Accepted Manuscript

On nonparametric estimation of the latent distribution for ordinal data

Sujit K. Ghosh, Christopher Burns, Daniel Prager, Li Zhang, Glen Hui

PII: S0167-9473(17)30209-8
DOI: https://doi.org/10.1016/j.csda.2017.10.001
Reference: COMSTA 6517

To appear in: *Computational Statistics and Data Analysis*

Received date : 28 September 2016
Revised date : 2 October 2017
Accepted date : 3 October 2017

Please cite this article as: Ghosh S.K., Burns C., Prager D., Zhang L., Hui G., On nonparametric estimation of the latent distribution for ordinal data. *Computational Statistics and Data Analysis* (2017), https://doi.org/10.1016/j.csda.2017.10.001

On Nonparametric Estimation of the Latent Distribution for Ordinal Data

Sujit K. Ghosh[1], Christopher Burns, Daniel Prager[2], Li Zhang and Glen Hui[3]

Abstract

Ordinal data collected in surveys often consist of numerical scores that have a natural ordering. Observed values of ordinal variables can be thought of as a manifestation of an underlying continuous latent variable that is related to the observed ordinal variable through a set of threshold values or "cut-points." The "cut-points" partition the latent variable into intervals corresponding to the observed levels of the ordinal variable. This latent distribution is of interest to researchers for purposes of descriptive statistics and statistical modeling. However, restrictive parametric assumptions about the latent distribution are often not adequate. A nonparametric model based on mixtures of scaled Beta distributions is presented and estimation is carried out using a version of Anderson-Darling statistic-based criteria which is shown to be computationally more efficient than likelihood based criteria. A Monte Carlo simulation shows that the proposed model and estimation method performs well and is robust against any underlying continuous distribution. Several empirical examples based on ordinal data from the household section of the Agricultural Resource Management Survey (ARMS) illustrate the versatility and adaptability of the method in practice.

Keywords: Anderson-Darling statistic; Bernstein polynomials; Quadratic programming; Survey data.

*Corresponding author
Email address: sujit.ghosh@ncsu.edu (Sujit K. Ghosh)
[1]Department of Statistics, NC State University
[2]Economic Research Service, USDA
[3]George Mason University