Conditional Independence; Random variables

CSE 312 Spring 26

Lecture 8

Conditional Independence

Definition. Two events A and B are **independent** conditioned on C if
$P(C) \neq 0$ and $P(A \cap B \mid C) = P(A \mid C) \cdot P(B \mid C)$.

- If $P(A \cap C) \neq 0$, equivalent to $P(B|A \cap C) = P(B \mid C)$
- If $P(B \cap C) \neq 0$, equivalent to $P(A|B \cap C) = P(A \mid C)$

Contrast to Plain Independence. Two events A and B are **independent** if

$$P(A \cap B) = P(A) \cdot P(B).$$

Example – Tossing Coins (1)

Suppose that Coin 1 has probability of heads 0.3
and Coin 2 has probability of head 0.9.
We choose one of the coins randomly with equal probability and flip that coin 2 times independently. What is the probability we get all heads?

C_i = coin i was selected

$P(H_1 H_2 | C_1) = P(H | C_1) P(H | C_1) = 0.3^2$

$P(H_1 H_2 | C_2) = P(H | C_2) P(H | C_2) = 0.9^2$

Tossing Coins (2)

Suppose that Coin 1 has probability of heads 0.3
and Coin 2 has probability of head 0.9.
We choose one of the coins randomly with equal probability and flip that coin 2 times independently. What is the probability we get all heads?

C_i = coin i was selected

$P(HH \cap C_1)$ $\quad$ $P(HH \cap C_2)$

$$P(HH) = P(HH \mid C_1) \cdot P(C_1) + P(HH \mid C_2) \cdot P(C_2)$$

Law of Total Probability (LTP)

$$= P(H \mid C_1)^2 P(C_1) + P(H \mid C_2)^2 P(C_2)$$

Conditional Independence

$$= 0.3^2 \cdot 0.5 + 0.9^2 \cdot 0.5 = 0.45$$

Tossing coins (3)

$P(HH) \neq P(H)^2 = 0.6^2 = 0.36$

Suppose that Coin 1 has probability of heads 0.3

and Coin 2 has probability of head 0.9.

We choose one coin randomly with equal probability and flip that coin 2 times independently. How does the probability we get all heads compare to $P(H)^2$?

$$P(H) = P(H|C_1)P(C_1) + P(H|C_2)P(C_2)$$
$$= 0.3 \cdot \frac{1}{2} + 0.9 \cdot \frac{1}{2} = 0.6$$

Example – Tossing coins (4)

Suppose that Coin 1 has probability of heads 0.3
and Coin 2 has probability of head 0.9.
We choose one coin randomly with equal probability and flip that coin 2 times independently. How does the probability we get all heads compare to $P(H)^2$?

$$P(HH \mid C_1) = P(H|C_1)^2$$

$$P(HH) \neq P(H)^2$$

$$P(HH) = P(H|C_1)^2\, P(C_1) + P(H \mid C_2)^2\, P(C_2) = 0.3^2 \cdot 0.5 + 0.9^2 \cdot 0.5 = 0.45$$

$$P(H) = P(H|C_1) \cdot P(C_1) + P(H|C_2) \cdot P(C_2) \doteq 0.6 \qquad P(HH) = 0.36$$

Pset 3 Coding – Naïve Bayes (Algorithm for Spam Detection)

Will compute probability an email is spam given the words in it

Say email contains words $\{x_1, x_2, \ldots, x_n\}$.
Want to compute probability it's spam: $\mathbb{P}(S \mid x_1, \ldots, x_n)$ ← Spam

$$\frac{\mathbb{P}(x_1, \ldots, x_n \mid S)\mathbb{P}(S)}{\mathbb{P}(x_1, \ldots, x_n)} = \frac{\mathbb{P}(x_1, \ldots, x_n \mid S)\mathbb{P}(S)}{\mathbb{P}(x_1, \ldots, x_n \mid S)\mathbb{P}(S) + \mathbb{P}(x_1, \ldots, x_n \mid H)\mathbb{P}(H)}$$

← LTP

Pset 3 Coding – Naïve Bayes (Algorithm for Spam Detection)

