STA 601 Homework 2 Solutions

Solutions provided by STA 601 TAs

22 September, 2019

Hoff exercises

3.1

a)

Since $Y_1, \ldots, Y_{100}$ are conditionally i.i.d. given θ, we can write their joint conditional distribution as a product of the individual conditional distributions:

$$\begin{aligned}
\Pr(Y_i = y_i|\theta) = \theta^{y_i}(1-\theta)^{1-y_i} &\implies \Pr(Y_1 = y_1, \ldots, Y_{100} = y_{100}|\theta) = \prod_{i=1}^{100} \theta^{y_i}(1-\theta)^{1-y_i} \\
&\implies \Pr(Y_1 = y_1, \ldots, Y_{100} = y_{100}|\theta) = \theta^{\sum_{i=1}^{100} y_i}(1-\theta)^{100-\sum_{i=1}^{100} y_i}
\end{aligned}$$

The probability of observing a particular (ordered) sum of the Y_i's is the same as the probability of observing any one of the corresponding joint outcomes. However, when we consider unordered sums (i.e. when we care about the total number of supporters, but not the identities of those who supported and those who didn't), we need to rescale each of the joint outcome probabilities according to the number of ways we could possibly observe y supporters of the policy out of the 100 individuals. For instance, there are $\binom{100}{2}$ ways we might observe $\sum_i Y_i = 2$, but only 100 ways we could observe $\sum_i Y_i = 1$.

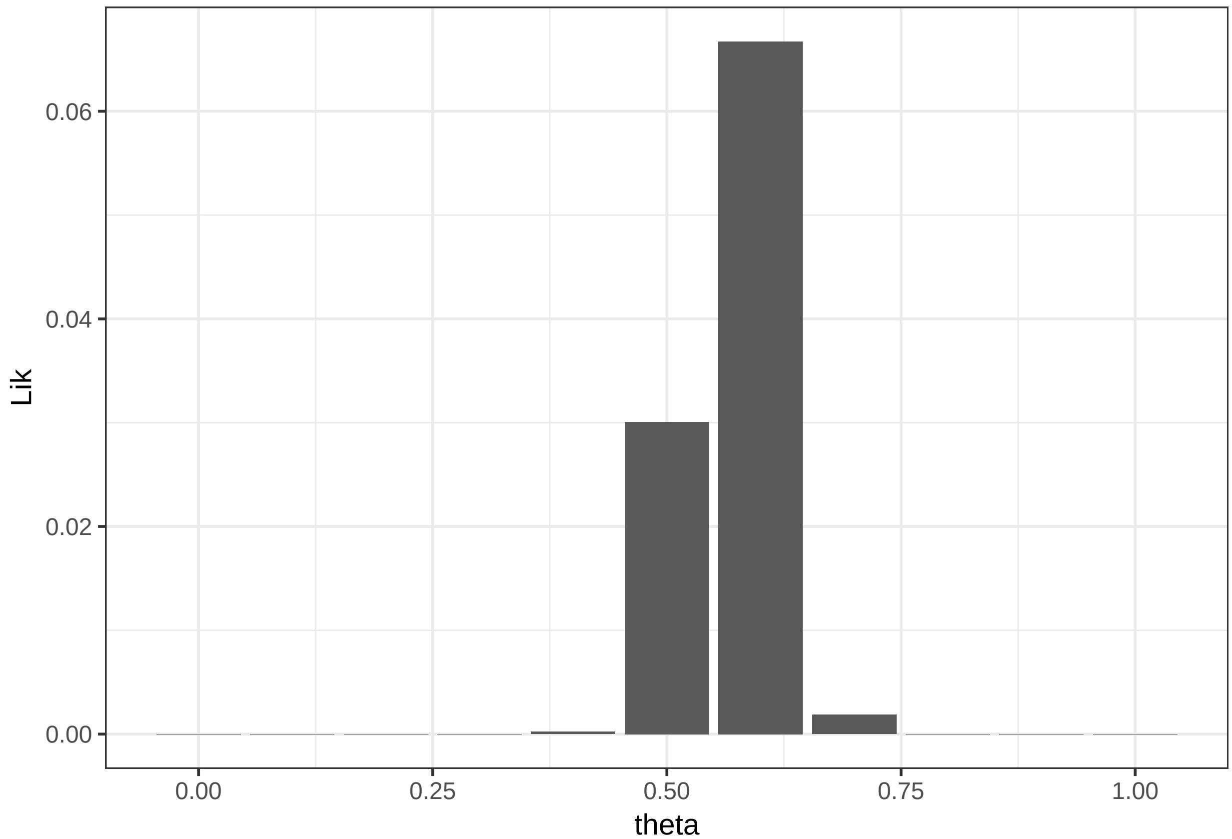
Thus,

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

which we recognize as the Binomial$(100, \theta)$ distribution.

