

Lecture 12. Convergence of sequences of random variables.

All of us have heard the following (or something like this): the relative frequency of an event in a very large number of repetitions of the experiment (under the same conditions) is approximately equal to its probability; or: the relative frequency of an event in n repetitions of an experiment becomes closer and closer to the probability of the event as the number n of repetitions grows.

For random variables, this can be reformulated as follows: if a real-valued random variable has a (finite) expectation, the arithmetic mean of its values in a large number of repetitions of our experiment is close to the expectation of our random variable; or: becomes closer and closer to it as the number of experiments grows.

We can see that this is, basically, the same thing if we consider, say, a random variable ξ taking three values: a , b , and c ; the arithmetic mean of the values of our random variable in n experiments is equal to a times the relative frequency of the event $\{\xi = a\}$, plus the same for b and c , and this is approximately equal to the expectation.

These are statements belonging not to *mathematics*: rather their place is at the borderline between mathematics and the extra-mathematical world: just outside mathematics.

But we can aspire to create a *mathematical model* for this.

Something has already been done in this direction: an idealized mathematical model for observing the values of a random variable in repeated experiments is a *sequence* of independent random variables; we are happy now that such a thing does exist (as well as an infinite sequence of independent events with any probabilities).

So we may aspire to have mathematical theorems stating that for a sequence of random variables $\xi_1, \xi_2, \dots, \xi_n, \dots$, under such and such conditions, the arithmetic mean $\frac{\xi_1 + \dots + \xi_n}{n}$ of the first n of them converges to... But here we stop for some time: we haven't specified what conditions are imposed on the random variables ξ_i , and we even don't know whether they all have the same expectation $E\xi_i$.

However, it is even more important that we haven't specified in what sense the convergence of random variables should be understood here.

You see, random variables are *functions* (of the sample point ω); and whereas we have, basically, only one concept of convergence for sequences of numbers, we have many types of convergence for *functions*: e. g., pointwise convergence; uniform convergence; or \mathbf{L}^p -convergence; etc. So we have to introduce and study some types of convergence for random variables.

It turns out that uniform convergence or convergence at all points do not play any significant role in probability theory. Let me introduce two other types of convergence.

Let $\eta_1, \eta_2, \dots, \eta_n, \dots$ be a sequence of random variables on a probability space (Ω, \mathcal{F}, P) ; ζ , a random variable on the same space. We say that the sequence η_n converges in probability to ζ (the notations: $\eta_n \rightarrow_P \zeta$ ($n \rightarrow \infty$), or: $(P) \lim_{n \rightarrow \infty} \eta_n = \zeta$) if for every positive ε

$$P\{|\eta_n - \zeta| < \varepsilon\} \rightarrow 1 \quad (n \rightarrow \infty). \quad (12.1)$$

Of course, (12.1) is equivalent to

$$\lim_{n \rightarrow \infty} P\{|\eta_n - \zeta| \geq \varepsilon\} = 0, \quad \varepsilon > 0. \quad (12.2)$$

We can define convergence in probability also for random variables taking values in an arbitrary metric space X , supposing the distance $\text{dist}(x, y)$ is a measurable function of the pair (x, y) , by

$$P\{\text{dist}(\eta_n, \zeta) < \varepsilon\} \rightarrow 1 \quad (n \rightarrow \infty); \quad (12.3)$$

but for simplicity's sake we'll stick to $X = \mathbb{R}^1$ or \mathbb{R}^n .