Will compute probability an email is spam given the words in it

Say email contains words $\{x_1, x_2, \ldots, x_n\}$.
Want to compute probability it's spam:

$$\mathbb{P}(S \mid x_1, \ldots, x_n)$$

$$\frac{\mathbb{P}(x_1, \ldots, x_n \mid S)\mathbb{P}(S)}{\mathbb{P}(x_1, \ldots, x_n)} = \frac{\mathbb{P}(x_1, \ldots, x_n \mid S)\mathbb{P}(S)}{\mathbb{P}(x_1, \ldots, x_n \mid S)\mathbb{P}(S) + \mathbb{P}(x_1, \ldots, x_n \mid H)\mathbb{P}(H)}$$

$$P(x_1, \ldots, x_n \mid S) = \frac{P(x_1 \ldots x_n \cap S)}{P(S)}$$

We will assume words in an email are conditionally independent, given email is spam or ham!

$$\begin{aligned}\mathbb{P}(x_1, \ldots, x_n, S) &= \mathbb{P}(x_1 \mid x_2, x_3, \ldots, x_n, S)\mathbb{P}(x_2 \mid x_3, \ldots, x_n, S) \ldots \mathbb{P}(x_{n-1} \mid x_n, S)\mathbb{P}(x_n \mid S)\mathbb{P}(S) \\ &\approx \mathbb{P}(x_1 \mid S)\mathbb{P}(x_2 \mid S) \ldots \mathbb{P}(x_{n-1} \mid S)\mathbb{P}(x_n \mid S)\mathbb{P}(S) \\ &= \mathbb{P}(S)\prod_{i=1}^{n}\mathbb{P}(x_i \mid S)\end{aligned}$$

Chain Rule

Coding problem will be released on Ed at 11am today.

Agenda (1)

- Random Variables
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)

Random Variables (Idea)

Often: We want to **capture quantitative properties** of the outcome of a random experiment, e.g.:

- *What is the total of two dice rolls?*
- *What is the number of coin tosses needed to see the first head?*
- *What is the number of heads among 5 coin tosses?*

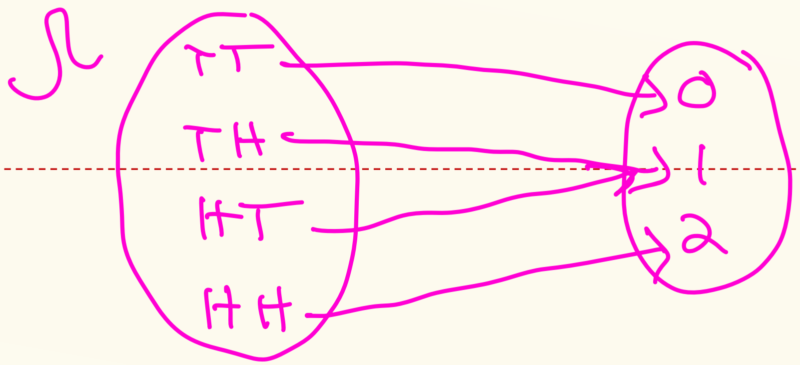
Random Variables

Definition. A **random variable (RV)** for a probability space $(\Omega, \mathbb{P})$ is a function $X: \Omega \to \mathbb{R}$.

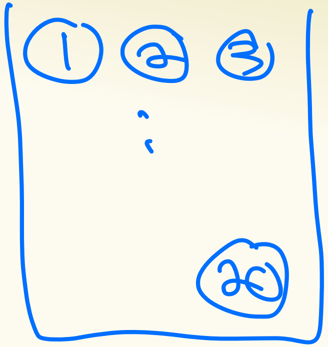
The set of values that X can take on is called its range/support Ω_X

Example. Number of heads in 2 independent coin flips $\Omega = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$

$\Omega_X = \{0, 1, 2\}$



Drawing Balls



20 balls labeled 1, 2, … , 20 in an urn

- Draw a subset of 3 uniformly at random
- What is $|\Omega|$?

$$\binom{20}{3}$$

Drawing Balls – new random variable

20 balls labeled 1, 2, … , 20 in an urn

- Draw a subset of 3 uniformly at random
- $|\Omega| = \binom{20}{3}$
- Let X = maximum of the 3 numbers on the balls
  - Example: $X(2, 7, 5) = 7$
  - Example: $X(15, 3, 8) = 15$

Drawing Balls – what is $|\Omega_X|$?

