STA 360 Lab 4: Exponential Data

Isaac Fan

19 February, 2021

Preliminaries

Please turn in a pdf version of this Rmd document on **Friday, February 19th** by 11:59 PM. Exercises 2 and 3 should be turned in and will be graded *for completion*. For derivations, you can attach a scan or picture of your work on paper, though we encourage you to type it directly into R using Latex.

Introduction

By now you have worked extensively with exponential family distributions whose parameters have natural prior distributions called *conjugate priors*. This lab will focus on a special continuous distribution: the $Exp(\lambda)$ distribution.

Likelihood and Prior

If $Y|\lambda \sim Exp(\lambda)$, then Y has density

$$\lambda e^{-\lambda y}\mathbf{1}(y > 0)$$

where $\lambda > 0$. Some comments:

- Note that the distribution of Y has support $\mathbb{R}^+$. So, if you have data that is not constrained to be only positive, fitting an exponential model will not make any sense.
- $E[Y] = 1/\lambda$ and $Var[Y] = 1/\lambda^2$. This means that there is a quadratic relationship between the mean and the variance.

Now, assume we have a sample of size n, such that

$$Y_i|\lambda \overset{iid}{\sim} Exp(\lambda).$$

Exercise 1: Conjugate Prior & Posterior Update

We will state the conjugate prior and derive the posterior during lab. Before we start, a few questions:

- What do we know about λ?

Lambda > 0

- How can we use this knowledge to think of possible prior distributions for λ?

Eliminate distributions that allow for negative lambdas

Exercise 2: The More General Case (For Completion)

Suppose instead that

$$Y_i|\alpha, \beta \overset{iid}{\sim} Gamma(\alpha, \beta)$$

where $\alpha > 0$ is known and $\beta > 0$.

1. What is the conjugate prior for β?

$$\beta \overset{iid}{\sim} Gamma(\alpha_0, \beta_0)$$

2. Under this conjugate prior, what is the posterior distribution for β?

$$\beta|Y_i = y_i \overset{iid}{\sim} Gamma(\alpha_0 + n\alpha, \beta_0 + \Sigma y_i)$$

Simulated Data Example

Next, we will simulate some data of size $n = 250$ and attempt to fit an exponential model with a conjugate prior assigned to λ. Observe that the actual data generating mechanism is $Y_i \sim Gamma(3, 5)$.

```
set.seed(360)
n <- 250
alpha <- 3
lambda <- 5
y <- rgamma(n, alpha, lambda)
```

Now, suppose we fit a model assuming that

$$Y_i|\lambda \overset{iid}{\sim} Exp(\lambda).$$

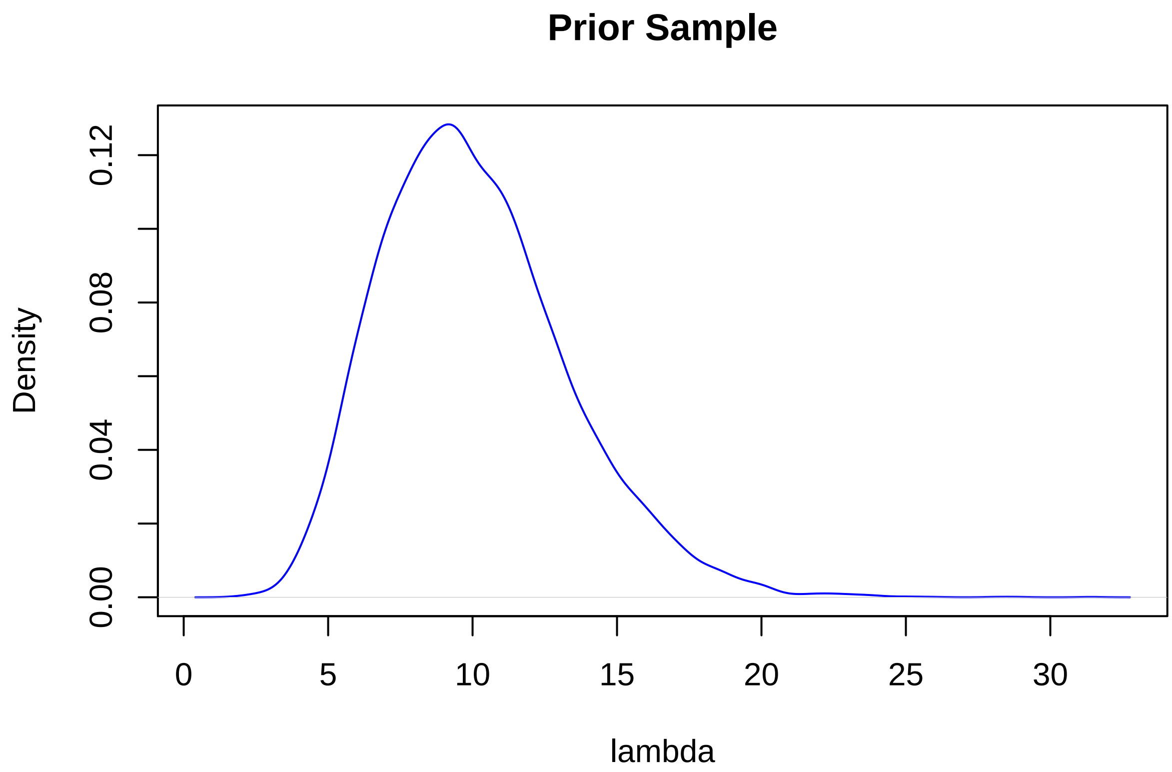
We know that the conjugate prior for λ is a Gamma distribution. Suppose we know from prior experiments that λ is most likely somewhere in the range $[1, 20]$. We might give λ the weakly informative prior

