関数型説明変数を伴う混合効果モデルにおける平滑化パラメータの選択

大阪府立大学 大学院工学研究科　道家悠太
大阪府立大学 大学院工学研究科　林　利治

第 1 節　序論

本研究では関数型線形混合効果モデル（functional linear mixed effect model, FLMM）において、勾配関数に関する平滑化パラメータを、データに応じて適応的に選択できることを例示する。

FLMM は混合効果モデル（LME）を基に、説明変数および固定効果と変量効果の勾配を関数へと一般化したモデルである。これによって、FLMM は目的変数と関数型説明変数（共変量の関数として表される説明変数）との関係を、変量効果（各観測群の特徴）を考慮しながら説明することができる。

FLMM はモデルパラメータの推定方法として REML-based EM アルゴリズムを用いる（Liu et al. (2017))。また、勾配関数（関数型の勾配）を推定する際は、LME と同様に、データへの過度な適合を避けるため罰則を課す。罰則の程度は平滑化パラメータにより定められ、その選択は、勾配関数やその他のモデルパラメータの推定精度に大きな影響を与える。

経時データの解析が行える LME は、これまでにも盛んに研究されてきた（Staniswalis and Lee (1998), Wu and Zhang (2006))。また経時データと同様に、離散的に観測される説明変数を共変量の関数として捉え、目的変数との関係を解析するモデル、関数型線形モデル（functional linear model, FLM）も研究されてきた（Guo (2002), Yao et al. (2003), Ramsay and Silverman (2005), Goldsmith et al. (2012))。FLMM は、LME と FLM のいずれの特徴も備えたモデルとして、Liu et al. (2017）が提案したモデルである。彼らは FLMM において、観測群の個数や観測回数、そして観測誤差の分散によるパラメータの推定精度への影響に関する研究をした。しかし、彼らは平滑化パラメータの選択方法に関して詳細には議論しておらず、また、彼らの手法では、データに対して適応的に平滑化パラメータを選択することが困難であると思われる。

そこで本研究では、FLMM を用いた実データの解析を念頭に置き、①平滑化パラメータをデータに対して適応的に選択する方法、②離散的な観測から関数型説明変数を近似する際の精度、について数値実験を通して研究する。主な結果として、推奨する手法（Wu and Zhang (2006）の BIC 規準を書き換えたもの）を用いれば、固定効果の勾配関数に関する平滑化パラメータを、データに応じて適応的に選択できることを例示する。

本論文の第 2 節では、FLMM に関する説明を行い、さらにはモデルパラメータ（勾配関数を含む）の推定手順を簡潔に紹介する。第 3 節では、離散的な観測からの関数型説明変数の近似に関する研究を、第 4 節では、平滑化パラメータの選択方法の比較研究を述べる。

第 2 節　関数型線形混合効果モデル

2-1 関数型線形混合効果モデル

この節では、本研究の対象となる関数型線形混合効果モデル（FLMM）に関する説明を行う。混合効果モデル（LME）は下の（2.1.1）で定義される（詳細は Staniswalis and Lee (1998), Guo (2002), Wu and Zhang (2006), 船渡川, 船渡川（2015）等を参照）

$$Y_i = (\alpha + \gamma_i) + \sum_{t=1}^{K} (\beta_t + b_{it}) X_{it} + \varepsilon_i \quad (i = 1,2,\dots,n) \tag{2.1.1}$$

ここで、LME (2.1.1) の説明変数 X_{it} $(t = 1,2,\dots,K)$ を共変量 $t \in [0,T]$ の関数 $X_i(t)$ に一般化する。同時に固定効果の勾配 β_t と変量効果の勾配 b_{it} も共変量の関数 $\beta(t)$, $b_i(t)$ とする。また、第 i 群では m_i 回観測が行われたとする $(i = 1,2,\dots,n)$。このとき、第 i 群に属する目的変数の j 番目の観測を Y_{ij}, 関数型の説明変数を $X_{ij}(t)$ と書き表す $(j = 1,2,\dots,m_i)$ 。以上の書き換えに伴い（2.1.1）は

$$Y_{ij} = \left(\alpha + \gamma_i\right) + \int_0^T \left(\beta(t) + b_i(t)\right) X_{ij}(t)\, dt + \varepsilon_{ij} \tag{2.1.2}$$

となる。（2.1.2）の誤差項 ε_{ij} は、$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ と仮定する。モデル（2.1.2）は関数型線形混合効果モデル（FLMM）と呼ばれる。

Liu et al. (2017) が用いた FLMM では、ベクトル化された LME と同様に、スカラーの固定効果と変量効果パラメータは $\boldsymbol{\alpha}$ (u 次元), $\boldsymbol{\gamma}_i$ (v 次元) とベクトル化され、それに伴い、$\boldsymbol{\alpha}$, $\boldsymbol{\gamma}_i$ に対応する説明変数もそれぞれ $\boldsymbol{W}_{ij}$, $\boldsymbol{Z}_{ij}$ とベクトル化されている。また、彼らは関数型説明変数 $\boldsymbol{X}_{ij}(t) = \left(X_{ij1}(t), \dots, X_{ijd}(t)\right)^t$ と勾配関数 $\boldsymbol{\beta}(t) = \left(\beta_1(t), \dots, \beta_d(t)\right)^t$, $\boldsymbol{b}_i(t) = \left(b_{i1}(t), \dots, b_{id}(t)\right)^t$ を d 次元化させ、それに伴い FLMM は

$$Y_{ij} = \left(\boldsymbol{W}_{ij}^t \boldsymbol{\alpha} + \boldsymbol{Z}_{ij}^t \boldsymbol{\gamma}_i\right) + \sum_{l=1}^{d} \left\{ \int_0^T \left(\beta_l(t) + b_{il}(t)\right) X_{ijl}(t)\, dt \right\} + \varepsilon_{ij} \tag{2.1.3}$$

