Chapter 8

Sums of independent random variables

This is based on [Billingsley, Section 22]

1. The strong law of large numbers

**Theorem 8.1** (Etemadi). If \(X_1, X_2, \ldots\) are pairwise independent identically distributed integrable random variables with mean \(m\) then \(\frac{1}{n}S_n \to m\) with probability one.

For a completely elementary (but not simple) proof of this result is [Billingsley, Theorem 22.1]. Instead we will exhibit several proof techniques which can be applied to prove laws of large numbers under various sets of assumptions. The proof techniques rely on Markov inequality, Borel-Cantelli lemma, decomposition into positive/negative parts, and truncation.

The following is known as a weak law of large numbers, and the proof is an exercise.

**Theorem 8.2.** If \(X_1, X_2, \ldots\) are pairwise independent with the same mean and uniformly bounded second moments, then \(\frac{1}{n}S_n \xrightarrow{P} m\).

**Proof.** Compute \(Var\frac{1}{n}S_n\).

**Theorem 8.3.** If \(X_1, X_2, \ldots\) are quadruple-independent with the same mean \(m\) and with uniformly bounded fourth moments then \(\frac{1}{n}S_n \to m\) with probability one.

**Proof.** This proof assumes that \(X_1, X_2, \ldots\) have the same distribution. You should figure out what needs to be modified if the distributions are not the same!!

Without loss of generality we can assume \(m = 0\). (Replace \(X_n\) by \(X_n - m\).) We will use Borel-Cantelli lemma to verify that for every \(\varepsilon > 0\), \(P(\frac{1}{n}|S_n| \geq \varepsilon \ i.o.) = 0\). To do so, we use Markov’s inequality,

\[
P\left(\frac{1}{n}|S_n| \geq \varepsilon\right) \leq \frac{E[S_n^4]}{\varepsilon^4 n^4}
\]
We note that by Laypunov (or Cauchy-Schwartz) inequality \( E(X_j^2) \leq \sqrt{EX_j^4} < \infty \), so for identically distributed random variables

\[
\mathbb{E}[S_n^4] = \sum_{j_1,j_2,j_3,j_4=1}^n \mathbb{E}[X_{j_1}X_{j_2}X_{j_3}X_{j_4}] = n\mathbb{E}(X_1^4) + 3n(n-1)(\mathbb{E}[X_1^2])^2 \leq Cn^2
\]

Thus \( \sum_n P\left( \frac{1}{n}|S_n| \geq \varepsilon \right) < \infty \). By Borel-Cantelli (Theorem 3.6) \( P\left( \frac{1}{n}|S_n| > \varepsilon \ i.o. \right) = 0 \), ending the proof. (See discussion of convergence with probability one in the proof of Proposition 4.11.) \( \square \)

**Theorem 8.4.** If \( X_1, X_2, \ldots \) are pairwise independent identically distributed square-integrable random variables with mean \( m \) then \( \frac{1}{n}S_n \to m \) with probability one.

**Proof.** The proof of Theorem 8.2 shows that \( \frac{1}{n^2}S_n^2 \to m \) with probability one. Writing \( X_j = X_j^+ - X_j^- \), when random variables have the same distribution, without loss of generality we may assume \( X_j \geq 0 \). Then for \( n \in \mathbb{N} \) choose \( k = k(n) \) such that \( k^2 \leq n < (k+1)^2 \). Notice that \( n/k^2 \to 1 \).

Since we know that \( \frac{1}{k^2}S_k^2 \to m \) with probability one, we have

\[
\frac{S_k^2}{n} \leq \frac{S_n}{n} \leq \frac{S_{(k+1)^2}}{n}
\]

And

\[
\frac{S_k^2}{n} = \frac{S_k^2}{k^2} \frac{k^2}{n} \to m, \quad \frac{S_{(k+1)^2}}{n} = \frac{S_{(k+1)^2}}{(k+1)^2} \frac{(k+1)^2}{n} \to m
\]

So by the squeeze theorem, \( \frac{S_n}{n} \to m \). \( \square \)

One can reduce moment assumptions by a suitable use of truncation.

---

(Omitted in 2018)

The following result assumes less than Theorem 8.3, but more than Theorem 8.1. It is harder to prove, but still easier to prove than Theorem 8.1.

**Theorem 8.5.** If \( X_1, X_2, \ldots \) are pairwise independent identically distributed random variables with mean \( m \), and \( E(|X_1|^{3/2}) < \infty \), then \( \frac{1}{n}S_n \to m \) with probability one.

**Proof.** The main steps in the proof are the same as in the proof of Theorem 8.1: are reduction to non-negative case, truncation, Borel-Cantelli lemma, and the use of subsequences. But the technicalities are somewhat simpler due to the stronger moment assumption.
1. The strong law of large numbers

Since \( m = m^+ - m^- \) and \( X_j = X_j^+ - X_j^- \) without loss of generality we can assume that \( X_j \geq 0 \).

