

変化係数を用いたがん死亡危険度の 年次変動要因の推測

富田 哲治¹・佐藤 健一¹・中山 晃志²・片野田 耕太³・
祖父江 友孝³・大瀧 慈¹

(受付 2010 年 12 月 24 日; 改訂 2011 年 3 月 15 日; 採択 4 月 28 日)

要 旨

がんなどの疾患の罹患・死亡リスクは、様々な背景要因の変化に伴って経年的に変動している。このような背景要因と経年変動の関係を解析することは、我が国のがんリスクの動向を把握し、今後のがん対策へ繋げていくために基礎となる重要なテーマである。がん死亡数は地域ごとに毎年集計されており、このような繰り返し観測されたデータは特に経時データと呼ばれている。本論文では、観測値駆動型モデルを用いて経時データの特性を考慮しつつ変化係数により説明変数の経時的な変化を推測するための解析方法を提案する。実データへの適用例として、日本における男性の大腸がん死亡数データに提案手法を適用し解析を行った。

キーワード：観測値駆動型モデル、一般化線形 ARMA、変化係数、一様性検定。

1. はじめに

がんなどの疾患の罹患・死亡リスクは、様々な背景要因の変化に伴って経年的に変動している。このような背景要因と経年変動の関係を解析することは、我が国のがんリスクの動向を把握し、今後のがん対策へ繋げていくために基礎となる重要なテーマである。がん死亡数は地域ごとに毎年集計され、その一部はウェブ上で一般公開もされている。このような、繰り返し観測されるデータは特に経時データと呼ばれており、経時データを扱う場合には観測時点間の相関を考慮することが重要である。また、経時データに対する回帰モデルにおいて、変化係数は時間とともに変化する背景要因の影響を評価でき、がん死亡危険度の年次変動に関わる背景要因の影響を定量的に評価し、視覚的に理解するのに有用である。

がん死亡数といった離散値の経時データの解析には、一般化線形モデル(Dobson, 1990)にランダム変数を導入することで時点間相関をモデル化した一般化線形混合モデル(例えば, Breslow and Clayton, 1993)がよく使われている。しかし、尤度関数が陽に表せないため未知母数推定に必要な計算量が膨大となり、観測時点や説明変数の数が比較的多い場合には適用が困難である(Chan and Ledolter, 1995; Durbin and Koopman, 2000; Jung and Liesenfeld, 2001)。観測値駆動型モデル(例えば, Zeger and Qaqish, 1988)は、ランダム変数の代わりに過去の観測値を用いたモデルであり、推移モデルとも呼ばれている(Diggle et al., 1994)。観測値駆動型モデルは、

¹ 広島大学 原爆放射線医科学研究所：〒734-8551 広島県広島市南区霞 1-2-3

² 国際医療福祉大学 福岡看護学部：〒810-0072 福岡県福岡市中央区長浜 1-3-1

³ 国立がんセンター がん対策情報センター：〒104-0045 東京都中央区築地 5-1-1

尤度関数が陽に得られ扱いやすいという利点があり、特に、Davis et al. (2003)は、自己回帰移動平均(AutoRegressive Moving Average, ARMA)に基づく観測値駆動型モデルを提案し、推定量の性質などについて報告している。

変化係数は、Hastie and Tibshirani (1993)によって提案され、時間とともに変化する説明変数の効果を視覚的に評価できるため、経時測定データの解析に極めて有用である。変化係数の推定法は、固定した時間の近傍データを用いて回帰を繰り返すカーネル平滑化法(例えば、Hoover et al., 1998; Tonda et al., 2011)が一般的だが、この方法では変化係数の同時信頼区間の構築や検定が困難であった。そこで、Sato and Yanagihara (2010)は変化係数の形状に線形構造を導入することで、連続値の経時データに対する変化係数の推測法を提案した。また、佐藤 他(2009)はLiang and Zeger (1986)による一般化推定方程式を用いた離散値の経時データにおける変化係数の推測法を提案し、富田 他(2010)は空間データに対して位置とともに変化する回帰係数(変化係数曲面)の推測法を提案した。

本稿では、観測値駆動型モデルに変化係数を導入し、時点間相関を考慮しつつ、がん死亡数危険度の年次変動要因を変化係数として視覚的に評価するための解析法の開発を試みる。変化係数を用いて、がんの危険・予防要因となる因子の効果の経時的な変化を捉えることは、がん対策の立案において有用であると考えられる。2節で、観測値駆動型モデルと変化係数を組み合わせることでモデルを記述し、3節および4節で推測法を提案する。5節において、日本における男性の大腸がん死亡数データを用いた解析例を紹介する。

2. 統計モデル

地域 i の t 年における観測死亡数を d_{it} とする ($i = 1, \dots, N; t = 1, \dots, T$)。過去の死亡数の観測履歴 $\mathcal{F}_{i,t-1} = \{d_{is} | s \leq t-1\}$ が与えられたとき、 d_{it} の条件付き分布は平均 μ_{it} のポアソン分布に従うとする、

$$(2.1) \quad d_{it} | \mathcal{F}_{i,t-1} \sim \text{Poisson}(\mu_{it}).$$

また、平均母数 μ_{it} は、 k 個の説明変数 $\mathbf{a}_i = (a_{i1}, \dots, a_{ik})'$ を用いて、

$$(2.2) \quad \mu_{it} = e^{W_{it}}, \quad W_{it} = \mathbf{a}_i' \boldsymbol{\beta}(t) + Z_{it},$$

