hw3 for stat341

Zhihong Zhang

Feb 9th,2017

Q: 2M2-2M3, 3M1-3M3

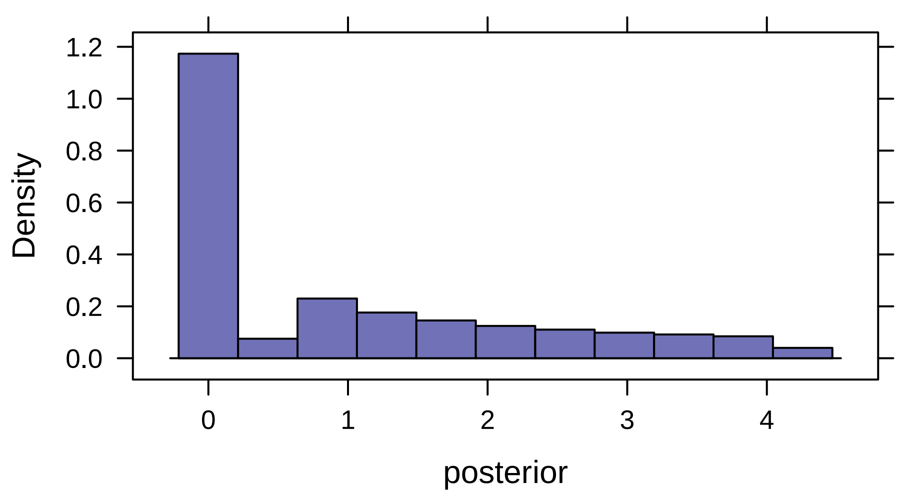
2M2.Q: Now assume a prior for p that is equal to zero when p < 0:5 and is a positive constant when p >=0:5. Again compute and plot the grid approximate posterior distribution for each of the sets of observations in the problem just above. **Solution:**

case(1) WWW

```
BinomGrid <-
  expand.grid(p = seq(0, 1, by = 0.001)) %>%       # create grid of values for p
  mutate(                                              # add additional variables
    prior =dunif(p,0.5,1),                              # uniform prior, value gets recycled
    likelihood = dbinom(3, size = 3, prob = p),            # binomial probabilty
    posterior_raw = prior * likelihood,                    # kernel of posterior
    posterior1 = posterior_raw / sum(posterior_raw),        # easy normalization
    posterior = posterior_raw / sum(posterior_raw) / 0.001 # fancy normalization
  )
head(BinomGrid)
```

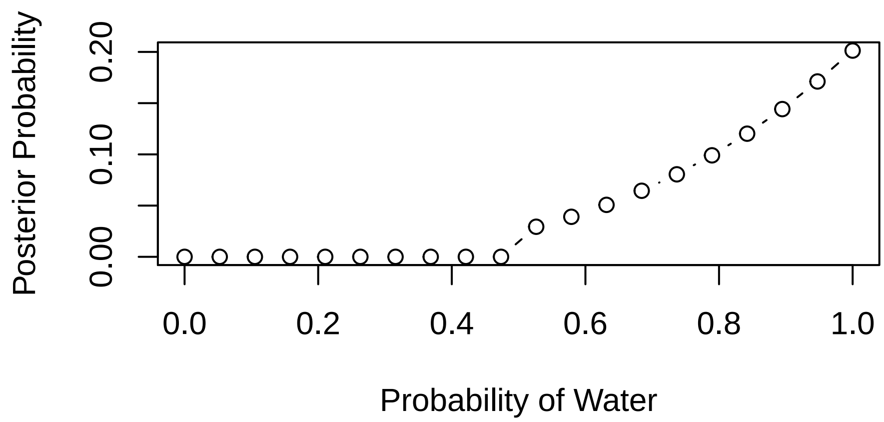
```
##       p prior likelihood posterior_raw posterior1 posterior
## 1 0.000     0   0.00e+00             0          0         0
## 2 0.001     0   1.00e-09             0          0         0
## 3 0.002     0   8.00e-09             0          0         0
## 4 0.003     0   2.70e-08             0          0         0
## 5 0.004     0   6.40e-08             0          0         0
## 6 0.005     0   1.25e-07             0          0         0
```

```
histogram(p~posterior,BinomGrid)
```



