CONVERGENCE OF RANDOM VARIABLES

Contents

1. Definitions
2. Convergence in distribution
3. The hierarchy of convergence concepts

1 DEFINITIONS

1.1 Almost sure convergence

Definition 1. *We say that X_n converges to X **almost surely** (a.s.), and write $X_n \overset{\text{a.s.}}{\to} X$, if there is a (measurable) set $A \subset \Omega$ such that:*

(a) $\lim_{n \to \infty} X_n(\omega) = X(\omega)$, for all $\omega \in A$;

(b) $\mathbb{P}(A) = 1$.

Note that for a.s. convergence to be relevant, all random variables need to be defined on the same probability space (one experiment). Furthermore, the different random variables X_n are generally highly dependent.

Two common cases where a.s. convergence arises are the following.

(a) The probabilistic experiment runs over time. To each time n, we associate a nonnegative random variable Z_n (e.g., income on day n). Let $X_n = \sum_{k=1}^{n} Z_k$ be the income on the first n days. Let $X = \sum_{k=1}^{\infty} Z_k$ be the lifetime income. Note that X is well defined (as an extended real number) for every $\omega \in \Omega$, because of our assumption that $Z_k \geq 0$, and $X_n \overset{\text{a.s.}}{\to} X$.

(b) The various random variables are defined as different functions of a single underlying random variable. More precisely, suppose that Y is a random variable, and let $g_n : \mathbb{R} \to \mathbb{R}$ be measurable functions. Let $X_n = g_n(Y)$ [which really means, $X_n(\omega) = g_n(Y(\omega))$, for all ω]. Suppose that $\lim_{n\to\infty} g_n(y) = g(y)$ for every y. Then, $X_n \overset{\text{a.s.}}{\to} X$. For example, let $g_n(y) = y + y^2/n$, which converges to $g(y) = y$. We then have $Y + Y^2/n \overset{\text{a.s.}}{\to} Y$.

When $X_n \overset{\text{a.s.}}{\to} X$, we always have

$$\phi_{X_n}(t) \to \phi_X(t), \qquad \forall\, t,$$

by the dominated convergence theorem. On the other hand, the relation

$$\mathbb{E}[X_n] \to \mathbb{E}[X]$$

is not always true; sufficient conditions are provided by the monotone and dominated convergence theorems. For an example, where convergence of expectations fails to hold, consider a random variable U which is uniform on $[0, 1]$, and let:

$$X_n = \begin{cases} n, & \text{if } U \leq 1/n, \\ 0, & \text{if } U > 1/n. \end{cases} \tag{1}$$

We have

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \lim_{n\to\infty} n\mathbb{P}(U \leq 1/n) = 1.$$

On the other hand, for any outcome ω for which $U(\omega) > 0$ (which happens with probability one), $X_n(\omega)$ converges to zero. Thus, $X_n \overset{\text{a.s.}}{\to} 0$, but $\mathbb{E}[X_n]$ does not converge to zero.

1.2 Convergence in distribution

Definition 2. *Let X and X_n, $n \in \mathbb{N}$, be random variables with CDFs F and F_n, respectively. We say that the sequence X_n converges to X **in distribution**, and write $X_n \overset{\text{d}}{\to} X$, if*

$$\lim_{n\to\infty} F_n(x) = F(x),$$

for every $x \in \mathbb{R}$ at which F is continuous.

(a) Recall that CDFs have discontinuities ("jumps") only at the points that have positive probability mass. More precisely, F is continuous at x if and only if $\mathbb{P}(X = x) = 0$.

(b) Let $X_n = 1/n$, and $X = 0$, with probability 1. Note that $F_{X_n}(0) = \mathbb{P}(X_n \leq 0) = 0$, for every n, but $F_X(0) = 1$. Still, because of the exception in the above definition, we have $X_n \xrightarrow{d} X$. More generally, if $X_n = a_n$ and $X = a$, with probability 1, and $a_n \to a$, then $X_n \xrightarrow{d} X$. Thus, convergence in distribution is consistent with the definition of convergence of real numbers. This would not have been the case if the definition required the condition $\lim_{n\to\infty} F_n(x) = F(x)$ to hold at every x.

