

STA 602 HW 3

William Tirone

1 (3.3)

a)

The below is possible because of independence of the priors. Thus, the posteriors are also independent since the priors and sampling models are independent as well.

$$\begin{aligned}p(\theta_A|y_1, \dots, y_{10}) &\propto p(\theta_A)p(y_1, \dots, y_{10}|\theta_A) \\&= c(y, a, b) \times \theta^{121}e^{-10\theta} \times \theta^{117}e^{-10\theta} \\&= c(y, a, b) \times \theta^{238}e^{-20\theta} \\&\sim \text{gamma}(239, 20) \\mean &= 239/20 = 11.95 \\var &= 239/400 = 0.5975\end{aligned}$$

from the above we can generalize to:

$$\begin{aligned}p(\theta|y_1, \dots, y_n) &\sim \text{gamma}(a + \sum_{i=1}^n Y_i, b + n) \\p(\theta_B|y_1, \dots, y_{13}) &\sim \text{gamma}(12 + 113, 1 + 13) = \text{gamma}(125, 14) \\mean &= 125/14 = 8.93 \\var &= 125/196 = 0.638\end{aligned}$$

```
cat('95% Interval for GAM(237,20) : ', qgamma(.025,293,20), qgamma(.975,293,20))
```

```
95% Interval for GAM(237,20) : 13.02039 16.3743
```

```
cat('95% Interval for GAM(125,14) : ', qgamma(.025,125,14), qgamma(.975,125,14))
```

```
95% Interval for GAM(125,14) : 7.432064 10.56031
```

b)

Based on the graph below, you would need a large value for α in your prior for $E(\theta_B|y_B)$ to be close to $E(\theta_A|y_A)$ - I'm not sure if there is an "intuitive" interpretation of this parameter, but if $E(\theta_B|y_B)$ was close to $E(\theta_A|y_A)$, that would mean y_B was a very unusual observation and had tumor counts much lower than expected.

```
# is posterior expectation just E(\theta | y)
n0 = seq(1,500,1)
yB = c(11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)
n = length(yB)
sum_yB = sum(yB)
means = c()

for (i in n0){

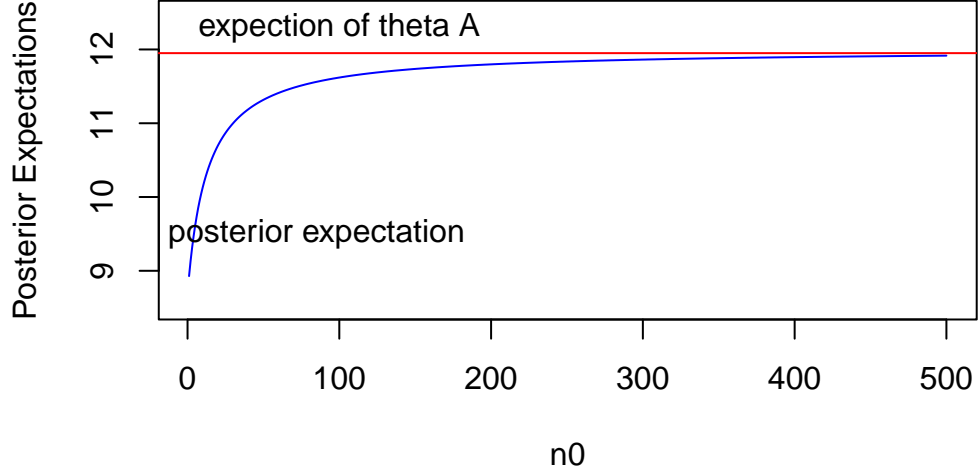
  # prior
  alpha = 12 * i
  beta = i

  # posterior
  posterior_alpha = alpha + sum_yB
  posterior_beta = beta + n

  posterior_expectation = posterior_alpha / posterior_beta
  means = append(means,posterior_expectation)

}

plot(n0,means,type='l',
     ylim=c(8.5,12.5),
     ylab='Posterior Expectations',
     col='blue')
abline(h=11.95 ,col='red')
text(85,9.5, "posterior expectation")
text(100,12.3, "expection of theta A")
```



c)

“Tumor count rates for type B mice are unknown, but type B mice are related to type A mice.” This seems to indicate that it does not make sense to have $p(\theta_A, \theta_B) = p(\theta_B) \times p(\theta_A)$ since that implies independence. So it seems like studies of the mice in group A could influence our priors for group B.

