Discrete Random Variables

MAT 152 – STATISTICAL METHODS I
LECTURE 2
INSTRUCTOR: DUSTIN ROTEN
FALL 2020

The Binomial Distribution

Characteristics of a binomial distribution:

1. There are a fixed number of trials/repetitions
2. There are two possible outcomes: successes/failures

(The probability of a "success" is *p.* The probability of a "failure" is q. So, $p + q = 1$.)

Each trial is called a **Bernoulli Trial.** (Mathematician from the 1600's)

Furthermore, the trials must be independent and repeated in the same way every time.

Examples:

Tossing a coin, (heads = "success", tails = "failure")

Probability of students submitting homework (submission = "success", tails = "failure")

The Binomial Distribution

The outcomes of a binomial experiment fit a **binomial probability distribution.**

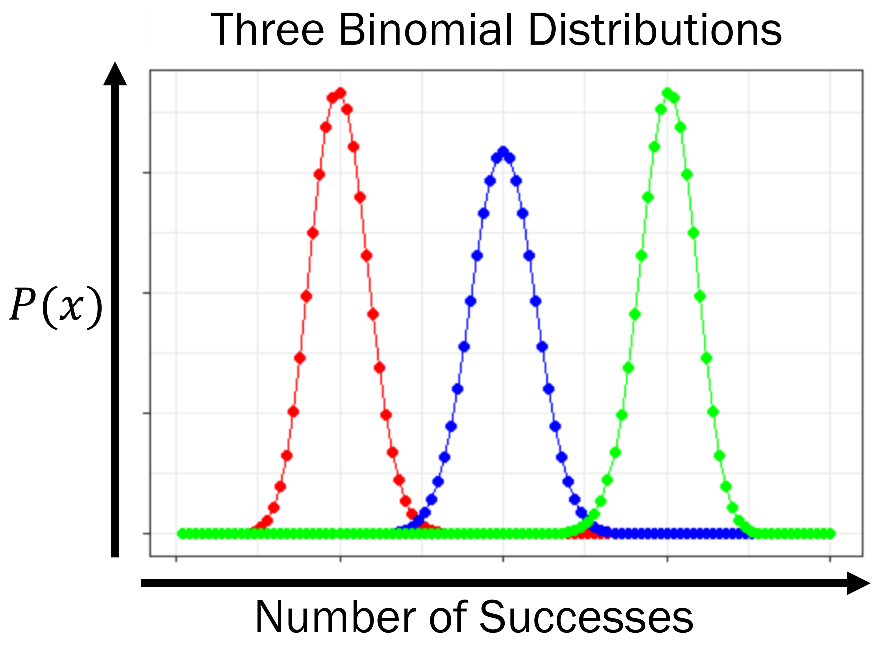
X = the number of successes obtained in n independent trials

mean: $\mu = np$

variance: $\sigma^2 = npq$

standard deviation: $\sigma = \sqrt{npq}$

The probability of success *(p)* determines the peak of a binomial distribution



Example (1)

Case #1: Flipping an unfair coin 100 times with only a 25% chance of success (heads). $(n = 100; p = 0.25; q = 0.75)$

