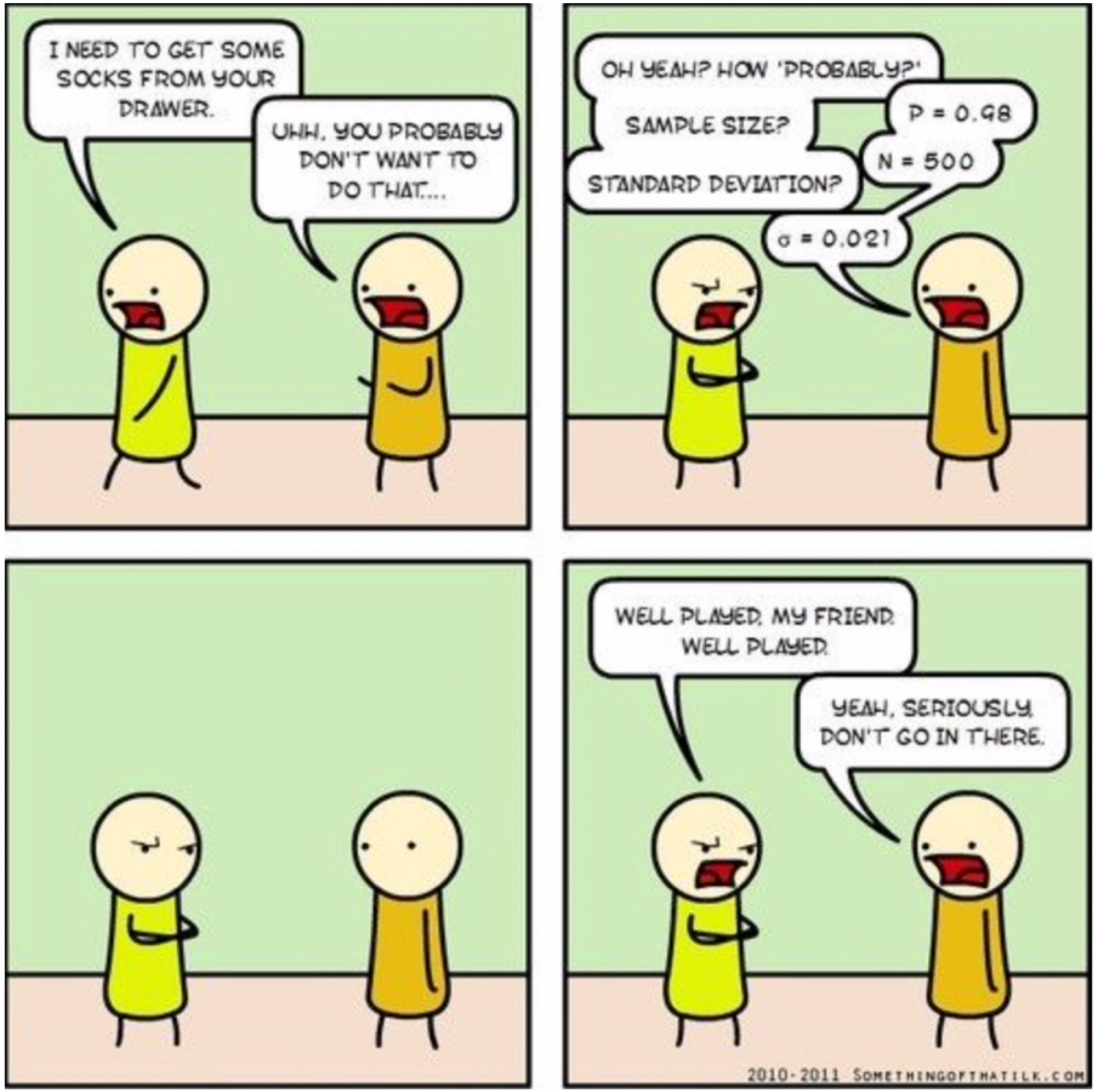
CSCI 3022: Intro to Data Science
Lecture 7: Discrete Random Variables and their Distributions

Rachel Cox
Department of Computer Science



Announcements & Reminders

- ☐ quizlet 02 - Due tonight at 11:59pm
- ☐ HW 2 - posted
  - Due next Friday, Feb. 14^{th}

What will we learn today?

