Statistics

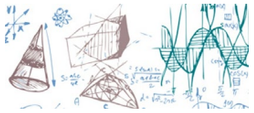
Computer Science

Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering

Agnieszka Obłąkowska-Mucha
Tomasz Szumlak

Faculty of Physics and Applied Computer Science
AGH University of Krakow

Experimental errors



- Is this statement correct from experimental point of view?

  A particle falling under the influence of gravity is subject to a constant acceleration of 9.8 m/s^2.

- For an experimental scientist this specification is incomplete.

  Better:

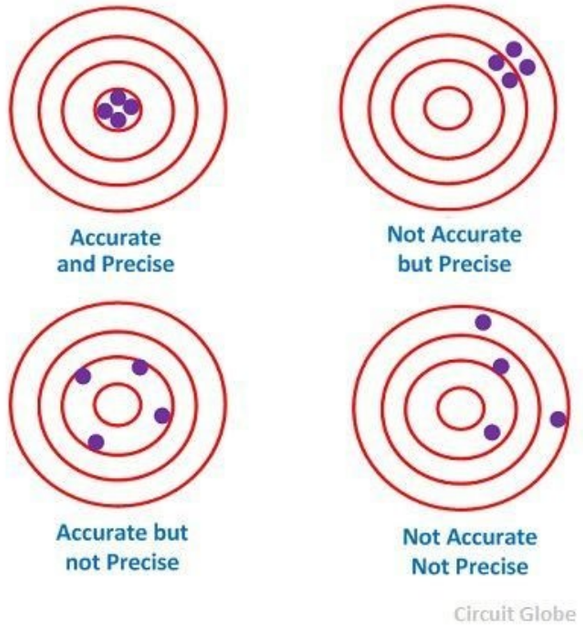
  A 5 g ball bearing falling under the influence of gravity in Room 108 was measured to be subject to a constant acceleration of 9.81 ± 0.03 m/ s^2.

- What kind of errors (we also call them uncertainties since errors sound bad) are associated with this statement?

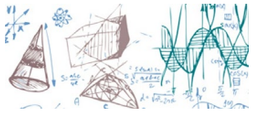
- Is this measurement:

  ✓ accurate?

  ✓ precise?



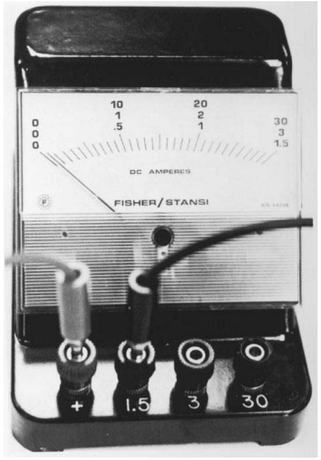
Experimental errors – two types



- Experimental uncertainty (error) generally can be classified as being of two types:
  - ✓ random or statistical error
  - ✓ systematic error
- **Statistical Error** - random errors result from unknown and unpredictable variations that arise in all experimental measurement situations.
  - ✓ There is no way to determine the magnitude or sign (+, too large; –, too small) of the error in any individual measurement.
  - ✓ Conditions in which random errors can result include:
    - Unpredictable fluctuations in temperature or line voltage.
    - Mechanical vibrations of an experimental setup.
    - Unbiased estimates of measurement readings by the observer.
- Repeated measurements with random errors give slightly different values each time. The effect of random errors may be reduced and minimized by improving and refining experimental techniques.

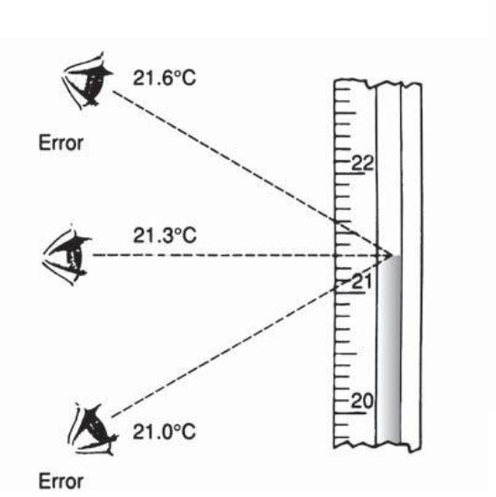
Systematic errors

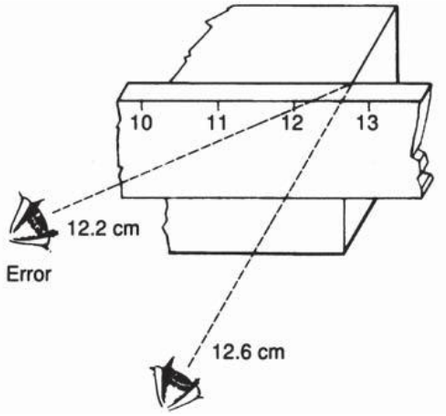
- **Systematic (Determinate) Errors**
  - ✓ Systematic errors are associated with particular measurement instruments or techniques, such as an improperly calibrated instrument or bias on the part of the observer.
  - ✓ The term systematic implies that the same magnitude and sign of experimental uncertainty are obtained when the measurement is repeated several times.
  - ✓ Determinate means that the magnitude and sign of the uncertainty can be determined if the error is identified.