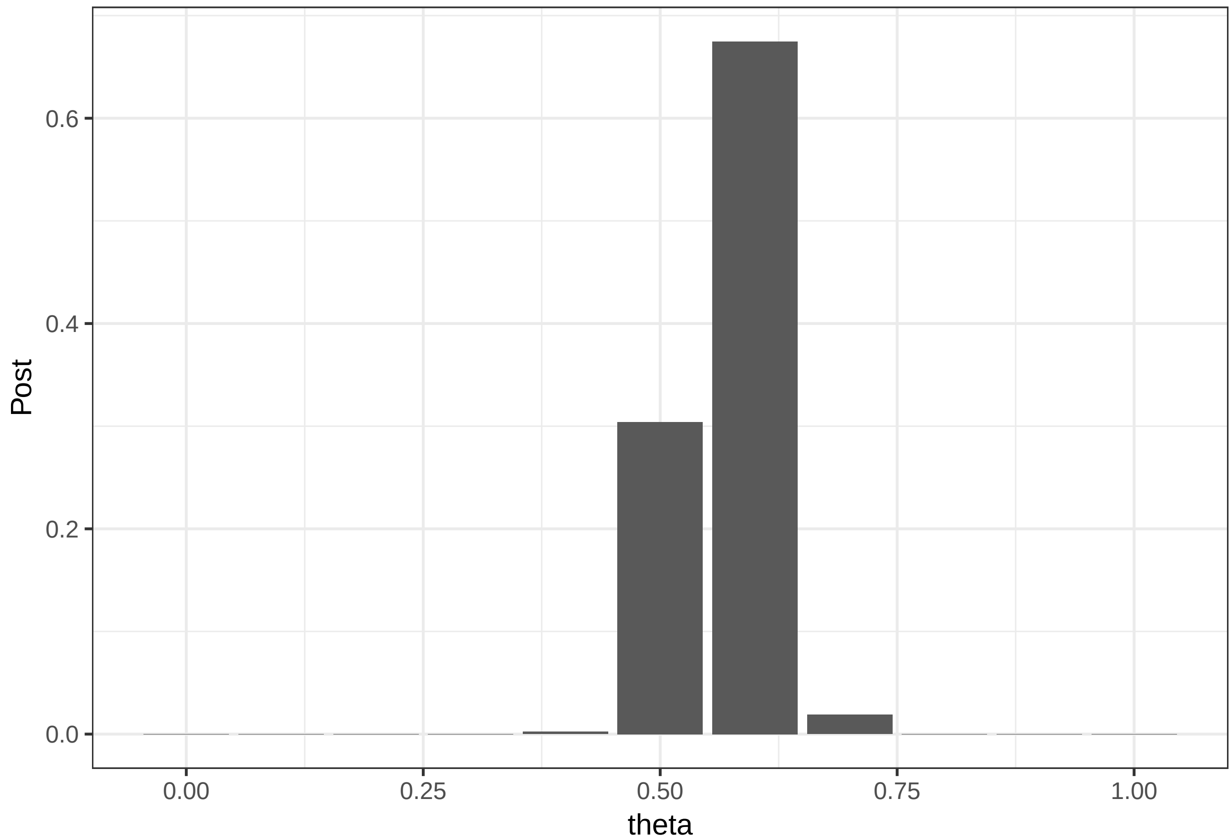
b)

```r
# Create theta test values
theta <- seq(0, 1, by = 0.1)
# Binomial PMF function
binom_fun <- function(n, y, theta){choose(n, y)*theta^(y)*(1-theta)^(n - y)}
# Calculate probabilities for each theta
theta_lik <- theta %>% plyr::aaply(1, function(t){binom_fun(100, 57, t)})
# Plot results
data.frame(theta = theta,
           Lik = theta_lik) %>%
  ggplot2::ggplot() +
  geom_bar(aes(x=theta, y=Lik), stat="identity")
```