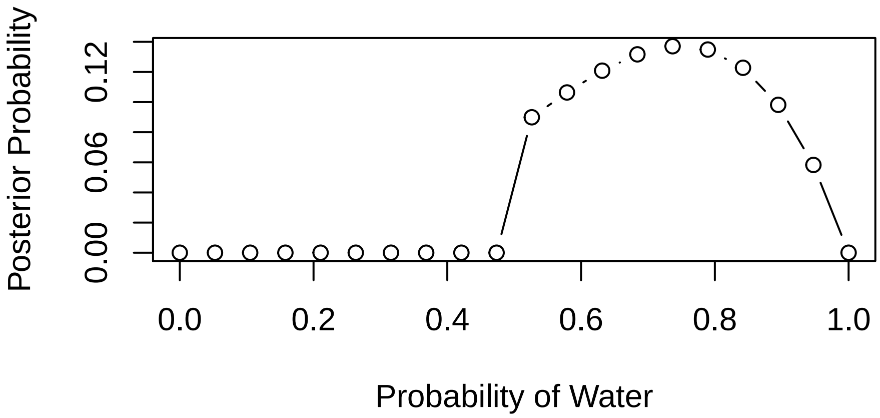
Another way

```r
# define grid
p_grid <- seq( from=0 , to=1 , length.out=20 )

# define prior
prior =dunif(p_grid,0.5,1)
likelihood = dbinom(3, size = 3, prob =p_grid )
posterior_raw = prior * likelihood                    # kernel of posterior
posterior = posterior_raw / sum(posterior_raw)
plot(p_grid, posterior, type = "b",
     xlab = "Probability of Water", ylab = "Posterior Probability")
```



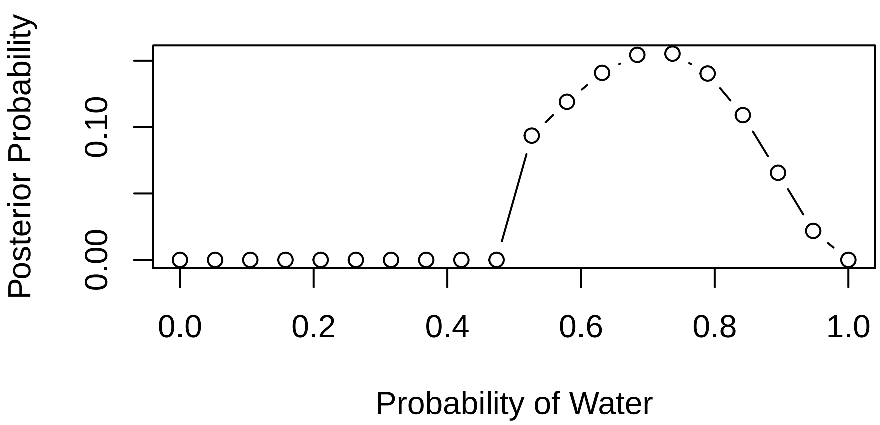
case(2)W, W, W, L

```r
p_grid <- seq( from=0 , to=1 , length.out=20 )

# define prior
prior =dunif(p_grid,0.5,1)
likelihood = dbinom(3, size = 4, prob = p_grid)
posterior_raw = prior * likelihood                    # kernel of posterior
posterior = posterior_raw / sum(posterior_raw)
plot(p_grid, posterior, type = "b",
     xlab = "Probability of Water", ylab = "Posterior Probability")
```



case(3)L, W, W, L, W, W, W

```
p_grid <- seq( from=0 , to=1 , length.out=20 )

# define prior
prior =dunif(p_grid,0.5,1)
likelihood = dbinom(5, size = 7, prob = p_grid)
posterior_raw = prior * likelihood                    # kernel of posterior
posterior = posterior_raw / sum(posterior_raw)
plot(p_grid, posterior, type = "b",
     xlab = "Probability of Water", ylab = "Posterior Probability")
```



2M3.Q: Suppose there are two globes, one for Earth and one for Mars. The Earth globe is 70% covered in water. The Mars globe is 100% land. Further suppose that one of these globes—you don't know which—was tossed in the air and produced a "land" observation. Assume that each globe was equally likely to be tossed. Show that the posterior probability that the globe was the Earth, conditional on seeing "land" (Pr(Earth|land)), is 0.23.

Solution:

posterior can be identified as prior*likelihood. Also it can be written $Pr(Earth|land) = \frac{Pr(Land|Earth)Pr(Earth)}{Pr(Land)} = \frac{Pr(Land|Earth)Pr(Earth)}{Pr(Land|Earth)Pr(Earth)+Pr(Land|Mars)Pr(Mars)}$ Since it is equally likely, prior is determined to be 0.5 Then it is $\frac{0.3*0.5}{(0.3*0.5)(1.0*0.5)}$

