CIRJE-J-242

制約された母数空間における推定

東京大学大学院経済学研究科
久保川達也

2012 年 6 月

制約された母数空間における推定

久保川 達也[*] [†]

平成 24 年 6 月 8 日

概 要

制約された母数空間における推定問題は, 許容性やミニマックス性などの統計的決定理論の視点から様々な興味深い問題を含んでいる. 例えば, 分散が既知の正規分布の平均の推定という簡単な問題を考えてみると, 平均が正に制約されているような片側制約の場合には標本平均はミニマックスになるが, 両側から制約されている場合にはもはやミニマックスでも許容的でもない. しかし, 分散が未知の場合には, 平均が両側から制約されていても標本平均はミニマックスになっている. このように, 母数空間への制約の入り方や推定問題の設定の仕方によって予想を覆すような現象が現れる. この論文では, 統計的決定理論の視点から母数空間が制約されているときの推定問題について解説する.

キーワード: 母数制約, ミニマックス性, ベイズ推定, 一般化ベイズ推定, 許容性, 統計的決定理論, Stein 恒等式, Gauss 発散定理, 不偏性, 共変性, 位置尺度分布族

1　はじめに

母数空間が制約されているときの統計的推測の問題は広い分野で研究され, その内容は Barlow, Bartholomew, Bremner, Brunk (1972) や Robertson, Wright, Dykstra (1988) で詳しく解説されている. その大半は, 仮説検定手法や最尤推定量を求めるためのアルゴリズムなどの研究に注がれてきたようである. 一方, 推定の理論的な研究は van Eeden (2006) の中で解説されているが, 統計的決定理論の立場に立った研究は簡単な問題設定のもとでは議論されてきたが, 現実的な制約条件が入った問題は制約の複雑さから扱いが困難になり, あまり発展してこなかったように思う. しかし, 例えば平均の順序制約の下で順序制約を利用した単調回帰 (isotonic regression) 推定量を考えてみると, 順序制約を利用しない推定量に比べ常によくなっていると予想されるが, 統計的決定理論の立場からすると, 実はその予想は正しくないことが Lee (1988), Hwang, Peddada (1994) によって示されている. このように, 制約された母数の推定には, 常識を覆すような興味深い現象が現れ, 解決するのに値する問題が残っているようである. 最近 Hartigan (2004) が導いた結果は極めて魅力的で画期的なもので, 順序制約を利用した一般化ベイズ推定量が順序制約を利用しない推定量を一様に改良することを, Stein の恒等式と Gauss の発散定理を用いて証明するのに成功した. また Marchand, Perron, Strawderman らによって新たな理論展開がなされ, 母数制約下での推定理論の世界が少しずつ広がってきている. この論文では, 制約された母数の点推定に焦点を絞り, 統計的決定理論の視点からこれまでの研究成果の解説を行い, 今後この分野に関心を持たれる方の参考になるようにまとめさせて頂いた.

制約された母数空間における点推定の, 統計的決定理論からの研究は, Katz (1961) が 1 母数の指数型分布族において母数制約されているときの一般化ベイズ推定量が許容的でミニマックスであ

[*]113-0033 東京都文京区本郷 7-3-1 東京大学大学院経済学研究科
[†]E-Mail: tatsuya@e.u-tokyo.ac.jp

ることを示して以来, 新たな理論展開を繰り返しながら発展してきた. その中から, この論文では, 次の４つのテーマについて研究成果をまとめ, その内容を解説する. (A) 一般化ベイズ推定量による改良, (B) 両側から制約されている場合の推定, (C) 正に制約された位置母数の和の推定, (D) 母数制約のもとでのミニマックス性

(A) 一般化ベイズ推定量による改良. 分散が既知の p-変量正規分布において, 平均が凸集合に制約されているとき, 平均の不偏推定量もしくは最良共変推定量は, 凸集合上の一様分布に対する一般化ベイズ推定量で改良できることを示したのが Hartigan (2004) の結果である. この証明には, Stein の恒等式と Gauss の発散定理が用いられており, 証明方法の美しさはこの分野の研究者に新たな創造の活力を与えてきた. Tsukuma, Kubokawa (2008) は, 平均が凸多面錐に制約されている場合に, 不偏推定量がミニマックスになり, その結果一般化ベイズ推定量がミニマックスになることを示した. また $p=1,2$ のとき一般化ベイズ推定量が許容的になるものの, $p\geq 3$ のとき非許容的になる, いわゆる Stein 現象を示した. Hartigan の方法は, 凸集合というかなり一般の制約を扱うことを可能にしている点で優れているが, (1) 分散が既知の正規分布に限定されること, (2) ２乗損失関数のときのみ証明可能, (3) 事前分布は一様分布のときのみ証明可能という限界がある. その限界を超える１つの試みが Kubokawa (1994a, b, 98, 99) により提案された IERD (Integral Expression of Risk Difference) 法である. Kubokawa (1994a) で与えた表現式を竹内啓氏が自然な導出方法へ修正したものである. Hartigan の方法のように一般の凸集合への制約を扱うことはできないが, IERD 法を用いることにより, 位置分布族や凸損失関数への拡張が可能になる. IERD 法は, James-Stein 推定量を改良する一般化ベイズ推定量の導出や分散に対するミニマックスな一般化ベイズ推定量の導出及び信頼区間への拡張など, 他の推定問題においても役立つ.

(B) 両側から制約されている場合の推定. 制約された母数の推定において興味深い結果の１つは, Casella, Strawderman (1981) により得られた結果で, 正規分布の平均が両側から制約されて閉区間に入る場合には, 不偏推定量はもはやミニマックスでなく, 区間の幅が小さいときには, 両端に等確率をもつ事前分布に対するベイズ推定量がミニマックスになるというものである. ベイズ推定量のリスクの上限を評価する素朴な証明方法なので, 分布の拡張などは簡単ではない. Marchand, Perron (2001) は, Casella, Strawderman の結果を多次元正規分布に拡張するとともに, ベイズ推定量が最尤推定量を改良することを示した.

(C) 正に制約された位置母数の和の推定. p 個の正規分布について, それぞれの平均が正に制限されているとき, p 個の平均の和を推定する問題は, 最近 Kubokawa (2012), Kubokawa, Strawderman (2011a, b) により, 一連の研究成果が得られた. そこでは, $p\geq 2$ のとき一般化ベイズ推定量がミニマックスにならないこと, $p\geq 5$ のとき最尤推定量がミニマックスにならないという非ミニマックス性が示されたが, その結果は一般の位置分布族において位置母数の線形結合の推定へ拡張され, ミニマックス性の必要十分条件が得られている. また許容的ミニマックス推定量の導出などもなされている.

この報告では, 制約された母数空間における推定問題について, (D) 母数制約のもとでのミニマックス性を含め, これまでの研究成果を解説する. 証明については, わかりやすい部分や興味深いものだけを取り上げ, 詳細は割愛させて頂く. 2 章は, 制約された母数空間上の一様分布に対する一般化ベイズ推定量が不偏推定量（最良共変推定量）を改良するための方法として, Hartigan の方法と IERD 法を説明する. 3 章では, 母数制約がないときの最良共変推定量が母数制約のもとでもミニマックスになることを, 簡単な場合に Girshick-Savage の方法を利用して証明し, 一般の群構造の入った推定問題でのミニマックス性の条件を与える. 凸多面錐に制約されている場合のミニマックス性について議論し, 一般化ベイズ推定量の許容性と非許容性について述べる. 4 章は, 両

側から制約されている場合の推定問題を扱う．5 章では, 正に制約された位置母数の線形結合の推定問題を考え, 一般化ベイズ推定量の非ミニマックス性, 最尤推定量の非ミニマックス性, 許容的ミニマックス推定量の導出, 分散の Stein 推定量との関係について説明する．

2　母数制約のもとでの一般化ベイズ推定量

確率変数 $\boldsymbol{X} = (X_1, \ldots, X_p)^t$ の確率密度関数が $f(\boldsymbol{x} - \boldsymbol{\mu})$ で与えられ, 位置ベクトル $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^t$ の母数空間 Ω が D に制約されている場合を考えよう．制約されていないときの母数空間 Ω は, p 次元ユークリッド空間である．母数 $\boldsymbol{\mu}$ が制約されていないときには, $\boldsymbol{\mu}$ の事前分布として Ω 上の一様分布を仮定すると, 得られる一般化ベイズ推定量は, 位置変換に関する最良共変推定量になり, これは定数リスクをもつ推定量である．母数空間が制約されているときには, $\boldsymbol{\mu}$ の事前分布として, 制約された空間 D 上に一様分布を想定してみると, 一般化ベイズ推定量が最良共変推定量を改良することが予想される．このことを示す方法として, Hartigan の方法と IERD 法の 2 つの方法が知られている．

位置ベクトルの推定においては, 推定量 $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_p)^t$ の良さを測るために, 2 乗損失関数 $\|\widehat{\boldsymbol{\mu}}(\boldsymbol{X}) - \boldsymbol{\mu}\|^2 = (\widehat{\boldsymbol{\mu}}(\boldsymbol{X}) - \boldsymbol{\mu})^t(\widehat{\boldsymbol{\mu}}(\boldsymbol{X}) - \boldsymbol{\mu}) = \sum_{i=1}^p (\widehat{\mu}_i(\boldsymbol{X}) - \mu_i)^2$ に関するリスク関数 $R(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) = E[\|\widehat{\boldsymbol{\mu}}(\boldsymbol{X}) - \boldsymbol{\mu}\|^2]$ を考える．この論文では, 特に断らない限り 2 乗損失に関するリスクを扱うこととする．母数 $\boldsymbol{\mu}$ の事前分布として, 制約された母数空間 D 上に一様分布を想定すると, 得られる一般化ベイズ推定量は,

$$\widehat{\boldsymbol{\mu}}^{GB}(\boldsymbol{X}) = \int_D \boldsymbol{\xi} f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} \Big/ \int_D f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}$$

で与えられる．これは, $\widehat{\boldsymbol{\mu}}^{GB} = \boldsymbol{X} - \boldsymbol{g}(\boldsymbol{X})$ とも書かれる．ここで,

$$\boldsymbol{g}(\boldsymbol{X}) = \frac{\int_D (\boldsymbol{X} - \boldsymbol{\xi}) f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}}{\int_D f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}}$$

である．$D = \Omega$ のときには, $\boldsymbol{g}(\boldsymbol{X})$ は定数になり，そのときの推定量を $\widehat{\mu}^U(\boldsymbol{X})$ と書くと, これは最良共変推定量になっている．

2.1　Hartigan の方法

Hartigan (2004) は, $\boldsymbol{X}$ の分布が正規分布 $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{I}_p)$ のときに, Stein の恒等式と Gauss の発散定理を用いて, $\widehat{\boldsymbol{\mu}}^{GB}$ が最良共変推定量を改良することを示した．このときの最良共変推定量は $\boldsymbol{X}$ となるので, $\widehat{\boldsymbol{\mu}}^{GB}$ のリスクが $\boldsymbol{X}$ のリスクより小さくなることを示せばよい．

補題 2.1 $\boldsymbol{X}$ の分布が正規分布 $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{I}_p)$ に従うとする．このとき, D 上の一様分布に対する一般化ベイズ推定量 $\widehat{\boldsymbol{\mu}}^{GB}(\boldsymbol{X})$ のリスク関数は次のように表される．

$$R(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}^{GB}) = p + E\Big[\frac{\int_D \boldsymbol{\nabla}_{\boldsymbol{\xi}}^t \{(\boldsymbol{\mu} - \boldsymbol{\xi}) f(\boldsymbol{X} - \boldsymbol{\xi})\} \mathrm{d}\boldsymbol{\xi}}{\int_D f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}}\Big] \tag{2.1}$$

ここで，$\boldsymbol{\nabla}_{\boldsymbol{\xi}} = (\partial/\partial\xi_1, \ldots, \partial/\partial\xi_p)^t$ である．

証明. まず, $\widehat{\boldsymbol{\mu}}^{GB}$ のリスク関数は,

$$R(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}^{GB}) = E[\|\widehat{\boldsymbol{\mu}}^{GB} - \boldsymbol{\mu}\|^2] = p - 2E[(\boldsymbol{X} - \boldsymbol{\mu})^t \boldsymbol{g}(\boldsymbol{X})] + E[\|\boldsymbol{g}(\boldsymbol{X})\|^2]$$

と表される. ここで, 積の期待値 $E[(\boldsymbol{X} - \boldsymbol{\mu})^t \boldsymbol{g}(\boldsymbol{X})]$ を評価するために, Stein の恒等式 (Stein (1981)) を用いる. これは, 部分積分を用いて期待値の中身を母数 $\boldsymbol{\mu}$ に依存しないように書き換えることのできる恒等式で, 具体的には,

$$E[(\boldsymbol{X} - \boldsymbol{\mu})^t \boldsymbol{g}(\boldsymbol{X})] = E[\boldsymbol{\nabla}_{\boldsymbol{X}}^t \boldsymbol{g}(\boldsymbol{X})]$$

