2)}X_{t-1} + \epsilon_t, & (k_1^* + 1 \leq t \leq k_2^*) \\ &= a^{(3)}X_{t-1} + \epsilon_t, & (k_2^* + 1 \leq t \leq n) \end{aligned} \quad (5)$$

と用意し、変化点を k_1, k_2 としたときの、(条件付) 最大対数尤度を $T_2(k_1, k_2)$ と記すことにする。

補題 1 モデル (1) が真であるという仮説のもとでモデル (5) を考えると,

$$|\hat{k}_1 - k^*| = O_P(1) \quad \text{あるいは} \quad |\hat{k}_2 - k^*| = O_P(1)$$

証明は (3) の導出を拡張して考えることによって得られる. この補題の結果を用いて $\max_{k_1 < k_2} T_2(k_1, k_2) - T(k^*)$ の漸近分布に関する補題を与える.

補題 2 モデル (1) が真であるという仮説のもと,

$$\begin{aligned} & \left| 2 \left\{ \max_{k_1 < k_2} T_2(k_1, k_2) - T(k^*) \right\} \right. \\ & \quad \left. - \left(\max \left[\sup_{1/n \leq t \leq 1-1/n} \left\{ \frac{B_{1,n}(t)^2}{t(1-t)} \right\}, \sup_{1/n \leq t \leq 1-1/n} \left\{ \frac{B_{2,n}(t)^2}{t(1-t)} \right\} \right] + \max_k Q_k \right) \right| \\ & = o_P(1), \end{aligned}$$

$$B_{j,n}(t) \stackrel{\text{dist.}}{=} B(t) \quad (j = 1, 2),$$

なる確率過程の独立な列 $\{B_{j,n}(t)\} (j = 1, 2)$ が存在する.

この補題 2 と (4) を組み合わせると, 変化点数 1 対 2 の尤度比検定統計量 $\max_{k_1 < k_2} T_2(k_1, k_2) - \max_k T(k)$ について以下の結果を得る.

定理 1 モデル (1) が真であるという仮説のもと,

$$\begin{aligned} & \left| \left[2 \left\{ \max_{k_1 < k_2} T_2(k_1, k_2) - \max_k T(k) \right\} \right]^{1/2} \right. \\ & \quad \left. - \max \left[\sup_{1/n \leq t \leq 1-1/n} \left\{ \frac{B_{1,n}(t)^2}{t(1-t)} \right\}^{1/2}, \sup_{1/n \leq t \leq 1-1/n} \left\{ \frac{B_{2,n}(t)^2}{t(1-t)} \right\}^{1/2} \right] \right| \\ & = o_P(1), \end{aligned}$$

$$B_{j,n}(t) \stackrel{\text{dist.}}{=} B(t) \quad (j = 1, 2),$$

なる確率過程の独立な列 $\{B_{j,n}(t)\} (j = 1, 2)$ が存在する.

これは自明に変化点数 m 対 $m+1$ の尤度比検定の場合に拡張でき, 収束のオーダーも精確に議論すれば以下が得られる.

命題 1 $h_j(n) \geq 1/n, l_j(n) \geq 1/n (j = 1, \dots, m+1)$ がある $0 < \epsilon^* \leq 1$ に対して

$$\limsup_{n \rightarrow \infty} n \{h_j(n) + l_j(n)\} \exp\{-(\log n)^{1-\epsilon^*}\} < \infty$$

を満たすとする. 変化点数 m のモデルが真であるという仮説のもと, 変化点数 m 対 $m+1$ の対数尤度比を $T_{m:m+1}$ と記せば,

$$\left| (2 T_{m:m+1})^{1/2} - \max_{1 \leq j \leq m+1} \left[\sup_{h_j(n) \leq t \leq 1-l_j(n)} \left\{ \frac{B_j(t)^2}{t(1-t)} \right\}^{1/2} \right] \right| = o_P[\exp\{-(\log n)^{1-\epsilon}\}],$$

$$B_{j,n}(t) \stackrel{\text{dist.}}{=} B(t) \quad (j = 1, \dots, m+1),$$

なる確率過程の独立な列 $\{B_{j,n}(t)\} (j = 1, \dots, m+1)$ が存在する.

従属確率変数列に対する対称統計量の極限定理について

武蔵工業大学工学部 金川 秀也

1 序

$\{\xi_j, j \geq 1\}$ を分布 μ に従う実数値確率変数列とする。また $u(x_1, x_2, \dots, x_m)$ を実対称関数とする。 $u(x_1, x_2, \dots, x_m)$ を核関数とする統計量を対称統計量と呼ぶ。本講演では特に核関数 u が退化型で確率変数列（標本）が従属性を持つ場合に、対称統計量の一種である U -統計量及び V -統計量の漸近的な性質について解説する。

U -統計量の U は unbiased、また V -統計量の V は von Mises の頭文字を取ったもので推定や検定、特にノンパラメトリックな場合に有効である。最近、このような統計量について詳しく解説された初めての邦文のテキストとして前園宣彦著「統計的推論の漸近理論」が出版された。詳しくはこちらを参照されたい。

例 1.1 U -統計量 (degree m) $U_n := \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < i_2 < \dots < i_m \leq n} u(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_m})$

標本平均 : $u(x) = x \quad (k=1)$

不偏分散 : $u(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2 \quad (k=2)$

例 1.2 V -統計量 (degree m) $V_n := n^{-m} \sum_{1 \leq i_1 \leq i_2 \leq \dots \leq i_m \leq n} u(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_m})$

2 Long-Range Dependence, Fractional Brownian Motion

$\{\xi_j, j \geq 1\}$ が強従属確率変数列の場合を考える。 $\{\xi_j, j \geq 1\}$ は強定常、平均 0 のガウス確率変数列で共分散を

$$r(k) = E[\xi_{k+1}\xi_1] = k^{-D}L(k), \quad 0 < D < 1$$

とする。ただし $L(k)$ は slowly varying。共分散が

$$\sum_{k=1}^{\infty} |r(k)| = \infty$$

の場合強従属 (long-range dependent) であるという。 $\{\xi_j, j \geq 1\}$ に対する degree 2 の U -統計量あるいは V -統計量に対して Dynkin-Mandelbaum (1983) とほぼ同様の極限定理が Dehling-Taqu (1989), (1991) によって示された。degree 2 の U -統計量と V -統計量に対してほぼ最終的な結論を得ているが、その証明の方法から容易に degree 3 以上について拡張することが出来ない。

また昨年出版された強従属性に関する論文集 Doukhan-Oppenheim-Taqu (2003) の中で、Hwai-Chung Ho and Taien Hsing は線形確率過程 (linear process) を用いて long-memory である確率変数列を構成し、それらの任意の degree を持つような非退化核関数に対する

U -統計量や V -統計量及びそれらを拡張した generalized U -統計量、 V -統計量の漸近正規性を示した。

$$X_n = \sum_{i=1}^{\infty} a_{n-i} \varepsilon_i, \quad n \geq 1,$$

ただし $\{\varepsilon_i, i \geq 1\}$ は i.i.d., mean 0, variance 1 とする。また

$$a_j = j^{-\alpha} L(j), \quad j \geq 1$$

とおく。ただし、 $L(j)$ は slowly varying function。 $\frac{1}{2} < \alpha < 1$ のとき $\{X_n\}$ は強従属性 (long-range dependence)、即ち

$$\sum_{n=1}^{\infty} |Cov(X_1, X_n)| = \infty$$

が成り立つ。このような線形確率過程は fractional ARIMA process をはじめ多くの有用な確率過程を含む。 $\{\varepsilon_i, i \geq 1\}$ が i.i.d. であるため $\{X_n\}$ の U -統計量や V -統計量に対して H -分解を使うことが出来、 $\{X_n\}$ が独立性あるいは弱従属性を持つ場合とほぼ同様に取り扱うことが出来る。しかし線形確率過程でない一般の強従属確率変数列に対しては独立性や弱従属性を取り扱う場合のような H -分解などの近似定理を用いることが出来ず、使用可能な道具立てが限られているためほとんど何もやられていない。

本講演では無限次元確率変数列の極限定理の独立あるいは弱従属を有するケースへの応用法について解説し、最後に強従属性を持つ場合にも有効であることを述べる。

参考文献

- [1] H. Dehling and M. S. Taqqu, The empirical process of some long-range dependent sequences with an application to U-statistics, Ann. Probability, 17 (1989), 1767-1783.
- [2] H. Dehling and M. S. Taqqu, Bivariate symmetric statistics of long-range dependent observations, J. Statistical Planning Inference, 28 (1991), 153-165.
- [3] P. Doukhan, G. Oppenheim and M. S. Taqqu (editors), Theory and Applications of Long-range Dependence, SBN 0-8176-4168-8. Birkhauser, Boston (2003).
- [4] T. E. Doukhan, B. Pasik-Duncan and J. Jakubowski, An approach to stochastic integration for fractional Brownian motion in a Hilbert space, preprint.
- [5] E. B. Dynkin and A. Mandelbaum, Symmetric statistics, Poisson point processes and multiple Wiener integrals, Ann. Statist., 11 (1983) 739-745.
- [6] Mwai-Chung Ho and T. Hsing, A Decomposition for generalized U-statistics of long-memory linear processes, in P. Doukhan, G. Oppenheim and M. S. Taqqu (eds.), Theory and Applications of Long-range Dependence, 143-155, Birkhauser, Boston (2003).
- [7] 前園宜彦、統計的推論の漸近理論、九州大学出版会、2001 年。

頑健性と効率性を同時に追い求める 適応的なパラメータ推定法

統計数理研究所 藤澤洋徳

1. はじめに

Windham (1995, JRSS) によって, 統計モデル $f_\theta(x) (= f(x; \theta))$ が対称でない場合にも有効なロバスト推定法が提案された. 彼の提案した方法は, 外れ値はその生起確率が小さい, という感覚に基づいた重み付き尤度を基本とし, 推定量の一致性を保証するために推定方程式にバイアス補正を行ったものである. Windham の研究の後に, Basu et al. (1998, Biometrika) や Jones et al. (2001, Biometrika) によって, 類似しているが別のロバスト推定法も提案され, 対応するダイバージェンスの構築や二つの推定方法の比較が行われた. ただし, Basu et al. の論文の中にもあるように, 良い重みを選ぶ方法の構築, という問題が残されている.

本報告では良い重みを選ぶ方法を提案する. そして, その方法がバイアスと分散の改良の調整を自動的に行っていることを指摘する. なお, その方法を提案する前に, 適当なダイバージェンスに基づいたロバスト推定法は, 外れ値の割合が多くても十分に働くことを指摘しておく.

2. ロバスト推定

まず次の量を準備しておく ($\beta > 0$):

$$d_\beta(g, f) = \int g f^\beta dx / (\int f^{1+\beta} dx)^{\beta/(1+\beta)}.$$

いま経験密度関数を $\bar{g}(x) = \sum_{i=1}^n \delta_{x_i}(x)/n$ とおく. このとき $\partial d_\beta(\bar{g}, f_\theta)/\partial \theta = 0$ が Windham によって提案された推定方程式に一致する. 整理すると以下で表現される:

$$\frac{\sum_{i=1}^n f(x_i; \theta)^\beta s(x_i; \theta)}{\sum_{i=1}^n f(x_i; \theta)^\beta} - \frac{\int f(x; \theta)^{1+\beta} s(x; \theta) dx}{\int f(x; \theta)^{1+\beta} dx} = 0.$$

ただし $s(x; \theta) = (\partial/\partial \theta) \log f(x; \theta)$ である. これは, 重み $f(x; \theta)^\beta$ に基づいた重み付きスコア関数のバイアスを補正して正規化したものである. そしてロバスト推定量は次で提案される: $\hat{\theta}_\beta = \arg \max_\theta d_\beta(\bar{g}, f_\theta)$.

なお $D_\beta(g, f) = d_\beta(g, g)^{1+\beta} - d_\beta(g, f)^{1+\beta}$ はダイバージェンスの基本性質を満たす. これは KL-divergence と正規方程式の関係に同じである. 特に $\beta \rightarrow 0$ のとき $D_\beta(g, f)$ は KL-divergence と一致する.

3. ロバスト性

ロバストだと考えられる推定量を提案したとき, そのロバスト性は, しばしば, 汚染分布 $g(x) = (1-\varepsilon)f_{\theta^*}(x) + \varepsilon\delta_{x^*}(x) (= (1-\varepsilon)f^* + \varepsilon\delta^*)$ に対する影響関数 (Huber, 1981; Hampel et al., 1986) に基づいて確認される. もちろんロバスト推定量 $\hat{\theta}_\beta$ は, そういう意味でロバストである.

