統計数理研究所　夏の大学院
統計思考院
2012年9月19日

アダプティブな切り取り型順位推定

三浦良造
統計数理研究所外来研究員
一橋大学名誉教授

アダプティブな切り取り型順位推定（位置母数推定）
Adaptive truncated rank estimates for location of symmetry.

背景となる理論。

：（first order efficiency):観測誤差の確率分布が既知であればそれに応じた最適な推定量が構成できる。漸近的に最小分散を達成する。

：(Robustness) :Contamination あるいはGross error modelとHuberのleast favorable distribution. 推定誤差の最大値を最小にするミニマックス解を与える。

適応型推定（三浦　1981年論文）

:上記のleast favorable distributionに応じたR-推定（切り取り型のスコアに基づく）を使う。

:上記の理解のもとで、未知の確率分布を"捕える"適応型＝「推定量候補の族」を用意する。

:Contaminationに関わる量を推定し、それを使って、族の中から一つを選ぶ。

:選び方は、推定された漸近分散が最小である推定量を選ぶ。漸近分散の推定には、順序統計量を使う。

:スコア=確率分布の関数形=推定量の最適性=（R-推定、M-推定、L-推定）

現在の問題意識との関連。

裾の問題。

:当時は裾の重さと位置母数推定のロバスト性に関心が集中した。現在の“金融工学”では、裾について;たとえば99%点の推定が論じられている。裾全体に対する定式化の議論はない？

:重い裾とその相互依存関係の世界をコピュラ関数で扱っているが、まだ発展途上段階の様相。これは2次元以上の話。

分布形

:当時は対称分布に関心が強かったが、現在は、ゆがみがある分布についても推論が盛ん？ ;skew symmetry.

:当時の基礎理論は、現在はどこに応用されているか？線形回帰モデルか？

統計分析ソフトでは、M-推定である切り取り型標本平均を線形回帰モデルに使ったロバスト回帰が見られるが、その他には？

観測の確率的独立性

:当時はiid(独立同一分布性)の仮定が主だったが、現在では、弱い依存性の仮定のもとで扱えるのではないか。

:有意義な応用の局面はあるのか？

今日の講義は、以下に挙げる三浦の3編の論文を参考にしています。
特に一つ目の論文の各所をコピーしてスライドに入れています。

: (1)."Adaptive Confidence Intervals for a Location Parameter"
Keiei Kenkyu, Vol.31, No.4.5.6, pp. 197-218, March 1981

: (2). "Spacing Estimation of the Asymptotic Variance of Trimmed Rank Estimators of Location"
Scandinavian Journal of Statistics, No.8, pp.48-54, 1981

: (3)."Spacing Estimation of the Asymptotic Variance of Rank Estimators" Proceedings of Golden Jubilee Conference of Indian Statistical Institute: Statistics; Applications and New Directions, pp.391-404, December 1981

ADAPTIVE CONFIDENCE INTERVALS FOR A LOCATION PARAMETER

Ryozo Miura

1. Introduction.

The nonparametric confidence interval for a location parameter which is derived from a rank test was introduced by Lehmann (1963). He derived it from the Wilcoxon's signed rank test and proved that the length of the derived confidence interval, multiplied by the square root of the sample size, is consistent with the product of two constants; One is the upper percentile point of the standard normal distribution and the other is the square root of the asymptotic variance of the point estimator derived from the Wilcoxon's signed rank test. Shorack (1970) extended the consistency result to the case where the confidence intervals are derived from general signed rank statistics.

In this paper we are concerned with the constructions of confidence intervals which work flexibly when the underlying distribution is vaguely known. We consider the situation which the gross error model in Huber (1964) describes with the arbitrarily fixed central distribution and the unknown amount of contamination, and construct the adaptive confidence interval which has a certain asymptotic optimality when the efficiency is measured in terms of the length of intervals.

まず、順位統計量から導かれる位置母数推定、信頼区間の構成について紹介する。

歴史的には、順位統計量は仮説検定に使われるために開発された。そこでは帰無仮説のもとでのパラメター値を統計量の中で使って、統計量の値がもっともらしい範囲にあるかどうかを観ていた。このパラメターの帰無仮説値を仮変数として動かせて、統計量がもっともらしい値をとるような仮変数値をパラメター推定値とすることで、推定量が定義できる。

順位統計量の確率分布は、観測誤差の確率分布関数形が未知であっても、求めることができる。これが長所である。また統計量の構成もこれを反映している。それが当時の“ノンパラメトリックス”である。

2. Signed Rank Statistics and Confidence Intervals.

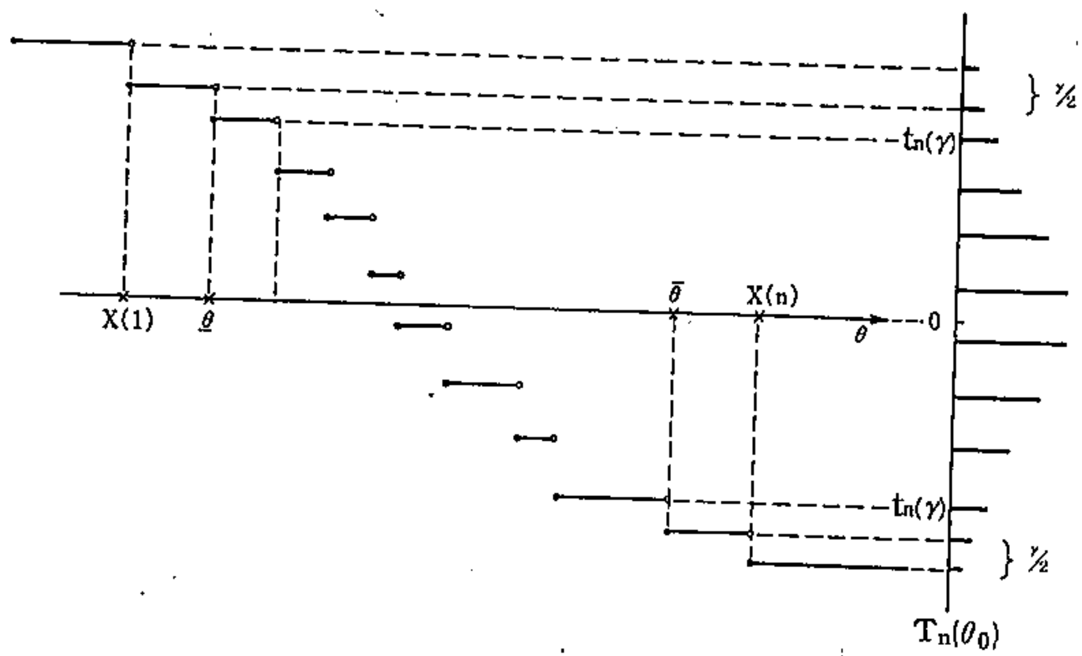
Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with the distribution function $F_0(\cdot - \theta_0)$, where θ_0 is the unknown location parameter to be estimated. Assume that F_0 is continuous and strictly increasing, and also that F_0 is symmetric about zero, i.e., $F_0(x) + F_0(-x) = 1$ for $-\infty < x < \infty$. F_0 is otherwise unknown. For $-\infty < \theta < \infty$ let $R_i(\theta)$ denote the rank of $|X_i - \theta|$ among $\{|X_j - \theta|, j = 1, 2, \ldots, n\}$, and sgn $(\cdot)$ denote the sign of the variable, i.e. sgn $(x) = 1$ or -1 according to $x > 0$ or $x \leqq 0$. Let $a(\cdot)$ be a score function, that is, a nondecreasing function defined on the unit interval $(0,1)$ with $a(0) = 0$. Assume $a(\cdot)$ is continuous and differentiable except at a finite number of points. The signed rank statistic evaluated at $|X_i - \theta|$, $i = 1, 2, \ldots, n$, or simply, evaluated at θ is defined as

