Statistical Methods

- Tilastolliset menetelmät

A 5 ECTS credit course autumn 2025

https://www.mv.helsinki.fi/home/osterber/statistics/

Lectures Prof. Kenneth Österberg,
email: kenneth.osterberg@helsinki.fi

Exercises PhD Fredrik Oljemark
email: fredrik.oljemark@helsinki.fi

lectures Fri 10-12 in Physicum D112
exercise sessions Fri 14-15 in Physicum D112
lectures: weeks 36-39, 41-42, 44-48 and 50
exercise sessions: starting week 38

This course aims to be:

Practical but (hopefully) still **precise,**

- give recipes & examples (students can suggest)
- give merits & limitations of methods
- explain background (avoid usage w/o understanding)
- exercises have a large weight (”learn by doing”)
- (hopefully) improves understanding & eases use

aimed for data analysis typical in physics:

- measurement – parameter & uncertainty estimate
- hypothesis testing

useful & supportive in research work however only an introduction to statistical methods:

- cannot cover all methods – give the broad picture!!
- any computational tool allowed: Matlab & Python tutorials (see homepage)
- covers (upon request) student questions & problems

N.B. For statistical methods in data analysis: practical methods can’t always be proven to be optimal but can be proven to be at least sensible !!

Things can get rather fast computationally intensive

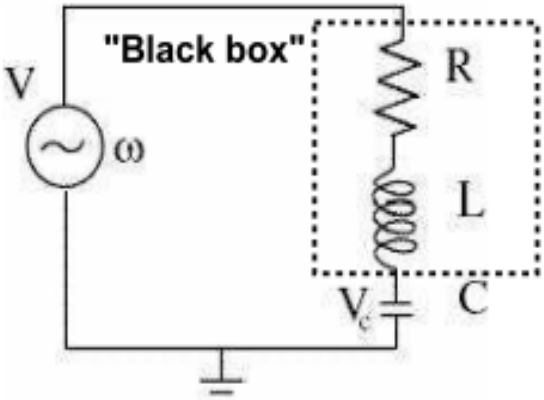
e.g. a simple electric circuit. Assume you built the circuit below ("Black box") & need to find out its resistance R & inductance L. Have to study response to the applied potential V_{AC} as function of the frequency ω. Specially uncertainties on ω, case (b) below, make things more complicated but needs to be taken into account in the determination of R & L and their uncertainties (!!).

5. An unknown electronics circuit ("Black box", see Fig. below) induces a signal measured at V_c. Given that the capacity C is known, one can determine the internal resistance R and inductance L of the "Black box" by measuring θ as function of the frequency ω:

$$\cot\theta = (L/R)\omega - (1/RC)\omega \quad \Rightarrow \quad y = \alpha_1 x - \alpha_2/x\,,$$

if $y \equiv \cot\theta$, $\alpha_1 \equiv \omega_0 L/R$, $\alpha_2 \equiv 1/(\omega_0 LR)$, $x \equiv \omega/\omega_0$ and $\omega_0 = 1$ rad/s. A measurement at five frequencies ω by connecting a known capacitor, $C = 0.02\ \mu$F, to the circuit, gave the following results:

$y \pm \sigma_y$	$x \pm \sigma_x$
-4.02±0.50	22000±440
-2.74±0.25	22930±470
-1.15±0.08	23880±500
1.49±0.09	25130±530
6.873±1.90	26390±540



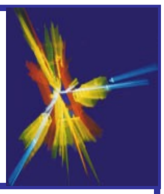
(a) Determine the L and R values (and their errors) for the "Black box" neglecting the uncertainties in x. Plot the covariance ellipse and compare extracted covariance to analytical result. What is the P-value of the fit?

(b) Determine the L and R values (and their errors) for the "Black box" taking into account both the uncertainties in x and y. Hint: Non-linear problem, solution must be found numerically. What is the P-value now?

The course content has significant overlap with PAP303 Statistical Inverse Methods, it is recommended to take only one of the two.

