Week 1 Problems

Juan Pablo Fonseca-Zamora

1. Suppose the globe tossing data (Lecture 2, Chapter 2) had turned out to be 3 water and 11 land. Construct the posterior distribution

Solution

Defining the a vector with the specified observations (3 Water and 11 land).

```
my_sample <- c(rep("W",3),rep("L",11))
my_sample
```

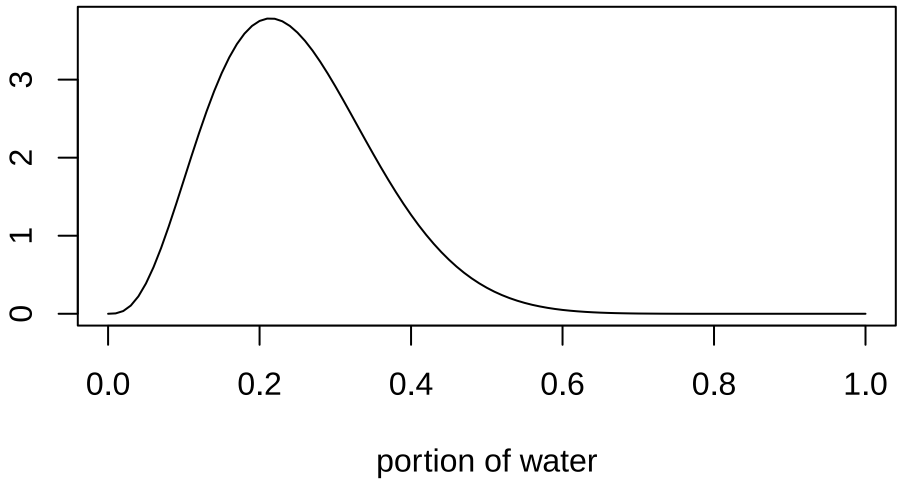
```
 [1] "W" "W" "W" "L" "L" "L" "L" "L" "L" "L" "L" "L" "L" "L"
```

Defining the function for computing the posterior based on a given sample:

```
compute_posterior <- function(p, the_sample) {
  # Counting the number of Water and Land occurrences
  W <- sum(the_sample == "W")
  L <- sum(the_sample == "L")
  # Calculating the density for p given the number of W and L and the prior
  dbeta(p, W + 1, L + 1)
}
```

Visualising the posterior

```
curve(
  compute_posterior(p = x, the_sample = my_sample),
  from = 0,
  to = 1,
  xlab = "portion of water",
  ylab = ""
)
```



2. Using the posterior distribution from **1**, compute the posterior predictive distribution for the next 5 tosses of the same globe. I recommend you use the sampling method.

Solution

Sampling values of *p* for the given number of W and L using the `rbeta` function.

```
p_samples <- rbeta(1e4, # Obtaining 10000 values of p
                   3 + 1, # The number of observed "W" in the sample
                   11 + 1) # The number of observed "L" in the sample
```

I modified the the function `sim_globe` from the book/lectures to return the number of W in each simulation.

```
sim_globe3 <- function(p, N) {
  # Simulating 5 tosses for a given p
  sim_sample  <- sample(
    c("W", "L"),
    size = N,
    replace = T,
    prob = c(p, 1 - p)
  )
  # Counting the number of W in the simulation
  return(sum(sim_sample == "W"))
}
```

A simulation is run for each value of p in the `p_samples` vector to obtain a posterior predictive distribution.

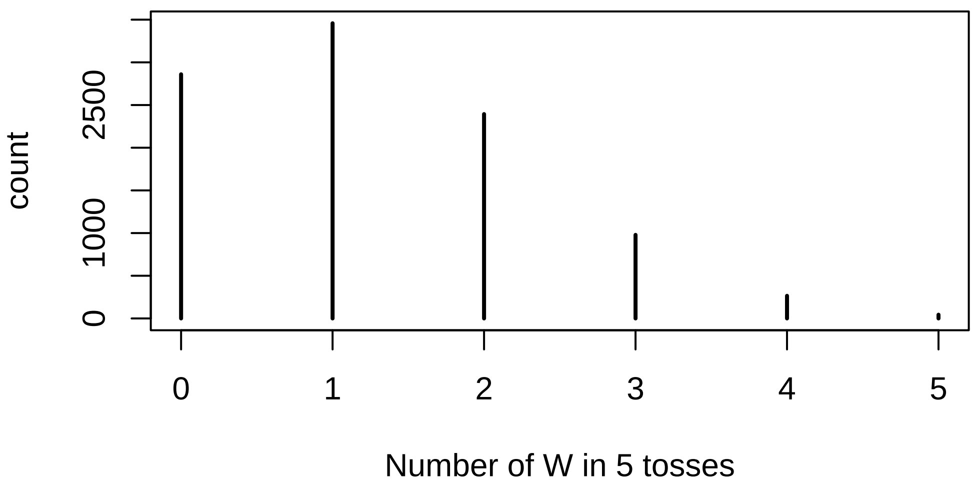
```
W_sim_5t <- vapply(p_samples, sim_globe3, 0, N = 5)
```

A summary and visualisation of the distribution:

```
table(W_sim_5t)
```

```
W_sim_5t
   0    1    2    3    4    5 
2860 3458 2394  979  265   44 
```

```
plot(table(W_sim_5t),
     xlab = "Number of W in 5 tosses",
     ylab = "count")
```



Optional 3. Suppose you observe $W = 7$ water points, but you forgot to write down how many times the globe was tossed, so you don't know the number of land points L.

Solution I defined a function to compute the probability density using a fixed number of successes $x = 7$ and fixed $p = 0.7$. Then the posterior is normalised.

```
compute_posterior_tosses = function(N,p = 0.7){
  post  <- dbinom(7, size = N, prob = p)
  post <- post / sum(post) # Normalising
}
```

A visual of the posterior for different number of tosses.

```
p_grid <- 0:100
posterior <-  compute_posterior_tosses(N = p_grid)
plot(p_grid,
```

```
    posterior,
    type = "l",
    xlim = c(0,30), # 0 to 30 to improve visualisation
    xlab = "Number of Tosses",
    ylab = "Density")
```