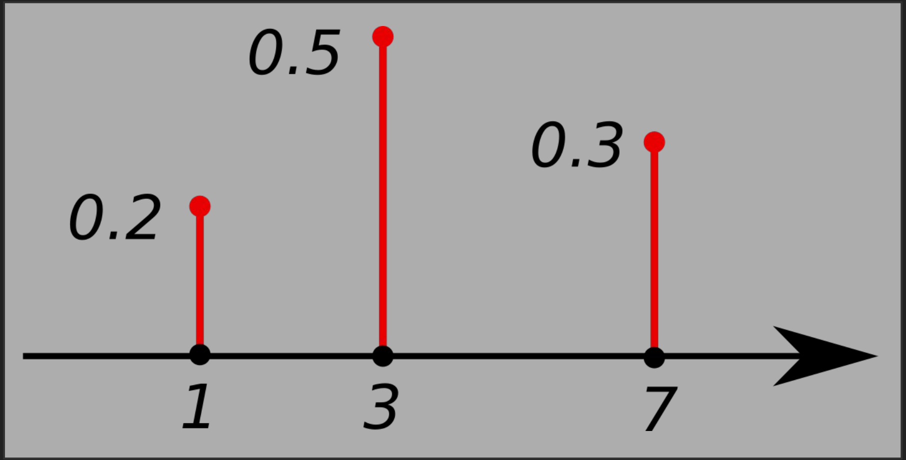
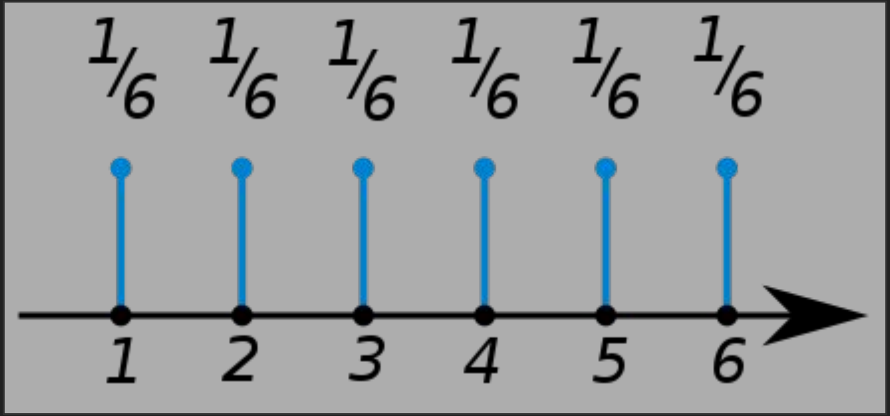
Tour de Distributions!

