Lecture 4. Lebesgue integral.

My students claim that they know everything about Lebesgue integral; but I am a little doubtful.

Let me remind you what Lebesgue integral with respect to an arbitrary measure is.

It will be convenient for us to consider both functions to be integrated and their integrals taking values in the *extended* real line $[-\infty, \infty]$, which is defined as the real line to which two extra points ∞ and $-\infty$ are added: $[-\infty, \infty] = \mathbb{R}^1 \cup \{\infty, -\infty\}$. In the real line we have many things defined: the order (inequalities), arithmetic operations, limits, the Borel σ-algebra $\mathcal{B}_{(-\infty, \infty)} = \mathcal{B}^1$. Let us define the corresponding things for the extended real line.

By definition, we have $-\infty < x < \infty$ for every $x \in \mathbb{R}^1$. So the order is defined. Of course, we can consider intervals of various kinds: for $-\infty \le a \le b \le \infty$

$$[a, b] = \{x \in [-\infty, \infty] \colon a \le x \le b\}, \tag{4.1}$$

$$[a, b) = \{x \in [-\infty, \infty] \colon a \le x < b\}, \tag{4.2}$$

etc. (and, say, $[-10, \infty] \ne [-10, \infty)$, in contrast with intervals in $\mathbb{R}^1$ – see Lecture 3). A *neighborhood* of a point $a \in \mathbb{R}^1$ in the extended real line is an arbitrary open interval containing this point, but a neighborhood of the point ∞ is an arbitrary interval $(a, \infty]$, $a < \infty$, and that of $-\infty$ is $[-\infty, b)$ (just as in a closed interval $[a, b]$ in the real line $\mathbb{R}^1$ we take $(c, b]$ as a neighborhood of its right end b).

If we have neighborhoods, we define *limits*; $\lim_{n\to\infty} a_n$, $a_n \in [-\infty, \infty]$, is, obviously, the same thing as the limit that we are accustomed to, except now we can have, say, sequences for which $a_n = \infty$; it is clear that $\lim_{n\to\infty} \infty = \infty$.

Now to the arithmetic operations: for numbers we leave them as they were, and we need to define these operations only if at least one of the operands is $\pm\infty$.

Operations with one operand: Quite naturally we take, by definition, that $|\pm\infty| = \infty$; and we take that $-(\pm\infty) = \mp\infty$.

Addition: for x being a *number*, we take

$$\infty + x = x + \infty = \infty, \quad \infty + \infty = \infty, \quad -\infty + x = x + (-\infty) = -\infty, \quad -\infty + (-\infty) = -\infty, \tag{4.3}$$

and

$$-\infty + \infty, \quad \infty - \infty \qquad \text{make no sense.} \tag{4.4}$$

These things are very natural, we knew them – but for infinite *limits.*

Subtraction: by definition, $a - b = a + (-b)$ for arbitrary $a, b \in [-\infty, \infty]$.

Multiplication:

$$x \cdot (\pm\infty) = \pm\infty \cdot x = \begin{cases} \pm\infty & \text{for } x > 0, \\ \mp\infty & \text{for } x < 0, \\ 0 & \text{for } x = 0 \end{cases} \tag{4.5}$$

(in particular, $-\infty \times \infty = -\infty$). The only thing that one wouldn't expect here is $\pm\infty \cdot 0 = 0$ (it is not necessarily so for limits); but this convention turns out useful.

The standard σ-algebra $\mathcal{B}_{[-\infty,\,\infty]}$ in the extended real line can be defined as one generated by open subsets, or one generated by intervals, or only by intervals with infinite left end, etc.:

$$\mathcal{B}_{[-\infty,\,\infty]} = \sigma\{\text{open subsets of } [-\infty,\,\infty]\} = \sigma\{[-\infty,\,a]\colon\ a \in \mathbb{R}^1\}, \tag{4.6}$$

or as the set of all Borel subsets of $(-\infty,\,\infty)$, plus the same sets with one or both of the points ∞, $-\infty$ added:

$$\mathcal{B}_{[-\infty,\,\infty]} = \mathcal{B}^1 \cup \bigl\{A \cup \{\infty\}\colon A \in \mathcal{B}^1\bigr\} \cup \bigl\{A \cup \{-\infty\}\colon A \in \mathcal{B}^1\bigr\} \cup \bigl\{A \cup \{\infty,\,-\infty\}\colon A \in \mathcal{B}^1\bigr\}. \tag{4.7}$$

So if we don't refer to any σ-algebra in $[-\infty,\,\infty]$, we mean $\mathcal{B}_{[-\infty,\,\infty]}$.

