1.) First Borel-Cantelli Lemma

We begin with some notation. If \( A_n, n \geq 1 \) is a sequence of subsets of \( \Omega \), then we define
\[
\limsup_{n \to \infty} A_n \equiv \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{ \omega \text{ that are in infinitely many } A_n \} \\
\liminf_{n \to \infty} A_n \equiv \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = \{ \omega \text{ that are in all but finitely many } A_n \}.
\]

It is common to write \( \limsup A_n = \{ \omega : \omega \in A_n \text{ i.o.} \} \). For example, \( X_n \to X \text{ a.s.} \) if and only if for all \( \epsilon > 0 \), \( P(|X_n - X| > \epsilon \text{ i.o.}) = 0 \). Our first key result is:

**Lemma (1.6.1): First Borel-Cantelli Lemma.** If \( \sum_{n=1}^{\infty} P(A_n) < \infty \), then \( P(A_n \text{ i.o.}) = 0 \).

**Proof:** Let \( N = \sum_k 1_{A_k} \) be the number of events that occur. By Fubini's theorem, \( E(N) = \sum_k P(A_k) < \infty \), so we must have \( N < \infty \text{ a.s.} \). □

Before we apply this, we first recall the following result from topology.

**Lemma (1.6.3):** \( y_n \to y \) if and only if every subsequence \( (y_{nm}; m \geq 1) \) contains a further subsequence \( (y_{nm(k)}; k \geq 1) \) that converges to \( y \).

**Theorem (1.6.2):** \( X_n \to X \) in probability if and only if every subsequence \( (X_{nm}; m \geq 1) \) contains a further subsequence \( (X_{nm(k)}; k \geq 1) \) that converges almost surely to \( X \).

**Proof:** Suppose that \( X_n \to X \) in probability and let \( \epsilon_k \downarrow 0 \). Given any sequence \( (n_m : m \geq 1) \), for each \( k \) there is an \( n_m(k) > n_{m(k-1)} \) such that \( P(|X_{nm(k)} - X| > \epsilon_k) \leq 2^{-k} \). Since
\[
\sum_{k=1}^{\infty} P(|X_{nm(k)} - X| > \epsilon_k) < \infty,
\]
the first Borel-Cantelli lemma implies that \( P(|X_{nm(k)} - X| > \epsilon_k \text{ i.o.}) = 0 \). Furthermore, since \( \epsilon_k \downarrow 0 \), this implies that \( P(|X_{nm(k)} - X| > \epsilon \text{ i.o.}) = 0 \) for every \( \epsilon > 0 \) and so \( X_{nm(k)} \to X \) almost surely.

To prove the converse, fix \( \delta > 0 \) and let \( y_n = P(|X_n - X| > \delta) \). We need to show that \( y_n \to 0 \). Given any subsequence \( (y_{nm}; m \geq 1) \), by assumption we know that the subsequence \( (X_{nm}; m \geq 1) \) contains a subsequence \( (X_{nm(k)}; k \geq 1) \) that converges to \( X \) almost surely. This implies that \( (y_{nm(k)}; k \geq 0) \) converges to 0 and so Lemma (1.6.3) implies that \( y_n \to 0 \). Since this holds for every \( \delta > 0 \), it follows that \( X_n \to X \) in probability. □
Corollary (1.6.4): If $f$ is continuous and $X_n \to X$ in probability, then $f(X_n) \to f(X)$ in probability. If $f$ is also bounded, then $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$.

Proof: If $X_{n(m)}$ is a subsequence, then Theorem (1.6.2) implies that there is a further subsequence $X_{n(m_k)} \to X$ almost surely. Since $f$ is continuous, $f(X_{n(m_k)}) \to f(X)$ almost surely and so (1.6.2) now implies that $f(X_n) \to f(X)$ in probability.

If $f$ is bounded, then the bounded convergence theorem implies that $\mathbb{E}f(X_{n(m_k)}) \to \mathbb{E}f(X)$ and the desired result follows from Lemma (1.6.3) with $y_k = \mathbb{E}f(X_k)$. □

2.) The Second Borel-Cantelli Lemma

The following example shows that the converse of the first Borel-Cantelli Lemma is false. Let $\Omega = (0, 1)$, $\mathcal{F} =$ Borel sets, $\mathbb{P} =$ Lebesgue measure. If $A_n = (0, 1/n)$, then $\lim \sup A_n = \emptyset$ even though $\sum_n \mathbb{P}(A_n) = \sum n^{-1} = \infty$.