$$\lambda \sim Gamma(10, 1).$$

Sample from the Prior

Below, I sample from the prior and make a density plot.

```
a_lambda = 10
b_lambda = 1
M <- 10000
prior_samp <- rgamma(M, a_lambda, b_lambda)
plot(density(prior_samp), col = "blue",
     main = "Prior Sample", xlab = "lambda")
```



Sampling from the Posterior

By now, we know that the Posterior under this model will be

$$\lambda|Y \sim Gamma\left(\alpha_\lambda + n, \beta_\lambda + \sum_{i=1}^{n} y_i\right).$$

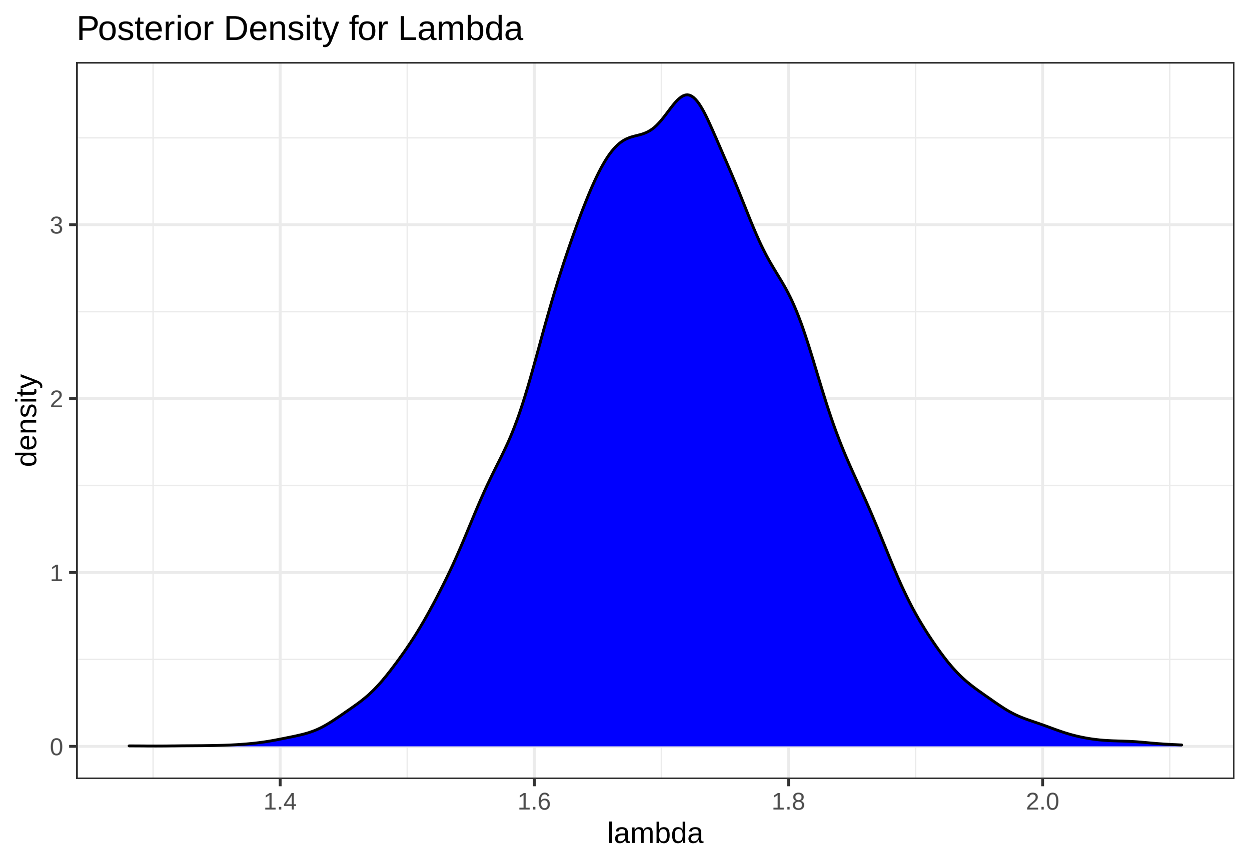
Let's draw a sample of size $R = 10000$ from this posterior distribution.

```
R <- 10000
post_samp <- rgamma(R, a_lambda + n, b_lambda + sum(y))

# making into df for ggplot
post.df <- data.frame(lambda = post_samp)
```