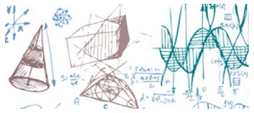
https://www.mhrd.org/cms

- Conditions from which systematic errors can result include
  - ✓ An improperly “zeroed” instrument,
  - ✓ A faulty instrument,
  - ✓ Personal error.



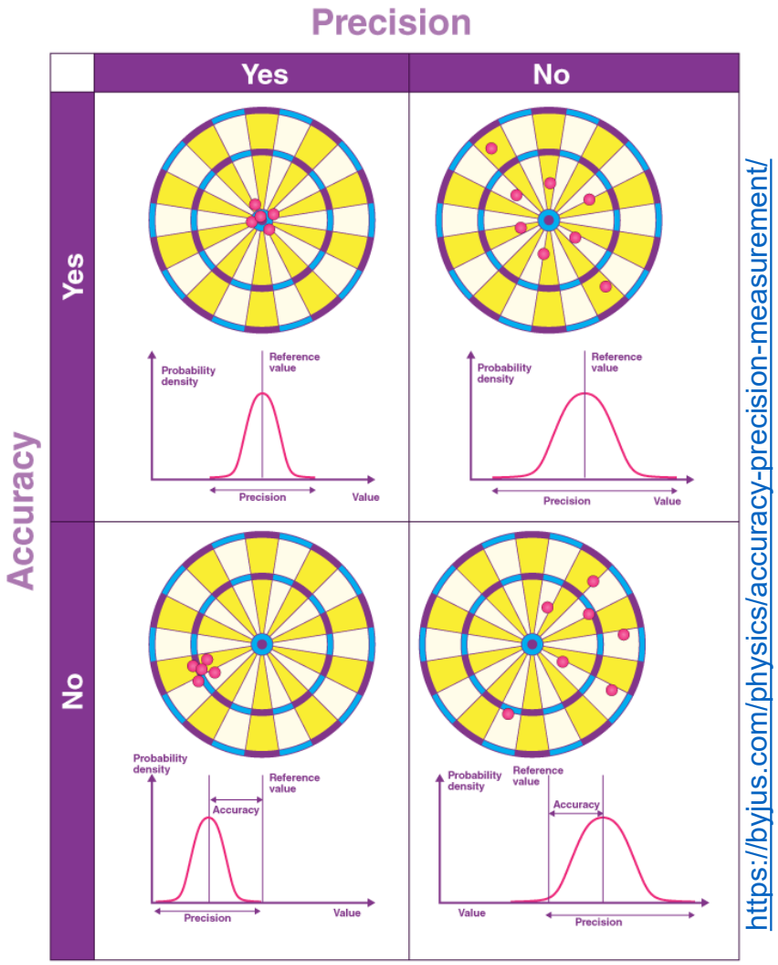


Experimental errors



- The **accuracy** of a measurement signifies how close it comes to the true (or accepted) value - that is, how nearly correct it is.
- **Precision** refers to the agreement among repeated measurements—that is, the “spread” of the measurements or how close they are together. The more precise a group of measurements, the closer together they are.

A large degree of precision does not necessarily imply accuracy.



https://byjus.com/physics/accuracy-precision-measurement/

Accuracy and precision

Example:

Two independent experiments give two sets of data with the expressed results and uncertainties of :

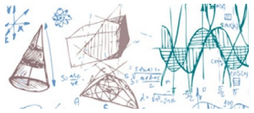
- ✓ 2.5 ± 0.1 cm
- ✓ 2.5 ± 0.2 cm.

- The first result is more precise than the second:
  - ✓ the spread in the first set of measurements is between 2.4 and 2.6 cm,
  - ✓ the spread in the second set of measurements is between 2.3 and 2.7 cm.

That is, the measurements of the first experiment are less uncertain than those of the second.

