Probability in Computing

LECTURE 19

CS 237

Last time

- Variance and its properties
- Discrete Distributions: Bernoulli, Uniform, and Binomial

Today

- Discrete Distributions: Binomial concluded, Geometric, Negative Binomial

Reminders

- HW 9 due Thursday

Reading

- LLM 19.4.6, P 3.1.5

Lecture slides are in Piazza Resources and will be on class web page later

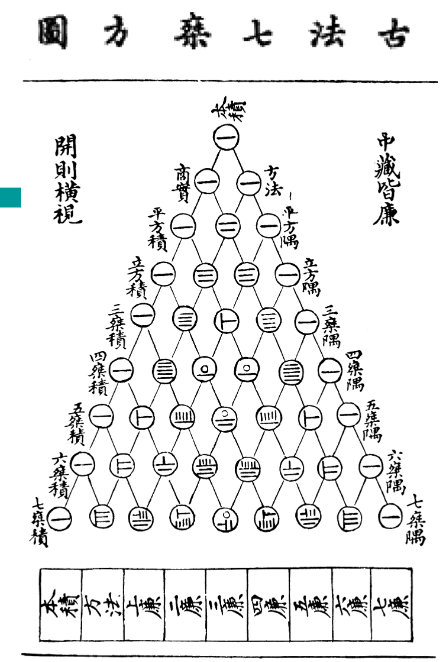
CS 237

Binomial Distribution

X ~ Binomial(N,p):

- N independent Bernoulli(p) trials
- X = number of successes
- Canonical experiment: Take a coin with probability of heads p, flip it N times and count the number of heads.
- Range(X) = {0, ..., N}
- Pr(X = k) =

$$\binom{N}{k} \cdot p^k \cdot (1-p)^{N-k}$$



Yang Hui's triangle
(Pascal's triangle)

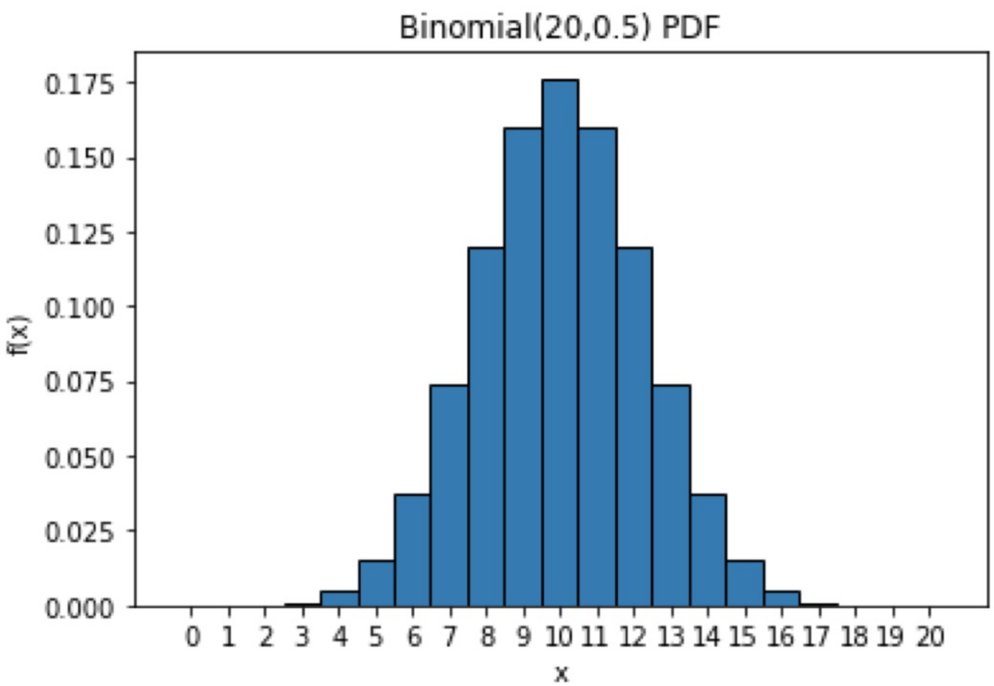
NOTE: Binomial is the **sum** of independent Bernoulli RVs:

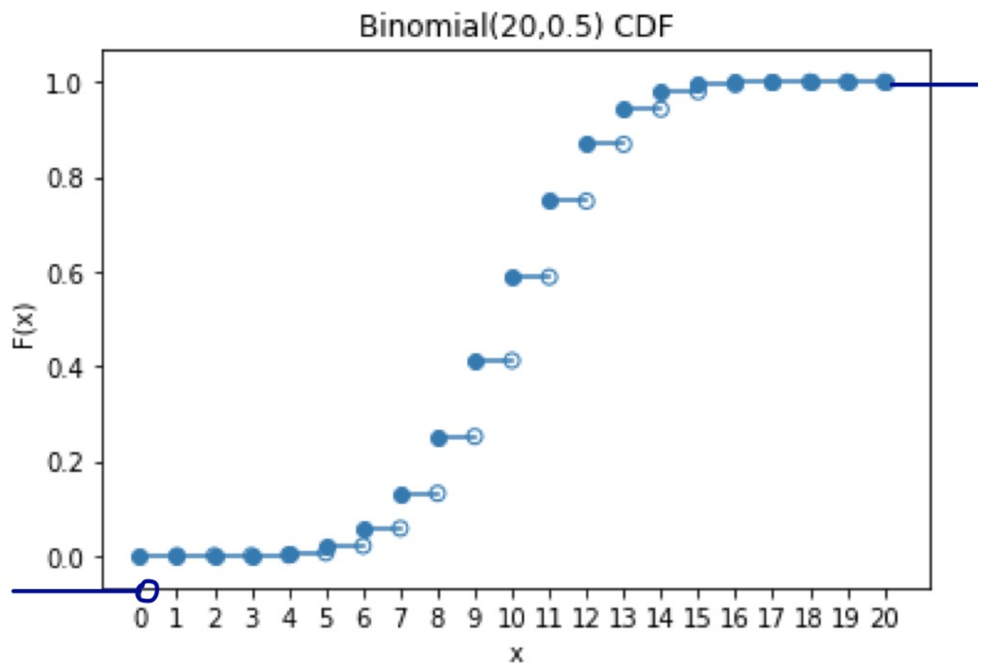
$X = X_1 + X_2 + ... + X_n$

Tophat Question One

Binomial Distribution

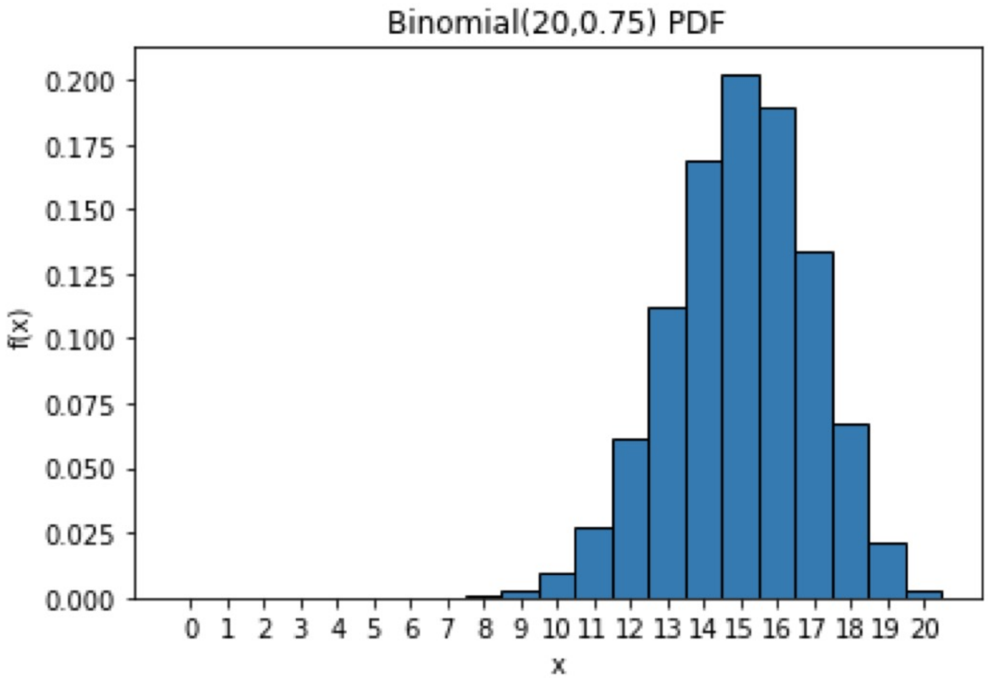
- X ~ Binomial(N, p) example

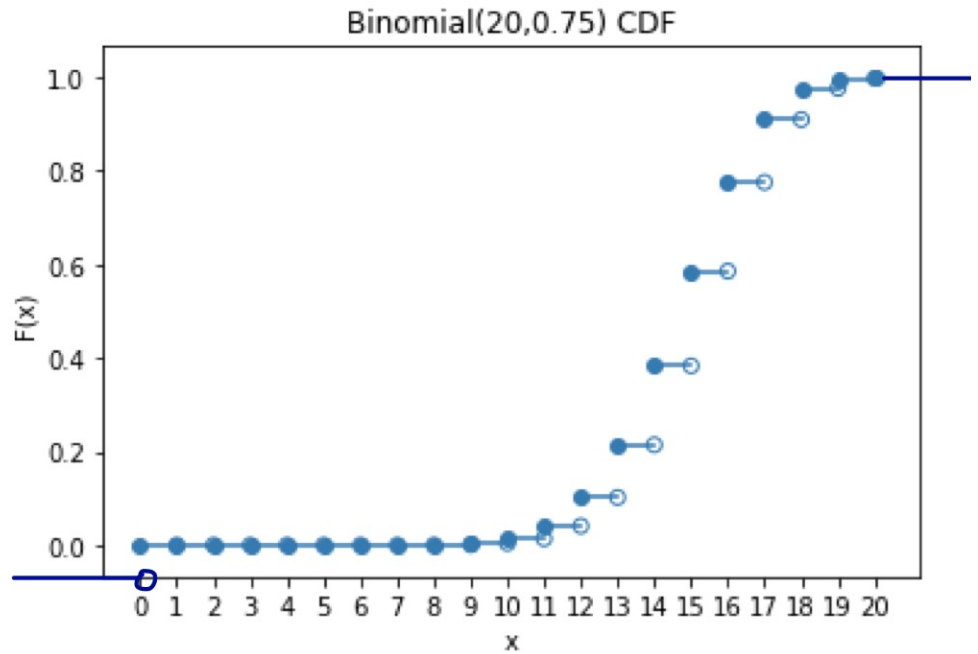




Binomial Distribution

- X ~ Binomial(N, p) example





Binomial Distribution

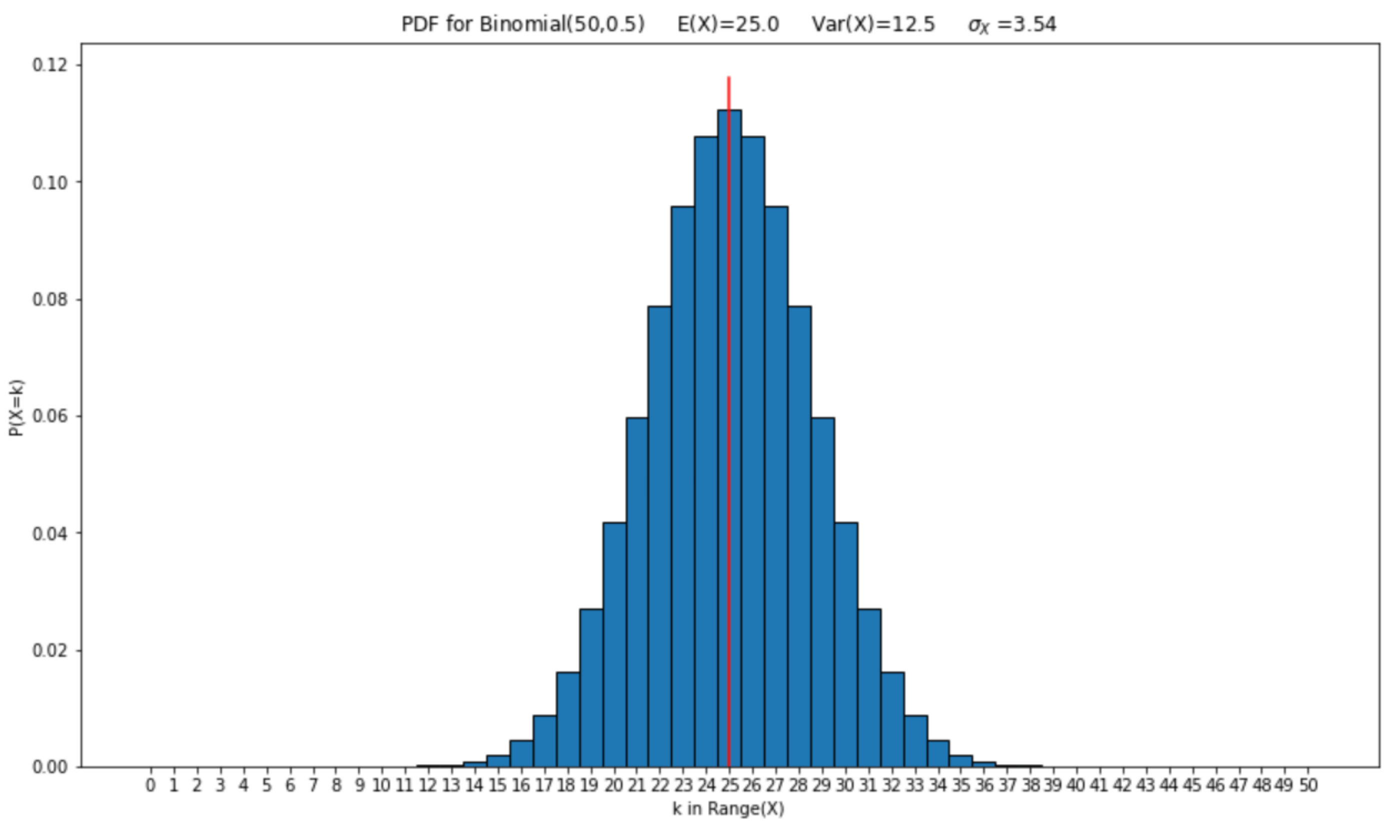
- Let $X \sim \text{Binomial}(N, p)$
- $E(X) = ?$

NOTE: Binomial is the **sum** of independent Bernoulli RVs:

$X = X_1 + X_2 + ... + X_n$

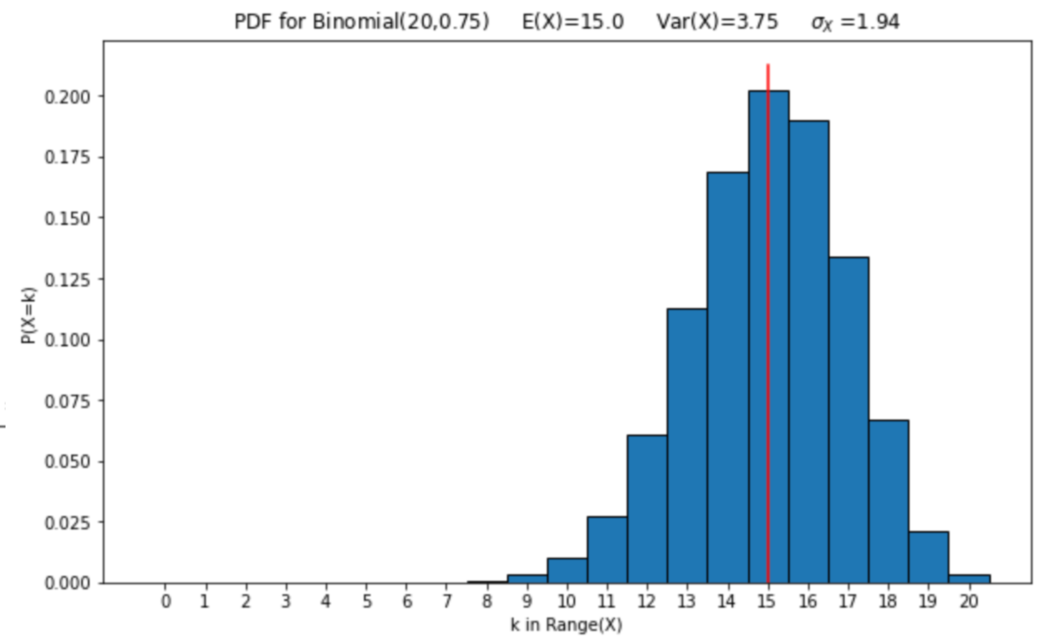
Binomial Distribution

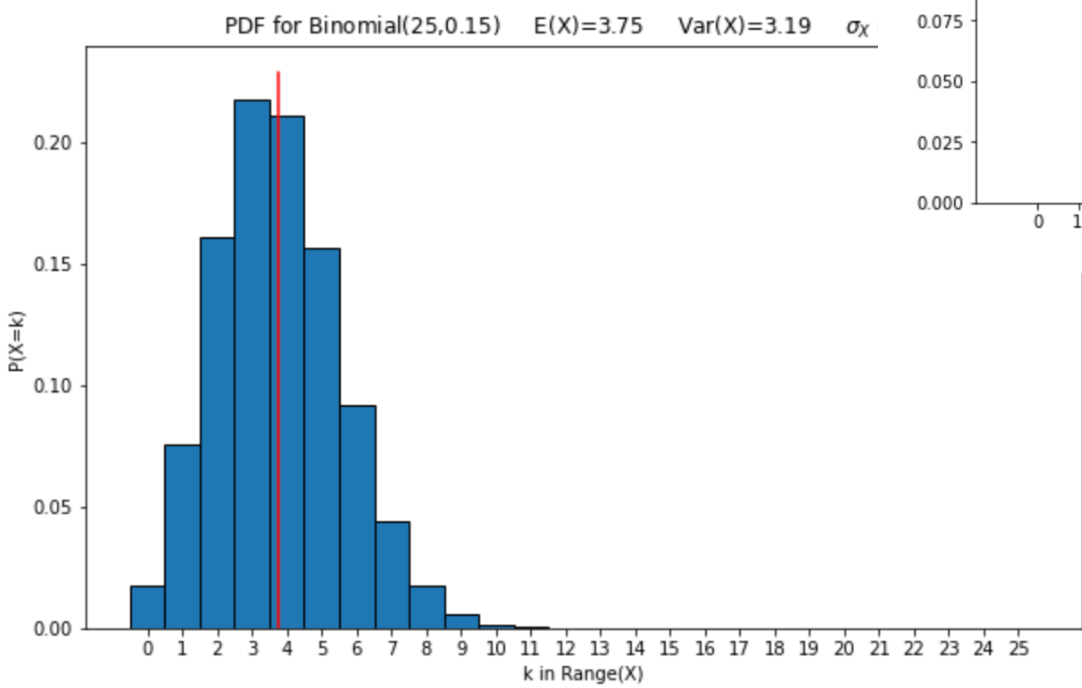
- Let X ~ Binomial(N, p)
- E(X) = $n * p$



