エントロピー法とマルチンゲール

西山陽一　　数理・推論研究系

1　研究概要

本研究室の目標は、生存解析や数理ファイナンスといった時間の流れに従って変化する現象を扱う統計解析の分野に新しい理論的貢献をすることです。使う道具をキーワード的に３つ挙げれば、確率場の理論・セミマルチンゲール理論・漸近的推測決定理論となります。

2　確率場の理論

T を任意の集合とします。確率場とは、一般には確率変数の集合 $\{X(t): t\in T\}$ のことを指します。形式的にはランダム現象であれば何でも含んでいるような概念ですが、特に $T=\{1,2,\ldots\}$ であるとき「時系列」と呼んだり、$T=[0,\infty)$ であるとき「確率過程」と呼んだりします。通常「確率場」というのは T が任意の集合である場合を指します。特に T が関数空間であるような場合を念頭に、セミパラメトリック・ノンパラメトリック統計手法のための基礎となる理論を研究しています。このような研究は、従来「独立同一分布 (i.i.d.)」の場合についてなされてきましたが、応用上その制約は強すぎるので、制約をマルチンゲールまで緩める（つまり理論を拡張する）ことを、本研究室では世界に先駆けて行いました（Nishiyama (1997, 1999, 2000, 2007)）。核心部分となるのは、メトリック・エントロピー条件です。

$$\int_0^1 \sqrt{\log N(T,\rho;\varepsilon)}d\varepsilon < \infty$$

3　セミマルチンゲール理論

上で出てきたマルチンゲールとは、「平均がゼロ」という概念を自然に拡張したものです。セミマルチンゲールとは、

増減のトレンド＋マルチンゲール

の形に分解される確率過程の総称で、豊富な応用例をもちます。また、それには Itô (1944) によって創始された確率解析の理論が適用できるので、時間の流れに従って変化する現象の統計解析に広く応用されています。

4　漸近的推測決定理論

Le Cam (1960) によって提起された局所漸近正規性という概念は、「i.i.d. の場合において成功してきた統計手法をどこまで拡張できるか」という問題に対するひとつの回答を与えるものであり、それをさらに無限次元統計モデルまで拡張する研究が 1980 年代に行われました。それを確率場の理論と融合する研究は van der Vaart and Wellner (1996) のモノグラフによってなされました。その成果のひとつは局所漸近ミニマックス定理です。

$$\sup_{I\subset H}\liminf_{n\to\infty}\sup_{h\in I} E_h\ell\left(\tau_n(T_n-\kappa_n(h))\right)\geq E\ell(G)$$

本研究室では、独自に開発したマルチンゲール確率場の理論を武器にその研究潮流に参入し、拡散過程・計数過程といった統計モデルにおける新しい統計手法の提案を行っています。

5　研究の例１：生存解析における離散的観測問題

計数過程の積強度モデルは、右センサリングを受けた観測値も扱える汎用性の高い統計モデルです。このモデルにおける漸近理論を創始したのは Aalen (1978) です。彼はマルチンゲール理論を用いて、今日 Nelson-Aalen 推定量と呼ばれるようになった推定量の汎関数の意味での漸近正規性を証明しました。ところが、その定義には確率積分が用いられているため、failure time data $T_1,T_2,\cdots$ が連続的に観測されていることが必要でした。しかし実際には、例えば観測期間が１年間で、failure time を日単位で観測すると、時間幅 1/365 に丸められたデータが採取されることになります。このような場合、従来の Nelson-Aalen 推定量は厳密には計算できません。この欠点を克服するために、本研究室では離散化 Nelson-Aalen 推定量を開発し、さらにその漸近正規性・漸近有効性を証明しました。

$$\widetilde{A}^n(x)=\sum_{i=1}^{m(n)} 1_{[0,x]}(t_i^n)Y_{t_{i-1}^n}^{n-}[N_{t_i^n}^n-N_{t_{i-1}^n}^n]$$

