Bayesian probit models for preference classification

An analysis of chess players' propensity for risk-taking

Lennart Oelschläger lennart.oelschlaeger@uni-bielefeld.de
Dietmar Bauer dietmar.bauer@uni-bielefeld.de

UNIVERSITÄT BIELEFELD

Summary

- Marketing, transportation, psychology, and other fields use probit models to analyse **discrete choice behavior** . We propose a latent class model extension that allows for the **classification of decider preferences** without requiring the explicit specification of the number of classes.
- The model is estimated in a **Bayesian framework** , and the class number is determined by a **Dirichlet process** .
- We apply the proposed method in the context of chess, where players are classified in three classes according to their **risk-taking propensity** .

Bayesian probit models

Probit models are commonly rooted in the **random utility framework** . They assume that deciders assign utility values to discrete choice alternatives and seek to maximize them. The utilities are modeled as a linear function of observable and unobservable factors, where the latter are assumed to follow a multivariate normal distribution. Specifically, decider n's choice $y_{nt} \in \{1, \dots, J\}$ at choice occasion t is explained through a matrix X_{nt} of choice characteristics as

$$y_{nt} = \arg\max U_{nt}, \quad U_{nt} = X_{nt}\beta + \varepsilon_{nt}, \quad \varepsilon_{nt} \sim N(0, \Sigma). \tag{1}$$

We assume that (1) has been normalized for level and scale. A Bayesian analysis requires the computation of the **posterior density**

$$\Pr(\beta, \Sigma \mid y, X) \propto \Pr(\beta, \Sigma) \times L(\beta, \Sigma \mid y, X). \tag{2}$$

For the prior $\Pr(\beta, \Sigma)$, it is convenient to employ independent conjugate distributions, i.e. the normal for β and the inverse Wishart for Σ. The probit likelihood is the product of independent multinomial distributions

$$L(\beta, \Sigma \mid y, X) = \prod \Pr(y_{nt} = \arg\max U_{nt}). \tag{3}$$

Evaluating (3) requires costly computations of the normal CDF due to the error specification in (1). Instead, we augment $(U_{nt})_{n,t}$ as parameters [1], following truncated normals, which yields a Gibbs sampling scheme to approximate (2).

We provide an **implementation** of the Gibbs sampler in R via the {RprobitB} package [5].

SCAN ME

Preference classification

To incorporate **preference heterogeneity** , we model random variation in the coefficient vector β across deciders using a Gaussian mixture with C classes:

$$\beta_n \sim \sum s_c N(b_c, \Omega_c), \tag{4}$$

where the weights $(s_c)_c$ are Dirichlet distributed with concentration $\delta > 0$. This

- provides an arbitrarily good approximation of the true underlying mixing distribution [4],
- and enables the classification of deciders with common expected preferences b_c and covariances Ω_c (our focus here).

To avoid the need to a priori select the number C of classes included, we impose a **Dirichlet process prior** $DP(G, \delta)$ on the distribution (4), where (assuming conjugate priors for b and Ω) the base distribution G is formed as the product of a normal and an inverse Wishart distribution [3]. The Dirichlet process integrates into the Gibbs sampler by iteratively updating $(b_c)_c$ and $(\Omega_c)_c$ using their posterior predictive distributions. The decider-specific assignments $z = (z_n)_n$ to either existing or new classes are updated via

$$\Pr(z_n = c \mid z_{-n}, \delta) = (N - 1 + \delta)^{-1} \cdot \begin{cases} |\{z_{-n} = c\}| & c = 1, \dots, C, \\ \delta & c = C + 1, \end{cases} \tag{5}$$

where z_{-n} denotes z excluding the n-th element, and N is the number of deciders.

The **impact of the concentration prior** δ on (5) diminishes as N increases, resulting in stable inference, as verified in our simulation:

	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$\delta = 10$
$N = 100$	1 (0.33)	2 (0.62)	2 (0.68)	3 (0.79)	4 (1.28)
$N = 1000$	3 (0.15)	3 (0.54)	3 (0.50)	4 (0.78)	5 (1.25)
$N = 6174$	3 (0.22)	3 (0.40)	3 (0.55)	3 (0.77)	4 (1.10)

Median C for varying N and δ with standard deviations. Choice data were simulated based on the application.

Application

We apply the proposed model to data from an online tournament hosted on www.lichess.org [2], where $N = 6174$ participants played multiple chess games with a time limit of one minute per game. A player whos time runs out looses the game automatically. Before the start of each round, players were presented with a **risky decision** : they could trade half of their clock time for the chance to earn one additional tournament point if they won the game.

The following **choice factors** potentially influence this decision:

- the player's rating and the rating difference to their opponent,
- whether they have the first-move advantage,
- the remaining tournament time,
- a winning streak (which yields extra points),
- whether they opted for the risky option in the previous round,
- whether they had lost in the previous round.