①②③④⑤ . . . . ⑱ ⑲ ⑳

20 balls labeled 1, 2, …, 20 in an urn

- Draw a subset of 3 uniformly at random
- Let X = maximum of the 3 numbers on the balls
  - Example: $X(2,7,5) = 7$
  - Example: $X(15,3,8) = 15$
- What is $|\Omega_X|$?

↑ # possible values X can take

https://pollev.com/ annakarlin185

A. 20^3
B. 20
C. 18
D. $\binom{20}{3}$

Example: Returning Homeworks

- Class with 3 students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

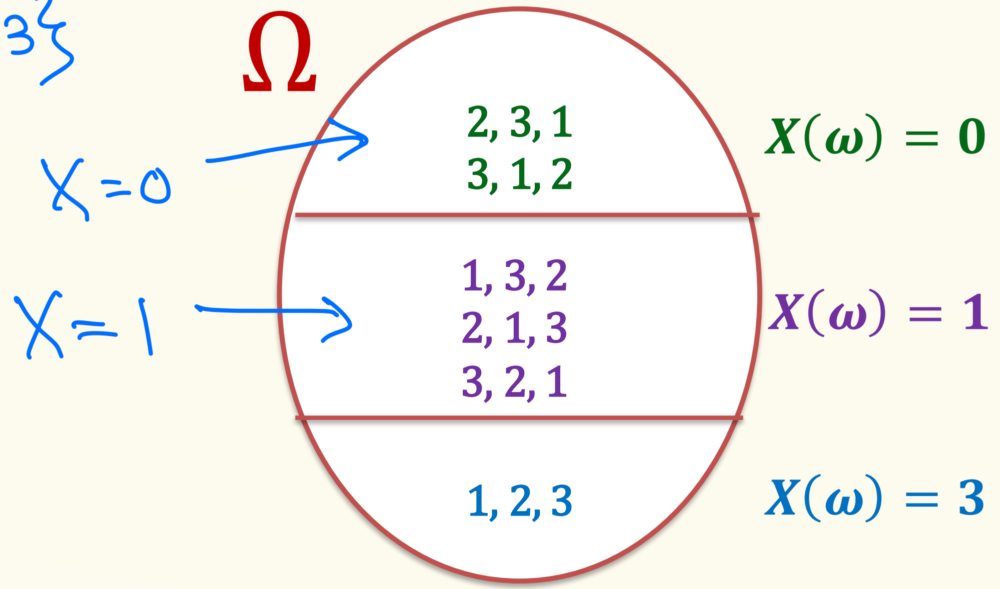
$\Pr(\omega)$	ω	$X(\omega)$
1/6	1, 2, 3	3
1/6	1, 3, 2	1
1/6	2, 1, 3	1
1/6	2, 3, 1	0
1/6	3, 1, 2	0
1/6	3, 2, 1	1

Example: Returning Homeworks (2)

- Class with 3 students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

$\Omega_X = \{0, 1, 3\}$

$\Pr(\omega)$	ω	$X(\omega)$
1/6	1, 2, 3	3
1/6	1, 3, 2	1
1/6	2, 1, 3	1
1/6	2, 3, 1	0
1/6	3, 1, 2	0
1/6	3, 2, 1	1



