Assignment #2

Elliot Smith

8/30/2018

Problem 1

Part a

Here our goal will be to minimize a to show that $a = E[\theta|y]$ is the unique Bayes estimate of θ:

$$\begin{aligned}
\frac{d}{da}E[L(a|y)] &= \frac{d}{da}\int L(\theta, a)p(\theta|y)d\theta \\
&= \frac{d}{da}\int (\theta - a)^2 p(\theta|y)d\theta \\
&= -2\int (\theta - a)p(\theta|y)d\theta \\
&= -2\left[\int \theta p(\theta|y)d\theta - a\int p(\theta|y)d\theta\right] \\
&= -2\big[E[\theta|y] - a\big]
\end{aligned}$$

$-2\big[E[\theta|y] - a\big] = 0$ when $a = E[\theta|y]$

To prove that it is a unique minimizing statistic, we must look at the second derivative:

$\frac{d}{da}(-2\big[E[\theta|y] - a\big]) = 2$

As $2 > 0$, this shows that it is a unique minimzing statistic.

Part b

Here our goal will be to show that for any median value of a, the derivative of $L(\theta, a)$ will evaluate to 0.

$$\begin{aligned}
\frac{d}{da}\big[E[L(a|y)]\big] &= \frac{d}{da}\left[\int_{-\infty}^{a}(a - \theta)p(\theta|y)d\theta + \int_{a}^{\infty}(\theta - a)p(\theta|y)d\theta\right] \\
&= \int_{-\infty}^{a}\frac{d}{da}(a - \theta)p(\theta|y)d\theta + \int_{a}^{\infty}\frac{d}{da}(\theta - a)p(\theta|y)d\theta \\
&= \int_{-\infty}^{a}p(\theta|y)d\theta + \int_{a}^{\infty}(-1)p(\theta|y)d\theta \\
&= \int_{-\infty}^{a}p(\theta|y)d\theta - \int_{a}^{\infty}p(\theta|y)d\theta \\
&= \frac{1}{2} - \frac{1}{2} \\
&= 0
\end{aligned}$$

As a result, it has been shown that any posterior median of θ is a Bayes estimate of θ.

Taking the second derivative we again get a positive number, thus again indicating that it is a minimizing statistic.

Part c

Here our goal will be to show that for any value of a, the derivative of $L(\theta, a)$ will evaluate to 0 where k_0 and k_1 are nonnegative numbers.

$$\begin{aligned}
\frac{d}{da}\big[E[L(a|y)]\big] &= \frac{d}{da}\left[\int_{-\infty}^{a} k_1(a-\theta)p(\theta|y)d\theta + \int_{a}^{\infty} k_0(\theta-a)p(\theta|y)d\theta\right] \\
&= \int_{-\infty}^{a} \frac{d}{da}k_1(a-\theta)p(\theta|y)d\theta + \int_{a}^{\infty} \frac{d}{da}k_0(\theta-a)p(\theta|y)d\theta \\
&= \int_{-\infty}^{a} k_1 p(\theta|y)d\theta + \int_{a}^{\infty} (-k_0)p(\theta|y)d\theta \\
&= \int_{-\infty}^{a} k_1 p(\theta|y)d\theta - \int_{a}^{\infty} k_0 p(\theta|y)d\theta \\
&= k_1\int_{-\infty}^{a} p(\theta|y)d\theta - k_0\int_{a}^{\infty} p(\theta|y)d\theta
\end{aligned}$$

Noting that: $k_0\int_a^{\infty} p(\theta|y)d\theta = k_0 - k_0\int_{-\infty}^{a} p(\theta|y)d\theta$

$$\begin{aligned}
k_1\int_{-\infty}^{a} p(\theta|y)d\theta - k_0\int_{a}^{\infty} p(\theta|y)d\theta &= k_1\int_{-\infty}^{a} p(\theta|y)d\theta - \left[k_0 - k_0\int_{-\infty}^{a} p(\theta|y)d\theta\right] \\
&= k_1\int_{-\infty}^{a} p(\theta|y)d\theta + k_0\int_{-\infty}^{a} p(\theta|y)d\theta - k_0 \\
&= (k_1+k_0)\int_{-\infty}^{a} p(\theta|y)d\theta - k_0
\end{aligned}$$

Now setting $\int_{-\infty}^{a} p(\theta|y)d\theta = \frac{k_0}{k_0+k_1}$ we get our result that any quantile is a Bayes estimate of θ.

