Lecture 8. Independence.

What is central to probability theory are the concepts of independence, dependence, and how dependence is characterized by means of conditional probabilities. The parts of probability theory that do not use these concepts should be considered as belonging not to probability theory strictly speaking, but rather to measure theory. Such is the very important concept of the *distribution* of a random variable, and various ways how to characterize distributions; the expectation; etc. (distributions are, in fact, measures carried to another space under a measurable mapping, expectations are Lebesgue integrals, etc.). I could have included in the lectures first all things "prior" to the central concepts of independence, dependence and conditional probabilities (and there is indeed much material here: characterization of distributions by means of characteristic functions, various types of convergence of random variables and of distributions, etc.), and come to these central concepts only afterwards; but I think the time has come for them now.

This concept of *independence* is applied to events, to whole *classes* of events, and also to random variables. In our advanced course of probability, let us start with random variables.

Two random variables ξ and η on the same probability space $(\Omega, \mathcal{F}, P)$, taking values in measurable spaces $(X, \mathcal{X})$, $(Y, \mathcal{Y})$ are called independent if for all $C \in \mathcal{X}$, $D \in \mathcal{Y}$

$$P\{\xi \in C, \eta \in D\} = P\{\xi \in C\} \cdot P\{\eta \in D\}. \tag{8.1}$$

Two events A, B are called independent if their indicator random variables $I_A = I_A(\omega)$, I_B are independent.

So this definition means that

$$P\{I_A \in C, I_B \in D\} = P\{I_A \in C\} \cdot P\{I_B \in D\} \tag{8.2}$$

for all C and D in the appropriate σ-algebra.

An indicator random variable I_A is a function $I_A : \Omega \mapsto \{0, 1\}$; or, as we can say, a measurable function from the measurable space $(\Omega, \mathcal{F})$ into the measurable space $(\{0, 1\}, \mathcal{P}\{0, 1\})$. The σ-algebra $\mathcal{P}\{0, 1\}$: that of *all* subsets of the two-point set $\{0, 1\}$ consists of exactly four sets: $\emptyset$, $\{0\}$, $\{1\}$, and $\{0, 1\}$. So we have to check (8.2) for four sets C and four sets D: check 16 equalities in all.

But if C or D is the empty set, the equality (8.2) reduces to $P(\emptyset) = P(\emptyset) \cdot P\{I_B \in D\}$, i.e., $0 = 0$, which is of course true and does not require any checking; and for $C = \{0, 1\}$ the event $\big\{\xi \in \{0, 1\}, \eta \in D\big\} = \{\eta \in D\}$, and (8.2) reduces to $P\{\eta \in D\} = P\big\{\xi \in \{0, 1\}\big\} \cdot P\{\eta \in D\} = P(\Omega) \cdot P\{\eta \in D\} = 1 \cdot P\{\eta \in D\}$, which is also true, and no checking is needed. So there remain *four* equalities:

$$P\{I_A = 1, I_B = 1\} = P\{I_A = 1\} \cdot P\{I_B = 1\}, \tag{8.3}$$

or simply

$$P(A \cap B) = P(A) \cdot P(B); \tag{8.4}$$

and also

$$P(A \cap B^c) = P(A) \cdot P(B^c), \quad P(A^c \cap B) = P(A^c) \cdot P(B), \quad P(A^c \cap B^c) = P(A^c) \cdot P(B^c). \tag{8.5}$$

In fact, each of the equalities (8.4) – (8.5) can be deduced from each other (I *prove* it when I teach a lower-level probability course), so we could reduce our definition of independence of two events to *one* equality; but we haven't taken the oath of necessarily reducing all our definitions to the minimum number of formulas – so we won't do it.

And two classes of events $\mathcal{A}$, $\mathcal{B} \subseteq \mathcal{F}$ (I feel a little uncomfortable denoting a class of subsets of the sample space with the same letter $\mathcal{B}$ that serves us to denote Borel σ-algebras in Euclidean spaces; but I hope you won't get confused) are called independent if for any $A \in \mathcal{A}$, $B \in \mathcal{B}$ the events A and B are independent.

