# Gibbs Sampling Activity

Instructions

1. Download this `.Rmd` file and open it in RStudio.

2. Update the author name and date.

3. Answer the questions below, using LaTeX within R Markdown to type up your answers.

4. When you finish, `Knit to PDF`.

Markov Chains

Fill in the code to construct an MCMC approximation of the Gamma-Poisson posterior corresponding to

$$Y_i \mid \lambda \overset{\text{ind}}{\sim} Pois(\lambda)$$

$$\lambda \sim Gamma(3, 1)$$

This plot represents our prior beliefs
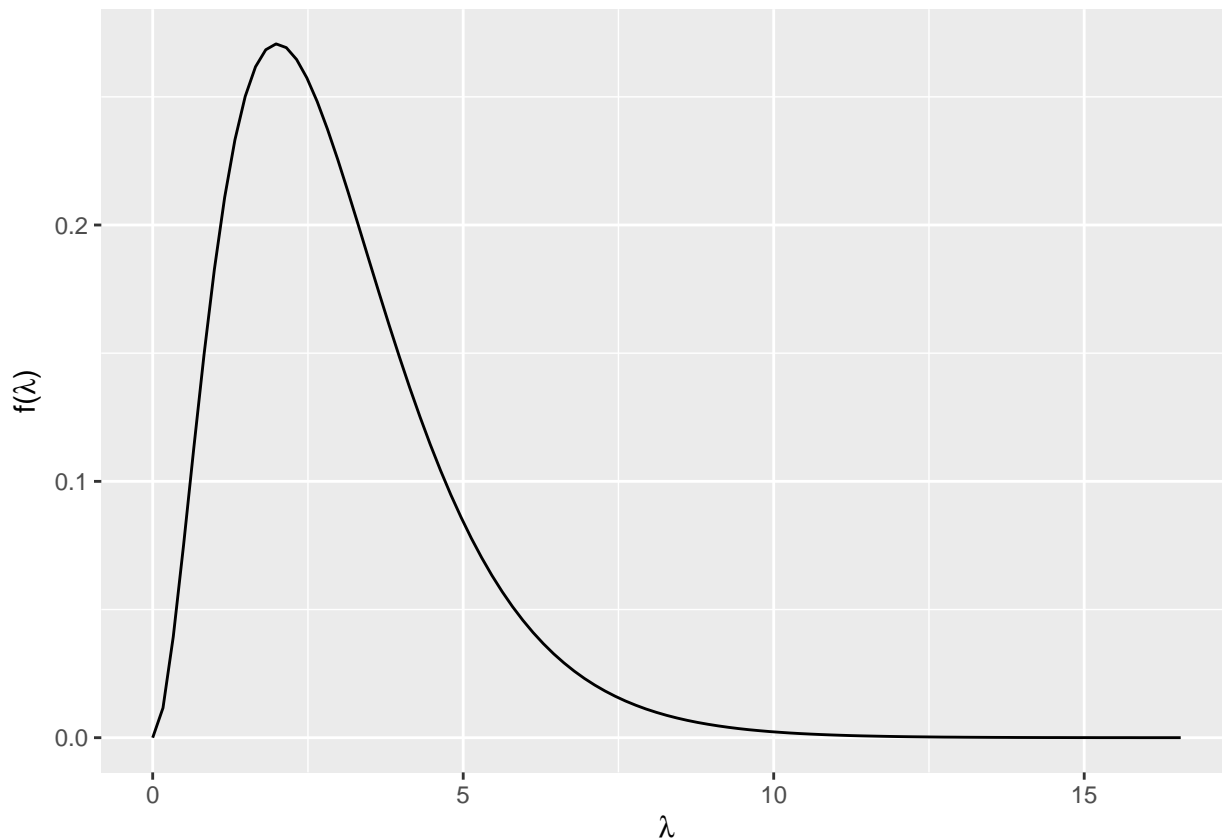
upon observing data $(Y_1, Y_2) = (2, 8)$. The code will require two terms you haven't seen, but you might guess how to use: poisson() and gamma(). Also, you code will incorporate a vector of (Y_1, Y_2) variables and observations as opposed to a single variable Y.

