(18)「確率過程と統計的漸近理論」に関する研究報告

Semiparametric Estimation in Copula Models

Hideatsu Tsukahara
Department of Economics, Seijo University

Introduction. Let $X^k = (X_1^k, \ldots, X_d^k)$, $k = 1, \ldots, n$ be independently and identically distributed random vectors with d-dimensional continuous distribution function $F = C_\theta(F_1, \ldots, F_d)$, where C_θ is a copula and $\theta = (\theta_1, \ldots, \theta_m)$ is an m-dimensional parameter. We consider two estimators based on the empirical copula: rank approximate M-estimator and minimum distance estimator. The large sample properties of these estimators can be derived from asymptotic results on the empirical copula.

Asymptotics for the empirical copula. Let $\xi^k = (\xi_1^k, \ldots, \xi_d^k)$, $k = 1, \ldots, n$ be independently and identically distributed random vectors with distribution function C which is a copula. Set

$$\mathbb{G}_n(u) \triangleq \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{\xi_1^k \le u_1, \ldots, \xi_d^k \le u_d\}}, \quad \mathbb{G}_{ni}(u_i) \triangleq \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{\xi_i^k \le u_i\}}.$$

Then the *empirical copula* is defined by

$$\mathbb{C}_n(u) \triangleq \mathbb{G}_n(\mathbb{G}_{n1}^{-1}(u_1), \ldots, \mathbb{G}_{nd}^{-1}(u_d)).$$

The empirical process $\mathbb{U}_n^C$ and the empirical copula process $\mathbb{D}_n^C$, both associated with C, are defined by

$$\mathbb{U}_n^C(u) \triangleq \sqrt{n}\big(\mathbb{G}_n(u) - C(u)\big), \quad \mathbb{D}_n^C(u) \triangleq \sqrt{n}\big(\mathbb{C}_n(u) - C(u)\big).$$

We then have the following asymptotic representation:

$$\mathbb{D}_n^C(u) = \mathbb{U}_n^C(u) - \sum_{i=1}^{d} C^i(u)\mathbb{U}_n^C(\mathbf{1}, u_i, \mathbf{1}) + R_n(u),$$

where $C^i(u) = \frac{\partial C(u)}{\partial u_i}$, $i = 1, \ldots, d$ and $\sup_u |R_n(u)| = o_P(1)$ as $n \to \infty$.

We can also prove the following maximal inequality: There exist positive constants K_1 and K_2, not depending on C, such that

$$\mathrm{P}(\sup_u |\mathbb{D}_n^C(u)| > r) \le K_1 e^{-K_2 r^2}, \quad r > 0,$$

for all $n = 1, 2, \ldots$.

Also, under certain regularity conditions on ϕ, it holds that

$$\int \phi \, d\mathbb{D}_n^C \xrightarrow{\mathrm{P}} \int \phi \, d\mathbb{U}^C + \sum_{i=1}^{d} \int \phi^i \, \mathbb{U}^C(\mathbf{1}, u_i, \mathbf{1}) \, dC$$

where $\phi^i = \frac{\partial \phi}{\partial u_i}$, $i = 1, \ldots, d$, and $\mathbb{U}^C$ is a C-pinned Brownian sheet. The random variable on the right-hand side has normal distribution with mean 0 and variance

$$\mathrm{var}\left(\phi(\xi) + \sum_{i=1}^{d} \int \phi^i(u) \, \mathbf{1}_{\{\xi_i \le u_i\}} \, dC(u)\right)$$

where $\xi \sim C$.

Rank Approximate M-estimate. Let $X^k = (X_1^k, \ldots, X_d^k)$, $k = 1, \ldots, n$ be a random sample on $F_\theta = C_\theta(F_1, \ldots, F_d)$. The parameter space is $\Theta \subset \mathbb{R}^m$. In this section we suppose that the copula C_θ has a density c_θ and c_θ is differentiable with respect to θ, and we write $\dot{c}_\theta = (\frac{\partial c_\theta}{\partial \theta_1}, \ldots, \frac{\partial c_\theta}{\partial \theta_m})'$.

Now suppose $\phi_\theta\colon \mathbb{R}^d \to \mathbb{R}^m$ is a function satisfying $\int \phi_\theta(u)\, dC_\theta(u) = 0$. Then any solution $\widehat{\theta}_n$ to the equation

$$\sum_{k=1}^{n} \phi_\theta\big(\mathbb{F}_{n1}(X_1^k), \ldots, \mathbb{F}_{nd}(X_d^k)\big) = 0$$

is called a *Rank Approximate M-estimator.*

Let us denote the left-hand side of the above equation by $S_n(\theta)$, then we have $S_n(\theta) = \int \phi_\theta(u)\, d\mathbb{C}_n(u)$. Assume that $\dot{\phi}_\theta(u) = \frac{\partial \phi_\theta}{\partial \theta}(u)$ $(m \times m)$ exists and satisfies certain regularity conditions. Then one can prove that, as $n \to \infty$, $\sqrt{n}\big(\widehat{\theta}_n - \theta\big)$ converges in distribution to a normal distribution with mean 0 and covariance matrix $A_\theta^{-1}\Sigma_\theta A_\theta^{-1}$, where, letting $\xi \sim C_\theta$,

$$A_\theta = \int \dot{\phi}_\theta\, dC_\theta, \quad \Sigma_\theta = \operatorname{var}\Big(\phi_\theta(\xi) + \sum_{i=1}^{d} \int \phi_\theta^i(u)(\mathbf{1}_{\{\xi_i \le u_i\}} - u_i)\, dC_\theta(u)\Big),$$

and $\phi_\theta^i(u) = \frac{\partial \phi_\theta(u)}{\partial u_i}$, $i = 1, \ldots, d$.

Minimum Distance Estimator. Suppose that the copula associated with the distribution function of $X^k = (X_1^k, \ldots, X_d^k)$ is in fact D. On the other hand, we have a given parametric family of copulas $\{C_\theta\colon \theta \in \Theta\}$ to fit the data. Let us define the *MD functional* T on the space of copulas by

$$T(D) \triangleq \arg\min_\theta \rho(D, C_\theta).$$

Here ρ is a distance between probabilities on $[0,1]^d$. In the present paper we consider the Cramér-von Mises distance:

$$\rho(C, D) = \int_{[0,1]^d} [C(u) - D(u)]^2\, du$$

As before, let $\mathbb{C}_n$ be the empirical copula based on the sample $X^1, \ldots, X^n$. Define $\widehat{\theta}_n = T(\mathbb{C}_n)$ and call it an *MD estimator.*

We can prove differentiability of T at C_θ under certain regularity conditions: i.e., $\forall \theta \in \Theta^\circ$ and $\forall D$ in a ρ-neighborhood of C_θ,

$$T(D) = \theta + \int \gamma_\theta(D - C_\theta)\, du + o(\rho(D, C_\theta))$$

where

$$\gamma_\theta(u) \triangleq \left[\int \delta_\theta \delta_\theta'\, du\right]^{-1} \delta_\theta(u)$$

Thanks to this differentiablility and the maximal inequality for the empirical copula, we can prove two results. One is a robustness property MD estimator: T_n converges locally uniformly to the estimand $T(D)$ when the true D is close to C_θ. The other one is asymptotic normality of MD estimator in locally uniform sense: under any sequence D_n in a $\sqrt{n}$-shrinking neighborhood of C_θ,

$$\sqrt{n}(T_n - T(D_n)) \xrightarrow{\mathcal{L}} N(0, \Gamma_\theta) \qquad \text{as } n \to \infty,$$

where

$$\Gamma_\theta = \operatorname{var}\left(\int \gamma_\theta(u)\Big[\mathbf{1}_{\{\xi \le u\}} - \sum_{i=1}^{d} C^i(u)\mathbf{1}_{\{\xi_i \le u_i\}}\Big]\, du\right)$$

Monte Carlo Method for pricing of Bermuda type derivatives

Shigeo Kusuoka
Graduate School of Mathematical Sciences, University of Tokyo

1 Introduction

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,\infty)}, P)$ be a filtered space with the usual condition, and $\{B_t\}_{t\in[0,\infty)}$ be a d-dimensinal Brownian motion. Let $T > 0$, and let $\sigma : [0,T] \times \mathbf{R}^D \to \mathbf{R}^D \times \mathbf{R}^d$ and $b : [0,T] \times \mathbf{R}^D \to \mathbf{R}^D$ be continuous functions. For each $s \in [0,T]$ and $x \in \mathbf{R}^D$, let $X(t;s,x)$, $t \in [s,T]$ be a solution of the following SDE.

$$X(t;s,x) = x + \int_s^t \sigma(r, X(r;s,x))dB_r + \int_s^t b(r, X(r;t,x))dr, \qquad t \in [s,T]. \tag{1}$$

We assume that the above SDE (1) has a path-wise unique solution for every $(s,x) \in [0,T] \times \mathbf{R}^D$.

Let $g : [0,T] \times \mathbf{R}^D \to \mathbf{R}$ be a continuous function with suitable conditions. Let $N \geq 2$ and let T_n, $n = 0, 1, \dots, N$, be positive numbers such that $0 = T_0 < T_1 < \dots < T_N = T$. Let $\mathcal{S}_n$, $n = 0, 1, \dots, N$, be the set of $\mathcal{F}_t$-stopping times taking value in $\{T_n, T_{n+1}, \dots, T_N\}$. Concerning the pricing of Bermuda type derivatives, we are interested in computing the following value functions.

$$v_n(x) = \sup\{E[g(\tau, X(\tau;s,x))]; \tau \in \mathcal{S}_n\}, \qquad n = 0, 1, \dots, N.$$

Let us define a probability measure $p_n(x,\cdot)$ over $\mathbf{R}^D$ for each $n = 0, 1, \dots, N-1$, and $x \in \mathbf{R}^D$ by

$$p_n(x,A) = P(X(T_{n+1};T_n,x) \in A), \qquad \text{for a Borel set } A \text{ in } \mathbf{R}^D,$$

and define a operator P_n, $n = 0, 1, \dots, N-1$, by

$$P_n f(x) = \int_{\mathbf{R}^D} f(y)p_n(x,dy) = E[f(X(T_{n+1};T_n,x))]$$

for a measurable function f on $\mathbf{R}^D$. Then v_n, $n = N, N-1, \dots, 0$, are given inductively by the following.

$$v_N(x) = g(T_N, x),$$
$$v_{n-1}(x) = (P_{n-1}v_n)(x) \vee g(T_{n-1}, x).$$

So the value function $v_0(x)$ is easily given mathematically. However, if D is not small, it is not easy to memorize a function on $\mathbf{R}^D$, and so it is not easy to compute $v_0(x)$.

Several people suggest a Monte-Carlo method to compute the value function. In this paper, we discuss and justify the method given by [1]. We assume the following assumption (A).

(A) D_n, $n = 0, 1, \dots, N-1$, are measurable sets in $\mathbf{R}^N$ such that $(P_n v_{n+1})(x) \geq g(T_n, x)$ for any $x \in \mathbf{R}^D \setminus D_n$.

Remark 1 (1) $D_n = \mathbf{R}^D$ *satisfies the assumption* (A).
(2) *If* $g(t,x) \geq 0$, *for any* $(t,x) \in [0,T] \times \mathbf{R}$, *then* $D_n = \{x \in \mathbf{R}^D; g(T_n, x) > 0\}$ *satisfies the assumption* (A).

Now let $L_n \geq 1$, $n = 0, 1, \dots, N-1$, and $\vec{X}_{n,\ell} = \{X_{n,\ell}(m)\}_{m=0}^N$, $\ell = 1, \dots, L_n$, $n = 0, 1, \dots, N-1$, are identically independent random vectors whose distribution is the same as the distribution of $\{X(T_m; 0, x)\}_{m=0}^N$. Let $K_n \geq 1$, $n = 0, 1, \dots, N-1$, and $\psi_{n,k}$, $k = 1, \dots, K_n$, $n = 0, 1, \dots, N-1$, are functions on $\mathbf{R}^D$. Then we define functions H_n, $n = N, N-1, \dots, 1, 0$, on $\mathbf{R}^D$ inductively by the following.

$$H_N(x) = 1.$$

When $\vec{H}_{n+1} = \{H_m\}_{m=n+1}^N$, are given we let

$$\sigma_{n,\ell} = \min\{m \geq n+1; H_m(X_{n,\ell}(m)) > 0\}, \qquad \ell = 1, \dots L_n.$$

Then we let $\{\tilde{a}_{n,k}\}_{k=1}^{K_n}$ be the minimizing point of the function

$$F_n(\{a_k\}_{k=1}^{K_n}) = \frac{1}{L_n} \sum_{\ell=1}^{L_n} |g(T_{\sigma_{n,\ell}}, X_{n,\ell}(\sigma_{n,\ell})) - \sum_{k=1}^{K_n} a_n \psi_{n,k}(X_{n,\ell}(n))|^2 1_{D_n}(X_{n,\ell}(n)).$$

Finally we define H_n by

$$H_n(x) = \begin{cases} g(T_n, x) - \sum_{k=1}^{K_n} \tilde{a}_{n,k} \psi_{n,k}(x), & x \in D_n \\ -1, & x \in \mathbf{R}^D \setminus D_n. \end{cases}$$

Then we let

$$\tilde{v}_0 = \frac{1}{L_0} \sum_{\ell=1}^{L_0} g(\sigma_{0,\ell}, X_{0,\ell}(\sigma_{0,\ell})).$$

We show that

$$\lim_{K_n} \lim_{L_n \to \infty} E[|v_0(x) - \tilde{v}_0|^2 \wedge 1] = 0.$$

References

[1] Longstaff, F., and E. Schwartz, Valueing American Options by Simulation: A simple Least-Squares Approach, The Review of Financial Studies, 14(2001), 113-147

Diffusions with measurement errors.

A. Gloter * & J. Jacod †

Let X be a 1-dimensional diffusion process on the time interval $[0,1]$, of the form

$$dX_t = b_t dt + \sqrt{c(\theta,t,X_t)}dW_t, \qquad \mathcal{L}(X_0) = \eta. \tag{1}$$

Here W is a standard Brownian motion, and η is an arbitrary initial law on $\mathbb{R}$, and b is a (non-anticipative) drift term which may depend on the path of X or W. The nonnegative function c depends on a parameter θ which, for simplicity, is assumed to belong to some (finite or infinite) open interval Θ of $\mathbb{R}$. We observe this process at times i/n for $i = 0, 1, \ldots, n$, but each observation is blurred by an error which is centered normal with variance ρ_n; in other words we observe the variables

$$Y_i^n = X_{i/n} + \sqrt{\rho_n}U_i, \qquad i = 0, \ldots, n, \tag{2}$$

where the U_i's are i.i.d. $\mathcal{N}(0,1)$ and independent of the process. Our aim is to estimate the parameter θ, knowing the noise level ρ_n, and on the basis of the observations (2).

This problem does not seem to have been studied so far, with the exception of a recent paper by Malyutov and Bayborodin where no attempt towards optimality is made. The model is a hidden Markov model for which a lot is known, but our situation differs in that usually hidden Markov chain are homogeneous ergodic and time goes to infinity, while here the sequence $X_{i/n}, i = 0, 1, \ldots$, has a transition kernel depending on n which degenerates as $n \to \infty$, while no ergodic property is relevant here.

