博士学位請求論文審査報告書

元山 斉

「有限母集団に於ける統計的汎関数について」

はじめに

標準的な統計理論では無限母集団からの標本抽出を仮定している。それは多くの場合、標本の大きさが母集団に比べ非常に小さいことによる。一方、官庁統計では母集団の10%、30%といった、非常に大きな標本を採ることが、しばしば見られる。即ち、標本は完全に、有限標本からの非復元抽出となり（計算の簡略化の為に用いられる）標準的な統計手法の正当性が疑われる。更に、近年ノンパラメトリック手法や外れ値に対する頑強な推定量の必要性が高まり、その漸近的な性質の研究には有限母集団における von Mises により提唱され、近年 Reeds や Fernfolz 等により発展してきた"統計的汎関数"の理論が有用である。そして、基本的な極限分布導出については Lindeberg-Feller の中心極限定理 (CLT) ではなく、Erdös と Rényi の CLT が必要となる。この様に本論文で元山氏は官庁統計という具体的な問題から派生した興味深い理論的問題を扱っている。更に極限に於いて連続分布を仮定する場合を考慮し、"平滑化"された経験分布関数に基づく漸近論を展開している。本論文は、4 章から成り、その構成は以下の通りである。

第 1 章　序章
第 2 章　平滑化経験分布関数の漸近的性質について
第 3 章　漸近正規性
第 4 章　モンテ・カルロ・シミュレーション

本論文の主な結果は第 2 章と 3 章で展開される有限母集団に於いて統計的汎関数で表現されうる推定量の漸近分布の導出である。以下、章ごとに内容を紹介するが、先ず主要結果から始める。

本論文の内容

第 1 章では、はじめに標本調査の意義を全数調査と比較して論じ、 標本調査の理論的枠組みである有限母集団モデルと そのモデルにおける漸近理論の取り扱いについて論じている。以下　$x_1, x_2, \cdots, x_N$ を有限母集団とする。

まず、本論文の考察対象となる統計的汎関数と平滑化統計的汎関数を定義している。

定義　$T_n = T_n(X_1, X_2, \cdots, X_n)$ を統計量とする。統計量 T_n が経験分布関数 F_n の汎関数 T として表わすことができ, さらに T を通じてのみ標本サイズ n に依存するとき、$T_n = T(F_n)$ を統計的汎関数とよぶ。

さらに平滑化統計的汎関数を定義するために、平滑化経験分布関数 (カーネル分布関数推定量) $\widetilde{F}_n$ を 平滑化統計的汎関数と定義する。

定義　カーネル分布関数推定量 $\widetilde{F}_n$ とは F_n とカーネル関数 k_n の畳み込み

$$\begin{aligned}\tilde{F}_n(x) &= F_n * k_n(x) = \int F_n(x-t)k_n(t)dt \\ &= \frac{1}{n}\sum_{i=1}^{n} K_n(x-X_i),\end{aligned}$$

である。ここで $K_n(x) = \int_{-\infty}^{x} k_n(t)dt$ とする。

汎関数の定義において経験分布関数 F_n をカーネル分布関数推定量 $\widetilde{F}_n$ で置き換えた 統計量を平滑化統計的汎関数と定義する。

その後、本論文の主要結果である 3 つの漸近正規性の結果を紹介している。

最初の定理は、平滑化していない 統計的汎関数の漸近正規性である。

定理 3.1. $X_1, \ldots, X_n$ を非復元単純無作為抽出による標本とする. T を統計的汎関数とし、τ を T から D[0,1] 空間上に誘導された統計的汎関数とする。τ が U で Hadamard 微分可能、影響関数は有界変動関数について Lebesgue-Stieltjes 積分可能であり、$\overline{\mathrm{IF}} = \mathrm{E}[\mathrm{IF}] = \sum_{i=1}^{N} \mathrm{IF}(x_i)/N$ かつ $0 < \sigma_N^2 = \mathrm{Var}\ [\mathrm{IF}] = \sum_{i=1}^{N}(\mathrm{IF}(x_i) - \overline{\mathrm{IF}})^2/N < \infty$ とする。

