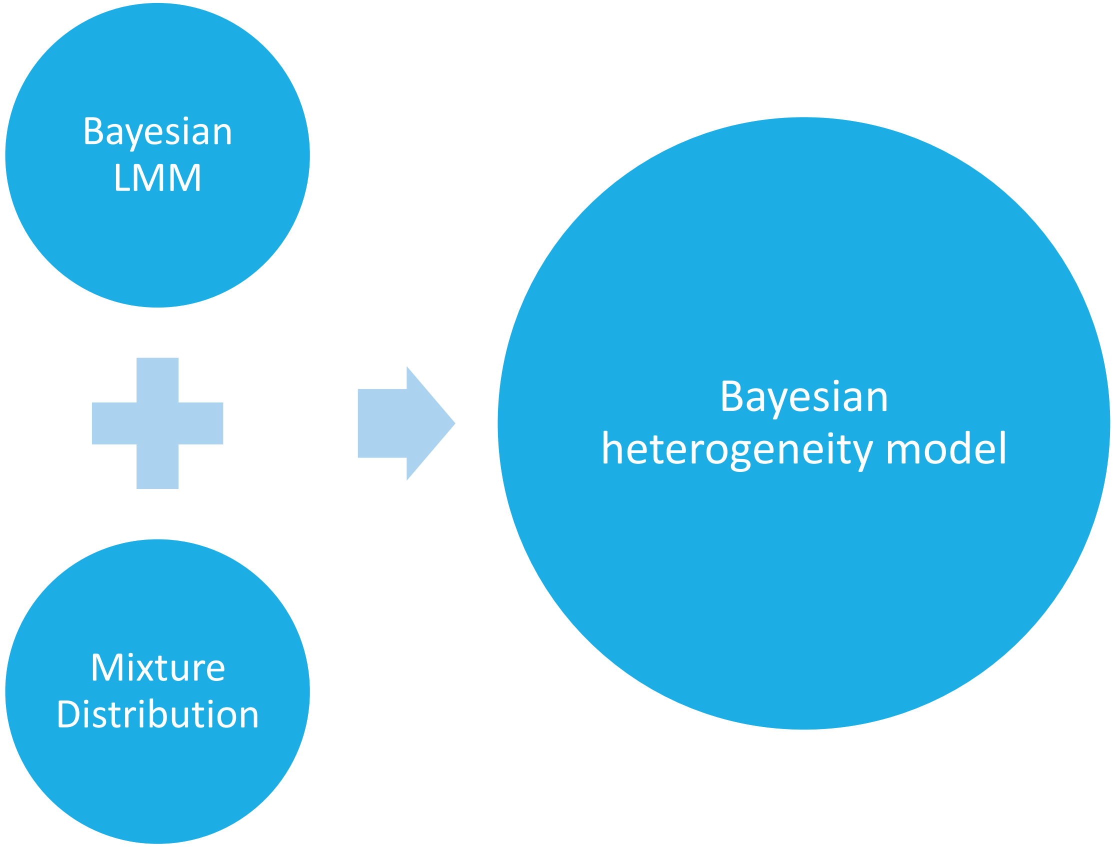
The use of mixture distributions in a Bayesian linear mixed effects model

Anirudh TOMER

Promoter: Professor Emmanuel LESAFFRE

KU LEUVEN

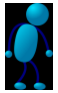
What are we dealing with?

