Contents

3 Infinite independent sequences

3a Independent events

Continuous probability spaces are needed here; triangle arrays do not help.

3a1 Definition. (a) Events $A_1, A_2, \dots$ are independent if for every n the events $A_1, \dots, A_n$ are independent;

(b) random variables $X_1, X_2, \dots$ are independent if for every n the random variables $X_1, \dots, X_n$ are independent;

(c) σ-algebras $\mathcal{A}_1, \mathcal{A}_2, \dots$ are independent if for every n the σ-algebras $\mathcal{A}_1, \dots, \mathcal{A}_n$ are independent.

The relation

$$\mathbb{P}\big(X_1 \in B_1, X_2 \in B_2, \dots\big) = \mathbb{P}\big(X_1 \in B_1\big)\mathbb{P}\big(X_2 \in B_2\big)\dots$$

holds for independent random variables X_n and Borel sets $B_n \subset \mathbb{R}$, but is of little use.

3a2 Exercise. Let $(\Omega, \mathcal{F}, P)$ be $(0,1)$ with Lebesgue measure, and $\beta_1, \beta_2, \dots : \Omega \to \{0,1\}$ binary digits;

$$\omega = \sum_{n=1}^{\infty} \frac{\beta_n(\omega)}{2^n}\,, \quad \liminf_n \beta_n(\omega) = 0\,.$$

Then β_n are independent random variables; also, $\beta_n = 1\!\!1_{A_n}$, and A_n are independent events of probability 0.5 each ("a fair coin tossed endlessly").

Prove it.

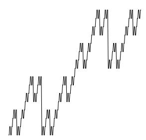
Treating β_n as random variables we observe that the random variable $U = \sum_n 2^{-n}\beta_n$ is distributed uniformly on $(0,1)$, that is, $F_U(u) = u$ for $0 \le u \le 1$.

We introduce random variables

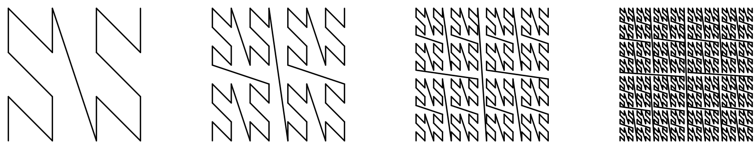
$$U_1 = \sum_{n=1}^{\infty} 2^{-n}\beta_{2n-1}\,, \quad U_2 = \sum_{n=1}^{\infty} 2^{-n}\beta_{2n}\,.$$

For each n the random vector $(\beta_1, \beta_3, \dots, \beta_{2n-1})$ is distributed like $(\beta_1, \beta_2, \dots, \beta_n)$; therefore $\sum_{k=1}^n 2^{-k}\beta_{2k-1}$ is distributed like $\sum_{k=1}^n 2^{-k}\beta_k$, that is, $F_{U_1}(u) = F_U(u)$ whenever u is dyadic (that is, of the form $k/2^n$); it follows that $F_{U_1} = F_U$. We see that U_1 is distributed uniformly on $(0,1)$; the same holds for U_2.

For each n the random vectors $(\beta_1, \beta_3, \dots, \beta_{2n-1})$ and $(\beta_2, \beta_4, \dots, \beta_{2n})$ are independent (think, why), therefore $F_{U_1,U_2}(u_1, u_2) = F_{U_1}(u_1)F_{U_2}(u_2)$ for all dyadic u_1, u_2, and for arbitrary u_1, u_2 as well. We see that U_1, U_2 are independent.[1]



Graph of U_1



Approximating the curve $\{(U_1(\omega), U_2(\omega)) : 0 < \omega < 1\}$

Similarly we may introduce $U_1, U_2, \dots$ by

$$U_n = \sum_{k=1}^{\infty} 2^{-k}\beta_{2^{n-1}(2k-1)}$$

and check that these are an infinite sequence of independent random variables, each distributed uniformly on $(0,1)$.[2]

Now, given $p_1, p_2, \dots \in [0,1]$, we may consider events $A_n = \{U_n \le p_n\}$ and check that they are independent, and $\mathbb{P}\big(A_n\big) = p_n$.

