カーネル型推定量を用いた統計量の漸近的性質について

Asymptotic properties of kernel type estimators

数学専攻　中村 広大
NAKAMURA,Koudai

1　はじめに

本論文ではノンパラメトリックな状況の下での確率密度関数や分布関数の推定と関連する統計的推測及びカーネル型推測法について議論していく．他方，統計量の一つのクラスである L-統計量に組み込まれた分布関数をカーネル型分布関数推定量で置き換えた推定量の有効性を議論し，従来の方法との比較を行なっている．また，確率点 $Q(p)$ を滑らかに推定する L-統計量に類似する推定量について，漸近的な性質と有効性を理論的に求め，シミュレーションによりその推定精度について考察する．

2　カーネル型推定量

2.1　密度関数と分布関数のカーネル型推定

カーネル法はノンパラメトリックな設定の下での密度関数の推定法として Akaike(1954),Rosenblatt(1956) 等によって以下で述べる経験分布関数 $F_n(\cdot)$ をもとに提案され，関連する研究が行われている推定法である．ここで経験分布関数は確率変数 $X_1, X_2, \ldots, X_n$ を無作為標本とし

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leqq x)$$

と定義される関数である．このとき密度関数として区間幅 h を用いて

$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h}$$

が提案されている．この $f_n(x)$ を観測数 n に依存させた区間幅 h を選び，カーネル関数をを用いることで滑らかさを持つカーネル型推定量が提案され密度関数と分布関数カーネル型推定量として $\widehat{f}_n(x)$ と $\widehat{F}_n(x)$ が

$$\widehat{f}_n(x) = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) dF_n(y), \widehat{F}_n(x) = \int_{-\infty}^{x} \widehat{f}_n(u)du$$

と定義されている．

3　U-統計量

3.1　L-統計量とその部分クラス

確率変数 $X_1, X_2, \ldots, X_n$ を母集団分布 $F(\cdot)$ からの無作為標本とし $X_{1;n} \leqq X_{2;n} \leqq \cdots \leqq X_{n;n}$ を対応する順序統計量とする．このときこれらの順序統計量の線形和

$$L_n = \sum_{i=1}^{n} c_{i;n} X_{i;n}$$

を L-統計量と呼ぶ．また，スコア関数 $J(u),(0 \leqq u \leqq 1)$ を用いた L-統計量のサブクラスに属する $T(F)=\int_0^1 F^{-1}(u)J(u)du$ の推定量

$$T(F_n)=\int_0^1 F_n^{-1}(u)J(u)du$$

についても研究が行われている．ここで

$$\begin{aligned}
\tilde{I}(X_i;u)&=I(X_i;u)-F(u), k(u,t)=F(min(u,t))-F(u)F(t),\\
\eta&=-\frac{1}{2}\int_{-\infty}^{\infty} J^{(1)}(u)F(u)(1-F(u))du,\\
\alpha_1(X_i)&=-\int_{-\infty}^{\infty} J(u)\tilde{I}(X_i;u)du, \alpha_2(X_i,X_j)=-\int_{-\infty}^{\infty} J^{(1)}(u)\tilde{I}(X_i;u)\tilde{I}(X_j;u)du.
\end{aligned}$$

とすると，$T(F_n)$ の漸近表現は

$$T(F_n)-T(F)=n^{-1}\eta+n^{-1}\sum_{i=1}^{n}\alpha_1(X_i)+n^{-2}\sum_{C_{n,2}}\alpha_2(X_i,X_j)+n^{-1/2}o_L(n^{-1/2}).$$

となる．ここで $o_L(n^{-1/2})$ は $P\{|o_L(n^{-1/2})| \geqq cn^{-1/2}(\log n)^{-1}\}=o(n^{-1/2})$ を満たす残差項である．

3.2　提案する推定量

推定量 $T(F_n)$ に対して F_n をカーネル型推定量 $\widehat{F}_n$ に置き換えた推定量 $T(\widehat{F}_n)$ は

$$\begin{aligned}
W_i(u)&=W\left(\frac{u-X_i}{h}\right)-E(W), b_n(u)=E(W)-F(u),\\
g_1(W_i)&=-\int_{-\infty}^{\infty} J(u)W_i(u)du, g_2(W_i,W_j)=-\int_{-\infty}^{\infty} J^{(1)}(u)W_i(u)W_j(u)du
\end{aligned}$$

とおき，バンド幅 h を $n^{-1/3} \leqq h \leqq n^{-1/4}$ と設定すると以下の漸近表現が求まる．ここで定義より

$$T(\widehat{F}_n)-T(F)=n^{-1}\eta+n^{-1}\sum_{i=1}^{n}g_1(W_i)+n^{-2}\sum_{C_{n,2}}g_2(W_i,W_j)-\int_{-\infty}^{\infty}J(F(u))b_n(u)du+n^{-1/2}o_L(n^{-1/2}).$$

が成り立つ．本研究では提案する推定量 $T(\widehat{F}_n)$ の漸近平均二乗誤差が大きくなり，経験分布関数を用いた推定量に比べ精度が悪いことが分かった．この結果よりカーネル型推定を乱用により漸近平均二乗誤差が大きくなることもあることを示しており注意する必要がある．

