

Applied Bayesian modeling - HW3

Score: The maximum number of points in this HW is 20 points, with 3 points extra credit. This HW counts for 1.5 HWs, so points are rescaled such that a “full pass” score is 150%. Specifically, for calculating your final HW score, the points will be rescaled to a maximum score of $(20+3)/20 \cdot 150\% = 172.5\%$.

Exercise 1: Fit a Bayesian model using brm and check and interpret the output (4 points)

We continue with IQ data, as introduced in HW2. For this HW, data set `iq_scores.csv` contains 10 iq-scores, sampled from a town in let's say, your favorite country.

Use brm to fit the following model to the IQ data:

$$\begin{aligned} y_i | \theta_i, \sigma^2 &\sim N(\theta_i, \sigma^2) (\text{independent}), \text{ for } i = 1, 2, \dots, n; \\ \mu &\sim N(100, 15^2); \\ \sigma &\sim \text{use the brm-default.} \end{aligned}$$

Then answer the following questions:

- (i) Plot a histogram of the posterior samples of μ and report a posterior point estimate and 80% CI for μ .
- (ii) Plot a histogram of the posterior samples of σ and report a posterior point estimate and 80% CI for σ .
- (iii) Can you report a posterior point estimate and 80% CI for μ/σ ? If yes, do so. If not, why not?

Note that the prior for μ can be specified with an additional argument in brm (as illustrated in optional material in module 5), as follows:

```
#mu_prior <- set_prior("normal(100,15)", class = "Intercept")
#fit <- brm(y ~ 1, prior = c(mu_prior),
#           your other usual arguments)
```

Exercise 2: Compare and contrast the MCMC diagnostics of two different model fits (4 points)

Continue with the IQ data from Q1 (with y = IQ scores) to fit the model as specified below. Present and briefly summarize resulting MCMC diagnostics for μ (traceplots, Rhat, effective sample sizes). Then and comment on whether this model fit can be used for summarizing information regarding μ . If not, why not?

```
#fit_bad <- brm(y ~ 1, data = dat,
#               chains = 4, iter = 200, cores = getOption("mc.cores", 4),
#               control = list(adapt_delta = 0.6, max_treedepth = 4)
#               # these are NOT recommended options, trying to create problems here!
#               )
```

Exercise 3: see separate pdf (8 points)

Exercise 4 (4 points)

Continue with the marriage data set from exercise 3.

The goal in this exercise is to predict the age at first marriage for a randomly sampled woman in an ethnic group for which we have not yet observed any data, using the model and data from exercise 3.

- (a) Obtain samples from the predictive posterior density for the age at first marriage and visualize the samples in a histogram. In your answer, include R code (that does NOT use the predict function from brms) as well as a write up in equations how you obtained the samples. Make sure to introduce notation first to explain what you're sampling.
- (b) Use the samples to construct a point prediction and 95% prediction interval for age at first marriage. In your answer, include the expression used for calculating the point prediction from the samples.
- (c) What is the probability that the observed age at first marriage will be greater than \bar{y} ? In your answer, include the expression used for calculating this probability from the samples.

Exercise 5 (extra credit, 3 points)

Obtain the joint distribution of $(y_1, y_2) | \mu_\alpha, \sigma_\alpha^2, \sigma_y^2$ in the setting where $j[1] \neq j[2]$ and in the setting where $j[1] = j[2]$.

Hints:

- First consider the univariate distribution of $y_i | \mu_\alpha, \sigma_\alpha^2, \sigma_y^2$.
- You may find it helpful to write y_i as the sum of parameters (thus random variables) μ_α , a parameter to capture the variability across counties, and a parameter to capture variability across observations.