intro to data science with probability & statistics

CSCI 3022

September 19, 2018

1. More discrete RVs
2. Common distributions

Department of Computer Science
UNIVERSITY OF COLORADO BOULDER

Dan Larremore

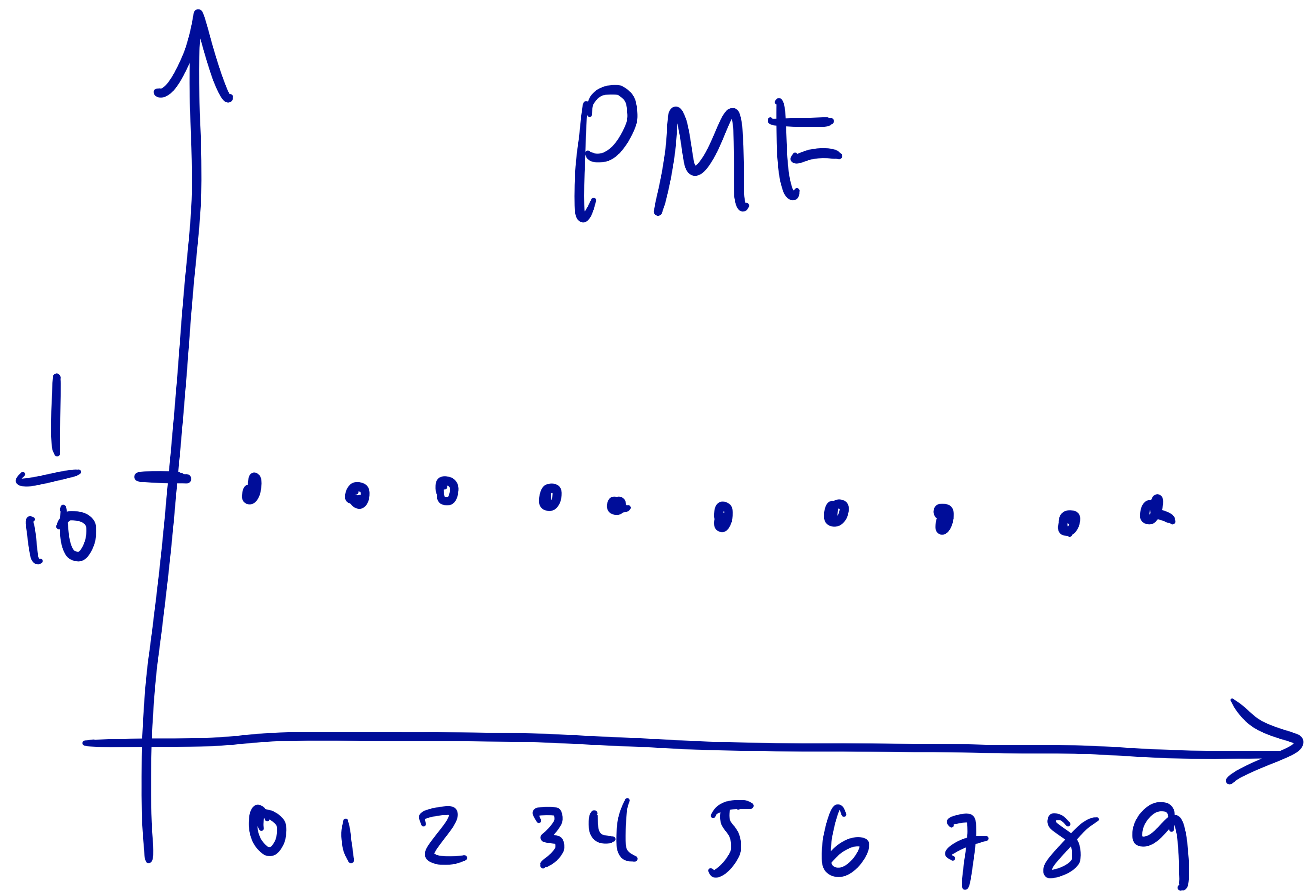
The Discrete Uniform Distribution

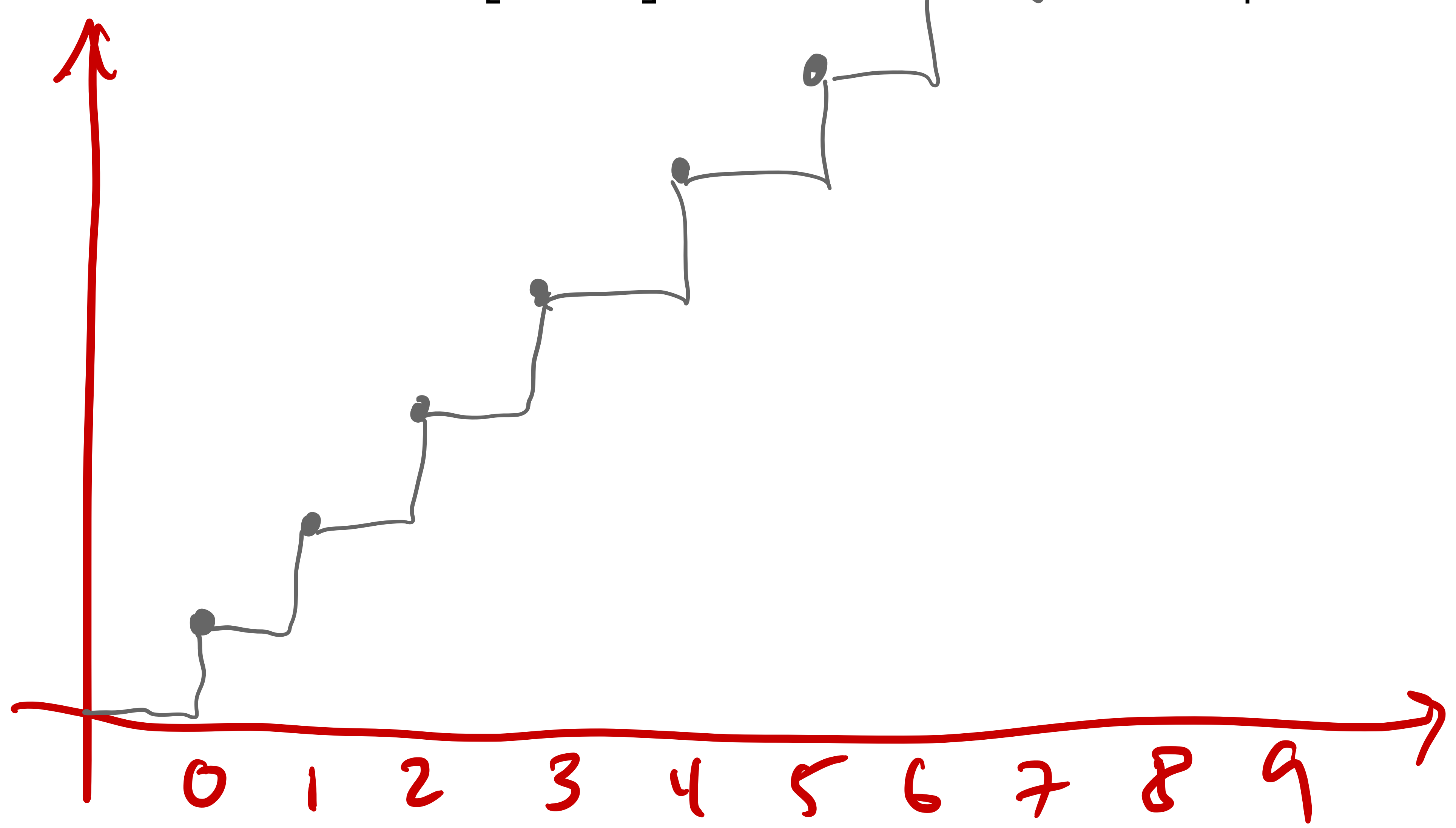
- **Definition**: the discrete *uniform* distribution assigns a probability mass of $\frac{1}{n}$ to each of n values in [a,b]. We write it as *unif(a,b).*

- **Ponder**: can you think of an example of a random experiment whose outcome is a discrete uniform distribution? What are *a* and *b*?

  rolling die $a = 1,\ b = 6,\ n = 6$

- **Plot**: the probability mass function for unif[0,9], *n*=10. Then plot the CDF.





The Bernoulli Distribution

- **Definition**: a discrete RV *X* has a *Bernoulli distribution* with parameter *p*, where $0 \le p \le 1$, if its probability mass function (PMF) is given by:

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

- We denote this distribution by *Ber*(p)

best joke (in-class only)