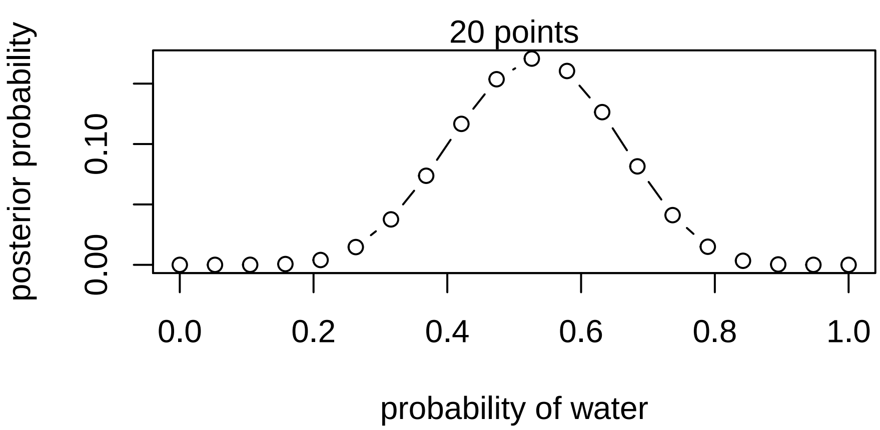
```
p<- 0.3*0.5/(0.3*0.5+1*0.5)
p
```

```
## [1] 0.2307692
```

by R, it is 0.2307

3M1. Q: Suppose the globe tossing data had turned out to be 8 water in 15 tosses. Construct the posterior distribution, using grid approximation. Use the same flat prior as before. **Solution:**

```
# define grid
p_grid <- seq( from=0 , to=1 , length.out=20 )

# define prior
prior <- rep( 1 , 20 )

likelihood = dbinom(8, size = 15, prob =p_grid )
posterior = prior * likelihood
posterior_raw = prior * likelihood                    # kernel of posterior
posterior = posterior_raw / sum(posterior_raw)        # easy normalization

plot( p_grid , posterior  ,type="b" , xlab="probability of water",ylab="posterior probability" )
mtext("20 points")
```



3M2.Q: Draw 10,000 samples from the grid approximation from above. Then use the samples to calculate the 90% HPDI for p. **Solve:** 10000

```
# define grid
p_grid <- seq( from=0 , to=1 , length.out=20 )

# define prior
prior <- rep( 1 , 20 )
```

```r
likelihood = dbinom(8, size = 15, prob =p_grid )
posterior = prior * likelihood
posterior_raw = prior * likelihood                    # kernel of posterior
posterior = posterior_raw / sum(posterior_raw)         # easy normalization
posterior2 = sample(p_grid, posterior, size = 1e4, replace = TRUE)
HPDI(posterior2, prob = 0.9)
```

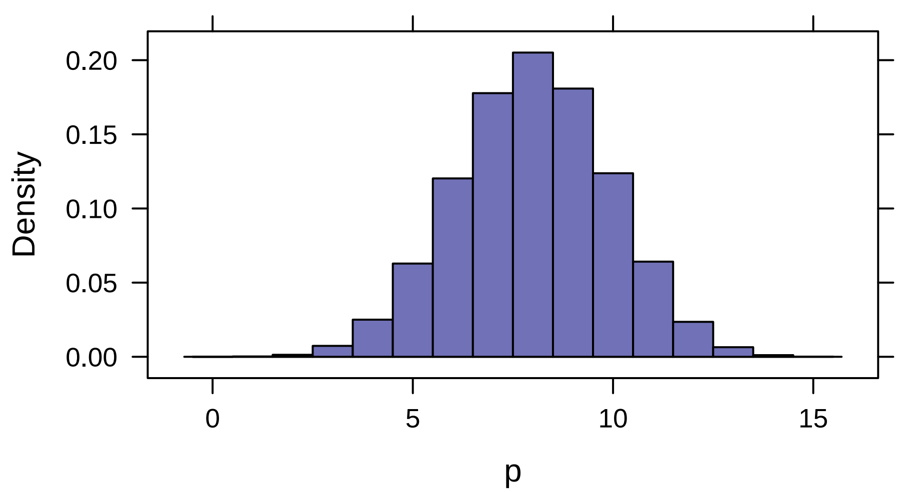
```
##      |0.9      0.9|
## 0.3157895 0.6842105
```

3M3.Q: Construct a posterior predictive check for this model and data. This means simulate the distribution of samples, averaging over the posterior uncertainty in p. What is the probability of observing 8 water in 15 tosses?

Solution:

Based on the R, probability is 0.20547.

```r
 p<-rbinom(1e5, size = 15, prob = 8/15)
histogram(~p,width=1)
```




```r
tally(~(p==8),format = "prop")
```

```
## (p == 8)
##    TRUE   FALSE
## 0.20512 0.79488
```