# Probability Theory and Statistics
## Lecture 13

Bence Csonka

Budapest University of Technology and Economics
*csonkab@edu.bme.hu*

October 28

# Multivariate Continuous Distributions and Moments

Given a continuous random vector $\underline{X} = (X_1, \ldots, X_n)$.
Question: How do we determine $\mathbb{E}(X_i)$ for $1 \leq i \leq n$?

# Multivariate Continuous Distributions and Moments

Given a continuous random vector $\underline{X} = (X_1, \ldots, X_n)$.
Question: How do we determine $\mathbb{E}(X_i)$ for $1 \leq i \leq n$?

Answer 1: We can already determine the density of $X_i$, denoted $f_{X_i}$.
Compute $\to \mathbb{E}(X_i) = \int_{-\infty}^{\infty} x_i\, f_{X_i}(x_i)\,\mathrm{d}x_i$ (if it exists).

## Theorem

*Let $\underline{X} = (X_1, \ldots, X_n)$ be a continuous random vector, and let $g\colon \mathbb{R}^n \to \mathbb{R}$ be a function such that*

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |g(x_1, \ldots, x_n)|\, f_{\underline{X}}(x_1, \ldots, x_n)\,\mathrm{d}x_1 \ldots \mathrm{d}x_n < \infty.$$

*Then*

$$\mathbb{E}(g(X_1, \ldots, X_n)) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n)\, f_{\underline{X}}(x_1, \ldots, x_n)\,\mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

# Multivariate Continuous Distributions and Moments

Answer 2: With the choice $g(x_1, \dots, x_n) = x_i$,

$$\mathbb{E}(X_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i \, f_{\underline{X}}(x_1, \dots, x_n) \, \mathrm{d}x_1 \dots \mathrm{d}x_n.$$

The two methods are the same; in Answer 1 we simply perform the integration with respect to $x_i$ later than the others.

# Multivariate Continuous Distributions and Moments

Answer 2: With the choice $g(x_1, \ldots, x_n) = x_i$,

$$\mathbb{E}(X_i) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} x_i \, f_{\underline{X}}(x_1, \ldots, x_n) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

(Example: $\mathbb{E}(X) = \mathbb{E}(Y) = \frac{2}{3}$, $\mathbb{E}(XY) = \frac{4}{9}$ in the $4xy$ example; see lecture.)

# Independence in the Continuous Case

Reminder:

**Definition**

*Let $X, Y : \Omega \to \mathbb{R}$ be random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $X$ and $Y$ are independent if, for all $x, y \in \mathbb{R}$, the events $\{X < x\}$ and $\{Y < y\}$ are independent.*

Equivalently,

$$F_{(X,Y)}(x, y) = F_X(x) F_Y(y), \qquad \forall x, y \in \mathbb{R}.$$

Independence of $n$ random variables is defined analogously, in accordance with mutual independence of events. In fact, it is still sufficient for independence that the joint distribution function factorizes as the product of the marginal distribution functions:

# Independence in the Continuous Case

Reminder:

**Definition**

*Let $X, Y : \Omega \to \mathbb{R}$ be random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $X$ and $Y$ are independent if, for all $x, y \in \mathbb{R}$, the events $\{X < x\}$ and $\{Y < y\}$ are independent.*

**Theorem**

*The random variables $X_1, \ldots, X_n$ are independent if and only if*

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = F_{X_1}(x_1) \ldots F_{X_n}(x_n)$$

*for all $x_1, \ldots, x_n \in \mathbb{R}$.*

*This holds in the discrete, continuous, and general cases.*

# Independence in the Continuous Case

Discrete case:

**Definition**

*Two discrete random variables $X$ and $Y$ are independent if and only if their joint mass function factorizes into the product of their marginal mass functions, i.e., for all $x, y \in \mathbb{R}$,*

$$p_{(X,Y)}(x,y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) = p_X(x)p_Y(y).$$

Continuous analogue: $X_1, \ldots, X_n$ are independent $\Leftrightarrow$ the joint density of $\underline{X} = (X_1, \ldots, X_n)$ factorizes into the product of the marginal densities of $X_1, \ldots, X_n$.

