UTMS 2007–20 October 15, 2007

Bootstrap for continuous-time processes

by

Masaaki Fukasawa

UNIVERSITY OF TOKYO

GRADUATE SCHOOL OF MATHEMATICAL SCIENCES

KOMABA, TOKYO, JAPAN

BOOTSTRAP FOR CONTINUOUS-TIME PROCESSES

MASAAKI FUKASAWA

Abstract. An Edgeworth expansion of a Studentized statistic for an ergodic regenerative strong Markov process is validated. A specific nonparametric bootstrap method is proposed and proved to be second-order correct in the light of the Edgeworth expansion, which is a variant of the regenerative block bootstrap designed for discrete-time Markov processes. One-dimensional diffusions and semi-Markov processes are treated as examples.

1. Introduction

An Edgeworth expansion (EE) for an ergodic continuous-time process was first validated by Yoshida [9] as a refinement of the martingale central limit theorem. Kusuoka and Yoshida [8] and Yoshida [10] extended it by a mixing-based approach, and Fukasawa [6] also did it by the regenerative method. As in the iid case, one may expect to obtain, for instance, second-order correct confidence intervals of estimators by utilizing the EE; however, in nonparametric contexts, no EE result has been available for Studentized statistics, which hampers the practical use of the EE theory for continuous-time processes. The first half of this paper extends the argument of Fukasawa [6] to validate the EE of a Studentized statistic, which is a counterpart of the result of Bertail and Clémençon [1] for Markov chains. Let $X = \{X_t\}$ be a regenerative strong Markov process with stationary distribution μ. Suppose that a path X_t, $0 \leq t \leq T$ is completely observed. We are interested in the asymptotic property of an estimator of type

$$\hat{\theta}_T = \frac{1}{T}\int_0^T f(X_t)dt$$

for $\theta = \mu[f]$ where f is a μ-integrable function. A nonparametric estimator $\hat{\sigma}_T$ for the asymptotic variance is proposed and the first-order EE

$$P[\sqrt{T}(\hat{\theta}_T - \theta)/\hat{\sigma}_T \leq z] = \Phi(z) + T^{-1/2}\phi(z)(a_1 + a_2(2z^2+1)) + O(T^{-1})$$

is validated, where Φ and ϕ are the distribution function and the density of the standard normal distribution respectively. Exploiting the above expansion, the latter half of this paper proves the second-order correctness of a variant of the regenerative block bootstrap proposed by Bertail and Clémençon [2, 3] for Markov chains. One-dimensional diffusions and semi-Markov processes are treated as examples. Seemingly, our bootstrap is the only resampling method so far proved to attain second-order correctness in the context of nonparametric inference for continuous-time processes.

Graduate School of Mathematical Science, University of Tokyo, JSPS research fellow [DC1].

2. Edgeworth expansion

2.1. **Regenerative method.** Here we give a rigorous formulation.

Definition 1. *A càdlàg process X is said to be regenerative if there exists an increasing sequence of finite random times $\{\tau_j\}_{j\geq 1}$ such that*

$$\{X_t\}_{0\leq t\leq\tau_1},\ \{X_t\}_{\tau_1\leq t\leq\tau_2},\ \cdots,\{X_t\}_{\tau_j\leq t\leq\tau_{j+1}},\ \cdots$$

are independent and

$$\{X_t\}_{\tau_1\leq t\leq\tau_2},\ \cdots,\{X_t\}_{\tau_j\leq t\leq\tau_{j+1}},\cdots$$

are identically distributed. The random time τ_j is called a j-th regenerative epoch.

We suppose that X is a regenerative strong Markov process with state space E. Denote by P_ν the probability or the expectation operator with respect to the initial condition $X_0\sim\nu$ for a given distribution ν on E. Let P_x stand for P_{δ_x} for $x\in E$. We suppose also that the regenerative epochs are given as

$$\tau_{j+1}=\inf\{t>\tau_j; X_t=x,\ \text{there exists } s\in(\tau_j,t) \text{ such that } X_s\in\hat{x}\}$$

where $\tau_0=0$ and x, $\hat{x}$ are a point and a closed set of E respectively such that $x\notin\hat{x}$. In particular, $\{\tau_j\}$ is a sequence of stopping times with respect to the canonical filtration of X and $X_{\tau_j}=x$ for all $j\geq 1$. Here the set $\hat{x}$ was introduced to assure $\tau_{j+1}>\tau_j$ a.s.. Since $L_j:=\tau_{j+1}-\tau_j$, $j\geq 1$ is an iid sequence, it holds that $\tau_j\to\infty$ a.s. as $j\to\infty$. The primary use of the regenerative method appears in a proof of the consistency of $\hat{\theta}_T$ as follows; by the law of large numbers for iid sequences, we have

