Modes of Convergence

Electrical Engineering 126 (UC Berkeley)

Spring 2018

There is only one sense in which a sequence of real numbers $(a_n)_{n\in\mathbb{N}}$ is said to *converge* to a limit. Namely, $a_n \xrightarrow[n\to\infty]{} a$ if for every $\varepsilon > 0$ there exists a positive integer N such that the sequence after N is always within ε of the supposed limit a. In contrast, the notion of convergence becomes somewhat more subtle when discussing convergence of functions. In this note we briefly describe a few **modes of convergence** and explain their relationship. Since the subject quickly becomes very technical, we will state many of the fundamental results without proof.

Throughout this discussion, fix a probability space Ω and a sequence of random variables $(X_n)_{n\in\mathbb{N}}$. Also, let X be another random variable.

1 Almost Sure Convergence

The sequence $(X_n)_{n\in\mathbb{N}}$ is said to **converge almost surely** or **converge with probability one** to the limit X, if the set of outcomes $\omega \in \Omega$ for which $X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)$ forms an event of probability one. In other words, all observed realizations of the sequence $(X_n)_{n\in\mathbb{N}}$ converge to the limit. We abbreviate "almost surely" by "a.s." and we denote this mode of convergence by $X_n \xrightarrow[n\to\infty]{\text{a.s.}} X$. Of course, one could define an even stronger notion of convergence in which we require $X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)$ for *every* outcome (rather than for a set of outcomes with probability one), but the philosophy of probabilists is to disregard events of probability zero, as they are never observed. Thus, we regard a.s. convergence as the strongest form of convergence.

One of the most celebrated results in probability theory is the statement that the sample average of identically distributed random variables, under *very* weak assumptions, converges a.s. to the expected value of their common distribution. This is known as the **Strong Law of Large Numbers (SLLN)**.

Theorem 1 (Strong Law of Large Numbers)**.** *If $(X_n)_{n\in\mathbb{N}}$ are pairwise independent and identically distributed with $\mathbb{E}[|X_1|] < \infty$, then $n^{-1}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{a.s.} \mathbb{E}[X_1]$.*

An example we will see later in the course is in the context of *discrete-time Markov chains*: does the fraction of time spent in a state converge a.s. to a value, and if so, to what value? Another example of a.s. convergence that we will study is the *asymptotic equipartition property* from information theory, and its relevance to coding. Finally, another question of interest comes from machine learning: if we use the *stochastic gradient descent* algorithm to minimize a function, do the iterates converge a.s. to the true minimizer of the function?

2 Convergence in Probability

Next, $(X_n)_{n\in\mathbb{N}}$ is said to **converge in probability** to X, denoted $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$, if for every $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow[n\to\infty]{} 0$. In other words, for any fixed $\varepsilon > 0$, the probability that the sequence deviates from the supposed limit X by more than ε becomes vanishingly small. We now seek to prove that a.s. convergence implies convergence in probability.

Theorem 2. *If* $X_n \xrightarrow[n\to\infty]{a.s.} X$*, then* $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$*.*

Proof. Fix $\varepsilon > 0$. Define $A_n := \bigcup_{m=n}^{\infty}\{|X_m - X| > \varepsilon\}$ to be the event that at least one of $X_n, X_{n+1}, \ldots$ deviates from X by more than ε. Observe that $A_1 \supseteq A_2 \supseteq \cdots$ decreases to an event A which has probability zero, since the a.s. convergence of the sequence $(X_n)_{n\in\mathbb{N}}$ implies that for all outcomes ω (outside of an event of probability zero), the sequence of real numbers $(|X_n(\omega) - X(\omega)|)_{n\in\mathbb{N}}$ is eventually bounded by ε. Thus,

$$\mathbb{P}(|X_n - X| > \varepsilon) \le \mathbb{P}(A_n) \xrightarrow[n\to\infty]{} \mathbb{P}(A) = 0. \qquad \square$$

However, the converse is not true.

