Chapter 12

Convergence, Sequences and Limit Theorems

Some of the most astonishing results in probability are related to the properties of sequences of random variables and the convergence of empirical distributions. From an engineering viewpoint, these results are important as they enable the efficient design of complex systems with very small probabilities of failure. Concentration behavior facilitates true economies of scale.

12.1   Types of Convergence

The premise on which most probabilistic convergence results lie is a sequence of random variables $X_1, X_2, \ldots$ and a limiting random variable X, all of which are defined on the same probability space. Recall that a random variable is a function of the outcome of a random experiment. The above statement stipulate that all the random variables listed above are functions of the outcome of a same experiment.

Statements that can be made about a sequence of random variables range from simple assertions to more intricate claims. For instance, the sequence may appear to move toward a deterministic quantity or to behave increasingly akin to a certain function. Alternatively, the CDFs of the random variables in the sequence may appear to approach a precise function. Being able to recognize specific patterns within the sequence is key in establishing converge

results. The various statement one can make about the sequence $X_1, X_2, \ldots$ lead to the different types of convergence encountered in probability. Below, we discuss briefly three types of convergence.

Example 105. *Suppose that $X_1, X_2, \ldots$ is a sequence of independent Gaussian random variables, each with mean m and variance σ^2. Define the partial sums*

$$S_n = \sum_{i=1}^{n} X_i, \tag{12.1}$$

and consider the sequence

$$S_1, \frac{S_2}{2}, \frac{S_3}{3}, \ldots \tag{12.2}$$

We know that affine transformations of Gaussian random variables remain Gaussian. Furthermore, we know that sums of jointly Gaussian random variables are also Gaussian. Thus, S_n/n possesses a Gaussian distribution with mean

$$\mathrm{E}\left[\frac{S_n}{n}\right] = \frac{\mathrm{E}[S_n]}{n} = \frac{\mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]}{n} = m$$

and variance

$$\mathrm{Var}\left[\frac{S_n}{n}\right] = \frac{\mathrm{Var}[S_n]}{n^2} = \frac{\mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n]}{n^2} = \frac{\sigma^2}{n}.$$

It appears that the PDF of S_n/n concentrates around m as n approaches infinity. That is, the sequence in (12.2) seems to become increasingly predictable.

Example 106. *Again, let $X_1, X_2, \ldots$ be the sequence described above, and let S_n be defined according to (12.1). This time, we wish to characterize the properties of*

$$S_1 - m, \frac{S_2 - 2m}{\sqrt{2}}, \frac{S_3 - 3m}{\sqrt{3}}, \ldots$$

From our current discussion, we know that $(S_n - nm)/\sqrt{n}$ is a Gaussian random variables. We can compute its mean and variance as follows

$$\mathrm{E}\left[\frac{S_n - nm}{\sqrt{n}}\right] = \frac{\mathrm{E}[S_n - nm]}{\sqrt{n}} = 0$$

$$\mathrm{Var}\left[\frac{S_n - mn}{\sqrt{n}}\right] = \frac{\mathrm{Var}[S_n - mn]}{n} = \frac{\mathrm{Var}[S_n]}{n} = \sigma^2.$$

No matter how large n is, the random variable $(S_n - nm)/\sqrt{n}$ has a Gaussian distribution with mean zero and variance σ^2. Intriguingly, the distributions remains invariant throughout the sequence.

12.1.1 Convergence in Probability

The basic concept behind the definition of *convergence in probability* is that the probability that a random variable deviates from its typical behavior becomes less likely as the sequence progresses. Formally, a sequence $X_1, X_2, \ldots$ of random variables converges in probability to X if for every $\epsilon > 0$,

$$\lim_{n \to \infty} \Pr\left(|X_n - X| \geq \epsilon\right) = 0.$$

In Example 105, the sequence $\{S_n/n\}$ converges in probability to m.

