2.6: Convergence

This is the first of several sections in this chapter that are more advanced than the basic topics in the first five sections. In this section we discuss several topics related to convergence of events and random variables, a subject of fundamental importance in probability theory. In particular the results that we obtain will be important for:

- Properties of distribution functions,
- The weak law of large numbers,
- The strong law of large numbers.

As usual, our starting point is a random experiment modeled by a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. So to review, Ω is the set of outcomes, $\mathscr{F}$ the σ-algebra of events, and $\mathbb{P}$ the probability measure on the sample space $(\Omega, \mathscr{F})$.

Basic Theory

Sequences of events

Our first discussion deals with sequences of events and various types of limits of such sequences. The limits are also event. We start with two simple definitions.

> Suppose that $(A_1, A_2, \ldots)$ is a sequence of events.
>
> 1. The sequence is *increasing* if $A_n \subseteq A_{n+1}$ for every $n \in \mathbb{N}_+$.
> 2. The sequence is *decreasing* if $A_{n+1} \subseteq A_n$ for every $n \in \mathbb{N}_+$.

Note that these are the standard definitions of increasing and decreasing, relative to the ordinary total order $\le$ on the index set $\mathbb{N}_+$ and the subset partial order $\subseteq$ on the collection of events. The terminology is also justified by the corresponding indicator variables.

> Suppose that $(A_1, A_2, \ldots)$ is a sequence of events, and let $I_n = \mathbf{1}_{A_n}$ denote the indicator variable of the event A_n for $n \in \mathbb{N}_+$.
>
> 1. The sequence of events is increasing if and only if the sequence of indicator variables is increasing in the ordinary sense. That is, $I_n \le I_{n+1}$ for each $n \in \mathbb{N}_+$.
> 2. The sequence of events is decreasing if and only if the sequence of indicator variables is decreasing in the ordinary sense. That is, $I_{n+1} \le I_n$ for each $n \in {}_+$.
>
> Proof

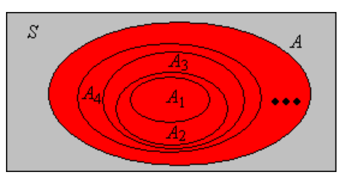

Figure 2.6.1: A sequence of increasing events and their union

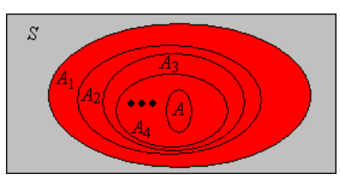

Figure 2.6.2: A sequence of decreasing events and their intersection

If a sequence of events is either increasing or decreasing, we can define the *limit* of the sequence in a way that turns out to be quite natural.

> Suppose that $(A_1, A_2, \ldots)$ is a sequence of events.
>
> 1. If the sequence is increasing, we define $\lim_{n\to\infty} A_n = \bigcup_{n=1}^\infty A_n$.
> 2. If the sequence is decreasing, we define $\lim_{n\to\infty} A_n = \bigcap_{n=1}^\infty A_n$.

Once again, the terminology is clarified by the corresponding indicator variables.

Suppose again that $(A_1, A_2, \ldots)$ is a sequence of events, and let $I_n = \mathbf{1}_{A_n}$ denote the indicator variable of A_n for $n \in \mathbb{N}_+$.

1. If the sequence of events is increasing, then $\lim_{n\to\infty} I_n$ is the indicator variable of $\bigcup_{n=1}^\infty A_n$
2. If the sequence of events is decreasing, then $\lim_{n\to\infty} I_n$ is the indicator variable of $\bigcap_{n=1}^\infty A_n$