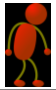


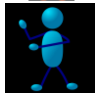
Bayesian heterogeneity model


Gender: Male, Age: 37


Gender: Female, Age: 23


Gender: Male, Age: 41

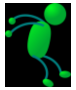

Gender: Female, Age: 59

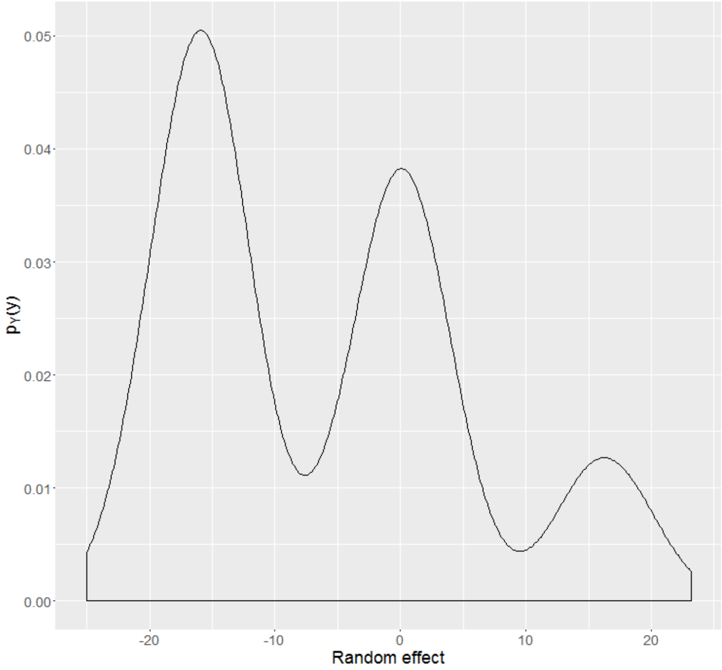

Gender: Female, Age: 35

. .

. .

. .


Gender: Male, Age: 19



Mathematical notation

$\boldsymbol{y_i} = \boldsymbol{X_i\beta} + \boldsymbol{Z_i b}_i + \boldsymbol{\varepsilon_i}$, where $1 \leq i \leq n,$

$\boldsymbol{y_i} = \left(y_{i1}, y_{i2}, \dots, y_{im_i}\right)^T$ is a vector of observations for the i^{th} subject taken at m_i time points,

$Xi = \left(x_{i1}^T, x_{i2}^T, \dots, x_{im_i}^T\right)^T$ is the $m_i \times (d + 1)$ design matrix for the i^{th} subject,

$\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_d)^T$ is a $(d + 1) \times 1$ vector of fixed effects with β_0 being the intercept,

$\boldsymbol{Z_i} = \left(z_{i1}^T, z_{i2}^T, \dots, z_{im_i}^T\right)^T$ is the $m_i \times q$ design matrix of covariates varying for a subject at each observation,

$\boldsymbol{b_i} = \left(b_{0i}, b_{1i}, \dots, b_{(q-1)i}\right)^T$ is a $q \times 1$ vector of random effects with b_{0i} being the random intercept.

The random effects $\boldsymbol{b_i} \sim \Sigma_{i=1}^{K} \eta_i N_q(\boldsymbol{b}_k^C, G_k),$ with b_k^C and G_k being the mean vector and covariance matrices for the k^{th} component in the mixture distribution respectively,

$\varepsilon_i = \left(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{im_i}\right)^T$ is a $m_i \times 1$ vector of measurement errors. The errors $\varepsilon_i \sim N_{m_i}(0, R_i)$ with the covariance matrix of errors R_i usually being $\sigma^2 I_{m_i}$.

The parameters $\boldsymbol{\beta}, \boldsymbol{\sigma^2}, \boldsymbol{b_1^C}, \boldsymbol{b_2^C}, \dots, \boldsymbol{b_k^C}, G_1, G_2, \dots G_k$ all have a prior distribution.

KU LEUVEN

What is the problem we are facing?

Estimation of component density parameters θ_G

Criteria for choice of number of components (DIC, Bayes Factor, Predictive methods)

Classification of observations into groups

Simulation study dataset

- 10 repeated measurements which are equally spaced for all subjects.
- Predictors were gender, birth year, time of measurement.
- Various versions of the dataset with very clearly distinguishable mixture components to almost fused mixture components were considered.
- No serial correlation considered.
- In some of the recent runs I also allowed component variances to differ.

DIC (deviance information criteria)

$DIC = \overline{D(\boldsymbol{\theta})} + p_D$

Where $p_D = \overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}})$ is the effective number of parameters in the model.

$D(\theta) = -2 \log p(\boldsymbol{y}|\boldsymbol{\theta}) + 2 \log f(\boldsymbol{y})$ is called the Bayesian deviance.