# Independence in the Continuous Case

Discrete case:

**Definition**

*Two discrete random variables $X$ and $Y$ are independent if and only if their joint mass function factorizes into the product of their marginal mass functions, i.e., for all $x, y \in \mathbb{R}$,*

$$p_{(X,Y)}(x, y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) = p_X(x)p_Y(y).$$

**Theorem**

*Let $X_1, \ldots, X_n$ be continuous random variables. They are independent if and only if $(X_1, \ldots, X_n)$ is a continuous random vector and*

$$f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$$

*holds for all $x_1, \ldots, x_n \in \mathbb{R}$.*

# Independence: Counterexample

In the following example:

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{10x}\,\mathrm{e}^{-x/10}, & \text{if } 0 < x < y < 2x, \\ 0, & \text{otherwise,} \end{cases}$$

independence does not hold. For instance,

- $f_X(50) > 0$,
- $f_Y(42) > 0$,
- but $f_{(X,Y)}(50,42) = 0$.

The formula for the joint density "apparently depends only on $x$", but the support (where it is nonzero) depends on the relation between $x$ and $y$.

# $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ when $X$ and $Y$ are independent

Previously:

**Theorem**

*If $X$ and $Y$ are independent random variables and $\mathbb{E}(XY)$, $\mathbb{E}(X)$, $\mathbb{E}(Y)$ exist, then*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \tag{1}$$

We proved this earlier only for simple random variables; now we prove it for the case when $(X, Y)$ is a continuous random vector (proof: see lecture).

(A previously mentioned) consequence: if $X$ and $Y$ are independent, then $\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$.
As with 6.1.4. Proposition, this holds for all random variables $X$, $Y$.
*Assuming $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$.

# Transforms of Independent Random Variables Are Independent

The following statement is not obvious, and we will not prove it here:

**Theorem**

*If $X$ and $Y$ are independent random variables, and $g$ and $h$ are continuous real functions, then $g(X)$ and $h(Y)$ are independent.*

We will soon use this frequently (in the topic of conditional expectation).

# Relationship Between Random Variables

We saw: if $X$ and $Y$ are independent random variables, they do not affect each other's values. For dependent random variables, however, the value of one can inform us about the value of the other.

Meteorological example:

- There is a "positive relationship" between daily average relative humidity and daily precipitation: if one is high, the other is likely high; if one is low, the other is likely low.
- There is a "negative relationship" between the end-of-month drought index (measuring soil dryness on an increasing scale) and the precipitation over the past month: if one is high, the other is expected to be low, and vice versa.

Measuring the linear relationship between two random variables $\rightarrow$ covariance, correlation.

# Relationship Between Random Variables

If $X$ and $Y$ are independent, the

$$\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$$

for every continuous real $g, h$ for which both sides are defined.

Question: With the choice $g(x) = h(x) = x$ (a linear function), when does

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

hold for not necessarily independent $X, Y$? We already mentioned that this can happen even when the variables are not independent.

# Covariance

Question: How far is $\mathbb{E}(XY)$ from $\mathbb{E}(X)\mathbb{E}(Y)$? To measure this, we introduce the following quantity.

**Definition**

*The covariance of random variables $X$ and $Y$ is defined by*

$$\operatorname{cov}(X, Y) \stackrel{\text{def}}{=} \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))),$$

*provided $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$.*

**Theorem**

*If $\operatorname{cov}(X, Y)$ is meaningful, then*

$$\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

(Proof)

**Theorem**

*If $\mathrm{cov}(X, Y)$ is meaningful, then*

$$\mathrm{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

(Proof)

Recall:
discrete case: $\mathbb{E}(XY) = \sum_{k\in\mathrm{Ran}(X)} \sum_{l\in\mathrm{Ran}(Y)} k\, l\, \mathbb{P}(X = k, Y = l)$,
jointly continuous case: $\mathbb{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{(X,Y)}(x, y)\, \mathrm{d}x\mathrm{d}y$.