で与えられる. ただし, $\boldsymbol{\nabla}_{\boldsymbol{X}}$ は $\boldsymbol{X}$ に関する微分作用素で, $\boldsymbol{\nabla}_{\boldsymbol{X}} = (\partial/\partial X_1, \ldots, \partial/\partial X_p)^t$ である. 具体的に, $\boldsymbol{\nabla}_{\boldsymbol{X}}^t \boldsymbol{g}(\boldsymbol{X})$ を計算すると,

$$\begin{aligned} -E[(\boldsymbol{X} - \boldsymbol{\mu})^t \boldsymbol{g}(\boldsymbol{X})] =& -E[\boldsymbol{\nabla}_{\boldsymbol{X}}^t \boldsymbol{g}(\boldsymbol{X})] \\ =& E\Big[-p + \frac{\int_D \|\boldsymbol{X} - \boldsymbol{\xi}\|^2 f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}}{\int_D f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}} - \|\boldsymbol{g}(\boldsymbol{X})\|^2\Big] \end{aligned}$$

となる. 一方, $(\boldsymbol{X} - \boldsymbol{\mu})^t (\boldsymbol{X} - \boldsymbol{\xi}) = \|\boldsymbol{X} - \boldsymbol{\xi}\|^2 - (\boldsymbol{\mu} - \boldsymbol{\xi})^t (\boldsymbol{X} - \boldsymbol{\xi})$ より,

$$\begin{aligned} &-(\boldsymbol{X} - \boldsymbol{\mu})^t \boldsymbol{g}(\boldsymbol{X}) \\ &\quad = -\frac{\int_D \|\boldsymbol{X} - \boldsymbol{\xi}\|^2 f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}}{\int_D f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}} + \frac{\int_D (\boldsymbol{\mu} - \boldsymbol{\xi})^t (\boldsymbol{X} - \boldsymbol{\xi}) f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}}{\int_D f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}} \end{aligned}$$

と表される. 従って,

$$R(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}^{GB}) = E\Big[\frac{\int_D (\boldsymbol{\mu} - \boldsymbol{\xi})^t (\boldsymbol{X} - \boldsymbol{\xi}) f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}}{\int_D f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}}\Big]$$

と書ける. ここで,

$$\int_D \boldsymbol{\nabla}_{\boldsymbol{\xi}}^t \{(\boldsymbol{\mu} - \boldsymbol{\xi}) f(\boldsymbol{X} - \boldsymbol{\xi})\} \mathrm{d}\boldsymbol{\xi} = -p \int_D f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} + \int_D (\boldsymbol{\mu} - \boldsymbol{\xi})^t (\boldsymbol{X} - \boldsymbol{\xi}) f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}$$

より, 補題 2.1 の結果を得る. ■

補題 2.1 の中で与えられたリスクの表現式について, その右辺の期待値の中の分子を評価するために Gauss の発散定理を用いる. 制約空間 D の境界を ∂D, 境界上の測度を $\mathrm{d}S(\cdot)$ で表し, ∂D 上の点 $\boldsymbol{\xi}$ における単位法線ベクトルを $\boldsymbol{n}_{\boldsymbol{\xi}}$ とすると, Gauss の発散定理により,

$$\int_D \boldsymbol{\nabla}_{\boldsymbol{\xi}}^t \{(\boldsymbol{\mu} - \boldsymbol{\xi}) f(\boldsymbol{X} - \boldsymbol{\xi})\} \mathrm{d}\boldsymbol{\xi} = \int_{\partial D} \boldsymbol{n}_{\boldsymbol{\xi}}^t (\boldsymbol{\mu} - \boldsymbol{\xi}) f(\boldsymbol{X} - \boldsymbol{\xi}) \mathrm{d}S(\boldsymbol{\xi})$$

が成り立つ. D が凸集合のときには, $\boldsymbol{\xi} \in \partial D$, $\boldsymbol{\mu} \in D$ に対して $\boldsymbol{n}_{\boldsymbol{\xi}}^t (\boldsymbol{\mu} - \boldsymbol{\xi}) \leq 0$ となるので, $R(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}^{GB}) \leq p$ が成り立つ. 従って, 次の定理を得る.

定理 2.1 *$\boldsymbol{X}$ の分布が正規分布 $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{I}_p)$ に従うとする. D が凸集合ならば, D 上の一様分布に対する一般化ベイズ推定量 $\widehat{\boldsymbol{\mu}}^{GB}$ は $\boldsymbol{X}$ を改良する.*

Hartigan の結果は, D が凸集合という, かなり一般的な制約空間に対して改良結果を導くことができる. しかし, 正規分布の仮定が本質的であり, 正規性の仮定を緩めると, Hartigan の方法は利用できない. 次の節で紹介する IERD 法は正規性の仮定を外すことのできる１つの方法である.

2.2 IERD 法

確率変数 X の密度関数が位置母数 μ をもつ分布 $f(x-\mu)$ に従っていて, 位置母数 μ が 1 次元で, $D_0 = \{\mu | \mu > 0\}$ に制約されている場合を考えよう. この場合, μ の最良共変推定量は, $c_0 = E[X-\mu] = \int uf(u)\mathrm{d}u$ に対して $\widehat{\mu}^U = X - c_0$ で与えられる. 母数が制約されているときに $\widehat{\mu}^U$ を改良する推定量のクラスとして

$$\widehat{\mu}_\phi(X) = X - \phi(X)$$

を考える. ここで, $\phi(\cdot)$ は絶対連続な関数である. $\widehat{\mu}_\phi(X)$ が $\widehat{\mu}^U$ を改良するための $\phi(\cdot)$ に関する条件を求めるために, IERD 法に基づいた次の補題が役立つ.

補題 2.2 $\phi(\cdot)$ が $\lim_{w\to\infty}\phi(w) = c_0$ をみたすならば, 2 つの推定量 $\widehat{\mu}_\phi$ と $\widehat{\mu}^U$ のリスク関数の差は, 積分を通して次のように書かれる.

$$\begin{aligned}\Delta \equiv& R(\mu, \widehat{\mu}^U) - R(\mu, \widehat{\mu}_\phi)\\ =& -2\int \Big\{\int_{-\infty}^{w} uf(u)\mathrm{d}u - \phi(w+\mu)\int_{-\infty}^{w} f(u)\mathrm{d}u\Big\}\phi'(w+\mu)\mathrm{d}w\end{aligned}$$

証明. 仮定より $\lim_{w\to\infty}\phi(w) = c_0$ だから, 推定量 $\widehat{\mu}_\phi$ と $\widehat{\mu}^U$ のリスク関数の差 Δ は,

$$\Delta = E\Big[\big[\big\{X - \phi(X+t) - \mu\big\}^2\big]_{t=0}^{\infty}\Big] = E\Big[\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}\big\{X - \phi(X+t) - \mu\big\}^2 \mathrm{d}t\Big]$$

と書けることがわかるので,

$$\Delta = -2\int\int_0^\infty \big\{x - \phi(x+t) - \mu\big\}\phi'(x+t)\mathrm{d}tf(x-\mu)\mathrm{d}x$$

となる. $w = x + t - \mu$, $u = w - t$, $(\mathrm{d}w = \mathrm{d}x,\ \mathrm{d}u = -\mathrm{d}t)$ と変数変換すると,

$$\begin{aligned}\Delta =& -2\int\int_0^\infty \big\{w - t - \phi(w+\mu)\big\}\phi'(w+\mu)f(w-t)\mathrm{d}t\mathrm{d}w\\ =& -2\int\int_{-\infty}^{w} \big\{u - \phi(w+\mu)\big\}\phi'(w+\mu)f(u)\mathrm{d}u\mathrm{d}w\end{aligned}$$

となり, 補題で与えられる式が成り立つ. ■

補題 2.2 から, $\widehat{\mu}_\phi$ が $\widehat{\mu}^U$ を改良するための条件を導くことができる.

定理 2.2 次の (a), (b) の条件を仮定する.

(a) $\phi(w)$ は w の非減少関数で, $\lim_{w\to\infty}\phi(w) = c_0$ とする.

(b) 不等式 $\phi(w) \geq \phi^{GB}(w)$ がみたされる. ただし, $\phi^{GB}(w)$ は次で与えられる.

$$\phi^{GB}(w) = \int_{-\infty}^{w} uf(u)\mathrm{d}u \Big/ \int_{-\infty}^{w} f(u)\mathrm{d}u$$

このとき, $\widehat{\mu}_\phi$ は $\widehat{\mu}^U$ を改良する.

証明. 条件 (a) より, 補題 2.2 の表現式が成り立ち, また $\phi(\cdot)$ の非減少性に注意すると,

$$\begin{aligned}\Delta =& -2\int\Big\{\int_{-\infty}^{w} uf(u)\mathrm{d}u - \phi(w+\mu)\int_{-\infty}^{w} f(u)\mathrm{d}u\Big\}\phi'(w+\mu)\mathrm{d}w \\ \geq& -2\int\Big\{\int_{-\infty}^{w} uf(u)\mathrm{d}u - \phi(w)\int_{-\infty}^{w} f(u)\mathrm{d}u\Big\}\phi'(w+\mu)\mathrm{d}w\end{aligned}$$

となることがわかる. 条件 (b) より この値は 0 となるので, $\Delta \geq 0$ が示される. ■

定理 2.2 の条件 (a), (b) をみたす推定量の例として, D_0 上の一様分布に対する一般化ベイズ推定量と打ち切り推定量があげられる.

例 2.1 (一般化ベイズ推定量) $\phi^{GB}(w)$ を w で微分することにより, $\phi^{GB}(w)$ が非減少であることがわかる. また $\lim_{w\to\infty}\phi^{GB}(w) = c_0$ より, 条件 (a) がみたされる. 条件 (b) は明らかであるので, 定理 2.2 より, $\widehat{\mu}^{GB} = X - \phi^{GB}(X)$ は X を改良することがわかる. $\widehat{\mu}^{GB}$ は, 制約された空間 D_0 上に一様分布を想定したときの一般化ベイズ推定量に一致する. すなわち,

$$\widehat{\mu}^{GB} = X - \phi^{GB}(X) = \int_0^\infty \xi f(X-\xi)\mathrm{d}\xi \Big/ \int_0^\infty f(X-\xi)\mathrm{d}\xi \tag{2.2}$$

である. Farrell (1964) は, $\widehat{\mu}^{GB}$ のミニマックス性と許容性を証明している. ■

例 2.2 (打ち切り推定量) $\phi^{TR}(w) = \min\{w, c_0\}$ とすると, これは w に関して非減少であり, また $\lim_{w\to\infty}\phi^{TR}(w) = c_0$ をみたしている. 例 2.1 より, $\phi^{GB}(w) \leq \min\{w, 0\}$ が成り立つことがわかるので, $\phi^{TR}(w) \geq \phi^{GB}(w)$ となり, 条件 (b) がみたされる. 従って, 定理 2.2 より, 打ち切り推定量

$$\widehat{\mu}^{TR} = X - \phi^{TR}(X) = \max\{X - c_0, 0\}$$

は, X を改良することがわかる. 正規分布のとき, 打ち切り推定量は最尤推定量になっている. ■

定理 2.2 は, 一般化ベイズ推定量や打ち切り推定量以外にも改良する推定量を与えることができるが, その推定量がベイズ推定量になっているかが興味あるところであり, Maruyama, Iwasaki (2005) はそれに関連した結果を与えている.

一般化ベイズ推定量 $\widehat{\mu}^{GB}$ のリスク関数を描いてみると, $\mu = 0$ の点でリスクの値がミニマックス・リスク $R(\mu, \widehat{\mu}^U)$ に一致することがわかる. この事実は, 補題 2.2 を用いれば解析的にも確かめられる. 実際, $\mu = 0$ の点で, 次の等式が成り立つ.

$$R(0,\widehat{\mu}^U) - R(0,\widehat{\mu}^{GB}) = -2\int\Big\{\int_{-\infty}^{w} uf(u)\mathrm{d}u - \phi^{GB}(w)\int_{-\infty}^{w} f(u)\mathrm{d}u\Big\}\{\phi^{GB}(w)\}'\mathrm{d}w = 0$$

命題 2.1 *$\mu = 0$ で, $R(0,\widehat{\mu}^U) = R(0,\widehat{\mu}^{GB})$ となり, $\widehat{\mu}^{GB}$ のリスクは $\widehat{\mu}^U$ のリスクに一致する.*

IERD 法を用いることによって, 改良結果が一般の位置分布族へ拡張できるようになるが, Hartigan の結果のように p-次元の凸集合へ拡張することは難しいようである. IERD 法の他の母数制約問題への適用は Kubokawa, Saleh (1998) で与えられている.

3　母数制約のもとでのミニマックス性

前節では, 制約された母数空間上の一様分布に対する一般化ベイズ推定量が最良共変推定量を改良することを示した. 最良共変推定量とは, 母数空間に制約がないときの最良共変推定量のことで, この場合, 適当な条件のもとでミニマックスになることが知られている. この節では, 最良共変推定量のミニマックス性が, 母数空間が制約されている場合にも維持されるための条件を与える.

