Lecture 5.

Random variable and distribution of probability

prof. dr hab.inż. Katarzyna Zakrzewska
Katedra Elektroniki, AGH
e-mail: zak@agh.edu.pl
http://home.agh.edu.pl/~zak

Outline:

- Concept of random variable
- Quantitative description of random variables
- Sample vs. population

The concept of random variable

Random variable is a function X, that attributes a real value x to a certain results of a random experiment.

$$\Omega = \{e_1, e_2, \ldots\}$$

$$X : \Omega \rightarrow R$$

$$X(e_i) = x_i \in R$$

Examples:

1) Coin toss: event 'head' takes a value of 1; event 'tails' - 0.
2) Products: event 'failure' - 0, well-performing – 1
3) Dice: '1' – 1, '2' – 2 etc.…
4) Interval [a, b]– a choice of a point of a coordinate 'x' is attributed a value, e.g. $\sin^2(3x+17)$ etc. ….

The concept of random variable

Random variable

Discrete

When the values of random variable X are isolated points on an number line

- **Toss of a coin**
- **Transmission errors**
- **Faulty elements on a production line**
- **A number of connections coming in 5 minutes**

Continuous

When the values of random variable cover all points of an interval

- **Electrical current, I**
- **Temperature, T**
- **Pressure, p**

Quantitative description of random variables

- Probability distributions and **probability mass functions** (for discrete random variables)
- **Probability density functions** (for continuous variables)
- Cumulative distribution function (distribution function for discrete and continuous variables)
- Characteristic quantities (expected value, variance, quantiles, etc.)

Distribution of random variable

Distribution of random variable (probability distribution for discrete variables) is a set of pairs (x_i, p_i) where x_i is a value of random variable X and p_i is a probability, that a random variable X will take a value x_i

Example 5.1

Probability mass function for a single toss of coin.
Event corresponding to heads is attributed $x_1=1$; tails means $x_2=0$.

$$x_1 = 1 \quad p(X=1) = p(x_1) = \frac{1}{2}$$

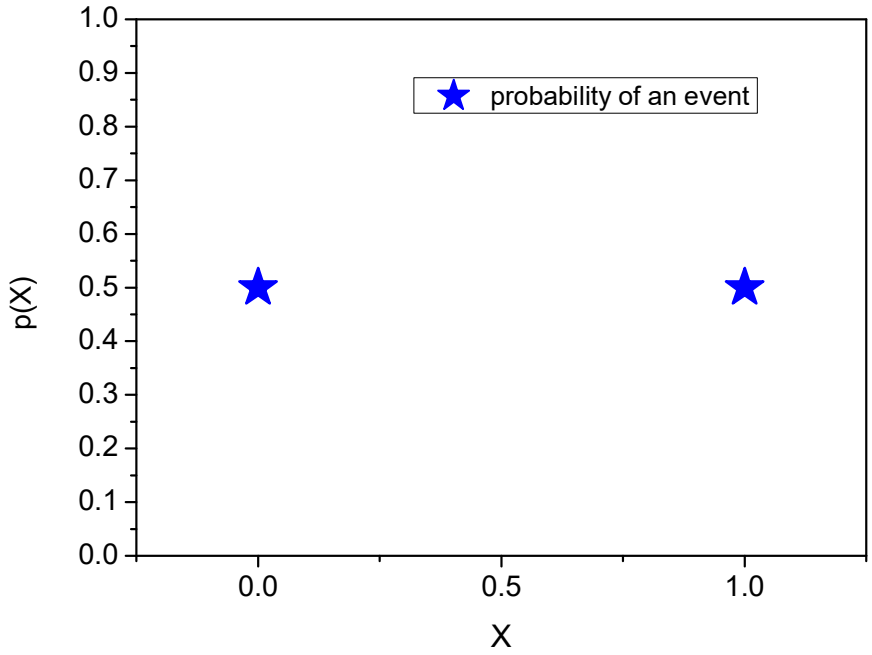
$$x_2 = 0 \quad p(X=0) = p(x_2) = \frac{1}{2}$$

Distribution of random variable

Example 5.1 cont.

Probability mass function for a single toss of coin is given by a set of the following pairs:

$$\{(1, \frac{1}{2}), (0, \frac{1}{2})\}$$



Random variable when discrete entails probability distribution also discrete.