影響関数に基づいたロバストの議論が行われる時は, 汚染割合 ε が小さいことも同時に仮定される. しかしながら, ロバスト推定量 $\hat{\theta}_\beta$ は, 汚染割合 ε が小さくなく

ともロバストに働くことが期待される．その理由は汚染分布の下で次が成り立つからである：

$$d_{\beta}(g, f_{\theta}) = (1 - \varepsilon)d_{\beta}(f^*, f_{\theta}) + \varepsilon d_{\beta}(\delta^*, f_{\theta}).$$

汚染部分 δ^* に関しては $d_{\beta}(\delta^*, f_{\theta}) \propto \int \delta^* f_{\theta}^{\beta} \approx 0$ が十分に期待される．なぜならば、推定の段階では $\theta \approx \theta^*$ が期待され、また、汚染部分 δ^* は目的分布 $f^* = f_{\theta^*}$ から見ると生起確率が小さい、という理由からである．ロバスト推定量 $\hat{\theta}_{\beta}$ は、統計モデル f_{θ} と真の分布 (汚染分布) g の距離を小さくすることを目指しながら、結果的に、目的分布 f^* との距離を小さくすることが期待される：

$$\hat{\theta}_{\beta} = \arg \max_{\theta} d_{\beta}(\bar{g}, f_{\theta}) \approx \arg \max_{\theta} d_{\beta}(f^*, f_{\theta}).$$

なお、上述の話は、目的分布 f^* が統計モデルのメンバーである必要もなく、汚染部分 δ^* も一点分布である必要もない．重要な性質は、 $d_{\beta}(\delta^*, f^*) \approx 0$ 、言い換えると、 $\int \delta^* f^* dx \approx 0$ という性質である．

4. チューニングパラメータの選択

いまロス関数を $L(g, f) = E_g(\rho(f))$ というタイプとしよう．このとき、交差確認法に基づいて、チューニングパラメータ β の選択方法として以下が提案できる：

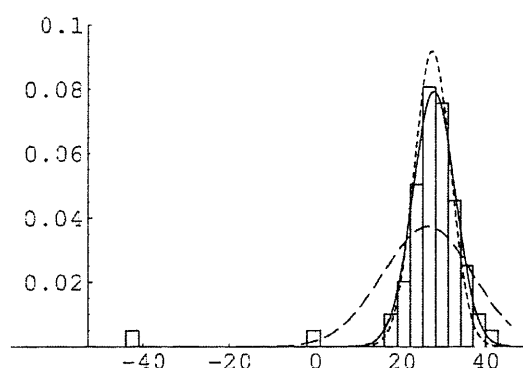
$$CV(\beta) = \frac{1}{n} \sum_{i=1}^n \rho(f(x_i; \hat{\theta}_{\beta}^{(-i)})), \quad \hat{\beta} = \arg \min_{\beta} CV(\beta).$$

残る問題はロス関数の選択である．KL タイプはロバストに推定されていないので使えない．ここではロバストに推定される $L(g, f) = -d_{\beta_0}(g, f)$ (e.g. $\beta_0 = 1$) を利用する．

なお、注意すべきこととして、一般的な枠組みでは、推定量 $CV(\beta)$ はリスク関数 $E[L(g, f_{\hat{\theta}_{\beta}})]$ の高次近似推定量だが、本当のターゲットとすべきリスク関数は $E[L(f^*, f_{\hat{\theta}_{\beta}})]$ である．上述のダイバージェンスの線形性は、この問題をも克服する．また、 $CV(\beta)$ の漸近展開を行うことで、それに基づくチューニングパラメータの選択は、バイアスと分散の改良の調整を自動的に行っていることが分かり、加えて、概パラメトリックモデルの下での漸近展開は、その調整の様子をよりクリアに見せる．

5. 数値例

右図は Newcomb light speed data に正規分布をあてはめた例で、破線は MLE に基づいたもの、点線は $\beta = 1$ という良く知られた L^2 によるロバスト法に近い方法によるもの、実線が今回提案した方法によるもの、である．破線は話にならないが、点線もロバストにし過ぎて裾のデータを軽視しているが、実線は適当な密度推定をしている．



MSEを改善する安定したリッジ回帰推定量について

丸山 祐造 東大・空間情報
W.E. Strawderman Rutgers 大

通常の線形モデルを考える．具体的には $Y = A\beta + \epsilon$ で Y は $N \times 1$ の観測ベクトル， A は既知の $N \times p$ デザイン行列 (ランクは p)， β は未知の $p \times 1$ 回帰係数ベクトル， ϵ は $N \times 1$ の誤差ベクトルとする． ϵ は，正規性をやや拡張し，0 回りの球対称分布で密度 $\sigma^{-N} f(\epsilon'\epsilon/\sigma^2)$ を持つと仮定する． σ は未知の尺度パラメータであり， $f(\cdot)$ は非負実数に対する非負関数とする．

β の最小二乗推定量 $\hat{\beta} = (A'A)^{-1}A'y$ の共分散行列は $\sigma^2(A'A)^{-1}$ に比例する． $A'A$ が非正則に近いとき， $\hat{\beta}$ は不安定になり，多重共線性の問題が生じる．一つの解決方法として，Hoerl and Kennard (1970) はリッジ回帰推定量 $\hat{\beta}_R(k) = (A'A + kI_p)^{-1}A'y$ (ここで k は正の実数) を提案した． $\hat{\beta}_R(k)$ は条件数の意味で $\hat{\beta}$ よりも安定した推定量である．条件数とは，数値的な安定性の指標であり，具体的には $\hat{\beta}$ や $\hat{\beta}_R(k)$ を線形方程式 $H\beta = A'y$ の解とみなし， $A'y$ の変動に対して， $H^{-1}A'y$ がどれだけ揺らぐかを測る量である．つまり

$$|\delta\hat{\beta}|/|\hat{\beta}| \leq \kappa(H)(|\delta A'y|/|A'y|)$$

($|\cdot|$ は，通常のユークリッドノルム) の $\kappa(H)$ が条件数であり， H を正定値行列とすると最大固有値と最小固有値の比として表される． $\hat{\beta}_R(k)$ が $\hat{\beta}$ よりも安定していることはこの定義から分かる．ただし k を理論的にデータから決める方法は容易でなく，さらには $\hat{\beta}_R(k)$ は kI_p を加えることに根拠はなく，一般に正定値行列 K を加える方が自然でクラスが広い．そこで，本発表では， $\hat{\beta}$ よりも平均二乗誤差 $E[(b - \beta)'(b - \beta)/\sigma^2]$ の意味で良い推定量 (ミニマクス) でかつ $\hat{\beta}$ よりも条件数の意味で安定している推定量を一般化リッジ回帰推定量 $\hat{\beta}_R(K) = (X'X + K)^{-1}A'y$ のクラスに対して提案することを目的とする．また正規性の下では一般化ベイズ推定量が提案出来る．この問題は Casella (1980,1985) や Strawderman (1978) などで考えられてきたが，完全な解答は得られていなかった．

問題を見やすくするために，適当な直交行列などを用いて正準化すると， X と Z がそれぞれ p 次元， n 次元 ($n = N - p$) 確率ベクトルで密度 $\prod d_i^{-1/2} \sigma^{-p-n} f(\{(x - \theta)'D^{-1}(x - \theta) + z'z\}/\sigma^2)$ ($P'(A'A)^{-1}P = D$, $D = \text{diag}(d_1, \dots, d_p)$, $d_1 \geq \dots \geq d_p$, $P'P = I_p$, $\theta = P'\beta$) のとき，二乗損失関数の下での θ の推定問題になる． X と $Z'Z (= S)$ は，それぞれ $P'\hat{\beta}$, $(y - A\hat{\beta})'(y - A\hat{\beta})$ である．

本発表では， $C = \text{diag}(c_1, \dots, c_p)$ ($c_i \geq 1, \forall i$)， $W = X'C^{-1}D^{-1}X/S$ として $\hat{\theta}_\phi = (I - W^{-1}\phi(W)C^{-1})X$ なるクラスを考える．対応する β の推定量 $\hat{\beta}_\phi$ は， $P\hat{\theta}_\phi = P(\text{diag}\{d_i^{-1}(1 - t/c_i)^{-1}\})^{-1}P'A'y(t = \phi(w)/w)$ であり，従ってその条件数は $\kappa(\hat{\beta}_\phi) = \frac{\max d_i(1-t/c_i)}{\min d_i(1-t/c_i)}$ である．以下に結果を列挙する．

Theorem 1. $\hat{\theta}_\phi$ is minimax if $\phi'(w) \geq 0$ and $0 \leq \phi \leq 2(n+2)^{-1}\{\sum(d_i/c_i)\}/\{\max(d_i/c_i)-2\}$.

Theorem 2. Suppose $0 \leq \phi(w)/w \leq t_0 < 1$ for any w . Then $\kappa(\hat{\beta}_\phi) \leq \kappa(\hat{\beta})$ for any w if either

1. if $c_p > c_1 \geq c_2 \geq \dots \geq c_{p-1}$ and $t_0 \leq \min\left(\frac{c_1 c_{p-1}(d_{p-1}-d_p)}{c_1 d_{p-1}-c_{p-1} d_p}, \frac{c_{p-1} c_p(d_1 d_{p-1}-d_p^2)}{c_p d_1 d_{p-1}-c_{p-1} d_p^2}\right)$ or
2. if $c_1 \leq c_2 \leq \dots \leq c_p$ and $t_0 \leq \min_{i>j}\left(\frac{c_i c_j(d_1 d_j-d_i d_p)}{c_i d_1 d_j-c_j d_i d_p}\right)$.

Theorem 3. Suppose $p \geq 3$ and $\sum\{d_i/d_1\} - 2 > 0$. If $d_1 > d_2$, let η_* be the unique root such that $\sum\{d_i/d_1\}^\eta = 2$ and let η_{**} be any value in $(1, \eta_*)$. If $d_1 = d_2$, let η_{**} be any value > 1 . Then if $c_i = (d_1/d_i)^{\eta_{**}-1}$, $u_+ = 2(n+2)^{-1}(\sum\{d_i/d_1\}^{\eta_{**}} - 2)$, and $v_+ = \min_{i>j}\left(\frac{c_i c_j(d_1 d_j-d_i d_p)}{c_i d_1 d_j-c_j d_i d_p}\right)$, the estimator $\hat{\theta}_\phi$ where $0 \leq \phi(w)/w \leq v_+$ for any w , $\phi(w)$ is increasing and $\lim_{w \rightarrow \infty} \phi(w) \leq u_+$, is minimax and condition number decreasing, further $c_1 \leq \dots \leq c_p$.

Theorem 4. Suppose $p \geq 3$ and $\sum\{d_i/d_1\} - 2 \leq 0$. If $p \geq 4$ let $\nu_* \in (0, 1)$ be the unique solution of $\sum_{i=1}^{p-1}\{d_i/d_1\}^\nu = 2$. Let ν_{**} be any value in $[0, \nu_*)$. If $p = 3$, choose $\nu_{**} = 0$. Then if $c_i = (d_i/d_{p-1})^{1-\nu_{**}}$ for $i = 1, 2, \dots, p-1$ and $c_p > c_1$, $u_- = 2(n+2)^{-1}\left(\sum_{i=1}^{p-1}\{d_i/d_1\}^{\nu_{**}} - 2 + \frac{c_1 d_p}{c_p d_1}\right)$ and $v_- = \min\left(\frac{c_1 c_{p-1}(d_{p-1}-d_p)}{c_1 d_{p-1}-c_{p-1} d_p}, \frac{c_{p-1} c_p(d_1 d_{p-1}-d_p^2)}{c_p d_1 d_{p-1}-c_{p-1} d_p^2}\right)$ the estimator $\hat{\theta}_\phi$ where $0 \leq \phi(w)/w \leq v_-$ for any w , $\phi(w)$ is increasing and $\lim_{w \rightarrow \infty} \phi(w) \leq u_-$, is minimax and condition number decreasing.

Theorem 5. The estimator $\left(I - \frac{\alpha}{\gamma(\alpha+1)+W}C^{-1}\right)X$ is minimax and condition number decreasing (and generalized Bayes under the normality) if either

1. under Theorem 3 setting, $\alpha \leq u_+$ and $\gamma \geq \alpha/\{(\alpha+1)v_+\}$, or
2. under Theorem 4 setting, $\alpha \leq u_-$ and $\gamma \geq \alpha/\{(\alpha+1)v_-\}$.

参考文献

- [1] Casella, G. (1980). *Ann. Statist.*, **8**, 1036-1056.
- [2] Casella, G. (1985). *J. Amer. Statist. Assoc.*, **80**, 753-758.
- [3] Hoerl, A.E. and Kennard, R.W. (1970). *Technometrics*, **12**, 55-68
- [4] Strawderman, W.E. (1978). *J. Amer. Statist. Assoc.*, **73**, 623-627.