3.1　Girshick-Savage の方法

まず, 簡単な例として, 2.2 節で扱った位置母数の推定問題 $X \sim f(x-\mu)$, $\mu > 0$, を取り上げ, 母数制約のもとで最良共変推定量 $\widehat{\mu}^U = X - c_0$ のミニマックス性がどのように示されるかをみてみたい. ここで用いられる方法は, Kubokawa (2004) で与えられたもので, Girshick, Savage (1951) の方法を, 母数が制約されている場合へ適用したものである.

命題 3.1 位置母数 μ が $\mu > 0$ に制約されている場合, $\widehat{\mu}^U$ はミニマックスになる.

証明. $A_k = \{\mu |\ 0 < \mu < k\}$, $k = 1, 2, \ldots,$ とし, 事前分布の列

$$\pi_k(\mu) = \begin{cases} k^{-1}, & \mu \in A_k \text{ のとき} \\ 0, & \text{その他}, \end{cases}$$

を考えると, 対応するベイズ推定量の列は

$$\widehat{\mu}_k^\pi = \widehat{\mu}_k^\pi(X) = \int_{A_k} af(X-a)\mathrm{d}a \int_{A_k} f(X-a)\mathrm{d}a$$

となり, そのときのベイズリスク関数は

$$r_k(\pi_k, \widehat{\mu}_k^\pi) = \frac{1}{k}\int_{A_k}\int \{\widehat{\mu}_k^\pi(x) - \mu\}^2 f(x-\mu)\mathrm{d}x\mathrm{d}\mu \tag{3.1}$$

と書ける. R_0 を $\widehat{\mu}^U$ のリスクとすると, これは定数であり, $r_k(\pi_k, \widehat{\mu}_k^\pi) \le r_k(\pi_k, \widehat{\mu}^U) = R_0$ が成り立つので, $\liminf_{k\to\infty} r_k(\pi_k, \widehat{\mu}_k^\pi) \ge R_0$ を示せば十分である. 変数変換 $z = x - \mu$, $(\mathrm{d}z = \mathrm{d}x)$, $t = a - \mu$, $(\mathrm{d}t = \mathrm{d}a)$ を行うと,

$$\begin{aligned}
\widehat{\mu}_k^\pi(x) - \mu =& \widehat{\mu}_k^\pi(z+\mu) - \mu \\
=& \int_{A_k} (a-\mu) f(z+\mu-a)\mathrm{d}a \int_{A_k} f(z+\mu-a)\mathrm{d}a \\
=& \int_{t+\mu\in A_k} tf(z-t)\mathrm{d}t \int_{t+\mu\in A_k} f(z-t)\mathrm{d}t
\end{aligned} \tag{3.2}$$

と書けることになる. ここで変数変換 $\xi = (2/k)(\mu - k/2)$, $(\mathrm{d}\xi = (2/k)\mathrm{d}\mu)$ を行う. この変数変換が証明のポイントになる. この変換により, A_k の範囲 $0 < \mu < k$ は $|\xi| < 1$ に移り, $0 < t + \mu < k$ は $-(k/2)(\xi+1) < t < (k/2)(1-\xi)$ に移る. そこで,

$$A_k^* = \{t |-(k/2)(\xi+1) < t < (k/2)(1-\xi)\}$$

とおくと, (3.2) と (3.1) は,

$$\widehat{\mu}_k^\pi(\boldsymbol{x}) - \mu = \int_{A_k^*} tf(\boldsymbol{z}-t)\mathrm{d}t \int_{A_k^*} f(\boldsymbol{z}-t)\mathrm{d}t \equiv \widehat{\mu}_k^*(\boldsymbol{z}|\xi), \tag{3.3}$$

$$r_k(\pi_k, \widehat{\mu}_k^\pi) = \frac{1}{2}\int_{|\xi|<1}\int \{\widehat{\mu}_k^*(\boldsymbol{z}|\xi)\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}\xi$$

となる．十分小さい $\varepsilon > 0$ と正値関数 $h(\xi)$ に対して，

$$\int_{|\xi|<1} h(\xi)\mathrm{d}\xi = \int_{|\xi|<1-\varepsilon} h(\xi)\mathrm{d}\xi + \int_{1-\varepsilon<|\xi|<1} h(\xi)\mathrm{d}\xi \geq \int_{|\xi|<1-\varepsilon} h(\xi)\mathrm{d}\xi,$$

となることに注意すると，

$$r_k(\pi_k, \widehat{\mu}_k^\pi) \geq \frac{1}{2}\int_{|\xi|<1-\varepsilon}\int \{\widehat{\mu}_k^*(\boldsymbol{z}|\xi)\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}\xi$$

が成り立つことがわかる．(3.3) で与えられる $\widehat{\mu}_k^*(\boldsymbol{z}|\xi)$ の中の積分範囲は，$A_k^* = \{t| -(k/2)(\xi+1) < t < (k/2)(1-\xi)\}$ であることに注意しよう．$|\xi| < 1-\varepsilon$ より，$1-\xi > 1-(1-\varepsilon) = \varepsilon > 0$, $1+\xi > 1+(-1+\varepsilon) = \varepsilon > 0$ となるので，$k \to \infty$ とすると端点 $(k/2)(1-\xi)$, $-(k/2)(1+\xi)$ はそれぞれ ∞, $-\infty$ に発散し，$\widehat{\mu}_k^*(\boldsymbol{z}|\xi)$ は $\widehat{\mu}^U(\boldsymbol{z})$ に収束する．Fatou の補題を用いると，

$$\begin{aligned}
\liminf_{k\to\infty} r_k(\pi_k, \widehat{\mu}_k^\pi) \geq & \liminf_{k\to\infty} \frac{1}{2}\int_{|\xi|<1-\varepsilon}\int \{\widehat{\mu}_k^*(\boldsymbol{z}|\xi)\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}\xi \\
\geq & \frac{1}{2}\int_{|\xi|<1-\varepsilon}\int \left\{\liminf_{k\to\infty} \widehat{\mu}_k^*(\boldsymbol{z}|\xi)\right\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}\xi \\
= & \frac{1}{2}\int_{|\xi|<1-\varepsilon}\mathrm{d}\xi \int \left\{\widehat{\mu}^U(\boldsymbol{z})\right\}^2 f(\boldsymbol{z})\mathrm{d}\boldsymbol{z} \\
= & (1-\varepsilon)R(\mu, \widehat{\mu}^U) = (1-\varepsilon)R_0
\end{aligned}$$

となり，$\varepsilon > 0$ の任意性から $\liminf_{k\to\infty} r_k(\pi_k, \widehat{\mu}_k^\pi) \geq R_0$ が成り立つ．以上から，命題 3.1 が示される． ■

3.2 ミニマックス性に対する一般的な条件

前節で用いた Girshick-Savage の方法を用いると，群構造の入った一般的な問題設定のもとで最良共変推定量がミニマックスなるための条件が得られる．ここでは，Hora, Buehler (1966) により与えられた枠組みで説明する．X を確率変数，$(\mathcal{X}, \mathcal{B})$ を X の可測空間，Θ を母数空間，$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ を母数により識別可能な確率分布族とし，次の仮定をみたすものとする．

(A1) 群 $\mathcal{G} = \{g\}$ と可測空間 $(\mathcal{G}, \mathcal{B}_G)$ があり，任意の $g \in \mathcal{G}$ と任意の $G \in \mathcal{B}_G$ に対して $\gamma(gG) = \gamma(G)$ をみたす左不変 Haar 測度 $\gamma(\cdot)$ が存在する．各 $g \in \mathcal{G}$ に対して，$P_{\overline{g}\theta}(gA) = P_\theta(A)$ となる Θ 上の 1 対 1 変換群 $\overline{g}$ がとれて，$\overline{\mathcal{G}} = \{\overline{g} : \overline{g} \in \mathcal{G}\}$ は可測であるとする．

(A2) $\mathcal{U}$ を可測空間とし，$\mathcal{X}$ と $\mathcal{G} \times \mathcal{U}$ の間の写像 $X \leftrightarrow (T, U)$ は 1 対 1 対応であり，gX は (gT, U) に対応するものとする．統計量 U は $\mathcal{G}$ について最大不変量である．

(A3) Θ と $\overline{\mathcal{G}}$ の間の写像 $\theta \leftrightarrow \overline{g}_\theta$ は 1 対 1 対応であり，$\overline{g}\theta$ は $\overline{g}\,\overline{g}_\theta$ に対応するものとする．また $\overline{g}_\theta$ は $\mathcal{G}$ では g_θ に対応するとする．

(A4) U を与えたときの T の条件付き確率密度関数 $p(g_\theta^{-1}t|u)$ が存在し，$A \in \mathcal{B}$ に対して，$P_\theta[A] = \int_A p(g_\theta^{-1}t|u)p(u)\gamma(\mathrm{d}t)\gamma(\mathrm{d}u)$ と表されるとする．ここで，$p(\cdot)$ は，$\mathcal{U}$ 上の測度 $\gamma(\cdot)$ に関する周辺密度関数とする．

測度 $\nu(\cdot)$ を $\nu(\mathrm{d}g) = \gamma(\mathrm{d}g^{-1})$ で定義すると, これは, 右不変 Haar 測度となる. このとき, 最良共変推定量は, 右不変 Haar 測度 $\nu(\mathrm{d}g)$ に対する一般化ベイズ推定量になる. これを $\hat{\theta}^U$ で表す. $\hat{\theta}^U$ がミニマックスになるための十分条件は, 群 $\mathcal{G}$ 上に確率測度の列 $\gamma_j(\cdot)$ で, 任意の $g \in \mathcal{G}$ と $\mathcal{G}$ 上の任意の有界な可測関数 ψ に対して $\lim_{j\to\infty} \int \{\psi(ag) - \psi(a)\}\gamma_j(\mathrm{d}a) = 0$ となることが知られている. しかし, 母数空間が制約されているときには, この結果を用いることができない. そこで, 前節の Girshick-Savage の方法を用いることによって, 母数が制約されているときのミニマックス性に関する条件を求めてみると次のようになる.

(A5) 母数空間 Θ が制約されており, この制約が $g_\theta \in D$ と表されるとする. また $D \subset \mathcal{G} \subset \boldsymbol{R}^r$ とする.

(A6) Ξ を r 次元の立方体 $[-1, 1]^r$ とする. 部分集合の列 D_k と 1 対 1 関数 $h_k(\cdot) : D_k \leftrightarrow \Xi$ で, 次の条件をみたすものが存在すると仮定する.

(A6-1) $\cup_{k=k_0}^{\infty} D_k = D$, $(k_0 \geq 1)$

(A6-2) $V(D_k) = \int_{D_k} \nu(\mathrm{d}g_\theta)$ とおき, $A \in \Xi$ に対して $\gamma_k(A) = \nu(h_k^{-1}(A))$ とおく. $\xi = h_k(g_\theta)$ とし, $I(\cdot)$ を指標関数とするとき, $h_k(D_k) = \prod_{i=1}^r [-1 + a_{i,k}, 1 + b_{i,k}]$ と表されて, 不等式

$$\int_{h_k(D_k)} f(\xi)\gamma_k(\mathrm{d}\xi)/V(D_k) \geq \frac{1}{2^r + c_k} \int I\Big(\xi \in \prod_{i=1}^r [-1 + a_{i,k}, 1 + b_{i,k}]\Big) f(\xi)\mathrm{d}\xi, \tag{3.4}$$

が成り立ち, $\lim_{k\to\infty} a_{i,k} = \lim_{k\to\infty} b_{i,k} = \lim_{k\to\infty} c_k = 0$, $(i = 1, \ldots, r)$ をみたすものとする.

(A6-3) $g_\theta g \in D_k$ すなわち $[h_k^{-1}(\xi)]g \in D_k$ が $g \in \widetilde{D}_k(\xi)$ と表現できるとする. 十分小さい $\varepsilon > 0$ と任意の $\xi \in \prod_{i=1}^r [-1 + a_{i,k} + \varepsilon, 1 + b_{i,k} - \varepsilon]$ に対して, ξ に依存しない部分集合の列 D_k^* で, $\cup_{k=k_1}^{\infty} D_k^* = \mathcal{G}$, $(k_1 \geq 1)$ と $D_k^* \subset \widetilde{D}_k(\xi)$ をみたすものが存在する.

定理 3.1 条件 **(A1)** から **(A6)** のもとで, 最良共変推定量 $\hat{\theta}^U$ は母数空間が制約されているときにもミニマックスになる.

定理 3.1 は, 命題 3.1 と Kubokawa, Marchand, Strawderm, Turcotte (2012) と同様にして示される. この定理と関連した統一的な結果として, Marchand, Strawderman (2012) は, ある推定量が制約されていない推定問題においてミニマックスならば制約されていてもミニマックス性を維持できるための一般的な条件を導出した.

