ECONOMICS MODULE 7 PART 2

LINEAR FUNCTIONS

In this section we begin our study of specific functions. At its simplest, a function relates one variable with another. So let us begin by defining a variable. An unknown value or an entity that can take different values is called a variable. The known value in an equation or a number that takes fixed values is called a constant. Thus a magnitude which does not change is a constant. In the context of an equation, when a constant is joined to a variable, then the a constant is called a coefficient. Sometimes in an equation a coefficient is denoted by a letter of the alphabet, rather than by a numeral. This letter is supposed to represent a constant, but doesn't have a fixed numerical value. These types of coefficients are called parameters.
The general form of the linear equation in one variable is

ax+b=c

where is the unknown or the variable and a, b and c are unspecified parameters. In the context of a specific equation, a, b and c can acquire specific numerical values, but we are speaking of general equations. We can solve this equation by expressing x in terms of the parameters.
Solving it we get,

$$x = \frac{c-b}{a}$$

The crucial points to remember about linear functions are:

a) all variables are raised to the power 1 and no other power, and
b) the slope of the graph (the line) remains constant at all point.

QUADRATIC FUNCTIONS

In the previous section, we looked at linear equations and functions. We saw that a linear equation has the general form +
= , where x is the variable

or unknown, and a,b, c are constants or parameters. This equation is called linear, because x is raised only to the power 1, and not to any lower or higher power.

A linear function has the general form = xy + ba . The graph is a straight line and the slope remains constant. This means that a one-unit change in x always increases or decreases y by the same amount (the slope can be negative). This may hold for some relationships in Economics, like a linear demand curve or supply curve. But there are many situations in economics where we need to deal with non-linear relationships, where a given change in x does not always lead to a constant change in y. In this section, we deal with a simple non-linear function, called quadratic function. We shall also discuss quadratic equations. A quadratic function, when graphed, gives a U-shaped curve. We will describe how to solve quadratic equations.

$ax^2 + bx + c = 0$, where a, b and c are constants.

To solve this equation, we begin by taking a outside the bracket, and we get

$$a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right) = 0$$

Next, we search for factors of the term inside the bracket $x^2 + \frac{b}{a}x + \frac{c}{a}$

Let us suppose the factors are $(x + C)$ and $(x + D)$, where C and D are two numbers. If these are the factors, then by definition we have,

$$(x + C)(x + D) = x^2 + \frac{b}{a}x + \frac{c}{a}$$

After multiplying both sides by a, this becomes

$$a(x + C)(x + D) \equiv ax^2 + bx + c$$

Since this is an identity, when the left-hand side becomes zero, the right-hand side also becomes zero. The left-hand side becomes zero when $x = -C$ or $x = -D$. These values are the solution to the quadratic equation $ax^2 + bx + c = 0$. There is a general formula for solving quadratic equations. Here the solutions are called *roots*. Given any quadratic equation $ax^2 + bx + c = 0$, the solution (roots) are given by the formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Cubic Functions

A Cubic function is expressed by a cubic equation which has the general form:

$$ax^3 + bx^2 + c\,x + d = 0$$

where a, b, c and d are constants. There is no general formula for solving a cubic function, but approximate solutions may be found graphically by plotting the graph and finding the points of intersection with the axes. Basic properties of the cubic function are:

1) The graph of cubic functions will have either no turning point or two turning points;

2) A cubic function will have either one root or three roots.

It is hard to factorize a cubic function, unlike a quadratic function. The shape of a cubic function is usually S-shaped. There are several contexts in which cubic functions are useful in Economics.

Like we considered a quadratic cost
function in the last section, we can have a cubic cost function as well. Consider a total cost function of the type $TC = 2q^3 + 15q^2 + 50q + 50$.

The curve of this
equation will be upward sloping, with a slightly flat surface in the beginning for lower values of q (till about q = 1.5).

Then there will be a range of q values
where the TC curve will be less flat, that is, the slope is more than that at the lowest values of q (from about q = 1.5 to q = 2.5).

It is in this middle range that the production is most efficient. At high levels of output, (particularly after q =5) the slope is very steep.

General Polynomial Functions

Quadratic and cubic functions belong to a group of functions called polynomials. The general form of a polynomial function of a single variable is:

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + a_{n-2} x^{n-2} + \dots a_1 x^1 + a_0 x^0$$

where, a0, a1, . . . , and are constants.

The 'degree' of the polynomial is given by the highest power of x in the expression. Therefore, a quadratic function is a polynomial of degree 2 and a cubic function is a polynomial of degree 3.
The Meaning of the word ‘polynomial’ is multi-term.

$$n=0, y=a_0; \quad n=1, y=a_0+a_1x; \quad n=2, y=a_0+a_1x+a_2x^2$$

In the case of Here, the first function is a constant function, the second a linear function and the third a quadratic function.
We are not going to study polynomials in depth, but you do need to be aware that they are continuous graphs, with no breaks or jumps. Therefore, it is quite safe to plot these graphs by joining the points as calculated and hence use the graphs to estimate roots and turning points.

Exponential Functions

The argument or the independent variable of an exponential function appears as an exponent. The general form of the univariate exponential function is:
$y= f(x) = b^x$
where b is called the base and is assumed to be greater than 1 (i.e., b> 1); and
x∈ R (the set of real numbers). When x= 0, y = b^0= 1,for any base. When b >1
then b
x monotonically increases with x.
Note: b > 1 restriction has been imposed, since with b = 0 or 1, we get constant functions y = 0 or y = 1, respectively. Also, with b< 0, we may get a function involving a complex value of the form√-
, as x ∈ R.Moreover, with 0 <b < 1,
we can always attain the base which is greater than 1 (for instance, consider b = 0.5 and x = 2, put these values in the exponential function to get y = 0.5^2⇒y = 2^{-2}. Here, we have b> 1 and x< 0. Since our x∈ R, such a result holds for an exponential function).
Exponential functions have a special role in economic analysis because of their use in calculating the growth of variables over time. Exponential functions also play an important role in a related problem—the calculation of the present value of a future payment.
Exponential functions are strictly monotonic and,

therefore, one-to-one. One-to-one functions have an inverse. The inverse of an
An exponential function is called a logarithmic function.
The properties of logarithmic functions are discussed in the next subsection. Logarithmic functions have a range of uses in economic analysis. These include the transformation of a non-linear relationship into a linear expression, which can be more easily evaluated; and the specification of an economic function with a constant elasticity.

