Bayes Estimation of Latent Class Mixed Multinomial Probit Models

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

For more information, download the full paper here.

UNIVERSITÄT BIELEFELD

Motivation

- Discrete choice models lie at the heart of many transportation models, e.g. the multinomial probit model.
- Modelling heterogeneity in preferences is indispensable in many applications and has been elaborated by imposing mixing distributions on the model coefficients.
- However, the literature does not provide much guidance for the specification of the mixing distribution apart from restrictive or computationally expensive strategies, e.g.
  - model selection on standard parametric distributions,
  - non-parametric approaches (cf. [1] and [2]).
- We present a new strategy, combining the ideas of
  1. a Bayesian framework (computational advantages),
  2. approximating mixing distributions by a mixture of normal distributions (high flexibility),
  3. weight-based updates on the number of latent classes (reduction of model assumptions).

Model Definition

Let person ns ($n=1,\dots,N$) utility of choice alternative j ($j=1,\dots,J-1$) at choice occasion t ($t=1,\dots,T$) be

$$U_{ntj} = W_{ntj}'\alpha + X_{ntj}'\beta_n + \varepsilon_{ntj},$$

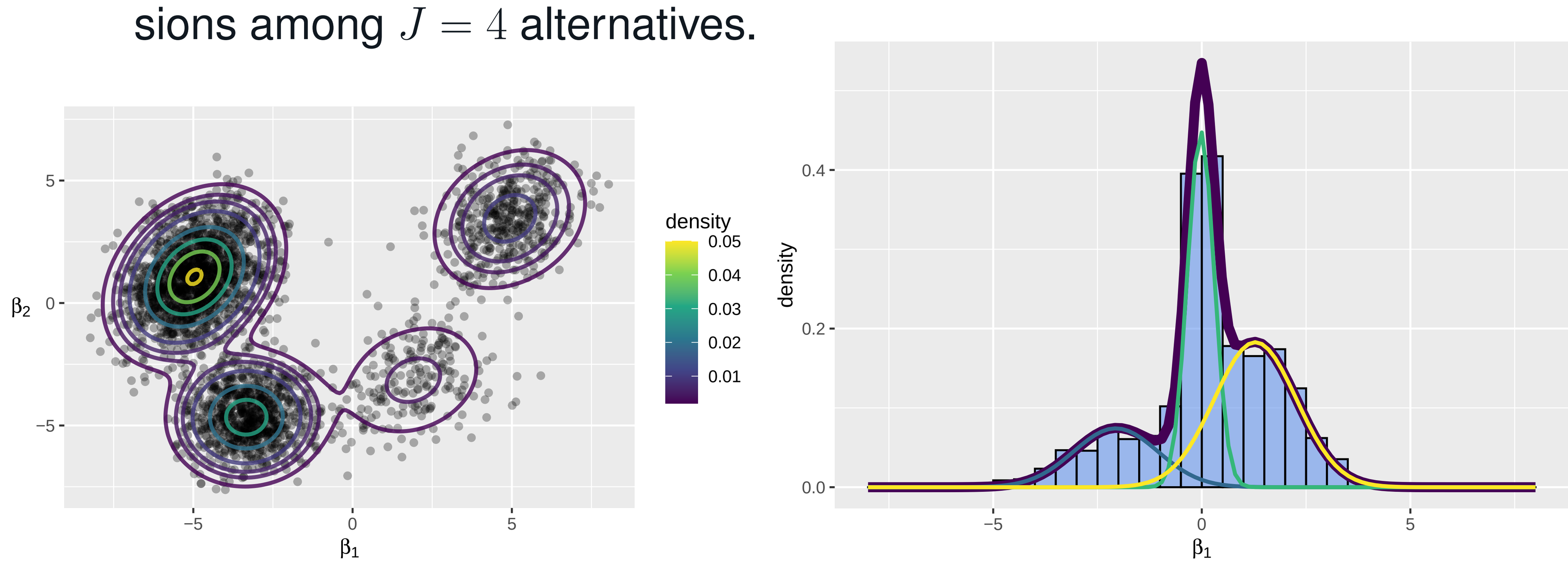
- where W_{ntj} (X_{ntj}) is a vector of P_f (P_r) differenced (wrt alternative J) characteristics of j as faced by n at t corresponding to the fixed (random and decision maker-specific) coefficient vector $\alpha \in \mathbb{R}^{P_f}$ ($\beta_n \in \mathbb{R}^{P_r}$),
- $(\varepsilon_{nt1},\dots,\varepsilon_{nt(J-1)})' \sim \text{MVN}_{J-1}(0,\tilde{\Sigma})$ with $\tilde{\Sigma}_{11}=1$.