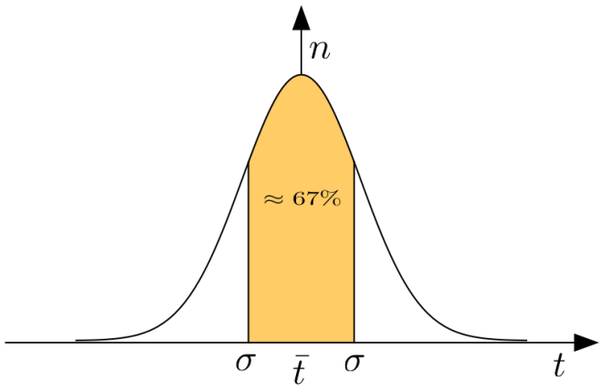
- Obtaining greater **accuracy** for an experimental value depends in general on minimizing **systematic errors**.
- Obtaining greater **precision** for an experimental value depends on minimizing **random errors.**

Estimating experimental errors

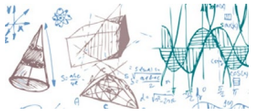


- Standard error:
  - ✓ for large number N of measurements of a single quantity t:

    σ – standard deviation (measure of the width), RMS.
- Single measurement with the result of N_k counts:
  - ✓ standard deviation is: $\sigma = \sqrt{N_k}$



Combination of errors



- Independent variables:

The measured quantity is derived from a number of measured, and usually independent variables , $x, y, z, \ldots$, connected to F, $F=f(x,y,z,.....)$.

The effect of uncertainties of the measured values $x, y, z, \ldots$, on F is:

$$\delta F = \left|\frac{\partial F}{\partial x}\right|\delta x + \left|\frac{\partial F}{\partial y}\right|\delta y + \left|\frac{\partial F}{\partial z}\right|\delta z + \ldots$$

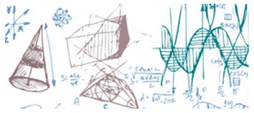
δx - the maximal error of the quantity x

the **standard error σ_F**:

$$\sigma_F = \sqrt{\left(\frac{\partial F}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial F}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial F}{\partial z}\right)^2 \sigma_z^2 + \ldots}$$

This assumes that all quantities $x,y,z,\ldots$ **are independent** of each other.

An experiment

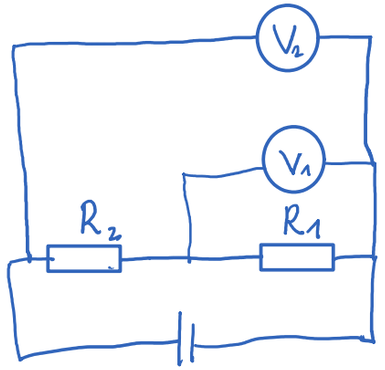


- Let’s try to analyse the following setup

  We want to measure unknown resistance R_2

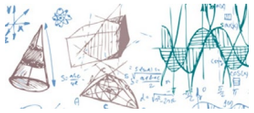
- What are the possible sources of systematic uncertainties?

  - ✓ Wrong calibration of voltmeters – may also be important if we use one or two devices
  - ✓ The calibration of the resistance meter
  - ✓ Type of power supply (AC – may introduce effects dependent on capacitances)
  - ✓ Resistors may be temperature dependent
  - ✓ Impedances of the voltmeters may be not large enough
  - ✓ Electromagnetic pick-up (ambient)



$$R_2 = \left(\frac{V_2 - V_1}{V_1}\right) R_1$$

Combining errors



- Let's consider the circuit from the previous slide and try to work out the error calculation from start to end, we assume that the following were measured:

$$R_1 = (2.2 \pm 0.1\ k\Omega) \pm 1\% \qquad V_1 = (1.0 \pm 0.02\ V) \pm 10\% \qquad V_2 = (1.5 \pm 0.02\ V) \pm 10\%$$

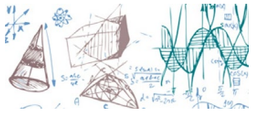
- The first error for each measurement is the **random** one, the second is the possible **systematic** error related to each meter
- Often, a single number is worked out from both errors, in that case we add them in quadrature. We get:

$$\sigma_{R_1}^{(Tot)} = \sqrt{\left(\sigma_{R_1}^{(Stat)}\right)^2 + \left(\sigma_{R_1}^{(Sys)}\right)^2} = \sqrt{(0.1)^2 + (2.21 \cdot 0.01)^2} = 0.1\ k\Omega$$

$$\sigma_{V_1}^{(Tot)} = \sqrt{\left(\sigma_{V_1}^{(Stat)}\right)^2 + \left(\sigma_{V_1}^{(Sys)}\right)^2} = \sqrt{(0.02)^2 + (1.02 \cdot 0.1)^2} = 0.10\ V\Omega$$

$$\sigma_{V_2}^{(Tot)} = \sqrt{\left(\sigma_{V_2}^{(Stat)}\right)^2 + \left(\sigma_{V_2}^{(Sys)}\right)^2} = \sqrt{(0.02)^2 + (1.52 \cdot 0.1)^2} = 0.15\ V\Omega$$

