Lecture 14.

Now to strong laws of large numbers.

Theorem 14.1 (E.Borel). *If* $\xi_1, \xi_2, ..., \xi_n, ...$ *is a sequence of independent random variables taking values* $0, 1$ *with probabilities* $1-p$, p (as in Theorem 13.6), *then*

$$\frac{\xi_1 + ... + \xi_n}{n} \to p \quad \text{almost surely} \quad (n \to \infty). \tag{14.1}$$

Proof. Let us apply Theorem 13.5 with $k = 4$. We have:

$$E\Big(\frac{\xi_1 + ... + \xi_n}{n} - p\Big)^4 = \frac{1}{n^4} \sum_{i,\,j,\,k,\,l} E(\xi_i - p)(\xi_j - p)(\xi_k - p)(\xi_l - p). \tag{14.2}$$

The expectations with i, j, k, l all different are equal to 0 because of Theorem 9.1 and $E(\xi_i - p) = 0$. Also are equal to 0 the expectations, say, with $i = j$, $i \neq k \neq l \neq i$; the only ones that are different from 0 are those with $i = j = k = l$ and those with two subscripts coinciding, and two others coinciding with each other, but not with the first ones. The number of terms of the form $E(\xi_i - p)^4$ is equal to n, that of the terms of the form $E(\xi_m - p)^2(\xi_r - p)^2$, $m \neq r$, is $3n(n-1)$, so the sum of expectations in (14.2) is equal to

$$n \cdot E(\xi_i - p)^4 + 3n(n-1) \cdot E(\xi_m - p)^2(\xi_r - p)^2\big|_{m \neq r} = n \cdot m_4 + 3n(n-1) \cdot m_2^2 \tag{14.3}$$

(the central moments are: $b_4 = E(\xi_i - p)^4 = (p - p^2)(1 - p + p^2)$, $b_2 = E(\xi_i - p)^2 = p - p^2$, though it is not important). What is important is that the sum is of order $O(n^2)$, the expectation (14.2) is of order $O(n^{-2})$, and the series (13.17) converges.

For $p = 1/2$ this result admits a number-theoretic interpretation:

In the binary representation of almost all (with respect to the Lebesgue measure) *real numbers the arithmetic mean of the first* n *digits converges to* $1/2$ *as* $n \to \infty$.

Of course, just the same proof is valid for

Theorem 14.2. *Let* $\xi_1, ..., \xi_n, ...$ *be a sequence of independent random variables. If their fourth central moments are bounded by the same constant*: $E(\xi_i - E\xi_i)^4 \leq C = \text{const} < \infty$, *then* (13.26) *holds.*

The only thing we should mention that was not there in the proof of Theorem 14.1 is that by the Schwartz inequality $E(\xi_i - E\xi_i)^2 \leq \sqrt{E(\xi_i - E\xi_i)^4} \leq \sqrt{C}$.

For independent *identically distributed* random variables the following theorem is true:

Kolmogorov's Theorem. *Let* $\xi_1, ..., \xi_n, ...$ *be a sequence of independent identically distributed random variables.*

If the expectation $E\xi_i$ *is finite, then the sequence of arithmetic means* $\dfrac{\xi_1 + ... + \xi_n}{n}$ *converges to it almost surely as* $n \to \infty$.

And conversely: *if this sequence has almost surely a finite limit*, *then the expectation* $E\xi_i$ *is finite.*

Also for *this* theorem we'll not give a proof now (this is hardly surprising: Kolmogorov's Theorem is stronger than Khinchin's, and we are not giving the proof of that one either).

The only thing about this theorem that we can decide now, based on the zero – one law, is: supposing a finite limit $\lim_{n\to\infty} \dfrac{\xi_1 + ... + \xi_n}{n}$ exists almost surely, can this limit not be a constant (almost surely)?

Think about this.

Before we go further, let us introduce expectations of vector-valued random variables.

If $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)$ is an n-dimensional random vector, we define $E\boldsymbol{\xi}$ as the vector $(E\xi_1, ..., E\xi_n) \in \mathbb{R}^n$ supposing the expectations $E\xi_i$ exist and are finite. The only difference with the one-dimensional case is that we don't consider infinite expectations: we haven't introduced (and it is not reasonable to introduce) such a thing as the n-dimensional space extended by adding some "infinite points".

