Central limit theorem

From Wikipedia, the free encyclopedia

In probability theory, the **central limit theorem** (**CLT**) states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution.[1][2] To illustrate what this means, suppose that a sample is obtained containing a large number of observations, each observation being randomly generated in a way that does not depend on the values of the other observations, and that the arithmetic average of the observed values is computed. If this procedure is performed many times, the central limit theorem says that the computed values of the average will be distributed according to the normal distribution (commonly known as a "bell curve"). A simple example of this is that if one flips a coin many times, the probability of getting a given number of heads should follow a normal curve, with mean equal to half the total number of flips.

The central limit theorem has a number of variants. In its common form, the random variables must be identically distributed. In variants, convergence of the mean to the normal distribution also occurs for non-identical distributions or for non-independent observations, given that they comply with certain conditions.

In more general usage, a **central limit theorem** is any of a set of weak-convergence theorems in probability theory. They all express the fact that a sum of many independent and identically distributed (i.i.d.) random variables, or alternatively, random variables with specific types of dependence, will tend to be distributed according to one of a small set of *attractor distributions*. When the variance of the i.i.d. variables is finite, the attractor distribution is the normal distribution. In contrast, the sum of a number of i.i.d. random variables with power law tail distributions decreasing as $|x|^{-\alpha-1}$ where $0 < \alpha < 2$ (and therefore having infinite variance) will tend to an alpha-stable distribution with stability parameter (or index of stability) of α as the number of variables grows.[3]

Contents

Central limit theorems for independent sequences

Classical CLT

Let $\{X_1, ..., X_n\}$ be a random sample of size n — that is, a sequence of independent and identically distributed random variables drawn from distributions of expected values given by μ and finite variances given by σ^2. Suppose we are interested in the sample average

$$S_n := \frac{X_1 + \cdots + X_n}{n}$$

of these random variables. By the law of large numbers, the sample averages converge in probability and almost surely to the expected value μ as $n \to \infty$. The classical central limit theorem describes the size and the distributional form of the stochastic fluctuations around the deterministic number μ during this convergence. More precisely, it states that as n gets larger, the distribution of the difference between the sample average S_n and its limit μ, when multiplied by the factor $\sqrt{n}$

(that is $\sqrt{n}(S_n - \mu)$), approximates the normal distribution with mean 0 and variance σ^2. For large enough n, the distribution

of S_n is close to the normal distribution with mean μ and

variance $\frac{\sigma^2}{n}$. The usefulness of the theorem is that the distribution of $\sqrt{n}(S_n - \mu)$ approaches normality regardless of the shape of the distribution of the individual X_i's. Formally, the theorem can be stated as follows:

> **Lindeberg–Lévy CLT.** Suppose $\{X_1, X_2, ...\}$ is a sequence of i.i.d. random variables with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$. Then as *n* approaches infinity, the random variables $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $N(0, \sigma^2)$:[4]

A distribution being "smoothed out" by summation, showing original density of distribution and three subsequent summations; see Illustration of the central limit theorem for further details.

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \xrightarrow{d} N(0, \sigma$$
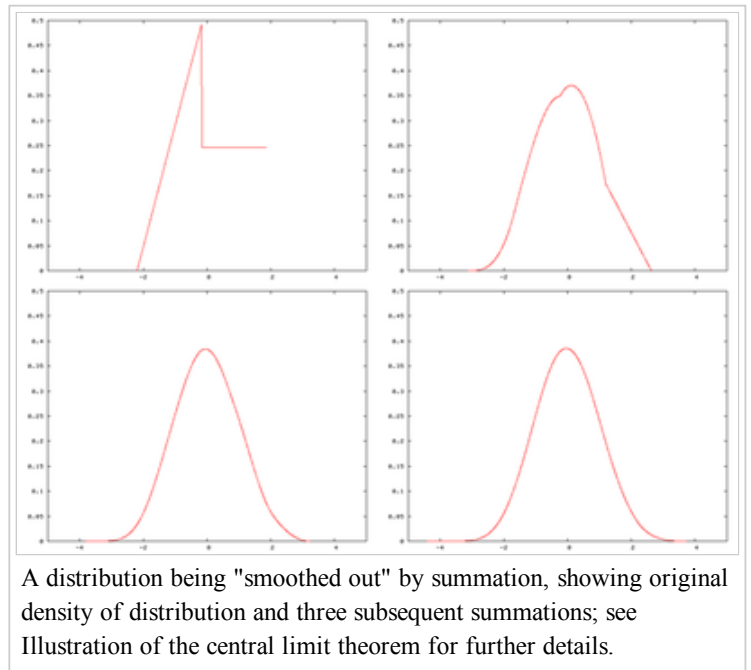
In the case σ > 0, convergence in distribution means that the cumulative distribution functions of $\sqrt{n}(S_n - \mu)$ converge pointwise to the cdf of the $N(0, \sigma^2)$ distribution: for every real number *z*,

$$\lim_{n\to\infty} \Pr[\sqrt{n}(S_n - \mu) \le z] = \Phi(z/\sigma),$$

where $\Phi(x)$ is the standard normal cdf evaluated at *x*. Note that the convergence is uniform in *z* in the sense that

$$\lim_{n\to\infty} \sup_{z\in\mathbf{R}} \left| \Pr[\sqrt{n}(S_n - \mu) \le z] - \Phi(z/\sigma) \right| = 0,$$

where sup denotes the least upper bound (or supremum) of the set.[5]

Lyapunov CLT

The theorem is named after Russian mathematician Aleksandr Lyapunov. In this variant of the central limit theorem the random variables X_i have to be independent, but not necessarily identically distributed. The theorem also requires that random variables $|X_i|$ have moments of some order (2 + δ), and that the rate of growth of these moments is limited by the Lyapunov condition given below.

> **Lyapunov CLT.**[6] Suppose $\{X_1, X_2, ...\}$ is a sequence of independent random variables, each with finite expected value μ_i and variance σ_i^2. Define

$$s_n^2 = \sum_{i=1}^{n} \sigma_i^2$$

If for some δ > 0, the *Lyapunov's condition*

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathrm{E}\left[|X_i - \mu_i|^{2+\delta} \right] = 0$$

is satisfied, then a sum of $(X_i - \mu_i)/s_n$ converges in distribution to a standard normal random variable, as *n* goes to infinity:

$$\frac{1}{s_n} \sum_{i=1}^{n} (X_i - \mu_i) \xrightarrow{d} \mathcal{N}(0, 1).$$

In practice it is usually easiest to check the Lyapunov's condition for $\delta = 1$. If a sequence of random variables satisfies Lyapunov's condition, then it also satisfies Lindeberg's condition. The converse implication, however, does not hold.