Let events $A_1, A_2, \dots$ be independent. The sum $S = \sum_{k=1}^{\infty} 1\!\!1_{A_k}$, the random number of occurred events, can be finite or infinite.

3a3 Theorem. [3] (a) If $\sum_{k=1}^{\infty} \mathbb{P}\big(A_k\big) < \infty$ then $S < \infty$ almost surely;
(b) if $\sum_{k=1}^{\infty} \mathbb{P}\big(A_k\big) = \infty$ then $S = \infty$ almost surely.

These (a) and (b) are called Borel-Cantelli lemmas. Independence matters for (b) but not (a). For independent events, $\mathbb{P}\big(S < \infty\big)$ is either 0 or 1, which is a special case of Kolmogorov's 0–1 law.

3a4 Exercise. Let $U_1, U_2, \dots$ be independent random variables, each distributed uniformly on $(-1,1)$. Then

[1] Instead of dyadic numbers and CDF we could use dyadic algebra; it generates the Borel σ-algebra.

[2] [W, Sect. 4.6].

[3] [KS, Sect. 7.1, Lemmas 7.3, 7.4]; [D, Sect. 1.6, (6.1) and (6.6)].

(a) the sequence $(nU_n)_{n=1}^{\infty}$ is dense in $\mathbb{R}$ a.s.;

(b) the sequence $(n^2U_n)_{n=1}^{\infty}$ is not dense, and moreover, $n^2|U_n| \to \infty$ a.s. Prove it.

If A_k are independent, $\mathbb{P}\big(A_k\big) \to 0$ but $\sum_k \mathbb{P}\big(A_k\big) = \infty$, then the indicators $X_k = 1\!\!1_{A_k}$ converge to 0 in $L_2(\Omega)$ but not almost surely; moreover, $\limsup_k X_k(\omega) = 1$ for almost all $\omega \in \Omega$. (There is a simpler, non-probabilistic example on $\Omega = (0,1)$.)

Proof of 3a3(a) (the first Borel-Cantelli lemma)

$$S_n = \sum_{k=1}^{n} 1\!\!1_{A_k} ; \quad \mathbb{E}\, S_n = p_1 + \dots + p_n \,, \quad p_k = \mathbb{P}\big(A_k\big) ;$$

$$\mathbb{P}\big(S_n > M\big) \uparrow \mathbb{P}\big(S > M\big) \quad \text{as } n \to \infty; \qquad (\text{wrong for “}\ge\text{”!})$$

$$\mathbb{P}\big(S_n > M\big) \le \frac{\mathbb{E}\, S_n}{M} = \frac{p_1 + \dots + p_n}{M} \le \frac{1}{M} \sum_k \mathbb{P}\big(A_k\big) ;$$

$$\mathbb{P}\big(S > M\big) \le \frac{1}{M} \sum_k \mathbb{P}\big(A_k\big) \downarrow 0 \quad \text{as } M \to \infty.$$

Another proof: the sequence S_n is increasing and $\mathbb{E}\, S_n$ is bounded, therefore $S_n \uparrow S < \infty$ a.s.

End of proof of 3a3(a) (the first Borel-Cantelli lemma)

Proof of 3a3(b) (the second Borel-Cantelli lemma)

(Clearly $\mathbb{E}\, S = \infty$, but we need much more...)

$$\mathbb{P}\big(S_n \le M\big) = \mathbb{P}\big(\mathrm{e}^{-S_n} \ge \mathrm{e}^{-M}\big) \le \frac{\mathbb{E}\,\mathrm{e}^{-S_n}}{\mathrm{e}^{-M}} = \mathrm{e}^{M} \prod_{k=1}^{n} \underbrace{\big(p_k \cdot \mathrm{e}^{-1} + (1-p_k)\cdot 1\big)}_{1-(1-\mathrm{e}^{-1})p_k} \le$$

$$\le \mathrm{e}^{M} \exp\Big(-\sum_{k=1}^{n} p_k \cdot (1-\mathrm{e}^{-1})\Big) \downarrow 0 \quad \text{as } n \to \infty; \qquad (\text{since } 1-\varepsilon \le \mathrm{e}^{-\varepsilon})$$

$$\mathbb{P}\big(S \le M\big) = 0 \quad \text{for all } M.$$

Another proof: $\mathbb{E}\,\mathrm{e}^{-S_n} \to 0$ (as before); $\mathbb{E}\,\mathrm{e}^{-S} \le \mathbb{E}\,\mathrm{e}^{-S_n}$; $\mathbb{E}\,\mathrm{e}^{-S} = 0$.

