San José State University

Math 161a: Applied Probability & Statistics

Lecture 4: Random variables

Prof. Guangliang Chen

Section 3.1 Random variables

Section 3.2 Probability distributions for discrete random variables

Introduction

Consider the following experiments:

- Flip a coin once;
- Flip a coin 5 times;
- Toss two dice;
- Select four numbers from 1:20, without replacement;
- Toss a coin repeatedly until a head first appears.

What are the outcomes of each experiment?

Some likely outcomes of each experiment:

- Flip a coin once; $\longrightarrow$ H, T
- Flip a coin 5 times; $\longrightarrow$ HHTTH, HHHHT
- Toss two dice; $\longrightarrow$ (3,1), (5,5), (2,6)
- Select four numbers from 1:20, without replacement; $\stackrel{\text{unordered}}{\longrightarrow}$ $\{6,9,17,2\}$, $\{20,7,12,16\}$
- Toss a coin repeatedly until a head first appears. $\longrightarrow$ H, TTH, TTTTTTH

It is often desirable to convert the outcomes to numbers in some way.

Informally, a random variable is a **numerical description** of the outcomes.

For example,

- Flip a coin once; $\longrightarrow$ $X = 1$ (H), 0 (T)
- Flip a coin 5 times; $\longrightarrow$ $X = \#$heads, $Y = \#$tails
- Toss two dice; $\longrightarrow$ $X =$ sum, $Y =$ absolute value of difference
- Select four numbers from 1:20 at random, without replacement; $\longrightarrow$ $X =$ maximum of the 4 numbers
- Toss a coin repeatedly until a head first appears. $\longrightarrow$ $X =$ total $\#$trials needed, $Y = \#$tails before the first head

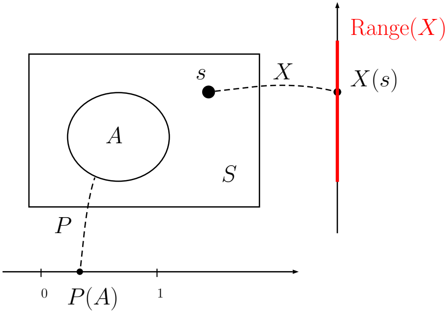
Definition of random variables

Def 0.1. A **random variable (r.v.)** associated to a sample space S is a **rule** that assigns a real number to each outcome of S:

$$X : S \mapsto \mathbb{R}.$$

The set of all possible values of X is called its range:

$$\text{Range}(X) = \{X(s) \mid s \in S\}.$$



Example 0.1. Find the range of the following random variables.

- Flip a coin once; $\longrightarrow$ $X =$ 1 (H), 0 (T)
- Flip a coin 5 times; $\longrightarrow$ $X = \#$heads
- Toss two dice; $\longrightarrow$ $X =$ sum, $Y =$ absolute value of difference
- Select four numbers from 1:20 at random, without replacement; $\longrightarrow$ $X =$ maximum of the 4 numbers
- Toss a coin repeatedly until a head first appears. $\longrightarrow$ $X =$ total $\#$trials needed, $Y = \#$tails before the first head

Answers:

- $\{0, 1\}$
- $\{0, 1, 2, 3, 4, 5\}$
- Range$(X) = \{2, 3, \ldots, 12\}$, Range$(Y) = \{0, 1, \ldots, 5\}$
- $\{4, 5, \ldots, 20\}$
- Range$(X) = \{1, 2, 3, \ldots\}$, Range$(Y) = \{0, 1, 2, \ldots\}$

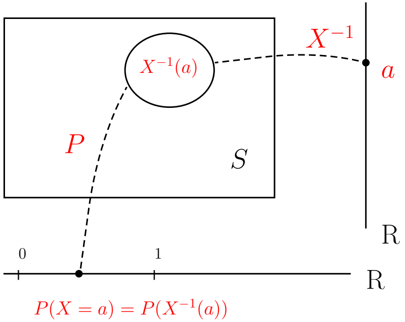
Preimages of a random variable are events

Def 0.2. Let $X : S \mapsto \mathbb{R}$ be a random variable. For any $a \in \mathbb{R}$, its **preimage** is defined as

$$X^{-1}(a) = \{s \in S \mid X(s) = a\}$$

Remark. Since $X^{-1}(a) \subseteq S$ is an event, we define

$$P(X = a) = P(X^{-1}(a)).$$



Example 0.2. Determine the following events:

- Flip a coin once; define $X = 1$ (H), 0 (T). $X^{-1}(1)$
- Toss two dice; define $X =$ sum. $X^{-1}(7)$
- Select four numbers from 1:20 at random, without replacement; define $X =$ maximum of the 4 numbers. $X^{-1}(3)$
  $X^{-1}(5)$
- Toss a coin repeatedly until a head first appears; define $X =$ total $\#$trials needed. $X^{-1}(3)$

Example 0.3. Find the following probabilities:

- Flip a fair coin once; define $X = 1$ (H), 0 (T). $P(X = 1)=$
- Toss two fair dice; define $X =$ sum. $P(X = 7)=$
- Select four numbers from 1:20 at random, without replacement; define $X =$ maximum of the 4 numbers. $P(X = 3) =$, $P(X = 5) =$
- Toss a fair coin repeatedly and independently until a head first appears; define $X =$ total #trials needed. $P(X = 3)=$

Example 0.4. Find the following probabilities:

- Toss two fair dice; define $X =$ sum. $P(X \leq 3) =$

  $P(X \geq 10)=$

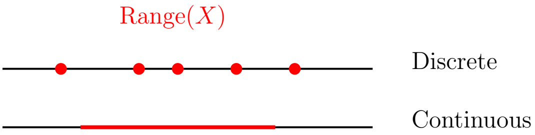
- Select four numbers from 1:20 at random, without replacement; define $X =$ maximum of the 4 numbers.

  $P(X \leq 5) =$

- Toss a fair coin repeatedly until a head first appears; define $X =$ total #trials needed. $P(X \leq 3)=$

Classification of random variables

Def 0.3. A random variable X is said to be **discrete** if it takes only a countable number of possible values, i.e., Range(X) is a finite or countably infinite set. Otherwise, it is said to be **continuous**.



Remark. Chapter 3 focuses on discrete random variables.

Example 0.5. Below are some examples of continuous random variables:

- Waiting time for your bus to come,
- Life time of electronic products
- A randomly selected SJSU student's height/weight/temperature
- Throwing a dart toward a board. Let X be the distance to the center, and Y the angle relative to the positive x-axis

Remark. Chapter 4 is about continuous random variables.

A joke

Two random variables were talking in a bar. They thought they were being discrete but I heard their chatter continuously.

Distribution of random variables

Informally speaking, the **probability distribution** of a random variable X refers to both

- the set of values it can take (range), and
- how often it takes those values (frequency).

The distribution of a discrete random variable can be fully characterized by a **probability mass function (pmf)**.

Def 0.4. Let X be a discrete random variable with range $\{x_1, x_2, \ldots\}$. The *probability mass function (pmf)* of X, denoted $f_X : \mathbb{R} \to \mathbb{R}$, is defined as

$$f_X(x) = \begin{cases} P(X = x_i), & \text{if } x = x_i, \text{ for } i = 1, 2, \ldots \\ 0, & \text{for all other } x. \end{cases}$$

For example, let X be the numerical outcome of a single toss of a fair coin (0 for tails and 1 for heads). Then its pmf is

$$f_X(x) = \begin{cases} \frac{1}{2}, & \text{if } x = 0 \\ \frac{1}{2}, & \text{if } x = 1 \\ 0, & \text{for all other } x. \end{cases}$$

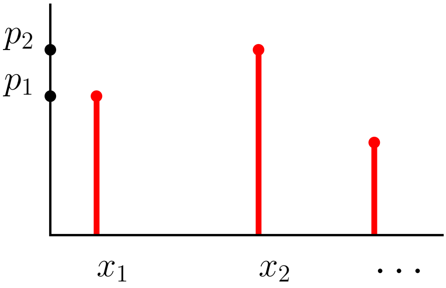
Displaying pmf

We may display the distribution of a discrete random variable using either a table or a plot consisting of spikes (line graph).

x	x_1	x_2	$\cdots$
$P(X = x)$	p_1	p_2	$\cdots$

(Notation: $p_i = f_X(x_i)$ for all i)

Important reminder:
f_X is defined everywhere on $\mathbb{R}$ (it takes the value 0 at locations not indicated in the table or plot).



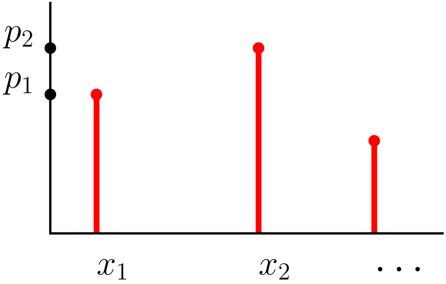
Find the pmf of X in each question below and display it in both ways.

Example 0.6 (Roll a fair die once)**.** Let X be the number obtained.

Example 0.7 (Roll a fair die twice)**.** Let X be the sum of the two numbers obtained.

Properties of a pmf f_X:

- It is nonnegative on $\mathbb{R}$: $f_X(x) \geq 0$ for all $x \in \mathbb{R}$
- It is positive (i.e., $f_X(x) > 0$) only in a countable number of locations, say $x_1, x_2, \ldots$
- The total sum is 1: $\sum_i f_X(x_i) = 1$.



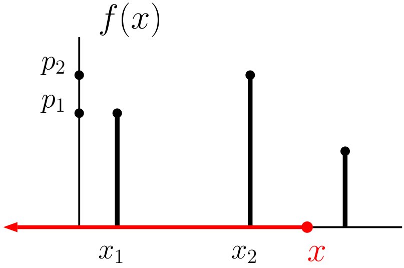
Conversely, any function satisfying all 3 conditions above is a pmf.

