# Probability Theory and Statistics
## Lecture 18

Bence Csonka

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

November 18

# Motivation

In the previous part of the lecture, the distribution of the random variables was assumed to be known. In contrast, in mathematical statistics, random variables correspond to measurement results, and therefore their distributions are not known precisely.

[1] $\vartheta$ and $\theta$ are two lowercase forms of the Greek letter „theta."

# Motivation

In the previous part of the lecture, the distribution of the random variables was assumed to be known. In contrast, in mathematical statistics, random variables correspond to measurement results, and therefore their distributions are not known precisely. The distribution depends on an unknown *parameter* $\vartheta$, whose possible values form a *parameter domain* $\theta \subseteq \mathbb{R}^d$ for some $d \geq 1$.[1]

[1] $\vartheta$ and $\theta$ are two lowercase forms of the Greek letter „theta."

# Examples 1

If we find a coin on the street and do not know whether it is fair, then it shows heads with an unknown probability $\vartheta \in \theta = [0, 1] \subset \mathbb{R}^1$. By tossing the coin several times, we can try to estimate $\vartheta$, or verify or reject the hypothesis that the coin is fair. The indicator variable

$$\mathbb{1}_{\{\text{the outcome of a given toss is head}\}}$$

thus has the unknown parameter $\vartheta$.

# Example 2

The number of accidents at railway crossings in Hungary during a given month can be assumed to follow a Poisson distribution quite accurately, since there are many drivers, each having a small probability of an accident, and the events are more or less independent. The Poisson distribution has an *unknown* parameter $\vartheta \in \theta = (0, \infty) \subset \mathbb{R}^1$.

# Example 3

The height of a randomly selected female student at BME can be modeled by a normal distribution.[2] In this case, both the mean $\mu \in \mathbb{R}$ and the variance $\sigma^2 > 0$ are unknown, so the parameter domain is

$$\theta = \{(\mu, \sigma^2) \in \mathbb{R}^2 \colon \sigma^2 > 0\} \subset \mathbb{R}^2.$$

[2] Similarly, a randomly selected male student's height can be modeled by a normal distribution with a different mean. However, the height of a randomly selected BME student (without conditioning on gender) cannot, since male and female averages differ, leading to a density with two local maxima — which is therefore not approximately normal.

# Basic concept

In mathematical statistics, it is typical that in order to study the unknown parameter, we *take a sample*, that is, we „generate" independent, identically distributed random variables $X_1, \dots, X_n$ following the distribution with the unknown parameter. For example:
We toss the coin $n$ times, and for each $i = 1, \dots, n$, let $X_i = 1$ if the $i$-th toss results in heads, and $X_i = 0$ otherwise. Then $X_1, \dots, X_n$ are independent, identically distributed indicator variables with parameter $\vartheta$.

# Basic concept

We define the concept of a sample without referring to the unknown parameter $\vartheta$.

## Definition

*Let $X_1, \ldots, X_n$ be independent, identically distributed random variables with possibly unknown marginal distributions. Then the random vector*

$$\mathbf{X} = (X_1, \ldots, X_n)$$

*is called an* **independent and identically distributed sample of size** *$n$ (abbreviated as an* **i.i.d. sample of size** *$n$).*

# Useful notations

Let $\mathbf{X} = (X_1, \dots, X_n)$ be an i.i.d. sample of size $n$, where the common distribution depends on a parameter $\vartheta \in \theta \subseteq \mathbb{R}^d$ for some $d \geq 1$. Then, for a given $\vartheta \in \theta$:

1. The distribution function of $X_1$ corresponding to parameter $\vartheta$ is denoted by $F_\vartheta$.
2. If $X_1$ has a density function under this parameter, it is denoted by $f_\vartheta$, i.e.
$$f_\vartheta(x) = \begin{cases} F'_\vartheta(x), & \text{if } F_\vartheta \text{ is differentiable at } x, \\ 0, & \text{otherwise.} \end{cases}$$
3. If $X_1$ is discrete under this parameter, its probability mass function is denoted by $p_\vartheta$, meaning that for $x \in \mathbb{R}$, $p_\vartheta(x)$ denotes the probability that $X_1 = x$, given that $\vartheta$ is the true parameter.

# Examples

- For the coin found on the street: $p_\vartheta(0) = 1 - \vartheta$ and $p_\vartheta(1) = \vartheta$, $\vartheta \in [0, 1]$.
- For the number of railway crossing accidents:

$$p_\vartheta(k) = \frac{\vartheta^k}{k!} e^{-\vartheta}, \qquad \vartheta > 0,\ k = 0, 1, \ldots.$$

- For the height of female BME students, the unknown parameter is $(\mu, \sigma^2)$, hence

$$f_{(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}, \quad \mu \in \mathbb{R},\ \sigma^2 > 0.$$

We leave the corresponding distribution functions to the reader.

