Infinitely often, Probability 1, Borel-Cantelli, and the Law of Large Numbers

Michael J. Neely
University of Southern California

Abstract

These notes give details on the probability concepts of "infinitely often" and "with probability 1." This is useful for understanding the Borel-Cantelli lemma and the strong law of large numbers.

I. SEQUENCES OF EVENTS

A. Probability experiment

A probability experiment has

1) A *sample space S*.
2) A *sigma algebra of events $\mathcal{F}$*.
3) A *probability measure $P : \mathcal{F} \to \mathbb{R}$*.

The sample space S is assumed to be a nonempty set. It can be finite or infinite and can consist of different types of objects, such as numbers, vectors, colors, sets, and so on. Elements of the sample space are called *outcomes*. The sigma algebra $\mathcal{F}$ is a set of subsets of S that satisfies structural properties specified in the next subsection. The subsets in $\mathcal{F}$ are called *events* and are precisely those subsets of S for which we define probabilities. The probability measure $P : \mathcal{F} \to \mathbb{R}$ defines a probability $P[A]$ to each event $A \in \mathcal{F}$. The probability measure must satisfy the three axioms of probability (specified, for example, in the Leon-Garcia text).

In particular, every element $\omega \in S$ is an outcome. It is not necessarily true that every subset $A \subseteq S$ is an event. Rather, if $A \subseteq S$ then

- If $A \in \mathcal{F}$ then the set A is an event and has probability $P[A]$.
- If $A \notin \mathcal{F}$ then the set A is not an event and $P[A]$ does not exist.

A subset $A \subseteq S$ that is an event is often called a *measurable set*. A subset $A \subseteq S$ that is not an event is often called a *non-measurable set*.

B. Sigma algebra

Let S be a nonempty set. A set $\mathcal{F}$ of subsets of S is called a *sigma algebra on S* if

1) $\phi \in \mathcal{F}$ (where ϕ is the empty set).
2) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$ (where A^c is the complement of A).
3) If $\{A_i\}_{i=1}^{\infty}$ is a countably infinite sequence of sets such that $A_i \in \mathcal{F}$ for all $i \in \{1, 2, 3, \ldots\}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Since we define a subset of S to be an event if and only if it is in $\mathcal{F}$, and since we define a probability $P[A]$ if and only if set A is an event, the above definition implies that: The empty set ϕ is an event and so $P[\phi]$ is defined (it is easy to prove that $P[\phi] = 0$ from the axioms of probability); If A is an event then A^c is an event and so both $P[A]$ and $P[A^c]$ are defined (it is easy to show that $P[A] + P[A^c] = 1$ from the axioms of probability); If $\{A_i\}_{i=1}^{\infty}$ is a sequence of events then $\cup_{i=1}^{\infty} A_i$ is also an event and so $P[\cup_{i=1}^{\infty} A_i]$ is defined. From the definition of a sigma algebra $\mathcal{F}$, it can be shown that if $\{A_i\}_{i=1}^{\infty}$ is a countably infinite sequence of sets in $\mathcal{F}$ then $\cap_{i=1}^{\infty} A_i$ is also in $\mathcal{F}$ (just use DeMorgan's law).

C. Interpretation

It is common to talk about events as being either "true" or "false." However, if an event is just a subset of S, what does it mean to say that an event can be "true"?

We can imagine that a probability experiment "randomly" produces a single outcome $\omega \in S$. Using this imagination, we can say that an event A is "true" (or that event A "occurs") if the randomly produced outcome ω is an element of A. An event A is "false" if the randomly produced outcome ω satisfies $\omega \notin A$. The probability that the randomly produced outcome is an element of A is equal to $P[A]$. This probability is defined for all $A \in \mathcal{F}$.

D. Limiting events

Let $\{A_n\}_{n=1}^{\infty}$ be an infinite sequence of subsets of S. Let A be another subset of S.

Definition 1: We say that $A_n \searrow A$ if:

- $A_n \supseteq A_{n+1}$ for all $n \in \{1, 2, 3, \ldots\}$.
- $\cap_{n=1}^{\infty} A_n = A$

Definition 2: We say that $A_n \nearrow A$ if:

- $A_n \subseteq A_{n+1}$ for all $n \in \{1, 2, 3, \ldots\}$.
- $\cup_{n=1}^{\infty} A_n = A$

Theorem 1: (Continuity of probability) Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of events. Let A be a subset of S.

a) If $A_n \searrow A$ then A is an event and $P[A_n] \searrow P[A]$.

b) If $A_n \nearrow A$ then A is an event and $P[A_n] \nearrow P[A]$.