# Covariance: Example

| $Y \backslash X$ | 0 | 1 | $p_Y$ |
|---|---|---|---|
| 0 | $\frac{1}{4}$ | 0 | $\frac{1}{4}$ |
| 1 | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{2}$ |
| 2 | 0 | $\frac{1}{4}$ | $\frac{1}{4}$ |
| $p_X$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 1 |

We toss a fair coin twice; $X = \mathbb{1}_{\{\text{the 2nd toss is heads}\}}$, $Y =$ the number of heads.
In this example, we already computed $\mathbb{E}(XY) = \frac{3}{4}$.
Furthermore, $\mathbb{E}(X) = \frac{1}{2}$, $\mathbb{E}(Y) = 1$, hence $\mathrm{cov}(X, Y) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$.

We will see: positive covariance $\rightarrow$ positive linear relationship between $X$ and $Y$.

# Properties of Covariance I

**Theorem**

*Let $X$ and $Y$ be random variables for which $\operatorname{cov}(X, Y)$ is meaningful.*

1. *If $Y$ is constant, then $\operatorname{cov}(X, Y) = 0$.*
2. *If $X$ and $Y$ are independent, then $\operatorname{cov}(X, Y) = 0$.*
3. *From $\operatorname{cov}(X, Y) = 0$ it does not necessarily follow that $X$ and $Y$ are independent.*

Proof:

1. See lecture (linearity of expectation).
2. We have already seen that $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Rearranging gives $\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$.

# Properties of Covariance I

**Theorem**

*Let $X$ and $Y$ be random variables for which $\operatorname{cov}(X, Y)$ is meaningful.*

1. *If $Y$ is constant, then $\operatorname{cov}(X, Y) = 0$.*
2. *If $X$ and $Y$ are independent, then $\operatorname{cov}(X, Y) = 0$.*
3. *From $\operatorname{cov}(X, Y) = 0$ it does not necessarily follow that $X$ and $Y$ are independent.*

Proof:

1. Simple counterexample: let $\operatorname{Ran}(X) = \{-1, 0, 1\}$, with probabilities $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$, respectively.
   Let $Y = |X|$. One can compute (see lecture) that $\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0 - 0 \cdot \frac{1}{2} = 0$, but $X$ and $Y$ are not independent, since, e.g.,

$$0 = \mathbb{P}(X = 1, Y = 0) \neq \mathbb{P}(X = 1)\mathbb{P}(Y = 0) = \frac{1}{8}.$$

# Properties of Covariance II

## Theorem

*Let $(X, Y, Z, W)$ be a random vector and $a, b, c, d \in \mathbb{R}$. Then the following hold, provided the quantities involved are well-defined:*

1. $\operatorname{cov}(X, X) = \mathbb{D}^2(X)$.
   *Consequence:* $\operatorname{cov}(X, X) \geq 0$ *and* $\operatorname{cov}(X, X) = 0 \Leftrightarrow X$ *is almost surely constant.*
2. *Symmetry:* $\operatorname{cov}(X, Y) = \operatorname{cov}(Y, X)$.
3. *Bilinearity (linearity in each variable):*

$$\operatorname{cov}(X+Y, Z+W) = \operatorname{cov}(X, Z)+\operatorname{cov}(X, W)+\operatorname{cov}(Y, Z)+\operatorname{cov}(Y, W).$$

# Properties of Covariance II

- An equivalent formulation of bilinearity:

$$\mathrm{cov}(X, bY + cZ) = b \cdot \mathrm{cov}(X, Y) + c \cdot \mathrm{cov}(X, Z),$$

