1. **What are the three axioms of Probability theory?**
2. The first axiom of probability is that the probability of any event is a non-negative real number.

P(A) >= 0 for all A ⊂ S or

0<=P(A)<=1

1. The second axiom of probability is that the probability of the entire sample space is one.

P(S) = 1

1. Additivity: The third axiom of probability deals with mutually exclusive events. If E1 and E2 are [mutually exclusive](https://www.thoughtco.com/mutually-exclusive-3126557), meaning that they have an empty intersection and we use U to denote the union, then

P(E1 U E2 ) = P(E1) + P(E2).

1. **What is a random variable?**

A ***random variable***, usually written *X*, is a variable whose possible values are numerical outcomes of a random phenomenon. There are two types of random variables, ***discrete*** and ***continuous***.

1. ***Discrete Random Variables***

A ***discrete random variable*** is one which may take on only a countable number of distinct values such as 0,1,2,3,4, ........ Discrete random variables are usually (but not necessarily) counts. If a random variable can take only a finite number of distinct values, then it must be discrete. Examples of discrete random variables include the number of children in a family, the Friday night attendance at a cinema, the number of patients in a doctor's surgery, the number of defective light bulbs in a box of ten.

The ***probability distribution*** of a discrete random variable is a list of probabilities associated with each of its possible values. It is also sometimes called the probability function or the probability mass function.

Suppose a random variable *X* may take *k* different values, with the probability that *X = xi* defined to be *P(X = xi) = pi*. The probabilities *pi* must satisfy the following:

***1:****0 < pi < 1 for each i*

***2:****p1 + p2 + ... + pk = 1.*

1. ***Continuous Random Variables***

A ***continuous random variable*** is one which takes an infinite number of possible values. Continuous random variables are usually measurements. Examples include height, weight, the amount of sugar in an orange, the time required to run a mile.

A continuous random variable is not defined at specific values. Instead, it is defined over an *interval* of values, and is represented by the ***area under a curve*** (in advanced mathematics, this is known as an *integral*). The probability of observing any single value is equal to 0, since the number of values which may be assumed by the random variable is infinite.

Suppose a random variable *X* may take all values over an interval of real numbers. Then the probability that *X* is in the set of outcomes *A, P(A)*, is defined to be the area above *A* and under a curve. The curve, which represents a function *p(x)*, must satisfy the following:

***1:****The curve has no negative values (p(x) > 0 for all x)*

***2:****The total area under the curve is equal to 1.*

A curve meeting these requirements is known as a ***density curve***.

1. **Define Probability Density Function for a continuous random variable**

A continuous random variable takes on an uncountably infinite number of possible values. For a discrete random variable X that takes on a finite or countably infinite number of possible values, we determined P(X = x) for all of the possible values of X, and called it the probability mass function ("p.m.f."). For continuous random variables, the probability that X takes on any particular value x is 0. That is, finding P(X = x) for a continuous random variable X is not going to work. Instead, we'll need to find the probability that X falls in some interval (a, b), that is, we'll need to find P(a < X < b). We'll do that using a probability density function ("p.d.f.").

**Definition.** The**probability density function** ("**p.d.f.***"*) of a continuous random variable *X* with support *S* is an integrable function *f*(*x*) satisfying the following:

(1) *f*(*x*) is positive everywhere in the support *S*, that is, *f*(*x*) > 0, for all *x* in *S*

(2) The area under the curve *f*(*x*) in the support *S* is 1, that is:

∫Sf(x)dx=1

(3) If *f*(*x*) is the p.d.f. of *x*, then the probability that *x* belongs to *A*, where *A* is some interval, is given by the integral of *f*(*x*) over that interval, that is:

P(X∈A) = ∫Af(x)dx

1. **Name any two discrete distributions and two continuous distributions**

Probability distributions are either *continuous probability distributions* or *discrete probability distributions*, depending on whether they define probabilities for continuous or discrete variables.

**Discrete distributions:** A [statistical distribution](http://mathworld.wolfram.com/StatisticalDistribution.html) whose variables can take on only discrete values.

A discrete distribution describes the probability of occurrence of each value of a discrete random variable. A discrete random variable is a random variable that has countable values, such as a list of non-negative integers.

With a discrete probability distribution, each possible value of the discrete random variable can be associated with a non-zero probability. Thus, a discrete probability distribution is often presented in tabular form.

