ガウス過程に基づくノンパラメトリックベイズ時系列整列

秋本　真治†* 　末松　伸朗† 　林　　朗† 　岩田　一貴†

A Nonparametric Bayesian Time Series Alignment Based on Gaussian Processes

Shinji AKIMOTO†*, Nobuo SUEMATSU†, Akira HAYASHI†, and Kazunori IWATA†

あらまし　時系列の整列問題に対するノンパラメトリックベイズアプローチを提案する．ある種の時系列データ集合は，共通の標準時系列に対する時間変動に基づいて生成されたとみなすことができる．このような時系列データ集合の解析では，しばしば，それぞれの時間変動が相殺されるよう整列されることが要求されるが，そのためには，標準時系列と，各時系列データの時間的変動を表す時間変換関数を同時に推定しなければならない．これを実現するため，本論文では，標準時系列や時間変換関数に対してガウス過程事前分布を仮定したモデルを考え，それに対するマルコフ連鎖モンテカルロ法を開発する．本手法は，メトロポリスヘイスティングス法で必要な，事前分布とよく整合する提案分布を，ガウス過程回帰を利用して構成することで実現される．

キーワード　時系列整列，ガウス過程，マルコフ連鎖モンテカルロ法，ノンパラメトリックベイズ

1. まえがき

ある種の時系列データ集合の各データは，一つの標準となる時系列に対して個別の変更が加えられたものを観測することで得らたものであるとみなすことができる．そして，標準時系列に加えられる一つの重要な変更が，時間的なずれである．例えば，人の身長が伸びる速度を考えると，一般的には，乳幼児期の第一次発育急進期と思春期の第二次発育急進期が顕著に見られるという共通の構造があるが，特に第二次発育急進期の始まる時期や継続期間には大きな個人差が存在する．このような時系列データ集合の平均をとることを考えると，各時系列のもつ構造的な特徴は平均操作により激しくぼかされることになる．したがって，例えば，第二次発育急進期のピークなどの構造的に対応する時刻をそろえるような，時間的なずれを考慮した時系列整列がしばしば必要とされる．

時系列整列問題に対し，[1] では，時系列の極大点，極小点，変曲点などの構造上の特徴的な点を見つけ，それらをそろえる手法を提案している．しかし，多くの実際の時系列データにおいて，適切な特徴的な点を安定して見つけるのは容易ではない．

特徴点を見つける必要のない方法として，時間的な変動などをパラメトリック，または，セミパラメトリックモデルで表現し，整列後の時系列に対して定義される誤差を最小化する手法が提案されている．例えば，[2] では比較的簡単なパラメトリックモデルを時間変動と振幅の倍率に用いて自乗誤差を最小化している．また，[3] では，時間変動をセミパラメトリックモデルでモデル化している．これについては **3.** でより詳しく述べる．[4], [5] などもセミパラメトリックモデルを用いている．

本研究では，標準時系列と時間変換関数の両方についてガウス過程事前分布を利用したモデルを考え，そのモデルに対するマルコフ連鎖モンテカルロ (MCMC) 法を開発することで，時系列整列問題へのノンパラメトリックベイズアプローチを実現する．

ベイズアプローチをとった先行研究としては [6] がある．この手法は，時間変換関数等をセミパラメトリックモデルである罰則付きスプラインで表現している．平滑化スプラインとガウス過程は近い関係にあり [10]，表現力は基本的には同等と考えられるので，適切に調整されれば提案手法と同程度の推定精度を達成できるはずである．しかし，罰則付きスプラインはセミパラメトリックモデルなので，データに合わせてノットの

† 広島市立大学大学院情報科学研究科，広島市
Graduate School of Information Sciences, Hiroshima City University, 3–4–1 Ozuka-higashi, Asaminami-ku, Hiroshima-shi, 731–3194 Japan
* 現在，アイコム株式会社第 2 設計部ソフト設計 2 課

配置を適切に指定しなければならない．また，罰則付きスプラインそのものは確率モデルではないので，ベイズアプローチを実現するためには何らかの方法で確率を導入しなければならない．[6] では [11] で開発された手法を用いているが，それは，一連のスプライン係数があるランダムウォークに従っているという仮定を導入した幾分複雑なものである．これらの問題は，ノンパラメトリックな確率過程モデルであるガウス過程を用いる提案手法には存在しない．

本論文は以降次のように構成されている．まず，**2.** で時系列整列問題の定式化を行い，**3.** で以降の議論に必要な関連研究について説明する．それから，**4.** でガウス過程回帰の基本事項を要約し，**5.** で提案手法を説明する．そして，**6.** で提案手法の有効性を検証するために行った実験を示し，**7.** で本論文をまとめる．

2. 問題の定式化

本論文で考える時系列の生成モデルの定式化を行う．

標準時系列 $f(t)$ に時間変換 $s_i(t)$ を施した後，定数 a_i を乗じて得られる連続時間の時系列を，時刻列 $\mathbf{t} = [t_1, \ldots, t_T]^{\mathrm{T}}$ で観測することにより得られたものが時系列データ $\mathbf{y}_i = [y_{i,1}, \ldots, y_{i,T}]^{\mathrm{T}}$ であると考える(注1)．すなわち，$i = 1, \ldots, n, j = 1, \ldots, T$ に対し，

$$y_{i,j} = a_i f(s_i(t_j)) + \varepsilon_{i,j} \tag{1}$$

であるとする．ここで，$\varepsilon_{i,j}$ は相互に独立な平均 0，分散 σ_ε^2 の正規分布に従う変動である．そして，このようにして生成された時系列データの集合 $\mathbf{y}_{1:n} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ が与えられたときに，標準時系列 $f(t)$ と各時間変換関数 $\{s_i(t)\}_{i=1}^n$ を推定する問題を考える．