2 (3.5)

a)

$$\begin{aligned}
 p(\theta|y) &\propto p(y|\theta)\tilde{p}(\theta) \\
 iid &\propto \prod_{i=1}^n c(\theta)h(y_i)e^{\theta t(y_i)} \sum_{i=1}^k w_k p_k(\theta) \\
 &\propto c(\theta)^n e^{\theta \sum t(y_i)} \sum_{i=1}^k w_k k(n_0, t_0) c(\theta)^{n_0} e^{n_0 t_0 \theta} \\
 &\propto \sum_{i=1}^k w_k k(n_0, t_0) c(\theta)^{n+n_0} e^{\theta(n_0 t_0 + \sum y_i)} \\
 &\propto \sum_{i=1}^k w_k k(n_0, t_0) p(\theta|n_0 + n, (n_0 t_0 + \sum y_i))
 \end{aligned}$$

b)

$$\begin{aligned}
 p(\theta|y_1, \dots, y_n) &\propto \tilde{p}(\theta)p(y_1, \dots, y_n|\theta) \\
 iid &\propto \sum_{k=1}^K w_k p_k(\theta) \prod_{i=1}^n \frac{e^{\theta y_i}}{y_i!} \\
 &\propto \sum_{k=1}^K w_k p_k(\theta) \frac{e^{n\theta \sum y_i}}{\prod y_i!} \\
 &\propto c(y, \alpha, \beta) \sum_{k=1}^K w_k \theta^{\sum y_i + \alpha - 1} e^{-\theta(\beta + n)} \\
 &\propto \sum_{k=1}^K w_k * \text{gamma}(\sum y_i + \alpha, n + \beta)
 \end{aligned}$$

3 (3.8)

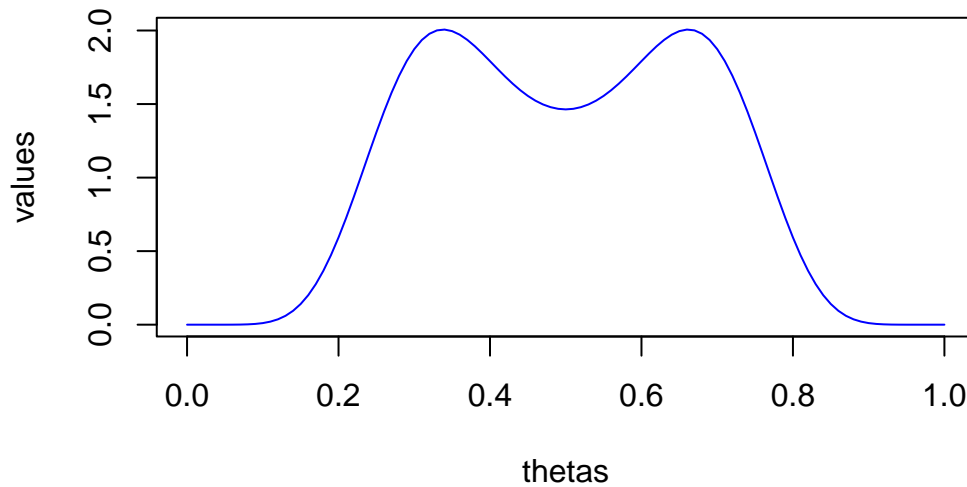
a)

