0.0.1 Borel-Cantelli lemmas
Let $A_n$, $n \in \mathbb{N}^+$ be events and $B (= A_n$ i.o., infinitely often) is the event that infinitely many among \{ $A_n$ \} occur:

$$B = \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m$$

(1)

Notice that $C_n = \bigcup_{m \geq n} A_m$ is the event that at least one event $A_m$ with $m \geq n$ occurs. Hence, $B = \bigcap_{n=1}^{\infty} C_n$ implies that $B \supset \{ \text{all } A_m \text{ occur.} \}$

(1) If $\sum P(A_n) < \infty$, $P(B) = 0$. That is, if $\sum P(A_n) < \infty$ infinitely many $A_n$ cannot occur simultaneously.

[Demo]  

$$P(B) \leq P(\bigcup_{m \geq n} A_m) \leq \sum_{m \geq n} P(A_m)$$  

(2)  

holds for any $n$. Since $\sum_{m=1}^{\infty} P(A_m)$ converges, the RHS must vanish in the large $n$ limit.

We could intuitively write $B = \bigcup_{K \in \mathbb{N}^+} A_{K_m}$, where $K = \{ K_m \}$ denotes an infinite subset of $\mathbb{N}^+$. Therefore,  

$$P(B) \leq \sum_{K} P(\bigcap_{K_m} A_{K_m}).$$  

(3)  

Each summand on the RHS is zero.

(2) If $A_n$ are independent, and if $\sum P(A_n) = \infty$, then $P(B) = 1$. That is, if for independent event $\sum P(A_n) = \infty$, then infinitely many $A_n$ simultaneously occur for sure.

[Demo] We consider  

$$B^c = \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_{m}^c$$

(4)  

and wish to show $P(B^c) = 0$. We see  

$$P(B^c) \leq \sum_{n=1}^{\infty} P(\bigcap_{m \geq n} A_{m}^c).$$  

(5)  

and thanks to independence  

$$P(\bigcap_{m \geq n} A_{m}^c) = \prod_{m \geq n} P(A_{m}^c) = \prod_{m \geq n} (1 - P(A_m)).$$  

(6)  

Notice that $1 - x \leq e^{-x}$. Therefore,  

$$P(\bigcap_{m \geq n} A_{m}^c) \leq \exp \left( - \sum_{m \geq n} P(A_m) \right).$$  

(7)  

This is zero, because $\sum_{m \geq n} P(A_m) = \infty$ for any $n$.

0.0.2 Weak law vs strong law
This is as usual, and can be proved by Chebyshev.
The reason why it is called ‘weak’ is: Let $Z_N = \{m - \varepsilon \leq S_N/N \leq m + \varepsilon\}$. The weak law says for any finite $n$ $P(Z_n)$ is positive. Therefore, the empirical expectation can wander for any $n$ away from $m$ significantly.

In contrast, the strong law implies $P(Z_n) = 1$ beyond some $n$.

In more detail, for any given $\varepsilon > 0$, there is $N$ such that $S_n/n \in (m - \varepsilon, m + \varepsilon)$ for $\forall n > N$. The weak law can get out of this interval for any $n$ with probability as large as $\sim 1/n^2$

In the weak case, it is the probability wrt $\Omega$, not for individual $\omega$.

**0.0.3 Weak law of large numbers for Bernoulli process**

The weak law:

For any $\varepsilon > 0$ \( \lim_{n \to \infty} P(|S_N/N - m| > \varepsilon) = 0 \) for a Bernoulli process, where $m$ is the expectation value. Chebyshev tells us that

\[
P(|S_N'/N| > \varepsilon) \leq E(S_N'^2)/\varepsilon^2N^2 = pq/\varepsilon^2N.
\]  

(8)

**0.0.4 Strong law of large numbers for Bernoulli process**

What we wish to show is that beyond some $n$ the probability for $|S_N'/N|$ to be large is zero. Therefore, we have only to show that $|S_N'/N|$ is large only for finitely many $N$’s. We should replace $\varepsilon$ with some $n$ dependent converging sequence to zero. We should use Borel-Cantelli (1) 0.0.1, but then the required sum does not converge. Therefore, we use the following trick:

First, we study the subsequence $S'_1, S'_4, S'_9, \ldots S'_{n^2}, \ldots$. (9) reads

\[
P(|S'_n/n^2| > \varepsilon_n) \leq pq/\varepsilon_n^2n^2.
\]  

Then, we can choose $\varepsilon_n$ to be decreasing, say, $\varepsilon_n = n^{-1/3}$:

\[
P(|S'_n/n^2| > n^{-1/3}) \leq pq/n^{-4/3}.
\]  

(10)

Thus, there is an $N$ beyond which $P(|S'_n/n^2| < pqn^{-1/3}) = 1$.

We must fill $n^2$ and $(n + 1)^2$. Take $n^2 < m < (n + 1)^2$. Obviously,

\[
|S'_m| \leq |S'_{n^2}| + (n + 1)^2 - n^2 = |S'_{n^2}| + 2n + 1,
\]  

(11)

but for sufficiently large $n$ (10) implies $|S'_{n^2}| < n^{5/3}$, so

\[
|S'_m| \leq n^{5/3} + 2n + 1,
\]  

(12)

We know $n < \sqrt{m}$, so

\[
|S'_m| \leq m^{5/6} + 2m^{1/2} + 1.
\]  

(13)

$m^{5/6} > 2\sqrt{m} + 1$ for large $m$ (obviously correct for $m \geq 3$):

\[
|S'_m/m| \leq 2m^{-1/6}
\]  

(14)

with probability 1 for sufficiently large $m$.
0.0.5 Strong law of large numbers under finite variance condition

This is easy to prove by mimicking the Bernoulli case.

0.0.6 Strong law of large numbers under \( E(|X|) < \infty \) (Kolmogorov 1933)

Kolmogorov’s strategy is to ‘regularize’ \( X_n \) first.

Let \( X_n^\pm = \max(\pm X_n, 0) \). Then, \( X = X^+ - X^- \). Since \(|X| = X^+ + X^-\), \( X_n^\pm \) are statistically independent sequences with finite averages. If we can prove the strong law for \( X^\pm \), it is also true for the original \( X \). Therefore, we may assume that \( X_n \geq 0 \).

\( X_n \) could be very large, but for sufficiently large \( n \) we may expect \( X_n < n \) is true for most \( n \). Indeed, as we see shortly, except for finitely many \( n \) \( X_n < n \). Therefore, we may assume \( 0 \leq X_n \leq n \) for all \( n \) without any loss of generality. We wish to use Borel-Cantelli 0.0.1(1):

\[
\sum_{n=1}^\infty P(X_n > n) = \sum_{n=1}^\infty \sum_{k=n}^\infty P(k \leq X_n < k + 1) = \sum_{k=1}^\infty \sum_{n=1}^k P(k \leq X_n < k + 1) \tag{15}
\]

Since \( P(k \leq X_n < k + 1) \) does not depend on \( n \), we have

\[
\sum_{n=1}^\infty P(X_n > n) = \sum_{k=1}^\infty kP(k \leq X_n < k + 1) < \sum_{k=1}^\infty k < E(X) < \infty. \tag{16}
\]

Therefore, \( X_n > n \) is true only for finitely many \( n \).

We may also subtract \( \langle X \rangle \) from \( X \), we may assume \( E(X) = 0 \).

What we wish to show is

\[
\frac{1}{N} \sum_{n=1}^N X_n \tag{17}
\]

converges. A sufficient condition is the convergence of

\[
\sum_{n=1}^\infty \frac{1}{n} X_n. \tag{18}
\]

This is Kronecker’s lemma 0.0.7. To show (18) we use Kolmogorov’s inequality 0.0.8, for \( \lambda > 0 \),

\[
P \left( \max_{k \leq m} (x_n/n + \cdots + x_k/k) > \lambda \right) \leq \frac{1}{\lambda^2} \sum_{k=n}^m \frac{1}{k^2} E(X_k^2) \tag{19}
\]
and the convergence of
\[ \sum_{k=1}^{\infty} \frac{1}{k^2} E(X_k^2). \] (20)

