

# 線形な変化係数における信頼区間の精密化

県立広島大学 経営情報学部 富田 哲治

広島大学 原爆放射線医科学研究所 佐藤 健一

**要旨** 経時データにおけ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  $\beta(t)$  は変化係数とよばれている。変化係数の推定は、カーネル平滑化の要領で、固定した時点周辺の近傍データに対して局所的な回帰を繰り返すことで推定され、その信頼区間も固定した各点毎に構築されるのが一般的であった。近年、Satoh and Yanagihara (2010) は、変化係数の関数形を線形に限定することで、 $t \in \mathbb{R}$  での同時信頼区間を提案した。本稿では、信頼区間を構築する領域を観測時点の範囲といった有限区間  $t \in [a, b]$  に限定することで、より精密な同時信頼区間の構築法を提案する。

## 1. はじめに

回帰分析において、時間とともに変化する回帰係数は変化係数とよばれ、Hastie and Tibshirani (1993) らによって提案された。横軸に時間  $t$  をとれば、変化係数  $\beta(t)$  は説明変数の効果を表す曲線として容易に視覚化できる。その推定は一般的には時間軸に沿った平滑化によって行われる(例えば、Hoover et al. 1998; Satoh and Ohtaki 2006; Tonda et al. 2011)。すなわち、固定された時間近傍毎のデータを用いて重回帰を繰り返し行い、得られた回帰係数を時間軸に沿って並べることで推定する。しかし、この方法では各点毎の信頼区間しか構成できず、関数としての変化係数曲線全体を包含する同時信頼区間を構成することは困難であった。そこで、Satoh and Yanagihara (2010) は成長曲線モデルにおいて変化係数を線形な基底  $\mathbf{x}(t) = (x_1(t), \dots, x_q(t))'$  で記述できる関数族  $\beta(t) = \boldsymbol{\theta}'\mathbf{x}(t)$  で考えた。例えば、多項式なら  $\mathbf{x}(t) = (1, t, \dots, t^{q-1})'$  である。このように線形な変化係数に限定することで、Satoh and Yanagihara (2010) は Rao (1973) で示された多変量解析の理論を利用して、 $t \in \mathbb{R}$  での変化係数の同時信頼区間を構成することを提案した。

線形な変化係数における未知母数  $\boldsymbol{\theta}$  は、共変量と基底  $\mathbf{x}(t)$  との交互作用として容易に推定可能である。簡単な例として、説明変数  $a$  の回帰係数が時間  $t$  とともに変化する変化係数  $\beta(t)$  とかけ、回帰モデルに  $\beta(t)a$  という項が存在する場合を考える。変化係数の関数形が時間について直線と仮定すると、基底  $\mathbf{x}(t) = (1, t)'$  を用いて  $\beta(t) = \boldsymbol{\theta}'\mathbf{x}(t) = \theta_1 + \theta_2 t$  とかける。したがって、 $\beta(t)a = \theta a + \theta_2 (ta)$  となり、交互作用項を考えることに等しい。ゆえに、線形な変化係数は後述する重回帰モデル以外のより一般的な回帰モデルでも適用可能であり、汎用性が高い。佐藤・柳原・加茂 (2009) では、Liang and Zeger (1986) の提案した一般化推定方程式の枠組みを利用し、

離散値の経時データにおける変化係数の推測へ適用した。また、富田・佐藤・柳原 (2010) では、空間上の位置によって変化する回帰係数を変化係数曲面として取り上げ、地理的加重回帰 (例えば、Fotheringham et al. 2002) と対比させながら線形な変化係数の推測を提案した。実データ解析への応用として、富田ら (2011) はがん死亡数の経時データに対して時系列解析で使われる系列相関構造を導入しがん死亡数の変動要因の経時変化を変化係数により記述した。富田ら (2010, 2012) および Tonda et al. (2012) では、コックス回帰 (1972, 1975) に位置によって変化する変化係数曲面を導入し、広島原爆被曝者コホートデータに対して、被曝時所在地により変化するハザード比に基づきリスク地図の推定を行った。また、加茂・富田・佐藤 (2011) は、観測年毎の年齢階級別死亡数のデータからがんの死亡率を年齢・時代空間上で変化する曲面として視覚化することを試みた。

本稿では、線形な変化係数  $\beta(t)$  の時間  $t$  の関数としての同時信頼区間を構築する領域を有限区間  $t \in [a, b]$  に限定することで、同時信頼区間の構築法について考える。回帰モデルにおける有限区間での同時信頼区間については、重回帰モデルの枠組みで Liu (2010) が詳しくまとめており、Liu et al. (2007) が重回帰モデルの比較に対するシミュレーションに基づく同時信頼区間の構築法、加藤・栗木 (2012) が 2 次多項式回帰モデルの正值性検定についてチューブ法に基づく方法を提案している。本稿では、多重比較法 (例えば、Bretz, Hothorn and Westfall 2011) を応用することで、より簡便で汎用的な構築法を提案する。以下、2 節では、Satoh and Yanagihara (2010) で提案された線形な変化係数の信頼区間の構築法と対比させて、本稿で提案する信頼区間の構築法について論じる。3 節では、Satoh and Yanagihara (2010) と同じデータを用いて、統計ソフト R (R Core Team 2012) での実行例と合わせて 2 節で示した信頼区間の比較を行い、提案法の有効性について調べる。4 節では、3 節の解析例に関連した数値実験を行う。

