Applied Mathematical Sciences, Vol. 20, 2026, no. 4, 157 - 164
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2026.919329

Transformation Based Procedure for Data with Compact Support

Marco Di Marzio

University of Chieti-Pescara, Italy

Stefania Fensore

University of Chieti-Pescara, Italy

Chiara Passamonti

University of Luxembourg, Luxembourg

Abstract

When we aim to estimate a density function with bounded support, the naive kernel density estimation becomes strongly biased in the boundary region due to the well-known estimated density overflow problem. The reflection method appears to be an efficient way to alleviate this issue. We propose a simple domain transformation that could favor an elegant application of such a principle. A small numerical experiment and a real case study involving suicide data end the work.

Mathematics Subject Classification: 62G05

Keywords: Boundary problem, Density estimation, Psychiatric treatment duration data

1 Introduction

In many applied statistical problems, the variable of interest is supported on a bounded domain. This is common in biomedical and social sciences, where

quantities such as durations, proportions, or physical measurements are naturally restricted to intervals. In such settings, nonparametric density estimation, and in particular kernel density estimation (KDE), suffers from well-known boundary bias issues.

The origin of this problem lies in the fact that classical kernel estimators allocate probability mass outside the support of the variable. As a consequence, the estimated density is distorted near the boundaries, especially when the true distribution is concentrated close to them. This effect is particularly evident for non-negative data, i.e. distributions supported on $[0, \infty)$, or for data defined on compact intervals.

A standard approach to mitigate boundary bias is the reflection method, originally proposed by [4] and further developed by [5].

The aim of this work is to propose an alternative approach based on domain transformation. By exploiting the geometric equivalence between bounded intervals and periodic domains, we transform the estimation problem into a circular setting, where boundaries are absent. This allows us to construct a density estimator that avoids boundary bias while preserving desirable statistical properties. For a comprehensive reference on circular statistics, see, for example, [3].

The paper is organized as follows: in Section 2 we recall some preliminaries about kernel density estimation and reflection method, and we also provide our transformation-based approach. A small simulation study and an analysis concerning psychiatric treatment durations are conducted in Sections 3 and 4, respectively.

2 Methodology

2.1 Preliminaries

Let X be a continuous random variable with unknown density f_X supported on a compact interval $[a, b] \subset \mathbb{R}$. Given a random sample $X_1, \ldots, X_n$, the classical kernel density estimator is defined as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1}$$

where K is a kernel function, which is typically smooth, symmetric, non-negative and integrating to 1. Also, $h > 0$ is the bandwidth parameter controlling the smoothness of the estimate.

Near the boundaries a and b, this estimator is biased because part of the kernel mass lies outside the support. A common correction is the reflection method, which augments the sample by reflecting observations across

the boundary. For instance, if the support is $[0, \infty)$, the augmented sample becomes

$$X_1, -X_1, X_2, -X_2, \ldots, X_n, -X_n.$$

The resulting estimator can be written as

$$\hat{f}(x) = \begin{cases} 2\hat{f}^*(x), & x \geq 0, \\ 0, & x < 0, \end{cases}$$

where $\hat{f}^*$ is the KDE computed on the augmented sample.

Despite its simplicity, this method has limitations. In particular, when the kernel has unbounded support, the reflection may not be complete, leading to estimators that do not integrate to one and therefore are not proper densities.

2.2 Reflection based on domain transformation

To overcome the limitations of the standard reflection method, we propose a transformation-based approach that maps the problem into a directional domain.

The first step of the procedure consists on transfering data from a linear domain to a circular one. Let X be supported on $[a, b] \in \mathbb{R}$. Given a random sample $X_1, \ldots, X_n$, we define the following transformation which maps the data onto the interval $[0, \pi]$, corresponding to the upper semicircle of a unit circle

$$\Theta_i = \pi \frac{X_i - a}{b - a}, \quad i = 1, \ldots, n.$$

As a second step, to eliminate boundary effects, we reflect the transformed data across the semicircle, obtaining $2n$ observations distributed over the entire circumference. Specifically, the reflection is performed by multiplying each point, expressed in Cartesian coordinates, by the reflection matrix

$$\mathtt{R} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

which flips the sign of the y-coordinate. The main advantage of this transformation, which is depicted in Figure 1, is related to the fact that this way we can work on a domain where boundaries are absent.

