Bayesian Project

王嘉雪 姜悦 师珑天 周劲松 谭致恒

Department of Statistics and Data Science, SUSTech

2024.6.7

Theoretical Derivation

One-way ANOVA:

$$y_{ij} = \beta_0 + u_j + e_{ij} \qquad e_{ij} \sim N(0, \sigma_e^2) \qquad u_j \sim N(0, \sigma_u^2)$$

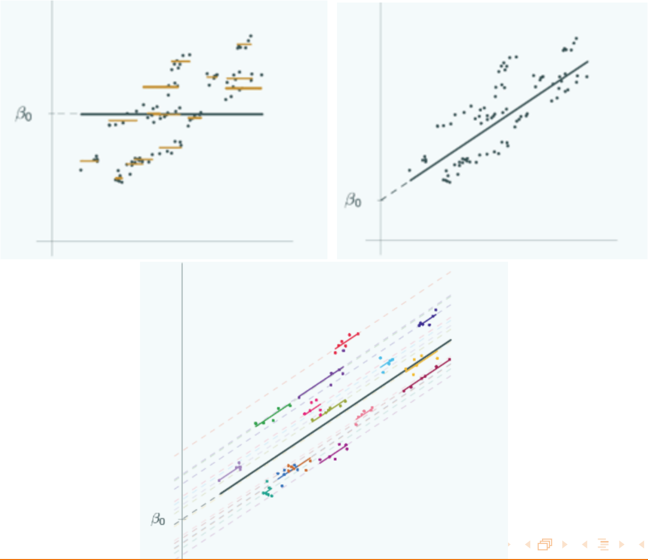
Regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad e_i \sim N(0, \sigma^2)$$

Random intercept model: The target of ANCOVA is to look at group effects on the response, while controlling for the effects of other continuous variables on the response.

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij} \qquad e_{ij} \sim N(0, \sigma_e^2) \qquad u_j \sim N(0, \sigma_u^2)$$

Theoretical Derivation



Theoretical Derivation

$$\begin{aligned} logit(\mu_{ij}) &= \beta_0 + \beta_1 Gender_{ij} + \beta_2 Race_{ij} + \beta_3 Age_{ij} + \beta_4 MHV_{1ij} \\ &+ \beta_5 MHV_{2ij} + \beta_6 MHV_{3ij} + \beta_7 INPTMHV_{ij} + u_i \\ &:= x_{ij}\beta + u_i \qquad fori = 1, \cdots, 500;\ j = 1, \cdots, 7, \end{aligned} \tag{1}$$

where $u_i \sim \mathcal{N}(0, e^{2\xi})$ with priors $\beta_j \sim \mathcal{N}(0, 100), \xi \sim \mathcal{N}(0, 100)$.

Define μ_{ij} as the probability of the subject taking drugs from 3 or more different groups, and define z_{ij} as whether the subject is taking drugs from 3 or more different groups.

$$\begin{aligned} \mu_{ij}|(\beta, u_i) &= logit^{-1}(x_{ij}\beta + u_i) \\ z_{ij}|\mu_{ij} &\sim Bernoulli(\mu_{ij}) \end{aligned} \tag{2}$$

Connection with logistic regression.