Our results go as follows. The "optimal" rate of convergence u_n depends on the behaviour of the sequence $n\rho_n$: up to taking subsequences it is no real restriction to assume that $n\rho_n$ converges in $[0,\infty]$ and, ruling out the totally uninteresting case where the sequence ρ_n itself is unbounded, three cases can occur:

$$\left.\begin{array}{lll} \textbf{Case 1} & n\rho_n \to u = 0 : & \text{take} \quad u_n = 1/\sqrt{n} \\ \textbf{Case 2} & n\rho_n \to u \in (0,\infty) : & \text{take} \quad u_n = 1/\sqrt{n} \\ \textbf{Case 3} & n\rho_n \to u = \infty, \quad \sup_n \rho_n < \infty : & \text{take} \quad u_n = (\rho_n/n)^{1/4}. \end{array}\right\} \tag{3}$$

We also set

$$\varphi_u(x,y) = \begin{cases} \frac{y^2}{2x^2} & \text{if } u = 0 \\ \frac{y^2(2+x/u)}{2\sqrt{u}x^{3/2}(4+x/u)^{3/2}} & \text{if } 0 < u < \infty \\ \frac{y^2}{8x^{3/2}} & \text{if } u = \infty. \end{cases} \tag{4}$$

*G.R.A.P.E., Université Montesquieu, av. Léon Duguit, 33608 Pessac, France (CNRS UMR 5113)

†Laboratoire de Probabilités et Modèles aléatoires, Université Paris 6, 4 place Jussieu, 75252 Paris, France (CNRS UMR 7599)

The first result concerns the case $b_t \equiv 0$ and $c(\theta,t,x) = c(\theta,t)$, that is when $X_t = \int_0^t \sqrt{c(\theta,s)}\, dW_s$. Our assumptions are, for a given θ:

Hypothesis ($\mathbf{H1_\theta}$) : The function $\zeta \mapsto c(\zeta,t)$ is twice differentiable with partial first and second derivatives denoted by $\dot c$ and $\ddot c$, and c, $\dot c$ and $\ddot c$ are continuous on $\Theta \times [0,1]$. Further the function $c(\theta,.)$ does not vanish, and the function $\dot c(\theta,.)$ is not identically 0.

Hypothesis ($\mathbf{H2_\theta}$) : The function $\dot c(\theta,.)$ does not vanish.

Hypothesis ($\mathbf{H3_\theta}$) : The set $F = \{s \in [0,1] : \dot c(s,\theta) = 0\}$ is the union of its connected components with positive length, plus a Borel set with Lebesgue measure equal to 0. Moreover the function $\dot c(\theta,.)$ is Hölder-continuous with some index $\alpha \in (0,1]$.

Theorem 1. *Assume ($H1_\theta$). We have the LAN property at point θ and with the rates u_n given above and the Fisher information*

$$I(\theta) = \int_0^1 \varphi_u(c(\theta,s), \dot c(\theta,s))ds, \tag{5}$$

in Cases 1 and 2 and also in Case 3 if further we have either ($H2_\theta$) , or ($H3_\theta$) and the sequence $n^{1-4\alpha}\rho_n$ is bounded (where α appears in ($H3_\theta$)).

Next we come back to the general equation (1). The assumptions are as follows:

Hypothesis (HS) : $(\theta,t,x) \rightsquigarrow c(\theta,t,x)$ is a function from $\Theta \times [0,1] \times I\!R \mapsto (0,\infty)$ which is twice continuously differentiable in θ and once continuously differentiable in t, and c, $\dot c$ and $\ddot c$ are twice continuously differentiable in x and continuous in t; the process $b = b_t(\omega)$ is optional and locally bounded (locally in time): hence in particular (1) admits a unique weak solution P_θ.

Hypothesis (HI) : (i) For $\zeta \neq \theta$ we have $P_\theta(c(\zeta,t,X_t) = c(\theta,t,X_t)\ \forall t \in [0,1]) = 0$.

(ii) We have $P_\theta(\dot c(\theta,t,X_t) = 0\ \forall t \in [0,1]) = 0$.

The natural extension of (5) is

$$I(\theta) = \int_0^1 \varphi_u(c(\theta,s,X_s), \dot c(\theta,s,X_s))ds. \tag{6}$$

Theorem 2. *Assume (HS), (HI), and also that we are in one of the three cases of (3). Then we can construct estimators $\widehat\theta_n$ which converge to θ in P_θ-probability, and the sequence $\frac{1}{u_n}(\widehat\theta_n - \theta)$ converges in law under P_θ to a variable which can be written as $U/\sqrt{I(\theta)}$, where U is $\mathcal{N}(0,1)$ and independent of $I(\theta)$ given by (6).*

A precise description of the estimators $\widehat\theta_n$ is a bit too complicated to be given here, but these estimators are indeed "explicit".

カウンターパーティ・リスクを加味した金利スワップの評価モデル

An Evaluation Method of Counter party Risk in Interest Rate Swaps

青沼 君明　　中山 季之
Kimiaki Aonuma　　Toshiyuki Nakayama
東京三菱銀行 金融商品開発部
The Bank of Tokyo-Mitsubishi, Ltd

1．概要

金利スワップは、カウンターパーティ・リスクが内包された商品であり、これらのカウンターパーティ・リスクを加味したプライシング・モデルの構築と、保有するスワップ・ポジションのリスク評価モデルの構築が実務上の重要な課題となっている。Jarrow and Turnbull は、カウンターパーティ・リスクを内包する金利スワップのプライシング・モデルを、デフォルトの推移確率を格子によって表現することによって示した。Jarrow and Turnbull の論文では、金利は deterministic であると仮定されている。それに対して本稿では金利はランダムであるとした。また、Jarrow and Turnbull の論文ではカウンター・パーティＡやＢのデフォルト確率を利払い日ごとに与えていた。これに対し本稿では、カウンター・パーティＡやＢのデフォルトを確率変数で表現してその生存関数によって各利払い日時点のデフォルト確率を表わしている。これは、生存関数を推定する方法は実務上よく発達しているという事実を考えてのことである。

2．理論的フレームワーク

カウンター・パーティ A が固定金利による利払い $\overline{R}$ を受取り、カウンター・パーティ B が変動金利による利払いを受け取るスワップ契約について考える。この金利スワップの利払い交換日を $T_1, T_2, \ldots, T_N$ $(T_n - T_{n-1} = \delta,\ n = 1,2,\ldots,N)$ とする。また、利払いの交換の基準となる変動金利の水準は、一期前の利払い交換日に確定するものとする。ここで δ は正の定数であり、期間 $(T_{n-1}, T_n]$ にデフォルトが起こったときの清算処理は、時点 T_n においてなされるものとする。

評価モデルは，完備確率空間 (Ω, F, P) 上で構成する。この空間上の1次元標準ブラウン運動 $(w(t))_{t\geq 0}$ を考え、それによるフィルトレーション $(\mathsf{G}_t)_{t\geq 0}$ を

$$\mathsf{G}_0 = \{A \in \mathsf{F};\ P(A) = 0 \ \ or \ \ P(A) = 1\}$$

$$\mathsf{G}_t = \mathsf{G}_0 \vee \sigma\{w(s);\ s \in [0,t]\},\quad t > 0$$

により定義する。このブラウン運動による HW 型モデルを使って、満期 T で 1 ドル支払われる無リスク割引債を定め、確率測度 P をリスク中立確率測度とみなす。この割引債によ

って時点T_jにおける$[T_j, T_{j+1}]$間のLibor rateを表わす確率変数$R(T_j)$を定める。

カウンターパーティAとBがデフォルトした時点を、$[0,\infty)$に値をとる確率変数τ_A, τ_B（ここでは独立性を仮定していない）で表わす。どちらかがデフォルトする時点を表わす確率変数を$\tau = \tau_A \wedge \tau_B$と表わす。デフォルトしたかどうかを表わす$N_A(t) = 1_{\{\tau_A \le t\}}, N_B(t) = 1_{\{\tau_B \le t\}}, N(t) = 1_{\{\tau \le t\}}$を導入する。金利のみならず、デフォルトの情報も込めたフィルトレーション$(\mathsf{F}_t)_{t\ge 0}$を$\mathsf{F}_t = \bigcap_{s>t}(\mathsf{G}_s \vee \sigma\{\tau_A \wedge s\} \vee \sigma\{\tau_B \wedge s\})$により定義する。

金利スワップの想定元本をN_Pとすると、時点T_jにおいてカウンター・パーティAが受け取る額は$V_P(T_j) = N_P\,\delta\left(\overline{R} - R(T_{j-1})\right)$で計算できる。カウンター・パーティA,Bがデフォルトしたときの回収率を、それぞれφ_A, φ_Bとし共に定数であるものとする。時点T_jにおけるカウンター・パーティAからみたスワップの残存価値（時点T_jより後の、デフォルトも考慮に入れたキャッシュ・フローの総価値）を$V(T_j)$と表わす。まず当然のことながら$V(T_N) = 0$とする。通常の意味でのSaving accountを表わす確率過程を$(\beta(t))_{t\ge 0}$とし、本稿では各$n = 1,2,\ldots,N$に対して

$$V(T_{n-1}) = E\left[\frac{\beta(T_{n-1})}{\beta(T_n)}\left(V_P(T_n) + V(T_n)\right)\left(1 - N(T_n)\right)\middle|\ \mathsf{F}_{T_{n-1}}\right]$$

$$+ E\left[\frac{\beta(T_{n-1})}{\beta(T_n)}\left(V_P(T_n) + V(T_n)\right)\varphi_B 1_{\{V_P(T_n)+V(T_n)<0\}}\left(1 - N_A(T_n)\right)N_B(T_n)\middle|\ \mathsf{F}_{T_{n-1}}\right]$$

$$+ E\left[\frac{\beta(T_{n-1})}{\beta(T_n)}\left(V_P(T_n) + V(T_n)\right)\varphi_A 1_{\{V_P(T_n)+V(T_n)<0\}}N_A(T_n)\left(1 - N_B(T_n)\right)\middle|\ \mathsf{F}_{T_{n-1}}\right]$$

$$+ E\left[\frac{\beta(T_{n-1})}{\beta(T_n)}\left(V_P(T_n) + V(T_n)\right)\left(\varphi_A 1_{\{V_P(T_n)+V(T_n)<0\}} + \varphi_B 1_{\{V_P(T_n)+V(T_n)>0\}}\right)N_A(T_n)N_B(T_n)\middle|\ \mathsf{F}_{T_{n-1}}\right]$$

によりスワップの残存価値をバックワードに定義し、いくつかの仮定の下でこれをフィルトレーション$(\mathsf{G}_t)_{t\ge 0}$による条件付き期待値と生存関数により表現した。

3．実際の数値計算と結論

最も簡易な場合についての数値実験を行った。第1に、A と B 両者のデフォルトを考慮に入れない場合のスワップ・レートが、両者とも格付け B の場合にデフォルトを加味すると2.5bp 程度低下した。第2に、スワップ・レートはカウンター・パーティ B すなわち固定金利の払い手のデフォルトの影響をより強く受けていることが実験で確かめられた。金利スワップでは、通貨スワップと異なり元本部分の交換がないため、カウンターパーティがディフォルトした場合の影響は少ない。しかし、ここで言っているのはあくまでも算出されるスワップ・レートの値に対する影響度であり、ポテンシャル・エクスポーシャーなどは大きくなるので注意が必要である。また、本稿では金利スワップの例で説明したが、同様のフレームワークは通貨スワップの場合にも適用可能である。

Application of tube formula to distributional problems in multiway layouts

栗木 哲 (S. Kuriki, 統計数理研究所・総合研究大学院大学)
竹村 彰通 (A. Takemura, 東京大学大学院情報理工学系研究科)

Standard results of matrix theory play a major role in conventional multivariate statistical analysis or categorical data analysis. In particular when data are summarized as a two-way table, many methods of data analysis have been based on the singular value decomposition of the data matrix. However data may not be summarized as a single matrix. In a factorial design or cross-classification, data are usually obtained as a multiway layout. Moreover even when data are summarized as a two-way table, usual matrix theory is not necessarily applicable. For example, in a two-way cross-classified table with ordinal row and column categories, ordinary matrix methods invariant with respect to permutations of rows or columns are not suitable.

Recently, an integral geometric method called the tube method has been actively developed. This method originates from Hotelling [5] and Weyl [16]. Sun [14] showed how the tube method can be used for deriving distributions of maxima of Gaussian random fields. A closely related technique is the Euler characteristic method developed mainly by Adler and Worsley. A detailed review of the Euler characteristic method and its relation to the tube method is given in Adler [1]. The Euler characteristic method has been extensively used for analyzing brain image data. In Takemura and Kuriki [15] we have established the equivalence of the two methods using an extended form of the Morse theorem (see also Kuriki and Takemura [7]).

Initially we have been investigating the coefficients of the $\bar{\chi}^2$ distribution appearing in order restricted inference from geometric viewpoint. During this investigation we have recognized that the tube method leads to approximation of the upper tail probability of maximum type statistics in the form of linear combination of χ^2 distributions, which can be regarded as a generalization of the $\bar{\chi}^2$ distribution. From this view point we applied the tube method to many distributional problems in conventional multivariate analysis, where matrix theory cannot be applied.

In this talk we review several applications of the tube method to distributional problems in multiway layouts by the authors [6, 8, 9] and by Ninomiya [11, 12]. We give a brief introduction to the tube method. Then we review four applications of the tube method to the testing problems below: (i) A test for interaction in three-way layout based on the three-way analogue of the largest singular value (Boik and Marasinghe [2]). (ii) A test for multivariate normality by searching a nonnormal direction proposed by Malkovich and Afifi [10]. (iii) Testing independence in ordered categorical data by maximizing row and column scores under order restriction (Goodman [3], Nishisato and Arri [13]). (iv) Detecting a change point in two-way layout with ordinal factors (Hirotsu [4]).

References

[1] Adler, R. J. On excursion sets, tube formulas and maxima of random fields. *Ann. Appl. Probab.*, 2000; **10**: 1–74.

[2] Boik, R. J. and Marasinghe, M. G. Analysis of nonadditive multiway classifications. *J. Amer. Statist. Assoc.*, 1989; **84**: 1059–1064.

[3] Goodman, L. A. The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *Ann. Statist.*, 1985; **13**: 10–69.

[4] Hirotsu, C. Two-way change-point model and its application. *Austral. J. Statist.*, 1997; **39**: 205–218.

[5] Hotelling, H. Tubes and spheres in n-spaces, and a class of statistical problems. *Amer. J. Math.*, 1939; **61**: 440–460.

[6] Kuriki, S. Asymptotic distribution of inequality restricted canonical correlation with application to tests for independence in ordered contingency tables. Research Memorandum No. 803, The Institute of Statistical Mathematics, 2001, submitted for publication.

[7] Kuriki, S. and Takemura, A. Distribution of the maximum of Gaussian random field: tube method and Euler characteristic method. *Proc. Inst. Statist. Math.*, 1999; **47**: 199–219 (in Japanese).

[8] Kuriki, S. and Takemura, A. Tail probabilities of the maxima of multilinear forms and their applications. *Ann. Statist.*, 2001; **29**, 328–371.

[9] Kuriki, S. and Takemura, A. Application of tube formula to distributional problems in multiway layouts. METR 01-10, Dept. of Mathematical Engineering and Information Physics, Univ. of Tokyo, 2001.

