**Lecture 10. More on independence.**

Now about independence of infinitely many random variables (events, whole classes of events). (But what we'll have here will be also true for finite number of random variables, events, classes of events.) First we go to the set-theoretic introduction to probability theory (which is the same as for measure theory).

We have introduced $\sigma$-algebras in the sample space $\Omega$ generated by classes of events; we can also introduce $\sigma$-algebras generated by (families of) random variables.

With every random variable, say, $\xi$, we can associate the $\sigma$-algebra of events (i. e., $\sigma$-algebra in the sample space $\Omega$ being a part of our fundamental $\sigma$-algebra $\mathcal{F}$) generated by it,

$$\sigma(\xi) = \big\{\{\xi \in C\} \colon C \in \mathcal{X}\big\}, \tag{10.1}$$

where $\mathcal{X}$ is the $\sigma$-algebra that we are considering in the space where $\xi$ takes its values.

Absolutely in the same way we can consider the $\sigma$-algebra generated by several random variables, say, $\xi_1$, $\xi_2$, and $\xi_3$:

$$\sigma(\xi_1,\, \xi_2,\, \xi_3) = \sigma\big\{\{\xi_1 \in C_1\}, \{\xi_2 \in C_2\}, \{\xi_3 \in C_3\} \colon C_1 \in \mathcal{X}_1, C_2 \in \mathcal{X}_2, C_3 \in \mathcal{X}_3\big\}; \tag{10.2}$$

this $\sigma$-algebra can also be written as

$$\sigma(\xi_1,\, \xi_2,\, \xi_3) = \big\{\{(\xi_1,\, \xi_2,\, \xi_3) \in C\} \colon C \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3\big\}. \tag{10.3}$$

Again, absolutely in the same way we can consider $\sigma$-algebras of events generated by infinitely many random variables, e. g.,

$$\sigma(\xi_\alpha,\, \alpha \in A) = \sigma\big\{\{\xi_\alpha \in C_\alpha\} \colon C_\alpha \in \mathcal{X}_\alpha\big\}; \tag{10.4}$$

only this $\sigma$-algebra cannot be written in a form similar to (10.3), because we haven't defined what an infinite direct product of $\sigma$-algebras is.

**Theorem 10.1.** *Let $\xi_\alpha$, $\alpha \in A$, be a family of independent random variables. Let $A_\beta$, $\beta \in B$, be a family of disjoint subsets of the index set $A$: $A_\beta \subseteq A$, $A_\beta \cap A_\gamma = \emptyset$ ($\beta \ne \gamma$).*

*Then the $\sigma$-algebras $\mathcal{S}_\beta = \sigma(\xi_\alpha,\, \alpha \in A_\beta)$, $\beta \in B$, are independent.*

This theorem makes perfect sense in the case of finite families of random variables: e. g., $A = \{1,\, 2,\, 3,\, 4,\, 5,\, 6\}$, $\xi_1$, $\xi_2$, ..., $\xi_6$ independent random variables, $B = \{1,\, 2,\, 3\}$, $A_1 = \{1,\, 2,\, 4\}$, $A_2 = \{3,\, 6\}$, $A_3 = \{5\}$; we conclude that the $\sigma$-algebras $\mathcal{S}_1 = \sigma(\xi_1,\, \xi_2,\, \xi_4\}$, $\mathcal{S}_2 = \sigma(\xi_3,\, \xi_6)$, $\mathcal{S}_3 = \sigma(\xi_5)$ are independent. Or: the random vectors $\boldsymbol{\xi}_1 = (\xi_1,\, \xi_2,\, \xi_4)$, $\boldsymbol{\xi}_2 = (\xi_3,\, \xi_6)$ and the random variable $\xi_5$ are independent. (Or, by Theorem 9.5: the random variables $\eta = \sqrt{\xi_1^2 + \xi_2^2 + \xi_4^2}$, $\zeta = \xi_3 + \xi_6$, and $\xi_5$ are independent.)

