【40】

氏名	秋仁善（ちゅう いん そん）
博士の専攻分野の名称	博士（理学）
学位記番号	第 10797 号
学位授与年月日	平成5年3月25日
学位授与の要件	学位規則第4条第1項該当 基礎工学研究科数理系専攻
学位論文名	**Nonparametric estimation of distribution function** **（分布関数のノンパラメトリック推定について）**
論文審査委員	（主査） 教授 白旗 慎吾 （副査） 教授 石井 恵一　教授 稲垣 宣生

論文内容の要旨

数理統計学のノンパラメトリック理論では，データの分布法則は未知として展開されているが，その未知な分布関数が推定できれば，効率の良いノンパラメトリック統計量が選択でき，したがって，データの持つ情報を無駄なく生かせるようになる。分布関数のもっとも標準的な推定量として経験分布関数がよく知られている。しかし推定したい分布関数が連続であると仮定したとき，その分布関数の連続性を無視している。そこで，従来用いられてきた標準的な方法に代って，近年，滑らかな核関数を用いる方法が脚光を浴び，理論的にその良さが証明されている。本論文では核関数による平滑化に基づいた推定量と平滑化されたブートストラップを用いる推定量に関する積分平均二乗誤差を中心として推定問題を考察する。

ある確率標本 $X_1, X_2, \cdots, X_n$ が未知の連続分布関数 $F(x)$ に従うとする。対応する経験分布関数を $F_n(x)$ とする。ここで，F_n は標本サイズ n 個の観測値の各点で確率 $1/n$ をもつ分布関数である。適当な分布関数 K を核関数とし，平滑化パラメータ h_n を用いる核型推定量（Kernel-type estimator）

$$\hat{F}_n(x)=\frac{1}{n}\Sigma K((x-X_i)/h_n)=\int_{-\infty}^{\infty}K((x-y)/h_n)\,dF_n(y)$$

が考えられる。h_n は積分二乗誤差を最小にする意味で適当な平滑化パラメータとする。核型推定量 $\hat{F}_n$ に対する様々な性質は多くの研究論文に発表されている。特に平滑化パラメータの推定問題に関してはいろいろな推定方法が考えられてきたが，分布関数の核型推定量は積分平均二乗誤差 $\int_{-\infty}^{\infty}E\{\hat{F}_n(x)-F(x)\}^2\,dF(x)$ の意味で経験分布関数を用いた推定方法より良いことが分かっている。ところが，積分平均二乗誤差の場合は平均して0となる項は無視されるが，もっと基本的な積分二乗誤差 $\int_{-\infty}^{\infty}\{\hat{F}_n(x)-F(x)\}^2\,dF(x)$ を用いた場合では正規分布に従う確率変数が表われ，積分平均二乗誤差の場合の平均して0となる項は確率的には無視できないことが分かる。本論文で，積分二乗誤差の意味では単純に経験分布関数によって推定を行うことと核型推定量を用いることに確率的には本質的な差がないことを示した。

次にブートストラップ法に基づいて一つの平滑化パラメータを選択する問題を考える。近年，経験分布関数を用いるブートストラップ推定方式よりも平滑化されたブートストラップ法に基づく推定量の方が良い場合かあることが何人かの著者によって明らかになった。本論文ではブートストラップ法を用いて最小の平滑化パラメータを選ぶことにする。すなわち，F_n^* を経験分布関数 F_n からのブートストラップ標本 $X_1^*, X_2^*, X, \cdots, X_n^*$ にもとづく経験分布関数とする。推定量 $\hat{F}_n$ のようにブートストラッピング核推定量（bootstrapping kernel estimator）

$$\hat{F}_n^*(x)=\frac{1}{n}\Sigma K(x-X_i^*)/h_n)=\int_{-\infty}^{\infty}K(x-y)/h_n)\,dF_n^*(y)$$

が構成できる。平滑化パラメータ h_n は推定量 $\hat{F}_n$ の場合と同じように選ばれる。しかし，この標準的なブートストラップ法は推定量のバイアス項 $E\hat{F}_n^*(x\mid x^*)-\hat{F}_n(x)$ が0になるため，ブートストラップ法の直接な応用はできない。それを避けるため別のアプローチを行なう。そのブートストラップ標本による適当な平滑化パラメータ g_n が選択できれば一つの核推定量 $\hat{F}_n(x)=1/n\Sigma_{i=1}^{n}K(x-X_i)/g_n$ が構成できる。したがって，バイアス項 $E\hat{F}_n^*(X)-\hat{F}_n^*(X)$ が把握でき，リサンプリング平滑化パラメータ g_n とともに誤差 $\hat{F}_n^*(x)-\hat{F}_n(x)$ に基づいて解析することによって $\hat{F}_n(x)-F(x)$ が推定できるようになる。このことからブートストラップ推定量 $\hat{F}_n^*$ の平均二乗誤差，積分平均二乗誤差，そして積分二乗誤差を評価することができる。本論文で，上の事実から最適のブートストラップ平滑化パラメータが選択可能であることを示した。

