STA 601 Homework 2

Lingyun Shao

Sep. 12, 2018

3.1

Sample survey: Suppose we are going to sample 100 individuals from a county (of size much larger than 100) and ask each sampled person whether they support policy Z or not. Let $Y_i = 1$ if person i in the sample supports the policy, and $Y_i = 0$ otherwise.

    a) Assume $Y_1, \cdots, Y_{100}$ are, conditional on θ, i.i.d. binary random variables with expectation θ. Write down the joint distribution of $Pr(Y_1 = y_1, \cdots, Y_{100} = y_{100}|\theta)$ in a compact form. Also write down the form of $Pr(\sum Y_i = y|\theta)$.

Since Y_i is a binary variable, it follows a Bernoulli distribution with parameter θ. Given θ, $Y_i, \cdots, Y_{100}$ are i.i.d., so we have

$$Pr(Y_1 = y_1, \cdots, Y_{100} = y_{100}|\theta)$$
$$=Pr(Y_1 = y_1|\theta) \cdots Pr(Y_{100} = y_{100}|\theta)$$
$$=\theta^{y_1}(1-\theta)^{1-y_1} \cdots \theta^{y_{100}}(1-\theta)^{1-y_{100}}$$
$$=\Pi_{i=1}^{100}\theta^{y_i}(1-\theta)^{1-y_i}$$
$$=\theta^{\Sigma y_i}(1-\theta)^{100-\Sigma y_i}$$

Obivously, $\sum Y_i \sim Binom(100, \theta)$, so we have

$$Pr(\sum Y_i = y|\theta)$$
$$=\binom{100}{y}\theta^y(1-\theta)^{100-y}$$

    b) For the moment, suppose you believed that $\theta \in \{0.0, 0.1, \cdots, 0.9, 1.0\}$. Given that the results of the survey were $\sum_{i=1}^{100} Y_i = 57$, compute $Pr(\sum Y_i = 57|\theta)$ for each of these 11 values of θ and plot these probabilities as a function of θ.

```
n=100 #sample size
y=57 #sum of yi
#computing
theta_dis = seq(0, 1, by = 0.1)
l_dis = dbinom( y, 100, theta_dis)
df_dis = data.frame(theta=theta_dis, likelihood=l_dis)
l_dis = df_dis$likelihood #prob for each theta
names(l_dis) = c('theta=0',theta_dis[-1])
round(l_dis, 4)
```

```
## theta=0     0.1     0.2     0.3     0.4     0.5     0.6     0.7     0.8
##  0.0000  0.0000  0.0000  0.0000  0.0002  0.0301  0.0667  0.0019  0.0000
##     0.9       1
##  0.0000  0.0000
```

```
#draw plot
ggplot(df_dis,aes(x=theta, y=likelihood)) +
  geom_bar(stat = 'identity') +
  labs(title = '(a) Pr(Y=57|theta)') +
  theme(plot.title = element_text(hjust = 0.5))
```