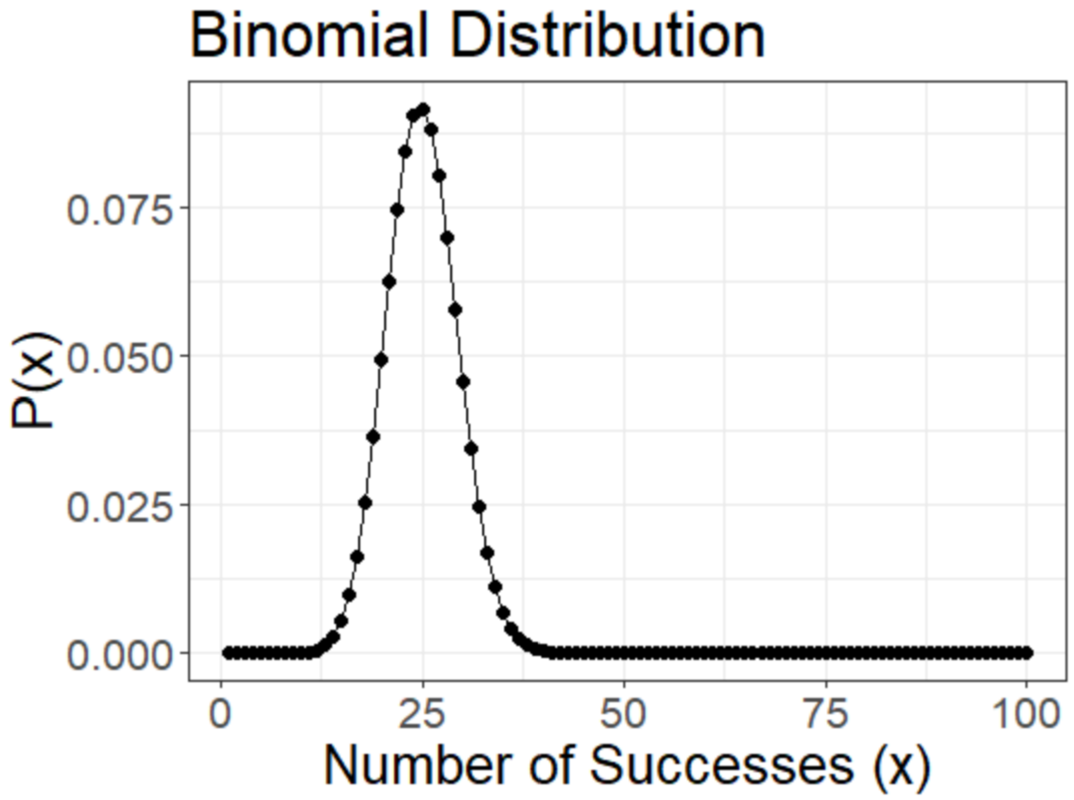
What is the expected number of successes?

$\mu = np = 100 \cdot 0.25 = 25$

25 successes are expected.

What is the standard deviation of the outcomes' probabilities?

$\sigma = \sqrt{\sigma^2} = \sqrt{100 \cdot 0.25 \cdot 0.75} \approx 4.33$



Example (2)

Case #2: Flipping a fair coin 100 times with a 50% chance of success (heads). $(n = 100; p = 0.5; q = 0.5)$

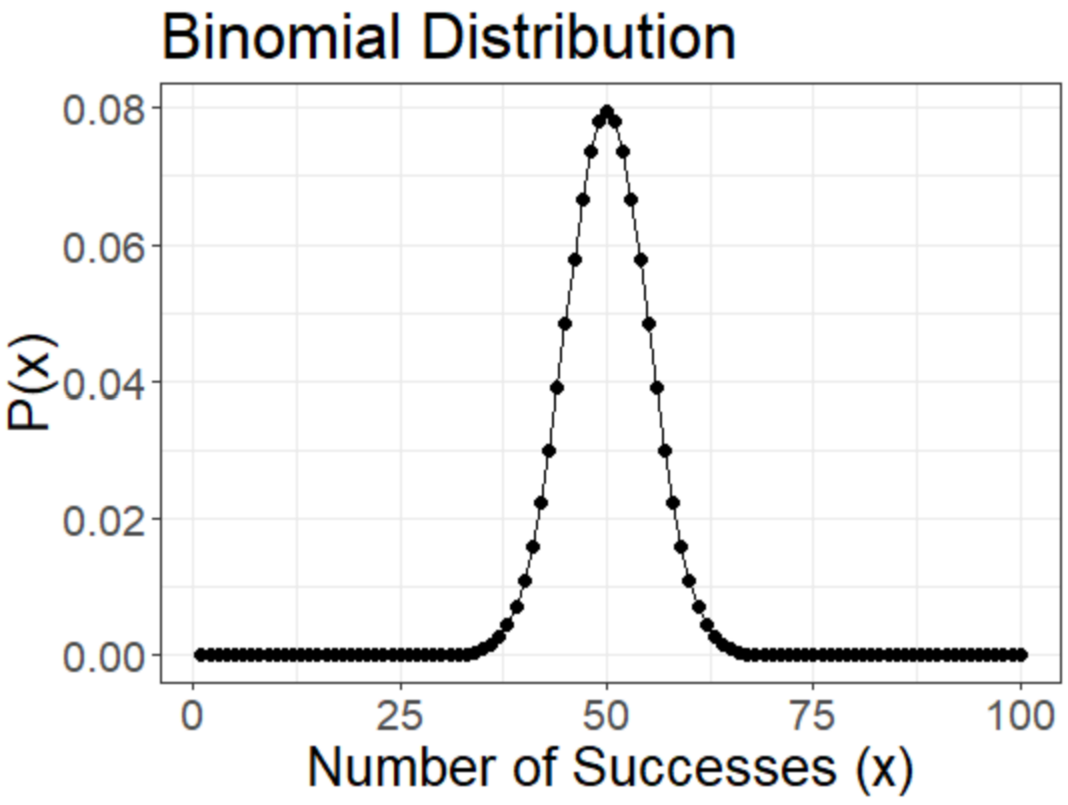
What is the expected number of successes?

