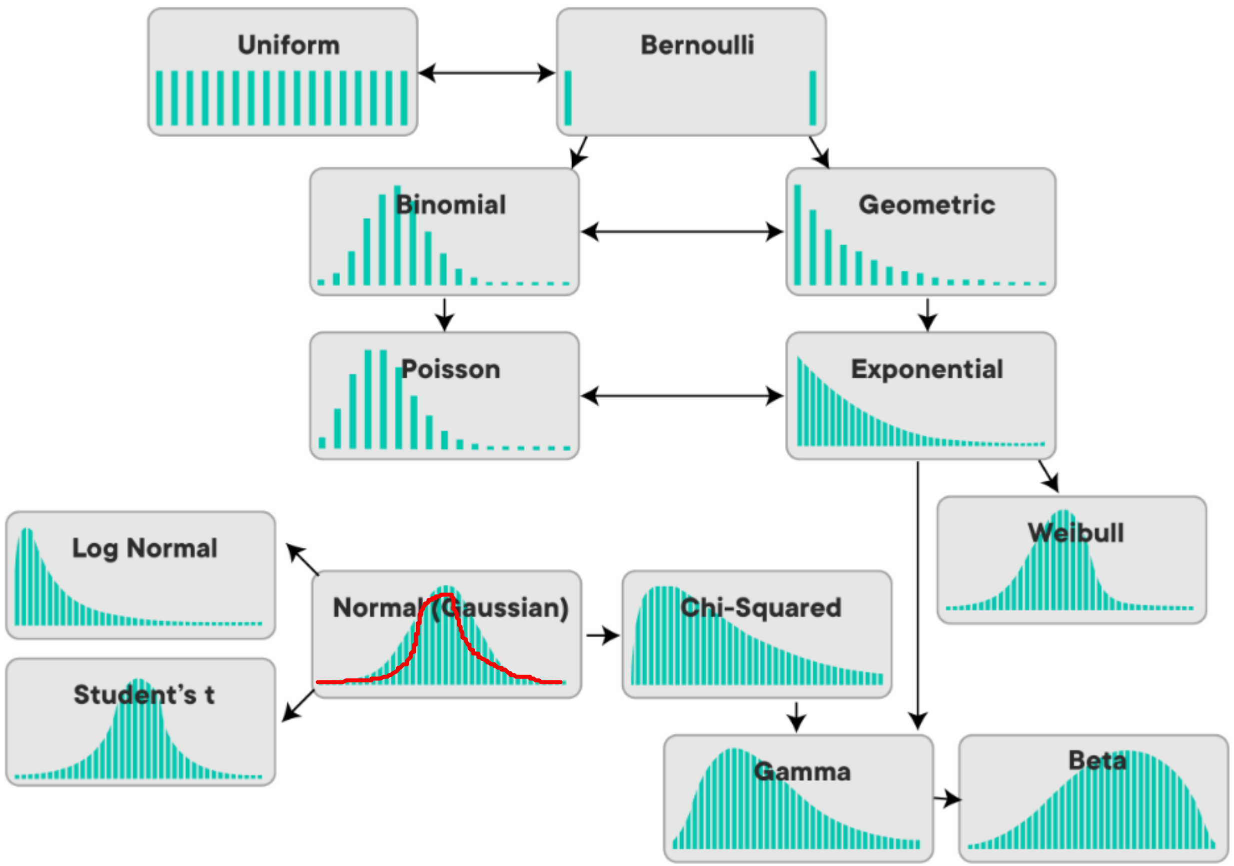
Probability Distribution



Outline

- What is probability distribution?
- Types of Probability Distribution?
  - ➢ Parametric and Non-parametric Probability Distribution
  - ➢ Types of Parametric Probability Distribution

    Discrete probability distributions with finite sample spaces:
    - ✓ Bernoulli distribution
    - ✓ Binomial distribution
    - ✓ Categorical distribution
    - ✓ Uniform distribution

    Discrete probability distributions with infinite sample spaces:
    - ✓ Geometric distribution
    - ✓ Poisson distribution
    - ✓ Skellam distribution
    - ✓ Power Law Distribution

    Continuous Probability Models
    - ✓ The Normal Family of Distributions
    - ✓ The Gamma Family of Distributions
    - ✓ The Beta Family of Distributions

What is probability distribution?

- The values of random variables along with the corresponding probabilities are the probability distribution of the random variable.
- Assume X is a random variable. A function P(X) is the probability distribution of X. Any function F defined for all real x by $F(x) = P(X \leq x)$ is called the distribution function of the random variable X.

Properties of Probability Distribution

- ✓ The probability distribution of a random variable X is $P(X = x_i) = p_i$ for $x = x_i$, and $P(X = x_i) = 0$ for $x \neq x_i$.
- ✓ The range of probability distribution for all possible values of a random variable is from 0 to 1, i.e., $0 \leq p(x) \leq 1$.

Shapes of Distributions

- Skew
- Kurtosis

Skew

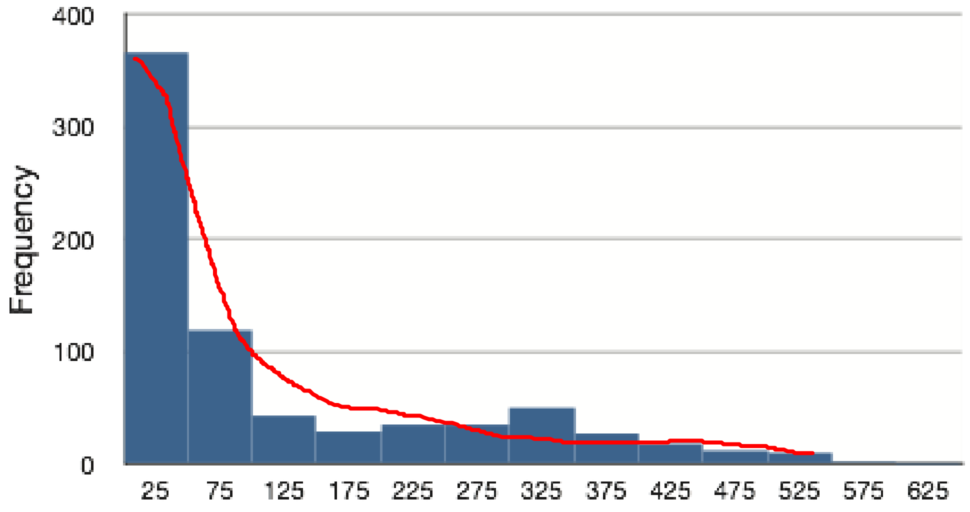


Figure 1. A distribution with a very large positive skew. This histogram shows the salaries of major league baseball players (in thousands of dollars).

- The relationship between skew and the relative size of the mean and median lead the statistician Pearson to propose the following simple and convenient numerical index of skew:

$$\frac{3(Mean - Median)}{\sigma}$$

The standard deviation of the baseball salaries is 1,390,922. Therefore, Pearson's measure of skew for this distribution is 3(1,183,417 - 500,000)/1,390,922 = 1.47.