## 2. 変化係数の信頼区間

線形な変化係数  $\beta(t) = \boldsymbol{\theta}'\mathbf{x}(t)$  に対して未知母数ベクトル  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$  の推定量  $\hat{\boldsymbol{\theta}}$  の漸近分布が  $\hat{\boldsymbol{\theta}} \sim N_q(\boldsymbol{\theta}, \Omega)$  として得られているとき、変化係数  $\beta(t)$  の推定量とその漸近分布は  $\hat{\beta}(t) = \hat{\boldsymbol{\theta}}'\mathbf{x}(t) \sim N(\beta(t), \lambda^2(t))$  で与えられる。ただし、 $\lambda(t) = \sqrt{\mathbf{x}(t)'\hat{\Omega}\mathbf{x}(t)}$  である。ここでは、 $\Omega$  の推定量を  $\hat{\Omega}$  とし変化係数  $\beta(t)$  に関する  $100(1 - \alpha)\%$  信頼区間、

$$\mathcal{I}_{1-\alpha}(t|u_\alpha) = \left[ \hat{\beta}(t) - u_\alpha \hat{\lambda}(t), \hat{\beta}(t) + u_\alpha \hat{\lambda}(t) \right], \quad \hat{\lambda}(t) = \sqrt{\mathbf{x}(t)'\hat{\Omega}\mathbf{x}(t)} \quad (1)$$

の構築法について考える。 $u_\alpha$  は信頼区間  $\mathcal{I}_{1-\alpha}(t|u_\alpha)$  の被覆確率  $\Pr(\beta(t) \in \mathcal{I}_{1-\alpha}(t|u_\alpha))$  を決める閾値である。 $\hat{\beta}(t)$  の漸近正規性から、標準正規分布の上側  $100\alpha\%$  点  $q_\alpha$  を用いて  $u_\alpha = q_{\alpha/2}$  とすることで、時点  $t$  を固定した時の被覆確率が  $\Pr(\beta(t) \in \mathcal{I}_{1-\alpha}(t|q_{\alpha/2})) \approx 1 - \alpha$  を満たす各点毎の信頼区間  $\mathcal{I}_{1-\alpha}(t|q_{\alpha/2})$  が構築される。

Satoh and Yanagihara (2010) は Rao (1973) の多変量解析の理論を用いて、変化係数の Wald 型統計量の上限を与える不等式、

$$\begin{aligned} \sup_{t \in \mathbb{R}} \left\{ \frac{\hat{\beta}(t) - \beta(t)}{\lambda(t)} \right\}^2 &= \sup_{t \in \mathbb{R}} \frac{\left\{ \mathbf{x}(t)'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right\}^2}{\mathbf{x}(t)'\hat{\Omega}\mathbf{x}(t)} \leq \sup_{\mathbf{x} \in \mathbb{R}^q} \frac{\left\{ \mathbf{x}'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right\}^2}{\mathbf{x}'\hat{\Omega}\mathbf{x}} \\ &= (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \hat{\Omega}^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightsquigarrow \chi_q^2, \end{aligned} \quad (2)$$

を導出し、自由度  $q$  の  $\chi^2$  分布の上側  $100\alpha\%$  点  $c_{q,\alpha}$  を用いて  $u_\alpha = \sqrt{c_{q,\alpha}}$  とすることで、 $t \in \mathbb{R}$  における被覆確率が  $\Pr(\beta(t) \in \mathcal{I}_{1-\alpha}(t|\sqrt{c_{q,\alpha}})) \geq 1 - \alpha$  を満たす同時信頼区間  $\mathcal{I}_{1-\alpha}(t|\sqrt{c_{q,\alpha}})$  を構築した。

本稿では、同時信頼区間を構築するために有限区間  $t \in [a, b]$  における被覆確率が  $\Pr(\beta(t) \in \mathcal{I}_{1-\alpha}(t|u_\alpha)) = 1 - \alpha$  を満たす変化係数  $\beta(t)$  の同時信頼区間を考える。そのためには、

$$\Pr(\mathcal{T} \leq u_\alpha) = 1 - \alpha, \quad \mathcal{T} = \sup_{t \in [a,b]} |\mathcal{T}(t)|, \quad \mathcal{T}(t) = \frac{\hat{\beta}(t) - \beta(t)}{\lambda(t)}, \quad (3)$$

を満たす閾値  $u_\alpha$  を求める必要があるが<sup>5</sup>、一般にこのような  $u_\alpha$  を求めることは困難である。そこで、ここでは、区間  $[a, b]$  を  $K$  個の時点に離散化した時点の集合、

$$I_K(a, b) = \{t_k, k = 1, \dots, K | a = t_1 < t_2 < \dots < t_K = b\}, \quad (4)$$

における  $\mathcal{T}(t)$  の値  $T_k = \mathcal{T}(t_k)$  ( $k = 1, \dots, K$ ) を用いて、 $\mathcal{T}$  を次式で近似する。

$$\mathcal{T} \approx T_{max} = \max_{t \in I_K(a,b)} |\mathcal{T}(t)| = \max\{|T_1|, |T_2|, \dots, |T_K|\}. \quad (5)$$

