スパース構造の下でのミニマックス予測分布

東京大・情報理工　金子 亮也
東京大・情報理工　駒木 文保

1　問題設定

X を観測量，Y を予測量として次の形のモデルを考える：

$$X \sim \mathcal{N}_n(\theta, \Sigma), \quad Y \sim \mathcal{N}_n(\theta, \tilde{\Sigma}), \quad \theta \in \{\theta \in \mathbb{R}^n : ||\theta||_0 \leq s_n\}. \tag{1}$$

ただし，$||\cdot||_0$ は非ゼロ要素の数を表す．本発表では，モデル (1) の下で X に基づく Y の予測分布 $\hat{p}(y \mid x)$ を構成する問題を考える．モデル (1) ではパラメータ空間がスパース制約によって制限されており，その具体的な応用例としては，DNA に内包される遺伝情報や複数地点での宇宙線観測データなど，非常に大きな次元を持つが意味のある情報は疎らにしか存在しない，という特徴を持つデータを念頭に置いている．

2　関連研究

Mukherjee and Johnstone [2] は，モデル (1) において $\Sigma = I_n, \tilde{\Sigma} = rI_n$ とする単純な設定の下で，Kullback-Leibler 損失に基づくリスクを考え，$s_n/n \to 0$ の極限で漸近的にミニマックス性を達成する予測分布を構成した．導出した漸近的ミニマックスリスクには観測量と予測量の分散比 r に依存する定数係数が陽に表れており，予測分布の推測問題の難しさと分散比との関係を定量的に表した結果と言える．また，結果の導出の際には，観測量と予測量の分散比が次元に依らず一定であることを用いて，1 次元の予測に帰着させて議論を行っており，その際に 1 次元 Bayes 予測分布のリスクを，対応する Bayes 推定量の 2 乗リスクによる積分表現 [1] を通じて推定リスクの評価へと帰着させている．

3　本研究

本発表では，[2] を含んだより一般的な設定を考える．具体的には，(1) において $\Sigma, \tilde{\Sigma}$ を互いに異なる正定値対角行列としたときに，漸近的にミニマックスな予測分布を構成することを考える．[2] においては上述の通り 1 次元の予測に帰着させて議論を行っていたが，本研究の設定では分散比が次元毎に異なるため同様の議論が適用できず，対応する結果が得られるか否かは自明ではない．そこで [1] の積分表現の結果を多次元正規分布において分散比が次元毎に異なる場合に拡張した等式を示す．この等式を経由して漸近的ミニマックスリスクを導出し，併せて漸近的にミニマックスな予測分布を具体的に構成する．導出した漸近的ミニマックスリスクには，[2] の結果に対応して，次元毎の分散比に依存した定数係数が陽に表れている．最後に数値例を紹介して，構成した予測分布の性能を議論する．

参考文献

[1] Brown, L. D., George, E. I., & Xu, X. (2008). Admissible predictive density estimation. *The Annals of Statistics, 36*(3), 1156-1170.

[2] Mukherjee, G., & Johnstone, I. M. (2015). Exact minimax estimation of the predictive density in sparse Gaussian models. *The Annals of Statistics, 43*(3), 937-961.