1 Kolmogorov’s maximal inequality

Let $X_1,\ldots,X_n$ be independent random variables, $\mathbb{E}X_k = 0$, $\text{Var}(X_k) < \infty$ for $k = 1,\ldots,n$. Consider $S_k = X_1 + \cdots + X_k$.

1.1 Lemma. $\mathbb{E}(\varphi(X_1,\ldots,X_k)S_n) = \mathbb{E}(\varphi(X_1,\ldots,X_k)S_k)$ for every $k < n$ and every bounded Borel function $\varphi : \mathbb{R}^k \to \mathbb{R}$.

Proof. Denoting by $\mu_k$ the distribution of $X_k$ we have $\int x \mu_k(dx) = 0$, thus

$$\mathbb{E}(\varphi(X_1,\ldots,X_k)S_n) = \int \mu_1(dx_1)\ldots\mu_n(dx_n) \varphi(x_1,\ldots,x_k)(x_1+\cdots+x_n) = \int \mu_1(dx_1)\ldots\mu_k(dx_k) \varphi(x_1,\ldots,x_k) \int \mu_{k+1}(dx_{k+1})\ldots\mu_n(dx_n) (x_1+\cdots+x_n) = \int \mu_1(dx_1)\ldots\mu_k(dx_k) \varphi(x_1,\ldots,x_k)(x_1+\cdots+x_k) = \mathbb{E}(\varphi(X_1,\ldots,X_k)S_k).$$

\[ \square \]

In terms of conditioning,

$$\mathbb{E}(\varphi(X_1,\ldots,X_k)S_n) = \mathbb{E}(\mathbb{E}(\varphi(X_1,\ldots,X_k)S_n|X_1,\ldots,X_k)) = \mathbb{E}(\varphi(X_1,\ldots,X_k)\mathbb{E}(S_n|X_1,\ldots,X_k)) = \mathbb{E}(\varphi(X_1,\ldots,X_k)S_k).$$

1.2 Exercise. $\mathbb{E}(\varphi(X_1,\ldots,X_k)S_n^2) \geq \mathbb{E}(\varphi(X_1,\ldots,X_k)S_k^2)$ for every $k < n$ and every bounded Borel function $\varphi : \mathbb{R}^k \to [0,\infty)$.

Prove it.

Hint: $\int \mu_{k+1}(dx_{k+1})\ldots\mu_n(dx_n)(x_1+\cdots+x_n)^2 \geq (\int \mu_{k+1}(dx_{k+1})\ldots\mu_n(dx_n)(x_1+\cdots+x_n))^2$.

1.3 Remark. More generally, the Jensen inequality gives $\mathbb{E}(\varphi(X_1,\ldots,X_k)\psi(S_n)) \geq \mathbb{E}(\varphi(X_1,\ldots,X_k)\psi(S_k))$ for every $k < n$, every bounded Borel function $\varphi : \mathbb{R}^k \to \mathbb{R}$ and every convex $\psi : \mathbb{R} \to \mathbb{R}$ (as long as the expectations exist). Especially, $\psi(s)$ may be $|s-a|$, or $(s-a)^+$, or $(s-a)^-$ for any $a \in \mathbb{R}$.

1.4 Theorem. For every $n$ and every $c > 0$,

$$\mathbb{P}\left(\max_{k=1,\ldots,n} |S_k| \geq c \right) \leq \frac{1}{c^2} \mathbb{E}S_n^2.$$
Proof. We introduce events $A_k = \{|S_1| < c, \ldots, |S_{k-1}| < c, |S_k| \geq c\}$ and apply 1.2 to their indicators:

$$E(1_{A_k} S^2_n) \geq E(1_{A_k} S^2_k) \geq c^2 P(A_k).$$

Summing up we get

$$E(1_{A} S^2_n) \geq c^2 P(A)$$

where $A = A_1 \cup \cdots \cup A_n = \{\max_k |S_k| \geq c\}$. \qed

1.5 Exercise. For an infinite sequence $(X_k)_k$, for every $c > 0$,

$$P\left(\sup_k |S_k| \geq c\right) \leq \frac{1}{c^2} \sum_{k=1}^{\infty} \text{Var} X_k.$$

Prove it.