$K$  が十分大きい時、 $T_{max}$  は  $\mathcal{T}$  のよい近似となる。このとき、

$$\Pr(T_{max} \leq u_\alpha) = 1 - \alpha \Leftrightarrow \Pr(|T_1| \leq u_\alpha, \dots, |T_K| \leq u_\alpha) = 1 - \alpha \quad (6)$$

が成り立つ。また、 $\mathbf{t} = (T_1, \dots, T_K)'$  は漸近的に多変量正規分布、

$$\mathbf{t} = (T_1, \dots, T_K)' \rightsquigarrow N_K(\mathbf{0}, R), \quad (7)$$

に従う。ここで、 $R = DC\Omega C'D$ 、 $C = (\mathbf{x}(t_1), \dots, \mathbf{x}(t_K))'$ 、 $D = \text{diag}(\lambda(t_1), \dots, \lambda(t_K))^{-1}$  である。したがって、式 (1) における閾値  $u_\alpha$  は上の多変量正規分布から近似的に求まる。今、 $\Phi_K(\mathbf{x}|R) = \Pr(-x \leq t_1 \leq x, \dots, -x \leq t_K \leq x|R)$  と定義すると、 $\Phi_K(\mathbf{x}|R)$  は多変量正規密度の  $K$  次元の多重積分

$$\Phi_K(\mathbf{x}|R) = \frac{1}{|R|^{1/2}(2\pi)^{K/2}} \int_{-x}^x \dots \int_{-x}^x e^{-\frac{1}{2}\mathbf{t}'R^{-1}\mathbf{t}} d\mathbf{t}, \quad (8)$$

で表される。したがって、 $\Phi_K(x_\alpha|R) = 1 - \alpha$  を満たす  $x_\alpha$  を用いて  $u_\alpha = x_\alpha$  とすることで、 $t \in [a, b]$  における被覆確率が  $\Pr(\beta(t) \in \mathcal{I}_{1-\alpha}(t|x_\alpha)) \approx 1 - \alpha$  を満たす同時信頼区間  $\mathcal{I}_{1-\alpha}(t|x_\alpha)$  が構築される。

多変量正規分布における多重積分の計算法については、Genz and Bretz (2009) が詳しく、Genz (1992, 1993) および Genz and Bretz (2002) のアルゴリズムに基づく方法が統計ソフト R の mvtnorm パッケージ (Hothorn, Bretz and Genz 2001) で提供されており、1000 次元までの計算に対応している。

### 3. 適用例

ここでは、Sato and Yanagihara (2010) の解析例で用いられた 2 つの経時データを用いて、

変化係数の信頼区間：(1) 各点毎の信頼区間  $\mathcal{I}_{1-\alpha}(t|q_{\alpha/2})$ , (2)  $t \in \mathbb{R}$  における同時信頼区間  $\mathcal{I}_{1-\alpha}(t|\sqrt{c_{q,\alpha}})$ , (3)  $t \in [a, b]$  における同時信頼区間  $\mathcal{I}_{1-\alpha}(t|x_{\alpha})$ , の比較を行い, 提案法の有効性について検証する.

### 3.1. 少年・少女の下顎枝の成長データ

このデータは Potthoff and Roy (1964) に掲載された少年 16 人および少女 11 人の下顎枝の長さ (mm) の成長データ (8 歳から 14 歳まで 2 年間隔で計 4 回測定) である. R では nlme パッケージに Orthodont として格納されており, distance が下顎枝の長さ, age が測定時年齢, Subject が個体識別番号, Sex が性別である. 以下, R のスクリプトによる実行例と合わせて説明する.

```
> library(nlme); data(Orthodont)
> levels(Orthodont$Sex) <- c("M","F"); Orthodont$Sex <- relevel(Orthodont$Sex, ref = "F");
> head(Orthodont)
Grouped Data: distance ~ age | Subject
  distance age Subject Sex
1     26.0   8     M01   M
2     25.0  10     M01   M
3     29.0  12     M01   M
4     31.0  14     M01   M
5     21.5   8     M02   M
6     22.5  10     M02   M
```

年齢  $t$  における下顎枝の長さを  $y(t)$  とし,  $a$  を性別 (女性なら  $a = 0$ , 男性なら  $a = 1$ ) を表す変数とする. このとき, 期待値  $E[y(t)]$  が変化係数  $\beta(t) = (\beta_1(t), \beta_2(t))'$  を用いて,

$$E[y(t)] = \mathbf{a}'\beta(t) = \beta_1(t) + \beta_2(t)a, \quad \beta_j(t) = \theta_j' \mathbf{x}(t), \quad (9)$$

