

13.2: Convergence and the Central Limit Theorem

The Central Limit Theorem

The *central limit theorem* (CLT) asserts that if random variable X is the sum of a large class of independent random variables, each with reasonable distributions, then X is approximately normally distributed. This celebrated theorem has been the object of extensive theoretical research directed toward the discovery of the most general conditions under which it is valid. On the other hand, this theorem serves as the basis of an extraordinary amount of applied work. In the statistics of large samples, the sample average is a constant times the sum of the random variables in the sampling process. Thus, for large samples, the sample average is approximately normal—whether or not the population distribution is normal. In much of the theory of errors of measurement, the observed error is the sum of a large number of independent random quantities which contribute additively to the result. Similarly, in the theory of noise, the noise signal is the sum of a large number of random components, independently produced. In such situations, the assumption of a normal population distribution is frequently quite appropriate.

We consider a form of the CLT under hypotheses which are reasonable assumptions in many practical situations. We sketch a proof of this version of the CLT, known as the Lindeberg-Lévy theorem, which utilizes the limit theorem on characteristic functions, above, along with certain elementary facts from analysis. It illustrates the kind of argument used in more sophisticated proofs required for more general cases.

Consider an independent sequence $\{X_n : 1 \leq n\}$ of random variables. Form the sequence of partial sums

$$S_n = \sum_{i=1}^n X_i \quad \forall n \geq 1 \text{ with } E[S_n] = \sum_{i=1}^n E[X_i] \text{ and } \text{Var}[S_n] = \sum_{i=1}^n \text{Var}[X_i]$$

Let S_n^* be the standardized sum and let F_n be the distribution function for S_n^* . The CLT asserts that under appropriate conditions, $F_n(t) \rightarrow \phi(t)$ as $n \rightarrow \infty$ for all t . We sketch a proof of the theorem under the condition the X_i form an iid class.

Central Limit Theorem (Lindeberg-Lévy form)

If $\{X_n : 1 \leq n\}$ is iid, with

$$E[X_i] = \mu, \text{Var}[X_i] = \sigma^2, \text{ and } S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

then

$$F_n(t) \rightarrow \phi(t) \text{ as } n \rightarrow \infty, \text{ for all } t$$

IDEAS OF A PROOF

There is no loss of generality in assuming $\mu = 0$. Let ϕ be the common characteristic function for the X_i , and for each n let ϕ_n be the characteristic function for S_n^* . We have

$$\varphi(t) = E[e^{itX}] \text{ and } \varphi_n(t) = E[e^{itS_n^*}] = \varphi^n(t/\sigma\sqrt{n})$$

Using the power series expansion of φ about the origin noted above, we have

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + \beta(t) \quad \text{where } \beta(t) = o(t^2) \text{ as } t \rightarrow 0$$

This implies

$$[\varphi(t/\sigma\sqrt{n}) - (1 - t^2/2n)] = [\beta(t/\sigma\sqrt{n})] = o(t^2/\sigma^2 n)$$

so that

$$n[\varphi(t/\sigma\sqrt{n}) - (1 - t^2/2n)] \rightarrow 0 \text{ as } n \rightarrow \infty$$

A standard lemma of analysis ensures

$$(1 - \frac{t^2}{2n})^n \rightarrow e^{-t^2/2} \text{ as } n \rightarrow \infty$$

so that

$$\varphi(t/\sigma\sqrt{n}) \rightarrow e^{-t^2/2} \text{ as } n \rightarrow \infty \text{ for all } t$$

By the convergence theorem on characteristic functions, above, $F_n(t) \rightarrow \phi(t)$.

—□

The theorem says that the distribution functions for sums of increasing numbers of the X_i converge to the normal distribution function, but it does not tell how fast. It is instructive to consider some examples, which are easily worked out with the aid of our m-functions.

Demonstration of the central limit theorem

Discrete examples

We first examine the gaussian approximation in two cases. We take the sum of five iid simple random variables in each case. The first variable has six distinct values; the second has only three. The discrete character of the sum is more evident in the second case. Here we use not only the gaussian approximation, but the gaussian approximation shifted one half unit (the so called continuity correction for integer-values random variables). The fit is remarkably good in either case with only five terms.