Next, we introduce truncation \( X_k = X'_k + X''_k \)

where \( X'_k = X_k I_{X_k \leq k} \) and denote \( S'_n = \sum_{k=1}^{n} X'_k \).

Since \( X_k \) are identically distributed and integrable,

\[
\sum_{k=1}^{\infty} P(X_k \neq X'_k) = \sum_{k} P(|X_k| > k) = \sum_{k} P(|X_1| > k) < \int_{0}^{\infty} P(|X_1| > t) dt = E(X_1) < \infty
\]

Therefore by the Borel-Cantelli Lemma \( P(X_k \neq X'_k \text{ i.o.}) = 0 \). This implies that \( \frac{1}{n} S_n - \frac{1}{n} S'_n \to 0 \) with probability one.

\[\text{(Omitted in 2018)}\]

Next we compute

\[
\text{(8.1)} \quad \text{Var}(S'_n) = \sum_{k=1}^{n} \text{Var}(X'_k) \leq \sum_{k=1}^{n} E \left( X'^2_k I_{|X_k| \leq k} \right) = \sum_{k=1}^{n} E \left( X'^2_k I_{|X_k| \leq k} \right) \leq nE(X'^2_1 I_{|X| \leq n}).
\]

We apply inequality (8.1) to subsequence \( \frac{1}{n^2} S''_{n^2} \). We get

\[
\sum_{n=1}^{\infty} \text{Var}(S''_{n^2}/n^2) \leq \sum_{n=1}^{\infty} \frac{n^2}{n^4} E(X'^2_1 I_{|X_1| \leq n^2}) = E \left( \frac{|X_1|^2}{n^2} \sum_{n^2 \geq |X_1|} \frac{1}{n^2} \right) \leq C + \int_{|X_1| \geq 4} \left( \frac{|X_1|^2}{n^2} \sum_{n^2 \geq |X_1|} \frac{1}{n^2} \right) dP \leq C + \frac{2}{3} E(|X_1|^{3/2}).
\]

Here we use the bound \( \frac{\sqrt{x}}{\sqrt{2n-1}} \leq \frac{\sqrt{x}}{\sqrt{4-1}} = 2 \) for \( x \geq 4 \). Therefore, Chebyshev’s inequality implies that

\[
\sum_{n} P\left( \frac{1}{n^2} |S''_{n^2} - E(S''_{n^2})| > \varepsilon \right) < \infty
\]

By another application of Borel-Cantelli, we see that \( \frac{1}{n^2} \left( S''_{n^2} - E(S''_{n^2}) \right) \to 0 \) with probability one.

Now we note that \( \frac{1}{n} E(S'_n) = \frac{1}{n} \sum_{k=1}^{n} m_k \) where \( m_k = E(X_1 I_{|X_1| \leq k}) \to m \) by Lebesgue’s dominated convergence theorem (Theorem 6.9). By Cesaro’s theorem \( 1 \) the sequence \( \frac{1}{n} E(S'_n) \) also converges to \( m \).
To conclude the proof, we write

\[(8.2) \quad \frac{1}{n}S_n - m = \frac{1}{n}(S_n - S_n') + \frac{1}{n}(S_n' - E(S_n')) + \left(\frac{1}{n} \sum_{k=1}^{n} m_k - m\right)\]

From (8.2) we therefore deduce that \(\frac{1}{n^2}S_n^2 \to m\) with probability one.

To conclude the proof, we note that every number \(n\) lies between two perfect squares. That is, we introduce the sequence \(k_n = \lfloor \sqrt{n} \rfloor\) so that \(k_n^2 \leq n \leq (k_n + 1)^2\).

Notice that \(k_n = k_{n+1}\) is possible, but \(k_n \to \infty\) eventually.

By the previous part of the proof, we know that \(\frac{1}{k_n^2}S_{k_n^2} \to m\) and \(\frac{1}{(k_n + 1)^2}S_{(k_n + 1)^2} \to m\) with probability one.

We now use \(X_j \geq 0\) to infer that

\[\frac{1}{n}S_n \leq S_{(k_n + 1)^2}\]

so \(\frac{1}{n}S_n\) is between

\[\frac{k_n^2}{(k_n + 1)^2} \frac{1}{k_n^2}S_{k_n^2} \text{ and } \frac{(k_n + 1)^2}{k_n^2} \frac{1}{(k_n + 1)^2}S_{(k_n + 1)^2}\]

Since \(\frac{k}{k+1} \to 1\) as \(k \to \infty\), the result follows.

\[\square\]

2. Kolmogorov’s zero-one law

(Omitted in 2018)

**Theorem 8.6.** Suppose \(X_1, X_2, \ldots\) are independent and \(A \in \mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \ldots)\). Then \(P(A) = 0\) or \(P(A) = 1\).

