Atlantis Studies in Uncertainty Modelling, volume 3

Joint Proceedings of the 19th World Congress of the International Fuzzy Systems Association (IFSA), the 12th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), and the 11th International Summer School on Aggregation Operators (AGOP)

fIRTree: An Item Response Theory Modeling of Fuzzy Rating Data

*Antonio Calcagnì

DPSS, University of Padova, via Venezia 8, 35131 Padova (Italy), `antonio.calcagni@unipd.it`

Abstract

In this contribution we describe a novel procedure to represent uncertainty in rating scales in terms of fuzzy numbers. Following the rationale of fuzzy conversion scale, we adopted a two-step procedure based on a psychometric model (i.e., Item Response Theory-based tree) to represent the process of answering survey questions. This provides a coherent context where fuzzy numbers, and the related fuzziness, can be interpreted in terms of decision uncertainty that usually affects the rater's response process. We reported results from a simulation study and an empirical application to highlight the characteristics and properties of the proposed approach.

Keywords: Fuzzy rating scale, Rating data, Item response theory, Decision uncertainty

1 Introduction

Over the recent years, fuzzy set theory and fuzzy statistics have been gradually adopted for modeling fuzziness in human rating data [7]. Typically, rating data emerge from the measurement of attitudes, motivations, satisfaction, and personality and sociodemographic constructs. In all these cases, as the measurement process involves human raters, data are often affected by fuzziness or imprecision [1]. Fuzziness in rating data has multiple sources, including the decision uncertainty that affects raters and their response process. Indeed, it is well-known that rating questions like "I am satisfied with my current work", using a graded discrete scale from "strongly disagree" to "strongly agree", entails some degrees of decision uncertainty in human raters. Fuzzy rating scales have been widely adopted with the purpose of quantifying fuzziness from rating data. Mainly, they can be constructed based on two different rationales, one requiring dedicated rating procedures by means of which fuzziness is measured explicitly (e.g., see [9]) or implicitly (e.g., see [6]), and the other using fuzzy conversion systems through which crisp rating data are converted into fuzzy data either by means of predefined fuzzy systems (e.g., see [18]) or on the basis of statistically-oriented models (e.g., see [20, 12]). Despite their differences, both the approaches aim at modeling fuzziness of rating data.

In this contribution, we describe fIRTree, a novel statistically-oriented procedure to model fuzziness from crisp rating data. The aim is to provide an approach to fuzzy rating data that incorporates a stage-wise cognitive formalization of the process that human raters use to answer survey questions. To this end, IRTrees are adopted, a novel class of Item Response Theory (IRT) models that formalize the steps needed by raters to answer multiple choice questions [3, 5]. In particular, fIRTree adopts a two-stage modeling system where IRTrees are fit on crisp data first, and then the estimated parameters are used as a bulding block for deriving triangular fuzzy numbers $\text{Trg}(c, l, r, \omega)$ for each rater and item combination. Although several alternatives are available, we resorted to adopt four-parameter triangular fuzzy number $\text{Trg}(c, l, r, \omega)$ because of their flexibility in representing decision rating uncertainty [10, 17].

The remainder of this contribution is organized as follows. Section 2 describes the rationale of IRTrees and provides the formal notation used throughout this paper. Section 3 presents the proposed fIRTree approach. Section 4 illustrates the results of a simulation study designed to assess the features of fIRTree whereas Section 5 reports a brief case study where fIRTree is applied on empirical data. Finally, Section 6 concludes this paper and provides a summary of the current findings.

Note that all the materials such as algorithms and datasets used throughout this paper are available to download at `https://github.com/antcalcagni/firtree/`.