A principal tool is the m-function *diidsum* (sum of discrete iid random variables). It uses a designated number of iterations of *mgsum*.

Example 13.2.1 First random variable

```
X = [-3.2 -1.05 2.1 4.6 5.3 7.2];
PX = 0.1*[2 2 1 3 1 1];
EX = X*PX'
EX = 1.9900
VX = dot(X.^2,PX) - EX^2
VX = 13.0904
[x,px] = diidsum(X,PX,5);           % Distribution for the sum of 5 iid rv
F = cumsum(px);                     % Distribution function for the sum
stairs(x,F)                          % Stair step plot
hold on
plot(x,gaussian(5*EX,5*VX,x),'-.') % Plot of gaussian distribution function
% Plotting details                  (see Figure 13.2.1)
```

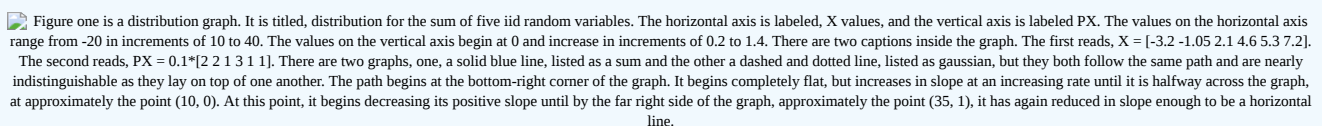


Figure 13.2.1. Distribution for the sum of five iid random variables.

Example 13.2.2 Second random variable

```
X = 1:3;
PX = [0.3 0.5 0.2];
EX = X*PX'
EX = 1.9000
EX2 = X.^2*PX'
EX2 = 4.1000
VX = EX2 - EX^2
VX = 0.4900
[x,px] = diidsum(X,PX,5);           % Distribution for the sum of 5 iid rv
F = cumsum(px);                     % Distribution function for the sum
stairs(x,F)                          % Stair step plot
```

```
hold on
plot(x,gaussian(5*EX,5*VX,x),'-.') % Plot of gaussian distribution function
plot(x,gaussian(5*EX,5*VX,x+0.5),'o') % Plot with continuity correction
% Plotting details (see Figure 13.2.2)
```

Figure two is a distribution graph. It is titled, distribution for the sum of five iid random variables. The horizontal axis is labeled, X values, and the vertical axis is labeled PX. The values on the horizontal axis range from 5 in increments of 1 to 15. The values on the vertical axis begin at 0 and increase in increments of 0.2 to 1.2. There are two captions inside the graph. The first reads, $X = [1\ 2\ 3]$. The second reads, $PX = [0.3\ 0.5\ 0.2]$. There are two graphs, one, a solid blue line, labeled step dbn fn, and the other a dashed and dotted line, labeled gaussian, but they both follow the same path. The step dbn fn is a series of horizontal line segments followed by vertical line segments in varying sizes that fit the shape of the smoother curve, the gaussian curve. A third labeled item is a series of small blue circles that sit at the upper corners of the steps of the solid lined curve, labeled, shifted gaussian. The path begins at the bottom-right corner of the graph. It begins completely flat, but increases in slope at an increasing rate until it is halfway across the graph, at approximately the point (5, 0). At this point, it begins decreasing its positive slope until by the far right side of the graph, approximately the point (14, 1), it has again reduced in slope enough to be a horizontal line.

Figure 13.2.2. Distribution for the sum of five iid random variables.

As another example, we take the sum of twenty one iid simple random variables with integer values. We examine only part of the distribution function where most of the probability is concentrated. This effectively enlarges the x-scale, so that the nature of the approximation is more readily apparent.