- ❑ Counting – Permutations & Combinations
- ❑ Bernoulli Distribution
- ❑ Binomial Distribution
- ❑ Uniform Distribution

- ❑ *A Modern Introduction to Probability and Statistics, Chapter 4*

Review from Last Time

A **discrete random variable** (r.v.) X is a function that maps the elements of the sample space Ω to a finite number of values $a_1, a_2, \ldots, a_n$ or an infinite number of values $a_1, a_2, \ldots$

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

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

Warm–Up 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? $\Omega = \{(1,1), (1,2) \ldots (6,1), (6,2) \ldots\}$
- What is the pmf of the random variable X?

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

$(2,3) \to 3$

$(6,2) \to 6$

· let a be the values that X can take on

This table defines our pmf function

a	1	2	3	4	5	6
$f(a)$	$1/36$	$3/36$	$5/36$	$7/36$	$9/36$	$11/36$

$$\sum_a f(a) = 1$$

Warm–Up 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? $\rightarrow P(X=2 \text{ or } X=4 \text{ or } X=6)$
- What is the probability that X is 3 or smaller?
- What is the complete cdf of X?

$$= P(X=2) + P(X=4) + P(X=6)$$
$$= f(2) + f(4) + f(6)$$
$$= \frac{3}{36} + \frac{7}{36} + \frac{11}{36}$$
$$\boxed{= \frac{21}{36}}$$

$$P(X \leq 3) = P(X=1) + P(X=2) + P(X=3)$$
$$= f(1) + f(2) + f(3)$$
$$= F(3)$$
$$= \frac{1}{36} + \frac{3}{36} + \frac{5}{36}$$
$$= 9/36$$

a	1	2	3	4	5	6
f(a)	1/36	3/36	5/36	7/36	9/36	11/36
F(a)	1/36	4/36	9/36	16/36	25/36	36/36

Permutations

in general: for n items $\boxed{n!}$

How many ways are there to order a set of 1 object? $\boxed{1}$

How many ways are there to order a set of 2 objects? AB, BA $\boxed{2 \text{ ways}}$

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

_AB, A_B, AB_

$3 \cdot 2 = \boxed{6 \text{ ways}}$

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

$P(n-1)$

$P(n) = n\,P(n-1)$ $\boxed{= n!}$

Permutations

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

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

$4 \cdot 3 \cdot 2 \cdot 1 \boxed{= 24 \text{ ways}} \quad (4!)$

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?

$$\underline{26} \cdot \underline{25} \cdot \underline{24}$$

$$P(26,3) = \frac{26 \cdot 25 \cdot 24 \cdot \boxed{23 \cdots 3 \cdot 2 \cdot 1}}{\boxed{23 \cdot 22 \cdot 21 \cdots 3 \cdot 2 \cdot 1}}$$

$$= 26 \cdot 25 \cdot 24$$

$$= \frac{26!}{(26-3)!}$$

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

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?

we don't care about order

ABC CAB are indistinguishable.

