A Bayesian beta linear model to analyze fuzzy rating responses

Un modello bayesiano di regressione beta per l'analisi di risposte imprecise

Antonio Calcagnì, Massimiliano Pastore, Gianmarco Altoè, Livio Finos

Abstract In this short paper, we describe a Bayesian beta linear model to analyse imprecise rating responses. The non-random imprecision is extracted from crisp responses via the Item Response Theory tree (IRtree) method and it is represented by means of beta fuzzy numbers. The parameters of the beta linear model are estimated using the adaptive Metropolis-Hastings algorithm, with the fuzzy likelihood function being used as empirical evidence for the imprecise observations. A real case study is used to show the characteristics of the fuzzy beta regression model.

Abstract *Questo contributo descrive l'applicazione del modello di regressione beta nell'analisi di dati imprecisi. In questo contesto, l'imprecisione è riferita ad una fonte non casuale di incertezza ed è calcolata mediante il metodo fuzzy-IRTree. Il risultato di tale pre-trattamento è una collezione di insiemi fuzzy di tipo beta. I parametri del modello di regressione beta sono stimati mediante l'algoritmo Metropolis-Hastings di tipo adattivo mentre l'evidenza empirica dei dati è espressa mediante una funzione di verosmiglianza imprecisa. Il contributo si chiude con l'analisi di un caso studio.*

Key words: Fuzzy rating data, Beta linear model, Bayesian data analysis

---

Antonio Calcagnì,
DPSS, University of Padova, e-mail: antonio.calcagni@unipd.it
GNCS Research Group, National Institute of Advanced Mathematics (INdAM)

Massimiliano Pastore,
DPSS, University of Padova, e-mail: massimiliano.pastore@unipd.it

Gianmarco Altoè,
DPSS, University of Padova, e-mail: gianmarco.altoe@unipd.it

Livio Finos,
DPSS, University of Padova, e-mail: livio.finos@unipd.it

1 Introduction

Rating data are widespread across disciplines dealing with human-based measurements. In these cases, since the measurement process is based on cognitive actors, the collected data are often affected by non-random imprecision or fuzziness. This type of uncertainty has multiple origins, including the semantic aspects of the items being rated and the individual-level decision uncertainty underlying the response process [1]. To give an example, consider the situation where the item "I am satisfied with my life" is rated through a scale ranging from "strongly disagree" to "strongly agree". A stage-wise response process is usually involved in responding to these types of items. In particular, in a first step cognitive and affective information about the item being rated are retrieved and integrated (opinion formation stage) until the second decision stage is triggered, which includes the selection of the final rating response (e.g., "strongly disagree"). Because of the integration of conflicting cognitive and affective information about the item, fuzziness arises from the conflicting demands of the opinion formation stage [2]. Over the recent years, several fuzzy rating scales have been proposed to quantify fuzziness from rating data, including both direct/indirect fuzzy rating scales and fuzzy conversion scales (for an extensive review, see [1]). While direct fuzzy rating scales quantifies fuzziness by mapping response process to fuzzy numbers directly, fuzzy indirect scales aim at turning standard crisp ratings into fuzzy numbers by means of statistically-based procedures (e.g., see [3]). Unlike for the previous case, here the aim is to represent as much information as possible from the rating process in terms of a more complex number representation. Once fuzzy numbers have been obtained, they can be analysed either by means of standard statistical approaches or by adopting fuzzy statistical methods devoted to this purpose (e.g., see [4]).

In this contribution, we describe an application of a Bayesian beta linear model to the analysis of IRTree-based fuzzy data, a novel type of fuzzy responses which treat fuzziness in terms of decision uncertainty [1]. The remainder of this short paper is as follows. Section 2 describes the fuzzy beta data. Section 3 exposes the Beta linear model along with the parameter estimation procedure. Finally, Section 4 concludes this contribution by illustrating the application of the proposed method to a real dataset.

2 Data

IRTree-based fuzzy data represent a particular type of fuzzy numbers which are the output of a psychometric-based fuzzy conversion method (i.e., the Item Response Theory tree approach). In particular, they are computed in a way that the imprecision encapsulated into the matrix of crisp rating data $\mathbf{Y}_{n\times J}$ is mapped onto beta fuzzy numbers using all the rater's responses $\mathbf{y}_i$ to the J items being rated. The reader can refer to [1] for technical details about the conversion system. In this context, data consist of a collection of n (raters) $\times$ J (items) beta fuzzy numbers:

$$\widetilde{\mathbf{Y}} = (\xi_{\widetilde{y}_{11}}, \ldots, \xi_{\widetilde{y}_{1J}}, \ldots, \xi_{\widetilde{y}_{nj}}, \ldots, \xi_{\widetilde{y}_{nJ}})$$

where

$$\xi_{\widetilde{y}_{ij}}(y) = \frac{1}{C}\, y^{a-1}(1-y)^{b-1} \tag{1}$$
$$a = 1 + ms \quad \text{and} \quad b = 1 + s(1-m)$$

