6 Infinite random sequences

6a Introductory remarks; almost certainty

There are two main reasons for entering continuous probability:

- infinitely high resolution;
- endless coin tossing.

Of course, both are theoretical idealizations.[64] Infinite resolution was discussed in Sect. 1c. Endless coin tossing was discussed in 1f4 and 2b5–2b7. Except for these digressions, Sections 1–5 are directed towards infinitely high resolution rather than endless coin tossing.[65] Now we turn to the latter (and its generalizations).

Almost certainty was introduced in Sect. 1c; recall the terminology:

$$\left.\begin{array}{l} A \text{ is negligible,} \\ A \text{ occurs almost never,} \\ \text{almost surely, } A \text{ does not occur} \end{array}\right\} \text{ when } P(A) = 0\,;$$

$$\left.\begin{array}{l} A \text{ is almost certain,} \\ A \text{ occurs almost always,} \\ \text{almost surely, } A \text{ occurs} \end{array}\right\} \text{ when } P(A) = 1\,.$$

Discrete probability gives us only trivial examples of almost certain events. Continuous probability gives better examples: let X have a continuous distribution, and x be a number,[66] then $X \ne x$ almost surely. Much deeper examples arise from (infinite) sequences of events or random variables, as we'll see soon.

Let a coin be tossed endlessly, giving independent identically distributed random variables $X_1, X_2, \dots$ each taking on two equiprobable values, say, $+1$ and -1 (or 0 and 1, if you like). What about $\lim_{n\to\infty} X_n$?[67]

Probably you believe that the limit does not exist. Why? Since there is a subsequence of $(+1)$, and another sequence of (-1). However, why they exist? What if X_n cease to change after some n? It seems unreasonable, but we need a proof. Consider an event[68] [69]

$$A = \{\, \exists n \; \forall m > n \; X_m = +1 \,\}\,;$$

we want to prove that $\mathbb{P}(A) = 0$. Introduce events

$$\begin{aligned} A_1 &= \{\, X_1 = +1,\ X_2 = +1,\ X_3 = +1,\ \dots \}\,, \\ A_2 &= \{\, X_2 = +1,\ X_3 = +1,\ \dots \}\,, \\ A_3 &= \{\, X_3 = +1,\ \dots \}\,, \end{aligned}$$

[64] We often prefer to idealize an unknown or irrelevant (high) resolution. Say, we prefer $\frac{d}{dx}\sin x = \cos x$ to $\frac{\sin(x+0.001)-\sin x}{0.001} = 0.999\,999\,833\,333\,341\ldots\cdot\cos x - 0.000\,499\,999\,958\ldots\cdot\sin x$. Similarly, we often prefer to move to infinity an unknown or irrelevant length of a (long) finite sequence.

[65] However, a number of general theorems are applicable to both cases.

[66] Non-random, of course.

[67] It is not the limit of frequency, just the limit of X_n itself.

[68] If you believe that its probability *tends* to 0, read Sect. 1c once again!

[69] An event is a subset of our probability space Ω; strictly speaking, we should write $A = \{\,\omega \in \Omega : \exists n \; \forall m > n \; X_m(\omega) = +1\,\}$, but probabilists usually omit ω.

and so on; $A_n = \{\, \forall m \ge n \; X_m = +1 \,\}$. We have $\mathbb{P}(A_1) = 0$ by the following argument:

$$\mathbb{P}(A_1) \le \mathbb{P}\big(X_1 = +1, \dots, X_n = +1\big) = \Big(\frac{1}{2}\Big)^n = \frac{1}{2^n}$$

for every $n = 1, 2, \dots$, therefore $\mathbb{P}(A_1) = 0$. The same argument[70] shows that $\mathbb{P}(A_2) = 0$, and similarly $\mathbb{P}(A_n) = 0$ for all n.

We have an increasing sequence of events, $A_1 \subset A_2 \subset \dots$ (think, why), and A is their union. We may say that $A = \lim_{n\to\infty} A_n$, according to the definition given after 2d9:[71]

(6a1)
$$\lim_{n\to\infty} A_n = \begin{cases} A_1 \cup A_2 \cup \dots & \text{if } A_1 \subset A_2 \subset \dots, \\ A_1 \cap A_2 \cap \dots & \text{if } A_1 \supset A_2 \supset \dots \end{cases}$$

Recall that probability depends continuously on an event, in the following sense:

(6a2)
$$\mathbb{P}\big(\lim_{n\to\infty} A_n\big) = \lim_{n\to\infty} \mathbb{P}(A_n)$$

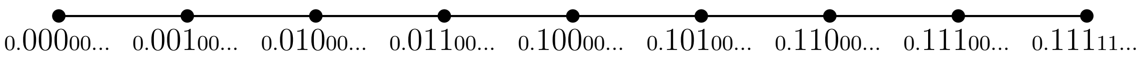
for every monotone sequence of events (see 2d9). So,

$$\mathbb{P}(A) = \mathbb{P}\big(\lim_{n\to\infty} A_n\big) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} 0 = 0\,.$$

Almost surely, there is no limit $\lim_{n\to\infty} X_n$ for the "coin tossing" sequence X_n.

We may treat X_n as binary digits[72] of a random point ω of $[0,1]$ (recall 2b5),

$$\omega = \big(0.X_1X_2\dots\big)_2\,.$$



That is, we may take $\Omega = [0,1]$ (with Lebesgue measure) as our probability space. Events A_n become subsets of $[0,1]$:

$$A_1 = \{1\}\,, \quad A_2 = \{\tfrac12, 1\}\,, \quad A_3 = \{\tfrac14, \tfrac12, \tfrac34, 1\}\,, \quad \dots$$

(think, why). Their limit is the set of all binary-rational points of $[0,1]$,

$$A = \lim_{n\to\infty} A_n = \big\{ \tfrac{k}{2^n} : n = 0, 1, 2, \dots,\; k = 0, 1, \dots, 2^n \big\}\,.$$

We have

$$\mathbb{P}(A) = 0\,; \quad \mathbb{P}\big([0,1] \setminus A\big) = 1\,.$$

Both A and $[0,1] \setminus A$ are dense in $[0,1]$, but A is negligible, while $[0,1] \setminus A$ is not.[73]

---

[70] Quite informally we could write $\mathbb{P}(A_1) = \big(1/2\big)^\infty = 0$, $\mathbb{P}(A_2) = \big(1/2\big)^{\infty-1} = 0$, and so on.

[71] It is a preliminary definition, applicable only for monotone sequences. A general definition will be given later (after (6b5)).

[72] Of course, now X_n takes on two values 0 and 1 (rather than ±1).

[73] Well, A is negligible since it is countable (and Lebesgue measure is nonatomic). Further we'll meet uncountable negligible sets, too.