[10] Malkovich, J. F. and Afifi, A. A. On tests for multivariate normality. *J. Amer. Statist. Assoc.*, 1973; **68**: 176–179.

[11] Ninomiya, Y. Construction of conservative testing for detection of two-way change-point. *Japanese Journal of Applied Statistics*, 2000; **29**: 117–130 (in Japanese).

[12] Ninomiya, Y. Construction of conservative testing for change-point problems in two-dimensional random fields. Research Memorandum No. 746, The Institute of Statistical Mathematics, 2000, submitted for publication.

[13] Nishisato, S. and Arri, P. S. Nonlinear programming approach to optimal scaling of partially ordered categories. *Psychometrika*, 1975; **40**: 525–548.

[14] Sun, J. Tail probabilities of the maxima of Gaussian random fields. *Ann. Probab.*, 1993; **21**, 34–71.

[15] Takemura, A. and Kuriki, S. On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains. *Ann. Appl. Probab.*, 2001, conditionally accepted.

[16] Weyl, H. On the volume of tubes. *Amer. J. Math.*, 1939; **61**: 461–472.

The Asymptotic Expansion Approach to Finance: Computation of Optimal Portfolio for Investment and Variance Reduction of Monte Carlo Simulations

Akihiko Takahashi and Nakahiro Yoshida

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

January 26, 2002

Abstract

We will propose a new computational scheme for the evaluation of the optimal portfolio for investment. Our method is based on an extension of the asymptotic expansion approach which has been recently developed for pricing problems of the contingent claims' analysis by Kunitomo-Takahashi(1992, 1998,2001), Yoshida(1992), Takahashi(1995,1999) Sorensen and Yoshida(1998) and Kashiwakura and Yoshida(2001). In particular, we will explicitly derive a formula of the optimal portfolio associated with maximizing utility from terminal wealth in a financial market with Markovian coefficients and give a numerical example for a power utility function. We will also show that our formula can be used for the variance reduction of the Monte Carlo method in computation of optimal portfolios.

1 Summary

We shall propose a new computational scheme for the evaluation of the optimal portfolios for investment. Our method is based on the asymptotic expansion approach, a unified method of efficient computation justified by Malliavin-Watanabe(1987) theory, which has been recently developed for pricing problems of the contingent claims' analysis by Kunitomo-Takahashi(1992,1995,1998), Yoshida(1992), Takahashi(1995,1999), Kunitomo and Kim(1999), Sorensen and Yoshida(1998) and Kashiwakura and Yoshida(2001). They have developed the method through deriving formulas for practical examples such as average options, basket options, and options with stochastic volatility and with stochastic interest rates in a Markovian setting, as well as bond options(swaptions), average options on interest rates, and average options on foreign exchange rates with stochastic interest rates in the Heath-Jarrow-Morton(1992) framework. In this paper, we extend the method for portfolio problems. In particular, we will explicitly derive the formula of the optimal portfolio associated with maximizing utility from terminal wealth in a financial market with Markovian coefficients, and give a numerical example for a power utility function. In general, it is quite difficult to compute an optimal portfolio explicitly when the investment oppotunity is stochastic in a multiperiod setting. The stochastic control approach initiated by Merton(1969,1971) gives a solution in terms of the derivatives of the value function: While the solution can be evaluated numerically based on the Hamilton-Jacobi-Bellman equation, the implementation is not easy especially for the case of multiple assets. In the martingale approach initiated by Karazas et al.(1987) and Cox and Huang(1989), Ocone and Karatzas(1991) proposed the representation of optimal portfolios by utilizing the Clark formula. Although their representation formulas were derived in general setting, explicit evaluation was obtained only for logarithmic utility functions or a financial market with deterministic coefficients, which were already known without their formulas. Starting with the Clark formula, we will present an explicit expression for the optimal portfolio in a financial market with Markovian coefficients which is more concrete but practically sufficient setting. Moreover, we will show that our formula can be used for the variance reduction of the Monte Carlo method in computation of optimal portfolios. Our method can be also extended to the optimal portfolios associated with maximizing utility from both consumption and terminal wealth, and to the hedging portfolios associated with contingent claims. The organization of this paper is as follows. In Section 2 we explain the problem of the optimal portfolio for investment and restate the problem in a Markovian setting. In Section 3 after explaining basic tools for the asymptotic expansion, we illustrate our method using a power utility function and derive the second order scheme explicitly. In Section 4 we also derive the second order scheme for general utility functions. In Section 5 we give a numerical example. In Section 6 we propose a new variance reduction menthod of Monte Carlo simulations for the computation of optimal portfolios. Finally, in Appendix we show the result of the third order scheme for a power utility function and discuss the validity of the asymptotic expansion for the numerical example.

Information geometry of estimators for diffusion processes with small noise

SEI Tomonari*, KOMAKI Fumiyasu†

Abstract

The second-order efficiency of bias-corrected estimators for diffusion processes with small noise is investigated from the viewpoint of information geometry. A bias-corrected estimator is second-order efficient if and only if the ancillary manifolds are orthogonal to the model and their embedding m-curvatures vanish. In particular, the maximum likelihood estimator is second-order efficient. It is essential that the statistics appeared in the expansions of estimators have asymptotic normality.

1. Diffusion processes with small noise and information geometry

Suppose that $X^\epsilon = \{X_t^\epsilon \mid t \in [0,T]\} \in \mathcal{C}_T$, where $\mathcal{C}_T$ is the set of continuous functions from $[0,T]$ to $\mathbf{R}$, is a diffusion process with small noise parameter $\epsilon \in [0,1]$

$$\mathrm{d}X_t^\epsilon = \mu(X_t^\epsilon, u)\mathrm{d}t + \epsilon \mathrm{d}W_t, \quad X_0^\epsilon = x_0,$$

where $W = \{W_t\}$ is a 1-dimensional standard Brownian motion, $\mu = \mu(\cdot,\cdot)$ is a smooth function and u is a m-dimensional unknown parameter in the parameter space $U \subset \mathbf{R}^m$, and the initial value x_0 is a constant independent of u. The log likelihood function ℓ_ϵ is given by

$$\ell_\epsilon(X,u) = \frac{1}{\epsilon^2}\int_0^T \mu \mathrm{d}X_t - \frac{1}{2\epsilon^2}\int_0^T \mu^2 \mathrm{d}t.$$

We denote this model by $\mathcal{P} = \{\ell_\epsilon(\cdot,u) | u \in U\}$ and introduce some geometrical notations as follows. See Amari [1] for details about information geometry. A tangent space of $\mathcal{P}$ at $u \in U$ is defined by $\mathcal{T}_u = \mathrm{span}\{\partial_a \ell_\epsilon(\cdot,u)\}$, where $\partial_a = \partial/\partial u^a$ for $a = 1,\cdots,m$. The Fisher information metric $\boldsymbol{g}_{ab}$ and e-, m-connection coefficients $\overset{\mathrm{e}}{\boldsymbol{\Gamma}}_{abc}$, $\overset{\mathrm{m}}{\boldsymbol{\Gamma}}_{abc}$ on $\mathcal{P}$ are defined by

$$\begin{aligned}
\boldsymbol{g}_{ab} &= \mathrm{E}_u[\partial_a \ell_\epsilon \partial_b \ell_\epsilon], \\
\overset{\mathrm{e}}{\boldsymbol{\Gamma}}_{abc} &= \mathrm{E}_u[\partial_a\partial_b \ell_\epsilon \partial_c \ell_\epsilon], \\
\overset{\mathrm{m}}{\boldsymbol{\Gamma}}_{abc} &= \overset{\mathrm{e}}{\boldsymbol{\Gamma}}_{abc} + \mathrm{E}_u[\partial_a \ell_\epsilon \partial_b \ell_\epsilon \partial_c \ell_\epsilon],
\end{aligned}$$

respectively. Since all of these quantities are order of ϵ^{-2}, we put $g_{ab} = \epsilon^2 \mathbf{g}_{ab}$, $\overset{\mathrm{e}}{\Gamma}_{abc} = \epsilon^2 \overset{\mathrm{e}}{\boldsymbol{\Gamma}}_{abc}$ and $\overset{\mathrm{m}}{\Gamma}_{abc} = \epsilon^2 \overset{\mathrm{m}}{\boldsymbol{\Gamma}}_{abc}$. The explicit expressions for these quantities have been obtained (See Sei and Komaki [6]). In the following, we use Einstein's summation convention.

2. The asymptotic expansion for the MLE

The maximum likelihood estimator (MLE) $\hat{u}$ with respect to an observed process X^ϵ has an asymptotic expansion:

$$\hat{u}^a \simeq u^a + \epsilon g^{ab}\tilde{y}_b + \epsilon^2\left(g^{ab}g^{cd}\tilde{y}_{bc}\tilde{y}_d - \frac{1}{2}\overset{\mathrm{m}}{\Gamma}_{cd}{}^a g^{ce}g^{df}\tilde{y}_e\tilde{y}_f\right),$$

where

$$\begin{aligned}
\tilde{y}_a &= \epsilon\partial_a \ell_\epsilon, \\
\tilde{y}_{ab} &= \epsilon\partial_a\partial_b\ell_\epsilon - \mathrm{E}_u[\epsilon\partial_a\partial_b\ell_\epsilon] - \overset{\mathrm{e}}{\Gamma}_{ab}{}^c\tilde{y}_c,
\end{aligned}$$

(g^{ab}) is the inverse matrix of (g_{ab}) and the symbol $\simeq$ means asymptotic equivalence. See Kutoyants [5] and Yoshida [7] for the validity of asymptotic expansions related to the MLE.

Using the expansion and asymptotic normality of $\tilde{y}_a$ and $\tilde{y}_{ab}$, we obtain the expansion of the mean square error (MSE) of the bias-corrected MLE $\hat{u}^*$. See Efron [2] and Amari [1] for the concept of bias correction.

*Department of Mathematical Engineering and Information Physics, School of Engineering, University of Tokyo.

†Department of Mathematical Informatics, School of Information Science and Technology, University of Tokyo.

Proposition 2.1 *The MSE of the bias-corrected MLE is given by*

$$\mathrm{E}_u[(\hat{u}^{*a}-u^a)(\hat{u}^{*b}-u^b)] \simeq \epsilon^2 g^{ab} + \frac{\epsilon^4}{2} C^{2ab},$$

where

$$\begin{aligned} C^{2ab} &= (\overset{\mathrm{m}}{\Gamma})^{2ab} + 2(\overset{\mathrm{e}}{\mathrm{H}}_{\mathcal{P}})^{2ab}, \\ (\overset{\mathrm{m}}{\Gamma})^{2ab} &= \overset{\mathrm{m}}{\Gamma}_{cd}{}^{a} \overset{\mathrm{m}}{\Gamma}_{ef}{}^{b} g^{ce} g^{df}, \\ (\overset{\mathrm{e}}{\mathrm{H}}_{\mathcal{P}})^{2ab} &= M_{cdef} g^{ac} g^{be} g^{df}, \\ M_{cdef} &= \mathrm{E}_u[\tilde{y}_{cd}\tilde{y}_{ef}]. \end{aligned}$$

3. The asymptotic expansions for Fisher-consistent estimators

In this section, we assume that the model $\mathcal{P}$ is a curved exponential family embedded in a full exponential family $\mathcal{E} = \{\ell_\epsilon(\cdot,\theta)\}$ defined by

$$\ell_\epsilon(X,\theta) = \epsilon^{-2}(\theta^i s_i(X) - \Psi(\theta,\epsilon)),$$

where θ is the n-dimensional natural parameter, $s(X)$ is the sufficient statistic corresponding to θ and $\Psi(\theta,\epsilon)$ is the potential function. The quantity ϵ plays a role as the dispersion parameter (Jørgensen [3]). See Küchler and Sørensen [4] for details about curved exponential families of stochastic processes. This assumption holds if the drift coefficient is in a linear form.

Let η be the expectation parameter: $\eta_i = \eta_i(u,\epsilon) = \mathrm{E}_u[s_i(X^\epsilon)]$, which depends on ϵ. We suppose that the asymptotic normality

$$\tilde{\eta}_i = \epsilon^{-1}(s_i(X) - \eta_i) \to N(0,(g_{ij})) \quad (1)$$

holds.

Let us consider an estimator expressed by $\hat{u} = T(\hat{\eta})$, where T is a smooth function from R^n to U. We assume that $\hat{u}$ is Fisher consistent, i.e., $u^a = T^a(\eta(u,\epsilon))$ holds for all $u \in U$ and ϵ. We define the ancillary manifold $\mathcal{A}_u$ by $\mathcal{A}_u = T^{-1}(\{u\})$. The notations about the ancillary manifolds follow Amari[1].

The estimator $\hat{u}$ is expanded as

$$\hat{u}^a \simeq u^a + \epsilon\{\partial^i T^a\}\tilde{\eta}_i + \frac{\epsilon^2}{2}\{\partial^i\partial^j T^a\}\tilde{\eta}_i\tilde{\eta}_j.$$

Here, $\partial^i T^a(\eta)$ $(a = 1,\cdots,m)$ are interpreted as vectors normal to the ancillary manifold $\mathcal{A}_u$. It can be shown that the estimator $\hat{u}$ is first-order efficient if and only if $\mathcal{A}_u$ and $\mathcal{T}_u$ are asymptotically orthogonal, i.e., $\partial^i T^a \simeq g^{ab}\partial_b\theta^i$.

From now on, we assume that $\hat{u}$ is first-order efficient and discuss second-order efficiency of $\hat{u}$. By using asymptotic normality (1) of $\tilde{\eta}$ and a relation $\partial^i\partial^j T^a = -A^{i\beta}A^{j\gamma}\overset{\mathrm{m}}{\Gamma}_{\beta\gamma}{}^{a}$, where $A^{i\beta} = \partial^i w^\beta$, the MSE of the bias-corrected estimator $\hat{u}^*$ is represented as follows.

Theorem 3.1 *The MSE of a bias-corrected first-order efficient estimator is*

$$\mathrm{E}_u[(\hat{u}^{*a}-u^a)(\hat{u}^{*b}-u^b)] \simeq \epsilon^2 g^{ab} + \frac{\epsilon^4}{2} C^{2ab}$$

where

$$\begin{aligned} C^{2ab} &= (\overset{\mathrm{m}}{\Gamma})^{2ab} + 2(\overset{\mathrm{e}}{\mathrm{H}}_{\mathcal{P}})^{2ab} + (\overset{\mathrm{m}}{\mathrm{H}}_{\mathcal{A}})^{2ab}, \\ (\overset{\mathrm{m}}{\mathrm{H}}_{\mathcal{A}})^{2ab} &= \overset{\mathrm{m}}{\Gamma}_{\kappa\lambda}{}^{a} \overset{\mathrm{m}}{\Gamma}_{\mu\nu}{}^{b} g^{\kappa\mu} g^{\lambda\nu}, \end{aligned}$$

and $(\overset{\mathrm{m}}{\Gamma})^{2ab}$ *and* $(\overset{\mathrm{e}}{\mathrm{H}}_{\mathcal{P}})^{2ab}$ *are given in proposition 2.1.*

The three terms in C^{2ab} are all non-negative and $(\overset{\mathrm{m}}{\Gamma})^{2ab}$ and $(\overset{\mathrm{e}}{\mathrm{H}}_{\mathcal{P}})^{2ab}$ are independent of the estimator. Therefore, $\hat{u}^*$ is second-order efficient if and only if $(\overset{\mathrm{m}}{\mathrm{H}}_{\mathcal{A}})^{2ab} \to 0$ as $\epsilon \to 0$. The quantity $(\overset{\mathrm{m}}{\mathrm{H}}_{\mathcal{A}})^{2ab}$ means square of the embedding m-curvature of $\mathcal{A}_u$. In particular, the bias-corrected MLE is second-order efficient.