先祖由来モデルとMDL原理に基づいた ハプロタイプブロック同定

藤澤洋徳 (統数研), 磯村実 (癌研究会), 江口真透 (統数研),
牛嶋大 (癌研究会), 宮田敏 (癌研究会), 松浦正明 (癌研究会)

1. はじめに

薬剤感受性などと SNP 遺伝子型の関係を調べる研究が現在でも続けられている。その成果は薬剤投与の事前審査などに大きく生かされることが期待されている。

ところで, SNP の存在する領域によっては, 薬剤感受性などを SNP 遺伝子型そのものと比較するよりも SNP から構成されるハプロタイプと比較する方が効率的であることが知られてきて, その方向の研究も進み始めている。そのときに重要なことは, ハプロタイプブロックの同定を行うことにある。以下ではハプロタイプブロックが存在する SNP 領域を想定して話を進める。

Daly et al. (2001, Nat.Genet.) は, ハプロタイプブロックの存在とその有効性を指摘し, ハプロタイプブロックがしばしば観測されることを報告した。ただし, ハプロタイプブロックの同定方法については, ややアドホックであった。その後に, Anderson & Novembre (2003, Am.J.Hum.Genet.) は, マルコフモデルに基づいて, その推移確率に組み換えの確率構造を導入したハプロタイプブロックの同定方法を提案した。本報告では, ハプロタイプブロックが起こりうる背後の現象をもモデル化して, 彼らの方法よりも検出力が高い方法を提案する。

2. ホットスポットとハプロタイプブロック

染色体に基づいて 2 本のハプロタイプが観測される。それが子孫に伝わる時の組み換えの状況を簡単に表したのが下図である。下図では組み換えが一箇所だけで起きているが, 実際には何箇所かで組み換えが起きたりして, それがハプロタイプの多様性を生じる。このような組み換えは起きやすいかほとんど起きないかである, ということが知られてきている。組み換えが起きやすい所をホットスポットと言い, ホットスポットとホットスポットとの間をハプロタイプブロックと言う。

$$\begin{array}{llll} \dots \text{CCCTGACG} \dots & \Rightarrow & \dots \text{CCCTGA} \quad \text{CG} \dots & \Rightarrow & \dots \text{CCCTGATA} \dots \\ \dots \text{GTTCCGTA} \dots & \Rightarrow & \dots \text{GTTCCG} \quad \text{TA} \dots & \Rightarrow & \dots \text{GTTCCGCG} \dots \end{array}$$

これまでの研究から, 子孫のハプロタイプのほとんどは, 数種類の先祖と考えられるハプロタイプの組み換えによる結果である, ということが知られてきている。この結果は, これまで SNP 毎に多様性を考慮してきた研究をハプロタイプブロック毎に考慮する研究へとシフトさせる契機となっている。

3. モデル

SNP の数を L とし, ハプロタイプが観測されていると想定して, それを $h = (h_1, \dots, h_L)$ で表すことにする。なお各 SNP には 2 つの塩基しか存在し得ないので, アルファベット順で 0 と 1 で表す。結果として $h_i = 0, 1$ である。上図のハプロタイプ CCCTGACG は $h = (0, 0, 0, 1, 1, 0, 0, 1)$ と表される。いま先祖ハプロタイプが A 個存在していたとして, それを $\mathcal{H}^* = \{h_1^*, \dots, h_A^*\}$ で表すことにし, その頻度を q_1, \dots, q_A で表すことにする。

いま簡単のためにホットスポットが一箇所だとしよう．それがSNP番号 l_1 と l_1+1 の間にあったとする．組み換えが起こったかどうかを $D = 1, 0$ で表すことにして，組み替え確率を $\lambda = \Pr(D = 1)$ とおく．また，結果的に生じる2つのハプロタイプブロックに応じて，ハプロタイプを分割し，次のような記号を用意する： $h = (h^{(1)}, h^{(2)})$ ， $h^{(1)} = (h_1, \dots, h_{l_1})$ ， $h^{(2)} = (h_{l_1+1}, \dots, h_L)$ ．さて観測されるハプロタイプは先祖ハプロタイプの組み換えによる結果であると仮定しよう．このときハプロタイプ h が観測される確率として次を想定する：

$$\begin{aligned}\Pr(H = h) &= \Pr(H = h | D = 0) \Pr(D = 0) + \Pr(H = h | D = 1) \Pr(D = 1) \\ &= (1 - \lambda) \sum_a q_a \delta(h = h_a^*) + \lambda \sum_a q_a \delta(h^{(1)} = h_a^{*(1)}) \sum_b q_b \delta(h^{(2)} = h_b^{*(2)}).\end{aligned}$$

このモデルはホットスポットの数が2つ以上の場合にも拡張できる．

構築した統計モデルは，厳密に組み換え現象を捉えたモデルではない，ということを指摘しておく．上記は一世代分だけを考えている．実際には先祖から現在までは何世代も経ているし，世代の途中で新しい先祖のような働きをするハプロタイプが出現することもある．また，突然変異などの結果として，先祖ハプロタイプからの組み換えだけでは説明できないようなハプロタイプも存在する．これらの扱いについては省略する．

上記は典型的な混合分布である．適当な潜在変数を設定し，完全モデルを作ること，最尤推定はEMアルゴリズムで実現できる．ただし，ホットスポットの数が多いと，その組み合わせの数が指数的に増え，EMアルゴリズムの各ステップで必要となる $\Pr(H = h)$ の計算量が爆発する．これに関しては，その確率を計算するための動的アルゴリズムを作って，計算量を減らした．

4. ハプロタイプブロックの同定

いま先祖ハプロタイプの集合を \mathcal{H}^* で表す．ホットスポットの場所の集合を \mathcal{S} で表す．ホットスポット集合の同定がハプロタイプブロックの同定に同値である．このとき一つの統計モデルが構築される．そのパラメータはEMアルゴリズムで推定できる．どのような \mathcal{H}^* と \mathcal{S} が良いかは，モデル選択規準MDLで決定する．なお，モデル選択規準として，AICやBICなどではなくMDLを用いたのは，確率構造と同時にハプロタイプ多様性も小さい方がハプロタイプブロック構造としては好ましく，MDLは両方とも扱えるからである．

実際には，特別なハプロタイプの扱い，先祖ハプロタイプ集合の決め方，外れ値の扱い，などの問題が残っているが，ここでは省略する．

5. シミュレーション研究とデータ解析

提案された方法のパフォーマンスを調べるために多くのシミュレーションを行った．その結果として，提案している方法は，ホットスポットを検出するパワーが高い，ことが示された．また，遺伝子配列の端の方を除けば，突然変異にも頑健であることが示された．

提案された方法を，癌研で採取されているデータやDaly et al. で解析されたのと同じデータに適用した．癌研のデータに適用した際には，それまでに行われてきた相関解析の結果の解釈を劇的にクリアにした．Daly et al. のデータに適用した際には，彼らの提案しているハプロタイプブロックとほぼ同じ構造を同定した．

競合危険モデルにおける原因別ハザード関数の推定

北海道大学大学院経済学研究科 鈴木晶夫

1. はじめに

競合危険モデルにおける原因特定ハザード関数のパラメトリック推定問題について考える。死亡原因として、互いに背反する k 個の死亡原因が考えられるとする。 $i = 1, 2, \dots, n$ に対して、個体 i の生存時間を表す確率変数を T_i 、死亡原因を表す確率変数を V_i (死亡原因全体の集合 $C = \{1, 2, \dots, k\}$ 上に値をとる)、打ち切りまでの時間を表す確率変数を Y_i とする。また、 $X_i = \min(T_i, Y_i)$ 、 $\delta_i = I(T_i \leq Y_i)$ とおく。ただし、 $I(A)$ を集合 A の定義関数とする。このとき、個体 i についての観測値は、 $(X_i = x_i, V_i = v_i, \delta_i = 1)$ または $(X_i = x_i, \delta_i = 0)$ のいずれかである。前者の場合には X_i は個体 i の生存時間を表し、後者の場合には打ち切り時間を表す。このようなデータを、競合危険データとよぶ。

死亡原因 $j \in C$ の原因特定ハザード関数は、

$$\lambda_0^{(j)}(t) = \lim_{\Delta t \rightarrow 0} \text{pr}(t \leq T_i < t + \Delta t, V_i = j | T_i \geq t) / \Delta t$$

により定義される。また、死亡原因 $j \in C$ の原因特定累積発生関数を $F_0^{(j)}(t) = \text{pr}(T_i \leq t, V_i = j)$ 、生存時間 T_i の累積分布関数を $F_0(t) = \text{pr}(T_i \leq t) = \sum_{j \in C} F_0^{(j)}(t)$ (生存関数 $\bar{F}_0 = 1 - F_0$) とする。

死亡原因全体を表す集合 C を、互いに背反する二つの部分集合 C_1 と C_2 に分割する ($C = C_1 \cup C_2$, $C_1 \cap C_2 = \emptyset$)。死亡原因の一部 C_1 のみに対して、原因特定ハザードのパラメトリックモデル

$$\mathcal{M} = \{ \{ \lambda^{(j)}(\cdot; \theta) \}_{j \in C_1}; \theta \in \Theta \} \quad \Theta \text{ はパラメータ空間}$$

を導入し、競合危険データに基づいて未知パラメータ θ を推定することを考える。死亡原因 C_2 の原因特定ハザードについては、パラメトリックモデルを仮定しない。例えば、 $C_1 = \{1\}$ 、 $C_2 = \{2, \dots, k\}$ とした場合には、死亡原因 1 の原因特定ハザード関数に対してのみパラメトリックモデルを仮定したことになる。また、モデル \mathcal{M} は、必ずしも C_1 に対する真の原因特定ハザード $\{ \lambda_0^{(j)}(\cdot) \}_{j \in C_1}$ を含むとは限らない。すなわち、モデル \mathcal{M} の misspecification が有り得るものとする。

モデル \mathcal{M} の misspecification を考慮するとき、推定すべきパラメータ値は、仮定したモデル \mathcal{M} のもとで、真のハザード関数に“最も近い”ハザードを与えるパラメータ値である。すなわち、

$$\eta(\theta) = \sum_{j \in C_1} \left[\int_0^\infty \log \lambda^{(j)}(t; \theta) dF_0^{(j)}(t) - \int_0^\infty \Lambda^{(j)}(t; \theta) dF_0(t) \right]$$

とおくとき、 $\eta(\theta)$ を最大化する $\theta \in \Theta$ の値が推定すべきパラメータ値である (これを θ_0^* とおく)。ただし、 $\Lambda^{(j)}(t; \theta) = \int_0^t \lambda^{(j)}(x; \theta) dx$, $j \in C_1$ とする。 θ_0^* に確率収束する推定量を一致推定量とよぶことにする。特に、モデル \mathcal{M} が正しいとき (真の原因特定ハザードを含むとき)、 θ_0^* は真のパラメータ値に他ならない。

2. MLE

モデル \mathcal{M} のもとで、パラメータ θ に対する対数尤度関数は

$$L_n(\theta) = \sum_{i=1}^n \sum_{j \in C_1} \left\{ \delta_i I(V_i = j) \log \lambda^{(j)}(X_i; \theta) - \Lambda^{(j)}(X_i; \theta) \right\}$$

で与えられる. MLE $\hat{\theta}_n$ は, $L_n(\theta)$ を最大化する $\theta \in \Theta$ として定義される. 打ち切り時間 Y_i の分布関数を G (生存関数 $\bar{G} = 1 - G$) とするとき, 大数の法則により, 平均対数尤度 $L_n(\theta)/n$ は,

$$\xi(\theta) = \sum_{j \in C_1} \left\{ \int_0^\infty \bar{G}(x) \log \lambda^{(j)}(x; \theta) dF_0^{(j)}(x) - \int_0^\infty \Lambda^{(j)}(x; \theta) dH(x) \right\}$$

に確率収束する. ただし, $H = 1 - \bar{F}_0 \bar{G}$. 特別な場合として, 打ち切りが起こらない場合 ($\bar{G} \equiv 1$ の場合) には, $\xi(\theta)$ を最大化する θ の値は θ_0^* に一致する. また, モデル \mathcal{M} が正しい場合にも, $\xi(\theta)$ を最大化する θ の値は, 打ち切り時間分布 G に依存せず, 真のパラメータ値に等しい. しかしながら, 一般には, $\xi(\theta)$ を最大化する θ の値は, 打ち切り時間分布 G に依存し, θ_0^* に一致しない. したがって, 一般には, MLE $\hat{\theta}_n$ は θ_0^* に対する一致性をもたない.

3. 近似MLE

通常の (死亡原因を考慮しない) 右側打ち切り生存時間データに基づくパラメトリック推定問題に対して, Oakes (1986) は近似MLEとよばれる推定量を提案した. Suzukawa *et al.* (2001) は, MLEと近似MLEの比較について検討した. ここでは, Oakes (1986) の近似MLEを競合危険問題に一般化する.

生存時間分布 F_0 と死亡原因 j の累積発生関数 $F_0^{(j)}$ のノンパラメトリック推定量をそれぞれ, $F_n, F_n^{(j)}$ とする. このとき, 近似対数尤度を

$$L_n^*(\theta) = n \sum_{j \in C_1} \left\{ \int_0^\infty \log \lambda^{(j)}(t; \theta) dF_n^{(j)}(t) - \int_0^\infty \Lambda^{(j)}(t; \theta) dF_n(t) \right\}$$

により定義する. 近似MLE $\hat{\theta}_n^*$ を, 近似対数尤度 $L_n^*(\theta)$ を最大化する θ として定義する. 近似対数尤度 $L_n^*(\theta)$ は, $\eta(\theta)$ における F_0 と $F_0^{(j)}$ をそれぞれ, それらのノンパラメトリック推定量で置き換えたものである.

近似対数尤度 $L_n^*(\theta)$ の第1項と第2項は, それぞれ Aalen-Johansen 積分, Kaplan-Meier 積分とよばれる統計量である. 前者の漸近的性質は Suzukawa (2002) により, また, 後者の漸近的性質は, Stute and Wang (1993), Stute (1995) などにより調べられた. これらと同様にして, 近似対数尤度 $L_n^*(\theta)$ の漸近的性質を示すことができる. また, それを用いて, 近似MLE $\hat{\theta}_n^*$ の漸近的性質が示される.

$\tau_{F_0} = \inf\{t : F_0(t) = 1\}$, $\tau_G = \inf\{y : G(y) = 1\}$ とおくとき, 条件 $\tau_{F_0} \leq \tau_G$ のもとで, $L_n^*(\theta)/n$ が $\eta(\theta)$ に強収束し, したがって, 近似MLE $\hat{\theta}_n^*$ が θ_0^* の一致推定量であることが示される. また, 適当な条件のもとで, $L_n^*(\theta)/\sqrt{n}$ の漸近正規性が示され, したがって, 近似MLE $\hat{\theta}_n^*$ の漸近正規性が示される.

モデル \mathcal{M} が正しいとき, MLEと近似MLEはいずれも真のパラメータの一致推定量である. しかし, 近似MLEのMLEに対する漸近相対効率率は1以下である. すなわち, モデル \mathcal{M} が正しいときには, 近似MLEによる情報損失があり, MLEが好ましい.