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f(X_t)dt=\lim_{N\to\infty}\frac{1}{\tau_N-\tau_1}\sum_{j=1}^{N-1}F_j=\frac{P_x[F_0]}{P_x[L_0]},$$

where

$$F_j=\int_{\tau_j}^{\tau_{j+1}} f(X_t)dt,\quad L_j=\tau_{j+1}-\tau_j \tag{1}$$

for $j\geq 0$. In order to check $\theta=P_x[F_0]/P_x[L_0]$, notice that the above convergence holds under any initial condition $X_0\sim\nu$ and that

$$P_\mu\left[\int_0^T f(X_t)dt\right]=T\mu[f] \tag{2}$$

for the stationary distribution μ. Hereafter we assume f not to be constant. It is also possible to exploit this regenerative argument in order to obtain the corresponding central limit theorem and EE;

Theorem 1. *Assume that there exists $p\geq 1$ such that the characteristic function of (F_0,L_0) under P_x is in $L^p(\mathbb{R}^2)$ and that*

$$P_\nu[F_0^2]+P_\nu[L_0^2]+P_x[F_0^4]+P_x[L_0^4]+P_x\left[\int_0^{\tau_1}\left|\int_0^t f(X_s)ds\right|^2 ds\right]<\infty. \tag{3}$$

Then, it holds

$$P_\nu[\sqrt{T}(\hat{\theta}_T-\theta)/\sigma\leq z]=\Phi(z)+T^{-1/2}\phi(z)(a_1+a_2(1-z^2))+O(T^{-1})$$

uniformly in $z \in \mathbb{R}$*, where*

$$\sigma^2 = P_x[\bar{F}_0^2]/\alpha, \quad \bar{F}_0 = F_0 - \theta L_0, \quad \alpha = P_x[L_0],$$

$$a_1 = P_\mu[\bar{F}_0] - P_\nu[\bar{F}_0], \quad a_2 = \frac{\kappa - 3\rho\sigma^2}{6\alpha\sigma^3}, \quad \kappa = P_x[\bar{F}_0^3], \quad \rho = P_x[\bar{F}_0 L_0].$$

Proof. If $\hat{x} = \{y\}$ for $y \in E$, the result is a special case of Theorem 1 of Fukasawa [6]. The general case is treated in the same way. □

2.2. **Studentized statistics.** The asymptotic variance σ defined in the preceding subsection is practically unknown in the nonparametric context. Hence we have to construct an estimator for it when constructing confidence intervals for instance. Here we propose an estimator which is a counterpart of $\sigma_n(f)$ in Bertail and Clémençon [1]. Let $M_T = \max\{j; \tau_{j+1} \leq T\}$ and

$$\hat{\sigma}_T^2 = \frac{\sum_{j=1}^{M_T} \left|F_j - \check{\theta}_T L_j\right|^2}{\sum_{j=1}^{M_T} L_j}, \quad \check{\theta}_T = \frac{\sum_{j=1}^{M_T} F_j}{\sum_{j=1}^{M_T} L_j}. \tag{4}$$

By the law of large numbers, we have $\check{\theta}_T \to \theta$ and $\hat{\sigma}_T \to \sigma$ a.s. as $T \to \infty$ since $M_T \to \infty$. Because of the regeneration-based construction, we can prove that the estimator $\hat{\sigma}_T$ admits the following EE;

Theorem 2. *Assume that there exists* $p \geq 1$ *such that the characteristic function of* (F_0, L_0) *under* P_x *is in* $L^p(\mathbb{R}^2)$ *and that*

$$P_\nu[F_0^2] + P_\nu[L_0^2] + P_x[F_0^{12}] + P_x[L_0^{12}] + P_x\left[\int_0^{\tau_1} \left|\int_0^t f(X_s)ds\right|^2 ds\right] < \infty. \tag{5}$$

Then, it holds

$$P_\nu[\sqrt{T}(\hat{\theta}_T - \theta)/\hat{\sigma}_T \leq z] = \Phi(z) + T^{-1/2}\phi(z)(a_1 + a_2(2z^2+1)) + O(T^{-1})$$

uniformly in $z \in \mathbb{R}$*, where* a_1 *and* a_2 *are defined in Theorem 1.*

Proof. See Section 4. □

The above two theorems correspond to Theorem 5.1 of Bertail and Clémençon [1]. It is noteworthy that we have a remainder term of $O(T^{-1})$ in Theorem 2, while the corresponding term is of $O(n^{-1}\log(n))$ in Theorem 5.1 of Bertail and Clémençon [1] where n corresponds to T. The reason why we can obtain such a better accuracy is that the regenerative block (F_j, L_j) has a bounded density in our continuous-time setting, so that we can exploit a more powerful result on the asymptotic expansion of iid sequence. See Section 4 for detail.

