

<修士論文>

関数ロジスティック回帰モデルに対する
スパース推定と
長距離走競技者の姿勢の分析

滋賀大学大学院
データサイエンス研究科
データサイエンス専攻

修了年度 : 2022年度

学籍番号 : 6021118

氏名 : 高田 拓弥

指導教員 : 松井 秀俊

提出年月日 : 2022年1月10日

目次

1	はじめに	1
2	関数データ	3
2.1	関数データ	3
2.2	関数データの表記法	3
2.3	関数データに対する統計量	4
2.4	関数データの解析手法	4
3	データの関数化	6
3.1	モデル	6
3.2	基底関数展開	6
3.2.1	B -スプライン	7
3.2.2	動径動径基底関数	8
3.3	モデルの推定	9
3.4	モデルの評価基準	11
3.4.1	交差検証法 CV	12
3.4.2	一般化交差検証法 GCV	12
4	スパース関数ロジスティック回帰	14
4.1	モデルと目的関数	14
4.2	パラメータの推定	15
4.3	モデル評価	19
4.3.1	赤池情報量基準 (AIC)	19
4.3.2	ベイズ情報量基準 (BIC)	19
5	適用例	20
6	まとめ	23

1 はじめに

関数データ解析 (Functional Data Analysis; FDA) は、経時的に観測された離散データを滑らかな関数として扱い、その関数からなるデータの集合 (関数データ集合) を解析する手法である (Ramsay and Silverman(1997,2005)). 近年は、観測や測定技術が急速に進歩しているため、より細かく多様なデータを取得することが可能になり、その中には各個体あるいは対象が時間や位置の変化に伴い繰り返し観測、測定されたデータも多く存在する。それに伴い、関数データ解析の需要は高まり、関数データ解析がデータ解析に応用された研究結果も多く報告されている。中でも、医療分野での応用例は多い。Kayano et al.(2016) の研究では、多発性硬化症患者に対してある治療を行った際の遺伝子発現量の経時測定データを扱い、関数ロジステック回帰モデルにより予後良好群と予後不良群とで発現量の経時変化に差異がある遺伝子を選択した。また Araki and Kawaguchi(2019) は、3次元のMRIで脳をスキャンしたデータに関数主成分分析を適用することで初期段階のアルツハイマー病患者の特徴を抽出している。最近では、Tang, Tiandong and Panpan(2020) が「COVID-19」に対しての研究結果を報告していて、アメリカ各州の「COVID-19」の1日ごとに観測される感染者数・死亡者数の経時データを関数化し、その地理的な位置関係を考慮した上で関数クラスタリングを行っている。

関数データ解析の中でも、関数回帰分析に関する研究は Ramsay and Silverman (1997) の登場以来、盛んに進んでいる。特に、目的変数が連続である場合の関数回帰モデルが多く研究されてきた。Yao, Müller and Wang (2005), Cai and Hall (2006), Hall and Horowitz (2007), Crambes, Kneip and Sarda (2009), and Yuan and Cai (2010) などが有名な例である。

また、目的変数が連続型でスカラーであり、説明変数に関数データを用いた場合の関数回帰モデルで、係数関数の局所的なスパース性に言及した論文は、James, Wang, Zhu(2009) が提案した「FLiRTI」法と Zhou, Wang, Wang(2013) が開発した2段法、Lin, Cao, Wang, Wang(2017) が提案した「SLoS」法がある。中でも「SLoS」法は、平滑であり、局所的にスパースな推定量を得るための優れた手法であり、計算が簡単で、統計的に安定した推定量を得られる。Xu, DU, Robertson, Senger(2021) は、「SLoS」法を一般化線形モデルの枠組みに拡張したものであり、係数関数を局

所的にゼロと推定するロジスティック回帰モデルについて研究している。

本研究は、これまで単変量の関数説明変数で行われてきた関数ロジスティック回帰における「SLoS」法を、複数の説明変数の枠組みに拡張することを目的とする。また、実際にその手法を長距離走者の走行データに当てはめることで、関数説明変数が複数の場合であっても、係数関数から、説明変数と目的変数の関係を容易に解釈することが可能であることを示す。

本論文の構成は以下の通りである。まず2節で関数データ解析の概要について述べる。3節では、離散データの関数化処理を行うために必要な、基底関数展開法に基づく関数化について述べる。ここでは正則化法による最尤法でパラメータの推定やモデルの評価基準についても説明する。4節では、3節で関数化処理をした関数データ集合を説明変数としたスパース関数ロジスティック回帰モデルについて述べ、3節と同様に正則化最尤法によるパラメータの推定とモデルの評価基準についても説明する。5節では、長距離走者の走行姿勢のデータに対してスパース関数ロジスティック回帰を行った結果を示す。最後に6節でまとめと今後の課題を述べる。

2 関数データ

本節では、関数データとはどのようなものであるかについて簡単に述べ、その表記法や関数データに基づく様々な統計量 (Ramsay and Silverman (1997,2005)), また関数化されたデータ集合を分析するための手法をいくつか紹介する。

2.1 関数データ

いま、各個体あるいは対象が時間や位置の変化に伴い繰り返し観測、測定されたデータがあるとする。このようなデータに対して、そのまま離散時点データとして扱うのではなく、一つの滑らかな関数を対応させて関数データとして処理し、得られた関数データ集合を一組のデータと見なし分析の対象とする方法が考えられる。このデータを関数データという。

2.2 関数データの表記法

いま、 n 個の各個体に対して、時間や位置などを表す N_i 個の観測点 $\{t_{i1}, \dots, t_{iN_i}\}$ ($i = 1, \dots, n$) において観測値 $\{x_{i1}, \dots, x_{iN_i}\}$ を得たとする。関数データ解析では、離散時点で観測されたデータの集合

$$\{(t_{ir}, x_{ir}); t_{ir} \in \mathcal{T} \subset \mathbb{R}, r = 1, \dots, N_i, i = 1, \dots, n\} \quad (2.1)$$

に替えて関数データ集合

$$\{x_i(t); t \in \mathcal{T} \subset \mathbb{R}, i = 1, \dots, n\} \quad (2.2)$$

をデータとして扱う。ただし、 \mathcal{T} は実数空間 \mathbb{R} の部分空間であり、 $x_i(t)$ は L^2 空間上の関数とする。2つの関数データ $x_1(t)$, $x_2(t)$ ($t \in \mathcal{T}$) の内積は、 $x_1(t)$ と $x_2(t)$ の \mathcal{T} 上の L^2 内積は以下で定義される。

$$\langle x_1(t), x_2(t) \rangle = \int_{\mathcal{T}} x_1(t)x_2(t)dt. \quad (2.3)$$

これより、関数データ $x_i(t)$ の L^2 ノルムは以下で与えられる。

$$\|x_i\| = \sqrt{\int_{\mathcal{T}} x_i^2(t)dt}. \quad (2.4)$$

2.3 関数データに対する統計量

関数データ $\{x_i(t); t \in \mathcal{T} \subset \mathbb{R}, i = 1, \dots, n\}$ に対する基本統計量は、次で与えられる。

$$\bar{x}(t) = \frac{1}{n} \sum_{i=1}^n x_i(t), \quad (2.5)$$

$$\text{var}_X(t) = \frac{1}{n-1} \sum_{i=1}^n \{x_i(t) - \bar{x}(t)\}^2. \quad (2.6)$$

