MATH 320: Probability

Lecture 7: Random Variables

Chapters 2 and 3: Distributions (2.1 and 3.1)

Why do we study statistics?

- The main purpose of studying statistics is because we want to study experiments and their outcomes.
- We want to analyze data from experiments numerically. But, outcomes are not always quantitative.
  - So we have to assign numbers to outcomes. Thus, random variables connect outcomes to numbers.
  - The advantage using random variables is that they are easily summarized.
- Intuitive definition: A **random variable** is a numerical quantity whose value depends on chance.

Types of random variables (RVs)

- Examples: Determine if each describes a RV.

  i.e. Is the outcome (a) is a number? (b) depends on chance?

  Discrete

  1. You are tossing a coin twice and will bet on the number of heads.
  2. You go to Las Vegas and begin to put quarters in a slot machine. Let X be the number of quarters you play in order to first win of any amount.
  3. You are tossing a coin twice and will bet on specific outcomes such as HT.

  Continuous

  4. A resident of Muncie is selected at random, and their height is measured.

	#	chance		Rv
1.	✓	✓	⇒	✓
2.	✓	✓	⇒	✗
3.	✗	✓	⇒	✓
4.	✓	✓	⇒	✓

- Similar to sample spaces, there are different kinds of random variables.

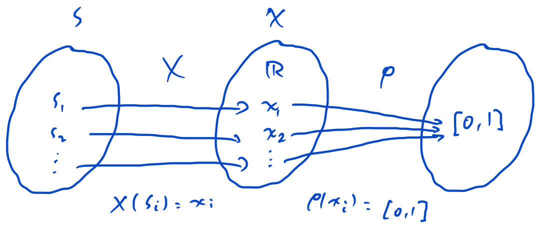
  This will be a very important distinction to make at the start of every single problem for the rest of the course.
- Random variables can be discrete (only distinct values are possible) or continuous (measured on a continuous scale).
  - When classifying a random variable as discrete or continuous, we are really just identifying the kind of mathematical model we will use.
  - Calculus-based mathematics is the most efficient way to analyze a random variable such as heights (which we may only measure as discrete to a certain precision).

Definitions and notation

- Functions *map* the input (domain, support) to the output (range).

- Our general definition of probability was a way to assign a probability $P(A)$ to any event A where all the axioms needed to be satisfied. This, more formally, is a function.

- A **random variable** is a function from a sample space S into real numbers.



	Random variable	Probability
Input:	outcome s_i	any real # x_i
Output:	any real # x_i	$[0,1]$
Maps:	$S \longrightarrow \mathbb{R}$	events $\longrightarrow [0,1]$

- Notation: We will use uppercase letters, such as X, Y, Z, ... to denote a random variable and lowercase letters, such as x, y, z, ... to denote a particular value that a random variable may assume.

- Definition: The set of possible values of X is the **range** of X, $\mathcal{X}$.

- Summary of notation:
  - X = Random variable.
  - x_i = Individual values of X.
  - $\mathcal{X}$ = Range of X $\rightarrow$ set of all $x_i = \{x_1, x_2, \ldots\}$ or $[x_a, x_b]$

- It is important to know the distinction between the outcomes in an experiment (sample space) and the range.

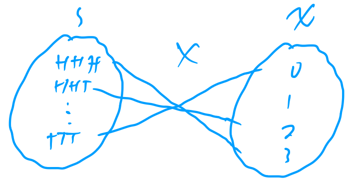
- Examples:

  1. Toss three fair coins and observe the results. Let X equal the number of heads obtained.

     (a) What is the sample space and range of X?

$S = \{HHH, HHT, \ldots, TTT\} \longrightarrow \mathcal{X} = \{0, 1, 2, 3\}$

(b) Show the connection between S and X.

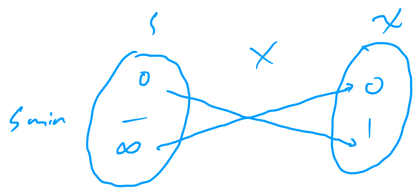


2. Let X be the time to failure for a machine part. Find the range.

$$\mathcal{X} : [0, \infty)$$

3. You are waiting for the bus to arrive. If it arrives in under 5 minutes, you will get on the bus. If not, you will walk to your destination.

   Let X be the random variable such that $X = 1$ if you get on the bus and $X = 0$ if you walk. Is X a continuous or discrete random variable?



sample space is continuous

Range of × is discrete $\Rightarrow$ RV discrete

- Types of random variables definitions

  X is a **discrete random variable** if the <u>range $\mathcal{X}$</u> is a finite or countable set.

  X is a **continuous random variable** if the <u>range $\mathcal{X}$</u> is an interval (or union of intervals) on the real number line.

Connection between random variables and probability

- We would like use random variables to express events, because we can calculate probabilities of events.
- Notation: $\{X = x\}$ is the set of <u>outcomes</u> in the sample space assigned the value x by the random variable X.

  $X = x$ means the random variable X was realized with a specific value x.

  So it is an <u>event</u>. As a result, we can compute the probability of $\{X = x\}$.
- Notation: We used to have events like $A \cap B$ or now $\{X = x\}$ in $P(\cdot)$, but we will now use $P(X = x)$ for simplicity.

Example: Continuing the previous three coin toss scenario, find the following events and their probabilities:

$\{X = 1\} =$ $\{HTT, THT, TTH\}$ $\rightarrow P(X=1) = 3/8$

$\{X = 3\} =$ $\{HHH\}$ $\rightarrow P(X=1) = 1/8$