尤度関数の一般化とモデル選択手法の性質

東京大学 大学院情報理工学系研究科　倉田 澄人

AIC や TIC に代表される情報量規準は, 主要項として対数尤度関数 (の負倍) を持つ. これらは, モデルがデータを発生させている真の確率分布からどの程度乖離しているかを KL divergence に基づいて評価しているものと捉えられ, ダイバージェンスを Taylor 展開等を用いて漸近的に見積もることで導出される.

一方, 主要項としては AIC と同じ形を持つベイズ情報量規準 BIC はその名の通りベイズ統計学の観点から導出された規準であり, こちらは周辺尤度関数

$$m_0(\boldsymbol{Y}) = \int f(\boldsymbol{Y} \,|\, \boldsymbol{\theta})\, \pi(\boldsymbol{\theta})\ d\boldsymbol{\theta}$$

の近似と看做せる. ここで, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ はそれぞれ独立に確率分布 $G_1, \ldots, G_n$ に従う確率変数 (同分布である必要は無い) とし, $f(\boldsymbol{Y} \,|\, \boldsymbol{\theta})$ は $\boldsymbol{Y}$ の同時分布, そして $\boldsymbol{\theta}$ は $\pi(\boldsymbol{\theta})$ を事前分布に持つ未知母数とする.

今回はこの「尤度」を, 基になるダイバージェンスに沿って拡張したモデル評価規準の性質を報告する. 具体的には周辺尤度の代わりに, より一般化された形

$$m_0(\boldsymbol{Y}) = \int \exp\left\{-n\, h_n(\boldsymbol{Y}\,;\boldsymbol{\theta})\right\}\, \pi(\boldsymbol{\theta})\ d\boldsymbol{\theta}$$

(一般化周辺尤度) を導入した場合について論じる. ここから導出されるモデル評価規準が最良と判ずるモデルは, 最小化すべき量 h_n を成すダイバージェンスに基づいた (疑似的) 事後確率を最大化するモデルと解釈出来る.

本研究では, 特に BIC の特長である「選択の一致性」と, それとは別の概念である「選択の頑健性」を中心に置いて考察する. ここで言う「頑健」とは, データを発生させている分布に, 真の分布とは異なる分布が混合した際の変動の少なさを意味し, 影響関数の観点を拡張しモデル評価へと導入したものである. 適当なダイバージェンスに基づいた一般化周辺尤度を用いた場合, BIC を自然に拡張した規準族が導出され, 大標本時の選択一致性を有しつつ, BIC 等と比較して異常な分布の混入 (これは一点分布に限定されない) に対して値が変動し難い頑健な選択規準が得られる. 本発表では以上の内容を理論的検討並びに数値実験例によって示す.

参考文献

[1] Ghosh, A. and Basu, A. (2016). Robust Bayes estimation using the density power divergence, *Annals of the Institute of Statistical Mathematics*, **68** (2), 413–437.

[2] Kass, R. E., Tierney, L., and Kadane, J. B. (1990). The validity of posterior expansions based on Laplace's method, *Bayesian and likelihood methods in statistics and econometrics*, **7**, 473–488.

[3] Kurata, S. and Hamada, E. (2019). On the consistency and the robustness in model selection criteria, *Communications in Statistics-Theory and Methods* (in press).