となる。これを Liu et al. (2017) は用いた。なお、変量効果パラメータ γ_i, $b_{il}(t)$と誤差 ε_{ij} の分布については後述する。

彼らは勾配関数 $\beta_l(t)$, $b_{il}(t)$を基底関数 $\boldsymbol{\phi}(t) = \left(\phi_1(t), \dots, \phi_{F_1}(t)\right)^t$, $\boldsymbol{\psi}(t) = \left(\psi_1(t), \dots, \psi_{F_2}(t)\right)^t$ を用いて $\beta_l(t) = \sum_{f_1=1}^{F_1} c_{f_1 l} \phi_{f_1}(t) = \boldsymbol{\phi}(t)^t \boldsymbol{c}_l$, $b_{il}(t) = \sum_{f_2=1}^{F_2} b_{i f_2 l} \psi_{f_2}(t) = \boldsymbol{\psi}(t)^t \boldsymbol{b}_{il}$ と近似した。ただし、$\boldsymbol{c}_l = \left(c_{1l}, \dots, c_{F_1 l}\right)^t$, $\boldsymbol{b}_{il} = \left(b_{i1l}, \dots, b_{iF_2 l}\right)^t$ である。

これに伴い、FLMM (2.1.3) は

$$Y_{ij} = \left(\boldsymbol{W}_{ij}^t \boldsymbol{\alpha} + \boldsymbol{Z}_{ij}^t \boldsymbol{\gamma}_i\right) + \sum_{l=1}^{d} \left\{ B_{ijl}^t \boldsymbol{c}_l + C_{ijl}^t \boldsymbol{b}_{il} \right\} + \varepsilon_{ij}$$

と書ける。ただし、$B_{ijl}=\int_0^T X_{ijl}(t)\boldsymbol{\phi}(t)\,dt,\quad C_{ijl}=\int_0^T X_{ijl}(t)\boldsymbol{\psi}(t)\,dt$ とおいた。また、σ_ε^2, $\boldsymbol{D}_0$, $\boldsymbol{D}_l\quad(l=1,2,\dots,d)$ を分散成分として、$\boldsymbol{\gamma}_i\sim N(0,\sigma_\varepsilon^2\boldsymbol{D}_0)$, $\boldsymbol{b}_{il}\sim N(0,\sigma_\varepsilon^2\boldsymbol{D}_l)$, $\varepsilon_{ij}\sim N(0,\sigma_\varepsilon^2)$ とする。そして、$\boldsymbol{\theta}=\left(\boldsymbol{\alpha}^t,\boldsymbol{c}_1^t,\dots,\boldsymbol{c}_d^t\right)^t$, $\boldsymbol{\vartheta}_i=\left(\boldsymbol{\gamma}_i{}^t,\boldsymbol{b}_{i1}^t,\dots,\boldsymbol{b}_{id}^t\right)^t$, $\widetilde{\boldsymbol{W}}_i=\left(\widetilde{W}_{i1},\dots,\widetilde{W}_{im_i}\right)^t$, $\widetilde{\boldsymbol{Z}}_i=\left(\tilde{Z}_{i1},\dots,\tilde{Z}_{im_i}\right)^t$ とおくと、(2.1.3) は、以下のように書き直される。

$$\boldsymbol{Y}_i=\widetilde{\boldsymbol{W}}_i\boldsymbol{\theta}+\widetilde{\boldsymbol{Z}}_i\boldsymbol{\vartheta}_i+\boldsymbol{\varepsilon}_i \tag{2.1.4}$$

なお、上式において $\widetilde{W}_i=\left(\boldsymbol{W}_{ij}^t,B_{ij1}^t,\dots,B_{ijd}^t\right)^t,\quad \tilde{Z}_i=\left(\boldsymbol{Z}_{ij}^t,C_{ijl}^t,\dots,C_{ijd}^t\right)^t$, $\boldsymbol{Y}_i=\left(Y_{i1},Y_{i2},\dots,Y_{im_i}\right)^t$, $\boldsymbol{\varepsilon}_i=\left(\varepsilon_{i1},\dots,\varepsilon_{im_i}\right)^t$ である。

2-2 平滑化パラメータと罰則項

FLMM のパラメータは REML−based EM アルゴリズム（詳細は Liu et al. (2017) 参照）を用いて推定する。推定の手順は LME のパラメータ推定の手順とおおよそ同じである。しかし、FLMM では推定対象に関数型のパラメータがあることから、尤度関数に罰則項を付加したものを最適化関数として扱う。Liu et al. (2017) は $\beta_l(t)$, $b_{il}(t)$ に対する罰則項をそれぞれ

$$\int_0^T\left(\frac{d^2\beta_l(t)}{dt^2}\right)^2dt=\boldsymbol{c}_l^tG_\beta\boldsymbol{c}_l\quad,\qquad\int_0^T\left(\frac{d^2b_{il}(t)}{dt^2}\right)^2dt=\boldsymbol{b}_{il}^tG_b\boldsymbol{b}_{il}$$