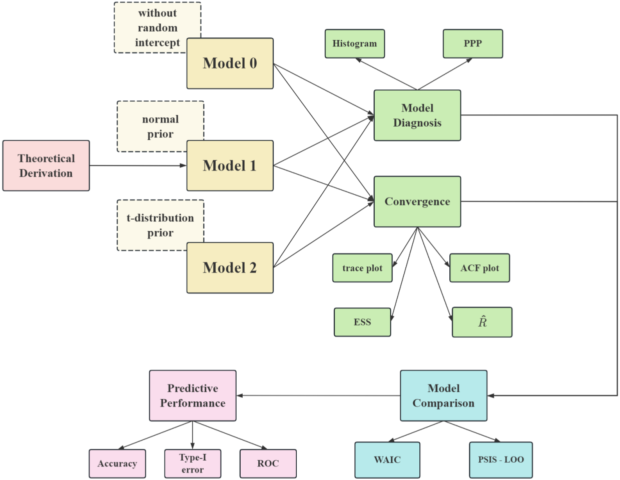
Theoretical Derivation

$$
\begin{aligned}
p(\beta, u_i, \xi, |z_{ij}) &\propto \prod_{i=1}^{500}\prod_{j=1}^{7} p(z_{ij}|\mu_{ij}) \cdot p(\beta) \cdot p(u_i|\xi) \cdot p(\xi) \\
&= \prod_{i=1}^{500}\prod_{j=1}^{7} \mu_{ij}^{z_{ij}} (1-\mu_{ij})^{1-z_{ij}} \cdot \left(\prod_{k=0}^{7} \frac{1}{\sqrt{2\pi\sigma}} \exp\{-\frac{\beta_k^2}{2\sigma^2}\} \right) \\
&\cdot \frac{1}{\sqrt{2\pi\sigma_u}} \cdot \exp\{-\frac{u_i^2}{2\sigma_u^2}\} \cdot \frac{1}{\sqrt{2\pi\sigma}} \exp\{-\frac{\xi^2}{2\sigma^2}\},
\end{aligned} \tag{3}
$$

where $\mu_{ij} = logit^{-1}(x_{ij}\beta + u_i)$, $\sigma = 10$ and $\sigma_u = e^{2\xi}$.

We have $8 + 1 + 500 = 509$ unknown parameters.

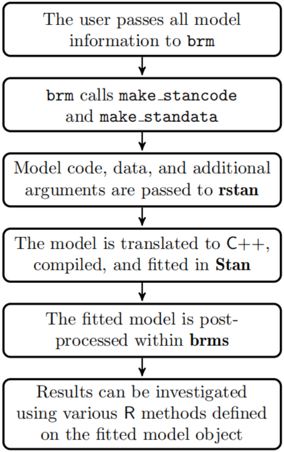
Structure


without
random
intercept
Model 0
normal
prior
Model 1
t-distribution
prior
Model 2
Theoretical
Derivation
Histogram
PPP
Model
Diagnosis
Convergence
trace plot
ACF plot
ESS
$\hat{R}$
Predictive
Performance
Model
Comparison
Accuracy
Type-I
error
ROC
WAIC
PSIS - LOO

1. Theoretical Derivation
2. Structure
3. Model Implement and Model Diagnosis
   - Introduction to brms R package
   - Model 1
   - Model 2 (t-distribution Prior)
   - Model 0 (Without Random Intercept)
4. Model Comparison
5. Predictive Performance

Introduction to brms R package

- Due to the limitations of R's computational accuracy and speed, we call the brms package, which can interact with C++ for computation.
- Stan implements Hamiltonian Monte Carlo and its extension, the No-U-Turn Sampler (NUTS). These algorithms converge much more quickly especially for high-dimensional models regardless of whether the priors are conjugate or not.



Model 1: Codes

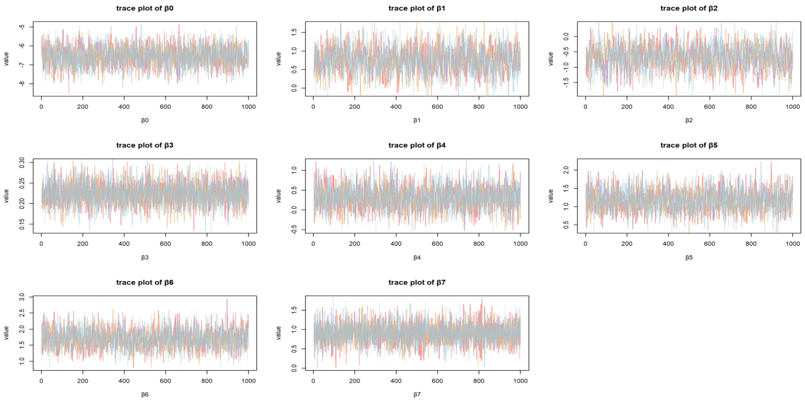
```
nlform <- bf(polypharmacy ~
    gender+race+age+mhv1+mhv2+mhv3+inptmhv+(1|id))

nlprior <- c(
  set_prior('normal(0, 10)', class = "Intercept"),
  set_prior('normal(0, 10)', class = "b"),
  set_prior('normal(0, exp(2*tau))', class = 'sd', group
      = 'id'), # here is where we add tau as a
      hyperparameter;
  set_prior("target += normal_lpdf(tau | 0, 10)", check =
      FALSE) # here is where we define the prior for tau;
)

stanvars <- stanvar(scode = " real<lower=0> tau;", # here
    is where we add the parameter for tau
                          block = "parameters")
```

