

Statistical Rethinking Notes - Chapter 2

Zachary Levonian

2022

```
library(rethinking)
```

Chapter 2

Notes on chapter 2.

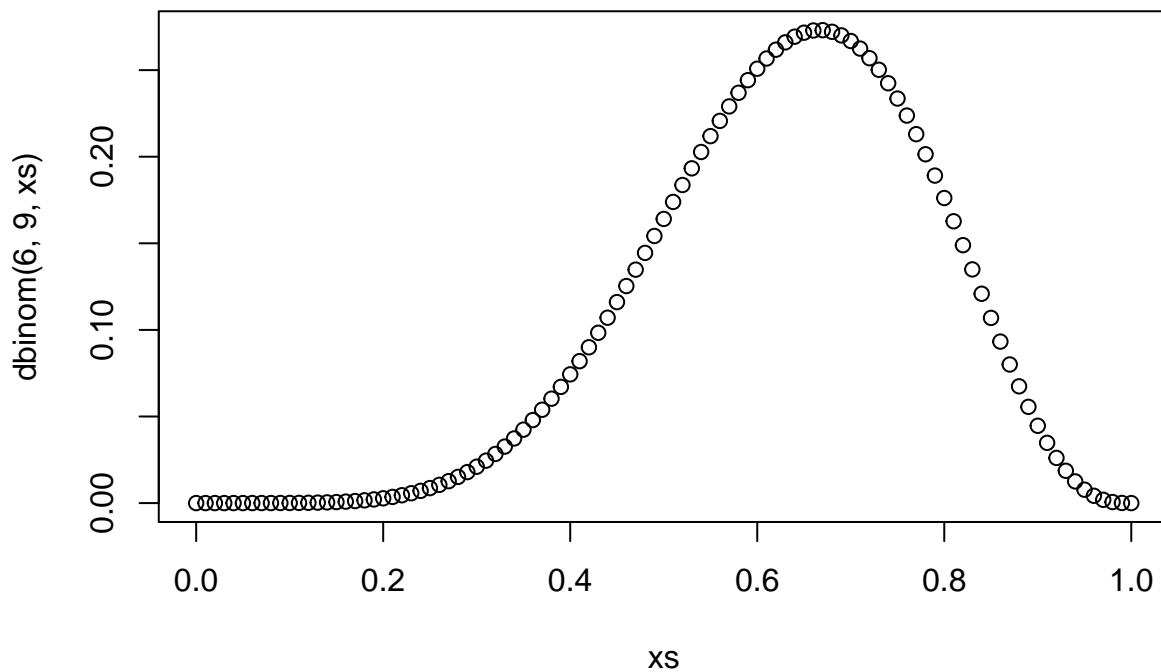
Bayesian data analysis: For each possible explanation of the data, Count all the ways the data can happen. Explanations with more ways to produce the data are more plausible.

$\Pr(W, L|p) = \frac{(W+L)!}{W!L!} p^W (1-p)^L$ where W is the number of water hits and L is the number of land hits.

