Lecture 11. The zero-one law.

Theorem 11.1 (the second Borel – Cantelli's Lemma). *Let* A_1, ..., A_n, ... *be an infinite sequence of* independent *events* (note that independence was not required in the first Borel – Cantelli Lemma). *If the series* $\sum_{i=1}^{\infty} P(A_i)$ *diverges* (its sum is equal to ∞), *then almost surely infinitely many events* A_i *occur*:

$$P\{\text{infinitely many of } A_i \text{ occur}\} = 1. \tag{11.1}$$

(In the first Lemma it was $\sum_{i=1}^{\infty} < \infty$, and $P\{\text{infinitely many of } A_i \text{ occur}\} = 0$.)

Proof. Let me show the proof assuming that the statement of Problem $\boxed{\mathbf{22}}$ is *proved* and not disproved. If it turns out that this statement is false (or if we cannot prove it), we'll invent another way to prove our theorem.

The event under the probability sign in (11.1) can be written as

$$\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i \tag{11.2}$$

(see formula (1 – 2.23)). There is something in Problem $\boxed{\mathbf{22}}$ about the intersection of an infinite sequence of events: can we use it here?

No: it is about an infinite sequence of *independent* events, and the events $\bigcup_{i=n+1}^{\infty} A_i$ that are intersected here are dependent. Indeed, they clearly form a non-increasing sequence: $\bigcup_{i=n}^{\infty} A_i \supseteq \bigcup_{i=n+1}^{\infty} A_i$ for all n; and the events $A \supseteq B$ *cannot* be independent: $P(A \cap B) = P(B) \neq P(A) \cdot P(B)$ except in the trivial cases of $P(A) = 1$, $P(B) = 0$.

But the infinite intersection (11.2) is the *limit* of a non-increasing sequence of events, so we can write:

$$P\Big(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i\Big) = \lim_{n\to\infty} P\Big(\bigcup_{i=n}^{\infty} A_i\Big). \tag{11.3}$$

How to find $P\big(\bigcup_{i=n}^{\infty} A_i\big)$?

Let us write this probability as 1 minus the probability of the opposite event:

$$P\Big(\bigcup_{i=n}^{\infty} A_i\Big) = 1 - P\Big(\big(\bigcup_{i=n}^{\infty} A_i\big)^c\Big) = 1 - P\Big(\bigcap_{i=n}^{\infty} A_i^c\Big). \tag{11.4}$$

If the statement of Problem $\boxed{\mathbf{22}}$ is true, we have

$$P\Big(\bigcap_{i=n}^{\infty} A_i^c\Big) = \prod_{i=n}^{\infty} P(A_i^c) \tag{11.5}$$

What can we say about this infinite product (which is defined, of course, as the limit $\lim_{m\to\infty} \prod_{i=n}^{m} P(A_i^c)$) based on the fact that the series $\sum_{i=1}^{\infty} P(A_i)$ diverges?

If infinitely many of $P(A_i)$ are equal to 1, there is at least one factor in the product (11.5) that is equal to 0, and the product is 0.

If all $P(A_i)$, $i > n$, are less than one, $P(A_i^c) > 0$, we can take the logarithm:

$$\ln\Bigl(\prod_{i=n}^{m} P(A_i^c)\Bigr) = \sum_{i=n}^{m} \ln P(A_i^c). \tag{11.6}$$

For all $x \in (0,\,1]$ we have: $\ln x \le x - 1$ (make a picture); so

$$\ln\Bigl(\prod_{i=n}^{m} P(A_i^c)\Bigr) \le \sum_{i=n}^{m} [P(A_i^c) - 1] = -\sum_{i=n}^{m} P(A_i), \tag{11.7}$$

$$\prod_{i=n}^{\infty} P(A_i^c) \le \lim_{m\to\infty} \exp\Bigl\{-\sum_{i=n}^{m} P(A_i)\Bigr\}. \tag{11.8}$$

The series $\sum_{i=1}^{\infty} P(A_i)$ diverges to $+\infty$, the series $\sum_{i=n}^{\infty} P(A_i)$ differs from this “longer” series only by its $n-1$ first terms, so this “shorter” series also diverges; and the limit as $m \to \infty$ of the exponent in (11.8) is equal to $-\infty$. So

$$\prod_{i=n}^{\infty} P(A_i^c) = 0, \tag{11.9}$$

and the probability (11.5) is also zero. So the probability (11.3) is the limit of ones, that is 1.

