Discrete Distributions

1. What is Discrete Distributions?

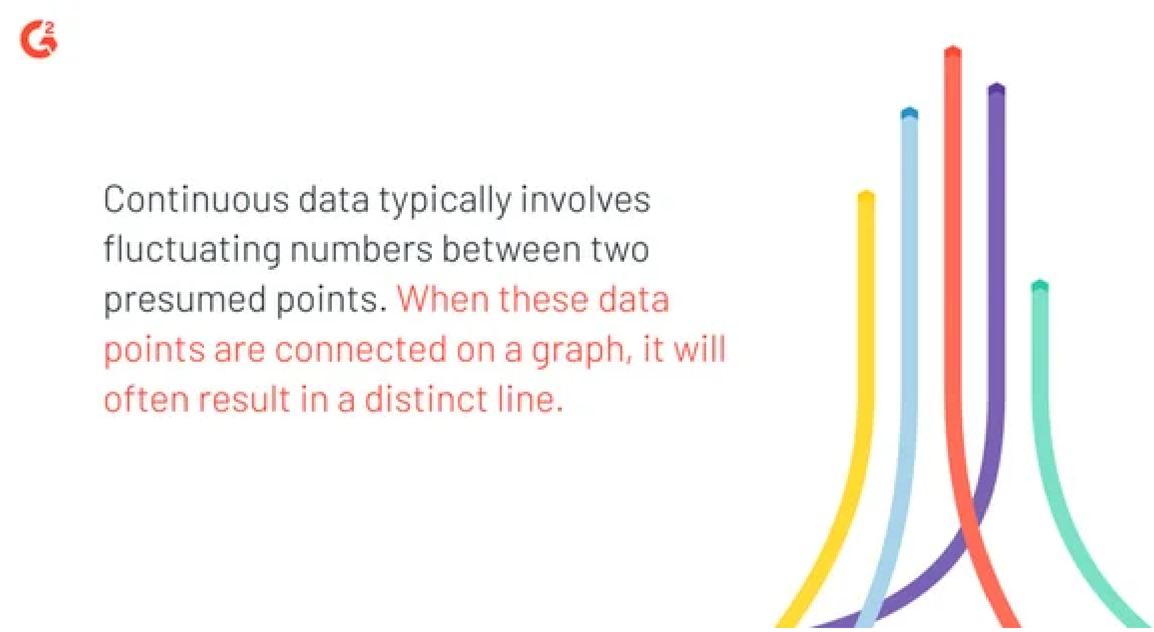
1.1. Discrete Data

Discrete data refers to countable, individualized, and nondivisible figures in statistics. These data points exist only in set increments.

