Development of Complex Curricula for Molecular Bionics and Infobionics Programs within a consortial* framework**

Consortium leader

PETER PAZMANY CATHOLIC UNIVERSITY

Consortium members

SEMMELWEIS UNIVERSITY, DIALOG CAMPUS PUBLISHER

The Project has been realised with the support of the European Union and has been co-financed by the European Social Fund ***

**Molekuláris bionika és Infobionika Szakok tananyagának komplex fejlesztése konzorciumi keretben

***A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

ELECTRICAL MEASUREMENTS

(Elektronikai alapmérések)

Uncertainty of measurements

A mérés bizonytalansága

Dr. Oláh András

Lecture 1 review

- Measurement in everyday life
- Pillars of the information technologies
- Course information
- History of measurements
- Trends in measurement technology
- Fundamentals and principles
- Modelling and measurements methods
- Structure of measuring systems
- The computer measuring systems (Labview)

Outline

- Definition and evaluation of uncertainty of measurements
- Some probabilty concepts
- Some statistics concepts
- Central Limit Theorem and its consequence
- Expanded uncertainty

Principles of measurements

- Measurement = comparision
- Each measuerement has some uncertainty!
- The measurement changes the investigated phenomenon! Fitting of the measurands and meaurement system.
- Calibration and certification

Error estimation = uncertainty of measurements

- **Error** is the difference between the measured value of a measurand and the true value of the measurand.
- The error is substituted by the uncertainty, because we also do not know the value of error (because we do not know the true value we can not determine the error)
- **Example**: If a value of a mass is given as (1.24 ± 0.13) kg, the actual value is asserted as very likely to be somewhere between 1.11 kg and 1.37 kg. The uncertainty is 0.13 kg and we note that uncertainty, like standard deviation, is a positive quantity. By contrast, an error may be positive or negative.
- In Chapter 1 we recognised that errors come in two flavours:
  - random (→Type A uncertainties)
  - Systematic (→Type B uncertainties)

Type A uncertainties

- Usually a sequence of repeated measurements giving slightly different values (because of random errors) is analysed by calculating the mean and then considering individual differences from this mean. The scatter of these individual differences is a rough indication of the uncertainty of the measurement: the greater the scatter, the more uncertain the measurement.
- The calculation of the mean, by summing the values and then dividing this sum by the number of values, is perhaps the simplest example of statistical analysis.
- However there are more sophisticated statistical methods and tools. For example: linear regression analyses for determination of temperature coefficients of resistor.

Type B uncertainties

- A Type B uncertainty may be determined by looking up specific information about a measurand such as that found in
  - a calibration report or
  - data book (instrument's specification).
- The information provided by these sources will remove the systematic error (correction/calibration) that would be present if we used only an approximate value. However, we then have to estimate the associated uncertainty ourselves, without benefit of either statistical analysis or a reported uncertainty.

Some probability concepts

- In case of „fair” dice with six sides

event	1	2	3	4	5	6
Occurence	105	99	102	96	98	100
Probability	0.175	0.165	0.17	0.16	0.1633	0.1667

Probability is a way of expressing knowledge or belief that an event will occur or has occurred.

Pr(dice=6)= ?
Pr(dice=1 or 6)= ?
Pr(dice<4)= ?

Some probability concepts (cont')

Mean

$$\mu = E(x) = \int_{-\infty}^{+\infty} x p(x)\,\mathrm{d}x.$$

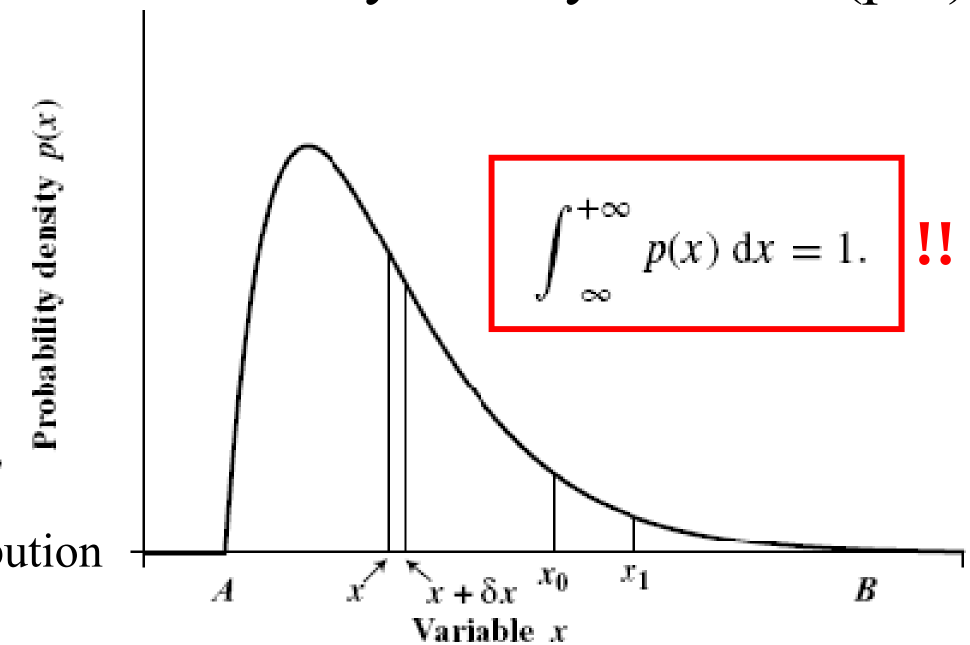
Second momentum

$$E(x^2) = \int_{-\infty}^{+\infty} x^2 p(x)\,\mathrm{d}x.$$

Variance:

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 p(x)\,\mathrm{d}x - \left(\int_{-\infty}^{+\infty} x p(x)\,\mathrm{d}x\right)^2,$$

Probability density function (pdf):



$$\int_{\infty}^{+\infty} p(x)\,\mathrm{d}x = 1.$$!!