2 IRTree models

Item Response Theory (IRT) trees are conditional linear models where rating data are modeled in terms of binary trees. They formalize the rating process as a sequence of conditional stages going through the tree to end nodes: in this schema, intermediate nodes represent specific cognitive components of the response process whereas terminal nodes represent the possible outcomes of the decision process. Figure 1 depicts the simplest IRT tree model for a rating scale with five choices (e.g., from 1: "strongly disagree"; to 5: "strongly agree"). The adoption of this schema would allow for the modeling of those situations where raters first decide whether or not provide their responses (Z_1) and, then, decide on the direction (Z_2) and strength of their answers (Z_3, Z_4). The probability of uncertain responses (i.e., $Y = 3$: "neither agree, neither disagree") is simply given by the probability to activate the first stage of the decision process, namely $\mathbb{P}(Y = 0) = \mathbb{P}(Z_1; \boldsymbol{\theta}_1)$. On the contrary, the probability of a negative extreme response (i.e., $Y = 1$: "strongly disagree") is computed as $\mathbb{P}(Y = 1) = \mathbb{P}(Z_1; \boldsymbol{\theta}_1)(1 - \mathbb{P}(Z_2; \boldsymbol{\theta}_2))(1 - \mathbb{P}(Z_3; \boldsymbol{\theta}_3))$. In this case, the latent random variables $\{Z_1, \dots, Z_4\}$ govern the sub-processes of the rater's response.
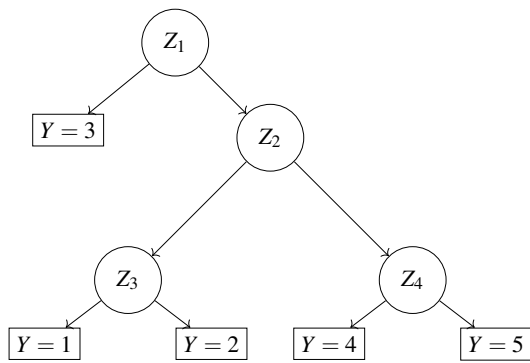

Figure 1: Example of IRTree used to model a common rating scale with five response categories.

As for more general IRT models, the probability to agree $\mathbb{P}(Y \in \{4,5\})$ or disagree $\mathbb{P}(Y \in \{1,2\})$ with an item can be represented as a function of a rater's latent trait and the specific content of the item [3]. More formally, let $i \in \{1, \dots, I\}$ and $j \in \{1, \dots, J\}$ be the indices for raters and items, respectively. Then, the final response variable $Y_{ij} \in \{1, \dots, m, \dots, M\} \subset \mathbb{N}$, with M being the maximum number of response cat-

egories, can be decomposed into N Boolean variables $Z_{ijn} \in \{0,1\}$, with $n \in \{1, \dots, N\}$ denoting the nodes of the tree. For a generic rater-item pair (i, j), the IRTree consists of the following equations:

$$\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta) \tag{1}$$

$$\pi_{ijn} = \mathbb{P}(Z_{ijn} = 1; \boldsymbol{\theta}_n) = \frac{\exp(\eta_{in} + \alpha_{jn})}{1 + \exp(\eta_{in} + \alpha_{jn})} \tag{2}$$

$$Z_{ijn} \sim \mathcal{B}er(\pi_{ijn}) \tag{3}$$

where $\boldsymbol{\theta}_n = \{\boldsymbol{\alpha}_j, \boldsymbol{\beta}_i\}$, with the arrays $\boldsymbol{\alpha}_j \in \mathbb{R}^N$ and $\boldsymbol{\eta}_i \in \mathbb{R}^N$ denoting the easiness of the item and the rater's latent trait. As for any IRT model, latent traits are modeled using a Gaussian distribution centered on zero and with covariance matrix $\boldsymbol{\Sigma}_\eta$. Overall, the probability for a given response category $\mathbb{P}(Y = m; \boldsymbol{\theta})$ is given by recursion from the single-branch probability equation as follows:

$$\begin{aligned}
\mathbb{P}(Y_{ij} = m) &= \prod_{n=1}^{N} \mathbb{P}(Z_{ijn} = t_{mn}; \boldsymbol{\theta}_n)^{t_{mn}} \\
&= \prod_{n=1}^{N} \left(\frac{\exp(\eta_{in} + \alpha_{jn}) t_{mk}}{1 + \exp(\eta_{in} + \alpha_{jn})} \right)^{\delta_{mn}}
\end{aligned} \tag{4}$$

where $t_{mn} \in \{0, 1, \mathrm{NA}\}$ is the entry of the mapping matrix $T_{M \times N}$ with $t_{mn} = 1$ indicating a connection from the m-th response category to the n-th node, $t_{mn} = 0$ or $t_{mn} = \mathrm{NA}$ indicating no connection at all, whereas $\delta_{mn} = 0$ if $t_{mn} = \mathrm{NA}$ and $\delta_{mn} = 1$ otherwise. IRTree models can be estimated either by means of general methods used for generalized linear mixed models, such as restricted or marginal maximum likelihood [3, 8], or using procedures for multidimensional IRT models, such as expectation-maximization algorithms [11]. By and large, IRTrees are flexible enough to model (i) simple situations like those requiring unidimensional latent variables (a single η for each node of the tree) or common item effects (a single α for each node of the tree), (ii) more complex scenario involving multidimensional high-order latent variables [3], and (iii) more general schemata requiring graded rating scales with more or less than five response categories [5].