Model 1: Codes

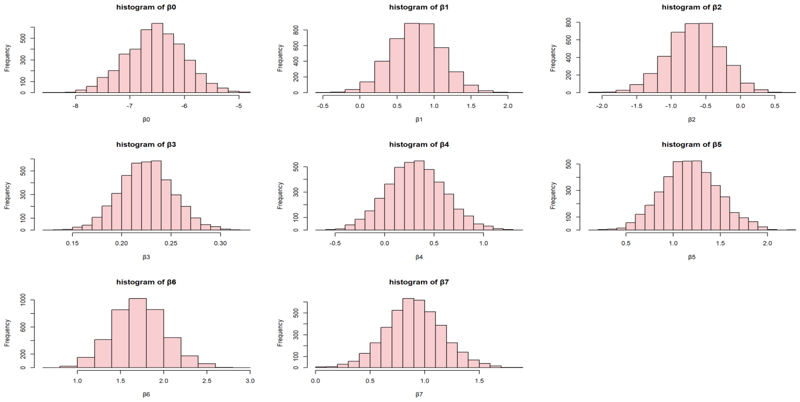
```

fit <- brm( formula = nlform,
            data = data1,
            family = bernoulli(link = 'logit'),
            prior = nlprior,
            stanvars = stanvars,
            warmup = 1000, iter = 2000, chains = 4,
            control = list(adapt_delta = 0.95),
            save_all_pars = TRUE
            )

fit <- add_criterion(fit, c("loo", "waic"))
```

Model 1: Trace Plot

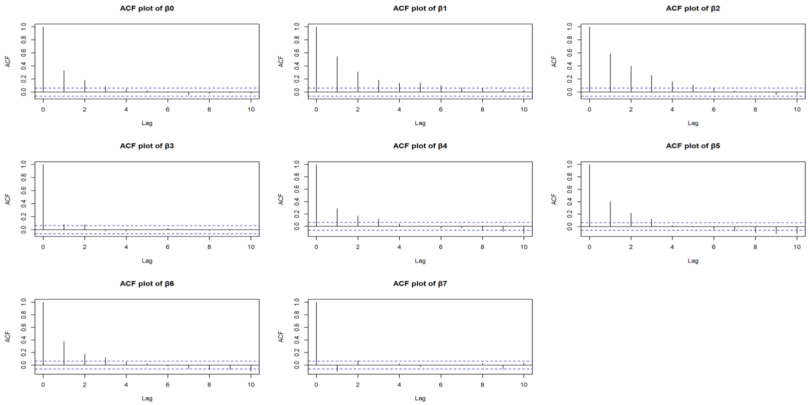
2,000 iterations (with 1,000 burn-in).