12.1.2 Mean Square Convergence

We say that a sequence $X_1, X_2, \ldots$ of random variables *converges in mean square* to X if

$$\lim_{n \to \infty} \mathrm{E}\left[|X_n - X|^2\right] = 0.$$

That is, the second moment of the difference between X_n and X vanishes as n goes to infinity. Convergence in the mean square sense implies convergence in probability.

Proposition 9. *Let $X_1, X_2, \ldots$ be a sequence of random variables that converge in mean square to X. Then, the sequence $X_1, X_2, \ldots$ also converges to X in probability.*

Proof. Suppose that $\epsilon > 0$ is fixed. The sequence $X_1, X_2, \ldots$ converges in mean square to X. Thus, for $\delta > 0$, there exists an N such that $n \geq N$ implies

$$\mathrm{E}\left[|X_n - X|^2\right] < \delta.$$

If we apply the Chebyshev inequality to $X_n - X$, we get

$$\Pr\left(|X_n - X| \geq \epsilon\right) \leq \frac{\mathrm{E}\left[|X_n - X|^2\right]}{\epsilon^2} < \frac{\delta}{\epsilon^2}$$

Since δ can be made arbitrarily small, we conclude that this sequence also converges to X in probability. $\square$

12.1.3 Convergence in Distribution

A sequence $X_1, X_2, \ldots$ of random variables is said to *converge in distribution* to a random variable X if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

at every point $x \in \mathbb{R}$ where $F_X(\cdot)$ is continuous. This type of convergence is also called *weak convergence.*

Example 107. *Let X_n be a continuous random variable that is uniformly distributed over $[0, 1/n]$. Then, the sequence $X_1, X_2, \ldots$ converges in distribution to 0.*

In this example, $X = 0$ and

$$F_X(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0. \end{cases}$$

Furthermore, for every $x < 0$, we have $F_{X_n}(x) = 0$; and for every $x > 0$, we have

$$\lim_{n \to \infty} F_{X_n}(x) = 1.$$

Hence, the sequence $X_1, X_2, \ldots$ converges in distribution to a constant.

12.2 The Law of Large Numbers

The law of large numbers focuses on the convergence of empirical averages. Although, there are many versions of this law, we only state its simplest form below. Suppose that $X_1, X_2, \ldots$ is a sequence of independent and identically distributed random variable, each with finite second moment. Furthermore, for $n \geq 1$, define the empirical sum

$$S_n = \sum_{i=1}^{n} X_n.$$

The law of large number asserts that the sequence of empirical averages,

$$\frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_n,$$

converges in probability to the mean $\mathrm{E}[X]$.

Theorem 4 (Law of Large Numbers). *Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean $\mathrm{E}[X]$ and finite variance. For every $\epsilon > 0$, we have*

$$\lim_{n \to \infty} \Pr\left(\left|\frac{S_n}{n} - \mathrm{E}[X]\right| \geq \epsilon\right) = \lim_{n \to \infty} \Pr\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mathrm{E}[X]\right| \geq \epsilon\right) = 0.$$

Proof. Taking the expectation of the empirical average, we obtain

$$\mathrm{E}\left[\frac{S_n}{n}\right] = \frac{\mathrm{E}[S_n]}{n} = \frac{\mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]}{n} = \mathrm{E}[X].$$

Using independence, we also have

$$\mathrm{Var}\left[\frac{S_n}{n}\right] = \frac{\mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n]}{n^2} = \frac{\mathrm{Var}[X]}{n}.$$

As n goes to infinity, the variance of the empirical average S_n/n vanishes. Thus, we showed that the sequence $\{S_n/n\}$ of empirical averages converges in mean square to $\mathrm{E}[X]$ since

$$\lim_{n \to \infty} \mathrm{E}\left[\left|\frac{S_n}{n} - \mathrm{E}[X]\right|^2\right] = \lim_{n \to \infty} \mathrm{Var}\left[\frac{S_n}{n}\right] = 0.$$