3. Continuous-time Regenerative Block Bootstrap

In this section, we assume X to be stationary, that is, $\nu = \mu$. Then, it holds $a_1 = 0$, so that we have

$$P_\mu[\sqrt{T}(\hat{\theta}_T - \theta)/\hat{\sigma}_T \leq z] = \Phi(z) + T^{-1/2}\phi(z)\hat{a}_2(2z^2+1) + O_p(T^{-1}),$$

where $\hat{a}_2$ is an arbitrary estimator for a_2 with $\sqrt{T}(\hat{a}_2 - a_2) = O(1)$. We can use for instance

$$\hat{a}_2 = \frac{\hat{\kappa} - 3\hat{\rho}\hat{\sigma}^2}{6\hat{\alpha}\hat{\sigma}^3}, \quad \hat{\alpha} = \frac{1}{M_T}\sum_{j=1}^{M_T} L_j, \quad (\hat{\kappa}, \hat{\rho}, \hat{\sigma}^2) = \frac{1}{M_T}\sum_{j=1}^{M_T}\left(\check{F}_j^3, \check{F}_j L_j, \check{F}_j^2/\hat{\alpha}\right),$$

where $\check{F}_j = F_j - \check{\theta}_T L_j$. In fact, since

$$P_\mu[|T - \alpha M_T| \geq \delta T] = O(T^{-1})$$

for $\delta \in (0, 1/2)$, using Kolmogorov's inequality, we have

$$\sup_{T>0} P_\mu \left[\left| \frac{\sqrt{T}}{M_T} \sum_{j=1}^{M_T} \{(F_j, L_j)^n - P_\mu[(F_1, L_1)^n]\} \right| > K \right] \to 0$$

as $K \to \infty$, where $n \in \mathbb{Z}_+^2$ with $|n| \leq 3$. Hence, the above expansion formula is practically of use to obtain second-order correct confidence intervals for instance by means of the Cornish-Fisher expansion. For the same purpose, it is then natural to expect that there corresponds a bootstrap method. Let $\mathcal{F}_T = \{(F_j, L_j)\}_{j=1,2,\dots,M_T}$ be the set of the observed regenerative blocks. Let (F_j^*, L_j^*), $j = 1, 2, \dots, M_T$ be an iid sequence and each (F_j^*, L_j^*) be uniformly distributed on $\mathcal{F}_T$. Here M_T and $\mathcal{F}_T$ are fixed conditionally to the observation $\{X_t\}_{0\leq t\leq T}$. Then, define bootstrap statistics $\check{\theta}_T^*$ and $\hat{\sigma}_T^*$ as

$$\check{\theta}_T^* = \frac{\sum_{j=1}^{M_T} F_j^*}{\sum_{j=1}^{M_T} L_j^*}, \quad \{\hat{\sigma}_T^*\}^2 = \frac{\sum_{j=1}^{M_T} \left|F_j^* - \check{\theta}_T^* L_j^*\right|^2}{\sum_{j=1}^{M_T} L_j^*}.$$

Further, put

$$T^* = \sum_{j=1}^{M_T} L_j^*.$$

Theorem 3. *Assume that there exists $p \geq 1$ such that the characteristic function of (F_0, L_0) under P_x is in $L^p(\mathbb{R}^2)$ and that (F_0, L_0) has finite moments of any order under P_x. Then,*

$$P_\mu^*[\sqrt{T^*}(\check{\theta}_T^* - \check{\theta}_T)/\hat{\sigma}_T^* \leq z] = \Phi(z) + T^{-1/2}\phi(z)(\hat{b} + \hat{a}_2(2z^2+1)) + O_p(T^{-1})$$

uniformly in $z \in \mathbb{R}$, where P_μ^* is the conditional probability given $\{X_t\}_{0\leq t\leq T}$ and $\hat{b} = \hat{\rho}/(2\hat{\alpha}\hat{\sigma})$. In particular,

$$P_\mu[\sqrt{T}(\theta_T - \theta)/\hat{\sigma}_T \leq z] = P_\mu^*[\sqrt{T^*}(\check{\theta}_T^* - \check{\theta}_T)/\hat{\sigma}_T^* - \hat{b}T^{-1/2} \leq z] + O_p(T^{-1})$$

uniformly in $z \in \mathbb{R}$.

