Probability and Mathematical Statistics

Miskolc, 2025.

Dr. Tamás Glavosits

Lecture 7

Random Vector Variables (Two-Dimensional Case)

1. Random Vector Variables

Two Interpretations of Random Vector Variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a Kolmogorov-type probability field.

1. If $\xi_1 : \Omega \to \mathbb{R}$, $\xi_2 : \Omega \to \mathbb{R}$ are random variables, then the function $(\xi_1, \xi_2) : \Omega \to \mathbb{R}^2$ is called a random vector variable.
2. The function $(\xi_1, \xi_2) : \Omega \to \mathbb{R}^2$ is called a random vector variable if, for all $x, y \in \mathbb{R}$,

$$(\xi_1 < x, \xi_2 < y) := \{\omega \in \mathbb{R} : \xi_1(\omega) < x \text{ and } \xi_2(\omega) < y\} \in \mathcal{F}.$$

The Two Interpretations Are Equivalent

1. $\Longrightarrow$ 2.: Let $x, y \in \mathbb{R}$ be arbitrary real numbers. Then, since ξ_1 and ξ_2 are random variables, we have $(\xi_1 < x) \in \mathcal{F}$ and $(\xi_2 < y) \in \mathcal{F}$. However, in this case

$$(\xi_1 < x, \xi_2 < y) = (\xi_1 < x) \cap (\xi_2 < y) \in \mathcal{F}.$$

2. $\Longrightarrow$ 1. Suppose that (ξ_1, ξ_2) is a random vector variable. We will show that then ξ_1 (the first component) is a random variable. Since for every $n \in \mathbb{Z}_+$ we have $(\xi_1 < x, \xi_2 < n) \in \mathcal{F}$, it follows that $\bigcup_{n=1}^{\infty} (\xi_1 < x, \xi_2 < n) \in \mathcal{F}$. Moreover,

$$\bigcup_{n=1}^{\infty} (\xi_1 < x, \xi_2 < n) = (\xi_1 < x) \cap \bigcup_{n=1}^{\infty} (\xi_2 < n) =$$
$$= (\xi_1 < x) \cap \Omega = (\xi_1 < x)\,.$$

from which we obtain that $(\xi_1 < x) \in \mathcal{F}$ for all $x \in \mathbb{R}$. In a similar way, it can be seen that ξ_2 (that is, the second component) is also a random variable.

An Important Property of Joint Distribution

Theorem

If (ξ_1, ξ_2) *is a random vector variable,* $(x, y) \in \mathbb{R}^2$*, then*

$$(\xi_1 = x, \xi_2 = y) = \{\omega \in \Omega | \xi_1(\omega) = x \text{ and } \xi_2(\omega) = y\} \in \mathcal{F}.$$

2. The Joint Distribution Function and Its Properties

Joint Distribution Function

Definition

If (ξ_1, ξ_2) is a random vector variable, then the function $\mathbb{F} : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\mathbb{F}(x, y) := \mathbb{P}\left(\xi_1 < x, \xi_2 < y\right) \qquad ((x, y) \in \mathbb{R})$$

is called the **joint distribution function** of the random vector variable (ξ_1, ξ_2).

Properties of the Joint Distribution Function

Theorem

Let (ξ_1, ξ_2) be a random vector variable with joint distribution function $\mathbb{F} : \mathbb{R}^2 \to \mathbb{R}$. Then $\mathbb{F}$ has the following properties:

1. *It is increasing in both of its variables and left-continuous.*
2. *(Instead of monotonic increase): If $a_1 < b_1$ and $a_2 < b_2$, then*

$$\mathbb{F}(b_1, b_2) - \mathbb{F}(a_1, b_2) - \mathbb{F}(b_1, a_2) + \mathbb{F}(a_1, a_2) \geqslant 0;$$

3. *Limits are statisfied:*

$$\lim_{x_1 \to -\infty} \mathbb{F}(x_1, x_2) = 0, \quad \lim_{x_2 \to -\infty} \mathbb{F}(x_1, x_2) = 0, \quad \lim_{\substack{x_1 \to +\infty \\ x_2 \to +\infty}} \mathbb{F}(x_1, x_2) = 1.$$

1. First, we show that the joint distribution function is **monotonically increasing** in both of its variables. This follows easily, for example, from the monotonicity of probability. Let $x_1 < x_2$ and $y \in \mathbb{R}$ be arbitrary. Then $(\xi_1 < x_1, \xi_2 < y) \subseteq (\xi_1 < x_2, \xi_2 < y)$, from which we get

$$\mathbb{F}(x_1, y) = \mathbb{P}(\xi_1 < x_1, \xi_2 < y) \leqslant \mathbb{P}(\xi_1 < x_2, \xi_2 < y) = \mathbb{F}(x_2, y).$$

