Probability Distributions

#CMSC320 #probability_statistics #M1

> ✎ **Definition**
>
> Distributions refer to the probability of occurrence for all possible outcomes. Distributions can be graphed. The kind of distribution associated with a set of events can tell us about the properties the events have and how to appropriately analyze them. (Note: the sum of all probabilities over an entire distribution is one)

Types of Distributions

### Know Well

Uniform

- **What it is**
    - Every possible outcome has the same probability (Can use integral to get sum of all probabilities).
- **Discrete vs. Continouos**
    - Continuous
- **What it looks like**
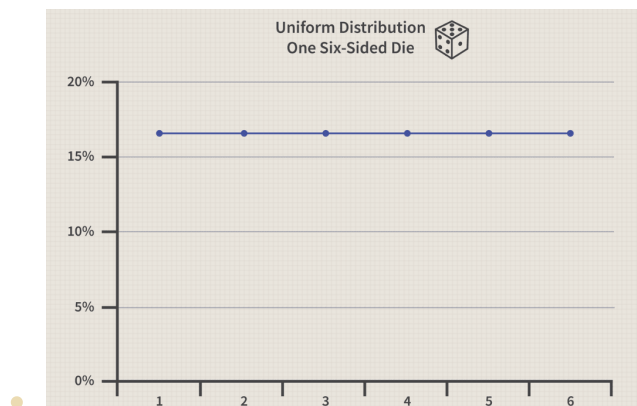
- **Examples**
   - Drawing a 6 from a deck of cards.

Bernoulli

- **What it is**
  - Distribution of probability for an event which takes on the value of zero or one (binary). Does not literally have to be zero or one, just means that the even has only two possible outcomes.
- **Discrete vs. Continouos**
  - Discrete
- **What it looks like**
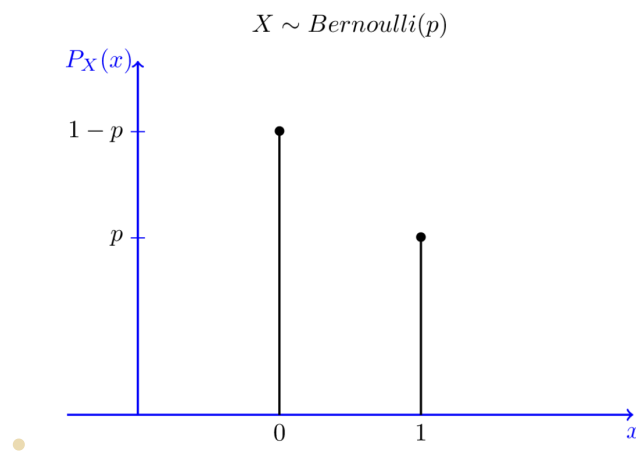
$$X \sim Bernoulli(p)$$

-
- **Examples**
  - Coin flip

Binomial

- **What it is**
  - The number of successes (or has the value one) in *n* independent Bernoulli trials.
  - Note: Bernoulli relates to distribution of single trial of event, binomial refers to multiple trials
- **Discrete vs. Continouos**
  - Discrete
- **What it looks like**
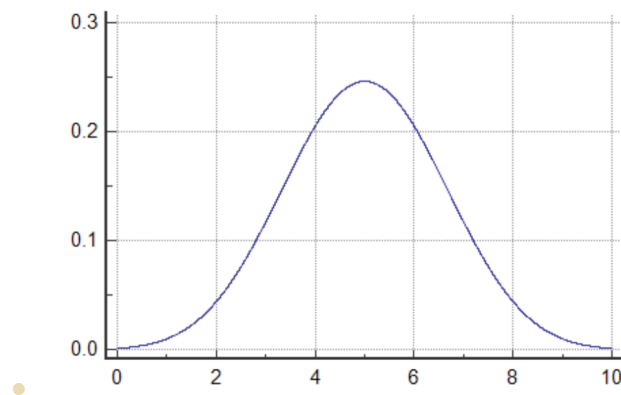
- **Examples**
    - Flipping coin 100 times.

Normal

- **What it is**
    - In this distribution, the probability of events cluster towards the center of all possible events and taper off towards the ends.
    - Note: shape of binomial and normal distribution are similar, but normal is continuous and binomial is discrete
- **Discrete vs. Continouos**
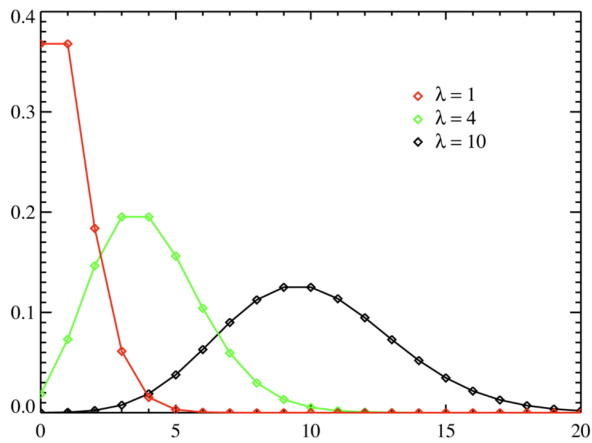   - Continuous
- **What it looks like**
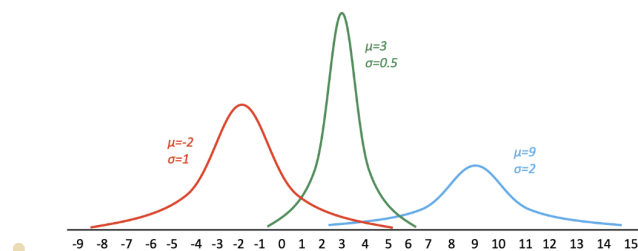
- **Examples**
    - Population height

Poisson

- **What it is**
    - Describes how many times an event will occur in a specified time frame.
- **Discrete vs. Continuous**
    - Discrete

- **What it looks like**

- **Examples**
  - Number of customers entering a strore every fifteen minutes
- **Note**
  - Zero-inflated poisson is a poisson distribution with a high number of zeroes. In this case the zeroes can be ignored.

