Lecture 9. Independence, continued.

Theorem 9.1. *Let* ξ_1, ..., ξ_n *be independent real-valued random variables. If* $E|\xi_1|$, ..., $E|\xi_n| < \infty$, *then* $E|\xi_1 \dots \xi_n|$ *is also finite, and*

$$E(\xi_1 \dots \xi_n) = E\xi_1 \cdot \dots \cdot E\xi_n. \tag{9.1}$$

Proof. For shortness, let us write it for $n = 2$. We'll use Fubini's Theorem. The left-hand side of (9.1) is

$$\begin{aligned} &\iint\limits_{X_1 \times X_2} x_1 x_2 \ \mu_{\xi_1, \xi_2}(dx_1\, dx_2) = \iint\limits_{X_1 \times X_2} x_1\, x_2 \ \mu_{\xi_1}(dx_1)\, \mu_{\xi_2}(dx_2) \\ &= \int\limits_{X_2} \Big[\int\limits_{X_1} x_1\, x_2 \ \mu_{\xi_1}(dx_1) \Big] \ \mu_{\xi_2}(dx_2) = \int\limits_{X_2} \Big[x_2 \cdot \int\limits_{X_1} x_1 \ \mu_{\xi_1}(dx_1) \Big] \ \mu_{\xi_2}(dx_2) \\ &= \int\limits_{X_2} (x_2 \cdot E\xi_1) \ \mu_{\xi_2}(dx_2) = E\xi_1 \cdot \int\limits_{X_2} x_2 \ \mu_{\xi_2}(dx_2) = E\xi_1 \cdot E\xi_2. \end{aligned} \tag{9.2}$$

Theorem 9.2. *Let* B *be an event*; $\mathcal{A}$, *a semi-algebra of events. If* B *is independent from* $\mathcal{A}$ (which means from every event in $\mathcal{A}$), *it is also independent from the* σ*-algebra* $\sigma(\mathcal{A})$ *generated by* $\mathcal{A}$.

Proof. Let us consider two measures on $\sigma(\mathcal{A}) \subseteq \mathcal{F}$:

$$m_1(A) = P(B \cap A), \quad m_2(A) = P(B) \cdot P(A), \qquad A \in \sigma(\mathcal{A}). \tag{9.3}$$

These measures coincide on $\mathcal{A}$, therefore by Theorem 6.1 they coincide on $\sigma(\mathcal{A})$.

Theorem 9.3. *Let* $\mathcal{A}$, $\mathcal{B}$ *be two semi-algebras of events. If they are independent, so are the* σ*-algebras* $\sigma(\mathcal{A})$, $\sigma(\mathcal{B})$ *generated by them.*

Proof. According to the previous theorem, every event $B \in \mathcal{B}$ is independent from every event in $\sigma(\mathcal{A})$. Now we consider, for a fixed $A \in \sigma(\mathcal{A})$, two measures on $\sigma(\mathcal{B})$:

$$n_1(B) = P(A \cap B), \quad b_2(B) = P(B) \cdot P(A), \qquad B \in \sigma(\mathcal{B}); \tag{9.4}$$

and we repeat the reasoning in Theorem 9.2.

Of course, we can also formulate (and prove – though the proof, being the same as before, is omitted)

Theorem 9.4. *If the semi-algebras* $\mathcal{A}_1$, ..., $\mathcal{A}_n$ *are independent, so are the* σ*-algebras* $\sigma(\mathcal{A}_1)$, ..., $\sigma(\mathcal{A}_n)$ *generated by them.*

Theorem 9.5. *Let* ξ_1, ..., ξ_n *be independent random variables taking values in measurable spaces* $(X_k, \mathcal{X}_k)$. *Let* f_k, $k = 1, ..., n$, *be* $\mathcal{X}_k$*-measurable functions on* X_k. *Then the random variables* $f_1(\xi_1)$, ..., $f_n(\xi_n)$ *are independent.*

Proof. Since $\{f_k(\xi_k) \in C_k\} = \{\xi_k \in f_k^{-1}(C_k)\}$, and $f_k^{-1}(C_k)$ belongs to the σ-algebra $\mathcal{X}_k$, the statement follows from (8.10).

Let me give several criteria for independence of random variables.

Theorem 9.6. *Let* $\xi_1, \ldots, \xi_n$ *be discrete random variables; let* p_{ξ_i} *be their individual probability mass functions:* $p_{\xi_i}(x_i) = P\{\xi_i = x_i\}$, *and* $p_{\xi_1, \ldots, \xi_n}$ *their joint probability mass function:*

$$p_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = P\{\xi_1 = x_1, \ldots, \xi_n = x_n\}. \tag{9.5}$$

The random variables $\xi_1, \ldots, \xi_n$ *are independent if and only if*

$$p_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = p_{\xi_1}(x_1) \cdot \ldots \cdot p_{\xi_n}(x_n) \tag{9.6}$$

for all $x_1, \ldots, x_n$.