と定めた。ただし、$G_{\boldsymbol{\beta}}=\int_0^T\left(\frac{d^2\boldsymbol{\phi}(t)}{dt^2}\right)\left(\frac{d^2\boldsymbol{\phi}(t)}{dt^2}\right)^t dt$, $G_{\boldsymbol{b}}=\int_0^T\left(\frac{d^2\boldsymbol{\psi}(t)}{dt^2}\right)\left(\frac{d^2\boldsymbol{\psi}(t)}{dt^2}\right)^t dt$ である。そして、各罰則項 $\boldsymbol{c}_l^tG_{\boldsymbol{\beta}}\boldsymbol{c}_l$, $\boldsymbol{b}_{il}^tG_{\boldsymbol{b}}\boldsymbol{b}_{il}$ に重み λ_l^β, $\lambda_l^b\quad(l=1,2,\dots,d)$ をつけ、対数尤度関数に加えた。この重み λ_l^β, λ_l^b が平滑化パラメータである。平滑化パラメータの適切な値を選択するための方法が、本研究の主題である。その方法に関しては、後の第 4 節で議論する。

ここまでの議論より、最適化関数 $H(\boldsymbol{\theta},\boldsymbol{\vartheta})$ は以下で定められる（Liu et al. (2017)）。

$$\begin{aligned}H(\boldsymbol{\theta},\boldsymbol{\vartheta})=&\sum_{i=1}^n\sum_{j=1}^{m_i}\frac{1}{2\sigma_\varepsilon^2}\left[Y_{ij}-\boldsymbol{W}_{ij}^t\boldsymbol{\alpha}-\boldsymbol{Z}_{ij}^t\boldsymbol{\gamma}_i-\sum_{l=1}^d\left(\boldsymbol{B}_{ijl}^t\boldsymbol{c}_l+\boldsymbol{C}_{ijl}^t\boldsymbol{b}_{il}\right)\right]^2\\&+\sum_{l=1}^d\left[\frac{\lambda_l^\beta}{2}\int_0^T\left\{\frac{d^2\beta_l(t)}{dt^2}\right\}^2dt+\frac{\lambda_l^b}{2\sigma_\varepsilon^2}\sum_{i=1}^n\int_0^T\left\{\frac{d^2b_{il}(t)}{dt^2}\right\}^2dt\right]+\frac{1}{2\sigma_\varepsilon^2}\sum_{l=1}^2\sum_{i=1}^n\boldsymbol{b}_{il}^tD_l^{-1}\boldsymbol{b}_{il}+\frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^n\boldsymbol{\gamma}_i^tD_0^{-1}\boldsymbol{\gamma}_i\end{aligned}\tag{2.2.1}$$

FLMM のパラメータ推定は REML−based EM アルゴリズムを用いるが、EM アルゴリズムは、推定結果が初期値設定に強く依存することが知られている（Biernacki et al. (2002), Micheal and Melnykov (2016)）。そのため、適切な初期値を定めておくことが重要である。本研究で行ったシミュレーション実験には、初期値設定を適切に定めるために、em-EM アルゴリズム（詳細は Biernacki et al. (2002) や Micheal and Melnykov (2016) 参照）を用いた。

第 3 節 離散的な観測からの関数型説明変数の近似

FLMM (2.1.2) の特徴でもある関数型説明変数 $X(t)$ は、連続的な観測 $\{X(t)\}_{t\in[0,T]}$ がふつう得られない。そこで、誤差を含む離散的な観測 $X'(t_p)=X(t_p)+\varepsilon_x(t_p)\quad(p=0,1,\dots P)$ から $\{X(t)\}_{t\in[0,T]}$ を近似する。これは離散的な観測値 $X'(t_0)$, $X'(t_1)$, … , $X'(t_P)$ を平滑化して $X(t)$

を近似することに相当する。ただし、観測誤差について $E\left[\varepsilon_x(t_p)\right]=0$, $Var\left[\varepsilon_x(t_p)\right]=\sigma_x^2$ とする。

有限個のデータを有効に用いるためにも、$X(t)$ の近似のために定めるべきパラメータの数は最小限に抑えたい。そこで、主成分分析（PCA）を基盤とした変換法、KL 変換（Staniswalis and Lee (1998), 石井 他（1998), Yao et al. (2003), Ramsay and Silverman (2005), Liu et al. (2017) 参照）を平滑化に用いる。

次の 3-1 節では、KL 変換を使った FLMM の関数型説明変数 $X(t)$ の近似方法を 2 種類説明する。そして 3-2 節では、それら 2 種類の近似方法を比較し、後述する Yao et al. (2003) が用いた方法の方が近似誤差が小さいことを数値例により示す。

3-1 FLMM 下での KL 変換

Liu et al. (2107) ほか、関数型のデータを離散値から近似する場合の多くは、固有方程式 $\int_0^T C(s,t)\varphi(t)\,dt=\lambda\varphi(s)$ を解く方針をとる（Yao et al. (2003), Ramsay and Silverman (2005), Goldsmith et al. (2012), Liu et al. (2017))。ここで、$C(s,t)=Cov(X(s),X(t))$ は共分散関数といわれ、$X(s),X(t)$ 間の共分散である。これは、離散的な観測を基に

$$C(t_i,t_j)\cong\hat{C}(t_i,t_j)=Cov\left(X'(t_i),X'(t_j)\right)\qquad(i,j=0,1,\ldots P)\tag{3.1.1}$$

と近似される。固有方程式の解として、固有値 λ, 固有関数 $\varphi(\cdot)$ が得られる。それを基に、KL 変換が行われる。詳細は Liu et al. (2017) を参照されたい。