Proof: To prove (a), suppose that $A_n \searrow A$. Then $A = \cap_{n=1}^{\infty} A_n$ and so A must be an event (it is the countable intersection of events). Further, for all positive integers n we have $A_n \supseteq A_{n+1} \supseteq A$ and so

$$P[A_n] \geq P[A_{n+1}] \geq P[A] \quad \forall n \in \{1, 2, 3, \ldots\}$$

It follows that $\{P[A_n]\}_{n=1}^{\infty}$ is a sequence of nonincreasing real numbers that are lower bounded by $P[A]$, and so $\lim_{n \to \infty} P[A_n]$ must exist and must be greater than or equal to $P[A]$. The fact that this limit is *exactly* $P[A]$ is proven in the Leon-Garcia book (using Axiom 3 of probability). The proof of part (b) is similar. □

E. Infinitely often and finitely often

Let $\{A_n\}_{n=1}^{\infty}$ be an infinite sequence of events. We say that events in the sequence occur "infinitely often" if A_n holds true for an infinite number of indices $n \in \{1, 2, 3, \ldots\}$. We say that events in the sequence occur "finitely often" if they do not occur infinitely often, that is, if A_n holds true for at most finitely many indices $n \in \{1, 2, 3, \ldots\}$. We formally define the set of all outcomes ω for which an infinite number of the events is true, and for which at most finitely many events are true, as follows:

$$\{A_n \quad i.o.\} = \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} A_n \tag{1}$$

$$\{A_n \quad f.o.\} = \{A_n \quad i.o.\}^c$$

$$= \cup_{m=1}^{\infty} \cap_{n=m}^{\infty} A_n^c \tag{2}$$

where the final equality holds by DeMorgan's law.

The right-hand-side of (1) can be read as follows: For all positive integers m, there exists a positive integer $n \geq m$ such that A_n is true. Some thought will convince you that this holds if and only if an infinite number of the events are true (see also the following lemma). The right-hand-side of (2) can be read as follows: There is a positive integer m such that A_n is false for all integers $n \geq m$. Some thought will convince you that this holds if and only if an at most finite number of the events occur.

A breakdown of the above paragraph is shown below:

$$\{A_n \quad i.o.\} = \underbrace{\cap_{m=1}^{\infty}}_{(\forall m \geq 1)} \underbrace{\cup_{n=m}^{\infty} A_n}_{(\exists n \geq m \text{ such that } A_n \text{ occurs})}$$

$$\{A_n \quad f.o.\} = \underbrace{\cup_{m=1}^{\infty}}_{(\exists m \geq 1)} \underbrace{\cap_{n=m}^{\infty} A_n^c}_{(\forall n \geq m \; A_n \text{ fails to occur})}$$

Here we are using the notation:

- "$\forall m \geq 1$" means "for all $m \geq 1$."
- "$\exists n \geq m$" means "there exists an n that is greater than or equal to m."

Lemma 1: We have

$$\{A_n \quad i.o\} = \{\omega \in S : \omega \in A_n \text{ for an infinite number of indices } n \in \{1, 2, 3, \ldots\}\}$$

$$\{A_n \quad f.o.\} = \{\omega \in S : \omega \in A_n \text{ for at most finitely many indices } n \in \{1, 2, 3, \ldots\}\}$$

Proof: The first equality can be shown by considering two cases (left as an exercise): Fix $\omega \in S$.

1) Suppose $\omega \in A_n$ for an infinite number of indices n. Show that $\cup_{n=m}^{\infty} A_n$ must be true for all $m \in \{1, 2, 3, \ldots\}$.
2) Suppose $\omega \in A_n$ for only finitely many indices n. Show that $\cup_{n=m}^{\infty} A_n$ must be false for some $m \in \{1, 2, 3, \ldots\}$.

The second equality can be obtained from the first equality by taking complements. □

Note that (1) defines $\{A_n \quad i.o\}$ as a countable intersection of a countable union of events, and so $\{A_n \quad i.o.\}$ must itself be an event. That is

$$\{A_n \quad i.o\} \in \mathcal{F}$$

Therefore its complement $\{A_n \quad f.o.\}$ is also an event. As events, they must have probabilities.