End of proof of 3a3(b) (the second Borel-Cantelli lemma)

Let A_k be equiprobable, of probability p each ("unfair coin").

3a5 Proposition. $\frac{1}{n}\big(\mathbb{1}_{A_1} + \dots + \mathbb{1}_{A_n}\big) \to p$ (as $n \to \infty$) almost surely.

This is a special case of the Strong Law of Large Numbers (see 3b2), but also of the following fact (less general and much simpler to prove).

3a6 Proposition. (Borel's strong law of large numbers) Let X_n be independent, identically distributed random variables such that $\mathbb{E}\, X_1^4 < \infty$, then $\frac{1}{n}(X_1 + \dots + X_n) \to \mathbb{E}\, X_1$ almost surely.

3a7 Exercise. (a) It is sufficient to prove 3a6 for $\mathbb{E}\, X_1 = 0$.
Let X_n be as in 3a6.

(b) If $\mathbb{E}\, X_1 = 0$ then $\mathbb{E}\,(X_1 + \dots + X_n)^4 \sim 3n^2 \big(\mathbb{E}\, X_1^2\big)^2$.
Prove it.

Proof of 3a6. Assuming $\mathbb{E}\, X_1 = 0$ and denoting $S_n = X_1 + \dots + X_n$ we have $\mathbb{E}\,\big(\frac{S_n}{n}\big)^4 = O(\frac{1}{n^2})$; $\sum_n \mathbb{E}\,\big(\frac{S_n}{n}\big)^4 < \infty$; $\sum_n \big(\frac{S_n}{n}\big)^4 < \infty$ a.s.; $\big(\frac{S_n}{n}\big)^4 \to 0$ a.s.; $\frac{S_n}{n} \to 0$ a.s. □

3a8 Exercise. For S_n as in 1a5 (the simple random walk),

$$S_n = o(n) \quad \text{a.s.}$$

Prove it.

Compare it with 1a5. Convergence in L_2 is rather evident, but almost everywhere convergence is not.

A real number $x \in (0,1)$ is called 10-normal, if its decimal digits $\alpha_1, \alpha_2, \dots$ defined by

$$x = \frac{\alpha_1}{10} + \frac{\alpha_2}{10^2} + \dots; \qquad \alpha_1, \alpha_2, \dots \in \{0,1,2,3,4,5,6,7,8,9\}$$

have equal frequencies, that is,

$$\frac{\#\{k \in [1,n] : \alpha_k = a\}}{n} \to \frac{1}{10} \quad \text{as } n \to \infty \quad \text{(for all } a\text{)}$$

and moreover, their combinations have equal frequencies, that is,

$$\frac{\#\{k \in [1,n] : \alpha_k = a_1, \alpha_{k+1} = a_2, \dots, \alpha_{k+l-1} = a_l\}}{n} \to \frac{1}{10^l} \quad \text{as } n \to \infty$$

for all $a_1, \dots, a_l$ and all l. Similarly, p-normal numbers are defined for any $p = 2, 3, \dots$. Finally, x is called normal, if it is p-normal for all p.

3a9 Proposition. Normal numbers exist.

Proposition 3a9 follows from Proposition 3a10.

3a10 Proposition. [1] Almost all numbers are normal.

That is, the set of all normal numbers is Lebesgue measurable, and its Lebesgue measure is equal to 1. This is Borel's normal number theorem (1909).

