Lecture 13. Convergence of sequences of random variables. Laws of large numbers.

Theorem 13.1. *We have* $\eta_n \to_P \zeta$ *if and only if from every sequence* n_k, $k = 1, 2, 3, \ldots$, *we can extract a subsequence* n_{k_j}, $j = 1, 2, 3, \ldots$, *such that*

$$\eta_{n_{k_j}} \to \zeta \quad \text{almost surely} \quad (j \to \infty). \tag{13.1}$$

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 $\eta_{n_{k_j}} \to^{\text{a.s.}} \zeta$, but $\eta_n \not\to_P \zeta$. This means that for some positive ε_0

$$P\{|\eta_n - \zeta| \geq \varepsilon_0\} \not\to 0 \qquad (n \to \infty). \tag{13.2}$$

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

$$P\{|\eta_{n_k} - \zeta| \geq \varepsilon_0\} \geq \delta, \qquad k = 1, 2, 3, \ldots. \tag{13.3}$$

By what we supposed, from the sequence n_k, $k = 1, 2, 3, \ldots$, we can extract a subsequence n_{k_j}, $j = 1, 2, 3, \ldots$, such that (13.1) holds. By Theorem 12.1 we get $\eta_{n_{k_j}} \to_P \zeta$, and $\lim_{j\to\infty} P\{|\eta_{n_{k_j}} - \zeta| \geq \delta\} = 0$. But this contradicts (13.3).

The properties of *almost-sure* convergence are very easy to derive: e. g. that a finite-valued almost-sure limit of η_n exists if and only if $\eta_n - \eta_m \to 0$ almost surely as m, $n \to \infty$: we simply disregard the zero-measure set of ω's at which $\eta_n - \eta_m \not\to 0$, and for all other ω's a finite $\lim_{n\to\infty} \eta_n(\omega)$ exists (how we define the limiting random variable on the exceptional zero-measure set does not matter).

An example of handling convergence *in probability*:

Theorem 13.2. *Let* $\eta_n \to_P \zeta$ $(n \to \infty)$. *If* f *is a continuous function, then* $f(\eta_n) \to_P f(\zeta)$ $(n \to \infty)$.

Proof: Proof 1. We have to prove that for every positive ε and δ for sufficiently large n

$$P\{|f(\eta_n) - f(\zeta)| \geq \varepsilon\} < \delta. \tag{13.4}$$

First we take a constant C such that

$$P\{|\zeta| > C\} < \delta/2. \tag{13.5}$$

Now, on the interval $[-C - 1, C + 1]$ the function f is *uniformly* continuous; so there exists a positive γ such that for every x, $y \in [-C - 1, C + 1]$, $|x - y| < \gamma$,

$$|f(x) - f(y)| < \varepsilon. \tag{13.6}$$

Let us take $\gamma < 1$.

For sufficiently large n,

$$P\{|\eta_n - \zeta| \geq \gamma\} < \delta/2 \tag{13.7}$$

Then, using the set inclusion

$$\{\omega\colon\, |f(\eta_n(\omega) - \zeta(\omega)| \geq \varepsilon\} \subseteq \{\omega\colon\, |\zeta(\omega)| \geq C\} \cup \{\omega\colon\, |\zeta(\omega)| < C,\, |\eta_n(\omega) - \zeta(\omega)| \geq \gamma\}, \tag{13.8}$$

we obtain (13.4).

Proof 2: By Theorem 13.1, it is sufficient to prove that from every sequence of natural numbers $n_i \to \infty$ we can select a subsequence n_{i_k} such that

$$f(\eta_{n_{i_k}}) \to f(\zeta) \quad \text{a. s.}. \tag{13.9}$$

A subsequence for which

$$\eta_{n_{i_k}} \to \zeta \quad \text{a. s.}, \tag{13.10}$$

which exists by the other part of the same theorem, will do.

Let us have some results about convergence of sequences of random variables in terms of *moments*. Before doing this, we return to some more elementary things in our course.

