Lab 2

Harry Bendekgey and Peter Brody-Moore
Math 153: Bayesian Statistics

February 15, 2018

Problem 1

We are looking to find an α and β that satisfies the scientist's prior. We first wrote the function to find alpha and beta for any inputted best guess and probability bounds (part b) and then used it to find the answer for part a. The values we found were $\alpha = 1.6$ and $\beta = 14.4$.

The function we wrote uses a while loop to check α levels using a step of 0.1. If the α and β levels inputted into the pbeta function are less than or equal to the confidence interval, then alpha and beta are returned. Else, it continues stepping until it finds a suitable combination, or reaches the inputted maximum level to be checked. This maxCheck is to ensure the while loop isn't infinite.

```
# function to find alpha and beta
find_alpha <- function(guess, bound,
                       confidence_level, maxCheck) {
  a <- 1;
  while(a < maxCheck) {
      if (pbeta(bound,a,9*a) <= confidence_level &&
          pbeta(bound,a+.1,9*(a+.1)) > confidence_level) {
            return(c(a,a*9))
      } else {
          a <- a + .1
      }
  }
  return(0);
}

# Find alpha and beta for part A
alphaBeta <- find_alpha(guess = 0.1, bound = 0.2,
                        confidence_level = .9, maxCheck = 100)
alphaBeta

## [1]  1.6 14.4
```