We say that the sequence of η_n converges to ζ *almost surely* if

$$P\{\lim_{n \rightarrow \infty} \eta_n = \zeta\} = P\{\omega: \lim_{n \rightarrow \infty} \eta_n(\omega) = \zeta(\omega)\} = 1, \quad (12.4)$$

or, equivalently,

$$P\{\eta_n \not\rightarrow \zeta \ (n \rightarrow \infty)\} = 0. \quad (12.5)$$

Of course, we have to check first that the ω -sets in (12.4) or (12.5) are indeed events. The set under the probability sign in (12.4) consists of all ω such that the limit *exists*, and is equal to $\zeta(\omega)$. The set

$$\{\omega: \lim_{n \rightarrow \infty} \eta_n(\omega) \text{ exists}\} \quad (12.6)$$

belongs to \mathcal{F} (is an event), and the limit is an \mathcal{F} -measurable function on this set (a fact from the set-theoretic introduction to measure theory, see Lecture 4, the text around formula (4.39)). So the set under the probability sign in (12.4) is one on which the \mathcal{F} -measurable function $\lim_{n \rightarrow \infty} \eta_n(\omega) - \zeta(\omega)$ belongs to the Borel set consisting of one point 0, and so this set belongs to \mathcal{F} (is an event).

Convergence in probability and almost sure convergence are considered also in measure theory, under the names of *convergence in measure* and *convergence almost everywhere* (remember that systematic disregard of sets of zero measure or events of zero probability is a common trait of measure theory and probability theory).

The types of convergence that we introduced have some properties that we are not accustomed to: namely, the limit in probability – or the almost-sure limit – is not unique, in general. Indeed, if we have in our probability space some non-empty events having probability 0 (and such is the situation, for example, always when we are considering continuous random variables), we can change the limiting random variable ζ arbitrarily on such a set of probability measure 0, and still it will be a version of the same limit. Let us say that a random variable ζ' is equivalent to the random variable ζ if $P\{\zeta' \neq \zeta\} = 0$. Then $\eta_n \rightarrow_P \zeta'$ if and only if $\eta_n \rightarrow_P \zeta$, and $\eta_n \rightarrow \zeta'$ almost surely if and only if $\eta_n \rightarrow \zeta$ almost surely. This is because for equivalent random variables ζ and ζ' we have:

$$\{|\eta_n - \zeta'| < \varepsilon\} \Delta \{|\eta_n - \zeta| < \varepsilon\} \subseteq \{\zeta' \neq \zeta\}, \quad (12.7)$$

$$|P\{|\eta_n - \zeta'| < \varepsilon\} - P\{|\eta_n - \zeta| < \varepsilon\}| \leq P\{\zeta' \neq \zeta\} = 0, \quad (12.8)$$

and also

$$|P\{\lim_{n \rightarrow \infty} \eta_n = \zeta'\} - P\{\lim_{n \rightarrow \infty} \eta_n = \zeta\}| \leq P\{\zeta' \neq \zeta\} = 0. \quad (12.9)$$

To overcome this difficulty (or: to go back to what we are accustomed to), we can consider these convergences not for random variables, but rather on the set of *equivalence classes* of random variables; or we can just state that the limits in these senses are not unique, but *almost* unique.

But we have to prove that if $\eta_n \rightarrow_P \zeta$ and $\eta_n \rightarrow_P \zeta'$, or if both $\eta_n \rightarrow \zeta$ and $\eta_n \rightarrow \zeta'$ almost surely, then $\zeta' \sim \zeta$ (ζ' is equivalent to ζ , which means that $P\{\zeta' \neq \zeta\} = 0$).

Suppose first that $\eta_n \rightarrow_P \zeta$ and $\eta_n \rightarrow_P \zeta'$. This means that (12.2) holds for every positive ε , and also

$$P\{|\eta_n - \zeta'| \geq \varepsilon\} \rightarrow 0 \quad (n \rightarrow \infty). \quad (12.10)$$

We have, for any positive ε and natural n :

$$\{|\zeta' - \zeta| \geq 2\varepsilon\} \subseteq \{|\eta_n - \zeta'| \geq \varepsilon\} \cup \{|\eta_n - \zeta| \geq \varepsilon\}, \quad (12.11)$$

so

$$\begin{aligned} P\{|\zeta' - \zeta| \geq 2\varepsilon\} &\leq P(\{|\eta_n - \zeta'| \geq \varepsilon\} \cup \{|\eta_n - \zeta| \geq \varepsilon\}) \\ &\leq P\{|\eta_n - \zeta'| \geq \varepsilon\} + P\{|\eta_n - \zeta| \geq \varepsilon\} \rightarrow 0 \quad (n \rightarrow \infty). \end{aligned} \quad (12.12)$$

