

方向データのノンパラメトリック密度推定とその漸近的性質

金沢大学人間社会環境研究科 鶴田靖人
金沢大学経済学経営学系 寒河江雅彦

1 はじめに

方向データのノンパラメトリック密度推定法について考える。方向データとは単位円周上の点もしくは平面上の単位ベクトルとして表されるデータである。本報告では, Marzio et al. (2009, 2011) に基づいて巻き込みコーシー・カーネル (WCK) の漸近的な性質を導く。また, 数値実験によってフォン・ミーゼス・カーネル (VMK) に基づく密度推定量を比較する。

2 定義

方向データのカーネル関数を $K_\kappa(\theta)$ とおく, 但し $\kappa > 0$ は集中度パラメータ (通常のバンド幅の逆数に対応する平滑化パラメータ) とする。標本 $\theta_1, \dots, \theta_n$ は独立同一分布 $f(\theta)$ に従うとする。カーネル密度推定量は,

$$\hat{f}(\theta; \kappa) := \frac{1}{n} \sum_{i=1}^n K_\kappa(\theta - \theta_i).$$

先行研究におけるカーネル関数としては主に VMK を用いている: すなわち,

$$K_\kappa(\theta) := \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos \theta\},$$

但し $I_p(\kappa)$ は p 次第 1 種変形ベッセル関数を表す。

3 MISE およびその漸近的性質

3.1 先行研究

Hall et al. (1987) は球面上で定義された密度関数のノンパラメトリック推定を議論した。その中で, 平均二乗誤差 (MISE) を導出し, その最適な収束レートは $O(n^{-4/5})$ であることを示した。Taylor (2008) は真の分布型をフォンミーゼス分布と仮定したうえで, VMK を用いたときの MISE を導出し, その最適な収束レートは $O(n^{-4/5})$ になることを示した。Marzio et al. (2009, 2011) はトラス上で定義された密度関数の推定問題を取扱い, MISE を導いた:

$$\text{AMISE} := \frac{\eta_2^2(K_\kappa)R(f'')}{4} + \frac{R(K_\kappa)}{n} \quad (1)$$

但し, $\eta_j(K_\kappa) := \int_0^{2\pi} \sin^j(\theta)K_\kappa(\theta)d\theta$, $R(g) := \int_0^{2\pi} \{g(\theta)\}^2 d\theta$. 加えて, VMK による (1) の最適な収束レートは $O(n^{-4/5})$ であることを示した。また, Garcia (2013) は Hall et al. (1987) の MISE に基づいて VMK を用いたプラグイン法を提案した。

3.2 巻き込みコーシー・カーネル

WCK は,

$$K_\rho(\theta) := \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta)},$$

但し, $\rho \in (0, 1)$ は分布の集中度を調節するパラメータとする。(1) の表現を用いて WCK による密度推定の漸近的性質を導くと,

$$\text{AMISE} = \frac{\{1 - \rho^2\}^2 R(f'')}{16} + \frac{1}{n\pi(1 - \rho^2)}. \quad (2)$$

ここで, $1 - \rho^2 = h$ とおくと, 次式が得られる,

$$\text{AMISE} = \frac{h^2 R(f'')}{16} + \frac{1}{n\pi h}, \quad (3)$$

その最適な収束レート $O(n^{-2/3})$ が導かれる。

4 数値実験

混合フォン・ミーゼス分布から発生させたデータに基づく WCK と VMK に基づく密度推定量のシミュレーション実験を行った。詳細な結果は当日発表する。

参考文献

- [1] Hall, P., Watson, G. S. and Cabrera, J. (1987). *Biometrika*, **74**, 751-762.
- [2] Taylor, C. C. (2008). *Computational Statistics & Data Analysis*, **52**, 3493-3500.
- [3] Di Marzio, M., Panzera, A. and Taylor, C. C. (2009). *Statistics and Probability Letters*, **79**, 2066-2075.
- [4] Di Marzio, M., Panzera, A. and Taylor, C. C. (2011). *Journal of Statistical Planning and Inference*, **141**, 2156-2173.
- [5] Garcia, E. (2013). *Electronic Journal of Statistics*, **7**, 1655-1685.