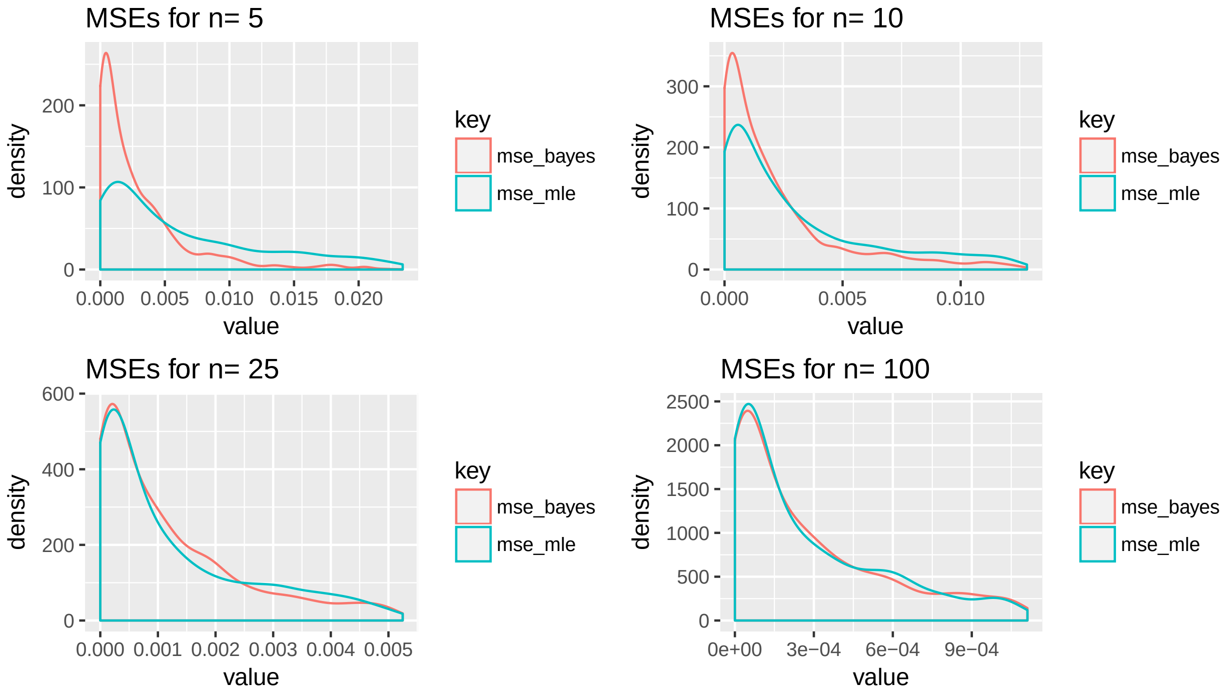
Problem 2

a) In Problem 2, we first compare the MSE of the posterior mean over the lifetime of estimation to the MSE of the freqentist MLE across various sample sizes. Looking at the graphs below, you can see that for the low values of n, the MSE of the posterior mean is lower on average than the MSE of the frequesntist MLE (densities are plotted, and the MSE of the posterior mean is highly clustered close to zero). Although this means that the posterior mean won't always be better than the MLE over the lifetime of the scientist, it is better more often. As n increases the two densities are relatively equivalent because of the asymptotic properties of the two estimators (MSE goes to zero as n goes infinity).

```r
set.seed(474747)
mse_mle <- c()
mse_bayes <- c()

compare_mse <- function(n,a=1.6,b=14.4) {
  for (i in 1:1000) {
    p <- rbeta(1,a,b)
    x <- sum(rbinom(n=n, size=1,prob=p))
    mse_mle[i] <- ((x / n) - p)^2
    mse_bayes[i] <- ((x+a)/(n+a+b) - p)^2
  }
  df <- tidyr::gather(data.frame(mse_mle,mse_bayes))
  return (ggplot(df, aes(x=value, color=key)) +
            geom_density() +
            scale_x_continuous(limits = c(0,mean(mse_mle) * 1.5))+
            labs(title=paste("MSEs for n=", n)))
}
```

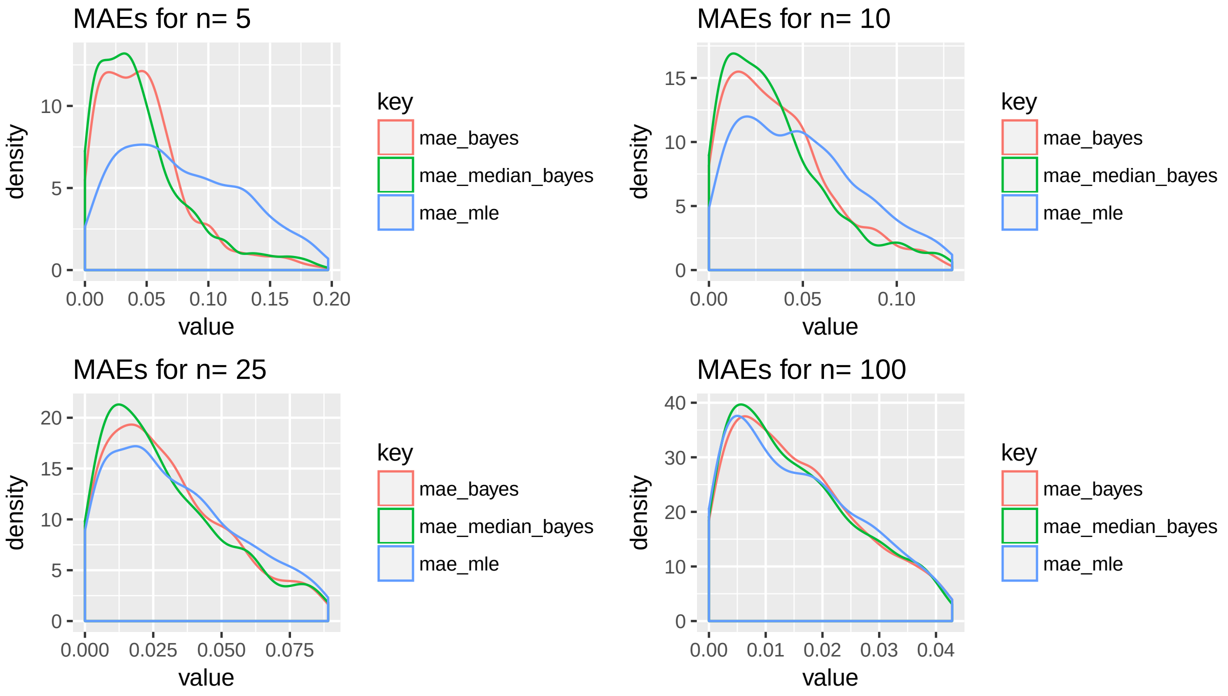
```r
ggarrange(compare_mse(5),compare_mse(10),
          compare_mse(25),compare_mse(100))
```



