Statistical Rethinking Notes - Chapter 2

Zachary Levonian

2022

```
library(rethinking)
```

Chapter 2

Notes on chapter 2.

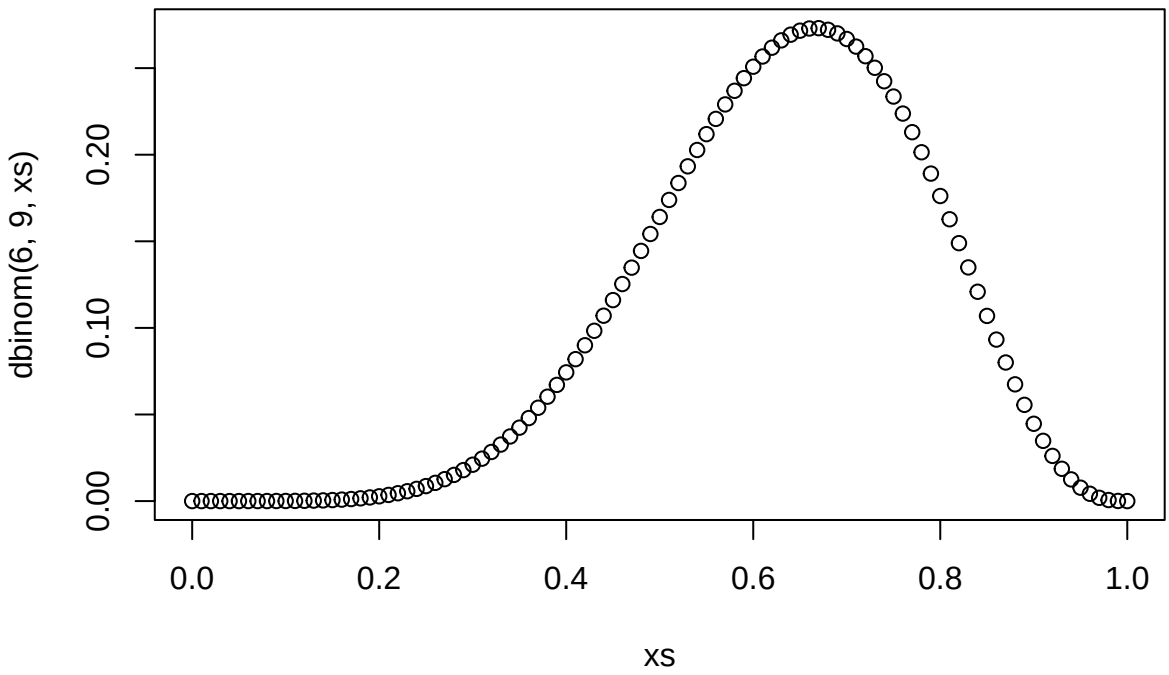
Bayesian data analysis: For each possible explanation of the data, Count all the ways the data can happen. Explanations with more ways to produce the data are more plausible.