Model 1: Histogram



Model 1: ACF Plot



Model 1: Results

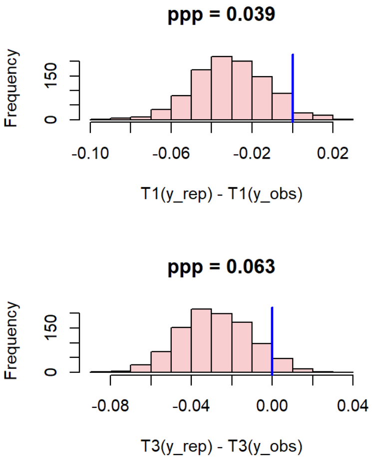
	Estimate	Error	$\hat{R}$	Bulk-ESS	Tail- ESS
Intercept	-6.53	0.53	1.00	2080	2405
gender	0.76	0.34	1.00	1220	2431
race	-0.67	0.39	1.00	1079	1884
age	0.22	0.03	1.00	3733	2895
mhv1	0.32	0.29	1.00	2306	2672
mhv2	1.18	0.30	1.00	1911	2517
mhv3	1.71	0.30	1.00	1943	2357
inptmhv	0.91	0.26	1.00	5387	2817
sd(Intercept)	2.48	0.17	1.00	995	1870

Model Diagnosis

Motivated by the adolescent-smoking example in the textbook, three test statistics are used here:

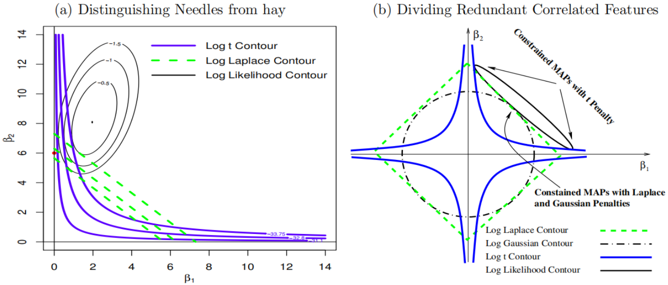
- $T_1(y)$: The percentage of subjects in the sample whose polypharmacy is always 0 in 7 years.
- $T_2(y)$: The percentage of subjects in the sample whose polypharmacy is always 1 in 7 years.
- $T_3(y)$: The percentage of subjects in the sample whose polypharmacy switch from 0 to 1 in 7 years.

Model 1: Diagnosis



Model 2 (t-distribution Prior)

Figure 1: Geometric illustrations of the properties of t penalty in MAP inference.

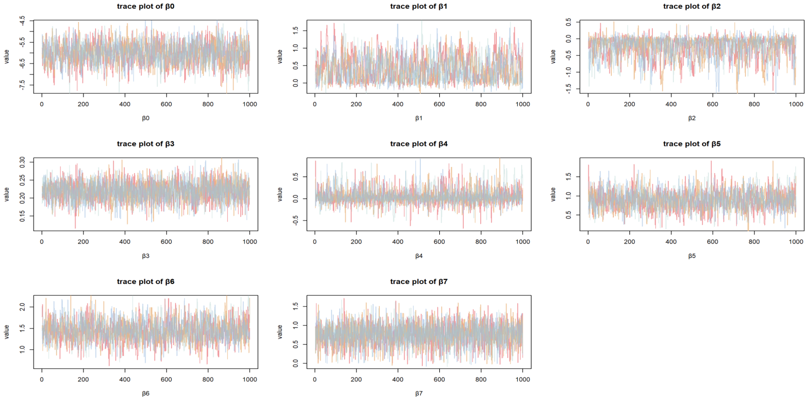


(Li and Yao, 2018)

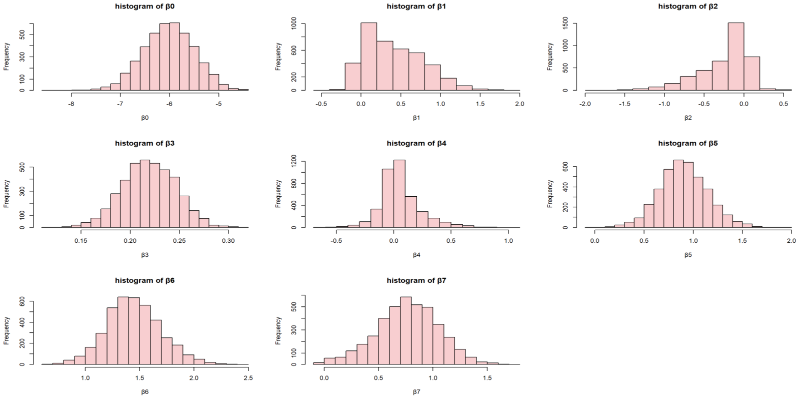
In model 2, we set priors of β_i to be $t(df = 0.5)$.