Combining errors



- So, how well we know the nominator of the circuit equation $V_2 - V_1$?
- We consider both cases without and with correlations. (BTW, we added systematic and statistical errors in quadrature assuming they are not correlated, what do you think about it?)
- Say, we used different meters – so no correlations are expected between the systematic uncertainties:

$$V_2 - V_1 = (1.50 \pm 0.15) - (1.0 \pm 0.10) = \mathbf{0.50 \pm 0.18\ V}$$

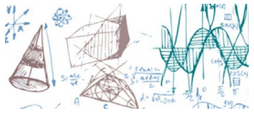
- If we use the same meter (the same systematic on result):

$$V_2 - V_1 = [(1.50 \pm 0.02) - (1.0 \pm 0.02)] \pm 10\% = (0.5 \pm 0.03) \pm 10\% = \mathbf{0.50 \pm 0.04\ V}$$

- It is convenient to remember that the systematic error can be treated as a scale factor, e.g.,: $f = 1.00 \pm 10\%$

$$M^{(True)} = M^{(Meas)} \cdot f$$

Combining errors



- So, for our case: $V_1^{(True)} - V_2^{(True)} = \left(V_1^{(Meas)} - V_2^{(Meas)}\right) f$
- There is something extremally interesting here to note! If the meter suffers from a systematic „zero error” (the zero of the scale is in fact not zero), this would have no impact on the final result (in case of subtraction)
- In turn, if we consider measuring the ratio: V_2/V_1 the scale error is now irrelevant! But the ratio is sensitive to the zero error.
- Let’s work out now the error for R_2
- We rewrite the circuit equation to be:

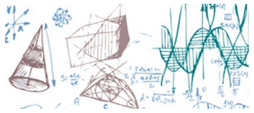
$$R_2 = \left(\frac{V_2}{V_1} - 1\right) R_1$$

- We start from the error on the voltage ratio

$$r = \frac{V_2}{V_1} = \frac{1.50 \pm 0.02}{1.0 \pm 0.02} = 1.50 \pm 0.03$$

$$\left(\left(\frac{\sigma_r}{r}\right)^2 = \left(\frac{\sigma_{V_1}}{V_1}\right)^2 + \left(\frac{\sigma_{V_1}}{V_1}\right)^2\right)$$

Combining errors



- Next step seems trivial

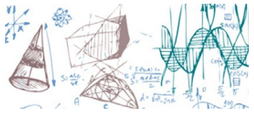
$$\frac{V_2}{V_1} - 1 = 0.50 \pm 0.03$$

- However, we should note that although the absolute errors are the same, obviously, the fractional error for this step is much larger!
- Finally, the $R_2 = 1.0 \pm 0.08\ k\Omega$ (mind the formula on the last slide)
- What about the case where the meters are different? Try to work it out yourself.
- And finally, what about using from start our formula for error propagation? (We followed it already when combining the sys and stat errors for respective variables) Try it out as well:

$$R_2 = f(V_1, V_2, R_1)$$

$$\sigma^2(R_2) = \sum_i \left(\frac{\partial R_2}{\partial X_i}\right)^2 \cdot \sigma^2(X_i)\,, X_i = V_1, V_2, R_1$$

Different experiments

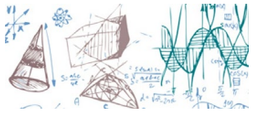


- Very often we face the following task: the same quantity was measured by different experiments with different precision. How should we calculate the combined result and its corresponding error?
- We use the following expressions (we measure R)

$$\boldsymbol{R} = \frac{\sum_i \frac{R_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}}, \qquad \boldsymbol{\frac{1}{\sigma_R^2}} = \sum_i \frac{1}{\sigma_i^2}$$

- Formally, it is so called weighted mean, where results from respective experiments are weighted by their variances (precisions).
- It is not so trivial to prove this is the best approach but we come back to this formula when discussing fitting and estimation. For now remember that we interpreted the variance as the measure of amount of information content.

Different experiments

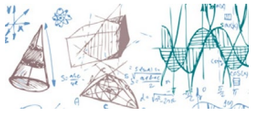


- Let's see the above in action
- Consider that all experiments feature the same precision $\sigma = \sigma_i$, then the best combined value for R would be just a simple mean (average) and it error: $\sigma_R = {}^{\sigma_i}/_{\sqrt{n_{exp}}}$. Note the difference between error for a given quantity and the error of its mean value ($\sqrt{n}$).
- Imagine we measured the same quantity using the same apparatus, but we repeated it N and M times respectively. Does the formula from the last slide still holds? Check it!

$$\bar{R}_N = \frac{1}{N}\sum_i R_i \,, \sigma_{R_N} = \frac{\sigma}{\sqrt{N}} \qquad \bar{R}_M = \frac{1}{M}\sum_i R_i \,, \sigma_{R_M} = \frac{\sigma}{\sqrt{M}}$$

$$\bar{R}_{N+M} = \frac{1}{N+M}\sum_j R_j \,, \sigma_{R_{N+M}} = \frac{\sigma}{\sqrt{N+M}}$$

