STA 602. HW03

Yicheng Shen

9/17/2022

1. PH 3.3

(a) Here we first derive the posterior of θ:

$$
\begin{aligned}
p(\theta|y_1,...y_n) &\propto p(\theta) \times p(y_i|\theta) \\
&= \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}\prod_{i=1}^{n}\frac{\theta^y e^{-\theta}}{(y_i)!} \\
&\propto \theta^{a-1}e^{-b\theta}\frac{\theta^{\sum_{i=1}^n (y_i)}e^{-\theta n}}{\prod_{i=1}^n (y_i)!} \\
&= \theta^{a-1}e^{-(b+n)\theta}\frac{\theta^{\sum_{i=1}^n (y_i)}}{\prod_{i=1}^n (y_i)!} \\
\text{Find the kernel of Gamma pdf} \quad &\propto \theta^{a-1+\sum_{i=1}^n (y_i)}e^{-(b+n)\theta} \\
p(\theta|y_i) &\sim Gamma(a+\sum_{i=1}^{n} y_i,\ b+n)
\end{aligned}
$$

Therefore, for θ_A, its posterior distribution is Gamma(120+117,10+10) = Gamma(237,20). The mean is $\frac{237}{20} = 11.85$, the variance is $\frac{237}{20^2} = 0.5925$, and the 95% CI is from 10.389 to 13.405.

For θ_B, its posterior distribution is Gamma(12+113,1+13) = Gamma(125,14). The mean is $\frac{125}{14} \approx 8.929$, the variance is $\frac{125}{14} \approx 0.638$, and the 95% CI is from 7.432 to 10.560.

```r
y_A <- c(12, 9, 12, 14, 13, 13, 15, 8, 15, 6)
sum(y_A)
```

```
## [1] 117
```

```r
qgamma(c(0.025, 0.975), 237, 20)
```

```
## [1] 10.38924 13.40545
```

```r
y_B <- c(11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)
sum(y_B)
```

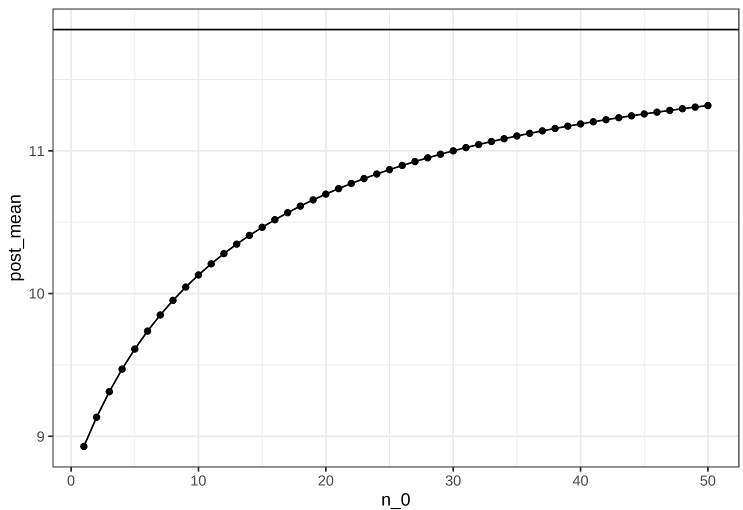
```
## [1] 113
```

```
qgamma(c(0.025, 0.975), 125, 14)
```

```
## [1]  7.432064 10.560308
```

(b) The prior of θ_B has to have large prior sample size (large n_0) in order for its posterior expectation to be close to that of θ_A. In other words, we need to have a very strong prior belief.

```
n_0 <- seq(1, 50, by = 1)
gamma_pois_table <- data.frame(n_0) %>%
  mutate(a = 12 * n_0, b = n_0) %>%
  mutate(post_mean = (a + sum(y_B)) / (b + length(y_B)))
ggplot(gamma_pois_table, aes(x = n_0 , y = post_mean)) + geom_line() +
  geom_point() + geom_hline(yintercept = 11.85) + theme_bw()
```