4　確率点の推定

4.1　経験分布関数の逆関数を用いた確率点の推定量

確率変数 $X_1,X_2,\ldots,X_n$ を互いに独立で同じ分布 $F(\cdot)$ に従う確率変数とする. このとき，$X_{1:n} \leqq \cdots \leqq X_{n:n}$ と小さい方から並べ替えたものを順序統計量とよぶ．経験分布関数の逆関数を用いた確率点の推定量 $\widehat{Q}(p)=\inf\{x;F_n(x) \geqq p\}$ は経験分布関数の性質から $\widehat{Q}(p) \approx X_{[np]+1:n}$ である. また，一様分布からの順序統計量の性質から $X_{[np]+1:n} \approx F^{-1}(U_{[np]+1:n})$ となるので $\widehat{Q}(p)$ の漸近平均二乗誤差は

$$AMSE(\widehat{Q}(p))=\frac{p(1-p)}{f^2(F^{-1}(p))n}=\frac{Q'(p)^2p(1-p)}{n}$$

となる.

4.2　確率点のカーネル型推定量

適切なバンド幅 h とカーネル関数 $K(\cdot)$ を選び，確率点の推定量である $\widehat{Q}_{p,h}$ は Falk(1984) によって

$$\widehat{Q}_{p,h} = \frac{1}{h}\int_0^1 F_n^{-1}(x)K\left(\frac{x-p}{h}\right)dx = \sum_{i=1}^{n} v_{i,n}X_{i:n},\quad v_{i,n} = \frac{1}{h}\int_{\frac{i-1}{n}}^{\frac{1}{n}} K\left(\frac{x-p}{h}\right)dx$$

と定義され，L- 統計量であることが分かっている. この推定量の漸近表現から期待値と分散を求めるために以下の定義を用意する.

[**定義**] 確率変数 Y_i を互いに独立で同一の分布 $U(0,1)$ にしたがうものとする．このとき

$$\begin{aligned}
\bar{Q}(p) &= \frac{1}{h}\int_0^1 F^{-1}(x)K\left(\frac{x-p}{h}\right)dx, \hat{I}_x(Y_1) = I(Y_1 \leqq p+h) - (p+h),\\
g_{1n}(Y_1) &= -\int_{-1}^1 Q'(p+hx)K(x)\hat{I}_x(Y_1)dx, \sigma_n^2 = \mathrm{Var}(g_{1n}(Y_1)),\\
d_{1n} &= \sigma_n^{-1}n^{-1/2}, d_{2n} = \sigma_n^{-1}n^{-3/2}h^{-1}, d_{3n} = \sigma_n^{-1}n^{-5/2}h^{-2},\\
g_{2n}(Y_1,Y_2) &= -\int_{-1}^1 Q'(p+hx)K'(x)\hat{I}_x(Y_1)\hat{I}_x(Y_2)dx,\\
g_{3n}(Y_1,Y_2,Y_3) &= -\int_{-1}^1 Q'(p+hx)K^{(2)}(x)\hat{I}_x(Y_1)\hat{I}_x(Y_2)\hat{I}_x(Y_3)dx\\
\hat{g}_{1n}(Y_1) &= -\frac{1}{2}\int_{-1}^1 Q'(p+hx)K^{(2)}(x)E[\hat{I}_x^2(Y_2)]\hat{I}_x(Y_1)dx, \hat{A}_{1n} = \sum_{i=1}^n \hat{g}_{1n}(Y_1),\\
A_{1n} &= \sum_{i=1}^n g_{1n}(Y_i), A_{2n} = \sum_{C_{n,2}} g_{2n}(Y_i,Y_j), A_{1n} = \sum_{C_{n,3}} g_{3n}(Y_i,Y_j,Y_k)
\end{aligned}$$

と定義する．またサポートを (-1,1) にもつ m 次オーダーカーネル $K(x)$ が

$$\int_{-1}^1 K(x)dx = 1,\ \int_{-1}^1 x^iK(x)dx = 0, (i=1,\ldots,m-1),\ \int_{-1}^1 x^mK(x)dx \neq 0.$$

を満たすことと，バンド幅 h を $h = n^{-1/4}(\log n)^{-1}$ とする．このとき 4 次オーダーカーネルと $\delta = Q'(p)(\frac{1}{2} - p) + \frac{1}{2}Q^{(2)}(p)p(1-p)$ を用いて

$$\sigma_n^{-1}\sqrt{n}(\widehat{Q}_{p,h} - \bar{Q}(p)) = d_{1n}A_{1n} + d_{2n}A_{2n} + d_{3n}A_{3n} + d_{3n}(n-1)\hat{A}_{1n} + \frac{\delta}{\sigma\sqrt{n}} + o_L(n^{-1/2})$$

が成り立つことが分かっている.(Maesono and Penev (2011) 参照)

