

Introduction to Data Science With Probability and Statistics

Lecture 8 & 9: Random Variables

CSCI 3022 - Summer 2020
Sourav Chakraborty
Dept. of Computer Science
University of Colorado Boulder

What will we learn today?

- Refresher & Random Variable & PMF, CDF

- Counting – Permutations & Combinations
 - Bernoulli Distribution
 - Binomial Distribution
 - Uniform Distribution
- A Modern Introduction to Probability and Statistics, Chapter 4*
- { lec 9}



Refresher

Example: Suppose that 1% of men over the age of 40 have prostate cancer. Also suppose that a test for prostate cancer exists with the following properties:

- 90% of people who have cancer will test positive
- 8% of people who do not have cancer will also test positive

What is the probability that a person who tests positive for cancer actually has cancer?

Refresher

Example: Suppose that 1% of men over the age of 40 have prostate cancer. Also suppose that a test for prostate cancer exists with the following properties:

- 90% of people who have cancer will test positive
- 8% of people who do not have cancer will also test positive

What is the probability that a person who tests positive for cancer actually has cancer?

Let;

C : {event of having cancer}

C^c : {event of not having cancer}

P : {event of testing positive}

P^c : {event of testing negative}

Refresher

Given; Prior : $P(c) = 0.01$; $P(P|c) = 0.90$; $P(P|c^c) = 0.08$

Evidence : positive test.

Posterior / updated belief : $P(c|P) = \frac{P(c \cap P)}{P(P)}$

$$\begin{aligned}\therefore P(c|P) &= \frac{P(c)P(P|c)}{P(c)P(P|c) + P(c^c)P(P|c^c)} \\ &= \frac{(0.01)(0.90)}{(0.01)(0.90) + (0.99)(0.08)} \approx 0.102 \approx \boxed{10\%}\end{aligned}$$

Random Variables

A **discrete random variable** is a function that maps elements of the sample space Ω to a finite number of values a_1, a_2, \dots, a_n or an infinite number of values a_1, a_2, \dots

Examples:

- Sum of the dice → Discrete R.V which can take values $\{2, 3, 4, \dots, 12\}$
- Difference of the dice → Discrete R.V with values $\{0, 1, 2, \dots, 5\}$
- Maximum of the dice → Discrete R.V with values $\{1, 2, 3, \dots, 6\}$
- Number of coin flips until we get a Heads Discrete R.V with values $\{0, 1, 2, \dots, \infty\}$
- Number of Heads in n flips ↴

Discrete R.V with values $\{1, 2, 3, \dots, n\}$

Probability Mass Function

A **probability mass function** (pmf) is the map between the random variable's values and the probabilities of those values. $f(a) = P(X = a)$

Each of the random variables' values has some probability mass (or weight) associated with it.

Because the pmf is a probability function, the sum of all the masses must be 1!

For e.g : A R.V 'X' is the result
of a die roll

$\therefore X$ can take $\{1, 2, \dots, 6\}$

$$\sum_{i=1}^n f(a_i) = 1$$

PMF; $P(X=1) = P(X=2) = \dots = P(X=6) = 1/6$

Probability Mass Function

Example: What is the pmf for the number of coin flips until a biased coin, with $P(H) = p$, comes up heads?

Probability Mass Function

Example: What is the pmf for the number of coin flips until a biased coin, with $P(H) = p$, comes up heads?