Lindeberg CLT

In the same setting and with the same notation as above, the Lyapunov condition can be replaced with the following weaker one (from Lindeberg in 1920).

Suppose that for every $\varepsilon > 0$

$$\lim_{n \to \infty} \frac{1}{s_n^2} \sum_{i=1}^{n} E\left[(X_i - \mu_i)^2 \cdot \mathbf{1}_{\{|X_i - \mu_i| > \varepsilon s_n\}} \right] = 0$$

where $\mathbf{1}_{\{\dots\}}$ is the indicator function. Then the distribution of the standardized sums $\dfrac{1}{s_n} \sum_{i=1}^{n} (X_i - \mu_i)$ converges towards the standard normal distribution N(0,1).

Multidimensional CLT

Proofs that use characteristic functions can be extended to cases where each individual X_i is a random vector in $\mathbf{R}^k$, with mean vector $\mu = E(X_i)$ and covariance matrix Σ (amongst the components of the vector), and these random vectors are independent and identically distributed. Summation of these vectors is being done componentwise. The multidimensional central limit theorem states that when scaled, sums converge to a multivariate normal distribution.[7]

Let

$$\mathbf{X_i} = \begin{bmatrix} X_{i(1)} \\ \vdots \\ X_{i(k)} \end{bmatrix}$$

be the k-vector. The bold in $\mathbf{X}_i$ means that it is a random vector, not a random (univariate) variable. Then the sum of the random vectors will be

$$\begin{bmatrix} X_{1(1)} \\ \vdots \\ X_{1(k)} \end{bmatrix} + \begin{bmatrix} X_{2(1)} \\ \vdots \\ X_{2(k)} \end{bmatrix} + \cdots + \begin{bmatrix} X_{n(1)} \\ \vdots \\ X_{n(k)} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} [X_{i(1)}] \\ \vdots \\ \sum_{i=1}^{n} [X_{i(k)}] \end{bmatrix} = \sum_{i=1}^{n} \mathbf{X_i}$$

and the average is

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{X_i} = \frac{1}{n} \begin{bmatrix} \sum_{i=1}^{n} X_{i(1)} \\ \vdots \\ \sum_{i=1}^{n} X_{i(k)} \end{bmatrix} = \begin{bmatrix} \bar{X}_{i(1)} \\ \vdots \\ \bar{X}_{i(k)} \end{bmatrix} = \bar{\mathbf{X}}_{\mathbf{n}}$$

and therefore

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\mathbf{X_i} - E(X_i)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{X_i} - \mu) = \sqrt{n} \left(\bar{\mathbf{X}}_{\mathbf{n}} - \mu \right).$$

The multivariate central limit theorem states that

$$\sqrt{n}\left(\overline{\mathbf{X}}_n - \mu\right) \overset{D}{\to} \mathcal{N}_k(0, \Sigma)$$

where the covariance matrix Σ is equal to

$$\Sigma = \begin{bmatrix} \mathrm{Var}\left(X_{1(1)}\right) & \mathrm{Cov}\left(X_{1(1)}, X_{1(2)}\right) & \mathrm{Cov}\left(X_{1(1)}, X_{1(3)}\right) & \cdots & \mathrm{Cov}\left(X_{1(1)}, X_{1(k)}\right) \\ \mathrm{Cov}\left(X_{1(2)}, X_{1(1)}\right) & \mathrm{Var}\left(X_{1(2)}\right) & \mathrm{Cov}\left(X_{1(2)}, X_{1(3)}\right) & \cdots & \mathrm{Cov}\left(X_{1(2)}, X_{1(k)}\right) \\ \mathrm{Cov}\left(X_{1(3)}, X_{1(1)}\right) & \mathrm{Cov}\left(X_{1(3)}, X_{1(2)}\right) & \mathrm{Var}\left(X_{1(3)}\right) & \cdots & \mathrm{Cov}\left(X_{1(3)}, X_{1(k)}\right) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}\left(X_{1(k)}, X_{1(1)}\right) & \mathrm{Cov}\left(X_{1(k)}, X_{1(2)}\right) & \mathrm{Cov}\left(X_{1(k)}, X_{1(3)}\right) & \cdots & \mathrm{Var}\left(X_{1(k)}\right) \end{bmatrix}.$$

Generalised theorem

The central limit theorem states that the sum of a number of independent and identically distributed random variables with finite variances will tend to a normal distribution as the number of variables grows. A generalization due to Gnedenko and Kolmogorov states that the sum of a number of random variables with a power-law tail (Paretian tail) distributions decreasing as $|x|^{-\alpha-1}$ where $0 < \alpha < 2$ (and therefore having infinite variance) will tend to a stable distribution $f(x; \alpha, 0, c, 0)$ as the number of summands grows.[8][9] If $\alpha>2$ then the sum converges to a stable distribution with stability parameter equal to 2, i.e. a Gaussian distribution.[10]

Central limit theorems for dependent processes

CLT under weak dependence

A useful generalization of a sequence of independent, identically distributed random variables is a mixing random process in discrete time; "mixing" means, roughly, that random variables temporally far apart from one another are nearly independent. Several kinds of mixing are used in ergodic theory and probability theory. See especially strong mixing (also called α-mixing) defined by $\alpha(n) \to 0$ where $\alpha(n)$ is so-called strong mixing coefficient.

A simplified formulation of the central limit theorem under strong mixing is:[11]

Theorem. Suppose that $X_1, X_2, \ldots$ is stationary and α-mixing with $\alpha_n = O(n^{-5})$ and that $E(X_n) = 0$ and $E(X_n^2) < \infty$. Denote $S_n = X_1 + \ldots + X_n$, then the limit

$$\sigma^2 = \lim_n \frac{E(S_n^2)}{n}$$

exists, and if $\sigma \neq 0$ then $S_n/\left(\sigma\sqrt{n}\right)$ converges in distribution to $N(0, 1)$.

In fact,

$$\sigma^2 = E(X_1^2) + 2\sum_{k=1}^{\infty} E(X_1 X_{1+k}),$$

where the series converges absolutely.

The assumption $\sigma \neq 0$ cannot be omitted, since the asymptotic normality fails for $X_n = Y_n - Y_{n-1}$ where Y_n are another stationary sequence.

There is a stronger version of the theorem:[12] the assumption $E(X_n^{12}) < \infty$ is replaced with $E(|X_n|^{2+\delta}) < \infty$, and the assumption $\alpha_n = O(n^{-5})$ is replaced with $\sum_n \alpha_n^{\frac{\delta}{2(2+\delta)}} < \infty$. Existence of such $\delta > 0$ ensures the conclusion. For encyclopedic treatment of limit theorems under mixing conditions see (Bradley 2005).