- The following measure is more commonly used. It is sometimes referred to as the third moment about the mean.

$$\sum \frac{(X-\mu)^3}{\sigma^3}$$

mean

Kurtosis

The following measure of kurtosis is similar to the definition of skew. The value “3” is subtracted to define “no kurtosis” as the kurtosis of a normal distribution. Otherwise, a normal distribution would have a kurtosis of 3.

$$\sum \frac{(X-\mu)^4}{\sigma^4} - 3$$

Types of Probability Distribution

- In statistics, you'll come across dozens of different types of probability distributions, like the binomial distribution, normal distribution and Poisson distribution. All of these distributions can be classified as either a continuous or a discrete probability distribution.

Parametric and Non-parametric

- Parametric (theoretical) probability distributions.

Note: parametric: assume a theoretical distribution (e.g., Gauss)

- Non-parametric: no assumption made about the distribution
- Advantages of assuming a parametric probability distribution:

✓Compaction: just a few parameters

✓Smoothing, interpolation, extrapolation

- **Parameter**: e.g.: μ, σ for **population** mean and standard deviation
- **Statistic**: estimation of parameter from **sample**: x, s sample mean and standard deviation

The probability distribution of a discrete random variable

- The probability distribution of a discrete random variable X is a list of each possible value of X together with the probability that X takes that value in one trial of the experiment.

- The probabilities in the probability distribution of a random variable X must satisfy the following two conditions:

  ➢Each probability P(x) must be between 0 and 1:

  $$0 \le P(x) \le 1$$

  The sum of all the possible probabilities is 1:

  $$\sum P(x) = 1$$

Probability “mass” function

- A ***discrete*** distribution or probability “mass” function (pmf) $p(X)$ is a set of probabilities, one for each value of X. More precisely, denoting *xi* as the values of X

$$p(x_i) = P[X = x_i]$$

for all values x_i of X

$$0 \le p(x_i) \le 1 \quad \text{for all } i$$

$$\sum_i p(x_i) = 1$$

Parameters of a discrete probability distribution

- The probability mass function has two kinds of inputs. The first is the outcome whose probability the function will return. The second is the **parameters** of the probability distribution.
- A common notation for writing probability mass functions is to put the outcome as the first input. Then you list all of its parameters, separated from the outcome by a semicolon:

(PROBABILITY MASS FUNCTION)

P(outcome; p1, p2, p3, ..., pN)

Here, "outcome" stands for an arbitrary element of the sample space. And p1, p2, ... stand for the first parameter, the second parameter, and so on.
A more compact way of expressing a probability mass function is: $P(x; m, n)$

This represents a probability distribution with two parameters, called *m* and *n*. The *x* stands for an arbitrary outcome of the random variable.
With all this background information in mind, let's finally take a look at some real examples of discrete probability distributions.

Commonly used discrete probability distributions

Discrete probability distributions with finite sample spaces:

- Bernoulli distribution
- Binomial distribution
- Categorical distribution
- Uniform distribution

Discrete probability distributions with infinite sample spaces:

- Geometric distribution
- Poisson distribution
- Skellam distribution

The Bernoulli distribution

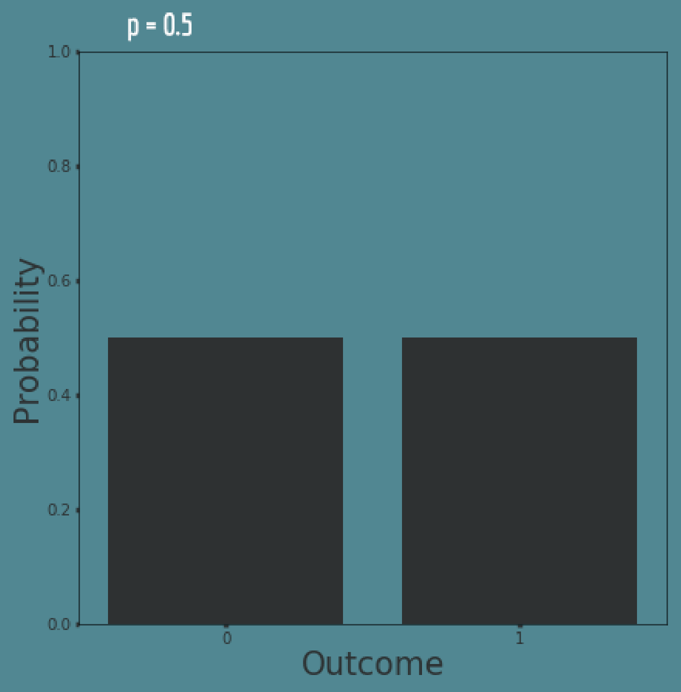
An example: A large pool of green balls and red balls.

Question: If you randomly draw one ball, what is the probability that it will be green?

Answer: Bernoulli distribution.

This is a distribution with a single parameter, often called **p** (a real number between 0 and 1) which represents the probability of one of the outcomes.

(BERNOULLI DISTRIBUTION)



(BERNOULLI DISTRIBUTION PMF)

$$P(x;p) = \begin{cases} p & if\ x = 1 \\ 1 - p & if\ x = 0 \end{cases}$$

Binomial Distribution

- Suppose a random experiment with exactly two outcomes is repeated n times independently. The probability of success is p and that of failure is q. Assume that out of these n times, we get success for x times and failure for the remaining i.e., n−x times. The total number of ways in which we can have success is ^nC_x. A random variable X will have a binomial distribution if

$$P(X = x) = p(x) = {}^nC_x\, p^x\, q^{n-x},$$

for x = 0, 1, … , n and P(X = x) = 0 otherwise. Here, q = 1 – p. Any such random variable X is binomial variate. A binomial trial is a set of n independent Bernoullian trials.

- **Conditions for Binomial Distribution:**

1) Each trial results in only two outcomes i.e., success and failure.
2) The number of trials 'n' is finite.
3) The trials are independent of each other.
4) The probability of success, p or that of failure, q is constant for each trial.

The binomial distribution

- **An example** : a pool of red and green balls.
- **Question:** You draw a ball at random and then throw it back inside the pool and mix the balls. If you repeat this 5 times, what is the probability that you will draw exactly 3 green balls?
- **Answer: Binomial distribution**. More generally, a binomial distribution is about the probability of getting *x* successes out of *n* independent trials, where each trial has a Bernoulli distribution with the same parameter p.
- Therefore, a binomial distribution has 2 parameters: **p** and **n**. Here p is the parameter of the Bernoulli distribution that defines each independent trial and n is the number of trials. In a way, the Bernoulli distribution is a special case of the binomial distribution. That is, a Bernoulli distribution is simply a binomial distribution with the parameter n equal to 1.