Theorem (1.6.6): The Second Borel-Cantelli Lemma: If the events $A_n$ are independent and $\sum \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(\lim \sup A_n) = 1$.

Proof: Let $M < N < \infty$. Using the inequality $1 - x \leq e^{-x}$ whenever $x \geq 0$, we have

$$\mathbb{P}\left(\bigcap_{n=M}^{N} A_n^c \right) = \prod_{n=M}^{N} (1 - \mathbb{P}(A_n)) \leq \prod_{n=M}^{N} \exp(-\mathbb{P}(A_n))$$

$$= \exp\left(-\sum_{n=M}^{N} \mathbb{P}(A_n)\right) \to 0.$$ 

So $\mathbb{P}(\bigcup_{n=M}^{N} A_n) = 1$ for every $M$. Since $\bigcup_{n=M}^{\infty} A_n \downarrow \lim \sup A_n$, it follows that $\mathbb{P}(\lim \sup A_n) = 1$. □

A typical application for the second B-C lemma is:

Theorem (1.6.7): If $X_1, X_2, \ldots$ are i.i.d. with $\mathbb{E}|X_1| = \infty$, then $\mathbb{P}(|X_n| \geq n \text{ i.o.}) = 1$. Consequently, if $S_n = X_1 + \cdots + X_n$, then $\mathbb{P}(\lim S_n/n \in (-\infty, \infty)) = 0$.

Proof: Since

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

the fact that $\mathbb{E}|X_1| = \infty$ along with the Second Borel-Cantelli lemma imply that $\mathbb{P}(|X_n| \geq n \text{ i.o.}) = 1$. To prove the second claim, let $C = \{\omega : \lim_{n \to \infty} S_n(\omega)/n \in (-\infty, \infty)\}$ and notice that

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}.$$
Since \( S_n/(n(n+1)) \to 0 \) on \( C \), it follows that on the set \( C \cap \{ \omega : |X_n| \geq n, \text{i.o.} \} \) we have
\[
\left| \frac{S_n}{n} - \frac{S_{n+1}}{n+1} \right| > \frac{2}{3} \text{ i.o.}
\]
contradicting the fact that \( \omega \in C \). This shows that
\[
C \cap \{ \omega : |X_n| \geq n, \text{i.o.} \} = \emptyset
\]
and so \( P(C) = 0 \). \( \square \)

With additional work, convergence rates can be estimated for the second Borel-Cantelli lemma.

**Theorem (1.6.8):** If \( A_1, A_2, \ldots \) are pairwise independent and \( \sum_{n=1}^{\infty} P(A_n) = \infty \), then
\[
\sum_{m=1}^{n} \frac{1_{A_m}}{\sum_{m=1}^{n} P(A_m)} \to 1 \text{ a.s.}
\]

**Proof:** Let \( X_m = 1_{A_m} \) and \( S_n = X_1 + \cdots + X_n \) and notice that \( \text{Var}(X_m) \leq \mathbb{E}X_m^2 = \mathbb{E}X_m \). Also, since the \( A_m \) are pairwise independent, the \( X_m \) are uncorrelated and so
\[
\text{Var}(S_n) = \sum_{m=1}^{n} \text{Var}(X_m) \leq \sum_{m=1}^{n} \mathbb{E}X_m = \mathbb{E}S_n.
\]
It then follows from Chebyshev’s inequality and our assumption that \( \mathbb{E}S_n \to \infty \) that
\[
(*) \quad P(\left| S_n - \mathbb{E}S_n \right| > \delta \cdot \mathbb{E}S_n) \leq \text{Var}(S_n)/(\delta \cdot \mathbb{E}S_n)^2 \to 0
\]
as \( n \to \infty \). This shows that \( S_n/\mathbb{E}S_n \to 1 \) in probability.