Example 1. The standard example of a sequence of random variables which converges in probability but not a.s. is the following. First, set $X_n = 0$ for all $n \in \mathbb{N}$. Then, for each $j \in \mathbb{N}$, pick an index N_j uniformly at random from $\{2^j, \ldots, 2^{j+1} - 1\}$ and set $X_{N_j} = 1$. Observe that $\mathbb{P}(X_n \neq 0) = \mathbb{P}(N_{\lfloor \log_2 n \rfloor} = n) = 2^{-\lfloor \log_2 n \rfloor} \xrightarrow[n\to\infty]{} 0$, so $X_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$. However, for *any* $\omega \in \Omega$, the sequence of real numbers $(X_n(\omega))_{n\in\mathbb{N}}$ takes on both values 0 and 1 infinitely often, so $(X_n(\omega))_{n\in\mathbb{N}}$ does not converge and hence $(X_n)_{n\in\mathbb{N}}$ does *not* converge a.s. to X.

As a consequence of the SLLN (Theorem 1) and Theorem 2, then if $(X_n)_{n\in\mathbb{N}}$ are pairwise independent and identically distributed with $\mathbb{E}[|X_1|] < \infty$, then $n^{-1}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[X_1]$. This is known as the **Weak Law of Large Numbers (WLLN)**.

The distinction between a.s. convergence and convergence in probability manifests itself in applications in the following way. If you have convergence in probability, then you know that the *probability* of a deviation of any particular size goes to zero, but you may indeed observe such deviations forever; if you had to pay a dollar for each ε-devation, you might end up paying up infinite dollars. In contrast, with a.s. convergence you are assured that for any observed realization, there will come a time in the sequence after which there will never be any such deviations, and thus you will only lose a finite amount of money.

3 Convergence in Distribution

Finally, the last mode of convergence that we will discuss is **convergence in distribution** or **convergence in law**. Here, $X_n \xrightarrow[n\to\infty]{\mathsf{d}} X$ if for each $x \in \mathbb{R}$ such that $\mathbb{P}(X = x) = 0$, we have $\mathbb{P}(X_n \le x) \xrightarrow[n\to\infty]{} \mathbb{P}(X \le x)$. Notice that unlike the previous two forms of convergence, convergence in distribution does not require all of the random variables to be defined on the same probability space.

First we explain why we require $\mathbb{P}(X_n \le x) \xrightarrow[n\to\infty]{} \mathbb{P}(X \le x)$ only at points x for which $\mathbb{P}(X = x) = 0$.

Example 2. Consider the sequence of constant random variables $(X_n)_{n\in\mathbb{N}}$, where we define $X_n := 2^{-n}$. We would like to assert that $X_n \xrightarrow[n\to\infty]{\mathsf{d}} X$, where $X := 0$. However, $\mathbb{P}(X_n \le 0) = 0$ for all $n \in \mathbb{N}$, whereas $\mathbb{P}(X \le 0) = 1$, so in particular $\mathbb{P}(X_n \le 0)$ does not converge to $\mathbb{P}(X \le 0)$. Notice that in this example, $\mathbb{P}(X = 0) = 1$, so we can fix this issue by only looking at points x for which $\mathbb{P}(X = x) = 0$, i.e., points at which the CDF of X is continuous.

In the following important special cases, convergence in distribution is easier to describe:

Theorem 3. *1. If $(X_n)_{n\in\mathbb{N}}$ and X take values in $\mathbb{Z}$, and if $\mathbb{P}(X_n = x) \xrightarrow[n\to\infty]{} \mathbb{P}(X = x)$ for all $x \in \mathbb{Z}$, then $X_n \xrightarrow[n\to\infty]{\mathsf{d}} X$.*

2. If $(X_n)_{n\in\mathbb{N}}$ and X are continuous random variables, and if $f_{X_n}(x) \xrightarrow[n\to\infty]{} f_X(x)$ for all $x \in \mathbb{R}$, then $X_n \xrightarrow[n\to\infty]{\mathsf{d}} X$.

