

# 論文の要旨

題目 Locally Adaptive Bayesian Smoothing using Shrinkage Priors  
(縮小事前分布を用いた局所適合的なベイズ平滑化)

氏名 鬼塚 貴広

ノイズを含む時系列データは、解釈性に乏しく、ノイズを除去して何らかの解釈可能なトレンドを見ることは実用上重要な問題である。本論文では、次の回帰モデル  $y_i = f(x_i) + \varepsilon_i$  ( $i = 1, \dots, n$ ) を考える。ただし、 $y = (y_1, \dots, y_n)^\top$  は観測値で、 $\varepsilon_1, \dots, \varepsilon_n$  は観測誤差を表す確率変数である。誤差分布に対して  $E[\varepsilon_i] = 0$  ( $i = 1, \dots, n$ ) を仮定すると、 $f(x)$  はデータ  $y$  の平均関数を表す。さらに、このモデルの特殊なモデルとして、以下の列モデルを考える。

$$y_i = \theta_i + \varepsilon_i, \quad i = 1, \dots, n.$$

ただし、 $\theta = (\theta_1, \dots, \theta_n)^\top = (f(x_1), \dots, f(x_n))^\top$  は  $n$  次元パラメータである。多くの研究では、データの平均トレンドである  $f(x_i) = \theta_i$  の推定に着目しており、その一つに  $L_1$  trend fitting (Kim et al., 2009) がある。具体的には、罰則付き二乗損失の最小化で定義される。また、二乗損失をチェック損失に置き換えることにより、Brantley et al. (2020) は quantile trend filtering を提案した。これにより、分位点のトレンド推定を実現させた。このような頻度論に基づく手法は、高速に点推定値を得ることができる点で有用である一方で、不確実性の評価が容易でないことや調整パラメータの選択が必要であるという問題点がある。このような問題を解決する一つの方法として、ベイズ法に基づく手法を与えることが考えられる。ベイズ法では、尤度と事前分布の組み合わせから事後分布を構成する手法であり、ベイズ法による trend filtering も考えられてきた。事後分布の信用区間に基づく不確実性の評価が可能であることや事前分布の設定によって  $L_1$  罰則に比べてより柔軟な平滑化が可能ながメリットとして挙げられる。本研究では、ベイズ法によるトレンドフィルタリングの手法として、3章で補正型変分ベイズ近似法を用いた時系列データに対する Bayesian quantile trend filtering と4章で空間データに対する Bayesian quantile trend filtering を提案し、5章では、時系列データの境界関数の推定に着目した Bayesian boundary trend filtering を提案した。提案手法のモデルは、いずれも次の形で書くことができる：

$$y_i = \theta_i + \varepsilon_i, \quad \varepsilon_i \sim p(\cdot), \quad D\theta \sim \pi(\cdot), \quad (i = 1, \dots, n).$$

ただし、 $p$  は誤差分布であり、 $\pi$  は  $D\theta$  に対する事前分布、 $D$  は差分行列である。本論文の構成は以下の通りである。

2章では、本論文全体に共通する内容として、二乗損失やチェック損失に基づく頻度論的手法の概要と、それに対応する正規尤度と非対称ラプラス尤度に基づくベイズ的手法についてまとめる。また、ベイズ法による縮小推定として、Bayesian lasso 型の事前分布と馬蹄事前分布を紹介する。

3章では、時系列データに対する Bayesian quantile trend filtering を提案する。特に、変分ベイズ近似を導入することで、高速な点推定値の計算と妥当な信用区間を構成するアルゴリズムを提案する。具体的には、尤度として非対称ラプラス尤度、差分に対する縮小事前分として馬蹄事前分布を用いたモデルにより、局所適合的な推定を実現させる。また、尤度と事前分布の混合表現を用いることで、MCMC 法の一つであるギブスサンプリングと変分ベイズ法の一つである mean-field