Logarithmic Functions

Any exponential function has an inverse since it is strictly monotonic and, therefore, one-to-one. The classes of functions that are inverses to exponential functions are logarithmic functions. Logarithmic functions are used in many different ways in Economics. This section defines these functions and shows some of their useful properties by illustrating their application in Economic problems.

Any point (i, j) in the exponential function y = bx
, has a corresponding point (j,
i) in the logarithmic function y = logb(x).
Like the exponential function, the
logarithmic function is strictly monotonic and increasing. Logarithmic functions are everywhere concave, while exponential functions are everywhere convex. You will learn about convex and concave functions.The domain of a logarithmic function is restricted to the set of positive real numbers, while the range of the function is the set of all real numbers, which is the converse of the case for exponential functions.

The definition of a logarithmic function is as follows.
The logarithmic function
y x = $\log_b$b =1, which is read as “y is the base b logarithm of x,” satisfies the relationship b^y = x. This definition of logarithms implieslog 1 b
b = for any base,
since b^1=b, also that **$\log_b b^x x = x$.** By the definition of an inverse function,
Economic models often employ a logarithmic transformation of the variables of the model. A logarithmic transformation is the conversion of a variable that

can take on different real positive values into its logarithm. In this section we demonstrate properties of logarithmic transformations and show why this is such a useful tool for economists. Economic models frequently include nonlinear relationships. For example, real money balances are represented by the quotient of nominal balances (M) over the price level (P)— (M/P), and the real exchange rate equals the product of the nominal exchange rate (E) and the foreign price level (P*) divided by the domestic price level (P)— (EP*/P). Nonlinear relationships among variables may be expressed as linear relationships among their logarithms. Multi-equation models that include products or quotients are more difficult to solve, than models that are linear in the variables of interest. Thus expressing these models in terms of the logarithms of their variables is often a useful strategy for making analysis more straightforward.

Applications of linear functions in everyday life

The applications of linear functions in everyday life are vast. Some common applications involve solving:

- Age problems
- Speed, time and distance problems
- Geometry problems
- Percentage and money problems
- Pressure and force problems
- Salary problems

These everyday problems are converted to mathematical forms to form linear equations, which are solved using various methods. These equations should clearly explain the relationship between the data and the variables.

Linear Algebra: Matrices,Vectors, Determinants.Linear Systems

Linear algebra is a fairly extensive subject that covers vectors and matrices, determinants,systems of linear equations, vector spaces and linear transformations, eigenvalue problems,and other topics. As an area of study it has a broad appeal in that it has many applications

in engineering, physics, geometry, computer science, economics, and other areas. It also contributes to a deeper understanding of mathematics itself.

Matrices, which are rectangular arrays of numbers or functions, and vectors are the main tools of linear algebra. Matrices are important because they let us express large amounts of data and functions in an organized and concise form. Furthermore, since matrices are single objects, we denote them by single letters and calculate them directly. All these features have made matrices and vectors very popular for expressing scientific and mathematical ideas.

Matrices, Vectors:Addition and Scalar Multiplication

The basic concepts and rules of matrix and vector algebra are introduced in Secs. 7.1 and 7.2 and are followed by linear systems (systems of linear equations), a main application,

. A matrix is a rectangular array of numbers or functions which we will enclose in brackets. For example,

$$\begin{bmatrix} 0.3 & 1 & -5 \\ 0 & -0.2 & 16 \end{bmatrix}, \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

$$\begin{bmatrix} e^{-x} & 2x^2 \\ e^{6x} & 4x \end{bmatrix}, \quad [a_1 \quad a_2 \quad a_3], \quad \begin{bmatrix} 4 \\ \frac{1}{2} \end{bmatrix}$$

are matrices. The numbers (or functions) are called entries or, less commonly, elements of the matrix. The first matrix in (1) has two rows, which are the horizontal lines of entries.

Furthermore, it has three columns, which are the vertical lines of entries. The second and third matrices are square matrices, which means that each has as many rows as columns—3 and 2, respectively. The entries of the second matrix have two indices, signifying their location within the matrix. The first index is the number of the row and the second is the number of the column, so that together the entry's position is uniquely identified. For

example, (read a two three) is in Row 2 and Column 3, etc. The notation is standard and applies to all matrices, including those that are not square.

Matrices having just a single row or column are called vectors. Thus, the fourth matrix in (1) has just one row and is called a row vector. The last matrix in (1) has just one column and is called a column vector. Because the goal of the indexing of entries was to uniquely identify the position of an element within a matrix, one index suffices for vectors, whether they are row or column vectors. Thus, the third entry of the row vector in (1) is denoted byMatrices are handy for storing and processing data in applications. Consider the following two common examples.

Vectors

A vector is a matrix with only one row or column. Its entries are called the components of the vector. We shall denote vectors by lowercase boldface letters a, b, or by its general components in brackets, , and so on. Our special vectors in (1) suggest that a (general) row vector is of the form

$$\mathbf{a} = [a_1 \quad a_2 \quad \cdots \quad a_n]. \qquad \text{For instance,} \qquad \mathbf{a} = [-2 \quad 5 \quad 0.8 \quad 0 \quad 1].$$

A **column vector** is of the form

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}. \qquad \text{For instance,} \qquad \mathbf{b} = \begin{bmatrix} 4 \\ 0 \\ -7 \end{bmatrix}.$$

Matrix Multiplication

Matrix multiplication means that one multiplies matrices by matrices. Its definition is standard but it looks artificial. Thus you have to study matrix multiplication carefully, multiply a few matrices together for practice until you can understand how to do it. Here then is the definition.

