

STA2201_Lab3

Alice Huang

25/01/2023

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 assume that the number of women who are happy, Y follows a Binomial distribution. $Y \sim \text{Binom}(n, \theta)$. Here $n = 129$.

The likelihood function is $\prod \binom{n}{x_i} \theta^{x_i} (1 - \theta)^{n - x_i} = (\prod \binom{n}{x_i}) \theta^{n\bar{x}} (1 - \theta)^{n^2 - n\bar{x}}$

The log-likelihood is $\ell = \log(\prod \binom{n}{x_i}) + n\bar{x} \log(\theta) + (n^2 - n\bar{x}) \log(1 - \theta)$

The score function is $\frac{\partial \ell}{\partial \theta} = n\bar{x} \frac{1}{\theta} + (n^2 - n\bar{x}) \frac{1}{1 - \theta} (-1)$

$$\frac{\partial \ell}{\partial \theta} = 0 \iff n\bar{x} \frac{1}{\theta} = (n^2 - n\bar{x}) \frac{1}{1 - \theta}$$

$$\bar{x} \frac{1}{\theta} = (n - \bar{x}) \frac{1}{1 - \theta}$$

$$\bar{x}(1 - \theta) = (n - \bar{x})\theta$$

$$\bar{x} - \bar{x}\theta = n\theta - \bar{x}\theta$$

$$\bar{x} = n\theta$$

The MLE is given by the proportion of successes

$$\hat{\theta} = \frac{\bar{x}}{n} = \frac{118}{129} \approx 0.91$$

Using the Wald method, the 95% confidence interval is $\hat{\theta} \pm \sqrt{\frac{\hat{\theta}(1 - \hat{\theta})}{n}}$

```
z <- qnorm(0.975)
theta_hat <- 118/129
var_theta_hat <- theta_hat*(1-theta_hat)/129
c(theta_hat - z*sqrt(var_theta_hat), theta_hat + z*sqrt(var_theta_hat))
```

```
## [1] 0.8665338 0.9629236
```

The 95% confidence interval is around (0.87, 0.96)

Question 2

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

$\text{Beta}(1,1)$ is the same as $\text{Unif}(0,1)$

From the lecture slides, the posterior density is $\theta|y \sim \text{Beta}(y+1, n-y+1) = \text{Beta}(119, 12)$.

$$E(\theta|y) = \frac{y+1}{y+1+n-y+1} = \frac{119}{131}$$

So we can estimate θ using $E(\theta|y) = \frac{119}{131}$

```
c(qbeta(0.025, 119, 12), qbeta(0.975, 119, 12))
```

```
## [1] 0.8536434 0.9513891
```

The 95% credible interval is given by (0.85, 0.95)

Question 3

Now assume a $\text{Beta}(10,10)$ 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?

This assumes that 9 successes and 9 failures. This is more information than the prior in Question 2. The prior in Question 2 is equivalent to $\text{Unif}(0,1)$ prior which assumes that everything in (0,1) is equally likely. However with a $\text{Beta}(10,10)$ prior, you assume that some values are more likely than others, which is extra information.