**Proof.** Of course, since the definition of independence of infinitely many objects (random variables, or events, or whole classes of such) are defined through independence of finitely many objects, we come to some statement about finite number of them.

We have to prove that for every finite set $\{\beta_1, ..., \beta_k\} \subseteq B$ the $\sigma$-algebras $\sigma(\xi_\alpha, \alpha \in A_{\beta_1})$, ..., $\sigma(\xi_\alpha, \alpha \in A_{\beta_k})$ are independent. The $\sigma$-algebra $\sigma(\xi_\alpha, \alpha \in A_{\beta_i})$ is generated by the semi-algebra $\big\{\{\xi_{\alpha_{i1}} \in C_{\alpha_{i1}}, ..., \xi_{\alpha_{ij_i}} \in C_{\alpha_{ij_i}}\}, \alpha_{i1}, ..., \alpha_{ij_i} \in A_{\beta_i}, C_{\alpha_{i1}} \in \mathcal{X}_{\alpha_{i1}}, ..., C_{\alpha_{ij_i}} \in \mathcal{X}_{\alpha_{ij_i}}\big\}$, so it is enough to prove that the events $\{\{\xi_{\alpha_{11}} \in C_{\alpha_{11}}, ..., \xi_{\alpha_{1j_1}} \in C_{\alpha_{1j_1}}\}, ..., \{\xi_{\alpha_{k1}} \in C_{\alpha_{k1}}, ..., \xi_{\alpha_{kj_k}} \in C_{\alpha_{kj_k}}\}$ are independent for every finite $k$, $j_1$, ..., $j_k$. And this means that

$$\begin{aligned} &P\{\xi_{\alpha_{11}} \in C_{\alpha_{11}}, ..., \xi_{\alpha_{1j_1}} \in C_{\alpha_{1j_1}}, ..., \xi_{\alpha_{k1}} \in C_{\alpha_{k1}}, ..., \xi_{\alpha_{kj_k}} \in C_{\alpha_{kj_k}}\} \\ &\quad = P\{\xi_{\alpha_{11}} \in C_{\alpha_{11}}, ..., \xi_{\alpha_{1j_1}} \in C_{\alpha_{1j_1}}\} \cdot ... \cdot P\{\xi_{\alpha_{k1}} \in C_{\alpha_{k1}}, ..., \xi_{\alpha_{kj_k}} \in C_{\alpha_{kj_k}}\} \end{aligned} \tag{10.5}$$

for every finite $k$, $j_1$, ..., $j_k$, $C_{\alpha_{ij}} \in \mathcal{X}_{\alpha_{ij}}$. But this follows from the independence of $\xi_\alpha$, because both sides are equal to

$$P\{\xi_{\alpha_{11}} \in C_{\alpha_{11}}\} \cdot ... \cdot P\{\xi_{\alpha_{1j_1}} \in C_{\alpha_{1j_1}}\} \cdot ... \cdot P\{\xi_{\alpha_{k1}} \in C_{\alpha_{k1}}\} \cdot ... \cdot P\{\xi_{\alpha_{kj_k}} \in C_{\alpha_{kj_k}}\}. \tag{10.6}$$

The mathematical concept of independent random variables is designed to serve as a mathematical model for quantities that are "physically" independent (this, in the extra-mathematical world). It would be very bad if it turned out that no such mathematical object as independent random variables exists (or independent random variables exist, but only under very severe restrictions): it would mean that the mathematical model is not valid. So we go to the question of existence.

For *finitely* many random variables the question for is solved easily:

**Theorem 10.2.** *Let* $\mu_1$, ..., $\mu_n$ *be distributions* (i. e. probability measures – with their largest value equal to 1) *on measurable spaces* $(X_1, \mathcal{X}_1)$, ..., $(X_n, \mathcal{X}_n)$. *Then there exists a probability space* $(\Omega, \mathcal{F}, P)$ *and independent random variables* $\xi_1$, ..., $\xi_n$ *on it such that the distribution of* $\xi_i$ *is* $\mu_i$.