Below is the density of these samples. What looks odd about it?

The lambda of five seems to not really be contained in the posterior.

```
ggplot(post.df, aes(x = lambda)) + geom_density(fill = "blue") +
  labs(title = "Posterior Density for Lambda")
```



Constructing a Credible Interval for λ

The α level credible interval (λ_u, λ_l), often referred to as the α level quantile-based interval, satisfies

$$P(\lambda < \lambda_l | y) = \alpha/2$$

and

$$P(\lambda > \lambda_u | y) = \alpha/2.$$

Since we are using a conjugate prior, we can compute a 95% credible interval using gamma quantiles:

```
lambda_l <- qgamma(p = 0.025, a_lambda + n, b_lambda + sum(y))
lambda_u <- qgamma(p = 0.975, a_lambda + n, b_lambda + sum(y))
c(lambda_l, lambda_u)
```

```
## [1] 1.507759 1.923137
```

We can also estimate this interval using the sample's quantiles (we'll come back to this when we study MCMC).

```
lambda_l <- quantile(post_samp, 0.025) # lower bound
lambda_u <- quantile(post_samp, 0.975) # upper bound
c(lambda_l, lambda_u)
```

```
##     2.5%    97.5%
## 1.503622 1.921419
```

The interpretation of the credible interval differs from the frequentist notion of a confidence interval. How should we interpret this result?

After observing data, we believe that there is a 95% probability that the parameter lamda is in the interval 1.503622 and 1.921419.

Posterior Predictive Checks

The posterior predictive distribution is

$$f(y_{new}|y) = \int f(y_{new}|\lambda)p(\lambda|y)d\lambda.$$

We can sample from this using the following scheme:

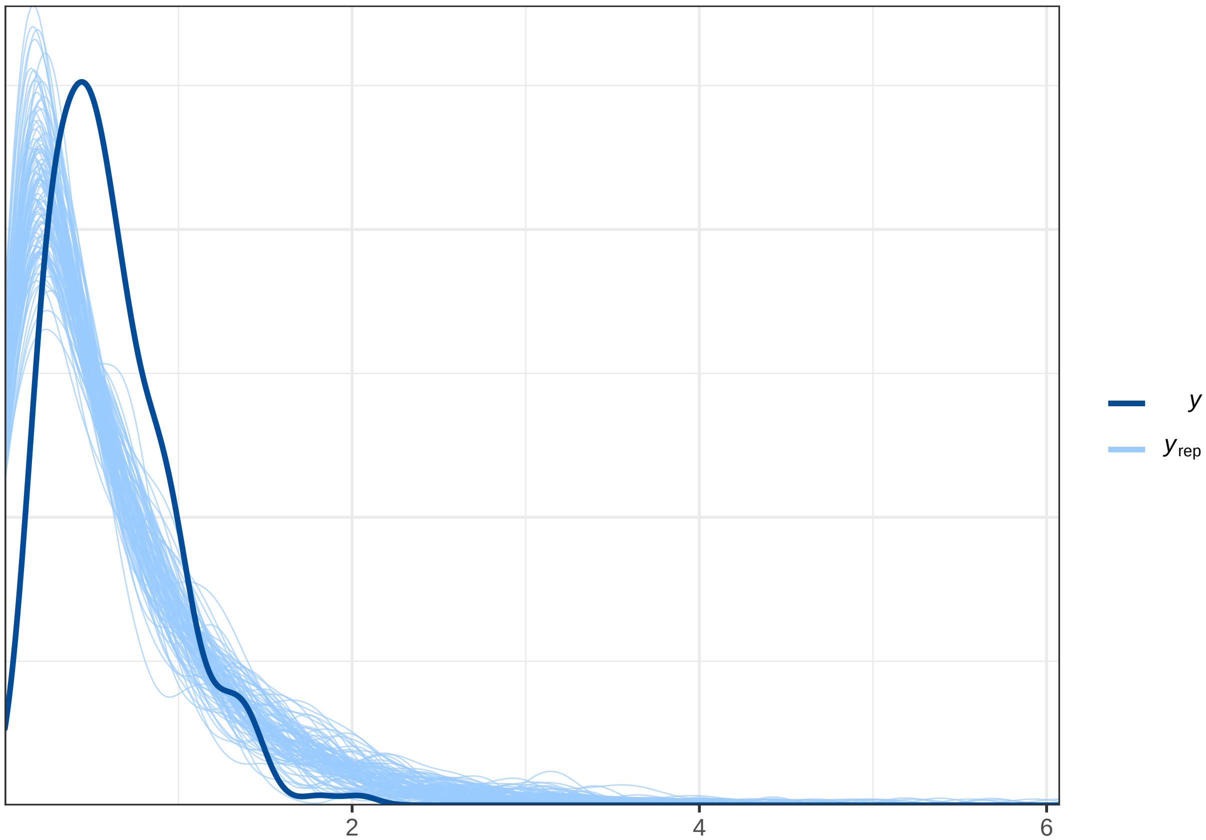
1. Sample $\lambda^* \sim p(\lambda|y)$.
2. Sample a vector of new observations from the likelihood, $y_{new} \sim f(\cdot|\lambda^*)$.

