

大規模スイッチング線形動的システムの確率的変分推論

Stochastic Variational Inference for Big Scale Switching Linear Dynamical System

岡 滉 河原 吉伸 鷲尾 隆
Akira Oka Yoshinobu Kawahara Takashi Washio

大阪大学 産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

Switching Linear Dynamical System is useful for analyzing complex time series data, but the inference with large scale datasets is really tough. Stochastic Variational Inference is new approach for approximating model parameters with using minibatches of small size. There is no other method for inferring big scale Switching Linear Dynamical System so we apply Stochastic Variational Inference to the model and we show the experimental result with artificial data.

1. はじめに

近年、時系列に沿った複雑なデータをモデル化する際にスイッチング線形動的システムが度々利用されている [1,2,3]. 例えば、蜂の動き [1] や、新生児の敗血症の検知 [2], 株価 [3] といった様々なデータに対して同モデルが適用されている。しかしながら、大規模なデータセットに対してこのモデルを用いて推定を行うことは困難であることが指摘されている [3].

他のモデルにおいても大規模なデータセットを用いたモデルの推定は盛んに研究されており、近年では D.Hoffman らはトピックモデルに対して確率的変分推論を用いて推定を行う手法を示している [4]. また、J.Johnson, J.Foti らはこの確率的変分推論を時系列モデルである隠れマルコフモデルに拡張し推定を行う手法を提案している [5,6].

本研究では、確率的変分推論をスイッチング線形動的システムに拡張することにより、従来の変分推論に比べて、より大規模なデータセットを用いた推論を可能にし、計算時間を短縮する手法について提案する。従来の変分推論では、推論の繰り返しに全データに対してモード、潜在変数の推定を必要とするが、本手法では、パラメータの推論の際に全データを用いるのではなく、全データからミニバッチをサンプルし、そのミニバッチのモード、潜在変数の推定のみを必要とするため、大規模なデータセットに対してより効率的な推論が可能となっている。

本稿の構成は、以下の通りである。まず 2.1. では、対象とするモデルであるスイッチング線形動的システムについて述べる。次に 2.2. では従来の推論手法について述べる。次に 3.1. ではスイッチング線形動的システムに変分推論を適用した場合について述べる。3.2. では更に、確率的変分推論をスイッチング線形動的システムに対して拡張し、パラメータを推定する手法について述べる。3.3. では具体的な更新方法と提案手法アルゴリズムについて示す。4. では提案手法を人工データに適用した評価実験結果を示し、5. において結論を述べる。

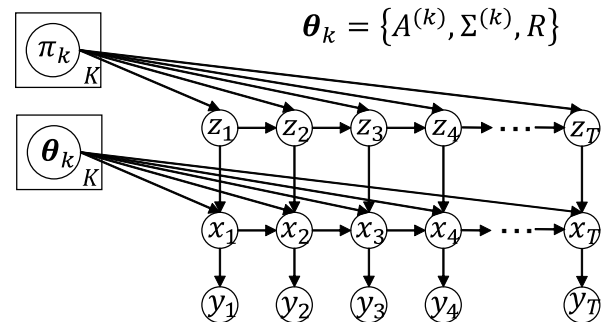


図 1: スwitching線形動的システムのグラフィカルモデル

2. 従来研究

2.1 スwitching線形動的システム

ある時刻 t の観測 $\mathbf{y}_t (\in \mathbb{R}^p)$ ($t = 1, \dots, T$) と、同時刻における潜在変数 $\mathbf{x}_t (\in \mathbb{R}^d)$ は、次式のような線形モデルで表されるとする。

$$\begin{aligned}\mathbf{x}_t &= \mathbf{A}^{(z_t)} \mathbf{x}_{t-1} + \mathbf{e}_t^{(z_t)} \\ \mathbf{y}_t &= \mathbf{C} \mathbf{x}_t + \mathbf{w}_t\end{aligned}$$