3 fIRTree model for rating data

Once an IRTree has been fit to a matrix $\mathbf{Y}_{I \times J}$ of crisp rating data and IRT parameters $\hat{\boldsymbol{\theta}} = \{\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\alpha}}\}$ have been estimated, fuzzy rating data can be obtained by computing the probabilistic model of the rater's multiverse $\mathcal{U}_{ij}$. This reflects the heterogeneity in the rater's pattern of responses and models all the possible responses the rater would reach if he or she could repeatedly answer a question in the same time and under the same

conditions. In this context, decision uncertainty is reflected by the estimated transition probabilities $\mathbb{P}(Z_n)$ $(n = 1, \ldots, N)$ under the general rule that more certain responses require easier transitions among the nodes of the three. Unlike other fuzzy conversion scales grounded on unidimensional IRT models (e.g., see [20]), our approach uses $\mathcal{U}_{ij}$ in order to map fuzzy numbers to the latent response process underlying the observed crisp responses. In this contribution, we used the four-parameter triangular fuzzy numbers [10, 17]:

$$\xi_{\widetilde{A}}(y; c, l, r, \omega) = \left(1 + \left(\frac{c-y}{y-l}\right)^{\omega}\right)^{-1} \cdot \mathbb{1}_{(l,c)}(y) \quad (5)$$

$$+ \left(1 + \left(\frac{r-y}{y-c}\right)^{-\omega}\right)^{-1} \cdot \mathbb{1}_{(c,r)}(y)$$

where $l < c < r$ and $\omega \in \mathbb{R}_0^+$. Unlike ordinary triangular fuzzy numbers, by adding the parameter ω to intensify ($\omega < 1$) or reduce ($\omega > 1$) the fuzziness of the fuzzy set they allow for a flexible representation of the fuzziness present in rating responses.

Given a pair (i, j), fIRTree consists of the following steps:

1. Get the estimates $\hat{\boldsymbol{\eta}}_{N \times 1}$ and $\hat{\boldsymbol{\alpha}}_{N \times 1}$ and plug-in them into Eq. (4) to get $\widehat{\mathbb{P}}(Y = m)$ for each $m \in \{1, \ldots, M\}$. This is the probabilistic model for the rater's multiverse $\mathcal{U}_{ij}$.

2. Compute expected value and variance for $\mathcal{U}_{ij}$:

$$c_{ij} = \sum_{y \in \{1, \ldots, M\}} = y \cdot \widehat{\mathbb{P}}(Y = y) \quad (6)$$

$$s_{ij} = \sum_{y \in \{1, \ldots, M\}} = (y - c_{ij})^2 \cdot \widehat{\mathbb{P}}(Y = y) \quad (7)$$

3. Compute left and right spreads of triangular fuzzy numbers using the following link equations [19]:

$$l_{ij} = c_{ij} - h_2, \quad r_{ij} = c_{ij} - h_2 + h_1$$
$$h_1 = \sqrt{3.5 v_{ij} - 3(c_{ij} - \mu_{ij})^2} \quad (8)$$
$$h_2 = \frac{1}{2}(h_1 + 3c_{ij} - 3\mu_{ij})$$
$$\mu_{ij} = (1 + c_{ij}(1/s_{ij}))/(2 + (1/s_{ij}))$$

4. Compute the intensification parameter as:

$$\omega_{ij} = \sum_{m=1}^{M} \widehat{\mathbb{P}}(Y_{ij} = m)^2 \quad (9)$$

5. The 4-tuple $\{c_{ij}, l_{ij}, r_{ij}, \omega_{ij}\}$ is used to compute the fuzzy number $\tilde{y}_{ij}$ associated to the observed rater's response y_{ij}.