Now we show that the joint distribution function is **left-continuous** in the first variable, that is,
$\mathbb{F}(x_0 - 0, y_0) = \mathbb{F}(x_0, y_0)$ for all real numbers x_0, y_0.
Since the joint distribution function is monotonically increasing in the first variable, the limits on both sides exist. Using the continuity of probability, we get

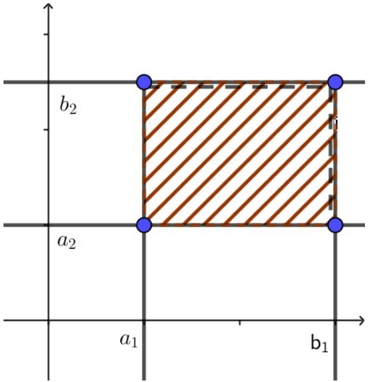
$$\mathbb{F}(x_0 - 0, y_0) = \lim_{n\to\infty} \mathbb{F}\left(x_0 - \frac{1}{n}, y_0\right) = \lim_{n\to\infty} \mathbb{P}\left(\left(\xi_1 < x_0 - \frac{1}{n}\right) \cap (\xi_2 < y_0)\right) =$$
$$= \mathbb{P}\left(\bigcup_{n=1}^{\infty}\left(\xi_1 < x_0 - \frac{1}{n}\right) \cap (\xi_2 < y_0)\right) = \mathbb{P}(\xi_1 < x_0, \xi_2 < y_0) =$$
$$= \mathbb{F}(x_0, y_0).$$

2. Proof

2. Let $a_1 < b_1$ and $a_2 < b_2$. Then

$$\mathbb{P}\left((\xi_1, \xi_2) \in [a_1, b_1[\times[a_2, b_2[\right) =$$
$$= \mathbb{F}(b_1, b_2) - \mathbb{F}(a_1, b_2) - \mathbb{F}(b_1, a_2) + \mathbb{F}(a_1, a_2)$$

from which the statement follows using the non-negativity of probability.



3. Proof

3. Analogously to the reasoning in point 2, we obtain

$$\lim_{x_1 \to -\infty} \mathbb{F}(x_1, x_2) = \mathbb{P}(\emptyset \cap (\xi_2 < x_2)) = \mathbb{P}(\emptyset) = 0,$$

$$\lim_{x_2 \to -\infty} \mathbb{F}(x_1, x_2) = \mathbb{P}((\xi_2 < x_2) \cap \emptyset) = \mathbb{P}(\emptyset) = 0,$$

$$\lim_{\substack{x_1 \to +\infty \\ x_2 \to +\infty}} \mathbb{F}(x_1, x_2) = \mathbb{P}((\xi_1 \in \mathbb{R}) \cap (\xi_2 \in \mathbb{R})) = \mathbb{P}(\Omega) = 1.$$

3. Discrete and Absolutely Continuous Random Vectors

Discrete Random Vectors

Definition

A (ξ_1, ξ_2) **random vector is called discrete** if there exists countable elements $(x_i, y_j) \in \mathbb{R}^2$ $(i = 1, 2, \dots, j = 1, 2, \dots)$ such that

$$\mathbb{P}\big((\xi_1, \xi_2) \in \{(x_i, y_j) \mid i = 1, 2, \dots, j = 1, 2, \dots\}\big) = 1,$$

(that is, the range of (ξ_1, ξ_2) is countable).

- The set $((x_i, y_j))$ is called the **support of the discrete random vector** (ξ_1, ξ_2),
- The sequence (p_{ij}) with $p_{ij} := \mathbb{P}(\xi_1 = x_i, \xi_2 = y_j)$ is called the **distribution of the discrete random vector** (ξ_1, ξ_2).

Properties of the Distribution of a Discrete Random Vector

Theorem

If (ξ_1, ξ_2) *is a random vector, then*

- $p_{ij} \geqslant 0$ *for all* $i = 1, 2, \ldots$ *and* $j = 1, 2, \ldots;$
- $\sum_i \sum_j p_{ij} = 1.$

Proof

Proof.