Taking the second derivative we again get a positive number, thus again indicating that it is a minimizing statistic.

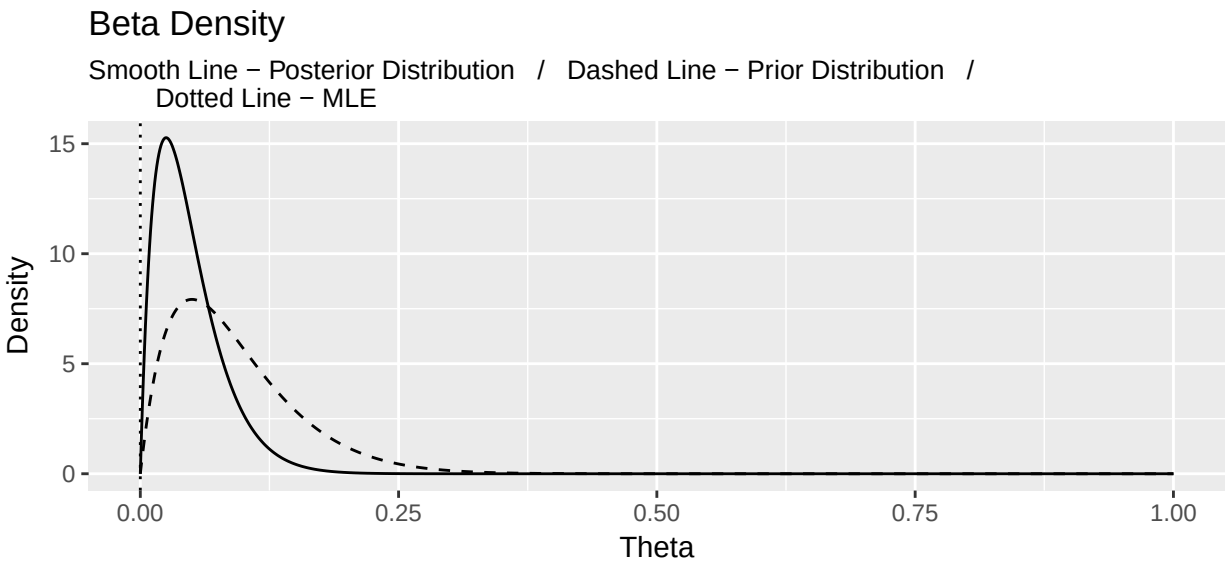
Problem 2

n = 20
Sampling Distribution: $y|\theta \sim Binomial(n = 20, \theta)$
Prior Distribution: $\theta \sim Beta(\alpha = 2, \beta = 20)$
Posterior Distribution:

$$\begin{aligned}
p(\theta|y) &= \binom{n}{y}\theta^y(1-\theta)^{n-y} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&= \binom{n}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} \\
&\propto Beta(y+\alpha, (n-y)+\beta) \\
&= Beta(y+2, (20-y)+20)
\end{aligned}$$

Part i

$y = 0$



From this graph we can see that while the MLE, $\hat{y} = \frac{y}{n}$ is not properly represented the true distribution due to the fact that there were no positive cases in this hospital, the posterior distribution, via the affect of the prior distribution, is impacting our results enough to render a more accurate result.

Part ii

```
## Posterior Credible Interval:  ( 0.002261283 , 0.1111897 )
```

```
## Frequentist Interval:  ( 0 , 0 )
```

From this result, we can see that the Frequestist result compares poorly ($(0, 0)$ is not really an interval at all) against the Bayesian posterior credible interval of θ. Again, the frequentist approach for this hospital is being adversely affected by the fact that there are no positive cases, and as such, we do not get a good representation of the true distribution of positive cases. Whereas the Bayesian approach factors in our prior beliefs on the number of positive cases, thus allowing us to interpret the results for this case.

