Entropy Methods and Martingales, with Applications to Statistics

西山　陽一

統計数理研究所

March 7, 2010

日本統計学会春季集会 2010 （青山学院大学にて）

講演の構成

1. エントロピー法とは

2. 丸められたデータに対する Donsker の定理

   (J.Japan Statist. Soc., 2008 および J.Japan Statist. Soc., 2009)

3. 拡散過程のセミパラメトリック推定

   (Ann. Statist., 2009)

（4.）ヒルベルト空間におけるマルチンゲール中心極限定理

   (番外編．時間が許せば話します)

１．エントロピー法とは

I.i.d. データに対する経験過程

$$X_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1_{(-\infty,x]}(Z_i) - P(-\infty,x]), \quad x \in \mathbb{R}$$

$$X_n(A) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1_A(Z_i) - P(A)), \quad A \in \mathcal{A} \subset \mathbf{B}(\mathcal{X})$$

$$X_n(\psi) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\psi(Z_i) - \int_{\mathcal{X}} \psi(z)P(dz)\right), \quad \psi \in \Psi \subset L_2(P)$$

「エントロピー法」＝確率場 $\psi \rightsquigarrow X_n(\psi)$ に対する理論の研究

Covering Number：
空間 Ψ を距離 ρ に関する ε-球で覆う為に必要な最小個数 $N(\varepsilon, \Psi, \rho)$.

Bracketing Number：
それを少し修正したもの：$N_{[\,]}(\varepsilon, \Psi, \rho)$. $N(\varepsilon, \Psi, \rho)$ より大きい.

I.i.d. の場合の最大不等式：ρ が $L_2(P)$-距離である場合,

$$E \sup_{\rho(\psi,\phi)\leq\delta} |X_n(\psi) - X_n(\phi)| \lesssim \int_0^\delta \sqrt{\log(1 + N_{[\,]}(\varepsilon, \Psi, \rho))}\, d\varepsilon.$$

よって緊密性の十分条件は

$$\int_0^1 \sqrt{\log N_{[\,]}(\varepsilon, \Psi, \rho)}\, d\varepsilon < \infty.$$

プロジェクト：この理論をマルチンゲールの場合に一般化する．

$||X_n||_\rho$ を quadratic modulus とするとき，

$$E \sup_{\rho(\psi,\phi)\le\delta} |X_n(\psi) - X_n(\phi)| 1_{\{||X_n||_\rho \le K\}} \lesssim K \int_0^\delta \sqrt{\log(1 + N_{[\,]}(\varepsilon, \Psi, \rho))}\, d\varepsilon.$$

よって緊密性の十分条件は

$$||X_n||_\rho = O_P(1) \quad \text{かつ} \quad \int_0^1 \sqrt{\log N_{[\,]}(\varepsilon, \Psi, \rho)}\, d\varepsilon < \infty.$$

統計的応用：

1. 弱収束定理による汎関数の意味での漸近正規性

2. 無限次元の漸近有効性

3. 適合度検定などへの応用

4. 従来の緊密性判定定理では困難な問題の解決

5. 無限次元推定量の収束率の導出

6. 無限次元局外パラメータの消去

今日の講演：

- ２節で 4. の一例（丸められたデータに対する Donsker の定理）
- ３節で 6. の一例（拡散過程のセミパラメトリック推定）

2．丸められたデータに対する Donsker の定理

$$
\begin{aligned}
D_n &= \sup_{t\in\mathbb{R}} n^{1/2}|\mathbb{P}_n(-\infty,t]-P_0(-\infty,t]| \\
D_n^\star &= \sup_{t\in\mathbb{R}} n^{1/2}|\mathbb{P}_n^\star(-\infty,t]-P_0(-\infty,t]| \\
D_n^{\star\star} &= \sup_{t\in\mathbb{R}} n^{1/2}|\mathbb{P}_n^\star(-\infty,t]-P_0^{n\star}(-\infty,t]|
\end{aligned}
$$

$$
\delta_n = \sup_k P_0(A_k^n)
$$

主定理 分布関数 $t \mapsto P_0(-\infty,t]$ は連続とする．

(i) もし $\delta_n = o(n^{-1/2})$ ならば $D_n^\star$ の収束が成立する．

(ii) もし $\delta_n = o(1)$ ならば $D_n^{\star\star}$ の収束が成立する．

では一番基本的な例に戻ろう.