(BINOMIAL DISTRIBUTION PMF)

$$P(x; p, n) = \binom{n}{x} p^x (1-p)^{n-x}$$

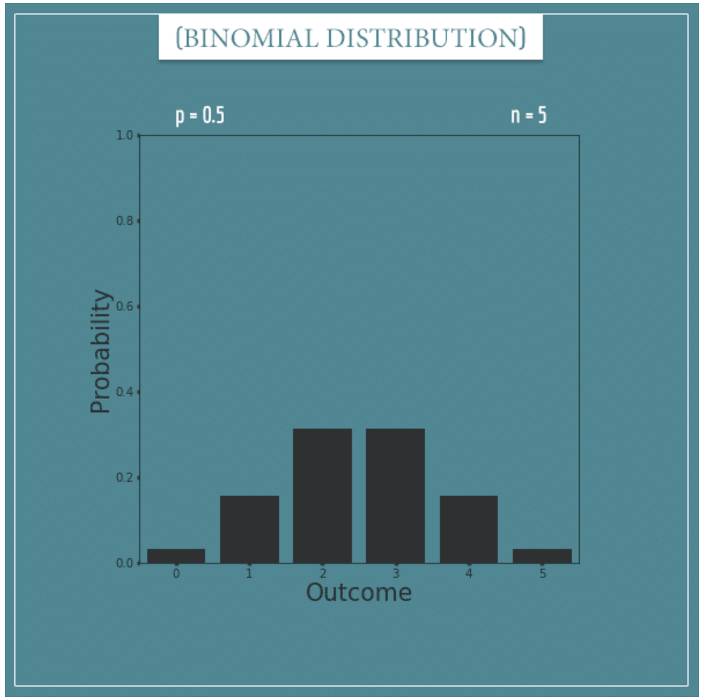
$_nC_x$

The first term is the binomial coefficient

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

$$P(x=2; p=0.3, n=3) = \binom{3}{2} \cdot 0.3^2 \cdot 0.7^1$$
$$= \frac{3!}{(3-2)! \cdot 2!} \cdot 0.3^2 \cdot 0.7^1$$
$$= 3 \cdot 0.09 \cdot 0.7 = 0.189$$

Namely, the number of ways in which you can arrange x objects into n slots.



For example, let's take a binomial distribution with p = 0.3 and n = 3. That is, we have a pool of 30% green balls and 70% red balls and we're drawing 3 balls at random. Let's say we want to calculate the probability of exactly 2 of them being green:

- The Bernoulli distribution is a special case of the binomial distribution where the parameter **n** is fixed to 1. Similarly, it is a special case of the categorical distribution where the number of possible outcomes is fixed to only 2.
- In case you're wondering, there is also a generalization which allows both more than 2 outcomes and more than 1 trial. This is called the **multinomial distribution**. The multinomial distribution is also encountered very frequently in a wide variety of domains, but I am going to leave its introduction for a future post.

The categorical distribution

- **An example:** a pool of red, green, blue, and black balls. And the percentages are:

Red: 30%; Green: 20%; Blue: 10%; Black: 40%

$P_0(R) = 0.3$ $\quad P_1 = 0.2$ $\quad P_2 = 0.1$ $\quad P_3 = 0.4$

- **Question:** If you draw a single ball, what is the probability of it being of a particular color?
- **Answer:** This probability is given by a **categorical distribution**. The categorical distribution describes random variables which have an arbitrary number of possible outcomes. This distribution has n parameters, where n is the number of possible outcomes. If we label the outcomes with the first n integers, then the parameters could be labeled $\mathbf{p_0, p_1, p_2, \ldots p_{n-1}}$.
- **Notice:** unlike the previous two distributions, the categorical distribution has a variable number of parameters!

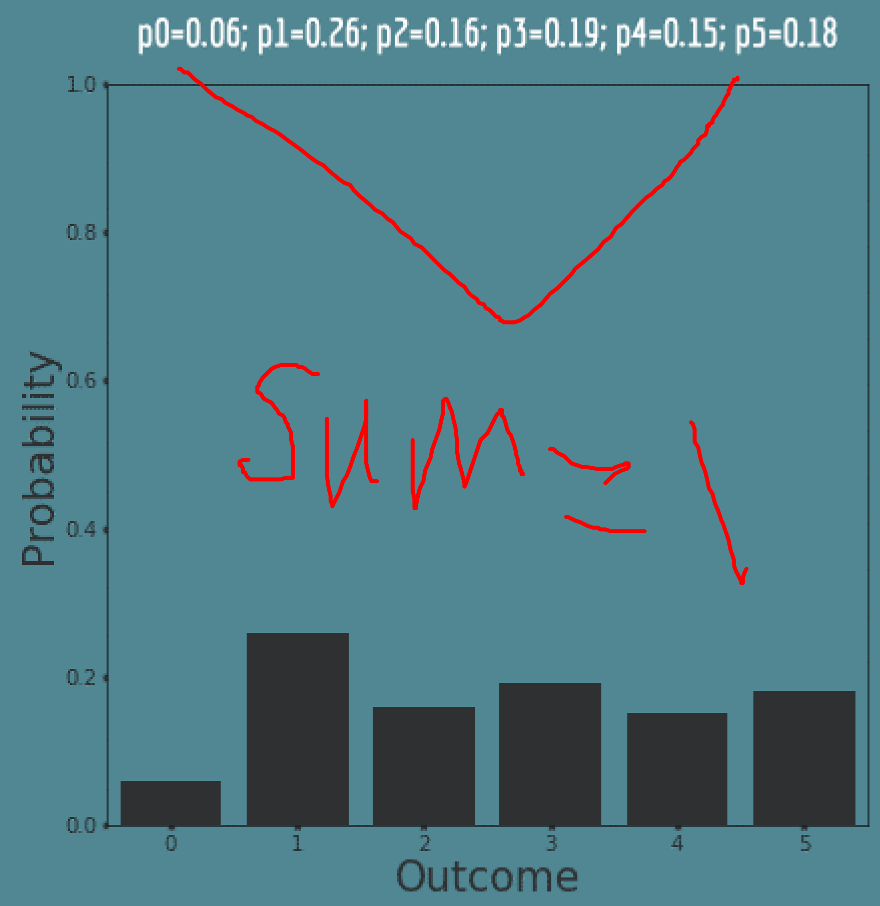
The categorical distribution

- Here’s a categorical distribution of a random variable with 6 possible outcomes and some arbitrary values for its 6 parameters:

(CATEGORICAL DISTRIBUTION PMF)

$$P(x; p_0, p_1, \dots, p_{n-1}) = p_x$$

(CATEGORICAL DISTRIBUTION)



The discrete uniform distribution

- **An example:** a pool of red, green, blue, and black balls. But this time they are all in the same proportion. That is, they are each 1/4 or 25% of the total number of balls in the pool.

  0.25

- **Question:** If you draw a single ball, what is the probability of it being of a particular color?
- Answer: **discrete uniform distribution.** The random variable can have any number of possible outcomes, but they all have to be equally likely.
- The discrete uniform distribution has a single parameter **n** which is equal to the number of possible outcomes.

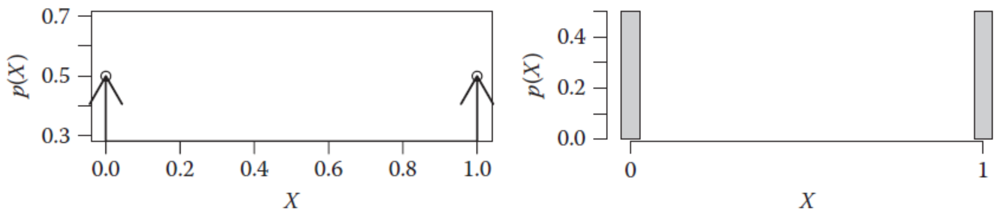
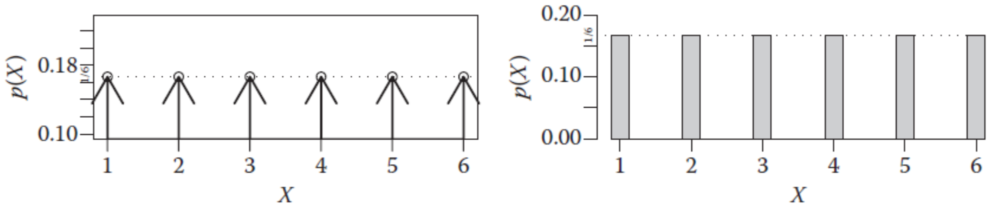


