

Tweedie GLM を用いた Queensland 州の降水量データ解析

統計数理研・データ科学 大西 俊郎

University of Southern Queensland, Faculty of Sciences Dr Peter Dunn

1. Introduction

近年史上最悪の干ばつに見舞われているオーストラリア・クイーンズランド州の降水量データを一般化線形モデルに基づき解析する．オーストラリア東部の降水量は SOI (Southern Oscillation Index) によって説明できるとされている (Stone and Auliciems, 1992)．SOI とはタヒチとダーウィンの 2 地点間の気圧の差を指標化したものである．本発表ではクイーンズランド州東南部の月別降水量と SOI の間の関係を調べる．

2. Tweedie 分布

当該データにはゼロを多く含むという特徴がある．このようなデータを取り扱うのに便利なのが Tweedie 分布と呼ばれる分布である．これはべき乗の分散関数 $v(\mu) = \mu^\xi/\tau$ をもつ指数型分布族 (τ は拡散パラメータ) である．べき指数 ξ が 1 より大きく 2 未満のとき，複合ポアソン分布 (ガンマ分布のポアソン・ミクスチャ) となり，ゼロをとる確率が正になることが知られている．Tweedie 分布の密度関数は複雑な関数を含んでいるが，十分な精度の数値的評価が可能なプログラムがある (Dunn and Smyth, 2007)．以下，密度関数を $p(x; \mu, \tau_0)$ と書く．

3. Tweedie 一般化線形モデル

第 k 観測地点の第 i 番目の降水量 x_{ki} について，次のような対数リンク回帰モデルを仮定する ($1 \leq k \leq K, 1 \leq i \leq n_k$)．

$$x_{ki} \sim p(x_{ki}; \mu_{ki}, \tau) \quad \text{with} \quad \log \mu_{ki} = \alpha_k + \beta \text{SOI}_{ki} + (\text{季節因子}) \quad (1)$$

切片パラメータ α_k は観測地点に依存する因子を表し，SOI による影響 (傾きパラメータ β) は一定としている．

4. 共役解析

Tweedie 一般化線形モデル (1) を Bayes の枠組みで扱う．このモデルは切片パラメータ α_k について共役事前分布が存在するという理論的に優れた性質を持つ．尤度関数が α_k に関して並進対称性を持ち，location family に属するある分布が共役事前分布となる．この分布は location family に共役解析可能性を要請すると得られるものである (Ohnishi and Yanagimoto, 2007)．

5. データ解析

経験 Bayes 推定を行う．具体的には周辺尤度を最大化するようなパラメータを求める．詳細については当日発表する予定である．

参考文献

- DUNN, P. K., AND SMYTH, G. K. (2007). Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Statistics and Computing*, To appear.
- OHNISHI, T. AND YANAGIMOTO, T. (2007). Conjugate location–dispersion families. *Journal of the Japanese Statistical Society*, To appear.
- STONE, R. C. AND AULICIEMS, A (1992). SOI phase relationships with rainfall in eastern Australia. *International Journal of Climatology*, **12**, 625–636.