$\mu = np = 100 \cdot 0.5 = 50$

50 successes are expected.

What is the standard deviation of the outcomes' probabilities?

$\sigma = \sqrt{\sigma^2} = \sqrt{100 \cdot 0.5 \cdot 0.5} \approx 5$



Example (3)

Case #3: Flipping an unfair coin 100 times with a 75% chance of success (heads). ($n = 100; p = 0.75; q = 0.25$)

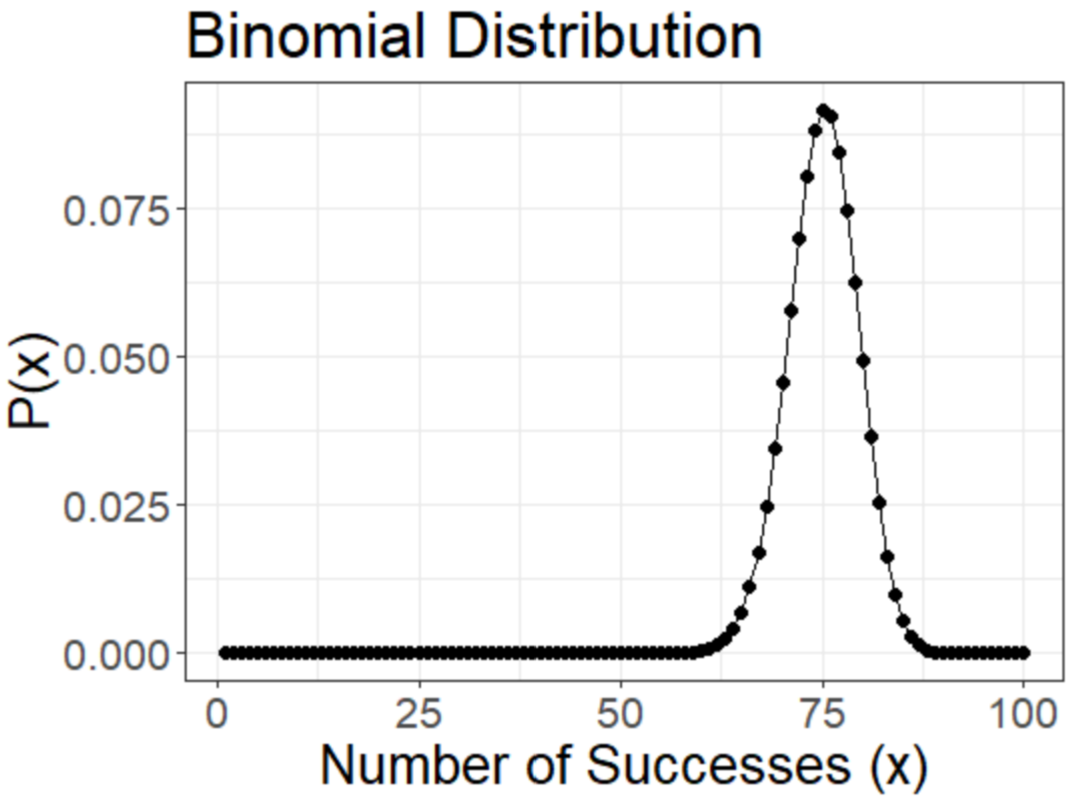
What is the expected number of successes?

$\mu = np = 100 \cdot 0.75 = 75$

75 successes are expected.