Is there an empirical test for the statement that $\omega \notin A$ almost surely? No. Any physical random choice of ω, being of a finite resolution, does not allow to decide, whether $\omega \in A$ or not. Similarly, any physical coin tossing process, being of finite length, is not enough for determining $\lim_{n\to\infty} X_n$. In this sense, "convergence of random sequences" is a formal mathematical theory with no empirical basis.

Then, why do we learn the elegant but groundless[74] theory? For a simple reason: it helps us to understand long finite sequences.

6a3 Exercise. Let $X_1, X_2, \dots : \Omega \to \mathbb{R}$ be independent identically distributed random variables. Can it happen that $X_n \to +\infty$?

Hint: consider the median $\mathrm{Me} = X^*(1/2)$; we have $\mathbb{P}\big(X_n > \mathrm{Me}\big) \le 1/2$. It follows that the event $A = \{\exists n\ \forall m > n\ X_m > \mathrm{Me}\}$ is of probability 0.

6a4 Exercise. Let $X_1, X_2, \dots : \Omega \to \mathbb{R}$ be independent identically distributed random variables having exponential distribution

$$\mathbb{P}(X_n \le x) = 1 - e^{-x} \quad \text{for } x > 0\,.$$

What is the probability that $X_n > 2^{-n}$ for all n?

Hint.[75]

$$\begin{aligned}\mathbb{P}\big(X_1 > 2^{-1}, X_2 > 2^{-2}, \dots\big) &= \mathbb{P}\big(X_1 > 2^{-1}\big) \cdot \mathbb{P}\big(X_2 > 2^{-2}\big) \cdot \ldots = \\ &= \exp\big(-\tfrac12\big) \cdot \exp\big(-\tfrac14\big) \cdot \ldots = \exp\big(-\big(\tfrac12 + \tfrac14 + \tfrac18 + \dots\big)\big) = e^{-1} \approx 0.37\,.\end{aligned}$$

6a5 Exercise. For the same X_n as before, what is the probability that $X_n > \frac1n$ for all n?

Hint.

$$\exp\big(-\big(1 + \tfrac12 + \tfrac13 + \tfrac14 + \dots\big)\big) = e^{-\infty} = 0\,.$$

You see, the event $\exists n\ X_n \le 2^{-n}$ has a non-degenerate probability $1 - \frac1e \approx 0.63$; in contrast, the event $\exists n\ X_n \le \frac1n$ occurs almost surely. In order to distinguish between the two cases, we need distinguish between convergent and divergent series. Recall some relevant arguments:

$$\underbrace{\dots \ll \frac{1}{2^n} \ll \dots \ll \frac{1}{n^3} \ll \frac{1}{n^2}}_{\text{convergence}} \ll \underbrace{\frac1n \ll \frac{1}{\sqrt{n}} \ll \frac{1}{\sqrt[3]{n}} \ll \dots \ll \frac{1}{\log n} \ll \dots}_{\text{divergence}}$$

$$\sum \frac{1}{1.01^n} < \infty\,; \quad \sum \frac{1}{n^{1.01}} < \infty\,; \quad \sum \frac1n = \infty\,; \quad \sum \frac{1}{n \log^2 n} < \infty\,; \quad \sum \frac{1}{n \log n} = \infty\,;$$

$$\sum a_n < \infty \quad \Longleftrightarrow \quad \sum 2^n a_{2^n} < \infty \quad \text{for } a_n \downarrow 0\,;$$

$$\sum f(a_n) < \infty \quad \Longleftrightarrow \quad \sum a_n < \infty \quad \text{whenever } f(0) = 0,\ f'(0) > 0, \text{ and } a_n \to 0+\,.$$

---

[74] Though groundless empirically, it is still well-founded mathematically. It is based on measure theory. Thus, it cannot lead to a contradiction (provided, of course, that measure theory is consistent).

[75] $\exp(a)$ is the same as e^a.

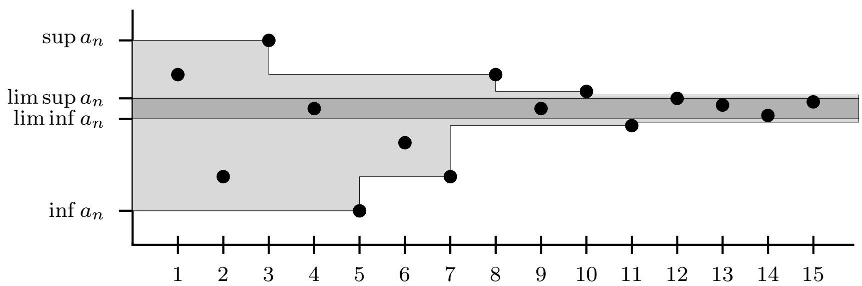
The case $f(a) = -\ln(1-a) = \ln\frac{1}{1-a}$ is especially important: for any $a_n \in [0,1)$

(6a6)
$$\prod(1-a_n) > 0 \iff \sum \ln(1-a_n) > -\infty \iff$$
$$\iff \sum \log\frac{1}{1-a_n} < \infty \iff \sum a_n < \infty\,.$$

6b Borel-Cantelli lemma

Sequences that do not converge are quite usual in probability theory. Having no limit, such a sequence has its *upper limit* (lim sup) and *lower limit* (lim inf). Given $a_1, a_2, \dots \in \mathbb{R}$, we define

(6b1)
$$\liminf_{n\to\infty} a_n = a_* = \sup_n \inf_{m\ge n} a_m = \lim_{n\to\infty} \inf(a_n, a_{n+1}, \dots)\,;$$
$$\limsup_{n\to\infty} a_n = a^* = \inf_n \sup_{m\ge n} a_m = \lim_{n\to\infty} \sup(a_n, a_{n+1}, \dots)\,.$$



In general,

(6b2)
$$-\infty \le \inf_n a_n \le \liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n \le \sup_n a_n \le +\infty\,.$$

If $\liminf a_n = \limsup a_n$ then $\lim a_n$ exists and is equal to both. Otherwise $\lim a_n$ does not exist.

Similarly, given events $A_1, A_2, \dots \subset \Omega$, we define

(6b3)
$$\liminf_{n\to\infty} A_n = A_* = \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m = \lim_{n\to\infty}\bigcap_{m=n}^{\infty} A_m\,;$$
$$\limsup_{n\to\infty} A_n = A^* = \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m = \lim_{n\to\infty}\bigcup_{m=n}^{\infty} A_m\,.$$

In other words,[76]

(6b4)
$$\omega \in \liminf_{n\to\infty} A_n \iff \exists n\ \forall m \ge n\ \ \omega \in A_m \iff$$
$$\iff \#\{m : \omega \notin A_m\} < \infty \iff \omega \in A_n \text{ eventually }\,;$$
$$\omega \in \limsup_{n\to\infty} A_n \iff \forall n\ \exists m \ge n\ \ \omega \in A_m \iff$$
$$\iff \#\{m : \omega \in A_m\} = \infty \iff \omega \in A_n \text{ infinitely often }\,.$$

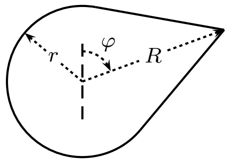
[76]Here $\#\{m : \dots\}$ stands for the number of such m.