References

[1] Amari, S. (1985). *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics 28, Springer-Verlag.

[2] Efron, B. (1975). *Annals of Statistics*, **3**, 1189-1242.

[3] Jørgensen, B. (1992). *International Statistical Review*, **60**, 5-20.

[4] Küchler, U. and Sørensen, M. (1997). *Exponential Families of Stochastic Processes*, Springer Series in Statistics, Springer-Verlag New York, Inc..

[5] Kutoyants, Yu. A. (1984). *Theory of Probability and its Applications* , **29**, 465-477.

[6] Sei, T. and Komaki, F. (2001). in preparation.

[7] Yoshida, N. (1992). *Probability Theory and Related Fields*, **92**, 275-311.

A flexible class of stochastic volatility models

Michael Sørensen, Department of Statistics and Operations Research, University of Copenhagen.

A generalization of the Black-Scholes model for the logarithm of an asset price

$$dX_t = (\mu + \nu\sigma^2)dt + \sigma dW_t,$$

that takes into account the empirical finding that the volatility σ^2 varies randomly over time is a stochastic volatility model:

$$dX_t = (\mu + \nu V_t)dt + \sqrt{V_t}dW_t. \tag{1}$$

Here the volatility V_t is a stochastic process that cannot be observed directly.

We shall in the following be interested in constructing stochastic volatility models where the volatility process V is stationary with a given distribution with density f. First, a diffusion process with linear drift and a marginal distribution with a given density function f can be obtained as the solution of

$$dV_t = -\theta(V_t - \xi)dt + \sqrt{v(V_t)}dB_t \tag{2}$$

where

$$v(x) = \frac{-\theta}{f(x)} \int_0^x (y - \xi)f(y)dy. \tag{3}$$

This is, of course, provided that (2) has a weak solution. It is not difficult to see that $v(x) > 0$ for all $x > 0$. Under the condition that there exist positive constants K, C and ϵ such that $f(x) \leq Kx^{-(2+\epsilon)}$ for $x \leq C$, we find that $\int_0^\infty v(x)f(x)dx < \infty$ and that the autocorrelation function of V is $\rho(u) = \exp(-\theta u)$. The advantage of a diffusion model with linear drift is that analytically it is relatively tractable. A problem is that it is a well-established empirical fact that the autocorrelation function of the volatility process decreases more slowly than a single exponential function. Under relatively weak regularity conditions, a diffusion model has an exponentially decreasing autocorrelation function. A sufficient condition is that it is ρ-mixing. For this reason, stochastic volatility models with a diffusion volatility process can usually not fit the autocorrelation of the volatility process well.

In applications where the autocorrelation of the volatility process is important, a solution is to use the following construction. We construct a volatility process with autocorrelation function

$$\rho(u) = \varphi_1 \exp(-\theta_1 u) + \cdots + \varphi_m \exp(-\theta_m u), \tag{4}$$

where $\theta_1 > \theta_2 > \cdots > \theta_m > 0$, $\varphi_i > 0$, and $\varphi_1 + \cdots + \varphi_m = 1$. Again we want a volatility process with a given marginal distribution. Suppose that this distribution is infinitely divisible and denote its density function by f and its characteristic function by $C(t)$. Let f_i, $i = 1, \ldots, m$, denote the density of the distribution with characteristic function $C(t)^{\varphi_i}$. Define the volatility process as the sum

$$V_t = V_t^{(1)} + \cdots + V_t^{(m)}, \tag{5}$$

where

$$dV_t^{(i)} = -\theta_i \left(V_t^{(i)} - \varphi_i \xi\right) dt + \sqrt{v_i(V_t^{(i)})} dB_t^{(i)}. \tag{6}$$

Here $B^{(1)}, \ldots, B^{(m)}$ are independent standard Wiener processes,

$$v_i(x) = \frac{-\theta_i}{f_i(x)} \int_0^x (y - \varphi_i \xi) f_i(y) dy,$$

and ξ is the expectation of f. Note that $\varphi_i \xi$ is the expectation of f_i, and that $\mathrm{Var}(V_t^{(i)}) = \varphi_i \mathrm{Var}(V_t)$. The volatility process V defined in this way has marginal distribution with density f and autocorrelation function (4).

Example 1. Suppose we want a volatility process with gamma distributed marginals, i.e.

$$f(x) = \frac{x^{\kappa-1} e^{-x/\beta}}{\beta^\kappa \Gamma(\kappa)},$$

where $\kappa > 0$ and $\beta > 0$. Then f_i is the gamma density with parameters $\kappa_i = \varphi_i \kappa$ and $\beta_i = \beta$, and

$$dV_t^{(i)} = -\theta_i \left(V_t^{(i)} - \varphi_i \alpha\beta\right) dt + \sqrt{2\beta\theta_i V_t^{(i)}} dB_t^{(i)}.$$

Example 2. Suppose we want a volatility process with inverse Gaussian distributed marginals, i.e.

$$f(x) = \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} x^{-\frac{3}{2}} \exp\left(-\frac{1}{2}\left(\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right)\right),$$

where $\lambda > 0$ and $\mu > 0$. Then f_i is the inverse Gaussian density with parameters $\lambda_i = \varphi_i^2 \lambda$ and $\mu_i = \varphi_i \mu$, and $V_t^{(i)}$ solves (6) with $\xi = \mu$ and

$$v_i(x) = \theta_i \sqrt{\frac{2\pi\mu^2}{\lambda}} e^{\varphi_i \lambda/\mu} x^{3/2} \exp\left(\frac{1}{2}\left(\frac{\lambda}{\mu^2} x + \varphi_i^2 \lambda x^{-1}\right)\right) \Phi\left(\varphi_i \sqrt{\frac{\lambda}{x}} + \sqrt{\frac{\lambda x}{\mu^2}}\right)$$

with Φ denoting the standard normal distribution function.

A more detailed account of the theory presented here can be found in Bibby and Sørensen (2002).

References

Bibby, B.M. and Sørensen, M. (2002). Flexible stochastic volatility models of the diffusion type. In preparation.

On inherited ergodicity in a class of partially observed stochastic models

Hiroki Masuda*
Graduate School of Mathematical Science, University of Tokyo. †

1 Contents

In this talk, we consider a moment estimation in a class of partially observed stochastic models with a certain stochastic structure. Our argument is motivated by the recent work of Genon-Catalot, Jeantheau and Laredo (2000, Bernoulli). With continuous time framework, our model consists of a hidden (unobservable) process X with an ε-Markov structure, an observed process Y and their driving Lévy processes $L^{(1)}$ and $L^{(2)}$. We assume that observed data is only $\{Y_{j\Delta} : j = 0, 1, \dots, n\}$ with a fixed (deterministic) time interval $\Delta > 0$. We suppose that a hidden process $X = (X_t)_{t\in\mathbf{R}_+}$ is strictly stationary and α-mixing (hence ergodic) and construct a discrete time observed process $y = (y_j)_{j\in\mathbf{Z}_+^*}$, $\mathbf{Z}_+^* := \{1, 2, \dots\}$, such that $y_j = Y_{j\Delta} - Y_{(j-1)\Delta}$. Then, under some regularity condition, y inherits ergodicity of X so that the usual moment method can be applied for this model. Of course, construction of moment estimators strongly depends on the concrete structure of the model.

As a special case for X, we treat a strictly stationary Ornstein-Uhlenbeck type (OU) process driven by a Lévy processes. Recently OU process based modeling have been proposed by the work of Barndorff-Nielsen and Shephard (2001, JRSS) mainly about stochastic volatility models in mathematical finance. They have not only flexibility for modeling purpose but also mathematical tractability. Under mild conditions, it can be seen that they possess exponentially decreasing β-mixing coefficients so that they exhibit ergodicity and then the moment estimation in their model can be verified.

Two types of examples including stochastic volatility models with numerical experiments are presented in one-dimensional framework.

2 A class of hidden ε-Markov model

We will consider a stochastic process $(X, Y) = \{(X_t, Y_t)\}_{t\in\mathbf{R}_+}$ driven by a process $L = \{(L_t^{(1)}, L_t^{(2)})\}_{t\in\mathbf{R}_+}$ on a given probability space $(\Omega, \mathcal{F}, P)$ where X and Y are $\mathbf{R}^{d_1}$- and $\mathbf{R}^{d_2}$-dimensional càdlàg stochastic processes, respectively. We assume that only Y is observable while X is unobservable. Moreover, we assume that $L^{(1)}$ and $L^{(2)}$ are $\mathbf{R}^{r_1}$ and $\mathbf{R}^{r_2}$-dimensional Lévy processes (càdlàg processes with independent increments, in other words, random walks in continuous time) starting at the origin, respectively. The distribution of (X, Y) (possibly also of $(L^{(1)}, L^{(2)})$) depends on unknown parameter θ which we want to estimate where Θ is a bounded domain in $\mathbf{R}^p$, $p \geq 1$. Let P_0 and E_0 denote the probability measure corresponding to the true value $\theta_0 \in \Theta$ and the expectation under P_0, respectively. For an $I \subseteq \mathbf{R}_+$, let $\mathcal{F}_I^X := \sigma[X_t : t \in I] \vee \mathcal{N}$ and $\mathcal{F}_I^{dL} := \sigma[L_t - L_s : s, t \in I] \vee \mathcal{N}$ where $\mathcal{N}$ denotes the σ-field generated by all the P-null sets.

Assume the following three statements about structure of our model.

A1 (X_0, Y_0) and $(L_t^{(1)}, L_t^{(2)})_{t\in\mathbf{R}_+}$ are mutually independent while X_0 and Y_0, and $L^{(1)}$ and $L^{(2)}$ may be correlated respectively.

*E-mail : flstf@ms.u-tokyo.ac.jp
†3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan.

A2 Under P_0, X is strictly stationary and α-mixing, that is,

$$\alpha_X(t) := \sup\{|P_0(A \cap B) - P_0(A)P_0(B)| : A \in \mathcal{F}^X_{[0]}, B \in \mathcal{F}^X_{[t,\infty)}\} \to 0$$

as t tends to infinity.

A3 $\sigma[X_t] \subseteq \mathcal{F}^X_{[s-\varepsilon,s]} \vee \mathcal{F}^{dL}_{[s,t]}$ and $\sigma[Y_t - Y_s] \subseteq \mathcal{F}^X_{[s-\varepsilon,s]} \vee \mathcal{F}^{dL}_{[s,t]}$ for some ε and any s and t such that $\varepsilon \geq 0$, $s, t \in \mathbf{R}_+$ and $0 \leq s \leq t$ where, for convention, $X_u = X_0$ for $u \in [-\varepsilon, 0]$.

As we have seen, we construct a discrete time observable process y from Y by

$$y_j = Y_{j\Delta} - Y_{(j-1)\Delta}, \qquad j \in \mathbf{Z}^*_+$$

where $\Delta > 0$ is a fixed deterministic sampling interval.

It can be shown that y inherits ergodicity from X under above assumptions so that the moment method can be applied with the aid of *Birkhoff's ergodic theorem* and *Ibragimov's central limit theorem for strictly stationary mixing processes* with additional assumptions.

Our argument is also valid in discrete time framework by modifying the assumptions by an obvious way.

3 Some concrete examples in one-dimensional framework

3.1 The random trend model with a hidden Gaussian process

Consider the model

$$\begin{cases} dX_t &= p(q - X_t)dt + \sqrt{r}dW_t, \\ dY_t &= X_t dt + dL_t \end{cases}$$

where W and L are one-dimensional Wiener process and Lévy process, respectively, which may be correlated. For example, $L_t = \rho W_t + \sqrt{1-\rho^2}\tilde{W}_t$ with $|\rho| \leq 1$ and an Wiener process $\tilde{W}$ which is independent of W. In this example, $X = (X_t)_{t\in\mathbf{R}_+}$ expresses a certain trend varying along time, and $Y = (Y_t)_{t\in\mathbf{R}_+}$ does a noisy observation. It follows that y possesses ergodicity.

3.2 The stochastic volatility model of Barndorff-Nielsen and Shephard

A strictly stationary one-dimensional OU process $X = (X_t)_{t\in\mathbf{R}_+}$ is expressed as

$$X_t = e^{-\lambda t}X_0 + \int_0^t e^{-\lambda(t-s)}dZ_t$$

with a one-dimensional Lévy process $Z = (Z_t)_{t\in\mathbf{R}_+}$ and a positive constant (regression parameter) λ. Under following conditions, X is ergodic with an exponentially decreasing β-mixing coefficient.

B1. The Lévy measure of Z has an absolutely continuous part whenever Z is non-Gaussian. If Z has no jumps, the Gaussian part of Z is nondegenerate.

B2. The stationary distribution μ of X has a first order absolute moment at least.

In multi-dimensional case, we need a further condition on the regression parameter for ergodicity of X.

Barndorff-Nielsen and Shephard (2001, JRSS) considered a stochastic volatility model with X as the squared volatility process. The dynamics is

$$\begin{cases} dX_t &= -\lambda X_t dt + dZ_{\lambda t}, \quad X_0 = \eta, \\ dY_t &= (\mu + \beta X_t)dt + \sqrt{X_t}dW_t, \quad Y_0 = y_0. \end{cases}$$

Here μ, $\lambda > 0$ and β are parameters, W is a standard Wiener process which is independent of a Lévy process Z containing some unknown parameters, η is a random variable distributed as a given marginal one of X, and y_0 is a random variable or a constant. The reason why the time scale of Z is different from other processes is to stay the stationary distribution of X hold whatever the value of λ is. Again ergodicity of y enables us to apply the moment method. ∎

WINDOWS 2001

Yu.A. Kutoyants, *Laboratoire de Statistique & Processus, Université du Maine, 72085 Le Mans, Cédex 9, FRANCE*
Tokyo, December 2001

We consider several problems of parameter estimation concerning ergodic diffusion process

$$\mathrm{d}X_t = S(\vartheta, X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad X_0, \quad 0 \le t \le T,$$

where $\vartheta \in \Theta = (\alpha, \beta)$ is the unknown parameter and the trend coefficient $S(\cdot, x)$ and diffusion coefficient $\sigma(x)^2$ are known functions, providing a weak solution of this equation. The goal of this work is to show how the choice of windows for observations of this process can help in the construction of estimators. The detailed proofs can be found in [3].

Optimal observation windows. Let us consider the problem of parameter estimation in the situation when we can observe the process X_t in the window $\mathcal{A}_\lambda$ of fixed length. The goal is to obtain an *optimal window in the class*, i.e., the window which provides the minimal asymptotic variance of the best estimator.

Therefore we have two problems. The first one is to construct the lower bound on the meansquare errors for the all choices of the windows and estimators and the second - to construct a window and an estimator which attains this bound.