So for an infinite sequence of independent events either finitely many of them occur almost surely, or almost surely infinitely many occur (according to whether the series $\sum_{i=1}^{\infty} P(A_i)$ converges or diverges). It turns out that this is a particular case of a more general statement, and the reason by which this general statement (which will be formulated later) can be applied here is that whether the event {infinitely many of events A_i occur} occurs is determined by the *tails* of the sequence: by its subsequence A_n, A_{n+1}, A_{n+2}, ..., for every natural n (because A_1, A_2, ..., A_{n-1} are only finitely many).

It is more convenient to speak about a sequence of *random variables* $\xi_1, \xi_2, \ldots, \xi_n, \ldots$; and we are going to do so (with the events A_i we can associate their indicator random variables I_{A_i}).

Now we go to the set-theoretic introduction to probability theory.

Let $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ be an infinite sequence of random variables (independent or not, does not matter now: we are in the set-theoretic introduction to probability theory, where the probability P and all things based on it, in particular, independence, are not admitted).

Let us introduce σ-algebras of events in the sample space Ω (i.e., σ-algebras being parts of the fundamental σ-algebra $\mathcal{F}$):

$$\mathcal{F}_{\le n} = \sigma(\xi_i,\ 1 \le i \le n), \tag{11.10}$$

generated by the random variables ξ_1, ..., ξ_n;

$$\mathcal{F}_{\geq n} = \sigma(\xi_i,\ i \geq n), \tag{11.11}$$

generated by the random variables ξ_n, ξ_{n+1}, ξ_{n+2}, ...

Now we introduce the *tail σ-algebra* $\mathcal{F}_{\geq\infty}$: the limit as $n \to \infty$ of the σ-algebras $\mathcal{F}_{\geq n}$ defined by (11.11). (The word "tail" is used when we are dealing with infinite sequences, and it means the parts of the sequences after their n-th element, for n being large. E. g., we can say that convergence of an infinite series depends only on its tails.)

What does it mean, the limit of a sequence of σ-algebras?

Remember that a σ-algebras are classes of sets (subsets of Ω, in our case); and a class of sets is just a *set* whose elements happen to be sets; and that we have defined what the limit of a sequence of sets is in the cases of a non-decreasing sequence and of a non-increasing one.

The sequence of σ-algebras $\mathcal{F}_{\geq n}$, $n = 1, 2, 3, ...$, is non-increasing:

$$\mathcal{F}_{\geq n} = \sigma(\xi_n, \xi_{n+1}, \xi_{n+2}, \xi_{n+3}, ...) \supseteq \mathcal{F}_{\geq n+1} = \sigma(\xi_{n+1}, \xi_{n+2}, \xi_{n+3}, ...), \tag{11.12}$$

because the first σ-algebra has one random variable, namely ξ_n, more in its generators.

So by definition

$$\mathcal{F}_{\geq\infty} = \bigcap_{n=1}^{\infty} \mathcal{F}_{\geq n}. \tag{11.13}$$

Note that $\mathcal{F}_{\geq\infty}$ is defined as *the limit* at infinity, and not as $\sigma(\xi_i,\ i \geq \infty)$: there is *no* random variable ξ_∞ (or random variables ξ_i numbered with indices that are greater than ∞: all i's are just natural numbers).

The intersection of any number of σ-algebras is again a σ-algebra; so $\mathcal{F}_{\geq\infty}$ is one.

There are no random variables ξ_i that participate in generating this σ-algebra: ξ_n is excluded at the stage that we include $\mathcal{F}_{\geq n+1}$ in our intersection. Does it mean that the σ-algebra $\mathcal{F}_{\geq\infty}$ is the same as the σ-algebra in Ω generated by and empty set of random variables – i. e., the σ-algebra $\{\Omega,\ \emptyset\}$ consisting of two events only? No, as this example shows:

Let A_1, A_2, ..., A_n, ... be a sequence of events. Let the random variables ξ_i be their indicators: $\xi_i = I_{A_i}$. Then the event

$$\{\text{infinitely many of } A_i,\ i = 1,\, 2,\, ... \text{ occur}\} \tag{11.14}$$

belongs to the tail σ-algebra $\mathcal{F}_{\geq\infty}$.

Indeed, the event (11.14) is the same as the event

$$\{\text{infinitely many of } A_i,\ i = n,\, n+1,\, n+2,\, ... \text{ occur}\} = \bigcap_{k=n}^{\infty} \bigcup_{i=k}^{\infty} A_i; \tag{11.15}$$

this event is represented by applying countable set-theoretic operations to events A_n, A_{n+1}, A_{n+2}, ..., and so it belongs to the σ-algebra $\sigma\{A_n,\ A_{n+1},\ A_{n+2}, ...\}$ generated by these events, which is the same as the σ-algebra $\sigma(\xi_n, \xi_{n+1}, \xi_{n+2}, ...) = \mathcal{F}_{\geq n}$ generated by their indicators.