$\bar{x}(t)$ は関数データ $\{x_1(t), \dots, x_n(t)\}$ の平均関数、 $\text{var}_X(t)$ は分散関数である。また、2つの時点 $t_1, t_2 (\in T)$ に対して相互的な関連をみるとき、次の統計量が用いられる。

$$\text{cov}_X(t_1, t_2) = \frac{1}{n-1} \sum_{i=1}^n \{x_i(t_1) - \bar{x}(t_1)\} \{x_i(t_2) - \bar{x}(t_2)\}, \quad (2.7)$$

$$\text{corr}_X(t_1, t_2) = \frac{\text{cov}_X(t_1, t_2)}{\sqrt{\text{var}_X(t_1)\text{var}_X(t_2)}}. \quad (2.8)$$

$\text{cov}_X(t_1, t_2)$, $\text{corr}_X(t_1, t_2)$ は、それぞれ $\{x_1(t), \dots, x_n(t)\}$ の2時点 t_1, t_2 の共分散、相関係数である。

2.4 関数データの解析手法

通常多変量解析を行う際に、観測時点が等間隔ではない場合や観測時点数が個体ごとに異なる場合、有益な情報を得ることは困難な場合がある。このようなデータに対して離散時点データとしてではなく、関数化したデータを扱うことによって、得られた関数データ集合を解析の対象とする手法が提案されている (Ramsay(1982), Ramsay and Silverman(2005)). この手法は関数データ解析と呼ばれ、観測時点が等間隔でない場合や、観測時点や時点数が個体ごとに異なる場合でも容易に解析できる。また関数データを微分することで、新たな情報を抽出し分析を行うことができるという利点もある。関数データ解析として、これまでに関数回帰分析、関数正準相関分析、関数主成分分析、関数データのクラスタリングなどの手法が提案されている。以降では、本論文で扱う関数ロジスティック回帰分析について簡単に紹介する。ロジスティック回帰分析は、結果となる変数と要因となる変数 (2値, または多値のカテゴリカル変数) の関係を調べ、それぞれの関係を明らかにする統計的

法である。要因となる変数を説明変数，結果となる変数を目的変数といい，説明変数が目的変数に与える影響の大きさを数値的に評価することができる。ここで通常のロジスティック回帰分析が説明変数と目的変数に離散時点で観測されたデータを用いるのに対し，関数ロジスティック回帰分析は説明変数に関数データを用いる解析手法である。本論文の4章で，このロジスティック回帰における係数関数を局所的にスパースに推定する方法について詳しく説明する。

3 データの関数化

関数データ解析を行う際、まず重要なことは離散時点データの関数化である。通常、観測されるデータは何らかの誤差を伴って得られるため、観測データから誤差を除去し、本質的な構造を適切に捉えた関数化を行う必要がある。本節では、平滑化と呼ばれる非線形回帰モデルによる関数化の手法を紹介する。ここでは、データの関数化のために基底関数展開に基づく近似を行う。また、パラメータの推定方法のための正則化法、そしてそれに伴う平滑化パラメータの選択基準をいくつか紹介する。

3.1 モデル

いま、 i 番目の個体が N_i 個の観測点 $\{t_{i1}, \dots, t_{iN_i}\}$ において観測値 $\{x_{i1}, \dots, x_{iN_i}\}$ を得たとする。ただし各個体の観測時点数は、個体ごとに必ず同一である必要はない。このときこれらのデータの関係は、 $\{t_{i1}, \dots, t_{iN_i}\}$ を説明変数、 $\{x_{i1}, \dots, x_{iN_i}\}$ を目的変数とみなした回帰モデル

$$x_{ir} = u_i(t_{ir}) + \varepsilon_{ir} \quad (3.1)$$

によって与えられるものとする。ここで、 $u_i(t)$ は真の関数、誤差 ε_{ir} は時点 r に関して互いに独立で平均 0、分散 $\sigma_{x_i}^2$ の正規分布に従う誤差項とする。関数回帰分析では $u_i(t)$ の推定値 $\hat{u}_i(t) \equiv x_i(t)$ を i 番目の個体の観測データ $\{(x_{ir}, t_{ir}); r = 1, \dots, N_i\}$ に対応する関数データとして扱う。

3.2 基底関数展開

(3.1) 式の $u_i(t_{ir})$ に線形性を仮定したモデルは、

$$x_{ir} = a_{i0} + a_{i1}t_{ir} + \varepsilon_{ir} \quad (3.2)$$

で表され、線形回帰モデルと呼ばれる。また、その線形回帰モデルに時点 t の k 次多項式を加えたモデルは、

$$x_{ir} = a_{i0} + a_{i1}t_{ir} + a_{i2}t_{ir}^2 + \dots + a_{iK}t_{ir}^K + \varepsilon_{ir} \quad (3.3)$$

で表され、多項式回帰モデルと呼ばれる。(3.2)式や(3.3)式のように、 $u_i(t)$ を $\{1, t, \dots, t^K\}$ のような基底関数の線形結合で仮定する手法は基底関数展開と呼ばれ、一般に以下の式で表される。

$$u_i(t) = \sum_{k=1}^K a_{ik} \phi_k(t) = \boldsymbol{\phi}^\top(t) \mathbf{a}_i. \quad (3.4)$$

ここで、 $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_K(t))^\top$ は K 次元基底関数ベクトル、 $\mathbf{a}_i = (a_{i1}, \dots, a_{iK})^\top$ は K 次元係数パラメータベクトルを表す。例えば、 $\phi_k(t) = t^k$ とし、 $\phi_0(t) = 1$ とその係数 a_{i0} を加えたものは多項式回帰モデルとなる。

ここで、線形回帰モデルや多項式回帰モデルは、データ解析においてよく用いられる手法であるが、モデルの柔軟性が低く、現象の構造を捉えるのに限界がある。よって関数データ化においては、より柔軟な基底関数が用いられる。ここでは代表的なものとして、 B -スプラインとガウス型動径基底関数を紹介する。なお、以後この節では個体番号を表す添字 i を省略して表記する。

3.2.1 B -スプライン

B -スプラインは節点と呼ばれる等間隔に配置された点において滑らかに連結する区分的多項式により構成される(井元, 小西(1999), Imoto and Konishi(2003))。いま、 d 次の B -スプライン関数 $B_k(t; d)$ を構成するために、節点 τ を

$$\tau_0 < \dots < \tau_d = t_0 < \dots < \tau_M = t_N < \dots < \tau_{M+d}$$

となるようにしてとる。そして、0次の B -スプライン関数を次で定義する。

$$B_k(t; 0) = \begin{cases} 1 & (\tau_{k-1} \leq t \leq \tau_k), \\ 0 & (\text{otherwise}). \end{cases}$$

そして、次の逐次計算法(de Boor(2003))によって d 次の B -スプライン関数は次で与えられる。

$$B_k(t; d) = \frac{t - \tau_{k-1}}{\tau_{k+d-1} - \tau_{k-1}} B_{k-1}(t; d-1) + \frac{\tau_{k+d} - t}{\tau_{k+d} - \tau_k} B_k(t; d-1)$$

このようにして得られる B -スプライン関数 $B_k(t; d)$ を基底関数 $\phi_k(t)$ とする。実際には3次の B -スプライン関数が多く用いられている。図1に、3次の B -スプライン関数の形を示す。