Part iii

```
## HPD Interval:  ( -4.872209 , -1.793283 )
```

Problem 3

Part i

My methodology for selecting my α and β for my $\theta_1, \theta_2 \sim Gamma(\alpha, \beta)$ distribution is as follows. First, using a Pew Research Center on Social & Demographic Trends article on average number of children per parent (http://www.pewsocialtrends.org/2015/05/07/family-size-among-mothers/), I deduced that the average number of children per parent for women who were in their 20s during the 1970s is 2.28, while the variance is 0.8905. These values were obtained by the results from the poll. Now that I had values for the prior mean and variance, I was able to algebraically solve for α and β via the following formulae: $EX = \frac{\alpha}{\beta}$ and $VarX = \frac{\alpha}{\beta^2}$.

```
## Gamma Parameters

## Alpha Parameter:  5.8376

## Beta Parameter:  2.5604
```

Part ii

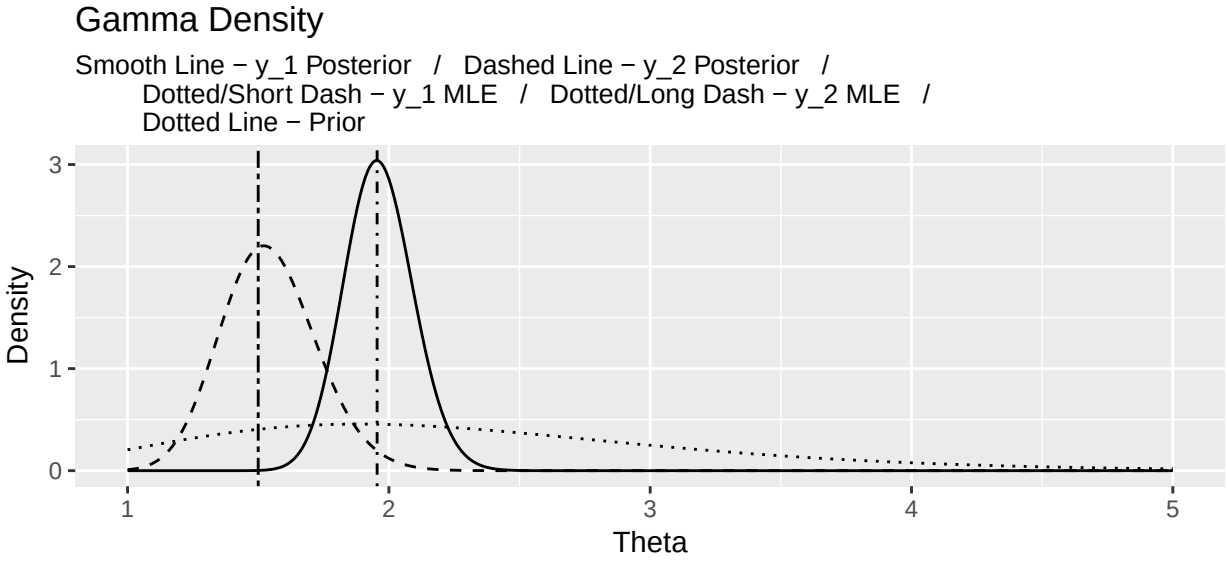
Posterior Distributions:

$$\begin{aligned} p(\theta|y) &= p(\theta_1, \theta_2|y_1, y_2) \\ &\propto p(y_1|\theta_1)p(y_2|\theta_2)p(\theta_1)p(\theta_2) \\ &= \frac{{\theta_1}^{y_1}e^{-\theta_1}}{y_1!}\frac{{\theta_2}^{y_2}e^{-\theta_2}}{y_2!}\frac{\beta^\alpha}{\Gamma(\alpha)}{\theta_1}^{\alpha-1}e^{-\beta\theta_1}\frac{\beta^\alpha}{\Gamma(\alpha)}{\theta_2}^{\alpha-1}e^{-\beta\theta_2} \\ &\propto {\theta_1}^{y_1}e^{-\theta_1}{\theta_2}^{y_2}e^{-\theta_2}{\theta_1}^{\alpha-1}e^{-\beta\theta_1}{\theta_2}^{\alpha-1}e^{-\beta\theta_2} \\ &= \theta_1^{y_1+\alpha-1}\theta_2^{y_2+\alpha-1}e^{-[\theta_1(1+\beta)+\theta_2(1+\beta)]} \end{aligned}$$