Course Material:

- ✓ Lecture notes
  (available on course web-page in pdf format).
- ✓ Selected lecture recordings from previous years
  (available on course web-page in mp4 format).
- ✓ G. Cowan: Statistical Data Analysis (Oxford University Press 1998)
  – highly recommended reference.
- ✓ Particle Data Group (PDG): reviews on probability, statistics, Monte Carlo techniques & machine learning (http://pdg.lbl.gov/)
  – compact & good summaries available on the web.
- ✓ Recent summaries on Interpretable machine learning & quantum computing (in particle physics)
  – additional reading on state-of-the-art methods.
- ✓ C. Walck: Handbook on statistical distributions for experimentalists
  (http://staff.fysik.su.se/~walck/suf9601.pdf)
  An opus on statistical distributions. If you know your physics distribution, useful for finding out distribution characteristics & how to generate it.



Course Outline:

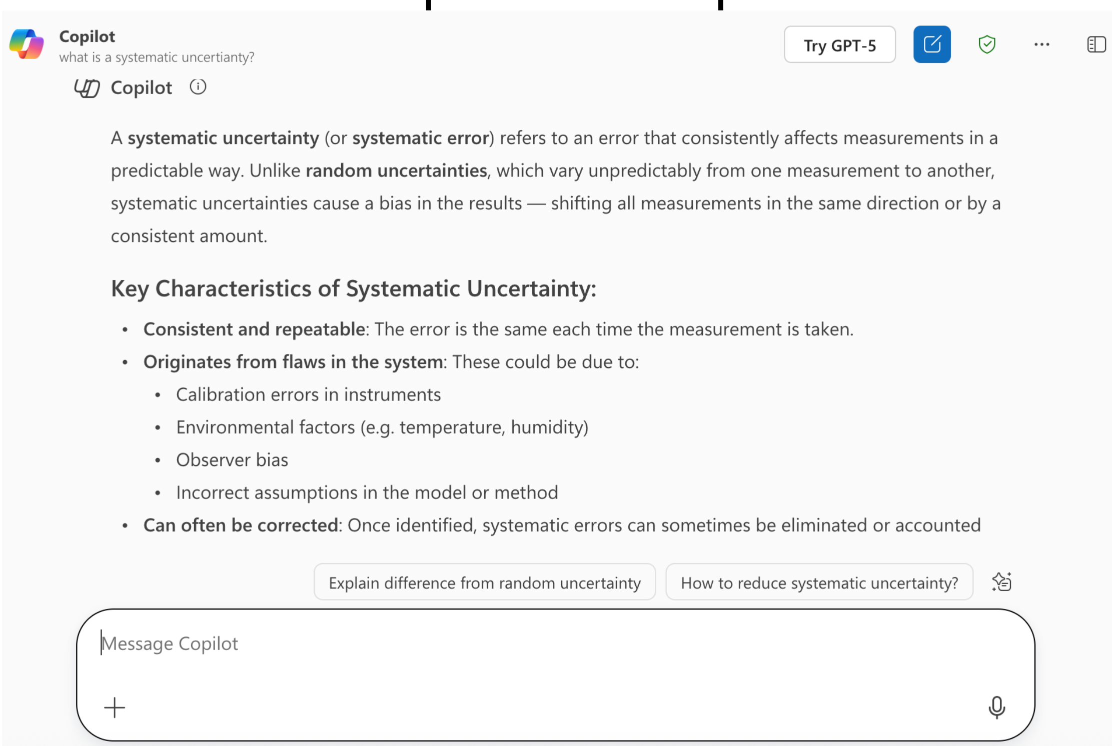
- Fundamental concepts: experimental uncertainties & their correct interpretation, frequentist & Bayesian interpretation of probability and common distributions.
- MC methods & statistical tests: Monte Carlo methods, the concept of hypothesis & test statistic, rejection of a hypothesis, discriminant analysis (including machine learning) and goodness-of-fit tests
- Parameter & uncertainty estimation: the concept of estimation, method of maximum likelihood & method of least squares
- Confidence interval & unfolding: classical confidence intervals and their interpretation, unfolding techniques

Course Grading:

- Exercises (~10 exercise papers), weight 50 %; given latest Thursday (on Moodle), to be return next Thursday 12.00.
- Home exam December/January, weight 50 % (time & date to be determined)
- General exam in February, weight 50 %
- **two best of the three used for course grade**

Use of generative AI:

- ✓ Follow University of Helsinki general rules on AI usage: https://studies.helsinki.fi/instructions/article/using-ai-support-learning?
- ✓ Usage of Large Language Models (LLMs) encouraged.
- ✓ **However: the usage of LLMs should be clearly stated including how the LLMs were used.**
- ✓ Recommendation: use Copilot provided by the university (https://copilot.cloud.microsoft/).
- ✓ **LLMs to be used as a tool** (& NOT as a black box).
- ✓ Can help you with ideas & methods (= library), coding & repetetive tasks (e.g. partial derivates).
- ✓ Always check yourself that the LLM suggestion makes sense. Explore the capabilities of LLMs.