(c) Note that this definition just involves the marginal distributions of the random variables involved. These random variables may even be defined on different probability spaces.

(d) Let Y be a continuous random variable whose PDF is symmetric around 0. Let $X_n = (-1)^n Y$. Then, every X_n has the same distribution, so, trivially, X_n converges to Y in distribution. However, for almost all ω, the sequence $X_n(\omega)$ does not converge.

(e) If we are dealing with random variables whose distribution is in a parametric class, (e.g., if every X_n is exponential with parameter λ_n), and the parameters converge (e.g., if $\lambda_n \to \lambda > 0$ and X is exponential with parameter λ), then we usually have convergence of X_n to X, in distribution.

(f) It is possible for a sequence of discrete random variables to converge in distribution to a continuous one. For example, if Y_n is uniform on $\{1, \ldots, n\}$ and $X_n = Y_n/n$, then X_n converges in distribution to a random variable which is uniform on $[0, 1]$.

(g) Similarly, it is possible for a sequence of continuous random variables to converge in distribution to a discrete one. For example if X_n is uniform on $[0, 1/n]$, then X_n converges in distribution to a discrete random variable which is identically equal to zero.

(h) If X and all X_n are continuous, convergence in distribution does not imply convergence of the corresponding PDFs. (Find an example, by emulating the example in (f).)

(i) If X and all X_n are integer-valued, convergence in distribution turns out to be equivalent to convergence of the corresponding PMFs: $p_{X_n}(k) \to p_X(k)$, for all k.

1.3 Convergence in probability

> **Definition 3.** *(a) We say that a sequence of random variables X_n (not neces-sarily defined on the same probability space) converges **in probability** to a real number c, and write $X_n \overset{i.p.}{\to} c$, if*
>
> $$\lim_{n \to \infty} \mathbb{P}(|X_n - c| \geq \epsilon) = 0, \qquad \forall\, \epsilon > 0.$$
>
> *(b) Suppose that X and X_n, $n \in \mathbb{N}$ are all defined on the same probability space. We say that the sequence X_n converges to X, in probability, and write $X_n \overset{i.p.}{\to} X$, if $X_n - X$ converges to zero, in probability, i.e.,*
>
> $$\lim_{n \to \infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0, \qquad \forall\, \epsilon > 0.$$

(a) When X in part (b) of the definition is deterministic, say equal to some constant c, then the two parts of the above definition are consistent with each other.

(b) The intuitive content of the statement $X_n \overset{i.p.}{\to} c$ is that in the limit as n in-creases, almost all of the probability mass becomes concentrated in a small interval around c, no matter how small this interval is. On the other hand, for any fixed n, there can be a small probability mass outside this interval, with a slowly decaying tail. Such a tail can have a strong impact on expected values. For this reason, convergence in probability does not have any im-plications on expected values. See for instance the example in Eq. (1). We have $X_n \overset{i.p.}{\to} X$, but $\mathbb{E}[X_n]$ does not converge to $\mathbb{E}[X]$.

(c) If $X_n \overset{i.p.}{\to} X$ and $Y_n \overset{i.p.}{\to} Y$, and all random variables are defined on the same probability space, then $(X_n + Y_n) \overset{i.p.}{\to} (X + Y)$. (Can you prove it?)

2 CONVERGENCE IN DISTRIBUTION

The following result provides insights into the meaning of convergence in dis-tribution, by showing a close relation with almost sure convergence.

Theorem 1. *Suppose that $X_n \xrightarrow{d} X$. Then, there exists a probability space and random variables Y, Y_n defined on that space with the following properties:*

(a) For every n, the random variables X_n and Y_n have the same CDF; similarly, X and Y have the same CDF.

(b) $Y_n \xrightarrow{a.s.} Y$.

For convergence in distribution, it makes no difference whether the random variables X_n are independent or not; they do not even need to be defined on the same probability space. On the other hand, almost sure convergence implies a strong form of dependence between the random variables involved. The idea in the preceding theorem is to preserve the marginal distributions, but introduce a particular form of dependence between the X_n, which then results in almost sure convergence. This dependence is introduced by generating random variables Y_n and Y with the desired distributions, using a *common random number generator*, e.g., a single random variable U, uniformly distributed on $[0, 1]$.