**Proof.** We take $\Omega = X_1 \times ... \times X_n$, $\mathcal{F} = \mathcal{X}_1 \times ... \times \mathcal{X}_n$, $P = \mu_1 \times ... \times \mu_n$, and, for $\omega = (x_1, ..., x_n)$,

$$\xi_i(\omega) = x_i, \qquad i = 1, ..., n. \tag{10.7}$$

The joint distribution of $\xi_1$, ..., $\xi_n$ clearly is $\mu_1 \times ... \times \mu_n$, the distribution of every $\xi_i$ is $\mu_i$:

$$\begin{aligned} &\mu_{\xi_i}(C) = P\{\xi_i \in C\} = (\mu_1 \times ... \times \mu_n)(X_1 \times ... \times X_{i-1} \times C \times X_{i+1} \times ... \times X_n) = \\ &\quad \mu_1(X_1) \cdot ... \cdot \mu_{i-1}(X_{i-1}) \cdot \mu_i(C) \cdot \mu_{i+1}(X_{i+1}) \cdot ... \cdot \mu_n(X_n) = 1 \cdot ... \cdot 1 \cdot \mu_i(C) \cdot 1 \cdot ... \cdot 1. \end{aligned} \tag{10.8}$$

The random variables $\xi_1$, ..., $\xi_n$ are independent because their joint distribution is the direct product of their individual distributions in $X_i$.

**Theorem 10.3.** *There exists a probability space and an* infinite *sequence of independent random variables* $\xi_1$, ..., $\xi_n$, ... *on it such that all of them have the distribution with probability* 1/2 *at each of the points* 0, 1:

$$P\{\xi_i = 0\} = P\{\xi_i = 1\} = 1/2. \tag{10.9}$$

**Proof.** Take $\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}_{[0,1)}$, $P$ being the one-dimensional Lebesgue measure, $\lambda_1$.

Each number $\omega \in \Omega = [0, 1)$ has a binary representation:

$$\omega = \frac{a_1}{2} + \frac{a_2}{4} + \frac{a_3}{8} + ... + \frac{a_n}{2^n} + ..., \qquad a_i = 0 \text{ or } 1. \tag{10.10}$$

Some numbers have two different binary representations, e. g., $\frac{3}{8} = \frac{0}{2} + \frac{1}{4} + \frac{1}{8} + \frac{0}{16} + ... + \frac{0}{2^n} + ... = \frac{0}{2} + \frac{1}{4} + \frac{0}{8} + \frac{1}{16} + \frac{1}{32} + ... + \frac{1}{2^n} + ...$ . For definiteness, take the binary representation with infinitely many zeros.

Let us define the function $\xi_i$ on the interval $[0, 1)$ by taking $\xi_i(\omega)$ being equal to the $i$-th binary digit in the representation (10.10): $\xi_i(\omega) = a_i$.

We have:

$$\{\omega\colon \xi_1(\omega) = 0\} = [0, 1/2), \qquad \{\omega\colon \xi_1(\omega) = 1\} = [1/2, 1); \tag{10.11}$$

$$\{\omega\colon \xi_2(\omega) = 0\} = [0, 1/4) \cup [1/2, 3/4)), \qquad \{\omega\colon \xi_2(\omega) = 1\} = [1/4, 1/2) \cup [3/4, 1); \tag{10.12}$$

generally,

$$\{\omega\colon \xi_n(\omega) = 0\} = \bigcup_{i=0}^{2^{n-1}-1} [\frac{i}{2^{n-1}}, \frac{2i+1}{2^n}), \qquad \{\omega\colon \xi_n(\omega) = 1\} = \bigcup_{i=1}^{2^{n-1}} [\frac{2i-1}{2^n}, \frac{i}{2^{n-1}}). \tag{10.13}$$