$$\frac{1}{n} \sum a\left(\frac{R_i(\theta)}{n+1}\right) \operatorname{sgn}(X_i - \theta) \quad \left(\equiv T_n(\theta)\right)$$

We note that $T_n(\theta)$ is a nonincreasing step function of θ since $a(\cdot)$ is non-decreasing. The jumps occur at $\theta=(X_{(i)}+X_{(j)})/2$ with the amount of jumps $2\{a((j-i)/(n+1))-a((j-i+1)/(n+1))\}$ for $i \leqq j$, $i, j=1, 2, \ldots, n$. Here $X_{(i)}$, $i=1, 2, \ldots, n$ denote the order statistics of X_i, $i=1, 2, \ldots, n$.

In Figure 1, a typical graph of $T_n(\theta)$ and a distribution of $T_n(\theta_0)$ with $\underline{\theta}_n$,

Figure 1. The graph of $T_n(\theta)$ as a function of θ.
The distribution of $T_n(\theta_0)$.



Since the distribution of $T_n(\theta_0)$ is calculated under symmetry of F_0, we define

$$t_n(\gamma) \leftrightharpoons \inf \{x: P_{\theta_0}(T_n(\theta_0)>x) = \gamma/2\}$$

for a certain set of numbers γ in (0, 1). Noting that the distribution of $T_n(\theta_0)$ is symmetric about zero, we define

$$\underline{\theta}_n = \sup \{\theta: T_n(\theta) > t_n(\gamma)\}$$

$$\bar{\theta}_n = \inf \{\theta: T_n(\theta) < -t_n(\gamma)\}.$$

Since $P_{\theta_0}\{\theta_0 < \underline{\theta}_n\} = P_{\theta_0}\{\theta_0 \geqq \bar{\theta}_n\} = \gamma/2$ from our definition, we have $1-\gamma = P_{\theta_0}\{\underline{\theta}_n \leqq \theta_0 < \bar{\theta}_n\}$. The absolute continuity of $\bar{\theta}_n$ (See Hodges & Lehmann 1963, Theorem 1) implies $1-\gamma = P_{\theta_0}\{\underline{\theta}_n \leqq \theta_0 \leqq \bar{\theta}_n\}$ so that $[\theta_n, \bar{\theta}_n]$ is a $(1-\gamma)\times 100\%$ (non adaptive) confidence interval for θ_0.

順位統計量がとる値は、パラメター値、確率分布関数形に依らず、決まっているので、
（統計量の確率分布は真のパラメター値のもとで決めるとして）、
上側・下側の％点を達成する仮変数値を信頼区間の単店とする。

上記はスコア関数を一般的な形で扱っている。次に例として具体的なスコア関数を観る。

Gross error modelで、母体となる確率分布をLogistic分布とした場合のleast favorable 分布(密度関数 f)に対応するスコア関数である。

$$-\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}$$

For the above g and any ε such that $0<\varepsilon<1$, define g_ε, the density function of the Huber's least favorable distribution, as follows.

$$g_\varepsilon(x)=\begin{cases}(1-\varepsilon)\,g(-x_0)\,e^{k(x+x_0)} & \text{for}\quad x\leqq -x_0\\ (1-\varepsilon)\,g(x) & \text{for}\quad -x_0<x\leqq x_0\\ (1-\varepsilon)\,g(x_0)\,e^{-k(x-x_0)} & \text{for}\quad x_0<x\end{cases}$$

where $x_0>0$ and $k>0$. x_0 and k are determined by the relations

(i) x_0 and $-x_0$ are the endpoints of the interval where $|g'/g|\leqq k$

(ii) $\displaystyle\int_{-x_0}^{x_0} g(x)\,dx+\frac{2g(x_0)}{k}=\frac{1}{1-\varepsilon}$.

3.3. An Example.

For the sake of applications of the adaptive procedure, we illustrate the simplest (trimmed Wilcoxon) case and give some comments.

When G is logistic, i.e. $G(x)=(1+e^{-x})^{-1}$ for $-\infty<x<\infty$, $J(t)=2t-1$ and $a(t)=t$ for $0\leqq t\leqq 1$.

For $0<\varepsilon<1$, we have by definition,

$$J_\varepsilon(t)=\begin{cases}(2\alpha-1)/(1-\varepsilon) & \text{for } 0\leqq t\leqq\alpha\\ (2t-1)/(1-\varepsilon) & \text{for } \alpha<t\leqq 1-\alpha\\ (1-2\alpha)/(1-\varepsilon) & \text{for } 1-\alpha<t\leqq 1\end{cases}$$

and

$$a_\varepsilon(t)=\begin{cases}t/(1-\varepsilon) & \text{for } 0\leqq t\leqq 1-2\alpha\\ (1-2\alpha)/(1-\varepsilon) & \text{for } 1-2\alpha<t\leqq 1.\end{cases}$$

The statistics with these scores are called "trimmed Wilcoxon".

