Probability and Mathematical Statistics

Miskolc, 2025.

Dr. Tamás Glavosits

Lecture 8

Random Variables II.
Notable Random Vector Variables

1. Mathematical tools

Transpose of Vectors

Definition

If $\mathbf{x} \in \mathbb{R}^n$, then

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

where $x_1, x_2, \ldots, x_n$ are called the components of the vector $\mathbf{x}$.
By the notation, it is clear that a vector in $\mathbb{R}^n$ is always a **column vector**.
If we need a **row vector** then we can transpose the vector.

$$\mathbf{x}^T := [x_1, x_2, \ldots, x_n].$$

Matrices

Definition

We say that $A \in \mathbb{R}^{n \times m}$ if

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

The a_{ij} elements are called the **components** of the matrix A. The matrix $A \in \mathbb{R}^{n \times m}$ has n rows and m columns. If $A \in \mathbb{R}^{n \times m}$, we say that the matrix A is type of $n \times m$.

Definition

A matrix A is called a **square** or **quadratic** matrix, if it is an $n \times n$ type matrix.

Transpose of a Matrix

Definition

Similarly to the transpose of vectors, we can define the transpose for matrices. If $A \in \mathbb{R}^{n \times m}$, then $A^T \in \mathbb{R}^{m \times n}$. The transpose changes the indices of the rows and columnes.

Example

$$A := \underbrace{\begin{bmatrix} 2 & 1 & 3 \\ 3 & 0 & 2 \end{bmatrix}}_{\in \mathbb{R}^{2 \times 3}} \qquad \Longrightarrow \qquad A^T = \underbrace{\begin{bmatrix} 2 & 3 \\ 1 & 0 \\ 3 & 2 \end{bmatrix}}_{\in \mathbb{R}^{3 \times 2}}$$

Multiplication of Matrices

Definition

If $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times l}$, then the matrices A and B can be multiplied and $C := AB \in \mathbb{R}^{m \times l}$.
The component c_{ij} of the matrix C is defined by

$$c_{ij} := \sum_{k=1}^{n} a_{ik} \cdot b_{kj}.$$

Example

Remark

It is worth noting that if $x, y \in \mathbb{R}^n$, then $x^T y$ and xy^T represent different things. Let us denote, for example,

$$x = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \qquad y = \begin{bmatrix} 2 \\ 0 \\ -3 \end{bmatrix}$$

Remark

- For $x^T y$, if $x, y \in \mathbb{R}^3$, then $x^T \in \mathbb{R}^{1\times 3}$, $y \in \mathbb{R}^{3\times 1}$, thus $x^T y \in \mathbb{R}^{1\times 1} = \mathbb{R}$.

$$x^T y = [1\ 3\ 2] \begin{bmatrix} 2 \\ 0 \\ -3 \end{bmatrix} = 1 \cdot 2 + 3 \cdot 0 + 2 \cdot (-3) = -4.$$

- For xy^T, if $x, y \in \mathbb{R}^3$, then $x \in \mathbb{R}^{3\times 1}$, $y^T \in \mathbb{R}^{1\times 3}$, thus $xy^T \in \mathbb{R}^{3\times 3}$.

$$xy^T = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} [2\ 0\ -3] = \begin{bmatrix} 1\cdot 2 & 1\cdot 0 & 1\cdot(-3) \\ 3\cdot 2 & 3\cdot 0 & 3\cdot(-3) \\ 2\cdot 2 & 2\cdot 0 & 2\cdot(-3) \end{bmatrix} = \begin{bmatrix} 1 & 0 & -3 \\ 6 & 0 & -9 \\ 4 & 0 & -6 \end{bmatrix}$$

Determinant of a square Matrix

Definition

- **Expansion along the i-th row** Let $A \in \mathbb{R}^{n \times n}$. Then

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} A_{ij} := \sum_{i=1}^{n} a_{ij} D_{ij},$$

  where A_{ij} denotes the (i,j)-th subdeterminant of the matrix A, obtained from A by removing its i-th row and j-th column, and taking the determinant of the remaining $(n-1) \times (n-1)$ type matrix.
  The determinant D_{ij} is called the (i,j)-th **adjugated subdeterminant** of the matrix A.