Proof. It suffices to treat a single base, for instance 10, and a single combination of digits, for instance "71":

$$\frac{\#\{k \in [1,n] : \alpha_k = 7, \alpha_{k+1} = 1\}}{n} \to \frac{1}{100} \quad \text{as } n \to \infty . \qquad (?)$$

Splitting these k into even and odd numbers we note that it suffices to treat the two cases separately; for instance, the odd case:

$$\frac{\#\{k : 2k-1 \le n, \alpha_{2k-1} = 7, \alpha_{2k} = 1\}}{n} \to \frac{1}{200} , \qquad (?)$$

or equivalently,

$$\frac{\#\{k : 2k-1 \le n, \alpha_{2k-1} = 7, \alpha_{2k} = 1\}}{\#\{k : 2k-1 \le n\}} \to \frac{1}{100} \quad \text{as } n \to \infty ,$$

which is a special case of 3a5. □

Do not think that the normality exhausts probabilistic properties of (digits of) real numbers.

3a11 Proposition. The series

$$\sum_{n=1}^{\infty} \frac{2\beta_n - 1}{n}$$

converges for almost all $x \in (0,1)$. (Here $\beta_1, \beta_2, \dots$ are the binary digits of x.)

By the way, the sum of the series, $f(x) = \sum \frac{(-1)^{\beta_n}}{n}$, is a terrible (but measurable) function. Especially,

$$\operatorname{mes}\{x \in (a,b) : f(x) \in (c,d)\} > 0$$

for *all* intervals $(a,b) \subset (0,1)$, $(c,d) \subset \mathbb{R}$ (here 'mes' stands for the Lebesgue measure). Surely we cannot draw its graph!

Here is a probabilistic counterpart of 3a11.

[1] [D, Sect. 6.2, Example 2.5].

3a12 Proposition. The series $\frac{X_1}{1} + \frac{X_2}{2} + \frac{X_3}{3} + \dots$ converges almost surely. (Here $X_1, X_2, \dots$ are independent random signs.)

Convergence in L_2 is rather evident, but almost everywhere convergence is not. See also 3b3 and 3b8.

Propositions 3a11 and 3a12 will be proved in Section 3b.

3a13 Proposition. A measurable function $f : \mathbb{R} \to \mathbb{R}$ satisfying $f(x + 2^{-n}) = f(x)$ for all $x \in \mathbb{R}$ and $n = 1, 2, \dots$ is constant almost everywhere.

In other words: there exists $a \in \mathbb{R}$ such that $f(x) = a$ for almost all x. (It need not hold for *all* x.) This is an analytical counterpart of the following probabilistic fact.

3a14 Proposition. Let $X_1, X_2, \dots$ be independent random signs, and a random variable Y be of the form $Y = f_n(X_n, X_{n+1}, \dots)$ for all n. Then Y is constant a.s.

This is a special case of Kolmogorov's 0–1 law (see 3b7).

3b Independent random variables

Let F and F_n be as in 1c2.

3b1 Theorem. [1]

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0 \quad \text{amost surely, as } n \to \infty \, .$$

This is the (strong form of) Glivenko-Cantelli theorem.

Proof. By 1c3, for every $\varepsilon > 0$ there exist m and $t_1 < \dots < t_m$ such that

$$\mu\big((-\infty, t_1)\big) \le \varepsilon \, , \; \mu\big((t_1, t_2)\big) \le \varepsilon \, , \; \dots \, , \; \mu\big((t_{m-1}, t_m)\big) \le \varepsilon \, , \; \mu\big((t_m, +\infty)\big) \le \varepsilon \, .$$

Similarly to the proof of 1c2, if $|\mu_n\big((-\infty, t_k]\big) - \mu\big((-\infty, t_k]\big)| \le \varepsilon$ and $|\mu_n\big((-\infty, t_k)\big) - \mu\big((-\infty, t_k)\big)| \le \varepsilon$ for all k then $\sup_t |F_n(t) - F(t)| \le 3\varepsilon$.

By 3a5, this happens eventually (almost surely), for every $\varepsilon > 0$ separately. Therefore, almost surely it holds for all $\varepsilon > 0$ simultaneously. $\square$

[1] [D, Sect. 1.7, (7.4)].

Strong law of large numbers

3b2 Theorem. [1] Let $X_1, X_2, \dots$ be independent identically distributed random variables. If $\mathbb{E}\,|X_1| < \infty$ then

$$\frac{X_1 + \dots + X_n}{n} \to \mathbb{E}\,X_1 \quad \text{a.s. as } n \to \infty\,.$$

This is the Strong Law of Large Numbers. Compare it with 1c1, 3a5 and 3a6. It appears that 3b2 is much harder to prove.