If we assume that all CDFs involved are continuous and strictly increasing, then we can let $Y_n = F_{X_n}^{-1}(U)$, and $Y = F_X^{-1}(U)$. This guarantees the first condition in the theorem. It then follows that $Y_n \xrightarrow{a.s.} Y$, as discussed in item (b) in Section 1.1. For the case of more general CDFs, the argument is similar in spirit but takes more work; see [GS], p. 315.

3 THE HIERARCHY OF CONVERGENCE CONCEPTS

Theorem 2. *We have*

$$[X_n \xrightarrow{a.s.} X] \implies [X_n \xrightarrow{i.p.} X] \implies [X_n \xrightarrow{d} X] \implies [\phi_{X_n}(t) \rightarrow \phi_X(t), \ \forall \, t].$$

(The first two implications assume that all random variables be defined on the same probability space.)

Proof:

(a) $[X_n \xrightarrow{a.s.} X] \implies [X_n \xrightarrow{i.p.} X]$:

We give a short proof, based on the DCT, but more elementary proofs are also possible. Fix some $\epsilon > 0$. Let

$$Y_n = \epsilon I_{\{|X_n - X| \geq \epsilon\}}.$$

5

If $X_n \overset{\text{a.s.}}{\to} X$, then $Y_n \overset{\text{a.s.}}{\to} 0$. By the DCT, $\mathbb{E}[Y_n] \to 0$. On the other hand,

$$\mathbb{E}[Y_n] = \epsilon \mathbb{P}\big(X_n - X| \geq \epsilon\big).$$

This implies that $\mathbb{P}\big(|X_n - X| \geq \epsilon\big) \to 0$, and therefore, $X_n \overset{\text{i.p.}}{\to} X$.

(b) $[X_n \overset{\text{i.p.}}{\to} X] \Rightarrow [X_n \overset{\text{d}}{\to} X]$:
The proof is omitted; see, e.g., [GS].

(c) $[X_n \overset{\text{d}}{\to} X] \Rightarrow [\phi_{X_n}(t) \to \phi_X(t), \; \forall \, t]$:
Suppose that $X_n \overset{\text{d}}{\to} X$. Let Y_n and Y be as in Theorem 1, so that $Y_n \overset{\text{a.s.}}{\to} Y$. Then, for any $t \in \mathbb{R}$,

$$\lim_{n \to \infty} \phi_{X_n}(t) = \lim_{n \to \infty} \phi_{Y_n}(t) = \lim_{n \to \infty} \mathbb{E}[e^{itY_n}] = \mathbb{E}[e^{itY}] = \phi_Y(t) = \phi_X(t),$$

where we have made use of the facts $Y_n \overset{\text{a.s.}}{\to} Y$, $e^{itY_n} \overset{\text{a.s.}}{\to} e^{itY}$, and the DCT. $\qquad \square$

At this point, it is natural to ask whether the converses of the implications in Theorem 2 hold. For the first two, the answer is, in general, "no", although we will also note some exceptions.

3.1 Convergence almost surely versus in probability

$[X_n \overset{\text{i.p.}}{\to} X]$ **does not imply** $[X_n \overset{\text{a.s.}}{\to} X]$:
Let X_n be equal to 1, with probability $1/n$, and equal to zero otherwise. Suppose that the X_n are independent. We have $X_n \overset{\text{i.p.}}{\to} 0$. On the other hand, by the Borel-Cantelli lemma, the event $\{X_n = 1, \; \text{i.o}\}$ has probability 1. Thus, for almost all ω, the sequence $X_n(\omega)$ does not converge to zero.

Nevertheless, a weaker form of the converse implication turns out to be true. If $X_n \overset{\text{i.p.}}{\to} X$, then there exists an increasing (deterministic) sequence n_k of integers, such that $\lim_{k \to \infty} X_{n_k} = X$, a.s. (We omit the proof.)

For an illustration of the last statement in action, consider the preceding counterexample. If we let $n_k = k^2$, then we note that $\mathbb{P}(X_{n_k} \neq 0) = 1/k^2$, which is summable. By the Borel-Cantelli lemma, the event $\{X_{n_k} \neq 0\}$ will occur for only finitely many k, with probability 1. Therefore, X_{n_k} converges, a.s., to the zero random variable.