Cumulative distribution function ↔ Probability distibution function

$$F(x) = Pr(X \le x)$$

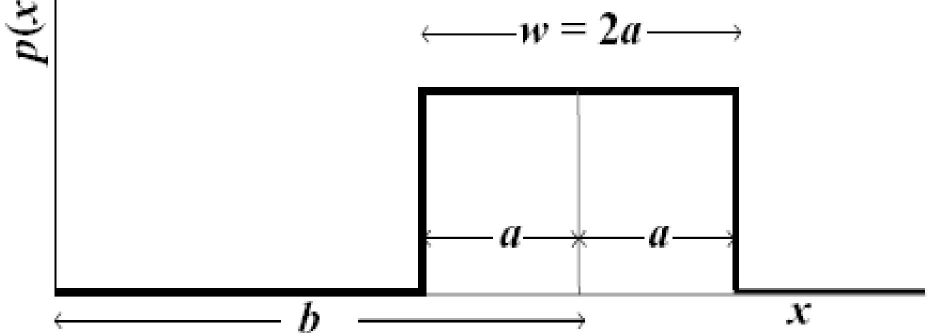
$$f(x) = \frac{dF(x)}{dx}$$

$$\Pr(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx$$

Uniform distribution

Probability density function (pdf):

$$p(x) = \begin{cases} 1/(2a), & b-a < x < b+a, \\ 0, & \text{for all other values of } x. \end{cases}$$



Mean:

$$\mu = \int_{-\infty}^{+\infty} x p(x)\,\mathrm{d}x = \frac{1}{2a}\int_{(b-a)}^{(b+a)} x\,\mathrm{d}x = \frac{1}{2a}\left[\frac{1}{2}x^2\right]_{(b-a)}^{(b+a)}$$

$$= \frac{1}{2a}\frac{1}{2}[(b+a)^2-(b-a)^2] = \frac{1}{4a}(4ba) = b.$$

Variance:

$$E(x^2) = \int_{-\infty}^{+\infty} x^2 p(x)\,\mathrm{d}x = \frac{1}{2a}\int_{(b-a)}^{(b+a)} x^2\,\mathrm{d}x = \frac{1}{2a}\left[\frac{1}{3}x^3\right]_{(b-a)}^{(b+a)}$$

$$= \frac{1}{2a}\frac{1}{3}[(b+a)^3-(b-a)^3] = \frac{1}{6a}[6b^2a+2a^3] = \underbrace{b^2+\frac{1}{3}a^2}.$$

$$\sigma^2 = b^2+\frac{1}{3}a^2-b^2 = \frac{1}{3}a^2, \quad \rightarrow \quad \boxed{\sigma = a/\sqrt{3}.}$$

Gaussian distribution (or normal distribution)

Probability density function (pdf):

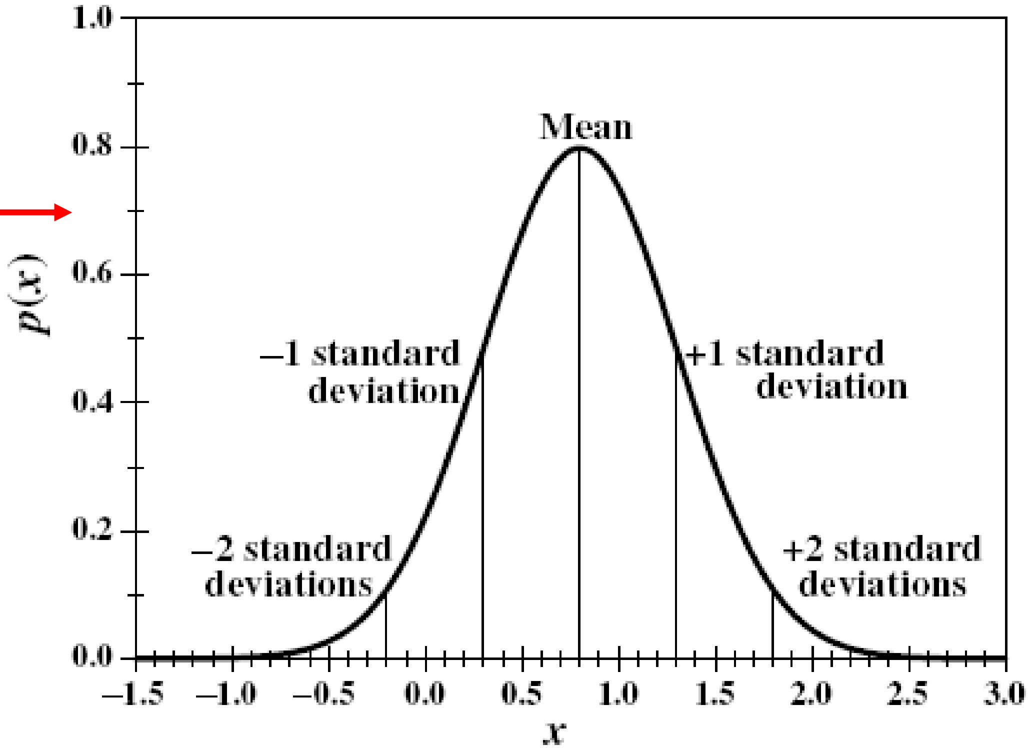
$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}.$$

Mean:

$$\int_{-\infty}^{+\infty} x p(x)\, dx = \mu$$

Variance:

$$\int_{-\infty}^{+\infty} x^2 p(x)\, dx - \mu^2 = \sigma^2.$$



Gaussian probability density with mean $\mu = 0.8$, standard deviation $\sigma = 0.5$.

