Quantitative Methods Edition

Timur Öztürk

Lecture 3: Probability and Statistical Calculations

April 25, 2025

Introduction

After learning about the data structure and basic syntax in R, we can now talk about probability. You should remember the theory behind today from your previous courses, but for a refresher I encourage you to have a look at the websites and further resources you can find in my GitHub repository.

The lecture material follows Chapter 2 of https://www.econometrics-with-r.org.

Probability I

A basic function to draw random samples from a specified set of elements is the function **sample()**, see **?sample**. We can use it to simulate the random outcome of a dice roll. Let's roll the dice!

```
sample(1:6, size=1)
```

We can also repeat this experiment multiple times. Some essential concepts, as you might remember:

```
# generate the vector of probabilities
probability <- rep(1/6, 6)

# plot the probabilities
plot(probability,
     xlab = "Outcomes",
     ylab="Probability",
     main = "Probability Distribution",
     pch=20)
```

Probability II

What about the PDFs?

```
sample(c("H", "T"), 1)
k <- 0:50
# assign the probabilities
probability <- dbinom(x = k,
                      size = 50,
                      prob = 0.5)
# plot the outcomes against their probabilities
plot(x = k,
    y = probability,
    ylab="Probability",
    main = "Probability Distribution Function",
    pch=20)
```

Probability III

And, of course, CDFs.

```
# compute cumulative probabilities
prob <- pbinom(q = k,
               size = 50,
               prob = 0.5)
# plot the cumulative probabilities
plot(x = k,
    y = prob,
    ylab="Probability",
    main = "Cumulative Distribution Function",
    pch=20)
```

Setting the Seed

One of the main reasons we use R is the reproducibility of our results. In order to be able to reproduce the same results, we need to set a seed. This is done by the function **set.seed()**.

```
# set seed for reproducibility
set.seed(1)

# compute the sample mean of 10000 dice rolls
mean(sample(1:6,
            10,
            replace = T))
```

Calculating Expected Value

```
# define functions
f <- function(x) 3 / x^4
g <- function(x) x * f(x)
h <- function(x) x^2 * f(x)

# compute area under the density curve
area <- integrate(f,
                  lower = 1,
                  upper = Inf)$value

# compute E(X)
EX <- integrate(g,
                lower = 1,
                upper = Inf)$value
```

Exercises

The exercises are based on

https://www.econometrics-with-r.org/2.3-exercises-2.html

we will continue there.

- Suppose you are the lottery fairy in a weekly lottery, where 6 out of 49 unique numbers are drawn. Draw the winning numbers of the week.
- Compute the value of the standard normal density function at the point 3.
- Generate 10 random numbers from a normal distribution with mean 2 and standard deviation 12.