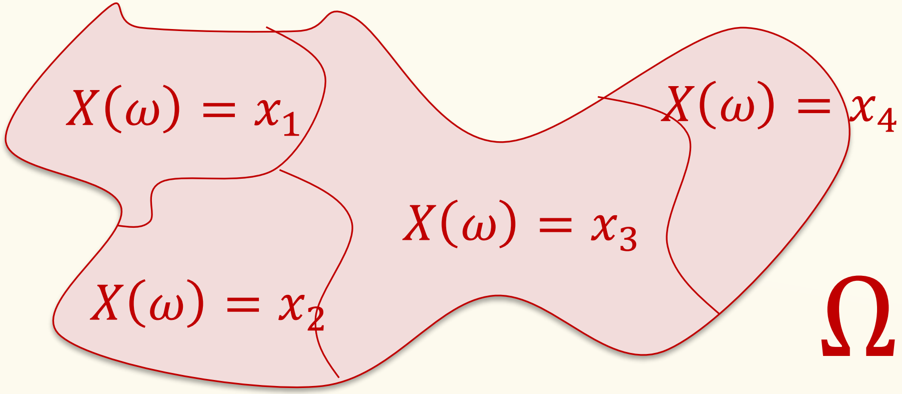
Random variables partition sample space

Definition. A **random variable (RV)** for a probability space $(\Omega, \mathbb{P})$ is a function $X: \Omega \to \mathbb{R}$.

The set of values that X can take on is called its range/support Ω_X

Random variables partition the sample space.

$\Omega_X = \{x_1, x_2, x_3, x_4\}$



Definition. For a RV $X: \Omega \to \mathbb{R}$, we define the event

$$\{X = x\} \stackrel{\text{def}}{=} \{\omega \in \Omega \mid X(\omega) = x\}$$

Agenda (2)

- Random Variables
- Probability Mass Function (pmf)
- Cumulative Distribution Function (CDF)

Probability Mass Function (PMF)

Definition. For a RV $X : \Omega \to \mathbb{R}$, we define the event

$$\{X = x\} \stackrel{\text{def}}{=} \{\omega \in \Omega \mid X(\omega) = x\}$$

The **probability mass function** (PMF) of X tells us the probabilities of these events, i.e., the probability that X takes each value in Ω_X

We use the notation

$p_X(x)$

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x\})$$

For the probability mass function

$$\sum_{x \in \Omega_X} \mathbb{P}(X = x) = 1$$

Probability Mass Function – example 1

Flipping two independent coins $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$

X = number of heads in the two flips

$X(HH) = 2$ $X(HT) = 1$ $X(TH) = 1$ $X(TT) = 0$

$\Omega_X = \{0, 1, 2\}$

What is the pmf of X?

$$p_X(x) = P(X = x) = \begin{cases} \frac{1}{4} & x = 0 \\ \frac{1}{2} & x = 1 \\ \frac{1}{4} & x = 2 \\ 0 & \text{otherwise} \end{cases}$$

Probability Mass Function

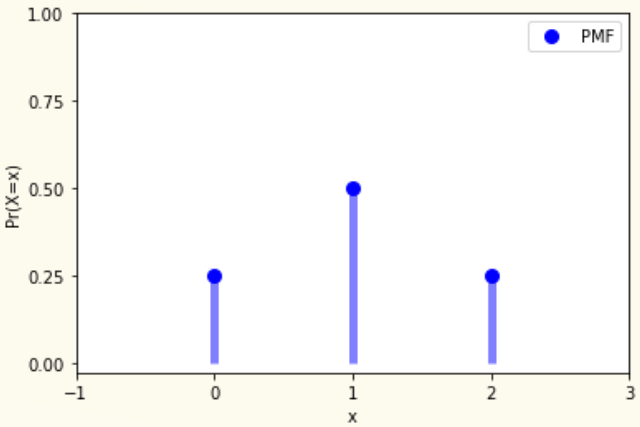
Flipping two independent coins $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$

$X = \text{number of heads in the two flips}$

$$X(HH) = 2 \qquad X(HT) = 1 \qquad X(TH) = 1 \qquad X(TT) = 0$$

$$\Omega_{\mathrm{X}} = \{0, 1, 2\}$$

$$\mathrm{p}_X(\mathrm{x}) = \Pr(X = x) = \begin{cases} \frac{1}{4}, & x = 0 \\ \frac{1}{2}, & x = 1 \\ \frac{1}{4}, & x = 2 \\ 0, & o.w. \end{cases}$$



RV Example

20 balls labeled 1, 2, …, 20 in a bin

- Draw a subset of 3 uniformly at random
- Let X = maximum of the 3 numbers on the balls

What is $p_X(20) = P(X = 20)$?