Proposition. *Under regularity conditions*

$$\varliminf_{\delta\to 0}\varliminf_{T\to\infty}\inf_{\{\bar{\mathbb{A}}_T,\,\bar{\vartheta}_T\}}\sup_{|\vartheta-\vartheta_0|<\delta} T\,\mathbf{E}_\vartheta\left(\bar{\vartheta}_{T,\bar{\mathbb{A}}_T}-\vartheta\right)^2 \ge \mathrm{I}(\vartheta_0,\mathbb{A}_0^*)^{-1},$$

where inf is taken over all possible choices of windows and estimators and the quantity $\mathrm{I}(\vartheta_0, \mathbb{A}_0^*)$ plays the role of Fisher information in this problem..

We call a strategy (window $\bar{\mathbb{A}}_T$+estimator $\bar{\vartheta}_{T,\bar{\mathbb{A}}_T}$) asymptotically optimal if it provides the equality and we propose an asymptotically optimal in this sense strategy.

Contaminated model. Let us suppose that the observed ergodic diffusion process is

$$\mathrm{d}X_t = S(\vartheta, X_t)\ \mathrm{d}t + h(X_t)\ \mathrm{d}t + \sigma(X_t)\ \mathrm{d}W_t,$$

where $h(\cdot)$ is unknown to the observer function of known support $\mathbb{A}$, i.e., $h(x) = 0$ if $x \in \mathbb{A}^c$.

We have once more two problems: to construct the lower bound on the risk of all estimators and then to find an estimator which attains this bound.

Let us fix $\vartheta_0, h_0(\cdot)$ and introduce the δ-vicinity V_δ of this fixed model

Proposition. (Höpfner - Kutoyants) *Suppose that the regularity conditions are fulfilled, then for all estimators* $\bar{\vartheta}_T$ *and* $\ell(\cdot) \in \mathscr{W}_p$

$$\varliminf_{\delta\to 0}\varliminf_{T\to\infty}\sup_{\vartheta,h(\cdot)\in V_\delta}\mathbf{E}_{\vartheta,h}\ell\left(\sqrt{T}\left(\bar{\vartheta}_T-\vartheta\right)\right)\geq\mathbf{E}\,\ell\left(\zeta\,\bar{\mathrm{I}}\left(\vartheta_0,h_0\right)^{-1/2}\right)$$

where $\bar{\mathrm{I}}(\vartheta,h)=\mathbf{E}_{\vartheta,h}\left(\frac{\dot{S}(\vartheta,\xi_*)}{\sigma(\xi_*)}\right)^2\chi_{\{|\xi_*|>A\}}>0.$

Then we propose an estimator $\hat{\vartheta}_{T,\mathbb{A}^c}$ which provides the equality in the given lower bound.

Similar result for nul-recurrent diffusion process are obtained in [1].

Change-point problem. Suppose that the observations are contaminated, i.e.,

$$\mathrm{d}X_t=-\operatorname{sgn}\left(X_t-\vartheta\right)\,\mathrm{d}t+h\left(X_t\right)\,\mathrm{d}t+\mathrm{d}W_t,\quad X_0,$$

where $h(\cdot)$ is unknown to the observer function. Let us fix $\gamma<1$ and denote by $\mathcal{H}_\gamma$ the set of functions $\mathcal{H}_\gamma=\{h(\cdot):\sup_x|h(x)|<\gamma\}$, i.e., the values of $h(\cdot)$ are in the window $\mathbb{B}=[-\gamma,\gamma]$.

Proposition.([2]) *The "MLE"* $\hat{\vartheta}_T$ *is uniformly in* $h(\cdot)\in\mathcal{H}_\gamma$ *consistent,* $\mathcal{L}_{h,\vartheta}\left\{T\left(\hat{\vartheta}_T-\vartheta\right)\right\}\Longrightarrow\mathcal{L}_{h,\vartheta}\{u^*\}$, *and the moments converge too.*

Here u^* is the random variable defined with the help of the limiting likelihood ratio.

Remark. Similar results are obtained in the problem *cusp* estimation, i.e., in the problem of parameter estimation by observations with the trend coefficient admitting the representation $S(\vartheta,x)=a\,|x-\vartheta|^p+r\,(x-\vartheta)$ where $p\in(0,1/2)$.

References

[1] Höpfner, R., Kutoyants Yu. A.,(2001) *On a problem of statistical inference in null recurrent diffusion*, to appear in Scandinavian J. Statistics

[2] Kutoyants Yu. A., (2000) *On parameter estimation for switching ergodic diffusion processes*, CRAS Paris, t. 330, Série 1, 925-930.

[3] Kutoyants Yu.A. (2001) *Statistical Inference for Ergodic Diffusion Processes*, forthcoming book.

General State Space Modeling for Complex Time Series

Genshiro Kitagawa
The Institute of Statistical Mathematics

1. General State Space Model and State Estimation

Consider a nonlinear non-Gaussian state space model for the time series y_n,

$$x_n = F_n(x_{n-1}, v_n), \qquad y_n = H_n(x_n, w_n), \tag{1}$$

where x_n is an unknown state vector, v_n and w_n are the system noise and the observation noise with densities $q_n(v)$ and $r_n(w)$, respectively. Two models in (1) are called the system model and the observation model, respectively. The initial state x_0 is assumed to be distributed according to the density $p_0(x)$. $F_n(x, v)$ and $H_n(x, w)$ are possibly nonlinear functions of the state and the noise.

The most important problem in state space modeling is the estimation of the state vector x_n from the observations, $Y_t \equiv \{y_1, \ldots, y_t\}$, since many important problems in time series analysis can be solved by using the estimated state vector. The problem of state estimation can be formulated as the evaluation of the conditional density $p(x_n|Y_t)$. Corresponding to the three distinct cases, $n > t$, $n = t$ and $n < t$, the conditional distribution, $p(x_n|Y_t)$, is called the predictor, the filter and the smoother, respectively.

For general state space models, the conditional distributions become non-Gaussian and their distributions cannot be completely specified by the mean vectors and the variance covariance matrices. Therefore, various types of approximations to or assumptions on the densities have been used to obtain recursive formulas for state estimation. However, the following non-Gaussian filter and smoother [2] can yield an arbitrarily precise posterior density.

[Non-Gaussian Filter]

$$p(x_n|Y_{n-1}) = \int p(x_n|x_{n-1})p(x_{n-1}|Y_{n-1})dx_{n-1}, \quad p(x_n|Y_n) = \frac{p(y_n|x_n)p(x_n|Y_{n-1})}{p(y_n|Y_{n-1})}, \tag{2}$$

where $p(y_n|Y_{n-1})$ is the predictive distribution of y_n and is defined by $\int p(y_n|x_n)p(x_n|Y_{n-1})dx_n$.

[Non-Gaussian Smoother]

$$p(x_n|Y_N) = p(x_n|Y_n)\int \frac{p(x_{n+1}|x_n)p(x_{n+1}|Y_N)}{p(x_{n+1}|Y_n)}dx_{n+1}. \tag{3}$$

The direct implementation of the formula requires computationally very costly numerical integration and can be applied only to lower dimensional state space models. To mitigate the computational burden, numerical methods based on Monte Carlo approximation of the distribution have been proposed [1, 3]. In the Monte Carlo filtering, we approximate each density function by many particles that can be considered as realizations from that distribution. Specifically, assume that each distribution is expressed by using m particles as follows: $\{p_n^{(1)}, \ldots, p_n^{(m)}\} \sim p(x_n|Y_{n-1})$, $\{f_n^{(1)}, \ldots, f_n^{(m)}\} \sim p(x_n|Y_n)$, $\{s_{n|N}^{(1)}, \ldots, s_{n|N}^{(m)}\} \sim p(x_n|Y_N)$. Namely, $p(x_n|Y_{n-1})$ is approximated by the probability function $\Pr(x_n = p_n^{(j)}|Y_{n-1}) = 1/m$, for $j = 1, \cdots, m$. Then it can be shown that a set of realizations expressing the one step ahead predictor $p(x_n|Y_{n-1})$ and the filter $p(x_n|Y_n)$ can be obtained recursively as follows.

[Monte Carlo Filter]

1. *Generate a random number* $f_0^{(j)} \sim p_0(x)$ *for* $j = 1, \ldots, m$.

2. *Repeat the following steps for* $n = 1, \ldots, N$.

   (a) *Generate a random number* $v_n^{(j)} \sim q(v)$, *for* $j = 1, \ldots, m$.

   (b) *Compute* $p_n^{(j)} = F(f_{n-1}^{(j)}, v_n^{(j)})$, *for* $j = 1, \ldots, m$.

   (c) *Compute* $\alpha_n^{(j)} = p(y_n|p_n^{(j)})$ *for* $j = 1, \ldots, m$.

   (d) *Generate* $f_n^{(j)}$, $j = 1, \ldots, m$ *by the resampling of* $p_n^{(1)}, \ldots, p_n^{(m)}$, *with the weights proportional to* $\alpha_n^{(1)}, \ldots, \alpha_n^{(j)}$.

The above algorithm for Monte Carlo filtering can be extended to smoothing by a simple modification. The details of the derivation of the algorithm is shown in [3].

2. Self-organizing State Space Model

The state space model, (1) and (2), usually contains several unknown parameters such as the variances of the noises and the coefficients of the functions F_n and H_n. The vector consisting of such unknown parameters is hereafter denoted by θ. If the non-Gaussian filter is implemented by the Monte Carlo filter, the sampling error sometimes renders the maximum likelihood method impractical. In this case, instead of estimating the parameter θ by the maximum likelihood method, we consider a Bayesian estimation by augmenting the state vector as $z_n = [x_n^T, \theta^T]^T$. The state space model for this augmented state vector z_n can be easily defined from the original state space model.

The marginal posterior densities of the parameter and of the original state can be obtained from the posterior distribution $p(z_n|Y_N)$ given the entire observations $Y_N = \{y_1, \cdots, y_N\}$. This method can be easily extended to a time-varying parameter situation where the parameter $\theta = \theta_n$ evolves with time n.

So far, we have considered the case when the parametric form of the model is given. As an extension of the self-organizing state space model, by using the prediction error in on-line modeling, we developed a method of constructing state space model without assuming a parametric form of the noise distribution. We note here that in the state space modeling, from the derivation of the non-Gaussian filter and smoother, we can use all the information in obtaining the predictive distribution

$$x_n \sim p(x_n|X_{n-1}, Y_{n-1}), \qquad y_n \sim p(y_n|X_n, Y_{n-1}). \tag{4}$$

It suggests that the noise density $q(v)$ can be estimated from Y_{n-1}. One way of estimating the noise distribution $q_n(v)$ is to use the following upadating formula

$$q_n(v) = (1-\alpha)q_{n-1}(v) + \alpha \sum_{j=1}^{K} s(v - \varepsilon_{n-1}^{(j)}), \tag{5}$$

where $\varepsilon_n^{(j)}$ is the prediction error at time n, $s(v)$ is an arbitrary density function and α is the discounting factor and satisfy $0 < \alpha < 1$. By the use of this updating formula and the Monte Carlo filter, it is possible to construct a self-organizing state space model that can automatically adapt to the change of noise distribution.

References

[1] N. J. Gordon, D. J. Salmond and A. F. M. Smith (1993), "Novel approach to nonlinear/non-Gaussian Bayesian state estimation", *IEE Proceedings-F*, Vol. 140, No. 2, 107–113.

[2] G. Kitagawa (1987), "Non-Gaussian state-space modeling of nonstationary time series" (with discussion). *Journal of the American Statistical Association,* Vol. 82, No. 400, 1032–1063.

[3] G. Kitagawa (1996), "Monte Carlo filter and smoother for non-Gaussian nonlinear state space model", *Journal of Computational and Graphical Statistics*, Vol. 5, 1-25.

[4] G. Kitagawa (1998), "Self-organizing state space model", *Journal of the American Statistical Association*, Vol. 93, No. 443, 1203–1215.

Conditional Asymptotic Expansion and Its Applications *

Nakahiro YOSHIDA
Graduate School of Mathematical Sciences, University of Tokyo,
3-8-1 Komaba, Meguro-ku, Tokyo 153, Japan.
e-mail: nakahiro@ms.u-tokyo.ac.jp

In this study, we considered conditional limit theorems and conditional asymptotic expansions. Our discussion is based on the Malliavin calculus. We provided asymptotic expansions in a general setting including the so-called small σ-models. In order to give a basis to the asymptotic expansion scheme for perturbed jump systems, we built an extension to the Watanabe theory in part. Finally, we derived the asymptotic expansions (*double Edgeworth expansions*) of conditional expectations.

The Malliavin calculus is nowadays recognized as an important instrument from a practical computational point of view in theoretical statistics, stochastic numerical analysis and mathematical finance as well as probability theory. It enables us to apply a usual differential calculus to *irregular* functionals, which very often appear, for example, as coverage probabilities, non-differentiable payoff functions, and so on.

The conditional expectation may be one of most irregular functionals. For a continuously distributed conditioning variable, it requires the analysis over a null set. Without doubt, the conditional stochastic calculus features in statistics: sufficient statistics in unbiased estimation and testing hypotheses (e.g., Lehmann-Scheffé theorem, Rao-Blackwell theorem, Neyman structure), conditional likelihood and conditional inference, conditionally Gaussian experiments as limits in LAMN situations, approximation formulas connected with the conditional distribution such as the p^* (magic) formula of Barndorff-Nielsen, filtering problems, recently introduced partial mixing, etc. In spite of the importance, conditional asymptotics does not seem to be so well founded as to fulfill the practical purpose.

We provided, in a general setting, asymptotic expansions under small perturbations. The small σ-theory has been well developed in statistics. Kutoyants [11] thoroughly investigated inference for diffusion type processes with small noises. Asymptotic expansions were presented by [21, 22, 23] by means of the Malliavin calculus and Prof Watanabe's theory. See also Sakamoto and Yoshida [12], Yoshida [24], and Uchida and Yoshida [18] for more statistical applications. As a byproduct, inspired by Professors Kunitomo and Takahashi's first-order approximation for a geometric Brownian motion, the asymptotic expansion scheme to compute the values of options was provided in [22]. There are many studies thereafter in this direction: Kunitomo and Takahashi [8, 9] Takahashi [14, 15], Kim and Kunitomo [6], Sørensen and Yoshida [13], Takahashi and Yoshida [16], Kashiwakura and Yoshida [5].

Recently, modeling with Lévy processes is attracting attention in financial statistics. In order to give a basis to the asymptotic expansion scheme for perturbed jump systems, we built an extension to the Watanabe theory. We adopted the Malliavin calculus formulated by Bichteler et al. [2]. Differently from the original form of Watanabe's theory [20] for Wiener functionals (also see Watanabe [19], Ikeda and Watanabe [4]), we do not use (have) Sobolev spaces of generalized

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

functionals in our setting. For this reason, we will go through by the generalized integral operator for Schwartz distributions.

After preparing asymptotic expansions for generalized expectations, it is straightforward to obtain our main results. The asymptotic expansion of conditional expectations will be derived together with a few variants. They are called the **double Edgeworth expansions**. In the present article, we only treat most simple double expansions. We will present other variants (e.g., Edgeworth-saddlepoint approximation) elsewhere by applying Schilder-type expansions of densities (cf. Kusuoka and Stroock [10], Takanobu and Watanabe [17]).

As for the role of the asymptotic expansion in the theoretical statistics, we refer the reader for example to Barndorff-Nielsen and Cox [1], Ghosh [3].

An application to the Kitagawa filter [7] was also considered.

References

[1] Barndorff-Nielsen, O.E., and Cox, D.R.(1994): Inference and Asymptotics. Chapman & Hall.

