Kolmogorov’s Maximal Inequality and 3-series Theorem

6.1 Kolmogorov’s maximal inequality

**Theorem 6.1 (Kolmogorov’s Maximal Inequality).** Let $X_1, \ldots, X_n$ be independent random variables with $\mathbb{E} X_i = 0, \mathbb{E} X_i^2 < \infty$ for all $1 \leq i \leq n$. Define $S_n = X_1 + \cdots + X_n$. Then

$$\mathbb{P} \left( \max_{1 \leq k \leq n} |S_k| \geq a \right) \leq \frac{\mathbb{E} S_n^2}{a^2}$$

for any $a > 0$.

**Proof Sketch:** Let $A_i = \{|S_1| < a, |S_2| < a, \ldots, |S_{i-1}| < a, |S_i| \geq a\}$, $i = 1, 2, \ldots, n$. Then $A_i \cap A_j = \phi$ for all $i \neq j$. Notice that $\mathbb{P} \left( \max_{1 \leq k \leq n} |S_k| \geq a \right) = \sum_{i=1}^n \mathbb{P}(|S_i| \geq a, \max_{1 \leq k < i} |S_k| < a)$. Then

$$\mathbb{E} S_n^2 \geq \sum_{i=1}^n \mathbb{E} S_n^2 1_{A_i} = \sum_{i=1}^n \mathbb{E}(S_i^2 + 2S_i(S_n - S_i) + (S_n - S_i)^2)1_{A_i} \geq \sum_{i=1}^n (a^2 \mathbb{P}(A_i) + 2 \mathbb{E} S_i 1_{A_i}(S_n - S_i)) = a^2 \mathbb{P} \left( \max_{1 \leq k \leq n} |S_k| \geq a \right).$$

In the last line we used the fact that $S_i 1_{A_i}$ is independent of $X_{i+1}, \ldots, X_n$. ■

**Theorem 6.2 (Basic $L^2$-convergence).** Let $X_1, X_2, \cdots$ be independent r.v.’s with $\mathbb{E}(X_i) = 0$ and $\sum_{i=1}^\infty \mathbb{V}(X_i) < \infty$. Then $S_n = \sum_{i=1}^n X_i$ converges in $L^2$ and a.s.

**Proof.** To see $L^2$-convergence, we will prove that $(S_n)_{n \geq 1}$ is Cauchy in $L^2$. It is enough to show that $||S_n - S_m||_2 < \varepsilon$ for all $n, m \geq N(\varepsilon)$. Suppose $n > m$, then we obtain

$$||S_n - S_m||_2^2 = \mathbb{E}(S_n - S_m)^2 = \mathbb{E}(X_{m+1} + X_{m+2} + \cdots + X_n)^2 = \mathbb{V}(X_{m+1} + X_{m+2} + \cdots + X_n) = \mathbb{V}(X_{m+1}) + \mathbb{V}(X_{m+2}) + \cdots + \mathbb{V}(X_n).$$

For any $\varepsilon > 0$, there exists $N = N(\varepsilon)$ with $\sum_{i=N}^\infty \mathbb{V}(X_i) < \varepsilon^2$, thus we have $||S_n - S_m||_2 < \varepsilon$ for all $n, m \geq N(\varepsilon)$ which implies that $(S_n)_{n \geq 1}$ is a Cauchy sequence in $L^2$ and therefore converges in $L^2$. Now, it remains to show a.s. convergence. In other words, we need to show $\mathbb{P}(S_n \text{ converges}) = 1$. Define $M_N = \max\{|S_n - S_m| : n, m \geq N\}$, then clearly $M_N$ is decreasing. Indeed, it is enough to
show that $M_N \xrightarrow{p} 0$ as $N \to \infty$. Let $\varepsilon > 0$ be given. Then,

$$\mathbb{P}(M_N > \varepsilon) = \mathbb{P}(\max_{n,m \geq N} |S_n - S_m| > \varepsilon) \leq \mathbb{P}(\max_{k > N} |S_k - S_N| > \varepsilon/2)$$

$$= \lim_{n \to \infty} \mathbb{P}(\max_{1 \leq k \leq n} |S_{N+k} - S_N| > \varepsilon/2)$$