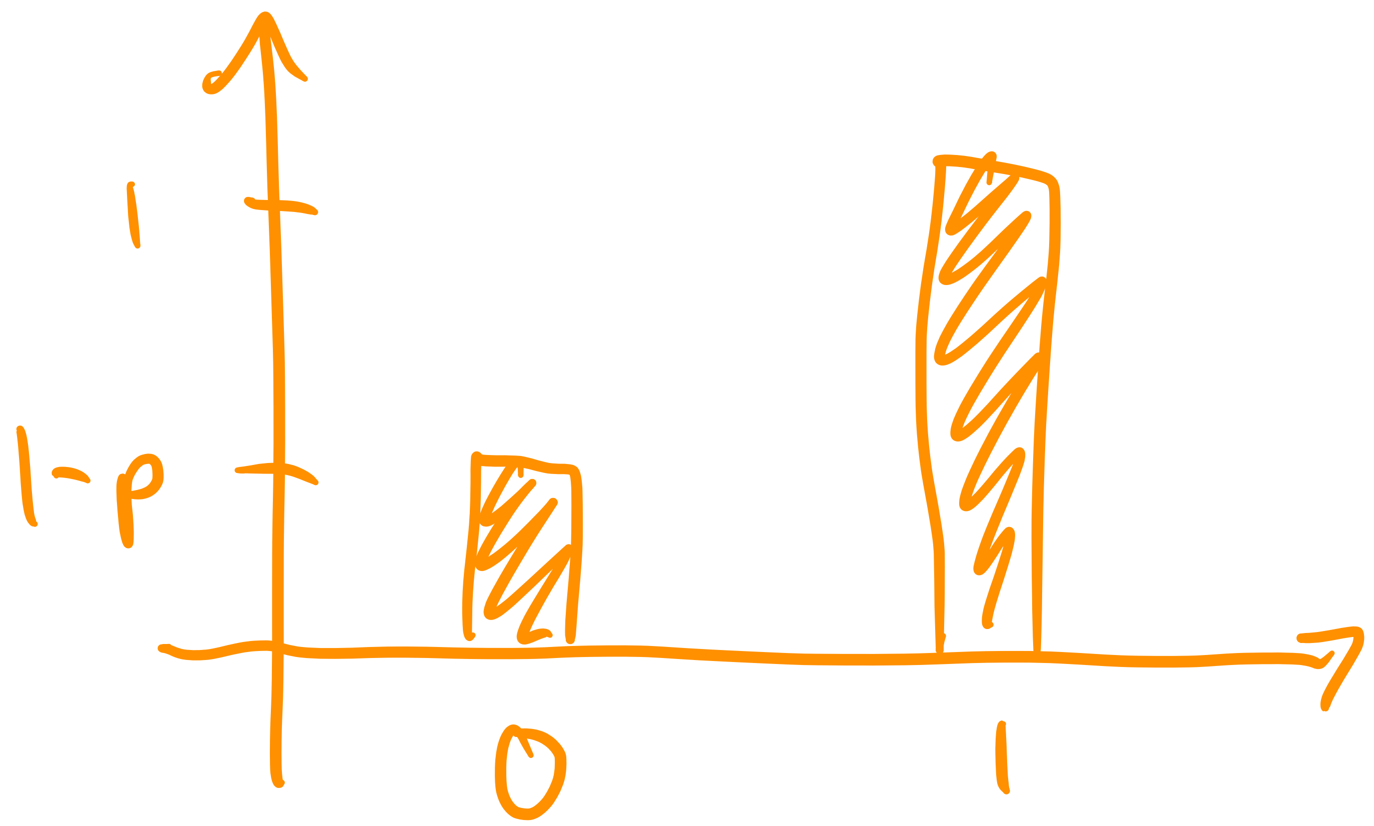
- Look closely—what does this PMF remind you of?