Let $X = \# \text{ flips till you get a Heads. } \Omega = \{H, TH, TTH, \dots\}$

$$X = \{1, 2, 3, \dots\}$$

Coin flips are independent

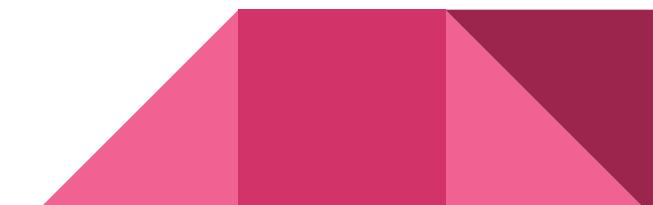
So; $P(H) = p$

$$P(TH) = P(T)P(H)$$

$$P(TTH) = P(T)P(T)P(H)$$

:

PMF	
x	$P(X=x)$
1	p
2	$(1-p)p$
3	$(1-p)^2p$
:	:



Cumulative Distribution Function

A **cumulative distribution function** (cdf) is a function whose value at a point a is the cumulative sum of probability masses up until that point a : $F(a) = P(X \leq a)$

Example: If I roll a single fair die, what is the cdf?

Question: What is the relationship between the pmf and the cdf?

Cumulative Distribution Function

A **cumulative distribution function** (cdf) is a function whose value at a point a is the cumulative sum of probability masses up until that point a : $F(a) = P(X \leq a)$

Example: If I roll a single fair die, what is the cdf? $X = \{1, 2, 3, 4, 5, 6\}$

- $P(X \leq 1) = P(X=1) = 1/6 \quad F(1)$
- $P(X \leq 2) = P(X=1) + P(X=2) = 2/6 \quad F(2)$
- :
 $P(X \leq 6) = P(X=1) + \dots + P(X=6) = 1 \quad F(6)$

Question: What is the relationship between the pmf and the cdf?

$$F(a) = \sum_{x \leq a} P(X=x)$$

Cumulative Distribution Function

Example: What is the probability that I roll two dice and they add up to at least 9?

Cumulative Distribution Function

Example: What is the probability that I roll two dice and they add up to at least 9?

Let X : sum of dice results.

$$P(X \geq 9) = P(X=9) + P(X=10) + P(X=11) + P(X=12) = \frac{10}{36}$$

$\frac{4}{36} \quad \frac{3}{36} \quad \frac{2}{36} \quad \frac{1}{36}$

OR;

$$\begin{aligned} P(X \geq 9) &= 1 - P(X < 9) \\ &= 1 - P(X \leq 8) \\ &= 1 - F(8) \\ &= 1 - \frac{26}{36} = \boxed{\frac{10}{36}} \end{aligned}$$

R.V Problem

Example: Suppose that you roll two fair, six-sided dice. Let X be a random variable representing the maximum of the two dice.

- What are the possible values that X can take?
- Which elements of the sample space map to which values of X ?
- What is the pmf of the random variable X ?

R.V Problem

Example: Suppose that you roll two fair, six-sided dice. Let X be a random variable representing the maximum of the two dice.

- What are the possible values that X can take? $X = \{1, 2, 3, 4, 5, 6\}$
- Which elements of the sample space map to which values of X ? $\Omega = \{(1,1), (1,2), \dots, (6,6)\}$
- What is the pmf of the random variable X ?

$$P(X=1) = 1/36; P(X=2) = 3/36, P(X=3) = 5/36 \dots$$

$\uparrow \quad \uparrow \quad \uparrow$
 $(1,1) \quad (2,1) (2,2) (1,1) \quad (1,3) (3,1) (2,3) (3,2) (3,3)$

x	1	2	3	4	5	6
$p(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

R.V Problem

Example (continued): Suppose that you roll two fair, six-sided dice. Let X be a random variable representing the maximum of the two dice.

- What is the probability that X is an even number?
- What is the probability that X is 3 or smaller?
- What is the complete cdf of X ?

R.V Problem

Example (continued): Suppose that you roll two fair, six-sided dice. Let X be a random variable representing the maximum of the two dice.

- What is the probability that X is an even number?
- What is the probability that X is 3 or smaller?
- What is the complete cdf of X ?



$$\begin{aligned} P(X=2) + P(X=4) + P(X=6) \\ \text{from table} \\ = 21/36 \end{aligned}$$

$$F(3) = \frac{9}{36}$$

x	1	2	3	4	5	6
$p(x)$	$1/36$	$3/36$	$5/36$	$7/36$	$9/36$	$11/36$
$F(x)$	$1/36$	$4/36$	$9/36$	$16/36$	$25/36$	$36/36$

Permutations

How many ways are there to order a set of 1 object?