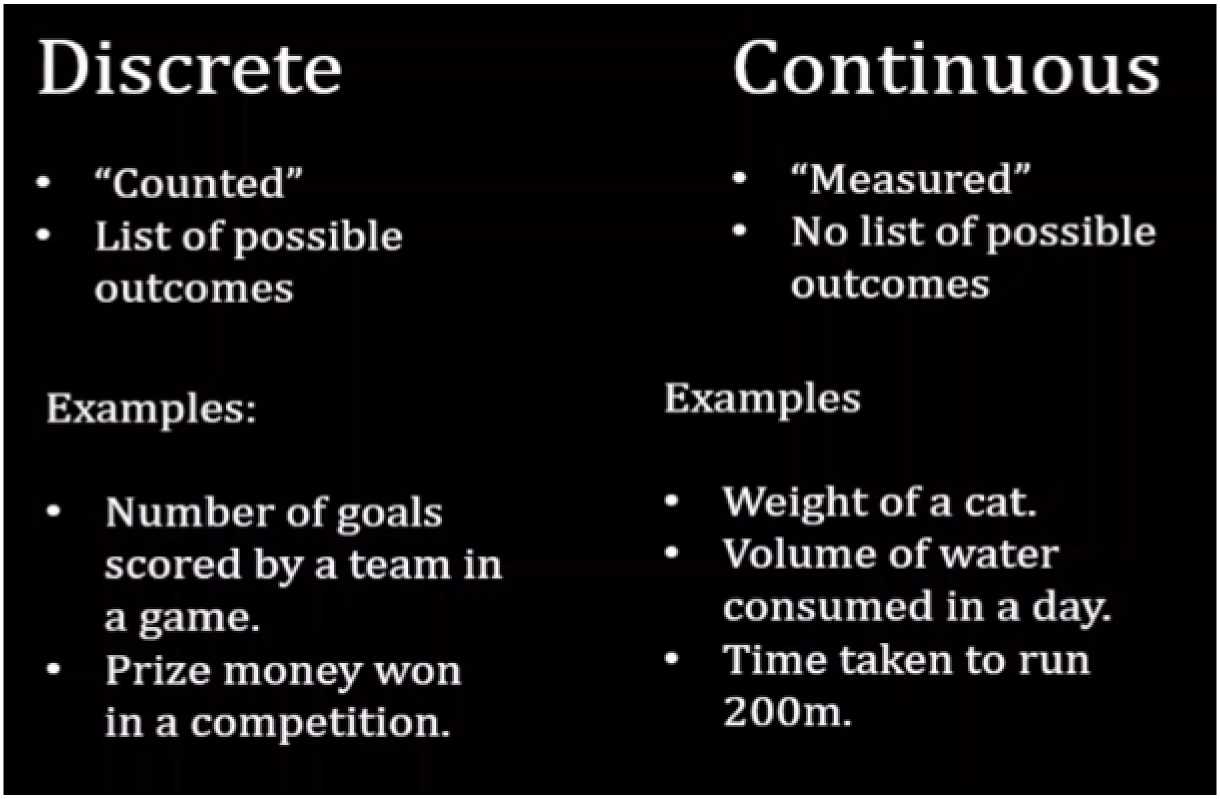
Discrete data represents discrete variables, which you can count in a finite amount of time. The key feature here is that these variables are countable instead of measurable.

1.2. Continuous Data

Continuous data is a type of numerical data that refers to the unspecified number of possible measurements between two realistic points.

