1 Kolmogorov's maximal inequality

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

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

Proof. Denoting by μ_k the distribution of X_k we have $\int x \, \mu_k(\mathrm{d}x) = 0$, thus

$$\mathbb{E}\left(\varphi(X_1, \dots, X_k) S_n\right) = \int \mu_1(\mathrm{d}x_1) \dots \mu_n(\mathrm{d}x_n) \, \varphi(x_1, \dots, x_k)(x_1 + \dots + x_n) =$$

$$\int \mu_1(\mathrm{d}x_1) \dots \mu_k(\mathrm{d}x_k) \, \varphi(x_1, \dots, x_k) \int \mu_{k+1}(\mathrm{d}x_{k+1}) \dots \mu_n(\mathrm{d}x_n) \, (x_1 + \dots + x_n)$$

$$= \int \mu_1(\mathrm{d}x_1) \dots \mu_k(\mathrm{d}x_k) \, \varphi(x_1, \dots, x_k)(x_1 + \dots + x_k) = \mathbb{E}\left(\varphi(X_1, \dots, X_k) S_k\right).$$

$\square$

In terms of conditioning,

$$\mathbb{E}\left(\varphi(X_1, \dots, X_k) S_n\right) = \mathbb{E}\left(\mathbb{E}\left(\varphi(X_1, \dots, X_k) S_n \,\middle|\, X_1, \dots, X_k\right)\right) =$$
$$= \mathbb{E}\left(\varphi(X_1, \dots, X_k) \mathbb{E}\left(S_n \,\middle|\, X_1, \dots, X_k\right)\right) = \mathbb{E}\left(\varphi(X_1, \dots, X_k) S_k\right).$$

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

Prove it.

Hint: $\int \mu_{k+1}(\mathrm{d}x_{k+1}) \dots \mu_n(\mathrm{d}x_n)(x_1 + \dots + x_n)^2 \ge$
$\left(\int \mu_{k+1}(\mathrm{d}x_{k+1}) \dots \mu_n(\mathrm{d}x_n)(x_1 + \dots + x_n)\right)^2$.

1.3 Remark. More generally, the Jensen inequality gives $\mathbb{E}\left(\varphi(X_1, \dots, X_k)\psi(S_n)\right) \ge \mathbb{E}\left(\varphi(X_1, \dots, X_k)\psi(S_k)\right)$ 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 $c > 0$,

$$\mathbb{P}\left(\max_{k=1,\dots,n} |S_k| \ge c\right) \le \frac{1}{c^2} \mathbb{E}\, S_n^2.$$

Proof. We introduce events $A_k = \{|S_1| < c, \dots, |S_{k-1}| < c, |S_k| \ge c\}$ and apply 1.2 to their indicators:

$$\mathbb{E}\left(\mathbf{1}_{A_k} S_n^2\right) \ge \mathbb{E}\left(\mathbf{1}_{A_k} S_k^2\right) \ge c^2 \mathbb{P}\left(A_k\right).$$

Summing up we get

$$\mathbb{E}\left(\mathbf{1}_A S_n^2\right) \ge c^2 \mathbb{P}\left(A\right)$$

where $A = A_1 \uplus \dots \uplus A_n = \{\max_k |S_k| \ge c\}$. □

Clearly, $\mathbb{E}\, S_n^2 = \sum_{k=1}^n \operatorname{Var} X_k$.

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

$$\mathbb{P}\left(\sup_k |S_k| \ge c\right) \le \frac{1}{c^2} \sum_{k=1}^{\infty} \operatorname{Var} X_k .$$

Prove it.

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

2 Random series

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

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

Then the series

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

converges a.s.