Because we have a posterior sample of size $10,000$, we could construct $10,000$ such samples. For simplicity, we'll just use 100 posterior predictive samples.

```r
# posterior predictive sampling
K <- 100
PP <- matrix(NA, nrow = K, ncol = n)
for (j in 1:K) {
  PP[j, ] <- rexp(n, post_samp[j])
}
```

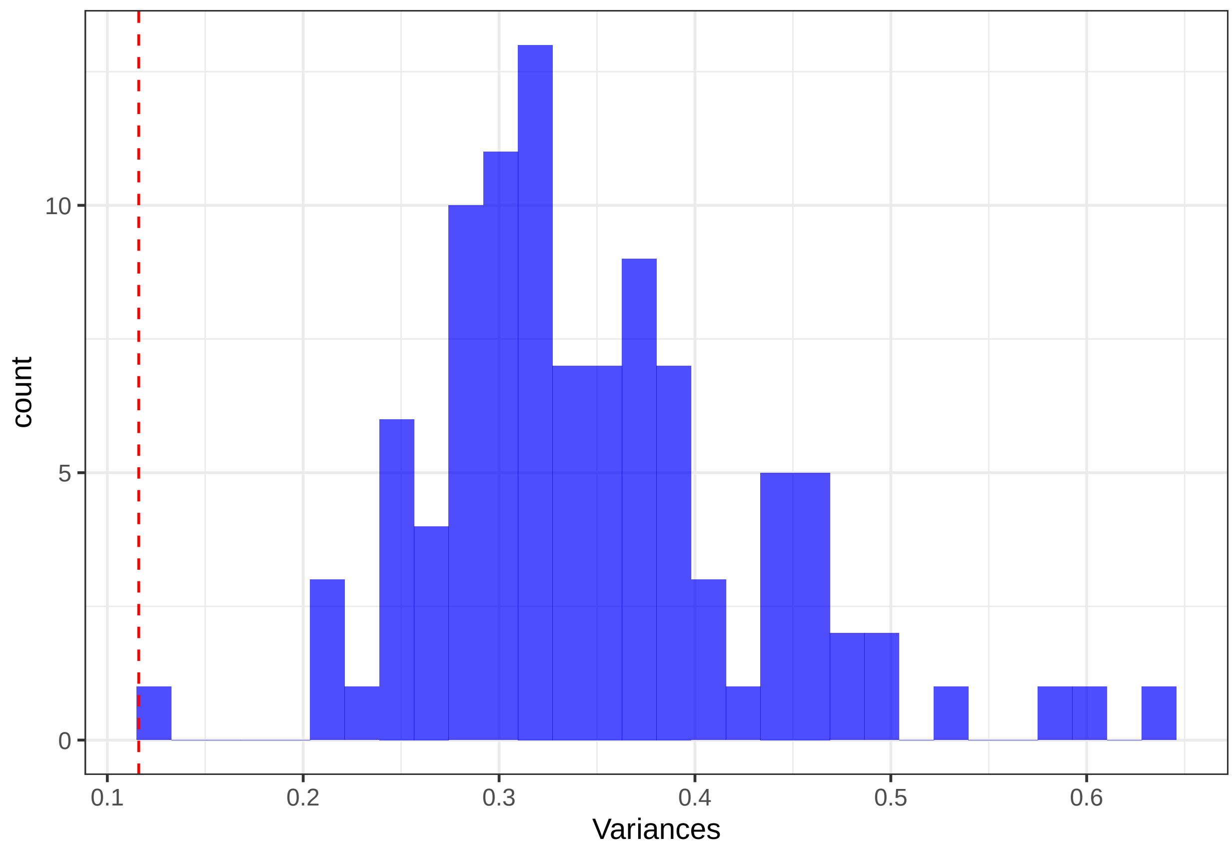
Posterior predictive checks evaluate how close our predictive samples are to the actual data. For instance, we could compare histograms:

```r
color_scheme_set("brightblue")
ppc_dens_overlay(y, PP)
```



