動的ノンパラメトリックベイズモデルにおける変分ベイズ推定

数学専攻 柳澤 圭介
Keisuke YANAGISAWA

概要

地球科学やマーケティング, スポーツなど様々な分野において, 状態空間モデルを用いた時系列データの解析が活用されるようになってきた．そのような中で, 複数の時系列データにおける状態の推移パターンを発見するために, 状態に階層構造を導入した様々なモデルが提案されてきている．例えば, Chiappa and Barber(2007) では, 混合分布を用いることにより, 状態の分類が可能なモデルを提案した．一般に, 混合分布を用いたモデルは, クラスタ数を与えた上でパラメータの推定を行う．それに対して, 近年, クラスタ数とパラメータを同時推定する, ディリクレ過程を導入した方法が注目されている．本研究では, ディリクレ過程混合モデルを用いることにより階層構造を導入し, データに基づいてクラスタ数を推定するモデルを考え, 変分ベイズ法によるパラメータ推定法を提案する．

1 動的ノンパラメトリックベイズモデル

状態空間モデルは, 観測された時系列データの時間的推移を直接モデル化するのではなく, 観測できない状態と呼ばれる潜在変数が時間的に推移し, 状態の値に応じて観測変数が観測されるとしたモデルである．状態の推移を表現するモデルをシステムモデルといい, 状態と観測変数の関係を表現するモデルを観測モデルという．

複数の時系列に対して状態空間モデルを用いた場合, 個体ごとに様々な状態のパターンが推定される．そこで, システムモデルにクラスタ構造を導入することにより, 状態の分類をすることが考えられる．Chiappa and Barber(2007) は, 状態ベクトルに混合分布を用いて階層構造を導入した状態空間モデルを提案した．複数の時系列データ $\boldsymbol{y}^i_{1:T} = \left\{\boldsymbol{y}^i_1, \ldots, \boldsymbol{y}^i_T\right\}, i = 1, \ldots, N$ が観測されたとき, このモデルは, 次のように書き表せる:

$$
\begin{aligned}
&\boldsymbol{y}^i_t = H\boldsymbol{x}^g_t + \boldsymbol{u}^i_t + \boldsymbol{w}^i_t, \boldsymbol{w}^i_t \sim N_l(\boldsymbol{w}^i_t; \boldsymbol{0}, R),\\
&\boldsymbol{x}^g_t = F^g\boldsymbol{x}^g_{t-1} + \boldsymbol{v}^g_t, \boldsymbol{v}^g_t \sim N_k(\boldsymbol{v}^g_t; \boldsymbol{0}, Q^g),\\
&\boldsymbol{u}^i_t \sim N_l(\boldsymbol{u}^i_t; \boldsymbol{0}, \Sigma_{u_i}), \boldsymbol{x}^g_0 \sim N_k(\boldsymbol{x}^g_0; \boldsymbol{\mu}^g_0, \Sigma^g_0).
\end{aligned}
$$

ここで, g は $\boldsymbol{y}^i_{1:T}$ が所属するクラスタを表しており, $1, \ldots, G$ のいずれかの値をとる．$\boldsymbol{x}^g_t$ は状態と呼ばれ, 直接には観測されない k 次元ベクトルである．$\boldsymbol{u}^i_t$ は, 個体ごとの変量効果を表すものである．H, F^g は, それぞれ $l \times k$, $k \times k$ の行列である．これは, 同じクラスタに属する個体は同じ状態の値をもつとしたモデルであり, 同一クラスタ内の個体ごとの差異は観測モデルの変量効果により表現される．

それに対して, 近年注目されている方法に, パラメータとクラスタ数を同時に推定するディリクレ過程混合モデルがある (Antoniak, 1974)．ディリクレ過程混合モデルでは, 観測データに応じて, クラスタ数を自動的に決定することができる．ディリクレ過程混合モデルを導入した線形ガウス状態空間モデルとして, 以下のようなモ