(c) Since the study has specified that type B mice are related to type A mice, knowing knowledge about population A should inform us about the prior of population B, for example setting a prior for B that is similar to A. However, the true Tumor count rates for type B mice are unknown parameters, therefore we should still view two populations as independent, hence it is valid to assume

$$p(\theta_A, \theta_B) = p(\theta_A) \times p(\theta_B)$$

2. PH 3.5

(a) From Section 3.3, we know that

$$p(y|\phi) = c(\phi)h(y)exp\{\phi t(y)\}$$
$$p(\phi|n_o, t_0) = \kappa(n_0, t_0 c(\phi))^{n_0} e^{n_0 t_0 \phi}$$

For the mixture prior, we have

$$\begin{aligned}
\tilde{p}(\theta) &= \sum_{k=1}^{K} w_k p_k(\theta) \quad \text{As defined} \\
&= \sum_{k=1}^{K} w_k p_k(\theta \mid n_{0,k}, t_{0,k}) \\
&= \sum_{k=1}^{K} \left(w_k \kappa(n_{0,k}, t_{0,k}) c(\phi)^{n_{0,k}} e^{n_{0,k} t_{0,k} \phi} \right)
\end{aligned}$$

So for the posterior, we have

$$\begin{aligned}
p(\phi \mid y_1, \ldots, y_n) &\propto \tilde{p}(\phi)\tilde{p}(y_1, \ldots, y_n \mid \phi) \\
&\propto \left[\sum_{k=1}^{K} \left(w_k \kappa(n_{0,k}, t_{0,k}) c(\phi)^{n_{0,k}} e^{n_{0,k} t_{0,k} \phi} \right) \right] \times \left[\prod_{i=1}^{n} h(y_i) c(\phi) e^{\phi t(y_i)} \right] \\
&= \left[\sum_{k=1}^{K} \left(w_k \kappa(n_{0,k}, t_{0,k}) c(\phi)^{n_{0,k}} e^{n_{0,k} t_{0,k} \phi} \right) \right] \times \left[c(\phi)^n e^{\phi \sum_{i=1}^{n} t(y_i)} \prod_{i=1}^{n} h(y_i) \right] \\
\text{Only keep related terms } &\propto \left[\sum_{k=1}^{K} \left(w_k \kappa(n_{0,k}, t_{0,k}) c(\phi)^{n_{0,k}} e^{n_{0,k} t_{0,k} \phi} \right) \right] \times \left[c(\phi)^n e^{\phi \sum_{i=1}^{n} t(y_i)}) \right] \\
&\propto \sum_{k=1}^{K} w'_k \kappa(n_{0,k}, t_{0,k}) c(\phi)^{n+n_{0,k}} e^{n_{0,k} t_{0,k} \phi} \times e^{\phi \sum_{i=1}^{n} t(y_i)} \\
&= \sum_{k=1}^{K} w'_k \kappa(n_{0,k}, t_{0,k}) c(\phi)^{n+n_{0,k}} e^{\phi \times (n_{0,k} t_{0,k} + \sum_{i=1}^{n} t(y_i))} \\
&\propto \sum_{k=1}^{K} w'_k \times p \left(\theta \mid n + n_0, \ n_0 t_0 + \sum_{i=1}^{n} t(y_i) \right)
\end{aligned}$$

Therefore, the expression shows that the general form of the posterior distribution is also a mixture and in the same conjugate class as the prior, with new weights.

(b) Now we plug in the pdf of Gamma prior

$$\begin{aligned}
\tilde{p}(\theta) &= \sum_{k=1}^{K} w_k p_k(\theta) \\
&= \sum_{k=1}^{K} w_k p_k(\theta \mid a_k, b_k) \\
&= \sum_{k=1}^{K} w_k \times Gamma(a_k, b_k) \\
&= \sum_{k=1}^{K} w_k \times \left(\frac{b_k^{a_k}}{\Gamma(a_k)} \theta^{a_k - 1} e^{-b_k \theta} \right)
\end{aligned}$$

