8. Borel-Cantelli Lemmas

Given a sequence of events $A_1, A_2, ... \in \mathcal{F}$, we define

$$\limsup_n A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{\omega : \omega \text{ is in infinitely many } A_n\},$$

which is often abbreviated as $\{A_n \text{ i.o.}\}$ where "i.o." stands for "infinitely often."

The nomenclature derives from the straight-forward identity $\limsup_{n\to\infty} 1_{A_n} = 1_{\limsup_n A_n}$.

One can likewise define the limit inferior by

$$\liminf_n A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = \{\omega : \omega \text{ is in all but finitely many } A_n\},$$

but little is gained by doing so since $\liminf_n A_n = \left(\limsup_n A_n^C\right)^C$.

To illustrate the utility of this notion, observe that $X_n \to X$ a.s. if and only if $P\left(|X_n - X| > \varepsilon \text{ i.o.}\right) = 0$ for every $\varepsilon > 0$.

Lemma 8.1 (Borel-Cantelli I)**.** *If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n \text{ i.o.}) = 0$.*

Proof. Let $N = \sum_{n=1}^{\infty} 1_{A_n}$ denote the number of events that occur. Tonelli's theorem (or MCT) gives

$$E[N] = \sum_{n=1}^{\infty} E[1_{A_n}] = \sum_{n=1}^{\infty} P(A_n) < \infty,$$

so it must be the case that $N < \infty$ a.s. $\square$

A nice application of the first Borel-Cantelli lemma is

Theorem 8.1. *$X_n \to_p X$ if and only if every subsequence $\{X_{n_m}\}_{m=1}^{\infty}$ has a further subsequence $\{X_{n_{m(k)}}\}_{k=1}^{\infty}$ such that $X_{n_{m(k)}} \to X$ a.s. as $k \to \infty$.*

Proof.

Suppose that $X_n \to_p X$ and let $\{X_{n_m}\}_{m=1}^{\infty}$ be any subsequence. Then $X_{n_m} \to_p X$, so for every $k \in \mathbb{N}$, $P\left(|X_{n_m} - X| > \frac{1}{k}\right) \to 0$ as $m \to \infty$. It follows that we can choose a further subsequence $\{X_{n_{m(k)}}\}_{k=1}^{\infty}$ such that $P\left(\left|X_{n_{m(k)}} - X\right| > \frac{1}{k}\right) \le 2^{-k}$ for all $k \in \mathbb{N}$. Since

$$\sum_{k=1}^{\infty} P\left(\left|X_{n_{m(k)}} - X\right| > \frac{1}{k}\right) \le 1 < \infty,$$

the first Borel-Cantelli lemma shows that $P\left(\left|X_{n_{m(k)}} - X\right| > \frac{1}{k} \text{ i.o.}\right) = 0$.

Because $\left\{\left|X_{n_{m(k)}} - X\right| > \varepsilon \text{ i.o.}\right\} \subseteq \left\{\left|X_{n_{m(k)}} - X\right| > \frac{1}{k} \text{ i.o.}\right\}$ for every $\varepsilon > 0$, we see that $X_{n_{m(k)}} \to X$ a.s.

To prove the converse, we first observe

Lemma 8.2. *Let $\{y_n\}_{n=1}^{\infty}$ be a sequence of elements in a topological space. If every subsequence $\{y_{n_m}\}_{m=1}^{\infty}$ has a further subsequence $\{y_{n_{m(k)}}\}_{k=1}^{\infty}$ that converges to y, then $y_n \to y$.*

Proof. If $y_n \nrightarrow y$, then there is an open set $U \ni y$ such that for every $N \in \mathbb{N}$, there is an $n \geq N$ with $y_n \notin U$, hence there is a subsequence $\{y_{n_m}\}_{m=1}^{\infty}$ with $y_{n_m} \notin U$ for all m. By construction, no subsequence of $\{y_{n_m}\}_{m=1}^{\infty}$ can converge to y, and the result follows by contraposition. □

Now if every subsequence of $\{X_n\}_{n=1}^{\infty}$ has a further subsequence that converges to X almost surely, then applying Lemma 8.2 to the sequence $y_n = P(|X_n - X| > \varepsilon)$ for an arbitrary $\varepsilon > 0$ shows that $X_n \to_p X$. □

Remark. Since there are sequences which converge in probability but not almost surely (e.g. Example 7.1), it follows from Theorem 8.1 and Lemma 8.2 that a.s. convergence does not come from a topology.
(In contrast, one of the homework problems shows that convergence in probability is metrizable.)

Theorem 8.1 can sometimes be used to upgrade results depending on almost sure convergence.