A discrete distribution with probability function P(x_k) defined over k=1, 2, ..., N has [distribution function](http://mathworld.wolfram.com/DistributionFunction.html)

|  |
| --- |
| D(x_n)=sum_(k=1)^nP(x_k) |

and [population mean](http://mathworld.wolfram.com/PopulationMean.html)

|  |
| --- |
| mu=1/Nsum_(k=1)^Nx_kP(x_k). |

**Example 1: -** Bernoulli Distribution

The Bernoulli distribution is a [discrete distribution](http://mathworld.wolfram.com/DiscreteDistribution.html) having two possible outcomes labelled by n=0 and n=1 in which n=1 ("success") occurs with probability *p* and n=0 ("failure") occurs with probability *q=1-p*, where 0<p<1. It therefore has [probability density function](http://mathworld.wolfram.com/ProbabilityDensityFunction.html)

|  |  |
| --- | --- |
| P(n)={1-p   for n=0; p   for n=1, |  |

which can also be written

|  |  |
| --- | --- |
| P(n)=p^n(1-p)^(1-n). |  |

The corresponding [distribution function](http://mathworld.wolfram.com/DistributionFunction.html) is

|  |
| --- |
| D(n)={1-p   for n=0; 1   for n=1. |

**Example 2: -** Binomial Distribution

The binomial distribution gives the [discrete probability distribution](http://mathworld.wolfram.com/DiscreteDistribution.html) P_p(n|N) of obtaining exactly *n* successes out of *N* [Bernoulli trials](http://mathworld.wolfram.com/BernoulliTrial.html) (where the result of each [Bernoulli trial](http://mathworld.wolfram.com/BernoulliTrial.html) is true with probability p and false with probability q=1-p). The binomial distribution is therefore given by

|  |  |  |  |
| --- | --- | --- | --- |
| P_p(n|N) | = | (N; n)p^nq^(N-n) |  |
| http://mathworld.wolfram.com/images/equations/BinomialDistribution/Inline9.gif | = | (N!)/(n!(N-n)!)p^n(1-p)^(N-n), |  |

where (N; n) is a [binomial coefficient](http://mathworld.wolfram.com/BinomialCoefficient.html). The above plot shows the distribution of *n* successes out of N=20 trials with p=q=1/2.

The probability of obtaining more successes than the n observed in a binomial distribution is

|  |  |
| --- | --- |
| P=sum_(k=n+1)^N(N; k)p^k(1-p)^(N-k)=I_p(n+1,N-n), |  |

where

|  |  |
| --- | --- |
| I_x(a,b)=(B(x;a,b))/(B(a,b)), |  |

B(a,b) is the [beta function](http://mathworld.wolfram.com/BetaFunction.html), and B(x;a,b) is the [incomplete beta function](http://mathworld.wolfram.com/IncompleteBetaFunction.html).

**Continuous distributions:** A statistical distribution for which the variables may take on a continuous range of values.

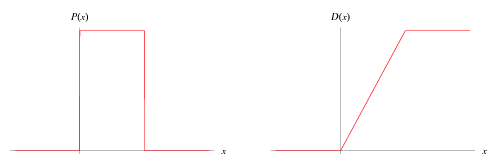
A continuous distribution describes the probabilities of the possible values of a continuous random variable. A continuous random variable is a random variable with a set of possible values (known as the range) that is infinite and uncountable.

Probabilities of continuous random variables (X) are defined as the area under the curve of its PDF. Thus, only ranges of values can have a nonzero probability. The probability that a continuous random variable equals some value is always zero.

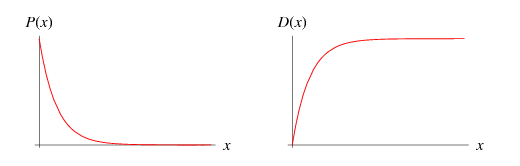
**Example 1: -** Uniform distribution

A uniform distribution, sometimes also known as a rectangular distribution, is a distribution that has constant probability.

|  |  |  |
| --- | --- | --- |
| P(x) | = | {0 for x<a; 1/(b-a) for a<=x<=b; 0 for x>b |
| D(x) | = | {0 for x<a; (x-a)/(b-a) for a<=x<=b; 1 for x>b. |