More variables



- Especially important for the inference is operating with samples of measurements (data), and usually we have $\boldsymbol{n}$ of them
- Define the CDF for this case:

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n)$$

- And the PDF in this case:

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F(x_1, x_2, \dots, x_n)$$

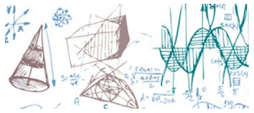
- Any marginal PDF of RV x_k

$$g_k(x_k) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_{k-1} dx_{k+1} \cdots dx_n$$

- … and the mean value for x_k

$$E[x_k] = \mu_k = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_k f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n$$

More variables



- And the same using the marginal PDF of x_k

$$E[x_k] = \mu_k = \int_{-\infty}^{+\infty} x_k g_k(x_k) dx_k$$

NICE!

- Now, in this convention let's write out the mean, variance and covariance

$$E[x_i] = \boldsymbol{\mu_i} \qquad E[(x_i - E[x_i])^2] = E[(x_i - \mu_i)^2] = \boldsymbol{\sigma_i^2}$$

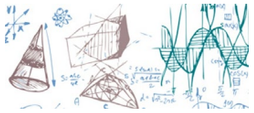
$$Cov(x_i, x_j) = E\big[(x_i - E[x_i])(x_j - E[x_j])\big] = E\big[(x_i - \mu_i)(x_j - \mu_j)\big] = \boldsymbol{c_{ij}}$$

- We can also introduce a pseudo-vector notation

$$\vec{x} = \{x_1, x_2, \dots, x_n\}, \vec{X} = \{X_1, X_2, \dots, X_n\} \qquad f(\vec{x}) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F(\vec{x})$$

Nice and compact!

More variables



- We can also put all our variances and covariances in one structure that we call **covariance matrix**

$$\mathcal{C} = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix} \qquad c_{ii} = \sigma_i^2, c_{ij} = c_{ji}$$

- Also, we can do similar thing („vectorisation”) for the means

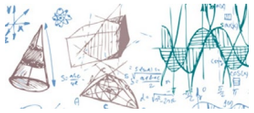
$$E[\vec{x}] = \vec{\mu} \qquad c_{ij} = E\left[(x_i - \mu_i)\left(x_j - \mu_j\right)\right]$$

$$\boldsymbol{\mathcal{C} = E[(\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^T]}$$

- The respective elements can be written explicitly (take 2 RV)

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \vec{x}^T = (x_1, x_2) \qquad \vec{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \vec{\mu}^T = (\mu_1, \mu_2)$$

More variables



- Now make the complete calculations

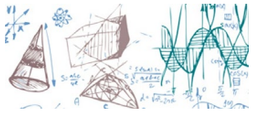
$$(\vec{x}-\vec{\mu})^T = (x_1-\mu_1, x_2-\mu_2), \vec{x}-\vec{\mu} = \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}$$

$$E[(\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^T] = \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}(x_1-\mu_1, x_2-\mu_2) =$$

$$= \begin{pmatrix} (x_1-\mu_1)(x_1-\mu_1) & (x_1-\mu_1)(x_2-\mu_2) \\ (x_2-\mu_2)(x_1-\mu_1) & (x_2-\mu_2)(x_2-\mu_2) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & c_{12} \\ c_{21} & \sigma_2^2 \end{pmatrix}$$

- We can use our new and compact notation to derive one super important rule in statistics: **error propagation formula**
- It combines variable change and multivariate functions of RV
- Interested already? Go to the next page!

Error propagation rule



- First some background… Imagine the following problem: in order to measure a quantity $\boldsymbol{y}$ (can be more than one of these) we measure $\boldsymbol{n}$ R.V.s $\boldsymbol{x_i}$:

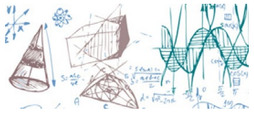
$$\vec{x} = (x_1, x_2, \dots, x_n)$$

- So, we can define a joint PDF, $f(\vec{x})$, but again usually we do not know its form but instead we can estimate respective mean values $\vec{\mu} = (\mu_1, \mu_2, \dots, \mu_n)$ and covariance matrix c_{ij}
- Ok, back to our $y(\vec{x})$. In principle we could follow the whole procedure of the variable change, but we can also just live with estimating the mean $\boldsymbol{E}[\boldsymbol{y}]$ and variance $\boldsymbol{V}[\boldsymbol{y}]$.
- The technique to be applied relies on using Taylor series expansion about the mean values of $\vec{x}$ (like asking what are the typical y value for typical x_i?)