permutations: $\frac{26!}{23!}$ combinations $= \frac{26!}{23!\,3!}$

combinations $= \frac{\text{permutations}}{r!}$

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

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

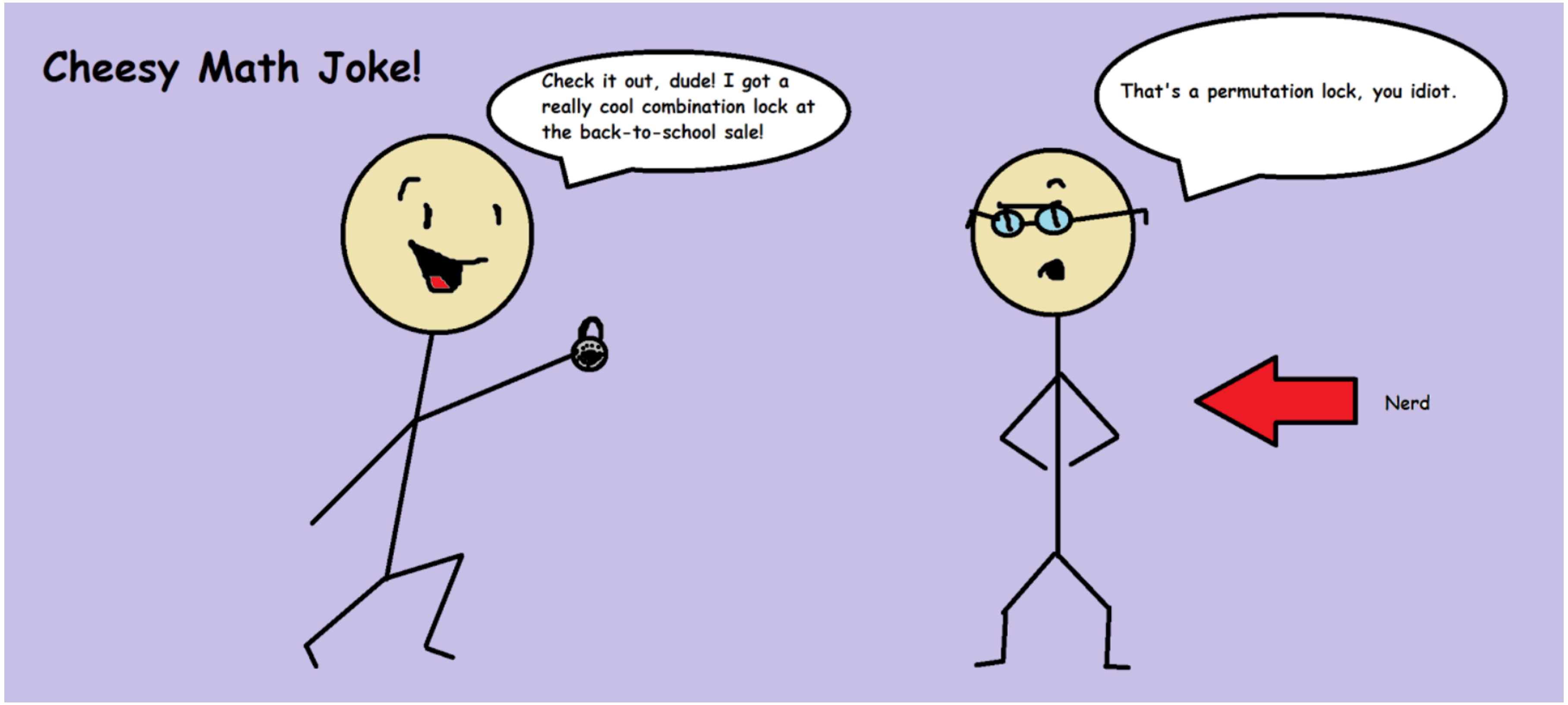
Combinations & Permutations

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?

$$\boxed{\# \text{ ways} = \binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}}$$

C C C C C C C _ _ _

_ _ _ C C C C C C C

C C C _ _ _ _ C C C C

"Derivation"