(a) Pr(Y=57|theta)
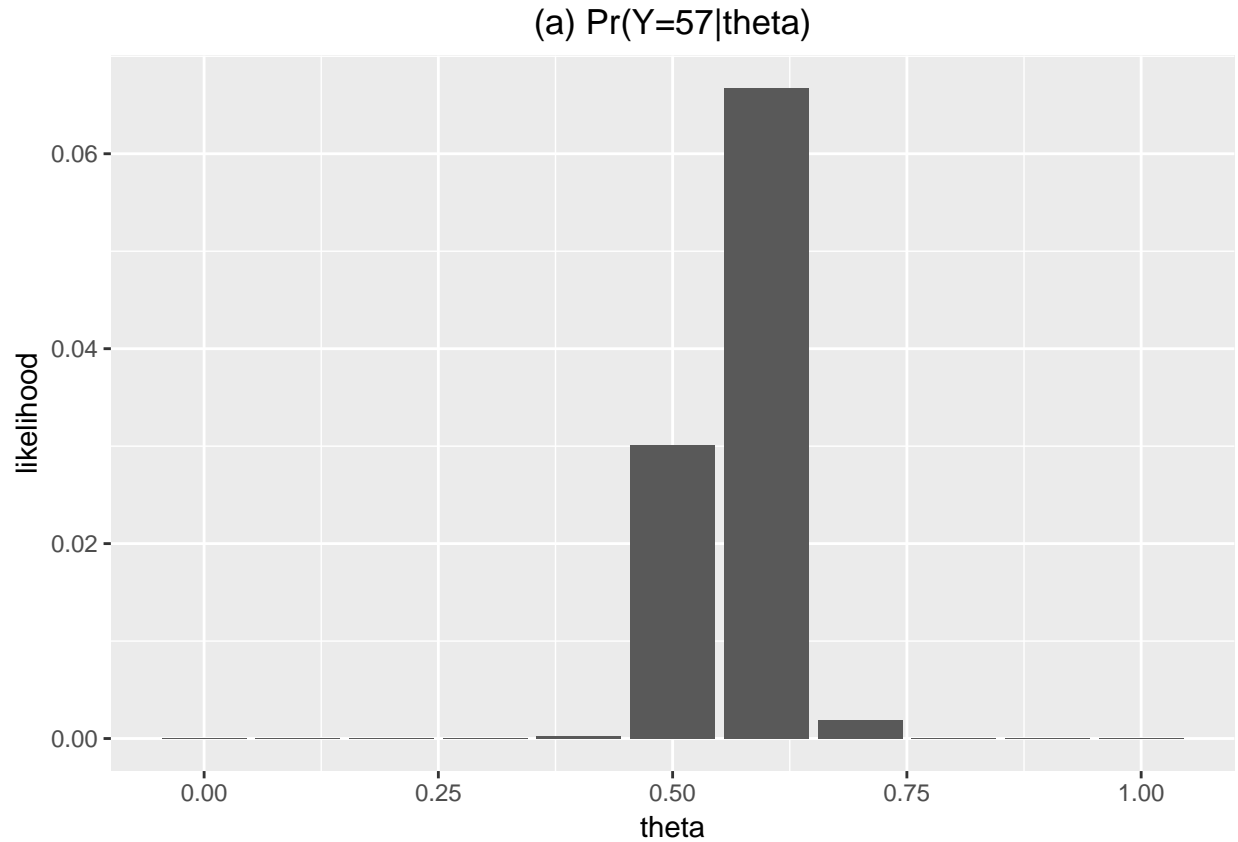
c) Now suppose you originally had no prior information to believe one of these θ-values over another, and so $Pr(\theta = 0.0) = Pr(\theta = 0.1) = \cdots = Pr(\theta = 0.9) = Pr(\theta = 1.0)$. Use Bayes' rule to compute $p(\theta | \sum_{i=1}^{100} Y_i = 57)$ for each θ-value. Make a plot of this posterior distribution as a function of θ.

According to Bayes' rule

$$p(\theta | \sum_{i=1}^{100} Y_i = 57) = \frac{Pr(\theta)Pr(\sum_{i=1}^{100} Y_i = 57|\theta)}{\sum_{\Theta} Pr(\theta)Pr(\sum_{i=1}^{100} Y_i = 57|\theta)}$$

we can compute the posterior probability in R

```
#computing
df_dis=df_dis %>%
  mutate(prior = rep(1/length(theta), length(theta))) %>%
  mutate(posterior = likelihood * prior/sum(likelihood * prior))
pos_dis = df_dis$posterior
```

```
names(pos_dis) = c('theta=0',theta_dis[-1])
round(pos_dis, 4) #posterior prob of each theta
```

```
## theta=0      0.1      0.2      0.3      0.4      0.5      0.6      0.7      0.8
##  0.0000   0.0000   0.0000   0.0000   0.0023   0.3041   0.6749   0.0187   0.0000
##     0.9        1
##  0.0000   0.0000
```

```
#draw plot
ggplot(df_dis, aes(theta, posterior)) +
  geom_bar(stat = 'identity') +
  labs(title = '(b) Posterior pmf of theta') +
  theme(plot.title = element_text(hjust = 0.5))
```