Then, we apply to the augmented data a circular version of the density estimator (1) which is defined as follows. Given a random sample of angles $\Theta_1, \ldots, \Theta_n$ from an unknown circular density f_Θ, the kernel estimator of f_Θ at $\theta \in [0, 2\pi)$ is given by

$$\hat{f}_\Theta(\theta; \kappa) = \frac{1}{n} \sum_{i=1}^{n} K_\kappa(\Theta_i - \theta),$$

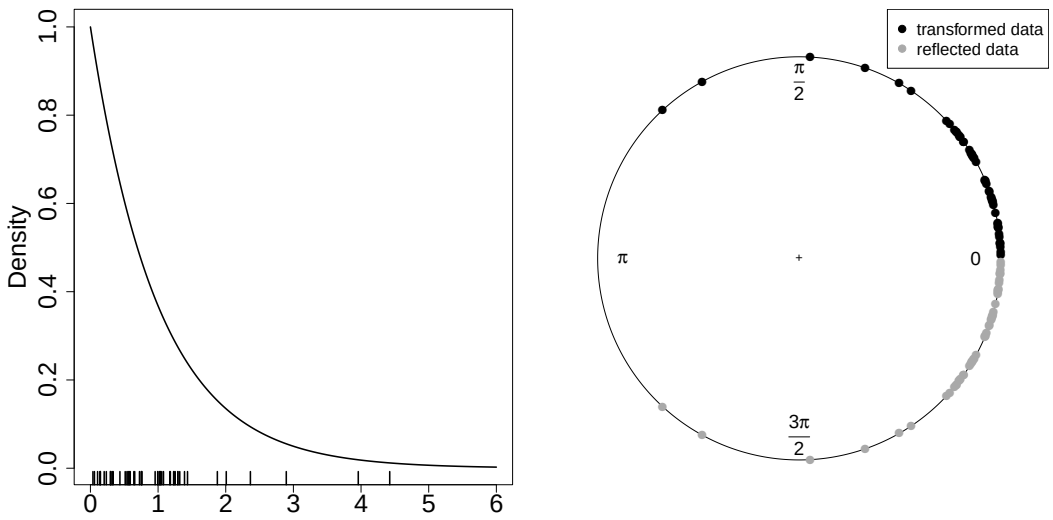

Figure 1: Left: data and population model in the original space. Right: data transferred on the circle (black) and reflected ones (grey).

where K_κ is a circular weight function, i.e. a periodic, unimodal, symmetric density function. The width of that neighbourhood is governed by the concentration parameter κ, which plays the inverse role of the bandwidth h. A common choice for K_κ is the von Mises kernel. See [2] for details.

As in the Euclidean reflection approach, we need, as an intermediate step, to restrict $\hat{f}_\Theta(\theta;\kappa)$ to the upper semicircle $[0,\pi]$, and then double it to obtain a bona fide density.

Finally, there is a back-transformation procedure, where the density is mapped back to the original domain $[a,b]$, yielding an estimator for f_X.

For illustrative purposes, the last three steps are depicted in Figure 2.

This method offers some advantages as follows. Firstly, standard smoothing degree selection rules for kernel density estimation could be adopted; then, this transformation procedure could be seen as a practical way to implement periodic estimators by avoiding their several implementation problems. Moreover, when the density has a compact support, the standard reflection approach requires $3n$ observations, while in our procedure we always have $2n$ data.

3 Simulation

To evaluate the performance of the proposed method, we conduct simulation experiments under two different boundary scenarios. Specifically, we consider distributions supported on $[0,1]$, which are Beta$(1,3)$ and Beta$(0.5,0.5)$. For a graphical illustration, see Fig. 3. To evaluate the difficulty of the estimation