Multiplication of a Matrix by a Matrix

The **product** $\mathbf{C} = \mathbf{AB}$ (in this order) of an $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ times an $r \times p$ matrix $\mathbf{B} = [b_{jk}]$ is defined if and only if $r = n$ and is then the $m \times p$ matrix $\mathbf{C} = [c_{jk}]$ with entries

$$\textbf{(1)} \qquad c_{jk} = \sum_{l=1}^{n} a_{jl} b_{lk} = a_{j1} b_{1k} + a_{j2} b_{2k} + \cdots + a_{jn} b_{nk} \qquad \begin{matrix} j = 1, \cdots, m \\ k = 1, \cdots, p. \end{matrix}$$

Linear Systems of Equations.

Gauss Elimination

We now come to one of the most important use of matrices, that is, using matrices to solve systems of linear equations. We showed informally, in Example 1 of Sec. 7.1, how to represent the information contained in a system of linear equations by a matrix, called the augmented matrix. This matrix will then be used in solving the linear system of equations. Our approach to solving linear systems is called the Gauss elimination method.

Since this method is so fundamental to linear algebra, the student should be alert.
A shorter term for systems of linear equations is just linear systems. Linear systems model many applications in engineering, economics, statistics, and many other areas. Electrical networks, traffic flow, and commodity markets may serve as specific examples of applications.

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

Fundamental Theorem for Linear Systems

(a) Existence. *A linear system of m equations in n unknowns* $x_1, \cdots, x_n$

(1)
$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

is **consistent**, *that is, has solutions, if and only if the coefficient matrix* **A** *and the augmented matrix* $\tilde{\mathbf{A}}$ *have the same rank. Here,*

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{A}} = \left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right]$$

(b) Uniqueness. *The system (1) has precisely one solution if and only if this common rank r of* **A** *and* $\tilde{\mathbf{A}}$ *equals n.*

Cramer's Rule Definition

Cramer's rule is one of the important methods applied to solve a system of equations. In this method, the values of the variables in the system are to be calculated using the determinants of matrices. Thus, Cramer's rule is also known as the determinant method.

Consider a system of linear equations with n variables $x_1, x_2, x_3, \ldots, x_n$ written in the matrix form AX = B.

Here,

A = Coefficient matrix (must be a square matrix)

X = Column matrix with variables

B = Column matrix with the constants (which are on the right side of the equations)

Now, we have to find the determinants as:

D = |A|, Dx1, Dx2, Dx3,..., Dxn

Here, Dxi for i = 1, 2, 3,..., n is the same determinant as D such that the column is replaced with B.

Thus,

x1 = Dx1/D; x2 = Dx2/D; x3 = Dx3/D; ....; xn = Dxn/D {where D is not equal to 0}

Let us consider two linear equations in two variables.

$a_1x + b_1y = c_1$

$a_2x + b_2y = c_2$

Let us write these two equations in the form of AX = B.

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

Here,

Coefficient matrix =

$$A = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$$

Variable matrix =

$$X = \begin{bmatrix} x \\ y \end{bmatrix}$$

Constant matrix =

$$B = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

$$D = |A| = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1$$

$$D_x = \begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix} = c_1b_2 - c_2b_1$$

$$D_y = \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} = a_1c_2 - a_2c_1$$

Therefore,

$x = D_x/D$

$y = D_y/D$

Inverse of a matrix and Cramer's rule

We are aware of algorithms that allow us to solve linear systems and invert a matrix. It turns out that determinants make it possible to find those by explicit formulas. For instance, if A is an n × n invertible matrix, then

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \cdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix}.$$

Note that the (i, j) entry of matrix (1) is the cofactor Aji (not Aij !). In fact the entry is Aji det(A)as we multiply the matrix by 1
det(A). **[We can divide by det(A) since it is not 0 for an invertible matrix.]**

Curiously, in spite of the simple form, formula (1) is hardly applicable for finding A-1 when n is large. This is because computing det(A) and the cofactors require too much time for such n. Notice that det(A) can be found as as soon as we know the cofactors, because of the cofactor expansion formula.

Cramer's Rule for x = A-1b We know that if Ax = b and A is nonsingular, then x = A-1b. Applying the formula A-1 = CT/det A

$$\mathbf{x} = \frac{1}{\det A} C^T \mathbf{b}.$$

gives us:

Cramer's rule gives us another way of looking at this equation. To derive this rule we break x down into its components. Because the i'th component of CTb is a sum of cofactors times some number, it is the determinant of some matrix Bj

$$x_j = \frac{\det B_j}{\det A},$$

ere B_j is the matrix created by starting with A and then replacing column j h **b**, so:

$$B_1 = \begin{bmatrix} \mathbf{b} & \begin{matrix}\text{last n-1} \\ \text{columns} \\ \text{of } A\end{matrix} \end{bmatrix} \quad \text{and}$$

$$B_n = \begin{bmatrix} \begin{matrix}\text{first n-1} \\ \text{columns} \\ \text{of } A\end{matrix} & \mathbf{b} \end{bmatrix}.$$

CHARACTERISTIC ROOTS AND VECTORS

DEFINITION OF CHARACTERISTIC ROOTS AND VECTORS

A matrix is a rectangular array of objects or
elements We will take these elements as being elements. We will take these elements as being
real numbers and indicate an element by its row
and column position.

• Let aij R denote an element of a matrix which
occupies the position of the ith row and jth
column.

• D t t i b it l l tt d it Denote a matrix by a capital letter and its
elements by the corresponding lower case letter.