Gaussian distribution (or normal distribution) (cont')

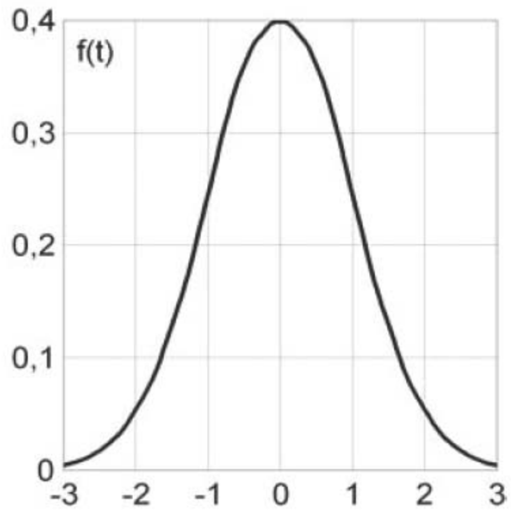
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

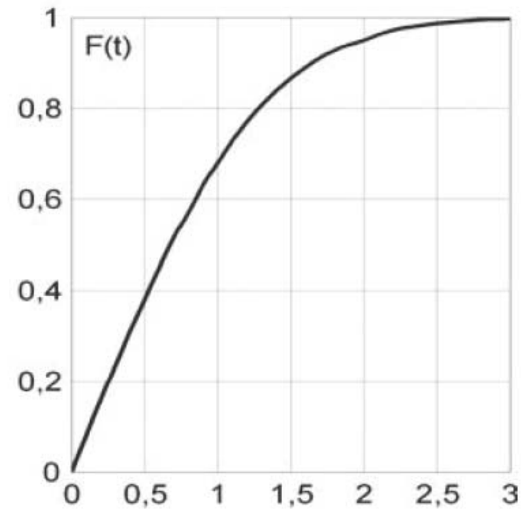
Normal distribution pdf

Standard normal distribution

$\sigma=1,\ \mu=0$

$\Phi_{st}(x)$





$$\Pr(\bar{x}-\sigma\leq x\leq\bar{x}+\sigma)=0.6826$$

Some statistical concepts

Let n denote the sample size, and x_i ($i = 1, 2, \ldots, n$) are the measured values that make up the sample, the mean and the second momentum are given by

$$E(x)$$

$$E(x_i) = \mu .$$

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

Variance and its statistics

$$E(x^2) \rightarrow s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$!!!

Relation:

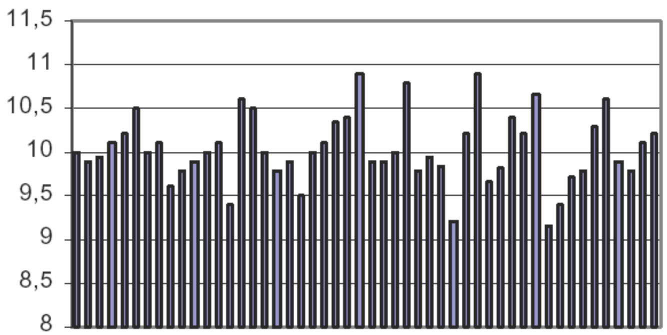
$$\sigma^2 = E(x_i^2) + \mu^2 - 2\mu^2 = E(x_i^2) - \mu^2.$$

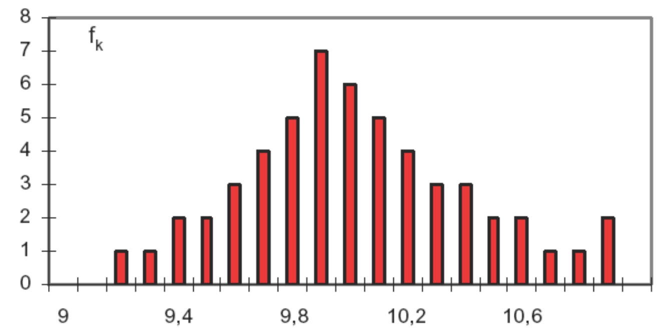
$$\sigma^2 = E[(x_i - \mu)^2]$$

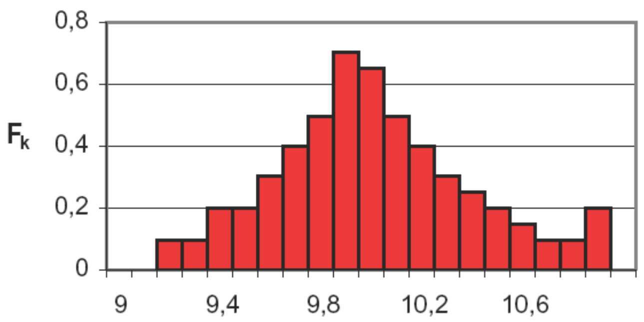
Standard deviation of avarage (for independent samples):

$$u(\bar{x}) = s/\sqrt{n}$$!!!