For the parameters, I assumed that we had previously observed 30 total flips. Thus, the expectation of the first part of the mixture is 1/3, the 20% portion is 1/2, and the last piece has an expectation of 2/3.

```

thetas = seq(0,1,0.01)
values = (.4 * dbeta(thetas,10,20)) + (.2 * dbeta(thetas,15,15)) + (.4 * dbeta(thetas,20,
plot(thetas,values,type='l',col='blue')

```



b)

U.S. Quarter, 2020, 27 heads, 23 tails

c)

$$\begin{aligned}
 P(\theta|y) &\propto P(y|\theta)P(\theta) \\
 &\propto \binom{50}{27} \theta^{27} (1-\theta)^{23} \left[\frac{4}{10} \frac{\Gamma(30)}{\Gamma(10)\Gamma(20)} \theta^{10-1} (1-\theta)^{20-1} + \frac{2}{10} \frac{\Gamma(30)}{\Gamma(15)\Gamma(15)} \theta^{15-1} (1-\theta)^{15-1} + \frac{4}{10} \frac{\Gamma(30)}{\Gamma(10)\Gamma(20)} \theta^{20-1} (1-\theta)^{10-1} \right] \\
 &\propto \binom{50}{27} \left[.4 \frac{\Gamma(30)}{\Gamma(10)\Gamma(20)} \theta^{36} (1-\theta)^{42} + .2 \frac{\Gamma(30)}{\Gamma(15)\Gamma(15)} \theta^{44} (1-\theta)^{42} + .4 \frac{\Gamma(30)}{\Gamma(10)\Gamma(20)} \theta^{46} (1-\theta)^{32} \right] \\
 &\propto \binom{50}{47} \left[.4 \frac{\Gamma(30)}{\Gamma(10)\Gamma(20)} \frac{\Gamma(37)\Gamma(43)}{\Gamma(37+43)} \text{Beta}(37, 43) \right. \\
 &\quad + .2 \frac{\Gamma(30)}{\Gamma(15)\Gamma(15)} \frac{\Gamma(42)\Gamma(43)}{\Gamma(42+43)} \text{Beta}(42, 43) \\
 &\quad \left. + .4 \frac{\Gamma(30)}{\Gamma(10)\Gamma(20)} \frac{\Gamma(47)\Gamma(33)}{\Gamma(47+33)} \text{Beta}(47, 33) \right]
 \end{aligned}$$

The weights were found below by multiplying the binomial sampling model from part b with the weighted priors from part a (this is displayed above, I worked this out by hand and it was quite lengthy). Then, with the kernel trick and dividing by the normalizing constant, we get weights that sum to 1. Then, the posterior is now weighted with respect to the different distributions below.

```

weight_1 = choose(50,27) * .4 * (1/beta(10,20)) * beta(37,43)
weight_2 = choose(50,27) * .2 * (1/beta(15,15)) * beta(42,43)
weight_3 = choose(50,27) * .4 * (1/beta(20,10)) * beta(47,33)

```

```

total = weight_1 + weight_2 + weight_3

```

```

weight_1 = weight_1 / total
weight_2 = weight_2 / total
weight_3 = weight_3 / total

```

```

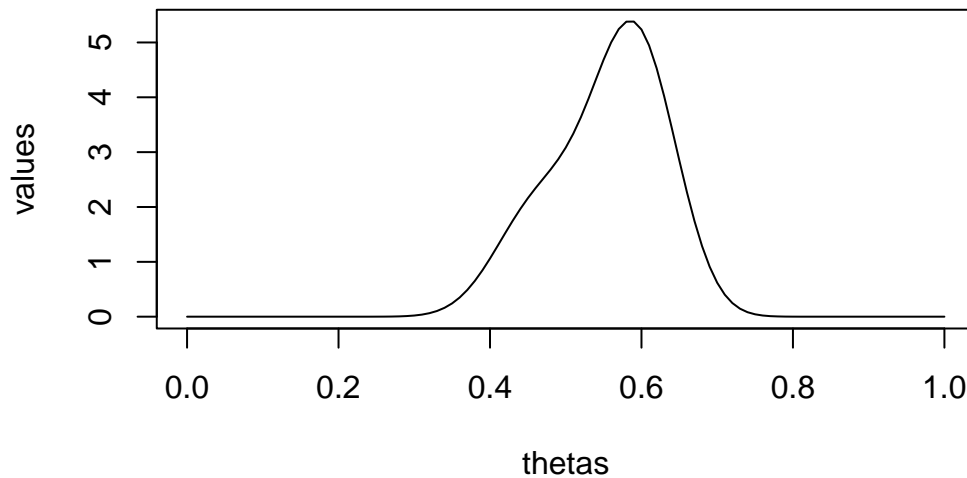
values = weight_1 * dbeta(thetas,37,43) + weight_2 * dbeta(thetas,42,43) + weight_3 * dbeta(thetas,47,33)

