

時間や変数空間上で変化する回帰係数について

佐藤 健一

広島大学原爆放射線医科学研究所

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。横軸に時間をとれば、変化係数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, Satoh and Ohtaki (2006), Tonda *et al.* (2011)). すなわち、固定された時間近傍ごと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連続的な関数の推定とする。

しかし、この方法では各点ごと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意味での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で記述できる関数族、例えば、Satoh, Yanagihara and Ohtaki (2003) で議論された直線、放物線および B スプラインなどに限定することで、Rao (1973) で示された多変量解析の理論を利用して変化係数曲線としての信頼区間を構成することを提案した。

この結果はより広い回帰に容易に拡張できる。変化係数の関数形を線形な基底に限定することは、回帰分析において説明変数と時間変数の交互作用の項を考えることに等しく、逆に時間変数との交互作用を考えている説明変数があれば時間変数の項を回帰係数としてまとめることで変化係数として解釈できる。簡単な例として説明変数 a の回帰係数 β が時間 t とともに変化し $\beta(t)$ 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。例えば、変化係数が時間について直線であれば $\beta(t) = b_1 + b_2t$ とかける。したがって、 $\beta(t)a = b_1a + b_2ta$ となり、交互作用項を考えることに等しい。それゆえ、変化係数の推定値は a および ta を説明変数とする回帰により得られる。逆に、交互作用項などを説明変数 a でまとめれば、 $b_1a + b_2ta = (b_1 + b_2t)a$ となり、 $b_1 + b_2t$ は説明変数 a の時間とともに変化する効果を表す変化係数として解釈可能となる。変化係数内の個別のパラメータ b_1 および b_2 の推定値を検討することも重要ではあるが、変化係数曲線として要約することで説明変数の効果の解釈は容易なものになる。

本稿では、交互作用を用いた変化係数の推定方法とその適用例を統計解析ソフト R (ver. 2.10.1) で実行可能なスクリプトとともに紹介していく。まず、線形重回帰モデルにおける変化係数の推定例を紹介する。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 として格納されている。

```
> library(nlme)
> head(Orthodont)
Grouped Data: distance ~ age | Subject
  distance age Subject  Sex
1     26.0   8     M01 Male
2     25.0  10     M01 Male
3     29.0  12     M01 Male
4     31.0  14     M01 Male
5     21.5   8     M02 Male
6     22.5  10     M02 Male
```

今、14 歳のときの顎枝の長さの性差に関心があるとすれば、少女なら 1、少年なら 0 をとる説明変数 (SexFemale) と定数項を用いて単回帰を行えばよい。

```

> Orthodont14 <- Orthodont[Orthodont$age==14,]
> res <- lm(distance~1+Sex, data=Orthodont14)
> summary(res)$coefficients
              Estimate Std. Error  t value    Pr(>|t|)
(Intercept) 27.468750   0.5582192 49.207821 2.081269e-26
SexFemale   -3.377841   0.8745614 -3.862326 7.050003e-04

```

その結果、少年に比べて少女の方が 3.4mm ほど短いことが分かる。なお、この結果は 2 群の平均の差の検定を等分散のもとで行うことに等しい。この例のように経時測定データの解析は最終時点での観測値のみを抽出し簡易な統計解析で代用されることが多い。さらに途中経過における性差にも関心があれば、4 つの観測時点それぞれにおいて単回帰を繰り返すことで横軸に年齢、縦軸に性差の変化の折れ線グラフを描ける。その延長上に時間とともに連続的に変化する性差を変化係数として表現することが考えられる。