All the ways to get 7 things correct : $C(10,7) = \binom{10}{7}$

\+ " " 8 " : $C(10,8)$

\+ " " 9 " : $C(10,9)$

\+ " " 10 " : $C(10,10)$

Combinations

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

$\boxed{C(10,2)}$

_ H _ _ _ _ H _ _ _

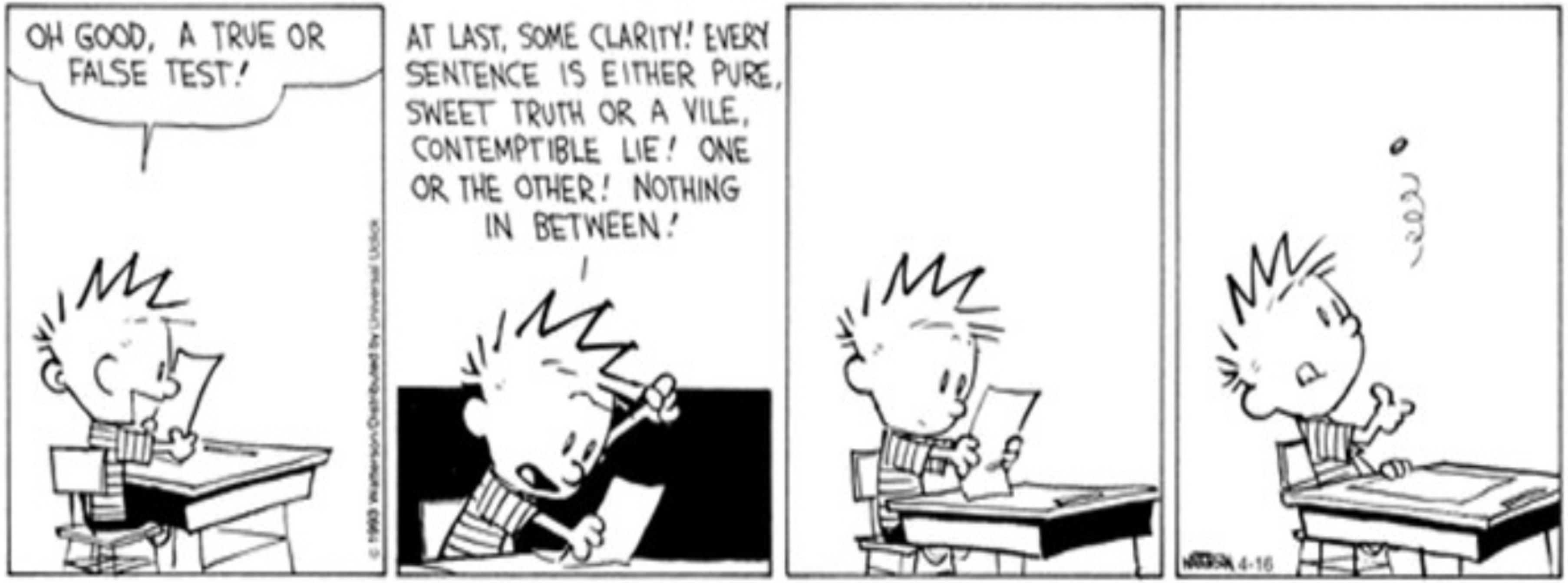
H H _ _ _ _ _ _ _ _

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

$\boxed{C(10,2) + C(10,1) + C(10,0)}$

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

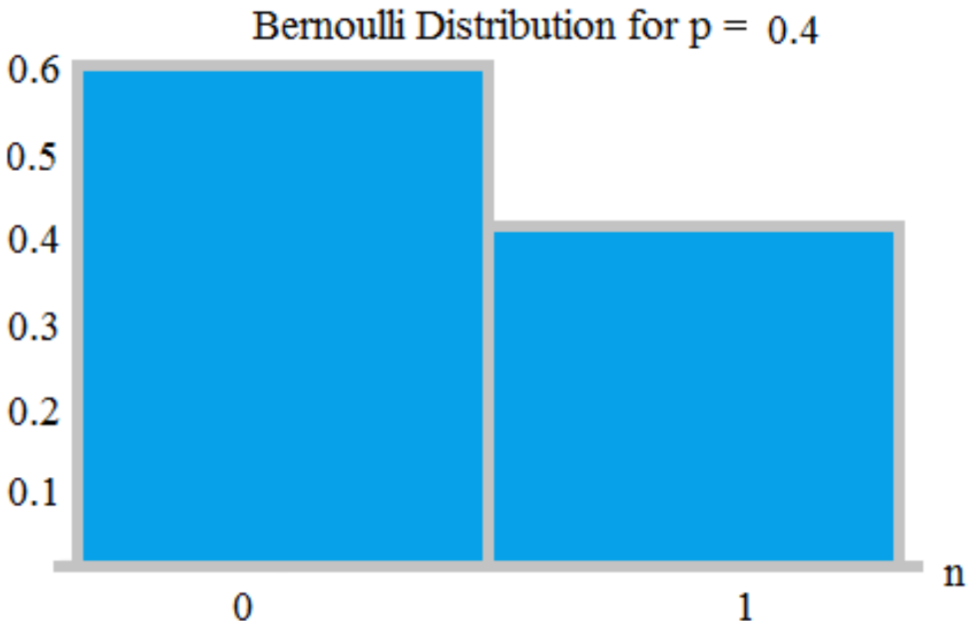
$$f(1) = p_X(1) = P(X = 1) = p \quad \text{and} \quad p_X(0) = P(X = 0) = 1 - p$$

We denote this distribution by Ber(p).

input

shorthand

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 Ber(p).