で記述されているとする。ここで、 $\boldsymbol{\beta}(t)$ は時間とともに変化する説明変数の効果を表す回帰係数であり、変化係数(Hastie and Tibshirani, 1993)と呼ばれている。一方、 Z_{it} 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、過去の観測値の関数として Z_{it} を記述する。 Z_{it} 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おり、Davis et al. (1999)で詳しくレビューされている。ここでは、 Z_{it} は過去全ての観測値の加重和で表されるものとする。つまり、

$$(2.3) \quad Z_{it} = \sum_{\ell=1}^{\infty} \pi_{\ell} e_{it-\ell}, \quad e_{it} = \frac{d_{it} - \mu_{it}}{\sqrt{\mu_{it}}}.$$

ここで、式(2.3)は無限次元の移動平均(Moving Average, MA)と呼ばれ、加重和の重みである π_{ℓ} は未知母数である。式(2.3)における無限個の未知母数 π_{ℓ} は自己回帰移動平均(AutoRegressive Moving Average, ARMA)を用いて有限個の未知母数で表現可能であることが知られている。ここでは、ARMA(p, q) フィルター:

$$\pi(B) = \sum_{\ell=1}^{\infty} \pi_{\ell} B^{\ell} = \frac{\theta(B)}{\phi(B)} - 1,$$

を用いて加重和の重み π_ℓ を有限個の未知母数で表すことにする. ただし, B は $Bd_{it} = d_{it-1}$ で定義されるラグ作用素であり,

$$\phi(B) = 1 - \sum_{\ell=1}^p \phi_\ell B^\ell, \quad \theta(B) = 1 + \sum_{\ell=1}^q \theta_\ell B^\ell,$$

は ARMA(p, q) における自己回帰と移動平均を規定するラグ作用素の多項式である. これより, Z_{it} は $p + q$ 個の未知母数 $\phi = (\phi_1, \dots, \phi_p)'$, $\theta = (\theta_1, \dots, \theta_q)'$ を用いて,

$$(2.4) \quad Z_{it} = \sum_{\ell=1}^p \phi_\ell (Z_{it-\ell} + e_{it-\ell}) + \sum_{\ell=1}^q \theta_\ell e_{it-\ell},$$

と表現される.

3. 変化係数の推定

本節では, 式(2.2)における変化係数 $\beta(t)$ の推定法について紹介する. 推定法の概要は, $\beta(t)$ の形状に線形構造を導入することで, 式(2.2)を Davis et al. (2003)の提案した一般化線形 ARMA モデルの枠組みに帰着させ, 尤度関数を陽に求めることにある.

まず, 変化係数 $\beta(t)$ の形状として $\beta(t) = \Xi \mathbf{x}(t)$ で表現されるクラスを考える. ここで, $\Xi = (\xi_1, \dots, \xi_k)'$ は $k \times r$ の未知母数行列であり, $\mathbf{x}(t)$ は r 個の既知の基底関数からなるベクトルである. 例えば, 変化係数の形状として $r-1$ 次多項式を用いるなら, $\mathbf{x}(t) = (1, t, \dots, t^{r-1})'$ となる. より滑らかな形状を仮定したい場合には, B スプライン基底やガウス基底などの基底を用いることも可能である. 基底については, 例えば, Satoh et al. (2003), Ruppert et al. (2003)の3章, 小西・北川(2004)の4章などを参照されたい. これにより, 変化係数 $\beta(t)$ を推定することは, 未知母数行列 Ξ を推定することと同値となる. 今, $\mathbf{x}_{it} = \mathbf{a}_i \otimes \mathbf{x}(t)$, $\xi = \text{vec}(\Xi)$ と定義すれば, 過去の観測値の履歴: $\mathcal{F}_{i,t-1}$ が与えられたときの d_{it} の条件付き期待値は,

$$\mu_{it} = e^{W_{it}}, \quad W_{it} = \mathbf{x}'_{it} \xi + Z_{it},$$

と書け, Davis et al. (2003)の提案した一般化線形 ARMA モデルで表される. モデルに含まれる未知母数は, 変化係数の形状を規定する $\xi = \text{vec}(\Xi)$, 時点間相関を記述する p 次の自己回帰に関する母数 $\phi = (\phi_1, \dots, \phi_p)'$ と q 次の移動平均に関する母数 $\theta = (\theta_1, \dots, \theta_q)'$ である.

Davis et al. (2003)の提案した尤度関数の構成法を参考に, 最尤法により未知母数 $\delta = (\xi', \theta', \phi)'$ の推定を行う. 式(2.1)より, 過去の観測値の履歴 $\mathcal{F}_{i,t-1}$ が与えられたときの d_{it} の分布はポアソン分布に従い, 密度関数を $f(d_{it} | \mathcal{F}_{i,t-1})$ とすれば, 対数尤度関数は $\sum_{i=1}^N \sum_{t=1}^T \log f(d_{it} | \mathcal{F}_{i,t-1})$ で与えられる. このとき, 未知母数を含まない項を除いた対数尤度関数 $\ell(\delta)$ は次式で与えられる.