To get almost sure convergence, we need to consider subsequences. Let \( n_k = \inf \{ n : \mathbb{E}S_n \geq k^2 \} \). Let \( T_k = S_{n_k} \) and observe that \( k^2 \leq \mathbb{E}T_k \leq k^2 + 1 \). Replacing \( n \) by \( n_k \) in (*) and using \( \mathbb{E}T_k \geq k^2 \) shows that
\[
P(\left| T_k - \mathbb{E}T_k \right| > \delta \mathbb{E}T_k) \leq 1/(\delta^2 k^2).
\]
So \( \sum_k P(\left| T_k - \mathbb{E}T_k \right| > \delta \mathbb{E}T_k) < \infty \) and thus the first Borel-Cantelli lemma implies that \( P(\left| T_k - \mathbb{E}T_k \right| > \delta \cdot \mathbb{E}T_k \text{ i.o.}) = 0 \). Since \( \delta > 0 \) is arbitrary, it follows that \( T_k/\mathbb{E}T_k \to 1 \) almost surely.

To show that \( S_n/\mathbb{E}S_n \to 1 \) almost surely, choose \( \omega \) such that \( T_k(\omega)/\mathbb{E}T_k \to 1 \) and observe that if \( n_k \leq n \leq n_{k+1} \) then
\[
\frac{T_k(\omega)}{\mathbb{E}T_{k+1}} \leq \frac{S_n(\omega)}{\mathbb{E}S_n} \leq \frac{T_{k+1}(\omega)}{\mathbb{E}T_k}.
\]
However, these inequalities can rewritten as
\[
\frac{\mathbb{E}T_k}{\mathbb{E}T_{k+1}} \cdot \frac{T_k(\omega)}{\mathbb{E}T_k} \leq \frac{S_n(\omega)}{\mathbb{E}S_n} \leq \frac{\mathbb{E}T_{k+1}}{\mathbb{E}T_k} \cdot \frac{T_{k+1}(\omega)}{\mathbb{E}T_{k+1}}
\]
and the desired result follows upon recognizing that \( \mathbb{E}T_{k+1}/\mathbb{E}T_k \to 1 \). \( \square \)
**Example: Record Values.** Let $X_1, X_2, \cdots$ be i.i.d. with continuous distribution function $F(x)$ and let $A_k = \{X_k > \sup_{j<k} X_j\}$ be the event that a record value occurs on the $k$’th trial. We will show that the $A_k$ are pairwise independent with $P(A_k) = 1/k$.

We first observe that because $F$ is continuous, $P(X_j = X_k) = 0$ for any $j \neq k$, so we can let $Y_1^n > Y_2^n > \cdots > Y_n^n$ be the (decreasing) order statistics of $X_1, \cdots, X_n$. This induces a random permutation of $\{1, \cdots, n\}$ defined by $\pi_n(i) = j$ if $X_i = Y_j^n$. Furthermore, because the joint distribution of the random variables $(X_1, \cdots, X_n)$ is not affected by the order of occurrence, it follows that $\pi_n$ is uniformly distributed over the set of $n!$ possible permutations.

To show pairwise independence, let $C_{n,i}$ be the set of permutations of $\{1, \cdots, n\}$ with the property that $\pi(i) < \min\{\pi(1), \cdots, \pi(i-1)\}$ and notice that

$$A_i = \{\pi_n \in C_{n,i}\}.$$ 

Since $\pi_n$ is uniformly distributed, it follows that

$$P(A_i) = \frac{|C_{n,i}|}{n!} = \frac{n(n-1)\cdots(i+1)(i-1)\cdots2\cdot1}{n!} = \frac{1}{i}.$$

Similarly, if $i < j$, then

$$P(A_i \cap A_j) = \frac{|C_{n,i} \cap C_{n,j}|}{n!} = \frac{n(n-1)\cdots(j+1)(j-1)\cdots(i+1)(i-1)\cdots2\cdot1}{n!} = \frac{1}{ij} = P(A_i)P(A_j)$$

and so $A_i$ and $A_j$ are independent.

Let $R_n = \sum_{m=1}^n 1_{A_m}$ be the number of record values at time $n$. Since $\sum_{m=1}^n P(A_m) \sim \log(n) \to \infty$, Theorem (1.6.8) implies that as $n \to \infty$

$$R_n / \log(n) \to 1 \text{ a.s.}$$