Experiments

Basically two different types of experiments:

- ”parameter determination” – determine numerical value of some physical quantity from data.
- ”hypothesis testing” – test whether a particular model or prediction consistent with data or not.

In physics, both occur commonly. N.B. a parameter determination involves also an uncertainty determination!

- not only interested in the result but also its uncertainty.

(the smaller the uncertainty $\rightarrow$ the more accurate experiment, the more selective & conclusive the answer; often not interested in parameters themselves but in their use as a test of a model ...)

- the design & the construction of an experiment mostly driven by the accuracy that one wants to obtain.

Numerical value of uncertainty crucial for interpretation. One can (schematically) obtain essentially 3 possible results of e.g. a measurement of the mass of the proton (m_p = (1.672 621 923 69 ± 0.000 000 000 51) $\cdot 10^{-27}$ kg):

- consistent: e.g. $(1.6727 \pm 0.0002) \cdot 10^{-27}$ kg
- inconsistent: e.g. $(1.6710 \pm 0.0002) \cdot 10^{-27}$ kg
- inconclusive: e.g. $(1.8 \pm 1.0) \cdot 10^{-27}$ kg

Course should provide you with methods & tools to determine a parameter & its uncertainty correctly.

Data consistent (or not) with prediction/model? Course provides possible methods to sort that out.

degree of randomness of a result of a measurement in physics usually quantified using probability.

mathematical definition of probability:

"Kolmogorov axioms": set S ("sample space") with subsets A, B

$$\text{For all } A \subset S, \quad P(A) \geq 0$$
$$\text{If } A \cap B = \emptyset \text{ (mutually exclusive)}, P(A \cup B) = P(A) + P(B)$$
$$P(S) = 1$$

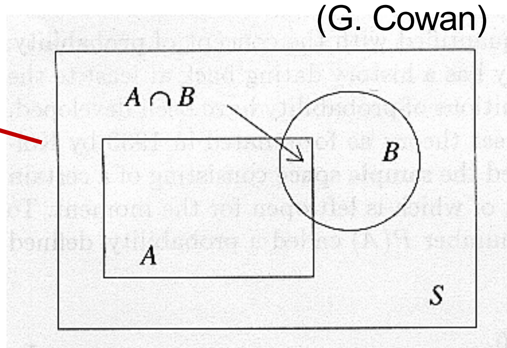
"random variable" has a specific value for each element of S.

"physical meaning" of probability P (in terms of frequency interpretation): if an element from sample space S drawn many times, obtain event A in a fraction $P(A)$ of the times
$\Rightarrow$ naturally true for all quantum mechanical phenomena.

define conditional probability of A given B ($P(B) > 0$) as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

(G. Cowan)



Subsets A and B independent if

$$P(A \cap B) = P(A)P(B)$$
$$\Rightarrow \quad P(A|B) = P(A)$$

NB! not confuse with mutually exclusive subsets i.e. $A \cap B = \emptyset$

$\Rightarrow$ Bayes' theorem:

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

from conditional probability definition since $P(A \cap B) = P(B \cap A)$
i.e. $P(A|B)P(B) = P(B|A)P(A)$

The law of total probability: Suppose S can be divided into disjoint subsets A_i such that $\cup_i A_i = S$

$$P(B) = P(\cup_i (B \cap A_i)) = \sum_i P(B \cap A_i) \Rightarrow P(B) = \sum_i P(B \mid A_i) P(A_i)$$

Modified Bayes' theorem:

$$\boxed{P(A \mid B) = \frac{P(B \mid A) P(A)}{\sum_i P(B \mid A_i) P(A_i)}}$$

N.B. Result depends on the $P(A_i)$'s, the prior probabilities!

