UNIVERSITY
OF BRESCIA

Department of Economics and Management

Ph.D. Program in
Metodi e Modelli per l'Economia e il Management
Analytics for Economics and Management (AEM)

Scientific sector:
STAT-01/A – Statistics
XXXVII Cycle

Advances in Mixture Models for Ordinal Data

Theoretical Insights and Model-based Clustering

Author:
Matteo Ventura

Supervisor:
Prof. Paola Zuccolotto – University of Brescia
Co-supervisor:
Prof. Julien Jacques – Lumière University Lyon 2

A.Y. 2023/2024

ABSTRACT

This doctoral thesis focuses on the analysis of ordinal data, specifically rating data, which has received limited attention in the literature and poses several challenges due to its unique characteristics. The first part of the thesis provides a comprehensive overview of the principal models used for analyzing ordinal data, beginning with the Generalized Linear Models framework and extending to more recent specialized distributions, including the CUB and BOS models.

The second part of this thesis presents the research conducted over the past two years, with a particular emphasis on the development of applications, theoretical insights, and new models within the framework of the CUB class introduced in the first section. This part follows a structured progression, allowing the reader to build on each new contribution as they are introduced.

The first contribution, presented in Chapter 3, extends the so-called CUM model, a specific approach for analyzing rating data from Semantic Differential scales. Originally proposed for general use and specifically developed for seven-point scales, this thesis introduces a novel adaptation of the CUM model for analyzing data from five-point Semantic Differential Scales. The performances of the model have been tested both with simulation studies and applications to real data.

The second contribution, presented in Chapter 4, compares the CUB and CUM models in the context of five- and seven-category scales. Specifically, this work aims to analytically investigate the conditions under which the CUB and CUM models are equivalent.

The third contribution, presented in Chapter 5, applies the CUM model to seven-point Semantic Differential Scales, with a dual aim. First, it demonstrates how the model works in practice and how it can be used to analyze ordinal data. Second, it offers a valuable contribution to both society and the city of Brescia (Italy), as this research was conducted within the "DS4BS — Data Science for Brescia" project, aimed at analyzing visitors' perceptions of the city's Art Gallery.

The fourth contribution, presented in Chapter 6, was developed during a visiting period at the ERIC Laboratory at the University Lumière Lyon 2 in France. This project introduced a Mixture Model for analyzing rating data within the CUB framework. Simulation studies were conducted to evaluate the model's performance, and it was subsequently applied to real data to demonstrate its practical application.

SOMMARIO

Questa tesi di dottorato si concentra sull'analisi dei dati ordinali, nello specifico dati di rating, un tipo di dati che ha ricevuto una limitata attenzione nella letteratura e che presenta diverse sfide a causa delle sue caratteristiche uniche. La prima parte della tesi fornisce una panoramica esaustiva dei principali modelli utilizzati per l'analisi dei dati ordinali, partendo dal contesto dei Modelli Lineari Generalizzati e arrivando a distribuzioni più recenti e specifiche, come i modelli CUB e BOS.

La seconda parte della tesi presenta la ricerca condotta negli ultimi due anni, con particolare enfasi sullo sviluppo di applicazioni, approfondimenti teorici e nuovi modelli all'interno del framework della classe CUB introdotto nella prima sezione. Questa parte segue una progressione strutturata, permettendo al lettore di costruire su ciascun nuovo contributo man mano che viene presentato.

Il primo contributo, presentato nel Capitolo 3, estende il cosiddetto modello CUM, un approccio specifico per l'analisi dei dati di valutazione provenienti da scale a differenziale semantico. Originariamente proposto per un uso generale e sviluppato specificamente per scale a sette punti, questa tesi introduce un adattamento innovativo del modello CUM per l'analisi di dati provenienti da scale a differenziale semantico con sette categorie. Le prestazioni del modello sono state testate sia con studi di simulazione che con applicazioni a dati reali.

Il secondo contributo, presentato nel Capitolo 4, confronta i modelli CUB e CUM nel contesto di scale con cinque e sette categorie. In particolare, questo lavoro mira a investigare analiticamente le condizioni in cui i modelli CUB e CUM sono equivalenti.

Il terzo contributo, presentato nel Capitolo 5, applica il modello CUM alle scale a differenziale semantico con sette categorie, con un duplice obiettivo. In primo luogo, dimostra come il modello funzioni nella pratica e come possa essere utilizzato per analizzare dati ordinali. In secondo luogo, offre un contributo utile sia alla società che alla città di Brescia (Italia), in quanto questa ricerca è stata condotta all'interno del progetto "DS4BS — Arts and Cultural Places", con l'obiettivo di analizzare le percezioni dei visitatori della Pinacoteca della città.

Il quarto contributo, presentato nel Capitolo 6, è stato sviluppato durante un periodo di visita presso il Laboratorio ERIC dell'Università Lumière Lyon 2 in Francia. Questo progetto ha introdotto un Modello Mistura per l'analisi dei dati di valutazione all'interno del framework CUB. Sono stati condotti studi di simulazione per valutare le prestazioni del modello, che è stato successivamente applicato a dati reali per dimostrarne l'applicazione pratica.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Part I

THEORY AND METHODS

The first part of this thesis sets the theoretical framework needed to understand the scientific contributions outlined in the subsequent section.

The initial chapter defines ordinal data and outlines their key characteristics. It also provides a brief overview of common methods for modeling this type of data, including approaches to handling cumulative probabilities in a regression framework, assuming the existence of a latent continuous variable, and developing customized distributions for modeling ordinal data.

The second chapter focuses on the CUB class of models, specifically designed for analyzing data from rating scales. The objective of these models is to measure two latent traits influencing raters' responses. The main works in this area are summarized, with particular attention to the basic CUB model, which set the basis for the whole class of models, and the CUM model which has been developed for analyzing data from Semantic Differential scales. Understanding these models is essential for comprehending the novel contributions explained in the subsequent part of the thesis.

1 MODELLING ORDINAL DATA DISTRIBUTIONS

Over the past few decades, there has been an increasing interest in the application of statistical sciences and methodologies in different disciplines like marketing, social sciences, and psychology. In most of these areas of study, research is made with the support of questionnaires, which are tools used to collect the respondents' opinions about a certain issue, by asking the respondents to rate their perceptions.

Latent trait theory provides the foundational framework for these questionnaires and the research based upon them. It facilitates the measurement of traits and abilities that are not directly observable. The theory posits that the observed responses of a person to items or questions are a result of the person's level of the underlying trait, which is said to be *latent* because it is not directly measurable or observable, but it can only be inferred from the responses to a set of items (Hambleton and Cook, 1977).

Therefore, the latent trait can be explained (or predicted) by observable and measurable characteristics of the respondents (Birnbaum, Lord, and Novick, 1968). The link between the observable traits and the latent trait is defined by a mathematical function; thus it is possible to refer to them as latent trait models. Several methods and techniques suitable for the framework of latent traits have been defined and are applied in several fields (Borsboom, Mellenbergh, and Van Heerden, 2003).

The psychologist Stanley Smith Stevens in 1946 proposed the well-known classification of variables which divides the variables into four types: *(i)* nominal variables represent categories without any inherent order (examples include gender or ethnicity); *(ii)* ordinal variables have a meaningful order or ranking, but the intervals between categories are not necessarily equal (e.g., education levels, customer satisfaction ratings); *(iii)* interval variables are similar to ordinal variables but with equal intervals between categories (e.g., temperature); *(iv)* ratio variables have equal intervals between categories like interval variables, but they also have a true zero point (e.g., height, weight, and income).

Considering the classification developed by Stevens, ordinal variables frequently arise in questionnaire-based research, where respondents provide subjective responses through rating scales. There are various types of rating scales (Nunnally and Bernstein, 1994): the Numeric Rating Scale (NRS) in which participants provide a numerical response to indicate their level of agreement or disagreement with a statement; the well-known Likert Scale (Likert, 1932) which typically presents a set of response options, asking participants to rate their level of agreement or disagreement; Semantic Differential Scale (Osgood, 1962; Osgood, Suci, and Tannenbaum, 1957) This scale asks respondents to rate a concept or object using pairs of opposite adjectives (e.g., Good - Bad, Sweet - Bitter, Cold - Warm).

Ordinality, as discussed earlier, is a fundamental attribute of measurement meanings, as established by Stevens in his seminal work on measurement theory in 1946 (Stevens, 1946). This unique characteristic necessitates the development of specialized probability distributions and models to account for it. Over the years, various approaches have been explored to create probability distributions that are well-suited for handling ordinal data. This thesis concentrates on a specific type of ordinal data: rating data ob-

tained from Likert and Semantic Differential scales. Three primary approaches suitable for handling such data can be distinguished (Biernacki and Jacques, 2016):

1. *Modeling Cumulative Probabilities*: this first approach involves modelling the cumulative probabilities associated with ordinal data, rather than the individual probabilities. This method takes into account the accumulation of probabilities as we move up the ordinal scale,
2. *Latent Continuous Variable Approach* the second approach is based on the concept of an underlying continuous latent random variable. It assumes that the observed ordinal data are essentially a discretization of this latent variable.
3. *Customized Distributions for Ordinal Data*: the third approach involves the creation of custom probability distributions tailored specifically for handling ordinal data. These distributions are designed with unique properties that align with the ordinal nature of the data, offering a more tailored and accurate representation.

The choice of the approach depends on the specific characteristics and requirements of the data being analyzed, as well as the objectives of the statistical analysis. In this chapter, a brief overview of the methods mentioned above is given.

1.1 MODELING CUMULATIVE PROBABILITIES

The first methodology for modeling ordinal data focuses on handling the cumulative probabilities associated with ordinal categories. This approach leverages the inherent order of the data to ensure that these probabilities increase monotonically as one progresses up the ordinal scale.

In scenarios involving binary response variables, logistic regression is the predominant model. For multinomial variables, an extension of logistic regression is the Baseline-category Logit model, which forms logits by pairing each category with a baseline category. However, this method is typically more suited to scenarios where the categories of the response variable are not ordered. This is because it does not account for the ordinality of the data, which is a crucial characteristic of ordinal variables.

To address ordinality more appropriately, the cumulative probabilities approach modifies logistic regression by applying transformations that consider the order of the categories. A common transformation is the logit transformation applied to the cumulative probabilities, enhancing the model's ability to capture the ordered nature of the data. Other transformations, such as probit or log-log, can also be utilized depending on the specific characteristics of the data and the analysis requirements.

Furthermore, two additional transformations are noteworthy when dealing with ordinal data: the logarithmic transformation applied to the odds of adjacent categories, and the use of the continuation ratio. These methods refine the approach by focusing on the relationships between consecutive categories, thus providing a more detailed modeling of the ordinal data's structure. This class of models requires a set of covariates to account for ordinality; otherwise, the resulting model would be a multinomial model.

In the following, frequentist approaches are explained, but also bayesian approaches have been developed for estimating these models. Detailed information are in Section 11 of the book by Agresti (2010), and in the book by Johnson and Albert (2006), both dedicated to dedicated to the analysis of ordinal data from a Bayesian perspective.

CUMULATIVE LOGIT MODELS Let R be an ordinal variable with m categories, whose realizations are identified by r, and let $\pi_1, \dots, \pi_m$ be the probabilities of observing each category. The *cumulative logits* are defined as:

$$\begin{aligned} \text{logit}[P(R \leqslant r)] &= \log \frac{P(R \leqslant r)}{1 - P(R \leqslant r)} && (1.1) \\ &= \log \frac{\pi_1 + \cdots + \pi_r}{\pi_{r+1} + \cdots + \pi_m} \quad r = 1, \dots, m\,. && (1.2) \end{aligned}$$

Given a set of p explanatory variables $\mathbf{X}$, a model which incorporates them and simultaneously uses all $r-1$ cumulative logits can be defined as follows:

$$\text{logit}[P(R_i \leqslant r)] = \alpha_r + \boldsymbol{\gamma}' \mathbf{x}_i \tag{1.3}$$

for $r = 1, \dots, m-1$. In this model, $\mathbf{x}_i$ is a column vector of the values of the explanatory variables for the ith subject, and $\boldsymbol{\gamma}$ is a column vector of parameters which describes the effects of each explanatory variables on the response variable R.

α_r is the intercept of the logit for cumulative probability r, and it is the log-odds of failing into or below category r when all the covariates are equal to 0. The values of α increase in r because $P(R \leqslant r)$ increases in r for each fixed value of $\mathbf{X}$, and the logit is an increasing function of its probability (Agresti, 2010).

In the model described by 1.3 the effects of $\boldsymbol{\gamma}$ are the same for each cumulative logit. Each coefficient γ represents the effect of the associated variable on the dependent variable. Namely, it represents the increase in log-odds of falling into or below the associated category r with a one-unit increase in the dependent variable x_q. Therefore, a positive slope indicates a tendency for the response level to increase as the variable increases.

The cumulative probabilities can also be expressed as follows:

$$P(R \leqslant r) = \frac{\exp(\alpha_r + \boldsymbol{\gamma}' \mathbf{x})}{1 + \exp(\alpha_r + \boldsymbol{\gamma}' \mathbf{x})}, \quad r = 1, \dots, m-1\,. \tag{1.4}$$

The general with multiple covariates defined in (1.3) satisfies:

$$\text{logit}[P(R \leqslant r \mid \mathbf{x}_1)] - \text{logit}[P(R \leqslant r \mid \mathbf{x}_2)]$$

$$= \log \frac{P(R \leqslant r \mid \mathbf{x}_1)/P(R > r \mid \mathbf{x}_1)}{P(R \leqslant r \mid \mathbf{x}_2)/P(R > r \mid \mathbf{x}_2)} = \boldsymbol{\gamma}'(\mathbf{x}_1 - \mathbf{x}_2)\,.$$

That is, the chances of obtaining a response $R \leqslant r$ at $\mathbf{x} = \mathbf{x}_1$ are $\exp[\boldsymbol{\gamma}'(\mathbf{x}_1 - \mathbf{x}_2)]$ times the odds at $\mathbf{x} = \mathbf{x}_2$. The log cumulative odds ratio is proportional to the distance between $\mathbf{x}_1$ and $\mathbf{x}_2$.

The parameters of the model can be estimated through Maximum Likelihood, as proposed by Walker and Duncan (1967), and then generalized for all models for cumulative probabilities by McCullagh (1980), but there exist also Bayesian approaches (Congdon, 2005).

By changing the Cumulative Link function, it is possible to obtain the so-called cumulative probit model that is defined as follows:

$$\Phi^{-1}[P(R \leqslant r)] = \alpha_r + \boldsymbol{\gamma}' \mathbf{x}, \qquad r = 1, \dots, m-1. \tag{1.5}$$

As in the cumulative logit model, the effect $\boldsymbol{\gamma}$ is the same for each cumulative probability.

ADJACENT-CATEGORIES LOGIT MODELS The *adjacent-categories logits* are formally defined as follows:

$$\text{logit}[P(R = r \mid R = r \text{ or } R = r+1)] = \log\frac{\pi_r}{\pi_{r+1}}, \quad r = 1, \dots, m-1. \tag{1.6}$$

That is equal to applying the well-known binary logit to the conditional probability of observing the response r, given the response outcome in category r or $r+1$.

If covariates are introduced to the model, it becomes as follows:

$$\text{logit}[P(R = r \mid R = r \text{ or } R = r+1)] = \alpha_r + \boldsymbol{\gamma}_r' \mathbf{x}, \quad r = 1, \dots, m-1. \tag{1.7}$$

In this model, the effects of the variables are described by local odds ratios rahter than cumulative odds ratios which describe the effect of the variable in the cumulative logit.

To ensure model parsimony, it can be supposed that the explanatory variables have similar effects for each logit, this allows to use a single parameter instead of $m-1$ parameters to describe that effect. Therefore, the model becomes:

$$\text{logit}[P(R = r \mid R = r \text{ or } R = r+1)] = \alpha_r + \boldsymbol{\gamma}' \mathbf{x}, \quad r = 1, \dots, m-1. \tag{1.8}$$

In this model, for predictor q, the coefficient γ_q represents the change in the log-odds of falling into category $r+1$ instead of falling into category r when x_q increases by one unit, keeping all the other covariates constant. Each odds-ratio is equal to $\exp(\gamma_q)$.

CONTINUATION-RATIO LOGITS The *continuation-ratio logits* are particularly useful when the outcome of the response variable can be determined by a sequential mechanism, meaning that an observation must potentially occur in category r before it can occur in a higher category. This happens, for example, with the survival of a person through different age periods.

The continuation-ratio logits can be defined in two ways according to the direction in which the sequential mechanism works. If the sequential mechanism is increasing, the continuation-ratio logits are defined as follows:

$$\log\frac{\pi_r}{\pi_{r+1} + \cdots + \pi_m}, \quad r = 1, \dots, m-1, \tag{1.9}$$

while, if the sequential mechanism decreases, they are defined as:

$$\log\frac{\pi_{r+1}}{\pi_1 + \cdots + \pi_r}, \quad r = 1, \dots, m-1. \tag{1.10}$$

The two formulations are not equivalent, but the most appropriate one has to be chosen according to the direction of the sequential mechanism characterising the ordinal variable.

Models with covariates can also be defined for continuation-ratio logits, with both a specific effect and an equal effect for each logit.

1.2 UNDERLYING CONTINUOUS VARIABLE APPROACH

Modeling ordinal variables through a underlying continuous variable approach involves conceptualizing the ordinal responses r_i as manifestations of an underlying continuous latent variable Y. Given an ordinal variable R measured on a response scale

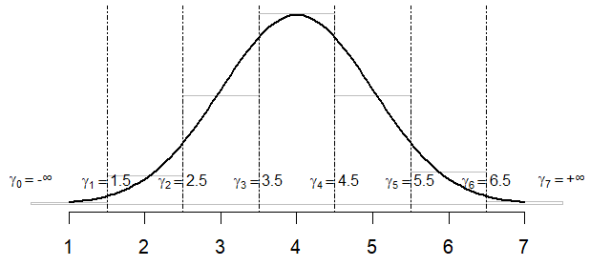


Figure 1.1: Example of the discretization of a continuous variable into an ordinal variable with $m = 7$ categories and $m + 1 = 8$ thresholds γ.

with m ordered categories, $r = 1, \ldots, m$, the vector $\boldsymbol{\gamma}$ contains $m + 1$ thresholds that segment the real line into m intervals. The observed value y_i of the random variable Y, in relation to $\boldsymbol{\gamma}$, determines the value of the observed ordinal response.

Each threshold in $\boldsymbol{\gamma}$ helps define the intervals on the real line, with the configuration $-\infty = \gamma_0 \leqslant \gamma_1 \leqslant \cdots \leqslant \gamma_m = \infty$. This setting ensures that the measured value y_i from the latent variable Y falls within a specific interval defined by these thresholds, thus determining the observed ordinal response. For example, if $\gamma_{r-1} \leqslant y_i \leqslant \gamma_r$, then the corresponding ordinal response is $y_i = r_i$.

The latent variable Y follows a specific probability distribution, characterised by a probability density function $f(y)$. The probability of observing the level c_r is then computed as the difference between the cumulative probabilities at these thresholds, formally expressed as $P(y_i = c_r = r_i) = F(y_{c_r}) - F(y_{c_{r-1}})$, where F denotes the cumulative distribution function associated with Y.

In statistical modelling of ordinal data, Gaussian distributions have been preferred for modelling Y (Johnson and Albert, 2006; McParland and Gormley, 2016). However, the Gaussian model assumes symmetry and light tails, which may not always fit well with real-world data, particularly when data exhibit skewness and heavier tails.

Recent developments in statistical methodology have introduced the discretized Beta distribution as an alternative to Gaussian models in the handling of ordinal data. This distribution is particularly beneficial due to its flexibility in shaping its form to fit the data better, accommodating varying levels of skewness and kurtosis. This adaptability makes the Beta distribution a good choice for modelling ordinal data, as it can more accurately reflect the underlying distributions of latent traits, especially in scenarios where traditional models like the Gaussian fail to provide an adequate fit (Fasola and Sciandra, 2015; Simone, 2022; Simone and Tutz, 2018; Tamhane, Ankenman, and Yang, 2002; Ursino, 2014; Ursino and Gasparini, 2018).

1.3 CUSTOMIZED DISTRIBUTIONS

Among the various approaches to modeling ordinal data, a third methodology has emerged, involving the definition of a model that allows the generation of customized distributions characterized by specific properties based on the aspects and features of

the ordinal variable one aims to emphasize. This approach goes beyond the use of standard models, offering greater flexibility in representing ordinal data. However, in the current literature, there are still limited examples of this innovative methodology.

Among the few known instances, two prominent models stand out: the CUB (Combination of discrete Uniform and shifted Binomial distribution) model (D'Elia and Piccolo, 2005; Piccolo, 2003) and the BOS (Binary Ordinal Search) model (Biernacki and Jacques, 2016). These models exemplify the exploration of tailored distributions to capture nuanced characteristics within ordinal variables, showcasing the potential of this approach.

1.3.1 *The CUB Class of models*

As the acronym suggests, the CUB model is a mixture model which combines a Shifted Binomial and a discrete Uniform random variable. The primary objective of the CUB model is to capture the cognitive processes occurring in respondents' minds when they are tasked with rating their opinions on specific statements. The fundamental assumption underlying the CUB model (D'Elia and Piccolo, 2005; Piccolo, 2003), and its associated class of models (Piccolo and Simone, 2019) is that the selection of a rating by a respondent is not only the outcome of an elicitation process, but there are other factors which influence rating chosen by the respondent. Indeed, the final rating is the result of a more complex process with involves two latent components, which are called *feeling* and *uncertainty* in the CUB framework.

The feeling component represents the rational part in the respondents' mind, which reflects consciousness and complete understanding of the item on which the respondent is asked to express a rating (Tourangeau, Rips, and Rasinski, 2000). The interpretation of the feeling can be different according to the topic of the questionnaire and of the item to which the respondent is asked to express his opinion. Therefore it can be related to a perception, an emotion, or a sensation.

The second component that is assumed to act on the final rating is the uncertainty, which represents the typical indecision that arises when attempting to position oneself on a rating scale (Zhou and Lange, 2009). This component can have different importance on the determination of the final rating and moreover it can be caused by different factors such as personal inclinations of the respondent, lack of knowledge about the topic the respondent is asked to rate his opinion, amount of time that is required for filling the questionnaire.

To model the joint impact of the feeling and uncertainty components on the final rating distribution, the CUB model has been developed as a mixture model that assigns weights to a shifted Binomial and a discrete Uniform, representing the feeling and uncertainty components, respectively.

These assumptions about the unconscious mechanisms influencing the final rating lay the foundation of a new paradigm (Piccolo, 2018) that has been the starting point for the development of numerous models within the so-called CUB Class, which find applications in various fields.

Further information about the CUB model and the CUB Class are given in the next chapter.

1.3.2 *The BOS Model*

The second customized distribution, known as the BOS model, was introduced by Biernacki and Jacques (2016). The acronym BOS stands for "Binary Ordinal Search", referring to the algorithm at the core of this model.

The BOS distribution is based on the assumption that the ordinal variable R results from a search process conducted within an ordered set of categories $\{1, \ldots, m\}$, where m represents the total number of categories. Initially, the exact value of the ordinal variable R is unknown. However, through repeated comparisons with elements from the set $\{1, \ldots, m\}$, the true value of μ is progressively uncovered.

An additional assumption is that during the search process, some erroneous results $r \in \{1, \ldots, m\}$ may occur. To reduce the possibility of such errors, it becomes important to minimize the number of comparisons made throughout the search process.

The Binary Search Algorithm, which accounts for the possibility of incorrect comparisons, iterates $m-1$ times, yielding an outcome $r \in \{1, \ldots, m\}$. Starting from an interval $\gamma_h = \{g_h^-, \ldots, g_h^+\} \subset \{1, \ldots, m\}$, the hth iteration is structured into three steps:

- Step 1: Select an element $y_h \in \gamma_h$ to divide the interval γ_h. The break point y_h is chosen uniformly from γ_h, hence:

$$P(y_h \mid \gamma_h) = \frac{1}{|\gamma_h|} \mathbb{I}(y_h \in \gamma_h), \tag{1.11}$$

  where $\mathbb{I}(\cdot)$ is the indicator function, and the length of the interval is given by $|\gamma_h| = g_h^+ - g_h^-$.

- Step 2: Determine the accuracy of the comparison between y_h and μ, denoted by $a_h \in \{0, 1\}$. The accuracy a_h is modeled as a Bernoulli random variable with parameter $\zeta \in [0, 1]$, such that:

$$P(a_h \mid \gamma_h; \zeta) = \zeta \mathbb{I}(a_h = 1) + (1 - \zeta)\mathbb{I}(a_h = 0), \tag{1.12}$$

  where $a_h = 0$ represents a blind comparison (i.e., μ is not used in the comparison), and $a_h = 1$ represents a perfect comparison (i.e., μ is used). Erroneous comparisons can only arise from blind comparisons ($a_h = 0$).

- Step 3: Based on the outcomes y_h and a_h, a new search interval $\gamma_{h+1} \subseteq \{\gamma_h^-, \gamma_h^=, \gamma_h^+\}$ is chosen, where:
  - $\gamma_h^- = \{g_h^-, \ldots, y_h - 1\}$ represents the interval to the left of the break point y_h,
  - $\gamma_h^= = \{y_h\}$ is the break point itself,
  - $\gamma_h^+ = \{y_h + 1, \ldots, g_h^+\}$ is the interval to the right of the break point.

  The selection of the new interval γ_{h+1} depends on whether the comparison was perfect or blind:
  - If the comparison is perfect ($a_h = 1$), the interval that contains μ is selected with certainty:

$$P(\gamma_{h+1} \mid y_h, \gamma_h, a_h = 1; \mu) = \mathbb{I}\left(\gamma_{h+1} = \underset{\gamma \in \{\gamma_h^-, \gamma_h^=, \gamma_h^+\}}{\operatorname{argmin}} \delta(\gamma, \mu)\right) \times \mathbb{I}\left(\gamma_{h+1} \in \{\gamma_h^-, \gamma_h^=, \gamma_h^+\}\right), \tag{1.13}$$

where $\delta(\gamma, \mu)$ represents the "distance" between the interval $\gamma = \{b^-, \ldots, b^+\}$ and the value μ, defined as:

$$\delta(\gamma, \mu) = \min\left(\left|\mu - b^-\right|, \left|\mu - b^+\right|\right).$$

– If the comparison is ($a_h = 0$), the new interval γ_{h+1} is chosen randomly, with a probability proportional to the size of the intervals:

$$P(\gamma_{h+1} \mid y_h, \gamma_h, a_h = 0) = \frac{|\gamma_{h+1}|}{|\gamma_h|}\mathbb{I}\left(\gamma_{h+1} \in \{\gamma_h^-, \gamma_h^=, \gamma_h^+\}\right). \tag{1.14}$$

The distribution of R can be obtained by marginalizing over α_h, using equations (1.12), (1.13), and (1.14):

$$\begin{aligned} P(\gamma_{h+1} \mid \gamma_h, y_h; \mu, \zeta) = {} & \zeta P(\gamma_{h+1} \mid y_h, \gamma_h, \alpha_h = 1; \mu) \\ & + (1 - \zeta) P(\gamma_{h+1} \mid y_h, \gamma_h, \alpha_h = 0). \end{aligned} \tag{1.15}$$

Next, α_h is marginalized over y_h, combining it with equation (1.11):

$$P(\gamma_{h+1} \mid \gamma_h; \mu, \zeta) = \sum_{y_h \in \gamma_h} P(\gamma_{h+1} \mid \gamma_h, y_h; \mu, \zeta) P(y_h \mid \gamma_h). \tag{1.16}$$

Finally, the distribution $P(R = r; \mu, \zeta)$ can be derived using equation (1.16):

$$\begin{aligned} P(R = r; \mu, \zeta) &= \sum_{\gamma_{m-1}, \ldots, \gamma_1} P(\gamma_m, \gamma_{m-1}, \ldots, \gamma_1; \mu, \zeta) \\ &= \sum_{\gamma_{m-1}, \ldots, \gamma_1} \prod_{h=1}^{m-1} P(\gamma_{h+1} \mid \gamma_h; \mu, \zeta) P(\gamma_h). \end{aligned} \tag{1.17}$$

The BOS model is defined by two key parameters: the location parameter μ and the precision parameter ζ. The parameter μ represents the mode of the distribution, which is always unique. Consequently, the BOS model cannot capture distributions with multiple adjacent modes. On the other hand, ζ controls the precision of the model: as ζ increases, the mode μ becomes increasingly pronounced. In the limiting case where $\zeta = 1$, the distribution collapses to a Dirac delta function centered at μ. Conversely, when $\zeta = 0$, the distribution becomes uniform over $1, \ldots, m$. Furthermore, the model is identifiable as long as $\zeta > 0$.

Additional properties of the BOS model, along with their proofs, can be found in the appendices of the foundational work by Biernacki and Jacques (2016).

2 THE CUB CLASS OF MODELS

Among the approaches to analyse rating data proposed in the literature, the one defined by the CUB (Combination of a discrete Uniform and shifted Binomial random variable) class (Piccolo and Simone, 2019) is particularly interesting. This class of models sets a framework based on the assumption that the final rating of the respondents is the outcome of a combination of two latent components: the *feeling* and the *uncertainty*. From a statistical point of view, these two components are modelled through a mixture distribution that combines a discrete Uniform and a shifted Binomial random variable, which respectively models the uncertainty and the feeling.

The forthcoming sections delve into an exploration of some models and paradigms falling within the CUB class, serving as foundational framework for both theoretical and practical advancements discussed in subsequent chapters. Specifically, Section 2.1 describes the fundamental CUB model (D'Elia and Piccolo, 2005; Piccolo, 2003). Subsequently, the Decision Process paradigm (Manisera, Zuccolotto, et al., 2014), which is a shared framework for various models within the CUB class, is delineated in section 2.2.1. Finally, section 2.2.2 introduces a model within the CUB class which is tailored to analyse data coming from Semantic Differential scales (Manisera and Zuccolotto, 2022).

2.1 THE BASIC CUB MODEL

The basic CUB model was originally proposed by D'Elia and Piccolo (2005) as a development of the Shifted Binomial (SB) (D'Elia, 2000) which considers rating data as the outcomes of a paired comparisons criterion and proposed to model the rank r assigned by a rater to an item among m as the realisations of a Shifted Binomial random variable, which is shifted since in ordinal scales the first category is usually represented by the number 1 (Piccolo, 2006).

The probability mass function of the shifted Binomial, therefore, is:

$$P(R = r \mid \xi) = \binom{m-1}{r-1}(1-\xi)^{r-1}\xi^{m-r}, \quad r = 1, \dots, m. \tag{2.1}$$

where the success parameter is $(1-\xi)$, which decreases as the positive feeling towards the item increases since it is assumed that $R = 1$ means "most preferred", and $R = m$ means "least preferred".

The paradigm which characterizes the CUB model and the related class of models states that the final rating chosen by a respondent is the result of the joint action of two latent components in his mind: the *feeling* and the *uncertainty*, which, respectively, refer to the subjective experience of an individual towards an object or an event, and the inherent indecision present in every human choice that can be due, for example, to the lack of information or knowledge about the same object or event. A second possible interpretation of the CUB model and its parameters is that this model describes the behavior of two clusters of the population: a cluster of persons who answer randomly, and a cluster of individuals who take a completely rational decision (Piccolo and Simone, 2019).

Given the observed rating $r = 1, \dots, m$, where m is the number of categories, according to the CUB model it is the realization of a mixed discrete random variable R with probability mass defined as:

$$P(R = r \mid \theta) = \pi P_B(r \mid \xi) + (1 - \pi) P_U(m), \tag{2.2}$$

where $\theta = (\xi, \pi)'$, with $\pi \in (0, 1]$ and $\xi \in [0, 1]$. Therefore, the basic CUB model has a parameter space given by $\theta = \{(0, 1] \times [0, 1]\}$. The random variable R is a mixture distribution of (i) a shifted Binomial random variable B with trial parameter m and probability mass function $P_B(r \mid \xi)$ and (ii) a discrete Uniform random variable U defined over the support $S = \{1, \dots, m\}$ and with probability mass function $P_U(m)$.

The probability mass function of the basic CUB model (2.2) can also be written as:

$$P(R = r \mid \theta) = \pi \left[\binom{m-1}{r-1} \xi^{m-r} (1 - \xi)^{r-1} \right] + (1 - \pi) \frac{1}{m}, \tag{2.3}$$

where $r = 1, 2, \dots, m$ is the rating chosen by the respondent. The basic CUB model has been proved to be identifiable for $m > 3$ (Iannario, 2010b).

The shifted Binomial component accounts for the feeling which is measured by the success parameter $1 - \xi$. The CUB model was initially proposed as a model for modelling rating data and, therefore, the parameter ξ was the most meaningful since, in a rank, lower numbers are assigned to the favourite objects (D'Elia and Piccolo, 2005). Then, the model has been widely used for analysing rating data, therefore the parameter $1 - \xi$ became more meaningful from an interpretation point of view (Piccolo and Simone, 2019). Indeed, in the shifted Binomial, this parameter represents the probability that in each paired comparison a category is preferred to the previous one (D'Elia, 2000). Hence, higher values of $1 - \xi$ denote a greater feeling toward the object or sentence being evaluated by the respondent.

The feeling parameter $1 - \xi$ impacts the asymmetry of the CUB distribution. The distribution exhibits a symmetric shape only when the feeling parameter is equal to 0.5. High values of the feeling parameter result in right-skewed asymmetry, indicating a stronger preference for higher categories; conversely, low values of the feeling parameter lead to left-skewed asymmetry, signifying a greater preference for lower categories.

The Uniform component, instead, addresses uncertainty, quantified by $1 - \pi$: elevated values of this quantity denote increased uncertainty. Similar to the feeling parameter, the uncertainty parameter also influences the distribution's shape. Specifically, the CUB distribution exhibits a flatter shape as $1 - \pi$ increases.

2.1.1 *Maximum Likelihood Estimation for the CUB Model*

The Maximum Likelihood estimates of the model can be easily obtained through EM algorithm (D'Elia and Piccolo, 2005; Dempster, Laird, and Rubin, 1977), commonly used for estimating mixture models. See Appendix A for a general description of the EM algorithm.

In order to implement the EM estimates of the parameters of the CUB model, the mixture component the rating data belongs to is considered unknown.

Let $R = (r_i)_{i=1,\dots,n}$ be a univariate ordinal random variable with m categories and $Z = (z_i)_{i=1,\dots,n}$ be a latent random variable distributed as a Bernoulli random variable, $Z \sim \text{Bin}(\pi)$, where $z_i = 1$ if the ith rather's preference comes from the kth distribution,

$z_i = 0$ otherwise. The observed data are $\mathbf{r} = (r_1, \ldots, r_n)'$, while the unobserved data are represented by $\mathbf{z} = (z_1, \ldots, z_n)'$. The complete data log-likelihood is:

$$\ell_c(\theta \mid \mathbf{r}, \mathbf{z}) = \sum_{i=1}^{n} z_i \ln[\pi P_B(r_i \mid \xi)] + (1 - z_i) \ln[(1-\pi) P_U(m)] \tag{2.4}$$

At each iteration, the EM algorithm alternates between an Expectation step and a Maximization step. The algorithm continues to iterate until it satisfies a predefined stopping criterion.

E-STEP (t-TH ITERATION) In the E-step, the expected value of the complete log-likelihood with respect to the conditional distribution of the indicator variable Z, given the data $\mathbf{r}$, and the parameters $\theta^{(t)}$ is computed:

$$Q(\theta, \theta^{(t)}) = \mathbb{E}_{Z|\mathbf{r},\theta^{(t)}}[\ell_c(\theta \mid \mathbf{r}, \mathbf{z})].$$

Let $P(z_i = 1) = \pi^{(t)}$, and $P(z_i = 0) = 1 - \pi^{(t)}$, the conditional expectation of Z, conditioned on the data $\mathbf{r}$ and the parameters $\theta^{(t)}$, $\mathbb{E}_{Z_i|\mathbf{r},\theta^{(t)}}$, is computed as follows:

$$\mathbb{E}_{Z_i=1|r_i,\theta^{(t)}} = \frac{\pi^{(t)} P_B(r_i \mid \xi^{(t)})}{\pi^{(t)} P_B(r_i \mid \xi^{(t)}) + (1-\pi^{(t)}) P_U(m)} = \tau_i^{(t)},$$

$$\mathbb{E}_{Z_i=0|r_i,\theta^{(t)}} = \frac{(1-\pi^{(t)}) P_U(m)}{\pi^{(t)} P_B(r_i \mid \xi^{(t)}) + (1-\pi^{(t)}) P_U(m)} = (1 - \tau_i^{(t)}).$$

Therefore, $Q(\theta, \theta^{(t)})$ becomes:

$$\begin{aligned} Q(\theta, \theta^{(t)}) = &\sum_{i=1}^{n} \tau_i^{(t)} \ln(\pi) + (1 - \tau_i^{(t)}) \ln(1-\pi) + \\ &+ \sum_{i=1}^{n} \tau_i^{(t)} \ln[P_B(r_i \mid \xi)] + (1 - \tau_i^{(t)}) \ln[P_U(m)] \end{aligned} \tag{2.5}$$

M-STEP (t-TH ITERATION) In this step, the expected value computed in the previous step is considered as a function of π and ξ, and therefore, $Q(\theta, \theta^{(t)})$ is maximized with respect to π and ξ in order to obtain the estimators of the parameters.

Equation (2.5) can be rewritten as:

$$Q(\theta, \theta^{(t)}) = Q_1^{(t)}(\pi) + Q_2^{(t)}(\xi), \tag{2.6}$$

such that

$$\begin{aligned} Q_1^{(t)}(\pi) &= \sum_{i=1}^{n} \tau_i^{(t)} \ln(\pi) + (1 - \tau_i^{(t)}) \ln(1-\pi), \\ Q_2^{(t)}(\xi) &= \sum_{i=1}^{n} \tau_i^{(t)} \ln[P_B(r_i \mid \xi)] + (1 - \tau_i^{(t)}) \ln[P_U(m)], \end{aligned}$$

which can be maximized separately to obtain the estimates of the parameters.

$Q_1^{(t)}(\pi)$ is maximized by solving the equation

$$\frac{\partial Q_1^{(t)}(\pi)}{\partial \pi} = \frac{\sum_{i=1}^{n} \tau_i^{(t)}}{\pi} - \frac{\sum_{i=1}^{n} (1 - \tau_i^{(t)})}{1 - \pi} = 0.$$

Therefore, the updated estimate of π is computed as follows:

$$\pi^{(t+1)} = \frac{\sum_{i=1}^{n} \tau_i^{(t)}}{n} = \frac{1}{n} \sum_{i=1}^{n} w_i, \tag{2.7}$$

The estimator of ξ is obtained by solving the first derivative with respect to ξ of the second component of equation (2.6), $Q_2^{(t)}(\xi)$:

$$\frac{\partial Q_2^{(t)}(\xi)}{\partial \xi} = \sum_{i=1}^{n} w_i \left[\frac{m - r_i}{\xi} - \frac{r_i - 1}{1 - \xi} \right] = 0,$$

so the updated estimate of ξ are obtained as follows:

$$\xi^{(t+1)} = \frac{m - \frac{\sum_{i=1}^{n} w_i r_i}{\sum_{i=1}^{n} w_i}}{m - 1} = \frac{m - \bar{R}_n(p)}{m - 1} \tag{2.8}$$

where $\bar{R}_n(p)$ is the average of the observed ranks weighted with the posterior probability that r_i is a realization of the shifted Binomial, given the current data.

The algorithm is initialized by choosing as starting values of the parameters $\pi^0 = \frac{1}{2}$ and $\xi^0 = (m - \bar{R}_n)/(m - 1)$, and it is stopped when a threshold $\epsilon = 10^{-10}$ is reached in the relative change of the log-likelihood: $|\ell(\xi^{(t+1)}), \pi^{(t+1)} - \ell(\xi^{(t)}, \pi^{(t)})| < \epsilon$.

The goodness of fit of the estimated CUB model can be assessed through a dissimilarity index which determines the percentage of observed data that should be changed for the model to perfectly fit the data. The dissimilarity index is defined as follows:

$$\text{Diss} = \frac{1}{2} \sum_{r=1}^{m} | f_r - P(R = r \mid \theta) |, \tag{2.9}$$

where f_r is the relative frequency of the category r. The index is a normalized index, meaning that $\text{Diss} \in [0, 1]$. Lower values of Diss mean better fitting (Iannario, 2009).

The estimated model can be represented on a Cartesian plane that represents the parameter space of the CUB models.

2.1.2 *The CUB model with Covariates*

The responses to rating scales are usually collected with a set of covariates that can be useful to explain the behavior of respondents and summarizes the available information about respondents.

A CUB model can be specified such that it accounts for the effect of covariates (Piccolo, 2006; Piccolo et al., 2003): it can be assumed that the uncertainty parameter π is a function of p covariates contained in the matrix $\mathbf{Y}$, and the feeling parameter is a function of q subjects' covariates contained in the matrix $\mathbf{W}$.

In order to specify a correspondence among the real-valued matrices $\mathbf{Y}$ and $\mathbf{W}$ and the parameters $\pi \in (0, 1]$ and $\xi \in [0, 1]$ it has been proposed a logistic mapping defined by:

$$\begin{cases} \text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{y}_i\boldsymbol{\beta}; \\ \text{logit}(\xi_i) = \ln\left(\frac{\xi_i}{1-\xi_i}\right) = \mathbf{w}_i\boldsymbol{\gamma}; \end{cases} \iff \begin{cases} \pi_i = \frac{1}{1+e^{-y_i\beta}}; \\ \xi_i = \frac{1}{1+e^{-\mathbf{w}_i\gamma}}; \end{cases} \tag{2.10}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the parameters vectors to be estimated and $\mathbf{y}_i$ and $\mathbf{w}_i$ are the row vectors of the matrices $\mathbf{Y}$ and $\mathbf{W}$ respectively, corresponding to the covariates values for the i-th subject. For easiness, $y_{i0} = w_{i0} = 1, i = 1, \ldots, n$.

The $\text{logit}(1-\pi_i)$ may also be interpreted as the log-odds that the i-th subject has a propensity to be uncertain rather than meditated in his/her choice (Piccolo and Simone, 2019).

Therefore, the general CUB model with covariates is defined as follows:

$$P(R = r \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{y}_i, \mathbf{w}_i) = \frac{1}{1+e^{-y_i\beta}}\left[\binom{m-1}{r-1}\frac{(e^{-\mathbf{w}_i\gamma})^{r-1}}{(1+e^{-\mathbf{w}_i\gamma})^{m-1}} + \frac{1}{m}\right] + \frac{1}{m} \tag{2.11}$$

It has to be noticed that even if this approach reminds the GLM framework, the CUB does not belong to this class of models because the CUB is not a member of the exponential family (Piccolo, 2006; Piccolo and Simone, 2019).

2.1.3 *The CUB model with shelter*

An important extension of CUB models addresses the situation where respondents disproportionately favor a specific category beyond what the model predicts. Various reasons have been proposed to explain why some individuals show a strong preference for a particular option: reducing mental effort, fatigue, habitual response patterns, desire for privacy, poorly phrased questions, the numbering of the scale, and more. Depending on the context, this category may be seen as safe, appealing, or politically correct. Generally, this preselected category is referred to as a "shelter option," as respondents choose it as a fallback instead of making a more thoughtful decision.

For a given shelter category $s \in \{1, \ldots, m\}$, $D_r^{(s)}$ is a random variable with mass concentrated in $R = s$:

$$D_r^{(s)} = \begin{cases} 1, \text{ if } r = s; \\ 0, \text{ otehrwise;} \end{cases} \qquad r = 1, \ldots, m. \tag{2.12}$$

It is possible to provide two different interpretations of the shelter effect. The first interpretation relates to the respondent adopting a two-step strategy: first, the respondent selects either a simplistic option (the shelter category) with probability δ, or a general response with probability $1-\delta$. If the respondent chooses the latter, the final score reflects a combination of their feelings and some degree of uncertainty (Iannario, 2012). This interpretation can be formally expressed as follows:

$$P(R = r \mid \pi, \xi, \delta) = \delta\left[D_r^{(s)}\right] + (1-\delta)\left[\pi P_B(r \mid \xi) + (1-\pi)P_U(r)\right]. \tag{2.13}$$

The second possible interpretation is that the respondent acts solely according to feeling with probability λ, or according to a non-motivated behavior with probability $1-\lambda$. If the respondent engages in non-motivated behavior, he/she may select a category at random with probability ϕ, or choose a shelter category with probability $1-\phi$.

Formally, this behavior is specified as follows:

$$P(R = r \mid \pi, \xi, \lambda, \phi) = \lambda P_B(r \mid \xi) + (1 - \lambda)\left[\phi P_U(r) + (1 - \phi) D_r^{(s)}\right]. \tag{2.14}$$

The first model should be chosen when the quantification of the shelter δ is required, while the second model should be chosen when it is supposed that the respondents follow a "satisfacing" behavior (Piccolo and Simone, 2019).

2.2 DEVELOPMENTS WITHIN THE CUB CLASS

Over the years, numerous researchers have contributed to the development and extension of the CUB model class, introducing various modifications and advancements. Recent contributions have focused on alternative approaches for modeling the feeling component, aiming to account for different response style characteristics related to this component. Notably, Manisera and Zuccolotto (2014a) introduced the Nonlinear CUB (NL-CUB) model, which incorporates non-constant transition probabilities between categories of the rating scale. This innovation implies that the respondent's path to the final rating is non-linear, offering a more flexible approach to understanding decision-making in ordinal data. The NL-CUB model allowed the authors to propose a potential formalization of a Decision Process that respondents may follow, providing a conceptual framework for developing new models within the CUB class. In fact, by modifying certain elements of the general paradigm, it is possible to generate new models that belong to this class. This paradigm will be thoroughly explained in Subsection 2.2.1, as it is essential for understanding the novel contributions to the CUB class presented in this thesis.

Piccolo (2015) further expanded the CUB class by proposing the CUBE model, which employs the discretized Beta-binomial distribution (Tripathi, Gupta, and Gurland, 1994) to account for an excess of inter-subject variability in ordinal data. Additionally, Tutz et al. (2017) introduced the CUP model, where the feeling component can be modeled using any ordinal model. More recently, Manisera and Zuccolotto (2022) proposed the CUM model (Combination of discrete Uniform and linearly transformed Multinomial random variable), which models the feeling component using a linearly transformed Multinomial distribution instead of the traditional shifted Binomial. This model will be discussed in detail in subsection 2.2.2, as the research presented in this thesis is particularly focused on it.

For the uncertainty component, Gottard, Iannario, and Piccolo (2016) proposed various alternatives to the traditional Uniform distribution to capture different uncertainty patterns. These include the trimmed Uniform, left/right bounded Uniform, Triangular, and Symmetric parabolic distributions. Simone and Tutz (2018) introduced the CAUB model (Combination of Adjusted Uniform and shifted Binomial), utilizing the discretized Beta distribution to account for preferences toward middle or extreme categories. Tutz and Schneider (2019) suggested the Beta-binomial distribution as a more flexible way to model the uncertainty component.

Additionally, an important advancement addresses the modeling of rating scales that include a "Don't know" option. Manisera and Zuccolotto (2014b) proposed adjusting the uncertainty parameter to account for this option, improving the accuracy of uncertainty representation in such cases.

Since the CUB model is inherently univariate, several multivariate extensions have been proposed to handle correlated ordinal variables. Andreis and Ferrari (2013), Cor-

duas (2015) and Corduas et al. (2011), suggested using copulas with CUB marginals to model bivariate distributions, allowing for the flexible capture of dependencies between two ordinal variables. In contrast, Ip and Wu (2024) proposed a model for bivariate correlated ordinal variables that does not rely on copulas, offering an alternative approach to the modeling of such dependencies.

Colombi and Giordano (2016) introduced a mixture of a Uniform distribution and a Sarmanov distribution (Sarmanov, 1966) with CUB marginals to account for the association among a respondent's answers to different items in a questionnaire. This approach distinguishes between two types of uncertainty: specific uncertainty, which relates to indecision on individual items, and global uncertainty, reflecting the respondent's overall hesitancy in completing the entire questionnaire.

Additionally, Simone, Tutz, and Iannario (2020) proposed a generalized mixture model designed to account for subjective heterogeneity in response behavior for multivariate ordinal data. This model incorporates random effects to capture individual differences in the propensity toward a structured or uncertain response style, providing a framework for modeling the diversity in how respondents approach rating tasks.

Several new approaches have combined the CUB model with other statistical methods to extend its use. These include Decision Trees (Cappelli, Simone, and Di Iorio, 2019; Simone, 2023; Simone, Cappelli, and Di Iorio, 2019), Latent Class analysis (Grilli et al., 2014), Structural Equation modeling (Carpita, Ciavolino, and Nitti, 2019), and Cluster Analysis (Biasetton et al., 2023; Corduas, 2010). Additionally, improvements have been made to the model's computational aspects (Cerulli et al., 2022; Simone, 2021), and it has also been applied in a Bayesian context (Deldossi and Paroli, 2012, 2015).

These methodological advancements have been successfully applied in various fields. For example, the CUB model has been used to analyze consumer preferences in the agri-food sector (Arboretti and Bordignon, 2016; Brentari, Manisera, and Zuccolotto, 2018; Capecchi et al., 2016; Cicia et al., 2010; Corduas, Cinquanta, and Ievoli, 2013; Gottard, Iannario, and Piccolo, 2016; Grilli et al., 2014; Lamonaca et al., 2022; Manisera and Zuccolotto, 2014a,b, 2016), and study medical data (Colombi and Giordano, 2016). It has also been applied in political science (Simone and Tutz, 2018), job satisfaction research (Gambacorta and Iannario, 2013; Punzo, Castellano, and Buonocore, 2018), consumer satisfaction analysis (Cafarelli et al., 2015), and educational sciences (Cafarelli and Crocetta, 2016; Iannario et al., 2020).