Now suppose $(X,\,\mathcal{X})$ is a measurable space, and m a measure on it (that is, rather on the σ-algebra $\mathcal{X}$). Let f be an $\mathcal{X}$-measurable function on X with values in the extended real line. How is the integral

$$\int_X f(x)\ m(dx) = \int_X f\ dm \tag{4.8}$$

defined?

The definition is done in several steps.

The first step: Let f be a measurable *simple* function; i. e., a measurable (of course, $\mathcal{X}$-measurable) function taking finitely many values: a linear combination of indicators of disjoint $\mathcal{X}$-sets with union $\bigcup_{i=1}^{n} A_i = X$:

$$f(x) = \sum_{i=1}^{n} c_i \cdot I_{A_i}(x), \qquad A_i \in \mathcal{X}, \tag{4.9}$$

(the indicator function I_A of a subset $A \subset X$ is defined by

$$I_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \in X \setminus A; \end{cases} \tag{4.10}$$

I am mentioning this because in different books this mathematical object is called and denoted differently – e. g., in Kolmogorov & Fomin's book it is called *characteristic function*, and denoted χ_A. But in probability theory the term "characteristic function" is used with quite a different meaning).

So the function $f(x)$ takes the value c_i on the set A_i. (See how the convention about $\infty \cdot 0$ being equal to 0 is useful here: otherwise we would get something absurd in the case of one of c_i being equal to ∞.)

We take, by definition,

$$\int_X f\ dm = \sum_{i=1}^{n} c_i \cdot m(A_i). \tag{4.11}$$

The sum (4.11) makes no sense if the function f takes values $+\infty$ and $-\infty$ on sets of positive m-measure; or, say, if $c_{i_1} = \infty$, $m(A_{i_1}) > 0$, $c_{i_2} < 0$, $m(A_{i_2}) = \infty$. So, the integral may be not defined for some measurable functions; we were prepared to this.

In Kolmogorov & Fomin's book a simple function is defined differently: as a *countable* linear combination of indicators. There are different approaches to constructing the Lebesgue integral; what matters is that they all lead in the end to the same concept of integral.

An unpleasantness may arise in this definition: the same function f can be represented in the form (4.9) in different ways. Say, the equality (4.9) holds for all x, and also

$$f(x) = \sum_{j=1}^{m} d_j \cdot I_{B_j}(x), \qquad B_j \in \mathcal{X}, \ \ B_j \cap B_k = \emptyset \ (j \neq k), \ \bigcup_{j=1}^{m} B_j = X. \tag{4.12}$$

Should we define the integral (4.8) by formula (4.11), or take

$$\int_X f \, dm = \sum_{j=1}^{m} d_j \cdot m(B_j)\,? \tag{4.13}$$

It is easily proved that if both (4.9) and (4.12) hold, the expressions (4.11) and (4.13) coincide (or both make no sense).

Normally, I am not going to give any indications of how the results belonging to measure theory – or to the set-theoretic introduction to it – are proved; but in this case I'll say something about it: this is so simple and natural. We consider the disjoint sets $A_i \cap B_j$, $i = 1, \ldots, n$, $j = 1, \ldots, m$; if the set $A_i \cap B_j$ is not empty, the function f takes on it the value $e_{ij} = c_i = d_j$. Let us consider the sum

$$\sum_{i,\,j} e_{ij} \cdot m(A_i \cap B_j) \tag{4.14}$$

(by our convention, the sum in which the range of the variable of summation is not shown is understood as being over all possible values of this variable: i from 1 to n, and j from 1 to m). If we take together the terms with the same i in (4.14), we get, using finite additivity of m, that this expression is equal to

$$\sum_{i=1}^{n} c_i \cdot \sum_{j=1}^{m} m(A_i \cap B_j) = \sum_{i=1}^{n} c_i \cdot m\Big(\bigcup_{j=1}^{m} (A_i \cap B_j)\Big), \tag{4.15}$$

i. e., to the expression (4.11). Collection together all terms with the same j, we get that the expression (4.14) is equal to (4.13) – which completes the proof.