Probability density function

Probability function is introduced for continuous variables; it is related to probability in the following way:

$$f(x)dx \equiv P(x \le X < x + dx)$$

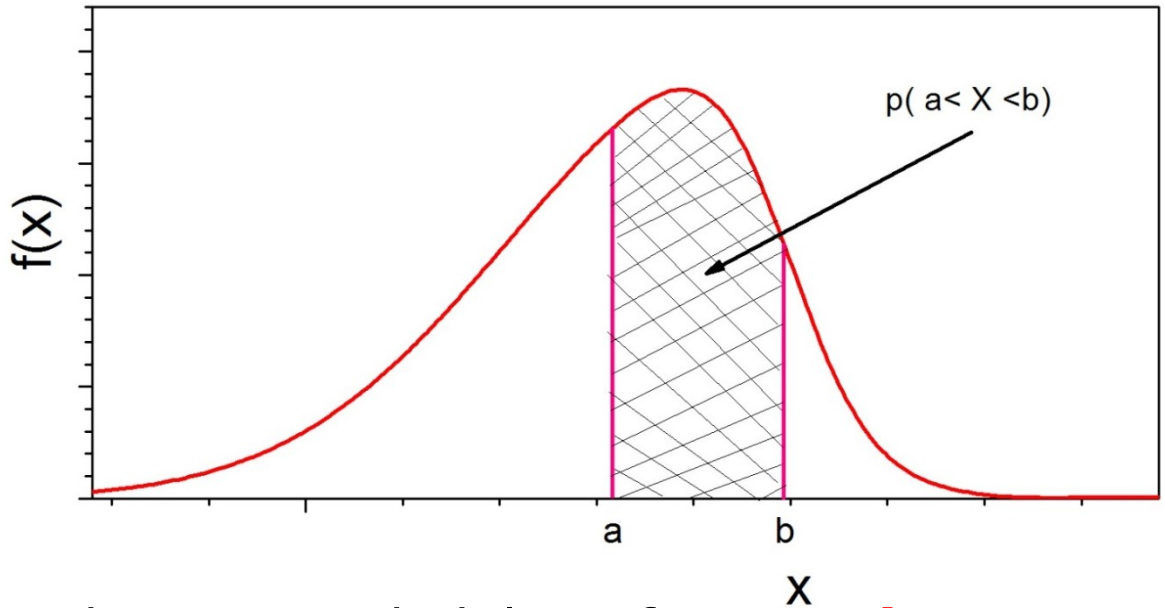
Properties of probability density function:

1. $f(x) \ge 0$
2. $f(x)$ *is normalized* $\int_{-\infty}^{+\infty} f(x)dx = 1$
3. f(x) has a measure of 1/x

Probability density function

Directly from a definition of probability density function f(x) we get a formula of calculating the probability that the random variable will assume a value within an interval of [a,b]:

$$P(a < X < b) = \int_a^b f(x)dx$$

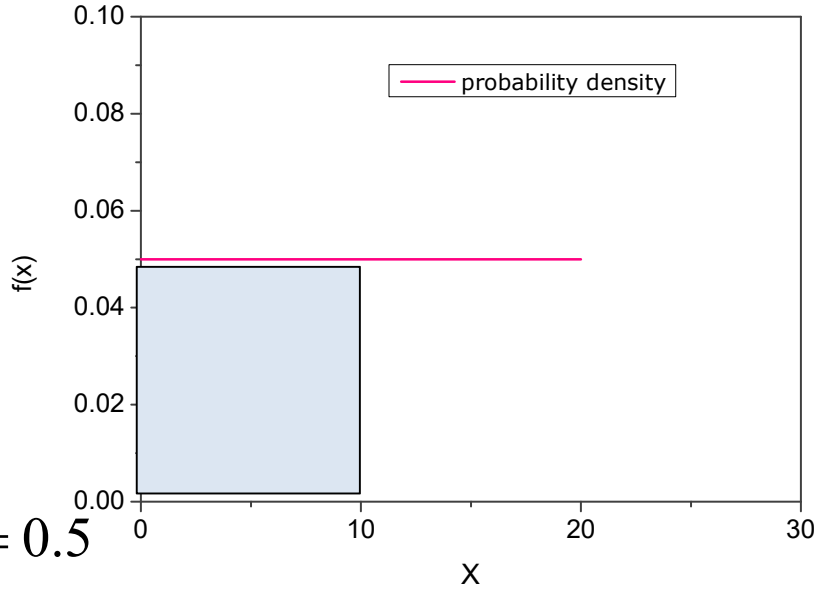


Question: what is a probability of x=a is **incorrect!!!**

Probability density function

Example 5.2

Let the continuous random variable X denote the current measured in a thin copper wire in mA. Assume that the range of X is [0, 20 mA], and assume that the probability density function of X is f(x)=0,05 for 0≤x≤20. What is the probability that a current measured is less than 10 mA.



$$P(0 \le X < 10) = \int_0^{10} f(x)dx = \int_0^{10} 0.05\, dx = 0.5$$

Quantitative description of random variables

- Cumulative distribution function (CDF) F(x) is a probability of an event that the random variable X will assume a value smaller than or equal to x (at most x)