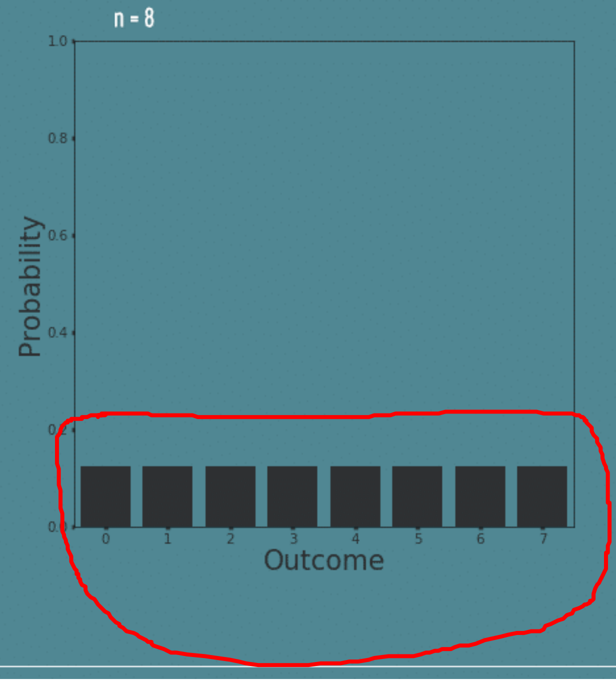
FIGURE 3.2 pmf of a discrete RV represented as a spike graph and as bar graph.



(UNIFORM DISTRIBUTION PMF)

$$P(x;n) = \frac{1}{n}$$

(UNIFORM DISTRIBUTION)



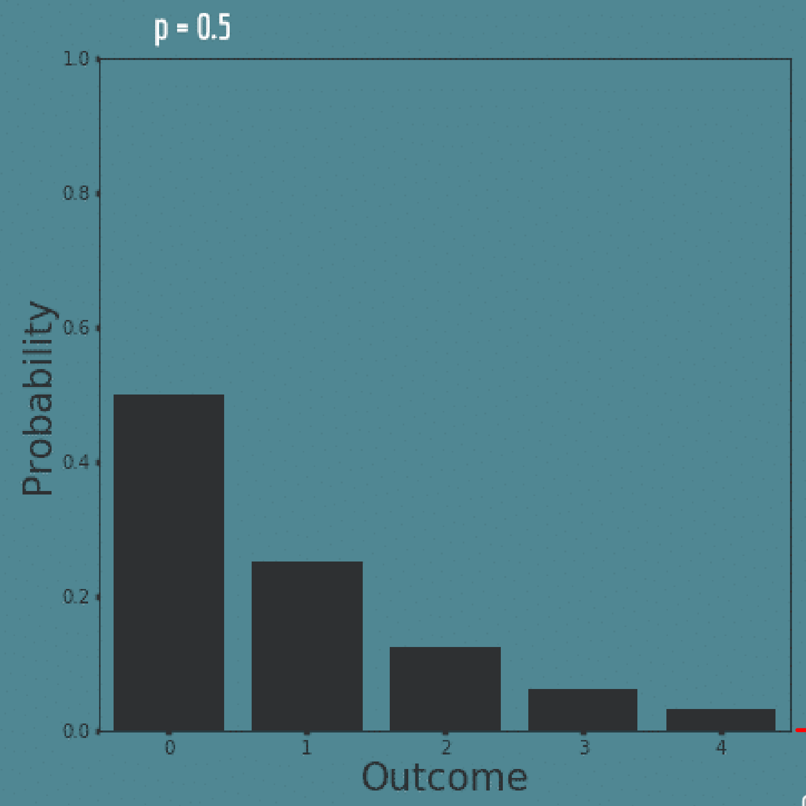
The geometric distribution

- As you can see, the probabilities decrease exponentially for each additional "failure" trial. Not being able to plot infinitely many outcomes isn't really a big problem because at some point the probabilities start getting very close to 0. So, for any geometric distribution, most of its probability mass will be concentrated over the first N outcomes, where N depends on the parameter p.
- Notice that the higher the value of p is, the easier it is to get a "success" trial and, therefore, the total probability mass will be concentrated over very few of the first outcomes.

(GEOMETRIC DISTRIBUTION PMF)

$$P(x;p) = (1-p)^x p$$

(GEOMETRIC DISTRIBUTION)



The Poisson distribution

- A Poisson distribution is the probability distribution that results from a Poisson experiment.
- **The Poisson distribution is a limiting case to the binomial distribution**
- Attributes of a Poisson Experiment:

1) The experiment results in outcomes that can be classified as successes or failures.
2) The average number of successes (μ) that occurs in a specified region is known.
3) The probability that a success will occur is proportional to the size of the region.
4) The probability that a success will occur in an extremely small region is virtually zero.
5) Note that the specified region could take many forms. For instance, it could be a length, an area, a volume, a period of time, etc.

(POISSON DISTRIBUTION PMF)

$$P(x;\mu) = \frac{\mu^x e^{-\mu}}{x!}$$

(BINOMIAL DISTRIBUTION PMF)

$$P(x;p,n) = \binom{n}{x} p^x (1-p)^{n-x}$$

The Poisson distribution

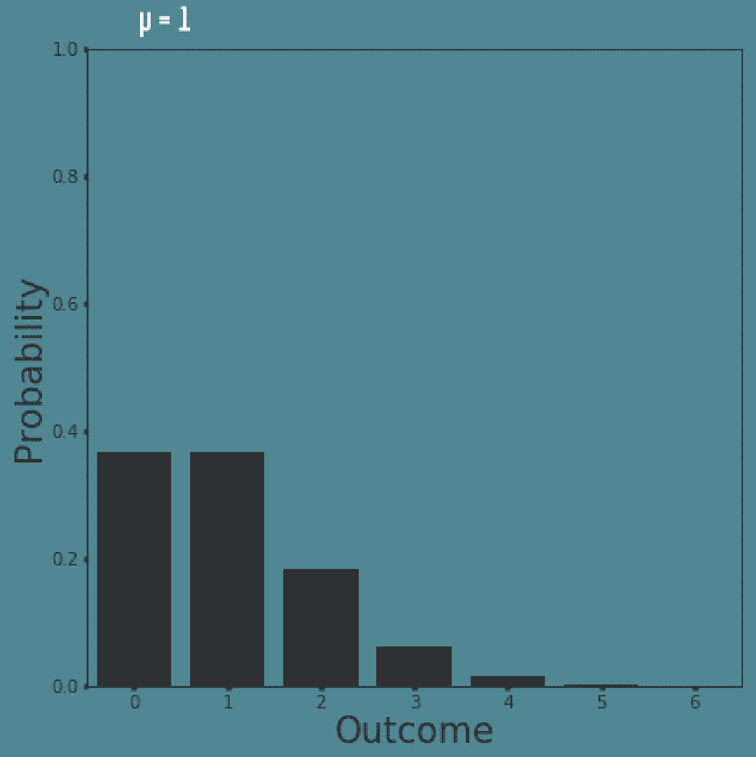
- A Poisson distribution is the probability distribution that results from a Poisson experiment.
- **The Poisson distribution is a limiting case to the binomial distribution**
- Attributes of a Poisson Experiment:

1) The experiment results in outcomes that can be classified as successes or failures.
2) The average number of successes (μ) that occurs in a specified region is known.
3) The probability that a success will occur is proportional to the size of the region.
4) The probability that a success will occur in an extremely small region is virtually zero.
5) Note that the specified region could take many forms. For instance, it could be a length, an area, a volume, a period of time, etc.