**Proof.** Consider \(\mathcal{F}_0 = \bigcup_{k=1}^{\infty} \sigma(X_1, X_2, \ldots, X_k)\). We first check that \(\mathcal{F}_0\) is a field, and that it generates \(\sigma(X_1, X_2, \ldots)\).

Next suppose \(A \in \mathcal{T}\). Then \(A \in \sigma(X_{k+1}, \ldots)\) for every \(k\), so \(A\) is independent of \(B \in \sigma(X_1, \ldots, X_k)\). So \(A\) is independent of \(\mathcal{F}_0\). Since \(\mathcal{F}_0\) is a \(\pi\)-system, hence \(A\) is independent of \(\sigma(\mathcal{F}_0)\). But that means that \(P(A)P(A) = P(A)\). \(\square\)
3. Kolmogorov’s Maximal inequality and its applications

**Theorem 8.7** (Kolmogorov’s maximal inequality). Suppose $X_1, X_2, \ldots, X_n$ are independent with mean 0 and finite variance. For $t > 0$,

\begin{equation}
P(\max_{1 \leq k \leq n} |S_k| > t) \leq \frac{\text{Var}(S_n)}{t^2}.
\end{equation}

**Proof.** We have a sequence of random variables $|S_1|, |S_2|, \ldots, |S_n|$ and we want to estimate the probability that at least one of them exceeds level $t$. The trivial estimate $P(\max_{1 \leq k \leq n} |S_k| > t) \leq \sum_{k=1}^{n} P(|S_k| > t)$ is not accurate enough due to multiple overlaps, so we must decompose the event \{max_{1 \leq k \leq n} |S_k| > t\} more carefully into disjoint sets. The trick is to look at where the first crossing of level $t$ occurs. Consider the (disjoint) events

$$A_k = \{|S_1| < t, \ldots, |S_{k-1}| < t, |S_k| \geq t\}.$$ 

Clearly, $P(\max_{1 \leq k \leq n} |S_k| > t) = \sum_{k=1}^{n} P(A_k)$.

Since the events are disjoint, we have

$$E(S_n^2) \geq \int_{\max_{1 \leq k \leq n} |S_k| \geq t} S_k^2 dP = \sum_{k=1}^{n} \int_{A_k} S_k^2 dP$$

$$= \sum_{k=1}^{n} \int_{A_k} (S_n - S_k + S_k)^2 dP = \sum_{k=1}^{n} \int_{A_k} ((S_n - S_k)^2 + S_k^2) dP + 2 \sum_{k=1}^{n} \int_{A_k} (S_n - S_k) S_k dP$$

$$\geq \sum_{k=1}^{n} \int_{A_k} S_k^2 dP + 2 \sum_{k=1}^{n} \int_{A_k} (S_n - S_k) S_k dP$$

$$\geq t^2 \sum_{k=1}^{n} P(A_k) + 2 \sum_{k=1}^{n} \int_{A_k} (S_n - S_k) S_k dP = t^2 P(\max_{1 \leq k \leq n} |S_k| \geq t) + 2 \sum_{k=1}^{n} \int_{A_k} (S_n - S_k) S_k dP$$

This is what we want, if we can show that the last sum vanishes. And this is indeed the case:

$$\int_{A_k} (S_n - S_k) S_k dP = \int_{\Omega} (S_n - S_k) S_k I_{A_k} dP = E(S_n - S_k) E(S_k I_{A_k}) = 0$$

as the event $A_k$ is $\sigma(X_1, \ldots, X_N)$-measurable, $S_k$ is also $\sigma(X_1, \ldots, X_k)$-measurable, but $(S_n - S_k) = X_{k+1} + \ldots + X_n$ is $\sigma(X_{k+1}, X_{k+2}, \ldots, X_n)$ -measurable. This shows that random variables $U = (S_n - S_k)$ and $V = S_k I_{A_k}$ are independent, and we can apply the formula $E(UV) = E(U)E(V)$. \hfill \Box

**Theorem 8.8** (Kolmogorov’s one-series theorem). Suppose that $\{X_n\}$ is an independent sequence, $E(X_n) = 0$ and $\sum_{n=1}^{\infty} \text{Var}(X_n) < \infty$. Then the series $\sum_{n=1}^{\infty} X_n$ converges with probability one.

**Proof.** Denote $S_n = \sum_{k=1}^{n} X_k$ From (8.3) we have

$$P(\max_{1 \leq k \leq r} |S_{n+k} - S_n| > \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{k=1}^{r} \text{Var}(X_{n+k})$$

Taking $r \to \infty$, (these are increasing events)

$$P(\sup_{k} |S_{n+k} - S_n| > \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} \text{Var}(X_{n+k})$$
Since the variances converge,
\begin{equation}
\lim_{n \to \infty} P(\sup_{k \geq 1} |S_{n+k} - S_n| > \varepsilon) = 0
\end{equation}

Now look at \(A_{n,\varepsilon} = \{\sup_{j,k \geq n} |S_j - S_k| > 2\varepsilon\}\).