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

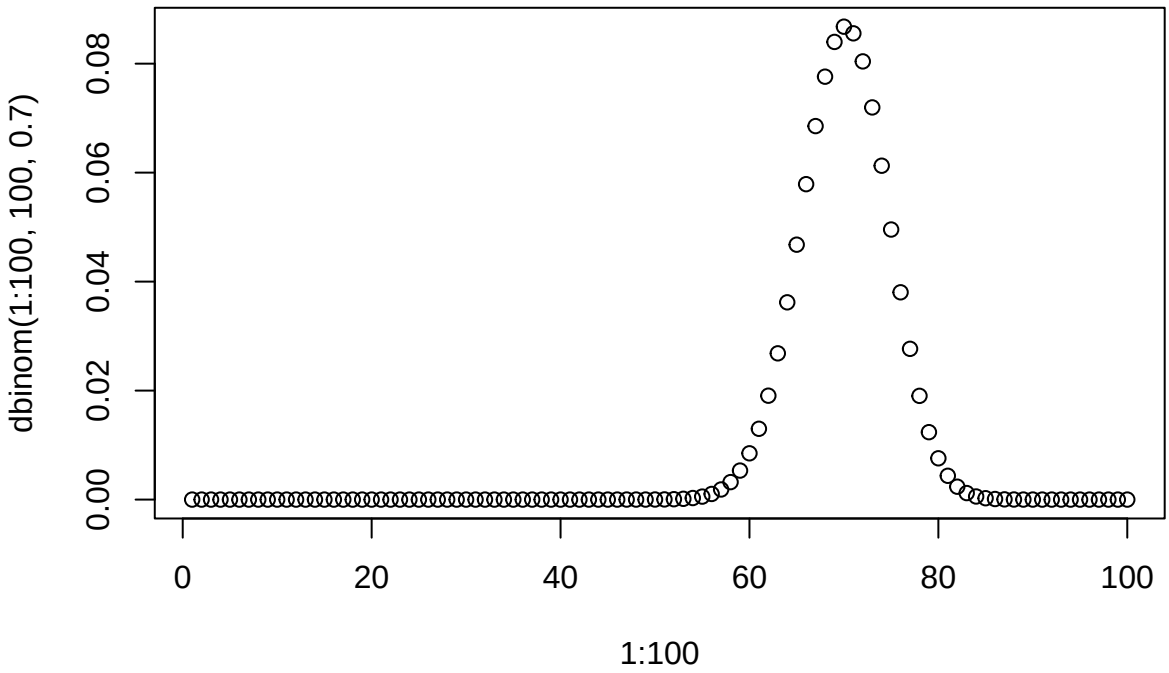
```
dbinom(6, 9, 0.7)
```

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

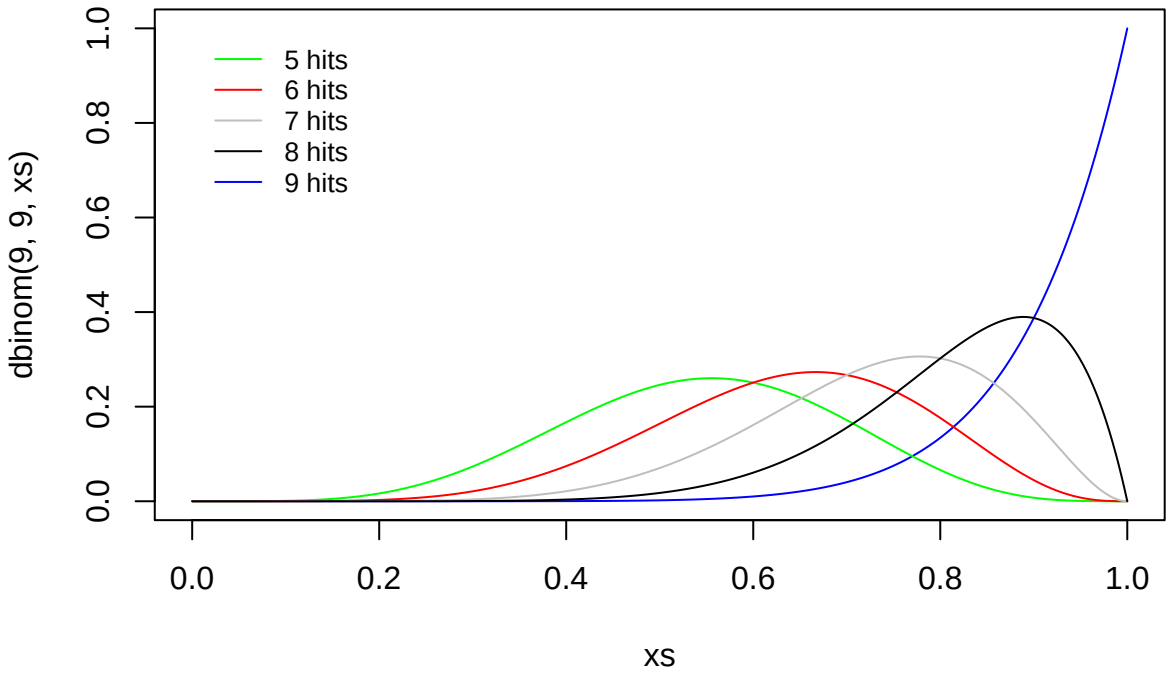
```
xs <- seq(0, 1, 0.01)
plot(xs, dbinom(6, 9, xs))
```




```
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)
```



Question 1: why is it okay to set the prior to 1 (rather than 1 / sum(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(water) = 1 to p(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(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."