(b) Posterior pmf of theta
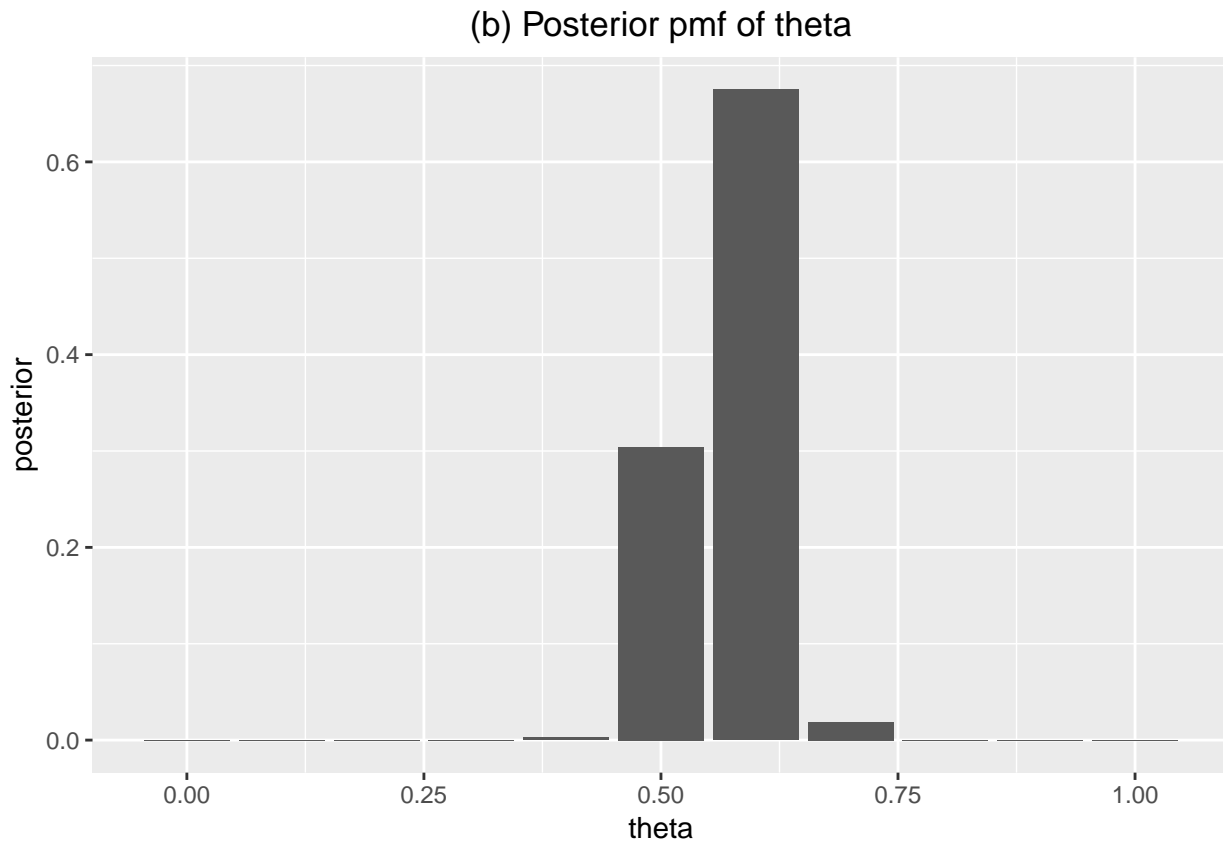
d) Now suppose you allow θ to be any value in the interval $[0, 1]$. Using the uniform prior density for θ, so that $p(\theta) = 1$, plot the posterior density $p(\theta) \times Pr(\sum_{i=1}^{n} Y_i = 57|\theta)$ as a function of θ.