平滑化の手法により観測時点でない年齢においても性差の評価を試みる。具体的には 8 歳から 14 歳まで 1 歳間隔で局所単回帰を繰り返しながら性差を推定していく。性差を直接図示することもできるが、ここでは少年の長さに性差を加えて少女の長さとして推定結果を図 1(左) に示す。

```

> t <- seq(8,14,1); b1 <- 0*t; b2 <- b1
> for(i in 1:length(t)){
+   w <- dnorm(Orthodont$age, mean=t[i], sd=1)
+   res <-lm(distance~1+Sex, weights=w, data=Orthodont)
+   b1[i] <- res$coeff[1]; b2[i] <- res$coeff[2]}
> plot(Orthodont$age, Orthodont$distance)
> lines(t,b1); lines(t,b1+b2,lty=2)

```

一方で変化係数を直線に限定すれば局所単回帰を利用することなく、線形重回帰により大域的に推定可能である。説明変数として必要なのは、定数項、少女を示す変数 (SexFemale)、年齢 (age)、そしてその積 (SexFemale:age) である。

```

> res <-lm(distance~(1+Sex)*(1+age), data=Orthodont)
> (b <- res$coefficients)
(Intercept)    SexFemale      age SexFemale:age
 16.3406250    1.0321023    0.7843750   -0.3048295
> b1 <- b[1]+b[3]*t; b2 <- b[2]+b[4]*t
> plot(Orthodont$age, Orthodont$distance)
> lines(t,b1); lines(t,b1+b2,lty=2)

```

推定結果を図 1(右) に示す。少年・少女ともに局所単回帰の推定結果と類似しているのが分かる。また線形重回帰の回帰係数の推定結果から、性差は年齢とともに変化する変化係数として $1.03 - 0.31 \times \text{age}$ と評価できる。特に年齢が 14 歳の場合には -3.2mm となり前述の -3.4mm に近い値となっている。このように変化係数を交互作用項として表現すれば線形重回帰の枠組みを用いて容易に推定できる。

変化係数の関数形を線形な基底で記述できるものに限定しているが、その柔軟性は基底の項数を増やせずことで高めることも可能である。例えば、変化係数に直線より 2 次の項を一つ多く含む放物線のほうが曲線としての自由度は高い。ここでは、変数減少法により年齢の 1 次の項が不要と判断された。その結果、性差は $-0.54 - 0.01 \times \text{age}^2$ と評価でき、年齢が 14 歳の時に -3.3mm と計算できる。

```

> res <-lm(distance~(1+Sex)*(1+age+I(age^2)), data=Orthodont)
> res <- step(res, trace=F)

```

```
> (b <- res$coefficients)
      (Intercept)      SexFemale      I(age^2) SexFemale:I(age^2)
      20.46390086     -0.54227077      0.03575277     -0.01411708
```