What is the standard deviation of the outcomes' probabilities?

$\sigma = \sqrt{\sigma^2} = \sqrt{100 \cdot 0.75 \cdot 0.25} \approx 4.33$



Binomial Distribution Notation

$X \sim B(n, p)$
"X is a random variable with a binomial distribution."
$n = \text{number of trials}$
$p = \text{probability of a success}$

We can calculate μ, σ^2, and σ. But what if we want to know the probability of a specific outcome?

How can specific outcomes like $P(x = 5)$ or $P(x \geq 5)$ be calculated?
(Technology may be needed! Microsoft Excel, R, Online Tools)

Calculating by Hand

Suppose you toss a coin three times.
Let $X =$ the number of heads that occurs. So, $x = 0, 1, 2, 3.$
What is the probability of two heads occurring?
$P(x = 2) = ?$

First, determine the total number of outcomes.
(Each coin has two options: heads or tails)
$|S| = 2 \times 2 \times 2 = 2^3 = 8$

Next, how many ways are there to get 2 heads?
$\{HHT, HTH, THH\}$

What is the probability of one of these event occurring?
$P(x = 2) = \frac{\text{outcomes of interest}}{\text{all possible outcomes}} = \frac{3}{8}$

Finding possible outcomes quickly using combinations: this is called "*n* choose *x*".

$\binom{n}{x} = \frac{n!}{x!(n-x)!}$

$\binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{3!}{2! \cdot 1!} = \frac{3 \cdot 2 \cdot 1}{(2 \cdot 1) \cdot (1)} = \frac{6}{2} = 3$

Now, formalize the probability: 2 successes and 1 failure.

$p^x \cdot q^{n-x}$

$\left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{3-2} = \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^1 = \frac{1}{8}$ (But there are 3 of them!)

Putting it all together:

$P(x = k) = \binom{n}{k} p^k q^{n-k}$

Example

A fair coin is flipped 15 times. Each flip is independent.

What is the probability of getting 6 heads?