$$\leq \lim_{n \to \infty} \frac{\text{Var}(S_{N+n} - S_N)}{(\varepsilon/2)^2}$$

$$= \lim_{n \to \infty} \left(\frac{2}{\varepsilon}\right)^2 \cdot \sum_{i=N+1}^{N+n} \text{Var}(X_i) = \left(\frac{2}{\varepsilon}\right)^2 \cdot \sum_{i=N}^{\infty} \text{Var}(X_i),$$

where the second inequality follows by Kolmogorov’s Maximal inequality. Hence, $\lim_{N \to \infty} \mathbb{P}(M_N > \varepsilon) = 0$ or $M_N \xrightarrow{p} 0$, thus we obtain $M_N \to 0$ a.s. Therefore, $\mathbb{P}((S_n)_{n \geq 1} \text{ is Cauchy}) = 1$. ■

Using basic $L^2$-convergence theorem, we can give a simple proof of SLLN under finite second moment. Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}X_1 = 0$ and $\text{Var}(X_1) < \infty$. By basic $L^2$-convergence theorem, we have $\sum_{k=1}^{n} X_k/k$ converges a.s as $n \to \infty$. Now, using Kronecker’s lemma (stated and proved below), we have $n^{-1} \sum_{k=1}^{n} X_k \to 0$ a.s.

**Lemma 6.3 (Kronecker’s lemma).** Let $a_n \uparrow \infty$ and $\sum_{i=1}^{n} x_i/a_i$ converges. Then

$$\frac{1}{a_n} \sum_{i=1}^{n} x_i \to 0$$

**Proof.** We have $b_n := \sum_{i=1}^{n} x_i/a_i \to b$, for some $b \in \mathbb{R}$, as $n \to \infty$. Moreover, $x_n = a_n(b_n - b_{n-1})$ so

$$\frac{1}{a_n} \sum_{i=1}^{n} x_i = \frac{1}{a_n} \sum_{i=1}^{n} (a_ib_i - a_ib_{i-1}) = \frac{1}{a_n} (a_nb_n + \sum_{i=1}^{n-1} (a_i - a_{i+1})b_i) = b_n - \frac{1}{a_n} \sum_{i=1}^{n-1} (a_{i+1} - a_i)b_i.$$

Let $d_i = a_{i+1} - a_i$. We know that $b_n \to b$ and thus

$$\frac{\sum_{i=1}^{n-1} d_i b_i}{\sum_{i=1}^{n-1} d_i} - b = \frac{\sum_{i=1}^{n-1} d_i(b_i - b)}{\sum_{i=1}^{n-1} d_i} \to 0$$

as $n \to \infty$, since $d_i \geq 0$ and $\sum_{i=1}^{\infty} d_i = \infty$. So this proves that $\frac{1}{a_n} \sum_{i=1}^{n} x_i \to 0$ completing the proof. ■

### 6.2 Applications of SLLN

**Theorem 6.4 (Renewal Theorem).** Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.’s with $X_i \geq 0$, $\mathbb{E}X_1 = \mu > 0$. Then

$$\frac{N_t}{t} \xrightarrow{a.s.} \frac{1}{\mu}$$

where $N_t = \sup\{n \mid X_1 + X_2 + \cdots + X_n \leq t\}$. 

Proof Sketch: By SLLN, if \( S_n = X_1 + X_2 + \cdots + X_n \), then \( \frac{S_n}{n} \xrightarrow{a.s.} \mu \) as \( n \to \infty \). We observe that

- \( N_t \uparrow \infty \) a.s. as \( t \to \infty \). (\( \{N_t \geq n\} = \{X_1 + X_2 + \cdots + X_n \leq t\} \)).
- \( \mathbb{P}\left( \frac{S_n}{n} \xrightarrow{t \to \infty} \mu, N_t \xrightarrow{t \to \infty} \infty \right) = 1 \) or \( \mathbb{P}\left( \frac{S_{N_t}}{N_t} \xrightarrow{t \to \infty} \mu, \frac{S_{N_{t+1}}}{N_{t+1}} \xrightarrow{t \to \infty} \mu \right) = 1 \).

By definition, we have \( S_{N_t} \leq t < S_{N_{t+1}} \) so that

\[
\frac{S_{N_t}}{N_t} \leq \frac{t}{N_t} < \frac{S_{N_{t+1}}}{N_t} = \frac{S_{N_{t+1}}}{N_{t+1}} \cdot \frac{N_{t+1}}{N_t}.
\]

Both upper and lower bounds converge to \( \mu \) a.s. and this completes the proof. \( \blacksquare \)

**Definition 6.5 (Empirical Distribution).** Let \( X_1, X_2, \cdots \) be a sequence of i.i.d. r.v.'s with distribution function \( F \). Define

\[
F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i \leq x} \text{ for all } x \in \mathbb{R}.
\]

\( F_n \) is called the empirical distribution function. By SLLN, \( F_n(x) \xrightarrow{a.s.} F(x) \), for each \( x \in \mathbb{R} \).