Since $\varepsilon > 0$ is arbitrary, we get:

$$P\{\zeta' \neq \zeta\} = P\left(\bigcup_{m=1}^{\infty} \{|\zeta' - \zeta| \geq 2/m\}\right) = \lim_{m \rightarrow \infty} P\{|\zeta' - \zeta| \geq 2/m\} = 0 \quad (12.13)$$

(because $\{\zeta' \neq \zeta\} = \bigcup_{m=1}^{\infty} \{|\zeta' - \zeta| \geq 2/m\}$).

For the almost-sure convergence it is *quite* simple:

$$\{\zeta' \neq \zeta\} \subseteq \{\lim_{n \rightarrow \infty} \eta_n \neq \zeta\} \cup \{\lim_{n \rightarrow \infty} \eta_n \neq \zeta'\}, \quad (12.14)$$

$$P\{\zeta' \neq \zeta\} \leq P\{\lim_{n \rightarrow \infty} \eta_n \neq \zeta\} + P\{\lim_{n \rightarrow \infty} \eta_n \neq \zeta'\} = 0. \quad (12.15)$$

It turns out that almost sure convergence is *stronger* than convergence in probability:

Theorem 12.1. *Let $\eta_n \rightarrow \zeta$ almost surely. Then $\eta_n \rightarrow_P \zeta$.*

Proof. Suppose $\eta_n \rightarrow \zeta$ almost surely. Then we have, for an arbitrary positive ε (see a similar formula (11.19)):

$$\{\omega: \eta_n(\omega) \rightarrow \zeta(\omega)\} = \bigcap_{\varepsilon > 0} \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} \{\omega: |\eta_i(\omega) - \zeta(\omega)| < \varepsilon\} \subseteq \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} \{\omega: |\eta_i(\omega) - \zeta(\omega)| < \varepsilon\}, \quad (12.16)$$

$$1 = P\{\eta_n \rightarrow \zeta\} \leq P\left(\bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} \{|\eta_i - \zeta| < \varepsilon\}\right). \quad (12.17)$$

So the probability in the right-hand side is equal to 1. Let us go to the complements:

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} \{|\eta_i - \zeta| \geq \varepsilon\}\right) = 0. \quad (12.18)$$

But the events $\bigcup_{i=n}^{\infty} \{|\eta_i - \zeta| \geq \varepsilon\}$ form a non-increasing sequence, so the $\bigcap_{n=1}^{\infty}$ in (12.18) is nothing but the limit $\lim_{n \rightarrow \infty}$ of these events, and the probability of the limit is equal to the limit of the probabilities:

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{i=n}^{\infty} \{|\eta_i - \zeta| \geq \varepsilon\}\right) = 0. \quad (12.19)$$

Now we use the fact that $\{|\eta_n - \zeta| \geq \varepsilon\} \subseteq \bigcup_{i=n}^{\infty} \{|\eta_i - \zeta| \geq \varepsilon\}$, and get:

$$\lim_{n \rightarrow \infty} P\{|\eta_n - \zeta| \geq \varepsilon\} \leq \lim_{n \rightarrow \infty} P\left(\bigcup_{i=n}^{\infty} \{|\eta_i - \zeta| \geq \varepsilon\}\right) = 0. \quad (12.20)$$

As a matter of fact, we did not prove that almost sure convergence is stronger than that in probability: we proved rather that it is *not weaker*. But in fact it *is* stronger: there are cases in which convergence in probability does take place, but not almost sure convergence.

Example (rather a whole class of such): Let $A_1, A_2, \dots, A_n, \dots$ be a sequence of independent events such that $\lim_{n \rightarrow \infty} P(A_i) = 0$, but $\sum_{i=1}^{\infty} P(A_i) = \infty$. Let $\xi_i, i = 1, 2, \dots, n, \dots$, be the corresponding indicators: $\xi_i = I_{A_i}$.