Proof. See Section 4. □

Example 1. Consider X to be given as a weak solution of the one-dimensional stochastic differential equation

$$dX_t = b(X_t)dt + c(X_t)dB_t, \quad X_0 \sim \mu,$$

where $B = \{B_t\}$ is a standard Brownian motion. Suppose that b is a locally integrable and that c and $1/c$ are locally bounded on $\mathbb{R}$. Assume that f is continuous on an interval I of $E = \mathbb{R}$ and that f is not constant on I. Applying Theorem 2 of Fukasawa [6], we conclude that the assumptions of Theorem 3 are satisfied if

$$\limsup_{|z|\to\infty} \frac{zb(z)}{c(z)^2} = -\infty$$

and both f and $1/c$ are of polynomial growth.

Example 2. Let (Y_j, T_j), $j = 0, 1, \dots$ be a Markov renewal process and $Z = \{Z_t\}_{t \geq 0}$ be the corresponding semi-Markov process with stationary distribution $\tilde{\mu}$. It holds by definition that

$$Z_t = Y_{N_t}, \quad N_t = \max\{j \geq 0; T_j \leq t\}.$$

The process $X_t = (Z_t, t - T_{N_t})$ is a Markov process with state space $E = S \times [0, \infty)$, where S is the state space of the embedded Markov chain $Y = \{Y_j\}$. Suppose that S is countable and that there exist $p(s; y_0, y_1)$ for $s \geq 0, y_0, y_1 \in S$ and $K > 0$ such that

$$P[T_{j+1} - T_j \leq t | Y_j = y_0, Y_{j+1} = y_1] = \int_0^t p(s; y_0, y_1) ds$$

and $p(s; y_0, y_1) \leq K$ for all $s \geq 0, y_0, y_1 \in S$. Suppose also that

$$\sup_{y_0, y_1 \in S} \int_0^\infty s^n p(s, y_0, y_1) ds < \infty$$

for each $n \in \mathbb{N}$. Consider an estimator of type

$$\hat{\theta}_T = \frac{1}{T} \int_0^T \tilde{f}(Z_t) dt$$

for $\theta = \tilde{\mu}[\tilde{f}]$, where $\tilde{f}$ is a bounded function on S. Then, it can be proved that $X = \{X_t\}$ is a strong Markov process and the assumptions of Theorem 3 hold provided that a recurrent time of Y has moments of any order.

4. Proof of theorems

4.1. **Proof of Theorem 2.** Consider the random vector

$$U_j = \left(\bar{F}_j, L_j - \alpha, \bar{F}_j^2 - \alpha\sigma^2, \bar{F}_j L_j - \rho, L_j^2 - P_x[L_0^2]\right)$$

where $\bar{F}_j = F_j - \theta L_j$. Putting

$$\bar{U}_n = \frac{1}{n} \sum_{j=1}^n U_j$$

and

$$\Sigma(x_1, \dots, x_5)^2 = \frac{x_3(x_2 + \alpha)^2 - 2x_1(x_2 + \alpha)(x_4 + \rho) + x_1^2(x_5 + P_x[L_0^2])}{(x_2 + \alpha)^3},$$

we have $\hat{\sigma}_T = \Sigma(\bar{U}_{M_T})$. Moreover an equivalence

$$\sqrt{T}(\hat{\theta}_T - \theta)/\hat{\sigma}_T \leq z \Leftrightarrow \sqrt{M_T} A^{K,H}(\bar{U}_{M_T}) \leq \hat{z}$$

holds, where

$$A^{k,h}(x_1, \dots, x_5) = \frac{\sigma k x_1 + h(\sigma - \Sigma(x_1, \dots, x_5))}{k\sigma\Sigma(x_1, \dots, x_5)}, \quad K = \frac{M_T}{T}, \quad H = \frac{\bar{F}_0 + R_T}{T}$$

and

$$R_T = \int_0^T f(X_t) dt - \theta T - \sum_{j=0}^{M_T} \bar{F}_j, \quad \hat{z} = \frac{z - \sqrt{T} H/\sigma}{\sqrt{K}}.$$

Lemma 1. *For sufficiently large m, $\bar{U}_m$ has a bounded density.*

Proof. By Theorem 19.1 of Bhattacharya and Rao [5], $\sum_{j=1}^n (\bar{F}_j, L_j)$ has a bounded density for sufficiently large n. Then, it is straightforward to see that $\bar{U}_m$ has a density for $m = 3n$ and that there exists $r > 1$ such that the density is in $L^r(\mathbb{R}^5)$. The lemma then obtained by using the discussion given after the proof of Theorem 19.1 of Bhattacharya and Rao [5]. □