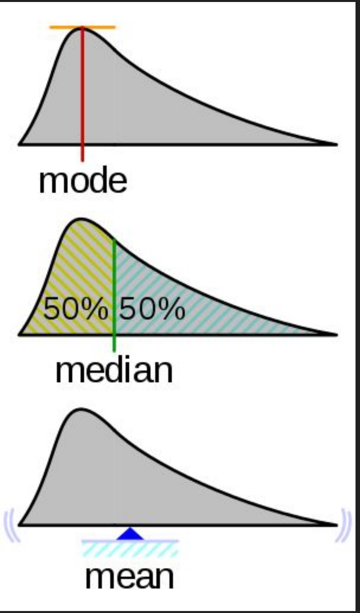
DC-DS-0603

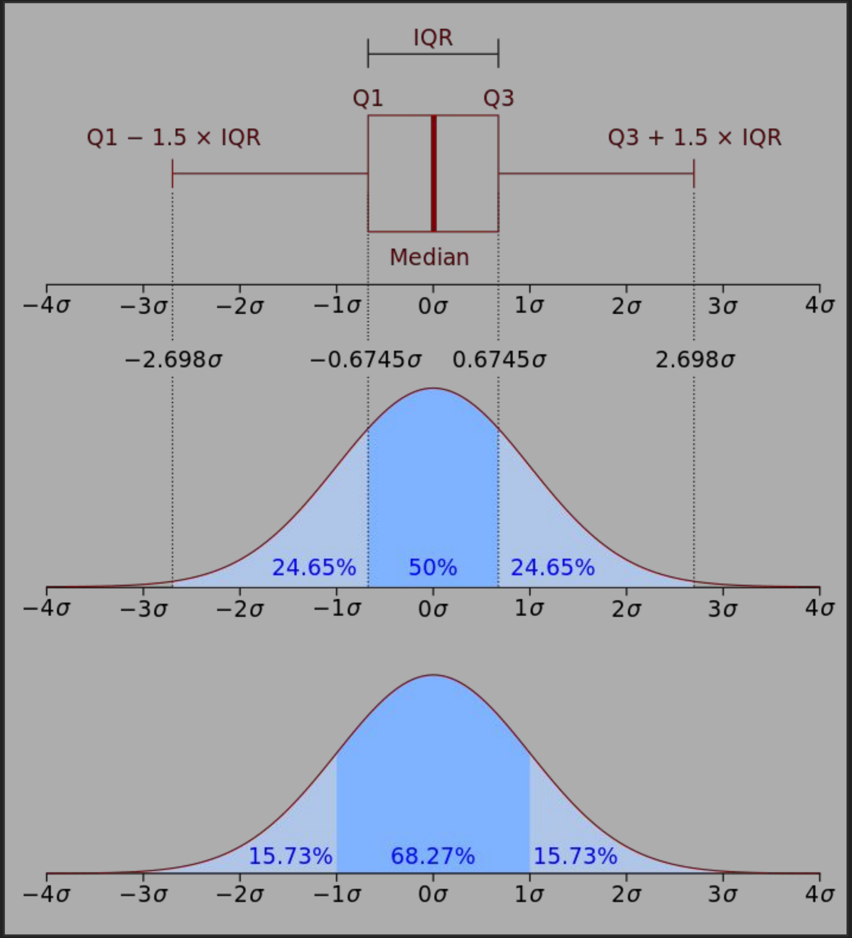
Probability Mass Function



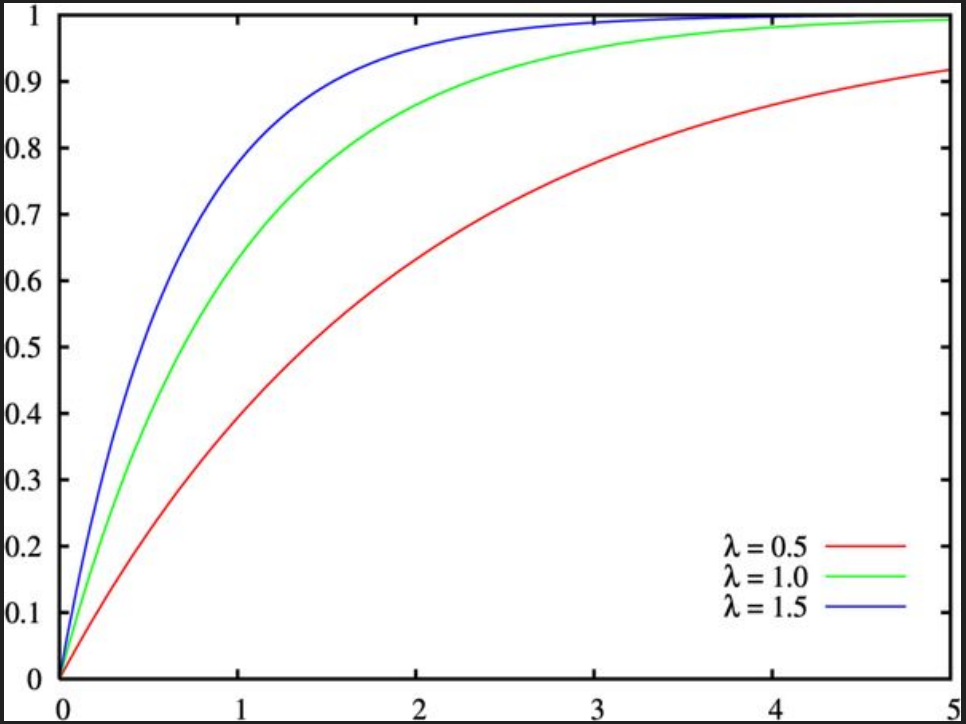


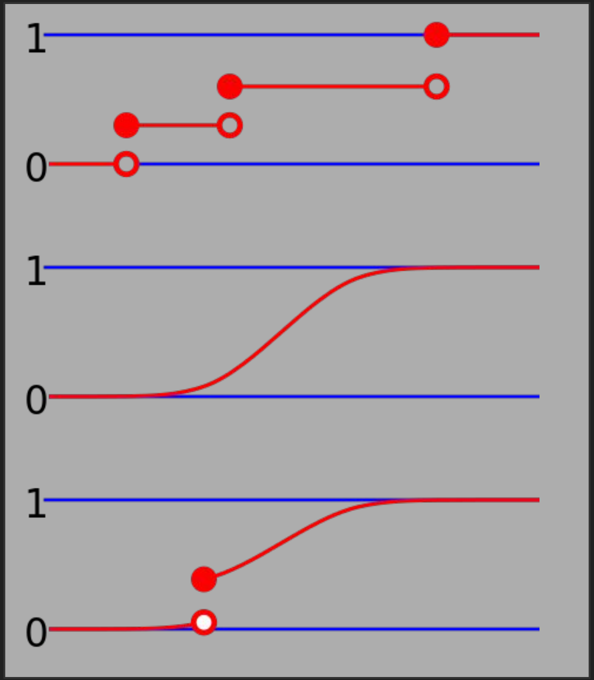
Probability Density Function





Cumulative Density Function