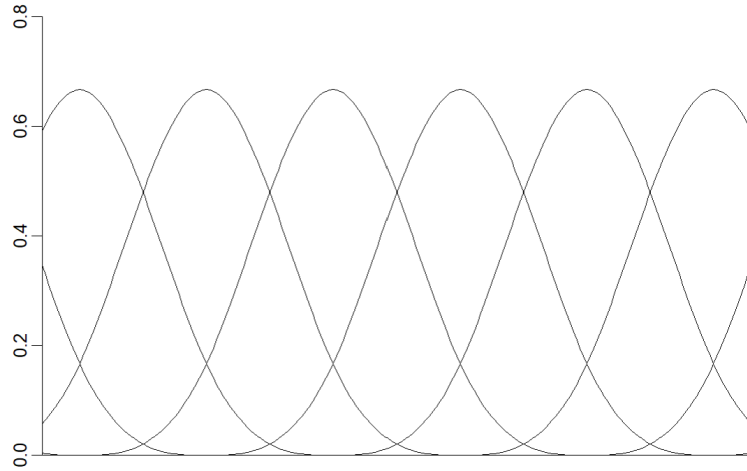


図1 3次B-スプライン

3.2.2 動径動径基底関数

動径基底関数は (3.4) 式に, $\phi_0(t) \equiv 1$ とその係数 a_0 を加えた形で与えられる. 多く用いられているのはガウス型基底関数であり, 特にハイパーパラメータを付与したものは次で与えられる (安道, 井元, 小西 (2001), Ando et al.(2005)).

$$\phi_k(t) = \exp\left(-\frac{\|t - c_k\|^2}{2vs_k^2}\right).$$

ただし, c_k は基底関数の中心を表すパラメータ, s_k^2 は基底関数の広がり程度を表すパラメータ $v(> 0)$ は基底の広がり程度をさらに調整するハイパーパラメータである. 図2にガウス型の基底関数の形を示す. c_k, s_k^2 は, 係数パラメータ a_k を推定する前にあらかじめ推定する方法が多く用いられている. ここでは k-means クラスタリング手法によって観測地点 t_1, \dots, t_N を基底関数の個数に相当する k 個のクラスタ C_1, \dots, C_K に分割し, 各クラスタ C_k に含まれる N_k 個のデータに基づいて次のように定義する.

$$\hat{c}_k = \frac{1}{N_k} \sum_{t_r \in C_k} t_r,$$

$$\hat{s}_k^2 = \frac{1}{N_k} \sum_{t_r \in C_k} (t_r - \hat{c}_k)^2.$$

ただし, N_k は k 番目のクラスタ C_k に含まれる t_r の要素数である. c_k, s_k^2 を推定するその他の方法として, t_r の分布が混合正規分布に従っているものと仮定し, EM アルゴリズムによって推定された各正規分布の平均, 分散をそれぞれ c_k, s_k^2 の推

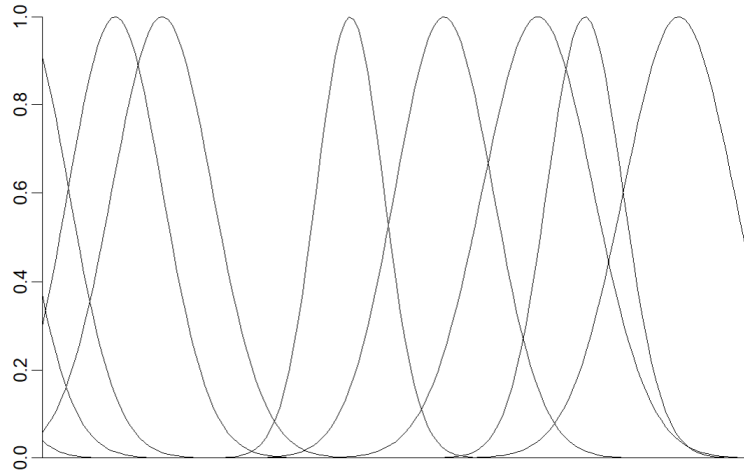


図2 動径基底関数

定値とする方法も提案されている (Hastie et al.(2001)). ガウス型基底関数の他に利用できる関数として thin-plate 型基底関数が挙げられている (Giorsi et al.(1995)). thin-plate 型基底関数は次の形で与えられる.

$$\phi_k(t) = \begin{cases} \left(\frac{\|t-c_k\|}{s_k}\right)^d & (d: \text{奇数}), \\ \left(\frac{\|t-c_k\|}{s_k}\right)^d \log\left(\frac{\|t-c_k\|}{s_k}\right) & (d: \text{偶数}). \end{cases}$$

ただし, d は t の次元とする. これらの動径基底関数は B -スプライン関数に比べて関数表現が簡易で, かつ t が高次元の場合にも関数データの分析に容易に適用が可能である.

3.3 モデルの推定

3.1 節, 3.2 節より, 誤差項正規分布を仮定した基底関数展開に基づくモデルは, x_r が互いに独立で平均 $\phi^\top(t_r)\mathbf{a}$, 分散 σ_x^2 の正規分布に従うことから, 次の確率密度関数で与えられる.

$$f(x_r|t_r; \mathbf{a}, \sigma_x^2) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left\{-\frac{(x_r - \phi^\top(t_r)\mathbf{a})^2}{(2\sigma_x^2)}\right\}.$$

従って、モデルの対数尤度関数は次で与えられる。

$$\begin{aligned}
 l(\mathbf{a}, \sigma_x^2) &= \sum_{r=1}^N \log f(x_r | t_r; \mathbf{a}, \sigma_x^2) \\
 &= \sum_{r=1}^N \log \left[\frac{1}{\sqrt{2\pi\sigma_x^2}} \exp \left\{ -\frac{(x_r - \boldsymbol{\phi}^\top(t_r)\mathbf{a})^2}{2\sigma_x^2} \right\} \right] \\
 &= -\frac{N}{2} \log(2\pi\sigma_x^2) - \frac{1}{2\sigma_x^2} \sum_{r=1}^N (x_r - \boldsymbol{\phi}^\top(t_r)\mathbf{a})^2.
 \end{aligned}$$

ここで、この対数尤度関数の最大化を行うことでパラメータの推定を行う。つまり、 $l(\mathbf{a}, \sigma_x^2) = l(\theta)$ とし、 $\partial l(\theta)/\partial \theta = 0$ を解くことでパラメータ \mathbf{a} 、 σ_x^2 は以下で表される。

$$\begin{aligned}
 \hat{\mathbf{a}} &= (\Phi^\top \Phi)^{-1} \Phi^\top \mathbf{x}, \\
 \sigma_x^2 &= \frac{1}{n} (\mathbf{x} - \Phi \hat{\mathbf{a}})^\top (\mathbf{x} - \Phi \hat{\mathbf{a}}).
 \end{aligned}$$

ただし、 $\mathbf{x} = (x_1, \dots, x_N)^\top$ は N 次元ベクトル、 $\Phi = (\boldsymbol{\phi}(t_1), \dots, \boldsymbol{\phi}(t_N))^\top$ は $N \times K$ 行列である。ここで、データ数に対して適切な基底関数の個数を設定しないと行列 $\Phi^\top \Phi$ の逆行列が不安定になる可能性がある。そこで、ここでは正則化法と呼ばれる、対数尤度関数に罰則項を付与した正則化対数尤度

$$l_\zeta(\mathbf{a}, \sigma_x^2) = l(\mathbf{a}, \sigma_x^2) - \frac{N\zeta}{2} \mathbf{a}^\top \Omega \mathbf{a}$$