Yao et al. (2003) と Staniswalis and Lee (1998) は、共分散関数を離散的な観測から近似する際、対角成分の値は誤差分散の大きさ分だけ増加することに着目した。

$$C(t_i,t_j)\cong\hat{C}(t_i,t_j)=Cov\left(X'(t_i),X'(t_j)\right)=Cov\left(X(t_i),X(t_j)\right)+\sigma_x^2\delta_{ij}\tag{3.1.2}$$

ただし、δ_{ij} はクロネッカーデルタである。彼らは、観測誤差の分散 σ_x^2 を $\hat{\sigma}_x^2$ と事前に推定することで、より精度の高い共分散関数の近似値 $Cov\left(X(t_i),X(t_j)\right)=Cov\left(X'(t_i),X'(t_j)\right)-\hat{\sigma}_x^2\delta_{ij}$ を得た。詳細は Yao et al. (2003) を参照されたい。

3-2 離散的な観測からの関数型説明変数の近似方法の比較

この節では、3-1 節にて紹介した 2 種類の近似方法を数値実験により比較する。以降でその実験の設定、そして結果を提示する。

Liu et al. (2017) が用いた近似手法との比較を行うために、この実験で用いる関数型説明変数は、彼らが行ったシミュレーション実験で用いたものと同一の設定にした。2 つの近似手法の比較には、評価規準として以下の RMISE (3.2.1) を用いる。これは、Liu et al. (2017) が勾配関数の推定精度の評価に用いた規準と同様である。

$$\mathrm{RMISE}\left(X_{ij}(t)\right) = \left[\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\int_0^T\left(X_{ij}^{(m)}(t) - \hat{X}_{ij}^{(m)}(t)\right)^2 dt}{\int_0^T X_{ij}^{(m)}(t)^2\, dt}\right]^{\frac{1}{2}} \tag{3.2.1}$$

ただし、$\hat{X}_{ij}^{(m)}(t)$ は m 回目（$m = 1,2,\dots,M$）の実験における関数型説明変数 $X_{ij}^{(m)}(t)$ の推定値である。また、M はシミュレーションの繰り返し数を表し、ここでは、$M = 200$と設定した。

前述した通り、実践的には連続的な観測 $X_{ij}(t)$ は得られず、離散的な観測点 t_p での誤差を含む関数値 $X'_{ij}(t_p) = X_{ij}(t_p) + \varepsilon_x(t_p)$ しか得られない。また、3-1 節で取り上げた、共分散関数を離散化する方針の下で行う近似法では、離散的な観測点 t_p（$p = 1,2,\dots,P$）における近似値しか得られない。そのため、RMISE 内の積分は近似されることになる。

RMISE 規準（3.2.1）を用いて、(a) Yao et al. (2003) が用いた近似法と (b) Liu et al. (2017) が用いた近似法とを比較した。

表 1 RMISEの比較結果

近似法	(a)	(b)
$\mathrm{RMISE}\left(\hat{X}_{ij}(t)\right)$	0.1573	0.5687

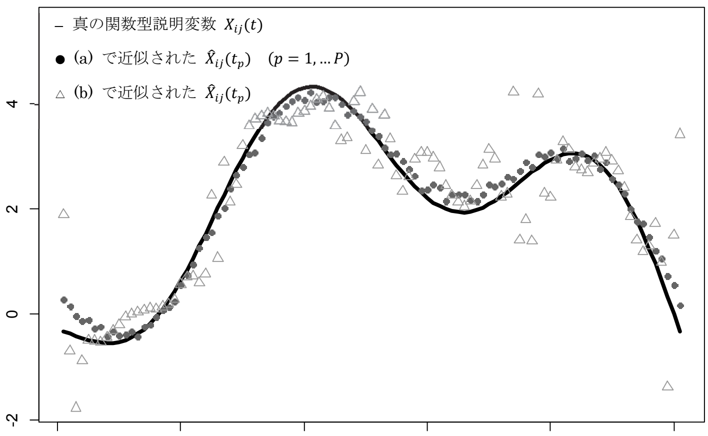

図 1 $\hat{X}_{ij}(t)$ の比較

以上の結果から、視覚的に見てもそして数値的にも Yao et al. (2003) が用いた近似法（a）の方が近似精度が高い、つまり離散化した共分散関数の対角成分を調整した上で KL 変換をする方が、関数型説明変数をより正確に近似できるといえる。この比較は Yao et al. (2003) や Ramsay and Silverman (2005) では検証されてこなかった。この結果を踏まえると、関数型説明変数を離散的な観測値から近似する際は、前者の近似法（a）を用いることが推奨される。

第 4 節 最適な平滑化パラメータの選択方法

FLMM でも、平滑化パラメータの選択はモデルパラメータの推定精度に大きな影響を及ぼす。

これまでにも、LME の形式で書けるモデルにおいて、一般化交差検証法規準や制限付き最尤法など、最適な平滑化パラメータを選択する方法の比較は行われてきた（Reiss and Odgen (2008), Aydin and Memmendli (2011))。しかし FLMM において、この議論はほとんど見られない。そこでこの節では、FLMM のモデルパラメータ推定時における最適な平滑化パラメータの選択方法について議論する。具体的には、4-1 節にて、比較対象である平滑化パラメータの選択方法を紹介し、さらに 4-2 節にて、各選択手法の比較実験について述べる。