[2] Bichteler, K., Gravereaux, J.-B., Jacod, J.: Malliavin calculus for processes with jumps. New York London Paris Montreux Tokyo: Gordon and Breach Science Publishers 1987

[3] Ghosh, J.K.: Higher order asymptotics. California: IMS 1994

[4] Ikeda, N., Watanabe, S.: Stochastic differential equations and diffusion processes, second edition. North-Holland/Kodansha, Tokyo (1990)

[5] Kashiwakura, K., Yoshida, N.: Asymptotic and hybrid expansions. in preparation (2001)

[6] Kim, Y.J., Kunitomo, N.: Pricing Options Under Stochastic Interest Rates. Asia Pacific Financial Markets, **6**, 49-70. (1999)

[7] Kitagawa, G.: Non-Gaussian state space modelling of non-stationary time series. J. Am. Statist. Assoc. **82**, 503-514 (1987)

[8] Kunitomo, N., Takahashi, A.: On Validity of the Asymptotic Expansion Approach in Contingent Claim Analysis. Discussion Paper No.98-F-6, Faculty of Economics, University of Tokyo. (1998)

[9] Kunitomo, N. and Takahashi, A.: The Aysmptotic Expansion Approach to the Valuation of Interest Rate Contingent Claims. Mathematical Finance, **11**, 117-151 (2001)

[10] Kusuoka, S., Stroock, D.W.: Precise asymptotics of certain Wiener functionals. J. Functional Analysis **99**, 1-74 (1991)

[11] Kutoyants, Yu. A.: Identification of dynamical systems with small noise. Kluwer (1994)

[12] Sakamoto, Y., Yoshida, N.: Expansion of perturbed random variables based on generalized Wiener functionals. J. Multivariate Anal., **59-1**, 34-59 (1996).

[13] Sørensen, M., Yoshida, N.: Random limit expansion for small diffusion processes. unpublished manuscript (2000)

[14] Takahashi, A.: Essays on the Valuation Problems of Contingent Claims. Ph.D. Dissertation, Haas School of Business, University of California, Berkeley 1995

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

[16] Takahashi, A., Yoshida, N.: Asymptotic Expansion Scheme for Optimal Portfolio. submitted (2001)

[17] Takanobu, S., Watanabe, S.: Asymptotic expansion formulas of the Schilder type for a class of conditional Wiener functional integrations. In: K.D. Elworthy and N. Ikeda (eds.) Asymptotic problems in probability theory: Wiener functionals and asymptotics, Proceedings of the Taniguchi International Symposium, Sanda and Kyoto, 1990, 194-241. UK: Longman 1993

[18] Uchida, M., Yoshida, N.: Information Criteria for Small Diffusions via the Theory of Malliavin-Watanabe. submitted (1999)

[19] Watanabe, S.: Malliavin's calculus in terms of generalized Wiener functionals. In: (Lect. Notes Control Inf. Sci., Vol. 49) Berlin Heidelberg New York: Springer (1983)

[20] Watanabe,S.: Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels. Ann. Probab. **15**, 1-39 (1987)

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

[22] Yoshida,N.: Asymptotic expansion for statistics related to small diffusions. J. Japan Statistical Society, **22**, No. 2, 139-159 (1992)

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

[24] Yoshida, N.: Asymptotic expansions for perturbed systems on Wiener space: maximum likelihood estimators. J. Multivariate Anal. **57**, 1-36 (1996)

[25] Yoshida, N.: Conditional asymptotics. preprint (2001)

Wavelet Methods for Time Series Analysis

Katsuto Tanaka
Hitotsubashi University, Tokyo

1. Introduction

The wavelet analysis started from quite a practical reason some 20 years ago. Since then it has been developed in fields as diverse as engineering, mathematics, image processing, statistics and so on. A special feature of the wavelet analysis is to look at the signal or data from two aspects, that is, the *time* and the *scale*, which enables us to see, not only the forest, but also the trees, so to speak.

In this report, concentrating on statistics, particularly on time series analysis, we review the methodology of the wavelet approach and seek to explore some applications to time series analysis. In Section 2 the idea of the wavelet analysis is explained from a statistical viewpoint. The detailed and basic ideas of the wavelet analysis can be found in Percival and Walden (2000), where discrete wavelet methods are extensively discussed for time series analysis.

Section 3 gives some useful examples in which the wavelet analysis offers a powerful tool, among which are as follows. The first is concerned with the wavelet-based estimation of the fractional differencing parameter in the so-called fractional ARIMA (ARFIMA) models. It is the power law nature of the spectrum of the long memory process that the wavelet analysis turns out to be successful. The second deals with spurious regression among fractionally differenced series and considers the wavelet method for detecting spurious regression. The third is concerned with wavelet-based tests for homogeneity of variance in long-memory time series, whose problem is related with testing for structural breaks.

2. DWT and MODWT

The DWT (discrete wavelet transform) has been devised as a need to analyze and synthesize discrete time series. It is the CWT (continuous wavelet transform) that has been initially developed. A wavelet $\psi(t)$ plays an important role in the definition of the CWT, which is a generic function, unlike trigonometric or exponential functions.

The CWT is different from the Fourie transform in that the former has two parameters, while the latter one parameter only. It is possible to restore the original signal from the resulting CWT under some restrictions, which is given by the ICWT (inverse

CWT).

The transition from the CWT to the DWT is not straightforward and takes a long way to arrive at. We leave the details of the derivation of the DWT to Percival and Walden (2000). An important thing is that we can accomplish the DWT in an efficient way by the so-called pyramid algorithm. The resulting wavelets are collected on a scale by scale basis.

The DWT is an orthonormal transform and is most often used, but it has several limitations (Percival and Walden (2000)), among which are

(a) The sample size T must be a power of 2.

(b) The number of wavelet coefficients, T_j, decreases by a factor of 2 for each increasing level of the transform so that wavelet coefficients $\mathbf{W}_j$ for j large cannot be used for inference purposes.

(c) The DWT coefficients are not circularly shift invariant.

These deficiencies can be overcome by using the MODWT (maximal overlap DWT) with some copmputational price to pay, although its computational burden is the same as the widely used fast Fourier transform. Note that transforms that are essentially the same as the MODWT have been discussed in the wavelet literature under the names 'non-decimated DWT', 'translation invariant DWT', 'time invariant DWT' and so on.

3. Some applications

We discuss some examples in which the wavelet method proves useful. Among the important and advantageous features associated with the DWT are (1) the decorrelation property, that is, the property that the DWT makes a strongly dependent sequence uncorrelated, which is possible even in nonstationary processes, and (2) the special relationship between the wavelet variance and the scale. These properties of the DWT enable us to analyze time series data more closely. More specifically, we can apply wavelet methods to the following problems.

(a) Wavelet-based estimation of the fractional differencing parameter d

(b) Wavelet solution to the spurious regression of fractionally differenced processes

(c) Wavelet-based test for homogeneity of variance in long-memory time series

Use of Stochastic Differential Equations in Financial Time Series Analysis

Tohru OZAKI

Institute of Statistical Mathematics, Tokyo and Graduate University of Advanced Studies

The innovation approach, in time series analysis, tries to find a causal model which "whiten" the observed time series data into Gaussian white noise. By doing so it provides us with a method to predict the series and to characterize the dynamics behind the time series with the identified causal model. The optimal smoother and the optimal filter of the variables used in the model for the time series are obtained at the same time from the identified causal model. This technique has been developed by Wold(1938), Kolmogorov(1941), Wiener(1949), Kalman(1960), Kailath(1968), Akaike(1969) and Box-Jenkins(1970) etc in the last century. Of course the causal model for the real data is unknown and we have to identify the model among the several candidates. The success of the approach in real application is closely dependent on the "causal model" we use with the innovation approach. Naturally the models employed with the approach were limited to linear models at the early stage of the development in 1940 to 1960s. However there is no reason why we have to use only linear causal models. Unfortunately developing useful nonlinear causal models in applications did not attract time series analysts' attention as much as applied scientists in meteorology and biology etc.. Emergence of chaos studies in 1970's was a challenge to the traditional approach of time series analysis. It is May(1976) which made the idea of deterministic chaos widely known among scientists related to time series analysis. This encouraged many scientists to pay more attention to models with an alternative nonlinear dynamics, which could reduce the uncertainty, i.e. the prediction errors, hopefully to be zero at an ultimate stage. In recent time series analysis, as several linear models have been taken over by rnewly developed nonlinear models, the original idea of innovation approach developed by Wiener, Kalman and Box-Jenkins seemed to have been put aside as well.

The innovation approach, however, was not completely thrown out in some of the group of time series analysts. Studies of nonlinear dynamic models, including those new animals such as deterministic and stochastic differential equation models and neural network models, have been pursued also from a stand point of the traditional innovation approach. In 1960-1970, it was widely known, in the traditional time series school, that nonlinear polynomial AR models are not appropriate for simulating the process although they improve prediction performance significantly compared with linear models. They are computationally easy to estimate but they always lead us to computational explosions if the models are simulated with Gaussian white noise of the estimated variance. Because of this, nonlinear time series models which are good for simulations as well as predictions start attracting people's attention in the traditional time series analysis in 1970's(Discussions to Campbell & Walker(1977) and Tong(1977), Ozaki & Oda(1978)). Here time series models which demonstrate some kind of nonlinear dynamic structure is analysts' main concern.

In this presentation we showed that the innovation approach can play an essential role in the assessment of the new animals such as chaos, neural network models and nonlinear differential equation models as well as of old animals such as linear and nonlinear time series models. These models are in the class of Markov models. One of the important nature of Markov models is that prediction errors are Gaussian when the trajectory of the process defined by the Markov model is continuous(Doob(1953)). This is the reason why innovation approach could still play an important role in the nonlinear time series modeling, where the data are usually sampled from continuous processes with sufficiently small time interval. We have seen that classic theories developed for Markov processes are useful in directing nonlinear time series modeling. Examples of nonlinear stochastic differential equation models successfully applied to real time series data and artificial data are also shown in the talk.

An application of the present idea to real financial time series analysis was presented. Here the time series is the currency exchange rate data between US Dollars and Japanese Yen. A combination of monitoring and control through an efficient dynamic model was shown to yield to a useful method in financial engineering similar to the PID control method in control engineering. A useful model for characterizing the market tendency of the price is the micro-market structure model (Bouchaud & Cont(1998), Iino & Ozaki(2000)). The model is given by the following stochastic differential equations:

$$d\phi = (\alpha_1 + \beta_1\phi)dt + \gamma_1 dw_1(t)$$
$$d\lambda = (\alpha_2 + \beta_2\lambda)dt + \gamma_2 dw_2(t)$$
$$dP = \phi\exp(\lambda)dt + \gamma_3 \exp(\lambda/2)dw_3(t)$$

Here P(t) is the price of the currency in the market. $dw_1(t)$, $dw_2(t)$, $dw_3(t)$ are increments of Brownian motion. The state variable $\phi(t)$ characterizes the state of the market, and whether the market is over-valued or under-valued: in other words, whether the currency (US Dollars) is over-valued or under-valued against the Japanese Yen. If $\phi(t) > 0$, it means the market is over-valued, and if $\phi(t) < 0$ the market is under-valued. $\exp(\lambda(t))$ represents the stiffness of the market, i.e. its inverse represents the liquidity of the market. If liquidity is large, stiffness is small, and the price of the market does not swing away much from the present price even though the currency is over-valued or under-valued. If we obtain good estimates of these state values from the past records of price data, it may provide us with useful information in controlling the currency allocations between US Dollars and Japanese Yen. Using this information we could design a rule for allocating currency between US Dollars and Yen. A model identification method was presented based on the innovation approach. Numerical results of the application of the identified model for allocation of the currencies between US Dollars and Japanese Yen was presented, where it was shown that significant gain was obtained even with very a simple allocation strategy based on the filtered estimate of the function $\phi(t)$.

マルコフ連鎖モンテカルロ法の最近の展開

大森裕浩*

1 マルコフ連鎖モンテカルロ (Markov Chain Monte Carlo) 法

1.1 データ拡大法/代替サンプリング

Tanner and Wong (1987) によって紹介された**データ拡大 (data augmentation) 法**は、**代替サンプリング (substitution sampling)** と基本的には同じもので、観測されるデータ Y を観測できないデータ Z によって拡大 (augment) させる手法である。Y と Z が与えられたとき事後分布 $\pi(\theta|Y,Z)$ が簡単に計算できるか、乱数の発生が簡単であることを仮定している。

1.2 ギブスサンプラー

未知の母数を $\theta=(\theta_1,\ldots,\theta_m)'$, 観測されたデータを Y とすると、事後分布 $\pi(\theta|Y)$ を求めることが難しいときに、他のパラメータを所与とした条件付き分布から反復的に θ を発生させることにより、事後分布からの標本を得るのが (systematic scan) **ギブスサンプリング (Gibbs sampling)** である。ギブスサンプラーを行うためには条件付分布から標本を簡単に発生できる必要があるが、それが困難な場合は Ritter and Tanner (1992) は格子点上で行う**格子点ギブスサンプラー (Griddy Gibbs sampler)** を提案した。Damien, Wakefield and Walker (1999) は条件付分布からのサンプリングが困難であっても潜在変数を導入することにより単純なサンプリングに帰着できる場合があることを示しているが、それ以外の場合では後述する棄却サンプリング連鎖 M-H アルゴリズム (A-R M-H) や Griddy Multiple-Try Metropolis アルゴリズムを用いることになる。

1.3 M-H (Metropolis-Hastings, メトロポリスーヘイスティングス) アルゴリズム

よく用いられる M-H アルゴリズムの 1 つに棄却サンプリング連鎖がある。$\pi(x)$ からの標本の発生が難しいためにすべての x について $\pi(x)<cg(x)$, (c は正の定数) を満たすような $g(x)$ を使って棄却サンプリングによって $\pi(x)$ からの標本の発生を行うものである。しかし、棄却サンプリングには $g(x)$ の近似精度の問題、上限 c の選択の問題がある。そこで、これを修正して $\pi(x)<cg(x)$ を仮定せず、 $f(x)=\min(\pi(x),cg(x))$ として独立連鎖を行えばよい。この他にも、 $\pi(x)$ が多峰分布であったり、状態空間が制約された空間であるときに有効な方法に Schmeiser and Chen (1991) による**ヒットエンドラン (hit-and-run) アルゴリズム**や、Liu, Liang and Wong (2000) による、点の周囲の情報も集めるために、点 x と y のそれぞれの周囲の点を k 個使って確率密度の比を比較する方法として **MTM (Multiple-Try Metropolis) アルゴリズム**などがある。

*東京大学大学院経済学研究科. 〒 113-0033 東京都文京区本郷 7-3-1 e-mail:omori@e.u-tokyo.ac.jp

2 マルコフ連鎖の収束判定の方法

実際にマルコフ連鎖モンテカルロ法を使うときには標本が均衡分布に収束するまでは、初期値に依存する期間であるとして棄て、それ以降の標本を用いて推論を行うことになる。その場合、反復を何回以上行えば初期値に依存せず、均衡分布に収束するのかという問題が生じる。収束に必要な反復回数を理論的に導く試みはなされているが、実用性に欠けているため現在の段階では、まず反復を行い得られた系列を用いてマルコフ連鎖が収束しているかどうかを検査するという方法がとられている。もちろん、Cowles, Roberts and Rosenthal (1999) で指摘されているようにどの方法も完全ではなく、いくつかの方法を併用しながら、収束の判定をしなくてはならない。収束の判定手法の包括的なサーベイはRobert and Casella (1999) や Mengersen, Robert and Guihenneuc-Jouyaux (1999) を、実験に基づく比較についてはCowles and Carlin (1996) や Brooks and Roberts (1999) を参照されたい。