We see that $\xi_i$ are indeed random variables (the inverse images $\xi_i^{-1}(C)$ are Borel sets: finite unions of intervals); and since the sums of the lengths of the intervals (10.11)–(10.13) are equal to 1/2, they all have the same distribution given by (10.9). The joint probability mass function of the first $n$ random variables is given by

$$\begin{aligned} &P\{\xi_1 = a_1, \xi_2 = a_2, ..., \xi_n = a_n\} \\ &\quad = P[\frac{a_1}{2} + \frac{a_2}{4} + ... + \frac{a_n}{2^n}, \frac{a_1}{2} + \frac{a_2}{4} + ... + \frac{a_n}{2^n} + \frac{1}{2^n}) = \frac{1}{2^n}, \quad a_i = 0 \text{ or } 1; \end{aligned} \tag{10.14}$$

this means that for every $n$ the random variables $\xi_1$, ..., $\xi_n$ are independent. Also are independent $\xi_{i_1}$, ..., $\xi_{i_k}$ with $i_j \neq i_l$ $(j \neq l)$ – this follows from the fact that $\xi_1$, ..., $\xi_{\max(i_1, ..., i_k)}$ are independent. And this means independence of the whole infinite sequence of $\xi_i$.

**Theorem 10.4.** *There exists a sequence of independent random variables all of them having the uniform distribution on the interval from* 0 *to* 1.

**Proof.** Take the random variables $\xi_1, ..., \xi_n, ...$ of Theorem 10.3, and take

$$\begin{aligned}
\eta_1 &= \frac{\xi_1}{2} + \frac{\xi_2}{4} + \frac{\xi_4}{8} + \frac{\xi_7}{16} + \frac{\xi_{11}}{32} + ..., \\
\eta_2 &= \frac{\xi_3}{2} + \frac{\xi_5}{4} + \frac{\xi_8}{8} + \frac{\xi_{12}}{16} + ..., \\
\eta_3 &= \frac{\xi_6}{2} + \frac{\xi_9}{4} + \frac{\xi_{13}}{8} + ..., \\
\eta_4 &= \frac{\xi_{10}}{2} + \frac{\xi_{14}}{4} + ..., \\
\eta_5 &= \frac{\xi_{15}}{2} + ..., \\
&\ldots .
\end{aligned} \tag{10.15}$$

(each of the random variables $\xi_i$ participating once and only once).

The series in (10.15) converge (they are dominated by the geometric series, and the sums are between 0 and 1); and their sums are $\mathcal{F}$-measurable because they are limits of measurable partial sums – see Lecture 4, around formula (4.39). And, by the way, we *feel* that the sum of finitely many (two is enough) random variables is again a random variable; but we haven't *proved* it. See Problem **10**.

Let us prove that each of the random variables $\eta_i$ has the uniform distribution on the interval from 0 to 1.

Do we include the ends? This does not matter because for $\eta_i = \sum_{j=1}^{\infty} \xi_{k_j}/2^j$ we have:

$$P\{\eta_i = 0\} = P\{\xi_{k_1} = 0, \xi_{k_2} = 0, ..., \xi_{k_n} = 0, ...\} = \prod_{j=1}^{\infty} P\{\xi_{k_j} = 0\} = \prod_{j=1}^{\infty} \frac{1}{2} = 0, \tag{10.16}$$

and the same for $P\{\eta_i = 1\}$ (I assume that the solution of Problem **22** is in the positive; if the statement is *dis*proved, we have to invent some other method to prove this minor point in our proof of Theorem 10.4).