According to Bayes' rule

$$p(\theta | \sum_{i=1}^{100} Y_i = 57)$$

$$= \frac{p(\theta)Pr(\sum_{i=1}^{100} Y_i = 57|\theta)}{\int_\Theta p(\theta)Pr(\sum_{i=1}^{100} Y_i = 57|\theta)d\theta}$$

$$\propto \theta^{57}(1-\theta)^{43}$$

Judging from the kernal of density, the posterior should follow $Beta(58, 44)$

```
#computing
df_cont = data.frame(theta = seq(0, 1, by=0.01))
df_cont = df_cont %>%
  mutate(prior = dbeta(theta, 1, 1)) %>%
  mutate(likelihood = dbinom(y, n, theta)) %>%
  mutate(p_times_l = prior * likelihood)

#draw plot
ggplot(df_cont, aes(theta, p_times_l)) +
  geom_line() +
  labs(title = '(c) p(theta) x Pr (y|theta)') +
  theme(plot.title = element_text(hjust = 0.5))
```


(c) p(theta) x Pr (y|theta)
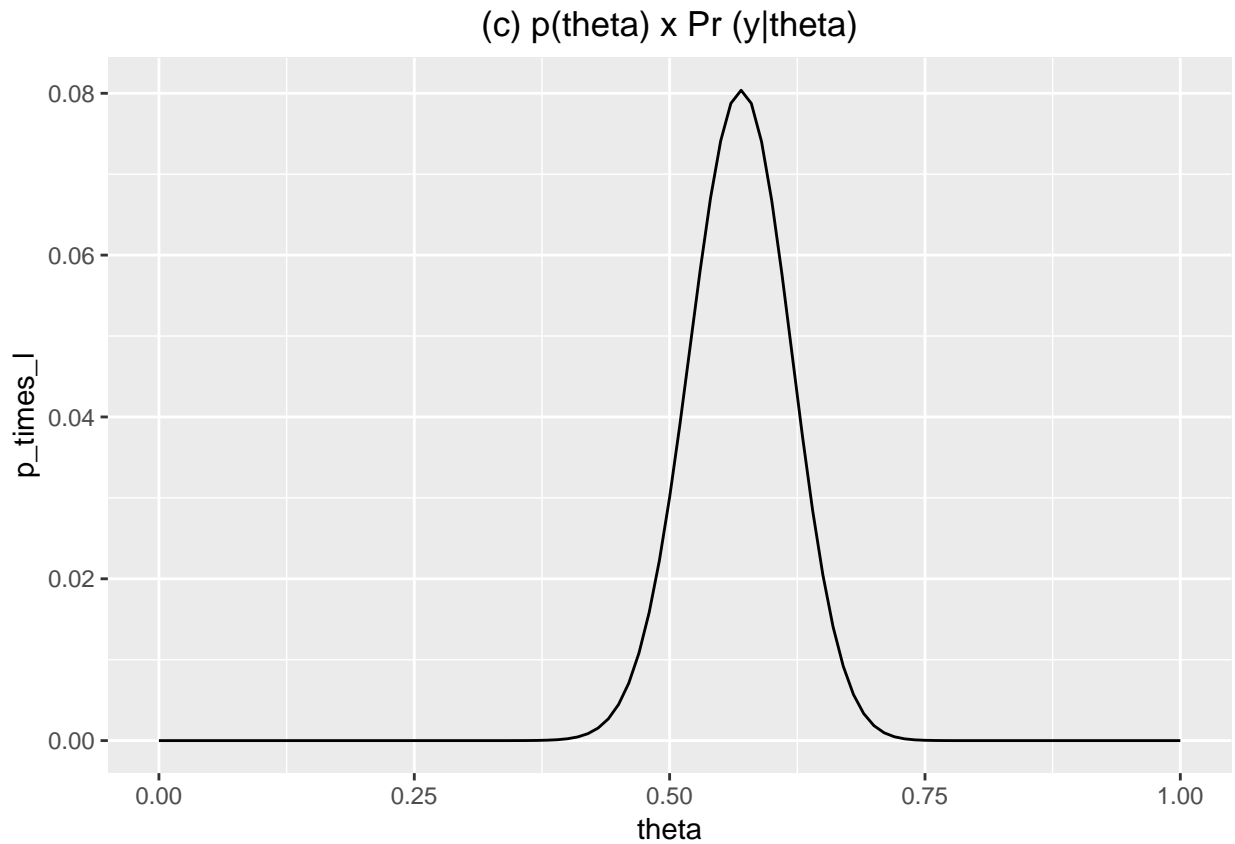
e) As discussed in this chapter, the posterior distribution of θ is $beta(1 + 57, 1 + 100 - 57)$. Plot the posterior density as a function of θ. Discuss the relationships among all of the plots you have made for this exercise.

```
#computing
a = 1 + y # parameter 1 of posterior
b = 1 + n - y # parameter 2 of posterior
df_cont = df_cont %>%
  mutate(posterior = dbeta(theta, a, b))

#draw plot
ggplot(df_cont, aes(theta, posterior)) +
  geom_line() +
  labs(title = '(d) Posterior pdf of theta') +
  theme(plot.title = element_text(hjust = 0.5))
```