How many ways are there to order a set of 2 objects?

How many ways are there to order a set of 3 objects?

What is a formula for the number of ways to order n objects?

Permutations

How many ways are there to order a set of 1 object? 1 way

How many ways are there to order a set of 2 objects? $A B, B A = 2! = \underline{\underline{2}}$ ways

How many ways are there to order a set of 3 objects? $3! = 1 \times 2 \times 3 = \underline{\underline{6}}$ ways

↓
 $A B C, A C B, B C A, B A C, C A B, C B A$

What is a formula for the number of ways to order n objects? $n!$

Permutations

- Counting **permutations** means counting the number of ways that a set of objects can be ordered.

Example: Parker, Nocona, Tatum, and Madeline line up at the ice cream truck. How many different orders could they stand in?

Permutations

- Counting **permutations** means counting the number of ways that a set of objects can be ordered.

Example: Parker, Nocona, Tatum, and Madeline line up at the ice cream truck. How many different orders could they stand in?

There are 4 persons; Ans = $4! = 1 \times 2 \times 3 \times 4 = 24$

Permutations

What is the general formula for r -permutations of n objects?

Example: How many 3-character strings can we make if each character is a distinct letter from the English alphabet?

Permutations

What is the general formula for r -permutations of n objects? $P(n, r) = \frac{n!}{(n-r)!}$

Example: How many 3-character strings can we make if each character is a distinct letter from the English alphabet?

$$\underline{26}, \underline{25}, \underline{24} \quad \therefore 26 \times 25 \times 24 \text{ ways}$$

$$P(26, 3) = \frac{26 \times 25 \times 24 \times 23 \times 22 \times \dots \times 1}{23 \times 22 \times \dots \times 1}$$
$$= \frac{26!}{(26-3)!}$$

Combinations

- Counting **combinations** means counting the number of ways that a set of objects can be combined into subsets.

Example: How many 3-character combinations can we make if each character is a distinct letter from the English alphabet?



don't care about order.

So, CAB, ABC, BAC are indistinguishable.

Permutations $\rightarrow \frac{26!}{23!}$ for combination divide it by all permutation size of 3 = 3!

$$= \frac{26!}{23! \times 3!}$$

Combinations

- Counting **combinations** means counting the number of ways that a set of objects can be combined into subsets.

Example: How many 3-character combinations can we make if each character is a distinct letter from the English alphabet?

General :