- Non-negativity follows directly from the fact that probabilities are non-negative.
- By σ-additivity (or finite additivity) of probability, we have

$$\sum_i \sum_j p_{ij} = \sum_i \sum_j \mathbb{P}\left(\xi_1 = x_i, \xi_2 = y_j\right) =$$

$$= \mathbb{P}\left(\bigcup_i \bigcup_j (\xi_1 = x_i, \xi_2 = y_j)\right) = \mathbb{P}(\Omega) = 1.$$

□

Absolutely Continuous Random Vectors

Definition

A random vector (ξ_1, ξ_2) is called **absolutely continuous** if there exists a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\mathbb{F}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, du\, dv \qquad \left((x, y) \in \mathbb{R}^2\right).$$

The function f is called the **joint density function** of the random vector (ξ_1, ξ_2).

Properties of the Density Function of a Random Vector

Theorem

If (ξ_1, ξ_2) is an absolutely continuous random vector with density function f, then

- $f(x, y) \geqslant 0$ *for all* $(x, y) \in \mathbb{R}^2$
  (that is, f can only take negative values on a set of Lebesgue measure zero);
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, dx\, dy = 1.$

4. Marginal Distributions of a Random Vector

Marginal Distributions

Definition

If (ξ_1, ξ_2) is a random vector, then the random variables ξ_1 and ξ_2 are called the **marginal distributions** of the vector (ξ_1, ξ_2).

Marginal Distribution Functions

Theorem

Let (ξ_1, ξ_2) be a random vector.

1. *Denote by $\mathbb{F}$ the joint distribution function of (ξ_1, ξ_2), and let $\mathbb{F}_1$ and $\mathbb{F}_2$ be the distribution functions of the marginals ξ_1 and ξ_2, respectively. Then*

$$\mathbb{F}_1(x) = \lim_{x_2 \to +\infty} \mathbb{F}(x, x_2), \qquad \mathbb{F}_2(y) = \lim_{x_1 \to +\infty} \mathbb{F}(x_1, y).$$

Proof.

1. Since $\mathbb{F}(\cdot,\cdot)$ is monotonically increasing in both variables, the limits
$$\lim_{x_1\to+\infty}\mathbb{F}(x_1,y), \qquad \lim_{x_2\to+\infty}\mathbb{F}(x,x_2)$$
exist and can be easily determined. By continuity of probability, we have
$$\begin{aligned}\lim_{x_2\to+\infty}\mathbb{F}(x,x_2)=\lim_{n\to+\infty}\mathbb{F}(x,n)&=\lim_{n\to+\infty}\mathbb{P}(\xi_1<x,\xi_2<n)\\ &=\mathbb{P}\Big((\xi_1<x)\cap\bigcup_{n=1}^{\infty}(\xi_2<n)\Big)\\ &=\mathbb{P}((\xi_1<x)\cap\Omega)=\mathbb{P}(\xi_1<x)=\mathbb{F}_1(x).\end{aligned}$$
The corresponding result for ξ_2 can be shown analogously.

□

Marginal Distributions in the Discrete Case

Theorem

2. *If* (ξ_1, ξ_2) *is a discrete random vector with values* (x_i, y_j) *and joint probabilities* $p_{ij} := \mathbb{P}(\xi_1 = x_i, \xi_2 = y_j)$*, then* ξ_1 *is a discrete random variable with values* $x_1, x_2, \ldots$ *and* ξ_2 *is a discrete random variable with values* $y_1, y_2, \ldots$. *Their distributions are*

$$p_{i\cdot} = \mathbb{P}(\xi_1 = x_i) = \sum_j p_{ij}, \qquad p_{\cdot j} = \mathbb{P}(\xi_2 = y_j) = \sum_i p_{ij},$$

for all $i = 1, 2, \ldots$ *and* $j = 1, 2, \ldots$.

Proof

Proof.

2. By σ-additivity of probability, we have

$$\begin{aligned}
p_{i\cdot} &= \mathbb{P}(\xi_1 = x_i) = \mathbb{P}\left((\xi_1 = x_i) \cap \Omega\right) \\
&= \mathbb{P}\left((\xi_1 = x_i) \cap \bigcup_j (\xi_2 = y_j)\right) \\
&= \mathbb{P}\left(\bigcup_j \left((\xi_1 = x_i) \cap (\xi_2 = y_j)\right)\right) \\
&= \sum_j \mathbb{P}\left(\xi_1 = x_i, \xi_2 = y_j\right) = \sum_j p_{ij}.
\end{aligned}$$

The analogous statement for ξ_2 follows similarly.