```

```

plot(thetas,values,type='l')

```



d)

1 Euro coin, 2021, 20 heads, 30 tails.

Below, I simply repeated the exercise but for a different number of heads obtained on 50 spins. I chose NOT to use a different prior because I think it's interesting seeing how the plot changes based on the same prior but using a different set of observed data from the sampling model. I also used a different country's currency than the first choice, so they are more likely to be independent and thus would not influence this prior.

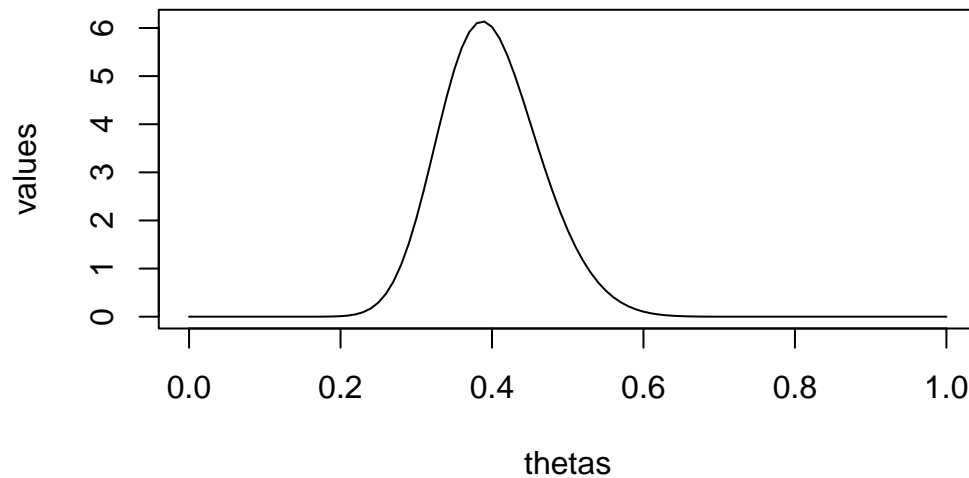
While above the data shifted towards the third prior, around $\theta = 0.6$, the data below shifted towards the "first" prior - getting a head 1/3 of the time. It's a noticeable change because 50 observations is more than the observations from the prior, thus the posterior dominates the thetas from the prior.

```
weight_1 = choose(50,20) * .4 * (1/beta(10,20)) * beta(30,50)
weight_2 = choose(50,20) * .2 * (1/beta(15,15)) * beta(35,45)
weight_3 = choose(50,20) * .4 * (1/beta(20,10)) * beta(40,40)
total = weight_1 + weight_2 + weight_3

weight_1 = weight_1 / total
weight_2 = weight_2 / total
weight_3 = weight_3 / total

values = weight_1 * dbeta(thetas,30,50) + weight_2 * dbeta(thetas,35,45) + weight_3 * dbet

plot(thetas,values,type='l')
```



4 (4.3)

a)