Proof. For our formulas to be shorter, let us consider the case of $n = 2$ (the only difference for larger n is longer formulas expressing the same).

Remember that independence means that

$$P\{\xi_1 \in C_1, \xi_2 \in C_2\} = P\{\xi_1 \in C_1\} \cdot P\{\xi_2 \in C_2\} \tag{9.7}$$

for C_1, C_2 belonging to the appropriate σ-algebras.

To deduce (9.6) from (9.7), we take one-point sets C_i: $C_1 = \{x_1\}$, $C_2 = \{x_2\}$; and here we are.

To work out the inverse implication, we use the fact that

$$\{\xi_1 \in C_1, \xi_2 \in C_2\} = \bigcup_{x_1 \in C_1,\, x_2 \in C_2} \{\xi_1 = x_1, \xi_2 = x_2\}, \tag{9.8}$$

the union being countable (a countable number of values for both – discrete! – random variables, and only a countable number of pairs), and the summands disjoint. So by countable additivity of probability

$$\begin{aligned} P\{\xi_1 \in C_1, \xi_2 \in C_2\} &= \sum_{x_1 \in C_1,\, x_2 \in C_2} P\{\xi_1 = x_1, \xi_2 = x_2\} = \sum_{x_1 \in C_1,\, x_2 \in C_2} p_{\xi_1, \xi_2}(x_1, x_2) \\ &= \sum_{x_1 \in C_1,\, x_2 \in C_2} p_{\xi_1}(x_1) \cdot p_{\xi_2}(x_2) = \sum_{x_1 \in C_1} p_{\xi_1}(x_1) \cdot \sum_{x_2 \in C_2} p_{\xi_2}(x_2) \end{aligned} \tag{9.9}$$

(the expression before the last is obtained from the last one by opening the parentheses).

Theorem 9.7. *Let* $\xi_1, \ldots, \xi_n$ *be continuous random variables* (or random vectors if you wish). *These random variables are independent if and only if their joint distribution is continuous with a density* (their joint density) $p_{\xi_1, \ldots, \xi_n}$ *satisfying the condition*

$$p_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = p_{\xi_1}(x_1) \cdot \ldots \cdot p_{\xi_n}(x_n) \tag{9.10}$$

for almost all $x_1, \ldots, x_n$ *with respect to the* n*-dimensional Lebesgue measure* (i. e., except a set of points $(x_1, \ldots, x_n)$ that has zero Lebesgue measure).

Of course it is not surprising that (9.10) has to be satisfied only *almost* everywhere: we know that there exist different *versions* of density, and each two differ from one another on a set of zero Lebesgue measure.

Proof of the theorem. Again let us write the proof for $n = 2$, and for one-dimensional ξ_1, ξ_2 rather than for random vectors (no difference in the proof except in the length of what we have to write).

That ξ_1, ξ_2 have a joint density (9.10) (or one differing from it only on a set of zero Lebesgue measure) means that

$$P\{(\xi_1,\, \xi_2) \in D\} = \iint\limits_D p_{\xi_1}(x_1) \cdot p_{\xi_2}(x_2)\, dx_1\, dx_2. \tag{9.11}$$

Let us show that from (9.11) follows (9.7). We have:

$$P\{\xi_1 \in C_1,\, \xi_2 \in C_2\} = \iint\limits_{C_1 \times C_2} p_{\xi_1}(x_1) \cdot p_{\xi_2}(x_2)\, dx_1\, dx_2 = \int\limits_{C_1} p_{\xi_1}(x_1)\, dx_1 \cdot \int\limits_{C_2} p_{\xi_2}(x_2)\, dx_2 \tag{9.12}$$

by Fubini's Theorem; and this is the right-hand side of (9.7).

The opposite: suppose (9.7) is satisfied. Let us consider the measure $\mu_{\xi_1,\,\xi_2}$ on $\mathcal{B}^2$ (the joint distribution of our random variables) and another measure ν on the same σ-algebra defined by

$$\nu(D) = \iint\limits_D p_{\xi_1}(x_1) \cdot p_{\xi_2}(x_2)\, dx_1\, dx_2. \tag{9.13}$$

Formula (9.7) means that these two measures coincide on the class $\mathcal{A}$ of all "rectangles" $C_1 \times C_2$, $C_i \in \mathcal{B}^1$.