3.3 位置尺度分布族の例

例 3.1 (位置分布族) 確率変数 $\boldsymbol{X} = (X_1, \ldots, X_n)$ が位置母数をもつ密度関数 $f(x_1 - \mu, \ldots, x_n - \mu)$ に従い, μ が $D = \{\mu \geq a_0\}$ に制約されているとする. この場合, $T = X_1$, $U = (X_2 - X_1, \ldots, X_n - X_1)$ に対応する. 定理 3.1 の条件 (A6) がみたされることを確かめてみよう.

$\mathcal{G} = \boldsymbol{R}$, $\gamma(\mathrm{d}\mu) = \nu(\mathrm{d}\mu) = \mathrm{d}\mu$, $D_k = \{\mu | a_0 < \mu < a_0 + k\}$, $V(D_k) = k$ とする. $\xi = h_k(\mu) = (2/k)(\mu - a_0) - 1$ とおくと, $h_k(D_k) = [-1, 1]$, $\gamma_k(\mathrm{d}\xi) = (k/2)\mathrm{d}\xi$, $\int_{h_k(D_k)} f(\xi_k)\gamma_k(\mathrm{d}\xi_k)/V(D_k) = (1/2)\int_{[-1,1]} f(\xi)\mathrm{d}\xi$ となり, (A6-2) が成り立つ.

$\xi \in [-1+\varepsilon, 1-\varepsilon]$ に対して, $\mu = h_k^{-1}(\xi) = a_0 + (k/2)(\xi+1)$ となるので, $\widetilde{D}_k(\xi) = \{[h_k^{-1}(\xi)]^{-1}g; g \in D_k\} = \{\mu - a_0 - (k/2)(\xi+1); a_0 < \mu < a_0 + k\} = (-(k/2)(\xi+1), (k/2)(1-\xi)) \supset (-(k/2)\varepsilon, (k/2)\varepsilon) \equiv D_k^*$ が成り立つ. $\lim_{k\to\infty} D_k^* = \boldsymbol{R}$ より, (A6-3) がみたされ, ミニマックス性が成り立つ. 尺度分布族についても, 確率変数を対数変換すると位置分布族に帰着されるので, 同様な結果が成り立つ. ■

次に，位置尺度分布族について考えてみよう．確率変数 $\boldsymbol{X} = (X_1, \ldots, X_n)$ が，密度関数 $\sigma^{-n} f((x_1-\mu)/\sigma, \ldots, (x_n-\mu)/\sigma))$ に従うとし，

$$D = \{(\mu, \sigma) | \mu > c_0\sigma + a_0,\ \sigma > b_0\}$$

に制約されているとする．ここで，a_0, b_0, c_0 は $b_0 \geq 0$, $-\infty \leq a_0, c_0 < \infty$ をみたす定数とし，$T = (|X_2 - X_1|, X_1)$, $U = ((X_2 - X_1)/|X_2 - X_1|, \ldots, (X_n - X_1)/|X_2 - X_1|)$ に対応する．

例 3.2 (片側制約の場合) [1] $a_0 > -\infty$, $b_0 > 0$ の場合．$c_0 = 0$ のときには $d_k = k$ とし，$c_0 \neq 0$ のときには $d_k = \log k$ とする．$k \to \infty$ のとき，次の性質をみたす．

(a) $c_0 = 0$ のとき，$\varepsilon > 0$ に対して $(k/d_k)d_k^{\varepsilon/2} \to \infty$

(b) $c_0 \neq 0$ のとき，$d_k/k \to 0$, $d_k \to \infty$

定理 3.1 の条件 (A6-1)-(A6-3) を調べる．この場合，$D = \{(\sigma, \mu)|a_0 + c_0\sigma < \mu, b_0 < \sigma\}$, $\mathcal{G} = \boldsymbol{R}^+ \times \boldsymbol{R}$, $D_k = \{(\sigma, \mu)|a_0 + c_0\sigma < \mu < a_0 + c_0\sigma + k, b_0 < \sigma < b_0 d_k\}$, $V(D_k) = k \log d_k$ が成り立つ．$\xi_1 = (2/\log d_k)\log(\sigma/b_0) - 1$, $\xi_2 = (2/k)(\mu - a_0 - c_0\sigma) - 1$ とおく．$\boldsymbol{\xi} = (\xi_1, \xi_2) = h_k((\sigma, \mu))$ とおくと，$h_k(D_k) = [-1, 1]^2$, $\gamma_k(\mathrm{d}\boldsymbol{\xi}) = \{(k \log d_k)/4\}\mathrm{d}\boldsymbol{\xi}$, $\int_{h_k(D_k)} f(\boldsymbol{\xi})\gamma_k(\mathrm{d}\boldsymbol{\xi})/V(D_k) = (1/4)\int_{[-1,1]^2} f(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}$ となり，(A6-2) が成り立つ．

任意の $\boldsymbol{\xi} \in [-1+\varepsilon, 1-\varepsilon]^2$ に対して，$(b, a) = h_k^{-1}(\boldsymbol{\xi})$ とおくと，$b = b_0 d_k^{(1+\xi_1)/2}$, $a = (k/2)(1+\xi_2) + a_0 + c_0 b_0 d_k^{(1+\xi_1)/2}$ となるので，$\{[h_k^{-1}(\boldsymbol{\xi})]^{-1}(\sigma, \mu); (\sigma, \mu) \in D_k\} = \{(\sigma/b, (\mu - a)/b); (\sigma, \mu) \in D_k\}$ となり，σ/b, $(\mu - a)/b$ が不等式

$$\begin{aligned} d_k^{-(1+\xi_1)/2} &< \frac{\sigma}{b} < d_k^{(1-\xi_1)/2}, \\ c_0\frac{\sigma}{b} - \frac{d_k^{-(1+\xi_1)/2}}{b_0}\{\frac{k}{2}(1+\xi_2) + c_0 b_0 d_k^{(1+\xi_1)/2}\} & \\ &< \frac{\mu - a}{b} < c_0\frac{\sigma}{b} + \frac{d_k^{-(1+\xi_1)/2}}{b_0}\{\frac{k}{2}(1-\xi_2) - c_0 b_0 d_k^{(1+\xi_1)/2}\} \end{aligned}$$

をみたす．$1 - \xi_i > \varepsilon$, $1 + \xi_i > \varepsilon$ に注意する．最初の不等式は，$d_k^{-\varepsilon/2} < \sigma/b < d_k^{\varepsilon/2}$ によりみたされ，これは，$d_k \to \infty$, $(k \to \infty)$ とすると，$(0, \infty)$ に広がる．また，2 番目の不等式も

$$\frac{k}{d_k^{(1+\xi_1)/2}}\left\{c_0\frac{\sigma}{b}\frac{d_k^{(1+\xi_1)/2}}{k} - \frac{\varepsilon}{2b_0} - \frac{c_0}{k}\right\} < \frac{\mu - a}{b} < \frac{k}{d_k^{(1+\xi_1)/2}}\left\{c_0\frac{\sigma}{b}\frac{d_k^{(1+\xi_1)/2}}{k} + \frac{\varepsilon}{2b_0} - \frac{c_0}{k}\right\}$$

によりみたされる．ここで，

$$\begin{aligned} \frac{\sigma}{b}\frac{d_k^{(1+\xi_1)/2}}{k} &< d_k^{(1-\xi_1)/2}\frac{d_k^{(1+\xi_1)/2}}{k} < \frac{d_k}{k}, \\ \frac{k}{d_k^{(1+\xi_1)/2}} &= \frac{k}{d_k}d_k^{(1-\xi_1)/2} > \frac{k}{d_k}d_k^{\varepsilon/2} \end{aligned}$$

に注意する．d_k は条件 (a) もしくは (b) をみたすので，$(\mu - a)/b$ の下限は $-\infty$, 上限は ∞ に発散する．従って，(A6-3) が成り立ち，ミニマックス性が成り立つ．[2] $a_0 = -\infty$, $b_0 > 0$ の場合，[3] $a_0 > -\infty$, $b_0 = 0$ の場合も同様に考えることができる． ■

例 3.3 (両側制約の場合) 定数 a_1, a_2, $(a_2 > a_1)$, 正の定数 b_0 に対して

$$D = \{(\sigma, \mu)|a_1 < \mu < a_2,\ 0 < \sigma < b_0\}$$

なる制約を考える．これは μ が両側から制約を受ける場合である．

この場合, $kb_0 > 1$ なる k に対して $D_k = \{(\sigma,\mu)|a_1 < \mu < a_2, b_0/k < \sigma < b_0\}$ とおくと $V(D_k) = (a_2 - a_1)\log k$ となる．$\xi_1 = (2/\log k)\log(\sigma/b_0) + 1$, $\xi_2 = \{2/(a_2 - a_1)\}\{\mu - (a_1 + a_2)/2\}$ とおく．$\boldsymbol{\xi} = (\xi_1, \xi_2) = h_k((\sigma,\mu))$ とおくと, $h_k(D_k) = [-1,1]^2$, $\gamma_k(\mathrm{d}\boldsymbol{\xi}) = \{(a_2 - a_1)\log k)/4\}\mathrm{d}\boldsymbol{\xi}$, $\int_{h_k(D_k)} f(\boldsymbol{\xi})\gamma_k(\mathrm{d}\boldsymbol{\xi})/V(D_k) = (1/4)\int_{[-1,1]^2} f(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}$ となるので, (A6-2) が成り立つ．

任意の $\boldsymbol{\xi} \in [-1+\varepsilon, 1-\varepsilon]^2$ に対して, $(b,a) = h_k^{-1}(\boldsymbol{\xi})$ とおく．このとき, $b = b_0 k^{(\xi_1 - 1)/2}$, $a = \{(a_2 - a_1)/2\}\xi_2 + (a_1 + a_2)/2$ となるので $\{[h_k^{-1}(\boldsymbol{\xi})]^{-1}(\sigma,\mu); (\sigma,\mu) \in D_k\} = \{(\sigma/b, (\mu - a)/b); (\sigma,\mu) \in D_k\}$ となり, σ/b, $(\mu - a)/b$ は不等式

$$k^{-(1+\xi_1)/2} < \frac{\sigma}{b} < k^{(1-\xi_1)/2},$$
$$-\frac{a_2 - a_1}{2b_0}(1+\xi_2)k^{-(\xi_1 - 1)/2} < \frac{\mu - a}{b} < \frac{a_2 - a_1}{2b_0}(1 - \xi_2)k^{-(\xi_1 - 1)/2}$$

をみたす．これらの不等式は, $k^{-\varepsilon/2} < \sigma/b < k^{\varepsilon/2}$,

$$-\frac{a_2 - a_1}{2b_0}\varepsilon k^{-\varepsilon/2} < \frac{\mu - a}{b} < \frac{a_2 - a_1}{2b_0}\varepsilon k^{-\varepsilon/2}$$

によってみたされ, $k \to \infty$ のとき範囲が $(0,\infty)$, $(-\infty,\infty)$ に広がるので, (A6-3) がみたされる．従って, ミニマックス性が成り立つ． ■

3.4 凸多面錐に制約されている場合のミニマックス性

平均の代表的な順序制約は凸多面錐への制約として捉えることができ, そのときの一般化ベイズ推定量の性質が Tsukuma, Kubokawa (2008) により議論された．確率変数 $\boldsymbol{X} = (X_1, \ldots, X_p)^t$ の分布が $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{I}_p)$ に従い, 平均ベクトル $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^t$ が凸多面錐

$$D = \{\boldsymbol{\mu}|\boldsymbol{r}_i^t\boldsymbol{\mu} \le \alpha_i,\ i = 1, \ldots, q\} = \{\boldsymbol{\mu}|\boldsymbol{R}\boldsymbol{\mu} \le \boldsymbol{\alpha}\}$$

に制約されているとする．ここで, $\boldsymbol{R} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_q)^t$ は $q \times p$ 行列, 既知でランク $q\ (\le p)$ とし, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)^t$ は既知のベクトルとする．この制約は, 例えば, $D_1 = \{\boldsymbol{\mu}|\mu_i \ge 0,\ i = 1, 2, \ldots, p\}$, $D_2 = \{\boldsymbol{\mu}|\mu_1 \le \mu_2 \le \cdots \le \mu_p\}$, $D_3 = \{\boldsymbol{\mu}|\mu_1 \le \mu_i,\ i = 2, 3, \ldots, p\}$, $D_4 = \{\boldsymbol{\mu}|\mu_1 \le \cdots \le \mu_\ell \ge \cdots \ge \mu_p\}$, $D_5 = \{\boldsymbol{\mu}|\mu_{i+1} - \mu_i \le \mu_{i+2} - \mu_{i+1},\ i = 1, \ldots, p-2\}$ を含んでいる．

制約された母数の推定方法としては, 制約された最尤推定量 (REML) や D 上に一様分布を想定したときの一般化ベイズ推定量が代表的で, 一般化ベイズ推定量は $\widehat{\boldsymbol{\mu}}^{GB}(\boldsymbol{X}) = \int_D \boldsymbol{\xi} f(\boldsymbol{X} - \boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}/\int_D f(\boldsymbol{X} - \boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}$ で与えられる．どちらが望ましいかについては, 例えば制約 D_2 を考えた場合, μ_i と μ_{i+1} の REML が同じ推定値を与える場合があるのに対して, 一般化ベイズ推定量はそれぞれの成分に対して異なっていてしかも単調増加する推定値を与える点で好ましいと思われるが, p 重積分を計算する必要が生ずる点で計算上の困難さを伴っている．まず, 定理 3.1 から, 次の命題が成り立つ．