いま影響関数が以下の Erdös と Rényi の Lindeberg 型の条件を満足すると仮定する:

任意の $\epsilon > 0$ に対して

$$\lim_{n,N-n\to\infty} \frac{\sum_{P_\epsilon}(\mathrm{IF}(x_i) - \overline{\mathrm{IF}})^2}{\sum_{i=1}^{N}(\mathrm{IF}(x_i) - \overline{\mathrm{IF}})^2} = 0,$$

但し P_ϵ は以下の不等式を満足する $P = \{1, \ldots, N\}$ の部分集合とする

$$|\mathrm{IF}(x_i) - \overline{\mathrm{IF}}| > \epsilon\sqrt{n(1-\frac{n}{N})}\,\sigma_N.$$

ここで、$n, N-n \to \infty$ のとき

$$\sqrt{n}(T(F_n) - T(F_{NS}) - \overline{\mathrm{IF}})/\sigma_{N,n} \xrightarrow{D} N(0,1),$$

但し、$\sigma_{N,n}^2 = N\sigma_N^2(1-n/N)/n(N-1) = (N-n)\sigma_N^2/n(N-1)$ である。

又、F_{NS} は、真の分布関数 F_N の近くにある十分に滑らかな分布関数である。

次の定理では、定理 3.1 と類似な条件下で平滑化した統計的汎関数の漸近正規性を証明している。

定理 3.2. $X_1, \ldots, X_n$ を非復元単純無作為抽出による標本、$\{k_n\}$ を有限な一次の積率をもつ正則なカーネル列とする。 T を統計的汎関数として τ を T から $D[0,1]$ 空間上に誘導された統計的汎関数とする。τ が U において Hadamard 微分可能であり影響関数は有界変動関数について Lebesgue-Stieltjes 積分可能、$\overline{\widetilde{\mathrm{IF}}} = \mathrm{E}[\widetilde{\mathrm{IF}}] = \sum_{i=1}^{N} \widetilde{\mathrm{IF}}(x_i)/N$ かつ $0 < \sigma_N^2 = \mathrm{Var}[\widetilde{\mathrm{IF}}] = \sum_{i=1}^{N}(\widetilde{\mathrm{IF}} - \overline{\widetilde{\mathrm{IF}}})^2/N < \infty$ とする。

いま、$\widetilde{\mathrm{IF}}$ が定理 3.1 と同様な Erdös と Rényi の Lindeberg 型の条件を満足すると仮定する。$n, N-n \to \infty$ のとき

$$\sqrt{n}(T(\widetilde{F}_n) - T(F_{NS}) - \overline{\widetilde{\mathrm{IF}}}/\sigma_{N,n} \xrightarrow{D} N(0,1),$$

但し $\overline{\widetilde{\mathrm{IF}}}$ は $\widetilde{\mathrm{IF}}$ の算術平均、$\sigma_{N,n}^2 = N\sigma_N^2(1-n/N)/n(N-1) = (N-n)\sigma_N^2/n(N-1)$ である。

影響関数 IF が有界な関数であるときは、平滑化した汎関数と平滑化していない汎関数の漸近分布は等しく、以下の定理を得る。

定理 3.3. $X_1, \ldots, X_n$ を非復元単純無作為抽出による標本とする。$\{k_n\}$ を有限な一次の積率をもつ正則なカーネル列とする。 T を汎関数とし、τ を T から $D[0,1]$ 空間上に誘導された統計的汎関数とする。τ が U で Hadamard 微分可能であり、 影響関数 IF が有界変動関数に関して Lebesgue-Stieltjes 積分可能な有界関数であり、$\overline{\mathrm{IF}} = \mathrm{E}[\mathrm{IF}] = \sum_{i=1}^{N} \mathrm{IF}(x_i)/N$ かつ $0 < \sigma_N^2 = \mathrm{Var}\ [\mathrm{IF}] = \sum_{i=1}^{N}(\mathrm{IF}(x_i) - \overline{\mathrm{IF}})^2/N < \infty$ とする。