In general,

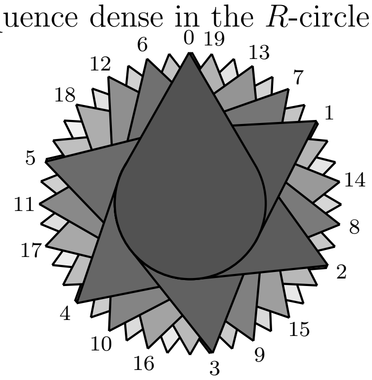
$$\emptyset \subset \bigcap_{n=1}^{\infty} A_n \subset \liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n \subset \bigcup_{n=1}^{\infty} A_n \subset \Omega\,. \tag{6b5}$$

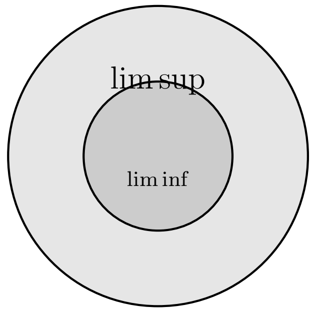
Now we are in position to generalize (6a1) for non-monotone sequences of events. By definition, if $\liminf A_n = \limsup A_n$ then $\lim A_n$ exists and is equal to both. Otherwise $\lim A_n$ does not exist. If $\lim A_n$ exists then $\mathbb{P}(\lim A_n) = \lim \mathbb{P}(A_n)$ by the sandwich argument (compare it with (6a2)).

A geometric example. Consider geometric figures of the following form:

$A(r, R, \varphi) \;\; =$ 

Let $\varphi_n = n - 1$ (which means $n - 1$ radians[77]), then vertices of $A_n = A(r, R, \varphi_n)$ are a non-periodic sequence dense in the R-circle:





A point of the small disk belongs to A_n for all n. A point of the annulus (between the two circles) belongs to A_n infinitely often, but not eventually (recall (6b4)). A point outside of the large disk belongs to no one of A_n. Thus,[78] $\cap A_n = \liminf A_n =$ (the r-disk), and $\limsup A_n = \cup A_n =$ (the R-disk). There is no $\lim A_n$. If you want all the six sets in (6b5) to differ, try $A_n = A(r_n, R_n, \varphi_n)$ with $r_n \uparrow r$, $R_n \downarrow R$, $r < R$.

6b6 Exercise. Consider "coin tossing" $X_1, X_2, \dots : \Omega \to \{0, 1\}$ and let $A_n = \{X_n = 1\} = \{X_n \ne 0\}$. Show that

$$\bigcap_{n=1}^{\infty} A_n = \Big\{ \sum_{n=1}^{\infty} (1 - X_n) = 0 \Big\}\,; \quad \bigcup_{n=1}^{\infty} A_n = \Big\{ \sum_{n=1}^{\infty} X_n > 0 \Big\}\,;$$

$$\liminf_{n\to\infty} A_n = \Big\{ \sum_{n=1}^{\infty} (1 - X_n) < \infty \Big\} = \big\{ X_n \to 1 \big\}\,;$$

$$\limsup_{n\to\infty} A_n = \Big\{ \sum_{n=1}^{\infty} X_n = \infty \Big\} = \big\{ X_n \not\to 0 \big\}\,.$$

Does $\lim_{n\to\infty} A_n$ exist? What about probability of the difference $\big(\limsup A_n\big) \setminus \big(\liminf A_n\big)$?

[77] Recall that the whole circle contains 2π (≈ 6.28) radians.

[78] There are some nuances concerning boundary points; I just ignore them.

Probably you know elementary relations for two indicators,[79]

$$\begin{aligned} A = B \cap C &\iff \mathbf{1}_A = \min\big(\mathbf{1}_B, \mathbf{1}_C\big)\,, \\ A = B \cup C &\iff \mathbf{1}_A = \max\big(\mathbf{1}_B, \mathbf{1}_C\big)\,. \end{aligned} \tag{6b7}$$

Now we have similar relations for infinite sequences of indicators:

$$\begin{aligned} A = \bigcap_{n=1}^{\infty} A_n &\iff \mathbf{1}_A = \inf_n \mathbf{1}_{A_n}\,; \\ A = \bigcup_{n=1}^{\infty} A_n &\iff \mathbf{1}_A = \sup_n \mathbf{1}_{A_n}\,; \\ A = \liminf_{n\to\infty} A_n &\iff \mathbf{1}_A = \liminf_{n\to\infty} \mathbf{1}_{A_n}\,; \\ A = \limsup_{n\to\infty} A_n &\iff \mathbf{1}_A = \limsup_{n\to\infty} \mathbf{1}_{A_n}\,. \end{aligned} \tag{6b8}$$

The following result, traditionally called "the first Borel-Cantelli lemma" (or "the first part of Borel-Cantelli lemma") is in fact an important theorem.

6b9 Theorem. For any[80] events $A_1, A_2, \dots$

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \quad\Longrightarrow\quad \mathbb{P}\Big(\limsup_{n\to\infty} A_n\Big) = 0\,.$$

Proof. First,

$$\mathbb{P}\bigg(\bigcup_{m=n}^{\infty} A_m\bigg) \le \sum_{m=n}^{\infty} \mathbb{P}(A_m)\,,$$

which is the limit (for $k \to \infty$) of

$$\mathbb{P}\big(A_m \cup A_{m+1} \cup \dots \cup A_{m+k}\big) \le \mathbb{P}(A_m) + \mathbb{P}(A_{m+1}) + \dots + \mathbb{P}(A_{m+k})\,.$$

Second,[81] [82]

$$\begin{aligned} \mathbb{P}\big(\limsup_{n\to\infty} A_n\big) &= \mathbb{P}\bigg(\lim_{n\to\infty} \bigcup_{m=n}^{\infty} A_m\bigg) = \\ &= \lim_{n\to\infty} \mathbb{P}\bigg(\bigcup_{m=n}^{\infty} A_m\bigg) \le \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}(A_m) = \\ &= \lim_{n\to\infty} \bigg(\sum_{m=1}^{\infty} \mathbb{P}(A_m) - \sum_{m=1}^{n-1} \mathbb{P}(A_m)\bigg) = \\ &= \sum_{m=1}^{\infty} \mathbb{P}(A_m) - \lim_{n\to\infty} \sum_{m=1}^{n-1} \mathbb{P}(A_m) = 0\,. \end{aligned}$$

---

[79] Indicators are functions, so, it is meant that $\mathbf{1}_{B\cap C}(\omega) = \min\big(\mathbf{1}_B(\omega), \mathbf{1}_C(\omega)\big)$ for each $\omega \in \Omega$.