It is very easy to check that if f and g both are measurable simple functions,

$$\int_X (f+g)\, dm = \int_X f \, dm + \int_X g \, dm \tag{4.16}$$

if both integrals $\int_X f \, dm$, $\int_X g \, dm$ make sense, unless one of them is equal to $+\infty$ and the other to $-\infty$; and that if $f(x) \leq g(x)$ for all x, then $\int_X f \, dm \leq \int_X g \, dm$ – if both integrals make sense.

The next step: we define the integral for all nonnegative measurable functions.

Suppose $f(x)$ is such a function, and

$$0 \le f_1(x) \le f_2(x) \le ... \le f_n(x) \le ... \tag{4.17}$$

(the inequalities are satisfied for all $x \in X$) is a sequence of nonnegative *simple* measurable functions such that

$$f(x) = \lim_{n\to\infty} f_n(x), \qquad x \in X. \tag{4.18}$$

The integrals $\displaystyle\int_X f_n(x)\, m(dx)$ *are* defined: no $\infty - \infty$ for *nonnegative* simple functions; and according to the above, they form a nondecreasing sequence. This sequence necessarily has a limit (finite or infinite); and we take

$$\int_X f\ dm = \lim_{n\to\infty} \int_X f_n\ dm. \tag{4.19}$$

Again an unpleasantness arises: the same measurable function may be represented in the form (4.18) using another nondecreasing sequence of simple functions:

$$f(x) = \lim_{n\to\infty} g_n(x), \tag{4.20}$$

where

$$0 \le g_1(x) \le g_2(x) \le ... \le g_n(x) \le ...; \tag{4.21}$$

should we define the integral as the limit (4.19), or as

$$\int_X f\ dm = \lim_{n\to\infty} \int_X g_n\ dm? \tag{4.22}$$

Using *countable* additivity of m, one proves that the limits (4.19) and (4.22) necessarily coincide.

Another unpleasantness that we may envisage here is that there may be no nondecreasing sequence (4.17) of simple functions with f as its limit (instead of too many different sequences, too few of them).

But in fact, there are always such sequences; one example:

$$f_n(x) = \sum_{i=0}^{n\cdot 2^n - 1} \frac{i}{2^n} \cdot I_{\{x:\, i/2^n \le f(x) < (i+1)/2^n\}}(x) + 0 \cdot I_{\{x:\, f(x) \ge n\}}(x). \tag{4.23}$$

Of course these functions are simple and nonnegative; drawing a picture, say, of $f_1(x)$ and $f_2(x)$ shows us that the sequence is nondecreasing; and the sets $\{x : i/2^n \le f(x) < (i+1)/2^n\}$, $\{x\colon f(x) \ge n\}$ clearly belong to $\mathcal{X}$.

So for every nonnegative measurable function its Lebesgue integral is defined.

It is easily proved that (4.16) holds for non-simple nonnegative measurable functions.

Finally, we define Lebesgue integral with respect to the measure m for measurable functions that take values of both signs.

If a measurable function $f(x)$ can be represented as the difference of two nonnegative measurable functions $g(x)$ and $h(x)$ with integrals at least one of which is not equal to $+\infty$:

$$f(x) = g(x) - h(x), \qquad \int_X g\, dm < \infty \quad \text{or} \quad \int_X h\, dm < \infty, \tag{4.24}$$

we take by definition

$$\int_X f\, dm = \int_X g\, dm - \int_X h\, dm. \tag{4.25}$$

If $f(x)$ cannot be represented as such a difference, we say that the integral $\int_X f\, dm$ does not exist.

Again it is possible that f is represented as such a difference in some other way:

$$f(x) = G(x) - H(x), \qquad \int_X G\, dm < \infty \quad \text{or} \quad \int_X H\, dm < \infty. \tag{4.26}$$

Then, should we take the value of the integral defined by (4.25), or

$$\int_X f\, dm = \int_X G\, dm - \int_X H\, dm? \tag{4.27}$$

If we prove that the right-hand side of (4.25) is necessarily equal to that of (4.27), our definition is OK. For this we use the equality (4.16) for arbitrary nonnegative measurable functions.