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

W_{it} と Z_{it} は δ に依存する関数であるが, 表記を簡単にするため以下省略して説明する. 対数尤度関数 $\ell(\delta)$ の最大化には Newton-Raphson 法といった数値最適化を用いて行う. 数値最適化に必要な $\ell(\delta)$ の1階および2階の導関数は, W_{it} の偏微分, Z_{it} と e_{it} に関する漸化式で得られ, 詳細な記述は付録にまとめた. Newton-Raphson 法の初期値は, $\xi = \hat{\xi}_{GLM}$, $\theta = \mathbf{0}$, $\phi = \mathbf{0}$, $\{e_{it} = 0 | t \leq 0\}$ とした. $\hat{\xi}_{GLM}$ は, 観測時点間の相関を独立とした通常の一般化線形モデルにおける推定値である. Davis et al. (2003, 2005)の結果から, δ の最尤推定量 $\hat{\delta}$ は漸近的に,

$$(3.1) \quad \hat{\delta} \sim N_{rk+p+q}(\hat{\delta}, \Omega), \quad \hat{\Omega} = - \left(\frac{\partial^2 \ell(\delta)}{\partial \delta \partial \delta'} \Big|_{\delta = \hat{\delta}} \right)^{-1},$$

が成り立つ. これより, 変化係数 $\beta(t) = (\beta_1(t), \dots, \beta_k(t))'$ の推定値およびその漸近分布は, 次式で与えられる.

$$(3.2) \quad \hat{\beta}_j(t) = \hat{\xi}_j' \mathbf{x}(t) \sim N(\beta_j(t), \lambda_j(t)), \quad j=1, \dots, k,$$

ただし, $\lambda_j(t) = \mathbf{x}(t)' \hat{\Omega}_{jj} \mathbf{x}(t)$, $\hat{\Omega}_{jj}$ は $\hat{\xi}_j$ の分散共分散行列に対応する Ω の部分行列である.

4. 信頼区間の構築と検定

ここでは, Satoh and Yanagihara (2010) が提案した連続値の経時データに対する変化係数の推測法を参考に, 変化係数 $\beta(t) = (\beta_1(t), \dots, \beta_k(t))'$ の各 $\beta_j(t)$ に関する信頼区間の構築と検定問題について扱う. $\hat{\xi}$ の分散共分散行列の推定量 $\hat{\Omega} = \{\hat{\Omega}_{ij}\}$ が得られているとき, 変化係数の推定量 $\hat{\beta}_j(t)$ の分散は $\hat{\Omega}_{jj}$ の推定量 $\hat{\Omega}_{jj}$ を用いて $\hat{\lambda}_j(t) = \mathbf{x}(t)' \hat{\Omega}_{jj} \mathbf{x}(t)$ で与えられる. このとき, Rao (1973) の 1 章 1f 節にある Cauchy-Schwarz の不等式を用いて,

$$(4.1) \quad \sup_{t \in \mathbb{R}} \frac{\{\hat{\beta}_j(t) - \beta_j(t)\}^2}{\hat{\lambda}_j(t)} \leq \sup_{\mathbf{x} \in \mathbb{R}^r} \frac{\{(\hat{\xi}_j - \xi_j)' \mathbf{x}\}^2}{\mathbf{x}' \hat{\Omega}_{jj} \mathbf{x}} \leq (\hat{\xi}_j - \xi_j)' \hat{\Omega}_{jj}^{-1} (\hat{\xi}_j - \xi_j),$$

が成り立つことが, Satoh and Yanagihara (2010) により示された. この不等式により, 変化係数 $\beta_j(t)$ の信頼領域の構成問題は, 変化係数 $\beta_j(t)$ の形状を規定する未知母数 ξ_j の問題に置き換えられ, 時間 t の関数である変化係数 $\beta_j(t)$ が時間 t に依存しない上限で評価することが可能となる. また, $\hat{\xi}_j$ の漸近正規性により, 右辺は自由度 r のカイ二乗分布 χ_r^2 に近似的に従う. したがって, $c_{r,\alpha}$ を χ_r^2 の上側 $100\alpha\%$ 点, すなわち $\Pr(\chi_r^2 \geq c_{r,\alpha}) = \alpha$ とすると, 変化係数 $\beta_j(t)$ の $100\alpha\%$ 信頼領域は次式で与えられる.

$$(4.2) \quad \mathcal{I}_{j,\alpha}(t) = \left[\hat{\beta}_j(t) - \sqrt{\hat{\lambda}_j(t) c_{r,\alpha}}, \quad \hat{\beta}_j(t) + \sqrt{\hat{\lambda}_j(t) c_{r,\alpha}} \right].$$

このとき, 式(4.2)の信頼区間の被覆確率は漸近的に $\Pr(\beta_j(t) \in \mathcal{I}_{j,\alpha}(t) : \forall t \in \mathbb{R}) \geq 1 - \alpha$ が成り立つ.

次に, 変化係数に関する検定問題を考える. まず, 変化係数 $\beta_j(t)$ の値がすべての時間 t において 0 であるという帰無仮説は,

$$(4.3) \quad H_0 : \beta_j(t) \equiv 0 \quad (\forall t \in \mathbb{R}),$$

と書ける. 素朴な検定統計量として $T_j(t) = \hat{\beta}_j(t)^2 / \hat{\lambda}_j(t)$ を考えると, $\Pr(T_j(t) \leq x : \forall t \in \mathbb{R}^2)$ を評価する必要があるが,

$$(4.4) \quad \Pr(T_j(t) \leq x : \forall t \in \mathbb{R}) = \Pr\left(\max_{t \in \mathbb{R}} T_j(t) \leq x\right),$$

が成り立つことに注意すれば, 信頼領域の構成と同様に $\max_{t \in \mathbb{R}} T_j(t) \leq T_j$, $T_j = \hat{\xi}_j' \hat{\Omega}_{jj}^{-1} \hat{\xi}_j$ が成り立つことが分かる. T_j は式(4.3)の帰無仮説のもとで漸近的に χ_r^2 に従うので, 帰無仮説 H_0 は, $T_j > c_{r,\alpha}$ ならば有意水準 $100\alpha\%$ で棄却される, もしくは, $\Pr(\chi_r^2 > T_j)$ から p 値を求めることが可能である.