It would seem that our posterior predictive samples do not resemble the data. Also, we could compare some sample statistics. For instance, here is a histogram of the sample variances of the data and posterior predictive samples, with a red-dotted line indicating the variance of the data.

```
v <- apply(PP, 1, var)
vars <- data.frame(Variances = c(v, var(y)))
ggplot(vars, aes(x = Variances)) + geom_histogram(alpha = 0.7, fill = "blue") +
  geom_vline(xintercept = var(y), col = "red", linetype = "dashed")
```



Clearly, our predictive model does not capture the variability within the actual data. That is where you come in.

Exercise 3 (For Completion)

Instead, let us now fit the model assuming that

$$Y_i|\lambda \overset{iid}{\sim} Gamma(3, \lambda).$$

Use the same prior that we used in the above analysis:

$$\lambda \sim Gamma(10, 1).$$

1. Using your answer to exercise 2, sample $10,000$ draws from the posterior distribution of λ.

```
post_samp_lambda <- rgamma(R, 3*n + 10, b_lambda + sum(y))
```

2. Provide a credible interval for λ and give the interpretation.

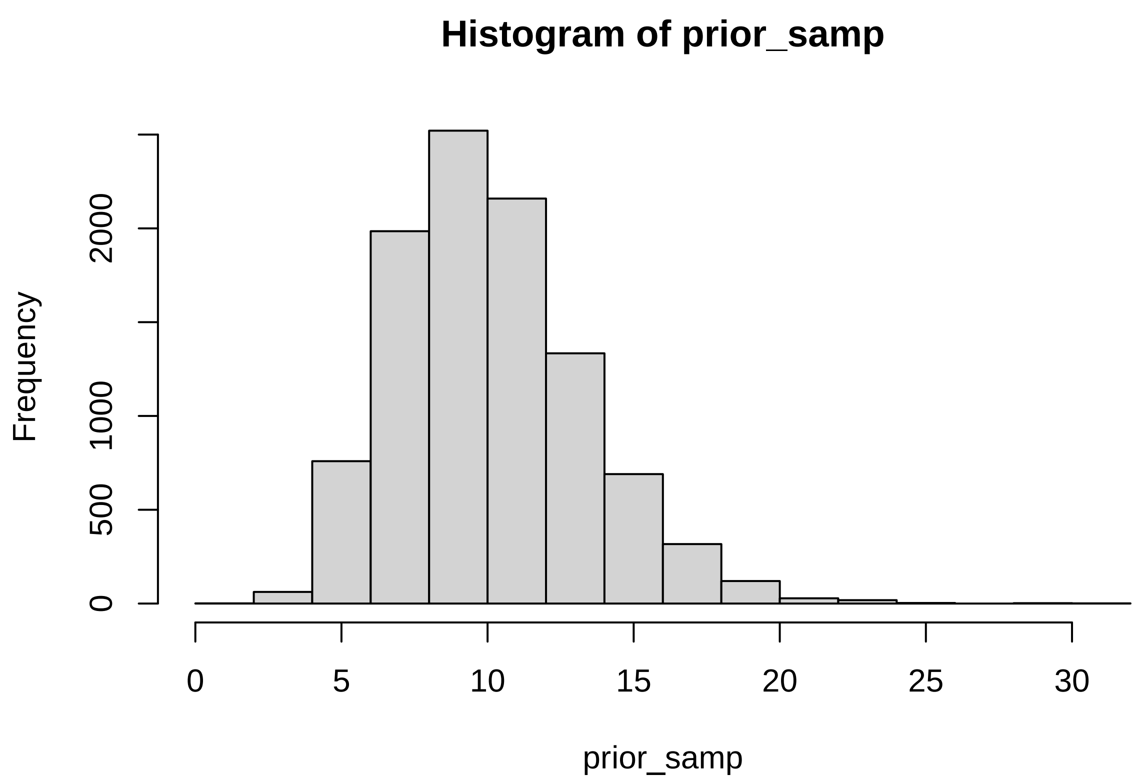
```
lambda_l_gamma <- qgamma(p = 0.025, 3*n + 10, b_lambda + sum(y))
lambda_u_gamma <- qgamma(p = 0.975, 3*n + 10, b_lambda + sum(y))
c(lambda_l_gamma, lambda_u_gamma)
```

```
## [1] 4.647251 5.357581
```

