Stein タイプの縮小推定量 とその応用

慶応大学総合政策学部 篠崎 信雄

Stein-type Shrinkage Estimators and their Applications

Nobuo SHINOZAKI

要 旨 多くの母平均を同時に推定するとき、平均2乗誤差の和を基準にすれば、Stein 推定量が通常の推定量を改良することはよく知られている。ここでは、Stein 推定量をさまざまな現実の問題に応用するために考えられてきた Stein タイプの縮小推定量について、その基本的考え方を明らかにし、どのような現実の問題に対して有効であり、どのような問題があるのかを論じる。また、経験的ベイズ推定量、平滑化との関連性をも明らかにし、信頼領域の問題についても触れる。さらに、応用例として都道府県庁所在都市の一世帯当りの平均教育費を推定する問題を取り上げ、回帰直線に向けて縮小する推定量とその有効性を示す。

1. はじめに

いま p 個の母数 $\theta_1, \theta_2, \dots, \theta_p$ を推定したいとき、それぞれの θ_i について、平均 θ_i 、分散 1 の正規分布 $N(\theta_i, 1)$ に従う確率変数 X_i が観測されるものとし、さらに、 $X_1, X_2, \dots, X_p$ は統計的に独立であるとする。(後で述べるように分散の値は同じ値ならばなんでも構わないし、さらに、各 $N(\theta_i, 1)$ から同じ大きさの標本が得られる場合には標本平均を X_i と考えればよい。) そのとき、 θ_i の自然な推定量は X_i 自身であり、これは最小分散不偏推定量であり、最尤推定量でもある。個別に、 X_i に基づいて θ_i を推定するという問題を考え、平均2乗誤差 (MSE)

$$E\{(\hat{\theta}_i - \theta_i)^2\}$$

を基準にするならば、 X_i よりも一様に (θ_i の値がどんな値のときも) MSE の小さい推定量は存在しないことが知られている。(つまり、2乗誤差損失の下で X_i は θ_i の許容的な推定量である。) ところが、MSE の和 (これをリスクとも呼ぶ)

$$R(\theta, \hat{\theta}) = \sum_{i=1}^p E\{(\hat{\theta}_i(X) - \theta_i)^2\}$$

Key words: Confidence region, Educational expenditure, Empirical Bayes estimator, Simultaneous estimation, Smoothing.

を基準にして、 $\theta = (\theta_1, \theta_2, \dots, \theta_p)^t$ (t は転置を表す) を同時に推定することを考えると、 $p \geq 3$ のとき

$$\delta(X) = \left\{ 1 - \frac{p-2}{\|X\|^2} \right\} X, \quad (1.1)$$

ここで $\|\cdot\|$ は通常のユークリッドノルムを表す、が自然な推定量 $X = (X_1, X_2, \dots, X_p)^t$ を改良することが James and Stein (1961) により示された。つまり、

$$R(\theta, \delta) < R(\theta, X), \quad \text{すべての } \theta \text{ に対して}$$

が成立することが示されたのである。推定量 (1.1) が有名な Stein 推定量であり、 $\{\cdot\}$ の中が非負ならば) 原点に向けて縮小する形になっている。 $\theta_1, \theta_2, \dots, \theta_p$ に何の関係も無いにも拘らず、3 つ以上の推定問題を一緒にすると、個々には改良できない推定量が改良されてしまい、しかも、 θ_i の推定量に一見何の情報を持たない X_j , $j \neq i$ がからんでくるという意味で、驚くべき結果であり、パラドクスとも呼べるものである。($p=2$ の場合には X よりも一様にリスクの小さい推定量は存在しないことが Stein (1956) により示されている。) しかしその後 Stein タイプの縮小推定量について、Efron と Morris の一連の論文を含め、おびただしい数の研究が発表され、応用上も大きな意味のあるものであることが認識されるようになり、実用にも供されている。(例えば Morris (1983) 等参照)

Stein タイプの縮小推定量の議論は、改良の基準、さまざまな推定量のクラス、正規分布以外の分布の場合についての議論を含め、いろいろな方向に発展している(例えば, Berger (1985), Sen, Kubokawa and Saleh (1989), Berger and Robert (1990), Brandwein and Strawderman (1990) 等を参照)。特に、多次元の場合(平均行列の推定)への拡張(Bilodeau and Kariya (1989), Konno (1990, 1991), Honda (1991))を含め、日本からの貢献も少なくない(例えば, Akaike (1977, 1980), Yuan-Tsung Chang (1982), Akai (1986), Shiraishi (1991) 等)。とてもすべてを述べるのは無理なので、ここでは MSE の和という基準の下での正規分布の平均の同時推定問題だけを取り上げる。位置母数の同時推定問題にまで枠組を広げても類似の結果が成立するので、正規分布の前提は必ずしも本質的ではないことだけは述べておこう。(Brown (1966), Shinozaki (1984), Brandwein and Strawderman (1990) 等参照)

2 節では、Stein タイプの縮小推定量の基本的性質と、どのような問題に応用できるのかについて述べ、3 節で経験的ベイズ (Empirical Bayes) 推定量としてあるいは平滑化 (Smoothing) の手法と見ることができることについて述べる。4 節では、応用に供されるために Stein 推定量に対しどのような修正変更が考えられているか、5 節では、応用上のいくつかの問題点について触れる。6 節では 1 つの応用例を与える。数学的詳細は省略し、考え方だけを紹介するにとどめる。

2. 基本的性質と問題の性格

ここでは Stein 推定量についての基本的結果とそれが応用されうる推定問題の特徴についてまとめとておく。

2.1. 経験的ベイズ推定量としての Stein 推定量

1 節で述べたように $p \geq 3$ のとき Stein 推定量は X を改良するが、この現象を説明する議論は

いくつかある (竹内 (1979), Stigler (1990), Brandwein and Strawderman (1990)), ここでは Stein 推定量が経験的ベイズ推定量として導かれることだけを述べておこう。

いま仮に, θ_i が独立に $N(0, \tau^2)$ に従うというベイズ的仮定を置く. そのときリスク $R(\theta, \hat{\theta})$ の θ についての期待値

$$E_{\theta}\{R(\theta, \hat{\theta})\}$$

を最小にする θ_i のベイズ推定量は事後分布の平均

$$\left\{1 - \frac{1}{1 + \tau^2}\right\} X_i, \quad i = 1, 2, \dots, p \quad (2.1)$$

で与えられる. τ^2 が未知であるとする, これを $X_1, X_2, \dots, X_p$ に基づいて推定することが考えられる. X_i は周辺分布としては, 独立に $N(0, 1 + \tau^2)$ に従うから, $\sum_{k=1}^p X_k^2 / (1 + \tau^2)$ が自由度 p の χ^2 分布に従うことがわかるので,

$$E\left\{\frac{1}{\sum_{k=1}^p X_k^2}\right\} = \frac{1}{(p-2)(1 + \tau^2)}$$

であるから, (2.1) において $1/(1 + \tau^2)$ を $(p-2)/\left(\sum_{k=1}^p X_k^2\right)$ で置き換えれば (1.1) の Stein 推定量が得られる.