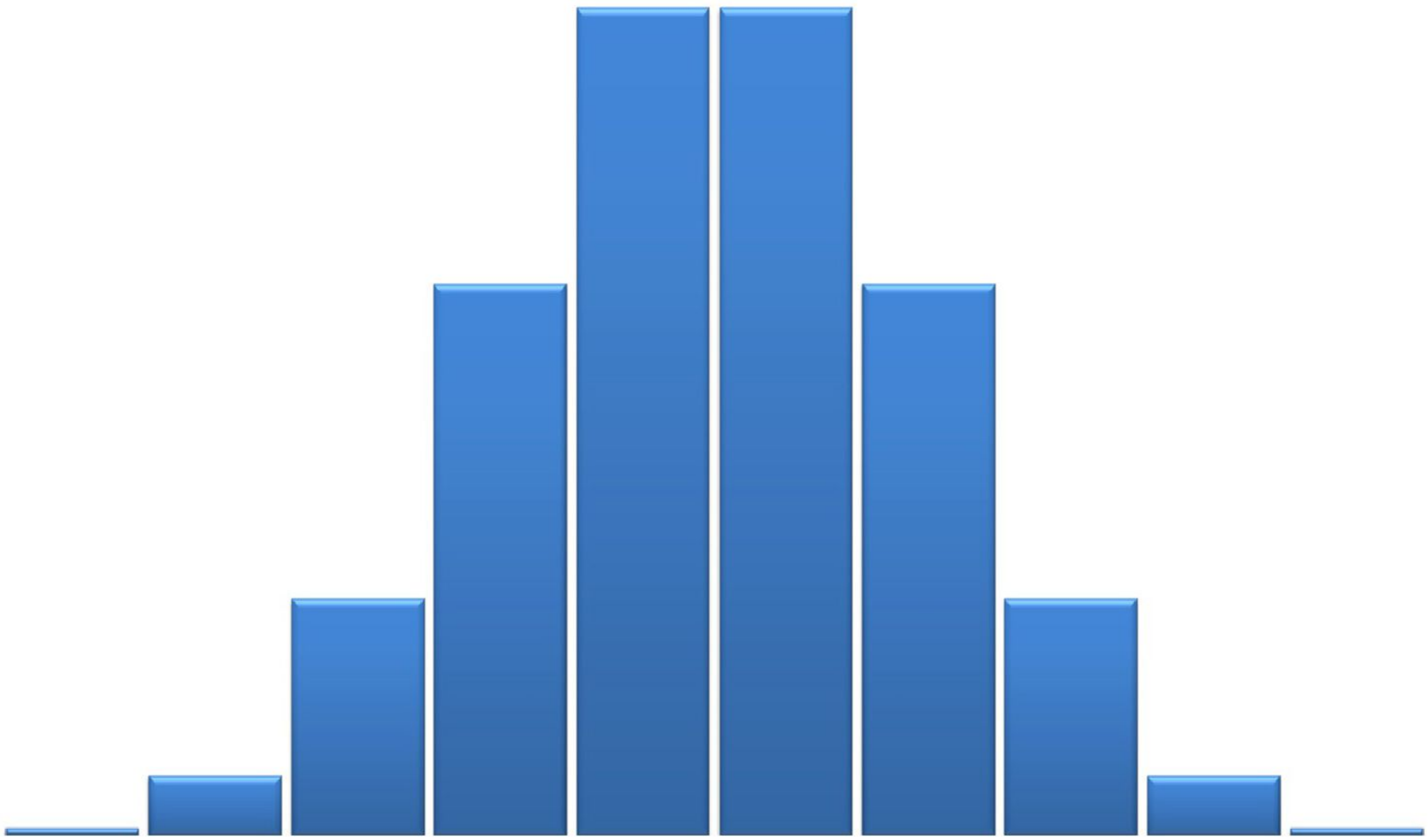


1.3. Discrete Data vs Continuous Data



1.4. Discrete Distribution

A **discrete distribution** is a probability distribution that depicts the occurrence of discrete (individually countable) outcomes, such as 1, 2, 3, yes, no, true, or false



2. The Expectation and Variance of a Discrete Random Variable

2.1. Expectation

The **expected value** of random variable X is often written as E(X) or μ or μX.

For a discrete random variable **the expected value** is calculated by summing the product of the value of the random variable and its associated probability, taken over all of the values of the random variable.

$$\mu = E(X) = \sum x_i f(x_i)$$

The formula means that we multiply each value, x, in the support by its respective probability, $f(x)$, and then add them all together. It can be seen as an average value but weighted by the likelihood of the value.

2.2. Variance

$$\sigma^2 = \mathrm{Var}(X) = \sum (x_i - \mu)^2 f(x_i)$$

The formula means that we take each value of x, subtract the expected value, square that value and multiply that value by its probability. Then sum all of those values.

There is an easier form of this formula we can use.

$$\sigma^2 = \mathrm{Var}(X) = \sum x_i^2 f(x_i) - E(X)^2 = \sum x_i^2 f(x_i) - \mu^2$$

The formula means that first, we sum the square of each value times its probability then subtract the square of the mean. We will use this form of the formula in all of our examples.

Note

	Discrete Random Variable	Continuous Random Variable
Mean (Expected Value)	$\mu = E(X) = \sum_{i=1}^{n} x f(x)$	$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$
Variance	$\sigma^2 = V(X) = \sum_{i=1}^{n} (x-\mu)^2 f(x)$	$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$
Standard Deviation (Standard Error)	$\sigma = \sqrt{\sigma^2}$	$\sigma = \sqrt{\sigma^2}$

Calcworkshop.com

3. Binominal Distribution

3.1. Definition

Binomial distribution is a statistical distribution that summarizes the probability that a value will take one of two independent values under a given set of parameters or assumptions.

1. Used in situations when you can count the number of successes given a fixed number of trials.

e.g.

- How many times a darts player hits the bullseye in 10 attempts.
- How 6's are rolled when a dice is rolled 30 times.

But not:

- How many goals a player scores in a game.
- How many taxis pass you by until one stops to pick you up.

2. Used in situations where there are just two possible outcomes.

e.g.

- How many customers buy your product out of the 40 that entered the store.
- How many 6's are rolled when a dice is rolled 30 times.

But not:

- How shots it takes for a golfer to complete an 18 hole course.
- How much money you have spent after making 5 purchases.

