(19)「確率統計学における漸近的方法－統計解析・金融工学・保険数理・確率数値解析への発展」に関する研究報告

Edgeworth Expansion and Regenerative Block Bootstrap for Ergodic Diffusion

東京大学大学院数理科学研究科　深澤 正彰

1　はじめに

ブートストラップとは1979年に Efron によって定式化されたリサンプリング手法である．独立同分布確率変数列の観測に対し適当な推定量を考えた時，その推定量の分布を評価するのは理論上，実用上共に重要な問題であるが，漸近正規性に基づく伝統的な正規近似に対し，ブートストラップはより適切かつ簡便な近似を与えている．

この成功を受けて，独立でない定常時系列に対しても，これに対応する手法が提案され，議論されてきた．本報告では，拡散過程に対するブートストラップ手法が提案され，その正当性が議論される．連続時間確率過程に対するブートストラップについては，特殊な枠組におけるものを除くと，筆者の知る限り未だ有効な手法が知られておらず，本報告で提案されるブートストラップは，二次の正確さが保証された最初のものとなる．

2　エッジワース展開

定義： $X=(X_t)$ を，ある適当な確率空間で定義された一次元拡散過程とし，$Z=(Z_t)$ を X の加法的汎関数とする．また，$\mathbb{R}$ 上の分布 μ に対し，μ が X の初期分布となるような確率測度，

$$P_\mu[\cdot]=\int_{\mathbb{R}} P_x[\cdot]\mu(dx)$$

を定義する．$\mathbb{R}$ 上の固定した2点，x,y に対し，マルコフ時刻の列 $\tau_m, m\in\mathbb{N}$ を以下のように定義しよう．

$$\tau_1=\inf\{t\geq 0\ ;\ X_t=x\},\quad \hat{\tau}_m=\inf\{t\geq\tau_m\ ;\ X_t=y\},\quad \tau_{m+1}=\inf\{t\geq\hat{\tau}_m\ ;\ X_t=x\}.$$

さらにブロック列，(F_m,L_m) を $F_m=Z_{\tau_{m+1}}-Z_{\tau_m}$, $L_m=\tau_{m+1}-\tau_m$ によって定義すると，$(F_m,L_m), m\in\mathbb{N}$ は P_μ のもと，独立同分布となる．$F_0=Z_{\tau_1}, L_0=\tau_1$ とおくと，(F_m,L_m) は (F_0,L_0) とも独立である．ここで x,y,μ は $P_x[\tau_2<\infty]=1, P_\mu[\tau_1<\infty]=1$ を満たす任意のものでよい．$G_m=F_m-L_m\beta/\alpha$ とおく．さて，定義されたブロック列に対し，モーメントを以下のように定義しよう．

$$\alpha=P_x[L_1],\quad \beta=(\beta_j)=P_x[F_1],\quad \rho=(\rho_j)=P_x[G_1L_1],\quad \Gamma=(\gamma_{i,j})=\mathrm{Var}_x[G_1]/\alpha,$$
$$\kappa_{i,j,k}=P_x[G_1^iG_1^jG_1^k],\quad \gamma_{i,j,k}=(\kappa_{i,j,k}-\gamma_{i,j}\rho_k-\gamma_{j,k}\rho_i-\gamma_{k,i}\rho_j)/\alpha.$$

定理： $A(\beta/\alpha)=0$ を満たす β/α の近傍で4回連続微分可能な関数 $A:\mathbb{R}^n\to\mathbb{R}$ に対し，X,Z,x,y についての適当な仮定のもと，

$$\sup_{\mathbb{B}\in\mathcal{B}(\mathbb{R})}\left|P_\mu[\sqrt{T}A(Z_T/T)/\sigma\in\mathbb{B}]-\int_{\mathbb{B}}\mathcal{P}_T(z)dz\right|$$

は $O(T^{-1})$ となる．ここで初期分布 μ は定常分布であり，

$$\mathcal{P}_T(z)=\phi(z)+\frac{1}{\sqrt{T}}\left\{\frac{A_1}{\sigma}z+\frac{A_2}{6\sigma^3}(z^3-3z)\right\}\phi(z),$$

$$A_1=\frac{1}{2}\sum_{i,j=1}^n a_{i,j}\gamma_{i,j}\quad ,A_2=\sum_{i,j,k=1}^n a_ia_ja_k\gamma_{i,j,k}+3\sum_{i,j,k,l=1}^n a_ia_ja_{k,l}\gamma_{i,k}\gamma_{j,l}\ ,$$

また,

$$a_j = \partial_j A(\beta/\alpha), \quad a_{i,j} = \partial_i \partial_j A(\beta/\alpha), \quad \sigma^2 = \sum_{i,j=1}^{n} a_i a_j \gamma_{i,j}.$$

である.

以下, X は, ある確率微分方程式,

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t$$

に従っているものとしよう. 定理の仮定に関して例えば次のような十分条件を得た.

命題: 係数 b は局所可積分, σ は局所有界で, $1/\sigma$ は多項式増大とする. さらに係数について,

$$\limsup_{|x|\to\infty} \frac{\mathrm{sgn}(x)b(x)}{\sigma(x)^2} < 0$$

が成立するとき, 任意の x, y と, 区間 $[x, y]$ 上一次独立な多項式増大連続関数を成分とするベクトル f に対して, $dZ_t = f(X_t)dt$ は定理の仮定を満たす.

3 ブートストラップ

以下, $A(\cdot)$ は, ある既知の $\tilde{A}(\cdot,\cdot)$ によって, $A(z) = \tilde{A}(z, \beta/\alpha)$ と表現でき, さらに $\sqrt{T}A(Z_T/T)$ の漸近分散が 1 である場合を考えよう. エッジワース展開に用いたブロック, $(F_m, L_m), m \in \mathbb{N}$ は独立同分布列であることに再び注意する. このブロック列に対して, ブートストラップ・リサンプリングを行おう. 観測 $(X_t)_{0\le t\le T}$ が含むブロックの数を M_T とおく. $\{(F_m.L_m)\}_{m=1}^{M_T}$ 上の一様分布に従う確率変数列 $\{(F_m^*, L_m^*)\}_{m=1}^{M_T}$ を考える. $\sqrt{T}A(Z_T/T)$ に関するブートストラップ推定量 A_T^* を

$$\sqrt{T^*}\hat{A}(Z_T^*/T^*)/\hat{\sigma} + \sum_{j=1}^{n} \hat{a}_j \hat{\rho}_j/(2\hat{\alpha}\hat{\sigma}\sqrt{T})$$

と定義する. ただし,

$$\hat{A}(\cdot) = \tilde{A}(\cdot, \hat{\beta}/\hat{\alpha}), \quad Z_T^* = \sum_{m=1}^{M_T} F_m^*, \quad T^* = \sum_{m=1}^{M_T} L_m^*$$

で, ^は モーメント, またはモーメントの関数である量を, 対応する標本モーメントで置き換えたものである. このブートストラップ推定量に対して, 以下の定理を得た.

定理: 先の定理の仮定に加え, 上の設定のもと,

$$\sup_{z\in\mathbf{R}} |P_\mu[\sqrt{T}A(Z_T/T) \le z] - P_\mu^*[A_T^* \le z]|$$

は $O_{P_\mu}(T^{-1})$ となる. ここで, P_μ^* は観測 $(X_t)_{0\le t\le T}$ のもとでの条件付き確率を表す. ただし $M_T = 0$ の場合は $P_\mu^*[A_T^* \le z] = 0$ と解釈する.

OU 過程, CIR 過程などの MLE に関して, そのスチューデント化が我々の定理の条件を満たすことが証明できる. かくして我々は, エルゴード性を持つ一次元拡散過程に対する推定量の分布を評価する上で, 二次の正確さを持つ有力な手法を手に入れたことになる.

Metropolis-Hastings algorithms with acceptance ratios of nearly 1

東京大学大学院数理科学研究科　　鎌谷研吾

概要

積分計算の方法として, モンテカルロ法は頻繁に用いられる手法で, 様々な種類のモンテカルロ法が存在し, 適用されている. いま, Π は $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 上の確率密度関数 π をもつ確率測度とし, $L^1(\Pi)$ 関数 f にたいして $\Pi(f)$ を求めたいとする. このときシンプルなモンテカルロ法では, 分布が Π に従う独立確率変数列 $X_1, \ldots, X_N$ を作り, 経験分布 Π_N から $\Pi_N(f)$ を計算する. すると対数の法則により $\Pi_N(f)$ は $\Pi(f)$ に almost surely に収束する. 一方このような独立確率変数列を作る事が困難な場合や, 計算時間を考えると有効でない場合には Π に収束するマルコフチェイン $X_1, \ldots, X_N$ によって経験分布を作る方法が用いられる. このような方法はマルコフチェインモンテカルロ法と呼ばれ, マルコフチェインの作り方によって $\Pi_N(f)$ が $\Pi(f)$ に収束する十分条件が研究されてきた.

Π に分布が収束するマルコフチェインを作る一般的方法はメトロポリス-ヘイスティングス法と呼ばれるもので, Metropolis et al. [4] によって 1950 年代から研究されているアルゴリズムである. ランダムウォークを次の様に修正する事により得られる. まず $X_0 = x \in \mathbb{R}^d$ を勝手にとり, 独立確率変数列 $W_1, W_2, \ldots \sim q(y)dy$ によって

$$X_n = \begin{cases} Y_n = X_{n-1} + W_n & \text{with prob } \alpha(X_{n-1}, Y_n) \\ X_{n-1} & \text{otherwise} \end{cases} \tag{1}$$

とする. ただし, $\alpha : \mathbb{R}^{2d} \to [0,1]$ は acceptance ratio とよばれ, $\alpha(x,y) = \min\{1, \pi(x)/\pi(y)\}$ と定義される. このマルコフチェインは Π を不変分布として持ち, $\Pi_N(f)$ が $\Pi(f)$ に収束する様な Π の十分条件が調べられている (Roberts & Tweedie [6] など).

90 年代中頃から, 上記の方法とは異なったメトロポリス-ヘイスティングス法も研究されている. $B = (B_t; t \in \mathbb{R}^+)$ を標準ブラウン運動とすると, $b(x) = 2^{-1}\nabla \log \pi(x)$ にたいして

$$dX_t = b(X_t)dt + dB_t; X_0 = x \tag{2}$$

の解は Π を不変分布としてもつ. これを利用し, この解のオイラー丸山の離散化を上記の様に修正する事により新しいメトロポリス-ヘイスティングス法を定める事が出来る. この方法はメトロポリス修正を施したランジュバン法などと呼ばれていて, 離散化の施し方に依ってまたいくつかの方法が存在し, $\Pi_N(f)$ が $\Pi(f)$ に収束する様な十分条件が調べられている (Stramer & Tweedie [7, 8] など).

メトロポリス-ヘイスティングス法では, さらに次元の異なる空間の和集合を扱える, リバーシブルジャンプ法という方法も存在する (Green [3], Andrieu et al. [1] など) パラメータの数の推定も含めた推定に用いられている. さらに最近では以上の様な時間一様のマルコフチェインだけでなく, 複雑なアルゴリズムを提案し研究をすることが盛んである (Roberts & Rosenthal [5], Andrieu & Moulines [2] など).

以上に挙げた以外にも, ベイズ理論に有用なギブスサンプラー等様々なモンテカルロ法が存在し, 場合に依って適度な方法を選ぶ事が必要である. 今回の発表では, Π の裾が重い場合を想定し, そ

のときに適したアルゴリズムを上記の確率微分方程式のオイラー-丸山の離散化を利用して構成する．このアルゴリズムの理論的な収束オーダーの計算と数値計算により有用な場合と不都合な場合を検証する．

参考文献

[1] C. Andrieu, P. M. Djuric, and A. Doucet. Model selection by MCMC computation. *Signal Processing*, (81):19–37, 2001.

[2] C. Andrieu and E. Moulines. On the Ergodicity Properties of some Adaptive MCMC Algorithms. 2005.

[3] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[4] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[5] G. O. Roberts and J. S. Rosenthal. Coupling and Ergodicity of Adaptive MCMC. March 2005.

[6] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithm, 1994.

[7] O. Stramer and R. L. Tweedie. Diffusions with given stationary distiributions and their discretizations. *Methodology and Computing in Applied Probabilities*, 1(3):283–306, 1999.

[8] O. Stramer and R. L. Tweedie. Self-Targeting Candidates for Metropolis-Hastings Algorithms. *Methodology and Computing in Applied Probabilities*, 1(3):307–328, 1999.

離散観測されるジャンプ型確率過程のジャンプ検出法

大阪大学基礎工学研究科　清水 泰隆

1. 研究の動機と概要

フィルター付き確率空間 $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ 上の d-次元確率過程

$$X_t(\omega) = \xi(\omega) + \alpha_t(\omega) + \int_0^t \int_E \eta(s, X_{s-}, z)\, d\mu(ds \times dz) \tag{1}$$

を考える．ここに，$E = \mathbb{R}^d \setminus \{0\}$, ξ はある確率変数，$\mu(dt \times dz)$ はポアソンランダム測度で，そのコンペンセイター ν が存在して $\nu(dt \times A) = f_\theta(z)\, dtdz = \lambda_\theta F_\theta(z)\, dtdz$ と表されると仮定する．ただし，$\nu([0,1] \times E) < \infty$. また，$\eta(t,x,z)$ は $\mathbb{R}_+ \times \mathbb{R}^d \times E$ 上の可測関数で，条件 $|\eta(t,x,z) - \eta(t,y,z)| \lesssim |x-y|$, $|\eta(t,x,z)| \lesssim \zeta(z)(1+|x|)$, かつ $\int_E \zeta^2(z)\, dz < \infty$ を満たすものとする．ただし，$A \lesssim B$ は，ある正の定数 C に対して $A \leq CB$ であることを表す．さらに，α は $\alpha_0 = 0$ なる $\mathscr{F}_t$-適合な確率過程で，ある $a, b > 0$ と，任意の $s, t \in \mathbb{R}^d$ に対して

$$\mathbb{E}|\alpha_t - \alpha_s|^a \lesssim |t-s|^{d+b} \tag{2}$$

を満たすとする．したがって，α_t は連続なパスを持つ確率過程である．

このようなジャンプ型確率過程は，ファイナンス等の応用分野で頻繁に用いられつつあり，現実的な要請として離散観測による統計推測は必須問題である．すなわち，離散観測 $\{X_{t_i^n}\}_{i=0}^n$, $t_i^n = ih_n$ によるパラメータ推定問題が実務上重要である．

この問題に対して，特に，X がマルコフ型のジャンプ型拡散過程の場合について，Shimizu and Yoshida [3] が漸近有効な推定関数を提案した．そこでの基本的なアイデアは，$\Delta X_i^n = X_{t_i^n} - X_{t_{i-1}^n}$ の絶対値の大小により，区間 $[t_{i-1}^n, t_i^n)$ におけるジャンプの有無を判定しようというものであり，$n \to \infty$ となるほどその判定は確かになる．ところが，n の大きさに対する判定の正確さは個々のモデルに依存しており，$|\Delta X_i^n| > r_n$ という判定条件を作るに際して，いかなる列 r_n を選ぶかという問題が応用上最も重要な課題となる．ところが，漸近論における解答は $r_n = O(h_n^\rho)$ であり，これでは具体的に r_n を決めることはできないのである．

本報告では，上記の問題に対してひとつの方法論を提示する．すなわち，実データ $\{X_{t_i^n}\}_{i=0}^n$ が与えられたとき，各観測区間 $[t_{i-1}^n, t_i^n)$ におけるジャンプの有無を判別するフィルター

$$\mathscr{H}_i^n(r_n) = \{\omega\, ;\, |\Delta X_i^n| > r_n\} \qquad r_n = O(h_n^\rho) \tag{3}$$

における r_n の具体的な選択方法を提案する．

2. フィルターの選択基準と選択のアルゴリズム

推定精度のよいフィルターを得るために，モデル (1) 内のあるパラメータに着目する．母数

$$\lambda = \int_0^1 \int_E \nu(ds \times dz)$$

はジャンプの回数の強度を表す母数であり，モデルの misspecification の下でもこの母数の一致推定量は構成できて，定義関数 I に対して，

$$\hat{\lambda}_n(r_n) = \frac{1}{n} \sum_{i=1}^n I(\mathscr{H}_i^n(r_n)) \tag{4}$$

となる．更に，この推定量は $\mathbb{E}[\hat{\lambda}_n(r_n)] \to \lambda\ (n \to \infty)$ となることが知られいる．

そこで，この推定量 $\hat{\lambda}_n(r_n)$ の，n を固定した下での真値 λ に対するバイアスができるだけ小さくなるように r_n を求めることが考えられる．

ここで，$\epsilon_n(r_n) = \dfrac{1}{n}\sum_{i=1}^{n} \mathbb{P}\{|\Delta\alpha_i^n| > r_n,\ J_i^n = 0\}$ とおく．ただし，$\Delta\alpha_i^n = \alpha_{t_i^n} - \alpha_{t_{i-1}^n}$，$J_i^n = \#\{t \in [t_{i-1}^n, t_i^n);\ |\Delta X_i^n| > 0\}$ である．次の命題に注意しよう．

命題 1 $h_n \to 0$ に対して，$\ell_n := h_n^{-1}\epsilon_n(r_n) - \int_{|z|<\frac{r_n}{2}} f(z)\,dz + O(\delta_n)$ とする．ただし，δ_n は十分早く 0 に収束する列である．このとき，次の不等式が成り立つ：

$$\tilde{\ell}_n \le e^{\lambda h_n} b(r_n) \le \ell_n. \tag{5}$$

ここに，$\tilde{\ell}_n := \ell_n - \displaystyle\int_{\frac{r_n}{2}<|z|\le 2r_n} f(z)\,dz$ である．

この命題より，$\bar{\ell}_n + \ell_n \approx 0$ となるように r_n を選ぶのことが，我々の問題に対する一つの考え方であろう．すなわち，

$$\begin{aligned} L_n(r_n) &:= 2\ell_n(r_n) - \int_{\frac{r_n}{2}<|z|\le 2r_n} f(z)\,dz \\ &= 2h_n^{-1}\epsilon_n(r_n) - int(r_n), \end{aligned} \tag{6}$$

（ただし，$int(r_n) = \int_{|z|\le\frac{r_n}{2}} f(z)\,dz + \int_{|z|\le 2r_n} f(z)\,dz$）に対して，$L_n(r_n) = 0$ となるような r_n を求めることが考えられる．$\epsilon_n(r)$ は r に関して単調減少で，$\epsilon_n(0) = 1$，$\lim_{r\to\infty}\epsilon_n(r) = 0$，また，$int(r)$ は正値単調増加であるから，このような解 r_n は一意に存在する．

X がジャンプ型拡散過程の場合には，局所ガウス近似によって $\epsilon_n(r_n)$ を近似できて，また，f に対しては Shimizu [1] で提案されたある種のカーネル推定量 $\hat{f}_n$ を用いることにより，$L(r_n)$ の近似を作ることができる．しかし，ここではまだ，$\epsilon_n(r_n)$ や $\hat{f}_n$ の中に，$\mathscr{H}_i^n(r_n)$ によって推定されるべき母数が含まれることに注意しなければならない．

そこで，まず適当に r_n の初期値 $r_n^{(0)}$ を選んでおき，それを用いて $\epsilon_n(r_n)$ や $\hat{f}_n$ の推定量である $\epsilon_n^{(0)}(r_n)$ や $\hat{f}_n^{(0)}$ を構成し，$L(r_n)$ の近似を作る．これを $L_n^{(1)}(r_n) = 2h_n^{-1}\epsilon_n^{(0)}(r_n) - int^{(0)}(r_n)$ と書こう．この下で $L_n^{(1)}(r_n) = 0$ の解を $r_n^{(1)}$ とおき，これを第一段階の閾値と呼ぶ．この操作を繰り返すことにより，**第 k-段階の閾値** $r_n^{(k)}$ を

$$L_n^{(k)}(r_n) = 2h_n^{-1}\epsilon_n^{(k-1)}(r_n) - int^{(k-1)}(r_n) = 0 \tag{7}$$

の解として定義する．$k \to \infty$ とすることで，$r_n^{(k)}$ はある定数に収束するであろうことが，数値実験で予想される（実際は $k = 10$ 程で十安定する）．この結果として，$\mathscr{H}_i^n(r_n^{(k)})$ を用いて作られた推定量の精度は非常に高まることが実験的に示される．

この方法は，広いモデルに応用でき非常に実際的な方法で，実務上強力な道具となるであろう．

参考文献

[1] Shimizu, Y. (2005) Density type estimation of Lévy measure for discretely observed diffusion processes with jumps, *to appear in JJSS*.

[2] Shimizu, Y. (2005) Model Selection for Lévy Characteristics in Discretely Observed Jump-Diffusion Models from the Aspect of Information Criteria, submitted.

[3] Shimizu, Y. and Yoshida, N. (2005) Estimation of Parameters for diffusion processes with jumps from discrete observations, *to appear in SISP*.

Mighty convergence of statistical random fields and applications to stochastic differential equations

Nakahiro YOSHIDA
Graduate School of Mathematical Sciences, University of Tokyo,

Analysis of the likelihood ratio is the first key step in order to investigate the performance of statistics. After the initiation of the local asymptotic normality by Le Cam and Hájek, a new paradigm of analysis was established by Ibragimov and Has'minskii [1, 2, 3]. That is, all asymptotic properties of statistics in likelihood analysis could be reduced in the convergence of the random field formed by the likelihood ratios. This program was successfully implemented mainly for i.i.d. settings and white Gaussian noise models.

Let $\mathcal{E}^\epsilon = \{\mathcal{X}^\epsilon, \mathcal{A}^\epsilon, P^\epsilon_\theta, \theta \in \Theta\}$ be a sequence of statistical experiments with $\epsilon \in (0,1]$. Θ denotes a parameter space in $\mathbb{R}^m$. $\varphi(\epsilon)$ is a positive normalizing factor tending to zero as $\epsilon \downarrow 0$. For a $\theta_0 \in \Theta$, define a random field Z_ϵ by $Z_\epsilon(u) = \frac{dP^\epsilon_{\theta_0+\varphi(\epsilon)u}}{dP^\epsilon_{\theta_0}}(X^\epsilon)$ for $u \in \mathbb{R}^m$.

We assume the finite-dimensional convergence: $Z_\epsilon \to^{d_f} Z$, where Z is a $\hat{C}(\mathbb{R}^m)$-valued random variable.[1] Then from the large deviation inequality

$$P^\epsilon\left[Z_\epsilon^{1/2}(u) \geq e^{-c|u|^\gamma}\right] \leq e^{-c|u|^\gamma} \tag{1}$$

for some $\gamma > 0$ and $c > 0$,[2] Ibragimov-Has'minskii [1, 2, 3] proved the weak convergence $(P^\epsilon)^{Z_\epsilon} \to \mathcal{L}\{Z\}$, moreover, the inequality $P^\epsilon\left[\sup_{u:|u|\geq r} Z_\epsilon(u) > e^{-c_1 r^\gamma}\right] \leq e^{-c_1 r^\gamma}$. If $\hat{u}$ uniquely attains the maximum of $Z(u)$, then for any sequence of the maximum likelihood estimator $\hat{\theta}_\epsilon$ for θ, $\hat{u}_\epsilon := \varphi(\epsilon)^{-1}(\hat{\theta}_\epsilon - \theta_0) \to^d \hat{u}$ and $P^\epsilon[f(\hat{u}_\epsilon)] \to P[f(\hat{u})]$ for every $f \in C_\uparrow(\mathbb{R}^m)$.[3]

Convergence of moments or sharp estimate of the error probability of an estimator plays an essential role in the key steps of theoretical statistics. For mean bias correction to an estimator, we need the existence of the mean of it. The expected value of the plug-in functional P^ϵ [log likelihood $(\hat{u}_\epsilon)$] is necessary in prediction theory and in construction of information criteria. For example, the number of parameters appearing in AIC as the correction term is nothing but the mean square of $\hat{u}_\epsilon$ [4] It is impossible to develop the higher-order asymptotic theory without the following type estimate: $\mathrm{P}\left[|\hat{\theta}_n - \theta_0| > n^{-\beta}\right] \leq n^{-N}$ for $\beta < 1/2$. Also, for the same estimates for Bayes estimators, we need an estimate of the large deviation probability for the likelihood ratio random field. The necessity of the polynomial type large deviation inequalities in the theory of asymptotic expansion of the estimators for stochastic processes motivates this article (cf. Yoshida [15], Sakamoto and Yoshida [7, 8]).

Kutoyants found that Ibragimov-Has'minskii's scheme could work for stochastic processes including diffusion type processes and point processes. Among his many results, Kutoyants established a complete theory for processes with small diffusions (Kutoyants [4]). Many applications of this methodology were also presented in Kutoyants [5]. We should note that the Ibragimov-Has'minskii-Kutoyants' theory was effectively used in derivation of asymptotic expansions ([12, 14]).

As the core of the theory, the large deviation inequality (1) is not easy to obtain for most of stochastic processes, even for nonlinear ergodic diffusions; asymptotic theory for small perturbed systems is special in this sense. As seen in Kutoyants [4], for ergodic diffusions, is necessary an explicit expression of the moment

[1] $\hat{C}(\mathbb{R}^m)$ is the sapce of continuous functions that tends to zero at the infinity.

[2] P^ϵ denotes $P^\epsilon_{\theta_0}$. By convention, P^ϵ functions as the expectation for a random variable.

[3] The set of continuous functions on $\mathbb{R}^m$ of at most polynomial growth.