命題 3.2 *D に制約された母数の推定問題において, 最良共変推定量 $\boldsymbol{X}$ はミニマックスになる．また定理 2.1 の Hartigan の結果から, 一般化ベイズ推定量 $\widehat{\boldsymbol{\mu}}^{GB}$ はミニマックスになる．*

$\boldsymbol{X}$ は明らかに非許容的であるが, $\widehat{\boldsymbol{\mu}}^{GB}$ の許容性についてはどうなるであろうか．1つの可能性として, $p \ge 3$ のとき, Stein 効果により $\widehat{\boldsymbol{\mu}}^{GB}$ は縮小推定量

$$\widehat{\boldsymbol{\mu}}^{SH} = \boldsymbol{\alpha}_* + \left\{1 - \frac{p-2}{\|\widehat{\boldsymbol{\mu}}^{GB}(\boldsymbol{X}) - \boldsymbol{\alpha}_*\|^2}\right\}(\widehat{\boldsymbol{\mu}}^{GB}(\boldsymbol{X}) - \boldsymbol{\alpha}_*)$$

により改良されるという予想が立てられる. ここで, $\boldsymbol{\alpha}_* = \boldsymbol{R}^t(\boldsymbol{R}\boldsymbol{R}^t)^{-1}\boldsymbol{\alpha}$ である. しかし, $\widehat{\boldsymbol{\mu}}^{GB}$ に含まれる積分範囲 D が複雑であるときには $\widehat{\boldsymbol{\mu}}^{SH}$ のリスク評価は簡単でない. Tsukuma, Kubokawa (2008) は, $q < p$ のとき $\boldsymbol{R}\boldsymbol{1}_p = \boldsymbol{0}_q$ が成り立ち, $\widehat{\boldsymbol{\mu}}^{GB}$ が p-次元ベクトル $\boldsymbol{q}(\boldsymbol{X})$ を用いて $\widehat{\boldsymbol{\mu}}^{GB}(\boldsymbol{X}) = \boldsymbol{X} + \boldsymbol{R}^t\boldsymbol{q}(\boldsymbol{X})$ と表され, $\boldsymbol{1}_p^t\widehat{\boldsymbol{\mu}}^{GB}(\boldsymbol{X}) = \boldsymbol{1}_p^t\boldsymbol{X}$ をみたすことに注意して, $\widehat{\boldsymbol{\mu}}^{GB}$ の非許容性の証明に成功した. また $p = 1, 2$ のときの許容性を制約 D_1 と D_2 のときに示している.

定理 3.2 $p \geq 3$ のとき, 凸多面錐 D に制約された一般化ベイズ推定量 $\widehat{\boldsymbol{\mu}}^{GB}$ は $\widehat{\boldsymbol{\mu}}^{SH}$ により改良される. すなわち, $\widehat{\boldsymbol{\mu}}^{SH}$ はミニマックスになる. 特に制約が D_1, D_2 で与えられるときには, $p = 1, 2$ のとき $\widehat{\boldsymbol{\mu}}^{GB}$ は許容的である.

この定理は, James, Stein (1961) の結果を凸多面錐へ拡張したものである. 制約 D_1 で, $p = 1$ のときの一般化ベイズ推定量の許容性は Katz (1961) により示されている.

4 両側から制約されている場合のミニマックス性と非ミニマックス性

前節で扱われた制約は, 直感的には片側からの制約で, もう片方が無限に広がる空間をもつことになるので, 最良共変推定量はミニマックス性を維持できると考えられる. では, 両側から制約を受ける場合にはミニマックス性を保持できないと予想されるが, これは正しいであろうか. 実は, この問題は, 分散が既知の場合と未知の場合とで異なった結果を与えることが知られている. こうした結果を含め, この節では, 両側から制約されている母数空間における推定問題について解説する.

4.1 最良共変推定量の非ミニマックス性と関連する結果

Casella, Strawderman (1981) は, X が $\mathcal{N}(\mu, 1)$ に従う確率変数とし, μ が両側から制約されるときの μ の推定問題を扱った. 例えば, μ が $B_d = \{\mu \,|\, |\mu| \leq d\}$ に制約されているとしよう. ここで d は正の定数である. 彼らは, μ の両端 $\mu = \pm d$ に等しい確率 $P[\mu = -d] = 1/2$, $P[\mu = d] = 1/2$ をもつ事前分布を考えた. このときのベイズ推定量は

$$\widehat{\mu}^{BU} = d\tanh(dX) = d\frac{e^{dX} - e^{-dX}}{e^{dX} + e^{-dX}} \tag{4.1}$$

と表されるが, Casella, Strawderman (1981) は, 直接 $\widehat{\mu}^{BU}$ のリスクを評価することによって, $d \leq 1.0567$ のときに $\widehat{\mu}^{BU}$ がミニマックスになることを証明した. このことから, 最良共変推定量 X はミニマックスでないことがわかる. この事実は多くの研究者の興味を引き, Bickel (1981), Berry (1990), Marchand, Perron (2001, 2005) などにより関連する研究がなされてきた.

$\widehat{\mu}^{BU}$ の ミニマックス性は, d が小さいとき, すなわち制約空間 B_d が狭いときに成り立つことからわかるように, 両側制約の研究は技術的に扱いが難しく未解決な問題が多いように思われる. B_d 上の一様分布を事前分布にとったときのベイズ推定量は

$$\widehat{\mu}^{FU} = \frac{\int_{-d}^{d} \mu \exp\{-(X-\mu)^2/2\}\mathrm{d}\mu}{\int_{-d}^{d} \exp\{-(X-\mu)^2/2\}\mathrm{d}\mu} = X - \frac{\int_{X-d}^{X+d} z\exp\{-z^2/2\}\mathrm{d}z}{\int_{X-d}^{X+d} \exp\{-z^2/2\}\mathrm{d}z}$$

で与えられる. Hartigan (2004) は $\widehat{\mu}^{FU}$ が X を改良することを示し, Marchand, Strawderman (2005), Kubokawa (2005b) は, 一般の位置分布族のもとで X を改良する推定量のクラスを求め,

その中に $\widehat{\mu}^{FU}$ が入っていることを示した. $\widehat{\mu}^{FU}$ はミニマックスではないが, 自然な形の推定量であり, すべての d に対して X を改良している.

平均 μ が B_d に制限されるときの最尤推定量は $\widehat{\mu}^{ML} = \min\{\max\{X, -d\}, d\}$ で与えられる. 2 乗損失のもとでは $\widehat{\mu}^{ML}$ は解析的でないので（一般化）ベイズ推定量で表現することができない. このことは, 一般論から $\widehat{\mu}^{ML}$ が許容的でないことを意味する. ところが, Iwasa, Moritani (1997) は, 絶対損失関数 $|\widehat{\mu} - \mu|$ に関しては $\widehat{\mu}^{ML}$ はある事前分布に対してベイズ推定量になっているという, 極めて興味深い結果を証明した. このことは, 解析的でない推定量 $\widehat{\mu}^{ML}$ であっても絶対損失のもとでは許容的になりうることを意味する.

4.2 Moors の方法

Moors (1981) は, 推定問題が対称性をもっているとき, 対称性の情報を用いることによって推定量を改良する方法を提案した. Marchand, Strawderman (2004) は, その方法を母数の両側制約の問題に適用した. 確率変数 X が位置母数をもつ密度関数 $f(x-\mu)$ に従い, $\mu \in B_d$ とし, $f(\cdot)$ は既知で対称な関数とする. 符号変換に関して共変な推定量のクラスは,

$$\widehat{\mu}_\phi(X) = \phi(|X|)X/|X| \tag{4.2}$$

と書かれる. このとき, $|X|$ を与えたときの条件付き期待値を考えることにより, $\widehat{\mu}_\phi$ のリスクは次のように表される.

補題 4.1 推定量 $\widehat{\mu}_\phi$ の 2 乗損失に関するリスクは,

$$R(\mu, \widehat{\mu}_\phi) = E[\mu^2 - \{a_{|X|}(\mu)\}^2] + E[\{\phi(|X|) - a_{|X|}(\mu)\}^2]$$

と表される. ここで, $a_w(\mu)$ は次で与えられる.

$$a_w(\mu) = \mu\frac{f(w-\mu) - f(w+\mu)}{f(w-\mu) + f(w+\mu)}$$

定理 4.1 集合 $A(w)$ を $A(w) = \{a_w(\mu) | -d \le \mu \le d\}$ とする. $\phi(w)$ の $A(w)$ への射影を $\phi_0(w)$ とおき, $\phi_0(w)$ をもった推定量 (4.2) を $\widehat{\mu}_{\phi_0}$ とおく. このとき, $R(\mu, \widehat{\mu}_{\phi_0}) \le R(\mu, \widehat{\mu}_\phi)$ が成り立つ.

正規分布のときには, $a_w(\mu) = \mu \tanh(\mu w)$ となり, $a_w(\mu)$ は $|\mu|$ に関して単調増加するので, $A(w) = [0, d\tanh(dw)]$ となる. (4.1) より, 両端点に等確率をもつ事前分布に対するベイズ推定量は, $\widehat{\mu}^{BU} = d\tanh(d|X|)X/|X|$ とかけるので, 区間 $A(|X|)$ の最大点は $\widehat{\mu}^{BU}$ により与えられることがわかる. 定理 4.1 より, (4.2) で与えられる推定量が $\widehat{\mu}^{BU}$ により改良される必要十分条件は $d\tanh(dw) \le \phi(w)$, $w > 0$, となる. 特に, 最尤推定量は $\widehat{\mu}^{ML} = \min(|X|, d)X/|X|$ と表されるので, $\widehat{\mu}^{BU}$ が $\widehat{\mu}^{ML}$ を改良する必要十分条件は, $d \le 1$ となることがわかる.

Marchand, Perron (2001) は, 確率変数 $\boldsymbol{X}$ が p-変量正規分布 $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{I}_p)$ に従う場合へ拡張した. ここで, $\boldsymbol{\mu}$ は, 半径 d の球 $B_{d,p} = \{\boldsymbol{\mu} \in \boldsymbol{R}^p | \|\boldsymbol{\mu}\| \le d\}$ へ制約されているとする. このとき, 直交共変な推定量は $\widehat{\boldsymbol{\mu}}_\phi(\boldsymbol{X}) = \phi(\|\boldsymbol{X}\|)\boldsymbol{X}/\|\boldsymbol{X}\|$ であり, 最尤推定量 $\widehat{\boldsymbol{\mu}}^{ML} = \min(d, \|\boldsymbol{X}\|)\boldsymbol{X}/\|\boldsymbol{X}\|$ もこのクラスに入る. Marchand, Perron (2001) は, $B_{d,p}$ の境界上の一様分布を事前分布とするベイズ推定量 $\widehat{\boldsymbol{\mu}}^{BU}$ を導出し, これが最尤推定量を改良するための十分条件は $d \le \sqrt{p}$ で与えられることを示した. Marchand, Perron (2005) は, この結果の球面対称分布への拡張を行った.

4.3 分散が未知の場合

平均が有界で分散 σ^2 が未知のとき, すなわち, 次のような制約空間を考えてみると,

$$\{(\mu,\sigma)||\mu| \le d,\ 0<\sigma<\infty\} \tag{4.3}$$

例 3.3 で示したように, 最良共変推定量はミニマックスになってしまう (Kubokawa (2005a)). 分散が既知のときにはミニマックスでないので, 分散が既知と未知の場合の間で最良共変推定量のミニマックス性が異なることを意味している. この現象は, μ が有界でも, μ/σ が有界でないことから生ずると思われる.

Kubokawa (2005a) は, μ/σ が有界である場合を考え, Marchand, Strawderman (2005) の結果を分散が未知のモデルへ拡張した. X, S は互いに独立な確率変数で $X \sim \mathcal{N}(\mu,\sigma^2)$, $S/\sigma^2 \sim \chi_m^2$ に従うとする. χ_m^2 は自由度 m のカイ 2 乗分布を表している. 推定量は損失関数 $(\widehat{\mu}-\mu)^2/\sigma^2$ に関して評価されるものとする. まず, $d>0$ に対して, 母数空間が

$$B_d^* = \{(\mu,\sigma)||\mu| \le d\sigma,\ 0<\sigma<\infty\},$$

に制約されている場合を考える. これは, 変動係数 μ/σ が両側から有界な区間 $|\mu/\sigma| \le d$ に制約されていることを意味する. Kubokawa (2005a) は, 絶対連続な関数 $\phi(\cdot)$ に対して

$$\widehat{\mu}_\phi = X - \sqrt{S}\frac{X}{|X|}\phi\left(|X|/\sqrt{S}\right) \tag{4.4}$$

なる形の, 尺度変換に関して共変な推定量のクラスを考え, Marchand, Strawderman (2005) の論法と IERD 法を用いて次の定理を示した.

