Lecture 5.

Let me introduce a term that will allow shorter formulations.

A measure m on a measurable space $(X, \mathcal{X})$ (in fact, on the σ-algebra $\mathcal{X}$) is called a *probability measure* if its total value: the value at the largest set, $m(X)$ is equal to 1. Every distribution of a random variable is a probability measure; and every probability measure m on a measurable space $(X, \mathcal{X})$ is the distribution of some random variable taking values in $(X, \mathcal{X})$ (in fact, of infinitely many random variables). Indeed, we take the probability space $(\Omega, \mathcal{F}, P)$ in this way: $\Omega = X$, $\mathcal{F} = \mathcal{X}$, $P = m$; we define on this space the random variable $\xi(\omega) = \omega$; and the distribution μ_ξ of this random variable will be nothing but the measure m: $\mu_\xi(C) = P\{\omega\colon \xi(\omega) \in C\} = P(C) = m(C)$.

I want to introduce two operations on distributions.

The first one is the *mixture* of a family of distributions.

Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be two measurable spaces. Let μ_y be a family of distributions on $(X, \mathcal{X})$ depending on a parameter $y \in Y$:

$$\text{for every } y \in Y, \quad \mu_y(A),\ A \in \mathcal{X}, \text{ is a measure, and } \mu_y(X) = 1. \tag{5.1}$$

Let this family be measurable in the parameter y:

$$\text{for every } A \in \mathcal{X}, \quad \mu_y(A) \text{ is a } \mathcal{Y}\text{-measurable function of } y \in Y. \tag{5.2}$$

Let ν be a probability distribution on $(Y, \mathcal{Y})$:

$$\nu(B), B \in \mathcal{Y}, \text{ is a measure, and } \nu(Y) = 1. \tag{5.3}$$

We define the mixture of distributions μ_y *with weight* ν by

$$\mu(A) = \int_Y \mu_y(A)\ \nu(dy), \qquad A \in \mathcal{X}. \tag{5.4}$$

Let us check that this is a probability distribution: that μ is a measure, and $\mu(X) = 1$.

That $\mu(A) \ge 0$ and $\mu(X) = 1$ is quite simple; countable additivity: for disjoint A_1, A_2, ..., A_n, ... $\in \mathcal{X}$ we have:

$$\begin{aligned} \mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \int_Y \mu_y\Big(\bigcup_{i=1}^{\infty} A_i\Big)\ \nu(dy) &= \int_Y \sum_{i=1}^{\infty} \mu_y(A_i)\ \nu(dy) \\ &= \sum_{i=1}^{\infty} \int_Y \mu_y(A_i)\ \nu(dy) = \sum_{i=1}^{\infty} \mu(A_i) \end{aligned} \tag{5.5}$$

(by (4.40)).

We can consider mixture of finitely many distributions, say of two:

$$\mu(A) = q_1 \cdot \mu_1(A) + q_2 \cdot \mu_2(A). \tag{5.6}$$

Such mixtures are often considered in applied statistics; e. g., could be used if the data are collected together for two different subgroups of a population, say, for men and women.

But mixtures of an infinite, and even uncountable number of distributions are also considered.

The second operation: putting a distribution through a filter.

Let μ be a probability distribution on $(X, \mathcal{X})$ (i. e., a measure on $\mathcal{X}$ with $\mu(X) = 1$); let $f(x)$, $x \in X$, be a nonnegative measurable function with $0 < \int_X f(x)\ \mu(dx) < \infty$.

We define a new distribution μ_f by

$$\mu_f(A) = \frac{\int_A f(x)\ \mu(dx)}{\int_X f(x)\ \mu(dx)} \tag{5.7}$$

(dividing by the integral over the whole space ensures that $\mu_f(X) = 1$).

A problem about this:

9 Let μ be the one-dimensional normal distribution with parameters (a, b) ($a \in (-\infty, \infty)$, $b \in (0, \infty)$); i. e., a continuous distribution on $\mathbb{R}^1$ with density

$$p(x) = \frac{1}{\sqrt{2\pi b}}\, e^{-(x-a)^2/2b} \tag{5.8}$$

(this function is positive, and it can be checked that its integral over the whole line is equal to 1).

What is the result of putting this distribution through the filter with $f(x) = e^{cx}$?

In what problems such operations on distributions arise, we'll see later.

In the previous lectures, discrete distributions and (absolutely) continuous distributions were introduced. Are there distributions that do not belong to these two classes?

Of course, there are: mixtures of discrete and continuous distributions: distributions given by formula (5.6), where q_1, $q_2 > 0$, $q_1 + q_2 = 1$, μ_1 is a discrete distribution, and μ_2 an aboslutely continuous one. The mixture is neither discrete nor continuous, because

$$\sum_{x \in \mathbb{R}^1} \mu\{x\} = q_1 \neq 0,\ 1, \tag{5.9}$$

while for discrete distributions it should be equal to 1, and for continuous, to 0.

Are there distributions on the real line that are not such mixtures?