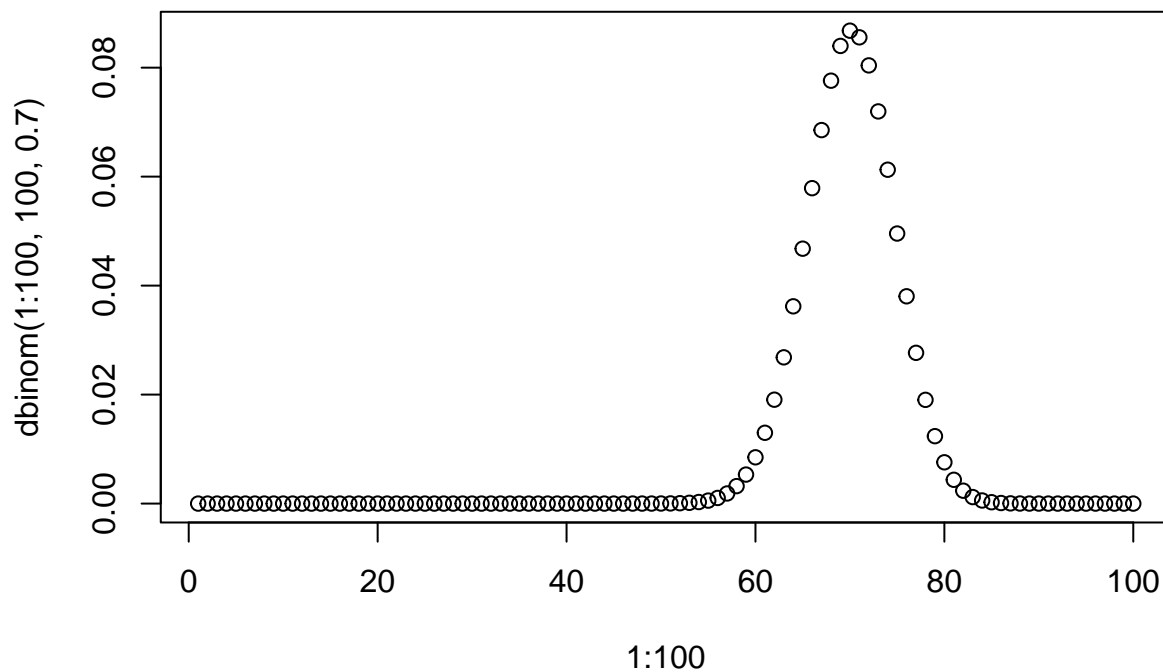
```
dbinom(6, 9, 0.7)
```

```
## [1] 0.2668279
```

```
xs <- seq(0, 1, 0.01)  
plot(xs, dbinom(6, 9, xs))
```

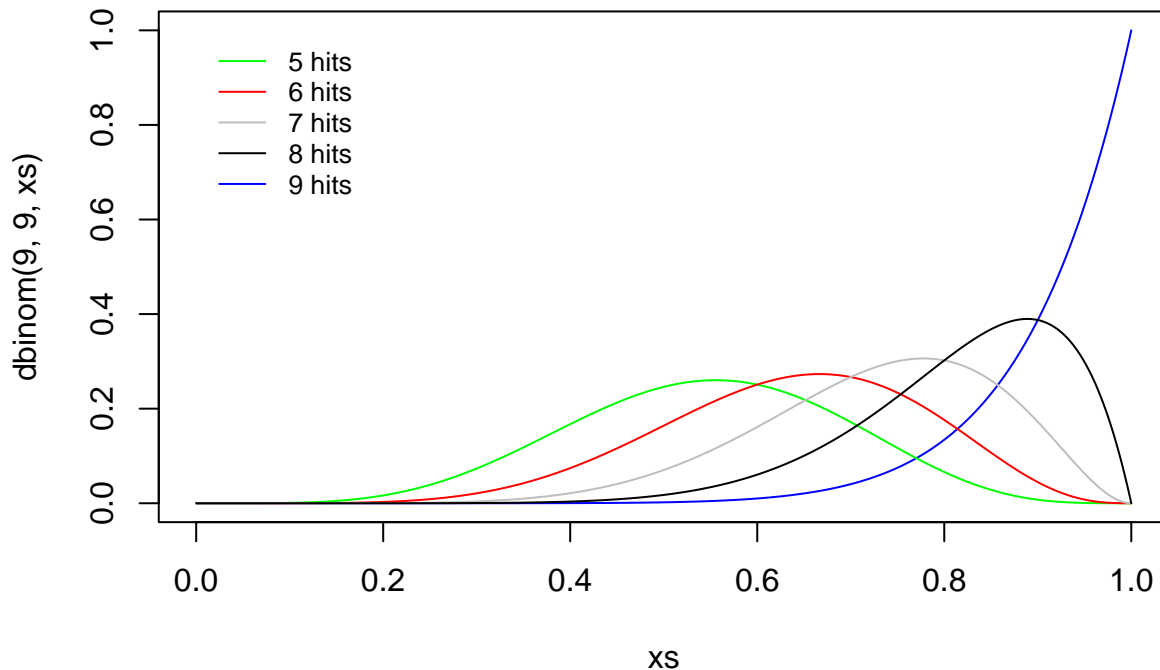


```
plot(1:100, dbinom(1:100, 100, 0.7))
```



```
xs <- seq(0, 1, 0.001)
plot(xs, dbinom(9, 9, xs), type="l", col="blue",
      main="Binomial density plot for different number of hits (of 9 total)")
lines(xs, dbinom(5, 9, xs), col="green")
lines(xs, dbinom(6, 9, xs), col="red")
lines(xs, dbinom(7, 9, xs), col="gray")
lines(xs, dbinom(8, 9, xs), col="black")
legend(0, 1,
       legend=c("5 hits", "6 hits", "7 hits", "8 hits", "9 hits"),
       col=c("green", "red", "gray", "black", "blue"),
       lty=1, cex=0.8,
       box.lty=0)
```