ただし、 $\mathbf{A}^{(z_t)} \in \mathbb{R}^{d \times d}$ は潜在変数の遷移モデルを表す行列、 $\mathbf{C} \in \mathbb{R}^{p \times d}$ は線形な観測モデルを表す行列である。ここで、 \mathbf{C} はモード遷移にかかわらず、固定された値であるため、 \mathbf{I} を単位行列として、 $\mathbf{C} = [\mathbf{I} \ \mathbf{O}]$ としても一般性を失わないことが知られており [3], \mathbf{C} についての推定は行わない。また、 $\mathbf{e}_t^{(z_t)} \in \mathbb{R}^d, \mathbf{w}_t \in \mathbb{R}^p$ はそれぞれ潜在変数遷移と観測に伴うガウス雑音を表す。 $\mathbf{A}^{(z_t)}, \mathbf{e}_t^{(z_t)}$ はモード z_t に依存しており、ガウス雑音 $\mathbf{e}_t^{(z_t)}$ の分散を $\Sigma^{(z_t)}$, \mathbf{w}_t の分散を \mathbf{R} とする。このモード z_t の遷移は次式のように表される。

$$Pr(z_{t+1} = e_k | \bar{\mathbf{z}}_t) = \pi_k^\top \bar{\mathbf{z}}_t$$

ただし、 $\bar{\mathbf{z}}_t \in \mathbb{R}^K$ (K はモード数) はその j 番目の要素が $Pr(z_t = e_j)$ (e_j は j 番目のモードを表す) となる確率ベクトルであり、時刻 t におけるモードの確率を表す。また $\pi_k \in \mathbb{R}^K$ はその j 番目の要素がモード e_j から e_k に遷移する確率となるような確率ベクトルである。

このモデルをグラフィカルモデルで表したものが図1である。

2.2 スイッチング線形動的システムの推論

スイッチング線形動的システムの推論の目的はモード \mathbf{Z} 、及び、潜在変数 \mathbf{X} の事後分布を与えられた観測 \mathbf{Y} から推定することにある。特に事後分布 $Pr(\mathbf{X}, \mathbf{Z} | \mathbf{Y})$ の十分統計量を見つけることが必要となる。もしモードのスイッチがないと仮定すれば、線形動的システムの推論 (RTS スムージング [8]) を用いることで観測 \mathbf{Y} より潜在変数 \mathbf{X} を簡単に推定することが可能となる。しかしながら、モードを考慮することによってモデルの複雑さは大幅に増し、正確な推定は困難を極める。もし仮に最初の潜在変数 \mathbf{x}_1 の分布が多変量正規分布だとすると、モード数を K として、 \mathbf{x}_2 は K 個の多変量正規分布の混合となり、 \mathbf{x}_t にいたっては K^t 個の多変量正規分布の混合となり、ある程度以上のシーケンスがある場合、明らかに計算が不可能であることがわかる。そのため近年ではおおよそその推定結果を得るための手法の研究が盛んに行われている。

中でも代表的な手法として、事後分布を近似し、制約を加える事によって解析的に求める手法である変分推論法 (変分ベイズ法) [9] や、マルコフ連鎖を利用することで有限個のパラメータをサンプリングより生成し、近似的に事後分布を求める手法であるマルコフ連鎖モンテカルロ法などがあり [10]、近年ではパラメータ次元を変更しながらサンプリングを行う手法も提案されている [11]。

3. スイッチング線形動的システムの確率的変分推論

3.1 変分推論

ここでの問題は前節で述べたモデルに基づいて事後確率分布 $Pr(\mathbf{X}, \mathbf{Z} | \mathbf{Y})$ を推定することである。この分布の厳密な計算は困難であるため、ここでは平均場近似に基づいて事後分布を近似することで計算を行う。