また, 変化係数 $\beta_j(t)$ の形状を構成する基底 $\mathbf{x}(t)$ の成分に 1 を含めれば, 変化係数が時間 t に依らず定数であるという一様性の仮説

$$(4.5) \quad H_0 : \beta_j(t) \equiv \text{const.} \quad (\forall t \in \mathbb{R})$$

の検定も可能である. 今, 基底を $\mathbf{x}(t) = (1, \mathbf{x}^{(g)}(t))'$, $\mathbf{x}^{(g)}(t) = (x_2(t), \dots, x_r(t))'$ とすれば, $\beta_j(t) = \xi_{j1} + \beta_j^{(g)}(t)$ と書ける. ただし, $\beta_j^{(g)}(t) = \xi_j^{(g)'} \mathbf{x}^{(g)}(t)$, $\xi_j^{(g)} = (\xi_{j2}, \dots, \xi_{jr})'$. これより,

変化係数に関する一様性の帰無仮説(4.5)は,

$$(4.6) \quad H_0: \beta_j^{(g)}(t) \equiv 0 \quad (\forall t \in \mathbb{R}),$$

で表される. 式(4.3)の帰無仮説の検定と同様に, $\xi_j^{(g)}$ の推定量を $\hat{\xi}_j^{(g)}$, $\hat{\xi}_j^{(g)}$ の分散共分散行列の推定量を $\hat{\Omega}_{jj}^{(g)}$, $\hat{\lambda}_j^{(g)}(t) = \mathbf{x}^{(g)}(t)' \hat{\Omega}_{jj}^{(g)} \mathbf{x}^{(g)}(t)$ とすれば, 素朴な検定統計量 $T_j^{(g)}(t) = \hat{\beta}_j^{(g)}(t)^2 / \hat{\lambda}_j^{(g)}(t)$ に対して, $\max_{t \in \mathbb{R}} T_j^{(g)}(t) \leq T_j^{(g)}$, $T_j^{(g)} = \hat{\xi}_j^{(g)'} \hat{\Omega}_{jj}^{(g)-1} \hat{\xi}_j^{(g)}$ が成り立つ. $T_j^{(g)}$ が式(4.6)の帰無仮説のもとで漸近的に χ_{r-1}^2 に従うので, 帰無仮説 H_0 は, $T_j^{(g)} > c_{r-1, \alpha}$ ならば有意水準 $100\alpha\%$ で棄却される, もしくは, $\Pr(\chi_{r-1}^2 > T_j^{(g)})$ から p 値を求めることが可能である.

5. 実データへの適用例

実データへの適用例として, 日本における男性の大腸がん死亡数データに提案手法を適用し解析を行った. データは1975年から2002年までの期間に都道府県別に毎年集計された28時点の経時データである. 期待死亡数算出のための基準集団には, 1975年から1994年の期間における全国平均を用いた. 説明変数には, 人口密度 [persons/km²], 日照量 [KWh/m²/day] および自然空間放射線 [nGy/h] を用いた, ただし全ての説明変数に関して各平均値でシフトしてある. 2節の表記に従って書けば, $\mathbf{a}_i = (a_{i1}, a_{i3}, a_{i2}, a_{i4})'$, $a_{i1} = 1$, $a_{i2} =$ 人口密度, $a_{i3} =$ 日照量, $a_{i4} =$ 自然空間放射線, $i = 1, \dots, 47$ となる. 人口密度は都会度を表す指標として用いた. 日照量に関して, Mizoue (2004) が横断的研究により消化器系がん死亡危険度との負の相関関係を報告しており, 同様な日照量データ(1961年から1990年の間の平年値)を用いた. 自然空間放射線は, Abe et al. (1981) の調査に基づく放射線医学総合研究所の公開データを用いた.

変化係数 $\beta(t)$ の形状を規定する基底関数として $r-1$ 次の多項式を用い, AIC を用いて最適な次数 (r, p, q) の選択を行った. 表1は, $r=4$ (つまり, 3次多項式を適合) の場合に, (p, q) をそれぞれ0から4まで動かしたときのAICの値である.

AICが最小となるのは $(r, p, q) = (4, 1, 1)$ であった. 他の r の値についても検討したが, AICが最小となるのは $(r, p, q) = (4, 1, 1)$ であった. 図1は, $(r, p, q) = (4, 1, 1)$ の場合の適合曲線である. 黒色の太線が適合曲線であり, 灰色の細線は観測値をつないだ折れ線である.

図2(a)から図2(d)は, $(r, p, q) = (4, 1, 1)$ の場合の変化係数 $\beta(t)$ の推定曲線である. 図2(a)が全ての説明変数の値が各平均値の場合の死亡危険度の時間変化の推定曲線, 図2(b)が人口密度の効果, 図2(c)が日照量の効果, 図2(d)が自然空間放射線の効果を表す変化係数の推定曲線に対応する. 上下の破線は, 各推定曲線の95%信頼区間である.

変化係数の有意性の検定および一様性の検定の結果を表2に示す. 表2の各セルの数値は検定統計量の値であり, 括弧の数値は対応する p 値である.