Let y_{nt} denote ns choice at t. Assuming rationality, $y_{nt} = \sum_{j=1}^{J-1} j \cdot 1\left(U_{ntj} = \max_i U_{nti} > 0\right) + J \cdot 1\left(U_{ntj} < 0 \text{ for all } j\right)$.

We approximate the mixing distribution of $(\beta_n)_n$ by a normal mixture, i.e. $\beta_n \sim \sum_{c=1}^{C} s_c \cdot \text{MVN}_{P_r}(b_c,\Omega_c)$. This is equivalent to introducing class allocation variables $(z_n)_n$ with $\text{Prob}(z_n = c) = s_c$ and $\beta_n \mid z,b,\Omega \sim \text{MVN}_{P_r}(b_{z_n},\Omega_{z_n})$.

Simulation Results

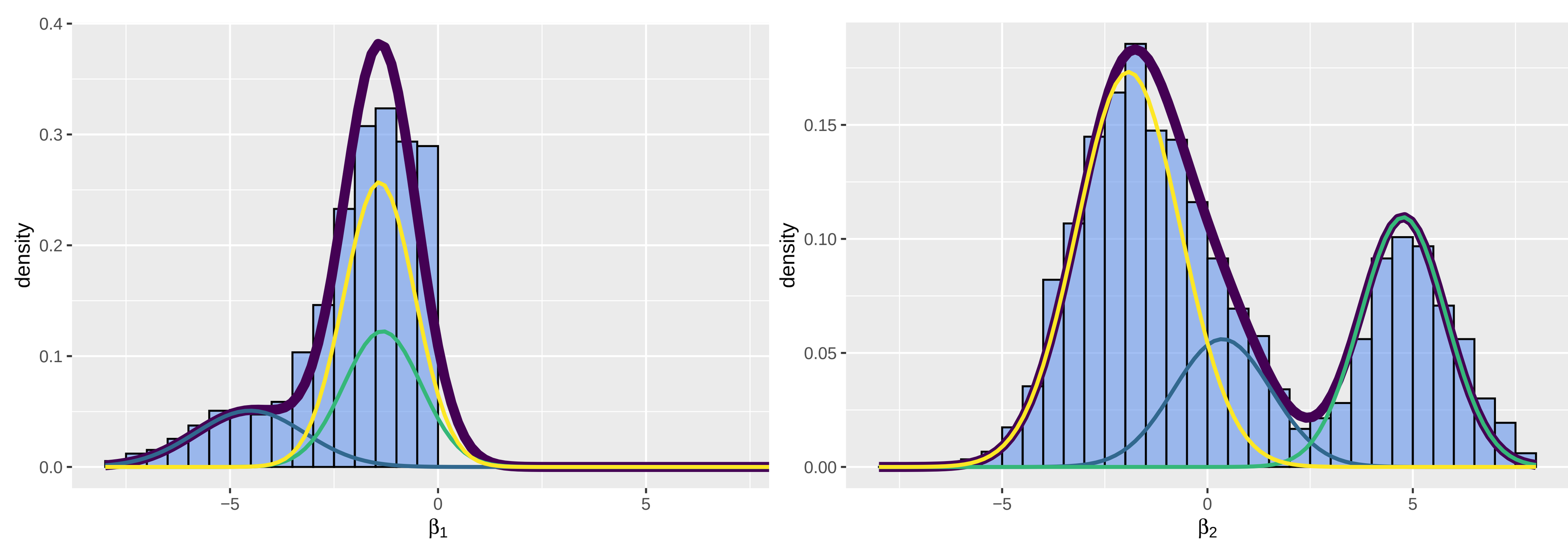
The proposed methodology to approximate underlying mixing distributions was tested in a series of simulations. Below, we exemplary present three of them, each having a data support of $N=3000$ individuals being observed on making decisions at $T=30$ choice occasions among $J=4$ alternatives.