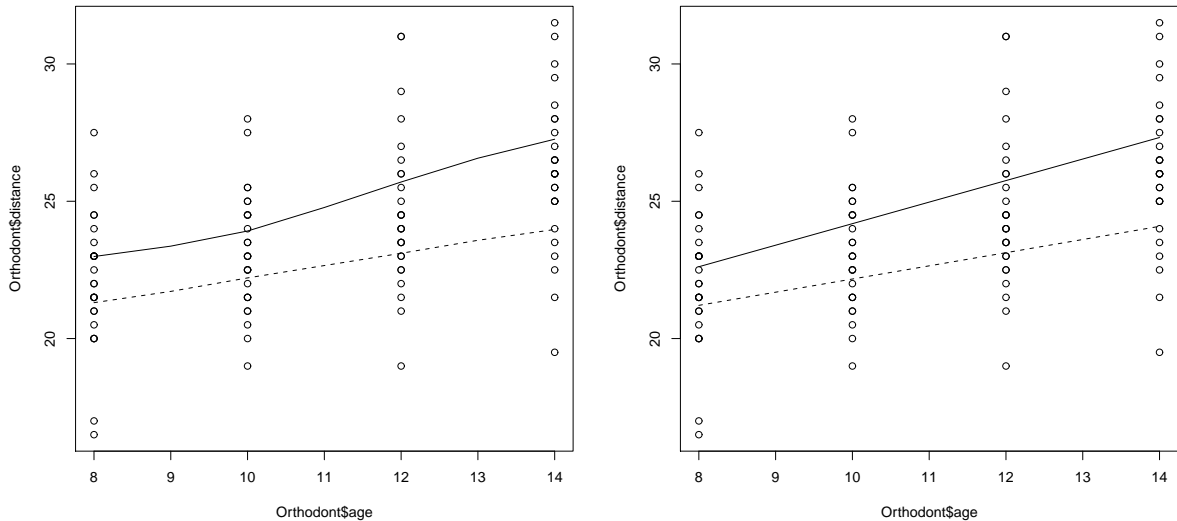


図 1: 下顎枝の長さの経時変化. (左) 局所単回帰, (右) 交互作用による線形重回帰. 実線が少年を破線が少女を表す.

経時測定データでは測定時点間の相関の存在を指摘されることが多い. 線形重回帰では個体ごとの4時点のデータは独立な観測値として扱われていたが, Liang and Zeger (1986) の提案した一般化推定方程式の枠組みを利用することで個体ごとの観測値間の相関を仮定したもとの回帰係数が推定できる. なお, 観測値間の相関構造としては自己回帰型など様々な構造を仮定できるが, ここでは特別な構造を仮定しなかった. 一般化推定方程式では適切なリンク関数を用いることで離散分布の回帰も可能である. 例えば, 佐藤・柳原・加茂 (2009) では Koch *et al.* (1977) が紹介した3時点で観測された2値の経時測定データに対して, オッズ比を線形な変化係数によって推定し, 曲線に対する信頼区間を導出している.

```
> library(gee)
> res <- gee(distance~(1+Sex)*(1+age), data=Orthodont,
+           id=Subject, corstr="unstructured")
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
      (Intercept)      SexFemale      age SexFemale:age
      16.3406250      1.0321023      0.7843750     -0.3048295
```

一般的な経時測定データにおいては個体ごとの測定時点が揃っていない場合も多い. その場合は Laird and Ware (1982) の提案した変量効果モデルを利用して変化係数を求めることができる. ここでは, ベースとなる少年の変化係数の切片と傾きに対してランダム効果を仮定した.

```

> library(nlme)
> res <- lme(distance~(1+Sex)*(1+age),
+           random=~1+age|Subject, data=Orthodont)
> res$coefficients$fixed
(Intercept)      SexFemale          age SexFemale:age
 16.3406250      1.0321023      0.7843750      -0.3048295

```

ここまでは、時間軸方向に対して変化する回帰係数を変化係数として扱ってきたが、Hastie and Tibshirani (1993) の定義では、単に複数の変数で張られる空間上で変化する回帰係数を変化係数としている。そこで、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、Brunsdon et al. (1996)) と対比させながら線形な変化係数の推測を提案した。位置情報を利用した変化係数の推定例として空間データ解析のパッケージ `spgwr` に収録されたデータ `columbus` を用いる。データは 1980 年のオハイオ州コロンバス市の 49 の地域における住居侵入窃盗および車両窃盗に関する千世帯あたりの犯罪発生率 (`crime`)、平均収入 (`income`)、平均家賃 (`housing`) および地域の位置情報 (`x`, `y`) で構成される。位置情報を利用しない場合の重回帰を行うと犯罪発生率に対する平均家賃の効果は -0.27 と推定される。

```

> library(spgwr)
> data(columbus); head(columbus)
  crime income housing      x      y
1 18.802 21.232 44.567 35.62 42.38
2 32.388  4.477 33.200 36.50 40.52
3 38.426 11.337 37.125 36.71 38.71
4  0.178  8.438 75.000 33.36 38.41
5 15.726 19.531 80.467 38.80 44.07
6 30.627 15.956 26.350 39.82 41.18
> res <- lm(crime~1+income+housing, data=columbus)
> res$coefficients
(Intercept)      income      housing
68.6188627    -1.5973038    -0.2739312

