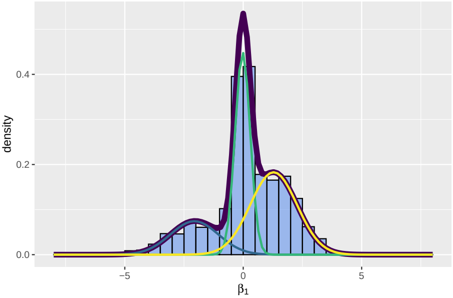


Approximating mixing distributions in probit models via a Bayesian approach

Lennart Oelschläger

November 24, 2020

The agenda

1. **Discrete choice**
   - Multinomial probit model
   - Mixing distributions
   - Latent class mixed multinomial probit model

2. **Bayesian framework**
   - Data augmentation
   - Priors
   - Gibbs sampler

3. **Latent class updating scheme**

4. **Simulations**

Discrete choice

Assume that we

- observe the choices of N decision makers (stated or revealed)
- which decide between J mutually exclusive alternatives
- at each of T choice occasions.

Example

Commute to the university ($J = 3$):

$$\underbrace{\begin{bmatrix} U_{\text{bicycle}} \\ U_{\text{car}} \\ U_{\text{helicopter}} \end{bmatrix}}_{\text{utility vector } U} = \underbrace{\begin{bmatrix} 0€ & 10min & 0g \\ 2€ & 5min & 80g \\ 500€ & 1min & 900g \end{bmatrix}}_{\text{choice characteristics } X} \cdot \underbrace{\begin{bmatrix} \beta_{\text{cost}} \\ \beta_{\text{time}} \\ \beta_{\text{emission}} \end{bmatrix}}_{\text{sensitivities } \beta} + \underbrace{\begin{bmatrix} \varepsilon_{\text{bicycle}} \\ \varepsilon_{\text{car}} \\ \varepsilon_{\text{helicopter}} \end{bmatrix}}_{\text{model's mistake } \varepsilon}$$

Discrete choice

Assume that we

- observe the choices of N decision makers (stated or revealed)
- which decide between J mutually exclusive alternatives
- at each of T choice occasions.

Example

Commute to the university ($J = 3$):

$$\underbrace{\begin{bmatrix} U_{\text{bicycle}} \\ U_{\text{car}} \\ U_{\text{helicopter}} \end{bmatrix}}_{\text{utility vector } U} = \underbrace{\begin{bmatrix} 0€ & 10min & 0g \\ 2€ & 5min & 80g \\ 500€ & 1min & 900g \end{bmatrix}}_{\text{choice characteristics } X} \cdot \underbrace{\begin{bmatrix} \beta_{\text{cost}} \\ \beta_{\text{time}} \\ \beta_{\text{emission}} \end{bmatrix}}_{\text{sensitivities } \beta} + \underbrace{\begin{bmatrix} \varepsilon_{\text{bicycle}} \\ \varepsilon_{\text{car}} \\ \varepsilon_{\text{helicopter}} \end{bmatrix}}_{\text{model's mistake } \varepsilon}$$

Multinomial probit model

Person n's utility U_{ntj} for alternative j at choice occasion t is modelled as

$$U_{ntj} = X_{ntj}'\beta + \varepsilon_{ntj}$$

for $n = 1, \dots, N$, $t = 1, \dots, T$ and $j = 1, \dots, J$, where (in the probit)

$$(\varepsilon_{nt1}, \dots, \varepsilon_{ntJ})' \sim \text{MVN}_J(0, \Sigma).$$

We have to normalize with respect to

- level (by taking utility differences, reference alternative J) and
- scale (by setting $\tilde{\Sigma}_{11} = 1$).

Let $y_{nt} = j$ denote that n chooses j at t. We have the link

$$y_{nt} = \sum_{j=1}^{J-1} j \cdot 1\left(U_{ntj} = \max_i U_{nti} > 0\right) + J \cdot 1\left(U_{ntj} < 0 \text{ for all } j\right).$$

Multinomial probit model

Person n's utility U_{ntj} for alternative j at choice occasion t is modelled as

$$U_{ntj} = X'_{ntj}\beta + \varepsilon_{ntj}$$

for $n = 1, \dots, N$, $t = 1, \dots, T$ and $j = 1, \dots, J$, where (in the probit)

$$(\varepsilon_{nt1}, \dots, \varepsilon_{ntJ})' \sim \text{MVN}_J(0, \Sigma).$$

We have to normalize with respect to

- level (by taking utility differences, reference alternative J) and
- scale (by setting $\tilde{\Sigma}_{11} = 1$).

Let $y_{nt} = j$ denote that n chooses j at t. We have the link

$$y_{nt} = \sum_{j=1}^{J-1} j \cdot 1\left(U_{ntj} = \max_i U_{nti} > 0\right) + J \cdot 1\left(U_{ntj} < 0 \text{ for all } j\right).$$

Multinomial probit model

Person n's utility U_{ntj} for alternative j at choice occasion t is modelled as

$$U_{ntj} = X'_{ntj}\beta + \varepsilon_{ntj}$$

for $n = 1, \dots, N$, $t = 1, \dots, T$ and $j = 1, \dots, J$, where (in the probit)

$$(\varepsilon_{nt1}, \dots, \varepsilon_{ntJ})' \sim \text{MVN}_J(0, \Sigma).$$

We have to normalize with respect to

- level (by taking utility differences, reference alternative J) and
- scale (by setting $\tilde{\Sigma}_{11} = 1$).

Let $y_{nt} = j$ denote that n chooses j at t. We have the link

$$y_{nt} = \sum_{j=1}^{J-1} j \cdot 1\left(U_{ntj} = \max_i U_{nti} > 0\right) + J \cdot 1\left(U_{ntj} < 0 \text{ for all } j\right).$$

Mixing distributions

Definition (Multinomial probit model)

Person n's utility U_{ntj} for j at t is modeled as

$$U_{ntj} = X_{ntj}'\beta + \varepsilon_{ntj}.$$

- different decision makers have different sensitivities
- Allowing for heterogeneity: $U_{ntj} = X_{ntj}'\beta_n + \varepsilon_{ntj}, \quad \beta_n \sim f(\beta)$, e.g.
  - $\beta_{\text{cost},n} \sim -\mathcal{LN}(\mu, \sigma^2)$
  - $\begin{bmatrix} \beta_{\text{time},n} \\ \beta_{\text{cost},n} \end{bmatrix} \sim \text{MVN}_2(b, \Omega)$
    (to capture correlation patterns: lower cost sensitivity correlated with higher time sensitivity for business travelers?)
- What is the "correct" f?