2.2.1 *A formalization of the Decision Process*

Following the CUB framework, Manisera and Zuccolotto (2014a) proposed a possible paradigm to formalize the Decision Process, which is based on the presence of the feeling and uncertainty components, which are supposed to derive from the unconscious combination in the respondents' minds of two separate reasoning approaches used to express their final choice, the above-mentioned feeling and the uncertainty component. The feeling approach follows a *feeling path*, which is a step-by-step process such that the rating r_T, chosen at the end of the process, is the result of T elementary judgements. Each elementary judgement corresponds to a step of the feeling path, which leads the respondent to a provisional rating.

Instead, the uncertainty component is related to the *uncertainty path*, which leads the individual to express a random rating.

The DP can be summarised in three main parts, i.e., the path followed in the feeling approach, the path followed in the uncertainty approach, and the final rating, which is the realization of the random variable R, generated as a mixture of the random variables R_T and Q with weight π. R_T is the random variable from which the last rating formulated at the end of the feeling path, r_T, is generated. Instead, Q is the random variable which generates the uncertainty judgement q.

The feeling approach consists of T steps which correspond to T elementary judgements, x_t, generated from an *i.i.d.* sequence of random variables $X_1, \dots, X_T$. At each step t, an accumulating function $f(\cdot)$, which maps the Cartesian product of the supports of $X_1, \dots, X_T$ into $\mathbb{R}$, summarizes the t past elementary judgements. Then, the accumulated judgements, $w_t = f(x_1, \dots, x_T)$, are generated at each step as the realizations of a sequence of random variables $W_1, \dots, W_T = f(X_1, \dots, X_T)$. Consequently, at each step t the accumulated judgement w_t is transformed into a provisional rating through the non-decreasing function $d(w_t)$ called "Likertization" function. Finally, the provisional ratings $r_1, \dots, r_T$, with $r_t = d(w_t)$, are generated as realizations of the sequence of random variables $R_1, \dots, R_T$, with $R_t = d(W_t)$ with support the space $(1, \dots, m)$.

The uncertainty approach consists of a value q, called uncertainty judgement, which is the realization of a random variable Q with support S in $(1, \dots, m)$.

Finally, the expressed rating r is the realization of a random variable R which is a mixture of R_T and Q, with weight π and $1 - \pi$.

The Decision Process formalised by Manisera and Zuccolotto (2014a) allows us to model elementary judgements both using continuous and discrete random variables. Since the models presented in this thesis are models for rating data, only discrete DP will be considered.

An interesting characteristic of the DP is the possibility to compute the transition probability, which is the probability of moving from a provisional rating $r_t = s$ to the provisional rating $r_{t+1} = s + 1$, denoted with $\phi_t(s) = P(R_{t+1} = s + 1 \mid R_t = s)$.

It is also useful to consider the average over t of the transition probabilities $\phi(s) = av_t(\phi_t(s))$, since this value can be interpreted as a perceived closeness between ratings s and $s + 1$: a high value of $\phi(s)$ means a small gap between the two ratings, therefore it is easy for the respondent to move from the rating s to the rating $s + 1$. Conversely, lower values of $\phi(s)$ indicate a large gap between the ratings.

The Decision Process paradigm of the basic CUB Model

According to the Decision Process previously described, in the feeling approach the respondents ask themselves for $T = m - 1$ times if they feel a positive sensation about the item they are asked to rate, and they give an instinctive response at each time. In the end, 1 plus the total of the positive responses, is the last rating r_T of the feeling process.

More formally, in the basic CUB model the T elementary judgements are generated from a sequence of random variables $X_1, \dots, X_T$ which follow a Bernoulli distribution with success parameter $1 - \xi$. Thus each elementary judgement $x_t \in \{0, 1\}$, where 1 means agreement with the evaluated item. The accumulating function $f(\cdot)$ which summarizes the t past elementary judgements at each step t, in the CUB model, is an additive function, meaning that $W_t = X_1 + \cdots + X_t$. Then, the accumulated judgements, w_t, are generated at each step as the realizations of a sequence of random variables $W_1, \dots, W_T$, which follow a Binomial distribution with parameters t and $1 - \xi$. Conse-

quently, at each step t the accumulated judgement w_t is transformed into provisional ratings r_t using the "Likertization" function $d(w_t) = w_t + 1$. Finally, the provisional ratings $r_1, \dots, r_T$, with $r_t = d(w_t)$, are generated as realizations of a Shifted Binomial distribution with parameters $t + 1$ and $1 - \xi$.

Instead, in the CUB model, the uncertainty judgement is the realization of the random variable Q which follows a Uniform distribution with support in the discrete space $(1, \dots, m)$.

Finally, as expressed by the probability mass function of the CUB model, the expressed rating r is the realization of a random variable R which is a mixture of shifted Binomial and a discrete Uniform random variable weighted by π.

Since the DP of the CUB model is linear, the transition probabilities are constant and equal to $\phi_t(s) = P(X_t = 1) = 1 - \xi$.

2.2.2 *The CUM Model for Semantic Differential Scales*

The Decision Process paradigm defined in the previous sections sets a possible starting point for the development of new models within the CUB class. Indeed, new models can be created by modifying the assumptions regarding the Decision Process that drives the choice of the respondents and the distributions that characterize the feeling and/or the uncertainty judgements.

A notable example of such development within the CUB Class is represented by the so-called CUM model (Manisera and Zuccolotto, 2022), where the acronym stands for Combination of discrete Uniform and Multinomial random variable.

The CUM model was specifically proposed for the analysis of data coming from Semanitc Differential scales (Osgood, 1962; Osgood, Suci, and Tannenbaum, 1957), which are one of the possible scales which are available to measure opinions and perceptions (Dawis, 1987). In this type of response scale, the extremes are represented by two opposite adjectives, and the central point of the scale represents neutrality. By positioning on the left or the right of the middle point, respondents express their agreement with one or the other adjective. With its type of rating scales, the DP of the respondent is assumed to start from the middle of the scale and then move upward or downward according to the prevalent sensations coming to their mind.

In the formulation of the CUB model through the Decision Process paradigm, the respondents are assumed to start from the bottom of the scale and move upward, instead, in the Semantic Differential scale, respondents are assumed to start from the middle point and move upward or downward depending on their feeling. The model was thought to model the ratings given on a multi-point semantic differential scale with an odd number of options, therefore the number of categories m is equal to $2k + 1$ where $k = T$, and it represents the number of steps that can be done on the feeling path.

Therefore, within the DP, the elementary judgements are generated by the sequence of random variables $\mathbf{X}_t = [X_{t,D}, X_{t,U}, X_t]$ which follow a Multinoulli distribution which parameters are ξ_D and ξ_U which respectively represents the probability of moving towards the lower or the upper part of the scale. The parameters are defined such that $\xi_D, \xi_U \leqslant 1$ and $\xi_D + \xi_U \leqslant 1$. Thanks to those constraints, it is possible to compute the probability of remaining still on the scale, which is the complement to one of the sum of ξ_D and ξ_U. When the elementary judgement determines a step towards the left part of the scale, $X_{t,D} = 1$ with probability ξ_D. The same happens with probability

ξ_U for $X_{t,U}$ when the elementary judgement determines a step towards the right part of the scale. Finally, when the elementary judgement determines that the respondent stays still in the middle of the scale, $X_t = 1$ with probability $1 - \xi_D - \xi_U$.

As for the CUB model, the accumulating function is additive, meaning that they are the T realizations of a sequence of random variables $M_1, \dots, M_T$, where each $\mathbf{M}_t = X_1 + \cdots + X_t$. Since the accumulating function is additive, the accumulated judgments m_t are generated at each step t as the realizations of the random variable $\mathbf{M}_t = [M_{t,D}, M_{t,U}, M_t]$ following a multinomial distribution with parameters $[\xi_D, \xi_U]$ and the trial parameter t.

The accumulated judgments are then transformed in provisional ratings through a likertization function that is a linear transformation of the Multinomial random variable $\mathbf{M}_t$. This transformation allows to modify the support of the random variable such that it ranges from 1 to m. The provisional ratings R_t follow a linear transformed Multinomial distribution and the final rating r_T derives from R_T which is described by the random variable W_k, which is defined as follows:

$$W_k = [-1\ 1\ 0]\,\mathbf{M}_k + k + 1. \tag{2.15}$$

By recalling that $k = T$, the Multinomial random variable $\mathbf{M}_k = \left[M_{k,D}, M_{k,U}, M_k\right]'$ represents the number of times a positive, negative or neutral basic judgement is formulated and the probability mass function of this random variable represents the probability of obtaining a specific number of positive, negative or neutral basic judgements. The probability mass function of $\mathbf{M}_k(k_1, k_2)$ is defined as a function of ξ_D, ξ_U and k:

$$P_M(\xi_D, \xi_U, k_1, k_2) = \frac{k!}{k_1! \cdot k_2! \cdot (k - k_1 - k_2)!} \cdot \xi_D^{k_1} \cdot \xi_U^{k_2} \cdot (1 - \xi_D - \xi_U)^{k - k_1 - k_2}, \tag{2.16}$$

where k_1 represents the number of negative elementary judgements, k_2 represents the number of positive judgements, and the quantity $(k - k_1 - k_2)$ represents the number of neutral judgements.

The probability mass function of W_k, instead, has to be specifically defined given the number of categories of the multi-point semantic differential scale. However, it can be generalised as follows:

$$P_W(W_k = k + h) = \begin{cases} \mathbf{M}_k\left(\max(-2\alpha, 0), \max(2\alpha, 0)\right) + \\ + \mathbf{M}_k\left(\max(-2\alpha + 1, 1), \max(2\alpha + 1, 1)\right) + \cdots + \\ + \mathbf{M}_k\left(d(k/2) - \alpha, d(k/2) + \alpha\right) \\ \text{if } \ h = 2\alpha + 1 \\ \\ \mathbf{M}_k\left(\max(-2\alpha + 1, 0), \max(2\alpha - 1, 0)\right) + \\ + \mathbf{M}_k\left(\max(-2\alpha + 2, 1), \max(2\alpha, 1)\right) + \cdots + \\ + \mathbf{M}_k\left(u(k/2) - \alpha, u(k/2) + \alpha - 1\right) \\ \text{if } \ h = 2\alpha \end{cases} \tag{2.17}$$

with $h \in \{-k + 1, -k + 2, \dots, k + 1\}$ and $\alpha \in \{-d(k/2), -d(k/2) + 1 \dots, u(k/2)\}$, where $d(\cdot)$ and $u(\cdot)$ denote rounding up and down; and $\mathbf{M}_k(k_1, k_2)$ denotes the probability that the Multinomial random variable with k trials is equal to $(k_1, k_2, k - k_1 - k_2)$.

Until now the feeling approach has been described. However, in the CUB framework, the final rating is supposed to be the result of the combination of feeling and

uncertainty. In the CUM model, as in the basic CUB model, the uncertainty approach is modelled by a discrete Uniform distribution, P_U, whose support is $\mathcal{S} = \{1, \dots, m\}$.

Finally, the random variable R can be obtained as a mixture of P_W and P_U:

$$P(R = r \mid \theta_{CUM}) = \pi P_W(r \mid \xi_D, \xi_U) + (1 - \pi) P_U, \tag{2.18}$$

where $\theta_{CUM} = (\pi, \xi_D, \xi_U)'$, in this section identified as θ to lighten the notation.

CUM7: The CUM Model for Seven-points Semantic Differential scales

In the work by Manisera and Zuccolotto (2022) the probability mass function of W_k for a CUM model for seven-points semantic differential scales is defined. In the following chapters, the CUM model for five-points semantic differential scales will be defined as a contribution of this thesis.

If a CUM model with 7 categories is considered, it means that the parameter k is equal to 3. Therefore, according to the realisations of $M_{3,D}$ and $M_{3,U}$, the random variable W_3 assumes the values reported in Table 2.1.

Table 2.1: Values assumed by the r.v. W_3 for the possible realizations of the Multinoulli r.v.s $M_{3,D}$ and $M_{3,U}$

W_3		$M_{3,U}$			
		0	1	2	3
$M_{2,D}$	0	4	5	6	7
	1	3	4	5	-
	2	2	3	-	-
	3	1	-	-	-

Recalling equation 2.17 the probability that $P_W = r$ with $r = 1, \dots, m$ can be computed as follows:

$$\begin{aligned}
P_W(R = k + h = 3 - 2 = \mathbf{1}) &= \mathbf{M_3}(\mathbf{3}, \mathbf{0}) && h = -2, \alpha = -1 \\
P_W(R = k + h = 3 - 1 = \mathbf{2}) &= \mathbf{M_3}(2, 0) && h = -1, \alpha = -1 \\
P_W(R = k + h = 3 + 0 = \mathbf{3}) &= \mathbf{M_3}(1, 0) + \mathbf{M_3}(2, 1) && h = 0, \alpha = 0 \\
P_W(R = k + h = 3 + 1 = \mathbf{4}) &= \mathbf{M_3}(0, 0) + \mathbf{M_3}(1, 1) && h = 1, \alpha = 0 \\
P_W(R = k + h = 3 + 2 = \mathbf{5}) &= \mathbf{M_3}(0, 1) + \mathbf{M_3}(1, 2) && h = 2, \alpha = 1 \\
P_W(R = k + h = 3 + 3 = \mathbf{6}) &= \mathbf{M_3}(0, 2) && h = 3, \alpha = 1 \\
P_W(R = k + h = 3 + 4 = \mathbf{7}) &= \mathbf{M_3}(0, 3) && h = 4, \alpha = 2
\end{aligned}$$

with $h \in \{-2, -1, \cdots, 3, 4\}$, $\alpha \in \{-1, 0, \dots, 2\}$, and $\mathbf{M_3}$ being a Multinomial random variable defined as in equation 2.16.

Maximum Likelihood Estimation for the CUM Model

Since the model is a finite mixture, the parameters are estimated using the EM algorithm, which is usual for such types of models. In the original work by Manisera and Zuccolotto (2022), the complete data log-likelihood is computed by introducing the unobserved random variable $Z : (z_i)\ i = 1, \dots, n$, where z_i is an indicator variable which follows a Bernoulli distribution, $Z \sim \text{Bin}(\pi)$, with $z_i = 1$ if the subject i gives a

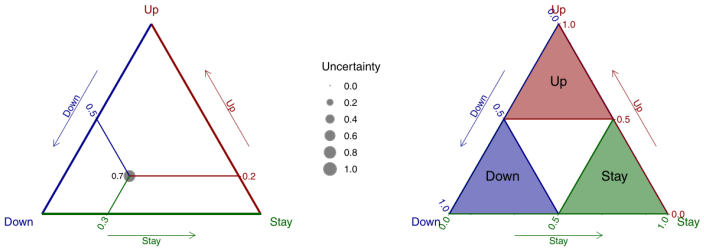


Figure 2.1: Left: graphical representation of a CUM model in a ternary plot. Right: Partition of the ternary plot into subtriangles.

rating derived from the linearly transformed Multinomial random variable, and $z_i = 0$ otherwise.

However, the same estimates can be obtained by maintaining the setting used in section 2.1.1, by replacing the probability mass function $P_B(r \mid \xi)$ with the probability mass function $P_W(r_i \mid \xi_D, \xi_U)$. Therefore, it is necessary to rewrite the second component in equation (2.6), which becomes:

$$Q_2^{(t)}(\xi_D, \xi_U) = \sum_{i=1}^{n} \tau_i^{(t)} \ln[P_W(r_i \mid \xi_D, \xi_U)] + (1 - \tau_i^{(t)}) \ln[P_U(m)].$$

The function $Q_2^{(t)}(\xi_D, \xi_U)$ is then optimized by solving the following system of nonlinear equations:

$$\begin{cases} \dfrac{\partial Q_2^{(t)}(\xi_D, \xi_U)}{\partial \xi_D} = \displaystyle\sum_{i=1}^{n} \frac{\tau_i^{(t)}}{P_W(r_i \mid \xi_D, \xi_U)} \frac{\partial P_W(r_i \mid \xi_D, \xi_U)}{\partial \xi_D} = 0 \\ \dfrac{\partial Q_2^{(t)}(\xi_D, \xi_U)}{\partial \xi_U} = \displaystyle\sum_{i=1}^{n} \frac{\tau_i^{(t)}}{P_W(r_i \mid \xi_D, \xi_U)} \frac{\partial P_W(r_i \mid \xi_D, \xi_U)}{\partial \xi_U} = 0 \end{cases} \tag{2.19}$$

where

$$\tau_i^{(t)} = \mathbb{E}_{Z_i=1|r,\theta^{(t)}} = \frac{\pi^{(t)} P_W(r_i|\xi_D^{(t)}, \xi_U^{(t)})}{\pi^{(t)} P_W(r_i|\xi_D^{(t)}, \xi_U^{(t)}) + (1 - \pi^{(t)}) P_U}.$$

The solution of this system cannot be obtained in closed form, therefore it has to be computed by using numerical methods (Remani, 2013). Additional details about the EM algorithm for the CUM model can be found in the seminal paper by Manisera and Zuccolotto (2022) and in Appendix B.

The triangular plot is a useful visual tool for interpreting and visualizing the parameters of estimated CUM models within the parameter space, as shown in Figure 2.1. Each edge of the triangle represents a different probability: the probability of moving towards the lower part of the scale, the probability of ascending to a higher position on the scale, and the probability of remaining still in the middle of the scale. The CUM model, instead, is represented by a point whose size is determined by the level of uncertainty, represented by the parameter π.

To simplify interpretation, the triangle is divided into sub-triangles: the red area indicates a dominant tendency to move upward, the blue area suggests a predominant inclination to move downward, and the green area denotes a stronger propensity to

remain at the midpoint of the scale. The white central area, meanwhile, indicates models where no particular directional inclination is dominant.

The goodness of fit for CUM models, instead, can be assessed through the dissimilarity index (2.9) defined for the CUB model, by substituting the probability mass function of the CUB model with the probability mass function for the CUM model. As it is shown in the seminal paper by Manisera and Zuccolotto (2022), the CUM model is particularly suitable for modeling bimodal distributions.

3

THE CUM MODEL FOR FIVE-POINTS SEMANTIC DIFFERENTIAL SCALES

CONTRIBUTIONS RELATED TO THIS CHAPTER:

Indexed Journals (WoS, Scopus)

[IJ1] Manisera M., Migliorati M., Ventura M., and Zuccolotto P., (2023) *A Mixture Model for the Analysis of Categorical Variables Measured on Five-point Semantic Differential Scales*, Austrian Journal of Statistics,53(3), 70-86.

From a methodological perspective, the CUM model was originally developed for the general situation of an odd number m of response categories, while simulations, case studies, and implementation in R were limited to $m = 7$, since the computations depend on the number of categories (Manisera and Zuccolotto, 2022).

The aim of the work presented in this chapter is to further investigate the functioning of the CUM model in the presence of a semantic differential response scale with $m = 5$ categories. In particular, a simulation study has been developed following the configuration in Manisera and Zuccolotto (2022) to compare results of $m = 5$ with those of $m = 7$, propose a case study, and adapt the R functions to cope with $m = 5$.

3.1 THE CUM MODEL FOR FIVE-POINT SEMANTIC DIFFERENTIAL SCALE

To develop this model, Semantic Differential Scales are considered, where 1 and 5 indicate two opposite adjectives, for example, sad and happy or completely dissatisfied and completely satisfied. First, if the rating scale has 5 categories, the feeling path has 2 steps, and starting from the middle option 3, the total number of steps $M_{k,D}$ towards the lowest rating can be 0, 1 or 2; the total number of steps $M_{k,U}$ towards the highest rating can be 0, 1 or 2. The linear transformation in (2.15) allows to obtain the ratings $1, ..., 5$ as in Table 3.1, that reports the values of the linearly transformed r.v. W_2 for the realizations of $M_{2,D}$ and $M_{2,U}$.

Table 3.1: Values assumed by the r.v. W_2 for the possible realizations of the Multinoulli r.v.s $M_{2,D}$ and $M_{2,U}$

W_2		$M_{2,U}$		
		0	1	2
$M_{2,D}$	0	3	4	5
	1	2	3	-
	2	1	-	-

Second, following Formula (2.17) applied to case $m = 5$, we have $h \in [-2+1, ..., 2+1] = [-1, 0, 1, 2, 3]$, $\alpha \in [-1, 0, 1]$, and the probability mass function of W_2 can be obtained as follows:

$$\begin{aligned}
P(W_2 = k + h = 2 - 1 = 1) &= \mathbf{M_2}(2, 0) && h = -1, \alpha = -1 \\
P(W_2 = k + h = 2 + 0 = 2) &= \mathbf{M_2}(1, 0) && h = 0, \alpha = 0 \\
P(W_2 = k + h = 2 + 1 = 3) &= \mathbf{M_2}(0, 0) + \mathbf{M_2}(1, 1) && h = 1, \alpha = 0 \\
P(W_2 = k + h = 2 + 2 = 4) &= \mathbf{M_2}(0, 1) && h = 2, \alpha = 1 \\
P(W_2 = k + h = 2 + 3 = 5) &= \mathbf{M_2}(0, 2) && h = 3, \alpha = 1
\end{aligned}$$

All the computations for the development of the EM algorithm for the Maximum Likelihood Estimation of the parameters, and the computation of the Information Matrix are shown in Appendix B.

3.2 THE SIMULATION STUDY

In this section, the results of a simulation study are shown. The study was carried out by fitting the CUM model with $m = 5$ (hereafter CUM5) to the data generated on the basis of the parameter values summarized in Table 3.2. The 18 scenarios are the same of the simulation study performed to evaluate the CUM model with $m = 7$ (CUM7) in Manisera and Zuccolotto (2022). For CUM5, iter $= 1000$ simulations with three different sample sizes ($n = 100$, $n = 500$, $n = 1000$ observations) are executed for each scenario. While for CUM7, only a sample size $n = 1000$ is considered. For more details on the simulation study of CUM7, see the seminal paper by Manisera and Zuccolotto (2022).

Table 3.2: Summary of parameter values used in the simulation study

	a			**b**			**c**		
	π	ξ_D	ξ_U	π	ξ_D	ξ_U	π	ξ_D	ξ_U
Case 1	0.3	0.2	0.1	0.5	0.2	0.1	0.7	0.2	0.1
Case 2	0.3	0.5	0.2	0.5	0.5	0.2	0.7	0.5	0.2
Case 3	0.3	0.8	0.1	0.5	0.8	0.1	0.7	0.8	0.1
Case 4	0.3	0.2	0.4	0.5	0.2	0.4	0.7	0.2	0.4
Case 5	0.3	0.4	0.5	0.5	0.4	0.5	0.7	0.4	0.5
Case 6	0.3	0.1	0.7	0.5	0.1	0.7	0.7	0.1	0.7

3.2.1 *CUM5 simulation study*

The ternary plots of the CUM5 simulation study for the 18 scenarios with sample size $n = 1000$ are represented in Figure 3.1. The plots for the cases with sample size $n = 100$ and $n = 500$ are reported in Figure B.1 and B.2, respectively, in Appendix B. The ternary plot represents ξ_U, on the red axis, labeled as "Up", ξ_D, on the blue axis, labeled as "Down", and $1 - \xi_D - \xi_U$ on the green axis, labeled as "Stay". Each estimated CUM model can be represented as a point in this plot, with coordinates given by the estimated parameters. The estimated uncertainty is included in the plot as the point size.

In general, the estimated values tend to be quite close to the true parameter value, except for Case 1a, where a portion of the estimated values tend to concentrate on a

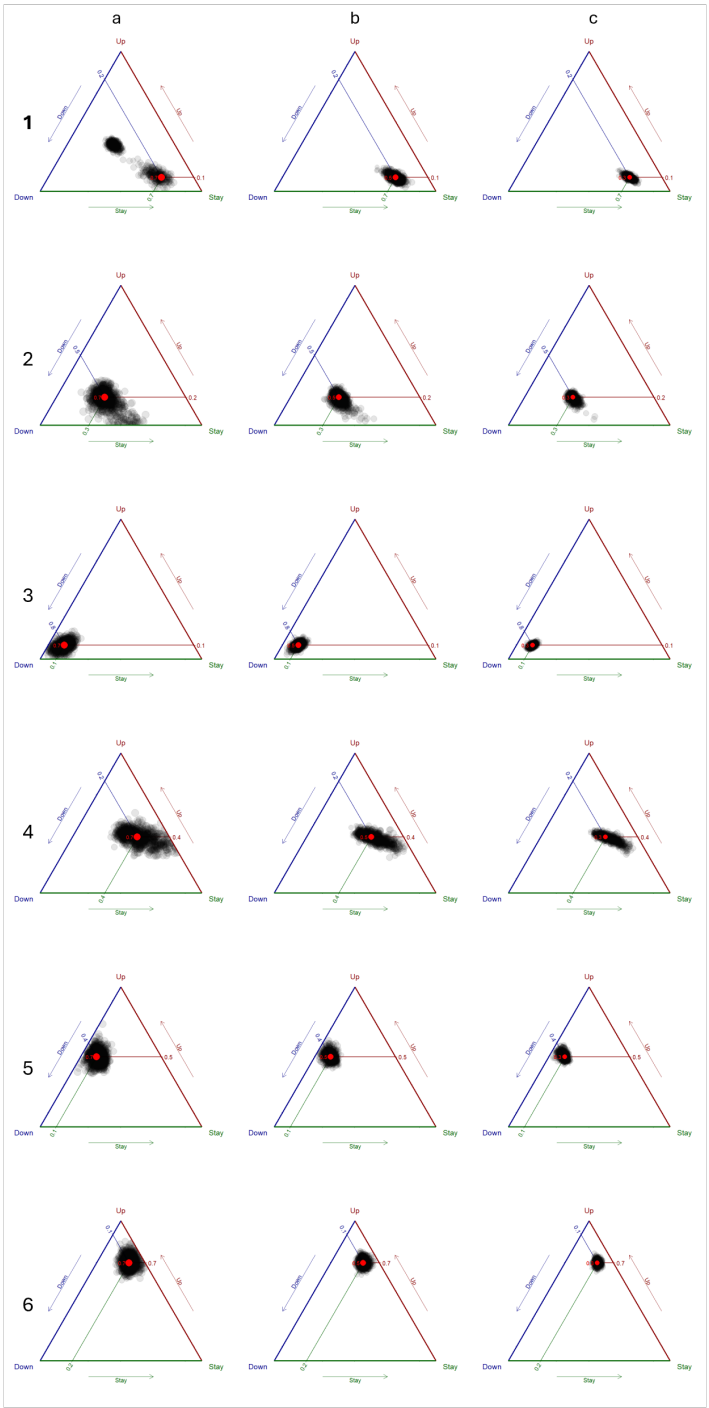


Figure 3.1: Ternary plots for CUM5 simulations ($n = 1000, iter = 1000$). Red dot: true parameter value; grey dots: estimated values.

wrong area of the parameter space. As will be clear in the next section, this pattern, although still present, tends to be less critical with CUM7. This topic will be addressed from a methodological point of view in the following chapter.

Averages and standard errors of the $\texttt{iter} = 1000$ estimated values for the three sample sizes are reported in Tables 3.3, 3.4, and 3.5.

Table 3.3: Averages and standard errors of the $\texttt{iter} = 1000$ estimated values with a sample size of 100 observations, CUM5

		a				**b**				**c**			
		π	ξ_D	ξ_U	diss	π	ξ_D	ξ_U	diss	π	ξ_D	ξ_U	diss
Case 1	True	0.3	0.2	0.1	-	0.5	0.2	0.1	-	0.7	0.2	0.1	-
	Average	0.5728	0.3272	0.2559	0.0344	0.6011	0.2444	0.1447	0.0297	0.7338	0.2099	0.1096	0.0257
	Std error	0.2188	0.105	0.1306	0.0256	0.179	0.0853	0.0998	0.0237	0.1249	0.05	0.0513	0.0201
Case 2	True	0.3	0.5	0.2	-	0.5	0.5	0.2	-	0.7	0.5	0.2	-
	Average	0.3734	0.4892	0.2028	0.0351	0.5223	0.5038	0.1845	0.0326	0.6825	0.4977	0.1761	0.0315
	Std error	0.1791	0.1449	0.1309	0.0273	0.1916	0.1012	0.0954	0.0245	0.1853	0.0777	0.0771	0.0236
Case 3	True	0.3	0.8	0.1	-	0.5	0.8	0.1	-	0.7	0.8	0.1	-
	Average	0.3668	0.7526	0.1139	0.0261	0.5319	0.7882	0.1028	0.0227	0.7052	0.8018	0.0987	0.0193
	Std error	0.1231	0.1001	0.0796	0.0197	0.1223	0.0762	0.0548	0.0171	0.1093	0.056	0.0405	0.0144
Case 4	True	0.3	0.2	0.4	-	0.5	0.2	0.4	-	0.7	0.2	0.4	-
	Average	0.393	0.223	0.4118	0.0394	0.5387	0.1925	0.3989	0.0367	0.6854	0.1776	0.3876	0.0378
	Std error	0.1922	0.1396	0.1398	0.0288	0.2037	0.1028	0.0907	0.0279	0.1885	0.0781	0.0803	0.0275
Case 5	True	0.3	0.4	0.5	-	0.5	0.4	0.5	-	0.7	0.4	0.5	-
	Average	0.4709	0.3631	0.4519	0.0282	0.6391	0.381	0.4839	0.0296	0.7758	0.3885	0.4919	0.0323
	Std error	0.2154	0.1291	0.1335	0.0214	0.2097	0.0772	0.0838	0.0218	0.1908	0.0522	0.0593	0.0247
Case 6	True	0.3	0.1	0.7	-	0.5	0.1	0.7	-	0.7	0.1	0.7	-
	Average	0.3594	0.1095	0.6731	0.0294	0.526	0.1007	0.6973	0.0249	0.7126	0.0994	0.6982	0.0225
	Std error	0.1387	0.09	0.1091	0.0215	0.1375	0.0633	0.0801	0.0186	0.1199	0.0508	0.0589	0.0174

Table 3.4: Averages and standard errors of the $\texttt{iter} = 1000$ estimated values with a sample size of 500 observations, CUM5

		a				**b**				**c**			
		π	ξ_D	ξ_U	diss	π	ξ_D	ξ_U	diss	π	ξ_D	ξ_U	diss
Case 1	True	0.3	0.2	0.1	-	0.5	0.2	0.1	-	0.7	0.2	0.1	-
	Average	0.5787	0.3248	0.2593	0.0182	0.5149	0.2043	0.1040	0.0118	0.7050	0.2017	0.1015	0.0118
	Std error	0.1807	0.0816	0.1093	0.0135	0.0783	0.0372	0.0372	0.0091	0.0551	0.0242	0.0222	0.0094
Case 2	True	0.3	0.5	0.2	-	0.5	0.5	0.2	-	0.7	0.5	0.2	-
	Average	0.3077	0.4975	0.1865	0.0140	0.4979	0.4987	0.1897	0.0139	0.6969	0.4993	0.1923	0.0136
	Std error	0.1034	0.0730	0.0785	0.0109	0.1030	0.0408	0.0537	0.0102	0.1011	0.0300	0.0406	0.0104
Case 3	True	0.3	0.8	0.1	-	0.5	0.8	0.1	-	0.7	0.8	0.1	-
	Average	0.3087	0.7976	0.0984	0.0105	0.5041	0.7999	0.0997	0.0092	0.7019	0.8002	0.0998	0.0079
	Std error	0.0647	0.0613	0.0430	0.0081	0.0571	0.0368	0.0268	0.0070	0.0489	0.0254	0.0188	0.0061
Case 4	True	0.3	0.2	0.4	-	0.5	0.2	0.4	-	0.7	0.2	0.4	-
	Average	0.3268	0.1951	0.4020	0.0161	0.5132	0.1939	0.3954	0.0153	0.6897	0.1866	0.3905	0.0154
	Std error	0.1100	0.0887	0.0640	0.0122	0.1203	0.0655	0.0452	0.0116	0.1285	0.0539	0.0399	0.0116
Case 5	True	0.3	0.4	0.5	-	0.5	0.4	0.5	-	0.7	0.4	0.5	-
	Average	0.3576	0.3887	0.4861	0.0135	0.5366	0.3971	0.4965	0.0128	0.7174	0.3992	0.5003	0.0129
	Std error	0.1089	0.0557	0.0579	0.0101	0.1264	0.0343	0.0422	0.0093	0.1327	0.0242	0.0328	0.0096
Case 6	True	0.3	0.1	0.7	-	0.5	0.1	0.7	-	0.7	0.1	0.7	-
	Average	0.3088	0.0984	0.6985	0.0118	0.5045	0.0985	0.6985	0.0111	0.7040	0.0988	0.6998	0.0106
	Std error	0.0703	0.0506	0.0625	0.0088	0.0648	0.0316	0.0379	0.0082	0.0572	0.0218	0.0270	0.0078

The quality of results has been assessed considering three indices: the estimated absolute bias $\widehat{\text{AB}}$, the mean squared error $\widehat{\text{MSE}}$ and the average diss index $\overline{\texttt{diss}}$. $\widehat{\text{AB}}$ and $\widehat{\text{MSE}}$ are averaged over π, ξ_D, ξ_U. These metrics are shown for all scenarios and for all sample sizes in Tables B.1, B.2, and B.3; while Table 3.6 shows the best and worst

Table 3.5: Averages and standard errors of the iter = 1000 estimated values with a sample size of 1000 observations, CUM5

		a				b				c			
		π	ξ_D	ξ_U	diss	π	ξ_D	ξ_U	diss	π	ξ_D	ξ_U	diss
Case 1	True	0.3	0.2	0.1	-	0.5	0.2	0.1	-	0.7	0.2	0.1	-
	Average	0.6229	0.3418	0.2827	0.0166	0.5035	0.2010	0.1005	0.0082	0.7033	0.2008	0.1005	0.0080
	Std error	0.1664	0.0709	0.0957	0.0115	0.0450	0.0223	0.0224	0.0062	0.0381	0.0170	0.0156	0.0061
Case 2	True	0.3	0.5	0.2	-	0.5	0.5	0.2	-	0.7	0.5	0.2	-
	Average	0.2957	0.4991	0.1864	0.0101	0.4999	0.4982	0.1946	0.0105	0.6972	0.5003	0.1965	0.0100
	Std error	0.0742	0.0490	0.0622	0.0076	0.0752	0.0277	0.0378	0.0079	0.0682	0.0199	0.0261	0.0077
Case 3	True	0.3	0.8	0.1	-	0.5	0.8	0.1	-	0.7	0.8	0.1	-
	Average	0.3004	0.8017	0.0982	0.0076	0.4982	0.8017	0.0997	0.0065	0.7003	0.8009	0.0995	0.0058
	Std error	0.0434	0.0435	0.0304	0.0056	0.0406	0.0253	0.0177	0.0052	0.0333	0.0181	0.0125	0.0044
Case 4	True	0.3	0.2	0.4	-	0.5	0.2	0.4	-	0.7	0.2	0.4	-
	Average	0.3098	0.1952	0.4011	0.0115	0.5031	0.1933	0.3947	0.0110	0.6935	0.1924	0.3952	0.0104
	Std error	0.0832	0.0713	0.0473	0.0088	0.0970	0.0530	0.0339	0.0085	0.0947	0.0398	0.0265	0.0082
Case 5	True	0.3	0.4	0.5	-	0.5	0.4	0.5	-	0.7	0.4	0.5	-
	Average	0.3294	0.3938	0.4963	0.0097	0.5153	0.3991	0.4990	0.0089	0.7023	0.4004	0.5011	0.0089
	Std error	0.0857	0.0393	0.0459	0.0076	0.0966	0.0233	0.0300	0.0068	0.1038	0.0159	0.0233	0.0065
Case 6	True	0.3	0.1	0.7	-	0.5	0.1	0.7	-	0.7	0.1	0.7	-
	Average	0.3004	0.0948	0.7020	0.0083	0.5039	0.1006	0.7001	0.0075	0.6991	0.1001	0.7009	0.0074
	Std error	0.0505	0.0376	0.0458	0.0065	0.0463	0.0217	0.0252	0.0058	0.0396	0.0156	0.0188	0.0053

Table 3.6: Summary results from the simulation study, CUM5

	Best			Worst		
	min(AB)	min(MSE)	min(diss)	max(AB)	max(MSE)	max(diss)
n = 100	0.0552	0.0056	0.0193	0.2007	0.2007	0.0394
	(Case 3c)	(Case 3c)	(Case 3c)	(Case 1a)	(Case 1a)	(Case 4a)
n = 500	0.0248	0.0011	0.0079	0.1929	0.0569	0.0181
	(Case 3c)	(Case 3c)	(Case 3c)	((Case 1a)	(Case 1a)	(Case 1a)
n = 1000	0.0006	0.0005	0.0058	0.2158	0.0665	0.0166
	(Case 3c)	(Case 3c)	(Case 3c)	(Case 1a)	(Case 1a)	(Case 1a)

results for each considered sample size. Case 1 was the worst performing, and Case 3 the best performing for all values of the sample size n. So, the specific configuration of the parameters ξ_D and ξ_U somehow affected the goodness of estimates.

3.2.2 *CUM7 simulation study*

The ternary plots of the CUM7 simulation study for the 18 scenarios are represented in Figure 3.2. The results of the CUM7 simulation study, assessed with the same metrics used for CUM5 (Tables 3.7, 3.8, and B.4), show an overall slightly better performance, mainly in terms of efficiency. Also with CUM7, Case 1a exhibits a problematic pattern, but, with respect to Figure 3.1, observations are more concentrated around the true parameter value.

3.3 CASE STUDY: THE INDIVIDUAL PERCEPTIONS OF MUSEUM VISITORS

In this section, the results of an application of the CUM model to real data are presented. The data were obtained by administering a questionnaire to the visitors of the Santa Giulia Museum in Brescia, Italy. The Santa Giulia Museum, included in the UNESCO World Heritage List, is the most important museum in Brescia and unique in

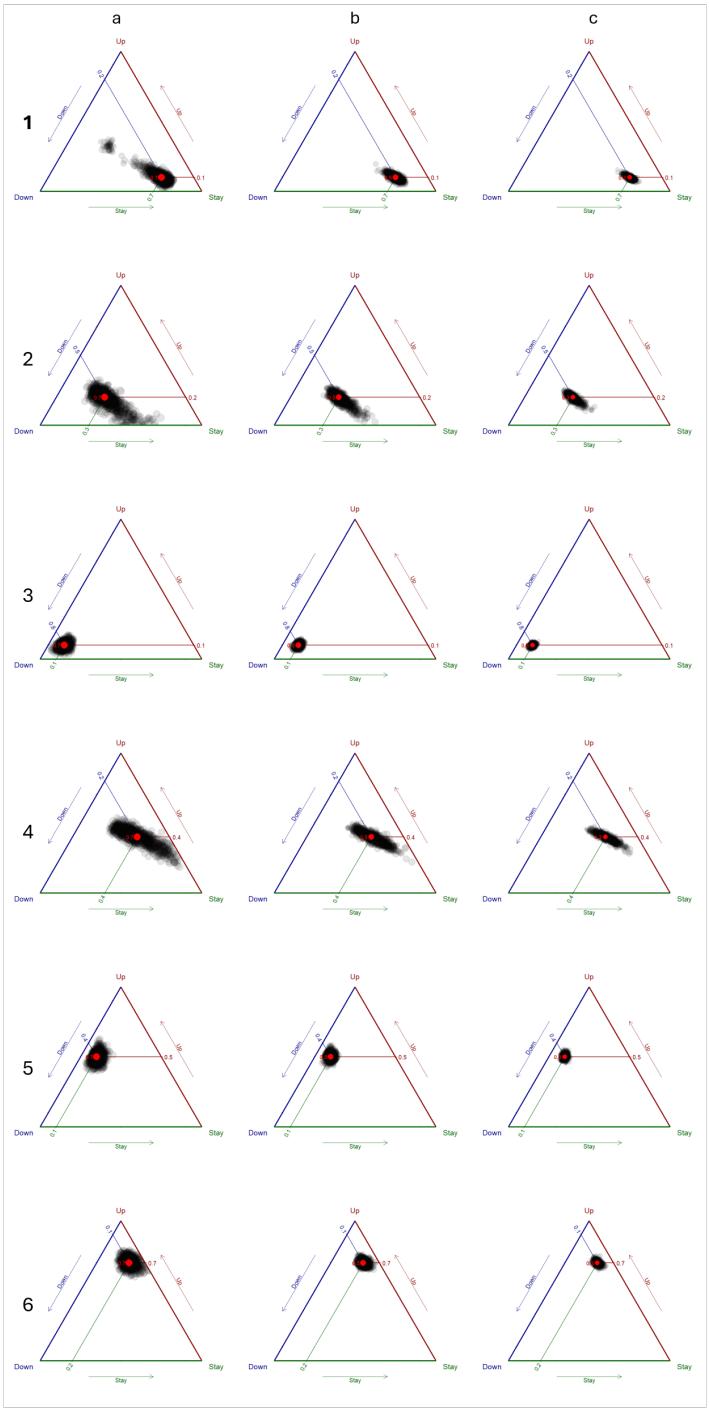


Figure 3.2: Ternary plots for CUM7 simulations ($n = 1000, iter = 1000$). Red dot: true parameter value; grey dots: estimated values.

Table 3.7: Averages and standard errors of the iter = 1000 estimated values, CUM7.

		a				b				c			
		π	ξ_D	ξ_U	diss	π	ξ_D	ξ_U	diss	π	ξ_D	ξ_U	diss
Case 1	True	0.3	0.2	0.1	-	0.5	0.2	0.1	-	0.7	0.2	0.1	-
	Average	0.3133	0.2105	0.1133	0.0199	0.5025	0.2002	0.1004	0.0185	0.7002	0.2004	0.1000	0.0174
	Std error	0.0519	0.0499	0.0521	0.0097	0.0367	0.0201	0.0188	0.0081	0.0319	0.0145	0.0137	0.0082
Case 2	True	0.3	0.5	0.2	-	0.5	0.5	0.2	-	0.7	0.5	0.2	-
	Average	0.2958	0.4918	0.1822	0.0206	0.4977	0.4966	0.1952	0.0205	0.6972	0.4982	0.1959	0.0205
	Std error	0.0601	0.0504	0.0648	0.0088	0.0556	0.0292	0.0367	0.0090	0.0505	0.0204	0.0274	0.0084
Case 3	True	0.3	0.8	0.1	-	0.5	0.8	0.1	-	0.7	0.8	0.1	-
	Average	0.3016	0.7994	0.0996	0.0172	0.5001	0.8014	0.0991	0.0158	0.6998	0.8003	0.0997	0.0148
	Std error	0.0352	0.0288	0.0230	0.0076	0.0339	0.0183	0.0147	0.0070	0.0291	0.0125	0.0098	0.0067
Case 4	True	0.3	0.2	0.4	-	0.5	0.2	0.4	-	0.7	0.2	0.4	-
	Average	0.3094	0.2016	0.4019	0.0218	0.5000	0.1955	0.3959	0.0211	0.6979	0.1965	0.3976	0.0212
	Std error	0.0685	0.0718	0.0598	0.0094	0.0659	0.0471	0.0403	0.0092	0.0625	0.0348	0.0294	0.0090
Case 5	True	0.3	0.4	0.5	-	0.5	0.4	0.5	-	0.7	0.4	0.5	-
	Average	0.3094	0.3971	0.4956	0.0213	0.5044	0.4007	0.4997	0.0215	0.6971	0.3999	0.5003	0.0211
	Std error	0.0577	0.0331	0.0349	0.0092	0.0619	0.0200	0.0230	0.0090	0.0595	0.0142	0.0155	0.0093
Case 6	True	0.3	0.1	0.7	-	0.5	0.1	0.7	-	0.7	0.1	0.7	-
	Average	0.3018	0.0969	0.6993	0.0180	0.4999	0.0983	0.7009	0.0174	0.6998	0.0987	0.7000	0.0164
	Std error	0.0412	0.0316	0.0307	0.0078	0.0368	0.0184	0.0188	0.0077	0.0313	0.0134	0.0149	0.0073

Table 3.8: Best and worst results for CUM7 simulation study

	min			max		
	$\widehat{AB}$	$\widehat{MSE}$	$\overline{diss}$	$\widehat{AB}$	$\widehat{MSE}$	$\overline{diss}$
Case	1c	3c	3c	1a	4a	4a
Value	0.0002	0.0004	0.0148	0.0124	0.0045	0.0218

Italy and in Europe due to its display concept and location. The data analyzed in this work were collected during the period April-July 2022 by a survey developed within the activities of the project "Data Science for Brescia (DS4BS) - Arts and cultural place".

The dataset contains 665 evaluations expressed by visitors about a question related to the easiness in visiting the museum. The adopted 7-point semantic differential scale ranges from "difficult" to "easy". The absolute frequency distribution of answers is displayed in Figure 3.3.

The results obtained through the application of CUM5 are compared to those derived from CUM7 and the traditional CUB approach. The case study aims to achieve two objectives: (1) to conduct a comparative assessment of the outcomes produced by the CUM and CUB methodologies, and (2) to evaluate the parameter estimates obtained using CUM5 and CUM7. Regarding the second objective, the analysis considers the original data collected on a 7-point semantic differential scale, as well as the same data after the merging of certain categories using two distinct strategies to create two different 5-point scales. The data are analyzed using CUM7 and CUM5 to verify the consistency of the results.

In order to adapt the dataset for CUM5 application, categories need to be reduced from seven to five. To this purpose, two different strategies have been implemented:

1. The first one is based on merging ratings two/three and five/six, to maintain the original ending and central categories of the scale. This dataset will be addressed as dataset-1 in the following.

2. The second one is based on merging ratings one/two/three, and leaving unchanged the other categories. This second choice is based on the analysis of fre-

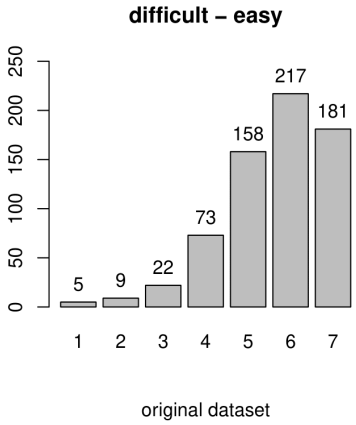


Figure 3.3: Absolute frequencies of the 7 original ratings for the question about easiness in visiting Santa Giulia museum

quencies of original ratings (see Figure 3.3), where categories in the right-hand side of the scale have an appreciably higher frequency than the other categories. This dataset will be addressed as dataset-2 in the following.

The rearranged datasets have the same number of observations as the original one, and the frequency distributions are displayed in Figure 3.4. The strategies implemented to reduce the number of categories from 7 to 5 are arbitrary, although motivated by reasons (1) related to a psychological argument, which states that respondents give significant importance to the extreme values of the scale and the middle value, and (2) suggested by data analysis. Other options could also be proposed. Alternatively, a dataset obtained from surveys with questions based on 5-point response scales could have been considered. However, in this case, it would not have been possible to fit the CUM7 model.

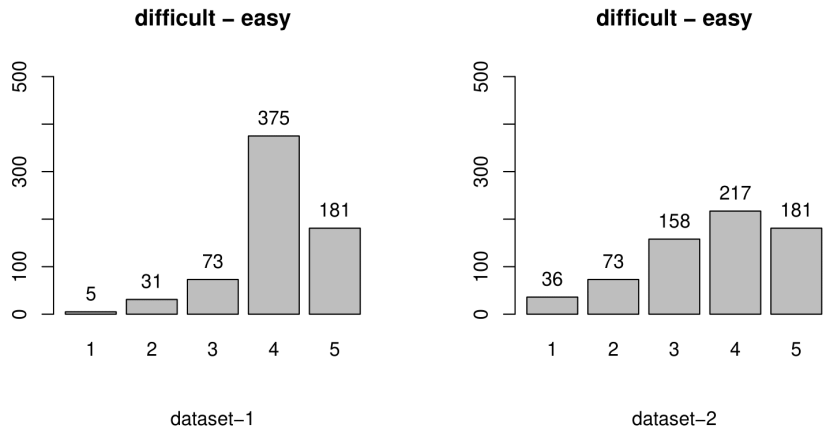


Figure 3.4: Absolute frequencies of dataset-1 and dataset-2 for the question about easiness in visiting the museum (5 ratings).

3.3.1 *Models for the dataset with 7 categories*

In this subsection, results of the application of CUB and CUM7 to the original dataset with 7 categories are described and compared. Table 3.9 reports estimated parameters; the ternary plot chart for CUM7 is displayed in Figure 3.5. Both the CUM7 and CUB models suggest a low level of uncertainty ($1 - \hat{\pi}$ is equal to 0.0509 and 0.0829, respectively). As for feeling, it is quite high for CUB model ($1 - \hat{\xi} = 0.7929$), while CUM7 recognises the presence of the different components of the assumed DP, namely the probability to move toward "easy" ($\hat{\xi}_U = 0.6527$), to move toward "difficult" ($\hat{\xi}_D = 0.0824$) and to stay still ($1 - \hat{\xi}_U - \hat{\xi}_D = 0.2649$).