[4] In the literature of asymptotic statistics, unfortunately, such an expectation has been very often assumed to exist without any mathematical backing. Rigorous treatment of this problem under a certain integrability assumption can be seen, e.g., in Uchida and Yoshida [10].

generating function of a kind of deviation in the statistical model, although it implies very strong results such as the convergence of moments once it is obtained; strong assumptions possibly produce strong results! Without such a strong assumption, the convergence of statistical random fields was proved in [11], however without moment convergence.

It is an important observation, as some of statisticians have been aware and really it was implicitly written in Ibragimov and Has'minskii's papers, that the exponential type large deviation inequality like (1) is much stronger than our use. The polynomial type large deviation inequality is sufficient to develop a theory. Here the polynomial type large deviation inequality means: $P^{\epsilon}\left[\sup_{u:|u|\geq r} Z_{\epsilon}(u) \geq r^{-N}\right] \leq \frac{C_N}{r^N} \quad (r>0)$ because the rate of the convergence of the probability is of or faster than a polynomial order. Some exponential in place of r^{-N} on the left-hand side is very often possible.

Kutoyants [6] presented a polynomial type large deviation inequality for one-dimensional diffusions by means of the local time. A polynomial type large deviation inequality in a more abstract setting of the partially locally asymptotically quadratic (PLAQ) sequence of experiments (without any special properties belonging to diffusion processes) was presented in 2004. [5] From this result, for stochastic processes including nonlinear non-Gaussian time series models and semimartingales as well as multi-dimensional diffusion processes, it is possible to obtain new convergence results, e.g., the convergence of moments of the M-estimator, and the asymptotic normality of the Bayes estimator and convergence of moments of it. Our results also provide the same consequences even for estimators based on sampled data from diffusions with/without jumps[6].

As an example, we discuss estimation problem (an MLE type estimator and Bayesian estimators) under a high-frequency sampling scheme and show the convergence of moments as well as the asymptotic normality of them. The author hopes that our approach enables us to connect the Ibragimov-Has'minskii-Kutoyants' program in the asymptotic decision theory to the *quasi*-likelihood analysis that is inevitable in statistics for sampled continuous-time processes.

Our setting includes the so-called non-ergodic experiments. There, the limit $\hat{u}$ of the estimator does not necessarily admit moments; for example, the limit distribution of the maximum likelihood estimator has the Cauchy distribution in a simple example in the non-ergodic statistics. So no theory that ensures all moments should exist. We shall deal with the order of moment more carefully than in the case of exponential type large deviation inequality.

References

[1] Ibragimov, I.A., Has'minskii, R.Z.: The asymptotic behavior of certain statistical estimates in the smooth case. I. Investigation of the likelihood ratio (Russian) Teor. Verojatnost. i Primenen. 17 (1972), 469–486.

[2] Ibragimov, I.A., Has'minskii, R.Z.: Asymptotic behavior of certain statistical estimates. II. Limit theorems for a posteriori density and for Bayesian estimates. (Russian) Teor. Verojatnost. i Primenen. 18 (1973), 78–93.

[3] Ibragimov, I.A., Has'minskii, R.Z.: Statistical estimation: asymptotic theory. Springer, New York 1981

[4] Kutoyants, Yu. A.: Parameter estimation for stochastic processes. Translated and edited by B.L.S.Prakasa Rao, Berlin: Herdermann 1984

[5] Kutoyants, Yu.: Identification of dynamical systems with small noise. Dordrecht Boston London: Kluwer 1994

[6] Kutoyants, Yu. A.: Statistical inference for ergodic diffusion processes. Springer Series in Statistics. Springer-Verlag London, Ltd., London, 2004

[7] Sakamoto, Y., Yoshida, N.: Higher order asymptotic expansions for a functional of a mixing process and applications to diffusion functionals. unpublished manuscript (1999)

[8] Sakamoto, Y., Yoshida, N.: Asymptotic expansion formulas for functionals of epsilon-Markov processes with a mixing property. Annals of the Institute of Statistical Mathematics, 56, 545-597 (2004)

[9] Shimizu, Y.; Yoshida, N.: Estimation of diffusion processes with jumps from discrete observations. (2002) preprint, to appear in Statistical Inference for Stochastic Processes

[10] Uchida, M., Yoshida, N.: Information criteria in model selection for mixing processes Statistical Inference for Stochastic Processes, 4, 73-98 (2001)

[11] Yoshida, N.: Asymptotic behavior of M-estimator and related random field for diffusion process. Annals of the Institute of Statistical Mathematics, 42, No. 2, 221-251 (1990)

[12] Yoshida, N: Asymptotic expansion for small diffusions via the theory of Malliavin-Watanabe. Prob. Theory Related Fields. 92, 275-311 (1992)

[13] Yoshida, N: Estimation for diffusion processes from discrete observation. J. Multivar. Anal. 41, 220-242 (1992)

[14] Yoshida,N.: Asymptotic expansion of Bayes estimators for small diffusions. Probab. Theory Relat. Fields 95, 429-450 (1993)

[15] Yoshida, N.: Malliavin calculus and asymptotic expansion for martingales. Probab. Theory Relat. Fields 109, 301-342 (1997)

[16] Yoshida, N.: Asymptotic expansion for general M-estimation for stochastic differential equation with jumps. *Zenkin Tenkai 2003*, Hiroshima International University (2003)

[17] Yoshida, N.: Polynomial type large deviation inequality and its applications. preprint (2005)

[5] Details are in [17].

[6] The asymptotic normality of the M-estimator is already known for jump-diffusions; see Shimizu and Yoshida [9].

オイラー標数法による直交不変ランダム行列の最大固有値分布の近似

統計数理研　　栗木 哲
東京大情報理工　竹村彰通

1. チューブ法とオイラー標数法

添字集合 $I \subset \mathbb{R}^n$ を持つ確率場 $X(t)$ $(t \in I \subset \mathbb{R}^n)$ の最大値分布 $P(x) = P(\sup_{t\in I} X(t) \geq x)$ を近似するための方法として，チューブ法とオイラー標数法が知られている．チューブ法とは，平均 0，分散 1 の正規確率場に対して適用されるものであり，最大値分布と，確率場 $X(\cdot)$ の相関構造から適当な方法で定義されるリーマン部分多様体 M のまわりの「チューブ」の体積とが 1 対 1 に対応することから，チューブの体積評価を通して最大値分布を評価するというものである．チューブ法は，正則条件 (サンプルパスが滑らか，M の臨界半径が正) の下で，その近似誤差が指数的微小量となるという意味において正当 (valid) であることが示されている．

一方のオイラー標数法は，イクスカーション集合 $A_x = \{t \in I \mid X(t) \geq x\}$ のオイラー標数の期待値 $\widehat{P}(x) = E[\chi(A_x)]$ を x が大きいときに最大値の上側確率 $P(x)$ の近似値として用いるという方法である．とくに $X(\cdot)$ が正規確率場の場合には，チューブ法と等価であることが示されている．したがって正規確率場に対するオイラー標数法は正当である．しかしオイラー標数法自体は非正規確率場に対しても適用しうるものであり，その場合その近似が正当であるか，すなわち誤差が微小であるかどうかは一般にはよく分かっていない．本報告では，オイラー標数法の誤差評価の一般論に向けた試みとして，一つの例題を通してオイラー標数法が正当であるかどうかを見てゆくことにする．

2. 直交不変な対称ランダム行列の最大固有値の分布

$p \times p$ 実対称ランダム行列 $W = (w_{ij})$ でルベーグ測度 $dW = \prod_{i\geq j} dw_{ij}$ に対して次の形の密度関数 $f(W)$ をもつものを考える．

$$f(W) = \prod_{i=1}^{p} g(l_i) \quad (l_1 \geq \cdots \geq l_p \text{ は } W \text{ の固有値}).$$

添字集合を単位球面 $\mathbb{S}^{p-1}$ とする確率場を 2 次形式 $X(h) = h'Wh$, $h \in \mathbb{S}^{p-1}$ によって定義し，その最大値，すなわち最大固有値の上側確率 $P(x) = P(\lambda_{\max}(W) \geq x)$ を考える．

$g(\cdot)$ に関して次の仮定をおく．

仮定 1 $a = \inf\{l \mid g(l) > 0\}$, $b = \sup\{l \mid g(l) > 0\}$ とおく．$g(\cdot)$ は (a, b) で非負かつ区分的連続．$c < b$ が存在し，$l \in (c, b)$ で $g(l) > 0$.

最大固有値の上側確率のオイラー標数法近似 $\widehat{P}(x)$ として次の形を得る．

定理 1 $\mathrm{tr}_j(\cdot)$ を j 次基本対称式とする．また $\Omega_i = \mathrm{Vol}(\mathbb{S}^{i-1}) = 2\pi^{i/2}/\Gamma(i/2)$ とおく．
(i) $b=\infty$ の場合．

$$\begin{aligned}\widehat{P}(x) &= \prod_{i=1}^{p}\frac{\Omega_i}{2}\sum_{j=0}^{p-1}(-1)^j\int_x^\infty l_1^{p-1-j}g(l_1)dl_1\int_{\infty>l_2>\cdots>l_p>a}\mathrm{tr}_j(l_2,\ldots,l_p)\\ &\quad\times\prod_{2\le i<j\le p}(l_i-l_j)g(l_2)\cdots g(l_p)dl_2\cdots dl_p.\end{aligned}$$

(ii) $b<\infty$ の場合．

$$\begin{aligned}\widehat{P}(x) &= \prod_{i=1}^{p}\frac{\Omega_i}{2}\sum_{j=0}^{p-1}(-1)^j\int_x^b (b-l_1)^j g(l_1)dl_1\int_{b>l_2>\cdots>l_p>a}\mathrm{tr}_{p-1-j}(b-l_2,\ldots,b-l_p)\\ &\quad\times\prod_{2\le i<j\le p}(l_i-l_j)g(l_2)\cdots g(l_p)dl_2\cdots dl_p.\end{aligned}$$

また式中の和 $\sum_{j=0}^{p-1}$ の各項は j が大きいほど $x\uparrow b$ のとき漸近的に微小となる．(すなわち和 $\sum_{j=0}^{p-1}$ は，$j=0$ の項を主項とする漸近展開を表している．)

オイラー標数法の近似誤差を $\Delta P(x)=\widehat{P}(x)-P(x)$ とおく．$x\uparrow b$ のときに $\Delta P(x)$ が $\widehat{P}(x)$ よりも漸近的に小さいとき，すなわち $\Delta P(x)=o(\widehat{P}(x))$ のとき，オイラー標数法近似は弱い意味で正当ということにする．また $x\uparrow b$ のときに $\Delta P(x)$ が漸近展開式 $\widehat{P}(x)$ の各項よりも漸近的に微小であるとき，このオイラー標数法近似は正当ということにする．

定理 2 仮定 1 の下で，$b=\infty, b<\infty$ の両者においてオイラー標数法は弱い意味で正当である．またさらに $p=2$ であるか，$p\ge 3$ かつ

$$\lim_{x\to\infty}x^{2p-2}g(x)=0 \qquad (b=\infty \text{ の場合}),$$
$$\lim_{x\uparrow b}(b-x)^{-(p-3)}g(x)=0 \qquad (b<\infty \text{ の場合})$$

であるならば，オイラー標数法は正当である．

$x\uparrow b$ のときの $g(x)\to 0$ の収束が十分に速ければ，すなわち W の密度関数がそのサポートの上界付近で十分に速く 0 に収束するならば，オイラー標数法は正当となる．

3. 例題

多変量解析において標準的な直交不変ランダム行列であるウィシャート行列 $W_p(n)$ $(n>p-1)$，多変量ベータ行列 $B_p(n_1/2,n_2/2)$ $(n_1,n_2>p-1)$，および逆ウィシャート行列 $W_p^{-1}(n)$ $(n>p-1)$ について，その最大固有値分布のオイラー標数法近似と誤差の漸近評価を具体的に与えることができる．定理 2 より，これらの 3 つの例のどの場合についてもオイラー標数法は弱い意味で正当であるが，とくにウィシャート行列ではいかなる p,n についても正当，多変量ベータ行列，逆ウィシャート行列ではそれぞれ $3p<n_2+5$，$3p<n+5$ のとき正当となる．近似公式の具体的な形と，その誤差の漸近的，数値的評価については当日説明する．

Asymptotic minimaxity of Bayesian prediction

Fumiyasu Komaki
Department of Mathematical Informatics
Graduate School of Information Science and Technology, the University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, JAPAN
komaki@mist.i.u-tokyo.ac.jp

Summary

We investigate decision theoretic properties such as minimaxity and admissibility of Bayesian predictive distributions in the framework of asymptotic theory. Kullback-Leibler divergence from the true distribution to a predictive distribution is adopted as a loss function.

1. Introduction

Suppose that we have a set of independent observations $x^{(N)} = (x(1), x(2), \ldots, x(N))$ from a distribution with density $p(x|\theta)$ that belongs to a model $\{p(x|\theta)|\theta = (\theta_1, \theta_2, \ldots, \theta_d) \in \Theta\}$. An unobserved variable $y := x(N+1)$ from the same distribution $p(y|\theta)$ is predicted by using a predictive density $\hat{p}(y; x^{(N)})$.

We adopt the Kullback-Leibler divergence $D\{p(y|\theta), \hat{p}(y; x^{(N)})\} := \int p(y|\theta) \log\{p(y|\theta) / \hat{p}(y; x^{(N)})\} \mathrm{d}y$, which has a natural information theoretic meaning, as a loss function. We evaluate the performance of predictive distributions by using the risk function $\mathrm{E}[D(p, \hat{p})|\theta] = \int p(x^{(N)}|\theta) \int p(y|\theta) \log\{p(y|\theta)/\hat{p}(y; x^{(N)})\} \mathrm{d}y \mathrm{d}x^{(N)}$.

A widely used method to construct a predictive density is to use a plug-in density $p(y|\hat{\theta}(x^{(N)}))$, where $\hat{\theta}(x^{(N)})$ is an appropriate estimator of θ. However, Bayesian predictive densities

$$p_\pi(y|x^{(N)}) = \frac{\int p(y|\theta) p(x^{(N)}|\theta) \pi(\theta) \mathrm{d}\theta}{\int p(x^{(N)}|\bar{\theta}) \pi(\bar{\theta}) \mathrm{d}\bar{\theta}}$$

have better performance than plug-in distributions in many examples (Aitchison and Dunsmore (1975), Geisser (1993), Komaki (1996)).

In the present paper, we investigate the use of shrinkage priors for constructing Bayesian predictive distributions asymptotically dominating those based on improper vague priors such as the Jeffreys prior.

In the present paper, constructing methods for shrinkage priors are introduced and properties of them are investigated from the viewpoint of information geometry by using the results of previous studies on asymptotic properties of predictive distributions (Vidoni (1995), Komaki (1996), Hartigan (1998), Datta et al. (2000)). The model $\{p(x|\theta)|\theta \in \Theta\}$ is regarded as a manifold, and the relation between differential geometric properties of the model manifold and the existence of shrinkage priors is studied.

2. Shrinkage priors asymptotically dominating the Jeffreys prior

First, we prepare some differential geometric notions and notations to be used. In the following, we assume that the model manifold M is a $d(\geq 2)$-dimensional connected and orientable C^∞ manifold. The parameter space Θ is regarded as a coordinate system of M. We use Einstein's summation convention: if an index occurs twice in any one term, once as an upper and once as a lower index, summation over that index is implied.

The Fisher metric tensor is defined by $g_{ij}(\theta) := \mathrm{E}[\partial_i \log p(x|\theta) \partial_j \log p(x|\theta)|\theta]$, where $\partial_i := \partial/\partial\theta^i$. The coefficients of the α-connection are defined by $\overset{\alpha}{\Gamma}{}_{ij}^{\ k}(\theta) := \Gamma_{ij}^{\ k}(\theta) - \frac{\alpha}{2} T_{ijl}(\theta) g^{kl}(\theta)$, where $\Gamma_{ij}^{\ k} := \frac{1}{2}(\partial_i g_{jl}(\theta) + \partial_j g_{li}(\theta) - \partial_l g_{ij}(\theta)) g^{kl}(\theta)$ are the coefficients of the Riemannian connection, $T_{ijk}(\theta) := \mathrm{E}[\partial_i \log p(x|\theta) \partial_j \log p(x|\theta) \partial_k \log p(x|\theta)|\theta]$ is the skewness tensor, and g^{ij} denotes the (i,j)-component of the inverse matrix of (g_{ij}), see Amari (1985) for details. The -1-connection and 1-connection are called the m-connection and e-connection and the coefficients of them are denoted by $\overset{\mathrm{m}}{\Gamma}{}_{ij}^{\ k}$ and $\overset{\mathrm{e}}{\Gamma}{}_{ij}^{\ k}$, respectively. The α-covariant derivative of a vector field v is defined by $\overset{\alpha}{\nabla}_i v^j = \partial_i v^j + \overset{\alpha}{\Gamma}{}_{ik}^{\ j} v^k$, and $\overset{0}{\nabla}$ and $\overset{1}{\nabla}$ are denoted by ∇ and $\overset{\mathrm{e}}{\nabla}$, respectively.

The Laplacian Δ on a manifold (M,g) endowed with a Riemannian metric g_{ij} is defined by $\Delta f = |g|^{-1/2} \partial_i (|g|^{1/2} g^{ij} \partial_j f) = \nabla_i (g^{ij} \partial_j f)$, where f is a real function on M, and $|g|$ is the determinant of the matrix (g_{ij}).

Theorem 1 (Komaki 2006). Let (M,g) be a model manifold endowed with the Fisher metric. If a Green function $G(\xi,\theta)$ on (M,g) exists, there exist Bayesian predictive distributions asymptotically dominating the Bayesian predictive distribution based on the Jeffreys prior $\pi_{\mathrm{J}}(\theta) \propto |g(\theta)|^{1/2}$. In particular, the Bayesian predictive distribution based on the prior $\pi_{\mathrm{G}}(\theta)\mathrm{d}\theta := G(\xi,\theta)\pi_{\mathrm{J}}(\theta)\mathrm{d}\theta$, where $\xi \in M$ is an arbitrary fixed point, asymptotically dominates the Bayesian predictive distribution based on the Jeffreys prior. □

From the relation $(1/2) g^{ij} \partial_i \log(f/\pi_{\mathrm{J}}) \partial_j \log(f/\pi_{\mathrm{J}}) - (\pi_{\mathrm{J}}/f)\Delta(f/\pi_{\mathrm{J}}) = -2(\pi_{\mathrm{J}}/f)^{1/2} \Delta (f/\pi_{\mathrm{J}})^{1/2}$, we have the following theorem. For the definition of superharmonic functions on Riemannian manifolds, see e. g. Ito (1964). A C^2 function is superharmonic if and only if $\Delta f \leq 0$.

Theorem 2 (Komaki 2006). Let $f(\theta)$ be a smooth prior density on a model manifold (M,g) endowed with the Fisher metric. The Bayesian predictive distribution based on $f(\theta)$ asymptotically dominates the Bayesian predictive distribution based on the Jeffreys prior $\pi_{\mathrm{J}}(\theta)$ if and only if $(f/\pi_{\mathrm{J}})^{1/2}$ is a non-constant positive superharmonic function on (M,g). □

References

AITCHISON, J. and DUNSMORE, I. R. (1975). *Statistical Prediction Analysis.* Cambridge: Cambridge University Press.

AMARI, S. (1985). *Differential-Geometrical Methods in Statistics.* New York: Springer-Verlag.

DATTA, G. S., MUKERJEE, R., GHOSH, M., and SWEETING, T. J. (2000). Bayesian prediction with approximate frequentist validity. *Annals of Statistics* **28** 1414-1426.

EISENHART, L. P. (1949). *Riemannian Geometry*, 2nd edition. Princeton: Princeton University Press.

GEISSER, S. (1993). *Predictive Inference: an Introduction.* New York: Chapman and Hall.

HARTIGAN, J. A. (1998). The maximum likelihood prior. *Annals of Statistics* **26** 2083-2103.

ITÔ, S. (1964). On existence of Green function and positive superharmonic functions for linear elliptic operators of second order. *J. Math. Soc. Japan* **16** 299-306.

KOMAKI, F. (1996). On asymptotic properties of predictive distributions. *Biometrika* **83** 299-313.

KOMAKI, F. (2006). Shrinkage priors for Bayesian prediction, *the Annals of Statistics*, **34**, to appear.

VIDONI, P. (1995). A simple predictive density based on the p^*-formula. *Biometrika* **82** 855-863.

ゲーム論的確率論におけるカルバック情報量

竹村 彰通 (東京大学大学院情報理工学系研究科)
公文 雅之 (統計数理研究所リスク解析戦略研究センター)
竹内 啓 (明治学院大学国際学部)

1. 測度論的確率論の役割

確率論は1930年代のコルモゴロフによる測度論的確率論の定式化により，その数学的基礎が明らかにされ，その上で一つの確立した数学分野として大きく発展して来た．測度論的確率論の定式化以前には，フォン・ミーゼスの「コレクチーフ」の概念など，頻度論的な確率論の基礎づけが様々な形で試みられていたが，いずれもあまり成功せず，測度論的確率論の定式化とともにそれらはほとんど忘れ去られることとなった．コルモゴロフによる測度論的確率論では，確率自体は「点」や「線」のように公理的に与えられ，その意味を問わないことが特徴となっている．確率を公理として与え，確率論を確率の操作の体系とすることによって，数学としての確率論が定式化されたと理解できる．伊藤清の『確率論の基礎』初版 (1944)[1] への序文にある

> “確率とは，ルベーグ測度である．” この言葉ほど確率の数学的本質を突いたものはない．

という言明が，このような事情を的確に表していると思われる．

ところが確率の実質的な意味を問わないという態度は一種の思考の停止をも意味している．実はコルモゴロフ自身，彼の確立した測度論的確率論の定式化に一種のためらいを感じていたのである．コルモゴロフ (1963)[2] の一節を引用すれば以下のようである．

> The set-theoretic axioms of probability theory, in whose formulation it was my lot to take part, allowed to eliminate most of the difficulties in constructing the mathematical apparatus appropriate for numerous applications of probabilistic methods, and so successfully, that the problem of finding the causes of the applicability of mathematical probability theory was felt by many researches to be of secondary importance.
>
> I have already expressed the point of view that the basis of the applicability of the mathematical theory of probability to random events of the real world is the *frequency approach to probability* in one form or another, which was so strongly advocated by von Mises.

コルモゴロフはこのような観点に立って，自身は Kolmogorv complexity の概念の研究に進んでいくのだが，これも必ずしも成功したとは言えなかったと思われる．

以上のような状況で，ほとんどの確率論研究者は測度論的確率論以外に確率論の体系はあり得ないと考えている．そこに現れたのが Vladimir Vovk と Glenn Shafer によるゲーム論的確率論の枠組である．

2. ゲーム論的確率論の定式化

Vladimir Vovk は1960年の生まれで，学位論文をコルモゴロフの指導のもとに書いたという意味で，コロモゴロフの最後の弟子の一人である．Vovk は90年代にゲーム論的確率論の構想を次第に具体化していたが，Glenn Shafer というすぐれた共著者を得て，2001年に “Probability and Finance, It's Only a Game!” という本を出版した．

そこでは，Skeptic とよばれる賭をする人と，Reality とよばれる賭の結果を定める人の，二人のプレーヤーの間のゲームを設定することにより，ゲームの結果として確率が定まることが示されている．注目すべきは，測度論無しに，大数の強法則，中心極限定理，重複対数の法則，さらに数理ファイナンスにおける価格付けの諸公式，などが証明される点にある．しかもそれらの証明は，しばしば測度論における対応する証

明より簡明である．実際，ゲーム論的な大数の強法則の証明はほとんど何も数学的な前提を必要とせず高校生でも理解できるものとなっており，測度論的な証明の複雑さと比較すると雲泥の差がある．

以下で，最も簡単なコイン投げの場合のゲームの具体形を紹介しよう．Skeptic の初期資金を $\mathcal{K}_0 = 1$ と基準化する．以下すべて金額は実数と考え，いくらでも小さい額の賭金が可能であるとする．ゲームのラウンドを n で表し，ゲームの繰り返しを $n = 1, 2, \ldots$ と表す．各ラウンドにおいて，まず Skeptic が自身の賭金 $M_n \in \mathbb{R}$ を明らかにする．それを知った上で，Reality は $x_n = 1$ あるいは $x_n = -1$ を選ぶ．この場合の Skeptic への支払い (ペイオフ) は $M_n x_n$ である．$\mathcal{K}_n$ をラウンド n 終了後の Skeptic の資金とすれば

$$\mathcal{K}_n = \mathcal{K}_{n-1} + M_n x_n$$

となる．以上で注意すべきは，Reality は $x_n = \pm 1$ の符号を M_n の符号と逆に選ぶことができるから，Skeptic は必ず損をするように思われるところにある．ところが実は次の形の定理 (大数の強法則) が成り立つ．

定理　Skeptic に戦略 $\mathcal{P}$ が存在し Reality が $\lim_{n\to\infty}(1/n)(x_1 + \cdots + x_n) = 0$ が成り立たないように行動する場合には必ず $\lim_{n\to\infty} \mathcal{K}_n = \infty$ となる．

最近我々[3][4] はこの定理に関する初等的かつ非常に簡明な証明を与えている．この定理は，極めて簡単なゲームの設定のもとで，実は Reality(現実) はあたかも確率的に行動しなければならない，ということを示している．ゲームの設定の中には確率的な要素は全くないにもかかわらず，確率的行動が帰結するところが重要である．

さらに [3][4] においては次の事実が示されている．それは大数法則が成り立たず $(1/n)(x_1 + \cdots + x_n) \not\to 0$ となる場合には，$\log \mathcal{K}_n$ は $\{x_1, x_2, \ldots\}$ の経験分布とリスク中立確率との間のカルバック情報量に支配される (精度良く近似できる) という事実である．ゲーム論的確率論では，このように測度論的確率論では測度 0 の集合として無視されてしまう標本空間上での確率過程の挙動の分析が可能となる．測度 0 の集合は言わば特異点の集合であり，特異点での挙動を記述できる枠組は非常に有効である．

ゲーム論的確率論の枠組は現時点ではまだ広く認知されているとは言いがたい．しかしながら，哲学的な貢献にとどまらず，ゲーム論的確率論の枠組に基づいて新しい予測アルゴリズム [8][9] が開発されるなど，工学的にも意味のある結果が得られてきており，筆者は今後の発展が多いに期待できるものと考えている．

参考文献

[1] 伊藤 清，確率論の基礎 [新版]，岩波書店，2004.

