STAT3120 Applied Bayesian Methods

Due Monday 22 August 2016 (on-campus by 3pm, distance by 5pm)

Late assignments will not be accepted without prior written permission of the lecturer.

Total marks: 50

Marking Criteria:

1. Answers must be written in clear English with all appropriate working and/or supporting computer output shown.
2. Raw computer output without explanatory text is unacceptable.
3. Students are required to understand the content in Weeks 1-4 to answer these questions.
4. As part of your workings you should include applicable R-code.

Question 1 Total: 15 marks

Law case (taken from Rossman and Short, 1995, Journal of Statistics Education).

Joseph Jameison was charged with multiple rapes in 1987 in Allegheny County in the US. DNA evidence from the scenes of the crimes revealed that the attacker displayed the genetic marker PGM2+1- in his DNA. This marker has only a 0.32% prevalence in the general population.

(a) Discuss how you might choose a prior probability that Jameison is guilty. Explain and defend your choice (do not set a specific number yet). 2 marks

Answer: From the description of the case we should highlight two important pieces of information. Firstly, the percentage of the population that carry the genetic marker displayed by the attacker is *very* small - 0.32%. Next, we should note what is omitted from the description of the genetic marker evidence: no values for the sensitivity or specificity of the test that determined the presence of the marker are stated. Therefore - at least in the toy context of this problem - we will assume that those values are both 100%. Obviously in reality this is impossible and we would have to take into account the accuarcy of the method/test used.

So, in summary, we are told that:

- The marker is *definitely* present.
- The attacker displayed the marker.
- Joseph Jameison is a suspect.

We are not given any reason's for why he is a suspect, so the only way we can decide on his guilt is by determining whether or not he displays the same genetic marker. This is an either/or question so it makes sense - without more evidence - to assign a 50/50 probabily of guilt. However, we should also take into account that the general principle of an assumption of innocence is a key tenet of most modern legal systems (certainly those in the anglosphere). Therefore it seems reasonable to suggest an intial probability of guilt $Pr(guilty) < 0.5$.

(b) Update your prior with the observed information. Display the full formula for Bayes' rule in terms of the as yet unknown prior Pr(guilty). 3 marks

does this change

Answer: We are looking for the folliwing formula:

$$Pr(G \mid M) = \frac{Pr(M \mid G)\,Pr(G)}{Pr(M \mid G)\,Pr(G) + Pr(M \mid G^c)\,Pr(G^c)}$$

Where

$$G := \text{Joseph Jameison (JJ) is Guilty}$$
$$M := \text{An individual carries the PGM2+1- marker}$$
$$Pr(M \mid G) := \text{JJ carries the PGM2+1- marker, given that he is guilty}$$
$$Pr(M \mid G^c) := \text{JJ carries the PGM2+1- marker, given that he is not guilty}$$

Now, given that the question states that the attacker *definitely has the marker*, the probability that JJ has the marker given that he is guilty is $Pr(M \mid G) = 1$.

The probability that JJ has the marker given that he is not guilty is $Pr(M \mid G^c) = (totalpopulation * prevalenceofthemarker) - 1/totalpopulation \approx 0.32\%$. So now we can restate the equation for Bayes rule as it applies here:

$$Pr(G \mid M) = \frac{Pr(M \mid G)\,Pr(G)}{Pr(M \mid G)\,Pr(G) + Pr(M \mid G^c)\,Pr(G^c)}$$
$$Pr(G \mid M) = \frac{(1)\,Pr(G)}{(1)\,Pr(G) + (0.0032)\,Pr(G^c)}$$
$$Pr(G \mid M) = \frac{Pr(G)}{Pr(G) + (0.0032) * Pr(G^c)}$$

(c) Display the posterior probability of guilt using a range of suitable values for the prior Pr(guilty). 3 marks

```
pr.guilty <- seq(0.1,1,by=0.1)
pr.marker <- 0.0032
pr.notguilty <- 1-pr.guilty
pr.post <- (1*pr.guilty)/(1*(pr.guilty) + pr.marker*(pr.notguilty))
prob.dens <- as.data.frame(cbind(pr.guilty, pr.post)) %>%
  transmute("Pr(Guilt)" =  pr.guilty, Posterior = signif(pr.post, digits = 5))
kable(prob.dens, "markdown")
```

Pr(Guilt)	Posterior
0.1	0.97201

Pr(Guilt)	Posterior
0.2	0.98736
0.3	0.99259
0.4	0.99522
0.5	0.99681
0.6	0.99787
0.7	0.99863
0.8	0.99920
0.9	0.99964
1.0	1.00000

Which doesn't shed much light on how the probability changes, other than telling us that it is above 90% for most values we can assign to $Pr(Guilt)$. Instead, lets simulate some more and plot the results.

```
pr.guilty <- seq(0,1,by=0.0001)
pr.marker <- 0.0032
pr.notguilty <- 1-pr.guilty
pr.post <- (1*pr.guilty)/(1*(pr.guilty) + pr.marker*(pr.notguilty))
prob.dens <- as.data.frame(cbind(pr.guilty, pr.post))

ggplot2::ggplot(prob.dens, aes(x=pr.guilty,y=pr.post)) +
  geom_line(data=prob.dens) +
  geom_hline(yintercept=0.975, linetype = 2, color = "red")
```
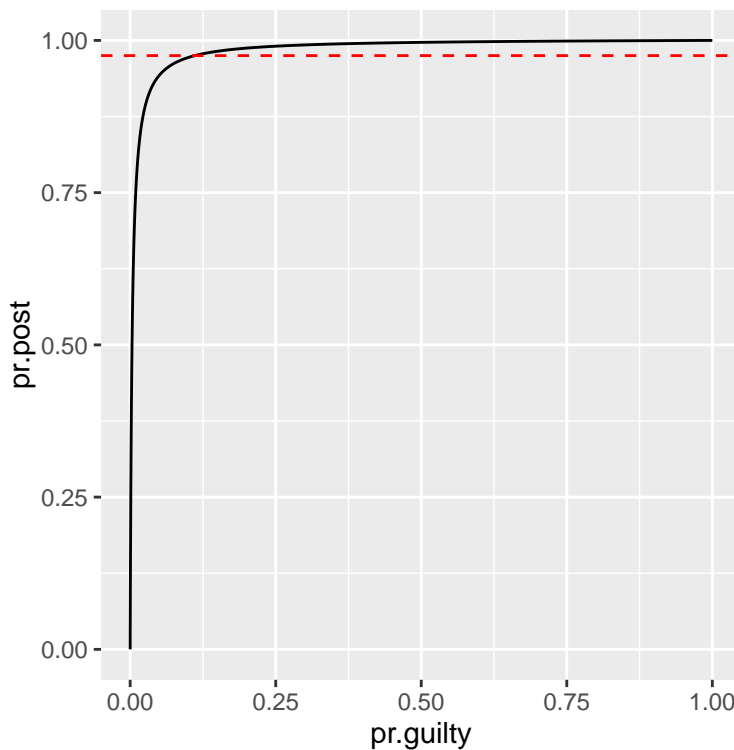

(d) Say that a jury wants to be at least 97.5% sure of guilt before returning a guilty verdict. What is the smallest prior probability of guilt that could be chosen to ensure this level of probability. 2 marks

3

First, we need to find the index of the first of the probabilities we calculated for the posterior density that is greater than 0.975. Then we find the corresponding value for the probabilty of guilt that generated that value at that index.

```
mpr.guilt <- which(pr.post > 0.975)
pr.guilty[mpr.guilt[1]]
```

```
## [1] 0.111
```

Therefore, the minimum probability of guilt that the can be chosen is $\approx 11.1\%$.

(e) Suppose there are 4,622 males living in Allegheny County. Given that we have found 1 male with PGM2+1- in his DNA, what is the probability that other males in the county also display this marker? 3 marks

Answer: We are not told that the distribution of the genetic marker amongst a population varies by sex. Therefore, we should assume that the 0.32% prevalence stated in the question is suitable to use here. In a population of $4,622$ we would expect to see:

```
n <- 4622; prev <- 0.0032;
pr.another <- ((n-1)*prev)/(n*prev)
pr.another
```

```
## [1] 0.9997836
```

So there is approximately a 99.98% chance that, given there is one male with the marker, there will be at least one more.

(f) Should Jameison be found guilty on the DNA evidence alone? Discuss. 2 marks

Question 2 Total : 10 marks

Part 1: ABC book distributors have a new book that they hope will be popular with their customers. To obtain an idea of how popular it will be they choose a random sample of 40 customers, ring them, give them a brief description and ask whether they would like to buy the book. 15 of the sample decided to buy the book. What inference can we make about the true proportion of the population of customers that will buy the new book based on this sample?

(a) Under a uniform prior, derive the posterior distribution for the unknown true proportion. 2 marks
Answer: First of all, what is described corresponds to a Binomial distribution.

$$p(\theta|y = 15) \propto Pr(y = 15|\theta)p(\theta)$$
$$\propto \theta^{15}(1 - \theta)^{25}$$

We know that the Bayes-Laplace ($Beta(1, 1)$) prior approximates a Uniform, flat prior over the interval $(0, 1)$. We also know that that the $Beta(\alpha, \beta)$ is a good substitute for the Binomial and will give us a $Beta$ posterior.
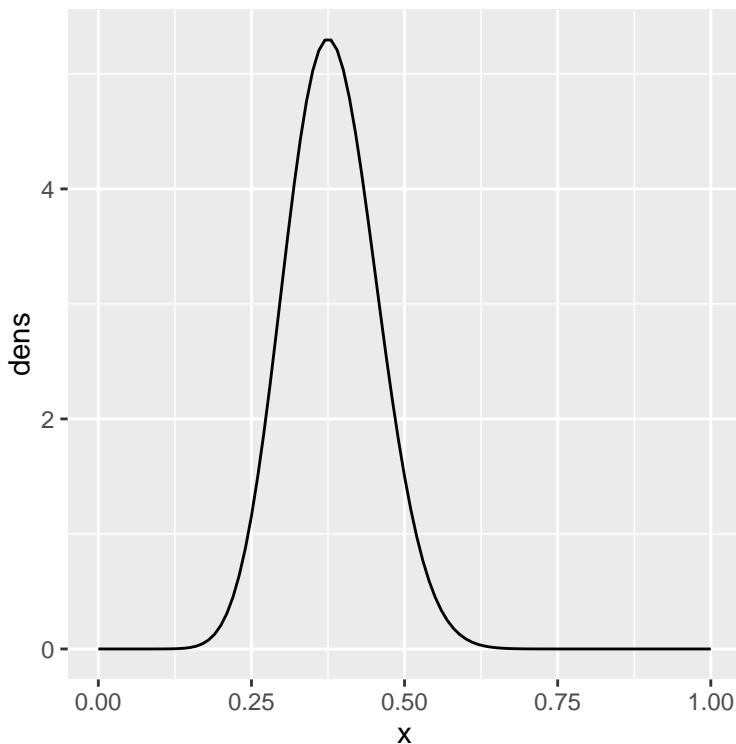
So the posterior works out to be

```
x <- seq(0,1,by = 0.01); a = 15; b = 40

dens <- dbeta(
  x,
  a + 1,
  b - a + 1
  )

posterior.dens.unif <- as.data.frame(cbind(x,dens))

ggplot(
  data = posterior.dens.unif,
  aes(x,dens)
  ) +
  geom_line()
```


(b) Is ABC likely to achieve 50% of their population of customers buying the new book? 2 marks

Answer: We can use the **pbeta** function to estimate the probability that ABC will meet their target.

```
a = 15; b = 40
pbeta(0.5,a + 1,b - a + 1,lower.tail=FALSE)
```

```
## [1] 0.0586376
```

implying that it is not very likely that ABC will achieve their target. We can confirm this be calulating the mean of a sample from the same distribution.

```
x <- seq(0,1,by = 0.01)
unif.data <- rbeta(
  10000,
  a + 1,
  b - a + 1
  )
mean(unif.data)
```

```
## [1] 0.381298
```

which implies that the most likely outcome for ABC - using the given prior and likelihood - is that $\approx 38\%$ of their customers will purchase the book.

We could also use **qbeta** to find a 90% CI.

```
beta.ci <- qbeta(c(0.95,0.05),a + 1, b - a + 1,lower.tail = FALSE)
names(beta.ci) <- c("5%","95%")
beta.ci
```

```
##         5%        95%
## 0.2623051 0.5061243
```

which backs up the result above.

Part 2: In past ABC new book promotions they have found that on average 23.5% of their customers buy the book being promoted. In fact, the proportion of customers buying new books after phone promotions has a histogram that closely matches a $Beta(4, 13)$ density.
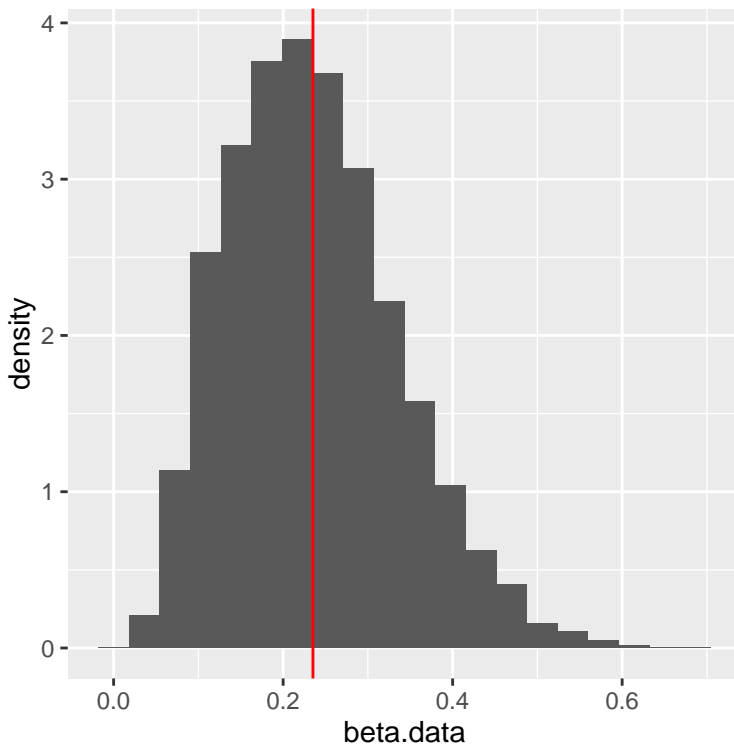
(c) Can we use this as a prior density? Discuss. 2 marks

Answer: We can check this by generating some random samples from a $Beta(4, 13)$ distribution and calculating the mean of those samples.

```
a = 3; b = 15

beta.data <- as.data.frame(
  rbeta(
    10000,
    a + 1,
    b - a + 1
    )
)

ggplot(beta.data, aes(beta.data, ..density..)) +
  geom_histogram(bins = 20) +
  scale_x_continuous() +
  geom_vline(xintercept = mean(beta.data[,1]), color = "red")
```
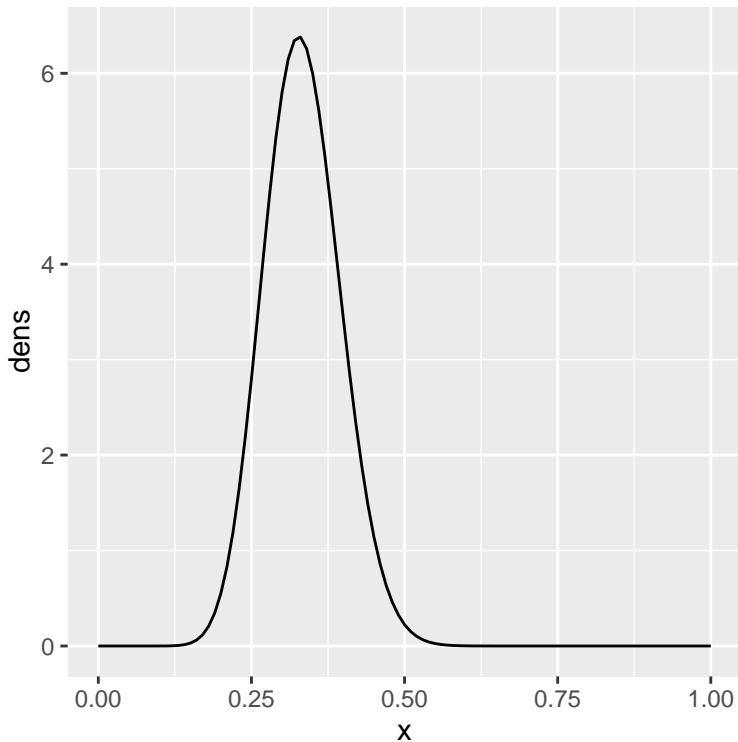
6

```
mean(beta.data[,1])
```

```
## [1] 0.235163
```