- **Expansion along the j-th column**.

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} A_{ij} = \sum_{i=1}^{n} a_{ij} D_{ij}.$$

Example of Calculating the Determinant of a Matrix

Let us calculate the determinant of the matrix

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 2 & 1 \\ 1 & 3 & 0 \end{bmatrix}$$

using expansion along the first and second rows, and along the second column. (For brevity, we write $|A|$ instead of $\det(A)$.)

- **Expansion along the first row**

$$\begin{vmatrix} 1 & 0 & 2 \\ 2 & 2 & 1 \\ 1 & 3 & 0 \end{vmatrix} = 1\underbrace{\begin{vmatrix} 2 & 1 \\ 3 & 0 \end{vmatrix}}_{-3} - 0\underbrace{\begin{vmatrix} 2 & 1 \\ 1 & 0 \end{vmatrix}}_{0} + 2\underbrace{\begin{vmatrix} 2 & 2 \\ 1 & 3 \end{vmatrix}}_{4} = 5;$$

Example of Calculating the Determinant of a Matrix

- Expansion along the second row

$$\begin{vmatrix} 1 & 0 & 2 \\ 2 & 2 & 1 \\ 1 & 3 & 0 \end{vmatrix} = -2\underbrace{\begin{vmatrix} 0 & 2 \\ 3 & 0 \end{vmatrix}}_{-6} + 2\underbrace{\begin{vmatrix} 1 & 2 \\ 1 & 0 \end{vmatrix}}_{-2} - 1\underbrace{\begin{vmatrix} 1 & 0 \\ 1 & 3 \end{vmatrix}}_{3} = 5;$$

- Expansion along the second column

$$\begin{vmatrix} 1 & 0 & 2 \\ 2 & 2 & 1 \\ 1 & 3 & 0 \end{vmatrix} = -0\underbrace{\begin{vmatrix} 2 & 1 \\ 1 & 0 \end{vmatrix}}_{0} + 2\underbrace{\begin{vmatrix} 1 & 2 \\ 1 & 0 \end{vmatrix}}_{-2} - 3\underbrace{\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix}}_{-3} = 5;$$

Multiplication Theorem, and Identity Matrix

Theorem

Multiplication Theorem

$$\det(AB) = \det(A)\det(B),$$

where $A, B \in \mathbb{R}^{n\times n}$ *for some* $n \geqslant 1$.

Definition

A matrix $I \in \mathbb{R}^{n\times n}$ (of type $n \times n$) is called the **identity matrix** if

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Invertible Matrices

Theorem

If $A \in \mathbb{R}^{n \times n}$ and I is the $n \times n$ identity matrix, then

$$A \cdot I = I \cdot A = A,$$

that is, I is the identity element of the (multiplicative) semigroup of $n \times n$ matrices. (The identity element in a semigroup exists uniquely.)

Definition

Let $A \in \mathbb{R}^{n\times n}$. If there exists a matrix $B \in \mathbb{R}^{n\times n}$ such that $A \cdot B = I$. Then the matrix A is called **invertible** or **non-singular**. The matrix B (if it exists) is said to be the inverse of matrix A is denoted by A^{-1}.

Theorem

A matrix $A \in \mathbb{R}^{n\times n}$ is invertible if and only if $\det(A) \neq 0$.

2. Random Vector Variables and Their Joint Distribution Function

Random Vector Variables (n-dimensional case)

Definition

The concept of a **random vector variable** can be interpreted in two ways:

- Let $\xi_1, \xi_2, \ldots, \xi_n$ be random variables; then the vector $(\xi_1, \xi_2, \ldots, \xi_n)$ is called a random vector variable.
- If $(\xi_1, \xi_2, \ldots, \xi_n) : \Omega \to \mathbb{R}^n$ is such that

$$(\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n) :=$$
$$= \{\omega \in \Omega \mid \xi_1(\omega) < x_1, \xi_2(\omega) < x_2, \ldots, \xi_n(\omega) < x_n\} \in \mathcal{F},$$

  for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, then the function $(\xi_1, \xi_2, \ldots, \xi_n)$ is called a **random vector variable**.