To show this convergence, note that
\[ \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{y=1}^{\infty} E(X, X \in [y-1, y)) \leq \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{y=1}^{\infty} y E(X, X \in [y-1, y)) \] (21)

Now, we exchange the order of the summations as
\[ \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{y=1}^{\infty} y E(X, X \in [y-1, y)) = \sum_{y=1}^{\infty} \left( \sum_{k=1}^{y} \frac{1}{k^2} \sum_{k=1}^{y} E(X, X \in [y-1, y]) \right) \] (22)

Notice that \( \sum_{y=1}^{\infty} \frac{1}{k^2} \sum_{y=1}^{\infty} y E(X, X \in [y-1, y)) \leq \sum_{y=1}^{\infty} \frac{1}{k^2} \sum_{y=1}^{\infty} y E(X, X \in [y-1, y)) = E(X) \) which is finite.

0.0.7 Kronecker's lemma
If \( \sum_{n=1}^{\infty} c_n/n \) converges, then \( (1/n) \sum_{i=1}^{n} c_n \rightarrow 0 \).
[Demo] Let \( c'_n = \sum_{k=1}^{n} c_k/k \) with \( c'_0 = 0 \). Then,
\[ \sum_{k=1}^{n} c_k = \sum_{k=1}^{n} \left( k(c'_k - c'_{k-1}) \right) = \sum_{k=1}^{n} kc'_k - \sum_{k=1}^{n} kc'_{k-1} = \sum_{k=1}^{n} kc'_k - \sum_{k=1}^{n} (k+1)c'_k = n c'_n - \sum_{k=1}^{n} c'_k. \] (24)

Therefore,
\[ \frac{1}{n} \sum_{i=1}^{n} c_n = c'_n - \frac{1}{n} \sum_{k=1}^{n} c'_k \] (25)

Let \( c'_n \rightarrow c_\infty \). Then
\[ \left| c'_\infty - \frac{1}{n} \sum_{k=1}^{n} c'_k \right| \leq \frac{1}{n} \sum_{k=1}^{n} |c'_k - c'_\infty| \] (26)

Since \( c'_k \) converges, for any \( \varepsilon > 0 \) there is \( N \) such that for \( n > N \) \( |c'_k - c'_\infty| < \varepsilon \). Therefore, (25) vanishes in the large \( n \) limit.

0.0.8 Kolmogorov's inequality
Let \( x_n \) be independent, \( E(X_n) = 0 \) and \( E(X_n^2) < \infty \).
\[ P \left( \max_{k \leq n} (x_1 + x_2 + \cdots + x_k) > \lambda \right) \leq \frac{1}{\lambda^2} \sum_{k=1}^{n} E(X_k^2) \] (27)
[Demo] The event $A = \{ \omega : \max_k |S_k| \geq c \}$ is the same as the event that there is $k$ such that $|S_k| \geq c$. Therefore, it can be partitioned as $A = \bigcup A_k$, where $A_k = \{ \omega : |S_k| \geq c, |S_1|, \ldots, |S_{k-1}| \}$.

What we should prove is

$$c^2 P(A) \leq \sum V(X_n). \quad (28)$$

We have, since $E(X_k) = 0$ and all $X_k$ are statistically independent,

$$\sum V(X_n) = E(S_n^2) \geq E(S_n^2 \chi_A) = \sum E(S_n^2 \chi_{A_k}). \quad (29)$$

We know $S_n - S_j$ and $S_j \chi_{A_j}$ are statistically independent,

$$E(S_n^2 \chi_{A_k}) = E((S_n - S_k + S_k)^2 \chi_{A_k}) \quad (30)$$
$$= E((S_n - S_k)^2 \chi_{A_k}) + 2E((S_n - S_k)S_k \chi_{A_k}) + E(S_k^2 \chi_{A_k}) \quad (31)$$
$$\geq E(S_k^2 \chi_{A_k}) \geq c^2 E(\chi_{A_k}) = c^2 P(A_k), \quad (32)$$

where we have used $E((S_n - S_k)S_k \chi_{A_k}) = E(S_n - S_k)E(S_k \chi_{A_k}) = 0$. This implies (29).