□

Marginal Densities

Theorem

3. *If* (ξ_1, ξ_2) *is an absolutely continuous random vector with joint density* $f : \mathbb{R}^2 \to \mathbb{R}$*, then* ξ_1 *and* ξ_2 *are absolutely continuous random variables with densities* f_1 *and* f_2 *given by*

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy, \qquad f_2(y) = \int_{-\infty}^{+\infty} f(x, y)\, dx,$$

for all $x \in \mathbb{R}$ *and* $y \in \mathbb{R}$*.*

Proof

Proof.

3. This follows easily from the previous theorem, where we derived the marginal distribution function from the joint distribution function using a limit:

$$\int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} f(u, v)\, dv \right) du = \lim_{y \to \infty} \mathbb{F}(x, y) = \mathbb{F}_1(x),$$

which implies

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, v)\, dv.$$

The statement for f_2 can be prooved analogously.

□

5. The Concept and Properties of Covariance

Covariance

We have already encountered the concept of covariance between two random variables ξ_1 and ξ_2, which is defined by

- $\operatorname{cov}(\xi_1, \xi_2) := \mathbb{E}\left[(\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))\right]$, and
- it can also be computed by $\operatorname{cov}(\xi_1, \xi_2) = \mathbb{E}(\xi_1 \xi_2) - \mathbb{E}(\xi_1)\mathbb{E}(\xi_2)$.

The computation of Expected Value $\mathbb{E}(\xi_1\xi_2)$

The expected value $\mathbb{E}(\xi_1\xi_2)$ in the above formula can be easily calculated using the transformation rule for random vectors:

- **In the discrete case:**

$$\mathbb{E}(\xi_1\xi_2) = \sum_i \sum_j x_i y_j p_{ij},$$

- **In the absolutely continuous case:**

$$\mathbb{E}(\xi_1\xi_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy\, f(x, y)\, dy\, dx.$$

Calculating the Covariance

The formula for covariance follows directly from the additivity and homogeneity (that is, linearity) of the expectation:

$$\begin{aligned}\operatorname{cov}(\xi_1, \xi_2) &= \mathbb{E}\big[(\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))\big] \\ &= \mathbb{E}\big[\xi_1\xi_2 - \xi_1\mathbb{E}(\xi_2) - \mathbb{E}(\xi_1)\xi_2 + \mathbb{E}(\xi_1)\mathbb{E}(\xi_2)\big] \\ &= \mathbb{E}(\xi_1\xi_2) - 2\mathbb{E}(\xi_1)\mathbb{E}(\xi_2) + \mathbb{E}(\xi_1)\mathbb{E}(\xi_2) \\ &= \mathbb{E}(\xi_1\xi_2) - \mathbb{E}(\xi_1)\mathbb{E}(\xi_2).\end{aligned}$$

Properties of Covariance

Theorem

Covariance is a symmetric bilinear form, and the associated quadratic form is positive semidefinite. More explicitly, this means:

1. ***Symmetric***,
2. ***Bilinear***,
   - ***Homogeneous in the first variable***,
   - ***Additive in the first variable***,
3. ***Positive semidefinite***,
4. ***Translation-invariant***,
5. ***Cauchy-Bunyakovsky-Schwarz inequality:*** *For random variables* ξ_1 *and* ξ_2 *with finite variances,*

$$|\mathrm{cov}(\xi_1, \xi_2)| \leq \mathbb{D}(\xi_1)\,\mathbb{D}(\xi_2).$$

Covariance is symmetric

Theorem

1. ***Symmetric***, *that is,* $\operatorname{cov}(\xi_1, \xi_2) = \operatorname{cov}(\xi_2, \xi_1)$.

Proof.

1. The symmetry of cov is evident from the definition, since

$$\begin{aligned}\operatorname{cov}(\xi_1, \xi_2) &= \mathbb{E}((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))) = \\ &= \mathbb{E}((\xi_2 - \mathbb{E}(\xi_2))(\xi_1 - \mathbb{E}(\xi_1))) = \operatorname{cov}(\xi_2, \xi_1).\end{aligned}$$

□

Covariance is bilinear

Theorem

2. ***Bilinear***, *that is, it is homogeneous and additive in both variables; due to symmetry, it is suffitient to justify the homogeneity and the additivity in the first variable.*
   - ***Homogeneity in the first variable:*** $\mathrm{cov}(\lambda\xi_1, \xi_2) = \lambda\mathrm{cov}(\xi_1, \xi_2)$.
   - ***Additivity in the first variable:*** $\mathrm{cov}(\xi_1 + \xi_2, \eta) = \mathrm{cov}(\xi_1, \eta) + \mathrm{cov}(\xi_2, \eta)$.