の最大化によってパラメータの推定を行う。ただし、 ζ は平滑化パラメータと呼ばれるハイパーパラメータであり、適切な ζ を選択することで安定したモデルを構築できる。ここで、 Ω は既知の $K \times K$ の非負値定符号行列であり、分析の目的に応じて使い分けられる。 Ω が単位行列の場合、ridge 回帰と呼ばれる回帰係数の変動を抑え曲線の凹凸が小さなモデルを選択することができる。また、隣り合うパラメータの差分を小さくする場合、以下の差分行列の積で表される。

$$\Omega = D_2^\top D_2, \quad D_2 = \begin{pmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & & & O \\ O & & & \ddots & & \\ & & & & 1 & -2 & 1 \end{pmatrix}.$$

そして、関数の曲率を考慮したものは、 (k, h) 番目に基底関数の 2 回微分の積の積分である

$$\Omega_{kh} = \int \frac{\partial^2 \phi_k(t)}{\partial t^2} \frac{\partial^2 \phi_h(t)}{\partial t^2} dt$$

を要素に持つ行列である。ただし、基底関数の形によっては計算が困難であるため、実際には単位行列や差分行列が使われることが多い。

そして、正則化最尤法によって得られるパラメータ $\hat{\mathbf{a}}$, $\hat{\sigma}_x^2$ の推定量はそれぞれ

$$\begin{aligned}\hat{\mathbf{a}} &= (\Phi^\top \Phi + N\zeta \hat{\sigma}_x^2 \Omega)^{-1} \Phi^\top \mathbf{x}, \\ \hat{\sigma}_x^2 &= \frac{1}{n} (\mathbf{x} - \Phi \hat{\mathbf{a}})^\top (\mathbf{x} - \Phi \hat{\mathbf{a}})\end{aligned}$$

である。 $\hat{\mathbf{a}}$, $\hat{\sigma}_x^2$ は互いの推定量に依存しているため、反復計算により推定値を得る。つまり、任意に定めた ζ と σ_x^2 から繰り返し $\hat{\mathbf{a}}$ と $\hat{\sigma}_x^2$ を計算し $\hat{\sigma}_x^2$ の更新値の差が十分小さくなったところで反復を終了する。

これより、 $u(t)$ の t_1, \dots, t_N における回帰関数の推定曲線 $\hat{u}(t) = (\hat{u}(t_1), \dots, \hat{u}(t_N))^\top$ は

$$\begin{aligned}u(t) &= \Phi \hat{\mathbf{a}} \\ &= \Phi (\Phi^\top \Phi + N\zeta \hat{\sigma}_x^2 \Omega)^{-1} \Phi^\top \mathbf{x} \\ &= \mathbf{S} \mathbf{x}\end{aligned}$$

で与えられる。ただし、 $S = \Phi (\Phi^\top \Phi + N\zeta \hat{\sigma}_x^2 \Omega)^{-1} \Phi^\top$ であり、 S は平滑化行列と呼ばれる。また、正則化最尤法によって推定されたモデルは次で与えられる。

$$f(\hat{x}_r | t_r; \hat{\mathbf{a}}, \hat{\sigma}_x^2) = \frac{1}{\sqrt{2\pi \hat{\sigma}_x^2}} \exp \left\{ -\frac{(x_r - \phi^\top(t_r) \hat{\mathbf{a}})^2}{(2\hat{\sigma}_x^2)} \right\}.$$

3.4 モデルの評価基準

基底関数法と正則化最尤法に基づく推定では、正則化パラメータ ζ 、基底関数の個数 K 、選択した罰則項に依存して推定量 $(\hat{\mathbf{a}}^\top, \hat{\sigma}_x^2)$ が決定される。平滑化パラメータが小さな値をとるに従い、推定された関数は与えられたデータに過度に依存し、変動の大きな曲線が得られる。逆に大きな値をとるにつれて当てはまりの悪い、変動の小さい直線に近い曲線が得られる。したがって、これらの適切な値の選択が重要である。本節では、これらを客観的に決定する基準を紹介する。

3.4.1 交差検証法 CV

交差検証法は、手元にあるデータのみではなく、未知のデータに対する予測精度を測るため、データを訓練データと検証データに分けることで疑似的にテストデータを作り、その予測精度を計算するための基準である。こうすることで過学習を防ぐことができ、汎化性能の高いモデルを構築できる。交差検証の中でもよく利用されるのは k -分割交差検証である。 k -分割交差検証は、データを k 個に分割してそのうち1つをテストデータとして残しておき、残りの $k-1$ 個を訓練データとしてモデルを構築する。そして、取り除いたテストデータで誤差を測ることで予測誤差の計算を行う。つまり以下の式を最小にするモデルを選択する。

$$CV_k = \frac{1}{k} \sum_{l=1}^k \frac{1}{N_l} \sum_{r \in T_l} \{x_r - \hat{u}^{(-l)}(t_r)\}.$$

ただし、 $\hat{u}^{(-l)}(t_r)$ は k 個に分割したデータ集合のうち l 番目のデータ集合を取り除いたデータを用いて構築されたモデルによって得られた $u(t_r)$ の推定量、 N_l は l 番目のデータ集合のデータ数である。また $k = N$ 場合、1つ抜き交差検証法と呼ばれ、以下の式で表される。

$$CV = \frac{1}{N} \sum_{r=1}^N \{x_r - \hat{u}^{(-r)}(t_r)\}. \quad (3.5)$$

ただし、 $\hat{u}^{(-r)}(t_r)$ は r 番目のデータを除いたデータを用いて構築されたモデルによって得られた $u(t_r)$ の推定量である。また、(3.5) 式は平滑化行列 S を用いて次のように書き換えることができる。

$$CV = \frac{1}{N} \sum_{r=1}^N \left(\frac{x_r - \hat{u}(t_r)}{1 - S_{rr}} \right)^2. \quad (3.6)$$

ただし、 S_{rr} はハット行列 S の第 rr 番目の成分である。これより、 N 回の反復推定を行う (3.5) 式に比べて計算量を大幅に削減することができる。

3.4.2 一般化交差検証法 GCV

一般化交差検証法は、(3.6) 式の分母に含まれる S_{rr} をハット行列 S の対角成分の平均値である trS/N で置き換えたものであり、次で与えられる。

$$GCV = \frac{1}{N} \sum_{r=1}^N \left(\frac{x_r - \hat{u}(t_r)}{1 - \frac{1}{N} trS} \right)^2.$$

一般化交差検証法では、この値を最小にするモデルを選択する。また、CVは過適合なモデルを選択するという傾向がある一方で、GCVはその問題点を解消できるとされている。

4 スパース関数ロジスティック回帰

本節では、複数の関数データの説明変数と、二値で表される目的変数との関係をモデル化する関数ロジスティック回帰モデルの係数関数に局所的なスパース性を取り入れたモデルについて紹介する。

4.1 モデルと目的関数

いま、目的変数と Q 変数の説明変数に対して、 n 組のデータ $\{(y_i, x_{iq}(t)); t \in \mathcal{T}, i = 1, \dots, n, q = 1, \dots, Q\}$ が得られたとする。ただし、 y_i は 0 か 1 であり、 $x_{iq}(t)$ は関数データである。

もし、 $y_i \sim B(p_i)$ であるとき、関数ロジスティック回帰モデルは以下で表される。

$$\log \left(\frac{p_i}{1 - p_i} \right) = \alpha + \sum_{q=1}^Q \int_0^T \beta_q(t) x_{iq}(t) dt. \quad (4.1)$$