3b3 Proposition. [2] (*Kolmogorov*) Suppose $X_1, X_2, \dots$ are independent random variables with $\mathbb{E}\,X_n = 0$. If $\sum \operatorname{Var}(X_n) < \infty$ then the series $\sum X_n$ converges almost surely.

Postponing the proof of 3b3 we first show that it implies 3b2.

Before treating random series, recall convergence of series $\sum a_n$ of real numbers ($a_n \in \mathbb{R}$), and do not confuse it with convergence of positive series ($a_n > 0$); do not write $\sum a_n < \infty$ instead of "$\sum a_n$ converges", and note that $\sum a_n$ can converge while $\sum b_n$ diverge even if $a_n/b_n \to 1$.

3b4 Lemma. (Kronecker) If $x_n \in \mathbb{R}$ are such that $\sum \frac{x_n}{n}$ converges then $\frac{x_1+\dots+x_n}{n} \to 0$.

Proof. (sketch) In terms of $y_n = \frac{x_n}{n}$, $x_n = ny_n$, it takes the form

$$\text{if } \sum y_n \text{ converges then } \frac{1}{n}y_1 + \frac{2}{n}y_2 + \dots + \frac{n}{n}y_n \to 0\,.$$

In terms of $S_n = y_1 + \dots + y_n$, $y_n = S_n - S_{n-1}$, it takes the form

$$\text{if } S_n \to S \text{ then } S_n - \frac{1}{n}(S_1 + \dots + S_{n-1}) \to 0\,,$$

which is easy to check. □

Proof of 3b2 (strong law of large numbers) assuming 3b3 (to be proved later)

By the first Borel-Cantelly lemma 3a3(a), $|X_n| \le n$ eventually, since $\sum_{n=1}^\infty \mathbb{P}\big(|X_n| > n\big) = \sum_{n=1}^\infty \mathbb{P}\big(|X_1| > n\big) \le \mathbb{E}\,|X_1| < \infty$.

[1] [D, Sect. 1.7, Items (7.1) and (8.6)]; [KS, Sect. 7.2, Th. 7.7].
[2] [D, Sect. 1.8, Th. (8.3)].

We introduce $Y_n = X_n \cdot 1\!\!1_{[-n,n]}(X_n)$ and note that $\frac{Y_1+\dots+Y_n}{n} - \frac{X_1+\dots+X_n}{n} \to 0$ almost surely, since $X_n - Y_n \to 0$ almost surely. Thus it is sufficient to prove that $\frac{Y_1+\dots+Y_n}{n} \to \mathbb{E}\, X_1$ a.s.

We introduce $Z_n = Y_n - \mathbb{E}\, Y_n$ and note that $\mathbb{E}\, Y_n = \mathbb{E}\,\big(X_1 \cdot 1\!\!1_{[-n,n]}(X_1)\big) \to \mathbb{E}\, X_1$, therefore $\frac{Y_1+\dots+Y_n}{n} - \frac{Z_1+\dots+Z_n}{n} = \frac{\mathbb{E}\,Y_1+\dots+\mathbb{E}\,Y_n}{n} \to \mathbb{E}\, X_1$. Thus it is sufficient to prove that $\frac{Z_1+\dots+Z_n}{n} \to 0$ a.s.

By 3b4, it is sufficient to prove that $\sum \frac{Z_n}{n}$ converges almost surely.

By 3b3, it is sufficient to prove that $\sum \operatorname{Var}\big(\frac{Z_n}{n}\big) < \infty$.

We have $\operatorname{Var} Z_n = \operatorname{Var} Y_n \le \mathbb{E}\, Y_n^2$; it remains to prove that $\sum \frac{1}{n^2} \mathbb{E}\, Y_n^2 < \infty$.