(d) Posterior pdf of theta
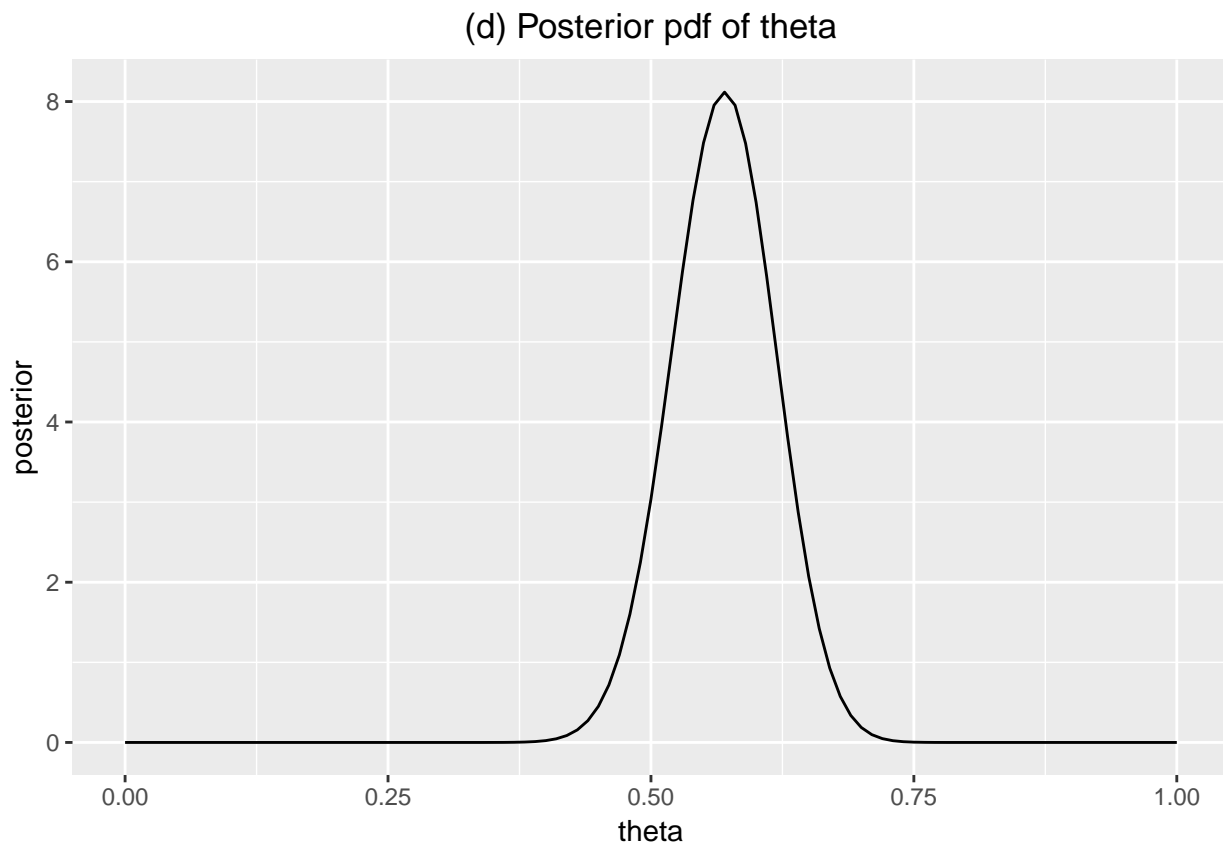
Plot (a) and (b) are both based on the prior belief that $\theta \in \{0.0, 0.1, \cdots, 0.9, 1.0\}$. Plot (a) is the probability of $\sum Y_i = 57$ given each θ. Plot (b) is the posterior pmf, which is derived by updating our prior belief using the information in plot (a) (observed data). Specifically, we need to use Bayes' Rule to calculate the posterior probability of each θ.