$$y(\vec{x}) = \sum_{k/0}^{k/\infty} \frac{y^{(k)}(\vec{x}=\vec{c})}{k!}(\vec{x}-\vec{c})^k$$

$$y(\vec{x}) = y(\vec{\mu}) + \left(\frac{\partial y}{\partial x_1}\right)_{x_1=\mu_1} (x_1 - \mu_1) + \cdots + \left(\frac{\partial y}{\partial x_n}\right)_{x_n=\mu_n} (x_n - \mu_n) + \cdots$$

Error propagation rule



- We stop the expansion after the first element

$$y(\vec{x}) \approx y(\vec{\mu}) + \sum_{l/1}^{l/n} \left(\frac{\partial y}{\partial x_l}\right)_{x_l=\mu_l} (x_l - \mu_l)$$

- In principle we could expand it about any point, but we use the fact that $E[(x_l - \mu_l)] = 0$ (**sneaky!**), $\boldsymbol{E[y] \approx y(\vec{\mu})}$

$\sigma_y^2 = E[y^2] - E^2[y]$, so… we just need to know the first term

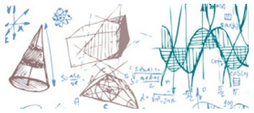
$$E[y^2(\vec{x})] \approx y^2(\vec{\mu}) + 2y(\vec{\mu}) \cdot \sum_{l/1}^{l/n} \left(\frac{\partial y}{\partial x_l}\right)_{x_l=\mu_l} E[(x_l - \mu_l)] +$$

$$+E\left[\left(\sum_{l/1}^{l/n} \left(\frac{\partial y}{\partial x_l}\right)_{x_l=\mu_l} (x_l - \mu_l)\right)\left(\sum_{j/1}^{j/n} \left(\frac{\partial y}{\partial x_j}\right)_{x_j=\mu_j} (x_j - \mu_j)\right)\right]$$

These are just numbers!

$$\sum_{l,j/1}^{l,j/n} \left(\frac{\partial y}{\partial x_l}\right)\left(\frac{\partial y}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} E[(x_l - \mu_l)(x_j - \mu_j)]$$

Error propagation rule



- Finally, we get

$$E[y^2(\vec{x})] \approx y^2(\vec{\mu}) + \sum_{l,j/1}^{l,j/n} \left(\frac{\partial y}{\partial x_l}\right)\left(\frac{\partial y}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} c_{lj}$$

Covariance matrix

$$\sigma_y^2 \approx \sum_{l,j/1}^{l,j/n} \left(\frac{\partial y}{\partial x_l}\right)\left(\frac{\partial y}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} c_{lj}$$

This is related to all $\boldsymbol{x_l}$

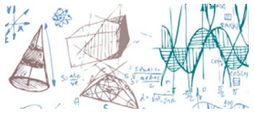
- And beyond… We can have many composite variables that depend on what we measure: $\vec{y} = \left(y_1(\vec{x}), y_2(\vec{x}), \dots, y_m(\vec{x})\right)$, so we can get a covariance matrix for all $\boldsymbol{y}$s:

$$u_{kl} \approx \sum_{i,j/1}^{i,j/n} \left(\frac{\partial y_k}{\partial x_i}\right)\left(\frac{\partial y_l}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} c_{ij}$$

Note the change of indices!

That's it! We propagate $\boldsymbol{x}$ erros to get $\boldsymbol{y}$ erros

Error propagation rule



- And an elegant matrix form

$$\mathcal{U} = \mathcal{T}\mathcal{C}\mathcal{T}^T$$

Transformation from $\boldsymbol{x} \rightarrow \boldsymbol{y}$

y uncertainty

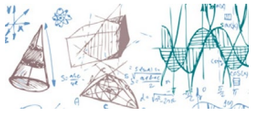
x uncertainty

$$t_{ij} = \left(\frac{\partial y_i}{\partial x_j}\right)_{\vec{x}=\vec{\mu}}$$

- Very often we deal with $\boldsymbol{n}$ measurements – which can be treated as independent R.V. (I.R.V.), the consequence is that all terms off the diagonal in the covariance matrix are 0, or we have some function that depend on $\boldsymbol{n}$ I.R.V. ($c_{ii} = \sigma_i^2, c_{ij} = 0$)

$$\sigma_y^2 \approx \sum_{j/1}^{j/n} \left(\frac{\partial y}{\partial x_j}\right)^2_{\vec{x}=\vec{\mu}} \sigma_j^2 \qquad u_{kl} \approx \sum_{j/1}^{j/n} \left(\frac{\partial y_k}{\partial x_j}\right)\left(\frac{\partial y_l}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} \sigma_j^2$$