(POISSON DISTRIBUTION PMF)

$$P(x; \mu) = \frac{\mu^x e^{-\mu}}{x!}$$

(POISSON DISTRIBUTION)



Poisson Formula. Suppose we conduct a Poisson experiment, in which the average number of successes within a given region is μ. Then, the Poisson probability is:

$$P(x; \mu) = (e^{-\mu}) (\mu^{x}) / x!$$

where *x* is the actual number of successes that result from the experiment, and *e* is approximately equal to 2.71828.

The Poisson distribution has the following properties:

- The mean of the distribution is equal to μ .
- The variance is also equal to μ .

P

An example : The average number of homes sold by the Acme Realty company is 2 homes per day.
Question: What is the probability that exactly 3 homes will be sold tomorrow?
Solution: This is a Poisson experiment in which we know the following:

P(

- $\mu = 2$; since 2 homes are sold per day, on average.
- $x = 3$; since we want to find the likelihood that 3 homes will be sold tomorrow.
- $e = 2.71828$; since *e* is a constant equal to approximately 2.71828.

We plug these values into the Poisson formula as follows:

$$P(x; \mu) = (e^{-\mu})\,(\mu^{x}) \;/\; x!$$
$$P(3; 2) = (2.71828^{-2})\,(2^{3}) \;/\; 3!$$
$$P(3; 2) = (0.13534)\,(8) \;/\; 6$$
$$P(3; 2) = 0.180$$

Thus, the probability of selling 3 homes tomorrow is 0.180 .

Poisson Calculator: Stat Trek Poisson Calculator

Cumulative Poisson Probability

The probability that the Poisson random variable is greater than some specified lower limit and less than some specified upper limit.

An example : Suppose the average number of lions seen on a 1-day safari is 5.
Question: What is the probability that tourists will see fewer than four lions on the next 1-day safari?
***Solution**:* This is a Poisson experiment in which we know the following:

- $\mu = 5$; since 5 lions are seen per safari, on average.
- $x = 0, 1, 2,$ or 3; since we want to find the likelihood that tourists will see fewer than 4 lions; that is, we want the probability that they will see 0, 1, 2, or 3 lions.
- $e = 2.71828$; since *e* is a constant equal to approximately 2.71828.

First, find the probability that tourists will see 0, 1, 2, or 3 lions. Thus, calculate the sum of four probabilities: P(0; 5) + P(1; 5) + P(2; 5) + P(3; 5). To compute this sum, we use the Poisson formula:

$$P(x \le 3, 5) = P(0; 5) + P(1; 5) + P(2; 5) + P(3; 5)$$
$$P(x \le 3, 5) = [(e^{-5})(5^0) / 0!] + [(e^{-5})(5^1) / 1!] + [(e^{-5})(5^2) / 2!] + [(e^{-5})(5^3) / 3!]$$
$$P(x \le 3, 5) = [(0.006738)(1) / 1] + [(0.006738)(5) / 1] + [(0.006738)(25) / 2] + [(0.006738)(125) / 6]$$
$$P(x \le 3, 5) = [0.0067] + [0.03369] + [0.084224] + [0.140375]$$
$$P(x \le 3, 5) = 0.2650$$

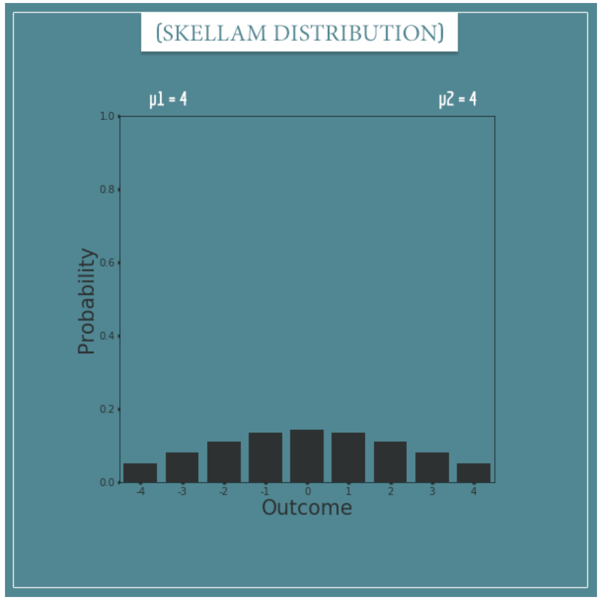
Thus, the probability of seeing at no more than 3 lions is 0.2650.

The Skellam distribution

- In short, a random variable having the **Skellam distribution** is the result of taking the difference between two independent random variables which have a Poisson distribution.
- In other words, if you have any two independent Poisson distributed random variables, their difference will be Skellam distributed.

(SKELLAM DISTRIBUTION PMF)

$$P(x;\mu_1,\mu_2) = e^{-(\mu_1+\mu_2)}\left(\frac{\mu_1}{\mu_2}\right)^{\frac{x}{2}} I_x(2\sqrt{\mu_1\mu_2})$$



Power Law Distribution

- The **power law** (also called the scaling law) states that a relative change in one quantity results in a proportional relative change in another. The simplest example of the law in action is a square; if you double the length of a side (say, from 2 to 4 inches) then the area will quadruple (from 4 to 16 inches squared). A power law distribution has the form $Y = k X^{\alpha}$, where:

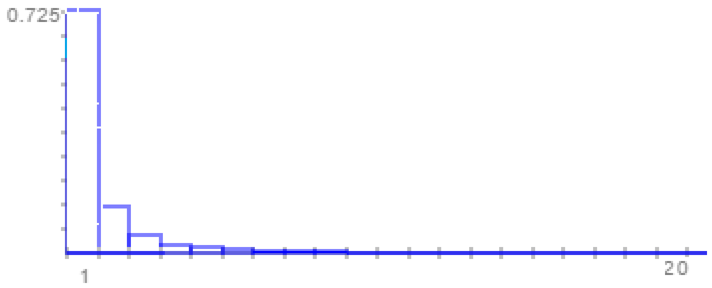
  X and Y are variables of interest,

  α is the law's exponent,

  k is a constant.