Lemma 2: For any sequence of events $\{A_n\}_{n=1}^{\infty}$ we have:

$$\cup_{n=m}^{\infty} A_n \searrow \{A_n \quad i.o\} \tag{3}$$

$$\cap_{n=m}^{\infty} A_n^c \nearrow \{A_n \quad f.o.\} \tag{4}$$

and so

$$\lim_{m \to \infty} P[\cup_{n=m}^{\infty} A_n] = P[A_n \quad i.o.]$$

$$\lim_{m \to \infty} P[\cap_{n=m}^{\infty} A_n^c] = P[A_n \quad f.o.]$$

Proof: Exercise. □

Corollary 1: For any sequence of events $\{A_n\}_{n=1}^{\infty}$ we have:

$$\limsup_{n \to \infty} P[A_n] \leq P[A_n \quad i.o.]$$

$$\liminf_{n \to \infty} P[A_n^c] \geq P[A_n \quad f.o.]$$

Proof: We have for all positive integers n

$$P[A_m] \leq P[\cup_{n=m}^{\infty} A_n]$$

Taking the $\limsup$ of both sides as $m \to \infty$ and using Lemma 2 gives the result. Similarly, for all positive integers n we have

$$P[A_m^c] \geq P[\cap_{n=m}^{\infty} A_n^c]$$

Taking the $\liminf$ of both sides gives the result. □

The above corollary implies the following general facts:

- If $P[A_n] \geq \epsilon$ for infinitely many indices $n \in \{1, 2, 3, \ldots\}$ then $P[A_n \quad i.o.] \geq \epsilon$.
- If $P[A_n^c] \leq \epsilon$ for infinitely many indices $n \in \{1, 2, 3, \ldots\}$ then $P[A_n \quad f.o.] \leq \epsilon$.

F. Preliminaries for Borel-Cantelli lemma

We need two facts about real numbers:
1) $\log(1 + x) \le x$ for all $x \in (-1, \infty)$.
2) If $\{x_i\}_{i=1}^{\infty}$ is a sequence of non-negative real numbers such that $\sum_{i=1}^{\infty} x_i < \infty$, then $\lim_{n \to \infty} \sum_{i=n}^{\infty} x_i = 0$.

The second fact is proven as follows: For all positive integers n we have:

$$\sum_{i=1}^{\infty} x_i = \sum_{i=1}^{n-1} x_i + \sum_{i=n}^{\infty} x_i$$

Taking a limit as $n \to \infty$ gives:

$$\sum_{i=1}^{\infty} x_i = \sum_{i=1}^{\infty} x_i + \lim_{n \to \infty} \sum_{i=n}^{\infty} x_i \tag{5}$$

Now, the equation $\infty = \infty + y$ is true for all $y \in \mathbb{R}$ (we cannot cancel ∞ from both sides to conclude $y = 0$). However, if $\sum_{i=1}^{\infty} x_i < \infty$, we cancel it from both sides of (5) to conclude:

$$0 = \lim_{n \to \infty} \sum_{i=n}^{\infty} x_i$$

G. Borel-Cantelli lemma

Lemma 3: (Borel-Cantelli) Let $\{A_n\}_{n=1}^{\infty}$ be an infinite sequence of events.
a) If $\sum_{n=1}^{\infty} P[A_n] < \infty$, then $P[A_n \quad f.o.] = 1$.
b) If $\sum_{n=1}^{\infty} P[A_n] = \infty$ *and if the events are mutually independent*, then $P[A_n \quad i.o.] = 1$.
Proof: (Part (a)) Suppose $\sum_{n=1}^{\infty} P[A_n] < \infty$. For all positive integers m:

$$\{A_n \quad i.o.\} \subseteq \cup_{n=m}^{\infty} A_n$$

Thus:

$$P[A_n \quad i.o.] \le P[\cup_{n=m}^{\infty} A_n] \le \sum_{n=m}^{\infty} P[A_n]$$

where the final inequality is the union bound. Taking a limit as $m \to \infty$ gives:

$$P[A_n \quad i.o.] \le \lim_{m \to \infty} \sum_{n=m}^{\infty} P[A_n] = 0 \tag{6}$$

where the final limit holds because $\sum_{n=1}^{\infty} P[A_n] < \infty$ (recall fact 2 in the previous subsection). Hence $P[A_n \quad i.o.] = 0$ and so $P[A_n \quad f.o.] = 1$. □
Proof: (Part (b)) Suppose $\sum_{n=1}^{\infty} P[A_n] = \infty$ and that the events are mutually independent. Fix m as a postive integer. By mutual independence we have:

$$P[\cap_{n=m}^{\infty} A_n^c] = \prod_{m=n}^{\infty} P[A_n^c] = \prod_{n=m}^{\infty} (1 - P[A_n])$$