Proof of the bilinearity of covariance

Proof.

2. We show that it is additive and homogeneous in both of its variables.
   - It is additive in the first variable, since by the additivity of the expectation we obtain that

$$\begin{aligned}\operatorname{cov}(\xi_1+\xi_2,\eta) &= \mathbb{E}\left((\xi_1+\xi_2-\mathbb{E}(\xi_1+\xi_2))(\eta-\mathbb{E}(\eta))\right) = \\ &= \mathbb{E}\left((\xi_1-\mathbb{E}(\xi_1))(\eta-\mathbb{E}(\eta))+(\xi_2-\mathbb{E}(\xi_2))(\eta-\mathbb{E}(\eta))\right) = \\ &= \mathbb{E}\left((\xi_1-\mathbb{E}(\xi_1))(\eta-\mathbb{E}(\eta))\right)+\mathbb{E}\left((\xi_2-\mathbb{E}(\xi_2))(\eta-\mathbb{E}(\eta))\right) = \\ &= \operatorname{cov}(\xi_1,\eta)+\operatorname{cov}(\xi_2,\eta).\end{aligned}$$

   - It is homogeneous in the first variable, since by the homogeneity of the expectation we get

$$\begin{aligned}\operatorname{cov}(\lambda\xi_1,\xi_2) &= \mathbb{E}\left((\lambda\xi_1-\mathbb{E}(\lambda\xi_1))(\xi_2-\mathbb{E}(\xi_2))\right) = \\ &= \mathbb{E}\left(\lambda(\xi_1-\mathbb{E}(\xi_1))(\xi_2-\mathbb{E}(\xi_2))\right) = \\ &= \lambda\mathbb{E}\left((\xi_1-\mathbb{E}(\xi_1))(\xi_2-\mathbb{E}(\xi_2))\right) = \lambda\operatorname{cov}(\xi_1,\xi_2).\end{aligned}$$

□

Covariance is positive semidefinite

Theorem

3. ***Positive semidefinite***: *the so-called quadratic form derived from it is positive semidefinite, that is,* $\operatorname{cov}(\xi, \xi) = \mathbb{D}^2(\xi) \geqslant 0$.

Proof.

3. (Positive semidefinite) Since the quadratic form derived from the covariance is exactly the variance, which is positive semidefinite, the statement is evident.

□

Covariance is translation invariant

Theorem

4. ***Translation invariant***, *that is, if* ξ_1, ξ_2 *are random variables and* c_1, c_2 *are real numbers, then*

$$\operatorname{cov}(\xi_1 + c_1, \xi_2 + c_2) = \operatorname{cov}(\xi_1, \xi_2).$$

Proof.

4. The proof follows easily from the property of the expectation with respect to translation:

$$\begin{aligned}
&\operatorname{cov}(\xi_1 + c_1, \xi_2 + c_2) = \\
&\quad = \mathbb{E}\left(((\xi_1 + c_1) - \mathbb{E}(\xi_1 + c_1))\left((\xi_2 + c_2) - \mathbb{E}(\xi_2 + c_2)\right)\right) = \\
&\quad = \mathbb{E}\left((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))\right) = \operatorname{cov}(\xi_1, \xi_2).
\end{aligned}$$

□

Cauchy–Bunyakovsky–Schwarz inequality

Theorem

5. *If ξ_1 and ξ_2 are random variables whose variances exist, then*

$$|\mathrm{cov}(\xi_1, \xi_2)| \leqslant \mathbb{D}(\xi_1)\mathbb{D}(\xi_2).$$

Proof.

5. (Cauchy–Bunyakovsky–Schwarz inequality.) If $\mathbb{D}^2(\xi_1) = 0$ and ξ_2 is arbitrary, then there exists $c \in \mathbb{R}$ such that $\mathbb{P}(\xi_1 = c) = 1$. By the translation invariance and additivity of the covariance in its first argument, we obtain

$$\mathrm{cov}(\xi_1, \xi_2) = \mathrm{cov}(c, \xi_2) = \mathrm{cov}(0, \xi_2) = 0,$$
$$\mathbb{D}(\xi_1)\mathbb{D}(\xi_2) = 0 \cdot \mathbb{D}(\xi_2) = 0;$$

thus both sides of the Cauchy–Bunyakovsky–Schwarz inequality are zero, that is, equality holds.