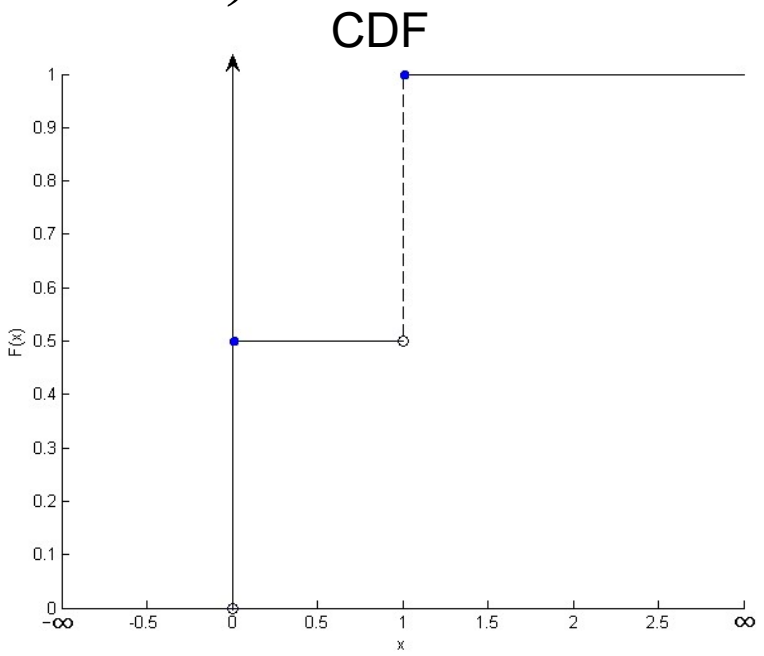
$$F(x) = P(X \leq x)$$

Example 5.1 cont.

CDF of coin toss:

$$F(x=0) = P(X \leq 0) = \frac{1}{2}$$

$$F(x=1) = P(X \leq 1) = 1$$



Properties of CDF

1. $0 \le F(x) \le 1$
2. $F(-\infty) = 0$
3. $F(+\infty) = 1$
4. $x \le y \Rightarrow F(x) \le F(y)$

non-decreasing function

5. F(x) has no unit
6. $f(x) = \frac{dF(x)}{dx}$ Relationship between cumulative distribution function and probability density (for continuous variable)

CDF of discrete variable

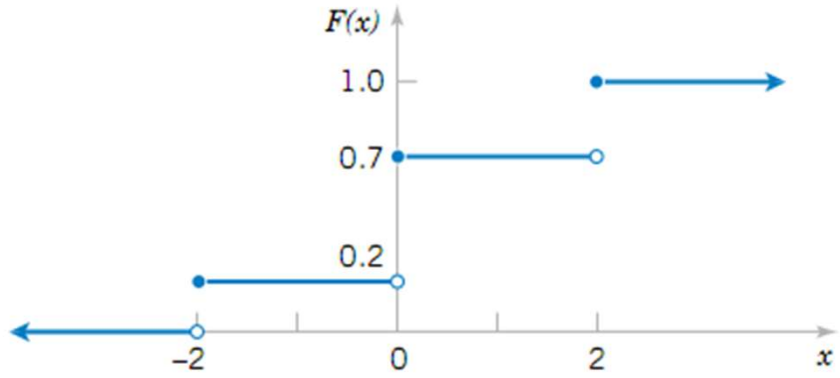
$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

f (x_i) – probability mass function

Example 5.3

Determine probability mass function of X from the following cumulative distribution function F(x)

$$F(x) = 0 \quad for \quad x < -2$$
$$0.2 \quad for \quad -2 \le x < 0$$
$$0.7 \quad for \quad 0 \le x < 2$$
$$1 \quad for \quad 2 \le x$$