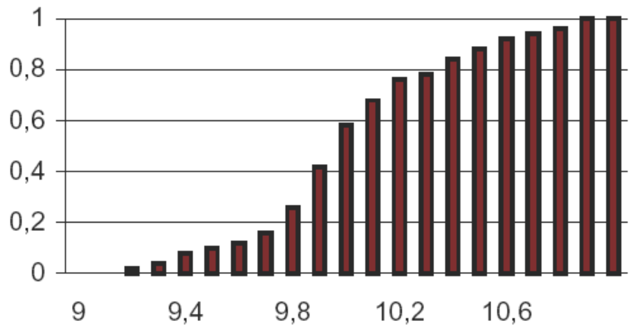
Some statistical concepts (cont')







Histogram



Cumulative histogram

$$f(x) = \frac{dF(x)}{dx}$$

$$f(x) = \Pr(x < X < x + dx) \longleftrightarrow F(x) = Pr(X \leq x)$$

Example (1)

- A particular probability density can be written $p(x) = Ax$ for the range $0 < x < 2$ and $p(x) = 0$ outside this range.
  - Sketch the graph of $p(x)$ versus x.
  - Determine the constant, A.
  - Calculate the probability that x lies between $x = 1$ and $x = 1.5$.
- A population consists of ten discrete values: 3, 3, 5, 5, 5, 6, 7, 8, 8, 8. Find the
  - mean,
  - standard deviation and
  - variance of these values.

Example (2)

- Six successive measurements of the number of airborne particles within a fixed volume of air within a clean room are made. The table shows the values obtained. Use these data to calculate
  - the variance and
  - the standard uncertainty in the number of particles.

Number of particles
137
114
88
102
95
102

Example (3)

- Ten samples of an oxide of nominally the same mass are heated in an oxygen-rich atmosphere for 1 hour. The mass of each sample increases by an amount shown in table. Using the data in table, calculate the variance and the standard uncertainty of the mass gain. Using the data in table, calculate the mean mass gain and standard uncertainty in the mean.
- Solutions: variance=0.305mg2, standard uncertainty=0.552mg mean= 11.85mg, uncertainty in the mean=0.17mg

Mass gain (mg)
12.5
11.2
11.8
11.8
12.1
11.5
11.0
12.1
11.7
12.8

Example (4)

- The thickness of an aluminium film deposited onto a glass slide is measured using a pro- filometer. The values obtained from six replicate measurements are shown in table. Using these data, calculate the variance and standard uncertainty in the film thickness. Using the data in table, calculate the mean thickness of the aluminium film and the standard uncertainty in the mean.
- Solutions: variance=906.7nm2, standard uncertainty=30.1nm, mean= 423nm, uncertainty in the mean=12nm

Thickness (nm)
420
460
400
390
410
460

Covariance and correlation

Covariance:

$$\text{covariance}(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n - 1},$$

$$\text{covariance}(x, y) = E[(x_i - \mu_x)(y_i - \mu_y)].$$

$$\begin{aligned}\text{covariance}(x, y) &= E(x_i y_i) - \mu_y E(x_i) - \mu_x E(y_i) + \mu_x \mu_y \\ &= E(x_i y_i) - \mu_x \mu_y\end{aligned}$$

Correlation:

$$r = \frac{\text{covariance}(x, y)}{\sqrt{\text{variance of } x \times \text{variance of } y}}$$

$$r = \frac{E\left[(x_i - \mu_x)(y_i - \mu_y)\right]}{\sigma_x \sigma_y} = \frac{\text{covariance}(x, y)}{\sigma_x \sigma_y},$$

$$r = \frac{E[(x_i - \mu_x)(y_i - \mu_y)]}{\sqrt{E[(x_i - \mu_x)^2]E[(y_i - \mu_y)^2]}}.$$

$$r = \frac{\sum_{i=1}^{n}[(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}}.$$

Correlation

Correlation between two linearly related variables, without random error:

Assuming that $\bar{y} = a + b\bar{x}$,

$$y_i - \bar{y} = (a + bx_i) - (a + b\bar{x}) = b(x_i - \bar{x}).$$

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = b\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

$$r = \frac{b\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[b^2\sum_{i=1}^{n}(x_i - \bar{x})^2\right]}} = \pm 1,$$

where the sign being positive if the slope *b* is positive, and negative if *b* is negative.

Example (5)

- Show for the data in the following table that the correlation between voltage and time is $r = +0.999\,48$!