Oakes, D. (1986). *Biometrics*, **42**, 177-182.

Stute, W. (1995). *Ann. Statist.*, **23**, 422-439.

Stute, W. and Wang, J. L. (1993). *Ann. Statist.*, **21**, 1591-1607.

Suzukawa, A. (2002). *J. Japan Statist. Soc.*, **32**, 77-93.

Suzukawa, A., Imai, H. and Sato, Y. (2001). *Ann. Inst. Statist. Math.*, **53**, 262-276.

Nonparametric kernel regression for multinomial data

中国短大・情報ビジネス 奥村 英則

島根大・総合理工 内藤 貫太

1. 序

多項回帰の問題を考える. 各共変量 x が与えられたときの $(r+1)$ 次元確率ベクトル $\mathbf{Y} = (Y_0, \dots, Y_r)^T$ が多項分布 $\text{Mn}(p_0(x), \dots, p_r(x); n(x))$ に従う, すなわち

$$\Pr(\mathbf{Y} = \mathbf{y}|x) = n(x)! \prod_{j=0}^r \frac{p_j(x)^{y_j(x)}}{y_j(x)!}$$

とする. ここで $\mathbf{y} = (y_0, \dots, y_r)^T$, $\sum_{j=0}^r p_j(x) = 1$, $\sum_{j=0}^r y_j(x) = n(x)$ である. 我々の目的は各項目における反応曲線 $p_j(x)$, $j = 0, \dots, r$ の推定である. 1つのアプローチとして多項ロジットモデルまたは多項プロビットモデルを使用したパラメトリック推定がある. 本報告では, kernel法による古典的な手法を特殊な場合として含む新しいノンパラメトリックアプローチを提案する.

2. kernel 推定量

K 個の共変量 $x_i = (i-1)/K$, $i = 1, \dots, K$ に対して, $\bar{Y}_{ij} = Y_j(x_i)/n(x_i)$, $i = 1, \dots, K$ とする. ここで $Y_j(x_i)$ は x_i が与えられたときの Y_j である. 推定量を構成するのに Read and Cressie (1988) で与えられているパラメータ λ を持つ power-divergence loss function を利用する. これはカルバック=ライブラー情報量やヘリンガー距離のタイプの損失関数を特殊な場合として含む. 各 x に対して局所的に power-divergence loss function を利用して構成した推定量の導出基準を

$$L_\lambda(\beta, \gamma) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^K \phi_h(x_i - x) \left[\sum_{j=0}^r \bar{Y}_{ij} \left\{ \left(\frac{\bar{Y}_{ij}}{\beta_j} \right)^\lambda - 1 \right\} + \gamma \left(1 - \sum_{j=0}^r \beta_j \right) \right],$$

と定義する. ここで $L_0(\beta, \gamma) \equiv \lim_{\lambda \rightarrow 0} L_\lambda(\beta, \gamma)$, $L_{-1}(\beta, \gamma) \equiv \lim_{\lambda \rightarrow -1} L_\lambda(\beta, \gamma)$, $\beta = (\beta_0, \dots, \beta_r)^T$, γ はラグランジュ乗数である. $\phi_h(x) = \phi(x/h)/h$, ϕ は kernel で, そのサポートは $[-1, 1]$ とする. 各 x に対して

$$(\hat{\beta}, \hat{\gamma}) = \arg \min_{\beta, \gamma} L_\lambda(\beta, \gamma)$$

とする. このとき $p_j(x)$ の推定量は $\hat{\beta}_j$ によって

$$\hat{p}_{j,\lambda}(x; h) = \hat{\beta}_j = \frac{(\sum_{i=1}^K \phi_h(x_i - x) \bar{Y}_{ij}^{\lambda+1})^{\frac{1}{\lambda+1}}}{\sum_{l=0}^r (\sum_{i=1}^K \phi_h(x_i - x) \bar{Y}_{il}^{\lambda+1})^{\frac{1}{\lambda+1}}}$$

で与えられる. ここで常に $0 \leq \hat{p}_{j,\lambda}(x; h) \leq 1$ を満たすことに注意する. 特に $\lambda = 0$ のとき, 古典的な Nadaraya-Watson 推定量となる. 実際の問題ではデータからパラメータ (λ, h) を決定する必要がある. 任意の (λ, h) に対する $\hat{\mathbf{p}}(\cdot; h) = (\hat{p}_{0,\lambda}(\cdot; h), \hat{p}_{1,\lambda}(\cdot; h), \dots, \hat{p}_{r,\lambda}(\cdot; h))^T$ の族は $\mathbf{p}(\cdot) = (p_0(\cdot), p_1(\cdot), \dots, p_r(\cdot))^T$ の推定量のある 1つのクラスを構成する. したがってパラメータの選択は推定量のクラスから推定量を選択することと考えられる. パラメータの選択はデータを生成する分布と推定量の良さを評価する損失関数に依存して決定すべきである.

3. 漸近的性質

推定量の良さを評価する基準として MISE (mean integrated square error) を採用する. $n_i = n(x_i), i = 1, \dots, K$ とおく. ある適当な正則条件の下で $\hat{p}_{j,\lambda}(x; h)$ のバイアスと分散の漸近的な評価はそれぞれ次式で与えられる:

$$\text{Bias}[\hat{p}_{j,\lambda}(x; h)] \simeq \frac{\lambda}{2n_1} b_{1j}(x) + h^2 b_{2j}(x), \quad V[\hat{p}_{j,\lambda}(x; h)] \simeq \frac{p_j(x)(1-p_j(x))R}{n_1 K h}.$$

ここで $\mu_2 = \int_{-1}^1 x^2 \phi(z) dz$, $R = \int_{-1}^1 \phi(z)^2 dz$ であって, $j = 0, 1, \dots, r$ に対して

$$\begin{aligned} b_{1j}(x) &= (1 - (1+r)p_j(x)), \\ b_{2j}(x) &= \frac{\mu_2}{2} \left\{ \lambda \left(\frac{p_j^{(1)}(x)^2}{p_j(x)} - p_j(x) \sum_{l=0}^r \frac{p_l^{(1)}(x)^2}{p_l(x)} \right) + p_j^{(2)}(x) \right\}. \end{aligned}$$

さらに, $n_1 K h^5 \rightarrow \rho^2$ を仮定する. ここで ρ はある正の定数である. このとき任意の $(r+1)$ 次元ベクトル $\alpha (\neq 0)$ に対して次の漸近正規性が得られる:

$$\sqrt{n_1 K h} \alpha^T \{ \hat{\mathbf{p}}(x; h) - \mathbf{p}(x) - \lambda (2n_1)^{-1} \mathbf{b}_1(x) \} \rightarrow_d N(\rho \alpha^T \mathbf{b}_2(x), \alpha^T \Sigma(x) \alpha).$$

ここで $\mathbf{b}_1(x) = (b_{10}(x), b_{11}(x), \dots, b_{1r}(x))^T$, $\mathbf{b}_2(x) = (b_{20}(x), b_{21}(x), \dots, b_{2r}(x))^T$ and $\Sigma(x) = R(\text{diag}(\mathbf{p}(x)) - \mathbf{p}(x)\mathbf{p}(x)^T)$ である. 仮定した条件の下で $\hat{p}_{j,\lambda}(x; h)$ のバイアスの右辺の第 1 項は第 2 項の $O(h^2)$ -バイアス関数より収束が遅くなるので

$$\tilde{p}_{j,\lambda}(x; h) = \hat{p}_{j,\lambda}(x; h) - (2n_1)^{-1} \lambda (1 - (1+r)\hat{p}_{j,\lambda}(x; h))$$

とおくと, これは $\text{Bias}[\hat{p}_{j,\lambda}(x; h)]$ の第 1 項をバイアスとして持たないバイアス修正推定量となり, さらにこのバイアス修正は分散の主要項に影響を与えないことがわかる. 分散の主要項は λ を含まず, バイアスの主要項のみに λ が含まれることに注意する. またバイアス修正推定量に対しても同様な漸近正規性が得られる.

4. パラメータの選択

推定量の良さを評価する基準として MISE (mean integrated square error) を使用する. バイアス修正推定量 $\tilde{p}_{j,\lambda}(x; h)$ の積分区間 $[\delta_1, 1 - \delta_2] (\subset [h, 1 - h])$ に対する MISE は

$$\text{MISE}[\tilde{p}_{j,\lambda}(\cdot; h)] \simeq \frac{h^4 \mu_2^2}{4} (B_{1j} \lambda^2 + 2B_{2j} \lambda + B_{3j}) + \frac{V_j R}{n_1 K h}$$

となる. ここで $B_{1j}, B_{2j}, B_{3j}, V_j$ は δ_1, δ_2 に依存した p_j の汎関数である. 従って, すべての $\tilde{p}_{\lambda j}(\cdot; h)$ の MISE の和 $\sum_{j=0}^r \text{MISE}[\tilde{p}_{j,\lambda}(\cdot; h)]$ を漸近的に最小にする最適な λ, h はそれぞれ

$$\lambda_{\text{opt}} = -\frac{B_2}{B_1}, \quad h_{\text{opt}} = \left(\frac{R}{\mu_2^2} \right)^{1/5} \left(\frac{B_1 V}{B_1 B_3 - B_2^2} \right)^{1/5} (n_1 K)^{-1/5}$$

で与えられる. ここで $B_1 = \sum_{j=0}^r B_{1j}$, $B_2 = \sum_{j=0}^r B_{2j}$, $B_3 = \sum_{j=0}^r B_{3j}$, $V = \sum_{j=0}^r V_j$ である. この結果に対して ROT(Rule of Thumb) や PI(Plug-in) の手法を適用すれば実際にデータに基づくパラメータ λ とバンド幅 h が構成できる.

5. シミュレーション

データに基づくパラメータ λ とバンド幅 h の選択および推定量の挙動に関するシミュレーション結果から我々のアプローチが有効であることを示した.

Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness-of-Fit Statistics for Discrete multivariate Data*, Springer-Verlag.

ロジスティック回帰分析におけるロジットの推定量の分布の近似について

帯広畜産大学・畜産 種市信裕
北海道教育大学・教育 関谷祐里

1. はじめに. 一般線形モデル (Nelder and Wedderburn [4]) を 2 項分布 $B(n, \pi)$,

$$\Pr(Y = y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}, \quad (y = 0, 1, \dots, n)$$

について考え, その連結関数を正準連結関数であるロジット関数 $\text{logit } \pi = \log\{\pi/(1 - \pi)\}$ で与える。ここで, N 個の異なるサブグループにおける反応数に対応した確率変数 Y_1, \dots, Y_N が互いに独立に $Y_\alpha \sim B(n_\alpha, \pi_\alpha)$, $(\alpha = 1, \dots, N)$ とすることに適用するとロジスティック回帰モデル (ロジットモデル) $\text{logit } \pi_\alpha = \mathbf{x}'_\alpha \boldsymbol{\beta}$, $(\alpha = 1, \dots, N)$ が得られる。ただし, $\mathbf{x}'_\alpha = (x_{\alpha 1}, \dots, x_{\alpha p})$, $(p < N)$ は共変量ベクトル, $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_p)$ とする。特に, $p = 2$, $x_{\alpha 1} = 1, x_{\alpha 2} = x_\alpha$, $(\alpha = 1, \dots, N)$ とすると用量反応モデルとして用いられるロジットモデル $\text{logit } \pi_\alpha = \beta_1 + \beta_2 x_\alpha$, $(\alpha = 1, \dots, N)$ となる。ロジスティック回帰分析においてはパラメータの最尤推定量 $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)'$ が漸近正規性を持つこと, つまり $\text{logit } \hat{\pi}_\alpha = \text{logit } \pi_\alpha(\hat{\boldsymbol{\beta}})$ が漸近正規性を持つことが知られているが, 本報告ではこの正規近似の改良の提案をおこなう。

2. 推定量の分布の性質. 最尤推定量の漸近的性質 (e.g. Serfling [5]), および Lawley [2], McCullagh [3] の結果に基づき, $\text{logit } \pi_\alpha$ の推定量 $\text{logit } \hat{\pi}_\alpha = \mathbf{x}'_\alpha \hat{\boldsymbol{\beta}}$ の分布に関して次の定理を得る。

定理 1 : $\sum_{\alpha=1}^N n_\alpha = N_T$ とするとき, 条件

$$n_j/N_T \rightarrow \mu_j \quad (0 < \mu_j < 1) \text{ for each } j, \quad \text{as } N_T \rightarrow \infty \quad (1)$$

のもとで, $\text{logit } \hat{\pi}_\alpha - \text{logit } \pi_\alpha \xrightarrow{L} N(0, \sigma_\alpha^2)$ as $N_T \rightarrow \infty$, $(\alpha = 1, \dots, N)$ ただし, $\sigma_\alpha^2 = \sum_{l=1}^p \sum_{m=1}^p x_{\alpha l} x_{\alpha m} \kappa^{l,m}$, $\kappa^{l,m}$ は (j, k) 成分が $\sum_{\alpha=1}^N x_{\alpha j} x_{\alpha k} n_\alpha \pi_\alpha (1 - \pi_\alpha)$ である $p \times p$ 行列 \mathbf{K} (Fisher 情報行列) の逆行列 \mathbf{K}^{-1} の (l, m) 成分であり, \xrightarrow{L} は法則収束を意味する。