定理 4.2 $\phi(w)$ が次の条件をみたすと仮定する.

(a) $\phi(0)=0$

(b) $\phi(w)$ は非減少関数とする.

(c) $\phi(w) \le \phi^U(w)$, ここで ϕ^U は $g(w)=\sqrt{w^2/(1+w^2)}$ に対して次で与えられる.

$$\begin{aligned}\phi^U(w) &= \int_0^\infty \int_{\sqrt{v}w-d}^{\sqrt{v}w+d} ue^{-u^2/2}v^{(m-1)/2}e^{-v/2}\mathrm{d}u\mathrm{d}v \int_0^\infty \int_{\sqrt{v}w-d}^{\sqrt{v}w+d} e^{-u^2/2}v^{m/2}e^{-v/2}\mathrm{d}u\mathrm{d}v \\ &= \frac{(1+w^2)^{-(m+1)/2}\int_0^\infty \left\{e^{dg(w)\sqrt{z}} - e^{-dg(w)\sqrt{z}}\right\} z^{(m+1)/2-1}e^{-z/2}\mathrm{d}z}{\int_w^\infty \int_0^\infty (1+y^2)^{-(m+3)/2}\left\{e^{dg(y)\sqrt{z}} - e^{-dg(y)\sqrt{z}}\right\} z^{(m+3)/2-1}e^{-z/2}\mathrm{d}z\mathrm{d}y}\end{aligned}$$

このとき, $\widehat{\mu}_\phi$ は X を改良する.

この結果から次のような推定量が X を改良することが示される.

[1] 事前分布 $\pi^{FU}(\mu,\sigma)$ に対する一般化ベイズ推定量 $\widehat{\mu}^{FU}$. $\pi^{FU}(\mu,\sigma)$ は, σ^2 を与えたときの μ の条件付き分布が, $\{\mu|-d\sigma<\mu<d\sigma\}$ 上の一様分布とし, σ^2 の密度が $\sigma^{-2}\mathrm{d}\sigma^2$ の形をしているとする. このとき, $\widehat{\mu}^{FU}$ は次のように書かれる.

$$\widehat{\mu}^{FU} = \frac{\int_0^\infty \int_{-d\sigma}^{d\sigma} \mu\sigma^{-(m+5)}e^{-\{(X-\mu)^2+S\}/(2\sigma^2)}\mathrm{d}\mu\mathrm{d}\sigma^2}{\int_0^\infty \int_{-d\sigma}^{d\sigma} \sigma^{-(m+5)}e^{-\{(X-\mu)^2+S\}/(2\sigma^2)}\mathrm{d}\mu\mathrm{d}\sigma^2}$$

[2] 事前分布 $\pi^{BU}(\mu,\sigma)$ に対する一般化ベイズ推定量 $\widehat{\mu}^{BU}$. $\pi^{BU}(\mu,\sigma)$ は, σ^2 を与えたときの μ の条件付き分布が $P[\mu=d\sigma|\sigma]=P[\mu=-d\sigma|\sigma]=1/2$ で与えられ, σ^2 の密度が $\sigma^{-1}\mathrm{d}\sigma^2$ とする. このとき, $\widehat{\mu}^{BU}$ は次で与えられる.

$$\widehat{\mu}^{BU}=d\frac{\int_0^\infty[e^{-(X-d\sigma)^2/(2\sigma^2)}-e^{-(X+d\sigma)^2/(2\sigma^2)}]\sigma^{-(m+3)}e^{-S/(2\sigma^2)}\mathrm{d}\sigma^2}{\int_0^\infty[e^{-(X-d\sigma)^2/(2\sigma^2)}+e^{-(X+d\sigma)^2/(2\sigma^2)}]\sigma^{-(m+4)}e^{-S/(2\sigma^2)}\mathrm{d}\sigma^2}$$

この場合は, X の改良は $d\le 1$ のとき成り立つ.

[3] 打ち切り推定量 $\widehat{\mu}^{TR}$. $e_m=\Gamma((m+1)/2)/\{\sqrt{2}\Gamma((m+2)/2)\}$ に対して, 次の形で与えられる.

$$\begin{aligned}\widehat{\mu}^{TR}&=\begin{cases}\min\{X,\ de_m\sqrt{S+X^2}\}, & \text{if } X>0,\\ \max\{X,\ -de_m\sqrt{S+X^2}\}, & \text{if } X<0,\end{cases}\\ &=X-\sqrt{S}\frac{X}{|X|}\max\left(\frac{|X|}{\sqrt{S}}-de_m\sqrt{1+\frac{X^2}{S}},\ 0\right)\end{aligned}$$

また, 制約空間 B_d^* のもとでは, X はつねに縮小推定量 $\widehat{\mu}^{SR}=d^2(1+d^2)^{-1}X$ により改良され, このリスクの上限は X のリスクより小さい. このことは, X が任意の d に対してミニマックスでも許容的でもないことを意味している.

5 正に制約された平均の和の推定

この節では, 各位置母数が正に制約されているときにその線形結合の推定問題を考える. 正規分布の平均に関して制約された母数空間上の一様分布に対する一般化ベイズ推定量を考えると, 定理 2.1, 命題 3.2 が同時推定の枠組みでミニマックスになることを示しているのに対して, 平均の和の推定を考えたときには必ずしもミニマックスにならないことを示してみよう. これは, Kubokawa (2012), Kubokawa, Strawderman (2011a, b) の一連の論文の中で与えられており, その内容を解説する.

5.1 一般化ベイズ推定量の非ミニマックス性

$X_1,\ldots,X_p$ を互いに独立な確率変数とし, 各 X_i の密度関数が $f_i(x_i-\mu_i)$ で与えられ, 位置母数が正に制約されている, すなわち $\mu_i>0$, $i=1,\ldots,p$ とする. $\boldsymbol{X}=(X_1,\ldots,X_p)^t$, $\boldsymbol{x}=(x_1,\ldots,x_p)^t$, $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_p)^t$, $f(\boldsymbol{x}-\boldsymbol{\mu})=\prod_{i=1}^p f_i(x_i-\mu_i)$ とし, $E[X_i^2]<\infty$ とする. このとき, μ_i の線形結合 $\theta=\sum_{i=1}^p a_i\mu_i=\boldsymbol{a}^t\boldsymbol{\mu}$ を推定する問題を考える. ここで, a_i は実数であり, $\boldsymbol{a}=(a_1,\ldots,a_p)^t$ とする. $c_i=E[X_i-\mu_i]$ とおくと $\widehat{\mu}_i^U=X_i-c_i$ が μ_i の不偏推定量なので, $\hat{\theta}^U=\sum_{i=1}^p a_i\widehat{\mu}_i^U$ が θ の不偏推定量になり, しかもミニマックスになっている. 制約された空間を $D=\{(\mu_1,\ldots,\mu_p)^t|\mu_i>0\}$ とおくと, D 上の一様分布に対する一般化ベイズ推定量は

$$\hat{\theta}^{GB}=\boldsymbol{a}^t\widehat{\boldsymbol{\mu}}^{GB}=\int_D\boldsymbol{a}^t\boldsymbol{\xi}f(\boldsymbol{X}-\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}/\int_D f(\boldsymbol{X}-\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}$$

で与えられる.

$\hat{\theta}^{GB}$ のミニマックス性を調べるために, まず $p=2$ の場合をみてみよう. $\hat{\theta}^{GB}$ が $\widehat{\mu}_i(\phi_i)=X_i-\phi_i(X_i)$ に対して $\hat{\theta}(\phi_1,\phi_2)=a_1\widehat{\mu}_1(\phi_1)+a_2\widehat{\mu}_2(\phi_2)$ なる推定量のクラスに入ることは (2.2) からわかる. このとき, 次の命題が成り立つ.

命題 5.1 $i=1,2$ に対して, $\phi_i(w)$ が次をみたすとする.
(a) $\phi_i(w)$ は非減少で, $\lim_{w\to\infty}\phi_i(w)=c_i$
(b) $\phi_i(w)\geq\phi_i^{GB}(w)=\int_{-\infty}^{w}uf_i(u)\mathrm{d}u/\int_{-\infty}^{w}f_i(u)\mathrm{d}u$
このとき, $a_1a_2\leq 0$ ならば, $\hat{\theta}(\phi_1,\phi_2)$ はミニマックスである.

証明. $\hat{\theta}^U$ がミニマックスになるので, $\hat{\theta}^U$ と $\hat{\theta}(\phi_1,\phi_2)$ のリスクの差を調べてみる. そこで,

$$
\begin{aligned}
R(\mu_1,&\mu_2,\hat{\theta}^U)-R(\mu_1,\mu_2,\hat{\theta}(\phi_1,\phi_2))\\
=&a_1^2\{R(\mu_1,\widehat{\mu}_1^U)-R(\mu_1,\widehat{\mu}_1(\phi_1))\}+a_2^2\{R(\mu_2,\widehat{\mu}_2^U)-R(\mu_2,\widehat{\mu}_2(\phi_2))\}\\
&-2a_1a_2E[\widehat{\mu}_1(\phi_1)-\mu_1]E[\widehat{\mu}_2(\phi_2)-\mu_2].
\end{aligned}
$$

となり, 定理 2.2 より, $R(\mu_i,\widehat{\mu}_i^U)\geq R(\mu_i,\widehat{\mu}_i(\phi_i))$, $i=1,2$, であることがわかる. また $E[\widehat{\mu}_i(\phi_i)-\mu_i]=E[X_i-\mu_i-\phi_i(X_i)]=c_i-E[\phi_i(X_i)]$ となり, 条件 (a) より $E[\widehat{\mu}_i(\phi_i)-\mu_i]\geq 0$ となるので, 命題が成り立つ. ■

特に, 一般化ベイズ推定量に対しては, 命題 2.1 より, $\mu_i=0$ で $R(0,\widehat{\mu}_i^U)=R(0,\widehat{\mu}_i^{GB})$ となるので,

$$
R(0,0,\hat{\theta}^U)-R(0,0,\hat{\theta}^{GB})=-2a_1a_2E_0[\widehat{\mu}_1^{GB}]E_0[\widehat{\mu}_2^{GB}]
$$

と書ける. 従って, 次の命題を得る.

命題 5.2 一般化ベイズ推定量 $\hat{\theta}^{GB}=a_1\widehat{\mu}_1^{GB}+a_2\widehat{\mu}_2^{GB}$ がミニマックスになる必要十分条件は $a_1a_2\leq 0$ である.

一般の p の場合には, $\Lambda_+=\{i|a_i>0\}$, $\Lambda_-=\{i|a_i<0\}$ に対して $\theta_+=\sum_{i\in\Lambda_+}a_i\mu_i$, $\theta_-=-\sum_{i\in\Lambda_-}a_i\mu_i$ とおくとき, θ は $\theta=\theta_+-\theta_-$ と書けることに注意して, 命題 5.2 を用いると, 次の結果を得る.

命題 5.3 一般化ベイズ推定量 $\hat{\theta}^{GB}$ がミニマックスになる必要十分条件は $p=1$ もしくは $p=2$ で $a_1a_2\leq 0$ である. 従って, $p\geq 3$ のとき, $\hat{\theta}^{GB}$ はミニマックスでない.

5.2 最尤推定量の非ミニマックス性

θ の推定として, 打ち切り推定量 $\widehat{\mu}_i^{TR}=\max\{X_i-c_i,0\}$ の線形結合 $\hat{\theta}^{TR}=\sum_{i=1}^{p}a_i\widehat{\mu}_i^{TR}$ を取り上げてみると, ミニマックス性の必要十分条件はどのような形で与えられるのだろうか. 特に, 正規分布のときには, $\hat{\theta}^{TR}$ は θ の最尤推定量になっているので, ミニマックス性が維持されているか否かを調べることは興味深い. $F_i(z)=\int_{-\infty}^{z}f_i(u)\mathrm{d}u$ に対して $B_i(\mu_i)=\int_{-\infty}^{c_i-\mu_i}(z-c_i+\mu_i)f_i(z)\mathrm{d}z=\int_{-\infty}^{c_i-\mu_i}F_i(z)\mathrm{d}z$, $L_i(\mu_i)=-\int_{-\infty}^{c_i-\mu_i}(z-c_i)F_i(z)\mathrm{d}z$ とおく.

命題 5.4 打ち切り推定量 $\hat{\theta}^{TR}$ がミニマックスになる必要十分条件は, すべての部分集合 $C_+\subset\Lambda_+$, $C_-\subset\Lambda_-$ に対して,

$$
\begin{aligned}
&\sum_{i\in C_+}\sum_{\ell\in C_+,\ell\neq i}a_ia_\ell B_i(0)B_\ell(0)\leq 2\sum_{i\in C_+}a_i^2L_i(0),\\
&\sum_{j\in C_-}\sum_{\ell\in C_-,\ell\neq j}a_ja_\ell B_j(0)B_\ell(0)\leq 2\sum_{j\in C_-}a_j^2L_j(0)
\end{aligned}
\tag{5.1}
$$

が成り立つことである.