Hint: it is not hard, but be careful; if in trouble, try $P\left(\sup_k |S_k| > c - \varepsilon\right)$.

2 Random series

2.1 Proposition. Let $X_1, X_2, \ldots$ be independent random variables, $E X_k = 0$, $\text{Var}(X_k) < \infty$ for all $k$, and

$$\sum_{k=1}^{\infty} \text{Var} X_k < \infty.$$

Then the series

$$\sum_{k=1}^{\infty} X_k$$

converges a.s.

Proof. Let $S_n = X_1 + \cdots + X_n$. It is sufficient to prove that $(S_n(\omega))_n$ is a Cauchy sequence for almost all $\omega$, that is,

$$\sup_{k,l \geq n} |S_k - S_l| \downarrow 0 \quad \text{a.s. as } n \to \infty,$$

or equivalently,

$$P\left(\sup_{k,l \geq n} |S_k - S_l| \geq 2\varepsilon\right) \downarrow 0 \quad \text{as } n \to \infty.$$
for every \( \varepsilon > 0 \). Using 1.5,

\[
\Pr \left( \sup_{k,l \geq n} |S_k - S_l| \geq 2\varepsilon \right) \leq \Pr \left( \sup_k |S_{n+k} - S_n| \geq \varepsilon \right) = \\
= \Pr \left( \sup_k |X_{n+1} + \ldots + X_{n+k}| \geq \varepsilon \right) \leq \frac{1}{\varepsilon^2} \sum_k \text{Var} X_{n+k} \downarrow 0
\]

as \( n \to \infty \).

\[
\square
\]

### 3 Martingale convergence

Given \( f \in L_2(0,1) \), we consider its orthogonal projection \( f_n \) to the \( 2^n \)-dimensional subspace of step functions,

\[
f_n(x) = 2^n \int_{2^{-n(k-1)}}^{2^{-n-k}} f(u) \, du \quad \text{for } x \in (2^{-n(k-1)}, 2^{-n-k}).
\]

In terms of binary digits \( \beta_1(x), \beta_2(x), \ldots \) of \( x \),

\[
x = \frac{\beta_1(x)}{2} + \frac{\beta_2(x)}{2^2} + \ldots, \quad \beta_k(x) \in \{0,1\},
\]

we have \( f_n(x) = g_n(\beta_1(x), \ldots, \beta_n(x)) \) for some \( g_n : \{0,1\}^n \to \mathbb{R} \). Note that

\[
g_k(b_1, \ldots, b_k) = \frac{1}{2} g_{k+1}(b_1, \ldots, b_k, 0) + \frac{1}{2} g_{k+1}(b_1, \ldots, b_k, 1)
\]

and moreover,

\[
g_k(b_1, \ldots, b_k) = 2^{-(n-k)} \sum_{b_{k+1}, \ldots, b_n} g_n(b_1, \ldots, b_k, b_{k+1}, \ldots, b_n)
\]

for \( k < n \).

Treating \((0,1)\) with Lebesgue measure as a probability space and \( \beta_1, \beta_2, \ldots \) as random variables we see that \( \beta_1, \beta_2, \ldots \) are independent, \( \Pr (\beta_k = 0) = 0.5 = \Pr (\beta_k = 1) \), and the random variables \( f_n = g_n(\beta_1, \ldots, \beta_n) \) satisfy

\[
\mathbb{E} (f_n | \beta_1, \ldots, \beta_k) = f_k \quad \text{for } k < n.
\]

Such sequences of random variables are called martingales. The differences \( f_n - f_{n-1} \) need not be independent, but still, we have a counterpart of 1.1 (It really means that \( f_k \) is the orthogonal projection of \( f_n \) to the \( 2^k \)-dimensional subspace...).
3.1 Lemma. $\mathbb{E}(\varphi(\beta_1, \ldots, \beta_k) f_n) = \mathbb{E}(\varphi(\beta_1, \ldots, \beta_k) f_k)$ for every $k < n$ and every function $\varphi : \{0, 1\}^k \to \mathbb{R}$.