We have
\[
A_{n,\varepsilon} = \{\sup_{j,k \geq n} |S_j - S_n + S_n - S_k| > 2\varepsilon\} \subset \sup_{j,k \geq n} |S_j - S_n| + |S_n - S_k| > 2\varepsilon
\]
\[
= \sup_{j \geq n} |S_j - S_n| + \sup_{k \geq n} |S_n - S_k| > 2\varepsilon \subset \{\sup_{j \geq n} |S_j - S_n| > \varepsilon\} \cup \{\sup_{k \geq n} |S_k - S_n| > \varepsilon\}
\]
\[
= \{\sup_{j \geq 1} |S_{n+j} - S_n| > \varepsilon\} \cup \{\sup_{k \geq 1} |S_{n+k} - S_n| > \varepsilon\}
\]

Since \(P(A \cup B) \leq P(A) + P(B)\), we see that (8.4) implies that \(P(A_{n,\varepsilon}) \to 0\). Since \(A_{n,\varepsilon} \ni \bigcap_n A_{n,\varepsilon}\), this shows that \(P(\bigcap_n A_{n,\varepsilon}) = 0\). So (by taking union over all rational \(\varepsilon > 0\)) we see that
\[
P(\forall \varepsilon > 0 \exists n \sup_{j,k \geq n} |S_j - S_k| > 2\varepsilon) = 0
\]

That is,
\[
P(\forall \varepsilon > 0 \exists n \forall j,k > n |S_j - S_k| < 2\varepsilon) = 1
\]

This means that there is \(\Omega_0 \subset \Omega\) of probability one such that for all \(\omega \in \Omega_0\) the sequence of numbers \(\{S_k(\omega)\}_{k \in \mathbb{N}}\) is a Cauchy sequence, i.e., \(\lim_{n \to \infty} S_n(\omega)\) exists. \(\square\)

**Corollary 8.9** (Kolmogorov’s two series theorem). Suppose that \(\{X_k\}\) is independent and that the following two series converge:
\begin{equation}
\sum_n E(X_n) \text{ converges, } \sum_n \text{Var}(X_n) < \infty
\end{equation}

Then \(\sum_n X_n\) converges.

**Proof.** \(\sum_{k=1}^n X_k = \sum_{k=1}^n (X_k - E(X_k)) + \sum_{k=1}^n E(X_k)\), so we get the sum of two convergent series. \(\square\)

**Corollary 8.10** (Kolmogorov’s three series theorem). Suppose that \(\{X_k\}\) is independent and that for some positive \(c > 0\) the following three series converge:
\begin{equation}
\sum_n P(|X_n| > c) < \infty, \sum_n E(X_n 1_{|X_n| \leq c}) \text{ converges, } \sum_n \text{Var}(X_n 1_{|X_n| \leq c}) < \infty
\end{equation}

Then \(\sum_n X_n\) converges.

**Proof.** Let \(X'_n = X_n 1_{|X_n| \leq c}\) denote the truncated random variables. Define \(m_n = E(X'_n)\). By Theorem 8.8, \(\sum_n (X'_n - m_n)\) converges with probability one. Since \(\sum_n m_n\) converges, therefore \(\sum_n X'_n\) converges. Now we note that
\[
\sum_n P(X_n \neq X'_n) = \sum_n P(|X_n| > c) < \infty
\]
so by Borel-Cantelli’s Lemma, \(P(X_n \neq X'_n \ i.o.) = 0\). Thus the series \(\sum_n X_n\) converges. \(\square\)
Example 8.1. It is well known that the harmonic series $\sum \frac{1}{n}$ diverges while the alternating series $\sum \frac{(-1)^n}{n}$ converges. It is somewhat comforting to know that the latter is more typical: Suppose $\varepsilon_k = \pm 1$ is an infinite sequence of signs such that every $n$-tuple of $2^n$ possible signs $(\pm 1, \pm 1, \ldots, \pm 1)$ is equally likely. Then the series $\sum \frac{\varepsilon_n}{n}$ converges with probability one.

Proof. It is easy to see that $\varepsilon_1, \varepsilon_2, \ldots$ are independent random variables with mean 0 and variance 1, compare Example 4.3. So with $X_n = \varepsilon_n/n$ the result follows from Theorem 8.8. □

Here is another proof of the strong law of large numbers - this proof uses joint independence, and second moments.

Corollary 8.11. Suppose $\{X_k\}$ is independent with the same mean $m$ and uniformly bounded (finite) variances. Then $\frac{1}{n}S_n \to m$ with probability one.