**Example 2: -** Exponential Distribution



Given a Poisson distribution with rate of change lambda, the distribution of waiting times between successive changes (with k=0) is

|  |  |  |  |
| --- | --- | --- | --- |
| D(x) | = | P(X<=x) |  |
| http://mathworld.wolfram.com/images/equations/ExponentialDistribution/Inline6.gif | = | 1-P(X>x) |  |
| http://mathworld.wolfram.com/images/equations/ExponentialDistribution/Inline9.gif | = | 1-e^(-lambdax), |  |

and the probability distribution function is

|  |
| --- |
| P(x)=D^'(x)=lambdae^(-lambdax). |

1. **Define expectation in the context of a continuous random variable**

The expected value µ = E(X) is a measure of location or central tendency

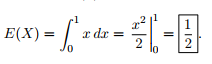
**Definition**: Let X be a continuous random variable with range [a, b] and probability density function f(x). The expected value of X is defined by,

So f(x) dx represents the probability that X is in an infinitesimal range of width dx around x. Thus we can interpret the formula for E(X) as a weighted integral of the values x of X, where the weights are the probabilities f(x) dx.

The expected value is also called the **mean** or **average.**

**Example 1**. Let X ∼ uniform(0, 1). Find E(X).

**Answer**: X has range [0, 1] and density f(x) = 1. Therefore,



The mean is at the midpoint of the range.

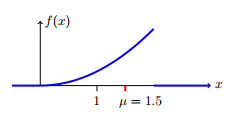
**Example 2.** Let X have range [0, 2] and density  Find E(X)

**Answer**:



**Does it make sense that this X has mean is in the right half of its range?**

**Answer**: Yes. Since the probability density increases as x increases over the range, the average value of x should be in the right half of the range.



1. **What is the value of the area under the normal curve?**

The graph of a normal distribution is a bell curve.

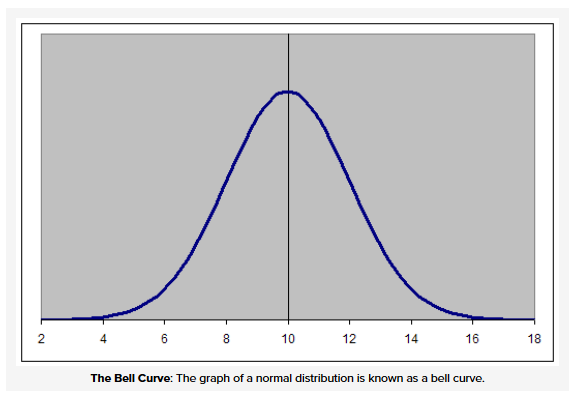
#### Key Points

* The mean of a normal distribution determines the height of a bell curve.
* The standard deviation of a normal distribution determines the width or spread of a bell curve.
* The larger the standard deviation, the wider the graph.
* Percentiles represent the area under the normal curve, increasing from left to right.

#### Key Terms

* **empirical rule**: That a normal distribution has 68% of its observations within one standard deviation of the mean, 95% within two, and 99.7% within three.
* **bell curve**: In mathematics, the bell-shaped curve that is typical of the normal distribution.
* **real number**: An element of the set of real numbers; the set of real numbers include the rational numbers and the irrational numbers, but not all complex numbers.

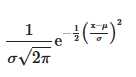
The graph of a normal distribution is a bell curve, as shown below.



The properties of the bell curve are as follows.

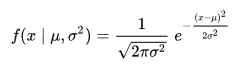
* It is perfectly symmetrical.
* It is unimodal (has a single mode).
* Its domain is all real numbers.
* The area under the curve is 1.

Different values of the mean and standard deviation determine the density factor. Mean specifically determines the height of a bell curve, and standard deviation relates to the width or spread of the graph. The height of the graph at any x value can be found through the equation:



# To Prove that Area under Standard Normal Curve is 1

The probability density of the normal distribution is:



μ = mean

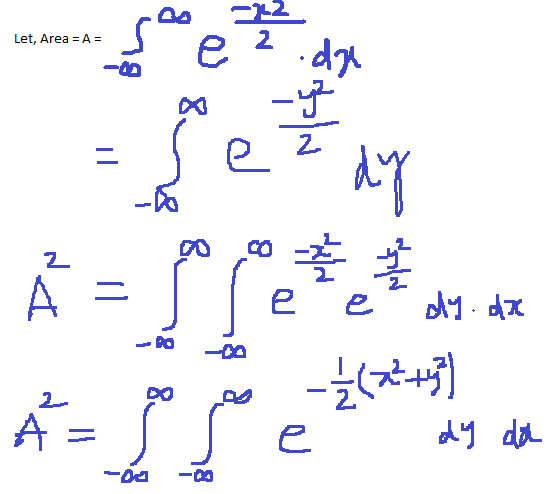
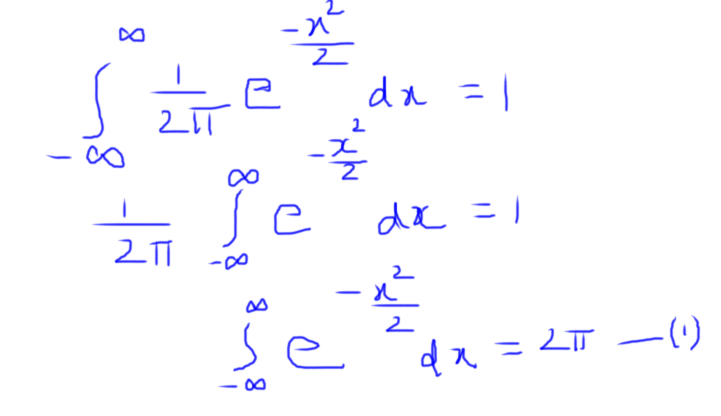
σ = standard deviation

σ² = variance

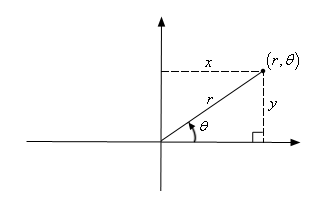
The simplest case of a normal distribution is known as the ***standard normal distribution***. This is a special case when **μ = 0 μ = 0 {\displaystyle \mu =0} and σ = 1**σ = 1 {\displaystyle \sigma =1} , and it is described by this [probability density function](https://en.wikipedia.org/wiki/Probability_density_function):

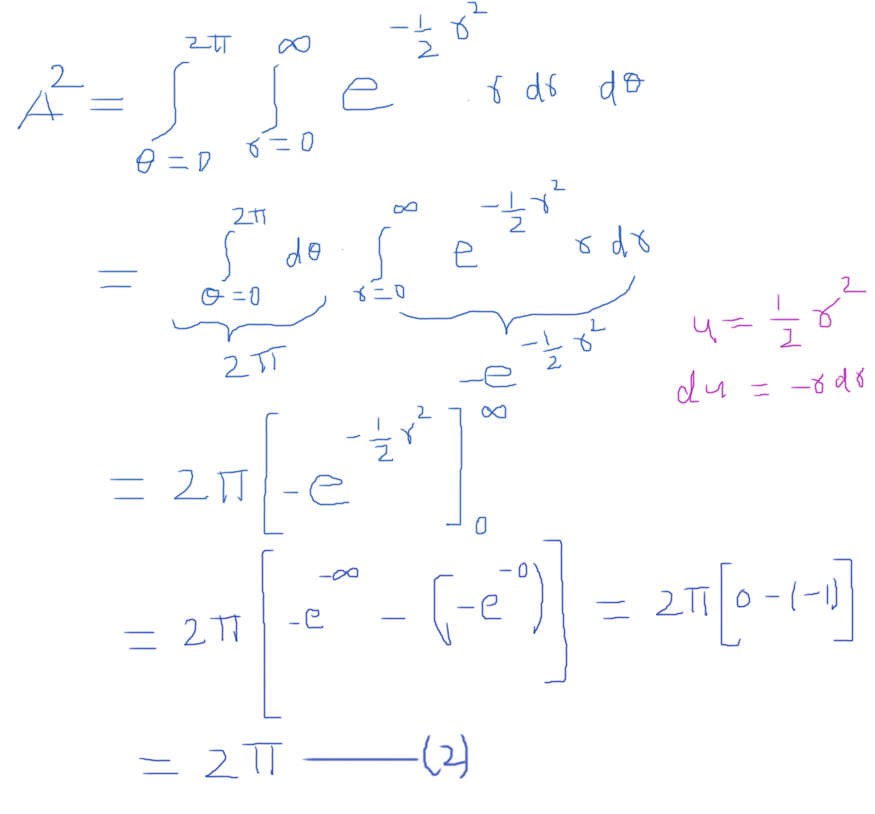


We have to prove that:



Let’s convert Cartesian (rectangular) co-ordinates x, y into polar (r, θ) form i.e. x² + y² = r²





From (1) and (2), we proved that, Area under curve is 1

1. **What are independent variables**

The variable which is not dependent on any other value(s)/variable(s) is called independent variable.

Let’s say, y = f(x) = 5x

Here x is independent variable

And y is dependent variable (dependent on x)

The value of y depends on the value chosen for x.

