Week 3: Intro to Bayes

Kishore Basu

28/01/23

```
library(ggplot2)
```

Question 1

Consider the happiness example from the lecture, with 118 out of 129 women indicating they are happy. We are interested in estimating θ, which is the (true) proportion of women who are happy. Calculate the MLE estimate $\hat{\theta}$ and 95% confidence interval.

We know that $Y|\theta\ Bin(129, \theta)$. So using maximum likelihood estimation, we can estimate the likelihood as

$$L(y;\theta) = \binom{n}{y}\ \theta^y(1-\theta)^{n-y}$$

and so our log-likelihood is given by

$$l(y;\theta) = \log\{\binom{n}{y}\ \theta^y(1-\theta)^{n-y}\} = \log\binom{n}{y} + y\log\theta + (n-y)\log(1-\theta),$$

differentiating to get the score function...

$$\frac{dl}{d\theta} = \frac{y}{\theta} - \frac{n-y}{1-\theta}$$

setting equal to zero, we solve for $\hat{\theta}$

$$0 = \frac{y}{\hat{\theta}} - \frac{n-y}{1-\hat{\theta}} 0 = (1-\hat{\theta})y - \hat{\theta}n + \hat{\theta}y\hat{\theta} = \frac{y}{n} = \frac{118}{129}$$

Thus, our MLE is about 0.914. To find a confidence interval, we need the variance so we take the derivative of the score function, and take the expectation to get $I(\theta)$.

$$\frac{d^2l}{d\theta^2} = -\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2} I(\theta) = -E_\theta(\frac{d^2l}{d\theta^2}) = \frac{\theta}{\theta^2} + \frac{n-\theta}{(1-\theta)^2}$$

since we know that $E(y) = \theta$. We can just invert and simplify to get the variance, subbing in $\hat{\theta}$.

$$Var(\hat{\theta}) = I(\hat{\theta})^{-1} = \left(\frac{1}{\hat{\theta}} + \frac{n-\hat{\theta}}{(1-\hat{\theta})^2}\right)^{-1} = \left(\frac{n}{\hat{\theta}(1-\hat{\theta})}\right)^{-1} Var(\hat{\theta}) = \frac{\hat{\theta}(1-\hat{\theta})}{n}$$

So finding the confidence interval is easy. Due to asymptotic normality, we use a normal distribution to get a confidence interval

```
theta_hat = 118/129
n = 129
sd_th = sqrt(theta_hat*(1-theta_hat)/n)
CI <- c(theta_hat - 1.96*sd_th, theta_hat + 1.96*sd_th)
CI
```

```
[1] 0.8665329 0.9629244
```

Question 2

Assume a Beta(1,1) prior on θ. Calculate the posterior mean for $\hat{\theta}$ and 95% credible interval.

Note that a Beta(1,1) prior has a density equal to 1. So our prior distribution is $p(\theta) = 1$. We calculate the posterior.

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)}{p(y)}$$

The denominator is a constant so we only look at proportionality to recognize the distribution. Thus,

$$p(\theta|y) \propto p(y|\theta) = \theta^y(1-\theta)^{n-y}$$

which is reminiscent of a beta distribution (since θ is the random quantity now) with parameters $\alpha = y+1$, $\beta = n-y+1$. We know that the expected value of a beta distribution is $\alpha/(\alpha+\beta)$, so we get

$$E(p(\theta|y)) = \frac{y+1}{n+2}$$

To calculate a credible interval, we can just take the quantiles of our beta distribution

```r
n = 129
y = 118
c(qbeta(0.025, shape1 = y + 1, shape2 = n - y + 1),
  qbeta(0.975, shape1 = y + 1, shape2 = n - y + 1)
)
```

```
[1] 0.8536434 0.9513891
```

Question 3

Now assume a Beta(10,10) prior on θ. What is the interpretation of this prior? Are we assuming we know more, less or the same amount of information as the prior used in Question 2?

If we assume a Beta(10,10) prior on θ, the interpretation is that there are the same number of successes (9) and failures (9). However, this is not necessarily the same, where we found equal number of successes and failures in the Beta(1,1) distribution. Since here we actually have information about the number of successes and failures as before, we are assuming to know more information than before. That is, we actually know that we have had some success and some failure, as opposed to Beta(1,1), where we have not observed anything at all.

Again, the prior can be found by way of proportionality. We know that

$$p(\theta|y) \propto p(y|\theta)p(\theta) = \theta^y(1-\theta)^{n-y}\theta^9(1-\theta)^9, p(\theta|y) \propto \theta^{y+9}(1-\theta)^{n-y+9}$$

which is clearly another Beta distribution when we account for the normalizing factor. In particular, it is a Beta distribution with Beta($\alpha = y + 10$, $\beta = n - y + 10$).