Proof. Subtracting $m$ if necessary, without loss of generality we assume that $m = 0$. Since $\text{Var}(\frac{1}{n}X_n) = \sigma^2/n^2$ and the series $\sum 1/n^2$ converges, by Theorem 8.8, $\sum \frac{1}{n}X_n$ converges with probability one.

We now use the so called Kronecker’s Lemma for numerical sequences

Lemma 8.12. If the series $\sum x_n/n$ converges then $\frac{1}{n}(x_1 + \cdots + x_n) \to 0$.

Proof. To prove Kronecker’s lemma, write $s_n = x_1 + \cdots + x_n$ and $t_n = \sum_{k=1}^n x_k/k$.

We have $x_k = k(t_k - t_{k-1})$ so

$$
\frac{1}{n}s_n = \frac{1}{n} \sum_{k=1}^n k(t_k - t_{k-1}) = \frac{1}{n} \sum_{k=1}^n \sum_{j=1}^k (t_k - t_{k-1}) = \frac{1}{n} \sum_{j=1}^n \sum_{k=j}^n (t_k - t_{k-1}) = \frac{1}{n} \sum_{j=1}^n (t_n - t_{j-1}) = t_n - \frac{1}{n} \sum_{j=1}^n t_{j-1}
$$

Now $t_n \to t_\infty$ by assumption and the average must have the same limit.

So $\frac{1}{n}s_n \to t_\infty - t_\infty = 0$. □

Corollary 8.13 (Kolmogorov’s strong law of large numbers). Suppose $\{X_n\}$ is independent identically distributed with mean $m$. Then $\frac{1}{n}S_n \to m$ with probability one.

Proof. As previously we consider $S'_n = \sum_{k=1}^n X'_k$ where $X'_k = X_kI_{|X_k|\leq k}$. As previously, $(S_n - S'_n)/n \to 0$, and $E(S'_n)/n \to m$, so we only need to show that $(S'_n - E(S'_n))/n \to 0$.

To prove will $(S'_n - E(S'_n))/n \to 0$ we use Kronecker’s Lemma. That is, we verify that $\sum (X'_k - E(X'_k))/k$ converges with probability one. To apply Kolmogorov’s convergence criterion (Theorem 8.8) we only need to verify that $\sum \frac{n}{k} \text{Var}(X'_k)/k^2 < \infty$. 

The main estimate is similar to the one we already saw in the proof of Theorem 8.1.

\[
\sum_{k=1}^{\infty} \frac{\text{Var}(X'_k)}{k^2} \leq \sum_{k=1}^{\infty} \frac{1}{k^2} E(X^2 I_{|X| \leq k}) = E\left( X^2_1 \sum_{k \geq |X| \vee 1} \frac{1}{k^2} \right)
\]

\[
= \int_{|X_1| \leq 1} |X^2_1| \sum_{k=1}^{\infty} \frac{1}{k^2} + \int_{|X_1| \geq 1} |X^2_1| \sum_{k \geq |X_1|} \frac{1}{k^2} \leq \pi^2/6 + \int_{|X_1| \geq 1} \left( |X_1|^2 \left( \frac{1}{X^2_1} + \int_{|X_1|}^{\infty} \frac{1}{t^2} dt \right) \right) dP
\]

\[
\leq \pi^2/6 + 1 + E(|X_1|)
\]

Here we used the fact that if \( n - 1 \leq |X_1| \leq n \) then \( 1/n^2 \leq 1/|X_1|^2 \), so

\[
\sum_{k \geq |X_1|} \frac{1}{k^2} = \frac{1}{n^2} + \sum_{k=n+1}^{\infty} \frac{1}{k^2} \leq \frac{1}{n^2} + \int_{n}^{\infty} \frac{1}{t^2} dt \leq \frac{1}{|X_1|^2} + \int_{|X_1|}^{\infty} \frac{1}{t^2} dt.
\]

\[\square\]

**Remark 8.1** (Levy’s theorem). It is known that for independent random variables \( \sum_{n} X_n \) converges in distribution iff it converges in probability iff it converges with probability one. See [Varadhan, Theorem 3.9] or (for one implication) [Billingsley, Theorem 22.7]. This result holds true also in infinite dimensional setting (Ito-Nisio)

4. Etemadi’s inequality and its application
Theorem 8.14. Suppose that $X_1, \ldots, X_n$ are independent. Then

\[(8.7)\]

\[
P\left( \max_{1 \leq k \leq n} |S_k| \geq 3t \right) \leq 3 \max_{1 \leq k \leq n} P(|S_k| \geq t)
\]