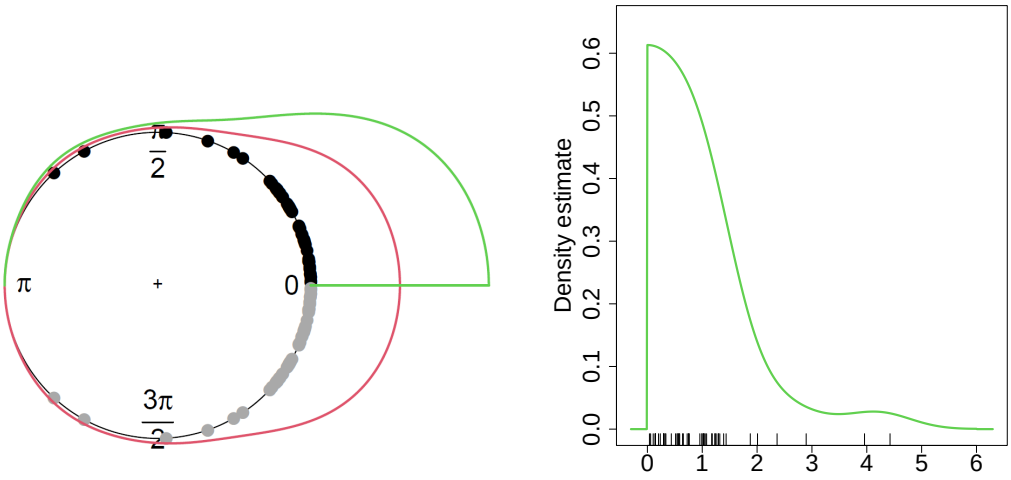

Figure 2: Left: KDE over the whole circle (red) and bona-fide density estimate on $[0, \pi]$ (green). Right: back-transformed estimate on the original domain.

procedure, we compute the roughness of the population model f, as

$$R(f) = \int_{\mathbb{R}} [f''(x)]^2 \, dx.$$

This quantity measures the overall curvature of the density, and a larger value implies a more challenging scenario.

For each model, we generate 1000 samples of size $n = 50, 300, 1000$. We compare the standard reflection method with the proposed transformation-based approach. Concerning the choice of the kernel, we use a Gaussian kernel for the standard method with bandwidth selected via unbiased cross-validation. For the proposed method, we use a von Mises kernel with smoothing degree selected via maximum likelihood cross-validation.

As the performance indicator, the Integrated Mean Integrated Squared Error (MISE) is used.

The results are reported in Table 3. It is shown that the proposed method consistently outperforms the classical reflection one. For example, for the Beta$(1, 3)$ distribution with $n = 1000$, the MISE is 3.34 for the reflection method and 2.71 for the proposed approach, reaching an improvement of 19%. Similarly, for the Beta$(0.5, 0.5)$ distribution with $n = 1000$, the MISE is reduced from 11.36 to 9.34, obtaining a gain of about 18%.

Moreover, we note that for the Beta(1,3) model, the roughness is modest, and the MISE values are relatively low for all sample sizes. In contrast, for the Beta(0.5,0.5) model, the roughness is extremely high, and the MISE values are substantially larger across all n.

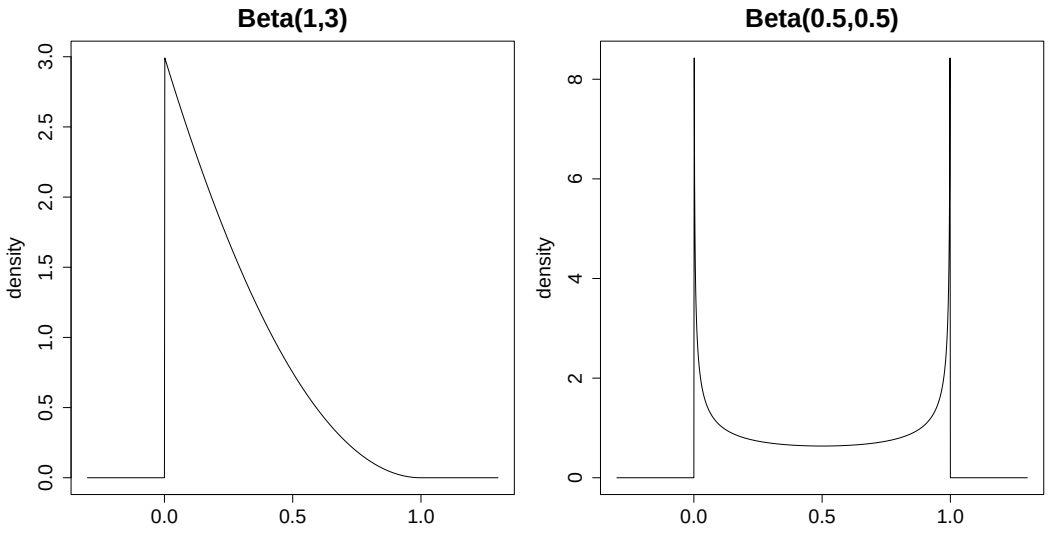

Figure 3: Simulation models.

Table 1: $MISE \times 100$ values of two estimators at different for 1000 samples of sample sizes n (best performances are highlighted in bold font).