Simulation 1 (left) and 2 (right)

In Simulation 1, the data originates from a latent class mixed multinomial probit model with $C=4$ classes, $P_f=1$ fixed and $P_r=2$ random coefficients. The updating scheme was initialised with 10 latent classes. Visibly, the true classes were reproduced.

Simulation 2 is based on a stylized real world case: The population is separable into distinct groups which have different views on a certain choice attribute (e.g. out-of-vehicle travel time). This can be translated into a latent class setting. Here, we considered 3 latent classes (disaffirmation, indifference and affirmation). Our methodology performs well in approximating such a complex mixing distribution.



Simulation 3

Another critical test of the approach constitutes its performance in applications with sign-restricted coefficients (e.g. for the alternative's price). In Simulation 3, $P_r=2$ random choice attributes were considered, the first of which was restricted to be non-positive. Visibly, a minor density mass (4.86%) is estimated also for positive values of the restricted coefficient. However, comparison tests indicate an acceptable approximation at a 5% significance level.

Bayesian Framework

Our Gibbs sampler builds upon the work of [3] and [4]. We apply conjugate and diffuse priors (their dependencies are abbreviated). Below, R denotes the number of iterations, B the number of discarded draws, and Q the thinning parameter. Within the second half of the burn-in period, the number of latent classes is updated.

Algorithm 1 Gibbs sampler

for $r=1,\dots,R$ **do**
  draw $(s_1,\dots,s_C) \mid z,\dots \sim D_C$ // C-dim. Dirichlet distribution
  draw z from its conditional distribution
  for $c=1,\dots,C$ **do**
    draw $b_c \mid z,\beta,\Omega\dots \sim \text{MVN}_{P_r}$ // P_r-dim. multivariate normal distribution
    draw $\Omega_c \mid z,\beta,b,\dots \sim W_{P_r}^{-1}$ // P_r-dim. inverse Wishart distribution
  for $n=1,\dots,N$ **do**
    draw $\beta_n \mid \Omega,b,X,\Sigma,U,W,\alpha \sim \text{MVN}_{P_r}$
    for $t=1,\dots,T$ **do**
      draw $U_{nt} \sim$ truncated MVN_{J-1} via a sub-Gibbs sampler (cf. [5])
  draw $\alpha \mid \beta,\Sigma,U,X,\dots \sim \text{MVN}_{P_f}$
  draw $\Sigma \mid U,W,\alpha,X,\beta,\dots \sim W_{J-1}^{-1}$
  if $B/2 < r \leq B$ **then**
    call Updating scheme with current draws
  if $(B < r \leq R) \wedge (r \mod Q = 0)$ **then**
    save current draws
normalize saved draws (cf. [6])

Algorithm 2 Updating scheme

for $c=1,\dots,C$ **do**
  if $s_c < \varepsilon_{\text{min}}$ **then**
    remove class c
  if $s_c > \varepsilon_{\text{max}}$ **then**
    split class c
  if $\|b_c - b_{c^*}\| < \varepsilon_{\text{distmin}}$ for any other class c^* **then**
    join classes c and c^* and average their parameters

References for Poster

[1] Kenneth Train. Mixed logit with a flexible mixing distribution. *Journal of choice modelling*, 19, 2016.

[2] Dietmar Bauer, Sebastian Büscher, and Manuel Batram. Non-parameteric estiation of mixed discrete choice models. *Second International Choice Modelling Conference in Kobe*, 2019.

[3] Robert McCulloch and Peter Rossi. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64, 1994.

[4] Luisa Scaccia and Edoardo Marcucci. Bayesian flexible modelling of mixed logit models. *Proceedings from the 19th International Conference on Computational Statistics*, 2010.

[5] John Geweke. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. *Comput. Sci. Statist.*, 23, 1998.

[6] Kosuke Imai and David A. van Dyk. A bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, 124, 2005.

This work has been financed partly by the Deutsche Forschungsgemeinschaft (German Research Foundation) which is gratefully acknowledged.

DFG Deutsche Forschungsgemeinschaft
German Research Foundation