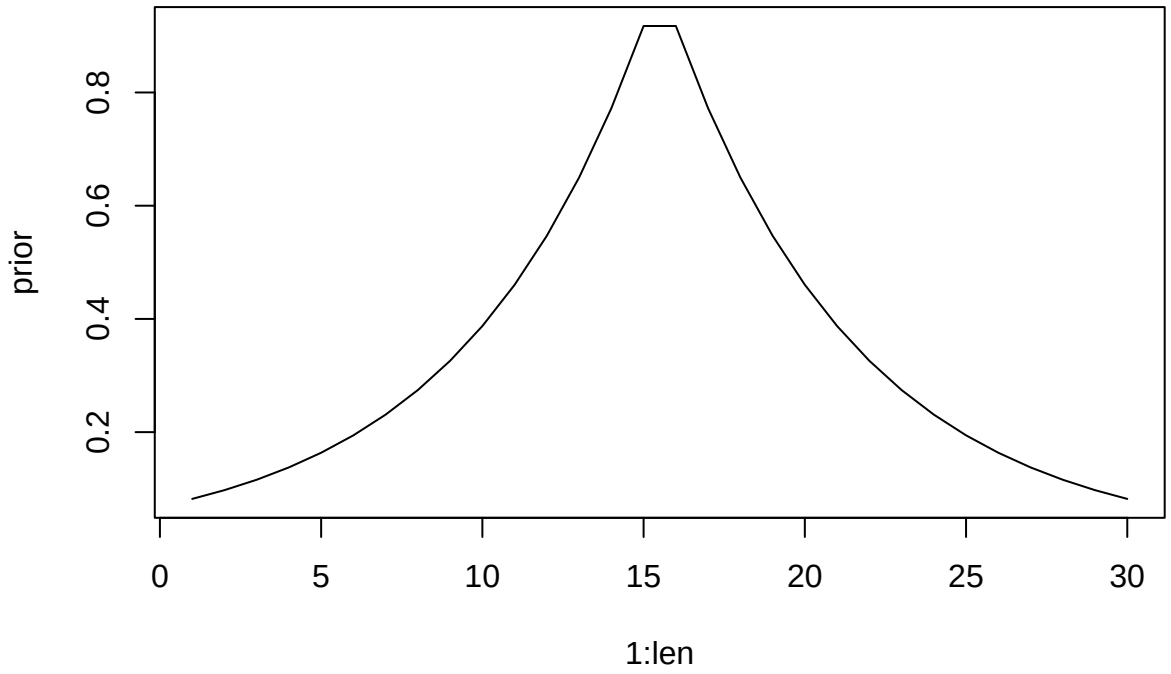
Book code

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

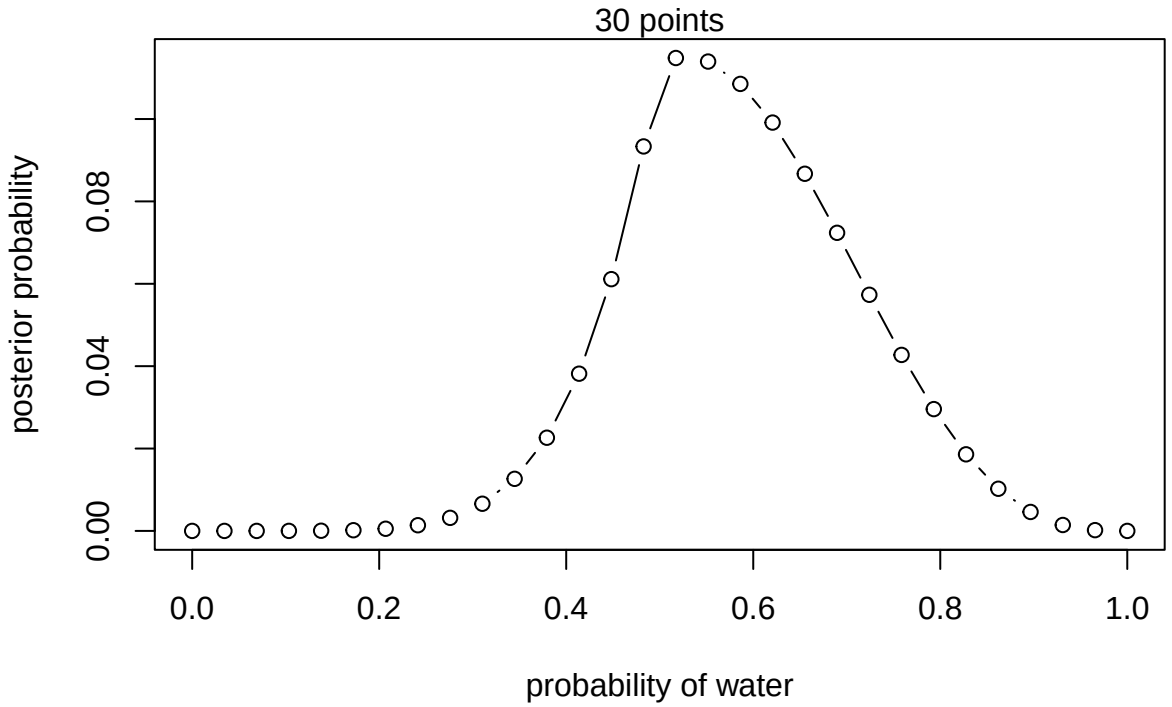
```r
prior <- exp( -5*abs( p_grid - 0.5 ) )

# 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')
```




```r
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
```

Homework

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"))
```