A note on "infinitely often," "probability 1," and the law of large numbers

Michael J. Neely
University of Southern California
http://www-bcf.usc.edu/~mjneely

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. Events

Consider a general sample space $\mathcal{S}$. Recall that an *event* is a subset of $\mathcal{S}$ that has a well defined probability. That is, a set A is an event if and only if $A \subseteq \mathcal{S}$ and $P[A]$ exists. Formally, a probability experiment introduces both a sample space $\mathcal{S}$ and a *sigma algebra* $\mathcal{F}$. The sigma algebra $\mathcal{F}$ contains all events, that is, it is the collection of all subsets of $\mathcal{S}$ for which probabilities are defined.

B. Can an event be "true"?

It is common to talk about events as being either "true" or "false." However, if an event is just a subset of $\mathcal{S}$, does it make sense to say that an event can be "true"? When we say that an event A is "true" (or that an event A "occurs"), we are imagining a probability experiment that produces an outcome $\omega \in \mathcal{S}$ for which $\omega \in A$. The event is "false" if the outcome satisfies $\omega \notin A$.

C. Shrinking a sequence of events

Let $\{A_n\}_{n=1}^{\infty}$ be an infinite sequence of events. Let A be another event. We say that $A_n \searrow A$ if:
- $A_n \supseteq A_{n+1}$ for all $n \in \{1, 2, 3, \ldots\}$.
- $\cap_{n=1}^{\infty} A_n = A$

We know that if $A_n \searrow A$ then $P[A_n] \searrow P[A]$.

D. Infinitely 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_i holds true for an infinite number of indices $i \in \{1, 2, 3, \ldots\}$. Define $\{A_i \quad i.o\}$ as the event that an infinite number of events A_i occur. Formally, this new event needs to be a subset of $\mathcal{S}$. Hence:

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

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

$$\cup_{i=n}^{\infty} A_i \searrow \{A_i \quad i.o\} \tag{2}$$

and so

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

Proof. To prove (2), note that:

$$\cup_{i=n}^{\infty} A_i \supseteq \cup_{i=n+1}^{\infty} A_i \quad \forall n \in \{1, 2, 3, \ldots\}$$

It remains to show that:

$$\cap_{n=1}^{\infty} \cup_{i=n}^{\infty} A_i = \{A_i \quad i.o.\} \tag{3}$$

This can be done by considering two cases (left as an exercise):
1) Case 1: Suppose $\{A_i \quad i.o\}$ is true. Show that $\cup_{i=n}^{\infty} A_i$ must be true for all $n \in \{1, 2, 3, \ldots\}$.
2) Case 2: Suppose $\{A_i \quad i.o\}$ is false. Show that $\cup_{i=n}^{\infty} A_i$ must be false for some $n \in \{1, 2, 3, \ldots\}$.

□

Note that equation (3) in the above proof shows that the event $\{A_i \quad i.o.\}$ can be written in terms of unions and intersections of the original events A_i. Formally, this means that the set $\{A_i \quad i.o.\}$ defined in (1) is indeed in the sigma algebra of events.

E. Finitely often

We say that A_i occurs "finitely often" if A_i holds true for an at most finite number of indices $i \in \{1, 2, 3, \ldots\}$. Specifically:

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

By taking complements in Lemma 1 we obtain:

$$\cap_{i=n}^{\infty} A_i^c \nearrow \{A_i \quad f.o.\} \implies \lim_{n\to\infty} P[\cap_{i=n}^{\infty} A_i^c] = P[A_i \quad f.o.]$$

Taking complements of (3) gives:

$$\cup_{n=1}^{\infty} \cap_{i=n}^{\infty} A_i^c = \{A_i \quad f.o.\}$$

F. Preliminaries for Borel-Cantelli lemma

We need two facts about real numbers:

1) $\log(1+x) \leq x$ 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:

$$\sum_{i=1}^{\infty} x_i = \sum_{i=1}^{\infty} x_i + \lim_{n\to\infty} \sum_{i=n}^{\infty} x_i \tag{4}$$

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 (4) to conclude:

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

G. Borel-Cantelli lemma

Lemma 2. *(Borel-Cantelli) Let $\{A_n\}_{n=1}^{\infty}$ be an infinite sequence of events.*

a) If $\sum_{i=1}^{\infty} P[A_i] < \infty$, then $P[A_i \quad f.o.] = 1$.