4-1 平滑化パラメータの選択方法

4-1 節では $\boldsymbol{Y} = (\boldsymbol{Y}_1^t, \dots, \boldsymbol{Y}_n^t)^t$, $\widetilde{\boldsymbol{W}} = \left(\widetilde{\boldsymbol{W}}_1^t, \dots, \widetilde{\boldsymbol{W}}_n^t\right)^t$, $\widetilde{\boldsymbol{Z}} = \left(\widetilde{\boldsymbol{Z}}_1^t, \dots, \widetilde{\boldsymbol{Z}}_n^t\right)^t$ を用いて（2.1.4）を書き直したモデルを用いる（Liu et al. (2017))。

$$\boldsymbol{Y} = \widetilde{\boldsymbol{W}}\boldsymbol{\theta} + \widetilde{\boldsymbol{Z}}\boldsymbol{\vartheta} + \boldsymbol{\varepsilon} \tag{4.1.1}$$

ただし、$\boldsymbol{\varepsilon} \sim N\left(0, \sigma_\varepsilon^2 \widetilde{D}_\vartheta\right)$, $\widetilde{D}_\vartheta = \mathrm{diag}\left(\widetilde{D}_0, \widetilde{D}_1, \dots, \widetilde{D}_d\right)$, $\widetilde{D}_l = \left(D_l^{-1} + \lambda_l^b G_b\right)^{-1}$ $(l = 1,2, \dots, d)$ と定める。モデルの誤差項 $\boldsymbol{\varepsilon}$ は $\boldsymbol{\varepsilon} \sim N(0, \sigma_\varepsilon^2 I_N$ ）とする。$N = \sum_{i=1}^n m_i$ であり、I_N は N 次単位行列を表す。

まずは Liu et al. (2018）が用いていた、各効果の勾配関数 $\beta(t)$, $b_i(t)$ に関する平滑化パラメータ $\boldsymbol{\lambda}^\beta, \boldsymbol{\lambda}^b$ ごとの選択方法を紹介する。

固定効果の勾配関数に関する平滑化パラメータ $\boldsymbol{\lambda}^\beta$ の選択は一般化交差検証規準（GCV 規準）を用いて行われる（Liu et al. (2017))。この GCV 規準は、LME 形式で書き表せるモデルに関する研究にて広く用いられてきた規準である（Ramsay and Silverman (2005), Wu and Zhang (2006), Reiss and Odgen (2008), Aydin and Memmendeli (2011))。GCV 規準は以下のように定められる。

$$\mathrm{GCV}\left(\boldsymbol{\lambda}^\beta\right) = \frac{\mathrm{SSE}\left(\boldsymbol{\lambda}^\beta\right)}{\mathrm{tr}\{I_N - S(\boldsymbol{\lambda}^\beta)\}^2} \tag{4.1.2}$$

ここで、$\mathrm{SSE}\left(\boldsymbol{\lambda}^\beta\right) = \sum_{i=1}^n \left\|\boldsymbol{Y} - \widetilde{\boldsymbol{W}}_i\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{Z}}_i\widehat{\boldsymbol{\vartheta}}_i\ \right\|^2$, $S\left(\boldsymbol{\lambda}^\beta\right) = \widetilde{\boldsymbol{W}}\left(\widetilde{\boldsymbol{W}}^t\widetilde{\boldsymbol{W}} + \tilde{G}_\theta\right)^{-1}\widetilde{\boldsymbol{W}}^t$ である。また、$\tilde{G}_\theta = \mathrm{diag}\left(\boldsymbol{0}_{u\times u}, \mathrm{diag}\left(\lambda_1^\beta, \dots, \lambda_d^\beta\right) \otimes\ G_\beta\right)$ である（$\otimes$ はクロネッカー積を表す)。この GCV 規準を最小化する $\boldsymbol{\lambda}^\beta$ が更新値として選択される。

対して、変量効果の勾配関数に関する平滑化パラメータ $\boldsymbol{\lambda}^b$ は制限付き最尤法（REML 法）を用いて選択される（Liu et al. (2017))。REML 法も GCV 規準と同様に、LME 形式で書き表せるモデルに関する研究にて広く用いられてきた（Wu and Zhang (2006), Reiss and Odgen (2008), Aydin and Memmendeli (2011))。モデル式（4.1.1）の下で REML 法を行い、$\boldsymbol{\lambda}^b$ を選択する。具体的には、以下の制限付き対数尤度 $l_{\mathrm{REML}}(\boldsymbol{\lambda}^b)$ を最大化する $\boldsymbol{\lambda}^b$ が選択される。

$$l_{\mathrm{REML}}\left(\boldsymbol{\lambda}^b\right) \propto -\frac{1}{2}\log\left|\widetilde{\boldsymbol{V}}\right| - \frac{1}{2}\log\left|\widetilde{\boldsymbol{W}}'\widetilde{\boldsymbol{V}}^{-1}\widetilde{\boldsymbol{W}}\right| - \frac{1}{2}\left(\boldsymbol{Y} - \widetilde{\boldsymbol{W}}\widehat{\boldsymbol{\theta}}\right)^t\widetilde{\boldsymbol{V}}^{-1}\left(\boldsymbol{Y} - \widetilde{\boldsymbol{W}}\widehat{\boldsymbol{\theta}}\right) \tag{4.1.3}$$

続いて、平滑化パラメータ（$\boldsymbol{\lambda}^\beta, \boldsymbol{\lambda}^b$）を同時に選択する方法を紹介する。その選択方法の基は、Wu and Zhang (2006）にて用いられた BIC 規準である。以降で、FLMM の文脈にて使えるよう