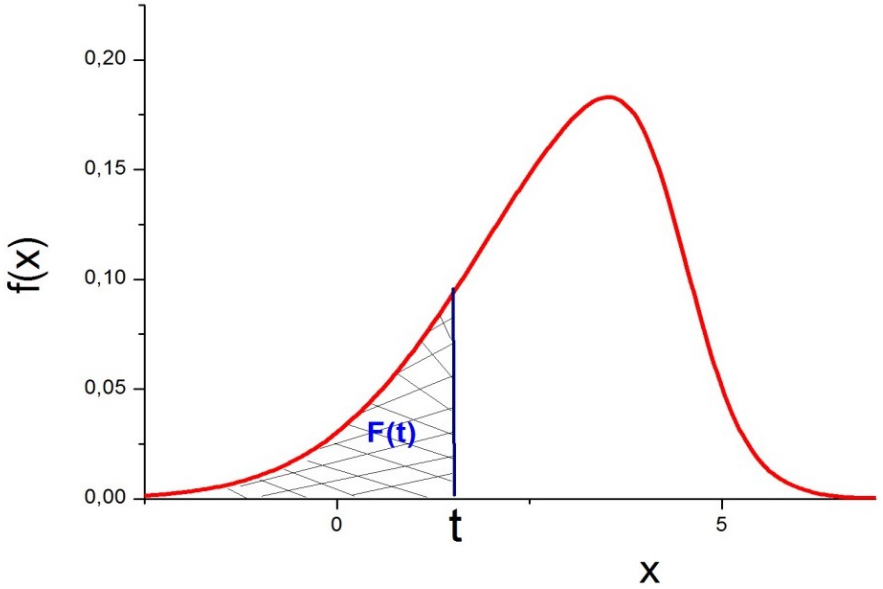
From the plot, the only points to receive f(x)≠0 are -2, 0, 2.

$$f(-2) = 0.2 - 0 = 0.2 \qquad f(0) = 0.7 - 0.2 = 0.5 \qquad f(2) = 1.0 - 0.7 = 0.3$$

CDF for continuous variable

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)\,dx$$

Cumulative distribution function F(t) of continuous variable is a non-decreasing continuous function and can be calculated as an area under density probability function f(x) over an interval from - ∞ to t.



Numerical descriptors

Parameters of

Position

- Quantile (e.g. median, quartile)
- Mode
- Expected value (average)

Dispersion

- Variance (standard deviation)
- Range

Numerical descriptors

Quantile x_q represents a value of random variable for which the cumulative distribution function takes a value of q.

$$F(x_q) = P(X \le x_q) = \int_{-\infty}^{x_q} f(u)\,du = q$$

Median i.e. $x_{0.5}$ is the most frequently used quantile.

In example 4.2 current I=10 mA is a median of distribution.

Example 5.4

For a discrete distribution : 19, 21, 21, 21, 22, 22, 23, 25, 26, 27 median is 22 (middle value or arithmetic average of two middle values)

Numerical descriptors

Mode represents the most frequently occurring value of random variable (x at which probability distribution attains a maximum)

Unimodal distribution has one mode (multimodal distributions – more than one mode)

In example 5.4: x_k= 19, 21, 21, 21, 22, 22, 23, 25, 26, 27 mode equals to 21 (which appears 3 times, i.e., the most frequently)

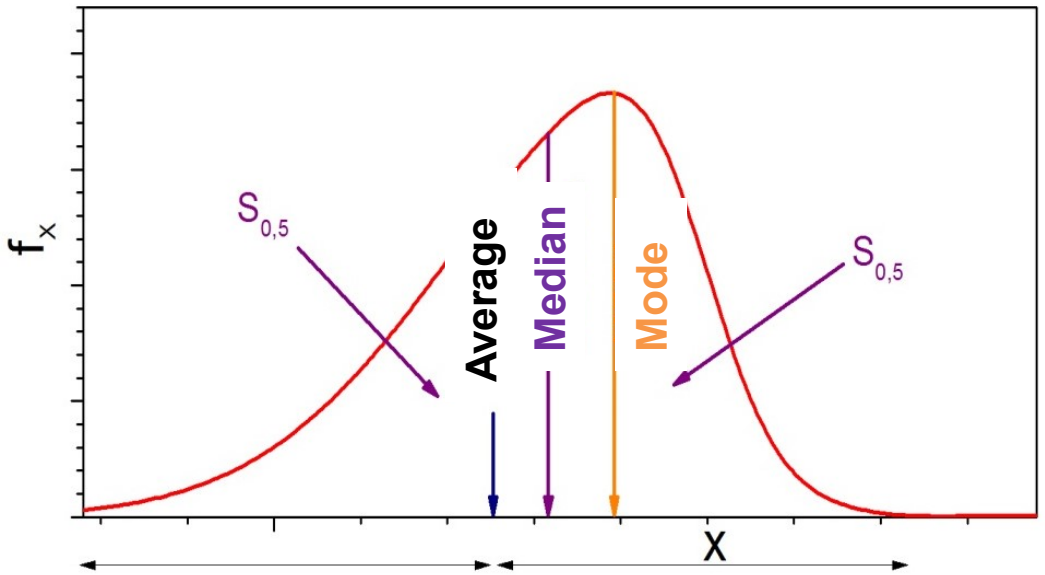
Average value

Arithmetic average:

x_i - belongs to a set of n – elements

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

In example 5.4: x_i = 19, 21, 21, 21, 22, 22, 23, 25, 26, 27, the arithmetic average is 22.7