b) If $\sum_{i=1}^{\infty} P[A_i] = \infty$ and if the events are mutually independent*, then $P[A_i \quad i.o.] = 1$.*

Proof. (Part (a)) Suppose $\sum_{i=1}^{\infty} P[A_i] < \infty$. For all positive integers n:

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

Thus:

$$P[A_i \quad i.o.] \leq P[\cup_{i=n}^{\infty} A_i] \leq \sum_{i=n}^{\infty} P[A_i]$$

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

$$P[A_i \quad i.o.] \leq \lim_{n\to\infty} \sum_{i=n}^{\infty} P[A_i] = 0 \tag{5}$$

where the final limit holds because $\sum_{i=1}^{\infty} P[A_i] < \infty$ (recall fact 2 in the previous subsection). Hence $P[A_i \quad i.o.] = 0$ and so $P[A_i \quad f.o.] = 1$. □

Proof. (Part (b)) Suppose $\sum_{i=1}^{\infty} P[A_i] = \infty$ and that the events are mutually independent. Fix n as a postive integer. We have:

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

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

$$\log(P[\cap_{i=n}^{\infty} A_i^c]) = \sum_{i=n}^{\infty} \log(1 - P[A_i]) \leq -\sum_{i=n}^{\infty} P[A_i] = -\infty$$

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

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

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

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

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

The number of events that occur is then $N = \sum_{i=1}^{\infty} I_i$. 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:

$$\mathbb{E}[N] = \sum_{i=1}^{\infty} \mathbb{E}[I_i] = \sum_{i=1}^{\infty} P[A_i]$$

H. *Coin flipping example*

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

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

Then:

- We know $\sum_{i=1}^{\infty} P[A_i] = \sum_{i=1}^{\infty} 1/i = \infty$. 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:

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

  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. SEQUENCES OF RANDOM VARIABLES

Let $\mathcal{S}$ be a sample space.

A. *Markov and Chebyshev inequalities*

Recall that:

Lemma 3. *(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 4. *(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. Random sequences defined by outcomes in $\mathcal{S}$

We want to treat infinite sequences of random variables $\{X_1, X_2, X_3, \ldots\}$. When working with random sequences, we know that the *entire sequence* is determined by the particular outcome ω in the sample space $\mathcal{S}$. That is, the sequence $\{X_1, X_2, X_3, \ldots\}$ can be written more formally as $\{X_1(\omega), X_2(\omega), X_3(\omega), \ldots\}$.

Suppose X is another random variable on the same probability space, so that X can be written as $X(\omega)$. We say that $\lim_{n\to\infty} X_n = X$ is true if the probability experiment produces an outcome $\omega \in \mathcal{S}$ such that $\lim_{n\to\infty} X_n(\omega) = X(\omega)$. Formally:

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

The complement event $\{\lim_{n\to\infty} X_n = X\}^c$ is the set of all outcomes ω for which either the limit does not exist or the limit exists but is not equal to $X(\omega)$.

We say that a sequence of random variables $\{X_n\}_{n=1}^{\infty}$ converges to a random variable X *with probability 1* if:

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

Probability 1 convergence is also called "almost sure" convergence. It is useful to remember the standard definition of a limit:

Definition 1. $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ *if for all* $\epsilon > 0$*, there exists an integer* k *such that:*

$$|X_n(\omega) - X(\omega)| \leq \epsilon \quad \forall n \geq k$$

C. Different types of convergence

Here are three useful types of convergence: (i) mean square convergence, (ii) convergence in probability, (iii) convergence with probability 1.

Definition 2. $X_n \to X$ in mean square *if:*

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

Definition 3. $X_n \to X$ in probability *if for all* $\epsilon > 0$*:*

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

Definition 4. $X_n \to X$ with probability 1 *if for all* $\epsilon > 0$*:*

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

equivalently, if $P[\lim_{n\to\infty} X_n = X] = 1$.

Subsection II-E shows why the two definitions given for "probability 1 convergence" are equivalent. Notice that probability 1 convergence implies convergence in probability. That is because, for all positive integers n,

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

and so

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

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.