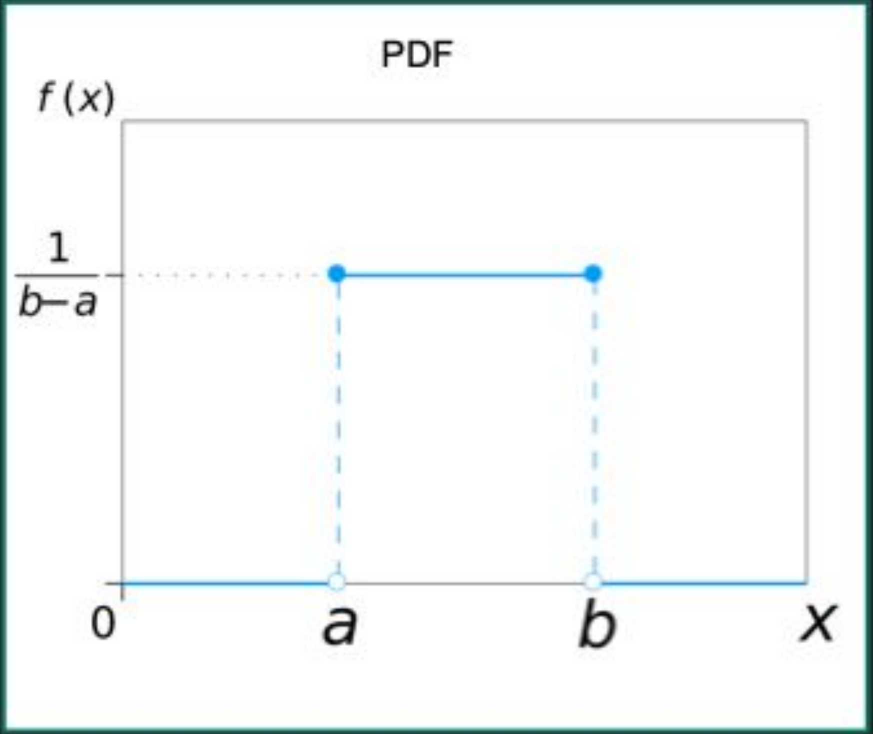


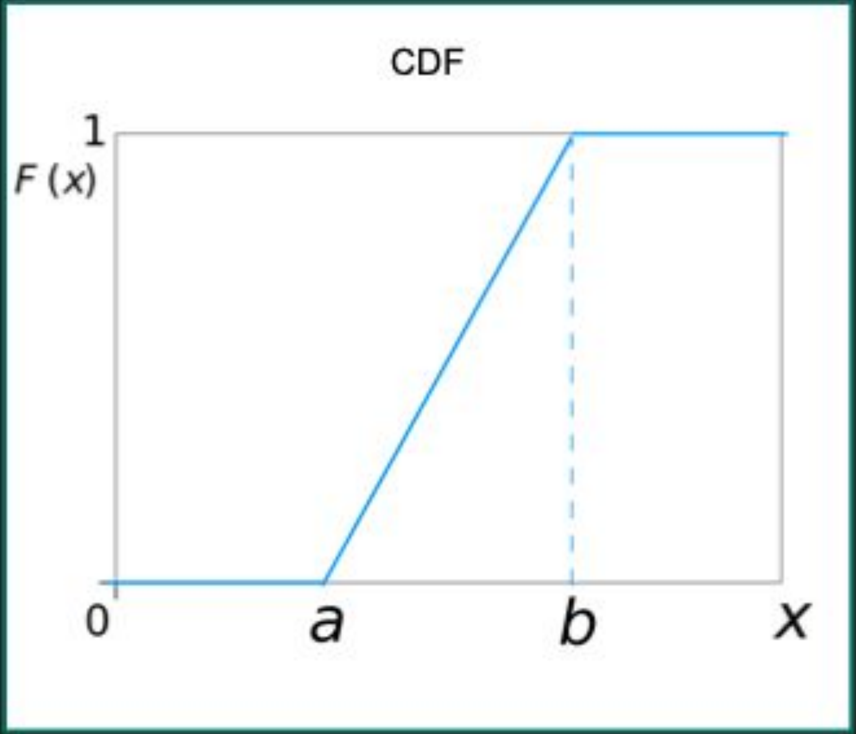


Your turn!

- Uniform
  - Mindy
  - Joe
  - Joey
- Bernoulli/Binomial
  - Cliff
  - Kate
  - Allan
- Normal (Standard Normal)
  - Anthony
  - Nateé
  - Tingting
- Poisson
  - Dmitry
  - Emefa
  - Tim
- Geometric/Exponential
  - Keita
  - Quinn
  - Phoebe
- z-test/Student's T-Test
  - Ngoc
  - Sean
  - Misha

Uniform Distribution





Notation: $U(a,b)$ or unif (a,b)

Parameters: $-\infty<a<b<\infty$

PDF: $1/(b-a)$ for x in (a,b)

CDF: 0 for $x < a$;

$(x-a)/(b-a)$ for x in (a,b)