Plot (c) and (d) are both based on the prior belief $\theta \sim Uniform(0, 1)$. In plot (c), the value of function $p(\theta) \times Pr(\sum_{i=1}^{n} Y_i = 57|\theta)$ indicates our posterior belief of θ, but it is not a posterior pdf since the integral over the support of θ is not equal to 1. However, we can simply get the result of plot (d) by normalizing the function in plot (c) to make its integral equal to 1. Specifically, we can dividing the values in plot (c) with $\int_{\Theta} p(\theta) \times Pr(\sum_{i=1}^{n} Y_i = 57|\theta)d\theta$. Plot (d) is the posterior pdf of θ.

3.2

Sensitivity analysis: It is sometimes useful to express the parameters a and b in a beta distribution in terms of $\beta_0 = a/(a+b)$ and $n_0 = a+b$, so that $a = \theta_0 n_0$ and $b = (1-\theta_0)n_0$. Reconsidering the sample survey data in Exercise 3.1, for each combination of $\theta_0 \in \{0.1, 0.2, \cdots, 0.9\}$ and $n_0 \in \{1, 2, 8, 16, 32\}$ find the corresponding a, b values and compute $Pr(\theta > 0.5 | \sum Y_i = 57)$ using a $beta(a,b)$ prior distribution for θ. Display the results with a contour plot, and discuss how the plot could be used to explain to someone whether or not they should believe that $\theta > 0.5$, based on the data that $\sum_{i=1}^{100} Y_i = 57$.

```
theta0 = seq(0.1, 0.9, by=0.1) #prior mean
n0 = c(1, 2, 8, 16, 32) #prior 'sample size'
a = n0 %*% t(theta0) #prior parameter 1
b = n0 %*% t(1 - theta0) #prior parameter 2
cname = c('theta0=0.1',seq(0.2, 0.9, by = 0.1))
rname = c('n0=1',2,8,16,32)
colnames(a) = cname
rownames(a) = rname
colnames(b) = cname
rownames(b) = rname
kable(a, caption = 'values of a') #display of a
```

Table 1: values of a

	theta0=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
n0=1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8
8	0.8	1.6	2.4	3.2	4.0	4.8	5.6	6.4	7.2
16	1.6	3.2	4.8	6.4	8.0	9.6	11.2	12.8	14.4
32	3.2	6.4	9.6	12.8	16.0	19.2	22.4	25.6	28.8

```
kable(b, caption = 'values of b') #display of b
```