Binomial Distribution

- Let X ~ Binomial(N, p)
- E(X) = $n * p$





Binomial Distribution

- Let $X \sim \text{Binomial}(N, p)$
- $\text{Var}(X) = ?$

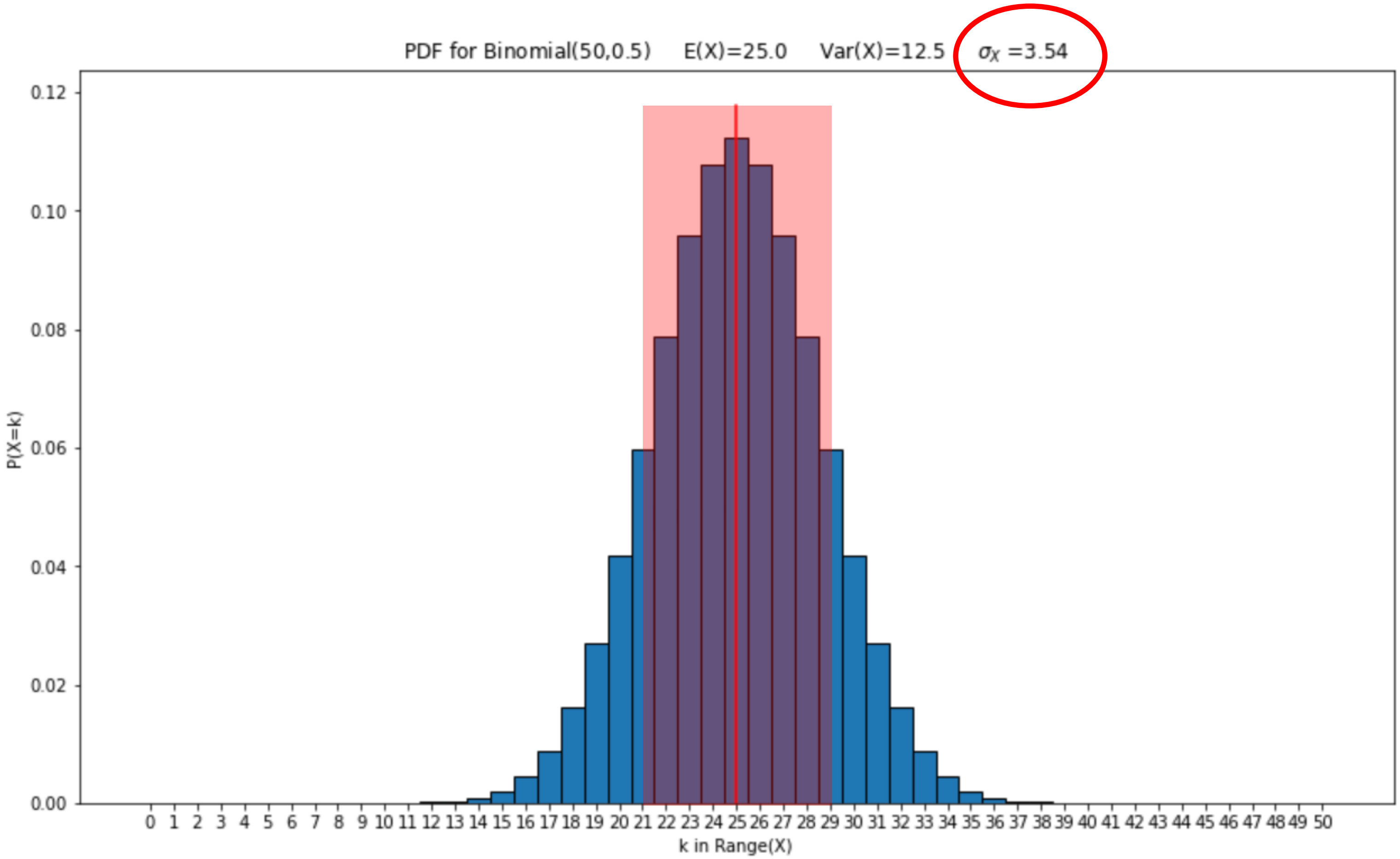
Tophat Question Two

NOTE: Binomial is the **sum** of independent Bernoulli RVs:

$X = X_1 + X_2 + \ldots + X_n$

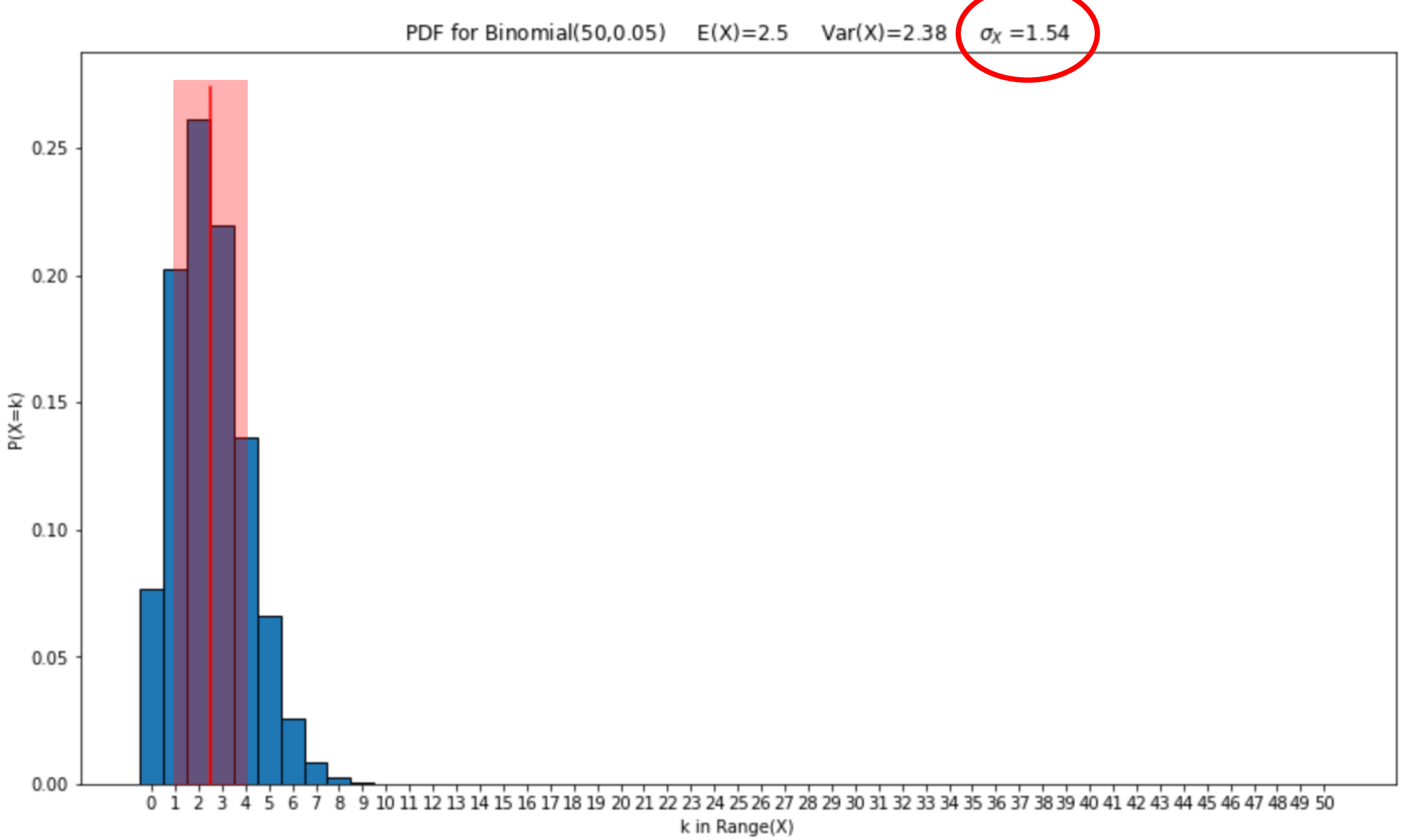
Binomial Distribution

- Let X ~ Binomial(N, p)
- Var(X) = $N * p * (1 - p)$



Binomial Distribution

- Let X ~ Binomial(N, p)
- Var(X) = $N * p * (1 - p)$



Binomial Distribution

The motivation for this distribution comes from the fact that many complex phenomena are composed of the additive effect of many small binary choices or events (Bernoulli Trials!); a vivid illustration of this can be seen in the Galton Board or Quicunx:

https://www.mathsisfun.com/data/quincunx.html

https://www.youtube.com/watch?v=J7AGOptcR1E

Binomial Distribution

The problem with the binomial is there is no simple way to calculate the PDF for large N:

$$\Pr(X = k) = \binom{N}{k} \cdot p^k \cdot (1-p)^{N-k} \qquad \Pr(X \le k) = \sum_{i=0}^{k} \binom{N}{i} p^i (1-p)^{N-i}$$

Example: There are about 20K genes in the human body. Supposing (very naively) that there is a 0.45 probability that a gene is dominant, what is the probability that 9000 are dominant?

$$\binom{20,000}{9,000} (0.45)^{9,000} (0.65)^{11,000}$$

The problem is calculating the binomial coefficient....

Binomial Distribution

```
math.comb(20000,9000)
```

```
710221065897948290123010077467417778567521588915058841117534090731622704232496308933461758317509424634977207693680961
```

This is ~7.1 x 10^{5975}.