Taking the $\log$ of both sides and using the fact that $\log(1 + x) \le x$ gives:

$$\log(P[\cap_{n=m}^{\infty} A_n^c]) = \sum_{n=m}^{\infty} \log(1 - P[A_n]) \le - \sum_{n=m}^{\infty} P[A_n] = -\infty$$

Thus, $P[\cap_{n=m}^{\infty} A_n^c] = 0$. This holds for all m. Thus:

$$0 = P[\cap_{n=m}^{\infty} A_n^c] \nearrow P[A_n \quad f.o.]$$

So $P[A_n \quad f.o.] = 0$. Hence, $P[A_n \quad i.o.] = 1$. □

Notice that $\sum_{n=1}^{\infty} P[A_n]$ represents the *expected number of events that occur*. Indeed, one can define indicator functions for each $n \in \{1, 2, 3, \ldots\}$ by

$$I_n = \begin{cases} 1 & \text{if event } A_n \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

The number of events that occur is then $N = \sum_{n=1}^{\infty} I_n$. Here, we are treating N as an *extended random variable* that is allowed to take values in the set $\{0, 1, 2, \ldots\} \cup \{\infty\}$. It can be shown that the expectation of a countably infinite sum of nonnegative random variables is equal to the sum of the expectations. Hence:

$$\mathbb{E}[N] = \sum_{n=1}^{\infty} \mathbb{E}[I_n] = \sum_{n=1}^{\infty} P[A_n]$$

H. Coin flipping example

Suppose we flip a sequence of coins. Let A_n be the event that coin flip n produces "heads." Suppose that:

$$P[A_n] = 1/n \quad \forall n \in \{1, 2, 3, \ldots\}$$

Then:

- We know $\sum_{n=1}^{\infty} P[A_n] = \sum_{n=1}^{\infty} 1/n = \infty$. Thus, if the coin flips are mutually independent, the Borel-Cantelli lemma ensures that, with probability 1, there will be an infinite number of heads. This may be surprising because the probability of heads shrinks down to 0. However, the probabilities do not shrink rapidly enough for their sum to be finite.
- Consider the sparsely sampled subsequence of events, sampled at indices that are perfect squares:

$$\{A_{n^2}\}_{n=1}^{\infty} = \{A_1, A_4, A_9, A_{16}, \ldots\}$$

  We know $\sum_{n=1}^{\infty} P[A_{n^2}] = \sum_{n=1}^{\infty} 1/n^2 < \infty$. Thus, the Borel-Cantelli lemma ensures that, with probability 1, only a finite number of perfect square indices produce heads. This holds regardless of whether or not the flips are mutually independent. Indeed, the probabilities $P[A_{n^2}]$ shrink rapidly enough for their sum to be finite.

II. RANDOM VARIABLES CONVERGING TO A LIMIT

Let S be a sample space and let $X : S \to \mathbb{R}$ be a random variable. For each $n \in \{1, 2, 3, \ldots\}$ let $X_n : S \to \mathbb{R}$ be additional random variables. Recall that the single outcome $\omega \in S$ determines the value of *all* random variables

$$X(\omega), X_1(\omega), X_2(\omega), \ldots$$

Define the A as the (possibly empty) set of all outcomes $\omega \in S$ for which $\lim_{n \to \infty} X_n(\omega) = X(\omega)$:

$$A = \{\omega \in S : \lim_{n \to \infty} X_n(\omega) = X(\omega)\}$$

Is A an event, that is, is $A \in \mathcal{F}$? Yes. This is shown in the next subsection where A is written in terms of countable unions and intersections of events.

A. The event $\{\lim_{n \to \infty} X_n = X\}$

For simplicity of notation we can write

$$A = \left\{ \lim_{n \to \infty} X_n = X \right\}$$

with the understanding that $A \subseteq S$. By the definition of a limit we have that $\omega \in A$ if and only if for all $\epsilon > 0$, there exists a positive integer m such that $|X_n(\omega) - X(\omega)| < \epsilon$ for all $n \geq m$. Of course, for every $\epsilon > 0$ we can find a positive integer k such that $1/k \leq \epsilon$. Therefore we can use the following equivalent but more convenient definition of a limit:

- $\omega \in A$ if and only if for all positive integers k there is a positive integer m such that $|X_n(\omega) - X(\omega)| < 1/k$ for all $n \geq m$.