Binomial density plot for different number of hits (of 9 total)



Question 1: why is it okay to set the prior to 1 (rather than $1 / \text{sum}(\text{prior})$)? (in the code example given in the lecture) (answer: because we will normalize after anyway, so it doesn't matter.)

Question 2: why is the evidence of a single W or L a line (and not some other shape)? (answer: garden of forking data; we assume a model where the number of ways that the true proportion is some value p is determined by the number of "paths" that end up at that proportion given the observed data.)

Question 3: can we choose a different likelihood function? More specifically: say there is some down-stream variable causally associated with the true probability of observing p . e.g. planetoids are either "land-likely" or "water-likely", where "water-likely" planets have a true distribution that is linearly decreasing from $p(\text{water}) = 1$ to $p(\text{water}) = 0$, while "land-likely" planets have the opposite. These planets occur at the same rate, so a flat prior is appropriate (i.e., absent data on whether a planetoid is land- or water-likely, there is a uniform probability of any proportion of water on that planet). In this case, it seems like maybe we want a different likelihood function! (Or should we? I think this is a false example, since land-likeliness needs to assign some probability mass to $p(\text{water})$, otherwise we shouldn't hold a uniform distribution.)

- "For each possible value of the unobserved variables, we need to define the relative number of ways — the probability — that the values of each observed variable could arise."
- "So that we don't have to literally count, we can use a mathematical function that tells us the right plausibility. In conventional statistics, a distribution function assigned to an observed variable is usually called a *likelihood*."
- "The probability of the data, often called the likelihood, provides the plausibility of an observation (data), given a fixed value for the parameters."
- On data vs parameters: "It is typical to conceive of data and parameters as completely different kinds of entities. Data are measured and known; parameters are unknown and must be estimated from data. Usefully, in the Bayesian framework the distinction between a datum and a parameter is not so fundamental. Sometimes we observe a variable, but sometimes we do not. In that case, the same distribution function applies, even though we didn't observe the variable. As a result, the same assumption can look like a "likelihood" or a "prior," depending upon context, without any change to the model."

Question 4: fun to think about the ball-drawing example. What if the data and the parameter are "flipped"?

Consider our data is that we are told by an employee that the proportion is 50/50. (In 2.1.2, “Combining other information”, they give an example where information is given by the factory and incorporated with prior draw data.)

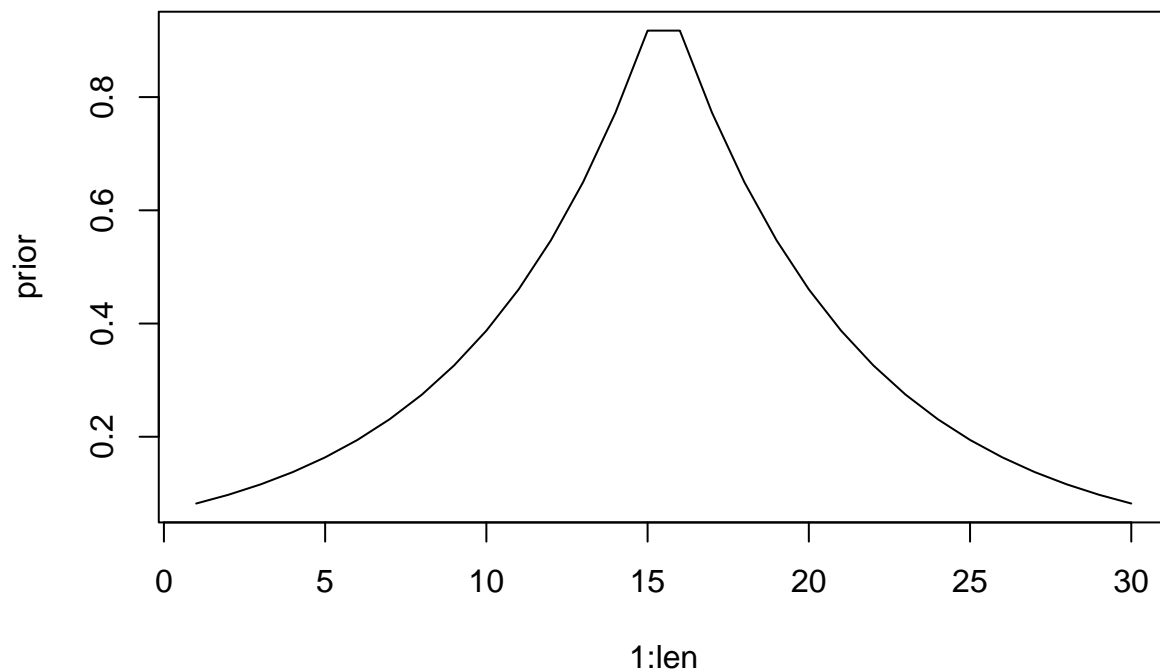
Let’s ask: what’s the plausibility of drawing a Blue vs a White marble given an observation about the true proportion? This problem is harder than the ones we’ve been solving in Chapter 2, since our OBSERVATION is continuous (a percentage/proportion). So, will need to come back to this question later, I think...