Mixing distributions

Definition (Multinomial probit model)

Person n's utility U_{ntj} for j at t is modeled as

$$U_{ntj} = X'_{ntj}\beta + \varepsilon_{ntj}.$$

- different decision makers have different sensitivities
- Allowing for heterogeneity: $U_{ntj} = X'_{ntj}\beta_n + \varepsilon_{ntj}, \quad \beta_n \sim f(\beta)$, e.g.
  - $\beta_{\text{cost},n} \sim -\mathcal{LN}(\mu, \sigma^2)$
  - $\begin{bmatrix}\beta_{\text{time},n} \\ \beta_{\text{cost},n}\end{bmatrix} \sim \text{MVN}_2(b, \Omega)$
    (to capture correlation patterns: lower cost sensitivity correlated with higher time sensitivity for business travelers?)
- What is the "correct" f?

Mixing distributions

Definition (Multinomial probit model)

Person n's utility U_{ntj} for j at t is modeled as

$$U_{ntj} = X_{ntj}'\beta + \varepsilon_{ntj}.$$

- different decision makers have different sensitivities
- Allowing for heterogeneity: $U_{ntj} = X_{ntj}'\beta_n + \varepsilon_{ntj}, \quad \beta_n \sim f(\beta)$, e.g.
  - $\beta_{\text{cost},n} \sim -\mathcal{LN}(\mu, \sigma^2)$
  - $\begin{bmatrix}\beta_{\text{time},n} \\ \beta_{\text{cost},n}\end{bmatrix} \sim \text{MVN}_2(b, \Omega)$
    (to capture correlation patterns: lower cost sensitivity correlated with higher time sensitivity for business travelers?)
- What is the "correct" f?

Mixing distributions

Definition (Multinomial probit model)

Person n's utility U_{ntj} for j at t is modeled as

$$U_{ntj} = X'_{ntj}\beta + \varepsilon_{ntj}.$$

- different decision makers have different sensitivities
- Allowing for heterogeneity: $U_{ntj} = X'_{ntj}\beta_n + \varepsilon_{ntj}, \quad \beta_n \sim f(\beta)$, e.g.
  - $\beta_{\text{cost},n} \sim -\mathcal{LN}(\mu, \sigma^2)$
  - $\begin{bmatrix}\beta_{\text{time},n} \\ \beta_{\text{cost},n}\end{bmatrix} \sim \text{MVN}_2(b, \Omega)$
    (to capture correlation patterns: lower cost sensitivity correlated with higher time sensitivity for business travelers?)
- What is the "correct" f?

Mixing distributions

Definition (Multinomial probit model)

Person n's utility U_{ntj} for j at t is modeled as

$$U_{ntj} = X'_{ntj}\beta + \varepsilon_{ntj}.$$

- different decision makers have different sensitivities
- Allowing for heterogeneity: $U_{ntj} = X'_{ntj}\beta_n + \varepsilon_{ntj}, \quad \beta_n \sim f(\beta)$, e.g.
  - $\beta_{\text{cost},n} \sim -\mathcal{LN}(\mu, \sigma^2)$
  - $\begin{bmatrix}\beta_{\text{time},n} \\ \beta_{\text{cost},n}\end{bmatrix} \sim \text{MVN}_2(b, \Omega)$
    (to capture correlation patterns: lower cost sensitivity correlated with higher time sensitivity for business travelers?)
- What is the "correct" f?

Latent class mixed multin. probit model

Definition

For $n = 1, \ldots, N$, $t = 1, \ldots, T$ and $j = 1, \ldots, J-1$,

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

where

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

Latent class mixed multin. probit model

Definition

For $n = 1, \ldots, N$, $t = 1, \ldots, T$ and $j = 1, \ldots, J-1$,

$$U_{ntj} = W'_{ntj}\alpha + X'_{ntj}\beta_n + \varepsilon_{ntj},$$

where

- $y_{nt} = \sum_{j=1}^{J-1} j \cdot 1\left(U_{ntj} = \max_i U_{nti} > 0\right) + J \cdot 1\left(U_{ntj} < 0 \text{ for all } j\right)$
- and

$$\beta_n \mid b, \Omega \sim \sum_{c=1}^{C} s_c \cdot \text{MVN}_{P_r}(b_c, \Omega_c)$$

$$\iff \quad \text{Prob}(z_n = c) = s_c \quad \text{and} \quad \beta_n \mid z, b, \Omega \sim \text{MVN}_{P_r}(b_{z_n}, \Omega_{z_n})$$

Latent class mixed multin. probit model

Definition

For $n = 1, \dots, N$, $t = 1, \dots, T$ and $j = 1, \dots, J-1$,

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

where

- $y_{nt} = \sum_{j=1}^{J-1} j \cdot 1\left(U_{ntj} = \max_i U_{nti} > 0\right) + J \cdot 1\left(U_{ntj} < 0 \text{ for all } j\right)$
- and

$$\beta_n \mid b, \Omega \sim \sum_{c=1}^{C} s_c \cdot \text{MVN}_{P_r}(b_c, \Omega_c)$$

$$\iff \quad \text{Prob}(z_n = c) = s_c \quad \text{and} \quad \beta_n \mid z, b, \Omega \sim \text{MVN}_{P_r}(b_{z_n}, \Omega_{z_n})$$

Data augmentation

- *"generate a variable that wasn't there before"*
- treat the latent utilities U as parameters
- conditional on the latent utilities, the model constitutes a standard Bayesian linear regression set-up ($U = X\beta + \varepsilon$)
- drawing from the posterior distribution becomes feasible without the need to evaluate any likelihood
- numerical advantages
  - avoids approximation of Gaussians
  - reduces curse of dimensionality (starting values)
  - faster?

Data augmentation

- *"generate a variable that wasn't there before"*
- treat the latent utilities U as parameters
- conditional on the latent utilities, the model constitutes a standard Bayesian linear regression set-up ($U = X\beta + \varepsilon$)
- drawing from the posterior distribution becomes feasible without the need to evaluate any likelihood
- numerical advantages
  - avoids approximation of Gaussians
  - reduces curse of dimensionality (starting values)
  - faster?

Data augmentation

- *"generate a variable that wasn't there before"*
- treat the latent utilities U as parameters
- conditional on the latent utilities, the model constitutes a standard Bayesian linear regression set-up ($U = X\beta + \varepsilon$)
- drawing from the posterior distribution becomes feasible without the need to evaluate any likelihood
- numerical advantages
  - avoids approximation of Gaussians
  - reduces curse of dimensionality (starting values)
  - faster?