Using the rule



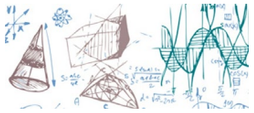
- Let's assume a very simple example: $y(\vec{x}) = x_1 + x_2$ and $y(\vec{x}) = x_1 \cdot x_2$, by applying the rule directly we have:

$$\sigma_{y+}^2 = \sum_{l,j/1}^{l,j/n} \left(\frac{\partial y}{\partial x_l}\right)\left(\frac{\partial y}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} c_{lj} = \sigma_1^2 + \sigma_2^2 + \mathbf{2c_{12}}$$

$$\sigma_{y.}^2 = y^2\left(\frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} + \frac{\mathbf{2c_{12}}}{x_1 x_2}\right)$$

- The covariance is sensitive to addition/subtraction **if RV are not independent!**
- So, we have a very nice tool to handle our data which will help us to calculate uncertainties. If the form of the transformations or the functions are not well approximated by linear formulas then our assumptions break and we should use the confidence interval instead (see future lectures!)

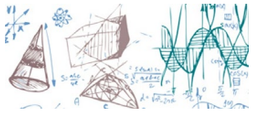
Something for ML Enthusiasts



- We saw, that when propagated errors the crucial role is played by the transformations (functions)
- One simple linear transformation is rotation in 2d space which is very popular in data analysis, computer vision etc.
- Formally we call it an orthogonal transformation
- When using ML to solve problems we find that we are having way too many variables – it would be useful to reduce them!
- One way to do that is **decorrelation**! This, as we see can be interpreted as just rotation.
- The task: we have n RVs $(x_1, x_2, \dots, x_n)$ and the covariance matrix has off-diagonal elements that are not all equal 0, we want to find a new set of RVs $(y_1, y_2, \dots, y_n)$ for which $u_{ij} = 0$
- We postulate it is always possible with a linear transformation like this

$$y_i = \sum_{j/1}^{j/n} t_{ij} x_j$$

Something for ML Enthusiasts



- Let's calculate the covariances for ys

$$u_{ij} = cov[y_i, y_j] = cov\left[\sum_{l/1}^{l/n} t_{il} x_l \, , \sum_{k/1}^{k/n} t_{jk} x_k\right] =$$

$$= \sum_{l,k/1}^{l,k/n} t_{il} t_{jk} cov[x_l, x_k] = \sum_{l,k/1}^{l,k/n} t_{il} c_{lk} t^{T}_{kj}$$

Pay attention to the indices!

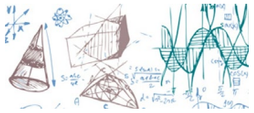
When tanspose: change the order of multiplication and inverse the indices!

- Ok guys… **we are back at the error propagation formula**!
- Our task then is to find a matrix $\mathcal{T}$ to make $\mathcal{U} = \mathcal{T}\mathcal{C}\mathcal{T}^T$ diagonal
- Very well known problem: first we need to find the e-vectors $\vec{\lambda}^{(i)}, i = 1,2, \dots, n$ of the covariance matrix $\mathcal{C}$

$$\mathcal{C}\vec{\lambda}^{(i)} = \lambda_i \vec{\lambda}^{(i)}$$

- In this procedure the e-vectors are determined up to a multiplicative factor, which can be set by requiring all $\vec{\lambda}^{(i)}$ should have unit length

Something for ML Enthusiasts



- When a matrix is symmetric the e-vecs are always orthogonal
- This is always true for the covariance matrix! So, we have

$$\vec{\lambda}^{(i)} \cdot \vec{\lambda}^{(j)} = \sum_{k/1}^{k/n} \lambda_k^i \lambda_k^j = \delta_{ij}$$

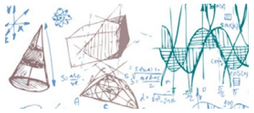
- We can proceed as follow: rows of the $\mathcal{T} (= \lambda_j^i)$ matrix are the e-vectors, and the columns of $\mathcal{T}^T (= \lambda_i^j)$ are the e-vectors, then

$$u_{ij} = \sum_{l,k/1}^{l,k/n} t_{il} c_{lk} t_{kj}^T = \sum_{l,k/1}^{l,k/n} \lambda_i^l \boldsymbol{c_{lk}} \lambda_j^k = \sum_{l/1}^{l/n} \lambda_i^l \boldsymbol{\lambda_j} \lambda_j^l = \lambda_j \vec{\lambda}^{(i)} \cdot \vec{\lambda}^{(j)}$$

$$u_{ij} = \lambda_j \delta_{ij}$$

- Variances of new RVs are expressed as e-values of the original covariance matrix $\mathcal{C}$ and $\mathcal{T}\mathcal{T}^T = 1$, thus, $\mathcal{T}^T = \mathcal{T}^{-1}$