1 for $x > b$

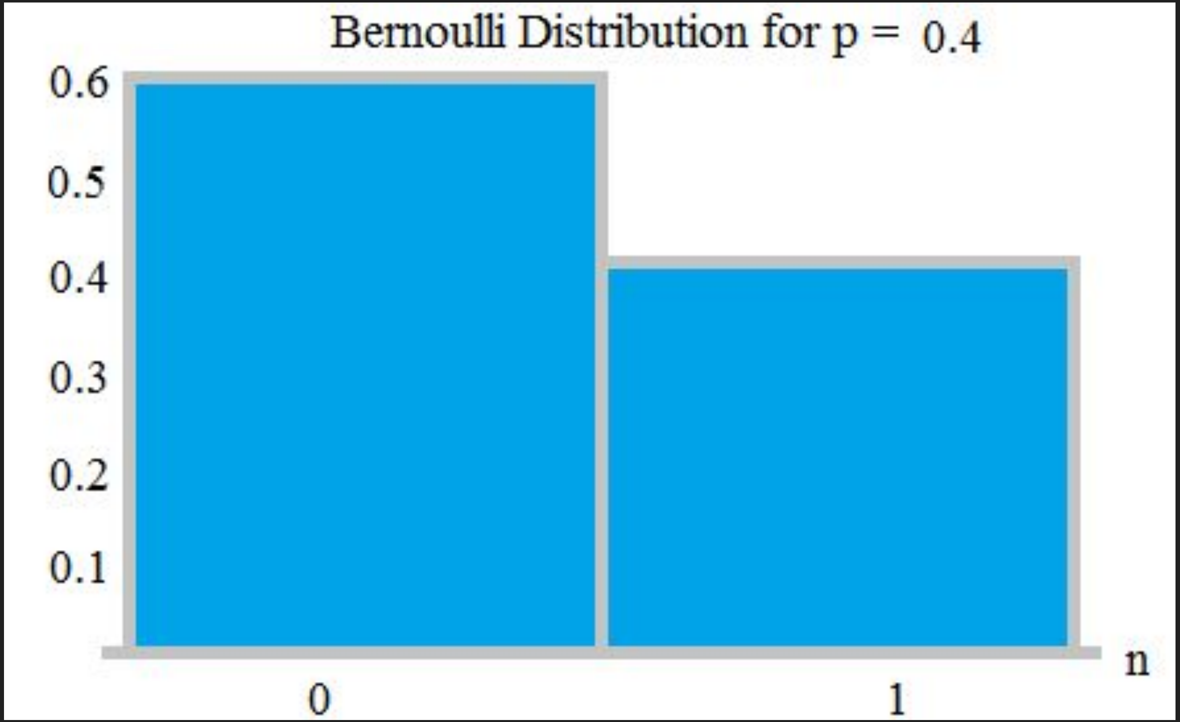
Mean: $(a+b)/2$

Median: $(a+b)/2$

Mode: any value in (a,b)