The relation of the constants α, ε, x_0 and k is given by the equations

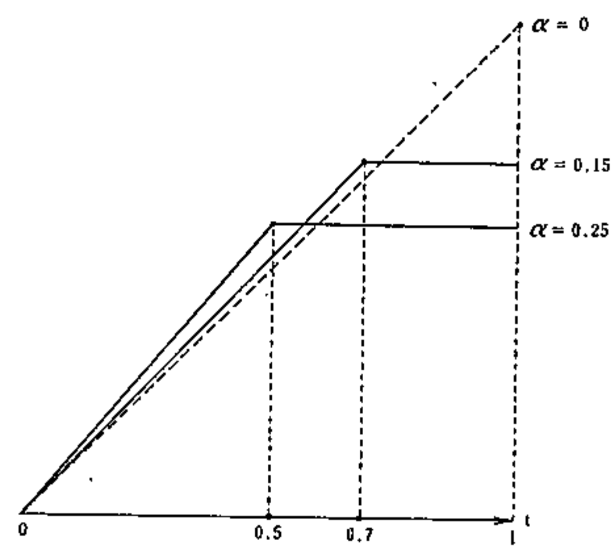
$$\varepsilon=(1-k)^2/(1+k^2), \quad x_0=\log((1+k)/(1-k))$$

and $k=\sqrt{(1-2\alpha)/(1+2\alpha)}$.

It is easily seen that when any one of the constants is given, it determines all the others uniquely. The above relation implies $1-\varepsilon=\sqrt{1-4\alpha^2}$. Therefore the trimmed score function for the present case is for $0<\alpha<1/2$,

$$a_\alpha(t)=\begin{cases} t/\sqrt{1-4\alpha^2} & \text{for } 0\leqq t\leqq 1-2\alpha \\ \sqrt{(1-2\alpha)/(1+2\alpha)} & \text{for } 1-2\alpha<t\leqq 1\,. \end{cases}$$

Figure 2. Graph of $a_\alpha(t)$.



α indicates the amount of trimming.

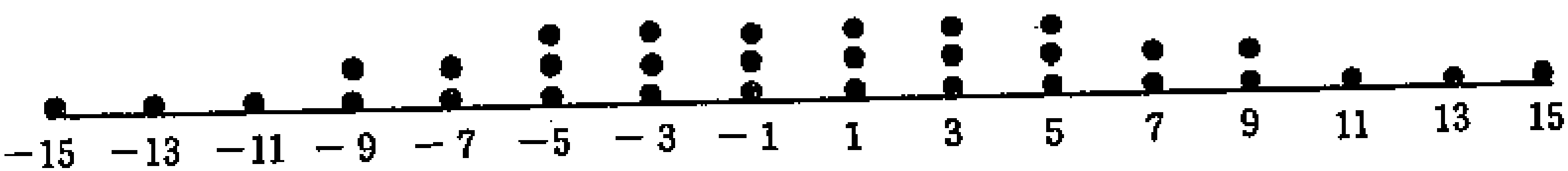
The graph of a_α is given in the Figure 2.

Figure 3 displays the probability distributions of $T_n(\theta_0)\times n$ in the two cases where the score functions are the above defined a_α with $\alpha=0$ and $\alpha=0.15$.

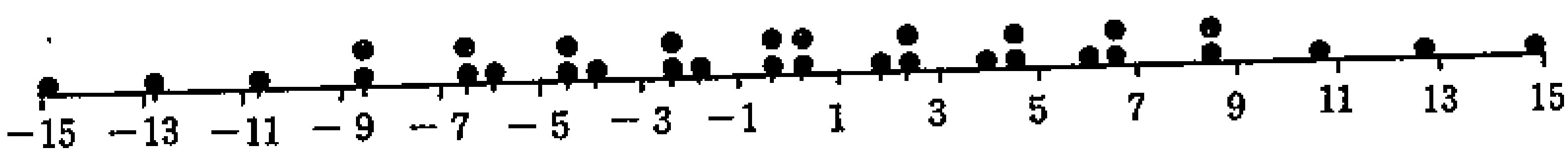
$$n+(\frac{n(n-1)}{2})$$:number of different values taken by Statistics

Figure 3. Distribution of $T_n(\theta_0)\times n$ $n=5$, $\alpha=0, 0.15$.

$\underline{\alpha=0}$



$\underline{\alpha=0.15}$ The absolute values which $T_n(\theta_0)\times n$ takes when $\alpha=0.15$ are 0.21 1.89 2.30 3.98 4.40 6.08 6.50 8.60 10.69 12.80 14.88.



数　学

第39巻 第3号　1987年7月 夏季号

日本数学会編集　岩波書店発売

$\bar{Y}_\alpha$ は古くから用いられていた推定量であるが，その理論的把握がここで初めてなされたと考えてよい．標本平均 $\bar{Y}$ は裾の重さ，粗大な誤差に影響されやすく $\bar{Y}_\alpha$ は $\bar{Y}$ をロバスト化したものとみられる．相対精度については

$$\inf\{ARE(\bar{Y}_\alpha, \bar{Y} : F) : F \text{ 対称分布}\} = (1-2\alpha)^2$$

であり，F を対称一山形に限れば右辺は $1/(1+4\alpha)$ である．

上の Φ を任意の G，ただし g を G の密度関数として $\log g(x)$ が Convex で滑らか，にとりかえても同様の議論ができ $\mathfrak{F}\{G : \varepsilon\}$ の least favorable distribution G_ε が一般的に定義される．ロジスティック分布 $G(x)=1/(1+e^{-x})$，$x\in R^1$ に対応する R-推定は前述の Wilcoxon 型推定 W (Hodges-Lehmann 推定とも呼ばれる.) である．W にも切り取り型 W_β が定義される．$J_{a_i}(\cdot)$ を用いて前述の手順で定義すれば

$$W_\beta = \text{median}\left\{\frac{Y_{(i)}+Y_{(j)}}{2} : [n\beta]+1 \leq i, j \leq n-[n\beta]\right\}$$