Data augmentation

- *"generate a variable that wasn't there before"*
- treat the latent utilities U as parameters
- conditional on the latent utilities, the model constitutes a standard Bayesian linear regression set-up ($U = X\beta + \varepsilon$)
- drawing from the posterior distribution becomes feasible without the need to evaluate any likelihood
- numerical advantages
  - avoids approximation of Gaussians
  - reduces curse of dimensionality (starting values)
  - faster?

Data augmentation

- *"generate a variable that wasn't there before"*
- treat the latent utilities U as parameters
- conditional on the latent utilities, the model constitutes a standard Bayesian linear regression set-up ($U = X\beta + \varepsilon$)
- drawing from the posterior distribution becomes feasible without the need to evaluate any likelihood
- numerical advantages
  - avoids approximation of Gaussians
  - reduces curse of dimensionality (starting values)
  - faster?

Data augmentation

- *"generate a variable that wasn't there before"*
- treat the latent utilities U as parameters
- conditional on the latent utilities, the model constitutes a standard Bayesian linear regression set-up ($U = X\beta + \varepsilon$)
- drawing from the posterior distribution becomes feasible without the need to evaluate any likelihood
- numerical advantages
  - avoids approximation of Gaussians
  - reduces curse of dimensionality (starting values)
  - faster?

Data augmentation

- *"generate a variable that wasn't there before"*
- treat the latent utilities U as parameters
- conditional on the latent utilities, the model constitutes a standard Bayesian linear regression set-up ($U = X\beta + \varepsilon$)
- drawing from the posterior distribution becomes feasible without the need to evaluate any likelihood
- numerical advantages
  - avoids approximation of Gaussians
  - reduces curse of dimensionality (starting values)
  - faster?

Data augmentation

- *"generate a variable that wasn't there before"*
- treat the latent utilities U as parameters
- conditional on the latent utilities, the model constitutes a standard Bayesian linear regression set-up ($U = X\beta + \varepsilon$)
- drawing from the posterior distribution becomes feasible without the need to evaluate any likelihood
- numerical advantages
  - avoids approximation of Gaussians
  - reduces curse of dimensionality (starting values)
  - faster?

Priors

We apply the following conjugate priors:

- $(s_1, \dots, s_C) \sim D_C(\delta)$, where $D_C(\delta)$ denotes the C-dimensional Dirichlet distribution with concentration parameter vector $\delta = (\delta_1, \dots, \delta_C)$,
- $\alpha \sim \text{MVN}_{P_f}(\psi, \Psi)$,
- $b_c \sim \text{MVN}_{P_r}(\xi, \Xi)$, independent for all c,
- $\Omega_c \sim W^{-1}_{P_r}(\nu, \Theta)$, independent for all c, where $W^{-1}_{P_r}(\nu, \Theta)$ denotes the P_r-dimensional inverse Wishart distribution with ν degrees of freedom and scale matrix Θ,
- and $\Sigma \sim W^{-1}_{J-1}(\kappa, \Lambda)$.

Parameters can be set based on previous estimation results or diffuse.

Priors

We apply the following conjugate priors:

- $(s_1, \dots, s_C) \sim D_C(\delta)$, where $D_C(\delta)$ denotes the C-dimensional Dirichlet distribution with concentration parameter vector $\delta = (\delta_1, \dots, \delta_C)$,
- $\alpha \sim \text{MVN}_{P_f}(\psi, \Psi)$,
- $b_c \sim \text{MVN}_{P_r}(\xi, \Xi)$, independent for all c,
- $\Omega_c \sim W_{P_r}^{-1}(\nu, \Theta)$, independent for all c, where $W_{P_r}^{-1}(\nu, \Theta)$ denotes the P_r-dimensional inverse Wishart distribution with ν degrees of freedom and scale matrix Θ,
- and $\Sigma \sim W_{J-1}^{-1}(\kappa, \Lambda)$.

Parameters can be set based on previous estimation results or diffuse.

Priors

We apply the following conjugate priors:

- $(s_1, \ldots, s_C) \sim D_C(\delta)$, where $D_C(\delta)$ denotes the C-dimensional Dirichlet distribution with concentration parameter vector $\delta = (\delta_1, \ldots, \delta_C)$,
- $\alpha \sim \text{MVN}_{P_f}(\psi, \Psi)$,
- $b_c \sim \text{MVN}_{P_r}(\xi, \Xi)$, independent for all c,
- $\Omega_c \sim W^{-1}_{P_r}(\nu, \Theta)$, independent for all c, where $W^{-1}_{P_r}(\nu, \Theta)$ denotes the P_r-dimensional inverse Wishart distribution with ν degrees of freedom and scale matrix Θ,
- and $\Sigma \sim W^{-1}_{J-1}(\kappa, \Lambda)$.

Parameters can be set based on previous estimation results or diffuse.

Gibbs sampler

Drawing from the conditional posteriors:

1. draw $(s_1, \dots, s_C) \mid \delta, z \sim D_C$
2. draw z from its conditional distribution
3. draw $b_c \mid \Xi, \Omega, \xi, z, \beta \sim \text{MVN}_{P_r}$
4. draw $\Omega_c \mid \nu, \Theta, z, \beta, b \sim W^{-1}_{P_r}$
5. draw $U \sim \text{TMVN}_{J-1}$ via sub-Gibbs sampler (Geweke, 1998)
   - $U_{nj} \mid U_{n(-j)}, y_{nt} \sim \text{MVN}_1 \cdot \begin{cases} 1(U_{nj} > \max(U_{n(-j)}, 0)) & \text{if } y_{nt} = j \\ 1(U_{nj} < \max(U_{n(-j)}, 0)) & \text{if } y_{nt} \neq j \end{cases}$
6. draw $\alpha \mid \Psi, \psi, W, \Sigma, U, X, \beta \sim \text{MVN}_{P_f}$
7. draw $\beta_n \mid \Omega, b, X, \Sigma, U, W, \alpha \sim \text{MVN}_{P_r}$
8. draw $\Sigma \mid \kappa, \Lambda, U, W, \alpha, X, \beta \sim W^{-1}_{J-1}(\kappa + NT, \Lambda + S)$
9. start again at 1

Gibbs sampler

Drawing from the conditional posteriors:

1. draw $(s_1, \dots, s_C) \mid \delta, z \sim D_C$
2. draw z from its conditional distribution
3. draw $b_c \mid \Xi, \Omega, \xi, z, \beta \sim \text{MVN}_{P_r}$
4. draw $\Omega_c \mid \nu, \Theta, z, \beta, b \sim W_{P_r}^{-1}$
5. draw $U \sim \text{TMVN}_{J-1}$ via sub-Gibbs sampler (Geweke, 1998)
   - $U_{ntj} \mid U_{nt(-j)}, y_{nt}, \dots \sim \text{MVN}_1 \cdot \begin{cases} 1(U_{ntj} > \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} = j \\ 1(U_{ntj} < \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} \neq j \end{cases}$
6. draw $\alpha \mid \Psi, \psi, W, \Sigma, U, X, \beta \sim \text{MVN}_{P_f}$
7. draw $\beta_n \mid \Omega, b, X, \Sigma, U, W, \alpha \sim \text{MVN}_{P_r}$
8. draw $\Sigma \mid \kappa, \Lambda, U, W, \alpha, X, \beta \sim W_{J-1}^{-1}(\kappa + NT, \Lambda + S)$
9. start again at 1

Gibbs sampler

Drawing from the conditional posteriors:

1. draw $(s_1, \dots, s_C) \mid \delta, z \sim D_C$
2. draw z from its conditional distribution
3. draw $b_c \mid \Xi, \Omega, \xi, z, \beta \sim \text{MVN}_{P_r}$
4. draw $\Omega_c \mid \nu, \Theta, z, \beta, b \sim W_{P_r}^{-1}$
5. draw $U \sim \text{TMVN}_{J-1}$ via sub-Gibbs sampler (Geweke, 1998)
   - $$U_{ntj} \mid U_{nt(-j)}, y_{nt}, \dots \sim \text{MVN}_1 \cdot \begin{cases} 1(U_{ntj} > \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} = j \\ 1(U_{ntj} < \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} \neq j \end{cases}$$
6. draw $\alpha \mid \Psi, \psi, W, \Sigma, U, X, \beta \sim \text{MVN}_{P_f}$
7. draw $\beta_n \mid \Omega, b, X, \Sigma, U, W, \alpha \sim \text{MVN}_{P_r}$
8. draw $\Sigma \mid \kappa, \Lambda, U, W, \alpha, X, \beta \sim W_{J-1}^{-1}(\kappa + NT, \Lambda + S)$
9. start again at 1

Gibbs sampler

Drawing from the conditional posteriors:

1. draw $(s_1, \ldots, s_C) \mid \delta, z \sim D_C$
2. draw z from its conditional distribution
3. draw $b_c \mid \Xi, \Omega, \xi, z, \beta \sim \text{MVN}_{P_r}$
4. draw $\Omega_c \mid \nu, \Theta, z, \beta, b \sim W_{P_r}^{-1}$
5. draw $U \sim \text{TMVN}_{J-1}$ via sub-Gibbs sampler (Geweke, 1998)
   - $U_{ntj} \mid U_{nt(-j)}, y_{nt}, \cdots \sim \text{MVN}_1 \cdot \begin{cases} 1(U_{ntj} > \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} = j \\ 1(U_{ntj} < \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} \neq j \end{cases}$
6. draw $\alpha \mid \Psi, \psi, W, \Sigma, U, X, \beta \sim \text{MVN}_{P_f}$
7. draw $\beta_n \mid \Omega, b, X, \Sigma, U, W, \alpha \sim \text{MVN}_{P_r}$
8. draw $\Sigma \mid \kappa, \Lambda, U, W, \alpha, X, \beta \sim W_{J-1}^{-1}(\kappa + NT, \Lambda + S)$
9. start again at 1

Gibbs sampler

Drawing from the conditional posteriors:

1. draw $(s_1, \dots, s_C) \mid \delta, z \sim D_C$
2. draw z from its conditional distribution
3. draw $b_c \mid \Xi, \Omega, \xi, z, \beta \sim \text{MVN}_{P_r}$
4. draw $\Omega_c \mid \nu, \Theta, z, \beta, b \sim W^{-1}_{P_r}$
5. draw $U \sim \text{TMVN}_{J-1}$ via sub-Gibbs sampler (Geweke, 1998)
   - $U_{ntj} \mid U_{nt(-j)}, y_{nt}, \dots \sim \text{MVN}_1 \cdot \begin{cases} 1(U_{ntj} > \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} = j \\ 1(U_{ntj} < \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} \neq j \end{cases}$
6. draw $\alpha \mid \Psi, \psi, W, \Sigma, U, X, \beta \sim \text{MVN}_{P_f}$
7. draw $\beta_n \mid \Omega, b, X, \Sigma, U, W, \alpha \sim \text{MVN}_{P_r}$
8. draw $\Sigma \mid \kappa, \Lambda, U, W, \alpha, X, \beta \sim W^{-1}_{J-1}(\kappa + NT, \Lambda + S)$
9. start again at 1

Gibbs sampler

Drawing from the conditional posteriors:

1. draw $(s_1, \ldots, s_C) \mid \delta, z \sim D_C$
2. draw z from its conditional distribution
3. draw $b_c \mid \Xi, \Omega, \xi, z, \beta \sim \text{MVN}_{P_r}$
4. draw $\Omega_c \mid \nu, \Theta, z, \beta, b \sim W^{-1}_{P_r}$
5. draw $U \sim \text{TMVN}_{J-1}$ via sub-Gibbs sampler (Geweke, 1998)
   - $U_{ntj} \mid U_{nt(-j)}, y_{nt}, \dots \sim \text{MVN}_1 \cdot \begin{cases} 1(U_{ntj} > \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} = j \\ 1(U_{ntj} < \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} \neq j \end{cases}$
6. draw $\alpha \mid \Psi, \psi, W, \Sigma, U, X, \beta \sim \text{MVN}_{P_f}$
7. draw $\beta_n \mid \Omega, b, X, \Sigma, U, W, \alpha \sim \text{MVN}_{P_r}$
8. draw $\Sigma \mid \kappa, \Lambda, U, W, \alpha, X, \beta \sim W^{-1}_{J-1}(\kappa + NT, \Lambda + S)$
9. start again at 1

Gibbs sampler

Drawing from the conditional posteriors:

1. draw $(s_1, \dots, s_C) \mid \delta, z \sim D_C$
2. draw z from its conditional distribution
3. draw $b_c \mid \Xi, \Omega, \xi, z, \beta \sim \text{MVN}_{P_r}$
4. draw $\Omega_c \mid \nu, \Theta, z, \beta, b \sim W^{-1}_{P_r}$
5. draw $U \sim \text{TMVN}_{J-1}$ via sub-Gibbs sampler (Geweke, 1998)
   - $U_{ntj} \mid U_{nt(-j)}, y_{nt}, \dots \sim \text{MVN}_1 \cdot \begin{cases} 1(U_{ntj} > \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} = j \\ 1(U_{ntj} < \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} \neq j \end{cases}$
6. draw $\alpha \mid \Psi, \psi, W, \Sigma, U, X, \beta \sim \text{MVN}_{P_f}$
7. draw $\beta_n \mid \Omega, b, X, \Sigma, U, W, \alpha \sim \text{MVN}_{P_r}$
8. draw $\Sigma \mid \kappa, \Lambda, U, W, \alpha, X, \beta \sim W^{-1}_{J-1}(\kappa + NT, \Lambda + S)$
9. start again at 1

Gibbs sampler

Drawing from the conditional posteriors:

1. draw $(s_1, \dots, s_C) \mid \delta, z \sim D_C$
2. draw z from its conditional distribution
3. draw $b_c \mid \Xi, \Omega, \xi, z, \beta \sim \text{MVN}_{P_r}$
4. draw $\Omega_c \mid \nu, \Theta, z, \beta, b \sim W^{-1}_{P_r}$
5. draw $U \sim \text{TMVN}_{J-1}$ via sub-Gibbs sampler (Geweke, 1998)
   - $U_{ntj} \mid U_{nt(-j)}, y_{nt}, \dots \sim \text{MVN}_1 \cdot \begin{cases} 1(U_{ntj} > \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} = j \\ 1(U_{ntj} < \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} \neq j \end{cases}$
6. draw $\alpha \mid \Psi, \psi, W, \Sigma, U, X, \beta \sim \text{MVN}_{P_f}$
7. draw $\beta_n \mid \Omega, b, X, \Sigma, U, W, \alpha \sim \text{MVN}_{P_r}$
8. draw $\Sigma \mid \kappa, \Lambda, U, W, \alpha, X, \beta \sim W^{-1}_{J-1}(\kappa + NT, \Lambda + S)$
9. start again at 1

Gibbs sampler

Drawing from the conditional posteriors:

1. draw $(s_1, \dots, s_C) \mid \delta, z \sim D_C$
2. draw z from its conditional distribution
3. draw $b_c \mid \Xi, \Omega, \xi, z, \beta \sim \text{MVN}_{P_r}$
4. draw $\Omega_c \mid \nu, \Theta, z, \beta, b \sim W_{P_r}^{-1}$
5. draw $U \sim \text{TMVN}_{J-1}$ via sub-Gibbs sampler (Geweke, 1998)
   - $U_{ntj} \mid U_{nt(-j)}, y_{nt}, \dots \sim \text{MVN}_1 \cdot \begin{cases} 1(U_{ntj} > \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} = j \\ 1(U_{ntj} < \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} \neq j \end{cases}$
6. draw $\alpha \mid \Psi, \psi, W, \Sigma, U, X, \beta \sim \text{MVN}_{P_f}$
7. draw $\beta_n \mid \Omega, b, X, \Sigma, U, W, \alpha \sim \text{MVN}_{P_r}$
8. draw $\Sigma \mid \kappa, \Lambda, U, W, \alpha, X, \beta \sim W_{J-1}^{-1}(\kappa + NT, \Lambda + S)$
9. start again at 1

Gibbs sampler

Drawing from the conditional posteriors:

1. draw $(s_1, \dots, s_C) \mid \delta, z \sim D_C$
2. draw z from its conditional distribution
3. draw $b_c \mid \Xi, \Omega, \xi, z, \beta \sim \text{MVN}_{P_r}$
4. draw $\Omega_c \mid \nu, \Theta, z, \beta, b \sim W^{-1}_{P_r}$
5. draw $U \sim \text{TMVN}_{J-1}$ via sub-Gibbs sampler (Geweke, 1998)
   - $U_{ntj} \mid U_{nt(-j)}, y_{nt}, \dots \sim \text{MVN}_1 \cdot \begin{cases} 1(U_{ntj} > \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} = j \\ 1(U_{ntj} < \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} \neq j \end{cases}$
6. draw $\alpha \mid \Psi, \psi, W, \Sigma, U, X, \beta \sim \text{MVN}_{P_f}$
7. draw $\beta_n \mid \Omega, b, X, \Sigma, U, W, \alpha \sim \text{MVN}_{P_r}$
8. draw $\Sigma \mid \kappa, \Lambda, U, W, \alpha, X, \beta \sim W^{-1}_{J-1}(\kappa + NT, \Lambda + S)$
9. start again at 1

Gibbs sampler

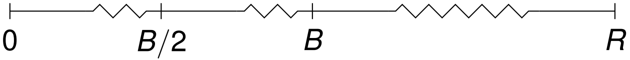
Drawing from the conditional posteriors:

1. draw $(s_1, \dots, s_C) \mid \delta, z \sim D_C$
2. draw z from its conditional distribution
3. draw $b_c \mid \Xi, \Omega, \xi, z, \beta \sim \text{MVN}_{P_r}$
4. draw $\Omega_c \mid \nu, \Theta, z, \beta, b \sim W^{-1}_{P_r}$
5. draw $U \sim \text{TMVN}_{J-1}$ via sub-Gibbs sampler (Geweke, 1998)
   - $U_{ntj} \mid U_{nt(-j)}, y_{nt}, \dots \sim \text{MVN}_1 \cdot \begin{cases} 1(U_{ntj} > \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} = j \\ 1(U_{ntj} < \max(U_{nt(-j)}, 0)) & \text{if } y_{nt} \neq j \end{cases}$
6. draw $\alpha \mid \Psi, \psi, W, \Sigma, U, X, \beta \sim \text{MVN}_{P_f}$
7. draw $\beta_n \mid \Omega, b, X, \Sigma, U, W, \alpha \sim \text{MVN}_{P_r}$
8. draw $\Sigma \mid \kappa, \Lambda, U, W, \alpha, X, \beta \sim W^{-1}_{J-1}(\kappa + NT, \Lambda + S)$
9. start again at 1

Gibbs sampler

Periods:

- $0, \ldots, B$ – discard draws
- $B/2, \ldots, B$ – latent class updating
- $B, \ldots, R$ – keep every Qth draw



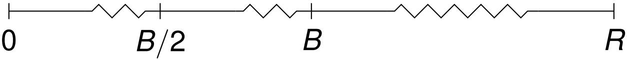
Normalization (Imai and van Dyk, 2005):