t (years)	V (µV/V)
0.79	2.2
1.89	2.5
3.17	2.8
4.62	3.2
5.96	3.5

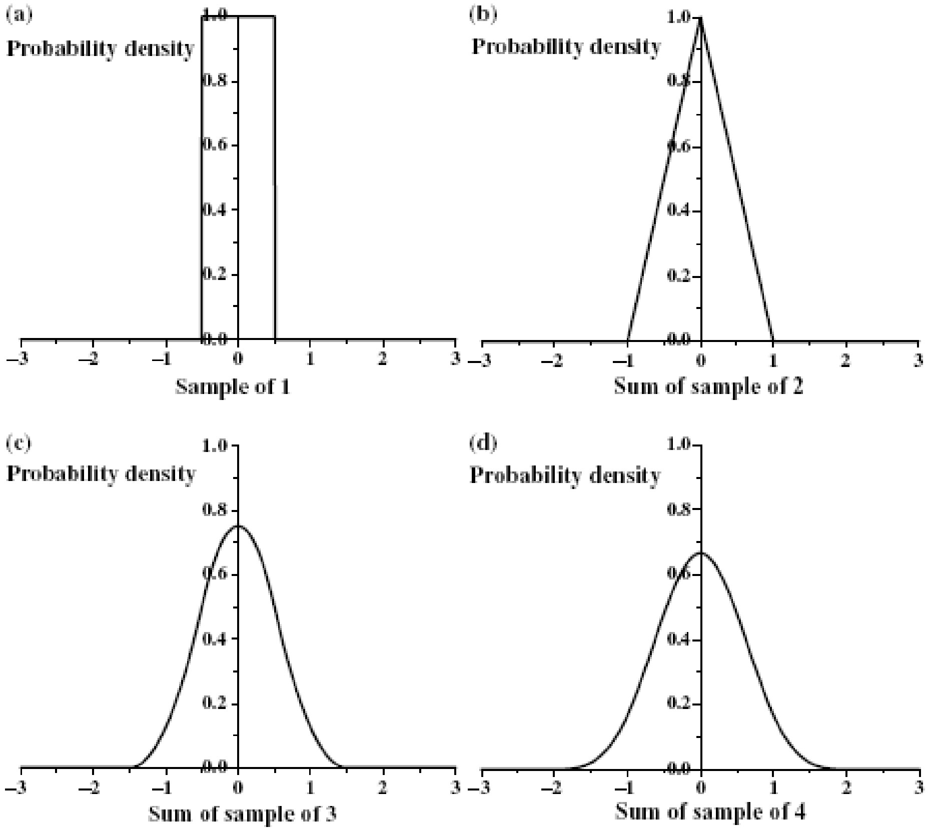
Central Limit Theorem (CLT)

Consider the case when the resultant value is composed from various values $Y = c_1X_1+c_2X_2+....$ determined with various probability distributions. In such case the **Central Limit Theorem** is helpful. This theorem states that the distribution of Y will be approximately *normal* with expected value and variance equal to

$$\mu(Y)=\sum_{i=1}^{N}c_i\mu(X_i)$$

$$\sigma^2(Y)=\sum_{i=1}^{N}c_i^2\sigma^2(X_i)$$

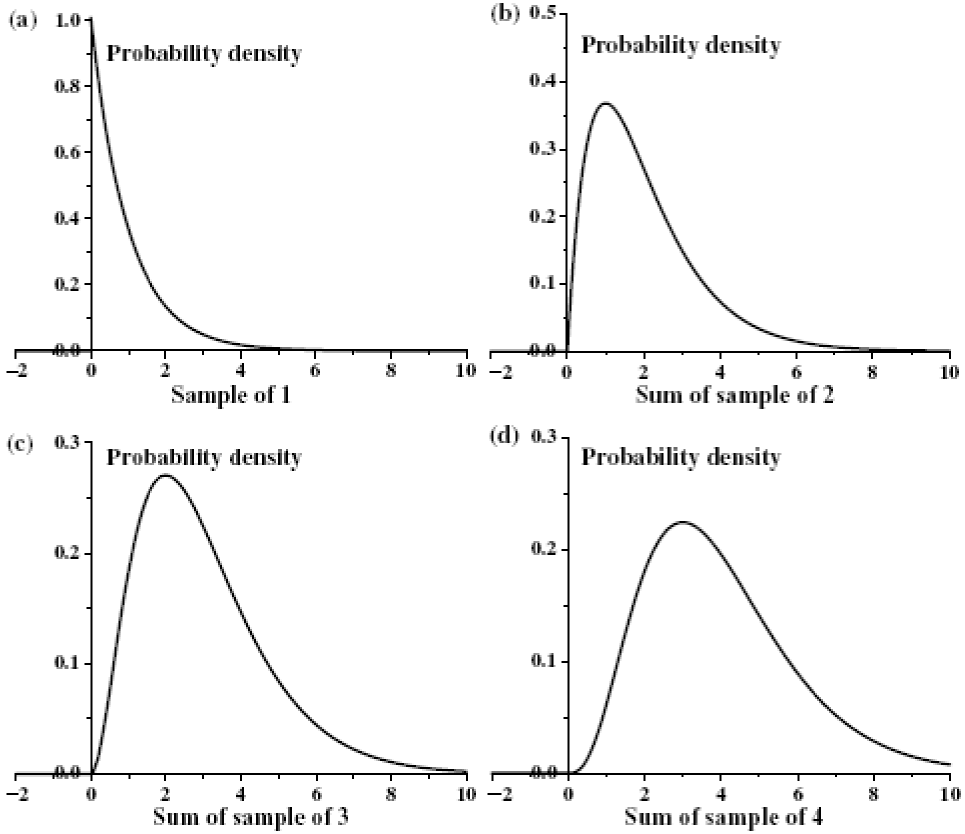
CLT example: uniform distribution



Probability density distributions of sums of samples consisting of one, two, three and four elements from a uniform distribution.

