ポアソン回帰モデルに対する回帰相関係数推定量の統計的性質

東京理科大学　黒沢 健
東京理科大学大学院　鈴木 進洋

与えられたデータに対して候補となる複数の統計モデルの中から, ある基準を用いて最も良い統計モデルを選択する必要がある. この統計モデルを選択する際に用いられる基準を **モデル評価測度** という. 本研究では, Generalized Linear Model (GLM) で用いられるモデル評価測度のうち, Zheng and Agresti [1] によって提案された**回帰相関係数** (RCC:Regression Correlation Coefficient)

$$RCC = cor(Y, E(Y|\boldsymbol{X}))$$

に注目する. RCC は確率変数ではなく母数であり, この推定量 (標本値) がモデル評価測度となる. RCC の推定量は, 重相関係数 R の一般化であり, 相関係数で定義されるため, 様々な GLM に対して適用できる. GLM は連結関数 φ を用いて

$$E(Y|\boldsymbol{X}) = \varphi^{-1}(\alpha + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}) \tag{1}$$

と書け, この $E(Y|\boldsymbol{X})$ と Y の相関によって定義される. GLM は非確率変数 X_i と関連付けられた Y_i を用いて (1) の $E(Y|\boldsymbol{X})$ を $E(Y_i)$ と書くことも多い. 本研究では (1) のように $\boldsymbol{X}$ は確率変数ベクトルとして扱う.

$\mathcal{P}(\lambda)$ をパラメータ λ を持つポアソン分布とする. 確率変数 $Y|\boldsymbol{X}$ が確率変数ベクトル $\boldsymbol{X}$ とパラメータ $\alpha, \boldsymbol{\beta}$ について

$$Y|\boldsymbol{X} \sim \mathcal{P}(\lambda), \quad E(Y|\boldsymbol{X}) = \exp(\alpha + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}) \tag{2}$$

となる時, 確率変数 $Y|\boldsymbol{X}$ はパラメータ λ のポアソン回帰モデルに従うという. このポアソン回帰モデルに対して, [2] では, RCC の陽な形を導出している.

定理 1 ([2]) *Σ を正定値行列, 確率変数 $Y|\boldsymbol{X}$ をポアソン回帰モデル (2) に従っている確率変数とする. 確率変数ベクトル $\boldsymbol{X}$ に対して $\boldsymbol{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$ を仮定すると, RCC は*

$$RCC(Y, \boldsymbol{X}; \alpha, \boldsymbol{\beta}) = \sqrt{\frac{E(Y)(\exp(\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}) - 1)}{1 + E(Y)(\exp(\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}) - 1)}}.$$

ただし $E(Y) = \exp\left(\alpha + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}\right)$ で与えられる.

更に, [2] では定理 1 で求めた RCC に $\alpha, \boldsymbol{\beta}$ の最尤推定量 $\hat{\alpha}, \hat{\boldsymbol{\beta}}$ を代入することで, RCC の新しい推定量 $\hat{\mathrm{R}}$ を提案している. 本研究ではこの $\hat{\mathrm{R}}$ の漸近的な性質について以下の定理を得た.

定理 2 *$Y|\boldsymbol{X}$ を (2) で与えられるパラメータ λ のポアソン回帰モデルに従う確率変数とし, 確率変数ベクトル $\boldsymbol{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$ とする. ただし, $\boldsymbol{\beta} \neq \boldsymbol{0}, \Sigma$ は正定値行列とする. ポアソン回帰モデルからのランダム標本のサンプルサイズを N とすると, RCC の推定量 $\hat{R}$ は*

$$\sqrt{N}(\hat{R} - RCC) \xrightarrow{d} \mathcal{N}(0, V_{\hat{R}}(\boldsymbol{\gamma})).$$

ただし,

$$V_{\hat{R}}(\boldsymbol{\gamma}) = (1 - RCC^2)^3 P(t), \qquad t = \exp(\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}), \quad P(t) = \frac{(1 + 4\log t)t^2 - 2t + 1}{4(t-1)}.$$

参考文献

[1] B. Zheng and A. Agresti. Summarizing the predictive power of a generalized linear model. *Statistics in Medicine*, Vol. 19, pp. 1771–1781, 2000.

[2] 高橋明史, 黒沢健. ポアソン回帰モデルにおける回帰相関係数. 第 27 回日本計算機統計学会シンポジウム, pp. 65–68, Nov. 2013.