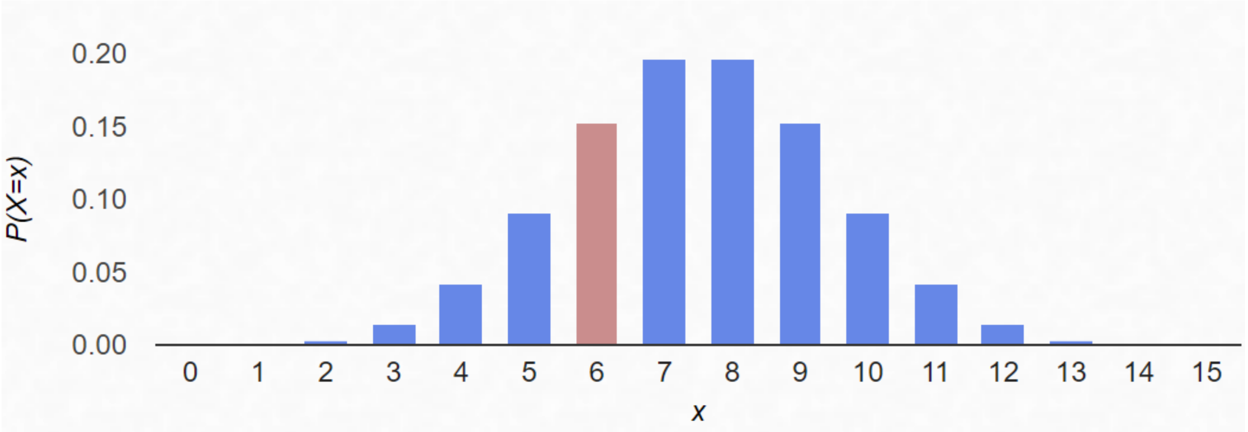
$P(x = 6) = 0.15274$ or 15.274%

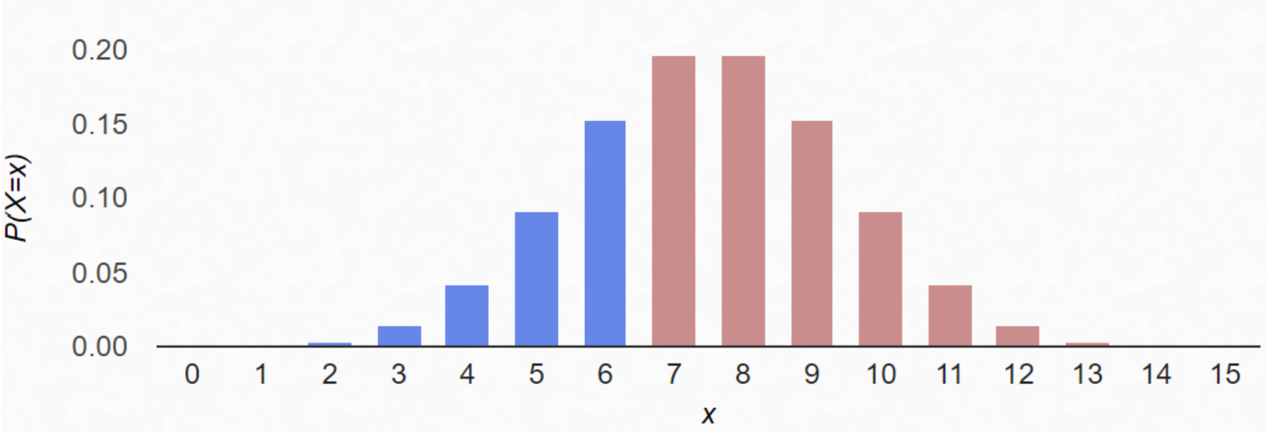
$$P(x = 6) = \binom{15}{6}(0.5)^6(0.5)^{15-6} = 0.15274$$

What is the probability of getting 7 or more heads?

$P(x \geq 7) = 0.69638$ or 69.638%

$$P(x \geq 7) = 1 - P(x \leq 6) = 1 - \sum_{i=1}^{6}\binom{15}{i}(0.5)^i(0.5)^{15-i}$$





Probabilities to be Calculated

For $X \sim B(n,p)$, consider the value $x = c$.

Possible values to be calculated:

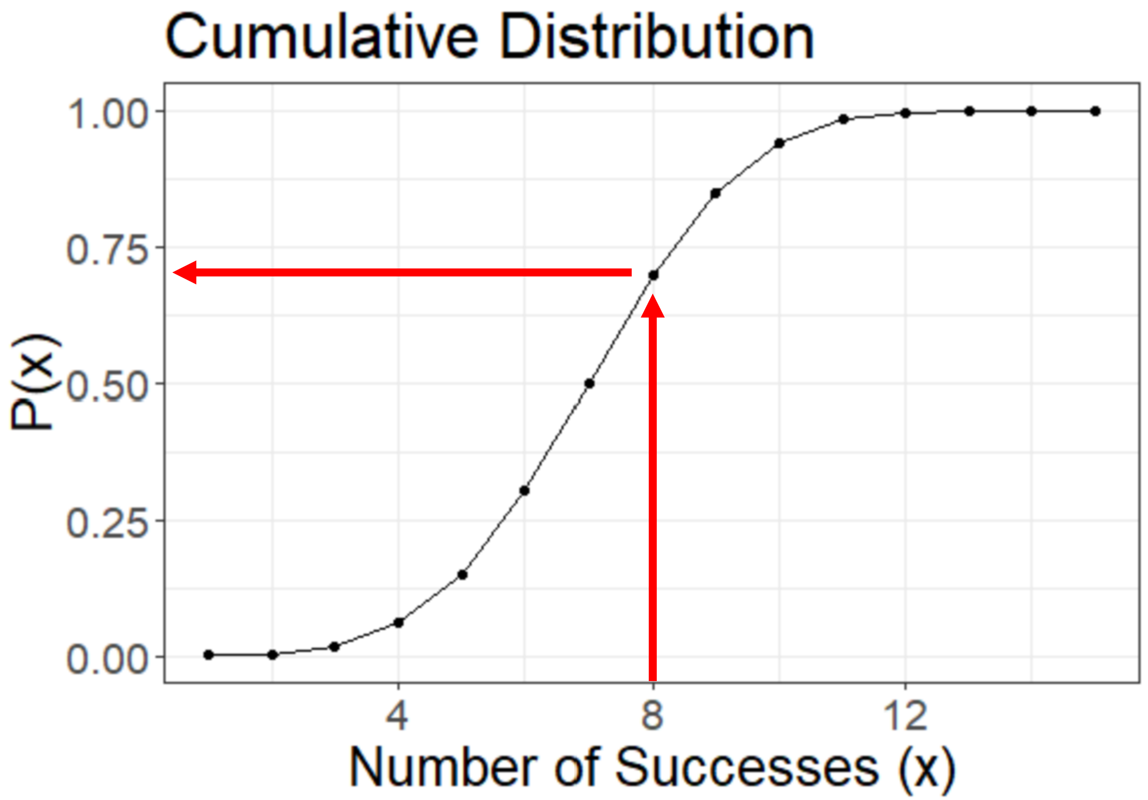
Probabilities	In words
$P(x = c)$	Probability of getting *c* successful outcomes
$P(x > c)$	Probability of getting more than *c* successful outcomes
$P(x < c)$	Probability of getting less than *c* successful outcomes
$P(x \geq c)$	Probability of getting at least *c* successful outcomes
$P(x \leq c)$	Probability of getting at most *c* successful outcomes