# Essential values

## Definition

*Let $\mathbf{X} = (X_1, \ldots, X_n)$ be an i.i.d. sample of size n, where the distribution of the sample elements depends on a parameter $\vartheta \in \theta \subseteq \mathbb{R}^d$. For $\vartheta \in \theta$ and $i = 1, \ldots, n$, if the density function $f_\vartheta$ exists, define*

$$S_{X_i}^{(\vartheta)} = \{x \in \mathbb{R} \mid f_\vartheta(x) > 0\} \subseteq \mathbb{R}.$$

If instead the probability mass function $p_\vartheta$ exists, define

$$S_{X_i}^{(\vartheta)} = \{x \in \mathbb{R} \mid p_\vartheta(x) > 0\} \subseteq \mathbb{R}.$$

In both cases, $S_{X_i}^{(\vartheta)}$ is called the **set of essential values of $X_i$** for the parameter $\vartheta$.[3]

[3] The set of essential values of $X_i$ (for a given $\vartheta$) is almost the same as the image of $X_i$ as a function $\Omega \to \mathbb{R}$. The image may be slightly larger, but all values outside $S_{X_i}^{(\vartheta)}$ occur with probability zero.

# Realization

**Definition**

*Let* $\mathbf{X} = (X_1, \ldots, X_n)$ *be an i.i.d. sample of size* $n$*, where the sample distribution depends on a parameter* $\vartheta \in \theta \subseteq \mathbb{R}^d$*. A vector* $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ *is called a (possible)* **realization** *of* $\mathbf{X} = (X_1, \ldots, X_n)$ *for parameter* $\vartheta \in \theta$*, if* $x_i \in S_{X_i}^{(\vartheta)}$ *holds for all* $i \in \{1, \ldots, n\}$*.*

# Example for realization

For the coin found on the street, the sequence $(x_1, \ldots, x_7) = (1, 0, 0, 0, 1, 1, 0)$ is a realization of the i.i.d. sample $\mathbf{X} = (X_1, \ldots, X_7)$ for all parameters $0 < \vartheta < 1$ (but not for $\vartheta = 0$ or $\vartheta = 1$).
For the i.i.d. sample $\mathbf{X} = (X_1, \ldots, X_5)$ uniformly distributed on $(0, \vartheta)$,

$$(x_1, \ldots, x_5) = (0.14, 0.79, 1.13, 1, 1.2)$$

is a possible realization if $1.2 < \vartheta < 2$, but not if $1 < \vartheta \leq 1.2$.

# Statistics

For an i.i.d. sample of size $n$, a *statistic* is any function of the sample elements which is symmetric, that is, it „depends on all sample elements in the same way." The following definition formalizes this property.

**Definition**

*Let* $\mathbf{X} = (X_1, \ldots, X_n)$ *be an i.i.d. sample of size n. If* $T\colon \mathbb{R}^n \to \mathbb{R}$ *is a symmetric function, that is,*

$$T(x_1, \ldots, x_n) = T(x_{\pi(1)}, \ldots, x_{\pi(n)})$$

*for all* $x_1, \ldots, x_n \in \mathbb{R}$ *and for every permutation* $\pi\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ *in the combinatorics chapter for an equivalent definition, then the random variable* $T(\mathbf{X}) = T(X_1, \ldots, X_n)$ *is called a* **statistic** *of* $X_1, \ldots, X_n$.

# Sample mean

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be an i.i.d. sample of size $n$. Then the quantity

$$\overline{X_n} = \frac{X_1 + \ldots + X_n}{n},$$

is called the **sample mean** of $\mathbf{X}$, and it is a statistic of the sample. If $\mathbf{x} = (x_1, \ldots, x_n)$ is a realization of $\mathbf{X} = (X_1, \ldots, X_n)$, then we denote the mean of the realization by $\overline{x_n}$:

$$\overline{x_n} = \frac{x_1 + \ldots + x_n}{n}.$$

# Corrected empirical variance

Let $\mathbf{X} = (X_1, \dots, X_n)$ be an i.i.d. sample of size $n$. Then

$$(S_n^*)^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X_n})^2 \tag{1}$$

is called the **corrected empirical variance** of $\mathbf{X}$,[4] and

$$S_n^* = \sqrt{(S_n^*)^2}$$

is called the **corrected empirical standard deviation** of $\mathbf{X}$.

[4] The word „empirical" means „based on observations." The term „corrected empirical variance" is also used.

# Property of mentioned statistics

**Theorem**

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}(X_1)$$

$$\mathbb{E}((S_n^*)^2) = \mathbb{D}^2(X_1).$$

(Proof)