If a matrix A is n x m, we write $\underset{nxm}{A}$

General

Example 1. $\underset{2\times2}{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

Example 2. $\underset{1\times1}{A} = [a_{11}]$

Example 3. $\underset{n\times n}{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$

Example 4. $\underset{2\times3}{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$

• **A matrix is said to be**
(i) square if # rows = # columns and a
square matrix is said to be
(ii) symmetric if if $a_{ij} = a_{ji}$ $\forall$i, j, i $\neq$ j.
The principle diagonal elements of a
square mat i A i b th l t trix A are given by the elements
aij, i = j.
• The principle diagonal is the ordered ntuple (a11,..., ann).
• The trace of a square matrix is defined as of a square matrix is defined as
the sum of the principal diagonal
elements. It is denoted $tr(A) = \Sigma_i a_{ii}$.

Example

• principal diagonal is (1,1,1), Tr(A) = 3

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 1 & 3 \\ 4 & 3 & 1 \end{bmatrix}$$

The Identity Matrix

• An identity matrix is a square matrix with ones in its principle diagonal and zeros elsewhere An n its principle diagonal and zeros elsewhere. An n n identity matrix is denoted In. For example

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Determinants

Definition. The minor of the element aij,

d t d |M denoted |Mij| i th d t i t f th | is the determinant of the

submatrix formed by deleting the ith row

and the jth column.

• Example: If A = [a Example: If A = [aij] is 3 x 3 then |M] is 3 x 3, then |M13|= a21a32 – a31a22. |M12| = a21a33 – a31a23.

Definition**. The cofactor of the element aij**

d t d |C |i b(1)i+j denoted |C |M | ij| given by (-1)i+j |Mij| .

Example: In the above 3 x 3 example

$|C_{13}| = a_{21}a_{32} - a_{31}a_2$

$|C_{12}| = -a_{21}a_{33} + a_{31}a_{23}$

Properties of Determinants

1. |A| =|A'|

2. The interchange of any two rows (or two col.) will change the sign of the determinant but will not change its determinant, but will not change its absolute value.

3. The multiplication of any p rows (or col) of a matrix A by a scalar k will change the value of matrix A by a scalar k will change the value of the determinant to kp |A|.

4. The addition (subtraction) of any multiple of any

row to (from) another row will leave the value
of the determinant unaltered, if the linear
combination is placed in the initial (the combination is placed in the initial (the
transformed) row slot. The same holds true if
we replace the word "row" by column.

5. If one row (col) is a multiple of another
row (l) th l f th d t i t ill (col), the value of the determinant will
be zero.

6. If A and B are square, then **|AB| = |A||B|.**

Vector Spaces, Linear Independence and Rank

- Define

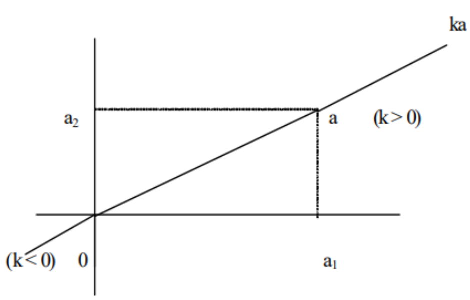
:f. An *n-component vector* a is an ordered n tuple of real numbers written as a row $(a_1 \ldots a_n) = a'$ or as a col $a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$. The $a_i, i = 1, \ldots, n$, are termed the components of the vector.

- The elements of such a vector can be viewed as the coordinates of a point in R^n or as the definition of the line segment connecting the origin and this point. We will take these as ordered n-tuples:

$$(a_1, \ldots, a_n) \in R^n$$

Two basic operations

- *Scalar multiplication*: $ka = (ka_1, \ldots, ka_n)$



Vector Space

• Def. A vector space is a collection of
vect th t i l d d th ti tors that is closed under the operations
of addition and scalar multiplication.

• Remark: Rn is a vector space.

• Def. A set of vectors span a vector space
if any vector in that space can be written if any vector in that space can be written
as a linear combination of the vectors in
that set.

Inverse Matrix

• Given an n n square matrix A, the
i ti fA d t dA 1 inverse matrix of A, denoted A i th t -1, is that
matrix which satisfies
A-1 A = A A-1 = In .
When such a matrix exists, A is said to be
nonsingular If A-1 . If A exists it is unique it is unique.

General Results for CharacteristicRoots and Vectors

- For a square matrix A, we have
- The product of the characteristic roots is equal to the i. The product of the characteristic roots is equal to the determinant of the matrix.
- The rank of A is equal to the number of nonzero characteristic roots.
- The characteristic roots of A2 are the squares of the
- characteristic roots A, but the characteristic vectors of both matrices are the same.
- . The characteristic roots of A-1 are the reciprocal of the characteristic roots of A, but the characteristic vectors of Both matrices are the same.

General Results on the Trace of a Matrix

- $tr(cA) = c(tr(A))$.
- $tr(A') = tr(A)$.
- $tr(A+B) = tr(A) + tr(B)$.
- $tr(I_k) = k$.
- $tr(AB) = tr(BA)$.

(Note this can be extended to any permutation: $tr(ABCD) = tr(BCDA) = tr(CDAB) = tr(DABC)$.

DEFINITIONS OF PROBABILITY

The term probability has been interpreted in terms of four definitions viz.,

1) Classical definition.

2) Axiomatic definition.

3) Empirical definition.

4) Subjective definition.

1) Classical Definition

The classical definition states that if an experiment consists of N outcomes

which are mutually exclusive, exhaustive and equally likely and NA of them

are favorable to an event A, then the probability of the event A (P (A)) is defined as

P (A) = NA / N

In other words, the probability of an event A equals the ratio of the number of outcomes NA favorable to A to the total number of outcomes. See the following example for a better understanding of the concept.

Example1: Two unbiased dice are thrown simultaneously. Find the probability that the product of the points appearing on the dice is 18.

There are 36 (N) possible outcomes if two dice are thrown simultaneously.

These outcomes are mutually exclusive, exhaustive and equally likely based

on the assumption that the dice are unbiased. Now we denote A: the product

of the points appearing on the dice is 18.