From this definition we immediately obtain:

$$\left\{ \lim_{n \to \infty} X_n = X \right\} = \cap_{k=1}^{\infty} \cup_{m=1}^{\infty} \cap_{n=m}^{\infty} \{|X_n - X| < 1/k\} \tag{7}$$

Indeed, the right-hand-side of the above equality can be read: For all positive integers k, there exists a positive integer m such that $|X_n - X| < 1/k$ for all $n \geq m$. A breakdown of this sentence is below:

$$\boxed{\left\{ \lim_{n \to \infty} X_n = X \right\} = \underbrace{\cap_{k=1}^{\infty}}_{(\forall k \geq 1)} \underbrace{\cup_{m=1}^{\infty}}_{(\exists m \geq 1)} \underbrace{\cap_{n=m}^{\infty} \{|X_n - X| < 1/k\}}_{(\forall n \geq m \text{ we have } |X_n - X| < 1/k)}}$$

Since X_n and X are random variables, it can be shown that for all $\epsilon > 0$ we have

$$\{\omega \in S : |X_n(\omega) - X(\omega)| < \epsilon\} \in \mathcal{F}$$

Thus, the right-hand-side of (7) is a countable intersection of a countable union of a countable intersection of events, and so it is indeed an event:

$$\left\{ \lim_{n \to \infty} X_n = X \right\} \in \mathcal{F}$$

For each $\epsilon > 0$ define

$$B(\epsilon) = \cup_{m=1}^{\infty} \cap_{n=m}^{\infty} \{|X_n - X| < \epsilon\}$$

It follows by definition of "finitely often" in (2) that

$$B(\epsilon) = \{|X_n - X| \geq \epsilon \quad f.o.\} \tag{8}$$

Thus, from (7),

$$\left\{ \lim_{n \to \infty} X_n = X \right\} = \cap_{k=1}^{\infty} B(1/k) \tag{9}$$
$$= \cap_{k=1}^{\infty} \{|X_n - X| \geq 1/k \quad f.o.\}$$

B. Probabilities from limits

Fix $\epsilon > 0$. It is worth mentioning that, by definition of an increasing sequence of events in Section I-D we have

$$\cap_{n=m}^{\infty} \{|X_n - X| < \epsilon\} \nearrow \cup_{m=1}^{\infty} \cap_{n=m}^{\infty} \{|X_n - X| < \epsilon\}$$

The event on the right-hand-side is, by definition of "finitely often" in (2), the event that $\{|X_n - X| \geq \epsilon\}$ finitely often. Thus

$$\cap_{n=m}^{\infty} \{|X_n - X| < \epsilon\} \nearrow \{|X_n - X| \geq \epsilon \quad f.o.\}$$

and so

$$\lim_{n \to \infty} P[\cap_{n=m}^{\infty} |X_n - X| < \epsilon] = P[\{|X_n - X| \geq \epsilon\} \quad f.o.] \tag{10}$$

and hence, by taking the limit of the complement events

$$\lim_{n \to \infty} P[\cup_{n=m}^{\infty} |X_n - X| \geq \epsilon] = P[\{|X_n - X| \geq \epsilon\} \quad i.o.] \tag{11}$$

C. Convergence almost surely

We say that $X_n \to X$ *almost surely* (also called "with probability 1") if

$$P\left[\lim_{n\to\infty} X_n = X\right] = 1$$

That is, if

$$P\left[\omega \in S : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right] = 1$$

Lemma 4: Let X and $\{X_n\}_{n=1}^{\infty}$ be random variables. Then $X_n \to X$ *almost surely if and only if for all $\epsilon > 0$* we have

$$P[\{|X_n - X| \geq \epsilon\} \quad f.o.] = 1$$

Proof: Define $B(\epsilon)$ as in (8). Suppose $P[\lim_{n\to\infty} X_n = X] = 1$. By (9) it holds that

$$P[\cap_{k=1}^{\infty} B(1/k)] = 1$$

Since $B(1/m) \supseteq \cap_{k=1}^{\infty} B(1/k)$ for all positive integers m, it holds that

$$P[B(1/m)] \geq P[\cap_{k=1}^{\infty} B(1/k)] = 1$$

and so $P[B(1/m)] = 1$ for all positive integers m. It follows that $P[B(\epsilon)] = 1$ for all $\epsilon > 0$.