デルが考えられる:

$$
\begin{aligned}
&\boldsymbol{y}_t^i = H\boldsymbol{x}_t^i + \boldsymbol{u}_t^i + \boldsymbol{w}_t^i, \boldsymbol{w}_t^i \sim N_l(\boldsymbol{w}_t^i; \mathbf{0}, R),\\
&\boldsymbol{x}_t^i = F^i\boldsymbol{x}_{t-1}^i + \boldsymbol{v}_t^i, \boldsymbol{v}_t^i \sim N_k(\boldsymbol{v}_t^i; \mathbf{0}, Q^i),\\
&\boldsymbol{u}_t^i \sim N_l(\mathbf{0}, \Sigma_{u_i}), \boldsymbol{x}_0^i \sim N_k(\boldsymbol{x}_0^i; \boldsymbol{\mu}_0^i, \Sigma_0^i),\\
&\boldsymbol{\psi}_i = \left\{\boldsymbol{\mu}_0^i, \Sigma_0^i, F^i, \boldsymbol{\alpha}^i, Q^i\right\} \sim \mathrm{DP}(\zeta, G_0).
\end{aligned}
$$

ここで, $\mathrm{DP}(\zeta, G_0)$ は集中度 ζ, 基底分布 G_0 のディリクレ過程を表し, 基底分布 G_0 は

$$
\begin{aligned}
G_0 = {}& N_k(\boldsymbol{\mu}_0; \mathbf{0}, \frac{1}{\xi_0}\Sigma_0)\mathrm{Wishart}(\Sigma_0^{0^{-1}}; \nu_0^0, B_0^0)\\
&\prod_{j=1}^{k} N_k(F_j; \mathbf{0}, \frac{1}{\alpha_j^0}Q^0)\mathrm{Wishart}(Q^{0^{-1}}; \nu_Q^0, B_Q^0)\mathrm{Gamma}(\alpha_j; \frac{1}{2}a_1^0, \frac{1}{2}b_{1j}^0)
\end{aligned}
$$

である. F_j は, 行列 F の第 j 列を表したものである. Wishart は, ウィッシャート分布を表す. また, ξ_0, ν_0^0, B_0^0, α_j^0, ν_Q^0, B_Q^0, a_1^0, b_{1j}^0 は既知とする. このようなディリクレ過程を導入した状態空間モデルを動的ノンパラメトリックベイズモデルとよぶ.

2 変分ベイズ法による推定

動的ノンパラメトリックベイズモデルにおけるパラメータ推定は, 一般に, マルコフ連鎖モンテカルロ法を用いたベイズ推定により行うことが多い. しかしながら, パラメータが高次元である場合などでは, 十分な精度を得るために多くの粒子を発生させる必要があり, 計算コストがかかることが知られている (Attias, 1999). そこで, 我々は変分ベイズ法を用いた事後分布の推定を提案する.

動的ノンパラメトリックベイズモデルにおいて, 変分ベイズ推定をするために二つの工夫を行う. まず, 有限次元の多項分布で近似する (Sethuraman, 1994; Blei and Jordan, 2006). そして, 所属クラスタを表すラベルとパラメータの変分分布が独立であると仮定する.

ディリクレ過程が無限次元の多項分布であることは, Sethuraman(1994) により知られている. このことを用いて, 変分ベイズ推定する際には, ディリクレ過程を G 次元の多項分布で近似する. この方法は, Truncated VB 法 (Blei and Jordan, 2006) と呼ばれる. すなわち, q を事後分布の近似分布としたとき, 以下のように仮定する:

$$
q(\tau_G = 1) = 1, \pi_g = 0, \quad \text{for } g > G.
$$

ここで, 観測モデルのパラメータ $\boldsymbol{\theta} = \{\boldsymbol{\beta}, H, R, \Sigma_{u_1}, \ldots, \Sigma_{u_N}\}$ に対して, 以下のような共役事前分布を導入する:

$$
\begin{aligned}
H &\sim \prod_{j=1}^{l} N_k(H_j; \mathbf{0}, \frac{1}{\beta_j}R_{jj}I_k),\ \beta_j \sim \mathrm{Gamma}(\beta_j; \frac{1}{2}a_2, \frac{1}{2}b_{2j}), j = 1, \ldots, l,\\
R^{-1} &\sim \mathrm{Wishart}(R^{-1}; \nu_1, B_R),\ \Sigma_{u_i}^{-1} \sim \mathrm{Wishart}(\Sigma_{u_i}^{-1}; \nu_{u_i}, B_{u_i}), i = 1, \ldots, N.
\end{aligned}
$$