3. Used in situations where the success of each trial is independent of all other trials.

e.g.

- The number tails from flipping a coin 12 times.
- The number of people who get skin cancer in a medical trial of 1000 people.

But not:

- The number of days it rains in the next 2 weeks.
- The number of games a team wins in the next 6 games.

3.2. Formula

Binomial Distribution Formula

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

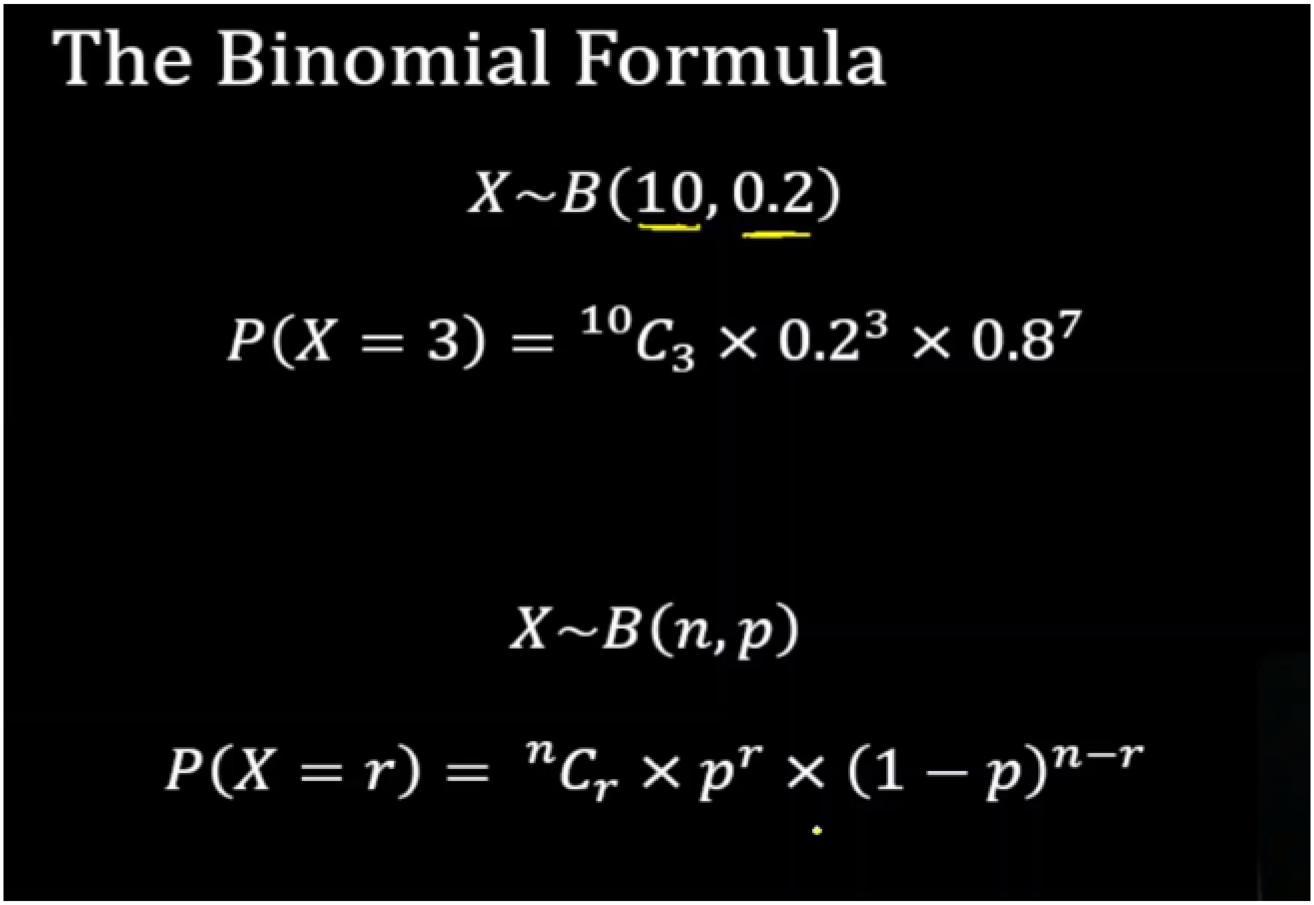
where

n = the number of trials (or the number being sampled)

x = the number of successes desired

p = probability of getting a success in one trial

q = 1 - p = the probability of getting a failure in one trial



4. Poisson Distribution

4.1. Definition

In statistics, a **Poisson distribution** is a probability distribution that is used to show how many times an event is likely to occur over a specified period.