Proof

1. If $s \in \bigcup_{n=1}^\infty A_n$ then $s \in A_k$ for some $k \in \mathbb{N}_+$. Since the events are increasing, $s \in A_n$ for every $n \ge k$. In this case, $I_n(s) = 1$ for every $n \ge k$ and hence $\lim_{n\to\infty} I_n(s) = 1$. On the other hand, if $s \notin \bigcup_{n=1}^\infty A_n$ then $s \notin A_n$ for every $n \in \mathbb{N}_+$. In this case, $I_n(s) = 0$ for every $n \in \mathbb{N}_+$ and hence $\lim_{n\to\infty} I_n(s) = 0$.
2. If $s \in \bigcap_{n=1}^\infty A_n$ then $s \in A_n$ for each $n \in \mathbb{N}_+$. In this case, $I_n(s) = 1$ for each $n \in \mathbb{N}_+$ and hence $\lim_{n\to\infty} I_n(s) = 1$. If $s \notin \bigcap_{n=1}^\infty A_n$ then $s \notin A_k$ for some $k \in \mathbb{N}_+$. Since the events are decreasing, $s \notin A_n$ for all $n \ge k$. In this case, $I_n(s) = 0$ for $n \ge k$ and hence $\lim_{n\to\infty} I_n(s) = 0$.

An arbitrary union of events can always be written as a union of increasing events, and an arbitrary intersection of events can always be written as an intersection of decreasing events:

Suppose that $(A_1, A_2, \ldots)$ is a sequence of events. Then

1. $\bigcup_{i=1}^n A_i$ is increasing in $n \in \mathbb{N}_+$ and $\bigcup_{i=1}^\infty A_i = \lim_{n\to\infty} \bigcup_{i=1}^n A_i$.
2. $\bigcap_{i=1}^n A_i$ is decreasing in $n \in \mathbb{N}_+$ and $\bigcap_{i=1}^\infty A_i = \lim_{n\to\infty} \bigcap_{i=1}^n A_i$.

Proof

1. Trivially $\bigcup_{i=1}^n A_i \subseteq \bigcup_{i=1}^{n+1} A_i$. The second statement simply means that $\bigcup_{n=1}^\infty \bigcup_{i=1}^n A_i = \bigcup_{i=1}^\infty A_i$.
2. Trivially $\bigcap_{i=1}^{n+1} A_i \subseteq \bigcap_{i=1}^n A_i$. The second statement simply means that $\bigcap_{n=1}^\infty \bigcap_{i=1}^n A_i = \bigcap_{i=1}^\infty A_i$.

There is a more interesting and useful way to generate increasing and decreasing sequences from an arbitrary sequence of events, using the *tail* segment of the sequence rather than the *initial* segment.

Suppose that $(A_1, A_2, \ldots)$ is a sequence of events. Then

1. $\bigcup_{i=n}^\infty A_i$ is decreasing in $n \in \mathbb{N}_+$.
2. $\bigcap_{i=n}^\infty A_i$ is increasing in $n \in \mathbb{N}_+$.

Proof

1. Clearly $\bigcup_{i=n+1}^\infty A_i \subseteq \bigcup_{i=n}^\infty A_i$
2. Clearly $\bigcap_{i=n}^\infty A_i \subseteq \bigcap_{i=n+1}^\infty A_i$

Since the new sequences defined in the previous results are decreasing and increasing, respectively, we can take their limits. These are the *limit superior* and *limit inferior*, respectively, of the original sequence.

Suppose that $(A_1, A_2, \ldots)$ is a sequence of events. Define

1. $\limsup_{n\to\infty} A_n = \lim_{n\to\infty} \bigcup_{i=n}^\infty A_i = \bigcap_{n=1}^\infty \bigcup_{i=n}^\infty A_i$. This is the event that occurs if an only if A_n occurs for infinitely many values of n.
2. $\liminf_{n\to\infty} A_n = \lim_{n\to\infty} \bigcap_{i=n}^\infty A_i = \bigcup_{n=1}^\infty \bigcap_{i=n}^\infty A_i$. This is the event that occurs if an only if A_n occurs for all but finitely many values of n.