Of course, we have all the usual rules: if $E\boldsymbol{\xi}$ and $E\boldsymbol{\eta}$ exist, then $E(\boldsymbol{\xi}+\boldsymbol{\eta})$ also exists, and

$$E(\boldsymbol{\xi} + \boldsymbol{\eta}) = E\boldsymbol{\xi} + E\boldsymbol{\eta}; \tag{14.4}$$

and if we write $\boldsymbol{\xi}$ as a column vector, and A is an $(m \times n)$-matrix (not a *random* matrix: just a matrix constant), we have:

$$EA\boldsymbol{\xi} = A \cdot E\boldsymbol{\xi} \tag{14.5}$$

(this is an m-dimensional vector).

And $E\boldsymbol{\xi}$ exists if and only if the expectation of the scalar random variable $|\boldsymbol{\xi}|$ is finite (is there anything to prove here?).

Of course, we can now introduce expectations of complex-valued random variables: for $\zeta = \xi + i\eta$ we take, by definition,

$$E\zeta = E\xi + i \cdot E\eta \tag{14.6}$$

if the expectations in the right-hand side are finite.

Quite naturally, $E\zeta$ exists if and only if $E|\zeta| < \infty$; for complex constants c_1, c_2

$$E(c_1 \cdot \zeta_1 + c_2 \cdot \zeta_2) = c_1 \cdot E\zeta_1 + c_2 \cdot E\zeta_2; \tag{14.7}$$

$$E\overline{\zeta} = \overline{E\zeta}; \tag{14.8}$$

$$|E\zeta| \le E|\zeta|; \tag{14.9}$$

and the complex version of Theorem 9.1 is true: *If* ζ_1, ..., ζ_n *are independent complex-valued random variables having expectations*, *then their product also have one*, *and*

$$E(\zeta_1 \cdot ... \cdot \zeta_n) = E\zeta_1 \cdot ... \cdot E\zeta_n. \tag{14.10}$$

We started with random vectors; but we did not have any analog of (14.10) for them. Here it is: *If* Ξ_1, ..., Ξ_n *are independent random matrices having expectations, then the expectation of their product is equal to the product of expectations*:

$$E(\Xi_1 \cdot \ldots \cdot \Xi_n) = E\Xi_1 \cdot \ldots \cdot E\Xi_n \tag{14.11}$$

(of course, if the sizes of the matrices are such that their product makes sense).

There are several similar devices for handling distributions of random variables (and random vectors, too).

If ξ is a discrete random variable taking nonnegative integer values 0, 1, 2, ..., n, ..., its *generating function* (or the generating function of the corresponding distribution, or of its probability mass function, i. e. of the sequence $p(0)$, $p(1)$, $p(2)$, ..., $p(n)$, ...) is, by definition,

$$P(s) = P_\xi(s) = Es^\xi = \sum_{k=0}^{\infty} p(k) \cdot s^k. \tag{14.12}$$

Of course, this expectation is finite at least for $|s| \le 1$.

We can define generating functions also for arbitrary sequences a_k, $k = 0, 1, 2, \ldots, n, \ldots$ by $A(s) = \sum_{k=0}^{\infty} a(k) \cdot s^k$, only it may not converge for all s, $|s| \le 1$. It may even diverge for all s except $s = 0$; of course, in this case the generating function is useless.

The *moment generating function* is defined, for an arbitrary random variable ξ, by

$$\Psi(z) = \Psi_\xi(z) = Ee^{z\xi} = \int_{-\infty}^{\infty} e^{zx}\, \mu_\xi(dx). \tag{14.13}$$

The integral may diverge (for some values of z no finite expectation (14.13) may exist); but it always converges for $z = 0$, and $\Psi(0) = 1$. Of course, if the integral (14.13) converges only for $z = 0$, the function $\Psi(z)$ is quite useless.

Why is this function called the moment generating function? We know that $e^{z\xi} = \sum_{k=0}^{\infty} \dfrac{z^k \xi^k}{k!}$; if this series admits term-by-term integration:

$$\Psi(z) = \sum_{k=0}^{\infty} \frac{E\xi^k}{k!}\, z^k, \tag{14.14}$$

we see that this function is the generating function – not precisely of the sequence of moments $m_k = E\xi^k$ – but of the sequence $m^k/k!$.

If ξ is a nonnegative integer-valued random variable, we have:

$$\Psi_\xi(z) = P_\xi(e^z): \tag{14.15}$$

this is how the moment-generating function is related to the generating function.

