**Mathematical Expectation**



We learned about some PDFs and how they are used to quantify the distribution of items in a population. In this section, we are going to summarize information in the population relative to the PDFs. Specifically, we are going to look at the “expected value” of a random variable and find it using its PDF. These expected values are going to represent parameters. Thus, they will summarize items in the population.

Expected value of a random variable

Given the use of a PDF for a population, we may want to know what value of Y we would expect on average to obtain. This is found through the expected value of Y, E(Y). This is our population mean μ!

Let Y be a random variable with PDF f(y). The population mean or expected value of Y is

μ = E(Y) = 

when Y is discrete and

μ = E(Y) = 

when Y is continuous.

Notes:

* In calculus you learned about something similar. It may have been called “moment about the y-axis”.
* You will often see μ written as μy to emphasize that the expected value is for the random variable Y. This notation is helpful when there are multiple random variables.
* In the discrete case, you can think of μ as a weighted average. The weight for each y is f(y) = P(Y = y).
* μ is a parameter because it is a numerical summary measure of the population.

Example: Let’s play Plinko! (plinko.xlsx)

Let X be a random variable denoting the amount won for one chip.[[1]](#footnote-1) Below are 5 different PDFs for the amount won.

|  |  |
| --- | --- |
|  | **Drop Plinko chip in column above:** |
|  | **$100** | **$500** | **$1,000** | **$0** | **$10,000** |
| **x** | **f(x)** | **f(x)** | **f(x)** | **f(x)** | **f(x)** |
| **$100** | 0.1667 | 0.1339 | 0.0822 | 0.0359 | 0.0176 |
| **$500** | 0.3571 | 0.3080 | 0.2204 | 0.1287 | 0.0882 |
| **$1,000** | 0.2976 | 0.2991 | 0.2862 | 0.2545 | 0.2353 |
| **$0** | 0.1429 | 0.1920 | 0.2796 | 0.3713 | 0.4118 |
| **$10,000** | 0.0357 | 0.0670 | 0.1316 | 0.2096 | 0.2471 |
|  |  |  |  |  |  |
| **** | $850.00 | $1,136.16 | $1,720.39 | $2,418.26 | $2,751.76 |

When dropping the chip above a $100 slot:

μ = E(X) =  100×0.1667 + 500×0.3571 + 1000×0.2976 + 0×0.1429 + 10000×0.0357 = 850

where T = {100, 500, 1000, 0, 10000}.

This is a good example of why you can think of μ as a weighted average.

Questions:

* What is the optimal place to drop the chip in order to maximize your expected winnings?
* Compare what actually occurred in one particular year to what is expected:

|  |  |
| --- | --- |
|  | **Drop Plinko chip in column above:** |
|  | **$100**  | **$500**  | **$1,000**  | **$0**  | **$10,000**  |
| **x** | **Count** | **Count** | **Count** | **Count** | **Count** |
| **$100**  | 0 | 3 | 2 | 5 | 2 |
| **$500**  | 0 | 3 | 12 | 5 | 6 |
| **$1,000**  | 1 | 3 | 12 | 19 | 11 |
| **$0**  | 0 | 2 | 9 | 21 | 8 |
| **$10,000**  | 0 | 1 | 4 | 9 | 4 |
|  |  |  |  |  |  |
| **Total chips** | 1 | 12 | 39 | 59 | 31 |
| **Average won** | $1,000  | $1,233  | $1,492  | $1,898  | $1,748  |

Why are these averages different from what we would expect?

* The sample size is small. As the sample size gets bigger, we would expect the sample average to approach the population mean.
* There is variability from one sample to the next. This implies that we would not expect the same number of observed values for other years when the game is played. More will be discussed about variability later.
* Possibly one of the underlying assumptions behind the calculation of the probabilities is incorrect. While we did not discuss these assumptions in class, the main one is that the probability of going to the left or right ONE slot is 0.5 as a plinko chip hits a peg on the board. If you watch the plinko game, you will notice that chips can often go more than ONE slot to the left or right after hitting a peg.

Example: Transaction time (TransExpect.R, TransExpect.ipynb)

The number of seconds between transactions (e.g., purchases) on a website can be represented by a PDF. Let X be a random variables representing the seconds. Suppose the PDF for X is





Find the expected number of seconds between transactions. In other words, this will be the average seconds.