Note that Eq. (9) is based on the difficulty of responding to an item, a measure which indicates the presence of uncertainty in the response process [14]. In this representation, when $\widehat{\mathbb{P}}(Y_{ij})$ approaches the uniform distribution, ω_{ij} will reach its minimum value and the fuzzy set will be intensified in order to reach the maximum fuzziness. By contrast, when $\widehat{\mathbb{P}}(Y_{ij})$ approaches a degenerated distribution where a single category has probability equals to one, ω_{ij} will reach the threshold of one and the fuzzy set will turn to the ordinary triangular shape. Hence, the cases where $\omega_{ij} < 1$ progressively indicate that the rater would be very hesitant in responding and he or she would be expected to show a relatively large uncertainty to make a final response. It should also be noted that fuzzy numbers are defined over the space of the means $\Omega(Y) \in [1, \ldots, M]$ instead of being defined over the discrete domain of the responses $\mathcal{R}(Y) \in \{1, \ldots, M\}$.

4 Simulation study

The aim of this brief simulation study is twofold. First, we will evaluate whether fIRTree can accurately recover fuzzy numbers. Second, we will provide an external validity check on the results provided by fIRTree in recovering decision uncertainty from crisp rating data. In particular, we assessed fIRTree using a controlled scenario based on simulated faking data (SGR) [13] where distinct levels of decision uncertainty were progressively increased from a baseline condition. Since it is well known that faking behaviors in rating situations may increase or decrease the overall decision uncertainty of the rater's response process, simulated faking can serve as a good candidate for studying uncertainty in rating process.[1] The whole simulation study has been performed on a remote HPC machine based on 16 Intel Xeon CPU E5-2630Lv3 1.80Ghz, 16x4 Gb Ram whereas computations and analyses have been performed in the R framework for statistical analyses.

Data generation model. Crisp rating data were generated using the simplest IRTree for five-point rating scales with a common latent trait and item parameters (i.e., η and α were the same across nodes) and the SGR model [13]. The latter allows for modulating the presence of faking patterns in the observed data. In particular, it defines the probability of a rating response Y_{ij} as being formed by a true unobserved component Y_{ij}^T and a faking component Y_{ij}^F. Simply stated, the

---

[1]Faking is a deliberate behavior through which respondents distort their responses towards ones they consider more favorable in order to give overly positive self-descriptions, to simulate physical or psychological symptoms as a way to obtain rewards, or to have access to advantageous work positions [21].

probability of answering an item is defined in terms of a conditional replacement distribution $\mathbb{P}(Y_{ij}^F|Y_{ij}^T;\boldsymbol{\theta})$, which models the hypothesized faking behavior of the respondent. The model $\mathbb{P}(Y_{ij}^F|Y_{ij}^T;\boldsymbol{\theta})$ is defined using a particular discrete beta distribution, which has been parameterized in order to represent a set of possible faking behaviors (e.g., faking-good, faking-bad) as well as their intensity (e.g., extreme, slight, uniform). For further information we refer the reader to [13]. For the purpose of this study, we manipulated one of the parameter of $\mathbb{P}(Y_{ij}^F|Y_{ij}^T;\boldsymbol{\theta})$, namely $\pi \in [0,1]$ which indicates the probability that an observed response will be replaced by a faking response.

Design. The study involved three factors: (i) $I \in \{50,100,150\}$, (ii) $J \in \{10,20\}$, (iii) $\pi \in \{0,0.25,0.50,0.75\}$. They were varied in a complete factorial design with a total of $3 \times 2 \times 4 = 24$ scenarios. For each combination, $B = 1000$ samples were generated which yielded to $1000 \times 24 = 24000$ new data as well as an equivalent number of parameters. For each condition, the number of response categories and decision nodes were held fixed ($M = 5$, $N = M - 1 = 4$) whereas the IRTree for five-point rating scales was adopted (see Fig. 1). Note that, the scenarios with $\pi = 0$ correspond to the baseline condition with no faking perturbation. This case has been used to assess the accuracy of fIRTree to recover fuzzy numbers (first aim of the study).