The concept of independence is related to an important concept of measure theory: that of *product of measures.*

But first let us take up a simple subject in the set-theoretic introduction to measure theory.

Suppose $\mathcal{X}$ is a σ-algebra in a space X, and $\mathcal{Y}$ in a space Y. The (*direct*) *product* of these σ-algebras is, by definition, the σ-algebra $\mathcal{X} \times \mathcal{Y}$ in the product space $X \times Y$ generated by all subsets of $X \times Y$ of the form $A \times B$, $A \in \mathcal{X}$, $B \in \mathcal{Y}$.

For example, the direct product of twice the same Borel σ-algebra $\mathcal{B}^1$ is the two-dimensional Borel σ-algebra $\mathcal{B}^2$ in the plane $\mathbb{R}^2$ (this needs a proof; but I omit it – anyway it is very simple, as almost everything in the set-theoretic introduction to measure theory and probability theory).

Theorem 8.1. *Let m be a σ-finite measure in the measurable space $(X, \mathcal{X})$, and n in the space $(Y, \mathcal{Y})$.*

Then there exists a unique measure $m \times n$ on the product space $(X \times Y, \mathcal{X} \times \mathcal{Y})$ such that

$$(m \times n)(A \times B) = m(A) \cdot n(B) \tag{8.6}$$

for all $A \in \mathcal{X}$, $B \in \mathcal{Y}$.

The measure $m \times n$ is called the product of the measures m and n.

In fact, we'll be using this theorem only for finite measures; I included the weaker requirement of σ-finiteness in order that the theorem should be applicable to the Lebesgue measure.

The **proof** (for finite measures) is based on Theorem 6.1: we define $m \times n$ on the semi-algebra of "rectangles" $A \times B$ by (8.2); and we check that this $m \times n$ is a *measure* (i. e., is countably additive). For σ-finite measures we construct $m \times n$ piece by piece (it being finite on each separate piece).

So, the two-dimensional Lebesgue measure λ_2 is the product $\lambda_1 \times \lambda_1$ of twice the same one-dimensional Lebesgue measure.

Now to independence. It can be formulated as follows: two random variables ξ, η on the same probability space are independent if the joint distribution $\mu_{\xi,\,\eta}$ of these two random variables (which is defined as the distribution of the random vector $\boldsymbol{\xi} = (\xi, \eta)$

taking values in the product space $(X \times Y, \mathcal{X} \times \mathcal{Y})$: $\mu_{\xi,\,\eta}(E) = \mu_{\boldsymbol{\xi}}(E) = P\{(\xi,\,\eta) \in E\}$, $E \in \mathcal{X} \times \mathcal{Y}$), if this joint distribution is the direct product of their individual distributions:

$$\mu_{\xi,\,\eta} = \mu_\xi \times \mu_\eta. \tag{8.7}$$

Theorem 8.2. *A random variable ξ taking values in the real line is independent from itself if and only if there is a value $a \in \mathbb{R}^1$ such that $\xi = a$ almost surely.*

Proof. The independence in question means that

$$P\{\xi \in C,\, \xi \in D\} = P\{\xi \in C\} \cdot P\{\xi \in D\} \tag{8.8}$$

for all $C,\, D \in \mathcal{B}^1$. The left-hand side is nothing but $P\{\xi \in C \cap D\}$.

The "if" part: both sides of (8.8) are equal to 1 if both $C \ni a$ and $D \ni a$, and to 0 if $C \not\ni a$ or $D \not\ni a$ (or both).

The "only if": Let us take $D = C$:

$$P\{\xi \in C,\, \xi \in C\} = P\{\xi \in C\} = P\{\xi \in C\} \cdot P\{\xi \in C\} = P\{\xi \in C\}^2. \tag{8.9}$$

The quadratic equation $x = x^2$ has only two solutions: $x = 0$ and $x = 1$; so for every Borel C we have either $P\{\xi \in C\} = 0$, or 1.