Proof. Let $B_k$ be the event that $k$ is the first index where $|S_k| \geq 3t$. Then

\[
P\left( \max_{1 \leq k \leq n} |S_k| \geq 3t \right) \leq P(|S_n| \geq t) + \sum_{k=1}^{n-1} P(B_k \cap |S_n| \leq t)
\]

\[
\leq P(|S_n| \geq t) + \sum_{k=1}^{n-1} P(B_k \cap |S_n - S_k| \geq 2t) = P(|S_n| \geq t) + \sum_{k=1}^{n-1} P(B_k)P(|S_n - S_k| \geq 2t)
\]

\[
\leq P(|S_n| \geq t) + \max_k P(|S_n - S_k| \geq 2t)
\]

\[
\leq P(|S_n| \geq t) + \max_k (P(|S_n| \geq t) + P(|S_k| \geq t)) \leq 3 \max_{1 \leq k \leq n} P(|S_k| \geq t)
\]

\[\square\]

Corollary 8.15. For an independent sequence $\{X_n\}$ the partial sums $S_n$ converge with probability one iff they converge in probability.

Proof. Suppose $S_n \xrightarrow{P} S$. We will show that $S_n$ is a Cauchy sequence with probability one.

Since $P(|S_{n+k} - S_n| > \varepsilon) \leq P(|S_{n+k} - S| > \varepsilon/2) + P(|S_n - S| > \varepsilon/2)$, from $S_n \xrightarrow{P} S$ we get

\[
\lim_{n \to \infty} \sup_{k \geq 1} P(|S_{n+k} - S_n| > \varepsilon) = 0
\]

But by Etemadi’s inequality

\[
P\left( \max_{1 \leq k \leq m} |S_{n+k} - S_n| > \varepsilon \right) \leq 3 \max_{1 \leq k \leq m} P|S_{n+k} - S_n| > \varepsilon/3)
\]

Thus

\[
P\left( \sup_{k \geq 1} |S_{n+k} - S_n| > \varepsilon \right) \leq 3 \sup_{k \geq 1} P|S_{n+k} - S_n| > \varepsilon/3)
\]

Thus

\[
\lim_{n \to \infty} P\left( \sup_{k \geq 1} |S_{n+k} - S_n| > \varepsilon \right) = 0
\]

This is (8.4), and the rest of proof is completed as before. \[\square\]

Exercise 8.1. Suppose $\{X_k\}$ are independent uniform $U(0,1)$ random variables. Prove that $\max_{1 \leq k \leq n} X_k \to 1$ with probability one.
Exercise 8.2. Suppose \( \{X_k\} \) are independent uniform \( U(0,1) \) random variables. Prove that \( X_1 X_2 \ldots X_n \to 0 \) with probability one. Hint: There are many proofs, but perhaps the easiest uses Markov’s inequality. (This is a repeat of an Exercise 4.14 that was solved differently!)

Exercise 8.3. Suppose \( \{X_k\} \) are independent uniform \( U(0,1) \) random variables. Prove that the geometric means \( \{\sqrt[n]{X_1 X_2 \ldots X_n}\}_{n \in \mathbb{N}} \) converge with probability one. (And find the limit!)

Exercise 8.4. Suppose \( \{X_k\} \) are independent uniform \( U(0,1) \) random variables. Prove that the geometric means \( \{\sqrt[n]{X_1 X_2 \ldots X_n}\}_{n \in \mathbb{N}} \) converges in mean. (And find the limit!)

Exercise 8.5. Let \( X_1, X_2, \ldots \) be identically distributed random variables with finite second moments. Show that \( n P(|X_1| > \varepsilon \sqrt{n}) \to 0 \) and \( n^{-1/2} \max_{k \leq n} |X_k| \overset{P}{\to} 0 \). Hint: Second moments imply that for every \( \varepsilon > 0 \) one can find \( K \) such that \( E|X|^2_{k=1} > K < \varepsilon \).

Exercise 8.6. Suppose \( \{X_n\} \) is independent identically distributed and integrable. Prove that \( \frac{1}{n} X_n \to 0 \) with probability one.

*Hint #1:* (7.4) can be used to show a more general fact that if \( E|X|^p < \infty \) then \( \frac{1}{\sqrt{n}} X_n \to 0 \) with probability one.

*Hint #2:* One can use the strong law of large numbers (which one?). This is a good practice exercise, although we omitted the proof of the theorem you need here, so this solution is "less complete"!

Exercise 8.7. Suppose \( \{X_n\} \) is independent exponentially distributed\(^2\). Prove that although \( P(\frac{1}{\log n} X_1 \to 0) = 1 \), we have \( P(\frac{1}{\log n} X_n \to 0) = 0 \).

Exercise 8.8. Suppose \( X_n \) are constructed iteratively by the following procedure: \( X_1 \) is uniform \( U(0,1) \), and for \( n \geq 1 \), \( X_{n+1} \) has uniform distribution on \((0, X_n)\). Show that
\[
\frac{1}{n} \log X_n
\]
and find the limit.