Now suppose $P[B(\epsilon)] = 1$ for all $\epsilon > 0$. Then $P[B(1/k)] = 1$ for all positive integers k and so

$$P[\cap_{k=1}^{\infty} B(1/k)] = 1$$

since the intersection of a countable number of probability-1 events is again a probability-1 event. It holds by (9) that $P[\lim_{n\to\infty} X_n = X] = 1$. □

Lemma 5: Let X and $\{X_n\}_{n=1}^{\infty}$ be random variables. Then $X_n \to X$ almost surely if and only if

$$\lim_{m\to\infty} P[\cup_{n=m}^{\infty}\{|X_n - X| \geq \epsilon\}] = 0 \quad \forall \epsilon > 0 \tag{12}$$

Proof: Suppose $X_n \to X$ almost surely. By the previous lemma we have for all $\epsilon > 0$ that

$$P[\{|X_n - X| \geq \epsilon\} \quad f.o.] = 1$$

and so

$$P[\{|X_n - X| \geq \epsilon\} \quad i.o.] = 0$$

Substituting the above equality into (11) yields (12).

Now suppose (12) holds. By (11) this implies that for all $\epsilon > 0$

$$P[\{|X_n - X| \geq \epsilon\} \quad i.o.] = 0$$

and so

$$P[\{|X_n - X| \geq \epsilon\} \quad f.o.] = 1$$

which implies $X_n \to X$ almost surely (by the previous lemma). □

D. Different types of convergence

Definition 3: $X_n \to X$ *in mean square* if:

$$\lim_{n\to\infty} \mathbb{E}\left[(X_n - X)^2\right] = 0$$

Definition 4: $X_n \to X$ *in probability* if for all $\epsilon > 0$:

$$\lim_{n\to\infty} P[|X_n - X| \geq \epsilon] = 0$$

Definition 5: $X_n \to X$ *almost surely* (also called *with probability 1*) if $P[\lim_{n\to\infty} X_n = X] = 1$. Equivalently (by Lemma 5), $X_n \to X$ almost surely if for all $\epsilon > 0$ we have

$$\lim_{n\to\infty} P[\cup_{i=n}^{\infty} |X_i - X| \geq \epsilon] = 0$$

Notice that convergence almost surely implies convergence in probability. That is because

$$\{|X_n - X| \geq \epsilon\} \subseteq \cup_{i=n}^{\infty}\{|X_n - X| \geq \epsilon\}$$

and so

$$P[|X_n - X| \geq \epsilon] \leq P[\cup_{i=n}^{\infty}\{|X_n - X| \geq \epsilon\}]$$

and if the right-hand-side goes to zero as $n \to \infty$, the left-hand-side must also. It can also be shown (via the Markov inequality) that mean square convergence implies convergence in probability. If the random variables X_n are deterministically bounded, say, always taking values in $[-M, M]$ for some positive integer M, then it can be shown that convergence in probability is equivalent to convergence in mean square.

There are examples where bounded random variables converge in probability, but not with probability 1 (try constructing such an example using the coin flips of Section I-H). There are also examples where (unbounded) random variables converge with probability 1, but not in mean square.

III. LAW OF LARGE NUMBERS

A. Markov and Chebyshev inequalities

Recall that:

Lemma 6: (Markov inequality) If X is a non-negative random variable then for all $\epsilon > 0$:

$$P[X \geq \epsilon] \leq \frac{\mathbb{E}[X]}{\epsilon}$$

Lemma 7: (Chebyshev inequality) If X is a random variable (possibly negative) with mean m and variance σ^2, then:

$$P[|X - m| \geq \epsilon] \leq \frac{Var(X)}{\epsilon^2}$$

The Chebyshev inequality can be proven from the Markov inequality by defining the non-negative random variable $Y = (X - m)^2$ and noting that $\{|X - m| \geq \epsilon\} = \{(X - m)^2 \geq \epsilon^2\}$.

B. Law of large numbers

Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of pairwise uncorrelated random variables with identical means and bounded variances:

$$\mathbb{E}[X_n] = m \quad \forall n \in \{1, 2, 3, \ldots\} \tag{13}$$
$$Var(X_n) \leq c \quad \forall n \in \{1, 2, 3, \ldots\} \tag{14}$$
$$Cov(X_i, X_j) = 0 \quad \forall i \neq j \tag{15}$$

The values m and c are assumed to be finite. The random variables can have different distributions. A special case is when the random variables are independent and identically distributed (i.i.d.), in which case they all have the same (finite) variance σ^2.

For positive integers n, define M_n as the average over the first n random variables:

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Notice that:

$$\mathbb{E}[M_n] = m \quad, \quad Var(M_n) \leq \frac{c}{n} \quad \forall n \in \{1, 2, 3, \ldots\}$$

The fact that the mean is always m and the variance shrinks to 0 suggests that the random variables M_n are converging to the mean m. In what sense does this convergence hold? The variance condition directly implies that the random variables X_n converge to m in the mean square sense. The Chebyshev inequality can be used to further show that convergence in probability holds. The *strong law of large numbers* makes the much deeper claim that convergence in fact holds *with probability 1*.