Let's examine an example:

Suppose the probabilities (for anyone) to have AIDS are:

$$P(\text{AIDS}) = 0.0002$$
$$P(\text{no AIDS}) = 0.9998$$

prior probabilities, i.e. before any test carried out

Consider an AIDS test: result is either + or −

$$P(+ \mid \text{AIDS}) = 0.9999$$
$$P(- \mid \text{AIDS}) = 0.0001$$

probabilities to (in)correctly identify AIDS infected person

$$P(+ \mid \text{no AIDS}) = 0.0001$$
$$P(- \mid \text{no AIDS}) = 0.9999$$

probabilities to (in)correctly identify person without AIDS

Suppose your AIDS test result is +. How worried should you be?

$$P(\text{AIDS} \mid +) = \frac{P(+ \mid \text{AIDS}) P(\text{AIDS})}{P(+ \mid \text{AIDS}) P(\text{AIDS}) + P(+ \mid \text{no AIDS}) P(\text{no AIDS})} =$$

$$\frac{0.9999 \cdot 0.0002}{0.9999 \cdot 0.0002 + 0.0001 \cdot 0.9998} = 0.67$$

posterior probability

Your viewpoint: my degree of belief that I have AIDS is 67 % **subjective**

Your doctor's viewpoint: 2/3 of people like you have AIDS **frequentist**

Interpretation of probability:

• **Frequentist interpretation**

$A,B,$ ... are possible outcomes of a repeatable experiment

$$P(A) = \lim_{N\to\infty} \frac{\text{number of occurances of outcome } A \text{ in } N \text{ experiments}}{N}$$

cf. quantum mechanics, radioactive decay… **classical statistics**

• **Subjective (or Bayesian) interpretation**

$A,B,$ ... are hypotheses (statements that are true or false)

$P(A) =$ degree of belief that hypothesis A is true **bayesian statistics**

Both interpretations consistent with the mathematical definition. Probability in data analysis: frequentist more natural but subjective used e.g. for non-repeatable phenomena:

Systematic uncertainties (same upon repetition) – the universe is open – the billionth digit of π is 7 – it rains in Paris tomorrow…

Frequentist vs. subjective probability:

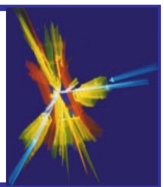
What does one mean by m_e = 520 $\pm$ 10 keV ?
(m_e is the mass of the electron)

frequentist: true or false (but one doesn't know which)
subjective (bayesian): 68 % (statement of knowledge)

i.e. $P(510 \leq m_e \leq 530$ keV$)$ = 0.68 (subjective) means:

My uncertainty that $510 \leq m_e \leq 530$ keV is same as uncertainty to draw a white ball out of container with 100 balls, 68 of which are white & rest black (cf. G. D'Agostini, CERN Yellow Report 99-03, 1999).
$\rightarrow$ **Calibration** by relating to some frequency (or symmetry etc...)

Problems of both interpretations



If a large group of Baysians say things like:

P(Finland gets a medal in ice hockey at 2026 Olympics) = 68 %
$P(0.22 \leq \Omega_M \leq 0.32)$ = 68 % (Ω_M = matter density in the universe)
P(Social Democrats wins Finnish parliamentary elections in 2027) = 68 %
then 68 % of these statements should end up being true.

NB! Calibration not always feasible, e.g.
P(Mr A. Known wins curling tournament in Ivalo January 2026) = ???

Possible frequentist interpretation: can $P(510 \leq m_e \leq 530$ keV) = 68 % mean, consider many universes with different values of m_e; 68 % of them will have m_e-values in [510,530] keV range (???)
$\Rightarrow$ usual stumbling block for the interpretation of a measurement

$P(m_e^{true}|m_e^{meas}) \propto e^{-(m_e^{true}-m_e^{meas})^2/2\sigma_{meas}^2}$ of certain result m_e^{meas}, no possibility to give any certain statements about m_e^{true}.

Note Bayes' theorem interpreted in subjective (bayesian) way:

$$P(\text{theory}\,|\,\text{result}) \propto \frac{P(\text{result}\,|\,\text{theory})}{P(\text{result})} \cdot P(\text{theory})$$

The (posterior) belief in a theory modified by the experimental result. Large **likelihood** *P*(theory|result) increases belief, small decreases. Receipe applies to many successive results.