In this representation, $m \in (0,1)$ is the mode of the set, $s \in (0,\infty)$ is the precision of the set, whereas C is a constant ensuring that $\max_{y \in \mathcal{Y}} \xi_{\widetilde{y}}(y) = 1$. Note that $\xi_{\widetilde{y}_{ij}}$ is a normal and convex fuzzy set which lies in the interval $(0,1) \subset \mathbb{R}$. Because of the parametric representation involved by beta fuzzy numbers, the observed fuzzy data can be represented using two matrices, namely the matrix of modes $\mathbf{M}_{n\times J}$ and that of precisions $\mathbf{S}_{n\times J}$. Figure 1 shows an example of beta rating responses. It should be remarked that, in view of the fuzzy-IRTree representation adopted here, m represents the most plausible rating choice, s is the precision of m (i.e., smaller values indicate larger levels of hesitation in the rating choice), and $\xi_{\widetilde{y}_{ij}}$ codifies the decision uncertainty in terms of fuzziness (the larger the fuzziness, the highest the decision uncertainty). Ideally, in the case of no decision uncertainty, the fuzziness would vanish and the true rating realization would be precisely observed.

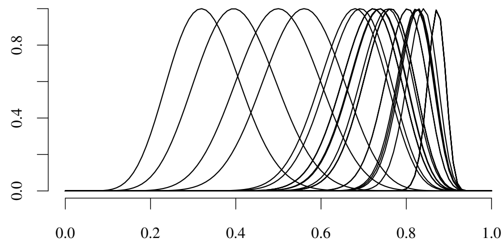

Fig. 1 An example of beta fuzzy responses.

3 Model and parameter estimation

To analyse fuzzy rating data we will adopt the Beta linear model proposed by [4], which is particularly well-suited for bounded rating responses. For a crisp collection of i.i.d. $(0,1)$-realizations $\mathbf{y} = (y_1, \ldots, y_n)$, the Beta density is as follows:

$$f_{\mathbf{Y}}(\mathbf{y};\mu,\phi) = \prod_{i=1}^{n} \frac{\Gamma(\phi_i)}{\Gamma(\phi_i\mu_i)\Gamma(\phi_i-\mu_i\phi_i)}\, y_i^{(\mu_i\phi_i-1)}(1-y_i)^{(\phi_i-\mu_i\phi_i-1)} \tag{2}$$

where

$$\mu = (1+\exp(\mathbf{X}\beta))^{-1} \quad \text{and} \quad \phi = \exp(\mathbf{Z}\gamma) \tag{3}$$

where $\mu \in (0,1)^n$ is the $n\times 1$ vector of location parameters and $\phi \in (0,\infty)^n$ the $n\times 1$ vector of precision parameters, which have been linearly expanded to account for the presence of covariates. To estimate model parameters $\theta = (\beta,\gamma)$, we use an adaptive Metropolis-Hastings algorithm where the transition distribution is approximated by means of a multivariate Normal distribution $q(\theta^{(t)}|\theta^{(t-1)}) = \mathcal{N}(;\theta^{(t-1)},\Sigma^{(t)})$, with the covariance matrix $\Sigma^{(t)}$ being adapted at each step by using a convenient subsample from the previous samples [5]. In this context, the acceptance ratio of the sampler is as follows:

$$\alpha^{(t)} = \frac{\mathcal{L}(\theta^{(t)};\mathbf{m},\mathbf{s})q(\theta^{(t-1)}|\theta^{(t)})f(\theta^{(t)})}{\mathcal{L}(\theta^{(t-1)};\mathbf{m},\mathbf{s})q(\theta^{(t)}|\theta^{(t-1)})f(\theta^{(t-1)})} \tag{4}$$

where $\mathcal{L}(\theta^{(t)};\mathbf{m},\mathbf{s})$ is the likelihood function for the fuzzy sample of data and $f(\theta^{(t)})$ is the prior density ascribed to the model parameters. In the case of i.i.d. and non-interactive fuzzy responses, the imprecise likelihood function is as follows [6, 7]:

$$\mathcal{L}(\theta^{(t)};\mathbf{m},\mathbf{s}) = \prod_{i=1}^{n}\int_0^1 \xi_{\widetilde{y}_i}(y;m_i,s_i)\frac{\Gamma(\phi_i)y^{(\mu_i\phi_i-1)}(1-y)^{(\phi_i-\mu_i\phi_i-1)}}{\Gamma(\phi_i\mu_i)\Gamma(\phi_i-\mu_i\phi_i)}\,dy \tag{5}$$

