MOMENTS OF THE ONE-DIMENSIONAL RANDOM VARIABLE

1. LESSON

Xabier Erdocia
Itsaso Leceta

BY NC SA

eman ta zabal zazu
UPV EHU

OBJECTIVES

- ✓ Understand and internalize the concept of one-dimensional random variable
- ✓ Be able to calculate the moments and the mean value of a random variable
- ✓ Calculation and adequate interpretation of position and dispersion measures using moments
- ✓ Be able to find the minimum probability that the random variable is within a range, using the differentiation of Tchebychev

OCW

eman ta zabal zazu
UPV EHU

INDEX

1.1. Concept of one-dimensional random variable
1.2. Mean value
1.3. Moments of the random variable
1.4. Measures of position. Dispersion measures. Fisher coefficients
1.5. Tchebychev's inequality

OCW

eman ta zabal zazu

UPV EHU

1.1. Concept of one-dimensional random variable

OCW

eman ta zabal zazu
UPV EHU

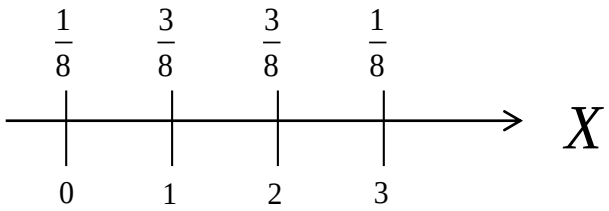
- Let be *E* a random essay and *S* its sample space. The **random variable** is a function in which to each possible result in *S* a real number is assigned.

$$X : S \longrightarrow \mathbb{R}$$

- The random variable can be discreet or continuous:
  - **Discrete**: when the number of images is finite or infinite quantifiable
  - **Continuous**: when the image can take all the values of a range

- Suppose the following essay: “flip a coin 3 times”.
- X random variable can be defined as “number of tails obtained”. The random variable assigns a real value to all the possible results obtained in the trial.
- The possible values of the random variable (outcomes) are:
  - X=0 ⟹ 3 heads ○○○
  - X=1 ⟹ 1 tail and 2 heads ✕○○ ○✕○ ○○✕
  - X=2 ⟹ 2 tails and 1 head ✕✕○ ○✕✕ ✕○✕
  - X=3 ⟹ 3 tails ✕✕✕

- The probability of each case can be calculated:
- The total amount of possible results is 8:
  - X=0 ⟹ ○○○ $P(X=0)=\frac{1}{8}$
  - X=1 ⟹ ⊠○○ ○⊠○ ○○⊠ $P(X=1)=\frac{3}{8}$
  - X=2 ⟹ ⊠⊠○ ○⊠⊠ ⊠○⊠ $P(X=2)=\frac{3}{8}$
  - X=3 ⟹ ⊠⊠⊠ $P(X=3)=\frac{1}{8}$



OCW

eman ta zabal zazu
UPV EHU

- **Probability distribution**: represents the distribution of probability among all possible results of a test for the defined random variable.

  - In order to analyse discrete distributions, the *probability function* is:

$$p(x) = P(X = x) / \forall x \in \mathbb{R}$$

    - That is, the probability that the variable will take a concrete value.
  - In order to analyse continuous distributions probability *density function* ($f(x)$) is defined in A interval:

$$P(X \in A) = \int_A f(x)dx$$

    - That is, it represents the distribution of the probability per unit of length in each point.
    - The following must be satisfied: $\int_{-\infty}^{\infty} f(x)dx = 1$

- In order to define random variables (discrete or continuous), *distribution function* ($F(x)$) is also used:

$$F(x) = P(X \leq x)$$

1.2. Mean value

OCW

eman ta zabal zazu
UPV EHU

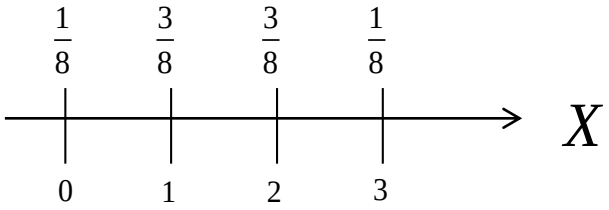
- **When the variable is discrete:**

  - Let be X a random one-dimensional variable that takes $\{x_1, x_2, \ldots, x_n\}$ values with a probability function p(x). By definition, the mean value is a numerical value calculated as:

$$E(X) = \sum_{i=1}^{n} x_i \, p(x_i)$$

    - The average value is also called mathematical expectation or expected value.