Next we show that convergence in probability implies convergence in distribution.

Theorem 4. *If $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$, then $X_n \xrightarrow[n\to\infty]{\mathsf{d}} X$.*

Proof. Let x be a point such that $\mathbb{P}(X = x) = 0$. Fix $\varepsilon > 0$. We can write

$$\begin{aligned}\mathbb{P}(X_n \le x) &= \mathbb{P}(X_n \le x,\ |X_n - X| < \varepsilon) + \mathbb{P}(X_n \le x,\ |X_n - X| \ge \varepsilon)\\ &\le \mathbb{P}(X \le x + \varepsilon) + \mathbb{P}(|X_n - X| \ge \varepsilon).\end{aligned}$$

Similarly, we write

$$\begin{aligned}\mathbb{P}(X \le x - \varepsilon) &= \mathbb{P}(X \le x - \varepsilon,\ |X_n - X| < \varepsilon) + \mathbb{P}(X \le x - \varepsilon,\ |X_n - X| \ge \varepsilon)\\ &\le \mathbb{P}(X_n \le x) + \mathbb{P}(|X_n - X| \ge \varepsilon).\end{aligned}$$

Combining the two bounds, we have

$$\mathbb{P}(X \le x - \varepsilon) - \mathbb{P}(|X_n - X| \ge \varepsilon) \le \mathbb{P}(X_n \le x) \le \mathbb{P}(X \le x + \varepsilon) + \mathbb{P}(|X_n - X| \ge \varepsilon).$$

Since $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$, then $\mathbb{P}(|X_n - X| \ge \varepsilon) \xrightarrow[n\to\infty]{} 0$, so the bound above tells us that eventually the sequence $(\mathbb{P}(X_n \le x))_{n\in\mathbb{N}}$ is trapped between $\mathbb{P}(X \le x - \varepsilon)$ and $\mathbb{P}(X \le x + \varepsilon)$. This is true for all $\varepsilon > 0$ and the CDF of X is continuous at x by assumption, so by taking $\varepsilon \to 0$, we conclude that $\mathbb{P}(X_n \le x) \xrightarrow[n\to\infty]{} \mathbb{P}(X \le x)$. □

The converse is not true: convergence in distribution does not imply convergence in probability. In fact, a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ can converge in distribution even if they are not jointly defined on the same sample space! (This is because convergence in distribution is a property only of their marginal distributions.) In contrast, convergence in probability requires the random variables $(X_n)_{n\in\mathbb{N}}$ to be jointly defined on the same sample space, and determining whether or not convergence in probability holds requires some knowledge about the joint distribution of $(X_n)_{n\in\mathbb{N}}$. Even when the random variables $(X_n)_{n\in\mathbb{N}}$ are jointly defined, it is possible to construct counterexamples:

Example 3. Let $X_0 \sim \mathrm{Uniform}[-1,1]$, and for each positive integer n, let $X_n := (-1)^n X_0$. Then, $X_n \overset{\mathsf{d}}{=} \mathrm{Uniform}[-1,1]$ for all $n \in \mathbb{N}$ because the $\mathrm{Uniform}[-1,1]$ distribution is symmetric around the origin, so convergence in distribution holds (for a silly reason: all of the marginal distributions are the same). However, $(X_n)_{n\in\mathbb{N}}$ does not converge in probability (think about why this is true).

Thus we have built a small hierarchy

$$\Big(X_n \xrightarrow[n\to\infty]{\text{a.s.}} X\Big) \xRightarrow{\textit{Theorem 2}} \Big(X_n \xrightarrow[n\to\infty]{\mathbb{P}} X\Big) \xRightarrow{\textit{Theorem 4}} \Big(X_n \xrightarrow[n\to\infty]{\mathsf{d}} X\Big).$$

We now precisely state the **Central Limit Theorem (CLT)**, which is an assertion about convergence in distribution.