4 Application

In the present application we aimed to investigate the predictors of sexual intimacy in a sample of $n = 450$ participants from Flanders (73% female, mean age 32.9 years, mean relationship length 7.68 years).[1] Because of its characteristics, assessing the determinants of sexual intimacy is a typical situation in which raters show some levels of decision uncertainty in providing their self-reported responses. The survey consisted of four questionnaires used to measure (i) the perceived sexual intimacy with the partner, (ii) the perceived partner responsiveness (i.e., the extent to which one experiences the partner as being responsive to emotional needs), (iii) the sexual desire, (iv) the avoidant attachment score (i.e., how ambivalent early developmental experiences affect the current relationship). The items have been measured on a 7-point Likert-type scale with response categories ranging from 1 (“definitely not”) to 7 (“yes, definitely”). The items associated with sexual intimacy have been fuzzified using the fuzzy-IRTree method [1] and the ensuing fuzzy beta responses

[1] The dataset is publicly available at `http://osf.io/adgw2/`. For further details about the survey, see [8].

have been averaged to form the final intimacy indicator (see Figure 1 for a subsample of fuzzy responses). Following the findings by [8], three additive Beta linear models M1-M3 have been defined to predict sexual intimacy (see Table 1). The models have been varied in terms of covariates for the term μ, whereas no covariates have been used to model the precision term (i.e., $\phi = \exp(\mathbf{1}\gamma)$). For all the models, diffuse Normal densities have been used for the priors $f(\beta) = \mathcal{N}(\beta; \mathbf{0}, \mathbf{I}10)$ and $f(\gamma) = \mathcal{N}(\gamma; 0, 3)$ and four parallel MCMC have been run with 20000 samples (5000 samples for the burn-in phase) by means of the R package `MHadaptive`. The final model has been chosen according to the LOO information criterion as implemented by the R package `loo` [9]. According to the Gelman and Rubin's convergence diagnostics, all the chains reached the convergence (i.e., $\hat{R} = 1.00$). Table 2 reports the posterior quantiles along with the 95% HDIs for the six model parameters. Figure 2 shows the marginal posterior densities for the model parameters whereas Figure 3 plots the predicted curves against the observed fuzzy data as a function of both continuous and categorical predictors. The posterior results suggest that sexual intimacy is predicted by the perceived partner responsiveness, with a slight decrease of the outcome for the case in which the partner is male. The other predictors seem to play a marginal role in predicting sexual intimacy, with a negative relationship between avoidant attachment style and sexual intimacy as expected.

Model	Covariates	LOOIC
M1	`partner_respo`, `sex_desire`	873.80
M2	`partner_respo`, `sex_desire`, `attach_avoid`	863.50
M3	`partner_respo`, `sex_desire`, `attach_avoid`, `gender_partner`	**857.00**

Table 1 Application: Models for the sexual intimacy fuzzy rating data. Note that M3 is the best model according to the lowest LOO-IC criterion.

	β_0	$\beta_{\texttt{partner\_respo}}$	$\beta_{\texttt{sex\_desire}}$	$\beta_{\texttt{attach\_avoid}}$	$\beta_{\texttt{gender\_partner:Male}}$	γ
min	-1.01	0.11	-0.01	-0.05	-0.26	3.90
mean	-0.59	0.13	0.02	-0.03	-0.11	4.35
max	-0.12	0.15	0.04	-0.00	0.03	4.90
0.95 HDI_{lb}	-0.83	0.12	0.01	-0.04	-0.20	4.10
0.95 HDI_{ub}	-0.36	0.14	0.03	-0.01	-0.04	4.68

Table 2 Application: Posterior quantiles and 95% HDI for the model parameters. Note that β_0 is the intercept of μ and codifies the level `gender_partner = Female`.