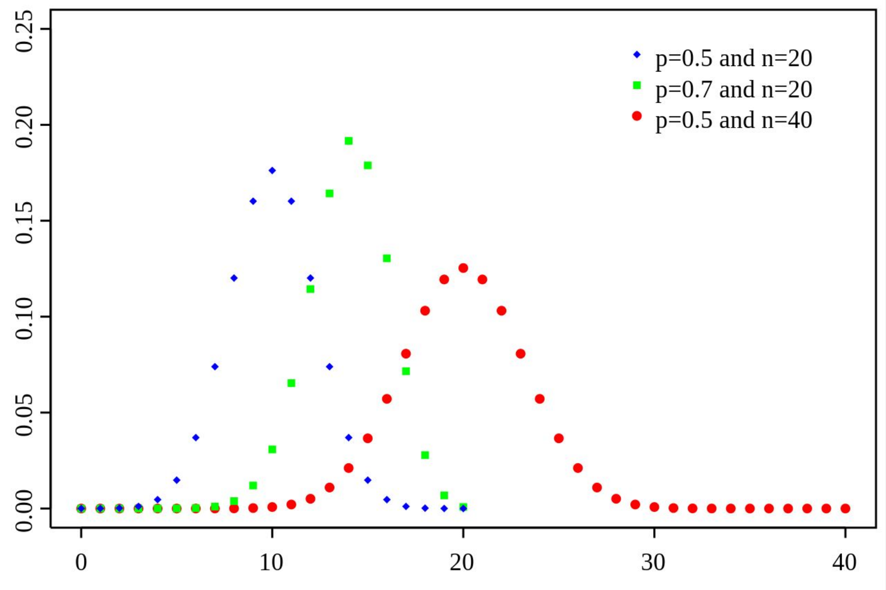
Variance: $1/12(a-b)^2$

Bernoulli and Binomial Distributions

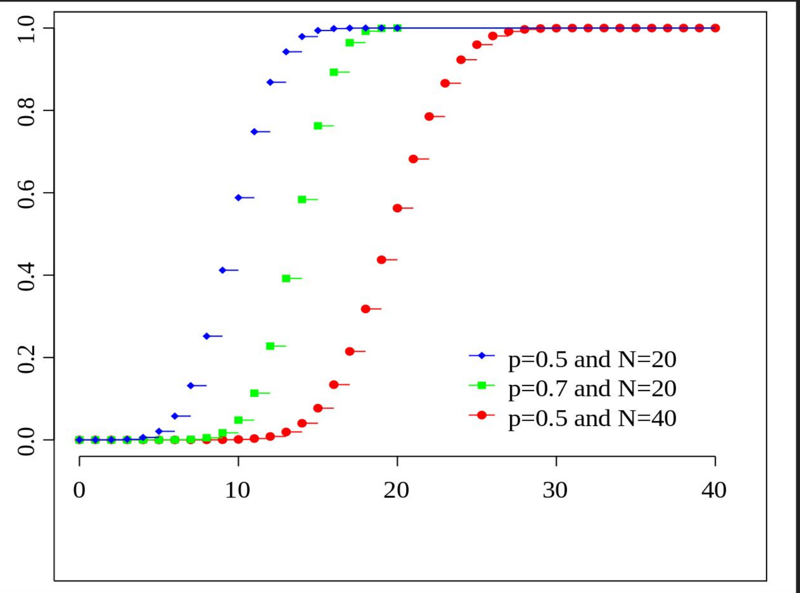


Bernoulli and Binomial Distributions

Binomial PDF



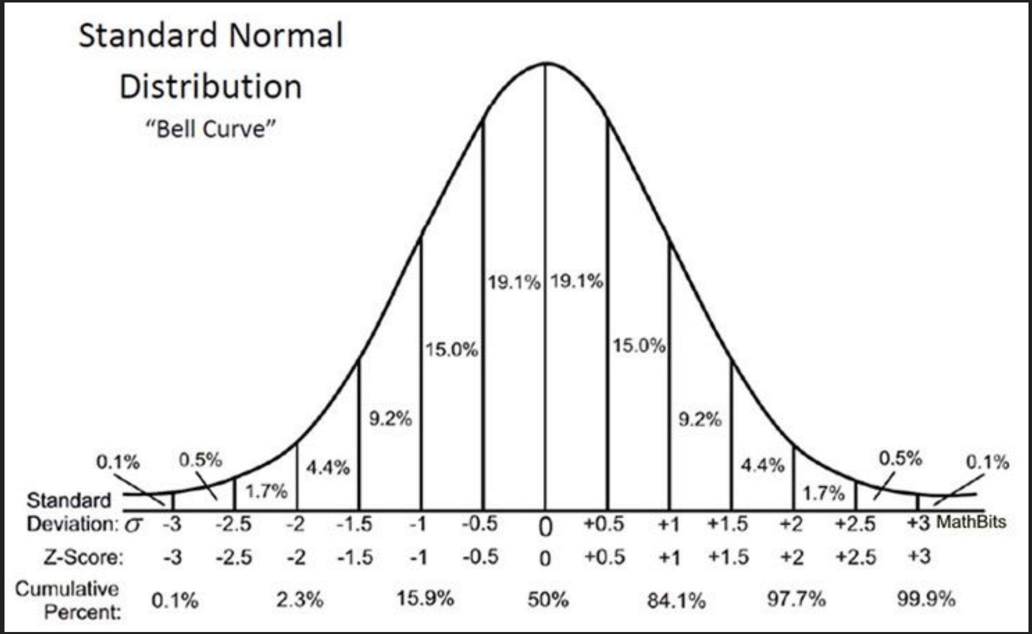
Binomial CDF



Bernoulli and Binomial Distributions

$$P(x) = \frac{N!}{x!(N-x)!} \pi^x (1-\pi)^{N-x}$$

Normal and Standard Normal Distributions



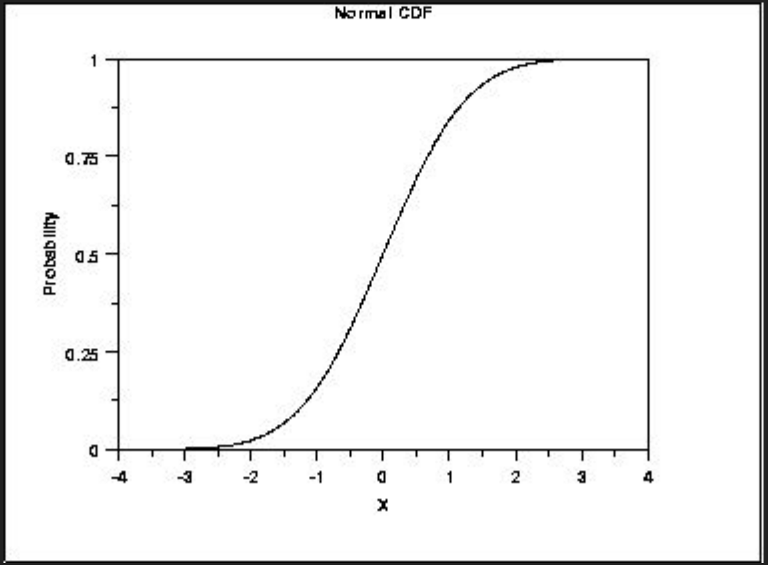
The **Normal Distribution** has:

- mean = median = mode
- symmetry about the center
- 50% of values less than the mean and 50% greater than the mean

Many things closely follow a Normal Distribution:

- heights of people
- size of things produced by machines
- errors in measurements
- blood pressure
- marks on a test