Arithmetic average

Many elements having the same value, we divide the set into classes containing n_k identical elements

Example 5.5

x_k	n_k	f_k
10.2	1	0.0357
12.3	4	0.1429
12.4	2	0.0714
13.4	8	0.2857
16.4	4	0.1429
17.5	3	0.1071
19.3	1	0.0357
21.4	2	0.0714
22.4	2	0.0714
25.2	1	0.0357
Sum	28	

$$\bar{x} = \frac{\sum_{k=1}^{p} n_k x_k}{n} = \sum_{k=1}^{p} f_k x_k$$

where: $f_k = \frac{n_k}{n}, p - number \quad of \quad classes \ (p \le n)$

Normalization condition $\sum_k f_k = 1$

$$\bar{x} = x_1 \cdot f_1 + x_2 \cdot f_2 + \ldots + x_n \cdot f_n =$$

$$= 10.2 \cdot 0.04 + 12.3 \cdot 0.14 + \ldots + 25.2 \cdot 0.04$$

$$\bar{x} = 15.77$$

Moments of distribution functions

Moment of the order k with respect to x_0

$$m_k(x_0) \equiv \sum_i (x_i - x_0)^k \, p(x_i)$$ for discrete variables

$$m_k(x_0) \equiv \int (x - x_0)^k \, f(x)\, dx$$ for continuous variables

The most important are the moments calculated with respect to $x_0=0$ (m_k) and $X_0=m_1$ the first moment (m_1 is called the expected value) – these are central moments μ_k.

Expected value

Symbols: m_1, E(X), μ, $\bar{x}$, $\hat{x}$

$$E(X) = \sum_i x_i p_i$$ for discrete variables

$$E(X) \equiv \int x\ f(x)\, dx$$ for continuous variables

Properties of E(X)

E(X) is a linear operator, i.e.:

1. $$E(\sum_i C_i X_i) = \sum_i C_i E(X_i)$$

In a consequence:

$$E(C) = C$$

$$E(CX) = CE(X)$$

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

2. For independent variables $X_1, X_2, \ldots X_n$

$$E(\prod_i X_i) = \prod_i E(X_i)$$

Variables are independent when:

$$f(X_1, X_2, \ldots, X_n) = f_1(X_1) f_2(X_2) \cdot \ldots \cdot f_n(X_n)$$

Properties of E(X)

3. For a function of X; Y= Y(X) the expected value E(Y) can be found on the basis of distribution of variable X without necessity of looking for distribution of f(y)

$$E(Y) = \sum_i \; y(x_i)\, p_i \qquad \text{for discrete variables}$$

$$E(Y) \equiv \int y(x)\; f(x)\, dx \qquad \text{for continuous variables}$$

Any moment $m_k(x_0)$ can be treated as an expected value of a function $Y(X)=(X-x_0)^k$

$$m_k(x_0) \equiv \int (x - x_0)^k \, f(x)\, dx = E((x - x_0)^k)$$

Variance

VARIANCE (dispersion) symbols: $\sigma^2(X)$, var(X), V(X), D(X).
Standard deviation σ(x)

$$\sigma^2(X) \equiv \sum_i p_i (x_i - E(X))^2$$ for discrete variables

$$\sigma^2(X) \equiv \int f(x)(x - E(X)^2\, dx$$ for continuous variables

Variance (or the standard deviation) is a measure of scatter of random variables around the expected value.

$$\sigma^2(X) = E(X^2) - E^2(X)$$

Properties of $\sigma^2(X)$

Variance can be calculated using expected values only:

1. $$\sigma^2(X) = E(X^2) - E^2(X)$$

In a consequence we get:

$$\sigma^2(C) = 0$$

$$\sigma^2(CX) = C^2\,\sigma^2(X)$$

$$\sigma^2(C_1X + C_2) = C_1^2\,\sigma^2(X)$$

2. For independent variables $X_1, X_2, \ldots X_n$

$$\sigma^2\left(\sum_i C_i X_i\right) = \sum_i C_i^2 \sigma^2(X)$$

Czebyszew inequality

Interpretation of variance results from Czebyszew theorem:

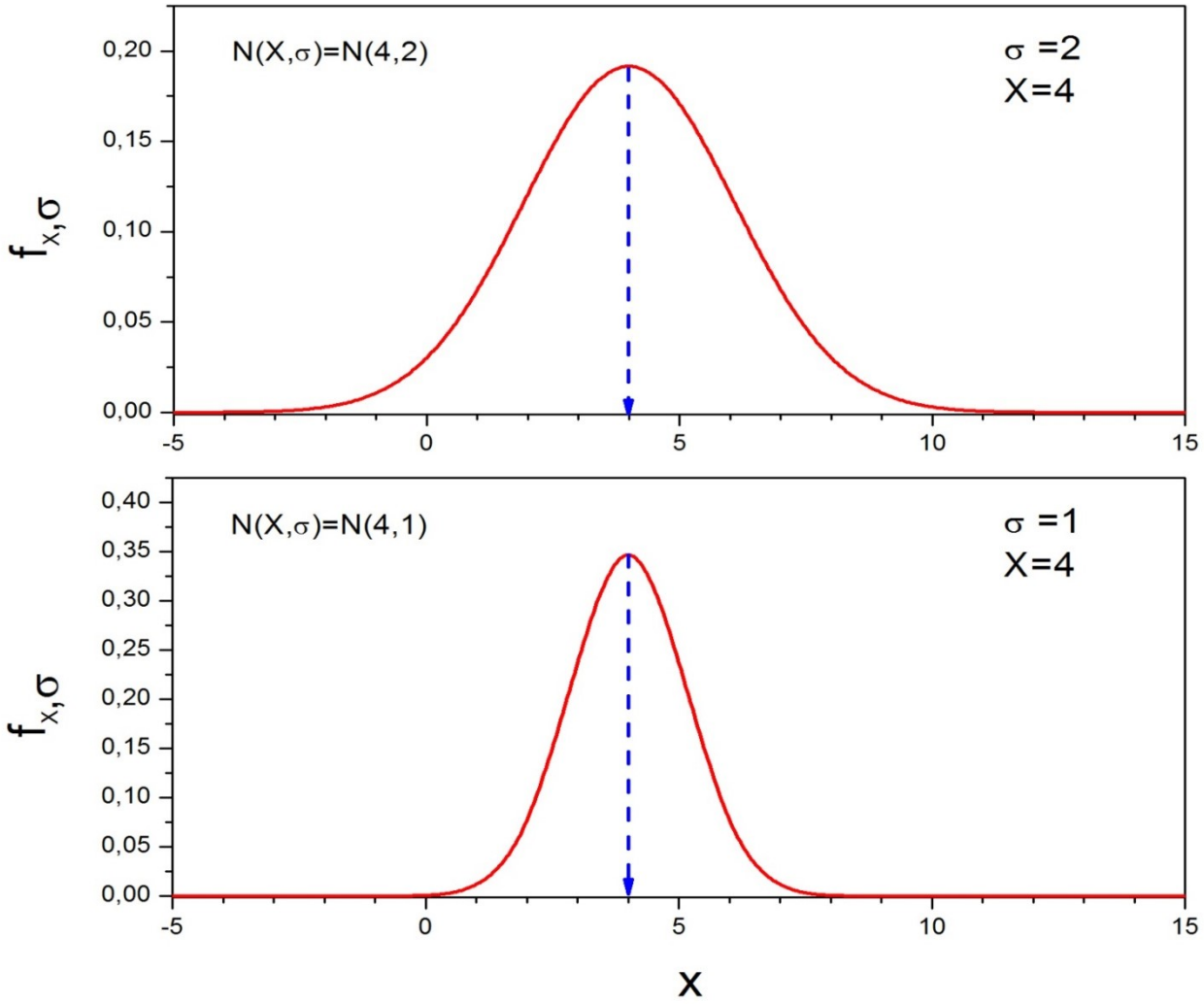
$$P\left(\left|X - E(X)\right| \geq a\,\sigma(X)\right) \leq \frac{1}{a^2}$$

Theorem:

Probability of the random variable X to be shifted from the expected value E(X) by a-times standard deviation is smaller or equal to $1/a^2$

This theorem is valid for all distributions that have a variance and the expected value. Number a is any positive real value.

