Elementary Probability Review
DATA 5600, Fall 2021

This is a review of elementary probability that will be useful for our study of regression for data analytics. It is based on coverage Wooldridge (2004).

Observational data sets econometrics apart from statistics. We will view an economic variable as an **outcome** from a **random process** not under the control of the researcher.

The descriptive term for this underlying mechanism is the **data–generating process**, or **DGP**.

We view the outcome variable X as a random variable because until it is observed we are not certain about its value.

An **experiment** is a procedure that can (at least in theory) be infinitely repeated and has a well–defined set of outcomes.

An example: flip a coin 10 time and count the number of heads. Each time the experiment is repeated the outcome will be an integer between 0 and 10.

A **random variable** is a variable that takes on numerical values and has an outcome that is determined by an experiment.

An example:

- An airline wants to decide how many reservations to book for a flight with 100 seats.
- If fewer than 100 people want reservations they should book them all.
- If more than 100 people want reservations a safe bet may be to only book 100. But not everyone will show up, resulting in lost revenue.
- If they book too many they will have to compensate passengers for having to bump them.

By convention, random variables are denoted by uppercase variables, such as X, Y, and Z.

The corresponding outcomes are denoted by lowercase letters x, y, z.

In the coin flipping example X denotes the number of heads in 10 flips. We don't know ahead of time what value X will take, but we know it will be in the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. A particular outcome may be $x = 7$.

Random variables are defined to take on numerical values. So even in the coin flipping example, where the outcomes are "heads" and "tails", we code the outcomes as follows:

- $X = 1$ for heads (success)
- $X = 0$ for tails (failure)

A random variable that only takes on values 0 and 1 is a **Bernoulli** (or **binary**) **random variable**.

A discrete random variable is one that takes on only a finite (or countably infinite) number of values. A binary variable is the simplest case of a discrete random variable. The only quantity that we need to completely describe its behavior is the probability that $X = 1$.

In the coin flipping example (if the coin is "fair") then

$$P(X = 1) = \tfrac{1}{2}$$

$$P(X = 0) = \tfrac{1}{2}$$

Consider again the airline's problem of booking seats on a flight. We can analyze this with several binary variables. For a randomly selected passenger define a binary variable as $X = 1$ if she shows up for the flight, and $X = 0$ otherwise. There is no reason to believe in this case that $P(X = 1) = \frac{1}{2}$, so we will define a *parameter* θ so that:

$$P(X = 1) = \theta$$

$$P(X = 0) = 1 - \theta$$

For example, if $\theta = 0.75$, then there is a 75% chance of the passenger showing up and 25% chance of not showing up. In a real–life business situation the actual value of θ is crucial in determining the airline's strategy.

Methods for *estimating* θ, given historical data on airline reservations is the subject of *mathematical statistics*.

Generally, a discrete random variable is completely described by listing the set of possible outcomes and the associated probability that it takes on each value.

If X has k possible values $\{x_1, x_2, \ldots, x_k\}$ then the probabilities $p_1, p_2, \ldots, p_k$ are defined by:

- $p_j = P(X = x_j)$, for $j = 1, 2, \ldots, k$
- $0 \leq p_j \leq 1$
- $\sum_{j=1}^{k} p_j = 1$

The **probability function** or (**PDF**) of X summarizes the information concerning the possible outcomes of X and the corresponding probabilities:

$$f(x_j) = p_j, \quad j = 1, 2, \ldots, k$$

For any real number x, $f(x)$ is the probability that the random variable X takes on the particular value x.

An example: Suppose that X is the number of free throws made by Larry Bird out of two attempts. X can take on the three values $\{0, 1, 2\}$. Assume the PDF of X is given by

$$f(0) = 0.20$$

$$f(1) = 0.44$$

$$f(2) = 0.36$$

We can calculate the probability that Larry Bird will make at least one free throw:

$$\begin{aligned} P(X \geq 1) &= P(X = 1) + P(X = 2) \\ &= 0.44 + 0.36 \\ &= 0.80 \end{aligned}$$

We can graph this discrete PDF as follows:

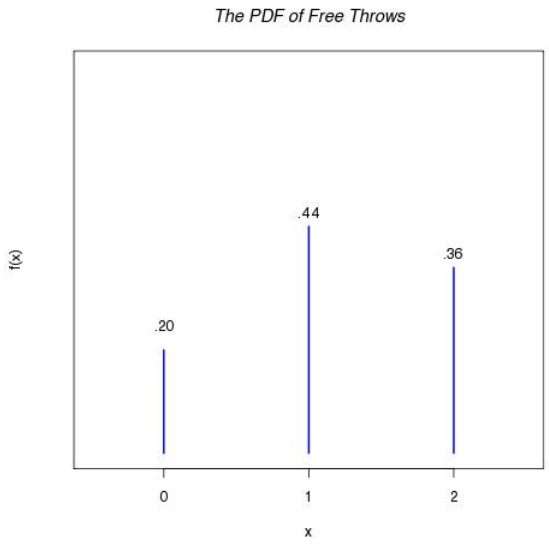

Figure 1: The Probability Density Function of Larry Bird Free Throws

When dealing with more than two random variables we subscript the PDF's as follows:

- f_x is the PDF of X
- f_y is the PDF of Y

A variable X is a **continuous random variable** if it takes on any real value with *zero* probability. A continuous random variable X can take on so many possible values that they are not countable, so logical consistency requires that each one has probability zero.

Examples:

- Prices
- Wages
- Interest rates
- Height
- Weight
- Waiting time