Example 13.2.3 Sum of twenty-one iid random variables

```
X = [0 1 3 5 6];
PX = 0.1*[1 2 3 2 2];
EX = dot(X,PX)
EX = 3.3000
VX = dot(X.^2,PX) - EX^2
VX = 4.2100
[x,px] = diidsum(X,PX,21);
F = cumsum(px);
FG = gaussian(21*EX,21*VX,x);
stairs(40:90,F(40:90))
hold on
plot(40:90,FG(40:90))
% Plotting details (see Figure 13.2.3)
```

Figure three is a distribution graph. It is titled, partial distribution for sum of 21 iid random variables. The horizontal axis is labeled, x-values, and the vertical axis is labeled PX. The values on the horizontal axis range in value from 40 to 90 at increments of 5, and the vertical axis ranges from 0 to 1 in increments of .1. There are two labeled equations. The first reads, $X = [0\ 1\ 3\ 5\ 6]$. The second reads, $PX = 0.1*[1\ 2\ 3\ 2\ 2]$. There are two graphs, one a smooth curve, labeled gaussian dbn, and the other a series of steps closely following the smooth curve, labeled Dbn for sum. Both graphs begin at the point (40, 0) at the bottom-left of the graph. The slope of the smooth curve is flat, and increases until approximately (70, 0.5). At this point, the graph continues increasing, but its slope begins decreasing until at approximately (90, 0.99), the path is again nearly flat. The steps follow the smooth curve along the same path.

Figure 13.2.3. Distribution for the sum of twenty one iid random variables.

Absolutely continuous examples

By use of the discrete approximation, we may get approximations to the sums of absolutely continuous random variables. The results on discrete variables indicate that the more values the more quickly the conversion seems to occur. In our next example, we start with a random variable uniform on (0, 1).

Example 13.2.4 Sum of three iid, uniform random variables.

Suppose $X \sim \text{uniform}(0, 1)$. Then $E[X] = 0.5$ and $\text{Var}[X] = 1/12$.

```
tappr
Enter matrix [a b] of x-range endpoints [0 1]
Enter number of x approximation points 100
Enter density as a function of t t<=1
Use row matrices X and PX as in the simple case
```

```
EX = 0.5;
VX = 1/12;
[z,pz] = diidsum(X,PX,3);
F = cumsum(pz);
FG = gaussian(3*EX,3*VX,z);
length(z)
ans = 298
a = 1:5:296; % Plot every fifth point
plot(z(a),F(a),z(a),FG(a),'o')
% Plotting details (see Figure 13.2.4)
```

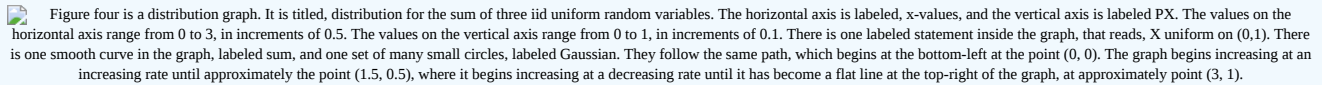


Figure 13.2.4. Distribution for the sum of three iid uniform random variables.

For the sum of only three random variables, the fit is remarkably good. This is not entirely surprising, since the sum of two gives a symmetric triangular distribution on (0, 2). Other distributions may take many more terms to get a good fit. Consider the following example.

Example 13.2.5 Sum of eight iid random variables

Suppose the density is one on the intervals (-1, -0.5) and (0.5, 1). Although the density is symmetric, it has two separate regions of probability. From symmetry, $E[X] = 0$. Calculations show $\text{Var}[X] = E[X^2] = 7/12$. The MATLAB computations are:

```
tappr
Enter matrix [a b] of x-range endpoints [-1 1]
Enter number of x approximation points 200
Enter density as a function of t (t<=-0.5)|(t>=0.5)
Use row matrices X and PX as in the simple case
[z,pz] = diidsum(X,PX,8);
VX = 7/12;
F = cumsum(pz);
FG = gaussian(0,8*VX,z);
plot(z,F,z,FG)
% Plotting details (see Figure 13.2.5)
```

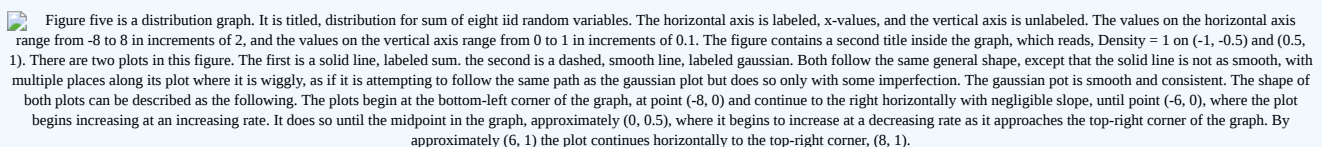


Figure 13.2.5. Distribution for the sum of eight iid uniform random variables.