μ = E(X) =  = 

Use integration by parts:

Let u = x dv = 

du = dx v =  = -e-x/30

Then



Note that by L’hopital’s rule,



Then



Thus, the number of seconds between transactions is 30 on average.

R:

> pdf.x <- function(x) {

 x \* 1/30\*exp(-x/30)

 }

> # E(X)

> integrate(f = pdf.x, lower = 0, upper = Inf)

30 with absolute error < 2.5e-05

Sage:



Expected values for functions of random variables

Sometimes, functions of random variables are of interest when finding the expected value. We will discuss shortly one very important example. In general, here is how the expected value can be found:

Let Y be a random variable with PDF f(y). The mean or expected value of the random variable g(Y) is

μg(y) = E[g(Y)] = 

when Y is discrete, and

μg(y) = E[g(Y)] = 

when Y is continuous.

Be very careful here! The function, g(Y), is NOT a PDF. It is a function of a random variable.

Example: Transaction time (TransExpect.R, TransExpect.ipynb)

Find E[|X-30|]. What does this mean in terms of the problem?







R: Notice how I was able to pass mu into pdf.abs().

> # E(|X - mu|)

> mu <- integrate(f = pdf.x, lower = 0, upper =

 Inf)$value

> mu

[1] 30

> pdf.abs <- function(x, mu) {

 abs(x - mu) \* 1/30\*exp(-x/30)

 }

> integrate(f = pdf.abs, lower = 0, upper = Inf, mu = mu)

22.07277 with absolute error < 0.00019

Sage: Notice the problem with the initial attempt.



Find E(X2). What does this mean in terms of the problem?

E(X2) =  = 

Use integration by parts twice to obtain:

E(X2) = 1800

Notice that E(X2) ≠ [E(X)]2 = E(X)2

R: Notice the potential problem with the initial attempt.

> pdf.xsq <- function(x) {

 x^2 \* 1/30\*exp(-x/30)

 }

> integrate(f = pdf.xsq, lower = 0, upper = Inf)

1800 with absolute error < 0.023

> integrate(f = pdf.xsq, lower = 0, upper = Inf, rel.tol

 = 0.00000001)

1800 with absolute error < 7.1e-07

Sage:



Specific case for g(Y): E(aY + b) = aE(Y) + b for some constants a and b.

All integrals are performed over all possible values of y



Example: Transaction time

While this example is not necessarily realistic, we can still use it to illustrate the result.

Find E(2X+1) = 2E(X) + 1 = 2×30 + 1 = 61.

Expected value for more than one random variable

Let X and Y be random variables with joint PDF f(x,y). The mean or expected value of the random variable g(X,Y) is

μg(x,y) = E[g(X,Y)] = 

when X and Y are discrete, and

μg(x,y) = E[g(X,Y)] = 

 when X and Y are continuous.

There is one particular case which will be important in the next section - when g(X,Y) = XY. In the continuous case, notice the difference between E(XY) and E(X)E(Y):





When are these two quantities equal?

Notice how E(X) and E(Y) are found: g(X,Y) = X or g(X,Y) = Y

Relating expected values back to samples

Example: A sample of times between transactions (Transaction.R, ExampleSample.csv)

Below is a sample that comes from a population characterized by the PDF of



> set1 <- read.csv(file = "ExampleSample.csv")

> head(set1)

 Observation x

1 1 8.782565

2 2 2.810164

3 3 71.201537

4 4 16.549732

5 5 23.581126

6 6 2.165690

> tail(set1)

 Observation x

995 995 38.96618

996 996 44.53581

997 997 58.59082

998 998 17.57703

999 999 34.00815

1000 1000 52.04435

Examine how close the sample mean is the E(X)!

> pdf.x <- function(x) {

 x \* 1/30\*exp(-x/30)

 }

> # E(X)

> integrate(f = pdf.x, lower = 0, upper = Inf)

30 with absolute error < 2.5e-05

> mean(set1$x)

[1] 28.93222

Questions:

* How could you estimate E(X2) using this sample and what would you expect it to be approximately?
* Will the sample mean tend to get closer or farther way from E(X) as the sample size increases?

1. Because changes in the amount won have occurred over time and will likely occur in the future, I decided to use these particular amounts. [↑](#footnote-ref-1)