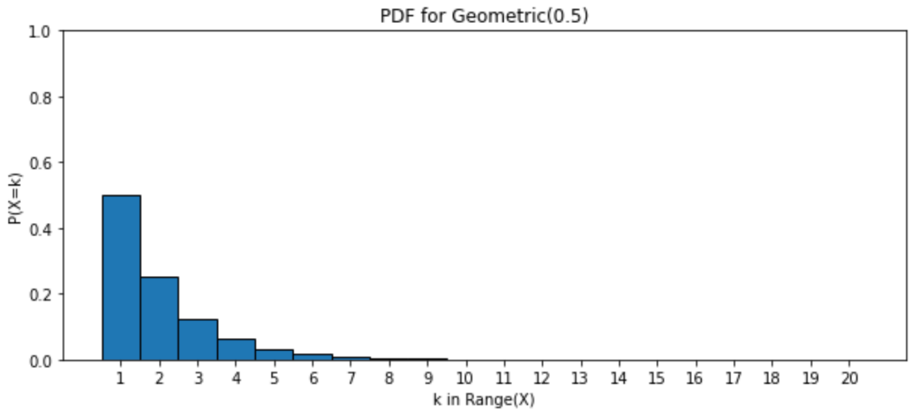
For comparison, there are ~10^{82} atoms in the entire known universe, "only"

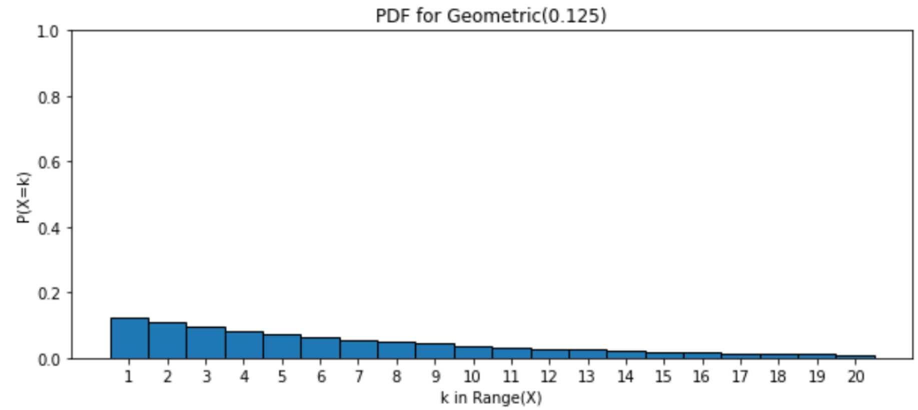
```
10000000000000000000
00000000000000000000
00000000000000000000
0000000000000000000
```

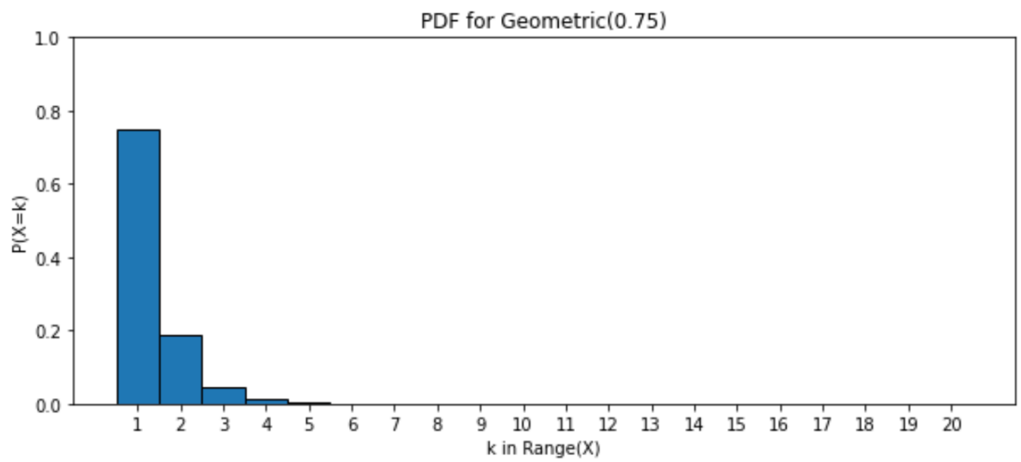
Geometric Distribution

- The Geometric Distribution models the following scenario
  - There are independent Bernoulli trials (possibly infinitely many)
  - Each trial is a success with probability p
  - The process stops at the first success
- X = number of trials until the first success (including the successful trial)
- Notation: $X \sim \text{Geometric}(p)$
- Canonical Problem: Take a coin whose probability of heads is p and count the number of flips until the first head.

Geometric Distribution







Geometric Distribution

- X ~ Geometric(p)
- PDF is simple to compute:

$$\Pr(X = k) \;=\; (1-p)^{k-1}p$$

$$\underbrace{\text{F F F F ..... F}}_{\text{k-1 failures before first success}}\ \text{S}$$

- So is CDF:

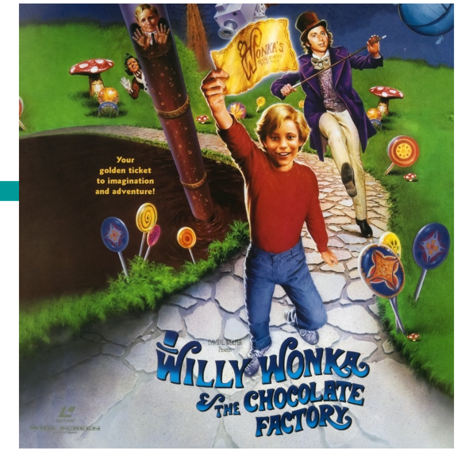
$$\Pr(X > k) \;=\; (1-p)^{k}$$

$$\underbrace{\text{F F F F ..... F}}_{\text{k failures so far....}}$$

$$\Pr(X \le k) \;=\; 1-(1-p)^{k}$$

CS 237

Ask the audience

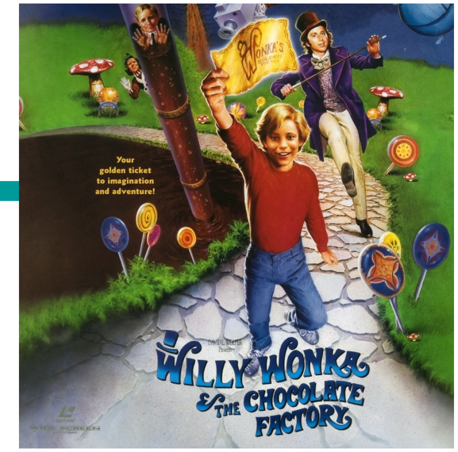


- Willy Wonka puts a golden ticket in each chocolate bar with probability 0.01, independently
- Charlie buys chocolate bars until he gets a golden ticket
- What is the Pr that Charlie buys exactly 5 chocolates?

(A) 0.99^4 (B) $0.99^4 \cdot 0.01$ (C) 0.99^5 (D) none of the above

CS 237

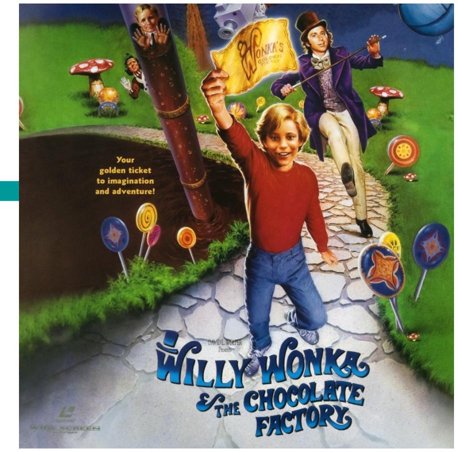
Ask the audience



- Willy Wonka puts a golden ticket in each chocolate bar with probability 0.01, independently
- Charlie buys chocolate bars until he gets a golden ticket
- What is the Pr that Charlie buys at least 5 chocolates?