2.2. 基本的事実

上のことから, $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$ が 0 に近いという“関連性”があるときに Stein 推定量は大きな改良を与えることが予想される. 実際, 通常の推定量 X のリスクが $\sum E\{(X_i - \theta_i)^2\} = p$ であるのに対し,

$$\sum_{i=1}^p E\{(\delta_i(X) - \theta_i)^2\} = p - E\left\{\frac{(p-2)^2}{\|X\|^2}\right\}$$

であることが示される. これは $\|\theta\|^2$ を通してのみ $\theta_1, \theta_2, \dots, \theta_p$ の関数であり, $\|\theta\|^2 = 0$ のとき 2 になるので, 特に p の値が大きければ大変大きな改良であるということになる. $\|\theta\|^2$ が増加すると, 単調に増加し p に近づいていく. (図 1 参照)

なお, Stein 推定量 (1.1) に現われる定数 $(p-2)$ は次のような意味で最良の選び方になっている. 一般に

$$\left\{1 - \frac{b}{\|X\|^2}\right\} X$$

という形の推定量を考えたとき, リスクを一様に最小にするのが $b = p-2$ であるのである.

ところで, $\|X\|^2 < p-2$ のとき, Stein 推定量は 0 を超えて X の符号を変えたものになる. これは不自然なので, $a^+ = \max(0, a)$ とするとき

$$\left\{1 - \frac{p-2}{\|X\|^2}\right\}^+ X$$

が用いられる. これを positive-part Stein 推定量と呼ぶが, 確かに Stein 推定量を改良するもの

図 1. $p=10$ のときのリスク $R(\theta, \delta)$

になっている。なお、Takada (1979) はその 1 つの特徴づけを与えており、Kubokawa (1991) は別の改良方法を与えている。

また、分散の値を 1 としたが、一般に σ^2 という既知の値であるとすれば、Stein 推定量は

$$\left\{1 - \frac{(p-2)\sigma^2}{\|X\|^2}\right\} X$$

と修正される。さらに、 σ^2 の値が未知の場合でも、 X と独立に $\sigma^2 \chi_n^2$ として分布する確率変数 S が観測されるとするならば

$$\left\{1 - \frac{p-2}{n+2} \frac{S}{\|X\|^2}\right\} X$$

とすればよい。これも c を定数としたとき $\{1 - cS/\|X\|^2\}X$ という形の推定量の中で最良のものである。従って、 σ^2 の値が既知であるかどうかは本質的な問題ではなく、既知とするとその値も議論の本質に影響しない。以降では、断らないかぎり σ^2 の値を 1 であるとしておく。

2.3. Stein タイプの縮小推定量が有効な推定問題

ここで、どのような状況において Stein タイプの推定量が有効に応用されうるかについて整理しておこう。

(1) 多くの同種類の量を同時に推定する。

形式的に考えれば、 $X_1, X_2, \dots, X_p$ は北海道のじゃがいも、青森のりんご、 $\dots$ 、愛媛のみかんの重量の測定値というようになんの関連性もないものであってもよい。しかし (3) でも触れるように、何らかの関連性が事前に想定できる場合にこそ Stein 推定量は有効であり、その意味で同じ種類の量についての同時推定という状況に応用されるものと考えるべきだろう。多くの地区について同じ特性値が観測される、時系列的にデータが得られる、さまざまなカテゴリーの患者群について処理効果を測る等々いろいろな場面が考えられる (Beslow (1990))。その意味で、回帰係数の同時推定 (これも変数変換を行えば既に議論した問題に帰着できる) に Stein タイプの推定量を適用するというのは、必ずしも適当でない場合もあるだろう。

(2) 元の推定量の分散が小さくない。

そもそも、 X_i の分散が小さい場合には、 θ_i の推定量としての X_i を用いることで満足してしまい、他の推定量を考えようもしないだろう。たとえ Stein タイプの推定量を用いたにしても、 X_i の分散が小さければ、大きな改良になるような $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$ の範囲は狭いものになり、有効ではない。分散が小さくないからこそ、不偏性を破ってまで改良の方向を探るのだと言えるだろう。

(3) ある種の先験情報が存在する。

Stein 推定量 $\delta(X)$ は θ が 0 に近いときに大きな改良を与える。その意味で θ が 0 に近いだろうと考えられる場合に使うべき推定量である。もしも θ が θ^0 という値に近いと考えるならば、 $X - \theta^0$ を改めて X と置いて Stein 推定量を作ればよい。つまり、positive-part をとれば

$$\theta^0 + \left\{ 1 - \frac{p-2}{\|X - \theta^0\|^2} \right\}^+ (X - \theta^0)$$

を用いることになるだろう。より一般に、 X から何らかの構造を引き去った量が 0 に近いという形の情報を取り入れた Stein タイプの推定量については次で述べるが、いずれにしても何らかの情報がなければ有効に用いることは困難であろう。

3. 経験的ベイズ推定量および平滑化法としての Stein 推定量

2.1 でも Stein 推定量を経験的ベイズ推定量として導いたが、3.1 ではより一般的に議論し、応用上も重要な例を与える。3.2 では平滑化法としての視点から Stein タイプの推定量をとらえる。

3.1. 経験的ベイズ推定量としての Stein タイプの縮小推定量

5.1 での議論の都合もあるので、まず一般的に、 X_i は $N(\theta_i, \sigma_i^2)$ に従い、 $X_1, X_2, \dots, X_p$ は独立であるという状況を考える。ここで σ_i^2 は既知の値とする。さらに θ_i は $N(y_i'\beta, \tau^2)$ に独立に従っているものと仮定する回帰モデルを想定してみよう。ここで、 $\beta = (\beta_1, \beta_2, \dots, \beta_m)'$ は m 次元の未知ベクトル、 $y_i = (y_{i1}, y_{i2}, \dots, y_{im})'$ は既知の共変量ベクトルであり、分散 τ^2 の値も未知であるとする。もしも、 β と τ^2 が既知であるならば、 θ_i の (2 乗誤差損失の下での) ベイズ推定量は事後分布の平均であり

$$\hat{\theta}_i^B = y_i'\beta + \left\{ 1 - \frac{\sigma_i^2}{\sigma_i^2 + \tau^2} \right\} (X_i - y_i'\beta), \quad i=1, 2, \dots, p \quad (3.1)$$

で与えられる。

β と τ^2 の値が未知であるとすれば、 X_i が周辺分布として、独立に $N(y_i'\beta, \sigma_i^2 + \tau^2)$ に従うことを基にして β と τ^2 の推定量を求め、それを (3.1) に代入し θ_i の経験的ベイズ推定量を得るという方法が考えられる。しかし σ_i^2 の値が異なる場合には、簡単に表現できる β と τ^2 の推定量を得ることは困難であり、反復計算により方程式を解くといった手続きが必要になる。(Berger(1985) 等参照。また、6 節でも多少触れる。なお、経験的ベイズ法の考え方自身はもっと複雑な場合でも適用できることを注意しておく。)

σ_i^2 の値が同じであるならば、簡単に表現される β と τ^2 の推定量を代入することで経験的ベイズ推定量を書き下すことができる。表現を簡単にするため $\sigma_i^2 = 1$ とし、 Y を y_i' を第 i 行とする p

$\times m$ 行列とすれば, β を $\hat{\beta} = (Y'Y)^{-1}Y'X$ で推定し, $p-m \geq 3$ とするとき $1/(1+\tau^2)$ を $(p-m-2)/\|X-Y\hat{\beta}\|^2$ で推定することにより

$$\hat{\theta} = Y(Y'Y)^{-1}Y'X + \left\{1 - \frac{p-m-2}{\|X-Y(Y'Y)^{-1}Y'X\|^2}\right\}(X - Y(Y'Y)^{-1}Y'X) \quad (3.2)$$

を得る.

