

IPS9 in R: Probability: The Study of Randomness (Chapter 4)

Bonnie Lin and Nicholas Horton (nhorton@amherst.edu)

July 22, 2018

Introduction and background

These documents are intended to help describe how to undertake analyses introduced as examples in the Ninth Edition of *Introduction to the Practice of Statistics* (2017) by Moore, McCabe, and Craig.

More information about the book can be found [here](#). The data used in these documents can be found under Data Sets in the Student Site. This file as well as the associated R Markdown reproducible analysis source file used to create it can be found at <https://nhorton.people.amherst.edu/ips9/>.

This work leverages initiatives undertaken by Project MOSAIC (<http://www.mosaic-web.org>), an NSF-funded effort to improve the teaching of statistics, calculus, science and computing in the undergraduate curriculum. In particular, we utilize the `mosaic` package, which was written to simplify the use of R for introductory statistics courses. A short summary of the R needed to teach introductory statistics can be found in the `mosaic` package vignettes (<http://cran.r-project.org/web/packages/mosaic>). A paper describing the `mosaic` approach was published in the *R Journal*: <https://journal.r-project.org/archive/2017/RJ-2017-024>.

Chapter 4: Probability: The Study of Randomness

This file replicates the analyses from Chapter 4: Probability: The Study of Randomness.

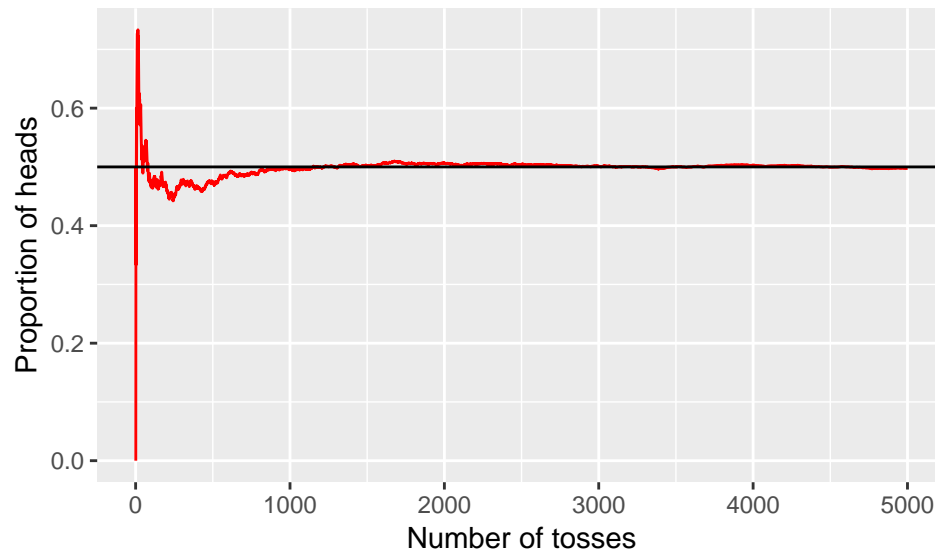
First, load the package that will be needed for this document:

```
library(mosaic)
```

Section 4.1: Randomness

The results of two trials of a coin toss simulation 5000 times are plotted as the proportion of heads, as shown in Figure 3.1 (page 216). We can emulate one trial of such simulation as a plot by typing:

```
tosses <- rbinom(n = 1:5000, size = 1, prob = 0.5)
x <- seq(1:5000)
cy <- cumsum(tosses)
phead <- (cy / x)
gf_line(phead ~ x, color = "red") %>%
  gf_labs(x = "Number of tosses", y = "Proportion of heads") %>%
  gf_hline(., yintercept = 0.5)
```



The parameters in the `rbinom()` function can be explained in the following way: `n = 1:5000` specifies the number of observations, while `size = 1` specifies the number of trials, with each trial getting a probability of success equal to 0.5. Another way to think about this is: Draw either 0 or 1, given there is a 50% chance of selecting either number, 5000 times.

Another useful function that is related to this is `xpbinom()`, which can be used to plot the probability

The R function, `runif()`, generates a random number between 0 and 1. We can demonstrate the code below:

```
runif(1)
```

```
## [1] 0.8074766
```

Since the default arguments in the function define the sample space to be all numbers between 0 and 1, all we need to specify is the number of random numbers we want outputted. Run the code above several times. Notice that every iteration gives you a different output. If you do not set a seed, every time you run the code, you will get a random number.