Let us consider in the interval from 0 to 1 – OK, let's take it with its left end but without the right end: $[0, 1)$ – let us consider in it the class of subsets

$$\begin{aligned}
\mathcal{A} = \{\emptyset, [0,1), [0, 1/2), [1/2, 1), [0, 1/4), [1/4, 1/2), [1/2, 3/4), [3/4, 1),& \\
[0, 1/8), [1/8, 1/4), [1/4, 3/8), ... ...\}&:
\end{aligned} \tag{10.17}$$

the empty set, and all intervals of length $1/2^n$ with ends being multiples of $k/2^n$, for $n = 0, 1, 2, 3, ...$ .

This class of sets is clearly a semi-algebra in $[0, 1)$ (say, the complement of the interval $\big[k/2^n, (k+1)/2^n\big)$ is the disjoint union of $2^n - 1$ intervals of the same length belonging to $\mathcal{A}$); this semi-algebra generates the Borel $\sigma$-algebra of our interval: $\mathcal{B}_{[0,1)} = \sigma(\mathcal{A})$, so by the uniqueness part of Theorem 6.1 the values of $\mu_{\eta_j}$ on the sets belonging to $\mathcal{A}$ determines this measure on $\big([0, 1), \mathcal{B}_{[0,1)}\big)$ uniquely (and we saw that this distribution is completely concentrated on this interval: $\mu_{\eta_i}[0, 1) = 1$, $\mu_{\eta_i}\big(\mathbb{R}^1 \setminus [0, 1)\big) = 0$).

If $\eta_i = \sum_{j=1}^{\infty} \xi_{k_j}/2^j$, the event

$$\{\eta_i \in [\frac{a_1}{2} + \frac{a_2}{4} + ... + \frac{a_n}{2^n}, \frac{a_1}{2} + \frac{a_2}{4} + ... + \frac{a_n + 1}{2^n})\} \tag{10.18}$$

can be written as

$$\begin{aligned} &\{\xi_{k_1} = a_1, \xi_{k_2} = a_2, ..., \xi_{k_n} = a_n\} \\ &\qquad \setminus \{\xi_{k_1} = a_1, \xi_{k_2} = a_2, ..., \xi_{k_n} = a_n, \xi_{k_{n+1}} = \xi_{k_{n+2}} = ... = 1\}. \end{aligned} \tag{10.19}$$

The set that we are subtracting here has zero probability (its probability is equal to the product of infinitely many (1/2)'s); and the probability (10.18) is equal to

$$P\{\xi_{k_1} = a_1\} \cdot P\{\xi_{k_2} = a_2\} \cdot ... \cdot P\{\xi_{k_n} = a_n\} = \frac{1}{2^n}, \tag{10.21}$$

as it should be for the uniform distribution on [0, 1).

Now to independence.

For every $k$ and for every $n$ the partial sums with $n$ summands for the infinite series defining $\eta_1$, ..., $\eta_k$ are independent. Now, independence survives limit passage – solve Problem **23**.

**Theorem 10.5.** *Let $\mu_1$, $\mu_2$, ..., $\mu_n$, ... be a sequence of probability distributions on the real line. Then there exists a probability space and a sequence of independent random variables $\zeta_1$, $\zeta_2$, ..., $\zeta_n$, ... on this space such that each $\zeta_i$ has $\mu_i$ as its distribution.*

**Proof.** We take the distribution functions $F_i(x) = \mu_i(-\infty, x]$ and their pseudo-inverses $G_i$; and we apply $G_i$ to the independent random variables $\eta_i$ of the previous theorem:

$$\zeta_i = G_i(\eta_i). \tag{10.22}$$

By Theorem 9.5, these random variables are independent; and by the second proof of Theorem 5.5 they have $F_i$ as their distribution functions.

This is a very important result; in particular, it shows that when we in Probability Theory say: "If $\xi_1$, $\xi_2$, ..., $\xi_n$, ... is a sequence of independent random variables with such and such distributions, then this or that is true", our statement is not only true because there is *no* such sequence. It would be *very* unpleasant if we did not have Theorem 10.5.