[80] Not just independent!

[81] Do you see, where the first part of the proof is used below?

[82] You see, tails $\sum_{m=n}^{\infty} a_n$ tend to 0 (when $n \to \infty$) for every convergent series $\sum_{n=1}^{\infty} a_n$. Think, what happens for a divergent series.

$\square$

Another proof. Introduce indicators $X_n = \mathbf{1}_{A_n}$, then $A_n = \{X_n = 1\}$ and $\limsup A_n = \{\sum X_n = \infty\}$. We have[83]

$$\mathbb{E}\,(X_1 + \dots + X_n) = \mathbb{E}\,X_1 + \dots + \mathbb{E}\,X_n = \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n)\,.$$

Markov inequality[84] gives

$$\mathbb{P}\big(X_1 + \dots + X_n \ge M\big) \le \frac{\mathbb{P}(A_1) + \dots + \mathbb{P}(A_n)}{M}$$

for every $M \in (0,\infty)$. The limit for $n \to \infty$ gives

$$\mathbb{P}\Big(\sum X_n \ge M \Big) \le \frac{1}{M} \sum \mathbb{P}(A_n)\,.$$

Another limit, for $M \to \infty$, gives

$$\mathbb{P}\Big(\sum X_n = \infty \Big) = 0\,.$$

$\square$

What about the converse,

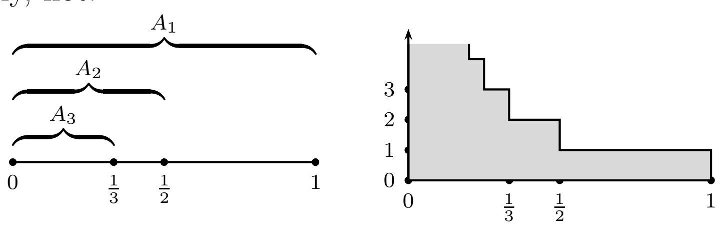
$$\sum_{n=1}^\infty \mathbb{P}(A_n) < \infty \quad \Longleftarrow \quad \mathbb{P}\Big(\limsup_{n\to\infty} A_n \Big) = 0 \quad ?$$

Check it for a simple case: $A_1 \supset A_2 \supset \dots$ Here, $\limsup A_n = \lim A_n$ and $\mathbb{P}(\limsup A_n) = \lim \mathbb{P}(A_n)$. Is it true that

$$\sum_{n=1}^\infty \mathbb{P}(A_n) < \infty \quad \Longleftarrow \quad \lim \mathbb{P}(A_n) = 0 \quad ?$$

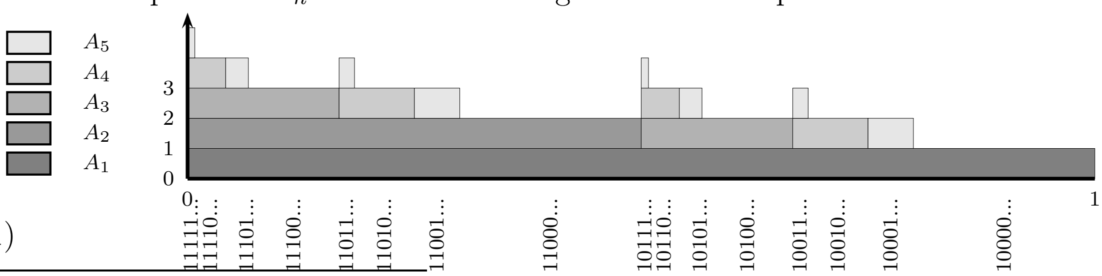
Evidently, not!

(6b10)



$\sum X_n < \infty$ almost sure, but $\mathbb{E} \sum X_n = +\infty\,.$

The case of independent A_n is more interesting and more complicated.



(6b11)

[83]In fact, $\mathbb{E} \sum X_n = \sum \mathbb{P}(A_n)$ by the monotone convergence theorem, but we do not need it.

[84]Recall it: $\mathbb{P}(X \ge M) \le \frac{1}{M}\mathbb{E}\,X$ for $X : \Omega \to [0,\infty)$, $M \in (0,\infty)$. Sketch of a proof: $M \cdot \mathbf{1}_{X\ge M} \le X$; thus $M \cdot \mathbb{P}(X \ge M) \le \mathbb{E}\,X$.

Continuing the process shown on (6b11) endlessly we get for the independent sum the same expectation as for the monotone sum (6b10); both are $1 + \frac{1}{2} + \frac{1}{3} + \dots = +\infty$. However, it is far from being evident, whether the function shown on (6b11) is finite almost everywhere, like (6b10), or not.

The following result, well-known as "the second Borel-Cantelli lemma" (or "the second part of Borel-Cantelli lemma") answers the question: the tower (6b11) is infinite almost everywhere!

6b12 Theorem. For any independent events $A_1, A_2, \dots$

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \quad \Longrightarrow \quad \mathbb{P}\Big(\limsup_{n\to\infty} A_n \Big) = 1 \, .$$

Proof. Introduce indicators $X_n = \mathbf{1}_{A_n}$, then $A_n = \{X_n = 1\}$ and $\limsup A_n = \{\sum X_n = \infty\}$. We have[85]

$$\begin{aligned} &\mathbb{E} \exp\big(-(X_1 + \dots + X_n)\big) = \mathbb{E}\left(e^{-X_1} \cdot \ldots \cdot e^{-X_n}\right) = \\ &= \left(\mathbb{E}\, e^{-X_1}\right) \cdot \ldots \cdot \left(\mathbb{E}\, e^{-X_n}\right) = \prod_{k=1}^{n} \Big(1 - \big(1 - \tfrac{1}{e}\big)\mathbb{P}(A_n) \Big) \, , \end{aligned}$$

since $\mathbb{E}\, e^{-X_k} = e^0 \cdot \mathbb{P}(X_k = 0) + e^{-1} \cdot \mathbb{P}(X_k = 1) = 1 \cdot (1 - \mathbb{P}(A_k)) + \frac{1}{e} \cdot \mathbb{P}(A_k)$. Thus,

$$\mathbb{E} \exp\big(-(X_1 + \dots + X_n)\big) \to 0 \quad \text{for } n \to \infty \, ,$$

since $\sum_{n=1}^{\infty} \big(1 - \frac{1}{e}\big)\mathbb{P}(A_n) = \infty$ (recall 6a6). Markov inequality gives for every $M \in (0, \infty)$

$$\mathbb{P}\Big(\exp\big(-(X_1 + \dots + X_n)\big) \ge e^{-M} \Big) \le \frac{\mathbb{E} \exp\big(-(X_1 + \dots + X_n)\big)}{e^{-M}} \, .$$

It follows that[86]

$$\mathbb{P}\bigg(\sum_{k=1}^{\infty} X_k \le M \bigg) \le e^M \mathbb{E} \exp\big(-(X_1 + \dots + X_n)\big) \, .$$

The limit for $n \to \infty$ gives

$$\mathbb{P}\bigg(\sum_{k=1}^{\infty} X_k \le M \bigg) = 0 \, .$$

Another limit, for $M \to \infty$, gives

$$\mathbb{P}\Big(\sum X_n < \infty \Big) = 0 \, .$$

□

[85] Did you note, where the independence is used?