事後分布を近似した分布を $Q(\mathbf{X}, \mathbf{Z})$ として、次のように分離が可能であると仮定する。

$$Pr(\mathbf{X}, \mathbf{Z} | \mathbf{Y}) \approx Q(\mathbf{X}, \mathbf{Z}) = Q(\mathbf{X})Q(\mathbf{Z}) \quad (1)$$

さらに、イェンゼンの不等式を用いて、尤度 $Pr(\mathbf{Y})$ は以下のような下限を持つことがわかる。

$$\begin{aligned} \log Pr(\mathbf{Y}) &= \log \int_{\mathbf{X}, \mathbf{Z}} Q(\mathbf{X}, \mathbf{Z}) \frac{Pr(\mathbf{Y}, \mathbf{X}, \mathbf{Z})}{Q(\mathbf{X}, \mathbf{Z})} d\mathbf{X} d\mathbf{Z} \\ &\geq \int_{\mathbf{X}, \mathbf{Z}} Q(\mathbf{X}, \mathbf{Z}) \log \frac{Pr(\mathbf{Y}, \mathbf{X}, \mathbf{Z})}{Q(\mathbf{X}, \mathbf{Z})} d\mathbf{X} d\mathbf{Z} \\ &\triangleq \mathcal{L}(Q) \end{aligned}$$

さらに、(1) より、上式は以下のように変形できる。

$$\begin{aligned} \mathcal{L}(Q) &= \int_{\mathbf{X}, \mathbf{Z}} Q(\mathbf{X})Q(\mathbf{Z}) \log Pr(\mathbf{Y}, \mathbf{X}, \mathbf{Z}) d\mathbf{X} d\mathbf{Z} \\ &\quad - \int_{\mathbf{X}, \mathbf{Z}} Q(\mathbf{X})Q(\mathbf{Z}) \log Q(\mathbf{X}, \mathbf{Z}) d\mathbf{X} d\mathbf{Z} \\ &= \int_{\mathbf{X}} Q(\mathbf{X}) \langle \log Pr(\mathbf{Y}, \mathbf{X}, \mathbf{Z}) \rangle_{Q(\mathbf{Z})} d\mathbf{X} \\ &\quad + \int_{\mathbf{Z}} Q(\mathbf{Z}) \langle \log Pr(\mathbf{Y}, \mathbf{X}, \mathbf{Z}) \rangle_{Q(\mathbf{X})} d\mathbf{Z} \\ &\quad - \int_{\mathbf{Z}} Q(\mathbf{Z}) \log Q(\mathbf{Z}) d\mathbf{Z} - \int_{\mathbf{X}} Q(\mathbf{X}) \log Q(\mathbf{X}) d\mathbf{X} \end{aligned}$$

ただし、 $\langle \cdot \rangle_Q$ は期待値を表す。したがって次式のように各変数ごとに更新を行うことで尤度の下限を最大化できる。

$$\begin{aligned} Q(\mathbf{X}) &= \frac{1}{c_x} \exp(\log Pr(\mathbf{X}, \mathbf{Z}, \mathbf{Y}))_{Q(\mathbf{Z})} \\ Q(\mathbf{Z}) &= \frac{1}{c_y} \exp(\log Pr(\mathbf{X}, \mathbf{Z}, \mathbf{Y}))_{Q(\mathbf{X})} \end{aligned}$$

ただし、 c_x, c_y は正規化定数である。

3.2 パラメータの確率的変分推論

パラメータの更新も前節の変分推論を用いることで可能であるが、前節の変分推論では変数の更新のたびに全データのモード、潜在変数を必要とする。大規模なデータセットである場合、冗長な情報を含むデータも多く存在し、その場合、計算コストは大きく、無駄となる。この問題に対処するために確率的変分推論 [4] では、ロビンソンモンローアルゴリズム [7] を用いた確率的な勾配法を用いることで下限の最大化を図り、計算量の削減を行っている。