- Consequence: "additive constants can be dropped; scalar factors pull out of both variables":

$$\mathrm{cov}(aX + b, cY + d) = \mathrm{cov}(aX, cY) = a \cdot c \cdot \mathrm{cov}(X, Y).$$

# Variance of a Sum

**Theorem**

*If $X, Y$ are random variables with $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$, then*

$$\mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y) + 2\,\mathrm{cov}(X,Y).$$

(Proof: $\mathbb{D}^2(X+Y) = \mathrm{cov}(X+Y, X+Y)$; apply bilinearity and symmetry of covariance.)

We see again: if $\mathrm{cov}(X,Y) = 0$, then $\mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$.
Previously known special case: if $X$ and $Y$ are independent, then $\mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$.
If $\mathrm{cov}(X,Y) \neq 0$, this fails!

Example: if $\mathrm{cov}(X,Y) = 1$, $\mathbb{D}^2(X) = 2$ and $\mathbb{D}^2(Y) = 3$, then what is $\mathrm{cov}(2X + Y - 42, 3Y + 100)$?

# Analogy with the Dot Product of Vectors

Covariance is (almost) analogous to the dot product of vectors.

Reminder: if $\underline{u} = (u_1, u_2, u_3), \underline{v} = (v_1, v_2, v_3) \in \mathbb{R}^3$, then the dot product of $\underline{u}$ and $\underline{v}$ is

$$\underline{u} \cdot \underline{v} = u_1 v_1 + u_2 v_2 + u_3 v_3.$$

In $n$ dimensions the definition is analogous, with a sum of $n$ terms.

### Properties

- Positive definite: $\underline{u} \cdot \underline{u} = u_1^2 + u_2^2 + u_3^2 \geq 0$ and $\underline{u} \cdot \underline{u} = 0 \Leftrightarrow \underline{u} = \underline{0}$.
- Symmetric: $\underline{u} \cdot \underline{v} = \underline{v} \cdot \underline{u}$.
- Bilinear: $(\underline{u} + \underline{v}) \cdot \underline{w} = \underline{u} \cdot \underline{w} + \underline{v} \cdot \underline{w}$ (and similarly in the other variable; scalar multiples pull out).

# Analogy with the Dot Product of Vectors

The covariance of random variables has the same properties, except that "additive constants do not matter," i.e., the covariance of a constant r.v. with anything is 0.

- Positive semidefinite: $\operatorname{cov}(X, X) \geq 0$ and $\operatorname{cov}(X, X) = 0 \Leftrightarrow \mathbb{P}(X = c) = 1$ for some $c \in \mathbb{R}$.
- Symmetric: $\operatorname{cov}(X, Y) = \operatorname{cov}(Y, X)$.
- Bilinear: $\operatorname{cov}(X + Y, Z) = \operatorname{cov}(X, Z) + \operatorname{cov}(Y, Z)$ (and similarly in the other variable).

# The CSB Inequality

For the dot product of spatial vectors, the Cauchy–Schwarz–Bunyakovsky (CSB) inequality holds: $\left|\underline{u} \cdot \underline{v}\right| \leq \sqrt{|\underline{u} \cdot \underline{u}| \cdot |\underline{v} \cdot \underline{v}|}$, with equality if and only if $\underline{u}$ and $\underline{v}$ are parallel (i.e., linearly dependent).

Equivalently, for $\underline{u}, \underline{v} \neq \underline{0}$:

$$-1 \leq \frac{\underline{u} \cdot \underline{v}}{\sqrt{(\underline{u} \cdot \underline{u}) \cdot (\underline{v} \cdot \underline{v})}} \leq +1.$$

- $= +1 \Rightarrow \underline{u}$ and $\underline{v}$ are parallel and in the same direction, i.e., $\underline{v} = a\underline{u}$ for some $a > 0$.
- $= -1 \Rightarrow \underline{u}$ and $\underline{v}$ are parallel and in opposite directions, i.e., $\underline{v} = a\underline{u}$ for some $a < 0$.
- $= 0$ if and only if $\underline{u}$ and $\underline{v}$ are orthogonal.