書き換えられた BIC 規準を説明する。

始めに、FLMM における対数尤度関数 $l(\boldsymbol{\lambda}^{\beta}, \boldsymbol{\lambda}^{b})$ を定義する。制限付き対数尤度（4.1.3）とは異なることに注意されたい。

$$l(\boldsymbol{\lambda}^{\beta}, \boldsymbol{\lambda}^{b}) = -\frac{1}{2}\log|\widetilde{\boldsymbol{V}}| - \frac{1}{2}(\boldsymbol{Y} - \widetilde{\boldsymbol{W}}\widehat{\boldsymbol{\theta}})^{t}\widetilde{\boldsymbol{V}}^{-1}(\boldsymbol{Y} - \widetilde{\boldsymbol{W}}\widehat{\boldsymbol{\theta}}) - \frac{N}{2}\log(2\pi) \tag{4.1.4}$$

次に、FLMM による目的変数の予測値を各効果（固定効果と変量効果）に分解する。

$$\widehat{\boldsymbol{Y}} = \widetilde{\boldsymbol{W}}\widehat{\boldsymbol{\theta}} + \widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{\vartheta}} = \widehat{\boldsymbol{Y}}_{\theta} + \widehat{\boldsymbol{Y}}_{\vartheta} \tag{4.1.5}$$

ここで、REML–based EM アルゴリズムを用いて推定されるモデルパラメータ $\boldsymbol{\theta}, \boldsymbol{\vartheta}$ の更新値を（4.1.1）に代入すると、

$$\widehat{\boldsymbol{Y}}_{\theta} = \widetilde{\boldsymbol{W}}\widehat{\boldsymbol{\theta}} = \left[\widetilde{\boldsymbol{W}}\left\{\widetilde{\boldsymbol{W}}^{t}\ \ \widetilde{\boldsymbol{V}}^{-1}\widetilde{\boldsymbol{W}} + G_{\theta}\right\}^{-1}\widetilde{\boldsymbol{W}}^{t}\ \ \widetilde{\boldsymbol{V}}^{-1}\right]\boldsymbol{Y} = \boldsymbol{A}_{\theta}\boldsymbol{Y},$$

$$\widehat{\boldsymbol{Y}}_{\vartheta} = \widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{\vartheta}} = \widetilde{\boldsymbol{Z}}\left(\widetilde{D}_{\vartheta}\ \ \widetilde{\boldsymbol{Z}}^{T}\ \ \ \widetilde{\boldsymbol{V}}^{-1}\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{Y}\right) = \left[\widetilde{\boldsymbol{Z}}\widetilde{D}_{\vartheta}\ \ \widetilde{\boldsymbol{Z}}^{T}\ \ \ \widetilde{\boldsymbol{V}}^{-1}(\boldsymbol{I}_{N} - \boldsymbol{A}_{\boldsymbol{\theta}})\right]\boldsymbol{Y} = \boldsymbol{A}_{\boldsymbol{\vartheta}}\boldsymbol{Y}$$

と書き直せる（$\boldsymbol{\theta}, \boldsymbol{\vartheta}$ の更新値の詳細については Liu et al. (2017) を参照）。よって $\boldsymbol{A}_{\theta}, \boldsymbol{A}_{\boldsymbol{\vartheta}}$ を用いることで、(4.1.1) は以下の様に変形できる。

$$\widehat{\boldsymbol{Y}} = \widetilde{\boldsymbol{W}}\widehat{\boldsymbol{\theta}} + \widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{\vartheta}} = (\boldsymbol{A}_{\theta} + \boldsymbol{A}_{\boldsymbol{\vartheta}})\boldsymbol{Y} \tag{4.1.6}$$

(4.1.6) より、$\boldsymbol{A}_{\theta}$ はハット行列の内、固定効果に対応する行列、そして、$\boldsymbol{A}_{\boldsymbol{\vartheta}}$ は変量効果に対応する行列と言える。これらの部分的なハット行列のトレースを、それぞれ df, df_{v}と定める。

$$\mathrm{df}(\boldsymbol{\lambda}^{\beta}) = \mathrm{tr}\left[\widetilde{\boldsymbol{W}}\left\{\widetilde{\boldsymbol{W}}^{t}\ \ \widetilde{\boldsymbol{V}}^{-1}\widetilde{\boldsymbol{W}} + G_{\theta}\right\}^{-1}\widetilde{\boldsymbol{W}}^{t}\ \ \widetilde{\boldsymbol{V}}^{-1}\right] \tag{4.1.7}$$

$$\mathrm{df}_{v}\left(\boldsymbol{\lambda}^{b}\right) = \mathrm{tr}\left[\widetilde{\boldsymbol{Z}}\widetilde{D}_{\vartheta}\ \ \widetilde{\boldsymbol{Z}}^{t}\ \ \ \widetilde{\boldsymbol{V}}^{-1}(\boldsymbol{I}_{N} - \boldsymbol{A})\right] \tag{4.1.8}$$

(4.1.4), (4.1.7), (4.1.8) より、本研究で用いる FLMM の下での BIC 規準は、

$$\mathrm{BIC}\left(\boldsymbol{\lambda}^{\beta}, \boldsymbol{\lambda}^{b}\right) = -2\ \ l\left(\boldsymbol{\lambda}^{\beta}, \boldsymbol{\lambda}^{b}\right) + \log(N)\left(\mathrm{df}(\boldsymbol{\lambda}^{\beta}) + \mathrm{df}_{v}(\boldsymbol{\lambda}^{b})\right) \tag{4.1.9}$$