本論文では、モード \mathbf{Z} 、潜在変数 \mathbf{X} は従来の変分推論から推定し、パラメータを確率的変分推論によって推定を行うため、本節ではパラメータ集合 θ を含んだ事後確率分布 $Pr(\mathbf{X}, \mathbf{Z}, \theta | \mathbf{Y})$ を考え、パラメータの更新式について導く。

前節と同様にイェンゼンの不等式を用いることで尤度 $Pr(\mathbf{Y})$ の下限は以下のように表される。

$$\begin{aligned} \log Pr(\mathbf{Y}) &\geq \int_{\mathbf{X}, \mathbf{Z}, \theta} Q(\mathbf{X}, \mathbf{Z}, \theta) \log \frac{Pr(\mathbf{Y}, \mathbf{X}, \mathbf{Z}, \theta)}{Q(\mathbf{X}, \mathbf{Z}, \theta)} d\mathbf{X} d\mathbf{Z} d\theta \\ &\triangleq \mathcal{L}_\theta(Q) \end{aligned}$$

ここで、計算の簡略化のため、 $Pr(\theta)$ 、 $Pr(\mathbf{Y}, \mathbf{X}, \mathbf{Z}, \theta)$ が指数族分布かつ共役であると仮定し、次式を得る。

$$\begin{aligned} \log Pr(\theta) &= \langle \eta_\theta, t_\theta(\theta) \rangle - A_\theta(\eta_\theta) \\ \log Pr(\mathbf{Z}, \mathbf{X}, \mathbf{Y} | \theta) &= \langle \eta_\theta, t_{zxy}(\mathbf{Z}, \mathbf{X}, \mathbf{Y}) \rangle \\ &\quad - A_{zxy}(\eta_{zxy}(\theta)) \end{aligned}$$

ただし、 $\langle \cdot \rangle$ は内積を表し、 η_θ は自然パラメータ、 $t_\theta(\theta)$ 、 $t_{zxy}(\mathbf{Z}, \mathbf{X}, \mathbf{Y})$ はそれぞれ十分統計量、 $A_\theta(\eta_\theta)$ 、 $A_{zxy}(\eta_{zxy}(\theta))$ はそれぞれログノーマライザを表す。さらに近似した分布 $Q(\theta)$ も同じ指数族分布かつ共役であると考え、 $\log Q(\theta) = \langle \tilde{\eta}_\theta, t_\theta(\theta) \rangle - A_\theta(\tilde{\eta}_\theta)$ とする。この仮定のもと、 $\mathcal{L}_\theta(Q)$ を変形し、後の勾配の計算において $\tilde{\eta}_\theta$ による偏微分を行うため、 $\tilde{\eta}_\theta$ に関わらない部分を const とすることで以下の式を得る。

$$\begin{aligned} \mathcal{L}_\theta(Q) &= \langle \eta_\theta + \langle t_{zxy}(\mathbf{Z}, \mathbf{X}, \mathbf{Y}) \rangle_{Q(\mathbf{Z})Q(\mathbf{X})}, \nabla A(\tilde{\eta}_\theta) \rangle \\ &\quad - \langle \tilde{\eta}_\theta, \nabla A(\tilde{\eta}_\theta) \rangle + A(\tilde{\eta}_\theta) + \text{const} \end{aligned}$$

$\tilde{\eta}_\theta$ についての自然勾配を $\tilde{\nabla}_{\tilde{\eta}_\theta}$ とすると、

$$\tilde{\nabla}_{\tilde{\eta}_\theta} \mathcal{L}_\theta(Q) = \eta_\theta + \langle t_{zxy}(\mathbf{Z}, \mathbf{X}, \mathbf{Y}) \rangle_{Q(\mathbf{Z})Q(\mathbf{X})} - \tilde{\eta}_\theta$$