Y の列ベクトルの張る空間を V とし, V への正射影行列を P とおけば, 推定量 (3.2) は

$$PX + \left\{1 - \frac{p-m-2}{\|(I-P)X\|^2}\right\}(I-P)X$$

と表される. これは X を V と V の直交補空間に属する2つの成分に分け, V に属する成分については通常の不偏推定量 PX を用い, V の直交補空間に属する成分については原点へ縮小する Stein 推定量 (1.1) を用いて推定するという形になっている. 見方を変えれば,

$$\left\{1 - \frac{p-m-2}{\|(I-P)X\|^2}\right\}X + \frac{p-m-2}{\|(I-P)X\|^2}PX$$

と表されるように, X と PX の重みつきの平均になっているとも言える.

$p-m \geq 3$ のとき推定量 (3.2) は X を改良するものとなっている. 共通の $\sigma_i^2 = \sigma^2$ の値が1でない場合, さらに, σ^2 の値が未知であるが, X と独立に $\sigma^2\chi_n^2$ として分布する統計量が観測される場合の議論は Stein 推定量 (1.1) と全く同様である.

例 1. V を $1=(1, 1, \dots, 1)'$ の生成する1次元の部分空間とし, $\bar{X} = \sum X_i/p$ とおけば, $p \geq 4$ のとき推定量 (3.2) は

$$\bar{X} + \left\{1 - \frac{p-3}{\sum (X_i - \bar{X})^2}\right\}(X_i - \bar{X}), \quad i=1, 2, \dots, p$$

となる. つまり X_i を総平均 $\bar{X}$ へ向けて縮小する推定量が得られる. (Lindley(1962)) これは $\theta_1, \theta_2, \dots, \theta_p$ が同じような値であろうと考えられるときに用いることのできる推定量である.

例 2. X_{ijk} が通常の繰り返しのある2元配置実験から得られる観測値であり, $X_{ijk} \sim N(\theta_{ij}, \sigma^2)$, $i=1, \dots, a, j=1, \dots, b, k=1, \dots, n$ とする. 交互作用はあっても小さいものだろうと予想される場合ならば“交互作用に対応する部分空間”の成分について0に縮小することが考えられる. つまり, $(a-1)(b-1) > 2$ のとき, 次の推定量を用いるのである.

$$\begin{aligned} \hat{\theta}_{ij} = \bar{X}_{i..}\bar{X}_{.j.} - \bar{X}... + \left\{1 - \frac{S}{ab(n-1)+2} \frac{(a-1)(b-1)-2}{\sum_{ij} (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}...)^2}\right\} \\ \cdot (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}...) \\ i=1, \dots, a, \quad j=1, \dots, b \end{aligned}$$

ここで, $S = \sum_{ijk} (X_{ijk} - \bar{X}_{ij.})^2$ は $\sigma^2\chi_{ab(n-1)}^2$ として分布していることに注意する.

以上の議論では空間を2つの直交する部分空間の直和に分解し, 一方だけに Stein 推定量を適用したが, 3つ以上の部分空間に分解し, さらに, 個別に Stein 推定量を適用することも考えられる.

3.2. 平滑化法 (スミージング) としての Stein タイプの縮小推定量

3.1 の議論において, もしも θ が $Y\beta$ と表せることに自信があるならば,

$$X = Y\beta + e$$

という通常の回帰模型について分析を行うことに帰着してしまう。その意味で Y の列ベクトルの張る空間に近づける Stein 推定量を用いることは、次のような状況において、データに基づき適応的に折衷案を採用することになっている：一方で何の構造も仮定せずに X_i 自身でその平均を推定するのではばらつきが大き過ぎるが、他方で回帰構造を想定し、それに偶然変動が加わったものとして X_i の挙動を説明することには完全な自信も無く大きな偏りを生じることが心配されるという状況である。そのような場合に、データが模型に適合していればその模型の構造に大きく近づけ、適合していなければ X をあまり動かさないというように、データに基づき2つを折衷して推定値を決めるのである。したがって、Stein 推定量を用いるのは、ラフな推定量 (X) を滑らかな解 (回帰平面) に向けて縮小することで平滑化することになっていると見ることができる。これは 3.1 で見たように Stein タイプの縮小推定量が経験的ベイズ推定量と見ることができ、経験的ベイズ法が平滑化の一つの有力な考え方である (柏木, 柳本 (1990)) ことから当然のことである。しかも MSE の和という基準で見ると、決してラフな推定量より悪くないという保証が与えられているのである。

しかし Stein タイプの縮小推定量の中には、3.1 で述べた推定量以外に平滑化法としての色合いの濃いものもある。Stein (1981) は 3 項移動平均の重みを決定する問題を、通常の推定量 X よりもよいものになるように次の形の縮小推定量を構成するという問題に定式化している。

$$\hat{\theta}_i = \begin{cases} \{1 - \lambda(X)\} X_i + \lambda(X) (X_{i-1} + X_{i+1}) / 2, & i \neq 1, p \\ \{1 - \lambda(X) / 2\} X_1 + \lambda(X) X_2 / 2, & i = 1 \\ \{1 - \lambda(X) / 2\} X_p + \lambda(X) X_{p-1} / 2, & i = p \end{cases}$$

ここで、 $p \times p$ 行列 $A = (a_{ij})$ を

$$a_{ij} = \begin{cases} 1/2, & i = j = 1 \text{ または } i = j = p \\ 1, & i = j \neq 1, p \\ -1/2, & |i - j| = 1 \\ 0, & |i - j| > 1 \end{cases}$$

とするとき、 $p \geq 5$ ならば $(\text{tr} A)I - 2A$ が正値定符号であることが示されるので

$$B = \{(\text{tr} A)I - 2A\}^{-1} A^2$$

とおくとき、 $\lambda(X) = 1/(X'BX)$ である。 $p \geq 5$ のとき $\hat{\theta}_i$ は、MSE の和という基準の下で X よりもよい推定量になることが示されている。

さらに Li and Hwang (1984) は Stein (1981) の議論を基にして、ノンパラメトリック回帰模型

$$X_i = f(u_i) + e_i, \quad u_i \in [0, 1], \quad i = 1, 2, \dots, n$$

について consistent であり、かつ、生のデータ $X = (X_1, \dots, X_n)'$ よりよい推定量であるという意味でロバストな推定量を構成している。

4. Stein 推定量の変形

Stein 推定量の MSE の和を小さくするという意味で通常の推定量を改良するものであるが、そのままでは、いくつかの点で望ましくない性質を持つ。その一つは個々の θ_i の推定量としては大きな MSE になりうるということである。(Efron and Morris (1972), Rao and Shinozaki (1979)) さらに、MSE の和が小さく元の推定量 X を大幅に改良するような θ の領域も必ずしも広いものとは言えないということも上げられる。ここでは、これらの点を克服すべく考えられている工夫について述べる。

4.1. 移動量を制限した (limited translation) 推定量

Stein 推定量がある θ_i の推定量として大きな MSE をもつのは、 $|\theta_i| = \|\theta\|$ が成り立ち、さらに $\|\theta\| = \sqrt{p}$ のときであって、 p が大きい場合その最大の MSE は $p/4$ 程にも大きくなることが知られている。(Efron and Morris (1972)) この欠点を克服するため Efron and Morris は Stein 推定量を修正して、どの座標も通常の推定量 X_i の値からあるあらかじめ指定された値以上は縮小しないという条件をつけた推定量を提案している。経験的ベイズの考え方により導かれているので、そのアイデアは明らかには見てとれないが、提案された推定量は次のようなものである：

$$\left\{1 - \frac{p-2}{\|X\|^2} f_D \left((p-2) \frac{X_i^2}{\|X\|^2} \right)\right\} X_i, \quad i=1, 2, \dots, p \quad (4.1)$$