It is easy to see that these two interpretations are equivalent.

Joint Distribution Function of Random Vector Variables

Definition

If $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ is a random vector variable, then the function $\mathbb{F} : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\mathbb{F}(x_1, x_2, \ldots, x_n) := \mathbb{P}\left(\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n\right)$$

(for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$) is called the **joint distribution function** of the random vector variable ξ.

Properties of the Joint Distribution Function

Theorem

If $\mathbb{F}$ is the distribution function of the random vector $(\xi_1, \xi_2, \ldots, \xi_n)$, then

1. *$\mathbb{F}$ is left-continuous in each of its variables;*
2. *If $a_1 < b_1, \ldots, a_n < b_n$, then*

$$\sum_{(\varepsilon_1, \ldots, \varepsilon_n) \in \{0,1\}^n} (-1)^{\varepsilon_1 + \cdots + \varepsilon_n} \mathbb{F}\left(\varepsilon_1 a_1 + (1-\varepsilon_1) b_1, \ldots, \varepsilon_n a_n + (1-\varepsilon_n) b_n\right) \geqslant 0;$$

3. *Limits:*
   - $\lim_{x_i \to -\infty} \mathbb{F}(x_1, x_2, \ldots, x_n) = 0$, *that is, if one of the variables tends to $-\infty$, the function $\mathbb{F}$ tends to 0.*
   - $\lim_{\substack{x_1 \to +\infty \\ x_2 \to +\infty \\ \cdots \\ x_n \to +\infty}} \mathbb{F}(x_1, x_2, \ldots, x_n) = 1$, *that is, if all variables tend to $+\infty$, the function $\mathbb{F}$ tends to 1.*

3. Types of Random Vector Variables

Discrete Random Vector Variables

Definition

A **random vector variable** $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ is called **discrete** if its set of values is countable. A discrete random vector variable $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ has

- **set of values**

  $$\{(x_{i_1}, x_{i_2}, \dots, x_{i_n}) \mid i_1 = 1, 2, \dots, i_2 = 1, 2, \dots, i_n = 1, 2, \dots\}$$

  such that

  $$\mathbb{P}\left((\xi_1, \xi_2, \dots, \xi_n) \in \{(x_{i_1}, x_{i_2}, \dots, x_{i_n})\}\right) = 1.$$

- **distribution**

  $$p_{i_1, i_2, \dots, i_n} := \mathbb{P}(\xi_1 = x_{i_1}, \xi_2 = x_{i_2}, \dots, \xi_n = x_{i_n})$$

  for all $i_1 = 1, 2, \dots, i_2 = 1, 2, \dots, i_n = 1, 2, \dots$.

Properties of the Joint Distributions of Discrete Random Vector Variables

Theorem

If $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ *is a discrete random vector variable with values*

$$\{x_{i_1}, x_{i_2}, \dots, x_{i_n} | i_1 = 1, 2, \dots; i_2 = 1, 2, \dots; \dots; i_n = 1, 2, \dots, \},$$

and with join distribution

$$p_{i_1 i_2 \dots i_n} = \mathbb{P}(\xi_1 = x_{i_1}, \xi_2 = x_{i_2}, \dots, \xi_n = x_{i_n}),$$

then

1. $p_{i_1 i_2 \dots i_n} \geqslant 0$,
2. $\sum_{i_1} \sum_{i_2} \cdots \sum_{i_n} p_{i_1 i_2 \dots i_n} = 1.$

Absolutely Continuous Random Vector Variables

Definition

A **random vector variable** $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ is called **absolutely continuous** if there exists a function $f : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathbb{F}(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(u_1, u_2, \dots, u_n) du_1 du_2 \dots du_n$$

for all $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.
The function f is called the **joint density function** of the absolutely continuous random vector variable.