The events favorable to 'A' are [(3, 6), (6, 3)] only, therefore, NA = 2.

According to classical definition of probability

P (A) = NA / N = 1/18

When none of the outcomes is favorable to the event A, NA= 0, P (A) also takes the value 0, in that case we say that event A is impossible.

There are many defects of the classical definition of probability. Unless the outcomes of an event are mutually exclusive, exhaustive and equally likely,classical definitions cannot be applied. Again, if the number of outcomes of an event is infinitely large, the definition fails. The phrase 'equally likely appearing in the classical definition of probability means equally probable,thus the definition is circular in nature.

2) Axiomatic Definition

In the axiomatic definition of probability, we start with a probability space 'S' where the set 'S' of abstract objects is called outcomes. The set S and its subsets are called events. The probability of an outcome A is by definition a number P

(A) assigned to A. Such a number satisfies the following axioms:

a) P (A) ≥ 0 i.e., P (A) is a nonnegative number.

b) The probability of the certain event S is 1, i.e., P (S) = 1.

c) If two events A and B have no common elements, or, A and B are mutually exclusive, the probability of the event (A U B) consisting of the outcomes that are in A or in B equals to sum of their probabilities:

P (A U B) = P (A) + P (B)

The axiomatic definition of probability is a relatively recent concept . However, the axioms and the results stated above had been used earlier.

Kolmogoroff's contribution was the interpretation of probability as an abstract concept and the development of the theory as a pure mathematical discipline.

We comment next on the connection between an abstract sample space and the underlying real experiment. The first step in model formation is between elements of S and experimental outcomes. The actual outcomes of a real experiment can involve a large number of observable characteristics. In the formation of the model, we select from these characteristics the one that is of interest in our investigation.

For example, consider the possible models of the throwing of an unbiased die by the 3 players X, Y and Z.

X says that the outcomes of this consist of six faces of the die, forming the sample space {1,2,3,4,5,6}.

Y argues that the experiment has only 2 outcomes, even or odd, forming the sample space {even, odd}

Z bets that the die will rest on the left side of the table and the face with one point will show. Her experiment consists of infinitely many points consisting of the six faces of the die and the coordinate of the table where the die rests finally.

3) Empirical Definition

In N trials of a random experiment if an event is found to occur m times, the relative frequency of the occurrence of the event is m/N. If this relative frequency approaches a limiting value p, as N increases indefinitely, then 'p' is called the probability of the event A.

$$P(A) = \lim_{N \to \infty} \left(\frac{m}{N} \right)$$

To give a meaning to the limit we must interpret the above formula as an assumption used to define P(A). This concept was introduced by Von Mises.However, the use of such a definition as a basis of deductive theory has not enjoyed wide acceptance.

4) Subjective Definition

In the subjective interpretation of probability, the number P (A) is assigned to a statement. A, which is a measure of our state of knowledge or belief

concerning the truth of A. These kinds of probabilities are most often used in our daily life and conversations. We often make statements like "I am 100% sure that I will pass the examination" i.e., P(of passing the examinations) = 1,

or "there is 50% chance that India will win the match against Pakistan" i.e., P(India will win the match against Pakistan)= ½

Theorem of Total Probability

If two events A and B are mutually exclusive, exhaustive and equally likely, then the occurrence of either A or B, (A U B) is given by the sum of their probability. Thus,

P (A U B) = P(A) + P(B)

This is also known as the **Addition Theorem.**

Proof: Let us assume that a random experiment has n possible outcomes which are mutually exclusive, exhaustive and equally likely. While m1 of them are favorable to A, m2 are favorable to B. By the classical definition of probability

P(A) = m1 / n and P(B) = m2 / n

Since A and B are mutually exclusive and exhaustive, the number of events favorable to the event (A U B) is given by m1 + m2, therefore,

P(A U B) = (m1 + m2) / n = (m1 / n) + (m2 / n) = P(A) + P(B) (proved)

Deductions from Theorem of Total Probability

1) Theorem of Complementary Event

If A denotes the occurrence of the event A, then Ac (read as 'compliment of

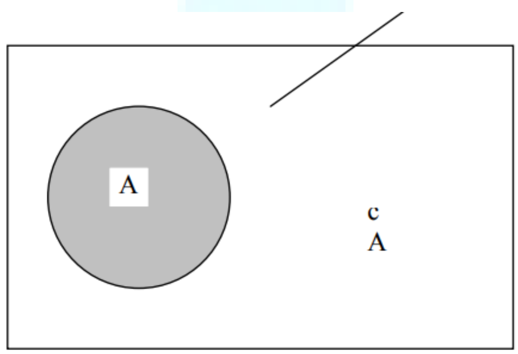
A') denotes non occurrence of the event A and P(A) = 1 - P(Ac

Since A and Ac are mutually exclusive and exhaustive events, S = {A,Ac

Applying the theorem of total probability we get,

P(S) = P(A) + P(Ac) = 1

or, P(Ac) = 1 - P(A)



Theorem of Compound Probability

The probability of occurrence of the event A and B simultaneously is given by

the product of the probability of the event A and conditional probability of the

event B given that A has actually occurred, which is denoted by P(A/B).

P(A/B) is given by the ratio of the number of events favorable to the event A

and B to the number of events favorable to the event A. Symbolically,
P(A I B) = P(A) × P(B/A).

Proof: Suppose a random experiment has n mutually exclusive, exhaustive
and equally likely outcomes among which m1, m2 and m12 are favorable to the
events A, B and (A I B) respectively.

P (A I B) = m12 / n

= m1/n × m12 / m1

= P(A) × P(B/A) (Proved).

This theorem is also known as the multiplication theorem.