Theorem 13.3: 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* $a > 0$:

$$P\{f(\xi) \geq a\} \leq \frac{Ef(\xi)}{a}. \tag{13.11}$$

Proof. Let η be the random variable defined by

$$\eta = \begin{cases} a & \text{if } f(\xi) \geq a, \\ 0 & \text{if } f(\xi) < a. \end{cases} \tag{13.12}$$

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

$$E\eta = a \cdot P\{f(\xi) \geq a\} \leq Ef(\xi). \tag{13.13}$$

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 (13.11)), is

$$P\{|\xi - E\xi| \geq b\} \leq \frac{\operatorname{Var}(\xi)}{b^2}, \qquad b > 0, \tag{13.14}$$

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

This is a particular case of (13.11) with $f(x) = (x - E\xi)^2$, $a = b^2$.

Theorem 13.4. *If for some even k the moment*

$$E(\eta_n - \zeta)^k \to 0 \qquad (n \to \infty), \tag{13.15}$$

then $\eta_n \to_P \zeta$ $(n \to \infty)$.

Proof. By the Chebyshev inequality (13.11) with $f(x) = x^k$, $a = \varepsilon^k$, we have, for every positive ε:

$$P\{|\eta_n - \zeta| \ge \varepsilon\} \le \frac{E(\eta_n - \zeta)^k}{\varepsilon^k} \to 0 \qquad (n \to \infty). \tag{13.16}$$

Theorem 13.5. *If for some even k we have*

$$\sum_{i=1}^{\infty} E(\eta_i - \zeta)^k < \infty, \tag{13.17}$$

then $\eta_n \to \zeta$ almost surely as $n \to \infty$.

Proof. Proof 1. If we have a convergent series $\sum_{i=1}^{\infty} a_i$ with nonnegative summands, there exists another convergent series $\sum_{i=1}^{\infty} a'_i$ whose terms converge to 0 slower than a_i: that is, there exists a sequence of positive numbers $b_i \to 0$ $(i \to \infty)$ such that $\sum_{i=1}^{\infty} a_i/b_i < \infty$. (In the general spirit of this course, I am giving no proof of this fact, which is not about probability theory, and even not about measure theory; the proof is simple, in fact.) So there exists a sequence of positive numbers $b_i \to \infty$ such that

$$\sum_{i=1}^{\infty} \frac{E(\eta_i - \zeta)^k}{b_i^k} < \infty. \tag{13.18}$$

By the Chebyshev inequality (13.11), we have:

$$\sum_{i=1}^{\infty} P\{|\eta_i - \zeta| \ge b_i\} \le \sum_{i=1}^{\infty} \frac{E(\eta_i - \zeta)^k}{b_i^k} < \infty. \tag{13.19}$$

By the first Borel – Cantelli Lemma, almost surely only finitely many of the events $\{|\eta_i - \zeta| \ge b_i\}$ will occur: for all ω except for a set of probability 0 we have $|\eta_i(\omega) - \zeta(\omega)| < b_i$ for all i with exception of a finite number of i's. Since $b_i \to 0$, convergence $\eta_n(\omega) \to \zeta(\omega)$ will hold for all such ω's.

Proof 2. Let us consider the following random variable taking values in the extended right half-line $[0, \infty]$:

$$\Xi = \sum_{i=1}^{\infty} (\eta_i - \zeta)^k. \tag{13.20}$$

We have:

$$E\Xi = \sum_{i=1}^{\infty} E(\eta_i - \zeta)^k < \infty. \tag{13.21}$$

It follows from this that almost surely $\Xi < \infty$ ($\Xi(\omega) < \infty$ except for a set of ω's of zero probability), the series $\sum_{i=1}^{\infty}(\eta_i - \zeta)^k$ converges almost surely. We know that the n-th term of a convergent series goes to 0 as $n \to \infty$; so almost surely

$$(\eta_n - \zeta)^k \to 0 \qquad (n \to \infty), \tag{13.22}$$

from which $\eta_n \to \zeta$ almost surely.