To demonstrate adjustments that you can make to the `runif()` call:

```
runif(2) #change the number of random numbers generated from this sample space
```

```
## [1] 0.03904229 0.11848192
```

```
set.seed(2018) #setting the seed to get reproducible results
```

```
runif(1)
```

```
## [1] 0.3361535
```

Now that we have selected a pseudorandom seed, anyone who runs this code should expect to see the output of `runif(1)` to be 0.34.

Section 4.2: Probability models

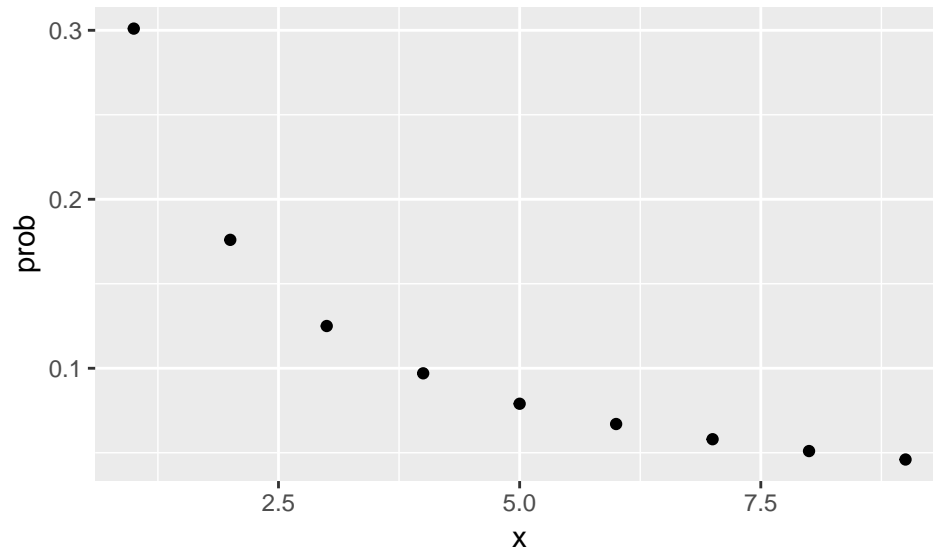
Section 4.3: Random variables

We can also display probability histograms that compare the probability model for equally likely random digits with the model given by Benford's law (page 237):

```

# Figure 4.5 (a)
eq_likely <- data.frame(Outcomes = rbinom(1000, 1:9, 0.111))
# Figure 4.5 (b)
benlaw <- c(0.301, 0.176, 0.125, 0.097, 0.079, 0.067, 0.058, 0.051, 0.046)
digits <- 1:9
bendata <- data.frame(prob = benlaw, x = digits)
gf_point(prob ~ x, data = bendata)

```

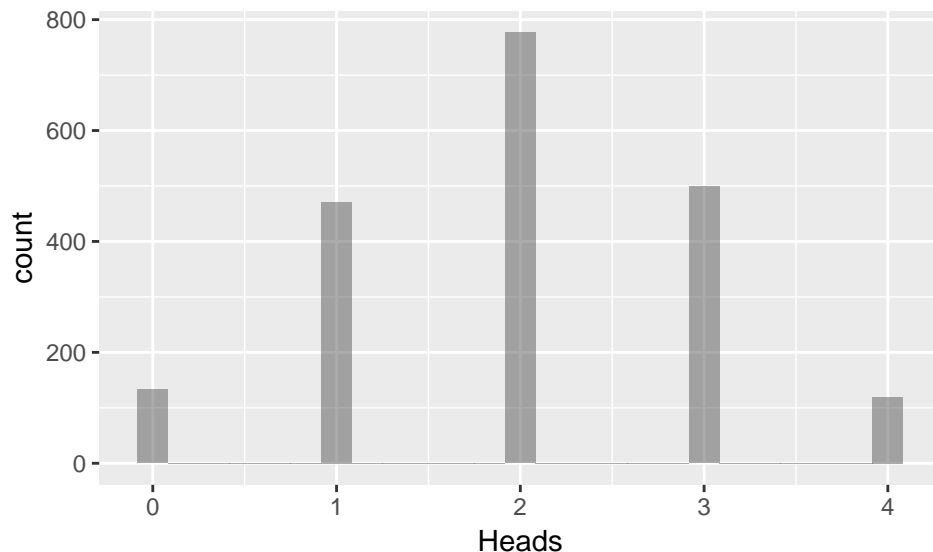


Below is the code to generate a probability histogram for the distribution of the number of heads in 2000 trials of tossing a coin four times as shown in Figure 4.7 (page 239):

```

trials <- data.frame(Heads = rbinom(2000, 4, 0.5))
gf_histogram(~ Heads, data = trials) # XX is there a way to make this look better?

```



Section 4.4: Means and variances of random variables

Section 4.5: General probability rules