Model 2: Trace Plot

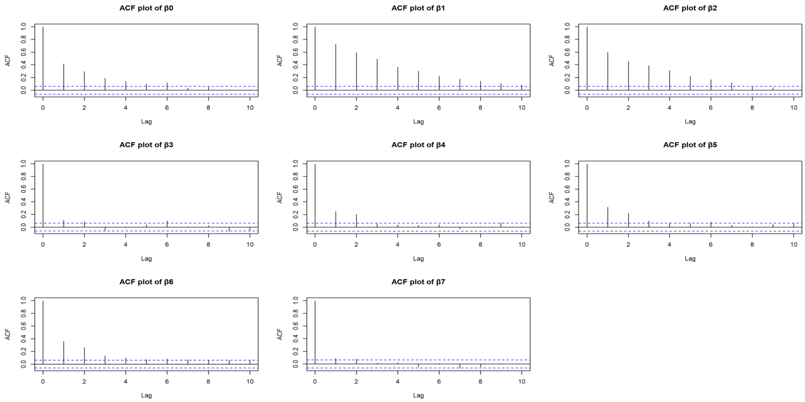
2,000 iterations (with 1,000 burn-in).



Model 2: Histogram



Model 2: ACF Plot

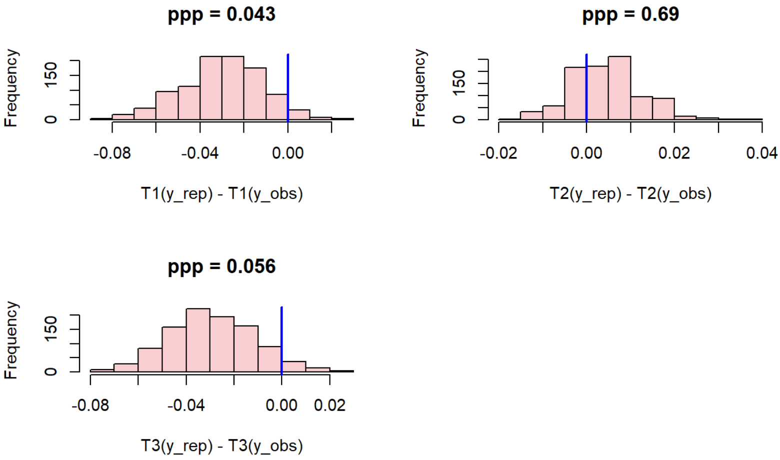


Model 2: Results

	Estimate	Error	$\hat{R}$	Bulk-ESS	Tail- ESS
Intercept	-6.02	0.51	1.00	1087	1778
gender	0.41	0.36	1.01	621	1191
race	-0.24	0.32	1.00	817	911
age	0.22	0.03	1.00	2722	2608
mhv1	0.06	0.18	1.00	1849	1439
mhv2	0.90	0.25	1.00	1284	1568
mhv3	1.45	0.25	1.00	1239	1827
inptmhv	0.77	0.29	1.00	3182	2988
sd(Intercept)	2.50	0.17	1.00	1046	2185

The 95% CI of gender,race and mhv1 contains 0, hence they are not significant at 0.05 significance level. This is consistent with t-distribution prior.

