

Statistical Rethinking — Week 1

Caio Borges Aguida Geraldes

January 11, 2022

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.

Answer:

```
1 grid.size <- 100
2 p.grid <- seq(from=0, to=1, length.out=grid.size)
3 prior <- rep(1, grid.size)
4 likelihood <- dbinom(4, size=15, prob=p.grid)
5 unstd.posterior <- likelihood * prior
6 posterior <- unstd.posterior / sum(unstd.posterior)
```

Listing 1: Calculating the posterior distribution with a constant prior after 15 globe tosses out of which water was observed 4 times.

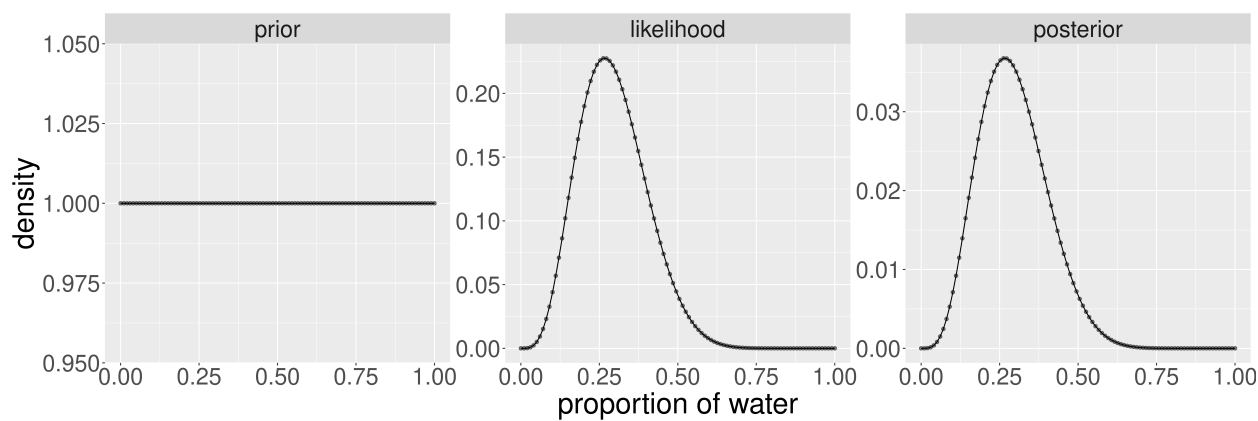


Figure 1: Grid approximation of the posterior distribution with a constant prior after 15 globe tosses out of which water was observed 4 times.

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.

Answer:

```
1 grid_size <- 100
2 p_grid <- seq(from=0, to=1, length.out=grid_size)
3 prior <- ifelse(p_grid < 0.5, 0, 1)
4 likelihood <- dbinom(4, size=6, prob=p_grid)
5 unstd.posterior <- likelihood * prior
6 posterior <- unstd.posterior / sum(unstd.posterior)
```

Listing 2: Calculating the posterior distribution with a constant prior that is 0 below 0.5 and 1 above after 6 globe tosses out of which water was observed 4 times.

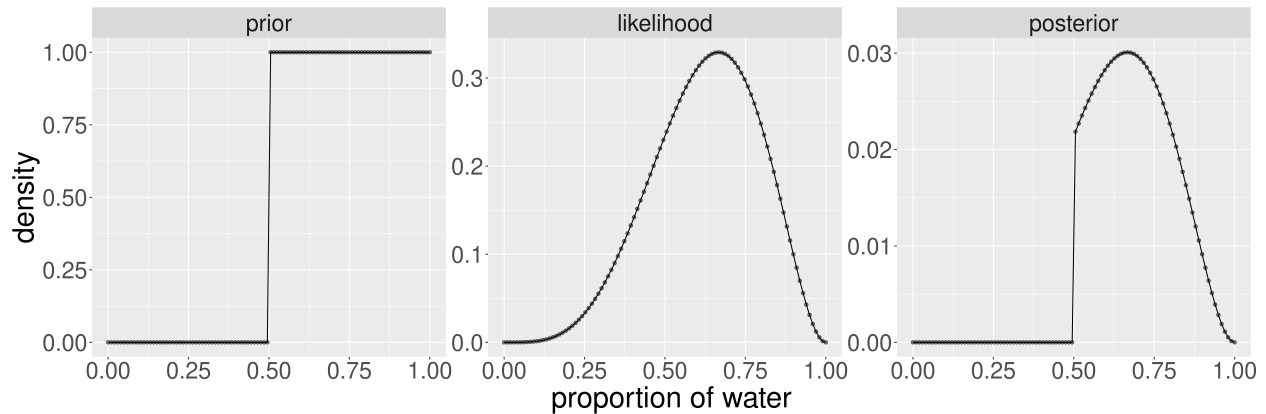


Figure 2: Grid approximation of the posterior distribution with a constant prior that is 0 below 0.5 and 1 above after 6 globe tosses out of which water was observed 4 times.

