Lecture 6.

It seems to me now that better I should formulate Theorem 5.5 a little differently:

Theorem 5.5′. *Let* $F(x)$, $-\infty < x < \infty$, *be a function satisfying the conditions* (5.15), (5.18) *and having finite limits* $F(-\infty) = \lim_{x\to-\infty} F(x)$, $F(\infty) = \lim_{x\to\infty} F(x)$.

Then there exists a finite measure μ *on the real line* (on its Borel σ-algebra $\mathcal{B}^1$, I mean) *such that for* $-\infty \le a \le b \le \infty$ *we have*: $\mu(a, b] = F(b) - F(a)$.

(So the total value of the measure μ is $\mu(\mathbb{R}^1) = F(\infty) - F(-\infty)$; if $F(-\infty) = 0$, $F(\infty) = 1$, the measure μ is a probability measure: the distribution of some random variable.)

Another variant of the same theorem, with interval $(A, B]$ instead of $(-\infty, \infty)$:

Theorem 5.5″. *Let* $F(x)$, $A \le x \le B$, *be a function satisfying the conditions* (5.15), (5.18) *and having finite limits* $F(A) = F(A^+) = \lim_{x\to A^+} F(x)$, $F(B^-) = \lim_{x\to B^-} F(x)$.

Then there exists a finite measure μ *on the interval* $(A, B]$ (on its Borel σ-algebra $\mathcal{B}(A, B]$, I mean) *such that for* $A \le a \le b \le B$ *we have*: $\mu(a, b] = F(b) - F(a)$.

This theorem can be applied to proving that the Lebesgue measure λ_1 exists: we take $F(x) \equiv x$, and get from Theorem 5.5″ the existence of a finite Lebesgue measure on every finite interval $(A, B]$; and then, uniting countably many finite intervals, we get the existence of the Lebesgue measure on the whole real line.

The proof of both Theorem 5.4 and 5.5′ is based on a theorem belonging to measure theory that I will formulate presently.

The theorem is that a finite measure defined on an *algebra* can be extended as a measure onto the σ-algebra generated by it; and this in a unique way.

However better I formulate another theorem (that is equivalent, in fact, to the statement I just formulated), which will be more convenient to use in the proofs of our theorems.

Before I do this, let me introduce a new class of classes of sets in addition to algebras and σ-algebras: *semi-algebras*.

Let X be a space. A class $\mathcal{A}$ of its subsets will be called a semi-algebra (in X) if

1) $X \in \mathcal{A}$ (the same as with *algebras*);

2) the intersection of two sets belonging to $\mathcal{A}$ must belong to $\mathcal{A}$: $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$ (in the definition of an algebra we required that $A \cup B \in \mathcal{A}$; this requirement could be replaced with $A \cap B \in \mathcal{A}$, leading to an equivalent definition);

3) if a set $A \in \mathcal{A}$, its complement $A^c = X \setminus A$ can be represented as a finite union of disjoint sets belonging to $\mathcal{A}$: $A^c = \bigcup_{i=1}^n B_i$, $B_i \cap B_j = \emptyset$ for $i \ne j$ (this is a weakening of the algebra requirement with every set its complement also belongs to the algebra).

Clearly, every algebra (and also every σ-algebra) is a semi-algebra. Also clearly the empty set belongs to every semi-algebra (the complement $X^c = \emptyset$ must be representable as a finite union of sets belonging to $\mathcal{A}$; and all summands in this union must be empty).

An example of a semi-algebra (not being an algebra): the class $\mathcal{A}$ of all semi-open intervals $(a,\, b]$ in the real line, $-\infty \le a \le b \le \infty$.

Indeed, $X = \mathbb{R}^1$ belongs to $\mathcal{A}$: $\mathbb{R}^1 = (-\infty,\, \infty]$ (by definition, $(-\infty,\, \infty] = \{x \in \mathbb{R}^1 : -\infty < x \le \infty\} = (-\infty,\, \infty) = \mathbb{R}^1$). (The empty set can also be considered as a semi-open interval: $\emptyset = (a,\, a]$, which is $\{x \in \mathbb{R}^1 :\ a < x \le a\}$: clearly empty.)

The intersection of two semi-open intervals is obviously either an interval of the same kind (better make a picture and look at it), or empty.

