



Mixture of shifted binomial distributions for rating data

Shaoting Li¹ · Jiahua Chen^{2,3}

Received: 19 April 2021 / Revised: 18 October 2022 / Accepted: 10 January 2023 /

Published online: 10 February 2023

© The Institute of Statistical Mathematics, Tokyo 2023

Abstract

Rating data are a kind of ordinal categorical data routinely collected in survey sampling. The response value in such applications is confined to a finite number of ordered categories. Due to population heterogeneity, the respondents may have several different rating styles. A finite mixture model is thus most suitable to fit datasets of this nature. In this paper, we propose a two-component mixture of shifted binomial distributions for rating data. We show that this model is identifiable and propose a numerically stable penalized likelihood approach for parameter estimation. We adapt an expectation-maximization algorithm for the penalized maximum likelihood estimation. Our simulation results show that the penalized maximum likelihood estimator is consistent and effective. We fit the proposed model and other models in the literature to some real-world datasets and find the proposed model can have much better fits.

Keywords Binomial distribution · Categorical data · Identifiability · EM algorithm · Mixture model · Ordinal data · Rating data

✉ Jiahua Chen
jhchen@stat.ubc.ca

¹ School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, Liaoning, China

² School of Mathematics and Statistics, Yunnan University, Kunming 650091, Yunnan, China

³ Department of Statistics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada