

# Homework 1

Bingying Liu (bl199)

January 22, 2020

## Question 1

1.  $p(y = \text{head}) = \sum_{\theta \in \text{fair, fake}} p(y|\theta) * p(\theta) = \frac{2}{3} * 0.5 + \frac{1}{3} * 1 = \frac{2}{3}$   
Therefore, the probability that it lands on heads is  $\frac{2}{3}$ .

2.

$$\begin{aligned} p(\theta = \text{fake} | y = \text{head}) &= \frac{p(\theta, y)}{p(y)} \\ &= \frac{p(y|\theta) * p(\theta)}{p(y)} \\ &= \frac{1 * \frac{1}{3}}{\frac{2}{3}} \\ &= \frac{1}{2} \end{aligned}$$

The probability that it's the fake coin is 50%.

## Question 2

a. Since  $p(x|y, z) = \frac{p(x, y, z)}{p(y, z)} = \frac{c * f(x, z) * g(y, z) * p(z)}{p(y, z)} = \frac{c * g(y, z) * p(z)}{p(y, z)} * f(x, z)$

Assuming c is a constant and since  $p(x|y, z)$  consists of multiplier  $\frac{c * g(y, z) * p(z)}{p(y, z)}$  and function  $f(x, z)$ , we can say that  $p(x|y, z)$  is a function of x and z.

b. Same reason as above: since  $p(y|x, z) = \frac{p(x, y, z)}{p(x, z)} = \frac{c * f(x, z) * g(y, z) * p(z)}{p(x, z)} = \frac{c * f(x, z) * p(z)}{p(x, z)} * g(y, z)$

Assuming c is a constant and since  $p(y|x, z)$  consists of multiplier  $\frac{c * f(x, z) * p(z)}{p(x, z)}$  and function  $g(y, z)$ , we can say that  $p(y|x, z)$  is a function of y and z.

c. Since we know  $p(x, y|z) = p(x|y, z) * p(y|z)$ , to prove  $p(x, y|z) = p(x|z) * p(y|z)$ , we only need to show that  $p(x|y, z) = p(x|z)$ .

$$\begin{aligned} p(x|z) &= \frac{p(x, z)}{p(z)} \\ &= \frac{\int p(x, y, z) dy}{\int p(x, y, z) dx dy} \\ &= \frac{c \int f(x, z) * g(y, z) * h(z) dy}{c \int f(x, z) * g(y, z) * h(z) dx dy} \text{ assuming } p(x, y, z) = c * f(x, z) * g(y, z) * h(z) \text{ and } c \text{ is a constant} \\ &= \frac{h(z) * G(z) * f(x, z)}{H(z)} \text{ we set } G(z) = \int g(y, z) dy \text{ and } H(z) = \int f(x, z) * g(y, z) * h(z) dx dy, \text{ they're both functions of } z \end{aligned}$$

Since we can treat  $\frac{h(z)G(z)}{H(z)}$  as a numerical value, we can say that  $p(x|z) \propto f(x, z)$ , which is the same as  $p(x|y, z) \propto f(x, z)$  proved in a).

Therefore, this equality is proved and we can say that X and Y are conditionally independent given Z.

## Question 3

Since the posterior distribution is  $\text{beta}(a + y, b + n - y)$ , we have the variance of posterior as:

$$\begin{aligned} V(\theta|y) &= \frac{(a + y)(b + n - y)}{(a + b + n)^2 * (a + b + n + 1)} \\ &= \frac{a + y}{a + b + n} * \frac{b + n - y}{a + b + n} * \frac{1}{a + b + n + 1} \\ &= E(\theta|y) * \frac{a + b + n - a - y}{a + b + n} * \frac{1}{a + b + n + 1} \\ &= E(\theta|y) * (1 - E(\theta|y)) * \frac{1}{a + b + n + 1} \\ &= \frac{E(\theta|y) * (1 - E(\theta|y))}{a + b + n + 1} \end{aligned}$$

## Question 4: Hoff 3.1

a.

- Joint distribution is the following:

$$\begin{aligned}
 p(Y_1 = y_1, \dots, Y_{100} = y_{100} | \theta) &= p(Y_1 = y_1 | \theta) * \dots * p(Y_{100} = y_{100} | \theta) \\
 &= \prod_{i=1}^n \theta^{y_i} * (1 - \theta)^{1-y_i} \\
 &= \theta^{\sum_{i=1}^n y_i} * (1 - \theta)^{n - \sum_{i=1}^n y_i} \\
 &= \theta^Y * (1 - \theta)^{n-Y}
 \end{aligned}$$

assuming  $Y = \sum_{i=1}^n y_i$ .

- Since  $\sum Y_i$  are the number of individuals who support policy Z from 100 individuals (Bernoulli trials), it follows that  $p(\sum Y_i = y | \theta)$  is a binomial distribution.

$$\begin{aligned}
 p(\sum Y_i = y | \theta) &= \binom{n}{y} \theta^y (1 - \theta)^{n-y} \\
 &= \binom{100}{y} \theta^y (1 - \theta)^{100-y}
 \end{aligned}$$

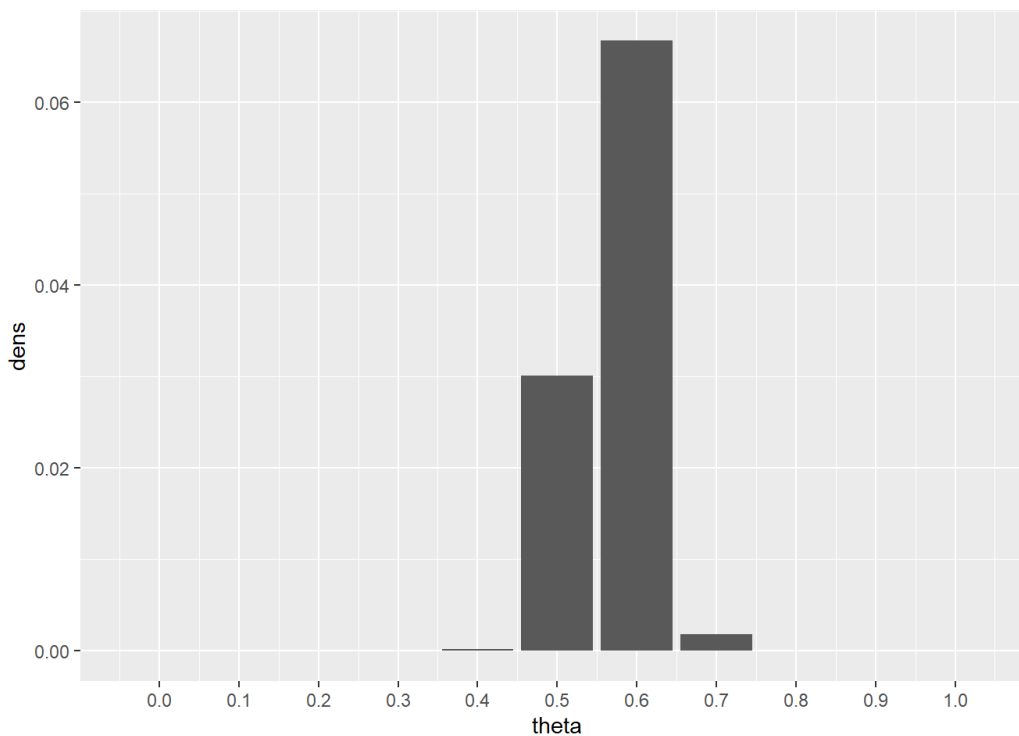
- b. Compute sampling distribution for each  $\theta$  and make a plot of the probabilities as a function of  $\theta$ .

```

Y = 57
N = 100
theta <- seq(from=0, to=1, by=0.1)
dens <- choose(N,Y) * theta^Y*(1-theta)^(N-Y) # sum of Y as sufficient statistics is a binomial distribution
df <- data.frame(theta, dens)

ggplot(df, aes(x = theta, y = dens)) +
  geom_bar(stat = 'identity') +
  scale_x_continuous(breaks = theta)

```

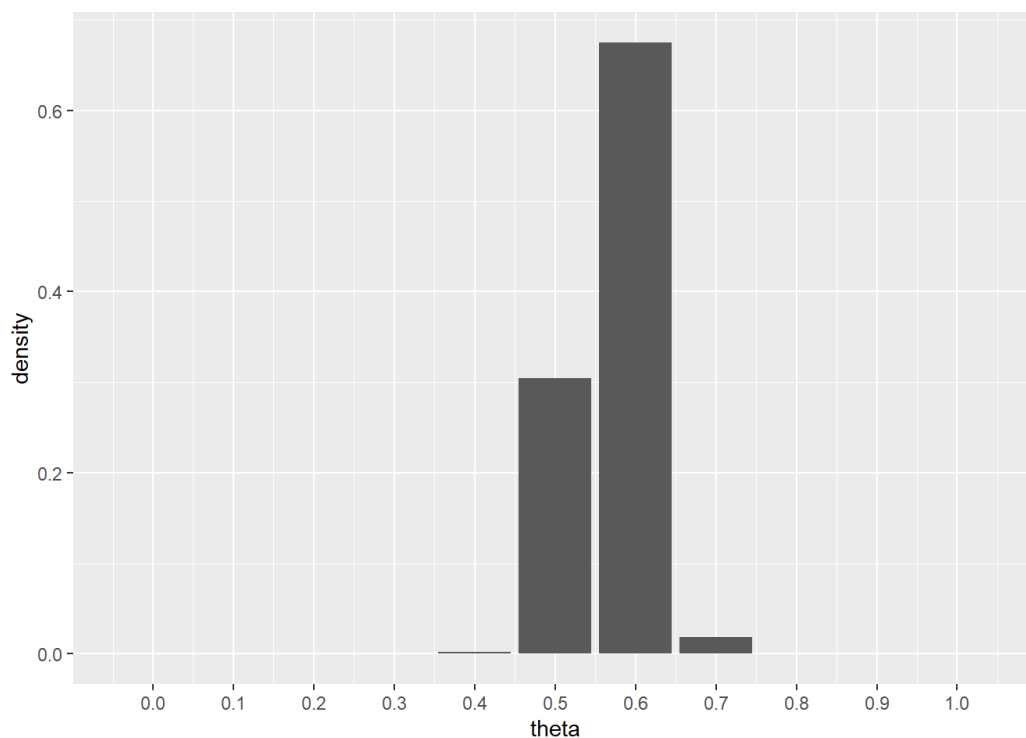


- c. Compute posterior distribution for each  $\theta$  and make a plot of this as a function of  $\theta$

$$\begin{aligned}
 p(\theta | \sum Y_i = 57) &= \frac{p(\sum Y_i = 57 | \theta)p(\theta)}{p(\sum Y_i = 57)} \\
 &= \frac{p(\sum Y_i = 57 | \theta)p(\theta)}{\sum_{\theta'} p(\sum Y_i = 57 | \theta')p(\theta')} \\
 &= \frac{p(\sum Y_i = 57 | \theta)}{\sum_{\theta'} p(\sum Y_i = 57 | \theta')} \\
 &\propto p(\sum Y_i = 57 | \theta)
 \end{aligned}$$