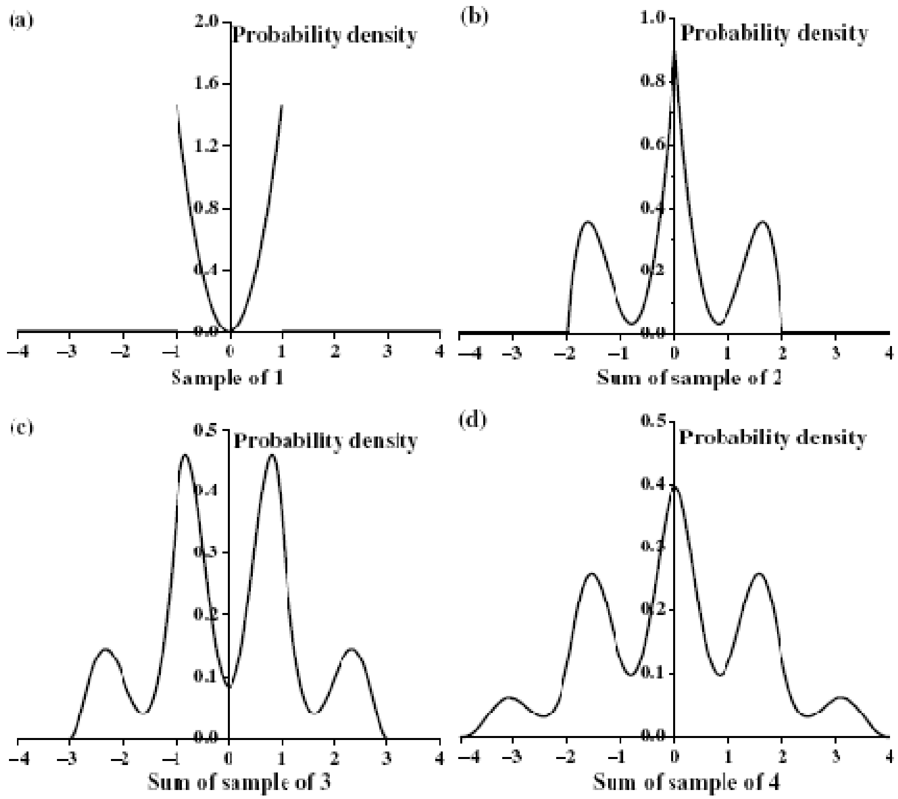
CLT example: exponential distribution

$$p(x) = \begin{cases} e^{-x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$



Probability density distributions of sums of samples consisting of one, two, three and four elements from a one-sided exponential distribution.

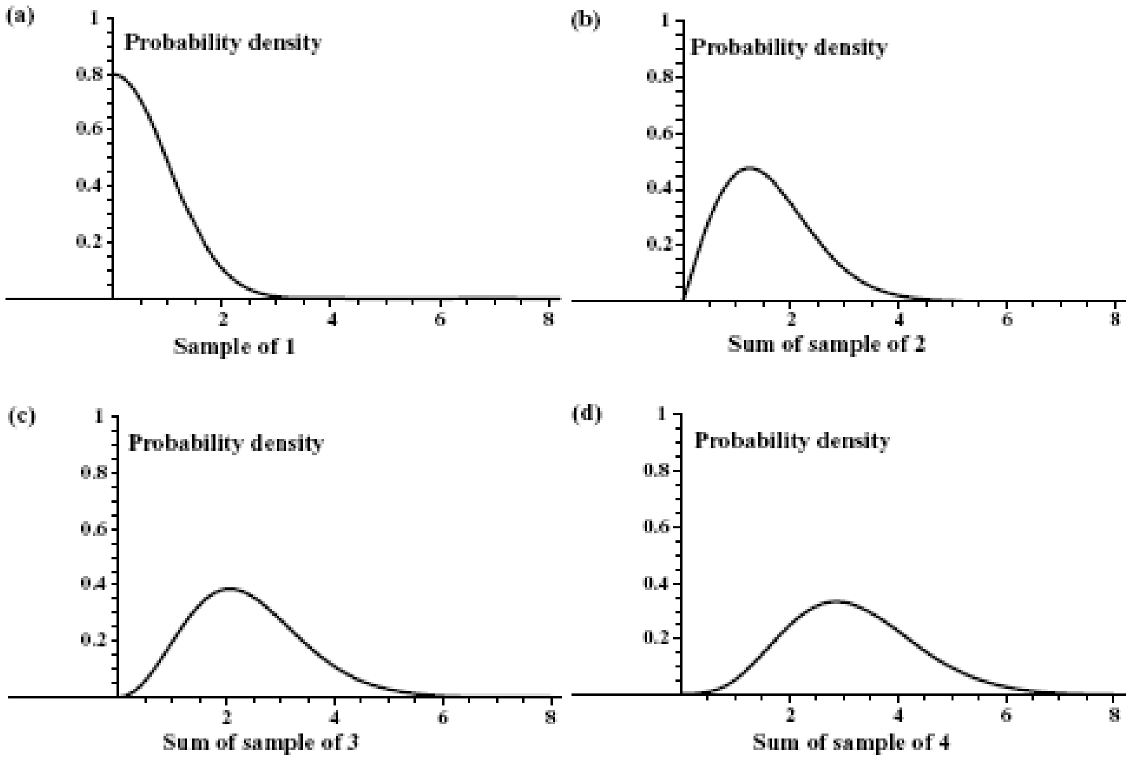
CLT example: parabolic distribution



Probability density distributions of sums of samples consisting of one, two, three and four elements from a central-dip parabolic distribution.

CLT example: truncated Gaussian distribution

$$p_{\mathrm{trunc}}(x) = \begin{cases} \sqrt{\dfrac{2}{\pi}}\mathrm{e}^{-x^2/2}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$



Probability density distributions of sums of samples consisting of one, two, three and four elements from a truncated Gaussian distribution.

Example (6)

- The probability density for a particular distribution is given by p(x) = 1 for $0 < x < +1$. For other values of x, p(x) = 0.
  - For this probability density, calculate the mean and standard deviation of the distribution.
  - Calculate the mean and standard deviation of the distribution of the mean of samples of two values drawn from this distribution. Use a uniform random-number generator (RAND function) to generate 2000 numbers in the interval 0 to 1. Taking these numbers in pairs, calculate the mean of each pair and create a column consisting of 1000 means.
  - Calculate the mean and standard deviation of the 1000 means – compare this with your answer for previous problem.