Poisson distributions are often used to understand independent events that occur at a constant rate within a given interval of time.

Conditions Required for The Poisson Dist

To model a situation with the Poisson Distribution, events must occur:

- At a constant average rate
- Independently of one another
- Singly (cannot occur simultaneously)

4.2. Formula

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Example

The Poisson Distribution

$X \sim Po(\lambda)$

$$P(X = r) = \frac{e^{-\lambda} \times \lambda^r}{r!}$$

A football team scores goals at a rate of 2.3 per game. Find the probability that they score 1 goals in their next game.

$G \sim Po(2.3)$

$$P(G = 1) = \frac{e^{-2.3} \times 2.3^1}{1!}$$

In a particular town it is known that there are, on average, 4 wrongful convictions per month. Each of these convictions is independent of the others, and they occur randomly.

Find the probability that there will be 6 wrongful convictions next month.

$$P(X = 6) = \frac{e^{-4} \times 4^6}{6!}$$

$$= 0.1042$$

5. Geometric Distribution

5.1. Definition

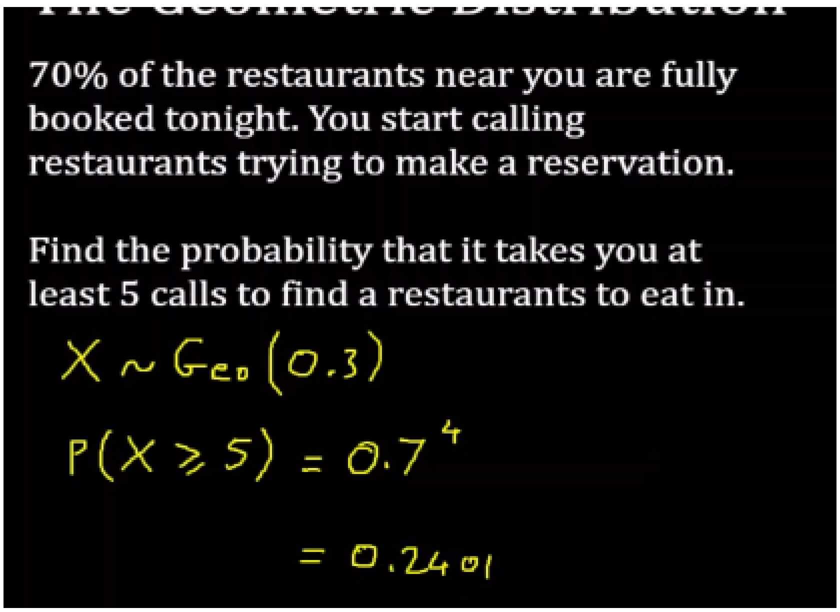
The **geometric distribution** represents the number of failures before you get a success in a series of Bernoulli trials.

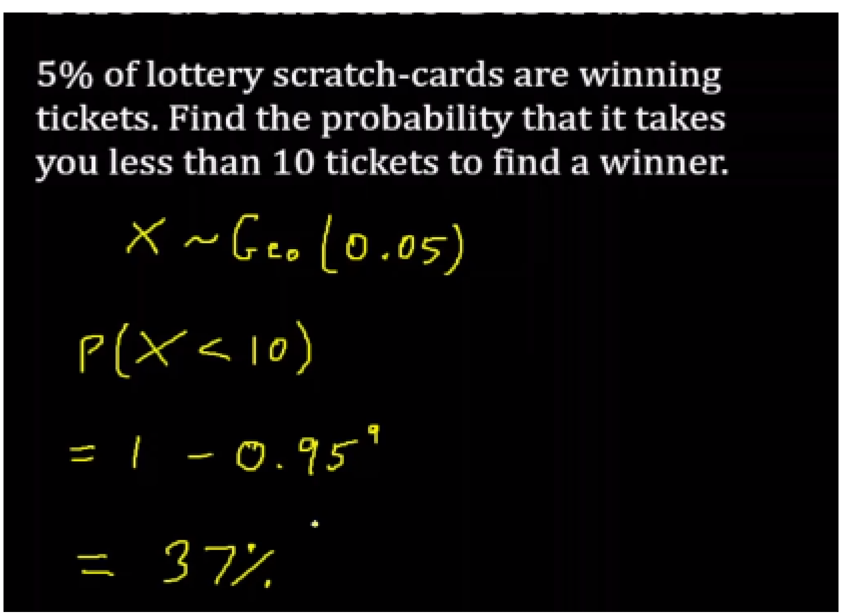
A **geometric distribution** can have an indefinite number of trials until the first success is obtained