```

一方で位置によって平均家賃の効果が変わるとして線形重回帰によって線形な変化係数を求めれば、図 2(左) のように表現できる。前述の平均家賃の効果 -0.27 は中央付近を通る等高線に対応する。富田・佐藤・柳原 (2010) では、効果の空間的な一様性に関する検定も提案している。

```

> res <- lm(crime~(1+x)*(1+y)*(1+income+housing), data=columbus)
> b <- res$coefficients[c(5,8,10,12)]
> x <- columbus$x; y <- columbus$y
> b.xy <- b[1]+b[2]*x+b[3]*y+b[4]*x*y
> library(akima)
> res <- interp(x, y, b.xy)
> plot(x,y); contour(res, add=T, labcex=1)

```

さらに、富田・佐藤・大谷ら (2010) は広島原爆被爆者の生存時間解析として位置情報を利用した Cox の比例ハザード回帰 (例えば、Kleinbaum and Klein (1995)) を考え、死亡リスクが従来の爆心地からの距離に依存するだけでなく、被爆位置によっても変化することを示した。また、加茂・富田・佐藤 (2011) で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ている。以

下に Epi パッケージに含まれるデンマークの男性の肺癌死亡数のデータ (lungDK) に対するポアソン回帰による解析例を示す. その結果, 死亡リスクは図 2(右) のように視覚化でき, 近年では高齢化にともなう死亡リスクが上昇していることが分かる.

```

> library(Epi); data(lungDK)
> a <- lungDK$A5; p <- lungDK$P5; head(lungDK)
  A5  P5  C5 up      Ax      Px      Cx D      Y
1 40 1943 1898 1 43.33333 1944.667 1901.333 52 336233.8
2 40 1943 1903 0 41.66667 1946.333 1904.667 28 357812.7
3 40 1948 1903 1 43.33333 1949.667 1906.333 51 363783.7
4 40 1948 1908 0 41.66667 1951.333 1909.667 30 390985.8
5 40 1953 1908 1 43.33333 1954.667 1911.333 50 391925.3
6 40 1953 1913 0 41.66667 1956.333 1914.667 23 377515.3
> res <- glm(D~offset(log(Y))+(a+I(a^2))*(p+I(p^2)),
+           family=poisson, data=lungDK)
> (b <- res$coefficients)
  (Intercept)      a      I(a^2)      p      I(p^2)
9.676628e+03 -3.342031e+02 1.500330e+00 -9.815017e+00 2.482855e-03
  a:p      a:I(p^2)      I(a^2):p I(a^2):I(p^2)
3.394524e-01 -8.606751e-05 -1.536293e-03 3.923490e-07
> RR <- 100000*res$fitted.values/lungDK$Y
> library(akima)
> res <- interp(a, p, RR, duplicate="mean")
> plot(a, p); contour(res, add=T, labcex=1)