Martingale difference CLT

Theorem. Let a martingale M_n satisfy

- $\dfrac{1}{n}\displaystyle\sum_{k=1}^{n} E\big((M_k - M_{k-1})^2 | M_1, \ldots, M_{k-1}\big) \to 1$ in probability as n tends to infinity,

- for every $\varepsilon > 0$, $\dfrac{1}{n}\displaystyle\sum_{k=1}^{n} E\big((M_k - M_{k-1})^2; |M_k - M_{k-1}| > \varepsilon\sqrt{n}\big) \to 0$ as n tends to infinity,

then $M_n / \sqrt{n}$ converges in distribution to N(0,1) as $n \to \infty$.[13][14]

Caution: The restricted expectation E(*X*; *A*) should not be confused with the conditional expectation
$E(X \mid A) = E(X; A)/\mathbf{P}(A)$.

Remarks

Proof of classical CLT

For a theorem of such fundamental importance to statistics and applied probability, the central limit theorem has a remarkably simple proof using characteristic functions. It is similar to the proof of a (weak) law of large numbers. For any random variable, *Y*, with zero mean and a unit variance (var(*Y*) = 1), the characteristic function of *Y* is, by Taylor's theorem,

$$\varphi_Y(t) = 1 - \frac{t^2}{2} + o(t^2), \quad t \to 0$$

where $o(t^2)$ is "little o notation" for some function of *t* that goes to zero more rapidly than t^2.

Letting Y_i be $(X_i - \mu)/\sigma$, the standardized value of X_i, it is easy to see that the standardized mean of the observations $X_1, X_2, \ldots, X_n$ is

$$Z_n = \frac{n\overline{X}_n - n\mu}{\sigma\sqrt{n}} = \sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}}$$

By simple properties of characteristic functions, the characteristic function of the sum is:

$$\varphi_{Z_n} = \varphi_{\sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}}}(t) = \varphi_{Y_1}\left(t/\sqrt{n}\right) \cdot \varphi_{Y_2}\left(t/\sqrt{n}\right) \cdots \varphi_{Y_n}\left(t/\sqrt{n}\right)$$

$$= \left[\varphi_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

so that, by the limit of the exponential function ($e^x = \lim(1 + x/n)^n$) the characteristic function of Z_n is

$$\left[\varphi_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n \to e^{-t^2/2}, \quad n \to \infty.$$

But this limit is just the characteristic function of a standard normal distribution *N*(0, 1), and the central limit theorem follows from the Lévy continuity theorem, which confirms that the convergence of characteristic functions implies convergence in distribution.

Convergence to the limit

The central limit theorem gives only an asymptotic distribution. As an approximation for a finite number of observations, it provides a reasonable approximation only when close to the peak of the normal distribution; it requires a very large number of observations to stretch into the tails.

If the third central moment $E((X_1 - \mu)^3)$ exists and is finite, then the above convergence is uniform and the speed of convergence is at least on the order of $1/n^{1/2}$ (see Berry-Esseen theorem). Stein's method[15] can be used not only to prove the central limit theorem, but also to provide bounds on the rates of convergence for selected metrics.[16]

The convergence to the normal distribution is monotonic, in the sense that the entropy of Z_n increases monotonically to that of the normal distribution.[17]

The central limit theorem applies in particular to sums of independent and identically distributed discrete random variables. A sum of discrete random variables is still a discrete random variable, so that we are confronted with a sequence of discrete random variables whose cumulative probability distribution function converges towards a cumulative probability distribution function corresponding to a continuous variable (namely that of the normal distribution). This means that if we build a histogram of the realisations of the sum of n independent identical discrete variables, the curve that joins the centers of the upper faces of the rectangles forming the histogram converges toward a Gaussian curve as n approaches infinity, this relation is known as de Moivre–Laplace theorem. The binomial distribution article details such an application of the central limit theorem in the simple case of a discrete variable taking only two possible values.

Relation to the law of large numbers

The law of large numbers as well as the central limit theorem are partial solutions to a general problem: "What is the limiting behaviour of S_n as n approaches infinity?" In mathematical analysis, asymptotic series are one of the most popular tools employed to approach such questions.

Suppose we have an asymptotic expansion of $f(n)$:

$$f(n) = a_1\varphi_1(n) + a_2\varphi_2(n) + O(\varphi_3(n)) \qquad (n \to \infty).$$

Dividing both parts by $\varphi_1(n)$ and taking the limit will produce a_1, the coefficient of the highest-order term in the expansion, which represents the rate at which $f(n)$ changes in its leading term.

$$\lim_{n\to\infty} \frac{f(n)}{\varphi_1(n)} = a_1.$$

Informally, one can say: "$f(n)$ grows approximately as $a_1 \varphi_1(n)$". Taking the difference between $f(n)$ and its approximation and then dividing by the next term in the expansion, we arrive at a more refined statement about $f(n)$:

$$\lim_{n\to\infty} \frac{f(n) - a_1\varphi_1(n)}{\varphi_2(n)} = a_2.$$

Here one can say that the difference between the function and its approximation grows approximately as $a_2 \varphi_2(n)$. The idea is that dividing the function by appropriate normalizing functions, and looking at the limiting behavior of the result, can tell us much about the limiting behavior of the original function itself.

Informally, something along these lines is happening when the sum, S_n, of independent identically distributed random variables, $X_1, ..., X_n$, is studied in classical probability theory. If each X_i has finite mean μ, then by the law of large numbers, $S_n/n \to \mu$.[18] If in addition each X_i has finite variance σ^2, then by the central limit theorem,

$$\frac{S_n - n\mu}{\sqrt{n}} \to \xi,$$

where ξ is distributed as N(0, σ^2). This provides values of the first two constants in the informal expansion

$$S_n \approx \mu n + \xi\sqrt{n}.$$

In the case where the X_i's do not have finite mean or variance, convergence of the shifted and rescaled sum can also occur with different centering and scaling factors:

$$\frac{S_n - a_n}{b_n} \to \Xi,$$

or informally

$$S_n \approx a_n + \Xi b_n.$$

Distributions Ξ which can arise in this way are called *stable*.[19] Clearly, the normal distribution is stable, but there are also other stable distributions, such as the Cauchy distribution, for which the mean or variance are not defined. The scaling factor b_n may be proportional to n^c, for any $c \geq 1/2$; it may also be multiplied by a slowly varying function of n.[10][20]

The law of the iterated logarithm specifies what is happening "in between" the law of large numbers and the central limit theorem. Specifically it says that the normalizing function $\sqrt{n \log \log n}$ intermediate in size between n of the law of large numbers and $\sqrt{n}$ of the central limit theorem provides a non-trivial limiting behavior.

Alternative statements of the theorem

Density functions

The density of the sum of two or more independent variables is the convolution of their densities (if these densities exist). Thus the central limit theorem can be interpreted as a statement about the properties of density functions under convolution: the convolution of a number of density functions tends to the normal density as the number of density functions increases without bound. These theorems require stronger hypotheses than the forms of the central limit theorem given above.