- F_0 は $[0,1]$ 上の一様分布とする.
- $\{Z_1, ..., Z_n\}$: i.i.d. with F_0.
- $\delta_n = 0.01$. $\{Z_1^\star, ..., Z_n^\star\}$ は δ_n まで丸められたデータとする.

$$
\begin{aligned}
Z_1 &= 0.67774205 & Z_1^\star &= 0.68 \\
Z_2 &= 0.81124449 & Z_2^\star &= 0.81 \\
& & \cdots & \\
Z_n &= 0.61694806 & Z_n^\star &= 0.62
\end{aligned}
$$

- $\widehat{F}_n$ は $\{Z_1, ..., Z_n\}$ の経験分布関数とする.
- $\widehat{F}_n^\star$ は $\{Z_1^\star, ..., Z_n^\star\}$ の経験分布関数とする.

Kolmogorov-Smirnov 統計量

$$D_n = \sup_{t \in [0,1]} n^{1/2} |\widehat{F}_n(t) - F_0(t)|$$

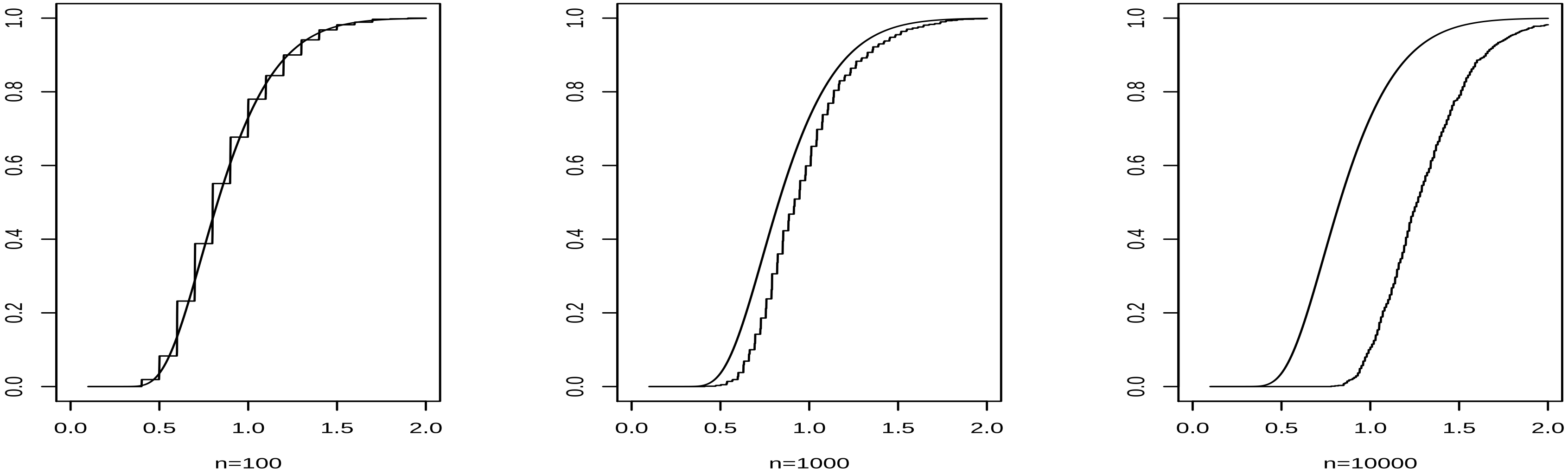
は $\sup_{u \in [0,1]} |B^\circ(u)|$ に分布収束する．ただし $u \rightsquigarrow B^\circ(u)$ は標準ブラウン橋.

一方，我々の結果によると，もし $\delta_n = o(n^{-1/2})$ ならば検定統計量

$$D_n^\star = \sup_{t \in [0,1]} n^{1/2} |\widehat{F}_n^\star(t) - F_0(t)|$$

も $\sup_{u \in [0,1]} |B^\circ(u)|$ に分布収束する.