[2] Andrei N. Kolmogorov. On tables of random numbers. *Sankhya*, Series A, 25:369–376. 1963.

[3] Masayuki Kumon and Akimichi Takemura. On a simple strategy weakly forcing the strong law of large numbers in the bounded forecasting game. Technical Report METR 05-20, University of Tokyo, 2005.

[4] Masayuki Kumon, Akimichi Takemura and Kei Takeuchi. Capital process and optimality properties of Bayesian Skeptic in the fair and biased coin games. Technical Report METR 05-32, University of Tokyo, 2005.

[5] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game!.* Wiley, New York, 2001.

[6] Akimichi Takemura and Taiji Suzuki. Game theoretic derivation of discrete distributions and discrete pricing formulas. Technical Report METR 05-25, University of Tokyo, 2005.

[7] 竹内 啓，賭けの数理と金融工学 ゲームとしての定式化，サイエンス社，2004.

[8] Vladimir Vovk, Akimichi Takemura, and Glenn Shafer. Defensive forecasting. in *Proceedings of the tenth international workshop on artificial intelligence and statistics.* R.G.Cowell and Z.Ghahramani editors, 365–372, 2005. (Available electronically at http://www.gatsby.ucl.ac.uk/aistats/)

[9] Vladimir Vovk, Ilia Nouretdinov, Akimichi Takemura, and Glenn Shafer. Defensive forecasting for linear protocols. in *Proceedings of the Sixteenth International Conference on Algorithmic Learning Theory* (ed. by Sanjay Jain, Hans Ulrich Simon, and Etsuji Tomita), LNAI 3734, Springer, Berlin, 459–473, 2005.

Bhattacharyya 型情報不等式について

田中 秀和　大阪府立大工学研究科
赤平 昌文　筑波大数理物質科学研究科

1. はじめに

統計的推定論において, 推定量のリスクを評価する際に情報不等式が重要な役割を果たす. 特に, 不偏推定量の分散に対する下界を与える式として, Cramér-Rao (C-R) の不等式及びこれを精密化した Bhattacharyya の不等式はよく知られている ([Z71]). また C-R の下界を達成する分布は指数型分布に限るということはよく知られているのに対し, Bhattacharyya の下界 (B-下界) を達成する分布については必ずしも明確にはなっていない. しかし, 最近, 指数型分布族の混合分布族等でも B-下界を達成することが分かってきた ([TA03], [T03a], [T03b]).

本論では位置, 尺度母数分布族において B-下界を達成する分布を特定する. また, 逐次推定の下で漸近不偏推定量の Bhattacharyya 下界を導出し, その達成についても論じる.

2. Bhattacharyya の下界

まず $\mathcal{X} \subset \mathbb{R}^r$ に対し, 標本空間 $(\mathcal{X}, \mathcal{B})$ 上の確率分布族 $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ は, ある σ-有限測度 μ に対して絶対連続とし, μ に対する確率密度関数を $f(x,\theta) := dP_\theta/d\mu$ $(\theta \in \Theta)$ とする. このとき, 標本 X に基づいて, 不偏推定可能な関数 $g(\theta)$ の推定問題について考える. ここで Θ を $\mathbb{R}^1$ の開区間, $\hat{g}(X)$ を $g(\theta)$ の不偏推定量とし, $f, g, \hat{g}$ に関して微分と積分の順序交換可能性等の適当な正則条件を仮定する.

いま, $\partial_l := \{\partial^l/(\partial\theta_1^{i_1} \cdots \partial\theta_r^{i_r}) | 0 \le i_1, \dots, i_r, i_1 + \cdots + i_r = l\}$ を微分作用素とし, $\boldsymbol{D}_k := {}^t(\partial_0, \partial_1, \dots, \partial_k)$, $\phi_k(x,\theta) := \boldsymbol{D}_k f(x,\theta)/f(x,\theta)$, $\mathrm{g}_k(\theta) := \boldsymbol{D}_k g(\theta)$ と表す. これらは $K(:= (k+r)!/(r!k!))$ 次ベクトルである. このとき (拡張) Fisher 情報量行列を $I_k^*(\theta) := E_\theta[\phi_k(X,\theta)\, {}^t\phi_k(X,\theta)]$ で定義する. このとき, 正則条件の下で, 任意の正の整数 k に対して, $I_k^*(\theta)$ が Θ で正則であれば, $E_\theta[\hat{g}^2(X)] \ge {}^t\mathrm{g}_k(\theta) I_k^{*-1}(\theta) \mathrm{g}_k(\theta)$ がすべての $\theta \in \Theta$ に対して成り立ち, 等号は $\hat{g}(x) = {}^t\tilde{a}_k(\theta)\phi_k(x,\theta)$ が μ-$a.a.x$ とすべての $\theta \in \Theta$ に対して成り立つ場合に限る. ただし, $\tilde{a}_k(\theta) = {}^t(a_{k0}(\theta), \dots, a_{k,K-1}(\theta)) := I_k^{*-1}(\theta)\mathrm{g}_k(\theta)$ とする. ここで, $I_k(\theta)$ を通常の Fisher 情報量行列とすると

$$I_k^*(\theta) = \begin{pmatrix} 1 & {}^t\boldsymbol{0} \\ \boldsymbol{0} & I_k(\theta) \end{pmatrix}$$

となり, $V_\theta(\hat{g}(X)) \ge ({}^t\partial_1, \dots, {}^t\partial_k) g(\theta) I_k^{-1}(\theta)\, {}^t({}^t\partial_1, \dots, {}^t\partial_k) g(\theta) =: B_k(\theta)$ となる. これが通常の Bhattacharyya (B-) 不等式であり, $B_k(\theta)$ は k 次の B-下界と呼ばれている. 特に $B_1(\theta)$ は C-R の下界と一致する. なお, k 次の B-下界 $B_k(\theta)$ は C-R 下界の精密化となっている.

3. 位置, 尺度母数分布族における B-下界

本節においては, [T03b] に従って, 位置, 尺度母数分布族における B-下界の達成について考える. (i) 位置母数分布族, つまり $f(x,\xi) = f(x-\xi)$, $\mathcal{X} = \Theta = \mathbb{R}^1$ の場合に次のことが成り立つ.

定理 1. 位置母数分布族で B-下界を達成する分布は, p.d.f. が次で与えられる正規分布と指数ガンマ分布に限る:

$$f(x,\xi) = \frac{1}{\sqrt{2\pi}b}\exp\left\{-\frac{(x-\xi-a)^2}{2b^2}\right\} \quad (x,\xi \in \mathbb{R}^1; a \in \mathbb{R}^1, b \in \mathbb{R}_+),$$
$$f(x,\xi) = \frac{|b|}{\Gamma(a)c^a}\exp\left\{-\frac{e^{b(x-\xi)}}{c} + ab(x-\xi)\right\} \quad (x,\xi \in \mathbb{R}^1; a, c \in \mathbb{R}_+, b \in \mathbb{R}^1 \setminus \{0\}).$$

(ii) 尺度母数分布族, つまり $f(x,\sigma) = f(x/\sigma)/\sigma$, $\mathcal{X} = \mathbb{R}^1$, $\Theta = \mathbb{R}_+$ の場合, $y_\pm := \log(\pm x)\chi_{\{x \gtrless 0\}}(x)$, $\tau := \log\sigma$ と変換することにより位置母数分布族の場合に帰着できて次のことを得る.

定理 2. 尺度母数分布族で B-下界を達成する分布は, p.d.f. が次で与えられる対数正規分布 $\mathrm{LN}(a,b^2)$ $(a \in \mathbb{R}^1, b \in \mathbb{R}_+)$ 同士の混合分布と拡張ガンマ分布 $\mathrm{Ga}(a,b,c)$ $(a, b \in \mathbb{R}_+, c \in \mathbb{R}^1 \setminus \{0\})$ 同士の混合分布に限る:

$$f(x,\sigma) = \frac{1}{\sqrt{2\pi}bx}\exp\left\{-\frac{1}{2b^2}\left(\log\frac{x}{\sigma} - a\right)^2\right\} \quad (x,\sigma \in \mathbb{R}_+; a \in \mathbb{R}^1, b \in \mathbb{R}_+),$$
$$f(x,\sigma) = \frac{|b|}{\Gamma(a)c^a\sigma}\left(\frac{x}{\sigma}\right)^{ab-1}\exp\left\{-\frac{1}{c}\left(\frac{x}{\sigma}\right)^b\right\} \quad (x,\sigma \in \mathbb{R}_+; a, c \in \mathbb{R}_+, b \in \mathbb{R}^1 \setminus \{0\}).$$

定理 3. 位置尺度母数分布族で B-下界を達成する分布は正規分布 $\mathrm{N}(\xi + a\sigma, b^2\sigma^2)$ に限る.

4. 漸近理論における B-下界

前節の議論から分かるように, 小標本の場合には一般の母集団分布族の下で, B-下界は達成され得ない. そこで, 大標本の場合を考えるが, そもそも B-下界は 2 次の次数まで考えることによって出現するものであり, この下界を達成することは難しい. それを可能にするためには逐次推定を考える必要があり, ここではそれについて論じる ([TA88]). まず, $X_1, X_2, \ldots, X_n, \ldots$ をたがいに独立に, いずれも (σ-有限測度 μ に関する) 密度 $f(x, \theta)$ をもつ分布に従う確率変数列とする. ただし, $\theta \in \Theta$ で Θ は開区間とする. ここで標本の大きさ n はある逐次法則に従って定められているものとする. いま, $\nu^*(\theta) := E_\theta(n)$ とし, $\nu^*(\theta)$ を θ について一様に大きくなるようにして, θ の逐次推定方式の B-下界を達成するという意味での 2 次の漸近有効性について考える. そこで $E_\theta(n) = \nu(\theta) + o(1)$, $V_\theta(n)/\nu(\theta) = O(1)$, $E_\theta(n^k)/\nu^k(\theta) = O(1)$ $(k = 2, 3, 4)$, $\{(\partial/\partial\theta)\nu(\theta)\}/\nu(\theta) = O(1)$, $\{(\partial^2/\partial\theta^2)\nu(\theta)\}/\nu(\theta) = O(1)$ とし, また, 各 $\theta \in \Theta$ について, $0 < I(\theta) := E_\theta[\{l^{(1)}(\theta, X)\}^2] = -E_\theta[l^{(2)}(\theta, X)] < \infty$ で, $J(\theta) := E_\theta[l^{(1)}(\theta, X) l^{(2)}(\theta, X)]$, $K(\theta) := E_\theta[\{l^{(1)}(\theta, X)\}^3]$, $M(\theta) := E_\theta[\{l^{(2)}(\theta, X)\}^2] - I^2(\theta)$, $N(\theta) := E_\theta[\{l^{(1)}(\theta, X)\}^2 l^{(2)}(\theta, X)] + I^2(\theta)$ とし, さらに適当な正則条件を仮定する.

定理 4. 適当な正則条件の下で, θ の任意の漸近不偏推定量 $\hat{\theta}_n$ (すなわち, $E_\theta(\hat{\theta}_n) = \theta + o(1/\nu)$) について

$$V_\theta\left(\sqrt{\nu(\theta)}(\hat{\theta}_n - \theta)\right) \geq \frac{1}{I(\theta)} + \frac{1}{2\nu(\theta)I^2(\theta)}\left\{\frac{J(\theta)+K(\theta)}{I(\theta)} + \frac{2\nu'(\theta)}{\nu(\theta)}\right\}^2 + o\left(\frac{1}{\nu(\theta)}\right)$$

が成り立つ.

次に, $\hat{\theta}_{ML}$ を標本 $(X_1, \ldots, X_n)$ に基づく (θ の) 最尤推定量 (MLE) とし, $\hat{\theta}^*_{ML}$ を $E_\theta(\hat{\theta}^*_{ML}) = \theta + o(1/\nu)$ となるように漸近的に偏り補正した MLE とする.

定理 5. 適当な正則条件の下で, $-\sum_{i=1}^n l^{(2)}(\hat{\theta}^*_{ML}, X_i) = \nu(\hat{\theta}^*_{ML}) I(\hat{\theta}^*_{ML}) + c(\hat{\theta}^*_{ML})$ となる n をとれば, $\hat{\theta}^*_{ML}$ の漸近分散は, 定理 4 の B-下界に一致し, その意味で逐次補正最尤推定方式は 2 次の漸近的有効である. ただし

$$c(\theta) = \frac{J(\theta)\nu'(\theta)}{I(\theta)\nu(\theta)} - \frac{\nu''(\theta)}{2\nu(\theta)} + \frac{1}{2I(\theta)}\{M(\theta) + N(\theta)\}$$

とする.

注意. 非逐次の場合には, $\hat{\theta}^*_{ML}$ の漸近分散は, B-下界に $\{I(\theta)M(\theta) - J^2(\theta)\}/\{nI^4(\theta)\}$ の項が追加されて, 補正 MLE の B-下界は達成されないが, 逐次の場合には停止則から情報を得ることによってその項が消えて, 逐次補正最尤推定方式の 2 次の漸近有効性がいえる. 上記の結果は, 局外母数が存在する場合に拡張され ([AT91]), また逐次差分尤度推定量の 2 次の漸近有効性も得られ, 線形結合母数の逐次推定にも拡張されている ([A95], [A97]).

参考文献

[A95] Akahira, M. (1995). The Bhattacharyya type bound for the asymptotic variance and the sequential discretized likelihood estimation procedure. *Sequential Analysis* **14**(3), 193–204.

[A97] Akahira, M. (1997). An information inequality bound for the asymptotic variance of sequential estimation procedures of a linearly combined parameter and its attainment. *Sequential Analysis* **16**(1), 47–63.

[AT91] Akahira, M. and Takeuchi, K. (1991). Second order asymptotic variance of sequential estimation procedures in the presence of nuisance parameters. *Sequential Analysis* **10**(1&2), 27–43.

[TA88] Takeuchi, K. and Akahira, M. (1988). Second order asymptotic efficiency in terms of asymptotic variances of the sequential maximum likelihood estimation procedures. In: *Statistical Theory and Data Analysis II, Proceedings of the Second Pacific Area Statistical Conference*, 191–196, North Holland, Amsterdam.

[T03a] Tanaka, H. (2003). On a relation between a family of distributions attaining the Bhattacharyya bound and that of linear combinations of the distributions from an exponential family. *Comm. Statist. Theory and Method.*, **32**(10), 1885–1896.

[T03b] Tanaka, H. (2003). On a location parameter family of distributions attaining the Bhattacharyya bound. *Proc. Sympo., Res. Inst. Math. Sci., Kyoto Univ.*, **1334**, 158–174.

[TA03] Tanaka, H. and Akahira, M. (2003). On a family of distributions attaining the Bhattacharyya bound. *Ann. Inst. Statist. Math.*, **55**(2), 309–317.

[Z71] Zacks, S. (1971). *The Theory of Statistical Inference.* Wiley, New York.

Nonparametric Methods of Estimating Integrated Multivariate Volatilities

星川 隼也 (横浜国立大学), 金谷 太郎 (広島大学)
永井 圭二 (横浜国立大学), 西山 慶彦 (京都大学)

1 序

計量ファイナンスにおける一つの問題として、拡散過程で記述されるような金融市場における高頻度ティックデータのように、各々の価格プロセスはより高頻度に観測することができるけれども異なる価格プロセスにおいては観測時点が非同期的であるような場合において、いかにして二次 (共) 変分 (ボラティリティ係数を積分した値) を推定するかという問題がある。特に一変量の場合はこの非同期性という問題は発生せず、従来より使われているノンパラメトリックな二次変分をとるという方法で高い推定精度を示すことができるが、多変量の場合は非同期性が障害となり従来の一変量の方法を自然に多変量へと拡張することができない。

本論文では、この多変量二次共変分の推定における非同期性を克服した Malliavin and Mancino(2002) 及び Hayashi and Yoshida(2005) の方法に注目し、従来の方法を無理やり拡張した方法を含めて、それらを理論、シミュレーション、実証の全側面から比較検証を行った。

2 問題設定

次のような n-次元伊藤過程を考える。

$$X_t = X_0 + \int_0^t \beta_s(\omega)ds + \int_0^t \sigma_s(\omega)dW_s$$

ここで β は $n\times 1$ 確率的ベクトル、 σ は $n\times r$ 確率的行列、$\{W_t,\{\mathscr{F}_t\}\}$ は r-次元ウィナー過程、 $\mathscr{F}_t := \sigma(W_s; s\le t)$ とする。ボラティリティ行列は $\Sigma := \sigma\sigma'$ として定義する。そのとき i 番目と j 番目の交差ボラティリティは次のようになる；

$$\Sigma^{ij}(t,\omega) := \sum_{k=1}^{r} \sigma_t^{ik}(\omega)\sigma_t^{jk}(\omega)$$

我々の推定したいものは、次のボラティリティを積分したものである；

$$\int_0^t \Sigma^{ij}(s,\omega)ds,$$

ここで我々は X_t^i を非同期的な時点 $\{t_k^i\}_{k=0}^{N_i}$ で観測する。

3 各推定量とその性質

1. 従来の方法

   $\mathrm{QV} = \sum_{k=1}^{m}(X_{t_k}^i - X_{t_{k-1}}^i)(X_{t_k}^j - X_{t_{k-1}}^j)$
   $E[\mathrm{QV}] = \int_0^t \Sigma^{ij}(s,\omega)ds$

   この推定量は不偏である。推定量の形から非同期性がする多変量の場合はそのまま使えない。

2. 線形補間を行い従来の方法を使った方法

$\mathrm{IQCOV} = \sum_{k=1}^{N_i}\sum_{l=1}^{N_j}(X^i_{t^i_k} - X^i_{t^i_{k-1}})(X^j_{t^j_l} - X^j_{t^j_{l-1}})w_{kl}$

$$w_{kl} = \begin{cases} (1 - \frac{t^j_{l-1}-t^i_{k-1}}{t^i_k - t^i_{k-1}})\frac{t^i_k - t^j_{l-1}}{t^j_l - t^j_{l-1}} & \text{if} \quad t^i_{k-1} \le t^j_{l-1} < t^i_k \le t^j_l \\ (1 - \frac{t^i_{k-1}-t^j_{l-1}}{t^j_l - t^j_{l-1}})\frac{t^j_l - t^i_{k-1}}{t^i_k - t^i_{k-1}} & \text{if} \quad t^j_{l-1} \le t^i_{k-1} < t^j_l \le t^i_k \\ \frac{t^j_l - t^j_{l-1}}{t^i_k - t^i_{k-1}} & \text{if} \quad t^i_{k-1} < t^j_{l-1} < t^j_l < t^i_k \\ \frac{t^i_k - t^i_{k-1}}{t^j_l - t^j_{l-1}} & \text{if} \quad t^j_{l-1} < t^i_{k-1} < t^i_k < t^j_l \\ 0 & \text{otherwise.} \end{cases}$$

この推定量は係数の絶対値が 1 より小さいことから原点方向へバイアスを持つ。

3. Hayashi and Yoshida(2005) の方法

$\mathrm{HY} = \sum_{k=1}^{N_i}\sum_{l=1}^{N_j}(X^i_{t^i_k} - X^i_{t^i_{k-1}})(X^j_{t^j_l} - X^j_{t^j_{l-1}})1_{\{(t^i_{k-1},t^i_k]\cap(t^j_{l-1},t^j_l]\neq\phi\}}$

$E[\mathrm{HY}] = \int_0^t \Sigma^{ij}(s,\omega)ds$

この方法は一変量の場合は従来の方法と同じ推定量となる。

4. Malliavin and Mancino(2002) の方法

$\mathrm{MM} = \sum_{k=1}^{N_i}\sum_{l=1}^{N_j}(X^i_{t^i_k} - X^i_{t^i_{k-1}})(X^j_{t^j_l} - X^j_{t^j_{l-1}})w_{kl}$

$$w_{kl} = \begin{cases} 1 & \text{If } t^i_k = t^j_l \\ \frac{\sin\frac{N(t^i_k - t^j_l)}{2}\cos\frac{(N+1)(t^i_k - t^j_l)}{2}}{N\sin\frac{(t^i_k - t^j_l)}{2}} & \text{otherwise} \end{cases}$$

Kanatani(2004) は Malliavin and Mancino の推定量が上式の形に書き換えられることを示した。係数の形から、一変量の場合は N を大きくしたとき漸近的に従来の方法と一致し、多変量の場合、観測時点の一致がほとんどなければ観測値に関わらず推定量は 0 へ収束する。

4 シミュレーション、実証、及び今後の展望

シミュレーションでは GARCH 及び EGARCH の離散化式を用いて 86400 個の代理離散過程を発生させ、観測数が約 1940,1440,720 個となるようにポアソン過程を作って観測時点を決め、その観測値を用いて各々の推定値、平均二乗誤差、バイアスを求めた。一変量の場合、QV が MM の N の値をどのようにとっても MM よりも良い結果を出した。しかし N の値を大きくすると MM の結果は QV に近づき、N=10000 ほどで両者はほぼ変わらぬ結果となった。多変量の場合、圧倒的に HY が圧倒的に結果がよく、IQCOV 及び MM の結果はひどく悪いものであった。また IQCOV 及び MM の原点方向へのバイアスも確認でき、MM の N の値を大きくすると結果はますます悪くなった。全体として理論から予想される結果がそのまま得られ、推定量として HY だけが妥当であり、その他の推定量は実用的なものとはいえない結果となった。

実証においては、日本国債先物データ (2000 年 3,6,9,12 月限月) を用いて各推定量を用いた推定値を時系列に求めた。交差ボラティリティについては隣り合う限月のものについて求めた。結果は漸近特性がまだ求められていないため信頼区間が構成できず、断定できるものではないが、視覚的には全く離れた軌跡を描き、理論、シミュレーションにおいて肯定される HY を基準に考えるならば、その他の推定量を用いることの危険性を示す一つの証明であるといえる。

今後の展望としては、各推定量の漸近特性を求めること、標本の加重積和を一般的な形で推定量としての特性を調べること、及び実務への応用が挙げられる。

最低積立金保証給付商品における
リスク尺度への漸近展開法の応用

亀山 敦史 (東京大学大学院数理科学研究科)
吉田 朋広 (東京大学大学院数理科学研究科)

概 要

変額年金保険等の商品の持つリスクの計量について考察をする．変額年金保険は，支払った保険料を株式や債券などに投資し，その資産の運用実績に応じて年金額などが変動する年金保険である．すなわち，運用の成果が上がれば受取れる年金額などが大きくなり，逆に期待したほどの運用成果が上がらなければ，受取れる年金額なども小さくなる．商品の特徴は保証のある点である．例えば，死亡保証では被保険者が積み立て期間中に死亡の場合，積立金の時価と最低保証額のどちらか大きい方の金額が支払われる．死亡しない場合，満期における積立金時価と最低保証額のどちらか大きい方が同様に支払われる．ここで，運用によって生じた最低保証額との損失の部分は保険会社が負担する．また，この年金部分に保証がつかない商品もある．今回は死亡を考えず，すなわち死亡保証は無視し満期において保険会社が保証を給付する際のリスクについて考察をする．まず，分位数によるリスク尺度を導入し，その計量を漸近展開を用いて計算する手法を述べる．さらに，条件付テイル期待値 (CTE) によるリスク尺度にも同様の方法を適用しその計算結果を最後に示す．

1 モデル化

確率変数 L^ϵ を金利 r で割り引かれた損失の現在価値とする．T を満期とし，G を T における保証額とする．危険保険料収入および死亡を無視した際の保証の現価は

$$L = \begin{cases} (G - F_T^\epsilon)e^{-rT} & (G \geq F_T^\epsilon) \\ 0 & (G < F_T^\epsilon) \end{cases}$$

である．ここで F_T^ϵ は T 時点におけるファンド (積立金) の価値であるとする．B を管理手数料を割り引く定数，株価過程を X_t^ϵ とし．$F_t^\epsilon = BX_t^\epsilon$ とする．確率過程 X_t^ϵ は確率微分方程式

$$\begin{aligned} dX_t^\epsilon &= V_0(X_t^\epsilon)dt + \epsilon V(X_t^\epsilon)dw_t \\ X_0 &= x_0, \quad t \in [0,T], \quad \epsilon \in (0,1] \end{aligned}$$

に従うとする．

2　分位数によるリスク尺度

確率変数 L^ϵ の分位数によるリスク尺度はパラメータ α $(0 \leq \alpha \leq 1)$ を用いて次で表すものとする:

$$\mathbb{V}_\alpha^\epsilon = \inf\{V; \Pr[L^\epsilon \leq V] \geq \alpha\}. \tag{1}$$

$\mathbb{V}_\alpha^\epsilon$ は満期において保証額 G がまかなえる確率が少なくとも α であるための最小額である．モデルは先に設定したものを用いる．

3　条件付テイル期待値によるリスク尺度

分位数によるリスク尺度に替わるものとして条件付テイル期待値 (CTE : Conditional Tail Expectation) によるリスク尺度がある．α を $0 \leq \alpha \leq 1$ なる定数とするおとき，CTE は分布の上側 $(1-\alpha)$ 点を超過する損害額の期待値として定められる．分位数に対するリスク尺度 V_α^ϵ を (1) 式で定義されるものとする．連続な損害分布に対して，パラメータ α の場合の CTE は，

$$CTE_\alpha(L^\epsilon, \mathbb{V}_\alpha^\epsilon) = E[L^\epsilon | L^\epsilon > \mathbb{V}_\alpha^\epsilon]$$

となる．モデルは先に設定したものと同様のものとする．

本講演ではこれら二つのリスク尺度に対して Yoshida[2] の定理を用いて漸近展開法を適用し，その結果が真値と近似可能であることを示す．そして最後に数値実験の結果を提示し，漸近展開法がこれらのリスク尺度に対して有効であることを示す．

参考文献

[1] Hardy, M. (2003) *"Investment Gurantees:Modeling and Risk Management for Equity-Linked Life　Insurance"* John-Wiley. 日本アクチュアリー会変額年金保険等の最低保証リスクに係わる研究ワーキンググループ訳 (2005) "投資型商品における最低保証給付の数理" 丸善プラネット.