Book code

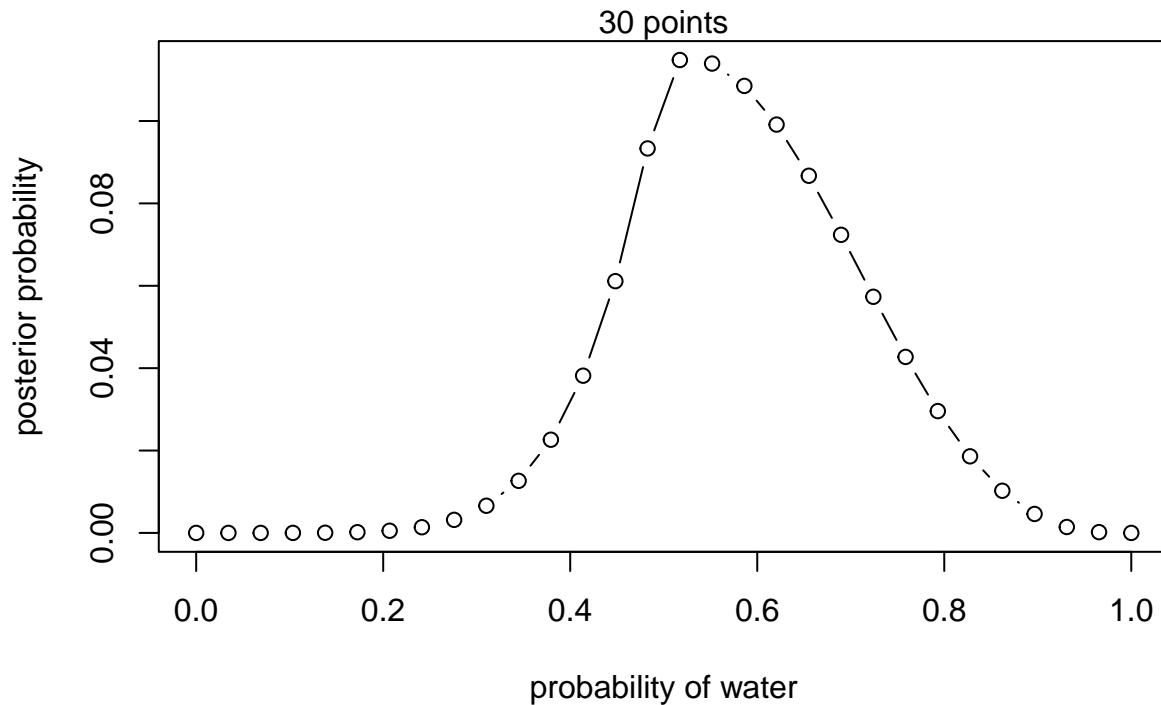
```
len <- 30
# define grid
p_grid <- seq(from=0, to=1, length.out=len )
# define prior
#prior <- rep(1 , len) # rep = repeat
prior <- exp( -5*abs( p_grid - 0.5 ) )

# compute likelihood at each value in grid
likelihood <- dbinom(6 , size=9, prob=p_grid)
# compute product of likelihood and prior
unstd.posterior <- likelihood * prior
# standardize the posterior, so it sums to 1
posterior <- unstd.posterior / sum(unstd.posterior)

# plot the prior
plot(1:len, prior, type='l')
```



```
plot( p_grid, posterior, type="b",
      xlab="probability of water" , ylab="posterior probability" )
mtext(sprintf("%d points", len))
```



```
globe.qa <- quap(
  alist(
    W ~ dbinom( W+L ,p) , # binomial likelihood
    p ~ dunif(0,1) # uniform prior
  ) ,
  data=list(W=6,L=3)
)
# display summary of quadratic approximation
precis( globe.qa )
```

```
##      mean      sd    5.5%    94.5%
## p 0.666667 0.1571337 0.415537 0.917797
```