$D_n^\star$ のシミュレーションを $n = 100, 1000, 10000$ および$\delta_n = 0.01$ で行った結果が次の図である.

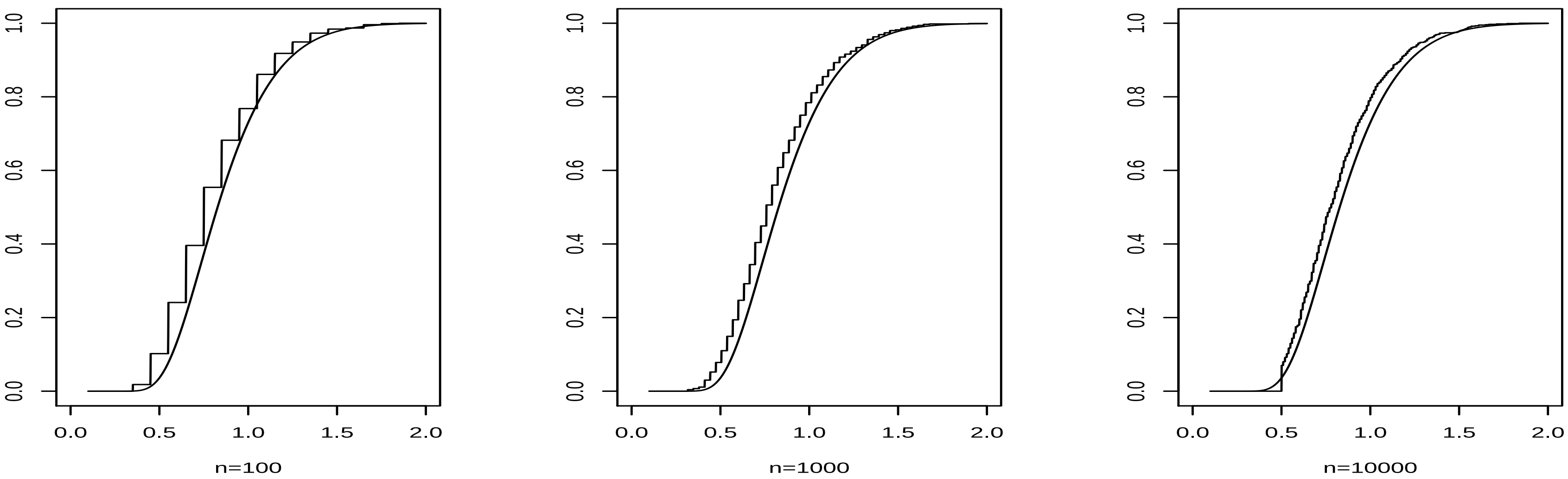


n が大きいとき $\sup_{u\in[0,1]} |B_u^\circ|$ の分布（実線）への収束は悪い．これは，仮定 $\delta_n = o(n^{-1/2})$ が成り立っていないからである.

この欠点を改善するため，検定統計量

$$D_n^{\star\star} = \sup_{t\in[0,1]} \sqrt{n}|\widehat{F}_n^{\star}(t) - F_{n,0}^{\star}(t)|$$

を導入する．ただし $F_{n,0}^{\star}$ は F_0 を「丸めた」ヴァージョンである．これは $\delta_n = o(1)$ という緩い条件のもとで収束する．



この事実は，確率過程

$$X_n(t) = n^{1/2}(\widehat{F}_n^\star(-\infty, t] - F_{n,0}^\star(-\infty, t])$$

の弱収束を示すことによって得られる．ただしそれは古典的な緊密性判定条件

$$\mathrm{E}|X_n(t) - X_n(t')|^p \le K|t - t'|^{1+r} \quad \text{for some } p, K, r > 0$$

では証明できない．なぜなら $P_0^{n\star}$ は離散分布だからである．

Billingsley (1968, page 133) は，この判定条件を，ある右連続関数 H に対する

$$\begin{aligned} &P(|X_n(t) - X_n(t_1)| \ge \lambda,\ |X_n(t_2) - X_n(t)| \ge \lambda) \\ &\le \frac{1}{\lambda^{2\gamma}}[H(t) - H(t_1)]^\alpha [H(t_2) - H(t)]^\alpha \end{aligned}$$