特に, $f_1(z) = \cdots = f_p(z) = f(z)$ の場合には, $K_f = 2L(0)/\{B(0)\}^2 + 1$ とおくと, 条件 (5.1) は,

$$\left(\sum_{i\in C_+} a_i\right)^2 \Big/ \sum_{i\in C_+} a_i^2 \le K_f, \qquad \left(\sum_{i\in C_-} a_i\right)^2 \Big/ \sum_{i\in C_-} a_i^2 \le K_f \tag{5.2}$$

と書き直すことができる. さらに, $|a_1| = \cdots = |a_p| = a$ となる場合を考えると, 条件 (5.2) は次のように簡単になる.

命題 5.5 $f_1(z) = \cdots = f_p(z) = f(z)$, $|a_1| = \cdots = |a_p| = a$ の場合を考える. このとき, 打ち切り推定量 $\hat{\theta}^{TR}$ がミニマックスになる必要十分条件は, $\max\{p_+,\ p_-\} \le K_f$ となる. ただし, p_+, p_- はそれぞれ Λ_+, Λ_- の個数であり, $p_+ + p_- = p$ をみたす.

正規分布のときを考えると, 標準正規分布の密度関数 $\phi(\cdot)$, 累積分布関数 $\Phi(\cdot)$ を用いて $B(\cdot)$, $L(\cdot)$ は $B(\mu_i) = \phi(\mu_i) - \mu_i\Phi(-\mu_i)$, $L(\mu_i) = \{\Phi(-\mu_i) + \mu_i B(\mu_i)\}/2$ と表される. $B(0) = 1/\sqrt{2\pi}$, $L(0) = 1/4$ より, $K_f = 2L(0)/\{B(0)\}^2 + 1 = \pi + 1$ となる.

命題 5.6 $\theta = \sum_{i=1}^p \mu_i$ の最尤推定量 $\hat{\theta}^{ML}$ がミニマックスになる必要十分条件は $p \le \pi + 1$ となる. すなわち, $p \le 4$ のとき $\hat{\theta}^{ML}$ はミニマックスであるが, $p \ge 5$ のときミニマックスでない.

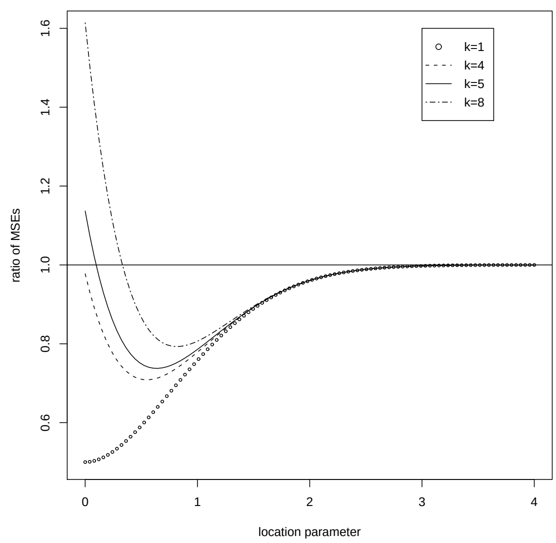

図 1: 最尤推定量と不偏推定量のリスクの比 $R(\mu, \hat{\theta}^{ML})/R(\mu, \hat{\theta}^{U})$ $(p = 1, 4, 5, 8,\ 0 < \mu < 4)$

$\mu_1 = \cdots = \mu_p = \mu$ の場合には, 最尤推定量のリスクは

$$R(\mu, \hat{\theta}^{ML}) = p - p\Phi(-\mu) - p\mu B(\mu) + p(p-1)\{B(\mu)\}^2$$

と表されるが, この値を $p = 1, 4, 5, 8,\ 0 < \mu < 4$ に対して図示したのが 図 1 である. この図から, 最尤推定量は $p = 1, 4$ のときミニマックスになるものの $p = 5, 8$ ではミニマックス・リスクの値を超えてしまうことが確かめられる. 命題 5.6 は, Shinozaki, Chang (1999) によって示された. Fernandez, Rueda, Salvador (2000) は他の分布への拡張を行っている. 正規分布以外にも様々な分布に対して K_f の値を求めることができ, ミニマックス性と非ミニマックス性の条件が与えられる. 代表的な分布については, 表 1 でまとめられている.

表 1: K_f の値と打ち切り推定量のミニマックス性・非ミニマックス性の条件

分布	K_f	ミニマックス性	非ミニマックス性
$\mathcal{N}(0,1)$	$\pi+1$	$p\le 4$	$p\ge 5$
$\mathcal{T}_3$	$\pi^2/2+1$	$p\le 5$	$p\ge 6$
両側指数分布	5	$p\le 5$	$p\ge 6$
ロジスティック分布	$\pi^2/\{6[\log(2)]^2\}+1$	$p\le 4$	$p\ge 5$
$\mathcal{U}(-1,1)$	11/3	$p\le 3$	$p\ge 4$
指数分布	$(e-1)^2$	$p\le 2$	$p\ge 3$

5.3 ミニマックス推定と許容的ミニマックス推定

これまでは, θ の一般化ベイズ推定量や最尤推定量が p が大きいときミニマックスにならないことを示してきた. ここでは, すべての p に対してミニマックス性が成り立つような打ち切り推定量と許容的な推定量を求めてみよう.

まず, ミニマックスな打ち切り推定量を求めるために, $\hat\theta_+^U=\sum_{i\in\Lambda_+}a_i\widehat{\mu}_i^U$, $\hat\theta_-^U=\sum_{i\in\Lambda_-}a_i\widehat{\mu}_i^U$ に対して, $\hat\theta_+^{TR*}=\max\{\hat\theta_+^U,0\}$, $\hat\theta_-^{TR*}=\max\{\hat\theta_-^U,0\}$ とおき,

$$\hat\theta^{TR*}=\hat\theta_+^{TR*}-\hat\theta_-^{TR*}$$

なる打ち切り推定量を考える. このとき, $\hat\theta^{TR*}$ のリスクは, $R(\boldsymbol{\mu},\hat\theta^{TR*})=E[(\hat\theta_+^{TR*}-\theta_+)^2]+E[(\hat\theta_-^{TR*}-\theta_-)^2]-2E[\hat\theta_+^{TR*}-\theta_+]E[\hat\theta_-^{TR*}-\theta_-]$ と書かれる. ここで, $\theta_+>0$, $\theta_->0$ より, $E[(\hat\theta_+^{TR*}-\theta_+)^2]\le E[(\hat\theta_+^U-\theta_+)^2]$, $E[(\hat\theta_-^{TR*}-\theta_-)^2]\le E[(\hat\theta_-^U-\theta_-)^2]$ であり, $E[\hat\theta_+^U]=\theta_+$ より

$$E[\hat\theta_+^{TR*}-\theta_+]=E[\hat\theta_+^{TR*}-\hat\theta_+^U]=E[\max\{\hat\theta_+^U,0\}-\hat\theta_+^U]=E[\max\{0,-\hat\theta_+^U\}]>0$$

となり, 同様にして $E[\hat\theta_-^{TR*}-\theta_-]>0$ となるので, 次の命題が成り立つ.

命題 5.7 打ち切り推定量 $\hat\theta^{TR*}$ はすべての $p\ge 1$ に対してミニマックスになる.

$\hat\theta^{TR*}$ はミニマックスになるものの, 一般化ベイズの形で表せないので非許容的である. そこで, 正規分布のもとで, 許容的でしかもミニマックスな推定量を求めてみよう. $i=1,\ldots,p$ に対して $X_i\sim\mathcal{N}(\mu_i,1)$, $\mu_i>0$, を仮定する. $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_p)^t$ の事前分布 $\pi^*(\boldsymbol{\mu})$ として, $i\in\Lambda_+$ に対しては確率 1 で $\mu_i=\alpha_i\xi_+$, $j\in\Lambda_-$ に対しては確率 1 で $\mu_j=\beta_j\xi_-$ とし, また ξ_+, ξ_- は 制約された空間 $\{(\xi_+,\xi_-)|\xi_+>0,\xi_->0\}$ 上を一様分布すると仮定する. ここで, $\alpha_i=a_i\sum_{j\in\Lambda_+}a_j/\sum_{j\in\Lambda_+}a_j^2$, $\beta_j=a_j\sum_{i\in\Lambda_-}a_i/\sum_{i\in\Lambda_-}a_i^2$ である. $A_2=\sum_{i\in\Lambda_+}\alpha_i^2$, $B_2=\sum_{i\in\Lambda_-}a_i^2$ とおき, $z_1=\hat\theta_+^U/\sqrt{A_2}$, $z_2=\hat\theta_-^U/\sqrt{B_2}$ とおく. このとき, 事前分布 π^* に対する θ の一般化ベイズ推定量は

$$\begin{aligned}\hat\theta^{GB*}=&\frac{\int_D(\sqrt{A_2}\xi_1-\sqrt{B_2}\xi_2)\exp\{-\|\boldsymbol{\xi}-\boldsymbol{z}\|^2/2\}\mathrm{d}\boldsymbol{\xi}}{\int_D\exp\{-\|\boldsymbol{\xi}-\boldsymbol{z}\|^2/2\}\mathrm{d}\boldsymbol{\xi}}\\ =&\sqrt{A_2}\{z_1-\phi^{GB*}(z_1)\}-\sqrt{B_2}\{z_2-\phi^{GB*}(z_2)\},\end{aligned}\tag{5.3}$$

と書かれる. ここで, $\phi^{GB*}(w)=\int_{-\infty}^w u\exp\{-u^2/2\}\mathrm{d}u/\int_{-\infty}^w\exp\{-u^2/2\}\mathrm{d}u$ である.

定理 5.1 事前分布 π^* に対する一般化ベイズ推定量 $\hat\theta^{GB*}$ は, すべての $p\ge 1$ に対してつねに許容的でしかもミニマックスになる.

5.4 分散の Stein 推定量との関係

最後に, 正規分布の平均が正に制限される問題が, 分散の Stein 推定と関係していること, 2 標本の正の平均の差の推定問題が分散の比の Stein 推定に関係していることを説明しよう.

通常の線形回帰モデル $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ を考える. ここで, $\boldsymbol{y}$ は n-次元の観測ベクトル, $\boldsymbol{X}$ は $n \times p$ 説明変数行列でランク p, $\boldsymbol{\beta}$ は p-次元係数ベクトル, 誤差項 $\boldsymbol{\epsilon}$ は $\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$ に従うとする. $\boldsymbol{\beta}$ の最小 2 乗推定量は $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$ であり, 残差平方和を $S = (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^t(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$ とおく. $m = n - p$ とおくと, σ^2 の不偏推定量は $\hat{\sigma}^{2U} = S/m$ であり, σ^2 の Stein 推定量は

$$\hat{\sigma}^{2S} = \min\{S/m,\ (S + \widehat{\boldsymbol{\beta}}^t\boldsymbol{X}^t\boldsymbol{X}\widehat{\boldsymbol{\beta}})/n\}$$

で与えられる. Stein (1964) の論法から, Stein 損失もしくはエントロピー損失のもとで $\hat{\sigma}^{2S}$ は $\hat{\sigma}^{2U}$ を一様に改良していることが示される. Rukhin (1992) は, この結果が, 正規分布の正の平均の推定において不偏推定量が最尤推定量により改良される結果と漸近的に同等であることを示した. 具体的には,

(B1) $p = n - d_n$, $d_n > 0$, $d_n = O(n^{\delta})$, $0 \le \delta < 1$

(B2) $\boldsymbol{X}^t\boldsymbol{X}/n$ が正定値行列に収束し, 次のような正の定数 θ が存在することを仮定する.

$$\lim_{n\to\infty} \sqrt{m}\boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta}/(n\sigma^2) = \sqrt{2}\theta$$

Y を $\mathcal{N}(\theta, 1)$ に従う確率変数とすると, これらの条件のもとで,

$$\begin{aligned}
&\sqrt{m}(\hat{\sigma}^{2U} - \sigma^2)/\sigma^2 \to -\sqrt{2}(Y - \theta),\\
&\sqrt{m}(\hat{\sigma}^{2S} - \sigma^2)/\sigma^2 \to -\sqrt{2}(\max\{Y, 0\} - \theta).
\end{aligned}$$

が示され, 正規分布の正の平均の推定問題に帰着される.