[2] Yoshida, N. (1992) "Asymptotic Expansion for Statistics Related To Small Diffusions"　*Journal of Japan Statistical Society*, **22**, 139-159.

Sampling-based estimation for Ornstein-Uhlenbeck processes with stable Lévy driver

九州大学大学院数理学研究院[†] 増田 弘毅[‡]

本講演では “1 次元非正規安定型オルンシュタイン-ウーレンベック過程 (OU 過程) から漸近的に高頻度な離散観測を得る” という統計モデルにおける尤度関数の漸近挙動を考える. 即ち, 背後に一種の連続時間モデルを想定し, それから所与の時点においてのみデータが観測されるという状況を扱う. 特に観測情報量のタイトネスを明確化する (推定量の収束率の明確化).

ここで扱う漸近手法は拡散過程の推測においては典型的なものである. 例えば, 時間区間 $[0,1]$ 上の $(i/n)_{i=0}^n$ でデータを得る場合というような, “漸近的に有界時間区間上の連続観測を得る統計モデル” の実現を可能とする.

一般にレヴィ駆動型 OU 過程は簡単な一次の自己回帰構造を持つわけだが, 安定分布は有限分散を持たないという事実により, 対象とする尤度関数の漸近挙動は正規型 OU 過程の場合のそれとはかなり異なるものとなる. 特に, 推定量の最適収束率は特殊な様相を呈することが分かった: ここでは観測情報量のタイトネスのために “point process convergence” の理論を本質的な道具として用いた; cf. Davis and Resnick (1985, AP).

1. 扱う統計モデル: 漸近手法

まず, $Z=(Z_t)_{t\in\mathbb{R}_+}$ で指数 $\alpha\in(0,2)$ の 1 次元非正規型安定過程で無限分解可能分布の特性量 $(0,0,\nu)$ に対応するものを表す. ここでレヴィ測度 $\nu(dz)=g(z)dz$ は, $(\delta^+,\delta^-)\neq(0,0)$ なる非負定数 δ^+, δ^- に対して

$$g(z)=\delta^+ z^{-1-\alpha}\mathbf{1}_{\{z>0\}}(z)+\delta^-|z|^{-1-\alpha}\mathbf{1}_{\{z<0\}}(z)$$

を通じて与えられる. このことを $\mathcal{L}(Z_1)=\mathcal{S}_\alpha(\delta^+,\delta^-)$ と表し, また $\phi_\alpha(\cdot;\delta^+,\delta^-)$ で $\mathcal{S}_\alpha(\delta^+,\delta^-)$ のルベーグ測度に関する密度関数を表す; 1 次元安定過程の解析に関しては Zolotarev (1986, monograph) 参照. $\int_{|z|\le 1}|z|\nu(dz)=\infty$ であることから, Z は任意の有界時間区間上, 確率 1 で無限回の (小さい) 飛躍を持ち, またウィーナー過程のように非有界変動な確率過程である.

定数 $\lambda,\gamma\in\mathbb{R}$ に対し, 安定型 OU 過程 $X=(X_t)_{t\in\mathbb{R}_+}$ は, 任意の $t\ge s\ge 0$ に対して

$$X_t=e^{-\lambda(t-s)}X_s+\gamma(1-e^{-\lambda(t-s)})+\int_s^t e^{-\lambda(t-u)}dZ_u$$

で定義される: X の $\mathbb{D}(\mathbb{R}_+;\mathbb{R})$ 上の像測度は凸集合 $\Theta\subset\mathbb{R}^5$ に含まれる未知母数

$$\theta=(\lambda,\alpha,\gamma,\delta^+,\delta^-)^\top$$

に依存する. 本報告では, X は所与の $n+1$ 時点 $(t_i^n)_{i=0}^n$, $t_i^n=ih_n$, においてのみ観測されると仮定する. ここで (h_n) は $h_n\to 0$ なる有界正数列である. 以上のもとで $(X_{t_i^n})_{i=0}^n$ の $\mathbb{R}^{n+1}$ 上の分布を P_θ^n で表し, 統計モデル $(P_\theta^n)_{\theta\in\Theta}$ を考える. X については通常の自己相関関数を定義できないが, 指数的 β-ミキシングであることは分かっているので弱従属過程といえる; cf. Masuda (2004, Bernoulli; 2005, preprint).

2. 対数尤度関数の表示

以下 $\alpha\neq 1$ を仮定する: (非対称) コーシーの場合も同様に考えることができるが, この場合は尤度の表示において他の場合と異なる特殊な状況が生じる. レヴィ積分の性質, 狭義安定分布のスケーリング性, および X のマルコフ性により, $(X_{t_i^n})_{i=0}^n$ の対数尤度 $\ell_n(\theta)$ は以下のように陽に表示できる:

$$\ell_n(\theta)=\sum_{i=1}^n\log\left\{c_{h_n}^{-1/\alpha}\phi_\alpha\left(Y_i^n;\delta^+,\delta^-\right)\right\}.$$

[†] 〒 812-8581 福岡市東区箱崎 6-10-1
[‡]Email: hiroki@math.kyushu-u.ac.jp. Personal URL: http://www.math.kyushu-u.ac.jp/~hiroki/

ここで $c_h=(\alpha\lambda)^{-1}\{1-\exp(-\alpha\lambda h)\}$, $Y_i^n=c_{h_n}^{-1/\alpha}\left\{X_{ih_n}-e^{-\lambda h_n}X_{(i-1)h_n}+\gamma(1-e^{-\lambda h_n})\right\}\overset{\mathcal{L}}{=}\mathcal{S}_\alpha(\delta^+,\delta^-)$ である. この表示は "安定型レヴィ過程固有の性質" を用いて導出されたものであり, Z が一般のレヴィ過程の場合には同様の表示式は得られないということに注意しておく (レヴィ駆動型 OU 過程の遷移確率は一般に無限分解可能であるが, その特性量までしか明確化できない).

3. $\ell_n(\theta)$の漸近挙動

ここで扱う統計モデルは未知母数に滑らかに依存しているため, スコア等計量 $U_n(\theta)=\partial_\theta l_n(\theta)$ および観測情報量 $J_n(\theta)=-\partial_\theta^2 l_n(\theta)$ が直ちに定義される; $U_n(\theta)$ は列ベクトルとみなす. 各 $n\in\mathbb{N}$ について $E_\theta^n[\{\partial_\lambda\ell_n(\theta)\}^2]=\infty(=-E_\theta^n[\partial_\lambda^2\ell_n(\theta)])$ となってしまうため, $\partial_\lambda\ell_n(\theta)$ の漸近挙動の導出が問題となる: $-E_\theta^n[\partial_\lambda^2\ell_n(\theta)]$ のタイトネス, 即ち λ の推定量の最適レートは導出できたが, 対応する極限分布の明確化にはまだ到達できていない.

$\theta^*=(\alpha,\gamma,\delta^+,\delta^-)$ と書き,

$$
\begin{aligned}
A_n^*(\alpha) &= \operatorname{diag}\left\{\sqrt{n}\log(1/h_n),\ \sqrt{n}h_n^{1-1/\alpha},\ \sqrt{n},\ \sqrt{n}\right\},\\
A_n(\alpha) &= \operatorname{diag}\left\{n^{1/\alpha}h_n^{1-1/\alpha},\ \sqrt{n}\log(1/h_n),\ \sqrt{n}h_n^{1-1/\alpha},\ \sqrt{n},\ \sqrt{n}\right\},
\end{aligned}
$$

を導入する. 以下の Theorems 1, 2 が本報告における主張である:

Theorem 1. λ を既知とするとき, θ^* について局所漸近正規性がレート $A^*(\alpha)$ で成立する; 漸近最適分散は, $\phi_\alpha(\cdot;\delta^+,\delta^-)$ を用いて明示的に与えられる.

Theorem 2. $J_n(\theta)$ について,

$$\{A_n(\theta)\}^{-1}\partial_\theta^2\ell_n(\theta)\{A_n(\theta)\}^{-1\top}=O_{P_\theta^n}(1)$$

が成立する.

4. 主張に関する注意点

1) 非正規型の場合には, 最適収束率はより複雑になることが分かった. 特に γ について, その最適収束率が α に依存することは興味深い. "安定型レヴィ過程 (+複号ポアソン過程 etc.) からの離散観測" という設定において Aït-Sahalia and Jacod (2004, preprint) はフィッシャー情報量行列の漸近挙動を様々な場合において調べているが, そこでは局所漸近正規性については議論されていない. また, 対称な安定レヴィ過程からの離散観測に基づいた $\delta^+(=\delta^->0)$ の最適収束率 $\sqrt{n}$ は Woerner (2001,2003) によって既に得られている. 我々の設定においても "L^2-可微分性, リンデベルグ条件, および大数の法則" が実際に確保できるのであるが, $\ell_n(\theta)$ の表示式の被加数が観測数 n に依存していることに注意したい. この事により, 例えば通常の L^2-可微分性の導出には注意を要する: ここでは iid の場合と異なり, 結局 2 次導関数まで見て示した.

2) Theorem 2 から以下が分かる: (i) $\alpha\in(0,1)$ のとき, $\lambda,\gamma,\alpha,\delta^\pm$ の順で速く収束する推定量の構成が望まれる; (ii) $\alpha\in(1,2)$ のときは, $\lambda,\alpha,\delta^\pm,\gamma$ の順で速く収束する推定量が構成が望まれる. 具体的な最適収束率は $A_n^*(\alpha)$ および $A_n(\theta)$ の対応する対角成分の通りである. ここで注意すべきは, $A_n(\alpha)$ の対角成分は, 任意の $\alpha\in(0,2)$ の対して, $h_n\to 0$ である限り全て $\to\infty$ となるという点である. 特に γ に関しては, $\mathcal{L}(Z_1)$ の裾が重いほど, 言い換えれば α が小さいほどその推定量は速く収束すべき, ということになる.

3) 更に, 上記の定理は "最終観測時点 nh_n の漸近挙動" については何も要求していないことに注意したい. これは正規型の場合と大きく異なる点である; 形式的に $\alpha=2$ とすれば, (λ,γ) の最適収束率は拡散過程の場合の $\sqrt{nh_n}$ に一致する.

4) ここでは安定分布固有の性質を駆使した為, Theorems 1,2 は残念ながら, "Z が他のレヴィ過程の場合どうなるか?" については訴える要素に欠ける. 同時に, 同じ理由で, 安定レヴィ駆動型以外の広範な X について局所漸近正規性の統一的な導出方法はない, とも断言できない. この点に関して鍵となるのが X の表示式において現れるレヴィ積分の分布である.

5) もし Theorem 2 において極限分布 (θ に依存する) が陽に求まり, またその分布収束が θ について一様であれば, Sweeting (1980,AS) の結果によって局所漸近混合正規性まで得られる.

Estimating a common slope of multiple strata in a compound Poisson distribution using a conjugate prior

統計数理研究所　大西 俊郎
統計数理研究所　柳本 武美

1. 問題設定

標本分布としてべき乗の分散関数をもつ指数型分布族を考える．

$$p(x;\,\mu) = \exp\{c(\mu)x - M(\mu)\}\,a(x) \quad (x \geq 0,\, \mu \in \mathbb{R}^+),$$
$$v(\mu) = \frac{1}{c'(\mu)} = \frac{\mu^{\xi}}{\tau_0} \quad (\tau_0 > 0,\, 1 \leq \xi \leq 2).$$

ここで，μ は平均パラメータ，$c(\mu)$ および $M(\mu)$ はそれぞれ正準パラメータおよびキュムラント関数を μ の関数とみたものである．べき指数 ξ およびパラメータ τ_0 は既知とする．この分布は，Tweedie モデルと呼ばれることもあり，$1 < \xi < 2$ のとき複合 Poisson 分布に一致することが知られている．アクチュアリアル・サイエンスの分野では支払保険金のモデルとして使われることがある．また，Poisson 分布（$\xi = 1,\, \tau_0 = 1$）とガンマ分布（$\xi = 2$）が含まれていることに注意されたい．この分布をベースとした K 個の層をもつ次のような対数リンク回帰モデルを考察する．

$$p(\boldsymbol{x};\,\boldsymbol{\alpha},\,\beta) = \prod_{k=1}^{K} p(\boldsymbol{x}_k;\,\alpha_k,\,\beta) = \prod_{k=1}^{K}\prod_{i=1}^{n_k} p(x_{ki};\,\exp(\alpha_k + \beta z_{ki})).$$

ここで，$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ および β はそれぞれ切片パラメータおよび共通の傾きパラメータであり，z_{ki} は共変量である．

本発表の目的は，$\boldsymbol{\alpha}$ に関する共役解析を通して共通の傾きパラメータ β に対する最適な推定関数を導出することである．

2. 共役事前分布

切片パラメータ α_k の上に次の事前分布を仮定する $(1 \leq k \leq K)$．

$$\pi(\alpha_k - \alpha_0;\,\delta n_k) = \frac{1}{K(\delta n_k)}\exp\left[-\delta n_k\left\{u(e^{\alpha_k-\alpha_0};\,2-\xi) - u(e^{\alpha_k-\alpha_0};\,1-\xi)\right\}\right],$$
$$u(x;\,\kappa) = \begin{cases} \log x & \text{for } \kappa = 0, \\ (x^{\kappa} - 1)/\kappa & \text{otherwise.} \end{cases}$$

ここで，α_0 および $\delta > 0$ は超パラメータ，$K(\delta n_k)$ は正規化定数である．

命題 2.1. 標本密度 $p(\boldsymbol{x}_k;\,\alpha_k,\,\beta)$ の下で事前密度 $\pi(\alpha_k - \alpha_0;\,\delta n_k)$ に対応する事後分布は $\pi(\alpha_k - \hat{\alpha}_k;\,\delta_k^* n_k)$ である．ここで，

$$\hat{\alpha}_k = \log\frac{\tau_0\sum_{i=1}^{n_k} x_{ki}e^{(1-\xi)\beta z_{ki}} + \delta n_k e^{-(1-\xi)\alpha_0}}{\tau_0\sum_{i=1}^{n_k} e^{(2-\xi)\beta z_{ki}} + \delta n_k e^{-(2-\xi)\alpha_0}},$$
$$\delta_k^* = \frac{1}{n_k}\left\{\tau_0\sum_{i=1}^{n_k} x_{ki}e^{(1-\xi)\beta z_{ki}} + \delta n_k e^{-(1-\xi)\alpha_0}\right\}^{2-\xi}\left\{\tau_0\sum_{i=1}^{n_k} e^{(2-\xi)\beta z_{ki}} + \delta n_k e^{-(2-\xi)\alpha_0}\right\}^{\xi-1}$$

したがって，事前密度 $\pi(\alpha_k - \alpha_0;\, \delta n_k)$ は共役である．

3. ピタゴラス関係

次の損失関数を採用し，切片パラメータ α_k に関する共役解析を議論する．

$$L(\check{\alpha}_k - \alpha_k) = u\big(\exp(\alpha_k - \check{\alpha}_k);\, 2-\xi\big) - u\big(\exp(\alpha_k - \check{\alpha}_k);\, 1-\xi\big).$$

以下，記法として $\mathrm{E}_p[\cdot]$，$\mathrm{E}_\pi[\cdot]$ および $\mathrm{E}_{\mathrm{post}}[\cdot]$ はそれぞれ標本分布，事前分布および事後分布での期待値を意味するものとする．

命題 3.1. 任意の推定量 $\check{\alpha}_k$ に対して変形ピタゴラス関係

$$\mathrm{E}_{\mathrm{post}}\Big[\, L(\check{\alpha}_k - \alpha_k) - L(\hat{\alpha}_k - \alpha_k) - \rho(\delta_k^* n_k) L(\check{\alpha}_k - \hat{\alpha}_k)\,\Big] = 0$$

が成り立つ．したがって，推定量 $\hat{\alpha}_k$ は損失関数 $L(\check{\alpha}_k - \alpha_k)$ の下で最適である．

4. 傾きパラメータに対する最適な推定関数

傾きパラメータ β に対するスコア関数を計算する．

$$\begin{aligned} S(\boldsymbol{x};\, \boldsymbol{\alpha},\, \beta) &= \sum_{k=1}^{K} S_k(\boldsymbol{x}_k;\, \alpha_k,\, \beta) \\ &= \tau_0 \sum_{k=1}^{K} \left\{ e^{(1-\xi)\alpha_k} \sum_{i=1}^{n_k} x_{ki} z_{ki} e^{(1-\xi)\beta z_{ki}} - e^{(2-\xi)\alpha_k} \sum_{i=1}^{n_k} z_{ki} e^{(2-\xi)\beta z_{ki}} \right\}. \end{aligned}$$

次の定理に基づき，このスコア関数の事後平均で β を推定することを提案する．

定理 4.1. スコア関数の事後平均 $\mathrm{E}_{\mathrm{post}}\left[S(\boldsymbol{x};\, \boldsymbol{\alpha},\, \beta) \right]$ は，不偏性条件 $\mathrm{E}_\pi\big[\, \mathrm{E}_p[\, g(\boldsymbol{x};\, \beta)\,]\big] = 0$ を満たす推定関数の中で，評価基準 $\mathrm{E}_\pi\big[\, \mathrm{E}_p[\, g^2(\boldsymbol{x};\, \beta)\,]\big] \,/\, \{\mathrm{E}_\pi\left[\, \mathrm{E}_p\left[\, \partial g(\boldsymbol{x};\, \beta)/\partial\beta\, \right]\right]\}^2$ の下で最適である．

次の命題で注目すべきは，最適な推定関数が前節の共役解析で得られた Bayes 推定量を用いて簡単に表現できることである．

命題 4.2. 最適な推定関数は

$$\mathrm{E}_{\mathrm{post}}\big[S(\boldsymbol{x};\, \boldsymbol{\alpha},\, \beta) \big] = \sum_{k=1}^{K} \rho(\delta_k^* n_k) S_k(\boldsymbol{x}_k;\, \hat{\alpha}_k,\, \beta)$$

である．特に，次が成り立つ．

$$\mathrm{E}_{\mathrm{post}}\big[S_k(\boldsymbol{x}_k;\, \alpha_k,\, \beta) \big] = \rho(\delta_k^* n_k) S_k(\boldsymbol{x}_k;\, \hat{\alpha}_k,\, \beta) \quad (1 \le k \le K).$$

情報量最小化を用いた過完備系の学習

早稲田大学 理工学部 村田 昇

文字認識などの高次元空間での判別や，ロボットなどの非線形性の強い制御など，工学的な問題では Wavelet 展開や多層神経回路網などの過剰な自由度を持つ基底関数の重ね合わせによる関数近似の方法が広く用いられている.

例えば三層パーセプトロンと呼ばれる学習機械は

$$f(\boldsymbol{x};\boldsymbol{\theta}) = \sum_{i=1}^{n} c_i \phi_c(\boldsymbol{a}_i \cdot \boldsymbol{x} - b_i)$$

で表される階層型の構造を持っている．ただし，$\boldsymbol{x}$ は d 次元の入力で，$\theta = (\boldsymbol{a}_i, b_i, c_i; i = 1, \ldots, n)$ は機械の入出力関係を決定するパラメタである．ここで中間表現である $z = \phi_c(\boldsymbol{a} \cdot \boldsymbol{x} - b)$ を基底関数と考えると，三層パーセプトロンによって実現される関数は，次の積分表現の有限和による近似としてとらえることができる.

$$f(\boldsymbol{x}) = \lim_{\varepsilon \to 0} \int_{R^{d+1}} \phi_c(\boldsymbol{a} \cdot \boldsymbol{x} - b) T(\boldsymbol{a}, b) e^{-\varepsilon|\boldsymbol{a}|^2} d\boldsymbol{a} db$$

ただし，$T(\boldsymbol{a}, b)$ は関数 ϕ_d による f の積分変換で

$$T(\boldsymbol{a}, b) = \frac{1}{(2\pi)^d C_{\phi_d,\phi_c}} \int_{R^d} \phi_d(\boldsymbol{a} \cdot \boldsymbol{x} - b) f(\boldsymbol{x}) d\boldsymbol{x}.$$

で定義され，二つの関数 $\phi_d, \phi_c \in L^1(R) \cap L^2(R)$ は，それぞれの Fourier 変換 $\hat{\phi}_d, \hat{\phi}_c$ を用いて定義される積分

$$C_{\phi_d,\phi_c} = \int_{R^d} |\omega|^{-d} \hat{\phi}_d(\omega) \hat{\phi}_c(\omega) d\omega,$$

が有界となるように選ばれているものとする．積分表現の収束は関数 f の性質に依存するが，$f \in L^1(R^d) \cap L^p(R^d)$ $(1 \le p < \infty)$ の場合は，L^p ノルムの意味で，$f \in L^1(R^d)$ かつ有界で一様連続の場合は，L^∞ ノルムの意味で収束する.

この積分表現においては $f(\boldsymbol{x})$ は R^d 上の関数，$T(\boldsymbol{a}, b)$ は R^{d+1} 上の関数であり，Fourier 変換とその逆変換のような一対一の対応関係とはなっていない．すなわち，ある関数 ϕ_c に対しては，C_{ϕ_d,ϕ_c} が有界となる条件を満しさえすれば ϕ_d は自由にとってよいので，$f(\boldsymbol{x})$ を再構成する $T(\boldsymbol{a}, b)$ には無数の候補が存在し，一意に決めることができない．同様な積分表現との対応付けは，Wavelet 展開や動径基底関数ネットワークなどの多層神経回路網でも成り立つが，こうした過剰な自由度をもつ基底関数系は過完備 (over-complete) と呼ばれ，工学では有限個の基底を用いた関数近似のために多く利用されている.

ところで，過完備な基底による積分表現の非一意性は，学習においては長所にも短所にもなりうる．学習問題の多くは，入力 $\boldsymbol{x}$ と出力 y の例題の集合 $\mathcal{D} = \{(\boldsymbol{x}_t, y_t); t = 1, \ldots, T\}$ から $\boldsymbol{x}$ と y の関数関係を推定する問題であるが，出力の予測 $f(\boldsymbol{x};\boldsymbol{\theta})$ と実際の出力 y の間の適当な損失 $L[f(\boldsymbol{x};\boldsymbol{\theta}), y]$ を用いて，例題にもとづくパラメタの最適化問題

$$\text{minimize:} \quad \sum_{t=1}^{T} L[f(\boldsymbol{x}_t;\boldsymbol{\theta}), y_t]$$

として定式化される．このとき過完備系を用いる長所の一つは，関数近似や数値積分などで問題になる「次元の呪い (curse of dimensionality)」を避けることが出来ることである．次元の呪いとは，例えば入力次元を d とした場合，近似の精度を k 倍に上げるためにはパラメタ数を k^d 倍程度に増やす必要があるという，精度が入力次元の大きさに指数的に依存する性質を指している．一方，過

完備系では豊富な関数の中から巧く基底を選択することによって，関数の入力値の次元によらずパラメタ数を線形的に増やす程度で近似精度を上げていくことができる．基底の選択は単純な貪欲算法 (greedy algorithm) を用いることで可能であるが，直感的には多次元数値積分において MCMC 法が台形則に比べて有利なように，近似対象となる関数の形状に適応させて，基底関数を規則的ではなく不規則に配置することによって良い近似を求めていると考えることができる．

一方，関数の表現に一意性がないことの短所は，最適化の難しさに現れる．非線形最適化問題において典型的な局所解が非常に多数存在することや外れ値に対する過学習など，学習結果の不安定性に密接に関係している．このような不安定性を回避する方法としては，最適化の目的関数に正則化項を付加することが一般的である．すなわち，適当な事前知識を仮定してパラメタ $\boldsymbol{\theta}$ の定義域の上に罰金 $P(\boldsymbol{\theta})$ を，あるいは推定される関数の滑らかさなどを仮定して Sobolev ノルムなどを利用した汎関数 $S[f(\boldsymbol{\theta})]$ を設定することによって，損失との同時最適化

$$\text{minimize:} \quad \sum_{t=1}^{T} L[f(\boldsymbol{x}_t;\boldsymbol{\theta}), y_t] + \lambda P(\boldsymbol{\theta})$$

$$\text{minimize:} \quad \sum_{t=1}^{T} L[f(\boldsymbol{x}_t;\boldsymbol{\theta}), y_t] + \lambda S[f(\boldsymbol{\theta})]$$

を行うことになる．ただし λ は，近似精度を決める損失 L と正則化因子である P や S のバランスを制御するハイパーパラメタであり，特別な知識がない限りは，情報量規準，経験ベイズ法，ブートストラップ法などを援用して決定する必要がある．

最適化の目的関数の決定における不定性を避ける方法の一つとして，本稿では，階層構造を持つ学習機械の確率化と情報量の最小化を提案する．学習機械の確率化は，中間表現としての出力値 z(あるいはその正規化したもの) を確率値とみなし，次の層への入力を $\{0,1\}$ の二値から確率的に選ぶことによって実現される．この結果，機械の出力 $f(\boldsymbol{x};\boldsymbol{\theta})$ は確率変数となるので，出力誤差 $R = y - f(\boldsymbol{x};\boldsymbol{\theta})$ も確率密度 $p(r)$ をもつ確率変数であると考えることができる．このとき出力誤差の自然な損失として，その情報量

$$H(R) = E\left[\log \frac{1}{p(R)}\right] = \int p(r) \log \frac{1}{p(r)} dr$$

を考えることができる．情報量は確率変数の実現値を伝送・保存する際に必要とされる平均的なコストを表している．同様に確率化された中間表現 Z の損失も情報量によって評価することができる．したがって機械全体の損失は，情報理論的に同一の意味をもつ情報量の和として定義でき，最適化は

$$\text{minimize: (出力誤差の情報量)} + \text{(中間表現の情報量)}$$

に従っておこなえばよいことがわかる．この考え方は，スパースコーディング (sparse coding) と呼ばれる確率過程の分解手法と関係し，画像や音声信号の分解に応用することによって，時系列の分解，圧縮，分類などのために有効な特徴抽出に利用できることを，実例をふまえて紹介する．

参考文献

[1] A. R. Barron, *Universal Approximation Bounds for Superpositions of a Sigmoidal Function*, IEEE Trans. Information Theory **39** (May 1993), 930–945.