である．$\bar{Y}_\alpha$ と W_β の相対精度については

$$\inf\{ARE(W_\beta, \bar{Y}_\alpha : F) : F \text{ 対称}\}$$
$$= \begin{cases} 0.864\times B(\beta)/(1-2\beta)^2(1-2\alpha)^2(1+4\alpha), & 0 \leq \alpha \leq \beta < 1/2 \\ 0.864\times B(\alpha)/(1-2\beta)^2(1-2\alpha)^2(1+4\alpha), & 0 \leq \beta \leq \alpha < 1/2 \end{cases}$$

当時の先行研究

In the case of the special score function: $a(t) = t$ for $0 \leqq t \leqq 1$, Lehmann (1963) proved that $\sqrt{n}\,(\bar{\theta}_n - \underline{\theta}_n)$ converges in probability to $t_\gamma/\sqrt{3}\int_{-\infty}^{\infty} f_0^2(x)\,dx$ as $n \to \infty$, where f_0 is the density function of F_0 and t_γ is the $100\,\gamma/2$ upper percentile point of the standard normal distribution. Under the conjecture of Huber (1970), Antille (1972) proved under some regularity conditions that the distribution of $\sqrt{n}\,\{\sqrt{n}\,(\bar{\theta}_n - \underline{\theta}_n)/2t_\gamma - 1/\sqrt{12}\int f_0^2(x)\,dx\}$ converges as $n \to \infty$ to normal with mean zero and variance $(\int f_0^3(x)\,dx - (\int f_0^2(x)\,dx)^2)/(3(\int f_0^2(x)\,dx)^4)$.

In the case of a general score function, Shorack (1970) proved the consistency result, using the everypath argument of the uniformly convergent empirical process. The asymptotic normality in this case has not been proved yet in the literature. In this paper we are concerned with the consistency result which is stated in the following theorem. Let $H_0(x) = F_0(x) - F_0(-x)$ for all $x \geqq 0$. For any real number b, let $F(x) = F_0(x + b/\sqrt{n})$ for all x and also let $H(x) = F(x) - F(-x)$ for all $x \geqq 0$. Then F and H are the distribution functions of $X_i - (\theta_0 + b/\sqrt{n})$ and $|X_i - (\theta_0 + b/\sqrt{n})|$ respectively. Let F_n and H_n be the enpirical distribution functions of $X_i - (\theta_0 + b/\sqrt{n})$ and $|X_i - (\theta_0 + b/\sqrt{n})|$, $i = 1, 2, \ldots, n$, respectively, i.e.

$$F_n(x) = 1/n \sum_{i=1}^{n} \mathrm{I}\ \{|X_i - (\theta_0 + b/\sqrt{n})| \leqq x\} \quad \text{for all } x$$

and

$$H_n(x) = 1/n \sum_{i=1}^{n} \mathrm{I}\ \{|X_i - (\theta_0 + b/\sqrt{n})| \leqq x\} \quad \text{for all } x \geqq 0$$

where $I\{\bullet\}$ is the indicator function.

Theorem 1.

Suppose that for any $\varepsilon > 0$, $a(\bullet)$ is of bounded variation on $[0, 1-\varepsilon)$ and also that $a(t) \leqq \text{constant} \times (1-t)^{-\eta}$ and $a'(t) \leqq \text{constant} \times (1-t)^{-3\eta}$ for some $0 < \eta < 1/2$. Suppose also that for any fixed $B > 0$, we have

(*) $\sqrt{n}\left\{\int_0^\infty a(H(x))\, dF(x) - \int_0^\infty a(H_0(x))\, dF_0(x)\right\}$ converges

as $n \to \infty$ to $b\Delta$ uniformly in $|b| \leqq B$ for some $\Delta > 0$.

Then $\sqrt{n}(\bar{\theta}_n - \underline{\theta}_n)$ converges in probability as $n \to \infty$ to $t_r\left\{\int_0^1 a^2(1-u)\, du\right\}^{1/2}/\Delta$.

Shorack (1970) also remarked some regularity conditions under which (*) holds with

$$\Delta = 2\int_0^\infty a'(2F_0(x) - 1)\, f_0^2(x)\, dx.$$

$\Delta \equiv \Delta(a)$ is called the efficacy of the signed rank statistic with the score function a, or simply the efficacy of a.

As Lehmann (1963) suggests, this theorem enables us to compare the efficiencies of two different sets of confidence intervals when the efficiency is measured in terms of the length of intervals. Then the asymptotic relative efficiency of the two sets of intervals is the ratio of limits of their length, that is,

$$\left\{ \frac{\int_0^1 a_1^2(u)\, du}{\int_0^1 a_2^2(u)\, du} \right\}^{1/2} \times \frac{\Delta(a_2)}{\Delta(a_1)}$$

where a_1 and a_2 are score functions of statistics from which the confidence intervals and derived. The comparison can also be made for the confidence intervals other than those derived from rank statistics. The asymptotic relative efficiency of the confidence intervals is the same as that of the corresponding point estimators of the location of symmetry. The important fact is that when f_0 is the density function of the underlying distribution the signed rank statistic with the score function

$$a(t) = -\frac{f_0'(F_0^{-1}((t+1)/2))}{f_0(F_0^{-1}((t+1)/2))}$$

derives the asymptotically efficient confidence interval.

真の密度関数とスコア

繰り返しになるが、M-推定、L-推定, R-推定について、それぞれ、ベストなスコアは真の密度関数を使って表わせること、さらに、そこで得られる3者の推定量は漸近的に同等であること（ただし、first order efficiencyの意味で）が分かっている。

例えば、Huberの本を参照。ベストと云うのは、漸近分散が最小であるモノ(フィッシャー情報量を全うする)。

スコアを推定する、あるいは ある族の中からadaptiveに選ぶ

The next theorem is an extension of Theorem 1 in the direction that the score funtion be a random element. Let C be the class of score functions which are continuous on the closed interval [0, 1], and continuously differentiable except at a finite number of points. Let the derivative a' of $a \in C$ be defined such that $a'(t) = \{a'(t^-)+a'(t^+)\}/2$ at the point t where a is not differentiable. We assume that $\sup_{a \in C} \sup_t |a(t)|$ and $\sup_{a \in C} \sup_t |a'(t)|$ are both bounded. Since $a(t)$ is defined also at $t = 1$, we write the score by $a(R_i(\theta)/n)$ instead of $a(R_i(\theta)/(n+1))$. This causes no change on the results of Theorem 1.