Theorems of this type are often called local limit theorems. See Petrov[21] for a particular local limit theorem for sums of independent and identically distributed random variables.

Characteristic functions

Since the characteristic function of a convolution is the product of the characteristic functions of the densities involved, the central limit theorem has yet another restatement: the product of the characteristic functions of a number of density functions becomes close to the characteristic function of the normal density as the number of density functions increases without bound, under the conditions stated above. However, to state this more precisely, an appropriate scaling factor needs to be applied to the argument of the characteristic function.

An equivalent statement can be made about Fourier transforms, since the characteristic function is essentially a Fourier transform.

Extensions to the theorem

Products of positive random variables

The logarithm of a product is simply the sum of the logarithms of the factors. Therefore when the logarithm of a product of random variables that take only positive values approaches a normal distribution, the product itself approaches a log-normal distribution. Many physical quantities (especially mass or length, which are a matter of scale and cannot be negative) are the products of different random factors, so they follow a log-normal distribution.

Whereas the central limit theorem for sums of random variables requires the condition of finite variance, the corresponding theorem for products requires the corresponding condition that the density function be square-integrable.[22]

Beyond the classical framework

Asymptotic normality, that is, convergence to the normal distribution after appropriate shift and rescaling, is a phenomenon much more general than the classical framework treated above, namely, sums of independent random variables (or vectors). New frameworks are revealed from time to time; no single unifying framework is available for now.

Convex body

Theorem. There exists a sequence $\varepsilon_n \downarrow 0$ for which the following holds. Let $n \geq 1$, and let random variables X_1, ..., X_n have a log-concave joint density f such that $f(x_1, ..., x_n) = f(|x_1|, ..., |x_n|)$ for all $x_1, ..., x_n$, and $E(X_k^2) = 1$ for all $k = 1, ..., n$. Then the distribution of

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

is ε_n-close to $N(0, 1)$ in the total variation distance.[23]

These two ε_n-close distributions have densities (in fact, log-concave densities), thus, the total variance distance between them is the integral of the absolute value of the difference between the densities. Convergence in total variation is stronger than weak convergence.

An important example of a log-concave density is a function constant inside a given convex body and vanishing outside; it corresponds to the uniform distribution on the convex body, which explains the term "central limit theorem for convex bodies".

Another example: $f(x_1, ..., x_n) = \text{const} \cdot \exp(-(|x_1|^\alpha + ... + |x_n|^\alpha)^\beta)$ where $\alpha > 1$ and $\alpha\beta > 1$. If $\beta = 1$ then $f(x_1, ..., x_n)$ factorizes into $\text{const} \cdot \exp(-|x_1|^\alpha)...\exp(-|x_n|^\alpha)$, which means independence of $X_1, ..., X_n$. In general, however, they are dependent.

The condition $f(x_1, ..., x_n) = f(|x_1|, ..., |x_n|)$ ensures that $X_1, ..., X_n$ are of zero mean and uncorrelated; still, they need not be independent, nor even pairwise independent. By the way, pairwise independence cannot replace independence in the classical central limit theorem.[24]

Here is a Berry–Esseen type result.

Theorem. Let $X_1, ..., X_n$ satisfy the assumptions of the previous theorem, then [25]

$$\left| \mathbb{P}\left(a \leq \frac{X_1 + \cdots + X_n}{\sqrt{n}} \leq b\right) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt \right| \leq \frac{C}{n}$$

for all $a < b$; here C is a universal (absolute) constant. Moreover, for every $c_1, ..., c_n \in \mathbf{R}$ such that $c_1^2 + ... + c_n^2 = 1$,

$$\left| \mathbb{P}\left(a \leq c_1 X_1 + \cdots + c_n X_n \leq b\right) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt \right| \leq C(c_1^4 + \cdots + c_n^4).$$

The distribution of $(X_1 + \cdots + X_n)/\sqrt{n}$ need not be approximately normal (in fact, it can be uniform).[26] However, the distribution of $c_1 X_1 + ... + c_n X_n$ is close to $N(0, 1)$ (in the total variation distance) for most of vectors $(c_1, ..., c_n)$ according to the uniform distribution on the sphere $c_1^2 + ... + c_n^2 = 1$.

Lacunary trigonometric series

Theorem (Salem–Zygmund). Let U be a random variable distributed uniformly on $(0, 2\pi)$, and $X_k = r_k \cos(n_k U + a_k)$, where

- n_k satisfy the lacunarity condition: there exists $q > 1$ such that $n_{k+1} \geq q n_k$ for all k,
- r_k are such that

$$r_1^2 + r_2^2 + \cdots = \infty \quad \text{and} \quad \frac{r_k^2}{r_1^2 + \cdots + r_k^2} \to 0,$$

- $0 \le a_k < 2\pi$.

Then[27][28]

$$\frac{X_1 + \cdots + X_k}{\sqrt{r_1^2 + \cdots + r_k^2}}$$

converges in distribution to $N(0, 1/2)$.

Gaussian polytopes

Theorem Let $A_1, ..., A_n$ be independent random points on the plane $\mathbf{R}^2$ each having the two-dimensional standard normal distribution. Let K_n be the convex hull of these points, and X_n the area of K_n Then[29]

$$\frac{X_n - \mathrm{E}(X_n)}{\sqrt{\mathrm{Var}(X_n)}}$$

converges in distribution to $N(0, 1)$ as n tends to infinity.

The same holds in all dimensions (2, 3, ...).

The polytope K_n is called Gaussian random polytope.

A similar result holds for the number of vertices (of the Gaussian polytope), the number of edges, and in fact, faces of all dimensions.[30]

Linear functions of orthogonal matrices

A linear function of a matrix M is a linear combination of its elements (with given coefficients), $M \mapsto \mathrm{tr}(AM)$ where A is the matrix of the coefficients; see Trace (linear algebra)#Inner product.

A random orthogonal matrix is said to be distributed uniformly, if its distribution is the normalized Haar measure on the orthogonal group $O(n, \mathbf{R})$; see Rotation matrix#Uniform random rotation matrices.

Theorem. Let M be a random orthogonal $n \times n$ matrix distributed uniformly, and A a fixed $n \times n$ matrix such that $\mathrm{tr}(AA^*) = n$, and let $X = \mathrm{tr}(AM)$. Then[31] the distribution of X is close to $N(0, 1)$ in the total variation metric up to $2\sqrt{3}/(n-1)$.

