Nonparametric estimation of a heaping mechanism for precise and heaped self-report data

Sandra D. Griffith[1,⋆], **Saul Shiffman**[2], **Daniel F. Heitjan**[1]

1. Department of Biostatistics & Epidemiology, University of Pennsylvania
2. Department of Psychology, University of Pittsburgh
⋆Contact author: sgrif@upenn.edu

Keywords: Digit preference, Coarse data, Rounded data, Smoking cessation, Measurement error

Open-ended numerical measures, often used in self-report to assess quantities or frequencies, exhibit a form of measurement error termed heaping. Heaping occurs when quantities are reported with varying levels of precision. Digit preference is a special case of heaping where the preferred values are round numbers. Daily cigarette counts, for example, commonly exhibit heaps at multiples of 20, and to a lesser extent, 2, 5, and 10, when measured by retrospective recall methods. Because heaping can introduce substantial bias to estimates, conclusions drawn from data subject to heaping are suspect. Several methods have been proposed to estimate the true underlying distribution from heaped data, but all depend on unverifiable assumptions about the heaping mechanism. A data set in which subjects reported cigarette consumption by both a precise method (ecological momentary assessment as implemented with a hand-held electronic device) and a more traditional, imprecise method (timeline followback, or periodic retrospective recall) motivates our method. We propose a nonparametric method to estimate the conditional distribution of the heaping mechanism given the precise measurement. We measure uncertainty in the heaping mechanism with a bootstrap approach. We describe our implementation of the method using *R* and graphically illustrate our results with the **ggplot2** package. Application to our data suggests that recall errors are a more important source of bias than actual heaping.

References

Griffith, S. D., S. Shiffman, and D. F. Heitjan (2009, November). A method comparison study of timeline followback and ecological momentary assessment of daily cigarette consumption. *Nicotine & tobacco research 11*(11), 1368–73.

Shiffman, S. (2009, September). How many cigarettes did you smoke? Assessing cigarette consumption by global report, Time-Line Follow-Back, and ecological momentary assessment. *Health psychology 28*(5), 519–26.

Wang, H. and D. F. Heitjan (2008, August). Modeling heaping in self-reported cigarette counts. *Statistics in medicine 27*(19), 3789–804.