と記述されているとする. ただし,  $\mathbf{a} = (1, a)'$  である. このとき,  $\beta_1(t)$  は女性の下顎枝の経時変化を表す変化係数,  $\beta_2(t)$  は性差の経時変化を表す変化係数である. Satoh and Yanagihara (2010) は変化係数の関数形に放物線を仮定し基底を  $\mathbf{x}(t) = (1, t, t^2)'$  とし, 未知母数  $\theta_j = (\theta_{j1}, \theta_{j2}, \theta_{j3})'$  ( $j = 1, 2$ ) の推定には個体毎の測定年齢間の相関を考慮した変量効果モデル (例えば, Laird and Ware 1982) を利用した. R では nlme パッケージの lme 関数で実行可能であり, fixef 関数と vcov 関数により  $\hat{\theta}$  と  $\Omega = \text{Cov}(\hat{\theta})$  の推定値が得られる. 式 (9) に基づく適合曲線 (太線) および観測値の折れ線 (細線) を図 1 の左図に示す.

```
> res <- lme(distance ~ (age + I(age^2)) * Sex, Orthodont, random = ~ age+I(age^2))
> (theta <- fixef(res))
      (Intercept)      age      I(age^2)      SexM      age:SexM I(age^2):SexM
17.043181818      0.542045455     -0.002840909      5.188068182     -0.874857955      0.053622159
> (Omega <- vcov(res))
      (Intercept)      age      I(age^2)      SexM      age:SexM I(age^2):SexM
(Intercept)      34.6007730     -6.38817510      0.285274321     -34.6007730      6.38817510     -0.285274321
age              -6.3881751      1.20104026     -0.054103751      6.3881751      -1.20104026      0.054103751
I(age^2)         0.2852743      -0.05410375      0.002459261     -0.2852743      0.05410375      -0.002459261
SexM            -34.6007730      6.38817510     -0.285274321      58.3888044     -10.78004549      0.481400416
age:SexM         6.3881751      -1.20104026      0.054103751     -10.7800455      2.02675544      -0.091300079
I(age^2):SexM   -0.2852743      0.05410375     -0.002459261      0.4814004     -0.09130008      0.004150004
```

測定年齢が 8 歳から 14 歳なので,  $t \in [8, 14]$  における被覆確率が 0.95 である同時信頼区間を考え

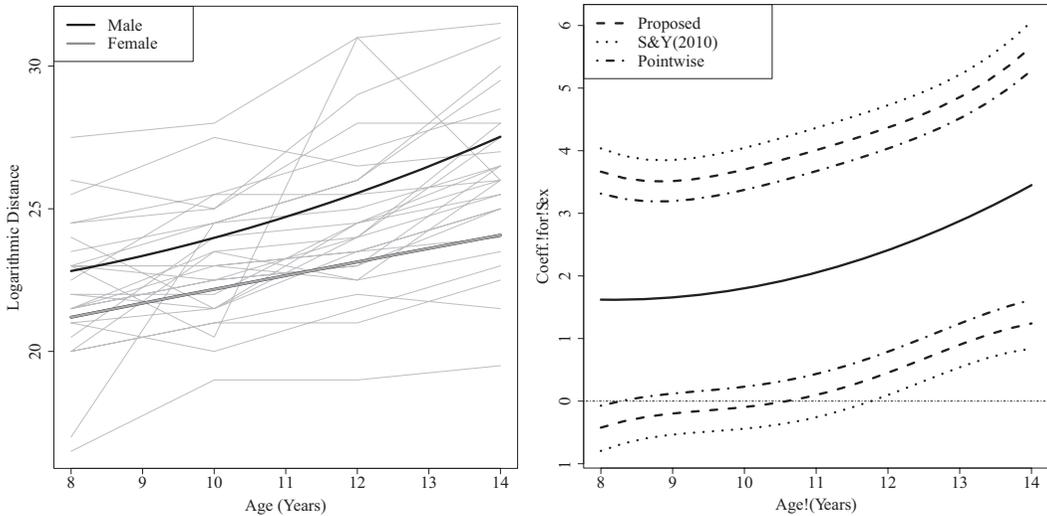


図 1. 少年・少女の下顎枝の成長データにおける適合曲線と変化係数の推定値. 左図：観測値の折れ線と適合曲線, 右図：性差の経時変化を表す変化係数.

る. ここでは, 区間  $[8, 14]$  を等間隔に分割し  $K = 20$  個の時点に離散化する. これより, `mvtnorm` パッケージの `qmvnorm` 関数を用いて, 同時信頼区間の閾値  $x_{0.05} = 2.365$  が計算される.

```
> ages <- seq(8, 14, length = 20); C <- cbind(0, 0, 0, 1, ages, ages^2)
> beta <- C %*% theta; s <- sqrt(diag(C %*% Omega %*% t(C)))
> R <- diag(1/s) %*% C %*% Omega %*% t(C) %*% diag(1/s)
> library(mvtnorm)
> x95 <- qmvnorm(p = 0.95, tail = "both.tails", corr = R)$quantile
> vc <- data.frame(Estimate = beta, lwr = beta - x95 * s, upr = beta + x95 * s)
```

比較のため, Satoh and Yanagihara (2010) が提案した  $t \in \mathbb{R}$  での同時信頼区間は次のスクリプトで計算され, その閾値は  $\sqrt{c_{3,0.05}} = 2.795$  である.

```
> c95 <- qchisq(0.95, df = 3)
> vc$g.lwr <- beta - sqrt(c95) * s; vc$g.upr <- beta + sqrt(c95) * s
```

また, カーネル平滑化における信頼区間と同様な, 時点  $t$  を固定した時の信頼区間は次のスクリプトで計算され, その閾値は  $q_{0.05/2} = 1.960$  である.

```
> q95 <- qnorm(0.975)
> vc$l.lwr <- beta - q95 * s; vc$l.upr <- beta + q95 * s
```