Measures. The accuracy of fIRTree to recover the parameters of the fuzzy numbers was measured using the average agreement index:

$$\text{PA} = \frac{1}{B}\sum_{b=1}^{B}\left(1 - \|\hat{\boldsymbol{\Theta}}^{(b)} - \boldsymbol{\Theta}^{(b)}\|^2 / \|\boldsymbol{\Theta}^{(b)}\|^2\right)$$

which indicates how much the recovered parameters $\hat{\boldsymbol{\Theta}}^{(b)}$ resemble the true parameters $\boldsymbol{\Theta}^{(b)}$ [16]. Values of PA closed to one indicate higher accuracy of the procedure in recovering the true parameters of the model. We computed PA indices for the parameters c, ω, and $r - l$ (total spread). The sensitivity of fIRTree to the distinct levels of decision uncertainty (as created by the SGR model) was instead measured by means of the average Kauffmann index:

$$K = \frac{1}{B}\sum_{b=1}^{B}\left(\frac{2}{\text{card}\left(\xi_{\tilde{A}}(y;c^{(b)},l^{(b)},r^{(b)},\omega^{(b)})\right)}\right.$$
$$\left.\cdot \sum_y |\xi_{\tilde{A}}(y;c^{(b)},l^{(b)},r^{(b)},\omega^{(b)}) - \delta(y)|\right)$$

with $\delta(y) = 1$ if $\xi_{\tilde{A}}(y;c^{(b)},l^{(b)},r^{(b)},\omega) \geq 0.5$ and zero otherwise. The index is a measure of the fuzziness of a set and it approaches one when the set is maximally fuzzy.

Procedure. Let i_h, j_t, π_q be distinct levels of factors I, J, π. Then, crisp rating data were generated according to the following procedure:

(a) $\boldsymbol{\eta}_{i_h} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$, $\boldsymbol{\alpha}_{j_t} \sim \mathcal{N}(\mathbf{1}\alpha_0,\mathbf{I}\sigma_\alpha)$, with $\alpha_0 = -1.75$ and $\sigma_\alpha = 0.25$

(b) Crisp rating data were simulated as follows: $y_{ij} \sim \mathcal{M}ult(M,\mathbb{P}(Y_{i_h,j_t}))$, with $\mathbb{P}(Y_{i_h,j_t})$ being computed via Eq. (4) using $\boldsymbol{\eta}_{i_h}$ and $\boldsymbol{\alpha}_{j_t}$, for each $i = 1,\ldots,i_h$ and $j = 1,\ldots,j_t$. Note that $\mathbb{P}(Y_{i_h,j_t})$ was also used to compute the true triangular fuzzy numbers by means of the procedure described in Section 3.

(c) For each $i = 1,\ldots,i_h$ and $j = 1,\ldots,j_t$, observed crisp data y_{ij} were perturbed by the SGR model using the faking parameter π_q

(d) The generated matrices of unperturbed ($\pi_q = 0$) and perturbed ($\pi_q > 0$) response data $\mathbf{Y}_{i_h \times j_t}$ were analysed using fIRTree. The nested decision tree with five response categories was adopted (see Figure 1). Since $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$ were simulated using the simplest model where latent traits and item parameters are invariant across nodes (e.g., see [3]), an IRTree with a common latent trait and common parameters was defined using the `IRTrees` R library [3]. The `glmmTMB` R package [4] was used to estimate the model parameters via marginal maximum likelihood. Once estimates were obtained, triangular fuzzy numbers were computed using the procedure described in Section 3, which yielded to four new matrices, namely $\mathbf{C}_{i_j \times j_t}$, $\mathbf{L}_{i_j \times j_t}$, $\mathbf{R}_{i_j \times j_t}$, $\mathbf{W}_{i_j \times j_t}$.

Results. Tables 1-2 show the results of the simulation study. With regards to the first aim of the study, the accuracy of fIRTree in recovering the true parameters of fuzzy numbers was higher in all the conditions. As expected, the accuracy increased as a function of sample size I and number of items J. Overall, fIRTree showed good accuracies in recovering the true parameters of the fuzzy numbers. With regards to the second aim of the study, when compared to the baseline with no faking condition ($\pi = 0$ in Table 2), fuzzy numbers computed via fIRTree behave according to the faking perturbation. As increasing in faking perturbation indicate higher decision uncertainty in rating data, fuzzy numbers progressively increased their fuzziness in the expected direction. Importantly, these results are unaffected by sample size i_h or number of items j_t. All in all, these results suggest that fIRTree accurately recovered fuzzy numbers and decision uncertainty when this is present in rating data.