This plot shows our likelihood based on our two data points

```
plot_poisson_likelihood(5)
```
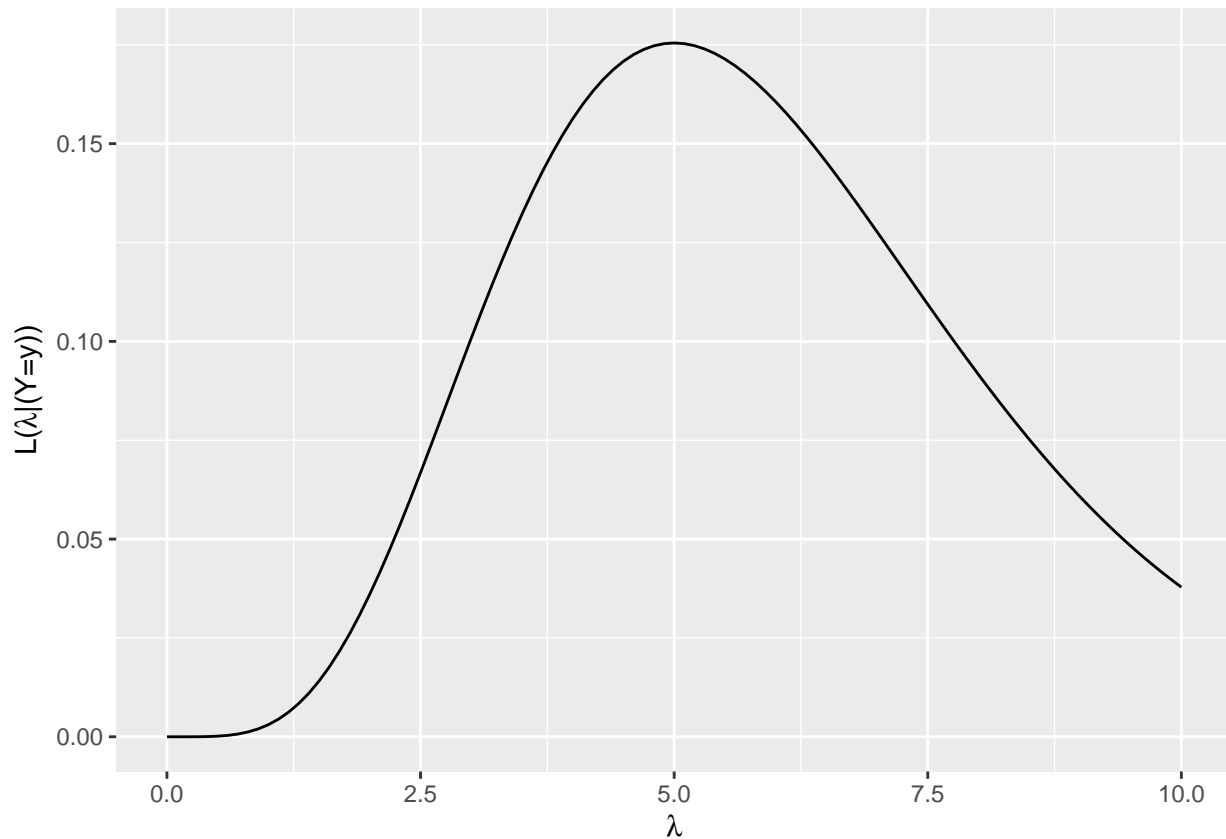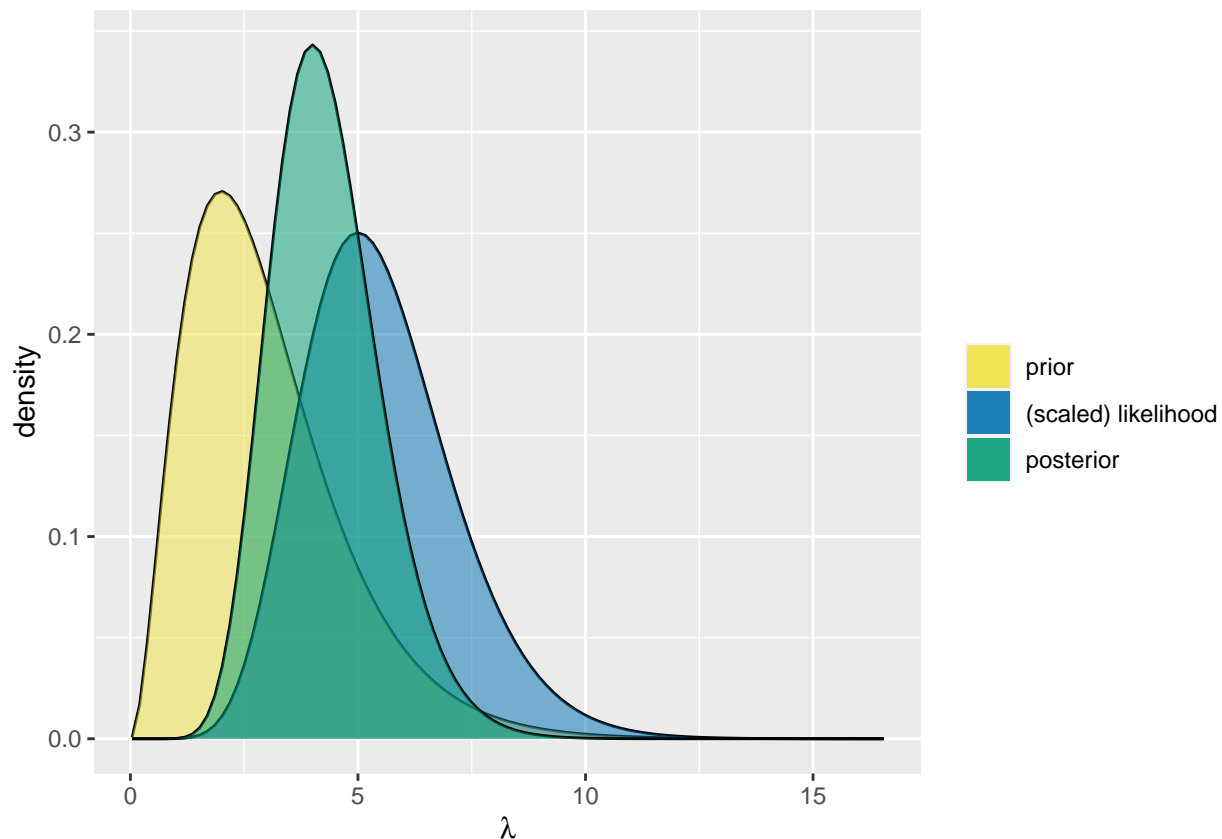
. Run four parallel chains for 10,000 iterations each (resulting in a sample size of 5000 per each chain)

```r
# STEP 1: DEFINE the model
gp_model <- "
  data {
    int<___> Y[2];
  }
  parameters {
    real<___> lambda;
  }
  model {
    Y ~ ___(___);
    lambda ~ ___(___, ___);
  }
"

# STEP 2: SIMULATE the posterior
#gp_sim <- ___(model_code = ___, data = list(___),
#              chains = 4, iter = 10000, seed = 84735)
```