Problem: *P*(theory|result) depends on assumed **prior** *P*(theory). Hence result not objective (as science supposed to be).

Usual assumption for a continuous parameter eg. $P(m_e)$ = flat arguments: value not known (problematic if value already ”known”), every value as believable as any other & all possibilities equal.
(statistically correct procedure is to test robustness under different priors)

Many measurements: *P*(theory)-dependence eventually removed.

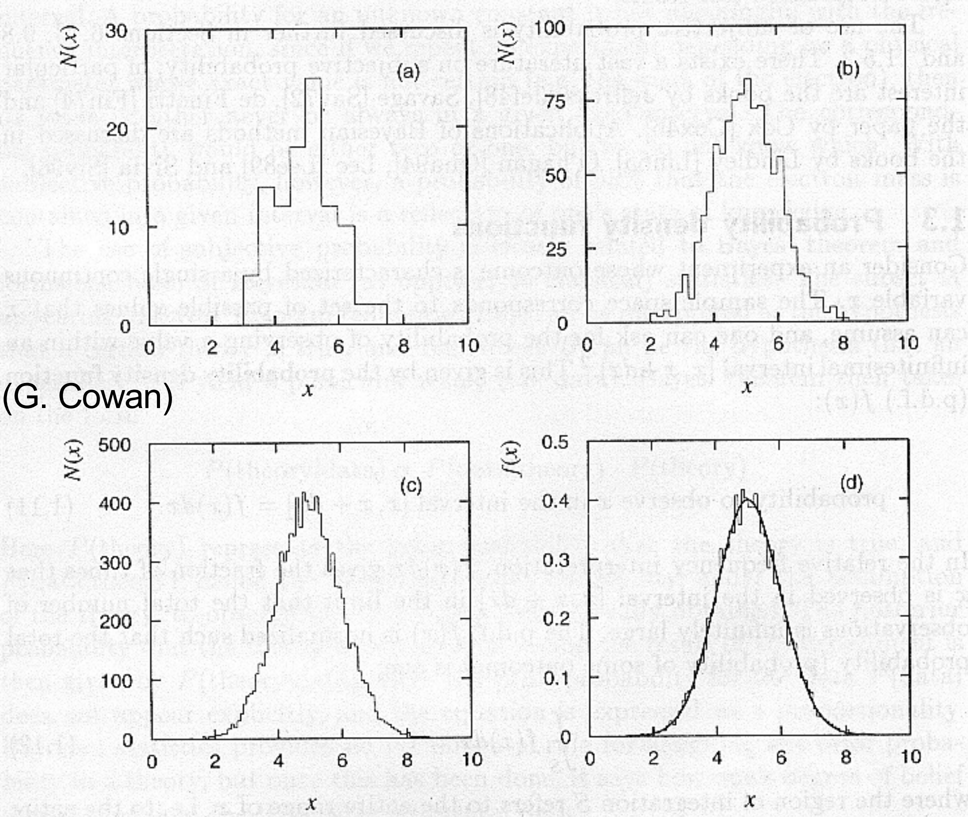
Probability density function

Let's examine meaning of measurement uncertainty σ. However let's first define the normalised distribution of repeated measurements of variable x, the probability density function (pdf) $f(x)$ (todennäköisyystiheysfunktio)

Sometimes only referred to as the "distribution" of variable x.

x can be either continuous or discrete & values of x confined to a finite (e.g. 0 – 1) or (semi-)infinite (e.g. 0 – ∞) range.

histogram: way of displaying number of entries within a sub-interval Δx (that is either constant or variable), $N(x)$, as a function of variable x. Δx optimised based on statistics to show distribution shape.