Model 2: Diagnosis



Model 0 (Without Random Intercept)

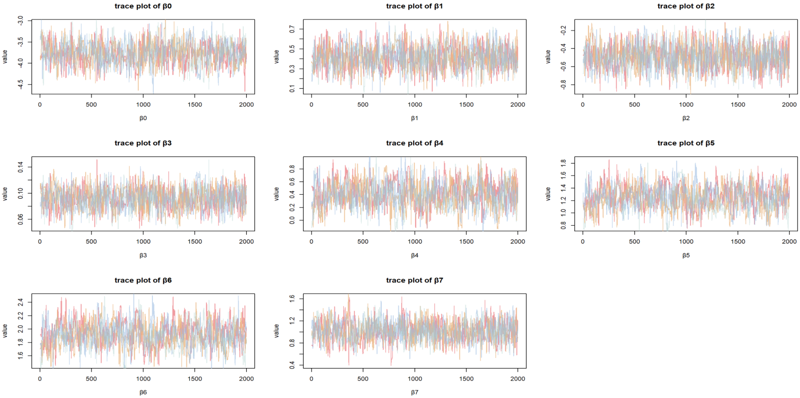
After discarding the random intercept, the model is

$$\begin{aligned} logit(\mu_{ij}) =& \beta_0 + \beta_1 Gender_i + \beta_2 Race_i + \beta_3 Age_{ij} + \beta_4 MHV_{1ij} \\ &+ \beta_5 MHV_{2ij} + \beta_6 MHV_{3ij} + \beta_7 INPTMHV_{ij}, \end{aligned} \tag{4}$$

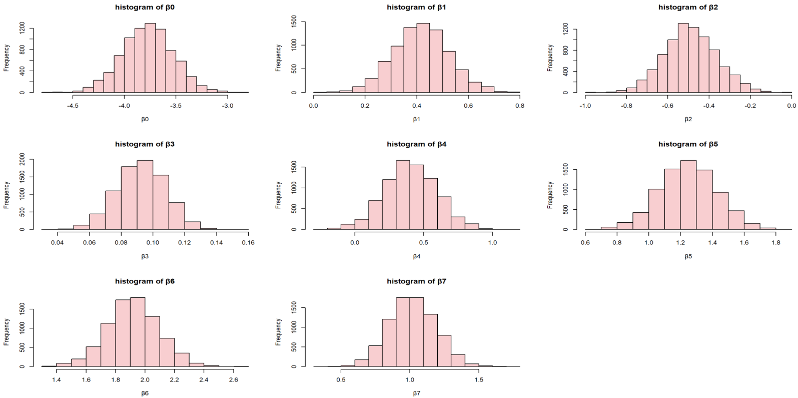
with prior being $\pi(\beta_j) \sim \mathcal{N}(0, \sigma_\beta^2)$.

Model 0: Trace Plot

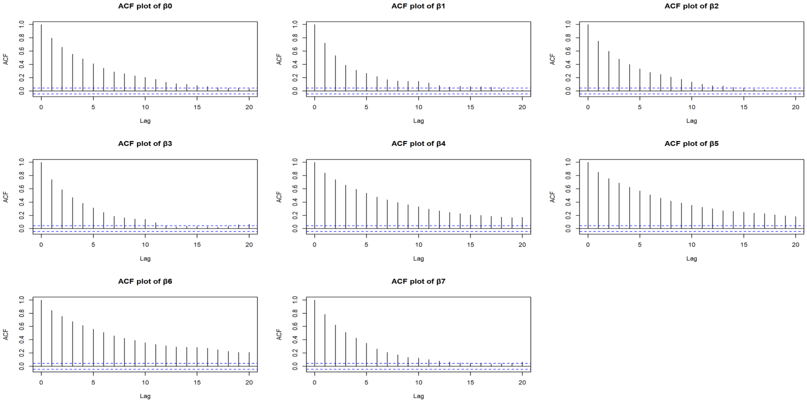
Note: 2,000 iterations (with 1,000 burn-in) cannot guarantee convergence, thus we perform 4,000 iterations (with 2,000 burn-in) here.