Book problems

2M1.

```
plot_posterior <- function(obs, len=1000) {
  W <- length(which(obs == "W"))
  L <- length(obs) - W

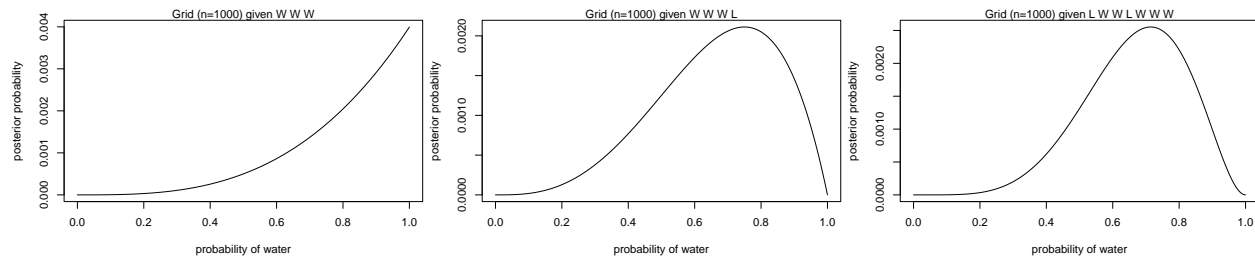
  # define grid
  p_grid <- seq(from=0, to=1, length.out=len )
  # define prior
  prior <- rep(1 , len) # rep = repeat

  # compute likelihood at each value in grid
  likelihood <- dbinom(W, size=W+L, prob=p_grid)
  # compute product of likelihood and prior
  unstd.posterior <- likelihood * prior
  # standardize the posterior, so it sums to 1
  posterior <- unstd.posterior / sum(unstd.posterior)
```

```

# plot the prior
plot(p_grid, posterior, type="l",
     xlab="probability of water",
     ylab="posterior probability"
)
mtext(sprintf("Grid (n=%d) given %s", len, paste(obs, collapse=" ")))
}
plot_posterior(c("W", "W", "W"))
plot_posterior(c("W", "W", "W", "L"))
plot_posterior(c("L", "W", "W", "L", "W", "W", "W"))

```



2M3. 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” ($\Pr(\text{Earth}|\text{land})$), is 0.23.

- Knowns: $\Pr(\text{Earth}) = 0.5$. $\Pr(L|\text{Earth}) = 0.3$.
- $\Pr(L|\text{Earth}) = \Pr(\text{Earth}|L) * \Pr(L) / \Pr(\text{Earth})$ by Bayes Theorem. Rearrange to find $\Pr(\text{Earth}|L) = \Pr(L|\text{Earth}) * \Pr(\text{Earth}) / \Pr(L)$. Replace knowns: $\Pr(\text{Earth}|L) = 0.3 * 0.5 / \Pr(L)$.
- What is $\Pr(L)$? $\Pr(L) = \Pr(L|\text{Earth}) * \Pr(\text{Earth}) + \Pr(L|\text{Mars}) * \Pr(\text{Mars}) = 0.3 * 0.5 + 1 * 0.5 = 0.65$.
- Thus, $\Pr(\text{Earth}|L) = 0.3 * 0.5 / 0.65 = 0.23$.

```
0.3 * 0.5 + 1 * 0.5
```

```
## [1] 0.65
```

```
0.3 * 0.5 / 0.65
```

