共変量の正規性の下での lasso の情報理論的リスク上界

九大・シス情 川喜田雅則， 九大・シス情 竹内純一

1．背景　過去の研究 [1] により，Kraft の不等式を満たす罰則関数を用いた罰則付き尤度は，用いた確率モデルに関するユニーバサル二段階語頭符号の符号長として解釈できて，罰則付き最尤推定量 $\hat{\theta}$ はいわゆる最小記述長 (MDL) 推定量と見なせることが知られている．さらに Rényi ダイバージェンスを損失関数としたときのリスクは上記の二段階符号の冗長度によって上から押さえられることが示された．しかしこの理論を用いるにはパラメータ空間を量子化する必要があり，厳密に罰則付き最尤推定のリスク評価をしたとは言えなかった．近年 Barron ら [2, 3] は**リスク妥当性**という概念を用いて量子化を介さず固定計画における lasso のリスク上界

$$E\left[\frac{\lambda}{2\sigma^2}\|X(\hat{\theta}-\theta^*)\|_2^2\right] \le E\left[\inf_{\theta\in\Theta}\left\{\frac{(\|y-X\theta^*\|_2^2-\|y-X\theta\|_2^2)}{2\sigma^2}+\sqrt{\frac{2n\log(4p)}{\sigma^2}}\|\theta\|_{w,1}+\frac{\log 2}{(1-\lambda)}\right\}\right]$$

を導出した．しかし固定計画では教師付き学習の汎化誤差を評価したことにはならない．彼らの理論を無作為計画の場合に拡張するには様々な困難に直面する．そのためか，これまで教師付き学習で無作為計画の場合のこの理論の研究はほとんどなく，[4] が著者の知る限り唯一の例である．

2．研究成果　本研究の目標は無作為計画の設定で lasso の情報理論的リスク上界を導くことである．無作為計画への拡張における最も困難な問題は次の二つである．一つは損失として用いる条件付き Rényi ダイバージェンスは通常の Rényi ダイバージェンスとは異なりパラメータの二次関数にはならず，複雑な形になることである．そのため [3] の技法を直接適用できない．この問題を一般の共変量の分布に対して解決することは難しいため，まず最初のステップとして共変量 x^n が平均 0 の正規分布に従うとして議論する．それでも条件付き Rényi ダイバージェンスはまだ複雑な形をしているが，テイラー展開を用いることで [3] のアイデアを適用可能な形にして解決した．しかしもう一つの困難はより深刻である． [2] の理論により量子化を回避するためには上記のリスク妥当性を満たす罰則関数を導出する必要があるが，今の設定ではリスク妥当性を満たす罰則関数は必ず何らかの形で共変量に依存する．しかし罰則関数が共変量に依存すると上記理論の主要な補題の一つが成立しなくなるため，リスク上界が導けないというジレンマに陥る．この問題は我々の設定に特別な問題ではなく教師付き学習を考える際に共通に現れる問題である．我々はこの問題を，x^n の範囲を都合の良い高確率集合（典型集合）$\mathscr{X}(\epsilon)$ に制約し，[3] の定理 II.3 の証明に工夫を加えることで解決した．その結果無作為計画の場合の lasso のリスク上界

$$E_1\Big[\bar{L}^\lambda(p_*, p_{\hat{\theta}(x^n,y^n)})\Big] \le E_1\left[\inf_{\theta}\left\{\log\frac{p_*(y^n|x^n)}{p_\theta(y^n|x^n)}+\widehat{V}(\theta)\right\}\right]-\frac{\log P_B}{1-\lambda}$$

の導出に成功した．ここで $\bar{L}^\lambda$ は条件付き Rényi ダイバージェンスを，E_1 は条件付き分布 $p_*(x^n, y^n|x^n \in \mathscr{X}(\epsilon))$ に関する期待値を指し，P_B は共変量が典型集合上に値を取る確率を指す．また $\widehat{V}$ は ℓ_1 ノルムに基づくリスク妥当な罰則関数を表す．さらに我々は PAC 学習タイプのリグレット上界も導出した．この PAC 学習タイプの上界については数値実験によって，その振る舞いを観察する．

参考文献

[1] A. R. Barron and T. M. Cover. Minimum complexity density estimation. *IEEE Transactions on Information Theory*, 37(4):1034–1054, 1991.

[2] A. R. Barron, C. Huang, J. Li, and X. Luo. MDL, penalized likelihood and statistical risk. In *Proceedings of IEEE Information Theory Workshop*, Porto, Portugal, May 4-9 2008.

[3] S. Chatterjee and A. Barron. Information theory of penalized likelihoods and its statistical implications. *arXiv'1401.6714v2 [math.ST] 27 Apr.*, 2014.

[4] K. Yamanishi. A learning criterion for stochastic rules. *Machine Learning*, 9(2-3):165–203, 1992.