Lab 3 STA2201

shaohan chang

29/01/23

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.

Answer:

Assume $Y \mid \theta \sim Bin(n, \theta)$ where Y is the number of women who report to be happy out of the sample of n women. Then, the MLE of θ is $\hat{\theta} = y/n = 118/129 \approx 0.91$. And the standard error of $\hat{\theta}$ is $\sqrt{\hat{\theta}(1-\hat{\theta})/n} \approx \sqrt{0.91 * (1-0.91)/129} \approx 0.025$. Thus, the 95% confidence interval of θ is $[0.91 - 1.96 * 0.025, 0.91 + 1.96 * 0.025]$, that is, $[0.86, 0.96]$.

```
theta0=0.5

Log_likelihood=function(theta){
  like=dbinom(118,129,theta,log = TRUE)
  return(-like)
}
results=nlm(Log_likelihood,p=theta0)
mle_theta=results$estimate
mle_theta-sqrt(mle_theta*(1-mle_theta)/129)*1.96
```

```
[1] 0.8665329
```

```
mle_theta+sqrt(mle_theta*(1-mle_theta)/129)*1.96
```

```
[1] 0.9629244
```

Question 2

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

Answer:

Note that $\theta \sim Beta(1,1)$, and $Y \mid \theta \sim Bin(n,\theta)$. Then, the posterior model will be $\theta \mid y \sim Beta(y+1, n-y+1)$, that is, $Beta(130, 12)$.

So, the posterior expected value is $\mathbb{E}(\pi \mid y) = \frac{y+1}{y+1+n-y+1} = \frac{y+1}{n+2} = \frac{118+1}{129+2} \approx 0.91$.

And, the posterior variance is $130 * 12/(130+12)^2(130+12+1)$.