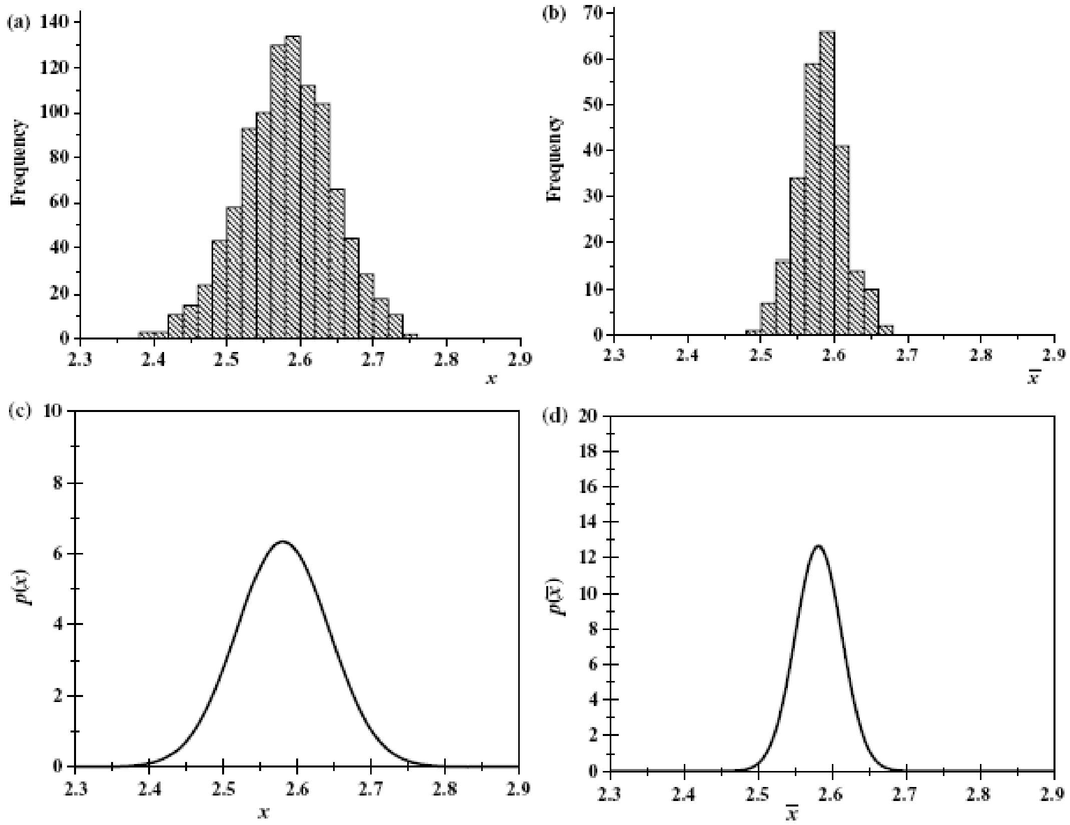
Example (6): the solution

A histogram of a software-generated Gaussian population of 1000 with assigned mean 2.5810 and assigned standard deviation 0.0630. The mean of the histogram is 2.5818; the standard deviation is 0.06277. (b) A histogram of means of 250 samples of size 4 from the population shown in (a). The mean of the histogram is 2.5818; the standard deviation is 0.031 94. The mean, ¯x, is calculated using

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

where f_i is the number of values in the i^{th} bin and x_i is the value of x corresponding to the mid-point of the i^{th} bin. (c) A Gaussian probability density distribution with mean 2.5810 and standard deviation 0.0630. (d) A probability density distribution of means of samples of size 4.

Example (6): the solution (cont')



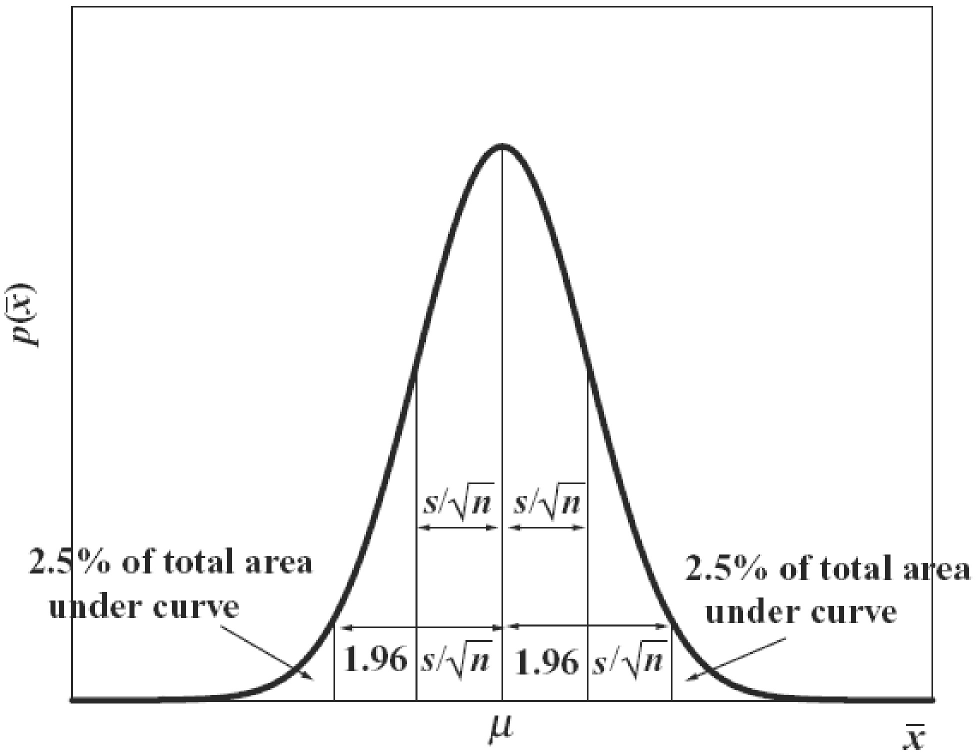
Consequence of CLT: the average can be approximated by a Gaussian random variable