である。そして、BIC 規準（4.1.9）を最小化する $\left(\boldsymbol{\lambda}^{\beta}, \boldsymbol{\lambda}^{b}\right)$ が更新値として選択される。

4-2 平滑化パラメータの選択方法の比較

比較実験に用いた FLMM を以下に示す。

$$Y_{i} = \left(\alpha + \gamma_{i}\right) + \int_{0}^{1}\left(\beta(t) + b_{i}(t)\right)X_{i}(t)\,dt + \varepsilon_{i} \quad (i = 1,2, \ldots ,100, m_{i} = 10) \tag{4.2.1}$$

固定効果、変量効果の真の勾配関数 $\beta(t), b_{i}(t)$ は次の（4.2.2）のようにおく。

$$\beta(t) = c_{1}\phi\left(\frac{t - 0.5}{c_{2}}\right), \qquad b_{i}(t) = \boldsymbol{\psi}'(t)\boldsymbol{b}_{i} \tag{4.2.2}$$

ただし、$\phi(t)$ は標準正規分布の確率密度関数であり、$c_{1} = 6,\ c_{2} = \frac{1}{8}$ とする。また、$\psi(t)$ はフーリエ基底関数からなる 3 次元ベクトルであり、さらに $\boldsymbol{b}_{i} \sim N(\boldsymbol{0}, I_{3})$ である。

4-1 節で紹介した平滑化パラメータの選択方法を踏まえて、以下の 2 種類の選択方法の比較する数値実験を行う。

選択方法（A）: BIC 基準（4.1.9）を最小化する $(\lambda^\beta, \lambda^b)$ を同時に選択する
選択方法（B）: GCV 規準（4.1.2）を最小化する λ^β を選択し
REML 規準（4.1.9）を最大化する λ^b を選択する

選択方法（A）は本研究で推奨する手法であり、対して選択方法（B）は Liu et al. (2017) が用いた手法である。比較のための実験では、両選択方法とも λ^β, λ^b の選択は、範囲 [exp(–16),exp(0)] を対数スケールで 2 刻みに探索することにより行う。

また、勾配関数 $\beta(t)$ の推定精度も以下の RMISE (4.2.2) で調べる。

$$\mathrm{RMISE}\left(\hat{\beta}(t)\right) = \left[\frac{1}{M}\sum_{m=1}^{M}\frac{\int_0^T \left(\hat{\beta}^{(m)}(t) - \beta(t)\right) dt}{\int_0^T \beta^2(t)\, dt}\right]^{\frac{1}{2}} \tag{4.2.2}$$

上式にて、$\hat{\beta}^{(m)}(t)$ は勾配関数 $\beta(t)$ の m 回目の実験での推定値を表す。さらに、$\hat{b}_i(t)$ に関しても同様に RMISE を調べる。

以下に、$M = 200$回の実験中にて各選択方法によって選択された平滑化パラメータ（λ^β, λ^b）の値の分布を表 2, 3 としてまとめた。そして、200 回の実験を通して得られた RMISE の値を表 4 にまとめた。

表 2　選択された平滑化パラメータ λ_β

$\log(\lambda_\beta)$ の値

	–16	–14	–12	–10	–8	–6	–4	–2	0
(A)	2	1		8	149	38	2		
(B)	41	2	3	16	48				90

表 3　選択された平滑化パラメータ λ_b

$\log(\lambda_b)$ の値

	–16	–14	–12	–10	–8	–6	–4	–2	0
(A)				11	189				
(B)	101	11	69	19					

表 4　RMISE の比較

	$\hat{\beta}(t)$	$\hat{b}_i(t)$
(A)	0.2467	0.7319
(B)	0.5090	0.8564
改善率	52%	14%

まずは、選択された平滑化パラメータ λ_β と推定された固定効果の勾配関数 $\hat{\beta}(t)$ との関係を調べる。表 2 より、方法 (A), (B) により最も多く選択された λ_β の値は $\lambda_\beta = \exp(-8), \exp(0)$ である。これらの値が選択されたときの各 $\hat{\beta}(t)$ を図 2, 3 に示す。これらの図より、この実験の設定においては、$\lambda_\beta = \exp(-0)$ であれば罰則が大きすぎて極端に強い平滑化が行われ、$\lambda_\beta = \exp(-8)$ であれば適切な程度の平滑化が行えていると考える。よって $\exp(-8)$ 付近の値 $(\exp(-10), \exp(-6))$ も適切な値であると考えられる。