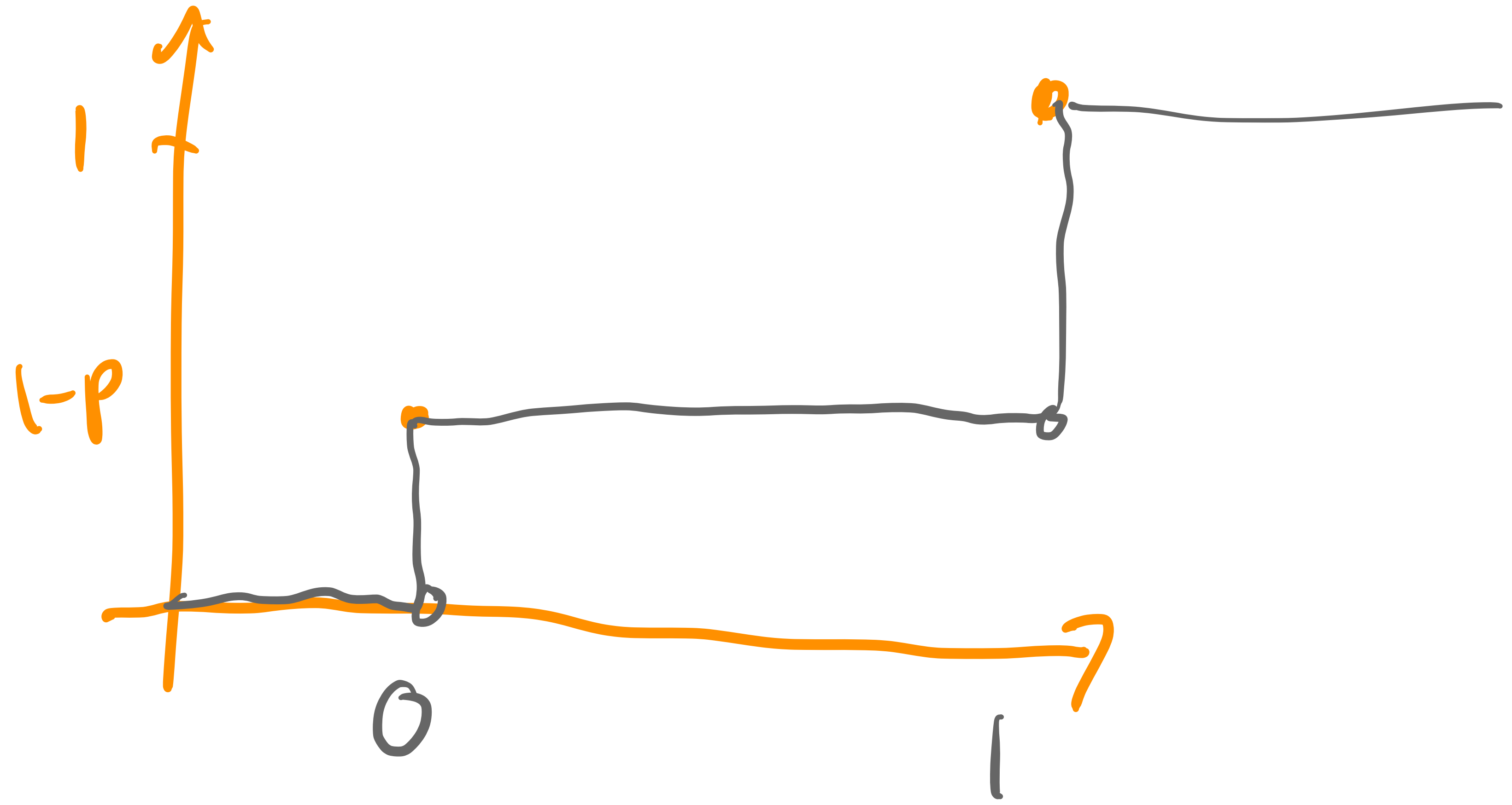
a biased coin

- What is the CDF of the Bernoulli distribution?

$F(0) = 1-p$

$F(1) = 1$





Counting

- Let's set aside the Bernoulli distribution for a moment and turn to counting.
- Believe it or not, *counting* comes up all over the place in probability, and therefore in CS, math, physics, engineering, data science, etc.
- Some counting is easy: how many integers are there in the interval [0,9]?
- We're interested in counting that requires math though: Ancsa, Brendan, Caterina, David, and Elly are standing in line. How many different orders could they stand in?
- If there are 10 problems on an exam and you need 7 or more correct to pass, how many different ways are there to pass the test?

Counting

- We'll talk about two key kinds of counting problems today:

- Counting **permutations** means counting the number of ways a set of objects can be ordered (or, more precisely, permuted).

Ancsa, Brendan, Caterina, David, and Elly are standing in line. How many different orders could they stand in?

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

If there are 10 problems on an exam and you need 7 or more correct to pass, how many different ways are there to pass the test?

Counting - Permutations

- How many ways are there of ordering 1 object? 1
- How many ways are there of ordering 2 objects? A, B or B, A $\rightarrow$ 2
- What about 3 objects?

  ABC BAC CAB

  ACB BCA CBA 6

- What is the formula for the number of possible permutations of *n* objects?

n = 5 objects

$$\underbrace{\quad}_{5 \text{ possible}}^{\#1} \quad \underbrace{\quad}_{4 \text{ possible}}^{\#2} \quad \underbrace{\quad}_{3 \text{ possible}}^{\#3} \quad \underbrace{\quad}_{2 \text{ possible}}^{\#4} \quad \underbrace{\quad}_{1 \text{ choice}}^{\#5}$$

$$5 \times 4 \times 3 \times 2 \times 1 = 5!$$

$$\boxed{\# \text{perms} = n!}$$

$$\boxed{\text{define: } 0! = 1}$$

Counting - Permutations 2

- Say there are 10 people in a race. How many ways are there of awarding the gold, silver, and bronze?

place	G	S	B	4	5	6	7	8	9	10

$10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 10!$

① $10 \cdot 9 \cdot 8 = 720$

② 10! total perms, but 7! I don't care about.

$$\frac{10!}{7!} = \frac{10 \cdot 9 \cdot 8 \cdot \cancel{7} \cdot \cancel{6} \cdot \cancel{5} \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2} \cdot \cancel{1}}{\cancel{7} \cdot \cancel{6} \cdot \cancel{5} \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2} \cdot \cancel{1}} = 10 \cdot 9 \cdot 8 = 720$$

Counting - Combinations

- We could approach the previous problem another way…

10 people… ask: how many ways could I possibly choose a subset of 3 of them?