下記スクリプトを実行し, 性差の経時変化を表す変化係数  $\beta_1(t)$  の推定値を実線で, 提案法である  $t \in [8, 14]$  における同時信頼区間  $\mathcal{I}_{0.95}(t|2.365)$  を破線, Satoh and Yanagihara (2010) による  $t \in \mathbb{R}$  における同時信頼区間  $\mathcal{I}_{0.95}(t|2.795)$  を点線, 時点  $t$  を固定したときの各点毎の信頼区間  $\mathcal{I}_{0.95}(t|1.960)$  を一点鎖線で表したグラフで表す(図 1 の右図).

```
> matplot(ages, vc, type = "l", lty = c(1, 2, 2, 3, 3, 4, 4), lwd = 2, col=1)
> abline(h = 0, lty = 4)
> legend("topleft", lty = c(2, 3, 4), lwd = 4, legend = c("Proposed", "S&Y(2010)", "Pointwise"))
```

図1の右図より、提案法による同時信頼区間はSatoh and Yanagihara (2010)よりも狭い領域であり、より精度の高い信頼区間が得られ、提案法の有効性が確認された。また、各点毎の信頼区間は2つの同時信頼区間に比べて狭くなっていることに注意されたい。

### 3.2. マウスの体重データ

次に、Watanabe et al. (1996)による放射線に被曝した父親を持つ子供の体重の経時データ(124匹, 測定回数のべ1,232回, 一匹平均9.9回)で考える。月齢 $t$ におけるマウスの体重の対数値を $y(t)$ とし、 $a_1$ を性別(雌なら $a_1 = 0$ , 雄なら $a_1 = 1$ )、 $a_2$ を父親の被曝線量(中性子線: 0, 50, 100, 200cGy)とし、 $\mathbf{a} = (1, a_1, a_2)'$ とする。このとき、期待値 $E[y(t)]$ が変化係数 $\beta(t) = (\beta_0(t), \beta_1(t), \beta_2(t))'$ を用いて、

$$E[y(t)] = \mathbf{a}'\beta(t) = \beta_0(t) + \beta_1(t)a_1 + \beta_2(t)a_2, \quad \beta_j(t) = \theta_j' \mathbf{x}(t), \quad (10)$$

と記述されているとする。 $\beta_0(t)$ は父親の被曝線量が0cGyの雌マウスの体重の経時変化を表す変化係数、 $\beta_1(t)$ は性差の経時変化を表す変化係数、 $\beta_2(t)$ は父親の被曝線量の効果の経時変化を表す変化係数である。変化係数の関数形に放物線を仮定し基底に $\mathbf{x}(t) = (1, t, t^2)'$ を用い、未知母数 $\theta_j = (\theta_{j1}, \theta_{j2}, \theta_{j3})'$  ( $j = 1, 2, 3$ )の推定にはマウス毎の測定月齢間の相関を考慮した変量効果モデルを利用した。前節と同様に3通りの信頼区間について検討する。

式(10)に基づく適合曲線(太線)およびマウス毎の観測値の折れ線(細線)を図2の左上図に示す。雄マウス(黒色)および雌マウス(灰色)の各4本の曲線は、それぞれ上から父親の被曝線量が0, 50, 100, 200cGyにおける適合曲線である。観測月齢が2.03から13.55なので、提案法に基づく信頼区間は $t \in [2.03, 13.55]$ で構築した。図2に、変化係数 $\beta_0(t)$ 、 $\beta_1(t)$ および $\beta_2(t)$ に対する3つの信頼区間を示す。右上図が被曝線量が0cGyの雌マウスの体重の経時変化を表す変化係数 $\beta_0(t)$ 、左下図が性差の経時変化を表す変化係数 $\beta_1(t)$ 、右下図が父親の被曝線量の効果の経時変化を表す変化係数 $\beta_2(t)$ である。図2より、提案法による同時信頼区間はSatoh and Yanagihara (2010)よりも狭い領域であり、より精度の高い信頼区間が得られ、提案法の有効性が確認された。

### 4. 数値実験

佐藤・柳原・加茂(2009)と同様な設定で数値実験を行い、標本数を変えたときの同時信頼区間の実際の被覆確率について調べる。簡単のため、数値実験の設定として3.1節のデータ解析の結果を利用することを考える。 $\mathbf{x}(t) = (1, t, t^2)'$ とし、式(9)における変化係数 $\beta_j(t) = \theta_j' \mathbf{x}(t)$ の真のパラメータ値を $\theta_1 = (17.043, 0.542, -0.003)'$ 、 $\theta_2 = (5.188, -0.875, 0.054)'$ とする。この設定のもとで、 $T_2 = \sup_{t \in [8, 14]} |T_2(t)|$ 、 $T_2(t) = (\hat{\beta}_2(t) - \beta_2(t)) / \lambda_2(t)$ の分布をモンテカルロ法により求め、Satoh and Yanagihara (2010)と提案法における同時信頼区間の実際の被覆確率 $\Pr(T_2 \leq \sqrt{c_{q,\alpha}})$ および $\Pr(T_2 \leq x_\alpha)$ を求める。

ここでは、 $\alpha = 0.05$ 、男性と女性の標本数を $n_k = 10, 20, 40, 100$  ( $k = 1, 2$ )、観測時点数を $p = 4, 8, 16$ とし、モンテカルロ法により $T_2$ の分布を以下の手順で求める。1) 式(9)に基づき