ここで

$$f_D(u) = \min \left\{ 1, \frac{D}{\sqrt{u}} \right\},$$

D は $0 \leq D \leq \sqrt{p-2}$ なる数で、 D の値を適当に選ぶことにより、個々の θ_i の推定量としての最大の MSE をコントロールすることになる。もちろん、 θ の推定量としてのリスクは Stein 推定量に比べ大きくなるのであるが、それほど大きくせず、個々の θ_i の最大の MSE を大幅に減少することができる。しかし推定量 (4.1) は滑らかな関数ではないので、 X を改良するものになるとは必ずしも言えない (Shinozaki (1986))。

一方、Stein (1981) は順序統計量に基づいて、別のタイプの移動量を制限した推定量を提案しているが、これについては次の 4.2 で触れる。(Dey and Berger (1983) も参照)

4.2. さまざまな縮小パターン

4.2.1. 打ち切り型の Stein 推定量

Stein 推定量 (1.1) が通常の推定量 X を大幅に改良する θ の領域は (分散の値によって変わるが) 必ずしも広いものとは言えない。リスクの改良分が $(p-2)^2 E(\|X\|^2)$ で与えられることからわかる通り、 θ_i の中に 1 つでも絶対値の大きなものがあれば、全体として大きな改良は望めない。(推定値自身が X のそれとほとんど変わらなくなってしまう。)

Stein (1981) の与えた打ち切り型の推定量は、このような状況に対処する工夫としても位置づけることができる。 $Z_i = |X_i|$, $i=1, 2, \dots, p$ とし、 $Z_1, Z_2, \dots, Z_p$ を大きさの順序に並べて得られる順序統計量を $Z_{(1)} < Z_{(2)} < \dots < Z_{(p)}$ で表す。 s を $3 \leq s < p$ なる整数とするとき、提案された推定量は

$$\delta(X) = X - g(X),$$

ここで

$$\begin{aligned} g_i(X) &= \frac{s-2}{\sum_k (X_k^2 \wedge Z_{(s)}^2)} X_i, \quad |X_i| \leq Z_{(s)} \\ &= \frac{s-2}{\sum_k (X_k^2 \wedge Z_{(s)}^2)} Z_{(s)} \operatorname{sign} X_i, \quad |X_i| > Z_{(s)} \end{aligned}$$

であり, $u \wedge v = \min(u, v)$ である. つまり, 絶対値の大きい $(p-s)$ 個については修正量を計算する上では大きな影響を与えないように, $Z_{(s)}$ × 符号で置き換えてしまうのである. このとき

$$E \|\delta(X) - \theta\|^2 = p - E \left\{ \frac{(s-2)^2}{\sum_k (X_k^2 \wedge Z_{(s)}^2)} \right\}$$

であることが示されるので, 多少 $((p-s)$ 個ほど) 絶対値の大きな θ_i があるときでも, ある程度の改良が確保されることになると言える. また, 個々の θ_i の推定量としての最大の MSE を小さくする効果も合わせ持っていることに注意しておく. s の選択については, 明確な方法はないが, p が大きいときの指針となる結果が Stein (1981), Dey and Berger (1983) に与えられている.

4.2.2. 適応型 Stein 推定量

Stein タイプの推定量を有効に用いるためには, θ が存在すると考える領域で大きな改良を与えるように推定量を作らなければならない. もちろん 3.1 の (θ が存在すると考えられる) 回帰平面に近づく縮小推定量もその 1 つである. しかし先験的な情報は, もっと漠然としたものであるかもしれない, さまざまな可能性がありうる. 3.2 で述べた移動平均へ向けて縮小する推定量もその 1 つであるとも言える.

George (1986a) は, それに向けて縮小すべき点あるいはより一般には部分空間の候補が複数個あるとき, それぞれに近づく Stein 推定量の重み付きの和として表される推定量を提案している. 原点に縮小すべきか総平均に向けて縮小すべきかあるいはまたある説明変数の 1 次式として説明できる部分を取り去ってから縮小を考えるべきかという問題は 1 つの例である. いま, K 個の異なる部分空間 $V_1, \dots, V_K$ があり, それぞれの直交補空間への正射影行列を $Q_1, \dots, Q_K$ とする. $q_k = \operatorname{rank} Q_k \geq 3, k=1, \dots, K$ とする. V_k にむけて縮小する Stein 推定量は

$$\delta^{(k)}(X) = (I - Q_k)X + \left\{ 1 - \frac{q_k - 2}{\|Q_k X\|^2} \right\} Q_k X, \quad k=1, 2, \dots, K$$

であるが, 提案された推定量は

$$\delta(X) = \sum_{k=1}^K \rho_k(X) \delta^{(k)}(X),$$

ここで

$$\rho_k(X) = \frac{w_k \|Q_k X\|^{-(q_k-2)}}{\sum_{i=1}^K w_i \|Q_i X\|^{-(q_i-2)}}$$

である. $w_k \geq 0$ は "重み" であり先験的な情報を反映して定められることになる.

$\delta(X)$ は $\|Q_k X\|^2$ が小さい, つまり改良分が大きいと考えられる $\delta^{(k)}(X)$ に対する重み $\rho_k(X)$ が大きくなるようになっており, θ が $V_1, \dots, V_K$ のどれかに近ければ意味のある改良を与える.

さらに, George (1986b) はいくつの変数のグループをまとめて Stein 推定量を作るべきか,

それとも個別に Stein 推定量を作るべきかという問題 (Efron and Morris (1973), Berger and Dey (1983) 参照) も取り扱うことができるように上の議論を拡張している。また, Akai (1989) もいくつかのグループに分類されるような場合に適応的に縮小する推定量を与えている。

5. いくつかの問題点

ここでは Stein タイプの縮小推定量を応用する際に, 一様な改良に固執すると, 大きな改良を望めない場合があること, さらに, 点推定に付随すべき信頼領域の設定が困難な問題であることについて述べる。

5.1. 一様な改良と経験ベイズ性

いま 3.1 の初めでも述べた状況を考える, つまり, X_i ($i=1, \dots, p$) が独立に $N(\theta_i, \sigma_i^2)$ に従うものとする。ここで σ_i^2 の値は既知であるとするが, それらの値が等しいと仮定してはいないことに注意する。仮に, θ_i ($i=1, \dots, p$) が独立に $N(0, \tau^2)$ に従うとすれば, ベイズ推定量は (3.1) の特別の場合 (y/β をすべて 0 にしてしまえばよい) として

$$\hat{\theta}_i^B = \left\{ 1 - \frac{\sigma_i^2}{\sigma_i^2 + \tau^2} \right\} X_i, \quad i=1, \dots, p$$

で与えられる。3.1 でも述べたように X_i の周辺分布を基にして τ^2 の推定値を得ることは可能であるので, それを $\hat{\tau}^2$ とすれば

$$\left\{ 1 - \frac{\sigma_i^2}{\sigma_i^2 + \hat{\tau}^2} \right\} X_i, \quad i=1, \dots, p \quad (5.1)$$