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

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

or equivalently,

$$\mathbb{P}\left(\sup_{k,l \ge n} |S_k - S_l| \ge 2\varepsilon\right) \downarrow 0 \quad \text{as } n \to \infty$$

for every $\varepsilon > 0$. Using 1.5,

$$\mathbb{P}\Big(\sup_{k,l\ge n} |S_k - S_l| \ge 2\varepsilon \Big) \le \mathbb{P}\Big(\sup_k |S_{n+k} - S_n| \ge \varepsilon \Big) =$$
$$= \mathbb{P}\Big(\sup_k |X_{n+1} + \dots + X_{n+k}| \ge \varepsilon \Big) \le \frac{1}{\varepsilon^2} \sum_k \operatorname{Var} X_{n+k} \downarrow 0$$

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,

$$f_n(x) = 2^n \int_{2^{-n}(k-1)}^{2^{-n}k} f(u)\,\mathrm{d}u \quad \text{for } x \in \big(2^{-n}(k-1), 2^{-n}k\big)\,.$$

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

$$x = \frac{\beta_1(x)}{2^1} + \frac{\beta_2(x)}{2^2} + \dots\,, \quad \beta_k(x) \in \{0,1\}\,,$$

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

$$g_k(b_1,\dots,b_k) = \tfrac12 g_{k+1}(b_1,\dots,b_k,0) + \tfrac12 g_{k+1}(b_1,\dots,b_k,1)$$

and moreover,

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

for $k < n$.

Treating $(0,1)$ with Lebesgue measure as a probability space and $\beta_1, \beta_2, \dots$ as random variables we see that $\beta_1, \beta_2, \dots$ are independent, $\mathbb{P}\big(\beta_k = 0\big) = 0.5 = \mathbb{P}\big(\beta_k = 1\big)$, and the random variables $f_n = g_n(\beta_1, \dots, \beta_n)$ satisfy

$$\mathbb{E}\big(f_n \,\big|\, \beta_1, \dots, \beta_k \big) = f_k \quad \text{for } k < n\,.$$

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. $\mathbb{E}\,\big(\varphi(\beta_1,\dots,\beta_k)f_n\big) = \mathbb{E}\,\big(\varphi(\beta_1,\dots,\beta_k)f_k\big)$ for every $k<n$ and every function $\varphi : \{0,1\}^k \to \mathbb{R}$.

Proof.

$$\begin{aligned}\mathbb{E}\,\big(\varphi(\beta_1,\dots,\beta_k)f_n\big) &= 2^{-n}\sum_{b_1,\dots,b_n}\varphi(b_1,\dots,b_k)g_n(b_1,\dots,b_n) = \\ &= 2^{-k}\sum_{b_1,\dots,b_k}\varphi(b_1,\dots,b_k)2^{-(n-k)}\sum_{b_{k+1},\dots,b_n} g_n(b_1,\dots,b_n) = \\ &= 2^{-k}\sum_{b_1,\dots,b_k}\varphi(b_1,\dots,b_k)g_k(b_1,\dots,b_k) = \mathbb{E}\,\big(\varphi(\beta_1,\dots,\beta_k)f_k\big)\,.\end{aligned}$$

□

In terms of conditioning,

$$\begin{aligned}\mathbb{E}\,\big(\varphi(\beta_1,\dots,\beta_k)f_n\big) &= \mathbb{E}\,\Big(\mathbb{E}\,\big(\varphi(\beta_1,\dots,\beta_k)f_n\,\big|\,\beta_1,\dots,\beta_k\big)\Big) = \\ &= \mathbb{E}\,\Big(\varphi(\beta_1,\dots,\beta_k)\mathbb{E}\,\big(f_n\,\big|\,\beta_1,\dots,\beta_k\big)\Big) = \mathbb{E}\,\big(\varphi(\beta_1,\dots,\beta_k)f_k\big)\,.\end{aligned}$$

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

Prove it.

Hint: similar to 1.2.

In fact, $\mathbb{E}\,\big(\varphi(\beta_1,\dots,\beta_k)\psi(f_n)\big) \ge \mathbb{E}\,\big(\varphi(\beta_1,\dots,\beta_k)\psi(f_k)\big)$ for convex ψ.

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

$$\mathbb{P}\,\Big(\max_{k=1,\dots,n}|f_k|\ge c\Big) \le \frac{1}{c^2}\mathbb{E}\,f_n^2\,.$$

Prove it.

Hint: similar to 1.4.

3.4 Exercise. For every $c>0$,

$$\mathbb{P}\,\Big(\sup_k|f_k|\ge c\Big) \le \frac{1}{c^2}\sup_k\mathbb{E}\,f_k^2\,.$$

Prove it.

Hint: similar to 1.5.