Now, we make a similar argument to the proof of Theorem 1 of Fukasawa [6]. For $\delta \in (0, 1/2)$,

$$P_\nu[\sqrt{T}(\hat{\theta}_T - \theta)/\hat{\sigma}_T \le z] = \sum_{m; |T-\alpha m| < \delta T} P_\nu[\sqrt{m}A^{K,H}(\bar{U}_m) \le \hat{z}; M_T = m] + O(T^{-1}).$$

By Lemma 1, $\sqrt{m}\bar{U}_m$ has a bounded density $p_m(u) = p_m(u_1, u_2, u_3, u_4, u_5)$. Hence, exploiting the strong Markov property, we have

$$\begin{aligned} &P_\nu[\sqrt{m}A^{K,H}(\bar{U}_m) \le \hat{z}; M_T = m] \\ &\qquad = \int \psi_m(u, f, l, r, t) P_x^{(\hat{R}_m(l,u_2),\tau_1)}(dr, dt) p_m(u) du P_\nu^{(F_0, L_0)}(df, dl), \end{aligned}$$

where $\psi_m(u, f, l, r, t)$ is the indicator function of the set

$$\left\{(u, f, l, r, t); \sqrt{m}A^{k(m),h(f,r)}(u/\sqrt{m}) \le \hat{z}_m(f, r), 0 \le \sqrt{m}(a_m - u_2) - l < t\right\},$$

$$a_m = \frac{T - \alpha m}{\sqrt{m}}, \quad k(m) = \frac{m}{T}, \quad h(f, r) = \frac{f + r}{T}, \quad \hat{z}_m(f, r) = \frac{z - \sqrt{T}h(f, r)/\sigma}{\sqrt{k(m)}},$$

and

$$\hat{R}_m(l, u_2) = \int_0^{T-l-\sqrt{m}u_2-\alpha m} f(X_s) ds. \tag{6}$$

Applying Theorem 19.2 of Bhattacharya and Rao [5], we have

$$\sup_{u \in \mathbb{R}^5} (1 + |u|^6) \left| p_m(u) - \phi_V(u) \left\{1 + \sum_{i=1}^4 m^{-i/2} p_k^V(u)\right\}\right| = o(m^{-2}),$$

where V is the covariance matrix of U_1, ϕ_V is the normal density with mean 0 and covariance V, and p_k^V are polynomials. We have then,

$$\begin{aligned} &P_\nu[\sqrt{T}(\hat{\theta}_T - \theta)/\hat{\sigma}_T \le z] \\ &\quad = \sum_{m; |T-\alpha m| < \delta T} \int \psi_m(u, f, l, r, t) \phi_V(u) \left\{1 + \sum_{i=1}^4 m^{-i/2} p_k^V(u)\right\} \\ &\qquad\qquad P_x^{(\hat{R}_m(l,u_2),\tau_1)}(dr, dt) du P_\nu^{(F_0, L_0)}(df, dl) + O(T^{-1}) \end{aligned}$$

by the same calculation as in Fukasawa [6].

Lemma 2. *For $(h, k) \in \mathbb{R} \times (0, \infty)$, there exists a sequence of polynomials $q_i^{h,k}$, $i = 1, 2, 3, 4$ such that*

$$\begin{aligned} &P_\nu\left[\sqrt{m}A^{k,h}(\bar{U}_m) \in S_1, \frac{1}{\sqrt{m}} \sum_{j=1}^m (L_j - \alpha) \in S_2\right] \\ &\quad = \int_{S_1 \times S_2} \phi_D(\xi, \eta) + \sum_{i=1}^4 m^{-i/2} q_i^{h,k}(\xi, \eta) \phi_D(\xi, \eta) d\xi d\eta + o(m^{-2}) \end{aligned}$$

uniformly in Borel sets $S_1, S_2 \subset \mathbb{R}$, and uniformly in (h,k) on compact sets of $\mathbb{R}\times(0,\infty)$, where Φ_D and ϕ_D are the distribution function and the density of the normal distribution with mean 0, covariance

$$D = D^{h,k} = c_1^{h,k}h + \begin{pmatrix} \alpha & \rho/\alpha \\ \rho/\alpha & P_x[|L_0-\alpha|^2] \end{pmatrix},$$

where $c_1^{h,k}$ is a matrix which is bounded in (h,k) on compact sets of $\mathbb{R}\times(0,\infty)$. Moreover, we have