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

Homework 1

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.

```

plot_posterior <- function(W, L, len=1000) {
  # define grid
  p_grid <- seq(from=0, to=1, length.out=len)
  # define prior
  prior <- rep(1, len) # rep = repeat

  # compute likelihood at each value in grid
  likelihood <- dbinom(W, size=W+L, prob=p_grid)
  # compute product of likelihood and prior
  unstd.posterior <- likelihood * prior
  # standardize the posterior, so it sums to 1
}

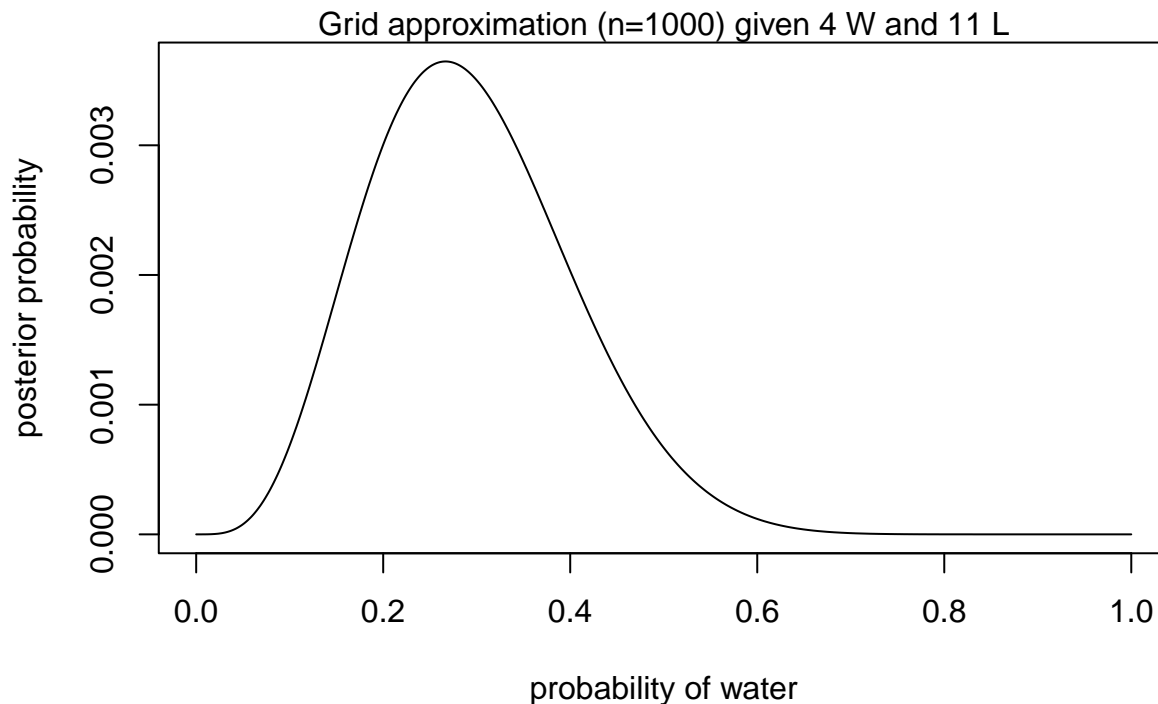
```

```

posterior <- unstd.posterior / sum(unstd.posterior)

# plot the prior
plot(p_grid, posterior, type="l",
     xlab="probability of water",
     ylab="posterior probability"
)
mtext(sprintf("Grid approximation (n=%d) given %d W and %d L", len, W, L))
}
plot_posterior(4, 11)

```



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.

(See 3)

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?

```

plot_posterior <- function(W, L, len=1000) {
  # define grid
  p_grid <- seq(from=0, to=1, length.out=len )
  # define prior
  prior <- ifelse(p_grid < 0.5, 0, 1)

  # compute likelihood at each value in grid
  likelihood <- dbinom(W, size=W+L, prob=p_grid)
  # compute product of likelihood and prior
  unstd.posterior <- likelihood * prior
  # standardize the posterior, so it sums to 1
  posterior <- unstd.posterior / sum(unstd.posterior)
}