**Theorem 6.6 (Glivenko-Cantelli Lemma).** We have

\[
||F_n - F||_{\infty} = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0.
\]

**Proof Sketch:*** Fix \( k \geq 1 \), define

\[
x_{j,k} = F^{-1} \left( \frac{j}{k} \right) = \inf \left\{ x : F(x) \geq \frac{j}{k} \right\}, \quad 1 \leq j < k.
\]

Define \( x_{0,k} = -\infty, x_{k,k} = \infty \). By SLLN,

\[
\mathbb{P}\left( F_n(x_{j,k}) \xrightarrow{a.s.} F(x_{j,k}) \forall 0 \leq j \leq k \right) = 1.
\]

Similarly,

\[
F_n(x-) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i < x} \xrightarrow{a.s.} F(x-).
\]

Thus

\[
\mathbb{P}\left( F_n(x_{j,k}-) - F(x_{j,k}-) \to 0, F_n(x_{j,k}) - F(x_{j,k}) \to 0 \forall 0 \leq j \leq k \right) = 1.
\]

Define

\[
\Delta_k^{(n)} = \max_{0 \leq j \leq k} \left\{ |F_n(x_{j,k}) - F(x_{j,k})|, |F_n(x_{j,k}-) - F(x_{j,k})| \right\}.
\]

Take \( x \in [x_{j,k}, x_{j+1,k}] \). Then,

\[
F_n(x_{j,k}) \leq F_n(x) \leq F_n(x_{j+1,k}) \implies F_n(x_{j,k}) - F(x) \leq F_n(x) - F(x) \leq F_n(x_{j+1,k}-) - F(x).
\]

Since

\[
|F_n(x_{j+1,k}-) - F(x)| \leq |F_n(x_{j+1,k}) - F(x_{j+1,k}^-)| + |F(x_{j+1,k}^-) - F(x)| \leq \Delta_k^{(n)} + \frac{1}{k},
\]

**Week 6: Kolmogorov’s Maximal Inequality and 3-series Theorem 6-3**
and
\[ |F_n(x_{j,k}) - F(x)| \leq |F_n(x_{j,k}) - F(x_{j,k})| + |F(x_{j,k}) - F(x)| \leq \Delta_{k}^{(n)} + \frac{1}{k}, \]
we have \( \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \Delta_{k}^{(n)} + \frac{1}{k} \). Note that \( \Delta_{k}^{(n)} \xrightarrow{a.s.} 0 \). Thus
\[ P \left( \limsup_{n} \|F_n - F\|_{\infty} \leq \frac{1}{k} \right) = 1 \quad \forall k \]
\[ \implies P \left( \limsup_{n} \|F_n - F\|_{\infty} \leq \frac{1}{k} \forall k \right) = 1, \]
\[ \implies P \left( \limsup_{n} \|F_n - F\|_{\infty} = 0 \right) = 1. \]

Remark. See "Bootstrap" in statistics for applications of Glivenko-Cantelli Lemma.

6.3 Tail events and Kolmogorov’s 3-Series Theorem

Definition 6.7 (Tail \( \sigma \)-field). Let \( X_1, X_2, \ldots \) be independent r.v.s. The \( \sigma \)-field
\[ \mathcal{T} = \bigcap_{k=1}^{\infty} \sigma(X_k, X_{k+1}, \ldots) \]
is called the tail \( \sigma \)-field.

Example 6.8. We have \( \{ \text{lim } X_n \text{ exists} \} \in \mathcal{T} \), \( \{ \text{lim sup}_n (X_1 + \ldots + X_n)/n \leq a \} \in \mathcal{T} \). However, the event \( \{ \text{lim sup}_n (X_1 + \ldots + X_n) \leq a \} \notin \mathcal{T} \).

Theorem 6.9 (Kolmogorov’s 0-1 Law). For all \( A \in \mathcal{T} \), \( P(A) \in \{0, 1\} \).