Theorem 2: (Law of large numbers with finite variance) Suppose $\{X_i\}_{i=1}^{\infty}$ are pairwise uncorrelated with identical means and bounded variances, so that (13)-(15) hold. In particular, suppose $\mathbb{E}[X_i] = m$ for all $i \in \{1, 2, 3, \ldots\}$. Define $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then:

a) $M_n \to m$ in mean square.
b) $M_n \to m$ in probability.
c) $M_n \to m$ almost surely (that is, with probability 1).
Proof: See next subsection. $\square$

Theorem 3: (Law of large numbers with infinite variance) Suppose $\{X_i\}_{i=1}^{\infty}$ are i.i.d. with finite mean m and with possibly infinite variance. Define $M_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then $M_n \to m$ almost surely (i.e., with probability 1).
Proof: See [1]. $\square$

The statement in Theorem 2 part (b) is referred to as the "weak law of large numbers." It is included for pedagogical reasons: Its proof is very simple and gives intuition behind the deeper proof for the strong law. The statement in Theorem 2 part (c) is a version of the "strong law of large numbers." This statement is more general than most statements of the strong law of large numbers because it allows the random variables to have different distributions and only requires them to be pairwise uncorrelated (rather than mutually independent). However, it does require all variances to be bounded. The usual statement of the strong law of large numbers is given in Theorem 3: This statement requires the random variables to be independent and identically distributed (i.i.d.), still requires them to have a finite mean, but allows for a possibly infinite variance.[1]

C. Proof of Theorem 2

Proof: (Theorem 2 part (a)) Since $\mathbb{E}[M_n] = m$ for all $n \in \{1, 2, 3, \ldots\}$, we have

$$\mathbb{E}\left[(M_n - m)^2\right] = Var(M_n) \leq \frac{c}{n} \to 0$$

and so $M_n \to m$ in mean square. $\square$

Proof: (Theorem 2 part (b)) Fix $\epsilon > 0$. By the Chebyshev inequality:

$$P[|M_n - m| \geq \epsilon] \leq \frac{Var(M_n)}{\epsilon^2} \leq \frac{c}{n\epsilon^2} \to 0$$

and so $M_n \to m$ in probability.

$\square$

Proof: (Theorem 2 part (c)) We first show that $M_{n^2} \to m$ almost surely, that is, we have convergence almost surely over the sparse subsequence where indices are perfect squares: $\{M_1, M_4, M_9, M_{16}, \ldots\}$. Fix $\epsilon > 0$. We have by the Chebyshev inequality:

$$P[|M_{n^2} - m| \geq \epsilon] \leq \frac{Var(M_{n^2})}{\epsilon^2} \leq \frac{c}{n^2\epsilon^2}$$

Since $\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$ we obtain

$$\sum_{n=1}^{\infty} P[|M_{n^2} - m| \geq \epsilon] < \infty$$

and so the Borel-Cantelli theorem implies

$$P[\{|M_{n^2} - m| \geq \epsilon\} \quad f.o.] = 1$$

This holds for all $\epsilon > 0$ and so Lemma 4 implies $M_{n^2} \to m$ almost surely.

It remains to show $M_n \to m$ almost surely (so that the same convergence holds over all indices n, not just indices that are perfect squares). For each $n \in \{1, 2, 3, \ldots\}$ define $k(n)$ as the largest integer such that $k(n)^2 \leq n$, and so

$$k(n)^2 \leq n < (k(n) + 1)^2 \quad \forall n \in \{1, 2, 3, \ldots\} \tag{16}$$

Then

$$M_n = \frac{1}{n} \sum_{i=1}^{k(n)^2} X_i + \frac{1}{n} \sum_{i=k(n)^2+1}^{n} X_i$$

---

[1] The i.i.d. assumption in Theorem 3 can be relaxed to "pairwise independent and identically distributed," see [1].

It suffices to show the following two limits hold almost surely (that is, with probability 1):

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{k(n)^2} X_i = m \tag{17}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=k(n)^2+1}^{n} X_i = 0 \tag{18}$$

To prove (17) we observe

$$\frac{1}{n} \sum_{i=1}^{k(n)^2} X_i = \frac{k(n)^2}{n} \frac{1}{k(n)^2} \sum_{i=1}^{k(n)^2} X_i$$

and since $\frac{k(n)^2}{n} \to 1$, and $\frac{1}{k(n)^2} \sum_{i=1}^{k(n)^2} X_i \to m$ almost surely (since we have already proven almost sure convergence over indices that are perfect squares) we conclude that (17) holds almost surely.

