Statistical Rethinking Homework

Courtney Robichaud
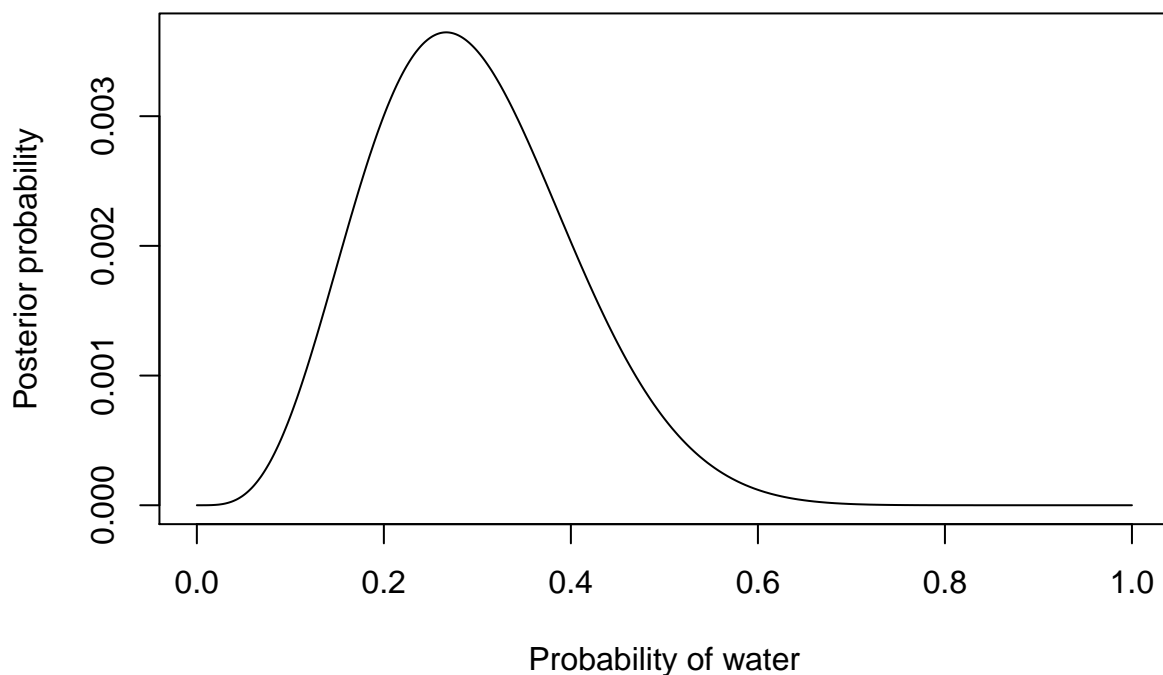
11/01/2022

Week 1

Question 1. Suppose the globe tossing data (Chapter 2) had turned out to be 4 water and 11 land. Construct the posterior distribution, using grid approximation. Use the same flat prior as in the book.

```r
p_grid <- seq(from = 0, to = 1,
              length.out = 1000)
prob_p <- rep(1,1000)
prob_data <- dbinom(4, 15, prob = p_grid)

posterior <- prob_data * prob_p
posterior <- posterior/sum(posterior)
```