5.2. Formula

Alt text

Example

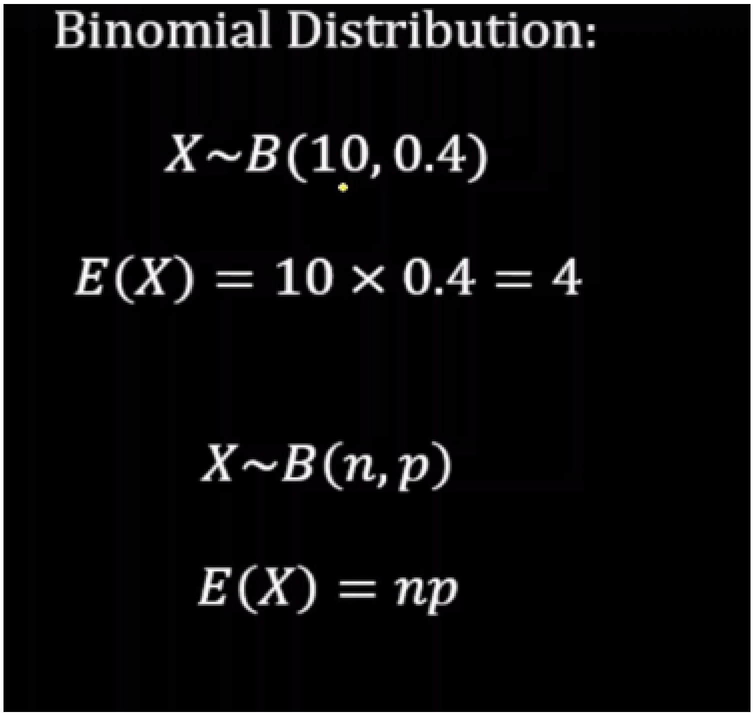




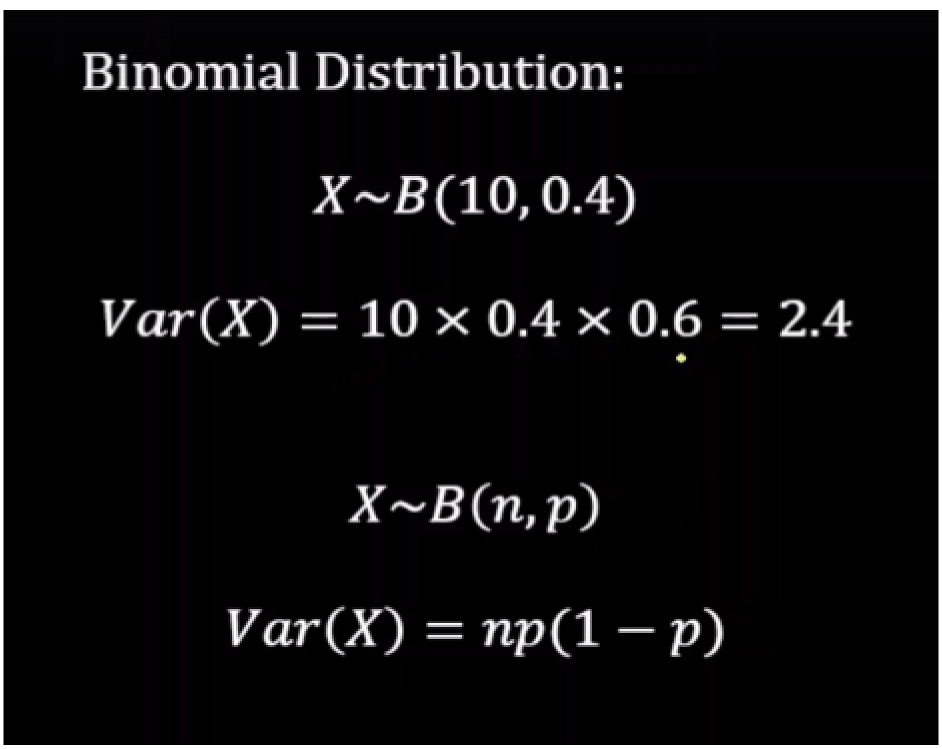
6. Expectation and Variance of Distributions

