

Robust parameter estimation of state space models

石塚 治也 総合研究大学院大学 統計科学専攻 博士課程(3年次編入学)2年

1. 概要

状態空間モデル(State Space Model: SSM)は潜在状態 $x_{1:T} = (x_1, \dots, x_T)$ 、観測 $y_{1:T} = [y_1, \dots, y_T]$ の時間発展を記述する統計モデルであり幅広い分野で応用されている。その主たる興味は状態推定や将来予測だが、これらのタスクはパラメータの推定精度に大きく左右される。SSMのパラメータ推定には最尤法が広く用いられるが、 $y_{1:T}$ に外れ値が含まれていると推定値がずれやすい。本報告では、この問題に対処するため、 γ -divergence [1]を活用することで外れ値が多く含まれていても妥当な推定が可能である推定量と、ロスの単調減少性が保証されるパラメータ推定アルゴリズムを提案する

2. 背景・問題設定

背景: ダイバージェンスにもとづく推定量

- ダイバージェンス D : 二つの確率分布 p, q の類似度を測る尺度
 - 性質: (i) $D(p, q) \geq 0$, (ii) $D(p, q) = 0 \Leftrightarrow p = q$
- 相互エントロピー d を使った D の分解: $D(p, q) = d(p, q) - d(p, p)$
 - 性質: (i) $d(p, q) \geq d(p, p)$, (ii) $d(p, p) = d(q, q) \Leftrightarrow p = q$
- パラメータ θ の推定: 相互エントロピー $d(g, f_\theta)$ を最小化するように θ を決定

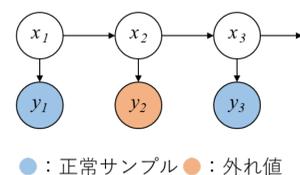
$$\hat{\theta} = \operatorname{argmin}_\theta d(g, f_\theta) = \operatorname{argmin}_\theta D(g, f_\theta)$$
 - 二つ目の等号は $d(g, g)$ が θ に依存しないことから得られる

問題: 観測 $y_{1:T}$ に外れ値が含まれる場合のロバスト推定量の構築

- 生成過程: 割合 ϵ ($0 < \epsilon < 1$) で観測に外れ値が混入

$$g(x_t | x_{t-1}) \sim \mathcal{N}(x_t; A^* x_{t-1}, \Sigma_x^*),$$

$$g(y_t | x_t) \sim (1 - \epsilon) \mathcal{N}(y_t; B^* x_t, \Sigma_y^*) + \epsilon \delta(y_t), \quad 0 \leq \epsilon \leq 1$$



- 統計モデル: $f_\theta(x_t | x_{t-1}) \sim \mathcal{N}(x_t; A x_{t-1}, \Sigma_x)$, $f_\theta(y_t | x_t) \sim \mathcal{N}(y_t; B x_t, \Sigma_y)$

研究目標: 上述の状況で、外れ値の影響を受けにくい θ の推定量を構築

3. 提案推定量

- 相互エントロピー: KL, γ -divergenceに関する相互エントロピーの和:

$$d(g, f) = d_{\text{KL}}(g(x_t | x_{t-1}), f_\theta(x_t | x_{t-1})) + d_\gamma(g(y_t | x_t), f_\theta(y_t | x_t))$$

$$d_{\text{KL}}(g(x_t | x_{t-1}), f_\theta(x_t | x_{t-1})) = \int g(x_t | x_{t-1}) \log \frac{g(x_t | x_{t-1})}{f_\theta(x_t | x_{t-1})} dx_t g(x_{t-1}) dx_{t-1}$$

$$d_\gamma(g(y_t | x_t), f_\theta(y_t | x_t)) = -\frac{1}{\gamma} \log \int \frac{g(y_t | x_t) f(y_t | x_t)^\gamma}{(\int f(y_t | x_t)^{1+\gamma} dy_t)^{\frac{\gamma}{1+\gamma}}} g(x_t) dy_t dx_t$$

- 外れ値を含む $g(y/x)$ に関する部分を γ -相互エントロピーで部分的にロバスト化
- γ : 統計的効率性とロバスト性をバランスするハイパーパラメータ
- ϵ が大きいても妥当な推定になることを理論的に保証 (既存研究は $\epsilon = 0$ の場合のみ保証)

- 実際のパラメータ推定

経験推定した相互エントロピーの最小化でパラメータを推定。ただし潜在変数 x に関しては未観測なので、関連した期待値は $y_{1:T}$ での事後期待値で代用。

$$\tilde{d}_{\text{KL}}(g(x_t | x_{t-1}), f_\theta(x_t | x_{t-1})) = \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\log f(x_t | x_{t-1}) | y_{1:T}]$$

$$\tilde{d}_\gamma(g(y_t | x_t), f_\theta(y_t | x_t)) = -\frac{1}{\gamma} \log \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[\frac{f(y_t | x_t)^\gamma}{(\int f(y_t | x_t)^{1+\gamma} dy_t)^{\frac{\gamma}{1+\gamma}}} \middle| y_{1:T} \right] \quad (*)$$

4. 統計的性質: 漸近正規性が成立

- [2] 生成モデル g が統計モデル f で表現できるとき、SSMのM推定量(下記)が漸近正規性を持つときの条件を導出
- $$\text{SSMのM推定量: } \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\psi(x_t, y_t; \theta) | y_{1:T}; \theta] = \mathbf{0} \quad \text{- 推定値が左の方程式の解になる}$$

i.e., 観測モデル $f_\theta(y|x)$ の推定方程式: (*) を指数ドメインへ変換、 θ について微分し0と置く

$$\frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[\frac{\partial}{\partial \theta} \frac{f_\theta(y_t | x_t)^\gamma}{(\int f_\theta(y_t | x_t)^{1+\gamma} dy_t)^{\frac{\gamma}{1+\gamma}}} \middle| y_{1:T}; \theta \right] = \mathbf{0}$$