Based on the statistic $\text{mean}(\text{data}) / \text{sd}$ and comparing the posterior samples to the observed data, this looks pretty good. It looks like the mean of the posterior samples is slightly lower than our observed data, but on average it's probably quite good.

```
yA = c(12, 9, 12, 14, 13, 13, 15, 8, 15, 6)
A_statistic = mean(yA) / sd(yA)

n1 = 10 # sample size

t.mc=NULL

for ( s in 1:1000) {

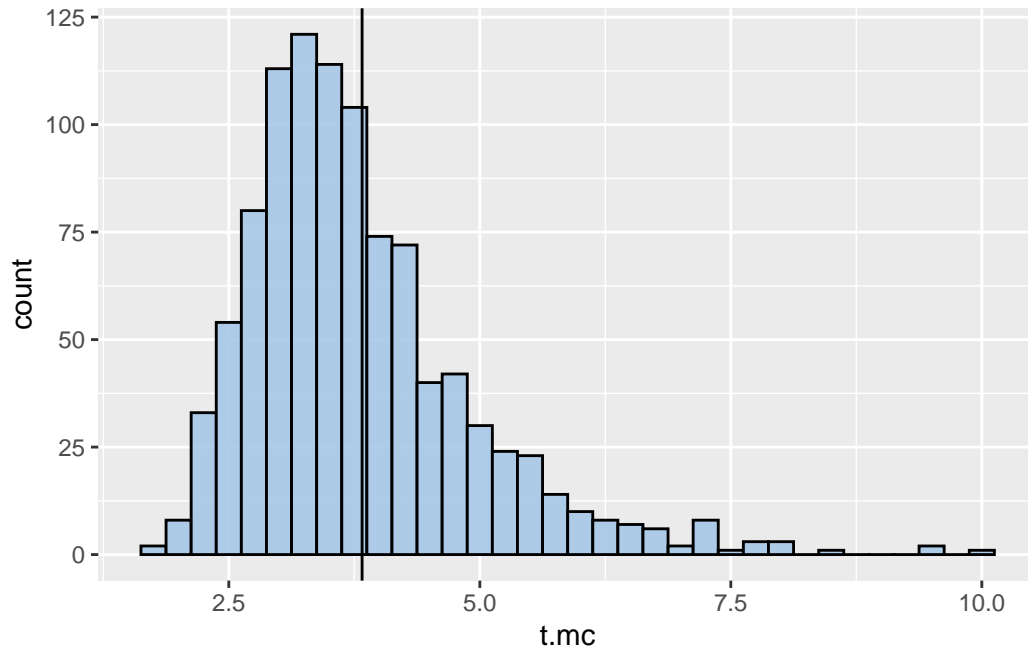
# theta1 is sample from the posterior, gam(239,20)
theta1 = rgamma(1,239,20) #samples from posterior

y1.mc = rpois(n1 , theta1) #dataset

t.mc = c(t.mc, mean(y1.mc) / sd(y1.mc))
}

df = as.data.frame(t.mc)
```

```
df |> ggplot(aes(x=t.mc)) +
  geom_histogram(binwidth=0.25, fill="#A7C7E7",color='black',alpha=0.9) +
  geom_vline(xintercept= A_statistic)
```



b)

Purely based on the statistic $\text{mean}(\text{data}) / \text{sd}$ and comparing that to the statistic from our posterior predictive datasets, it looks like a very poor fit, most likely because the original data was not actually independent.

```
yB = c(11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)
B_statistic = mean(yB) / sd(yB)

n1 = 13 # sample size

t.mc2=NULL

for ( s in 1:1000) {

  # theta1 is sample from the posterior, gamma(125,14)
  theta1 = rgamma(1,125,14) #samples from posterior
```



