経時離散データに対するベースラインを特定しない変化係数の推測について

県立広島大・経営情報学部 冨田哲治，札幌医科大・医療人育成センター 加茂憲一， 広島大・原医研 佐藤健一

1．はじめに： 経時データに対する回帰モデルにおいて，時間 t とともに変化する共変量の効果 $\boldsymbol{\beta}(t)$ は変化係数 (Hastie and Tibshirani, 1993, *J. Roy. Statist. Soc. Ser. B*) と呼ばれる．Satoh and Tonda (2016, *Amer. J. Math. Management Sci.*) は連続値データに対して，ベースラインの形状を特定しない変化係数の推測法を提案した．本稿では，計数データに対して同様な変化係数の推測法を提案する．

2．統計モデル： 個体 i $(i = 1, \ldots, N)$ について，時点 t で観察された計数値を $d_i(t)$ とする $(t = 1, \ldots, T)$. 過去の観測値の履歴 $\mathcal{F}_{i,t-1} = \{d_i(s) | s \le t-1\}$ が与えられたとき，$d_i(t)$ の条件付き分布が平均 $\mu_i(t)$ のポアソン分布で記述されているとする．

$$d_i(t) | \mathcal{F}_{i,t-1} \sim \mathrm{Poisson}\left(\mu_i(t)\right), \quad \mu_i(t) = O_i(t)\lambda(t)e^{\boldsymbol{a}_i' \boldsymbol{\beta}(t) + Z_i(t)} \tag{1}$$

ただし，$O_i(t)$ はオフセット，$\lambda(t)$ はベースラインの経時変化，$\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_k(t))'$ は共変量 $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{ik})'$ に対する変化係数である．一方，$Z_i(t)$ は観測値の観測時点間の相関をモデル化するための変数である．本研究では，尤度関数が陽に得られ扱いやすいことから，観測値駆動型モデルを用いて $Z_i(t)$ を記述する．$Z_i(t)$ の構造の違いによって様々な観測値駆動型モデルが提案されている．ここでは，次式で表される GLARMA 構造 (Dunsmuir and Scott, 2015, *J. Stat. Softw.*) を用いた．

$$Z_i(t) = \sum_{\ell=1}^{p} \phi_\ell \left\{Z_i(t-\ell) + e_i(t-\ell)\right\} + \sum_{\ell=1}^{q} \theta_\ell e_i(t-\ell), \quad e_i(t) = \frac{d_i(t) - \mu_i(t)}{\sqrt{\mu_i(t)}}. \tag{2}$$

3．変化係数の推定法： 変化係数の関数形に $\boldsymbol{\beta}(t) = \Xi \boldsymbol{x}(t)$ を仮定する，ただし，$\Xi = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k)'$ は $k \times r$ の未知母数行列，$\boldsymbol{x}(t) = (x_1(t), \ldots, x_r(t))'$ は $\boldsymbol{\beta}(t)$ の関数形を規定する既知の基底関数である．次の対数尤度関数を用いて，ベースラインの経時変化 $\lambda(t)$ の関数形を特定せず，変化係数に関する未知母数 Ξ と系列相関に関する未知母数 $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$ を推定する．

$$\log L\left(\Xi, \boldsymbol{\phi}, \boldsymbol{\theta}\right) = \sum_{t=1}^{T} \left\{ \sum_{i=1}^{N} d_i(t) W_i(t) - n_t \log \left(\sum_{i=1}^{N} O_i(t) e^{W_i(t)} \right) \right\} + \mathrm{Const.} \tag{3}$$

ただし，$W_i(t) = \boldsymbol{a}_i' \boldsymbol{\beta}(t) + Z_i(t)$, $n_t = \sum_{i=1}^{N} d_i(t)$, $Z_i(t)$ は式 (2) で定義される．

4．適用例： 人口動態統計に基づき都道府県別に集計された 1975 年から 2012 年の 38 年間における男性の大腸がん死亡数データに対する適用例について紹介する．全期間の全国平均が基準集団の期待死亡数をオフセットとし，共変量は人口密度 [persons/km^2] と日照量 [KWh/m^2/day]，変化係数の関数形に 2 次曲線を仮定した．提案法を適用し，ベースラインの関数形を特定することなく人口密度と日照量の効果を推定した結果が図 1 と図 2 である．図 3 は変化係数の推定値を用いて事後的に推定したベースラインの経時変化である．

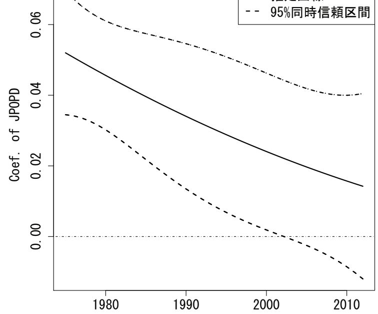

図 1．人口密度の効果

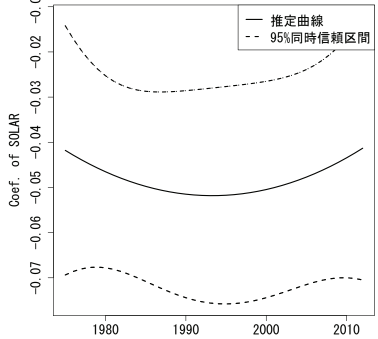

図 2．日照量の効果

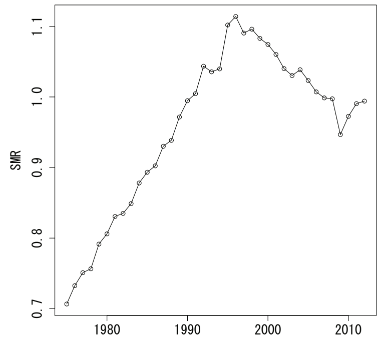

図 3．ベースライン