Conditional

- Conditional on knowing allocations.
- Conditional on knowing the random effect.

Observed

- Likelihood based on observed part of the data.

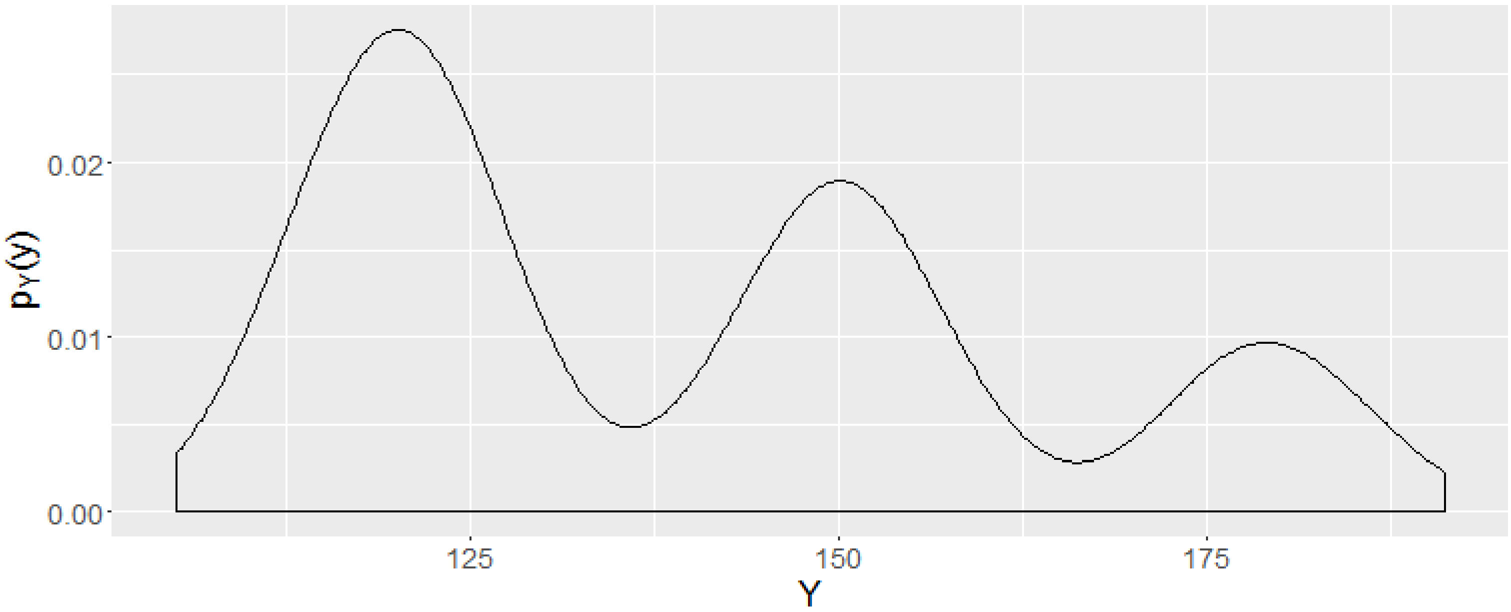
Complete

- Model observed and unobserved data both.
- Random effect and allocations are unobserved.

Results for 3 well separated components

Fitted Components	DIC 1	DIC 2	DIC 4	DIC 5	DIC 7
1	10531.02 (pD = 15.39)	10528.17 (pD = 12.53)	25837.71 (pD = 26.65)	25966.21 (pD = 155.15)	9213.49 (pD = 112.43)
2	10622.86 (pD = 8.59)	10642.80 (pD = 28.52)	26228.60 (pD = 42.98)	26636.60 (pD = 450.99)	9166.69 (pD = 76.05)
3	13938.77 (pD = 4.41)	13966.63 (pD = 32.27)	22875.79 (pD = 63.78)	23118.10 (pD = 306.08)	9247.58 (pD = 157.10)
4	12364.91 (pD = 8.12)	12387.15 (pD = 30.36)	23767.86 (pD = 84.17)	23965.74 (pD = 282.05)	9250.44 (pD = 160.53)
5	14371.40 (pD = 149.29)	14256.81 (pD = 34.71)	23542.39 (pD = 87.64)	22971.57 (pD = -483.18)	9258.069 (pD = 166.94)
6	14521.31 (pD = 241.9265)	14081.84 (pD = -197.55)	23761.01 (pD = 99.70)	21755.54 (pD = -1905.77)	9255.67 (pD = 165.18)