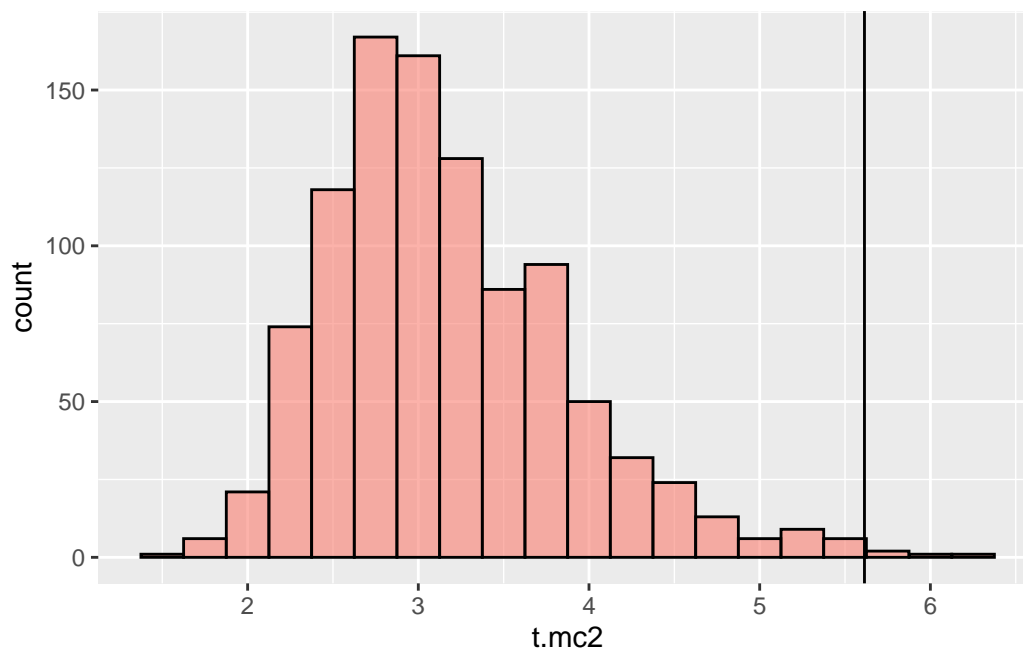
```

y1.mc = rpois(n1 , theta1) #dataset

t.mc2 = c(t.mc2, mean(y1.mc) / sd(y1.mc))
}

df = as.data.frame(t.mc2)
df |> ggplot(aes(x=t.mc2)) +
  geom_histogram(binwidth=0.25, fill="salmon",color='black',alpha=0.6) +
  geom_vline(xintercept= B_statistic)

```



5

a)

We know from lecture / past homework problems that a gamma prior is conjugate with the a poisson sampling model:

$$\begin{aligned}
 p(\theta|x) &\propto \text{gamma}(\sum x_i + \alpha, n + \beta) \\
 &\text{with the data given in the problem:} \\
 &\propto \text{gamma}(210, 11)
 \end{aligned}$$

With squared error loss, the Bayes estimate is $E(\theta|x)$, the posterior mean. Thus:

$$E(\theta|x) = \frac{210}{11} = 19.09$$

With absolute error loss, the Bayes estimate is the posterior median:

```
# posterior median
qgamma(.5,210,11)
```

```
[1] 19.06061
```

b)

Intuitively, we may violate the poisson assumption if the call center has some kind of seasonality, whether yearly or through the day / week / month. For example, they may receive more calls after 5 P.M. when people are off work, so it's hard to say in this scenario when we just have total calls over the number of days.

However, in our case, if we simulate data based on the posterior, our mean is not far off from the center of the means of the simulated data sets. With that in mind, in our example in a vacuum, it seems like the model is at least a decent fit.

```
statistic = 20
n1=10

t.mc3=NULL

for ( s in 1:1000) {

  # theta1 is sample from the posterior, gam(239,20)
  theta1 = rgamma(1,210,11) # samples from posterior

  y1.mc = rpois(n1 , theta1) # fake dataset

  t.mc3 = c(t.mc3, mean(y1.mc))
}

df = as.data.frame(t.mc3)
df |> ggplot(aes(x=t.mc3)) +
  geom_histogram(binwidth=0.25, fill="purple",color='blue',alpha=0.6) +
  geom_vline(xintercept= statistic)
```