10 choose 3

Note: 3! = 6 ways of ordering those 3.

$$\text{\# of possible GSB} = [10 \text{ choose } 3] \times 3! = \frac{10!}{7!}$$

$$[10 \text{ choose } 3] = \frac{10!}{7!\,3!}$$

$$n \text{ choose } k = \frac{n!}{(n-k)!\,k!}$$

Counting - Combinations

- The previous slide shows us how to use permutations to get to combinations.
- If I have *n* objects and I choose *k* of them, I can do this in *n* choose *k* ways.
- Various notations: $\binom{n}{k}$ $\cancel{nCk}$ $\frac{n!}{k!(n-k)!}$ $C_{n,k}$

formula

{n \choose k}

Counting - Combinations

- The previous slide shows us how to use permutations to get to combinations.

- If I have n objects and I choose k of them, I can do this in n choose k ways.

- Various notations: $\binom{n}{k}$ $\quad nCk \quad$ $\frac{n!}{k!(n-k)!}$ $\quad C_{n,k}$

- So, how many ways are there to choose 3 runners from a total of 10?

$$\binom{10}{3} = \frac{10!}{(10-3)!\,3!} = \frac{10!}{7!\,3!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = \frac{720}{6} = 120$$

Counting - Combinations

- Let's connect some dots.
- Let's compute: $\binom{1}{1}$ $\frac{1!}{(1-1)!\,1!} = 1$
- $\binom{2}{1}$, $\binom{2}{2}$ $\frac{2!}{(2-1)!\,1!} = 2$ $\frac{2!}{(2-2)!\,2!} = 1$
- $\binom{3}{1}$, $\binom{3}{2}$, $\binom{3}{3}$
- Now $\binom{1}{0}$, $\binom{2}{0}$, $\binom{3}{0}$ and $\binom{0}{0}$
- What do you notice?

Binomial Coefficients

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

Pascal's Triangle

$$1$$
$$1 \quad 1$$
$$1 \quad 2 \quad 1$$
$$1 \quad 3 \quad 3 \quad 1$$
$$1 \quad 4 \quad 6 \quad 4 \quad 1$$

Combinations applied!

- If there are 10 problems on an exam and you need 7 or more correct to pass, how many different ways are there to pass the test?

Pass: 7 or 8 or 9 or 10

$$\binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10} = \boxed{176}$$

120

$$\frac{10!}{(10-8)!\,8!} = \frac{\cancel{10}^{5} \cdot 9}{\cancel{2} \cdot 1} = 45$$

10

$$\frac{10!}{(10-9)!\,9!} = \frac{10!}{(10-1)!\,1!} = \binom{10}{1}$$

1

wow!!

The Binomial Distribution

- Suppose I sum 5 Bernoulli random variables. This sum is a RV too!
- It takes on values in the interval [0,5].
- What is the PMF of the sum of 5 Bernoulli RVs? Let's build it up!

The Binomial Distribution

- The sum of Bernoulli RVs is the **Binomial Distribution** (see top of slide).
- It is the distribution of the number of “heads” you’ll get when flipping a coin n times.
- Note that it is parameterized by the number of flips n and the Bernoulli parameter p. So we call this *Bin(n,p)*.

The Binomial Distribution

- The sum of Bernoulli RVs is the **Binomial Distribution** (see top of slide).
- It is the distribution of the number of “heads” you’ll get when flipping a coin *n* times.
- Note that it is parameterized by the number of flips n and the Bernoulli parameter p. So we call this *Bin(n,p)*.

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

- **Practice**: what’s the probability that a biased coin with *p*=0.8 comes up heads 7 times in 10 flips?