Proof

1. From the definition, the event $\limsup_{n\to\infty} A_n$ occurs if and only if for each $n \in \mathbb{N}_+$ there exists $i \ge n$ such that A_i occurs.
2. From the definition, the event $\liminf_{n\to\infty} A_n$ occurs if and only if there exists $n \in \mathbb{N}_+$ such that A_i occurs for every $i \ge n$.

Once again, the terminology and notation are clarified by the corresponding indicator variables. You may need to review limit inferior and limit superior for sequences of real numbers in the section on Partial Orders.

Suppose that $(A_1, A_2, \ldots)$ is a sequence of events, and et $I_n = \mathbf{1}_{A_n}$ denote the indicator variable of A_n for $n \in \mathbb{N}_+$. Then

1. $\limsup_{n\to\infty} I_n$ is the indicator variable of $\limsup_{n\to\infty} A_n$.
2. $\liminf_{n\to\infty} I_n$ is the indicator variable of $\liminf_{n\to\infty} A_n$.

Proof

1. By the result above, $\lim_{n\to\infty} \mathbf{1}\left(\bigcup_{i=n}^\infty A_i\right)$ is the indicator variable of $\limsup_{n\to\infty} A_n$. But $\mathbf{1}\left(\bigcup_{i=n}^\infty A_i\right) = \max\{I_i : i \ge n\}$ and hence $\lim_{n\to\infty} \mathbf{1}\left(\bigcup_{i=n}^\infty A_i\right) = \limsup_{n\to\infty} I_n$.
2. By the result above, $\lim_{n\to\infty} \mathbf{1}\left(\bigcap_{i=n}^\infty A_i\right)$ is the indicator variable of $\liminf_{n\to\infty} A_n$. But $\mathbf{1}\left(\bigcap_{i=n}^\infty A_i\right) = \min\{I_i : i \ge n\}$ and hence $\lim_{n\to\infty} \mathbf{1}\left(\bigcap_{i=n}^\infty A_i\right) = \liminf_{n\to\infty} I_n$.

Suppose that $(A_1, A_2, \ldots)$ is a sequence of events. Then $\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$.

Proof

If A_n occurs for all but finitely many $n \in \mathbb{N}_+$ then certainly A_n occurs for infinitely many $n \in \mathbb{N}_+$.

Suppose that $(A_1, A_2, \ldots)$ is a sequence of events. Then

1. $\left(\limsup_{n\to\infty} A_n\right)^c = \liminf_{n\to\infty} A_n^c$
2. $\left(\liminf_{n\to\infty} A_n\right)^c = \limsup_{n\to\infty} A_n^c$.

Proof

These results follows from DeMorgan's laws.

The Continuity Theorems

Generally speaking, a function is *continuous* if it preserves limits. Thus, the following results are the *continuity theorems of probability*. Part (a) is the continuity theorem for increasing events and part (b) the continuity theorem for decreasing events.

Suppose that $(A_1, A_2, \ldots)$ is a sequence of events.

1. If the sequence is increasing then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^\infty A_n\right)$
2. If the sequence is decreasing then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \mathbb{P}\left(\bigcap_{n=1}^\infty A_n\right)$

Proof

1. Let $B_1 = A_1$ and let $B_i = A_i \setminus A_{i-1}$ for $i \in \{2, 3, \ldots\}$. Note that the collection of events $\{B_1, B_2, \ldots\}$ is pairwise disjoint and has the same union as $\{A_1, A_2, \ldots\}$. From countable additivity and the definition of infinite series,
$$\mathbb{P}\left(\bigcup_{i=1}^\infty A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^\infty B_i\right) = \sum_{i=1}^\infty \mathbb{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^n \mathbb{P}(B_i) \tag{2.6.1}$$
But $\mathbb{P}(B_1) = \mathbb{P}(A_1)$ and $\mathbb{P}(B_i) = \mathbb{P}(A_i) - \mathbb{P}(A_{i-1})$ for $i \in \{2, 3, \ldots\}$. Therefore $\sum_{i=1}^n \mathbb{P}(B_i) = \mathbb{P}(A_n)$ and hence we have $\mathbb{P}\left(\bigcup_{i=1}^\infty A_i\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$.
The construction in the continuity theorem for increasing events
The construction in the continuity theorem
2. The sequence of complements $(A_1^c, A_2^c, \ldots)$ is increasing. Hence using part (a), DeMorgan's law, and the complement rule we have
$$\mathbb{P}\left(\bigcap_{i=1}^\infty A_i\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^\infty A_i^c\right) = 1 - \lim_{n\to\infty} \mathbb{P}(A_n^c) = \lim_{n\to\infty} \left[1 - \mathbb{P}(A_n^c)\right] = \lim_{n\to\infty} \mathbb{P}(A_n) \tag{2.6.2}$$