に置き換えているが，関数 H は全ての n に共通にとらなければならない．

我々の方法によると

$$\int_0^1 \sup_{n\in\mathbb{N}} \sqrt{\log N_{[\,]}(\varepsilon,\Psi,L_2(P_0^{n\star}))}d\varepsilon < \infty;$$

でよい．幸いにしてブラケティング数 $N_{[\,]}(\varepsilon,\Psi,L_2(Q))$ は Q に依存せずに押さえられる．実際， $\Psi = \{1_{(-\infty,t]} : t\in\mathbb{R}\}$ のとき，ある定数 $K>0$ が存在して 任意の 確率測度 Q に対し

$$\log N_{[\,]}(\varepsilon,\Psi,L_2(Q)) \le K\left(\frac{1}{\varepsilon}\right)$$

が成り立つ.（例えば van der Vaart and Wellner (1996) を見よ.）

３．拡散過程のセミパラメトリック推定

１次元の拡散過程を考える.

$$X_t = X_0 + \int_0^t S(X_s;\theta)ds + \int_0^t \sigma(X_s;h)dW_s,$$

ただし $s \rightsquigarrow W_s$ は標準ブラウン運動.

- $\{S(\cdot;\theta);\theta\in\Theta\}$, ただし $\Theta\subset\mathbb{R}^d$ はコンパクト.
- $\{\sigma^2(\cdot;h);h\in H\}$, ただし (H,d_H) は全有界距離空間.

目標は θ_0 の推定を，モデルが局外パラメータ h をもっている場合に行うことである．手順は

- まず d_H-一致推定量 $\widehat{h}_n$ を構成する.
- そして Z-推定量 $\widehat{\theta}_n$ （すなわち推定方程式 $\Psi_n(\theta,\widehat{h}_n)=0$ の解）が漸近正規かつ漸近有効であることを示す.

連続観測の場合：

拡散係数は既知としてよいので，h_0 と書くことにし，最尤推定量 $\widehat{\theta}_T$ は $\dot{\ell}_T(\theta) = 0$ の解である．ただし

$$\dot{\ell}_T(\theta) = \frac{1}{T}\int_0^T \frac{\dot{S}(X_t;\theta)}{\sigma^2(X_t;h_0)}[dX_t - S(X_t;\theta)dt],$$

であり，$\dot{S}$ は S の θ に関する微分を表す.

離散観測の場合：

$$\{0 = t_0^n < t_1^n < \cdots < t_n^n\}$$

拡散係数も推定しなければならない.

- Florens-Zmirou (1989), Yoshida (1992) and Kessler (1997) らは H が有限次元である場合を考察した.
- 我々は無限次元の (H, d_H) を考える.
- サンプリングスキームとしては

$$\Delta_n = \max_{1 \le i \le n} |t_i^n - t_{i-1}^n| = o((t_n^n)^{-1}) \quad \text{and} \quad t_n^n \to \infty,$$

を仮定する．前者は $n\Delta_n^2 \to 0$ とほぼ同じ.

- 注意 1 ：Kessler (1997) の仮定 $n\Delta_n^p \to 0$ の方が弱い.
- 注意 2 ：先行研究は有限次元パラメータ $\widehat{h}_n$ の一致性のみならず漸近分布も導いている.
- しかしながら：我々の研究のポイントは (H, d_H) が無限次元である場合に θ の**漸近有効**推定を行うことである.

ただし**漸近有効**とは，規格化された残差 $\sqrt{t_n^n}(\widehat{\theta}_n - \theta_0)$ が，連続観測であって $h = h_0$ が既知の場合と同じ漸近分布をもつことである. よって局所漸近正規族理論（Le Cam 理論）の意味で漸近有効である.