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

$$\mathbb{P}\,\Big(\sup_k|f_{n+k}-f_n|\ge c\Big) \le \frac{1}{c^2}\sup_k\mathbb{E}\,|f_{n+k}-f_n|^2\,. \tag{3.5}$$

3.6 Proposition. The sequence $(f_n)_n$ converges almost everywhere.

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

$$\|f_0\|^2 + \|f_1 - f_0\|^2 + \dots + \|f_n - f_{n-1}\|^2 = \|f_n\|^2 \le \|f\|^2 \,.$$

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), $\mathbb{P}\big(\sup_k |f_{n+k} - f_n| \ge \varepsilon\big) \to 0$ 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. $\square$

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

$$f_k(x) = \tfrac12 f_{k-1}(x) + \tfrac12 f_{k-1}(x + 2^{-k})$$

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 $f\dots$ 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

$$\int_0^1 \varphi(x) f_n(x)\,\mathrm{d}x = \int_0^1 \varphi(x) f_k(x)\,\mathrm{d}x \,.$$

Proof.

$$\int_0^1 \varphi(x) f_n(x)\,\mathrm{d}x = 2^n \int_0^{2^{-n}} \varphi(x) f_n(x)\,\mathrm{d}x = 2^n \sum_{j=1}^{2^{k-n}} \int_{(j-1)2^{-k}}^{j2^{-k}} \varphi(x) f_n(x)\,\mathrm{d}x =$$

$$= 2^k \int_0^{2^{-k}} \varphi(x) \bigg(2^{-(k-n)} \sum_{j=1}^{2^{k-n}} f_n(x + j2^{-k}) \bigg)\,\mathrm{d}x =$$

$$= 2^k \int_0^{2^{-k}} \varphi(x) f_k(x)\,\mathrm{d}x = \int_0^1 \varphi(x) f_k(x)\,\mathrm{d}x \,.$$

$\square$

Treating $(0,1)$ with Lebesgue measure as a probability space and f_n, φ as random variables, we have

$$\mathbb{E}\left(\varphi f_n\right) = \mathbb{E}\left(\varphi f_k\right).$$

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

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

$$\mathbb{E}\left(\psi(\beta_{k+1}, \dots) g_n(\beta_{n+1}, \dots)\right) = \mathbb{E}\left(\mathbb{E}\left(\psi(\beta_{k+1}, \dots) g_n(\beta_{n+1}, \dots) \,\middle|\, \beta_{k+1}, \dots\right)\right) =$$
$$= \mathbb{E}\left(\psi(\beta_{k+1}, \dots) \mathbb{E}\left(g_n(\beta_{n+1}, \dots) \,\middle|\, \beta_{k+1}, \dots\right)\right) = \mathbb{E}\left(\psi(\beta_{k+1}, \dots) g_k(\beta_{k+1}, \dots)\right).$$

4.2 Exercise. $\mathbb{E}\left(\varphi f_n^2\right) \ge \mathbb{E}\left(\varphi f_k^2\right)$ for $\varphi(\cdot) \ge 0$.
Prove it.

In fact, $\mathbb{E}\left(\varphi \psi(f_n)\right) \ge \mathbb{E}\left(\varphi \psi(f_k)\right)$ for convex ψ.

4.3 Lemma.
$$\mathbb{P}\left(\max_{k=n,\dots,n+m} |f_k| \ge c\right) \le \frac{1}{c^2} \mathbb{E}\, f_n^2.$$

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

$$\mathbb{E}\left(\mathbf{1}_{A_k} f_n^2\right) \ge \mathbb{E}\left(\mathbf{1}_{A_k} f_k^2\right) \ge c^2 \mathbb{P}\left(A_k\right).$$

Summing up we get
$$\mathbb{E}\left(\mathbf{1}_A f_n^2\right) \ge c^2 \mathbb{P}\left(A\right)$$
where $A = A_1 \uplus \dots \uplus A_n = \{\max_{k=n,\dots,n+m} |f_k| \ge c\}$. $\square$

It follows that
$$\mathbb{P}\left(\sup_{k \ge n} |f_k| \ge c\right) \le \frac{1}{c^2} \mathbb{E}\, f_n^2.$$

4.4 Exercise. The sequence $(f_n)_n$ converges almost everywhere.
Prove it.
Hint: similar to 3.6.