ただし，この問題には次のような不定性があることに注意されたい．ある単調な時間変換関数 $\phi(t)$ によって $f'(t) = f(\phi(t))$ と定義される $f'(t)$ を標準時系列だと考えても，時間変換関数を $s_i'(t) = \phi^{-1}(s_i(t))$ と変更すれば式 (1) と同様の関係が成立する．したがって，標準時系列関数と時間変換関数は一意には定まらず，一つの単調な時間変換関数 $\phi(t)$ だけの不定性が存在している．

3. 関 連 研 究

本論文では，DTW (Dynamic Time Warping) [7] と CR (Continuous Registration) [3], [9] を実験における比較対象とする．ここではこれらの手法の概要を述べる．[6] で提案された手法は提案手法に一番近いが，多くの調整すべきパラメータを含む複雑な手法であり比較が容易でないため，ここでは比較対象としなかった．

DTW は，音声認識の分野で開発された，2 時系列の整列法である．DTW は，2 時系列の観測時刻の対応を示す，離散的なワーピングパスの空間上で，動的計画法によって誤差を最小化する解を見つける．許されるワーピングパスの中から最適解を見つけることができるが，ノイズによる値の変動を無理に時間変動によって説明するような妥当性の低い解を見つけてしまう場合があるという問題が知られている [8]．

CR は，連続時間領域での最適化問題という形で時系列の整列問題を扱う．そのため，時系列データ $\mathbf{y}_i$ が与えられたときには，それを平滑化したスプライン曲線などをあらかじめ求め，それらを対象に整列を行う．時間変換関数 $s_i(t)$ は，二乗可積分関数 $w_i(t)$ を使って

$$s_i(t) = C_0 + C_1 \int_0^t \exp\left(\int_0^{t'} w_i(t'')dt''\right) dt' \tag{2}$$

で定義される単調関数とする．ここで，C_0, C_1 は $s_i(t)$ の値域を調整する実定数である．そして，目標時系列 $x(t)$ が与えられたときに，誤差基準

$$F[s_i(t); x(t)] = \int \{x(t) - y_i(s_i(t))\}^2 dt + \rho \int w_i^2(t)dt \tag{3}$$

を最小化する $w_i(t)$ を見つける．ここで，$y_i(t)$ は $\mathbf{y}_i$ の平滑化により得られた関数である．$F[s_i(t); x(t)]$ の最小化は，$w_i(t)$ を B スプラインなどで表現し，最適化手法によりその係数を調整することで行われる．式 (3) 右辺第 2 項は，$s_i(t)$ の曲率に関するペナルティ項で，その係数 ρ はペナルティ項の寄与を調整するためのパラメータである．

複数時系列の整列に関しては，

- 現在の $\{s_i(t)\}$ による平均時系列 $x(t)$ の計算，
- $F[s_i(t); x(t)]$ の最小化による各 $\{s_i(t)\}$ の更新

を定常に至るまで繰り返す Procrustes fitting process を用いることが提案されている．

(注1)：分かりやすさのために，本論文では全ての時系列が同じ時刻列で観測されたと仮定して論じるが，提案手法はそれぞれの時系列が異なる時刻列で観測された場合にも適用可能である．

4. ガウス過程回帰

我々の提案する手法を説明する上で必要となるガウス過程回帰 [10] の基本的事項を本節にまとめる．

まず，ガウス過程は次のように定義される．

定義 1. 確率的に定まる実数上の関数（若しくは，確率過程）$g(t)$ について，任意の自然数 p と任意の $t_1, \ldots, t_p \in \boldsymbol{R}$ に対して，$\mathbf{g} = [g(t_1), \ldots, g(t_p)]^{\mathrm{T}}$ が p 変数正規分布に従うとき $g(t)$ はガウス過程に従うといわれる．

ガウス過程は，

$$m(t) = \mathbb{E}[g(t)],$$
$$k(t, t') = \mathbb{E}[\{g(t) - m(t)\}\{g(t') - m(t')\}]$$

で定義される平均関数 $m(t)$ と共分散関数 $k(t, t')$ により完全に指定される．ここで $\mathbb{E}[\cdot]$ は期待値を表す．そしてこのとき，

$$g(t) \sim \mathcal{GP}(m(t), k(t, t'))$$

と表記される．

未知関数 $g(t)$ にガウス過程事前分布を仮定し，観測値に基づいて $g(t)$ の事後分布を得る技術はガウス過程回帰と呼ばれる．

$g(t)$ の $t_1, \ldots, t_p$ におけるノイズ付きの観測値 $\mathbf{y} = [y_1, \ldots, y_p]^{\mathrm{T}}$ が与えられたとき，$t_1^*, \ldots, t_q^*$ における値 $g(t_1^*), \ldots, g(t_q^*)$ を推測したいとする．ここで，$j = 1, \ldots, p$ に対して

$$y_j = g(t_j) + \varepsilon_j$$

であり，ε_j は他の変数と独立な平均 0，分散 $\sigma_{\varepsilon_j}^2$ の正規分布に従うノイズである．ただし，どの j についても $\sigma_{\varepsilon_j}^2 = 0$, つまり，$y_j$ はノイズなしの観測値であっても構わない．このとき，$\mathbf{g}(\mathbf{t}^*) = [g(t_1^*), \ldots, g(t_q^*)]^{\mathrm{T}}$ と表記すると，ガウス過程の定義より

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{g}(\mathbf{t}^*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(\mathbf{t}) \\ \mathbf{m}(\mathbf{t}^*) \end{bmatrix}, \begin{bmatrix} K(\mathbf{t}, \mathbf{t}) + \Lambda & K(\mathbf{t}, \mathbf{t}^*) \\ K(\mathbf{t}^*, \mathbf{t}) & K(\mathbf{t}^*, \mathbf{t}^*) \end{bmatrix} \right) \tag{4}$$