Cumulative Distribution

A fair coin is flipped 15 times. Each flip is independent.

What is the probability of at most 8 heads?
(What is the probability of a 1, 2, 3, 4, 5, 6, 7, or 8)
$P(x \leq 8) = 0.6963806 \approx 69.64\%$

What is the probability of more than 8 heads?
$P(x > 8) = ?$
$1 - P(x \leq 8) = 1 - 0.6963806 = 0.3036194 \approx 30.36\%$



Special Case: Geometric Distribution

The probability of rolling a three when a fair dice is thrown is $\frac{1}{6}$. (Multiple rolls are independent!)

What is the probability of not rolling a three until the 5th roll?

$X = \text{the number of rolls until a 3 is rolled}$

$$p = \frac{1}{6}; q = \frac{5}{6}$$

Number of trials:	$P(X = k) = q^{k-1} \cdot p$
Number of failures:	$P(Y = k) = q^k \cdot p$

$$P(x = 5) = ?$$

$$P(x = 5) = \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = 0.0804$$

Geometric Distribution

Three main characteristics of a geometric distribution:

1. There are one or more Bernoulli Trials with all failures except the last one.
2. The number of trials could go on forever.
3. The probability of a success (and failure) are the same for each trial.

$X \sim G(p)$; "X is a random variable with a geometric distribution"

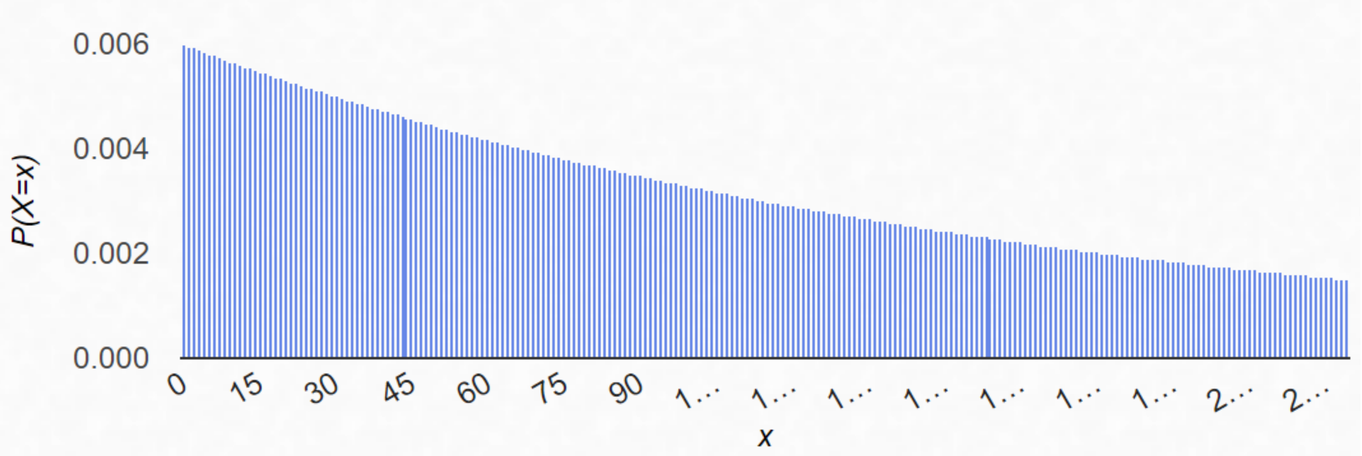
mean: $\mu = \frac{1}{p}$

variance: $\sigma^2 = \left(\frac{1}{p}\right)\left(\frac{1}{p} - 1\right)$

standard deviation: $\sigma = \sqrt{\left(\frac{1}{p}\right)\left(\frac{1}{p} - 1\right)}$