		C	$R-L$	W
	$I = 50$	0.779 (0.088)	0.897 (0.038)	0.965 (0.013)
$J = 10$	$I = 150$	0.839 (0.027)	0.916 (0.017)	0.981 (0.003)
	$I = 500$	0.851 (0.016)	0.921 (0.011)	0.983 (0.002)
	$I = 50$	0.805 (0.078)	0.908 (0.035)	0.965 (0.014)
$J = 20$	$I = 150$	0.883 (0.028)	0.942 (0.014)	0.985 (0.004)
	$I = 500$	0.9 (0.018)	0.951 (0.008)	0.988 (0.002)

Table 1: Simulation study: Accuracy of fIRTree to recover fuzzy numbers computed in terms of PA index [16]. Values of PA closed to one indicate that the model accurately recover the true parameters of the fuzzy numbers. Standard deviation over $B = 1000$ samples is reported in parenthesis for each scenario. Note that C indicates modes of the fuzzy sets, $R-L$ indicates the total spread of the set, whereas W stands for the intensification parameter.

		$\pi = 0$	$\pi = 0.25$	$\pi = 0.5$	$\pi = 0.75$
	$I = 50$	0.617 (0.075)	0.724 (0.041)	0.784 (0.019)	0.815 (0.01)
$J = 10$	$I = 150$	0.602 (0.058)	0.716 (0.026)	0.779 (0.01)	0.812 (0.006)
	$I = 500$	0.603 (0.062)	0.717 (0.028)	0.78 (0.008)	0.813 (0.004)
	$I = 50$	0.613 (0.079)	0.72 (0.045)	0.781 (0.022)	0.814 (0.01)
$J = 20$	$I = 150$	0.599 (0.062)	0.713 (0.03)	0.776 (0.011)	0.811 (0.006)
	$I = 500$	0.6 (0.066)	0.714 (0.032)	0.777 (0.011)	0.812 (0.004)

Table 2: Simulation study: Kauffman index of the computed fuzzy numbers as a function of the intensity of the faking perturbation π. Values of the index closed to one indicate a higher degree of fuzziness in the fuzzy sets. Standard deviation over $B = 1000$ samples is reported in parenthesis for each scenario.

5 Case study

In this section we will shortly describe the application of fIRTree on a real case study involving rating data from the Personality Need for Structure (PNS) questionnaire [15]. The questionnaire includes 11 items rated on a six-point rating scale (from "strongly disagree" to "strongly agree") and it was designed to measure a person's desire for certainty and decisiveness. The sample consists of $I = 497$ respondents, 58% women, 32.4 years old on average ($SD = 11.3$, range = $18 - 70$), all of them living in the United States. Data were originally collected by [5]. A subset of $J = 5$ items was selected for the sake of illustration.[2] The first step of the analysis consisted in the definition of an IRTree model for the six-point rating scale. To this end, we used the tree suggested by [2] (see Figure 2). It formally represents a response model where (i) the node M determines whether the respondent has a weak A_w or strong A_s attitude/trait toward the item, (ii) the node A_s activates the strength/extremity E of the rating response, by selecting either low values or higher
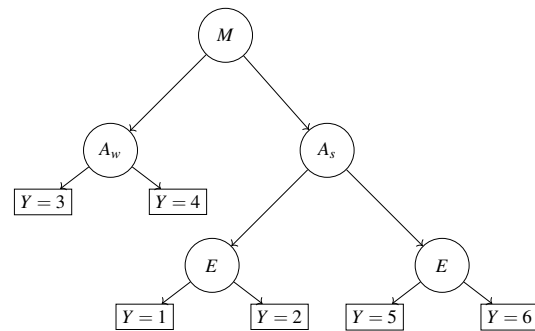

Figure 2: Case study: IRTree used for the six-point rating scale.

values of the scale.

The tree with $N = 4$ was defined using the R library `irtree` [3]. Next, the IRTree was fit on the data matrix $\mathbf{Y}_{497 \times 5}$ by means of the R library `glmmTMB` [4]. The estimated parameters $\hat{\boldsymbol{\eta}}_{I \times N}$ and $\hat{\boldsymbol{\alpha}}_{J \times N}$ (see Tables 3-4) were then used to compute the estimated probabilistic models of the multiverses $\{\widehat{\mathbb{P}}(Y = m)\}_{m=1}^{M}$ for each rater $i = 1, \dots, I$ and item $j = 1, \dots J$. Finally, fuzzy rating data were obtained using the procedure described in Section 3. Figures 3-4 shows some of the estimated fuzzy numbers. In particular, Figure 3 (left panel) shows a prototypical case in which the final crisp response $y_{98,4} = 4$ does not correspond to the

---

[2]The items were as follows: 1. *It upsets me to go into a situation without knowing what I can expect from it*, 2. *I enjoy having a clear and structured mode of life*, 3. *I like to have a place for everything and everything in its place*, 4. *I don't like situations that are uncertain*, 5. *I find that a consistent routine enables me to enjoy life more*.