ここで、 α は切片項であり、 $\beta_q(t)$ は滑らかな係数関数である。ここでは $\beta_q(t)$ の局所的なスパース化に興味があり、 $[0, T]$ の中の未知の部分領域の $t \in \mathcal{Z}_q \subset \mathcal{T}$ で、 $\beta_q(t) = 0$ と推定したい。 \mathcal{Z}_q は関数 $x(t)$ が二値の応答変数 y に関する情報を持たない領域を表す。したがって、 \mathcal{Z}_q の同定は $\beta_q(\cdot)$ の推定と同様に重要である。

モデル (4.1) の対数尤度関数は以下で表すことができる。

$$l(\beta) = \sum_{i=1}^n \left(y_i \left\{ \alpha + \sum_{q=1}^Q \int_0^T \beta_q(t) x_{iq}(t) dt \right\} - \log \left[1 + \exp \left\{ \alpha + \sum_{q=1}^Q \int_0^T \beta_q(t) x_{iq}(t) dt \right\} \right] \right). \quad (4.2)$$

ここで、滑らかで局所的にスパースな係数関数 $\beta_q(t)$ を推定するために、 $\beta_q(t)$ に対して滑らかさを制御するペナルティ項と、局所的なスパース性を制御する L_1 ノルム $\int_0^T |\beta_q(t)| dt$ に対するペナルティ項について検討する。最終的な目的関数となるペナルティ付き対数尤度関数は、以下で表される。

$$J(\beta) = -l(\beta) + \gamma \sum_{q=1}^Q \int_0^T \{\beta_q^{(m)}(t)\}^2 dt + \lambda \sum_{q=1}^Q \int_0^T |\beta_q(t)|^2 dt. \quad (4.3)$$

ここで、 γ と λ は、正のハイパーパラメータであり、 β の平滑性と局所的なスパース性をそれぞれ評価する。 m は、係数関数の平滑化の対象となる微分階数を示す。 β の推定値 $\hat{\beta}$ は以下で定義される目的関数を最小にする β である。

$$\begin{aligned} \hat{\beta}_q(t) = \arg \min_{\beta_q(t)} & \left\{ - \sum_{i=1}^n \left(y_i \left\{ \alpha + \sum_{q=1}^Q \int_0^T \beta_q(t) x_{iq}(t) dt \right\} \right. \right. \\ & \left. \left. - \log \left[1 + \exp \left\{ \alpha + \sum_{q=1}^Q \int_0^T \beta_q(t) x_{iq}(t) dt \right\} \right] \right) \right. \\ & \left. + \gamma \sum_{q=1}^Q \int_0^T \{\beta_q^{(m)}(t)\}^2 dt + \lambda \sum_{q=1}^Q \int_0^T |\beta_q(t)|^2 dt \right\}. \end{aligned} \quad (4.4)$$

4.2 パラメータの推定

本節では、ニュートン-ラフソン法を用いて、(4.4)式の目的関数を最適化する方法を説明する。まず、説明変数 $x_{iq}(t)$ 、係数関数 $\beta_q(t)$ を基底関数展開により、以下で表す。

$$x_{iq}(t) = \sum_{k=1}^K a_{iqk} \phi_k(t) = \boldsymbol{\phi}^\top(t) \mathbf{a}_{iq}, \quad (4.5)$$

$$\beta_q(t) = \sum_{l=1}^L b_{ql} \psi_l(t) = \boldsymbol{\psi}^\top(t) \mathbf{b}_q. \quad (4.6)$$

ただし、 $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_K(t))^\top$ 、 $\boldsymbol{\psi}(t) = (\psi_1(t), \dots, \psi_L(t))^\top$ は K 、 L 次元の $(d+1)$ 次 B -スプライン基底関数ベクトルである。また、 $\mathbf{a}_{iq} = (a_{iq1}, \dots, a_{iqK})^\top$ 、 $\mathbf{b}_q = (b_{q1}, \dots, b_{qL})^\top$ は、 K 、 L 次元の係数パラメータベクトルである。ここで、 $\psi_l(t)$ は、 $M+1$ 個の等間隔なノット ($0 = \tau_0 < \tau_1 < \dots < \tau_M = T$) で配置されるように定義する。(4.5)式と(4.6)式より、(4.2)式は以下で表すことができる。

$$\begin{aligned} l(\mathbf{b}^*) &= \sum_{i=1}^n \left(y_i \left\{ \alpha + \sum_{q=1}^Q \mathbf{a}_{iq}^\top \int_0^T \boldsymbol{\phi}(t) \boldsymbol{\psi}^\top(t) dt \mathbf{b}_q \right\} \right. \\ & \left. - \log \left[1 + \exp \left\{ \alpha + \sum_{q=1}^Q \mathbf{a}_{iq}^\top \int_0^T \boldsymbol{\phi}(t) \boldsymbol{\psi}^\top(t) dt \mathbf{b}_q \right\} \right] \right) \\ &= \sum_{i=1}^n \left[y_i \left(\alpha + \sum_{q=1}^Q \mathbf{U}_{iq}^\top \mathbf{b}_q \right) - \log \left\{ 1 + \exp \left(\alpha + \sum_{q=1}^Q \mathbf{U}_{iq}^\top \mathbf{b}_q \right) \right\} \right] \\ &= \sum_{i=1}^n \left[y_i \left(\mathbf{U}_i^{*\top} \mathbf{b}^* \right) - \log \left\{ 1 + \exp \left(\mathbf{U}_i^{*\top} \mathbf{b}^* \right) \right\} \right]. \end{aligned} \quad (4.7)$$

ここで、 $G(\tilde{\beta}_q)$ は $\beta_q(t)$ に依存しない変数とする。また、 $\|\beta_{q[j]}\|_2^2$ は以下で計算可能である。

$$\|\beta_{q[j]}\|_2^2 = \int_{\tau_{j-1}}^{\tau_j} \beta_q^2(t) dt = \mathbf{b}_q^\top \int_{\tau_{j-1}}^{\tau_j} \boldsymbol{\psi}(t) \boldsymbol{\psi}^\top(t) dt \mathbf{b}_q = \mathbf{b}_q^\top \mathbf{W}_j \mathbf{b}_q. \quad (4.11)$$

ここで、 $\mathbf{W}_j = \int_{\tau_{j-1}}^{\tau_j} \boldsymbol{\psi}(t) \boldsymbol{\psi}^\top(t) dt$ であり、 $L \times L$ 行列である。ただし、 $j \leq l, h \leq j+d$ のとき $w_{lh} = \int_{\tau_{j-1}}^{\tau_j} \psi_l(t) \psi_h(t) dt$ であり、そうでなければ $w_{lh} = 0$ である。(4.11) 式を (4.10) 式に代入すると、以下のようになる。

$$\begin{aligned} \sum_{j=1}^M p_\lambda \left(\frac{\|\beta_{q[j]}\|_2}{\sqrt{T/M}} \right) &= \frac{\lambda}{2\sqrt{T/M}} \mathbf{b}_q^\top \left(\sum_{j=1}^M \frac{1}{\|\tilde{\beta}_{q[j]}\|_2} \mathbf{W}_j \right) \mathbf{b}_q + G(\tilde{\beta}_q) \\ &= \frac{\lambda}{2\sqrt{T/M}} \mathbf{b}_q^\top \tilde{\mathbf{W}} \mathbf{b}_q + G(\tilde{\beta}_q). \end{aligned} \quad (4.12)$$