Since we may assume iid priors, we get:

$p(\theta_1|y_1) \sim Gamma(y_1 + \alpha, n_1 + \beta)$

$p(\theta_2|y_2) \sim Gamma(y_2 + \alpha, n_2 + \beta)$



As we can see from the plot, we have a fairly weakly informative prior; its very hard to tell whether or not our prior distribution is influencing either of our posterior distributions in a meaningful way. Our MLE mean results appear to align very well with our posterior distributions.

Part iii

```
## Posterior Credible Interval, theta_1:  ( 1.707723 , 2.22927 )
```

```
## Posterior Credible Interval, theta_2:  ( 1.187905 , 1.89459 )
```

As we can see from our results, the two intervals do overlap! This would lead us to believe that it is not a forgone conclusion that a mother from group y_1 would have more children than a mother from group y_2; though our intervals imply than we can be more confident that a mother in group y_1 will have more children on average.

Part iv

```
## Posterior Credible Interval, theta_1 - theta_2:  ( -0.02385043 , 0.8436399 )
```

As we can see from our results, the interval of $\theta_1 - \theta_2$ does include zero! This reinforces our previous point, that a randomly chosen mother from y_2 may have more children than a randomly chosen mother from y_1, though it is significantly less likely than the other way around.

Part v

```
## Posterior Probability:  0.9688
```

This provided value dictates the likelihood that a value falls within a particular range in the distribution. In this case, the provided value states how likely that the true answer lies in the range, which in this case, is greater than zero. Thus, in this case of birth rates, we can be very confident that an individual mother from y_1 will have more children than a mother from y_2, however, it is still possible to randomly choose a case where a mother from y_2 would have more children than a mother randomly chosen from y_1.

The difference between the classical Frequensist approach of a hypothesis test and our Bayesian approach is that the Frequentists believe that our statistic in question is an unknown, but fixed, number, whereas Bayesians believes that it is an unknown distribution. As a result, the Bayesian approach allows us to say whether or not a true answer lies in a given range, while the Frequentist approach relies heavily on weakly defined constraints ("data more extreme") to determine the probability that we are making an error when accepting our Null Hypothesis as true (such as whether a parameter equals a pre-defined value).