The distribution function $F_\xi(x) = P\{\xi \in (-\infty,\, x]\}$ takes only two values, 0 and 1; its limit at $-\infty$ is equal to 0, and 1 at $+\infty$. So the function F_ξ can increase only by a jump: there exists a number a such that $F_\xi(a^-) = 0$, and $F_\xi(a) = F_\xi(a^+) = 1$. And we have: $P\{\xi = a\} = F_\xi(a) - F_\xi(a^-) = 1$.

The theorem remains true for random vectors with values in $\mathbb{R}^n$; only I cannot see how to prove it using the n-dimensional distribution function (I have told you that the distribution function is not a good tool in the multidimensional case).

Now, by definition, n random variables $\xi_1,\, \xi_2,\, \ldots,\, \xi_n$ defined on the same probability space $(\Omega,\, \mathcal{F},\, P)$ and taking values in measurable spaces $(X_i,\, \mathcal{X}_i)$, $i = 1,\, 2,\, \ldots,\, n$, are independent if

$$P\{\xi_1 \in C_1, \ldots, \xi_n \in C_n\} = P\{\xi_1 \in C_1\} \cdot \ldots \cdot P\{\xi_n \in C_n\}. \tag{8.10}$$

This can be formulated in the language of direct products of measures and distributions: the random variables $\xi_1,\, \xi_2,\, \ldots,\, \xi_n$ are independent if and only if their joint distribution $\mu_{\xi_1,\, \ldots,\, \xi_n}$ is the direct product of their individual distributions:

$$\mu_{\xi_1,\, \ldots,\, \xi_n} = \mu_{\xi_1} \times \ldots \times \mu_{\xi_n}. \tag{8.11}$$

It's true we haven't introduced the direct product of $n > 2$ measures. It can be defined quite anew, using Theorem 6.1, starting with "parallelepipeds"; but also we can define $\mu_{\xi_1} \times \mu_{\xi_2} \times \mu_{\xi_2}$ just as $(\mu_{\xi_1} \times \mu_{\xi_2}) \times \mu_{\xi_3}$, or in any other order, e. g.: $\mu_{\xi_1} \times (\mu_{\xi_2} \times \mu_{\xi_3})$: it can be proved that all these direct products are the same measure in $(X_1 \times X_2 \times X_3,\, \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$; and so on.

Now to independence of events: n events $A_1, \ldots, A_n$ (also in the same probability space, i. e., all A_i belonging to the same σ-algebra $\mathcal{F}$) are independent if the indicator random variables $I_{A_1}, \ldots, I_{A_n}$ are – or, deciphering this: if

$$P(B_1 \cap B_2 \cap \ldots \cap B_n) = P(B_1) \cdot P(B_2) \cdot \ldots \cdot P(B_n), \tag{8.12}$$

where for every $i = 1, 2, \ldots, n$, the event $B_i = A_i$ or A_i^c. Formula (8.12) is the short expression for 2^n formulas $P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_n)$, $P(A_1^c \cap A_2 \cap \ldots \cap A_n) = P(A_1^c) \cdot P(A_2) \cdot \ldots \cdot P(A_n)$, ..., $P(A_1^c \cap A_2^c \cap \ldots \cap A_n^c) = P(A_1^c) \cdot P(A_2^c) \cdot \ldots \cdot P(A_n^c)$. These 2^n formulas *cannot* be reduced to one, say, $P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_n)$; they *can be* reduced to a smaller number, say, for $n = 10$ to 1013 instead of $2^{10} = 1024$; but I cannot see why should we aspire to this reduction. Finally, n classes of events $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n \subseteq \mathcal{F}$ are independent, by definition, if arbitrary events $A_1 \in \mathcal{A}_1, \ldots, A_n \in \mathcal{A}_n \subseteq \mathcal{F}$ are independent.