Which is $\approx 23.5\%$. Therefore, we can justifiably use $Beta(4, 13)$ as a prior distribution.

```
x <- seq(0,1,by = 0.01); a1 <- 15; b1 <- 40; a2 <-  3; b2 <-  15

dens <- dbeta(
  x,
  (a1 + a2) + 1,
  (b1 + b2) - (a1 + a2) + 1
  )

posterior.dens.beta <- as.data.frame(cbind(x, dens))

ggplot(
  data = posterior.dens.beta,
  aes(x,dens)
  ) +
  geom_line()
```

```r
beta.data <- rbeta(
  10000,
  (a1 + a2) + 1,
  (b1 + b2) - (a1 + a2) + 1
  )

mean(beta.data>0.5)
```
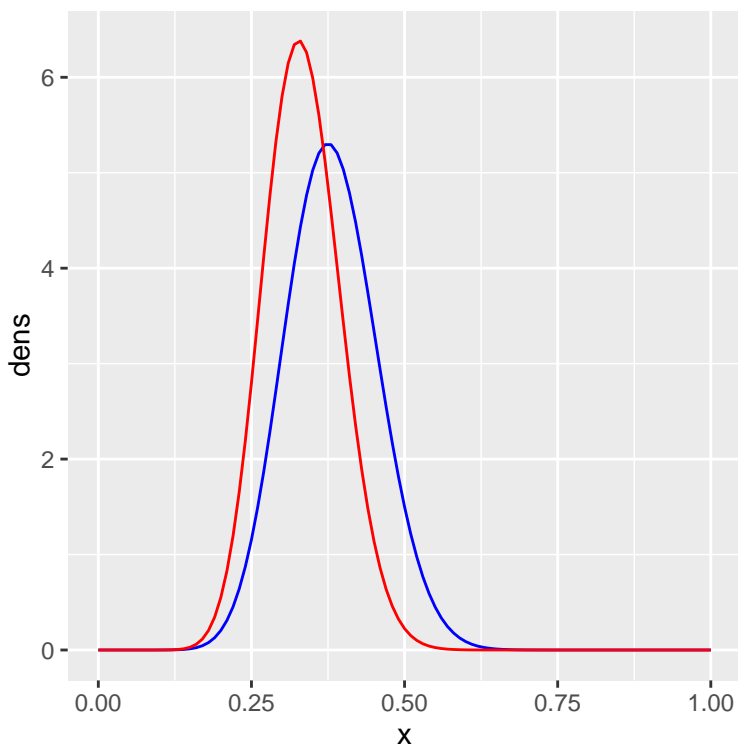
```
## [1] 0.0055
```

```r
mean(beta.data)
```

```
## [1] 0.3342675
```

(e) How did this prior information affect the likelihood of at least 50% of ABCs population of customers buying the new book? 2 marks

```r
ggplot() +
  geom_line(
    data = posterior.dens.unif,
    aes(x,dens),
    color = "blue"
    ) +
  geom_line(
    data = posterior.dens.beta,
    aes(x,dens),
    color = "red"
    )
```

8

```r
mean(beta.data>0.5)
```

```
## [1] 0.0055
```

```r
mean(beta.data)
```

```
## [1] 0.3342675
```


Question 3 Total: 4 marks

Suppose you have a Beta(6,6) prior on the probability θ that a coin will yield a head when spun in a specified manner. The coin is independently spun 5 times. All you are told is that all 5 results were the same, i.e. either 5 heads or 5 tails. Derive your posterior density (up to a proportionality constant) for θ and sketch it.

Answer: We are given that the prior probability of θ is proportional to a $Beta(6,6)$ distribution:

$$f(\theta) \propto \theta^5 (1-\theta)^5$$

```r
theta.prior <- seq(0,1,0.01)
prior.density  <- theta.prior^5*(1-theta.prior)^5
prior.data <- as.data.frame(cbind(theta.prior, prior.density))

ggplot(
  prior.data,
  aes(theta.prior, prior.density)) +
  geom_line()
```
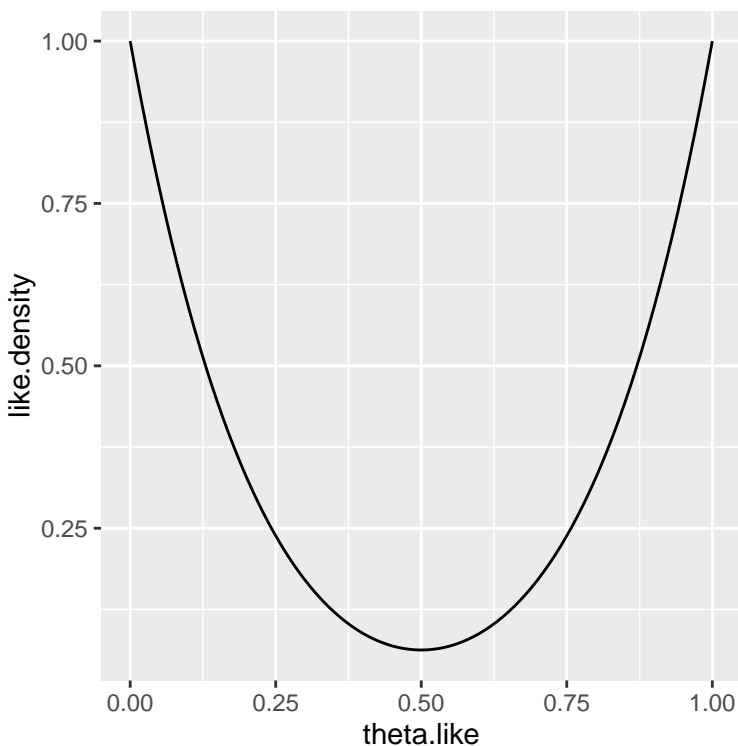
9
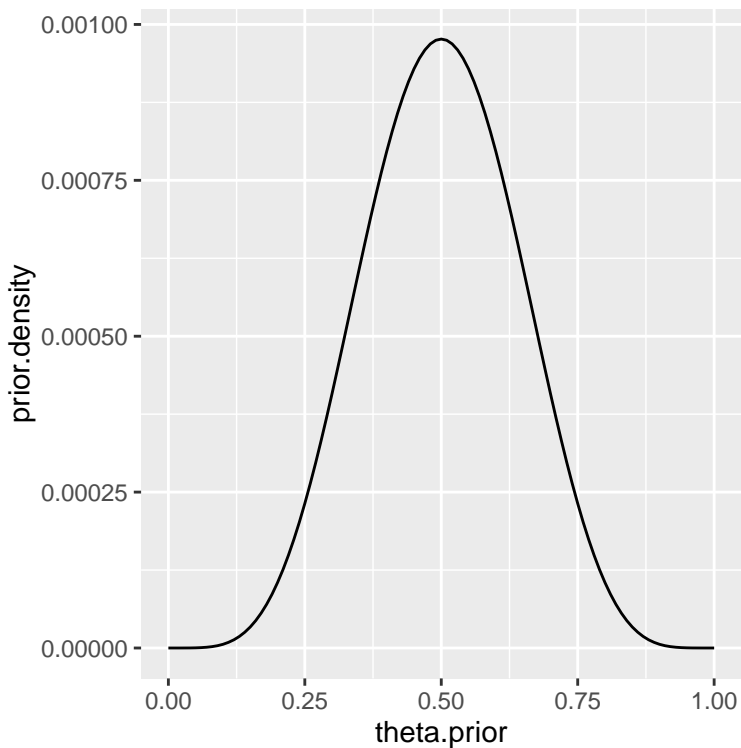
We are also given that the data could represent only two possible outcomes - either all heads or all tails. Therefore, the likelihood of the data is equal to:

$$Pr(x \mid \theta) = \binom{5}{0}(1-\theta)^5 + \binom{5}{5}\theta^5$$
$$= (1-\theta)^5 + \theta^5$$

```
theta.like <- seq(0,1,0.01)
like.density  <- (1-theta.like)^5  + theta.like^5
like.data <- as.data.frame(cbind(theta.like, like.density))

ggplot(
  like.data,
  aes(theta.like, like.density)) +
  geom_line()
```

So, we can derive the posterior by multiplying the prior by the likelihood:

$$f(\theta|x) \propto Pr(x|\theta)f(\theta)$$
$$\propto \theta^5(1-\theta)^5 * \left((1-\theta)^5 + \theta^5\right)$$
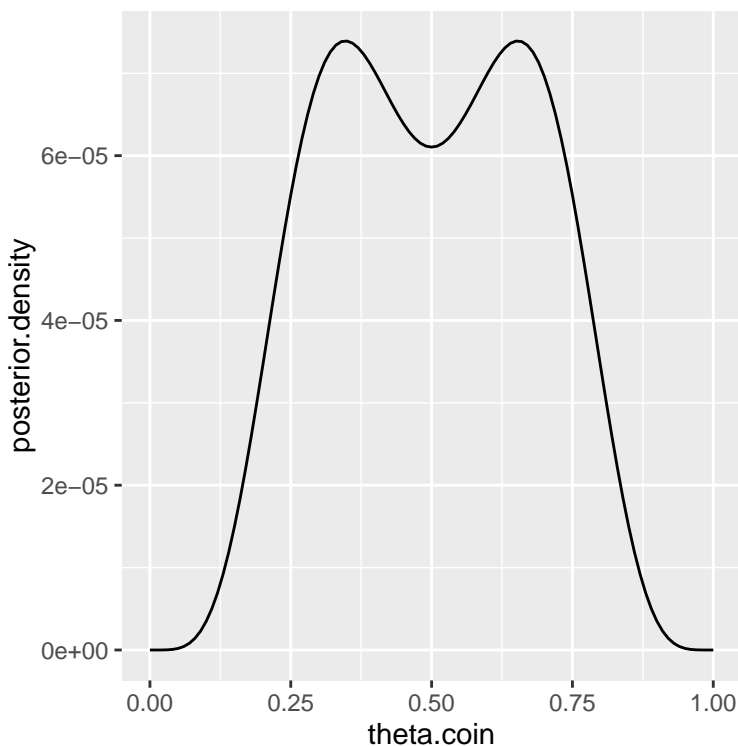$$\propto \theta^5(1-\theta)^{10} + \theta^{10}(1-\theta)^5$$