This plot updates our prior based on our data to form our posterior

```r
plot_gamma_poisson(3,1,10,2)
```

Gibbs Sampling

Suppose we have data from a normal distribution where both the mean **and** variance are unknown. For convenience, we'll parameterize this model in terms of the *precision* $\gamma = \frac{1}{\sigma^2}$ instead of the variance σ^2.

$$Y \mid \mu, \gamma \sim N\left(\mu, \frac{1}{\gamma}\right)$$

Suppose we put the following *independent* priors on the mean μ and precision γ:

$$\mu \sim N(m, v)$$

$$\gamma \sim \mathrm{Gamma}(a, b)$$

1. Write down the joint posterior distribution for μ, γ. Does this look like a recognizable probability distribution?

ANSWER:

$$g(\mu, \gamma \mid y) \propto f(y \mid \mu, \gamma) f(\mu, \gamma)$$
$$= \ldots$$
$$= \gamma^{\frac{1}{2}+a-1} e^{-\frac{1}{2}\gamma(y-\mu)^2 + -\frac{1}{2v}(\mu-m)^2 - b\gamma}$$
$$= \gamma^{\frac{1}{2}+a-1} e^{-\frac{1}{2}\left[\gamma(y-\mu)^2 + \frac{1}{v}(\mu-m)^2 + 2b\gamma\right]}$$
$$= \gamma^{\frac{1}{2}+a-1} e^{-\frac{1}{2}\left[\gamma y^2 - 2\mu y\gamma + \gamma\mu^2 + \mu^2/v - 2m\mu/v + m^2/v + 2b\gamma\right]}$$
$$= \gamma^{\frac{1}{2}+a-1} e^{-\frac{1}{2}\left[\gamma(y^2+2b) - 2\mu(y\gamma + m/v) + \mu^2(\gamma+1/v) + m^2/v\right]}$$
$$\propto \gamma^{\frac{1}{2}+a-1} e^{-\frac{1}{2}\left[\gamma(y^2+2b) - 2\mu(y\gamma + \frac{m}{v}) + \mu^2(\frac{1}{v}+\gamma)\right]}$$

You should have answered "no" to Question 1, meaning that we can't use our usual techniques here to find Bayes estimators for μ or γ since we don't have a recognizable posterior distribution. Instead, we'll use a computational technique known as *Gibbs Sampling* to generate samples from this posterior distribution. Gibbs Sampling is particularly useful when we have more than one parameter, and the basic idea involves reducing our problem to a series of calculations involving one parameter at a time. In order to perform Gibbs Sampling, we need to find the conditional distributions

$$g(\mu \mid y, \gamma) \propto f(y \mid \mu, \gamma) f(\mu)$$

$$g(\gamma \mid y, \mu) \propto f(y \mid \mu, \gamma) f(\gamma)$$

We will use these conditional distributions to sample from the joint posterior $g(\mu, \gamma \mid y)$ according to the following algorithm:

(1) Start with initial values $\mu^{(0)}, \gamma^{(0)}$.