[2] E. J. Candès and D. L. Donoho, *Recovering Edges in Ill-Posed Inverse Problems Optimality of Curvelet Frames*, Ann. Statist. **30** (2000), 784–842.

[3] ———, *Curvelets and Curvilinear Integrals*, J. Approx. Theory **113** (2000), 59–90.

[4] F. Girosi, M. Jones, and T. Poggio, *Regularization Theory and Neural Networks Architectures*, Neural Computation **7** (1995), 219–269.

[5] M. S. Lewicki and T. J. Sejnowski, *Learning overcomplete representations*, Neural Computation **12** (2000), 337–365.

[6] N. Murata, *An integral representation with ridge functions and approximation bounds of three-layered network*, Neural Networks **9** (1996), 947–956.

A New Computational Scheme for Computing Greeks by the Asymptotic Expansion Approach

内田善彦 (阪大経済学研究科) ＜高橋明彦 (東大経済学研究科)、松岡竜佑 (東京海上日動火災保険株式会社) との共同研究＞

1 漸近展開法のファイナンスへの応用の概要

原資産株価過程 S_t とする。定数パラメータ $\varepsilon(0<\varepsilon\leq 1)$ を導入し $S_t^{(\varepsilon)}$ 以下を満たすとする。

$$\begin{cases} dS_t^{(\varepsilon)} &= \mu S_t^{(\varepsilon)}dt+\varepsilon\sigma\left(S_t^{(\varepsilon)}\right)dw_t \\ S_0^{(\varepsilon)} &= s_0\,(>0) \end{cases}. \tag{1}$$

ただし、w_t は 1 次元ブラウン運動、$\mu=r-q$、$r\ q$ はは無リスク金利、配当率とし定数である。また、σ は有界な微分をもつ滑らかな関数とする。$S_t^{(\varepsilon)}$ は時刻 t、D_∞ 上、$\varepsilon\downarrow 0$ で漸近展開

$$S_t^{(\varepsilon)}=A_{0t}+\varepsilon A_{1t}+\frac{\varepsilon^2}{2}A_{2t}+\cdots \tag{2}$$

を持つ。$y=(A_{0T}-K)/\varepsilon$,$g_1=A_{1T}$, $g_2=A_{2T}/2$ とし、数学的条件を Kunitomo and Takahashi[2003] の *Assumption 6.2* と同様に仮定すると、 満期 T、満期時点のペイオフ $(S_T-K)^+$ を持つプレーン・バニラ・タイプのヨーロピアン・コール・オプション価格の漸近展開は以下。

$$\begin{aligned} e^{-rT}E\left[\left(S_T^{(\varepsilon)}-K\right)_+\right] &= \varepsilon\, e^{-rT}E\left[(y+g_1)\,1_{\{g_1\geq -y\}}\right]+\varepsilon^2 e^{-rT}E\left[g_2\,1_{\{g_1\geq -y\}}\right] \\ &\quad +\varepsilon^2 e^{-rT}E\left[(y+g_1)\,g_2\,\delta_{-y}(g_1)\right]+O\left(\varepsilon^3\right), \end{aligned}$$

ただし $E[\cdot]$ はリスク中立測度を用いた期待値、$\delta_x(\cdot)$ はディラックのデルタ関数。さらに Takahashi[1999] に従い次を得る。ただし、Σ、c は定数、また $f=-c\Sigma$。

Theorem 1. *The asymptotic expansion of the price of the plain vanilla European call option at time zero with maturity* T, $C_E^{(\varepsilon)}(0,T)$, *is represented by*

$$C_E^{(\varepsilon)}(0,T) \;=\; \varepsilon\, e^{-rT}\left(y\,N\left(\frac{y}{\sqrt{\Sigma}}\right)+\Sigma\, n\,[y;0,\Sigma]\right)+\varepsilon^2 e^{-rT}f\,y\,n\,[y;0,\Sigma]+O(\varepsilon^3), \tag{3}$$

where $n\,[x;0,\Sigma]$ *is the density function of* $N\,(0,\Sigma)$ *and* $N\,(x)$ *is cumulative distribution function of* $N\,(0,1)$.

なお、平均型ヨーロピアン・コール・オプションについても同様の議論ができる。

2 漸近展開を用いた分散減少法

以下の議論は主に Takahashi and Yoshida[2005] に基づいた漸近展開を用いた分散減少法の概要である。$X_u(s,y)\,(s\leq u\leq T)$ が以下の確率積分方程式に従うとする。

$$X_u^{(\varepsilon)}(t,y)=y+\int_t^u V_0\left(X_{s-}^{(\varepsilon)}(t,y),\varepsilon\right)ds+\int_t^u V\left(X_{s-}^{(\varepsilon)}(t,y),\varepsilon\right)dw_s. \tag{4}$$

$u\,(0,y)\equiv E\left[f\left(X_T^{(\varepsilon)}(0,y)\right)\right]$ を計算することを考えるとき、N 個のサンプルを用いた普通の (crude) モンテカルロ法における推定値は以下。

$$G\,(n,N)=G\,(n,N;\omega_1,\cdots,\omega_N)\equiv\frac{1}{N}\sum_{j=1}^{N}f\left(\bar{X}_T^{(\varepsilon)}(\omega_j)\right). \tag{5}$$

$u(0,y)$ の新しい推計値として以下を導入する。なお、$E\left[\hat{f}\left(X_T^{(0)}(0,y)\right)\right]$ は漸近展開を用いれば解析的に計算できる（ハイブリッド法)。

$$\begin{aligned} G^*(\varepsilon,n,N) &= G^*(\varepsilon,n,N;\omega_1,\cdots,\omega_N) \\ &\equiv E\left[\hat{f}\left(X_T^{(0)}(0,y)\right)\right]+\frac{1}{N}\sum_{j=1}^{N}\left(f\left(\bar{X}_T^{(\varepsilon)}(\omega_j)\right)-\hat{f}\left(\bar{X}_T^{(0)}(\omega_j)\right)\right). \end{aligned} \tag{6}$$

3 デルタ

例えば、$\phi(S_T) = (S_T - K)^+$ といった、満期 T 時点のペイオフ ϕ を持つ派生商品を考える。この商品の価格は

$$u(s_0) = E\left[e^{-rT}\phi(S_T)\right] = e^{-rT}E\left[\phi(S_T)\right].$$

と書ける。さらに、以下が成り立つ

$$u'(s_0) = e^{-rT}E\left[\phi'(S_T)\,Y_T\right], \tag{7}$$

ただし $Y_t \equiv \frac{\partial S_t}{\partial s_0}$ (stochastic flow) は以下の確率微分方程式を満たす。

$$\begin{cases} dY_t = r(t)\,Y_t\,dt + \sigma^{*\prime}(S_t)\,Y_t\,dw_t \\ Y_0 = 1 \end{cases} \tag{8}$$

式 (7) の導出は 例えば Fournié et al.[1999] または Imamura et al.[2004] に詳しい。$Y_t^{(\varepsilon)}$ を $\frac{\partial S_t^{(\varepsilon)}}{\partial s_0}$ で定義する。stochastic flow の漸近展開により派生商品価格のデルタが計算できる。

Proposition 1. *The asymptotic expansion of the stochastic flow is represented by:*

$$Y_t^{(\varepsilon)} = \frac{\partial A_{0t}}{\partial s_0} + \varepsilon\frac{\partial A_{1t}}{\partial s_0} + \varepsilon^2\frac{\partial A_{2t}}{\partial s_0} + \cdots, \tag{9}$$

in D_∞ *as* $\varepsilon \downarrow 0$.

Theorem 2. *Delta of European Call Option* $D(0,T)$ *is represented by*

$$D(0,T) = e^{-rT}\left\{\varepsilon^2(A + By + Cy^2 + Dy^3)n[y;0,\Sigma] + \varepsilon(d\cdot N(\frac{y}{\sqrt{\Sigma}}) + E\cdot n[y,0,\Sigma])\right\} + O\left(\varepsilon^3\right), \tag{10}$$

where A,B,C,D and E are constants.

デルタの計算にはハイブリッド法が適用できる。この際、以下を用いればよい。

$$\hat{f}(x) = \varepsilon\{(d + d_1x) + \varepsilon(d_2 + d_3x^2)\}1_{\{x\geq -y\}} + \varepsilon^2 d_4, \tag{11}$$

ただし d、d_1、d_2、 d_3、 d_4 は定数。また、この議論は平均型ヨーロピアン・コール・オプションのデルタ計算にも適用できる。

4 ベガおよび数値計算

同様の議論はベガの計算にも適用できる。原資産過程が BS プロセスの場合と CEV プロセスの場合について数値計算を行い、提案したアルゴリズムが効果的に機能することを確認した。

参考文献

[1] Fournié, Eric, Jean-Michel Lasry, Jérôme Lebuchoux, Pierre-Louis Lions and Nizar Touzi, "Application of Malliavin calculus to Monte Carlo Method in Finance," *Finance and Stochastics* 3, 1999, pp.391-412.

[2] Imamura, S, Takahashi, A., and Uchida, Y. (2004), "On Risk Sensitivity Analysis of Options Based on Malliavin Calculus," IMES Discussion Paper Series 2004-J-24, Bank of Japan (in Japanese).

[3] Kunitomo, N. and Takahashi, A. (2003), "Foundation of Mathematical Finance-Application of Malliavin Calculus and Asymptotic Expansion Method-," Toyo-Keizai (in Japanese).

[4] Takahashi, A. (1999), "An Asymptotic Expansion Approach to Pricing Contingent Claims," *Asia-Pacific Financial Markets*, Vol.6, 115-151.

[5] Takahashi, A. and Yoshida, N.(2005), "Monte Carlo Simulation with Asymptotic Method," Discussion Paper Series, CIRJE-F-335, Faculty of Economics, The University of Tokyo, Forthcoming in *The Journal of Japan Statistical Society*.

A control variate method for the Monte Carlo simulation of densities of diffusions

Ahmed Kebaier and Arturo Kohatsu-Higa
Universite de Le Mans and Osaka University

In this article we consider a hypoelliptic diffusion with smooth coefficients. Under this condition one proves using Malliavin Calculus that the density of the diffusion exists and is smooth. Gaussian type upper bounds for the densities and its derivatives can also be obtained.

Since the articles of Bally-Talay

1.V. Bally, D. Talay. "The law of the Euler scheme for stochastic differential equations (II): convergence rate of the density", Monte Carlo Methods and Applications, 2, 1996, p. 93-128

2.V. Bally and D. Talay, The law of the Euler scheme for stochastic differential equations I: convergence rate of the distribution function, Probab. Theory Related Fields 104 (1996) 43-60.

one knows that the approximation of the density of a diffusion is possible using the Euler scheme.

Nevertheless in the case of diffusions the approximation does not necessarily have a smooth density as the diffusion coefficient can be zero conditioned on the past information.

Therefore one needs to perturb slightly the Euler-Maruyama scheme in order to obtain a smooth random variable. Furthermore this is needed to prove the stability of the Malliavin covariance matrix.

In 1. the approximation result obtained by the authors had various resrictions that were proved not to be necessary in

3. A. Kohatsu-Higa. High order Ito-Taylor approximations to heat kernels. Journal of Mathematics of Kyoto University, vol 37, 1, 129-151, 1997.

With this result it was proven that the simulation approach was possible. Nevertheless as it is well known from kernel density estimation methods the approximation using kernels needs to have a fine tuning of parameters. In this setting various methods become necessary to reduce the amount of calculations in a Monte Carlo setting.

Recently Nakahiro Yoshida, Naoto Kunitomo, Akihiko Takahashi, Yoshihiko Uchida, between others have proposed in a series of articles a control variate method using the idea of asymptotic expansions of Malliavin Calculus.

See for example, the following articles:

"Monte Carlo Simulation with Asymptotic Method," Working Paper Series CARF-F-011 (CIRJE-F-335), Center for Advanced Research in Finance, The University of Tokyo, 2005, forthcoming in Journal of Japan Statistical Society

"An Asymptotic Expansion Scheme for Optimal Investment Problems," Statistical Inference for Stochastic Processes, Vol.7, No.2, 153-188, 2004

"Applications of the Asymptotic Expansion Approach based on Malliavin-Watanabe Calculus in Financial Problems," Stochastic Processes and Applications to Mathematical Finance, World Scientific, 195-232, 2004 (

"On Validity of the Asymptotic Expansion Approach in Contingent Claim Analysis," Annals of Applied Probability, Vol.13, No.3, 914-952, 2003

"An Asymptotic Expansion Scheme for the Optimal Portfolio for Investment," Mathematical Economics, Kokyuroku 1215, Research Institute for Mathematical Sciences, Kyoto University, 2001 (

"A New Computational Scheme for Computing Greeks by the Asymptotic Expansion Approach" Working Paper CARF-F-044 (CIRJE-F-366), Center for Advanced Research in Finance, The University of Tokyo, 2005 forthcoming in Asia-Pacific Financial Markets (Special Issue on Mathematical Finance)

"New Acceleration Schemes with the Asymptotic Expansion in Monte Carlo Simulation," Working Paper CARF-F-012 (CIRJE-F-298), Center for Advanced Research in Finance, The University of Tokyo, 2004 forthcoming in Advances in Mathematical Economics, Springer

Recently, Ahmed Kebaier (universite Marne-LaVallee, Universite LeMans) has introduced a control variate method in the article

Statistical Romberg Extrapolation: A New Variance Reduction Method and Applications to Option Pricing. Ahmed Kebaier. To appear in Annals of Applied Probability.

This method is based on the idea of pilot studies. That is, one considers a pre-estimation with a smaller number of simulations which one uses for a control variate method.

A. Kebaier shows that there is reduction of the order $1/2$ in the number of simulations compared with the original method proposed by Bally and Talay.

We present the convergence analysis for a control variate method for diffusion density simulation using kernel density methods.

The control variance method is based on the well known idea of "pilot studies" in order to have an idea of the quantity to be simulated.

We prove that there is variance reduction if we use superkernels. Otherwise the variance reduction is not achieved.

This premilinary result is joint work with Ahmed Kebaier (U. Le Mans) and extends his previous result about this control variate method in the on regular case.

We also compute the optimal superkernel in the sense of asymptotic minimal variance of the estimator. It is proven that this has to be of exponential type combined with a polynomial type kernel.

To prove our results we need to use the results of Jacod-Kurtz-Protter on the weak convergence of the normalized error of the Euler scheme such as the ones found in

Jacod, Jean; Protter, Philip Asymptotic error distributions for the Euler method for stochastic differential equations. Ann. Probab. 26 (1998), no. 1, 267–307.

Kurtz, Thomas G.; Protter, Philip E. Weak convergence of stochastic integrals and differential equations. II. Infinite-dimensional case. Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995), 197–285, Lecture Notes in Math., 1627, Springer, Berlin, 1996.

Kurtz, Thomas G.; Protter, Philip E. Weak convergence of stochastic integrals and differential equations. Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995), 1–41, Lecture Notes in Math., 1627, Springer, Berlin, 1996.

Kurtz, Thomas G.; Protter, Philip Characterizing the weak convergence of stochastic integrals. Stochastic analysis (Durham, 1990), 255–259, London Math. Soc. Lecture Note Ser., 167, Cambridge Univ. Press, Cambridge, 1991.

Kurtz, Thomas G.; Protter, Philip Wong-Zakai corrections, random evolutions, and simulation schemes for SDEs. Stochastic analysis, 331–346, Academic Press, Boston, MA, 1991.

Kurtz, Thomas G.; Protter, Philip Weak limit theorems for stochastic integrals and stochastic differential equations. Ann. Probab. 19 (1991), no. 3, 1035–1070.

On a pricing rule in asymmetric information

Shigeyoshi OGAWA
Dept.of Mathematical Sciences
Ritsumeikan University
525-8577 Shiga,
JAPAN
ogawa-s@se.ritsumei.ac.jp

Monique PONTIER
Labo.de statist.et probabilités
Univ. Paul Sabatier
31 062 TOULOUSE cedex 04
FRANCE
pontier@cict.fr

Abstract

We are interested in an extension of the Kyle and Back' model on the market with insider traders [2],[1]. That is to say a model for the market with a continuous time risky asset and a-symmetrical information. The market is supposed to consist of three financial agents : the market maker, an insider trader (who knows a random variable V which will be revealed at final time) and a non informed agent.

In 1985 Kyle [2] defined an equilibrium problem. On a Gaussian financial market in discrete time, there are three agents: a market maker, an insider trader, a non informed agent (noise trader). The market maker has to define a rule price in such way that an equilibrium does exist between the traders. Back [1] extended this model to continuous time. Then N. El Karoui and K.Cho [3] relaxed the Gaussian hypothesis in Kyle's model using fine tools in stochastic control (cf. [5], [6]). After these works on Kyle and Back's model, K.Cho [4] delivered a new version of that model, also relaxing the Gaussian hypothesis. In these four papers, the non informed agent is supposed to be non strategic and so he/she is called "noise trader". As for the Cho's paper, we like to ask the question: what happens if the non informed agent tries to be strategic instead of being only "noisy"?

In this lecture we are concerned with these questions. Following our joint work [7], we are to give our recent results on the extension of the Kyle-Back's model to such more realistic case, where we assume that also the non informed agent is *strategic* in such sense that this noise trader uses a utility function to optimize his/her strategy. Our main results are as follows;

(1) We prove the existence of an equilibrium price when the insider trader and the non informed agent are risk-neutral.

(2) We will show that if such an equilibrium exists, then the non informed agent's optimal strategy is *do nothing*, in other words to be non strategic.

References

[1] K. BACK, "Insider trading in continuous time", Rev. Financial Studies, 5(3), 1992, 387-409.

[2] A.S.KYLE, "Continuous auctions and insider trading", Econometrica 53, 1985, 1315-1335.

[3] K.H.CHO and N. EL KAROUI, "Insider trading and nonlinear equilibria:uniqueness: single auction case" Annales d'économie et de statistique, 60, 21-41, 2000.

[4] K.H.CHO, "Continuous auctions and insider trading :uniqueness and risk aversion", Finance and Stochastics, 7-1, 47-71, 2003.

[5] Nicole EL KAROUI, Les aspects probabilistes du contrôle stochastique, L. N. in Maths, Ecole d'été de Saint Flour 1979, Springer, Berlin, 1981.

[6] W. H. FLEMING and R. W. RISHEL, Deterministic and Stochastic Optimal Control, Springer, Berlin, 1975.

[7] M.PONTIER and S.OGAWA, "Pricing rule in asymmetric information", *(to appear)* 2005

拡散過程の期待値の近似：ディリクレ条件付きの場合

楠岡成雄 (東京大学大学院数理科学研究科)

Stratonovich 型の SDE

$$dX^i(t,x) = \sum_{i=0}^{d} V_i(X(t,x)) \circ dB^i(t), \qquad X(0,x) = x \in \mathbf{R}^N \tag{1}$$

を考える。ただし、$B^0(t) = t$, $B(t) = (B^1(t), \ldots, B^d(t))$ は d-次元ブラウン運動、$V_i, i = 0, 1, \ldots, d$, は $\mathbf{R}^N$ 上の滑らかなベクトル場である。必ずしも滑らかでない関数 $f : \mathbf{R}^d \to \mathbf{R}$ に対して $(P_t f)(x) = E[f(X(t,x))]$ の値を高い精度で求めることが、ファイナンスでは重要な問題となる。f が滑らかならば、確率 Taylor 展開が有効となるが、f が滑らかでなくともブラウン運動の滑らかさに対する影響についてマリアバン解析が情報を与えてくれる、また、方程式 (1) はある意味で常微分方程式体系なので、ベクトル場の作るリー環が重要な情報を握っている。この２つを考慮した新しい方法が楠岡近似と呼ばれる方法である。このバージョンの一つとして、使いやすい方法を二宮-Victoir が見いだした。それは以下のようなものである。

$\mathbf{R}^N$ 上のベクトル場 W に対して ODE

$$\frac{d}{dt} y(t,x) = W(y(t,x)), \qquad y(0,x) = x$$

を考え、$y(1,x)$ を $\exp(W)(x)$ と表すことにする。また、 $C_b(\mathbf{R}^N)$ 上の線形作用素 $U(t;W)$ を

$$(U(t;W)f)(x) = f(\exp(tW)(x))$$

で表すことにする。

$s > 0$ に対して $C_b(\mathbf{R}^N)$ 上の線形作用素 $Q_{(s)}$ を

$$\begin{aligned}
&(Q_{(s)}f)(x) \\
&\doteq \frac{1}{2} E[(U(\frac{s}{2}V_0)U(B^1(s);V_1)\cdots U(B^d(s);V_d)U(\frac{s}{2};V_0)(x))f)(x)] \\
&\quad + \frac{1}{2} E[(U(\frac{s}{2}V_0)U(B^d(s);V_d)\cdots U(B^1(s);V_1)U(\frac{s}{2};V_0)(x))f)(x)]
\end{aligned}$$

で定義する。

この時、以下が証明される。

$T > 0$、$f : \mathbf{R}^N \to \mathbf{R}$ 有界可測とする。この時、条件 UFG の下で、ある定数 c が存在し、

$$(Q^n_{(T/n)} f)(x_0) = E[f(X(T,x_0))] + \frac{c}{n^2} + O(n^{-3})$$

となる。

$a(n) = (Q^n_{(T/n)} f)(x_0)$ とおくと、この量はほぼ、nd 次元の積分を計算することに他ならず、近似的に求める手段は色々と存在する（現在は 準乱数を用いた方法が最も効率的である。）$a(n)$ 自体が近似値であるが、上記結果を用いて、Romberg Extrapolation を行うと、例えば

$$\frac{4}{3} a(2n) - \frac{1}{3} a(n) = E[f(X(T,x_0))] + O(n^{-3})$$

となり、n^{-3} のオーダーの誤差の計算が可能となる。

上記の結果を示すためには、シンボル $v_0, v_1, \ldots, v_d$, の生成する単位元を持つ $\mathbf{R}$ 上の非可換多元環 $\mathbf{R}\langle\langle A \rangle\rangle$ が用いられる。

線形作用素 P_s 及び二宮-Victor 法に現れる線形作用素 $Q_{(s)}$ に対応する $\mathbf{R}\langle\langle A \rangle\rangle$ の元はそれぞれ

$$\eta_0(s) = \exp(s(v_0 + \frac{1}{2}\sum_{i=1}^{d} v_i^2))$$

$$\eta_1(s) = \frac{1}{2}\exp(\frac{s}{2}v_0)\exp(\frac{s}{2}v_1^2))\cdots\exp(\frac{s}{2}v_d^2)\exp(\frac{s}{2}v_0) + \frac{1}{2}\exp(\frac{s}{2}v_0)\exp(\frac{s}{2}v_d^2))\cdots\exp(\frac{s}{2}v_1^2)\exp(\frac{s}{2}v_0)$$

である。重要なポイントは、

(1) $\langle \eta_i(s) \rangle$, $i = 0, 1$, の v_0 の係数が共に s

(2) $\eta_0(s) - \eta_1(s)$ の重み 5 までの元の係数が 0

となることである。

$$Q^n_{(s)} = P_{ns} + \sum_{k=0}^{n-1} Q^k_{(s)}(Q_{(s)} - P_s)P_{(n-k-1)s}$$

$$= P_{ns} + \sum_{k=0}^{n-1} P_{ks}(Q_{(s)} - P_s)P_{(n-k-1)s} + \sum_{k=1}^{n-1}\sum_{\ell=0}^{k-1} Q^\ell_{(s)}(Q_{(s)} - P_s)P_{(k-\ell-1)s}(Q_{(s)} - P_s)P_{(n-k-1)s}$$

であることを用いて上記の結果が示せる。

ファイナンスにおいてはノックアウト・ノックイン条件の付いたデリバティブが存在する。その価格の計算は大ざっぱに言って次の量の計算に帰着する。

$$(P_t f)(x) = E[f(X(t,x)),\ \min_{s\in[0,t]} X^1(s,x) \geq 0], \qquad x \in \mathbf{R}^N,\ x^1 \geq 0,\ t \geq 0$$

ただし、X は SDE 1 の解である。

$u(t,x) = (P_t f)(x)$ とおくと $u(t,x)$ は粗く言って次の PDE の解となる。

$$\frac{\partial}{\partial t}u(t,x) = Lu(t,x), \qquad x^1 > 0, t > 0$$

$$u(t,(0,x')) = 0, \qquad t > 0,\ x' \in \mathbf{R}^{N-1}$$

境界条件は大域的な影響を与えるために局所的な議論だけでは近似は不可能になる。

今、簡単のために以下の条件を課す。

(仮定 D)

$$v_1^1(x) = 1, \qquad v_1^i(x) = 0, \quad i = 2, \ldots, N,$$

$$v_i^1(x) = 0, \qquad i = 0, 2, 3, \ldots, d$$

この仮定の下では

$$X^1(t,x) = x^1 + B^1(t)$$

となる。

この時、部分マリアバン解析によりこの時、UFG 条件の下で $P_t f$ の滑らかさの情報が得られる。条件 (D) の下では楠岡近似の考え方を適用することが可能となる。この時、注目すべき代数的対象は

$$E[\exp(-tv_0)\hat{X}(t)\exp(-B^1(t)v_1)|\ \min_{s\in[0,t]}(x^1 + B^1(s)) > 0,\ x + B^1(t) = y^1], \qquad t, x^1, y^1 > 0$$

となる。

そして $\mathcal{L}((A))$-値確率変数 $Z(t,x^1,y^1)$, $t, x^1, y^1 > 0$ で

$$\langle Z(t,x,y), v_0\rangle = t \qquad \langle Z(t,x,y), v_1\rangle = y^1$$

$$j_m(E[\hat{X}(t)|\ \min_{s\in[0,t]}(x^1 + B^1(s)) > 0,\ x + B^1(t) = y^1] = j_m(E[\exp(Z(t,x,y))|\ \min_{s\in[0,t]}(x^1 + B^1(s)) > 0,\ x + B^1(t) = y^1]$$

を満たす良いものを見つけることが問題となる。

現在のところ二宮-Victoir 法を変形することで 4 次までの近似法を見つけている。

On tightness of ℓ^∞-valued local martingales with infinitely many jumps: metric and partitioning entropy approach

西山陽一（統計数理研究所）

1. 序.