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

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

$$u(\bar{x}) = s/\sqrt{n}$$

$$x = \bar{x} \pm u(\bar{x})$$



Confidence interval, the expanded uncertainty

The result of measurement $\overline{x}$ is determined with the uncertainty $\pm u$ around the estimated value $\overline{x}$ with the level of confidence (1- α):

$$\Pr\left(\overline{x} - u \le x \le \overline{x} + u\right) = 1 - \alpha$$

Standard deviation!

The current value

k=1

Avarage value

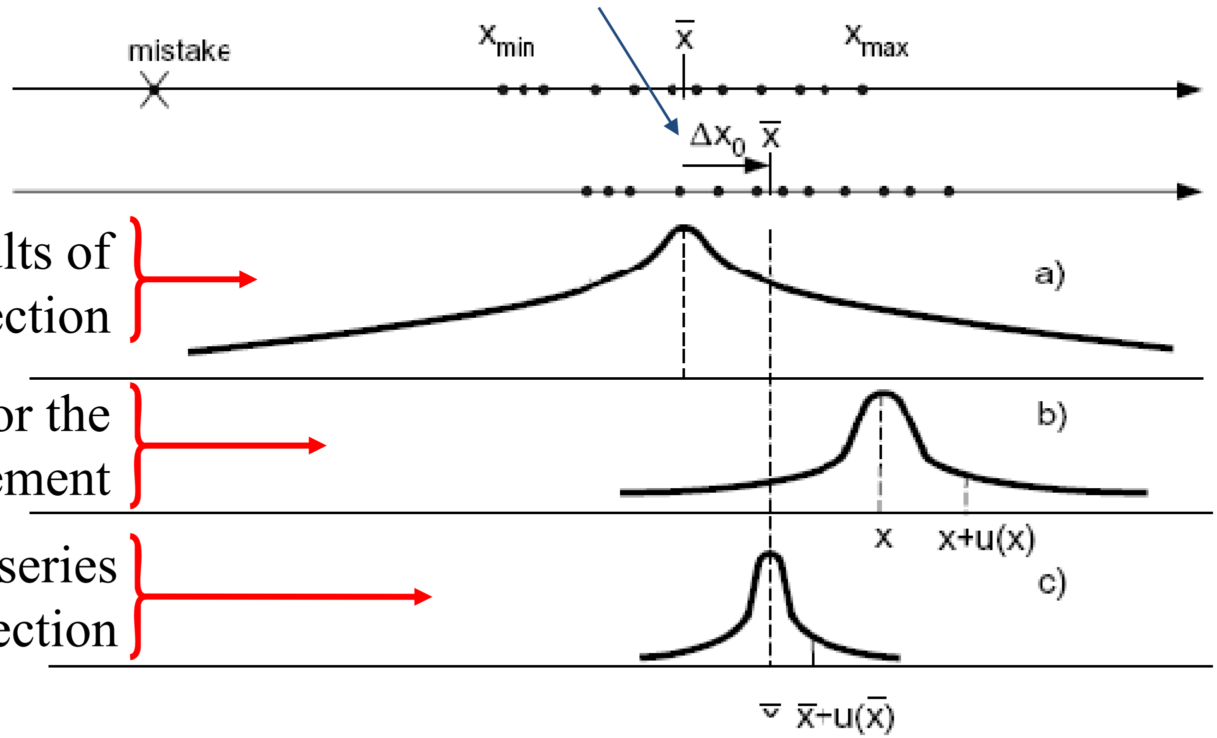
$$\Pr\left(\overline{x} - \sigma \le x \le \overline{x} + \sigma\right) = 0.6826 \qquad u = k \cdot \sigma\left(\overline{x}\right)$$

Normal distribution (according to CLT)

Thus the probability that the result of observation is in the range around the mean value (expected value) is 68.26%. Similarly, we can calculate that this probability for the dispersion ± 2 σ is 95.44% and for ± 3 σ is 99.73%. We can say that the result of measurement is very close to estimated value if the uncertainty is 3 σ.

„Cymbals” – Hanna Oláh (four years old)

System error (=constans offset)



pdf for the results of measurement without correction

distribution function for the result of measurement

pdf of the mean value of series of measurements with correction

$$x = \bar{x} \pm u(\bar{x})$$

$$u(\bar{x}) = s/\sqrt{n} \Leftarrow s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \Leftarrow \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

uncertainty of avarage

Summary

- The errors are conveniently categorised as random or systematic, the GUM („Guide to the expression of uncertainty in measurement”.) introduces the new terms ‘Type A’ and ‘Type B’ to categorise uncertainties.
- The Type A uncertainty is necessary to perform the statistical analysis. It requires certain number of measurements.
- In probability theory, the central limit theorem (CLT) states conditions under which the mean of a sufficiently large number of independent random variables will be approximately normally distributed.
- The result of measurement is determined with the uncertainty $\pm u$ around the estimated value with a given level of confidence
- In some cases (health service, military industry, etc.) there is a need for increased accuracy. In this case the GUM proposes to substitute the standard uncertainty u by the expanded uncertainty ku.
- ***Next lecture:*** *Measurement of voltage and time*