(A) 0.99^4 (B) $0.99^4 \cdot 0.01$ (C) 0.99^5 (D) none of the above

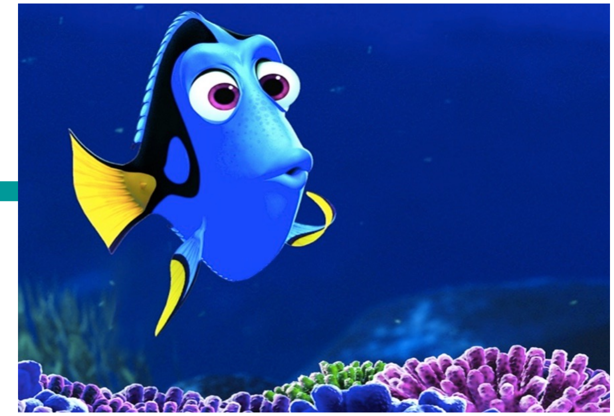
CS 237 Ask the audience



- Suppose Charlie already bought 5 chocolates, none of which had a golden ticket
- What is the Pr that he buys at least 5 more chocolates?

(A) 0.99^4 (B) $0.99^4 \cdot 0.01$ (C) 0.99^{10} (D) none of the above

Memoryless Property



- Theorem Let X ~ Geometric(p).

$$P(\underbrace{X > n+m}_{\text{It will take > n+m total flips.}} \mid \underbrace{X > m}_{\text{You've flipped it m times.}}) = P(\underbrace{X > n}_{\text{It will take > n more flips.}})$$

- Version for a 10-year old:

You flip a coin until heads. The coin doesn't remember how many tails have occurred, so you start over exactly the same with every flip.

- Proof for a 10-year:

Suppose I'm flipping a coin to get heads and while I'm doing it, someone walks in the room, not knowing how many tails I've gotten already. Why would the coin behave differently with this person watching?

Memoryless Property: Version for CS 237

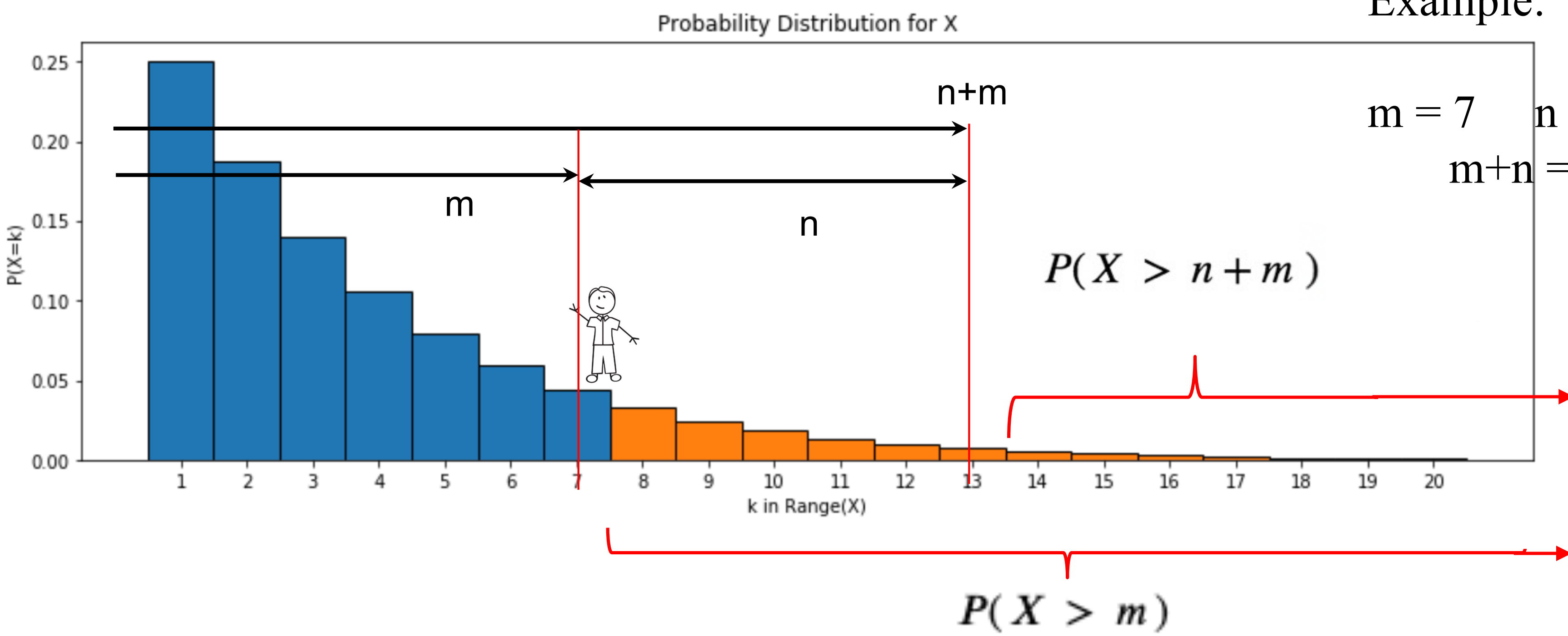
Theorem A random variable X is called memoryless if, for any n, m ≥ 0,

$$P(X > n+m \mid X > m) = P(X > n)$$

Fact: For any probability p, X ~ Geometric(p) has the memoryless property.

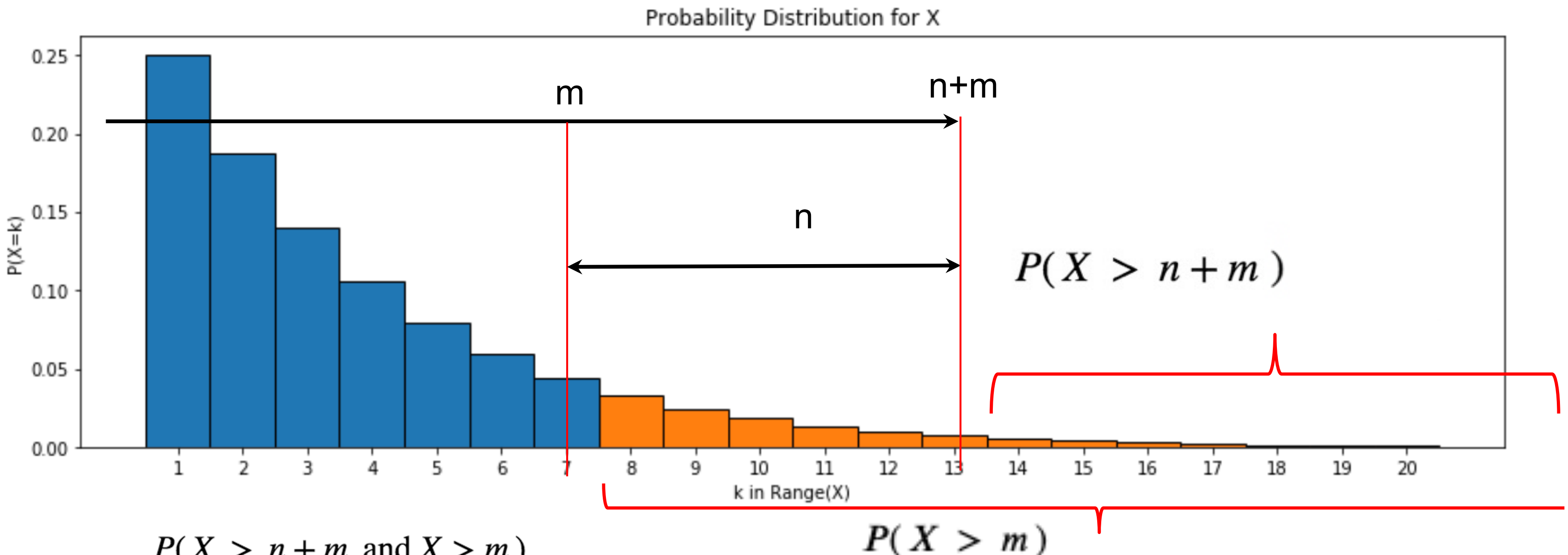
Example:

m = 7   n = 6
m+n = 13



(In fact, the Geometric is the only discrete distribution with this property; a continuous version of the Geometric, called the Exponential, is the other one.)