Now let's sketch the distribution

```
theta.coin <- seq(0,1,0.01)
posterior.density  <- theta.coin^5*(1-theta.coin)^10 + theta.coin^10*(1-theta.coin)^5
coin.data <- as.data.frame(cbind(theta.coin, posterior.density))

ggplot() +
  geom_line(
    data = coin.data,
    aes(theta.coin, posterior.density))
```

Question 4 Total: 9 marks

A treatment vs control group example discussed at various times in the literature is based on 11 successes in 11 treatments (i.e. 11/11), and a single failure in the control group (i.e. 0/1).

(a) Find a classical solution to the question of whether the treatment is successful. Bare output from a statistical package is acceptable, but without understanding the procedure, it may be difficult to obtain full marks at (c). 3 marks

```
##              outcome
## group       Failure Success
##    Control        1       0
##    Treatment      0      11
```

```
## Warning in chisq.test(tab, correct = FALSE): Chi-squared approximation may
## be incorrect
```

```
##    statistic      p.value parameter                     method
## 1         12 0.0005320055         1 Pearson's Chi-squared test
```

Utilising a classical solution, we find that the p-value returned by the χ^2 test is $\approx 5.3200551 \times 10^{-4} < \alpha = 0.05$, which implies that we should reject H_0: that the treatment result is significant.

(b) Apply a Bayesian approach to this question, trying both the Bayes-Laplace and Jeffreys priors. 4 marks

Answer: We should define two Binomial distributions, one for each treatment group: "Treatment" and "Control".

$$X_{Treatment} \sim Binomial(11, \theta_T)$$
$$X_{Control} \sim Binomial(1, \theta_C)$$

We know that a good conjugate prior to use for the Binomial distribution is a Beta distribution. So let's define two conjugate priors for our data.

$$Pr(\theta_T) \sim Beta(\alpha_T, \beta_T)$$
$$Pr(\theta_C) \sim Beta(\alpha_C, \beta_C)$$

Our posterior distributions for each would then both be Beta distributions:

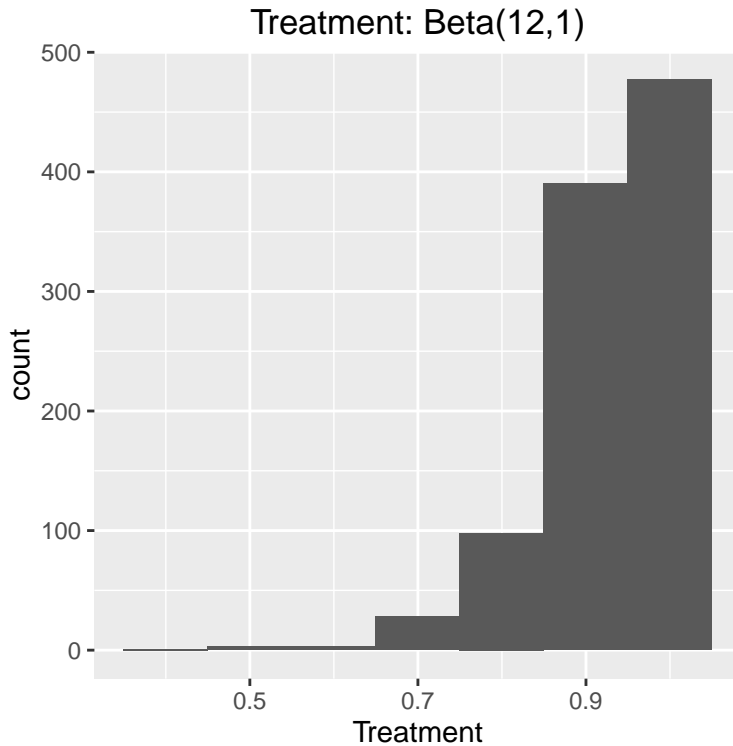$$Pr(\theta_T|X_T) \sim Beta(\alpha = 11 + \alpha_T, \beta = 0 + \beta_T)$$
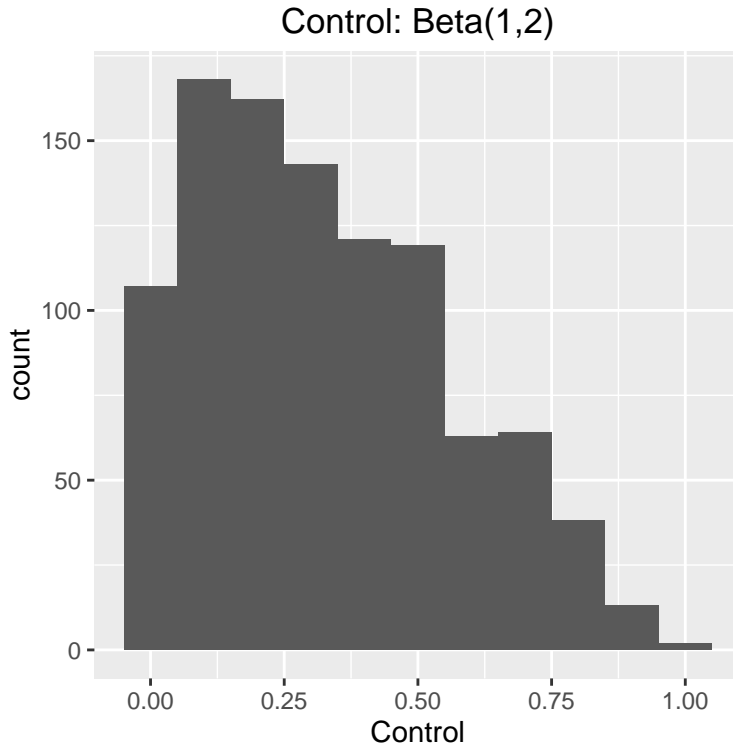$$Pr(\theta_C|X_C) \sim Beta(\alpha = 0 + \alpha_C, \beta = 1 + \beta_C)$$