Well separated but unequal proportions

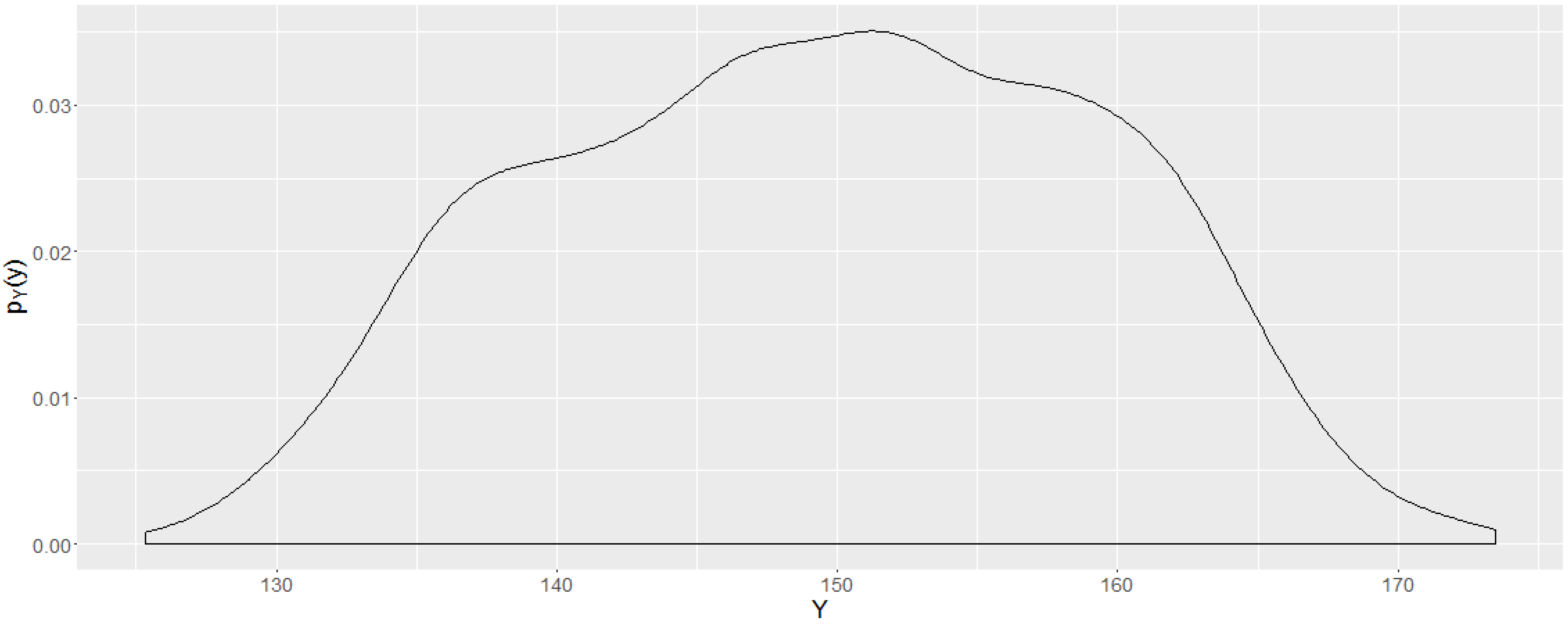


Results for 3 well separated components, un

Fitted Components	DIC 1	DIC 2	DIC 4	DIC 5	DIC 7
1	7028.43 (pD = 11.87)	7035.24 (pD = 18.68)	17040.36 (pD = 25.78)	17076.01 (pD = 61.43)	6110.43 (pD = 21.68)
2	7101.44 (pD = 14.22)	7101.873 (pD = 14.66)	17270.44 (pD = 44.16)	17433.49 (pD = 207.21)	6061.03 (pD = -21.71)
3	7750.93 (pD = 9.28)	7770.39 (pD = 28.73)	15382.12 (pD = 64.10)	15586.91 (pD = 268.89)	6189.61 (pD = 106.94)
4	7783.76 (pD = 7.63)	7854.453 (pD = 78.32)	15658.16 (pD = 77.92)	16037.37 (pD = 457.14)	6189.28 (pD = 105.68)
5	7624.78 (pD = -162.35)	7816.59 (pD = 29.45)	15880.63 (pD = 85.80)	12865.95 (pD = -2928.87)	6191.27 (pD = 108.37)