For example, you are asked to show in your homework that the assumptions in Fatou's lemma and the dominated convergence theorem can be weakened to require only convergence in probability.

To get a feel for how this works, we prove

Theorem 8.2. *If f is continuous and $X_n \to_p X$, then $f(X_n) \to_p f(X)$. If, in addition, f is bounded, then $E[f(X_n)] \to E[f(X)]$.*

Proof. If $\{X_{n_m}\}$ is a subsequence, then Theorem 8.1 guarantees the existence of a further subsequence $\{X_{n_{m(k)}}\}$ which converges to X a.s. Since limits commute with continuous functions, this means that $f(X_{n_{m(k)}}) \to f(X)$ a.s. The other direction of Theorem 8.1 now implies that $f(X_n) \to_p f(X)$.
If f is bounded as well, then the dominated convergence theorem yields $E\left[f\left(X_{n_{m(k)}}\right)\right] \to E[f(X)]$.
Applying Lemma 8.2 to the sequence $y_n = E[f(X_n)]$ establishes the second part of the theorem.
(Since f is bounded, the same argument shows that $f(X_n) \to f(X)$ in L^1.) □

We will now use the first Borel-Cantelli lemma to prove a weak form of the Strong Law of Large Numbers.

Theorem 8.3. *Let $X_1, X_2, ...$ be i.i.d. with $E[X_1] = \mu$ and $E\left[X_1^4\right] < \infty$. If $S_n = X_1 + ... + X_n$, then $\frac{1}{n}S_n \to \mu$ almost surely.*

Proof. By taking $X_i' = X_i - \mu$, we can suppose without loss of generality that $\mu = 0$. Now

$$E\left[S_n^4\right] = E\left[\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{j=1}^{n} X_j\right)\left(\sum_{k=1}^{n} X_k\right)\left(\sum_{l=1}^{n} X_l\right)\right] = E\left[\sum_{1 \leq i,j,k,l \leq n} X_i X_j X_k X_l\right].$$

By independence, terms of the form $E\left[X_i^3 X_j\right]$, $E\left[X_i^2 X_j X_k\right]$ and $E\left[X_i X_j X_k X_l\right]$ are all zero (since the expectation of the product is the product of the expectations).

The only non-vanishing terms are thus of the form $E\left[X_i^4\right]$ and $E\left[X_i^2 X_j^2\right]$, of which there are n of the former and $3n(n-1)$ of the latter (determined by the $\binom{n}{2}$ ways of picking the indices and the $2\binom{4}{2}$ ways of picking which two of the four sums gave rise to the smaller and larger indices).

Because $E\left[X_i^2X_j^2\right] = E\left[X_i^2\right]^2 \le E\left[X_i^4\right]$, we have

$$E\left[S_n^4\right] \le nE\left[X_1^4\right] + 3n(n-1)E\left[X_1^2\right]^2 \le Cn^2$$

where $C = 3E\left[X_1^4\right] < \infty$ by assumption.

It follows from Chebychev's inequality that

$$P\left(\frac{1}{n}|S_n| > \varepsilon\right) = P\left(|S_n|^4 > (n\varepsilon)^4\right) \le \frac{C}{n^2\varepsilon^4},$$

hence

$$\sum_{n=1}^{\infty} P\left(\frac{1}{n}|S_n| > \varepsilon\right) \le C\varepsilon^{-4}\sum_{n=1}^{\infty}\frac{1}{n^2} < \infty.$$

Therefore, $P\left(\frac{1}{n}|S_n| > \varepsilon \text{ i.o.}\right) = 0$ by Borel-Cantelli, so, since $\varepsilon > 0$ was arbitrary, $\frac{1}{n}S_n \to 0$ a.s. □

The converse of the Borel-Cantelli lemma is false in general:

Example 8.1. Let $\Omega = [0,1]$, $\mathcal{F} =$ Borel sets, $P =$ Lebesgue measure, and define $A_n = (0, \frac{1}{n})$. Then $\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty}\frac{1}{n} = \infty$ and $\limsup_{n\to\infty} A_n = \emptyset$.

However, if the $A_n's$ are independent, then we have

Lemma 8.3 (Borel-Cantelli II)**.** *If the events $A_1, A_2, \ldots$ are independent, then $\sum_{n=1}^{\infty} P(A_n) = \infty$ implies $P(A_n\ i.o.) = 1$.*

Proof. For each $n \in \mathbb{N}$, the sequence $B_{n,1}, B_{n,2}, \ldots$ defined by $B_{n,k} = \bigcap_{m=n}^{n+k} A_m^C$ decreases to $B_n := \bigcap_{m=n}^{\infty} A_m^C$.