Table 2: values of b

	theta0=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
n0=1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
2	1.8	1.6	1.4	1.2	1.0	0.8	0.6	0.4	0.2
8	7.2	6.4	5.6	4.8	4.0	3.2	2.4	1.6	0.8
16	14.4	12.8	11.2	9.6	8.0	6.4	4.8	3.2	1.6
32	28.8	25.6	22.4	19.2	16.0	12.8	9.6	6.4	3.2

```
z=pbeta(0.5, a+y, b+n-y, lower.tail = FALSE) #posterior prob of theta>0.5
contour(n0,theta0,z,col=gray(0.3),
        lwd=2, main = 'Contour of Pr(theta>0.5|Y=57)',
        xlab ='n0', ylab = 'theta0')
```

Contour of Pr(theta>0.5|Y=57)
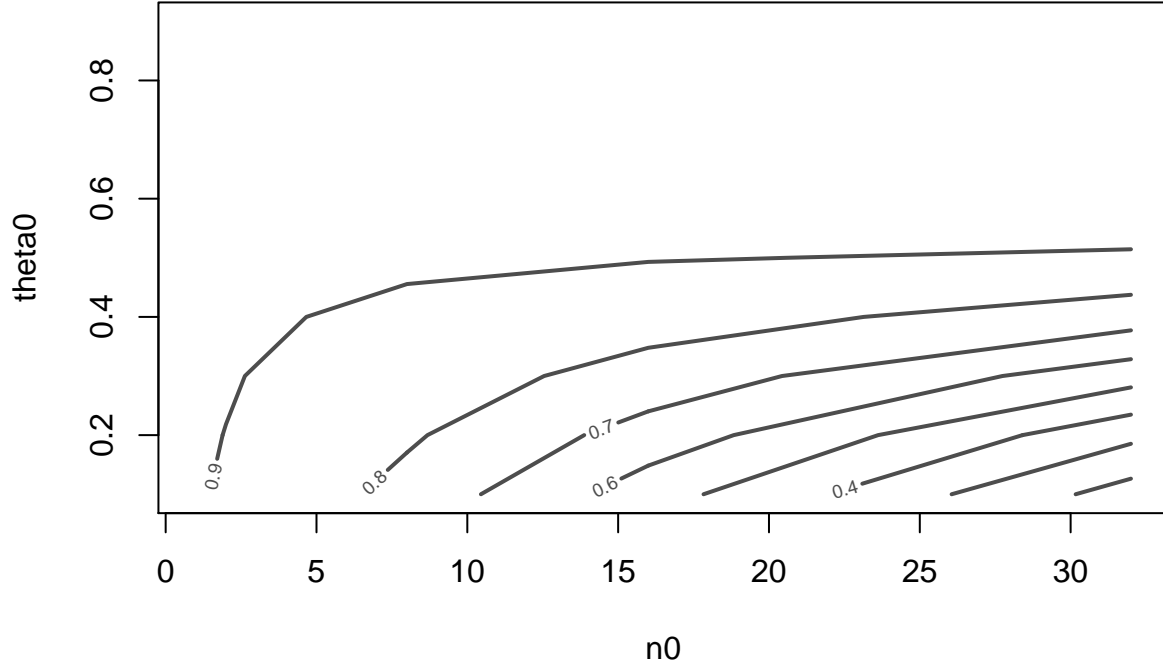
As the contour shows, when our prior weight, n_0 is very large and prior mean, θ_0 is small, the the posterior probability of $Pr(\theta > 0.5 | \sum Y_i = 57)$ is small and we should not believe $\theta > 0.5$. However, as long as our prior belief θ_0 is not too small, $Pr(\theta > 0.5 | \sum Y_i = 57)$ will be quite large and we should believe that $\theta > 0.5$ given $\sum Y_i = 57$.