(2) Sample $\mu^{(t+1)} \sim g(\mu \mid y, \gamma = \gamma^{(t)})$.

(3) Sample $\gamma^{(t+1)} \sim g(\gamma \mid y, \mu = \mu^{(t+1)})$.

(4) Repeat many times.

It turns out that the resulting $\mu^{(0)}, \mu^{(1)}, \ldots, \mu^{(N)}$ and $\gamma^{(0)}, \gamma^{(1)}, \ldots, \gamma^{(N)}$ are samples from the joint posterior distribution $g(\mu, \gamma \mid Y)$, and we can use these sampled values to estimate quantities such as the posterior mean of each parameter $\hat{E}(\mu \mid y) = \frac{1}{N}\sum_{i=1}^{N}\mu^{(i)}$, $\hat{E}(\gamma \mid y) = \frac{1}{N}\sum_{i=1}^{N}\gamma^{(i)}$. Note that in practice we typically remove the initial iterations, known as the "burn-in" period: e.g., $\hat{E}(\mu \mid y) = \frac{1}{N-B}\sum_{i=B}^{N}\mu^{(i)}$.

2. Show that the conditional distributions $g(\mu \mid y, \gamma), g(\gamma \mid y, \mu)$ are proportional to $f(y \mid \mu, \gamma)f(\mu), f(y \mid \mu, \gamma)f(\gamma)$, respectively, as stated above.

ANSWER:
$$g(\mu \mid y, \gamma) = \frac{f(\mu, y, \gamma)}{f(y, \gamma)}$$
$$\propto f(\mu, y, \gamma), \text{ since } f(y, \gamma) \text{ doesn't depend on } \mu$$
$$= \ldots$$

3. Use this result to show that $\mu \mid y, \gamma \sim N\left(\frac{y\gamma + \frac{m}{v}}{\gamma + \frac{1}{v}}, \left[\gamma + \frac{1}{v}\right]^{-1}\right)$ and $\gamma \mid y, \mu \sim \text{Gamma}\left(\frac{1}{2} + a, \frac{1}{2}(y-\mu)^2 + b\right)$.

ANSWER:

$$g(\mu \mid y, \gamma) \propto f(y \mid \mu, \gamma) f(\mu)$$
$$= \left[(2\pi)^{-\frac{1}{2}} \gamma^{\frac{1}{2}} e^{-\frac{1}{2}\gamma(y-\mu)^2} \right] \left[(2\pi v)^{-\frac{1}{2}} e^{-\frac{1}{2v}(\mu-m)^2} \right]$$
$$= \dots$$
$$= \propto e^{-\frac{1}{2\left(\gamma+\frac{1}{v}\right)^{-1}} \left[\mu - \left(\frac{y\gamma + \frac{m}{v}}{\gamma + \frac{1}{v}} \right) \right]^2}$$

4. Suppose that we choose the following hyperparameters for our prior distributions—$m = 5, v = 5, a = 5, b = 5$—and that we observe $y = 10$. Write code to implement this Gibbs Sampler.

ANSWER:

5. Look at a histogram of your posterior samples for μ, γ and $\sigma^2 = \frac{1}{\gamma}$.

ANSWER:

6. Estimate the posterior mean and median of μ.

ANSWER:

7. Find a 90% credible interval for μ, and estimate the probability that $\mu > 2$.

ANSWER:

8. Ucomment the code to create a *trace plot* showing the behavior of the samples over the N iterations.Describe what these trace plots tell us.

ANSWER:

```
#iterations <- 1:N

#par(mfrow=c(1,3))
#plot(mus ~ iterations, xlab = 'Iteration', ylab = expression(mu), type = 'l')
#plot(gams ~ iterations, xlab = 'Iteration', ylab = expression(gamma), type = 'l')
#plot(1/gams ~ iterations, xlab = 'Iteration', ylab = expression(sigma^2), type = 'l')
```

9. As mentioned above, in practice we usually pick a burn-in period of initial iterations to remove. This decision is often motivated by the fact that, depending on your choice of starting value, it may take awhile for your chain of samples to look like it is "mixing" well. Play around with your choice of starting value above to see if you can find situations in which a burn-in period might be helpful.