論文審査の結果の要旨

$X_1,\cdots,X_n$ を連続な分布関数 $F(x)$ からの無作為標本とする。統計的問題に対する多くのパラメトリックな接近では少数のパラメータを除き関数 $F(x)$ は既知とされるが，ノンパラメトリックな接近では $F(x)$ は未知として処理される。ただし，ノンパラメトリックな接近でも $F(x)$ が分かっていればより効率のよい手法が導かれ，データの持つ情報を無駄なく引き出すことが可能となる。したがって，$F(x)$ を効率よく推定することは重要な問題である。

本論文では分布関数 $F(x)$ の推定問題を考察している。第1章は導入部分である。

第2章では経験分布関数と Kernel 型推定量の比較を行う。従来分布関数の推定では経験分布関数 $F_n(x)=\Sigma I(X_i\leqq x)/n$ を用いるのが通常であった。ただし，経験分布関数は $F(x)$ が連続であるという情報を無視している。そこで Kernel 型推定量 $\hat{F}_n(x)=\Sigma\ K(\frac{x-X_i}{h_n})/n$ が提案された。ここで $K(x)$ は適当な正則条件を満たす連続な分布関数である。$h=h_n$ は0に収束する定数列であり，平滑化パラメータと呼ばれる。$\hat{F}_n(x)$ は積分平均2乗誤差 (IMSE)

$$\mathrm{IMSE}(\hat{F}_n)=\int E(\hat{F}_n(x))-F(x)^2)dF(x)$$

の意味では $F_n(x)$ より良いことがすでに多くの研究で示されている。ところが実際のデータに適用している現場では $\hat{F}_n(x)$ の優位性は疑問視されていた。そこで IMSE ではなく，平均をとらない積分2乗誤差 (ISE)

$$\mathrm{ISE}(\hat{F}_n)=\int(\hat{F}_n(x)-F(x))^2\,dF(x)$$

での比較を試みた。ISE は確率変数で表現した誤差であり，より正確な誤差解析が可能となる。本論文では Kernel 型が IMSE の意味で優位にある h に対しても $n(\mathrm{ISE}(\hat{F}_n)-\mathrm{ISE}(F_n))$ が漸近的に平均0の正規分布に従うことを示した。したがって，IMSE の意味で $\hat{F}_n(x)$ の優位は平均0のランダムな項を無視した結果であり，$\hat{F}_n(x)$ と $F_n(x)$ に本質的な差はないことを示した。したがって，推定量に連続性を要求すれば $\hat{F}_n(x)$，要求しなければ計算の容易さから $F_n(x)$ を用いることが合理的であることが分かる。

第3章では Kernel 型推定量における平滑化パラメータを選択する問題を扱う。Kernel 型推定量には Kernel 関数 $K(x)$ と平滑化パラメータ h が解析する者によって選択されるが，Kernel 関数 $K(x)$ は適当な広い範囲から選択しても推定の良さにはあまり影響せず，平滑化パラメータの選択が重要な問題である。平滑化パラメータを選択するためには誤差を評価する必要があるが，誤差は真の分布関数 $F(x)$ に関係するので，それを推定しなければならない。誤差 $\hat{F}_n^*(x)-F(x)$ を推定する方法としてブートストラップ法がよく知られている。ブートストラップ推定量 $\hat{F}_n^*(x)$ を用い，$\hat{F}_n^*(x)-\hat{F}_n(x)$ によって推定するのが標準的な方法である。ところが，$\hat{F}_n(x)$ は bias のある推定量であるが，この方法では bias が検出できない。そこでブートストラップのための平滑化パラメータ $g=g_n$ を導入し，それによる推定量 $\hat{F}_n^*(x)$ と推定量 $\hat{F}_n(x)$ の差 $\hat{F}_n^*(x)-\hat{F}_n(x)$ によって誤差が推定できることを示した。これにより，Kernel 型推定量の誤差が解析でき，かつ最良の平滑化パラメータが選択できる。

以上のように本論文は分布関数の推定理論とその応用に寄与するものであり，博士の学位論文として価値あるものと認める。