To get convergence in probability, we apply the Chebyshev inequality, as we did in Proposition 9,

$$\Pr\left(\left|\frac{S_n}{n} - \mathrm{E}[X]\right| \geq \epsilon\right) \leq \frac{\mathrm{Var}\left[\frac{S_n}{n}\right]}{\epsilon^2} = \frac{\mathrm{Var}[X]}{n\epsilon^2},$$

which clearly goes to zero as n approaches infinity. $\qquad\square$

Example 108. *Suppose that a die is thrown repetitively. We are interested in the average number of times a six shows up on the top face, as the number of throws becomes very large.*

Let D_n be a random variable that represent the number on the nth roll. Also, define the random variable $X_n = \mathbf{1}_{\{D_n=6\}}$. Then, X_n is a Bernoulli random variable with parameter $p = 1/6$, and the empirical average S_n/n is equal to the number of times a six is observed divided by the total number of rolls. By the law of large numbers, we have

$$\lim_{n \to \infty} \Pr\left(\left|\frac{S_n}{n} - \frac{1}{6}\right| \geq \epsilon\right) = 0.$$

That is, as the number of rolls increases, the average number of times a six is observe converges to the probability of getting a six.

12.2.1 Heavy-Tailed Distributions*

There are situations where the law of large numbers does not apply. For example, when dealing with *heavy-tailed distributions*, one needs to be very careful. In this section, we study Cauchy random variables in greater details. First, we show that the sum of two independent Cauchy random variables is itself a Cauchy random variable.

Let X_1 and X_2 be two independent Cauchy random variables with parameter γ_1 and γ_2, respectively. We wish to compute the PDF of $S = X_1 + X_2$. For continuous random variable, the PDF of S is given by the convolution of $f_{X_1}(\cdot)$ and $f_{X_2}(\cdot)$. Thus, we can write

$$f_S(x) = \int_{-\infty}^{\infty} f_{X_1}(\xi) f_{X_2}(x - \xi) d\xi$$

$$= \int_{-\infty}^{\infty} \frac{\gamma_1}{\pi \left(\gamma_1^2 + \xi^2\right)} \frac{\gamma_2}{\pi \left(\gamma_2^2 + (x - \xi)^2\right)} d\xi.$$

This integral is somewhat difficult to solve. We therefore resort to complex analysis and contour integration to get a solution. Let C be a contour that goes along the real line from $-a$ to a, and then counterclockwise along a semicircle centered at zero. For a large enough, Cauchy's residue theorem requires that

$$\oint_C f_{X_1}(z) f_{X_2}(x - z) dz = \oint_C \frac{\gamma_1}{\pi \left(\gamma_1^2 + z^2\right)} \frac{\gamma_2}{\pi \left(\gamma_2^2 + (x - z)^2\right)} dz \quad (12.3)$$

$$= 2\pi i \left(\text{Res}(g, i\gamma_1) + \text{Res}(g, x + i\gamma_2)\right)$$

where we have implicitly defined the function

$$g(z) = \frac{\gamma_1 \gamma_2}{\pi^2 \left(\gamma_1^2 + z^2\right) \left(\gamma_2^2 + (z - x)^2\right)}$$

$$= \frac{\gamma_1 \gamma_2}{\pi^2 (z - i\gamma_1)(z + i\gamma_1)(z - x + i\gamma_2)(z - x - i\gamma_2)}.$$

Only two residues are contained within the enclosed region. Because they are simple poles, their values are given by

$$\text{Res}(g, i\gamma_1) = \lim_{z \to i\gamma_1} (z - i\gamma_1) g(z) = \frac{\gamma_2}{2i\pi^2 \left((x - i\gamma_1)^2 + \gamma_2^2\right)}$$

$$\text{Res}(g, x + i\gamma_2) = \lim_{z \to x + i\gamma_2} (z - x - i\gamma_2) g(z) = \frac{\gamma_1}{2i\pi^2 \left((x + i\gamma_2)^2 + \gamma_1^2\right)}.$$