Even when we can establish that there are systematic differences in the value of the dependent variable for different values of the independent variable, we do not know what caused these differences. In an experiment, the analyst manipulates one variable (the so-called “independent variable”) and observes what happens to the response (the “dependent variable”). If the manipulations are properly designed, then one can be confident that observed effects are caused by the independent variable.

Many sorts of data consist of sums of independent Bernoulli Random Variables, called *Bernoulli trials*. The probability function is called the *Binomial Distribution*. Each random variable may equal 0 or 1. When a Bernoulli trial equals 1 it is called a “success”. Let X1, X2, … Xn be a series of independent random variables and the sum of these variables is X = X1 + X2 + … + Xn. The support of this random variable is the set of integers from 0 to n.

**PROBABILITY**

Consider an **event E** which is a possible outcome of a random experiment. We denote by **P(E)** the probability of this event, and think of it intuitively as the limit, as the number of trials becomes large, of the ratio of the number of times E occurred to the number of times the experiment was tried. The joint event that A and B and C, etc., occurred is denoted by ABC… , and the probability of this joint event, by P(ABC…). If these events A, B, C, etc., are **mutually *independent***, which means that the occurrence of any one of them bears no relation to the occurrence of any other, the probability of the joint event is the product of the probabilities of the simple events. That is,

P(ABC. . . ) = P(A)P(B)P(C)...

if the events A, B, C, etc., are mutually independent. Actually, the mathematical definition of independence is the reverse of this statement, but the result of consequence is that independence of events and the multiplicative property of probabilities go together.

1. **What does a set of mutually exclusive and collectively exhaustive set of events constitute?**

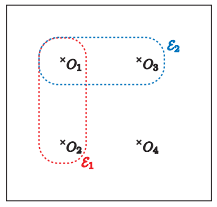


Figure 1: A Venn diagram representation of the sample space of (noutcomes = 4) possible outcomes, {O1, O2, O3, O4}, and two events E1 (red dashed line) and E2 (blue dashed line).

Let us consider two events, E1 and E2 (not necessarily the events of Figure 1).

If E1 ∩ E2 = ∅ (the empty set), we say that events E1 and E2 are **mutually exclusive**.

In words, E1 and E2 are mutually exclusive if they share no outcomes; equivalently, E1 and E2 are mutually exclusive if “E1 happens” implies “E2 does not happen” (and of course also the converse) — E1 and E2 cannot both happen. As already indicated, for any event E, E and Ē are mutually exclusive. A set of M events Em, 1 ≤ m ≤ M, is mutually exclusive if each pair of events in the set of events is mutually exclusive: no two events in the set of events share an outcome; equivalently, in any given experiment, at most only one event in the set of events can happen. In our particular example of Figure 1, the events E1, {O3}, and {O4} are mutually exclusive, however the events E1 and E2 are not mutually exclusive.

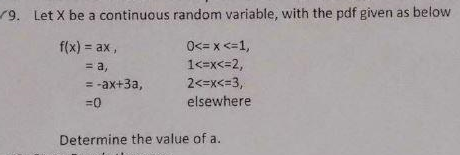
Let us again consider any two events, E1 and E2 (not necessarily the events of Figure 1).

