Homework 2

Zach Calhoun

Due at 5:00 PM EDT on Friday, 14 January

Total pts: 10 (reproducibility) + 30 (Q1) + 20 (Q2) + 40 (Q3) = 100

1. *Lab component* (30 points total) Please refer to lab 2 and complete tasks 3—5.

(10) Task 3

```r
# This code was mostly taken from the lab

# set a seed
set.seed(123)
# create the observed data
obs.data <- rbinom(n = 100, size = 1, prob = 0.01)

### Bernoulli LH Function ###
# Input: obs.data, theta
# Output: bernoulli likelihood

# This function takes the observed data and a theta and
# calculates the Bernoulli likelihood function
myBernLH <- function(obs.data, theta) {
  N <- length(obs.data)
  x <- sum(obs.data)
  LH <- (theta^x)*(1-theta)^{N-x}
  return(LH)
}

### Plot LH for a grid of theta values ###
# Create the grid #
# Store the LH values
# Create the Plot

# Get the possible theta values to pass into the function.
theta.sim <- seq(from = 0, to = 1, length.out=1000)
```
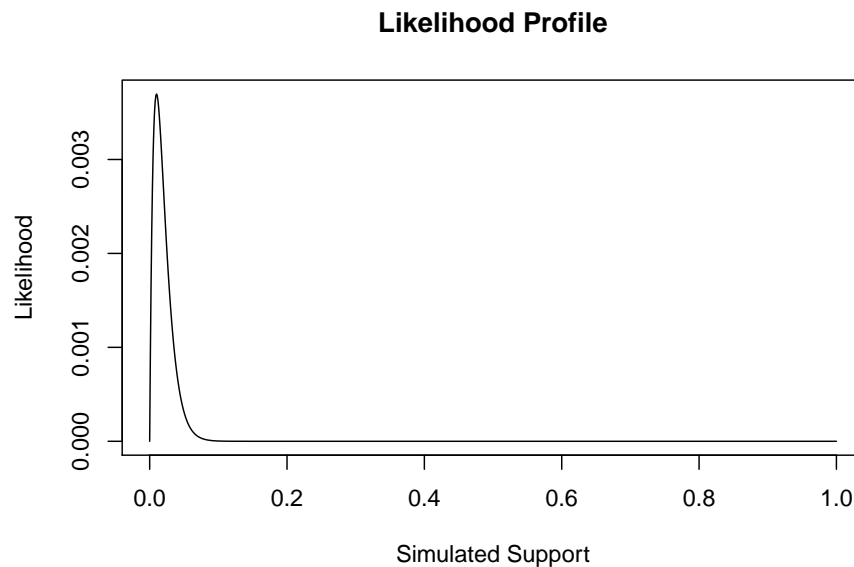
```
# Calculate the likelihood for the range of thetas.
sim.LH <- myBernLH(obs.data, theta = theta.sim)

# Plot the simulated likelihood
plot(theta.sim, sim.LH, type="l", main="Likelihood Profile",
     xlab="Simulated Support", ylab="Likelihood")
```

Likelihood Profile

(10) Task 4 Write a function that takes as its inputs prior parameters `a` and `b` for the Beta-Bernoulli model and the observed data, and produces the posterior parameters you need for the model. **Generate and print** the posterior parameters for a non-informative prior i.e. $(a,b) = (1,1)$ and for an informative case $(a,b) = (3,1)$}.

The posterior parameters for the Beta-Bernoulli model are denoted by $\sum X + a$ and $b + n - \sum X$, where n denotes the number of observations. Thus, we just need to create a function that returns these values as an array.

```
posteriorParams <- function(a, b, obs.data) {

  ## posteriorParams
  # Input: a, b, observed data
  # Output: posterior parameters

  # Get the length of the data
```

```
  n <- length(obs.data)
  # Get the sum of the observed data
  x <- sum(obs.data)

  # Output the parameters
  return(c(a+x, b+n-x))
}

print(posteriorParams(1,1,obs.data))
```

```
## [1]   2 100
```

```
print(posteriorParams(3,1,obs.data))
```

```
## [1]   4 100
```