が経験ベイズ推定量である。

ところで, X_i が独立に $N(\theta_i, \sigma_i^2)$ に従うとき, $\sum_{i=1}^p E\{(\hat{\theta}_i - \theta_i)^2\}$ (重みなしの MSE の和) を基準にしたとき, 改良となる推定量のクラスはいろいろ与えられているが (Berger (1982) 参照), $p \geq 4$ になるとそのクラスには自由度が増し, 簡単には述べきれない (Bhattacharya (1966), Shin-ozaki (1974))。しかし, 例えば Berger (1976) の与えている推定量は

$$\left\{ 1 - \frac{p-2}{\sum_{k=1}^p X_k^2 / \sigma_k^2} \frac{1}{\sigma_i^2} \right\} X_i, \quad i=1, \dots, p \quad (5.2)$$

であり, 本質的には, 縮小する割合は相対的に σ_i^2 の大きいものほど小さくなるという性質を持つ。これは (5.1) が σ_i^2 が大きいほど縮小する割合が大きいのに対し, 全く逆である。さらに, 推定量 (5.2) は, σ_i^2 の中に 1 つでも飛び離れて小さい値のものがあると, 縮小する割合は小さくなってしまい, 推定値は X_i のそれとほとんど変わらなくなってしまう傾向があり, 従って改良分はとるに足らないものになってしまうということになる。

つまり, 等分散ではない場合には, 経験的ベイズ推定量と一様な改良を与える Stein タイプの縮小推定量には食い違いが生じ, Stein タイプの縮小推定量は, 分散のアンバランスがひどければ, とるに足らない改良しか与えないということである。確かに一様な改良を望むのは厳しい態度であり, 大変消極的とも言える推定方式に結びついてしまう。この問題に対しては, とるに足らない一様な改良よりは, 大幅な改良の可能性のある経験的ベイズ推定量を採用するという方向が大

勢のようである。しかし経験的ベイズ推定量は一樣な改良になる保証などないし、大きな改悪になる可能性もあるのであるから、(必ずしも MSE の和という基準にこだわらずとも)何らかの形で悪くないものであることを確認していくことが必要であろう。

なお、これまで等分散ではないが、基準を MSE の和とした場合を考えたが、議論は重み付きの MSE の和

$$\sum_{i=1}^p w_i E(\hat{\theta}_i - \theta_i)^2,$$

ここで $w_i, i=1, 2, \dots, p$ は正の数、を用いた場合に簡単に拡張される。この場合でもベイズ推定量は変わらないが、推定量 (5.2) は例えば

$$\left\{ 1 - \frac{p-2}{\sum X_k^2 / (w_k \sigma_k^4)} \frac{1}{\sigma_i^2 w_i} \right\} X_i, \quad i=1, 2, \dots, p$$

と修正される。 w_i の指定のされ方によっては、ますますベイズ推定量の形から遠ざかることになってしまう。

また、一樣な改良になるという要請を緩めてベイズ性との接点を探るという方向も可能である。(Chen (1988))

5.2. 信頼領域の問題

Stein タイプの縮小推定量があまり使われないとするならば、その 1 つの大きな原因は推定量に伴うべき信頼区間、信頼領域が有用なものとして明確に与えられないということであろう。ここでは、信頼領域の構成についての試みを紹介しておこう。 X_i が独立に $N(\theta_i, 1)$ に従うとき、 θ の信頼領域として X を中心とする半径 c の球

$$C^0(X) = \{\eta : \|\eta - X\| \leq c\},$$

を選び、 c が $P\{X_p^2 \leq c^2\} = 1 - \alpha$ を満たすものとすれば、これは信頼係数 $1 - \alpha$ の信頼領域である。この $C^0(X)$ を改良する信頼領域を構成することを考えよう。いま信頼領域 $C(X)$ が $C^0(X)$ の改良であるとは、

$$(I) \quad P_\theta\{\theta \in C(X)\} \geq P_\theta\{\theta \in C^0(X)\}, \quad \text{すべての } \theta \text{ に対して}$$

$$(II) \quad C(X) \text{ の体積} \leq C^0(X) \text{ の体積} \quad \text{すべての } X \text{ に対して}$$

の双方が成り立ち、かつ、(I) の厳密な不等号がある θ に対して成り立つかまたは (II) の厳密な不等号がルベグ測度が 0 でないある集合に属するすべての X に対して成り立つかどうかであることをいう。 $p \geq 3$ のとき改良が存在することは知られていたが (Stein (1962), Brown (1966), Joshi (1967) 参照)、具体的に改良となる $C(X)$ はずっと与えられていなかった。

Hwang and Casella (1982, 1983) は信頼領域 $C(X)$ として positive-part Stein 推定量

$$\left\{ 1 - \frac{a}{\|X\|^2} \right\}^+ X$$

を中心とする $C^0(X)$ と同じ半径の球を考え、 a が $0 < a \leq a^*(p)$ を満たすとき、それが改良となることを示した。ここで $a^*(p)$ は p により定まる数であり、ある種の方程式の解として数値的にその値が求められているが、 $p-2$ よりも小さい値である。しかしこの信頼領域は $C^0(X)$ と同じ体積であり、名目上の信頼係数も同じものでしかない ($\|\theta\|$ が小さいときその信頼領域が θ を含む確

率は $1-\alpha$ よりかなり大きい、 $\|\theta\| \rightarrow \infty$ とともにその確率は $1-\alpha$ に近づいていく。

信頼係数を同じに保ったまま体積の小さい信頼領域を得る試みもなされている。Casella and Hwang (1983) は中心を $a=p-2$ と選んだ positive-part Stein 推定量とし、半径が $\|X\|$ の値により変化する球として表される信頼領域を考え、数値的に改良になることを検証している。また、Shinozaki (1989) は球全体を原点に向けて、縮小する（その結果得られる領域は球でも楕円でもない）ことで改良となるものを解析的に与えている。

いずれにしても、改良となる信頼領域を構成することは困難な問題であり、分散 σ^2 の値が未知の場合の扱いを含め、まだ実用には程遠いと言わなければならない。

6. 応用例

6.1. 問題とその定式化

やや古いものではあるが、Carter and Rolph (1974) や Fay and Herriot (1979) は興味深い Stein タイプの推定量の応用例を与えている。ここでは、後者にやや似通った問題であるが、地区別の平均教育費の同時推定を取り上げてみよう。

表1の第IV列に平成元年の家計調査年報による都道府県庁所在都市別の（以下都道府県を単に県と呼ぶ）勤労者世帯の一世帯当たり年平均1か月間の教育費が与えられている。これを $Z_i, i=1, 2, \dots, 47$, としよう。（ ）の中は集計世帯数である。世帯毎に教育費は大きくばらつくことが想像され、この程度の大きさの世帯数に基づく推定では、あまり信頼の置けるものではないと推測される。試みに、家計調査年報を基に四国四県について昭和60年から平成元年までの5年間の教育費を表2に与えてみた。ここでも（ ）の中は集計世帯数である。県別にあるいは県間の差異に着目して時系列として眺めても、大まかな傾向すらつかみにくい程、ばらつきが大きいと判断できるだろう。

そこで Stein タイプの推定量を用いることを考えてみよう。いくつかのブロックに分け、ブロックの平均に近づける、家計調査年報に与えられている過去数年間の同種の数字を用いて回帰を行い、その回帰平面に近づける等いろいろな可能性が考えられるが、ここでは全国消費実態調査報告の同種の数字を用いて単回帰を行い、その回帰推定値に近づけることを試みた。（1人の査読者が指摘されたように、定数項を削除し、原点を通る回帰直線を想定の方が自然であるかもしれない。）全国消費実態調査は5年おきに行われるものであるが、表1の第II列に与えられているのは昭和59年度全国消費実態調査報告による勤労者世帯の一世帯当たりの1か月の教育費である。（ ）の中は集計世帯数である。これは県庁所在都市だけでなく県全体の集計結果であるが、集計世帯数は家計調査報告のそれよりずっと大きく、安定した数値であると考えられる。