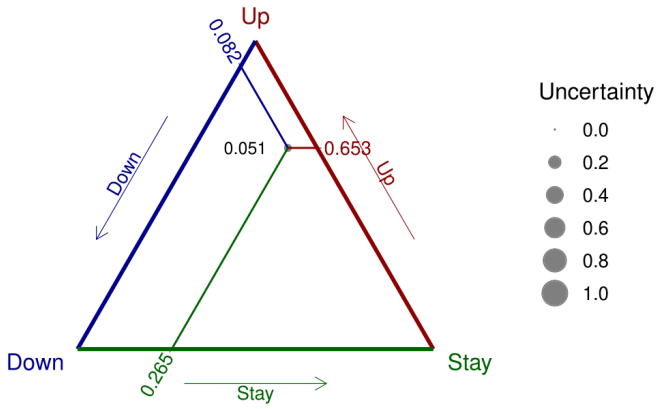
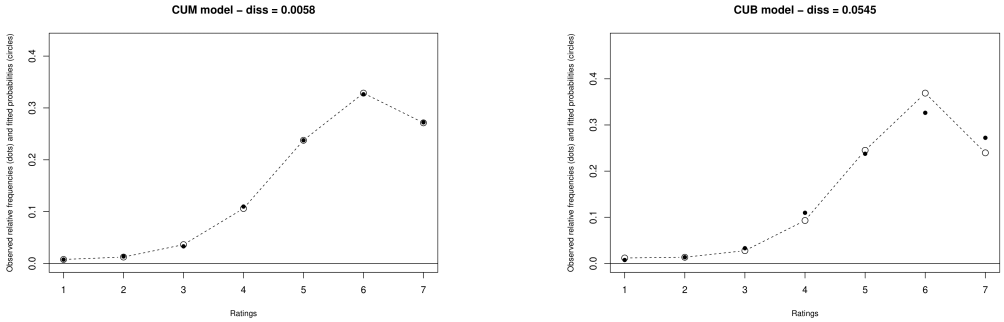


Figure 3.5: Ternary plot of CUM7 model.

Plots of observed versus fitted frequencies for CUM7 and CUB are shown in Figure 3.6: CUM7 exhibits a better fit than CUB. Diss, BIC and AIC indices for the two models (Table 3.10) are all lower for CUM7 than for CUB, suggesting that the improved goodness of fit of CUM7 justifies the additional parameter.

Table 3.9: Estimated parameters (standard errors in parenthesis) - CUM7 and CUB models fitted to the original dataset with 7 ratings

CUM7			**CUB**	
π	ξ_D	ξ_U	π	ξ
0.9491	0.0824	0.6527	0.9171	0.2071
(0.0202)	(0.0130)	(0.0140)	(0.0218)	(0.0077)



(a) Observed vs. fitted frequencies for CUM7

(b) Observed vs. fitted frequencies for CUB

Figure 3.6: Original dataset: Observed vs. fitted frequencies for CUM7 (left) and CUB (right) models

Table 3.10: Diss index, BIC and AIC - CUM7 and CUB fitted to the original dataset with 7 ratings

	diss	BIC	AIC
CUM7	0.0058	2030.230	2016.524
CUB	0.0545	2033.283	2024.284

3.3.2 *Models for the datasets with 5 categories: dataset-1*

In this subsection, CUM5 and CUB are used to fit the ratings of dataset-1. For the CUB approach, a model with shelter on the fourth category was also used, but the shelter parameter turned out to be not significant. Table 3.11 reports the estimated parameters for CUM5 and CUB models; for CUM5, the ternary plot is in Figure 3.7. Also with these data, both the CUM5 and CUB models suggest a low level of uncertainty ($1-\hat{\pi}$ is equal to 0.0151 and 0.0174, respectively). As for feeling, it is again quite high for CUB model ($1-\hat{\xi}=0.7668$), and also in this case CUM5 recognises the presence of the different components of the assumed DP, with estimated values consistent with those obtained with CUM7 fitted to the original dataset ($\hat{\xi}_U=0.5764$, $\hat{\xi}_D=0.0245$ and $1-\hat{\xi}_U-\hat{\xi}_D=0.3991$).

Plots of observed versus fitted frequencies are displayed in Figure 3.8. Both CUM5 and CUB are not able to model the large observed frequency in the fourth rating, and this reflects on high values of the diss index, which however is lower for CUM5. Also in this case, according to diss index, BIC and AIC (Table 3.12), the CUM model outperforms the others.

Table 3.11: Estimated parameters (standard errors in parenthesis) - CUM5 and CUB models fitted to dataset-1

CUM5			**CUB**	
π	ξ_D	ξ_U	π	ξ
0.9849	0.0245	0.5764	0.9826	0.2332
(0.0249)	(0.0089)	(0.0150)	(0.0145)	(0.0090)

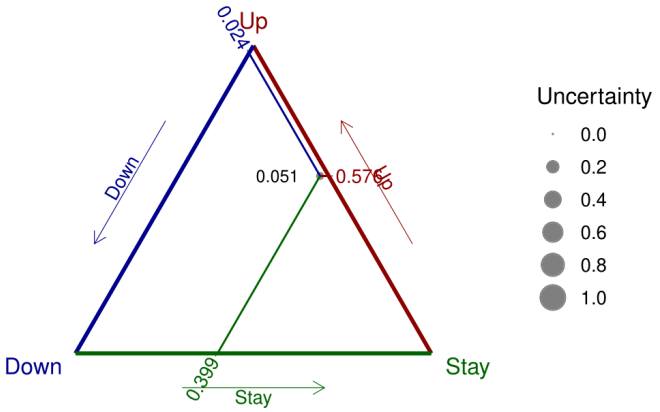
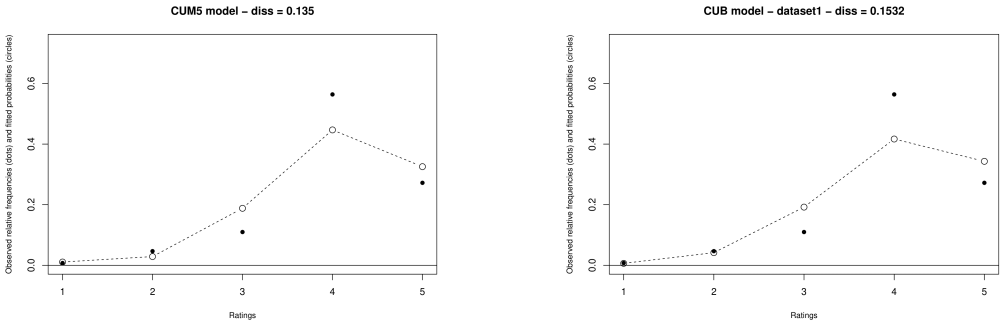


Figure 3.7: Ternary plot of CUM5 model, dataset-1.



(a) Observed vs. fitted frequencies for CUM5

(b) Observed vs. fitted frequencies for CUB

Figure 3.8: Dataset-1: Observed vs. fitted frequencies for CUM5 (left) and CUB (right) models

Table 3.12: Diss index, BIC and AIC - CUM5 and CUB models fitted to dataset-1

	diss	BIC	AIC
CUM5	0.1350	1539.24	1525.74
CUB	0.1532	1545.10	1536.10

3.3.3 *Models for the datasets with 5 categories: dataset-2*

In this subsection, CUM5 and CUB models are fitted to dataset-2. In this case the observed frequencies do not suggest the presence of any shelter effect, so only the basic CUB model is used. The parameter estimates are in Table 3.13; the ternary plot of CUM is shown in Figure 3.9. The different aggregation of categories proposed in dataset-2 generates data with higher uncertainty, as confirmed by both the CUM5 and CUB approaches ($1-\hat{\pi}$ is equal to 0.2078 and 0.2874, for CUM5 and CUB, respectively). However, no appreciable difference emerges for feeling, whose parameters have quite similar estimates to those obtained with dataset-1, both for CUB model ($1-\hat{\xi}=0.7248$), and for CUM5, except for a slightly higher probability of moving toward "difficult" ($\hat{\xi}_U = 0.5393$, $\hat{\xi}_D = 0.1276$ and $1-\hat{\xi}_U-\hat{\xi}_D = 0.3331$). The higher value of $\hat{\xi}_D$ can be justified by the fact that the aggregation rule generating dataset-2 (merging of the original categories 1-2-3) moves the middle position of the scale upward. Since the CUM DP assumes that the respondent's reasoning begins from the middle position, if this position is moved upward, a higher probability of moving toward "difficult" can reasonably be expected. So, the different aggregation of categories seems to have modified only the uncertainty assessment, but the feeling measurement remains consistent with that obtained with the other datasets, denoting a very robust assessment of this component.

Figure 3.10 displays the plots of observed versus fitted frequencies, suggesting a good fit for both the models. According to diss, BIC and AIC indices (Table 3.14), the CUM model outperforms the other with respect to diss and AIC, while in this case the lowest BIC is reached by CUB.

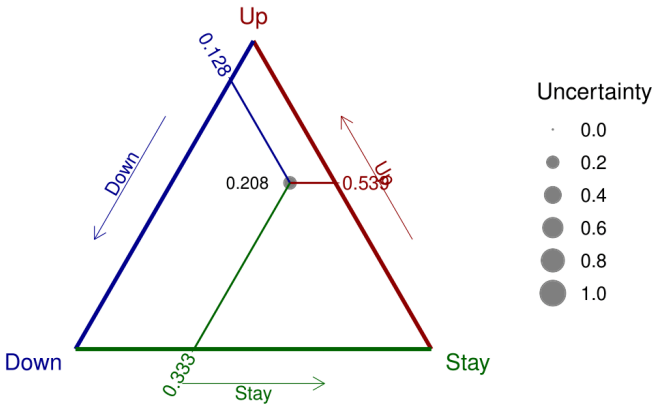
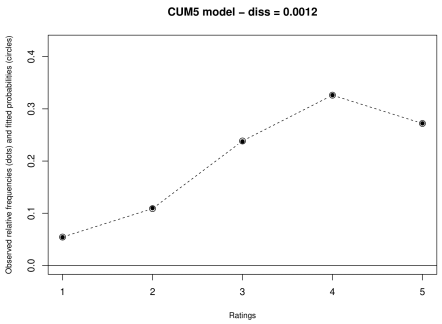
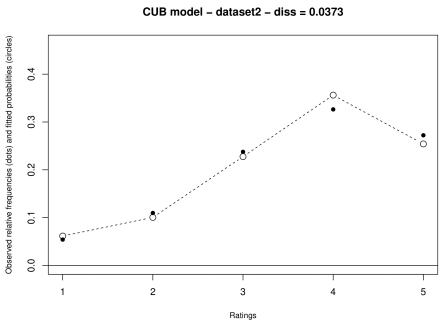


Figure 3.9: Ternary plot of CUM5 model fitted to dataset-2.



(a) Observed vs. fitted frequencies for CUM5



(b) Observed vs. fitted frequencies for CUB

Figure 3.10: Dataset-2: Observed vs. fitted frequencies for CUM5 (left) and CUB (right) models

Table 3.13: Estimated parameters (standard errors in parenthesis) - CUM5 and CUB models fitted to dataset-2

CUM5			**CUB**	
π	ξ_D	ξ_U	π	ξ
0.7922	0.1276	0.5393	0.7126	0.2752
(0.0599)	(0.0260)	(0.0201)	(0.0441)	(0.0142)

Table 3.14: Diss index, BIC and AIC - CUM5 and CUB models fitted to dataset-2

	diss	BIC	AIC
CUM5	0.0012	1963.29	1949.79
CUB	0.0373	1960.79	1951.79

In this work the CUM model is extended to the case of 5 categories.

The results from a simulation considering the application of CUM5 to 18 scenarios based on different parameter values are reported and compared to the results from an equivalent CUM7 simulation, showing how fitting measures are similar in the two cases, with a single case that requires further investigation.

A case study concerned with the the evaluation of the visitor experience at the Santa Giulia Museum is proposed. The original dataset with 7 ratings was analysed via CUM7 and CUB. Then, the original dataset has been transformed into a 5-point scale, following two different strategies, and analysed by means of CUM5 and CUB. The idea was to compare the CUM to the CUB approach and also to check the sensitivity of CUM5 estimation on collapsing rules by considering different aggregations of the original dataset with 7 categories into two datasets with 5 categories.

From the point of view of the comparison of the different approaches, in general the CUM model outperforms CUB, apart from a single case where CUB exhibits a lower BIC value. So, the additional parameter of CUM with respect to CUB seems to provide significant information and goodness-of-fit improvement.

As for parameter estimates, in the examined case study, the different aggregation rule of the response categories impacts on the uncertainty measurement, while the assessment of the feeling component remains stable, except for a small, easily interpretable, difference. Both the CUM and CUB models exhibit generally low sensitivity to various collapsing rules, with the CUM model showing slightly greater sensitivity compared to the CUB model. Future research could explore this further through simulation studies, focusing on the sensitivity of the estimated parameters of the CUM model under different collapsing rules.

In a single simulation case estimates far from actual parameters values have been obtained. This raises concerns about the identifiability of the CUM5 model over the entire parameter space which is proved by th e bimodality of the parameters estimates that are reported in Figure B.3 and will be further investigated in the following chapter, where the investigation of the equivalence of CUB and CUM model has been a starting point for proving and having a better knowledge about the identifiability of CUM5.

Additionally, to avoid any problem, it may be better to estimate the model by first verifying the uniqueness of estimates using different starting values for the EM algorithm, as done in this work.

4

EXPLORING THE EQUIVALENCE BETWEEN CUB AND CUM MODELS

CONTRIBUTIONS RELATED TO THIS CHAPTER:

Indexed Journals (WoS, Scopus)

[IJ2] Ventura M., Macis A., Manisera M., and Zuccolotto P., *On the Equivalence of two Mixture Models for Rating Data*, Advances in Statistical Analysis, *In press*

International Conferences

[IC1] Macis A., Ventura M., Manisera M., Zuccolotto P., *Modeling Rating Data: Exploring the Relationship between CUB and CUM Models*, 29th Nordic Conference in Mathematical Statistics (NORDSTAT), Gothenburg (Sweden), 19-22 June 2023.

[IC2] Macis A, Ventura M., Manisera M. Zuccolotto P., *Exploring the Equivalence of two Mixture Models for Rating Data in the CUB class*, Joint Conferences Data Science, Statistics and Visualisation (DSSV) and European Conference on Data Analysis (ECDA), Antwerp (Belgium), 5-7 July 2023.

Many rating scales contain a middle option that represents a neutral or indifferent response between two extremes. For example, when respondents are asked to rate their satisfaction degree with a given item, they can be asked to use a response scale with 7 ratings, from 1='very dissatisfied' to 7='very satisfied', with middle category 4='neither satisfied nor dissatisfied'. In such cases, the same reasoning applied to multi-point Semantic Differential scales can also be extended to these response scales. Conceptually, there is little distinction in the decision-making approach when defining an assessment in multi-point Semantic Differential scales and Likert-type response scales. Thus, both the CUB and CUM models can be utilised to analyse rating data from these scales, resulting in possibly different interpretations.

The work presented in this chapter aims at examining the possibility of the existence of a CUB able to reproduce a CUM model, and vice versa, given their shared framework. Specifically, the possible existence of CUB and CUM models yielding the same theoretical probabilities is investigated. In mathematical terms, this means studying the equivalence between the CUB and CUM models to gain deeper knowledge of these models from an interpretative perspective. Furthermore, the objective is to potentially establish a foundation for investigating the identifiability of CUM models.

To achieve this goal, a preliminary definition of model equivalence for discrete probability distributions is proposed, with a distinction between local and global, unidirectional and bidirectional properties.

Let $\ddot{P}(X \mid \ddot{\theta})$ and $\tilde{P}(X \mid \tilde{\theta})$ be two discrete probability distributions with the same support $\mathcal{S}_X$ and parameter vectors $\ddot{\theta} = (\ddot{\theta}_1, \ldots, \ddot{\theta}_p)' \in \ddot{\Theta}$ and $\tilde{\theta} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_q)' \in \tilde{\Theta}$, respectively. The statistical models described by $\ddot{P}$ and $\tilde{P}$ are assumed to be identifiable, *i.e.* different values of the parameters must generate different probability distributions.

Equivalence between models has been already investigated in the literature (Bork et al., 2021; McCullagh, 2002; Robitzsch, 2021). For instance, talking about Structural Equation Models, two models can be considered *observationally equivalent* if a model can generate the same probability distribution generated by the other one (Paulino and Bragança Pereira, 1994). Moreover, a distinction between *local* and *global equivalence* can be done: two models are said to be *globally equivalent* if there exists a function that links every parameter of a model into the parameters of another model. Conversely, if the function maps only from one subset of the parameters of a model into a subset of the parameters of another model, it is defined as *local equivalence* (Hancock and Mueller, 2013).

This work is based on a specific definition that distinguishes between local and global properties along with two main concepts of equivalence: unidirectional and bidirectional.

Definition: Local Unidirectional Equivalence
$\ddot{P}$ is unidirectionally equivalent to $\tilde{P}$ in $\ddot{\theta}$, if there exists $\tilde{\theta}$ such that

$$\ddot{P}(X = x \mid \ddot{\theta}) = \tilde{P}(X = x \mid \tilde{\theta}). \tag{4.1}$$

for all $x \in \mathcal{S}_X$. $\diamondsuit$

When local equivalence holds across the entire parameter space, global unidirectional equivalence is obtained, hereafter referred to simply as unidirectional equivalence, and is defined as follows.

Definition: Unidirectional Equivalence
$\ddot{P}$ is unidirectionally equivalent to $\tilde{P}$ if local unidirectional equivalence holds for all $\ddot{\theta} \in \ddot{\Theta}$. $\diamondsuit$

Unidirectional equivalence can be formalized through a function f, which maps from a parameter space $\ddot{\Theta}$ to another parameter space $\tilde{\Theta}$:

$$f\colon\ \ddot{\Theta} \to \tilde{\Theta},$$

such that

$$\ddot{P}(X = x \mid \ddot{\theta}) = \tilde{P}(X = x \mid f(\ddot{\theta})) \quad \forall x \in \mathcal{S}_X. \tag{4.2}$$

From the assumed identifiability property of the statistical models described by $\ddot{P}$ and $\tilde{P}$, the function f is an injective function, i.e., different elements in $\ddot{\Theta}$ are mapped to different elements in $\tilde{\Theta}$. In other words, for each pair $\ddot{\theta}_1, \ddot{\theta}_2 \in \ddot{\Theta}$, $\ddot{\theta}_1 \neq \ddot{\theta}_2$, we have $f(\ddot{\theta}_1) \neq f(\ddot{\theta}_2)$.

On the contrary, surjectivity is not ensured. When f is also surjective, it is bijective, thus invertible, and this implies that unidirectional equivalence also holds for $\tilde{P}$ to $\ddot{P}$. In this case, we have:

$$f^{-1}: \tilde{\Theta} \to \ddot{\Theta},$$

such that

$$\ddot{P}(X = x \,|\, f^{-1}(\tilde{\theta})) = \tilde{P}(X = x \mid \tilde{\theta}) \quad \forall x \in \mathcal{S}_X. \tag{4.3}$$

Definition: Bidirectional Equivalence
$\ddot{P}$ and $\tilde{P}$ are bidirectionally equivalent if $\ddot{P}$ is unidirectionally equivalent to $\tilde{P}$ and $\tilde{P}$ is unidirectionally equivalent to $\ddot{P}$. $\diamondsuit$

Bidirectional equivalence, referred to hereafter simply as equivalence, can be formalized through the same function f defined by (4.2), under the assumption that f is invertible, so as to ensure (4.3).

4.2 Equivalence of CUB and CUM models

The standard CUB model was introduced for analysing rating data measured on Likert response scales, where the respondents rate their perceptions by positioning themselves among a set of ordered categories. The model can be viewed as the result of a DP underlying the final rating of the respondents that starts from the bottom of the scale and then moves upward, based on the different sensations coming to their minds. Conversely, the CUM model assumes that the DP of the respondents starts from the middle of the multi-point Semantic Differential response scale. However, since it is common for Likert response scales to have a neutral option between two opposite extremes, conceptually, both the two decision-making approaches and the two corresponding models (CUB and CUM) can be used for expressing the probability of the responses. Thus, it becomes interesting to analytically investigate the equivalence of the two models, by solving the following equation:

$$P_{CUB}(R = r \mid \theta_{CUB}) = P_{CUM}(R = r \mid \theta_{CUM}) \quad \forall\, r = 1, 2, \ldots, m, \tag{4.4}$$

that is, questioning if there exist two parameter vectors $\theta_{CUB} \in \Theta_{CUB}$ and $\theta_{CUM} \in \Theta_{CUM}$ ensuring that the two models CUB and CUM yield the same expected frequencies of the responses.

In other words, the equivalence of CUB and CUM models was investigated by following the paradigm described in Section 4.1, with $\ddot{P}$ and $\tilde{P}$ being the probability mass functions of the CUB and CUM model, respectively, and $\mathcal{S}_X = \{1, 2, \cdots, m\}$. The formalization of the function f defining unidirectional equivalence as in (4.2) is then

$$f: \Theta_{CUB} \to \Theta_{CUM},$$

such that

$$P_{CUB}(R = r \mid \theta_{CUB}) = P_{CUM}(R = r \mid f(\theta_{CUB})) \quad \forall\, r = 1, 2, \ldots, m. \tag{4.5}$$

Some propositions are hereafter analytically proved to show if a function f obeying definition (4.5) exists, and its properties.

4.2.1 *CUB and CUM models with* $m = 5$

In this section, the analysis is limited to the case $m = 5$. While the probability mass function for the CUB model can be easily obtained for any m as shown in (2.3), the probability mass function for the CUM model must be explicitly calculated for a given $m = 2k+1$ starting from equations (2.17) and (2.18). When $m = 5$, the probability mass function for the variable W is the one derived in Section 3.1.

To explore unidirectional equivalence of CUB to CUM, the probability mass functions of the two models were equated, as in equation (4.4). Being π_B and π_M the parameter π of the CUB and the CUM model, respectively, the case $\pi_B = \pi_M = \pi$ has been examined firstly. Under this assumption, (4.4) can be simply written as

$$\binom{5-1}{r-1}\xi^{5-r}(1-\xi)^{r-1} = P_W(r \mid \xi_D, \xi_U), \quad \forall\, r = 1, 2, \ldots, 5 \tag{4.6}$$

from which the following system has been obtained:

$$\begin{cases} r = 1: & \xi^4 = \xi_D^2 \\ r = 2: & 4\xi^3(1-\xi) = 2\xi_D(1-\xi_D-\xi_U) \\ r = 3: & 6\xi^2(1-\xi)^2 = (1-\xi_D-\xi_U)^2 + 2\xi_D\xi_U \\ r = 4: & 4\xi(1-\xi)^3 = 2\xi_U(1-\xi_D-\xi_U) \\ r = 5: & (1-\xi)^4 = \xi_U^2 \end{cases} \tag{4.7}$$

that admits one and only one solution, $\xi_D = \xi^2$ and $\xi_U = (1-\xi)^2$, thus proving the following Proposition 1, with (4.4) being verified if $\theta_{CUB} = (\xi, \pi)'$ and $\theta_{CUM} = f(\theta_{CUB}) = (\xi^2, (1-\xi)^2, \pi)'$.

Proposition 1 *Given* $m = 5$, *for every* $\theta_{CUB} = (\xi, \pi)'$ *the parameter vector* $\theta_{CUM} = (\xi^2, (1-\xi)^2, \pi)$ *verifies (4.4).*

When $\pi_B \neq \pi_M$, equation (4.4) generates the system:

$$\begin{cases} r = 1: & \pi_B\xi^4 + U_B = \pi_M\xi_D^2 + U_M \\ r = 2: & 4\pi_B\xi^3(1-\xi) + U_B = 2\pi_M\xi_D(1-\xi_D-\xi_U) + U_M \\ r = 3: & 6\pi_B\xi^2(1-\xi)^2 + U_B = \pi_M[(1-\xi_D-\xi_U)^2 + 2\xi_D\xi_U] + U_M \\ r = 4: & 4\pi_B\xi(1-\xi)^3 + U_B = 2\pi_M\xi_U(1-\xi_D-\xi_U) + U_M \\ r = 5: & \pi_B(1-\xi)^4 + U_B = \pi_M\xi_U^2 + U_M \end{cases} \tag{4.8}$$

where $U_B = (1-\pi_B)P_U = \dfrac{(1-\pi_B)}{5}$, and $U_M = (1-\pi_M)P_U = \dfrac{(1-\pi_M)}{5}$.

The parameters ξ_D and ξ_U can be obtained as a function of ξ, π_B, and π_M by solving the first and the last equation of the system (4.8)

$$\begin{aligned} \xi_D &= \sqrt{\frac{\pi_B}{\pi_M}\xi^4 + \frac{\pi_M - \pi_B}{5\pi_M}} = \sqrt{\delta(0.5-x)^4 + \beta} \\ \xi_U &= \sqrt{\frac{\pi_B}{\pi_M}(1-\xi)^4 + \frac{\pi_M - \pi_B}{5\pi_M}} = \sqrt{\delta(0.5+x)^4 + \beta} \end{aligned} \tag{4.9}$$

where $\delta = \pi_B/\pi_M$, $\beta = (1-\delta)/5$, $x = 0.5 - \xi$, and then verifying if conditions (4.9) are also solutions of the other three equations for some ξ, π_B, π_M. For illustrative purposes, an example is shown in Figure 4.1, where the probability mass functions of CUB (solid line) and CUM (dashed line) for $r = 2, 3, 4$ are plotted against ξ. The values of ξ_D and ξ_U are set as in (4.9), so that (4.4) is always verified for $r = 1, 5$. The values of ξ for which the solid lines intersect the dashed ones are solutions of the single equations in system (4.8). The aim is to identify one or more values of ξ that ensure common solutions to the three equations. The example in Figure 4.1 is obtained for $\pi_B = 0.1$ and $\delta = 0.7$: in this case there is no value of ξ that satisfies the required condition.

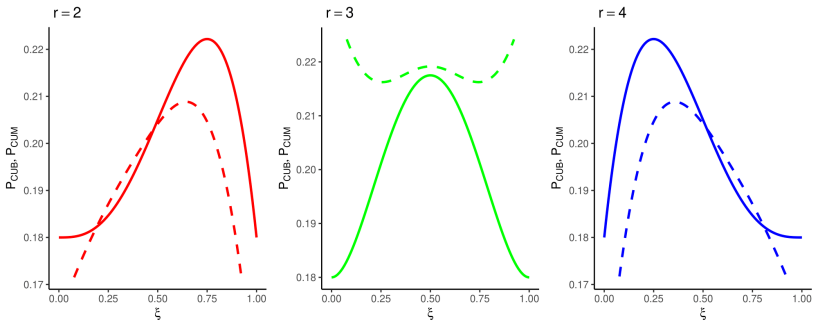


Figure 4.1: Probability mass functions of CUB (solid lines) and CUM5 (dashed lines), as functions of ξ, with ξ_D and ξ_U as in (4.9), $\pi_B = 0.1$, $\delta = 0.7$.

Before exploring the behaviour of the three pairs of curves, the prerequisites for the existence of the expressions (4.9) are established; that is, the conditions that ensure that the expressions under the square root are greater than or equal to 0 and that ξ_D and ξ_U take acceptable values ($\xi_D, \xi_U \in [0, 1]$ and $\xi_D + \xi_U \leqslant 1$). The analysis can be separated into two cases:

1. $\pi_B < \pi_M$ $(0 < \delta < 1)$
   In this case, the expressions under the square root are always greater than 0, so it is only necessary to ensure that ξ_D and ξ_U take acceptable values. To do that, then, the following function has been studied:

$$y = \xi_D + \xi_U = \sqrt{\delta(0.5 - x)^4 + \beta} + \sqrt{\delta(0.5 + x)^4 + \beta} \tag{4.10}$$

   which turns out to be a parabola with the vertex at $x = 0$ (*i.e* at $\xi = 0.5$), as explained in Appendix C. To determine the conditions that guarantee $\xi_D, \xi_U \in [0, 1]$ and $\xi_D + \xi_U \leqslant 1$, it is sufficient to identify the values of ξ, π_B, and π_M that ensure the parabola lies below the line $y = 1$ within the assumed support ($x \in [-0.5, 0.5]$). Through algebraic manipulation, the following conditions obtained:

   a) if $0 < \delta \leqslant 1/5$, then ξ_D and ξ_U are defined for all ξ (with $\xi_D + \xi_U = 1$ if $\delta = 1/5$ and $\xi = 0$ or $\xi = 1$);

   b) if $1/5 < \delta < 1$, then ξ_D and ξ_U are defined for $\xi \in I(0.5)$, where $I(0.5)$ is an interval centered at 0.5 with an amplitude dependent on ξ and δ.

   Examples of cases (a) and (b) are shown in Figure 4.2, with $\delta = 0.15$ (top) and $\delta = 0.5$ (bottom) respectively ($\pi_B = 0.1$).

2. $\pi_B > \pi_M$ ($\delta > 1$) this case, it is first necessary to ensure that the expressions under the square root are greater than or equal to 0, which requires the following conditions:

$$\begin{aligned} \xi &\geqslant \sqrt[4]{\frac{\pi_B - \pi_M}{5\pi_B}} = \sqrt[4]{\frac{\delta - 1}{5\delta}} = s_1 \\ \xi &\leqslant 1 - \sqrt[4]{\frac{\pi_B - \pi_M}{5\pi_B}} = 1 - \sqrt[4]{\frac{\delta - 1}{5\delta}} = s_2 \end{aligned} \tag{4.11}$$

that is $s_1 \leqslant \xi \leqslant s_2$. It is easy to show that $s_1 \leqslant s_2$ for $\delta \leqslant 16/11$. In addition, $\xi_D, \xi_U \in [0, 1]$ and $\xi_D + \xi_U \leqslant 1$ is also guaranteed, because it can be shown that the function (4.10) completely lies under the line $y = 1$ in the assumed support if $\delta > 1$. So, we obtain

   (c) if $1 < \delta \leqslant 16/11$, then ξ_D and ξ_U are defined for $\xi \in [s_1, s_2]$;

   (d) if $\delta > 16/11$, then there does not exist any value of ξ ensuring the existence of ξ_D and ξ_U.

   One example of case (c) is shown in Figure 4.3, with $\delta = 1.15$, so that $s_1 \simeq 0.37$ and $s_2 \simeq 0.63$ ($\pi_B = 0.5$).

In view of the difficulty of determining a closed-form solution for system (4.8), a naive approach based on a graphical analysis can be adopted. So, the first aim is to ascertain the existence of values of ξ, π_B and π_M that guarantee the joint satisfaction of equations 2-4 of the system (4.8), with ξ_D and ξ_U expressed as in (4.9). To do that, equations 2 and 4 of (4.8) are considered and it can be noticed that, for a given π_B, $P_B(R = 2)$ and $P_B(R = 4)$, considered as function of ξ, are reflections of each other over the axis $\xi = 0.5$. In other words, when expressions (4.9) are valid, for the functions

$$\begin{aligned} B_1(\xi \mid r = 2, \pi_B) &= P_B(R = 2) = 4\pi_B \xi^3 (1 - \xi) + U_B \\ B_2(\xi \mid r = 4, \pi_B) &= P_B(R = 4) = 4\pi_B \xi (1 - \xi)^3 + U_B \end{aligned}$$

we have $B_1(0.5 - \xi) = B_2(\xi - 0.5)$.

The same holds, for a given π_M, for $P_M(R = 2)$ and $P_M(R = 4)$, considered as function of ξ we have:

$$\begin{aligned} M_1(\xi \mid r = 2, \pi_M) &= P_M(R = 2) = 2\pi_M \xi_D (1 - \xi_D - \xi_U) + U_M \\ M_2(\xi \mid r = 4, \pi_M) &= P_M(R = 4) = 2\pi_M \xi_U (1 - \xi_D - \xi_U) + U_M \end{aligned}$$

with $M_1(0.5 - \xi) = M_2(\xi - 0.5)$.

The values of ξ for which $B_1(\xi) = M_1(\xi)$ are then specular (with respect to $\xi = 0.5$) to the values of ξ for which $B_2(\xi) = M_2(\xi)$. In other words, for given π_B and π_M, if $B_1(0.5 - x) = M_1(0.5 - x)$, then it is also $B_2(0.5 + x) = M_2(0.5 + x)$. As a consequence, in order to have common solutions to equations 2 and 4 of (4.8), we must have:

$$B_1(0.5 - x) = M_1(0.5 - x) \quad \text{and} \quad B_1(0.5 + x) = M_1(0.5 + x), \tag{4.12}$$

which implies

$$B_2(0.5 + x) = M_2(0.5 + x) \quad \text{and} \quad B_2(0.5 - x) = M_2(0.5 - x). \tag{4.13}$$

So, roughly speaking, the common solutions to equations 2 and 4 of (4.8) can only be either an even number of solutions symmetrical with respect to $\xi = 0.5$, or the single

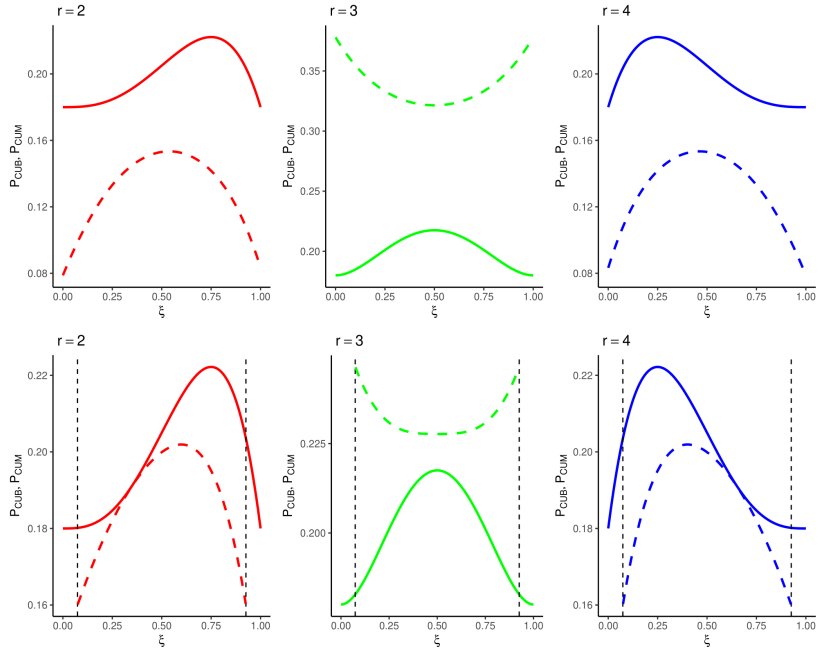


Figure 4.2: Probability mass functions of CUB (solid lines) and CUM (dashed lines), as functions of ξ, with ξ_D and ξ_U as in (4.9). Top: example of case (a), with $\pi_B = 0.1$ and $\delta = 0.15$. Bottom: example of case (b), with $\pi_B = 0.1$ and $\delta = 0.5$ (black vertical dashed lines denote the extremes of the interval $I(0.5)$ where ξ_D and ξ_U are defined).

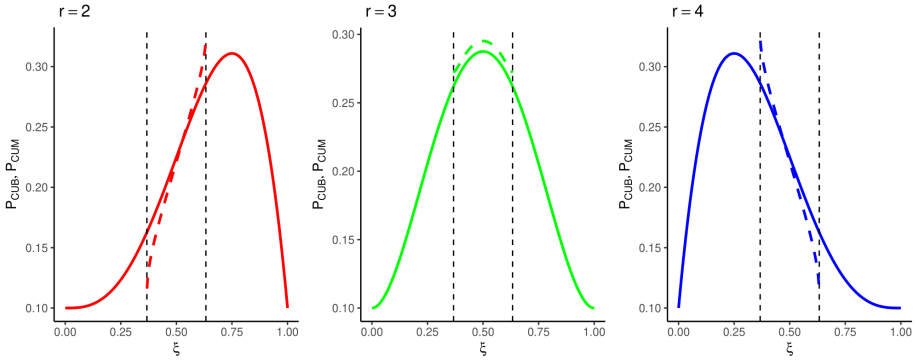


Figure 4.3: Probability mass functions of CUB (solid lines) and CUM (dashed lines), as functions of ξ, with ξ_D and ξ_U as in (4.9). Example of case (c), with $\pi_B = 0.5$ and $\delta = 1.15$ (black vertical dashed lines denote the extremes of the interval $[s_1, s_2]$ where ξ_D and ξ_U are defined).

solution $\xi = 0.5$. The behavior of equation 2 over a fine grid of values $(\pi_B; \pi_M)$ has been explored. Examples of the graphics obtained for $\pi_M = 0.63$ and π_B ranging from 0.28 to 0.9 are in Figures 4.4 and 4.5. A total of 2500 graphs have been examined. From these graphs it can be conjectured that there are at most two solutions, and they are never symmetric around $\xi = 0.5$. So, the only possible common solution with equation 4 is $\xi = 0.5$.

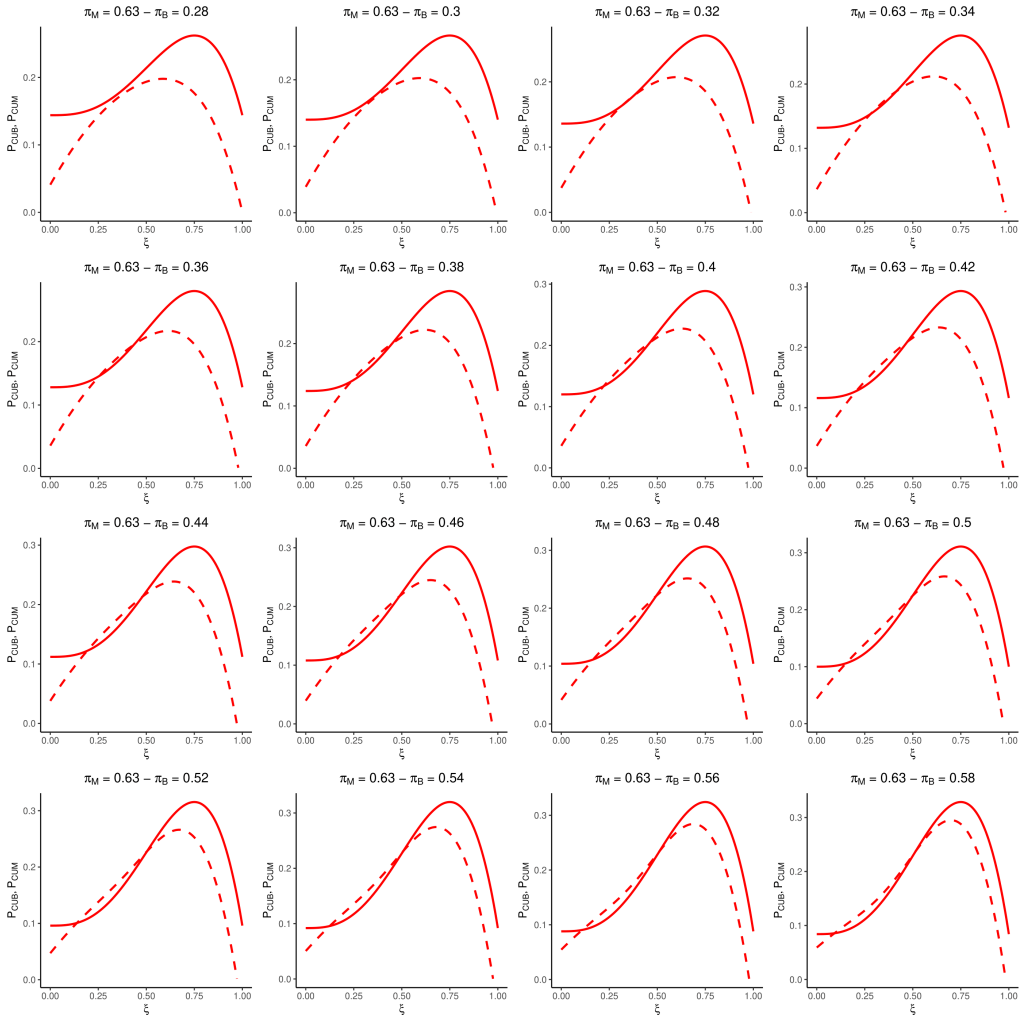


Figure 4.4: Probability mass functions of CUB (solid lines) and CUM (dashed lines), as functions of ξ, for $r = 2$, with ξ_D and ξ_U as in (4.9) and for $\pi_M = 0.63$ and π_B ranging from 0.28 to 0.58.

With $\xi = 0.5$, according to (4.9), we have $\xi_D = \xi_U$, so that equations 2 and 4 are identical, given by

$$4 \cdot 0.5^4 \pi_B + U_B = 2\pi_M \xi_D (1 - 2\xi_D) + U_M$$

that, with some simple algebra, is transformed into:

$$4 \cdot 0.5^4 \delta + \beta = 2\sqrt{0.5^4 \delta + \beta} - 4(0.5^4 \delta + \beta). \tag{4.14}$$

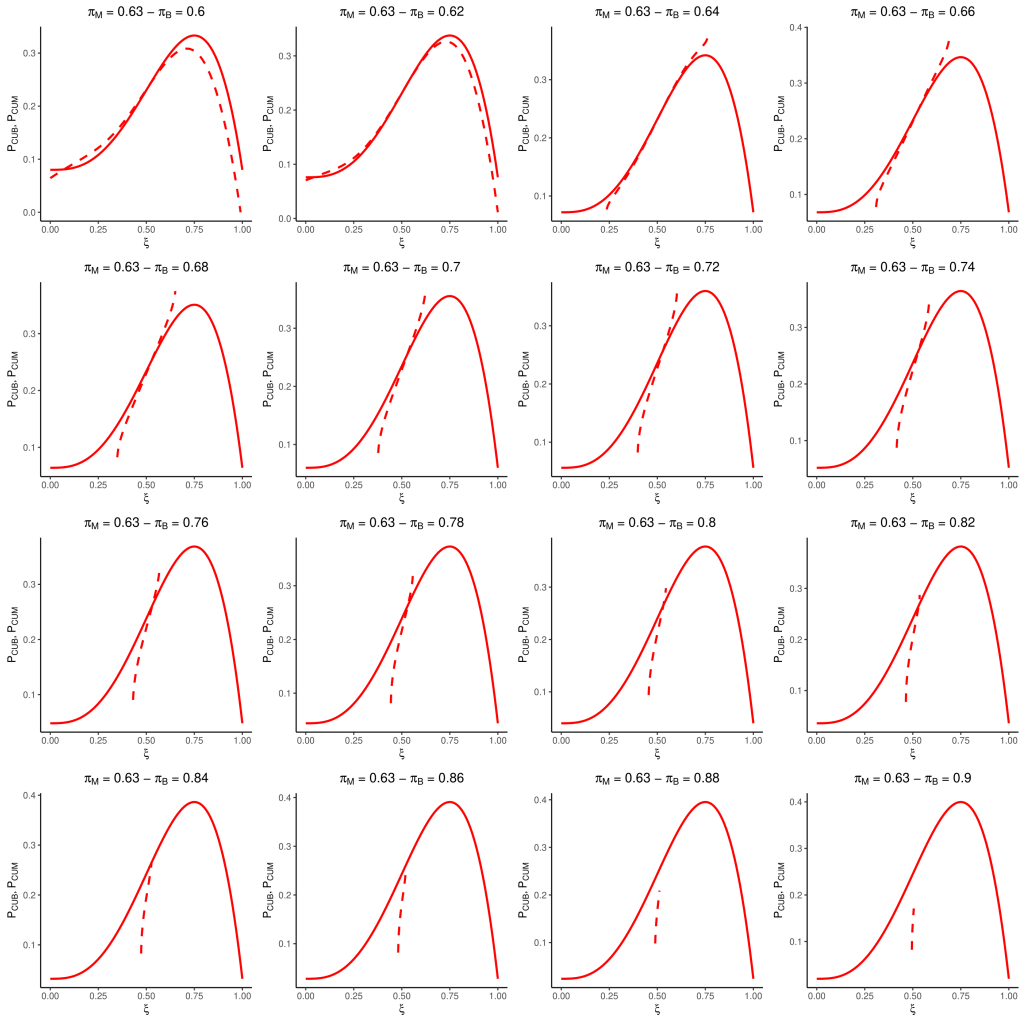


Figure 4.5: Probability mass functions of CUB (solid lines) and CUM (dashed lines), as functions of ξ, for $r = 2$, with ξ_D and ξ_U as in (4.9) and for $\pi_M = 0.63$ and π_B ranging from 0.5 to 0.9.

Equation (4.14) is solved for $\delta = 1$ (which leads back to the case $\pi_B = \pi_M$, described in Proposition 1) and $\delta = 4/5$, which implies $\xi_D = \xi_U = 0.3$ for $\xi = 0.5$. It is easy to show that this configuration is also solution of equation 3 in the system (4.8), as visualized in Figure 4.6 and explained in Appendix C.2.

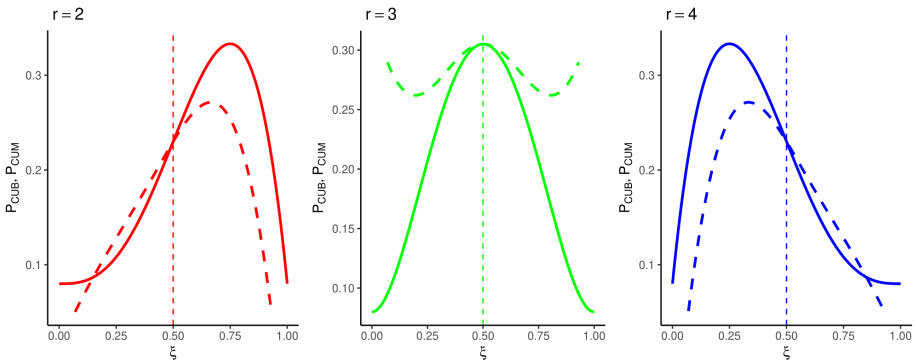


Figure 4.6: Probability mass functions of CUB (solid lines) and CUM (dashed lines), as functions of ξ, with ξ_D and ξ_U as in (4.9), $\pi_M = 0.75$, $\pi_B = \delta\pi_M$, where $\delta = 4/5$: the value $\xi = 0.5$ is a common solution to the three equations.

Now Proposition 2 can be formalized.

Proposition 2 *Given* $m = 5$, *for every* $\theta_{CUB} = (\xi, \pi)'$, $\theta_{CUB} \neq (0.5, \pi \leqslant 4/5)'$, *the relation (4.4) is satisfied only by the parameter vector* θ_{CUM} *defined in Proposition 1. For* $\xi = 0.5$ *and* $\pi \leqslant 4/5$, *the relation (4.4) is satisfied both by the parameter vector* θ_{CUM} *defined in Proposition 1 and by the parameter vector* $\theta_{CUM} = (0.3, 0.3, 5\pi/4)'$.

From Propositions 1 and 2 it follows that, in the case $m = 5$, CUB is unidirectionally equivalent to CUM, provided that its parameter space is restricted to

$$\bar{\Theta}_{CUB} = \Theta_{CUB} \backslash \bar{\delta}$$

where $\bar{\delta} = \{\theta_{CUB} : \theta_{CUB} = (0.5, \pi_B)', \pi_B \leqslant 4/5\}$. The formalization of the function f defining unidirectional equivalence as in (4.2) is then:

$$f : \bar{\Theta}_{CUB} \to \Theta_{CUM},$$

with $f(\theta_{CUB}) = f((\xi, \pi_B)') = (\xi^2, (1-\xi)^2, \pi_B)'$, for all $\theta_{CUB} \in \bar{\Theta}_{CUB}$.

4.2.2 *CUB and CUM models with* $m = 7$

In the case $m = 7$, (4.4) leads to the following system, corresponding to system (4.8) in the case $m = 5$:

$$\begin{cases} r=1: & \pi_B\xi^6+u_B=\pi_M\xi_D^3+u_M \\ r=2: & 6\pi_B\xi^5(1-\xi)+u_B=3\pi_M\xi_D^2(1-\xi_D-\xi_U)+u_M \\ r=3: & 15\pi_B\xi^4(1-\xi)^2+u_B=3\pi_M\xi_D[(1-\xi_D-\xi_U)^2+\xi_D\xi_U]+u_M \\ r=4: & 20\pi_B\xi^3(1-\xi)^3+u_B=\pi_M(1-\xi_D-\xi_U)[(1-\xi_D-\xi_U)^2+6\xi_D\xi_U]+u_M \\ r=5: & 15\pi_B\xi^2(1-\xi)^4+u_B=3\pi_M\xi_U[(1-\xi_D-\xi_U)^2+\xi_D\xi_U]+u_M \\ r=6: & 6\pi_B\xi(1-\xi)^5+u_B=3\pi_M\xi_U^2(1-\xi_D-\xi_U)+u_M \\ r=7: & \pi_B(1-\xi)^6+u_B=\pi_M\xi_U^3+u_M \end{cases} \tag{4.15}$$

The same steps illustrated for the case $m=5$ have been followed, resulting in the outcomes summarised in Propositions 3 and 4 below.

Proposition 3 *Given* $m=7$, *for every* $\theta_{CUB}=(\xi,\pi)'$ *the parameter vector,* $\theta_{CUM}=(\xi^2,(1-\xi)^2,\pi)$ *verifies (4.4).*

The values of ξ_D and ξ_U that are the solution of the first and last equation of the system (4.15) are

$$\begin{aligned} \xi_D &= \sqrt[3]{\frac{\pi_B}{\pi_M}\xi^6+\frac{\pi_M-\pi_B}{7\pi_M}}=\sqrt[3]{\delta(0.5-x)^6+\beta} \\ \xi_U &= \sqrt[3]{\frac{\pi_B}{\pi_M}(1-\xi)^6+\frac{\pi_M-\pi_B}{7\pi_M}}=\sqrt[3]{\delta(0.5+x)^6+\beta} \end{aligned} \tag{4.16}$$

In this case, it is necessary to verify whether the conditions (4.16) are also solutions to the other five equations for some ξ,π_B,π_M. For illustrative purposes, an example is shown in Figure 4.7, where the probability mass functions of CUB and CUM for $r=2,3,4,5,6$ are plotted against ξ. The values of ξ_D and ξ_U are set as in (4.16), so that (4.4) is always verified for $r=1,7$.