図2(b)から, 人口密度の効果は常に正值であることから, 地方(人口密度の低い地域)に比べて都会(人口密度の高い地域)の死亡リスクが経年的に高いことが分かる. また, 地方と都会

表1. AICによる (p, q) の選択 ($r=4$ の場合).

$p \setminus q$	0	1	2	3	4
0	10324	10096	9994	9937	9903
1	10089	9734	9735	9737	9739
2	9984	9735	9738	9739	9741
3	9926	9737	9739	9741	9744
4	9891	9739	9741	9743	9738

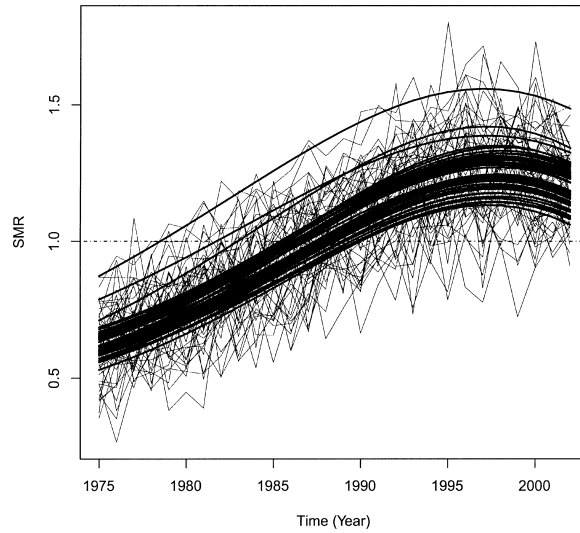
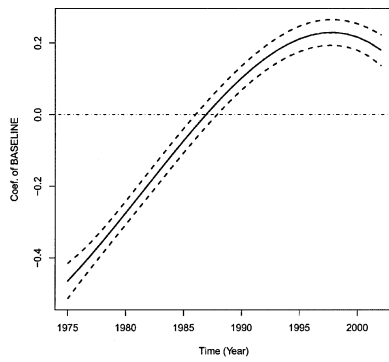
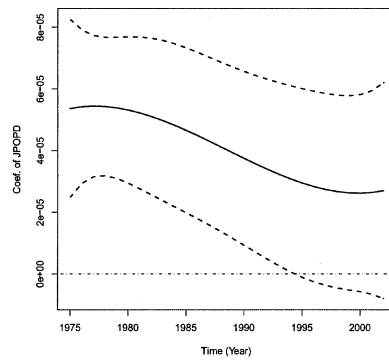


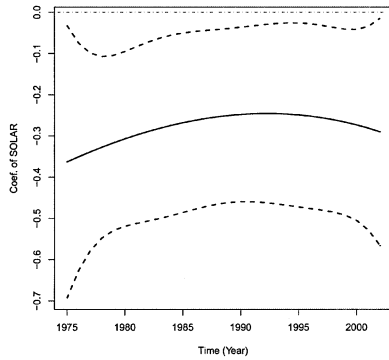
図 1. $(r, p, q) = (4, 1, 1)$ における適合曲線.



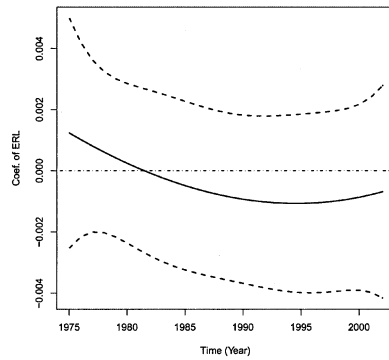
(a) 時間効果



(b) 人口密度の効果



(c) 日照量の効果



(d) 自然空間放射線の効果

図 2. $(r, p, q) = (4, 1, 1)$ における変化係数 $\beta(t)$ の推定曲線 (実線) と 95% 信頼区間 (破線).

表 2. $(r, p, q) = (4, 1, 1)$ における変化係数に関する検定.

	$H_0 : \beta_j(t) \equiv 0$	$H_0 : \beta_j(t) \equiv \text{const.}$
人口密度	55.693 ($p < 0.001$)	7.183 ($p = 0.066$)
日照量	26.662 ($p < 0.001$)	1.261 ($p = 0.738$)
自然空間放射線	3.165 ($p = 0.531$)	3.139 ($p = 0.371$)

の差は経年的に見て減少傾向であるが、表 2 の一様性の検定結果から、5%有意とはならなかった。図 2 (c) から、日照量の効果は常に負値であることから、日照量の低い地域は高い地域に比べて死亡リスクが経年的に高く、この結果は Mizoue (2004) の報告と一致する。また、表 2 の一様性の検定結果から、日照量の効果は経年的に定数であることが示唆された。図 2 (d) から、自然空間放射線の効果は常に 0 近くにあること、表 2 の検定結果から、自然空間放射線の死亡危険度への寄与はみられなかった。

以上の解析結果をまとめると、都会および日照量の少ない地域での死亡危険度が高い傾向があることが示唆された。また、このような傾向は経年的に変化していないと考えられ、その結果、死亡危険度の適合曲線はほぼ平行となった。日照量のがん死亡危険度に関する効果は、日光浴により生成されるビタミン D が関与していると考えられている (Mizoue, 2004; Garland and Garland, 2006; Boscoe and Schymura, 2006)。