まず一般論として、推定されたスコアの場合を扱う。
推定されたスコアを使った場合の位置母数推定は、推定スコアの収束先のスコアを使った位置母数推定に収束する、という命題。

Theorem 2.

Let $\hat{a}$ be a function of $X_1, \ldots, X_n$, and $\hat{a} \in C$. Assume that $\sup_{0 \leq t \leq 1} |\hat{a}(t) - a_0(t)| \xrightarrow{P} 0$ as $n \to \infty$ for some $a_0 \in C$. Assume also that C is uniformly equicontinuous on [0, 1] and that the class of a', $a \in C$ is uniformly equicontinuous on the differentiable intervals, i.e. for any $\varepsilon > 0$ there exists $\delta > 0$, which is independent of $a \in C$, such that $|s-t| < \delta$ implies $|a'(t) - a'(s)| < \varepsilon$ for any s, t in any one of the differentiable intervales of a, for all $a \in C$. We assume that F_0 has a bounded positive density on its support. Then $\sqrt{n}(\bar{\theta}_n(\hat{a}) - \underline{\theta}_n(\hat{a}))$ converges in probability as $n \to \infty$ to $t_r \left\{ \int_0^1 a_0^2(1-u)\, du \right\}^{1/2} / \Delta(a_0)$ where $\Delta(a_0) = 2\int_0^\infty a_0'(2F_0(x) - 1) f_0^2(x)\, dx$.

Proof. Let B be any fixed positive number.

For $\theta = \theta_0 + \dfrac{b}{\sqrt{n}}$,

$$
\begin{aligned}
\hat{T}_n(\theta) &\overset{\leftarrow}{=} \frac{1}{n}\sum_{i=1}^{n} \hat{a}\left(\frac{R_i(\theta)}{n}\right) \operatorname{sgn}(X_i-\theta) \\
&= \frac{1}{n}\sum_{i:X_i>\theta} \hat{a}\left(\frac{R_i(\theta)}{n}\right) - \frac{1}{n}\sum_{i:X_i<\theta} \hat{a}\left(\frac{R_i(\theta)}{n}\right) \\
&= 2\left\{\frac{1}{n}\sum_{i:X_i>\theta} \hat{a}\left(\frac{R_i(\theta)}{n}\right) - \frac{1}{2n}\sum_{i=1}^{n} \hat{a}\left(\frac{i}{n}\right)\right\} \\
&= 2\left\{\int_0^\infty \hat{a}(H_n(x))\, dF_n(x) - \frac{1}{2n}\sum_{i=1}^{n} \hat{a}\left(\frac{i}{n}\right)\right\} \\
&\overset{\rightarrow}{=} 2\,\hat{L}_n(b)\,.
\end{aligned}
$$

First we show that the statistic $\sqrt{n}\,\hat{T}_n(\theta)$ as a function of b coverges in probability to $2\,(Y_0+b\Delta)$ as $n\to\infty$ uniformly in $|b| \leqq B$ where Y_0 is the limit variable of $\sqrt{n}\left\{\int_0^\infty a_0(H_n(x))\, dF_n(x) - \int_0^\infty a_0(H(x))\, dF(x)\right\}$.

For simplicity we assume in the proof that each though we do not mention at each step. The method in the proof is that of Theorem 1, and the notations U, U_n, ρ, Γ_n, $\xrightarrow{\circ}$ are taken from Shorack (1972). The definitions of them are referred to the paper.

Now $\sqrt{n}\,\hat{L}_n(b) = I_n + II_n + III_n + IV_n$

where
$$I_n = \sqrt{n}\left\{\int_0^\infty \hat{a}(H_n(x))\,dF_n(x) - \int_0^\infty \hat{a}(H(x))\,dF(x)\right\}$$
$$-\sqrt{n}\left\{\int_0^\infty a_0(H_n(x))\,dF_n(x) - \int_0^\infty a_0(H(x))\,dF(x)\right\}$$
$$II_n = \sqrt{n}\left\{\int_0^\infty a_0(H_n(x))\,dF_n(x) - \int_0^\infty a_0(H(x))\,dF(x)\right\}$$
$$III_n = \sqrt{n}\left\{\int_0^\infty \hat{a}(H(x))\,dF(x) - \frac{1}{2n}\sum_{i=1}^{n} \hat{a}\left(\frac{i}{n}\right)\right\}$$
$$-\sqrt{n}\left\{\int_0^\infty a_0(H(x))\,dF(x) - \frac{1}{2n}\sum_{i=1}^{n} a_0\left(\frac{i}{n}\right)\right\}$$
$$IV_n = \sqrt{n}\left\{\int_0^\infty a_0(H(x))\,dF(x) - \frac{1}{2n}\sum_{i=1}^{n} a_0\left(\frac{i}{n}\right)\right\}.$$

I_n is decomposed into five parts, and each of them are examined in (i) through (v).

$$
\begin{aligned}
I_n = & \sqrt{n}\int_0^\infty \hat{a}(H(x))\, d(F_n(x)-F(x)) \\
& + \sqrt{n}\int_0^\infty \hat{a}'(H(x))\,(H_n(x)-H(x))\, dF(x) \\
& + \sqrt{n}\int_0^\infty \hat{a}'(H(x))\,(H_n(x)-H(x))\, d(F_n(x)-F(x)) \\
& + \sqrt{n}\int_0^\infty \{\hat{a}(H_n(x))-\hat{a}(H(x))-\hat{a}'(H(x))\,(H_n(x)-H(x))\}\, dF_n(x) \\
& - \sqrt{n}\left\{\int_0^\infty a_0(H_n(x))\, dF_n(x)-\int_0^\infty a_0(H(x))\, dF(x)\right\} \\
& \fallingdotseq I_{1,n}+I_{2,n}+I_{3,n}+I_{4,n}-I_{5,n}\,.
\end{aligned}
$$

$$\text{(i)}\quad I_{1,n} = \int_0^\infty \hat{a}(H(x))\, dU_n(F)$$

$$= \hat{a}(1)\, U_n(1) - \hat{a}(0)\, U_n(F(0)) - \int_0^\infty U_n(F(x))\, d\hat{a}(H(x))$$

$$= -\int_0^\infty U_n(F(x))\, \hat{a}'(H(x))\, dH(x)\,.$$

Since

$$\left| \int_0^\infty U(F(x)\, \hat{a}'(H(x)\, dH(x) - \int_0^\infty U_n(F(x))\, \hat{a}'(H(x))\, dH(x) \right|$$

$$\leqq \rho\,(U, U_n) \int_0^\infty \hat{a}'(H(x))\, dH(x)$$

$$= \rho\,(U, U_n) \int_0^1 \hat{a}'(t)\, dt$$

$$\leqq \rho\,(U, U_n)\, M_C \xrightarrow{\theta} 0 \quad \text{as } n \to \infty$$

where $M_C = \sup\,\{a'(t);\, t \in [0,1],\, a \in C\}$, we have

$$I_{1,n} - \left\{ -\int_0^\infty U(F(x))\, \hat{a}'(H(x))\, dH(x) \right\} \xrightarrow{\theta} 0 \quad \text{as } n \to \infty.$$