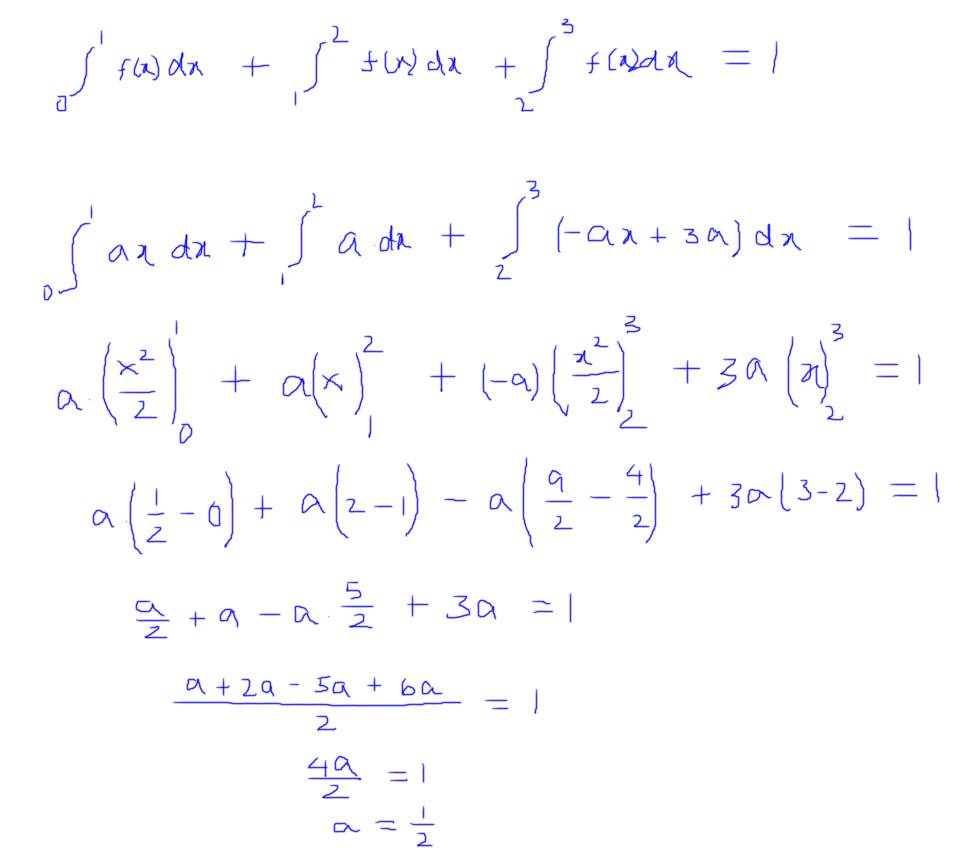
If E1 ∪ E2 is the entire sample space, we say that E1 and E2 are **collectively exhaustive**.

In words, E1 and E2 are collectively exhaustive if all possible outcomes of the experiment appear in either E1 or E2; equivalently, in any given experiment, either E1 or E2 must happen. For any event E, E and Ē are collectively exhaustive. A set of M events Em, 1 ≤ m ≤ M, is collectively exhaustive if the union of all the events in the set is the entire sample space; equivalently, in any given experiment, at least one of the events in the set of events must happen. In our particular example of Figure 1, the events E3 ≡ E1 ∪ E2, {O3}, and {O4} are collectively exhaustive, however the events E1 and E2 are not collectively exhaustive.

A pair of events, or a set of events, may be both mutually exclusive and collectively exhaustive. In this case, the result of an experiment is one event, and only one event, in the set of events: we know that at most one event can happen since the set of events is mutually exclusive; we note that at least one event must happen since the set of events is outcome collectively exhaustive. An important case is the set of n events {O1}, {O2}, . . . , {O noutcomes }. Clearly this set of events is both mutually exclusive and collectively exhaustive: an experiment can yield only one outcome, hence mutually exclusive; an experiment must yield ¯ at least one outcome, hence collectively exhaustive. The pair of events E and Ē (for any event E) is another important example of a set of mutually exclusive and collectively exhaustive events. Finally, the event ∅ — the empty set — and the sample space are yet another example of mutually exclusive and collectively exhaustive events.

1. **Let X be a continuous random variable, with the pdf given as below, determine the value of a.**





* 1. F(x) = ax, 0<=x <= 1

Substituting the value of x in f(x)

x=0 🡺 f(x) = ax = 0 🡺 (0,0)

x=1 🡺 f(x) = ax = a = ½ 🡺 (1, ½)

* 1. F(x) = a, 1<=x<=2

Substituting the value of x in f(x)

X=1 🡺 f(x) = a = ½ 🡺 (1, ½)

X=2 🡺 f(x) = a = ½ 🡺 (2, ½ )

* 1. F(x) = -ax+3a, 2<=x<=3

Substituting the value of x in f(x)

X=2 🡺 f(x) = -2a + 3a = a = ½ 🡺 (2, ½ )

X=3 🡺 f(x) = -3a + 3a = 0 🡺 (3,0)

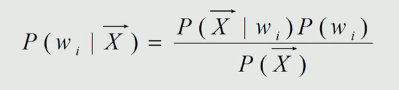
Let’s find the Area under the curve and see whether it equals to 1 or not

Area under the curve = Area of first triangle + area of reactance + area of second triangle

Area = (½ \* 1 \* 0.5) + (1 \* 0.5) + (½ \* 1 \* 0.5) = ¼ + ½ + ¼ = 1

1. **State Bayes’ theorem**

Bayes’ theorem is a way to figure out [conditional probability](http://www.statisticshowto.com/what-is-conditional-probability/). Conditional probability is the probability of an event happening, given that it has some relationship to one or more other events. For example, your probability of getting a parking space is connected to the time of day you park, where you park, and what conventions are going on at any time. Bayes’ theorem is slightly more nuanced. In a nutshell, it gives you the actual [probability](http://www.statisticshowto.com/probability-and-statistics/probability-main-index/)of an event given information about tests.