3 モデルの選択

モデル選択の基準としては周辺尤度 (marginal likelihood) やその比であるベイズ比 (Bayes Factor) がよく使われる。Raftery (1996) によると重点的サンプリングの方法では g を注意深くとらなければ精度の悪い推定値になってしまうため、Chib (1995), Chib and Jeliazkov (2001) はそういった問題のない、周辺尤度の推定方法を提案した。一方、モデルのサンプリングも含める計算方法に Green (1995) による**リバーシブルジャンプ (reversible jump) アルゴリズム**がある。現在の状態が (M_i, θ_i) であるとき

(i) まずモデル選択のために提案された密度 $q_1(M_i, M_j)$ に従ってモデル M_j を選択する。
(ii) $i = j$ ならば通常の M-H アルゴリズムかギブスサンプラーで新しい θ_i を発生させる。
(iii) $i \neq j$ ならば、まず確率変数 u_i を提案された分布 $q_2(u_i|\theta_i, M_i, M_j)$ に従って発生させる。また g_{ij} を (θ_i, u_i) から (θ_j, u_j) への単射 (u_i, u_j は $\dim(\theta_i) + \dim(u_i) = \dim(\theta_j) + \dim(u_j)$ を満たす) としたとき、$(\theta_j, u_j) = g_{ij}(\theta_i, u_i)$ を確率

$$\min\left\{1, \frac{\pi(y|\theta_j, M_j)\pi(\theta_j|M_j)\pi(M_j)}{\pi(y|\theta_i, M_i)\pi(\theta_i|M_i)\pi(M_i)} \times \frac{q_1(M_j, M_i)q_2(u_j|\theta_j, M_j, M_i)}{q_1(M_i, M_j)q_2(u_i|\theta_i, M_i, M_j)} \times \left|\frac{\partial g_{ij}(\theta_i, u_i)}{\partial(\theta_i, u_i)}\right|\right\}$$

で採択する。

4 ソフトウェア

BUGS(Bayesian inference Using Gibbs Sampling) は、イギリスの Institute of Public Health の MRC Biostatistics Unit により開発されたシステムで (現在無料で提供されている) 事後分布の平均とその標準誤差、標準偏差、95%信用区間のほか、標本経路、事後密度関数のカーネル推定値、標本自己相関関数などを出力してくれる。詳しい収束の判定は得られた標本を一旦ファイルに出力してから **CODA** (Convergence Diagnostic and Output Analysis Software) や **BOA** (Bayesian Output Analysis) というプログラムで行う。BUGS は現在 1 次元の条件付分布からしか乱数を発生できないので、標本の自己相関が高くてブロック化を必要とするようなサンプリングはまだできない。

Asymptotics of Rank Order Statistics for ARCH Residual Empirical Processes

AJAY CHANDRA[1] AND MASANOBU TANIGUCHI[2]

Department of Mathematical Science
Graduate School of Engineering Science
Osaka University

December, 2001

1. Introduction

Classical time series models assume a constant one-period forecast variance. In order to overcome this implausible assumption, Engle (1982) introduced a class of ARCH(p) models, which allows the conditional forecast variance to change overtime as a function of the past values. Since then, ARCH related models have become perhaps the most popular and extensively studied financial econometric models (Engle (1995), Gouriéroux (1997), Chandra and Taniguchi (2001a,b)). Moreover, Giraitis *et. al* (2000) discussed a class of ARCH(∞) models, which includes that of ARCH(p) model as a special case, and established sufficient conditions for the existence of a stationary solution and its explicit representation. For ARCH(p) model, Horváth and Kokoszka (2001) derived the asymptotic distribution of the empirical process based on the squared residuals which is considered of fundamental importance for statistical analysis. Then they showed that, unlike the residuals of ARMA models, these residuals do not behave in this context like asymptotically independent random variables, and the asymptotic distribution involves a term depending on estimators of the volatility parameters of the model. Also Lee and Taniguchi (2000) proved the local asymptotic normality for ARCH(∞), and discussed the residual empirical process for ARCH(1) with stochastic mean.

In the i.i.d. settings, two-sample problem is one of the important statistical problems. For this problem, a class of rank order statistics plays a prominent role since it provides locally most powerful rank tests. The study of the asymptotic properties based on such rank order statistics is one of the most essential parts of nonparametric statistics. Many authors have contributed to its development, and numerous theorems have been formulated to show the asymptotic normality of a properly normalized rank order statistic in many testing problems. The classical limit theorem which generated much interest in this area is the celebrated Chernoff-Savage (1958) theorem. This theorem has been used to study the asymptotic power and power efficiency of a class of two-sample tests. Further refinements on the conditions of the theorem, extensions and related results, for example, are due to Hájek and Šidák (1967), Pyke and Shorack (1968) and Puri and Sen (1993). More specifically, the Chernoff-Savage theorem given in Puri and Sen (1993), is formulated under less stringent conditions on the score generating functions.

The present paper discusses the asymptotic theory of the two-sample rank order statistics $\{T_N\}$ for ARCH residual empirical processes based on the techniques of Puri and Sen (1993) and Horváth and Kokoszka (2001). Since the asymptotics of the residual empirical processes are different from those for the usual ARMA case, the limiting distribution of $\{T_N\}$ is greatly different from that of ARMA case (of course i.i.d. case). Based on the asymptotic

[1]chandra@sigmath.es.osaka-u.ac.jp
[2]taniguti@sigmath.es.osaka-u.ac.jp

results, we evaluate the asymptotic relative efficiencies, construct confidence intervals and study robustness for various residual densities in some two-sample problems. These studies illuminate some interesting characteristics of ARCH residuals in comparison with the i.i.d. settings.

Asymptotic expansion for hidden Markov models

Yuji Sakamoto*

Hiroshima International University

1 Hidden Markov model

Let us consider the following hidden Markov model:

$$X_t = X_0 + \int_0^t V_0(X_s, \theta)ds + \int_0^t V(X_s, \theta)dw_s, \tag{1}$$

$$Y_t = y_0 + \int_0^t S_0(X_s, \theta)ds + L_t, \tag{2}$$

where θ is an unknown parameter varing in a parameter space $\Theta \subset \mathbb{R}^p$ (the dimension of θ is p), V_0, V and S_0 are real-valued given functions on $\mathbb{R} \times \Theta$, $w = (w_t)_{t\in\mathbb{R}_+}$ is a one-dimensional standard Wiener process, $L = (L_t)_{t\in\mathbb{R}_+}$ is a one-dimensional Lévy process, X_0 is a random variable independent of w and L, and y_0 is a given constant. The distributions of X_0 and L may also depend on θ. Suppose that (i) X is strictly stationary and the moments of X_t up to any order exist for any $t \in \mathbb{R}_+$, (ii) w and L are independent. Note that the increment $L_{t+s} - L_t$ has the same distribution as L_s.

Suppose that we can observe $Z_j^\Delta = (Y_{\Delta j} - Y_{\Delta(j-1)})/\Delta$, $j = 1, \ldots, n$, for a fixed $\Delta > 0$, while X cannot be observed. In this article, we will consider the moment estimator $\hat{\theta}_n$ based on $Z^\Delta = (Z_j^\Delta)_{j=1}^n$.

Example 1. (*L is a Wiener process*) Let $L_t = \sqrt{\sigma}\tilde{w}_t$, where $\tilde{w} = (\tilde{w}_t)$ is a standard Wiener process.

Example 2. (*Normal inverse Gaussian Lévy process*) Let L_1 be distributed as a normal inverse Gaussian distribution $NIG(\alpha(\theta), \beta(\theta), \delta(\theta), \mu(\theta))$ for given functions $(\alpha, \beta, \delta, \mu) : \Theta \to \{0 \le |\beta| < \alpha, \delta > 0, \mu \in \mathbb{R}\}$.

Example 3. (*hidden Ornstein-Uhlenbeck diffusion*) Let $X = (X_t)_{t\in\mathbb{R}_+}$ be a stationay diffusion process satisfying $dX_t = p(q - X_t)dt + \sqrt{r}w_t$.

Example 4. (*hidden Cox-Ingersoll-Ross process*) Let $X = (X_t)_{t\in\mathbb{R}_+}$ be a stationay diffusion process satisfying $dX_t = p(q - X_t)dt + \sqrt{r}\sqrt{X_t}dw_t$.

For the model (1) and (2), the first order property of the moment estimator studied by Masuda(2001). For NIG-Lévy motion, see Barndorff-Nielsen(1998), and for the application of HMM, see Genon-Catalot, et.al.(2000).

2 Asymptotic expansion

For any lag set $\mathscr{K} = (k_j)_{j=1,\ldots,m}$, $k_j \in \mathbb{Z}_+$ †, let

$$\mu(\mathscr{K}; \theta) = E_\theta\left[(Z_1^\Delta - E_\theta[Z_1^\Delta])(Z_{1+k_1}^\Delta - E_\theta[Z_1^\Delta]) \cdots (Z_{1+k_1+\cdots+k_m}^\Delta - E_\theta[Z_1^\Delta])\right].$$

For the case where $\mathscr{K}$ is the empty set ϕ, we denote $E[Z_1^\Delta]$ by $\mu(\phi; \theta)$. In order to estimate θ, one may construct a estimating equation : $\psi_{n,j}(\mathscr{K}_j; \theta) = 0, \quad j = 1, \ldots, p$ for lag sets $(\mathscr{K}_j)_{j=1}^p$ such that $A_j = j$, $\mathscr{K}_j = \phi$, $j = 1, \ldots, d$. Here $\psi_{n,1}(\phi; \theta) = \sum_{i=1}^n (Z_i^\Delta - \mu(\phi; \theta))$, and

$$\psi_{n,j}(\mathscr{K}_j; \theta) = \sum_{i=1}^{n-\bar{k}_{j,m_j}} \left\{\prod_{l=0}^{m_j}(Z_{1+\bar{k}_{j,l}}^\Delta - \mu(\phi; \theta)) - \mu(\mathscr{K}_j; \theta)\right\}$$

for $j = 2, \ldots, p$, where $\mathscr{K}_j = (k_{j,l})_{l=1}^{m_j}$, $\bar{k}_{j,0} = 0$, and $\bar{k}_{j,l} = k_{j,1} + \cdots + k_{j,l}$. In general, it is known that under some regularity conditions, M-estimator $\hat{\theta}_n$ corresponding to the estimating function $\psi_n = (\psi_{n,a})_{a=1}^p$ admit

*Fax:81-823-70-4852. E-mail: y-sakamo@he.hirokoku-u.ac.jp

†$\mathbb{Z}_+$ denotes the set of non-negative integers.

the second order asymptotic expansion: for any measurable function f having at most polynomial growth order, there exist constants $C > 0$, $\epsilon > 0$, and a positive definite matrix $\hat{g} = (\hat{g}^{ab})_{a,b=1}^{p}$ such that

$$\left| E[f(\sqrt{n}(\hat{\theta}_n - \theta))] - \int_{\mathbb{R}^p} q_{n,1}(z)dz \right| \leq c\omega(f, Cn^{-(\epsilon+1)/2}, \hat{g}^{ab}) + o(n^{-1/2}) \tag{3}$$

where $q_{n,1}$ and ω are defined by

$$q_{n,1}(z) = \phi(z; g^{ab})\left(1 + \frac{1}{6\sqrt{n}} c^{abc} h_{abc}(z; g^{ab}) + \frac{1}{\sqrt{n}} \tilde{\mu}^{a;}{}_{cd} g^{cd} h_a(z; g^{ab})\right),$$

$$\omega(f, r, \hat{g}^{ab}) = \int_{\mathbb{R}^p} \sup\{|f(x+y) - f(x)| \ : \ |y| \geq r\}\phi(x; \hat{g}^{ab})dx,$$

for the nomal density $\phi(x; \Sigma^{ab})$ with mean 0 and covariance matrix $\Sigma = (\Sigma^{ab})_{a,b=1}^{p}$ and the first and third order Hermite polynomials $h_a(x; \Sigma^{ab})$ and $h_{abc}(x; \Sigma^{ab})$ generated by $\phi(x; \Sigma^{ab})$. The positive matrix $g = (g^{ab})_{a,b=1}^{p}$ and the coefficients c^{abc} and $\tilde{\mu}^{a;}{}_{bc}$ are defined in terms of the moments of ψ and its derivatives. For the rigorous derivation of the asymptotic expansion above, see Sakamoto-Yoshida(1998), which dealed with the expansion up to third order.

As for the moment estimators, the corresponding estimating function becomes $\psi_{n,a}(\mathcal{K}_a; \theta)$ defined for suitably choosen lag sets $\mathcal{K}_a$, and it is rather easy to show the validity of the asymptotic expansion.

Theorem 1. *Let Θ be a open convex subset of $\mathbb{R}^p$. Suppose that X is strictly stationary, has moments up to arbitrary order, and α-mixing with the mixing coefficient decreasing exponentially. Moreover, suppose that η is smooth, and that $\partial\eta/\partial\theta$ is non-singular. Then, if the distribution of L_1 has the smooth absolutely continuous part of the Lebesugue decomposition, the moment estimator $\hat{\theta}_n$ admits the second order asymptotic expansion given by (3). $\hat{\theta}_n^* = \hat{\theta}_n - \tilde{\mu}^{a;}{}_{cd}(\hat{\theta}_n) g^{cd}(\hat{\theta}_n)$ is a bias-corrected version of $\hat{\theta}_n$, i.e., then $E[\sqrt{n}(\hat{\theta}_n^* - \theta)] = o(n^{-1/2})$.*

Example 5. (*OU-diffusion & Weiner*) Let X and Y be processes considered in Example 3 and 1, respectively. Suppose that q, r and σ are unknown, p is known. In this case, the moment estimators $\hat{q}_n$, $\hat{r}_n$, $\hat{\sigma}_n$ are given by

$$\hat{q}_n = \bar{Z}_n^{\Delta}, \quad \hat{r}_n = 2\Delta^2 p^3 (1 - e^{-\Delta p})^{-2} \hat{\mu}(1), \quad \hat{\sigma}_n = \Delta\left(\hat{\mu}(0) + \frac{\hat{r}_n}{\Delta^2 p^3}(1 - \Delta p - e^{-\Delta p})\right).$$

Their bias-corrected versions are

$$\hat{q}_n^* = \hat{q}_n, \quad \hat{r}_n^* = \hat{r}_n + \frac{1}{n}\frac{2\,\Delta\,p}{(1 - e^{-\Delta\,p})^2}\left(\hat{r}_n + p^2\,\hat{\sigma}_n\right), \quad \hat{\sigma}_n^* = \hat{\sigma}_n + \frac{1}{n}\frac{(3 - 2\,\Delta\,p - 4e^{-\Delta\,p} + e^{-2\,\Delta\,p})}{(1 - e^{-\Delta\,p})^2\,p^2}\left(\hat{r}_n + p^2\,\hat{\sigma}_n\right).$$

Example 6. (*CIR-diffusion & NIG-Lévy*) Let X and Y be processes defined in Example 4 and 1, respectively. Suppose that q, σ are unkown, p, r is known and $\beta = \mu = 0$. Then the moment estimators and their bias-corrected versions are given by

$$\hat{q}_n = \bar{Z}_n^{\Delta}, \quad \hat{\sigma}_n = \Delta\left(\hat{\mu}(0) + \frac{\hat{q}_n r}{\Delta^2 p^3}(1 - \Delta p e^{-\Delta p})\right), \qquad \hat{q}_n^* = \hat{q}_n, \quad \hat{\sigma}_n^* = \hat{\sigma}_n + \frac{1}{n}\left(\hat{\sigma}_n + \frac{r}{p^2}\hat{q}_n\right).$$

References

[1] Barndorff-Nielsen, O.E.(1998): *Finance and Stochastics.*, **2** 41–68.

