

# homework3

PSTAT 115, Spring 2021

Due on May 23, 2021 at 11:59 pm

---

## Problem 1. Rejection Sampling the Beta distribution. (15 pts)

Assume we did not have access to the `rbeta` function for sampling from a Beta, but we were able to evaluate the density, `dbeta`. This is a very common setting in Bayesian statistics, since we can always evaluate the (proportional) posterior density  $p(\theta | y) \propto p(y | \theta)p(\theta)$  but we don't have immediate access to a method for sampling from this distribution.

1. Let  $p(x)$  be a  $\text{Beta}(3, 9)$  density,  $q_1(x)$  a  $\text{Uniform}(0, 1)$  density, and  $q_2(x)$  a  $\text{Normal}(\mu = 0.25, \sigma = 0.15)$  density.
2. Use rejection sampling to sample from  $p(x)$  by proposing samples from  $q_1(x)$ . To do so, first find  $M_1 = \max_x p(x)/q_1(x)$  using the `optimize` function and set `lower=0`, `upper=1`, and `maximum = TRUE` (since we are maximizing not minimizing, the default).  $M$  will be the value in the `objective` argument returned by `optimize` (`maximum` tells us where the maximum occurs, but not what height it achieves). Propose 10000 samples and keep only the accepted samples.
3. Use rejection sampling to sample from  $p(x)$  by proposing samples from  $q_2(x)$ . To do this you need to find  $M_2 = \max_x p(x)/q_2(x)$  as above. Propose 10000 samples and keep only the accepted samples.
4. Plot the  $p(x)$ ,  $M_1q_1(x)$  and  $M_2q_2(x)$  all on the same plot and verify visually that the scaled proposal densities “envelope” the target,  $p(x)$ . Set the `xlimits` of the plot from 0 to 1. Use different color lines for the various densities so are clearly distinguishable. (5 pts)
5. Which rejection sampler had the higher rejection rate? Why does this make sense given the plot from the previous part? This means when proposing 10000 samples from each proposal, the Monte Carlo error of our approximation will be higher when proposing from \_\_\_\_\_ (choose  $q_1$  or  $q_2$ ). (5 pts)
6. Report the variance of  $\text{Beta}(3, 9)$  distribution by computing the variance of the beta samples. How does this compare to the theoretical variance (refer to the probability cheatsheet). (5 pts)

## Problem 2. Frequentist Coverage of The Bayesian Posterior Interval. (35 pts)

Suppose that  $y_1, \dots, y_n$  is an IID sample from a  $\text{Normal}(\mu, 1)$ . We wish to estimate  $\mu$ .

**2a.** For Bayesian inference, we will assume the prior distribution  $\mu \sim \text{Normal}(0, \frac{1}{\kappa_0})$  for all parts below. Remember, from lecture that we can interpret  $\kappa_0$  as the pseudo-number of prior observations with sample mean  $\mu_0 = 0$ . State the posterior distribution of  $\mu$  given  $y_1, \dots, y_n$ . Report the lower and upper bounds of the 95% quantile-based posterior credible interval for  $\mu$ , using the fact that for a normal distribution with standard deviation  $\sigma$ , approximately 95% of the mass is between  $\pm 1.96\sigma$ . (5 pts)

Type your answer here, replacing this text.

**2b.** Plot the length of the posterior credible interval as a function of  $\kappa_0$ , for  $\kappa_0 = 1, 2, \dots, 25$  assuming  $n = 10$ . Report how this prior parameter effects the length of the posterior interval and why this makes intuitive sense. (10 pts)

```
# YOUR CODE HERE
```

**2c.** Now we will evaluate the *frequentist coverage* of the posterior credible interval on simulated data. Generate 1000 data sets where the true value of  $\mu = 0$  and  $n = 10$ . For each dataset, compute the posterior 95% interval endpoints (from the previous part) and see if the interval covers the true value of  $\mu = 0$ . Compute the frequentist coverage as the fraction of these 1000 posterior 95% credible intervals that contain  $\mu = 0$ . Do this for each value of  $\kappa_0 = 1, 2, \dots, 25$ . Plot the coverage as a function of  $\kappa_0$ . (5 pts)

```
# YOUR CODE HERE
```