Although the sum of eight random variables is used, the fit to the gaussian is not as good as that for the sum of three in [Example 13.2.4](#). In either case, the convergence is remarkable fast—only a few terms are needed for good approximation.

Convergence phenomena in probability theory

The central limit theorem exhibits one of several kinds of convergence important in probability theory, namely convergence *in distribution* (sometimes called *weak convergence*). The increasing concentration of values of the sample average random variable

A_n with increasing n illustrates convergence *in probability*. The convergence of the sample average is a form of the so-called *weak law of large numbers*. For large enough n the probability that A_n lies within a given distance of the population mean can be made as near one as desired. The fact that the variance of A_n becomes small for large n illustrates convergence *in the mean* (of order 2).

$$E[|A_n - \mu|^2] \rightarrow 0 \text{ as } n \rightarrow \infty$$

In the calculus, we deal with sequences of numbers. If $\{a_n : 1 \leq n\}$ is a sequence of real numbers, we say the sequence converges iff for N sufficiently large a_n approximates arbitrarily closely some number L for all $n \geq N$. This unique number L is called the limit of the sequence. Convergent sequences are characterized by the fact that for large enough N , the distance $|a_n - a_m|$ between any two terms is arbitrarily small for all $n, m \geq N$. Such a sequence is said to be *fundamental* (or *Cauchy*). To be precise, if we let $\epsilon > 0$ be the error of approximation, then the sequence is

- Convergent iff there exists a number L such that for any $\epsilon > 0$ there is an N such that

$$|L - a_n| \leq \epsilon \text{ for all } n \geq N$$

- Fundamental iff for any $\epsilon > 0$ there is an N such that

$$|a_n - a_m| \leq \epsilon \text{ for all } n, m \geq N$$

As a result of the completeness of the real numbers, it is true that any fundamental sequence converges (i.e., has a limit). And such convergence has certain desirable properties. For example the limit of a linear combination of sequences is that linear combination of the separate limits; and limits of products are the products of the limits.

The notion of convergent and fundamental sequences applies to sequences of real-valued functions with a common domain. For each x in the domain, we have a sequence

$\{f_n(x) : 1 \leq n\}$ of real numbers. The sequence may converge for some x and fail to converge for others.

A somewhat more restrictive condition (and often a more desirable one) for sequences of functions is *uniform convergence*. Here the uniformity is over values of the argument x . In this case, for any $\epsilon > 0$ there exists an N which works for all x (or for some suitable prescribed set of x).

These concepts may be applied to a sequence of random variables, which are real-valued functions with domain Ω and argument ω . Suppose $\{X_n : 1 \leq n\}$ is a sequence of real random variables. For each argument ω we have a sequence $\{X_n(\omega) : 1 \leq n\}$ of real numbers. It is quite possible that such a sequence converges for some ω and diverges (fails to converge) for others. As a matter of fact, in many important cases the sequence converges for all ω except possibly a set (event) of probability zero. In this case, we say the sequence converges *almost surely* (abbreviated a.s.). The notion of uniform convergence also applies. In probability theory we have the notion of *almost uniform convergence*. This is the case that the sequence converges uniformly for all ω except for a set of arbitrarily small probability.

The notion of convergence *in probability* noted above is a quite different kind of convergence. Rather than deal with the sequence on a pointwise basis, it deals with the random variables as such. In the case of sample average, the “closeness” to a limit is expressed in terms of the probability that the observed value $X_n(\omega)$ should lie close to the value $X(\omega)$ of the limiting random variable. We may state this precisely as follows:

A sequence $\{X_n : 1 \leq n\}$ converges to X in probability, designated $X_n \xrightarrow{P} X$ iff for any $\epsilon > 0$.

$$\lim_n P(|X - X_n| > \epsilon) = 0$$

There is a corresponding notion of a sequence fundamental in probability.

The following schematic representation may help to visualize the difference between almost-sure convergence and convergence in probability. In setting up the basic probability model, we think in terms of “balls” drawn from a jar or box. Instead of balls, consider for each possible outcome ω a “tape” on which there is the sequence of values $X_1(\omega), X_2(\omega), X_3(\omega), \dots$.