- 適当な正則条件のもとで、この推定方程式が上述の条件を満たすことを証明

5. パラメータ推定: MMアルゴリズム

- MMアルゴリズム: 推定ロス L に対して以下の性質を満たす優関数 h を最小化

$$L(\theta) \geq h(\theta | \theta^{(a)}), \quad L(\theta^{(a)}) = h(\theta^{(a)} | \theta^{(a)})$$

- $h(\theta; \theta^{(a)})$ を繰り返し最小化することで、ロスの単調減少性を保証

- θ_y に関する優関数: 重み付き対数尤度に似た形

$$h(\theta_y; \theta_y^{(a)}) = -\frac{1}{z_y^{(a)}} \sum_{t=1}^T \mathbb{E}[w_t^{(a)} \log f_{\theta_y}(y_t | x_t) | y_{1:T}] + \frac{1}{1+\gamma} \log \int f_{\theta_y}(y_t | x_t)^{1+\gamma} dy_t$$

- 重み w : 観測 y が分布の中心から遠いと小さくなる \Rightarrow 外れ値の影響を軽減

6. シミュレーション実験

- トイモデル: 定常線形ガウス状態空間モデル

$$x_t = \rho x_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(\epsilon_t | 0, \sigma_\epsilon^2)$$

$$y_t = x_t + \eta_t, \quad \eta_t \sim \mathcal{N}(\eta_t | 0, \sigma_\eta^2)$$

- $T = [250, 500, \dots, 2000]$ で状態、観測の系列を生成

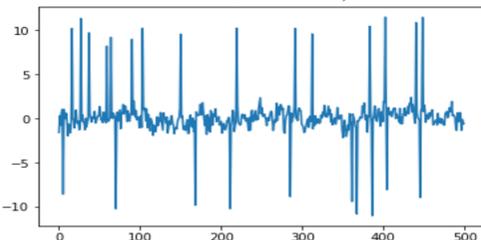
- 外れ値の生成: 確率 ϵ ($0 \leq \epsilon \leq 1$) で外れ値を生成

$$y_{\text{out}} = y_t + \epsilon_{\text{out}}, \quad \epsilon_{\text{out}} \sim \mathcal{N}(\epsilon_{\text{out}}; \mu_{\text{out}}, \sigma_{\text{out}}^2)$$

- $\mu_{\text{out}} = 10, \sigma_{\text{out}}^2 = 0.2$

- $\epsilon = [0.0, 0.05, \dots, 0.25]$ でシフト

サンプル時系列 ($T=500, \epsilon=0.1$)



- 100 seedの推定値の平均値を可視化

図3: MLEとのパラメータ推定精度の比較: 外れ値の割合が増加しても、真値に近い値を推定

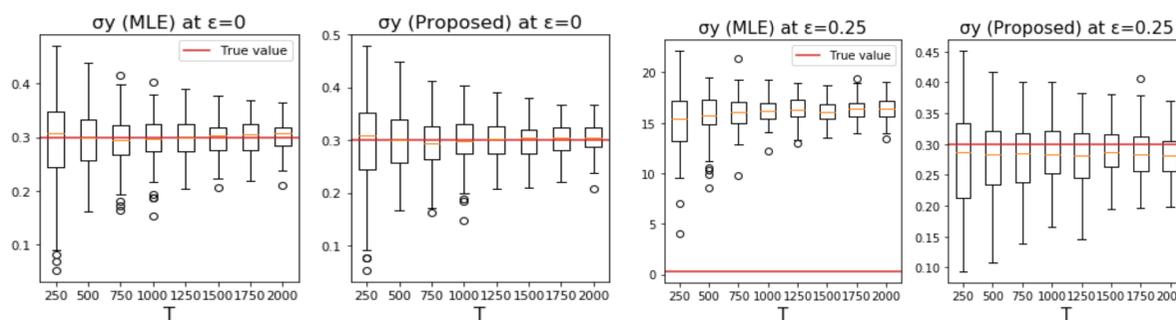
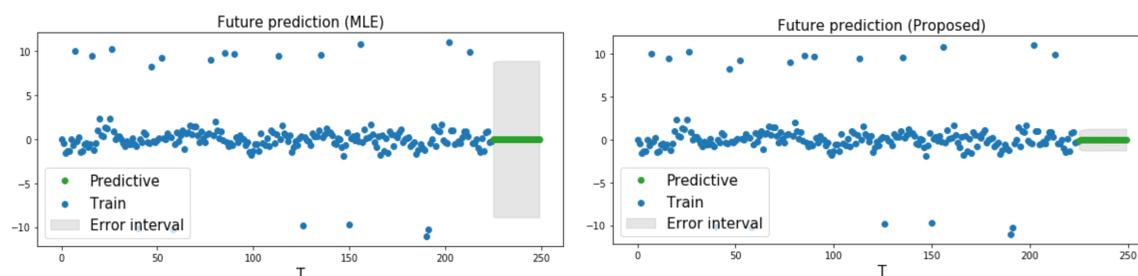


図4: held-outデータの予測における信頼区間: MLEと比較して、よりcredibleな予測を実現



今後の課題

- 非線形・非ガウスモデルへの数値アルゴリズムの拡張、
- 実データ実験、
- オンライン学習への対応 (現代的なデータは系列長が長くバッチ学習が難しい)

Reference

- Fujisawa, H. and Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis*, Vol.99, 2053-2081.
- Jensen, J.L. (2011). Asymptotic normality of M-estimators in nonhomogeneous hidden Markov models. *J. Appl. Prob. Spec.* 48A:295-306.