From (4.24) and (4.26) we have, for all $x \in X$:

$$g(x) + H(x) = h(x) + G(x); \tag{4.28}$$

for these nonnegative functions we have:

$$\int_X g\, dm + \int_X H\, dm = \int_X h\, dm + \int_X G\, dm. \tag{4.29}$$

If this expression is finite, we just subtract the number $\int_X h\, dm + \int_X H\, dm$ from both sides, and get

$$\int_X g\, dm - \int_X h\, dm = \int_X G\, dm - \int_X H\, dm. \tag{4.30}$$

If the expression (4.29) is equal to $+\infty$, then either $\int_X g\, dm$, or $\int_X H\, dm$ must be equal to $+\infty$ – but not both of them: by our conditions (4.24), (4.26) at least two of $\int_X g\, dm$, $\int_X h\, dm$, $\int_X G\, dm$, $\int_X H\, dm$ must be finite. If $\int_X g\, dm = \infty$, then $\int_X H\, dm$ and

$\int_X h\,dm$ must be finite, $\int_X G\,dm = \infty$, and so both sides of (4.30) are equal to $+\infty$. If it is $\int_X H\,dm$ that is infinite, we obtain in the same way that both sides of (4.30) are equal to $-\infty$, and this equality still holds.

Can an arbitrary measurable function f be represented as $f(x) = g(x) - h(x)$, where nonnegative g and h do not necessarily satisfy the condition (4.24)? – Yes, we can take

$$g(x) = \big(f(x)\big)_+, \qquad h(x) = \big(f(x)\big)_-, \tag{4.31}$$

where the functions y_+, y_- of extended-real argument y are defined by

$$y_+ = \begin{cases} y, & y \ge 0, \\ 0, & y < 0, \end{cases} \qquad y_- = \begin{cases} 0, & y \ge 0, \\ -y, & y < 0. \end{cases} \tag{4.32}$$

These functions are continuous and so Borel-measurable, and so the functions (4.31) are $\mathcal{X}$-measurable.

The functions (4.31) are the smallest possible of all nonnegative functions h, g such that $f = g - h$; so the integral $\int_X f\,dm$ exists if and only if at least one of the integrals $\int_X (f)_+\,dm$, $\int_X (f)_-\,dm$ is finite.

We call a measurable function f *integrable* (with respect to the measure m) if its integral is finite. It is easy to see that a measurable function f is integrable if and only if

$$\int_X |f(x)|\; m(dx) < \infty. \tag{4.33}$$

So, in contrast with the (improper) Riemann integral, a function cannot be Legesgue-integrable without being *absolutely* integrable.

The integral thus defined has all the usual good properties, such as

$$\int_X c \cdot f(x)\; m(dx) = c \cdot \int_X f(x)\; m(dx) \tag{4.34}$$

for every constant $c \in \mathbb{R}^1$; if the integrals of f_1, f_2 make sense,

$$\int_X (f_1 + f_2)\,dm = \int_X f_1\,dm + \int_X f_2\,dm \tag{4.35}$$

unless it is $-\infty + \infty$ or $\infty - \infty$; also, even if it is not the "usual" good properties,

$$\int_X f\,d(c \cdot m) = c \cdot \int_X f\,dm \tag{4.36}$$

for $c \in [0, \infty)$,

$$\int_X f\,d(m_1 + m_2) = \int_X f\,dm_1 + \int_X f\,dm_2 \tag{4.37}$$

(unless...); and if $0 \le f_1(x) \le f_2(x) \le ... \le f_n(x) \le ...$ is a nondecreasing sequence of nonnegative measurable functions, then

$$\int_X \lim_{n\to\infty} f_n(x)\ m(dx) = \lim_{n\to\infty} \int_X f_n(x)\ m(dx). \tag{4.38}$$

This follows – not immediately, but quite naturally – from the way the Lebesgue integral is defined for nonnegative functions by formula (4.19).

By the way, do we need here to require in addition that the function $\lim_{n\to\infty} f_n(x)$ should be measurable? It turns out that for any, not necessarily nondecreasing, sequence of measurable functions $f_n(x)$ the set

$$\{x\colon \lim_{n\to\infty} f_n(x) \text{ exists}\} \tag{4.39}$$

belongs to the σ-algebra $\mathcal{X}$; and the function $\lim_{n\to\infty} f_n(x)$ is $\mathcal{X}$-measurable on this set. In our case of a nondecreasing sequence the limit does exist for every x, the set (4.39) is the whole X, and the limit is automatically measurable.