The continuity theorems can be applied to the increasing and decreasing sequences that we constructed earlier from an arbitrary sequence of events.

Suppose that $(A_1, A_2, \ldots)$ is a sequence of events.

1. $\mathbb{P}\left(\bigcup_{i=1}^\infty A_i\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right)$
2. $\mathbb{P}\left(\bigcap_{i=1}^\infty A_i\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=1}^n A_i\right)$

Proof

These results follow immediately from the continuity theorems.

Suppose that $(A_1, A_2, \ldots)$ is a sequence of events. Then

1. $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=n}^\infty A_i\right)$
2. $\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=n}^\infty A_i\right)$

Proof

These results follows directly from the definitions, and the continuity theorems.

The next result shows that the countable additivity axiom for a probability measure is equivalent to *finite additivity* and the continuity property for increasing events.

Temporarily, suppose that $\mathbb{P}$ is only finitely additive, but satisfies the continuity property for increasing events. Then $\mathbb{P}$ is countably additive.

Proof

Suppose that $(A_1, A_2, \ldots)$ is a sequence of pairwise disjoint events. Since we are assuming that $\mathbb{P}$ is finitely additive we have

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) \tag{2.6.3}$$

If we let $n \to \infty$, the left side converges to $\mathbb{P}\left(\bigcup_{i=1}^\infty A_i\right)$ by the continuity assumption and the result above, while the right side converges to $\sum_{i=1}^\infty \mathbb{P}(A_i)$ by the definition of an infinite series.

There are a few mathematicians who reject the countable additivity axiom of probability measure in favor of the weaker finite additivity axiom. Whatever the philosophical arguments may be, life is certainly much harder without the continuity theorems.

The Borel-Cantelli Lemmas

The *Borel-Cantelli Lemmas*, named after Emil Borel and Francessco Cantelli, are very important tools in probability theory. The first lemma gives a condition that is sufficient to conclude that infinitely many events occur with probability 0.

First Borel-Cantelli Lemma. Suppose that $(A_1, A_2, \ldots)$ is a sequence of events. If $\sum_{n=1}^\infty \mathbb{P}(A_n) < \infty$ then $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0$.

Proof

From the result above on limit superiors, we have $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=n}^\infty A_i\right)$. But from Boole's inequality, $\mathbb{P}\left(\bigcup_{i=n}^\infty A_i\right) \le \sum_{i=n}^\infty \mathbb{P}(A_i)$. Since $\sum_{i=1}^\infty \mathbb{P}(A_i) < \infty$, we have $\sum_{i=n}^\infty \mathbb{P}(A_i) \to 0$ as $n \to \infty$.

The second lemma gives a condition that is sufficient to conclude that infinitely many *independent* events occur with probability 1.

Second Borel-Cantelli Lemma. Suppose that $(A_1, A_2, \ldots)$ is a sequence of independent events. If $\sum_{n=1}^\infty \mathbb{P}(A_n) = \infty$ then $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1$.