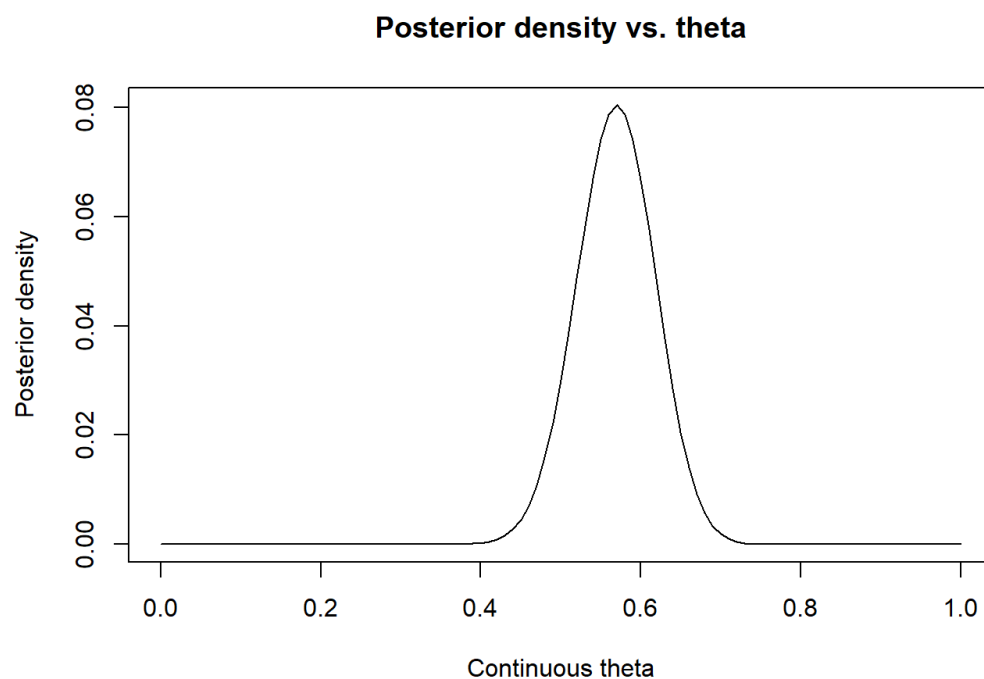
The second to last step is derived because we have a uniform prior so  $p(\theta) = p(\theta')$  and the last step is derived because we know that the denominator is a constant.

```
prior <- dbeta(theta,1,1)
posterior <- prior * dens/sum(dens)
ggplot(df, aes(x = theta, y = posterior)) +
  geom_bar(stat = 'identity') +
  scale_x_continuous(breaks = theta)+
  ylab('density')
```



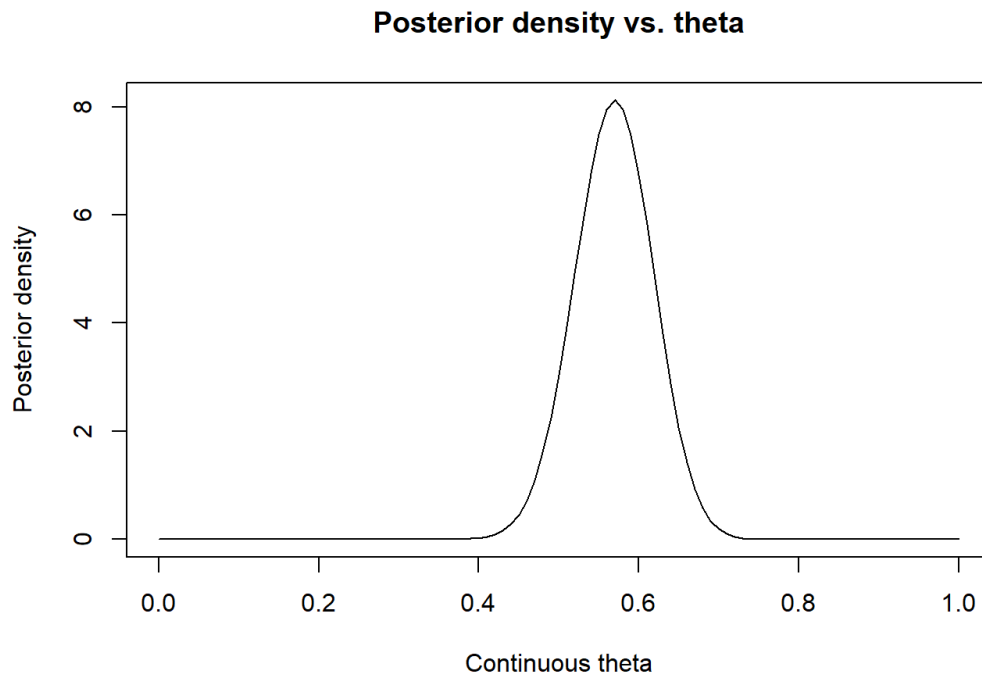
d. Plot the posterior density as a function of  $\theta$  using  $p(\theta) \times P(\sum Y_i = 57 \mid \theta)$ .

```
theta.any <- seq(from=0, to=1, by=0.01)
posterior_wlcons <- choose(N,Y) * theta.any^Y*(1-theta.any)^(N-Y)
plot(theta.any, type='l',posterior_wlcons, xlim=c(0,1),xlab = "Continuous theta", ylab = "Posterior density", main="Posterior density vs. theta")
```