(G. Cowan)

Fig. 1.2 Histograms of various numbers of observations of a random variable x based on the same p.d.f. (a) $n = 100$ observations and a bin width of $\Delta x = 0.5$. (b) $n = 1000$ observations, $\Delta x = 0.2$. (c) $n = 10000$ observations, $\Delta x = 0.1$. (d) The same histogram as in (c), but normalized to unit area. Also shown as a smooth curve is the p.d.f. according to which the observations are distributed. For (a–c), the vertical axis $N(x)$ gives the number of entries in a bin containing x. For (d), the vertical axis is $f(x) = N(x)/(n\Delta x)$.

ideally pdf (for a continious x) = histogram with:

- infinite data sample
- zero bin width
- normalised to unity

$$f(x) = \frac{N(x)}{n\Delta x}$$

n = total # of entries

$$\int f(x) = 1$$

For a discrete variable, pdf of course discrete as well: probability to observe value $x_i = f_i$

$$\sum_i f_i = 1$$

Random variables:

Normalised distribution of repeated measurements of a continuous variable x (label for an element in sample space S):

$$P(x \text{ found in } [x, x+dx]) = f(x)dx$$

$$\Rightarrow f(x) = \textbf{probability density function} \text{ (pdf)}$$

(todennäköisyys-tiheysfunktio)

In frequentist interpretation, $f(x)dx$ gives fraction of times x observed in $[x, x+dx]$ for n observations when $n \to \infty$.

normalization: $\int_S f(x)dx = 1 \quad (\text{e.g. } S = [-\infty, \infty])$,

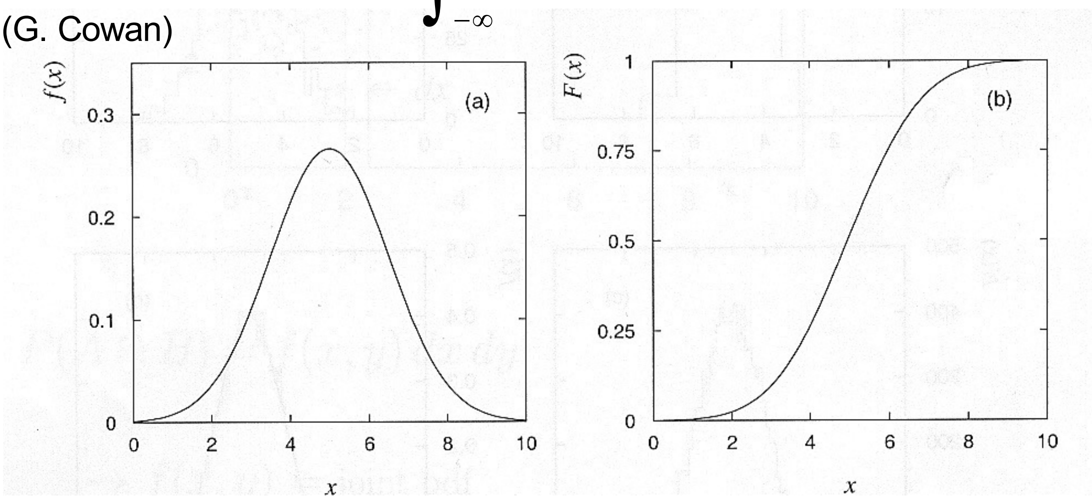
where S = entire x range, i.e. true value of x somewhere in S.

Define the **cumulative distribution function**:

$$F(x) = \int_{-\infty}^{x} f(x')dx'$$

(kertymäfunktio)

(G. Cowan)



A useful concept related to cumulative distribution are α**–points**:

$x_\alpha = F^{-1}(\alpha)$ most common α–point the **median** $x_{½}$ (mediaani)

For a discrete variable x_i ($P(x_i)$ = probability to observe value x_i):

$$f_i = P(x_i) \qquad \sum_i f_i = 1 \qquad F(x) = \sum_{x_i \le x} P(x_i)$$

Consider continuous random variable x with pdf $f(x)$
define the **expectation or mean value & variance** as:

$$E[x] = \int_{-\infty}^{\infty} x\, f(x)dx \equiv \mu \quad V[x] = E[(x - E[x])^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \equiv \sigma^2$$

Variance measures spread of x around mean μ.
The standard deviation (hajonta) σ given by $\sqrt{V[x]}$

For a function $a(x)$ with pdf $g(a)$:

$$E[a] = \int_{-\infty}^{\infty} a\, g(a)da = \int_{-\infty}^{\infty} a(x)\, f(x)dx \equiv \mu_a \quad \text{(equivalent but } \neq a(E[x])\text{)}$$

$$V[a] = E[(a - E[a])^2] = \int_{-\infty}^{\infty} (a-\mu_a)^2 g(a)da = \int_{-\infty}^{\infty} (a(x)-\mu_a)^2 f(x)dx \equiv \sigma_a^2$$