Deductions from Theorem of Total Probability

The occurrence of one event, say, B may be associated with the occurrence or non-occurrence of another event, say, A. This in turn implies that we can think of B to be composed of two mutually exclusive events (A I B) and (AcI B). Applying the theorem of total probability

P(B) = P(A I B) + P(Ac I B)= P(A) × P(B/A) + P(Ac) × P(B/Ac)… [using theorem of compound probability]

1) Extension of Compound Probability Theorem

The above theorem can be extended to include the cases when there are three or more events. Suppose there are three events A, B and C, then

P(A I B I C) = P(A) × P(B/A) × P(C/(A I B)

And so on for more than three events.

CONDITIONAL PROBABILITY AND
CONCEPT OF INDEPENDENCE

15.6.1 Conditional Probability

From the theorem of compound probability we can get the probability of one
even, say, event B conditioned on some other event, say A. As we have

discussed earlier, this is symbolically written as P(B/A). From the theory of
compound probability, we know that
P(A I B) = P(A) × P(B/A)
or, P(B/A) = P(A I B) / P(A) provided that P(A) ≠ 0.
Example 4: Find out the probability of getting the Ace of hearts when one card is drawn from a well-shuffled pack of cards given the fact that the card is
red.
Let A denotes the event that the card is red and B denotes the event that the
the card is the Ace of hearts. Then clearly we are interested in finding P (B/A).
From the theorem of conditional of probability
P (B/A) = P(A I B) / P(A) = (1/52)/(26/52) = 1/26
Concept of Independent Events
Two events A and B are said to be statistically independent if the occurrence
one event is not affected by the occurrence of another event. Similarly, several events are said to be independent, mutually independent or statistically
independent if the occurrence of one event is not affected by the supplementary knowledge of the occurrence of other events. These imply that
P(B/A) = P(B/Ac) = P(B)
Therefore, from the theorem of compound probability, we get
P (A I B) = P(A) × P(B/A)
= P (A) × P (B)
Similarly, for three events we have the following results is that events are mutually or statistically independent
P(A I B I C) = P(A) × P(B) × P(C) along with

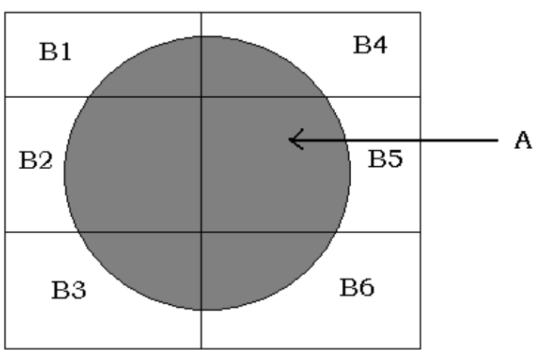
P (A I B) = P (A) × P (B)
P (C I B) = P (C) × P (B)
P (C I A) = P (C) × P (A)

For more events A, B, C, D to be mutually independent following should hold:

P(A I BI C I D) = P(A) × P(B) × P(C) × P(D) along with
P(A I B I C) = P(A) × P(B) × P(C)
P(A I B I D) = P(A) × P(B) × P(D)
P(D I B I C) = P(D) × P(B) × P(C)
P(A I D I C) = P(A) × P(D) × P(C)
P (AI B) = P (A) × P (B)

BAYES' THEOREM AND ITS APPLICATION

Suppose an event A can occur if and only if one of the mutually exclusive events B1, B2, B3,................., Bn occurs. If the unconditional probabilities P(B1), **P(B2), P(B3),........, P(Bn) are known and the conditional probabilities are P(A/B1),**



P(A /B3),.........., P(A /Bn) are also known. Then the conditional probability P(Bi/A) could be calculated when A has actually occurred.

P(A) = 1

i=ΣP(AI Bi) = 1
i=ΣP(Bi) P(A/Bi)
P(Bi/A) = P(Bi I A) / P(A) = P(A /Bi)×P(A) / 1

i=ΣP(Bi) P(A/Bi), therefore

P(Bi/A) = P(A /Bi)×P(A) / 1n

i=ΣP(Bi) P(A/Bi)

This is known as Bayes' theorem. This is a very strong result in the theory of probability. An example will illustrate the theorem more vividly.

Empirical Distributions

An empirical distribution is one for which each possible event is assigned a probability derived from experimental observation. It is assumed that the events are independent and the sum of the probabilities is 1.

An empirical distribution may represent either a continuous or a discrete distribution. If it represents a discrete distribution, then sampling is done "on step". If it represents a continuous distribution, then sampling is done via

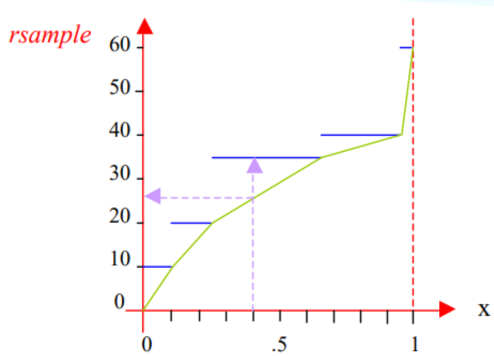
"interpolation". The way the table is described usually determines if an empirical distribution is to be handled discretely or continuously; e.g.,

discrete description

value	probability
10	.1
20	.15
35	.4
40	.3
60	.05

continuous description

value	probability
$0 - 10^-$	.1
$10 - 20^-$	.15
$20 - 35^-$	.4
$35 - 40^-$	.3
$40 - 60^-$	.05



Discrete Distributions

To put a little historical perspective behind the names used with these distributions, James Bernoulli (1654-1705) was a Swiss mathematician whose book Ars Conjectandi (published posthumously in 1713) was the first significant book on probability; it gathered together the ideas on counting, and among other things provided a proof of the binomial theorem. Siméon-Denis Poisson (1781- 1840) was a professor of mathematics at the Faculté des Sciences whose 1837 text

Recherchés sur la probabilité des jugements en matière criminelle et en matière civile introduisant la discrete distribution now call the Poisson distribution. Keep in mind that scholars such as these evolved their theories with the objective of providing sophisticated abstract models of real-world phenomena (an effort which, among other things, gave birth to the calculus as a major modeling tool)

I. Bernoulli Distribution

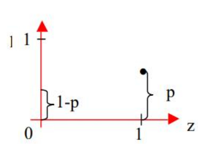
A Bernoulli event is one for which the probability the event occurs is p and the probability the event does not occur is 1-p; i.e., the event is has two possible outcomes (usually viewed as success or failure) occurring with probability p and 1-p, respectively. A Bernoulli trial is an instantiation of a Bernoulli event. So long as the probability of success or failure remains the same from trial to trial (i.e., each trial is independent of the others), a

The sequence of Bernoulli trials is called a Bernoulli process. Among other conclusions that could be reached, this means that for n trials, the probability of n successes is pn

.