Then we derive the posterior distribution (the sampling model is a series of poisson):

$$
\begin{aligned}
p(\theta \mid y_1, \ldots, y_n) &\propto \tilde{p}(\theta)\tilde{p}(y_1, \ldots, y_n \mid \theta) \\
&\propto \left[\sum_{k=1}^{K} w_k \left(\frac{b_k^{a_k}}{\Gamma(a_k)}\theta^{a_k-1}e^{-b_k\theta}\right)\right] \times \left[\prod_{i=1}^{n}\frac{1}{y_i!}\theta^{y_i}e^{-\theta}\right] \\
\text{Only keep related terms } &\propto \left[\sum_{k=1}^{K} w_k \left(\theta^{a_k-1}e^{-b_k\theta}\right)\right] \times \left[\theta^{\sum y_i}e^{-n\theta}\right] \\
&= \sum_{k=1}^{K} w_k' \left(\theta^{a_k+\sum y_i-1}e^{-(b_k+n)\theta}\right) \\
\text{Recognize Gamma Kernel} &= \sum_{k=1}^{K} w_k' \times Gamma\left(a_k + \sum y_i, b_k + n\right) \\
&\propto \sum_{k=1}^{K} w_k' \times p\left(\theta \,\middle|\, a_k + \sum y_i,\ b_k + n\right)
\end{aligned}
$$

So from the Gamma kernel, it is shown that the posterior distribution is also a mixture of Gamma, but with different weights.

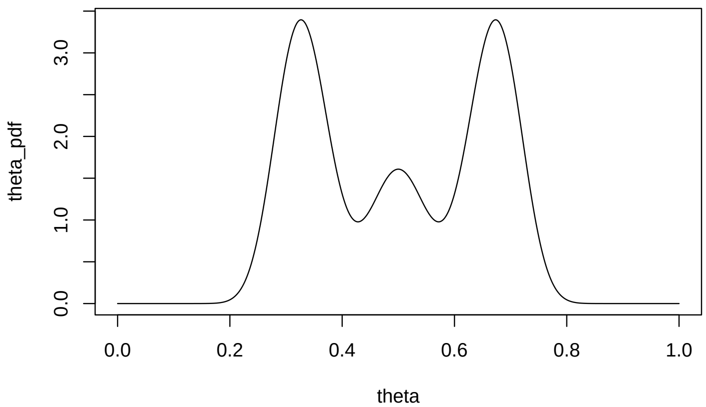
3. PH 3.8

(a) Since our prior specification comes from observing the long-run frequencies of coin flipping, this prior should be quite strong, in other words, having large prior sample sizes (for example 100).

Also by description, the prior should centers around $\frac{1}{2}$ 20% of time, and centers around $\frac{1}{4}$ or $\frac{1}{4}$ 80% of time. So here we could specify this prior using a mixture of three Beta distributions.

$$
\begin{aligned}
p(\theta) &= \frac{1}{5}\frac{\Gamma(100)}{\Gamma(50)\Gamma(50)}\theta^{49}(1-\theta)^{49} + \frac{2}{5}\frac{\Gamma(100)}{\Gamma(33)\Gamma(67)}\theta^{32}(1-\theta)^{66} + \frac{2}{5}\frac{\Gamma(100)}{\Gamma(67)\Gamma(33)}\theta^{66}(1-\theta)^{32} \\
&= \frac{1}{5}p_1(\theta) + \frac{2}{5}p_2(\theta) + \frac{2}{5}p_3(\theta) \\
\text{Where } & p_1(\theta) = Beta(50, 50),\ p_2(\theta) = Beta(33, 67),\ p_3(\theta) = Beta(67, 33)
\end{aligned}
$$

```r
theta <- seq(from = 0, to = 1, length.out = 1000)
theta_pdf <- 0.2 * dbeta(theta, 50,50) + 0.4 * dbeta(theta, 33, 67) + 0.4 * dbeta(theta, 67,33)
plot(theta, theta_pdf, type = "l")
```



(b) Coin information: year is 2017 and denomination is quarter dollar.

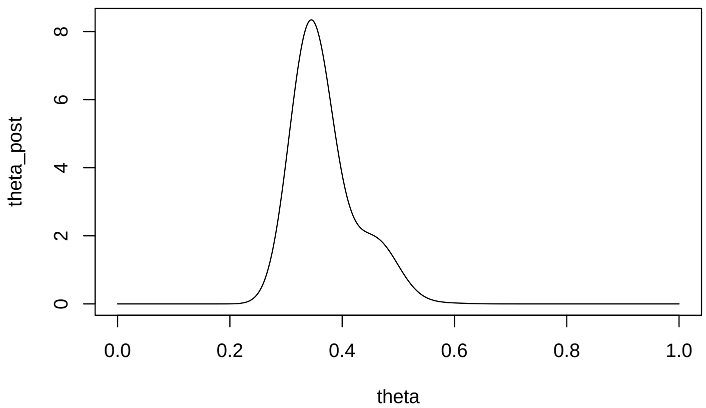
Total Tests: 50; total successes (heads): 19; total failures (tails): 31