P(X) is the probability distribution for feature X in the entire population. Also called unconditional density function (or evidence)

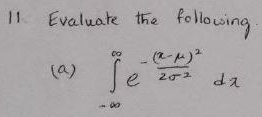
P(wi) is the prior probability that a random sample is a member of the class Ci.

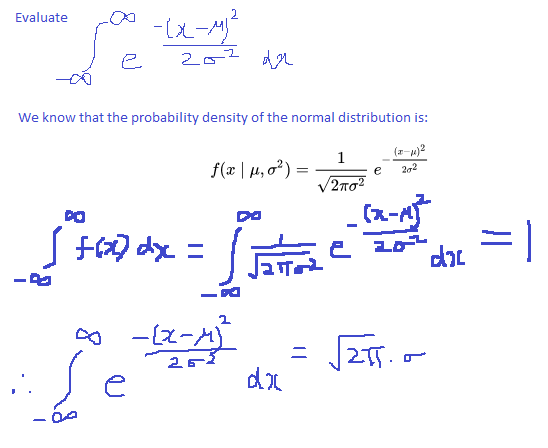
P(X | wi) is the class conditional probability (or likelihood) of obtaining feature value X given that the sample is from class wi. It is equal to the number of times (occurrence) of X, if it belongs to class wi.

The goal is to measure: P(wi | X) – Measured-conditioned or Posteriori probability, from the above three values.

This is the Probability of any vector X being assigned to class wi.

1. **Evaluate the following:**



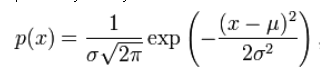


The function f(x) is called the probability density function (pdf).

The pdf always satisfies the following properties:

i. f(x) ≥ 0 (f is nonnegative).

ii. (This is equivalent to: P(−∞ < X < ∞) = 1).

Normal Distribution = 

P(x) tells us the height of the normal distribution curve.

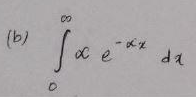
Mean = μ (-∞ < μ < ∞)

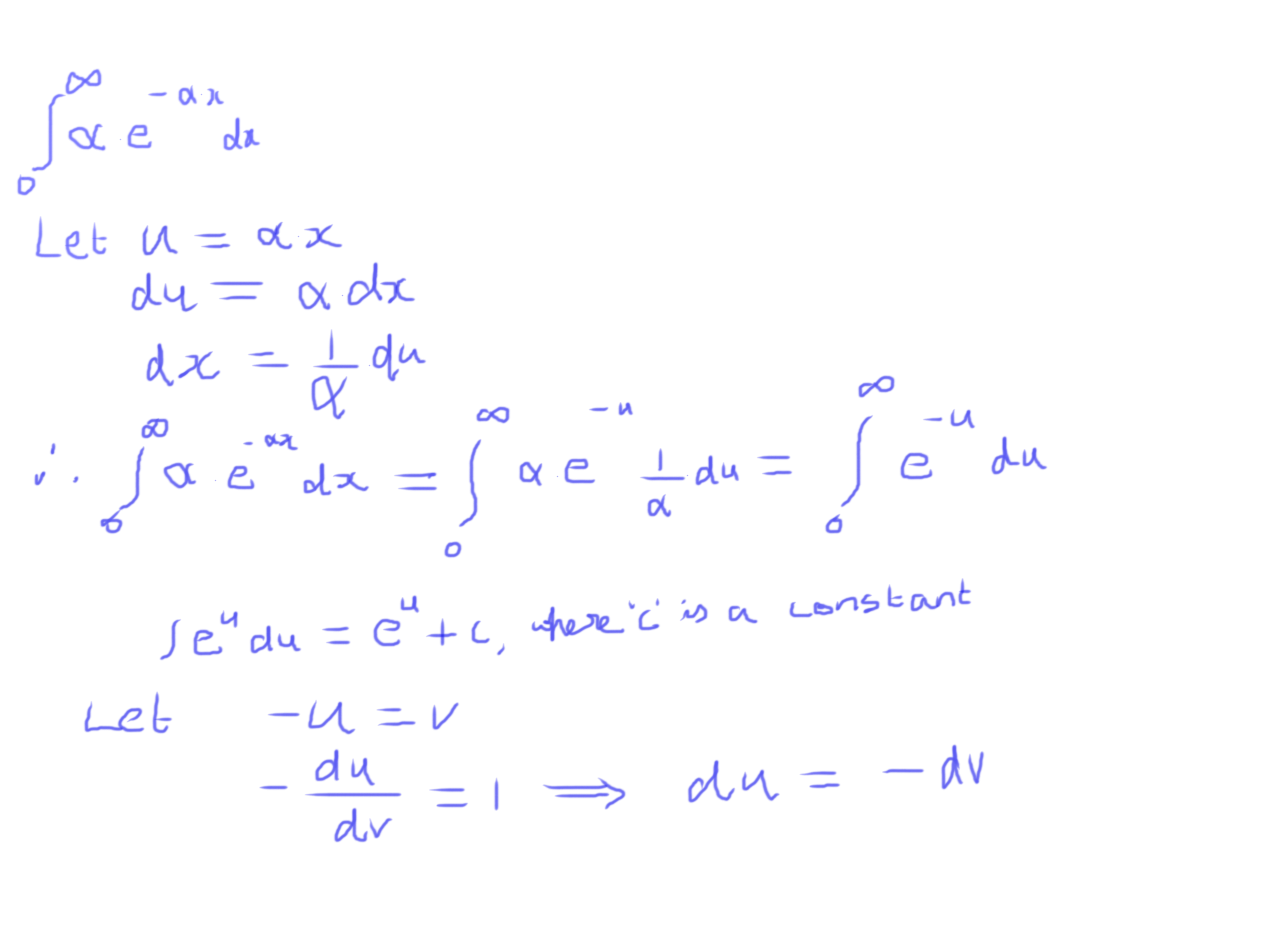
Standard Deviation = σ ( σ > 0)