We'll return to this question after considering *distribution functions.*

Let ξ be a real-valued random variable. Its distribution function is a function on $(-\infty, \infty)$ defined by

$$F(x) = F_\xi(x) = P\{\xi \leq x\}. \tag{5.10}$$

Clearly the distribution function depends only on the distribution μ of a random variable:

$$F(x) = F_\mu(x) = \mu(-\infty, x]. \tag{5.11}$$

Theorem 5.1. *We have, for all* $-\infty < a \leq b < \infty$:

$$P\{a < \xi \leq b\} = F_\xi(b) - F_\xi(a). \tag{5.12}$$

The **proof** is very simple: Clearly, $(-\infty, b] = (-\infty, a] \cup (a, b]$, and these intervals are disjoint; the same holds for their inverse images under the mapping ξ:

$$\{\xi \leq b\} = \{\xi \leq a\} \cup \{a < \xi \leq b\}, \tag{5.13}$$

and by (finite) additivity of P,

$$P\{\xi \leq b\} = P\{\xi \leq a\} + P\{a < \xi \leq b\}, \tag{5.14}$$

which leads to (5.12).

Theorem 5.2. *If* $F(x)$ *is the distribution function of a random variable* ξ, *then*

$$F(x) \text{ is non-decreasing,} \tag{5.15}$$

$$F(\infty) = 1, \tag{5.16}$$

$$F(-\infty) = 0, \tag{5.17}$$

$$F(c^+) = F(c) \qquad \text{for } c \in (-\infty, \infty), \tag{5.18}$$

$$F(c^-) = P\{\xi < c\} \tag{5.19}$$

($F(\infty)$, $F(-\infty)$, $F(c^+)$, $F(c^-)$ are the notations for the limits $\lim_{x\to\infty} F(x)$, $\lim_{x\to-\infty} F(x)$, the right-hand limit at c: $\lim_{x\to c^+} F(x)$, and the left-hand limit $\lim_{x\to c^-} F(x)$. Note that the formula (5.10) does not define $F(x)$ on the *extended* real line).

Formula (5.18) means that any distribution function is continuous from the right at every point of the real axis.

Proof. The statement (5.15) follows from Theorem 5.1. For a monotone function, limits at $\pm\infty$ and one-sided limits at every finite point necessarily exist, so the limits (5.16) – (5.19) do exist.

The limit at $+\infty$ is equal to the limit along every sequence of numbers going to ∞; e. g.,

$$F(\infty) = \lim_{n\to\infty} F(n) = \lim_{n\to\infty} P\{\xi \leq n\}. \tag{5.20}$$

The sequence of events $B_n = \{\xi \le n\}$ is clearly non-decreasing, so by Theorem $1-2.1$ we have:

$$F(\infty) = P(\lim_{n\to\infty} \{\xi \le n\}). \tag{5.21}$$

What is the limit of this sequence of events? It is a non-decreasing sequence, so the limit is the union:

$$\lim_{n\to\infty} \{\xi \le n\} = \bigcup_{i=1}^{\infty} \{\xi \le i\}. \tag{5.22}$$

But this union clearly is the whole sample space Ω: for every sample point $\omega \in \Omega$, there exists at least one natural number i such that $i > \xi(\omega)$ (or $i \ge \xi(\omega)$). So

$$F(\infty) = P(\Omega) = 1. \tag{5.23}$$

Now to (5.17):

$$\begin{aligned} F(-\infty) = \lim_{n\to\infty} F(-n) = \lim_{n\to\infty} P\{\xi \le -n\} = P(\lim_{n\to\infty} \{\xi \le -n\}) \\ = P\big(\bigcap_{i=1}^{\infty} \{\xi \le -i\}\big) = P(\emptyset) = 0 \end{aligned} \tag{5.24}$$

(the events $\{\xi \le -n\}$ form a *non-increasing* sequence with (clearly) empty intersection).

To (5.18) (with a non-increasing sequence):

$$F(c^+) = \lim_{n\to\infty} P\{\xi \le c + 1/n\} = P\big(\bigcap_{i=1}^{\infty} \{\xi \le c + 1/i\}\big) = P\{\xi \le c\} = F(c) \tag{5.25}$$

(the event $\{\xi \le c\}$ occurs if and only if all events $\{\xi \le c + 1/i\}$ occur: the "if" part by limit passage in the inequality $\xi \le c + 1/i$, and the "only if": $\xi \le c \;\Rightarrow\; \xi \le c + 1/i$ for all natural i). Finally, (5.19) (again a non-decreasing sequence of events):

$$\bigcup_{i=1}^{\infty} \{\xi \le c - 1/i\} = \{\xi < c\} \tag{5.26}$$

(if $\xi(omega)$ is $\le$ than at least one $c - 1/i$, it is certainly less than c; and if $\xi(\omega)$ is less than c, there exists a natural number i such that $c - 1/i \in \big(\xi(\omega), c\big)$, and ω belongs to the event in the left-hand side of (5.26)), so

$$F(c^-) = \lim_{n\to\infty} P\{\xi \le c - 1/n\} = P\big(\bigcup_{i=1}^{\infty} \{\xi \le c - 1/i\}\big) = P\{\xi < c\}. \tag{5.27}$$

From (5.10) and (5.19) we obtain easily:

$$P\{\xi = c\} = P\{\xi \le c\} - P\{\xi < c\} = F(c) - F(c^-) = F(c^+) - f(c^-); \tag{5.28}$$

this is the *jump* of the distribution function F at the point c.