が成り立つ．ここで，$\mathcal{N}(\mu, \Sigma)$ は平均ベクトル μ，共分散行列 Σ の正規分布を表し，$\mathbf{t} = [t_1, \ldots, t_p]^{\mathrm{T}}, \mathbf{t}^* = [t_1^*, \ldots, t_q^*]^{\mathrm{T}}$．また，$K(\mathbf{t}, \mathbf{t}^*)$ 等は (i, j) 成分が $k(t_i, t_j^*)$ の行列等であり，$\mathbf{m}(\mathbf{t})$ 等は $[m(t_1), \ldots, m(t_p)]^{\mathrm{T}}$ のベクトル等である．更に，Λ は対角成分が $\sigma_{\varepsilon_1}^2, \ldots, \sigma_{\varepsilon_p}^2$ の対角行列である．

そして，

$$\begin{aligned} \boldsymbol{\mu}^* &= \mathbf{m}(\mathbf{t}^*) \\ &\quad + K(\mathbf{t}^*, \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \Lambda]^{-1}(\mathbf{y} - \mathbf{m}(\mathbf{t})), \\ \Sigma^* &= K(\mathbf{t}^*, \mathbf{t}^*) \\ &\quad - K(\mathbf{t}^*, \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \Lambda]^{-1} K(\mathbf{t}, \mathbf{t}^*) \end{aligned}$$

とおくと，正規分布の性質を使って式 (4) より

$$\mathbf{g}(\mathbf{t}^*) \mid \mathbf{t}, \mathbf{y}, \mathbf{t}^* \sim \mathcal{N}(\boldsymbol{\mu}^*, \Sigma^*)$$

であることが示される．この関係が任意の $\mathbf{t}^*$ について成り立つのだから

$$\begin{aligned} m^*(t) &= m(t) \\ &\quad + K(t, \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \Lambda]^{-1}(\mathbf{y} - \mathbf{m}(\mathbf{t})), \\ k^*(t, t') &= k(t, t') \\ &\quad - K(t, \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \Lambda]^{-1} K(\mathbf{t}, t') \end{aligned} \tag{5}$$

を定義すると

$$g(t) \mid \mathbf{t}, \mathbf{g}(\mathbf{t}) \sim \mathcal{GP}(m^*(t), k^*(t, t'))$$

である，つまり，$g(t)$ の事後分布は平均関数 $m^*(t)$，共分散関数 $k^*(t, t')$ のガウス過程であることが分かる．

5. 提 案 手 法

式 (1) のモデルにおいて，

$$f(t) \sim \mathcal{GP}(m_f(t), k_f(t, t'))$$
$$s_i(t) \overset{\mathrm{iid}}{\sim} \mathcal{GP}(m_s(t), k_s(t, t'))$$
$$a_i \overset{\mathrm{iid}}{\sim} \mathcal{N}_{>0}(\mu_a, \sigma_a^2)$$

を仮定する．ここで，$\mathcal{N}_{>0}(\mu_a, \sigma_a^2)$ は定義域が正の実数である平均 μ_a，分散 σ_a^2 の切断正規分布である．a_i の事前分布に切断正規分布を用いるのは，後の **5.2** の議論で分かるように，a_i のベイズ推定について共役事前分布となっているからである．時間変換関数は，単調増加関数に限定して推定が行われるのが一般的であ

Algorithm 1 Proposed algorithm

1: Initialize $\mathbf{s}_i^{(0)}, a_i^{(0)}$ for $i = 1, \ldots, n$ and $\sigma_\varepsilon^{2(0)}$
2: **for** $m \leftarrow 1$ **to** M **do**
3: Draw $f^{(m)}(t)$ from the posterior of $f(t)$ given $a_{1:n}^{(m-1)}, \mathbf{s}_{1:n}^{(m-1)}, \mathbf{y}_{1:n}, \sigma_\varepsilon^{2(m-1)}$
4: **for** $i \leftarrow 1$ **to** n **do**
5: Draw $a_i^{(m)}$ from the posterior of a_i given $a_{-i}^{(m)}, \mathbf{s}_{1:n}^{(m-1)}, f^{(m)}(t), \mathbf{y}_{1:n}, \sigma_\varepsilon^{2(m-1)}$
6: **end for**
7: **for** $i \leftarrow 1$ **to** n **do**
8: Draw $\mathbf{s}_i^{(m)}$ from the posterior of $\mathbf{s}_i$ given $a_{1:n}^{(m)}, \mathbf{s}_{-i}^{(m)}, f^{(m)}, \mathbf{y}_{1:n}, \sigma_\varepsilon^{2(m-1)}$
9: **end for**
10: Draw $\sigma_\varepsilon^{2(m)}$ from the posterior of σ_ε^2 given $a_{1:n}^{(m)}, \mathbf{s}_{1:n}^{(m)}, f^{(m)}, \mathbf{y}_{1:n}$
11: **end for**

るが，上記仮定ではこの制限がないことに注意されたい．したがって，本手法で推定される時間変換関数は，単調増加とならない場合もある．ガウス過程では，ノイズなしの観測値が得られた点以外の点における周辺分布は正規分布であり，正規分布の台は実数全体にわたるため，ガウス過程は減少関数の可能性を排除できない．