In fact, $\sum_{n=1}^\infty \frac{1}{n^2} \mathbb{E}\, Y_n^2 \le 2\mathbb{E}\, |X_1|$, since $\forall y \quad \sum_{k=1}^\infty \frac{y^2}{k^2} \cdot 1\!\!1_{[-k,k]}(y) \le 2|y|$. Indeed, for $y \in (n-1, n]$ we have

$$\sum_{k=n}^\infty \frac{1}{k^2} y^2 \le 2|y| \quad \Longleftarrow \quad \sum_{k=n}^\infty \frac{1}{k^2} \le \frac{2}{n} \le \frac{2}{y} \quad \Longleftarrow$$
$$\frac{1}{n^2} + \sum_{k=n+1}^\infty \frac{1}{k^2} \le \frac{1}{n^2} + \int_n^\infty \frac{\mathrm{d}x}{x^2} = \frac{1}{n^2} + \frac{1}{n} \le \frac{2}{n}.$$

End of proof of 3b2 assuming 3b3

Kolmogorov's maximal inequality

The following result is needed for 3b3.

3b5 Proposition. Let $X_1, \dots, X_n$ be independent random variables, $\mathbb{E}\, X_k = 0$ and $\mathbb{E}\, X_k^2 < \infty$ for $k = 1, \dots, n$. Then, for every $c > 0$,

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

(Here $S_n = X_1 + \dots + X_n$.)

3b6 Remark. Evidently, $\mathbb{P}\big(|S_k| \ge c\big) \le \frac{\mathbb{E}\, S_k^2}{c^2} \le \frac{\mathbb{E}\, S_n^2}{c^2}$, thus, $\max_k \mathbb{P}\big(|S_k| \ge c\big) \le \frac{\mathbb{E}\, S_n^2}{c^2}$. However, Kolmogorov's result is much stronger! Also, evidently, $\mathbb{P}\big(\max_k |S_k| \ge c\big) \le \sum_k \mathbb{P}\big(|S_k| \ge c\big) \le \frac{1}{c^2} \sum_k \mathbb{E}\, S_k^2$, but it does not help: the latter may grow as n (try $X_2 = X_3 = \dots = 0$).

Here is the first proof, for the discrete case; it shows the idea[1] used afterwards in the second, general proof. (We do it for the quadratic function, but the proofs work for every convex function.)

$$\mathbb{E}\big(S_n \,\big|\, X_1,\dots,X_k\big) = S_k\,, \quad \text{thus} \quad \mathbb{E}\big(S_n^2 \,\big|\, X_1,\dots,X_k\big) \ge S_k^2\,;$$

$$\text{(by conditional Jensen, or just conditional } \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2 \ge 0)$$

$$\text{introduce disjoint events} \quad A_k = \{|S_1| < c, \dots, |S_{k-1}| < c, |S_k| \ge c\}\,;$$

$$\mathbb{E}\big(S_n^2 \,\big|\, A_k\big) \ge c^2\,; \quad \mathbb{E}\left(S_n^2 \mathbb{1}_{A_k}\right) \ge c^2 \mathbb{P}\big(A_k\big)\,;$$

$$\mathbb{E}\left(S_n^2 \mathbb{1}_{A_1 \uplus \dots \uplus A_n}\right) \ge c^2 \mathbb{P}\big(A_1 \uplus \dots \uplus A_n\big)\,;$$

$$\mathbb{E}\,S_n^2 \ge c^2 \mathbb{P}\big(\max_k |S_k| \ge c\big)\,.$$

Here is the second (final) proof.

Proof of 3b5

We introduce disjoint events A_k as before and prove that $\mathbb{E}\left(S_n^2 \mathbb{1}_{A_k}\right) \ge c^2 \mathbb{P}\big(A_k\big)$ as follows. We have

$$\mathbb{E}\left(S_n^2 \mathbb{1}_{A_k}\right) = \int_{B_k \times \mathbb{R}^{n-k}} (x_1 + \dots + x_n)^2 \, \mu_1(\mathrm{d}x_1) \dots \mu_n(\mathrm{d}x_n)\,,$$

where

$$B_k = \{(x_1,\dots,x_k) : |x_1| < c, \dots, |x_1 + \dots + x_{k-1}| < c, |x_1 + \dots + x_k| \ge c\}\,;$$

we rewrite the integral as

$$\int_{B_k} \mu_1(\mathrm{d}x_1) \dots \mu_k(\mathrm{d}x_k) \int_{\mathbb{R}^{n-k}} (x_1 + \dots + x_n)^2 \, \mu_{k+1}(\mathrm{d}x_{k+1}) \dots \mu_n(\mathrm{d}x_n)\,;$$