Example

Only 0.6% of the population has blood type AB-. Assuming that blood types are “well mixed” within the United States, about how many people would you expect to test before an AB- blood type is found?



Expected value: $\mu = \frac{1}{p} = \frac{1}{0.006} \approx 167$

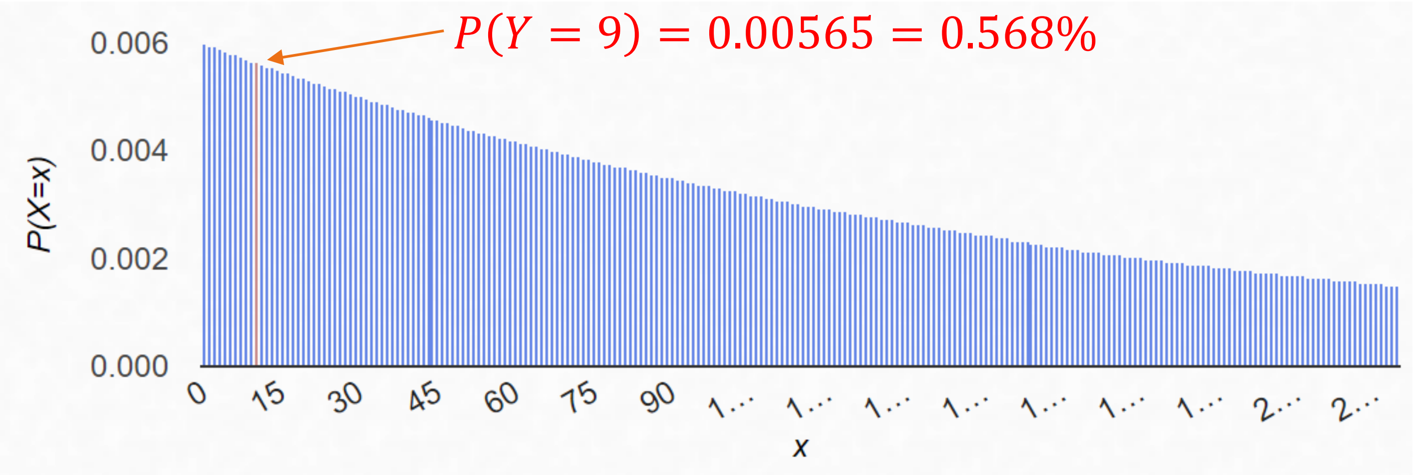
What is the standard deviation?

$$\sigma = \sqrt{\sigma^2} = \sqrt{\left(\frac{1}{p}\right)\left(\frac{1}{p} - 1\right)} = \sqrt{\left(\frac{1}{0.006}\right)\left(\frac{1}{0.006} - 1\right)} = 166.2$$

Example (cont.)

Only 0.6% of the population has blood type AB-. What is the probability of the 10th test revealing a blood type of AB-?

Given $X \sim G(0.006)$, find $P(X = 10)$.



A Quick Review

Binomial Distribution: $X \sim B(n, p)$

- There are a fixed number of trials/repetitions
- There are two possible outcomes: successes/failures
- Each independent experiment is called a Bernoulli Trial
- mean: $\mu = np$; standard deviation: $\sigma = \sqrt{npq}$

Geometric Distribution: $X \sim G(p)$

- There are one or more independent Bernoulli Trials with all failures except the last one.
- The number of trials is NOT fixed (they can go on forever).
- mean: $\mu = \frac{1}{p}$; standard deviation: $\sigma = \sqrt{\left(\frac{1}{p}\right)\left(\frac{1}{p} - 1\right)}$