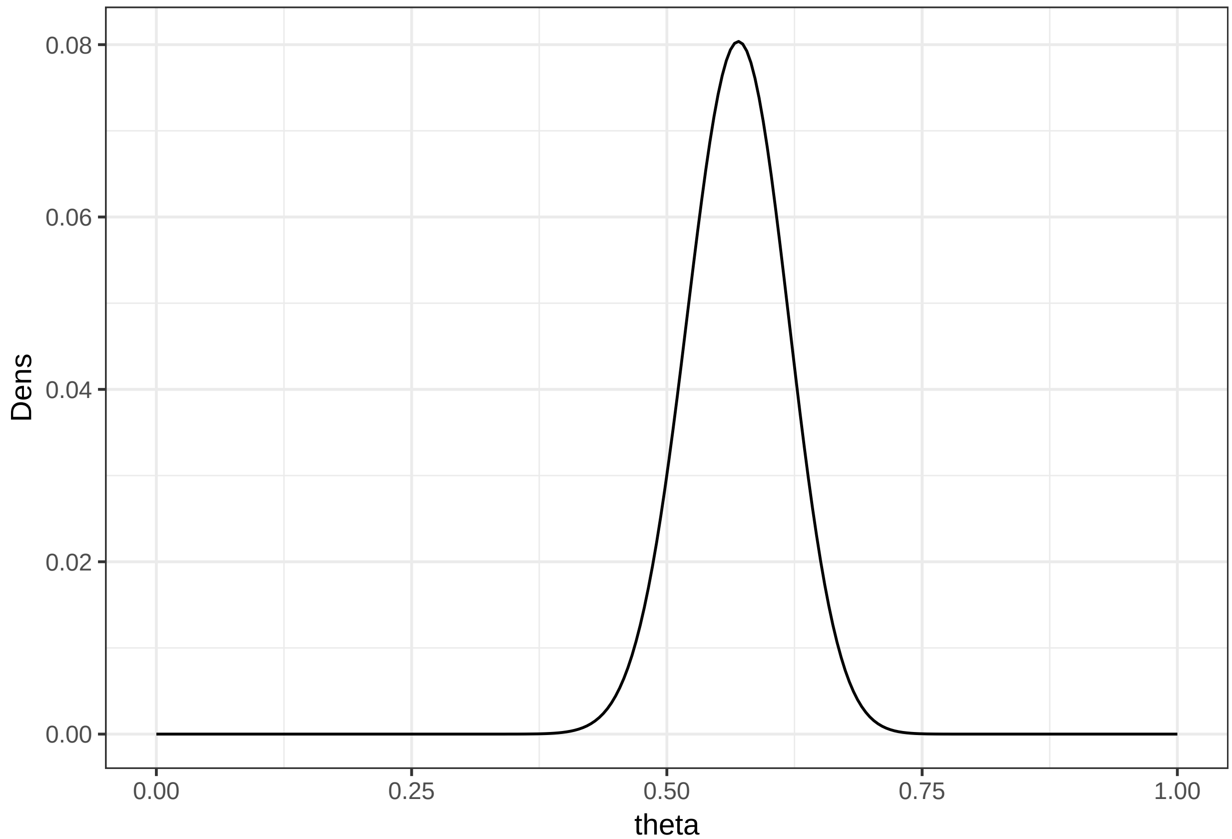
c)

```
# Calculate posterior using Bayes' Rule
# Since outcomes have equal probability, we can cancel the prior
# in the numerator and denominator
theta_post <- theta_lik / sum(theta_lik)
# Plot results
data.frame(theta = theta,
           Post = theta_post) %>%
  ggplot2::ggplot() +
  geom_bar(aes(x=theta, y=Post), stat="identity")
```