To prove (18) we observe

$$\frac{1}{n} \sum_{i=k(n)^2+1}^{n} X_i = \frac{1}{n} \sum_{i=k(n)^2+1}^{n} (X_i - m) + \frac{1}{n} \sum_{i=k(n)^2+1}^{n} m$$

$$= \frac{1}{n} \sum_{i=k(n)^2+1}^{n} (X_i - m) + m\left(\frac{n - k(n)^2}{n}\right)$$

and since by (16) we have

$$\frac{n - k(n)^2}{n} \to 0$$

it suffices to prove

$$\frac{1}{n} \sum_{i=k(n)^2+1}^{n} (X_i - m) \to 0 \quad \text{(almost surely)}$$

Define $V_n = \frac{1}{n} \sum_{i=k(n)^2+1}^{n}(X_i - m)$. Fix $\epsilon > 0$. Since $\mathbb{E}[V_n] = 0$ we have

$$P[|V_n| \geq \epsilon] \overset{(a)}{\leq} \frac{Var(V_n)}{\epsilon^2}$$

$$\overset{(b)}{=} \frac{\sum_{i=k(n)^2+1}^{n} Var(X_i)}{n^2\epsilon^2}$$

$$\overset{(c)}{\leq} c\frac{n - k(n)^2}{n^2\epsilon^2}$$

$$\overset{(d)}{\leq} c\frac{(k(n)+1)^2 - k(n)^2}{k(n)^4\epsilon^2}$$

$$= c\frac{2k(n)+1}{k(n)^4\epsilon^2}$$

$$\overset{(e)}{\leq} c\frac{2k(n) + k(n)}{k(n)^4\epsilon^2}$$

$$= \frac{3c}{k(n)^3\epsilon^2}$$

where (a) holds by the Chebyshev inequality; (b) holds because $\{X_i - m\}_{i=1}^{\infty}$ are pairwise uncorrelated; (c) holds because $Var(X_i) \leq c$ for all i; (d) holds because $n^2 \geq k(n)^4$ and $n < (k(n)+1)^2$; (e) holds because $k(n) \geq 1$. Thus

$$\sum_{n=1}^{\infty} P[|V_n| \geq \epsilon] \leq \frac{3c}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{k(n)^3} < \infty \tag{19}$$

which holds because (16) implies $k(n) > -1 + \sqrt{n}$ for all $n \in \{1, 2, 3, \ldots\}$ and so

$$\sum_{n=2}^{\infty} \frac{1}{k(n)^3} \leq \sum_{n=2}^{\infty} \frac{1}{(-1 + \sqrt{n})^3} < \infty$$

By the Borel-Cantelli theorem, (19) implies $|V_n| \geq \epsilon$ finitely often (with probability 1). This holds for all $\epsilon > 0$ and so Lemma 4 ensures that $V_n \to 0$ almost surely. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

D. Distinctions between the law of large numbers and the central limit theorem

Let $\{X_1, X_2, X_3, \ldots\}$ be i.i.d. with finite mean and variance given by $\mathbb{E}[X_1] = m$, $Var(X_1) = \sigma^2$. Assume $\sigma^2 > 0$. Define:

$$G_n = \frac{1}{\sqrt{n\sigma^2}} \sum_{i=1}^{n} (X_i - m)$$

$$M_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

It is important to know how to derive the following facts:

$$\mathbb{E}[G_n] = 0 \quad , \quad Var(G_n) = 1 \quad \forall n \in \{1, 2, 3, \ldots\}$$

$$\mathbb{E}[M_n] = m \quad , \quad Var(M_n) = \frac{\sigma^2}{n} \quad \forall n \in \{1, 2, 3, \ldots\}$$

Since $Var(M_n)$ converges to 0, it is easy to remember that M_n converges to the constant m (with probability 1). Since $Var(G_n) = 1$ for all n, it is easy to see that G_n does not converge to a constant. In fact, G_n does not converge to a random variable either: It oscillates without converging to anything as n increases. However, the central limit theorem says that the *distribution* of G_n converges to a Gaussian distribution with zero mean and unit variance:

$$\lim_{n \to \infty} P[G_n \leq x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt \quad \forall x \in \mathbb{R}$$

REFERENCES

[1] N. Etemadi. An elementary proof of the strong law of large numbers. Z. *Wahrscheinlichkeitstheorie verw. Geiete*, vol. 5:pp. 119–122, 1981.