Properties of the Joint Density Function

Theorem

If $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ is an absolutely continuous random vector variable, then its joint density function $f : \mathbb{R}^n \to \mathbb{R}$ has the following properties:

- $f(x_1, x_2, \ldots, x_n) \geqslant 0$ *for all* $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ *(this is more of a convention than a property);*
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(u_1, u_2, \ldots, u_n) du_1 du_2 \ldots du_n = 1.$

4. Expectation Vector and Covariance Matrix of a Random Vector Variable

Expectation Vector

Definition

For $\xi = [\xi_1, \xi_2, \ldots, \xi_n]^T$, the **expectation vector of the random vector variable** is

$$\mathbb{E}(\xi) := [\mathbb{E}(\xi_1), \mathbb{E}(\xi_2), \ldots, \mathbb{E}(\xi_n)]^T .$$

Properties of the Expectation Vector

Theorem

Let $\xi = [\xi_1, \xi_2, \ldots, \xi_n]^T$ *be random vector variables,* $A \in \mathbb{R}^{n \times n}$, $\boldsymbol{b} \in \mathbb{R}^n$. *The expectation has the following properties:*

1. ***Additivity***: $\mathbb{E}(\xi + \eta) = \mathbb{E}(\xi) + \mathbb{E}(\eta)$,
2. ***Homogeneity***: $\mathbb{E}(A\xi) = A\mathbb{E}(\xi)$,
3. ***Property concerning the translation***: $\mathbb{E}(\xi + \boldsymbol{b}) = \mathbb{E}(\xi) + \boldsymbol{b}$.

Covariance Matrix

Definition

For $\xi := [\xi_1, \xi_2, \ldots, \xi_n]^T$, the **covariance matrix of the random vector variable** is defined as

$$\mathrm{var}(\xi) = \begin{bmatrix} \mathbb{D}^2(\xi_1) & \mathrm{cov}(\xi_1, \xi_2) & \ldots & \mathrm{cov}(\xi_1, \xi_n) \\ \mathrm{cov}(\xi_2, \xi_1) & \mathbb{D}^2(\xi_2) & \ldots & \mathrm{cov}(\xi_2, \xi_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(\xi_n, \xi_1) & \mathrm{cov}(\xi_n, \xi_2) & \ldots & \mathbb{D}^2(\xi_n) \end{bmatrix}.$$

Thus, $\mathrm{var}(\xi) \in \mathbb{R}^{n \times n}$, whose (i, j)-th component is $\mathrm{cov}(\xi_i, \xi_j)$.

Properties of the Covariance Matrix

Theorem

Let $\xi := [\xi_1, \xi_2, \ldots, \xi_n]^T$ *be a random vector variable. Then the matrix* $\mathrm{var}(\xi) \in \mathbb{R}^{n \times n}$ *has the following properties:*

1. ***Positive semidefinite****, i.e., for all* $x \in \mathbb{R}^n$, $x^T \mathrm{var}(\xi) x \geqslant 0$.
2. ***Quadratically homogeneous****, i.e., for all* $A \in \mathbb{R}^{n \times n}$, $\mathrm{var}(A\xi) = A\mathrm{var}(\xi)A^T$.
3. ***Translation invariant****, i.e., for all* $B \in \mathbb{R}^n$, $\mathrm{var}(\xi + B) = \mathrm{var}(\xi)$.

5. Notable Discrete Random Vector Variables

The Multinomial Distribution

Definition

Let $r \geqslant 2$, and let $B_1, B_2, \ldots, B_r$ be a complete system of events with probabilities $p_1, p_2, \ldots, p_r$. We perform n independent trials and count how many times each of the events $B_1, B_2, \ldots, B_r$ occurs.
Denoted as: $\xi = [\xi_1, \xi_2, \ldots, \xi_r]^T \sim \text{Multinom}(n, p_1, p_2, \ldots, p_r)$.
Distribution:

$$p_{k_1, k_2, \ldots, k_r} = \mathbb{P}(\xi_1 = k_1, \xi_2 = k_2, \ldots, \xi_r = k_r) = \frac{n!}{k_1! \cdot k_2! \cdot \cdots \cdot k_r!} p_1^{k_1} p_2^{k_2} \ldots p_r^{k_r}$$

where $0 \leqslant k_i \leqslant n$ $(i = 1, 2, \ldots, r)$ and $k_1 + k_2 + \cdots + k_r = n$.
Expectation vector and covariance matrix:

$$\mathbb{E}(\xi_i) = np_i, \qquad \mathbb{D}^2(\xi_i) = np_i(1 - p_i), \qquad \operatorname{cov}(\xi_i, \xi_j) = -np_i p_j \qquad (i \neq j).$$

Multivariate Hypergeometric Distribution

Definition

Drawing balls without replacement from an urn with more than two colors. We use the following notations:
N: the total number of balls,
r: the number of colors,
$s_1, s_2, \ldots, s_r$: the number of balls of each color ($s_1 + s_2 + \cdots + s_r = N$),
n: the number of balls drawn.

