Binary logarithm

In mathematics , the binary logarithm (log2 n) is the power to which the number 2 must be raised to obtain the value n . That is , for any real number x ,
<formula>
For example , the binary logarithm of 1 is 0 , the binary logarithm of 2 is 1 , the binary logarithm of 4 is 2 , and the binary logarithm of 32 is 5 .
The binary logarithm is the logarithm to the base 2 . The binary logarithm function is the inverse function of the power of two function . As well as log2 , alternative notations for the binary logarithm include lg , ld , lb , and (with a prior statement that the default base is 2) log .
Historically , the first application of binary logarithms was in music theory , by Leonhard Euler : the binary logarithm of a frequency ratio of two musical tones gives the number of octaves by which the tones differ . Binary logarithms can be used to calculate the length of the representation of a number in the binary numeral system , or the number of bits needed to encode a message in information theory . In computer science , they count the number of steps needed for binary search and related algorithms . Other areas in which the binary logarithm is frequently used include combinatorics , bioinformatics , the design of sports tournaments , and photography .
Binary logarithms are included in the standard C mathematical functions and other mathematical software packages . The integer part of a binary logarithm can be found using the find first set operation on an integer value , or by looking up the exponent of a floating point value . The fractional part of the logarithm can be calculated efficiently .

History

The powers of two have been known since antiquity ; for instance they appear in Euclid 's Elements , Props . IX.32 (on the factorization of powers of two) and IX.36 (half of the Euclid ? Euler theorem , on the structure of even perfect numbers) . And the binary logarithm of a power of two is just its position in the ordered sequence of powers of two . On this basis , Michael Stifel has been credited with publishing the first known table of binary logarithms in 1544 . His book Arthmetica Integra contains several tables that show the integers with their corresponding powers of two . Reversing the rows of these tables allow them to be interpreted as tables of binary logarithms .
Earlier than Stifel , the 8th century Jain mathematician Virasena is credited with a precursor to the binary logarithm . Virasena 's concept of ardhacheda has been defined as the number of times a given number can be divided evenly by two . This definition gives rise to a function that coincides with the binary logarithm on the powers of two , but it is different for other integers , giving the 2 @-@ adic order rather than the logarithm .
The modern form of a binary logarithm , applying to any number (not just powers of two) was considered explicitly by Leonhard Euler in 1739 . Euler established the application of binary logarithms to music theory , long before their more significant applications in information theory and computer science became known . As part of his work in this area , Euler published a table of binary logarithms of the integers from 1 to 8 , to seven decimal digits of accuracy .

Definition and properties

The binary logarithm function may be defined as the inverse function to the power of two function , which is a strictly increasing function over the positive real numbers and therefore has a unique inverse . Alternatively , it may be defined as ln n / ln 2 , where ln is the natural logarithm , defined in any of its standard ways . Using the complex logarithm in this definition allows the binary logarithm to be extended to the complex numbers .
As with other logarithms , the binary logarithm obeys the following equations , which can be used to simplify formulas that combine binary logarithms with multiplication or exponentiation :
<formula>
<formula>

<formula>
For more , see list of logarithmic identities .

= = Notation = =

In mathematics , the binary logarithm of a number n is often written as log2 n . However , several other notations for this function have been used or proposed , especially in application areas .
Some authors write the binary logarithm as lg n , the notation listed in The Chicago Manual of Style . Donald Knuth credits this notation to a suggestion of Edward Reingold , but its use in both information theory and computer science dates to before Reingold was active . The binary logarithm has also been written as log n with a prior statement that the default base for the logarithm is 2 . Another notation that is sometimes used for the same function (especially in the German scientific literature) is ld n , from Latin logarithmus dualis . The DIN 1302 , ISO 31 @-@ 11 and ISO 80000 @-@ 2 standards recommend yet another notation , lb n . According to these standards , lg n should not be used for the binary logarithm , as it is instead reserved for log10 n .

= = Applications = =

= = = Information theory = = =

The number of digits (bits) in the binary representation of a positive integer n is the integral part of 1 + log2 n , i.e.
<formula>
In information theory , the definition of the amount of self @-@ information and information entropy is often expressed with the binary logarithm , corresponding to making the bit the fundamental unit of information . However , the natural logarithm and the nat are also used in alternative notations for these definitions .