Simulation model	$R(f)$	n	Reflection method	Our proposal
Beta(1,3)	36	50	6.79	**5.77**
		300	4.08	**3.03**
		1000	3.34	**2.71**
Beta(0.5,0.5)	2×10^6	50	17.14	**16.25**
		300	11.54	**10.70**
		1000	11.36	**9.34**

4 Application to suicide data

We apply the proposed method to real data concerning psychiatric treatment durations.

One aspect of the well-established link between mental illness and increased suicide rates that has garnered attention in the psychiatric literature is the way in which suicide risk is influenced by the duration of treatment. Specifically, we investigate treatment duration data (see [1]), where estimates assigning weight to negative values are considered inappropriate.

Data refers to the lengths (in days) of 86 spells of psychiatric treatment undergone by patients who were used as controls in a study of suicide risks.

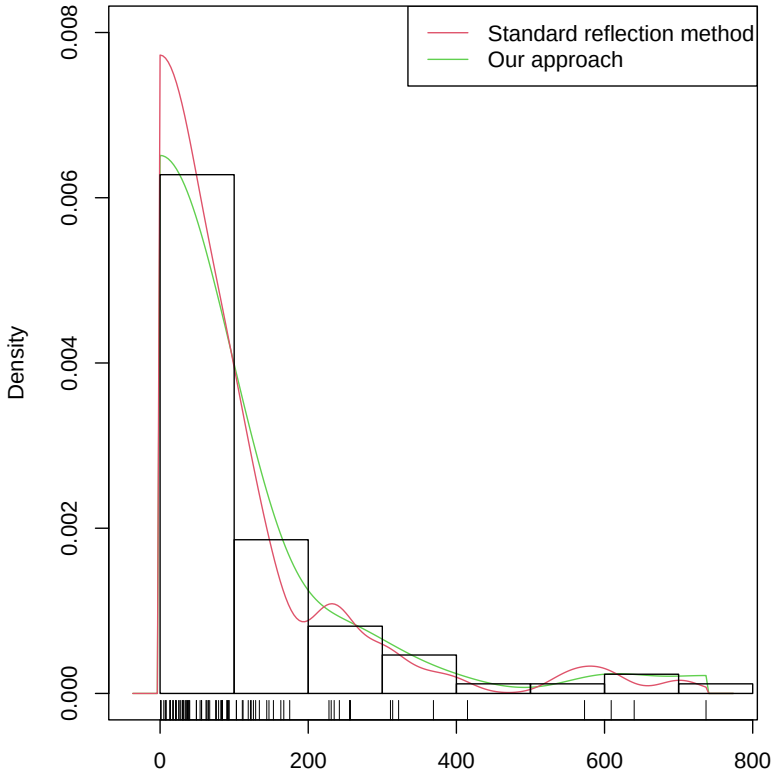

Figure 4: Histogram and density estimates on suicide data.

We compare our transformation-based procedure along with the standard reflection density estimator. The smoothing parameter in the first case is selected using the function `bw.cv.ml.circular` from the R package `circular`, which implements cross-validation based on the Kullback–Leibler loss, while, for the standard reflection approach, the bandwidth is selected by the rule-of-thumb.

The histogram of the data, along with the obtained density estimates are depicted in Figure 4. We can see that the estimate based on our approach is closer to the histogram, especially near zero, where boundary effects are most pronounced. In contrast, the standard reflection method exhibits noticeable distortion in this region.

References

[1] Copas, J. B., & Fryer, M. J., Density estimation and suicide risks in psychiatric treatment, *Journal of the Royal Statistical Society, Series A: Statistics in Society*, **143** (2) (1980), 167-176. https://doi.org/10.2307/2981988

[2] Hall, P., Watson, G. S., & Cabrera, J., Kernel density estimation with spherical data, *Biometrika*, **74** (4) (1987), 751-762. https://doi.org/10.1093/biomet/74.4.751

[3] Jammalamadaka, S. R., & Sengupta, A., *Topics in circular statistics*, (Vol. 5), World Scientific, 2001. https://doi.org/10.1142/4031

[4] Schuster, E. F., Incorporating support constraints into nonparametric estimators of densities, *Communications in Statistics-Theory and Methods*, **14** (5) (1985), 1123-1136. https://doi.org/10.1080/03610928508828965

[5] Silverman, B. W., *Density estimation for statistics and data analysis*, Routledge, 2018. https://doi.org/10.1201/9781315140919

Received: March 25, 2026; Published: April 28, 2028