D. A useful fact about limits

Let $\{X_n\}_{n=1}^{\infty}$ be random variables and let X be another random variable, all on the same probability space $\mathcal{S}$. It can be shown that:

$$\{\lim_{n\to\infty} X_n = X\} = \cap_{k=1}^{\infty}\{|X_i - X| > 1/k \quad f.o.\} \tag{6}$$

The right-hand-side is the event that, for each positive integer k, the inequality $|X_i - X| > 1/k$ holds for only a finite number of indices i, so that there is an n for which $|X_i - X| \leq 1/k$ for all $i \geq n$. Taking complements of (6) gives:

$$\{\lim_{n\to\infty} X_n = X\}^c = \cup_{k=1}^{\infty}\{|X_i - X| > 1/k \quad i.o.\} \tag{7}$$

As a minor detail: There is nothing special about the $1/k$ function used in (6). That can equally be replaced by any positive function $g(k)$ that satisfies $\lim_{k\to\infty} g(k) = 0$.

E. Equivalent definitions of "with probability 1"

Lemma 5. *The following are equivalent statements, all meaning that $X_n \to X$ with probability 1.*

1) *Statement 1:*
$$P[\lim_{n\to\infty} X_n = X] = 1$$
2) *Statement 2: For all $\epsilon > 0$ we have:*
$$\lim_{n\to\infty} P[\cup_{i=n}^{\infty}\{|X_i - X| > \epsilon\}] = 0$$
3) *Statement 3: For all $\epsilon > 0$ we have:*
$$P[|X_i - X| > \epsilon \quad i.o.] = 0$$

Proof. ($2 \iff 3$) To understand the equivalence between statements 2 and 3, fix $\epsilon > 0$ and observe
$$\cup_{i=n}^{\infty}\{|X_i - X| > \epsilon\} \searrow \{|X_i - X| > \epsilon \quad i.o.\}$$
Hence
$$\lim_{n\to\infty} P\left[\cup_{i=n}^{\infty}\{|X_i - X| > \epsilon\}\right] = P[|X_i - X| > \epsilon \quad i.o.]$$
Thus, the probabilities in statements 2 and 3 are the same, so the two statements are equivalent. □

Proof. ($3 \implies 1$) Suppose statement 3 holds, so for all $\epsilon > 0$ we have $P[|X_i - X| > \epsilon \quad i.o.] = 0$. We want to show that $P[\lim_{n\to\infty} X_n = X] = 1$. It suffices to show that $P[(\lim_{n\to\infty} X_n = X)^c] = 0$. By (7):
$$\begin{aligned}
P[(\lim_{n\to\infty} X_n = X)^c] &= P[\cup_{k=1}^{\infty}\{|X_i - X| > 1/k \quad i.o.\}] \\
&\overset{(a)}{\leq} \sum_{k=1}^{\infty} P[|X_i - X| > 1/k \quad i.o.] \\
&= \sum_{k=1}^{\infty} 0 = 0
\end{aligned}$$
where (a) holds by the union bound. □

Proof. ($1 \implies 3$) Suppose statement 1 holds, so that $P[\lim_{n\to\infty} X_n = X] = 1$. We want to show statement 3 holds. It suffices to show that for all positive integers m we have $P[|X_i - X| > 1/m \quad i.o.] = 0$. Fix a positive integer m. Then:
$$\begin{aligned}
1 &= P[\lim_{n\to\infty} X_n = X] \\
&\overset{(a)}{=} P[\cap_{k=1}^{\infty}\{|X_i - X| > 1/k \quad f.o.\}] \\
&\leq P[|X_i - X| > 1/m \quad f.o.]
\end{aligned}$$
where (a) holds by (6). It follows that $P[|X_i - X| > 1/m \quad f.o.] = 1$, and so the probability of its complement is 0. □

III. Law of large numbers

Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of mutually independent and identically distributed (i.i.d.) random variables with mean m and variance σ^2. For now, assume both the mean and variance are finite. For positive integers n, define L_n as the average over the first n random variables:
$$L_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$
Notice that:
$$\mathbb{E}[L_n] = m \quad , \quad Var(L_n) = \frac{\sigma^2}{n} \quad \forall n \in \{1, 2, 3, \dots\}$$
The fact that the mean is always m and the variance shrinks to 0 suggests that the random variables L_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 1. *(Strong law of large numbers) Suppose $\{X_i\}_{i=1}^{\infty}$ are i.i.d. with finite mean and variance m and σ^2. Define $L_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then:*

a) *$L_n \to m$ in probability.*
b) *$L_n \to m$ with probability 1.*
c) *Both (a) and (b) hold even if $\sigma^2 = \infty$.*

d) Parts (a)-(c) hold if "i.i.d." is replaced by "identically distributed and pairwise independent."