taking into account that (for every a)

$$\int_{\mathbb{R}^{n-k}} \underbrace{(a + x_{k+1} + \dots + x_n)^2}_{=a^2 + 2a(x_{k+1} + \dots + x_n) + (x_{k+1} + \dots + x_n)^2} \mu_{k+1}(\mathrm{d}x_{k+1}) \dots \mu_n(\mathrm{d}x_n) \ge a^2$$

we get

$$\dots \ge \int_{B_k} \mu_1(\mathrm{d}x_1) \dots \mu_k(\mathrm{d}x_k) \underbrace{(x_1 + \dots + x_k)^2}_{\ge c^2} \ge c^2 \mathbb{P}\big(A_k\big)\,.$$

End of proof of 3b5

[1] This idea, "stopping", will be the tenor of Part 2 of the course.

Proof of 3b3

We'll prove that the partial sums S_n are a Cauchy sequence a.s., that is,

$$\lim_n \sup_{k,l\ge n} |S_k - S_l| = 0 \quad \text{a.s.}$$

These suprema, being a decreasing (in n) sequence, converge a.s.; in order to prove that their limit vanishes a.s. it is sufficient to prove that

$$\forall \varepsilon > 0 \quad \mathbb{P}\Big(\sup_{k,l\ge n} |S_k - S_l| > 2\varepsilon \Big) \xrightarrow[n\to\infty]{} 0 \,.$$

We have, using 3b5,

$$\mathbb{P}\Big(\sup_{k,l\ge n} |S_k - S_l| > 2\varepsilon \Big) \le \mathbb{P}\Big(\sup_{k\ge n} |S_k - S_n| > \varepsilon \Big) =$$

$$= \lim_m \underbrace{\mathbb{P}\Big(\max_{k=n,\dots,n+m} |S_k - S_n| > \varepsilon \Big)}_{\le \frac{1}{\varepsilon^2}\mathbb{E}\,(X_{n+1}^2+\dots+X_{n+m}^2)} \le \frac{1}{\varepsilon^2} \sum_{k=n+1}^{\infty} \operatorname{Var} X_k \xrightarrow[n\to\infty]{} 0 \,.$$

End of proof of 3b3

The proof of 3b2 (strong law of large numbers) is now complete.

Zero-one law

3b7 Proposition. Let $X_1, X_2, \dots$ be independent random variables, and a random variable Y be of the form $Y = f_n(X_n, X_{n+1}, \dots)$ for all n. Then Y is constant a.s.

This is a form of Kolmogorov's 0–1 law. (See also 3a14.) Basically, it holds because every measurable function of $X_1, X_2, \dots$ is approximately a measurable function of $X_1, \dots, X_n$ (see 3b14).

3b8 Exercise. Let $X_1, X_2, \dots$ be independent random variables, and $S_n = X_1 + \dots + X_n$. Then the following events are of probability 0 or 1 each:

S_n converge;
S_n are bounded;
S_n are bounded from above;
S_n are bounded from below.

Deduce it from 3b7.

Recall the σ-algebras generated by random variables: $\sigma(X)$, $\sigma(X,Y)$ etc.; $\sigma(X,Y)$ consists of sets of the form $\{\omega : (X(\omega),Y(\omega)) \in B\}$ for Borel $B \subset \mathbb{R}^2$. Rewriting $(X(\omega),Y(\omega)) \in B$ as $1\!\!1_B(X(\omega),Y(\omega)) = 1$ we see that a $\sigma(X,Y)$-measurable indicator function is of the form $\varphi(X,Y)$ where φ is a Borel measurable indicator function on $\mathbb{R}^2$. It follows (but not immediately) that the same holds for $\mathbb{R}$-valued (rather than $\{0,1\}$-valued) functions (the Doob-Dynkin lemma); this is why $\sigma(X,Y)$-measurable functions are often called measurable functions of X,Y. Similarly, $\sigma(X_1,X_2,\dots)$-measurable functions are often called measurable functions of $X_1,X_2,\dots$ Here $\sigma(X_1,X_2,\dots)$ is the least σ-algebra making all X_k measurable. Denoting $\mathcal{F}_n^\infty = \sigma(X_n,X_{n+1},\dots)$ we have

$$\mathcal{F}_n^\infty \downarrow \text{(the tail } \sigma\text{-algebra)} = \bigcap_n \mathcal{F}_n^\infty \,.$$

Measurability w.r.t. the tail σ-algebra is measurability w.r.t $\mathcal{F}_n^\infty$ for every n. It holds for Y of 3b7 and 3a14.

