

# 経時測定カウントデータの解析 (OD モデルを用いた対数線型変化係数の推定)

富田哲治

広島大学 原爆放射線医科学研究所 計量生物研究分野

## 1. はじめに

経時測定データとは、個体毎に繰り返し測定されたデータに観測時点の情報が得られているデータのことである．同一個体に対して繰り返し測定しているため、測定値間に正の相関が認められる．そのため、解析時には同一個体内の測定値間の相関構造を考慮した解析方法を用いる必要がある．身長や体重といった連続変量の経時測定データに対する解析法は、多変量正規分布の理論を用いて重回帰分析を拡張した手法として多く提案されている一方、罹患数や死亡数といった離散変量の経時測定データに対する解析法は、扱いやすい多変量離散分布がないため工夫が必要である．本研究では、時系列データ解析における離散値データの解析法を、経時測定カウントデータの解析に応用することを考え、その研究結果について実データの解析を交えて報告する．

## 2. 統計モデル

$y_{it}$  を個体  $i$  の時刻  $t$  における観測値とする ( $i = 1, \dots, N; t = 1, \dots, T$ ). 過去の観測値の履歴  $\mathcal{F}_{i,t-1} = \{y_{is} | s \leq t-1\}$  が与えられたとき、 $y_{it}$  が以下のようなポアソン分布に従っているとする、

$$y_{it} | \mathcal{F}_{i,t-1} \sim \text{Poisson}(\mu_{it}), \quad \mu_{it} = e^{W_{it}}, \quad W_{it} = \mathbf{a}_i' \boldsymbol{\beta}(t) + Z_{it}.$$

ここで、 $\mathbf{a}_i = (a_{i1}, \dots, a_{ik})'$  は時間に依存しない個体間説明変数、 $\boldsymbol{\beta}(t) = (\beta_1(t), \dots, \beta_k(t))'$  は時間と共に変化する回帰係数を表わす変化係数、 $Z_{it}$  は次のような全ての過去の観測値の加重和で表わされているとする、

$$Z_{it} = \sum_{\ell=1}^{\infty} \pi_{\ell} e_{it-\ell}, \quad e_{it} = \frac{y_{it} - \mu_{it}}{\sqrt{\mu_{it}}}.$$

このような無限次数の移動平均過程  $\text{MA}(\infty)$  は、有限個のパラメータで表現することができ、ここでは、以下のような  $\text{ARMA}(p, q)$  過程を用いる、

$$Z_{it} = \sum_{\ell=1}^p \phi_{\ell} (Z_{it-\ell} + e_{it-\ell}) + \sum_{\ell=1}^q \theta_{\ell} e_{it-\ell}.$$

## 3. 未知パラメータの推定

未知パラメータ  $\boldsymbol{\beta}(t) = (\beta_1(t), \dots, \beta_k(t))'$ ,  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)'$ ,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$  の推定は、Davis et al. (2003, Biometrika) と Satoh and Yanagihara (2009, AJMMS) の手法を組み合わせることで可能である．今、変化係数に線形構造、 $\boldsymbol{\beta}(t) = \Xi \mathbf{x}(t)$ , を仮定する．ここで、 $\Xi = (\xi_1, \dots, \xi_k)'$  は  $k \times r$  の未知パラメータ行列、 $\mathbf{x}(t)$  は時間に関する基底を表わす  $r \times 1$  ベクトルとする、例えば、基底に  $r-1$  次の多項式を用いた場合は  $\mathbf{x}(t) = (1, t, \dots, t^{r-1})'$  となる．このとき、 $\boldsymbol{\beta}(t)$  の推定は  $\Xi$  の推定と同値となる． $\mathbf{x}_{it} = \mathbf{a}_i \otimes \mathbf{x}(t)$ ,  $\boldsymbol{\xi} = \text{vec}(\Xi')$  とすると、 $W_{it}$  は  $\boldsymbol{\xi}$  を用いて  $W_{it} = \mathbf{x}_{it}' \boldsymbol{\xi} + Z_{it}$  と表せる．これは、Davis et al. (2003) の提案した GLARMA モデルと同じ形である．したがって、Davis et al. (2003) と同様な推定方法 (最尤法) で  $\boldsymbol{\delta} = (\boldsymbol{\xi}', \boldsymbol{\phi}', \boldsymbol{\theta}')$  を推定できる． $\ell_{it}(\boldsymbol{\delta}) = \log f(y_{it} | \mathcal{F}_{i,t-1})$ ,  $f(\cdot)$  は  $y_{it} | \mathcal{F}_{i,t-1}$  の密度関数とすると、対数尤度は  $\sum_{i=1}^N \sum_{j=1}^T \ell_{it}(y_{it} | \mathcal{F}_{i,t-1})$  で表わされ、定数項を無視したものを  $\ell(\boldsymbol{\delta})$  とすると、 $\ell(\boldsymbol{\delta})$  は次のように書ける、

$$\ell(\boldsymbol{\delta}) = \sum_{i=1}^N \sum_{j=1}^T \left( y_{it} W_{it}(\boldsymbol{\delta}) - e^{W_{it}(\boldsymbol{\delta})} \right), \quad W_{it}(\boldsymbol{\delta}) = \mathbf{x}_{it}' \boldsymbol{\xi} + Z_{it}(\boldsymbol{\delta}).$$

$\ell(\delta)$  の最大化には，Newton-Raphson 法などの数値最適化を適用する．最尤推定量  $\hat{\delta}$  は漸近正規性をもち，

$$\hat{\delta} \rightsquigarrow N_{rk+p+q}(\delta, \Omega), \quad \hat{\Omega} = - \left( \frac{\partial^2 \ell(\delta)}{\partial \delta \partial \delta'} \Big|_{\delta=\hat{\delta}} \right)^{-1},$$