Normal and Standard Normal Distributions



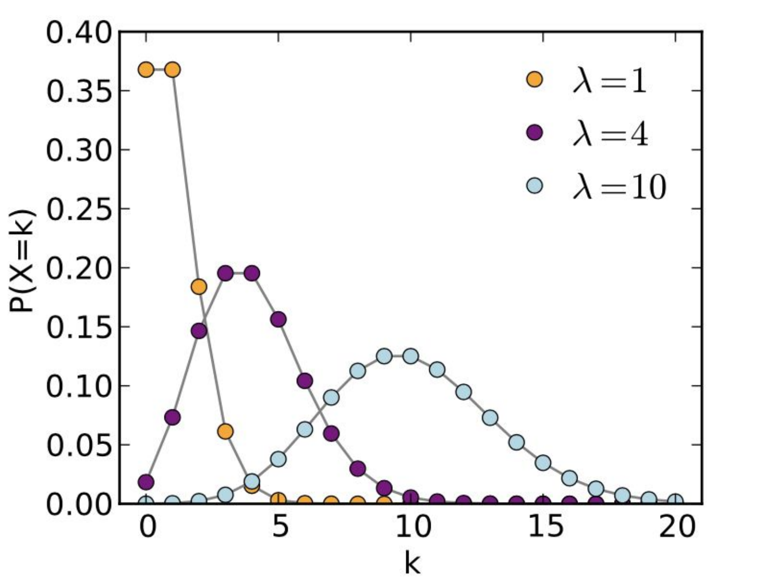
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$

Poisson Distribution

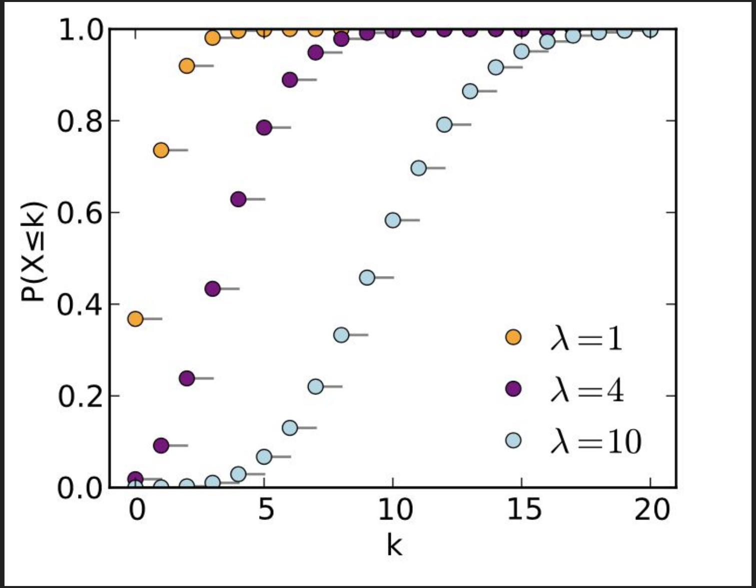
$$P(k \text{ events in interval } t) = e^{-rt}\frac{(rt)^k}{k!}$$

- Allows us to calculate the probability of a given event happening by examining the mean number of events that happen in a given time period

PMF



CDF



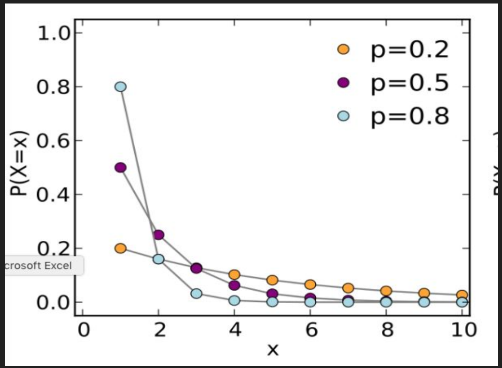
Geometric and Exponential Distributions

Geometric (discrete outcomes): the number of failures before you get a success in a series of Bernoulli trials.

The three assumptions are:

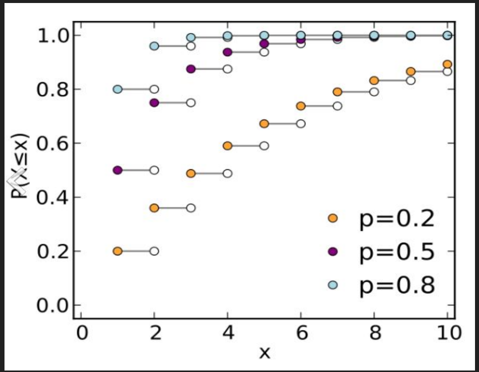
- There are two possible outcomes for each trial (success or failure).
- The trials are independent.
- The probability of success is the same for each trial.