(10) Task 5

Create two plots, one for the informative and one for the non-informative case to show the posterior distribution and superimpose the prior distributions on each along with the likelihood. What do you see? Remember to turn the y-axis ticks off since superimposing may make the scale non-sense.

```
th  <- seq(0,1, length=500)
# Non-informative case
a <- 1
b <- 1
x <- sum(obs.data)
n <- length(obs.data)
prior <- dbeta(th, a, b)
like <- dbeta(th, x+1, n-x+1)
postParams <- posteriorParams(a, b, obs.data)
post <- dbeta(th, postParams[1], postParams[2])
plot(th,like, type="l", ylab="Density", lty=2, lwd=3,
     xlab=expression(theta), yaxt="n",
     main="With a non-informative prior")
lines(th, prior, lty=1, lwd=3)
lines(th, post, col="blue", lty=3, lwd=3)
legend(0.5, 20, c("Likelihood", "Prior", "Posterior"),
       col=c("black", "black", "blue"), lty=c(2,1,3), lwd=c(3,3,3))
```
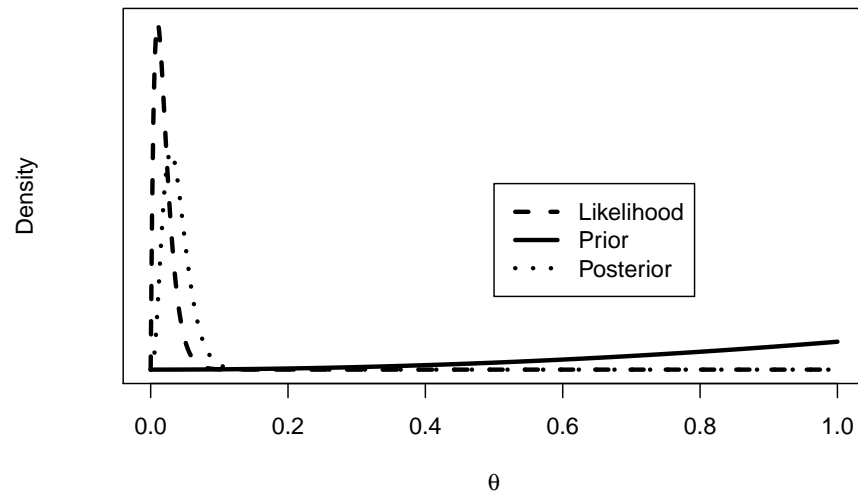
With a non–informative prior
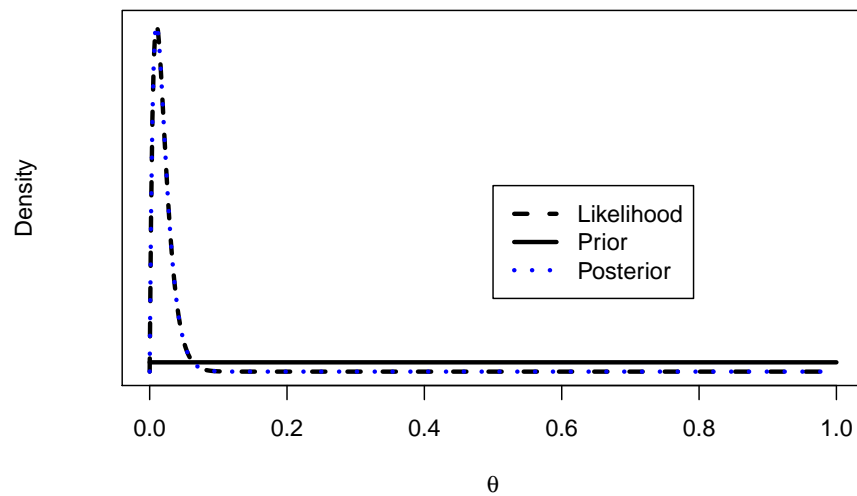

```
# Informative case
a <- 3
b <- 1
prior <- dbeta(th, a, b)
postParams <- posteriorParams(a, b, obs.data)
post <- dbeta(th, postParams[1], postParams[2])
plot(th,like, type="l", ylab="Density", lty=2, lwd=3,
     xlab=expression(theta), yaxt="n",
     main="With an informative prior")
lines(th, prior, lty=1, lwd=3)
lines(th, post, lty=3, lwd=3)
legend(0.5, 20, c("Likelihood", "Prior", "Posterior"),
       lty=c(2,1,3), lwd=c(3,3,3))
```

With an informative prior

When there is a non-informative prior, then the posterior is the same as the likelihood. With the informative prior, the posterior has its peak skewed in the direction of the prior. However, since the data is large enough, the posterior is not too different from the likelihood. As more data is added, the posterior approaches the likelihood.