証明に使われる数学についての注意（1）

U_n がUに収束すること。

[0,1]区間上を一様分布に従う、独立同一分布性を持つ確率変数から作られるU_n のUへの収束は、独立同一分布性を同一分布に従うが、しかし、独立ではなくて、弱い依存性(strongly mixing)を持つ場合にも証明されている。Shao&Yu(1996), Louhichi(2000)を見よ。

この1996年、2000年の結果を用いると、U_n のUへの収束を用いる上記の定理を弱い依存性を持つ場合に拡張できるのではないか。

(これが今年の「数学」の論稿に書いてある趣旨です)

In the proof of Theorem 2, we can also see the asymptotic normality of the adaptive point estimator $\hat{\theta}_n(\hat{a})$ of θ_0 which is defined as the minimizer of $|T_n(\theta)|$, or equivalently of $\sqrt{n}\,(\hat{\theta}_n(\hat{a})-\theta_0)$ which is defined as the b which minimizes $|\hat{L}_n(b)|$. We can further see in the proof that $\sqrt{n}\,(\bar{\theta}_n(\hat{a})-\bar{\theta}_n(a_0)) \xrightarrow{P} 0$ as $n\to\infty$.

3. Adeptive Confidence Intervals.

An adaptive statistical procedure in general consists of a family $C = \{s_\lambda, \lambda \epsilon \Lambda\}$ of simple (non-adaptive) procedures and a selection rule (or an estimator) $\hat{\lambda}$ of λ which is a function of observations. This procedure may well be applied when it is certain that one of the members in C has an optimal performance, but not certain of which one.

3.1. Family of trimmed score functions.

When F_0 is only approximately known, a gross error model $F_0=(1-\varepsilon)G+\varepsilon H$ can be considered, where G is a known distribution, H is an unknown distribution and ε is a number such that $0 \leqq \varepsilon \leqq 1$ and indicates the amount of contamination by gross errors. The adaptive procedure in this section is concerned with this ε. The construction of a family of trimmed score function is due to Huber (1964) and Jaeckel (1971 a).

Let G be a symmetric distribution function. Assume that G has a convex support and a twice continuously differentiable density function g such that $-\log g$ is convex on the support of G.

For the above g and any ε such that $0<\varepsilon<1$, define g_ε, the density function of the Huber's least favorable distribution, as follows.

$$g_\varepsilon(x)=\begin{cases}(1-\varepsilon)\,g(-x_0)\,e^{k(x+x_0)} & \text{for} \quad x\leqq -x_0\\ (1-\varepsilon)\,g(x) & \text{for} \quad -x_0<x\leqq x_0\\ (1-\varepsilon)\,g(x_0)\,e^{-k(x-x_0)} & \text{for} \quad x_0<x\end{cases}$$

where $x_0>0$ and $k>0$. x_0 and k are determined by the relations

(i) x_0 and $-x_0$ are the endpoints of the interval where $|g'/g|\leqq k$

(ii) $\displaystyle\int_{-x_0}^{x_0} g(x)\,dx+\frac{2g(x_0)}{k}=\frac{1}{1-\varepsilon}$.

The score function which is best for g_ε is the trimmed score function.

In correspondence to the above g_ε, define

$$a_{g_\varepsilon}(t) = J_{g_\varepsilon}\left(\frac{t+1}{2}\right) \quad \text{for } 0 \leqq t \leqq 1$$

where

$$J_{g_\varepsilon}(u) = -\frac{g_\varepsilon'(G_\varepsilon^{-1}(u))}{g_\varepsilon(G_\varepsilon^{-1}(u))}$$

$$= \begin{cases} -k & \text{for } 0 \leqq u \leqq \alpha \\ -\dfrac{g'(G_\varepsilon^{-1}(u))}{g(G_\varepsilon^{-1}(u))} & \text{for } \alpha < u \leqq 1-\alpha \\ k & \text{for } 1-\alpha < u \leqq 1 . \end{cases}$$

and $\quad \alpha = \int_{x_0}^{\infty} g_\varepsilon(x)\, dx.$

G_ε is the distribution function of the density g_ε and relates to G by

$$G_\varepsilon^{-1}(t) = G^{-1}\left(\frac{2t-\varepsilon}{2(1-\varepsilon)}\right) \quad \text{for } 0 \leqq t \leqq 1 .$$

We note that $\quad 2G(-x_0) = 1 - \dfrac{1-2\alpha}{1-\varepsilon}$.

We write a_ε or a_α instead of a_{g_ε}, and J_ε or J_α instead of J_{g_ε} for the notational simplicity.

Let α_1, α_2 be any numbers such that $0<\alpha_1<\alpha_2<1/2$. Let $C = \{a_\alpha, \alpha \in [\alpha_1, \alpha_2]\}$.

Trimmed scoreを使った場合の推定量の推定誤差の漸近分散を推定する。

Trimmingの量の範囲を決めて、推定量の族を用意する。対応する推定誤差分散の推定値が最小であるTrimming量に対応する切り取り型推定を選ぶ。

（漸近分散の推定については,三浦論文(2),(3)を参照）

3.2. Selection rules.

For all $\alpha \in [\alpha_1, \alpha_2]$, the asymptotic length of the confidence interval derived from the signed rank statistic with the score function a_α, multiplied by $\sqrt{n}$, is $t_\gamma \left\{\int_0^1 a_\alpha^2(u)\, du\right\}^{1/2} / \Delta(a_\alpha)$ by Theorem 1. Our interest is to choose α which minimizes $\sigma(\alpha) \leftrightharpoons \left\{\int_0^1 a_\alpha^2\, du\right\}^{1/2} / \Delta(\alpha)$, where $\Delta(\alpha) \equiv \Delta(a_\alpha)$ for simplicity. Let α_0 be such that $\sigma(\alpha_0) = \underset{\alpha \in [\alpha_1, \alpha_2]}{\text{minimum}}\, \sigma(\alpha)$. Note that such α_0 exists by the continuity of $\sigma(\alpha)$ on the compact set $[\alpha_1, \alpha_2]$. Since the numerator is known, $\Delta(\alpha)$ is the quantity to be estimated. The unknown factor in $\Delta(\alpha)$ is F_0.