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*

$$\frac{\xi_1 + \dots + \xi_n}{n} - \frac{E\xi_1 + \dots + E\xi_n}{n} \longrightarrow 0 \qquad (n \to \infty), \tag{13.23}$$

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*), (13.23) can be rewritten as

$$\frac{\xi_1 + \dots + \xi_n}{n} \longrightarrow E\xi_i \qquad (n \to \infty); \tag{13.24}$$

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 (13.23) means

$$\frac{\xi_1 + \dots + \xi_n}{n} - \frac{E\xi_1 + \dots + E\xi_n}{n} \to_P 0 \qquad (n \to \infty), \tag{13.25}$$

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

$$\frac{\xi_1 + \dots + \xi_n}{n} - \frac{E\xi_1 + \dots + E\xi_n}{n} \to 0 \quad \text{a.s.} \quad (n \to \infty). \tag{13.26}$$

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.

Let us start with some theorems belonging to the group of *weak* laws of large numbers.

Theorem 13.6 (Bernoulli's Theorem). *Let* $\xi_1, ..., \xi_n, ...$ *be a sequence of independent random variables taking values* 0 *and* 1 *with probabilities* $1-p$ *and* p. *Then*

$$\frac{\xi_1 + ... + \xi_n}{n} \to_P p \qquad (n \to \infty). \tag{13.27}$$

We will not be proving this theorem now, but better formulate a more general one belonging to the same group:

Theorem 13.7 (Chebyshev's Theorem). *Let* $\xi_1, ..., \xi_n, ...$ *be a sequence of independent* (not necessarily identically distributed) *random variables with finite expectations and variances. Let the sequence of variances be bounded*:

$$\mathrm{Var}(\xi_i) \le C = \mathrm{const} < \infty \tag{13.28}$$

for all natural n. *Then* (13.25) *is satisfied.*

Before we prove this theorem, let me formulate another one of the same group:

Khinchin's Theorem (proved in the 1930's). *Let* $\xi_1, ..., \xi_n, ...$ *be a sequence of independent identically distributed random variables with finite expectation* $E\xi_i$. *Then*

$$\frac{\xi_1 + ... + \xi_n}{n} \to_P E\xi_i \qquad (n \to \infty). \tag{13.29}$$

Clearly Bernoulli's Theorem is a particular case of both Chebyshev's and Khinchin's Theorem. We'll prove Khinchin's Theorem later; and we'll prove Chebyshev's Theorem in the next lecture.

The following wasn't in Lecture # 13; but I'm including it in this lecture note.

Proof of Chebyshev's Theorem. We apply Theorem 13.4 with $k = 2$ (that is, the classical Chebyshev's inequality) to the random variables $\eta_n = \frac{\xi_1 + ... + \xi_n}{n} - \frac{E\xi_1 + ... + E\xi_n}{n}$ and $\zeta \equiv 0$. We have:

$$E\Big(\frac{\xi_1 + ... + \xi_n}{n} - \frac{E\xi_1 + ... + E\xi_n}{n}\Big)^2 = \mathrm{Var}\Big(\frac{\xi_1 + ... + \xi_n}{n}\Big) = \frac{1}{n^2} \cdot E\Big(\sum_{i=1}^{n} (\xi_i - E\xi_i)\Big)^2. \tag{13.30}$$

Opening the parentheses, we get:

$$E\Big(\sum_{i=1}^{n} (\xi_i - E\xi_i)\Big)^2 = \sum_{i=1}^{n} E(\xi_i - E\xi_i)^2 + \sum_{i \ne j} E\big((\xi_i - E\xi_i) \cdot (\xi_j - E\xi_j)\big). \tag{13.31}$$

By Theorem 9.1, the expectations with $i \ne j$ are equal to 0, and the variance of the sum of independent random variables is equal to the sum of their variances. By (13.28), we have:

$$\mathrm{Var}\Big(\frac{\xi_1 + ... + \xi_n}{n}\Big) \le \frac{n \cdot C}{n^2} = \frac{C}{n} \to 0 \qquad (n \to \infty), \tag{13.32}$$

from which convergence in probability follows.