ST 516: Foundations of Data Analytics

Random Variables Introduction

Definition

A **random variable** is a value assigned to each possible outcome of a random trial.

We often use capital letters like X, Y, and Z to denote random variables.

Note that for any given random trial, there are many, many possible random variables we could define. Here are a few examples.

Random Variables Example 1

- Random Trial: Roll a die
- Outcome space (set of all possible outcomes): The faces '1', '2', '3', '4', '5', and '6'.
- Potential Random Variables:

Random Variables Example 1

- Random Trial: Roll a die
- Outcome space (set of all possible outcomes): The faces '1', '2', '3', '4', '5', and '6'.
- Potential Random Variables:

$$X = \text{The number showing on the die}$$

Random Variables Example 1

- Random Trial: Roll a die
- Outcome space (set of all possible outcomes): The faces '1', '2', '3', '4', '5', and '6'.
- Potential Random Variables:

$$X = \text{The number showing on the die}$$

$$Y = \begin{cases} 10 & \text{If the number showing on the die is odd} \\ 3.14 & \text{If the number showing on the die is even} \end{cases}$$

Random Variables Example 1

- Random Trial: Roll a die
- Outcome space (set of all possible outcomes): The faces '1', '2', '3', '4', '5', and '6'.
- Potential Random Variables:

$$X = \text{The number showing on the die}$$

$$Y = \begin{cases} 10 & \text{If the number showing on the die is odd} \\ 3.14 & \text{If the number showing on the die is even} \end{cases}$$

$$Z = \begin{cases} -4 & \text{If the number showing on the die is '1' or '6'} \\ 1 & \text{If the number showing on the die is '2' or '5'} \\ 6 & \text{If the number showing on the die is '3' or '4'} \end{cases}$$

Random Variables Example 1

You could imagine that each of the potential random variables on the previous slides is the amount of money (in \$) that you win or lose if that particular outcome occurs.

Then each random variable represents a different game. Which 'game' (that is, the games represented by X, Y, and Z) would you most like to play, and why?

We will talk more about this when we discuss 'means' or 'averages'.

Random Variables Example 2

- Random Trial: Select a randomly chosen student from OSU
- Outcome space (set of all possible outcomes): All students at OSU
- Potential Random Variables:

Random Variables Example 2

- Random Trial: Select a randomly chosen student from OSU
- Outcome space (set of all possible outcomes): All students at OSU
- Potential Random Variables:

$$X = \text{GPA of chosen student}$$

Random Variables Example 2

- Random Trial: Select a randomly chosen student from OSU
- Outcome space (set of all possible outcomes): All students at OSU
- Potential Random Variables:

$$X = \text{GPA of chosen student}$$

$$Y = \text{Number of siblings of chosen student}$$

Random Variables Example 2

- Random Trial: Select a randomly chosen student from OSU
- Outcome space (set of all possible outcomes): All students at OSU
- Potential Random Variables:

$$X = \text{GPA of chosen student}$$

$$Y = \text{Number of siblings of chosen student}$$

$$Z = \begin{cases} 1 & \text{if chosen student is male} \\ 0 & \text{otherwise} \end{cases}$$

Random Variables Discussion

In many trials (or experiments) we may be interested in several random variables. For instance, if we randomly select an OSU student, we might want to record their GPA, number of credits taken, and their class-year.

In a clinical trial for a potential headache medicine, we might perform a randomized experiment by randomly assigning subjects to a new pharmaceutical or to placebo. For each subject, we might want to record their weight, blood pressure, smoking status, and their response to the treatment (drug or placebo).

The point is that you are not restricted to choosing just *a single* random variable for a given trial–in many cases, we consider several. We will later talk about how random variables can be related to each other.

Random Variable Support

Definition

The set of all possible values that a random variable can take on is called the **support** of that random variable.

In the previous example, what is the support of each of the random variables?

$$X = \text{GPA of chosen student}$$

$$Y = \text{Number of siblings of chosen student}$$

$$Z = \begin{cases} 1 & \text{if chosen student is male} \\ 0 & \text{otherwise} \end{cases}$$

Random Variable Support

First random variable:

$$X = \text{GPA of chosen student}$$

Random Variable Support

First random variable:

$$X = \text{GPA of chosen student}$$

Since GPAs can be any value between 0 and 4.0, the support of the random variable X is $(0, 4.0)$.

Random Variable Support

Second random variable:

$$Y = \text{Number of siblings of chosen student}$$

Random Variable Support

Second random variable:

$$Y = \text{Number of siblings of chosen student}$$

Since a person's number of siblings can be any non-negative integer, the support of the random variable Y is $\{0, 1, 2, 3, \ldots\}$.

Random Variable Support

Third random variable:

$$Z = \begin{cases} 1 & \text{if chosen student is male} \\ 0 & \text{otherwise} \end{cases}$$

Random Variable Support

Third random variable:

$$Z = \begin{cases} 1 & \text{if chosen student is male} \\ 0 & \text{otherwise} \end{cases}$$

There are only two possible cases for the value of this third random variable: either the student is male or the student is not male. Therefore, the support of the random variable Z is $\{0, 1\}$.

Random Variable Classification: Discrete vs. Continuous

- A random variable that can take on any of a continuous set of values is called a **continuous random variable**.
- A random variable that can only take on certain separate values is called a **discrete random variable**.

Another way of saying this same thing is to talk about the support of the random variable:

- A random variable whose support is a continuous interval is continuous.
- A random variable whose support is a collection of separate values is discrete.

The following examples may help make this distinction more clear.

Discrete vs. Continuous Random Variable Examples

Example:

- Random Trial: Randomly select a student at OSU
- Outcome space (set of all possible outcomes): {All students at OSU}
- Potential Random Variables:

$$X = \text{GPA of chosen student}$$

$$Y = \text{Number of siblings of chosen student}$$

$$Z = \begin{cases} 1 & \text{if chosen student is male} \\ 0 & \text{otherwise} \end{cases}$$

Classify each of these random variables as discrete or continuous.

Discrete vs. Continuous Random Variable Examples

First random variable:

$$X = \text{GPA of chosen student}$$

Support: $(0, 4.0)$

Discrete vs. Continuous Random Variable Examples

First random variable:

$$X = \text{GPA of chosen student}$$

Support: $(0, 4.0)$

Since the support of the random variable X is a continuous interval, X is a continuous random variable.

Discrete vs. Continuous Random Variable Examples

Second random variable:

$$Y = \text{Number of siblings of chosen student}$$

Support: $\{0, 1, 2, \ldots, \}$

Discrete vs. Continuous Random Variable Examples

Second random variable:

$$Y = \text{Number of siblings of chosen student}$$

Support: $\{0, 1, 2, \ldots,\}$

Since the support of the second random variable Y is a collection of separate values, Y is a discrete random variable.

Discrete vs. Continuous Random Variable Examples

Third random variable:

$$Z = \begin{cases} 1 & \text{if chosen student is male} \\ 0 & \text{otherwise} \end{cases}$$

Support: $\{0, 1\}$

Discrete vs. Continuous Random Variable Examples

Third random variable:

$$Z = \begin{cases} 1 & \text{if chosen student is male} \\ 0 & \text{otherwise} \end{cases}$$

Support: $\{0, 1\}$

Since the support of the third random variable Z is a collection of separate values, Z is a discrete random variable.