The Binomial Distribution

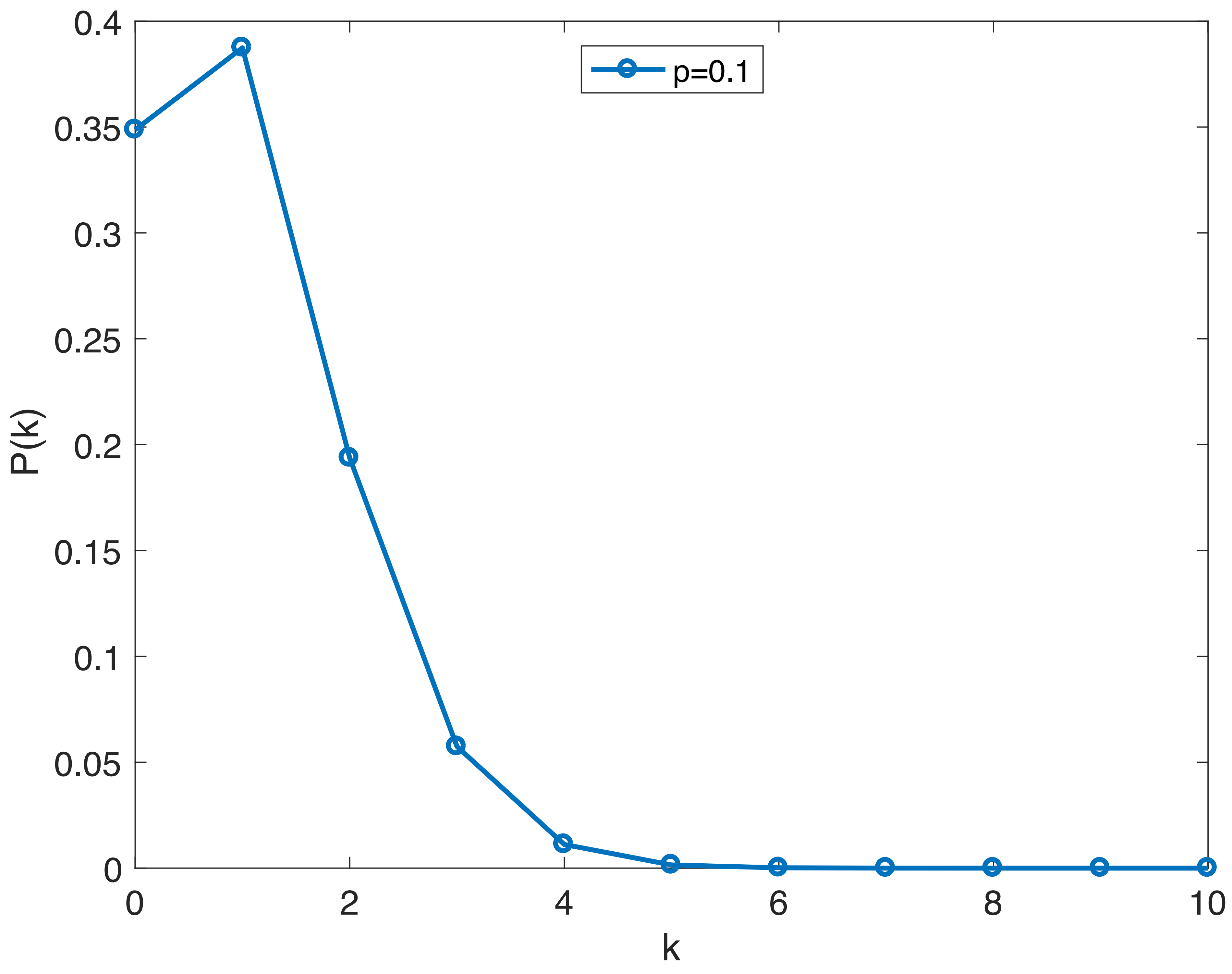
- **Practice**: You are the caretaker of 50 octopuses at the aquarium. 100% of the octopuses are magnificent and intelligent. They *get* you. And you get them. It's a great job, and you feel lucky to have it. However, each octopus has a 12.5% probability of being grumpy on any given day. Today, you choose 20 octopuses at random to put in the Display Habitat.

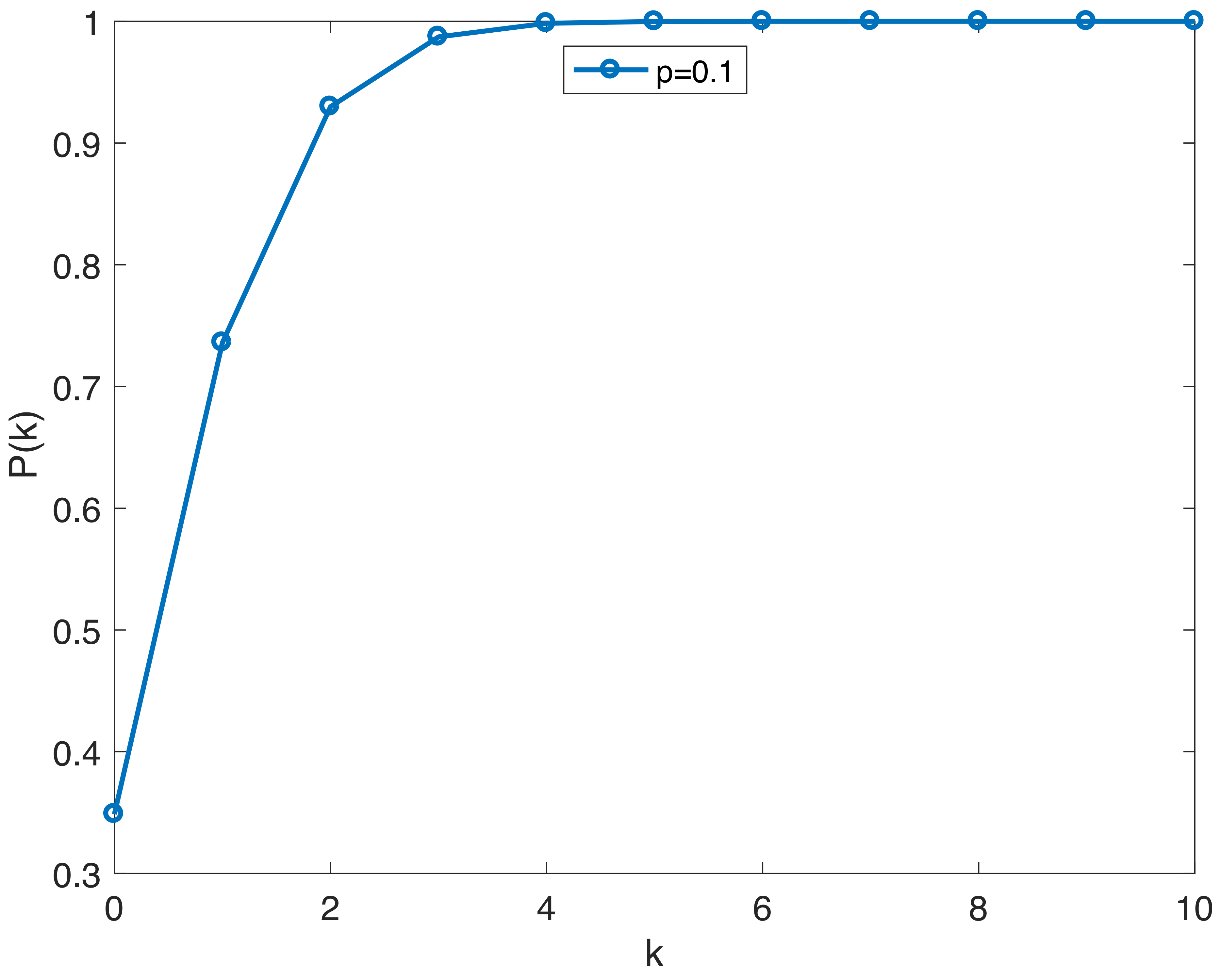
- What is the probability that 5 of the Display Habitat octopuses will be grumpy?