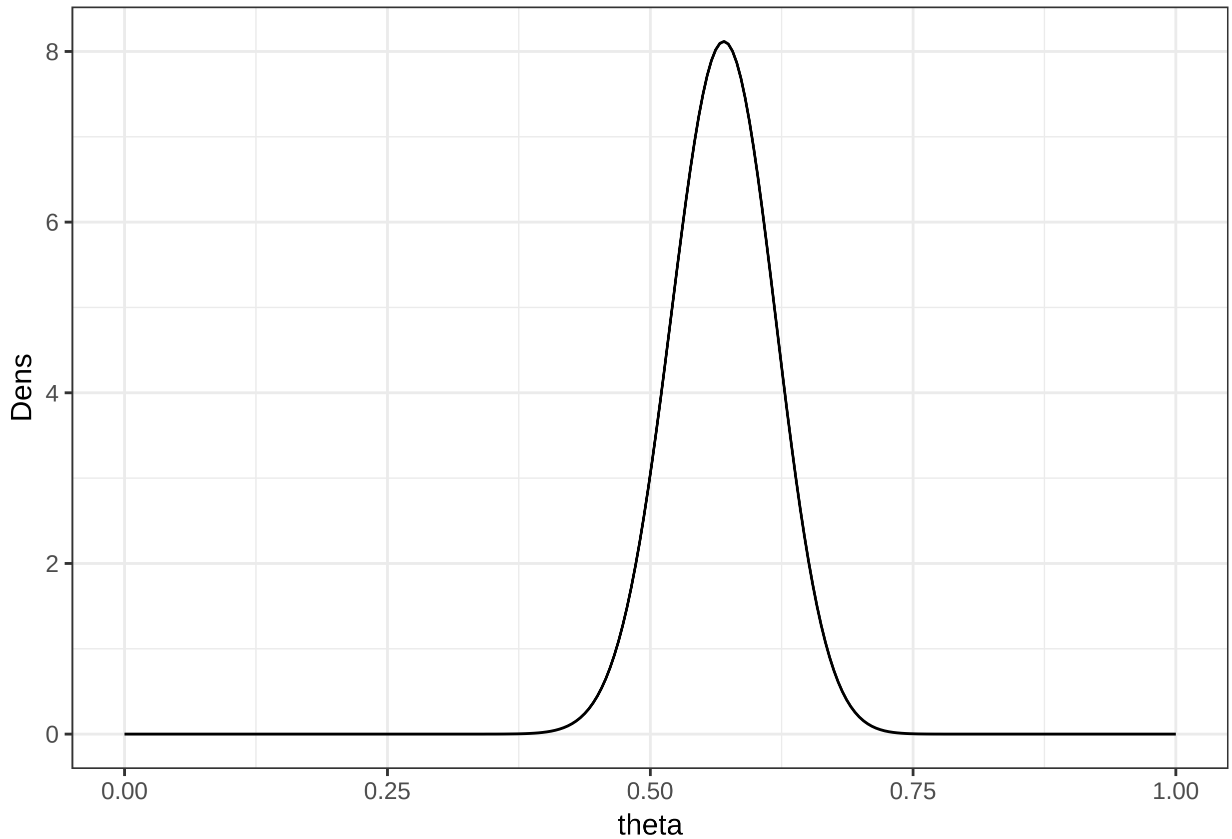
d)

```
# Generate many thetas uniformly in [0,1]
unif_theta <- seq(0, 1, length.out = 250)
# Calculate posterior density with p(theta) = 1
post_dens <- binom_fun(100, 57, unif_theta)
# Plot results
data.frame(theta = unif_theta,
           Dens = post_dens) %>%
  ggplot2::ggplot() +
  geom_line(aes(x=theta, y=Dens))
```



e)

```
# Verify that posterior is the beta
beta_dist <- dbeta(unif_theta, 1 + 57, 1 + 100 - 57)
# Plot results
data.frame(theta = unif_theta,
           Dens = beta_dist) %>%
  ggplot2::ggplot() +
  geom_line(aes(x=theta, y=Dens))
```



Plot discussion

In part (b) we plotted values from the data-generating model. This is alternatively referred to as the data likelihood given θ or the sampling distribution of the data given θ. In part (c), we plotted the proper posterior distribution given a discrete equiprobable prior on θ, which was proportional to the data likelihood given θ.