e. Plot the posterior density of as a function of  $\theta$  using  $\theta = \text{beta}(1 + 57, 1 + 100 - 57)$ .

```
posterior_beta <- dbeta(theta.any, 1+Y,1+N-Y)
plot(theta.any, type='l',posterior_beta, xlim=c(0,1), xlab = "Continuous theta", ylab = "Posterior density", main="Posterior density vs. theta")
```



Relationships among all of the plots:

- b) is the sampling distribution and c) is the posterior distribution with fewer theta values on the x-axis. They have the same shape because uniform prior doesn't influence posterior calculation, so the posterior distribution is completely influenced by data itself. Posterior density is after normalization, so the pictures are in different scales.
- d) is the posterior density before normalization by  $p(\sum Y_i = 57) = 0.099$ . e) is the posterior density after normalization. (d) and (e) have the same shape because prior  $p(\theta)$  which is a uniform distribution, doesn't influence the posterior calculation.

Question 5: Hoff 3.2f

```

N = 100
exp.theta = function(n0, theta0, y) {
  (N / (n0 + N)) * (y / N) + (n0 / (n0 + N)) * theta0
}
Theta0 = rev(seq(0, 1, by = 0.1))
N0 = seq(0, 32, by = 0.5)

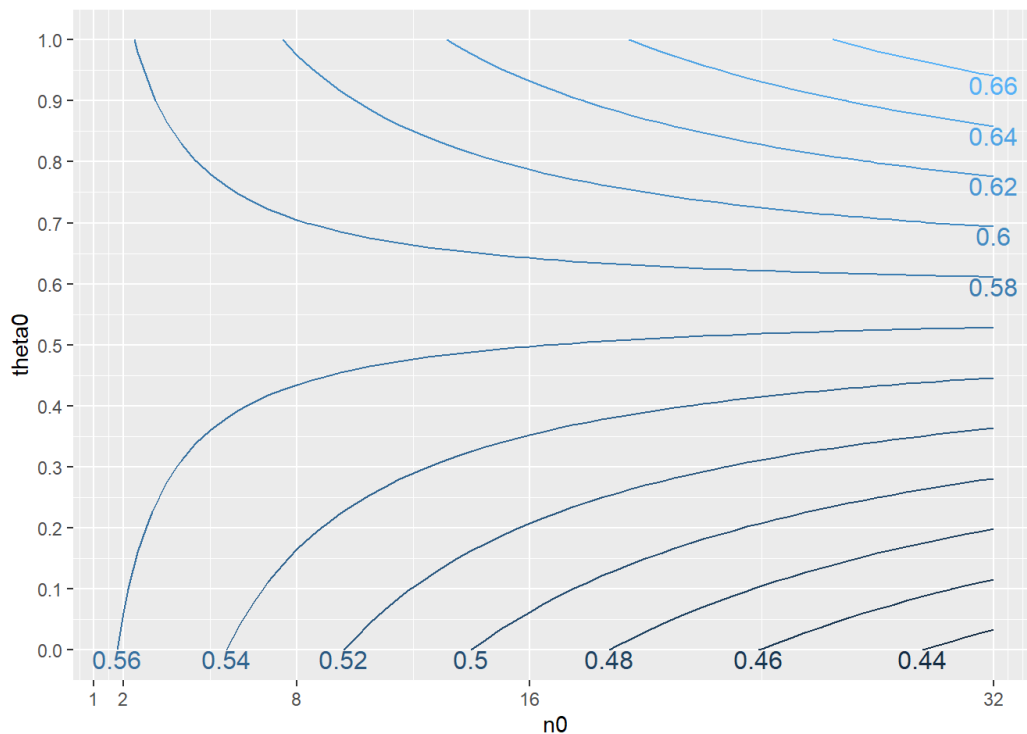
# Find a and b values
a_formula <- function(n0, theta0){
  n0 * theta0
}
da <- outer(Theta0, N0, function(n0, theta0) a_formula(n0, theta0))
rownames(da) = Theta0
colnames(da) = N0
df_a = melt(da)
colnames(df_a) = c('theta0', 'n0', 'a')
df_a['b'] <- df_a['n0']-df_a['a']

# Find posterior theta values
y = 57
d = outer(Theta0, N0, FUN = function(theta0, n0) exp.theta(n0, theta0, 57))