$$c(n,r) = \frac{n!}{r!(n-r)!} = \frac{P(n,r)}{r!}$$

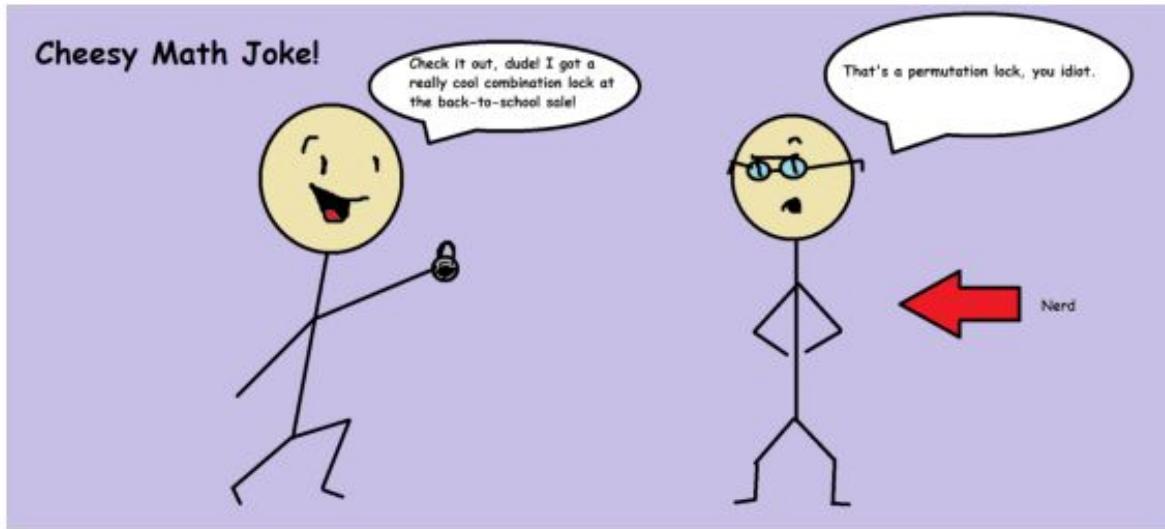
Combinations & Permutations

Combinations: Order does not matter.

$$C(n, r) = \frac{P(n, r)}{r!} = \frac{n!}{(n - r)! r!}$$

Permutations: Order does matter.

$$P(n, r) = \frac{n!}{(n - r)!}$$



Combinations

Example: If there are 10 problems on an exam, and you need at least 7 correct to pass, how many different ways are there to pass?

Combinations

Example: If there are 10 problems on an exam, and you need at least 7 correct to pass, how many different ways are there to pass?

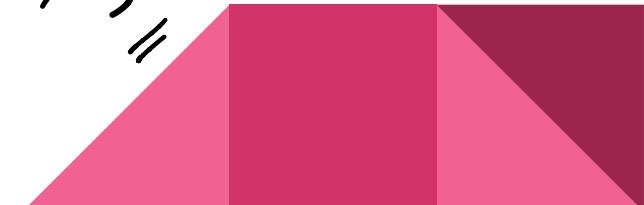
To pass, no. of correct answers need to be 7, 8, 9 or 10

But there are multiple ways to select 7 correct problems. = $C(10, 7)$

$$\underline{\underline{C}}, \underline{\underline{C}}, \underline{\underline{C}}, \underline{\underline{C}}, \underline{\underline{C}}, \underline{\underline{C}}, \underline{\underline{C}}, \underline{\underline{-}}, \underline{\underline{-}}, \underline{\underline{-}}$$

Final ans;

$$C(10, 7) + C(10, 8) + C(10, 9) + C(10, 10) =$$



Combinations

Example: A coin is flipped 10 times. How many possible outcomes have exactly 2 Heads?

Example: A coin is flipped 10 times. How many possible outcomes have 2 Heads or fewer?

Combinations

Example: A coin is flipped 10 times. How many possible outcomes have exactly 2 Heads?