[86] You see, $\sum_1^{\infty} X_k \ge \sum_1^n X_k$.

So, for *independent* events the problem is solved:

$$\text{(6b13)} \qquad \begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \quad &\Longrightarrow \quad \mathbb{P}\bigl(\limsup_{n\to\infty} A_n\bigr) = 0\,, \\ \sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \quad &\Longrightarrow \quad \mathbb{P}\bigl(\limsup_{n\to\infty} A_n\bigr) = 1\,. \end{aligned}$$

Note that intermediate values (between 0 and 1) are excluded.

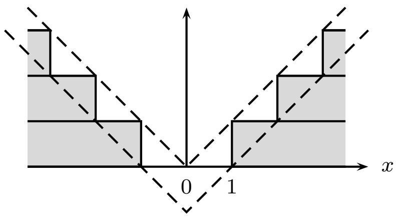
A corollary: for any sequence $X_1, X_2, \dots$ of i.i.d.[87] random variables,

$$\text{(6b14)} \qquad \begin{aligned} \mathbb{E}\,|X_1| < \infty \quad &\Longrightarrow \quad \frac{X_n}{n} \xrightarrow[n\to\infty]{} 0 \quad \text{almost surely}\,; \\ \mathbb{E}\,|X_1| = \infty \quad &\Longrightarrow \quad \limsup_{n\to\infty} \frac{|X_n|}{n} = \infty \quad \text{almost surely}\,. \end{aligned}$$

An explanation. First, for any random variable X,

$$\mathbb{E}\,|X| < \infty \quad \Longleftrightarrow \quad \sum_{n=1}^{\infty} \mathbb{P}\bigl(|X| > n\bigr) < \infty\,.$$

Moreover, $\mathbb{E}\,|X| - 1 \le \sum \mathbb{P}(|X| > n) \le \mathbb{E}\,|X|$ according to a "sandwich" argument:



Now, Borel-Cantelli lemma gives[88]

$$\bigl(\mathbb{E}\,|X_1| = \infty\bigr) \quad \Longleftrightarrow \quad \bigl(\,|X_n| > n \text{ infinitely often}\,\bigr)\,.$$

6b15 Exercise. Complete the explanation, prove (6b14).

Hint. $\mathbb{E}\,|X_1| < \infty \quad \Longleftrightarrow \quad \mathbb{E}\,|cX_1| < \infty$ for any $c \in (0, \infty)$.

The *normal* distribution is especially important. Let $X_1, X_2, \dots$ be i.i.d. $\mathrm{N}(0,1)$ random variables. Then $\mathbb{E}\,|X_1| < \infty$, therefore $X_n/n \to 0$ almost surely.[89] Moreover, the density $f_X(x) = \text{const} \cdot \exp(-x^2/2)$ tends to 0 (for $x \to \infty$) *exponentially* fast, which ensures that $\int x^k f_X(x)\,dx < \infty$ for each k. Thus, for instance, $\mathbb{E}\,|X_1|^{10} < \infty$. Applying (6b14) to the sequence $X_1^{10}, X_2^{10}, \dots$ we get $X_n^{10}/n \to 0$, that is, $X_n/\sqrt[10]{n} \to 0$ almost sure. It is much more than $X_n/n \to 0$. Still more, consider $\mathbb{E}\,\exp(cX_1^2)$; it is finite for $c < 1/2$ but infinite for $c \ge 1/2$ (check it). Therefore, $\exp(X_n^2/2) > n$ infinitely often, that is, $|X_n| > \sqrt{2 \ln n}$

[87] i.i.d. = independent, identically distributed.

[88] You see, $\mathbb{P}\bigl(|X_n| > n\bigr) = \mathbb{P}\bigl(|X_1| > n\bigr)$.

[89] Do you think that, say, $X_n/\ln\ln n$ also tends to 0, just because $\ln\ln n \to \infty$? Wait a little…

infinitely often, and $\limsup_{n\to\infty}(|X_n|/\sqrt{2\ln n}) \ge 1$. On the other hand, taking c a bit less than $1/2$ we get, say, $|X_n| \le \sqrt{2.02\ln n}$ eventually, thus, $\limsup_{n\to\infty}(|X_n|/\sqrt{2\ln n}) \le 1.01$. It means that

$$\limsup_{n\to\infty} \frac{|X_n|}{\sqrt{2\ln n}} = 1 \quad \text{almost sure} \tag{6b16}$$

for independent random variables $X_1, X_2 \dots$ having the normal distribution with the mean 0 and the variance 1. In fact, $\limsup X_n/\sqrt{2\ln n} = 1$ and $\liminf X_n/\sqrt{2\ln n} = -1$ a.s.

6c Modes of convergence

After all, does $X_n/\sqrt{2\ln n}$ converge to 0, or not? It depends...

6c1 Exercise. For every random variable $X : \Omega \to \mathbb{R}$,

$$\mathbb{E}\,|X| = 0 \quad \Longleftrightarrow \quad \mathbb{P}\big(\, X = 0 \,\big) = 1\,.$$

Prove it. Hint: $\mathbb{P}\big(\, X \ne 0 \,\big) = \lim_{\varepsilon\to 0} \mathbb{P}\big(\, |X| \ge \varepsilon \,\big)$; also, $\mathbb{P}\big(\, |X| \ge \varepsilon \,\big) \le \mathbb{E}\,|X|/\varepsilon$.

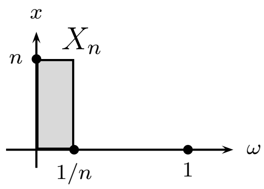
6c2 Exercise. Let $X_1, X_2, \dots : \Omega \to \mathbb{R}$ be random variables. Then

(a) if Ω is finite then

$$\mathbb{P}\big(\, X_n \to 0 \,\big) = 1 \quad \Longrightarrow \quad \mathbb{E}\,|X_n| \to 0\,;$$

(b) in general, it does not hold.