Proof. We will prove that \( \mathcal{T} \perp \mathcal{T} \) (the tail \( \sigma \)-field is independent of itself). Then for all \( A \in \mathcal{T} \),
\[ P(A) = P(A \cap A) = P(A) \cdot P(A) \]
which implies \( P(A) \in \{0, 1\} \). The proof is in 4 steps.

Step 1. \( \sigma(X_1, \ldots, X_{k-1}) \perp \sigma(X_{l+1}, X_{l+2}, \ldots) \) for all \( l \geq k \), since \( X_i \)'s are independent.

Step 2. In particular, \( \sigma(X_1, \ldots, X_{k-1}) \perp \mathcal{T} \) for all \( k \geq 2 \).

Step 3. \( \sigma(X_1, \ldots) \perp \mathcal{T} \).

Step 4. \( \mathcal{T} \perp \mathcal{T} \).

Proof of step 1: Clearly \( \sigma(X_1, \ldots, X_{k-1}) \perp \sigma(X_l, \ldots, X_{l+j}) \) for all \( l \geq k, j \geq 0 \), which implies that
\[ \sigma(X_1, \ldots, X_{k-1}) \perp \bigcup_{j=0}^{\infty} \sigma(X_l, \ldots, X_{l+j}). \]
By \(\pi - \lambda\) theorem, \(\sigma(X_1, \ldots, X_{k-1}) \perp \sigma(\cup_{j=0}^{\infty} \sigma(X_i, \ldots, X_{i+j})) = \sigma(X_i, X_{i+1}, \ldots)\).

**Proof of step 2:** Note that \(\mathcal{F} \perp \mathcal{G}\) implies that any sub \(\sigma\)-field of \(\mathcal{F}\) is independent of \(\mathcal{G}\). Clearly \(\mathcal{T} \subseteq \sigma(X_k, X_{k+1}, \ldots)\). So by \((*)\), \(\sigma(X_1, \ldots, X_{k-1}) \perp \mathcal{T}\).

**Proof of step 3:** We have \(\bigcup_{k=1}^{\infty} \sigma(X_1, \ldots, X_{k-1}) \perp \mathcal{T}\). Thus by \(\pi - \lambda\) theorem, \(\sigma(X_1, \ldots) \perp \mathcal{T}\). Clearly \(\mathcal{T} \subseteq \sigma(X_1, X_2, \ldots)\). So by \((*)\), we have the proof. \(\blacksquare\)

**Theorem 6.10 (Kolmogorov’s 3-series Theorem).** Let \(X_1, X_2, \ldots\) be independent. Fix \(b > 0\). Consider the three deterministic series:

\[
(i) \sum_{i=1}^{n} \mathbb{P}(|X_i| > b), \quad (ii) \sum_{i=1}^{n} \mathbb{E}(X_i \mathbb{1}_{|X_i| \leq b}), \quad (iii) \sum_{i=1}^{n} \text{Var}(X_i \mathbb{1}_{|X_i| \leq b}).
\]

Then \(\sum_{i=1}^{n} X_i\) converges a.s. if and only if \((i), (ii), (iii)\) converge.

**Proof.** We will first prove the if part. Write \(X_i = U_i + v_i + W_i\), where

\[
U_i := X_i \mathbb{1}_{|X_i| > b}, \quad v_i := \mathbb{E}X_i \mathbb{1}_{|X_i| \leq b}, \quad \text{and} \quad W_i := X_i \mathbb{1}_{|X_i| \leq b} - v_i.
\]

By convergence of \((ii)\), we have \(\sum_{i \geq 1} v_i < \infty\). We will prove that both \(\lim_{n \to \infty} \sum_{i=1}^{n} U_i\) and \(\lim_{n \to \infty} \sum_{i=1}^{n} W_i\) exists a.s. By Borel-Cantelli Lemma and convergence of series \((i)\), we have \(\mathbb{P}(|X_i| > b \text{ i.o.}) = 0\) which implies that \(\lim_{n \to \infty} \sum_{i=1}^{n} U_i\) exists a.s. Convergence of \((iii)\) and Basic \(L^2\) convergence implies \(\lim_{n \to \infty} \sum_{i=1}^{n} W_i\) exists a.s.