T, T, T, H, T, H, T, T, T, T

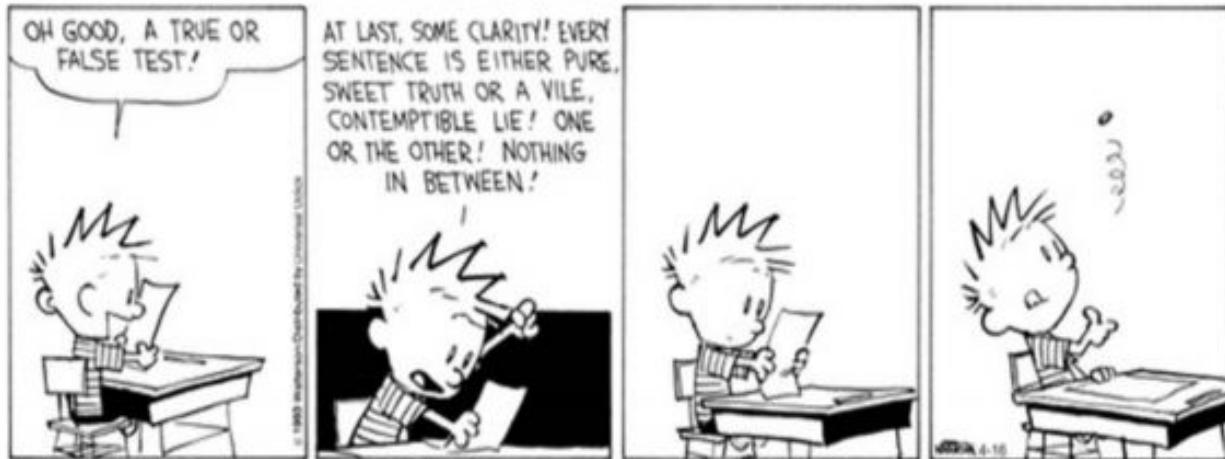
$$C(10, 2) / \binom{10}{2} = ^{10}C_2$$

Example: A coin is flipped 10 times. How many possible outcomes have 2 Heads or fewer?

$$^{10}C_0 + ^{10}C_1 + ^{10}C_2$$

The Bernoulli Distribution

- ❖ The Bernoulli distribution is used to model experiments with only two possible outcomes; often referred to as “success” and “failure”, and encoded as 1 and 0, respectively.



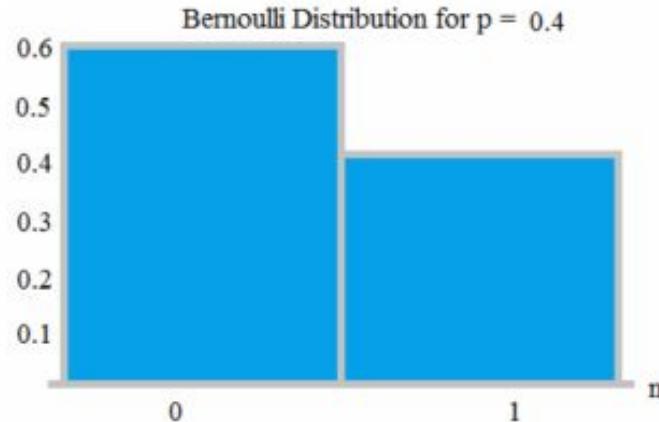
The Bernoulli Distribution

A discrete random variable X has a **Bernoulli distribution** with parameter p , where $0 \leq p \leq 1$, if its probability mass function is given by

$$f(1) = p_X(1) = P(X = 1) = p \quad \text{and} \quad p_X(0) = P(X = 0) = 1 - p$$

We denote this distribution by $\text{Ber}(p)$.

Example:



The Bernoulli Distribution

A discrete random variable X has a **Bernoulli distribution** with parameter p , where $0 \leq p \leq 1$, if its probability mass function is given by

$$f(1) = p_X(1) = P(X = 1) = p \quad \text{and} \quad p_X(0) = P(X = 0) = 1 - p$$

We denote this distribution by $\text{Ber}(p)$.

Describing the pmf with a single equation:

$$p_X(1) = p$$

$$p_X(0) = 1 - p$$

$$p_X(x) = ? \quad p^x (1-p)^{1-x}$$

Sum of Bernoulli Random Variables

Example: Suppose that you show up to a quiz completely unprepared. The quiz has 5 questions, each with 4 multiple choice options. You decide to guess each answer in a completely random way. What is the probability that you get 3 questions correct?

Sum of Bernoulli Random Variables

Example: Suppose that you show up to a quiz completely unprepared. The quiz has 5 questions, each with 4 multiple choice options. You decide to guess each answer in a completely random way. What is the probability that you get 3 questions correct?

prob. of success for a particular question : 0.25
" " failure " " " " " : 0.75 $\{ \because 1 - 0.25 \}$

$$c(5,3) \times (0.25)^3 \times (0.75)^2$$

$\frac{C}{C}, \frac{C}{C}, \frac{C}{C}, \frac{C}{C}, \frac{C}{C}$ etc.