$= P(\text{largest of 3} = 20) = \frac{|E|}{\binom{20}{3}}$

Poll:

https://pollev.com/ annakarlin185

A. $\binom{20}{2} / \binom{20}{3}$

B. $\binom{19}{2} / \binom{20}{3}$

C. $19^2 / \binom{20}{3}$

D. $19 \cdot 18 / \binom{20}{3}$

Agenda (3)

- Random Variables
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)

Cumulative Distribution Function (CDF)

CDF

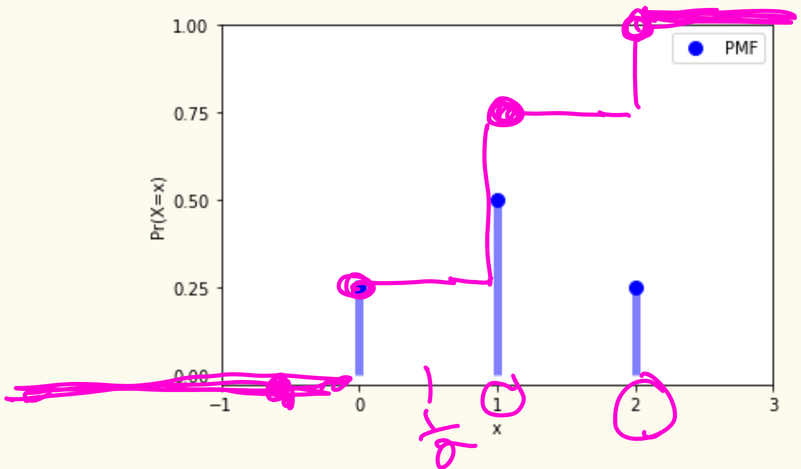
Definition. For a RV $X: \Omega \to \mathbb{R}$, the cumulative distribution function of X specifies for any real number x, the probability that $X \leq x$.

$$F_X(x) = \Pr(X \leq x)$$

pmf

Recall the **probability mass function** of X, where X is the number of heads in 2 independent coin tosses.

$$p_X(x) = \Pr(X = x) = \begin{cases} \frac{1}{4}, & x = 0 \\ \frac{1}{2}, & x = 1 \\ \frac{1}{4}, & x = 2 \\ 0, & o.w. \end{cases}$$