Finally, the complement of a finite interval is either the union of two disjoint semi-infinite intervals ($n = 2$; draw a picture), the complement of a semi-infinite interval is a semi-open interval of the opposite kind (say, $(-\infty,\, 5]^c = (5,\, \infty] = (5,\, \infty)$: $n = 1$); and the complement of the interval $(-\infty,\, \infty]$ is the empty interval.

Semi-algebras are more convenient to consider than arbitrary systems of subsets of X in that the algebra $\alpha(\mathcal{A})$ generated by a semi-algebra $\mathcal{A}$ consists of finite disjoint unions of sets belonging to $\mathcal{A}$. In fact, we have considered the algebra generated by the semi-algebra of semi-open intervals: I think, our first example of an algebra was that of finite unions of intervals.

Theorem 6.1. *Let $\mathcal{A}$ be a semi-algebra of subsets of a space X; and let m be a finite measure on $\mathcal{A}$. Then there exists a unique measure $\tilde{m}$ on the σ-algebra $\sigma(\mathcal{A})$ generated by $\mathcal{A}$ such that*

$$\tilde{m}(C) = m(C) \qquad \text{for all } C \in \mathcal{A}. \tag{6.1}$$

In other words, a measure m can be extended, and this in a unique way, from a semi-algebra to the σ-algebra generated by it.

The proof of this result belonging to the pure measure theory is, according to the general spirit of these lectures, not given. But let us look how our Theorems 5.4, 5.5′ are proved using Theorem 6.1.

Proof of Theorem 5.4. Suppose μ, ν are two distributions (probability measures) on $(\mathbb{R}^1,\, \mathcal{B}^1)$; and suppose that the corresponding distribution functions coincide:

$$F_\mu(x) = \mu(-\infty,\, x] = F_\nu(x) = \nu(-\infty,\, x] \tag{6.2}$$

for all $x \in (-\infty,\, \infty)$.

Let us consider the semi-algebra $\mathcal{A}$ of all semi-closed intervals in the real line (not in the *extended* real line $[-\infty,\, \infty]$). By Theorem 5.3,

$$\mu(a,\, b] = F_\mu(b) - F_\mu(a) = F_\nu(b) - F_\nu(a) = \nu(a,\, b], \tag{6.3}$$

so the measures μ and ν coincide on the semi-algebra $\mathcal{A}$. By the *uniqueness* of extension, we have that $\mu(C) = \nu(C)$ for all C belonging to the σ-algebra $\sigma(\mathcal{A}) = \mathcal{B}^1$ generated by this semi-algebra.

Theorem 5.5 is more complicated.

Proof of Theorem 5.2. I am going to write here *two* proofs.

Proof 1: Take $\Omega = (-\infty, \infty)$, $\mathcal{F} = \mathcal{B}^1$, $\xi(\omega) = \omega$. Let us define the set function μ on the semi-algebra $\mathcal{A}$ by

$$\mu(a,\, b\,] = F(b) - F(a). \tag{6.4}$$

We have to prove that μ, defined on $\mathcal{A}$, is a measure: i.e., it is nonnegative (obvious by (5.15)), and countably additive:

$$(a,\, b\,] = \bigcup_{i=1}^{\infty} (a_i,\, b_i],\ (a_i,\, b_i] \cap (a_j,\, b_j] = \emptyset\ (i \neq j)\ \Rightarrow\ \mu(a,\, b\,] = \sum_{i=1}^{\infty} \mu(a_i,\, b_i]. \tag{6.5}$$

It is sufficient to prove that

$$\mu(a,\, b\,] \geq \sum_{i=1}^{\infty} \mu(a_i,\, b_i] \tag{6.6}$$

and that

$$\mu(a,\, b\,] \leq \sum_{i=1}^{\infty} \mu(a_i,\, b_i]. \tag{6.7}$$

To prove (6.6) it is sufficient to prove that for every n

$$\mu(a,\, b\,] \geq \sum_{i=1}^{n} \mu(a_i,\, b_i], \tag{6.8}$$

because the right-hand side of (6.6) is the limit of the right-hand side of (6.8). Since there are finitely many (disjoint) intervals in (6.8), we can number them in an increasing order – the sum does not depend on the order of summation:

$$a \leq a_1 \leq b_1 \leq a_2 \leq b_2 \leq \ldots \leq a_{n-1} \leq b_{n-1} \leq a_n \leq b_n \leq b. \tag{6.9}$$