Joint distribution:

$$p_{k_1,k_2,\ldots,k_r} = \mathbb{P}(\xi_1 = k_1, \xi_2 = k_2, \ldots, \xi_r = k_r) = \frac{\binom{s_1}{k_1}\binom{s_2}{k_2}\ldots\binom{s_r}{k_r}}{\binom{N}{n}}$$

where $0 \leqslant k_i \leqslant s_i$ $(i = 1, 2, \ldots, r)$ and $k_1 + k_2 + \cdots + k_r = n$.

Expectation vector and covariance matrix:

$$\mathbb{E}(\xi_i) = n\frac{s_i}{N}, \quad \mathbb{D}^2(\xi_i) = n\frac{N-n}{N-1}\frac{s_i}{N}\left(1 - \frac{s_i}{N}\right), \quad \operatorname{cov}(\xi_i, \xi_j) = -\frac{N-n}{N-1}\frac{s_i}{N}\frac{s_j}{N}$$

$(i \neq j)$.

6. Notable Absolutely Continuous Random Vector Variables

Multidimensional Uniform Distribution

We will consider only the two-dimensional case; the multidimensional uniform distribution can be defined analogously.

Definition

Let $D \subseteq \mathbb{R}^2$ be a connected bounded open (or closed) set. We say that the vector variable $\xi = [\xi_1, \xi_2]^T$ is uniformly distributed over D if its density function is

$$f(x, y) = \begin{cases} C, & \text{if } (x, y) \in D \\ 0, & \text{otherwise,} \end{cases}$$

where $C = \frac{1}{m(D)}$, with $m(D)$ denotes the area of the set D.

n-dimensional Standard Normal Distribution

Definition

Let $\eta_1, \eta_2, \ldots, \eta_n$ be completely independent standard normal distributions. Then the random vector $\eta = [\eta_1, \ldots, \eta_n]^T$ is called an **n-dimensional standard normal random vector**. Denoted by: $\eta \sim \mathcal{N}(0, I)$.

Joint density function:

$$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}x^T x\right) \qquad (x \in \mathbb{R}^n).$$

Expectation vector and covariance matrix:

$$\mathbb{E}(\eta) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \mathrm{var}(\eta) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

n-dimensional Non-degenerate Normal Distribution

Definition

Let η be an n-dimensional standard normal random vector, $A \in \mathbb{R}^{n \times n}$ with $\det(A) \neq 0$, and $m \in \mathbb{R}^n$. Then the random variable $\xi \sim \eta + m$ is called an n-dimensional normal distributed random vector. Denoted by $\xi \sim \mathcal{N}(m, D)$, where $D = AA^T$.

Joint density function:

$$f_\eta(x) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(D))^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-m)^T D^{-1}(x-m)\right)$$

$(x \in \mathbb{R}^n)$.

Expectation vector and covariance matrix:

$$\mathbb{E}(\xi) = m, \qquad \mathrm{var}(\xi) = AA^T.$$

2-dimensional Normal Random Variables

Theorem

Let (ξ_1, ξ_2) *be a normally distributed random variable with parameters:* $m_1 = \mathbb{E}(\xi_1)$, $\sigma_1^2 = \mathbb{D}^2(\xi_1)$, $m_2 = \mathbb{E}(\xi_2)$, $\sigma_2^2 = \mathbb{D}^2(\xi_2)$, $\mathrm{r} = \mathrm{r}(\xi_1, \xi_2)$. *Then the joint density function is:*

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\mathrm{r}^2}} \cdot \exp\left(-\frac{1}{2(1-\mathrm{r}^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2} + 2\mathrm{r}\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right]\right)$$

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

End of Lecture 8