Prove it. Hint: (a) $\max_\omega |X_n(\omega)| \to 0$; (b)



What happens if Ω is countable? What if Ω has both a discrete part and a continuous part?

6c3 Exercise. Let $X_1, X_2, \dots : \Omega \to \mathbb{R}$ be random variables. Then

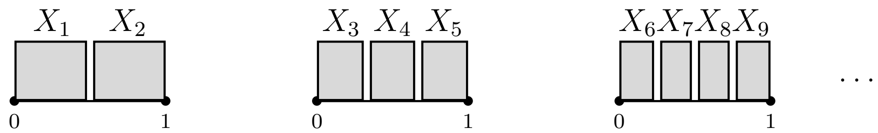
(a) if Ω is finite or countable then

$$\mathbb{E}\,|X_n| \to 0 \quad \Longrightarrow \quad \mathbb{P}\big(\, X_n \to 0 \,\big) = 1\,;$$

(b) in general, it does not hold.

Prove it. Hint: (a) $|X(\omega)| \le \mathbb{E}\,|X|/\mathbb{P}\big(\,\{\omega\}\,\big)$;

(b)



What happens if Ω has both a discrete part and a continuous part?

For a sequence of numbers $x_1, x_2, \dots \in \mathbb{R}$ the condition "$x_n \to 0$" is unambiguous. In contrast, for a sequence of random variables we have several *nonequivalent* interpretations of "$X_n \to 0$", that is, several modes of convergence.

6c4 Definition. Let $X, X_1, X_2, \ldots : \Omega \to \mathbb{R}$ be random variables.

(a) $X_n \to X$ almost surely, if[90]

$$\mathbb{P}\big(\{\omega \in \Omega : X_n(\omega) - X(\omega) \xrightarrow[n\to\infty]{} 0\}\big) = 1\,;$$

(b) $X_n \to X$ in square mean, if $\mathbb{E}\,|X|^2 < \infty$ and

$$\mathbb{E}\,|X_n - X|^2 \xrightarrow[n\to\infty]{} 0\,;$$

(c) $X_n \to X$ in absolute mean, if $\mathbb{E}\,|X| < \infty$ and

$$\mathbb{E}\,|X_n - X| \xrightarrow[n\to\infty]{} 0\,;$$

(d) $X_n \to X$ in probability,[91] if for every $\varepsilon > 0$

$$\mathbb{P}\big(|X_n - X| > \varepsilon\big) \xrightarrow[n\to\infty]{} 0\,.$$

6c5 Exercise. Let X_n be indicators, $X_n = \mathbf{1}_{A_n}$, and $X = 0$. Show that each one of (b), (c), (d) is equivalent to $\mathbb{P}(A_n) \to 0$, while (a) is not. What happens for independent A_n?

6c6 Exercise. Let $c_n \to \infty$, $X_n = c_n \mathbf{1}_{A_n}$, and $X = 0$. Show that

$$\begin{aligned}
&\text{(b)} \quad \iff \quad c_n^2 \mathbb{P}(A_n) \to 0,\\
&\text{(c)} \quad \iff \quad c_n \mathbb{P}(A_n) \to 0,\\
&\text{(d)} \quad \iff \quad \mathbb{P}(A_n) \to 0 \qquad \text{(irrespective of } c_n),
\end{aligned}$$

and

$$\text{(a)} \quad \iff \quad \mathbb{P}(\limsup A_n) = 0 \qquad \text{(irrespective of } c_n).$$

Show by examples that there are no two equivalent conditions among (a), (b), (c), (d).

6c7 Exercise. (b) $\Longrightarrow$ (c) $\Longrightarrow$ (d) for any $X, X_1, X_2, \ldots$

Prove it. Hint: $\mathbb{E}\,|X_n - X|^2 - (\mathbb{E}\,|X_n - X|)^2 = \operatorname{Var}(|X_n - X|) \ge 0$; also, $\mathbb{P}\,\big(|X_n - X| \ge \varepsilon\big) \le \mathbb{E}\,|X_n - X|/\varepsilon$.

6c8 Lemma. (a) $\Longrightarrow$ (d) for any $X, X_1, X_2, \ldots$

Proof. Almost surely $X_n - X \to 0$, therefore, $|X_n - X| \le \varepsilon$ eventually. Introduce events

$$A_n = \{|X_n - X| \le \varepsilon,\ |X_{n+1} - X| \le \varepsilon,\ \ldots\}\,,$$

then $A_1 \subset A_2 \subset \ldots$ and $\mathbb{P}(\lim A_n) = 1$. It follows that $\lim \mathbb{P}(A_n) = 1$. However, A_n is incompatible with $|X_n - X| > \varepsilon$; thus, $\mathbb{P}(|X_n - X| > \varepsilon) \le 1 - \mathbb{P}(A_n) \to 0$. □

[90] It can be shown that the set $\{\omega \in \Omega : X_n(\omega) - X(\omega) \xrightarrow[n\to\infty]{} 0\}$ is measurable.

[91] Analysts say "in measure".

So,

$$\begin{array}{ccccc} & & & & \text{(a)} \\ & & & & \Downarrow \\ \text{(b)} & \Longrightarrow & \text{(c)} & \Longrightarrow & \text{(d)} \end{array}$$

All the 4 modes (a)–(d) are modes of convergence of random variables, not distributions. Say, for the "coin tossing" sequence $X_1, X_2, \dots$ distribution functions F_n of X_n are all the same, $F_1 = F_2 = \dots = F$, thus, $F_n \xrightarrow[n\to\infty]{} F$ trivially. However, X_n does not converge.[92] Do not confuse convergence of distributions and convergence of random variables!

6c9 Definition. (a) A sequence $(F_1, F_2, \dots)$ of distribution functions *converges weakly* to a distribution function F, if $F_n(x) \xrightarrow[n\to\infty]{} F(x)$ for every x such that F is continuous at x.

(b) A sequence $X_1, X_2, \dots$ of random variables *converges in distribution* to a random variable X, if $F_{X_n} \xrightarrow[n\to\infty]{} F_X$ weakly.[93]

Item (b), "convergence in distribution", is rather illogical; you see, convergence of distributions should not be ascribed to random variables. Say, for the "coin tossing" sequence $X_1, X_2, \dots$ we may say that $X_n \to X_1$ in distribution, as well as $X_n \to X_2$ in distribution! Still, the terminology is widely used. Moreover, people say "$X_n \to F$ in distribution", having in mind "$F_{X_n} \to F$ weakly". For instance, a claim of the form "$X_n \to \mathrm{N}(0, 1)$ in distribution" appears often in limit theorems.