Describing the pmf with a single equation:

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

$p_X(x) =$?

$$\boxed{p_x(x) = p^x \cdot (1-p)^{1-x}}$$

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?

Think of each question as a Bernoulli trial

$p(\text{success}) = 1/4$

$p(\text{failure}) = 3/4$

Random variable X = # of questions that you get correct.

$X = \{0, 1, 2, 3, 4, 5\}$

What is $P(X=3)$?

Sum of Bernoulli Random Variables

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

$$p(X=0) = \left(\frac{3}{4}\right)^5$$

What is the probability that you get 1 problem correct?

5 ways to get 1 problem correct

C _ _ _ _

_ C _ _ _

_ _ C _ _

_ _ _ C _

_ _ _ _ C

$$p(X=1) = \binom{5}{1} \cdot \left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^4$$

$$\boxed{p(X=1) = 5\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^4}$$

prior

$$\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)\left(\frac{3}{4}\right)\left(\frac{3}{4}\right)\left(\frac{3}{4}\right) +$$

Sum of Bernoulli Random Variables

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

$$\boxed{P(X=3) = \binom{5}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2}$$

Sum of Bernoulli Random Variables → A Binomial Distribution!

<u>Example</u> : What is the probability that you get *k* questions correct out of *n* problems total?

$p(\text{correct}) = p$

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

Binomial Distribution

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

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

shorthand

We denote this distribution by Bin(n, p).