$$\begin{aligned} q_{h,1}(\xi,\eta)\phi_D(\xi,\eta) &= (3a_2 + c_2^{h,k}h)\partial_\xi\phi_D(\xi,\eta) + (2\alpha a_2 - \rho/(2\sigma) + c_3^{h,k}h)\partial_\xi^3\phi_D(\xi,\eta) \\ &\quad + c_4^{h,k}\partial_\xi^2\partial_\eta\phi_D(\xi,\eta) + c_5^{h,k}\partial_\xi\partial_\eta^2\phi_D(\xi,\eta) + c_6^{h,k}\partial_\eta^3\phi_D(\xi,\eta), \end{aligned}$$

where $c_i^{h,k}$, $i = 2,\dots,6$ are constants which are bounded in (h,k) on compact sets of $\mathbb{R}\times(0,\infty)$. Besides, $c_i^{0,k}$, $i=1,\dots,6$ and $q_i^{0,k}$, $i=1,\dots,4$ do not depend on k.

Proof. Use the same argument as the proof of Theorem 2.1 of Bhattacharya and Ghosh [4]. The uniformity in (h,k) follows from the fact that $\bar U_m$ does not depend on (h,k) and $A^{k,h}$ does continuously. The expression of constants follows from

$$\mathrm{Cov}\left(\sqrt{m}A^{k,h}(\bar U_m), \frac{1}{\sqrt{m}}\sum_{k=1}^m (L_j-\alpha)\right) = D + O(m^{-1})$$

and

$$\begin{aligned} P_\nu[\sqrt{m}A^{k,h}(\bar U_m)] &= -\frac{1}{\sqrt{m}}(3a_2 + c_2^{h,k}h) + O(m^{-1}), \\ P_\nu[\{\sqrt{m}A^{k,h}(\bar U_m)\}^3] &= \frac{1}{\sqrt{m}}\left\{\frac{27\rho}{2\sigma} - \frac{7\kappa}{2\sigma^3} + c_7^{h,k}h\right\} + O(m^{-1}), \end{aligned}$$

where $c_7^{h,k}$ is a constant uniformly bounded in (h,k) on compact sets of $\mathbb{R}\times(0,\infty)$. □

Now, changing variable as

$$(\xi,\eta,\zeta_1,\zeta_2,\zeta_3) = (\sqrt{m}A^{k(m),h(f,r)}(u/\sqrt{m}), u_2, u_3, u_4, u_5),$$

using Taylor's expansion, and integrating in $(\zeta_1,\zeta_2,\zeta_3)$, we have

$$\begin{aligned} &\int \psi_m(u,f,l,r,t)1_{|h|\le h_0}\phi_V(u)\left\{1 + \sum_{i=1}^4 m^{-i/2}p_k^V(u)\right\} P_x^{(\hat R_m(l,u_2),\tau_1)}(dr,dt)du \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\hat z}\int \psi_{m,2}(\eta,l,t)1_{|h|\le h_0}\phi_{\tilde D}(\xi,\eta)\left\{1 + \sum_{i=1}^4 m^{-i/2}\tilde q_i(\xi,\eta)\right\} \\ &\qquad P_x^{(\hat R_m(l,\eta),\tau_1)}(dr,dt)d\xi d\eta + o(m^{-2}) \end{aligned}$$

for any $h_0 > 0$, where $\tilde D$ is a matrix, $\tilde q_i$ are polynomials and $\psi_{m,2}$ is the indicator function of the set $\{(\eta,l,t); 0 \le \sqrt{m}(a_m-\eta) - l < t\}$. Note that $|T-\alpha m| < \delta T$ implies $(1-\delta)/\alpha < k(m) < (1+\delta)/\alpha$. In the light of Lemma 2, we can take

$$\tilde D = D^{h,k}, \quad \tilde q_i = q_i^{h,k}$$

for $k = k(m)$ and $h = h(f,r)$.