- If the sequence of random variable converges a.s. to a random variable X , then there is a set of “exceptional tapes” which has zero probability. For all other tapes, $X_n(\omega) \rightarrow X(\omega)$. This means that by going far enough out on *any* such tape, the values $X_n(\omega)$ beyond that point all lie within a prescribed distance of the value $X(\omega)$ of the limit random variable.
- If the sequence converges in probability, the situation may be quite different. A tape is selected. For n sufficiently large, the probability is arbitrarily near one that the observed value $X_n(\omega)$ lies within a prescribed distance of $X(\omega)$. This says nothing about the values $X_m(\omega)$ on the selected tape for any larger m . In fact, the sequence on the selected tape may very well diverge.

It is not difficult to construct examples for which there is convergence in probability but pointwise convergence for *no* ω . It is easy to confuse these two types of convergence. The kind of convergence noted for the sample average is convergence in probability (a “weak” law of large numbers). What is really desired in most cases is a.s. convergence (a “strong” law of large numbers). It turns out that for a sampling process of the kind used in simple statistics, the convergence of the sample average is almost sure (i.e., the strong law holds). To establish this requires much more detailed and sophisticated analysis than we are prepared to make in this treatment.

The notion of mean convergence illustrated by the reduction of $\text{Var}[A_n]$ with increasing n may be expressed more generally and more precisely as follows. A sequence $\{X_n : 1 \leq n\}$ converges in the mean of order p to X iff

$$E[|X - X_n|^p] \rightarrow 0 \text{ as } n \rightarrow \infty \text{ designated } X_n \xrightarrow{L^p} X; \text{ as } n \rightarrow \infty$$

If the order p is one, we simply say the sequence converges in the mean. For $p = 2$, we speak of *mean-square convergence*.

The introduction of a new type of convergence raises a number of questions.

1. There is the question of fundamental (or Cauchy) sequences and convergent sequences.
2. Do the various types of limits have the usual properties of limits? Is the limit of a linear combination of sequences the linear combination of the limits? Is the limit of products the product of the limits?
3. What conditions imply the various kinds of convergence?
4. What is the relation between the various kinds of convergence?

Before sketching briefly some of the relationships between convergence types, we consider one important condition known as *uniform integrability*. According to the property (E9b) for integrals

$$X \text{ is integrable iff } E[I_{\{|X_t| > a\}} |X_t|] \rightarrow 0 \text{ as } a \rightarrow \infty$$

Roughly speaking, to be integrable a random variable cannot be too large on too large a set. We use this characterization of the integrability of a single random variable to define the notion of the uniform integrability of a class.

Definition

An arbitray class $\{X_t : t \in T\}$ is *uniformly integrable* (abbreviated u.i.) with respect to probability measure P iff

$$\sup_{t \in T} E[I_{\{|X_t| > a\}} |X_t|] \rightarrow 0 \text{ as } a \rightarrow \infty$$

This condition plays a key role in many aspects of theoretical probability.

The relationships between types of convergence are important. Sometimes only one kind can be established. Also, it may be easier to establish one type which implies another of more immediate interest. We simply state informally some of the important relationships. A somewhat more detailed summary is given in PA, Chapter 17. But for a complete treatment it is necessary to consult more advanced treatments of probability and measure.

Relationships between types of convergence for probability measures

Consider a sequence $\{X_n : 1 \leq n\}$ of random variables.

- It converges almost surely iff it converges almost uniformly.
- If it converges almost surely, then it converges in probability.
- It converges in mean, order p , iff it is uniformly integrable and converges in probability.
- If it converges in probability, then it converges in distribution (i.e. weakly).

Various chains of implication can be traced. For example

- Almost sure convergence *implies* convergence in probability *implies* convergence in distribution.
- Almost sure convergence and uniform integrability *implies* convergence in mean p .

We do not develop the underlying theory. While much of it could be treated with elementary ideas, a complete treatment requires considerable development of the underlying measure theory. However, it is important to be aware of these various types of convergence, since they are frequently utilized in advanced treatments of applied probability and of statistics.

This page titled [13.2: Convergence and the Central Limit Theorem](#) is shared under a [CC BY 3.0](#) license and was authored, remixed, and/or curated by [Paul Pfeiffer](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.