So we have:

$$\{\text{infinitely many of } A_i,\ i = 1, 2, \ldots \text{ occur}\} \in \mathcal{F}_{\geq n} \tag{11.16}$$

for every n, and the event (11.14) belongs to the intersection of these σ-algebras:

$$\{\text{infinitely many of } A_i,\ i = 1, 2, \ldots \text{ occur}\} \in \mathcal{F}_{\geq\infty}. \tag{11.17}$$

We can produce many more examples of "tail" events belonging to the σ-algebra $\mathcal{F}_{\geq\infty}$; I am going to show one more example, and see Problems $\boxed{\mathbf{25}}$–$\boxed{\mathbf{31}}$.

Let $\xi_1, \xi_2, ..., \xi_n, ...$ be a sequence of real-valued random variables. Let us prove that the event

$$\{\lim_{n\to\infty} \xi_n = a\} \tag{11.18}$$

belongs to the tail σ-algebra $\mathcal{F}_{\geq\infty}$ (a is an arbitrary real number).

We know that $\lim_{n\to\infty} \xi_i(\omega) = a$ means that for every $\varepsilon > 0$ there exists a natural k such that for every $i \geq k$ we have $|\xi_i(\omega) - a| < \varepsilon$; so

$$\{\omega\colon \lim_{n\to\infty} \xi_n(\omega) = a\} = \bigcap_{\varepsilon>0} \bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} \{\omega\colon |\xi_i(\omega) - a| < \varepsilon\}. \tag{11.19}$$

We feel a little uncomfortable having an *uncountable* intersection here: $\bigcap_{\varepsilon>0}$; applying uncountable operations to events may, in general, produce a set not belonging to our σ-algebra $\mathcal{F}$ of events. But we can take as ε all numbers of the form $1/m$ and rewrite the event as follows:

$$\{\omega\colon \lim_{n\to\infty} \xi_n(\omega) = a\} = \bigcap_{m=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} \{\omega\colon |\xi_i(\omega) - a| < 1/m\}. \tag{11.20}$$

Now, this is clearly and event, and even one belonging to $\mathcal{F}_{\geq 1}$, because all events $\{|\xi_i - a| < 1/m\}$ belong to it.

We can also rewrite this event in the form

$$\{\lim_{n\to\infty} \xi_n = a\} = \bigcap_{m=1}^{\infty} \bigcup_{k=n}^{\infty} \bigcap_{i=k}^{\infty} \{|\xi_i - a| < 1/m\} \tag{11.21}$$

for an arbitrary natural n; all events $\{|\xi_i - a| < 1/m\}$ here belong to the σ-algebra $\mathcal{F}_{\geq n}$, so the event $\{\lim_{n\to\infty} \xi_n = a\}$ belongs to the σ-algebra $\mathcal{F}_{\geq n}$ for every natural n; and so also to their intersection, the tail σ-algebra $\mathcal{F}_{\geq\infty}$.

Theorem 11.2 (the $0-1$ law). *Let $\xi_1, \xi_2, ..., \xi_n, ...$ be a sequence of* independent *random variables. Then every event belonging to the tail σ-algebra $\mathcal{F}_{\geq\infty}$ either has probability* 1, *or* 0.

The **proof** is based on a theorem from measure theory (which I am giving without proof):

Theorem 11.3. *Let* $\mathcal{A}$ *be an* algebra *in a space* X; *let* m *be a finite measure on the measurable space* $\big(X,\, \sigma(\mathcal{A})\big)$ (that is, on the σ-algebra generated by our algebra).

Then for every set $A \in \sigma(\mathcal{A})$ *and for every positive* ε *there exists a set* $A_\varepsilon \in \mathcal{A}$ *such that the* m*-measure of their symmetric difference is less than* ε:

$$m(A \,\Delta\, A_\varepsilon) < \varepsilon. \tag{11.22}$$

(The symmetric difference is the union of the two differences: $A \,\Delta\, A_\varepsilon = (A \setminus A_\varepsilon) \cup (A_\varepsilon \setminus A)$.)

In short: every set belonging to the σ-algebra generated by the algebra $\mathcal{A}$ can be approximated with arbitrary accuracy (in the sense explained in (11.22)) by sets from the algebra.

Let me show how this theorem is used to prove the zero-one law.

Let us take

$$\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_{\le n}. \tag{11.23}$$

This class of sets is an *algebra* in Ω (but not a σ-algebra: otherwise, since it contains all events of the form $\{\xi_i \in C_i\}$, it would contain the smallest σ-algebra $\mathcal{F}_{\ge 1}$ containing all such events – which is not the case).