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?

Answer:

First, we draw 10000 samples from the posterior distribution:

```
1 samples <- sample( p_grid , size=1e4 , replace=TRUE , prob=posterior )
```

Listing 3: Sampling from the posterior distribution and plotting Figure 3.

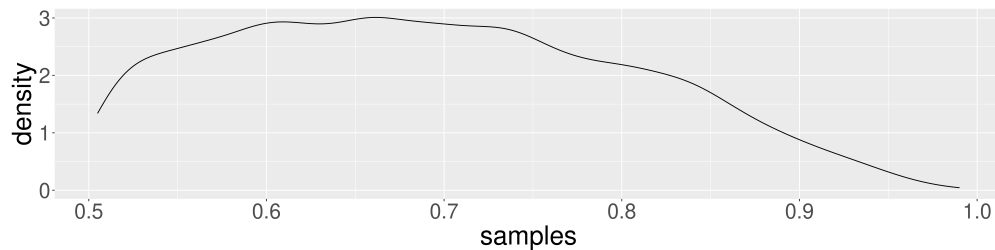


Figure 3: Density of samples drawn from the posterior distribution of Figure 2.

From the samples drawn from the posterior distribution, we calculate the 89% percentile and HPDI interval with the functions `rethinking::PI` and `rethinking::hpdi`.

```
1 percentile <- rethinking::PI(samples, .89)
2 #      5%      94%
3 # 0.5252525 0.8787879
4 print(str_c("Percentile width: ", as.numeric(percentile[2] - percentile[1])))
5 # [1] "Percentile width: 0.353535353535354"
6 hpdi <- rethinking::HPDI(samples, .89)
7 #      |0.89      0.89|
8 # 0.5050505 0.8383838
9 print(str_c("HPDI width: ", as.numeric(hpdi[2] - hpdi[1])))
10 # [1] "HPDI width: 0.333333333333333"
```

Listing 4: Calculating the 89% percentile and HPDI intervals of the samples in Figure 3.

This way, the 89% percentile interval is described in Table 1, and the percentile width is 0.35, as shown in Figure 4; and the 89% HPDI interval is described in Table 2, and the hpdi width is 0.35, as shown in Figure 5. The percentile interval has a higher width than that of HPDI. This is due to the fact of the percentile interval attributing even weights for every p while capturing the central 89% probability, whereas the HPDI finds the densest interval in the distribution by searching the narrowest possible interval with the highest density.

5%	94%
0.5252525	0.8787879

Table 1: 89% percentile interval

$ 0.89$	$0.89 $
0.5050505	0.8383838

Table 2: 89% HPDI interval

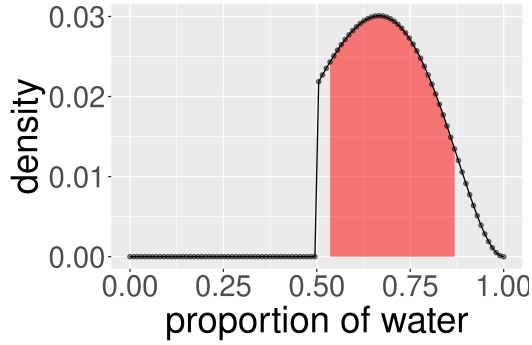


Figure 4: 89% percentile interval of the posterior distribution resulting from Figure 2.

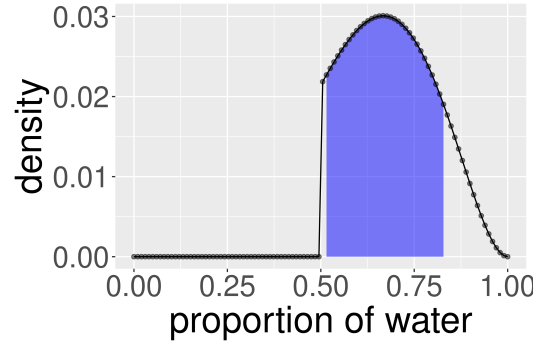


Figure 5: 89% HPDI of the posterior distribution resulting from Figure 2.

The intervals do not account for the region of the graph influenced by the prior distribution assumed, i.e. the region with 0 plausibility where the proportion of water < 0.5 . Without the visualization of the samples Figure 3 and of the posterior Figure 4 and Figure 5, it is possible to assume a less skewed posterior distribution with a much more symmetric profile.