6. おわりに

本論文では、都道府県別がん死亡数の経時測定データに対して、観測値駆動型モデルを導入することで観測時点間相関を考慮しつつ、変化係数を用いてがん死亡危険度の変動要因の推測を行った。変化係数の推定は、カーネル平滑化法 (Hoover et al., 1998 など) を用いるのが一般的であるが、本論文では、Sato and Yanagihara (2010) の推測法を参考に、変化係数の形状に線形構造を仮定した推測法を適用した。線形構造を仮定することで、変化係数に関する同時信頼区間の構築や時間変化の有無に関する仮説検定が可能となった。経時データにおける観測値の時点間相関を考慮したモデルとして、ランダム変量を用いた一般化線形混合モデル (GLMM) に代表されるランダム効果モデル (例えば、Breslow and Clayton, 1993) と、一般化推定方程式 (GEE) に代表される周辺モデル (例えば、Liang and Zeger, 1986) がよく使われている。GLMM では、導入するランダム変数の数だけ陽に解けない多重積分を数値計算で評価する必要があり、観測時点や説明変数の数が比較的多いデータに対する適用は困難となる。GEE は、計算量が少なく適用は容易であるものの、観測時点が多い場合には分散共分散に関する未知パラメータ数が増大し推定が不安になり易く、また、尤度が得られないことから、AIC といったモデル選択規準も利用できない。本論文では、ランダム変量の代わりに過去の観測値を用いた観測値駆動型モデルを適用した。観測値駆動型モデルを用いる利点は、少数の未知パラメータで時点間相関のモデル化が可能であり、尤度関数が陽に表現可能であることから計算量が少なく、実データへの適用が容易であることにある (Davis et al., 1999, 2003, 2005)。観測値駆動型モデルの導入により、より効率よく変化係数の推測が可能となった。また、尤度関数が陽に得られ容易に計算可能であることから、AIC といった従来の変数選択の手法を用いることができた (表 1)。

次に、提案手法を日本における男性の大腸がん死亡数データに適用した結果について考察を行う。本論文では、がん死亡危険度の変動要因として人口密度、日照量、自然空間放射線について検討を行った。変化係数に関する一様性の検定結果は 5% 有意ではないものの (表 2)、人口密度の効果に経年的な減少傾向が見られた (図 2 (b))。これは、近年になって都会と田舎の

地域差が小さくなっていると考えられる。日照量のがん死亡危険度への影響に関する報告は、Mizoue (2004), Garland and Garland (2006) および Boscoe and Schymura (2006) らにより報告されているが、これらの報告は横断データの解析報告である。本論文では、経時データに対して変化係数を推定し、効果の時間変化の有無についても検討を行った。経時データ解析により、日照量の効果に時間変化が認められず経年的にほぼ一定であることが分かった(表 2, 図 2 (c))。

最後に、今回扱ったがん統計データは、都道府県という比較的大きな構成単位の集計データであったが、より細かな地域差に関心がある場合には、市区町村や第2次医療圏などを構成単位としたがん統計データの解析が必要となる。このような解析を行う際には、時点間相関に加えて空間相関を適切にモデルに導入することが重要となり、空間相関にも対応した解析法の開発は今後の課題である。

付録： $\ell(\delta)$ の導関数に関する漸化式

対数尤度関数 $\ell(\delta)$ の導関数は次式で与えられる：

$$\frac{\partial \ell}{\partial \delta} = \sum_{i=1}^N \sum_{t=1}^T (y_{it} - e^{W_{it}}) \frac{\partial W_{it}}{\partial \delta},$$

$$\frac{\partial^2 \ell}{\partial \delta \partial \delta'} = \sum_{i=1}^N \sum_{t=1}^T \left\{ (y_{it} - e^{W_{it}}) \frac{\partial^2 W_{it}}{\partial \delta \partial \delta'} - e^{W_{it}} \frac{\partial W_{it}}{\partial \delta} \frac{\partial W_{it}}{\partial \delta'} \right\},$$

ただし、

$$\frac{\partial W_{it}}{\partial \delta} = \begin{bmatrix} \mathbf{x}_{it} \\ \mathbf{0}_{p+q} \end{bmatrix} + \frac{\partial Z_{it}}{\partial \delta}, \quad \frac{\partial^2 W_{it}}{\partial \delta \partial \delta'} = \frac{\partial Z_{it}}{\partial \delta \partial \delta'}.$$

Z_{it} の 1 階微分は以下のような漸化式から得られる：

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

各成分は、ある添字 a に対して、

$$\frac{\partial Z_{it}}{\partial \xi_a} = \sum_{\ell=1}^p \phi_{\ell} \left(\frac{\partial Z_{it-\ell}}{\partial \xi_a} + \frac{\partial e_{it-\ell}}{\partial \xi_a} \right) + \sum_{\ell=1}^q \theta_{\ell} \frac{\partial e_{it-\ell}}{\partial \xi_a},$$

$$\frac{\partial Z_{it}}{\partial \phi_a} = (Z_{it-a} + e_{it-a}) + \sum_{\ell=1}^p \phi_{\ell} \left(\frac{\partial Z_{it-\ell}}{\partial \phi_a} + \frac{\partial e_{it-\ell}}{\partial \phi_a} \right) + \sum_{\ell=1}^q \theta_{\ell} \frac{\partial e_{it-\ell}}{\partial \phi_a},$$

$$\frac{\partial Z_{it}}{\partial \theta_a} = \sum_{\ell=1}^p \phi_{\ell} \left(\frac{\partial Z_{it-\ell}}{\partial \theta_a} + \frac{\partial e_{it-\ell}}{\partial \theta_a} \right) + e_{it-a} + \sum_{\ell=1}^q \theta_{\ell} \frac{\partial e_{it-\ell}}{\partial \theta_a}.$$