Using Taylor's expansion around $h = 0$, we have

$$
\begin{aligned}
&P_\nu[\sqrt{T}(\hat{\theta}_T - \theta)/\hat{\sigma}_T \leq z] \\
&= \sum_{m;|T-\alpha m|<\delta T} \int_{-\infty}^{\infty}\int_{-\infty}^{\hat{z}} \psi_{m,2}(\eta,l,t)\phi_{D^0}(\xi,\eta)\left\{1+\sum_{i=1}^{4} m^{-i/2}q_i^0(\xi,\eta)\right\} \\
&\qquad\qquad P_x^{(\hat{R}_m(l,\eta),\tau_1)}(dr,dt)d\xi d\eta P_\nu^{(F_0,L_0)}(df,dl) + O(T^{-1}) \\
&= \sum_{m;|T-\alpha m|<\delta T} \int_{-\infty}^{\infty}\int_{-\infty}^{\hat{z}} \psi_{m,2}(\eta,l,t)\phi_{D^0}(\xi,\eta)\left\{1+m^{-1/2}q_1^0(\xi,\eta)\right\} \\
&\qquad\qquad P_x^{(\hat{R}_m(l,\eta),\tau_1)}(dr,dt)d\xi d\eta P_\nu^{(F_0,L_0)}(df,dl) + O(T^{-1}),
\end{aligned}
$$

where

$$D^0 = D^{0,k}, \quad q_i^0 = q_i^{0,k}$$

which do not depend on k by Lemma 2. Here we used the fact

$$
\begin{aligned}
&\sum_{m;|T-\alpha m|<\delta T} \int \psi_{m,2}(\eta,l,t)|h(f,r)|e^{-\epsilon\eta^2} P_x^{(\hat{R}_m(l,\eta),\tau_1)}(dr,dt)d\eta P_\nu^{(F_0,L_0)}(df,dl) \\
&= \sum_{m;|T-\alpha m|<\delta T} m^{-1/2}\int 1_{\{0\leq T-\eta<t\}} \exp\left\{-\epsilon\left|\frac{\eta-\alpha m - l}{\sqrt{m}}\right|^2\right\} \\
&\qquad\qquad |h(f,r)|P_x^{(R(\eta),\tau_1)}(dr,dt)d\eta P_\nu^{(F_0,L_0)}(df,dl)
\end{aligned}
$$

where

$$R(\eta) = \int_0^{T-\eta} f(X_t)dt$$

and

$$
\begin{aligned}
&\sum_{m;|T-\alpha m|<\delta T} m^{-1/2}\exp\left\{-\epsilon\left|\frac{\eta-\alpha m-l}{\sqrt{m}}\right|^2\right\} \\
&\leq ||1-\delta|T/\alpha|^{-1/2}\left|3+\int_{\mathbb{R}}\exp\left\{-\epsilon\left|\frac{\eta-\alpha u - l}{\sqrt{|1+\delta|T/\alpha}}\right|^2\right\}du\right| = O(1)
\end{aligned}
$$

uniformly in (l,η) for any $\epsilon > 0$, and

$$
\begin{aligned}
&\int |h(f,r)|1_{\{0\leq T-\eta<t\}}P_x^{(R(\eta),\tau_1)}(dr,dt)d\eta P_\nu^{(F_0,L_0)}(df,dl) \\
&\leq T^{-1}P_\nu[|F_0|]P_x[\tau_1] + T^{-1}P_x\left[\int_0^{\tau_1}\left|\int_0^v f(X_s)ds\right|dv\right] = O(T^{-1}).
\end{aligned}
$$

The rest of the proof is the same as Steps 2 and 3 of Section 4.1 of Fukasawa [6].

4.2. **Proof of Theorem 3.** Let

$$\hat{A}(x_1,\ldots,x_5) = \frac{x_1}{\sqrt{x_3+\hat{\alpha}\hat{\sigma}^2 - \frac{2x_1(x_4+\hat{\rho})}{x_2+\hat{\alpha}} + \frac{x_1^2(x_5+\hat{v}_L)}{(x_2+\hat{\alpha})^2}}}$$

and

$$\bar{X}^* = \frac{1}{M_T}\sum_{j=1}^{M_T}\left(\check{F}_j^*, L_j^* - \hat{\alpha}, |\check{F}_j^*|^2 - \hat{\alpha}\hat{\sigma}^2, \check{F}_j^* L_j^* - \hat{\rho}, \{L_j^*\}^2 - \hat{v}_L\right)$$

where $\check{F}_j^* = F_j^* - \check{\theta}_T L_j^*$ and $\hat{v}_l = \frac{1}{M_T}\sum_{j=1}^{M_T} L_j^2$. Notice that

$$\sqrt{T^*}(\check{\theta}_T^* - \check{\theta}_T)/\hat{\sigma}_T^* = \sqrt{M_T}\hat{A}(\bar{X}^*).$$

Although M_T is a random variable, the proof of Theorem 5.1 of Hall [7] remains valid for $n = M_T$. Hence, we have

$$\begin{aligned} &P_\mu^*\left[\sqrt{T^*}(\check{\theta}_T^* - \check{\theta}_T)/\hat{\sigma}_T^* \leq z\right] \\ &= \Phi(z) + \{\hat{\alpha}M_T\}^{-1/2}\phi(z)\left\{(\hat{b} + 3\hat{a}_2)z + 2\hat{a}_2(z^2-1)\right\} + O(M_T^{-1}) \end{aligned}$$

a.s.. It suffices then to observe that

$$\{\hat{\alpha}M_T\}^{-1/2} = T^{-1/2} + O(T^{-1}).$$

References

[1] P. Bertail and S. Clémençon. Edgeworth expansions of suitably normalized sample mean statistics for atomic Markov chains. *Probability Theory and Related Fields*, 130:388–414, 2004.