	M		A_w		A_s		E	
	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	$\hat{\theta}$	$\sigma_{\hat{\theta}}$	$\hat{\theta}$	$\sigma_{\hat{\theta}}$
α_1	-1.19	0.12	-0.40	0.13	-1.04	0.25	0.33	0.20
α_2	-0.88	0.11	0.53	0.13	0.79	0.22	0.05	0.18
α_3	-0.56	0.10	0.25	0.14	-0.18	0.20	0.47	0.17
α_4	-1.50	0.12	0.46	0.12	0.51	0.26	0.06	0.22
α_5	-0.71	0.11	0.05	0.13	-0.28	0.20	0.04	0.17

Table 3: Case study: Estimates ($\hat{\theta}$) and standard errors ($\sigma_{\hat{\theta}}$) for item parameters $\hat{\boldsymbol{\alpha}}_{J \times N}$.

	η_1	η_2	η_3	η_4	$\hat{\sigma}_\eta$
η_1	1.00				0.75
η_2	-0.20	1.00			0.95
η_3	0.15	0.94	1.00		1.16
η_4	-0.76	0.41	0.12	1.00	0.29

Table 4: Case study: Estimated correlation matrix and standard deviations ($\hat{\sigma}_\eta$) for the latent traits $\hat{\boldsymbol{\eta}}_{I \times N}$.
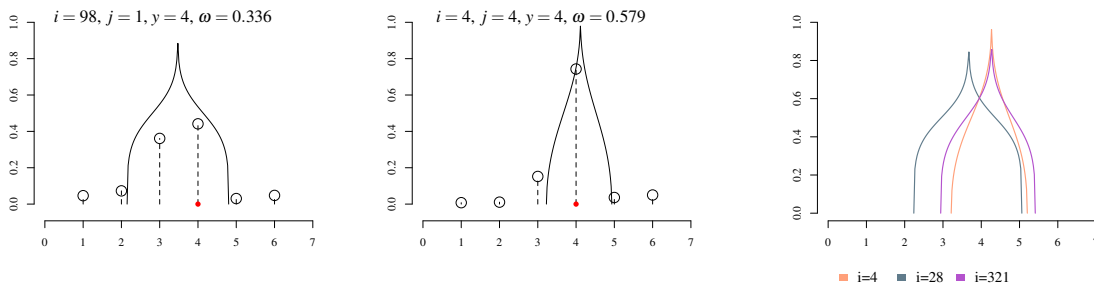

Figure 3: Case study: Estimated probabilistic model $\{\hat{\mathbb{P}}(Y = m)\}_{m=1}^{M}$ (black dashed vertical lines) for two rater's multiverses on two items and the associated fuzzy number (black dashed curves). The red circle indicates the final crisp responses y_{ij}. Note that probability masses and fuzzy membership functions are overlapped over the same domain $\Omega(Y)$ for the sake of comparison.

Figure 4: Case study: Estimated fuzzy numbers for three raters on item $j = 2$.

mode of the estimated fuzzy number. This can happen because fuzzy numbers encapsulate the decision uncertainty that affected the rater's response process, a feature which would otherwise be lost if crisp responses were solely considered. On the contrary, Figure 3 (right panel) depicts a situation with lower degree of decision uncertainty.

6 Conclusion

In this contribution we presented fIRTree, a novel procedure to represent human rating data in terms of fuzzy numbers. Similarly for other procedures aiming at turning crisp ratings into fuzzy ratings (e.g., see

[20]), our approach is based on a two-step rationale through which a psychometric rating model (IRtree) is defined and fit on empirical data first, and then the estimated parameters are used to compute 4-tuple triangular fuzzy numbers $\text{Trg}(c, l, r, \omega)$ for each rater and item combination. Although other fuzzy numbers can be adopted to this purpose (e.g., trapezoidal, beta, gaussian), $\text{Trg}(c, l, r, \omega)$ are flexible enough to represent the fuzziness of the rater's response process [10, 17]. A simulation study has been designed to assess the accuracy of fIRTree to recover fuzziness of rating data and a case study has been used to illustrate how fIRTree can be applied on empirical cases.

