Chapter 3.1 Introduction

Jim Albert and Monika Hu

Chapter 3 Conditional Probability

The Three Card Problem

- Suppose one has three cards – one card is blue on both sides, one card is pink on both sides, and one card is blue on one side and pink on the other side.
- Suppose one chooses a card and place it down showing "blue". What is the chance that the other side is also blue?
- This is a famous conditional probability problem. One is given certain information – here the information is that one side of the card is blue – and one wishes to determine the probability that the other side is blue.

A simulation experiment

- One can obtain the correct answer by simulating this experiment many times.
- Think of this experiment as first choosing a card, and then choosing a side from the card.
- There are three possible cards, which we call "Blue", "Pink" and "mixed". For the blue card, there are two blue sides; for the pink card, there are two pink sides, and the "mixed" card has a blue side and a pink side.

A R simulation

- Define a data frame `df` with two variables `Card` and `Side`.
- `sample()` function randomly chooses a card and a side by choosing a random row from the data frame. Repeat the experiment 1000 times.

```
df <- data.frame(Card = c("Blue", "Blue",
                          "Pink", "Pink",
                          "Mixed", "Mixed"),
                 Side = c("Blue", "Blue",
                          "Pink", "Pink",
                          "Blue", "Pink"))
cards <- df[sample(6, size = 1000, replace = TRUE), ]
```

Table of outcomes

```
table(cards$Card, cards$Side)
```

```
##
##          Blue Pink
##   Blue    346    0
##   Mixed   162  164
##   Pink      0  328
```

What did we learn?

- One observed “side is blue” and one are interested in the probability of the event “card is blue”.
- In this experiment, the blue side was observed $346 + 162 = 508$ times.
- Of these blue sides, the card was blue 346 times.
- So the probability the other side is blue is approximated by 346 / 508 which is close to the exact probability of $2/3$.

Selecting slips of paper

- Suppose one has a box that has 6 slips of paper labeled with the numbers 2, 4, 6, 8, 10, and 12.
- One selects two slips at random without replacement from the box.
- Here are the ${}_6C_2 = 15$ possible outcomes

$S = \{$(2, 4), (2, 6), (2, 8), (2, 10), (2, 12), (4, 6), (4, 8), (4, 10), (4, 12) (6, 8), (6, 10), (6, 12), (8, 10), (8, 12), (10, 12)$\}$.

Finding a probability

- Suppose one are interested in the probability the sum of the numbers on the two slips is 14 or higher.
- Assume that the 15 outcomes listed above are equally likely.
- One sees there are 9 outcomes where the sum is 14 or higher and so

$$Prob(\text{sum } 14 \text{ or higher}) = \frac{9}{15}.$$

New information

- Suppose one is given some new information – both of the numbers on the slips are single digits.
- Given this information, one now has only six possible outcomes.
- This new sample space is called the reduced sample space based on the new information.

$S = \{(2, 4), (2, 6), (2, 8), (4, 6), (4, 8), (6, 8)\}$

A conditional probability

- One finds the probability *Prob*(sum is 14 or higher) given that both of the slip numbers are single digits.
- Since there is only one way of obtaining a sum of 14 or higher in our new sample space, one sees

$$\mathit{Prob}(\mathrm{sum}\,14\,\mathrm{or}\,\mathrm{higher}) = \frac{1}{6}.$$

- Notation: we write

$\mathit{Prob}(\text{sum is 14 or higher} \mid \text{both numbers are single digits}).$

- Initially, the probability of 14 and higher was high $(9/15)$, but given the new information, the probability dropped to $1/6$.