Question 4

Create a graph in ggplot which illustrates

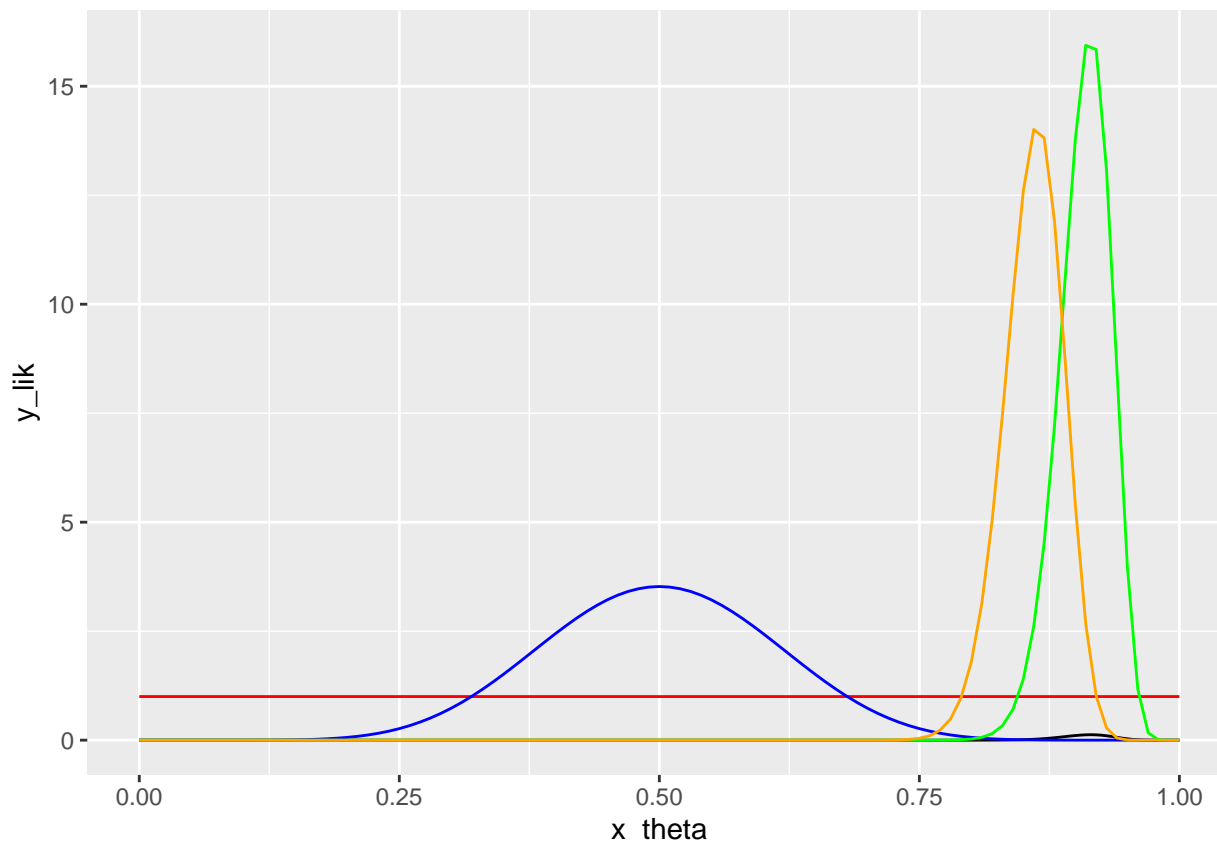
- The likelihood (easiest option is probably to use `geom_histogram` to plot the histogram of appropriate random variables)
- The priors and posteriors in question 2 and 3 (use `stat_function` to plot these distributions)

Comment on what you observe.

The posterior density $\pi(\theta|x)$ for $x \sim \text{Binom}(n, \theta)$ and $\theta \sim \text{Beta}(\alpha, \beta)$ is $\pi(\theta|x) \sim \text{Beta}(\alpha + x, n - x + \beta) = \text{Beta}(10 + 118, 129 - 118 + 10) = \text{Beta}(128, 21)$

```
library(tidyverse)
likelihood <- function(theta, n, y){
  result = choose(n,y)*(theta^(y))*(1-theta)^((n -y))
  return(result)
}

x_theta = seq(from = 0, to=1, by=0.005)
y_lik = likelihood(x_theta, 129, 118)
df = data.frame(x_theta, y_lik)
ggplot(df, aes(x_theta, y_lik)) + geom_line() +
  stat_function(fun="dbeta", args = list(1,1), col="red") +
  stat_function(fun=dbeta, args = list(10, 10), col="blue") +
  stat_function(fun="dbeta", args = list(119, 12), col="green") +
  stat_function(fun="dbeta", args = list(128, 21), col="orange")
```



The posterior from Question 2 has a higher mean than the posterior from Question 3. Both posteriors appear to be Beta distributions.

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, and
- A subjective/informative prior based on your best knowledge

Solution

One possible noninformative prior is $\text{Unif}(0,1)$, which assumes that students are equally likely to have any level of improvement.

A subjective/informative prior is $\text{Beta}(21, 81)$, which assumes that students will improve by 20 more successful shots than the beginning of the month.