- Example: lets take the previous example of flipping a coin 3 times:

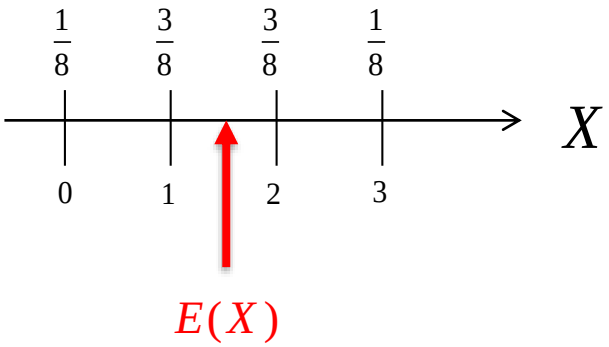


The mean value is the following:

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}$$

- It can be observed that the mean value is the **center of gravity** of the distribution mass of the variable

$$E(X)=0\cdot\frac{1}{8}+1\cdot\frac{3}{8}+2\cdot\frac{3}{8}+3\cdot\frac{1}{8}=\frac{3}{2}$$



If we assumed that the X-axis is a weightless bar in which the places masses are equal to p(x) and the bar is held in point E(X), this bar would be in balance.

OCW
eman ta zabal zazu
UPV EHU

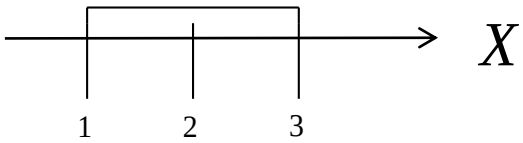
- **When the variable is continuous**:

  - Let be X random one-dimensional variable with f(x) density function. By definition, the mean value is a numerical value and is calculated as:

$$E(X) = \int_{-\infty}^{\infty} x\,\mathrm{f}(x)dx$$

    - Example: X variable is defined by the density function:

$$\mathrm{f}(x) = \begin{cases} \dfrac{x^2}{3} & -1 < x < 2 \\ 0 & \text{other cases} \end{cases}$$



OCW

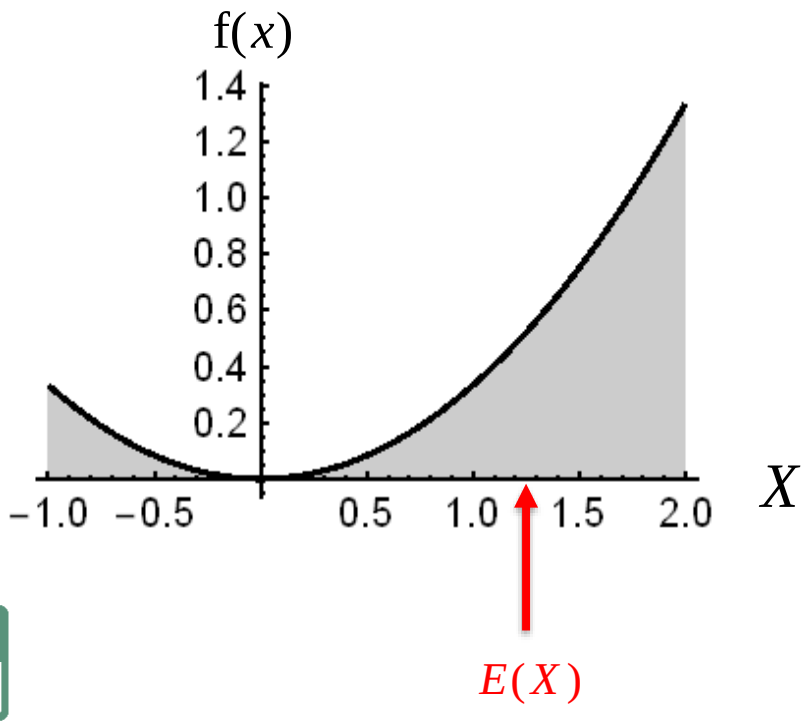
eman ta zabal zazu
UPV EHU

1.2. Mean value

- The mean value is the following:

$$f(x)=\begin{cases}\dfrac{x^2}{3} & -1<x<2\\ 0 & \text{other cases}\end{cases}$$

$$E(X)=\int_{-1}^{2} x\frac{x^2}{3}dx=\frac{5}{4}$$



In continuous variables the interpretation is the same, the mass changes continuously and E(X) is the center of gravity

OCW
eman ta zabal zazu
UPV EHU

1.3. Moments of the random variable

OCW

eman ta zabal zazu
UPV EHU

- The k^{th} statistical moment of a random variable is defined by the *level* and the *application point*.

$$\alpha_{k,a}$$

where k is the level and a is the application point.