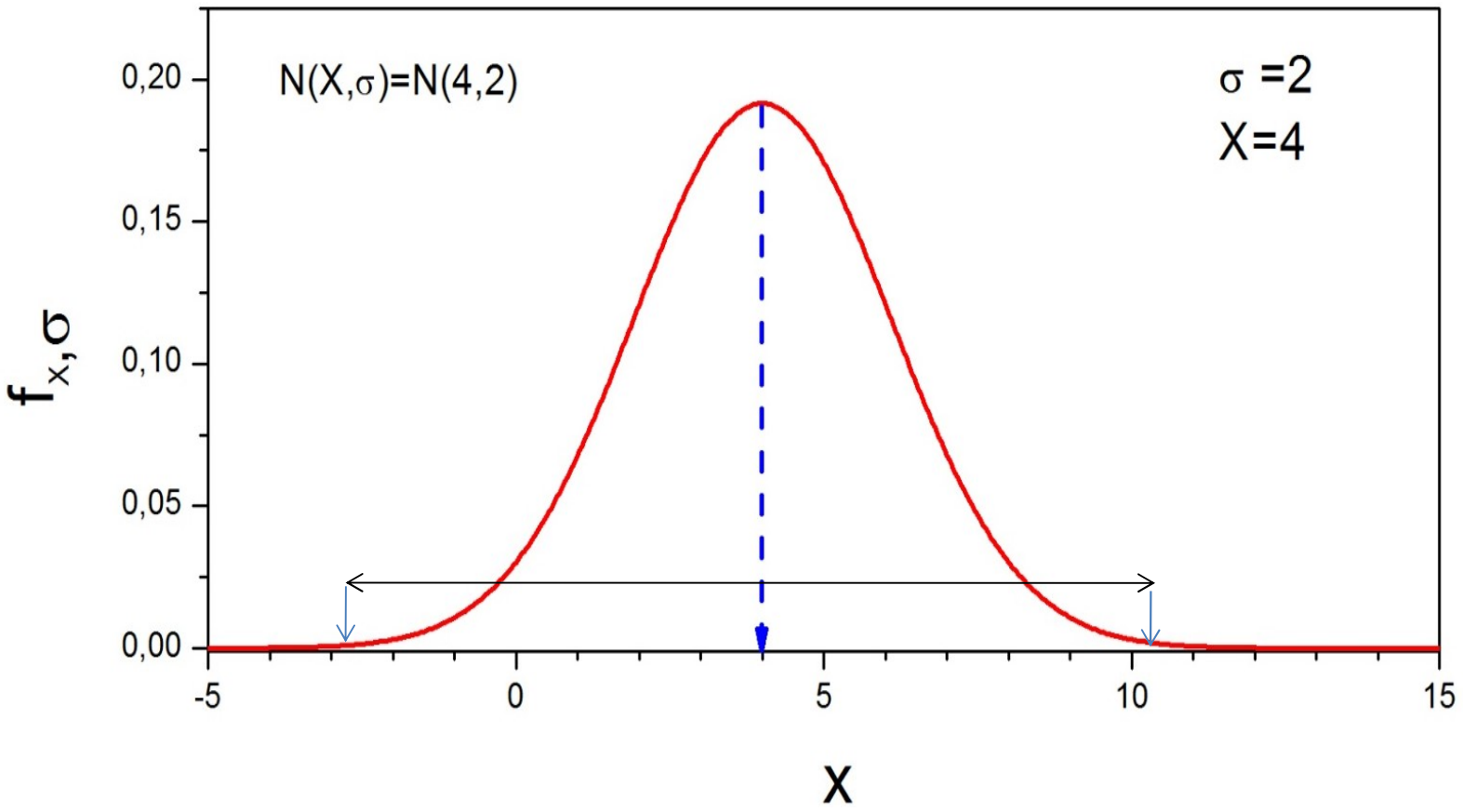
Variance as a measure of data scatter



Big scatter of data

Smaller scatter of data

Range as a measure of scatter



RANGE = x_{max} - x_{min}

Skewness and kurtosis

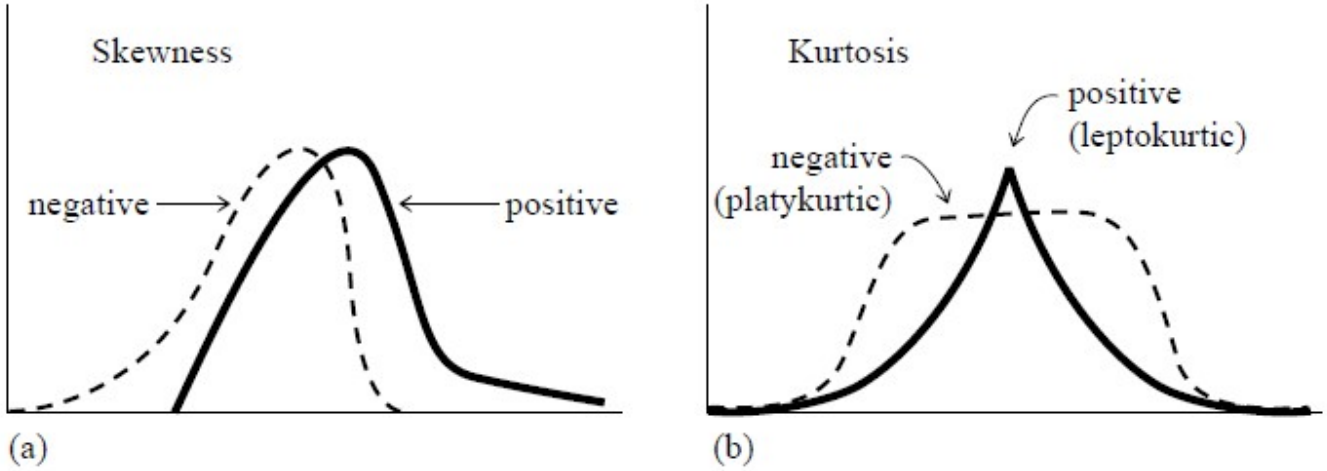


Figure 14.1.1. Distributions whose third and fourth moments are significantly different from a normal (Gaussian) distribution. (a) Skewness or third moment. (b) Kurtosis or fourth moment.