rownames(d) = Theta0
colnames(d) = N0

df = melt(d)
colnames(df) = c('theta0', 'n0', 'theta')

p = ggplot(df, aes(x = n0, y = theta0, z = theta)) +
  geom_contour(aes(colour = ..level..)) +
  scale_x_continuous(breaks = c(1, 2, 8, 16, 32), labels = c(1, 2, 8, 16, 32)) +
  scale_y_continuous(breaks = Theta0)
library(directlabels)
direct.label(p, method = 'bottom.pieces')

```



We can explain to someone to believe that  $\theta > 0.5$  by quantifying their prior beliefs about the proportion with two factors:  $\theta_0$ , prior expectation, and  $n_0$ , the “sample size” of observed individuals that contributed to the initial estimate.

## Question 6: Trials we need

```

# This is a function to find what value of n is required
find_n <- function(shape_1, shape_2){
  seq1 <- c()
  for (i in 100:3000){
    cdf <- pbeta(0.0015, shape1 = shape_1, shape2 = shape_2+i)
    seq1 <- c(seq1, cdf)
  }
  n <- Position(function(x) x > 0.95, seq1)
  return(n+100)
}

beta_1 <- find_n(1,666) #1231
beta_2 <- find_n(0.05,33.33) #146
beta_3 <- find_n(1.6,407.4) #2311
beta_4 <- find_n(1.05,497) #1564

# Since Unif(0,0.1) is a truncated prior, we use F(z)-F(a)/F(b)-F(a) to calculate its cdf
seq2 <- c()
for (i in 100:3000){
  cdf <- pbeta(0.0015, shape1 = 1, shape2 = 1+i)/pbeta(0.1, shape1 = 1, shape2 = 1+i)
  seq2 <- c(seq2, cdf)
}
n <- Position(function(x) x > 0.95, seq2)
beta_5 <- n+100 #199

# Print out the number of samples we need
beta_1

```

```
## [1] 1331
```

```
beta_2
```

```
## [1] 146
```

```
beta_3
```

```
## [1] 2311
```

```
beta_4
```

```
## [1] 1564
```

```
beta_5
```

```
## [1] 1996
```