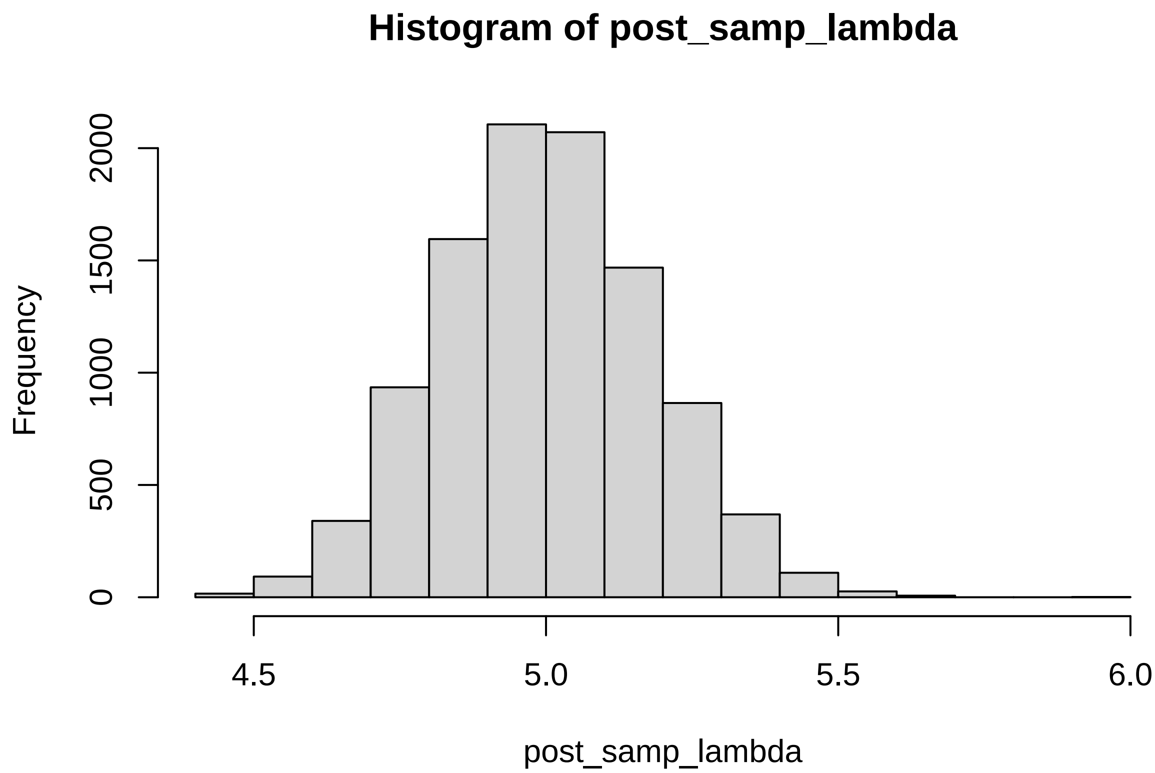
After observing data, we believe that there is a 95% probability that the parameter lamda is in the interval 4.65 and 5.36.

3. Plot a histogram of the prior sample, then a histogram of the posterior sample. What do you see?

```r
hist(x = prior_samp)
```




```r
hist(x = post_samp_lambda)
```