A Bernoulli distribution is the pair of probabilities of a Bernoulli event, which is too simple to be interesting. However, it is implicitly used in "yesno" decision processes where the choice occurs with the same probability from trial to trial (e.g., the customer chooses to go down aisle 1 with probability p) and can be case in the same kind of mathematical notation used to describe more complex distributions:

$$p(z) = \begin{cases} p^z(1-p)^{1-z} & \text{for } z = 0,1 \\ 0 & \text{otherwise} \end{cases}$$



The expected value of the distribution is given by

$$E(X) = (1-p)\cdot 0 + p\cdot 1 = p$$

The standard deviation is given by

$$\sqrt{p\cdot(1-p)}$$

Binomial Distribution

The Bernoulli distribution represents the success or failure of a single Bernoulli trial. The Binomial Distribution represents the number of successes and failures in n independent Bernoulli trials for some given value of n. For example, if a manufactured item is defective with probability p, then the binomial distribution represents the number of successes and failures in a lot of n items. In particular, sampling from this distribution gives a count of the number of defective items in a sample lot. Another
An example is the number of heads obtained in tossing a coin n times.

The binomial distribution gets its name from the binomial theorem which states that the binomial

$$(a+b)^n = \sum_0^n \binom{n}{k} a^k b^{n-k} \text{ where } \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

It is worth pointing out that if a = b = 1, this becomes

$$(1+1)^n = 2^n = \sum_0^n \binom{n}{k}$$

Yet another viewpoint is that if S is a set of size n, the number of k element subsets of S is given by $\frac{n!}{k!(n-k)!} = \binom{n}{k}$

This formula is the result of a simple counting analysis: there are

$$n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

ordered ways to select k elements from n (n ways to choose the 1st item, (n-1) the 2nd , and so on). Any given selection is a permutation of its k elements, so the underlying subset is counted k! times. Dividing by k! eliminates the duplicates.

Note that the expression for 2^n counts the total number of subsets of an n-element set.

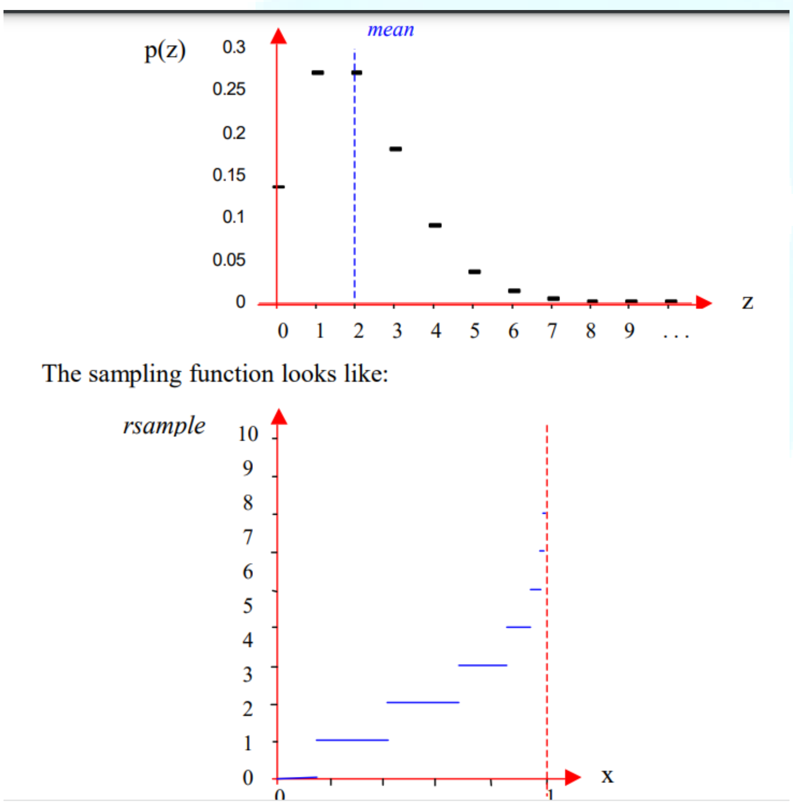
Poisson Distribution (values n = 0, 1, 2, . . .)

The Poisson distribution is the limiting case of the binomial distribution where p ® 0 and n ® ¥. The expected value E(X) = l where np ® l as p ® 0 and n ® ¥. The standard deviation is l . The pdf is given by

This distribution dates back to Poisson's 1837 text regarding civil and criminal matters, in effect scotching the tale that its first use was for modeling deaths from the kicks of horses in the Prussian army. In addition to modeling the number of arrivals over some interval of time (recall the relationship to the exponential distribution; a Poisson process has exponentially distributed interarrival times),

The distribution has also been used to model the number of defects on a manufactured article. In general the Poisson distribution is used for situations where the probability of an event occurring is very small, but the number of trials is very large (so the event is expected to actually occur a few times).