It follows that

$$
2\pi i \left(\mathrm{Res}(g, i\gamma_1) + \mathrm{Res}(g, x + i\gamma_2) \right)
$$
$$
= \frac{\gamma_2}{\pi \left((x - i\gamma_1)^2 + \gamma_2^2 \right)} + \frac{\gamma_1}{\pi \left((x + i\gamma_2)^2 + \gamma_1^2 \right)}
$$
$$
= \frac{(\gamma_1 + \gamma_2)}{\pi \left((\gamma_1 + \gamma_2)^2 + x^2 \right)}
$$

The contribution of the arc in (12.3) vanishes as $a \to \infty$. We then conclude that the PDF of S is equal to

$$
f_S(x) = \frac{(\gamma_1 + \gamma_2)}{\pi \left((\gamma_1 + \gamma_2)^2 + x^2 \right)}.
$$

The sum of two independent Cauchy random variables with parameters γ_1 and γ is itself a Cauchy random variable with parameter $\gamma_1 + \gamma_2$.

Let $X_1, X_2, \ldots$ form a sequence of independent Cauchy random variables, each with parameter γ. Also, consider the empirical sum

$$
S_n = \sum_{i=1}^{n} X_n.
$$

Using mathematical induction and the aforementioned fact, it is possible to show that S_n is a Cauchy random variable with parameter $n\gamma$. Furthermore, for $x \in \mathbb{R}$, the PDF of the empirical average S_n/n is given by

$$
\frac{f_{S_n}(nx)}{\left| \frac{1}{n} \right|} = \frac{n^2\gamma}{\pi \left(n^2\gamma^2 + (nx)^2 \right)} = \frac{\gamma}{\pi \left(\gamma^2 + x^2 \right)},
$$

where we have used the methodology developed in Chapter 9. Amazingly, the empirical average of a sequence of independent Cauchy random variables, each with parameter γ, remains a Cauchy random variable with the same parameter. Clearly, the law of large numbers does not apply to this scenario. Note that our version of the law of large numbers requires random variables to have finite second moments, a condition that is clearly violated by the Cauchy distribution. This explains why convergence does not take place in this situation.

12.3 The Central Limit Theorem

The *central limit theorem* is a remarkable result in probability; it partly explains the prevalence of Gaussian random variables. In some sense, it captures

the behavior of large sums of small, independent random components.

Theorem 5 (Central Limit Theorem). *Let $X_1, X_2, \ldots$ be independent and identically distributed random variables, each with mean $\mathrm{E}[X]$ and variance σ^2. The distribution of*

$$\frac{S_n - n\mathrm{E}[X]}{\sigma\sqrt{n}}$$

converges in distribution to a standard normal random variable as $n \to \infty$. In particular, for any $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \Pr\left(\frac{S_n - n\mathrm{E}[X]}{\sigma\sqrt{n}} \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} \, d\xi.$$

Proof. Initially, we assume that $\mathrm{E}[X] = 0$ and $\sigma^2 = 1$. Furthermore, we only study the situation where the moment generating function of X exists and is finite. Consider the log-moment generating function of X,

$$\Lambda_X(s) = \log M_X(s) = \log \mathrm{E}\left[e^{sX}\right].$$

The first two derivatives of $\Lambda_X(s)$ are equal to

$$\frac{d\Lambda_X}{ds}(s) = \frac{1}{M_X(s)} \frac{dM_X}{ds}(s)$$

$$\frac{d^2\Lambda_X}{ds^2}(s) = -\left(\frac{1}{M_X(s)}\right)^2 \left(\frac{dM_X}{ds}(s)\right)^2 + \frac{1}{M_X(s)} \frac{d^2 M_X}{ds^2}(s).$$