variational Bayes (MFVB) approximation による 2 つの手法により事後分布の近似を考える。MFVB は MCMC と異なり、サンプリングに基づいて事後分布を構成するのではなく、カルバック・ライブラー擬距離の意味で事後分布を近似する分布を最適化手法によって求める手法であり、高速に事後分布を近似できることが利点である。一方で、MFVB による信用区間は MCMC による信用区間に比べて、狭い信用区間を与えてしまうという問題が知られている (e.g. Blei et al., 2017)。さらに、MCMC や解析的に事後分布を得る場合であっても、モデルが誤特定されている下では、事後分布に基づく信用区間が妥当でない (真値の被覆確率が適切でない) ことが知られており、例えば、一般化ベイズの枠組みで Syring and Martin (2019) によってアルゴリズムに基づく補正手法が提案されている。本研究では、このような問題を解決するために、Bayesian quantile trend filtering に対する変分事後分布の信用区間の補正アルゴリズムを与える。数値実験と実データ解析を通して、既存手法と提案手法の性能を比較する。

4 章では、データにグラフ構造を仮定することによって空間データに対する Bayesian quantile trend filtering を提案する。グラフ構造の仮定は、areal data に対する空間統計モデルの一つである同時自己回帰 (SAR) モデルや条件付自己回帰 (CAR) モデルなどと同様の設定であり、既存手法では過剰に平滑化してしまうという問題がある。この問題を解決するために、具体的には、差分行列をグラフ構造に拡張し、3 章と同様に、非対称ラプラス尤度と差分に対する縮小事前分布を仮定するモデルを与え、局所適合的な空間分位平滑化を実現する。効率的なアルゴリズムを構成するために、shadow prior と呼ばれる事前分布を導入し、ギブスサンプリングアルゴリズムを提案する。提案手法の有用性は、数値実験と実データ解析によって示す。特に、提案手法を各地点で複数観測のあるデータに適用できるモデルに拡張して定義し、2013 年から 2017 年の東京 23 区の年間暴力犯罪件数の空間データに対して適用することで、複数年を通した潜在的な暴力犯罪の空間分位トレンドを局所適合的に推定できることを確認する。

5 章では、時系列データに対する Bayesian boundary trend filtering を提案する。上側境界トレンドの推定は、誤差分布に、片側の台  $(-\infty, 0)$  を仮定することに対応する。なお、このようなモデルは、目的変数  $y$  の台が未知パラメータに依存する形となり、非正則モデルと呼ばれる。本研究では、特に、切断正規尤度を仮定し、差分に対する事前分布として、3 章と同様に、馬蹄事前分布を仮定することでより柔軟な平滑化を与える。提案モデルにおいて、切断正規尤度と条件付き正規事前分布による完全条件付き分布は切断正規事前分布となり、ギブスサンプリングを構成するためには、 $n$  次元切断正規分布からのサンプリングが必要になる。しかし、切断正規分布からのサンプリングは次元が高くなると困難になることが知られており、この問題を解決するため、本研究では切断正規尤度の近似を導入し、効率的なアルゴリズムを与える。また、実データ解析での境界トレンドの推定においては、トレンドに単調増加制約を課すことが自然な場合も多いため、単調制約を緩めた近単調制約の下でのモデルも提案する。提案手法の有用性は、数値実験と実データ解析によって示す。

最後に、本論文のまとめと今後の課題について論じる。

## 参考文献

- [1] Brantley, H.L., Guinness, J., Chi, E.C. (2020), Baseline drift estimation for air quality data using quantile trend filtering. *The Annals of Applied Statistics*, **14**(2), 585–604.

- [2] Onizuka, T., Hashimoto, S. and Sugasawa, S. (2024), Fast and Locally Adaptive Bayesian Quantile Smoothing using Calibrated Variational Approximations. *Statistics and Computing*, **34**, 15.
- [3] Onizuka, T., Hashimoto, S. and Sugasawa, S. (2024), Locally Adaptive Spatial Quantile Smoothing: Application to Monitoring Crime Density in Tokyo. *Spatial Statistics*, **59**, 100793.
- [4] Onizuka, T., Iwashige, F. and Hashimoto, S. (2024), Bayesian Boundary Trend Filtering. *Computational Statistics and Data Analysis*, **191**, 107889.
- [5] Syring, N., Martin, R. (2019), Calibrating general posterior credible regions. *Biometrika*, **106**(2), 479–486.