Graphically, with Λ = 2, it appears as:



Geometric Distribution

The geometric distribution gets its name from the geometric series:

$$\text{for } r<1,\ \sum_{0}^{\infty} r^n = \frac{1}{1-r},\quad \sum_{0}^{\infty} n\cdot r^n = \frac{r}{(1-r)^2},\quad \sum_{0}^{\infty}(n+1)\cdot r^n = \frac{1}{(1-r)^2}$$

various flavors of the geometric series

The pdf for the geometric distribution is given by

$$p(z) = \begin{cases} (1-p)^{z-1}\cdot p & \text{for } z=1,2,\dots \\ 0 & \text{otherwise} \end{cases}$$

Like the exponential distribution, it is "memoryless" (and is the only discrete distribution with this property; see the discussion of the exponential distribution).

Its expected value is given by

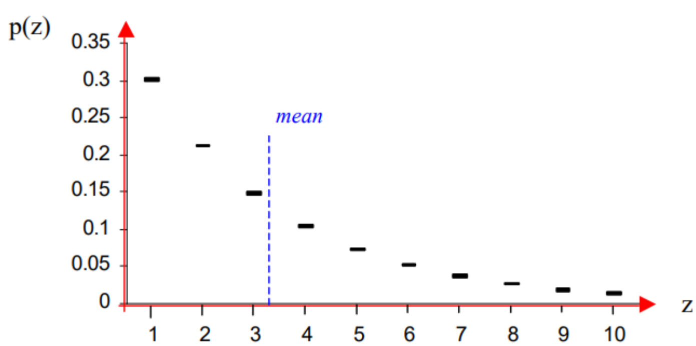
$$E(X) = \sum_{1}^{\infty} z(1-p)^{z-1}\cdot p = p\cdot\frac{1}{(1-1+p)^2} = \frac{1}{p}$$

(by applying the 3rd form of the geometric series).

The standard deviation is given by $\frac{\sqrt{1-p}}{p}$.

p

A plot of the geometric distribution with p = 0.3 is given by



Negative Binomial Distribution

The negative binomial distribution is a discrete probability distribution of the number of failures in a sequence of iid Bernoulli trials with probability of success p before a specified (non-random) number of successes (denoted r) occurs. It is also called Pascal Distribution (when r is an integer).

For x+r Bernoulli trials with success probability p, the negative binomial gives the probability of x failures and r successes, with a success on the last trial.

Its pmf:

Its pmf:

$$f(x|p,r) = \frac{\Gamma(x+r)}{\Gamma(r)x!}p^r(1-p)^x$$

Expected value:

$$\frac{r(1-p)}{p}$$

Variance:

$$\frac{r(1-p)}{p^2}$$

Normal Distribution

The normal(μ,σ2) distribution is used for continuous random variables that can take any value -∞≤x<∞.

The normal distribution is immensely useful because of the central limit theorem, which states that, under mild conditions, the mean of many random variables independently drawn from the same distribution is distributed approximately normally, irrespective of the form of the original distribution

Given $Y \sim N(\mu, \sigma^2)$:

Its probability density function,

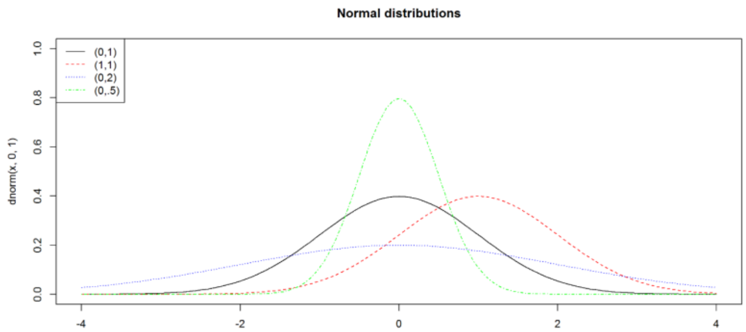
$$f(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

Expected Value:

$$E(Y) = \mu$$

Variance:

$$Var(Y) = \sigma^2$$



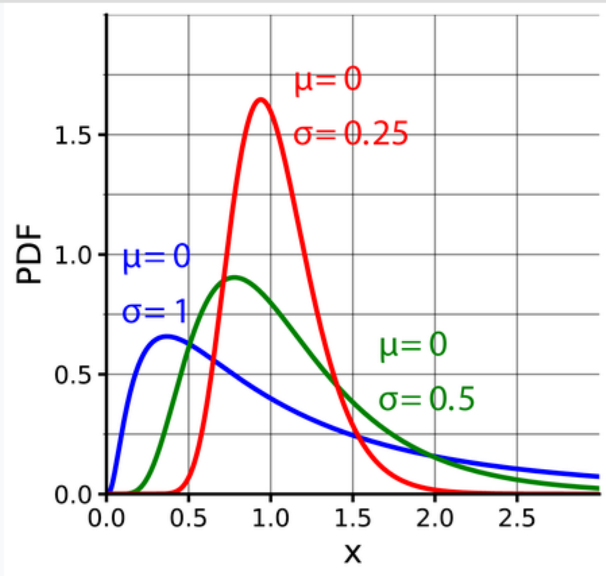
Log-normal distribution

In probability theory, a log-normal (or lognormal) distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable X is log-normally distributed, then Y = ln(X) has a normal distribution.[1][2] Equivalently, if Y has a normal distribution, then the exponential function of Y, X = exp(Y), has a log-normal distribution. A random variable which is log-normally distributed takes only positive real values. It is a convenient and useful model for measurements in exact and engineering sciences, as well as medicine, economics and other topics (e.g., energies, concentrations, lengths, prices of financial instruments, and other metrics).

The distribution is occasionally referred to as the Galton distribution or Galton's distribution, after Francis Galton.[3] The log-normal distribution has also been associated with other names, such as McAlister, Gibrat and Cobb–Douglas.

Log-normal

Probability density function



Identical parameter μ but differing parameters σ

A log-normal process is the statistical realization of the multiplicative product of many independent random variables, each of which is positive. This is justified by considering the central limit theorem in the log domain (sometimes called Gibrat's law). The log-normal distribution is the maximum entropy probability distribution for a random variate X—for which the mean and variance of ln(X) are specified.