Exercise 8.9. Suppose \( X_n \) are independent exponential with parameter \( \lambda_n > 0 \). For which \( \lambda_n \) we have \( X_n \to 0 \) in mean? In mean square? In probability? With probability one?

Exercise 8.10. Suppose \( X_n \) are independent random variables such that \( X_n \to X \) with probability one. Show that \( X \) cannot take two different values.

Exercise 8.11. Suppose \( \{X_n\} \) are independent with mean 0 and finite variance. Let \( S_n = \sum_{k=1}^n X_1 X_2 \ldots X_k \). Adapt the proof of Kolmogorov’s maximal inequality to estimate \( P(\max_{1 \leq k \leq n} |S_k| \geq t) \).

Exercise 8.12. Suppose \( \{X_n\} \) are independent identically distributed with mean 0 and finite variance. Show that the series \( \sum_{n=1}^{\infty} \frac{1}{n} X_n X_{n+1} \) converges almost surely.

Exercise 8.13. Suppose \( \{X_n\} \) is independent identically distributed square-integrable with mean \( E(X_k) = 0 \). Let \( \{c_n\} \) be a bounded sequence of numbers. Modify the proof of the law of large numbers (which one?) to show that \( \frac{1}{n} \sum_{k=1}^{n} c_k X_k \to 0 \) with probability one.

Exercise 8.14. Suppose \( X_n \) are independent identically distributed integrable with symmetric distribution: \( X_1 \) has the same law as \( -X_1 \). Prove that the series \( \sum_n \frac{1}{n} X_n \) converges with probability one. Hint: Corollary 8.10 is in fact iff.

\(^2\)That is, \( P(X_n \leq x) = 1 - e^{-x} \) for \( x > 0 \)
Exercise 8.15. Modify our proof of Theorem 8.1 under extra moments to show that $\frac{1}{n}S_n \to m$ with probability one under additional moment assumption that $E(|X_1|^{1+\delta}) < \infty$ for some $\delta > 0$. Hint: Consider the subsequence $\frac{1}{n^p}S_{np}$.

Exercise 8.16. Modify our proof of Theorem 8.1 under extra moments to show that under the assumption $E(|X_1|) < \infty$ we have $\frac{1}{n}S_{2n} \to m$ with probability one. Hint: An important step is to use (8.1) to show that $\sum_n \text{Var}(S_{2n}^2)/4^n < \infty$, see [Billingsley, page 283].

Exercise 8.17. Suppose $\{X_k\}$ are independent identically distributed and integrable. Show that $\frac{1}{n}S_n$ is uniformly integrable, so $\frac{1}{n}S_n \to m$ in $L_1$. That is, $E[\frac{1}{n}S_n - m] \to 0$.

Hint: Without loss of generality we can assume that mean is zero. Use symmetry: $\int \frac{1}{n}S_n > a dP = \int |S_n| > an X_k dP$ for $k \leq n$ to show that $\int \frac{1}{n}S_n > a dP \to 0$ as $n \to \infty$. Deduce from this that

$$\lim_{a \to \infty} \sup_n \int \frac{1}{n}S_n > a \frac{1}{n}S_n dP = 0$$

Exercise 8.18. Let $X_1, X_2, \ldots$ be independent random variables. Show that $A = \{\omega : \frac{1}{\pi}(X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)) \to 17\}$ is a tail event.

Exercise 8.19. Let $X_1, X_2, \ldots$ be independent random variables. Show that

$$A = \{\omega : \sum_{n=1}^{\infty} X_n(\omega) \text{ converges} \}$$

is a tail event.

---

3Etemadi’s proof allows $\delta = 0$ by using geometric subsequences!
Bibliography

[Gut]  A. Gut, Probability: a graduate course
[Proshan-Shaw]  S M. Proshan and P. Shaw, Essential of Probability Theory for Statistitcians,  
CRC Press 2016
Index

$L_1$ metric, 11
$L_2$ metric, 11
$L_p$-norm, 67
$\lambda$-system, 25
$\pi$-system, 25
$\sigma$-field, 16
$\sigma$-field generated by $X$, 45
distribution of a random variable, 46