Question 4

Create a graph in ggplot which illustrates

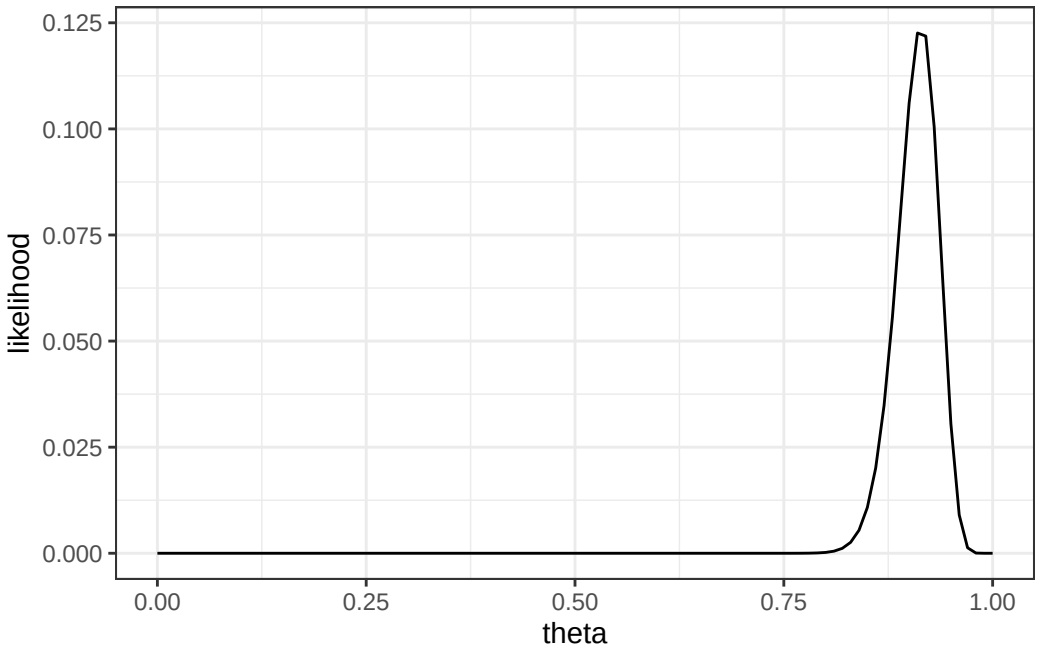
- The likelihood (easiest option is probably to use `geom_histogram` to plot the histogram of appropriate random variables)
- The priors and posteriors in question 2 and 3 (use `stat_function` to plot these distributions)

Comment on what you observe.

First, we plot the likelihood $L(y;\theta) = \binom{129}{118}\ \theta^{118}(1-\theta)^{11}$

```r
# Plot the likelihood
theta <- seq(0,1,by = 0.01)
l <- function(theta){
  return(choose(n, y) * theta^(y)*(1-theta)^(n-y))
}

df <- data.frame(theta = theta, likelihood = l(theta))
ggplot(data = df, aes(x = theta, y = likelihood)) +
  geom_line() + theme_bw()
```

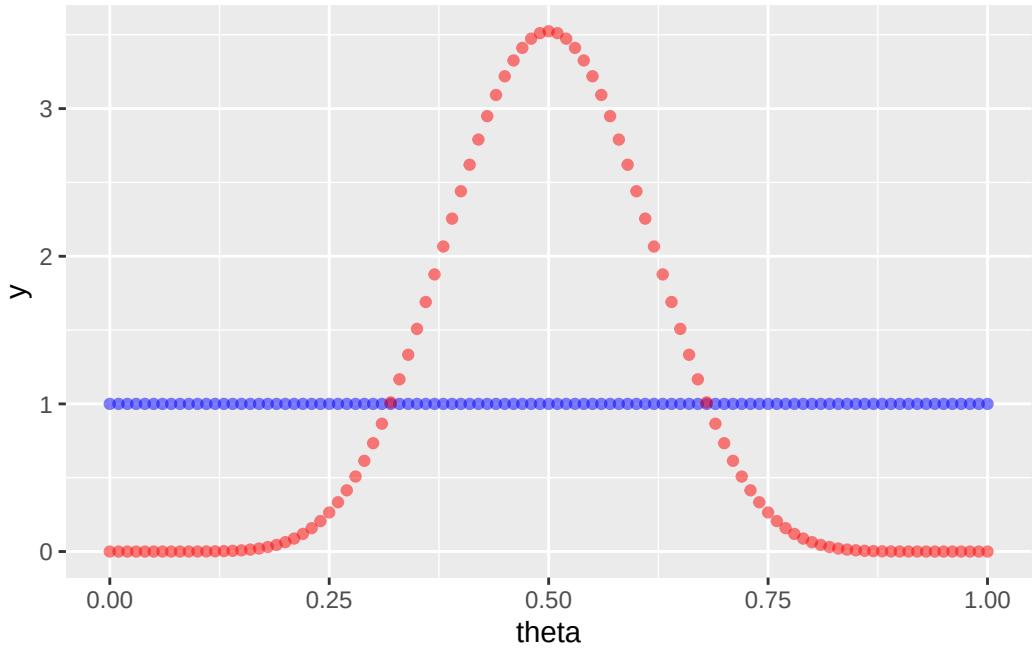


Then, we plot the two priors.

```r
df2 <- data.frame(theta = theta)
ggplot(data = df2, aes(x = theta)) +
  stat_function(fun = dbeta, args = c(1,1),
                geom = "point", color = "blue",
                fill = "blue", alpha = 0.5, lab = 'Prior 1') +
  stat_function(fun = dbeta, args = c(10,10),
                geom = "point", color = "red",
                fill = "red", alpha = 0.5, lab = 'Prior 2')
```

```
Warning in stat_function(fun = dbeta, args = c(1, 1), geom = "point", color =
"blue", : Ignoring unknown parameters: `lab`
```

