論文の内容の要旨

論文題目：Essays on Shrinkage Methods in Statistics

題目和訳：統計学における縮小化法に関する研究

氏名：加藤賢悟

—博士論文の要旨—

本論文は，近年統計学のさまざまな場面において注目を集めている縮小化法に関して，著者が博士課程において研究した成果をまとめたものである．本論文では，主に回帰係数の推定・予測分布の構成という 2 種類の問題に焦点を当てる．

まず，回帰係数の推定問題においては，縮小化法は一般に "損失関数+(チューニングパラメータ)×(ペナルティ)" の最小化問題の形で定式化されることが多い．"損失関数+(チューニングパラメータ)×(ペナルティ)" の最小化問題の形で定式化される縮小化法は正則化法とも呼ばれる．'縮小' 化法と呼ぶ理由は，無制約の推定値に比べて何らかの意味でより小さい推定値を返すからである．縮小化法を用いる利点として，回帰係数をいくらか縮小することよってモデルの予測精度の向上が期待できる点と，特定のペナルティを用いることで推定と変数選択を同時に実行できる点があげられる．

Tibshirani (1996) により提案された Lasso は，線形回帰モデルの係数を推定する手法であり，その推定値は 2 乗損失関数に ℓ_1 ペナルティを課した評価関数の最小化解で定義される．このとき，ℓ_1 ペナルティの性質から，Lasso はいくつかの係数を正確にゼロに縮小する．すなわち，Lasso は推定と変数選択を同時に実行できるのである．また，Lasso に対しては解のパスを計算する効率的なアルゴリズムが知られており，計算量の観点からも，Lasso が従来の変数選択法と比べて魅力的な手法であるといえることも指摘しておく．近年，特に生物データなどは，共変量が数が従来のデータと比較して極めて大きくなる傾向にある．従って，Lasso の持つ計算上の利点は非常に重要な特徴であるといえる．

Lasso が提案されて以降，用途に応じた様々なペナルティが多くの研究者によって提案されている．また，2 乗損失以外の損失関数に対して ℓ_1 ペナルティなどを課す試みも多く見受けられる．このような縮小化法を実際に実行する際，以下の課題があると思われる：

- チューニングパラメータの選択.
- 解のパスの計算.
- 実データに対するパフォーマンス.

本論文の第2章から第4章はそれぞれ以上の課題に焦点を当てている．まず第2章では，凸ペナルティ付き最小2乗法に対して，自由度 (degrees of freedom) と呼ばれる量の不偏推定量を微分幾何的なアプローチを用いて導出した．導出された推定量を用いて，C_p 規準やAICといったチューニングパラメータの選択規準が導出される．Lassoに対しては，Zou et al. (2007) が ($p \leq n$ の下で) 自由度の不偏推定量を導出しており，本研究は彼らの結果を拡張したものとみなすこともできる．ただし，本研究はより一般的な制約に対して，自由度の不偏推定量を (原則として計算可能な) 制約集合に関する幾何的な量を用いて表しており，Zou et al. (2007) のアプローチとはかなり異なるものであることを指摘しておく．また，共変量の数が標本数よりも多いケース (すなわち，$p > n$ のケース) において，Lassoの自由度の不偏推定値を明示的に計算する方法を示した．第3章は区分的に線形な損失関数と ℓ_1 ペナルティまたはブロック ℓ_∞ ペナルティを持つ正則化問題の計算上の側面を扱っている．区分的に線形な損失関数としては，分位点回帰において用いられる "check" 関数や，SVMにおいて用いられる "hinge" 関数などがあげられる．第3章ではパラメトリック改訂単体法を紹介し，同方法がこのタイプの正則化問題に対して解のパスを計算する安定的かつ効率的な手法を与えることを示す．主要な貢献として，問題の特性を活かした具体的な計算上の工夫を提案した点があげられる．例えば ℓ_1 罰則付き分位点回帰に対しては，Li and Zhu (2008) が正則化パスを計算するアルゴリズムを提案してるが，彼らのアルゴリズムは正則化パスの一意性といった強い仮定に依存している．パラメトリック単体法はデータに対する仮定を必要としない点で彼らのアルゴリズムより有利であるといえる．また，シミュレーション実験を通じて，正則化パスの変化点の個数が標本数・共変量の数を変えたときにどのように変化するか調べた．第4章では分位点回帰と ℓ_1 罰則付き分位点回帰を自動車保険のデータに対して適用した．データ解析の結果から，ℓ_1 罰則付き分位点回帰が保険請求額のリスク要因の特定化に役立つことがわかった．また，第4章ではデータ解析に加えて，分位点回帰と ℓ_1 罰則付き分位点回帰の理論的な性質をまとめた．

縮小化法は予測分布の構成においても有効な手法である．ここで，d 次元の多変量正規分布 $N_d(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_d)$ の予測問題を考えよう．損失関数としては，Kullback-Leibler divergenceを用いることにする．分散既知の場合，一様分布 $\pi_U(\boldsymbol{\mu}) \propto 1$ に基づくベイズ予測分布 $\widehat{p}_U$ が定数リスクを持つミニマクスな予測分布となる．Komaki (2001) は縮小型事前分布 $\pi_S(\boldsymbol{\mu}) = \|\boldsymbol{\mu}\|^{-(d-2)}$ に基づくベイズ予測分布が $d \geq 3$ のとき $\widehat{p}_U$ を改良することを示した．'縮小' と呼ぶ理由は一様分布に比べて π_S がより原点方向に重みをおいているからである．Komaki (2001) の結果はいわゆる "スタイン現象" が分布予測においても現れることを示しており，興味深い．

第5章では平均と分散が未知の多変量正規分布に対する予測問題を扱う．Komaki (2001) の結果以降，多変量正規分布の予測問題を扱った研究はいずれも分散が既知であることを仮定している．しかしながら，通常分散は未知であることから，平均･分散がともに未知の多変量正規分布に対してよりよい予測分布を導くことは意義のあることであるといえる．分散が未知であるとき，右不変な事前分布 $\pi_R(\boldsymbol{\mu}, \sigma) = 1/\sigma$ に基づくベイズ予測分布 $\widehat{p}_R$ が定数リスクを持つミニマクスな予測分布となる．本研究では縮小型事前分布 $\pi_{LT}(\boldsymbol{\mu}, \sigma) = \|\boldsymbol{\mu}\|^{-(d-2)}/\sigma$ に基づくベイズ予測分布が $d \geq 3$ のとき $\widehat{p}_R$ を改良することを示した．なお，証明に用いた確率順序のテクニックは点推定の文脈においても見られなかった手法であり，このような新しいテクニックを導入した点も副次的な貢献と言える．