が成り立つ．これより，変化係数  $\beta(t) = (\beta_1(t), \dots, \beta_k(t))'$  の推定量とその漸近分布は，

$$\hat{\beta}_j(t) = \hat{\xi}_j' x(t) \rightsquigarrow N(\beta_j(t), \lambda_j(t)), \quad j = 1, \dots, k,$$

で得られる．ただし， $\lambda_j(t) = x(t)' \Omega_{jj} x(t)$ ， $\Omega_{jj}$  は  $\hat{\xi}_j$  の共分散行列に対応する  $\Omega$  の部分行列である． $\Omega_{jj}$  の推定値を  $\hat{\Omega}_{jj}$ ， $\hat{\lambda}_j(t) = x(t)' \hat{\Omega}_{jj} x(t)$  とすると，Sato and Yanagihara (2009) の結果から，

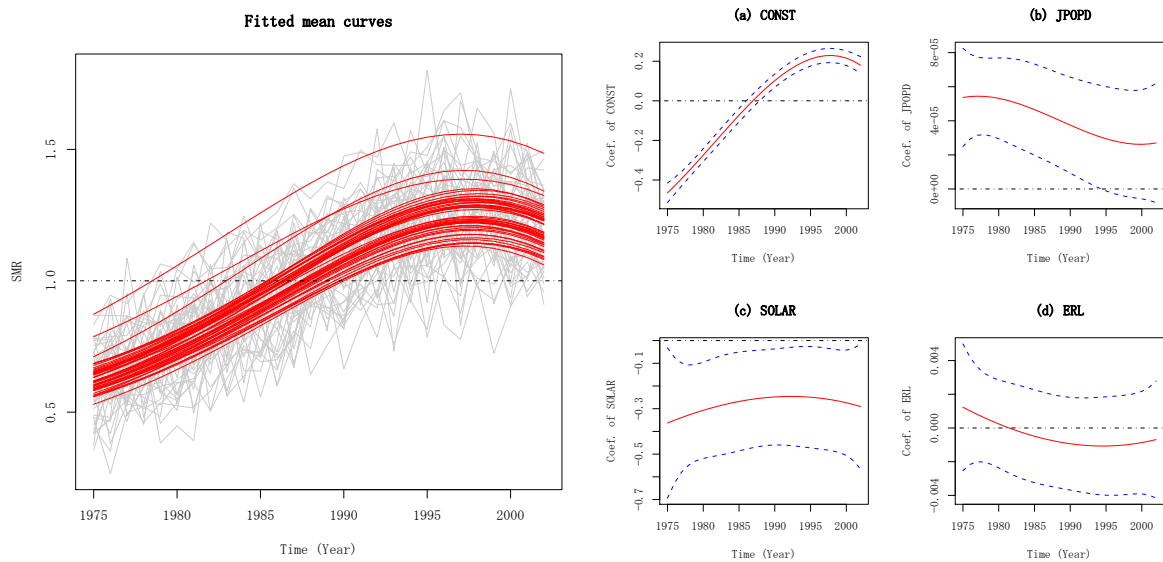
$$\sup_{t \in R} \frac{\{\hat{\beta}_j(t) - \beta_j(t)\}^2}{\hat{\lambda}_j(t)} \leq \sup_{x \in R^q} \frac{\{(\hat{\xi}_j - \xi_j)' x\}^2}{x' \hat{\Omega}_{jj} x} \leq (\hat{\xi}_j - \xi_j)' \hat{\Omega}_{jj}^{-1} (\hat{\xi}_j - \xi_j) \rightsquigarrow \chi_q^2.$$

これより， $c_{q,\alpha}$  を  $\chi_q^2$  の上側  $100\alpha\%$  点とすると，変化係数  $\beta_j(t)$  の  $100(1-\alpha)\%$  信頼区間が次式で得られる．

$$\left[ \hat{\beta}_j(t) - \sqrt{\hat{\lambda}_j(t) c_{q,\alpha}}, \quad \hat{\beta}_j(t) + \sqrt{\hat{\lambda}_j(t) c_{q,\alpha}} \right].$$

#### 4. 適用例

実データへの適用例として，1975 年から 2002 年における日本男性の大腸がんの都道府県別死亡数の経時測定データ（出典：祖父江・片野田，国立がんセンター）を提案手法を用いて解析する．説明変数に，都会度の指標として 1985 年の人口密度 [人/km<sup>2</sup>]，Mizoue (2004, Health Phys.) が消化器系がんの死亡危険度との関連性を報告した日照量 [kwh/m<sup>2</sup>/day]，Abe et al. (1981, J. Nucl. Sci. Technol.) で報告された自然空間放射線 [nGy/h] を用いた．時間に関する基底に 3 次多項式， $Z_{it}$  の記述に ARMA(1, 1) 過程を用いて推定された，死亡危険度の推定曲線，各説明変数の変化係数の推定曲線および 95 % 信頼区間を下図に示す．信頼区間を含めて，人口密度の変化係数曲線が正値を，日照量の変化係数曲線が負値になっていることから，都市部および日照量の少ない地域で死亡危険度が高くなる傾向があることが示唆される．



【左図】提案手法で推定された死亡危険度の曲線と観測値を繋いだ折れ線，【右図 (a)】ベースライン（説明変数の値が全て平均値）の推定曲線，【右図 (b)】人口密度の変化係数の推定曲線，【右図 (c)】日照量の変化係数の推定曲線，【右図 (d)】自然空間放射線の変化係数の推定曲線，ただし，点線は 95 % 信頼区間．