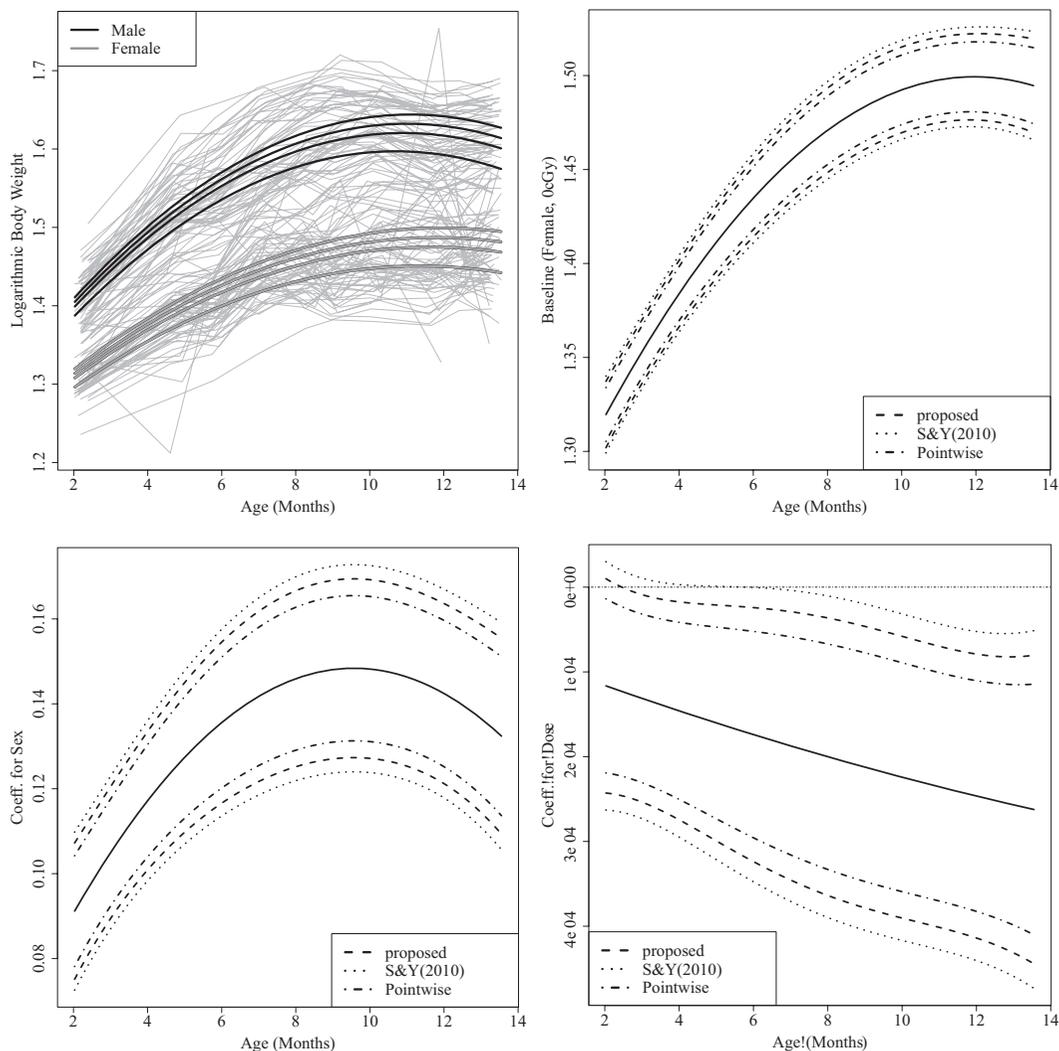


図 2. マウスの体重データにおける適合曲線と変化係数の推定値. 左上図: 観測値の折れ線および適合曲線, 右上図: 父親の被曝線量が 0cGy の雌マウスの体重の経時変化を表す変化係数, 左下図: 性差の経時変化を表す変化係数, 右下図: 父親の被曝線量の効果の経時変化を表す変化係数.

$(n_1 + n_2) \times p$  個の観測値の乱数を生成する, 2)  $\hat{\beta}_2(t)$  と  $\hat{\lambda}_2(t)$  を求める, 3)  $\mathcal{T}_2 = \sup_{t \in [8, 14]} |T_2(t)|$  を求める, 4) 1)–3) を 10,000 回繰り返し,  $\Pr(\mathcal{T}_2 \leq \sqrt{c_{q,\alpha}})$  と  $\Pr(\mathcal{T}_2 \leq x_\alpha)$  の値を評価する. なお, 1) における乱数生成には 3.1 節の変量効果モデルを用い, 観測時点は区間  $[8, 14]$  を等間隔に  $p$  個に分割した. 表 1 に, 数値実験に基づく被覆確率  $\Pr(\mathcal{T}_2 \leq \sqrt{c_{q,\alpha}})$  と  $\Pr(\mathcal{T}_2 \leq x_\alpha)$  の比較結果を示す. 表 1 より, Satoh and Yanagihara (2010) の被覆確率はすべて 95% を超えており, 標本数が増えても改善されていない. 一方, 提案法は標本数が少ないと被覆確率が 95% を下回る場合があるものの, 標本数が多い場合ではほぼ 95% に近い値となり改善されたといえる.