ただし、 $\tilde{\mathbf{W}} = \sum_{j=1}^M \|\tilde{\beta}_{q[j]}\|_2^{-1} \mathbf{W}_j$ である。したがって、スパース性に関するペナルティ項は、次で近似できる。

$$\begin{aligned} \lambda \sum_{q=1}^Q \int_0^T |\beta_q(t)| dt &\approx \sum_{q=1}^Q \frac{\lambda\sqrt{T/M}}{2} \mathbf{b}_q^\top \tilde{\mathbf{W}} \mathbf{b}_q + \sum_{q=1}^Q \frac{\lambda\sqrt{T/M}}{2} G(\tilde{\beta}) \\ &= \mathbf{b}^{*\top} \tilde{\mathbf{W}}^* \mathbf{b}^* + \sum_{q=1}^Q \frac{\lambda\sqrt{T/M}}{2} G(\tilde{\beta}). \end{aligned} \quad (4.13)$$

ここで、 $\tilde{\mathbf{W}}^*$ は以下を満たす $(LQ+1) \times (LQ+1)$ 行列である。

$$\tilde{\mathbf{W}}^* = \frac{\lambda\sqrt{T/M}}{2} \begin{pmatrix} 0 & & & & \\ & \tilde{\mathbf{W}} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \tilde{\mathbf{W}} \end{pmatrix}.$$

(4.7) 式, (4.8) 式, (4.13) 式を (4.4) 式の目的関数に代入し、 \mathbf{b}^* によらない部分を除いた式は、次で与えられる。

$$\begin{aligned} J(\mathbf{b}^*) &= - \sum_{i=1}^n \left[y_i \left(\mathbf{U}_i^{*\top} \mathbf{b}^* \right) + \log \left\{ 1 + \exp \left(\mathbf{U}_i^{*\top} \mathbf{b}^* \right) \right\} \right] \\ &\quad + \mathbf{b}^{*\top} \mathbf{V}^* \mathbf{b}^* + \mathbf{b}^{*\top} \tilde{\mathbf{W}}^* \mathbf{b}^* \end{aligned} \quad (4.14)$$

ここで、ペナルティ付き尤度 J を \mathbf{b}^* について偏微分すると、以下のようになる。

$$\begin{aligned}
\frac{\partial J(\mathbf{b}^*)}{\partial \mathbf{b}^*} &= -\sum_{i=1}^n \mathbf{U}_i^* [y_i - P(\mathbf{U}_i^*; \mathbf{b}^*)] + \mathbf{V}^* \mathbf{b}^* + \tilde{\mathbf{W}}^* \mathbf{b}^* \\
&= -\mathbf{U}^{*\top} (\mathbf{y} - \mathbf{c}) + \mathbf{V}^* \mathbf{b}^* + \tilde{\mathbf{W}}^* \mathbf{b}^* \\
\frac{\partial^2 J(\mathbf{b}^*)}{\partial \mathbf{b}^* \partial \mathbf{b}^{*\top}} &= \sum_{i=1}^n \mathbf{U}_i^* \mathbf{U}_i^{*\top} P(\mathbf{U}_i^*; \mathbf{b}^*) [1 - P(\mathbf{U}_i^*; \mathbf{b}^*)] + \mathbf{V}^* + \tilde{\mathbf{W}}^* \\
&= \mathbf{U}^* \mathbf{D} \mathbf{U}^{*\top} + \mathbf{V}^* + \tilde{\mathbf{W}}^*
\end{aligned} \tag{4.15}$$

ただし、 $P(\mathbf{U}_i^*; \mathbf{b}^*) = \frac{\exp(\mathbf{U}_i^{*\top} \mathbf{b}^*)}{1 + \exp(\mathbf{U}_i^{*\top} \mathbf{b}^*)}$, $\mathbf{y} = (y_1, \dots, y_n)^\top$, $\mathbf{c} = (P(\mathbf{U}_1^*; \mathbf{b}^*), \dots, P(\mathbf{U}_n^*; \mathbf{b}^*))^\top$, $D = \text{diag}(d_{i,i})$, $1 \leq i \leq n$, $d_{i,i} = P(\mathbf{U}_i^*; \mathbf{b}^*)(1 - P(\mathbf{U}_i^*; \mathbf{b}^*))$ である。よって、ニュートン-ラフソンの更新式は以下のようになる。

$$\begin{aligned}
\mathbf{b}^{*(new)} &= \mathbf{b}^{*(old)} - \left(\frac{\partial^2 J(\mathbf{b}^*)}{\partial \mathbf{b}^* \partial \mathbf{b}^{*\top}} \right)^{-1} \frac{\partial J(\mathbf{b}^*)}{\partial \mathbf{b}^*} \\
&= \mathbf{b}^{*(old)} + (\mathbf{U}^* \mathbf{D} \mathbf{U}^{*\top} + \mathbf{V}^* + \tilde{\mathbf{W}}^*)^{-1} \\
&\quad [\mathbf{U}^{*\top} (\mathbf{y} - \mathbf{c}) - \mathbf{V}^* \mathbf{b}^{*(old)} - \tilde{\mathbf{W}}^* \mathbf{b}^{*(old)}]
\end{aligned} \tag{4.16}$$

\mathbf{b}^* の推定値 $\hat{\mathbf{b}}^*$ を求め、係数関数の推定量 $\beta_q(t)$ を得るためのアルゴリズムは以下のようになる。

1. (4.3) 式の局所的なスパース項を除いた尤度を目的関数として、最適化した際の推定値を初期推定値 $\hat{\mathbf{b}}^{*(0)}$ とする。
2. $\hat{\mathbf{b}}^{*(k)}$ を用いて、 $\tilde{\mathbf{W}}^*$ を更新し、(4.16) 式を通して、 $\hat{\mathbf{b}}^{*(k+1)}$ を更新する。
3. 収束するまでステップ 2 を繰り返す。
4. 最終的な出力 $\hat{\mathbf{b}}^{*(k+1)}$ を用いて $\hat{\mathbf{b}}^*$ とし、係数関数の推定値 $\hat{\beta}_q(t) = \psi^\top(t) \hat{\mathbf{b}}_q$ を計算する。ただし、 $\hat{\mathbf{b}}^*$ のうち、閾値より小さな推定値は 0 に設定する。

また、確率 $P(\mathbf{U}_i^*; \mathbf{b}^*)$ が、0 や 1 に近いために数値が不安定になることを防ぐために、ステップ 2 において δ 以下の確率は δ に、 $1 - \delta$ を越える確率は $1 - \delta$ に設定している。本論文では、 $\delta = 10^{-5}$ と設定した。B-スプラインは特に指定しない限り 3 次スプラインを用い、等間隔にノットを配置した 30 個程度の基底関数を使用する。なお、一般に粗さペナルティを使用する場合、 M は十分に大きい限り、重要ではないとされている。

4.3 モデル評価

推定された関数ロジスティック回帰モデルは，ハイパーパラメータ γ, λ に依存しており，この適切な値を選択することが重要である．ハイパーパラメータ γ, λ を選択するモデルの評価基準として，ここでは赤池情報量基準 (AIC)，ベイズ情報量基準 (BIC) を紹介する．

4.3.1 赤池情報量基準 (AIC)

