Central Limit Theorem

Independence & Benford’s Law

CLT: Important Requirements

CLT requires the sampled variables to be independent and identically distributed.

Independence:
Two events are independent if the
occurrence of one event does not affect the
chances of the occurrence of the other event

Is there independence in real life?

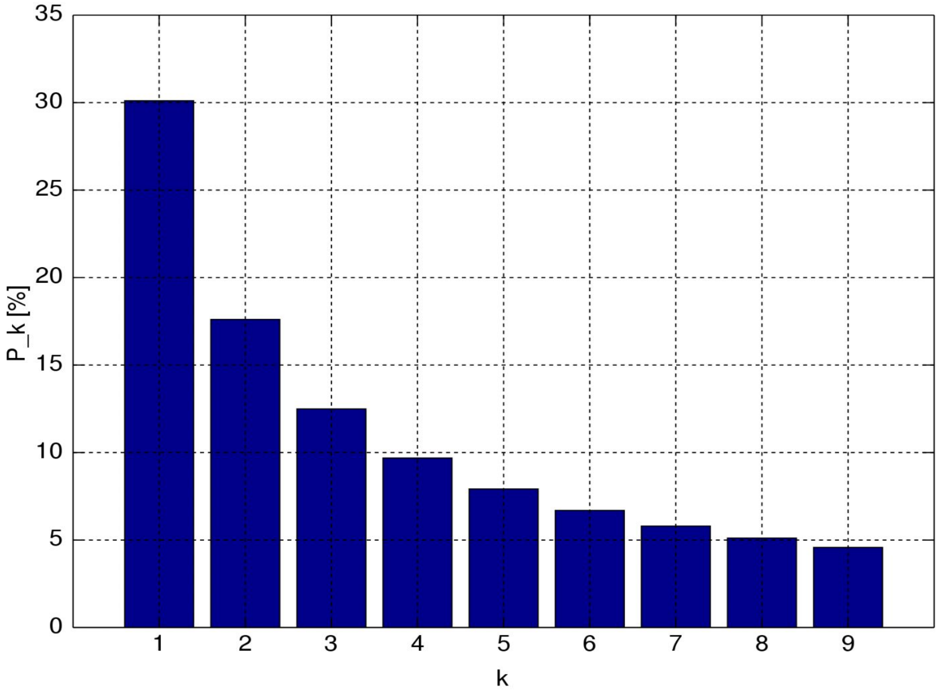
X1 (A, B, C, XI, 陆, ⚕, ♃, ♵, weather, …)

X2 (D, E, F, V, 拾, 🛳, ⛪, ☊, weather, …)

Benford’s Law

Benford's law is an observation about the leading digits of the numbers found in real-world data sets. It states that the leading digit being 1 is more frequent than 2 and 2 more than 3 and so on.

Distribution of first digits according to Benford’s Law



Pr(d) = log(1+ 10/d)

Leading digit of heights of the 58 tallest structures in the world

Leading digit	m		ft		Per Benford's law
	Count	%	Count	%	
1	24	41.4%	16	27.6%	30.1%
2	9	15.5%	8	13.8%	17.6%
3	7	12.1%	5	8.6%	12.5%
4	6	10.3%	7	12.1%	9.7%
5	1	1.7%	10	17.2%	7.9%
6	5	8.6%	4	6.9%	6.7%
7	1	1.7%	2	3.4%	5.8%
8	4	6.9%	5	8.6%	5.1%
9	1	1.7%	1	1.7%	4.6%

Leading digit of 1st 100 elements of Fibonacci Series

Leading Digit	Frequency
1	0.3
2	0.18
3	0.13
4	0.09
5	0.08
6	0.06
7	0.05
8	0.07
9	0.04

Benford’s Law compliance: Benford’s law is prominent when the data has several orders of magnitude, like exponential distribution. It needs the distribution to be widely spread.

Examples that follow Benford’s Law:

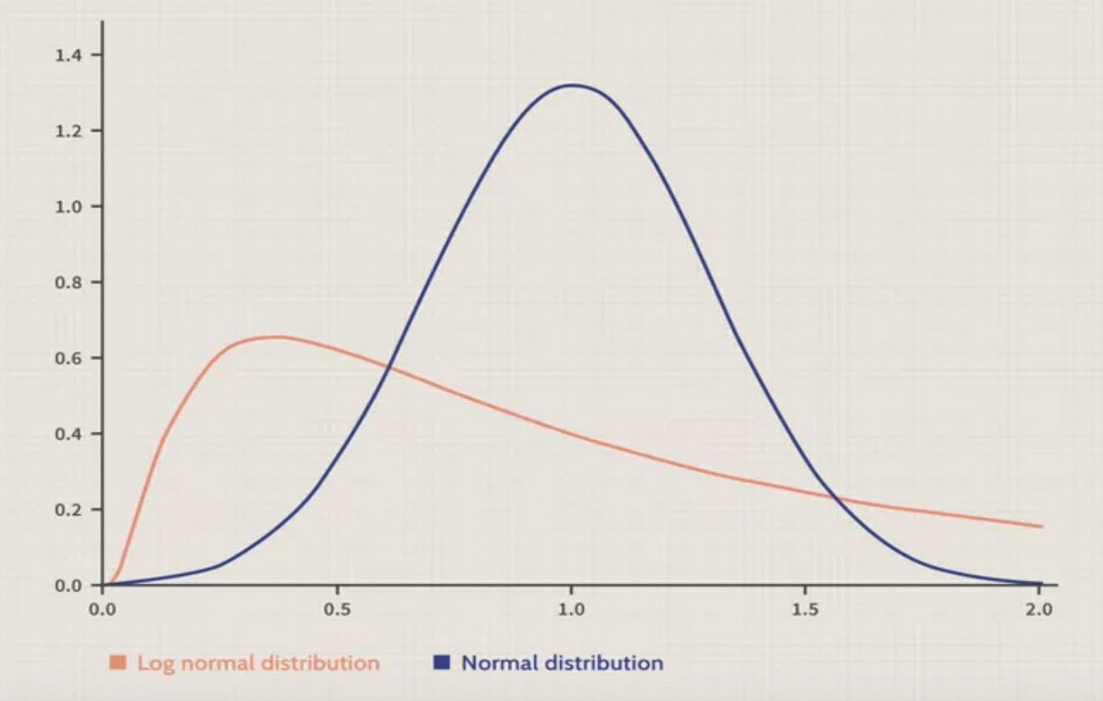
1) Fibonacci series
2) Factorials
3) Powers of two

Examples that don’t follow Benford’s Law:

1) Lists of local telephone numbers
2) Distributions that do not span several orders of magnitude will not follow Benford's law
   Ex: Height

How is CLT related to Benford’s Law?

Central limit theorem says that multiplying more and more random variables will create a log-normal distribution with larger and larger variance, so eventually it covers many orders of magnitude.



Log-normal distributions are positively skewed with long right tails due to low mean values and high variances in the random variables.

Applications of Benford’s Law

1) Accounting fraud detection: Based on a plausible assumption that people who fabricate figures tend to distribute their digits fairly uniformly, a simple comparison of first-digit frequency distribution from the data with the expected distribution according to Benford's law ought to show up any anomalous results.

2) Criminal trials

3) Election data