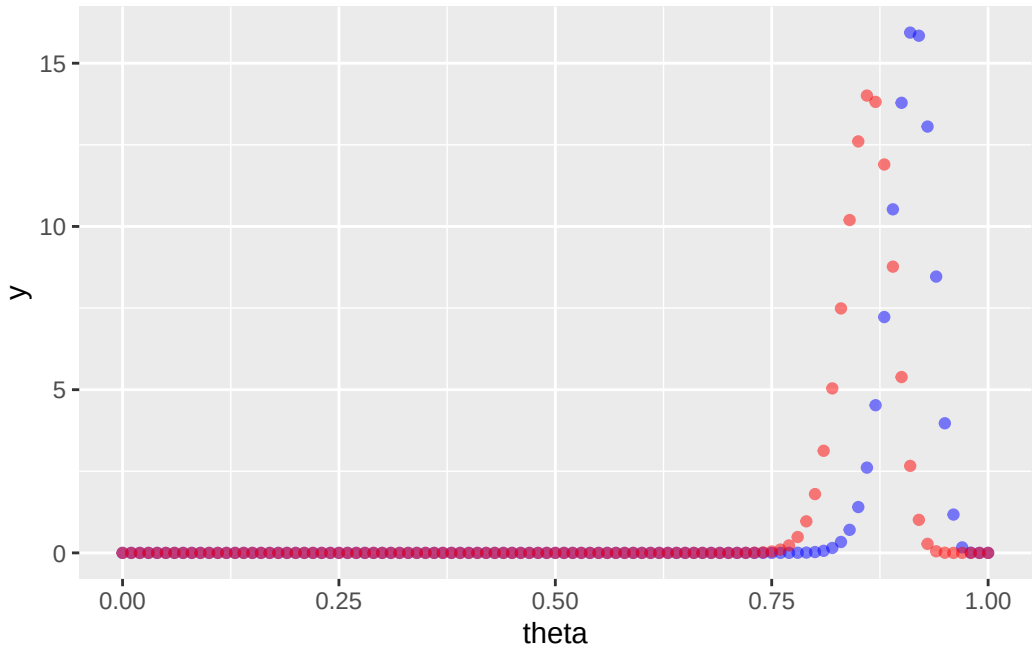
```
Warning in stat_function(fun = dbeta, args = c(10, 10), geom = "point", :
Ignoring unknown parameters: `lab`
```



where the first prior is in blue and the second is in red.

Now to plot the two posteriors:

```
df2 <- data.frame(theta = theta)
ggplot(data = df2, aes(x = theta)) +
  stat_function(fun = dbeta, args = c(y+1,n-y+1),
                geom = "point", color = "blue",
                fill = "blue", alpha = 0.5) +
  stat_function(fun = dbeta, args = c(y + 10,n - y + 10),
                geom = "point", color = "red",
                fill = "red", alpha = 0.5)
```



Again, red reflects the Beta(10,10) prior, and blue represents the Beta(1,1) prior. So when we have a prior of Beta(10,10), our distribution has less density near 1. However, with a 'uninformative' prior, our posterior is closer to 1. So when we don't include much prior knowledge, our model tells us that females are happier than if we have prior knowledge that they are mostly around $\theta = 0.5$. Again, this makes sense as now we are assuming that at baseline it is 50/50, imparting a prior bias.

Question 5

(No R code required) A study is performed to estimate the effect of a simple training program on basketball free-throw shooting. A random sample of 100 college students is recruited into the study. Each student first shoots 100 free-throws to establish a baseline success probability. Each student then takes 50 practice shots each day for a month. At the end of that time, each student takes 100 shots for a final measurement. Let θ be the average improvement in success probability. θ is measured as the final proportion of shots made minus the initial proportion of shots made.

Given two prior distributions for θ (explaining each in a sentence):

- A noninformative prior

A noninformative prior would place equal density on all outcomes, even the (admittedly unlikely) scenario that players gets worse - this could be encapsulated by a Uniform(-1,1) prior.

- A subjective/informative prior based on your best knowledge

Players are more likely to get better than worse, so positive values of theta are more likely, so I could do a Normal(0.4, 0.2) prior (note that this makes values that are impossible such as theta $>$ 1 or theta $<$ -1 occur with very small probability).