Since the functions of both pairs $P_B(R=2),P_B(R=6)$ and $P_M(R=2),P_M(R=6)$ are reflections of each other with respect to $\xi=0.5$, and the same holds for the pairs $P_B(R=3),P_B(R=5)$ and $P_M(R=3),P_M(R=5)$, the naive approach is followed again, consisting of a graphical search for the solutions of $P_B(R=2),P_M(R=2)$ (Figure 4.8), resulting in the conclusion that the only possible common solution to equations 2 and 6 is $\xi=0.5$.

With $\xi=0.5$, according to (4.16), we have $\xi_D=\xi_U$, so that equations 2 and 6 of (4.15) are identical, given by

$$6\cdot 0.5^6\delta+\beta=3\left(\sqrt[3]{0.5^6\delta+\beta}\right)^2-6\sqrt[3]{0.5^6\delta+\beta}$$

that is solved for $\delta=1$ (which leads back to the case $\pi_B=\pi_M$),

$$\delta=\frac{64\left(1579+15\sqrt{555}\right)}{107653}=1.148803$$

which is not acceptable, as, according to (4.16), it leads to negative values of ξ_D and ξ_U in $\xi=0.5$, and

$$\delta=\frac{64\left(1579-15\sqrt{555}\right)}{107653}=0.7286364 \tag{4.17}$$

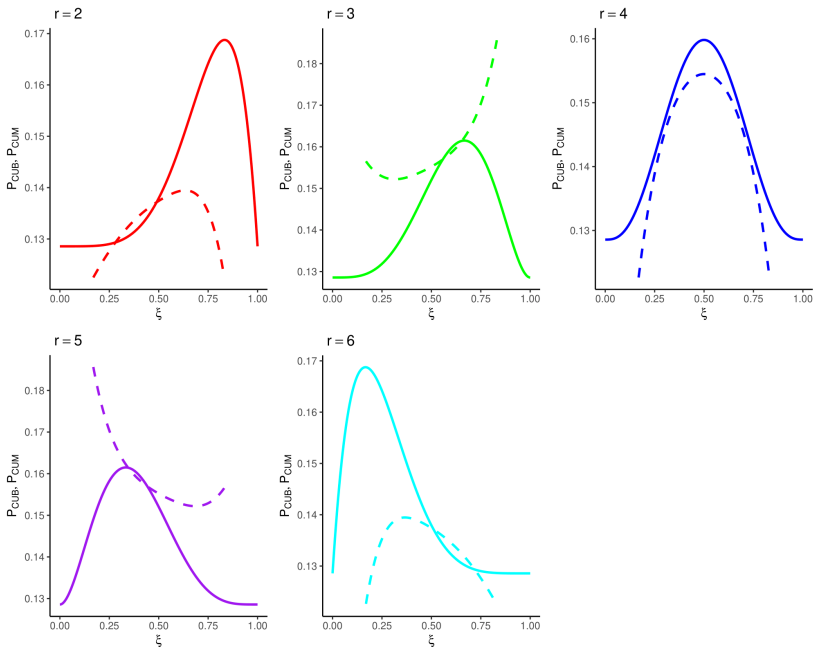


Figure 4.7: Probability mass functions of CUB (solid lines) and CUM (dashed lines), as functions of ξ, with ξ_D and ξ_U as in (4.16), $\pi_B = 0.1$, $\delta = 0.7$.

which implies $\xi_D = \xi_U \simeq 0.3688$ for $\xi = 0.5$ and whose effect is displayed in Figure 4.9, which clearly shows that the value $\xi = 0.5$, when (4.17) holds, is not a common solution to all the equations of system (4.15). This, of course, can also rather simply be verified analytically as shown in Appendix C.3 .

Proposition 4 *Given $m = 7$, for every $\theta_{CUB} = (\xi, \pi)'$, the relation (4.4) is satisfied only by the parameter vector θ_{CUM} defined in Proposition 3.*

From Propositions 3 and 4 it follows that, in the case $m = 7$, CUB is unidirectionally equivalent to CUM, without any restriction on its parameter space. The formalization of the function f defining unidirectional equivalence as in (4.2) is then

$$f : \Theta_{CUB} \to \Theta_{CUM},$$

with $f(\theta_{CUB}) = f((\xi, \pi)') = (\xi^2, (1-\xi)^2, \pi)'$, for all $\theta_{CUB} \in \Theta_{CUB}$.

4.2.3 *Properties of the functions describing unidirectional equivalence*

The functions described in Subsections 4.2.1 and 4.2.2 to formalise the unidirectional equivalence of CUB to CUM in the two cases $m = 5$ and $m = 7$ have the following properties:

- they are injective functions.
  In fact, since CUB models have proven to be identifiable for $m > 3$ (Iannario, 2010a), $\nexists\ \theta^{(1)}_{CUB}, \theta^{(2)}_{CUB}$, $\theta^{(1)}_{CUB} \neq \theta^{(2)}_{CUB}$, such that $f(\theta^{(1)}_{CUB}) = f(\theta^{(2)}_{CUB})$;

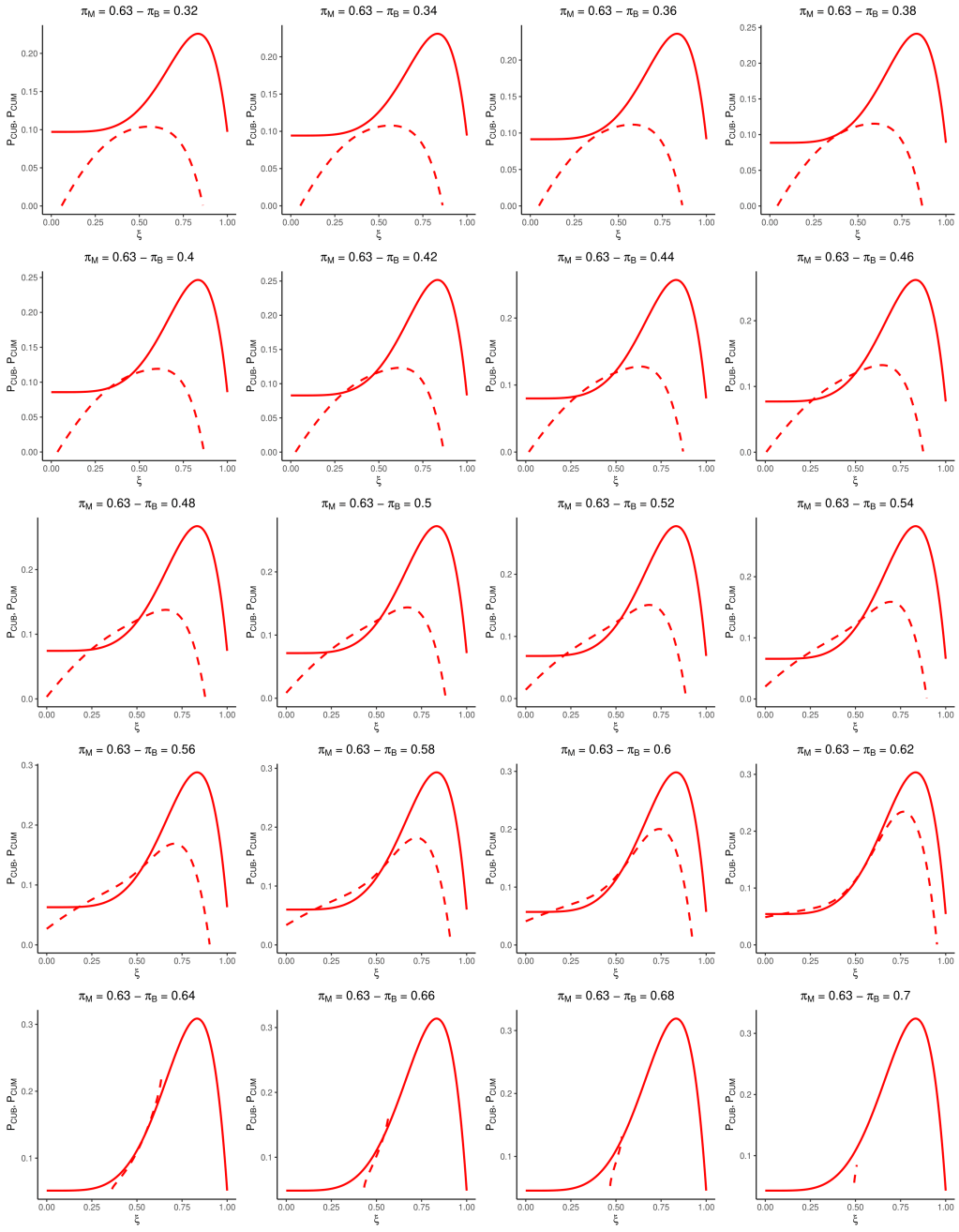


Figure 4.8: Probability mass functions of CUB (solid lines) and CUM (dashed lines), as functions of ξ, for $r = 2$, with ξ_D and ξ_U as in (4.9) and for $\pi_M = 0.63$ and π_B ranging from 0.32 to 0.7.

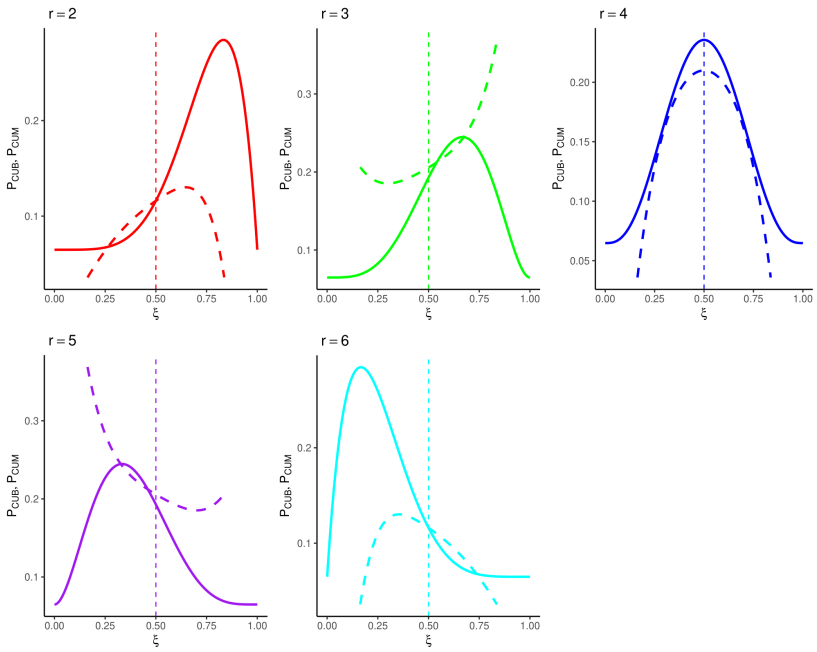


Figure 4.9: Probability mass functions of CUB (solid lines) and CUM (dashed lines), as functions of ξ, with ξ_D and ξ_U as in (4.16), with $\pi_M = 0.75$, $\pi_B = \delta\pi_M$, where δ is expressed as in (4.17): the value $\xi = 0.5$ is a common solution to equations 2 and 6, but not to the others.

- they are not surjective functions.
  In fact, $\exists\ \theta_{CUM}$ such that, for at least two values of r, $P_{CUB}(R = r \mid \theta_{CUB}) \neq P_{CUM}(R = r \mid \theta_{CUM})$ for all θ_{CUB}, as easily proven in the following.

To prove the non-surjectivity of the functions, it is recalled that there exists a relation between the parameters ξ_D and ξ_U that guarantees the unidirectional equivalence of CUB to CUM within the restricted parameter space $\bar{\Theta}_{CUB}$.

In detail, both for $m = 5$ and $m = 7$ the function f is such that $f(\theta_{CUB}) = f((\xi, \pi)') = (\xi^2, (1-\xi)^2, \pi)'$, which implies the relation

$$\xi_U = (1 - \sqrt{\xi_D})^2. \tag{4.18}$$

So, all the pairs (ξ_D, ξ_U) that do not obey (4.18) imply a CUM model that is not the (unidirectional) equivalent of any CUB. Thus, bidirectional equivalence between CUB and CUM does not hold.

Within the parameter space of CUM, a curve can be observed, representing CUM models with feeling parameters that follow the relation (4.18). Namely, the red curve represents the CUM models that yield equal theoretical probabilities of CUB models (i.e., the subspace of Θ_{CUM} where local equivalence to CUB holds) and will be called *equivalence curve* hereafter. The equivalence curve a discontinuity point when $m = 5$ and is continuous when $m = 7$, as illustrated in Figure 4.10.

The patterns of ξ_D, ξ_U, and $1 - \xi_D - \xi_U$ following relation (4.18) are visualised as functions of ξ in Figure 4.11.

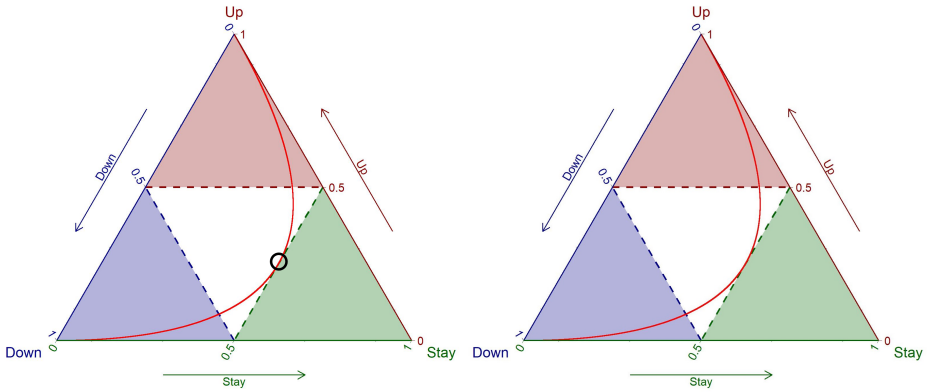


Figure 4.10: CUM parameter space with the subset of CUM models that are equivalent to a CUB model (red line), in the case with $m = 5$ (left) and $m = 7$ (right).

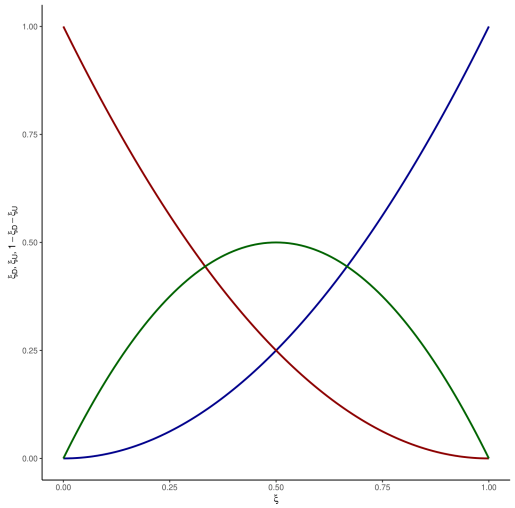


Figure 4.11: Pattern followed by ξ_D (blue), ξ_U (red) and $1 - \xi_D - \xi_U$ (green), as functions of ξ, for the CUM models on the red curve in Figure 4.10.

In this section, the results obtained from data simulated using three selected examples of data generating processes are presented. These examples are of interest for a better understanding and a more immediate interpretation of how the formal correspondence between the models can be practically exploited.

The three generating processes have been selected such that they all characterize a CUM model with $m = 7$ categories in a specific point in the parameter space: Cases 1 and 2 lie respectively far and close to the equivalence curve while Case 3 lies on the curve, and hence it has an equivalent CUB model. The set of the parameters chosen for the analysis is reported in Table 4.1.

Table 4.1: Set of parameters of the CUM model for each selected data generating process, and set of parameters of the equivalent CUB model.

DGP	CUM			CUB	
	ξ_D	ξ_U	π_M	ξ	π_B
Case 1	0.40	0.50	0.80	-	-
Case 2	0.70	0.10	0.80	-	-
Case 3	0.15	0.37	0.80	0.39	0.80

For each examined case, $iter = 1000$ datasets of $n = 1000$ observations were generated. The data were then fitted using both the CUM and the CUB models. Figure 4.12 shows the estimates of ξ_D, ξ_U, and ξ in the CUM triangular parameter space (top panels), the distribution of the estimates of π_B and π_M (middle panels), and one single example of the distribution of the observed frequencies (grey bars) compared to the theoretical frequencies of the two models obtained with the estimated parameters (bottom panels).

For each case, the difference between the BIC of the CUM model (BICCUM) and the BIC of the CUB model (BICCUB) was also calculated. The distribution of this difference is reported in Figure 4.13 (top panels). Negative values indicate that BICCUM $<$ BICCUB, suggesting that the CUM model should be preferred. Conversely, positive values indicate a preference for CUB according to BIC. The same analysis was conducted for the Diss index, and the distributions are presented in Figure 4.12 (bottom panels). The interpretation of this difference is the same as for the BIC.

In the first case, the decision process followed by the respondents is clearly the one which is assumed to be modelled by the CUM model. Consequently, using the CUB model results in inaccurate estimates for both the parameters ξ and π_B. From a practical perspective, both the Diss index and the BIC strongly indicate a preference for the CUM model, as the distributions of the BIC and Diss differences are entirely in the negative range.

In the second case, the estimates for the parameters ξ and π are not substantially different between the models. The description of feeling and uncertainty made by the two models through the estimated parameters, notwithstanding the differences in interpretation that the two models imply for the decision process, are quite consistent

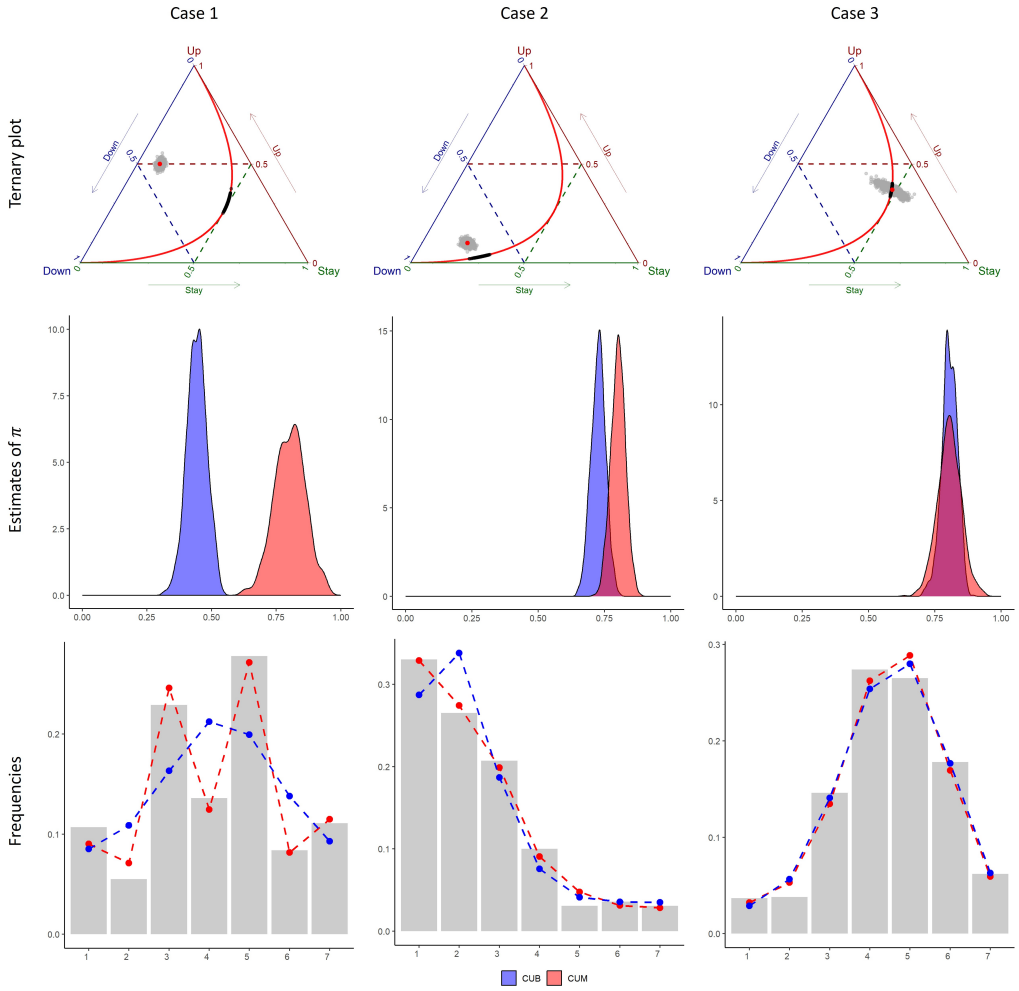


Figure 4.12: Top: ternary plot reporting the estimates of ξ_D and ξ_U and ξ. The theoretical CUM model is represented with a red point. Middle: distributions of the estimates of π_B and π_M. Bottom: example of observed and theoretical frequencies of CUB and CUM.

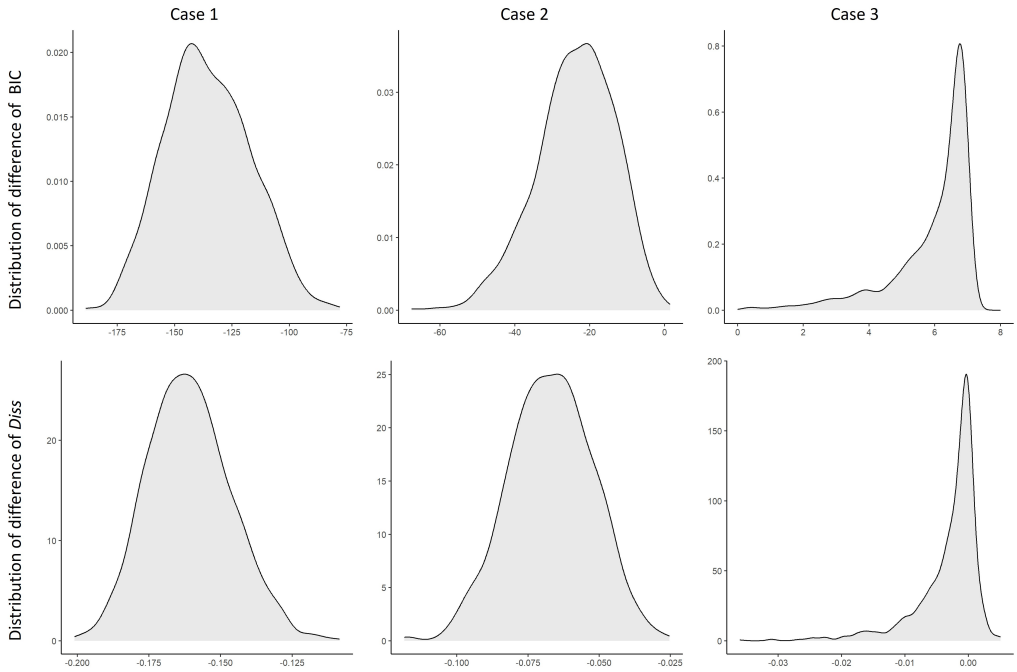


Figure 4.13: Top row: distributions of the differences between the BIC of the CUM model and the BIC of the CUB model. Bottom row: distributions of the differences between the Diss index of the CUM model and the Diss index of the CUB model.

which each other. However, both the Diss index and the BIC indicate that the CUM model has to be preferred to the CUB model.

Finally, in the third case, the Diss index does not identify significant differences between the two models, and in 99% of cases, the BIC suggests a preference for the CUB model. Additionally, when examining the parameter estimates, the CUB model is preferable due to its parameter estimates having lower variability compared to those of the CUM model.

4.4 DISCUSSION

In this study a definition of model equivalence for discrete probability distributions has been formalized, distinguishing between local/global and unidirectional/bidirectional properties. According to this definition, the equivalence of CUB and CUM model has been investigated by focusing on the two cases of Likert scales with $m = 5$ and $= 7$ categories, which are very commonly adopted choices in practice. It has been found that unidirectional equivalence of CUB to CUM holds for the whole CUB parameter space in case $m = 7$, while some restrictions apply in the case $m = 5$. Instead, the CUM model is equivalent to CUB, only locally, specifically when the relationship (4.18) is valid, so bidirectional equivalence between the two models does not hold. From the point of view of the probability distribution of the responses, this means that CUM is able to cover a wider set of situations, as could also be suggested by the higher number of parameters of CUM with respect to CUB.

However, this study has also revealed that some serious concerns might be raised against CUM model in the case $m = 5$. In fact, having found a set of CUB models to which correspond two CUM models, formally proves that CUM is not identifiable is some areas of its parameter space. In detail, it has been found that for $m = 5$,

$$P_{CUM}(R = r \mid \theta^{(1)}_{CUM}) = P_{CUM}(R = r \mid \theta^{(2)}_{CUM}) \quad \forall r = 1, 2, \ldots, 5$$

if $\theta^{(1)}_{CUM} = (0.5^2, 0.5^2, \pi)$ and $\theta^{(2)}_{CUM} = (0.3, 0.3, 5\pi/4)$. Further investigations proved that the non-identifiability area extends to all cases where $\xi_D = \xi_U$, provided that the two uncertainty parameters have a certain ratio (depending on ξ_D and ξ_U) between them.

So far, any similar evidence for the case $m = 7$ has not been found.

However, this issue is worth a specific in-depth analysis, which will be the topic of our future research.

Accordin to these results, if possible, Likert scales with $m = 7$ should be preferred. In any case, it is advisable to perform parameter estimation by initialising the EM algorithm with a wide set of randomly defined starting values and checking that the algorithm does not converge to different estimated values.

With respect to the equivalence of CUB to CUM, this study proves the very interesting fact that, for each probability distribution modelled by CUB, there exists a CUM model able to reproduce the same distribution. Recalling the different unconscious Decision Process implied by the two models, this evidence opens the fascinating prospect of being able to give a double interpretation to each set of responses. The first intriguing point is that equivalent models have the same uncertainty parameters. So, the differences only pertain to the feeling component of the decision process. According to Manisera and Zuccolotto (2022), the CUB model implicitly assumes that the respondent starts his reasoning from the bottom of the Likert scale and the feeling parameter $1 - \xi$ measures his propensity of moving upward. The CUM model, instead, is built thinking to a respondent who starts from the middle and moves upward or downward with propensities given by ξ_D and ξ_U, respectively. Nevertheless, it has been proven that in certain cases, these two different approaches can lead to the same probability distribution of the responses. This can provide interesting insights into the psychological processes underlying the expression of ratings.

5

EVALUATION OF THE SYNETHETIC EXPERIENCE IN MUSEUMS THROUGH MIXTURE MODELS

CONTRIBUTIONS RELATED TO THIS CHAPTER:

Indexed Journals (WoS, Scopus)

[IJ3] Ventura M., Macis A., Manisera M., and Zuccolotto P., *A Mixture Model to Analyse Synesthetic Experience in Museums using Multi-point Semantic Differential Scales*, *Under review*

International Conferences

[IC3] Ventura M., Macis A., Manisera M., Zuccolotto P., *Synesthetic Experience in Museums: a Statistical Analysis based on Semantic Differential Scales*, 29th Nordic Conference in Mathematical Statistics (NORDSTAT), Gothenburg (Sweden), 19-22 June 2023.

[IC4] Ventura M., Macis A., Manisera M., Zuccolotto P., *A Mixture Modelling Approach to Enhance the Multisensory Experience of Museum Visitors*, Joint Conferences Data Science, Statistics and Visualisation (DSSV) and European Conference on Data Analysis (ECDA), Antwerp (Belgium), 5-7 July 2023.

Questionnaires have been particularly popular for evaluating customers' satisfaction by asking them to rate their opinion on a response scale. Moreover, questionnaires and rating scales are widely used for evaluating visitors' experiences in various settings, including museums. Galleries, expositions, and museums aim at inspiring and educating visitors through their collections. Therefore, it is important to map visitors' experiences, behaviours, and attitudes in cultural places to enhance the knowledge of organizations and decision-makers for understanding how to improve the impact and relevance of cultural institutions.

An original approach for understanding visitors' sensations is based on the concepts of synesthesia and ideasthesia (Cho, 2021). The former is a phenomenon where the stimulation of one sense causes the simultaneous triggering of another sense, such as seeing colours when smelling fragrances (Simner, 2012). The latter refers to the phenomenon where mental concepts trigger sensory experiences, such as feeling a sensation of heaviness when thinking about a difficult problem (Jürgens and Nikolić, 2012). These concepts have been used to explain how visitors experience museums and how their perceptions and emotions are influenced by the sensory qualities of the exhibits. They can also be used to explain what are the emotions, sensations, and perceptions that characterise their visit and how they are generated and conditioned by the sensory qualities of the visit (Bitgood, 2011; Falk and Dierking, 2016).

Various response scales are available for rating the sensations of the visitors (Dawis, 1987). Multi-point Semantic Differential scales are particularly suitable for measuring individuals' emotional responses to stimuli, since when using these scales, the visitors are asked to rate their experience on a multi-point scale between two bipolar adjectives (e.g., sad and happy, loud and quiet, bored and exited, etc.) (Osgood, 1962; Osgood, Suci, and Tannenbaum, 1957).

The work presented in this chapter aims at investigating the synesthetic experience of museum visitors by applying statistical models such as CUB and CUM. Specifically, the study focuses on analyzing the emotions, sensations, and feelings of visitors in three different rooms of the Pinacoteca Tosio Martinengo in Brescia, Italy. This research is part of the Data Science for Brescia (DS4BS) project, which aims to promote data-driven decision-making and innovation in the city by supporting the development of data science skills and tools for local stakeholders.

5.1 THE SYNESTHETIC VISITOR EXPERIENCE

In this section, the basic CUB and the CUM model are applied to data collected within the DS4BS – Arts and Cultural Places project. The project is a collaborative initiative between academic researchers, museum professionals, and city officials aimed at leveraging the power of data science and statistics to improve the experience of visitors to museums and cultural institutions in the city of Brescia, Italy.

The research explained in this work is part of the DS4BS project, and its primary goal is to gain a better understanding of visitors' sensory experience in museums. Indeed, while vision is the most stimulated sense during museum visits, the artworks and other visual elements present in a museum can also stimulate other senses and sensations. The synesthetic experience of visitors was evaluated by studying perceptions related to touch, smell, taste, and vision using semantic differential scales, and different aspects for each sense were investigated. This research allowed us to gain a deeper understanding of visitors' sensory experiences in museums and to gain insights into the factors that contribute to these experiences. The insights gained from this analysis can be used to inform the development of strategies and interventions aimed at enhancing visitors' experiences and improving the overall quality of cultural institutions in the city.

The data related to the perceptions of visitors were collected by administering a questionnaire to the visitors of the Tosio-Martinengo Art Gallery in Brescia, Italy. The study was conducted over a period of several weeks. The questionnaire consisted of several questions, but attention was focused on question number 4, which aims to investigate the synesthetic experience of visitors by measuring perceptions related to touch, smell, taste, and vision using multi-point semantic differential scales. Three different aspects for each sense were measured, as shown in Table 5.1.

The questionnaire was administered in three different rooms, each one characterized by a different wall colour. Specifically, the room dedicated to sacred works dating back to the early 16th century is characterized by walls covered in blue velvet, while the room housing the works of the 16th-century painters Moretto, Savoldo, and Lotto features walls covered in red velvet. The room dedicated to portraits is covered with green velvet.

Out of the 1024 questionnaires administered, 32.9% were filled out in the blue room, 32.6% in the red room, and 34.5% in the green room.

Table 5.1: Question: Below you will find pairs of opposite adjectives. Observing these pairs, darken the box that best corresponds to your perception of the visit to the room.

Touch	Rough	☐	☐	☐	☐	☐	☐	☐	Soft
	Angular	☐	☐	☐	☐	☐	☐	☐	Rounded
	Sticky	☐	☐	☐	☐	☐	☐	☐	Fluid
Smell	Claustrophobic	☐	☐	☐	☐	☐	☐	☐	Airy
	Antique	☐	☐	☐	☐	☐	☐	☐	New
	Foul	☐	☐	☐	☐	☐	☐	☐	Aromatic
Taste	Bitter	☐	☐	☐	☐	☐	☐	☐	Sweet
	Spicy	☐	☐	☐	☐	☐	☐	☐	Fruity
	Tasteless	☐	☐	☐	☐	☐	☐	☐	Savory
Vision	Glacial	☐	☐	☐	☐	☐	☐	☐	Tropical
	Pale	☐	☐	☐	☐	☐	☐	☐	Radiant
	Cloudy	☐	☐	☐	☐	☐	☐	☐	Clear

Since gaining information about the experience of visitors using traditional statistical techniques would have been difficult, the collected data in each room were then analyzed using the CUM model, chosen for its suitability in analyzing data obtained from Semantic Differential scales, which has been compared to the results obtained with the CUB model, considered as a benchmark.

These models' performances were compared using the *Diss* index and the Akaike Information Criterion (Akaike, 1973). The results are presented in Table 5.3. For the item *Glacial - Tropical* in the Green room, the results are not shown because in that case, the estimates of the CUM model are not stable, as explained in more detail below.

The comparison between the CUB model, which has two parameters, and the CUM model, which has three, is not entirely equitable. While the interpretation and the Decision Processes underlying the two models differ, a more balanced comparison can be made by evaluating the CUM model against the CUB model with a shelter option. The CUB model with shelter is particularly meaningful when applied to cases where the presence of a shelter option is clearly highlighted by the frequency distribution. Accordingly, based on the frequency distributions provided in Appendix D, specific items from each room were selected for the analysis. The parameter estimates obtained using the CUB model with shelter are summarized in Table 5.2.

Overall, the CUM model showed better performances, compared to the CUB model, regardless of the evaluation criterion used to assess the Goodness of Fit. However, in most of the cases analyzed, the CUB model with shelter demonstrated a slight advantage over the CUM model.

The scale *Glacial - Tropical* in the Red room demonstrates notable differences in the `Diss` index when comparing the CUB and CUM models. The index is 0.1516 for the CUB model, whereas it decreases to 0.0346 for the CUM model. This is clearly represented in Figure 5.1, where the observed relative frequencies and the fitted probabilities for a relevant case in each room are reported. Another interesting example, albeit less evident, is observed with the scale *Sticky - Fluid* in the Green room, where the CUM model yields a `Diss` index of 0.0535, while the CUB model produces a slightly higher

Table 5.2: CUB model with shelter – Parameter estimates with Standard Error reported in parentheses and evaluation of the fit of the model in each selected case, with AIC values highlighted in bold to indicate better performances compared to CUM model.

Red room						
	Shelter	δ	$1-\xi$	$1-\pi$	AIC	Diss
Angular – Rounded	7	0.1620	0.6642	0.7428	1257.82	0.0448
		(0.0307)	(0.0572)	(0.0775)		
Claustrophobic – Airy	7	0.1718	0.6663	0.7433	**1253.22**	0.0396
		(0.0310)	(0.0577)	(0.0777)		
Bitter – Sweet	4	0.1083	0.7564	0.8606	**1281.56**	0.0304
		(0.0299)	(0.0881)	(0.0656)		
Glacial – Tropical	4	0.1968	0.6385	0.5338	1182.12	0.0644
		(0.0371)	(0.0339)	(0.0800)		
Blue room						
	Shelter	δ	$1-\xi$	$1-\pi$	AIC	Diss
Rough – Soft	7	0.1857	0.8315	0.3763	**1086.73**	0.0382
		(0.0600)	(0.0289)	(0.0522)		
Angular – Rounded	7	0.2044	0.7451	0.5010	**1188.61**	0.0549
		(0.0370)	(0.0333)	(0.0649)		
Sticky – Fluid	7	0.2730	0.7445	0.2576	**1079.08**	0.0435
		(0.0377)	(0.0234)	(0.0556)		
Claustrophobic – Airy	7	0.3433	0.7523	0.2174	**1006.27**	0.0069
		(0.0388)	(0.0234)	(0.0543)		
Cloudy – Clear	7	0.2661	0.8484	0.3516	**1003.86**	0.0262
		(0.0671)	(0.0299)	(0.0532)		

index of 0.0908 (Figure 5.1). Furthermore, in the Blue room, a relevant example is observed with the scale *Claustrophobic - Airy*, where the CUM model has a Diss index of 0.0396, while the CUB model has a notably higher index of 0.1433 (Figure 5.1).

In general, the *Diss* index values obtained from fitting the CUM model are consistently below 0.1, indicating that the proportion of data that require to be modified to obtain a perfect fit is less than 10%. Only in two items in the Green room, the *Diss* index is greater than 0.1, specifically for the scales *Foul - Aromatic* (0.1571) and *Tasteless - Savory* (0.1282).

The scale *Spicy - Fruity* in the Red and in the Green room is a case where the CUB model has slightly better performances than the CUM model according to the *Diss* index. The better fit of the CUM model can be attributed to its enhanced flexibility, which stems from the inclusion of two feeling parameters. Despite the increased complexity compared to the CUB model, the AIC supports the preference for the greater flexibility offered by the CUM model.

The results of the estimates, presented in Table 5.4, can be effectively represented in the parameter space of the models. The fitted CUB models can be visualized in their parameter space, which can be represented by a Cartesian plane with $1-\hat{\pi}$ on the x-axis, and $1-\hat{\xi}$ on the y-axis. The fitted CUM models can be represented on a triangular plot with the parameter ξ_D on the blue edge of the triangle, the parameter ξ_U on the red edge, and the quantity $1-\xi_D-\xi_U$ on the green edge. The uncertainty, instead, is represented through the size of the point that identifies the model in its parameter space. To make the plots easier to be understood, the triangular plot can be divided into three sub-triangles: the triangle in the upper part represents the subset of the parameter space with $\xi_U > 0.5$, the sub-triangle located in the bottom-left part represents the part

Table 5.3: Evaluation of the fit of CUB and CUM models in each room, with AIC values highlighted in bold to indicate better performances.

	Red room				**Green room**				**Blue room**			
	CUB		CUM		CUB		CUM		CUB		CUM	
	Diss	*AIC*	*Diss*	*AIC*	*Diss*	*AIC*	*Diss*	*AIC*	*Diss*	*AIC*	*Diss*	*AIC*
Rough – Soft	0.0333	**1153.71**	0.0286	1155.63	0.0698	**1321.80**	0.0598	1322.24	0.0771	1090.9943	0.0353	**1086.78**
Angular – Rounded	0.0761	1263.60	0.0593	**1260.91**	0.0445	**1360.67**	0.0427	1362.55	0.1050	1200.5086	0.0691	**1192.24**
Sticky – Fluid	0.0824	1184.16	0.0754	**1179.92**	0.0908	1290.41	0.0535	**1281.96**	0.1191	1102.4353	0.0607	**1083.96**
Claustrophobic – Airy	0.0572	1257.87	0.0503	**1256.48**	0.1037	1321.95	0.0911	**1317.33**	0.1433	1039.3723	0.0396	**1010.70**
Antique – New	0.0760	1282.91	0.0636	**1281.74**	0.0264	**1327.48**	0.0259	1329.10	0.0310	**1307.9947**	0.0229	1309.50
Foul – Aromatic	0.1358	1165.05	0.1037	**1158.77**	0.0901	1246.85	0.0473	**1240.31**	0.2154	1114.7281	0.1571	**1094.13**
Bitter – Sweet	0.0963	**1290.22**	0.0938	1292.20	0.0472	**1370.92**	0.0414	1372.14	0.1509	1216.6806	0.0908	**1193.45**
Spicy – Fruity	0.0464	**1299.65**	0.0492	1301.40	0.0554	**1347.80**	0.0557	1349.80	0.1169	1296.8325	0.1033	**1294.97**
Tasteless – Savory	0.1098	1176.08	0.0599	**1165.00**	0.0799	1302.32	0.0501	**1298.69**	0.1787	1221.2657	0.1282	**1204.22**
Glacial – Tropical	0.1516	1210.29	0.0346	**1179.20**	-	-	-	-	0.0822	1299.3332	0.0533	**1295.98**
Pale – Radiant	0.0870	1277.60	0.0696	**1276.23**	0.0593	**1375.36**	0.0390	1375.76	0.0579	1260.4521	0.0476	**1259.78**
Cloudy – Clear	0.0617	**1231.70**	0.0413	1232.12	0.0796	**1362.52**	0.0780	1365.30	0.0569	1007.3284	0.0264	**1004.30**

of the parameter space with $\xi_D > 0.5$, and the sub-triangle in the bottom-right part of the triangular plot represents the parameter space with $1 - \xi_D - \xi_U > 0.5$. Figures 5.3, 5.5, and 5.6 report the fitted models for the three rooms in their respective parameter spaces. The CUB parameter space is on the left and the triangular plot reporting the CUM models is on the right.

The CUM model has a complex likelihood surface, which causes stability issues in the estimates via the EM algorithm. Therefore, to assess the stability and uniqueness of the estimates, a grid of starting points for the algorithm was defined. Four pairs of values for ξ_D and ξ_U, and three values of π were considered. This grid represents twelve points in the parameter space of the CUM model (Figure 5.2). These points were used as initializations for the EM algorithm. Through this method, the reliability of the estimates was confirmed.

By observing the results represented in Figure 5.3, a parallelism between the CUB and the CUM models can be established. The triangular plot in Figure 5.3 shows that for eight out of twelve items, respondents exhibit a stronger inclination toward the upper part of the scale. This observation is further supported by the CUB parameter space shown in Figure 5.3, where the corresponding points for these eight items lie in the upper part of the plane due to higher values of the feeling parameter. This suggests that respondents are more inclined to provide higher ratings. The results unveil that respondents perceive the room as clear and luminous. Upon further contemplation, they feel sensations related with an airy and aromatic ambience, accompanied by flavours that are both fruity and savoury. The overall immersive experience is completed with the feeling of a sense of roundness and softness.

n the bottom-left sub-triangle, the CUM model fitted for the item *Antique - New* is found, indicating that respondents tend to position themselves towards the lower part of the scale. This behaviour is also captured by the CUB model, as the corresponding point for that scale is located in the lower part of the CUB parameter space. This result suggests that respondents perceive an odour reminiscent of a classic and old-fashioned room.

The explanatory power of the CUM model becomes evident when examining the position of the models fitted for the items *Spicy - Fruity*, *Bitter - Sweet*, and *Glacial - Tropical*. Although their positions in the CUB parameter space (Figure 5.3) suggest that

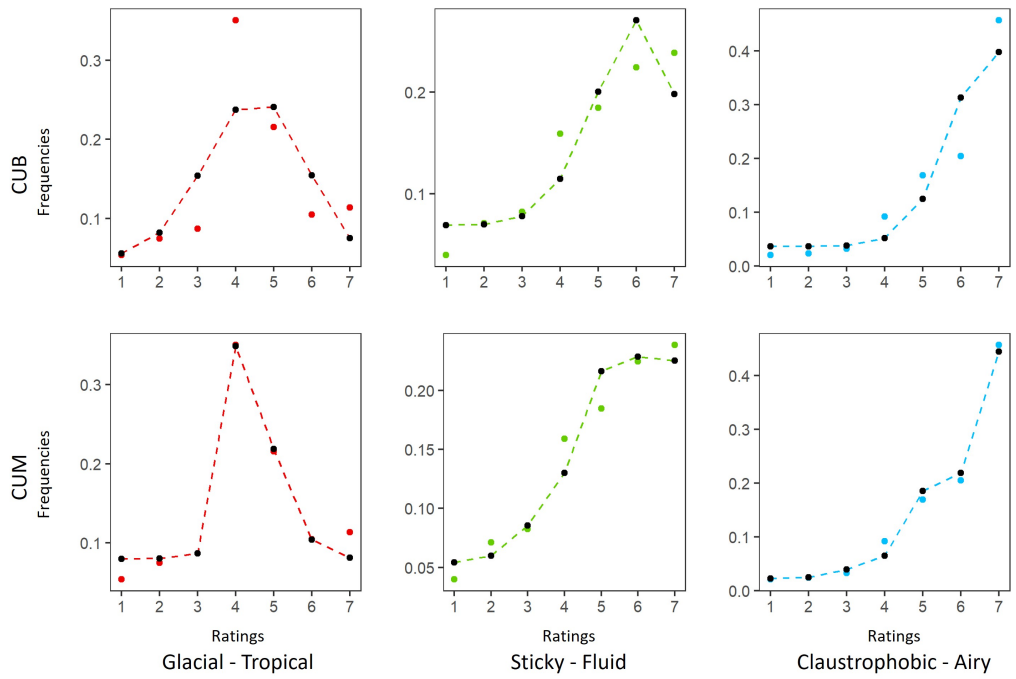


Figure 5.1: Observed relative frequencies (coloured dots) and fitted probabilities (black dots) for three notable examples in each one of the three rooms for both CUB and CUM models.

respondents tend to position themselves in the middle of the scale, the results in the triangular plot reveal a difference in the meaning of this tendency.

Specifically, the point related to the scale *Glacial - Tropical* is positioned in the bottom-right subtriangle, suggesting that the propensity of remaining still in the middle of the scale dominates on the other inclinations. The interpretation, in that case, is the same as for the CUB model. However, in the case of the *Spicy - Fruity* and *Bitter - Sweet* scales, the CUM model allows for a deeper understanding of respondents' feelings. Contrary to the interpretation suggested by the CUB model, the CUM model indicates that no single feeling dominates in these cases. The feeling parameters for these items in the CUM model are nearly identical (Table 5.4).

The procedure described above for assessing the stability of the estimates was performed also for the Green room. In that case, for the item *Glacial - Tropical*, the procedure identified two different estimates, as shown in Figure 5.4. The tern of parameters $\{\pi, \xi_D, \xi_U\}$ that identifies each model is A = $\{0.2894, 0.0999, 0.2269\}$, and the other one is identified by the tern B = $\{0.4621,$
$0.2742, 0.3801\}$. As a result, this item was excluded from the interpretation of the results in the Green room.

The results obtained for the Green room reveal a stronger inclination toward the upper part of the scale for seven items. Similarly to the Red room, visitors in the Green room experience a delightful ambience of airiness and radiance, accompanied by a pleasant aroma that evokes fruity and intense flavors. The findings associated with the items *Antique - New* and *Spicy - Fruity* suggest that the respondents tend to move towards the lower part of the scale, reflecting their perception of an ancient smell and

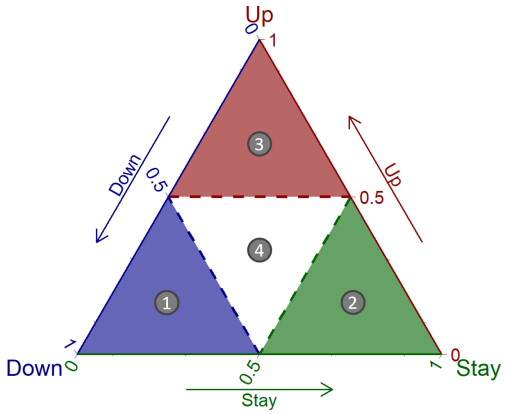
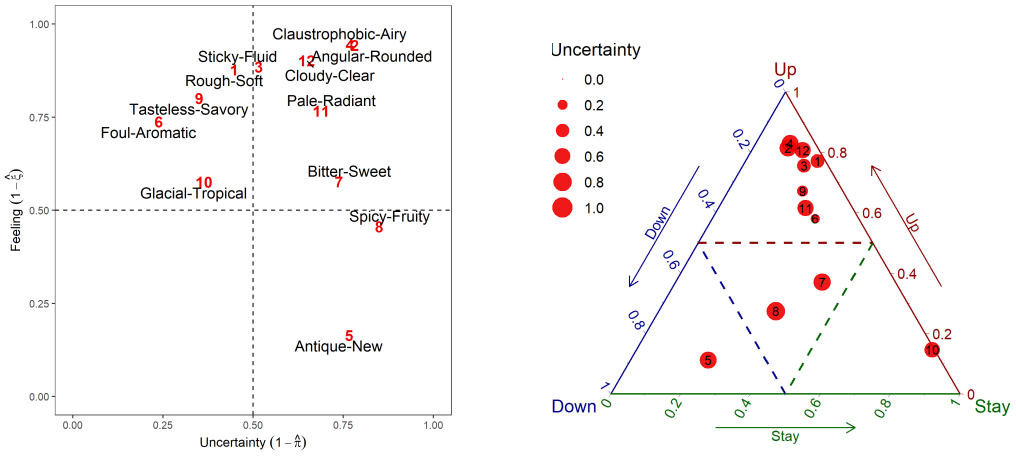


Figure 5.2: Starting points for the EM algorithm. For ξ_D and ξ_U the set of values {0.10, 0.20, 0.33, 0.70} was considered, and for π the set of values {0.25, 0.5, 0.75} was considered.



(a) CUB model parameter space.

(b) CUM model parameter space.

Figure 5.3: CUB (left) and CUM (right) models for the Red room represented in their parameter spaces. The labels of the items are associated with the number reported both on the left and right panels.