The goal of this problem is to see how a conjugate model relates to real data. You will get practice deriving a posterior distribution that you have not seen before, plotting densities as we did in class, and seeing a connection to real data. Finally, you will get practice thinking about when the model below might be appropriate in practice.

2. (20 points total) *The Exponential-Gamma Model* We write $X \sim Exp(\theta)$ to indicate that X has the Exponential distribution, that is, its p.d.f. is

$$p(x|\theta) = Exp(x|\theta) = \theta \exp(-\theta x) \mathbb{K}(x > 0).$$

The Exponential distribution has some special properties that make it a good model for certain applications. It has been used to model the time between events (such as neuron spikes, website hits, neutrinos captured in a detector), extreme values such as maximum daily rainfall over a period of one year, or the amount of time until a product fails (lightbulbs are a standard example).

Suppose you have data $x_1, \ldots, x_n$ which you are modeling as i.i.d. observations from an Exponential distribution, and suppose that your prior is $\theta \sim Gamma(a, b)$, that is,

$$p(\theta) = Gamma(\theta|a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta) \mathbb{K}(\theta > 0).$$

(a) (5) Derive the formula for the posterior density, $p(\theta|x_{1:n})$. Give the form of the posterior in terms of one of the most common distributions (Bernoulli, Beta, Exponential, or Gamma).

$$p(\theta|x_{1:n}) \propto p(x_{1:n}|\theta)p(\theta) \tag{1}$$

$$\propto (\prod_{i=1}^{n} \theta \exp(-\theta x_i)) * \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta) \tag{2}$$

$$\propto (\prod_{i=1}^{n} \theta \exp(-\theta x_i))\theta^{a-1} \exp(-b\theta) \tag{3}$$

$$\propto \theta^n \exp(-\theta * \sum_{i=1}^{n} x_i)\theta^{a-1} \exp(-b\theta) \tag{4}$$

$$\propto \theta^{n+a-1} \exp(-\theta(\sum_{i=1}^{n} x_i + b)) \tag{5}$$

$$\propto Gamma(n + a, \sum_{i=1}^{n} x_i + b) \tag{6}$$

First, we can get rid of the marginal distribution $p(x)$ in 1, so we just look at the likelihood and the prior. In step 2, I write out the complete likelihood for

6

$x_{1:n}$ times the prior. In 3, I remove the constant $\frac{b^a}{\Gamma(a)}$, then in step 4, I rewrite the likelihood product as an exponent of sums, which then allows me to combine terms in step 5. After looking at 5, I see that I have the same form as the gamma distribution, so the posterior is proportional to a $Gamma(n + a, \sum_{i=1}^{n} x_i - b)$.

(b) (5) Why is the posterior distribution a *proper* density or probability distribution function?

In this case, we know that the posterior distribution is a proper density because it is in the form of our prior. Since we have conjugacy, we know that the posterior is proper.

(c) (5) Now, suppose you are measuring the number of seconds between lightning strikes during a storm, your prior is $Gamma(0.1, 1.0)$, and your data is