本研究では $f(t), \{s_i(t), a_i\}_{i=1}^n$ に対するギブスサンプラーを考えるが，$\{s_1(t), \ldots, s_n(t)\}$ については，観測時刻列 $\mathbf{t} = [t_1, \ldots, t_T]^{\mathrm{T}}$ における値しか必要ないので，$\mathbf{s}_{1:n} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ について考えればよい．ただし，$\mathbf{s}_i = [s_i(t_1), \ldots, s_i(t_T)]^{\mathrm{T}}$ である．したがって，Algorithm 1 に示したアルゴリズムにより，このモデルに対するギブスサンプラーを構成できる．ただし，Algorithm 1 中，$a_{1:n}^{(m)} = \{a_1^{(m)}, \ldots, a_n^{(m)}\}$，$a_{-i}^{(m)} = \{a_1^{(m)}, \ldots, a_{i-1}^{(m)}, a_{i+1}^{(m-1)}, \ldots, a_n^{(m-1)}\}$，$\mathbf{s}_{-i}^{(m)} = \{\mathbf{s}_1^{(m)}, \ldots, \mathbf{s}_{i-1}^{(m)}, \mathbf{s}_{i+1}^{(m-1)}, \ldots, \mathbf{s}_n^{(m-1)}\}$ である．

続く四つの小節で，Algorithm 1 の 3，5，8，10 行のサンプリングそれぞれについて説明する．

5.1 $f(t)$ のサンプリング

式 (1) より，$y_{i,j}/a_i$ は，分散 $\sigma_\varepsilon^2/a_i^2$ のノイズで乱された $f(s_i(t_j))$ の観測値である．したがって，$a_{1:n}, \mathbf{s}_{1:n}, \mathbf{y}_{1:n}$ が与えられたときの $f(t)$ の事後分布はガウス過程であり，その平均関数と共分散関数は，$n \times T$ の点 $\{s_i(t_j)\}$ における観測値 $\{y_{i,j}/a_i\}$ が与えられたときの式 (5) に相当する式から求めることができる．

ただし，実際に関数そのものをサンプリングすることはできないので，**6.** の実験で用いた実装では，比較的密にサンプルを生成してスプライン補間を行うことで必要な点における値を得る．原理的には $f(t)$ の値が必要な時刻におけるサンプルのみを順次取り出すことでスプライン補間の利用を避けることができるが，アルゴリズムが複雑になり計算時間が増大する．

5.2 a_i のサンプリング

μ_a, σ_a^2 は既知と仮定しているので，条件付き独立性より

$$\begin{aligned} p(a_i \mid a_{-i}, \mathbf{s}_{1:n}, f(t), \mathbf{y}_{1:n}) &= p(a_i \mid \mathbf{s}_i, f(t), \mathbf{y}_i) \\ &\propto p(\mathbf{y}_i|a_i, f(t), \mathbf{s}_i)p(a_i) \\ &= \mathcal{N}(\mathbf{y}_i|a_i\mathbf{f}(\mathbf{s}_i), \sigma_\varepsilon^2 I) \times \mathcal{N}_{>0}(a_i|\mu_a, \sigma_a^2) \end{aligned} \tag{6}$$

である．そして，$\mathcal{N}(\mathbf{y}_i|a_i\mathbf{f}(\mathbf{s}_i), \sigma_\varepsilon^2 I)$ を a_i の関数として捉えると

$$\begin{aligned} \mathcal{N}(\mathbf{y}_i|a_i\mathbf{f}(\mathbf{s}_i), \sigma_\varepsilon^2 I) &= \mathcal{N}(a_i\mathbf{f}(\mathbf{s}_i)|\mathbf{y}_i, \sigma_\varepsilon^2 I) \\ &\propto \mathcal{N}(a_i|\mu, \sigma^2) \end{aligned}$$

である．ここで，μ と σ^2 は，$\mathbf{f}(\mathbf{s}_i), \mathbf{y}_i, \sigma_\varepsilon^2 I$ から決まる平均と分散とする．すると，式 (6) より

$$\begin{aligned} p(a_i \mid a_{-i}, &\mathbf{s}_{1:n}, f(t), \mathbf{y}_{1:n}) \\ &\propto \mathcal{N}(a_i|\mu, \sigma^2)\mathcal{N}(a_i|\mu_a, \sigma_a^2)[\![a_i > 0]\!] \\ &\propto \mathcal{N}_{>0}(a_i|\mu_a', \sigma_a'^2) \end{aligned}$$

となり，再び正の実数上の切断正規分布となることが分かる．ここで，$[\![a > 0]\!]$ は正の実数の指示関数である．結果的に得られる切断正規分布の分散は

$$\sigma_a'^2 = \left(\frac{\mathbf{f}(\mathbf{s}_i)^{\mathrm{T}}\mathbf{f}(\mathbf{s}_i)}{\sigma_\varepsilon^2} + \frac{1}{\sigma_a^2} \right)^{-1}$$

であり，平均は上記分散を使って

$$\mu_a' = \sigma_a'^2 \left(\frac{\mathbf{y}_i^{\mathrm{T}}\mathbf{f}(\mathbf{s}_i)}{\sigma_\varepsilon^2} + \frac{\mu_a}{\sigma_a^2} \right)$$

である．a_i のサンプルはこの切断正規分布から取り出される．

5.3 $\mathbf{s}_i$ のサンプリング

$\mathbf{s}_i$ の条件付き確率は，

$$
\begin{aligned}
p(\mathbf{s}_i|a_{1:n}, \mathbf{s}_{-i}, f(t), \mathbf{y}_{1:n}) &= p(\mathbf{s}_i|a_i, f(t), \mathbf{y}_i) \\
&\propto p(\mathbf{y}_i|\mathbf{s}_i, a_i, f(t)) p(\mathbf{s}_i) \\
&= \mathcal{N}(\mathbf{y}_i|a_i \mathbf{f}(\mathbf{s}_i), \sigma_\varepsilon^2 I) p(\mathbf{s}_i) \qquad (7)
\end{aligned}
$$

である．上式の右辺の第一因子は正規分布であるが，$\mathbf{s}_i$ について見ると，非線形関数 $f(\cdot)$ を通して関係しているために標準的な分布とはならない．