Also, since the A_m's (and thus their complements) are independent, we have

$$\begin{aligned} P(B_{n,k}) &= P\left(\bigcap_{m=n}^{n+k} A_m^C\right) = \prod_{m=n}^{n+k} P\left(A_m^C\right) \\ &= \prod_{m=n}^{n+k}(1 - P(A_m)) \le \prod_{m=n}^{n+k} e^{-P(A_m)} = e^{-\sum_{m=n}^{n+k} P(A_m)} \end{aligned}$$

where the inequality is due to the Taylor series bound $e^{-x} \ge 1 - x$ for $x \in [0,1]$.

Because $\sum_{m=n}^{\infty} P(A_m) = \infty$ by assumption, it follows from continuity from above that

$$P(B_n) = \lim_{k\to\infty} P(B_{n,k}) \le \lim_{k\to\infty} e^{-\sum_{m=n}^{n+k} P(A_m)} = 0,$$

hence $P\left(\bigcup_{m=n}^{\infty} A_m\right) = P\left(B_n^C\right) = 1$ for all $n \in \mathbb{N}$.

Since $\bigcup_{m=n}^{\infty} A_m \searrow \limsup_{n\to\infty} A_n = \{A_n \text{ i.o.}\}$, another application of continuity from above gives

$$P\left(A_n \text{ i.o.}\right) = \lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} A_m\right) = 1.$$

□

Taken together, the Borel-Cantelli lemmas show that if $A_1, A_2, \ldots$ is a sequence of independent events, then the event $\{A_n \text{ i.o.}\}$ occurs either with probability 0 or probability 1.

Thus if $A_1, A_2, \ldots$ are independent, then $P(A_n \text{ i.o.}) > 0$ implies $P(A_n \text{ i.o.}) = 1$.

It follows from the second Borel-Cantelli lemma that infinitely many independent trials of a random experiment will almost surely result in infinitely many realizations of any event having positive probability.

For example, given any finite string from a finite alphabet (e.g. the complete works of Shakespeare in chronological order), an infinite string with characters chosen independently and uniformly from the alphabet (produced by the proverbial monkey at a typewriter, say) will almost surely contain infinitely many instances of said string.

Similarly, many leading cosmological theories imply the existence of infinitely many universes which may be regarded as being i.i.d. with the current state of our universe having positive probability. If any of these theories is true, then Borel-Cantelli says that there are infinitely many copies of us throughout the multiverse having this discussion!

A more serious application demonstrates the necessity of the integrability assumption in the strong law.

Theorem 8.4. *If $X_1, X_2, ...$ are i.i.d. with $E|X_1| = \infty$, then $P(|X_n| \geq n \text{ i.o.}) = 1$.*
Thus if $S_n = \sum_{i=1}^{n} X_i$, then $P\left(\lim_{n\to\infty} \frac{S_n}{n} \text{ exists in } \mathbb{R}\right) = 0$.

Proof. Lemma 7.2 and the fact that $G(x) := P(|X_1| > x)$ is nonincreasing give

$$E|X_1| = \int_0^\infty P(|X_1| > x)\,dx \leq \sum_{n=0}^{\infty} P(|X_1| > n) \leq \sum_{n=0}^{\infty} P(|X_1| \geq n).$$

Because $E|X_1| = \infty$ and the $X_n'^s$ are i.i.d., it follows from the second Borel-Cantelli lemma that $P(|X_n| \geq n \text{ i.o.}) = 1$.

To establish the second claim we will show that $C = \left\{\lim_{n\to\infty} \frac{S_n}{n} \text{ exists in } \mathbb{R}\right\}$ and $\{|X_n| \geq n \text{ i.o.}\}$ are disjoint, hence $P(|X_n| \geq n \text{ i.o.}) = 1$ implies $P(C) = 0$.

To this end, observe that

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{(n+1)S_n - n(S_n + X_{n+1})}{n(n+1)} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}.$$

Now suppose that $\omega \in C$. Then it must be the case that $\lim_{n\to\infty} \frac{S_n(\omega)}{n(n+1)} = 0$, so there is an $N \in \mathbb{N}$ with $\left|\frac{S_n(\omega)}{n(n+1)}\right| < \frac{1}{2}$ whenever $n \geq N$.

If $\omega \in \{|X_n| \geq n \text{ i.o.}\}$ as well, then there would be infinitely many $n \geq N$ with $\frac{|X_n(\omega)|}{n} \geq 1$.