Proof.

\[
\mathbb{E}(\varphi(\beta_1, \ldots, \beta_k) f_n) = 2^{-n} \sum_{b_1, \ldots, b_n} \varphi(b_1, \ldots, b_k) g_n(b_1, \ldots, b_n) = 2^{-k} \sum_{b_1, \ldots, b_k} \varphi(b_1, \ldots, b_k) 2^{-(n-k)} \sum_{b_{k+1}, \ldots, b_n} g_n(b_1, \ldots, b_n) = 2^{-k} \sum_{b_1, \ldots, b_k} \varphi(b_1, \ldots, b_k) g_k(b_1, \ldots, b_k) = \mathbb{E}(\varphi(\beta_1, \ldots, \beta_k) f_k).
\]

In terms of conditioning,

\[
\mathbb{E}(\varphi(\beta_1, \ldots, \beta_k) f_n) = \mathbb{E}(\mathbb{E}(\varphi(\beta_1, \ldots, \beta_k) f_n | \beta_1, \ldots, \beta_k)) = \mathbb{E}(\varphi(\beta_1, \ldots, \beta_k) \mathbb{E}(f_n | \beta_1, \ldots, \beta_k)) = \mathbb{E}(\varphi(\beta_1, \ldots, \beta_k) f_k).
\]

3.2 Exercise. $\mathbb{E}(\varphi(\beta_1, \ldots, \beta_k) f_n^2) \geq \mathbb{E}(\varphi(\beta_1, \ldots, \beta_k) f_k^2)$ for every $k < n$ and every $\varphi : \{0, 1\}^k \to [0, \infty)$.

Prove it.

Hint: similar to 1.2.

In fact, $\mathbb{E}(\varphi(\beta_1, \ldots, \beta_k) \psi(f_n)) \geq \mathbb{E}(\varphi(\beta_1, \ldots, \beta_k) \psi(f_k))$ for convex $\psi$.

3.3 Exercise. For every $n$ and every $c > 0$,

\[
\mathbb{P}\left(\max_{k=1, \ldots, n} |f_k| \geq c\right) \leq \frac{1}{c^2} \mathbb{E} f_n^2.
\]

Prove it.

Hint: similar to 1.3.

3.4 Exercise. For every $c > 0$,

\[
\mathbb{P}\left(\sup_k |f_k| \geq c\right) \leq \frac{1}{c^2} \sup_k \mathbb{E} f_k^2.
\]

Prove it.

Hint: similar to 1.3.

Applying it to $f - f_n$ (in place of $f$) we get

\[
\mathbb{P}\left(\sup_k |f_{n+k} - f_n| \geq c\right) \leq \frac{1}{c^2} \sup_k \mathbb{E} |f_{n+k} - f_n|^2.
\]
3.6 Proposition. The sequence \((f_n)_n\) converges almost everywhere.

Proof. The differences \(f_n - f_{n-1}\) are mutually orthogonal, thus

\[
\|f_0\|^2 + \|f_1 - f_0\|^2 + \cdots + \|f_n - f_{n-1}\|^2 = \|f_n\|^2 \leq \|f\|^2.
\]

It follows that \(\sum_{k=n}^{\infty} \|f_{k+1} - f_k\|^2 \to 0\) as \(n \to \infty\). Therefore \(\sup_k \mathbb{E}|f_{n+k} - f_n|^2 \to 0\) as \(n \to \infty\). By (3.5), \(\mathbb{P}(\sup_k |f_{n+k} - f_n| \geq \varepsilon) \to 0\) as \(n \to \infty\) for every \(\varepsilon > 0\). Similarly to the proof of 2.1 we conclude that \((f_n(x))_n\) is a Cauchy sequence for almost all \(x\).

In fact, \(\lim_{n} f_n = f\).