```

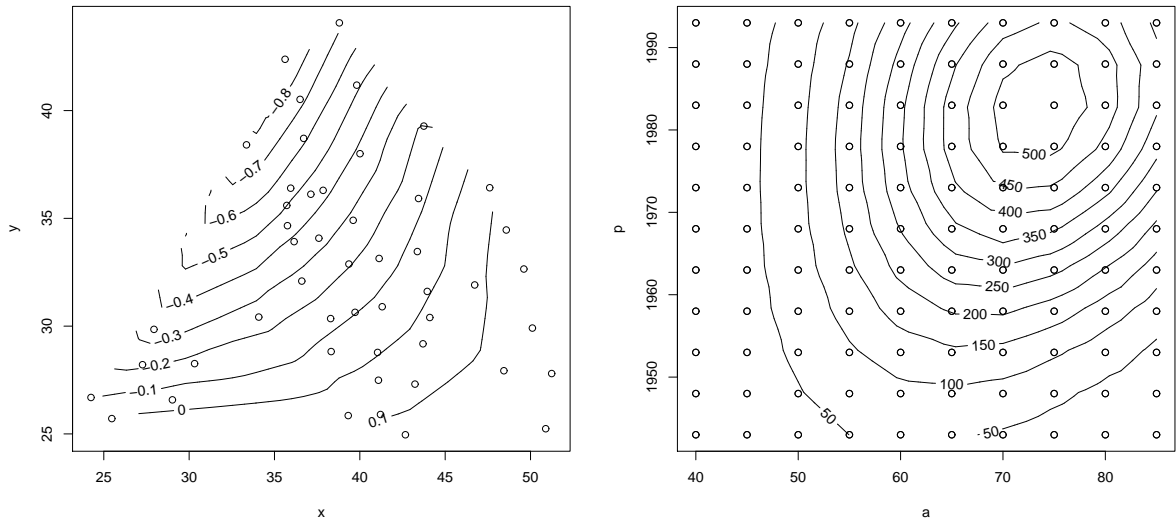


図 2: 変数空間上の変化係数の推定結果. (左) コロンバス市の 49 の地域における犯罪発生率に対する平均家価格の効果の地理分布. (右) デンマークの男性における年齢-時代空間上の肺癌死亡リスク (10 万人当たりの推定死亡数).

Brunsdon, C., Fotheringham, A. S. and Charlton, M. (1996): Geographically weighted regression: a method for exploring spatial non-stationarity, *Geographical Analysis* **28**, 281?-298.

- Hastie, T. and Tibshirani, R. (1993): Varying-coefficient models, *Journals of the Royal Statistical Society Series B* **55**, 757–796.
- 加茂憲一, 富田哲治, 佐藤健一. (2011): 交互作用モデルに基づく年齢・時代・コホート効果の検証, 第21回日本疫学会学術総会講演集, **21**, 213.
- Kleinbaum, D. G. and Klein, M. (2005): *Survival Analysis: A Self-Learning Text (Statistics for Biology and Health)* (2nd ed.), Springer.
- Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H. and Lehnen, R. G. (1977): A general methodology for analysis of experiments with repeated measurements of categorical data, *Biometrics* **33**, 133–158.
- Laird, N. M. and Ware, J. H. (1982): Random-effects models for longitudinal data, *Biometrika* **38**, 963–974.
- Liang, K. Y. and Zeger, S. L. (1986): Longitudinal data analysis using generalized linear models, *Biometrika* **73**, 13–22.
- Potthoff, R. F. and Roy, S. N. (1964): A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika* **51**, 313–326.
- Rao, C. R. (1973): *Linear Statistical Inference and Its Applications* (2nd ed.), John Wiley.
- Satoh, K., Yanagihara, H. and Ohtaki, M. (2003): Bridging the gap between B-spline and polynomial regression model, *Communications in Statistics - Computation and Simulation* **32**, 179–190.
- Satoh, K. and Ohtaki, M. (2006): Nonparametric growth curve model with local linear approximation, *Communications in Statistics - Theory and Methods* **35**, 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**, 19–29.
- Satoh, K. and Yanagihara, H. (2010): Estimation of varying coefficients for a growth curve model, *American Journal of Mathematical and Management Sciences*, in press.
- 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.
- 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): 広島原爆被爆者コホートにおける被爆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**, 185–188.
- Tonda, T., Satoh, K., Nakayama, T., Katanoda, K., Sobue, T. and Ohtaki, M. (2011): A non-parametric mixed-effects model for cancer mortality, *Australian and New Zealand Journal of Statistics*, in press.

〒734-8551 広島市南区霞 1-2-3 総合研究棟 4F 計量生物
佐藤健一 (Tel. 082-257-5857)
E-mail: ksatoh@hiroshima-u.ac.jp
<http://home.hiroshima-u.ac.jp/ksatoh/>