We believe that fIRTree can improve the way to get fuzzy numbers from rating data. fIRTree does not require dedicated rating tools or measurement settings and it can directly be applied on crisp rating data collected using standard formats of rating scales. Despite its simplicity, it provides a coherent meaning system to compute and interpret fuzzy numbers in measurement

contexts involving human ratings.

References

[1] B. Berglund, G. B. Rossi, J. T. Townsend, L. R. Pendrill, Measurement with persons: theory, methods, and implementation areas, Psychology Press, 2012.

[2] R. Berkachy, L. Donzé, Testing hypotheses by fuzzy methods: A comparison with the classical approach, in: Applying Fuzzy Logic for the Digital Economy and Society, Springer International Publishing, 2019, pp. 1–22.

[3] P. D. Boeck, I. Partchev, IRTrees: Tree-based item response models of the GLMM family, J. Stat. Soft. 48 (Code Snippet 1).
URL https://doi.org/10.18637%2Fjss.v048.c01

[4] M. E. Brooks, K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Maechler, B. M. Bolker, glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling, The R Journal 9 (2) (2017) 378–400.

[5] U. Böckenholt, Measuring response styles in likert items., Psychological Methods 22 (1) (2017) 69–83.

[6] A. Calcagnì, L. Lombardi, Dynamic fuzzy rating tracker (dyfrat): a novel methodology for modeling real-time dynamic cognitive processes in rating scales, Applied soft computing 24 (2014) 948–961.

[7] I. Couso, C. Borgelt, E. Hullermeier, R. Kruse, Fuzzy sets in data analysis: from statistical foundations to machine learning, IEEE Computational Intelligence Magazine 14 (1) (2019) 31–44.

[8] P. De Boeck, M. Bakker, R. Zwitser, M. Nivard, A. Hofman, F. Tuerlinckx, I. Partchev, et al., The estimation of item response models with the lmer function from the lme4 package in r, Journal of Statistical Software 39 (12) (2011) 1–28.

[9] S. d. l. R. de Sáa, M. Á. Gil, G. González-Rodríguez, M. T. López, M. A. Lubiano, Fuzzy rating scale-based questionnaires and their statistical analysis, IEEE Transactions on Fuzzy Systems 23 (1) (2014) 111–126.

[10] J. Dombi, T. Jónás, Approximations to the normal probability distribution function using operators of continuous-valued logic, Acta Cybernetica 23 (3) (2018) 829–852.

[11] M. Jeon, P. D. Boeck, A generalized item response tree model for psychological assessments, Behavior Research Methods 48 (3) (2015) 1070–1085.

[12] Q. Li, Indirect membership function assignment based on ordinal regression, Journal of Applied Statistics 43 (3) (2016) 441–460.

[13] L. Lombardi, M. Pastore, M. Nucci, A. Bobbio, Sgr modeling of correlational effects in fake good self-report measures, Methodology and Computing in Applied Probability 17 (4) (2015) 1037–1055.

[14] X.-B. Meng, J. Tao, N.-Z. Shi, An item response model for likert-type data that incorporates response time in personality measurements, Journal of Statistical Computation and Simulation 84 (1) (2014) 1–21.

[15] S. L. Neuberg, J. T. Newsom, Personal need for structure: Individual differences in the desire for simpler structure., Journal of personality and social psychology 65 (1) (1993) 113.

[16] M. E. Timmerman, H. A. Kiers, Three-way component analysis with smoothness constraints, Computational statistics & data analysis 40 (3) (2002) 447–470.

[17] Z. E. Tóth, T. Jónás, R. V. Dénes, Applying flexible fuzzy numbers for evaluating service features in healthcare–patients and employees in the focus, Total Quality Management & Business Excellence 30 (sup1) (2019) S240–S254.

[18] P. Vonglao, Application of fuzzy logic to improve the likert scale to measure latent variables, Kasetsart Journal of Social Sciences 38 (3) (2017) 337–344.

[19] T. M. Williams, Practical use of distributions in network analysis, Journal of the Operational Research Society 43 (3) (1992) 265–270.

[20] S.-C. Yu, B. Wu, Fuzzy item response model: a new approach to generate membership function to score psychological measurement, Quality and Quantity 43 (3) (2009) 381.

[21] M. J. Zickar, R. E. Gibby, C. Robie, Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory, Organizational Research Methods 7 (2) (2004) 168–190.