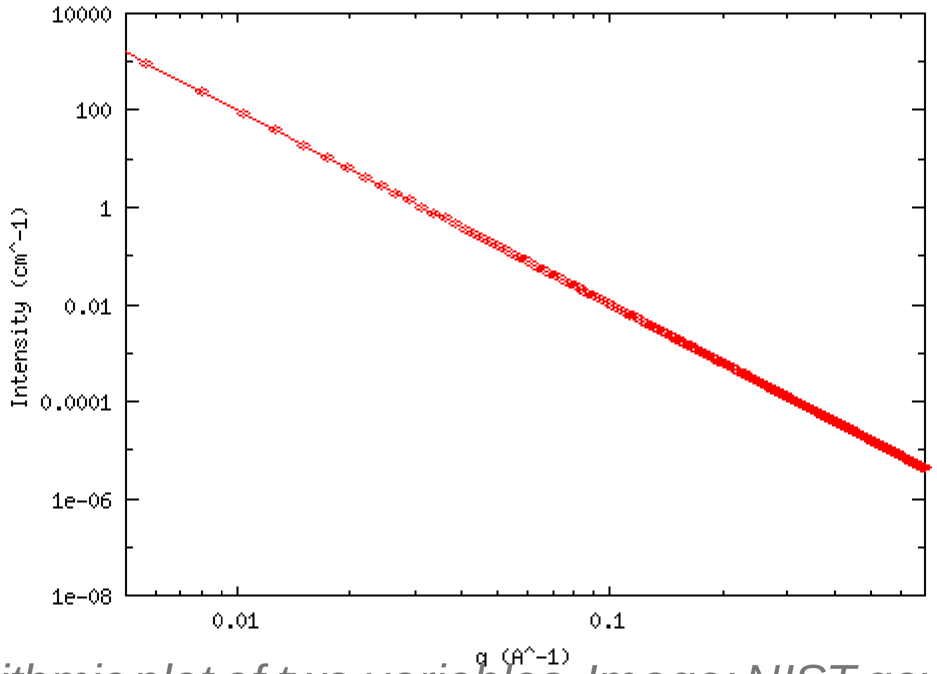
  Any inverse relationship like $Y = X^{-1}$ is also a power law, because a change in one quantity results in a negative change in another.

https://www.statisticshowto.com/power-law/#:~:text=The%20power%20law%20can%20be,while%20larger%20occurrences%20are%20rare.



- Examples
- The power law can be used to describe a phenomenon where a small number of items is clustered at the top of a distribution (or at the bottom), taking up 95% of the resources. In other words, it implies a small amount of occurrences is common, while larger occurrences are rare. For example, where the distribution of income is concerned, there are very few billionaires; the bulk of the population holds very modest nest eggs.
- A cluster of values dominates at one end of the graph.

- If you plot two quantities against each other with logarithmic axes and they show a linear relationship, this indicates that the two quantities have a power law distribution.
- Other examples of phenomena with this type of distribution:
  - Distribution of income,
  - Magnitude of earthquakes,
  - Size of cities according to population,
  - Size of corporations,
  - Trading volumes on the stock market,
  - word frequencies.



Logarithmic plot of two variables. Image: NIST.gov.

- Discrete probability distributions are usually described with **a frequency distribution table, or other type of graph or chart**. For example, the following chart shows the probability of rolling a die. All of the die rolls have an equal chance of being rolled (one out of six, or 1/6). This gives you a discrete probability distribution of:

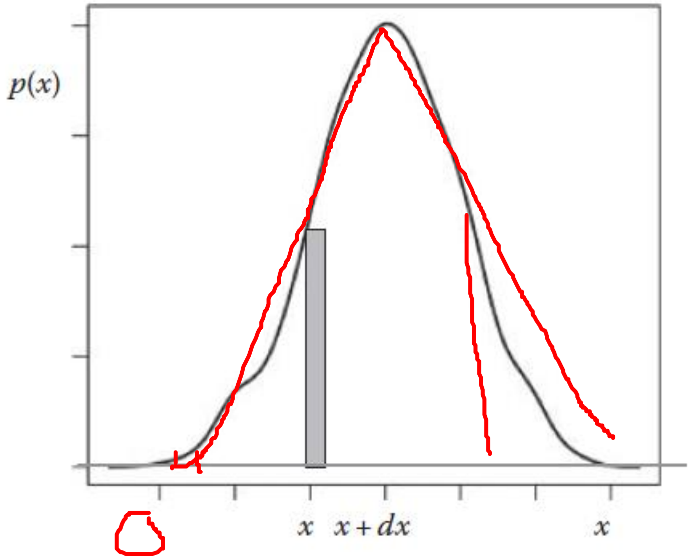
- **Continuous probability distributions are expressed with a formula (a Probability Density Function)** describing the shape of the distribution.

mass

Continuous Probability Models

Probability Density

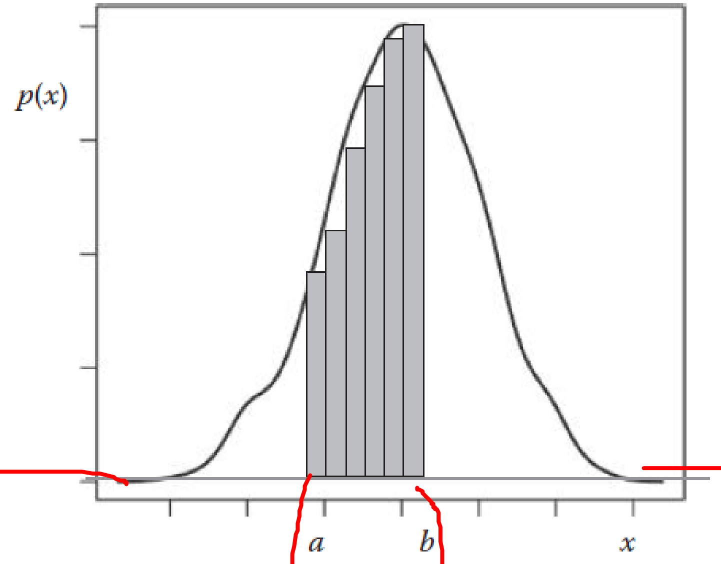
- Distributions for continuous variables are called continuous distributions.
- Probability density function (pdf)
- Probability is area under the curve in between two values separated by a very small difference. where $p(x)$ is always positive or zero, that is $p(x) \geq 0$. Probability is the area under the curve in between the two values x and $x + dx$.

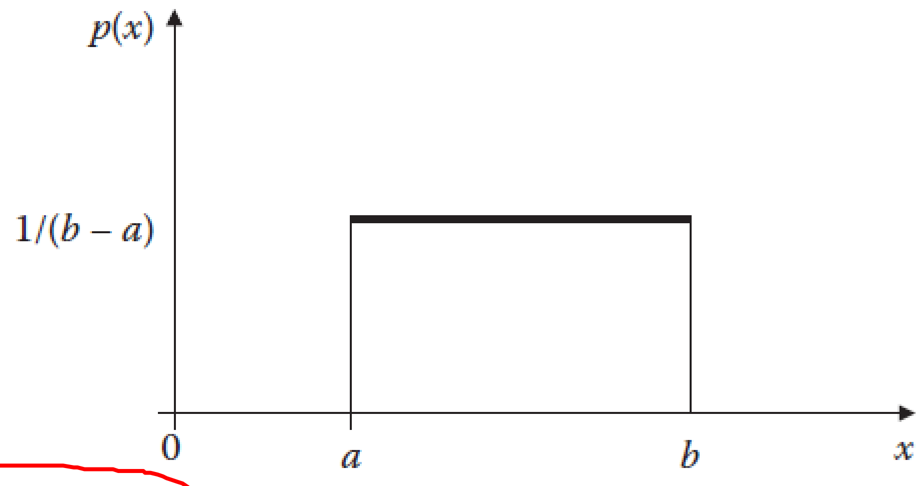


- The probability of a value being in an interval of *X* between *a* and *b* can be found using the integral

$$P[a < X \le b] = \int_a^b p(x)dx$$

$$\int_{-\infty}^{+\infty} p(x)dx = 1$$





pdf of a uniform RV.