References

1. Calcagnì, A., Cao, N., Rubaltelli, E., Lombardi, L.: A psychometric modeling approach to fuzzy rating data. Fuzzy Sets and Systems (2022)

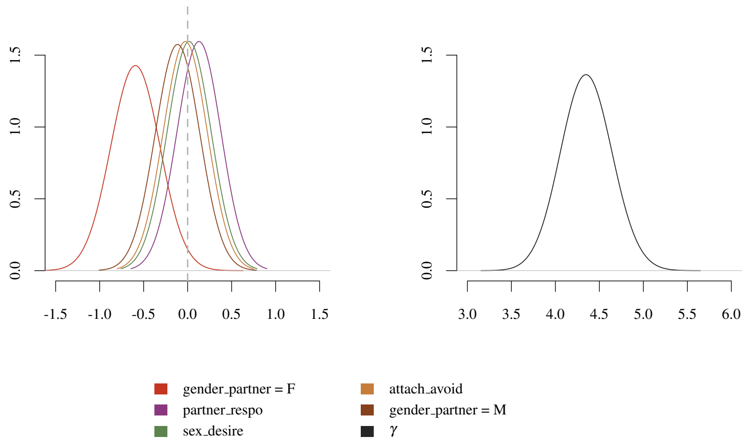

Fig. 2 Application: Marginal posterior densities for the model parameters.

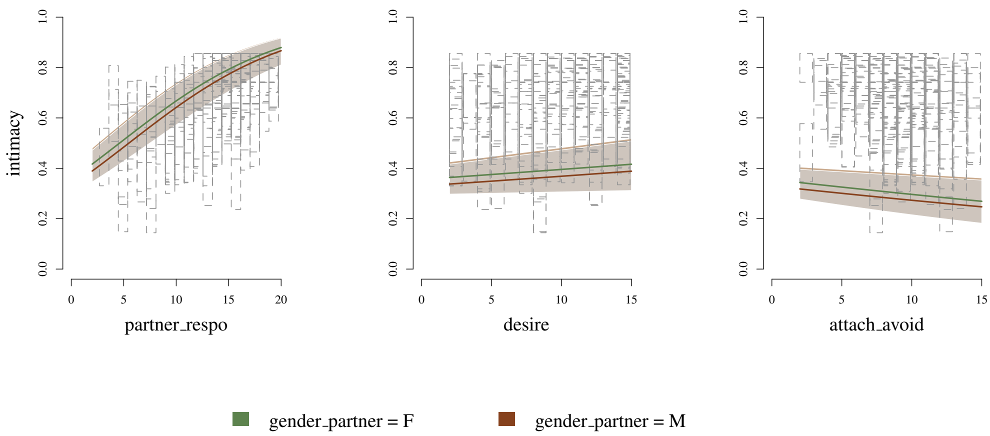

Fig. 3 Application: Observed fuzzy data for sexual intimacy as a function of the categorical predictor `gender_partner` (colors in the panels) and the three continuous predictors (panels). Fitted curves correspond to posterior means (see Table 2) whereas shadows represent the posterior 95% HDI of the predicted curves. Note that rectangles represent α-cuts of the observed fuzzy data with $\alpha = 0.5$, i.e. $\mathbf{y}_i^\alpha = \left[\min\left(\{y \in [0,1] : \xi_{\widetilde{y}_i}(y) > 0.5\}\right), \max\left(\{y \in [0,1] : \xi_{\widetilde{y}_i}(y) > 0.5\}\right)\right]$

2. Leary, M.R., Kowalski, R.M.: Impression management: A literature review and two-component model. Psychological bulletin **107**(1), 34 (1990)
3. Yu, S.-C., Wu, B.: Fuzzy item response model: a new approach to generate membership function to score psychological measurement. Quality and Quantity **43**(3), 381–390 (2009)
4. Coppi, R., Gil, M.A., Kiers, H.A.: The fuzzy approach to statistical analysis. Computational

statistics & data analysis **51**(1), 1–14 (2006)
5. Haario, H., Saksman, E., Tamminen, J.: An adaptive metropolis algorithm. Bernoulli, 223–242 (2001)
6. Gil, M., Corral, N., Gil, P.: The minimum inaccuracy estimates in $\chi 2$ tests for goodness of fit with fuzzy observations. Journal of Statistical Planning and Inference **19**(1), 95–115 (1988)
7. Calcagnì, A., Lombardi, L.: Modeling random and non-random decision uncertainty in ratings data: a fuzzy beta model. AStA Advances in Statistical Analysis, 1–29 (2021)
8. Van Lankveld, J.J., Dewitte, M., Verboon, P., van Hooren, S.A.: Associations of intimacy, partner responsiveness, and attachment-related emotional needs with sexual desire. Frontiers in Psychology **12** (2021)
9. Vehtari, A., Gelman, A., Gabry, J.: Practical bayesian model evaluation using leave-one-out cross-validation and waic. Statistics and computing **27**(5), 1413–1432 (2017)