次に, 2 つの回帰モデル $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$, $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_i^2\boldsymbol{I}_n)$, $i = 1, 2$ を考えよう. 変数や $\widehat{\boldsymbol{\beta}}_i$, S_i を上述のものと同様に定義し, 分散の比 $\rho = \sigma_2^2/\sigma_1^2$ の推定問題を考える. σ_i^2 の不偏推定量を $\hat{\sigma}_i^{2U} = S_i/m$, Stein 推定量を $\hat{\sigma}_i^{2S} = \min\{\hat{\sigma}_i^{2U}, (S_i + \widehat{\boldsymbol{\beta}}_i^t\boldsymbol{X}_i^t\boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_i)/n\}$ とし, 比に対する 2 つの推定量を $\hat{\rho}^U = \hat{\sigma}_2^{2U}/\hat{\sigma}_1^{2U}$, $\hat{\rho}^S = \hat{\sigma}_2^{2S}/\hat{\sigma}_1^{2S}$ とするとき, Kubokawa (1994b), Kubokawa, Srivastava (1996) は $\hat{\rho}^S$ が $\hat{\rho}^U$ を一様に改良することを示した. 上記の (B1) と次を仮定する.

(B2′) $i = 1, 2$ に対して, $\boldsymbol{X}_i^t\boldsymbol{X}_i/n$ は正定値行列に収束し, 次をみたす正の定数 θ_i が存在する.

$$\lim_{n\to\infty} \sqrt{m}\boldsymbol{\beta}_i^t\boldsymbol{X}_i^t\boldsymbol{X}_i\boldsymbol{\beta}_i/(n\sigma_i^2) = \sqrt{2}\theta_i$$

このとき Y_i を $Y_i \sim \mathcal{N}(\theta_i, 1)$, $i = 1, 2$, なる確率変数とすると, 上述の条件のもとで

$$\begin{aligned}
&\sqrt{m}(\hat{\rho}^U - \rho) \to (\sigma_2^2/\sigma_1^2)\{(Y_1 - Y_2) - (\theta_1 - \theta_2)\},\\
&\sqrt{m}(\hat{\rho}^S - \rho) \to (\sigma_2^2/\sigma_1^2)\{(\max(Y_1, 0) - \max(Y_2, 0)) - (\theta_1 - \theta_2)\}
\end{aligned}$$

が示される. これは, 分散の比の Stein 推定が正規分布の正の平均の差の推定に漸近的に帰着されることを意味しており, 正の平均の差の推定を考える動機付けを与えている. 正の平均の和の推定は, 分散の積の Stein 推定に対応していると思われるが, 正の平均の和の推定については, 一般化ベイズ推定量は不偏推定量を改良できない一方で, 最尤推定量は改良できることを前節で示した. 分散の積の推定に関しては, Stein 型推定量の優越性は示せるのだろうかは興味深い問題である.

6 おわりに

この論文では, 位置母数分布族において位置母数に不等式制約が入っているときの推定問題を中心に, ミニマックス性や許容性, ベイズ的性質など統計的決定理論の視点から, これまでの研究成果を解説してきた.

位置母数以外にも不等式制約, 順序制約が想定される問題として, 分散に関する制約も考えられる. 例えば, 変量効果モデルを考えると, いくつかの分散の間に不等式が自然に組み込まれることになる. Tsukuma, Kubokawa (2011) は, こうした問題に対して改良するための方法論を展開している. 多変量の変量効果モデルを考えると, 共分散行列の間に不等式制約が入ることになるが, この問題については Srivastava, Kubokawa (1999), Kubokawa, Tsai (2006) で議論されている.

最近では, 予測分布の予測問題への拡張も取り上げられ, Fourdrinier, Marchand, Righi, Strawderman (2011) は, 分散が既知の多変量正規分布のもとでは, 点推定の一連の結果はすべて予測問題へ移し替えることができることを示した. 従って, この論文の中で与えられた平均の制約に関する結果は予測問題においても成り立つことになる. しかし, この結果は分散が未知の正規分布や他の分布へ適用することができない. Kubokawa, *et al*. (2012) は, 一般的な群が埋め込まれている予測問題において, 母数空間が制約されているときのミニマックス性の条件を与えている. また位置母数が制約されているときの一般化ベイズ予測分布のミニマックス性を示している.

2 節で解説したように, 制約された母数空間上に一様分布をもつ事前分布に対する一般化ベイズ推定量が最良共変推定量を改良するための方法として提示された Hartigan の方法と IERD 法は大変興味深い. Hartigan の方法の限界は分散が既知の正規分布の場合でしかも事前分布が一様分布の場合しか利用できないことにある. この限界を超える手法の確立は可能なのかは興味深い問題であり, 今後の研究に期待したい.

謝辞. 編集者及び査読者から貴重なコメントを頂きました. ここに深く感謝申し上げます. 本研究は, 科学研究費補助金 19200020, 21540114 から研究助成を受けております.

参考文献

[1] Barlow, R.E., Bartholomew, J.M., Bremner, J.M., and Brunk, H.D. (1972). *Statistical Inference under Order Restrictions, the Theory and Application of Isotonic Regression*, Wiley, New York.

[2] Berry, J.C. (1990). Minimax estimation of a bounded normal mean vector. *J. Multivariate Anal.*, **35**, 130-139.

[3] Bickel, P.J. (1981). Minimax estimation of the mean of normal distribution when the parameter space is restricted. *Ann. Statist.*, **9**, 1301-1309.

[4] Casella, G. and Strawderman, W.E. (1981). Estimating a bounded mean. *Ann. Statist.*, **9**, 870-879.

[5] Farrell, R.H. (1964). Estimators of a location parameter in the absolutely continuous case. *Ann. Math. Statist.*, **35**, 949-998.

[6] Fernandez, M.A., Rueda, C., and Salvador, B. (2000). Parameter estimation under orthant restrictions. *Canadian J. Statist.*, **28**, 171-181.

[7] Fourdrinier, D., Marchand, E., Righi, A., and Strawderman, W.E. (2011). On improved predictive density estimation with parametric constraints. *E. J. of Statist.*, **5**, 172-191.

[8] Girshick, M.A. and Savage, L.J. (1951). Bayes and minimax estimates for quadratic loss functions. In *Proc. Second Berkeley Symp. Math. Statist. Probab.*, **1**, 53-74. University of California Press, Berkeley.

[9] Hartigan, J.A. (2004). Uniform priors on convex sets improve risk. *Statist. Probab. Letters*, **67**, 285-288.

[10] Hora, R.B., and Buehler, R.J. (1966). Fiducial theory and invariant estimation. *Annals of Mathematical Statistics*, **37**, 643-656.

[11] Hwang, J.T., and Peddada, S.D. (1994). Confidence intervals subject to order restrictions. *Ann. Statist.*, **22**, 67-93.

[12] Iwasa, M. and Moritani, Y. (1997). A note on the admissibility of the maximum likelihood estimator for a bounded normal mean. *Statist. Probab. Letters*, **32**, 99-105.

[13] James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, **1**, 361-379. University of California Press, Berkeley.

[14] Katz, M. (1961). Admissible and minimax estimates of parameters in truncated spaces. *Ann. Math. Statist.*, **32**, 136-142..

[15] Kubokawa, T. (1994a). A unified approach to improving equivariant estimators. *Ann. Statist.*, **22**, 290-299.

[16] Kubokawa, T. (1994b). Double shrinkage estimation of ratio of scale parameters. *Ann. Inst. Statist. Math.*, **46**, 95-119.

[17] Kubokawa, T. (1998). The Stein phenomenon in simultaneous estimation: A review. In *Applied Statistical Science* III (S.E. Ahmed, M. Ahsanullah and B.K. Sinha, eds.), 143-173, NOVA Science Publishers, New York.

[18] Kubokawa, T. (1999). Shrinkage and modification techniques in estimation of variance and the related problems: A review. *Commun. Statist. - Theory and Methods*, **28**, 613-650.

[19] Kubokawa, T. (2004). Minimaxity in estimation of restricted parameters. *Journal of the Japanese Statistical Society*, **34**, 229-253.

[20] Kubokawa, T. (2005a). Estimation of a mean of a normal distribution with a bounded coefficient of variation. *Sankhyā*, **67**, 499-525.

[21] Kubokawa, T. (2005b). Estimation of bounded location and scale parameters. *Journal of the Japanese Statistical Society*, 35, 221-249.

[22] Kubokawa, T. (2012). Minimax estimation of linear combinations of restricted location parameters. *Contemporary Developments in Bayesian Analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman*, IMS Collections, Vol. 8, 24-41, Institute of Mathematical Statistics.

[23] Kubokawa, T., Marchand, T., Strawderman, W.E. and Turcotte, J.-P. (2012). Minimaxity in predictive density estimation with parametric constraints. Discussion Paper Series, CIRJE-F-843.

[24] Kubokawa, T. and Saleh, A.K.MD.E. (1998). Estimation of location and scale parameters under order restrictions. *J. Statistical Research*, **28**, 41-51.

[25] Kubokawa. T. and Srivastava, M.S. (1996). Double shrinkage estimators of ratio of variances. In: *Multidimensional Statistical Analysis and Theory of Random Matrices* (Editors: A.K. Gupta and V.L. Girko), 139-154, VSP, Netherlands. Proceedings of the Sixth Eugene Lukacs Symposium, Bowling Green.

[26] Kubokawa, T. and Strawderman, W. E. (2011a). Non-minimaxity of Linear Combinations of Restricted Location Estimators and Related Problems. *J. Statist. Plan. Inf.*, **141**, 2141-2155.

[27] Kubokawa, T. and Strawderman, W.E. (2011b). A unified approach to non-minimaxity of sets of linear combinations of restricted location estimators. *J. Multivariate Anal.*, **102**, 1429-1444.

[28] Kubokawa, T. and Tsai, M.-T. (2006). Estimation of covariance matrices in fixed and mixed effects linear models. *J. Multivariate Anal.*, **97**, 2242-2261.

[29] Lee, C.I.C. (1988). The quadratic loss of order restricted estimators for several treatment means and a control mean. *Ann. Statist.*, **16**, 751-758.

[30] Marchand, E. and Perron, F. (2001). Improving on the MLE of a bounded normal mean. *Ann. Statist.*, **29**, 1078-1093.

[31] Marchand, E. and Perron, F. (2005). Improving on the mle of a bounded location parameter for spherical distributions. *J. Multivariate Anal.*, **92**, 227-238.

[32] Marchand, E. and Strawderman, W.E. (2004). Estimation in restricted parameter space: A review. A Festshrift for Herman Rubin, Institute of Mathematical Statistics, Lecture Notes - Monograph Series, **45**, 21-44.

[33] Marchand, É., and Strawderman, W. E. (2005). On improving on the minimum risk equivariant estimator of a location parameter which is constrained to an interval or a half-interval. *Annals of the Institute of Statistical Mathematics*, **57**, 129-143.

[34] Marchand, É., and Strawderman, W.E. (2012). A unified minimax result for restricted parameter spaces. *Bernoulli*, **18**, 635-643.

[35] Maruyama, Y. and Iwasaki, K. (2005). Sensitivity of minimaxity and admissibility in the estimation of a positive normal mean. *Ann. Inst. Statist. Math.*, **57**, 145-156.

[36] Moors, J.J.A. (1981). Inadmissibility of linearly invariant estimators in the truncated parameter spaces. *J. American Statist. Soc.*, **76**, 910-915.

[37] Robertson, T., Wright, F.T., and Dykstra, R.L. (1988). *Order Restricted Statistical Inference*, Wiley, New York.

[38] Rukhin, A. (1992). Asymptotic risk behavior of mean vector and variance estimators and the problem of positive normal mean. *Ann. Inst. Statist. Math.*, **44**, 299-311.

[39] Shinozaki, N., and Chang, Y.-T. (1999). A comparison of maximum likelihood and best unbiased estimators in the estimation of linear combinations of positive normal means. *Statist. Decisions*, **17**, 125-136.

[40] Srivastava, M.S. and Kubokawa, T. (1999). Improved nonnegative estimation of multivariate components of variance. *Ann. Statist.*, **27**, 2008-2032.

[41] Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.*, **16**, 155-160.

[42] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, 1135–1151.

[43] Tsukuma, H. and Kubokawa, T. (2008). Stein phenomenon in estimation of means restricted to a polyhedral convex cone. *J. Multivariate Anal.*, **99**, 141-164.

[44] Tsukuma, H. and Kubokawa, T. (2011). Modifying estimators of ordered positive parameters under the Stein loss. *J. Multivariate Anal.*, **102**, 164-181.

[45] van Eeden, C. (2006). *Restricted parameter space problems - Admissibility and minimaxity properties.* Lecture Notes in Statistics, **188**, Springer.

Estimation in Restricted Parameter Spaces

Tatsuya Kubokawa

University of Tokyo

Abstract

In estimation of parameters in restricted spaces, several interesting and surprising results have been developed from a decision-theoretic point of view. For instance, in estimation of a normal mean with a known variance, the sample mean is minimax in the case that the mean is bounded from one side, but it is not minimax in the case of the mean bounded from both sides. Nevertheless, the sample mean becomes minimax even in the case of the mean bounded from both sides if the variance of the normal distribution is unknown. This surprising example inspires us to study more about estimation of the restricted parameter. In this paper, we review several topics in estimation of restricted parameters and explain the interesting results and phenomena derived in the literature from a decision-theoretic aspect.

Key words

Restricted parameter, minimaxity, Bayes estimation, generalized Bayes estimation, admissibility, statistical decision theory, Stein identity, unbiasedness, invariance, location-scale family