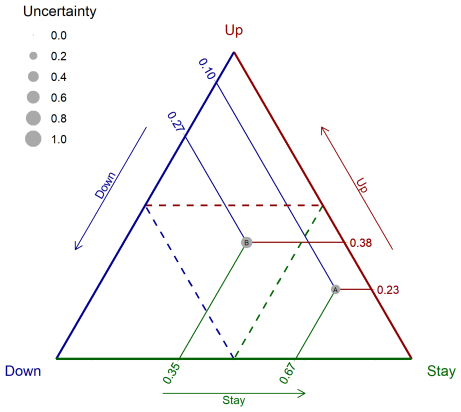
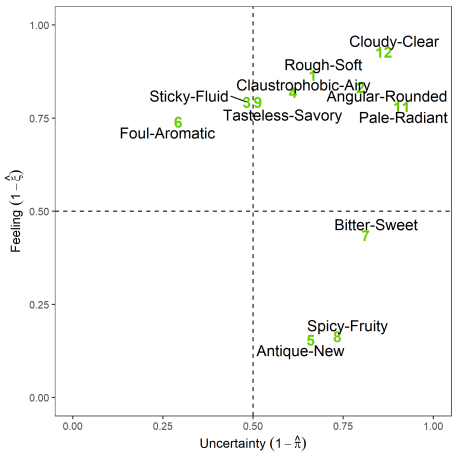
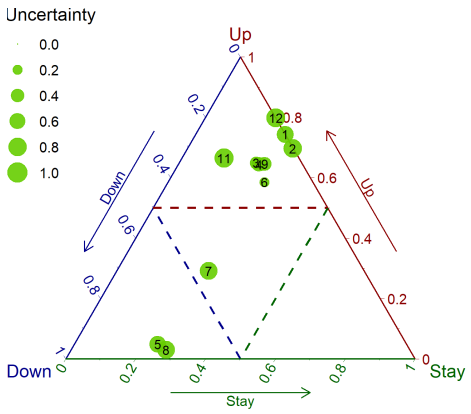


Figure 5.4: Parameter estimates for the scale *Glacial - Tropical* in the Green room identified by the grid procedure for initializing the EM algorithm.

a spicy taste. This evidence can be observed in Figure 5.5. However, the CUB and the CUM model offer a different point of view on the interpretation of the results for the item *Bitter - Sweet*. According to the CUB model, there is a distinct inclination to move in the middle of the scale. Conversely, the feeling parameters of the CUM model suggest that no specific feeling dominates in the respondents' perceptions.



(a) CUB model parameter space.

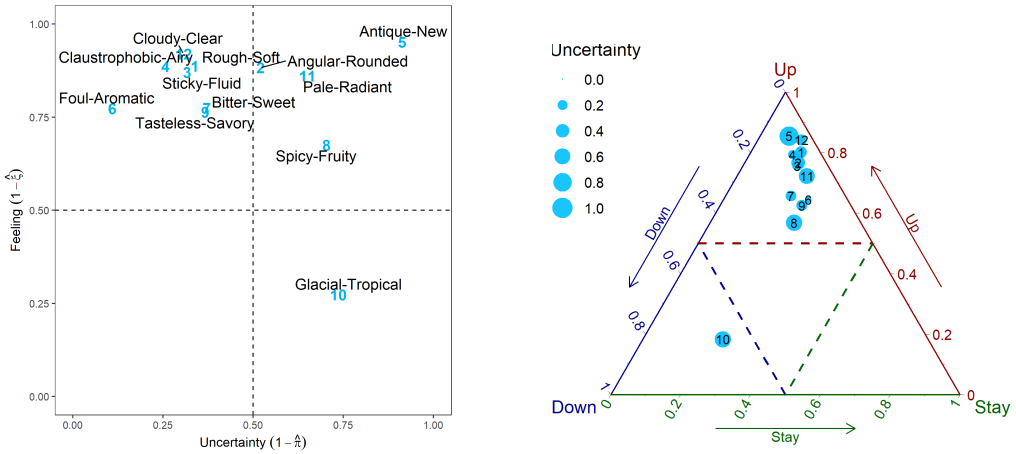


(b) CUM model parameter space.

Figure 5.5: CUB (left) and CUM (right) models for the Green room represented in their parameter spaces. The labels of the items are associated with the number reported both on the left and right panels.

The interpretation of the results in the Blue room remains consistent, regardless of whether the CUB or CUM model is considered. In fact, respondents demonstrate a consistent tendency to move towards the upper part of the scale for eleven items. However, item *Glacial - Tropical* highlights that the respondents tend to provide lower ratings and

show no inclination towards the upper part of the scale, as observed from the perspective of the CUB model. Conversely, according to the CUM model's perspective, this result indicates a stronger inclination towards the lower part of the scale. In both cases, the underlying implication remains the same: when respondents are in the Blue room, they experience sensations associated with an icy and glacial environment.



(a) CUB model parameter space.

(b) CUM model parameter space.

Figure 5.6: CUB (left) and CUM (right) models for the Blue room represented in their parameter spaces. The labels of the items are associated with the number reported both on the left and right panels.

These results provide valuable insights into the sensory experiences of visitors when they explore the three rooms of the art gallery. The findings indicate that the vibrant colors of the walls not only stimulate the sense of sight but also trigger various additional sensory perceptions. However, the aim of this study was to investigate what were the specific sensations triggered by specific colors on the walls. It can be concluded that color influences the sensory experience of visitors when it elicits distinct sensations that consistently differ across the rooms.

By observing the results regarding the first three items of the questionnaire, which were designed to assess tactile perceptions evoked by room colors, it is possible to notice that the results in the three rooms are similar to each other. In fact, the respondents appear to be more inclined to move toward the upper end of the scale, indicating that color does not have a significant effect on tactile perceptions.

The same considerations apply to olfactory perceptions as well. The only olfactory perception that shows a significant difference is the one assessed to the *Antique - New* scale in the Green and Red rooms, where respondents tend to move towards the lower end of the scale, differently from the Blue room.

Notable differences are evident across the rooms when it comes to the gustatory and visual experience. Specifically, variations are observed in the *Bitter - Sweet* and *Spicy - Fruity* scales for taste, indicating divergent preferences among respondents. Additionally, noticeable distinctions emerge in the *Glacial - Tropical* scale for vision, highlighting varying perceptions and preferences in relation to the room aesthetics.

Therefore, as a result, all the rooms create an impression of clarity and radiance, evoking sensations of pleasant tastes and aromas.

To investigate the synesthetic experience elicited by the visit to these rooms, the focus was placed on the items *Glacial - Tropical*, *Antique - New*, *Bitter - Sweet*, and *Spicy - Fruity*. The results for these items are reported in Figure 5.7, where the CUM models estimated in each room are presented and differentiated by color.

The findings indicate that when respondents visit the Red room, they perceive an aroma reminiscent of something classical, ancient, and old-fashioned. This similar olfactory experience is also observed when they are in the Green room. However, what distinguishes the Green room is that the color itself has an additional impact on their multisensory experience, specifically their sense of taste, indeed, respondents report perceiving a unique combination of bitter and spicy flavors. Finally, the Blue room creates an impression of being in a glacial environment, and the respondents perceive a smell of newness and experience a pleasant taste characterized by sweet and fruity notes.

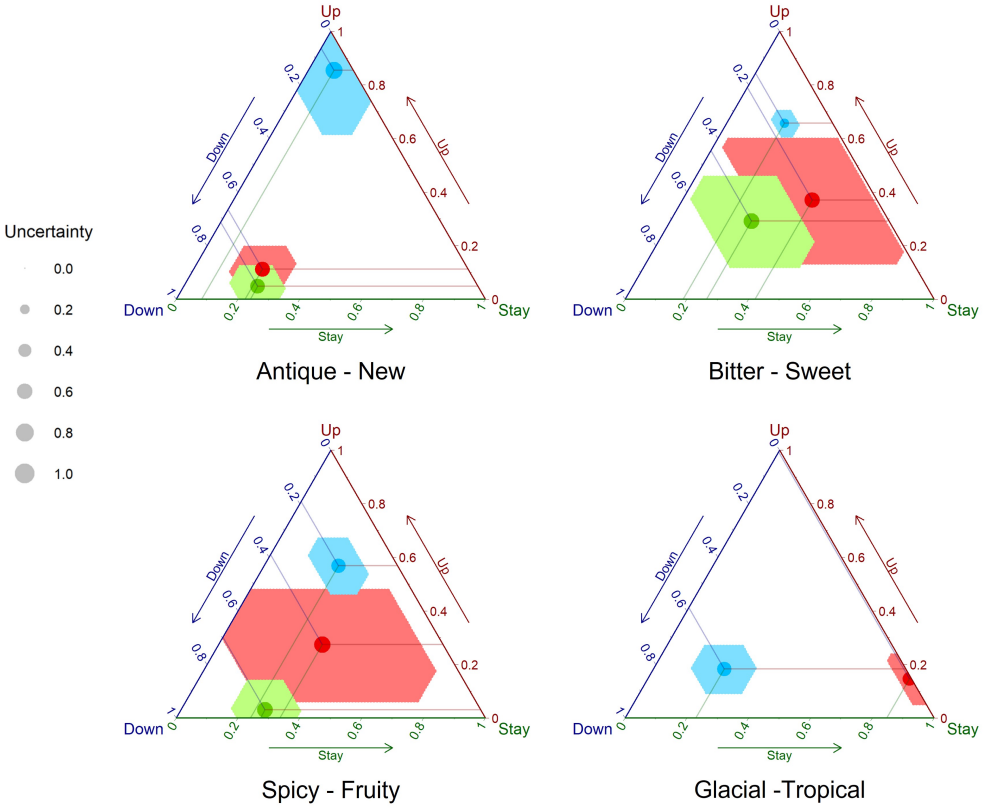


Figure 5.7: Triangular plots reporting the CUM models fitted for the three Rooms for the four items considered to evaluate the synesthetic experience. The color of the points corresponds to the color of the room. The area represents the intersection of the 95% Confidence Intervals around the estimated parameters for each CUM model.

5.1.1 *CUB model with shelter option*

The comparison between CUB and CUM models can be considered

5.2 DISCUSSION

This study aimed to analyse the sensations experienced by visitors at the Pinacoteca Tosio Martinengo, an art gallery located in Brescia, Italy. The visitors' experience was defined as synesthetic experience, as the study was based by considering the concepts of synesthesia and ideasthesia. The research yielded interesting results regarding sensations and perceptions related to various senses. To explore these aspects, a questionnaire was administered, and the data were analyzed using two notable models for rating data: the CUB and CUM models. The former served as a benchmark, while the latter proved suitable for data derived from Semantic Differential response scales.

Both the models are univariate models and allow the measurement of two latent traits: the feeling and the uncertainty which jointly act on the final rating chosen by the respondent.

Our study has demonstrated that the CUM model allows for a different point of view in the interpretation of the results, compared to the CUB model. Specifically, the CUM model enables us to gain deeper insight into the nature of cases where the feeling parameter in the CUB model assumes intermediate values. Notably, the CUM model distinguishes between instances where there is a tendency to remain stationary without significant movement up or down the scale and cases where no specific inclination dominates. However, it is worth noting that the current version of the CUM model lacks the capability to include covariates, which could significantly enhance our understanding of the results. Further research is needed to develop an extended version of the CUM model that incorporates covariates, thereby providing more comprehensive insights into the data.

From our analyses emerged that colors evoke different sensations and perceptions in visitors. Therefore, the color blue should be considered to evoke a sensation associated with something new, sweet, and linked to a glacial image. Conversely, for evoking a bitter or spicy taste, the color green may be more appropriate, potentially leading to an association with an antique scent.

These results are relevant for decision-makers involved in shaping and enhancing the visitor experience within museums and cultural institutions. By integrating these insights with tangible aspects of the museum visit, such as satisfaction levels and visitor expectations, decision-makers can make informed decisions on how to curate spaces, design exhibitions, and select color schemes that optimize the multisensory journey for visitors.

Table 5.4: Parameter Estimates for CUB and CUM Models in the three rooms, with Standard Errors shown in parentheses.

	Red room					**Green room**					**Blue room**				
	CUB		**CUM**			**CUB**		**CUM**			**CUB**		**CUM**		
	$1-\pi$	$1-\xi$	$1-\pi$	ξ_D	ξ_U	$1-\pi$	$1-\xi$	$1-\pi$	ξ_D	ξ_U	$1-\pi$	$1-\xi$	$1-\pi$	ξ_D	ξ_U
Rough – Soft	0.4477	0.8779	0.4396	0.0224	0.7713	0.6658	0.8662	0.6725	0.0000	0.7438	0.3381	0.8873	0.3004	0.0540	0.8014
	(0.0458)	(0.0142)	(0.0525)	(0.0254)	(0.0249)	(0.0475)	(0.0202)	(0.0538)	(0.0563)	(0.0359)	(0.0426)	(0.0120)	(0.0434)	(0.0160)	(0.0200)
Angular – Rounded	0.7823	0.9435	0.6934	0.0856	0.8146	0.7979	0.8328	0.8114	0.0025	0.6969	0.5204	0.8832	0.4276	0.0795	0.7664
	(0.0482)	(0.0308)	(0.0666)	(0.0374)	(0.0545)	(0.0528)	(0.0388)	(0.0634)	(0.0767)	(0.0680)	(0.0535)	(0.0201)	(0.0581)	(0.0209)	(0.0289)
Sticky – Fluid	0.5158	0.8853	0.4216	0.0687	0.7557	0.4826	0.7933	0.3696	0.1303	0.6492	0.3166	0.8709	0.2021	0.0875	0.7564
	(0.0519)	(0.0193)	(0.0626)	(0.0226)	(0.0307)	(0.0581)	(0.0213)	(0.0656)	(0.0258)	(0.0283)	(0.0494)	(0.0152)	(0.0477)	(0.0162)	(0.0209)
Claustrophobic – Airy	0.7688	0.9457	0.6938	0.0717	0.8297	0.6106	0.8195	0.4749	0.1251	0.6442	0.2570	0.8867	0.1565	0.0830	0.7938
	(0.0465)	(0.0268)	(0.0652)	(0.0372)	(0.0536)	(0.0604)	(0.0299)	(0.0800)	(0.0339)	(0.0362)	(0.0449)	(0.0130)	(0.0401)	(0.0136)	(0.0185)
Antique – New	0.7664	0.1654	0.6818	0.6649	0.1124	0.6599	0.1551	0.6388	0.7120	0.0485	0.9135	0.9525	0.8903	0.0627	0.8550
	(0.0636)	(0.0522)	(0.0782)	(0.0582)	(0.0457)	(0.0525)	(0.0253)	(0.0639)	(0.0410)	(0.0401)	(0.0442)	(0.0639)	(0.0590)	(0.0808)	(0.1200)
Foul – Aromatic	0.2381	0.7385	0.1741	0.1246	0.5806	0.2924	0.7402	0.2018	0.1386	0.5852	0.1089	0.7737	0.0662	0.1120	0.6448
	(0.0505)	(0.0143)	(0.0505)	(0.0203)	(0.0223)	(0.0533)	(0.0156)	(0.0553)	(0.0242)	(0.0232)	(0.0389)	(0.0116)	(0.0305)	(0.0145)	(0.0182)
Bitter – Sweet	0.7383	0.5773	0.7134	0.2094	0.3703	0.8124	0.4359	0.7321	0.4450	0.2919	0.3717	0.7752	0.2473	0.1560	0.6568
	(0.0700)	(0.0470)	(0.1112)	(0.0986)	(0.1196)	(0.0705)	(0.0670)	(0.1048)	(0.0812)	(0.0860)	(0.0586)	(0.0180)	(0.0575)	(0.0204)	(0.0239)
Spicy – Fruity	0.8498	0.4566	0.8081	0.3901	0.2745	0.7338	0.1647	0.7325	0.6976	0.0300	0.7036	0.6754	0.6138	0.1915	0.5685
	(0.0722)	(0.0878)	(0.1072)	(0.1598)	(0.1067)	(0.0531)	(0.0318)	(0.0691)	(0.0556)	(0.0585)	(0.0673)	(0.0403)	(0.0814)	(0.0394)	(0.0518)
Tasteless – Savory	0.3510	0.8011	0.2625	0.1149	0.6715	0.5132	0.7935	0.4369	0.1078	0.6457	0.3672	0.7640	0.2493	0.1407	0.6249
	(0.0540)	o(.0166)	(0.0547)	(0.0207)	0.0240)	(0.0580)	(0.0223)	(0.0652)	(0.0277)	(0.0305)	(0.0610)	(0.0185)	(0.0600)	(0.0204)	(0.0242)
Glacial – Tropical	0.3639	0.5762	0.5607	0.0071	0.1466	-	-	-	-	-	0.7359	0.2738	0.6237	0.5878	0.1837
	(0.0583)	(0.0187)	(0.0529)	(0.0193)	(0.0270)	-	-	-	-	-	(0.0687)	(0.0531)	(0.0865)	(0.0513)	(0.0449)
Pale – Radiant	0.6883	0.7669	0.6154	0.1345	0.6157	0.9118	0.7816	0.8643	0.2147	0.6653	0.6503	0.8611	0.5834	0.0775	0.7225
	(0.0643)	(0.0377)	(0.0770)	(0.0418)	(0.0474)	(0.0587)	(0.1169)	(0.0834)	(0.0997)	(0.1264)	(0.0546)	(0.0265)	(0.0690)	(0.0352)	(0.0409)
Cloudy – Clear	0.6480	0.9024	0.6131	0.0474	0.8072	0.8622	0.9282	0.8525	0.0000	0.7982	0.3065	0.9203	0.2631	0.0331	0.8421
	(0.0493)	(0.0209)	(0.0565)	(0.0295)	(0.035)	(0.0411)	(0.0318)	(0.0478)	(0.1242)	(0.1063)	(0.0395)	(0.0105)	0.0417)	(0.0125)	(0.0183)

6 THE MULTIVARIATE LATENT CLASS CUB MODEL

CONTRIBUTIONS RELATED TO THIS CHAPTER:

Indexed Journals (WoS, Scopus)

[IJ4] Ventura M., Jacques J., Zuccolotto P., *Model-based Clustering of Multivariate Rating Data accounting for Feeling and Uncertainty*, *Under review*

Conference Proceedings

[CP1] Ventura M., Jacques J., Zuccolotto P. (2024) Clustering Multivariate Rating Data within the CUB Framework. Short paper in A. Pollice and P. Mariani (Eds.), 52nd Scientific Meeting of the Italian Statistical Society – Bari, 17–20 June 2024, In press.

International Conferences

[IC5] Ventura M., Jacques J., Zuccolotto P., *Clustering Multivariate Rating Data within the CUB Framework*, 52nd Scientific Meeting of the Italian Statistical Society (SIS), Bari (Italy), 17-20 June 2024.

[IC6] Ventura M., Jacques J., Zuccolotto P., *A Mixture of Multivariate CUB Models for Clustering Rating Data*, 30th Summer Working Group on Model Based Clustering (WMBC), Bertinoro (Italy), 22-26 July 2024.

6.1 INTRODUCTION

Latent Class models, also known as Finite Mixture models, play a crucial role in accounting for unobserved heterogeneity within groups of subjects exhibiting different latent behaviours. These models are instrumental in uncovering hidden structures within data sets across various research domains (Hagenaars and McCutcheon, 2002; Peel and MacLahlan, 2000), and find applications in cluster analysis by providing a flexible framework that can generalize classical geometric clustering methods (Ingrassia, Minotti, and Vittadini, 2012).

Clustering methods are widely used in several research areas for analysing multivariate data and can handle different types of variables. In recent decades, research in several fields such as psychology, marketing, sociology, educational sciences, and behavioural sciences has focused their research efforts on measuring, analyzing, and studying latent constructs like perceptions, sensations, opinions, tastes, and attitudes related to specific topics or items. Information about these hidden traits is commonly obtained by administering questionnaires to people and asking the respondents to

express their opinions. Questionnaires can be designed to collect different types of information through variables expressed on different response scales (Stevens, 1946).

Rating questionnaires are a common tool to investigate respondents' opinions, where individuals evaluate a given subject or item by selecting a score from a predefined set of ordered alternatives on a response scale. The resulting data are typically classified as ordinal categorical variables, which poses statistical challenges that require specialized methods and models tailored to accommodate their unique properties (Agresti, 2010, 2012).

Compared to other types of variables, the ordinal ones have received less attention, especially in the model-based clustering context. It is also common practice among researchers and practitioners to treat ordinal data as numerical, thereby ignoring their categorical nature and using models specifically designed for continuous data analysis, assuming more or less explicitly that the ratings derive from the discretization of an underlying latent continuous variable. Notable approaches include the Binary Ordinal Search (BOS) algorithm (Biernacki and Jacques, 2016) and models assuming ordinal data as a generalization of latent Gaussian distributions (Corneli, Bouveyron, and Latouche, 2020; Grün and Dolnicar, 2016; McParland and Gormley, 2016; Ranalli and Rocci, 2017). Another model-based approach considers logit models for taking into account the presence of covariates (Preedalikit et al., 2024).

Although there are numerous contributions and extensions to the CUB class even in a clustering context (Biasetton et al., 2024), to the best of my knowledge a model-based clustering approach for multivariate data within this framework has not yet been proposed.

This chapter proposes a model-based approach that employs the CUB model to cluster multivariate rating data, utilizing its assumptions and properties. Since the CUB model primarily addresses univariate data, it is extended to handle multivariate scenarios, culminating in a Latent Class extension of Multivariate CUB models, abbreviated as MLC-CUB, specifically designed for clustering multivariate rating data.

The chapter is organized as follows: in Section 6.2 the Multivariate Latent Class CUB is defined and the EM algorithm for the Maximum Likelihood Estimation of the parameters of the model is derived. In Section 6.3, a simulation study is performed to assess the performance of the model, while in Section 6.4 the identifiability of the model is discussed and a tool is proposed for detecting the possible existence of multiple models fitting the same data. Subsequently, in Section 6.5, the model is compared with several mixture models commonly used for analyzing ordinal data. After demonstrating the effectiveness of the proposed model with simulated data, its application to real-world data is presented in Sections 6.6 and 6.7. Finally, the results and potential further developments are discussed in Section 6.8.

6.2 THE MULTIVARIATE LATENT CLASS CUB MODEL

In this section, the Multivariate Latent Class CUB (MLC-CUB) model for clustering multivariate rating data is proposed. In the literature on CUB models there are some proposals of multivariate CUB models (Andreis, Ferrari, et al., 2013; Colombi and Giordano, 2016; Ip and Wu, 2024; Simone, Tutz, and Iannario, 2020), but using these solutions for clustering purposes would lead to an overparametrized model. Therefore, a more parsimonious model is proposed, which consists of a mixture of multivariate CUB models, under the assumption of conditional independence. This assumption,

commonly used for categorical data, states that, conditionally on belonging to cluster k, the J ordinal responses of an individual are independently drawn from J univariate CUB models.

Let $\mathbf{R} = (R_j)_{j=1,\dots,J}$ be a multivariate ordinal variable, where the jth component is an ordinal variable with m_j categories. Let ω_k be the mixing proportion of cluster k, such that $\omega_k > 0$ and $\sum_{k=1}^{K} \omega_k = 1$, and let $\mathbf{Z} : (Z_k)_{k=1,\dots,K}$ be a group indicator variable distributed as a one-order Multinomial distribution, $\mathbf{Z} \sim \mathcal{M}(1; \omega_1, \dots, \omega_k)$, such that $Z_k = 1$ if the preferences come from cluster k, and $Z_k = 0$ otherwise. Given the conditional independence assumption, conditionally on cluster k and being $\mathbf{r} = (r_j)_{j=1,\dots,J}$, the distribution of $\mathbf{R}$ is defined as follows:

$$P(\mathbf{R} = \mathbf{r} \mid Z_k = 1, \boldsymbol{\pi}_k, \boldsymbol{\xi}_k) = \prod_{j=1}^{J} \Big[\pi_{jk} P_B(r_j \mid \xi_{jk}) + (1 - \pi_{jk}) P_U(m_j)\Big], \tag{6.1}$$

where $\boldsymbol{\xi}_k = (\xi_{jk})_{j=1,\dots,J}$ and $\boldsymbol{\pi}_k = (\pi_{jk})_{j=1,\dots,J}$.

Therefore the marginal distribution of $\mathbf{R}$ is then:

$$P(\mathbf{R} = \mathbf{r} \mid \boldsymbol{\xi}, \boldsymbol{\pi}, \boldsymbol{\omega}) = \sum_{k=1}^{K} \omega_k \prod_{j=1}^{J} \Big[\pi_{jk} P_B(r_j \mid \xi_{jk}) + (1 - \pi_{jk}) P_U(m_j)\Big], \tag{6.2}$$

with $\boldsymbol{\xi} = (\xi_{jk})_{j=1,\dots,J;k=1,\dots,K}$, $\boldsymbol{\pi} = (\pi_{jk})_{j=1,\dots,J;k=1,\dots,K}$, $\boldsymbol{\omega} = (\omega_k)_{k=1,\dots,K}$.

Grilli et al. (2014), who proposed the Latent Class CUB (LC-CUB) model, pointed out that including a uniform component in the mixture of univariate CUB models can cause identifiability issues. To address this, they proposed keeping the parameter π constant across the different groups. This approach is reasonable because in real applications various factors influencing uncertainty might be common among subjects from different groups. However, applying this method to the case presented in this chapter would mean assuming that $\pi_{jk} = \pi_j$, implying that every multivariate observation could originate either from a multivariate independent shifted Binomial or a multivariate independent Uniform random variable which, conditionally on cluster k, are defined as $\prod_{j=1}^{J} P_B(\xi_j)$ and $\prod_{j=1}^{J} P_U(m_j)$, respectively. This assumption suggests that respondents are either entirely uncertain or entirely certain about all the questions in the questionnaire. This is a strong and unrealistic assumption that could also weaken an essential foundation of the CUB class of models, which is that respondents' answers are influenced by both their feelings and a degree of uncertainty. To preserve this essential aspect of the CUB framework, the uncertainty parameter π_{jk} is allowed to vary across different clusters and items.

In the paper by Grilli et al. (2014), the LC-CUB model is fitted by using numerical methods. In Appendix E, the EM algorithm for the Maximum Likelihood Estimation of the parameters of the LC-CUB is explained and used as a preliminary study for the development of the EM algorithm for the model presented in this chapter.

6.2.1 *Maximum Likelihood Estimation with EM algorithm*

Let us consider a n-sample $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_n)$ where each $\mathbf{r}_i$ is an observation of the multivariate rating variable $\mathbf{R}$ which is independently and identically distributed, and drawn from an MLC-CUB distribution $P(\mathbf{R} \mid \boldsymbol{\xi}, \boldsymbol{\pi}, \boldsymbol{\omega})$. In addition, let us consider $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iK})$ and $\mathbf{V}_i = (V_{i1}, \dots, V_{iJ})$ which are the latent variables associated

to each subject i. Since the parameters of finite mixture models cannot be estimated in a close form, an effective way to overcome this issue is represented by the EM algorithm (Dempster, Laird, and Rubin, 1977), which allows us to obtain the maximum likelihood estimates of the model parameters by maximizing the complete log-likelihood function. This function is defined by introducing two latent allocation variables in the model: $\mathbf{Z} : (Z_k)_{k=1,\dots,K}$ is a random variable distributed as a one-order Multinomial distribution, $\mathbf{Z} \sim \mathcal{M}(1; \omega_1, \dots, \omega_k)$, such that $Z_{ik} = 1$ if the ith rater preferences come from cluster k, and $Z_{ik} = 0$ otherwise. The second allocation variables $\mathbf{V} : (V_j)_{j=1,\dots,J}$, is dependent on the allocation variable $\mathbf{Z}$ and it is distributed as a Bernoulli with parameter π_{jk}. The observation $V_{ij} = 1$ if the preference of the ith rater for the jth item comes from a shifted Binomial of parameter ξ_{jk}, and $V_{ij} = 0$ if it belongs to a Uniform random variable. Therefore, the complete log-likelihood is defined as follows:

$$\ell_c(\theta \mid \mathbf{R}, \mathbf{Z}, \mathbf{V}) = \sum_{i=1}^{n} \sum_{k=1}^{K} Z_{ik} \Big\{ \ln(\omega_k) + \sum_{j=1}^{J} V_{ij} \big[\ln(\pi_{jk}) + \ln \big[P_B(R_{ij} \mid \xi_{jk}) \big] \big] + \sum_{j=1}^{J} (1 - V_{ij}) \big[\ln(1 - \pi_{jk}) + \ln P_U(m_j) \big] \Big\}, \tag{6.3}$$

where $\theta' = (\xi', \pi', \omega')$.

Starting from the initial set of parameters generated at random, $\theta^{(0)} = (\xi, \pi, \omega)^{(0)}$, the algorithm alternates between an expectation step and a maximization step.

At the t-th iteration, the Expectation step consists of evaluating the expected value of the complete log-likelihood conditioned on the observed data $\mathbf{r}$ and on the parameters $\theta^{(t)}$) computed in the previous step. Therefore, computing the conditional expectation of Equation (6.3), for all $i = 1, \dots, n$, $j = 1, \dots, J$, $k = 1, \dots K$, lead to compute the conditional expectation of Z_{ik} as follows:

$$\mathbb{E}_{Z_{ik} \mid \mathbf{r}_i; \theta^{(t)}} =$$

$$= \frac{\omega_k^{(t)} \prod_j \left[\pi_{jk}^{(t)} P_B(r_{ij} \mid \xi_{jk}^{(t)}) + (1 - \pi_{jk}^{(t)}) P_U(m_j) \right]}{\sum_{k'=1}^{K} \omega_{k'}^{(t)} \prod_j \left[\pi_{jk'}^{(t)} P_B(r_{ij} \mid \xi_{jk'}^{(t)}) + (1 - \pi_{jk'}^{(t)}) P_U(m_j) \right]}$$

$$= \tau_{ik}(\mathbf{r}_i; \theta^{(t)}) = \tau_{ik}^{(t)}.$$

Then, the conditional expectation of the product $Z_{ik} V_{ij}$ is computed:

$$\mathbb{E}_{Z_{ik} V_{ij} \mid r_{ij}; \theta^{(t)}} =$$

$$= \frac{\pi_{jk}^{(t)} P_B(r_{ij} \mid \xi_{jk}^{(t)})}{\pi_{jk}^{(t)} P_B(r_{ij} \mid \xi_{jk}^{(t)}) + (1 - \pi_{jk}^{(t)}) P_U(m_j)} \cdot \tau_{ik}(\mathbf{r}_i; \theta^{(t)}) =$$

$$= \nu_{ik}(r_{ij}; \theta^{(t)}) \cdot \tau_{ik}(\mathbf{r}_i; \theta^{(t)}) = \eta_{ijk}^{(t)}.$$

The maximization step, instead, consists of computing the new maximum likelihood estimates $\theta^{(t+1)}$ of the parameters of the mixture, by maximizing according to θ the function:

$$Q(\theta, \theta^{(t)}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \tau_{ik}^{(t)} \Big\{ \ln(\omega_k) + \sum_{j=1}^{J} \eta_{ijk}^{(t)} \big[\ln(\pi_{jk}) + \ln\big[P_B(r_{ij} \mid \xi_{jk})\big]\big] + \\ + \sum_{j=1}^{J} (1 - \eta_{ijk}^{(t)}) \big[\ln(1 - \pi_{jk}) + \ln P_U(m_j) \big] \Big\} \tag{6.4}$$

By computing the first derivatives of the function $Q(\theta, \theta^{(t)})$, the maximum likelihood estimators are obtained as follow:

$$\pi_{jk}^{(t+1)} = \frac{\sum_{i=1}^{n} \eta_{ijk}^{(t)}}{\sum_{i=1}^{n} \tau_{ik}^{(t)}}, \quad \xi_{jk}^{(t+1)} = \frac{\sum_{i=1}^{n} \eta_{ijk}^{(t)} (m_j - r_{ij})}{\sum_{i=1}^{n} \eta_{ijk}^{(t)} (m_j - 1)}, \quad \omega_k^{(t+1)} = \frac{\sum_{i=1}^{n} \tau_{ik}^{(t)}}{n}. \tag{6.5}$$

The algorithm is stopped when a threshold $\epsilon > 0$ is reached in the relative change of the log-likelihood: $\mid \ell(\boldsymbol{\xi}^{(t+1)}, \boldsymbol{\pi}^{(t+1)}, \boldsymbol{\omega}^{(t+1)}) - \ell(\boldsymbol{\xi}^{(t)}, \boldsymbol{\pi}^{(t)}, \boldsymbol{\omega}^{(t)}) \mid < \epsilon$.

To avoid convergence to a local maximum, the algorithm is run multiple times from different random starting points. Among these iterations, the one yielding the highest log-likelihood is selected as the final outcome of the model. Additional details about the development of the EM algorithm for the MLC-CUB model can be found in Appendix E.

6.2.2 *Selection of the number of clusters*

To determine the optimal number of clusters K, the Bayesian Information Criterion (BIC) (Schwarz, 1978) is employed. This criterion helps in selecting the model that best balances goodness of fit and model complexity. The BIC for a number of clusters K is defined as:

$$\text{BIC}_K = -2 \ln \ell(\theta \mid \mathbf{r}) + (2KJ + K) \ln(n)$$

where $\ell(\theta \mid \mathbf{r})$ is the observed likelihood of the model, that is

$$\ell(\theta \mid \mathbf{r}) = \prod_{i=1}^{n} \Big\{ \sum_{k=1}^{K} \omega_k \prod_{j=1}^{J} \Big[\pi_{jk} P_B(r_j \mid \xi_{jk}) + (1 - \pi_{jk}) P_U(m_j) \Big] \Big\}.$$

To select the best number of clusters, the model must be executed for all $K = 1, \ldots, max_K$, where max_K is the maximum number of clusters that can reasonably be expected to be present in the data. The model with the lowest BIC_K is then selected.

6.3 SIMULATION STUDY

This section aims at evaluating through simulation studies the performance of MLC-CUB model for clustering purposes regarding the influence of sample size and the robustness to noise. The efficiency of BIC in selecting the number of clusters was also evaluated. All the simulations and applications of the model in this work have been performed with R statistical software (R Core Team, 2024).

In all the experiments reported in this section, the EM algorithm is run 10 times with uniform random starting parameters, and with a threshold for the relative change of the log-likelihood equal to 10^{-5}. Among the 10 runs, the one with the maximum log-likelihood has been kept.

The following three experiments were conducted by simulating 100 multivariate samples with sample sizes of $n = \{100, 500, 1000\}$. Each sample was generated from a trivariate ($J = 3$) mixture of three CUB components ($K = 3$), with each dimension having seven categories ($m_j = 7$ for all j). The parameters were selected to ensure low uncertainty since it is the case of most interest in practical applications, and the ξ parameters were sufficiently distinct for each variable. The mixture proportions were set to 0.25 and 0.75. The MLC-CUB parameters used for generating the datasets are presented in Table 6.1.

Table 6.1: Set of parameters chosen for generating the simulated data sets.

	$k = 1$			$k = 2$		
ω		0.25			0.75	
	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
π	0.80	0.90	0.60	0.60	0.80	0.70
ξ	0.30	0.20	0.10	0.70	0.80	0.70

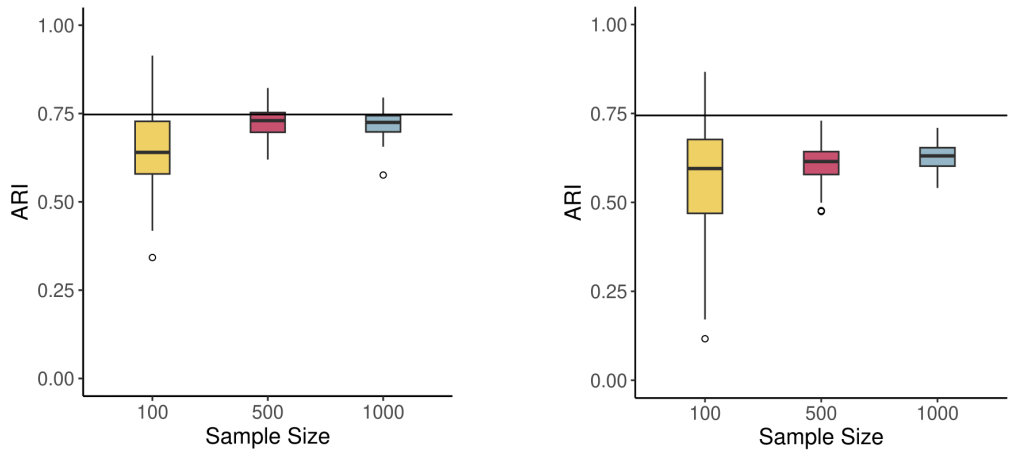
The time required for estimating one MLC-CUB model, with the procedure described above, on a data set with these characteristics and 1000 observations is around 14 minutes (the R code was run on an AMD Ryzen 5 5600U with Radeon Graphics 2.30 GHz and 40GB RAM.). This computing time has been obtained with prototype code which will be optimized and included in the R package that is under development.

6.3.1 *Influence of sample size*

This simulation study aims to investigate the influence of sample size on parameter estimates and the quality of the estimated partitions. This effect is studied both by looking at the distributions of the parameter estimates, at the goodness of fit of the CUB distributions by item on a cluster basis, and at the quality of the estimated partitions which is assessed by using the Adjusted Rand Index (ARI) (Rand, 1971). Values close to 1 indicate that the partition provided by the algorithm closely matches the simulated one, while values close to 0 suggest that the two partitions are no better aligned than random chance would predict. The ARI results are shown in Figure 6.1a, where an optimal ARI value (approximately $\simeq 0.75$) is indicated by a horizontal line. The optimal ARI has been computed by performing one step of the clustering algorithm with the true parameters on a simulated data set with 10000 observations and then comparing the obtained clustering with the real one.

The boxplots show that increasing the sample size enhances the quality of clustering. For a sample size of 100, there is greater variability in the ARI. In contrast, for sample sizes of 500 and 1000, the ARI values are more consistent and closely aligned with the optimal ARI. The median ARI values for sample size equal to 500 and 1000 are almost

equal to the optimal one. Those results suggest that larger sample sizes generally improve clustering accuracy, however, a medium sample size of 500 is still adequate to produce partitions of good quality.



(a) Simulation 1 – Influence of sample size without noise.

(b) Simulation 2 – Influence of sample size with noisy data.

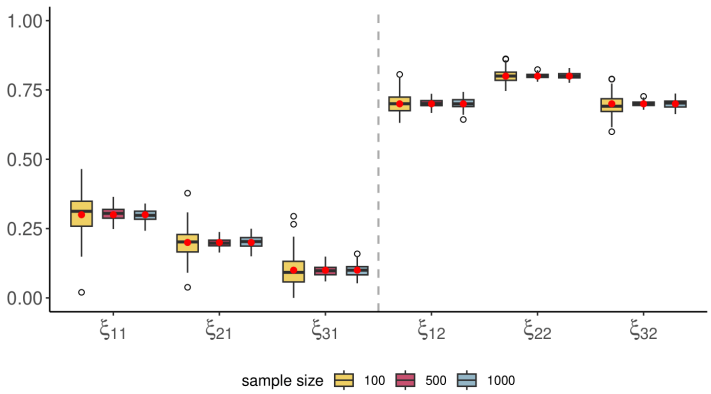
Figure 6.1: Adjusted Rand Index (ARI) distribution on 100 samples. The horizontal line represents the optimal ARI.

In Figures 6.2a, 6.3a, and 6.4a, the estimates of the parameters ξ_{jk}, π_{jk}, and ω_k are shown, respectively. These boxplots clearly illustrate that as the sample size increases, the variability of the estimates decreases. Noteworthy, with a medium and a large sample size, the median values of all parameter distributions coincide with the true parameter values, represented by red circles. It has to be noted that the estimates of the parameter π_{jk} with a small sample size are not so precise. Indeed, in this case, the median of the parameter estimates sometimes does not coincide with the real value of the estimate. It is important to highlight that the distributions of the parameters estimated on small sample size, in this case $n = 100$, exhibit some outliers, indicating that with small sample sizes, the model encounters difficulties in accurately estimating the parameters. Instead, looking at the boxplots of the parameter estimates on data sets of size 500 and 1000, the distributions show lower variability and less outliers. This already happens with $n = 500$, confirming that a medium sample size is enough to obtain good parameter estimates.

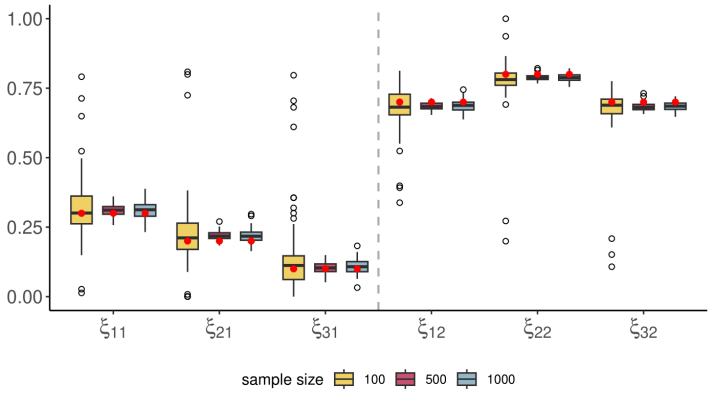
The results reported in Table 6.2 confirm that the quality of the estimates gets better with increasing sample size, indeed, the Diss index decreases with increasing sample size, meaning that the observed distribution is closer to the theoretical one.

6.3.2 *Robustness to noise*

This section aims to evaluate the performance and behaviour of the model when it is fitted on noisy data. Typically, noise is simulated by adding data from a Uniform distribution to the dataset. However, in this specific case, introducing noise through a Uniform random variable is not possible, as it is a component of the mixture that characterizes the CUB model since it models the uncertainty component. Therefore,



(a) Simulation 1 – Effect of sample size without noise.

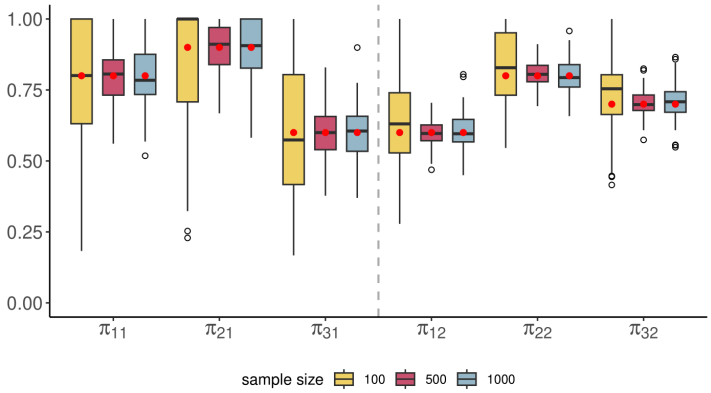


(b) Simulation 2 – Effect of sample size with noisy data ($\phi = 0.1$).

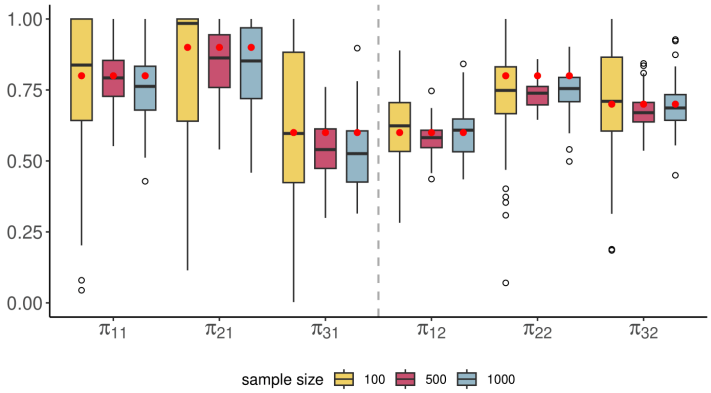
Figure 6.2: Boxplots representing the estimates of the parameter ξ_{jk} obtained on 100 data sets. The real values of the parameters are represented by the red circle.

Table 6.2: Average *Diss* index of CUB models for each item on a cluster basis. The average dissimilarity refers to the *Diss* indexes computed at each replication between the observed and the theoretical distributions for each item in each cluster. The standard deviation is reported in parentheses.

	n = 100		**n = 500**		**n = 1000**	
	Cluster 1	**Cluster 2**	**Cluster 1**	**Cluster 2**	**Cluster 1**	**Cluster 2**
j = 1	0.1341	0.0893	0.0725	0.0403	0.0581	0.0301
	(0.0563)	(0.0354)	(0.0253)	(0.0143)	(0.0188)	(0.0096)
j = 2	0.1230	0.0775	0.0756	0.0380	0.0661	0.0306
	(0.0538)	(0.0356)	(0.0247)	(0.0141)	(0.0201)	(0.0111)
j = 3	0.1204	0.0868	0.0656	0.0368	0.0455	0.0293
	(0.0455)	(0.0308)	(0.0240)	(0.0142)	(0.0165)	(0.0120)

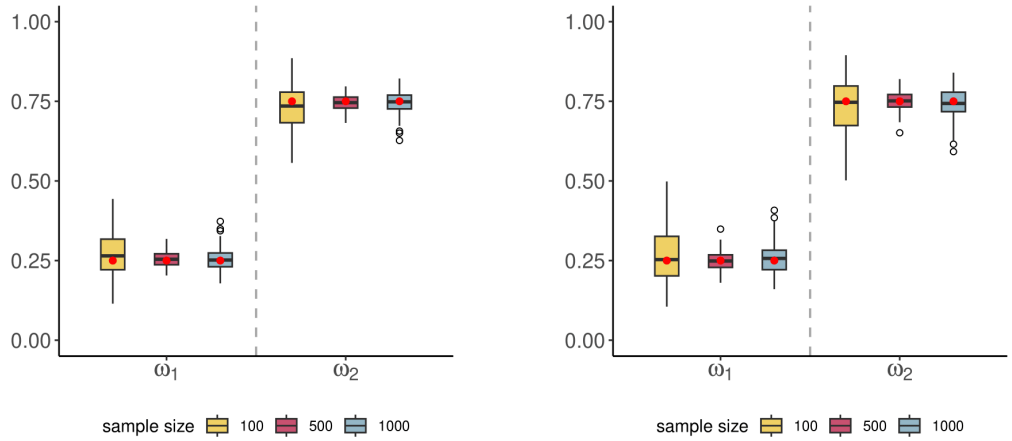


(a) Simulation 1 – Effect of sample size without noise.



(b) Simulation 2 – Effect of sample size with noisy data ($\phi = 0.1$).

Figure 6.3: Boxplots representing the estimates of the parameter π_{jk} obtained on 100 data sets. The real values of the parameters are represented by the red circle.



(a) Simulation 1 – Effect of sample size without noise.

(b) Simulation 2 – Effect of sample size with noisy data ($\phi = 0.1$).

Figure 6.4: Boxplots representing the estimates of the parameter ω_k obtained on 100 data sets. The real values of the parameters are represented by the red circle.

noise was introduced in the simulated data by replacing some observations from a CUB component with observations from a discretized Gaussian distribution.

Datasets were simulated with a noise proportion of 10% ($\phi = 0.1$). As in the previous simulations, the results were evaluated using the ARI (see Figure 6.1b) and by examining the distributions of the parameter estimates (see Figures 6.2b, 6.3b, and 6.4b). It is noteworthy that the variability of the ARI decreases with sample size, and the median value of the ARI approaches the optimal value as the sample size increases; however, in presence of noise the median value of ARI never reaches the optimal value. By looking at the distributions of the parameter estimates, it can be noticed that the parameter π_{jk} remains the most challenging to estimate as demonstrated by its high variability. Interestingly, for the parameters ξ_{jk} and π_{jk}, even though the variability of the estimates decreases with increasing sample size, sometimes the median values of the parameter estimates do not coincide with the real ones, even though the deviation from the real value is very low. The only exception is the parameter ω_k, whose distribution is not much affected by the noise.

Therefore, as expected, the quality of clustering and the accuracy of the parameter estimates decreases as the amount of noise increases, indicating that the model is sensitive to noise.

6.3.3 *Selection of number of clusters*

This section aims at evaluating the ability of the model to identify the correct number of clusters.To do this, the BIC was used, as described in section 6.2.2, in a setting consisting of 6 different scenarios obtained by varying either the sample size $n \in 100, 500, 1000$ or the amount of noise $\phi \in 0, 0.1$. For each scenario, 100 datasets were drawn. The results are reported in Table 6.3.

Looking at the results of the model selection, it can be concluded that the performance of the model in selecting the correct number of clusters ($K = 2$) is overall good,

Table 6.3: Frequency of selection of the number of clusters K as the best number of clusters among 100 simulated data sets, for increasing number of observations and increasing amount of noise. The maximum number of clusters that has been considered is $max_K = 5$ and the actual number of clusters is $K = 2$.

	Scenario $\phi = 0$					Scenario $\phi = 0.1$				
n/K	1	2	3	4	5	1	2	3	4	5
100	18	82	-	-	-	56	44	-	-	-
500	-	100	-	-	-	-	100	-	-	-
1000	-	100	-	-	-	-	100	-	-	-

especially when there is no noise. Indeed, it can be noticed that only in the case of a small sample size ($n = 100$) the model doesn't select the right number of clusters every time both with and without noise. Indeed, in the case without noise, the model selects the right number of clusters in the 82% of times, while this percentage decreases to 44% of times in the data sets with the addition of noise.

However, two facts have to be noticed: the first is that with a higher sample size the model always selects the correct number of clusters; the second fact that has to be noticed is that a medium sample size ($n = 500$) is already enough to select the correct number of clusters in all the cases, both with and without noise.

6.4 MODEL IDENTIFIABILITY

As already discussed in section 6.2, the parameter π is allowed to vary both across variables and clusters. This inevitably leads to identifiability issues due to the uniform component since it is equal for every cluster, therefore when one observation r_{ij} comes from the Uniform random variable (*i.e.* $V_{ij} = 0$), it is impossible to determine to which cluster it belongs (Titterington, Smith, and Makov, 1985). However, in the multivariate case, the probability of encountering this problem is lower than in the univariate case, since it should be enough that one observation r_{ij} comes from a shifted Binomial to determine to which cluster it belongs. Additionally, it is hypothesized that identifiability issues are less likely when the values of π_{jk} are high, indicating low uncertainty. This problem is extensive and complex, and will thus be the subject of further research. Nevertheless, the identifiability issue also affects real data analysis. Therefore, an empirical procedure is proposed to detect identifiability problems in real datasets.