Cumulative Functions (cdf)

- The “cumulative” density function (cdf) at a given value are defined by “accumulating”
- All probabilities up to that value. Accumulation is simply an integration in the case of continuous RV.

$$F(x) = P[X \le x] = \int_{-\infty}^{x} p(s)ds$$

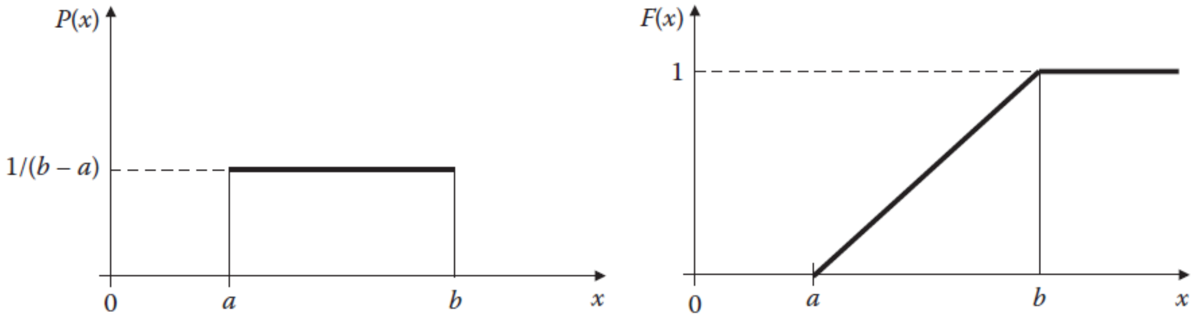
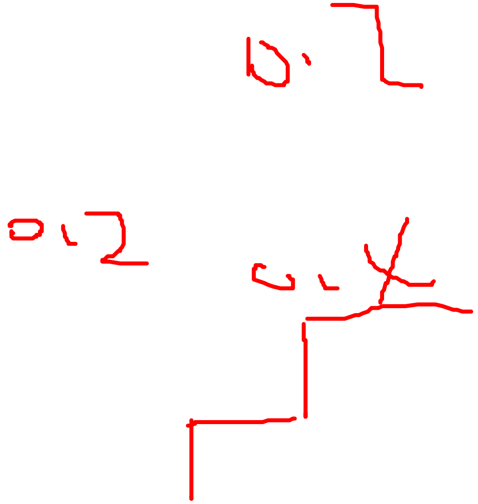
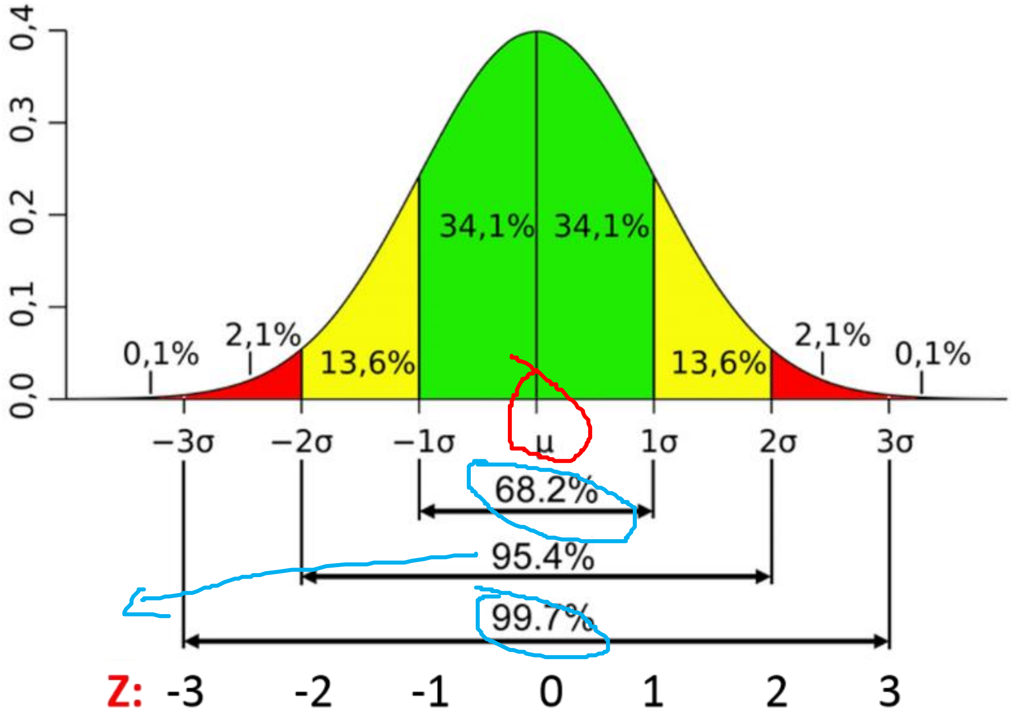


FIGURE 3.8 pdf and cdf of a uniform RV. Integration of a constant yields a linear increase (ramp function).



The Normal Probability Distribution

- An infinite number of distributions with differing means (μ) and standard deviations (σ).
- Calculate a probability for a range of outcomes instead of calculating exact probability for an outcome
- Symmetric and centered on the mean
- Mean is same as the median and mode.
- Three sigma rule or the 68-95-99.7 rule

