カーネル密度推定における適切な台の推定の必要性

九州大学大学院数理学府　森 山　卓

1. カーネル密度推定量における境界バイアス問題と先行研究

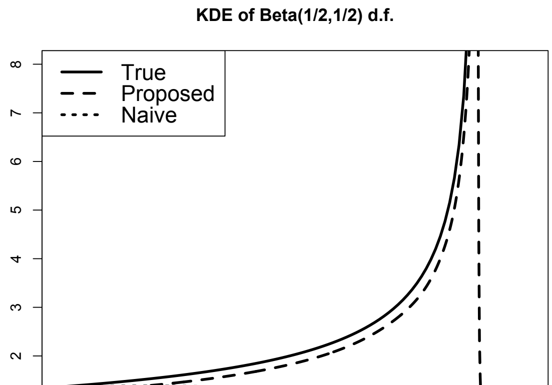
確率的現象はその分布関数によって一意に表され，分布の推定量からは有益な情報を抽出できる．ノンパラメトリックな推定量としてはカーネル平滑化法を用いたカーネル密度推定量が広く知られており，その性質は Tsybakov (2009) [1] によくまとめられている．

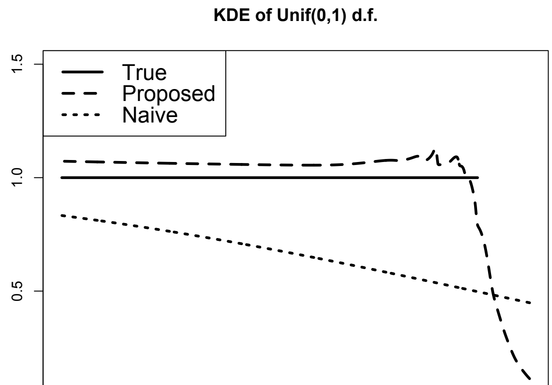
さて真の（1次元）密度 f の台が実数全体を覆わない場合には，カーネル型推定量にはいわゆる境界バイアス問題が存在する．これにより台の境界付近で，密度推定量に $O(1)$ のバイアスが生じ，推定量は一致性すら失いかねない．この問題について，真の台 $supp(f)$ が既知である場合を対象として近年世界的に研究が進められている．一方で真の台が未知であっても，境界バイアスは起こりうるにもかかわらず，必ずしも特別な注意が払われていない．これはおそらく，未知の台に適合する密度推定法については Hall & Park (2002) [2] を除くとほとんど先行研究が見当たらず，問題が広く認識されていないからと考えられる．

本講演ではまずこの境界バイアス問題を取り上げ，数値実験を用いて通常のカーネル型推定量が '精確でない' 結果を返す例を紹介する．その後，カーネル密度推定法における適切な台の推定の必要性を主張し，台の推定を手法に組み込んだ新たな密度（分布）推定法を提案する．下の図は真の密度（True），提案手法による密度推定量（Proposed）及び通常のカーネル密度推定量（Naive）の数値実験結果を描写したものである．左はパラメータ $(\frac{1}{2}, \frac{1}{2})$ のベータ分布，右は $[0, 1]$ 上の一様分布についての結果であり，標本数 $n = 300$ の 1000 回の平均値をプロットしている．

2. 未知の台に適合する提案手法

提案手法は真の台が未知であっても，ある条件下で境界を検知し，密度推定量はバイアスを縮小し，一致性を回復することを示す．実は新たな密度推定量は，真の台の境界付近における，ある局所期待損失の最小解と漸近的に一致する統計量として導出される．このことから Hall & Park (2002) [2] の手法と異なり，台の推定量は用いられる境界バイアス縮小法を考慮する意味で正当性を持つことが示される．





3. 参考文献

[1] Tsybakov, A. B. (2009). *Introduction to nonparametric estimation.* Springer Series in Statistics. Springer, New York.

[2] Hall, P. and Park, B. U. (2002). New methods for bias correction at endpoints and boundaries. *Annals of Statistics*, 1460-1479.