さらに、全データを長さ L の M 個のミニバッチにわけたとき、 $\mathbf{Y} = \{\mathbf{y}^{(m)}\}_{m=1}^M$ とすると、

$$\begin{aligned} \tilde{\nabla}_{\tilde{\eta}_\theta} \mathcal{L}_\theta(Q) &= \eta_\theta - \tilde{\eta}_\theta \\ &\quad + \sum_{m=1}^M \langle t_{zxy}(\mathbf{z}^{(m)}, \mathbf{x}^{(m)}, \mathbf{y}^{(m)}) \rangle_{Q(\mathbf{Z})Q(\mathbf{X})} \end{aligned}$$

Algorithm 1 スイッチング線形動的システムの確率的変分推定アルゴリズム

Input: positive integer s .

- 1: Initialize $\eta_{\theta}, \langle \mathbf{x}_0 \rangle, \langle \Sigma^{z_0} \rangle, \langle \mathbf{z}_t \rangle$ and choose stepsize $\rho^{(n)}, n = 1, 2, \dots$
- 2: **loop**
- 3: Sample a subchain $\mathbf{y}^{(m)} \in \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(M)}$
- 4: **for** $i = 1$ to s **do**
- 5: Compute the locally-linearized parameter $\Sigma_{t,i}, \mathbf{A}_{t,i}$
- 6: Compute $\hat{\Sigma}_t, \hat{\mathbf{A}}_t, \hat{\mathbf{x}}_0$ from $\langle \mathbf{z}_t \rangle$ using Eqs.(3).
- 7: Estimate $\langle \mathbf{x}_t \rangle$ by KarmanFilterSmoothing($\hat{\Sigma}_t, \hat{\mathbf{A}}_t, \hat{\mathbf{x}}_0$).
- 8: Compute $\langle \mathbf{q}_t \rangle$ from $\langle \mathbf{x}_t \rangle$ using Eqs.(4).
- 9: Estimate $\langle \mathbf{z}_t \rangle$ by ForwardBackward($\langle \mathbf{q}_t \rangle, \theta$)
- 10: **end for**
- 11: Compute $\langle t_{zxy}(\mathbf{z}^{(m)}, \mathbf{x}^{(m)}, \mathbf{y}^{(m)}) \rangle_{Q(\mathbf{Z})Q(\mathbf{X})}$ from $\langle \mathbf{z}_t \rangle, \langle \mathbf{x}_t \rangle$.
- 12: Update η_{θ} from $\langle t_{zxy}(\mathbf{z}^{(m)}, \mathbf{x}^{(m)}, \mathbf{y}^{(m)}) \rangle_{Q(\mathbf{Z})Q(\mathbf{X})}$ using Eqs.(2).
- 13: **end loop**

ここで、ロビンソンモンローアルゴリズム [7] を用いることによって、イテレーション回数を n 、ステップサイズを $\rho^{(n)}$ として以下の更新式が求まる。

$$\begin{aligned} \tilde{\eta}_{\theta} &\leftarrow (1 - \rho^{(n)})\tilde{\eta}_{\theta} \\ &+ \rho^{(n)} \left(\eta_{\theta} + s \cdot \langle t_{zxy}(\mathbf{z}^{(m)}, \mathbf{x}^{(m)}, \mathbf{y}^{(m)}) \rangle_{Q(\mathbf{Z})Q(\mathbf{X})} \right) \end{aligned} \quad (2)$$

ただし、 $s \triangleq |\mathbf{Y}|/|\mathbf{y}^{(m)}|$ とした。

3.3 スイッチング線形動的システムの確率的変分推定アルゴリズム

ここでは各事後分布 $Q(\mathbf{X}), Q(\mathbf{Y})$ の具体的更新手法及び、アルゴリズムについて述べる。2.2 による変分推定を全データに対して行う場合、データ数が大きくなった際に、計算コストが爆発的に増える。そこで、全データをミニバッチに区切り、イテレーションごとにランダムなミニバッチをサンプリングし、そのミニバッチ内のモード \mathbf{Z} 、潜在変数 \mathbf{X} の推定を従来の変分推論を s 回繰り返すことで推定した後、得られたモード、潜在変数を利用して確率的にパラメータの変分推論を行う。