Cumulative distribution function (cdf)

A different way of characterizing the distribution of a random variable is through specifying all the cumulative probabilities.

Def 0.5. The *cdf* of a r.v. X, denoted $F_X : \mathbb{R} \to \mathbb{R}$, is defined by

$$F_X(x) = P(X \leq x), \quad \forall x \in \mathbb{R}.$$



$$F(x) = p_1 + p_2$$

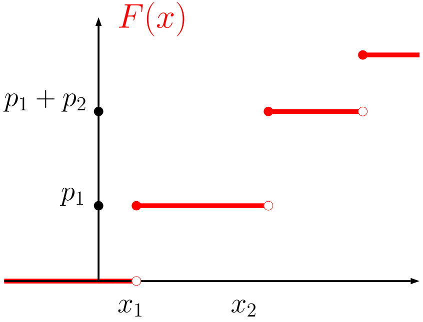
Remark. The cdf is also defined everywhere on $\mathbb{R}$.

pmf = "individual contributions";
cdf = "cumulative contributions"

The cdf can also be displayed as a table or graph.

cdf table

x	x_1	x_2	$\cdots$
$P(X \leq x)$	p_1	$p_1 + p_2$	$\cdots$



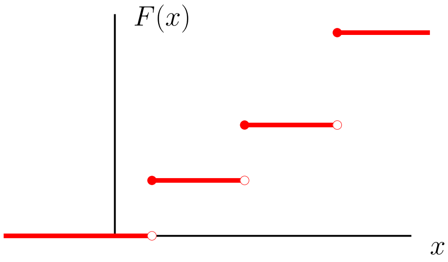
Note that the value of the cdf between two neighboring points is not zero, but determined by the left neighbor.

Example 0.8 (Roll a fair die once)**.** Let X be the number obtained. Find the cdf of X.

Properties of a cdf $F(x)$:

- $\lim_{x\to-\infty} F(x) = 0$,
  $\lim_{x\to\infty} F(x) = 1$.
- $F(x)$ is nondecreasing.
- $F(x)$ is right-continuous.

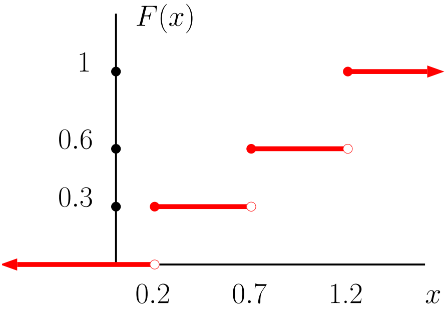
The converse is also true.

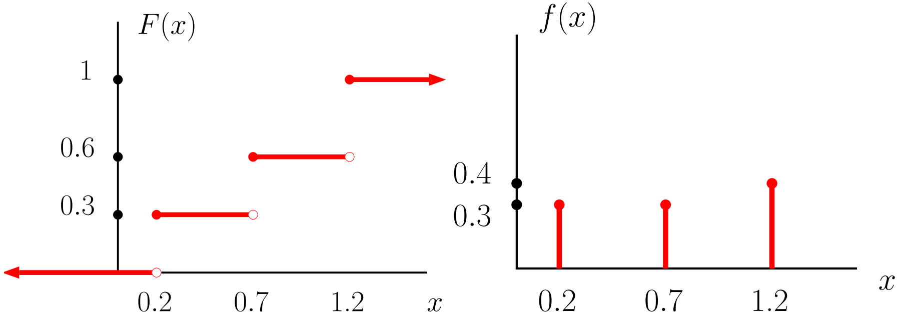


The cdf of a discrete random variable X is a step function.

In next chapter we will see that the cdf of a continuous X is a continuous curve (satisfying the three conditions above).

Example 0.9. Find the pmf corresponding to the cdf given below.




F(x)
1
0.6
0.3
0.2
0.7
1.2
x
f(x)
0.4
0.3
0.2
0.7
1.2
x

Example 0.10. For the pmf on the previous slide, find

- $P(X < 0.2), P(X \leq 0.2), P(X > 0.2), P(X \geq 0.2)$
- $P(X \leq 1), P(X < 1)$
- $P(0.2 < X \leq 1.2)$

Summary

We presented the following concepts:

- **Random variables**: $X : S \mapsto \mathbb{R}$
- **Range of** X: $\{X(s) \in \mathbb{R} \mid s \in S\}$
- Classification of X based on its range: **discrete** (countable range) or **continuous** (interval range)
- Description of distribution of X by either of the following
  - **pmf**: $f_X(x) = P(X = x)$ for any $x \in \mathbb{R}$
  - **cdf**: $F_X(x) = P(X \leq x)$ for any $x \in \mathbb{R}$

- **Tabular representation** of pmf and cdf:

x	x_1	x_2	x_3	$\cdots$
$f_X(x)$	p_1	p_2	p_3	$\cdots$
$F_X(x)$	p_1	p_1+p_2	$p_1+p_2+p_3$	$\cdots$

- **Graphical representations**