We have, for every $\varepsilon \in (0, 1)$:

$$P\{|\xi_n - 0| \geq \varepsilon\} = P\{\xi_n = 1\} = P(A_n) \rightarrow 0 \quad (n \rightarrow \infty), \quad (12.21)$$

so $\xi_n \rightarrow_P 0$; but almost surely there occur infinitely many of the events A_i , so almost surely among $\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega), \dots$ there are infinitely many 1's, and $\xi_n(\omega) \not\rightarrow 0$. So not only don't we have almost sure convergence, but almost surely we have non-convergence (the probability $P\{\xi_n \rightarrow 0\}$ is not equal to 1, but rather to 0).

So, convergence in probability does not generally imply almost sure convergence. But a weaker result can be proved:

Theorem 12.2. *Let $\eta_n \rightarrow_P \zeta$ ($n \rightarrow \infty$). Then there exists a subsequence $\eta_{n_i}, n_1 < n_2 < \dots < n_k < \dots$, such that $\eta_{n_k} \rightarrow \zeta$ ($k \rightarrow \infty$) almost surely.*

Proof. Let us take a sequence of positive $\varepsilon_m \rightarrow 0$ ($\varepsilon_m = 1/m$ would do). Since

$$P\{|\eta_n - \zeta| \geq \varepsilon_1\} \rightarrow 0 \quad (n \rightarrow \infty), \quad (12.22)$$

there is a sequence $n_1^1 < n_2^1 < \dots < n_k^1 < \dots$ such that

$$P\{|\eta_{n_k^1} - \zeta| \geq \varepsilon_1\} < 1/2^k. \quad (12.23)$$

Now we go to ε_2 : we have:

$$P\{|\eta_{n_k^1} - \zeta| \geq \varepsilon_2\} \rightarrow 0 \quad (k \rightarrow \infty). \quad (12.24)$$

From the sequence $n_1^1 < n_2^1 < \dots < n_k^1 < \dots$ we can extract a subsequence $n_1^2 < n_2^2 < \dots < n_k^2 < \dots$ ($\{n_1^2, n_2^2, \dots, n_k^2, \dots\} \subseteq \{n_1^1, n_2^1, \dots, n_k^1, \dots\}$) such that

$$P\{|\eta_{n_k^2} - \zeta| \geq \varepsilon_2\} < 1/2^k. \quad (12.25)$$

In the same way we take a subsubsequence $n_1^3 < n_2^3 < \dots < n_k^3 < \dots$, $\{n_1^3, n_2^3, \dots, n_k^3, \dots\} \subseteq \{n_1^2, n_2^2, \dots, n_k^2, \dots\}$, such that

$$P\{|\eta_{n_k^3} - \zeta| \geq \varepsilon_3\} < 1/2^k; \quad (12.26)$$

etc.

We get an infinite sequence of sequences $n_1^j < n_2^j < \dots < n_k^j < \dots$, of which each next one is a subsequence of the previous one, with

$$P\{|\eta_{n_k^j} - \zeta| \geq \varepsilon_j\} < 1/2^k. \quad (12.27)$$

Trying to take the intersection of all these sequences $\bigcap_{j=1}^{\infty} \{n_1^j, n_2^j, \dots, n_k^j, \dots\}$ wouldn't do: in all probability this intersection is empty. So we take the sequence

$$n_k = n_k^k, \quad k = 1, 2, 3, \dots \quad (12.28)$$

(this standard trick is called *the diagonal process*: we write our sequences $n_1^j, n_2^j, \dots, n_k^j, \dots$ one under another, and take the sequence of elements n_k^k standing on the "diagonal"). The sequence $n_1, n_2, \dots, n_k, \dots$ is, clearly, a subsequence of the first sequence $n_1^1, n_2^1, \dots, n_k^1, \dots$; it is a subsequence of the second sequence $n_1^2, n_2^2, \dots, n_k^2, \dots$ starting with the term n_2 ; etc.