もちろん、Stein タイプの推定量を用いる可能性を考慮する以前に、さまざまな回帰構造を想定して回帰推定量自身を用いる可能性も検討されねばならないことである。例えば、過去5年間の情勢にそれほど変化がないと考えるならば、5年間の県別の平均値に回帰し推定することも考えられるだろう。しかし、ここではそのような構造に確信も持てないので、1つの単回帰構造を用い、その回帰推定値に近づける Stein タイプの推定量を用いてみるということである。

表 1. 都道府県（庁所在都市）別一世帯当りの 1 か月の平均教育費と推定値（単位は円）

	I	II	III	IV	V
1	北海道（札幌）	6992(1537)	9472(1704)	10029(69)	11144
2	青森（青森）	9145(330)	12678(412)	9351(61)	12698
3	岩手（盛岡）	6551(363)	9282(425)	12661(61)	11242
4	宮城（仙台）	9553(499)	13541(547)	12276(68)	13591
5	秋田（秋田）	7567(334)	9242(402)	12207(60)	12045
6	山形（山形）	9396(342)	11231(423)	14899(63)	13966
7	福島（福島）	8036(480)	11241(602)	12234(65)	12432
8	茨城（水戸）	9486(646)	17818(761)	12229(60)	13563
9	栃木（宇都宮）	10267(440)	15358(499)	14080(59)	14483
10	群馬（前橋）	8991(451)	12924(555)	14448(57)	13561
11	埼玉（浦和）	13499(1662)	20746(1889)	18515(71)	17545
12	千葉（千葉）	13520(1432)	19784(1596)	19718(81)	17826
13	東京（都区部）	16669(1592)	20826(1798)	19828(235)	19804
14	神奈川（横浜）	14938(1635)	24371(1784)	25319(105)	20073
15	新潟（新潟）	8342(638)	11416(709)	14297(63)	13033
16	富山（富山）	10263(435)	13689(457)	16402(67)	14890
17	石川（金沢）	10023(430)	18122(438)	16529(62)	14707
18	福井（福井）	8474(399)	12918(424)	11445(52)	12670
19	山梨（甲府）	11253(345)	13274(410)	15333(51)	15420
20	長野（長野）	8501(547)	11858(615)	12116(62)	12782
21	岐阜（岐阜）	9832(506)	17950(533)	13104(58)	13989
22	静岡（静岡）	9497(953)	14306(999)	18829(52)	14524
23	愛知（名古屋）	11689(1553)	15666(1667)	15372(78)	15717
24	三重（津）	11934(517)	17805(618)	13572(57)	15590
25	滋賀（大津）	11696(448)	14384(493)	12953(67)	15236
26	京都（京都）	15084(507)	18411(573)	13794(45)	17849
27	大阪（大阪）	12778(2142)	17380(2293)	15017(104)	16258
28	兵庫（神戸）	12273(1404)	19884(1524)	18085(59)	16592
29	奈良（奈良）	15407(438)	23590(497)	15732(65)	18218
30	和歌山（和歌山）	10858(346)	16898(362)	15279(59)	15122
31	鳥取（鳥取）	7715(377)	12052(426)	11905(58)	12117
32	島根（松江）	6973(405)	8836(410)	9390(60)	11034
33	岡山（岡山）	9027(510)	11885(524)	13862(60)	13501
34	広島（広島）	10218(806)	17141(818)	18431(63)	15151
35	山口（山口）	7180(527)	12422(573)	9250(67)	11134
36	徳島（徳島）	10232(332)	14429(398)	13487(59)	14358
37	香川（高松）	7854(360)	14002(439)	11055(60)	12076
38	愛媛（松山）	8673(395)	12172(406)	16837(63)	13692
39	高知（高知）	10360(334)	11026(339)	10276(42)	14009
40	福岡（福岡）	10287(998)	13501(1013)	16033(72)	14860
41	佐賀（佐賀）	9704(358)	13750(405)	15038(56)	14210
42	長崎（長崎）	8335(363)	9451(404)	13733(54)	12914
43	熊本（熊本）	8953(390)	13396(418)	13005(49)	13307
44	大分（大分）	6493(392)	11322(400)	12879(67)	11249
45	宮崎（宮崎）	6563(332)	8952(373)	12102(61)	11165
46	鹿児島（鹿児島）	7873(392)	10330(469)	14524(56)	12648
47	沖縄（那覇）	9920(395)	14459(446)	14028(113)	14190

I 都道府県とその県庁所在地

II 昭和 59 年度全国消費実態調査による都道府県別勤労者世帯の一世帯当り 1 か月の教育費

III 平成元年度全国消費実態調査による都道府県別勤労者世帯の一世帯当り 1 か月の教育費

IV 平成元年度家計調査年報による都道府県庁所在都市別勤労者世帯の一世帯当り年平均 1 か月の教育費

V Stein タイプの推定量による IV の平均値の推定値

II, III, IV の () 内はすべて集計世帯数である。

表 2. 家計調査年報による昭和 60 年から平成元年までの四国 4 県の県庁所在都市の平均教育費 () 内は集計世帯数, 単位は円

県 \ 年度	昭和 60 年	昭和 61 年	昭和 62 年	昭和 63 年	平成元年
徳島 (徳島)	11441(56)	8388(55)	7898(60)	12177(57)	13487(59)
香川 (高松)	12298(62)	10889(66)	8803(62)	11160(55)	11055(60)
愛媛 (松山)	7595(60)	13009(64)	14291(66)	16729(63)	16837(63)
高知 (高知)	9390(52)	10827(53)	12108(53)	11278(47)	10276(42)

6.2. 推定量

推定量を構成する際の 1 つの問題は変数変換の可能性である. 1 つの尤もらしい考え方は, 分散を安定化するため, 対数をとる

$$X_i = \log Z_i \sim N(\theta_i, \sigma^2/n_i), \quad i=1, 2, \dots, 47 \quad (6.1)$$

と想定することであろう, ここで n_i は各県の集計世帯数を表す. つまり, 元の Z_i のままでは, その標準偏差は平均に比例して大きくなると想定することになる.

上の仮定を認めると, 次には σ^2 の値をどのように推定するかが問題になる. 元の世帯毎のデータがあれば別であるが, 表 1 の第 IV 列のデータだけでは σ^2 の情報を得ることはできない. そこで, いろいろと問題点のある荒っぽい方法ではあるが, 次のように考えた. 表 2 のように, 各県毎に年度別の Z_i の値が得られている. 昭和 60 年から平成元年までの 5 年間のデータを, 物価変動を考慮し平成元年の物価水準での値に修正したうえで, 大きさ 5 の標本とみなした. それらの対数をとった上で各県 (i) 毎に不偏分散を計算し, n_i 倍すれば σ^2 の推定量 $\hat{\sigma}_{(i)}^2$, つまり, $\sigma^2 \chi_i^2/4$ として分布すると考える量が得られることになる. (細かいことではあるが, n_i の値は年度により多少変動するので, 5 年間の平均値を用いた.) それら 47 個の $\hat{\sigma}_{(i)}^2$ の算術平均値 $\bar{\sigma}^2 = 1.3073$ を以て σ^2 を置きかえることにした.