Z_{it} の 2 階微分も以下のような漸化式から得られる：

$$\frac{\partial^2 Z_{it}}{\partial \delta \partial \delta'} = \sum_{\ell=1}^p \left\{ \frac{\partial \phi_{\ell}}{\partial \delta} \left(\frac{\partial Z_{it-\ell}}{\partial \delta'} + \frac{\partial e_{it-\ell}}{\partial \delta'} \right) + \left(\frac{\partial Z_{it-\ell}}{\partial \delta} + \frac{\partial e_{it-\ell}}{\partial \delta} \right) \frac{\partial \phi_{\ell}}{\partial \delta'} \right. \\ \left. + \phi_{\ell} \left(\frac{\partial^2 Z_{it-\ell}}{\partial \delta \partial \delta'} + \frac{\partial^2 e_{it-\ell}}{\partial \delta \partial \delta'} \right) \right\} \\ + \sum_{\ell=1}^q \left\{ \frac{\partial \theta_{\ell}}{\partial \delta} \frac{\partial e_{it-\ell}}{\partial \delta'} + \frac{\partial e_{it-\ell}}{\partial \delta} \frac{\partial \theta_{\ell}}{\partial \delta'} + \theta_{\ell} \frac{\partial^2 e_{it-\ell}}{\partial \delta \partial \delta'} \right\}.$$

各成分は、ある添字 a, b に対して、

$$\begin{aligned} \frac{\partial^2 Z_{it}}{\partial \xi_a \partial \xi_b} &= \sum_{\ell=1}^p \phi_\ell \left(\frac{\partial^2 Z_{it-\ell}}{\partial \xi_a \partial \xi_b} + \frac{\partial^2 e_{it-\ell}}{\partial \xi_a \partial \xi_b} \right) + \sum_{\ell=1}^q \theta_\ell \frac{\partial^2 e_{it-\ell}}{\partial \xi_a \partial \xi_b}, \\ \frac{\partial^2 Z_{it}}{\partial \xi_a \partial \phi_b} &= \frac{\partial Z_{it-b}}{\partial \xi_a} + \frac{\partial e_{it-b}}{\partial \xi_a} + \sum_{\ell=1}^p \phi_\ell \left(\frac{\partial^2 Z_{it-\ell}}{\partial \xi_a \partial \phi_b} + \frac{\partial^2 e_{it-\ell}}{\partial \xi_a \partial \phi_b} \right) + \sum_{\ell=1}^q \theta_\ell \frac{\partial^2 e_{it-\ell}}{\partial \xi_a \partial \phi_b}, \\ \frac{\partial^2 Z_{it}}{\partial \xi_a \partial \theta_b} &= \sum_{\ell=1}^p \phi_\ell \left(\frac{\partial^2 Z_{it-\ell}}{\partial \xi_a \partial \theta_b} + \frac{\partial^2 e_{it-\ell}}{\partial \xi_a \partial \theta_b} \right) + \frac{\partial^2 e_{it-b}}{\partial \xi_a} + \sum_{\ell=1}^q \theta_\ell \frac{\partial^2 e_{it-\ell}}{\partial \xi_a \partial \theta_b}, \\ \frac{\partial^2 Z_{it}}{\partial \phi_a \partial \phi_b} &= \frac{\partial Z_{it-a}}{\partial \phi_b} + \frac{\partial e_{it-a}}{\partial \phi_b} + \frac{\partial Z_{it-b}}{\partial \phi_a} + \frac{\partial e_{it-b}}{\partial \phi_a} \\ &\quad + \sum_{\ell=1}^p \phi_\ell \left(\frac{\partial^2 Z_{it-\ell}}{\partial \phi_a \partial \phi_b} + \frac{\partial^2 e_{it-\ell}}{\partial \phi_a \partial \phi_b} \right) + \sum_{\ell=1}^q \theta_\ell \frac{\partial^2 e_{it-\ell}}{\partial \phi_a \partial \phi_b}, \\ \frac{\partial^2 Z_{it}}{\partial \phi_a \partial \theta_b} &= \frac{\partial Z_{it-a}}{\partial \theta_b} + \frac{\partial e_{it-a}}{\partial \theta_b} + \sum_{\ell=1}^p \phi_\ell \left(\frac{\partial^2 Z_{it-\ell}}{\partial \phi_a \partial \theta_b} + \frac{\partial^2 e_{it-\ell}}{\partial \phi_a \partial \theta_b} \right) \\ &\quad + \frac{\partial e_{it-b}}{\partial \phi_a} + \sum_{\ell=1}^q \theta_\ell \frac{\partial^2 e_{it-\ell}}{\partial \phi_a \partial \theta_b}, \\ \frac{\partial^2 Z_{it}}{\partial \theta_a \partial \theta_b} &= \sum_{\ell=1}^p \phi_\ell \left(\frac{\partial^2 Z_{it-\ell}}{\partial \theta_a \partial \theta_b} + \frac{\partial^2 e_{it-\ell}}{\partial \theta_a \partial \theta_b} \right) + \frac{\partial e_{it-a}}{\partial \theta_b} + \frac{\partial e_{it-b}}{\partial \theta_a} + \sum_{\ell=1}^q \theta_\ell \frac{\partial^2 e_{it-\ell}}{\partial \theta_a \partial \theta_b}. \end{aligned}$$

e_{it} の 1 階微分と 2 階微分は次式で与えられる:

$$\begin{aligned} \frac{\partial e_{it}}{\partial \delta} &= - \left(e^{\frac{1}{2}W_{it}} + \frac{1}{2}e_{it} \right) \frac{\partial W_{it}}{\partial \delta}, \\ \frac{\partial^2 e_{it}}{\partial \delta \partial \delta'} &= - \left(e^{\frac{1}{2}W_{it}} + \frac{1}{2}e_{it} \right) \frac{\partial^2 W_{it}}{\partial \delta \partial \delta'} + \frac{1}{4}e_{it} \frac{\partial W_{it}}{\partial \delta} \frac{\partial W_{it}}{\partial \delta'}. \end{aligned}$$