Bernoulli random variables, 50
Binomial distribution, 17, 81
bivariate cumulative distribution function, 30
Bonferroni’s correction, 18
Boole’s inequality, 18
Borel $\sigma$-field, 45
Borel sigma-field, 16
Cantelli’s inequality, 74
cardinality, 9
Cauchy distribution, 115
Cauchy-Schwarz inequality, 66
centered, 72
Central Limit Theorem, 119
characteristic function, 111
characteristic function – continuity theorem, 115
Characteristic functions – uniqueness, 114
Characteristic functions – inversion formula, 114
Chebyshev’s inequality, 65
complex numbers, 110
conjugate exponents, 68
continuity condition, 14
converge in $L_p$, 69
converge in mean square, 69
convergence in distribution, 53, 101
converges in distribution, 125
converges in probability, 50
converges pointwise, 7
converges uniformly, 7
converges with probability one, 51
convex function, 66
copula, 56
correlation coefficient, 67
countable additivity, 14
covariance matrix, 128
cumulative distribution function, 26, 47
cylindrical sets, 32, 33
cylindrical sets, 32
DeMorgan’s law, 8
density function, 29, 82
diadic interval, 133
discrete random variable, 81
discrete random variables, 49
equal in distribution, 47
events, 13, 17
expected value, 62, 77
Exponential distribution, 82
exponential distribution, 29
Fatou’s lemma, 79
field, 13
finite dimensional distributions, 32
finitely-additive probability measure, 14
Fubini’s Theorem, 88
Geometric distribution, 81
Hölder’s inequality, 68, 83
inclusion-exclusion, 18
independent $\sigma$-fields, 37
independent events, 37
independent identically distributed, 50
independent random variables, 48
indicator functions, 9
induced measure, 46
infinite number of tosses of a coin, 133
integrable, 77
intersection, 8
Jensen’s inequality, 66
joint cumulative distribution function, 30
joint distribution of random variables, 47
Kolmogorov’s maximal inequality, 93
Kolmogorov’s one series theorem, 94
| **Kolmogorov’s three series theorem**, 95 |
| **Kolmogorov’s two series theorem**, 95 |
| **Kolmogorov’s zero-one law**, 93 |
| **Kolmogorov–Smirnov metric**, 10, 11 |
| **Kronecker’s Lemma**, 95 |

| **Lévy distance**, 107 |
| **law of X**, 46 |
| **Lebesgue’s dominated convergence theorem**, 79, 80 |
| **Lebesgue’s dominated convergence theorem – used**, 81, 92, 103, 116 |
| **Levy’s metric**, 11 |
| **Levy’s theorem**, 96 |
| **Lindeberg condition**, 121 |
| **Lyapunov’s condition**, 122 |
| **Lyapunov’s inequality**, 66 |

| **marginal cumulative distribution functions**, 30 |
| **Markov’s inequality**, 65 |
| **maximal inequality**, Etemadi’s, 96 |
| **maximal inequality, Kolmogorov’s**, 93 |
| **mean square convergence**, 84 |
| **measurable function**, 45 |
| **measurable rectangle**, 87 |
| **metric**, 10 |
| **metric space**, 10 |
| **Minkowski’s inequality**, 67 |
| **Minkowski’s inequality**, 83 |
| **moment generating function**, 64, 85 |
| **moments**, 63 |
| **Monotone Convergence Theorem**, 77 |
| **multivariate normal**, 127 |
| **multivariate normal distribution**, 128 |
| **multivariate random variable**, 45 |

| **negative binomial distribution**, 18 |
| **normal distribution**, 29 |

| **Poisson distribution**, 18, 81 |
| **Polya’s distribution**, 18 |
| **Portmanteau Theorem**, 103 |
| **power set**, 7 |
| **probability**, 13 |
| **probability measure**, 14 |
| **probability space**, 13, 17 |
| **product measure**, 88 |

| **quantile function**, 48, 103 |

| **random element**, 45 |
| **random variable**, 45 |
| **random vector**, 45 |

| **sample space**, 13 |
| **Scholle’s theorem**, 101 |
| **section**, 87 |
| **semi-algebra**, 15 |
| **semi-ring**, 15 |
| **sigma-field generated by \( \mathcal{A} \)**, 16 |
| **simple random variable**, 61 |
| **simple random variables**, 49 |

| **Skorohod’s theorem**, 103 |
| **Slutsky’s Theorem**, 102 |
| **Standard normal density**, 82 |
| **stochastic process with continuous trajectories**, 47 |
| **stochastic processes**, 46 |
| **stochastically bounded**, 57 |
| **symmetric distribution**, 98 |

| **tail \( \sigma \)-field**, 38 |
| **Tail integration formula**, 89 |
| **Taylor polynomials**, 109 |
| **tight**, 57 |
| **tight probability measure**, 19 |
| **Tonelli’s theorem**, 88 |
| **total variation metric**, 11 |
| **truncation of r.v.**, 56 |

| **uncorrelated**, 72 |
| **uniform continuous**, 28 |
| **Uniform density**, 82 |
| **uniform discrete**, 28 |
| **uniform singular**, 28 |
| **uniformly integrable**, 80, 105 |
| **union**, 8 |

| **variance**, 64 |

| **Wasserstein distance**, 11 |
| **weak convergence**, 53 |
| **weak law of large numbers**, 72 |

| **zero-one law**, 38, 93 |