Less Common

Power Law

- **What it is**
  - In this distribution a small set of events have a large probability while the remaining set of events have a small probability
- **Discrete vs. Continuous**
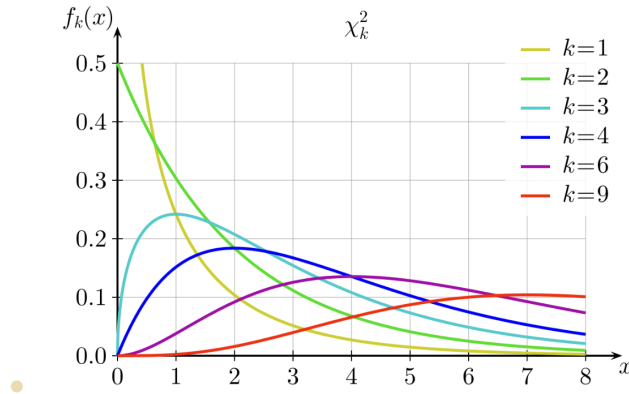 - Continuous
- **What it looks like**

- **Examples**
 - Wealth distribution in the U.S.

Chi-Squared

- **What it is**
  - Distribution of sum of squared random variables. Used when there are many categorical variables.
- **Discrete vs. Continuous**
  - Continuous
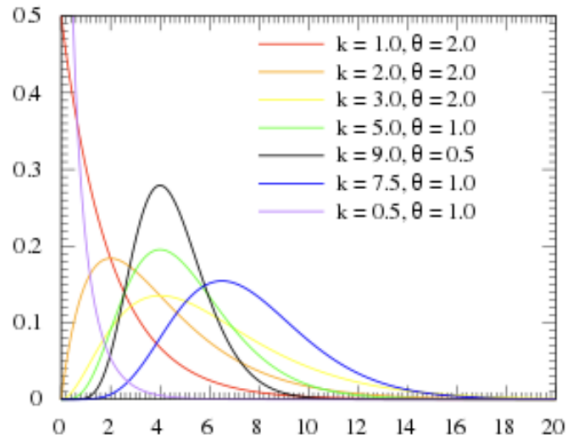- **What it looks like**

-

Exponential

- **What it is**
  - Usually concerns that amount of time until a specific event happens. (can think of poisson as having exponential between each event)
- **Discrete vs. Continuous**
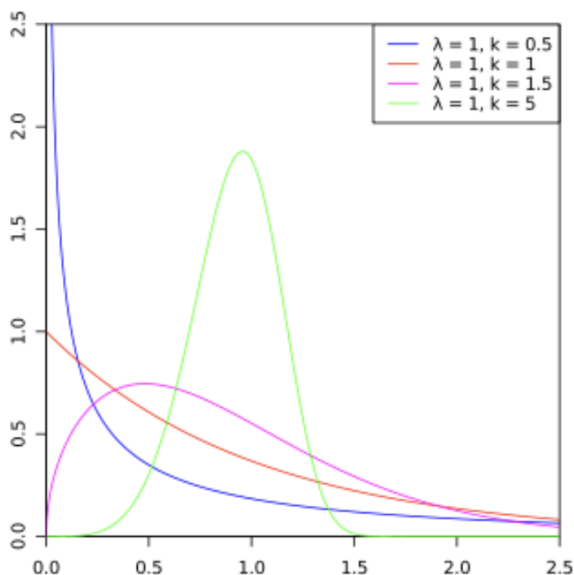 - Continuous
- **What it looks like**

-
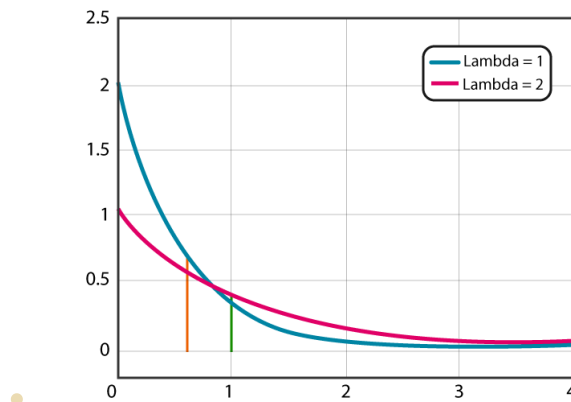
Gamma

- **What it is**
  - Wait time for a fixed number of events
- **Discrete vs. Continuous**
  - Continuous
- **What it looks like**

  -

Weibull

- **What it is**
  - Describes a waiting time for one event, if that event becomes more or less likely with time.
- **Discrete vs. Continuous**
  - Continuous
- **What it looks like**

  -

Central Limit Theorem

- The central limit theorem states that as sample size becomes significantly large, the distribution of events approaches the normal distribution.

Extra Reading

1. Ippolito, P. P. (2021, August 24). *Probability distributions in data science*. Medium. Retrieved March 20, 2023, from https://towardsdatascience.com/probability-distributions-in-data-science-cce6e64873a7