定理 2 : 条件 (1) のもとで, $E(\text{logit } \hat{\pi}_\alpha) = A_\alpha + N_T^{-1} B_\alpha + o(N_T^{-1})$, $V(\text{logit } \hat{\pi}_\alpha) = N_T^{-1} C_\alpha + o(N_T^{-1})$, $E[\{\text{logit } \hat{\pi}_\alpha - E(\text{logit } \hat{\pi}_\alpha)\}^3] = N_T^{-2} D_\alpha + o(N_T^{-2})$, と評価できる。ただし, $A_\alpha = \text{logit } \pi_\alpha$,

$$B_\alpha = -(1/2) \lim_{N_T \rightarrow \infty} N_T \sum_{j=1}^p \sum_{q=1}^p \sum_{r=1}^p \sum_{s=1}^p x_{\alpha j} \kappa^{j,q} \kappa^{r,s} \kappa_{q,r,s}, \quad C_\alpha = \lim_{N_T \rightarrow \infty} N_T \sigma_\alpha^2,$$

$$D_\alpha = \lim_{N_T \rightarrow \infty} N_T^2 \sum_{j=1}^p \sum_{l=1}^p \sum_{m=1}^p \sum_{q=1}^p \sum_{r=1}^p \sum_{s=1}^p x_{\alpha j} x_{\alpha l} x_{\alpha m} \kappa^{j,q} \kappa^{l,r} \kappa^{m,s} \kappa_{q,r,s},$$

$$\kappa_{q,r,s} = \sum_{\alpha=1}^N x_{\alpha q} x_{\alpha r} x_{\alpha s} n_\alpha \pi_\alpha (1 - \pi_\alpha)(1 - 2\pi_\alpha), \quad (q, r, s = 1, \dots, p)$$

とする。

3. 正規化変換によるロジットの正規近似の改良. 定理 1, 定理 2 を用いて Konishi [1] および Taneichi et. al.[6] により導出された正規化変換の一般的公式に基づき $\text{logit } \hat{\pi}_\alpha$ に関して以下の正規化変換を得ることができる。

表 1: $\beta_1^* = -2, \beta_2^* = 3$ の場合の $n = 40, 60, 80, 100, 120$ に対する $C_1^j(0.05) \times 10^5$ ($j = 1, 2, 3$) の値

n	40	60	80	100	120
$C_1^1(0.05)$	1280	-136	295	586	384
$C_1^2(0.05)$	885	-403	61	314	204
$C_1^3(0.05)$	885	-383	72	314	214

$$g_1(\text{logit } \hat{\pi}_\alpha) = \begin{cases} \sqrt{\frac{N_T}{C_\alpha}} \left[\frac{A_\alpha}{\eta_\alpha} \left\{ \left(\frac{\text{logit } \hat{\pi}_\alpha}{A_\alpha} \right)^{\eta_\alpha} - 1 \right\} - \frac{1}{N_T} (B_\alpha + \frac{1}{2} C_\alpha \xi_\alpha) \right] & (\eta_\alpha \neq 0) \\ \sqrt{\frac{N_T}{C_\alpha}} \left[A_i \log \frac{\text{logit } \hat{\pi}_\alpha}{A_\alpha} - \frac{1}{N_T} (B_\alpha + \frac{1}{2} C_i \xi_\alpha) \right] & (\eta_i = 0), \end{cases}$$

$$g_2(\text{logit } \hat{\pi}_\alpha) = \sqrt{\frac{N_T}{C_\alpha}} \left\{ \frac{1}{\xi_\alpha} \left(e^{\xi_\alpha (\text{logit } \hat{\pi}_\alpha - A_\alpha)} - 1 \right) - \frac{1}{N_T} \left(B_\alpha + \frac{1}{2} C_\alpha \xi_\alpha \right) \right\}$$

ただし, $\xi_\alpha = -D_\alpha/3C_\alpha^2$, $\eta_\alpha = A_\alpha \xi_\alpha + 1$ とする。これらは, 以下の性質 $g_j(\text{logit } \hat{\pi}_\alpha) \xrightarrow{L} N(0, 1)$, as $N_T \rightarrow \infty$, ($j = 1, 2$), および $E[\{g_j(\text{logit } \hat{\pi}_\alpha) - E(g_j(\text{logit } \hat{\pi}_\alpha))\}^3] = o(N_T^{-1/2})$, ($j = 1, 2$) を満たし正規分布への近似が改良される。

4. 数値計算による近似の性能の考察

モデルとして $p = 2, N = 8, n_1 = \dots = n_8 = n$ を考え, 二項乱数を用いたモンテカルロ実験をおこなう。第一ロジットの推定値 $\text{logit } \hat{\pi}_1$ に関して $A_1^1 = (\text{logit } \hat{\pi}_1 - \text{logit } \pi_1)/\sigma_1^2$, $A_1^2 = g_1(\text{logit } \hat{\pi}_1)$, $A_1^3 = g_2(\text{logit } \hat{\pi}_1)$, としたときに $B = 10000$ 回の繰り返しに対し A_1^j , ($j = 1, 2, 3$) が $N(0, 1)$ の上側 5%点 $l_{0.05}$ を越える回数を $B_1^j(0.05)$, ($j = 1, 2, 3$) とするときに $N(0, 1)$ への近似の良さとして指標 $C_1^j(0.05) = B_1^j(0.05)/B - 0.05$, ($j = 1, 2, 3$) を用い, パラメータの真の値が $\beta_1^* = -2, \beta_2^* = 3$ の場合にその値を示したのが表 1 である。

参考文献

- [1] Konishi, S.: *Ann. Statist.*, **19** (1991), 2209–2225.
- [2] Lawley, D.N.: *Biometrika*, **43** (1956), 295–303.
- [3] McCullagh, P.: *Tensor methods in statistics*, (1987), Chapman and Hall.
- [4] Nelder, J.A. and Wedderburn, R.W.M.: *J. R. Statist. Soc., A*, **135** (1972), 370–384.
- [5] Serfling, R.J.: *Approximation theorems of mathematical statistics*, (1980), Wiley.
- [6] Taneichi, N., Sekiya, Y., Suzukawa, A., and Imai, H.: *Commun. Statist.-Theory Meth.*, **31** (2002), 163–179.

A CONVEX COMBINATION of TWO-SAMPLE U -STATISTICS

鹿児島大学理学部 大和 元
鹿児島高等予備校 戸田 光一郎

1. 2 標本 U -統計量の凸結合

F と G を実軸上の連続な分布関数とし、母数 $\theta = \theta(F, G)$ を F and G の estimable parameter(regular functional) とする。すなわち、可測関数(核, kernel) $h(x_1, \dots, x_{k_1}; y_1, \dots, y_{k_2})$ があって、

$$\theta = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x_1, \dots, x_{k_1}; y_1, \dots, y_{k_2}) \prod_{i=1}^{k_1} dF(x_i) \prod_{j=1}^{k_2} dG(y_j)$$

と表されるものとする。ただし、 $h(x_1, \dots, x_{k_1}; y_1, \dots, y_{k_2})$ はそれぞれ x_1, \dots, x_{k_1} と y_1, \dots, y_{k_2} に関し対称で、正整数 k_1 と k_2 は上の式を満たす最小の整数で degree である。 X_1, \dots, X_{n_1} と Y_1, \dots, Y_{n_2} をそれぞれ分布 F 及び G からの大きさ n_1, n_2 の標本とする。

kernel h に対する 2 標本 U -統計量 U_{n_1, n_2} は次で与えられる：

$$U_{n_1, n_2} = \binom{n_1}{k_1}^{-1} \binom{n_2}{k_2}^{-1} \sum_{(n_1, k_1)} \sum_{(n_2, k_2)} h(X_{i_1}, \dots, X_{i_{k_1}}; Y_{j_1}, \dots, Y_{j_{k_2}}),$$

ここで、和 $\sum_{(n_1, k_1)}$ は $1 \leq i_1 < \dots < i_{k_1} \leq n_1$ を満たす i_1, \dots, i_{k_1} 上で取られる。

$w(\alpha_1, \dots, \alpha_j; k)$ は与えられた k について、 $j = 1, \dots, k$ および $\alpha_1 + \dots + \alpha_j = k$ を満たす正の整数 $\alpha_1, \dots, \alpha_j$ の非負で対称な関数とする。 $w(\alpha_1, \dots, \alpha_j; k)$'s の少なくとも 1 つは正と仮定する。 $d(k, j) = \sum_{\alpha_1 + \dots + \alpha_j = k}^+ w(\alpha_1, \dots, \alpha_j; k)$, $j = 1, 2, \dots, k$, とし、 $D(n, k) = \sum_{j=1}^k d(k, j) \binom{n}{j}$ とおく。 $j_1 = 1, \dots, k_1$ 及び $j_2 = 1, \dots, k_2$ に対し、kernel $h_{(j_1, j_2)}$ を

$$\begin{aligned} & h_{(j_1, j_2)}(x_1, \dots, x_{j_1}; y_1, \dots, y_{j_2}) \\ &= \frac{1}{d(k_1, j_1)d(k_2, j_2)} \sum_{r_1 + \dots + r_{j_1} = k_1}^+ \sum_{s_1 + \dots + s_{j_2} = k_2}^+ w(r_1, \dots, r_{j_1}; k_1) w(s_1, \dots, s_{j_2}; k_2) \\ & \quad \times h(\underbrace{x_1, \dots, x_{r_1}}_{r_1}, \dots, \underbrace{x_{j_1}, \dots, x_{r_{j_1}}}_{r_{j_1}}; \underbrace{y_1, \dots, y_{s_1}}_{s_1}, \dots, \underbrace{y_{j_2}, \dots, y_{s_{j_2}}}_{s_{j_2}}) \end{aligned}$$

とし、この kernel に対する 2 標本 U -統計量を $U_{n_1, n_2}^{(j_1, j_2)}$ とする。

θ の推定量として、次で与えられる 2 標本 U -統計量の凸結合を導入する：

$$Y_{n_1, n_2} = \frac{1}{D(n_1, k_1)D(n_2, k_2)} \sum_{j_1=1}^{k_1} \sum_{j_2=1}^{k_2} d(k_1, j_1)d(k_2, j_2) \binom{n_1}{j_1} \binom{n_2}{j_2} U_{n_1, n_2}^{(j_1, j_2)}.$$

例えば、 $w(1, 1, \dots, 1; k) = 1$ and $w(\alpha_1, \dots, \alpha_j; k) = 0$ for $j = 1, \dots, k-1$ の場合、 Y_{n_1, n_2} は 2 標本 U -統計量 U_{n_1, n_2} になる。他方、 $j = 1, \dots, k$ および $r_1 + \dots + r_j = k$ を満たす正の整数 $\alpha_1, \dots, \alpha_j$ に対して、 $w(\alpha_1, \dots, \alpha_j; k) = k! / (\alpha_1! \dots \alpha_j!)$ の場合、 Y_{n_1, n_2} は 2 標本 V -統計量に一致する： $V_{n_1, n_2} = n_1^{-k_1} n_2^{-k_2} \sum_{i_1=1}^{n_1} \dots \sum_{i_{k_1}=1}^{n_1} \sum_{j_1=1}^{n_2} \dots \sum_{j_{k_2}=1}^{n_2} h(X_{i_1}, \dots, X_{i_{k_1}}; Y_{j_1}, \dots, Y_{j_{k_2}})$ 。

2. 漸近的性質

$d(k, k) > 0$, すなわち $w(1, \dots, 1; k) > 0$ とする。このとき, $\delta_k = kd(k, k-1)/d(k, k)$ とおくと

$$\frac{d(k, k)}{D(n, k)} \binom{n}{k} = 1 - \frac{\delta_k}{n} + O\left(\frac{1}{n^2}\right), \quad \frac{d(k, k-1)}{D(n, k)} \binom{n}{k-1} = \frac{\delta_k}{n} + O\left(\frac{1}{n^2}\right)$$

が成り立つ。また, $r_1 + \dots + r_{j_1} = k_1$ ($j_1 = 1, \dots, k_1$) および $s_1 + \dots + s_{j_2} = k_2$ ($j_2 = 1, \dots, k_2$) を満たす正の整数 r_1, \dots, r_{j_1} および s_1, \dots, s_{j_2} に対して

$$E\left\{h\left(\underbrace{X_1, \dots, X_1}_{r_1}, \dots, \underbrace{X_{j_1}, \dots, X_{j_1}}_{r_{j_1}}; \underbrace{Y_1, \dots, Y_1}_{s_1}, \dots, \underbrace{Y_{j_2}, \dots, Y_{j_2}}_{s_{j_2}}\right)\right\}^2 < \infty$$

とする。このとき,

$$Y_{n_1, n_2} - \theta = U_{n_1, n_2} - \theta + R_{n_1, n_2}^1, \quad E[R_{n_1, n_2}^1]^2 \leq \frac{C}{N}$$

が成り立つ。これにより, U_{n_1, n_2} の漸近的性質から Y_{n_1, n_2} の漸近的性質が導かれる。なお, 以下の極限操作では $N = n_1 + n_2$ とし, $0 < p < 1$ に対し, $p_{n_1, n_2} = n_1/N \rightarrow p$ となるように $n_1, n_2 \rightarrow \infty$ とする。

$h^{(1,0)}(x_1) = E[h(x_1, X_2, \dots, X_{k_1}; Y_1, \dots, Y_{k_2})]$ および $h^{(0,1)}(y_1) = E[h(X_1, \dots, X_{k_1}; y_1, Y_2, \dots, Y_{k_2})]$ とし, $\zeta_{1,0}^2 = \text{Var}[h^{(1,0)}(X_1)]$ および $\zeta_{0,1}^2 = \text{Var}[h^{(0,1)}(Y_1)]$ とおく。

