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

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

$p_X(x)$

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?

p = prob of tails
$1-p$ = prob of heads

$S = \{HHH, HHT, \ldots, TTT\}$

$X(\omega) = x \rightarrow x = 0, 1, 2, 3$

$P(X=0) = P(HHH) = (1-p)\cdot(1-p)(1-p) = (1-p)^3$

$P(X=1) = P(HHT \text{ or } HTH \text{ or } THH)$
$= (1-p)(1-p)p + (1-p)p(1-p) + p(1-p)(1-p)$
$= 3p(1-p)^2$

$P(X=2) = P(TTH \text{ or } THT \text{ or } HTT)$
$= P(TTH) + P(THT) + P(HTT)$
$= p\cdot p(1-p) + p(1-p)p + (1-p)p\cdot p$
$= 3p^2(1-p)$

$P(X=3) = P(TTT) = p\cdot p\cdot p = p^3$

$$\boxed{P_X(x) = \binom{3}{x} p^x (1-p)^{3-x} \quad \text{for } x = 0, 1, 2, \text{ or } 3}$$

#tails

$\binom{3}{1} = 3$ $\binom{3}{2} = 3$
$\binom{3}{0} = 1$ $\binom{3}{3} = 1$

Remarks on the pmf

- A pmf $p_X(x)$ must satisfy the following properties:
  - $0 \leq p_X(x) \leq 1$ for all x.
  - $\sum_{\{\text{all } x\}} p_X(x) = 1.$ $P(X=0)+P(X=1)+P(X=2)+P(X=3)=1$
- 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?

$S = \{HHH, HHT \ldots \quad TTT\}$

$P_X(x) = \binom{3}{x} p^x (1-p)^{3-x}$

4 tosses: for $x = 0, 1, 2, 3,$ or 4

$P_X(x) = \binom{4}{x} p^x (1-p)^{4-x}$

BINOMIAL PMF: (family)

for 2 option outcomes.

defined x as # of pre-defined outcome (1 of the 2)

$$\boxed{P_{\underline{X}}(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, 2, 3, \ldots, n}$$

Bernoulli family of RVs

binomial = $\sum$ bernoulli

> **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?

$n = 1$

$$p_X(x) = \binom{1}{x} p^x (1-p)^{1-x} \quad \text{for } x = 0, 1$$

$\binom{1}{0} = 1 \quad \binom{1}{1} = 1$

$$\boxed{p_X(x) = p^x (1-p)^{1-x} \quad \text{for } x = 0, 1}$$

$p_X(0) = 1 - p$

$p_X(1) = p$

$$p_X(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \\ 0 & \text{otherwise} \end{cases}$$

all other values 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.

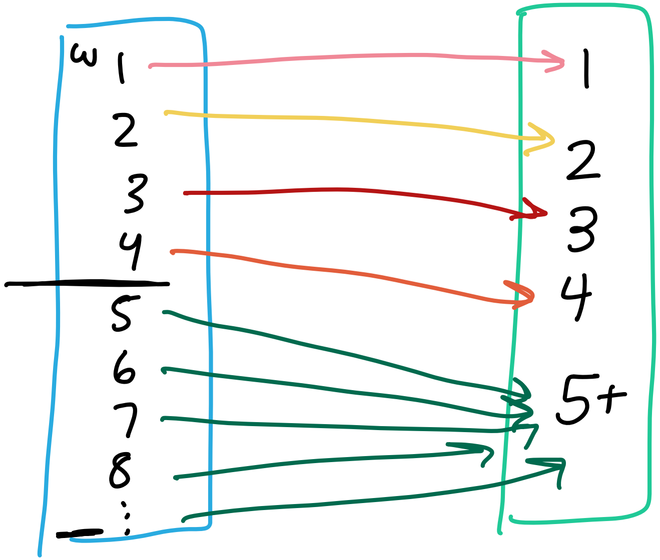
Size	1	2	3	4	5 or more
Percent	28%	35%	15%	13%	9%