From this we easily obtain a description of the distribution function of a discrete distribution: the distribution function grows by jumps of size $p(x^i) = P\{\xi = x^i\}$ at the points x^i that are the values taken by the random variable (function) ξ; and on the intervals between the jumps the function remans constant. Since every distribution function is continuous from the right, these intervals are open on the left, but closed on the right (make a picture). (Note that these intervals (say, $[x^i, x^{i+1})$) are of a different sort from the intervals $(a, b]$ mentioned in Theorem 5.1.)

This is what we tell our students in an elementary course of probability theory; and this is the truth, but not the *whole* truth: in fact, for a discrete random variable there may be no such thing as *intervals between the jumps*: e. g., if the random variable takes all rational values (and the set of rational numbers is, as we know, countable), there are no such intervals: in every small interval there are infinitely many rational points.

The distribution function of a discrete random variable can be written in the form

$$F(x) = \sum_{i:\, x^i \le x} p(x^i); \tag{5.29}$$

for a continuous random variable the corresponding formula is

$$F(x) = \int_{-\infty}^{x} p(u)\, du, \tag{5.30}$$

where p is the probability density.

Theorem 5.3. *For all* $-\infty \le a \le b \le \infty$ *we have*:

$$P\{a < \xi \le b\} = F_\xi(b) - F_\xi(a), \tag{5.31}$$

$$P\{a < \xi < b\} = F_\xi(b^-) - F_\xi(a), \tag{5.32}$$

$$P\{a \le \xi \le b\} = F_\xi(b) - F_\xi(a^-), \tag{5.33}$$

$$P\{a \le \xi < b\} = F_\xi(b^-) - F_\xi(a^-), \tag{5.34}$$

where $F_\xi(\infty)$, $F_\xi(-\infty)$, $F_\xi(c^-)$ *denote the limits, and as* $F_\xi(\infty^-)$, $F_\xi(-\infty^-)$ *we take* $F_\xi(\infty)$, $F_\xi(-\infty)$, *that is* 1, 0.

That is, if the inequality at an end of the interval from a to b is not such as in (5.31), we replace the value of the function F_ξ at this end with the left limit at it (or do nothing if the limit is infinite, and it is all the same whether we have the $<$ or the $\le$ inequality at it).

Proof. The first statement is almost that of Theorem 5.1, except that it allows now $a = -\infty$ or $b = +\infty$ (or both). Here is how we check it for $a \in (-\infty, \infty)$, $b = \infty$:

$$P\{a < \xi \le \infty\} = P\{\xi > a\} = P(\{\xi \le a\}^c) = 1 - P\{\xi \le a\} = 1 - F_\xi(a) = F_\xi(\infty) - F_\xi(a). \tag{5.35}$$

The equalities (5.32) – (5.34) are checked using the fact that $F_\xi(\infty) = 1$, $F_\xi(-\infty) = 0$ (for b or a being infinite), and the equality (5.28). Say, (5.32) for $b \in (-\infty, \infty)$:

$$\begin{aligned} P\{a < \xi < b\} &= P\{a < \xi \le b\} - P\{\xi = b\} \\ &= F_\xi(b) - F_\xi(a) - [F_\xi(b) - F_\xi(b^-)] = F_\xi(b^-) - F_\xi(a). \end{aligned} \tag{5.36}$$

Now we formulate two theorems (big and serious theorems) about distribution functions.

Theorem 5.4. *Let* μ *and* ν *be two distributions* (two probability measures) *on the real line* $\mathbb{R}^1$; *let* F_μ, F_ν *be the corresponding distribution functions*:

$$F_\mu(x) = \mu(-\infty, x], \qquad F_\nu(x) = \nu(-\infty, x]. \tag{5.37}$$

If $F_\mu(x) = F_\nu(x)$ *for all* $x \in (-\infty, \infty)$, *then* $\mu = \nu$ (i. e., the distribution function determines the distribution uniquely).

Theorem 5.5. *Let* $F(x)$, $-\infty < x < \infty$, *be a function satisfying the conditions* (5.15) – (5.18) ((5.19) is not a property of a function: while (5.15) – (5.18) are only about a function F, (5.19) is about this function and the probability measure P).

Then there exists a probability space $(\Omega, \mathcal{F}, P)$ *and a random variable* ξ *on it such that its distribution function is the prescribed function* F:

$$F_\xi(x) = P\{\xi \le x\} = F(x), \qquad -\infty < x < \infty. \tag{5.38}$$

About how Theorems 5.4 and 5.5 are proved I will speak in the next lecture.