以下の性質が成り立つ:

$$N^{1/2}(Y_{n_1, n_2} - \theta) \xrightarrow{D} N(0, p^{-1}k_1^2\zeta_{1,0}^2 + (1-p)^{-1}k_2^2\zeta_{0,1}^2) \quad (n_1, n_2 \rightarrow \infty),$$

$$\sup_{-\infty < x < \infty} \left| P(Y_{n_1, n_2} - \theta \leq x) - P(U_{n_1, n_2} - \theta \leq x) \right| \leq \frac{C}{N}.$$

$E[|h^{(1,0)}(X_1)|^3] < \infty$ and $E[|h^{(0,1)}(Y_1)|^3] < \infty$ and $|\sqrt{p_{n_1, n_2}} - \sqrt{p}| \leq CN^{-1/2}$ を仮定する。 Z_1, Z_2 は独立で, $N(0, 1)$ に従う確率変数とする。 $\xi_{n_1, n_2}^2 = n_1^{-1}k_1^2\zeta_{1,0}^2 + n_2^{-1}k_2^2\zeta_{0,1}^2$ とおく。このとき,

$$\sup_{-\infty < x < \infty} \left| P\left(N^{1/2}(Y_{n_1, n_2} - \theta) \leq x\right) - P\left(p^{-1/2}k_1\zeta_{1,0}Z_1 + (1-p)^{-1/2}k_2\zeta_{0,1}Z_2 \leq x\right) \right| \leq \frac{C}{N^{1/2}},$$

$$\sup_{-\infty < x < \infty} \left| P\left(N^{1/2}\left\{p^{-1}k_1^2\zeta_{1,0}^2 + (1-p)^{-1}k_2^2\zeta_{0,1}^2\right\}^{-1/2}(Y_{n_1, n_2} - \theta) \leq x\right) - \Phi(x) \right| \leq \frac{C}{N^{1/2}},$$

$$\sup_{-\infty < x < \infty} \left| P\left(\xi_{n_1, n_2}^{-1/2}(Y_{n_1, n_2} - \theta) \leq x\right) - \Phi(x) \right| \leq \frac{C}{N^{1/2}}.$$

Denker, M. (1985), Asymptotic distribution theory in nonparametric statistics. Friedr. Vieweg & Sohn, Braunschweig.

Koroljuk, V. S. & Borovskich, Yu. V. (1994), Theory of U-statistics. Kluwer Academic Publishers, Dordrecht.

Lee, A. J. (1990), U-statistics. Marcel Dekker, New York.

Nomachi, T., Kondo, M. & Yamato, H. (2002), *Scientiae Mathematicae Japonicae*, **56**, 95–106.

Toda, K. & Yamato, H. (2001), *J. Japan Statist. Soc.*, **31**, 225–237.

Statistical Estimation of Optimal Portfolios for Dependent Returns of Assets

早稲田大学大学院 理工学研究科 D1 白石博

早稲田大学 理工学部 谷口正信

時間 t における m 個の資産のランダムなリターンを $\mathbf{X}(t) = (X_1(t), \dots, X_m(t))'$ とし、その期待値および分散共分散行列をそれぞれ $\mu = E\{\mathbf{X}(t)\}$ 、 $\Sigma = Cov(\mathbf{X}(t))$ とする。また、portfolio weight を $\alpha = (\alpha_1, \dots, \alpha_m)'$ で表す。このとき、過去、種々の基準から optimal portfolio weight が提案されたが、これらは統一的な形

$$g(\mu, \Sigma)$$

で表される。ここに、 $g: \mathbf{R}^{m+m(m+1)/2} \rightarrow \mathbf{R}^{m-1}$ なる関数。

先行文献では、リターンが i.i.d. or stationary process と仮定して optimal portfolio weight を統計的に推定するものがあるが、推定量の最適性を論じたものは皆無である。本報告ではリターンに時系列構造を仮定し、 $g(\mu, \Sigma)$ の漸近最適推定を報告する。

$\{\mathbf{X}(t) = (X_1(t), \dots, X_m(t))'; t \in \mathbf{Z}\}$ が Gaussian m -vector stationary process で、平均 $\mu = (\mu_1, \dots, \mu_m)'$ 、共分散関数 $\mathbf{R}(k) = E\{(\mathbf{X}(t) - \mu)(\mathbf{X}(t+k) - \mu)'\}$ 、スペクトル行列 $\mathbf{f}(\lambda) = \{f_{jk}(\lambda)\}$ を持つとする。

さらに次を仮定する。

Assumption 1.

$$\sum_{l=-\infty}^{\infty} \|\mathbf{R}(l)\|_E < \infty$$

$\mathbf{X}(1), \dots, \mathbf{X}(T)$ が観察されたとする。そのとき $\mathbf{R}(0) = \Sigma = \{\sigma_{ij}\}$ 、 $\theta = (\mu', vech(\Sigma)')'$ とし、

$$\begin{aligned} \hat{\mu} &= \frac{1}{T} \sum_{t=1}^T \mathbf{X}(t) \\ \hat{\Sigma} &= \frac{1}{T} \sum_{t=1}^T (\mathbf{X}(t) - \hat{\mu})(\mathbf{X}(t) - \hat{\mu})' \\ \hat{\theta} &= (\hat{\mu}', vech(\hat{\Sigma})')'. \end{aligned}$$

とする。また、次の行列を導入する。

$$\begin{aligned} \Omega_1 &= 2\pi \mathbf{f}(0), \quad (m \times m) - matrix \\ \Omega_2 &= \left\{ 2\pi \int_{-\pi}^{\pi} \{f_{\alpha_1 \alpha_3}(\lambda) \overline{f_{\alpha_2 \alpha_4}(\lambda)} + f_{\alpha_1 \alpha_4}(\lambda) \overline{f_{\alpha_2 \alpha_3}(\lambda)}\} d\lambda \right. \\ &\quad \left. ; \alpha_1, \alpha_2, \alpha_3, \alpha_4 = 1, \dots, m, \alpha_1 \geq \alpha_2 \text{ and } \alpha_3 \geq \alpha_4 \right\}, \quad (m(m+1)/2 \times m(m+1)/2) - matrix, \end{aligned}$$

Assumption 2. 関数 $g(\theta)$ は、連続的に微分可能とする。

Theorem 1. Assumptions 1 and 2 の下で、

$$\sqrt{T}(g(\hat{\theta}) - g(\theta)) \xrightarrow{L} N\left(0, \left(\frac{\partial g}{\partial \theta}\right)' \begin{pmatrix} \Omega_1 & \mathbf{0} \\ \mathbf{0} & \Omega_2 \end{pmatrix} \left(\frac{\partial g}{\partial \theta}\right)\right), (as\ T \rightarrow \infty)$$

ここに、

$$\left(\frac{\partial g}{\partial \theta'}\right) = \left(\frac{\partial g}{\partial \mu'}, \frac{\partial g}{\partial vech(\Sigma)'}\right) = \left(\frac{\partial g}{\partial \mu_1}, \dots, \frac{\partial g}{\partial \mu_k}, \frac{\partial g}{\partial \sigma_{11}}, \dots, \frac{\partial g}{\partial \sigma_{kk}}\right).$$

とする。

Assumption 3.

- (i) $\mathbf{f}(\lambda)$ は、 $\eta = (\eta_1, \dots, \eta_q)' \in \mathcal{H} \subset \mathbf{R}^q$ でパラメトライズされる。
- (ii) $\mathbf{A}^{(j)}(l) \equiv \int_{-\pi}^{\pi} \partial \mathbf{f}_{\eta}(\lambda) / \partial \eta_j d\lambda$, $j = 1, \dots, q, l \in \mathbf{Z}$ に対して、 $\sum_{l=-\infty}^{\infty} \|\mathbf{A}^{(j)}(l)\|_E < \infty$ が成り立つ。
- (iii) $q \geq m(m+1)/2$.

Assumption 4. (λ に無関係な) 正の定数 c が存在して、 $\mathbf{f}_{\eta}(\lambda) - c\mathbf{I}_m$ は半正定値とする。ここに、 \mathbf{I}_m は $m \times m$ の単位行列とする。

Assumption 5. 時系列 Fisher 情報行列 $\mathcal{F}(\eta)$ は正定値とする。

ここで、

$$\Phi(\lambda) = (vec\{\psi_{11}(\lambda)\}, \dots, vec\{\psi_{m1}(\lambda)\}, vec\{\psi_{22}(\lambda)\}, \dots, vec\{\psi_{mm}(\lambda)\})$$

ただし、 $\psi_{ab}(\lambda) = \frac{1}{2}\{\mathbf{e}_a \mathbf{e}_b' + \mathbf{e}_b \mathbf{e}_a'\}$ (\mathbf{e}_a は第 a 成分が 1、その他の成分は 0 の m 次元ベクトル) とする。

Theorem 2. Assumptions 1-5 の下では、 $g(\hat{\theta})$ が漸近有効であることと、次を満たす λ に無関係な行列 C が存在することは同値である。

$$\{\mathbf{f}_{\eta}(\lambda)' \otimes \mathbf{f}_{\eta}(\lambda)\} \Phi(\lambda) = (vec\{\partial \mathbf{f}_{\eta}(\lambda) / \partial \eta_1\}, \dots, vec\{\partial \mathbf{f}_{\eta}(\lambda) / \partial \eta_q\}) C.$$

定理 2 より、リターンが時系列である場合、 $g(\cdot, \cdot)$ の推定量は一般に漸近有効でないことが判明した。

次に、どのような時系列モデルで、漸近有効でなくなるか、その度合いの大きさを

$$INE \equiv \det \left[\left\{ \text{Asymptotic variance of } g(\hat{\theta}) \right\} - \mathcal{F}(\eta)^{-1} \right]$$

を数値的に評価した例を報告する。

最後に、quasi-Gaussian maximum likelihood estimator を使って漸近有効推定量が構成できることも報告する。

Discriminant Analysis for Multivariate Non-Gaussian Locally Stationary Processes.

早稲田大学 理工学部 蛭川 潤一

多変量解析においては、多くの判別解析の手法が詳細に研究されてきた。多変量解析における古典的な判別の技法を時系列データに応用することは、実用上興味深い。Shumway(1982) は定常時系列における様々な判別問題について、包括的な研究を行った。Non-Gaussian 多変量定常時系列データ間の判別について、Kakizawa et al.(1998) は Kullback-Leibler 情報測度や Chernoff 情報測度を含む判別手法を提案し、地震と探鉱の爆発の分類の問題に応用した。

定常時系列データについての解析はよく確立されて来たが、実世界の現象を記述するのに定常時系列モデルは妥当ではない。Dahlhaus(1996) は重要な非定常過程のクラスである局所定常過程を提案し、統計的推論の漸近理論を確立した。本報告では、Kullback-Leibler 情報測度や Chernoff 情報測度を含む非線形な time-varying スペクトル測度の積分汎関数を用いた一般化された判別基準を提案し、その判別手法の最適性に関する asymptotics を明らかにする。また、LAN アプローチによる真に non-Gaussian optimal な判別基準も提案する。

定義 1. 確率過程列 $\mathbf{X}_{t,T} = (X_{t,T}^{(1)}, \dots, X_{t,T}^{(m)})'$, $(t = 2-N/2, \dots, 1, \dots, T, \dots, T+N/2; T, N \geq 1)$ が transfer function 行列 \mathbf{A}° の局所定常過程とは、表現

$$\mathbf{X}_{t,T} = \int_{-\pi}^{\pi} \exp(i\lambda t) \mathbf{A}_{t,T}^\circ(\lambda) d\xi(\lambda) \quad (1)$$

が存在することである。ここに、 $\xi(\lambda)$ は適当な条件を満たす定常スペクトル測度であり、 $\varepsilon_t := \int_{-\pi}^{\pi} \exp(i\lambda t) d\xi(\lambda)$ は、ホワイトノイズベクトル。また、 \mathbf{A}° は次の条件を満たす。

定数 K と 2π -周期行列値関数 $\mathbf{A} : [0, 1] \times \mathbb{R} \rightarrow \mathbb{C}^{m \times m}$ が存在して、 $\mathbf{A}(u, -\lambda) = \overline{\mathbf{A}(u, \lambda)}$ であり

$$\sup_{t, \lambda} \left| \mathbf{A}_{t,T}^\circ(\lambda)_{a,b} - \mathbf{A}\left(\frac{t}{T}, \lambda\right)_{a,b} \right| \leq KT^{-1} \quad (2)$$

for all $a, b = 1, \dots, m$ and $T \in \mathbb{N}$ を満たす。ここに、 $\mathbf{A}(u, \lambda)$ は u に関して連続である。以後、 $\mathbf{f}(u, \lambda) := \mathbf{A}(u, \lambda) \Omega \mathbf{A}(u, \lambda)^*$, $\Omega = E(\varepsilon_t \varepsilon_t')$ を time-varying スペクトル密度関数行列と呼ぶ。

さて、 $\mathbf{X}_{t,T}$ は $\text{MA}(\infty)$ 表現

$$\mathbf{X}_{t,T} = \sum_{s=-\infty}^{\infty} \mathbf{a}_{t,T}(s) \varepsilon_{t-s} \quad (3)$$

を持つ、つまり、

$$\mathbf{A}_{t,T}^\circ(\lambda) = \sum_{s=-\infty}^{\infty} \mathbf{a}_{t,T}(s) \exp(-i\lambda s) \quad (4)$$

