

# 混合正規分布の均一性検定におけるベイズ検定統計量の漸近挙動

東京工業大学 情報理工学院 仮屋 夏樹、渡辺 澄夫

所与のデータが複数の分布の混合からなる確率分布から得られたか、あるいは単一の確率分布から得られたかを判定する均一性の検定は、基礎・応用の双方の観点から重要な問題である。

この問題の重要な特徴として、混合分布は最尤推定量の漸近正規性が成り立つための条件を満たさない特異モデルであり、そのため尤度比検定の検定統計量がカイ二乗分布に法則収束しないことが挙げられる [1]。

したがって、モデルが正則モデルであることを仮定して構築されてきた従来の仮説検定の手法は、特異モデルを用いた均一性の検定に適用することは出来ない。そのため、混合分布による均一性の検定のための新たな手段として、現在までに尤度比検定に補正を加えた修正尤度比検定の提案 [2] や、シミュレーションにより帰無仮説によるデータの当てはまりと対立仮説によるデータの当てはまりを評価する D 検定 [3] などの提案がなされてきた。

本発表ではこれらとは異なるアプローチとしてベイズ仮説検定の枠組みに基づき、確率モデルとして混合正規分布を用いた場合の均一性検定を対象に、検定統計量の従う確率分布の挙動の議論を行う。

ベイズ仮説検定の枠組みでは、帰無仮説及び対立仮説は確率モデルを表すパラメータの事前分布として与えられる。検定統計量は周辺尤度の比で与えられ、所与のデータを  $\{X_n\}$ 、パラメータ  $w$  により定まる確率モデルを  $p(X|w)$ 、帰無仮説、対立仮説に相当する事前分布をそれぞれ  $\varphi_0(w), \varphi_1(w)$  とすれば

$$L(X^n) = \frac{\int \varphi_1(w) \prod_i p(X_i|w) dw}{\int \varphi_0(w) \prod_i p(X_i|w) dw}$$

と表される。ベイズ仮説検定ではこの周辺尤度比を検定統計量とした仮説検定が最強力検定を与えることが知られている [4]。

したがって、特異モデルを用いたベイズ仮説検定の構築には周辺尤度比の挙動を明らかにすることが重要であり、過去には時系列解析を題材とした研究などがある [5]。しかし、均一性検定においてベイズ仮説検定の検定統計量である周辺尤度比の挙動について十分な議論はされてこなかった。

今回我々はクラスタ数  $K = 2$  の混合正規分布を確率モデルとした均一性の検定を対象に、周辺尤度比の評価を解析的・数値的に行った。特に確率モデルの特異性が周辺尤度比の挙動に影響を及ぼす領域の解析を行うため、パラメータのデータ数に関するスケール則を適用して漸近挙動の評価を行った。

解析の結果モデルの特異性を反映しカイ二乗分布と異なる周辺尤度比の漸近挙動が得られた。発表では解析結果と棄却域・検出力の挙動、対立仮説のハイパーパラメータが棄却域・検出力に及ぼす影響を報告する。

## 参考文献

- [1] J.A.Hartigan “A Failure of likelihood asymptotics for normal mixtures” Proc. of the. Berkeley Conference in Honor of J.Neyman and J.Kiefer, Vol.2, 807-810 (1985)
- [2] H. Chen, J. Chen, J.D. Kalbfleisch, ”A modified likelihood ratio test for homogeneity in finite mixture models” J. R. Statist. Soc. B, 63, 19-29 (2001)
- [3] R. Charnigo, J. Sun ”Testing Homogeneity in a Mixture Distribution via the L2 Distance Between Competing Models”, Journal of the American Statistical Association, 99:466, 488-498 (2004)
- [4] 渡辺澄夫「ベイズ統計の理論と方法」コロナ社, 2012
- [5] 藤原香織、渡辺澄夫「特異モデルにおけるベイズ検定と時系列解析への応用」電子情報通信学会誌 D Vol.J91-D No.4 pp. 889-896(2008)