The distribution of the Adjusted Rand Index (ARI) computed between the partition obtained on the original sample and the partition obtained on each bootstrap sample is an effective tool for detecting models affected by identifiability issues. To illustrate the procedure, three datasets with a sample size of $n = 1000$ were generated to ensure that the sample size did not affect the estimation:

- **Dataset 1 – No identifiability problem**: To observe the behavior of an identifiable model, a dataset identical to the one used in previous simulations was generated, as it was reasonably confident that this model does not exhibit identifiability problems, given that the parameter estimates are accurate. This dataset consisted of three ordinal variables with 7 categories, following an MLC-CUB distribution with parameters as shown in Table 6.1.

- **Dataset 2 – Identifiability problem with two clusters**: To investigate whether the value of π_{jk} influences the identifiability of the model, a dataset with three ordinal variables was generated, each variable with 7 categories, following a two-component MLC-CUB distribution characterized by low values of the parameter π_{jk}. The parameters used are specified in Table 6.4.
- **Dataset 3 – Identifiability problem with three clusters**: To examine whether the number of clusters affects identifiability problems, a dataset with three ordinal variables was generated, each with 7 categories, following a three-component MLC-CUB distribution characterized by low values of the parameter π_{jk}. The parameters used are specified in Table 6.5.

Table 6.4: Dataset 2 – Set of parameters chosen for generating the simulated data.

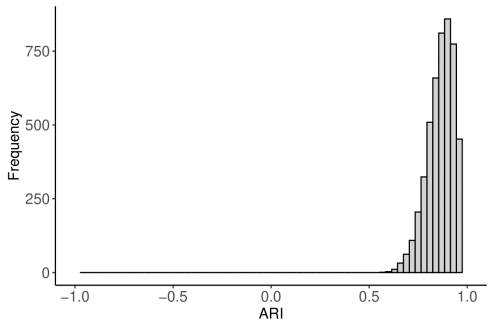
	$k=1$			$k=2$		
ω	0.25			0.75		
	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
π	0.20	0.10	0.25	0.30	0.35	0.40
ξ	0.30	0.20	0.10	0.70	0.80	0.70

Table 6.5: Dataset 3 – Set of parameters chosen for generating the simulated data.

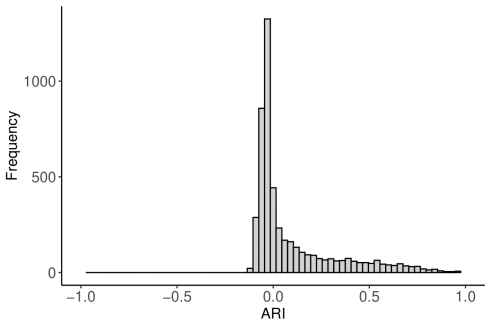
	$k=1$			$k=2$			$k=3$		
ω	0.20			0.50			0.30		
	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
π	0.10	0.10	0.35	0.20	0.25	0.10	0.15	0.17	0.22
ξ	0.50	0.20	0.10	0.70	0.80	0.70	0.10	0.60	0.90

Each dataset was bootstrapped, and an MLC-CUB model was estimated for each bootstrap sample. The Adjusted Rand Index (ARI) was then computed between the partition obtained from the original sample and the partition obtained from each bootstrap sample. The results of this procedure demonstrate that identifiability problems can be identified by examining the ARI distributions, which are shown in Figure 6.1.

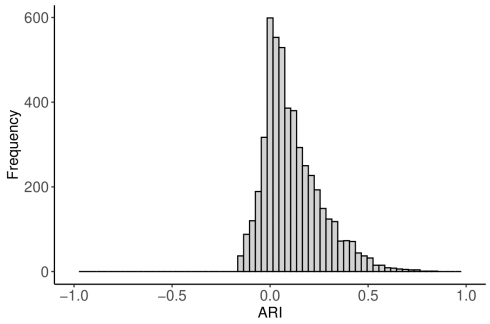
Indeed, when there are no identifiability problems the distribution of the ARI is left-skewed and close to 1 (Figure 6.5a), while when there are identifiability problems the distribution of ARI is right-skewed and far from 1 (Figures 6.5b and 6.5c). Although this study does not establish whether the degree of skewness can measure the severity of identifiability problems, it does suggest that a left-skewed ARI distribution indicates stable results. Therefore, it is reasonable to infer that there are no identifiability problems in the model fitted to a specific dataset when the ARI distribution is left-skewed and has values close to 1. Notably, when the same experiment was conducted using a Gaussian Mixture Model (GMM) (Banfield and Raftery, 1993), which is known to have no identifiability issues (if label switching is omitted, as it has no impact on ARI), the resulting distribution, shown in Figure 6.6a, was very close to the one in Figure 6.5a.



(a) Dataset 1 – Distribution of ARI for high values of π_{jk} and two clusters



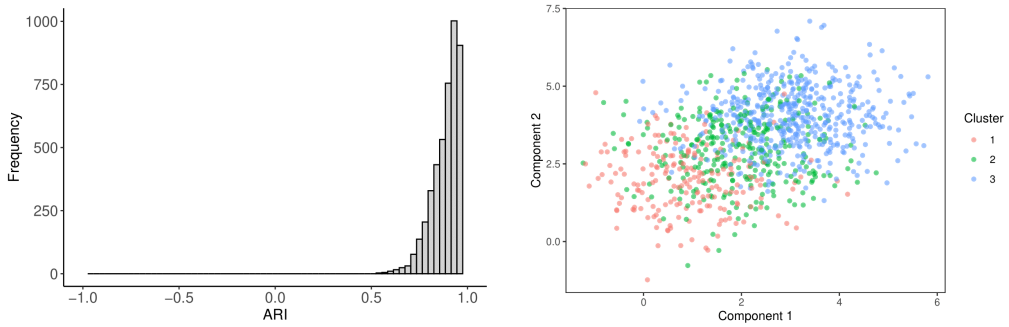
(b) Dataset 2 – Distribution of ARI for low values of π_{jk} and two clusters



(c) Dataset 3 – Distribution of ARI for low values of π_{jk} and three clusters

Figure 6.5: Distribution of the ARI obtained as pairwise comparison of the classifications of 100 models fitted on three different bootstrapped data sets.

Note that the GMM model has been estimated using the `Mclust` package (Scrucca et al., 2016). Such ARI distribution, instead, is due to the presence of clusters that are overlapped, as it can be noticed by looking at Figure 6.6b.



(a) Distribution of ARI of the bootstrapped GMM.

(b) Representation of the clusters on the first two dimensions.

Figure 6.6: Results of the clustering performed with GMM.

6.5 COMPARISON WITH STATE OF THE ART METHODS

In the literature, several models have been developed specifically to model ordinal data in a model-based setting, which are competitors to the proposed model. The first competitor is the BOS model (Biernacki and Jacques, 2016), which models ordinal data using a Binary Ordinal Search Algorithm. This model can be fitted in R using the dedicated package `ordinalClust` (Selosse, Jacques, and Biernacki, 2021). Another competitor is the clustMD model (McParland and Gormley, 2016), designed for clustering mixed data. However, it can also be used for modelling only ordinal data, which are assumed to be manifestations of a continuous latent variable. A similar latent variable approach was adopted by Corneli, Bouveyron, and Latouche (2020), who proposed a mixture model for co-clustering ordinal data. This model can be fitted using the dedicated package `ordinalLBM`, assuming that the column cluster is singular.

It is important to note that the estimation performances of the BOS model are influenced by the number of categories of ordinal variables. Indeed, the model is efficient for a number of categories up to $m = 8$ (Biernacki and Jacques, 2016). Moreover, the BOS model doesn't manage variables with different numbers of categories. In this comparison study, the BOS model and the ClustMD model could not be considered due to convergence problems with the type of data generated.

In this comparative study, the proposed MLC-CUB model is evaluated against three alternative models: the Ordinal Latent Block Model (OLBM) (Corneli, Bouveyron, and Latouche, 2020), the Multinomial Mixture Model (MMM) (Everitt, 1984; Lebret et al., 2015), and the Gaussian Mixture Model (GMM) (Banfield and Raftery, 1993). The OLBM is tailored for analyzing ordinal data by focusing on latent block modelling to uncover underlying group structures in the data set. In this case, it is fitted with the package `OrdinalLBM` (Corneli, Bouveyron, and Latouche, 2020) by assuming one column cluster. The MMM treats ordinal data as nominal by ignoring their ordinal nature and clustering them using mixtures of multivariate multinomials. It is fitted using the `Rmixmod` package (Lebret et al., 2015). In contrast, the GMM treats ordinal data

as continuous, assuming they are generated from a mixture of multivariate Gaussian distributions. The GMM is fitted using the `Mclust` package (Scrucca et al., 2016).

To perform the comparison study, 100 datasets were generated with a sample size of $n = 1000$ using the same simulation setting described in Section 6.3. The MLC-CUB model, the OLBM, the GMM, and the MMM were fitted to each dataset, with the number of clusters fixed at $K = 2$. The clustering performances of each model were compared using the ARI, as shown in Figure 6.7.

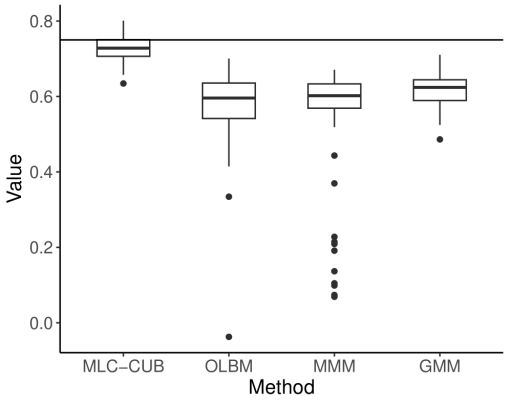


Figure 6.7: ARI for Gaussian Mixture Model, OrdinalLBM Model, Multinomial Model and MLC-CUB Model. The horizontal line represents the optimal ARI.

The results show that the model outperforms the others, the median ARI value approaches the optimal ARI. The ARI of the competing models is on average the same, but lower than the median ARI of MLC-CUB. Moreover, the ARI of OLBM and MMM show high variability.

6.6 Case study: evaluation of the university orientation service

In this section, a case study is presented. The data comes from the data set `univer`, which is publicly available in the package `CUB` (Iannario, Piccolo, and Simone, 2020). The data set contains the answers from a sample survey on students' evaluation of the Orientation services of the University of Naples Federico II. The data were collected in 2002 and consisted of 2179 observations. Participants were asked to express their level of satisfaction on a seven-point Likert scale (1 = "very unsatisfied", 7 = "extremely satisfied") on the following elements of evaluation:

(Q1) Level of satisfaction about the acquired information

(Q2) Level of satisfaction about the willingness of the staff

(Q3) Level of satisfaction about the opening hours

(Q4) Level of satisfaction about the competence of the staff

(Q5) Level of global satisfaction

The MLC-CUB model was fitted on the `univer` dataset, considering the five variables described above. The results of the model were then compared with those of the competitor models using the ARI and the BIC. The results are reported in Table 6.6.

Table 6.6: University data set – Values of the BIC for the MLC-CUB and its competitors for 1 to 7 clusters. The best ones are highlighted in bold.

Model	K = 1	K = 2	K = 3	K = 4	K = 5	K = 6	K = 7
MLC-CUB	33891.10	30253.74	28769.16	28301.50	**28228.88**	28251.68	28289.33
OLBM	34977.40	31948.27	30309.65	29718.57	29173.19	**28676.43**	29022.91
MMM	33796.73	29705.67	28223.80	27471.74	**27292.52**	27370.96	27561.38
GMM	31042.78	30670.30	29754.40	29286.27	29099.78	**29069.44**	29153.94

The results show that both the MLC-CUB and MMM models identify 5 clusters in the data, whereas the GMM and OLBM models identify 6 clusters. Among all the models, the Multinomial Mixture Model demonstrates the best performance in terms of BIC, having the lowest BIC value. The MLC-CUB model follows with the second-best BIC value.

The optimal MLC-CUB model is characterised by 5 components, with the estimated parameters for each cluster illustrated in Figure 6.8. This plot is designed in line with the typical graphical representation of the basic CUB model, which is reported on a Cartesian plane that represents the parameter space of the CUB (D'Elia and Piccolo, 2005). In the representation considered here, each univariate CUB model which constitutes the MLC-CUB model is plotted on the same Cartesian plane, where the x-axis reports the level of uncertainty and the y-axis reports the level of feeling.

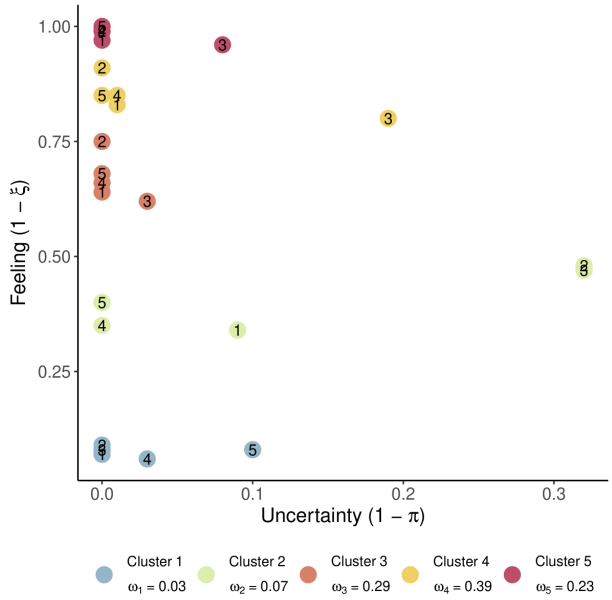


Figure 6.8: University data set – Representation of the model parameters which characterize each cluster of the university data. Each number represents the associated question in the questionnaire.

Table 6.7: University data set – Diss index for each item of the questionnaire at cluster basis.

Item	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Q1	0.0819	0.6172	0.2560	0.1629	0.0189
Q2	0.0174	0.5092	0.1511	0.0930	0.0208
Q3	0.0229	0.2868	0.1620	0.0467	0.0369
Q4	0.0803	0.6136	0.2073	0.1880	0.0158
Q5	0.1168	0.6178	0.2894	0.3098	0.0089

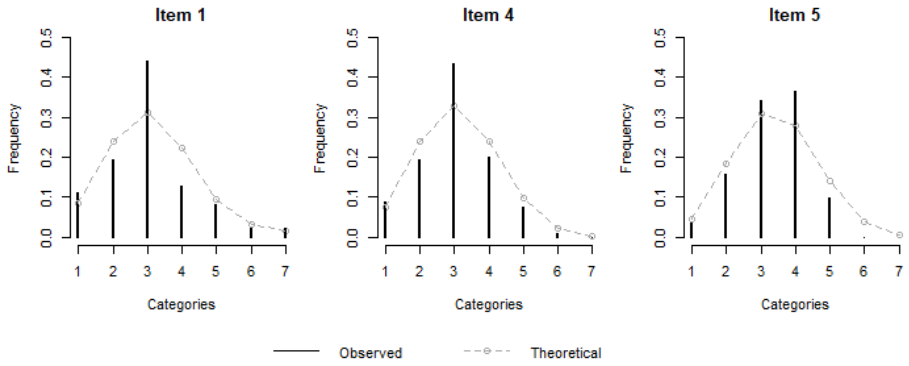


Figure 6.9: University data set - Observed and theoretical frequencies for items 1, 4, and 5 in cluster 2.

By examining the results, it can be concluded that three main clusters (clusters 3, 4, 5) are present, all characterized by low uncertainty and generally high levels of satisfaction. Among these three clusters, cluster 3 has the lowest level of satisfaction, although students in this group still exhibit medium-high satisfaction with university services.

Clusters 4 and 5 contain students with the highest levels of satisfaction. In particular, cluster 5 includes students who are completely satisfied with the university's orientation services, comprising 23% of the sample analyzed.

Clusters 1 and 2 are smaller compared to the others (see the legend of Figure 6.8), but they are worth noting. Cluster 1 includes students who are not satisfied at all, while cluster 2 consists of students with medium-low levels of satisfaction, indicating that they are not completely satisfied with the university's orientation services. Additionally, students in cluster 2 exhibit higher levels of uncertainty in their responses to questions 2 (willingness) and 3 (office hours).

The goodness of fit for the CUB distributions, evaluated on a cluster basis for each item, is presented in Table 6.7. In certain clusters, the Diss index exhibits low values, indicating a strong fit. However, for items in clusters 2 and 3, the fit of the CUB distribution is poorer, as highlighted by high Diss index values across all items in these clusters and by Figure 6.9 where the distributions of the three items with the highest Diss index are reported. These distributions show a pick for a specific category, that in the CUB framework is called *shelter option* and in the univariate setting is modeled by the CUB model with shelter proposed by Iannario (2012). To assess the stability of the results and whether there were identifiability issues, the procedure described

in section 6.4 has been performed. The distribution of the pairwise ARI between the classifications fitted on 100 bootstrapped data sets is reported in Figure 6.10.

The shape of the distribution is closely similar to the one reported in Figure 6.5a, suggesting that there are no identifiability problems in the estimated model.

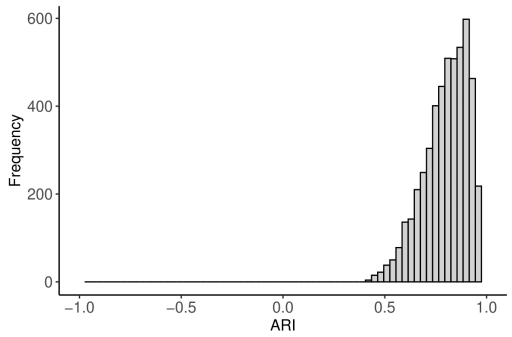


Figure 6.10: University data set – Distribution of the pairwise ARI indexes between the classification fitted on bootstrapped data.

6.7 CASE STUDY: EVALUATION OF THE SERVICES IN PUBLIC KINDERGARTENS

In this section, an application of the model to real data is presented, focusing on the evaluation of services in public kindergartens (Brentari, Carpita, and Zuccolotto, 2006). The data were collected during a study sponsored by the Municipality of Brescia, Italy, in 2004. The purpose of the survey was to evaluate the quality of service provided by the 22 municipal kindergartens. The questionnaire, divided into sections, consisted of questions where respondents expressed their agreement or disagreement on a four-point Likert scale, where 1 corresponded to "strongly disagree" and 4 corresponded to "strongly agree." There was also a fifth option, "Don't know" (DK), for respondents who were unsure of their judgment. The aspects on which parents were asked to provide their opinions included key characteristics of the service offered by the schools: the educational area, the relational area, and the organizational area. A total of 1337 responses were obtained. After deleting the observations that contained a DK answer or missing values, a dataset with 883 observations was obtained.

The focus was placed on the evaluation of the organizational area of the schools, assessed through the following questions:

(Q1) Do you think the school environment is suitable for the activities conducted there?

(Q2) Do you find the role of school aides (parents, cooks, etc.) to be adequate?

(Q3) Are you satisfied with the quality and variety offered by the school cafeteria?

(Q4) Are the reception hours for children adequate for your needs?

(Q5) Do you consider the management of school entry and exit adequate?

(Q6) Are you satisfied with the level of supervision provided by the school for the children?

(Q7) Are you interested in the initiatives promoted for parents?

(Q8) Do you find parental participation in school activities adequate?

The obtained results have been compared with those obtained using other state-of-the-art methods described in section 6.5. The optimal number of clusters for each method was determined using the BIC. The results are reported in Table 6.8. Three methods suggest that the best number of clusters is 3. However, one of those clusters is very small and does not add any valuable information to the interpretation of the results. Moreover, the difference in BIC between two and three clusters is low. Therefore, to enhance the interpretability of the results, the model with 2 clusters was chosen for presentation.

Table 6.8: Kindergarten data set – Values of the BIC for the MLC-CUB and its competitors for 1 to 7 clusters. The best ones are highlighted in bold.

Model	K = 1	K = 2	K = 3	K = 4	K = 5	K = 6	K = 7
MLC-CUB	13350.31	12767.05	**12766.37**	12809.55	12888.92	12981.69	13070.47
OLBM	14084.18	13795.02	13786.37	**13773.98**	13859.51	13913.55	13909.78
MMM	13371.38	12714.84	**12693.21**	12741.10	12817.28	12909.22	13031.63
GMM	14206.66	13588.47	**12526.45**	12641.75	12065.22	11620.49	10466.07

The results of MLC-CUB are reported in Figure 6.11. By examining the results, it can be concluded that there are two clusters of almost the same size (see the legend of Figure 6.11). The first one is characterized by a higher level of feeling and, therefore, satisfaction with respect to the organizational level of the kindergarten system in Brescia. Both clusters show slightly lower levels of satisfaction in response to questions 7 and 8, which pertain to the level of parental participation and the parent-school relationship. In particular, both clusters exhibit a relatively low level of uncertainty.

The goodness of fit of CUB models for each item at cluster basis has been measured through the Diss index. The results are reported in Table 6.9 and they shows a good fitting, especially for cluster 1.

In this application, the stability of the results was also assessed by fitting the model on 100 bootstrapped datasets. The identifiability of the estimated model was checked by examining the distribution of the pairwise ARI (see Figure 6.12). The distribution has a shape similar to that reported in Figure 6.5a, suggesting that the parameters estimates are stable and the estimated model is not compromised by identifiability problems.

Table 6.9: Kindergarten data set – Diss index for each item of the questionnaire at cluster basis.

Item	**Cluster 1**	**Cluster 2**	**Item**	**Cluster 1**	**Cluster 2**
Q1	0.0062	0.0862	Q5	0.0150	0.0991
Q2	0.0093	0.1467	Q6	0.0104	0.1394
Q3	0.0158	0.1021	Q7	0.0064	0.0138
Q4	0.0077	0.0515	Q8	0.0542	0.0928

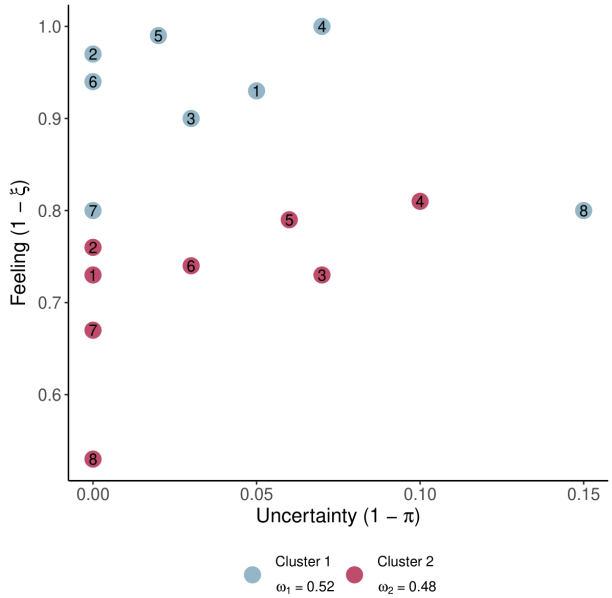


Figure 6.11: Kindergarten data set – Representation of the model parameters which characterize each cluster of the Kindergarten data. Each number represents the associated question in the questionnaire.

6.8 DISCUSSION

A mixture model for clustering multivariate rating data, which belongs to the CUB class of models, was presented. Under the assumption of conditional independence, the model is a mixture of multivariate CUB models. This allows us to capture the latent components that characterize the CUB framework: feeling and uncertainty.

An EM algorithm for the Maximum Likelihood Estimation of the model parameters has been defined and implemented. The performance of the algorithm was evaluated through simulation studies. The model demonstrated good performance when the level of uncertainty was low (i.e., the weight of the uniform component was low). However, it exhibited some drawbacks related to identifiability issues when the uncertainty was high. This problem was explored, and an empirical tool for identifying it was proposed. However, this issue will be the focus of further and more in-depth studies to better understand the conditions under which this problem arises and to determine which types of constraints on model parameters could be useful for ensuring identifiability.

Applications to real data demonstrated that the identifiability problem does not affect the model's performance in practice. In addition, these applications highlighted the valuable interpretability of the results in terms of understanding the latent traits that influence how people respond to questionnaires.

In the future, the model is planned to be enhanced by incorporating the possibility of handling missing data imputation within the EM algorithm. Additionally, the "Don't know" option in rating data will be addressed, following the approach proposed by Manisera and Zuccolotto (2014b), and the option to account for the shelter choice will be added, following the proposal by Iannario (2012).

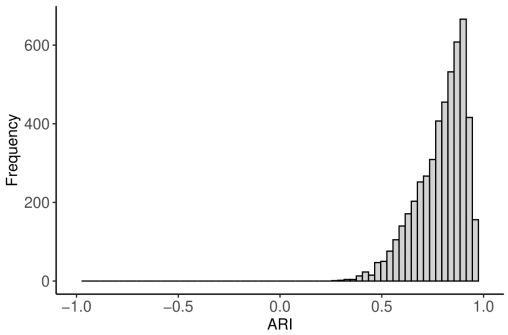


Figure 6.12: Kindergarten data set – Distribution of the pairwise ARI indexes between the classification fitted on boostrapped data.

The model is also planned to be extended in various ways. This includes incorporating other models within the CUB class to capture specific aspects and traits of response styles, such as using the discretized Beta distribution (Simone and Tutz, 2018). Furthermore, while independence among ordinal variables was assumed in this work, there are plans to relax this assumption using copulas, as proposed for the bivariate case by Corduas et al. (2011) and Andreis and Ferrari (2013). Another extension involves developing a multilevel Latent Class model for clustering multilevel rating data. Furthermore, a dedicated R package for this model is currently under development.

Part II

APPENDICES

The second part of this thesis shows my scientific contributions to the field of statistical modeling and analysis, specifically focusing on Mixture Models for the analysis of rating data.

The first work presented in this section constitutes a methodological advancement in the case of the CUM model. As explained in the first part of the thesis, the probability mass function of the CUM model poses computational challenges, since it needs the development of an EM algorithm for parameter estimation according to the number of ordinal variable categories. In this contribution, a CUM model tailored for analyzing five-point Semantic Differential scales is developed. In addition, a comprehensive simulation study was conducted to evaluate the performance and behavior of the model under various conditions.

The second scientific contribution refers to an analytical investigation into the conditions of the equivalence between CUB and CUM models, both of which belong to the same family of models. This chapter aims to study the theoretical relations between these models, providing valuable insights for their interpretation.

The third chapter shifts the focus to an empirical application of these methodologies within the"Data Science for Brescia - Arts and Cultural Place" project. Through the use of multi-point Semantic Differential scales, the chapter explores visitors' perceptions of the Art Gallery of Brescia, uncovering intriguing insights into the multisensory nature of the museum experience.

Lastly, the fourth scientific contribution introduces a novel model within the CUB class. This innovative approach regards the development of a mixture of multivariate CUB models tailored for model-based clustering of multivariate rating data. Through a simulation study, the chapter evaluates the behavior and performance of the model in various scenarios. Additionally, real-world applications are shown to highlight the practical utility of this model in real-world contexts.

A

FINITE MIXTURE MODELS AND EM ALGORITHM

The CUB model is a mixture distribution of a shifted Binomial and a discrete Uniform distribution. Therefore, this model belongs to the wider class of Finite Mixture models (FMM) (McLachlan, Lee, and Rathnayake, 2019). FMMs have been widely used in the last years to model in a computationally convenient way complex distributions of data, with applications to different fields like agriculture, biology, physics, social sciences and psychology, economics, and engineering. Finite Mixture modeling is especially used to provide descriptive models for distributions where a single component is considered inappropriate. However, FMMs have been introduced and integrated in various statistical techniques such as cluster and latent class analyses, discriminant analysis, image analysis, and survival analysis.

In this chapter, the Finite Mixture Models are introduced and defined. Subsequently, the Expectation Maximization algorithm is introduced for the Maximum Likelihood Estimation of FMMs.

A.1 FINITE MIXTURE MODELS

Finite Mixture models assume that the data are generated from a mixture of several probability distributions, where each distribution corresponds to a different subpopulation or latent class within the data. The mixture model is called "finite" because it involves a finite number of components, each represented by a distinct probability distribution. The fundamental idea behind FMMs is to model the overall data distribution as a weighted sum of these component distributions, capturing the inherent diversity within the data.

An FMM assumes that the overall Probability Density Function of the observed data represented by the J-dimensional random vector $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is a weighted sum of K densities, each associated with a specific subpopulation. The mode can be formally defined as:

$$f(\mathbf{x}; \boldsymbol{\theta}) = \sum_{k=1}^{K} \omega_k f_k(\mathbf{x}; \boldsymbol{\psi}_i), \tag{A.1}$$

where $\boldsymbol{\theta} = (\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K; \omega_1, \ldots, \omega_k)$ denotes the vector of unknown parameters. The mixing proportions ω_k are defined such that $\omega_k > 0$ and $\sum_{k=1}^{K} = 1$, while $f_k(\mathbf{x}, \boldsymbol{\psi}_k)$ are the densities of components, each characterized by a specific set of parameters $\boldsymbol{\psi}_k$.

There are several well-known FMMs such as those whose component densities belong to the same parametric family, such as Gaussian Mixture Models (GMM) (Celeux and Govaert, 1995), the t-distribution mixture Mixture Models (Peel and McLachlan, 2000), and the skew normal Mixture Models (Lin, Lee, and Yen, 2007). Other types of FMMs assume that the densities of the components come from different parametric families (Coretto and Hennig, 2011). Finally, in non-parametric FMMs, no assumptions are made about the form of the function (Benaglia, Chauveau, and Hunter, 2009).

A.2 EXPECTATION MAXIMIZATION ALGORITHM

The inference of FMMs originally proposed by Pearson **metti ref** was the method of moments, which has been popular until the advent of the Expectation Maximization (EM) algorithm that facilitated the ML estimates of FMMs The fitting of Finite Mixtures, indeed, is a well-known problem, since it is impossible to maximize the log-likelihood function of equation (A.1) in a closed form.

To define the EM algorithm for the Maximum Likelihood Estimation of the parameters of an FMM, the latent variable $\mathbf{Z}$ distributed as a one-order Multinomial distribution has to be introduced. The latent variable $\mathbf{Z}$ is encoded such that $z_{ik} = 1$ if $\mathbf{x}_i$ belongs to the kth component, and $z_{ik} = 0$ otherwise. The complete-data log-likelihood function $\ell_c(\theta; \mathbf{X}, \mathbf{Z})$ can be written as follows:

$$\begin{aligned}
\ell_c(\theta; \mathbf{X}, \mathbf{Z}) &= \log \prod_{i=1}^{n} f(\mathbf{x}_i, z_i; \theta) \\
&= \log \prod_{i=1}^{n} \prod_{k=1}^{K} \Big[\omega_k f_k(\mathbf{x}_i; \boldsymbol{\psi}_k) \Big]^{z_{ik}} \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \Big\{ \ln \omega_k + \ln f_k(\mathbf{x}_i; \boldsymbol{\psi}_k) \Big\}
\end{aligned} \tag{A.2}$$

The EM algorithm consists of two main steps that are iterated until convergence: the Expectation and the Maximization steps. The value of the parameters vector θ estimated at each t-th iteration is denoted as $\theta^{(t)}$.

E-STEP This step consists in computing the expectation of $\ell_c(\theta, \mathbf{X}, \mathbf{Z})$ conditional on the data $\mathbf{X}$ using the parameter vector $\theta^{(t)}$ of the previous iteration. This expectation is denoted by $Q(\theta; \theta^{(t)})$:

$$\begin{aligned}
Q(\theta; \theta^{(t)}) &= \mathbb{E}[\ell_c(\theta, \mathbf{X}, \mathbf{Z}) \mid \mathbf{X}; \theta^{(t)}] \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} \mathbb{E}[z_{ik} \mid \mathbf{X}; \theta^{(t)}] \{ \ln \omega_k^{(t)} + \ln f_k(\mathbf{x}_i; \boldsymbol{\psi}_k^{(t)}) \} \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} \eta_{ik}^{(t)} \{ \ln \omega_k^{(t)} + \ln f_k(\mathbf{x}_i; \boldsymbol{\psi}_k^{(t)}) \},
\end{aligned} \tag{A.3}$$

where

$$\eta_{ik}^{(t)} = \mathbb{E}[z_{ik} \mid \mathbf{X}; \theta^{(t)}] = \frac{\omega_k^{(t)} f_k(\mathbf{x}_i; \boldsymbol{\psi}_k^{(t)})}{\sum_{k'=1}^{K} \omega_{k'}^{(t)} f_{k'}(\mathbf{x}_i; \boldsymbol{\psi}_{k'}^{(t)})}, \tag{A.4}$$

where $\eta_{ik}^{(t)}$ is the posterior probability that the observation $\mathbf{x}_i$ comes from the k-th component at iteration t.

M-STEP In this step, the estimate of θ that maximizes the function $Q(\theta; \theta^{(t)})$ is computed.

Considering that $Q(\theta, \theta^{(t)})$ can be written as:

$$Q(\theta^{(t+1)}; \theta^{(t)}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \eta_{ik}^{(t)} \ln \omega_k^{(t)} + \sum_{i=1}^{n} \sum_{k=1}^{K} \eta_{ik}^{(t)} f_k(x_i; \psi_k^{(t)}), \tag{A.5}$$

the mixing proportions ω and the parameters ψ can be estimated independently. Specifically, the estimate of the parameters vector ψ can be obtained by computing the first order derivative of the second component of the function $Q(\psi; \psi^{(t)})$ with respect to the parameters of the vector ψ:

$$\sum_{i=1}^{n} \sum_{k=1}^{K} \eta_{ik}^{(t)} \frac{\partial \ln f_k(x_i; \psi_k)}{\partial \psi}. \tag{A.6}$$

The EM algorithm is a powerful tool for estimating the parameters of an FMM because the solution to Equation A.6 is often straightforward to obtain, as it can typically be solved in closed form.

The estimation of the mixing proportions ω is computed by maximizing the function $Q(\theta, \theta^{(t)})$ with respect to ω subject to the constraint $\sum_{k=1}^{K} \omega_k = 1$. This maximiazion can be solved through a constrained optimization using Lagrangian multipliers. The estimates of $\omega_k^{(t+1)}$ are computed as follows:

$$\omega_k^{(t+1)} = \frac{\sum_{i=1}^{n} \eta_{ik}^{(t)}}{n}. \tag{A.7}$$

The Expectation and the Maximization steps are iterated until the difference $\ell(\theta^{(t+1)}) - \ell(\theta^{(t)})$ is less than an arbitrarily small amount ϵ. This stopping criterion can be used because the likelihood function $\ell(\theta)$ increases with each iteration of the EM algorithm (Dempster, Laird, and Rubin, 1977), which means that:

$$\ell(\theta^{(t+1)}) \geqslant \ell(\theta^{(t)}). \tag{A.8}$$

The EM algorithm can sometimes converge to a local maximum, making it crucial to initialize the algorithm using a range of starting values for the parameter vector θ or by exploring different initial partitions of the data into K groups. The second approach can be implemented by randomly dividing the data into K groups and then estimating the component parameters. Alternatively, the initial partitions can be determined using clustering methods, such as the k-means algorithm (Coleman et al., 1999).

APPENDIX TO CHAPTER 3

B

In this appendix, the development of the EM algorithm for obtaining the Maximum Likelihood Estimates and the Information Matrix of the CUM5 model is explained.

First of all, it has to be recalled that, given the number of categories, m, and the collection of ratings of n respondents, $\mathbf{R}$, the log-likelihood of a CUM model is defined as follows (Manisera and Zuccolotto, 2022):

$$\ell(\theta \mid \mathbf{R}) = \sum_{i=1}^{n} \ln \left\{ \pi P_W(r_i \mid \xi_D, \xi_U) + (1-\pi) P_U \right\}. \tag{B.1}$$

The log-likelihood is then maximized by using the EM algorithm (Nelder and Mead, 1965). For computing the Maximum Likelihood Estimates of the parameters, the complete data log-likelihood has to be defined.

B.1 EM ALGORITHM FOR THE CUM5 MODEL

Given $\mathbf{r} = (r_1, \ldots, r_n)'$ the realizations of the mixture, and given the unobserved data $\mathbf{z} = (z_1, \ldots, z_n)'$ which is the realization of a random variable Z_i and it is an indicator that is equal to 1 if the subject i gives rating derived from the feeling component, 0 otherwise; the complete data log-likelihood is defined as follows:

$$\ell_c(\theta \mid \mathbf{r}, \mathbf{z}) = \sum_{i=1}^{n} \left\{ z_i \ln[\pi P_W(r_i \mid \xi_D, \xi_U)] + (1-z_i) \ln[(1-\pi) P_U] \right\}. \tag{B.2}$$

The EM algorithm proceeds iteratively maximizing equation (B.1) by a step-by-step maximization of (B.2).

E-STEP (TTH ITERATION) The function that has to be maximized in order to compute the EM estimates of the parameters is the following:

$$Q(\theta \mid \theta^{(t)}) = \mathbb{E}_{Z|r_i,\theta^{(t)}}[\ell_c(\theta \mid \mathbf{r}, \mathbf{z})]. \tag{B.3}$$

It can also be rewritten as

$$\sum_{i=1}^{n} \left\{ \mathbb{E}_{Z_i|r_i,\theta^{(t)}} \ln \left[\pi^{(t)} P_W(r \mid \xi_D^{(t)}, \xi_U^{(t)}) \right] + 1 - \mathbb{E}_{Z_i|r_i,\theta^{(t)}} \ln \left[\frac{1-\pi^{(t)}}{m} \right] \right\},$$

where $\pi^{(t)}$, $\xi_D^{(t)}$ and $\xi_U^{(t)}$ are the current parameter estimates, and $\mathbb{E}_{Z_i|r_i,\theta^{(t)}}$ estimates the probability that r_i is expressed according to the feeling component, given the rating r_i of the ith subject, using the current parameter estimates:

$$\mathbb{E}_{Z_i|r_i,\theta^{(t)}} = \frac{\pi^{(t)} P_W(r_i|\xi_D^{(t)}, \xi_U^{(t)})}{\pi^{(t)} P_W(r_i|\xi_D^{(t)}, \xi_U^{(t)}) + (1-\pi^{(t)}) P_U} = \tau_i^{(t)}.$$

Therefore, the current conditional expectation of (B.3) is:

$$Q(\theta, \theta^{(t)}) = \sum_{i=1}^{n} \left\{ \tau_i^{(t)} \ln \left[\pi P_W(r_i \mid \xi_D, \xi_U) \right] + (1 - \tau_i^{(j)}) \ln \left[(1-\pi) P_U \right] \right\} \tag{B.4}$$

M-STEP (JTH ITERATION) In this step, the function $Q(\theta, \theta^{(t)})$ is maximized with respect to π, ξ_D, and ξ_U in order to obtain the estimators of the parameters. Equation (B.5) can be rewritten as follows:

$$\begin{aligned} Q(\theta, \theta^{(t)}) = &\sum_{i=1}^{n} \tau_i^{(t)} \ln(\pi) + (1 - \tau_i^{(t)}) \ln(1-\pi) + \\ &+ \sum_{i=1}^{n} \tau_i^{(t)} \ln \left[P_W(r_i \mid \xi_D, \xi_U) \right] + (1 - \tau_i^{(t)}) \ln(P_U), \end{aligned} \tag{B.5}$$

therefore, $Q(\theta, \theta^{(t)})$ can be expressed as a sum of two functions depending on π and the couple ξ_D, ξ_U, respectively:

$$Q(\theta, \theta^{(t)}) = Q_1^{(t)}(\pi) + Q_2^{(t)}(\xi_D, \xi_U), \tag{B.6}$$

where

$$Q_1^{(t)}(\pi) = \sum_{i=1}^{n} \tau_i^{(j)} \ln(\pi) + (1 - \tau_i^{(j)}) \ln(1-\pi),$$

and

$$Q_2^{(t)}(\xi_D, \xi_U) = \sum_{i=1}^{n} \tau_r^{(t)} \ln P_W(r_i | \xi_D, \xi_U) - (1 - \tau_i^{(t)}) \ln(P_U),$$

so that $Q_1^{(t)}(\pi)$ and $Q_2^{(t)}(\xi_D, \xi_U)$ can be maximized separately to obtain the estimates of the parameters. $Q_1^{(t)}(\pi)$ is maximized by solving the equation

$$\frac{\partial Q_1^{(t)}(\pi)}{\partial \pi} = \sum_{i=1}^{n} \frac{(\tau_i^{(t)} - \pi)}{\pi(1-\pi)} = 0$$

so by solving the equation

$$\pi = \frac{1}{n} \sum_{i=1}^{n} \tau_i^{(t)}.$$

The updated estimate of π is then a weighted mean of the $\tau_i^{(t)}$ computed in the E-step:

$$\pi^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \tau_i^{(t)}. \tag{B.7}$$

As far as it concerns ξ_D and ξ_U, the function $Q_2^{(t)}(\xi_D, \xi_U)$ is optimized by solving the following system of nonlinear equations

$$\begin{cases} \frac{\partial Q_2^{(t)}(\xi_D, \xi_U)}{\partial \xi_D} = \sum_{r=1}^{m} \frac{\tau_i^{(t)}}{P_W(r_i | \xi_D, \xi_U)} \frac{\partial P_w(r | \xi_D, \xi_U)}{\partial \xi_D} = 0 \\ \frac{\partial Q_2^{(t)}(\xi_D, \xi_U)}{\partial \xi_U} = \sum_{r=1}^{m} \frac{\tau_i^{(t)}}{P_W(r_i | \xi_D, \xi_U)} \frac{\partial P_W(r_i | \xi_D, \xi_U)}{\partial \xi_U} = 0 \end{cases} \tag{B.8}$$

The derivatives $\frac{\partial P_W(r_i|\xi_D,\xi_U)}{\partial \xi_D}$ and $\frac{\partial P_W(r_i|\xi_D,\xi_U)}{\partial \xi_U}$ can be easily obtained with simple algebra, as it is going to be explained in the following section. The updated estimates $\xi_D^{(t+1)}$ and $\xi_U^{(t+1)}$ are then given by the values of ξ_D and ξ_U that satisfy the system (B.8) within the boundaries of the parameter space; the solution is not closed-form, but it must be searched by numerical methods.

First and Second derivatives of $Q_2^{(t)}(\xi_D, \xi_U)$ *with respect to* ξ_D *and* ξ_U

In this appendix some mathematical details about the partial derivatives of $P_W(r|\xi_D, \xi_U)$ for each value of $r = 1, \ldots, 5$ are given. These derivatives are used in the M-step of the EM algorithm for the estimates of the parameters of a CUM model with $m = 5$.

$$P_W(1|\xi_D, \xi_U) = \binom{2}{2,0,0}\xi_D^2 = \xi_D^2$$
$$\frac{\partial P_W(1|\xi_D, \xi_U)}{\partial \xi_D} = 2\xi_D$$
$$\frac{\partial^2 P_W(1|\xi_D, \xi_U)}{\partial \xi_D^2} = 2$$
$$\frac{\partial P_W(1|\xi_D, \xi_U)}{\partial \xi_U} = 0$$
$$\frac{\partial^2 P_W(1|\xi_D, \xi_U)}{\partial \xi_U^2} = 0$$
$$\frac{\partial^2 P_W(1|\xi_D, \xi_U)}{\partial \xi_D \partial \xi_U} = 0$$

* * *

$$P_W(2|\xi_D, \xi_U) = \binom{2}{1,0,1}\xi_D(1 - \xi_D - \xi_U) = 2\xi_D - 2\xi_D^2 - 2\xi_D\xi_U$$
$$\frac{\partial P_W(2|\xi_D, \xi_U)}{\partial \xi_D} = 2 - 4\xi_D - 2\xi_U$$
$$\frac{\partial^2 P_W(2|\xi_D, \xi_U)}{\partial \xi_D^2} = -4$$
$$\frac{\partial P_W(2|\xi_D, \xi_U)}{\partial \xi_U} = -2\xi_D$$
$$\frac{\partial^2 P_W(2|\xi_D, \xi_U)}{\partial \xi_U^2} = 0$$
$$\frac{\partial^2 P_W(2|\xi_D, \xi_U)}{\partial \xi_D \partial \xi_U} = -2$$

* * *

$$
\begin{aligned}
P_W(3|\xi_D,\xi_U) &= \binom{2}{0,0,2}(1-\xi_D-\xi_U)^2 + \binom{2}{1,1,0}\xi_D\xi_U \\
&= (1-\xi_D-\xi_U)^2 + 2\xi_D\xi_U
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial P_W(3|\xi_D,\xi_U)}{\partial \xi_D} &= -2(1-\xi_D-\xi_U) + 2\xi_U \\
&= -2 + 2\xi_D + 4\xi_U
\end{aligned}
$$

$$
\frac{\partial^2 P_W(3|\xi_D,\xi_U)}{\partial \xi_D^2} = 2
$$

$$
\begin{aligned}
\frac{\partial P_W(3|\xi_D,\xi_U)}{\partial \xi_U} &= -2(1-\xi_D-\xi_U) + 2\xi_D \\
&= -2 + 4\xi_D + 2\xi_U
\end{aligned}
$$

$$
\frac{\partial^2 P_W(3|\xi_D,\xi_U)}{\partial \xi_U^2} = 2
$$

$$
\frac{\partial^2 P_W(3|\xi_D,\xi_U)}{\partial \xi_D \partial \xi_U} = 4
$$

* * *

$$
\begin{aligned}
P_W(4|\xi_D,\xi_U) &= \binom{2}{0,1,1}\xi_U(1-\xi_D-\xi_U) \\
&= 2\xi_U - 2\xi_D\xi_U - 2\xi_U^2
\end{aligned}
$$

$$
\frac{\partial P_W(4|\xi_D,\xi_U)}{\partial \xi_D} = -2\xi_U
$$

$$
\frac{\partial^2 P_W(4|\xi_D,\xi_U)}{\partial \xi_D^2} = 0
$$

$$
\frac{\partial P_W(4|\xi_D,\xi_U)}{\partial \xi_U} = 2 - 2\xi_D - 4\xi_U
$$

$$
\frac{\partial^2 P_W(4|\xi_D,\xi_U)}{\partial \xi_U^2} = -4
$$

$$
\frac{\partial^2 P_W(4|\xi_D,\xi_U)}{\partial \xi_D \partial \xi_U} = -2
$$

* * *

$$P_W(5|\xi_D,\xi_U) = \binom{2}{0,2,0}\xi_U^2 = \xi_U^2$$
$$\frac{\partial P_W(5|\xi_D,\xi_U)}{\partial \xi_D} = 0$$
$$\frac{\partial^2 P_W(5|\xi_D,\xi_U)}{\partial \xi_D^2} = 0$$
$$\frac{\partial P_W(5|\xi_D,\xi_U)}{\partial \xi_U} = 2\xi_U$$
$$\frac{\partial^2 P_W(5|\xi_D,\xi_U)}{\partial \xi_U^2} = 2$$
$$\frac{\partial^2 P_W(5|\xi_D,\xi_U)}{\partial \xi_D \partial \xi_U} = 0$$

B.2 INFORMATION MATRIX

In this section, it is shown how to obtain the observed Information Matrix $\hat{I}(\theta)$ for CUM5 model.

$$\begin{aligned}
\hat{I}_{\pi\pi} = n_1 &\left[\frac{\xi_D^2 - \frac{1}{m}}{\pi\left(\xi_D^2 - \frac{1}{m}\right) + \frac{1}{m}}\right]^2 + \\
&+ n_2 \left[\frac{2\xi_D(1-\xi_D-\xi_U) - \frac{1}{m}}{\pi\left(2\xi_D(1-\xi_D-\xi_U) - \frac{1}{m}\right) + \frac{1}{m}}\right]^2 + \\
&+ n_3 \left[\frac{(1-\xi_D-\xi_U)^2 + 2\xi_D\xi_U - \frac{1}{m}}{\pi\left((1-\xi_D-\xi_U)^2 + 2\xi_D\xi_U - \frac{1}{m}\right) + \frac{1}{m}}\right]^2 + \\
&+ n_4 \left[\frac{2\xi_U(1-\xi_D-\xi_U) - \frac{1}{m}}{\pi\left(2\xi_U(1-\xi_D-\xi_U) - \frac{1}{m}\right) + \frac{1}{m}}\right]^2 + \\
&+ n_5 \left[\frac{\xi_U^2 - \frac{1}{m}}{\pi\left(\xi_U^2 - \frac{1}{m}\right) + \frac{1}{m}}\right]^2
\end{aligned}$$

* * *

$$\hat{I}_{\xi_D\xi_D} = \pi\Bigg\{\pi n_1\left[\frac{2\xi_D}{\pi\left(\xi_D^2-\frac{1}{m}\right)+\frac{1}{m}}\right]^2 - n_1\frac{2}{\pi\left(\xi_D^2-\frac{1}{m}\right)+\frac{1}{m}}+$$
$$+\pi n_2\left[\frac{2-4\xi_D-2\xi_U}{\pi\left(2\xi_D-2\xi_D^2-2\xi_D\xi_U-\frac{1}{m}\right)+\frac{1}{m}}\right]^2+$$
$$-n_2\frac{-4}{\pi\left(2\xi_D-2\xi_D^2-2\xi_D\xi_U-\frac{1}{m}\right)+\frac{1}{m}}+$$
$$+\pi n_3\left[\frac{-2+2\xi_D+4\xi_U}{\pi\left((1-\xi_D-\xi_U)^2+2\xi_D\xi_U\right)-\frac{1}{m}\right)+\frac{1}{m}}\right]^2+$$
$$-n_3\frac{2}{\pi\left((1-\xi_D-\xi_U)^2+2\xi_D\xi_U\right)-\frac{1}{m}\right)+\frac{1}{m}}+$$
$$+\pi n_4\left[\frac{-2\xi_U}{\pi\left(2\xi_U-2\xi_D\xi_U-2\xi_U^2\right)-\frac{1}{m}\right)+\frac{1}{m}}\right]^2 - n_4\cdot 0+$$
$$+n_5[\pi * 0^2 - 0]\Bigg\}$$