AIC は，想定したモデルと真のモデルとの距離を Kullback-Leibler 情報量 (Kullback and Leibler(1951)) で測り，モデルの平均対数尤度の推定量をモデル評価に用いたものである (Akaike(1973))．Hastie and Tibshirani(1990) は，正則化法によって推定されたモデルの評価基準として，AIC におけるバイアス補正項，すなわち自由パラメータ数を有効自由度 df に置き換えた，次の基準を用いる．

$$\text{AIC} = -2l(\mathbf{b}^*) + 2df$$

この値を最小にするモデルを最適なモデルとする．ただし，(4.3) 式を対数尤度関数とするロジスティック回帰モデルの有効自由度 df は，以下で表される．

$$df = \text{tr}\{\mathbf{U}^{*\top}(\mathbf{U}^{*\top} \mathbf{D} \mathbf{U}^* + \gamma \mathbf{V}^* + \lambda \tilde{\mathbf{W}}^*)^{-1} \mathbf{U}^{*\top} \mathbf{D}\}. \quad (4.17)$$

4.3.2 ベイズ情報量基準 (BIC)

BIC は，モデルの事後確率に対応する周辺尤度を積分のラプラス法によって近似することによって得られたものである．AIC と同様に，自由パラメータ数を (4.17) 式の有効自由度 df に置き換えた，次の基準を用いる．

$$\text{BIC} = -2l(\mathbf{b}^*) + 2df \log(n)$$

この値を最小にするモデルを最適なモデルとする．

5 適用例

ここで扱うデータは、大阪医科薬科大学の土井博文氏より提供いただいたデータであり、高等学校陸上競技部で長距離を専門とする学生が、20km 程度のペース走を行った際に取得されたデータである。研究対象者には、走行前に骨盤ベルトを用いてセンサーを取り付けてペース走に参加してもらい、3方向 (x : 上下, y : 左右, z : 前後) の加速度を観測している。またセンサーから取得される加速度データの他に、研究対象者が感じる疲労度についても観測した。ここでは、疲労度合いを自覚的運動強度 (Rate of Perceived Extention:RPE) のひとつである Borg scale(15 スケール: Borg-PRE) を用いて評価する。研究対象者は1km 毎に区間タイムの確認時に、疲労度合いについての報告を行うことで観測した(土井(2022))。

走行周期 1 サイクルごとの観測時点数はばらばらであるため、500 時点へとダウンサンプリングして扱う。また、疲労度合いについては、得られた Borg-RPE の Borg 9 ~ 11(「かなり楽である」~「楽である」)を「疲労なし」、Borg 13 ~ 15(「ややきつい」~「きつい」)を「疲労あり」に分類した。Borg 9 ~ 11, 13 ~ 15 に該当しない場合、疲労ありか疲労なしかを判定するには不確かなので、今回は除外している。その結果、最終的に解析対象となるデータ数は 84 (1, 2, 3, 18, 19, 20, 21km 時点それぞれの走行周期 12 サイクル分) であり、加速度のデータを $x_{iqa}(i = 1, \dots, 84, q = 1, 2, 3, a = 1, \dots, 500)$ と表記する。図 3 は、解析対象となる走行周期全 84 サイクルであり、疲労なしと分類されたものを青色、疲労ありと分類されたものを赤色で色分けしている。ここでは、走行周期内で地面に接地しけることで加減速の大きくなる立脚期、地面に接地せず外力の影響を受けないため加減速の小さい跳躍期が確認できる。また、図 4 は、疲労なしとありそれぞれの各変数の平均関数を表す。

本解析では、走者の走行周期 1 サイクルの中における加速度と疲労の関係性をモデリングすることを考える。そこで、走行周期全 84 サイクル分をデータ解析の対象として、3 方向 (x, y, z) に対する加速度を関数データ化したものを説明変数 $x_{iq}(t)$ 、疲労の有無を目的変数 $y_i \in \{0, 1\}$ とし、提案手法のスパース関数ロジスティック回帰モデルを構築する。ただし、 $x_{iq}(t)$ の基底の個数を $K = 50$ 、回帰モデルの係数関数 $\beta_q(t)$ の基底の個数を $L = 30$ とする。パラメータ λ と γ は、それぞれ

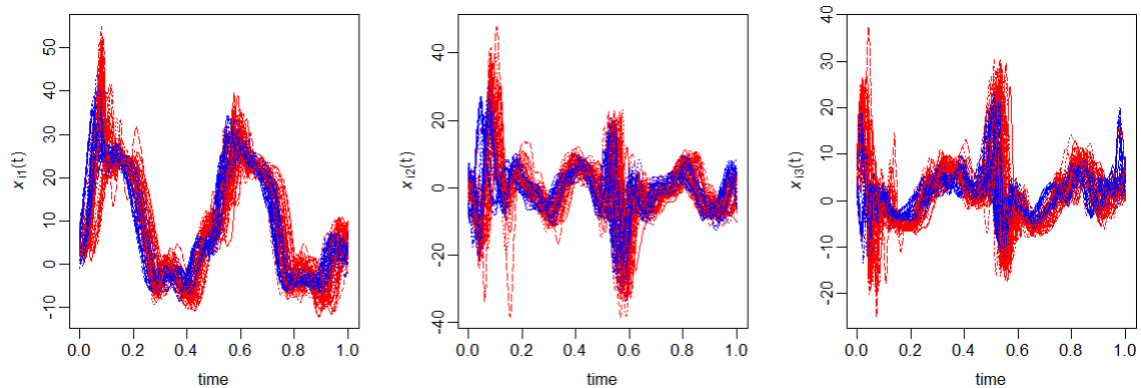


図3 左図：x字方向の加速度，中図：y軸方向の加速度，右図：z軸方向の加速度．青色は疲労なし，赤色は疲労ありである．

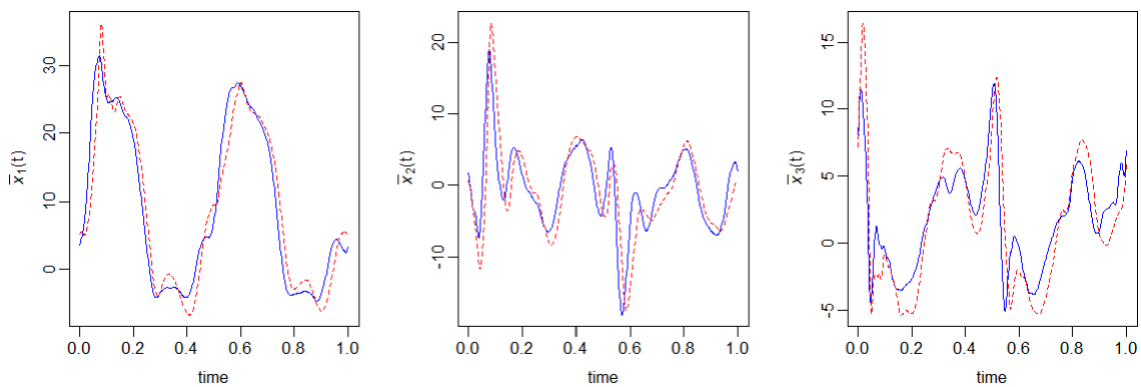


図4 左図：x字方向の加速度の平均，中図：y軸方向の加速度の平均，右図：z軸方向の加速度の平均．青色は疲労なし，赤色は疲労ありである．

$\{1, 2, 3, \dots, 30\}$ と $\{0, 1\} \times 10^{-5}$ のグリッドから BIC で選択を行った．

提案するスパースロジスティック回帰を走行データに適用した結果、 $\hat{\beta}_q(t)$ の推定値は図5のようになった．まず、 $\hat{\beta}_1(t)$ 、 $\hat{\beta}_2(t)$ 、 $\hat{\beta}_3(t)$ のいずれにおいても、推定された係数関数の非ゼロの区間は、図4の平均関数でも違いとして表れていて、その正負も一致しているように見える．また、 $\hat{\beta}_3(t)$ を見ると、係数関数の絶対値が大きい区間がいくつか見られるため、疲労を特徴づける変数としてz軸(前後)方向への加速度の影響が大きいと考えられる．一方で、 $\hat{\beta}_2(t)$ を見ると半分以上の区間で係数関数がゼロと推定されている．これについては、 $x_2(t)$ が、他の2変数と比べて、疲労なしとありで平均関数の差が小さいことから、係数関数がゼロの区間が広く推定されているためと予想される．これらのことより、提案手法は疲労に対して関連のある区間とそうでない区間を分けて推定したうえで、相関の大小・正負