Question 2. Now suppose the data are 4 water and 2 land. Compute the posterior again, but this time use a prior that is zero below p = 0.5 and a constant above p =0.5. This corresponds to prior information that a majority of the Earth's surface is water.

```
prior2 <- ifelse( p_grid < 0.5, 0, 1)
p_likelihood <- dbinom(4, 6, prob = p_grid)

posterior2 <- p_likelihood * prior2
posterior2 <- posterior2/sum(posterior2)
```
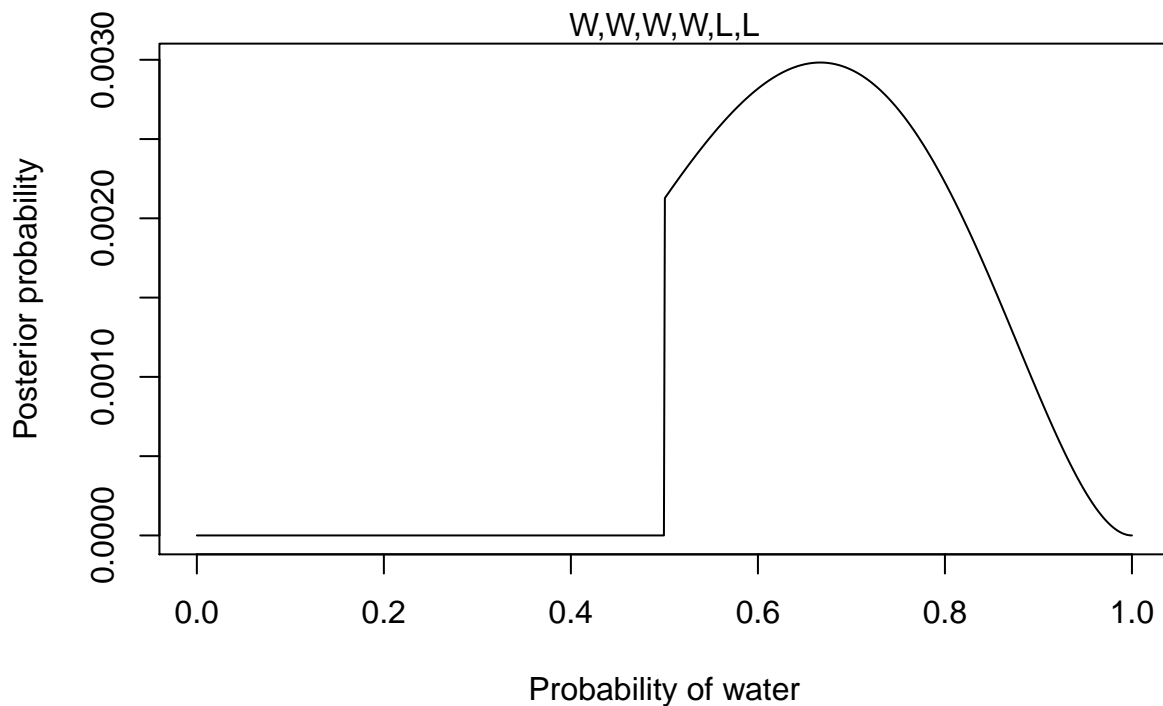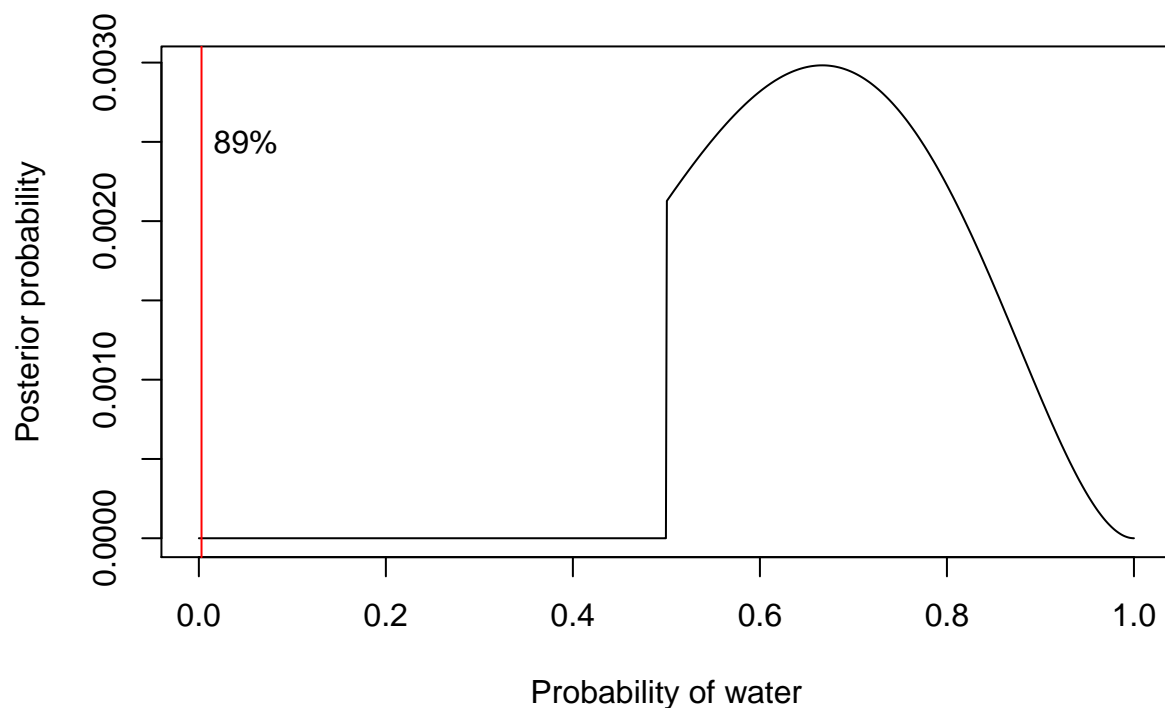

Question 3. For the posterior distribution from 2, compute 89% percentile and HPDI intervals. Compare the widths of these intervals. Which is wider? Why? If you had only the information in the interval, what might you misunderstand about the shape of the posterior distribution?

```
quantile(posterior2, 0.89)
```

```
##             89%
## 0.002860699
```

2

```r
HPDI(posterior2, 0.5)
```

```
## |0.5 0.5|
##    0    0
```

```r
HPDI(posterior2, 0.7)
```

```
##          |0.7         0.7|
## 0.000000000 0.002169826
```

```r
HPDI(posterior2, 0.89)
```

```
##          |0.89        0.89|
## 0.000000000 0.002864337
```

```r
HPDI(posterior2, 0.5) < quantile(posterior2, 0.89)
```

```
## |0.5 0.5|
## TRUE TRUE
```

It looks like the 89% interval contains nearly all of the data (is wider), compared to the HPDI values (though I am not confident in my approach). The issue with these approaches is that they do not give information about the shape of the posterior distribution, which assumes there is no land on Earth.