(c) The derivation

$$
\begin{aligned}
p(\theta|y_i) &\propto \frac{1}{5}p_1(\theta) \times p(y_i|\theta) + \frac{2}{5}p_2(\theta) \times p(y_i|\theta) + \frac{2}{5}p_3(\theta) \times p(y_i|\theta) \\
&\propto \frac{1}{5}p_1(\theta)\left[\binom{50}{19}\theta^{19}(1-\theta)^{31}\right] + \frac{2}{5}p_2(\theta)\left[\binom{50}{19}\theta^{19}(1-\theta)^{31}\right] + \frac{2}{5}p_3(\theta)\left[\binom{50}{19}\theta^{19}(1-\theta)^{31}\right] \\
\text{Weights calcualted below } &\propto 0.186591p_1(\theta|y_i) + 0.4483797p_2(\theta|y_i) + 0.003033463p_3(\theta|y_i) \\
\text{Where } &\; p_1(\theta|y_i) = Beta(69, 81),\; p_2(\theta|y_i) = Beta(52, 98),\; p_3(\theta|y_i) = Beta(86, 64)
\end{aligned}
$$

We can see from the posterior distribution of θ that with new data our posterior belief seems to think θ is more likely to be below 0.5.

```r
demon <- (0.4*beta(52,98)/beta(33,67) + 0.2*beta(69,81)/beta(50,50) + 0.4*beta(86, 64)/beta(67,33))
w1 <- 0.2*beta(69,81)/beta(50,50) / demon
w2 <- 0.4*beta(52,98)/beta(33,67) / demon
w3 <- 0.4*beta(86,64)/beta(67,33)/ demon
theta_post <- w1 * dbeta(theta, 69,81) + w2 * dbeta(theta, 52, 98) + w3 * dbeta(theta, 86,64)
plot(theta, theta_post, type = "l")
```



(d) This time, I used another quarter dollar, but it was made in 1986.

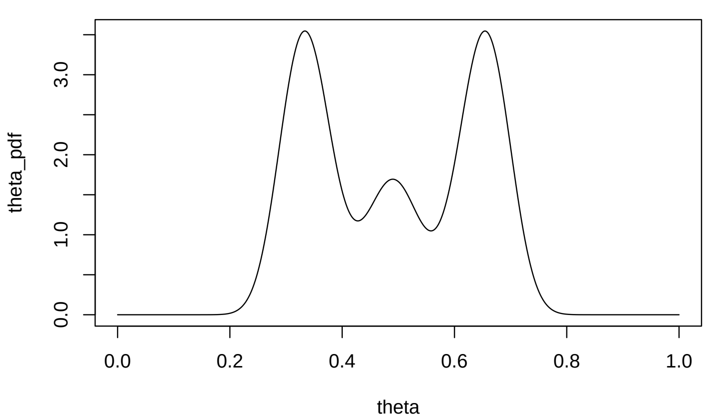
Since they share the same denomination, they are probably similar, but still not identical. I would use a prior for this one that is similar to the previous one's posterior by adding a prior sample size of 10 (not very strong) with prior success of 4 (because last experiment was 19 out of 50).

$$
\begin{aligned}
p(\theta) &= \frac{1}{5}\frac{\Gamma(110)}{\Gamma(54)\Gamma(56)}\theta^{53}(1-\theta)^{55} + \frac{2}{5}\frac{\Gamma(110)}{\Gamma(37)\Gamma(73)}\theta^{36}(1-\theta)^{72} + \frac{2}{5}\frac{\Gamma(110)}{\Gamma(71)\Gamma(39)}\theta^{70}(1-\theta)^{38} \\
&= \frac{1}{5}p_1(\theta) + \frac{2}{5}p_2(\theta) + \frac{2}{5}p_3(\theta)
\end{aligned}
$$

Where $p_1(\theta) = Beta(54, 56),\ p_2(\theta) = Beta(37, 73),\ p_3(\theta) = Beta(71, 39)$

The prior distribution is shown below (prior mean is 0.491 because of information from last time)

```
theta <- seq(from = 0, to = 1, length.out = 1000)
theta_pdf <- 0.2 * dbeta(theta, 54, 56) + 0.4 * dbeta(theta, 37, 73) + 0.4 * dbeta(theta, 73, 39)
plot(theta, theta_pdf, type = "l")
```