連続観測の場合の

$$\dot{\ell}_T(\theta) = \frac{1}{T}\int_0^T \frac{\dot{S}(X_t;\theta)}{\sigma^2(X_t;h_0)}[dX_t - S(X_t;\theta)dt]$$

を離散化して,

$$\Psi_n(\theta,h) = \frac{1}{t_n^n}\sum_{i=1}^n \frac{\dot{S}(X_{t_{i-1}^n};\theta)}{\sigma^2(X_{t_{i-1}^n};h)}[X_{t_i^n} - X_{t_{i-1}^n} - S(X_{t_{i-1}^n};\theta)|t_i^n - t_{i-1}^n|]$$

を提案する．ただし h_0 は未知の h によって置き換えた．その「カンペンセイター」は

$$\widetilde{\Psi}_n(\theta,h) = \frac{1}{t_n^n}\sum_{i=1}^n \frac{\dot{S}(X_{t_{i-1}^n};\theta)}{\sigma^2(X_{t_{i-1}^n};h)}\left[\int_{t_{i-1}^n}^{t_i^n} S(X_t;\theta_0)dt - S(X_{t_{i-1}^n};\theta)|t_i^n - t_{i-1}^n|\right]$$

である．すなわち，差 $\Psi_n(\theta,h) - \widetilde{\Psi}_n(\theta,h)$ はマルチンゲールである．

キーポイントは

- 確率場 $(\theta,h) \rightsquigarrow r_n(\Psi_n(\theta,h)-\widetilde{\Psi}_n(\theta,h))$ の弱収束,
- $(\theta,h) \rightsquigarrow \widetilde{\Psi}_n(\theta,h)$ の (θ_0,h_0) の周りでの微分可能性.

ラフに言って，我々の結果は

- $h \mapsto \sigma^2(\cdot;h)$ が d_H に関しリプシッツ連続
- $\int_0^1 \sqrt{\log N(\varepsilon,H,d_H)}d\varepsilon < \infty$
- h_0 に対する d_H-一致推定量 $\widehat{h}_n$ の存在

の仮定のもとで $r_n(\widehat{\theta}_n-\theta_0)$ の漸近分布が導出できる．$\widehat{h}_n$ の一致性は別途証明せねばならないが，今の場合は可能である.

The Jain-Marcus theorem for martingales

- (H, d_H)：距離空間.
- 各 $h \in H$ に対し，$\{\xi_i^{n,h}\}_{i=1,2,\dots}$ は実数値のマルチンゲール差分列.

ここで

$$|\xi_i^{n,h} - \xi_i^{n,h'}| \le K_i^n d_H(h, h') \quad h, h' \in H.$$

を仮定する．このとき，緊密性判定条件は

$$\sum_{i=1}^n E[|K_i^n|^2 | \mathcal{F}_{i-1}^n] = O_P(1) \qquad \int_0^1 \sqrt{\log N(\varepsilon, H, d_H)} d\varepsilon < \infty.$$

番外編．ヒルベルト空間のマルチンゲール中心極限定理

Kolmogorov-Smirnov 検定は ℓ^∞ 空間における弱収束に基づいている．

$$\sup_{x\in\mathbb{R}} |\sqrt{n}(\widehat{F}_n(x) - F_0(x))| \to^d \sup_{t\in[0,1]} |B_t^\circ|$$

Cramér-von Mises 検定については L_2 空間における弱収束で十分である．

$$\int_{\mathbb{R}} |\sqrt{n}(\widehat{F}_n(x) - F_0(x))|^2 dF_0(x) \to^d \int_0^1 |B_t^\circ|^2 dt$$

とある問題を解決することをモチベーションとして，このような考察にいたりました．すなわち，ヒルベルト空間における弱収束理論が欲しい！

- $\mathbb{H}$: 実ヒルベルト空間
  - $\langle , \rangle$：内積
  - $||\cdot||$：ノルム
  - $\{e_j : j \in J\}$：完全正規直交系
- 各 $n \in \mathbb{N}$ に対し，$\{\xi_i^n\}_{i=1,2,\ldots}$ は確率基 $(\Omega^n, \mathcal{F}^n, \{\mathcal{F}_i^n\}_{i=0,1,2,\ldots}, P^n)$ の上で定義された $\mathbb{H}$ に値をとる確率変数の三角列
  - ξ_i^n の可測性は仮定しない．単なる Ω^n から $\mathbb{H}$ への写像である
  - 各 $h \in \mathbb{H}$ に対し，$\langle \xi_i^n, h\rangle$ が $\mathcal{F}_i^n$-可測かつ $E^n[\langle \xi_i^n, h\rangle | \mathcal{F}_{i-1}^n] = 0$ almost surely であることを仮定する．

定理. 次の (i) – (iii) を仮定する：

(i) $\sum_{i=1}^n E^n[\langle \xi_i^n, h\rangle^2 | \mathcal{F}_{i-1}^n] \to^p C(h) \quad \forall h \in \mathbb{H}$.