Interpretation: the ratio measures the "degree of linear dependence" (of course, two nonparallel nonzero vectors are always linearly independent).

Why are these true? Because $\underline{u} \cdot \underline{v} = |\underline{u}||\underline{v}| \cos(\alpha)$, where $\alpha$ is the angle between the two vectors.

# CSB Inequality and Correlation

$$-1 \leq \frac{\underline{u} \cdot \underline{v}}{\sqrt{|\underline{u} \cdot \underline{u}| \cdot |\underline{v} \cdot \underline{v}|}} \leq +1, \qquad \underline{u}, \underline{v} \neq 0.$$

The analogue of this ratio for covariance is correlation:

## Definition

*Let $X$ and $Y$ be non-constant random variables. If $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$, then the correlation of $X$ and $Y$ is*

$$\mathrm{corr}(X, Y) \overset{\text{def.}}{=} \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{cov}(X, X)\mathrm{cov}(Y, Y)}} = \frac{\mathrm{cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)}.$$

# Correlation

Correlation always lies between $-1$ and $1$ ("$-100\%$ and $+100\%$"). The following statements are analogues of the CSB inequality, taking into account the "additive constants do not matter":

$$-1 \leq \operatorname{corr}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)} \leq 1.$$

if $X$ and $Y$ are independent, then $\operatorname{corr}(X, Y) = 0$. From $\operatorname{corr}(X, Y) = 0$ we generally cannot conclude that $X, Y$ are independent; typically it only rules out a *linear* relationship between them.
Terminology: $\operatorname{corr}(X, Y) > 0$: $X$ and $Y$ are positively correlated; $\operatorname{corr}(X, Y) = 0$: uncorrelated; $\operatorname{corr}(X, Y) < 0$: negatively correlated.

# Correlation: Example

| $Y \backslash X$ | 0 | 1 | $p_Y$ |
|---|---|---|---|
| 0 | $\frac{1}{4}$ | 0 | $\frac{1}{4}$ |
| 1 | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{2}$ |
| 2 | 0 | $\frac{1}{4}$ | $\frac{1}{4}$ |
| $p_X$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 1 |

We toss a fair coin twice; $X = \mathbb{1}_{\{\text{the 2nd toss is heads}\}}$, $Y =$ the number of heads. We saw: $\operatorname{cov}(X, Y) = \frac{1}{4}$.
Furthermore, $\mathbb{E}(X) = \frac{1}{2}$, $\mathbb{E}(Y) = 1$, hence

$$\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

and

$$\mathbb{D}^2(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \frac{1}{2} \cdot 1^2 + \frac{1}{4} \cdot 2^2 - 1^2 = \frac{1}{2}.$$

# Correlation: Example

| $Y \backslash X$ | 0 | 1 | $p_Y$ |
|---|---|---|---|
| 0 | $\frac{1}{4}$ | 0 | $\frac{1}{4}$ |
| 1 | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{2}$ |
| 2 | 0 | $\frac{1}{4}$ | $\frac{1}{4}$ |
| $p_X$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 1 |

Thus

$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)} = \frac{\frac{1}{4}}{\frac{1}{2}\frac{1}{\sqrt{2}}} = \frac{\sqrt{2}}{2}.$$

$X$ and $Y$ are positively correlated (what is the reason?).

In the weather example:
Humidity and precipitation $\rightarrow$ positively correlated; drought index and precipitation $\rightarrow$ negatively correlated.

A frequent misunderstanding in some social-science contexts: a large (near 1) correlation between two random quantities usually indicates not causation, but a **linear** relationship.
It may happen, e.g., that neither of two positively correlated quantities causes the other; rather, both follow from a third cause.