

# 分位点回帰を用いた条件付き密度関数と条件付き最頻値の推定

大田 浩史 東京大学大学院経済学研究科  
原 聡 国立情報学研究所, JST, ERATO, 河原林巨大グラフプロジェクト

## 1 はじめに

統計学において最頻値は標本の分布の特徴を表すひとつの要約統計量であり、分布の非対称性や外れ値の影響を受けにくい頑健なものとして知られている。裾が厚い分布・歪度が大きい分布については、期待値よりも最頻値が情報を持っていると考えられ、最頻値は期待値が捉えられない分布の構造を捉えうる。

ノンパラメトリック回帰分析においては、被説明変数  $Y$  を説明変数  $X$  で条件付けたときの mean function および quantile function の統計的推測に関する研究が多くなされてきたが、近年 [2] で modal function の推定法が提案されている。[2] の方法はカーネル密度推定法に依存し、 $X$  の次元の増加に伴って推定のレート・精度が大きく落ち、次元が 3 次元程度でも利用に制限があるため、何らかの方法で次元削減を行い、高次元データに対応できるようにしたい。そこで本研究では、条件付き密度関数および最頻値を分位点回帰を経由して推定する方法を提案し、次元削減を図った。

## 2 モデル

一段階目で、分位点関数が密度関数の情報を有することに基づいて、分位点回帰を利用した条件付き密度関数の推定量を構成する。( $y_\tau = F_{Y|X}^{-1}(\tau)$ ,  $\hat{Q}(\cdot|x)$  は分位点関数  $Q(\cdot|x)$  の推定量,  $h = h_n \downarrow 0$ ,  $n \rightarrow \infty$  はバンド幅.)

$$\hat{f}(y_\tau|x) := \frac{2h}{\hat{Q}(\tau+h|x) - \hat{Q}(\tau-h|x)} \quad (*)$$

$Y$  を  $X = x$  で条件付けたときの最頻値  $m(x)$  の推定量  $\hat{m}(x)$  は、(\*) に基づく条件付き密度関数の推定量を最大化する点で定義される。

$$\hat{m}(x) := \arg \max_{y_\tau} \hat{f}(y_\tau|x).$$

$X$  の次元を  $p$  とすると、推定量が一致性を持つためには、[2] では  $nh^{p+5}/\log n \rightarrow \infty$  の条件が必要であるが、提案法では  $nh^2/\log n \rightarrow \infty$  でよい。

さらに、本研究では高次元スパースな仮定 [1] の下で  $p = p_n \rightarrow \infty$  となる状況においても  $\hat{m}(x)$  の一致推定の条件を導出し、次元の増加が収束レートに与える悪影響の改善ができたことを理論的に確認した。

## 参考文献

- [1] BELLONI, A., AND CHERNOZHUKOV, V.  $\ell_1$ -penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics* 39, 1 (2011), 82–130.
- [2] CHEN, Y.-C., GENOVESE, C. R., TIBSHIRANI, R. J., AND WASSERMAN, L. Nonparametric modal regression. *The Annals of Statistics* 44, 2 (2016), 489–514.