# 有限母集団における統計的汎関数について
# 学位申請論文要旨

元山　斉

**論文の目的と主要結果**

本稿の企図は，標本調査の枠組みでの統計的汎関数の漸近正規性を示すことにある．

新聞や日々のニュースで取り上げげられる世論調査などの社会調査，テレビの視聴率や商品の売り上げ調査などの企業の実施する市場調査，工業品の品質検査，特定の地域における生物の個体総数の調査など，数多くのデータ収集・作成の現場において標本調査は，しばしば行われている．また，我々が行う経済分析の対象となる企業・世帯等のデータの多く, 特に個票データは標本調査によって得られている．また，集計データの代表である国民経済計算 (SNA) や物価指標である消費者物価指数 (CPI) などのいわゆる二次統計（加工統計）についても，国民経済計算（内閣府経済社会総合研究所）の家計消費支出の推計において全国消費実態調査（総務省総計局）の結果を用いており，消費者物価指数の作成においても家計調査（総務省統計局）で支出額の多い指数品目を選び，消費支出全体に占める各品目に対する支出の割合（ウェイト）を家計調査における年平均 1 か月間の 1 世帯当たり品目別支出金額を用いて推計し，価格については小売物価統計調査（総務省統計局）を用いるなど，その作成過程は標本調査によって得られたデータに基づいている．

標本調査により得られた統計量は，標本抽出誤差の影響を受けるため，その誤差の程度を調べ統計量の精度を評価する必要がある．

本稿では，標本調査によって得られる統計量の中でも，統計的汎関数と呼ばれる，標本平均および $M$-統計量，$L$-統計量のすべてを含む包括的な統計量のクラスに対して，漸近正規性の一般的結果を与える．

統計的汎関数の定義は以下である．

**定義 1.** $T_n = T_n(X_1, X_2, \ldots, X_n)$ 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.

ここで，経験分布関数 $F_n$ とは，母集団 $x_1, \ldots, x_N$ からの非復元無作為標本 $X_1, \ldots, X_n$ により

$$F_n(x, \omega) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty, x]}(X_i).$$

と定義される関数である．

本稿では加えて，平滑化統計的汎関数についても漸近正規性を示す．もし母集団の分布関数が漸近的に滑らかな分布に近づいていくような状況では，分布関数を推定する上で，素朴な経験分布関数よりむしろ滑らかにした推定量を用いた方が良いのではないかと考えられる．

本稿では，独立同一分布標本における Fernholz (1993, 1997) の先行研究に従い，汎関数の定義において経験分布関数 $F_n$ を以下で述べる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．

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 $\widetilde{F}_n = F_n * k_n$ で得られる．即ち

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

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

漸近正規性の結果を述べる前に，必要となる記号と概念を定義する．

統計的汎関数 $T$ と分布関数 $F_N$ について, $T$ の $F_N$ における影響関数 (影響曲線) $\mathrm{IF}_{T,F_N}$ とは以下で定義される実数値関数である

$$\mathrm{IF}_{T,F_N}(x) = \frac{d}{dt}T(F_N + t(\Delta_x - F_N))|_{t=0}$$

ここで $\Delta_x$ は 1 点 $x$ において確率 1 をとる分布の分布関数である．影響関数は，新たに $x$ という値が観測されたときの統計量への影響を評価していると解釈することができる．

次に，統計的汎関数 $T$ を，それと同値な $D[0,1]$ 上の汎関数 $\tau$ への誘導を行う．母集団 $\mathcal{P} = \{x_1, x_2, \ldots, x_N\}$ の $T$-変換を $T\mathcal{P} = \{F_N(x_1), F_N(x_2), \ldots, F_N(x_N)\}$ と定義する．

この「母集団」$T\mathcal{P}$ からの「標本」$F_N(X_1), \ldots, F_N(X_n)$ の経験分布関数を $U_n$ とすると，経験分布関数 $F_n$ は

$$\begin{aligned}
F_n(x) &= \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,F_N(x)]}(F_N(X_i)) = U_n \circ F_N(x)
\end{aligned}$$