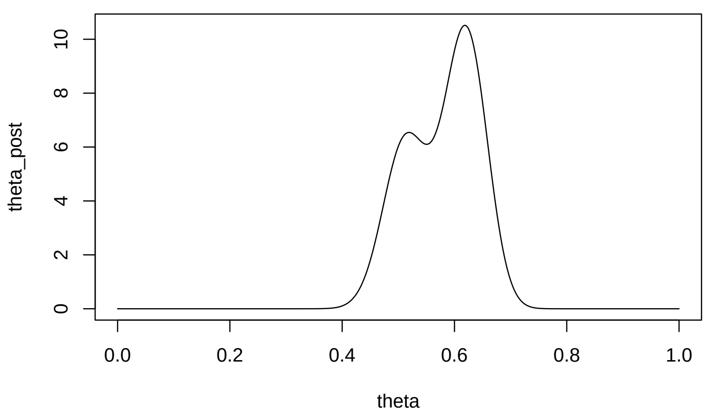
New Experiment Result: Total = 50, Successes: 28 The the posterior for this one is

$$
\begin{aligned}
p(\theta|y_i) &\propto \frac{1}{5}p_1(\theta) \times p(y_i|\theta) + \frac{2}{5}p_2(\theta) \times p(y_i|\theta) + \frac{2}{5}p_3(\theta) \times p(y_i|\theta) \\
&\propto \frac{1}{5}p_1(\theta) \times \left[\binom{50}{28}\theta^{28}(1-\theta)^{22}\right] + \frac{2}{5}p_2(\theta) \times \left[\binom{50}{28}\theta^{28}(1-\theta)^{22}\right] + \frac{2}{5}p_3(\theta) \times \left[\binom{50}{28}\theta^{28}(1-\theta)^{22}\right] \\
&\propto 0.6183211p_1(\theta|y_i) + 0.0000497(\theta|y_i) + 0.9984771p_3(\theta|y_i)
\end{aligned}
$$

Where $p_1(\theta|y_i) = Beta(82, 78),\ p_2(\theta|y_i) = Beta(65, 95),\ p_3(\theta|y_i) = Beta(99, 61)$

The posterior distribution is shown. Now the posterior belief of θ seems to be more concentrated to higher values (because of the new data).

```
demon <-(0.4*beta(65,95)/beta(33,67) + 0.2*beta(82,78)/beta(37, 73)  + 0.4*beta(99, 61)/beta(71,39))
w1 <- 0.2*beta(82,78)/beta(54,56) / demon
w2 <- 0.4*beta(65,95)/beta(33,67)/ demon
w3 <- 0.4*beta(99, 61)/beta(71,39)/ demon
theta_post <- w1 * dbeta(theta, 82, 78) + w2 * dbeta(theta, 65, 95) + w3 * dbeta(theta, 99, 61)
plot(theta, theta_post, type = "l")
```



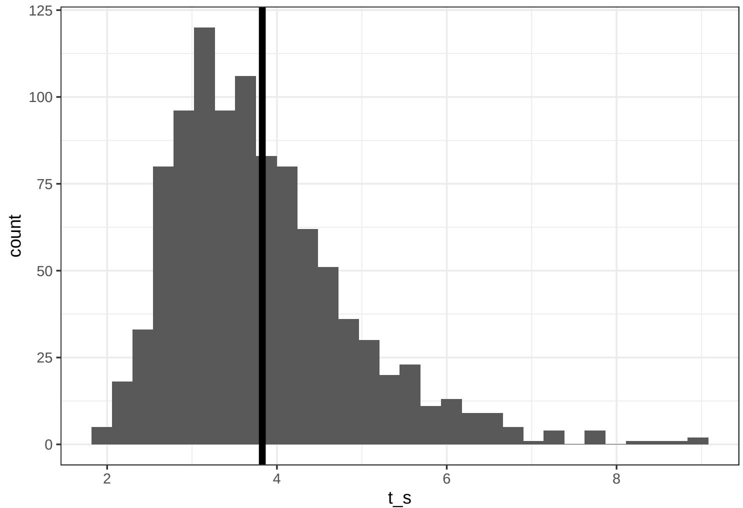
4. PH 4.3

(a) The tumor data with y_A and y_B is already loaded in question 3.3.