But this would mean that $\left|\frac{S_n(\omega)}{n} - \frac{S_{n+1}(\omega)}{n+1}\right| = \left|\frac{S_n(\omega)}{n(n+1)} - \frac{X_{n+1}(\omega)}{n+1}\right| > \frac{1}{2}$ for infinitely many n, so that the sequence $\left\{\frac{S_n(\omega)}{n}\right\}_{n=1}^{\infty}$ is not Cauchy, contradicting $\omega \in C$. □

Our next example is a typical application where the two Borel-Cantelli lemmas are used together to obtain results on the limit superior of a (suitably scaled) sequence of i.i.d. random variables.

Example 8.2. Let $X_1, X_2, ...$ be a sequence of i.i.d. exponential random variables with rate 1 (so that $X_i \geq 0$ with $P(X_i \leq x) = 1 - e^{-x}$).

We will show that

$$\limsup_{n\to\infty} \frac{X_n}{\log(n)} = 1 \text{ a.s.}$$

First observe that

$$P\left(\frac{X_n}{\log(n)} \geq 1\right) = P\left(X_n \geq \log(n)\right) = P\left(X_n > \log(n)\right) = e^{-\log(n)} = \frac{1}{n},$$

so

$$\sum_{n=1}^{\infty} P\left(\frac{X_n}{\log(n)} \geq 1\right) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty,$$

thus, since the $X_n's$ are independent, the second Borel-Cantelli lemma implies that $P\left(\frac{X_n}{\log(n)} \geq 1 \text{ i.o.}\right) = 1$, and we conclude that $\limsup_{n\to\infty} \frac{X_n}{\log(n)} \geq 1$ almost surely.

On the other hand, for any $\varepsilon > 0$,

$$P\left(\frac{X_n}{\log(n)} \geq 1+\varepsilon\right) = P\left(X_n > (1+\varepsilon)\log(n)\right) = \frac{1}{n^{1+\varepsilon}},$$

which is summable, so it follows from the first Borel-Cantelli lemma that $P\left(\frac{X_n}{\log(n)} \geq 1+\varepsilon \text{ i.o.}\right) = 0$. Since $\varepsilon > 0$ was arbitrary, this means that $\limsup_{n\to\infty} \frac{X_n}{\log(n)} \leq 1$ almost surely, and the claim is proved.

We conclude with a cute example in which an a.s. convergence result cannot be upgraded to pointwise convergence.

Example 8.3. We will show that for any sequence of random variables $\{X_n\}_{n=1}^{\infty}$, one can find a sequence of real numbers $\{c_n\}_{n=1}^{\infty}$ such that $\dfrac{X_n}{c_n} \to 0$ a.s., but that in general, no such sequence can be found such that the convergence is pointwise.

The first statement is an easy application of the first Borel-Cantelli lemma: Given $\{X_n\}_{n=1}^{\infty}$, let $\{c_n\}_{n=1}^{\infty}$ be a sequence of positive numbers such that $P\left(|X_n| > \frac{c_n}{n}\right) \leq 2^{-n}$. Such a sequence can be found since $P\left(|X_n| > x\right) \to 0$ as $x \to \infty$. Then

$$\sum_{n=1}^{\infty} P\left(\left|\frac{X_n}{c_n}\right| > \frac{1}{n}\right) \leq 1 < \infty,$$

so for all $\varepsilon > 0$, $P\left(\left|\frac{X_n}{c_n}\right| > \varepsilon \text{ i.o.}\right) \leq P\left(\left|\frac{X_n}{c_n}\right| > \frac{1}{n} \text{ i.o.}\right) = 0$, hence $\dfrac{X_n}{c_n} \to 0$ a.s.

The interesting observation is that we cannot always choose $\{c_n\}_{n=1}^{\infty}$ so that the convergence is pointwise. To see this, let $\mathcal{C}$ denote the Cantor set. Since $\mathcal{C}$ has the cardinality of the continuum, there is a bijection $f : \mathcal{C} \to \{\{a_n\}_{n=1}^{\infty} : a_n \in \mathbb{N} \text{ for all } n\}$.

Define the random variables $\{X_n\}_{n=1}^{\infty}$ on $[0,1]$ with Borel sets and Lebesgue measure by

$$X_n(\omega) = \begin{cases} f(\omega)_n + 1, & \omega \in \mathcal{C} \\ 1, & \omega \notin \mathcal{C} \end{cases}.$$

For any sequence $\{c_n\}_{n=1}^{\infty}$, the sequence $\{\widetilde{c}_n\}_{n=1}^{\infty}$ defined by $\widetilde{c}_n = \lceil |c_n| \rceil$ is equal to $f(\omega')$ for some $\omega' \in \mathcal{C}$, hence $\left|\dfrac{X_n(\omega')}{c_n}\right| > 1$ for all n, so there is no sequence of reals for which the convergence is sure.