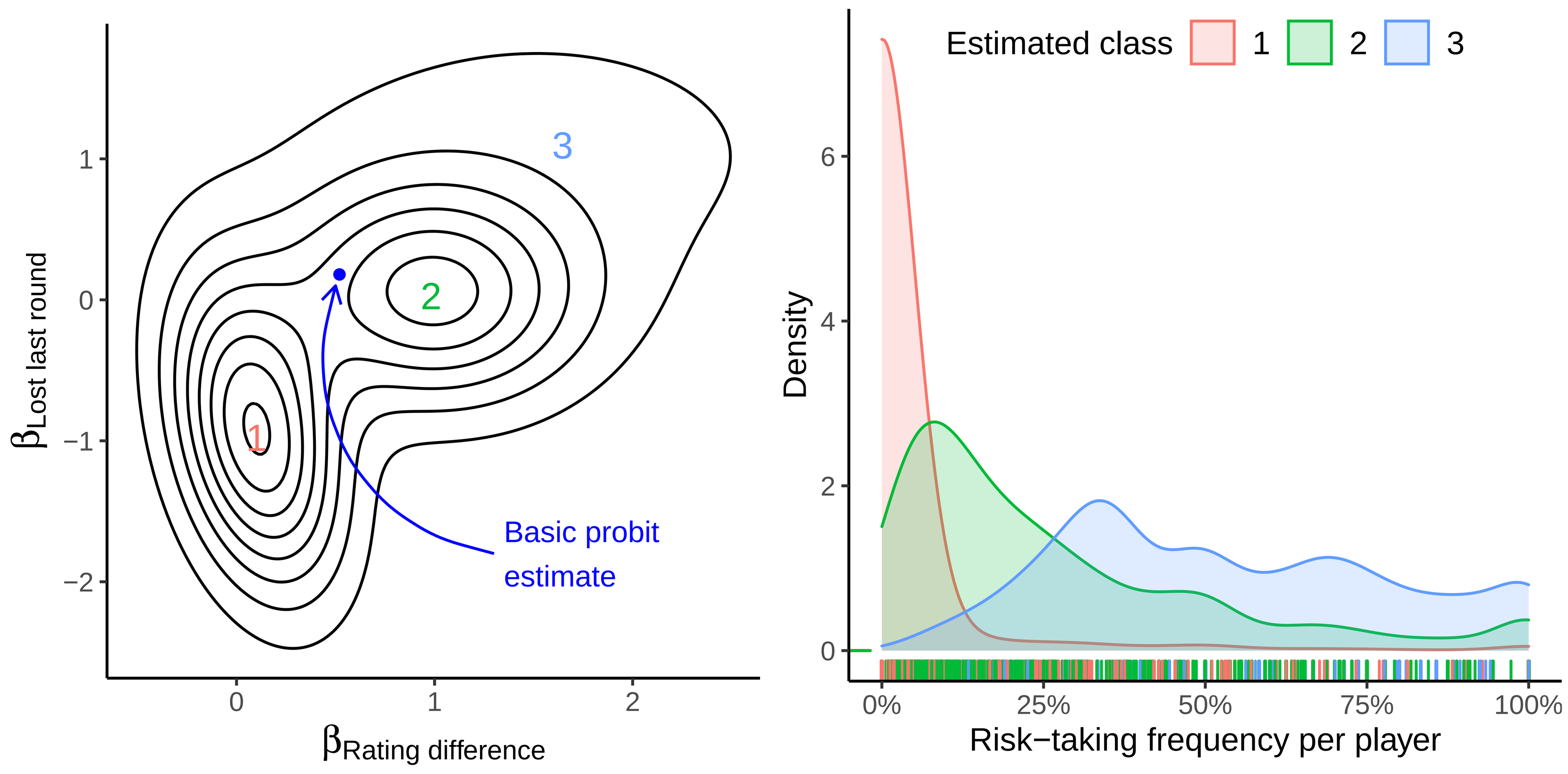
Model results

Factor	Latent class probit			Basic probit
Intercept	-2.05 (0.03)			-1.94 (0.01)
Rating	-0.11 (0.01)			-0.08 (0.01)
Having first move	-0.04 (0.02)			-0.02 (0.01)
Minutes remaining	0.04 (0.01)			0.04 (0.01)
On a winning streak	-0.27 (0.03)			-0.21 (0.02)
Took risk last round	1.21 (0.02)			1.82 (0.02)
	Class 1	Class 2	Class 3	
Proportion	54% (0.03)	36% (0.04)	10% (0.03)	
Lost last round	-0.98 (0.09)	0.03 (0.08)	1.10 (0.18)	0.18 (0.01)
Rating difference	0.10 (0.02)	0.98 (0.06)	1.65 (0.22)	0.52 (0.01)

Change in utility for taking the risk (ceteris paribus). Reported are the marginal posterior means with standard deviations. We fit both models using 5000 Gibbs iterations and set the concentration $\delta = 1$.

The latent class model converged to three classes that characterize different types of players:

- **Type 1 players** are risk-averse, rarely choosing the risky option against lower-rated opponents or after losing in the previous round.
- **Type 2 players** decide independently of the previous game's outcome.
- **Type 3 players** take more risks, with a higher likelihood of choosing the risky option after a loss and favoring it against weaker opponents.



Using the relative frequencies of the class allocation z, we can **classify each player** . For example, the tournament winner is of type 2 with a probability of 78%, while the runner-up is of type 1 with a probability of 94%.

References

[1] **Imai, K. and van Dyk, D. A.** (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, 124.2, 311–334.

[2] **Lichess API** (2023). https://lichess.org/api/tournament/RibHfoX6. Data accessed on 17 February 2023.

[3] **Neal, R. M.** (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2), 249–265.

[4] **Oelschläger, L. and Bauer, D.** (2021). Bayes Estimation of Latent Class Mixed Multinomial Probit Models. *TRB 100th Annual Meeting*.

[5] **Oelschläger, L., Bauer, D., Büscher, S., and Batram, M.** (2022). *RprobitB: Bayesian Probit Choice Modeling*. R package version 1.1.2, accessible via https://CRAN.R-project.org/package=RprobitB.