Example – Two fair independent coin flips

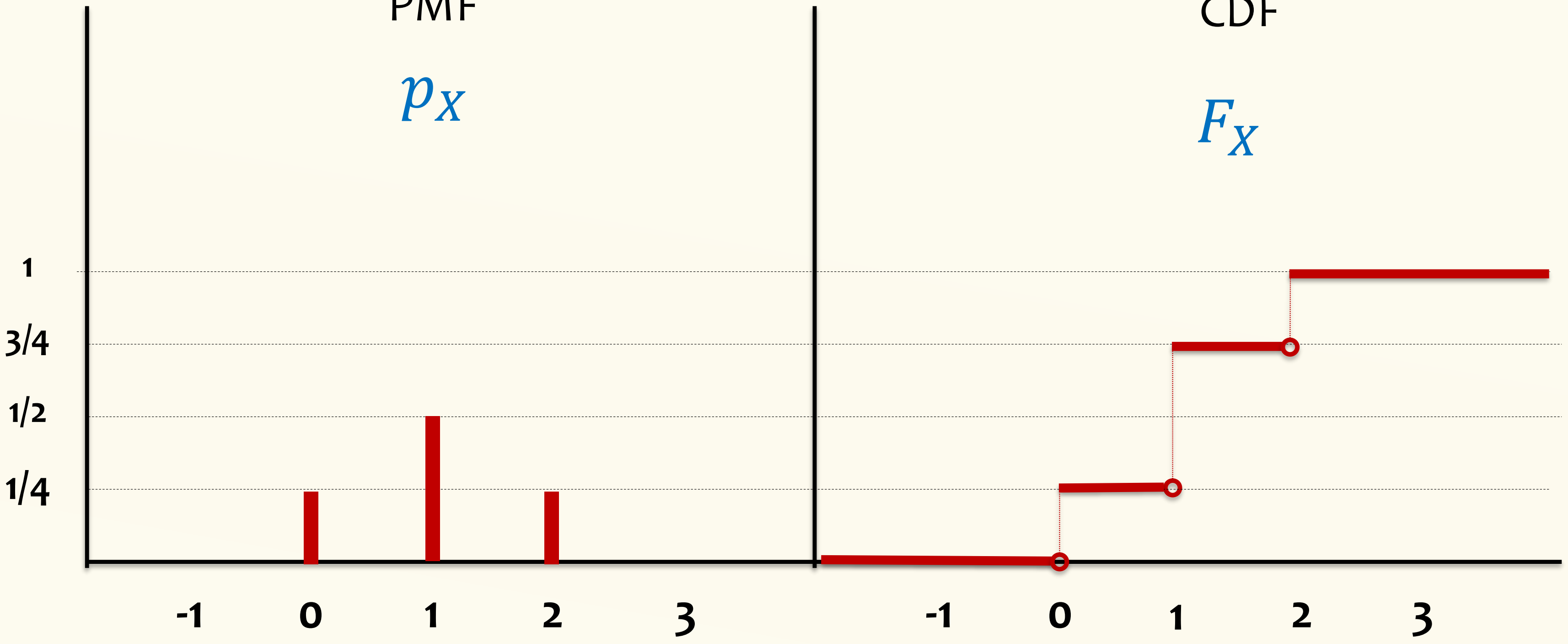
X = number of heads

Probability Mass Function
PMF

p_X

Cumulative Distribution Function
CDF

F_X



Cumulative Distribution Function (pmf and CDF for two coin tosses)

Definition. For a RV $X: \Omega \to \mathbb{R}$, the cumulative distribution function of X specifies for any real number x, the probability that $X \leq x$.

$$\mathrm{F}_X(x) = \Pr(X \leq x)$$

Go back to 2 coin clips, where X is the number of heads

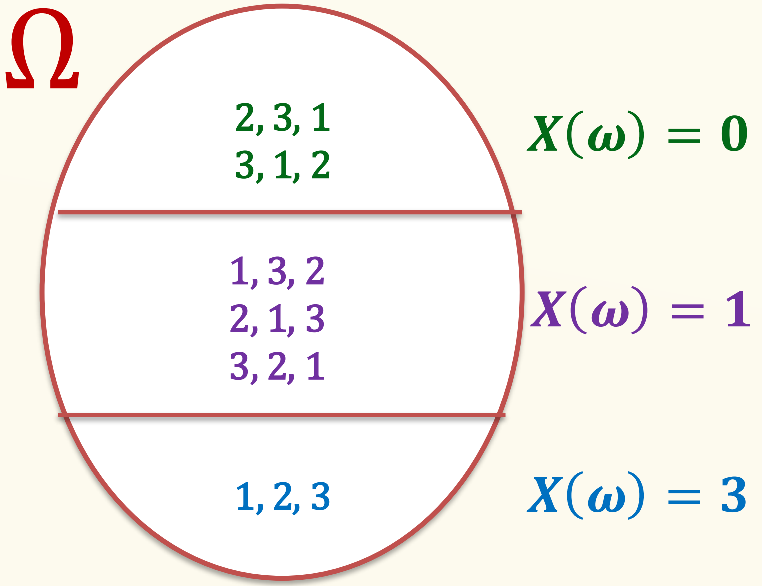
$$\mathrm{p}_X(\mathrm{x}) = \Pr(X = x) = \begin{cases} \frac{1}{4}, & x = 0 \\ \frac{1}{2}, & x = 1 \\ \frac{1}{4}, & x = 2 \\ 0, & o.w. \end{cases}$$

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & 0 \leq x < 1 \\ \frac{3}{4}, & 1 \leq x < 2 \\ 1, & 2 \leq x \end{cases}$$

Example: Returning Homeworks – Partition of sample space

- Class with 3 students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

$\Pr(\omega)$	ω	$X(\omega)$
1/6	1, 2, 3	3
1/6	1, 3, 2	1
1/6	2, 1, 3	1
1/6	2, 3, 1	0
1/6	3, 1, 2	0
1/6	3, 2, 1	1