表 1. 数値実験による被覆確率の比較結果

$p$	$n_k$	$\Pr(\mathcal{I}_2 \leq \sqrt{c_{q,\alpha}})$	$\Pr(\mathcal{I}_2 \leq x_\alpha)$
4	10	98.0	93.1
4	20	98.0	94.4
4	40	98.3	94.8
4	100	98.4	94.9
8	10	96.8	92.7
8	20	97.9	93.9
8	40	98.3	94.6
8	100	98.7	95.0
16	10	97.4	92.8
16	20	98.3	94.1
16	40	98.6	94.6
16	100	98.8	94.9

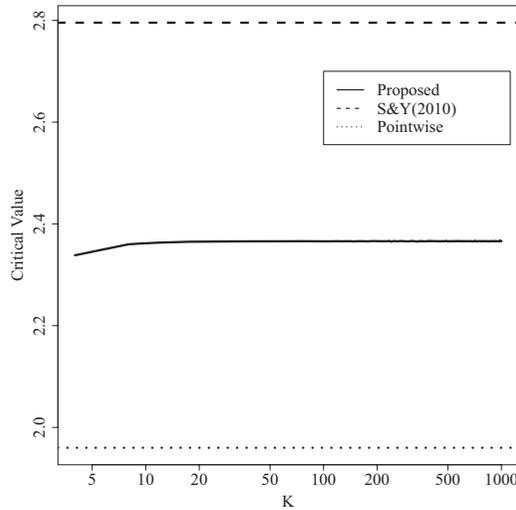


図 3. 時点数  $K$  による閾値  $x_{0.05}$  の変化. 横軸は対数軸であり, 実線が提案法で用いる閾値  $x_{0.05}$ , 破線が Satoh and Yanagihara (2010) で用いる閾値  $\sqrt{c_{3,0.05}}$ , 点線が各点毎の信頼区間での閾値  $q_{0.05/2}$  である.

## 5. おわりに

適用例および数値実験により, 連続値の経時データにおける線形な変化係数 (Satoh and Yanagihara 2010) において, 本稿で提案する同時信頼区間の有効性を示した. 提案法の有効性を理論的に示すことは今後の課題である. 本稿で提案した, 有限区間  $t \in [a, b]$  における同時信頼区間の構築法は, 線形な変化係数  $\beta(t) = \boldsymbol{\theta}'\boldsymbol{x}(t)$  における未知母数  $\boldsymbol{\theta}$  の推定量  $\hat{\boldsymbol{\theta}}$  の漸近正規性  $N_q(\boldsymbol{\theta}, \Omega)$  と共分散行列の一致推定量  $\hat{\Omega}$  に基づいている. したがって, より一般の回帰モデルで適用可能であり, 本稿の適用例で紹介した連続値の経時データのみならず, 離散値の経時データ, 空間データ, 生存時間データ等に対する回帰モデルにおいても適用可能であり, 汎用性は高いと思われる.

最後に,  $u_\alpha$  の計算にあたって,  $K$  個の時点への離散化に基づく  $K$  次元の多変量正規分布の分

位点で近似することの妥当性について検討する。3.1 節の少年・少女の下顎枝の成長データに対して、 $K$  の値を大きくしたときの  $x_\alpha$  の値を図 3 に示す。図 3 は、 $K$  の値を 4 (観測時点数) から 1000 (qmvnorm 関数で計算可能な次元数の上限) の範囲で  $x_{0.05}$  の値を実線で描いたものである。比較のため、 $t \in \mathbb{R}$  における同時信頼区間の閾値  $\sqrt{c_{3,0.05}} = 2.795$  を破線、カーネル平滑化における信頼区間と同様な時点  $t$  を固定した時の各点毎の信頼区間の閾値  $q_{0.05/2} = 1.960$  を点線で示す。図 3 より、観測時点数に比べてある程度の大きな  $K$  を用いれば近似は妥当であると思われる。

## 謝 辞

本論文の修正にあたり、建設的なご意見と有益なコメントを頂いた 2 人の査読者および編集委員の方々に深く感謝致します。この研究の一部は、文部科学省科学研究費若手 (B) 課題番号 23790694 および 23700337 の援助を受けています。

## 参 考 文 献

- Bretz, F., Hothorn, T. and Westfall, P. (2011): *Multiple Comparisons Using R*, Boca Raton, Chapman and Hall/CRC.
- Cox, D. R. (1972): Regression models and life-tables (with discussion), *J. Roy. Statist. Soc. Ser. B* **34**(2), 187–220.
- Cox, D. R. (1975): Partial likelihood, *Biometrika* **62**(2), 269–276.
- Fotheringham, A. S., Brunson, C. and Charlton, M. (2002): *Geographically Weighted Regression: the Analysis of Spatially Varying Relationship*, Chichester, Wiley.
- Genz, A. (1992): Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics* **1**, 141–150.
- Genz, A. (1993): Comparison of methods for the computation of multivariate normal probabilities, *Computing Science and Statistics* **25**, 400–405.
- Genz, A. and Bretz, F. (2002): Methods for the computation of multivariate  $t$ -probabilities. *Journal of Computational and Graphical Statistics* **11**, 950–971.
- Genz, A. and Bretz, F. (2009): *Computation of Multivariate Normal and  $t$ -Probabilities*, Springer-Verlag, Berlin Heidelberg.
- Hastie, T. and Tibshirani, R. (1993): Varying-coefficient models, *J. Roy. Statist. Soc. Ser. B* **55**(4), 757–796.
- Hothorn, T., Bretz, F. and Genz, A. (2001): On multivariate  $t$  and Gauss probabilities in R. *R Newsletter* **1**, 27–29.
- 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**(4), 809–822.
- 加茂憲一, 富田哲治, 佐藤健一 (2011): 年齢-時代平面上における癌死亡リスクの視覚化, *統計数理* 59(2), 217–237.
- 加藤直広, 栗木哲 (2012): 2 次多項式回帰曲線の正値性検定, *応用統計学* 41(1), 1–15.
- Laird, N. M. and Ware, J. H. (1982): Random-effects models for longitudinal data, *Biometrika* **38**(4), 963–974.
- Liang, K. Y. and Zeger, S. L. (1986): Longitudinal data analysis using generalized linear models, *Biometrika* **73**(1), 13–22.
- Liu, W. (2010): *Simultaneous Inference for Regression*, Boca Raton, Chapman and Hall/CRC.
- Liu, W., Jamshidian, M., Zhang, Y., Bretz, F. and Han, X. (2007): Some new methods for the comparison of two linear regression models, *Journal of Statistical Planning and Inference* **137**(1), 57–67.
- Potthoff, R. F. and Roy, S. N. (1964): A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika* **51**(3&4), 313–326.
- R Core Team (2012): *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Rao, C. R. (1973): *Linear Statistical Inference and Its Applications* (2nd ed.), John Wiley & Sons, Inc.
- Satoh, K. and Ohtaki, M. (2006): Nonparametric growth curve model with local linear approximation, *Comm. Statist. Theory Methods* **35**(4), 641–648.