Theorem 5 (Central Limit Theorem)**.** *If $(X_n)_{n\in\mathbb{N}}$ is a sequence of i.i.d. random variables with common mean μ and finite variance σ^2, then*

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{\mathsf{d}} Z,$$

where Z is a standard Gaussian random variable. Explicitly, for all $x \in \mathbb{R}$,

$$\mathbb{P}\Big(\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \le x\Big) \xrightarrow[n\to\infty]{} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{z^2}{2}\Big)\,\mathrm{d}z.$$

The CLT plays a huge role in statistics, where it is used to provide asymptotic confidence intervals. Similarly, statisticians work towards proving convergence in distribution to other common distributions in statistics, such as the *chi-squared distribution* or the *t distribution*.

Another example of convergence in distribution is the **Poisson Law of Rare Events**, which is used as a justification for the use of the Poisson distribution in models of rare events.

Theorem 6 (Poisson Law of Rare Events)**.** *If $X_n \sim \mathrm{Binomial}(n, p_n)$ where $p_n \xrightarrow[n\to\infty]{} 0$ such that $np_n \xrightarrow[n\to\infty]{} \lambda > 0$, then $X_n \xrightarrow[n\to\infty]{\mathsf{d}} X$, where $X \sim \mathrm{Poisson}(\lambda)$.*

In fact, many other situations (especially concerning balls and bins) have Poisson limits, and Poisson limits are used in popular *random graph* models.

4 Miscellaneous Results

4.1 Continuous Mapping

One of the most useful results is presented below:

Theorem 7 (Continuous Mapping)**.** *Let f be a continuous function.*

1. *If $X_n \xrightarrow[n\to\infty]{a.s.} X$, then $f(X_n) \xrightarrow[n\to\infty]{a.s.} f(X)$.*
2. *If $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$, then $f(X_n) \xrightarrow[n\to\infty]{\mathbb{P}} f(X)$.*

3. *If* $X_n \xrightarrow[n\to\infty]{\mathsf{d}} X$*, then* $f(X_n) \xrightarrow[n\to\infty]{\mathsf{d}} f(X)$*.*

A typical application is to analyze a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ by applying a log or exp transformation, which is useful when showing the convergence of $(\log X_n)_{n\in\mathbb{N}}$ or $(\exp X_n)_{n\in\mathbb{N}}$ is easier than showing that the original sequence $(X_n)_{n\in\mathbb{N}}$ converges.

4.2 Convergence of Expectation

In general, none of the above modes of convergence imply that $\mathbb{E}[X_n] \xrightarrow[n\to\infty]{} \mathbb{E}[X]$. As an example, let $U \sim \text{Uniform}[0,1]$ and let $X_n := n\mathbb{1}\{U \le n^{-1}\}$. Then, $X_n \xrightarrow[n\to\infty]{\text{a.s.}} 0$, but $\mathbb{E}[X_n] = 1$ for all n. In more advanced treatments of probability theory, convergence of expected values is quite important, and there are a number of technical tools called *convergence theorems* used to justify convergence of expectations. Although we will not need them, we will state them here.

Theorem 8 (Convergence Theorems)**.** *Suppose* $X_n \xrightarrow[n\to\infty]{a.s.} X$*.*

1. *(Monotone Convergence) If* $0 \le X_1 \le X_2 \le X_3 \le \cdots$*, then* $\mathbb{E}[X_n] \xrightarrow[n\to\infty]{} \mathbb{E}[X]$*.*

2. *(Dominated Convergence) If there exists a random variable* $Y \ge 0$ *with* $\mathbb{E}[Y] < \infty$ *and* $|X_n|, |X| \le Y$ *for all* n*, then* $\mathbb{E}[X_n] \xrightarrow[n\to\infty]{} \mathbb{E}[X]$*.*