Memoryless Property: Version for CS 237



Proof:

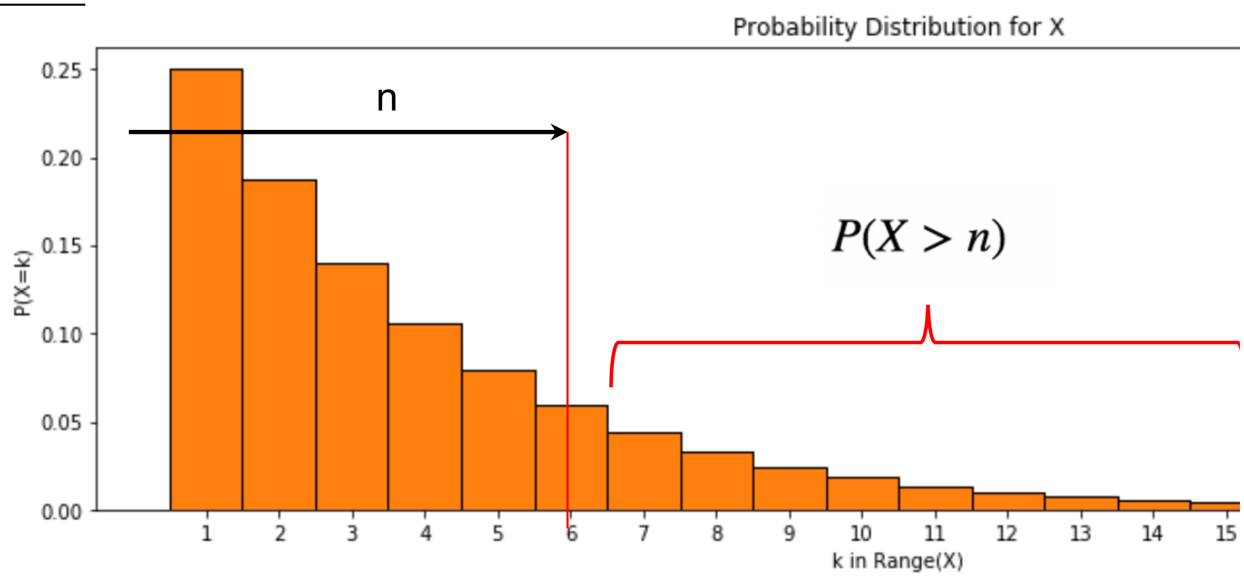
$$P(X > n+m \mid X > m) = \frac{P(X > n+m \text{ and } X > m)}{P(X > m)}$$

$$= \frac{P(X > n+m)}{P(X > m)}$$

$$= \frac{(1-p)^{(n+m)}}{(1-p)^{m}}$$

$$= (1-p)^{n}$$

$$= P(X > n)$$



Geometric Distribution: The Memoryless Property

Suppose in Park Street a train arrives every 15 minutes, and the probability that an arriving train is for the Green line is 1/5. Suppose you have just arrived at Park Street and are waiting for a Green Line train.

<u>What is the probability that you will have to wait at least 45 minutes (3 train arrivals) for a Green Line train?</u>

Consider these three situations:

(1) When you arrive, you just miss a Green Line train. How long to wait?

(2) When you arrive, you see a Red Line train leaving. How long to wait?

(3) Five trains have come and gone, one of them a Green Line train, but you missed it because you were reading your phone. How long to wait?

How are these different? BACK TO TOPHAT!

Geometric Distribution

- The memoryless property also simplifies the analysis of the Geometric
- $X \sim \text{Geometric}(p)$
- $E(X) = ?$
- $Var(X) = ?$

Geometric Distribution

Derivation using law of total expectation (aka case analysis)

On the board:

$$E(X) = E(X|X=1)\cdot \Pr(X=1) + E(X|X>1)\cdot \Pr(X>1)$$

$$= 1\cdot p + E(X|X>1)\cdot(1-p)$$

$$= p + (E(X)+1)\cdot(1-p)$$

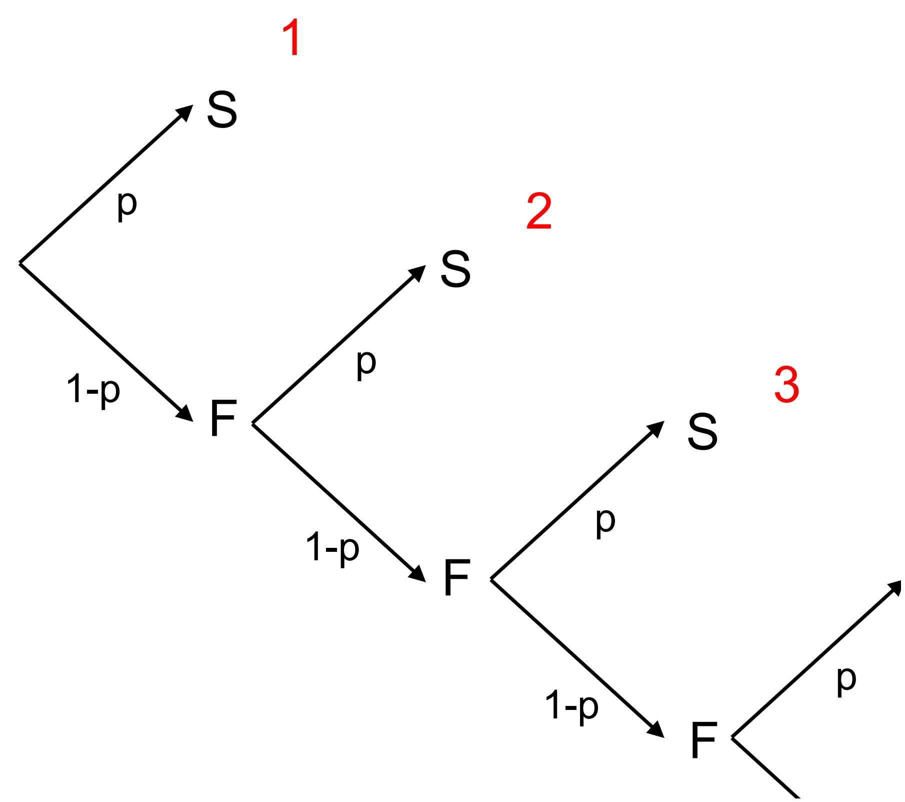
$$= p + E(X) - E(X)\cdot p + 1 - p$$

$$E(X) = E(X) - E(X)\cdot p + 1$$

$$0 = -E(X)\cdot p + 1$$

$$E(X)\cdot p = 1$$

$$E(X) = \frac{1}{p}$$



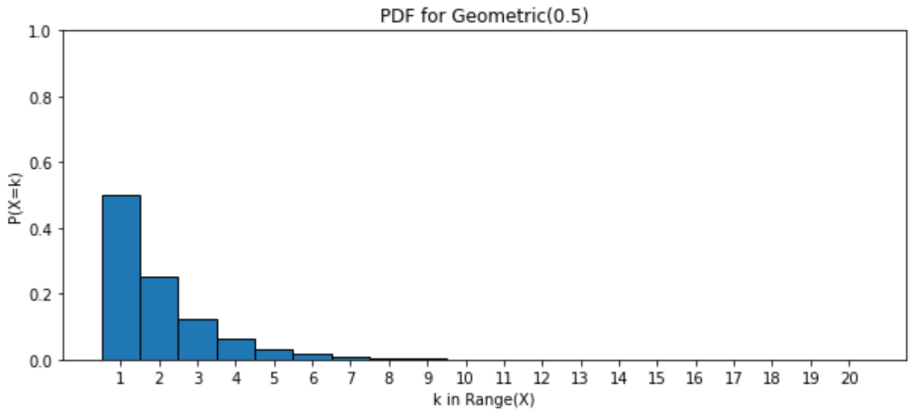
Geometric Distribution

- X ~ Geometric(p)