We have by (5.15):

$$F(a) \leq F(a_1),\ F(b_1) \leq F(a_2),\ \ldots,\ F(b_{n-2}) \leq F(a_{n-1}),\ F(b_{n-1}) \leq F(a_n),\ F(b_n) \leq F(b), \tag{6.10}$$

$$\begin{gathered} F(b_1) - F(a_1) \leq F(b_1) - F(a), \quad F(b_2) - F(a_2) \leq F(b_2) - F(b_1), \ldots, \\ F(b_{n-1}) - F(a_{n-1}) \leq F(b_{n-1}) - F(b_{n-2}), \quad F(b_n) - F(a_n) \leq F(b) - F(b_{n-1}). \end{gathered} \tag{6.11}$$

Adding these inequalities, we get:

$$\sum_{i=1}^{n} [F(b_i) - F(a_i)] \leq F(b) - F(a) \tag{6.12}$$

(all other terms cancel), and this is (6.8).

To prove (6.7) it is sufficient to prove that for every $\varepsilon > 0$

$$\mu(a,\, b\,] < \sum_{i=1}^{\infty} \mu(a_i,\, b_i] + \varepsilon. \tag{6.13}$$

The proof uses the right continuity of the function F, and also compactness of finite closed intervals.

For a given $\varepsilon > 0$, we find a compact interval (a closed interval with finite ends) $[a', b'] \subset (a, b]$ such that $\mu(a', b'] > \mu(a, b] - \varepsilon/2$. It is done so: if $b = +\infty$, we take a finite $b' > a$ so that $|F(b') - F(+\infty)| < \varepsilon/4$ (this can be done by the definition of $F(+\infty)$). If $b < \infty$, we take $b' = b$. Then, if $a = -\infty$, we take a', $-\infty < a' < b'$ so that $|F(a') - F(-\infty)| < \varepsilon/4$; and if $a > -\infty$, we take a' close to a on the right so that $|F(a') - F(a)| < \varepsilon/4$ (this can be done because of the right continuity of F).

Then, for every $i = 1, 2, \ldots$ we take an open interval $(a_i, b_i') \supseteq (a_i, b_i]$ such that $\mu(a_i, b_i'] < \mu(a_i, b_i] + \varepsilon/2^{i+1}$: if $b_i = +\infty$, we take $b_i' = b_i = +\infty$ (indeed, $(a_i, +\infty) = \{x \in \mathbb{R}^1 : a_i < x < \infty\} = (a_i, +\infty] = \{x \in R^1 : a_i < x \le \infty\}$), and if $b_i < \infty$, we use the right continuity of F at b_i.

Now we have:

$$[a', b'] \subseteq \bigcup_{i=1}^{\infty} (a_i, b_i') : \tag{6.14}$$

a compact interval is covered by some number of open intervals. We know that then there exists a finite number of them that still cover $[a', b']$:

$$[a', b'] \subseteq \bigcup_{i=1}^{n} (a_i, b_i') \tag{6.15}$$

for some finite n. Among these intervals exists one, (a_{i_1}, b'_{i_1}), containing a'; if this interval contains b' ($b'_{i_1} > b'$), we stop, otherwise we consider another interval (a_{i_2}, b'_{i_2}) containing b'_{i_1}; etc. Finally we come to a finite chain of intervals (a_{i_k}, b'_{i_k}) such that

$$a_{i_1} < a' < b'_{i_1}, \quad a_{i_2} < b'_{i_1} < b'_{i_2}, \quad \ldots, \quad a_{i_m} < b'_{i_{m-1}} < b' < b'_{i_m} \tag{6.16}$$

(draw a picture).

Using the fact that the function F is nondecreasing, we obtain:

$$\begin{aligned} F(b') - F(a') &= F(b'_{i_1}) - F(a') + F(b'_{i_2}) - F(b'_{i_1}) + \ldots + F(b') - F(b_{i_{m-1}}) \\ &\le F(b'_{i_1}) - F(a_{i_1}) + F(b'_{i_2}) - F(a_{i_2}) + F(b'_{i_m}) - F(a_{i_m}) \le \sum_{i=1}^{\infty} [F(b_i') - F(a_i)], \end{aligned} \tag{6.17}$$

$$\mu(a, b] < \mu(a', b'] + \varepsilon/2 \le \sum_{i=1}^{\infty} \mu(a_i, b_i'] + \varepsilon/2 < \sum_{i=1}^{\infty} [\mu(a_i, b_i] + \varepsilon/2^{i+1}] + \varepsilon/2, \tag{6.18}$$

and we get (6.13).