であるとし、transfer function 行列は適当な正則条件を持つと仮定する。Time-varying スペクトル密度関数 $\mathbf{f}(u, \lambda)$ の kernel 型ノンパラメトリック推定量として、

$$\hat{\mathbf{f}}_T(u, \lambda) = \int_{-\pi}^{\pi} W_T(\lambda - \mu) \mathbf{I}_N(u, \mu) d\mu, \quad (5)$$

を用いる。ここに、 $W_T(\omega) = M \sum_{\nu=-\infty}^{\infty} W(M(\omega + 2\pi\nu))$ は適当な条件を満たす weight function であり、 $\mathbf{I}_N(u, \lambda)$ は時刻 u における局所ペリオドグラム行列である。

Kullback-Leibler 情報測度や Chernoff 情報測度は、次の非線形な time-varying スペクトル測度の積分汎関数に一般化される。

$$D_H(\mathbf{f}_j; \mathbf{f}_k) = \frac{1}{2} \int_0^1 \int_{-\pi}^{\pi} H(\mathbf{f}_j(u, \lambda) \mathbf{f}_k^{-1}(u, \lambda)) \frac{d\lambda}{2\pi} du \quad (6)$$

ここに、 $H(\cdot)$ はある行列値関数であり、関数 $H(\mathbf{Z})$ は、 $\mathbf{Z} = \mathbf{E}_m$ (単位行列) で唯一の最小値を持つ。 $H(\mathbf{Z})$ の候補として、

$$H_K(\mathbf{Z}) = \text{tr}\{\mathbf{Z}\} - \log|\mathbf{Z}| - m, \quad H_{B_\alpha}(\mathbf{Z}) = \log|\alpha\mathbf{Z} + (1-\alpha)\mathbf{E}_m| - \alpha \log|\mathbf{Z}| \quad (7)$$

の他に、quadratic function

$$H_Q(\mathbf{Z}) = \frac{1}{2} \text{tr}(\mathbf{Z} - \mathbf{E}_m)^2 \quad (8)$$

等が考えられる。

今、観測系列 $\mathbf{X}_T = (\mathbf{X}'_{2-N/2,T}, \dots, \mathbf{X}'_{1,T}, \dots, \mathbf{X}'_{T,T}, \dots, \mathbf{X}'_{T+N/2,T})'$ が time-varying スペクトル密度関数 $\mathbf{f}_j(u, \lambda)$ で特徴づけられる 2 つの既知のカテゴリ Π_j , $j = 1, 2$ の内、どちらのクラスに属しているかを判別したいとする。 \mathbf{X}_T のスペクトルとカテゴリ Π_j の距離を $D_H(\hat{\mathbf{f}}_T; \mathbf{f}_j)$ で測る。判別関数 D_H を

$$D_H = D_H(\hat{\mathbf{f}}_T; \mathbf{f}_2) - D_H(\hat{\mathbf{f}}_T; \mathbf{f}_1) \quad (9)$$

で定義して、 $D_H > 0$ の時 Π_1 に、 $D_H \leq 0$ の時 Π_2 に分類する。 D_H に基づく誤判別確率は $P_{D_H}(2|1) = \Pr(D_H \leq 0 | \Pi_1)$ 、 $P_{D_H}(1|2) = \Pr(D_H > 0 | \Pi_2)$ となる。

定理 1. 適当な正則条件を仮定し、ルベーク測度正の集合上で $\mathbf{f}_1(u, \lambda) \neq \mathbf{f}_2(u, \lambda)$ とする。そのとき、 Π_j の下で、 $D_H \xrightarrow{P} (-1)^{j+1} D_H(\mathbf{f}_j; \mathbf{f}_k)$ であり、

$$\sqrt{T}\{D_H + (-1)^j D_H(\mathbf{f}_j; \mathbf{f}_k)\} \xrightarrow{D} N(0, V_H^2(j, k)), \quad \text{as } T \rightarrow \infty \quad (10)$$

である。ここに、

$$V_H^2(j, k) = \int_0^1 \left[\frac{1}{4\pi} \int_{-\pi}^{\pi} \text{tr}\{\mathbf{Q}_{j,k}(u, \lambda) \mathbf{f}_j(u, \lambda)\}^2 d\lambda + \frac{1}{64\pi^4} \sum_{a,b,c,d} \kappa_{a,b,c,d} \gamma_{a,b}^{(j,k)}(u) \gamma_{c,d}^{(j,k)}(u) \right] du,$$

$$\mathbf{Q}_{j,k}(u, \lambda) = \mathbf{f}_k^{-1}(u, \lambda) [H^{(1)}(\mathbf{f}_j(u, \lambda) \mathbf{f}_k^{-1}(u, \lambda))]',$$

$$\Gamma_H^{(j,k)}(u) = \{\gamma_{a,b}^{(j,k)}(u)\}_{a,b=1,\dots,m} = \int_{-\pi}^{\pi} \mathbf{A}_j^*(u, \lambda) \mathbf{Q}_{j,k}(u, \lambda) \mathbf{A}_k(u, \lambda) d\lambda$$

である。

この定理より、 D_H の一致性が示されるが、その他、 Π_1 が Π_2 と contiguous である時の asymptotic Gaussian optimality や Gaussian robustness を報告する。また、Non-Gaussian LAN に基づく真に、asymptotic optimal な判別基準も報告する。

References

1. Dahlhaus, R., 1996. On the Kullback-Leibler information divergence of locally stationary processes. *Stoch. Processes and Their Applications* **62** 139-168.
2. Kakizawa, Y., Shumway, R. H. and Taniguchi, M., 1998. Discrimination and Clustering for Multivariate Time Series. *J. Amer. Statist. Assoc.* **93**. 328-340.
3. Sakiyama, K. and Taniguchi, M., 2004. Discriminant Analysis for Locally Stationary Processes. To appear in *J. Multivariate Anal.*
4. Shumway, R. H., 1982. Discriminant analysis for time series. *Handbook of Statistics. Vol.2.* 1-46. Amsterdam: North-Holland.

Influence of Liquidity on the Optimal Strategy

九州大学大学院経済学研究院 松本 浩一

数理ファイナンスの古典的な理論では、市場において金融商品が自由に取引できることが仮定されている。しかし、現実の市場では、その仮定が適切でない流動性の低い金融商品も多数存在する。流動性の影響は無視できない問題である。本報告では最適投資戦略に対する流動性の影響を解析した。

$(\Omega, \mathcal{F}, P, \{\mathcal{F}_t; 0 \leq t \leq T\})$ を通常の状態を満す確率空間とし、 P の下での $\{\mathcal{F}_t\}$ -Brown 運動を $\{B(t); 0 \leq t \leq T, B(0) = 0\}$, 強度 λ の $\{\mathcal{F}_t\}$ -Poisson 過程を $\{P(t); 0 \leq t \leq T, P(0) = 0\}$ とする。安全資産の価格過程 $\beta(t)$ と危険資産の価格過程 $S(t)$ はそれぞれ、以下を満たすとする。

$$\begin{aligned} d\beta(t) &= r\beta(t)dt, \quad \beta(0) = 1, \\ dS(t) &= \mu S(t)dt + \sigma S(t)dB(t), \quad S(0) = S_0. \end{aligned}$$

ここで、 S_0, r, μ, σ は正の定数であり、 $r < \mu$ を満たす。

総資産 $W(t)$ を持ち、安全資産と危険資産で運用する投資家を考える。投資家は運用効率を上げるために、危険資産の投資比率 $X(t)$ を $v(t)$ に変更する取引を試みる。しかし、流動性が低いために取引は失敗する可能性があり、Poisson 過程のジャンプ時刻にのみ取引は成功する。

投資家の制御する $v(t)$ を投資戦略と呼び、それに従って決定される $W(t), X(t)$ を $W(t; v), X(t; v)$ と書くこととする。なお、投資家に許容される投資戦略を以下の空間に限定する。

$$\mathcal{V}[t, T] = \{v | v \text{ は可予測, } t \leq s \leq T \text{ において } 0 \leq v(s) \leq 1\}.$$

このことは、投資家は危険資産の空売りや借金による危険資産への投資ができないことを意味する。

本報告では最適投資戦略 $v^\lambda(t)$ と値関数 $V^\lambda(t, x, w)$ を解析した。最適投資戦略とは投資家の期待効用 $E[\log W(T; v)]$ を最大にする投資戦略であり、値関数とは

$$V^\lambda(t, x, w) = \sup_{v \in \mathcal{V}[t, T]} E[\log(W(T; v)) | \mathcal{F}_t] |_{(X(t), W(t)) = (x, w)}$$

である。なお、取引が常に成功するときの最適投資戦略 $v^\infty(t)$ と値関数 $V^\infty(t, x, w)$ は、以下になることが Merton (1971) によって示されている。

$$\begin{aligned} v^\infty(t) &= x_0, \\ V^\infty(t, x, w) &= \log w + \left(r + \frac{(\mu - r)^2}{2\sigma^2} \right) (T - t). \end{aligned}$$

ここで、 $x_0 = (\mu - r)/\sigma^2$ であり、 $0 < x_0 < 1$ とする。

本報告で示した主要結果は次の通りである．詳細は Matsumoto (2003) を参照して欲しい．

[定理 1] 最適投資戦略は存在し，値関数は以下によって与えられる．

$$V^\lambda(t, x, w) = \log w + A^\lambda(T - t, x)$$

ここで，

$$\begin{aligned} A^\lambda(t, x) &= \int_0^t e^{-\lambda s} K(s, x) ds + \lambda \int_0^t \sup_{0 \leq y \leq 1} \left(\int_0^s e^{-\lambda u} K(u, y) du \right) ds, \\ K(t, y) &= E[f(Y^y(t))], \\ f(y) &= (\mu - r)y + r - \frac{1}{2}\sigma^2 y^2, \\ Y^y(t) &= \frac{yS(t)/S_0}{yS(t)/S_0 + (1 - y)\beta(t)}. \end{aligned}$$

特に λ が十分に大きければ，最適投資戦略は一意で，適当な定数 C に対して以下が成立する．

$$|v^\lambda(t) - x_0| \leq C \frac{1}{\lambda}, \quad 0 \leq t \leq T.$$

[定理 2] 有界で連続な関数 $h_i^* : [0, \infty) \rightarrow \mathbf{R}$, $i \geq 1$ が存在し，任意の $n \in \mathbf{N}$ に対して次を満たす定数 $C_n > 0$, $\lambda_n > 0$ が存在する． $0 \leq t \leq T, \lambda \geq \lambda_n$ に対して

$$\begin{aligned} \left| v^\lambda(T - t) - \left(x_0 + \sum_{i=1}^n \frac{h_i^*(\lambda t)}{\lambda^i} \right) \right| &\leq C_n \frac{1}{\lambda^{n+1}}, \\ \left| V^\lambda(T - t, v^\lambda(T - t), w) - V^\lambda(T - t, x_0 + \sum_{i=1}^n \frac{h_i^*(\lambda t)}{\lambda^i}, w) \right| &\leq C_n \frac{1}{\lambda^{2n+3}}. \end{aligned}$$

[定理 3] $0 \leq t \leq T$ とする． $\lambda \rightarrow \infty$ のとき， $0 \leq x \leq 1$ で一様に

$$\begin{aligned} V^\lambda(t, x, w) &\rightarrow V^\infty(t, x, w), \\ \lambda(V^\infty(t, x, w) - V^\lambda(t, x, w)) &\rightarrow \frac{1}{2}\sigma^2(x - x_0)^2 + \frac{1}{2}\sigma^4 x_0^2(1 - x_0)^2(T - t). \end{aligned}$$

なお，本報告では最適投資戦略の数値計算結果や漸近展開による近似精度についても言及した．

[参考文献]

Matsumoto, K. (2003) : Optimal Portfolio of the Low Liquid Asset, Doctoral Thesis, Graduate School of Mathematical Science, the University of Tokyo
Merton, R.C. (1971) : Optimum Consumption and Portfolio Rules in a Continuous Time Model, *Journal of Economic Theory*, 3, 373-413

Test for parameter change in diffusion processes by cusum statistics based on one-step estimators

ソウル大学 Sangyeol Lee

統計数理研究所 西山陽一

東京大学大学院数理科学研究科 吉田朋広

1. 主定理

パラメータ空間 Θ は \mathbb{R} の有界凸開集合であるとする。次の確率微分方程式によって与えられる 1-次元定常拡散過程を考える：

$$dX_t = V_0(X_t, \theta)dt + V(X_t)dW_t.$$

ただし $t \rightsquigarrow W_t$ は 1-次元標準ウィナー過程であるとする。我々は、仮説

H_0 : 真値 $\theta_0 \in \Theta$ は変化しない vs

H_1 : H_0 でない

を検定したい。以下、我々は CUSUM 検定統計量を導入し、 H_0 のもとでの漸近挙動を導出する。

ここで、次のような仮定をおく。ただし $\delta^k = \frac{d^k}{d\theta^k}$ と表す。

Assumption 1 確率過程 $t \rightsquigarrow X_t^{\theta_0}$ は定常で exponential rate の α -混合過程である。

Assumption 2 各 $x \in \mathbb{R}$ に対し、関数 $\theta \rightsquigarrow V_0(x, \theta)$ は 4 回連続的微分可能である。さらに、 $E_\theta[|\frac{\delta V_0}{V}(X_0^\theta, \theta)|^2] > 0$ が各 $\theta \in \Theta$ に対して成り立つ。

Assumption 3 $E_{\theta_0}[|\frac{\delta V_0}{V}(X_0^{\theta_0}, \theta_0)|^{2q}] < \infty$ がある $q > 1$ に対して成り立つ。