この表現より $\widehat{Q}_{p,h} - Q(p)$ は $\widehat{Q}_{p,h} - Q(p)$ をバイアス項 $d_n = \bar{Q}(p) - Q(p)$ を用いて

$$\widehat{Q}_{p,h} - Q(p) = \widehat{Q}_{p,h} - \bar{Q}(p) + d_n = \widehat{Q}_{p,h} - \bar{Q}(p) + o(n^{-1})$$

となる．ここで $W(x)$ と $L(x)$ を $W(x) = \int_{-1}^x K(t)dt, L(x) = \int_{-1}^x W(t)dt$ を定義し，計算に用いるカーネルを 4 次オーダーカーネル, バンド幅を $h = n^{-1/4}(\log n)^{-1}$ とすると,

$$AMSE(\widehat{Q}_{p,h}) = \frac{Q'(p)^2p(1-p)}{n} + \frac{2h}{n}Q'(p)^2\left(-\frac{1}{2} + \frac{1}{2}\int_{-1}^1 W^2(x)dx\right)$$

を得る．ここで $\widehat{Q}(p)$ の漸近平均二乗誤差と比較すると，$\widehat{Q}_{p,h}$ の漸近二乗誤差の二項目が適切なカーネルとバンド幅を選択することで負になる．よって漸近平均二乗誤差がより小さくなる $\widehat{Q}_{p,h}$ が確率点 $Q(p)$ の推定量として優れているのではないかと理論的に考察できる.

5 シミュレーション

5.1 各種分布の確率点の推定

5.1.1 適用プロセス

確率点 $Q(p)$ の推定量 $\widehat{Q}(p)$ と $\widehat{Q}_{p,h}$ の平均二乗誤差 (MSE) を比べ推定量としてどちらが優れているかを議論する．以下の手順で MSE の導出を行なっていく．まず，$\widehat{Q}(p)$ について説明する．
(1) ある分布からの n 個の乱数を大きさ順に並べた $X_{1:n},\ldots,X_{n:n}$ を用意する．
(2) 4.2 節より $\widehat{Q}(p)\approx X_{[np]+1:n}$ であるから n 個の内の $[np]+1$ 番目を $\widehat{Q}(p)$ とする．
(3) (2) で求めた $\widehat{Q}(p)$ に対して $Q(p)$ との二乗誤差 $(\widehat{Q}(p)-Q(p))^2$ を計算する．
(4) (1)〜(3) を 1000 回行う．一回の試行ごとに求めた二乗誤差の総和を試行回数 1000 で割った $MSE(\widehat{Q}(p))$ を得る．

次に $\widehat{Q}_{p,h}$ について説明する．
(1) ある分布からの n 個の乱数を大きさ順に並べた $X_{1:n},\ldots,X_{n:n}$ を用意する．ここは $\widehat{Q}(p)$ で用いたものを使用する．
(2′) 4.3 節より $\widehat{Q}_{p,h}$ の L- 統計量による表現を用いる．ここでバンド幅 h_n は $h_n=n^{-1/4}(\log n)^{-1}$, カーネル関数 $K(\cdot)$ はシミュレーション毎に決めるものとする．
(3′) (2′) で求めた $\widehat{Q}_{p,h}$ に対して $Q(p)$ との二乗誤差 $(\widehat{Q}_{p,h}-Q(p))^2$ を計算する．
(4) (1)〜(3’) を 1000 回行う．一回の試行ごとに求めた二乗誤差の総和を試行回数 1000 で割ったものを $MSE(\widehat{Q}_{p,h})$を得る．

6 おわりに

本研究では，$T(F)$ の推定について $T(F_n)$ と $T(\widehat{F}_n)$ の漸近平均二乗誤差による比較からカーネル型推定をすることで推定が悪くなることを確認した．また，$\widehat{Q}(p)$ と $\widehat{Q}_{p,h}$ の推定精度を漸近平均二乗誤差を調べることで理論的に比較し, カーネル型推定量を用いたものが推定としてより優れていることが分かった．さらに，シミュレーションにおいても理論的な比較と同様の結果を得た．今後の展望として，カーネル型推定量を用いた確率点の推定において今回使用した $\widehat{Q}_{p,h}$ の他に $\widehat{F}_n(x)$ を用いた推定量 $Q^*(p)=\inf\{x:\widehat{F}_n(x)\geqq p\}$ との推定の比較など行っていきたい．

参考文献

[1] Akaike, H.(1954) An approximation to the density function. *Ann. Inst.Statist. Math.,***6**,pp.127-132.

[2] Falk,M.(1984).Relative deficiency of kernel quantile estimators. *The Annals Statist,* 12(1),261-268.

[3] 前園宜彦.(2019) ノンパラメトリック統計. 共立出版

[4] Maesono, Y.(2007) An Edgeworth expansion and a normalizing transformation for L-statistics.Bulletin of Informatics and Cybernetics, Vol.39, pp.25-43.

[5] Maesono, Y. and Penev, S.(2011). Edgeworth expansion for the kernel quantile estimator. *Annals of the Institute of Statistical Mathematics.***63**, pp.617-644.

[6] Rosenblatt, M.(1956).Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist..,***27**, pp.832-837.