Basically because the sample mean $\bar{Y} = 0.57$ is larger than 0.5, so either when our prior belief $\theta_0 > 0.5$, then our posterior belief will definitely be in favor of $\theta > 0.5$, or when our prior belief $\theta_0 < 0.5$, then only a small θ_0 and large n_0 may drive us away from believing $\theta > 0.5$ given $\sum Y_i = 57$.

Math Problem

Verify the following

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)} \tag{1}$$

where $B(a, b) = \int_0^1 \theta^{a-1}(1 - \theta)^{b-1}d\theta$ and $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx$.

Proof 1:

$$\Gamma(a)\Gamma(b) = \int_0^\infty t^{a-1}e^{-t}dt \int_0^\infty s^{a-1}e^{-s}ds$$

Let $t = z\theta$, $s = z(1-\theta)$, then $t + s = z$. Since $t \in (0,\infty)$, $s \in (0,\infty)$ so $z \in (0,\infty)$, $\theta \in (0,1)$.

$$J(z,\theta) = \begin{vmatrix} \frac{\partial t}{\partial z} & \frac{\partial t}{\partial \theta} \\ \frac{\partial s}{\partial z} & \frac{\partial s}{\partial \theta} \end{vmatrix} = -z$$

$$\Gamma(a)\Gamma(b)$$
$$= \int_0^\infty t^{a-1}e^{-t}dt \int_0^\infty s^{b-1}e^{-s}ds$$
$$= \int_0^1 \int_0^\infty (z\theta)^{a-1}(z(1-\theta))^{b-1}e^{-z}|J|dzd\theta$$
$$= \int_0^\infty z^{a+b-1}e^{-z}dz \int_0^1 \theta^{a-1}(1-\theta)^{b-1}d\theta$$
$$= \Gamma(a+b)B(a,b)$$

Thus, we have proved $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Proof 2:

Let $x = t^2$, $t > 0$, $x \in (0,\infty)$, so $t \in (0,\infty)$. By substituting x with t^2, we can have the equivalent expression of $\Gamma(a)$

$$\Gamma(a) = \int_0^\infty t^{2a-2}e^{-t^2}dt^2 = 2\int_0^\infty t^{2a-1}e^{-t^2}dt \tag{2}$$

Using the form of $\Gamma(a)$ in (2), then the right hand side of (1) is

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \frac{4\int_0^\infty x^{2a-1}e^{-x^2}dx \int_0^\infty y^{2b-1}e^{-y^2}dy}{2\int_0^\infty x^{2a+2b-1}e^{-x^2}dx} = \frac{2\int_0^\infty \int_0^\infty x^{2a-1}y^{2b-1}e^{-x^2-y^2}dxdy}{\int_0^\infty x^{2a+2b-1}e^{-x^2}dx}$$

Let $x = r\cos(\theta)$, $y = r\sin(\theta)$, $\theta \in [0,2\pi]$, $r \in (0,\infty)$, then $J(r,\theta) = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = r$

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$
$$= \frac{2\int_0^{2\pi}\int_0^\infty (r\cos\theta)^{2a-1}(r\sin\theta)^{2b-1}e^{-r^2}|J|drd\theta}{\int_0^\infty x^{2a+2b-1}e^{-x^2}dx}$$
$$= 2\int_0^{2\pi}(\cos\theta)^{2a-1}(\sin\theta)^{2b-1}d\theta \frac{\int_0^\infty r^{2a+2b-1}e^{-r^2}dr}{\int_0^\infty x^{2a+2b-1}e^{-x^2}dx}$$
$$= 2\int_0^{2\pi}(\cos\theta)^{2a-1}(\sin\theta)^{2b-1}d\theta$$
$$= \int_0^1 (\cos^2\theta)^{a-1}(\sin^2\theta)^{b-1}d(\cos^2\theta)$$
$$= \int_0^1 z^{a-1}(1-z)^{b-1}dz \qquad (\text{let } z = \cos^2\theta)$$
$$= B(a,b)$$