[2] P. Bertail and S. Clémençon. Regenerative block bootstrap for Markov chains. *Bernoulli*, 12, no.4:689–712, 2006.

[3] P. Bertail and S. Clémençon. Second-order properties of regeneration-based bootstrap for atomic Markov chains. *Test*, 16:109–122, 2007.

[4] R.N. Bhattacharya and J.K. Ghosh. On the validity of the formal Edgeworth expansion. *Annals of Statistics*, 6:434–451, 1976.

[5] R.N. Bhattacharya and R.R. Rao. *Normal approximation and asymptotic expansions*. Wiley, New York, 1976.

[6] M. Fukasawa. Edgeworth expansion for ergodic diffusions. *Probability Theory and Related Fields*, 2007, to appear.

[7] P. Hall. *The bootstrap and Edgeworth expansion*. Springer Verlag, New York, 1992.

[8] S. Kusuoka and N. Yoshida. Malliavin calculus, geometric mixing, and expansion of diffusion functionals. *Probability Theory and Related Fields*, 116:457–484, 2000.

[9] N. Yoshida. Malliavin calculus and asymptotic expansion for martingales. *Probability Theory and Related Fields*, 109:301–342, 1997.

[10] N. Yoshida. Partial mixing and Edgeworth expansion. *Probability Theory and Related Fields*, 129:559–624, 2004.

Graduate School of Mathematical Science, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo, 153-8914, Japan

E-mail address: `fukasawa@ms.u-tokyo.ac.jp`

UTMS

2007–10 Yasufumi Osajima: *General asymptotics of Wiener functionals and application to mathematical finance.*

2007–11 Shuichi Iida: *Adiabatic limits of η-invariants and the meyer functions.*

2007–12 Ken-ichi Yoshikawa: *$K3$ surfaces with involution, equivariant analytic torsion, and automorphic forms on the moduli space II: a structure theorem.*

2007–13 M. Bellassoued, D. Jellali and M. Yamamoto: *Stability estimate for the hyperbolic inverse boundary value problem by local Dirichlet-to-Neumann map.*

2007–14 M. Choulli and M. Yamamoto: *Uniqueness and stability in determining the heat radiative coefficient, the initial temperature and a boundary coefficient in a parabolic equation.*

2007–15 Yasuo Ohno, Takashi Taniguchi, and Satoshi Wakatsuki: *On relations among Dirichlet series whose coefficients are class numbers of binary cubic forms.*

2007–16 Shigeo Kusuoka, Mariko Ninomiya, and Syoiti Ninomiya: *A new weak approximation scheme of stochastic differential equations by using the Runge-Kutta method.*

2007–17 Wuqing Ning and Masahiro Yamamoto: *The Gel'fand-Levitan theory for one-dimensional hyperbolic systems with impulsive inputs.*

2007–18 Shigeo Kusuoka and Yasufumi Osajima: *A remark on the asymptotic expansion of density function of Wiener functionals.*

2007–19 Masaaki Fukasawa: *Realized volatility based on tick time sampling.*

2007–20 Masaaki Fukasawa: *Bootstrap for continuous-time processes.*

The Graduate School of Mathematical Sciences was established in the University of Tokyo in April, 1992. Formerly there were two departments of mathematics in the University of Tokyo: one in the Faculty of Science and the other in the College of Arts and Sciences. All faculty members of these two departments have moved to the new graduate school, as well as several members of the Department of Pure and Applied Sciences in the College of Arts and Sciences. In January, 1993, the preprint series of the former two departments of mathematics were unified as the Preprint Series of the Graduate School of Mathematical Sciences, The University of Tokyo. For the information about the preprint series, please write to the preprint series office.

ADDRESS:
Graduate School of Mathematical Sciences, The University of Tokyo
3–8–1 Komaba Meguro-ku, Tokyo 153-8914, JAPAN
TEL +81-3-5465-7001 FAX +81-3-5465-7012