The simplest case of Definition 6c9 is the degenerate case: each F_n is concentrated at a point a_n, that is, $F_n(x) = \begin{cases} 1 & \text{for } x \ge a_n, \\ 0 & \text{for } x < a_n. \end{cases}$ Accordingly, $X_n = a_n$ almost sure. Here, X_n converges in distribution, if and only if a_n converges to a number $a \in (-\infty, +\infty)$, and then $F_n \to F$ weakly, where $F(x) = \begin{cases} 1 & \text{for } x \ge a, \\ 0 & \text{for } x < a. \end{cases}$ Accordingly, $X_n \to X$ in distribution, where $X = a$ almost sure. Note that $\mathbb{P}(X_n = a)$ need not converge to $\mathbb{P}(X = a)$; this is why the convergence $F_n \to F$ is called weak.

Clearly, $X_n \to X$ in distribution if and only if $X_n^* \to X^*$ in distribution. The latter appears to be equivalent to

$$X_n^*(p) \xrightarrow[n\to\infty]{} X^*(p)$$

for every $p \in (0, 1)$ such that X^* is continuous at p. (Do not think that $X_n^*(p\pm) \to X^*(p\pm)$.)

6c10 Lemma. (a) If $X_n \to X$ in probability then $X_n \to X$ in distribution.

(b) If $X_n \to 0$ in distribution then $X_n \to 0$ in probability.

(I give no proof.)

6c11 Exercise. (a) If $X_n \to X$ in absolute mean then $\mathbb{E}\, X_n \to \mathbb{E}\, X$.

(b) The converse is generally wrong.

Prove it. Hint: (a) $-|X_n - X| \le X_n - X \le |X_n - X|$; take the expectation. (b) It may happen that $\mathbb{E}\,(X_1 - X) = 0$ but $\mathbb{E}\,|X_n - X| \ne 0$; try $X_1 = X_2 = \dots$

[92] To any limit, in any mode.

[93] Of course, $F_X(x) = \mathbb{P}(X \le x)$, and $F_{X_n}(x) = \mathbb{P}(X_n \le x)$.

Here are two famous theorems of measure theory (formulated here for probability measures, in probabilistic language). (For $\Omega = (0,1)$ we may think in terms of mes_2...)

6c12 Theorem (Monotone Convergence Theorem)**.** Let $X, X_1, X_2, \ldots : \Omega \to \mathbb{R}$ be random variables such that

$$\mathbb{P}\big(X_n \uparrow X \big) = 1 \quad \text{and} \quad \mathbb{E}\,|X_1| < \infty\,.$$

Then[94]

$$\lim_{n\to\infty} \mathbb{E}\, X_n = \mathbb{E}\, X \in (-\infty, +\infty]\,.$$

6c13 Theorem (Dominated Convergence Theorem)**.** Let $Y, X, X_1, X_2, \ldots : \Omega \to \mathbb{R}$ be random variables such that

$$\mathbb{P}\big(X_n \to X \big) = 1\,, \quad \mathbb{P}\big(|X_n| \le Y \big) = 1\,, \quad \text{and} \quad \mathbb{E}\,|Y| < \infty\,.$$

Then[95]

$$\lim_{n\to\infty} \mathbb{E}\, X_n = \mathbb{E}\, X\,.$$

For $\Omega = (0,1)$ these facts may be deduced from 5c13. Think also about the bounded case, especially, indicators...

6c14 Exercise. If Y is non-integrable then it may happen that $\mathbb{P}\big(X = 0 \big) = 1$ but $\mathbb{E}\, X_n \to \infty$.

Give an example. Hint: recall 6c6.

6c15 Exercise. If $\mathbb{E}\,|X| < \infty$ then there exist bounded random variables $X_1, X_2, \ldots$ such that $\mathbb{E}\,|X_n - X| \to 0$. Prove it.

Recall also the moments as the derivatives of the MGF...

6d Laws of large numbers

6d1 Theorem (Weak law of large numbers)**.** Let $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ be independent, identically distributed random variables, $\mathbb{E}\,|X_1| < \infty$, $\mu = \mathbb{E}\, X_1$. Then

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n\to\infty]{} \mu \quad \text{in absolute mean.}$$

Note that convergence in absolute mean implies convergence in probability (and in distribution). The square-integrable case, $\mathbb{E}\,|X_1|^2 < \infty$, is easy:

$$\text{(6d2)} \qquad \mathrm{Var}\left(\frac{X_1 + \dots + X_n}{n} \right) = \frac{\mathrm{Var}(X_1)}{n} \to 0\,,$$

[94] Existence of the (finite or infinite) limit is evident due to monotonicity.

[95] Note that $|X| \le |Y|$, thus X must be integrable.

which gives convergence in square mean, the more so, in absolute mean. The general case follows via approximation:[96]

$$X_k = Y_k + Z_k\,, \quad \mathbb{E}\,|Y_k|^2 < \infty\,, \quad \mathbb{E}\,|Z_k| \le \varepsilon\,;$$

(6d3) $$\mathbb{E}\left|\frac{X_1+\dots+X_n}{n} - \mu\right| \le \underbrace{\mathbb{E}\left|\frac{Y_1+\dots+Y_n}{n} - \mu\right|}_{\to|\mathbb{E}\,Y-\mu|\le\varepsilon} + \underbrace{\mathbb{E}\left|\frac{Z_1+\dots+Z_n}{n}\right|}_{\le\varepsilon}\,.$$

6d4 Theorem (Strong law of large numbers)**.** Let $X_1, X_2, \dots : \Omega \to \mathbb{R}$ be independent, identically distributed random variables, $\mathbb{E}\,|X_1| < \infty$, $\mu = \mathbb{E}\,X_1$. Then

$$\frac{X_1+\dots+X_n}{n} \xrightarrow[n\to\infty]{} \mu \quad \text{almost surely, and in absolute mean.}$$

Several proofs are well-known; no one is simple enough for being reproduced here. The most natural proof (for my opinion) is given by theory of martingales. It combines a clever observation that

$$\mathbb{E}\left(X_1 \,\middle|\, S_n, S_{n+1}, \dots \right) = \frac{1}{n} S_n$$

(here $S_n = X_1 + \dots + X_n$) and a deep theorem:

For every integrable random variable X and every random variables $Y_1, Y_2, \dots$ (quite arbitrary, not at all independent), random variables $\mathbb{E}\left(X \,\middle|\, Y_n, Y_{n+1}, \dots \right)$ converge almost surely, and in absolute mean.

It remains to note that the limit is just μ by the weak law.

(Normal numbers and singular measures may be mentioned here...)