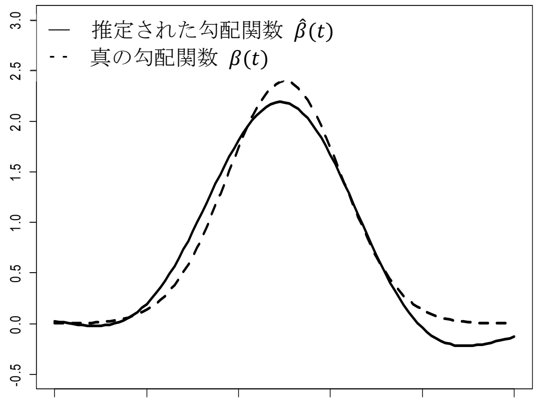

図 2　$\hat{\beta}(t)$　with　$\lambda_\beta = \exp(-8)$

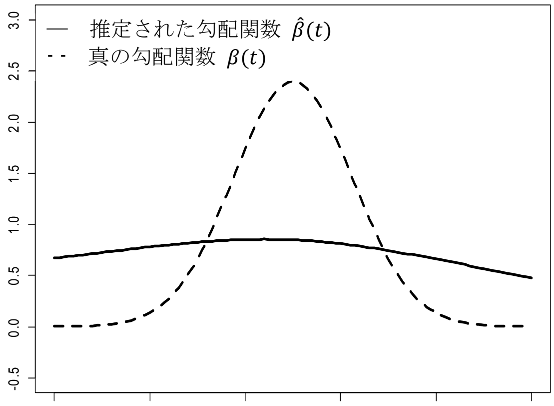

図 3　$\hat{\beta}(t)$　with　$\lambda_\beta = \exp(0)$

選択方法（B）を用いた場合、総実験回数のうち約半数（90 回/ 200 回）は $\lambda_\beta = \exp(-0)$ を選択し、極端に強い平滑化を行っている。対して選択方法（A）を用いた場合、一度も $\lambda_\beta = \exp(-0)$ を選択することなく、全実行のうち大半（195 回 / 200 回）は $\lambda_\beta = \exp(-10)$，$\exp(-8)$，$\exp(-6)$ のいずれかの値を選択し、適切な平滑化が行えている。

上で言及した平滑化の程度は表 4 の RMISE の大きさにも影響している。平滑化パラメータの選択方法として（A）を用いた場合の $\hat{\beta}(t)$ の RMISE は、(B) を用いた場合よりも 50%以上減少している。このことから、最適な λ_β を選択したいならば、Liu et al. (2017) が用いた GCV 基準 (選択方法（B)) を用いて選択するよりも、本研究で推奨する BIC 規準（選択方法（A)）を用いて選択する方が良いといえる。

さらに、選択された平滑化パラメータ λ_b と推定された変量効果の勾配関数 $\hat{b}_i(t)$ との関係を調べる。表 3 に着目すると、λ_β と同様に、方法（A), (B) とで選択される λ_b の様子が大きく異なっている。方法（A）を用いると、大半の実験（189 回 / 200 回）において $\lambda_\beta = \exp(-8)$ を選択し、方法 (B)を用いると、多くの場合 $\lambda_\beta = \exp(-10), \exp(-12)$（各 101 回, 69 回 / 200 回）を選択する。

表 4 を見ると、方法（A）を用いた場合は方法（B）を用いた場合よりも 15%近く $\hat{b}_i(t)$ RMISE が減少している。このことから、λ_b の選択にも、Liu et al. (2017) が用いた選択方法（B）よりも、本研究で推奨している方法（A）を用いる方が、妥当な値を安定して選択できるといえる。

第 5 節　結果のまとめ

本研究では、BIC 基準（4.1.9）に基づく方法（A）を用いて同時に $(\lambda_\beta, \lambda_b)$ を選択する方が、Liu et al. (2017) のように GCV や REML を用いて個別に選択するよりも、適切な値を選択でき、より適切な平滑化を行えることを例示できた（表 2, 3, 図 2, 3 参照)。また、勾配関数の推定精度が上昇したことも示せた（表 4 参照)。また、関数型説明変数を離散的な観測から再現する手法に関しても、Yao et al. (2003) の提案手法の方が Liu et al. (2017) が用いていた手法よりも近似精度が高く、推奨されることを例示した（表 1，図 1 参照)。

参考文献

[1] Aydin, D. and Memmedli, M. (2011). Optimum smoothing parameter selection for penalized least squares in form of linear mixed effect models, *Optimization* **61**,459–476.

[2] Biernacki, C., Celeux, G. and Govaert, G. (2002). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models, *Computational Statistics and Data Analysis* **41**, 561–575.

[3] Goldsmith, J., Crainiceanu, C.M., Caffo, B. and Reich, D. (2012). Longitudubal Penalized Functional Regression for Cognitive Outcomes on Neuronal Tract Measurements, *Journal of the Royal Statistical Society Ser.C*, **61**, 453–469.

[4] Guo, W. (2002). Functional Mixed Effects Models, *BIOMETRICS* **58**, 121–128.

[5] Liu, B., Wang, L. and Cao, J. (2017). Estimating functional linear mixed-effects regression models, *Computational Statistics and Data Analysis* **106**, 153–164.

[6] Micheal, S., Melnykov, V. (2016). An effective strategy for initializing the EM algorithm in finite mixture models, *Advances in Data Analysis and Classification* **10**, 563–583.

[7] Ramsay, J.O. and Siverman, B. W. (2005). *Functional Data Analysis,* Springer

[8] Reiss, P.T. and Ogden, R. T. (2008). Smoothing parameter selection for a class of semiparametric linear models, *Journal of the Royal Statistical Society* **71**, 505–523.

[9] Staniswalis, J.G. and Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data, *Journal of the American Statistical Association* **93**, 1403–1418.

[10] Wu, H. and Zhang, J.T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*, Wiley.

[11] Yao, F., Muller, H-G., Clifford, A.J., Dueker, S.R., Follet, J., Lin, Y., Buchholz, B.A. and Vogal, J.S. (2003). Shrinkage Estimation for Functional Principal Component Scores, with Application to the Population Kinetics of Plasma Folate, *Biometrics* **59,** 676–685

[12] 石井 健一郎, 上田 修功, 前田 英作, 村瀬 洋 (1998). わかりやすいパターン認識, オーム社.

[13] 船渡川 伊久子, 船渡川 隆 (2015). 統計解析スタンダード 経時データ解析, 朝倉書店.