Now, a Bayes-Laplace prior $\equiv Beta(1, 1)$. Assigning this to both of our distributions gives us two posterior distributions.

$$Pr(\theta_T|X_T) \sim Beta(\alpha = 11 + 1, \beta = 0 + 1) \implies Beta(12, 1)$$
$$Pr(\theta_C|X_C) \sim Beta(\alpha = 0 + 1, \beta = 1 + 1) \implies Beta(1, 2)$$
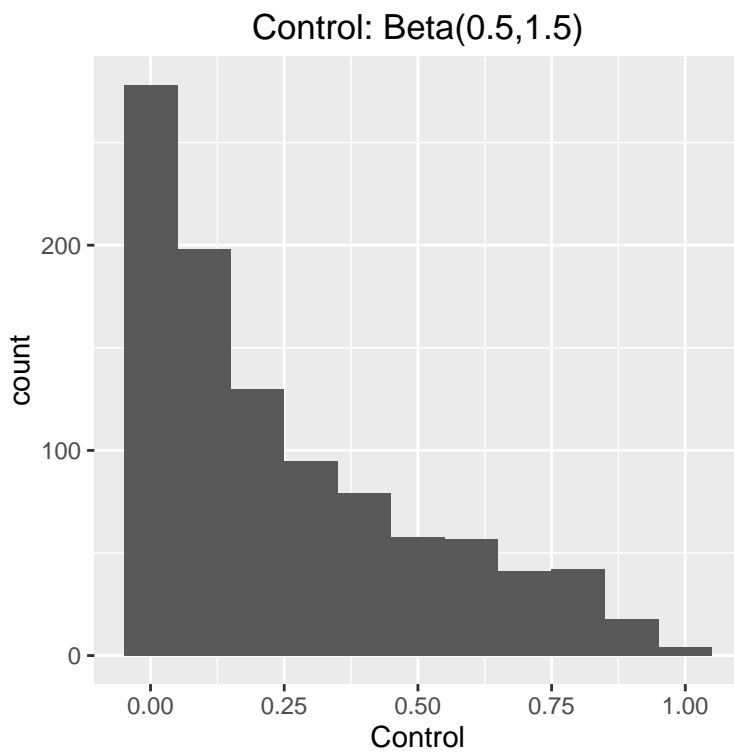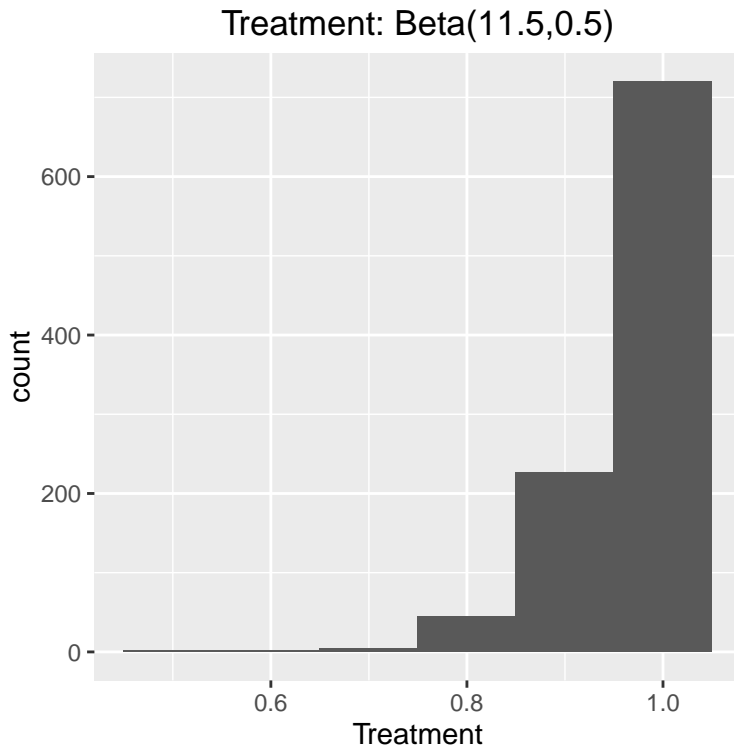
Bayes-Laplace Priors

Treatment: Beta(12,1)

Control: Beta(1,2)

A Jeffrey's prior however is $\equiv Beta(1/2, 1/2)$. Assigning this to both of our distributions gives us two more posterior distributions.

$$Pr(\theta_T | X_T) \sim Beta(\alpha = 11 + 0.5, \beta = 0 + 0.5) \quad \implies \quad Beta(11.5, 0.5)$$
$$Pr(\theta_C | X_C) \sim Beta(\alpha = 0 + 0.5, \beta = 1 + 0.5) \quad \implies \quad Beta(0.5, 1.5)$$

Jeffrey's Priors

Treatment: Beta(11.5,0.5)

Control: Beta(0.5,1.5)

(c) Briefly explain possible reasons behind the various differences. 2 marks

Question 5 Total: 4 marks

A random sample of n students is drawn from a large population, and their weights are measured. The average weight of the n sampled students is $\bar{y} = 150$ pounds. Assume the weights in the population are Normally distributed with unknown mean θ and known standard deviation of 20 pounds. Suppose your prior distribution for θ is Normal with mean 180 and known standard deviation of 40 pounds.