In the case of y_A, we simulate 1000 sample groups with $\theta^{(s)}$, and calculate the sample mean and sd for each. In the plotted histgram, the observed mean(y_A)/sd(y_A) is also shown as a vertical line.

```
set.seed(991109)
a <- 120; b <- 10
theta_A = rgamma(1000, a + sum(y_A), b + length(y_A))
# generate posterior predictive datasets
t_s <- sapply(theta_A, function(theta) {
  yi = rpois(10, theta)
  t_s = mean(yi) / sd(yi)
  t_s })
```

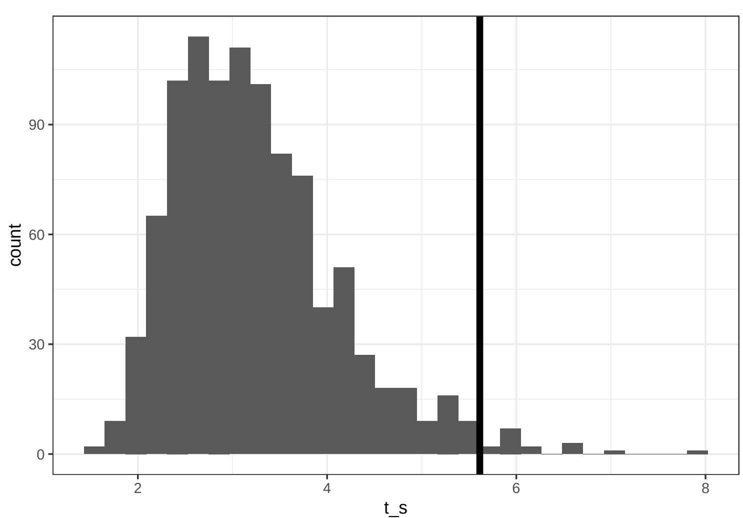
```
ggplot(data.frame(t_s), aes(t_s)) + geom_histogram(bins = 30) +
  theme_bw() + geom_vline(xintercept = mean(y_A) / sd(y_A), size = 2)
```



From the plot, it seems that the observed statistic (3.8) is a reasonable value among the spread of the posterior predictive distribution of $t^{(s)}$. Therefore, the fit of Poisson model should be acceptable in this case.

(b) The same procedures are repeated for y_B here:

```
set.seed(8848)
a <- 12; b <- 1
theta_B <- rgamma(1000, a + sum(y_B), b + length(y_B))
t_s <- sapply(theta_B, function(theta) {
  yi = rpois(10, theta)
  t_s = mean(yi) / sd(yi)
  t_s })
ggplot(data.frame(t_s), aes(t_s)) + geom_histogram(bins = 30) +
  theme_bw() + geom_vline(xintercept = mean(y_B) / sd(y_B), size = 2)
```



Now notice that the observed statistic (5.6) seems to be an outlier among the simulated data (above the 97.5% quantile = 5.3), so it means that the Poisson model does not seem to fit well.

5.

(a) We know that the sampling data is $\sum_{n=1}^{n} X_i = 200,\ n = 10$, and our prior of θ is $Gamma(10, 1)$.

Therefore, we could derive the posterior to be

$$p(\theta|x) = Gamma(10 + 200, 1 + 10) = Gamma(210, 11)$$

Now we construct the Bayes estimates under absolute error loss:

$$\begin{aligned}
\delta^*(\mathbf{x}) &= \text{argmin}_a E(L(\theta, a)|\mathbf{x}) \\
&= \text{argmin}_a \int_{-\infty}^{\infty} |\theta - a| p(\theta|\mathbf{x}) \mathbf{d}\theta \\
&= \text{argmin}_a \left[\int_{\delta^*(\mathbf{x})}^{\infty} |\theta - a| p(\theta|\mathbf{x}) \mathbf{d}\theta + \int_{-\infty}^{\delta^*(\mathbf{x})} |\theta - \mathbf{a}| \mathbf{p}(\theta|\mathbf{x}) \mathbf{d}\theta \right] \\
&= \text{argmin}_a \left[\int_{\delta^*(\mathbf{x})}^{\infty} (\theta - a) p(\theta|\mathbf{x}) \mathbf{d}\theta + \int_{-\infty}^{\delta^*(\mathbf{x})} (\mathbf{a} - \theta) \mathbf{p}(\theta|\mathbf{x}) \mathbf{d}\theta \right] \\
&\text{Now we differentiate with respect to } \delta^*(\mathbf{x}) \\
0 &= \int_{\delta^*(\mathbf{x})}^{\infty} -p(\theta|\mathbf{x}) \mathbf{d}\theta + \int_{-\infty}^{\delta^*(\mathbf{x})} \mathbf{p}(\theta|\mathbf{x}) \mathbf{d}\theta \\
\int_{\delta^*(\mathbf{x})}^{\infty} p(\theta|\mathbf{x}) \mathbf{d}\theta &= \int_{-\infty}^{\delta^*(\mathbf{x})} p(\theta|\mathbf{x}) \mathbf{d}\theta \\
\text{Since } \int_{-\infty}^{\infty} p(\theta|\mathbf{x}) \mathbf{d}\theta = 1, &\int_{\delta^*(\mathbf{x})}^{\infty} p(\theta|\mathbf{x}) \mathbf{d}\theta = \mathbf{\frac{1}{2}} \\
\text{Check with 2nd derivative } &\ 2p(\theta|\mathbf{x}) > 0 \text{ So it is minimum}
\end{aligned}$$

Therefore, the Bayes estimates under absolute error loss is in fact the posterior median, which means

$$\delta^*(\boldsymbol{x}) = \text{median}[Gamma(210, 11)] = 19.06061$$

```
qgamma(0.5,210,11)
```

```
## [1] 19.06061
```

According to the lecture, the Bayes estimates under squared error loss is proved to be the posterior mean under the finite and well-defined constraint.

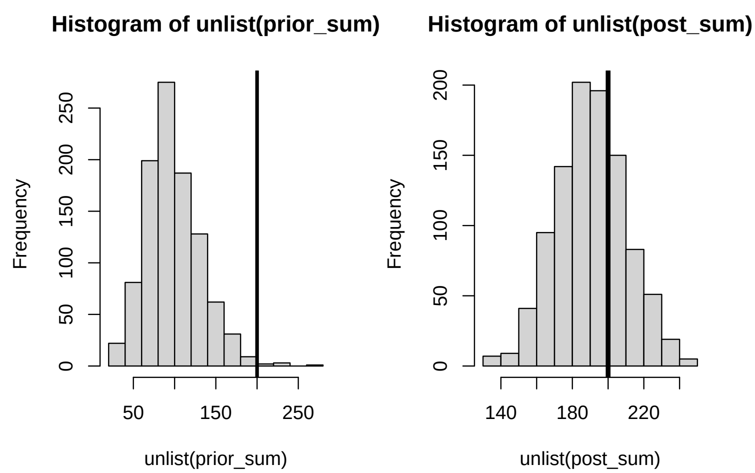
Therefore, we use the mean of Gamma(210, 11) here

$$\delta^*(\boldsymbol{x}) = \text{mean}[Gamma(210, 11)] = E(\theta|\boldsymbol{x}) = \frac{210}{11} = 19.09$$

(b) The prior is $Gamma(10, 1)$, which means that the prior mean of θ is 10.

The observed data has a sum of 200 counts over $n = 10$ days, so on average that is 20 calls received per day. And we assume it is a Poisson model.

```r
set.seed(10086)
y_prior_draw <- list()
y_post_draw <- list()
prior_sum <- list()
post_sum <- list()
for (i in 1:1000)
{
  y_prior_draw[[i]] <- rpois(n = 10, lambda =  rgamma(1, 10, 1))
  y_post_draw[[i]] <- rpois(n = 10, lambda =  rgamma(1, 210, 11))
  prior_sum[[i]] <- sum(y_prior_draw[[i]])
  post_sum[[i]] <- sum(y_post_draw[[i]])
}
par(mfrow = c(1, 2))
hist(unlist(prior_sum))
abline(v = 200, lwd = 3)
hist(unlist(post_sum))
abline(v = 200, lwd = 4)
```



Although the prior is probably not a good guess, considering our Bayes estimates under absolute error loss and squared error loss are 19.06 and 19.09 respectively and the posterior predictive distribution above, the posterior is reasonable for the data set observed, so it is a reasonable model overall.