ただし, I_k は k 次元単位行列である. また, 事後分布を近似する変分分布に以下の仮定を導入する:

$$q(\boldsymbol{x}_{0:T}^{1:G}, \boldsymbol{\psi}, \boldsymbol{\tau}, \boldsymbol{u}_{1:T}^{1:N}, \boldsymbol{z}^{1:N}, \boldsymbol{\theta}) = \prod_{g=1}^{G} \{q(\boldsymbol{x}_{0:T}^{g}|\boldsymbol{z}^{1:N}) q(\boldsymbol{\psi}^g) q(\tau_g)\} q(\boldsymbol{z}^{1:N}) q(\boldsymbol{\theta}),$$

$$q(\boldsymbol{\psi}^g) = q(\boldsymbol{\mu}_0^g|\Sigma_0^{g-1}) q(\Sigma_0^{g-1}) \prod_{j=1}^{k} q(F_j^g|\alpha_j^g, Q^g) q(\alpha_j^g) q(Q^g),$$

$$q(\boldsymbol{z}^{1:N}) = \prod_{i=1}^{N} q(\boldsymbol{z}^i), q(\boldsymbol{\theta}) = \prod_{j=1}^{l} q(H_j|\beta_j, R) q(\beta_j) q(R) \prod_{i=1}^{N} q(\Sigma_{u_i}).$$

以上二つの工夫により, 動的ノンパラメトリックベイズモデルにおける変分ベイズ法の目的関数は以下のように書き表せる:

$$F[q] = \left\langle \log \frac{p(\boldsymbol{u}_{1:T}^{1:N}, \boldsymbol{x}_{0:T}^{1:G}, \boldsymbol{y}_{1:T}^{1:N}, \boldsymbol{\theta}, \boldsymbol{\psi})}{q(\boldsymbol{u}_{1:T}^{1:N}, \boldsymbol{x}_{0:T}^{1:G}, \boldsymbol{\theta}, \boldsymbol{\psi})} \right\rangle_{q(\boldsymbol{u}_{1:T}^{1:N}, \boldsymbol{x}_{0:T}^{1:G}, \boldsymbol{\theta}, \boldsymbol{\psi})}.$$

ただし, $\langle f(x) \rangle_{q(x)}$ は $f(x)$ の $q(x)$ に対する期待値

$$\langle f(x) \rangle_{q(x)} = \int f(x) q(x) dx$$

を表すものとする.

この汎関数 $F[q]$ は, 観測データを固定したとき対数尤度を用いて以下の式で表される:

$$\log p(\boldsymbol{y}_{1:T}^{1:N}) = F[q] + \mathrm{KL}(q(\boldsymbol{x}_{0:T}^{1:G}, \boldsymbol{u}_{1:T}^{1:N}, \boldsymbol{\theta}, \boldsymbol{\psi}) || p(\boldsymbol{x}_{0:T}^{1:G}, \boldsymbol{u}_{1:T}^{1:N}, \boldsymbol{\theta}, \boldsymbol{\psi} | \boldsymbol{y}_{1:T}^{1:N}).$$

ただし, $\mathrm{KL}(q||p)$ は, 分布の近さを表すカルバック・ライブラー距離である. すなわち, 変分ベイズ法における目的関数である汎関数 $F[q]$ を最大にすることと, パラメータの事後分布と変分分布のカルバック・ライブラー距離を最小にすることは等しい.