(ii) $\exists \delta > 0$ s.t. $\sum_{i=1}^n E^n[||\xi_i^n||^{2+\delta} | \mathcal{F}_{i-1}^n] \to^p 0$.

(iii) $\exists \zeta$ ($\mathbb{H}$ に値をとる確率変数) s.t. $E||\zeta||^2 < \infty$ かつ

$$\sum_{i=1}^n E^n\langle \xi_i^n, e_j\rangle^2 \le E\langle \zeta, e_j\rangle^2, \quad \forall n, \forall j \in J.$$

このとき,

$$\sum_{i=1}^n \xi_i^n \to^d G \quad \text{in } \mathbb{H}, \quad \text{ただし } \langle G, h\rangle \sim \mathcal{N}(0, C(h)).$$

注意： 仮定 (iii) は，もしも ξ_i^n が $\frac{1}{\sqrt{n}}\zeta_i$ の形をしていて ζ_i が定常かつ $E||\zeta_1||^2 < \infty$ ならば満たされている.

References

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

[2] Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley, New York.

[3] Donsker, M.D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **23** 277-281.

[4] Doob, J.L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **20** 303-403.

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

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

[7] Florens-Zmirou, D. (1989). Approximate discrete time schemes for statistics of diffusion processes. *Statistics.* **20** 547-557.

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

[9] Jain, N. and Marcus, M. (1975). Central limit theorems for $C(S)$-valued random variables. *J. Funct. Anal.* **19** 216-231.

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

[11] Negri, I. and Nishiyama, Y. (2009). Goodness of fit test for ergodic diffusion processes. *Ann. Inst. Statist. Math.* **61** 919-928.

[12] Nishiyama, Y. (1996). A central limit theorem for ℓ^∞-valued martingale difference arrays and its application. *Preprint* **971**, Dept.of Mathematics, Utrecht Univ. Available at `http://www.ism.ac.jp/~nisiyama/`

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

[14] Nishiyama, Y. (1999). A maximal inequality for continuous martingales and M-estimation in a Gaussian white noise model. *Ann. Statist.* **27** 675-696.

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

[16] Nishiyama, Y. (2000b). *Entropy Methods for Martingales.* CWI Tract **128** Centrum voor Wiskunde en Informatica, Amsterdam.

[17] Nishiyama, Y. (2007). On the paper "Weak convergence of some classes of martingales with jumps". *Ann. Probab.* **35** 1194-1200.

[18] Nishiyama, Y. (2008a). Nonparametric estimation and testing time-homogeneity for processes with independent increments. *Stochastic Process. Appl.* **118** 1043-1055.

[19] Nishiyama, Y. (2008b). Donsker's theorem for discretized data. *J. Japan Statist. Soc.* **38** 505-515.

[20] Nishiyama, Y. (2009a). Asymptotic theory of semiparametric Z-estimators for stochastic processes with applications to ergodic diffusions and time series. *Ann. Statist.* **37** 3555-3579.

[21] Nishiyama, Y. (2009b). A note on semiparametric estimation for ergodic diffusion processes. *Research Memorandum* **1089**, Inst. Statist. Math. Available at `http://www.ism.ac.jp/~nisiyama/`

[22] Nishiyama, Y. (2009c). Two sample problem for rounded data. *J. Japan Statist. Soc.* **39** 233-238.

[23] Nishiyama, Y. (2010). A martingale central limit theorem in Hilbert space and its applications. *Research Memorandum* **1116**, Inst. Statist. Math. *Submitted for publication.*

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

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

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

[27] Yoshida, N. (1992). Estimation for diffusion processes from discrete observations. *J. Multivariate Anal.* **41** 220-242.