- Σ is drawn from the unrestricted space of symmetric, positive-definite matrices, therefore the samples lack identification
- normalize $\alpha^{(i)}/\sqrt{(\Sigma^{(i)})_{11}}$, $b_c^{(i)}/\sqrt{(\Sigma^{(i)})_{11}}$, $\Omega_c^{(i)}/(\Sigma^{(i)})_{11}$, $\Sigma^{(i)}/(\Sigma^{(i)})_{11}$

Gibbs sampler

Periods:

- $0, \ldots, B$ – discard draws
- $B/2, \ldots, B$ – latent class updating
- $B, \ldots, R$ – keep every Qth draw



Normalization (Imai and van Dyk, 2005):

- Σ is drawn from the unrestricted space of symmetric, positive-definite matrices, therefore the samples lack identification
- normalize $\alpha^{(i)}/\sqrt{(\Sigma^{(i)})_{11}}$, $b_c^{(i)}/\sqrt{(\Sigma^{(i)})_{11}}$, $\Omega_c^{(i)}/(\Sigma^{(i)})_{11}$, $\Sigma^{(i)}/(\Sigma^{(i)})_{11}$

Latent class updating scheme

Within the second half of the burn-in period, every 50th iteration:

- Remove class c, if $s_c < \varepsilon_{\text{min}}$.
- Split class c into two classes c_1 and c_2, if $s_c > \varepsilon_{\text{max}}$. The class means b_{c_1} and b_{c_2} of the new classes c_1 and c_2 are shifted in opposite directions from the class mean b_c of the old class c in the direction of the highest variance.
- Join two classes c_1 and c_2 to one class c, if $\|b_{c_1} - b_{c_2}\| < \varepsilon_{\text{distmin}}$. The parameters of c are assigned by adding the values of s from c_1 and c_2 and averaging the values for b and Ω.

Latent class updating scheme

Within the second half of the burn-in period, every 50th iteration:

- Remove class c, if $s_c < \varepsilon_{\text{min}}$.
- Split class c into two classes c_1 and c_2, if $s_c > \varepsilon_{\text{max}}$. The class means b_{c_1} and b_{c_2} of the new classes c_1 and c_2 are shifted in opposite directions from the class mean b_c of the old class c in the direction of the highest variance.
- Join two classes c_1 and c_2 to one class c, if $\|b_{c_1} - b_{c_2}\| < \varepsilon_{\text{distmin}}$. The parameters of c are assigned by adding the values of s from c_1 and c_2 and averaging the values for b and Ω.

Latent class updating scheme

Within the second half of the burn-in period, every 50th iteration:

- Remove class c, if $s_c < \varepsilon_{\text{min}}$.
- Split class c into two classes c_1 and c_2, if $s_c > \varepsilon_{\text{max}}$. The class means b_{c_1} and b_{c_2} of the new classes c_1 and c_2 are shifted in opposite directions from the class mean b_c of the old class c in the direction of the highest variance.
- Join two classes c_1 and c_2 to one class c, if $\|b_{c_1} - b_{c_2}\| < \varepsilon_{\text{distmin}}$. The parameters of c are assigned by adding the values of s from c_1 and c_2 and averaging the values for b and Ω.

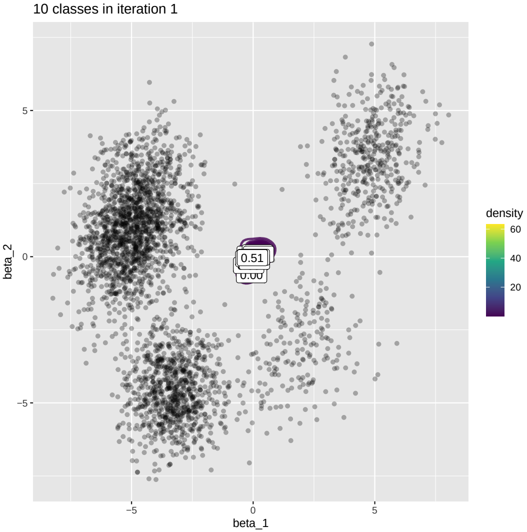
Latent class updating scheme

Within the second half of the burn-in period, every 50th iteration:

- Remove class c, if $s_c < \varepsilon_{\text{min}}$.
- Split class c into two classes c_1 and c_2, if $s_c > \varepsilon_{\text{max}}$. The class means b_{c_1} and b_{c_2} of the new classes c_1 and c_2 are shifted in opposite directions from the class mean b_c of the old class c in the direction of the highest variance.
- Join two classes c_1 and c_2 to one class c, if $\|b_{c_1} - b_{c_2}\| < \varepsilon_{\text{distmin}}$. The parameters of c are assigned by adding the values of s from c_1 and c_2 and averaging the values for b and Ω.