KU LEUVEN

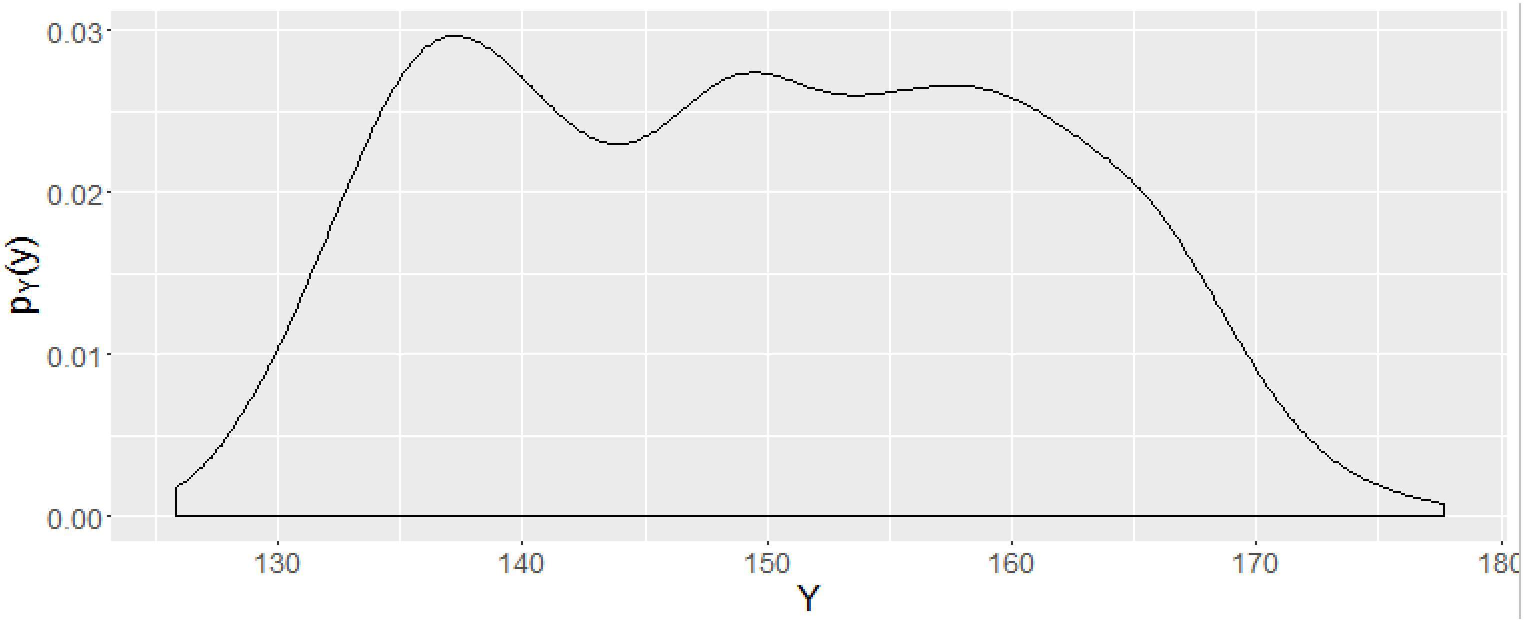
Partially fused components



Results for 3 partially fused components

Fitted Components	DIC 1	DIC 2	DIC 4	DIC 5	DIC 7
1	11652.52 (pD = 16.86)	11631.82 (pD = -3.85)	22250.66 (pD = 25.57)	22449.76 (pD = 224.67)	9252.85 (pD = 165.91)
2	13050.51 (pD = -0.66)	13045.97 (pD = -5.20)	23105.93 (pD = 42.46)	23751.53 (pD = 688.06)	9235.00 (pD = 157.26)
3	15383.37 (pD = -99.51)	14436.66 (pD = -1046.22)	23209.09 (pD = 62.97)	24137.68 (pD = 991.56)	9238.93 (pD = 161.37)
4	15701.01 (pD = 77.61)	14779.17 (pD = -844.23)	23517.28 (pD = 75.05)	22195.75 (pD = -1246.48)	9230.14 (pD = 153.80)
5	16149.49 (pD = 295.3527)	15155.12 (pD = -699.01)	23812.6 (pD = 87.08)	22494.82 (pD = -1230.70)	9237.79 (pD = 159.73)
6	16806.97 (pD = 964.85)	15324.64 (pD = -517.49)	24151.71 (pD = 99.13)	22906.95 (pD = -1145.64)	9238.55 (pD = 162.99)

Partially fused components



KU LEUVEN

Results for 3 partially fused components

Fitted Components	DIC 1	DIC 2	DIC 4	DIC 5	DIC 7
1	7478.89 (pD = 13.23)	7476.65 (pD = 10.99)	15304.93 (pD = 26.31)	15428.91 (pD = 150.29)	6183.127 (pD = 99.80)
2	8300.28 (pD = 6.79)	8383.95 (pD = 90.47)	15582.35 (pD = 43.18)	15931.90 (pD = 392.74)	6181.79 (pD = 104.65)
3	9284.91 (pD = -245.00)	9046.34 (pD = -483.57)	15528.08 (pD = 62.31)	15995.44 (pD = 529.68)	6185.74 (pD = 108.12)
4	10047.85 (pD = 182.57)	9762.34 (pD = -102.94)	15670.82 (pD = 76.11)	13523.96 (pD = -2070.75)	6186.42 (pD = 109.33)