Therefore, the credible interval of $\hat{\theta}$ is $[0.85, 0.95]$, and the posterior expected value is about 0.91.

```r
p <- 0.025
qbeta(p = p, shape1 = 119, shape2 = 12)
```

```
[1] 0.8536434
```

```r
p <- 0.975
qbeta(p = p, shape1 = 119, shape2 = 12)
```

```
[1] 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?

Answer:

A $Beta(10, 10)$ prior means that there are half of 129 women indicates they are happy.

Here, we are assuming we know more information than the prior used in Question 2. In Question 2, the $Beta(1, 1)$ is exactly the uniform distribution on $(0, 1)$. The uniform prior is non-informative. $Beta(10, 10)$ in Question 3 is a symmetric bell, it provides more information.

Question 4

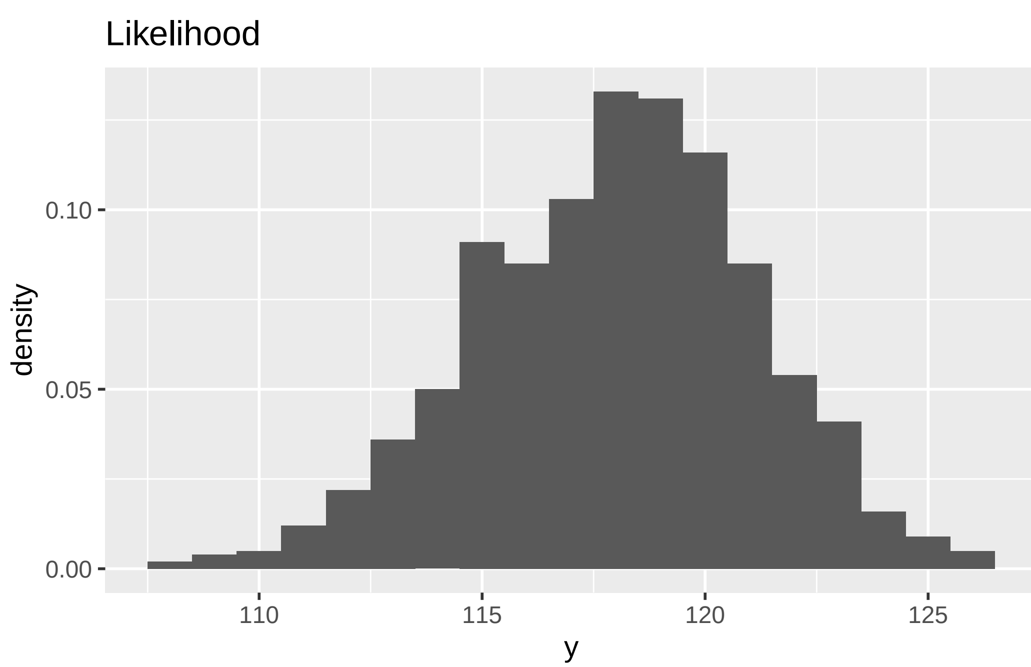
Create a graph in ggplot which illustrates

- 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)

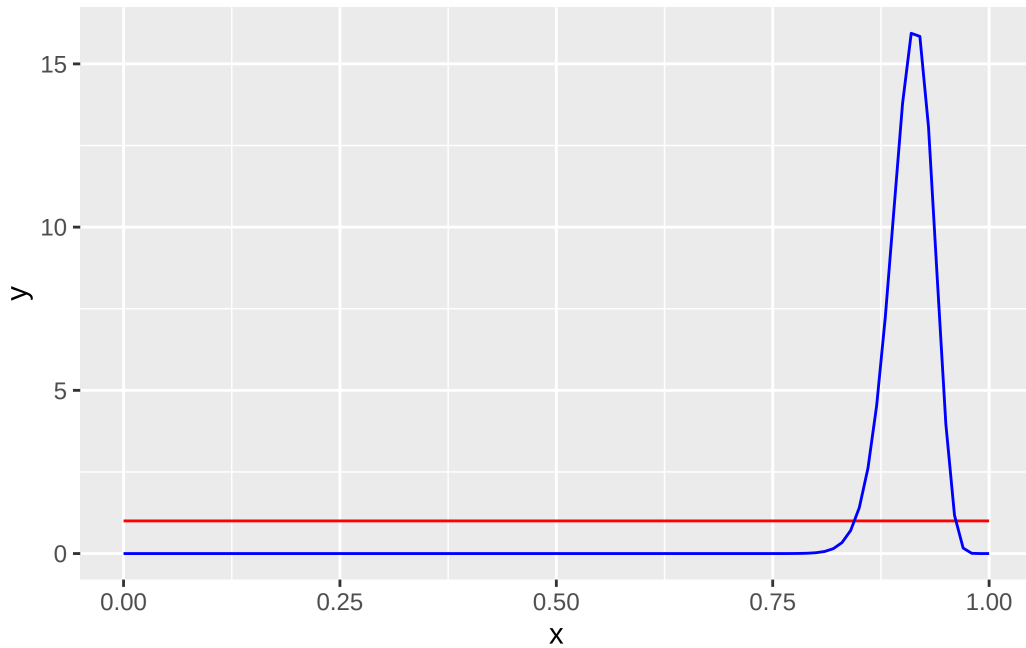
Comment on what you observe.

Answer:

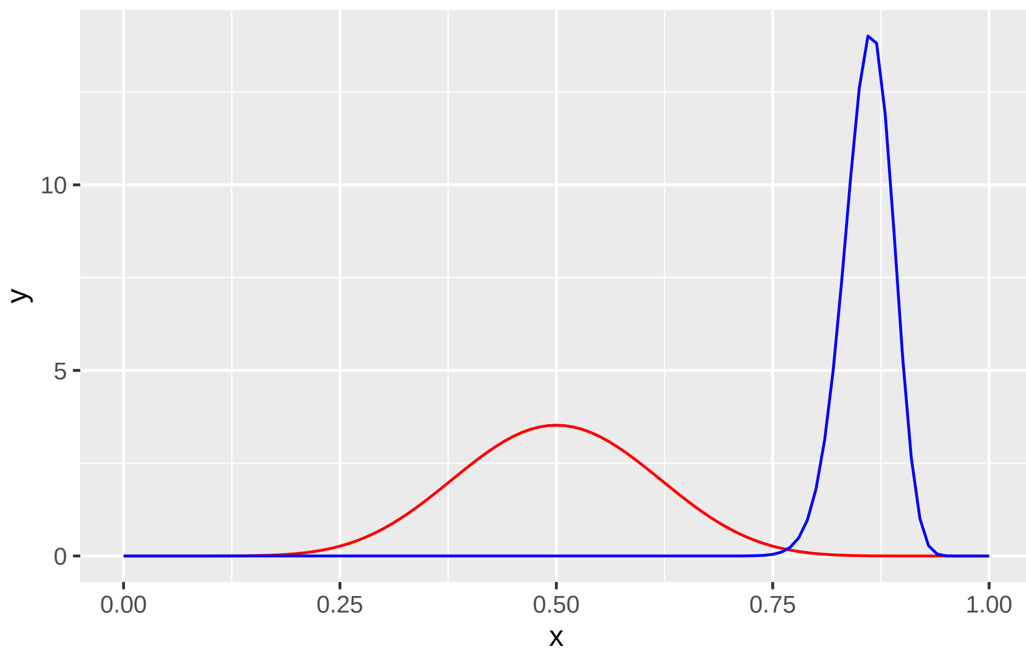
```
#likelihood of Y | theta
library(ggplot2)
library(dplyr)
n=129
theta=118/n
y=rbinom(1000,n,theta)

ggplot()+geom_histogram(aes(x=y,y=..density..),stat ="bin",binwidth = 1)+
  ggtitle("Likelihood")
```



```
#Q2
f <- ggplot(data.frame(x=c(0,1)),aes(x))
f+stat_function(fun=dunif,color='red',show.legend = TRUE)+
  stat_function(fun=dbeta,args = list(shape1=119,shape2=12),
                color='blue',show.legend = TRUE)
```




```
#Q3
f+stat_function(fun=dbeta,args = list(shape1=10,shape2=10),
                color='red',show.legend = TRUE)+
  stat_function(fun=dbeta,args = list(shape1=128,shape2=21),
                color='blue',show.legend = TRUE)
```



Comments:

With the $Beta(1, 1)$ prior, the posterior mean is closed to the MLE of θ, which is decided by the likelihood. However, with the $Beta(10, 10)$ prior, the posterior mean is a balanced result of the prior mean and the MLE.

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

Answer:

A noninformative prior: uniform distribution on $(0, 1)$.

For the noninformative prior uniform distribution on $(0, 1)$, For the average improvement in success probability of basketball free throws, a uniform distribution on $(0, 1)$ is used as the noninformative prior. The presumption behind this prior is that the true value of theta could be any number between 0 and 1.

A subjective/information prior : Beta(10,10)

By the knowledge, For the subjective/informative prior we usually assume the expected mean of the improvement in success probability is 0.5 at first. Thus, a Beta(10,10) is given as an informative prior. Additionally, The researcher's hypothesis claims that a Beta(10,10) prior represents the initial increase in the success probability of a basketball free throw at 0.5.