Example: Returning Homeworks –pmf and CDF

$\Pr(\omega)$	ω	$X(\omega)$
1/6	1, 2, 3	3
1/6	1, 3, 2	1
1/6	2, 1, 3	1
1/6	2, 3, 1	0
1/6	3, 1, 2	0
1/6	3, 2, 1	1

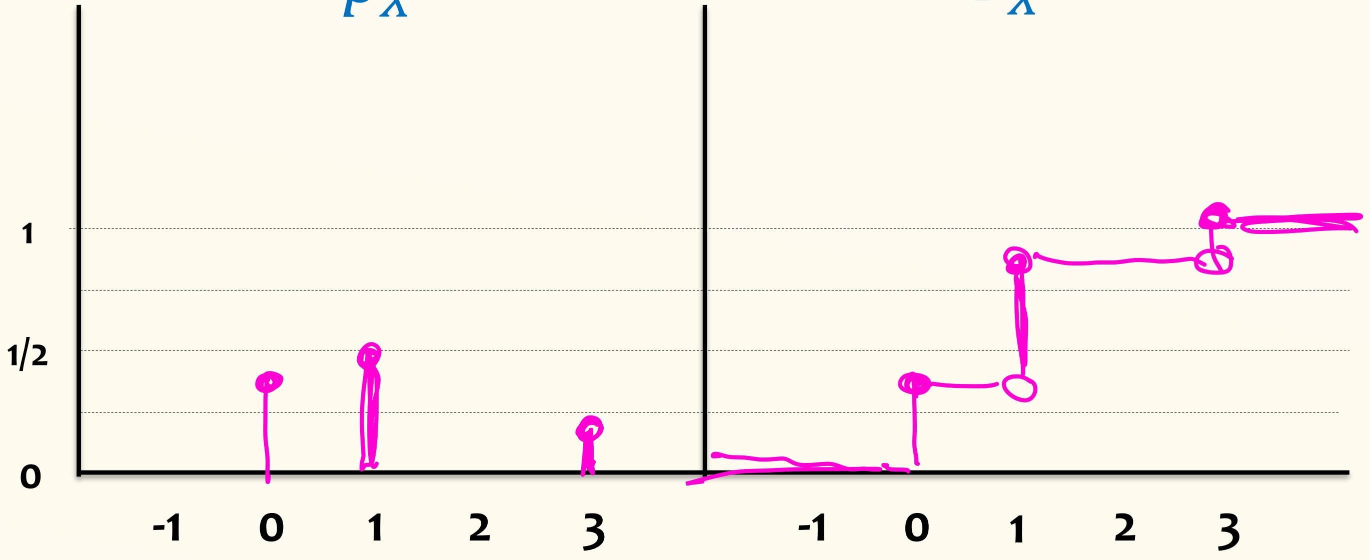
- Class with 3 students, randomly hand back homeworks.

All permutations equally likely.

- **Let X be the number of students who get their own HW**

Probability Mass Fn
PMF p_X

Cumulative Distribution Fn
CDF F_X



$$p_X(0) = P(X = 0) = 1/3$$

$$p_X(1) = P(X = 1) = 1/2$$

$$p_X(2) = P(X = 3) = 1/6$$

Example – Number of Heads

We flip n coins, independently, each heads with probability p

$\Omega = \{\mathrm{HH\cdots HH, HH\cdots HT, HH\cdots TH}, \ldots, \mathrm{TT\cdots TT}\}$ $\qquad |\Omega| = 2^n$

$X =$ # of heads $\qquad \Omega_X = \{0, 1, 2, \ldots, n\}$

$p_X(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

$P(\text{\# H's is } k)$

TH

$P(\mathrm{HHH \cdots H\,TTT \cdots T}) = p^k (1-p)^{n-k}$

Example – Number of Heads - pmdf

We flip n coins, independently, each heads with probability p

$$\Omega = \{\mathrm{HH}\cdots\mathrm{HH}, \mathrm{HH}\cdots\mathrm{HT}, \mathrm{HH}\cdots\mathrm{TH}, \dots, \mathrm{TT}\cdots\mathrm{TT}\}$$

X = # of heads

$$p_X(k) = P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

# of sequences with k heads

Prob of sequence w/ k heads