Collecting these results and evaluating the functions at zero, we get $\Lambda_X(0) = 0$, $\frac{d\Lambda_X}{ds}(0) = \mathrm{E}[X] = 0$, and $\frac{d^2\Lambda_X}{ds^2}(0) = \mathrm{E}\left[X^2\right] = 1$. Next, we study the log-moment generating function of $S_n/\sqrt{n}$. Recall that the expectation of a product of independent random variables is equal to the product of their individual expectations. Using this property, we get

$$\log \mathrm{E}\left[e^{sS_n/\sqrt{n}}\right] = \log \mathrm{E}\left[e^{sX_1/\sqrt{n}} \cdots e^{sX_n/\sqrt{n}}\right]$$

$$= \log\left(M_X\left(\frac{s}{\sqrt{n}}\right) \cdots M_X\left(\frac{s}{\sqrt{n}}\right)\right)$$

$$= n\Lambda_X\left(sn^{-1/2}\right).$$

To explore the asymptotic behavior of the sequence $\{S_n/\sqrt{n}\}$, we take the limit of $n\Lambda_X\left(sn^{-1/2}\right)$ as $n \to \infty$. In doing so, notice the double use of L'Hôpital's

rule,

$$\lim_{n\to\infty} \frac{1}{n^{-1}} \Lambda\left(sn^{-1/2}\right) = \lim_{n\to\infty} \frac{1}{n^{-2}} \frac{d\Lambda}{ds}\left(sn^{-1/2}\right) \frac{sn^{-3/2}}{2}$$

$$= \lim_{n\to\infty} \frac{s}{2n^{-1/2}} \frac{d\Lambda}{ds}\left(sn^{-1/2}\right)$$

$$= \lim_{n\to\infty} \frac{s}{n^{-3/2}} \frac{d^2\Lambda}{ds^2}\left(sn^{-1/2}\right) \frac{sn^{-3/2}}{2}$$

$$= \lim_{n\to\infty} \frac{s^2}{2} \frac{d^2\Lambda}{ds^2}\left(sn^{-1/2}\right) = \frac{s^2}{2}.$$

That is, the moment generating function of $S_n/\sqrt{n}$ converges point-wise to $e^{s^2/2}$ as $n \to \infty$. In fact, this implies that

$$\Pr\left(\frac{S_n}{\sqrt{n}} \le x\right) \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi$$

as $n \to \infty$. In words, $S_n/\sqrt{n}$ converges in distribution to a standard normal random variable. The more general case where $E[X]$ and $Var[X]$ are arbitrary constants can be established in an analog manner by proper scaling of the random variables $X_1, X_2, \ldots$ □

In the last step of the proof, we stated that point-wise convergence of the moment generating functions implies convergence in distribution. This is a sophisticated result that we quote without proof.

12.3.1   Normal Approximation

The central limit theorem can be employed to approximate the CDF of large sums. Again, let

$$S_n = X_1 + \cdots + X_n$$

where $X_1, X_2, \ldots$ are independent and identically distributed random variables with mean $E[X]$ and variance σ^2. When n is large, the CDF of S_n can be estimated by approximating

$$\frac{S_n - nE[X]}{\sigma\sqrt{n}}$$

as a standard normal random variable. More specifically, we have

$$F_{S_n}(x) = \Pr(S_n \le x) = \Pr\left(\frac{S_n - nE[X]}{\sigma\sqrt{n}} \le \frac{x - nE[X]}{\sigma\sqrt{n}}\right)$$

$$\approx \Phi\left(\frac{x - nE[X]}{\sigma\sqrt{n}}\right),$$

where $\Phi(\cdot)$ is the CDF of a standard normal random variable.

Further Reading

1. Ross, S., *A First Course in Probability*, 7th edition, Pearson Prentice Hall, 2006: Chapter 8.

2. Bertsekas, D. P., and Tsitsiklis, J. N., *Introduction to Probability*, Athena Scientific, 2002: Sections 7.2–7.4.

3. Miller, S. L., and Childers, D. G., *Probability and Random Processes with Applications to Signal Processing and Communications*, 2004: Chapter 7.