Indeed, clearly $\Omega \in \mathcal{A}$ (it belongs to every summand in the union (11.23)). Now, about the complement: Let $A \in \mathcal{A}$. By its definition (11.23), this means that there exists a natural n such that $A \in \mathcal{F}_{\le n}$. Since $\mathcal{F}_{\le n}$ is a σ-algebra, we have $A^c \in \mathcal{F}_{\le n}$, and of course this complement belongs to the union (11.23).

Finally, about the union of two (or finitely many) sets. Let $A,\, B \in \mathcal{A}$; this means that there exist natural n and m such that $A \in \mathcal{F}_{\le n}$, $B \in \mathcal{F}_{\le m}$. Without restriction of generality we can assume that $m \ge n$. Clearly $\mathcal{F}_{\le n} \le \mathcal{F}_{\le m}$, because the second σ-algebra is generated by (the same or) a larger number of random variables; so we have also $A \in \mathcal{F}_{\le m}$. From $A \in \mathcal{F}_{\le m}$, $B \in \mathcal{F}_{\le m}$ we obtain: $A \cup B \in \mathcal{F}_{\le m} \subseteq \mathcal{A}$.

Clearly, the σ-algebra $\mathcal{F}_{\ge 1}$ is generated by the algebra $\mathcal{A}$.

By Theorem 11.3, for every $A \in \mathcal{F}_{\ge 1}$ and every positive ε there exists an event $A_\varepsilon \in \mathcal{A}$ such that (11.22) is satisfied. In other words, for every event $A \in \mathcal{F}_{\ge 1}$ and every $\varepsilon > 0$ there exists a natural n and an event $A_\varepsilon \in \mathcal{F}_{\le n}$ such that (11.22) holds.

Let us apply this to an event A belonging to the tail σ-algebra $\mathcal{F}_{\ge\infty}$ ($\subseteq \mathcal{F}_{\ge 1}$): for every positive ε there exist an event $A_\varepsilon \in \mathcal{F}_{\le n}$ such that we have (11.22).

But since the event A belongs to $\mathcal{F}_{\ge\infty}$, it belongs also to the σ-algebra $\mathcal{F}_{\ge n+1}$ (because of the definition (11.13): clearly, $\bigcap_{k=1}^{\infty} \mathcal{F}_{\ge k} \subseteq \mathcal{F}_{\ge n+1}$).

So we have: $A_\varepsilon \in \mathcal{F}_{\le n}$, $A \in \mathcal{F}_{\ge n+1}$. These σ-algebras are generated by disjoint sets of independent random variables (the first one by ξ_1, ..., ξ_n, and the second by ξ_{n+1}, ξ_{n+2}, ξ_{n+3}, ...). So these events are independent:

$$P(A \cap A_\varepsilon) = P(A) \cdot P(A_\varepsilon). \tag{11.24}$$

We have:

$$P(A_\varepsilon) = P(A) - P(A \setminus A_\varepsilon) + P(A_\varepsilon \setminus A) \tag{11.25}$$

(make a picture to see why this is true). Since $P(A\,\Delta A_{\,\varepsilon}) = P(A \setminus A_\varepsilon) + P(A_\varepsilon \setminus A) < \varepsilon$, we have:

$$|P(A_\varepsilon) - P(A)| < \varepsilon. \tag{11.26}$$

And since

$$(A \cap A_\varepsilon)\,\Delta\,A = \big((A \cap A_\varepsilon) \setminus A\big) \cup \big(A \setminus (A \cap A_\varepsilon)\big) = \emptyset \cup (A \setminus A_\varepsilon), \tag{11.27}$$

we have also

$$P\big((A \cap A_\varepsilon)\,\Delta\,A\big) < \varepsilon, \tag{11.28}$$

$$|P(A \cap A_\varepsilon) - P(A)| < \varepsilon. \tag{11.29}$$

Multiplying (11.26) by $P(A)$, we get:

$$|P(A) \cdot P(A_\varepsilon) - P(A)^2| \le P(A) \cdot \varepsilon \le \varepsilon; \tag{11.30}$$

and this, together with (11.24) and (11.29), yields:

$$|P(A) - P(A)^2| < 2\varepsilon. \tag{11.31}$$

Since ε was an *arbitrary* positive number, we conclude that

$$P(A)^2 = P(A). \tag{11.32}$$

This quadratic equation has only two solutions: $P(A) = 0$ and $P(A) = 1$.

The theorem is proved.

Many concrete things follow from this theorem: that a series with independent summands can either converge almost surely, or diverge almost surely; that the distribution function of the random variable $\eta = \overline{\lim}_{n\to\infty} \dfrac{\xi_1 + ... + \xi_n}{n}$ can only take values 0 or 1 (what can you say about such a random variable?); etc.