The class $\mathcal{A}$ is a semi-algebra in $\mathbb{R}^2$. Indeed, $\mathbb{R}^2$ belongs to this class of all "rectangles":

$$\mathbb{R}^2 = \mathbb{R}^1 \times \mathbb{R}^1, \qquad \mathbb{R}^1 \in \mathcal{B}^1, \qquad \mathbb{R}^2 \in \mathcal{A}; \tag{9.14}$$

the intersection of two "rectangles" is again one:

$$A = C_1 \times C_2 \in \mathcal{A},\ B = D_1 \times D_2 \in \mathcal{A} \ \Rightarrow\ A \cap B = (C_1 \cap D_1) \times (C_2 \cap D_2) \in \mathcal{A}; \tag{9.15}$$

and the complement of a "rectangle" can be represented as a union of a finite number of disjoint "rectangles":

$$(C_1 \times C_2)^c = (C_1^c \times C_2) \cup (C_1 \times C_2^c) \cup (C_1^c \times C_2^c) \tag{9.16}$$

(make a picture), where $C_1^c \times C_2$, $C_1 \times C_2^c$, $C_1^c \times C_2^c \in \mathcal{A}$.

By the (uniqueness part of) the Theorem 6.1 about extension of measures, the (finite) measures $\mu_{\xi_1,\,\xi_2}$ and ν coincide on the σ-algebra $\sigma(\mathcal{A})$, which is nothing but $\mathcal{B}^2$.

So (9.11) is satisfied.

Theorem 9.8. *Real-valued random variables* $\xi_1, ..., \xi_n$ *are independent if and only if their joint distribution function is equal to the product of their individual one-dimensional distribution functions*:

$$F_{\xi_1, ..., \xi_n}(x_1, ..., x_n) = F_{\xi_1}(x_1) \cdot ... \cdot F_{\xi_n}(x_n), \qquad x_1, ..., x_n \in (-\infty, \infty). \tag{9.17}$$

Proof. The "only if" part: we take in the equality

$$P\{\xi_1 \in C_1, ..., \xi_n \in C_n\} = P\{\xi_1 \in C_1\} \cdot ... \cdot P\{\xi_n \in C_n\} \tag{9.18}$$

$C_1 = (-\infty, x_1], ..., C_n = (-\infty, x_n]$.

The "if" part – to write less, let it again be $n = 2$. Let (9.17) be satisfied. We mentioned that

$$\begin{aligned} &P\{a_1 < \xi_1 \le b_1, a_2 < \xi_2 \le b_2\} \\ &\qquad = F_{\xi_1, \xi_2}(b_1, b_2) - F_{\xi_1, \xi_2}(a_1, b_2) - F_{\xi_1, \xi_2}(b_1, a_2) + F_{\xi_1, \xi_2}(a_1, a_2). \end{aligned} \tag{9.19}$$

According to (9.17), this is equal to

$$\begin{aligned} &F_{\xi_1}(b_1) \cdot F_{\xi_2}(b_2) - F_{\xi_1}(a_1) \cdot F_{\xi_2}(b_2) - F_{\xi_1}(b_1) \cdot F_{\xi_2}(a_2) + F_{\xi_1}(a_1) \cdot F_{\xi_2}(a_2) \\ &\qquad = [F_{\xi_1}(b_1) - F_{\xi_1}(a_1)] \cdot [F_{\xi_2}(b_2) - F_{\xi_2}(a_2)]. \end{aligned} \tag{9.20}$$

So we have:

$$\mu_{\xi_1, \xi_2}\big((a_1, b_1] \times (a_2, b_2]\big) = \mu_{\xi_1}(a_1, b_1] \cdot \mu_{\xi_2}(a_2, b_2]: \tag{9.21}$$

the two measures μ_{ξ_1, ξ_2} and $\mu_{\xi_1} \times \mu_{\xi_2}$ coincide on all rectangles. Since all rectangles (finite and infinite) form a semi-algebra in $\mathbb{R}^2$, and since the Borel σ-algebra $\mathcal{B}^2$ is generated by all rectangles, by the uniqueness part of the extension theorem these two measures coincide on $\mathcal{B}^2$.

We can reformulate Theorems 9.6, 9.7, 9.8 differently:

Theorem 9.6′. *Let* $p_{\xi_1, ..., \xi_n}(x_1, ..., x_n)$ *be the joint probability mass function of* n *random variables. The random variables* $\xi_1, ..., \xi_n$ *are independent if and only if this joint probability mass function factorizes into* n *factors of which the first one depends on* x_1 *only, the second on* $x_2, ...,$ *the last one on* x_n:

$$p_{\xi_1, ..., \xi_n}(x_1, ..., x_n) = f_1(x_1) \cdot ... \cdot f_n(x_n). \tag{9.22}$$

We do not presuppose here that $f_i(x_i) = p_{\xi_i}(x_i)$ (and it may not be so).