- The statistical moment is the mean value of the difference between the variable and the application point, to the power ok k.

$$\alpha_{k,a} = E(X-a)^k$$

- When a distribution is unknown, the different characteristics can give an overview of it. The moments are these characteristics of the distribution.

OCW

eman ta zabal zazu
UPV EHU

Therefore,

- For discrete variables :

$$\alpha_{k,a} = E(X-a)^k = \sum_{i=1}^{n} (x_i - a)^k \, p(x_i)$$

- For continuous variables:

$$\alpha_{k,a} = E(X-a)^k = \int_{-\infty}^{\infty} (x-a)^k \, \mathrm{f}(x)dx$$

- The statistical moments can be classified according to the origin:
  - $a = 0$ raw moments

$$a = 0 \Rightarrow \alpha_{k,0} = \alpha_k \qquad \alpha_k = E(X)^k$$

  - $E(X)$ central moments

$$a = E(X) \Rightarrow \alpha_{k,E(X)} = \mu_k \qquad \mu_k = E(X - E(X))^k$$

  - $a \neq 0$ moments

$$a \neq 0 \ \wedge \ a \neq E(X) \Rightarrow \alpha_{k,a} \qquad \alpha_{k,a} = E(X - a)^k$$

1.4. Measures of position. Dispersion measures. Fisher coefficients.

OCW

eman ta zabal zazu
UPV EHU

- By quantifying the values of the moments we can obtain measures of position, dispersion or form.

MEASURES OF CENTRAL TENDENCY:

➢ **Mean**

The mean is the first raw moment.

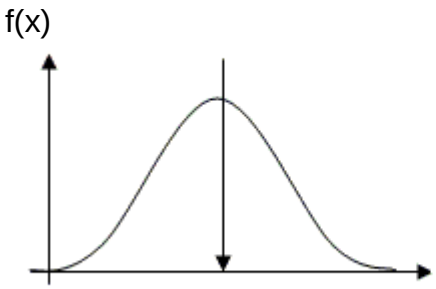
$$k = 1 \Rightarrow \alpha_{1,0} = \alpha_1 \qquad \alpha_1 = E(X)^1 = E(X)$$

As indicated, the mean represents the position of the center of gravity of the distribution mass.

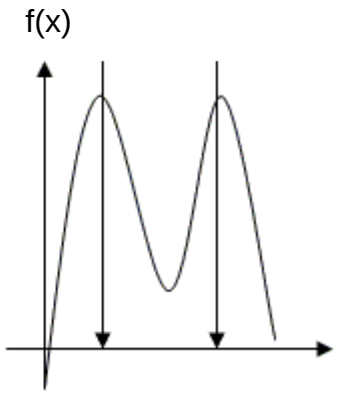
➢ Mode

Mode is a value that appears most often in a distribution.

That is, the value of the variable that makes maximum the probability function or the density function (discrete or continuous depending on the random variable). For example in a continuous case:



One mode value



Two mode values

➢ Median

The median is the value of the random variable that divides the distribution into two. The median is the value of any x that satisfies:

$$P(X \leq x) = 0.5 \quad \text{and} \quad P(X \geq x) = 0.5$$

MEASURES OF DISPERSION:

The measures of central tendency describe in some way the centre of the distribution. However, the measures of dispersion describe how disperse are the values.

- **Variance**

The variance of a random variable is the second central moment.

$$a = E(X) \Rightarrow \alpha_{2,E(X)} = \mu_2 \qquad \mu_2 = \sigma^2 = E(X - E(X))^2$$

- If the variable is continuous:

$$\sigma^2 = E(X - E(X))^2 = \int_{-\infty}^{\infty} (x - E(X))^2 \, \mathrm{f}(x) dx$$

- If the variable is discrete:

$$\sigma^2 = E(X - E(X))^2 = \sum_{i=1}^{n} (x_i - E(X))^2 \, p(x_i)$$

The variance can be interpreted as the moment of inertia of the distribution with respect to the mean. The low values of the variance indicate that the mass of the variable is concentrated around the average. However, high values of variance indicate that the mass is dispersed.