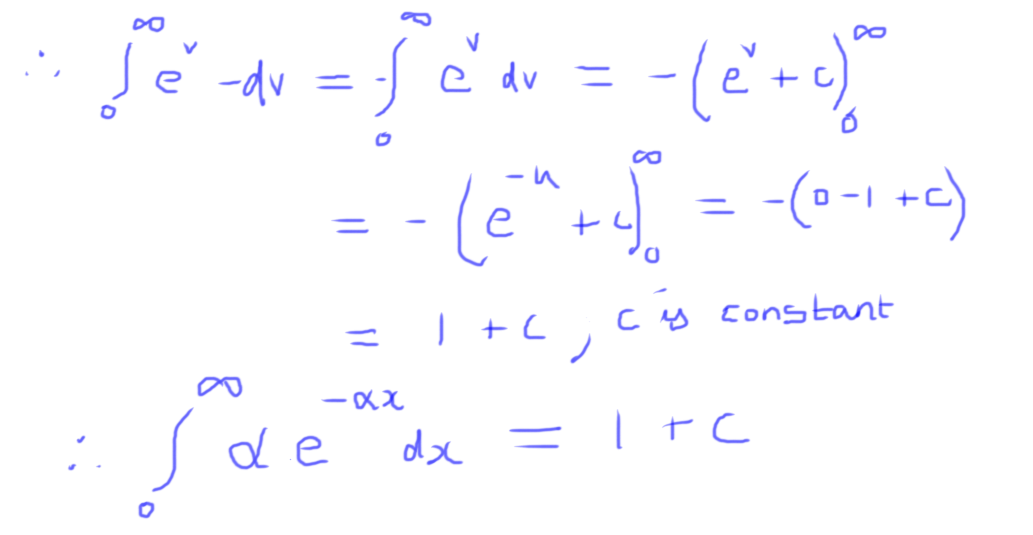
Variance = σ2 ( σ > 0)

The distribution is symmetric about μ, μ represents the mean of the probability distribution and also the median of the probability distribution

The standard normal distribution is a normal distribution with mean 0 and variance 1.







**References:**

<http://www.stat.yale.edu/Courses/1997-98/101/ranvar.htm>

<http://mathworld.wolfram.com/DiscreteDistribution.html>

<http://support.minitab.com/en-us/minitab-express/1/help-and-how-to/basic-statistics/probability-distributions/supporting-topics/basics/continuous-and-discrete-probability-distributions/>

<https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading6a.pdf>

<https://www.khanacademy.org/math/precalculus/prob-comb/dependent-events-precalc/v/independent-events-1>

<https://ocw.mit.edu/courses/mechanical-engineering/2-086-numerical-computation-for-mechanical-engineers-fall-2014/nutshells-guis/MIT2_086F14_Probability.pdf>