Because of $\{n_1^2, n_2^2, \dots, n_k^2, \dots\} \subseteq \{n_1^1, n_2^1, \dots, n_k^1, \dots\}$, every term in the second sequence is \geq than the corresponding term in the first sequence:

$$n_k^2 \geq n_k^1; \quad (12.29)$$

in general,

$$n_k^1 \leq n_k^2 \leq \dots \leq n_k^k \leq \dots \quad (12.30)$$

For every $k \geq m$ the number $n_k = n_k^k$ is also an element of the sequence $n_1^m, n_2^m, n_3^m, \dots$ under some number j :

$$n_k = n_j^m. \quad (12.31)$$

Because of (12.30), this number j is $\geq k$.

So we have, for $k \geq m$:

$$P\{|\eta_{n_k} - \zeta| \geq \varepsilon_m\} = P\{|\eta_{n_j^m} - \zeta| \geq \varepsilon_m\} \leq 1/2^j \leq 1/2^k. \quad (12.32)$$

For every natural m , the series $\sum_{k=1}^{\infty} P\{|\eta_{n_k} - \zeta| \geq \varepsilon_m\}$ converges, because its terms, starting with the m -th, are dominated by the geometric sequence $1/2^k$; so by the first Borel–Cantelli Lemma, almost surely only finitely many of the events $\{|\eta_{n_k} - \zeta| \geq \varepsilon_m\}$ occur:

$$P\left(\bigcap_{k=1}^{\infty} \bigcup_{i=k}^{\infty} \{|\eta_{n_k} - \zeta| \geq \varepsilon_m\}\right) = 0. \quad (12.33)$$

The probability of the union of countably many events of probability 0 also has zero probability:

$$P\left(\bigcup_{m=1}^{\infty} \bigcap_{k=1}^{\infty} \bigcup_{i=k}^{\infty} \{|\eta_{n_k} - \zeta| \geq \varepsilon_m\}\right) = 0; \quad (12.34)$$

or, passing to opposite events (the complements):

$$P\left(\bigcap_{m=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} \{|\eta_{n_k} - \zeta| < \varepsilon_m\}\right) = 1. \quad (12.35)$$

But the event under the sign of probability here is the same as $\{\omega : \eta_{n_k}(\omega) \rightarrow \zeta(\omega) \text{ (} k \rightarrow \infty)\}$. So we have established the almost sure convergence for the subsequence η_{n_k} .

Theorem 12.3. *We have $\eta_n \rightarrow_P \zeta$ if and only if from every sequence n_k , $k = 1, 2, 3, \dots$, we can extract a subsequence n_{k_j} , $j = 1, 2, 3, \dots$, such that*

$$\eta_{n_{k_j}} \rightarrow \zeta \quad \text{almost surely} \quad (j \rightarrow \infty). \quad (12.36)$$

Proof. The “only if” part is just Theorem 12.2 applied to the subsequence η_{n_k} , which clearly converges in probability to ζ . Let us prove the “if” part.

Suppose that from every subsequence η_{n_k} we can extract an almost surely convergent subsubsequence $\eta_{n_{k_j}} \xrightarrow{\text{a.s.}} \zeta$, but $\eta_n \not\rightarrow_P \zeta$. This means that for some positive ε_0

$$P\{|\eta_n - \zeta| \geq \varepsilon_0\} \not\rightarrow 0 \quad (n \rightarrow \infty). \quad (12.37)$$

This, in its turn, means that there exists a positive δ and a sequence $n_1 < n_2 < \dots < n_k < \dots$ such that

$$P\{|\eta_{n_k} - \zeta| \geq \varepsilon_0\} \geq \delta, \quad k = 1, 2, 3, \dots \quad (12.38)$$

By what we supposed, from the sequence n_k , $k = 1, 2, 3, \dots$, we can extract a subsequence n_{k_j} , $j = 1, 2, 3, \dots$, such that (12.36) holds. By Theorem 12.1 we get $\eta_{n_{k_j}} \rightarrow_P \zeta$, and $\lim_{j \rightarrow \infty} P\{|\eta_{n_{k_j}} - \zeta| \geq \delta\} = 0$. But this contradicts (12.38).