Sum of Bernoulli Random Variables

Example (continued): What is the probability that you get 0 questions correct?

What is the probability that you get 1 problem correct?

Sum of Bernoulli Random Variables

Example (continued): What is the probability that you get 0 questions correct?

$$c(5,0) \cdot (0.25)^0 \cdot (0.75)^5$$

What is the probability that you get 1 problem correct?

$$c(5,1) \cdot (0.25)^1 \cdot (0.75)^4$$

Sum of Bernoulli Random Variables

Example (continued): What is the probability that you get 3 questions correct?

Sum of Bernoulli Random Variables

Example (continued): What is the probability that you get 3 questions correct?

Sum of Bernoulli Random Variables → A Binomial Distribution!

Example : What is the probability that you get k questions correct out of n problems total?

Sum of Bernoulli Random Variables → A Binomial Distribution!

Example : What is the probability that you get k questions correct out of n problems total?

$$c(n, k) \cdot p^k (1-p)^{n-k}$$

p : probability of success .

Binomial Distribution

A discrete random variable X has a **Binomial Distribution** with parameters n and p , where $n = 1, 2, \dots$ and $0 \leq p \leq 1$, if its probability mass function is given by

$$p_X(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for } k = 0, 1, 2, \dots, n$$

We denote this distribution by $\text{Bin}(n, p)$.

So, last problem can be modelled as; $Q_i = \begin{cases} 1, & \text{if correct} \\ 0, & \text{if incorrect} \end{cases}$
every question $Q_i \sim \text{Ber}(0.25)$; $\therefore Q_i = \begin{cases} 1, & \text{if correct} \\ 0, & \text{if incorrect} \end{cases}$

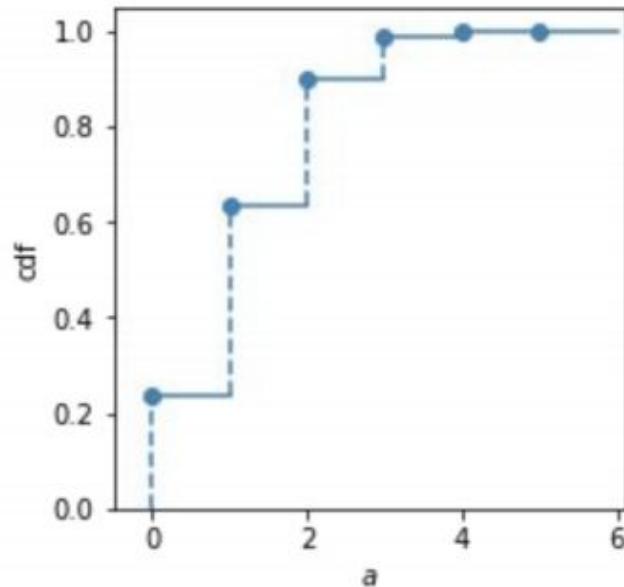
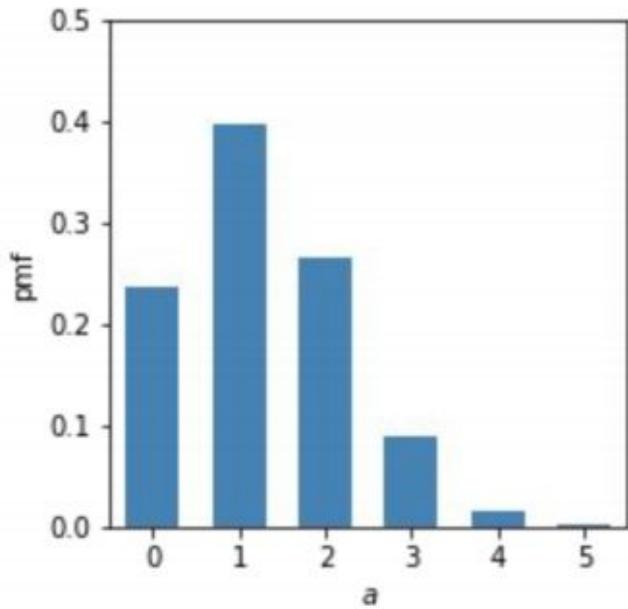
$$\therefore X : \# \text{ correct questions} = Q_1 + Q_2 + Q_3 + Q_4 + Q_5$$