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

(1) $\underline{\omega \in S}$

$S = \{1, 2, 3, 4, 5, 6, 7, 8, \underline{\ldots}\}$

(2)



$$X(\omega) = \begin{cases} \omega & \omega = 1, 2, 3, 4 \\ 5+ & \omega = 5, 6, 7, 8, \ldots \end{cases}$$

$\omega \geq 5$ include 5.5

★ What is really holding us back from saying $\omega \geq 5$

Household size (3/5)

Example 4

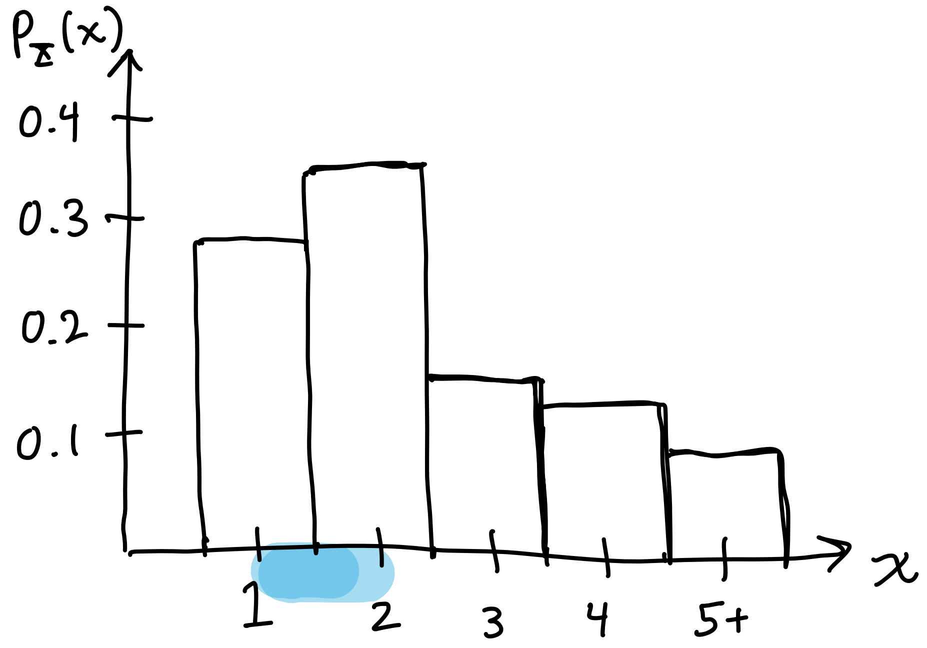
The table below shows household sizes in 2019. Data are from the U.S. Census.

Size	**1**	**2**	**3**	**4**	**5 or more**
Percent	28%	35%	15%	13%	9%
	0.28	0.35	0.15	0.13	0.09

3. Do the values in the table create a pmf? Why or why not?

4. Make a plot of the pmf

4



3

$0 \le P_X(x) \le 1$? yes!

$\sum_{\{\text{for all } x\}} P_X(x) = 1$

$0.28 + 0.35 + 0.15 + 0.13 + 0.09 = 1$

implies

$\Rightarrow$ creates viable 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

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

$$P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$$

(in coin toss ex)

☆ r code intro

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.

$F_X(1) = P(X \le 1) = P(X = 1) = 0.28$

$F_X(2) = P(X \le 2) = P(X=1) + P(X=2)$
$= 0.28 + 0.35$
$= 0.63$

$F_X(3) = P(X \le 3) = \underbrace{P(X=1) + P(X=2)}_{F_X(2)} + P(X=3)$
$= 0.63 + 0.15 = 0.78$

$F_X(x) = \sum_{\{\text{all } y,\ y \le x\}} p_X(y)$

$x \rightarrow 3$, $y = 1, 2, 3$

$F_X(4) = F_X(3) + P(X=4)$
$= 0.78 + 0.13$
$= 0.91$

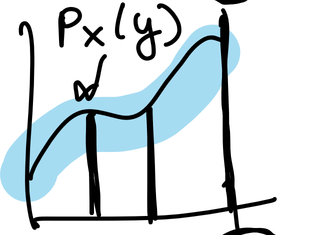
$F_X(5+) = F_X(4) + P(X=5+)$
$= 0.91 + 0.09$
$= 1$