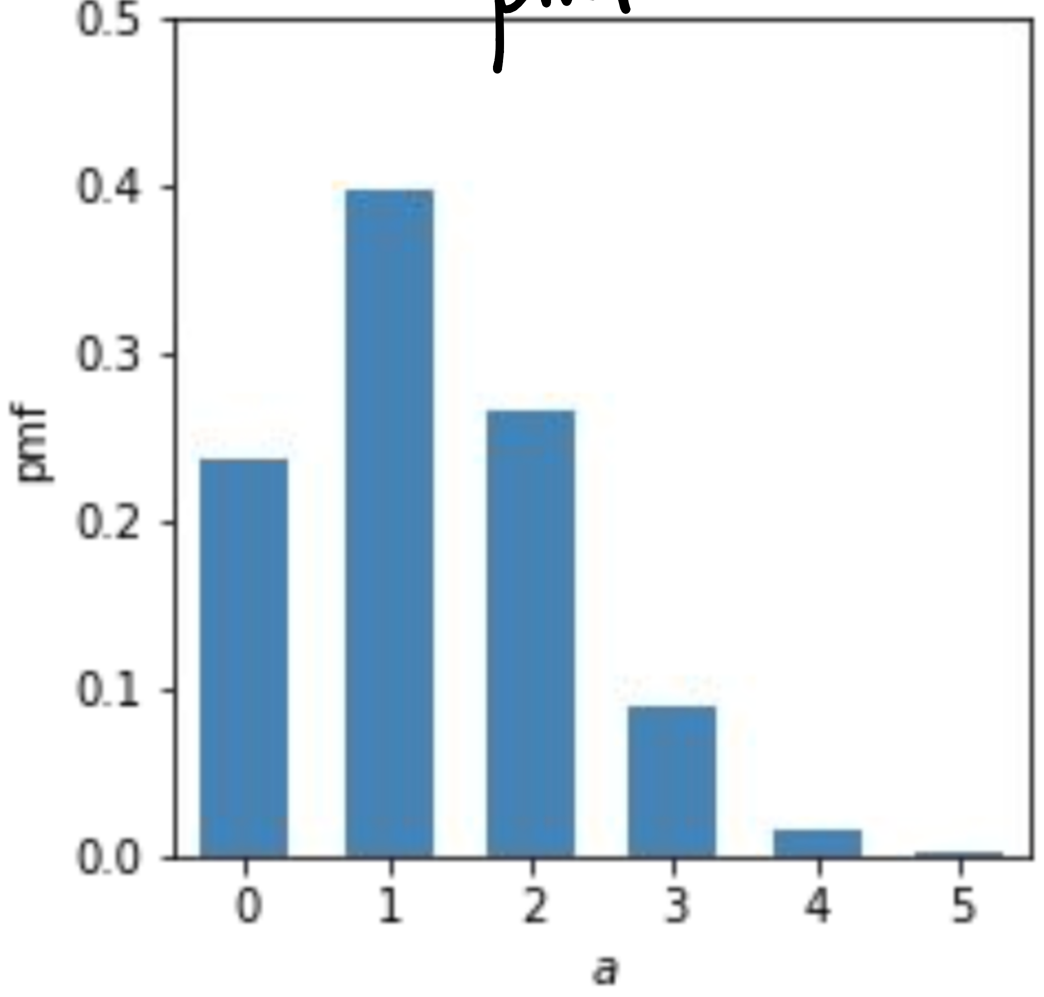
parameters

$\binom{n}{k}$ is the binomial coefficient.