The same theorem about monotone limit passage can be reformulated as follows: if $g_1(x)$, $g_2(x)$, ..., $g_n(x)$, ... is a sequence of nonnegative measurable functions, then

$$\int_X \sum_{i=1}^{\infty} g_i(x)\ m(dx) = \sum_{i=1}^{\infty} \int_X g_i(x)\ m(dx). \tag{4.40}$$

Here we take $f_n(x) = \sum_{i=1}^{n} g_i(x)$ (of course, we have to prove that this finite sum is $\mathcal{X}$-measurable).

For sequences that are not necessarily nondecreasing, (4.38) may not hold; there are several theorems about *limit passage under the integral sign*; I don't want to speak about this now.

Concrete case: For the integral with respect to Lebesgue measure λ_n in the space $(\mathbb{R}^n, \mathcal{B}^n)$ the notation

$$\int_{\mathbb{R}^n} f(x)\ \lambda_n(dx) = \int_{\mathbb{R}^n} f(x)\ dx = \underset{\mathbb{R}^n}{\int \cdots \int} f(x_1, ..., x_n)\ dx_1 ... dx_n \tag{4.41}$$

is used. If the Riemann integral $\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, ..., x_n)\ dx_1 ... dx_n$ converges absolutely, then this integral coincides with the corresponding Lebesgue integral (4.41).

Another concrete class of examples: If X is an arbitrary set, the *counting measure* $\#$ is defined on the σ-algebra $\mathcal{P}(X)$ of all subsets of X by

$$\#(A) = \text{number of points } x \in A \tag{4.42}$$

($\#(A) = \infty$ if the set A is infinite). The Lebesgue integral $\int_X f(x)\ \#(dx)$ is nothing but the sum

$$\sum_{x \in X} f(x), \tag{4.43}$$

with the necessary stipulations for the case of an uncountable sum. For simplicity, let us restrict ourselves to nonnegative integrands f. The integral is taken to be equal to $+\infty$ if there are uncountably many points x with $f(x) > 0$; otherwise: zero summands in (4.43) do not count; if the sum without zeros is finite, it is just the finite sum, and if it is countably infinite, it is understood as the sum of an infinite series (this sum *does not depend on the order of summation* because the summands are all of the same sign: an elementary theorem from Analysis).

Now, how is the integral $\int_A f\,dm$ over some subset $A \subset X$, and not over the whole X, defined? This simplest way to do so is to take by definition, for $A \in \mathcal{X}$,

$$\int_A f(x)\ m(dx) = \int_X I_A(x)\ m(dx). \tag{4.44}$$

This definition leads to the same result as if we consider simple functions *on* A, then arbitrary nonnegative measurable functions on A, etc.

If the set function n is defined by formula (3.27) for all $C \in \mathcal{X}$, it is necessarily countably additive: for disjoint A_1, A_2, ..., A_n, ...

$$\begin{aligned} n\Big(\bigcup_{i=1}^{\infty} A_i\Big) &= \int\limits_{\bigcup_{i=1}^{\infty} A_i} f(x)\ m(dx) = \int_X I_{\bigcup_{i=1}^{\infty} A_i}(x) \cdot f(x)\ m(dx) \\ &= \int_X \sum_{i=1}^{\infty} I_{A_i}(x) \cdot f(x)\ m(dx) = \sum_{i=1}^{\infty} \int_X I_{A_i}(x) \cdot f(x)\ m(dx) = \sum_{i=1}^{\infty} n(A_i) \end{aligned} \tag{4.45}$$

(proved above for *nonnegative* integrands; so something remains here to be proved).

So a set function n having a density with respect to a measure m is necessarily countably additive. (In Kolmogorov & Fomin's book countably additive set functions are called *charges* – because such functions form a mathematical model for electric charges, $n(A)$ being the total charge, positive or negative, carried by the region A).

A density $f(x)$ of n with respect to m is, generally, not unique: if we change the function $f(x)$ arbitrarily on a non-empty set A having zero m-measure (such sets may not exist for a measure m; they certainly do exist for the Lebesgue measure), the integrals $\int_C f\ dm$ do not change, and the changed f is still a version of density.

