Numer. Math. Theor. Meth. Appl. Vol. **19**, No. 1, pp. 240-274
doi: 10.4208/nmtma.OA-2025-0038 February 2026

Density Estimation via Mixture Discrepancy and Moments

Zhengyang Lei[1], Lirong Qu[2], Sihong Shao[1,*] and Yunfeng Xiong[2]

[1] *CAPT, LMAM and School of Mathematical Sciences, Peking University, Beijing 100871, China*
[2] *School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China*

Received 10 April 2025; Accepted (in revised version) 18 December 2025

Abstract. With the aim of generalizing histogram statistics to higher dimensional cases, density estimation via discrepancy based sequential partition (DSP) has been proposed to learn an adaptive piecewise constant approximation defined on a binary sequential partition of the underlying domain, where the star discrepancy is adopted to measure the uniformity of particle distribution. However, the calculation of the star discrepancy is NP-hard and it does not satisfy the reflection invariance and rotation invariance either. To this end, we use the mixture discrepancy and the comparison of moments as a replacement of the star discrepancy, leading to the density estimation via mixture discrepancy based sequential partition (DSP-mix) and density estimation via moment-based sequential partition (MSP), respectively. Both DSP-mix and MSP are computationally tractable and exhibit the reflection and rotation invariance. Numerical experiments in reconstructing beta mixtures, Gaussian mixtures and heavy-tailed Cauchy mixtures up to 30 dimension are conducted, demonstrating that MSP can maintain the same accuracy compared with DSP, while gaining an increase in speed by a factor of two to twenty for large sample size. DSP-mix can achieve satisfactory accuracy and boost the efficiency in low-dimensional tests ($d \le 6$), but might lose accuracy in high-dimensional problems due to a reduction in partition level.

AMS subject classifications: 62G07, 62G05, 65M50, 11K38, 44A60

Key words: Density estimation, adaptive sequential partition, discrepancy, Hausdorff moment problem, Cauchy mixtures.

1. Introduction

As a fundamental problem in statistics, density estimation aims at constructing an estimate of their common density function $q(\mathbf{y})$ for given N observations $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$.

*Corresponding author. *Email addresses:* `sihong@math.pku.edu.cn` (S. Shao), `leizy@stu.pku.edu.cn` (Z. Lei), `lirong.qu@mail.bnu.edu.cn` (L. Qu), `yfxiong@bnu.edu.cn` (Y. Xiong)

It has various applications in uncertainty quantification [11], Bayesian inference [3] and scientific computing [5, 19, 29, 30]. A simple and classical approach is the histogram statistics as a uniform piecewise constant density estimator. Unfortunately, it is rather difficult to apply the histogram to high-dimensional problems as both bin size and required sample size grow exponentially in dimension, which is well known as the curse of dimensionality (CoD) [23, 30].

This paper focuses on nonparametric approaches for multivariate density estimation, especially for the case when the dimension d is moderately large, say, 6–30 [18, 20]. The benefit of nonparametric methods is their ability to achieve estimation optimality for any input distribution as more data are observed. A widely used method is the kernel density estimation (KDE), treating each data point as the center of a kernel function like the Gaussian functions and smoothed splines [23]. But the accuracy of the kernel estimator becomes very sensitive to the choice of the window size and the shape of the kernel [28], and the magic lies in how to balance the bias and variance of the estimator, like using the smoothing properties of linear diffusion processes [2] or choosing adaptive bandwidths [28, 31]. Another approach is based on tensor decompositions, with huge advantages in that the computational cost of the construction, the storage requirements and the operations required for conditional distribution method sampling from the distributional approximation all scale linearly with dimension [5, 19, 25]. However, it still relies heavily on the low-rank assumption of the underlying density, whereas the detection and characterization of local features of multivariate density estimation are usually prohibitive [18]. Machine learning approaches have recently been drawing a growing attention, e.g., the density estimation through normalizing flows [24] and deep generative neural networks [17]. In practice, they require extensive tuning to perform well [15].

Here we mainly discuss a data-driven density estimation method, termed tree-based density estimation. This features a class of estimators which employs simple and flexible binary partitions to adapt to the underlying density function, along with the decision tree using stopping times in a data-driven way [3]. The decision tree based method has a great potential to overcome the burden imposed by the high dimensionality, especially when the dimension is only moderately large and the density function exhibits certain spatial features that can be leveraged of [15]. Essentially, it is equivalent to cluster the samples into small nonintersecting sets, each supported by a tree-structured density [16]. An adaptive nonparametric density estimation can be constructed by either the Bayesian sequential partition (BSP) [15, 18] or the discrepancy based sequential partition [14]. Compared with BSP, the computational cost of DSP is much cheaper due to its greedy construction, and is fully capable to handle higher dimension [29].

To be more specific, DSP uses the star discrepancy, a concept from the quasi-Monte Carlo method, to measure the uniformity of the observations [4, 10]. For a point set $(\mathbf{y}_1, \ldots, \mathbf{y}_n) \subset [0,1]^d$, its star discrepancy reads

$$D^*(\mathbf{y}_1, \ldots, \mathbf{y}_n) = \sup_{\boldsymbol{u} \in [0,1]^d} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{[0,\boldsymbol{u})}(\mathbf{y}_i) - \mathrm{vol}([0,\boldsymbol{u}))\right|, \tag{1.1}$$