謝 辞

本論文作成にあたって、ニューサウスウェールズ大学の William Dunsmuir 教授には貴重なコメントを頂いた。記して深く感謝申し上げたい。また、本誌査読者の方からは論文改訂にあたって有益なコメントを頂いた。ここに感謝いたします。本研究は、厚生労働省科研費「がん罹患・死亡動向の実態把握の研究」からの援助を受けた。

参 考 文 献

- Abe, S., Fujitaka, K., Abe, M. and Fujimoto, K. (1981). Extensive field survey of natural radiation in Japan, *Journal of Nuclear Science and Technology*, **18**, 21–45.
- Boscoe, F. P. and Schymura, M. J. (2006). Solar ultraviolet-B exposure and cancer incidence and mortality in the United States, 1993–2002, *BMC Cancer*, Art. No. 264.
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, **88**, 9–25.
- Chan, K. S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts, *Journal of the American Statistical Association*, **90**, 242–252.
- Davis, R. A., Dunsmuir, W. T. M. and Wang, Y. (1999). Modelling time series of count data, *Asymp-*

- otics, Nonparametrics and Time Series* (ed. S. Ghosh), 63–114, Marcel Dekker, New York.
- Davis, R. A., Dunsmuir, W. T. M. and Streett, S. B. (2003). Observation driven models for Poisson counts, *Biometrika*, **90**, 777–790.
- Davis, R. A., Dunsmuir, W. T. M. and Streett, S. B. (2005). Maximum Likelihood estimation for an observation driven model for poisson counts, *Methodology and Computing in Applied Probability*, **7**, 149–159.
- Diggle, P. J., Liang, K-Y. and Zeger, S. L. (1994). *Analysis of Longitudinal Data*, Oxford University Press, New York.
- Dobson, A. (1990). *Introduction to Generalized Linear Models*, Chapman and Hall, London.
- Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-Gaussian observation based on state space models from both classical and Bayesian perspectives (with discussion), *Journal of the Royal Statistical Society: Series B*, **62**, 3–59.
- Garland, C. F. and Garland, F. C. (2006). Do sunlight and vitamin D reduce the likelihood of colon cancer?, *International Journal of Epidemiology*, **35**, 217–220.
- Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models, *Journal of the Royal Statistical Society: Series B*, **55**, 757–796.
- Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data, *Biometrika*, **85**, 809–822.
- Jung, R. C. and Liesenfeld, R. (2001). Estimating time series models for count data using efficient importance sampling, *Allgemeines Statistisches Archiv*, **85**, 387–407.
- 小西貞則, 北川源四郎(2004). 『情報量規準』, 朝倉書店, 東京.
- Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13–22.
- Mizoue, T. (2004). Ecological study of solar radiation and cancer mortality in Japan, *Health Physics*, **87**, 532–538.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed., John Wiley, New York.
- Ruppert, D., Wand, M. and Carroll, R. (2003). *Semiparametric Regression*, Cambridge University Press, New York.
- Satoh, K. and Yanagihara, H. (2010). Estimation of varying coefficients for a growth curve model, *American Journal of Mathematical and Management Sciences* (in press).
- Satoh, K., Yanagihara, H. and Ohtaki, M. (2003). Bridging the gap between *B*-spline and polynomial regression model, *Communications in Statistics—Theory and Methods*, **32**, 179–190.
- 佐藤健一, 柳原宏和, 加茂憲一(2009). 離散分布の経時測定データにおける線形な変化係数の推測について, *応用統計学*, **38**, 1–11.
- 富田哲治, 佐藤健一, 柳原宏和(2010). 空間データに対する交互作用モデルを用いた変化係数曲面の推測について, *応用統計学*, **39**, 59–70.
- Tonda, T., Satoh, K., Nakayama, T., Katanoda, K., Sobue, T. and Ohtaki, M. (2011). A nonparametric mixed-effects model for cancer mortality, *Australian and New Zealand Journal of Statistics* (in press).
- Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach, *Biometrics*, **44**, 1019–1031.

Statistical Analysis of Time Trend in Cancer Mortality Using Time Varying Coefficient

Tetsuji Tonda¹, Kenichi Satoh¹, Teruyuki Nakayama², Kota Katanoda³,
Tomotaka Sobue³ and Megu Ohtaki¹

¹Research Institute for Radiation Biology and Medicine, Hiroshima University

²Faculty of Nursing, International University of Health and Welfare

³Research Center for Cancer Prevention and Screening, National Cancer Center

Disease risk, which is usually quantified by incidence or mortality, may have patterns of change by region or time due to varying factors such as level of exposure to environmental hazards or lifestyle habits of residents. Our aim is to assess how regional variations of cancer mortality change over time and what characteristics influence such changes using longitudinal cancer mortality data. In this paper, we develop a statistical method for longitudinal cancer mortality data using an observation driven model with time-varying coefficient. As an illustration based on real data, we apply the proposed method to a set of prefecture-specific longitudinal data on mortality from males' large bowel cancer in Japan.