図 2 は 47 個の $\hat{\sigma}_{(i)}^2$ の値から描かれるヒストグラムを $1.3073 \times \chi_i^2/4$ の密度関数と重ねて描いたものである. 比較的良好なあてはまりを示しており, これで見える限り (6.1) の想定も大きな無理のあるものとする必要はないであろう.

さて, 推定量としては, 表 1 の第 II 列に与えられている量の対数をとる, これを説明変数として単回帰を行い, 回帰直線へ縮小する Stein タイプの推定量を用いる. ここでの問題は X_i の分散が

$$\sigma_i^2 = \frac{\sigma^2}{n_i}, \quad i=1, 2, \dots, 47$$

と異なることである. 5.1 で述べたように, 分散が等しくない場合一様な改良を与える推定量と経験的ベイズ推定量とは異なるが, ここでは Fay and Herriot (1979) に沿って経験的ベイズ推定量を用いてみる.

独立に

$$N(\alpha + \beta Y_i, \tau^2)$$

に従うという事前分布を想定すれば, X_i の周辺分布は, 独立に

図2. $\hat{\alpha}_{ii}^2$ のヒストグラムとあてはめた密度関数

$$N(\alpha + \beta Y_i, \sigma_i^2 + \tau^2)$$

に従うというものになる。 X を第 i 成分が X_i の 47 次の列ベクトル, Y を第 i 行が $y_i' = (1, Y_i)$ である 47×2 の行列とし, V を対角成分として $\sigma_i^2 + \tau^2$ をもつ 47 次の対角行列とする. $\alpha + \beta Y_i$ の推定量は, τ^2 の値がわかっているとすれば,

$$X_i^* = y_i' (Y' V^{-1} Y)^{-1} Y' V^{-1} X, \quad i = 1, 2, \dots, 47 \quad (6.2)$$

で求められる. 一方

$$E \left\{ \sum_{i=1}^{47} \frac{(X_i - X_i^*)^2}{\sigma_i^2 + \tau^2} \right\} = 45 \quad (6.3)$$

である.(重みつき最小 2 乗法を用いたときの重みつき残差平方和に関する基本的結果である.)そこで (6.3) の左辺の期待値の記号を取り去り

$$\sum_{i=1}^{47} \frac{(X_i - X_i^*)^2}{\sigma_i^2 + \tau^2} = 45 \quad (6.4)$$

を満たすことを要請することにする.

従って, (6.2) と (6.4) を連立させて, τ^2 の推定値と X^* の値とを求めるのである.(計算は反復を必要とすることになるが, 詳細は Fay and Herriot (1979) を参照されたい.) τ^2 の推定量 $\hat{\tau}^2$ と X_i^* の値が得られれば, (3.1) に相当する Stein タイプの推定量として

$$\hat{\theta}_i = X_i^* + \left\{ 1 - \frac{\sigma_i^2}{\sigma_i^2 + \hat{\tau}^2} \right\} (X_i - X_i^*), \quad i = 1, 2, \dots, 47 \quad (6.5)$$

が得られる. $\hat{\tau}^2$ の値は 0.00449 となり, $\sigma_i^2 = \sigma^2/n_i$ の値 ($n_i = 60$ で約 0.02179) に比べ小さい値になったので, X_i を X_i^* に向けて大きく縮小する結果になっている. もちろん θ_i を逆変換した $e^{\hat{\theta}_i}$ が各県の一世代当りの平均教育費の推定量である. このようにして得られた推定値が表 1 の第 V 列に与えられている.

6.3. 評価

得られた推定量がどの程度の改良を与えるものであるかを多少なりとも評価するため, 表 1 の第 III 列に与えられている平成元年度全国消費実態調査による県別の一世帯当りの一か月間の教育費を用いることにしよう. () 内に与えられている集計世帯数は IV 列のそれに比べおおむね少なくとも 7 倍程度の大きさであるので, その意味ではこれらの方がずっと信頼の置けるもので

ある。しかし、これは全県についてのデータから求められた値であり、調査方法も異なるので、III がIVのデータのとられた集団の真の平均値を与えるなどとはもちろん言えないが、大まかな目安を与えるものであると考えるのに無理はないだろう。なお、III列のデータとIV列のデータの単純平均はそれぞれ14366と14287でほとんど差はない。つまり、大都市部の方が全県の場合よりも教育費支出が大きくなるという傾向は、この2組のデータを比較する限り全く無いということである。

あくまでも1つの大まかな目安を与えるものに過ぎないということを強調した上で、III列とIV列（通常の推定量）およびIII列とV列（Steinタイプの推定量）について差の2乗和の値を示せば、前者が 3.774×10^8 、後者が 2.157×10^8 であった。つまりSteinタイプの推定量の方が4割以上も小さい値になっている。逆変換をせずに対数尺度のままで差の2乗和の計算を行っても1.913と1.031、さらに、対数尺度のままで分散 σ_i^2 の逆数で重みづけをした差の2乗和を計算しても88.49と48.88とほぼ同様の傾向を示している。またIII列とIV列、および、III列とV列について順位相関係数を計算してみると、それぞれ、0.611と0.847となり、やはりSteinタイプの推定量がIIIのパターンに近いものを与えていることがわかる。いずれにしても、かなりの程度の改良を示唆するものと受け取ってよいだろう。

また、想定した回帰構造に十分な自信があるのなら、 σ_i^2 の逆数を重みとした重みつき最小2乗法による回帰推定量を用いるであろうが、その場合の推定値とIII列との差の2乗和は 2.255×10^8 であった。（対数尺度のままの場合には1.043であった。）Steinタイプの推定量の方が多少小さいが、ほとんど差はない。 $\hat{\theta}_i$ が X_i を X_i^* に向けて大きく縮小する結果になっているので納得の行くことである。

7. 結 び

点推定量としてのStein推定量についての議論はかなり成熟を遂げ、文献にもSteinタイプの縮小推定量を用いた応用例が数多く現われるようになっている（Morris (1983), Breslow (1990) 参照）。多少気になるのは5.1でも述べた通り（等分散でない場合を含め）複雑な状況では、MSEの和という基準の下で一様な改良となる推定量などつukれない、あるいは、一様な改良を与える推定量と経験的ベイズ推定量が異なるものであるということが起こるが、実際に用いられているのは経験的ベイズの考え方による推定方法が多いようである。もしもその正当化が、Stein推定量は一様な改良を与えるものであり（これは大変大きなセールスポイントである）、その考え方に沿った推定法であるからというものであるとすれば、これはやや羊頭狗肉の感がしないでもない。2つの行き方の接点を探るような研究が未だ必要なようにも思われる。

Stein推定量を用いるのは、ただ単にデータを要約するという場面ではないだろう。さらに、やや誇張して言えば、むしろ予測や決定としての側面を持つ場面であることも多いように思える。そのような場面であれば、Steinタイプの縮小推定量を用いて成功するか（したか）どうかはある程度ははっきりすることもあるであろうし、そのような経験が有用な方法だけを残していくことになるだろう。日本でも官庁や民間会社でSteinタイプの推定量を用いた、あるいは用いることを検討しているという話を風の便りには聞くが、多くの場面で役立てられることを願っている。

謝 辞 本稿は、応用統計学会1991年度年会での講演の予稿に加筆・訂正したものです。年会で

は、討論者の後藤昌司先生(塩野義製薬)をはじめ何人かの方々から有益なご意見を頂きました。編集理事および査読者の方々からも適切なご指摘・ご助言を頂戴しました。ここに感謝の意を表させていただきます。

