Download codes for this lecture at:

Lecture 2 codes: https://www.dropbox.com/s/bnvk63bbwhgemhq/Lec2.R?dl=0

Review

We went over the basics of `R` with Rstudio cloud in Lecture 1. We also started learning about *probability* and *randomness*, and how to simulate experiments in `R`.

Let's review with a little more formality. Suppose we flip a p-coin, and let Y be the outcome: either 1 (heads) or 0 (tails). Then

$$\Pr(Y = 1) = p\,.$$

If we don't know what p is, then just seeing one coin-toss we can't understand what value of p we have obtained.

Suppose we flip the coin n times: $Y_1, Y_2, \dots, Y_n$, then for each Y_i,

$$\Pr(Y_i = 1) = p\,.$$

So now we can borrow information from n different coin-tosses. A good guess of what p would be $\hat{p}$, where we define $\hat{p}$ as

$$\hat{p} := \frac{\text{Number of heads}}{\text{Total coin tosses}} = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}\,.$$

Thus, a good guess of p is the average of the observed numbers. Such a guess, that is based on random events, is called an *estimate*.

Using this, try the Worksheet 1, question 1 and 2 again.

Random draws in (a, b)

When the number of outcomes is countable, probabilities are relatively easy to understand. For example in tossing a coin, there are two outcomes. In rolling a regular die, there are six outcomes.

But what about when the number of outcomes are uncountably infinite?

For example, consider the experiment of drawing a random number between $(0, 1)$. How many outcomes are possible?

Let's try to do such experiments in `R`:

```r
## Uniform draws between (0,1)
runif(1, min = 0, max = 1)
```

```r
## Uniform draws between (-4, 6)
runif(1, min = -4, max = 6)
```

How do we check, whether these are uniform draws?

```r
## Q: How do you check whether draws are uniform??
## A: Get many draws and then visualize!
uni_draws = runif(1000, min = 0, max = 1)
hist(uni_draws, main = "Histrogram of Uniform draws (0,1)")
```

When you draw the above histogram plot in `R`, you will see how the numbers drawn are *distributed* between $(0, 1)$.

Q: What is the probability that a sampled number between $(0, 1)$ *is 0.5?*

A: Clearly, there are infinite choices of numbers between $(0, 1)$ or any interval. The number of ways to generate the event is 1, and the number of possible outcomes are ∞. So:

$$\Pr(\text{number is } 0.5) = 0!$$

And infact, this is true for all numbers. So for uncountably infinite problems, we *cannot* study probabilities in the same way as before. Instead, we define sets.

Consider the event: $A =$ number is in a set $(0, .2)$. Then,

$$\Pr(A) = \frac{.2 - 0}{1 - 0} = .2\,.$$

In general, when obtaining uniform samples from (a, b). The probability of obtaining a number between $(a_1 - b_1)$ is

$$\Pr(A) = \frac{b_1 - a_1}{b - a}\,.$$

Naturally, $a_1 > a$ and $b_1 < b$. Let's do this more formally.

A little more formally

Let $X \sim \text{Uniform}(0, 1)$. That is, let X be a uniformly chosen number between 0 and 1. Define Y

$$Y = \mathbb{I}\left\{a_1 < X < b_1\right\}\,.$$

Then Y is either 0 or 1. Since the outcomes now are either 1 or 0, we can view 1 as a success (heads) and 0 as a failure (tails). That is, the event Y is like a coin-toss where

$$\Pr(Y = 1) = \Pr(a_1 < X < b_1) := p \quad (say)\,.$$

Let's do this in R!

```
### Simulate the experiment Y = I( 0 < X < .2)
# where X = Uniform(0,1).
uni_draws = runif(1, min = 0, max = 1)

# Check whether uni_draw < .2
# multiplying by 1, makes it a number, rather than TRUE/FALSE
check <- 1*(uni_draws < .2)
check
```

For some chosen n, we get different

$$X_1, X_2, X_3, \ldots, X_n \sim \text{Uniform}(0, 1)$$

For each $i = 1, \ldots, n$, set

$$Y_i = \mathbb{I}\{a_1 < X_i < b_1\}\ .$$

Based on each of the X_i's were drawn, some Y_i's will be 0 and some will be 1. If we forget that the X's were involved, then the Ys essentially behave like coin-tosses (potentially unfair). So we can obtain an estimate of p, as before. Let's try it in R for $(a, b) = (0, 1)$ and $(a_1, b_1) = (0, .20)$.

```
## Verify the Pr(Y = 1)
n = 1000
uni_draws = runif(n, min = 0, max = 1)

# Check whether uni_draw < .2
# multiplying by 1, makes it a number, rather than TRUE/FALSE
check <- 1*(uni_draws < 0.2)
check

## Estimate of Pr(Y = 1)
mean(check)
```