Model 0: Histogram



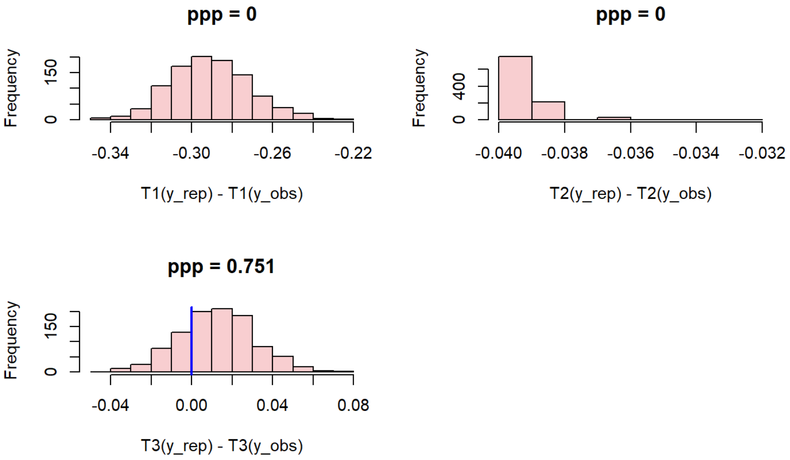
Model 0: ACF Plot



Model 0: Results

	Estimate	Error	$\hat{R}$	Bulk-ESS	Tail- ESS
Intercept	-3.77	0.25	1.01	563	933
gender	0.42	0.11	1.00	697	1540
race	-0.49	0.13	1.01	602	1053
age	0.09	0.02	1.01	698	1095
mhv1	0.41	0.19	1.01	328	660
mhv2	1.25	0.18	1.02	298	538
mhv3	1.92	0.18	1.02	282	708
inptmhv	1.02	0.17	1.00	621	1315

Model 0: Diagnosis



PSIS-LOO

PSIS-LOO is short for "Pareto smoothed importance sampling leave-one-out cross validation".

$$lpd = \sum_{i=1}^{n} \log(p(y_i|y)) = \sum_{i=1}^{n} \log(\int p(y_i|\theta)p(\theta|y)d\theta) \qquad (5)$$

$$lpd_{LOO} = \sum_{i=1}^{n} \log(p(y_i|y_{-i})) = \sum_{i=1}^{n} \log(\int p(y_i|\theta)p(\theta|y_{-i})d\theta) \qquad (6)$$

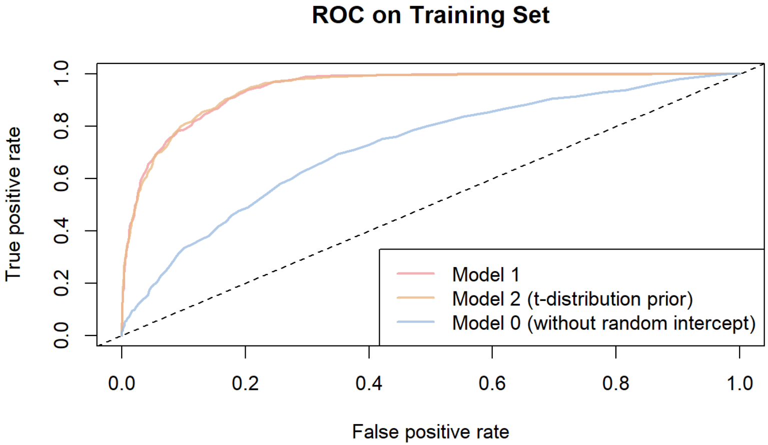
Model Comparison

Model 1 is the most recommended.

	Δelpd_{WAIC}	Δelpd_{LOO}
model 1	0.0	0.0
model 2 (t-distribution prior)	-2.9	-3.2
model 0 (without random intercept)	-509.4	-520.8

(Note: Δ denotes for "difference")

ROC Curve


ROC on Training Set
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
Model 1
Model 2 (t-distribution prior)
Model 0 (without random intercept)

Thanks!