b) In part b, we do a relatively similar analysis to part a, except we add in the posterior median and use the mean absolute error as a metrics of comparison. In this case, the posterior mean still outperforms the MLE for low values of n in regards to absolute error. However, the posterior median outperforms both of these because its density is clustered most closely to zero for the lower values of n. Sometimes the posterior mean and MLE will be the best in the scientist's lifetime, but the posterior median will be the best most often. All three still converge to relatively equal densities for large values of n becuase of the asymptotic properties.

```r
mae_mle <- c()
mae_bayes <- c()
mae_median_bayes <- c()
compare_mae <- function(n,a=1.6,b=14.4) {
  for (i in 1:1000) {
    p <- rbeta(1,a,b)
    x <- sum(rbinom(n=n, size=1,prob=p))
    mae_mle[i] <- abs((x / n) - p)
    mae_bayes[i] <- abs((x+a)/(n+a+b) - p)
    mae_median_bayes[i] <- abs((a - 1/3 + x)/(a + b - 2/3 + n) - p)
  }
  df <- tidyr::gather(data.frame(
                      mae_mle,mae_bayes,mae_median_bayes))
  return (ggplot(df, aes(x=value, color=key)) +
            geom_density() +
            scale_x_continuous(limits = c(0,mean(mae_mle) * 2)) +
            labs(title=paste("MAEs for n=", n)))
}
```

```r
ggarrange(compare_mae(5),compare_mae(10),
          compare_mae(25),compare_mae(100))
```



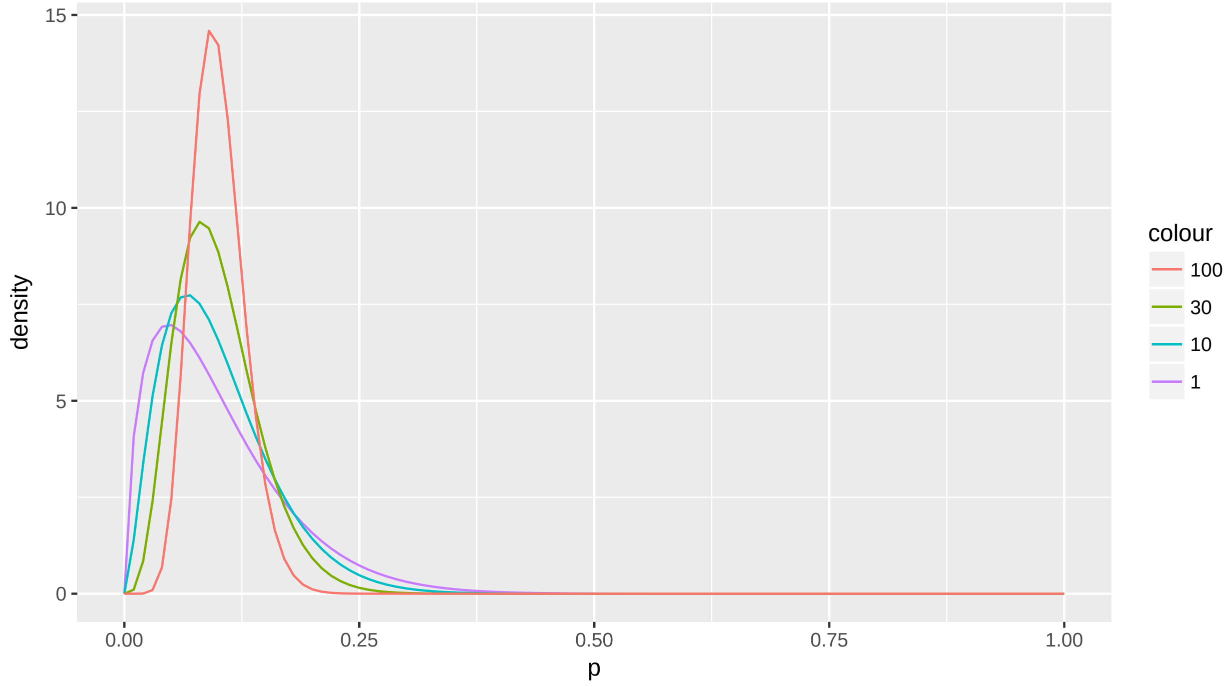
Problem 3

```
post_dist <- function(p_hat, n, alpha, beta) {
  return(function(x) dbeta(x,alpha+p_hat*n, beta+n-p_hat*n))
}

graph_posts <- function(p_hat, alpha = 1.6, beta = 14.4) {
  plot <- ggplot(data.frame(x = c(0, 1)), aes(x)) +
    stat_function(fun = post_dist(p_hat, 1, alpha, beta),
                  aes(col="red")) +
    stat_function(fun = post_dist(p_hat, 10, alpha, beta),
                  aes(col="orange")) +
    stat_function(fun = post_dist(p_hat, 30, alpha, beta),
                  aes(col="green")) +
    stat_function(fun = post_dist(p_hat, 100, alpha, beta),
                  aes(col="blue")) +
    scale_color_discrete(labels = c("100","30","10","1")) +
    labs(x = "p", y = "density")
  return (plot)
}
```