Give your posterior distribution for θ. (Note that your answer will be a function of n.)

We are given the following:

$$\bar{y} \mid \theta \sim N(\mu, \frac{20^2}{n})$$

$$\theta \sim N(180, 40^2)$$

So we are looking for the following:

$$Pr(\theta \mid \hat{y}) = N(\theta \mid \mu_n, \tau_n^2)$$

From class we know that

$$\mu_n = \frac{\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}}{frac1\tau_0^2 + \frac{n}{\sigma^2}}$$

***

Question 6 Total: 8 marks

Consider the Uniform$(0, \theta)$ distribution, i.e.

$$p(x \mid \theta) = \frac{1}{\theta}, 0 < x < \theta$$

from which a random sample $\{x_1, ..., x_n\}$ is obtained. Note that $y_n = \max\{x_1, ..., x_n\}$ is a sufficient statistic.

(a) Derive the posterior distribution $p(\theta|x_1, ..., x_n) = p(\theta|y_n)$, based on the noninformative prior $p(\theta) \propto 1/\theta$. Be careful to use the correct likelihood and range! 3 marks

Answer: Firstly, we have that that the density of the uniform distribtion $\mathcal{U}(0, \theta)$ is:

From this we can calculate the Likelihood function be taking the product of $n\times$ the uniform density of a single r.v.

$$L(\theta \mid x) = \begin{cases} \prod_{i=1}^{n} \frac{1}{\theta^i} & \max\{x_1, x_2, \cdots, x_n\} \leq \theta \\ 0 & otherwise \end{cases}$$

(b) Plot this posterior for $y_n = 5$ and $n = 20$. 2 marks

Answer: With $y_n = \max\{x_1, x_2, \cdots, x_n\} = 5$ and $n = 20$ we have a posterior distribution of

$$Pr(\theta \mid x) = \begin{cases} \frac{1}{\theta^{21}} & \max\{x_1, x_2, \cdots, x_{20}\} \leq \theta \\ 0 & otherwise \end{cases}$$
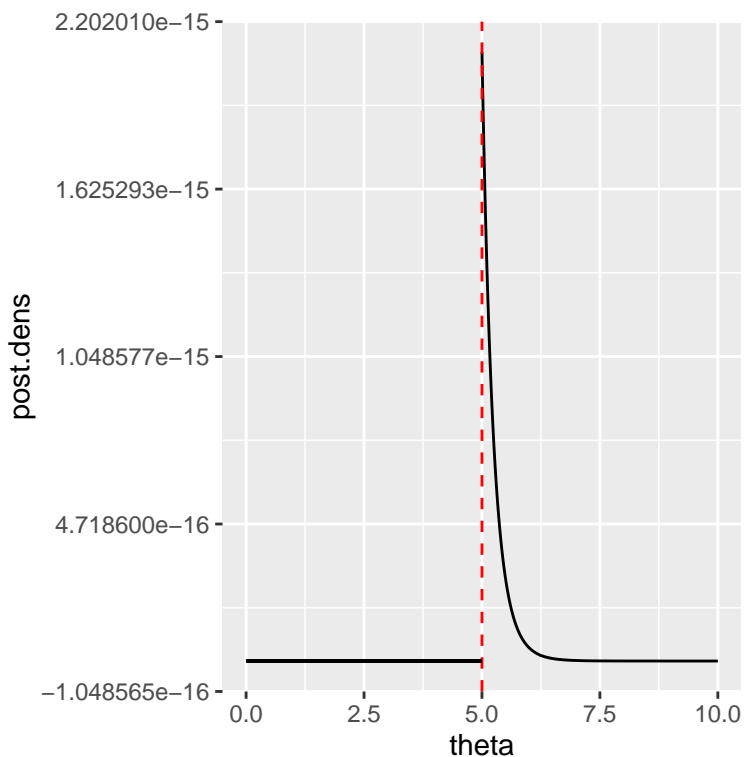
16

```
y <- 5; n <- 20; a <- 5

theta <- seq(y,10, by = 0.001)
post.dens <- 1/(theta^(n+1))
dens <- as.data.frame(cbind(theta,post.dens))

theta.0 <- as.data.frame(seq(0,10, by = 0.001))

  ggplot2::ggplot() +
    geom_line(
      data = dens,
      aes(theta,post.dens)
      ) +
    geom_vline(
      xintercept=y,
      color= "red",
      linetype = 2
      ) +
    geom_segment(
      data = theta.0,
      x=0,
      xend=y,
      y=0,
      yend=0
      ) +
    xlim(0,10)
```


From this it should be obvious that the MLE of θ is achieved when $\theta = y_n$.

17

(c) For the data at (b), derive 95% credible intervals, both central and HPD. Which would you prefer and why? 3 marks

End of Assignment 1