

ロジスティック回帰モデルにおける $\log F$ 分布の紹介

医学・疫学分野では、2値アウトカムと説明変数の関係性を推定する際にロジスティック回帰モデルがしばしば用いられる。しかし、サンプルサイズが小さい場合やイベントが稀なとき、スパースデータバイアスが発生することが問題に挙げられている [1]。この状況下では、最尤法による推定値にバイアスが入り、正しく推定されない、または推定値が存在しない場合がある [2]。

そのような状況の対処法の1つとして、 $\log F$ 分布による罰則化を行うという手法が提案されている [3]。 $\log F$ 分布は、 F 分布に対数をとった分布であり、順序統計量と繋がりがあることが知られている [4]。Greenland らは、 $\log F$ 分布による罰則項を課した罰則付き対数尤度関数を考えることでロジスティック回帰モデルにおける推定値のバイアスを取り除き、パラメータ推定の改善を行う手法を提案し、その有用性を実データへの適用を通じて確認した [3]。

本抄読会では、ロジスティック回帰モデルへの $\log F$ 分布の適用方法を確認した上で、実データへの適用例を紹介する [3]。

参考文献

- [1] Greenland S, Mansournia M.A, and Altman D.G. Sparse data bias: a problem hiding in plain sight. *BMJ*, 352:i1981, 2016.
- [2] Albert A and Anderson J.A. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71(1):1–10, 1984.
- [3] Greenland S and Mansournia M.A. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Statistics in Medicine*, 34(23):3133–3143, 2015.
- [4] Jones M.C. Families of distributions arising from distributions of order statistics. *Test*, 13(1):1–43, 2004.