The last thing: suppose we have an infinite family ξ_α, $\alpha \in A$, of random variables. We say that ξ_α, $\alpha \in A$, are independent if for every finite subset $\{\alpha_1, \ldots, \alpha_n\} \subseteq A$ the random variables $\xi_{\alpha_1}, \ldots, \xi_{\alpha_n}$ are independent (of course we mean that $\alpha_i \neq \alpha_j$ for $i \neq j$ – no random variable could be independent with itself, see Theorem 8.2). Independence of an infinite family of events B_α, $\alpha \in A$, and of an infinite system of classes of events $\mathcal{A}_\alpha \subseteq \mathcal{F}$ is defined in an obvious way.

Now we are going to consider some properties of independent objects, and of products of measures. The first theorem I'll formulate here is Fubini's Theorem; but before formulating this important theorem of the theory of measure and integration, I'll formulate another, a simpler one, belonging to the set-theoretic introduction to measure theory:

Theorem 8.3. *Let $(X, \mathcal{X})$, $(Y, \mathcal{Y})$, $(Z, \mathcal{Z})$ be measurable spaces. Let $f(x, y)$ be a function $X \times Y \mapsto Z$ that is measurable with respect to $\mathcal{X} \times \mathcal{Y}$, $\mathcal{Z}$. Then for every $x \in X$ the function $f(x, y)$ is measurable in y with respect to $\mathcal{Y}$, $\mathcal{Z}$, and for every fixed y it is measurable in x.*

The **proof** is standard, and very simple; I'll skip it.

Theorem 8.4 (Fubini's Theorem). *Let m and n be σ-finite measures on $(X, \mathcal{X})$, $(Y, \mathcal{Y})$. Let $f(x, y)$ be an $\mathcal{X} \times \mathcal{Y}$-measurable function of $x \in X$, $y \in Y$ taking values in the extended real line $[-\infty, \infty]$.*

1) *If $f \geq 0$, the function*

$$\int_Y f(x, y)\; n(dy) \tag{8.13}$$

is measurable in x, and the function

$$\int_X f(x, y)\; m(dx) \tag{8.14}$$

is measurable in y; and

$$\int\limits_{X \times Y} f(x, y)\; (m \times n)(dx\, dy) = \int\limits_X \Big[\int\limits_Y f(x, y)\; n(dy) \Big]\; m(dx) = \int\limits_Y \Big[\int\limits_X f(x, y)\; m(dx) \Big]\; n(dy). \tag{8.15}$$

2) *For a function* f *taking values of both signs, the set of* x *for which the integral* (8.13) *exists belongs to the* σ*-algebra* $\mathcal{X}$, *and the function* (8.13) *is measurable in* x *on this set*; *the same for the integral* (8.14). *If one of the integrals*

$$\int\limits_{X\times Y} |f(x,\, y)|\; (m\times n)(dx\, dy), \qquad \int\limits_X \Big[\int\limits_Y |f(x,\, y)|\; n(dy)\Big]\; m(dx),$$

$$\int\limits_Y \Big[\int\limits_X |f(x,\, y)|\; m(dx)\Big]\; n(dy) \tag{8.16}$$

converges, then the integrals (8.13), (8.14) *converge almost everywhere*, *and formula* (8.15) *holds if we define the functions* (8.13), (8.14) *arbitrarily where they are not defined* (which is on sets of zero measure).

This rule is the reason for notations: we denote integration over the product space by two (or more, for multiple products) integrals if the integration variable is written in coordinates; and the differential of the product measure (used in integration, or in writing the density) is, traditionally, written as the product of differentials of the factor measures:

$$(m\times n)(dx\, dy) = m(dx)\; n(dy). \tag{8.17}$$

So formula (8.16) is rewritten as

$$\iint\limits_{X\times Y} f(x,\, y)\; m(dx)\, n(dy) = \int\limits_X \big[\int\limits_Y f(x,\, y)\; n(dy)\big]\; m(dx) = \dots . \tag{8.18}$$

Most often we use Fubini's Theorem to change the order of integration.