[2] Genon-Caralot, V., Jeantheau, T. and Laredo(2000): *Bernoulli.*, **6** 1051–1079.

[3] Masuda, H.(2001): *Master Thesis. Graduate School of Mathematical Science, University of Tokyo.*

[4] Sakamoto, Y., and Yoshida, N.(1998): *submitted.*

A Generalized SSAR Model and Predictive Distribution with an Application to VaR *

Naoto Kunitomo†
and
Seisho Sato ‡

December 2001

Summary

In our talk we have proposed to use a class of nonlinear time series models, which is a generalization of the SSAR (simultaneous switching autoregressive) models, and develop a new approach to deal with the asymmetrical sample paths of time series. Earlier, we have introduced the stationary and nonstationary SSAR time series models and discussed their statistical properties in some detail (Kunitomo and Sato (1996, 1999, 2000), Sato and Kunitomo (1996)). The SSAR models have been developed for applications in econometric analyses including the disequilibrium econometric models and the time series models with adjustments in financial markets. Although the SSAR models have been discussed in econometric applications, there are some interesting new aspects for statistical nonlinear time series modelling. We are trying to extend the SSAR models to a class of GSSAR models in time series analysis and discuss some possible applications. Since the most important application of time series models is prediction, we shall discuss some related applications based on the predictive distribution of the GSSAR models. In particular we shall propose a new estimation method of the percentile

*Our talk was based on the paper which was originally written for our presentation at the 3rd US-JAPAN conference on time series analysis held in June 2001 at Kyoto University under the title "A Generalized SSAR Model and its Applications". We thank Professor Genshiro Kitagawa for his useful discussions on our previous studies on the SSAR models, which lead us to work on the present study. The full text of this paper has been available as Discussion Paper CIRJE-F-122 from http://www.e.u-tokyo.ac.jp/cirje/research/dp/2001/list.htm.

†Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, JAPAN.

‡Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, JAPAN.

points of the predictive distribution such as the median when it is not necessarily symmetrical and the resulting volatility function of the time series can be also asymmetrical. It is essentially the same problem as the estimation of Value-at-Risk (VaR) in recent financial risk management. In the standard VaR procedure, however, it has been often assume that the underlying return process is a sequence of uncorrelated random variables. (See JP Morgan (1996) or Jorion (2000) for its details.)

There can be other approaches in the estimation problem of asymmetrical volatility function and VaR. In particular some non-linear statistical time series models along the line of Nelson (1991) and Harvey and Shephard (1996) in the econometric or financial analysis of the asymmetrical volatility functions have been known and there have been many related studies already appeared, which are closely related to the applications we shall investigate in this paper. However, our approach to the problem of estimating asymmetrical volatility function and measuring VaR is a simple but different one from the existing literatures in this respect.

We have introduced a generalized univariate SSAR (GSSAR) model and discuss some examples. Then we have investigated some properties of the GSSAR models including the geometric ergodicity and the estimation problems of the GSSAR models. Also we shall discuss the predictive distribution in the GSSAR models and the estimation problem of percentiles of the predictive distribution such as the median. As an application we shall discuss the prediction based on the median and the estimation of VaR in the financial risk management and give a case study on the interest rates futures market in Japan. The full text of ourtalk has been available as Discussion Paper CIRJE-F-122 from http://www.e.u-tokyo.ac.jp/cirje/research/dp/2001/list.htm.

Key Words

Asymmetrical Sample Paths, Generalized SSAR Model, Predictive Distribution, Predictive Median, Transformation Models, Value at Risk (VaR).

Large deviations limit theorems for the kernel spectral density estimator

北海道大学大学院経済学研究科　　柿沢　佳秀

1. はじめに　正規定常過程 $\{Y_t : t \in \mathbf{Z}\}$ は平均 0 で自己共分散関数 $\gamma(h) = E(Y_tY_{t+h}) = \gamma(-h)$ をもつとする．仮定 $\sum_{h=-\infty}^{\infty}|\gamma(h)| < \infty$ の下で連続なスペクトル

$$f(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma(h)\,\mathrm{e}^{-ih\lambda} = \frac{1}{2\pi}\gamma(0) + \frac{1}{\pi}\sum_{h=1}^{\infty}\gamma(h)\cos(h\lambda)$$

が定義される．観測系列 $Y_1,\ldots,Y_T$ に基づいたカーネル型スペクトル推定量

$$\begin{aligned}\widehat{f}_T(\alpha) &= \int_{\Pi} K_M(\alpha-\lambda)I_T(\lambda)\,d\lambda \\ &= \int_{\Pi}\frac{1}{2}\{K_M(\alpha-\lambda)+K_M(\alpha+\lambda)\}I_T(\lambda)\,d\lambda\end{aligned}$$

を考えよう．ここに $I_T(\lambda)$ はピリオドグラム

$$I_T(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T}Y_t\,\mathrm{e}^{it\lambda}\right|^2 = \frac{1}{2\pi T}\sum_{s,t=1}^{T}Y_sY_t\,\mathrm{e}^{i\lambda(s-t)},$$

$K_M(\lambda)$ は非負で原点対称な連続関数 $K(\theta)$ $(\theta\in\mathbf{R})$ に基づいたスペクトルウィンドウ

$$K_M(\lambda) = M\sum_{\nu=-\infty}^{\infty}K\{M(\lambda+2\pi\nu)\}$$

である．カーネル関数 $K(\theta)$ の台が $\Pi=[-\pi,\pi]$ であること，及び，$\int_{\Pi}K(\nu)\,d\nu = 1$ を仮定する．また $T\to\infty$ のとき $M = M(T)$ は $1/M + M/T \to 0$ を満たすとする．$\widehat{f}_T(\alpha)$ は原点対称で周期 2π をもつから，以後断ることなく $\alpha\in[0,\pi]$ と仮定する．

$\sqrt{T/M}\{\widehat{f}_T(\alpha) - f(\alpha)\}$ の漸近正規性は知られている．本報告では $\widehat{f}_T(\alpha)$ に関する裾確率を扱う．

2. 準備　(s,t) 要素が

$$\gamma(s-t) = \int_{\Pi}f(\lambda)\mathrm{e}^{i\lambda(s-t)}\,d\lambda$$

である $T\times T$ 対称行列を $\Sigma_{T,f}$ と書くことにすれば，カーネル型スペクトル推定量は $Y = (Y_1,\ldots,Y_T)' \sim N_T(\mathbf{0},\Sigma_{T,f})$ の 2 次形式であることに注意しよう:

$$\widehat{f}_T(\alpha) = \frac{M}{T}\,Y'Q_{T,\alpha}Y\,.$$

ここで $Q_{T,\alpha}$ は $T\times T$ 対称行列で，その (s,t) 要素は

$$q_{M,\alpha}(s-t) = \int_{\Pi}\frac{1}{2\pi}\widetilde{K}_M(\lambda;\alpha)\mathrm{e}^{i\lambda(s-t)}\,d\lambda\,,\quad \widetilde{K}_M(\lambda;\alpha) = \frac{1}{2M}\{K_M(\lambda-\alpha)+K_M(\lambda+\alpha)\}\,.$$

大偏差定理では積率母関数が必要である. 今の場合

$$\frac{M}{T}\log E(\mathrm{e}^{tY'Q_{T,\alpha}Y}) = -\frac{M}{2T}\log\det(I_T - 2t\Sigma_{T,f}Q_{T,\alpha})$$

の極限関数を $f(\lambda), K(\lambda)$ $(\lambda \in \mathbf{R})$ のリプシッツ連続性の下で求めることができる.

3. 結果

定理 1 $\alpha = \alpha(T) \in [0,\pi]$ が次の条件を満たすとする:

(C1) $\pi/M \le \alpha \le \pi - \pi/M$, かつ, $\lim_{T\to\infty}\alpha = \omega \in [0,\pi]$.

このとき, $0 < x < x_0$ に対して

$$\begin{aligned}
\lim_{T\to\infty}\frac{M}{T}\log P[\hat{f}_T(\alpha) - f(\alpha) > x] &= -\Gamma_{f(\omega)}(x)\,,\\
\lim_{T\to\infty}\frac{M}{T}\log P[\hat{f}_T(\alpha) - f(\alpha) < -x] &= -\Gamma_{f(\omega)}(-x)\,,\\
\lim_{T\to\infty}\frac{M}{T}\log P[|\hat{f}_T(\alpha) - f(\alpha)| > x] &= -\Gamma_{f(\omega)}(x)\,.
\end{aligned}$$

なお, 同様の結果が

(C2) $0 \le \alpha \le \pi/M$, かつ, $\lim_{T\to\infty}\alpha M = \omega \in [0,\pi]$

あるいは

(C3) $\pi - \pi/M \le \alpha \le \pi$, かつ, $\lim_{T\to\infty}(\pi-\alpha)M = \omega \in [0,\pi]$

に対しても示される.

定理 1 から $||\hat{f}_T - f||_\infty = \sup_{\lambda\in[0,\pi]}|\hat{f}_T(\lambda) - f(\lambda)|$ の裾確率を評価することもできる.

定理 2 $0 < x < x_0'$ に対して

$$\lim_{T\to\infty}\frac{M}{T}\log P[||\hat{f}_T - f||_\infty > x] = -\Lambda(x)\,.$$

$\Lambda(x)$ は定理 1 の右辺で得られた (C1),(C2),(C3) の関数の最小値として定義される. なお $x \searrow 0$ における挙動は次のようである.

系 3 $C_U = \max_{x\in[0,\pi]} f(x)$ が (i) $\max\{f(0), f(\pi)\} < C_U/\sqrt{2}$ を満たすとき

$$\lim_{x\searrow 0}\lim_{T\to\infty}\frac{M}{Tx^2}\log P[||\hat{f}_T - f||_\infty > x] = -\frac{1}{4\pi C_U^2\int_\Pi\{K(\nu)\}^2\,d\nu}\,,$$

(ii) $C_U/\sqrt{2} \le \max\{f(0), f(\pi)\} \le C_U$ を満たすとき

$$\lim_{x\searrow 0}\lim_{T\to\infty}\frac{M}{Tx^2}\log P[||\hat{f}_T - f||_\infty > x] = -\frac{1}{8\pi[\max\{f(0), f(\pi)\}]^2\int_\Pi\{K(\nu)\}^2\,d\nu}\,.$$

On semiparametric estimation of a fractional difference parameter

F.J.Hidalgo(London School of Economics)
Yoshihiro Yajima(University of Tokyo)

We study two estimators of d in the model $f(\lambda) \sim C\lambda^{-2d}$ as $\lambda \to 0+$, where $f(\lambda)$ denotes the spectral density of a covariance stationary linear process x_t. We show that one of the estimators achieves the optimal semiparametric rate of convergence, whereas the other has a rate of convergence as close as desired to the optimal rate. Moreover , we show that the estimators are asymptotically normal with a variance , which does not depend on any unknown parameter.

Define the discrete Fourier transform and periodogram of x_t by $a(\lambda) = (2\pi n)^{-1/2} \sum_{t=1}^{n} x_t e^{it\lambda}$ and $I(\lambda) = |\alpha(\lambda)|^2$. Then define $\hat{f}_p$ by

$$\hat{f}_p = \hat{f}(\lambda_p) = \begin{cases} \frac{1}{m+1} \sum_{j=-m/2}^{m/2} I_{j+p} & \text{if } m < 2p \\ \frac{2}{m} \sum_{j=1}^{m/2} I_{j+p} & \text{if } 0 < 2p \leq m \end{cases}$$

where $\lambda_p = \frac{2\pi p}{n}, p = 1, \ldots, n-1$ is a Fourier frequency and $I_j = I(\lambda_j)$.

Then first we define $\tilde{d}$ by

$$\tilde{d} = h_w \left(\frac{1}{k} \sum_{p=1}^{k} w_p \log \hat{f}_p - \left(\frac{1}{k} \sum_{p=1}^{k} w_p \right) \log \hat{f}_{k+1} \right),$$

where $w(u)$ with $u \in (0,1)$ is a positive weight function and $h_w = (-2 \int_0^1 w(u) \log(u) du)^{-1}$ and $w_p = w(p/k)$.

Next let

$$b(d) = h_w \left(\frac{1}{k} \sum_{p=1}^{k} w_p \log \left(\frac{\bar{f}_p}{f_p^*} \right) - \left(\frac{1}{k} \sum_{p=1}^{k} w_p \right) \log \left(\frac{\bar{f}_{k+1}}{f_{k+1}^*} \right) \right),$$

where

$$\bar{f}_p = \bar{f}(\lambda_p) = \begin{cases} \frac{C}{m+1} \sum_{j=-m/2}^{m/2} \lambda_{j+p}^{-2d} & \text{if } m < 2p \\ \frac{2C}{m} \sum_{j=1}^{m/2} \lambda_{j+p}^{-2d} & \text{if } 0 < 2p \leq m \end{cases}$$

and $f_p^* = C\lambda_p^{-2d}$.

Finally we difine the first estimator by

$$\hat{d} = \tilde{d} - b(\bar{d}),$$

where $\bar{d}$ is a preliminary estimator of d , say $\tilde{d}$ but with a different bandwidth number $k^{1/2}$ instead of m for $\hat{f}_p$. And we define the second estimator by

$$\hat{d}^* = \bar{v}^{-1} \frac{1}{m} \sum v_p \tilde{d}_p$$

where

$$\tilde{d}_p = h_w \left(\frac{1}{p} \sum_{l=1}^{p} w_l \log \hat{f}_l - \left(\frac{1}{p} \sum_{l=1}^{p} \right) \log \hat{f}_{p+1} \right),$$

and $w(l/p)$, $v_p = v(p/m)$, $\bar{v} = m^{-1} \sum_{p=1}^{m} v_p$ and the bandwidth of $\hat{f}_p$ is m_1 insted of m.

3.Theoretical results

Theorem 1 *Under some assumptions, as* $n \to \infty$, $m^{1/2}(\hat{d} - d)$ *converges to* $N(0, h_w^2 \Phi_w^2)$ *in distribution where* $\Phi_w = \int_0^1 w(u)du$.

Theorem 2 *Under some assumptions, as* $n \to \infty$, $m^{1/2}(\hat{d}^* - d)$ *converges to* $N(0, h_w^2 (\int_0^1 v(x)dx)^{-2} \Phi^2)$ *in distribution where* $\Phi^2 = \int_0^1 ((v(u) \int_0^1 w(x)dx) - \int_u^1 x^{-1} w(u/x) v(x)dx)^2 du$.

$\hat{d}^*$ achieves the optimal semiparametric rate of convergence, whereas $\hat{d}$ has a rate of convergence as close as desired to the optimal rate.