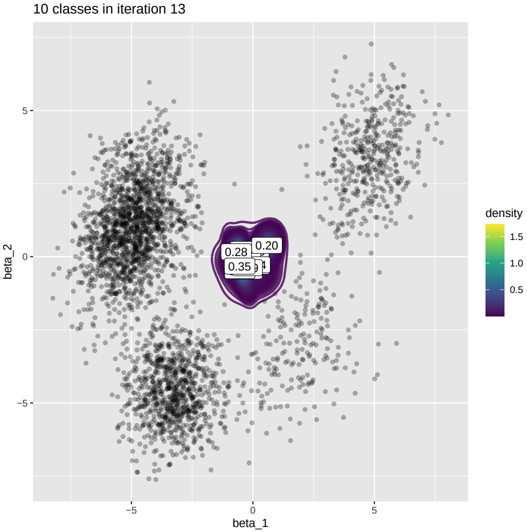
Simulation 1

$P_r = 2$ with 4 latent classes



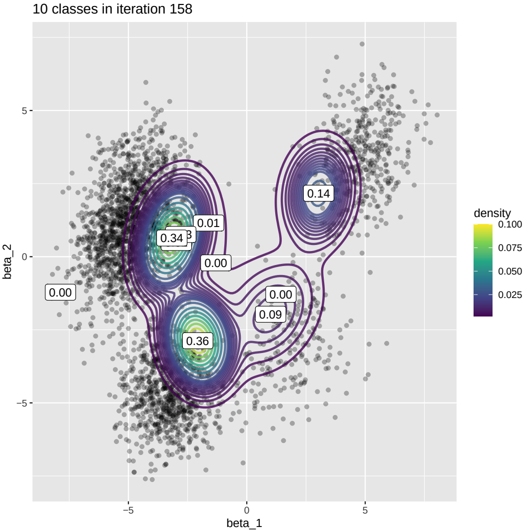
Simulation 1

$P_r = 2$ with 4 true latent classes



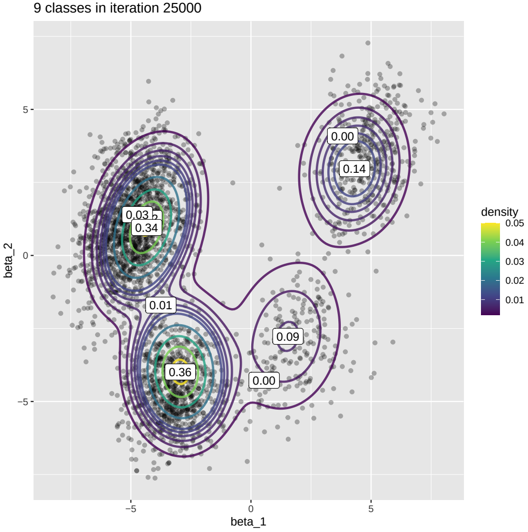
Simulation 1

$P_r = 2$ with 4 true latent classes



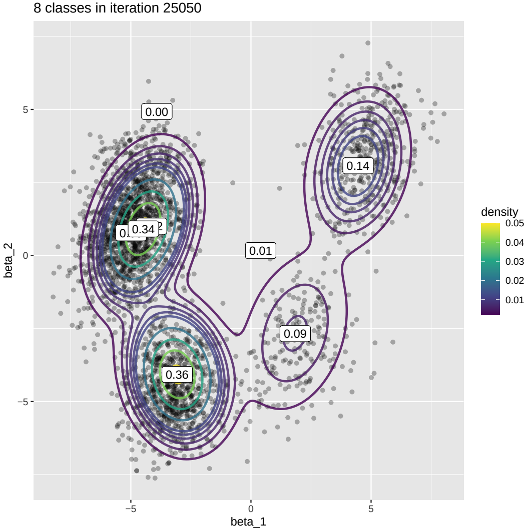
Simulation 1

$P_r = 2$ with 4 true latent classes



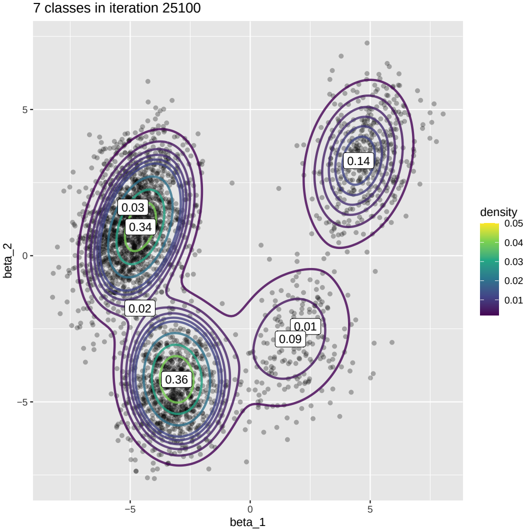
Simulation 1

$P_r = 2$ with 4 true latent classes



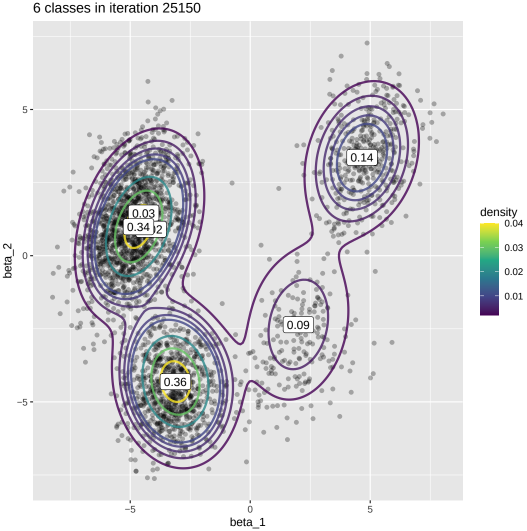
Simulation 1

$P_r = 2$ with 4 true latent classes



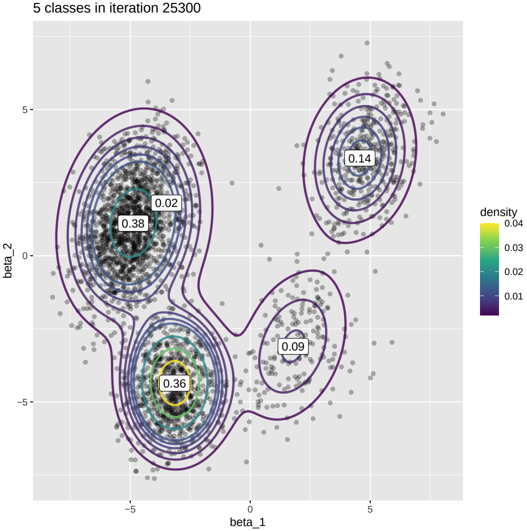
Simulation 1

$P_r = 2$ with 4 true latent classes



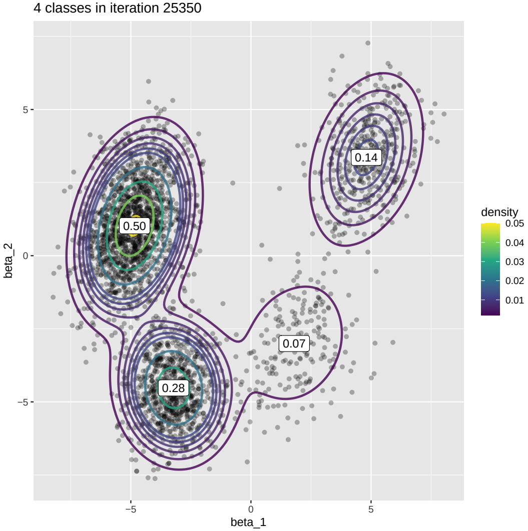
Simulation 1

$P_r = 2$ with 4 true latent classes



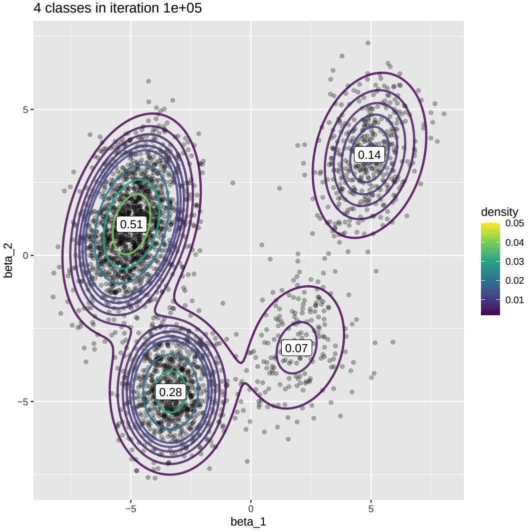
Simulation 1

$P_r = 2$ with 4 true latent classes



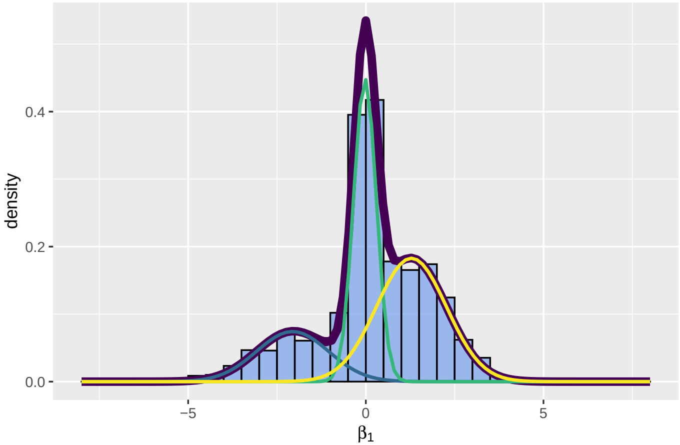
Simulation 1

$P_r = 2$ with 4 true latent classes



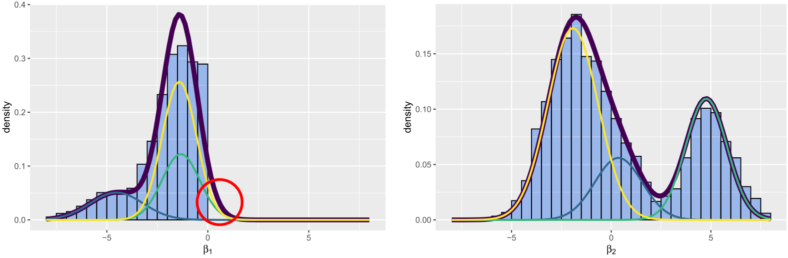
Simulation 2

$P_r = 1$, controversial choice attribute (e.g. out-of-vehicle travel time)



Simulation 3

$P_r = 2$, sign-restricted choice attribute (e.g. cost)



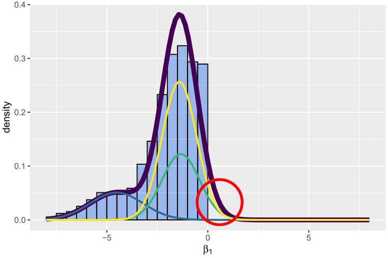
Estimated mixing distribution (posterior mean as point estimate):

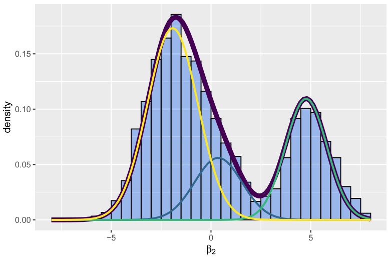
$$\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} \sim 0.16 \cdot \text{MVN}_2\left(\begin{pmatrix}-4.52\\ 0.36\end{pmatrix}, \begin{pmatrix}1.72 & -0.28\\ \cdot & 1.41\end{pmatrix}\right) + 0.29 \cdot \text{MVN}_2\left(\begin{pmatrix}-1.33\\ 4.77\end{pmatrix}, \begin{pmatrix}0.86 & 0.13\\ \cdot & 1.08\end{pmatrix}\right)$$