Equivalent to (arithmetic) mean & variance when $N \to \infty$:

$$\bar{x} = \frac{1}{N}\sum_i x_i \quad \text{or generally } \bar{a} = \frac{1}{N}\sum_i a(x_i)$$

$$s^2(x) = \frac{1}{N-1}\sum_i (x_i - \bar{x})^2 \text{ or generally } s^2(a) = \frac{1}{N-1}\sum_i (a(x_i) - \bar{a})^2$$

Covariance of two random variables x & y defined as:

$$V_{xy} = E[(x-\mu_x)(y-\mu_y)] = E[xy] - \mu_x\mu_y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)dxdy - \mu_x\mu_y$$

Covariance (or uncertainty) matrix V_{xy} often denoted by cov[x,y]

By construction the covariance matrix V_{ab} symmetric in a & b
and diagonal elements $V_{aa} = \sigma_a^2$ (i.e. the variances) positive.

Correlations

To express the correlation between two random variables, x & y, in a dimensionless way, **the correlation coefficient**, $\rho_{xy,}$ used

$$\rho_{xy} = \mathrm{cov}[x,y]/\sigma_x\sigma_y, \qquad -1 \le \rho_{xy} \le 1$$

If x,y independent, i.e. $f(x,y) = f_x(x)f_y(y)$, then

$$E[xy] = \int\int xy\ f(x,y)dxdy = \mu_x\mu_y \quad \Rightarrow \quad \mathrm{cov}[x,y] = 0$$

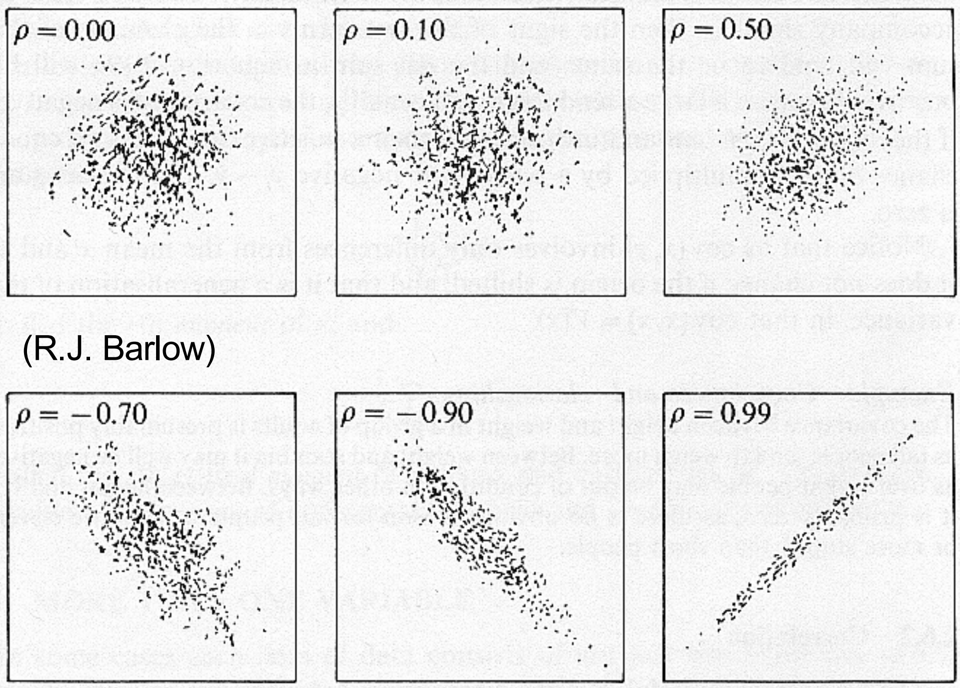
i.e. x & y 'uncorrelated' NB! not always true the other way.