□

If ξ_1 is arbitrary and $\mathbb{D}^2(\xi_2) = 0$, then, analogously to the previous way, we obtain that $\operatorname{cov}(\xi_1, \xi_2) = 0$ and $\mathbb{D}(\xi_1)\mathbb{D}(\xi_2) = 0$, hence the Cauchy–Bunyakovsky–Schwarz inequality again holds with equality. Now assume that $\mathbb{D}^2(\xi_1) \neq 0$ and $\mathbb{D}^2(\xi_2) \neq 0$. Define the real quadratic polynomial $g(\lambda) := \mathbb{D}^2(\lambda\xi_1 + \xi_2)$. Then, using the properties of the variance, we have

$$0 \leqslant g(\lambda) = \underbrace{\mathbb{D}^2(\xi_1)}_{a}\lambda^2 + \underbrace{2\operatorname{cov}(\xi_1, \xi_2)}_{b}\lambda + \underbrace{\mathbb{D}^2(\xi_2)}_{c},$$

which shows that the polynomial $g(\lambda)$ either has one double real root or no real roots at all. Therefore, the discriminant D of $g(\lambda)$ is nonpositive, that is

$$D = b^2 - 4ac = (2\operatorname{cov}(\xi_1, \xi_2))^2 - 4\mathbb{D}^2(\xi_1)\mathbb{D}^2(\xi_2) \leqslant 0,$$

whence we obtain the Cauchy–Bunyakovsky–Schwarz inequality.

6. The relationship between uncorrelatedness and independence

Independent random variables

Definition

The random variables ξ_1 and ξ_2 are called **independent** if for all $x, y \in \mathbb{R}$,

$$\mathbb{P}(\xi_1 < x, \xi_2 < y) = \mathbb{P}(\xi_1 < x)\mathbb{P}(\xi_2 < y).$$

Characterizations of independence of random variables

Theorem

The random variables ξ_1 and ξ_2 are independent if and only if one of the following equivalent conditions holds:

1. ***Expressed by distribution functions:*** $\mathbb{F}(x, y) = \mathbb{F}_1(x)\mathbb{F}_2(y)$ *for all* $x, y \in \mathbb{R}$, *their joint distribution function equals the product of their marginal distribution functions;*
2. ***In the discrete case:*** $p_{ij} = p_{i\cdot}p_{\cdot j}$ *for all* $i = 1, 2, \ldots,$ $j = 1, 2, \ldots,$ *that is, the joint probability mass function equals the product of the marginals;*
3. ***In the absolutely continuous case:*** $f(x, y) = f_1(x)f_2(y)$ *for all* $x, y \in \mathbb{R}$, *that is, the joint density function equals the product of the marginal density functions.*

Proof

1. Let $x, y \in \mathbb{R}$ be arbitrary. Then

$$\mathbb{F}(x, y) = \mathbb{P}(\xi_1 < x, \xi_2 < y) = \mathbb{P}(\xi_1 < x)\mathbb{P}(\xi_2 < y) = \mathbb{F}_1(x)\mathbb{F}_2(y).$$

2. We show that if ξ_1 and ξ_2 are independent discrete random variables, then $p_{ij} = p_{i\cdot}\, p_{\cdot j}$.

$$p_{ij} = \mathbb{P}(\xi_1 = x_i, \xi_2 = y_j) = \lim_{n\to\infty}\lim_{m\to\infty} \mathbb{P}\left(\xi_1 \in \left]x_i - \frac{1}{n}, x_i\right], \xi_2 \in \left]y_j - \frac{1}{m}, y_j\right]\right) =$$

$$= \lim_{n\to\infty} \mathbb{P}\left(\xi_1 \in \left]x_i - \frac{1}{n}, x_i\right]\right) \lim_{m\to\infty} \mathbb{P}\left(\xi_2 \in \left]y_j - \frac{1}{m}, y_j\right]\right) =$$

$$= \mathbb{P}(\xi_1 = x_i)\mathbb{P}(\xi_2 = y_j) = p_{i\cdot}\, p_{\cdot j}.$$

Now we show the converse statement: if $p_{ij} = p_{i\cdot}\, p_{\cdot j}$, then ξ_1 and ξ_2 are independent.

$$\mathbb{F}(x, y) = \mathbb{P}(\xi_1 < x, \xi_2 < y_j) = \sum_{x_i < x,\ y_j < y} p_{ij} = \sum_{x_i < x} p_{i\cdot} \sum_{y_j < y} p_{\cdot j} =$$

$$= \mathbb{P}(\xi_1 < x)\mathbb{P}(\xi_2 < y_j) = \mathbb{F}_1(x)\mathbb{F}_2(y).$$