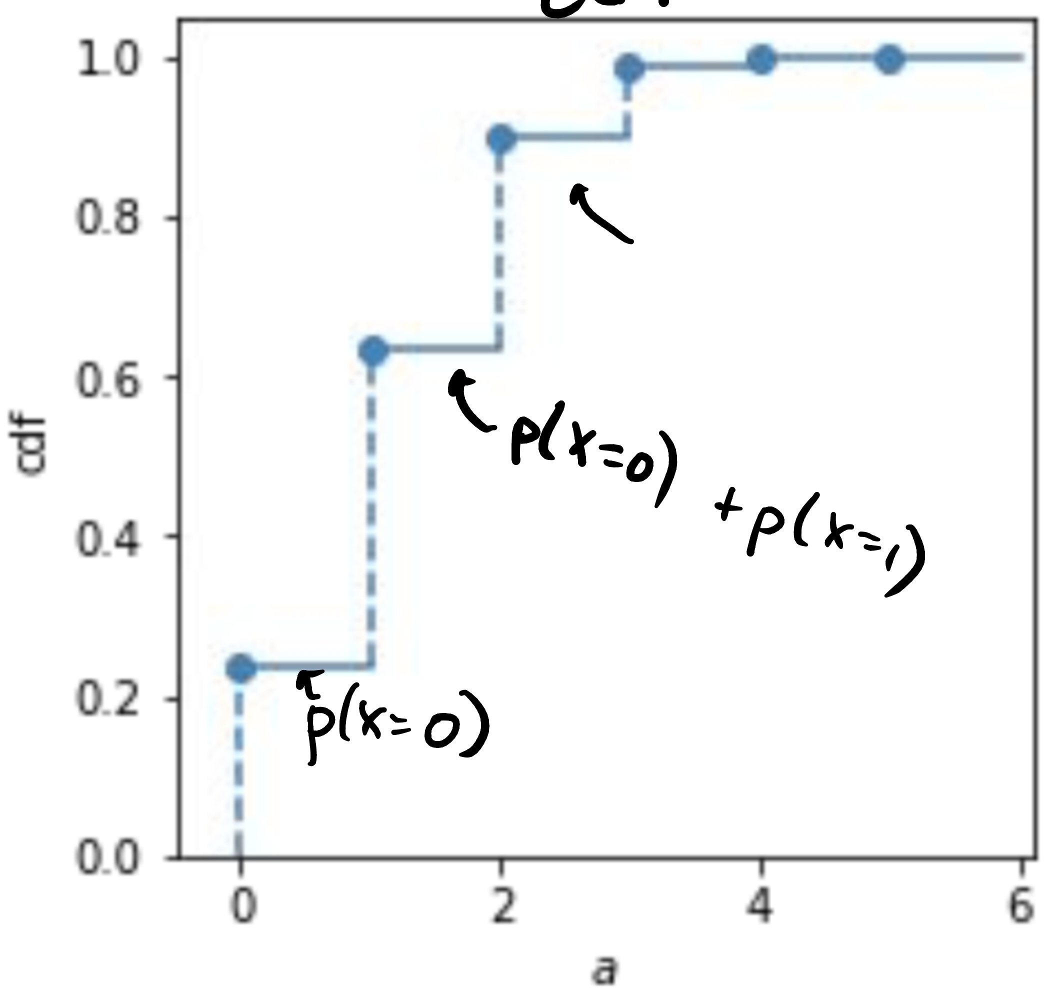
Binomial Distribution

Example: n=5, p=0.25

pmf

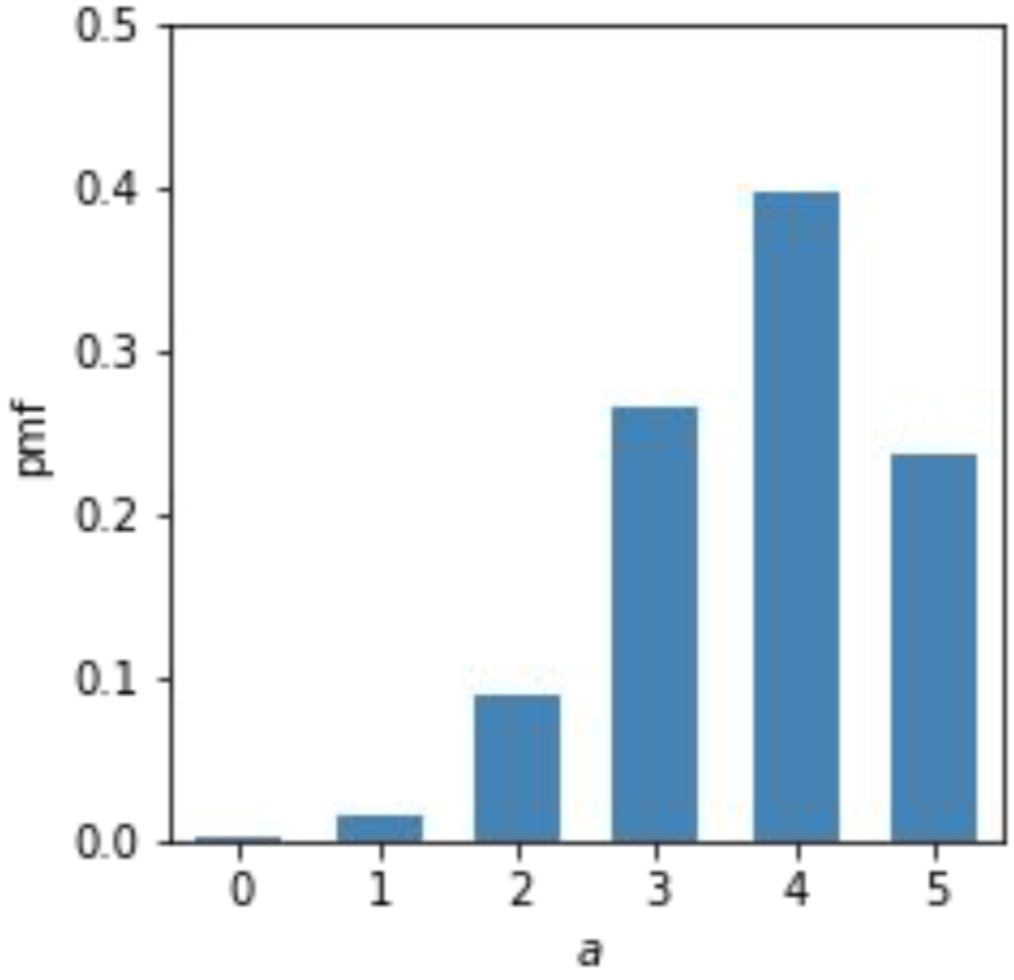


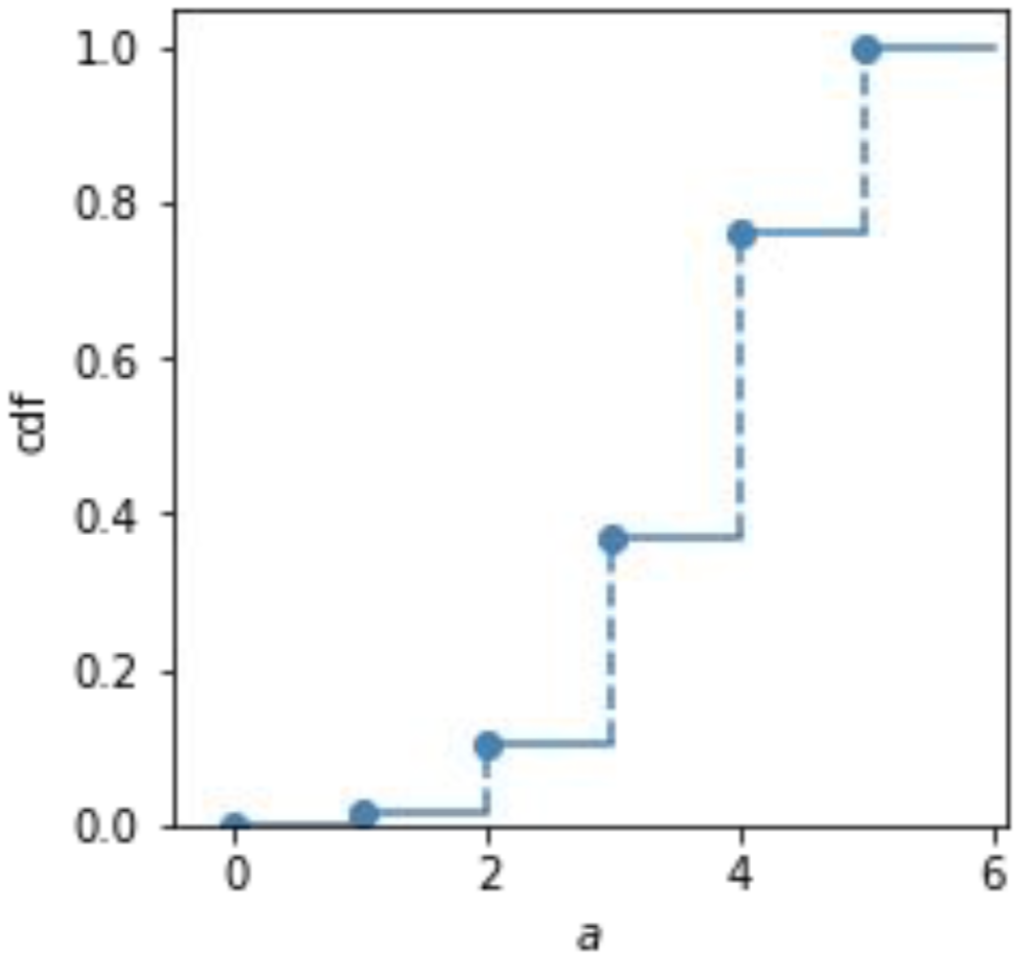
cdf



Binomial Distribution

Example: n=5, p=0.75





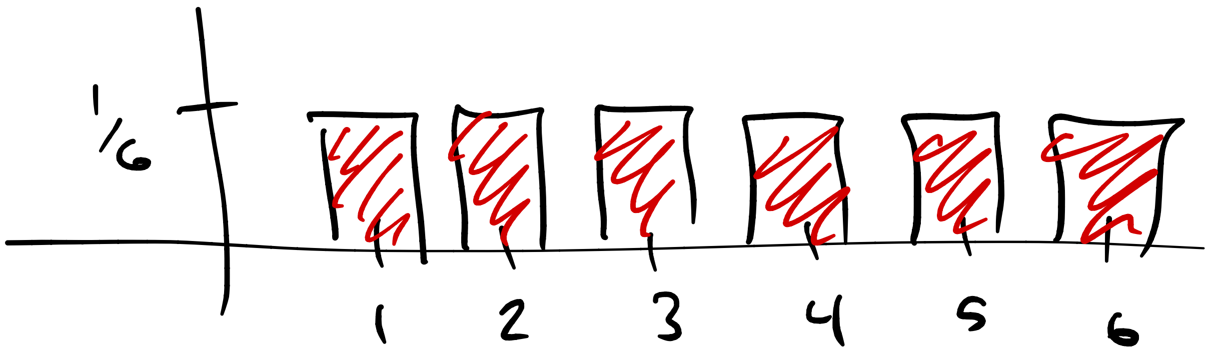
Discrete Uniform Distribution

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

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

Example : What is the distribution of a fair die?

$X = 1, 2, 3, 4, 5, 6$



Next Time:

- More Random Variables and Coin flipping