But the generating function can be used only for a narrow class of nonnegative integer-valued random variables; the moment-generating function only for such random variables for which the integral (14.13) converges; we would like to have a similar device that can be used for *all* random variables. Such is the *characteristic function.*

The characteristic function of a random variable ξ (or that of its distribution μ) is, by definition, the complex-valued function

$$f(t) = f_\xi(t) = E e^{it\xi} = f_\mu(t) = \int_{-\infty}^{\infty} e^{itx}\,\mu(dx). \qquad (14.16)$$

The expectation (14.16) is defined (the integral converges) at least for all real values of t, because $E|e^{it\xi}| = E1 = 1 < \infty$.

It is clear that the characteristic function, for real values of its argument t, is just the moment-generating function considered for complex values of its argument z, namely, on the imaginary axis:

$$f(t) = \Psi(it). \qquad (14.17)$$

For discrete random variables the characteristic function can be written as

$$f(t) = \sum_k p(x_k) \cdot e^{ix_k t}; \qquad (14.18)$$

for integer-valued ones,

$$f(t) = \sum_{k=-\infty}^{\infty} p(k) \cdot e^{ikt}. \qquad (14.19)$$

This is a *Fourier series* in the complex form, with Fourier coefficients $p(k)$; it is a periodic function with period 2π.

Example: if ξ has the Poisson distribution with parameter a, we have:

$$f(t) = \sum_{k=0}^{\infty} \frac{a^k e^{-a}}{k!} \cdot e^{ikt} = e^{-a} \cdot \sum_{k=0}^{\infty} \frac{(a \cdot e^{ikt})^k}{k!} = e^{-a} \cdot e^{a\, e^{ikt}} = e^{a(e^{ikt}-1)}. \qquad (14.20)$$

If a random variable takes values that are multiples of some $h > 0$, its characteristic function also is represented by a Fourier series:

$$f(t) = \sum_{k=-\infty}^{\infty} p(kh) \cdot e^{ikht}. \qquad (14.21)$$

Such a function is periodic with period $T = 2\pi/h$; and knowing it, we can find the values of the probability mass function, which are the Fourier coefficients:

$$p(kh) = \frac{1}{2\pi/h} \int_{-\pi/h}^{\pi/h} e^{-ikht} f(t)\; dt. \qquad (14.22)$$

Another example – for a non-discrete random variable: Let ξ have the normal distribution with parameters $(0,\ b)$. We have:

$$f(t) = Ee^{it\xi} = E\sum_{k=0}^{\infty}\frac{(it\xi)^k}{k!} = \int_{\Omega}\sum_{k=0}^{\infty}\frac{i^k\cdot t^k\cdot\xi^k}{k!}\ dP. \tag{14.23}$$

By Problem $\boxed{\mathbf{32}}$ we know what the moments $m_k = E\xi^k$ are; so it seems that we can evaluate the integral (14.23) as the sum of an infinite series containing these moments. But, in general, the integral of the sum of an infinite series of functions does not have to be equal to the sum of the integrals: this is a question of whether limit passage under the integral sign is possible.

We have spoken about this in Lecture 4 (about the Lebesgue integral), but only for monotone limit passages (or series consisting of nonnegative functions, see formula (4.40)), which was closely related to the definition of Lebesgue integral. The time has come to formulate a couple of other theorems about limit passage. You are supposed to know these theorems; but anyway I'll give their precise formulation in the form in which I am planning to use them:

Theorem 14.3 (dominated-convergence theorem). *Let $(X,\ \mathcal{X})$ be a measurable space, and m a measure on it; let $f_1(x)$, $f_2(x)$, ..., $f_n(x)$, ... be a sequence of measurable functions such that $\lim_{n\to\infty} f_n(x)$ exists almost everywhere. If all functions $f_k(x)$ are dominated by one and the same integrable function g:*

$$|f_k(x)| \le g(x) \quad \text{for all } k \text{ and } x, \qquad \int_X g(x)\ m(dx) < \infty, \tag{14.24}$$

(the functions $f_n(x)$ may be extended-real, or complex, or vector-valued), *then*

$$\int_X \lim_{n\to\infty} f_n(x)\ m(dx) = \lim_{n\to\infty}\int_X f_n(x)\ m(dx) \tag{14.25}$$

(there may be a set of zero m-measure on which the function $\lim_{n\to\infty} f_n(x)$ is not defined; in this case we define it there arbitrarily so that the function to integrate is defined everywhere: it does not affect the value of the integral).