6.1. Binominal Distribution

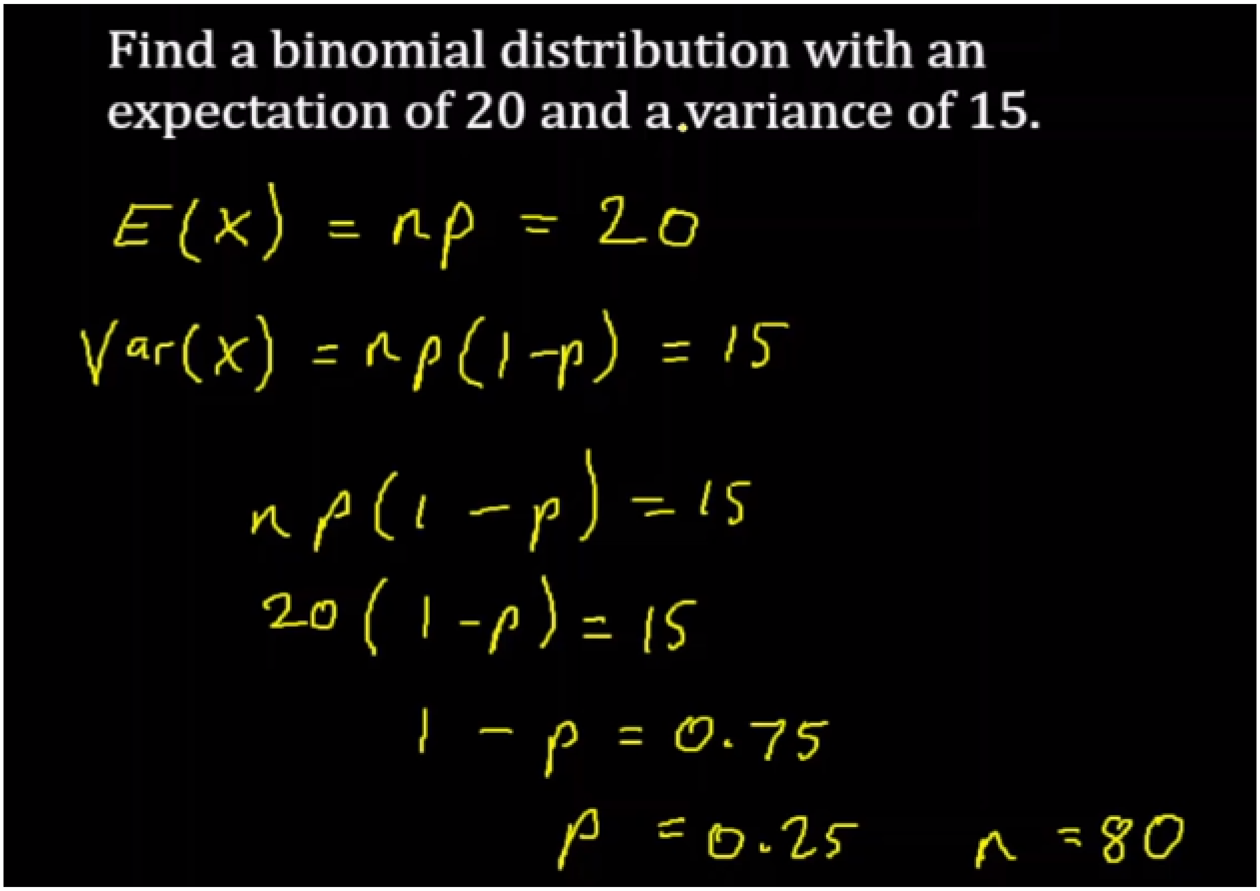
Expectation



Variance

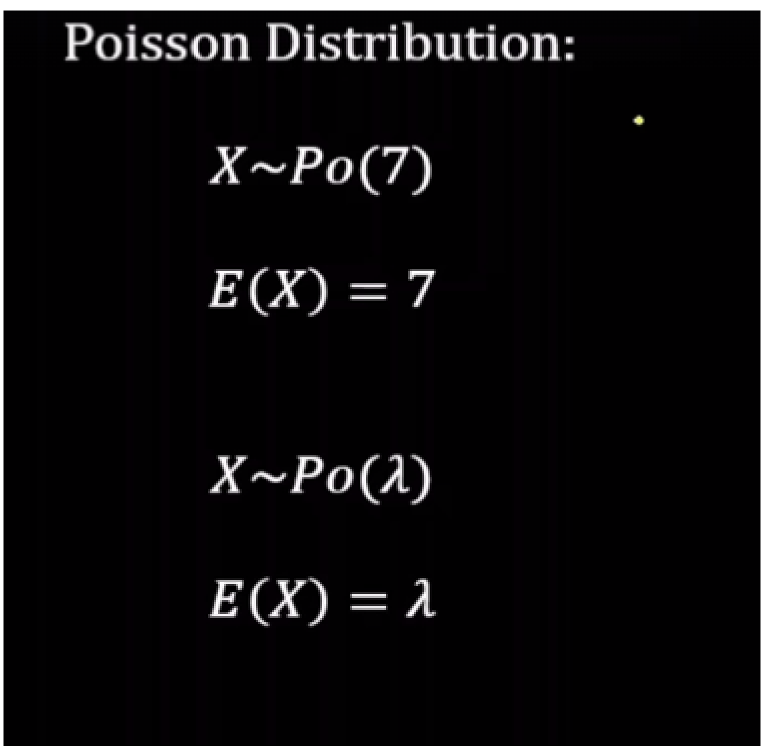


Example

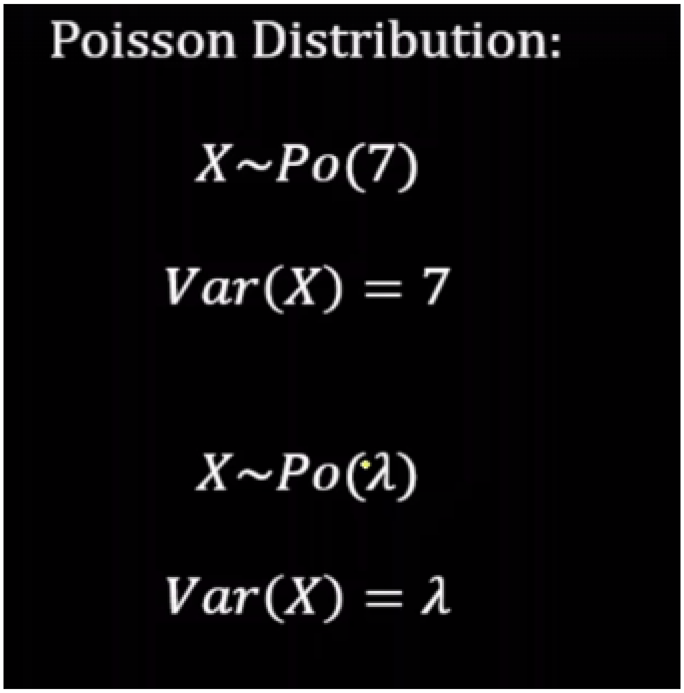


6.2. Poisson Distribution

Expectation



Variance



7. Approximations

The **Poisson distribution** is actually a limiting case of a **Binomial distribution** when the number of trials, n, gets very large and p, the probability of success, is small.

As a rule of thumb, `if n ≥ 100 and np ≤ 10`, the **Poisson distribution** (taking `λ=np`) can provide a very good approximation to the **Binomial distribution**.