Schweder (1975) proposed the estimator of $\Delta(\alpha)$

$$\hat{\Delta}_1(\alpha) = \frac{1}{n^2} \sum_i \sum_k J_\alpha'\left(\frac{i}{n}\right) \frac{1}{h(n)}\, W\left(\frac{X(i)-X(k)}{h(n)}\right)$$

based on the window method. Another estimator

$$\hat{\Delta}_2(\alpha) = \frac{1}{n} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} J_\alpha'\left(\frac{i}{n}\right) \frac{2\delta/n}{X(i+\delta)-X(i-\delta)}$$

based on the spacing method was proposed by Miura (1976, 1980). It was proved that $|\hat{\Delta}_2(\alpha) - \Delta(\alpha)| \xrightarrow{P} 0$ as $n \to \infty$, uniformly in $\alpha \in [\alpha_1, \alpha_2]$ under the stronger conditions on a and F_0.

2. Asymptotic distribution for a fixed score function

Let, $Y_i = \dfrac{\{U(i+\Delta)-U(i-\Delta)\}}{2\Delta/n}$ for $i = K+1, \ldots, n-K$.

Lemma 2 : $sup\{|Y_i-1| : i = K+1, \ldots, n-K\} \leqslant O_p\left(\Delta^{-\frac{1}{2}}\left(log\ \dfrac{n}{\Delta}\right)^{\frac{1}{2}}\right).$

Proof : By an application of Lèvy's model (see, e.g., Ito and Mckean, 1965, page 36) and Theorem 4.5.2. of Csörgö and Révész (1981), where the definition of B_n also appears; we have

$$\begin{aligned}
Y_i-1 &= \frac{\sqrt{n}}{2\Delta}\left[\sqrt{n}\left\{U(i+\Delta)-\frac{i+\Delta}{n+1}\right\}-B_n\left(\frac{i+\Delta}{n+1}\right)\right] \\
&\quad -\frac{\sqrt{n}}{2\Delta}\left[\sqrt{n}\left\{U(i-\Delta)-\frac{i-\Delta}{n+1}\right\}-B_n\left(\frac{i-\Delta}{n+1}\right)\right] \\
&\quad +\frac{\sqrt{n}}{2\Delta}\left\{B_n\left(\frac{i+\Delta}{n+1}\right)-B_n\left(\frac{i-\Delta}{n+1}\right)\right\}-\frac{1}{n+1} \\
&\leqslant \frac{\sqrt{n}}{2\Delta}\times O(n^{-\frac{1}{2}}\log n)+\frac{\sqrt{n}}{2\Delta}\times O\left(n^{-\frac{1}{2}}\Delta^{\frac{1}{2}}\left(\log\frac{n}{\Delta}\right)^{\frac{1}{2}}\right) \\
&\leqslant O\left(\Delta^{-\frac{1}{2}}\left(\log\frac{n}{\Delta}\right)^{\frac{1}{2}}\right).
\end{aligned}$$

証明に使われる数学についての注意（2）

上記のΔ（α)の推定には、順序統計量（の差）が使われている。順序統計量の差の収束は、iidの場合には、当時よく研究された。

この命題を弱い依存性がある場合に扱える可能性が見えている。Biao Wuの論文では、弱い依存性がある場合の、Bahadur表現を扱っている。

Δ_1（α) 、Δ_2（α)における順序統計量の使い方は異なるが、前者の方が扱いやすいように見受けられる。

Let $\hat{\alpha}$ be such that $\hat{\sigma}(\hat{\alpha}) = \underset{\alpha \in [\alpha_1, \alpha_2]}{\text{minimum}} \ \hat{\sigma}(\alpha)$ where $\hat{\sigma}(\alpha) \leftleftarrows \left\{ \int_0^1 a_\alpha{}^2 \, du \right\}^{1/2} / \hat{\Delta}(\alpha)$. Since $\sigma(\alpha)$ is continuous in α, Lemma 3 of Jaeckel (1971 b) implies $\hat{\alpha} \xrightarrow{P} \alpha_0$ as $n \to \infty$, and hence $\sup_{0 \leq t \leq 1} | a_{\hat{\alpha}}(t) - a_{\alpha_0}(t) | \xrightarrow{P} 0$ as $n \to \infty$. Therefore, by Theorem 2, we have the following statement.

Theorem 3.

Under the assumptions of Theorem 1 in Miura (1980) and Theorem 2 in the section 2, we have for any γ, $\sqrt{n}\ (\bar{\theta}_n(a_{\hat{\alpha}}) - \underline{\theta}_n(a_{\hat{\alpha}})) \xrightarrow{P} t_\gamma \cdot \sigma(\alpha_0)$ as $n \to \infty$. In fact $\sqrt{n}\ (\bar{\theta}_n(a_{\hat{\alpha}}) - \bar{\theta}_n(a_{\alpha_0})) \xrightarrow{P} 0$ and $\sqrt{n}\ (\underline{\theta}_n(a_{\hat{\alpha}}) - \underline{\theta}_n(a_{\alpha_0})) \xrightarrow{P} 0$ as $n \to \infty$.

Theorem 3 means that when the best α is unknown, but is known to lie between α_1 and α_2, the adaptive procedure is asymptotically as good as knowing and using the best α. The adaptive procedure is also an asymptotic minimax solution to Huber's problem with the general G, and is asymptotically efficient when $F_0 \equiv G_{\alpha 0}$.

We may obtain an adaptive confidence interval in a direct manner. For each n we have a finite number of α's which provides different signed rank statistic and hence different confidence intervals. Then we pick up, as the adaptive confidence interval, the one which has the minimum length among the finite number of confidence intervals. This is apparently the best for all n and all $\omega \in \Omega$. This procedure, however, requires an enormous amount of computation. Theorem 3 yields also that our adaptive confidence interval based on the estimation of efficacy Δ, which requires far less computation, is asymptotically as good as the all-the-time best confidence interval as well.