- 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), 1–11.
- Satoh, K. and Yanagihara, H. (2010): Estimation of varying coefficients for a growth curve model, *Amer. J. Math. Management Sci.* **30**(3&4), 243–256.
- 富田哲治, 佐藤健一, 大谷敬子, 佐藤裕哉, 丸山博文, 川上秀史, 星正治, 大瀧慈 (2010): 広島原爆, 被曝者コホートにおける, 被曝時所在地に基づく死亡危険度地図作成の試み, 長崎医学会誌 85(特集号), 185–188.
- 富田哲治, 佐藤健一, 大谷敬子, 佐藤裕哉, 丸山博文, 川上秀史, 田代聡, 星正治, 大瀧慈 (2012): 広島原爆, 被曝者コホート 1970–2010 年におけるリスク地図の推定, 広島医学 65(4), 255–258.
- Tonda, T., Satoh, K., Nakayama, T., Katanoda, K., Sobue, T. and Ohtaki, M. (2011): A nonparametric mixed-effects model for cancer mortality, *Aust. N. Z. J. Stat.* **53**(2), 247–256.
- Tonda, T., Satoh, K., Otani, K., Sato, Y., Maruyama, H., Kawakami, H., Tashiro, S., Hoshi, M. and Ohtaki, M. (2012): Investigation on circular asymmetry of geographical distribution in cancer mortality of Hiroshima atomic bomb survivors based on risk maps: analysis of spatial survival data, *Radiation and Environmental Biophysics* **51**(2), 133–141.
- 富田哲治, 佐藤健一, 中山晃志, 片野田耕太, 祖父江友孝, 大瀧 慈 (2011): 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, 統計数理 59(2), 205–215.
- 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(2&3), 59–70.
- Watanabe, H., Takahashi, T., Lee, J., Ohtaki, M., Roy, G., Ando, Y., Yamada, K., Gotoh, T., Kurisu, K., Fujimoto, N., Satow, Y. and Ito, A. (1996): Influence of paternal  $^{252}\text{Cf}$  Neutron exposure on abnormal Sperm, embryonal Lethality, and Liver tumorigenesis in  $F_1$  offspring of mice, *Japanese Journal of Cancer Research* **87**, 51–57.

(2012 年 10 月 29 日受付 2013 年 2 月 12 日最終修正 2 月 13 日採択)

著者連絡先: 〒 734–8558 広島市南区宇品東 1–1–71  
 富田 哲治 (Tel. 082–251–9788)  
 E-mail: ttetsuji@pu-hiroshima.ac.jp

# Improvement of Confidence Interval for Linear Varying Coefficient

Tetsuji Tonda<sup>1,\*</sup> and Kenichi Satoh<sup>2</sup>

<sup>1</sup> Faculty of Management and Information Systems, Prefectural University of Hiroshima

<sup>2</sup> Research Institute for Radiation Biology and Medicine, Hiroshima University

## Abstract

Varying coefficient, which might be varying on time, can be used for visualizations or interpretations of covariate effects. In general, varying coefficient is usually estimated by kernel smoothing method, which is essentially repetition of local multiple regression for data around the fixed time. Therefore, a pointwise confidence interval is constructed for varying coefficient. Recently, Satoh and Yanagihara (2010) developed statistical inference for linear varying coefficient. They proposed a simultaneous confidence interval for  $t \in \mathbb{R}$  as a function of time. In this paper, we construct a simultaneous confidence interval for finite interval of time  $t \in [a, b]$  and improve an accuracy of confidence interval for varying coefficient. The proposed method can be applied for general regression model, and can be easily implemented by statistical software R. We compare the proposed confidence interval with that of Satoh and Yanagihara (2010) using longitudinal data in Potthoff and Roy (1964) and Watanabe et al. (1996).

**Key words:** linear varying coefficient, longitudinal data, pointwise confidence interval, simultaneous confidence interval.

\*Corresponding author

E-mail address: ttetsuji@pu-hiroshima.ac.jp (Tetsuji Tonda)

Received October 29, 2012; Received in final form February 12, 2013; Accepted February 13, 2013.