6　研究の例２：拡散過程の適合度検定

適合度検定における Kolmogorov-Smirnov 検定統計量が漸近的に分布不変である（漸近分布が標準 Brown 橋のシンプルな汎関数となる）ことは、特に i.i.d. の場合には統計学が生まれた 20 世紀初頭から知られています。一方で拡散過程は、株価の変動を表す数理モデルとして近年脚光を浴びています。ところが、拡散過程に対する適合度検定は、Kolmogorov-Smirnov 検定統計量を単純に拡張しただけでは漸近的にも分布不変にならないので、つい最近に至るまで未解決問題でした。そこで本研究室では、score marked empirical process のアイデアを用いて、2006 年に（連続観測の場合について）ドリフトの検定統計量であって漸近的分布不変であるものを構成することに成功しました。さらにその研究を発展させ、離散観測やいわゆる tick time 観測の場合について、ドリフトのみならず拡散係数に対しても漸近的分布不変な検定統計量を構成しました。

7　研究の例３：拡散過程のセミパラメトリック推定

確率微分方程式

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

は、ドリフト係数 S をトレンドの微係数とし、拡散係数 σ によってノイズの構造を表すようなセミマルチンゲールの統計モデルです。従来の理論では、未知母数 θ と h はともに有限次元であると仮定されていました。ところが、仮にドリフト係数の未知母数 θ の推定のみに興味がある場合、拡散係数の未知母数 h の入り方にはなるべく特別な構造を仮定せず、高次元の未知母数としておく方が、モデルのミススペシフィケーションを避けるという観点からは望ましいです。そこで、本研究室では、エントロピー法の技術をセミパラメトリック推測理論に応用し、本モデルにおいて無限次元の攪乱母数 h の存在のもとで有限次元の未知母数 θ の漸近有効推定を行うことに世界で初めて成功し、論文 [9] として *Ann. Statist.* 誌に発表しました。

8　研究の例４：丸められたデータの解析

統計的漸近理論において最も基本的な定理のひとつが Donsker (1952) の定理です。ただし、そこではデータが実数として連続的に観測できるという仮想的な仮定が置かれていました。しかしながら、実際のデータは常に切り上げ、切り捨て、あるいは四捨五入といった丸めが施されて採取されます。従来の理論ではこの影響は無視して解析が行われてきました。本研究室では、この影響を無視せずにエントロピー法を用いて丸められたデータを解析する方法を研究しました。その結果の一部は、論文 [7] と [8] で公表しました。前者は 2009 年 9 月に第２３回日本統計学会小川研究奨励賞を受賞しました。

参考文献

[1] Negri, I. and Nishiyama, Y. (2009). Goodness of fit test for ergodic diffusion processes. *Ann. Inst. Statist. Math.* **61** 919-928.
[2] Negri, I. and Nishiyama, Y. (2010). Goodness of fit test for ergodic diffusions by tick time sample scheme. *Stat. Inference Stoch. Process.* **13** 81-95.
[3] Nishiyama, Y. (1999). A maximal inequality for continuous martingales and M-estimation in a Gaussian white noise model. *Ann. Statist.* **27** 675-696.
[4] Nishiyama, Y. (2000). Weak convergence of some classes of martingales with jumps. *Ann. Probab.* **28** 685-712.
[5] Nishiyama, Y. (2007). On the paper "Weak convergence of some classes of martingales with jumps". *Ann. Probab.* **35** 1194-1200.
[6] Nishiyama, Y. (2007). Nonparametric inference in multiplicative intensity model by discrete observation. *To appear in Ann. Inst. Statist. Math.*
[7] Nishiyama, Y. (2008). Donsker's theorem for discretized data. *J. Japan Statist. Soc.* **38** 505-515.
[8] Nishiyama, Y. (2009). Two sample problem for rounded data. *J. Japan Statist. Soc.* **39** 233-238.
[9] Nishiyama, Y. (2009). Asymptotic theory of semiparametric Z-estimators for stochastic processes with applications to ergodic diffusions and time series. *Ann. Statist.* **37** 3555-3579.
[10] van der Vaart, A.W. and Wellner, J.A. (1996). Weak Convergence and Empirical Processes: With Applications to Statistics. Springer.