4 Backwards martingale convergence

Given \(f \in L_2(0,1)\), we consider its orthogonal projection \(f_n\) to the subspace of \(2^{-n}\)-periodic functions,

\[
f_n(x) = 2^{-n} \sum_{k:0<x+2^{-n}k<1} f(x + 2^{-n}k) \quad \text{for } x \in (0,1).
\]

Note that

\[
f_k(x) = \frac{1}{2} f_{k-1}(x) + \frac{1}{2} f_{k-1}(x + 2^{-k})
\]

and moreover,

\[
f_k(x) = 2^{-(k-n)} \sum_{j=1}^{2^{k-n}} f_n(x + 2^{-k}j)
\]

for \(n < k\).

The following fact is evident if we are sure that \(f_n\) is indeed the orthogonal projection of \(f\) . . . but let us prove it anyway.

4.1 Lemma. Let \(n < k\), and \(\varphi : (0,1) \to \mathbb{R}\) be a \(2^{-k}\)-periodic bounded Borel function. Then

\[
\int_{0}^{1} \varphi(x)f_n(x) \, dx = \int_{0}^{1} \varphi(x)f_k(x) \, dx.
\]

Proof.

\[
\int_{0}^{1} \varphi(x)f_n(x) \, dx = 2^n \int_{0}^{2^{-n}} \varphi(x)f_n(x) \, dx = 2^n \sum_{j=1}^{2^{k-n}} \int_{(j-1)2^{-k}}^{j2^{-k}} \varphi(x)f_n(x) \, dx = 2^n \int_{0}^{2^{-k}} \varphi(x) \left( \frac{1}{2} \sum_{j=1}^{2^{k-n}} f_n(x + j2^{-k}) \right) \, dx = 2^n \int_{0}^{2^{-k}} \varphi(x)f_k(x) \, dx = \int_{0}^{1} \varphi(x)f_k(x) \, dx.
\]
Treating $(0, 1)$ with Lebesgue measure as a probability space and $f_n, \varphi$ as random variables, we have

$$E(\varphi f_n) = E(\varphi f_k).$$

In terms of (non-elementary!) conditioning (and binary digits),

$$f_n = g_n(\beta_{n+1}, \beta_{n+2}, \ldots), \quad \varphi = \psi(\beta_{k+1}, \beta_{k+2}, \ldots);$$

$$E(\psi(\beta_{k+1}, \ldots) g_n(\beta_{n+1}, \ldots)) = E(E(\psi(\beta_{k+1}, \ldots) g_n(\beta_{n+1}, \ldots) | \beta_{k+1}, \ldots)) =$$

$$E(\psi(\beta_{k+1}, \ldots) E(g_n(\beta_{n+1}, \ldots) | \beta_{k+1}, \ldots)) = E(\psi(\beta_{k+1}, \ldots) g_k(\beta_{k+1}, \ldots)).$$

**4.2 Exercise.** $E(\varphi f_n^2) \geq E(\varphi f_k^2)$ for $\varphi(\cdot) \geq 0$.

Prove it.

In fact, $E(\varphi \psi(f_n)) \geq E(\varphi \psi(f_k))$ for convex $\psi$.

**4.3 Lemma.**

$$P\left(\max_{k=n, \ldots, n+m} |f_k| \geq c\right) \leq \frac{1}{c^2} E f_n^2.$$

**Proof.** We introduce events $A_k = \{|f_k| \geq c, |f_{k+1}| < c, \ldots, |f_{n+m}| < c\}$ and apply [4.2] to their indicators:

$$E(1_{A_k} f_n^2) \geq E(1_{A_k} f_k^2) \geq c^2 P(A_k).$$

Summing up we get

$$E(1_A f_n^2) \geq c^2 P(A)$$

where $A = A_1 \cup \cdots \cup A_n = \{\max_{k=n, \ldots, n+m} |f_k| \geq c\}$. \hfill \qed

It follows that

$$P\left(\sup_{k \geq n} |f_k| \geq c\right) \leq \frac{1}{c^2} E f_n^2.$$

**4.4 Exercise.** The sequence $(f_n)_n$ converges almost everywhere.

Prove it.

Hint: similar to 3.6.