= = = Combinatorics = = =

Although the natural logarithm is more important than the binary logarithm in many areas of pure mathematics such as number theory and mathematical analysis , the binary logarithm has several applications in combinatorics :
Every binary tree with n leaves has height at least log2 n , with equality when n is a power of two and the tree is a complete binary tree . Relatedly , the Strahler number of a river system with n tributary streams is at most log2 n + 1 .
Every family of sets with n different sets has at least log2 n elements in its union , with equality when the family is a power set .
Every partial cube with n vertices has isometric dimension at least log2 n , and has at most 1 / 2 n log2 n edges , with equality when the partial cube is a hypercube graph .
According to Ramsey 's theorem , every n @-@ vertex undirected graph has either a clique or an independent set of size logarithmic in n . The precise size that can be guaranteed is not known , but the best bounds known on its size involve binary logarithms . In particular , all graphs have a clique or independent set of size at least 1 / 2 log2 n (1 ? o (1)) and almost all graphs do not have a clique or independent set of size larger than 2 log2 n (1 + o (1)) .
From a mathematical analysis of the Gilbert ? Shannon ? Reeds model of random shuffles , one can show that the number of times one needs to shuffle an n @-@ card deck of cards , using riffle shuffles , to get a distribution on permutations that is close to uniformly random , is approximately 3 / 2 log2 n . This calculation forms the basis for a recommendation that 52 @-@ card decks should be shuffled seven times .

= = = Computational complexity = = =

The binary logarithm also frequently appears in the analysis of algorithms , not only because of the frequent use of binary number arithmetic in algorithms , but also because binary logarithms occur in the analysis of algorithms based on two @-@ way branching . If a problem initially has n choices for its solution , and each iteration of the algorithm reduces the number of choices by a factor of two , then the number of iterations needed to select a single choice is again the integral part of log2 n . This idea is used in the analysis of several algorithms and data structures . For example , in binary search , the size of the problem to be solved is halved with each iteration , and therefore roughly log2 n iterations are needed to obtain a problem of size 1 , which is solved easily in constant time . Similarly , a perfectly balanced binary search tree containing n elements has height log2 (n + 1) ? 1 .

The running time of an algorithm is usually expressed in big O notation , which is used to simplify expressions by omitting their constant factors and lower @-@ order terms . Because logarithms in different bases differ from each other only by a constant factor , algorithms that run in O (log2 n) time can also be said to run in , say , O (log13 n) time . The base of the logarithm in expressions such as O (log n) or O (n log n) is therefore not important and can be omitted . However , for logarithms that appear in the exponent of a time bound , the base of the logarithm cannot be omitted . For example , O (2log2 n) is not the same as O (2ln n) because the former is equal to O (n) and the latter to O (n0.6931 ...) .

Algorithms with running time O (n log n) are sometimes called linearithmic . Some examples of algorithms with running time O (log n) or O (n log n) are :

Average time quicksort and other comparison sort algorithms
Searching in balanced binary search trees
Exponentiation by squaring
Longest increasing subsequence

Binary logarithms also occur in the exponents of the time bounds for some divide and conquer algorithms , such as the Karatsuba algorithm for multiplying n @-@ bit numbers in time O (nlog2 3) , and the Strassen algorithm for multiplying n × n matrices in time O (nlog2 7) . The occurrence of binary logarithms in these running times can be explained by reference to the master theorem .

= = = Bioinformatics = = =

In bioinformatics , microarrays are used to measure how strongly different genes are expressed in a sample of biological material . Different rates of expression of a gene are often compared by using the binary logarithm of the ratio of expression rates : the log ratio of two expression rates is defined as the binary logarithm of the ratio of the two rates . Binary logarithms allow for a convenient comparison of expression rates : a doubled expression rate can be described by a log ratio of 1 , a halved expression rate can be described by a log ratio of ? 1 , and an unchanged expression rate can be described by a log ratio of zero , for instance .

Data points obtained in this way are often visualized as a scatterplot in which one or both of the coordinate axes are binary logarithms of intensity ratios , or in visualizations such as the MA plot and RA plot that rotate and scale these log ratio scatterplots .

= = = Music theory = = =

In music theory , the interval or perceptual difference between two tones is determined by the ratio of their frequencies . Intervals coming from rational number ratios with small numerators and denominators are perceived as particularly euphonious . The simplest and most important of these intervals is the octave , a frequency ratio of 2 : 1 . The number of octaves by which two tones differ is the binary logarithm of their frequency ratio .

To study tuning systems and other aspects of music theory that require finer distinctions between tones , it is helpful to have a measure of the size of an interval that is finer than an octave and is additive (as logarithms are) rather than multiplicative (as frequency ratios are) . That is , if tones x

, y , and z form a rising sequence of tones , then the measure of the interval from x to y plus the measure of the interval from y to z should equal the measure of the interval from x to z . Such a measure is given by the cent , which divides the octave into 1200 equal intervals (12 semitones of 100 cents each) . Mathematically , given tones with frequencies f1 and f2 , the number of cents in the interval from f1 to f2 is

<formula>

The millioctave is defined in the same way , but with a multiplier of 1000 instead of 1200 .

= = = Sports scheduling = = =