Geometric Distribution PMF



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

Geometric DIstribution CDF



$1-(1-p)^k$

Geometric and Exponential Distributions

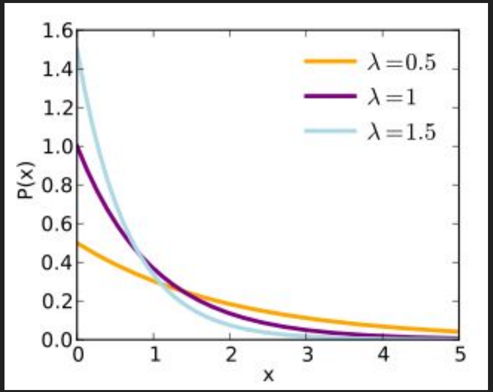
Describes the probability distribution of the amount of time it takes before an event occurs.

Assume: some variable X distributed through a Poisson process:

- This variable has been occurring at some fixed rate over a period of time;
- The event never occurs more than once per interval.

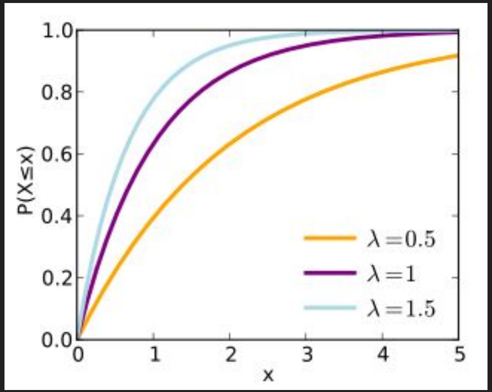
E.g. The number of hits a website receives in an hour; the years between major floods; death rate by age

Exponential Distribution PDF



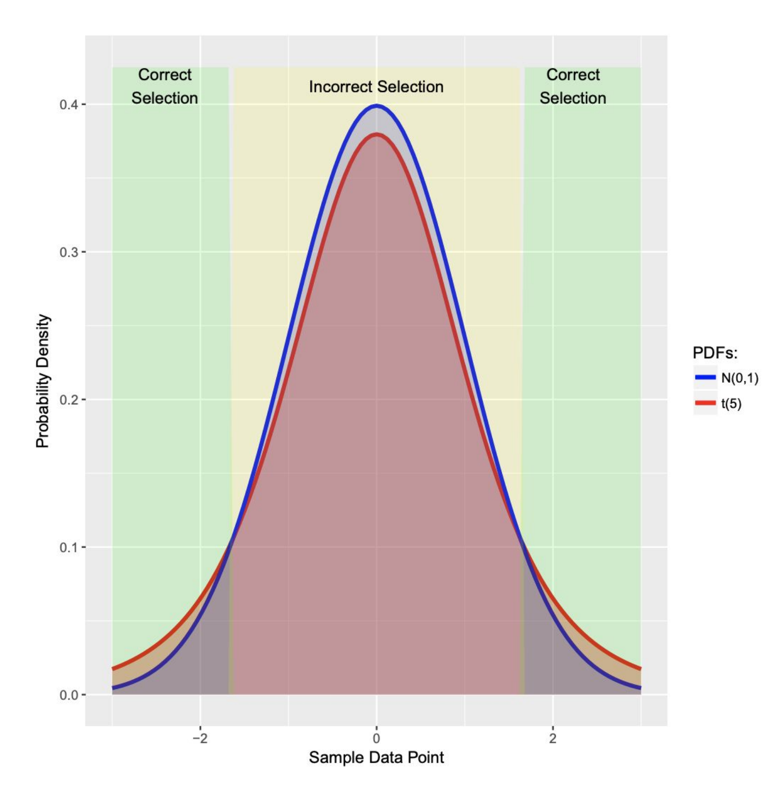
$\lambda e^{-\lambda x}$

Exponential Distribution CDF



$1-e^{-\lambda x}$

Z-test and Student's T-Test



$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$