Subsequences

Theorem. Let random variables $X_1, X_2, \ldots \in L_2(\Omega)$ be such that $X_n \to 0$ weakly in $L_2(\Omega)$ and $X_n^2 \to 1$ weakly in $L_1(\Omega)$. Then there exist integers $n_1 < n_2 < \ldots$ such that $\left(X_{n_1} + \cdots + X_{n_k} \right) / \sqrt{k}$ converges in distribution to $N(0, 1)$ as k tends to infinity.[32]

Tsallis statistics

A generalization of the classical central limit theorem to the context of Tsallis statistics has been described by Umarov, Tsallis and Steinberg[33] in which the independence constraint for the i.i.d. variables is relaxed to an extent defined by the q parameter, with independence being recovered as $q \to 1$. In analogy to the classical central limit theorem, such random variables with fixed mean and variance tend towards the q-Gaussian distribution, which maximizes the Tsallis entropy under these constraints. Umarov, Tsallis, Gell-Mann and Steinberg have defined similar generalizations of all symmetric alpha-stable distributions, and have formulated a number of conjectures regarding their relevance to an even more general central limit theorem.[34]
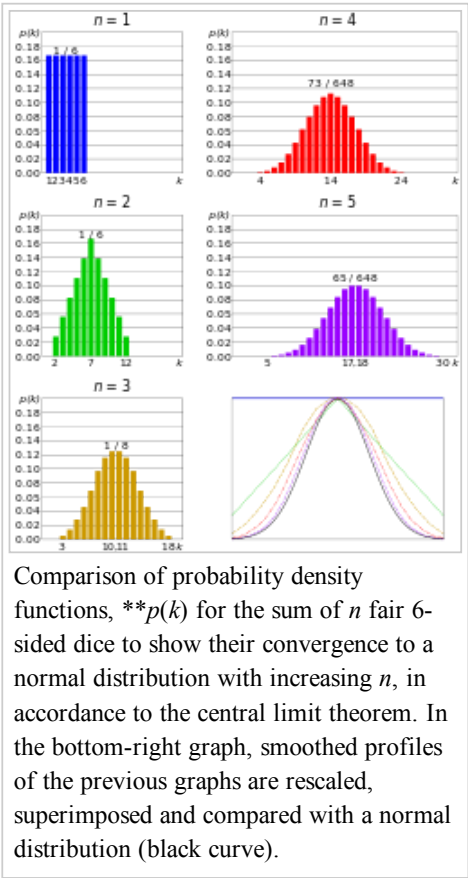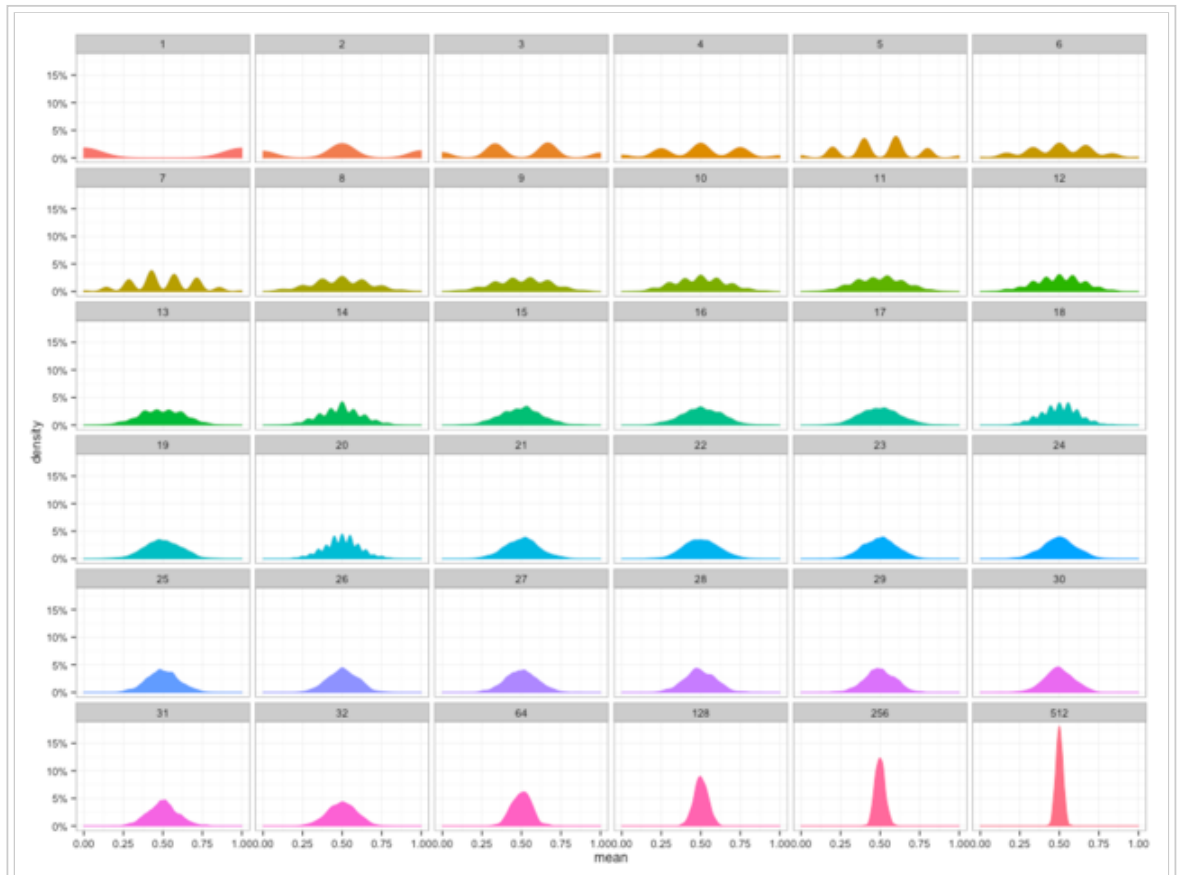
Random walk on a crystal lattice

The central limit theorem may be established for the simple random walk on a crystal lattice (an infinite-fold abelian covering graph over a finite graph), and is used for design of crystal structures. [35][36]

Applications and examples

Simple example

A simple example of the central limit theorem is rolling a large number of identical, unbiased dice. The distribution of the sum (or average) of the rolled numbers will be well approximated by a normal distribution. Since real-world quantities are often the balanced sum of many unobserved random events, the central limit theorem also provides a partial explanation for the prevalence of the normal probability distribution. It also justifies the approximation of large-sample statistics to the normal distribution in controlled experiments.

Comparison of probability density functions, **$p(k)$ for the sum of n fair 6-sided dice to show their convergence to a normal distribution with increasing n, in accordance to the central limit theorem. In the bottom-right graph, smoothed profiles of the previous graphs are rescaled, superimposed and compared with a normal distribution (black curve).