本手法の擬似コードをアルゴリズム 1 に示す。

3.3.1 $Q(\mathbf{X})$ の更新

固定された $Q(\mathbf{Z})$ の期待値を用いて \mathbf{X} に関する対数尤度を評価することで計算される。非線形性を扱うため、各モードに対して局所線形化を施し、モーメントマッチング [12] を用いた近似によってこの計算が可能となる。具体的には以下の手順を踏む。

通常の線形動的システムにおける推論と同様の推論を行うために拡張カルマンフィルタ [13] と同様の局所線形化を施し、更にモードに伴う組合せ爆発への対処としてモーメントマッチングを用いる。線形化されたスイッチング線形動的システムは以下の様に表わされる。

$$\begin{aligned} \mathbf{x}_{t+1} &= \hat{\mathbf{A}}_t \mathbf{x}_t + \bar{\mathbf{e}}_t \\ \mathbf{y}_t &= \mathbf{C} \mathbf{x}_t + \mathbf{w}_t \end{aligned}$$

ただし、 $\bar{\mathbf{e}}_t$ はガウス雑音であり、その分散は $\hat{\Sigma}_t$ であるとす

る。このとき、これらのパラメータは次式のように計算される。

$$\begin{aligned} \hat{\Sigma}_{L-1}^{-1} &= \langle \Sigma_i^{-1} \rangle_{Q(\mathbf{z}_i)} \\ \hat{\Sigma}_t^{-1} &= \langle \Sigma_i^{-1} \rangle_{Q(\mathbf{z}_t)} + \langle \mathbf{A}_{t,i}^{\top} \Sigma_i^{-1} \mathbf{A}_{t,i} \rangle_{Q(\mathbf{z}_{t+1})} \\ &\quad - \hat{\mathbf{A}}_t^{\top} \hat{\Sigma}_1^{-1} \hat{\mathbf{A}}_t \quad (0 < t < L-1) \\ \hat{\Sigma}_0^{-1} &= \langle \Sigma_{0,i}^{-1} \rangle_{Q(\mathbf{z}_0)} + \langle \mathbf{A}_{0,i}^{\top} \Sigma_i^{-1} \mathbf{A}_{0,i} \rangle_{Q(\mathbf{z}_1)} \\ &\quad - \hat{\mathbf{A}}_1^{\top} \hat{\Sigma}_1^{-1} \hat{\mathbf{A}}_1 \\ \hat{\mathbf{A}}_t &= \hat{\Sigma}_t \langle \Sigma_i^{-1} \mathbf{A}_{t,i} \rangle_{Q(\mathbf{z}_t)} \\ \hat{\mathbf{x}}_0 &= \hat{\Sigma}_0 \langle \Sigma_{0,i}^{-1} \mathbf{x}_{0,i} \rangle_{Q(\mathbf{z}_t)} \end{aligned} \quad (3)$$

$\Sigma_{t,i}, \mathbf{A}_{t,i}$ はそれぞれ各モード i に関して局所線形化されたパラメータを表す。こうして得られたパラメータを用いて線形動的システムと同様の推論を行うことで $\langle \mathbf{x}_t \rangle$ が得られる、