Assumption 4 各 $k = 1, 2, 3$ に対し、確率過程の族 $\{\frac{\delta^k V_0}{V}(X_s, \theta); \theta \in \Theta\}$ が、一様大数の法則における確率過程の族 $\{g(s, \theta); \theta \in \Theta\}$ に課せられた条件を満たす。さらに、

$$E_{\theta_0} \left[\sup_{\theta \in \Theta} \left| \frac{\delta^k V_0}{V^2}(X_0, \theta) V_0(X_0, \theta_0) \right| \right] < \infty, \quad \text{for every } k = 1, 2, 3$$

および

$$E_{\theta_0} \left[\sup_{\theta \in \Theta} \left| \frac{\delta^k V_0 \delta^l V_0}{V^2}(X_0, \theta) \right| \right] < \infty, \quad \text{for every } k, l = 0, 1, 2, 3 \text{ such that } 1 \leq k + l \leq 3.$$

が成り立つ。

対数尤度が

$$\ell_T^{\theta_0}(\theta) = \int_0^T \frac{V_0}{V^2}(X_t^{\theta_0}, \theta) dX_t^{\theta_0} - \frac{1}{2} \int_0^T \left(\frac{V_0}{V} \right)^2 (X_t^{\theta_0}, \theta) dt$$

によって与えられるという事実に基づき、我々は $\psi_T^{\theta_0}(\theta)$ と $\eta(\theta)$ を次のように定義する：

$$\begin{aligned} \psi_T^{\theta_0}(\theta) &:= \int_0^T \frac{\delta V_0}{V^2}(X_t^{\theta_0}, \theta) dX_t^{\theta_0} - \int_0^T \frac{V_0 \delta V_0}{V^2}(X_t^{\theta_0}, \theta) dt; \\ \eta(\theta) &:= -E_\theta \left[\left(\frac{\delta V_0}{V} \right)^2 (X_0^\theta, \theta) \right]. \end{aligned}$$

いま、我々はワンステップ推定量に基づく CUSUM 検定手続きを考える。我々は、初期推定量 $\hat{\theta}_T^0$ が次を満たすと仮定する。

Assumption 5 $\theta_* \in \Theta$ を固定する。 $T \in [0, 1]$ に対し、 $\hat{\theta}_T^0 = \theta_*$ とおく。 $T \in (1, \infty)$ に対し、 $\sup_{T \in (1, \infty)} r_T |\hat{\theta}_T^0 - \theta_0| < \infty$ 、 P_{θ_0} -almost surely が、ある $r_T > 0$ であって $T^{1/4} r_T^{-1} \rightarrow 0$ as $T \rightarrow \infty$ であるようなものに対して成り立つ。

我々は、ワンステップ推定量 $\hat{\theta}_T$ を

$$\hat{\theta}_T = \hat{\theta}_T^0 - [T\eta(\hat{\theta}_T^0)]^{-1} \psi_T^{\theta_0}(\hat{\theta}_T^0)$$

と定義し、確率過程 $u \rightsquigarrow S_T^u$ を

$$S_T^u = u\sqrt{T}(\hat{\theta}_{uT} - \hat{\theta}_T), \quad \forall u \in [0, 1]$$

と定義する。次が我々の主定理である。

Theorem 1 *Assumption 1 – 5 が成り立つと仮定する。このとき、 H_0 のもとで、 $T \rightarrow \infty$ とするとき、確率過程の列 $u \rightsquigarrow S_T^u$ は 平均ゼロのガウス過程 $u \rightsquigarrow Y^u$ であって $EY^u Y^v = -\eta(\theta_0)^{-1}\{(u \wedge v) - uv\}$ を満たすものに $\ell^\infty([0, 1])$ 空間の中で分布収束する。*

この定理と連続写像定理から、我々は次の検定統計量の漸近分布を導出することができる。

$$\hat{S}_T = \sup_{u \in [0, 1]} u^2 T(-\eta(\hat{\theta}_T)) |\hat{\theta}_{uT} - \hat{\theta}_T|^2.$$

Corollary 2 *Assumption 1 – 5 が成り立つと仮定する。このとき、 H_0 のもとで、確率変数の列 \hat{S}_T は $T \rightarrow \infty$ とするとき $\sup_{u \in [0, 1]} |W_u^\circ|^2$ に分布収束する。ただし $u \rightsquigarrow W_u^\circ$ は標準ブラウン橋である。*

2. Remarks

我々の結果を多次次元 (J 次元) の場合に一般化できる。この場合、検定統計量 \hat{S}_T は

$$\hat{S}_T = \sup_{u \in [0, 1]} u^2 T(\hat{\theta}_{uT} - \hat{\theta}_T)' (-\eta(\hat{\theta}_T)) (\hat{\theta}_{uT} - \hat{\theta}_T)$$

と置き換えられる。ただし $-\eta(\theta)$ はフィッシャー情報行列である。Corollary 2 における極限は $S^J = \sup_{u \in [0, 1]} \sum_{j=1}^J |W_u^{\circ, j}|^2$ によって置き換えられる。ただし $u \rightsquigarrow (W_u^{\circ, 1}, \dots, W_u^{\circ, J})'$ は J -次元標準ブラウン橋である。その分布を解析的に計算するのは容易ではないが、Lee et al. (2003) はモンテカルロシミュレーションによる数表を与えている。(彼らの Table 1 を見よ。)

上の結果は、独立同一分布に従う確率変数列の場合の結果と比較されるべきであろう。Csörgő and Horváth(1997) は尤度比に基づく検定統計量を提案し、極限

$$\sup_{\epsilon_n \leq u \leq 1 - \epsilon_n} \sqrt{\frac{\sum_{j=1}^J |W_u^{\circ, j}|^2}{u(1-u)}}$$

を導出している。ただし ϵ_n はゼロに収束する定数列であるが、分母に $u(1-u)$ があったためこのような切り取りが必要だったのである。一方、我々の極限分布では切り取りを行う必要はない。

Csörgő, M. & L. Horváth (1997). Wiley.

Lee, S., J. Ha, O. Na, & S. Na (2003). *Scand. J. Statist.* **30**, 781-796.

First and higher-order asymptotic properties of estimators for stochastic differential equations with jumps *

Nakahiro YOSHIDA

Graduate School of Mathematical Sciences, University of Tokyo,
3-8-1 Komaba, Meguro-ku, Tokyo 153, Japan.

December 2, 2004

1 Estimation for stochastic differential equations with/without jumps finitely/infinately many on compacts based on continuous/discrete observations

For estimation of stochastic differential equations without jumps, estimation schemes have been provided and the asymptotic behaviors have been established. Rigorous asymptotic theory in estimation problem by sampled data was essentially initiated by Prakasa Rao (1983, 1988). Prakasa Rao provided a least square estimator and established asymptotic results under the “rapidly increasing experimental design (RIED)” condition. Yoshida (1992) proposed a maximal likelihood type estimator and proved the joint convergence in law of drift-diffusion estimators with correct rates under a lower sampling rate than RIED. This form of asymptotic results are succeeded by other studies: Kessler (1997) improved it to lower sampling rates. For a stochastic differential equation with jumps, recently Shimizu and Yoshida (to appear in SISP) gave a maximum likelihood type estimator and proved asymptotic normality of it.

A general M-estimator unifying those approaches and the asymptotic normality and expansion were presented in Yoshida (Zenkintenkai 2003 (2003), DYNSTOCH: Inference for Partially Observed Processes at Copenhagen(2004)). Our estimator is general, and hence, it includes the maximal likelihood type estimator as well as other estimators such as a moment estimator. Misspecified cases are also in our scope. So we start with a stochastic differential equation as a generator of the “true” data, independently of the parameters to be estimated.

Let (E, \mathcal{E}) be a measurable space. μ denotes a Poisson random measure on $\mathbf{R}_+ \times E$ with intensity measure $\nu(dt, dv) = dt \times \lambda(dv)$ for some σ -finite measure λ on E , and let $\tilde{\mu} = \mu - \nu$ be the compensated Poisson random measure. Moreover, w^β denote independent Wiener processes. We consider a d -dimensional stochastic process $X = (X_t)$ which satisfies the stochastic integral equation

$$X_t = X_0 + \int_0^t A(X_s) ds + \sum_{\beta=1}^r \int_0^t B_\beta(X_s) dw_s^\beta + \int_0^t \int_E C(X_{s-}, v) \tilde{\mu}(ds, dv), \quad (1)$$

where $A : \mathbf{R}^d \rightarrow \mathbf{R}^d$, $B_\beta : \mathbf{R}^d \rightarrow \mathbf{R}^d$ and $C : \mathbf{R}^d \times E \rightarrow \mathbf{R}^d$.

In a usual correctly specified parametric case, the coefficients and ν of (1) depend on unknown parameters, and (1) coincides with the equation corresponding to the true value of the parameter vector. Processes with infinitely many jumps on compacts are within our scope.

*This work was in part supported by the Research Fund for Scientists of the Ministry of Science, Education and Culture, and by Cooperative Research Program of the Institute of Statistical Mathematics.

The data set we can observe consists of the discrete-time observations $(X_{t_i})_{i=0}^n$ from the complete realization $(X_t)_{t \in \mathbf{R}_+}$. Here $t_i = ih$ ($i = 0, 1, \dots, n$) with $h = h_n$. We denote by $\theta = (\theta', \theta'') \in \Theta$ the unknown parameter we want to estimate from the data $(X_{t_i})_{i=0}^n$, where $\Theta = \Theta' \times \Theta''$ is an open rectangle of $\mathbf{R}^{m'} \times \mathbf{R}^{m''}$. The unknown parameter has two parts. In a standard situation, for example, θ' denotes a parameter which is commonly involved in the drift and jump parts, and θ'' denotes a parameter in the diffusion part. Since θ' can be multi-dimensional, this parameterization includes a special case where there are distinct parameters in drift and diffusion.

In the previous conferences, the author showed asymptotic expansion as well as asymptotic normality for an estimator, i.e., a mapping from $(X_{t_i})_{i=0}^n$ to Θ . It can be constructed quite freely, thus, we can choose an estimating function independently from (1). Let $\bar{A}^n \in \mathcal{F}(\mathbf{R}^d \times \Theta'; \mathbf{R}^d)$, and set

$$e_i(\theta') \equiv e_{n,i}(\theta') = \Delta_i X - h \bar{A}_{i-1}^n(\theta'),$$

where $\Delta_i X = X_{t_i} - X_{t_{i-1}}$ and $\bar{A}_{i-1}^n(\theta') = \bar{A}^n(X_{t_{i-1}}, \theta')$; this is our common rule of notation. We considered an M-estimator, a root of the following random functions:

$$\begin{aligned} \psi_a(\theta) &= \sum_{i=1}^n A_a^n(X_{t_{i-1}}, \Delta_i X, \theta) [e_i(\theta')] \\ &\quad + \sum_{i=1}^n \{C_a^n(X_{t_{i-1}}, \Delta_i X, \theta') - \bar{C}_a^n(X_{t_{i-1}}, \Delta_i X, \theta')\}, \\ \psi_b(\theta) &= \sum_{i=1}^n \{B_b^n(X_{t_{i-1}}, \Delta_i X, \theta'') [e_i(\theta')^{\otimes 2}] h^{-1} - \bar{B}_b^n(X_{t_{i-1}}, \Delta_i X, \theta'')\}. \end{aligned}$$

2 Polynomial type large deviation inequality

In order to obtain the most rigid consistency (i.e., consistency of *any* sequence of M-estimators), estimate of the residual term in a stochastic expansion, and convergence of the moments, we need an estimate

$$P \left[|\hat{\theta}_n - \theta_0| \geq \varphi_n^\beta \right] \leq n^{-M}.$$

We prepare the following notation. a bounded open set Θ , in \mathbf{R}^m , $C = \text{diameter}(\Theta)$, $\Xi = \Theta \times \mathcal{T}$, $\mathbb{H}_T : \Omega \times \Xi \rightarrow \mathbf{R}^{C^3}$ in θ , $\mathbb{U}_T(\theta_0) = \{u \in \mathbf{R}^m; \theta_0 + a_T(\theta_0)u \in \Theta\}$, $a_T(\theta_0) \in \text{GL}(m)$,

$$Z_T(u; \theta_0, \tau) = \exp \{ \mathbb{H}_T(\theta_0 + a_T(\theta_0)u, \tau) - \mathbb{H}_T(\theta_0, \tau) \}.$$

We work under the Partial LAQ property at a $\theta_0 \in \Theta$:

$$Z_T(u; \theta_0, \tau) = \exp \left(\Delta_T(\theta_0, \tau)[u] - \frac{1}{2} \Gamma(\theta_0, \tau)[u, u] + r_T(u, \theta_0, \tau) \right)$$

for $(u, \tau) \in \mathbb{U}_T(\theta_0) \times \mathcal{T}$.

Let $V_T(r, \theta_0) = \{u \in \mathbb{U}_T(\theta_0); r \leq |u|\}$. We will assume four moment conditions. Under mild conditions, we have

Theorem 1. *Given $L > 0$, under the above assumptions, there exists a constant $C_L > 0$ such that*

$$P_{\theta_0, \tau_0} \left[\sup_{(u, \tau) \in V_T(r, \theta_0) \times \mathcal{T}} Z_T(u; \theta_0, \tau) \geq e^{-r^{2-\mu_2}/2} \vee e^{-r^{2-\mu_4}/2} \right] \leq \frac{C_L}{r^L}$$

for any $T > 0$ and $r > 0$.

We discussed an illustrative example only to explain the idea, while it is too simple to show the real advantage of our result:

$$X_t = X_0 + \theta_1 t + \sqrt{\theta_2} w_t.$$