```

```

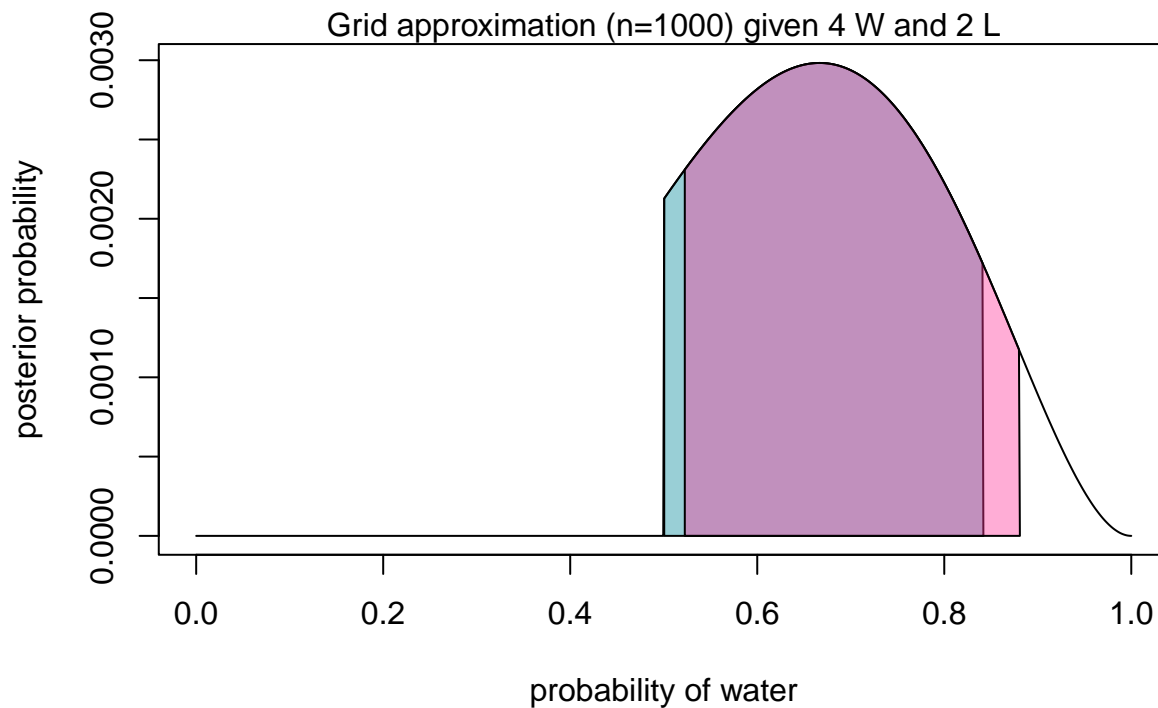
# plot the prior
plot(p_grid, posterior, type="l",
     xlab="probability of water",
     ylab="posterior probability"
)
mtext(sprintf("Grid approximation (n=%d) given %d W and %d L", len, W, L))

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

x <- p_grid
# in green, the highest posterior density interval
bounds <- HPDI(samples, prob=0.89)
print(bounds)
lb <- bounds[1]
ub <- bounds[2]
polygon(c(x[x]>=lb & x<ub], ub, lb), c(posterior[x>=lb & x<ub], 0, 0), col=alpha("#008999", 0.4))

# in pink, the percentile interval
bounds <- PI(samples, prob=0.89)
print(bounds)
lb <- bounds[1]
ub <- bounds[2]
polygon(c(x[x]>=lb & x<ub], ub, lb), c(posterior[x>=lb & x<ub], 0, 0), col=alpha("#FF3399", 0.4))
}
plot_posterior(4, 2)

```



```

##      |0.89      0.89|
## 0.5005005 0.8418418
##      5%      94%
## 0.5225225 0.8808809

```


The HPDI is narrower than the percentile interval. That's because the censored nature of the function pushes more of the probability mass towards the peak, making the higher-density region narrower. The percentile intervals mask the discontinuity around 0.5.

4. (Optional) Suppose there is bias in sampling so that Land is more likely than Water to be recorded. Specifically, assume that 1-in-5 (20%) of Water samples are accidentally recorded instead as "Land". First, write a generative simulation of this sampling process. Assuming the true proportion of Water is 0.70, what proportion does your simulation tend to produce instead? Second, using a simulated sample of 20 tosses, compute the unbiased posterior distribution of the true proportion of water.