3.3.2 $Q(\mathbf{Z})$ の更新

$Q(\mathbf{Z})$ の更新は、前小節から得られたパラメータを用いて計算される変分パラメータ $q_t(i)$ を観測とする HMM の推論を行うことによって計算される。 $q_t(i)$ は以下のように計算される。

$$\begin{aligned} \log q_0(i) &= -\frac{1}{2} \langle (\mathbf{x}_0 - \mathbf{x}_{0,i})^{\top} \Sigma_{0,i}^{-1} (\mathbf{x}_0 - \mathbf{x}_{0,i}) \rangle \\ &\quad - \frac{1}{2} \log |\Sigma_{0,i}| \\ \log q_t(i) &= -\frac{1}{2} \langle (\mathbf{x}_t - \mathbf{A}_{t,i} \mathbf{x}_{t,i})^{\top} \Sigma_{t,i}^{-1} (\mathbf{x}_t - \mathbf{A}_{t,i} \mathbf{x}_{t,i}) \rangle \\ &\quad - \frac{1}{2} \log |\Sigma_{t,i}| \end{aligned} \quad (4)$$

こうして得られた $q_t(i)$ を用いて HMM の推定量としてモードの推定量 $\langle \mathbf{z}_t \rangle$ が得られる。

3.3.3 パラメータの更新

それぞれのパラメータの事前分布を以下のように設定する。

$$\begin{aligned} \pi_i &= \{\pi_{i1}, \dots, \pi_{ik}\} \sim Dir(\boldsymbol{\alpha}^{(i)}) \\ \phi &= (A^{(k)}, \Sigma^{(k)}) \sim MNIW(\mathbf{M}, \mathbf{C}, \mathbf{S}, h) \\ R &\sim IW(r_0, \mathbf{R}_0) \end{aligned}$$

それぞれに対してこれまでに求められた $\langle \mathbf{x}_t \rangle, \langle \mathbf{z}_t \rangle$ を用いて十分統計量を計算し、(2) 式に当てはめることで更新式を得ることができる。例えば遷移パラメータについては以下の更新式が得られる。

$$\begin{aligned} \tilde{\boldsymbol{\alpha}}^{(i)} &\leftarrow (1 - \rho^{(n)})\tilde{\boldsymbol{\alpha}}^{(i)} + \rho^{(n)}(\boldsymbol{\alpha}^{(i)} + s \cdot \hat{\mathbf{t}}_{\boldsymbol{\alpha}}^{(i)}) \\ (\hat{\mathbf{t}}_{\boldsymbol{\alpha}}^{(i)})_k &\triangleq \left\langle \sum_{t=1}^{L-1} \mathbb{I}[z_t^{(m)} = i, z_{t+1}^{(m)} = k] \right\rangle_{Q(z_{1:L}^{(m)})} \\ &= \sum_{t=1}^{L-1} Q(z_t^{(m)} = i, z_{t+1}^{(m)} = k) \end{aligned}$$

4. 評価実験

人工データを用いて実験を行った。用意したデータは観測の次元 $p = 3$ 、潜在変数の次元 $d = 4$ 、モード数 $K = 2$ とし、 $T = 30000$ のシークエンスを生成し、ランダムなサブシークエンス 90% を訓練データとして、残りの 10% をテスト用データとして用いる手続きを 10 回行った。パフォーマンス評価のため以下の様な、テストデータに対するおよその対数尤度を用

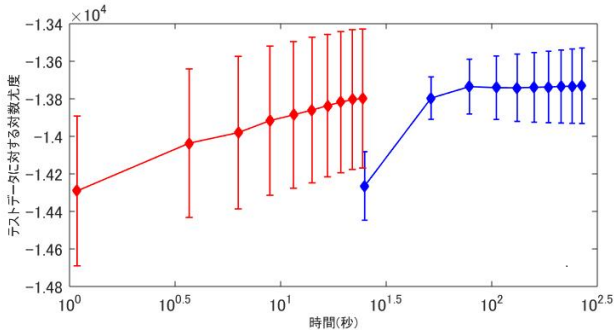


図 2: 提案手法 (赤線) と従来手法 (青線) の計算時間 (横軸), テストデータに対する対数尤度 (縦軸) の比較

いた.