$$= \sum_{i=1}^5 Q_i$$

$$\therefore X \sim \text{Bin}(5, 0.25)$$

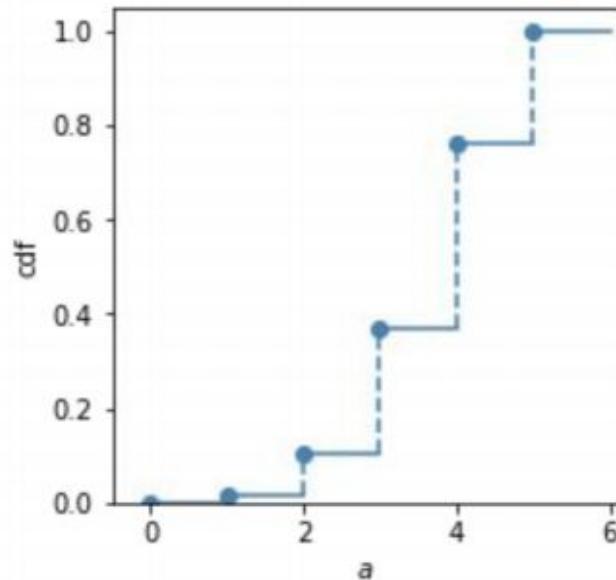
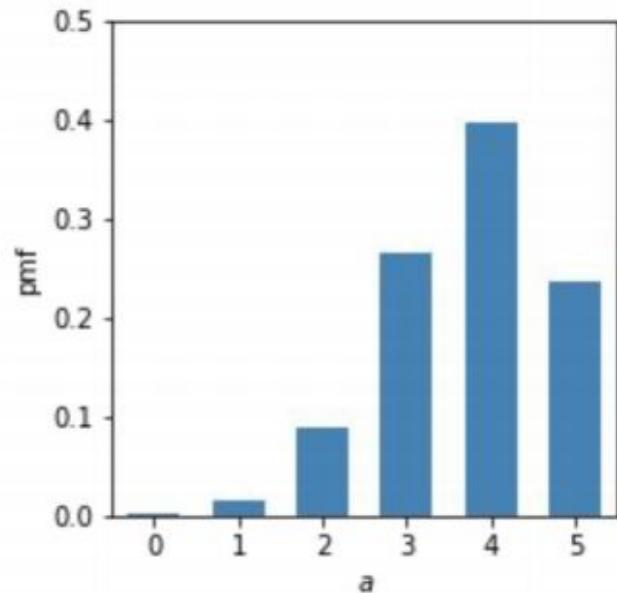
Binomial Distribution

Example: $n=5, p=0.25$



Binomial Distribution

Example: $n=5, p=0.75$



Discrete Uniform Distribution

A discrete random variable X has a **Discrete Uniform Distribution** with parameters a and b , where $n = b - a + 1$ if:

$$p_X(k) = \frac{1}{n} \quad \text{for } k = a, a + 1, a + 2, \dots, b$$

Example : What is the distribution of a fair die?

Discrete Uniform Distribution

A discrete random variable X has a **Discrete Uniform Distribution** with parameters a and b , where $n = b - a + 1$ if:

$$p_X(k) = \frac{1}{n} \quad \text{for } k = a, a+1, a+2, \dots, b$$

Example : What is the distribution of a fair die?

Let X : Result of a die roll : $X = 1, 2, 3, 4, 5, 6$

pMF : $P(X=1) = P(X=2) = \dots = P(X=6) = \frac{1}{6}$

$$X \sim U(1, 6)$$