➢ Standard deviation

The standard deviation is the positive square root of the variance.

$$\sigma = \sqrt{\sigma^2}$$

FISHER COEFFICIENT

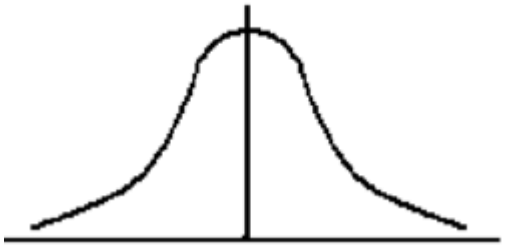
They describe the shape of the distribution.

- **Asymmetry** coefficient: skewness

It gives information about the symmetry of the distribution.

$$\gamma_1 = \frac{1}{\sigma^3} E(X - E(X))^3 = \frac{\mu_3}{\sigma^3}$$

- When the distribution is symmetrical: $\gamma_1 = 0$



- Positive skew: $\gamma_1 > 0$

- Negative skew: $\gamma_1 < 0$

OCW

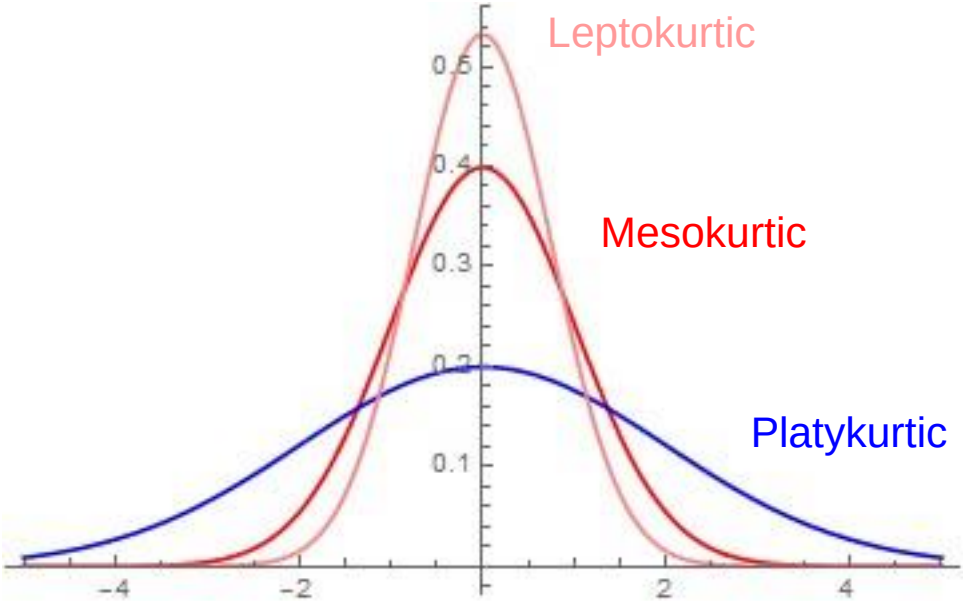
eman ta zabal zazu
UPV EHU

- **Kurtosis** coefficient

It gives information about the "peakedness" of the distribution.

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

- When the distribution is mesokurtic: $\gamma_2 = 0$
- When the distribution is leptokurtic: $\gamma_2 > 0$
- When the distribution is platykurtic: $\gamma_2 < 0$



1.5. Tchebychev's inequality

eman ta zabal zazu
UPV EHU

- Be X the random variable with E(X) mean and σ^2 variance finite values. The Tchebychev's inequality based on the differentiation of Markov is:

$$P(E(X)-k\sigma < X < E(X)+k\sigma) \geq 1-\frac{1}{k^2} \quad \forall k > 0$$

$$P(|\, X - E(X)\,| < k\sigma) \geq 1-\frac{1}{k^2} \quad \forall k > 0$$

- When the distribution of the variable is unknown, using Tchebychev's inequality we can calculate the minimum probability that the random variables are within a range. This probability value can be calculated. For this, we need to know the mean and variance of the random variable.

- Through this inequality we will obtain information only when $\forall k > 1$.

