Law of averages
Statistical sampling
Estimation of expected value

Dr. Miklós BERTA
Department of Physics and Chemistry
Széchenyi István University

SZÉCHENYI ISTVÁN EGYETEM
SZE - GYŐR

Fizika és Kémia Tanszék

Measurement as a statistical sampling

- With the measurement, we take a *finite sample* from among the infinite possible values of the random variable described by the distribution function.
- Mathematical statistics seeks answers to questions such as *how likely it is to infer the distribution function of the original random variable based on finite sampling.*

Chebyshev inequality

- If the arbitrary random variable *X* has an expected value (*m*) and standard deviation (*D*), as well as any $k > 1$ real number, then the probability that *X* deviates from its expected value in absolute value by more than k times its standard deviation is at most $\frac{1}{k^2}$.

$$P(|X - m| \geq k.D) \leq \frac{1}{k^2}$$

Example: We are producing bearing balls on a production line, and we have determined that the expected diameter of the balls is 5 mm with a standard deviation of 0.1 mm. What is the maximum probability that if one of the manufactured bearing balls is selected, its diameter will be greater than 5.3 mm? What is this probability if we know that the diameter of the balls produced is a Gaussian-distributed random variable?

Solution: A diameter greater than 5.3 mm deviates from the expected value by more than 3 times the standard deviation. Based on the Chebyshev inequality, the required probability is:

$$P((d-5)\geq 0{,}3)\leq \frac{1}{3^2}=0{,}11.$$

If we assume the Gaussian distribution, then:

$$P((d-5)\geq 0{,}3)=1-F_G(3)=0{,}0013.$$

Note: The $F_G(3)$ is the value of the standard Gaussian distribution at 3. MATLAB calculates this with the normcdf(3) command.

Law of averages

- Let $X_1, X_2, \ldots . X_N$ random variables are independent, with the same expected value m and the same standard deviation D, and let ε be any positive real number. For the arithmetic mean

$$\overline{X}_N = \frac{X_1 + X_2 + \ldots . + X_N}{N}$$

is valid:

$$\lim_{n \to \infty} P(|\overline{X}_N - m| < \varepsilon) = 1.$$

Consequence

- If the value of a quantity is measured *N* times, the *arithmetic mean* of the measured values will be closer and closer to the "*true value*" of the quantity if the number *N* of measure-ments is increased.
- From the Chebyshev inequality, it is always possible to determine the maximum probability that the arithmetic mean of a given *N* number of measurements differs from the "true value" by at most k times of the standard deviation of the measured values.

Statistical estimations

- We have seen that the arithmetic mean, in a statistical sense, tends to the expected value if we increase the number of measu-rements.
- That is, the arithmetic mean is a *statistical estimate* of the expected value.
- A statistical estimate is called *unbiased* if it tends to the estimated value as N increa-ses.

Estimation of standard deviation

- The

$$s_0^{*2} = \frac{\sum_{i=1}^{N} (X_i - m)^2}{N - 1}$$

quantity is called the *corrected empirical standard deviation*, and it can be shown that *it gives an unbiased estimate of the standard deviation of the probability distri-bution*.

Estimation of the standard deviation of the arithmetic mean

- Random errors occurring during measurements are assumed to be normally distributed.
- If a quantity is measured several times, the arithmetic mean of the measured values is an unbiased estimate of the quantity to be measured.
- *Since the arithmetic mean is also calculated from measurement results, it is itself a probability variable!*

Estimation of the standard deviation of the arithmetic mean

- According to the central limit distribution theorem: The sum of N independent variables (μ_0 , σ_0) with the same probability distribution is a probability variable with a normal distribution in the limiting case $N \to \infty$, the expected value of which is $\mu = N. \mu_0$, and its standard deviation is $\sigma^2 = N. \sigma_0^2$.
- Based on the definition of the arithmetic mean:

$$\sum_{i=1}^{N} x_i = N\overline{X}_N$$

Estimation of the standard deviation of the arithmetic mean

- We know that the expected value of the former sum is:

$$\left\langle \sum_{i=1}^{N} x_i \right\rangle = N.\mu_0 = \left\langle N\overline{X}_N \right\rangle = N\left\langle \overline{X}_N \right\rangle \Rightarrow \boxed{\left\langle \overline{X}_N \right\rangle = \mu_0}$$

- We know that the square of the standard deviation of the former sum is:

$$\sigma^2\left(\sum_{i=1}^{N} x_i\right) = N\sigma_0^2 = \sigma^2(N\overline{X}_N) = N^2\sigma^2(\overline{X}_N) \Rightarrow \boxed{\sigma^2(\overline{X}_N) = \frac{\sigma_0^2}{N}}$$

<u>Note:</u>

$$\sigma^2(N\overline{X}_N) = \int_{\forall x} (N\overline{X}_N - \mu(N\overline{X}_N))^2 \rho_G(x)dx =$$

$$= N^2 \int_{\forall x} (\overline{X}_N - \mu(\overline{X}_N))^2 \rho_G(x)dx = N^2\sigma^2(\overline{X}_N)$$

Measurement theory application

- If the measured values of a quantity *X* are denoted by series (x_i), then the unbiased estimate of the expected value of the quantity to be measured is the $\overline{X}_N$ arithmetic mean of the measured values:

$$\overline{X}_N = \frac{\sum_{i=1}^{N} x_i}{N},$$

with the following standard deviation:

$$s_{\overline{X}}^{*2} = \frac{\sum_{i=1}^{N} (x_i - \overline{X}_N)^2}{N(N-1)} = \frac{s_0^{*2}}{N}.$$

Example: The length of a rod was measured several times. The measured values are as follows:

i	1	2	3	4	5	6	7	8	9
L_i [cm]	102	103	105	101	102	104	105	102	104

Estimate the length of the rod and the "error" of the measurement.

Solution: (using MATLAB)

$$L = (103.1111 \pm 0.4843)\ \text{cm}$$

Example: We measured the swing time of a pendulum. The following 15 values were obtained. Based on these, determine the swing time of the pendulum and the uncertainty of the measurement.

N	t[s]
1	29,729
2	26,711
3	27,787
4	26,457
5	27,158
6	27,200
7	28,801
8	34,318
9	28,157
10	26,264
11	27,404
12	28,274
13	28,964
14	29,513
15	28,193
Mean value	27,9
Standard deviation	0,3

Solution: (using MATLAB)