と表すことができるため，統計的汎関数 $T$ は以下のように分布関数 $U_n$ の汎関数 $\tau$ によって表わすことができる

$$\tau(U_n) = T(F_n) = T(U_n \circ F_N).$$

$[0,1]$ 上の任意の分布関数 $G$ に対して

$$\tau(G) = T(G \circ F_N) \tag{1}$$

と定義する．すると, 固定した $F_N$ に対して, 汎関数 $T$ は $[0,1]$ 上の分布関数の空間における汎関数 $\tau$ へ誘導される．故に, われわれは関心を，右連続で左極限を持つ $[0,1]$ 上の実数値関数の集合 $D[0,1]$ の要素の $[0,1]$ 上の分布関数に限定することができる．

以上の定義のもと，本論文で示す主要結果は次の 3 つの漸近正規性である．

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，$\tau$ を関係式 ( 1 ) に基づいて $T$ から $D[0,1]$ 空間上に誘導された統計的汎関数とする．$\tau$ が $U$ で Hadamard 微分可能であり，影響関数 IF が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 IF が以下の Erdös と Rényi による以下の Lindeberg 型の条件を満足すると仮定する:

任意の $\epsilon > 0$ に対して

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

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

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

そのとき $n, N-n \to \infty$ のとき

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

となる．ここで，$\sigma_{N,n}^2 = N\sigma_N^2(1 - n/N)/n(N-1) = (N-n)\sigma_N^2/n(N-1)$ である．

2 番目の定理は，平滑化した統計的汎関数の漸近正規性についてである．

**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．$T$ を統計的汎関数として $\tau$ を ( 1 ) に基づいて $T$ から $D[0,1]$ 空間上に誘導された統計的汎関数とする．$\tau$ が $U$ において Hadamard 微分可能であり影響関数 IF 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，$\widetilde{\mathrm{IF}}$ が以下の Erdös と Rényi による Lindeberg 型の条件を満足すると仮定する:

任意の $\epsilon > 0$ に対して

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

ここで $P_\epsilon$ は以下の不等式を満たす $P=\{1,\ldots,N\}$ の部分集合とする

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

ここで $\overline{\widetilde{\mathrm{IF}}}$ は $\widetilde{\mathrm{IF}}$ の算術平均である.

すると $n, N-n\to\infty$ のとき

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

となる．ここで，$\sigma_{N,n}^2=N\sigma_N^2(1-n/N)/n(N-1)=(N-n)\sigma_N^2/n(N-1)$ 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，$\tau$ を関係式 ( 1 ) に基づいて $T$ から $D[0,1]$ 空間上に誘導された統計的汎関数とする．$\tau$ 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．

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

任意の $\epsilon>0$ に対して

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

ここで $P_\epsilon$ 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 $\overline{\mathrm{IF}}$ は IF の算術平均である.

そのとき $n, N-n\to\infty$ のとき

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

となる．ここで，$\sigma_{N,n}^2=N\sigma_N^2(1-n/N)/n(N-1)=(N-n)\sigma_N^2/n(N-1)$ である．

**各章の構成**

本論文の構成は以下である：

- 第 1 章 序論
- 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 統計的汎関数の漸近正規性
- 第 4 章 モンテ・カルロ・シミュレーション

第 1 章では，標本調査の意義，標本調査の理論的枠組みである有限母集団モデルと有限母集団モデルにおける漸近理論の取り扱いについて論じ，統計的汎関数の定義，平滑化経験分布関数とそれに基づいた平滑化統計的汎関数の定義を行い，本稿の主要結果である漸近正規性の結果について述べる．

第 2 章では，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を示す．本章の結果は主に，独立同一分布標本に対しての Fernholz(1991) の結果の非復元無作為抽出への拡張である．

第 3 章では，本稿の中心結果である，非復元無作為抽出のもと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を証明する．そのため統計的汎関数を関数空間上の Taylor 展開に基づき線型の部分と剰余項に分割する．本章では，関数空間上での微分の定義を述べ，線型項について漸近正規性の成り立つ条件を精査し，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示し，最終的に統計的汎関数の漸近正規性を証明している．本章の主要結果は，独立同一分布標本について Reeds(1976) および Fernholz(1993) が示した汎関数の中心極限定理の非復元無作為抽出への拡張である．

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，3 種類の大きさの異なる母集団に対し 2 種類の異なる抽出率で，メディアンと四分位偏差の 2 つの統計量に対し，正規近似で求まる信頼区間の妥当性を検証した．そして，母集団および標本のサイズが十分に大きいとき，正規近似が非常に精度の高いものであることを確認する．