Theorem 9.7′. *Let* $p_{\xi_1, ..., \xi_n}(x_1, ..., x_n)$ *be the joint probability density of* n *random variables* $\xi_1, ..., \xi_n$. *These random variables are independent if and only if this joint density factorizes* almost everywhere with respect to the Lebesgue measure *into* n *factors of which the first one depends on* x_1 *only, the second on* $x_2, ...,$ *the last one on* x_n – which is expressed by the same formula (9.22) (but with a different meaning).

Theorem 9.8′. *Let* $F_{\xi_1, ..., \xi_n}(x_1, ..., x_n)$ *be the joint distribution function of* $\xi_1, ..., \xi_n$. *These random variables are independent if and only if this joint distribution* *function factorizes into* n *factors of which the first one depends on* x_1 *only, the second on* x_2*, ..., the last one on* x_n*:*

$$F_{\xi_1,\,...,\,\xi_n}(x_1,\,...,\,x_n) = G_1(x_1)\cdot ...\cdot G_n(x_n) \tag{9.23}$$

(again we do not presuppose anything).

The **proof** is similar for all three theorems; let us take on the second – in the case of $n = 2$ for less writing, and real-valued random variables ξ_1, ξ_2 rather than random vectors.

By Problem $\boxed{\mathbf{20}}$, ξ_1, ξ_2 necessarily are continuous random variables with densities

$$p_{\xi_1}(x_1) = \int\limits_{-\infty}^{\infty} p_{\xi_1,\,\xi_2}(x_1,\,x_2)\;dx_2, \qquad p_{\xi_2}(x_2) = \int\limits_{-\infty}^{\infty} p_{\xi_1,\,\xi_2}(x_1,\,x_2)\;dx_1. \tag{9.24}$$

If these random variables are independent, we can apply Theorem 9.7 and get that (9.22) is satisfied almost everywhere with respect to the two-dimensional Lebesgue measure λ_2 with $f_i(x_i) = p_{\xi_i}(x_i)$.

If (9.22) is satisfied almost everywhere, we get by (9.24) (almost everywhere with respect to the one-dimensional Lebesgue measure):

$$\begin{aligned} p_{\xi_1}(x_1) &= \int\limits_{-\infty}^{\infty} f_1(x_1)\cdot f_2(x_2)\;dx_2 = \Bigl(\int\limits_{-\infty}^{\infty} f_2(x_2)\;dx_2\Bigr)\cdot f_1(x_1),\\ p_{\xi_2}(x_2) &= \int\limits_{-\infty}^{\infty} f_1(x_1)\cdot f_2(x_2)\;dx_1 = \Bigl(\int\limits_{-\infty}^{\infty} f_1(x_1)\;dx_1\Bigr)\cdot f_2(x_2), \end{aligned} \tag{9.25}$$

and almost everywhere

$$p_{\xi_1,\,\xi_2}(x_1,\,x_2) = f_1(x_1)\cdot f_2(x_2) = \Bigl(\int\limits_{-\infty}^{\infty} f_2(x_2)\;dx_2\Bigr)\cdot f_1(x_1)\cdot\Bigl(\int\limits_{-\infty}^{\infty} f_1(x_1)\;dx_1\Bigr)\cdot f_2(x_2), \tag{9.26}$$

because the constant

$$\Bigl(\int\limits_{-\infty}^{\infty} f_2(x_2)\;dx_2\Bigr)\cdot\Bigl(\int\limits_{-\infty}^{\infty} f_1(x_1)\;dx_1\Bigr) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} p_{\xi_1,\,\xi_2}(x_1,\,x_2)\;dx_1\,dx_2. = 1 \tag{9.27}$$

In the n-dimensional case it is

$$\begin{aligned} &\Bigl(\int f_2(x_2)\;dx_2\cdot ...\cdot \int f_n(x_n)\;dx_n\Bigr)\cdot\Bigl(\int f_1(x_1)\;dx_1\cdot \int f_3(x_3)\;dx_3\cdot ...\cdot\int f_n(x_n)\;dx_n\Bigr)\times\\ &\qquad\times ...\cdot\Bigl(\int f_1(x_1)\;dx_1\cdot ...\cdot\int f_{n-1}(x_{n-1})\;dx_{n-1}\Bigr)\\ &\qquad\qquad = \Bigl(\int\cdots\int\limits_{\mathbb{R}^n} p_{\xi_1,\,...,\,\xi_n}(x_1,...,x_n)\;dx_1\,...\,dx_n\Bigr)^{n-1} = 1. \end{aligned} \tag{9.28}$$

In Theorem 9.8′, we get the one-dimensional distribution function $F_{\xi_i}(x_i)$ as the limit of the joint n-dimensional distribution function as all variables x_j except for $j = i$ go to $+\infty$.