Proof 2 (let me give this proof in the case of Theorem 5.5: for *probability* measures). Let us take $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}_{(0,1)}$, and as the probability P let us take the Lebesgue measure: $P = \lambda_1$. Suppose for a minute that the function $F(x)$ has an inverse, F^{-1}, defined on (0, 1). Let ξ be a random variable defined by $\xi(\omega) = F^{-1}(\omega)$. Since the function F^{-1} is increasing, for every $x \in (-\infty, \infty)$ the set

$$\{\omega\colon F^{-1}(\omega) \le x\} = \{\omega \in (0,1)\colon \omega \le F(x)\} = (0, F(x)]. \tag{6.19}$$

So we have:

$$F_\xi(x) = P\{\omega\colon F^{-1}(\omega) \le x\} = P(0, F(x)] = F(x). \tag{6.20}$$

So our theorem is proved at least for functions F having an inverse defined on (0, 1).

Now the function F may have horizontal parts in its graph and so no inverse function, or it may have jumps, and then the inverse function is not defined for some values of the argument in (0, 1). Nevertheless we can use the same idea, taking a *pseudo-inverse*:

$$G(y) = \min\{x\colon F(x) \ge y\}, \qquad 0 < y < 1 \tag{6.21}$$

(draw a graph of a function F satisfying the conditions (5.15) – (5.18), having jumps and horizontal stretches, and that of the corresponding function G; the minimum exists because the function F is right-continuous); the equality

$$\{\omega\colon G(\omega) \le x\} = \{\omega \in (0, 1)\colon \omega \le F(x)\} = (0, F(x)], \tag{6.22}$$

replacing (6.19), is still true, and the random variable $\xi(\omega) = G(\omega)$ has F as its distribution function.

The second proof seems to be much simpler, avoiding the reasoning about finite covering of a compact interval and the like. However it relies on the existence of the Lebesgue measure, and just the same kind of reasoning as in the first proof is used to prove the existence of the Lebesgue measure.

Note that in both proofs the sample space Ω and the σ-algebra $\mathcal{F}$ of events are taken in a standard way, independently from the given function F. In the first proof the random variable $\xi(\omega) = \omega$ is also taken in a standard way, while the probability P depends on F; in the second one, just the opposite: the probability P is taken to be the Lebesgue measure λ_1, and the random variable ξ is taken dependent on F: the inverse, or the quasi-inverse of this function.

In the two-dimensional case instead of intervals $(a, b]$ we take rectangles $(a_1, b_1] \times (a_2, b_2]$; the two-dimensional distribution function of a random vector $\boldsymbol{\xi} = (\xi_1, \xi_2)$ is defined as

$$F(x_1, x_2) = F_{\xi_1, \xi_2}(x_1, x_2) = P\{\xi_1 \le x_1, \xi_2 \le x_2\}; \tag{6.23}$$

instead of formula (5.12) we have:

$$\begin{aligned} P\{(\xi_1, \xi_2] \in (a_1, b_1] \times (a_2, b_2]\} &= P\{a_1 < \xi_1 \le b_1, a_2 < \xi_2 \le b_2\} \\ &= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, b_1); \end{aligned} \tag{6.24}$$

instead of properties (5.15) – (5.18) we have the following:

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, b_1) \ge 0 \text{ for all } -\infty < a_i \le b_i < \infty, \tag{6.25}$$

$$F(\infty,\,\infty) = 1, \tag{6.26}$$

$$F(-\infty,\,x_2) = F(x_1,\,-\infty) = 0, \tag{6.27}$$

$$\text{the function } F \text{ is right-continuous in both of its arguments;} \tag{6.28}$$

and Theorems 5.4 and 5.5 are true with properties (6.25)–(6.28) instead of (5.15)–(5.18). The first proof of Theorems 5.5 remains essentially the same (only there is much more to write), but its second proof does not work: there is no such thing as the inverse of a function $F : \mathbb{R}^2 \mapsto \mathbb{R}^1$.

In the *multi*dimensional case the formula corresponding to (6.24) has 2^n summands with alternating signs.

However, whereas distribution functions are a convenient enough instrument to work with one-dimensional distributions, two- and more- dimensional distribution functions are a much less convenient tool; so I am mentioning the facts about them only in small font.