そこで，$\mathbf{s}_i$ については，メトロポリス・ヘイスティングス (MH) 法を使ってサンプリングを行う．つまり，提案分布に従って候補 $\mathbf{s}'_i$ を生成し，MH 法の定める受理確率に従ってそれを受理する．

MH 法では，適切な提案分布を選ばなければ効果的なサンプリングが実現できない．$\mathbf{s}_i$ のサンプリングにおいては，$\mathbf{s}_i$ がガウス過程に従う関数 $s_i(t)$ のサンプルであるため，要素間に強い相関があることを考慮して提案分布を考えなければならない．その要素間の強い相関を考慮せず，例えば，$\mathbf{s}_i$ の要素に独立な正規分布に従う変動を加えるような提案分布を考えると，変動の大きさがほとんど無視できるほど小さくなければ受理される確率はほぼ 0 となってしまう．この事情は，変動を加える要素を一つに限定しても変わりない．

また，$\mathbf{s}_i$ の次元は，時系列の長さに等しいので，一般的に比較的高次元である．そのため，全部の要素に一斉に変更を加えるような提案分布では，変更の受理がごくまれにしか起こらない．

そこで，本研究では移動窓型のサンプリング法を用いる．すなわち，$t_j, \ldots, t_{j+\lambda-1}$ を更新対象となる幅 λ の窓とすると，まず，窓外の $s_i(t_1), \ldots, s_i(t_{j-1}), s_i(t_{j+\lambda}), \ldots, s_i(t_T)$ をノイズなしの観測値，窓内の $s_i(t_j), \ldots, s_i(t_{j+\lambda-1})$ をノイズ付きの観測値とみなしたときの $s_i(t)$ の事後分布を (5) より得る．そして，その事後分布にしたがって，窓内の $s_i(t_j), \ldots, s_i(t_{j+\lambda-1})$ の提案値をサンプリングし，MH 法に従って受理，または，棄却する．窓は 1 観測時刻ずつ移動して $\mathbf{s}_i$ 全体について更新を行う．

この提案分布では，窓の幅 λ により提案分布により生じる変動の大きさを調整することができるため，それを適切に設定することにより，効率的なマルコフ連鎖のシミュレーションが可能となる．

5.4 σ_ε^2 のサンプリング

モデル (1) より，$a_{1:n}^{(m)}, \mathbf{s}_{1:n}^{(m)}, f^{(m)}, \mathbf{y}_{1:n}$ が与えられたとき，$i = 1, \ldots, n$，$j = 1, \ldots, T$ に対する $y_{i,j} - a_i^{(m)} f^{(m)}(t_j)$ は，平均 0，分散 σ_ε^2 の正規分布に従うサンプルとなる．したがって，σ_ε^2 の事後分布は，平均既知，分散未知の正規分布に関するベイズ推定より得られる（例えば [12]）．

このベイズ推定における σ_ε^2 の共役事前分布は逆ガンマ分布であるが，本論文の実験では，逆ガンマ分布の極限として得られる変則分布 (improper distribution)

$$
p(\sigma_\varepsilon^2) \propto (\sigma_\varepsilon^2)^{-1}
$$

を用いる．

6. 実　　験

提案手法の有効性を実験的に検証する．

本論文の実験を通じて，標準時系列と時間変換関数の平均関数を，それぞれ，$m_f(t) \equiv 0$，$m_s(t) = t$ とする．また，共分散関数は，

$$
k(t, t') = \sigma^2 \exp\left(-\frac{1}{2l^2}(t - t')^2\right)
$$

の形のものを用いる．この共分散関数は，1 点の周辺分散 σ^2 とスケールパラメータ l をパラメータとしてもつ．標準時系列に関するこれらのパラメータを l_f, σ_f^2，時間変換関数に関するものを l_s, σ_s^2 と記す．

6.1 人工データ実験

モデル (1) に従って複数の時系列データを生成し，それらを整列させる実験を行い，DTW，CR，提案手法の 3 手法で整列の精度を比較する．

DTW 及び CR との比較を行うため，本実験ではどの時系列の始端と終端も対応している，つまり $s_i(t_1) = t_1, s_i(t_T) = t_T$ $(i = 1, \ldots, n)$ とする．また，モデル (1) における振幅の倍率 a_i は考慮しない，つまり $a_i = 1$ $(i = 1, \ldots, n)$ とする．

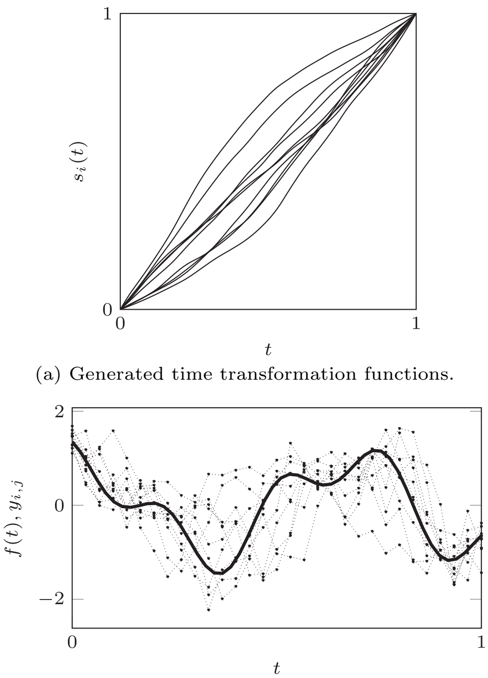
6.1.1 人工データの生成

時系列データを生成するには，ランダムな時間変換関数 $s_i(t)$ の生成が必要である．これには，式 (2) で表される Ramsay [13] の提案する単調増加関数のクラスを用いる．式 (2) を使うと，$w_i(t)$ をランダムに生成することで時間変換関数のサンプルが得られる．本実験では，時間を $[0, 1]$ の範囲で考えるので，$w_i(t)$ を $[0, 1]$ 区間を 20 等分した各区間で一定の値をとる関数として定義する．そして，各区間の値は，平均 0，分散 16 の正規分布からサンプリングする．

標準時系列は

$$
f(t) = \cos(3\pi t) + \frac{1}{2} \sin\left[3\pi\left(\frac{5}{2}t + \frac{1}{4}\right)\right]
$$

とした．



(a) Generated time transformation functions.