Theorem 4.1. *Let the measure* m *on* $(X, \mathcal{X})$ *be* σ*-finite* (which means that there exists an infinite sequence of sets $B_i \in \mathcal{X}$ such that $\bigcup_{i=1}^{\infty} B_i = X$, and $m(B_i) < \infty$). *Then the density of a countably additive set function* n *with respect to* m *is* almost *unique* (of course supposing that it exists); *i. e., if the measurable functions* f_1, f_2 *are two versions of this density*:

$$\int_C f_1(x)\ m(dx) = \int_C f_2(x)\ m(dx) = n(C) \tag{4.46}$$

for every $C \in \mathcal{X}$, *then*

$$f_1(x) = f_2(x) \qquad \text{almost everywhere,} \tag{4.47}$$

that is,

$$m\{x\colon\ f_1(x) \neq f_2(x)\} = 0. \tag{4.48}$$

The notations for (every version of) the density of n with respect to m will be

$$f(x) = \frac{n(dx)}{m(dx)} = \frac{dn}{dm}(x). \tag{4.49}$$

Why such notations are used, we'll discuss later.

It is clear that if n has a desity with respect to m, then for $C \in \mathcal{X}$

$$m(C) = 0 \ \Rightarrow\ n(C) = 0. \tag{4.50}$$

A countably additive set function n is called *absolutely continuous* with respect to m if this implication is satisfied. The notation for absolute continuity is

$$n \ll m. \tag{4.51}$$

The following theorem (a real big one, not so easy to prove) is proved in measure theory:

Theorem 4.2 (Radon – Nikodym's Theorem: see, e. g., Kolmogorov & Fomin's book, Theorem 2 of Section 34). *Let a measure m on $(X, \mathcal{X})$ be σ-finite, and let n be a finite countably additive set function on $(X, \mathcal{X})$ that is absolutely continuous with respect to m. Then n has a density $f(x)$ with respect to m.*

Note that the Lebesgue measure λ_n is σ-finite ($B_i = \{x\colon\ |x| \leq i\}$); so a density with respect to the Lebesgue measure $\dfrac{\mu_\xi(dx)}{\lambda_n(dx)} = \dfrac{\mu_\xi(dx)}{dx}$ exists if and only if μ_ξ is absolutely continuous with respect to the Lebesgue measure. (Usually, for shortness, we speak of just *continuous* distributions. Random variables having such distributions are called continuous random variables.)

It turns out that the description of discrete distributions by means of their "probability mass functions" $p_\xi(x) = P\{\xi = x\}$ is also a description involving the density – not with respect to the Lebesgue measure, but with respect to the counting measure #:

$$p_\xi(x) = \frac{\mu_\xi(dx)}{\#(dx)}. \tag{4.52}$$

Indeed, this means that for every set $C \in \mathcal{X}$

$$P\{\xi \in C\} = \int_C p(x)\ \#(dx). \tag{4.53}$$

By (4.43), this is the same as

$$P\{\xi \in C\} = \sum_{x \in C} p(x), \tag{4.54}$$

and this is just the formula (3.23).

This is a reason for treating probability mass functions of discrete distributions and probability densities of (absolutely) continuous ones in parallel ways (e. g., the statistical maximum-likelihood estimates are defined by the same formulas in the discrete and in the continuous case).

Note that since the counting measure $\#$ is not, generally, σ-finite, the above results about existence and (almost) uniqueness of the density cannot be applied. As for uniqueness, it still is there, and even without "almost" (because there are no non-empty sets with $\#(A) = 0$); but it does not follow from $\mu_\xi(\emptyset) = 0$ that this distribution is a discrete one. (Nothing at all can follow from $\mu_\xi(\emptyset) = 0$, because this a general property of all measures.)

Sometimes it is reasonable to consider densities of probability distributions not with respect to the Lebesgue measure, or to the counting measure $\#$, but with respect to some other measures. For example, if we consider random variables taking values in an infinite-dimensional space: in such spaces no infinite-dimensional Lebesgue measure can be defined. It is worth while to consider densities $\dfrac{d\mu_\eta}{d\mu_\xi}$ of one distributions with respect to the others.