Another simulation with binomial distribution. 0 and 1 were generated and their means calculated for different sample sizes, from 1 to 512. It's possible to see that when the sample increases, the mean distribution tend to be more centered and with thinner tails.
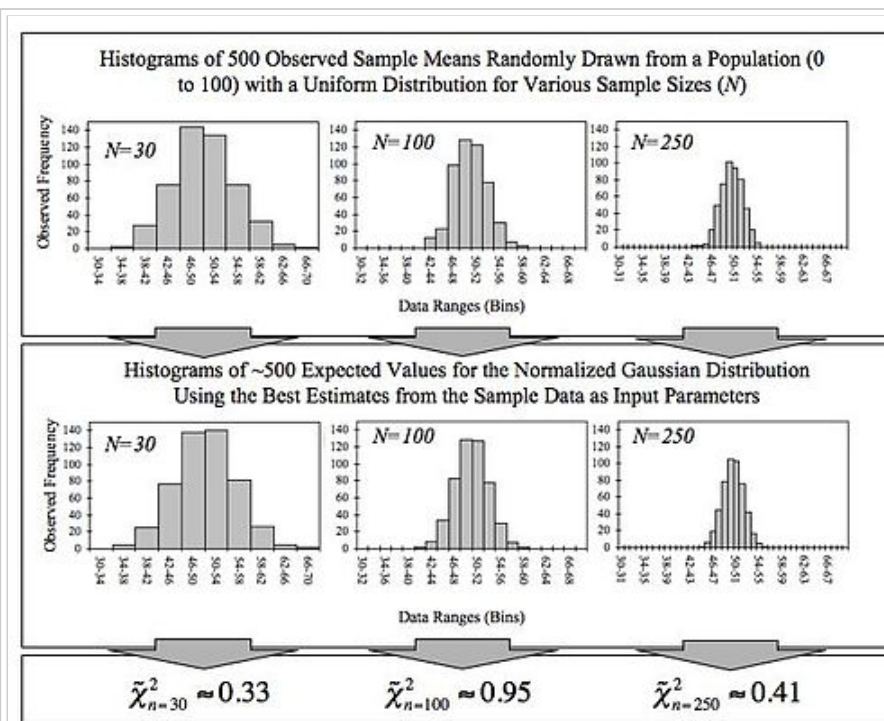
This figure demonstrates the central limit theorem. The sample means are generated using a random number generator, which draws numbers between 0 and 100 from a uniform probability distribution. It illustrates that increasing sample sizes result in the 500 measured sample means being more closely distributed about the population mean (50 in this case). It also compares the observed distributions with the distributions that would be expected for a normalized Gaussian distribution, and shows the chi-squared values that quantify the goodness of the fit (the fit is good if the reduced chi-squared value is less than or approximately equal to one). The input into the normalized Gaussian function is the mean of sample means (~50) and the mean sample standard

deviation divided by the square root of the sample size ($\sim 28.87/\sqrt{n}$), which is called the standard deviation of the mean (since it refers to the spread of sample means).

Real applications

Published literature contains a number of useful and interesting examples and applications relating to the central limit theorem.[37] One source[38] states the following examples:

- The probability distribution for total distance covered in a random walk (biased or unbiased) will tend toward a normal distribution.
- Flipping a large number of coins will result in a normal distribution for the total number of heads (or equivalently total number of tails).

From another viewpoint, the central limit theorem explains the common appearance of the "Bell Curve" in density estimates applied to real world data. In cases like electronic noise, examination grades, and so on, we can often regard a single measured value as the weighted average of a large number of small effects. Using generalisations of the central limit theorem, we can then see that this would often (though not always) produce a final distribution that is approximately normal.

In general, the more a measurement is like the sum of independent variables with equal influence on the result, the more normality it exhibits. This justifies the common use of this distribution to stand in for the effects of unobserved variables in models like the linear model.

Regression

Regression analysis and in particular ordinary least squares specifies that a dependent variable depends according to some function upon one or more independent variables, with an additive error term. Various types of statistical inference on the regression assume that the error term is normally distributed. This assumption can be justified by assuming that the error term is actually the sum of a large number of independent error terms; even if the individual error terms are not normally distributed, by the central limit theorem their sum can be well approximated by a normal distribution.

Other illustrations

Given its importance to statistics, a number of papers and computer packages are available that demonstrate the convergence involved in the central limit theorem.[39]

History

Tijms writes:[40]

> The central limit theorem has an interesting history. The first version of this theorem was postulated by the French-born mathematician Abraham de Moivre who, in a remarkable article published in 1733, used the normal distribution to approximate the distribution of the number of heads resulting from many tosses of a fair coin. This finding was far ahead of its time, and was nearly forgotten until the famous French mathematician Pierre-Simon Laplace rescued it from obscurity in his monumental work *Théorie analytique des probabilités*, which was published in 1812. Laplace expanded De Moivre's finding by approximating the binomial distribution with the normal distribution. But as with De Moivre, Laplace's finding received little attention in his own time. It was not until the nineteenth century was at an end that the importance of the central limit theorem was discerned, when, in 1901, Russian mathematician Aleksandr Lyapunov defined it in general terms and proved precisely how it worked mathematically. Nowadays, the central limit theorem is considered to be the unofficial sovereign of probability theory.

Sir Francis Galton described the Central Limit Theorem as:[41]

> I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error". The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The

huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.

The actual term "central limit theorem" (in German: "zentraler Grenzwertsatz") was first used by George Pólya in 1920 in the title of a paper.[42][43] Pólya referred to the theorem as "central" due to its importance in probability theory. According to Le Cam, the French school of probability interprets the word *central* in the sense that "it describes the behaviour of the centre of the distribution as opposed to its tails".[43] The abstract of the paper *On the central limit theorem of calculus of probability and the problem of moments* by Pólya[42] in 1920 translates as follows.

The occurrence of the Gaussian probability density $1 = e^{-x^2}$ in repeated experiments, in errors of measurements, which result in the combination of very many and very small elementary errors, in diffusion processes etc., can be explained, as is well-known, by the very same limit theorem, which plays a central role in the calculus of probability. The actual discoverer of this limit theorem is to be named Laplace; it is likely that its rigorous proof was first given by Tschebyscheff and its sharpest formulation can be found, as far as I am aware of, in an article by Liapounoff. [...]

A thorough account of the theorem's history, detailing Laplace's foundational work, as well as Cauchy's, Bessel's and Poisson's contributions, is provided by Hald.[44] Two historical accounts, one covering the development from Laplace to Cauchy, the second the contributions by von Mises, Pólya, Lindeberg, Lévy, and Cramér during the 1920s, are given by Hans Fischer.[45] Le Cam describes a period around 1935.[43] Bernstein[46] presents a historical discussion focusing on the work of Pafnuty Chebyshev and his students Andrey Markov and Aleksandr Lyapunov that led to the first proofs of the CLT in a general setting.