- What is the probability that all the Display Habitat octopuses will not be grumpy?

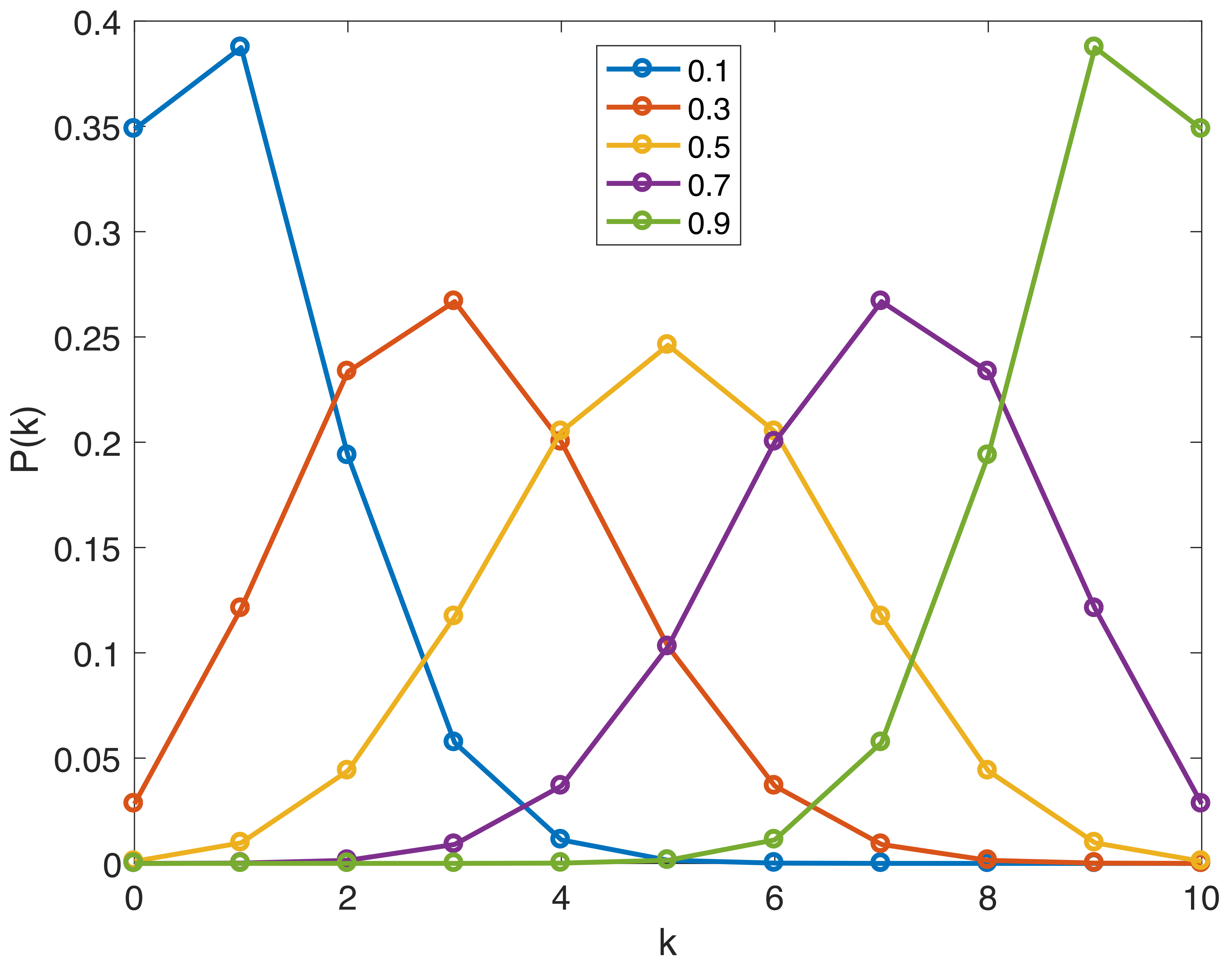
- What are we assuming when we answer these Qs using the binomial distr.?