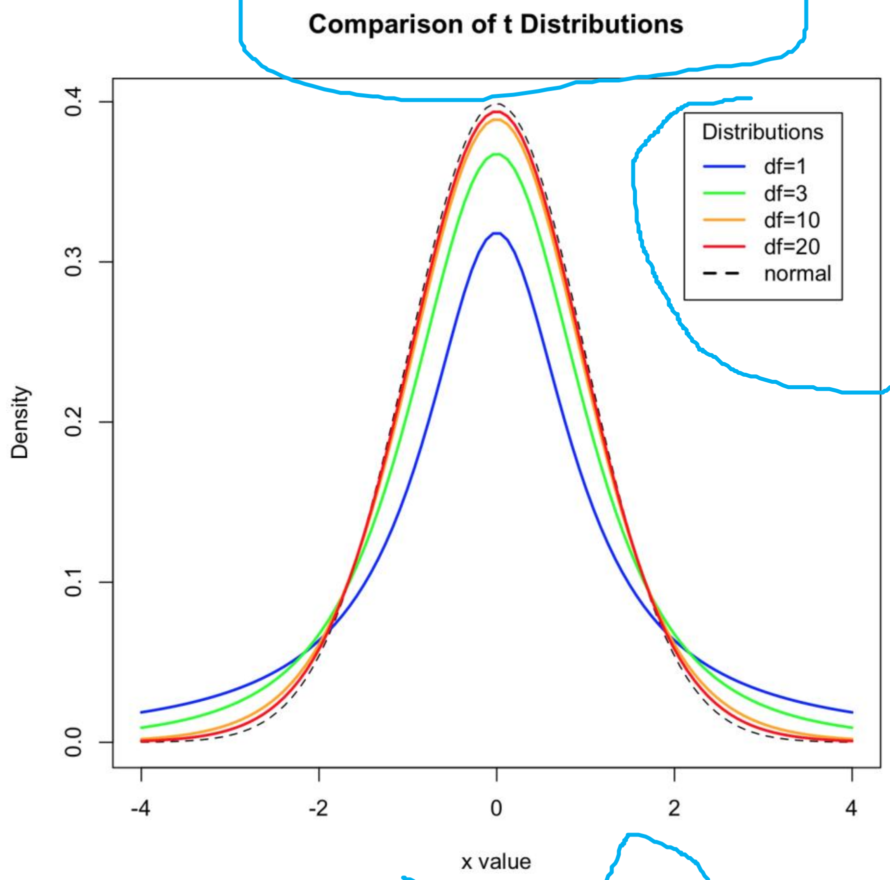


Normal Probability Density Function

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

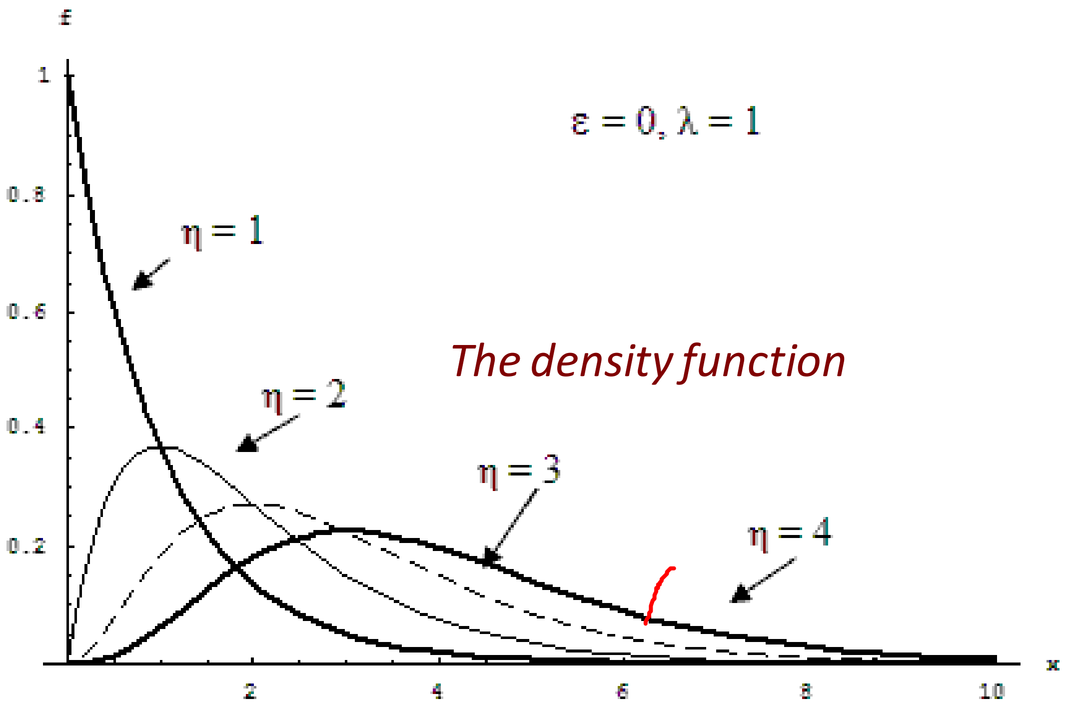
Any and all normal distributions can be converted to the standard normal distribution using the equation below. The z-score equals an X minus the population mean (μ) all divided by the standard deviation (σ).

$$Z = \frac{X - \mu}{\sigma}$$



The Gamma Family of Distributions

- It covers any specified average, standard deviation and skewness.



$$f\left(x \middle| \varepsilon,\lambda,\eta\right)=\begin{cases} 0 & x \le \varepsilon \\ \dfrac{1}{\lambda^{\eta}\Gamma(\eta)}(x-\varepsilon)^{\eta-1}e^{-\frac{x-\varepsilon}{\lambda}} & x > \varepsilon \end{cases}$$

Parameters:	Location:	ε	$-\infty < \varepsilon < \infty$
	Scale:	λ	$\lambda > 0$
	Shape:	η	$\eta > 0$

Bounds:	Bounded below by ε.

As the skewness goes to zero, both the gamma and negative gamma distributions limit to the normal distribution.

Moments of Gamma Distribution

Mean: $\varepsilon + \lambda\eta$

Standard Deviation: $\lambda\sqrt{\eta}$

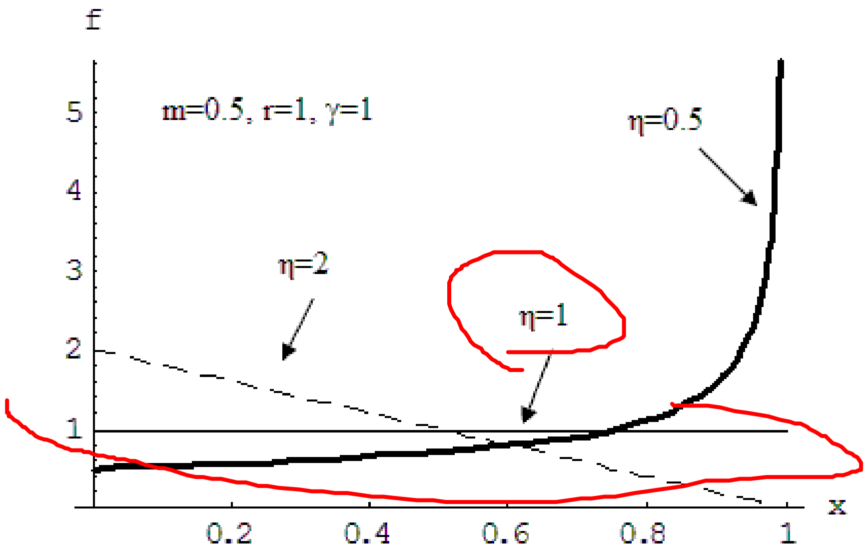
Skewness: $\dfrac{2}{\sqrt{\eta}}$ which is always positive

Kurtosis: $3 + \dfrac{6}{\eta}$

The Beta Family of Distributions

A 4-parameter distribution that is represented by a region between the gamma curve and the impossible region on a skewness-kurtosis plot.

The density function



$$f(x|m,r,\gamma,\eta)=\begin{cases} 0 & x < m-\frac{r}{2} \text{ or } x > m+\frac{r}{2} \\ \frac{\Gamma(\gamma+\eta)}{r\Gamma(\gamma)\Gamma(\eta)}\left(\frac{x-m+\frac{r}{2}}{r}\right)^{\gamma-1}\left(\frac{m+\frac{r}{2}-x}{r}\right)^{\eta-1} & m-\frac{r}{2}\le x\le m+\frac{r}{2} \end{cases}$$

Parameters:	Location:	m	$-\infty < m < \infty$
	Scale:	r	$r > 0$
	Shape:	γ	$\gamma > 0$
		η	$\eta > 0$

Bounds: Bounded below by m-r/2 and above by m+r/2.

The uniform distribution is a special case.

Moments of Beta Distribution

Mean: $$m + r\left(\frac{\gamma}{\gamma+\eta} - \frac{1}{2}\right)$$

Standard Deviation: $$r\sqrt{\frac{\eta\gamma}{(\eta+\gamma)^2(\eta+\gamma+1)}}$$

Skewness: $$\frac{2(\eta-\gamma)\sqrt{\gamma+\eta+1}}{\sqrt{\eta\gamma}(\eta+\gamma+2)}$$

Kurtosis: $$\frac{3(\eta+\gamma+1)\left[2(\eta+\gamma)^2+\gamma\eta(\eta+\gamma-6)\right]}{\eta\gamma(\eta+\gamma+2)(\eta+\gamma+3)}$$