Proof

Note first that $1 - x \le e^{-x}$ for every $x \in \mathbb{R}$, and hcnce $1 - \mathbb{P}(A_i) \le \exp[-\mathbb{P}(A_i)]$ for each $i \in \mathbb{N}_+$. From the results above on limit superiors and complements,

$$\mathbb{P}\left[\left(\limsup_{n\to\infty} A_n\right)^c\right] = \mathbb{P}\left(\liminf_{n\to\infty} A_n^c\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=n}^\infty A_i^c\right) \tag{2.6.4}$$

But by independence and the inequality above,

$$\mathbb{P}\left(\bigcap_{i=n}^\infty A_i^c\right) = \prod_{i=n}^\infty \mathbb{P}\left(A_i^c\right) = \prod_{i=n}^\infty [1 - \mathbb{P}(A_i)] \le \prod_{i=n}^\infty \exp[-\mathbb{P}(A_i)] = \exp\left(-\sum_{i=n}^\infty \mathbb{P}(A_i)\right) = 0 \tag{2.6.5}$$

For independent events, both Borel-Cantelli lemmas apply of course, and lead to a *zero-one law*.

If $(A_1, A_2, \ldots)$ is a sequence of independent events then $\limsup_{n\to\infty} A_n$ has probability 0 or 1:

1. If $\sum_{n=1}^\infty \mathbb{P}(A_n) < \infty$ then $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0$.
2. If $\sum_{n=1}^\infty \mathbb{P}(A_n) = \infty$ then $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1$.

This result is actually a special case of a more general zero-one law, known as the *Kolmogorov zero-one law*, and named for Andrei Kolmogorov. This law is studied in the more advanced section on measure. Also, we can use the zero-one law to derive a calculus theorem that relates infinite series and infinte products. This derivation is an example of the *probabilistic method*—the use of probability to obtain results, seemingly unrelated to probability, in other areas of mathematics.

Suppose that $p_i \in (0, 1)$ for each $i \in \mathbb{N}_+$. Then

$$\prod_{i=1}^\infty p_i > 0 \text{ if and only if } \sum_{i=1}^\infty (1 - p_i) < \infty \tag{2.6.6}$$

Proof

We can easily construct a probability space with a sequence of independent events $(A_1, A_2, \ldots)$ such that $\mathbb{P}(A_i) = 1 - p_i$ for each $i \in \mathbb{N}_+$. The result then follows from the proofs of the two Borel-Cantelli lemmas.

Our next result is a simple application of the second Borel-Cantelli lemma to independent replications of a basic experiment.

Suppose that A is an event in a basic random experiment with $\mathbb{P}(A) > 0$. In the compound experiment that consists of independent replications of the basic experiment, the event "A occurs infinitely often" has probability 1.

Proof

Let p denote the probability of A in the basic experiment. In the compound experiment, we have a sequence of independent events $(A_1, A_2, \ldots)$ with $\mathbb{P}(A_n) = p$ for each $n \in \mathbb{N}_+$ (these are "independent copies" of A). But $\sum_{n=1}^\infty \mathbb{P}(A_n) = \infty$ since $p > 0$ so the result follows from the second Borel-Cantelli lemma.

Convergence of Random Variables

Our next discussion concerns two ways that a sequence of random variables defined for our experiment can "converge". These are fundamentally important concepts, since some of the deepest results in probability theory are limit theorems involving random variables. The most important special case is when the random variables are real valued, but the proofs are essentially the same for variables with values in a metric space, so we will use the more general setting.