$$\text{Skew}(x_1 \ldots x_N) = \frac{1}{N}\sum_{j=1}^{N}\left[\frac{x_j - \overline{x}}{\sigma}\right]^3 \qquad \text{Kurt}(x_1 \ldots x_N) = \left\{\frac{1}{N}\sum_{j=1}^{N}\left[\frac{x_j - \overline{x}}{\sigma}\right]^4\right\} - 3$$

Sample vs. population

A population consists of the totality of the observations with which we are concerned

In any particular problem, the population may be small, large but finite, or infinite. The number of observations in the population is called the **size** of the population.
For example:

- the number of underfilled bottles produced on one day by a soft-drink company is a population of finite size,
- the observations obtained by measuring the carbon monoxide level every day is a population of infinite size.

We often use a **probability distribution** as a **model** for a population.

For example, a structural engineer might consider the population of tensile strengths of a chassis structural element to be normally distributed with mean and variance.
We could refer to this as a **normal population** or a normally distributed population

Sample vs. population

In most situations, it is impossible or impractical to observe the entire population. For example, we could not test the tensile strength of all the chassis structural elements because it would be too time consuming and expensive.
Furthermore, some (perhaps many) of these structural elements do not yet exist at the time a decision is to be made, so to a large extent, we must view the population as **conceptual.**
Therefore, we depend on a subset of observations from the population to help make decisions about the population.

A **sample** is a subset of observations selected from a population.

Sample vs. population

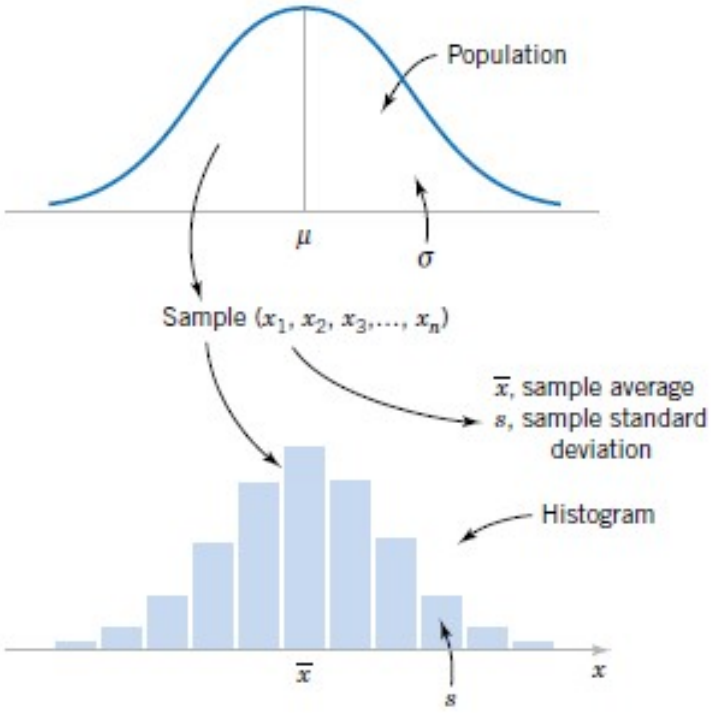


Figure 6-3 Relationship between a population and a sample.

A **statistic** is any function of the observations in a random sample.

Sample vs. population

For statistical methods to be valid, the sample must be **representative** of the population.

- It is often tempting to select the observations that are most convenient as the sample or to exercise judgment in sample selection. These procedures can frequently introduce **bias** into the sample, and as a result the parameter of interest will be consistently underestimated (or overestimated) by such a sample.
- Furthermore, the behavior of a judgment sample cannot be statistically described.

To avoid these difficulties, it is desirable to select a **random sample** as the result of some chance mechanism. Consequently, the selection of a sample is a random experiment and each observation in the sample is the observed value of a random variable. The observations in the population determine the probability distribution of the random variable.

Practical ways of calculating variance

Variance of n-element sample:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right]$$

$$\bar{x} - average$$

Variance of N-element population :

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

$$\mu - expected \quad value$$

Practical ways of calculating standard deviation

Standard deviation of sample (or: standard uncertainty):

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

Standard deviation (population):

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu\right)^2}$$