本報告では Nishiyama (2000a, 2000b) [今後 "N-00a" および "N-00b" と略す] の結果を拡張する。$(E,\mathcal{E})$ は Blackwell 空間であるとする。各 $n \in \mathbb{N}$ に対し μ^n は $\mathbb{R}_+ \times E$ 上の整数値ランダム測度であるとし、その予測可能カンペンセイターを ν^n とする。Ψ は一般の集合であるとし、それによって添字づけられた $\Omega^n \times \mathbb{R}_+ \times E$ 上の予測可能関数の族 $\mathcal{W}^n = \{W^{n,\psi} : \psi \in \Psi\}$ が与えられたとする。τ^n は有限停止時刻であるとする。我々は確率過程の列 $(t,\psi) \rightsquigarrow X_t^{n,\psi}$ を考える。ただし

$$X_t^{n,\psi} = W^{n,\psi} * (\mu^n - \nu^n)_t \qquad \forall t \in \mathbb{R}_+ \ \forall \psi \in \Psi.$$

N-00a はこの列の $n \to \infty$ としたときの弱収束を考慮した。すなわち、いわゆる quadratic modulus が確率有界で、かつ、ある分割エントロピー条件が満たされるならば、その列の $\ell^\infty([0,1] \times \Psi)$-空間における緊密性が従うことを証明した。しかしながら、N-00a は確率過程 $t \rightsquigarrow \overline{W}^n * \nu_t^n$ が局所可積分であることと、$\nu^n([0,\tau^n] \times E) < \infty$ が確率 1 で成り立つことを仮定した。ただし "$\overline{W}^n = \sup_{\psi \in \Psi} |W^{n,\psi}|$" である。[引用符の意味は、厳密には、可測性に対する考慮が必要である、ということである。] この仮定は確率過程のジャンプが有限区間の中では有限個しか起こらないということを導いてしまう。そのような場合も十分に豊富な応用を産み出すが、局所マルチンゲールの一般論の観点からは望ましくない仮定であった。

本報告の主たる貢献は、この仮定を次の 2 条件に置き換えることである：確率過程 $t \rightsquigarrow (|\overline{W}^n|^2 \wedge \overline{W}^n) * \nu_t^n$ は局所可積分である；Ψ は可算であって、ある種の分離条件を満たす。

我々はまず確率過程 $(t,\psi) \rightsquigarrow X_t^{n,\psi}$ が実際に ℓ^∞-空間に値をとるための十分条件を提示し、つぎにその緊密性を証明する。さらに、連続マルチンゲールを考慮した N-00b の結果と組み合わせることにより、より一般的な形

$$X_t^{n,\psi} = M_t^{n,\psi} + W^{n,\psi} * (\mu^n - \nu^n)_t,$$

をした局所マルチンゲールに対する緊密性判定条件を得る。ただし $t \rightsquigarrow M_t^{n,\psi}$ は連続マルチンゲールである。

2. Nishiyama (2000a) への追加.

上記のように、N-00a の 3 節の冒頭の記号を用いる。その Definition 2.2 で出てきた "DFP" の定義も思い出されたい。ここで我々は次のような条件を導入する。

Definition. Asymptotically separating. ある Ψ のある DFP $\Pi = \{\Pi(\varepsilon)\}_{\varepsilon\in(0,\Delta_\Pi]}$ が Ψ を漸近的に分離するとは、任意の有限部分集合 $F \subset \Psi$ に対してある正の数 ε_F が存在して、各 $\varepsilon \in (0, \varepsilon_F]$ に対し、各分割集合 $\Psi(\varepsilon; k)$ が F の点としては高々 1 個しか含まないときにいう。

Remark. これは強い要請ではない。Ψ 上にある距離 ρ が定義されている状況を考えよう。もしも各分割集合 $\Psi(\varepsilon; k)$ (ただし $\Psi = \bigcup_{k=1}^{N_\Pi(\varepsilon)} \Psi(\varepsilon; k)$) が $\mathrm{Diam}(\Psi(\varepsilon; k), \rho) \le \varepsilon$ を満たすならば、その DFP は漸近的に Ψ を分離することが容易にわかる。

さて、N-00a の仮定 (3.1) と (3.2) のかわりに、我々は次のいずれかを仮定する。(実際、Case A は N-00a の (3.1)+(3.2) と同じである。)

Case A: 過程 $t \rightsquigarrow \overline{W}^n * \nu_t^n$ は局所可積分で、かつ、$\nu^n([0, \tau^n] \times E) < \infty$ が確率 1 で成り立つ;

Case B: 過程 $t \rightsquigarrow (|\overline{W}^n|^2 \wedge \overline{W}^n) * \nu_t^n$ は局所可積分で、かつ、Ψ は可算で Π は Ψ を漸近的に分離する DFP である。

これらの準備に基づき、我々は次の主張をえる。

Theorem. N-00a の Theorem 3.2 および 3.4 と同じ主張が Case A だけでなく Case B のもとでも成り立つ。

3. Lipschitz weight の場合.

ここでは重み関数 (すなわち、確率積分の被積分関数) が Lipschitz 連続である場合への注意を与える。すなわち、我々は次の形をした場合を考える:

$$M_t^{n,\psi} = \int_0^t K_s^{n,\psi} dN_s^n \quad \text{such that} \quad |K_t^{n,\psi}(\omega) - K_t^{n,\phi}(\omega)| \le L_t^n(\omega)\rho_c(\psi, \phi),$$

ここに $t \rightsquigarrow N_t^n$ は連続マルチンゲール;

$$|W^{n,\psi}(\omega, t, z) - W^{n,\phi}(\omega, t, z)| \le H^n(\omega, t, z)\rho_d(\psi, \rho).$$

この場合、緊密性の判定条件は次のようなシンプルな形になる:

$$\int_0^{\tau^n} |L_t^n|^2 d\langle N^n\rangle_t + |H^n|^2 * \nu_{\tau^n}^n = O_{P^n}(1);$$

$$\int_0^1 \sqrt{\log N(\Psi, \rho_c; \varepsilon)} d\varepsilon + \int_0^1 \sqrt{\log N(\Psi, \rho_d; \varepsilon)} d\varepsilon < \infty.$$

ここに $N(\Psi, \rho; \varepsilon)$ は被覆数である。このことを見るには、DFP $\Pi = \{\Pi(\varepsilon)\}_{\varepsilon\in(0,\Delta_\Pi]}$ を ε-球によって生成されるものにとればよい。

引用文献.

Nishiyama, Y. (2000a). Weak convergence of some classes of martingales with jumps. *Ann. Probab.* **28** 685-712.

Nishiyama, Y. (2000b). *Entropy Methods for Martingales.* CWI Tract **128**, CWI, Amsterdam.

Covariance estimation of nonsynchronously observed diffusion processes

林　高樹 (HAYASHI, Takaki)*　　吉田　朋広 (YOSHIDA, Nakahiro)
慶應大学大学院経営管理研究科　　東京大学大学院数理科学研究科

We consider the problem of estimating the (integrated) covariance/correlation of two diffusion-type processes when they are observed at discrete times in a *nonsynchronous* manner. This kind of situation occurs typically in *high-frequency* financial data. The popular approach in the literature, the *realized covariance/correlation* estimator (e.g., [1], [2]), which is based on regularly spaced, synchronous data, is problematic because the choice of regular interval size and data interpolation scheme may lead to unreliable estimation. In the empirical finance literature, dependency of (synchronous data based) correlation measurements on sampling frequency has been indeed reported, known as the *Epps effect* ([3]).

Estimation problems of the diffusion parameter for diffusion processes based on discrete-time samples have been well studied in statistics. See, e.g., [9], [4]; however, nonsynchronicity seems to have been rarely treated.

To tackle the nonsynchronicity problem, in 2003 we proposed a new estimation procedure that is free of any 'synchronization' processing of original data. Suppose P^l follows the one-dimensional Itô process

$$dP_t^l = \mu^l(t)dt + \sigma^l(t)dW_t^l, \quad P_0^l = p^l,\ l = 1,2, \tag{0.1}$$

with $d\langle W^1, W^2\rangle_t = \rho(t)dt$, where $\rho(\cdot) \in (-1,1)$ is an unknown function, $p^l > 0$ is a constant, $\sigma^l(\cdot) > 0$ is a bounded (possibly unknown) function, $\mu^l(\cdot)$ is a progressively measurable (possibly unknown) function. Let $T \in (0,\infty)$ be an arbitrary terminal time for observing P^ls; for instance, $T = 1$ day in a typical application. We are interested in estimating $\langle P^1, P^2\rangle_T = \int_0^T \sigma^1(t)\sigma^2(t)\rho(t)dt$.

Suppose $(I^i)_{i=1,2,\ldots}$ and $(J^i)_{i=1,2,\ldots}$ are random intervals, reading from left to right, each of which partitions $(0,T]$. Let $T^{1,i} := \inf\{t \in I^{i+1}\}$ represent the ith observation time of P^1, and $T^{2,i} := \inf\{t \in J^{i+1}\}$ that of P^2, $i \geq 0$. Let n be an index representing the "size" of (I^i) and (J^i). Let $r_n := \max_{1\leq i<\infty} |I^i| \vee \max_{1\leq j<\infty} |J^j|$

Then, we proved that the statistic

$$U_n := \sum_{i,j} \left(P^1_{T^{1,i}} - P^1_{T^{1,i-1}}\right)\left(P^2_{T^{2,j}} - P^2_{T^{2,j-1}}\right) 1_{\{I^i \cap J^j \neq \emptyset\}} \tag{0.2}$$

consistently estimates the unknown $\langle P^1, P^2\rangle_T$ as $r_n \to 0$ ([8] for details).

In the case when the volatilities and correlation are constant, i.e., $\rho_t \equiv \rho$ and $\sigma_t^l \equiv \sigma^l$ for some $\rho \in (-1,1)$ and $\sigma^l > 0$, $l = 1,2$. The correlation ρ can be estimated consistently with the following statistics:

$$R_n^{(1)} := \frac{1}{T}\sum_{i,j} \frac{\left(P^1_{T^{1,i}} - P^1_{T^{1,i-1}}\right)\left(P^2_{T^{2,j}} - P^2_{T^{2,j-1}}\right)}{\sigma^1\sigma^2} 1_{\{I^i\cap J^j\neq\emptyset\}} \quad (\sigma^l \text{ are known}),$$

$$R_n^{(2)} := \frac{\sum_{i,j}\left(P^1_{T^{1,i}} - P^1_{T^{1,i-1}}\right)\left(P^2_{T^{2,j}} - P^2_{T^{2,j-1}}\right) 1_{\{I^i\cap J^j\neq\emptyset\}}}{\left\{\sum_i \left(P^1_{T^{1,i}} - P^1_{T^{1,i-1}}\right)^2\right\}^{1/2}\left\{\sum_j \left(P^2_{T^{2,j}} - P^2_{T^{2,j-1}}\right)^2\right\}^{1/2}} \quad (\sigma^l \text{ are unknown/known}).$$

[5] extended [8] and showed consistency of the estimators is preserved when the underlying processes are *continuous semimartingales* and the observation times are *stopping times*.

In the talk, we presented advances of the theory ([6], [7]). In particular, we discussed (i) *asymptotic normality* of the covariance estimator U_n, joint with the realized volatilities of P^1 and P^2, and, as its direct application, (ii) asymptotic normality of the correlation estimators $R_n^{(1)}$ and $R_n^{(2)}$, accompanied by a few examples.

Key words: diffusions; high-frequency data; nonsynchronicity; realized volatility

参考文献

[1] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.*, 96:42–55, 2001.

[2] O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized covariation: High-frequency based covariance, regression and correlation in financial economics. *Econometrica*, 72:885–925, 2004.

[3] T. W. Epps. Comovements in stock prices in the very short run. *J. Amer. Statist. Assoc.*, 74:291–298, 1979.

[4] V. Genon-Catalot and J. Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Insti. Henri Poincaré, Probab. Statist.*, 29(1):119–151, 1993.

[5] T. Hayashi and S. Kusuoka. Nonsynchronous covariation measurement for continuous semimartingales. Preprint 2004-21, Grad. Sch. of Math. Sci., Univ. of Tokyo (submitted), 2004.

[6] T. Hayashi and N. Yoshida. Asymptotic normality of nonsynchronous covariance estimators for diffusion processes. submitted, 2004.

[7] T. Hayashi and N. Yoshida. Estimating correlations with missing observations in continuous diffusion models. submitted, 2005.

[8] T. Hayashi and N. Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379, 2005.

[9] N. Yoshida. Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.*, 41:220–242, 1992.

Comparison of discrimination rules for diffusion processes *

YUJI SAKAMOTO[1]

[1] *Faculty of Human Environment, Hiroshima International University*
555-36 Gakuendai, Kurose-cho, Kamo-Gun, Hiroshima 724-0695 Japan

Key words and phrases : asymptotic expansion, diffusion process, discriminant analysis

1 Discrimination rules

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

$$\begin{cases} dX_t^{(k)} = V_0(X_t^{(k)}, \theta_k)dt + V(X_t^{(k)})dw_t^{(k)} \\ X_0^{(k)} \sim \nu_{\theta_k} \end{cases}, \tag{1.1}$$

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

$$d_B^{\theta_0}(\theta_1, \theta_2) = \ell(\theta_1; (X_t^{(0)})_{t\in[0,T_0]}) - \ell(\theta_2; (X_t^{(0)})_{t\in[0,T_0]})$$

and ℓ is the log-likelihood function defined by

$$\ell(\theta; (X_t)_{t\in[0,T]}) = \log \frac{d\nu_\theta}{dx}(X_0) + \int_0^T V_0'(VV')^{-1}(X_t, \theta)dX_t - \frac{1}{2}\int_0^T V_0'(VV')^{-1}V_0(X_t, \theta)dt.$$

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 particular, K is equal to 0 if $\pi_1 = \pi_2$ and $c(1|2) = c(2|1)$.

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:

$$d_W^{\theta_0}(\theta_1, \theta_2) = d_B(\hat\theta_1^{(1)}, \hat\theta_2^{(2)}).$$

Another rule for unknown parameter case is given by a likelihood ratio test. Suppose that for each $k = 0,1,2$, $X^{(k)} = (X_t^{(k)})_{t\in[0,T_k)}$ is a d-dimensional diffusion process satisfying (1.1) with a parameter

*AMS 1991 subject classification : 62E20, 62M05.

$\theta_k \in \Theta$ and an r-dimensional standard Wiener process $w^{(k)} = (w_t^{(k)})_{t\in[0,\infty)}$. In addition, suppose that driving processes $w^{(k)}$ $k = 0, 1, 2$, are independent of each other. Then the likelihood ratio test statistic of $H : \theta_0 = \theta_1$ against $K : \theta_0 = \theta_2$ is given by

$$d_Z^{\theta_0}(\theta_1,\theta_2) = \ell^{(0)}(\hat{\theta}_1^{(0,1)}) + \ell^{(1)}(\hat{\theta}_1^{(0,1)}) + \ell^{(2)}(\hat{\theta}_2^{(2)}) - \ell^{(0)}(\hat{\theta}_2^{(0,2)}) - \ell^{(1)}(\hat{\theta}_1^{(1)}) - \ell^{(2)}(\hat{\theta}_2^{(0,2)}),$$

where $\ell^{(k)}(\theta) = \ell(\theta; (X_t^{(k)})_{t\in[0,T_k]}))$, and $\hat{\theta}_1^{(0,1)}$ and $\hat{\theta}_2^{(0,2)}$ are the maximum likelihood estimators defined by

$$(\delta_a\ell^{(0)}(\hat{\theta}_1^{(0,1)}) + \delta_a\ell^{(1)}(\hat{\theta}_1^{(0,1)}))_{a=1}^p = 0, \quad (\delta_a\ell^{(0)}(\hat{\theta}_2^{(0,2)}) + \delta_a\ell^{(2)}(\hat{\theta}_2^{(0,2)}))_{a=1}^p = 0.$$

Note that $\hat{\theta}_1^{(0,1)}$ is not consistent under the hypothesis K, while $\hat{\theta}_2^{(0,2)}$ is not consisitent under H.

2 Expansions

Let

$$\begin{aligned}\bar{\nu}_{a_1\cdots a_m}(\theta;(k)) =& \frac{1}{T}E_{\theta_k}[\ell^{(k)}_{a_1\cdots a_m}(\theta)], \quad \bar{\nu}_{a_1\cdots a_m,b_1\cdots b_{m'}}(\theta;(k)) = \frac{1}{T}E_{\theta_k}[\ell^{(k)}_{a_1\cdots a_m}(\theta)\ell^{(k)}_{b_1\cdots b_{m'}}(\theta)],\\ \nu_{a_1\cdots a_m}(k) =& \bar{\nu}_{a_1\cdots a_m}(\theta_k;(k)), \quad \nu_{a_1\cdots a_m,b_1\cdots b_{m'}}(k) = \bar{\nu}_{a_1\cdots a_m,b_1\cdots b_{m'}}(\theta_k;(k)),\end{aligned}$$

$g_{ab}(k) = -\nu_{ab}(k)$, and $(g^{ab}(k)) = (g_{ab}(k))^{-1}$. Suppose that $d_B^{\theta_0}(\theta_1,\theta_2)$ has a density $p_B^{\theta_0}(\cdot;\theta_1,\theta_2)$ w.r.t the Lebesgue measure.

Theorem 1. *Under some regularity conditions, it holds that as T_1 and $T_2 \to \infty$ with T_0 fixed,*

$$\begin{aligned}P[d_W^{\theta_0}(\theta_1,\theta_2) < x] =& P[d_B^{\theta_0}(\theta_1,\theta_2) < x] + E[S_W^{\theta_0}(\theta_1,\theta_2) \mid d_B^{\theta_1}(\theta_1,\theta_2) = x]p_B^{\theta_0}(x;\theta_1,\theta_2)\\ &+ \partial_x(E[U_W^{\theta_0}(\theta_1,\theta_2) \mid d_B^{\theta_1}(\theta_1,\theta_2) = x]p_B^{\theta_0}(x;\theta_1,\theta_2)) + o(\frac{1}{T}),\end{aligned}$$

where

$$\begin{aligned}S_W^{\theta_0}(\theta_1,\theta_2) =& \sum_{k=1}^{2}(-1)^k T_k^{-1}\Big(\ell_a^{(0)}(\theta_k)g^{aa'}(k)g^{bc}(k)\big(\nu_{a'b,c}(k) + \frac{1}{2}\nu_{a'bc}(k)\big) + \ell_{ab}^{(0)}(\theta_k)g^{ab}(k)\Big),\\ U_W^{\theta_0}(\theta_1,\theta_2) =& \sum_{k=1}^{2} T_k^{-1}g^{ab}(k)\ell_a^{(0)}(\theta_k)\ell_b^{(0)}(\theta_k).\end{aligned}$$

Theorem 2. *Under some regularity conditions, it holds that as T_1 and $T_2 \to \infty$ with T_0 fixed,*

$$P[d_Z^{\theta_0}(\theta_1,\theta_2) < x] = P[d_W^{\theta_0}(\theta_1,\theta_2) < x] + E[S_Z^{\theta_0}(\theta_1,\theta_2) \mid d_B^{\theta_1}(\theta_1,\theta_2) = x]p_B^{\theta_0}(x;\theta_1,\theta_2) + o(\frac{1}{T}),$$

where

$$S_Z^{\theta_0}(\theta_1,\theta_2) = T_0\sum_{k=1}^{2}(-1)^k T_k^{-1}g^{ab}(k)\bar{\nu}_a(\theta_k,(0))\Big(\ell_b^{(0)}(\theta_k) - T_0\bar{\nu}_b(\theta_k;(0))\Big).$$

Strong Ignorability が成立しない場合の傾向スコアを用いた周辺分布のパラメータ推定について

星野 崇宏　倉田 博史　繁桝 算男 (東京大学大学院総合文化研究科)

1 目的

Rosenbaum &Rubin(1983) は、無作為割り当てが行えない研究（観察研究）において無作為割り当ての近似を行うために、傾向スコア (Propensity score) の概念を提案した。傾向スコアは共変量を与えたときに各被験者が各条件（群）に割り当てられる確率である。割り当てや標本抽出が共変量の値に依存する場合は、傾向スコアを用いることで共変量の分布を群間で共通にすることが出来るため、無作為割り当ての近似が可能になる。このような性質から傾向スコアは医学や経済学における準実験データでの因果解析法として近年非常に頻繁に利用されている。しかし、既存の傾向スコアによる調整法の問題点として、Strong Ignorability の仮定、すなわち結果変数 $\boldsymbol{Y}_j$ と割り当て変数 Z が共変量 $\boldsymbol{X}$ を所与として独立

$$\boldsymbol{Y}_j \perp\!\!\!\perp Z|\boldsymbol{X} \quad j=1,2. \tag{1}$$

という強い仮定に依存している。本研究の目的は、割り当てが共変量だけでなく結果変数にも依存している場合に、結果変数の周辺分布の母数推定を行う方法の開発と、その漸近的性質を示すことにある。

2 モデル設定

簡単のために条件（群）の数を 2 とする。このとき各ユニット（被験者）は理論上、2 個の条件ごとに結果変数の値を有する。しかし観測されるのはそのユニットが属する条件における結果変数のみとする。具体的には、y_{ij}, x_i, z_i をそれぞれ、第 i 被験者の第 j 条件における結果変数の値、共変量の値、(1 なら第一群、0 なら第二群を示す) 割り当て変数とする。このとき、結果変数では欠測が起っているみなすことができる。ここで、$w(y_i, x_i, \alpha) = p(z_i = 1|y_i, x_i, \alpha)$ は結果変数と共変量を所与としたときに第 i 被験者の y_{i1} が観測される確率であり、α は母数ベクトルとする。ここで、 y_{ij} の周辺分布を $p(y_{ij}|\theta_j(\theta))$ $(i = 1\cdots N)$ と置く。

ここで、全確率変数の同時分布は $p(y_1, y_2, z, x)$ で表現されるが、この分布を仮定することはできないものとする。さらに、一般の傾向スコア解析同様、共変量を所与とした時の結果変数の条件付分布 $p(y_1, y_2|x)$ を仮定することが難しいとする。仮定できるのは、割り当ての分布 $p(z|y_1, y_2, x)$ と結果変数 y_1, y_2 の周辺分布だけだとする。

3 擬似条件付分布を用いた推定方程式

結果変数がすべて観測された場合の推定方程式を

$$S(\alpha) = \sum_{i=1}^{N} S(\alpha|z_i, y_i, x_i) = \sum_{i=1}^{N} \frac{\partial}{\partial \alpha}\{z_i w(y_{i1}, y_{i2}, x_i, \alpha) + (1-z_i)(1-w(y_{i1}, y_{i2}, x_i, \alpha))\}, \tag{2}$$

$$\begin{aligned} S(\theta|\alpha) &= \sum_{i=1}^{N} S(\theta|y_i, x_i, z_i, \alpha) \\ &= \sum_{i=1}^{N} \frac{\partial}{\partial \theta}\Big[\frac{z_i}{w(y_{i1}, y_{i2}, x_i, \alpha)} \log p(y_{i1}|\theta_1(\theta)) + \frac{1-z_i}{1-w(y_{i1}, y_{i2}, x_i, \alpha)} \log p(y_{i2}|\theta_2(\theta))\Big], \end{aligned} \tag{3}$$