A curious footnote to the history of the Central Limit Theorem is that a proof of a result similar to the 1922 Lindeberg CLT was the subject of Alan Turing's 1934 Fellowship Dissertation for King's College at the University of Cambridge. Only after submitting the work did Turing learn it had already been proved. Consequently, Turing's dissertation was never published.[47][48][49]

See also

- Asymptotic equipartition property
- Benford's law – Result of extension of CLT to product of random variables.
- Central limit theorem for directional statistics – Central limit theorem applied to the case of directional statistics
- Delta method – to compute the limit distribution of a function of a random variable.
- Erdős–Kac theorem – connects the number of prime factors of an integer with the normal probability distribution
- Fisher–Tippett–Gnedenko theorem – limit theorem for extremum values (such as $\max\{X_n\}$)
- Tweedie convergence theorem – A theorem that can be considered to bridge between the central limit theorem and the Poisson convergence theorem

[50]

Notes

1. http://www.math.uah.edu/stat/sample/CLT.html
2. Rice, John (1995), *Mathematical Statistics and Data Analysis* (Second ed.), Duxbury Press, ISBN 0-534-20934-3)
3. Voit, Johannes (2003), *The Statistical Mechanics of Financial Markets*, Springer-Verlag, p. 124, ISBN 3-540-00978-7
4. Billingsley (1995, p. 357)
5. Bauer (2001, Theorem 30.13, p.199)
6. Billingsley (1995, p.362)
7. Van der Vaart, A. W. (1998), *Asymptotic statistics*, New York: Cambridge University Press, ISBN 978-0-521-49603-2, LCCN 98015176
8. Voit, Johannes (2003). "Section 5.4.3". *The Statistical Mechanics of Financial Markets*. Texts and Monographs in Physics. Springer-Verlag. ISBN 3-540-00978-7.
9. Gnedenko, B. V.; Kolmogorov, A. N. (1954). *Limit distributions for sums of independent random variables*. Cambridge: Addison-Wesley.
10. Uchaikin, Vladimir V.; Zolotarev, V. M. (1999). *Chance and stability: stable distributions and their applications*. VSP. pp. 61–62. ISBN 90-6764-301-7.

11. Billingsley (1995, Theorem 27.4)
12. Durrett (2004, Sect. 7.7(c), Theorem 7.8)
13. Durrett (2004, Sect. 7.7, Theorem 7.4)
14. Billingsley (1995, Theorem 35.12)
15. Stein, C. (1972), "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables", *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*: 583–602, MR 402873, Zbl 0278.60026
16. Chen, L.H.Y., Goldstein, L., and Shao, Q.M (2011), *Normal approximation by Stein's method*, Springer, ISBN 978-3-642-15006-7
17. Artstein, S.; Ball, K.; Barthe, F.; Naor, A. (2004), "Solution of Shannon's Problem on the Monotonicity of Entropy", *Journal of the American Mathematical Society* **17** (4): 975–982, doi:10.1090/S0894-0347-04-00459-X
18. Rosenthal, Jeffrey Seth (2000) *A first look at rigorous probability theory*, World Scientific, ISBN 981-02-4322-7.(Theorem 5.3.4, p. 47)
19. Johnson, Oliver Thomas (2004) *Information theory and the central limit theorem*, Imperial College Press, 2004, ISBN 1-86094-473-6. (p. 88)
20. Borodin, A. N. ; Ibragimov, Il'dar Abdulovich; Sudakov, V. N. (1995) *Limit theorems for functionals of random walks*, AMS Bookstore, ISBN 0-8218-0438-3. (Theorem 1.1, p. 8)
21. Petrov, V.V. (1976), "7", *Sums of Independent Random Variables*, New York-Heidelberg: Springer-Verlag
22. Rempala, G.; Wesolowski, J. (2002). "Asymptotics of products of sums and *U*-statistics" (PDF). *Electronic Communications in Probability* **7**: 47–54. doi:10.1214/ecp.v7-1046.
23. Klartag (2007, Theorem 1.2)
24. Durrett (2004, Section 2.4, Example 4.5)
25. Klartag (2008, Theorem 1)
26. Klartag (2007, Theorem 1.1)
27. Zygmund, Antoni (1959), *Trigonometric series, Volume II*, Cambridge. (2003 combined volume I,II: ISBN 0-521-89053-5) (Sect. XVI.5, Theorem 5-5)
28. Gaposhkin (1966, Theorem 2.1.13)
29. Bárány & Vu (2007, Theorem 1.1)
30. Bárány & Vu (2007, Theorem 1.2)
31. Meckes, Elizabeth (2008), "Linear functions on the classical matrix groups", *Transactions of the American Mathematical Society* **360** (10): 5355–5366, arXiv:math/0509441, doi:10.1090/S0002-9947-08-04444-9
32. Gaposhkin (1966, Sect. 1.5)
33. Umarov, Sabir; Tsallis, Constantino; Steinberg, Stanly (2008), "On a q-Central Limit Theorem Consistent with Nonextensive Statistical Mechanics" (PDF), *Milan j. Math.* (Birkhauser Verlag) **76**: 307–328, doi:10.1007/s00032-008-0087-y, retrieved 2011-07-27.
34. Umarov, Sabir; Tsallis, Constantino; Gell-Mann, Murray; Steinberg, Stanly (2010), "Generalization of symmetric α-stable Lévy distributions for $q > 1$", *J Math Phys*. (American Institute of Physics) **51** (3): 033502, doi:10.1063/1.3305292, PMC 2869267, PMID 20596232.
35. Kotani, M.; Sunada, Toshikazu (2003). *Spectral geometry of crystal lattices* **338**. Contemporary Math. pp. 271–305. ISBN 978-0-8218-4269-0.
36. Sunada, Toshikazu (2012). *Topological Crystallography ---With a View Towards Discrete Geometric Analysis---*. Surveys and Tutorials in the Applied Mathematical Sciences **6**. Springer. ISBN 978-4-431-54177-6.
37. Dinov, Christou & Sanchez (2008)
38. SOCR CLT Activity (http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_GCLT_Applications) wiki
39. Marasinghe, M., Meeker, W., Cook, D. & Shin, T.S.(1994 August), "Using graphics and simulation to teach statistical concepts", Paper presented at the Annual meeting of the American Statistician Association, Toronto, Canada.
40. Henk, Tijms (2004), *Understanding Probability: Chance Rules in Everyday Life*, Cambridge: Cambridge University Press, p. 169, ISBN 0-521-54036-4
41. Galton F. (1889) *Natural Inheritance* , p. 66 (http://galton.org/cgi-bin/searchImages/galton/search/books/natural-inheritance/pages/natural-inheritance_0073.htm)
42. Pólya, George (1920), "Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem", *Mathematische Zeitschrift* (in German) **8** (3–4): 171–181, doi:10.1007/BF01206525
43. Le Cam, Lucien (1986), "The central limit theorem around 1935", *Statistical Science* **1** (1): 78–91, doi:10.2307/2245503
44. Hald, Andreas *A History of Mathematical Statistics from 1750 to 1930* (http://www.gbv.de/dms/goettingen/229762905.pdf), Ch.17.
45. Fischer, Hans (2011), *A History of the Central Limit Theorem: From Classical to Modern Probability Theory*, Sources and Studies in the History of Mathematics and Physical Sciences, New York: Springer, doi:10.1007/978-0-387-87857-7, ISBN 978-0-387-87856-0, MR 2743162, Zbl 1226.60004 (Chapter 2: The Central Limit Theorem from Laplace to Cauchy: Changes in Stochastic Objectives and in Analytical Methods, Chapter 5.2: The Central Limit Theorem in the Twenties)
46. Bernstein, S.N. (1945) *On the work of P.L.Chebyshev in Probability Theory*, Nauchnoe Nasledie P.L.Chebysheva. Vypusk Pervyi: Matematika. (Russian) [The Scientific Legacy of P. L. Chebyshev. First Part: Mathematics, Edited by S. N. Bernstein.] Academiya Nauk SSSR, Moscow-Leningrad, 174 pp.
47. Hodges, Andrew (1983) *Alan Turing: the enigma*. London: Burnett Books., pp. 87-88.
48. Zabell, S.L. (2005) *Symmetry and its discontents: essays on the history of inductive probability*, Cambridge University Press. ISBN 0-521-44470-5. (pp. 199 ff.)
49. Aldrich, John (2009) "England and Continental Probability in the Inter-War Years", *Electronic Journ@l for History of Probability and Statistics*, vol. 5/2, Decembre 2009 (http://www.jehps.net/decembre2009.html). (Section 3)
50. Jørgensen, Bent (1997). *The theory of dispersion models*. Chapman & Hall. ISBN 978-0412997112.