ASIDE

$F_X(x)$

$\int_0^{x} p_X(y)\, dy$

$p_X(x)\,dx$



$x = 2$, $x = 3$

$$F_X(x) = \begin{cases} 0 & x < 1 \\ 0.28 & 1 \le x < 2 \\ 0.63 & 2 \le x < 3 \\ 0.78 & 3 \le x < 4 \\ 0.91 & 4 \le x < 5 \\ 1 & x \ge 5 \end{cases}$$

Household size (5/5)

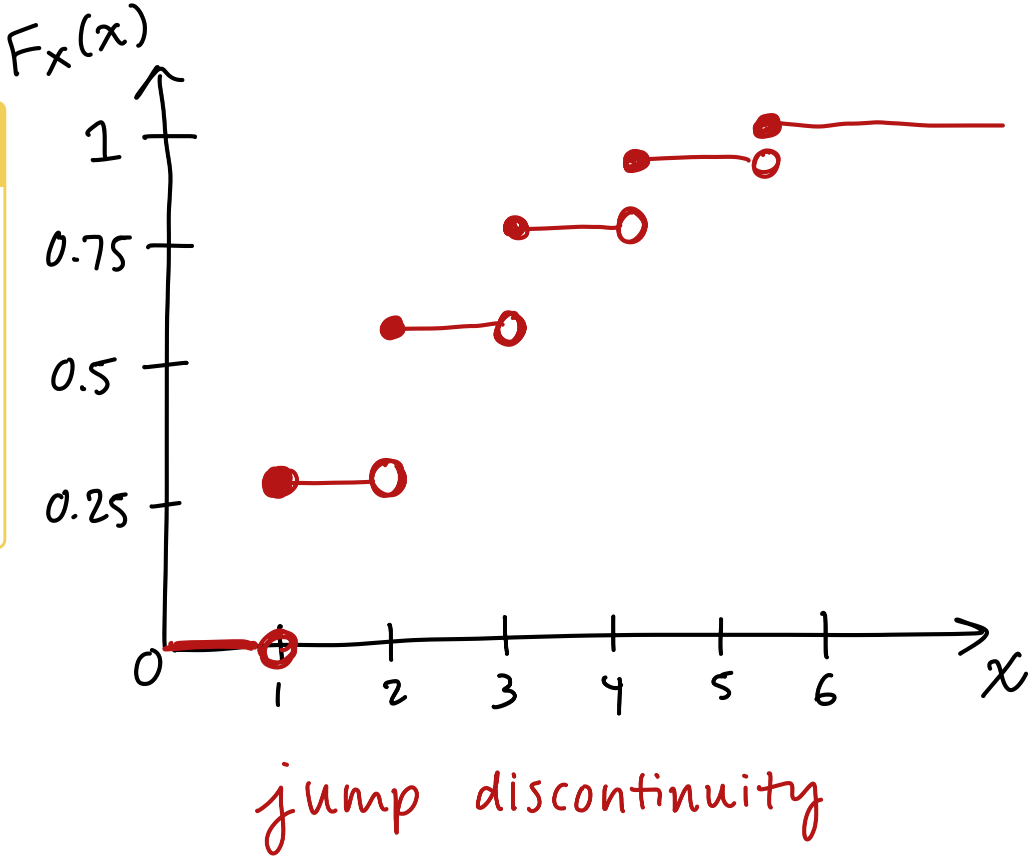
Example 4

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.

$$F_X(x) = \begin{cases} 0 & x < 1 \\ 0.28 & 1 \le x < 2 \\ 0.63 & 2 \le x < 3 \\ 0.78 & 3 \le x < 4 \\ 0.91 & 4 \le x < 5 \\ 1 & x \ge 5 \end{cases}$$



Properties of *discrete* CDFs

• jump discontinuity

- $F(x)$ is increasing or flat (never decreasing)
- $\min_{x} F(x) = 0$
- $\max_{x} F(x) = 1$
- CDF is a step function

piecewise