In the next lecture I am planning to start on *laws of large numbers*.

Everybody who hasn't studied probability theory has heard about the law of large numbers: what exactly it is, they don't know, but they know that it is some law related to probability theory. People with more mathematical skills can be more precise: some *theorem* in probability theory.

This is not entirely true. In fact, *the Law of Large Numbers* is a common name for a whole *group* of theorems of probability theory. Their formulations differ from one another, but share the general pattern:

Let a sequence of random variables $\xi_1, \dots, \xi_n, \dots$ be such that expectations $E\xi_i$ exist; let such and such conditions be satisfied (the conditions are different in different theorems belonging to this group). Then

$$\frac{\xi_1 + \dots + \xi_n}{n} - \frac{E\xi_1 + \dots + E\xi_n}{n} \longrightarrow 0 \quad (n \rightarrow \infty), \quad (12.39)$$

where the long arrow \longrightarrow denotes some type of convergence of random variables.

If the random variables ξ_i have all the same distribution (are *identically distributed*), (12.39) can be rewritten as

$$\frac{\xi_1 + \dots + \xi_n}{n} \longrightarrow E\xi_i \quad (n \rightarrow \infty); \quad (12.40)$$

but there are theorems of this group in which the random variables ξ_i have different distributions, and different expectations.

We introduced two types of convergence: convergence in probability, and almost sure convergence, the first being weaker than the second; accordingly we have *weak* laws of large numbers, in which (12.39) means

$$\frac{\xi_1 + \dots + \xi_n}{n} - \frac{E\xi_1 + \dots + E\xi_n}{n} \xrightarrow{P} 0 \quad (n \rightarrow \infty), \quad (12.41)$$

and *strong* laws of large numbers that state that under certain conditions

$$\frac{\xi_1 + \dots + \xi_n}{n} - \frac{E\xi_1 + \dots + E\xi_n}{n} \rightarrow 0 \quad \text{a.s.} \quad (n \rightarrow \infty). \quad (12.42)$$

In most of laws of large numbers the conditions that we haven't yet specified include the requirement that $\xi_1, \xi_2, \dots, \xi_n, \dots$ should be independent; but there are some such laws that are formulated for dependent random variables.

But before we go to laws of large numbers, we return to some more elementary things in our course.

Theorem 12.4: the Chebyshev(-type) inequality. Let ξ be a random variable with values in (X, \mathcal{X}) . If $f(x)$ is a measurable function on X with nonnegative values (i.e., with values in $[0, \infty)$ – or in $[0, \infty]$), we have for every $a > 0$:

$$P\{f(\xi) \geq a\} \leq \frac{Ef(\xi)}{a}. \quad (12.43)$$

Proof. Let η be the random variable defined by

$$\eta = \begin{cases} a & \text{if } f(\xi) \geq a, \\ 0 & \text{if } f(\xi) < a. \end{cases} \quad (12.44)$$

It is clear that $\eta(\omega) \leq f(\xi(\omega))$ for all ω , so

$$E\eta = a \cdot P\{f(\xi) \geq a\} \leq Ef(\xi). \quad (12.45)$$

The inequality that Chebyshev proved (not that there was much to prove) and used, which we'll call *Chebyshev's inequality* (whose generalization is inequality (12.43)), is

$$P\{|\xi - E\xi| \geq b\} \leq \frac{\text{Var}(\xi)}{b^2}, \quad b > 0, \quad (12.46)$$

where $\text{Var}(\xi)$ is the variance of the random variable ξ : $\text{Var}(\xi) = E(\xi - E\xi)^2$.

This is a particular case of (12.43) with $f(x) = (x - E\xi)^2$, $a = b^2$.