

Multiple Imputation for Kernel Density Estimation

Ryan Elmore*

Department of Statistics, University of Denver
and

Sean Kerr

Department of Mathematics and Statistics, Loyola University Chicago
and

Gregory J. Matthews

Department of Mathematics and Statistics, Loyola University Chicago

May 27, 2025

Abstract

The text of your abstract. 200 or fewer words.

Keywords: 3 to 6 keywords, that do not appear in the title

*The authors gratefully acknowledge *please remember to list all relevant funding sources in the non-anonymized (unblinded) version.*

1 Introduction

Assessing the distribution of an observed variable is an issue of keen interest to many areas of statistical practice. Density estimates are often part of exploratory data analysis (EDA) to help an analyst understand how the observed data are distributed. In addition, many statistical procedures rely on specific distributional assumptions about the data: classic parametric tests such as a t-test rely on the assumption that the data come from a normal population; Traditional linear regression analysis assumes the model residuals follow a normal distribution; data that follows a multimodal distribution may lead a practitioner to perform mixture modeling.

When data are fully observed there are a host of methods for estimating the probability density function; here we focus on kernel density estimation.

Ryan T. Elmore (2006) Dubnicka (2009) Based on Horvitz-Thompson estimator (Horvitz & Thompson (1952))

Saloui: <https://www.tandfonline.com/doi/abs/10.1080/03610926.2016.1205618>

2 Sim settings

Sample size? Number of imputations? Imputation method: pmm, norm, etc?

True Distribution of Y: 1. Normal 2. Chi squared 3. Mixture of Normals

Missingness: 1. MCAR 2. MAR 3. MNAR?????

Missing rate: 1. Low 15ish 2. Medium 30ish 3. High 50ish

Combining Rules: 1. stacking and then one density estimate 2. estimating and the combining

3 Methods

4 Real data

Fish: Aging the fish you have to kill the fish.

AIDS:

5 Conclusion

6 Disclosure statement

The authors have the following conflicts of interest to declare (or replace with a statement that no conflicts of interest exist).

7 Data Availability Statement

Deidentified data have been made available at the following URL: XX.

SUPPLEMENTARY MATERIAL

8 BibTeX

We encourage you to use BibTeX. If you have, please feel free to use the package natbib with any bibliography style you're comfortable with. The .bst file agsm has been included here for your convenience.

References

- Dubnicka, S. R. (2009), ‘Kernel Density Estimation with Missing Data and Auxiliary Variables’, *Australian and New Zealand Journal of Statistics* **51**, 247–270.
- Horvitz, D. & Thompson, D. (1952), ‘A generalization of sampling without replacement from a finite universe’, *Journal of the American Statistical Association* **47**, 663–685.
- Ryan T. Elmore, Peter Hall, V. S. T. (2006), ‘Nonparametric Density Estimation from Covariate Information’, *Journal of the American Statistical Association* **101**(474), 701–711.