ここで上と同様、影響関数 IF が Erdös と Rényi による Lindeberg 型条件を満足すると仮定する:この時 $n, N-n \to \infty$ ならば

$$(T(\widetilde{F}_n) - T(F_{NS}) - \overline{\mathrm{IF}})/\sigma_{N,n} \xrightarrow{D} N(0,1),$$

但し、$\sigma_{N,n}^2 = N\sigma_N^2(1-n/N)/n(N-1) = (N-n)\sigma_N^2/n(N-1)$ である。

さて、第 2 章と第 3 章で、本論文の主要結果が述べられている。

第 2 章では、始めに平滑化統計的汎関数の理論を構成する上で必要となる平滑化経験分布関数の漸近的性質が示されている。カーネルに正則条件：

ある正の実数列 $\{b_n\}$ (但し $b_n = o(n^{-1/2})$) が存在して

$$\int_{|t|>b_n} |k_n(t)|dt = o(n^{-1/2}) \quad \text{as} \quad n \to \infty$$

を課した基での平滑化経験分布関数と経験分布関数の乖離のオーダーは第 3 章で統計的汎関数のテイラー展開の剰余項をコントロールする為に必要である（命題 2.1）。一方、Glivenko-Cantelli 型の結果（命題 2.2）はそれ自身興味深い。以下主な結果だけ列挙しておく：

命題 2.1　$n, N \to \infty$ のとき

$$\sqrt{n} \parallel \widetilde{F}_n - F_n \parallel \to 0 \quad \text{a.s.}$$

更に $n, N-n \to \infty$ のとき

$$\sqrt{n}\sqrt{N/(N-n)} \parallel \widetilde{F}_n - F_n \parallel \to 0 \quad \text{a.s.}$$

これらより、Glivenko-Cantelli 型の一様強収束結果を得る：

命題 2.2　$n, N \to \infty$ のとき

$$D_n \equiv \sup_{-\infty<x<\infty} |F_n(x) - F_N(x)| \to 0 \quad \text{a.s.}$$

その他の結果として興味深いものは以下の二つである。まず、平滑化経験分布関数に対し、以下の結果が示されている。

命題 2.3　$n, N \to \infty$ のとき

$$\parallel \tilde{F}_n(x) - F_N(x) \parallel \to 0 \quad \text{a.s.}$$

一方、バイアスのオーダーは

命題 2.4

$$\sqrt{n} \parallel F_n^*(x) - F_N(x) \parallel = o(1)$$

で与えられている、ここで、 $F_n^*(x)$ は平滑化された真の分布関数である。

本章は、Glivenko-Cantelli 型の定理を除いて、主として独立同一標本の Fernholz (1991) の内容の非復元無作為抽出への拡張である。

第 3 章では、本論文の主要結果である、非復元無作為抽出の基での統計的汎関数および平滑化統計的汎関数の漸近正規性が証明されている。そのために、先ず統計的汎関数を関数空間上の Taylor 展開を用い線型の部分と剰余項に分割している。

本章は関数空間上での微分の定義から始まる。 線型ノルム空間上の微分の定義は複数存在するが、本論文では以下の定義を用いている。

定義 (Reeds (1976), Fernholz(1983)). τ を線形ノルム空間 $\mathbf{V}$ の開部分集合 $\mathcal{A}$ で定義された汎関数とし $G \in \mathcal{A}$ とする。

1. 汎関数 τ が G で Gateaux 微分可能とは、$\mathbf{V}$ で定義された連続な線型汎関数 τ_G' が存在し

$$\lim_{t \to 0} \frac{\tau(G+tH) - \tau(G) - \tau_G'(tH)}{t} = 0 \tag{1}$$

が各 $H \in \mathbf{V}$ に対して成立することである。このとき τ_G' を G における τ の Gateaux 微係数と呼ぶ。

2. 汎関数 τ が G で Hadamard (compact) 微分可能とは、任意のコンパクトな部分集合 $K \subset \mathbf{V}$ に対して、(1) が $H \in K$ に対して一様に成り立つことである。この線型汎関数を G における τ の Hadamard 微係数と呼ぶ。