* * *

$$\hat{I}_{\xi_U\xi_U} = \pi\Bigg\{n_1[\pi * 0^2 - 0]+$$
$$+\pi n_2\left[\frac{-2\xi_D}{\pi\left(2\xi_D-2\xi_D^2-2\xi_D\xi_U-\frac{1}{m}\right)+\frac{1}{m}}\right]^2 - n_2\cdot 0+$$
$$+\pi n_3\left[\frac{-2+4\xi_D+2\xi_U}{\pi\left((1-\xi_D-\xi_U)^2+2\xi_D\xi_U-\frac{1}{m}\right)+\frac{1}{m}}\right]^2+$$
$$-n_3\frac{2}{\pi\left((1-\xi_D-\xi_U)^2+2\xi_D\xi_U-\frac{1}{m}\right)+\frac{1}{m}}+$$
$$+\pi n_4\left[\frac{2-2\xi_D-4\xi_U}{\pi\left(2\xi_U-2\xi_D\xi_U-2\xi_U^2-\frac{1}{m}\right)+\frac{1}{m}}\right]^2+$$
$$-n_4\frac{-4}{\pi\left(2\xi_U-2\xi_D\xi_U-2\xi_U^2-\frac{1}{m}\right)+\frac{1}{m}}+$$
$$+\pi n_5\left[\frac{2\xi_U}{\pi\left(\xi_U^2-\frac{1}{m}\right)+\frac{1}{m}}\right]^2 - n_5\frac{2}{\pi\left(\xi_U^2-\frac{1}{m}\right)+\frac{1}{m}}\Bigg\}$$

* * *

$$
\begin{aligned}
\hat{I}_{\pi\xi_D} = n_1 & \left[\frac{\pi(\xi_D^2 - \frac{1}{m})2\xi_D}{\left[\pi(\xi_D^2 - \frac{1}{m}) + \frac{1}{m}\right]^2} - \frac{2\xi_D}{\pi\left(\xi_D^2 - \frac{1}{m}\right) + \frac{1}{m}} \right] + \\
& + n_2 \Big[\frac{\pi\left(2\xi_D - 2\xi_D^2 - 2\xi_D\xi_U - \frac{1}{m}\right)(2 - 4\xi_D - 2\xi_U)}{\left[\pi(2\xi_D - 2\xi_D^2 - 2\xi_D\xi_U - \frac{1}{m}) + \frac{1}{m}\right]^2} + \\
& - \frac{2 - 4\xi_D - 2\xi_U}{\pi\left(2\xi_D - 2\xi_D^2 - 2\xi_D\xi_U - \frac{1}{m}\right) + \frac{1}{m}} \Big] + \\
& + n_3 \Big[\frac{\pi\left((1 - \xi_D - \xi_U)^2 + 2\xi_D\xi_U - \frac{1}{m}\right)(-2 + 2\xi_D + 4\xi_U)}{\left[\pi((1 - \xi_D - \xi_U)^2 + 2\xi_D\xi_U - \frac{1}{m}) + \frac{1}{m}\right]^2} + \\
& - \frac{-2 + 2\xi_D + 4\xi_U}{\pi((1 - \xi_D - \xi_U)^2 + 2\xi_D\xi_U - \frac{1}{m}) + \frac{1}{m}} \Big] + \\
& + n_4 \Big[\frac{\pi\left(2\xi_U - 2\xi_D\xi_U - 2\xi_U^2 - \frac{1}{m}\right)(-2\xi_U)}{\left[\pi\left(2\xi_U - 2\xi_D\xi_U - 2\xi_U^2 - \frac{1}{m}\right) + \frac{1}{m}\right]^2} + \\
& - \frac{-2\xi_U}{\pi\left(2\xi_U - 2\xi_D\xi_U - 2\xi_U^2 - \frac{1}{m}\right) + \frac{1}{m}} \Big] + \\
& + n_5(0 - 0)
\end{aligned}
$$

* * *

$$
\begin{aligned}
\hat{I}_{\pi\xi_U} = & \, n_1(0 - 0) + \\
& + n_2 \Big[\frac{\pi\left(2\xi_D - 2\xi_D^2 - 2\xi_D\xi_U - \frac{1}{m}\right)(-2\xi_D)}{\left[\pi(2\xi_D - 2\xi_D^2 - 2\xi_D\xi_U - \frac{1}{m}) + \frac{1}{m}\right]^2} + \\
& - \frac{-2\xi_D}{\pi\left(2\xi_D - 2\xi_D^2 - 2\xi_D\xi_U - \frac{1}{m}\right) + \frac{1}{m}} \Big] + \\
& + n_3 \Big[\frac{\pi\left((1 - \xi_D - \xi_U)^2 + 2\xi_D\xi_U - \frac{1}{m}\right)(-2 + 4\xi_D + 2\xi_U)}{\left[\pi\left((1 - \xi_D - \xi_U)^2 + 2\xi_D\xi_U - \frac{1}{m}\right) + \frac{1}{m}\right]^2} + \\
& - \frac{(-2 + 4\xi_D + 2\xi_U)}{\pi\left((1 - \xi_D - \xi_U)^2 + 2\xi_D\xi_U - \frac{1}{m}\right) + \frac{1}{m}} \Big] + \\
& + n_4 \Big[\frac{\pi\left(2\xi_U - 2\xi_D\xi_U - 2\xi_U^2 - \frac{1}{m}\right)(2 - 2\xi_D - 4\xi_U)}{\left[\pi\left(2\xi_U - 2\xi_D\xi_U - 2\xi_U^2 - \frac{1}{m}\right) + \frac{1}{m}\right]^2} + \\
& - \frac{2 - 2\xi_D - 4\xi_U}{\pi\left(2\xi_U - 2\xi_D\xi_U - 2\xi_U^2 - \frac{1}{m}\right) + \frac{1}{m}} \Big] + \\
& + n_5 \left[\frac{\pi\left(\xi_U^2 - \frac{1}{m}\right)2\xi_U}{\left[\pi\left(\xi_U^2 - \frac{1}{m}\right) + \frac{1}{m}\right]^2} - \frac{2\xi_U}{\pi\left(\xi_U^2 - \frac{1}{m}\right) + \frac{1}{m}} \right]
\end{aligned}
$$

* * *

$$
\begin{aligned}
\hat{I}_{\xi_D \xi_U} = \pi \Bigg\{ & n_1 (0-0) + \\
& + n_2 \Bigg[\frac{\pi (2 - 4\xi_D - 2\xi_U)(-2\xi_D)}{\left[\pi(2\xi_D - 2\xi_D^2 - 2\xi_D \xi_U - \frac{1}{m}) + \frac{1}{m}\right]^2} + \\
& - \frac{-2}{\pi \left(2\xi_D - 2\xi_D^2 - 2\xi_D \xi_U - \frac{1}{m}\right) + \frac{1}{m}} \Bigg] + \\
& + n_3 \Bigg[\frac{\pi (-2 + 2\xi_D + 4\xi_U)(-2 + 4\xi_D + 2\xi_U)}{\left[\pi \left((1 - \xi_D - \xi_U)^2 + 2\xi_D \xi_U - \frac{1}{m}\right) + \frac{1}{m}\right]^2} + \\
& - \frac{4}{\pi \left((1 - \xi_D - \xi_U)^2 + 2\xi_D \xi_U - \frac{1}{m}\right) + \frac{1}{m}} \Bigg] + \\
& + n_4 \Bigg[\frac{\pi (-2\xi_U)(2 - 2\xi_D - 4\xi_U)}{\left[\pi \left(2\xi_U - 2\xi_D \xi_U - 2\xi_U^2 - \frac{1}{m}\right) + \frac{1}{m}\right]^2} + \\
& - \frac{-2}{\pi \left(2\xi_U - 2\xi_D \xi_U - 2\xi_U^2 - \frac{1}{m}\right) + \frac{1}{m}} \Bigg] + \\
& + n_5 (0-0) \Bigg\}
\end{aligned}
$$

B.3 SIMULATIONS RESULTS

In this section, additional results of the simulations with smaller sample sizes are reported.

Table B.1: Quality metrics for results of the simulation study with a sample size of 100 observations, CUM5

	a			**b**			**c**		
	$\widehat{\text{AB}}$	$\widehat{\text{MSE}}$	$\overline{\text{diss}}$	$\widehat{\text{AB}}$	$\widehat{\text{MSE}}$	$\overline{\text{diss}}$	$\widehat{\text{AB}}$	$\widehat{\text{MSE}}$	$\overline{\text{diss}}$
Case 1	0.2007	0.0636	0.0344	0.1000	0.0212	0.0297	0.0611	0.0074	0.0257
Case 2	0.1211	0.0252	0.0351	0.1029	0.0189	0.0326	0.0911	0.0157	0.0315
Case 3	0.0866	0.0128	0.0261	0.0690	0.0083	0.0227	0.0552	0.0056	0.0193
Case 4	0.1283	0.0284	0.0394	0.1072	0.0206	0.0367	0.0961	0.0163	0.0378
Case 5	0.1392	0.0379	0.0282	0.1084	0.0256	0.0296	0.0877	0.0162	0.0323
Case 6	0.0921	0.0145	0.0294	0.0758	0.0100	0.0249	0.0616	0.0069	0.0225

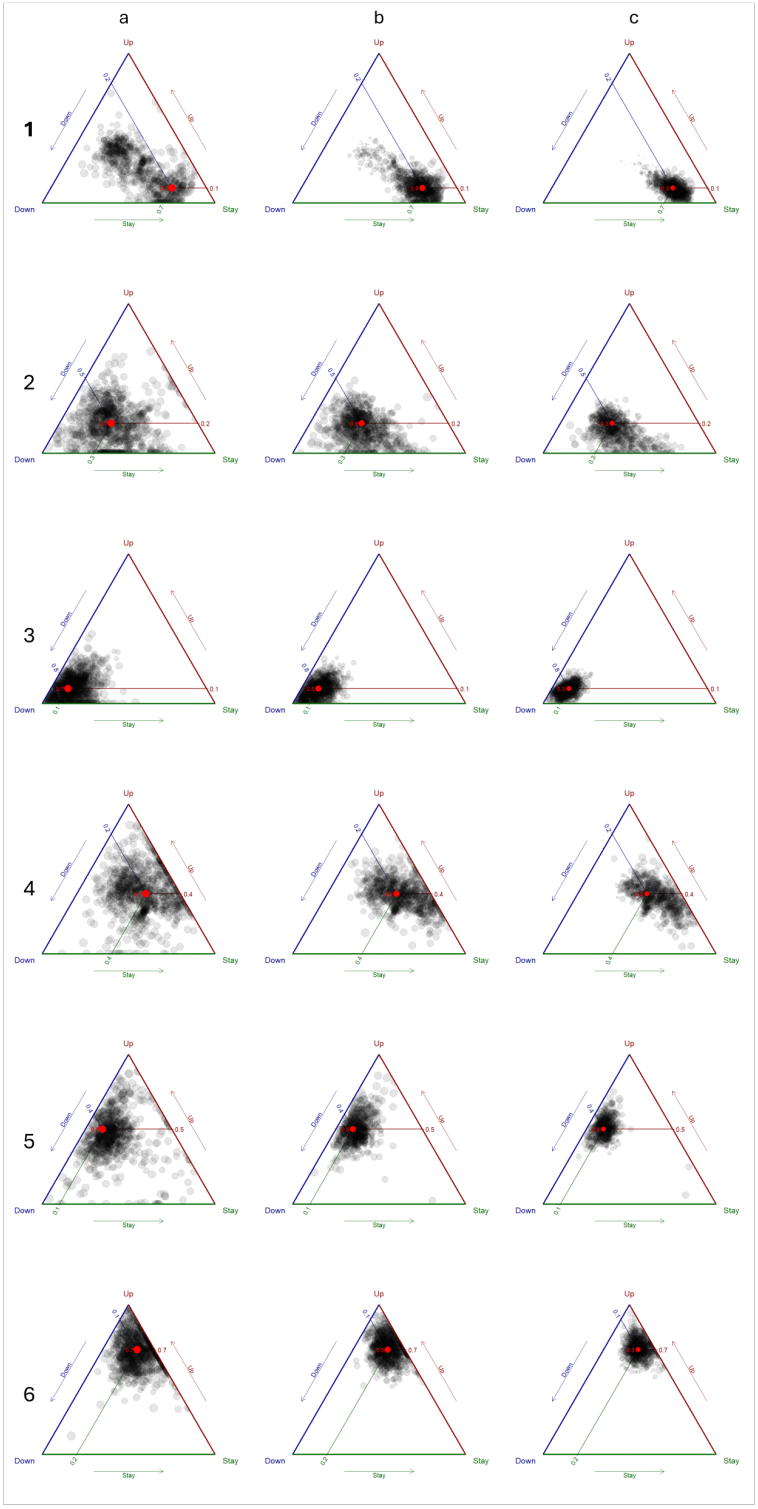


Figure B.1: Ternary plots for CUM5 simulations ($n = 100, iter = 1000$). Red dot: true parameter value; grey dots: estimated values.

Table B.2: Quality metrics for results of the simulation study with a sample size of 500 observations, CUM5

	a			**b**			**c**		
	$\widehat{\text{AB}}$	$\widehat{\text{MSE}}$	$\overline{\text{diss}}$	$\widehat{\text{AB}}$	$\widehat{\text{MSE}}$	$\overline{\text{diss}}$	$\widehat{\text{AB}}$	$\widehat{\text{MSE}}$	$\overline{\text{diss}}$
Case 1	0.1929	0.0569	0.0181	0.0368	0.0029	0.0117	0.0271	0.0014	0.0118
Case 2	0.0663	0.0074	0.0139	0.0519	0.0051	0.0139	0.0447	0.0043	0.0136
Case 3	0.0455	0.0033	0.0105	0.0324	0.0018	0.0092	0.0248	0.0011	0.0079
Case 4	0.0713	0.0083	0.0162	0.0622	0.0070	0.0153	0.0603	0.0071	0.0154
Case 5	0.0620	0.0073	0.0135	0.0553	0.0068	0.0128	0.0517	0.0065	0.0129
Case 6	0.0489	0.0038	0.0117	0.0355	0.0022	0.0111	0.0281	0.0015	0.0106

Table B.3: Quality metrics for results of the simulation study with a sample size of 1000 observations, CUM5

	a			**b**			**c**		
	$\widehat{\text{AB}}$	$\widehat{\text{MSE}}$	$\overline{\text{diss}}$	$\widehat{\text{AB}}$	$\widehat{\text{MSE}}$	$\overline{\text{diss}}$	$\widehat{\text{AB}}$	$\widehat{\text{MSE}}$	$\overline{\text{diss}}$
Case 1	0.2158	0.0665	0.0166	0.0017	0.0010	0.0082	0.0015	0.0007	0.0080
Case 2	0.0063	0.0040	0.0101	0.0024	0.0026	0.0105	0.0022	0.0019	0.0100
Case 3	0.0013	0.0016	0.0076	0.0013	0.0009	0.0065	0.0006	0.0005	0.0058
Case 4	0.0052	0.0048	0.0115	0.0050	0.0045	0.0110	0.0063	0.0038	0.0104
Case 5	0.0131	0.0040	0.0097	0.0058	0.0037	0.0089	0.0013	0.0039	0.0089
Case 6	0.0025	0.0020	0.0083	0.0015	0.0011	0.0075	0.0007	0.0007	0.0074

Table B.4: Quality metrics for results of the simulation study, CUM7

	a			**b**			**c**		
	$\widehat{\text{AB}}$	$\widehat{\text{MSE}}$	$\overline{\text{diss}}$	$\widehat{\text{AB}}$	$\widehat{\text{MSE}}$	$\overline{\text{diss}}$	$\widehat{\text{AB}}$	$\widehat{\text{MSE}}$	$\overline{\text{diss}}$
Case 1	0.0124	0.0028	0.0199	0.0010	0.0007	0.0185	0.0002	0.0005	0.0174
Case 2	0.0101	0.0036	0.0206	0.0035	0.0018	0.0205	0.0029	0.0012	0.0205
Case 3	0.0009	0.0009	0.0172	0.0008	0.0008	0.0158	0.0003	0.0004	0.0148
Case 4	0.0043	0.0045	0.0218	0.0029	0.0027	0.0211	0.0027	0.0020	0.0212
Case 5	0.0056	0.0019	0.0213	0.0018	0.0016	0.0215	0.0011	0.0013	0.0211
Case 6	0.0019	0.0012	0.0180	0.0009	0.0007	0.0174	0.0005	0.0005	0.0164

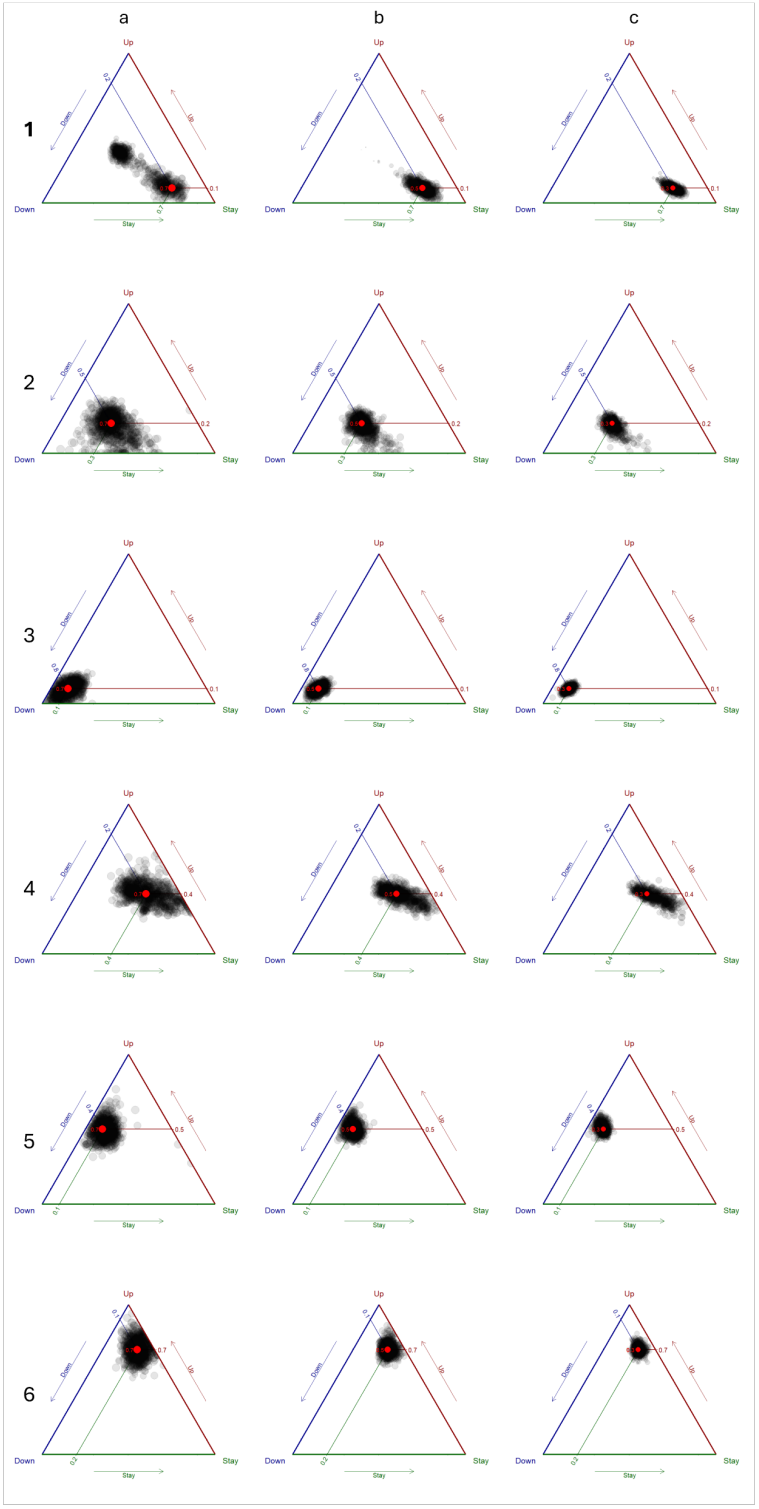


Figure B.2: Ternary plots for CUM5 simulations ($n = 500, iter = 1000$). Red dot: true parameter value; grey dots: estimated values.

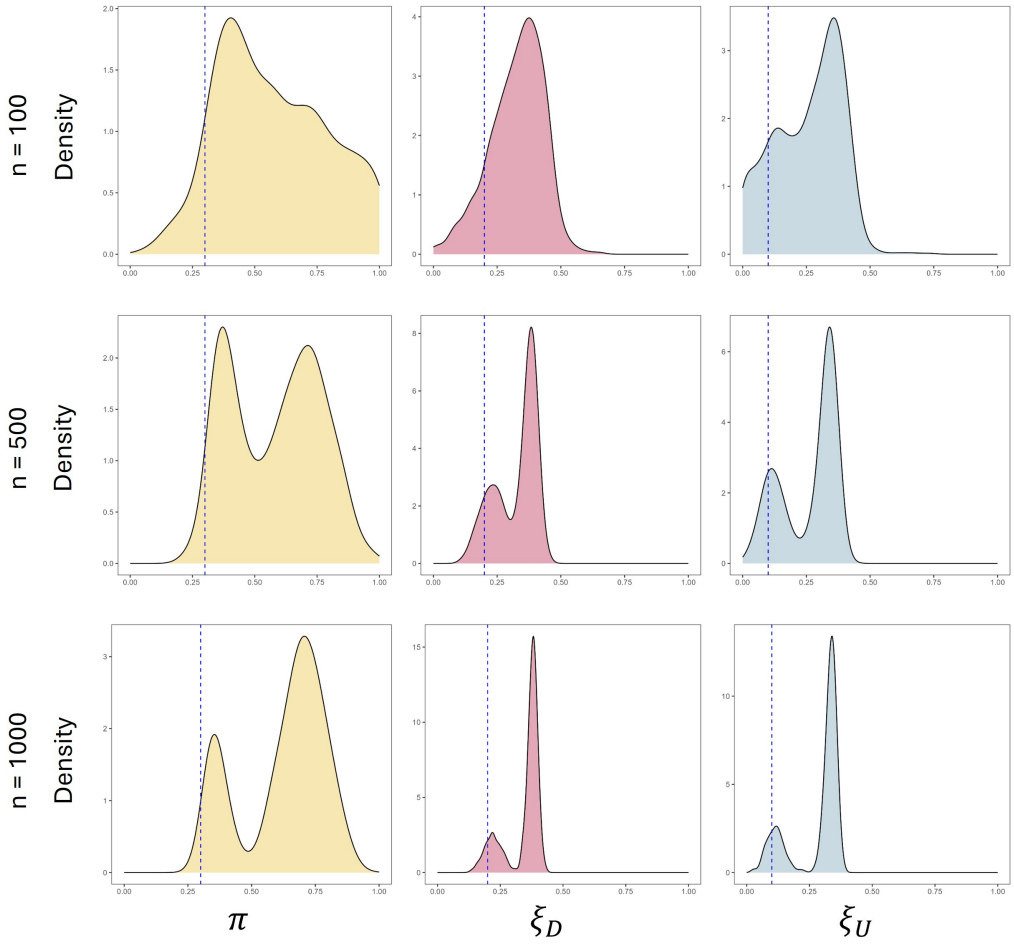


Figure B.3: CUM5 – Distributions of the estimated parameters for case 1a for sample size $n = \{100, 500, 1000\}$. The true values of the parameters are $\pi = 0.3$, $\xi_D = 0.2$, and $\xi_U = 0.1$.

C

APPENDIX TO CHAPTER 4

In this appendix, the algebra developed for obtaining the results shown in Chapter 4 is shown.

C.1 CONDITIONS FOR WHICH ξ_D AND ξ_U ARE IN THE DOMAIN

To find the interval of values of π_B and π_M for which $\xi_D + \xi_U \leqslant 1$ when $\pi_B < \pi_M$, it is necessary to study the parabola described by equation $y = \xi_D + \xi_U$, whose components can be managed as shown to make calculations simpler. Therefore, recalling equation 4.10, the function to study is:

$$y = \xi_D + \xi_U = \sqrt{\delta(0.5-x)^4 + \beta} + \sqrt{\delta(0.5+x)^4 + \beta}, \tag{C.1}$$

where $\delta = \frac{\pi_B}{\pi_M}$, $\beta = (1-\delta)/5$, and $x = 0.5 - \xi$.

The function is symmetric with respect to $x = 0$ since $f(x) = f(-x)$, and its domain is $[-0.5, 0.5]$, since $\xi \in [0, 1]$.

The derivative of function C.1 is computed as follows:

$$\begin{aligned}\frac{\partial y}{\partial x} =& -\frac{1}{2}\left[\delta(0.5-x)^4 + \beta\right]^{\frac{1}{2}} 4(0.5-x)^3 + \\ &+ \frac{1}{2}\left[\delta(0.5+x)^4 + \beta\right]^{\frac{1}{2}} 4(0.5+x)^3\end{aligned}$$

This derivative is equal to 0 if $x = 0$, and is greater than zero if $x > 0$. Therefore, we can conclude that the function C.1 is symmetric in $x = 0$ (i. e., $\xi = 0.5$).

Now let's focus on the vertex of the parabola and let's compute when $y < 1$. The function C.1 in $x = 0$ can be rewritten as follows:

$$y = \sqrt{\frac{\delta}{2^4} + \beta} + \sqrt{\frac{\delta}{2^4} + \beta} = \frac{1}{2}\sqrt{\delta + 2^4\beta},$$

and the inequality becomes:

$$\begin{aligned}\frac{1}{2}\sqrt{\delta + 2^4\beta} &< 1 \\ \frac{1}{4}\left(\delta + 2^4\beta\right) &< 1\end{aligned}$$

$$\begin{aligned}\frac{1}{4}\left(\frac{\pi_B}{\pi_M} + \frac{2^4}{5} - \frac{2^4\pi_B}{5\pi_M}\right) &< 1 \\ \frac{5\pi_B + 2^4\pi_M - 2^4\pi_B - 20\pi_M}{20\pi_M} &< 0\end{aligned}$$

Since the denominator is always greater than zero, only the numerator can be considered, which is always lower than zero. Therefore it can be concluded that $y < 1 \ \forall \ x$.

In the following, the behaviour of the function at the domain boundaries is studied. Since the function is symmetric, only the behaviour in $x = 0.5$ can be studied and then extended for the case with $x = -0.5$.

In $x = 0.5$ the function becomes:

$$\sqrt{\beta} + \sqrt{\delta + \beta},$$

and the inequality can be solved as follows:

$$\begin{gathered}
\sqrt{\beta} + \sqrt{\delta + \beta} < 1 \\
\sqrt{\frac{1}{5}(1-\delta)} + \sqrt{\delta + \frac{1}{5}(1-\delta)} < 1 \\
\frac{1}{5}(1-\delta) + \delta + \frac{1}{5}(1-\delta) + \sqrt{\frac{\delta}{5}(1-\delta) + \frac{1}{25}(1-\delta)^2} < 1 \\
\frac{1-\delta+5\delta+1-\delta}{5} + 2\sqrt{\frac{\delta(1-\delta)}{5} + \frac{(1-\delta)^2}{25}} < 1 \\
2\sqrt{\frac{5\delta - 5\delta^2 + 1 + \delta^2 - 2\delta}{25}} < 1 - \frac{2+3\delta}{5} \\
\sqrt{\frac{-4\delta^2 + 3\delta + 1}{25}} < \frac{3-3\delta}{10} \\
\frac{-4\delta^2 + 3\delta + 1}{25} < \frac{9(1+\delta^2-2\delta)}{100} \\
\frac{-16\delta^2 + 12\delta + 4 - 9 - 9\delta^2 + 18\delta}{100} < 0
\end{gathered}$$

The numerator becomes $-5\delta^2 + 6\delta - 1 < 0$. Which happens for $0 < \delta < 1/5$.

Considering that $\delta = \pi_B/\pi_M$, it can be concluded that when $\pi_B < \pi_M$:

- ξ_D and ξ_U are defined for all ξ if $0 < \delta \leqslant 1/5$, i. e. $0 < \pi_B \leqslant \frac{\pi_M}{5}$;
- ξ_D and ξ_U are defined for $\xi \in I(0.5)$ if $5 < \delta < 1$, i. e. $\frac{\pi_M}{5} < \pi_B < \pi_M$.

C.2 PROOF THAT SYSTEM 4.8 HAS TWO SOLUTIONS IN $\xi = 0.5$

In $\xi = 0.5$:

$$\xi_D = \xi_U = \sqrt{0.5^4\delta + \frac{1-\delta}{5}}. \tag{C.2}$$

Considering this relation, equations 2 and 4 of the system 4.8 are equal and can be rewritten as follows:

$$4\pi_B 0.5^4 + \frac{\pi_M - \pi_B}{5} = 2\pi_M \xi_D (1 - 2\xi_D)$$
$$4\delta 0.5^4 + \frac{1-\delta}{5} = 2\xi_D - 4\xi_D^2$$
$$4\delta 0.5^4 + \frac{1-\delta}{5} = 2\sqrt{0.5^4\delta + \frac{1-\delta}{5}} - 4\delta 0.5^5 - \frac{4(1-\delta)}{5}$$
$$8\delta 0.5^4 + 1 - \delta = 2\sqrt{0.5^4\delta + \frac{1-\delta}{5}}$$
$$\frac{1}{4}\delta\frac{1}{2} - \frac{\delta}{2} = \sqrt{0.5^4\delta + \frac{1-\delta}{5}}$$
$$\left(\frac{1}{2} - \frac{1}{4}\delta\right)^2 = \frac{1}{16}\delta + \frac{1}{5} - \frac{1}{5}$$
$$\frac{1}{4} + \frac{1}{16}\delta^2 - \frac{1}{4}\delta - \frac{1}{16}\delta - \frac{1}{5} + \frac{1}{5}\delta = 0$$
$$\frac{1}{16}\delta^2 - \frac{9}{80}\delta + \frac{1}{20} = 0$$
$$\delta_{1,2} = \frac{9}{10} \pm \sqrt{\frac{81}{80^2} - \frac{1}{80}} = \frac{9}{10} \pm \frac{1}{10} \qquad \delta_1 = 1;\ \delta_2 = \frac{4}{5}$$

By substituting the values of δ in equation C.2, we can find that when $\delta = 1$, i. e.$\pi_B = \pi_M$, $\xi = 0.25$; while when $\delta = 4/5$, i. e.$\pi_M = 5\pi_B/4$, $\xi = 0.3$. Meaning that there are two sets of parameters for which equation 2 and 4 of system 4.8 can be solved.

To conclude that the system 4.8 has two solutions in $\xi = 0.5$, the same has to be checked for equation 3 of the system. It is simple to show that this equation is solved for $\delta = \frac{4}{5}$, therefore it can be concluded that the system has two solutions.

C.3 Proof that system 4.15 has only one solution in $\xi = 0.5$

Recalling the notation reported in equations 4.16, in $\xi = 0.5$:

$$\xi_D = \xi_U = \sqrt[3]{\delta 0.5^6 + \beta}. \tag{C.3}$$

First of all the solutions of equations 2 and 6 of system 4.15 are computed by substituting ξ_D and ξ_U as expressed in equation C.3. Equation 2 and 6 in this case are equal and becomes:

$$6\pi_B 0.5^6 + \frac{1-\pi_B}{7} = 3\pi_M \left(\sqrt[3]{\delta 0.5^6 + \beta}\right)^2 \left(1 - 2\sqrt[3]{\delta 0.5^6 + \beta}\right) + \frac{1-\pi_M}{7}$$
$$6\pi_B 0.5^6 + \frac{1-\pi_B}{7} - \frac{1-\pi_M}{7} = 3\pi_M \left(\sqrt[3]{\delta 0.5^6 + \beta}\right)^2 \left(1 - 2\sqrt[3]{\delta 0.5^6 + \beta}\right)$$

Dividing by π_M:

$$6\frac{\pi_B}{\pi_M} 0.5^6 + \frac{\pi_M - \pi_B}{7\pi_M} - \frac{1-\pi_M}{7} = 3\left(\sqrt[3]{\delta 0.5^6 + \beta}\right)^2 \left(1 - 2\sqrt[3]{\delta 0.5^6 + \beta}\right).$$

Recalling that, for the case $m = 7$, $\delta = \pi_B/\pi_M$ and $\beta = \frac{1}{7}(1-\delta)$, the previous equation can be rewritten as follows:

$$6\delta 0.5^6 + \beta + 6\delta 0.5^6 + 6\beta = 3\sqrt[3]{\left(\delta 0.5^6 + \beta\right)^2}$$

which, by applying some simple algebra becomes:

$$\frac{1}{3} - \frac{13}{48}\delta = \sqrt[3]{\left(\frac{1}{7} - \frac{57}{448}\delta\right)^2}$$
$$\left(\frac{1}{3} - \frac{13}{48}\delta\right)^3 = \left(\frac{1}{7} - \frac{57}{448}\delta\right)^2.$$

After some computations, the equation can be reduced to a third order polynomial:

$$-107653\delta^3 + 309765\delta^2 - 292224\delta + 90112 = 0.$$

Since we know that system 4.15 has solutions for $\pi_B = \pi_M$, i.e.$\delta = 1$, we can conclude that 1 is a zero of the polynomial and we can factorize it by using Ruffini's rule. Therefore the polynomial can be rewritten as follows:

$$(\delta - 1)(-107653\delta^2 + 202112\delta - 90112) = 0$$

and by considering the second factor, the second and third values of δ which solve the equation 2 and 6 can be found:

$$\delta_{1,2} = \frac{64(1579 \pm 15\sqrt{555})}{107653}.$$

To check if these are solutions of system 4.15, the values of δ should be substituted in equations 3, 4, and 5 of the system[1]. This shows that there exists only one solution of the system, which is $\delta = 1$.

1 Given the length of the computations of the solutions of these equations, this passage can be done with the aid of the mathematical software Mathematica (Wolfram Research, 2024)

APPENDIX TO CHAPTER 5

D

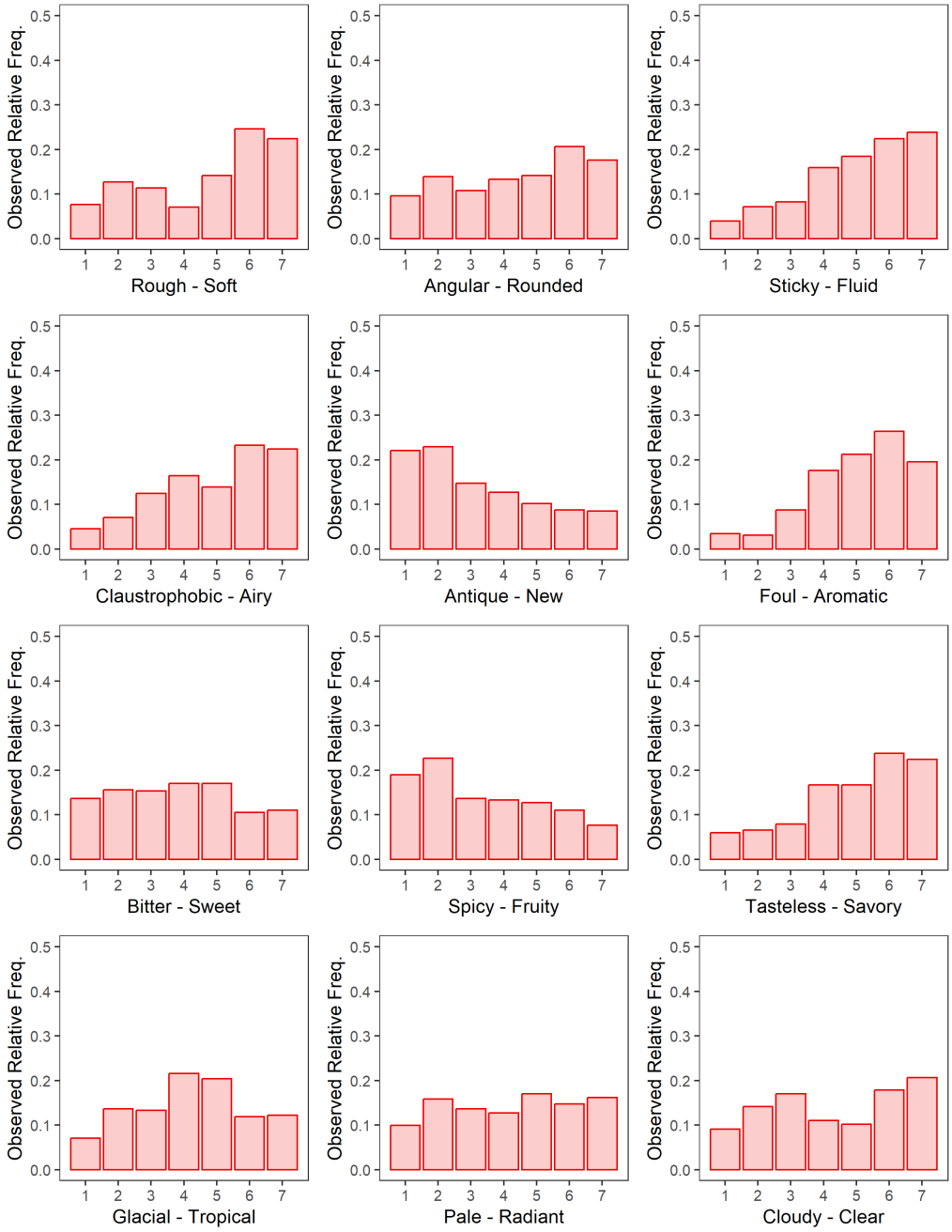


Figure D.1: Relative frequencies distributions of the answers given by the respondents in the Red room.

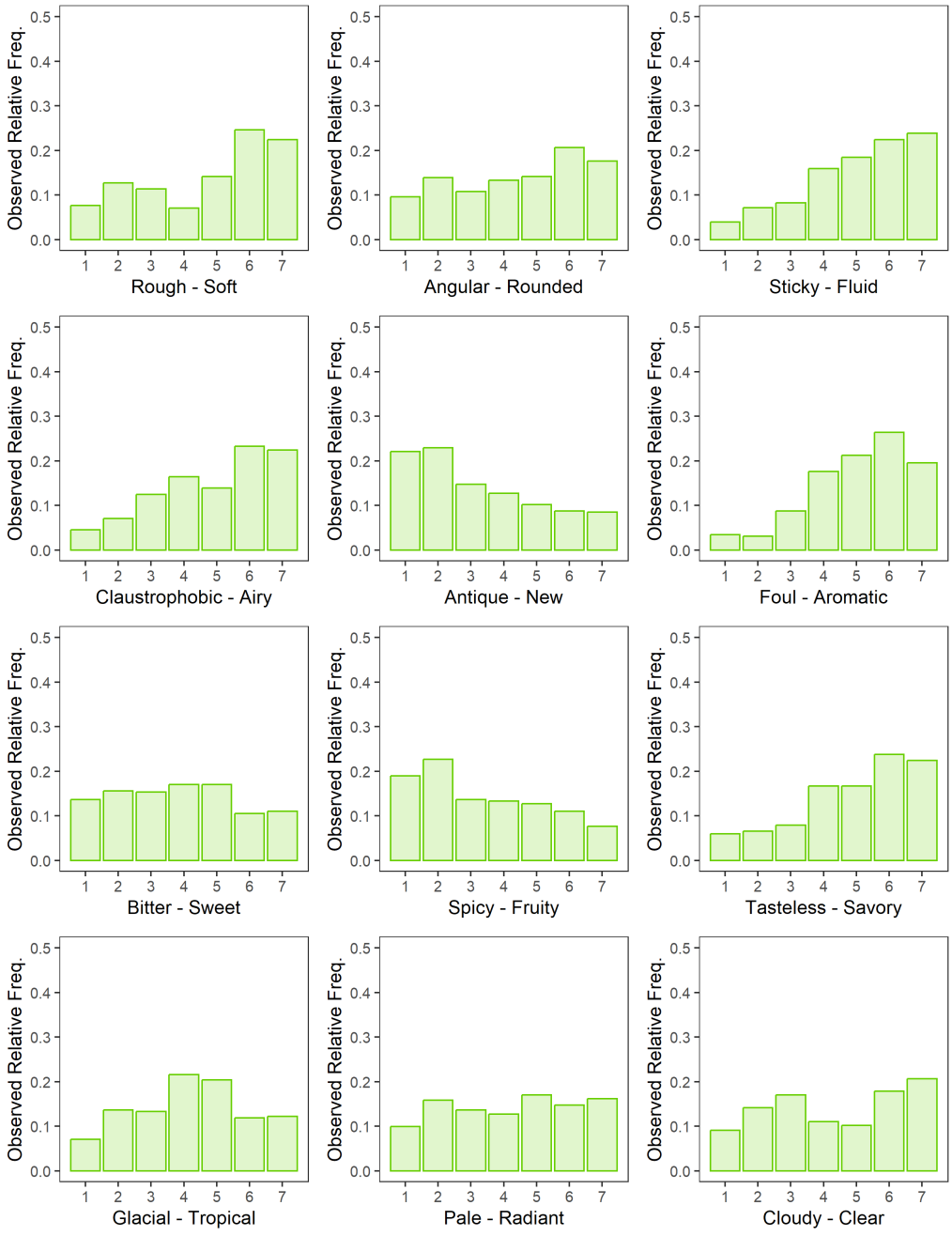


Figure D.2: Relative frequencies distributions of the answers given by the respondents in the Green room.

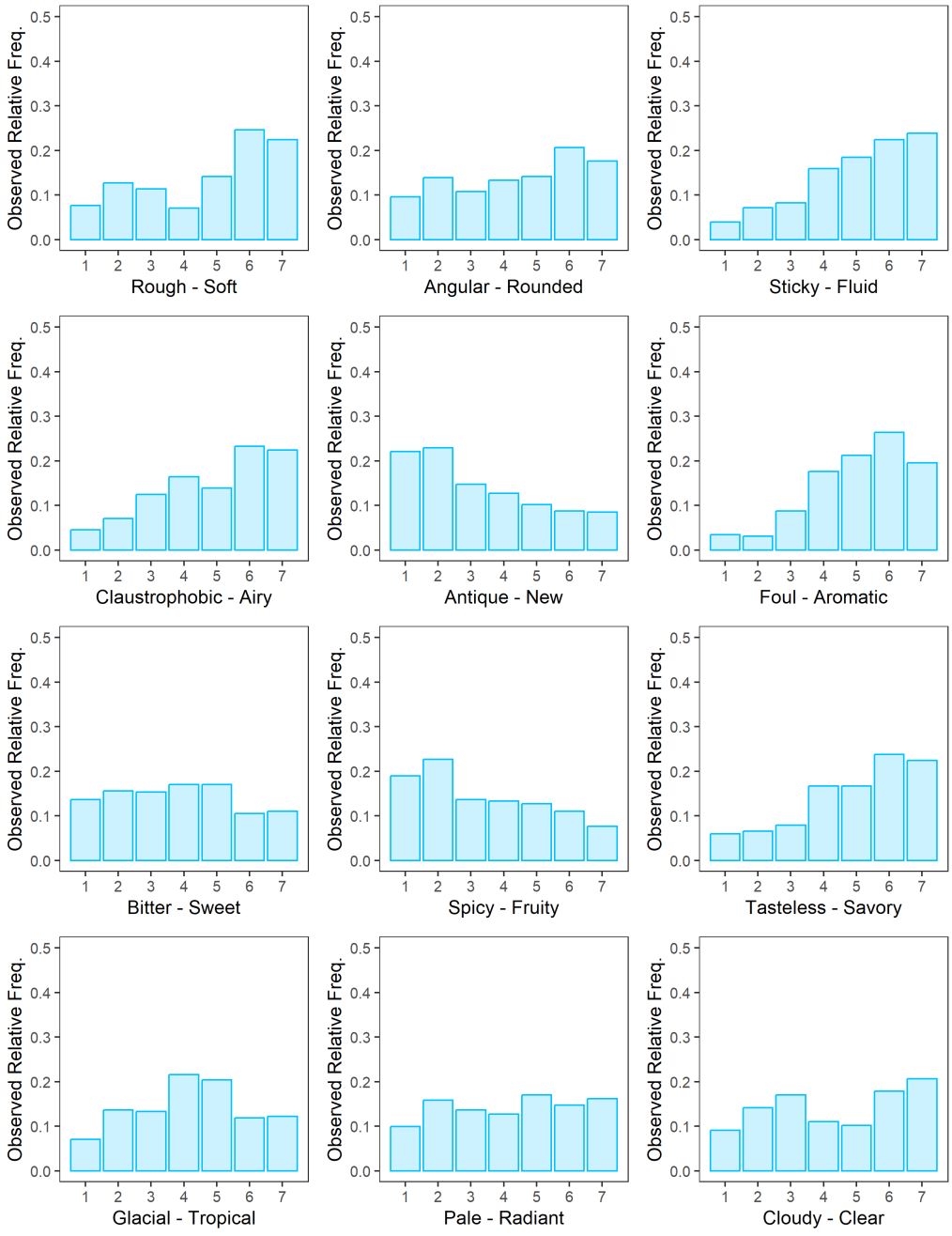


Figure D.3: Relative frequencies distributions of the answers given by the respondents in the Blue room.

E

APPENDIX TO CHAPTER 6

E.1 EM ALGORITHM FOR LATENT CLASS CUB MODEL

In the original paper by Grilli et al. (2014), the estimates of the parameters of the Latent Class CUB (LC-CUB) model are obtained by direct optimization through numerical derivatives. In this section, the Maximum Likelihood (ML) estimation of the model parameters through EM algorithm, is proposed.

Let R be a univariate ordinal random variable with m categories, and let ω_k being the mixing proportion of class K such that $\omega_k > 0$ and $\sum_{k=1}^{K} \omega_k = 1$. The LC-CUB model is defined as follows:

$$P(R = r \mid \theta_{LCCUB}) = \pi \sum_{k=1}^{K} \omega_k P_B(\xi_k) + (1-\pi) P_U. \tag{E.1}$$

with $\theta_{LCCUB} = (\pi, \boldsymbol{\xi}, \boldsymbol{\omega})$ where π is assumed to be constant across the classes in order to make the model identifiable; $\boldsymbol{\omega} = (\omega_k)$, and $\boldsymbol{\xi} = (\xi_k)$, with $k = 1, \ldots, K$. In this section, $\theta_{LCCUB} = \theta$ for the sake of simplicity.

Like for the CUB model, in order to obtain the ML estimates of the parameters, the latent variable $\mathbf{Z} : (Z_k)$ with $k = 1, \ldots, K+1$ has to be introduced. The latent variable is distributed as a one-order Multinomial distribution, $\mathbf{Z} \sim \mathcal{M}(1; \pi\omega_1, \ldots, \pi\omega_K, (1-\pi))$ whose realizations are expressed by z_{ik} and are such that $z_{ik} = 1$ if the ith rater's preference comes from the kth random variable, and $z_{ik} = 0$ otherwise. Given the data $\mathbf{r} = (r_1, \ldots, r_n)'$, the complete log-likelihood is defined as:

$$\ell_c(\theta \mid \mathbf{r}, \mathbf{Z}) = \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \{\ln(\pi_k) + \ln(P_k(r_i \mid \xi_k, \omega_k))\}, \tag{E.2}$$

where

$$P_k(r_i; \xi_k, \omega_k) = \begin{cases} P_k(r_i \mid \xi_k) = \omega_k P_B(r_i \mid \xi^k) & \text{if} \quad k = 1, \ldots, K \\ P_{K+1}(r_i) = P_U(r_i) & \text{if} \quad k = K+1 \end{cases}$$

E-STEP (t-TH ITERATION) The function which has to be maximized in order to obtain the EM estimates of the parameters is:

$$Q(\theta, \theta^t) = \mathbb{E}_{\mathbf{Z}|\mathbf{r}, \theta^{(t)}}[\ell_c(\theta \mid \mathbf{r}, \mathbf{Z})].$$

The conditional expectation can be computed for a given value of $k = 1, \ldots, K$ is computed as:

$$\mathbb{E}_{Z_{ik}|r_i, \theta^{(t)}} = \frac{\pi^{(t)} \omega_k^{(t)} P_B(r_i \mid \xi_k^{(t)})}{\pi^{(t)} \sum_{k=1}^{K} \omega_k^{(t)} P_B(r_i \mid \xi_k^{(t)}) + (1-\pi^{(t)}) P_U(r_i)} = \tau_{ik}^{(t)}, \tag{E.3}$$

while when $k = K+1$, the conditional expectation is equal to:

$$\mathbb{E}_{Z_{i(K+1)},r_i,\theta^{(t)}} = \frac{(1-\pi^t)P_U(r_i))}{\pi^{(t)}\sum_{k=1}^{K}\omega_k^{(t)}P_B(r_i \mid \xi_k^{(t)}) + (1-\pi^{(t)})P_U(r_i)} = \tau_{i(K+1)}^{(t)}$$

such that $\sum_{k=1}^{K}\tau_{ik}^{t} + \tau_{(K+1)i}^{t} = 1$.