Code Appendix

```
### Load in the necessary packages

library(ggplot2)
library(coda)



##### Problem 2 #####



## Set the parameter values

y <- 0
n <- 20
alpha <- 2
beta <- 20

## Set the theta values

theta <- seq(from = 0, to = 1, length.out = 1000)



### Part i ###



## Plot the posterior distribution

post_beta_dens <- dbeta(x = theta, shape1 = y + alpha, shape2 = (n - y) + beta)

ggplot() +
    geom_line(aes(x = theta, y = post_beta_dens), linetype = 1) +
    geom_line(aes(x = theta, y = dbeta(x = theta, shape1 = alpha, shape2 = beta)), linetype = 2) +
    geom_vline(aes(xintercept = (y / n)), linetype = 3) +
    labs(x = "Theta", y = "Density", title = "Beta Density",
         subtitle = "Smooth Line - Posterior Distribution   /   Dashed Line - Prior Distribution   /
         Dotted Line - MLE")



### Part ii ###



## Draw a sample from my posterior distribution

post_beta_rand <- rbeta(n = 10000, shape1 = y + alpha, shape2 = (n - y) + beta)

## Compute the HPD interval
```

```r
bayes_int <- as.vector(HPDinterval(as.mcmc(post_beta_rand), prob = 0.95)[1, 1:2])

## Calculate the MLE

y_hat <- (y / n)

## Compute the frequentist interval

freq_int <- c((y_hat) - (1.96 * (sqrt(y_hat * (1 - y_hat)) / 2)),
              (y_hat) + (1.96 * (sqrt(y_hat * (1 - y_hat)) / 2)))



### Part iii ###



## Compute the log odds of our posterior beta distrbution

odds_samp <- log(post_beta_rand / (1 - post_beta_rand))

## Compute the HPS interval

bayes_int_odds <- as.vector(HPDinterval(as.mcmc(odds_samp), prob = 0.95)[1, 1:2])



##### Problem 3 #####



### Part i ###



## Sample mean and variance

samp_mean <- ((1 * 21) + (2 * 43) + (3 * 23) + (4 * 13)) / (21 + 43 + 23 + 13)
samp_var <- var(c(rep(1, 21), rep(2, 43), rep(3, 23), rep(4, 13)))

## Compute the gamma parameters

find_gamma_params <- function(mu, sigma_sq) {

    beta <- mu / sigma_sq
    alpha <- beta * mu
    return(c(alpha, beta))

}

gamma_params <- round(find_gamma_params(samp_mean, samp_var), 4)
```

```
### Part ii ###



## Set theta values

theta <- seq(from = 1, to = 5, by = 0.01)

## Set the sample values

y_1 <- 217
y_2 <- 66
n_1 <- 111
n_2 <- 44
lambda_1 <- 217 / 111
lambda_2 <- 66 / 44

## Generate samples from the posterior distributions

post_dens_y1 <- dgamma(x = theta, shape = y_1 + gamma_params[1], rate = gamma_params[2] + n_1)
post_dens_y2 <- dgamma(x = theta, shape = y_2 + gamma_params[1], rate = gamma_params[2] + n_2)
prior_dens <- dgamma(x = theta, shape = gamma_params[1], rate = gamma_params[2])

## Generate the plot

ggplot() +
    geom_line(aes(x = theta, y = post_dens_y1), linetype = 1) +
    geom_line(aes(x = theta, y = post_dens_y2), linetype = 2) +
    geom_line(aes(x = theta, y = prior_dens), linetype = 3) +
    geom_vline(aes(xintercept = lambda_1), linetype = 4) +
    geom_vline(aes(xintercept = lambda_2), linetype = 6) +
    labs(x = "Theta", y = "Density", title = "Gamma Density",
         subtitle = "Smooth Line - y_1 Posterior   /   Dashed Line - y_2 Posterior   /
         Dotted/Short Dash - y_1 MLE   /   Dotted/Long Dash - y_2 MLE   /
         Dotted Line - Prior")



### Part iii ###



## Draw a sample from each posterior distribution

post_gamma_y1 <- rgamma(n = 10000, shape = y_1 + gamma_params[1], rate = n_1 + gamma_params[2])
post_gamma_y2 <- rgamma(n = 10000, shape = y_2 + gamma_params[1], rate = n_2 + gamma_params[2])

## Compute the HPD intervals

hpd_int_y1 <- as.vector(HPDinterval(as.mcmc(post_gamma_y1), prob = 0.95)[1, 1:2])
hpd_int_y2 <- as.vector(HPDinterval(as.mcmc(post_gamma_y2), prob = 0.95)[1, 1:2])

## Output the results
```

```r
cat("Posterior Credible Interval, theta_1: ", "(", hpd_int_y1[1], ",", hpd_int_y1[2], ")")
cat("Posterior Credible Interval, theta_2: ", "(", hpd_int_y2[1], ",", hpd_int_y2[2], ")")



### Part iv ###



## Get the difference of our random samples

post_diff_gamma <- post_gamma_y1 - post_gamma_y2

## Compute the HPD interval

diff_int <- as.vector(HPDinterval(as.mcmc(post_diff_gamma), prob = 0.95)[1, 1:2])

## Output the results

cat("Posterior Credible Interval, theta_1 - theta_2: ", "(", diff_int[1], ",", diff_int[2], ")")



### Part v ###



post_prob <- mean(post_diff_gamma > 0)
```