Thus, suppose that (S, d) is a metric space, and that $\mathscr{S}$ is the corresponding Borel σ-algebra (that is, the σ-algebra generated by the topology), so that our measurable space is $(S, \mathscr{S})$. Here is the most important special case:

For $n \in \mathbb{N}_+$, is the n-dimensional *Euclidean space* is $(\mathbb{R}^n, d_n)$ where

$$d_n(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}, \quad \boldsymbol{x} = (x_1, x_2 \ldots, x_n),\ \boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n \tag{2.6.7}$$

Euclidean spaces are named for Euclid, of course. As noted above, the one-dimensional case where $d(x, y) = |y - x|$ for $x,\ y \in \mathbb{R}$ is particularly important. Returning to the general metric space, recall that if $(x_1, x_2, \ldots)$ is a sequence in S and $x \in S$, then $x_n \to x$ as $n \to \infty$ means that $d(x_n, x) \to 0$ as $n \to \infty$ (in the usual calculus sense). For the rest of our discussion, we assume that $(X_1, X_2, \ldots)$ is a sequence of random variable with values in S and X is a random variable with values in S, all defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

We say that $X_n \to X$ as $n \to \infty$ *with probability 1* if the *event* that $X_n \to X$ as $n \to \infty$ has probability 1. That is,

$$\mathbb{P}\{\omega \in S : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\} = 1 \tag{2.6.8}$$

Details

We need to make sure that the definition makes sense, in that the statement that X_n converges to X as $n \to \infty$ defines a valid event. Note that X_n does *not* converge to X as $n \to \infty$ if and only if for some $\epsilon > 0$, $d(X_n, X) > \epsilon$ for infinitely many $n \in \mathbb{N}_+$. Note that if the this condition holds for a given $\epsilon > 0$ then it holds for all smaller $\epsilon > 0$. Moreover, there are arbitrarily small rational $\epsilon > 0$ so X_n does not converge to X as $n \to \infty$ if and only if for some rational $\epsilon > 0$, $d(X_n, X) > \epsilon$ for infinitely many $n \in \mathbb{N}_+$. Hence

$$\{X_n \to X \text{ as } n \to \infty\}^c = \bigcup_{\epsilon \in \mathbb{Q}_+} \limsup_{n \to \infty} \{d(X_n, X) > \epsilon\} \tag{2.6.9}$$

where $\mathbb{Q}_+$ is the set of positive rational numbers. A critical point to remember is that this set is countable. So, building a little at a time, note that $\{d(X_n, X) > \epsilon\}$ is an event for each $\epsilon \in \mathbb{Q}_+$ and $n \in \mathbb{N}_+$ since X_n and X are random variables. Next, the limit superior of a sequence of events is an event. Finally, a countable union of events is an event.

As good probabilists, we usually suppress references to the sample space and write the definition simply as $\mathbb{P}(X_n \to X \text{ as } n \to \infty) = 1$. The statement that an event has probability 1 is usually the strongest affirmative statement that we can make in probability theory. Thus, convergence with probability 1 is the strongest form of convergence. The phrases *almost surely* and *almost everywhere* are sometimes used instead of the phrase *with probability 1*.

Recall that metrics d and e on S are *equivalent* if they generate the same topology on S. Recall also that convergence of a sequence is a *topological property*. That is, if $(x_1, x_2, \ldots)$ is a sequence in S and $x \in S$, and if d, e are equivalent metrics on S, then $x_n \to x$ as $n \to \infty$ relative to d if and only if $x_n \to x$ as $n \to \infty$ relative to e. So for our random variables as defined above, it follows that $X_n \to X$ as $n \to \infty$ with probability 1 relative to d if and only if $X_n \to X$ as $n \to \infty$ with probability 1 relative to e.

The following statements are equivalent:

1. $X_n \to X$ as $n \to \infty$ with probability 1.
2. $\mathbb{P}[d(X_n, X) > \epsilon \text{ for infinitely many } n \in \mathbb{N}_+] = 0$ for every rational $\epsilon > 0$.
3. $\mathbb{P}[d(X_n, X) > \epsilon \text{ for infinitely many } n \in \mathbb{N}_+] = 0$ for every $\epsilon > 0$.
4. $\mathbb{P}[d(X_k, X) > \epsilon \text{ for some } k \ge n] \to 0$ as $n \to \infty$ for every $\epsilon > 0$.