Therefore, the function $Q(\theta, \theta^{(t)})$ becomes:

$$Q(\theta, \theta^{(t)}) = \sum_{i=1}^{n}\sum_{k=1}^{K+1}\tau_k(r_i \mid \theta^{(t)})\{\ln(\pi^{(t)}) + \ln[P_k(r_i \mid \xi_k, \omega_k)], \tag{E.4}$$

whose terms are then factorized as follows:

$$\begin{aligned}
Q(\theta, \theta^{(t)}) = &\sum_{i=1}^{n}\left[\sum_{k=1}^{K}\tau_{ik}^{(t)}\ \ln(\pi^{(t)}) + \tau_{i(K+1)}^{(t)}\ \ln(1-\pi^{(t)})\right] + \\
&+ \sum_{i=1}^{n}\left[\sum_{k=1}^{K}\tau_{ik}^{(t)}\ \ln\omega_k[P_B(r_i \mid \xi_k)] + \tau_{i(K+1)}^{(t)}\ln[P_U(r_i)]\right] = \\
&= \underbrace{\sum_{i=1}^{n}\left[\sum_{k=1}^{K}\tau_{ik}^{(t)}\ \ln(\pi^{(t)}) + \tau_{i(K+1)}^{(t)}\ \ln(1-\pi^{(t)})\right]}_{Q_1^{(t)}(\pi)} + \\
&+ \underbrace{\sum_{i=1}^{n}\sum_{k=1}^{K}\tau_{ik}^{(t)}\ \ln\omega_k}_{Q_2^{(t)}(\omega_k)} + \underbrace{\sum_{i=1}^{n}\left[\sum_{k=1}^{K}\tau_{ik}^{(t)}\ln P_B(r_i \mid \xi_k) + \tau_{i(K+1)}^{(t)}\ln[P_U(r_i)]\right]}_{Q_3^{(t)}(\xi_k)}
\end{aligned}$$

M-STEP (t-TH ITERATION) Following the procedure explained in chapter 2, it is simple to maximize the function $Q(\theta, \theta^{(t)})$ in order to derive the estimators of π and ξ_k . Also for the LC-CUB model, the function $Q(\theta, \theta^{(t)})$ can be written as:

$$Q(\theta, \theta^{(t)}) = Q_1^{(t)}(\pi) + Q_2^{(t)}(\omega_k) + Q_3^{(t)}(\xi_k) \tag{E.5}$$

The estimators of the parameters π, ω_k, and ξ_k can be simply obtained by maximizing the functions $Q_1^{(t)}(\pi)$, $Q_2^{(t)}(\omega_k)$, and $Q_3^{(t)}(\xi_k)$, respectively.

The maximisation of the function $Q_1^{(t)}(\pi)$ with respect to π allows us to obtain the updated estimate of the parameter:

$$\frac{\partial Q_1^{(t)}(\pi)}{\partial\pi} = \frac{\sum_{i=1}^{n}\sum_{k=1}^{K}\tau_{ik}^{(t)}}{\pi} - \frac{\sum_{i=1}^{n}\tau_{i(K+1)}^{(t)}}{1-\pi} = 0,$$

and after some simple algebra, the updated estimate of π is computed as follows:

$$\pi^{(t+1)} = \frac{\sum_{i=1}^{n}\sum_{k=1}^{K}\tau_{ik}^{(t)}}{\sum_{i=1}^{n}\left[\sum_{k=1}^{K}\tau_{ik}^{(t)} + \tau_{i(K+1)}^{(t)}\right]},$$

which can be written as:

$$\pi^{(t+1)} = \frac{\sum_{i=1}^{n}\sum_{k=1}^{K}\tau_{ik}^{(t)}}{n} \tag{E.6}$$

since $\sum_{k=1}^{K}\tau_{ik}^{(t)} + \tau_{i(K+1)}^{(t)} = 1$.

The parameter ω_k is obtained as result of a constrained maximization:

$$\sum_{i=1}^{n}\tau_{ik}^{(t)}\,\ln(\omega_k) \qquad \text{s. t.} \qquad \sum_{k=1}^{K}\omega_k = 1,$$

therefore, the Lagrangian function to maximise is:

$$\mathcal{L}(\omega_k;\lambda) = \sum_{i=1}^{n}\tau_{ik}^{(t)}\,\ln(\omega_k) - \lambda\left(\sum_{k=1}^{K}\omega_k - 1\right). \tag{E.7}$$

The first order derivative of the Lagrangian function with respect to ω_k is:

$$\frac{\partial\mathcal{L}(\omega_k;\lambda)}{\partial\omega_k} = \frac{\sum_i^n \tau_{ik}^{(t)}}{\omega_k} - \lambda = 0, \tag{E.8}$$

which is equal to:

$$\sum_{i}^{n}\tau_{ik}^{(t)} = \lambda\omega_k.$$

Then the sum over k is computed:

$$\sum_{i}^{n}\sum_{k=1}^{K}\tau_{ik}^{(t)} = \lambda\sum_{k=1}^{K}\omega_k,$$

where $\sum_{k=1}^{K}\omega_k = 1$, therefore $\lambda = \sum_i^n\sum_{k=1}^{K}\tau_{ik}^{(t)}$, and the first order derivative (E.8) can be written as:

$$\frac{\sum_i^n \tau_{ki}^t}{\omega_k} - \sum_{i}^{n}\sum_{k=1}^{K}\tau_{ki}^t = 0.$$

Now ω_k can be easily computed by solving the equation:

$$\omega_k^{(t+1)} = \frac{\sum_i^n \tau_{ki}^t}{\sum_i^n\sum_{k=1}^{K}\tau_{ki}^t}. \tag{E.9}$$

Finally, the parameter ξ_k is obtained by maximising the function $Q_3^{(t)}(\xi_k)$:

$$\frac{\partial Q_3^{(t)}(\xi_k)}{\partial\xi_k} = \sum_{i=1}^{n}\tau_{ik}^{(t)}\left[\frac{m-r_i}{\xi_k} - \frac{r_i-1}{1-\xi_k}\right] = 0,$$

which can be solved with some trivial algebra in order to obtain the ML estimator of the parameter ξ_k:

$$\xi^{(t+1)} = \frac{\sum_{i=1}^{n}\tau_{ik}^{(t)}r_i - m\sum_{i=1}^{n}\tau_{ik}^{(t)}}{-(m-1)\sum_{i=1}^{n}\tau_{ik}^{(t)}} = \frac{m - \frac{\sum_{i=1}^{n}\tau_{ik}^{(t)}r_i}{\sum_{i=1}^{n}\tau_{ik}^{(t)}}}{m-1},$$

where $\frac{\sum_{i=1}^{n} \omega_{ki}^{t} r_i}{\sum_{i=1}^{n} \omega_{ki}^{t}}$ is equal to $\bar{R}_k(p)$, which is the average of the ratings weighted with the posterior probability that r_i is a realization of the kth Shifted Binomial distribution, given the current data. Therefore ξ^{t+1} can be written as follows:

$$\xi^{(t+1)} = \frac{m - \bar{R}_k(p)}{m-1}. \tag{E.10}$$

The EM algorithm developed in this section can be considered a starting point for the development of a suitable EM algorithm for an LC-CUB model with covariates, as stated by Grilli et al. (2014), which can affect or the conditional CUB distributions, either the latent class proportions.

E.2 EM ALGORITHM FOR MULTIVARIATE LATENT CLASS CUB MODEL

In this section, more details are given about the development of the EM algorithm for the Maximum Likelihood Estimates of the Multivaraite Latent Class CUB (MLC-CUB) model.

First, let's recall the MLC-CUB as defined in equation (6.2):

$$P(\mathbf{R} = \mathbf{r} \mid \theta) = \sum_{k=1}^{K} \omega_k \prod_{j=1}^{J} \Big[\pi_{jk} P_B(r_j \mid \xi_{jk}) + (1-\pi_{jk}) P_U(m_j)\Big], \tag{E.11}$$

where $\theta = (\xi', \boldsymbol{\pi}', \boldsymbol{\omega}')$ with $\xi = (\xi_{jk})_{j=1,\dots,J;k=1,\dots,K}$, $\boldsymbol{\pi} = (\pi_{jk})_{j=1,\dots,J;k=1,\dots,K}$, $\boldsymbol{\omega} = (\omega_k)_{k=1,\dots,K}$.

To develop the EM algorithm, the complete log-likelihood is defined by introducing two latent allocation variables in the model: $\mathbf{Z} : (Z_k)_{k=1,\dots,K}$ is a random variable distributed as a one-order Multinomial distribution, $\mathbf{Z} \sim \mathcal{M}(1; \omega_1, \dots, \omega_k)$, such that $Z_{ik} = 1$ if the ith rater preferences come from cluster k, and $Z_{ik} = 0$ otherwise. The second allocation variables $\mathbf{V} : (V_j)_{j=1,\dots,J}$, is dependent on the allocation variable $\mathbf{Z}$ and it is distributed as a Bernoulli with parameter π_{jk}. The observation $V_{ij} = 1$ if the preference of the ith rater for the jth item comes from a shifted Binomial of parameter ξ_{jk}, and $V_{ij} = 0$ if it belongs to a Uniform random variable. Therefore, the complete log-likelihood is defined as follows:

$$\ell_c(\theta \mid \mathbf{R}, \mathbf{Z}, \mathbf{V}) = \prod_{i=1}^{n} \prod_{k=1}^{K} \Big\{ \omega_k \prod_{j=1}^{J} \big[\pi_{jk} P_B(R_{ij} \mid \xi_{jk})\big]^{V_{ij}} \big[(\pi_{jk}) P_U(m_j)\big]^{(1-V_{ij})} \Big\}^{Z_{ik}} =$$

$$\sum_{i=1}^{n} \sum_{k=1}^{K} Z_{ik} \ln(\omega_k) + \sum_{j=1}^{J} Z_{ik} \ln \Big\{ \big[\pi_{jk} P_B(R_{ij} \mid \xi_{jk})\big]^{V_{ij}} \big[(1-\pi_{jk}) P_U(m_j)\big]^{(1-V_{ij})} \Big\} =$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} Z_{ik} \ln(\omega_k) + \sum_{j=1}^{J} Z_{ik} V_{ij} \ln \big[\pi_{jk} P_B(R_{ij} \mid \xi_{jk})\big] +$$
$$+ Z_{ik}(1 - V_{ij}) \ln \big[(1-\pi_{jk}) P_U(m_j)\big] =$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} Z_{ik} \ln(\omega_k) + \sum_{j=1}^{J} Z_{ik} V_{ij} \ln(\pi_{jk}) + Z_{ik} V_{ij} \ln \big[P_B(R_{ij} \mid \xi_{jk})\big] +$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} Z_{ik} \ln(\omega_k) + \sum_{j=1}^{J} Z_{ik} V_{ij} \ln(\pi_{jk}) + Z_{ik} V_{ij} \ln\left[P_B(R_{ij} \mid \xi_{jk})\right] +$$
$$+ Z_{ik}(1 - V_{ij}) \ln(1 - \pi_{jk}) + Z_{ik}(1 - V_{ij}) \ln\left[P_U(m_j)\right] =$$

$$\ell_c(\theta \mid \mathbf{R}, \mathbf{Z}, \mathbf{V}) = \sum_{i=1}^{n} \sum_{k=1}^{K} Z_{ik} \Big\{ \ln(\omega_k) + \sum_{j=1}^{J} V_{ij} \big[\ln(\pi_{jk}) + \ln\left[P_B(R_{ij} \mid \xi_{jk})\right] \big] + \sum_{j=1}^{J} (1 - V_{ij}) \big[\ln(1 - \pi_{jk}) + \ln P_U(m_j) \big] \Big\}. \tag{E.12}$$

Starting from the initial set of parameters generated at random, $\theta^{(0)} = (\boldsymbol{\xi}, \boldsymbol{\pi}, \boldsymbol{\omega})^{(0)}$, the algorithm alternates between an expectation step and a maximization step.

E-STEP (t-TH ITERATION) Consists of evaluating the expected value of the complete log-likelihood conditioned on the observed data $\mathbf{r}$ and on the parameters $\theta^{(t)})$ computed in the previous step.

Therefore, computing the conditional expectation of Equation (E.12), for all $i = 1, \ldots, n$, $j = 1, \ldots, J$, $k = 1, \ldots K$, lead to compute the conditional expectation of Z_{ik} as follows:

$$\mathbb{E}_{Z_{ik} \mid \mathbf{r}_i; \theta^{(t)}} =$$
$$= \frac{\omega_k^{(t)} \prod_j \left[\pi_{jk}^{(t)} P_B(r_{ij} \mid \xi_{jk}^{(t)}) + (1 - \pi_{jk}^{(t)}) P_U(m_j) \right]}{\sum_{k'=1}^{K} \omega_{k'}^{(t)} \prod_j \left[\pi_{jk'}^{(t)} P_B(r_{ij} \mid \xi_{jk'}^{(t)}) + (1 - \pi_{jk'}^{(t)}) P_U(m_j) \right]}$$
$$= \tau_{ik}(\mathbf{r}_i; \theta^{(t)}) = \tau_{ik}^{(t)}.$$

Then, the conditional expectation of the product $Z_{ik} V_{ij}$ is computed:

$$\mathbb{E}(Z_{ik} V_{ij} \mid r_{ij}; \theta^{(t)}) =$$
$$= \frac{\pi_{jk}^{(t)} P_B(r_{ij} \mid \xi_{jk}^{(t)})}{\pi_{jk}^{(t)} P_B(r_{ij} \mid \xi_{jk}^{(t)}) + (1 - \pi_{jk}^{(t)}) P_U(m_j)} \cdot \tau_{ik}(\mathbf{r}_i; \theta^{(t)}) =$$
$$= \nu_{ik}(r_{ij}; \theta^{(t)}) \cdot \tau_{ik}(\mathbf{r}_i; \theta^{(t)}) = \eta_{ijk}^{(t)}.$$

M-STEP (t-TH ITERATION) The maximization step consists of computing the new maximum likelihood estimates $\theta^{(t+1)}$ of the parameters of the mixture, by maximizing according to θ the function:

$$Q(\theta, \theta^{(t)}) = \sum_{1=1}^{n} \sum_{k=1}^{K} \tau_{ik}^{(t)} \Big\{ \ln(\omega_k) + \sum_{j=1}^{J} \eta_{ijk}^{(t)} \big[\ln(\pi_{jk}) + \ln\left[P_B(r_{ij} \mid \xi_{jk})\right] \big] + \sum_{j=1}^{J} (1 - \eta_{ijk}^{(t)}) \big[\ln(1 - \pi_{jk}) + \ln P_U(m_j) \big] \Big\}, \tag{E.13}$$

Similarly as for the LC-CUB model, the terms of function (E.13) can be factorized as follows:

$$Q(\theta,\theta^{(t)}) = \underbrace{\sum_{i=1}^{n}\sum_{k=1}^{K}\tau_{ik}^{(t)}\left\{\sum_{j=1}^{J}\eta_{ijk}^{(t)}\ln(\pi_{jk})+(1-\eta_{ijk}^{(t)})\ln(1-\pi_{jk})\right\}}_{Q_1^{(t)}(\pi_{jk})} +$$
$$+\underbrace{\sum_{i=1}^{n}\sum_{k=1}^{K}\tau_{ik}^{(t)}\ln(\omega_k)}_{Q_2^{(t)}(\omega_k)} +$$
$$+\underbrace{\sum_{i=1}^{n}\sum_{k=1}^{K}\tau_{ik}^{(t)}\left\{\sum_{j=1}^{J}[\eta_{ijk}^{(t)}\ln\left[P_B(r_{ij}\mid\xi_{jk})\right]+(1-\eta_{ijk}^{(t)})\ln P_U(m_j)]\right\}}_{Q_3^{(t)}(\xi_{jk})}.$$

Therefore, the function (E.13) can be written as:

$$Q(\theta,\theta^{(t)}) = Q_1^{(t)}(\pi_{jk}) + Q_2^{(t)}(\omega_k) + Q_3^{(t)}(\xi_{jk}). \tag{E.14}$$

The estimators of the parameters π_{jk}, ω_k, and ξ_{jk} can be obtained by maximizing the functions $Q_1^{(t)}(\pi)$, $Q_2^{(t)}(\omega_k)$, and $Q_3^{(t)}(\xi_k)$, respectively.

The maximisation of the function $Q_1^{(t)}(\pi)$ with respect to π_{jk} allows us to obtain the updated estimate of the parameter:

$$\frac{\partial Q_1^{(t)}(\pi_{jk})}{\partial\pi_{jk}} = \sum_i\left[\frac{\tau_{ik}^{(t)}\eta_{ijk}^{(t)}}{\pi_{jk}} - \frac{\tau_{ik}^{(t)}(1-\eta_{ijk}^{(t)})}{1-\pi_{jk}}\right] = 0$$

The solution of this equation can be easily obtain by simple algebra:

$$\sum_i\left[\tau_{ik}^{(t)}\eta_{ijk}^{(t)} - \tau_{ik}^{(t)}\pi_{jk}\right] = 0$$

and thus:

$$\pi_{jk}^{(t+1)} = \frac{\sum_i \tau_{ik}^{(t)}\eta_{ijk}^{(t)}}{\sum_i \tau_{ik}^{(t)}}$$

The parameter ω_k is obtained through a constrained maximization of the function $Q_2^{(t)}(\omega_k)$ under the constraint $\sum_{k=1}^{K}\omega_k = 1$. Therefore, the following Lagrangian function has to be maximized:

$$\mathcal{L}(\omega_k;\lambda) = \sum_{i=1}^{n}\tau_{ik}^{(t)}\ln(\omega_k) - \lambda\left(\sum_{k=1}^{K}\omega_k - 1\right). \tag{E.15}$$

The first order derivative of the Lagrangian function is:

$$\frac{\partial\mathcal{L}(\omega_k;\lambda)}{\partial\omega_k} = \frac{\sum_{i=1}^{n}\tau_{ik}^{(t)}}{\omega_k} - \lambda = 0, \tag{E.16}$$

which is equal to:

$$\sum_{i=1}^{n} \tau_{ijk}^{t} = \lambda \omega_k.$$

The sum over k leads to:

$$\sum_{i=1}^{n} \sum_{k=1}^{K} \tau_{ik}^{(t)} = \lambda \sum_{k=1}^{K} \omega_k.$$

Considering that $\sum_{k=1}^{K} \omega_k = 1$, $\lambda = \sum_{i=1}^{n} \sum_{k=1}^{K} \tau_{ik}^{(t)}$ can be computed.

Therefore, the previous derivative becomes:

$$\frac{\sum_{i=1}^{n} \tau_{ik}^{(t)}}{\omega_k} - \sum_{i=1}^{n} \sum_{k=1}^{K} \tau_{ik}^{(t)} = 0,$$

thus:

$$\omega_k^{(t+1)} = \frac{\sum_{i=1}^{n} \tau_{ik}^{(t)}}{\sum_{i=1}^{n} \sum_{k=1}^{K} \tau_{ik}^{(t)}} = \frac{\sum_{i=1}^{n} \tau_{ik}^{(t)}}{n}. \tag{E.17}$$

The maximization of the function $Q_3^{(t)}(\xi_{jk})$ allows us to obtain the estimator of the parameter ξ_{jk}:

$$\frac{\partial Q_3^{(t)}(\xi_{jk})}{\partial \xi_{jk}} = \sum_{i} \tau_{ik}^{(t)} \eta_{ijk}^{(t)} \left[\frac{m_j - r_{ij}}{\xi_{jk}} - \frac{r_{ij} - 1}{1 - \xi_{jk}} \right] = 0$$

After some simple algebra, the following equation can be obtained:

$$\sum_{i} \tau_{ik}^{(t)} \eta_{ijk}^{(t)} \left[\frac{m_j}{\xi_{jk}} + \frac{1}{1 - \xi_{jk}} \right] = \sum_{i} \tau_{ik}^{(t)} \eta_{ijk}^{(t)} \left[\frac{r_{ij}}{\xi_{jk}(1 - \xi_{jk})} \right] = 0$$

The estimator of the parameter ξ_{jk} can be obtained by solving the previous equation:

$$\xi_{jk}^{(t+1)} = \frac{\sum_i \tau_{ik}^{(t)} \eta_{ijk}^{(t)} (m_j - r_{ij})}{\sum_i \tau_{ik}^{(t)} \eta_{ijk}^{(t)} (m_j - 1)}. \tag{E.18}$$

E.3 CODE FOR FITTING THE MULTIVARIATE LATENT CLASS CUB MODEL

```
EM <- function(R, m, K,
               tol = 10e-10, EM.iter = 20,
               max_iter = 1500) {
  library(dplyr)
  #------------------------------
  # Loglikelihood function
  #------------------------------

  LogLik.function <- function(omega.mat, pi.tens,
                              t, R, m, xi.tens,
                              K, J, n, uniform.j) {
    CUB.mix <- array(rep(NA, K*J*n), c(K,J,n))
```

```r
  for(i in 1:n){
   for (k in 1:K){
    for(j in 1:J){
     CUB.mix[k,j,i] <- pi.tens[k,j,t]*dbinom(R[i,j] - 1,
                                    m[j]-1,
                                    1 - xi.tens[k,j,t]) +
       (1-pi.tens[k,j,t])*uniform.j[i,j]
    }
   }
  }


  # Product over j
  multiCUB.j <- apply(CUB.mix, c(3, 1), prod)

  # Product with omega_k
  multiCUB.jk <- matrix(nrow = n, ncol = K)
  for (k in 1:K) {
   multiCUB.jk[,k] <- omega.mat[t,k]*multiCUB.j[,k]
  }

  MLCCUBk <- apply(multiCUB.jk, 1, sum)

  # log of the sum over K
  MLCCUBi <- log(MLCCUBk)
  LL <- sum(MLCCUBi)
  res <- list(multiCUB.w_k = multiCUB.jk,
           MLCCUBk = MLCCUBk,
           MLCCUBi = MLCCUBi,
           LL = LL)
  return(res)
}


#----------------------------------------
# Function for computing eta
#----------------------------------------

eta_ijk <- function(omega.mat, pi.tens,
              t, R, m, xi.tens,
              K, J, n, uniform.j) {

  eta.ijk <- array(rep(NA, K*J*n), c(K,J,n))
  for(i in 1:n){
   for (k in 1:K){
    for(j in 1:J){
     eta.ijk[k,j,i] <- pi.tens[k,j,t]*dbinom(R[i,j] - 1,
                        m[j]-1,
                          1 - xi.tens[k,j,t])/ (pi.tens[k,j,t]*dbinom(R[i,j] - 1,
                               m[j]-1, 1- xi.tens[k,j,t]) +
                               (1-pi.tens[k,j,t])*uniform.j[i,j])
    }
   }
```

```
  }

  return(eta.ijk)
}


results.list <- list()
max.LL <- NA
# set.seed(seed)
for (em in 1:EM.iter) {
  Niter <- 0
  iter <- 0
  t = 2
  m = m # Number of categories
  n = nrow(R) # Number of obs
  J = ncol(R) # Number of variables
  K = K # Number of clusters
  tol = tol

  max_retries <- 10# Numero massimo di tentativi
  retry_count <- 0# Contatore dei tentativi
  retry <- FALSE

  while (retry_count < max_retries) {
    tryCatch({

  #--------------------#
  #  INITIALIZATION #
  #--------------------#

  # Starting values of pi_jk
  # K rows, J columns, t blocks
  pi.tens = array(rep(NA, K*J*t),dim = c(K,J,t))
  pi.tens[,,t] <- matrix(runif(K*J),
                   ncol = J, nrow = K)

  pi.mat <- matrix(NA, nrow=t, ncol = (J*K))
  pi.mat[t, ] <- as.vector(pi.tens[,,t])
  colnames(pi.mat) <- paste0("pi.",
                      rep(1:J, each = K),
                      rep(1:K, times = J),
                      sep ="")

  # Starting values of xi_jk
  # K rows, J columns, t blocks
  xi.tens = array(rep(NA, K*J*t),dim =c(K,J,t))
  xi.tens[,,t] <- matrix(runif(K*J), ncol = J, nrow = K)

  xi.mat <- matrix(NA, nrow=t, ncol = (J*K))
  xi.mat[t, ] <- as.vector(xi.tens[,,t])
  colnames(xi.mat) <- paste0("xi.",
                      rep(1:J, each = K),
                      rep(1:K, times = J),
```

```
                      sep ="")

  # starting values of omega
  omega.mat = matrix(nrow = t, ncol = K)
  omega.mat[t,] <- rep(1/K,K)

  # Uniform distribution
  uniform.j = matrix(ncol = J, nrow = n)
  for (j in 1:J) {
   uniform.j[,j] <- rep(1/m[j], n)
  }

  #--------------------
  # LogLikelihood
  #--------------------
  LL <- -Inf
  LL.list <- LogLik.function(omega.mat,
                      pi.tens, t, R, m,
                      xi.tens, K, J, n,
                      uniform.j)
  LL[t] <- unlist(LL.list$LL)

  #-------------------------#
  #    EM ALGORITHM  #
  #-------------------------#

  while ((LL[t]-LL[t-1]) >= tol && iter < max_iter) {
   Niter <- t-1
   iter <- iter + 1
   #-----------------------------
   # EXPECTATION
   #-----------------------------

   tau.mat <- matrix(NA, nrow = n, ncol = K)
   for(i in 1:n){
    for (k in 1:K){
     tau.mat[i,k] = LL.list$multiCUB.w_k[i,k]/LL.list$MLCCUBk[i]
    }
   }

   eta.tens <- eta_ijk(omega.mat,
                  pi.tens, t, R, m,
                  xi.tens, K, J, n,
                  uniform.j)

   t = t+1

   #-----------------------------------
   # MAXIMIZATION
   #-----------------------------------

   # Pi parameter
   if (dim(pi.tens)[3] < t) {
```

```
170         depth_to_add <- t - dim(pi.tens)[3]
171         new_depth <- array(NA, dim = c(dim(pi.tens)[1], dim(pi.tens)[2],
                depth_to_add))
172         pi.tens <- abind::abind(pi.tens, new_depth, along = 3)
173       }
174
175       pi.tens.n <- array(NA, dim = c(K, J, n))
176       for (i in 1:n) {
177         pi.tens.n[, , i] <- tau.mat[i,] * eta.tens[, , i]
178       }
179
180       pi.tens[, , t] <- apply(pi.tens.n, c(1, 2), sum) / colSums(tau.mat)
181
182       if (t > nrow(pi.mat)) {
183         pi.mat <- rbind(pi.mat, matrix(NA, nrow = t - nrow(pi.mat), ncol = K*J))
184       }
185
186       pi.mat[t, ] <- as.vector(pi.tens[,,t])
187
188       # Xi parameters
189       if (dim(xi.tens)[3] < t) {
190         depth_to_add <- t - dim(xi.tens)[3]
191         xi.tens <- abind::abind(xi.tens, array(NA, dim = c(dim(xi.tens)[1],
                dim(xi.tens)[2], depth_to_add)), along = 3)
192       }
193
194       for (j in 1:J) {
195         for (k in 1:K) {
196           xi.tens[k, j, t] <-
197             sum(tau.mat[, k] * eta.tens[k, j, ] * (m[j] - R[, j])) /
198             sum(tau.mat[, k] * eta.tens[k, j, ] * (m[j] - 1))
199         }
200       }
201
202       if (t > nrow(xi.mat)) {
203         xi.mat <- rbind(xi.mat, matrix(NA, nrow = t - nrow(xi.mat), ncol = K * J))
204       }
205
206       xi.mat[t, ] <- as.vector(xi.tens[, , t])
207
208       # Omega parameter
209
210       if (t > nrow(omega.mat)) {
211         omega.mat <- rbind(omega.mat,
212                            matrix(NA, nrow = t - nrow(omega.mat),
213                                   ncol = K))
214       }
215
216       omega.mat[t,] <- apply(tau.mat,2, sum)/n
217
218       # LogLikelihood
219       LL.list <- LogLik.function(omega.mat,
220                                  pi.tens, t, R, m,
```

```
                        xi.tens, K, J, n,
                        uniform.j)
      LL[t] <- unlist(LL.list$LL)
    }

     }, error = function(e) {
       retry_count <- retry_count + 1

       # Set retry = T to repeat the while loop with different parameters
       iter <- 0
       retry <- TRUE
       cat("No convergence - Retry - N. of tries: ", retry_count, "\n")
     })
     if (retry) {
       retry <- FALSE # Reset retry to FALSE
       next # Skip to the next iteration
     }

     # If no error, break
     break
    }

    colnames(omega.mat) <- paste("w.", 1:K, sep = "")
    params_table = cbind(pi.mat, xi.mat, omega.mat, LL)
    colnames(tau.mat) <- paste("tau.", 1:K, sep = "")
    class <- max.col(tau.mat)

    results.list[[em]] <- list(params_table = params_table,
                        xi.conv = xi.mat,
                        pi.conv = pi.mat,
                        omega.conv = omega.mat,
                        xi.est = as.matrix(xi.tens[, , dim(xi.tens)[3]]),
                        pi.est = as.matrix(pi.tens[, , dim(pi.tens)[3]]),
                        omega.est = as.matrix(tail(omega.mat, 1)),
                        taus = tau.mat,
                        class = class,
                        LogLik_vec = LL,
                        LogLik = tail(LL,1),
                        Niter = iter,
                        AIC = 2*(K*J*2+K)-2*tail(LL,1),
                        BIC = (K*J*2+K)*log(n)-2*tail(LL,1))
    max.LL[em] <- tail(LL,1)
    cat("Number of clusters:", K, "- EM initialization number:",
       em, "- N. of iterations:", tail(Niter, 1), "\n")
  }

  best.result <- results.list[[which.max(max.LL)]]
  results <- best.result
  return(results)
}
```

BIBLIOGRAPHY

Agresti, Alan (2010). *Analysis of ordinal categorical data*. Vol. 656. John Wiley & Sons.

Agresti, Alan (2012). *Categorical data analysis*. Vol. 792. John Wiley & Sons.

Akaike, Hirotsugu (1973). "Information Theory and an Extension of the Maximum Likelihood Principle." In: *Proceedings of the 2nd International Symposium on Information Theory*. Budapest: Akademia Kiado, pp. 267 –281.

Andreis, Federico and Pier Alda Ferrari (2013). "On a copula model with CUB margins." In: *Quaderni di Statistica* 15, pp. 33–51.

Andreis, Federico, Pier Alda Ferrari, et al. (2013). "A proposal for the multidimensional extension of CUB models." In: *Cladag 2013: 9th meeting of the classification and data analysis group: book of abstracts*. CLEUP, pp. 15–18.

Arboretti, Rosa and Paolo Bordignon (2016). "Consumer preferences in food packaging: CUB models and conjoint analysis." In: *British Food Journal* 118.3, pp. 527–540.

Banfield, Jeffrey D and Adrian E Raftery (1993). "Model-based Gaussian and non-Gaussian clustering." In: *Biometrics*, pp. 803–821.

Benaglia, Tatiana, Didier Chauveau, and David R Hunter (2009). "An EM-like algorithm for semi-and nonparametric estimation in multivariate mixtures." In: *Journal of Computational and Graphical Statistics* 18.2, pp. 505–526.

Biasetton, Nicolò, Marta Disegna, Elena Barzizza, and Luigi Salmaso (2023). "A new adaptive membership function with CUB uncertainty with application to cluster analysis of Likert-type data." In: *Expert Systems with Applications* 213.118893.

Biasetton, Nicolò, Pierpaolo D'Urso, Marta Disegna, and Luigi Salmaso (2024). "Cub model-based clustering of Likert-type data with a tourist satisfaction application." In: *Annals of Operations Research*, pp. 1–16.

Biernacki, Christophe and Julien Jacques (2016). "Model-based clustering of multivariate ordinal data relying on a stochastic binary search algorithm." In: *Statistics and Computing* 26, pp. 929–943.

Birnbaum, Allan, Frederic M Lord, and Melvin R Novick (1968). "Statistical theories of mental test scores." In: *Some latent trait models and their use in inferring an examinee's ability. Addison-Wesley, Reading, MA*.

Bitgood, Stephen (2011). *Social Design in Museums: The Psychology of Visitor Studies: Collected Essays*. Vol. 1. MuseumsEtc, Edinburgh.

Bork, Riet van, Mijke Rhemtulla, Lourens J Waldorp, Joost Kruis, Shirin Rezvanifar, and Denny Borsboom (2021). "Latent variable models and networks: Statistical equivalence and testability." In: *Multivariate behavioral research* 56.2, pp. 175–198.

Borsboom, Denny, Gideon J Mellenbergh, and Jaap Van Heerden (2003). "The theoretical status of latent variables." In: *Psychological review* 110.2, p. 203.

Brentari, Eugenio, Maurizio Carpita, and Paola Zuccolotto (2006). *Qualità e cus- tomer satisfaction nei servizi: un'indagine statistica nelle scuole dell'infanzia del Comune di Brescia*. Vol. 472. FrancoAngeli.

Brentari, Eugenio, Marica Manisera, and Paola Zuccolotto (2018). "Modelling preceived variety in a choice process with nonlinear cub." In: *Proceedings of the International Conference ASMOD2018*. Federico II University Press, pp. 69–76.

Cafarelli, Barbara and Corrado Crocetta (2016). "An evaluation of the student satisfaction based on CUB models." In: *Topics in Theoretical and Applied Statistics*. Springer, pp. 73–83.

Cafarelli, Barbara, Vittoria Pilone, Amalia Conte, Daniela Gammariello, and Matteo Alessandro Del Nobile (2015). "Development of consumer acceptable products using CUB analysis: an example with burgers from dairy cattle." In: *Journal of Sensory Studies* 30.5, pp. 413–424.

Capecchi, Stefania, Isabella Endrizzi, Flavia Gasperi, and Domenico Piccolo (2016). "A multi-product approach for detecting subjects' and objects' covariates in consumer preferences." In: *British Food Journal* 118.3, pp. 515–526.

Cappelli, Carmela, Rosaria Simone, and Francesca Di Iorio (2019). "CUBREMOT: a tool for building model-based trees for ordinal responses." In: *Expert systems with applications* 124, pp. 39–49.

Carpita, Maurizio, Enrico Ciavolino, and Mariangela Nitti (2019). "The MIMIC–CUB model for the prediction of the economic public opinions in Europe." In: *Social Indicators Research* 146, pp. 287–305.

Celeux, Gilles and Gérard Govaert (1995). "Gaussian parsimonious clustering models." In: *Pattern recognition* 28.5, pp. 781–793.

Cerulli, Giovanni, Rosaria Simone, Francesca Di Iorio, Domenico Piccolo, and Christopher F Baum (2022). "Fitting mixture models for feeling and uncertainty for rating data analysis." In: *The Stata Journal* 22.1, pp. 195–223.

Cho, Jun Dong (2021). "A study of multi-sensory experience and color recognition in visual arts appreciation of people with visual impairment." In: *Electronics* 10.4, pp. 470–506.

Cicia, Gianni, Marcella Corduas, Teresa Del Giudice, and Domenico Piccolo (2010). "Valuing consumer preferences with the CUB model: a case study of fair trade coffee." In: *International Journal on Food System Dynamics* 1.1, pp. 82–93.

Coleman, Dan, Xioapeng Dong, Johanna Hardin, David M Rocke, and David L Woodruff (1999). "Some computational issues in cluster analysis with no a priori metric." In: *Computational Statistics & Data Analysis* 31.1, pp. 1–11.

Colombi, Roberto and Sabrina Giordano (2016). "A class of mixture models for multidimensional ordinal data." In: *Statistical Modelling* 16.4, pp. 322–340.

Congdon, Peter (2005). *Bayesian models for categorical data*. John Wiley & Sons.

Corduas, Marcella (2010). "Assessing similarity of rating distributions by Kullback-Leibler divergence." In: *Classification and multivariate analysis for complex data structures*. Springer, pp. 221–228.

Corduas, Marcella (2015). "Analyzing bivariate ordinal data with CUB margins." In: *Statistical Modelling* 15.5, pp. 411–432.

Corduas, Marcella, Luciano Cinquanta, and Corrado Ievoli (2013). "The importance of wine attributes for purchase decisions: A study of Italian consumers' perception." In: *Food Quality and Preference* 28.2, pp. 407–418.

Corduas, Marcella et al. (2011). "Modelling correlated bivariate ordinal data with CUB margins." In: *Quaderni di statistica* 13, pp. 109–119.

Coretto, Pietro and Christian Hennig (2011). "Maximum likelihood estimation of heterogeneous mixtures of Gaussian and uniform distributions." In: *Journal of Statistical Planning and inference* 141.1, pp. 462–473.

Corneli, Marco, Charles Bouveyron, and Pierre Latouche (2020). "Co-clustering of ordinal data via latent continuous random variables and not missing at random entries." In: *Journal of Computational and Graphical Statistics* 29.4, pp. 771–785.

D'Elia, Angela (2000). "A Shifted Binomial Model for Rankings." In: *Proceedings of the 15th International Workshop on Statistical Modelling*. Bilbao: Servicio Editorial de la Universidad del Pais Vasco, pp. 412 –416.

D'Elia, Angela and Domenico Piccolo (2005). "A mixture model for preferences data analysis." In: *Computational Statistics & Data Analysis* 49.3, pp. 917–934.

Dawis, Renè V. (1987). "Scale construction." In: *Journal of Counseling Psychology* 34.4, pp. 481–489.

Deldossi, Laura and Roberta Paroli (2012). "Inference on the CUB model: an MCMC approach." In: *Classification and Data Mining*. Springer, pp. 19–26.

Deldossi, Laura and Roberta Paroli (2015). "Bayesian variable selection in a class of mixture models for ordinal data: a comparative study." In: *Journal of Statistical Computation and Simulation* 85.10, pp. 1926–1944.

Dempster, Arthur P, Nan M Laird, and Donald B Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm." In: *Journal of the royal statistical society: series B (methodological)* 39.1, pp. 1–22.

Everitt, Brian S (1984). *Introduction to Latent Variable Models*. London, England.

Falk, John H and Lynn D Dierking (2016). *The museum experience revisited*. Routledge. ISBN: 9781611320459.

Fasola, Salvatore and Mariangela Sciandra (2015). "New flexible probability distributions for ranking data." In: *Advances in statistical models for data analysis*. Springer, pp. 117–124.

Gambacorta, Romina and Maria Iannario (2013). "Measuring job satisfaction with CUB models." In: *Labour* 27.2, pp. 198–224.

Gottard, Anna, Maria Iannario, and Domenico Piccolo (2016). "Varying uncertainty in CUB models." In: *Advances in Data Analysis and Classification* 10, pp. 225–244.

Grilli, Leonardo, Maria Iannario, Domenico Piccolo, and Carla Rampichini (2014). "Latent class CUB models." In: *Advances in Data Analysis and Classification* 8, pp. 105–119.

Grün, Bettina and Sara Dolnicar (2016). "Response style corrected market segmentation for ordinal data." In: *Marketing Letters* 27, pp. 729–741.

Hagenaars, Jacques A and Allan L McCutcheon (2002). *Applied latent class analysis*. Cambridge University Press.

Hambleton, Ronald K and Linda L Cook (1977). "Latent trait models and their use in the analysis of educational test data." In: *Journal of educational measurement*, pp. 75–96.

Hancock, Gregory R and Ralph O Mueller (2013). *Structural equation modeling: A second course*. Iap.

Iannario, M. (2010a). "On the identifiability of a mixture model for ordinal data." In: *Metron* LXVIII, pp. 87–94.

Iannario, Maria (2009). "Fitting measures for ordinal data models." In: *Quaderni di statistica* 11, pp. 46–79.

Iannario, Maria (2010b). "On the identifiability of a mixture model for ordinal data." In: *Metron* 68.1, pp. 87–94.

Iannario, Maria (2012). "Modelling shelter choices in a class of mixture models for ordinal responses." In: *Statistical Methods & Applications* 21, pp. 1–22.

Iannario, Maria, Marica Manisera, Domenico Piccolo, and Paola Zuccolotto (2020). "Ordinal data models for no-opinion responses in attitude survey." In: *Sociological Methods & Research* 49.1, pp. 250–276.

Iannario, Maria, Domenico Piccolo, and Rosaria Simone (2020). *The R package CUB: a class of mixture models for ordinal rating data*.

Ingrassia, Salvatore, Simona C Minotti, and Giorgio Vittadini (2012). "Local statistical modeling via a cluster-weighted approach with elliptical distributions." In: *Journal of classification* 29, pp. 363–401.

Ip, Ryan HL and Ka Yui Karl Wu (2024). "A mixture distribution for modelling bivariate ordinal data." In: *Statistical Papers*, pp. 1–36.

Johnson, Valen E and James H Albert (2006). *Ordinal data modeling*. Springer Science & Business Media.

Jürgens, Uta Maria and Danko Nikolić (2012). "Ideaesthesia: Conceptual processes assign similar colours to similar shapes." In: *Translational Neuroscience* 3, pp. 22–27.

Lamonaca, Emilia, Barbara Cafarelli, Crescenza Calculli, and Caterina Tricase (2022). "Consumer perception of attributes of organic food in Italy: A CUB model study." In: *Heliyon* 8.3.

Lebret, Rémi, Serge Iovleff, Florent Langrognet, Christophe Biernacki, Gilles Celeux, and Gérard Govaert (2015). "Rmixmod: The R package of the model-based unsupervised, supervised, and semi-supervised classification Mixmod library." In: *Journal of Statistical Software* 67, pp. 1–29.

Likert, Rensis (1932). "A technique for the measurement of attitudes." In: *Archives of psychology*.

Lin, Tsung I, Jack C Lee, and Shu Y Yen (2007). "Finite mixture modelling using the skew normal distribution." In: *Statistica Sinica*, pp. 909–927.

Manisera, Marica and Paola Zuccolotto (2014a). "Modeling rating data with nonlinear CUB models." In: *Computational Statistics & Data Analysis* 78, pp. 100–118.

Manisera, Marica and Paola Zuccolotto (2014b). "Modeling "don't know" responses in rating scales." In: *Pattern Recognition Letters* 45, pp. 226–234.

Manisera, Marica and Paola Zuccolotto (2016). "Treatment of "don't know" responses in a mixture model for rating data." In: *Metron* 74, pp. 99–115.

Manisera, Marica and Paola Zuccolotto (2022). "A mixture model for ordinal variables measured on semantic differential scales." In: *Econometrics and Statistics* 22, pp. 98–123.

Manisera, Marica, Paola Zuccolotto, et al. (2014). "Nonlinear CUB models: the R code." In: *Stat & App*, pp. 205–223.

McCullagh, Peter (1980). "Regression models for ordinal data." In: *Journal of the Royal Statistical Society: Series B (Methodological)* 42.2, pp. 109–127.

McCullagh, Peter (2002). "What is a statistical model?" In: *The Annals of Statistics* 30.5, pp. 1225–1310.

McLachlan, Geoffrey J, Sharon X Lee, and Suren I Rathnayake (2019). "Finite mixture models." In: *Annual review of statistics and its application* 6.1, pp. 355–378.

McParland, Damien and Isobel Claire Gormley (2016). "Model based clustering for mixed data: clustMD." In: *Advances in Data Analysis and Classification* 10.2, pp. 155–169.

Nelder, John A and Roger Mead (1965). "A simplex method for function minimization." In: *The computer journal* 7.4, pp. 308–313.

Nunnally, Jum C and Ira H Bernstein (1994). "Psychometric Theory New York." In: *NY: McGraw-Hill*.

Osgood, Charles E (1962). "Studies on the generality of affective meaning systems." In: *American Psychologist* 17.1, p. 10.

Osgood, Charles Egerton, George J Suci, and Percy H Tannenbaum (1957). *The measurement of meaning*. 47. University of Illinois press.

Paulino, Carlos Daniel Mimoso and Carlos Alberto de Bragança Pereira (1994). "On identifiability of parametric statistical models." In: *Journal of the Italian Statistical Society* 3, pp. 125–151.

Peel, David and Geoffrey MacLahlan (2000). "Finite mixture models." In: *John & Sons*.

Peel, David and Geoffrey J McLachlan (2000). "Robust mixture modelling using the t distribution." In: *Statistics and computing* 10, pp. 339–348.

Piccolo, Domenico (2003). "On the moments of a mixture of uniform and shifted binomial random variables." In: *Quaderni di Statistica* 5.1, pp. 85–104.

Piccolo, Domenico (2006). "Observed information matrix for MUB models." In: *Quaderni di Statistica* 8.1, pp. 33–78.

Piccolo, Domenico (2015). "Inferential issues on CUBE models with covariates." In: *Communications in Statistics-Theory and Methods* 44.23, pp. 5023–5036.

Piccolo, Domenico (2018). "A new paradigm for rating data models." In: *Book of short papers SIS 2018*. Pearson Publisher New York, pp. 1–12.

Piccolo, Domenico and Rosaria Simone (2019). "The class of CUB models: statistical foundations, inferential issues and empirical evidence." In: *Statistical Methods & Applications* 28, pp. 389–435.

Piccolo, Domenico et al. (2003). "Computational issues in the EM algorithm for ranks model estimation with covariates." In: *Quaderni di Statistica* 5, pp. 27–48.

Preedalikit, Kemmawadee, Daniel Fernández, Ivy Liu, Louise McMillan, Marta Nai Ruscone, and Roy Costilla (2024). "Row mixture-based clustering with covariates for ordinal responses." In: *Computational Statistics* 39.5, pp. 2511–2555.

Punzo, Gennaro, Rosalia Castellano, and Mirko Buonocore (2018). "Job satisfaction in the "Big Four" of Europe: Reasoning between feeling and uncertainty through CUB models." In: *Social Indicators Research* 139, pp. 205–236.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Ranalli, Monia and Roberto Rocci (2017). "A model-based approach to simultaneous clustering and dimensional reduction of ordinal data." In: *psychometrika* 82, pp. 1007–1034.

Rand, William M (1971). "Objective criteria for the evaluation of clustering methods." In: *Journal of the American Statistical association* 66.336, pp. 846–850.

Remani, Courtney (2013). "Numerical methods for solving systems of nonlinear equations." In: *Lakehead University Thunder Bay, Ontario, Canada* 77.

Robitzsch, Alexander (2021). "About the equivalence of the latent D-scoring model and the two-parameter logistic item response model." In: *Mathematics* 9.13, p. 1465.

Sarmanov, Oleg Vasil'evich (1966). "Generalized normal correlation and two-dimensional Fréchet classes." In: *Doklady Akademii Nauk*. Vol. 168. 1. Russian Academy of Sciences, pp. 32–35.

Schwarz, Gideon (1978). "Estimating the dimension of a model." In: *The annals of statistics*, pp. 461–464.

Scrucca, Luca, Michael Fop, T Brendan Murphy, and Adrian E Raftery (2016). "mclust 5: clustering, classification and density estimation using Gaussian finite mixture models." In: *The R journal* 8.1, p. 289.

Selosse, Margot, Julien Jacques, and Christophe Biernacki (2021). "ordinalClust: An R Package to Analyze Ordinal Data." In: *The R Journal* 12.2, pp. 173–188.

Simner, Julia (2012). "Defining synaesthesia." In: *British journal of psychology* 103.1, pp. 1–15.
Simone, Rosaria (2021). "An accelerated EM algorithm for mixture models with uncertainty for rating data." In: *Computational Statistics* 36.1, pp. 691–714.
Simone, Rosaria (2022). "On finite mixtures of Discretized Beta model for ordered responses." In: *TEST* 31.3, pp. 828–855.
Simone, Rosaria (2023). "Uncertainty Diagnostics of Binomial Regression Trees for Ordered Rating Data." In: *Journal of Classification* 40, pp. 1–27.
Simone, Rosaria, Carmela Cappelli, and Francesca Di Iorio (2019). "Modelling marginal ranking distributions: the uncertainty tree." In: *Pattern Recognition Letters* 125, pp. 278–288.
Simone, Rosaria and Gerhard Tutz (2018). "Modelling uncertainty and response styles in ordinal data." In: *Statistica Neerlandica* 72.3, pp. 224–245.
Simone, Rosaria, Gerhard Tutz, and Maria Iannario (2020). "Subjective heterogeneity in response attitude for multivariate ordinal outcomes." In: *Econometrics and Statistics* 14, pp. 145–158.
Stevens, Stanley Smith (1946). "On the theory of scales of measurement." In: *Science* 103.2684, pp. 677–680.
Tamhane, Ajit, Bruce Ankenman, and Ying Yang (2002). "The beta distribution as a latent response model for ordinal data (I): estimation of location and dispersion parameters." In: *Journal of Statistical Computation and Simulation* 72.6, pp. 473–494.
Titterington, David Michael, Adrian FM Smith, and Udi E Makov (1985). *Statistical analysis of finite mixture distributions*. Wiley.
Tourangeau, Roger, Lance J Rips, and Kenneth Rasinski (2000). *The psychology of survey response*. Cambridge University Press.
Tripathi, Ram C, Ramesh C Gupta, and John Gurland (1994). "Estimation of parameters in the beta binomial model." In: *Annals of the Institute of Statistical Mathematics* 46, pp. 317–331.
Tutz, Gerhard and Micha Schneider (2019). "Flexible uncertainty in mixture models for ordinal responses." In: *Journal of Applied Statistics* 46.9, pp. 1582–1601.
Tutz, Gerhard, Micha Schneider, Maria Iannario, and Domenico Piccolo (2017). "Mixture models for ordinal responses to account for uncertainty of choice." In: *Advances in Data Analysis and Classification* 11, pp. 281–305.
Ursino, Moreno (2014). "Ordinal data: a new model with applications." In: *PhD Thesis, Politecnico di Torino, Italy*.
Ursino, Moreno and Mauro Gasparini (2018). "A new parsimonious model for ordinal longitudinal data with application to subjective evaluations of a gastrointestinal disease." In: *Statistical methods in medical research* 27.5, pp. 1376–1393.
Walker, Strother H and David B Duncan (1967). "Estimation of the probability of an event as a function of several independent variables." In: *Biometrika* 54.1-2, pp. 167–179.
Wolfram Research, Inc. (2024). *Mathematica, Version 14.0*. Champaign, IL. URL: `https://www.wolfram.com/mathematica`.
Zhou, Hua and Kenneth Lange (2009). "Rating movies and rating the raters who rate them." In: *The American Statistician* 63.4, pp. 297–307.