We are interested in the evolution of our posterior density as we see more data. Our prior, the Beta distribution, has a nice property: all posteriors will also be Beta distributions. We begin by choosing a smart prior ($\alpha = 1.6, \beta = 14.4$ for $p = .1$, the same prior that matched the scientist's expectations earlier in this lab) and see how our posterior density changes as we see more and more data.

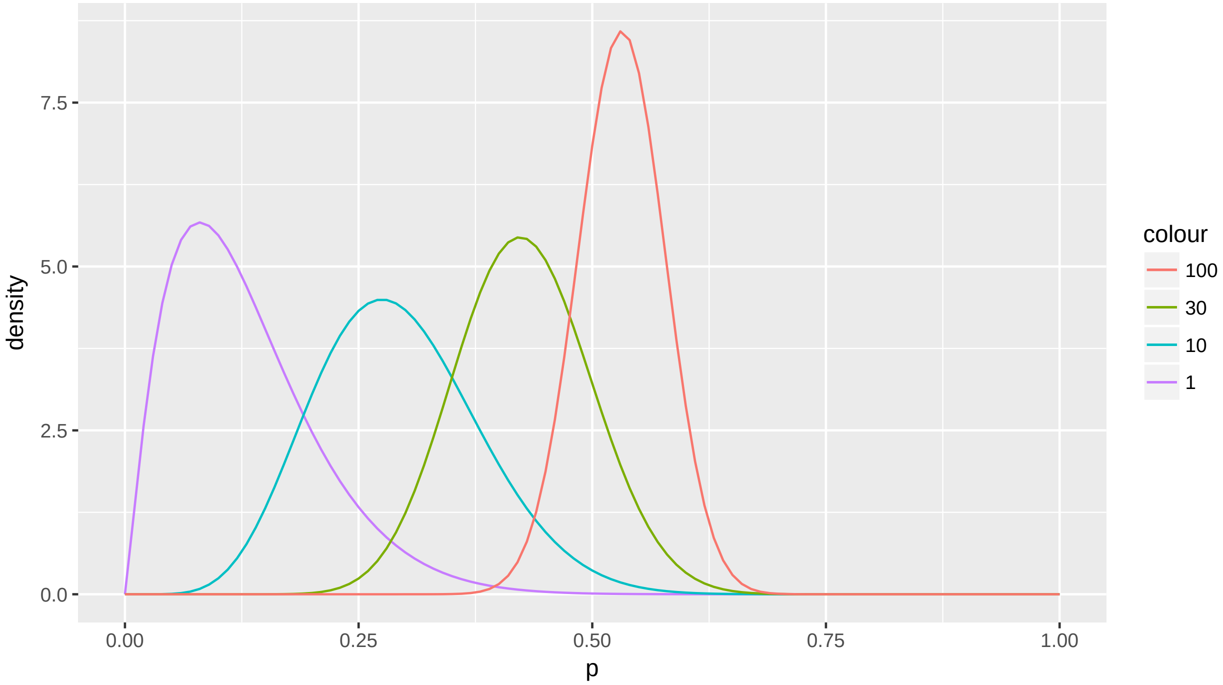
```
graph_posts(.1)
```



As expected, over time we get more sure of ourselves, as our posterior gets tighter. At $n = 100$, we're pretty confident that $p = .1$ which is corroborated by both our prior and the data.

But what happens when we have a bad choice of prior? Here we will keep the same prior but let $\hat{p}_l = p = 0.6$. We see what happens to our posterior over time:

```
graph_posts(.6)
```



In one way, this behaves as expected: we expected the prior to overtime become overrun by the data and the posterior to asymptotically approach symmetry around $p = 0.6$. However, I did not expect the distributions to get tighter; one might expect that as we see data that contradicts our prior, we get less sure of ourselves and the posterior distribution gets wider. But at every step of the way, the posterior distribution is getting tigher, which is obvious when viewed through a mathematical lens: at each step, both α and β is getting larger, reflecting a tighter and tighter density distribution.