汎関数 $F[q]$ を最大にするには, 汎関数の極値問題を解けばよい. いま, q は事後分布の近似であるため, 以下の制約条件の下で $F[q]$ を最大にする:

$$\int q(z^i) dz^i = 1, \quad \int q(\boldsymbol{u}_{1:T}^i) d\boldsymbol{u}_{1:T}^i = 1, \quad \int q(\boldsymbol{x}_{0:T}^g) d\boldsymbol{x}_{0:T}^g = 1,$$
$$\int q(\boldsymbol{\theta}) d\boldsymbol{\theta} = 1, \int q(\boldsymbol{\psi}) d\boldsymbol{\psi} = 1.$$

この極値問題の解は, オイラー・ラグランジュ方程式を解くことにより得られる. これらの方程式の解である変分分布は, 相互に依存する関係がある. したがって, 反復法により求める.

以上より, 変分ベイズ推定のアルゴリズムは次のように表される (アルゴリズム 1).

アルゴリズム 1 動的ノンパラメトリックベイズモデルにおける変分ベイズ推定

$Y^i \leftarrow \sum_{t=1}^T \boldsymbol{y}_t^i \boldsymbol{y}_t^{i\,T}$; パラメータの初期値を与える．

repeat

カルマンフィルタと固定区間平滑化により $q(\boldsymbol{x}_t^g)$, $q(\boldsymbol{u}_t^i)$ を推定する． ▷ VB E-step

$V_1^g \leftarrow \sum_{t=1}^T \left\langle \boldsymbol{x}_t^g \boldsymbol{x}_t^{g\,T}\right\rangle$; $V_2^g \leftarrow \sum_{t=1}^{T-1}\left\langle \boldsymbol{x}_t^g\boldsymbol{x}_t^{g\,T}\right\rangle$; $V_u^i \leftarrow \sum_{t=1}^T\left\langle \boldsymbol{u}_t^i\boldsymbol{u}_t^{i\,T}\right\rangle$; $C_x^g \leftarrow \sum_{t=1}^T \langle \boldsymbol{x}_t^g \boldsymbol{x}_{t-1}^{g\,T}\rangle$;

$C_{xu}^{gi} \leftarrow \sum_{t=1}^T\langle \boldsymbol{x}_t^g\boldsymbol{u}_t^{i\,T}\rangle$; $C_{xy}^{gi} \leftarrow \sum_{t=1}^T \langle\boldsymbol{x}_t^g\rangle \boldsymbol{y}_t^{i\,T}$; $C_{uy}^i \leftarrow \sum_{t=1}^t \langle \boldsymbol{u}_t^i\rangle \boldsymbol{y}_t^{i\,T}$;

$\mathcal{M}_1 \leftarrow Y - \langle H\rangle C_{xy}^{gi} - C_{xy}^{gi\,T}\langle H\rangle^T - C_{uy}^i - C_{uy}^{i\,T} + \langle H\rangle V_1^g \langle H\rangle^T + \langle H\rangle C_{xu}^{gi} + C_{xu}^{gi}\langle H\rangle^T + V_u^i$;

$\mathcal{M}_2 \leftarrow V_1^g - \langle F^g\rangle C_x^g - C_x^g\langle F^g\rangle^T + \langle F\rangle V_2^g\langle F\rangle^T$; $\mathcal{M}_3 \leftarrow \langle \boldsymbol{x}_0^g\boldsymbol{x}_0^{g\,T}\rangle - 2\langle\boldsymbol{\mu}_0^g\rangle\langle\boldsymbol{x}_0^g\rangle^T + \langle \boldsymbol{\mu}_0^g\boldsymbol{\mu}_0^{g\,T}\rangle$;

$\rho_i^g \leftarrow -\frac{1}{2}(\langle\log|Q^g|\rangle + \langle \log|R|\rangle + \langle\log|\Sigma_0^g|\rangle + \mathcal{M}_1+\mathcal{M}_2+\mathcal{M}_3 + \langle\log\tau_g\rangle + \langle\log(1-\tau_g)\rangle)$

$\bar{z}_{ig} = q(z_{ig}=1) \leftarrow \exp(\rho_i^g)/(\sum_g \exp(\rho_{ig}))$; $N^g \leftarrow \sum_{i=1}^N \bar{z}_{ig}$

$\omega_{g,1} \leftarrow 1 + N^g$; $\omega_{g,2}\leftarrow \zeta + \sum_{i=1}^N\sum_{j=g+1}^G \bar{z}_{ij}$ ▷ VB M-step