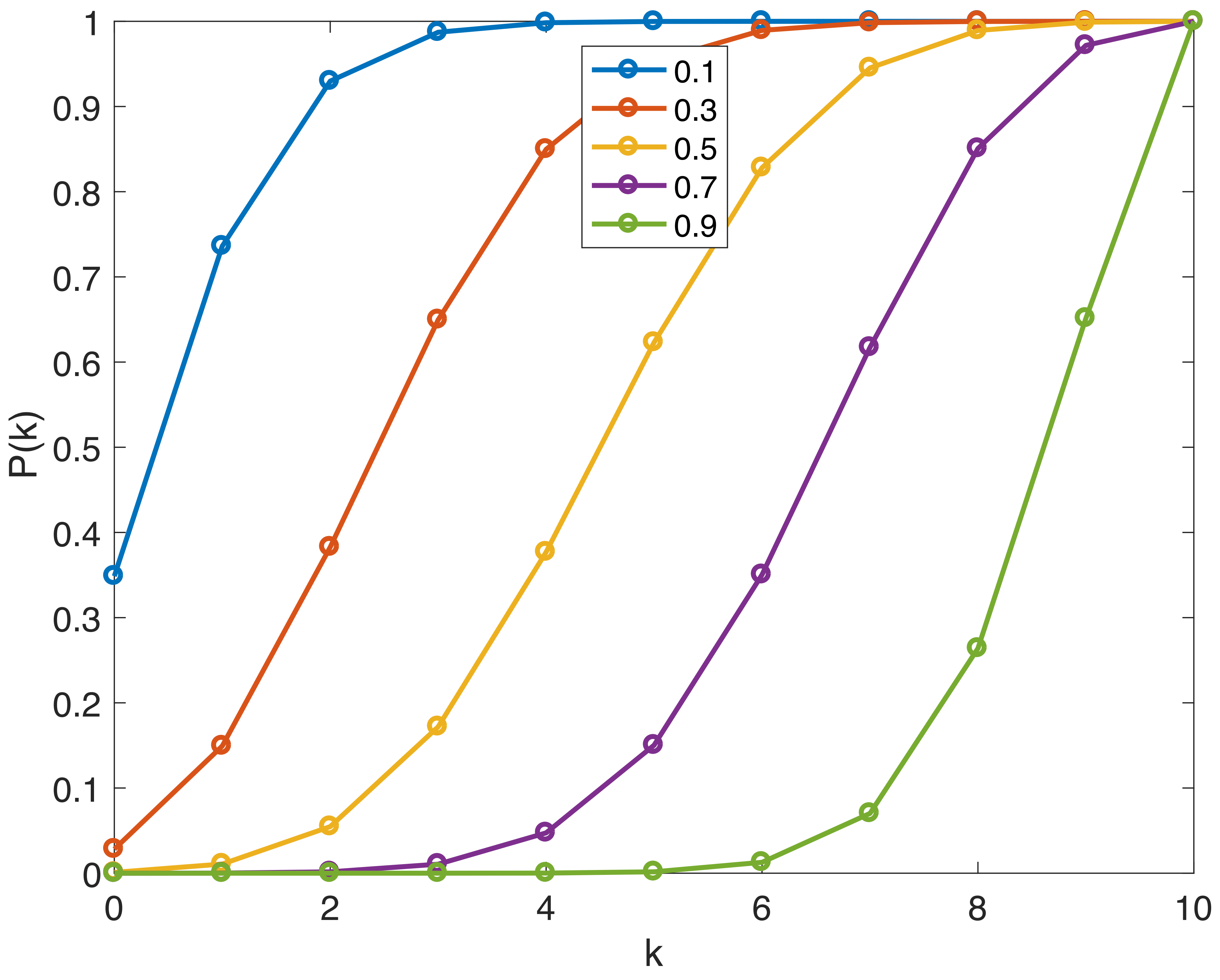
Properties of the Binomial PMF





Properties of the Binomial PMF





Properties of the Binomial PMF