(b) $f(t)$ and generated time series.

図 1　人工時系列データセットの例

Fig. 1　An example of synthesized time series dataset.

10 本の時系列データの整列を 1 試行とする．ある試行で生成された 10 の時間変換関数を図 1 (a) に示す．また，これに対応する 10 の時系列データを図 1 (b) に，標準時系列とともに示す．図中の太線が標準時系列であり，点線で結ばれた点が $\{y_{i,j} | i = 1, \ldots, 10, j = 1, \ldots, 31\}$ である．観測時刻は $[0, 1]$ を 30 等分する 31 時刻であり，ノイズの分散 σ_ε^2 は 0.3^2 である．

6.1.2　評価方法

整列の精度を比較する上でまず思いつくのは，標準時系列 $f(t)$ の推定精度である．しかし，**2.** で述べたように，時系列の整列問題に存在する不定性のために，$f(t)$ の推定精度が必ずしもよい指標とはいえない．また，DTW ではそれが得られないという問題もある．

そこで，本実験では，2 時系列データ間の時間の対応を示すワーピングパスの推定精度を測る．$\mathbf{y}_i$ と $\mathbf{y}_j$ の間のワーピングパスは $s_j^{-1}(s_i(t))$ により得られる．そして，推定されたワーピングパスの真のワーピングパスに対する誤差は，二つのワーピングパスにより挟まれる部分の面積で評価することにする．

6.1.3　実験条件

DTW，CR と比較するため，提案手法においては $s_i(0) = 0$，$s_i(1) = 1$，$a_i = 1$（$i = 1, \ldots, 10$）は固定する．そして，5 万ステップの MCMC シミュレーションを行って最後の 1 万ステップの時間変換関数を平均することで $s_i(t)$ の事後平均を得る．用いた共分散関数に関するパラメータは，$l_f = 0.2, \sigma_f^2 = 25$, $l_s = 0.5, \sigma_s^2 = 0.1$ であり，窓幅 λ は 25 とした．

3. で述べたように，CR では，時系列データの平滑化関数をまず求めて，それを整列に用いる．したがって，その際に使われる平滑化パラメータが結果に大きく影響を及ぼすが，本実験では，平均誤差が最小となるようにパラメータを調整した．また，Procrustes fitting process を十分収束するまで繰り返した．CR の実装は Ramsay 等により公開されている fda パッケージ [14] を用いた．

DTW は，Classical DTW とも呼ばれる最も基本的なものである．つまり，ワーピングパスでは，それぞれの時系列の時刻が，どちらか一方のみが 1 時刻進む，若しくは，両方が 1 時刻進むかのいずれかのみが許される．また，誤差は，ワーピングパスで対応する時刻の観測値の差の自乗和で定義される．

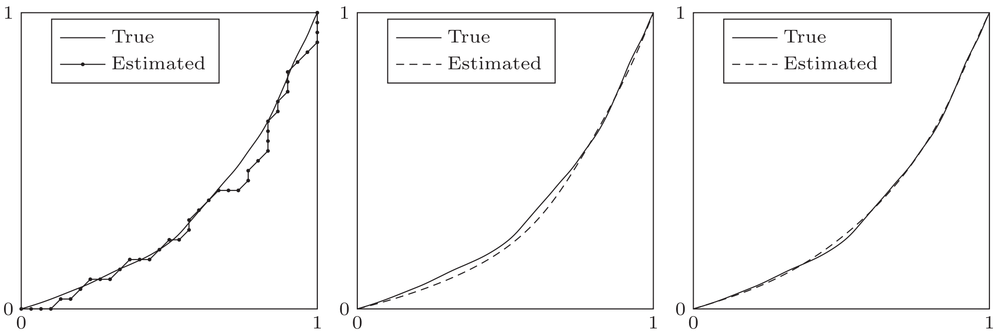
10 本の時系列データの整列を行うのを 1 試行とするので，45 のワーピングパスが推定される．それら，45 のワーピングパスの誤差の平均値を 1 試行の誤差とし，100 試行の結果により評価を行う．

6.1.4　実験結果

図 2 に，真のワーピングパスと推定された三つのワーピングパスの例を示す．この例では，DTW は真のワーピングパスから大きく逸脱しているところが見られる一方，提案手法では非常に精度良く推定できている．

100 試行で得られた誤差の箱ひげ図を図 3 に示す．図より，DTW, CR, 提案手法の順に誤差が減少していることが分かる．この図で切り込み (notch) は互いに重なっておらず，互いの中央値は 95%程度の信頼度で有意な差があることが示唆される [15]．

提案手法が CR よりも高い精度を実現できるのは，ベイズの枠組みに従った確率論的手法であるためと思われる．高い精度を実現するためには，一般的には，標準時系列や時間変換関数をより柔軟なモデルで表現することが望ましい．しかし，柔軟なモデルを用いると，過適合を抑制するための正則化が必要になる．CR では，前処理として行われる平滑化と，誤差汎関



(a) DTW　　(b) CR　　(c) Our method

図 2　真のワーピングパスと推定されたワーピングパスの差
Fig. 2　Difference between estimated warping path and true one.

