

## モデル選択の不確実性を考慮した Bayesian Model Averaging

【背景】疾患関連マーカーの探索及び診断・予測モデルの開発は、様々な病気の早期発見や治療、予防のためにも重要な課題である。通常、予測モデルの構築に際しては、結果変数の予測に有用な説明変数をステップワイズ法などにより選択し、予測能力が最も良いと考えられるモデルを一つ選択する。以後、選択されたモデルが正しいモデルという前提の下で推測が進められる。しかし、このような方法ではモデル選択における不確実性が考慮されていないため、信頼区間を狭く推定してしまい、 $\alpha$ エラーが名目水準以下に保たれなくなる。

Bayesian Model Averaging (BMA)は、モデル選択の不確実性を推測・予測に考慮する一つの方法である。BMA は、観測データ $\mathbf{D}$ に対し、あるモデル $M_i$ の下での推定値 $\Delta$ の事後分布 $p(\Delta|M_i, \mathbf{D})$ をそのモデルの事後確率 $p(M_i|\mathbf{D})$ で重みづけし、考慮するすべてのモデルについて平均化した事後分布を用いて推測を行う。

【目的】BMA の方法論について整理し、従来の変数選択法と比較することでその性能について検討する

【方法】対数スコア関数を用いて予測性能を比較した場合、BMA はどの単一のモデルに基づく予測性能よりも優れていることが理論的に示されている。この性質は、特に予測性能の高いいくつかのモデルだけに関して平均化した予測分布に関するシミュレーション実験で示されており、被覆確率もほぼ名目上の水準が担保されることが確認されている。さらに、ステップワイズ方法よりも真の関連がある変数を選択する傾向にあり、 $p$  値に基づく変数選択は $\alpha$ エラーの名目水準を大きく上回る一方で、BMA によるある変数をモデルに含める事後確率は直感的な解釈とほとんど一致することが示されている。

【今後の展望】疾患関連マーカーの探索及び診断・予測モデルの開発といった観点からその応用可能性について検討したい。

### 参考文献

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian Averaging Models. *Statistical Science*, 14(4), 382–417.

Viallefont, V., Raftery, A. E., & Richardson, S. (2001). Variable selection and Bayesian model averaging in case-control studies. *Statistics in Medicine*, 20(21), 3215–3230. <https://doi.org/10.1002/sim.976>