Chapter 8: Probability Mass Functions (pmf’s) and Cumulative Distribution Functions (cdf’s)

Meike Niederhausen and Nicky Wakim

2023-10-09

Table of contents

Learning Objectives

1. Calculate probabilities for discrete random variables
2. Calculate and graph a probability mass function (pmf)
3. Calculate and graph a cumulative distribution function (CDF)

What is a probability mass function?

Definition: probability distribution or probability mass function (pmf)

The **probability distribution** or **probability mass function** (**pmf**) of a discrete r.v. X is defined for every number x by

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}(\text{all } \omega \in S : X(\omega) = x)$$

Let's demonstrate this definition with our coin toss

Example 1

Suppose we toss 3 coins with probability of tails p. If X is the random variable counting the number of tails, what are the probabilities of each value of X?

Remarks on the pmf

- A pmf $p_X(x)$ must satisfy the following properties:
  - $0 \leq p_X(x) \leq 1$ for all x.
  - $\sum_{\{\text{all } x\}} p_X(x) = 1.$
- Some distributions depend on parameters
  - Each value of a parameter gives a different pmf
  - In previous example, the number of coins tossed was a parameter
    - We tossed 3 coins
    - If we tossed 4 coins, we'd get a different pmf!
  - The collection of all pmf's for different values of the parameters is called a *family* of pmf's

Binomial family of RVs

Example 2

Suppose you toss n coins, each with probability of tails p. If X is the number of tails, what is the pmf of X?

Bernoulli family of RVs

Example 3

Suppose you toss 1 coin, with probability of tails p. If X is the number of tails, what is the pmf of X?

Household size (1/5)

Example 4

The table below shows household sizes in 2019. Data are from the U.S. Census.

Size	1	2	3	4	5 or more
Percent	28%	35%	15%	13%	9%

1. What is the sample space for household sizes?
2. Define the random variable for household sizes.
3. Do the values in the table create a pmf? Why or why not?
4. Make a plot of the pmf.
5. Write the cdf as a function.
6. Graph the cdf of household sizes in 2019.

Household size (2/5)

Example 4

The table below shows household sizes in 2019. Data are from the U.S. Census.

Size	1	2	3	4	5 or more
Percent	28%	35%	15%	13%	9%

1. What is the sample space for household sizes?
2. Define the random variable for household sizes.

Household size (3/5)

Example 4

The table below shows household sizes in 2019. Data are from the U.S. Census.

Size	1	2	3	4	5 or more
Percent	28%	35%	15%	13%	9%

3. Do the values in the table create a pmf? Why or why not?
4. Make a plot of the pmf

What is a cumulative distribution function?

Definition: cumulative distribution function (CDF)

The **cumulative distribution function** (**cdf**) of a discrete r.v. X with pmf $p_X(x)$, is defined for every value x by

$$F_X(x) = \mathbb{P}(X \le x) = \sum_{\{\text{all } y:\ y \le x\}} p_X(y)$$

Household size (4/5)

Example 4

The table below shows household sizes in 2019. Data are from the U.S. Census.

Size	1	2	3	4	5 or more
Percent	28%	35%	15%	13%	9%

5. Write the cdf as a function.

Household size (5/5)

Example 4

The table below shows household sizes in 2019. Data are from the U.S. Census.

Size	**1**	**2**	**3**	**4**	**5 or more**
Percent	28%	35%	15%	13%	9%

6. Graph the cdf of household sizes in 2019.

Properties of *discrete* CDFs

- $F(x)$ is increasing or flat (never decreasing)
- $\min_x F(x) = 0$
- $\max_x F(x) = 1$
- CDF is a step function