3.2 Convergence in probability versus in distribution

The converse turns out to be false in general, but true when the limit is deterministic.

$[X_n \xrightarrow{\text{d}} X]$ **does not imply** $[X_n \xrightarrow{\text{i.p.}} X]$:

Let the random variables X, X_n be i.i.d. and nonconstant random variables, in which case we have (trivially) $X_n \xrightarrow{\text{d}} X$. Fix some ϵ. Then, $\mathbb{P}(|X_n - X| \geq \epsilon)$ is positive and the same for all n, which shows that X_n does not converge to X, in probability.

$[X_n \xrightarrow{\text{d}} c]$ **implies** $[X_n \xrightarrow{\text{i.p.}} c]$:

The proof is omitted; see, e.g., [GS].

3.3 Convergence in distribution versus characteristic functions

Finally, the converse of the last implication in Theorem 2 is always true. We know that equality of two characteristic functions implies equality of the corresponding distributions. It is then plausible to hope that "near-equality" of characteristic functions implies "near equality" of corresponding distributions. This would be essentially a statement that the mapping from characteristic functions to distributions is a continuous one. Even though such a result is plausible, its proof is beyond our scope.

Theorem 3. Continuity of inverse transforms: *Let X and X_n be random variables with given CDFs and corresponding characteristic functions. We have*

$$[\phi_{X_n}(t) \to \phi_X(t), \ \forall \, t] \quad \Rightarrow \quad [X_n \xrightarrow{\text{d}} X].$$

The preceding theorem involves two separate conditions: (i) the sequence of characteristic functions ϕ_{X_n} converges (pointwise), and (ii) the limit is the characteristic function associated with some other random variable. If we are only given the first condition (pointwise convergence), how can we tell if the limit is indeed a legitimate characteristic function associated with some random variable? One way is to check for various properties that every legitimate characteristic function must possess. One such property is continuity: if $t \to t^*$, then (using dominated convergence),

$$\lim_{t \to t^*} \phi_X(t) = \lim_{t \to t^*} \mathbb{E}[e^{itX}] = \mathbb{E}[e^{it^* X}] = \phi_X(t^*).$$

It turns out that continuity at zero is all that needs to be checked.

> **Theorem 4. Continuity of inverse transforms:** *Let X_n be random variables with characteristic functions ϕ_{X_n}, and suppose that the limit $\phi(t) = \lim_{n \to \infty} \phi_{X_n}(t)$ exists for every t. Then, either*
>
> *(i) The function ϕ is discontinuous at zero (in this case X_n does not converge in distribution); or*
>
> *(ii) There exists a random variable X whose characteristic function is ϕ, and $X_n \overset{\mathrm{d}}{\to} X$.*

To illustrate the two possibilities in Theorem 4, consider a sequence $\{X_n\}$, and assume that X_n is exponential with parameter λ_n, so that $\phi_{X_n}(t) = \lambda_n/(\lambda_n - it)$.

(a) Suppose that λ_n converges to a positive number λ. Then, the sequence of characteristic functions ϕ_{X_n} converges to the function ϕ defined by $\phi(t) = \lambda/(\lambda - it)$. We recognize this as the characteristic function of an exponential distribution with parameter λ. In particular, we conclude that X_n converges in distribution to an exponential random variable with parameter λ.

(b) Suppose now that λ_n converges to zero. Then,

$$\lim_{n \to \infty} \phi_{X_n}(t) = \lim_{n \to \infty} \frac{\lambda_n}{\lambda_n - it} = \lim_{\lambda \downarrow 0} \frac{\lambda}{\lambda - it} = \begin{cases} 1, & \text{if } t = 0, \\ 0, & \text{if } t \neq 0. \end{cases}$$

Thus, the limit of the characteristic functions is discontinuous at $t = 0$, and X_n does not converge in distribution. Intuitively, this is because the distribution of X_n keeps spreading in a manner that does not yield a limiting distribution.

6.436J / 15.085J Fundamentals of Probability
Fall 2008