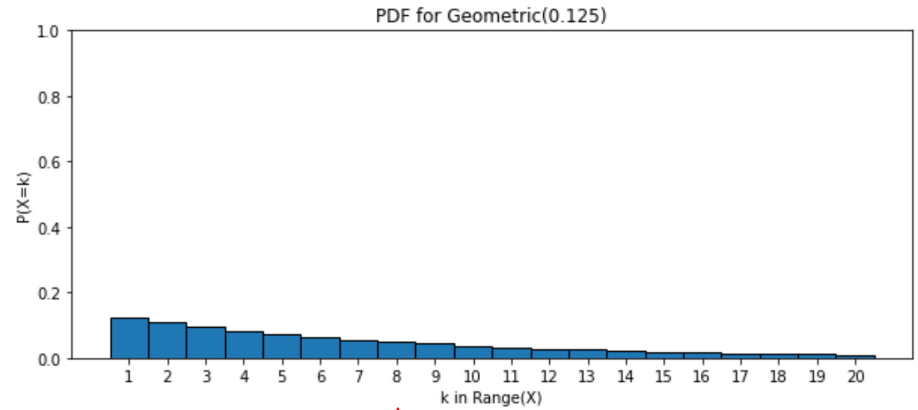
$$E(X) = \frac{1}{p}$$

$$Var(X) = \frac{1-p}{p^2}$$

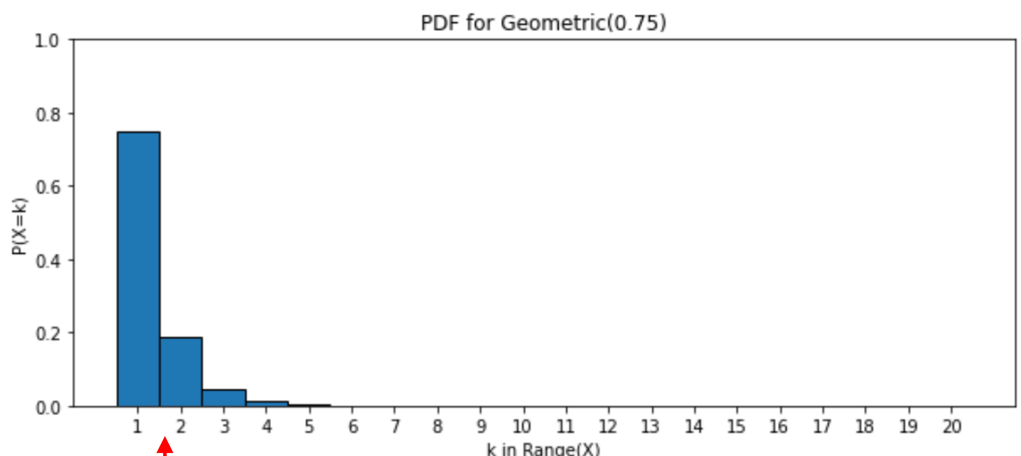
Geometric Distribution



$\mu = 2$



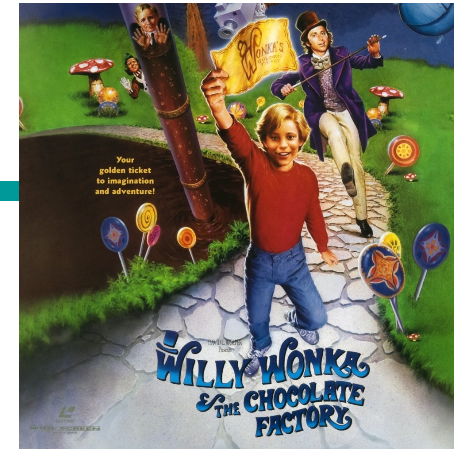
$\mu = 8$



$\mu = 4/3$

CS 237

Ask the audience



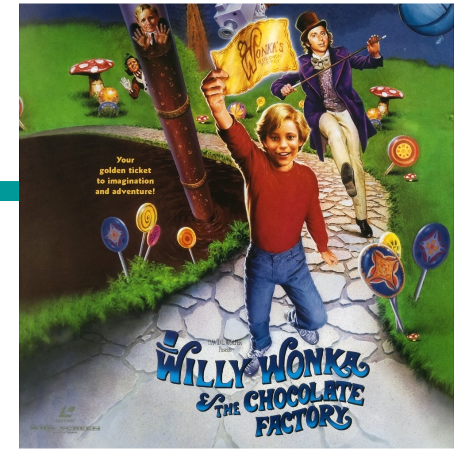
- Willy Wonka puts a golden ticket in each chocolate bar with probability 0.01, independently
- Charlie buys chocolate bars until he gets a golden ticket
- Let C = number of chocolates that Charlie buys

What is E(C)?

(A) 2 (B) 3 (C) 100 (D) none of the above

CS 237

Ask the audience



- Willy Wonka puts a golden ticket in each chocolate bar with probability 0.01, independently
- Charlie buys chocolate bars until he gets a golden ticket
- He's already bought 5 chocolates and no golden ticket.
- How many more bars does he need to buy in expectation?

(A) 2 (B) 3 (C) 100 (D) none of the above

Negative Binomial Distribution

A natural generalization of the Geometric is to ask how long before r >= 1 successes...

Formally, if Y ~ Bernoulli(p), and

X = "The number of trials of Y until the first r successes"

then we say that X is distributed according to the Negative Binomial Distribution with parameters r and p, and write this as:

$$X \sim \text{NegativeBinomial}(r, p)$$

Example: An oil company conducts a geological study that indicates that an exploratory oil well should have a 20% chance of striking oil. In order to be profitable, they must strike oil 5 times a year. What is the probability that the 5^{th} strike comes on the 30th well drilled?

Negative Binomial Distribution

X ~ NegativeBinomial(r, p)

$$\Pr(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

$$E(X) = \frac{r}{p}$$

$$Var(X) = \frac{r(1-p)}{p^2}$$

NOTE: Negative Binomial is the **sum** of independent Geometric RVs:

$X = X_1 + X_2 + \ldots + X_r$

Negative Binomial Distribution

On the board:

If it takes exactly k trials to get the rth success:

```
1                                              k
FFF...SFFF...SFFF...SFFF ....  FFF...SFFF.. S
      1      2      3                       r
```

then the probability of any specific sequences of k trials, with r successes and k-r failures, is

$$p^r (1-p)^{k-r}$$

Since the last trial is S, then we count the number of such sequences by choosing r-1 successes out of k-1 trials, yielding:

$$\Pr(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

CS 237

Structure of Discrete Distributions

	Single		Sum Of
Counting:	Bernoulli(p)	$\Rightarrow$	Binomial(N,p)
	$\Downarrow$		
Waiting:	Geometric(p)	$\Rightarrow$	NegativeBinomial(r,p)