Proof

From the details in the definition above, $\mathbb{P}(X_n \to X \text{ as } n \to \infty) = 1$ if and only if

$$\mathbb{P}\left(\bigcup_{\epsilon \in \mathbb{Q}_+} \{d(X_n, X) > \epsilon \text{ for infinitely many } n \in \mathbb{N}_+\}\right) = 0 \tag{2.6.10}$$

where again $\mathbb{Q}_+$ is the set of positive rational numbers. But by Boole's inequality, a countable union of events has probability 0 if and only if every event in the union has probability 0. Thus, (a) is equivalent to (b). Statement (b) is clearly equivalent to (c) since there are arbitrarily small positive rational numbers. Finally, (c) is equivalent to (d) by the continuity result in above.

Our next result gives a fundamental criterion for convergence with probability 1:

If $\sum_{n=1}^\infty \mathbb{P}[d(X_n, X) > \epsilon] < \infty$ for every $\epsilon > 0$ then $X_n \to X$ as $n \to \infty$ with probability 1.

Proof

By the first Borel-Cantelli lemma, if $\sum_{n=1}^\infty \mathbb{P}[d(X_n, X) > \epsilon) < \infty$ then $\mathbb{P}[d(X_n, X) > \epsilon \text{ for infinitely many } n \in \mathbb{N}_+) = 0$. Hence the result follows from the previous theorem.

Here is our next mode of convergence.

We say that $X_n \to X$ as $n \to \infty$ *in probability* if

$$\mathbb{P}[d(X_n, X) > \epsilon] \to 0 \text{ as } n \to \infty \text{ for each } \epsilon > 0 \tag{2.6.11}$$

The phrase *in probability* sounds superficially like the phrase *with probability 1*. However, as we will soon see, convergence in probability is much weaker than convergence with probability 1. Indeed, convergence with probability 1 is often called *strong* *convergence*, while convergence in probability is often called *weak convergence*.

If $X_n \to X$ as $n \to \infty$ with probability 1 then $X_n \to X$ as $n \to \infty$ in probability.

Proof

Let $\epsilon > 0$. Then $\mathbb{P}[d(X_n, X) > \epsilon] \le \mathbb{P}[d(X_k, X) > \epsilon \text{ for some } k \ge n]$. But if $X_n \to X$ as $n \to \infty$ with probability 1, then the expression on the right converges to 0 as $n \to \infty$ by part (d) of the result above. Hence $X_n \to X$ as $n \to \infty$ in probability.

The converse fails with a passion. A simple counterexample is given below. However, there is a partial converse that is very useful.

If $X_n \to X$ as $n \to \infty$ in probability, then there exists a subsequence $(n_1, n_2, n_3 \ldots)$ of $\mathbb{N}_+$ such that $X_{n_k} \to X$ as $k \to \infty$ with probability 1.

Proof

Suppose that $X_n \to X$ as $n \to \infty$ in probability. Then for each $k \in \mathbb{N}_+$ there exists $n_k \in \mathbb{N}_+$ such that $\mathbb{P}[d(X_{n_k}, X) > 1/k] < 1/k^2$. We can make the choices so that $n_k < n_{k+1}$ for each k. It follows that $\sum_{k=1}^\infty \mathbb{P}[d(X_{n_k}, X) > \epsilon] < \infty$ for every $\epsilon > 0$. By the result above, $X_{n_k} \to X$ as $n \to \infty$ with probability 1.

Note that the proof works because $1/k \to 0$ as $k \to \infty$ and $\sum_{k=1}^\infty 1/k^2 < \infty$. Any two sequences with these properties would work just as well.

There are two other modes of convergence that we will discuss later:

- Convergence in distribution.
- Convergence in mean,

Examples and Applications

Coins

Suppose that we have an infinite sequence of coins labeled $1, 2, \ldots$ Moreover, coin n has probability of heads $1/n^a$ for each $n \in \mathbb{N}_+$, where $a > 0$ is a parameter. We toss each coin in sequence one time. In terms of a, find the probability of the following events:

1. infinitely many heads occur
2. infinitely many tails occur

Answer

Let H_n be the event that toss n results in heads, and T_n the event that toss n results in tails.