とおく。しかし、実際は割り当て変数の値に応じて結果変数が欠測になるので、上記の推定方程式を利用することはできない。そこで、2つの関数 $S(\alpha|z_i, y_i^{obs}, x_i, \theta^*, \alpha^*) = E_{y_i^{mis}|z_i,x_i,\theta^*,\alpha^*}[S(\alpha|z_i, y_i, x_i)]$ と $S(\theta|y_i^{obs}, x_i, z_i, \theta^*, \alpha^*) = E_{y_i^{mis}|z_i,x_i,\theta^*,\alpha^*}[S(\theta|y_i, x_i, z_i, \alpha)]$ を定義する。但し、上付き記号 mis と obs は欠測された結果変数と観測された結果変数を示し、

$$p(y_{ij}|y_{ik}, z, x, \theta, \alpha) = \frac{p(y_{ik}|y_{ij}, x_i)p(z_i|y_i, x_i, \alpha)p(y_{ij}|x_i)}{\int p(y_{ik}|y_{ij}, x_i)p(z_i|y_i, x_i, \alpha)p(y_{ij}|x_i)dy_{ij}}, \quad j=1, k=2 \;\; or \;\; j=2, k=1 \tag{4}$$

ここで、$S_{obs}(\alpha|\theta, \alpha) = \sum_{i=1}^N S(\alpha|z_i, y_i^{obs}, x_i, \theta^*, \alpha^*)$, $S_{obs}(\theta|\theta, \alpha) = \sum_{i=1}^N S(\theta|y_i^{obs}, x_i, z_i, \theta^*, \alpha^*)$ は unbiased estimating equation であるため、この推定方程式から、EM algorithm と同様に期待値計算と解の導出によって母数の一致推定量を得ることができる (Rosen, Jiang and Tanner, 2000)。しかし、観測データを所与としたときの欠測された結果変数の条件付分布 $p(y_{ij}|y_{ik}, z, x, \theta, \alpha)$ の計算には共変量を所与とした時の結果変数の条件付分布 $p(y|x)$ を仮定することが出来ないために計算ができない。そこで、結果変数の擬似条件付分布

$$p^*(y_{ij}|y_{ik}, z_i, x_i, \theta, \alpha) = \frac{p(y_{ij}|\theta_j(\theta))p(y_{ik}|\theta_k(\theta))p(z_i|y_i, x_i, \alpha)}{p(y_{ik}, z_i|x_i)} \quad j=1, k=2 \; or \; j=2, k=1. \tag{5}$$

を利用した期待値計算を行い、以下の関数を定義する：

$$\begin{aligned} S^q(\alpha|z_i, y_i^{obs}, x_i, \theta^*, \alpha^*) &= \frac{\partial}{\partial\alpha}\{\int z_i w(y_{i1}, y_{i2}, x, \alpha)p^q(y_{i2}|y_{i1}, z_i, x_i, \theta^*, \alpha^*)dy_{i2} \\ &\quad + \int (1-z_i)(1-w(y_{i1}, y_{i2}, x, \alpha))p^q(y_{i1}|y_{i2}, z_i, x_i, \theta^*, \alpha^*)dy_{i1}\}, \end{aligned} \tag{6}$$

$$\begin{aligned} S^q(\theta|y_i^{obs}, x_i, z_i, \theta^*, \alpha^*) &= \frac{\partial}{\partial\theta}[\int \frac{z_i}{w(y_{i1}, y_{i2}, x_i, \alpha^*)}\log p(y_{i1}|\theta_1(\theta))p^q(y_{i2}|y_{i1}, z_i, x_i, \theta^*, \alpha^*)dy_{i2} \\ &\quad + \int \frac{1-z_i}{1-w(y_{i1}, y_{i2}, x_i, \alpha^*)}\log p(y_{i2}|\theta_2(\theta))p^q(y_{i1}|y_{i2}, z_i, x_i, \theta^*, \alpha^*)dy_{i1},] \end{aligned} \tag{7}$$

3 本研究で利用する Expectaiton-Solution アルゴリズム

$l = 0, 1, 2, \cdots$ に対して、以下の2ステップを繰り返し実行する：

Expectation Step Metropolis-Hastings アルゴリズムを利用して、y_{i1} を $p^*(y_{i1}|y_{i2}, z_i, x_i, \theta, \alpha)$ から発生させる (y_{i2} も同様)。Monte Carlo E step (Wei and Tanner, 1990) を利用して、以下の関数を計算する。

$$S_{obs}^q(\alpha|\theta^l, \alpha^l) = \sum_{i=1}^N S^q(\alpha|z_i, y_i^{obs}, x_i, \theta^*, \alpha^*), \qquad S_{obs}^q(\theta|\theta^l, \alpha^l) = \sum_{i=1}^N S^q(\theta|y_i^{obs}, x_i, z_i, \theta^*, \alpha^*) \tag{8}$$

Solution Step

$$S_{obs}^q(\alpha|\theta^l, \alpha^l) = 0, \;\; and \;\; S_{obs}^q(\theta|\theta^l, \alpha^l) = 0. \tag{9}$$

となる α を α^{l+1}、θ を θ^{l+1} とする。

ここで $S^q(\alpha|\theta, \alpha)$, $S^q(\theta|\theta, \alpha)$ が unbiased estimating equation であることが証明できるため、上記のアルゴリズムで得られる解は、一致性と漸近正規性を有することを証明できる。

参考文献

[1] Rosen, O., Jiang, W., and Tanner, M.A. (2000). Mixtures of marginal models. *Biometrika,* 87, 391–404.

[2] Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70, 41–55.

[3] Wei, G.C.G., and Tanner, M.A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association,* 85 , 699–704.

単調変換による4次キュムラントの除去

若木 宏文（広島大学大学院理学研究科）

1　3次キュムラントの除去

確率変数列 X_n の分布関数が、

$$\mathrm{P}(X_n \le x) = \Phi(x) + \frac{1}{\sqrt{n}}\phi(x)(ax^2+b) + O(\frac{1}{n})$$

と展開されるとする。多項式

$$t(x) = x + \frac{1}{\sqrt{n}}(ax^2+b)$$

によって、

$$t(x) = x + \frac{1}{\sqrt{n}}(ax^2+b) + \frac{a^2}{3n}x^3 \tag{1}$$

となるが、t は、単調ではないため、例えば、母数 θ の信頼区間を得るために $X_n = n^{-1/2}(\hat{\theta}-\theta)/\hat{\sigma}$ に適用する場合、被覆確率と区間幅に単調関係が成立しない。

Hall (1992) は、$t(x)$ に、3次の項を追加した、

$$t_1(x) = x + \frac{1}{\sqrt{n}}(ax^2+b) + \frac{a^2}{3n}x^3$$

によって (1) を満たす単調変換を提案した。$\tilde{t}(x) = t(x) + O(\frac{1}{n})$ を満たす単調変換は、他にも多く考えられるが、t_1 は、逆関数が陽に与えられるという利点がある。

無作為標本に基づく推測において、多くの基準化、あるいは Student 化された統計量 (T) は

$$\log \mathrm{E}[\exp(isT)] = \frac{1}{2}(is)^2 + \frac{1}{\sqrt{n}}\left\{a_1(is)+a_3(is)^3\right\} + \frac{1}{n}\left\{a_2(is)^2+a_4(is)^4\right\} + O(\frac{1}{n\sqrt{n}}) \tag{2}$$

のように展開される。ここで、a_1, a_3 は、3次キュムラントのみに、また、a_2, a_4 は、3次および4次キュムラントに依存する。Hall (1992) の単調3次変換において、3次キュムラントをその $\sqrt{n}$–一致推定量で置き換えても、(1) は成り立ち、これによって、正規近似に対する、3次キュムラントの影響を除去することができる。本報告では、漸近展開式の2次 $(O(\frac{1}{n}))$ の項まで利用し、正規近似の精度を上げるような単調変換を導出する。

2　多項式変換による近似の改良

確率変数列 $\{X_n\}_{n=1,2,\cdots}$ の極限分布関数を F とし、X_n の分布関数 F_n が次のように展開されるとする。

$$F_n(x) := \mathrm{P}(X_n \le x) = F(x) - f(x)\{n^{-d}g_1(x) + n^{-2d}g_2(x)\} + O(n^{-3d})$$

ここで、$d > 0,\ f(x) = F'(x)$ である。このとき、F_n の上側 $100\alpha\%$ 点の Cornish–Fisher 展開は、F の上側 α 点を x_α として、$q(x_\alpha)$ によって与えられる。ここで、

$$
\begin{aligned}
q(x) &= x + n^{-d}h_1(x) + n^{-2d}h_2(x),\\
h_1(x) &= g_1(x),\ h_2(x) = g_2(x) + \frac{\{g_1(x)\}^2 f'(x)}{2f(x)} + g_1(x)g_1'(x),
\end{aligned}
$$

である。多くの場合、F は標準正規分布、または、χ^2 分布であり、g_1, g_2 は多項式である。そこで、以下、h_1, h_2 は多項式であるとする。

正数 M に対して n を十分大きくとれば、$-M \le x \le M$ で、$q(x)$ は狭義単調増加となり、その逆関数を $t(x) = q^{-1}(x)$ は、区間 $(q(-M), q(M))$ において

$$t(x) = x - n^{-d}h_1(x) + n^{-2d}\{h_1(x)h_1'(x) - h_2'(x)\} + O(n^{-3d}) \tag{3}$$

と展開される。

3 単調変換

Hall (1992) の単調変換は、次のように一般化される。

$$t_1(x : h_1) = \int_0^x \left[1 - n^{-d}h_1'(t) + \frac{1}{4}n^{-2d}\{h_1'(t)\}^2\right] dt \tag{4}$$

と定義すると、$t_1(x : h_1)$ は単調で、かつ、$t(x : h_1) = t_1(x) + O_p(n^{-2d})$ より、

$$\mathrm{P}(t_1(X_n : h_1) \ge x_\alpha) = \alpha + O(n^{-2d}) \tag{5}$$

を満たす。$t_1(X_n : h_1)$ の分布の漸近展開を導出し、パーセント点の Cornish–Fisher 展開の逆変換の展開式を、あらためて t と考えて (4) を用いると極限分布による近似誤差のオーダーを $O(n^{-3d})$ とすることができる。しかし、例えば、Hall の３次変換に続けて、４次キュムラントを含む、次のオーダーの項まで除去しようとすると、５次多項式を用いることとなり、逆変換は数値計算に拠らざるをえない。

$t_1(x : h_1)$ の逆変換 $q_1(x : h_1) = t_1^{-1}(x : h_1)$ が陽に求まるとする。$q(x : h_1)$ の漸近展開は、Cornish–Fisher 展開と $O(n^{-d})$ の項まで一致する。$q(x)$ と $t(x)$ の $O(n^{-d})$ の項とは、符号が逆になっているのみなので、$q_1(x : -h_1)$ も、単調でかつ、(5) を満たす。さらに、$t_1(x : \frac{1}{2}h_1)$ と $q_1(x : -\frac{1}{2}h_1)$ の合成関数も、(5) を満たすことが容易にわかるが、一般に、任意の多項式 $p_1(x)$ を用いて、$t_1(x : p_1 + h_1)$ と $q_1(x : p_1)$ との合成変換もまた、単調でかつ、(5) を満たす。p_1 をうまく選ぶことで、$O(n^{-2d})$ まで除去することが可能であり、特に、(2) の展開を持つ場合に対しては、Hall 型の３次変換と、その逆変換の合成変換によって４次キュムラントの影響を除去できることがわかった。一般に３次、４次キュムラントは未知であるが、一致推定量を利用することで、極限分布による近似精度を同じオーダーまで保つことができる。例として、単調変換を利用した母平均の信頼区間の構成法を示す。

Asymptotic efficiency of conditional least squares estimators for ARCH models

Tomoyuki AMANO and Masanobu TANIGUCHI*

Department of Mathematical Sciences, School of Science and Engineering, Waseda University, Tokyo, 169-8555, Japan

Abstract

The conditional least squares (CL) estimators proposed by Tjϕstheim (1986) are important and fundamental. If we apply the CL estimators to the square transformed ARCH models, they have a simple explict form, and do not depend on the innovation distribution. Since the CL's are not asymptotically efficient in general we give a necessary and sufficient condition that CL is asymptotically efficient based on the LAN approach. Next, a measure of efficiency for CL is introduced. Then numerical evaluations of the inefficiency for various nonlinear time series models are given. They elucidate some interesting features of CL.

Keywords: ARCH model; Conditional least squares estimator; asymptotically efficiency, LAN approach

1. Introduction

Traditional time series models such as ARMA models assume a constant one-period forecast variance. However, in actual practice, this assumption is often violated, especially in economic time series. In order to circumvent this difficulty, Engle (1982) and Bollerslev (1986) introduced, respectively, the autoregressive conditional heteroscedastic (ARCH) model and the generalized ARCH (GARCH) model. Since then, a great number of theoretical and empirical studies have been conducted for them (c.f.,Engle, 1995, Linton, 1993 and Taniguchi and Kakizawa, 2000). Recently, Giraitis, Kokoszaka and Leipus (2000) introduced a class of ARCH(∞) models, which includes the ARCH and GARCH models as special cases, and provided sufficient conditions for the existence of a stationary solution and its explict representation.

LeCam (1960) established the most important and sophisticated foundation of the general statistical asymptotic theory. He introduced the concept of local

*Corresponding author. +81-3-5286-8095.

E-mail addresses: tomtochami@ruri.waseda.jp (T. Amano), taniguchi@waseda.jp (M. Taniguchi)

asymptotic normality (LAN) for the likelihood ratio of general statistical models. Once LAN is proved, the asymptotic optimality of estimators and tests is described in terms of the LAN property. Lee and Taniguchi (2005) established the LAN for a class of ARCH(∞)-SM models, which include ARCH models. Then they described the asymptotic optimality of estimators in terms of LAN.

Tjϕstheim (1986) proved a consistency and asymptotic normality of the conditional least squares estimators for stationary processes. The CL for ARCH is written as a simple linear form and we can construct it without knowing the distribution of the innovation density. Since the CL is the most fundamental estimator for the parameter of ARCH model, it is important to investigate its goodness based on LAN. This paper provide a necessary and sufficient condition for the CL to be asymptotically efficient in the sense of LAN and discuss the inefficiency of it.

This paper is organized as follows. In Section 2 we explain the LAN for ARCH(q), and give the lower bound of the asymptotic variance for estimators. Comparing the asymptotic variance of CL with the lower bound, we evaluate the inefficiency of CL. Section 3 provides numerical studies of the inefficiency for various models. The results elucidate some interesting features of the asymptotics of CL. Proof of the main theorem is relegated to Section 4.

References

[1] Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica. 50, 987–1007.

[2] Kholevo, A.S., 1969. On estimates of regression coefficients, Theor. Prob. Appl. 14, 79-104.

[3] Le Cam, L., 1960. Locally asymptotically normal families of distributions. Univ. CaliforniaPubl. Statist. 3, 37–98.

[4] Lee, S., Taniguchi, M., 2005. Asymptotic theory for ARCH-SM models: LAN and residual empirical processes, Statist. Sinica. 15, 215–234.

[5] Linton, O., 1993. Adaptive estimation in ARCH models, Econometric Theory. 9, 539–569.

[6] Taniguchi, M., Kakizawa, Y., 2000. *Asymptotic Theory of Statistical Inference for Time Series.* Springer, New York.

[7] Tjostheim, 1986. Dag Estimation in nonlinear time series models, Stochastic Process. Appl. 21, 251–273.

Statistical Estimation of Optimal Portfolios for Locally Stationary Returns of Assets

Hiroshi Shiraishi
Waseda University
Masanobu Taniguchi
Waseda University

Abstract

This paper discusses the asymptotic property of estimators for optimal portfolios when the returns are vector-valued locally stationary processes. First, we derive the asymptotic distribution of a nonparametric portfolio estimator based on the kernel method. Optimal bandwidth and kernel function are given by minimizing the mean squares error of it. Next, assuming parametric models for non-Gaussian locally stationary processes, we prove the LAN theorem. We propose a parametric portfolio estimator $\hat{g}$ based on a quasi-maximum likelihood estimator. Then it is shown that $\hat{g}$ is asymptotically efficient based on the LAN. Numerical studies are provided to investigate the accuracy of the portfolio estimators parametrically and nonparametrically. They illuminate some interesting features of them.

1. Introduction

In the theory of portfolio analysis, optimal portfolios are determined by the mean $\boldsymbol{\mu}$ and variance Σ of the portfolio return. Several authors proposed estimators of the optimal portfolios as the functions of the sample mean $\hat{\boldsymbol{\mu}}$ and the sample variance $\hat{\Sigma}$ for independent returns of assets. However, empirical studies show that financial return processes are often dependent. From this point of view, Basak, Jagannathan and Sun (2002) showed the consistency of optimal portfolio estimators when the portfolio returns are stationary processes. Furthermore, Shiraishi and Taniguchi (2005) discussed the asymptotic efficiency for optimal portfolios when returns are non-Gaussian stationary processes. Although they dropped Gaussianity and independence for return processes, many empirical studies show that real time series data are generally non-stationary. To overcome this problem, Dahlhaus (1996a) proposed an important class of locally stationary processes, and discussed the kernel methods for local estimators of the covariance structure. Regarding parametric approach, Dahlhaus (1996b) developed asymptotic theory for Gaussian locally stationary process, and derived the LAN property. Furthermore Hirukawa and Taniguchi (2005) dropped the Gaussian assumption, i.e., they proved the LAN theorem for non-Gaussian locally stationary process with mean vector $\boldsymbol{\mu} = 0$, and applied the results to the asymptotic estimation.

In this paper, denoting the optimal portfolios by a function $g = g(\boldsymbol{\mu}, \Sigma)$ of $\boldsymbol{\mu}$ and Σ, we discuss the asymptotic property of estimators $\hat{g} = g(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$ when the returns are vector-valued locally stationary processes. Since the returns are non-stationary, the quantities $\boldsymbol{\mu}, \Sigma, \hat{\boldsymbol{\mu}}, \hat{\Sigma}$ depend on the time when the portfolio is constructed.

Section 2 gives the asymptotic distribution of a nonparametric estimator $\hat{g}$ based on the kernel method when the returns are Gaussian locally stationary process. We evaluate the mean squares error of the estimators, and determine the optimal bandwidth and kernel function. Assuming parametric structure $\boldsymbol{\mu}_\theta$ and Σ_θ, section 3 proves the LAN theorem for non-Gaussian locally stationary process with non-zero mean vector. In this setting we propose a parametric estimator $\hat{g} = g(\boldsymbol{\mu}_{\hat{\theta}}, \Sigma_{\hat{\theta}})$ where $\hat{\theta}$ is a quasi-maximum likelihood estimator of θ. Then it is shown that $\hat{g}$ is asymptotically efficient. Numerical studies are provided to confirm the accuracy of $\hat{g}$ nonparametrically and parametrically. They show some interesting features of $\hat{g}$.

Throughout this paper, if $\{X_n\}$ is a sequence of random vectors which converges in distribution to a random vector X, then we write $X_n \xrightarrow{\mathcal{D}} X$. The 'vec' operator transforms a matrix into a vector by stacking the columns, and the 'vech' operator transforms a symmetric matrix into a vector by stacking the elements on and below the main diagonal. We denote the set of all positive integers, the set of all real numbers, and the set of all complex numbers by $\mathbf{N}, \mathbf{R}$ and $\mathbf{C}$, respectively.

References

[1] Basak, G. & Jagannathan, R. & Sun, G. (2002) A direct test for the mean variance efficiency of a portfolio. *Journal of Economic Dynamics and Control* **26**, 1195-1215.

[2] Brockwell, P. J. and Davis, R. A. (1987) *Time Series:Theory and Methods.* New York: Springer.

[3] Brillinger,D. R. (1981) *Time Series:Data Analysis and Theory*, expanded. San Francisco: Holden-Day.

[4] Dahlhaus, R. (1996a) Asymptotic statistical inference for nonstationary processes with evolutionary spectra. *Athens Conference on Applied Probability and Time Series 2. Lecture Notes in Statist.* **115** 145-159. New York: Springer.

[5] Dahlhaus, R. (1996b) Maximum likelihood estimation and model selection for locally stationary processes. *J. Nonparameter. Statist.* **6** 171-191.

[6] Dahlhaus, R. (2000) A likelihood approximation for locally stationary processes. *The Annals of Statistics.* **28** 1762-1794.

[7] Hirukawa, J. and Taniguchi, M. (2005) LAN theorem for non-Gaussian locally stationary processes and its applications *J. Statist. Plan. Inf.* To be published.

[8] Shiraishi, H. and Taniguchi, M. (2005) Statistical Estimation of Optimal Portfolios for non-Gaussian Dependent Returns of Assets. *Waseda University Time Series Discussion Paper.*

[9] Taniguchi, M & Kakizawa, Y. (2000) *Asymptotic Theory of Statistical Inference for Time Series.* New York: Springer.

観測値が次元より少ない場合の多変量分散分析法と判別法

中央大学・理工　　藤越　康祝

多変量線形仮説の検定, あるいは, 正準判別法において, 標本数が次元より小さいと群内平方和積和行列 W が特異になり, 通常の方法を利用することができなくなる. このような場合の1つの解決策として, Srivastava and Fujikoshi(2005), Srivastava(2005) は, W の逆行列をムーアペンローズ逆行列 W^+ で置き換えてた方法を提案し, 統計的性質を調べている. 本報告では, これらの結果を紹介している.

$N \times p$ の観測行列 Y についての多変量線形回帰モデル

$$Y = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)' = X\Xi + E$$

を考える. ここに, X はランク $k(< N)$ の $N \times k$ 既知計画行列, Ξ は $k \times p$ の未知パラメータ行列である. 誤差行列 $E = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N)'$ の各行は独立同一で平均ゼロ, 共分散行列 Σ の正規分布, $\boldsymbol{e}_i \sim N_p(\boldsymbol{0}, \Sigma)$ に従うとする. この論文では, 観測値が次元より少ない, $N \leq p$ の場合を考える. このような状況は, マイクロアレイデータのように僅かの個体に対して数千の遺伝子発現を扱う場合, などで生じる.

線形仮説の検定問題

$$H \ : \ C\Xi = 0 \quad \text{vs} \quad A \ : \ C\Xi \neq 0$$

を考える. ここに, C はランク $q(\leq k)$ の $q \times k$ の定数行列である. 仮説による平方和積和行列 B, および, 誤差による平方和積和行列 W はそれぞれ

$$B = N\left(C\hat{\Xi}\right)'\left[CGC'\right]^{-1}C\hat{\Xi}, \quad W = (Y - X\hat{\Xi})'(Y - X\hat{\Xi})$$

で与えられる. ここに, $G = \left[N^{-1}X'X\right]^{-1}$, $\hat{\Xi} = (X'X)^{-1}X'Y$ は Ξ の最尤推定量あるいは最小2乗推定量である. Gは正定値であると仮定する. 正規性のもとでWはウィシャート分布 $W \sim W_p(\Sigma, n)$, B は非心ウィシャート分布 $B \sim W_p(\Sigma, q, N\eta\eta')$ に従い, これらは互いに独立である. ここに, $n = N - k$, $\eta = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_q) = (C\Xi)'(CGC')^{-\frac{1}{2}}$.

$p \to \infty$ の場合の漸近理論においては,

$$a_i = \frac{\text{tr } \Sigma^i}{p}; \ i = 1, \ldots, 4$$

とおくとき,

$$0 < \lim_{p \to \infty} a_i = a_{i0} < \infty \ , i = 1, \ldots, 4$$

を仮定する. このとき, 推定量

$$\hat{a}_1 = (\text{tr } W)/np, \quad \hat{a}_2 = \frac{1}{(n-1)(n+2)p}\left[\text{tr } W^2 - \frac{1}{n}(\text{tr } W)^2\right]$$

は, n および$p \to \infty$ のとき, a_i の一致推定量であることが知られている (Srivastava(2004)). したがって, $\hat{b} = \hat{a}_1^2/\hat{a}_2$ も $b = a_1^2/a_2$ の推定量として同様な性質をもつ.

$N < p$ のとき, W が特異になるのが, Dempster(1959) の一般化

$$\tilde{T}_1 = \frac{(pq)^{-1}\ \mathrm{tr}\ B}{\hat{a}_1} = \frac{n\ \mathrm{tr}\ B}{q\ \mathrm{tr}\ W}$$

を利用することができる. $\tilde{T}_1$ のパワーを調べるため, 漸近的に同等である

$$T_1 = \left[2q\hat{a}_2(1+n^{-1}q)\right]^{-\frac{1}{2}}\left[\frac{\mathrm{tr}\ B}{\sqrt{p}} - \frac{q}{\sqrt{n}}\frac{\mathrm{tr}\ W}{\sqrt{np}}\right]$$

を考える. 尤度比検定, Lawley-Hotelling 検定, Bartlett-Nanda-Pillai 検定に関しては, W の逆行列をムーアペンローズ逆行列 W^+ で置き換えた検定統計量

$$T_2 = -p\hat{b}\log\Pi_{i=1}^{q}\ (1+c_i)^{-1}, \quad T_3 = p\hat{b}\sum_{i=1}^{q} c_i, \quad T_4 = p\hat{b}\sum_{i=1}^{q}\frac{c_i}{1+c_i}$$

を考える. ここに, c_i は BW^+ のゼロでない固有値である. まず, $p \to \infty$ のとき, 検定統計量 T_2, T_3, T_4 は漸近的に同等であることを示す. したがって, 最終的には, T_1 と T_2 について考える. これら２つの検定統計量の仮説および対立仮説のもとでの漸近分布が求められる. とくに局所対立仮説のもとで, p が大きく, n が小さいと検定 T_2 が T_1 より良いことが示される. 詳細については, Srivastava and Fujikoshi(2005) を参照されたい. $p/n \to c, 0 \le c < 1$ の場合のパワー比較については, Fujikoshi et al.(2004) を参照されたい.