3b9 Proposition (Kolmogorov's 0-1 law)**.** [1] If X_n are independent then the tail σ-algebra is trivial.

Denoting $\mathcal{F}_1^n = \sigma(X_1,\dots,X_n)$ we have $\mathcal{F}_1^n \uparrow \sigma(X_1,X_2,\dots) = \mathcal{F}_1^\infty$ in the sense that $\mathcal{F}_1^\infty$ is the least σ-algebra that contains all $\mathcal{F}_1^n$. That is, $\mathcal{F}_1^\infty = \sigma(\mathcal{E})$ where $\mathcal{E} = \cup_n \mathcal{F}_1^n$.

3b10 Exercise. (a) $\mathcal{E}$ is an algebra;
(b) $\mathcal{E}$ need not be a σ-algebra.
Prove it.
Hint: (b) try binary digits.

By 1b6, $\mathcal{E}$ is dense in $\sigma(\mathcal{E})$, that is,

$$\inf_{E \in \mathcal{E}} P(A \triangle E) = 0 \quad \text{for all } A \in \sigma(\mathcal{E}) \tag{3b11}$$

whenever $\mathcal{E}$ is an algebra (not just $\cup_n \sigma(X_1,\dots,X_n)$).

3b12 Exercise. If a σ-algebra is independent of (all events of) an algebra $\mathcal{E}$ then it is independent of $\sigma(\mathcal{E})$.
Prove it.

Proof of Kolmogorov's 0-1 law. Independence of $\mathcal{F}_1^n$ and $\mathcal{F}_{n+1}^\infty$ implies independence of $\mathcal{F}_1^n$ and the tail σ-algebra for every n. By 3b12 the tail σ-algebra is independent of $\mathcal{F}_1^\infty$, therefore, of itself! □

[1] [D, Sect. 1.8, (8.1)]; [W, Th. 4.11].

3a13, 3a14 and 3b7 follow.

Here is another useful consequence of (3b11).

3b13 Exercise. Let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}$ be sub-σ-algebras, and $\mathcal{F}_\infty = \sigma(\mathcal{F}_1, \mathcal{F}_2, \dots)$. Then

$$L_2(\mathcal{F}_\infty) \text{ is the closure of } \bigcup_n L_2(\mathcal{F}_n)\,.$$

Prove it.

Hint: for an indicator function in $L_2(\mathcal{F}_\infty)$ use (3b11); their linear combinations approximate every bounded function.

In particular,

$$L_2\big(\sigma(X_1, X_2, \dots)\big) \text{ is the closure of } \bigcup_n L_2\big(\sigma(X_1, \dots, X_n)\big) \tag{3b14}$$

whenever $X_1, X_2, \dots$ are random variables (not just independent).

Some more applications of zero-one law (and CLT).

3b15 Exercise. For the simple random walk $(S_n)_n$,

(a) $\sup_n |S_n| = \infty$ a.s.;

(b) $\liminf_n S_n = -\infty$ and $\limsup_n S_n = \infty$ a.s.;

(c) $\sup\{n : S_n = 0\} = \infty$ a.s.

Prove it.

Hint: (a) $\max_k (S_{kn+n} - S_{kn}) = n$; (b) use (a) and 3b8; (c) use (b).

3b16 Exercise. For the simple random walk $(S_n)_n$,

$$\liminf_n \frac{S_n}{\sqrt{n}} = -\infty \quad \text{and} \quad \limsup_n \frac{S_n}{\sqrt{n}} = \infty \quad \text{a.s.}$$

Prove it.

Hint: $\sup_k \frac{S_{2^{k+1}} - S_{2^k}}{2^{k/2}} = \infty$ (using 2a1).