DSSV 2023 DATA SCIENCE, STATISTICS & VISUALISATION
ECDA 2023 EUROPEAN CONFERENCE ON DATA ANALYSIS

Exploring the Equivalence of two Mixture Models for Rating Data in the CUB class

Ambra Macis[1], Matteo Ventura[1], Marica Manisera[1], and Paola Zuccolotto[1]

[1]University of Brescia, Department of Economics and Management, C.da Santa Chiara, 50, 25122, Brescia, Italy

Questionnaires are commonly used in social, psychological, and marketing research to assess latent traits like respondents' perceptions, opinions, and attitudes, which are usually measured through rating scales, such as the Likert and Semantic Differential scales. Ratings provided by respondents are typically ordinal data, meaning that the categories have a natural ordering. Analyzing ordinal data can be challenging because traditional statistical methods do not account for the inherent nature and properties of ordinal variables. To address this issue, specific models have been developed, such as the CUB (Combination of discrete Uniform and shifted Binomial random variables) class of models, which is a notable example in the literature. The CUB model, introduced by Piccolo and D'Elia in 2005, is designed to model rating data measured on a Likert-type response scale. The model assumes that respondents start their decision-making process from the bottom of the scale and move upward based on their sensations. The underlying decision process is characterized by two components: feeling and uncertainty. The feeling component represents the strength of respondents' preferences for a particular item or attribute, while the uncertainty component reflects the degree of uncertainty associated with their responses. The shifted Binomial distribution models the feeling component, while the discrete Uniform distribution models the uncertainty component. By modeling the decision process underlying rating data, the CUB model provides researchers with a powerful tool for identifying the factors that influence respondents' ratings [1]. The CUB model has been modified and extended by various researchers, resulting in the development of the CUB class of models, which is a flexible framework for analyzing rating data.

The decision process underlying ratings can vary depending on the type of the response scale used. To gain valuable insights into the underlying attitudes, beliefs, and opinions of respondents, it is essential to develop appropriate models that take into account the specific features of each response scale. For example, multi-point Semantic Differential scales ask respondents to rate a concept or object by positioning themselves on a scale between two bipolar adjectives. In such cases, respondents may start their decision process from the middle of the response scale and move up or down based on their sensations. To analyze ordinal data measured on Semantic Differential scales, an extension of the CUB paradigm, called CUM model (Combination of discrete Uniform and linearly transformed Multinomial random variable), has been proposed [2]. The CUM model assumes that respondents start their decision process from the middle of the Semantic Differential scale and estimates three latent components: the propensity of formulating a judgment shifted towards the left extreme of the scale, the propensity of obtaining a judgment shifted towards the right extreme of the scale, and the uncertainty component.

It is worth noting that many rating scales, such as the commonly used Likert scale, include a middle option that indicates a neutral or indifferent response between two extremes. In this case, the same reasoning that is assumed for Semantic Differential scales can then be extended to these scales. As a result, there is a conceptual overlap between Semantic Differentials and Likert scales, and both the CUB and the CUM models can be applied to analyze rating data from these scales.

In our research, we aim to investigate if there is also an analytical overlap between the CUB and CUM models. Specifically, we explore the relationships between the basic CUB and the CUM models by focusing on the possible overlapping of their parametric spaces. By doing so, we want to provide new insights into the relationship between the basic CUB and the CUM models, which can be useful in analyzing ordinal data from various rating scales.

Keywords: rating, ordinal data, semantic differential scales, mixture model

[1] A. D'Elia and D. Piccolo, "A mixture model for preferences data analysis," *Computational Statistics & Data Analysis*, vol. 49, no. 3, pp. 917–934, 2005.

[2] M. Manisera and P. Zuccolotto, "A mixture model for ordinal variables measured on semantic differential scales," *Econometrics and Statistics*, vol. 22, pp. 98–123, 2022.