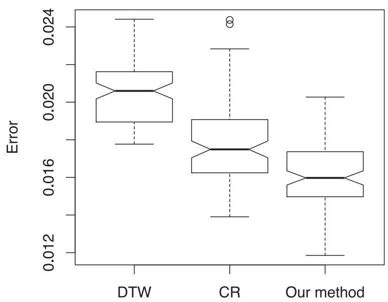


図 3　誤差の箱ひげ図
Fig. 3　Box and whisker plot of the errors.

数 (3) に含まれる正則化項でこれを実現しているが，それらは多分に便宜的であり必然性に乏しい．

6.2　成長速度曲線

実データに対する整列を行い，CR の結果と比較する．fda パッケージの例にある成長速度曲線の整列問題を対象とする．54 人の女性について記録された，1 歳から 18 歳までの身長データ [16] に基づいて得られた成長速度曲線の整列問題である．

3. で述べたように CR では，まず時系列データを平滑化した曲線が求められる．今の場合，データは身長であり，整列の対象は成長速度曲線なので，データに対する平滑化スプラインの結果を微分してそれを得る．CR により得られた整列結果を図 4 (a)～(c) に示す．図中の灰色の線は整列された成長速度曲線であり，太線はそれらの平均である．図 4 (a)～(c) は，成長速度曲線を求めるために行った平滑化スプラインのパラメータのみが異なる．平滑化にはいずれも六次の基底関数を用いているが，roughness penalty で使われる微分の階数 d と roughness penalty の係数 c はそれぞれ図の下に示された値を使っている．

一方，提案手法では，身長の時系列データに対する三次スプライン補間から，各観測年齢における成長速度を求め，整列を行う．MCMC シミュレーションは 2 万ステップ行い，後半の 1 万ステップから時間変換関数 $\{s_i(t)\}$ と，振幅の倍率 $\{a_i\}$ の事後平均を求めた．結果を図 4 (d)～(f) に示す．図中の灰色の線は整列された成長速度時系列であり，太線は，サンプルから得られた標準時系列 $f(t)$ の事後平均である．共分散関数のパラメータのうち l_f は各図に示した値を用い，その他は共通で $\sigma_f^2 = 2500$，$l_s = 6$，$\sigma_s^2 = 4$ である．窓幅 λ は 25 とした．

図 4 では，上下の標準時系列（平均時系列）がおおむね似た形状を示すようにパラメータが選ばれている．左の列の図 4 (a) と 4 (d) は最も滑らかな結果であり，乳幼児期の第一次発育急進期と思春期の第二次発育急進期だけが顕著であり，他の部分では滑らかに推移している．

第一次発育急進期と第二次発育急進期の間には，一度成長が加速する時期があることが知られており mid-growth spurt などと呼ばれている [17]．図 4 の中央の列はこの mid-growth spurt に相当すると思われるピークが顕著な場合である．図 4 (b) の 4 歳ごろのピークと，図 4 (e) の 7 歳ごろのピークがそれに当たると思われる．図 4 (e) には 3 歳ごろにも顕著なピークが存在しているが図 4 (b) には見られない．この差異は，このピークに対応する成長速度の高まりが，平滑化スプラインにより均されいるためだと思われる．実際，roughness penalty の係数がより小さな図 4 (c) では，これに対応すると思われるピークが 2 歳ごろに存在する．そして，図 4 の右の列は，より屈曲した標

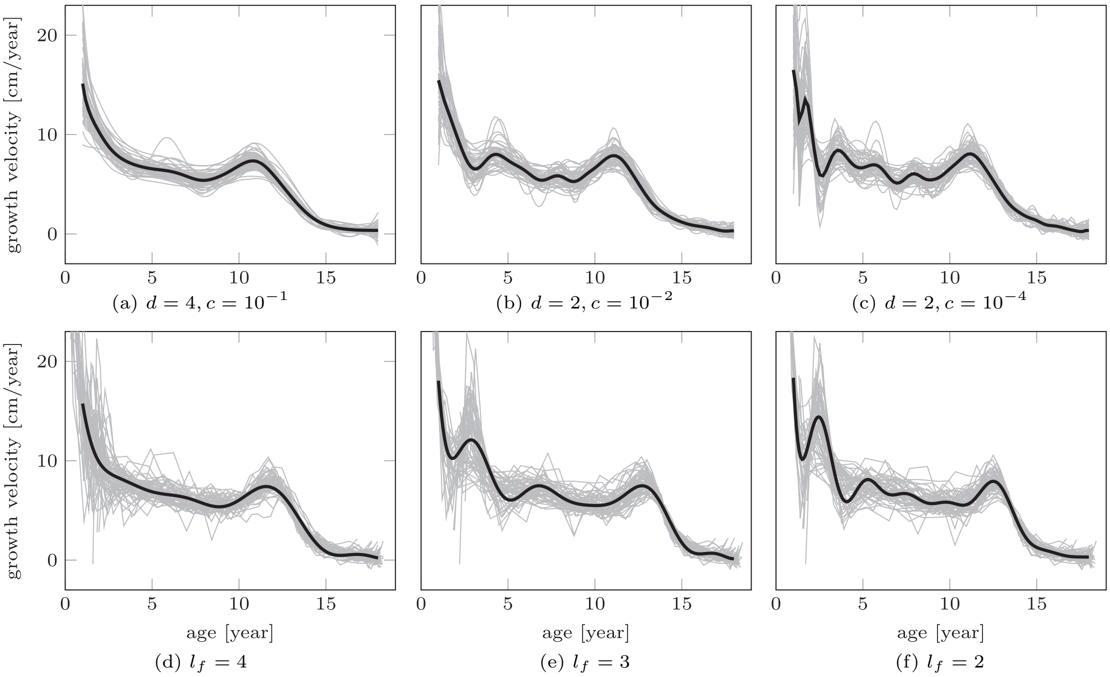


図 4　整列された成長速度時曲線
Fig. 4　Aligned growth velocity curves.

準時系列が得られた場合である．

上記のように，パラメータを適当に選ぶことで CR と提案手法はおおむね似た形状の標準時系列が得られることが確認できたが，全体的な傾向として，CR の方が提案手法よりも各ピークが低年齢側にずれているという差異も見られた．[17] では女子の mid-growth spurt は，平均的には 6.5 から 7 歳で起きると推定されており，提案手法の結果の方がこの知見との一致がよい．**2.** で述べたように，時系列整列問題には時間変換関数 $\phi(t)$ だけの不定性が存在するにもかかわらず，提案手法でこの一致を見たのは，時間変換関数の平均を恒等変換 $s_i(t) = t$ としており，それに近い時間変換関数を好む傾向があるためと推測される．CR でも，ペナルティ項の係数を大きくしてより滑らかな時間変換関数を好むように設定すれば，恒等変換に近いものを好むようになるとも考えられるので実際に試したが，目立った変化は生じなかった．

7. む す び

時系列整列問題に対する，ノンパラメトリックベイズ手法を提案した．この手法では，共通の標準時系列に対して，それぞれ異なる時間変換が施されたものを繰り返し観測することで一連の時系列データが得られるとみなす．そして，標準時系列と各時系列データに対応する時間変換関数にガウス過程を仮定し，MCMC 法を利用することで，標準時系列と全ての時間変換関数を同時に推定する．

人工データを用いて，DTW や CR と整列の精度を比較する実験を行い，提案手法が高い精度の整列を実現できることを確認した．また，実データに対しても実験を行い，結果を CR と比較したところ，CR ではほとんど見られない mid-growth spurt に対応すると思われる山が提案手法では得られた．

現在，ガウス過程事前分布パラメータ等を既知であるとしているが，今後，それらの推定を MCMC に組み込むことを検討する予定である．

文　献

[1] A. Kneip and T. Gasser, “Statistical tools to analyze data representing a sample of curves,” The Annals of Statistics, vol.20, no.3, pp.1266–1305, 1992.

[2] W. Härdle and J.S. Marron, “Semiparametric comparison of regression curves,” The Annals of Statistics, vol.18, no.1, pp.63–89, 1990.

[3] J.O. Ramsay and X. Li, “Curve registration,” J. Royal Statistical Society B, vol.60, no.2, pp.351–363, 1998.

[4] K. Wang and T. Gasser, "Synchronizing sample curves nonparametrically," The Annals of Statistics, vol.27, no.2, pp.439–460, 1999.
[5] D. Gervini and T. Gasser, "Self-modelling warping functions," J. Royal Statistical Society B, vol.66, pp.959–971, 2004.
[6] D. Telesca and L.Y.T. Inoue, "Bayesian hierarchical curve registration," J. American Statistical Association, vol.103, no.481, pp.328–339, 2008.
[7] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," IEEE Trans. Acoust. Speech Signal Proces., vol.ASSP-26, no.1, pp.43–49, 1978.
[8] E.J. Keogh and M.J. Pazzani, "Derivative dynamic time warping," Proc. First SIAM International Conference on Data Mining (SDM'2001), 2001.
[9] J.O. Ramsay and B.W. Silverman, Functional Data Analysis, Springer, 1997.
[10] C.E. Rasmussen and C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.
[11] S. Lang and A. Brezger, "Bayesian p-splines," J. Computational and Graphical Statistics, vol.13, no.1, pp.183–212, 2004.
[12] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin, Bayesian Data Analysis, 2nd edition, Chapman & Hall, 2003.
[13] J.O. Ramsay, "Estimating smooth monotone functions," J. Royal Statistical Society B, vol.60, no.2, pp.365–375, 1998.
[14] J.O. Ramsay, G. Hooker, and S. Graves, Functional Data Analysis With R and MATLAB, Springer, 2009.
[15] R. McGill, J.W. Tukey, and W.A. Larsen, "Variations of box plots," The American Statistician, vol.32, no.1, pp.12–16, Feb. 1978.
[16] R.D. Tuddenham and M.M. Snyder, "Physical growth of california boys and girls from birth to eighteen years," University of California publications in child development, vol.1, no.2, 1954.
[17] L. Molinari, R.H. Largo, and A. Prader, "Analysis of the growth spurt at age seven," Helvetica Paediatrica acta, vol.35, no.4, pp.325–334, Sept. 1980.

（平成 24 年 6 月 1 日受付，9 月 14 日再受付）

秋本　真治

平 22 広島市立大・情報科学卒．平 24 同大大学院情報科学研究科博士前期課程了．同年，アイコム（株）入社，現在に至る．

末松　伸朗　（正員）

昭 63 九大・理・物理卒．平 2 同大大学院修士課程了．同年,（株）富士通研究所入社．平 6 より広島市立大学情報科学部助手．現在は，同大学大学院情報科学研究科准教授．博士（工学）．人工知能学会，情報処理学会各会員．

林　　朗　（正員）

昭 49 京大・理・数学卒．同年，日本 IBM（株）入社．昭 63 ブラウン大学計算機科学科修士課程了．平 3 テキサス大学オースティン校計算機科学科博士課程了．同年，九州工業大学情報工学部客員助教授．平 6 広島市立大学情報科学部教授．現在は，同大学大学院情報科学研究科教授．人工知能学会，情報処理学会，AAAI，ACM，IEEE 各会員．

岩田　一貴　（正員）

平 12 名工大・工・知能情報システム卒．平 14 同大大学院工学研究科博士前期課程了．平 17 京大大学院情報学研究科博士後期課程了．平 14～17 日本学術振興会特別研究員 DC1．平 17 広島市立大学情報科学部助手．現在は，同大学大学院情報科学研究科講師．平 17 IEEE 関西支部 Student Paper Award 受賞．IEEE 会員．