

Week 3: Intro to Bayes

27/01/23

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.

Assume $Y|\theta \sim \text{Bin}(n, \theta)$.

$$L(\theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y} \log(\theta) = \log \binom{n}{y} + y \log \theta + (n - y) \log(1 - \theta) \frac{\partial l}{\partial \theta} = \frac{y}{\theta} + \frac{n - y}{1 - \theta} (-1)$$

setting the derivative to 0 gives

$$\begin{aligned} \frac{y}{\theta} &= \frac{n - y}{1 - \theta} \\ (n - y)\theta &= y(1 - \theta) \\ n\theta - y\theta &= y - y\theta \\ n\theta &= y \\ \hat{\theta} &= \frac{y}{n} \end{aligned}$$

to get the variance

$$\begin{aligned} \text{Var}(\hat{\theta}) &= \text{Var}\left(\frac{y}{n}\right) \\ &= \frac{1}{n^2} \text{Var}(y) \\ &= \frac{1}{n^2} n\theta(1 - \theta) \\ &= \frac{\theta(1 - \theta)}{n} \end{aligned}$$

$$\hat{\text{Var}}(\hat{\theta}) = \frac{\hat{\theta}(1 - \hat{\theta})}{n}$$

```

n=129
y=118
theta_hat=y/n
upper=theta_hat+1.96*sqrt(theta_hat*(1-theta_hat)/n)
lower=theta_hat-1.96*sqrt(theta_hat*(1-theta_hat)/n)
sprintf("MLE estimate of theta is %f",theta_hat)

```

```
[1] "MLE estimate of theta is 0.914729"
```

```

sprintf("confidence interval is (%f , %f)", lower, upper)

```

```
[1] "confidence interval is (0.866533 , 0.962924)"
```

Question 2

Assume a $\text{Beta}(1,1)$ prior on θ . Calculate the posterior mean for $\hat{\theta}$ and 95% credible interval. The $\text{Beta}(1, 1)$ distribution is the same as the $\text{Uniform}(0, 1)$ distribution. Hence from the lecture we know that the posterior distribution is $p(\theta|y) \sim \text{Beta}(y+1, n-y+1)$

```

pos_mean=(y+1)/(y+1+n-y+1)
sprintf("posterior mean is %f",pos_mean)

```

```
[1] "posterior mean is 0.908397"
```

```

sprintf("credible interval is (%f,%f)", qbeta(0.25,y+1,n-y+1),qbeta(0.975,y+1,n-y+1))

```

```
[1] "credible interval is (0.892696,0.951389)"
```

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?

The interpretation for a $\text{Beta}(10,10)$ prior is: what is the most likely probability given $\alpha-1=10-1=9$ success (happy), and $\beta-1=10-1=9$ of failures (not happy). The mean of $\text{Beta}(10,10)$ and $\text{Beta}(1,1)$ are both 0.5, meaning that using $\text{Beta}(10,10)$ or $\text{Beta}(1,1)$ prior gives the same expectation of the probability of being happy, hence we have the same amount of information as the prior used in Question 2.

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)

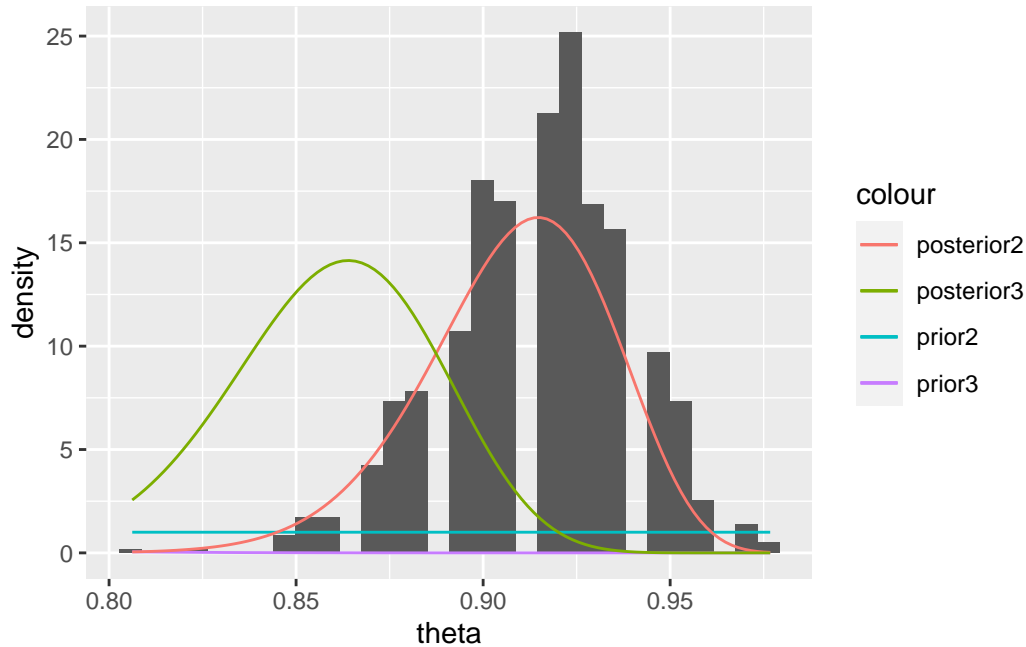
Comment on what you observe.

prior Q2: $\text{Beta}(1,1)$. posterior Q2: $\text{Beta}(y+1, n-y+1)$. prior Q3: $\text{Beta}(10,10)$. posterior Q3: $\text{Beta}(10+y, 10+n-y)$

```
library(ggplot2)

set.seed(1)
df=data.frame(num_happy=rbinom(n=1000,size=129,prob=118/129))
df$theta=df$num_happy/129
ggplot(data=df)+
  geom_histogram(aes(x = theta, y=..density..))+
  stat_function(aes(x = theta, color="prior2"),fun=dbeta,args=list(shape1=1,shape2=1))+
  stat_function(aes(x = theta, color="prior3"),fun=dbeta,args=list(shape1=10,shape2=10))+
  stat_function(aes(x = theta, color="posterior2"),fun=dbeta,
               args=list(shape1=118+1,shape2=129-118+1))+
  stat_function(aes(x = theta, color="posterior3"),fun=dbeta,
               args=list(shape1=118+10,shape2=129-118+10))
```

``stat_bin()`` using ``bins = 30``. Pick better value with ``binwidth``.



We see that the posterior distribution looks like a combination of the likelihood and prior distribution. Posterior in Q2 is more closely aligned with the likelihood, suggesting that prior used in Q2 is weakly-informative.

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.

A noninformative prior could be $\text{Uniform}(0,1)$ since it assigns equal probability to all possible θ hence does not provide more information about θ other than the data itself.

θ should be in the range $(-1,1)$, so an informative prior could be a tanh function $f(\theta) = \tanh(\theta) = \frac{e^\theta - e^{-\theta}}{e^\theta + e^{-\theta}}$ which has the range $(-1,1)$.