Case Study I

Math 315, Fall 2019

Due 22 October by 4:00 p.m.

Your task

The data set `test_scores.csv` contains 29 test scores from the first exam in Math 215 (out of 100 possible points). Your task is to conduct a Bayesian analysis of these data in order to answer the following:

i. Estimate the average exam I score, μ.
ii. Estimate the standard deviation of the exam I scores, σ.
iii. Predict the exam I test score for a randomly selected future Math 215 student.

```r
scores <- read.csv("http://aloy.rbind.io/data/test_scores.csv")
```

Model requirements

- Use a *truncated normal distribution* for the likelihood, where $Y_i \sim \mathcal{N}(\mu, \sigma^2)$ with $Y_i \in (a, b)$, $-\infty < a < b < \infty$.
- Use a prior distribution where μ and σ^2 are assumed independent, $\mu \sim \mathcal{N}(\theta, \tau^2)$ for $\mu \in (a_0, b_0)$ and $\sigma^2 \sim InverseGamma(\gamma, \delta)$ for $\sigma^2 \in (a_1, b_1)$.
- Specify values for θ, τ, γ, and δ, along with the constraints, in your prior elicitation process.

Truncated normal details

The truncated normal density has the form

$$f(y|\mu, \sigma^2, a, b) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)}{\Phi(b, \mu, \sigma^2) - \Phi(a, \mu, \sigma^2)},$$

where $\Phi(x, \mu, \sigma^2)$ is the normal cumulative distribution function evaluated at x.

The expected value and variance of a truncated normal distribution can be written in terms of μ, σ^2, a, and b. In particular, we have

$$\mathrm{E}(Y) = \mu + \sigma \cdot \frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

$$\mathrm{Var}(Y) = \sigma^2 \left[1 + \frac{\frac{a-\mu}{\sigma} \cdot \phi\left(\frac{a-\mu}{\sigma}\right) - \frac{b-\mu}{\sigma} \cdot \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} - \left(\frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\right)^2\right]$$

In R, you can install and load the **truncnorm** package to obtain the `dtruncnorm()`, `ptruncnorm()`, `qtruncnorm()`, and `rtruncnorm()` functions. Additionally, the expected value and variance can be calculated using the `etruncnorm()` and `vtruncnorm()` functions, respectively.

Advice

- Constraints on the likelihood are not handled the same way as constraints on the prior distribution.
- The normalizing constants on the priors matter in this problem. Be careful with your bookkeeping.
- To restrict the range of the inverse gamma distribution, you will need to add an additional normalizing constant to ensure that the density integrates to 1 on the restricted range, (a_1, b_1)

Report requirements

Summarize your analysis in a 3-6 page report (double spaced, 11pt, one-inch margins, including figures). Your report MUST have the follow sections and contents:

1. Introduction: Briefly describe the problem, the data, and your objectives
2. Model specification: Describe the Bayesian model you propose and other relevant details, including a clear outline and justification of your prior specification. For this case study, I recommend plotting the marginal priors and joint prior distribution to help justify your specification.
3. Computation: Give the details of the algorithms you use, if any. This could be a brief overview of the steps required for the grid approximation and/or Monte Carlo sampling. If you derive the posterior in closed form, this may not be necessary.
4. Results: Analyze and discuss your posterior distribution. For this case study, I recommend plotting the marginal posteriors and joint posterior distribution to support your discussion.
5. Prediction: Forecast (with uncertainty measures) the exam I score of a future Math 215 student

Think of this case study as a "mini-project." This means that everything must be typed, proofread, labeled, captioned, and referenced, as appropriate. If you develop an interesting computational way to solve a problem, feel free to include pseudo (or actual) code as an appendix, but, in general, code is not required in the body of the report. Please email me your code for the case study with the subject line: "`Math 315 case study: <Last Name>, <Last Name>`," where you fill in the last names of your group as appropriate.