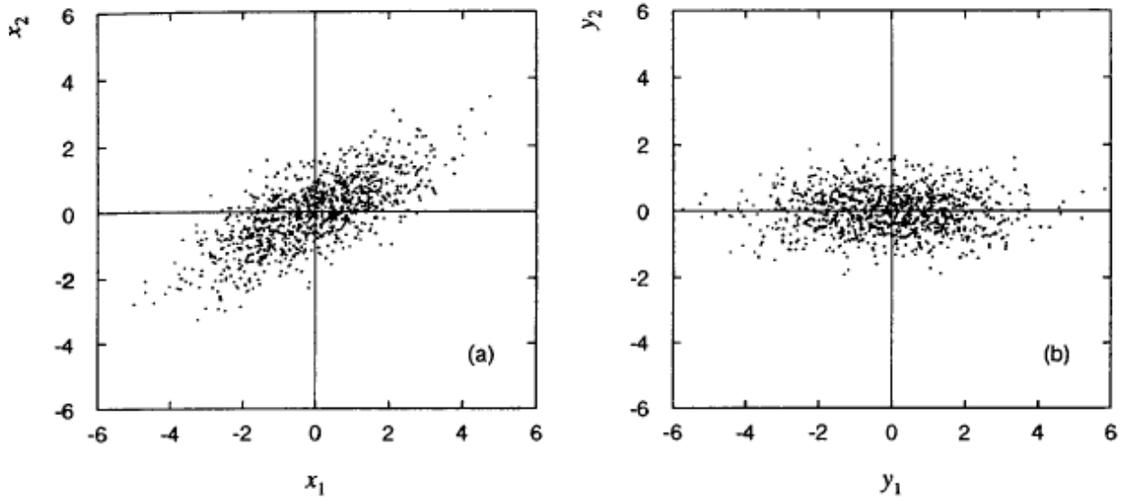
Something for ML Enthusiasts

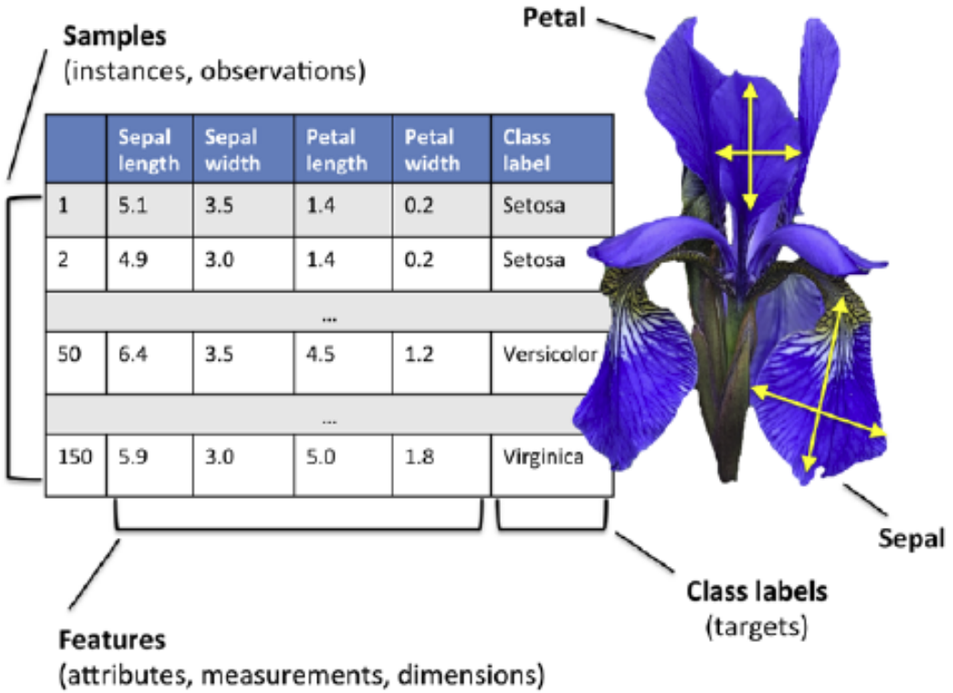


- For two dimensions this is a simple calculation, for more we just use computer programs. In the case of 2d it can be shown:

$$\mathcal{C} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \qquad \vec{\lambda}^{(1)} = \begin{pmatrix} cos\theta \\ sin\theta \end{pmatrix} \qquad \vec{\lambda}^{(2)} = \begin{pmatrix} -sin\theta \\ cos\theta \end{pmatrix}$$

$$\theta = \frac{1}{2} tg^{-1}\left(\frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}\right)$$





	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

$x^{(i)} = \left(x_1^{(i)}, \ldots, x_n^{(i)}\right)$ - one instance

$$x_j = \begin{pmatrix} x_j^{(1)} \\ . \\ . \\ . \\ x_j^{(m)} \end{pmatrix}$$

Something for ML Enthusiasts