1. $\mathbb{P}(\limsup_{n\to\infty} H_n) = 1$, $\mathbb{P}(\limsup_{n\to\infty} T_n) = 1$ if $a \in (0, 1]$
2. $\mathbb{P}(\limsup_{n\to\infty} H_n) = 0$, $\mathbb{P}(\limsup_{n\to\infty} T_n) = 1$ if $a \in (1, \infty)$

The following exercise gives a simple example of a sequence of random variables that converge in probability but not with probability 1. Naturally, we are assuming the standard metric on $\mathbb{R}$.

Suppose again that we have a sequence of coins labeled $1, 2, \ldots$, and that coin n lands heads up with probability $\frac{1}{n}$ for each n. We toss the coins in order to produce a sequence $(X_1, X_2, \ldots)$ of independent indicator random variables with

$$\mathbb{P}(X_n = 1) = \frac{1}{n}, \ \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}; \quad n \in \mathbb{N}_+ \tag{2.6.12}$$

1. $\mathbb{P}(X_n = 0 \text{ for infinitely many } n) = 1$, so that infinitely many tails occur with probability 1.
2. $\mathbb{P}(X_n = 1 \text{ for infinitely many } n) = 1$, so that infinitely many heads occur with probability 1.
3. $\mathbb{P}(X_n \text{ does not converge as } n \to \infty) = 1$.
4. $X_n \to 0$ as $n \to \infty$ in probability.

Proof

1. This follow from the second Borel-Cantelli lemma, since $\sum_{n=1}^\infty \mathbb{P}(X_n = 0) = \infty$

2. This also follows from the second Borel-Cantelli lemma, since $\sum_{n=1}^{\infty} \mathbb{P}(X_n = 1) = \infty$.
3. This follows from parts (a) and (b). Recall that the intersection of two events with probability 1 still has probability 1.
4. Suppose $0 < \epsilon < 1$. Then $\mathbb{P}(|X_n - 0| > \epsilon) = \mathbb{P}(X_n = 1) = \frac{1}{n} \to 0 \quad \text{as } n \to \infty$.

Discrete Spaces

Recall that a measurable space $(S, \mathscr{S})$ is *discrete* if S is countable and $\mathscr{S}$ is the collection of all subsets of S (the power set of S). Moreover, $\mathscr{S}$ is the Borel σ-algebra corresponding to the *discrete metric* d on S given by $d(x, x) = 0$ for $x \in S$ and $d(x, y) = 1$ for distinct $x, y \in S$. How do convergence with probability 1 and convergence in probability work for the discrete metric?

Suppose that $(S, \mathscr{S})$ is a discrete space. Suppose further that $(X_1, X_2, \ldots)$ is a sequence of random variables with values in S and X is a random variable with values in S, all defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Relative to the discrete metric d,

1. $X_n \to X$ as $n \to \infty$ with probability 1 if and only if $\mathbb{P}(X_n = X \text{ for all but finitely many } n \in \mathbb{N}_+) = 1$.
2. $X_n \to X$ as $n \to \infty$ in probability if and only if $\mathbb{P}(X_n \neq X) \to 0$ as $n \to \infty$.

Proof

1. If $(x_1, x_2, \ldots)$ is a sequence of points in S and $x \in S$, then relative to metric d, $x_n \to x$ as $n \to \infty$ if and only if $x_n = x$ for all but finitely many $n \in \mathbb{N}_+$.
2. If $\epsilon \ge 1$ then $\mathbb{P}[d(X_n, X) > \epsilon] = 0$. If $\epsilon \in (0, 1)$ then $\mathbb{P}[d(X_n, X) > \epsilon] = \mathbb{P}(X_n \neq X)$.

Of course, it's important to realize that a discrete space can be the Borel space for metrics other than the discrete metric.