$$+\ 0.55 \cdot \text{MVN}_2\left(\begin{pmatrix}-1.41\\ -1.92\end{pmatrix}, \begin{pmatrix}0.72 & 0.15\\ \cdot & 1.59\end{pmatrix}\right)$$

Thanks for listening, questions please!

Simulation 3

$P_r = 2$, sign-restricted choice attribute (e.g. cost)





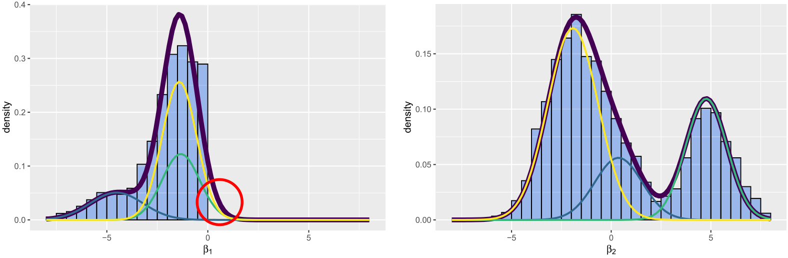
Estimated mixing distribution (posterior mean as point estimate):

$$\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} \sim 0.16 \cdot \text{MVN}_2\left(\begin{pmatrix}-4.52\\ 0.36\end{pmatrix}, \begin{pmatrix}1.72 & -0.28\\ \cdot & 1.41\end{pmatrix}\right) + 0.29 \cdot \text{MVN}_2\left(\begin{pmatrix}-1.33\\ 4.77\end{pmatrix}, \begin{pmatrix}0.86 & 0.13\\ \cdot & 1.08\end{pmatrix}\right)$$
$$+\ 0.55 \cdot \text{MVN}_2\left(\begin{pmatrix}-1.41\\ -1.92\end{pmatrix}, \begin{pmatrix}0.72 & 0.15\\ \cdot & 1.59\end{pmatrix}\right)$$

Thanks for listening, questions please!

Simulation 3

$P_r = 2$, sign-restricted choice attribute (e.g. cost)



Estimated mixing distribution (posterior mean as point estimate):

$$\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} \sim 0.16 \cdot \text{MVN}_2\left(\begin{pmatrix}-4.52\\ 0.36\end{pmatrix}, \begin{pmatrix}1.72 & -0.28\\ \cdot & 1.41\end{pmatrix}\right) + 0.29 \cdot \text{MVN}_2\left(\begin{pmatrix}-1.33\\ 4.77\end{pmatrix}, \begin{pmatrix}0.86 & 0.13\\ \cdot & 1.08\end{pmatrix}\right)$$
$$+\ 0.55 \cdot \text{MVN}_2\left(\begin{pmatrix}-1.41\\ -1.92\end{pmatrix}, \begin{pmatrix}0.72 & 0.15\\ \cdot & 1.59\end{pmatrix}\right)$$

Thanks for listening, questions please!