3. We show that if ξ_1 and ξ_2 are independent absolutely continuous random variables, then $f(x,y) = f_1(x) f_2(y)$. Let $x_0, y_0 \in \mathbb{R}$. Then

$$\mathbb{F}(x_0, y_0) = \mathbb{F}_1(x_0)\mathbb{F}_2(y_0) = \left(\int_{-\infty}^{x_0} f_1(x)\, dx\right)\left(\int_{-\infty}^{y_0} f_2(y)\, dy\right) =$$
$$= \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f_1(x) f_2(y)\, dy\, dx,$$

while, on the other hand,

$$\mathbb{F}(x_0, y_0) = \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f(x,y)\, dy\, dx.$$

Hence (by measure-theoretic arguments) this is possible only if $f(x,y) = f_1(x) f_2(y)$.
Next, we show that if ξ_1 and ξ_2 are absolutely continuous random variables such that $f(x,y) = f_1(x) f_2(y)$, then ξ_1 and ξ_2 are independent. Let $x_0, y_0 \in \mathbb{R}$. Then

$$\mathbb{F}(x_0, y_0) = \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f(x,y)\, dy\, dx = \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f_1(x) f_2(y)\, dy\, dx =$$
$$= \left(\int_{-\infty}^{x_0} f_1(x)\, dx\right)\left(\int_{-\infty}^{y_0} f_2(y)\, dy\right) = \mathbb{F}_1(x_0)\mathbb{F}_2(y_0).$$

Theorem concerning to the relationship between Independence and Uncorrelatedness

Theorem

If ξ_1 *and* ξ_2 *are independent random variables such that the expectations* $\mathbb{E}(\xi_1)$, $\mathbb{E}(\xi_2)$, *and* $\mathbb{E}(\xi_1\xi_2)$ *exist, then*

$$\mathbb{E}(\xi_1\xi_2) = \mathbb{E}(\xi_1)\mathbb{E}(\xi_2).$$

That is $\operatorname{cov}(\xi_1, \xi_2) = 0$.
However, the converse statement is not true in general.

Proof

The proof relies on the transformation rule for random variables. To compute $\mathbb{E}(\xi_1\xi_2)$, we use the function $g : \mathbb{R}^2 \to \mathbb{R}$ defined by $g(x, y) := xy$.

- In the discrete case, we have

$$\mathbb{E}(\xi_1\xi_2) = \sum_{ij} x_i y_j p_{ij} = \sum_{ij} x_i y_j p_{i\cdot} p_{\cdot j} = \sum_i x_i p_{i\cdot} \sum_j y_j p_{\cdot j}$$
$$= \mathbb{E}(\xi_1)\mathbb{E}(\xi_2).$$

- In the absolutely continuous case, we obtain

$$\mathbb{E}(\xi_1\xi_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y)\, dx\, dy =$$
$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf_1(x)f_2(y)\, dx\, dy$$
$$= \int_{-\infty}^{+\infty} xf_1(x)\, dx \int_{-\infty}^{+\infty} yf_2(y)\, dy = \mathbb{E}(\xi_1)\mathbb{E}(\xi_2).$$

Relationship between Independence and Uncorrelatedness

From the previous results, the relationship between independence and uncorrelatedness is clear. Independence of random variables implies that they are uncorrelated. This follows directly from the covariance formula and the fact that the expected value of the product of independent random variables equals the product of their expectations. That is, if ξ_1 and ξ_2 are independent, then

$$\mathrm{cov}(\xi_1, \xi_2) = \mathbb{E}(\xi_1 \xi_2) - \mathbb{E}(\xi_1)\mathbb{E}(\xi_2) = \mathbb{E}(\xi_1)\mathbb{E}(\xi_2) - \mathbb{E}(\xi_1)\mathbb{E}(\xi_2) = 0.$$

The converse is not true: two random variables can be uncorrelated without being independent, as the following example illustrates.

Uncorrelated but not Independent

Example

We provide an example of two random variables that are uncorrelated but not independent.

- Discrete case: Let (ξ_1, ξ_2) be a discrete random vector with joint distribution given in the table below:

$\xi_1 \backslash \xi_2$	$y_1 = 1$	$y_2 = 3$	$y_3 = 5$
$x_1 = 4$	0.1	0.2	0.1
$x_2 = 6$	0.1	0.4	0.1

A straightforward calculation shows that $\mathbb{E}(\xi_1) = \mathbb{E}(\xi_2) = \mathbb{E}(\xi_1 \xi_2) = 0$, so ξ_1 and ξ_2 are uncorrelated. However, they are not independent, since the product of the marginal distributions does not reproduce the joint distribution.