Here is a sketch of a proof assuming that the moment generating function (4f4) of X_1 is finite in a neighborhood of 0. (The case of bounded X_1 is thus covered.) By 4f7,

$$\frac{d}{dt}\bigg|_{t=0} \mathrm{MGF}_{X_1}(t) = \mu\,,$$

therefore for every $\varepsilon > 0$

$$\mathrm{MGF}_{X_1}(t) < e^{(\mu+\varepsilon)t}$$

provided that $t > 0$ is small enough. We apply Markov inequality[97] to $e^{t(X_1+\dots+X_n)}$:

$$\mathbb{P}\left(\frac{X_1+\dots+X_n}{n} \ge \mu + \varepsilon \right) = \mathbb{P}\left(e^{t(X_1+\dots+X_n)} \ge e^{tn(\mu+\varepsilon)} \right) \le$$
$$\le \frac{\mathbb{E}\,e^{t(X_1+\dots+X_n)}}{e^{tn(\mu+\varepsilon)}} = \left(\frac{\mathbb{E}\,e^{tX_1}}{e^{t(\mu+\varepsilon)}} \right)^n = q^n\,,$$

where $q = \frac{\mathrm{MGF}_{X_1}(t)}{e^{t(\mu+\varepsilon)}} < 1$. By the first Borel-Cantelli lemma 6b9, $\frac{X_1+\dots+X_n}{n} < \mu + \varepsilon$ for large n, almost surely. It holds for all $\varepsilon > 0$, thus

$$\limsup_{n\to\infty} \frac{X_1+\dots+X_n}{n} \le \mu \quad \text{a.s.}$$

Similarly, $\mathrm{MGF}_{X_1}(t) < e^{(\mu-\varepsilon)t}$ for some $t < 0$, which leads to $\liminf \dots \ge \mu$.

[96] Consider $Y_k = \begin{cases} X_k & \text{if } |X_k| \le M, \\ 0 & \text{otherwise,} \end{cases}$ where M is large enough.

[97] It was also used in the second proof of 6b9.

6e Central limit theorem

Here is probably the most famous theorem of the whole probability theory.

6e1 Theorem (Central Limit Theorem)**.** Let $X_1, X_2, \dots : \Omega \to \mathbb{R}$ be independent, identically distributed random variables, $\mathbb{E}\, X_1 = \mu$, $\operatorname{Var} X_1 = \sigma^2 \in (0, \infty)$. Then

$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{} \mathrm{N}(0,1) \quad \text{in distribution.}$$

Here $\mathrm{N}(0,1)$ is the standard normal distribution. In other words,

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \le z \Big) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\, du \tag{6e2}$$

for all $z \in \mathbb{R}$. One says that $X_1 + \dots + X_n$ is *asymptotically normal* (for $n \to \infty$).

We cannot hope for convergence in probability, since S_m and S_n are nearly independent for $n \gg m \gg 1$.

Several proofs (of the central limit theorem) are well-known; no one is simple. I give a sketch of one proof in the form of several lemmas.

Recall the Poisson distribution $\mathrm{P}(\lambda)$,

$$\begin{gathered} X \sim \mathrm{P}(\lambda) \quad \Longleftrightarrow \quad \mathbb{P}\big(X = k \big) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ for } k = 0, 1, 2, \dots ; \\ \mathbb{E}\, X = \lambda\,, \qquad \sigma_X = \sqrt{\lambda}\,; \quad \lambda \in [0, \infty)\,. \end{gathered} \tag{6e3}$$

6e4 Lemma. $\mathrm{P}(\lambda)$ is asymptotically normal for $\lambda \to \infty$. That is, $\lim_{\lambda\to\infty} \mathbb{P}\big(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \le z \big) = \Phi(z)$, where $X_\lambda \sim \mathrm{P}(\lambda)$.

Hint. Use the Stirling formula

$$k! \sim \sqrt{2\pi k} k^k e^{-k} \quad \text{for } k \to \infty\,.$$

6e5 Lemma. Let $N_\lambda \sim \mathrm{P}(\lambda)$ be a random variable independent of $X_1, X_2, \dots$ Introduce

$$S_n = X_1 + \dots + X_n\,, \quad S_{N_\lambda} = X_1 + \dots + X_{N_\lambda}\,.$$

Then

$$\mathbb{E} \left| \frac{S_{N_n} - N_n\mu}{\sigma\sqrt{n}} - \frac{S_n - n\mu}{\sigma\sqrt{n}} \right|^2 \xrightarrow[n\to\infty]{} 0\,.$$

Hint. On one hand, $N_n/n \to 1$ in absolute mean (by the weak law of large numbers). On the other hand,

$$\mathbb{E} \left| \frac{S_m - m\mu}{\sigma\sqrt{n}} - \frac{S_n - n\mu}{\sigma\sqrt{n}} \right|^2 \le \left| \frac{m}{n} - 1 \right|;$$

and $\mathbb{E}\,(\dots) = \mathbb{E}\,\big(\mathbb{E}\,\big(\dots \,\big|\, N_n\,\big)\big)$.

6e6 Corollary. The following two conditions are equivalent.

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to \mathrm{N}(0,1) \quad \text{in distribution;}$$
$$\frac{S_{N_n} - N_n\mu}{\sigma\sqrt{n}} \to \mathrm{N}(0,1) \quad \text{in distribution.}$$

Hint. If two random variables are close in square mean (or in absolute mean, or even in probability) then their distributions are close.

6e7 Lemma. The central limit theorem holds when X_1 takes on a finite number of values only.

Hint. Let X_1 take on just two values x_1 and x_2. We have $S_{N_n} = x_1 N'_n + x_2 N''_n$ where N'_n is the number of $k \in \{1,\dots,N_n\}$ such that $X_k = x_1$; similarly, N''_n and x_2. Due to remarkable properties of Poisson distribution, random variables N'_n and N''_n are independent and have Poisson distributions, $N'_n \sim \mathrm{P}\big(n\mathbb{P}\,(\,X_1 = x_1\,)\big)$ and $N''_n \sim \mathrm{P}\big(n\mathbb{P}\,(\,X_1 = x_2\,)\big)$. So, S_{N_n} is the sum of two independent random variables, each being approximately normal by 6e4.

Finally, the general case (of the central limit theorem) follows from 6e7 via approximation:

(6e8)
$$X_k = Y_k + Z_k\,, \quad Y_k \text{ takes on a finite number of values,} \quad \mathbb{E}\,|Z_k|^2 \le \varepsilon\sigma^2\,, \quad \mathbb{E}\,Z_k = 0\,;$$
$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} = \underbrace{\frac{Y_1 + \dots + Y_n - n\mu}{\sigma\sqrt{n}}}_{\to \mathrm{N}(0,1)} + \underbrace{\frac{Z_1 + \dots + Z_n}{\sigma\sqrt{n}}}_{\mathbb{E}\,|\dots|^2 \le \varepsilon}\,.$$