- Example: Be a random variable with mean 50 and variance 16. Calculate the minimum probability that the variable is within the range (42, 58).

$$P(E(X)-k\sigma < X < E(X)+k\sigma) \geq 1-\frac{1}{k^2} \quad \forall k > 0$$

$$P(42 < X < 58) \geq 1-\frac{1}{k^2} \quad \forall k > 0$$

$$42 = E(X)-k\sigma \qquad 42 = 50-k4 \qquad k = 2 \quad \text{or}$$

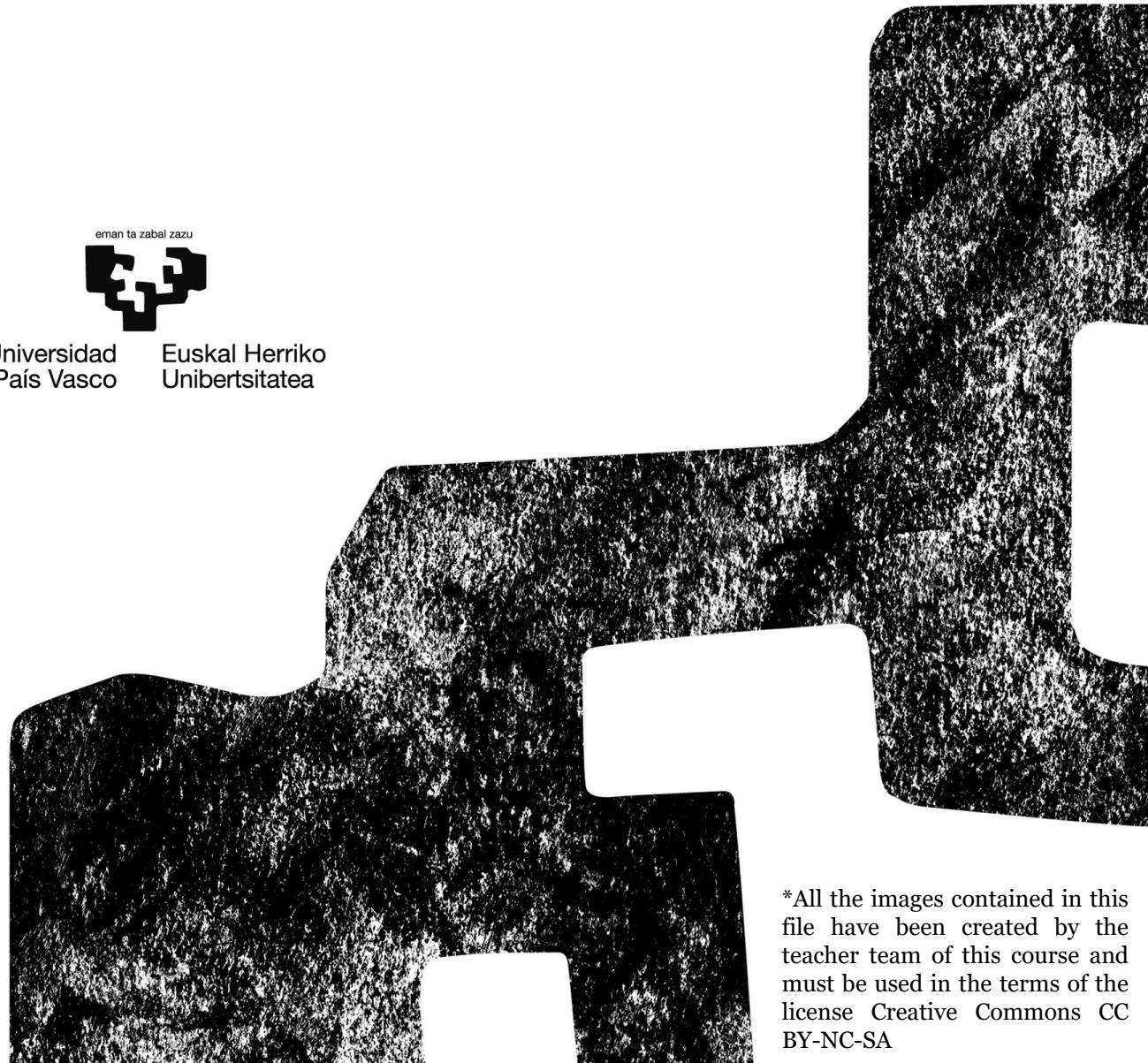
$$58 = E(X)+k\sigma \qquad 58 = 50+k4 \qquad k = 2$$

$$P(42 < X < 58) \geq 1-\frac{1}{2^2} = 0.75$$

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea



*All the images contained in this file have been created by the teacher team of this course and must be used in the terms of the license Creative Commons CC BY-NC-SA