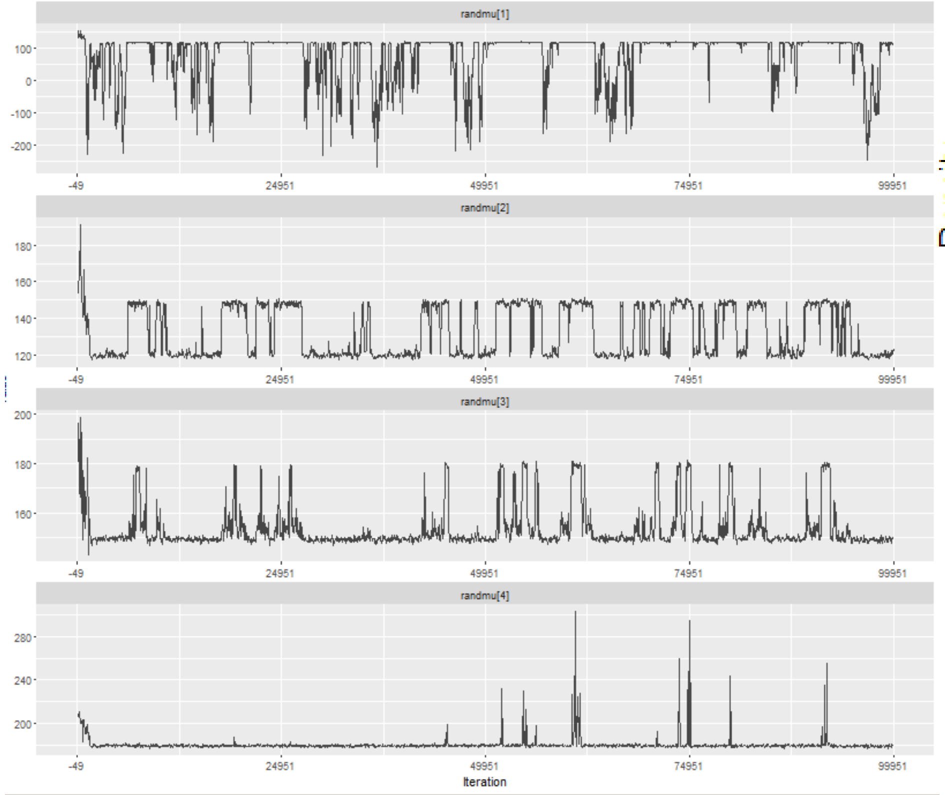
Discussion

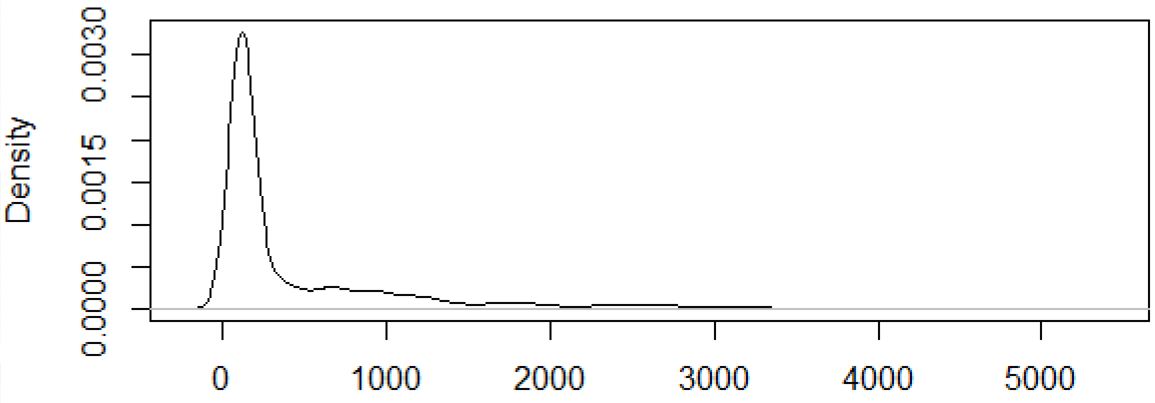
- DIC 4 works best when components are well separated
- pD in DIC1 gives an indication when components are well separated?
- Interpretation of the DIC
- Some simulations are incorrect as JAGS does not throw away burn in part

Posterior predictive checks

- The various summary measures I tried (based on Frühwirth-Schnatter, 2006):
  mean, variance, skewness, kurtosis and even higher order moments of the observed data
- In higher dimensions: Mardia's skewness and kurtosis and a few more test statistics
- In one dimension the posterior of component means can be checked
- More detailed questions: Are the mixture proportions correctly identified?
- Why some of the posterior checks aren't working? Perhaps checking for model fit is not a good approach because with more components it is never bad.

Mahalanobis distance of observed data to components





Bayes Factor

Number of components (Original/Fitted)	2	3	4	5	6
3	-5201.17	-5095.25	-5098.50	-5092.53	-5104.56
5	-8453.78	-8725.04	-8662.46	-8617.31	-8608.85

- Implementation based on Chib (1995)
- Loose pattern: Under fitting gives smaller marginal log likelihood

What remains?

More simulations/Confirmatory simulations.

- Double check my code.
- More complex scenarios.

Modelling the blood donor data.

- I doubt I will be able to run so many observations on my computer.
- I am using some parallel processing packages for post MCMC processing.

Answer why things work, when I don’t expect them to? Or why they don’t when I expect them to work?

- I do lack some of the theoretical foundations.

A few more things

Changing title of the thesis

- Bayesian model selection for Bayesian heterogeneity model.
- Bayesian model selection for Bayesian linear mixed models having mixture distributed random effects.

Thank you MCMC

- I get to pursue other activities between work.
- Have you heard about "beepr"?

Feedback and questions from the Panel