The statement in part (a) 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.

Proof. (LLN part (a)) Fix $\epsilon > 0$. By the Chebyshev inequality:

$$P[|L_n - m| > \epsilon] \leq \frac{Var(L_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0$$

□

Proof. (LLN part (b)) Suppose the result of part (b) holds whenever the random variables $\{X_n\}_{n=1}^{\infty}$ are nonnegative (the next subsection shows it is indeed true in the nonnegative case). For general (possibly negative) random variables $\{X_n\}_{n=1}^{\infty}$ define:

$$X_n^+ = \max[X_n, 0] \quad , \quad X_n^- = \max[-X_n, 0]$$

Then $X_n = X_n^+ - X_n^-$, and X_n^+ and X_n^- are nonnegative. Because $\{X_n\}_{n=1}^{\infty}$ are i.i.d. with finite means and variances, the random variables $\{X_n^+\}_{n=1}^{\infty}$ are also i.i.d. with finite means and variances, as are the variables $\{X_n^-\}_{n=1}^{\infty}$. Define $m^+ = \mathbb{E}\left[X_1^+\right]$ and $m^- = \mathbb{E}\left[X_1^-\right]$. Then:

$$m = \mathbb{E}\left[X_1\right] = \mathbb{E}\left[X_1^+\right] - \mathbb{E}\left[X_1^-\right] = m^+ - m^-$$

Thus, with probability 1:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i^+ - \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i^- = m^+ - m^- = m$$

□

For a proof of parts (c)-(d), see [1].

A. *Extensions*

The following exercise extends the weak law of large numbers to treat random variables with possible dependencies and with different distributions.

Exercise 1. *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of random variables with finite means and variances. Suppose:*

- $\mathbb{E}\left[X_i\right] = m$ *for all* $i \in \{1, 2, 3, \ldots\}$.
- $Var(X_i) = \sigma_i^2$ *for all* $i \in \{1, 2, 3, \ldots\}$.
- *The random variables are pairwise uncorrelated, so that* $\mathbb{E}\left[X_i X_j\right] = m^2$ *whenever* $i \neq j$.
- $\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 = 0$.

a) Prove that $\frac{1}{n}\sum_{i=1}^{n} X_i \to m$ in probability. Hint: Just follow the standard proof of the weak law.
b) Prove that $\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 = 0$ holds whenever $\sigma_i^2 \leq \sigma_{max}^2$ for all $i \in \{1, 2, 3, \ldots\}$, where σ_{max}^2 is a finite constant.

Exercise 2. *Consider the same scenario as the previous exercise, but replace the condition $\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 = 0$ with the more stringent condition $\sigma_i^2 \leq \sigma_{max}^2$ for all $i \in \{1, 2, 3, \ldots\}$, where σ_{max}^2 is a finite constant. Define $L_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Prove that L_{n^2} converges to m with probability 1. Hint: Prove for all $\epsilon > 0$ that $P[|L_{n^2} - m| > \epsilon \quad i.o.] = 0$.*

If the average of a sequence X_n converges to a constant m when sampled over a sparse subsequence of times that correspond to perfect squares, does that mean the average converges to m? It does in the special case when the sequence X_n is *non-negative*. Specifically, suppose $\{X_1, X_2, X_3, \ldots\}$ is a non-negative sequence. Define $L_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Suppose we know that $L_{n^2} \to m$ for some finite constant m. We want to show that $L_n \to m$. For any positive integer k, define $n[k]$ as the largest integer n such that $n^2 \leq k$. Thus, for all positive integers k we have

$$n[k]^2 \leq k < (n[k]+1)^2$$

Then for any positive integer k we have (because the X_i values are non-negative):