共分散行列が等しい k 群の判別分析においては, 正準判別変量が重要な役割を演じている. この変量は, 群間平方和積和行列 B および群内平方和積和行列を W とするとき. $W^{-1}B$ の固有ベクトルを用いて定義される. しかし, 全標本数が次元より少ないと W が特異になり利用できなくなる. この場合の１つの解決策として, Srivastava(2005) は W の逆行列の代わりにムーアペンロズ逆行列を用いることを提案している. この方法の漸近的性質や、２群の場合の誤判別確率の評価, などにも関心があが, これらについては Srivastava(2005) による結果がある.

参考文献

[1] Dempster, A.P. (1958). A high dimensional two sample significance test. *Ann. Math. Stat.,* **29**, 995-1010.

[2] Fujikoshi, Y., Himeno, T. and Wakaki, H. (2004). Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large compared to the sample size. *J. Japan Statist. Soc.,* **34**, 19-26.

[3] Srivastava, M.S. (2004). Multivariate theory for analyzing high dimensional data. *Tech Report #0401,* University of Toronto.

[4] Srivastava, M. S. (2005). Minimum distance classification rules for high dimensional data. To appear in *J. Multivariate Analysia.*

[5] Srivastava, M. S. and Fujikoshi, Y. (2005). Multivariate analysis of variance with fewer observations than the dimension. To appear in *J. Multivariate Analysia.*

U-統計量の分散推定量の漸近表現と二乗誤差

九州大学経済学研究院　前園宜彦

1．U-統計量の分散推定量

$X_1, \ldots, X_n$ を互いに独立で同じ分布 $F_\theta(x)$ に従う確率変数とし，$h(x_1, \cdots, x_r)$ を成分の入れ替えに対して不変な r 次の対称なカーネルとする．このとき U-統計量は

$$\hat{\theta} = U_n = \binom{n}{r}^{-1} \sum_{C_{n,r}} h(X_{i_1}, \cdots, X_{i_r}) \tag{1}$$

と定義される．ただし $\sum_{C_{n,r}}$ は $1 \leq i_1 < \cdots < i_r \leq n$ を表す．この U-統計量は標本平均や標本不偏分散などの統計量を含む重要な統計量のクラスである．本報告では U_n の分散 $V(U_n)$ の推定量の漸近表現を求め，それを元に漸近平均二乗誤差による比較を行った．

簡単のために $r = 2$ の場合を報告した．このとき分散は

$$n\sigma_n^2 = 4\xi_1^2 + \frac{1}{n-1}\xi_2^2$$

となる．$n\sigma_n^2$ のジャックナイフ分散推定量は

$$V_J = n\hat{\sigma}_J^2 = (n-1)\sum_{i=1}^{n}(U_n^{(i)} - U_n)^2$$

で与えられる．ここで $U_n^{(i)}$ は元のデータから X_i を除いて計算される U_n の値である．

また $4\xi_1^2$ の推定の観点から Sen(1960, Calucutta Statist. Assoc. Bull., 1977, Ann. Statist.) は次の推定量

$$V_S = \frac{4}{n-1}\sum_{i=1}^{n}(S_i - U_n)^2$$

を提案している．ただし

$$S_i = \frac{1}{n-1}\sum_{j=1, \neq i}^{n} h(X_i, X_j)$$

である．

一般にジャックナイフ分散推定量は正のバイアスを持つことが知られており，V_J はバイアスを持つ (Efron and Stein(1981, Ann. Statist.))．Hinkley(1978, B.K.) はジャックナイフ分散推定量のバイアス修正を一般の統計量について議論しており，U-統計量の場合は

$$Q_{i,j} = nU_n - (n-1)(U_n^{(i)} + U_n^{(j)}) + (n-2)U_n^{(i,j)}$$

とおくとき

$$V_C = V_J - \frac{1}{n+1}\sum_{C_{n,2}} Q_{i,j}^2$$

で与えられる．ただし $U_n^{(i,j)}$ は元のデータから X_i と X_j を除いて計算される U_n の値である．

Schucany and Bankson(1989, Austral. J. Statist.) は $n\sigma_n^2$ の不偏推定量 V_U を提案している.

ブートストラップ分散推定量（理論値）は，経験分布関数 $F_n(x)$ を使って

$$V_B = \int \Big[\frac{2}{n(n-1)} \sum_{i<j} h(x_i, x_j) - \int h(y_1, y_2) dF_n(y_1) dF_n(y_2)\Big]^2 \prod_{k=1}^{n} F_n(x_k)$$

である. Shirahata and Sakamoto(1992) はシミュレーションでブートストラップ分散推定量の良さを議論している.

これらの推定量はすべて $4\xi_1^2$ に概収束することも示されている.

2. 漸近表現と平均二乗誤差

2章で述べた分散推定量について H-分解を用いて漸近表現を求め，漸近二乗誤差を求めた. 次の記号を定義する.

$$V_n = \frac{4}{n}\sum_{i=1}^{n} f_1(X_i) + \frac{8}{n^2}\sum_{C_{n,2}} f_2(X_i, X_j) + \frac{8}{n^3}\sum_{C_{n,3}} f_3(X_i, X_j, X_k).$$

このとき次の定理が成り立つ.

[定理 1]. もしある $\varepsilon > 0$ に対して $E|h(X_1, X_2)|^{4+\epsilon} < \infty$ ならば

$$V_J = V_n + \frac{8}{n^2}\sum_{i=1}^{n} \delta(X_i) + n\sigma_n^2 + \frac{b_J}{n} + R_{1;n},$$

$$V_S = V_n + \frac{8}{n^2}\sum_{i=1}^{n} \{\delta(X_i) - f_1(X_i)\} + n\sigma_n^2 + \frac{b_S}{n} + R_{2;n},$$

$$V_C = V_n + \frac{4}{n^2}\sum_{i=1}^{n} \delta(X_i) + n\sigma_n^2 + R_{3;n},$$

$$V_U = V_n + \frac{4}{n^2}\sum_{i=1}^{n} \delta(X_i) + n\sigma_n^2 + R_{4;n},$$

$$V_B = V_n + \frac{1}{n^2}\sum_{i=1}^{n} f_0(X_i) + n\sigma_n^2 + \frac{b_B}{n} + R_{5;n}$$

が成り立つ.

この漸近表現を使うと，漸近平均二乗誤差を次のように求めることができる.

$$\begin{aligned}
mse(V_J) &= \frac{16}{n}E[f_1^2(X_1)] + \frac{1}{n^2}\{b_J^2 + 64E[f_1(X_1)\delta(X_1)] + 32E[f_2^2(X_1, X_2)]\},\\
mse(V_S) &= \frac{16}{n}E[f_1^2(X_1)] + \frac{1}{n^2}\{b_S^2 + 64E[f_1(X_1)(\delta(X_1) - f_1(X_1))]\\
&\qquad +32E[f_2^2(X_1, X_2)]\},\\
mse(V_C) &= \frac{16}{n}E[f_1^2(X_1)] + \frac{1}{n^2}\{32E[f_1(X_1)\delta(X_1)] + 32E[f_2^2(X_1, X_2)]\},\\
mse(V_B) &= \frac{16}{n}E[f_1^2(X_1)] + \frac{1}{n^2}\{b_B^2 + 96E[f_1(X_1)\delta(X_1)] - 160E[f_1^2(X_1)]\\
&\qquad +8E[f_1(X_1)\tilde{\delta}(X_1)] + 32E[f_2^2(X_1, X_2)]\}.
\end{aligned}$$

これらを利用して，n^{-2} の項までの平均二乗誤差を議論することができるようになった.

Selecting models with different spectral density matrix structures by cross validated log likeliood criterion

東北大学大学院経済学研究科　松田安昌
東京大学大学院経済学研究科　矢島美寛
London School of Economics Howell Tong

1. 序

多変量定常時系列におけるノンパラメトリックあるいはセミパラメトリックなモデルはスペクトル密度行列の制約として表現できる場合が多い. 本論文ではモデル選択規準のひとつとしてよく知られているクロスバリデーション法を, 複数の制約からひとつの制約を選ぶ方法として利用した場合の理論的性質について考察する.

2. モデルの定式化

$\{X_{t,a}, a = 1, \ldots, r\}$ を $r-$次元定常過程, $f(\lambda,\ -\pi \leq \lambda \leq \pi)$ をそのスペクトル密度行列とする. $G(\theta, y) = (G_{ab}(\theta, y))$ を $r \times r$行列, θはパラメータベクトルとする. $\mathcal{F}$をスペクトル密度行列の全体すなわち各周波数λで非負定値エルミート行列となる行列値関数の全体とする. そのなかでGによって導入される制約をみたすスペクトル密度関数の全体を

$$\mathcal{F}_G = \{g(\lambda) \in \mathcal{F} | \exists\theta, g(\lambda) = G(\theta, g(\lambda))\} \tag{1}$$

とおく.

3. CVLL 規準

$\{\boldsymbol{X}_t = (X_{t,a}, a = 1, \ldots, r)', T = 1, 2, \ldots, n\}$ を観測値としてそのフーリエ変換, ピリオドグラムを

$$W(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \boldsymbol{X}_t \exp(-i\lambda t),\ I(\lambda) = W(\lambda)\overline{W}(\lambda)'$$

とおく. 以下ではλとしてフーリエ周波数$\lambda_j = (2\pi j)/n, j = -[(n-1)/2], \ldots, -1, 0, 1, \ldots, [n/2]$ を考える.

定義 1(CVLL 規準)

$$\mathrm{CVLL}(G) = \sum_{j=1}^{[n/2]} \log\det(\hat{g}_{-j}(\lambda_j) + \mathrm{tr}(I(\lambda_j)\hat{g}_j^{-1}(\lambda_j)))$$

ここで

$$
\begin{aligned}
\hat{g}_{-j}(\lambda_j) &= G(\hat{\theta}, \hat{f}_{-j}(\lambda_j)) \\
\hat{f}_{-j}(\lambda_j) &= (\sum_{k=-m/2}^{m/2} w_k)^{-1} \sum_{k=-m/2, k\neq 0}^{m/2} w_k I(\lambda_{j+k})
\end{aligned}
$$

とする．また$\hat{\theta}$はθの推定量である．

4. CVLLの漸近的性質

$f(\lambda)$を真のスペクトル密度行列として，$\mathcal{F}_G$を2つのカテゴリー，Category I:$\{G|f(\lambda) \notin \mathcal{F}_G\}$ Category II:$\{G|f(\lambda) \in \mathcal{F}_G\}$ に分類する．

仮定

A1. $\{\boldsymbol{X}_t\}$は正規定常過程

A2. $f(\lambda)$, $-\pi \leq \lambda < \pi$は正定値行列

A3. $f(\lambda)$, $-\pi < \lambda < \pi$は2回連続微分可能

A4. $m = O(n^{\beta})$, $1/2 < \beta < 3/4$ また$u(x)$は$[-1/2, 1/2]$上で正値をとる連続関数で$w_k = u(k/m)$, $k = -m/2, \ldots, -1, 0, 1, \ldots, m/2$とする．

A5. GがCategory IIに属し，$f(\lambda) = G(\theta_0, f(\lambda))$が成立するとき$\hat{\theta} - \theta_0 = O_p(n^{-1/2})$をみたす．

このとき以下の定理が成立する．

定理

A1-A5の仮定の下で， $n \to \infty$のときCVLLがCategory Iに属するモデルを選ぶ確率は0へ収束する．Category IIのモデルの中でAMISE(G)を最小にするモデルを選ぶ確率が1に収束する．ここで

$$\text{AMISE}(G) = p - \lim_{n\to\infty} \frac{m}{n} \sum j = 1^{[n/2]} \text{tr}\{\hat{g}_{-j}(\lambda_j) - f(\lambda_j)\hat{f}^{-1}(\lambda_j)\}.$$

とする．

5. 例

(1)のような定式化が可能なモデルの例として，成分時系列の独立性，セパラブル・モデル，時間可逆性，グラフィカルモデリングなどがある．またこれらのモデルではAMISE(G)を最小にするモデルが，Category IIに属するモデルの中で最も冗長度の小さいコンパクトなモデルである．すなわちCVLLによるモデル選択は一致性をもつ．ただしCVLLの一致性が任意のモデル選択問題について成立するか否かはまだ未解決であり，それを明らかにすることが今後の課題である．

年金保険と派生証券の理論

2005 年 12 月

国友直人
東京大学経済学研究科

室井芳史
日本銀行金融研究所

1. 英国 Equitable Life 社の例

日本の保険業界においても、近年になりようやく死亡率リスクに直接的には関連しない資産価値リスクや金利リスクなどの金融リスクに連動する年金保険契約などが研究・開発・販売されるようになっている。近代的な生命保険の発祥の地として知られている英国では 1970 年代〜1980 年代に生命保険契約者が「将来において契約したファンドを年金に変換できる権利を保障する」という年金保険契約の保証を伴う生命保険がかなり販売されていた。この種の生命保険契約は年金保険に関する一種の派生証券と見なすことができる。

ところが 1990 年代になり英国の金利水準が下降する中で、この種の年金保険の権利行使が保険契約者にとってきわめて有利となり、結果として英国を代表する世界最古の保険会社 Equitable Life は 1990 年代に経営危機に陥った (例えば Boyle and Hardy (2003) を参照)。

2. 年金保険の派生証券

ここで言及した英国で販売されていた年金保険の派生証券契約とは、例えば派生証券契約の部分のみを取り出せば、将来時刻 T 以降に年金支払い（ペイオフ）が

$$S(T)\max\left[(\frac{a_{65}(T)}{g}-1),0\right] \tag{1}$$

で与えられる派生証券である。ここで $S(T)$ は将来時刻 T における保険ファンドの価値、$a_{65}(T)$ は 65 歳（年金開始時点）における市場年金価値、g は実定数である。ここで例えばファンドとして株式に投資する場合には、将来の年金の市場価値は、将来の株価と金利水準、さらに生存確率にも依存する一種のプット・オプション契約と見なせよう。

3. 漸近展開法の応用

ここで様々に考え得る年金保険の派生証券契約についての一般的評価理論を考察する。特に漸近展開法 (the asymptotic expansion method) と呼ばれている方法に注目するが、これは確率解析学・数理統計学でよく知られている漸近展開を連続確率過程モデルに拡張し、Yoshida (1992)、Kunitomo=Takahashi (2001, 2003a, 2003b)、Kunitomo=Kim (2001) などがファイナンス分野へ応用した方法である。本報告では年金保険の派生証券の分野についても漸近展開法が有効に適用できるか否かを含め、派生証券を巡る様々な問題を議論する。連続確率過程の漸近展開法は Yoshida (1992) が示しているように数学的には Malliavin 解析の応用である。特に漸近展開を正確に正当化するには Kunitomo=Takahashi (2003) で議論したような数学的な展開が必要であることを指摘する。

4. 参考文献

Boyle, P. and M. Hardy (2003), "Guaranteed Annuity Options," *ASTIN Bulletin*, 33(2), 125-152.

Kunitomo, N. and Y. Kim (2001),"Effects of Stochastic Interest Rates and Volatility on Contingent Claims," Discussion Paper CIRJE-F-129, Graduate School of Economics, forthcoming in *Japanese Economics Review*, Blackwell.

Kunitomo, N. and A. Takahashi (2001),"The Asymptotic Expansion Approach to The Valuation of Interest Contingent Claims," Mathematical Finance, Vol.11-1, 117-151.

Kunitomo, N. and A. Takahashi (2003a),"On Validity of the Asymptotic Excpansion Approach in Contingent Claims," Annals of Applied Probability, Vol.13-3, 914-952.

国友直人・高橋明彦 (2003b), 「数理ファイナンスの基礎：マリアバン解析と漸近展開の応用」, 東洋経済新報社。

Yoshida, N. (1992),"Asymptotic Expansions for Statistics Related to Small Diffusions," Journal of Japan Statistocal Society, Vol.22, 139-159.

確率微分方程式の統計推測

内田 雅之　　九州大学大学院数理学研究院

1 SDEモデル

次の確率微分方程式 (SDE) によって定義される 1 次元拡散過程 X を考える.

$$dX_t = b(X_t, \alpha)dt + \sigma(X_t, \beta)dw_t, \quad t \in [0, T], \quad X_0 = x_0, \tag{1}$$

ここで, $b : \mathbf{R} \times \Theta_\alpha \to \mathbf{R}$ (ドリフト), $\sigma : \mathbf{R} \times \Theta_\beta \to \mathbf{R}$ (拡散係数), w は 1 次元標準 Wiener 過程, Θ_α と Θ_β は, それぞれ $\mathbf{R}^p$ と $\mathbf{R}^q$ のコンパクト部分集合. $\theta := (\alpha, \beta) \in \Theta := \Theta_\alpha \times \Theta_\beta$. $C_\uparrow^{k,3}(\mathbf{R} \times \Theta)$ を以下の 2 つの条件を満たす関数 f の空間とする: (i) $f(x, \theta)$ は $\mathbf{R} \times \Theta$ 上で定義された $\mathbf{R}$-値関数で, x について k 回連続微分可能で, すべての $p = 0, 1, \ldots, k$ に対して,$\partial_x^p f(x, \theta)$ は θ について 3 回連続微分可能. (ii) すべての $p = 0, 1, \ldots, k$, $|\nu| = 0, 1, 2, 3$, さらに, すべての x に対して $\sup_{\theta\in\Theta} |\delta^\nu \partial_x^p f| \le C(1 + |x|)^C$ となる $C > 0$ が存在する. ここで, $\nu = (\nu_1, \cdots, \nu_l)$ は multi-index であり, $l = \dim(\Theta)$, $|\nu| = \nu_1 + \cdots + \nu_l$, $\delta^\nu = \delta_1^{\nu_1} \cdots \delta_l^{\nu_l}$, $\delta_j = \partial/\partial\theta^j$, $j = 1, \cdots, l$ とする. ν と δ はパラメータ空間 Θ に依存していることに注意. つまり, Θ_α に対しては, $\nu = (\nu_1, \cdots, \nu_p)$, $\delta_j = \partial/\partial\alpha^j$ であり, Θ_β に対しては, $\nu = (\nu_1, \cdots, \nu_q)$, $\delta_j = \partial/\partial\beta^j$ である. $(p+q) \times (p+q)$ の行列 $D_n = (D_n^{ij})_{i,j=1,\ldots,(p+q)}$ を $D_n^{ii} = 1/\sqrt{nh_n}$ (for $i = 1, \ldots, p$), $D_n^{ii} = 1/\sqrt{n}$ (for $i = p+1, \ldots, p+q$), $D_n^{ij} = 0$ (others) とする. さらに, $(p+q) \times (p+q)$ の (Fisher 情報) 行列を

$$I(\theta_0) = \begin{pmatrix} (I_b^{ij}(\theta_0))_{i,j=1,\ldots,p} & 0 \\ 0 & (I_\sigma^{ij}(\theta_0))_{i,j=1,\ldots,q} \end{pmatrix},$$

ただし, $I_b^{ij}(\theta_0) = \int_\mathbf{R} \frac{\delta_{\alpha_i} b(x,\alpha_0) \delta_{\alpha_j} b(x,\alpha_0)}{\sigma^2(x,\beta_0)} \mu_{\theta_0}(dx)$, $I_\sigma^{ij}(\theta_0) = \frac{1}{2} \int_\mathbf{R} \frac{\delta_{\beta_i} \sigma(x,\beta_0) \delta_{\beta_j} \sigma(x,\beta_0)}{\sigma^2(x,\beta_0)} \mu_{\theta_0}(dx)$.

Assumption 1 *(i)* $[0, T]$ 上で方程式 *(1)* の強い解が一意に存在する. *(ii)* すべての x と β に対して, ある定数 $c > 0$ が存在して, $c \le \sigma^2(x, \beta) \le 1/c$. *(iii)* すべての θ に対して, 拡散過程 X はエルゴード的である. その不変測度を μ_θ とする. すべての $p \ge 0$ に対して, $\int_\mathbf{R} |x|^p \mu_\theta(dx) < \infty$. *(iv)* すべての $p \ge 0$ に対して, $\sup_t E[|X_t|^p] < \infty$. *(v)* $b(x, \alpha) \in C_\uparrow^{2,3}(\mathbf{R} \times \Theta_\alpha)$, $\sigma(x, \beta) \in C_\uparrow^{2,3}(\mathbf{R} \times \Theta_\beta)$. *(vi)* $I(\theta_0)$ は正則である. *(vii)* すべての x に対して, $b(x, \alpha) = b(x, \alpha_0) \Rightarrow \alpha = \alpha_0$. すべての x に対して, $\sigma(x, \beta) = \sigma(x, \beta_0) \Rightarrow \beta = \beta_0$.

本講演では離散観測データ $\mathbf{X}_n = (X_{t_0}, X_{t_1}, X_{t_2}, \ldots, X_{t_n})$, ここで $t_k = kh_n$, $nh_n = T$, に基づいた統計推測について, (i) パラメータ推定, (ii) モデル選択に話題を限定して, 解説を行った. 本報告では asymptotics を $h_n \to 0$, $T = nh_n \to \infty$ and $nh_n^2 \to 0$ as $n \to \infty$ とする.

2 パラメータ推定

$\mathbf{X}_n$ の近似尤度関数 (局所正規近似) として, 次のコントラスト関数 g_n を考える.

$$g_n(\mathbf{X}_n, \theta) = \sum_{k=1}^n g(X_{t_{k-1}}, X_{t_k}, \theta), \tag{2}$$

$$g(x, y, \theta) = -\frac{1}{2}\log(2\pi h_n) - \log\sigma(x, \beta) - \frac{(y - x - h_n b(x, \alpha))^2}{2h_n \sigma^2(x, \beta)}. \tag{3}$$

最大コントラスト推定量を $\hat{\theta}_n = \arg\sup_\theta g_n(\mathbf{X}_n, \theta)$ で定義する．この時，最大コントラスト推定量 $\hat{\theta}_n$ にして以下が成り立つ (Yoshida (1992), Kessler (1997), Yoshida (2005)).

Theorem 1 *Assumption 1* の下, $nh_n^2 \to 0$ の時,

$$D_n^{-1}(\hat{\theta}_n - \theta_0) \to^d N(0, I^{-1}(\theta_0)).$$

Remark 1 一般に g_n は次の意味で対数尤度関数 l_n の近似になっていない *(Uchida (2005))*.

3 モデル選択

真のモデルを

$$dX_t = B(X_t)dt + S(X_t)dw_t, \quad t \in [0, T], \qquad X_0 = x_0$$

として, M 個の統計モデル: $i = 1, 2, \ldots, M$ に対して,

$$dX_t = b_i(X_t, \alpha_i)dt + \sigma_i(X_t, \beta_i)dw_t, \quad t \in [0, T], \qquad X_0 = x_0$$

を考える．ただし, 統計モデルは真のモデルを含むと仮定する (specified parametric model). すなわち, $i = 1, \ldots, M$ に対して, 真値 $\theta_{i,0} = (\alpha_{i,0}, \beta_{i,0})$ が存在して,

$$b_i(x, \alpha_{i,0}) = B(x), \quad \sigma_i(x, \beta_{i,0}) = S(x).$$

データ $\mathbf{X}_n$ に対して, そのコピーを $\mathbf{Z}_n$ とし, 統計モデル i における対数尤度関数と局所正規近似に基づくコントラスト関数をそれぞれ, l_i, g_i とする．g_i から得られる最大コントラスト推定量を $\hat{\theta}_i$ とし, $\alpha_{i,0} \in \Theta_{\alpha_i} \subset \mathbf{R}^{p_i}$, $\beta_{i,0} \in \Theta_{\beta_i} \subset \mathbf{R}^{q_i}$, $\Theta_i = \Theta_{\alpha_i} \times \Theta_{\beta_i}$ とする．

$$\begin{aligned}
\mathrm{DAIC}(i,j) &:= -2\left[g_i(\mathbf{X}_n, \hat{\theta}_i(\mathbf{X}_n)) - g_j(\mathbf{X}_n, \hat{\theta}_j(\mathbf{X}_n)) - \dim(\Theta_i) + \dim(\Theta_j)\right], \\
\mathrm{DKL}(i,j) &:= -2\left[E_{\mathbf{Z}_n}[l_i(\mathbf{Z}_n, \hat{\theta}_i(\mathbf{X}_n))] - E_{\mathbf{Z}_n}[l_j(\mathbf{Z}_n, \hat{\theta}_j(\mathbf{X}_n))]\right].
\end{aligned}$$

と定義する．各モデル i について Assumption 1 が成り立ち, さらに正則条件の下で, $nh_n^2 \to 0$ の時,

$$E_{\mathbf{X}_n}[\mathrm{DAIC}(i,j) - \mathrm{DKL}(i,j)] = o(1).$$

$\mathrm{DAIC}(i,j)$ は赤池 (Akaike (1993, 1994)) によって提案された AIC の差に対応する．

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium in Information Theory*, Petrov, B.N. & Csaki, F. eds., Akademiai Kaido, Budapest, 267-281.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Auto. Control* **AC-19**, 716-723.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211–229.

Uchida, M. (2005). A note on AIC for ergodic diffusion processes from discrete observations. (submitted).

Yoshida, N. (1992). Asymptotic expansion for statistics related to small diffusions. *J. Japan Statist. Soc.* **22**, 139-159.

Yoshida, N. (2005). Polynomial type large deviation inequality and its applications. preprint.