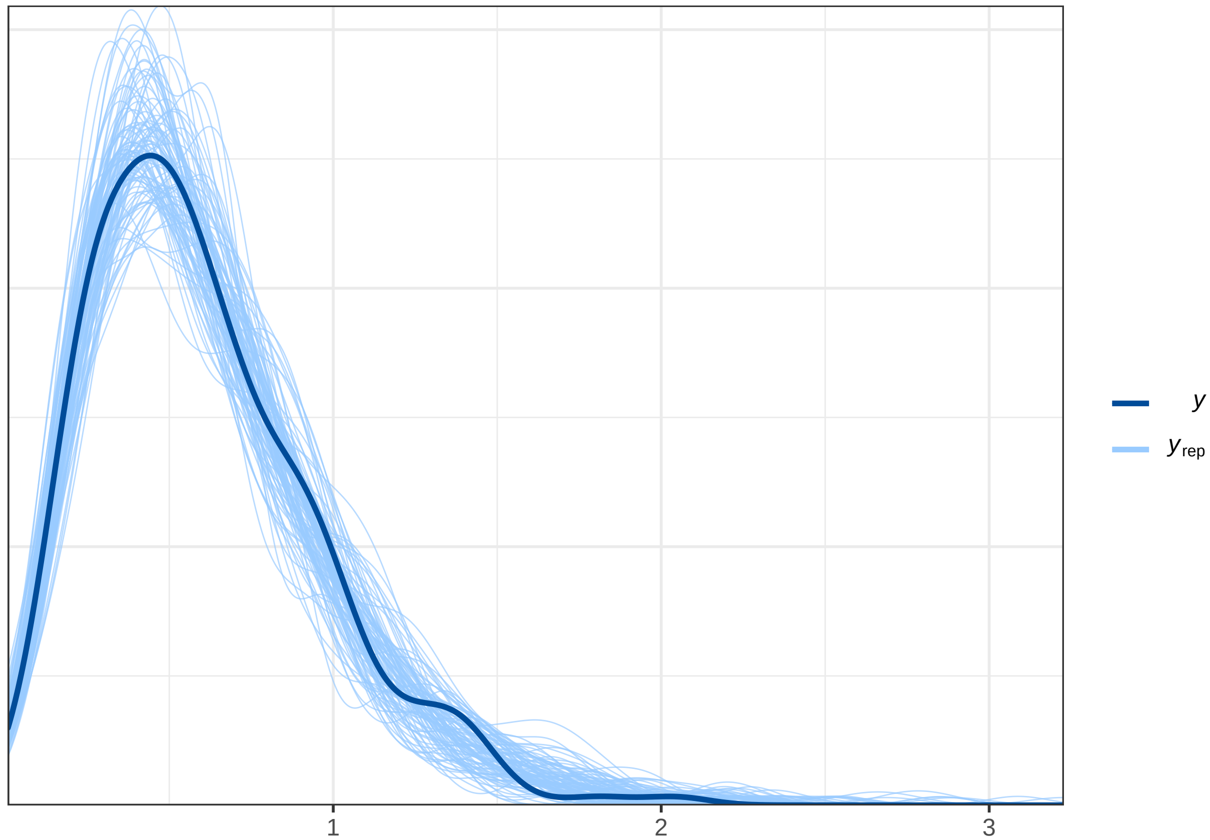


The posterior is much less skewed than the prior and the posterior appears the be centered around 5.

4. Generate 100 samples from the posterior predictive distribution and compare these (visually) to the actual data.

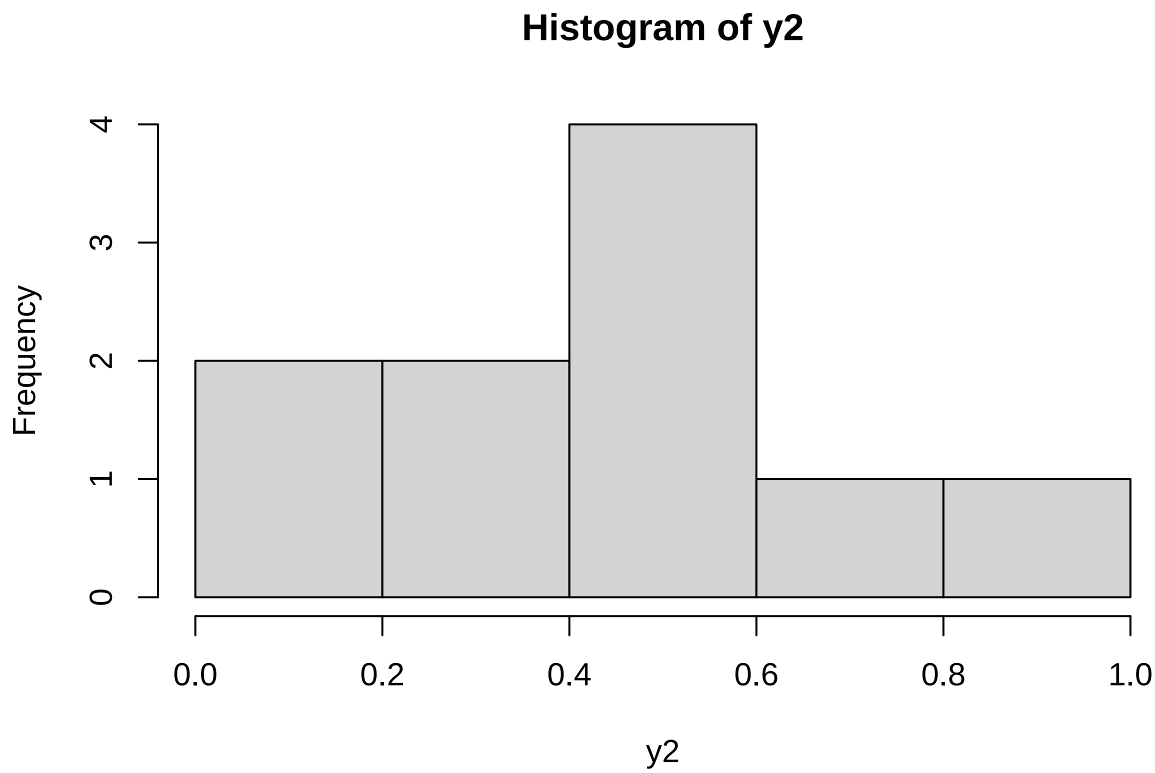
```
# posterior predictive sampling

PP2 <- matrix(NA, nrow = K, ncol = 250)
for (j in 1:K) {
  PP2[j, ] <- rgamma(n,3, post_samp_lambda[j])
}
```

```
color_scheme_set("brightblue")
ppc_dens_overlay(y, PP2)
```