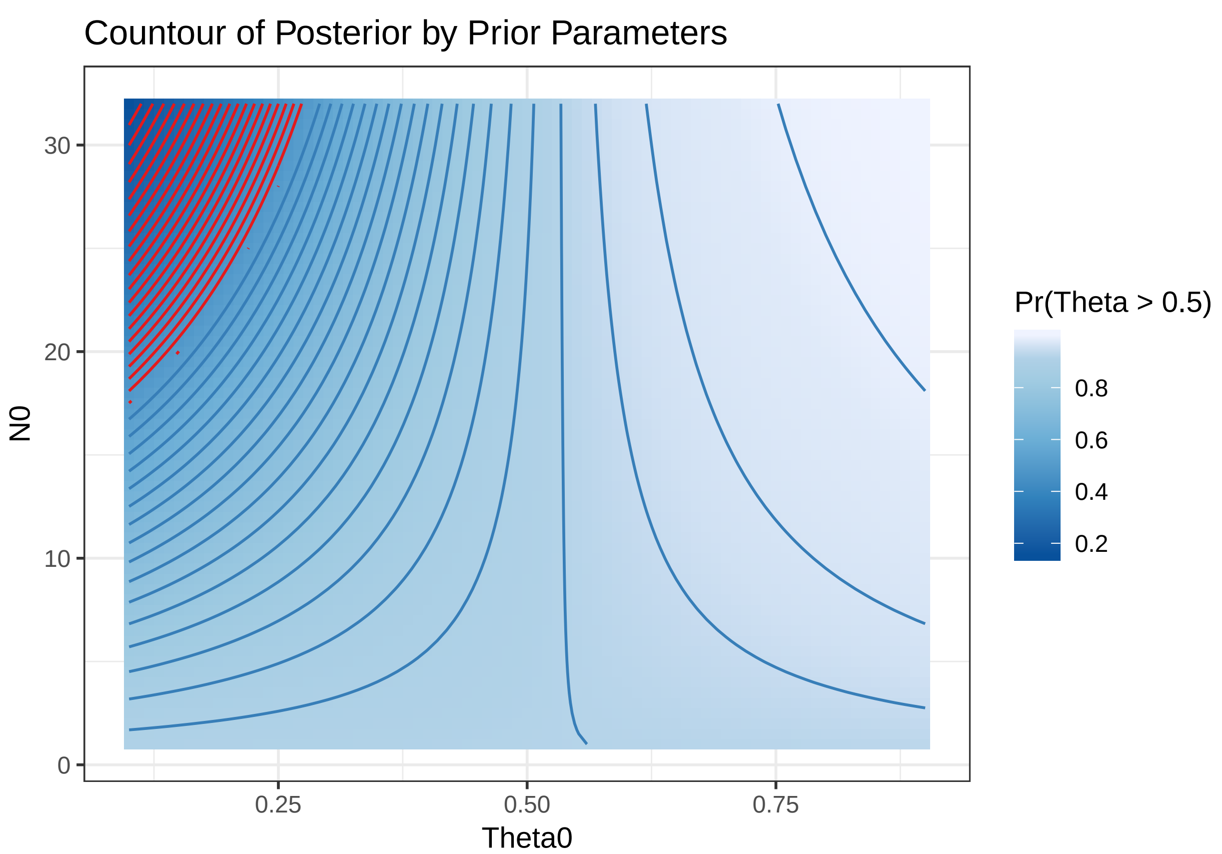
In part (d) we plotted the *unnormalized* improper posterior distribution of θ given our observed data under a continuous uniform prior for θ. Since $p(\theta)$ was equal to 1, this was essentially the continuous analog of what we did in part (b), i.e. plotting the likelihood. Part (e) displays the proper normalized posterior for θ given the data, which is proportional, but not equal to the likelihood plotted in part (d). Part (e) is the continuous analog of part (c).

3.2

```
# Define theta_0 and n_0
theta_0 <- seq(0.1, 0.9, by=0.01)
n_0 <- seq(1, 32, by=0.5)
# Generate combinations of theta_0 and n_0; calculate a,b
df <- expand.grid(theta_0 = theta_0, n_0 = n_0) %>%
        dplyr::mutate(a = n_0*theta_0,
                      b = n_0 - a)
# Find posterior distributions
p_df <- df %>%
            plyr::ddply(.(theta_0, n_0, a, b), function(x){
              a <- x$a
```

```r
                b <- x$b
                posterior <- (1 - pbeta(0.5, a + 57, b + 100 - 57))
                return(data.frame(p=posterior))
              })

p_df %>%
  ggplot2::ggplot() +
  geom_tile(aes(x=theta_0, y=n_0, fill=p)) +
  geom_contour(aes(x=theta_0, y=n_0, z=p, colour = p > 0.5),
               bins=20) +
  scale_fill_distiller(values=c(0,0.1,0.2,
                                0.3, 0.4, 0.5,
                                0.6, 0.7, 0.8,
                                0.9, 0.925, 0.95,
                                0.99,1.0),
                       name = "Pr(Theta > 0.5)") +
  scale_colour_brewer(guide=F, palette = "Set1") +
  labs(x="Theta0", y="N0",
       title="Countour of Posterior by Prior Parameters")
```



I would explain to someone viewing this plot that unless their prior beliefs about the proportion of supporters of this policy were very strong, it would be a better bet to say that $\theta > 0.5$ given the data we observed. If this person had weak or even moderate prior information (the regions around the dark blue contour lines), then the probability of $\theta > 0.5$ would be fairly high.

If this person's prior belief was very strong – roughly equivalent to having observed a hypothetical previous sample of 25 people for which the proportion of policy-supporters was less than a quarter (red-colored contours) – then they should bet on $\theta < 0.5$. Otherwise, $\theta > 0.5$ has higher probability than $\theta < 0.5$.