3. 汎関数 τ が G で Fréchet 微分可能とは、任意の有界な部分集合 $B \subset \mathbf{V}$ に対して、(1) が $H \in B$ に対して一様に成り立つことである。この線型汎関数を G における τ の Fréchet 微係数と呼ぶ。

上記の関数空間上の微分を行うことによって、統計的汎関数を Taylor 近似すると、統計的汎関数は線型項と剰余項に分割することができる。

本章では、始めに線型項について漸近正規性成立の十分条件を吟味している (Erdös-Rényi の条件)。

その後、剰余項が十分に速いオーダーで 0 に確率収束することを内測度を用いた議論で示し、汎関数の Hadamard 微分可能性が成立するときの、統計的汎関数の漸近正規性を証明し、上述の 3 定理を得ている。

本章は、Reeds (1976)、Fernholz (1983, 1993) の結果の有限母集団への拡張、および Campbell(1980) の議論が精緻化されている。

第 4 章では、漸近正規性の妥当性をモンテ・カルロ・シミュレーションによって調べている。母集団として経済社会データで多く観測される分布の歪みを表現した対数正規分布を採用し、標本抽出による影響をみるため、その母集団から非復元無作為抽出を 100,000 回行い、正規近似より求められた信頼区間の妥当性を検証している。

検討の対象とした汎関数は、対数正規分布のような歪みのあるデータに対して頑健な、位置および広がりの尺度としてしばしば用いられるメディアンと四分位偏差の 2 つの統計量である。

また、標本の大きさと抽出率の違いによる影響を比較するために、母集団の大きさ1000から10,000 までの1,000刻みの10種類の母集団に対して、10%と30%という2種類の異なる抽出率で検証し、以下の結果を得た。
1. 標本と母集団が共に大きい場合は、特に母集団のサイズが5000を超えると、片側の場合も両側の場合も区間の近似の精度はかなりよい。
2. 標本数が少ないときは、抽出率10%の方が抽出率30%よりもよい精度を示している。
3. 信頼係数が高い場合の精度が、低い場合とくらべて相対的によい。
母集団のサイズが十分に大きく特に5000を超えると、 正規近似の精度が非常に高いものであることが確認される。

評価

以上、本論文の概要について述べてきた。本論文の価値は以下の２点に集約できる。
１．現実の標本調査に於ける問題から大変興味深い理論的な問題の発見と解決。
２．平滑化された経験分布関数の基での統計的汎関数の漸近理論を有限母集団の仮定下で厳密に証明。
特に、本問題が内包する各種の数学的な問題を、一つずつ丹念・克明に解決してきた元山氏の力量は十分評価できる。之までにも類似の結果は出ていたが、それらは単に直観的・発見的な証明に終始していた事に比べ、本論文の数学的厳密性は特筆される。また、そのような技術的な問題以外にも平滑化された経験分布関数に関する結果はオリジナルなものである。統計的推測に関し、元山氏の想定した複雑な状況下でも、現実にはコンピューターを用いた数値計算やシュミレーションで十分との考えもあり得るが、その利便性や問題の理解の為に、この様な漸近理論は十分価値がある。第4章、シミュレーションでは比較的簡単な統計量について考察しただけであったが、より複雑な推定量についても同様の考察を加えること。理論面では高次の漸近展開の可能性等について未だ解決すべき問題はあるが、本論文は博士学位を授与するに十分な内容を持ったものと考えられる。したがって、審査員一同は、所定の試験結果と上記論文評価に基づき、元山斉氏に一橋大学博士（経済学）の学位を授与することが適当と判断する。

２００８年7月９日

審査員

石村 直之
黒住 英司
高橋 一 （審査委員長）
田中 勝人
美添 泰人