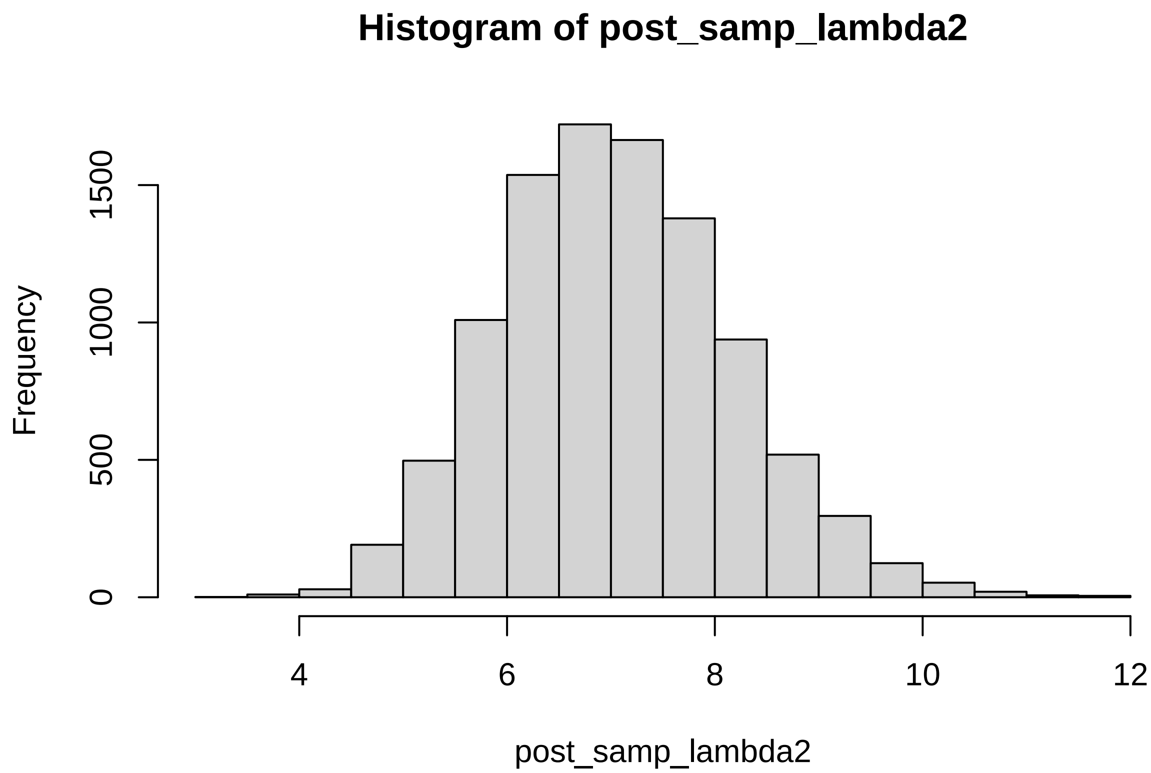


5. Set $n = 10$, sample a new vector of data y of size n from the Gamma distribution (ie, using `rgamma(n = 10, 3, 5)`), and simulate a sample of $10,000$ draws from the posterior distribution for λ. Compare this histogram to your posterior histogram from question 3. What differences do you see? Why do you see these differences?

```
y2 <- rgamma(10, 3, 5)
post_samp_lambda2 <- rgamma(R, 3*10 + 10, b_lambda + sum(y2))
hist(x = y2)
```




```r
hist(post_samp_lambda2)
```



It is centered around a higher lambda, this is probably because with a smaller sample size comes greater variance, and the expected value of the gamma distribution increases (since E(x) = a/b) and b involves summing the y's.