It was presented in Miura (1979) that

$$\hat{\Delta}_2(\alpha, \underline{X+a}) = \hat{\Delta}_2(\alpha, \underline{X}) \qquad \text{for all real numbers } a$$

and

$$\hat{\Delta}_2(\alpha, \underline{X/b}) = \frac{1}{b}\,\hat{\Delta}_2(\alpha, \underline{X}) \qquad \text{for } b \neq 0$$

This means that the change of shift and scale of the observations does not affect the selection $\hat{\alpha}$ of α, i.e.

$$\hat{\alpha}(\underline{X+a}) = \hat{\alpha}(\underline{X}) \text{ and } \hat{\alpha}(\underline{X/b}) = \hat{\alpha}(\underline{X}).$$

Therefore, the adaptive confidence interval is shift and scale equivariant, i.e.

$$[\underline{\theta}_n(\hat{\alpha}, \underline{X+a}), \bar{\theta}_n(\hat{\alpha}, \underline{X+a})] = [\underline{\theta}_n(\hat{\alpha}, \underline{X})+a, \bar{\theta}_n(\hat{\alpha}, \underline{X})+a]$$
for all a

and $$[\underline{\theta}_n(\hat{\alpha}, \underline{X/b}), \bar{\theta}_n(\hat{\alpha}, \underline{X/b})] = [\frac{1}{b}\,\underline{\theta}_n(\hat{\alpha}, \underline{X}), \frac{1}{b}\,\bar{\theta}_n(\hat{\alpha}, \underline{X})]$$
for all $b \neq 0$.

今後の試み

:一般論として、線形回帰モデルの偶然誤差項を古典的粗大誤差モデルで表わして、回帰母数推定を切り取り型順位推定により行う(切り取り型標本平均型は現在のソフトウエアにも入っているが、それは)M-推定である）。応用の具体例としては、ロジスティック分布を母体とするleast favorable分布を使うとよい。

:上記は偶然誤差項が対称分布に従うとしている。これを歪みがある分布として扱うには、一般化されたレーマン対立仮説モデル(変換モデル)を導入するのも一案である(このモデルは明日のトピックです)。Fをleast favorable分布として、適応型にすることも原理的には可能ではないか。

: iidの仮定を弱い依存性の仮定に置き換えたい。

:有意義な応用の局面はあるのか？

切り取り型標本平均と切り取り型ウイルコクソンの漸近分散の比較

三浦論文(3)を参照。

インド統計研究所創立50周年大会の論文集に収録されている。

Adaptive rank estimators of ξ. Let J_α be the score function corresponding to the Huber's least favorable distribution when the central distribution under contamination is logistic, (see Huber, 1964) i.e. for $0 \leqslant \alpha < \frac{1}{2}$

$$J_\alpha(t) = \begin{cases} (2\alpha-1)/\sqrt{1-4\alpha^2} & \text{for } 0 \leqslant t \leqslant \alpha \\ (2t-1)/\sqrt{1-4\alpha^2} & \text{for } \alpha < t \leqslant 1-\alpha \\ (2\alpha-1)/\sqrt{1-4\alpha^2} & \text{for } 1-\alpha < t \leqslant 1. \end{cases}$$

The rank estimator of ξ based on this score function is called a trimmed Wilcoxon estimator. Denote it by W_α. The form of its asymptotic variance is reduced to

$$\sigma^2(W_\alpha) = (1-2\alpha)^2(1+4\alpha) \Big/ \left\{12 \int_\alpha^{1-\alpha} f(F^{-1}(t))dt\right\}^2.$$

Comparisons with the adaptive trimmed mean. Let $\bar{X}_\alpha$ denote the trimmed mean and $\bar{X}_{\hat{\alpha}}$ denote the Jaeckel's adaptive trimmed mean.

C1. Define the asymptotic efficiency of W_α with respect to $\bar{X}_\beta$ by the ratio of their asymptotic variances.

$$e_F(W_\alpha, \bar{X}_\beta) = \frac{\sigma_F^2(\bar{X}_\beta)}{\sigma_F^2(W_\alpha)} = \left\{ \int_\beta^{1-\beta} \{F^{-1}(t)\}^2 dt + 2\beta\{F^{-1}(\beta)\}^2 \right\} \left\{ \int_\alpha^{1-\alpha} f(F^{-1}(t))dt \right\}$$
$$\times 12/(1-2\beta)^2(1-2\alpha)^2(1+4\alpha).$$

It may be worthwhile to state here the following result by Miura (1976) since it is not well known. Its proof is in the spirit of Hodges and Lehmann (1956) and Bickel (1965), and is omitted here.

Let $\mathscr{F}$ be the class of all the symmetric (about zero) distributions satisfying the regularity conditions for the asymptotic normality of the estimates of location. Then,

$$\sup\{e_F(W_\alpha, \bar{X}_\beta) : F \in \mathscr{F}\} = \infty$$

$$\inf\{e_F(W_\alpha, \bar{X}_\beta) : F \in \mathscr{F}\} = \begin{cases} 0{\cdot}864 \times B(\beta)/(1-2\beta)^2(1-2\alpha)^2(1+4\alpha) \\ \qquad\qquad \text{for } 0 \leqslant \alpha \leqslant \beta < \frac{1}{2} \\ 0{\cdot}864 \times B(\alpha)/(1-2\beta)^2(1-2\alpha)^2(1+4\alpha) \\ \qquad\qquad \text{for } 0 \leqslant \beta \leqslant \alpha < \frac{1}{2} \end{cases}$$

where

$$B(t) = \left\{1+4t+\frac{4}{3} \times \sqrt{3t(t+1)}\right\}\left\{8t^2+1-\frac{2}{3} \times (1+4t)\sqrt{3t(t+1)}\right\}^2$$

$$\text{for } 0 \leqslant t < \tfrac{1}{2}.$$

By a numerical tabulation of the lower bound, we find that for each given β the lower bound is maximized at $\alpha = \beta$ and this maximum value increases strictly to one as $\alpha = \beta$ increases to $\frac{1}{2}$. We find the same thing on the other way around except for small values of α ($< 0{\cdot}04$) where the maximum is attained around $\beta = 0{\cdot}04$.

ご静聴ありがとうございました。