$$\begin{aligned} Pr(\mathbf{y}_{test} | \mathbf{y}_{train}) &= \int \int Pr(\mathbf{y}_{test} | \pi, \theta) Pr(\pi, \theta | \mathbf{y}_{train}) d\pi d\theta \\ &\simeq \langle Pr(\mathbf{y}_{test} | \pi, \theta) \rangle_{Q(\pi)Q(\theta)} \end{aligned}$$

提案手法である確率的変分推定と従来手法であるバッチ式の変分推定を用いて, テストデータに対する対数尤度を評価し, テストデータに対する対数尤度を縦軸に, 計算時間を横軸に取り, 中央値と誤差範囲を示したものが図 2 である. ただし, 提案手法では一つのミニバッチのシークエンス数 $L = 500$, モード, 及び, 潜在変数推定のための変分推定のイタレーション回数 $s = 3$ として実験を行い, どちらの手法でも 10 回変数を更新するまで行った. また, パラメータ Σ に関しては, 推定がうまくいかなかったため, どちらの手法でも固定した値を用いている. 図 2 より, 提案手法の計算時間が従来手法であるバッチ式のものに比べ大幅に早いことがわかる. 一方で, 従来手法に比べ誤差範囲が大きくなっている. これは選んだミニバッチのモードに偏りが有り, うまく学習できない場合があったためではないかと思われる.

5. まとめ

本稿では, 大規模なスイッチング線形動的システムの推論手法について述べた. 従来の変分推論の問題点である, イタレーションごとに全モード, 潜在変数の推定が必要となっていた事に着目し, 確率的にパラメータを変分推論する手法について提案した. 評価実験よりシークエンス数の長い大規模なデータセットにおいて提案手法が有効的な推論を行えることを示した. 今後は提案手法で推定がうまくいかなかった一部のパラメータの推定, 及び, 人工データではなく現実問題に対して提案手法を適用することを目標とする.

参考文献

- [1] S. Oh, J. Rehg, T. Balch, and F. Dellaert. *Learning and inferring motion patterns using parametric segmental switching linear dynamic systems*, IJCV, **77**(13), 2008, 103-124.
- [2] C. K. Williams, and I. Stanculescu. *Automating the Calibration of a Neonatal Condition Monitoring System*, In Artificial Intelligence in Medicine, Springer, 2011, 240–249.

- [3] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, *Nonparametric Bayesian Learning of Switching Linear Dynamical Systems*, Neural Information Processing Systems, 2008.
- [4] D. Hoffman, M. Blei, Chong. Wang, and John. Paisley, *Stochastic variational inference*, Statistical Analysis and Data Mining, **14**, 2013, 13031347.
- [5] M. J. Johnson and A. S. Willsky, *Stochastic variational inference for Bayesian time series models*, ICML, 2014.
- [6] N. Foti, J. Xu, D. Laird, and E. B. Fox, *Stochastic variational inference for hidden Markov models*, Advances in Neural Information Processing Systems, **27**, 2014.
- [7] H. Robbins, and S. Monro, *A stochastic approximation method*, The annals of Mathematical Statistics, **22**(3), 2014, 400-407.
- [8] B. D. O. Anderson and J. B. Moore, *Optimal filtering*, 1979.
- [9] H. Attias, *Learning parameters and structure of latent variable models by variational Bayes*, Uncertainty in Artificial intelligence, 1999.
- [10] D. Gamerman, *Markov chain Monte Carlo*, Chapman & Hall, 1979.
- [11] S. Richardson and P. Green, *Bayesian analysis of mixtures with an unknown number of components*, Journal of the Royal Statistical Society B, **59**, 1997, 731-792.
- [12] A. Smith and U. Markov, *Bayesian detection and estimation of jumps in linear systems*, Analysis and Optimization of Stochastic Systems, 1980.
- [13] V. pavlovic and J. M. Rehg, *Impact of dynamic model learning on classification of human motion*, Proc. of IEEE 2000 Conf. on Computer Vision and Pattern Recognition, 2000, 1788-1795.