References

- Bárány, Imre; Vu, Van (2007), "Central limit theorems for Gaussian polytopes", *Annals of Probability* (Institute of Mathematical Statistics) **35** (4): 1593–1621, arXiv:math/0610192, doi:10.1214/009117906000000791
- Bauer, Heinz (2001), *Measure and Integration Theory*, Berlin: de Gruyter, ISBN 3110167190
- Billingsley, Patrick (1995), *Probability and Measure* (Third ed.), John Wiley & sons, ISBN 0-471-00710-2
- Bradley, Richard (2007), *Introduction to Strong Mixing Conditions* (First ed.), Heber City, UT: Kendrick Press, ISBN 0-9740427-9-X
- Bradley, Richard (2005), "Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions" (PDF), *Probability Surveys* **2**: 107–144, arXiv:math/0511078v1, doi:10.1214/154957805100000104
- Dinov, Ivo; Christou, Nicolas; Sanchez, Juana (2008), "Central Limit Theorem: New SOCR Applet and Demonstration Activity", *Journal of Statistics Education* (ASA) **16** (2)
- Durrett, Richard (2004), *Probability: theory and examples* (4th ed.), Cambridge University Press, ISBN 0521765390
- Gaposhkin, V.F. (1966), "Lacunary series and independent functions", *Russian Mathematical Surveys* **21** (6): 1–82, doi:10.1070/RM1966v021n06ABEH001196.
- Klartag, Bo'az (2007), "A central limit theorem for convex sets", *Inventiones Mathematicae* **168**, 91–131.doi:10.1007/s00222-006-0028-8 Also arXiv.
- Klartag, Bo'az (2008), "A Berry-Esseen type inequality for convex bodies with an unconditional basis", *Probability Theory and Related Fields*. doi:10.1007/s00440-008-0158-6 Also arXiv.

External links

- Simplified, step-by-step explanation of the classical Central Limit Theorem. (http://www.quantumfieldtheory.info/CentralLimitTheorem.pdf) with histograms at every step.
- Hands-on explanation of the Central Limit Theorem in tutorial videos from Khan Academy (https://www.khanacademy.org/math/probability/statistics-inferential/sampling_distribution/v/central-limit-theorem), with many examples
- Hazewinkel, Michiel, ed. (2001), "Central limit theorem", *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- Animated examples of the CLT (http://www.statisticalengineering.com/central_limit_theorem.html)
- Central Limit Theorem (http://www.vias.org/simulations/simusoft_cenlimit.html) interactive simulation to experiment with various parameters
- CLT in NetLogo (Connected Probability — ProbLab) (http://ccl.northwestern.edu/curriculum/ProbLab/CentralLimitTheorem.html) interactive simulation w/ a variety of modifiable parameters
- General Central Limit Theorem Activity (http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_GeneralCentralLimitTheorem) & corresponding SOCR CLT Applet (http://www.socr.ucla.edu/htmls/SOCR_Experiments.html) (Select the Sampling Distribution CLT Experiment from the drop-down list of SOCR Experiments (http://wiki.stat.ucla.edu/socr/index.php/About_pages_for_SOCR_Experiments))
- Generate sampling distributions in Excel (http://www.indiana.edu/~jkkteach/ExcelSampler/) Specify arbitrary population, sample size, and sample statistic.
- MIT OpenCourseWare Lecture 18.440 *Probability and Random Variables*, Spring 2011, Scott Sheffield Another proof. (http://ocw.mit.edu/courses/mathematics/18-440-probability-and-random-variables-spring-2011/lecture-notes/MIT18_440S11_Lecture31.pdf) Retrieved 2012-04-08.
- CAUSEweb.org (http://www.causeweb.org) is a site with many resources for teaching statistics including the Central Limit Theorem
- The Central Limit Theorem (http://demonstrations.wolfram.com/TheCentralLimitTheorem/) by Chris Boucher, Wolfram Demonstrations Project.
- Weisstein, Eric W., "Central Limit Theorem" (http://mathworld.wolfram.com/CentralLimitTheorem.html), *MathWorld*.
- Animations for the Central Limit Theorem (http://animation.yihui.name/prob:central_limit_theorem) by Yihui Xie using the R package animation (http://cran.r-project.org/package=animation)
- Teaching demonstrations of the CLT: **clt.examp** function in Greg Snow (2012). *TeachingDemos: Demonstrations for teaching and learning. R package version 2.8.*
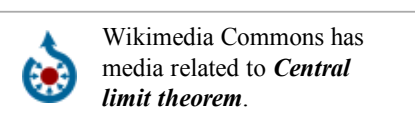
Wikimedia Commons has media related to *Central limit theorem*.

- This page was last modified on 1 February 2016, at 04:05.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using