The same in the language of infinite series:

Theorem 14.4. *Let $(X,\ \mathcal{X})$ be a measurable space, and m a measure on it; let $f_1(x)$, $f_2(x)$, ..., $f_n(x)$, ... be a sequence of measurable functions such that*

$$\sum_{k=1}^{\infty}\int_X |f_k(x)|\ m(dx) < \infty. \tag{14.26}$$

Then the series $\sum_{k=1}^{\infty} f_k(x)$ converges almost everywhere, and

$$\int_X \sum_{k=1}^{\infty} f_k(x)\ m(dx) = \sum_{k=1}^{\infty}\int_X f_k(x)\ m(dx). \tag{14.27}$$

Theorem 14.5 (Fatou's Lemma). *Let $(X, \mathcal{X})$ be a measurable space, and m a measure on it; let $f_1(x)$, $f_2(x)$, ..., $f_n(x)$, ... be a sequence of nonnegative measurable functions. Then*

$$\int_X \underline{\lim}_{n\to\infty} f_n(x)\, m(dx) \le \underline{\lim}_{n\to\infty} \int_X f_n(x)\, m(dx). \tag{14.28}$$

It is difficult to remember whether it is $\le$, or $\ge$; and whether it is the lower limits, or the upper ones. To remember this, I consider some examples: first, a sequence of functions with integrals equal, say to 1, but with a limit that is zero almost everywhere, e. g., $X = (0, 1)$, m the Lebesgue measure λ_1, $f_k(x) = (k + 1)x^k$: so the sign must be $\le$. And a second example with $f_{2m-1}(x) = 1$ in the left half of the interval from 0 to 1, and $= 0$ in its right half, and $f_{2m}(x) = 1 - f_{2m-1}(x)$: we have $\int_0^1 f_k(x)\, dx = 1/2$, $\underline{\lim}_{n\to\infty} f_n(x) \equiv 0$, $\overline{\lim}_{n\to\infty} f_n(x) \equiv 1$: so it has to be lower limits $\underline{\lim}$.

We won't need Fatou's Lemma right now; but we are going to use Theorem 14.4.

We have:

$$E\xi^{2m} = 1 \cdot 3 \cdot \ldots \cdot (2m - 1)\, b^m; \tag{14.29}$$

and by Schwarz's inequality

$$E|\xi|^{2m+1} \le \sqrt{E\xi^{2m} \cdot E\xi^{2m+2}} = 1 \cdot 3 \cdot \ldots \cdot (2m - 1) \cdot \sqrt{2m + 1} \cdot b^{m+1/2}. \tag{14.30}$$

The series

$$\sum_{k=0}^{\infty} \frac{E|\xi t|^k}{k!} \tag{14.31}$$

is divided into two:

$$\begin{aligned} \sum_{m=0}^{\infty} \frac{E(\xi t)^{2m}}{(2m)!} &= \sum_{m=0}^{\infty} \frac{1 \cdot 3 \cdot \ldots \cdot (2m-1) \cdot b^m t^{2m}}{(2m)!} \\ &= \sum_{m=0}^{\infty} \frac{(bt^2)^m}{2 \cdot 4 \cdot \ldots \cdot (2m)} = \sum_{m=0}^{\infty} \frac{(bt^2/2)^m}{m!} = e^{bt^2/2} < \infty, \end{aligned} \tag{14.32}$$

and

$$\begin{aligned} \sum_{m=0}^{\infty} \frac{E|\xi t|^{2m+1}}{(2m+1)!} &\le \sum_{m=0}^{\infty} \frac{1 \cdot 3 \cdot \ldots \cdot (2m-1) \cdot \sqrt{2m+1} \cdot b^{m+1/2} |t|^{2m+1}}{(2m+1)!} \\ &\le \sum_{m=0}^{\infty} \frac{1 \cdot 3 \cdot \ldots \cdot (2m-1) \cdot b^{m+1/2} |t|^{2m+1}}{(2m)!} = b^{1/2}|t| \cdot e^{bt^2/2} < \infty; \end{aligned} \tag{14.33}$$

so by Theorem 14.4 we have:

$$f(t) = Ee^{it\xi} = \sum_{k=0}^{\infty} \frac{i^k t^k E\xi^k}{k!} = \sum_{m=0}^{\infty} \frac{i^{2m} t^{2m} \cdot (2m-1)!! \cdot b^m}{(2m)!} = \sum_{m=0}^{\infty} \frac{(-bt^2/2)^m}{m!} = e^{-bt^2/2}. \tag{14.34}$$