$$(x_1, \ldots, x_8) = (20.9, 69.7, 3.6, 21.8, 21.4, 0.4, 6.7, 10.0).$$

Plot the prior and posterior p.d.f.s. (Be sure to make your plots on a scale that allows you to clearly see the important features.)

```
th = seq(0,1, length=1000)
a = 0.1
b = 1
prior <- dgamma(th, a, b)

x = c(20.9, 69.7, 3.6, 21.8, 21.4, 0.4, 6.7, 10.0)
x.sum = sum(x)
x.n = length(x)
post = dgamma(th, x.n + a, x.sum + b)
plot(th,prior, type="l", ylab="Density", lty=2, lwd=3,
     xlab=expression(theta), yaxt="n", ylim=c(0,30))

lines(th, post, lty=1, lwd=3)
legend(0.5, 10, c("Prior", "Posterior"),lty=c(2,1), lwd=c(3,3))
```
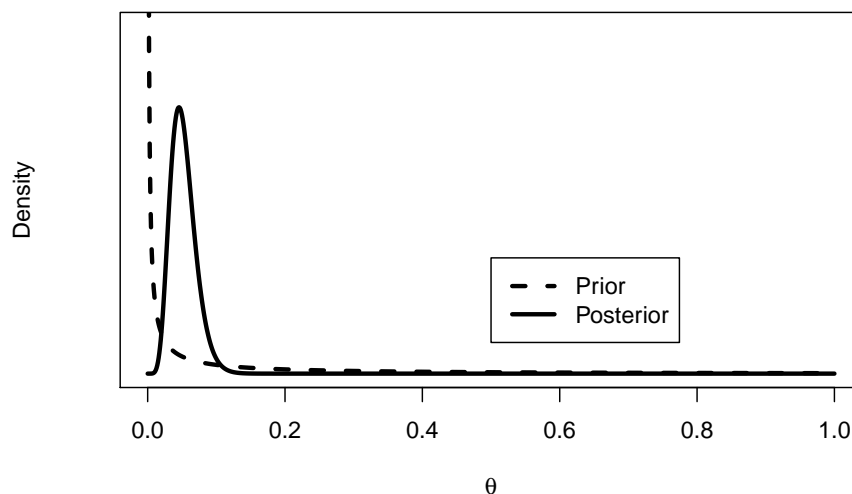
(d) (5) Give a specific example of an application where an Exponential model would be reasonable. Give an example where an Exponential model would NOT be appropriate, and explain why.

The exponential model has support $[0, +\inf]$, with a decaying probability of occurring as x increases. An appropriate application might be modeling the time between orders in a coffee shop. We would expect that there would be a high density between 0 and the mean, with decaying density following the mean. This is reasonable because there will likely be a lot of times when the coffee shop has near 0 time between orders, but also slow periods where orders do not occur as often.

An inappropriate example would be when an event almost never occurs in rapid succession (because the mode in an exponential distribution is 0). For example, if we wanted to measure the time between bus stops along a route, we would assume that there is some minimum time greater than zero, and some maximum time. In this case, the mode is probably closer to the mean.

The goal of this problem is to introduce you to a new family of distributions, get more practice deriving the posterior, and work with a posterior predictive distribution on your own for the first time. This will be an intense problem, so reach out if you're having trouble!

3. (40 points total) *Priors, Posteriors, Predictive Distributions (Hoff, 3.9)* An unknown quantity $Y|\theta$ has a Galenshore(a, θ) distribution if its density is given by

$$p(y|\theta) = \frac{2}{\Gamma(a)} \theta^{2a} y^{2a-1} e^{-\theta^2 y^2}$$

for $y > 0, \theta > 0, a > 0$. Assume for now that a is known and θ is unknown and a random variable. For this density,

$$E[Y] = \frac{\Gamma(a+1/2)}{\theta\Gamma(a)}$$

and

$$E[Y^2] = \frac{a}{\theta^2}.$$

(a) (10) Identify a class of conjugate prior densities for θ. Assume the prior parameters are c and d. That is, state the distribution that θ should have with parameters c, d such that the resulting posterior is conjugate. Plot a few members of this class of densities.

To find a suitable conjugate prior, we need to reduce the Galenshore distribution to only show which components of the equation will show up when solving for the posterior. We can write that as:

$$p(\theta) \propto \theta^{2a} \exp(\theta^2 y^2)$$

This looks pretty similar to a gamma distribution, however, since this equation contains θ^2, it is probably easier to just assume a Galenshore prior, since that function contains the same terms along with the θ^2 form. We can confirm whether this prior is conjugate by solving for the posterior in part b. We can plot several examples of this function as follows:

```
th = seq(0,10,length=500)

dgalenshore <- function(th, c, d) {
  return((2/gamma(c))*th^(2*d)*th^(2*c-1)*exp(-d^2*th*2))
}

plot(th, dgalenshore(th, 1,1), type="l", lty=2, lwd=3,
     xlab=expression(theta), ylab="galenshore", ylim=c(0,10))

c <- seq(0.1, 10, length=10)
d <- seq(0.1, 10, length=10)
```
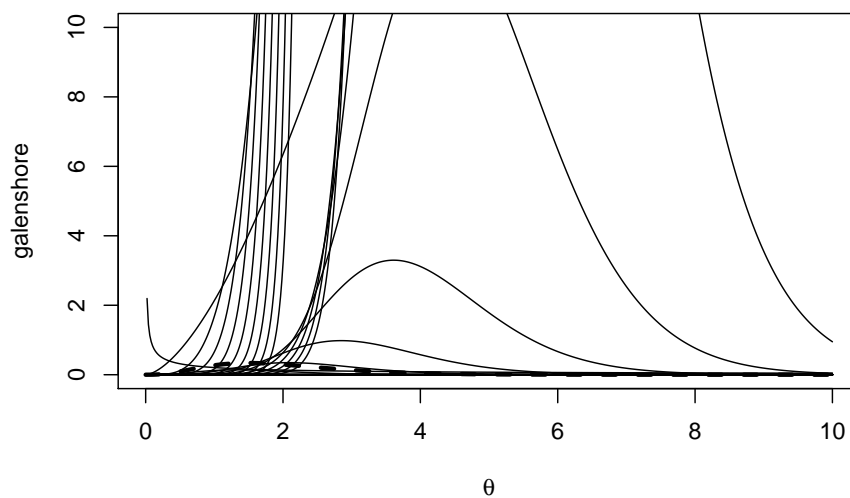
```
for(i in c) {
  for(j in d) {
    lines(th, dgalenshore(th, i, j))
  }
}
```


As seen in the graph above, the value of the galenshore increases very quickly as theta increases,

(b)   (5) Let $Y_1, \ldots, Y_n \overset{iid}{\sim}$ Galenshore(a, θ). Find the posterior distribution of $\theta | y_{1:n}$ using a prior from your conjugate class.

If we set the prior $p(\theta)$ to be Galenshore(c, d), then we can derive the posterior distribution as follows:

$$p(\theta|Y) = p(Y|\theta)p(\theta) \tag{7}$$

$$= (\prod_{i=1}^{n} \frac{2}{\Gamma(a)} \theta^{2a} y_i^{2a-1} e^{-\theta^2 y_i^2}) \frac{2}{\Gamma(c)} d^{2c} \theta^{2c-1} e^{-d^2\theta^2} \tag{8}$$

$$\propto (\prod_{i=1}^{n} \theta^{2a} e^{-\theta^2 y_i^2}) \theta^{2c-1} e^{-d^2\theta^2} \tag{9}$$

$$\propto \theta^{2na} \exp(-\theta^2 \sum_{i=1}^{n} y_i^2) \theta^{2c-1} \exp(-d^2\theta^2) \tag{10}$$

$$\propto \theta^{2na+2c-1} \exp(-\theta^2 (\sum_{i=1}^{n} y_i^2 + d^2)) \tag{11}$$

$$\propto \theta^{2(na+c)-1} \exp(-\theta^2 (\sqrt{\sum_{i=1}^{n} y_i^2 + d^2})^2) \tag{12}$$

$$\propto Galenshore(na + c, \sqrt{\sum_{i=1}^{n} y_i^2 + d^2}) \tag{13}$$

(c) (10) Show that

$$\frac{p(\theta_a|y_{1:n})}{p(\theta_b|y_{1:n})} = \left(\frac{\theta_a}{\theta_b}\right)^{2(an+c)-1} e^{(\theta_b^2 - \theta_a^2)(d^2 + \sum y_i^2)},$$

where

$$\theta_a, \theta_b \sim \text{Galenshore}(c, d).$$

Identify a sufficient statistic.

To show this, we simply need to write out the full forms of each posterior, then simplify, as follows:

$$\frac{p(\theta_a|y_{1:n})}{p(\theta_b|y_{1:n})} = \frac{Galenshore(na + c, \sqrt{\sum_{i=1}^{n} y_i^2 + d^2})}{Galenshore(na + c, \sqrt{\sum_{i=1}^{n} y_i^2 + d^2})} \tag{14}$$

$$= \frac{\frac{2}{\Gamma(na+c)}(\sum_{i=1}^{n} y_i^2 + d^2)^{an+c} \theta_a^{2(an+c)-1} \exp(-\theta_a^2(\sum_{i=1}^{n} y_i^2 + d^2))}{\frac{2}{\Gamma(na+c)}(\sum_{i=1}^{n} y_i^2 + d^2)^{an+c} \theta_b^{2(an+c)-1} \exp(-\theta_b^2(\sum_{i=1}^{n} y_i^2 + d^2))} \tag{15}$$