**2d.** Repeat the 1c but now generate data assuming the true  $\mu = 1$ . (5 pts)

```
# YOUR CODE HERE
```

**2e.** Explain the differences between the coverage plots when the true  $\mu = 0$  and the true  $\mu = 1$ . For what values of  $\kappa_0$  do you see closer to nominal coverage (i.e. 95%)? For what values does your posterior interval tend to overcover (the interval covers the true value more than 95% of the time)? Undercover (the interval covers the true value less than 95% of the time)? Why does this make sense? (10 pts)

Type your answer here, replacing this text.

### Problem 3. Bayesian inference for the normal distribution in Stan. (50pts)

Create a new Stan file by selecting “Stan file” in the Rstudio menu. Save it as `IQ_model.stan`. We will make some basic modifications to the template example in the default Stan file for this problem. Consider the IQ example used from class. Scoring on IQ tests is designed to yield a  $N(100, 15)$  distribution for the general population. We observe IQ scores for a sample of  $n$  individuals from a particular town,  $y_1, \dots, y_n \sim N(\mu, \sigma^2)$ . Our goal is to estimate the population mean in the town. Assume the  $p(\mu, \sigma) = p(\mu | \sigma)p(\sigma)$ , where  $p(\mu | \sigma)$  is  $N(\mu_0, \sigma/\sqrt{\kappa_0})$  and  $p(\sigma)$  is  $\text{Gamma}(a, b)$ . Before you administer the IQ test you believe the town is no different than the rest of the population, so you assume a prior mean for  $\mu$  of  $\mu_0 = 100$ , but you aren’t to sure about this a priori and so you set  $\kappa_0 = 1$  (the effective number of pseudo-observations). Similarly, a priori you assume  $\sigma$  has a mean of 15 (to match the intended standard deviation of the IQ test) and so you decide on setting  $a = 15$  and  $b = 1$  (remember, the mean of a Gamma is  $a/b$ ). Assume the following IQ scores are observed:

```
y <- c(70, 85, 111, 111, 115, 120, 123)
n <- length(y)
```

**3a.** Make a scatter plot of the posterior distribution of the median,  $\mu$ , and the precision,  $1/\sigma^2$ . Put  $\mu$  on the x-axis and  $1/\sigma^2$  on the y-axis. What is the posterior relationship between  $\mu$  and  $1/\sigma^2$ ? Why does this make sense? *Hint:* review the lecture notes. (10pts)

Type your answer here, replacing this text.

**3b.** You are interested in whether the mean IQ in the town is greater than the mean IQ in the overall population. Use Stan to find the posterior probability that  $\mu$  is greater than 100. (20pts)

```
library(rstan)
y <- c(70, 85, 111, 111, 115, 120, 123)
n <- length(y)
```

```
# YOUR CODE HERE
```

**3c.** You notice that two of the seven scores are significantly lower than the other five. You think that the normal distribution may not be the most appropriate model, in particular because you believe some people in this town are likely have extreme low and extreme high scores. One solution to this is to use a model that is more robust to these kinds of outliers. The [Student's t distribution](#) and the [Laplace distribution](#) are two so called “heavy-tailed distribution” which have higher probabilities of outliers (i.e. observations further from the mean). Heavy-tailed distributions are useful in modeling because they are more robust to outliers. Fit the model assuming now that the IQ scores in the town have a Laplace distribution, that is  $y_1, \dots, y_n \sim \text{Laplace}(\mu, \sigma)$ . Create a copy of the previous stan file, and name it “IQ\_laplace\_model.stan”. *Hint:* In the Stan file you can replace `normal` with `double_exponential` in the model section, another name for the Laplace distribution. Like the normal distribution it has two arguments,  $\mu$  and  $\sigma$ . Keep the same prior distribution,  $p(\mu, \sigma)$  as used in the normal model. Under the Laplace model, what is the posterior probability that the median IQ in the town is greater than 100? How does this compare to the probability under the normal model? Why does this make sense? (20pts)

```
# YOUR CODE HERE
```