For the only if part, we need to use Central Limit Theorem, which states that for a sequence of iid mean zero, variance one rvs \(X_1, X_2, \ldots\), the sequence \(n^{-1/2} S_n\) converges in distribution to \(N(0, 1)\) and will be proved later. \(\blacksquare\)

Another application of Central Limit Theorem and truncation + subsequence technique is the following result, which will not be proved here. The crucial idea is that \(S_n/\sqrt{\sigma}\) converges in distribution to \(N(0, 1)\) and for a rv \(X \sim N(0, 1)\) we have \(\mathbb{P}(X \geq t) = \mathbb{P}(X \leq -t) \leq e^{-t^2/2}, t > 0\).

**Theorem 6.11 (Law of Iterated Logarithms (LIL)).** Let \(X_1, X_2, \ldots\) be i.i.d. with mean 0, \(\text{Var}(X_1) = 1\). Define \(S_n = \sum_{i=1}^{n} X_i\). Then

\[
\limsup_{n} \frac{S_n}{\sqrt{2n \log \log n}} = 1 \text{ a.s.}
\]

\[
\liminf_{n} \frac{S_n}{\sqrt{2n \log \log n}} = -1 \text{ a.s.}
\]

A sequence \(\{t_n\}_{n \geq 1}\) is **subadditive**, if \(t_{m+n} \leq t_m + t_n\) for all \(m, n \geq 1\). The following lemma for a deterministic sequence can be generalize to random sequences to give another proof of SLLN.

**Theorem 6.12 (Subadditive limit theorem).** If \(\{t_n\}_{n \geq 1}\) is subadditive, then

\[
\frac{t_n}{n} \to \inf_{m \geq 1} \frac{t_m}{m} \text{ as } n \to \infty.
\]
Proof. Clearly, $\liminf_n t_n/n \geq \inf_{m\geq 1} t_m/m$. Note that, 
\[
\frac{t_{mn+k}}{mn+k} \leq \frac{nt_m + t_k}{mn+k} = \frac{t_m + t_k}{m} + \frac{t_k}{mn},
\]
Thus, 
\[
\limsup_n \frac{t_{mn+k}}{mn+k} \leq \frac{t_m}{m}.
\]
Fix $m \geq 1$ and $0 \leq k < m$, then 
\[
\limsup_n \frac{t_{mn+k}}{mn+k} \leq \frac{t_m}{m}.
\]
Thus, 
\[
\limsup_n \frac{t_n}{n} \leq \frac{t_m}{m}.
\]
Since $m$ is arbitrary, we have 
\[
\limsup_{n \to \infty} \frac{t_n}{n} \leq \inf_{m\geq 1} \frac{t_m}{m}.
\]

**Corollary 6.13.** Let $X_1, X_2, \ldots$ be i.i.d. r.v.s. Define $S_n = X_1 + X_2 + \cdots + X_n, n \geq 1$. Then,
\[
\kappa(a) := \lim_{n \to \infty} -\frac{1}{n} \log P(S_n \geq a) \text{ exists and is non-negative (need not be finite).}
\]

**Proof.** We use the fact that 
\[
P(S_{n+m} \geq (n+m)a) \geq P(S_{m} \geq ma, S_{n+m} - S_m \geq na) \geq P(S_m \geq ma) \cdot P(S_{n} \geq na)
\]
and thus the sequence $t_n := -\log P(S_n \geq a) \geq 0$ is subadditive. 

In the next lecture, we will study conditions under which $\kappa(a)$ is positive and finite. The random version of the subadditive limit theorem is stated below without proof.

**Theorem 6.14 (Subadditive ergodic theorem).** Let $\{X_{m,n} \mid n > m \geq 0\}$ be a collection of rvs indexed by integers $n > m \geq 0$, such that
\[
X_{0,n} \leq X_{0,m} + X_{m,n} \text{ a.s. for } 0 < m < n.
\]
Assume that

(a) The joint distributions of $\{X_{m+1,m+k+1}, k \geq 1\}$ are the same as those of $\{X_{m,m+k}, k \geq 1\}$ for each $m \geq 0$,

(b) For each $k \geq 1$, $\{X_{nk,(n+1)k}, n \geq 1\}$ is a stationary process and

(c) For each $n$, $E|X_{0,n}| < \infty$ and $EX_{0,n} \geq -cn$ for some constant $c$.

Then, $\frac{1}{n}X_{0,n}$ converges a.s. and in $L^1$ to a r.v. $X$ and $EX \in [-c, \infty)$. Moreover, if the processes in (b) are ergodic, then $X$ is a constant a.s.