$$= \frac{\theta_a^{2(an+c)-1} \exp(-\theta_a^2(\sum_{i=1}^{n} y_i^2 + d^2))}{\theta_b^{2(an+c)-1} \exp(-\theta_b^2(\sum_{i=1}^{n} y_i^2 + d^2))} \tag{16}$$

$$= \left(\frac{\theta_a}{\theta_b}\right)^{2(an+c)-1} \exp(\theta_b^2 - \theta_a^2(\sum_{i=1}^{n} y_i^2 + d^2)) \tag{17}$$

11

We can see based on the above that the constants cancel out, yielding the equation originally provided, when simplified. In this case, the sufficient statistic is $\sum_{i=1}^{n} y_i^2$.

(d) (5) Determine $E[\theta|y_{1:n}]$.

Given that $E[Y] = \frac{\Gamma(a+1/2)}{\theta\Gamma(a)}$, we can plug in the parameters derived from part b to get $E[\theta|y_{1:n}]$:

$$E[\theta|y_{1:n}] = E\left[Galenshore(na + c, \sqrt{\sum_{i=1}^{n} y_i^2 + d^2})\right] \quad (18)$$

$$= \frac{\Gamma(na + c + 1/2)}{\sqrt{\sum_{i=1}^{n} y_i^2 + d^2}\,\Gamma(na + c)} \quad (19)$$

(e) (10) Show that the form of the posterior predictive density

$$p(y_{n+1}|y_{1:n}) = \frac{2y_{n+1}^{2a-1}\Gamma(an + a + c)}{\Gamma(a)\Gamma(an + c)} \frac{(d^2 + \sum y_i^2)^{an+c}}{(d^2 + \sum y_i^2 + y_{n+1}^2)^{(an+a+c)}}.$$

$$p(y_{n+1}|y_{1:n}) = \int p(y_{n+1}|\theta)p(\theta|y_{1:n})d\theta$$

$$= \int \frac{2}{\Gamma(a)}\theta^{2a}y_{n+1}^{2a-1}\exp(-\theta^2 y_{n+1}^2)\frac{2}{\Gamma(an + c)}(\sum_{i=1}^{n} y_i^2 + d^2)^{an+c}\theta^{2(an+c)-1}\exp(-\theta^2(\sum_{i=1}^{n} y_i^2 + d^2))d\theta$$

$$= \int \frac{4(\sum_{i=1}^{n} y_i^2 + d^2)^{an+c}y_{n+1}^{2a-1}}{\Gamma(a)\Gamma(an + c)}\theta^{2(a+an+c)-1}\exp(-\theta^2(\sum_{i=1}^{n} y_i^2 + d^2 + y_{n+1}^2))d\theta$$

The first fractional term is a constant, so we can move out the integral to get:

$$= \frac{4(\sum_{i=1}^{n} y_i^2 + d^2)^{an+c}y_{n+1}^{2a-1}}{\Gamma(a)\Gamma(an + c)}\int \theta^{2(a+an+c)-1}\exp(-\theta^2(\sum_{i=1}^{n} y_i^2 + d^2 + y_{n+1}^2))d\theta$$

We can then recognize the integral as a form of the Gamma function since we are integrating from $(0, +\inf]$, where:

$$\frac{\Gamma(a + an + c)}{2 * (\sum_{i=1}^{n} y_i^2 + d^2 + y_{n+1}^2)^{(an+a+c)}} = \int \theta^{2(a+an+c)-1}\exp(-\theta^2(\sum_{i=1}^{n} y_i^2 + d^2 + y_{n+1}^2))d\theta$$

Noticing this allows us to write out $p(y_{n+1}|y_{1:n})$ completely as:

$$p(y_{n+1}|y_{1:n}) = \frac{2y_{n+1}^{2a-1}\Gamma(an + a + c)}{\Gamma(a)\Gamma(an + c)} \frac{(d^2 + \sum y_i^2)^{an+c}}{(d^2 + \sum y_i^2 + y_{n+1}^2)^{(an+a+c)}}$$