Uncorrelated but not Independent

Example

- Continuous case: Let (ξ_1, ξ_2) be a random vector with joint density function

$$f(x,y) = \begin{cases} \frac{1}{4}(1 + x^3 y - xy^3), & \text{if } -1 \le x \le 1, -1 \le y \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

A simple calculation shows that $\mathbb{E}(\xi_1) = \mathbb{E}(\xi_2) = \mathbb{E}(\xi_1 \xi_2) = 0$, so ξ_1 and ξ_2 are uncorrelated. However, $\xi_1 \sim U(-1,1)$ and $\xi_2 \sim U(-1,1)$, so they are not independent, because the product of the marginal densities does not reproduce the joint density.
This example can be found in Géza Denkinger's collection of examples.

2nd Example

Example

We spin two spinners, one red and one blue. Each spinner has numbers 1, 2, 3 with equal probability. Let ξ denote the sum of the two numbers, and η the number obtained from the blue spinner. Determine the value of $\operatorname{cov}(\xi, \eta)$.

Solution: We use the additivity of covariance in the first variable and the fact that ξ_1 and ξ_2 are independent, hence uncorrelated.

$$\operatorname{cov}(\xi, \eta) = \operatorname{cov}(\xi_1 + \xi_2, \xi_2) = \operatorname{cov}(\xi_1, \xi_2) + \operatorname{cov}(\xi_2, \xi_2) =$$
$$= 0 + \operatorname{cov}(\xi_2, \xi_2) = \mathbb{D}^2(\xi_2).$$

Solution

Thus it is enough to calculate the variance of the random variable ξ_2, which is straightforward:
The distribution of ξ_2 is:

x_i	1	2	3
p_i	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

From this we get

$$\mathbb{E}(\xi_2) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} = \frac{1+2+3}{3} = 2,$$
$$\mathbb{E}(\xi_2^2) = 1^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{3} + 3^2 \cdot \frac{1}{3} = \frac{1+4+9}{3} = \frac{14}{3},$$
$$\mathbb{D}^2(\xi_2) = \mathbb{E}(\xi_2^2) - (\mathbb{E}(\xi_2))^2 = \frac{14}{3} - 4 = \frac{2}{3},$$

Thus $\operatorname{cov}(\xi, \eta) = \frac{2}{3}$.

7. The Correlation Coefficient

Correlation Coefficient

Definition

If ξ_1 and ξ_2 are random variables with positive standard deviations, then the

$$\mathrm{r}(\xi_1, \xi_2) := \frac{\mathrm{cov}(\xi_1, \xi_2)}{\mathbb{D}(\xi_1)\mathbb{D}(\xi_2)}$$

is called the **correlation coefficient** of the random variables ξ_1 and ξ_2.

Properties of the Correlation Coefficient

Theorem

1. $|\mathrm{r}(\xi_1, \xi_2)| \leqslant 1$,
2. *If* λ_1, λ_2 *are nonzero real numbers and* c_1, c_2 *are arbitrary real numbers, then*

$$\mathrm{r}(\lambda_1\xi_1 + c_1, \lambda_2\xi_2 + c_2) = \mathrm{sgn}(\lambda_1\lambda_2)\, r(\xi_1, \xi_2),$$

where sgn *is the sign function, defined by*

$$\mathrm{sign}(x) := \begin{cases} -1, \text{ if } x < 0 \\ 0, \text{ if } x = 0 \\ 1, \text{ if } x > 0 \end{cases}$$

for all $x \in \mathbb{R}$.

Proof

1. It is a direct consequence of the Cauchy-Bunyakovsky-Schwarz inequality.
2. It easily follows from the properties of covariance and standard deviation:

$$\mathrm{r}(\lambda_1\xi_1 + c_1, \lambda_2\xi_2 + c_2) = \frac{\mathrm{cov}(\lambda_1\xi_1 + c_1, \lambda_2\xi_2 + c_2)}{\mathbb{D}(\lambda_1\xi_1 + c_1)\mathbb{D}(\lambda_2\xi_2 + c_2)} =$$
$$= \frac{\lambda_1\lambda_2}{|\lambda_1\lambda_2|}\frac{\mathrm{cov}(\xi_1, \xi_2)}{\mathbb{D}(\xi_1)\mathbb{D}(\xi_2)} = \mathrm{sgn}(\lambda_1\lambda_2)\mathrm{r}(\xi_1, \xi_2).$$

End of Lecture 7