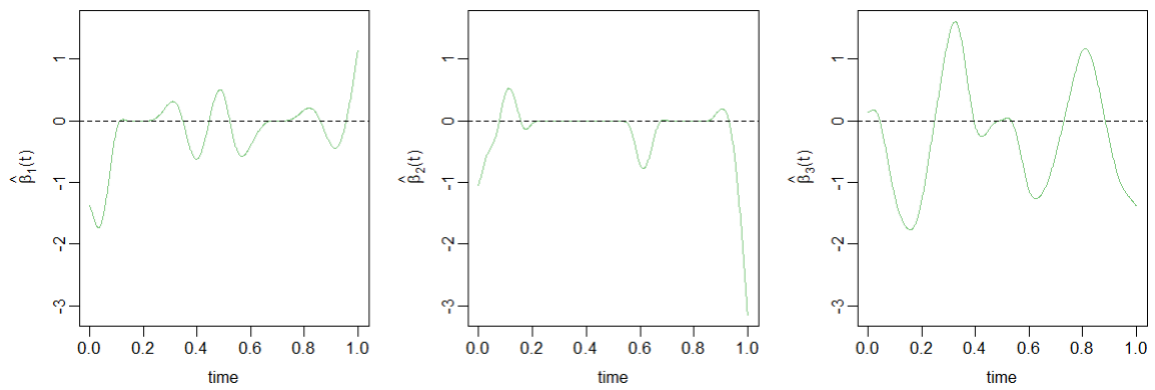


図5 走行データを可視化．左図：x 字方向の加速度，中図：y 軸方向の加速度，右図：z 軸方向の加速度

についても適切に推定することができると考えられる．

6 まとめ

本論文では、まず、2節で関数データについて述べ、3節で関数化の手法として基底展開法に基づく平滑化について述べた。4節では、複数の説明変数を用いたスパース関数ロジスティック回帰モデルについて述べ、5節で4節で提案したモデルを長距離走者の走行データに適用した。

今回の分析では、関数説明変数が複数であるロジスティック回帰モデルにおいて、係数関数を局所的にスパースに推定する方法について検討した。本提案手法は、モデルの推定と同時に、目的変数と関連のある説明変数の区間とそうでない区間を分けることができ、相関の大小・正負についても検討することが可能である。したがって、関数説明変数が一つである場合と比べて、より複雑な状況下でも適切なモデリングを行うことが可能である。

今後の展望としては、関数説明変数に対する係数関数の一部をゼロと推定する本提案手法について、係数関数の全区間をゼロと推定することのできる groupLasso に関するペナルティ項も対数尤度関数に含めることで、目的変数に対して影響のある説明変数をより適切に取捨選択できる新たなモデルを開発することが可能であると考えられる。また、ロジスティック回帰モデル以外の一般化線形モデルで係数関数を局所的にスパースに推定するなど、新たな研究の余地はある。

謝辞

本研究を遂行するにあたり、御指導を頂きました松井秀俊准教授に感謝を申し上げます。また、大阪医科薬科大学の土井博文氏には研究に必要なデータを頂きました。厚く御礼を申し上げ、感謝する次第です。この論文は私のこれからの更なる発展へ役立てたいと思います。

参考文献

- [1] 安道知寛・井元清哉・小西貞則 (2001). 動径基底関数ネットワークに基づく非線形回帰モデルとその推定. *応用統計学* **30**(1), 19-35.
- [2] Ando, T., Konishi, S. and Imoto, S. (2005). Nonlinear regression modeling via regularized radial basis function networks, To appear in *Journal of Statistical Planning and Inference*.
- [3] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csaki), 267-281, Akademiai Kiado, Budapest. (Reproduced in *Breakthroughs in Statistics, Vol.1, Foundations and Basic Theory* (eds. S. Kotz and N.L. Johnson), Springer-Verlag, New York, (1992).)
- [4] Araki, Y. (2005). Statistical Modeling for Functional Data via Regularized Basis Expansions. Dissertation, Graduate School of Mathematics, Kyushu University.
- [5] Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *Annals of Statistics* **34**, 2159–217.
- [6] Crambes, C., Kneip, A. and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *Annals of Statistics* **37**, 35–72.
- [7] de Boor C. (2001). *A Practical Guide to Splines Revised Edition*. Springer-Verlag.
- [8] 土井博文. (2022). 長距離走競技者の疲労や故障を推計するためのウェアラブルモーションセンサ信号の分析に関する基礎的研究.
- [9] Girosi, F., Jones, M. and Poggio, T. (1995). Regularization theory and neural network architectures, *Neural Computation* **7**, 219-269.

- [10] Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics* **35**, 70–91.
- [11] Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Chapman and Hall, London.
- [12] Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer-Verlag, New York.
- [13] 井元清哉・小西貞則 (1999). *B-スプラインによる非線形回帰モデルと情報量規準*. *統計数理* **47**(2), 359-373.
- [14] Imoto, S. and Konishi, S. (2003). Selection of smoothing parameters in B-spline nonparametric regression models using information criteria, *Ann. Inst. Statist. Math* **55**(4), 671-687.
- [15] James, G. M., Wang, J. and Zhu, J. (2009). Functional linear regression that’s interpretable. *Annals of Statistics* **37**, 2083–2108.
- [16] Kayano, M., Matsui, H., Yamaguchi, R., Imoto, S., and Miyano, S. (2016). (Equally contributed) Gene set differential analysis of time course expression profiles via sparse estimation in functional logistic model with application to time-dependet biomarker detection. *Biostatistics* **17**, 235-248.
- [17] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.* **22**, 79-86.
- [18] Lin, Z., Cao, J., Wang, L. and Wang, H. (2017). Locally Sparse Estimator for Functional Linear Regression Models. *Journal of Computational and Graphical Statistics* **26**, 306–318.
- [19] Ramsay, J. O. (1982) When the data are functions. *Psychometrika* **47**, 379-396.
- [20] Ramsay, J. O. and Silverman, B.W. (1997). *Functional Data Analysis*. SpringerVerlag, New York.
- [21] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis Second Edition*. Springer-Verlag, New York.
- [22] Tang, C., Wang, T. and Zhang, P. (2020). Functional data analysis: An application to COVID-19 data in the United Atates. arXiv: 2009.08363.
- [23] Xu, y., Du, P., Robertson, .J and Senger, R. (2021). Sparse logistic regression on functional data. arXiv: 2106.10583.

- [24] Yao, F. and Müller, H.-G. (2010). Functional quadratic regression. *Biometrika* **97**, 49-64.
- [25] Yuan, M. and Cai, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Annals of Statistics* **38**, 3412–3444.
- [26] Zhou, J., Wang, N. Y. and Wang, N. (2013). Functional Linear Model with Zero-Value Coefficient Function at Sub-Regions. *Statistics Sinica* **23**, 25–50.