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

Thus:

$$\left[\frac{n[k]^2}{(n[k]+1)^2}\right] \underbrace{\frac{1}{n[k]^2}\sum_{i=1}^{n[k]^2} X_i}_{L_{n[k]^2}} \leq \frac{1}{k}\sum_{i=1}^{k} X_i \leq \left[\frac{(n[k]+1)^2}{n[k]^2}\right] \underbrace{\frac{1}{(n[k]+1)^2}\sum_{i=1}^{(n[k]+1)^2} X_i}_{L_{(n[k]+1)^2}}$$

Hence, for all positive integers k:

$$\left[\frac{n[k]^2}{(n[k]+1)^2}\right] L_{n[k]^2} \leq \frac{1}{k}\sum_{i=1}^{k} X_i \leq \left[\frac{(n[k]+1)^2}{n[k]^2}\right] L_{(n[k]+1)^2}$$

Note that $\lim_{k\to\infty} n[k] = \infty$. Taking a limit as $k \to \infty$ and using the fact that $L_{n^2} \to m$ gives:

$$m \leq \lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} X_i \leq m$$

B. *Where is independence used?*

The main idea behind the law of large numbers is that the variance of a sum of pairwise uncorrelated random variables is equal to the sum of the variances. So, why is "independence" so important?

The discussion after Exercise 2 shows that, in the special case when the random variables are non-negative and have finite variance, the assumption that the variables are independent can be replaced by the weaker assumption that they are pairwise uncorrelated. However, independence is used when extending this result to general (possibly negative) random variables. Why? If X_1 and X_2 are uncorrelated, their positive parts X_1^+ and X_2^+ are not necessarily uncorrelated. However, if X_1 and X_2 are *independent*, then indeed their positive parts X_1^+ and X_2^+ are *also independent* (and hence uncorrelated). Likewise, their negative parts X_1^- and X_2^- are independent (and hence uncorrelated).

Both "independence" and "identically distributed" are used in the proof of [1] that extends to the case of infinite variance.

The strong law of large numbers only requires the random variables to be pairwise independent. Mutual independence is not needed. However, most textbooks state the result in terms of "i.i.d." random variables, which assumes mutual independence. This is because several other important results use i.i.d. random variables (including the central limit theorem). It is easiest to consistently use "i.i.d." rather than changing the assumptions to fit each different result.

IV. What are the most important things to remember?

The mean of a random variable X is denoted $\mathbb{E}[X]$. If the mean is finite, the variance is defined by:

$$Var(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

A. *Mean and variance facts*

- $\mathbb{E}[X_1 + ... + X_n] = \mathbb{E}[X_1] + ... + \mathbb{E}[X_n]$.
- $Var(aX) = a^2 Var(X)$.
- If $\{X_1, ..., X_n\}$ are pairwise uncorrelated, then $Var(X_1 + ... + X_n) = Var(X_1) + ... + Var(X_n)$.
- Since "i.i.d." implies "pairwise uncorrelated," the same variance-sum result holds for i.i.d. random variables.

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

Let $\{X_1, X_2, X_3, ...\}$ 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:[1]

$$\begin{aligned} C_n &= \frac{1}{\sqrt{n\sigma^2}}\sum_{i=1}^{n}(X_i - m) \\ L_n &= \frac{1}{n}\sum_{i=1}^{n} X_i \end{aligned}$$

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

$$\mathbb{E}[C_n] = 0 \quad , \quad Var(C_n) = 1 \quad \forall n \in \{1, 2, 3, \dots\}$$
$$\mathbb{E}[L_n] = m \quad , \quad Var(L_n) = \frac{\sigma^2}{n} \quad \forall n \in \{1, 2, 3, \dots\}$$

Since $Var(L_n)$ converges to 0, it is easy to remember that L_n converges to the constant m (with probability 1). Since $Var(C_n) = 1$ for all n, it is easy to see that C_n does not converge to a constant. In fact, C_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 C_n converges to a Gaussian distribution with zero mean and unit variance:

$$\lim_{n\to\infty} P[C_n > x] = Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

References

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

[1] The letter C is used to remind you of the $\boxed{C}$ *entral limit theorem*. The letter L is used to remind you of the $\boxed{L}$ *aw of large numbers*.