$\langle\boldsymbol{\mu}_0^g\rangle \leftarrow N^g/(N^g+\xi_0^g)\langle \boldsymbol{x}_0^g\rangle$; $\Delta_0^g \leftarrow N^g(\langle \boldsymbol{x}_0^g\boldsymbol{x}_0^{g\,T}\rangle - 2\langle\boldsymbol{\mu}_0^g\rangle\langle\boldsymbol{x}_0^g\rangle^T + \langle\boldsymbol{\mu}_0^g\boldsymbol{\mu}_0^{g\,T}\rangle)$

$\langle\Sigma_0^g\rangle \leftarrow (N^g + \nu_0^g + 1)\Delta_0^{g-1}$; $\langle\alpha_j^g\rangle \leftarrow \frac{1}{4}(a_1^g+1)(b_{1,j} + \mathrm{tr}\ \langle Q^g\rangle^{-1}\langle F_jF_j^T\rangle)$

$\Delta_Q^g \leftarrow V_1^g - \langle F^g\rangle C_x^{g\,T} - C_x^g\langle F^g\rangle^T + \langle F\rangle V_2^g\langle F\rangle^T + \sum_j\langle\alpha_j\rangle\langle F_jF_j^T\rangle + B_Q$

$\mathcal{A}_j^g \leftarrow \langle\alpha_j^g\rangle I_k + N^g V_2^g$ $\langle F\rangle \leftarrow \mathcal{A}_j^{g-1}N^gC_x^{g\,T}$; $\langle F_jF_j^T\rangle \leftarrow \mathcal{A}_j^{g-1}\langle Q^g\rangle + \langle F_j\rangle\langle F_j\rangle^T$

$\Delta_R \leftarrow \sum_i(Y^i - C_{uy}^{i\,T} - C_{uy}^i + V_u^i) - \sum_{i,g}\bar{z}_{ig}(\langle H\rangle C_{xy}^{gi} - C_{xy}^{gi\,T}\langle H\rangle^T + \langle H\rangle C_{xu}^{gi} + C_{xu}^{gi}\langle H\rangle^T) + \sum_g N^g\langle H\rangle V_1^g\langle H\rangle^T + B_R + \mathrm{diag}(\mathrm{tr}\ \langle\beta_1\rangle\langle H_1H_1^T\rangle, \dots, \mathrm{tr}\ \langle\beta_l\rangle\langle H_lH_l^T\rangle)$

$\langle R\rangle \leftarrow (\nu_R + T\sum_g N^g + l)\Delta_R^{-1}$; $\langle\beta_j\rangle \leftarrow \frac{1}{4}(a_2+1)(b_{2j} + \mathrm{tr}\ \langle R\rangle_{jj}^{-1}\langle H_jH_j^T\rangle)$;

$\langle H_j\rangle \leftarrow (\langle\beta_j\rangle+1)\langle R\rangle_{jj}^{-1}\sum_{i,g}\bar{z}_{ig}(C_{xy}^{gi} - C_{ux}^{gi\,T})\boldsymbol{e}_j$; $\langle H_jH_j^T\rangle \leftarrow (\langle\beta_j\rangle+1)^{-1}\langle R\rangle_{jj}I_k + \langle H_j\rangle\langle H_j\rangle^T$;

$\langle\Sigma_{u_i}\rangle \leftarrow (\nu_{u_i}+T)(V_u^i + B_{u_i})^{-1}$;

until $F[q]$ converged

参考文献

Antoniak, C. E. (1974). Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The annals of statistics*, 2(6):1152–1174.

Attias, H. (1999). Inferring parameters and structure of latent variable models by variational bayes. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 21–30. Morgan Kaufmann Publishers Inc.

Blei, D. and Jordan, M. (2006). Variational inference for dirichlet process mixtures. *Bayesian Analysis*, 1(1):121–143.

Chiappa, S. and Barber, D. (2007). Dirichlet mixtures of bayesian linear gaussian state-space models: a variational approach. Technical Report 161, Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4(2):639–650.