参 考 文 献

- Akaike, H. (1977): An extension of the method of maximum likelihood and the Stein's problem. *Ann. Inst. Statist. Math.* **29**, 153-164.
- Akaike, H. (1980): Ignorance prior distribution of a hyperparameter and Stein's estimator. *Ann. Inst. Statist. Math.* **32**, 171-178.
- Akai, T. (1986): Simultaneous estimation of location parameters of the distribution with finite support. *Ann. Inst. Statist. Math.* **38**, 85-99.
- Akai, T. (1989): Simultaneous estimation of means of classified normal observations. *Ann. Inst. Statist. Math.* **41**, 485-502.
- Berger, J. (1976): Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* **4**, 223-226.
- Berger, J. (1982): Selecting a minimax estimator of a multivariate normal mean. *Ann. Statist.* **10**, 81-92.
- Berger, J. (1985): *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.
- Berger, J. and Dey, D.K. (1983): Combining coordinates in simultaneous estimation of normal means. *J. Statist. Plan. Inf.* **8**, 143-160.
- Berger, J. and Robert, C. (1990): Subjective hierarchical Bayes estimation of a multivariate normal mean: On the frequentist interface. *Ann. Statist.* **18**, 617-651.
- Bhattacharya, P.K. (1966): Estimating the mean of a multivariate normal population with general quadratic loss function. *Ann. Math. Statist.* **37**, 1819-1824.
- Bilodeau, M. and Kariya, T. (1989): Minimax estimators in the normal MANOVA model. *J. Multivariate Anal.* **28**, 260-270.
- Brandwein, A.C. and Strawderman, W.E. (1990): Stein estimation: The spherically symmetric case. *Statistical Science* **5**, 356-369.
- Breslow, L. (1990): Biostatistics and Bayes. *Statistical Science* **5**, 269-298.
- Brown, L.D. (1966): On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37**, 1087-1135.
- Carter, G. and Rolph, J. (1974): Empirical Bayes methods applied to estimating fire alarm probabilities. *J. Amer. Statist. Assoc.* **69**, 880-885.
- Casella, G. and Hwang, J.T. (1983): Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Amer. Statist. Assoc.* **78**, 688-698.
- Chen, S.Y. (1988): Restricted risk Bayes estimation for the mean of the multivariate normal distribution. *J. Multivariate Anal.* **24**, 207-217.
- Dey, D.K. and Berger, J. (1983): On truncation of shrinkage estimators in simultaneous estimation of normal means. *J. Amer. Statist. Assoc.* **78**, 865-869.
- Efron, B. and Morris, C. (1971): Limiting the risk of Bayes and empirical Bayes estimators — Part I: The Bayes case. *J. Amer. Statist. Assoc.* **66**, 807-815.
- Efron, B. and Morris, C. (1972): Limiting the risk of Bayes and empirical Bayes estimators — Part II: The empirical Bayes case. *J. Amer. Statist. Assoc.* **67**, 130-139.
- Efron, B. and Morris, C. (1973a): Stein's estimation rule and its competitors — An empirical Bayes approach. *J. Amer. Statist. Assoc.* **68**, 117-130.
- Efron, B. and Morris, C. (1973b): Combining possibly related estimation problems. *J. Roy. Statist. Soc. Ser. B* **35**, 379-421.
- Efron, B. and Morris, C. (1975): Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.* **70**, 311-331.
- Fay, R.E. III and Herriot, R.A. (1979): Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.* **74**, 269-277.
- George, E.I. (1986a): Minimax multiple shrinkage estimation. *Ann. Statist.* **14**, 188-205.
- George, E.I. (1986b): Combining minimax shrinkage estimators. *J. Amer. Statist. Assoc.* **81**, 437-445.

- Honda, T. (1991): Minimax estimators in the Manova model for arbitrary quadratic loss and unknown covariance matrix. *J. Multivariate Anal.* **36**, 113-120.
- Hwang, J.T. and Casella, G. (1982): Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist.* **10**, 868-881.
- Hwang, J.T. and Casella, G. (1983): Improved set estimators for a multivariate normal mean. *Statist. Decisions* (supplement issue) **1**, 3-16.
- Hui, S.L. and Berger, J. (1983): Empirical Bayes estimation of rates in longitudinal studies. *J. Amer. Statist. Assoc.* **78**, 753-760.
- James, W. and Stein, C. (1961): Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1**, 361-380. Univ. California Press.
- Joshi, V.M. (1967): Inadmissibility of the usual confidence sets for the mean of a multivariate normal population. *Ann. Math. Statist.* **38**, 1868-1875.
- 柏木宣久, 柳本武美 (1990): 平滑化法とそのソフトウェア. 応用統計学 **19**, 79-92.
- Konno, Y. (1990): Families of minimax estimators of matrix of normal means with unknown covariance matrix. *J. Japan Statist. Soc.* **20**, 191-201.
- Konno, Y. (1991): On estimation of a matrix of normal means with unknown covariance matrix. *J. Multivariate Anal.* **36**, 44-55.
- Kubokawa, T. (1991): An approach to improving the James-Stein estimator. *J. Multivariate Anal.* **36**, 121-126.
- Li, K. and Hwang, J.T. (1984). The data-smoothing aspect of Stein estimates. *Ann. Statist.* **12**, 887-897.
- Lindley, D.V. (1962): Discussion of "Confidence sets for the mean of a multivariate normal distribution", by C. Stein. *J. Roy. Statist. Soc. Ser. B* **24**, 285-287.
- Morris, C. (1983): Parametric empirical Bayes inference: Theory and applications. *J. Amer. Statist. Assoc.* **78**, 47-65.
- Rao, C.R. and Shinozaki, N. (1979): Precision of individual estimators in simultaneous estimation of parameters. *Biometrika* **65**, 23-30.
- Sen, P.K., Kubokawa, T. and Saleh, A.K. (1989): The Stein paradox in the sense of Pitman measure of closeness. *Ann. Statist.* **17**, 1375-1386.
- Shinozaki, N. (1974): A note on estimating the mean vector of a multivariate normal distribution with general quadratic loss function. *Keio Engng. Rep.* **27**, 105-112.
- Shinozaki, N. (1984): Simultaneous estimation of location parameters under quadratic loss. *Ann. Statist.* **12**, 322-335.
- Shinozaki, N. (1986): Some modifications of Efron-Morris estimator and their minimaxity. *J. Japan Statist. Soc.* **16**, 191-201.
- Shinozaki, N. (1989): Improved confidence sets for the mean of a multivariate normal distribution. *Ann. Inst. Statist. Math.* **41**, 331-346.
- Shiraishi, T. (1991): On positive-part shrinkage R- and M-estimation in one-way ANOVA. *J. Japan Statist. Soc.* **21**, 61-72.
- Stein, C. (1956): Inadmissibility of the usual estimator of the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1**, 197-206. Univ. California Press.
- Stein, C. (1962): Confidence sets for the mean of a multivariate normal distribution (with discussion) *J. Roy. Statist. Soc. Ser. B* **24**, 265-296.
- Stein, C. (1981): Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9**, 1135-1151.
- Stigler, S.M. (1990): A Galtonian perspective on shrinkage estimators. *Statistical Science* **5**, 147-155.
- Takada, Y. (1979): Stein's positive part estimator and Bayes estimator. *Ann. Inst. Statist. Math.* **31**, 177-183.
- 竹内 啓 (1979): Stein 推定量の意味とその応用. 応用統計学 **8**, 81-95.
- Yuan-Tsung Chang (1982): Stein-type estimators for parameters in truncated spaces. *Keio Sci. Tech. Reports* **35**, 185-193.

(1991年8月30日受付 10月15日最終修正)