Equivalent to well-known variables with similar names:

$$\mathrm{cov}(x,y) = \frac{1}{N}\sum_i (x_i - \bar{x})(y_i - \bar{y}) = \overline{xy} - \bar{x}\cdot\bar{y}$$

$$\rho_{xy} = \frac{\mathrm{cov}(x,y)}{\sigma_x\sigma_y} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x\sigma_y} \qquad \rho_{xy} \in [-1,1]$$

Examples of distributions of 2 variables with different correlation coefficients



(R.J. Barlow)

Fig. 2.4 Scatter plots showing examples of correlation. The scales and origin of the axes are irrelevant (see text) and are therefore not shown.

Two types of uncertainties related to any measurement:

• random/statistical uncertainties (satunnaiset/tilastolliset epävarmuudet) – inability of measurement system to give infinitely accurate answer (due to measurement precision | sampling statistics | phenomenon itself) – $\propto 1/\sqrt{\text{statistics}}$

• systematic uncertainties (systemaattiset epävarmuudet) – some feature/uncertainty of method, apparatus, calibration, environment etc ... influencing result ((un)knowingly to experimentalist) – (usually) not statistics dependent N.B. often knowledge & ability to determine systematic uncertainty increases with statistics ⇒ decreased systematic uncertainty.

(sometimes 3 given: statistical, systematic & gross (karkeita). Gross refers here to carelessness in measurement or malfunction of apparatus).

E.g. determine the decay constant λ of a radioactive source based on decay rate (-dN/dt) and sample weight. $-dN/dt = \lambda N$
Statistical uncertainty comes from the number of events in time interval. Possible systematic uncertainties are detector efficiency, background sources (& subtraction), sample purity, clock & balance calibration etc...

Systematic distinct w.r.t. statistical uncertainties – no decrease with statistics + measurement at different points usually affected in same direction (points ”non-independent”)

No simple rules/prescriptions for eliminating systematic uncertainties but common sense, careful thinking & experience usually helps to identify & estimate sources.

Ideally minimize effect from systematics sources during the design of the experiment (which is not always possible).

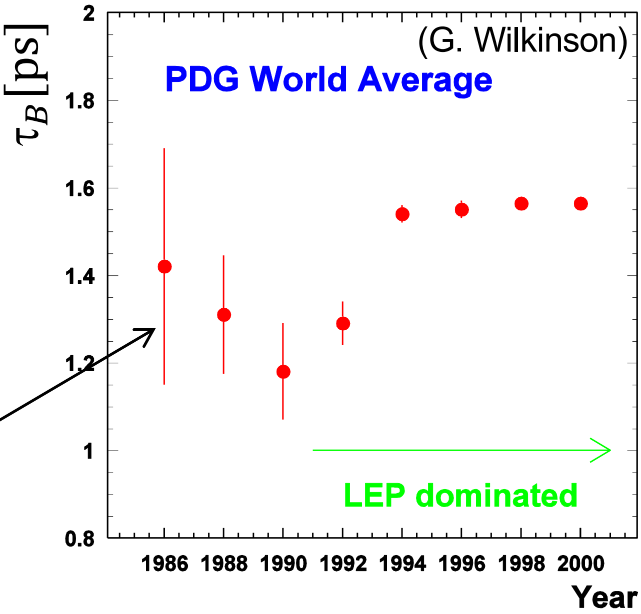
Systematic uncertainties

Generally accepted method for treating systematic uncertainties: <u>eliminate</u> those effects you can, <u>correct</u> (large) ”biases” introduced by effects you can estimate & <u>evaluate</u> remaining (+ uncertainties of the corrections of the ”biases”) in terms of statistical-like uncertainties.

$$m_p = (1.6727 \pm 0.0001\,(\mathrm{stat}) \pm 0.0002\,(\mathrm{syst})) \cdot 10^{-27}\ \mathrm{kg}$$

Some suggestions related to systematic uncertainties:

- Pay special attention to all numbers that enter result directly & crosscheck them (like efficiency, calibration etc..)!!
- Make analysis in subsamples as a function of variables that should have no influence on result (e.g. time, rate etc..)
- Use alternative analysis methods & possible constraints (e.g. energy & momentum conservation, coincidences etc...)
- Never underestimate your own ”subjectivity” (tendency to ”know” the answer) / (self)criticism useful & mostly constructive



Average B hadron (i.e. particle containing a b quark) lifetime vs year. Note jump 1992 → 1994.

A good rule of thumb:
a) estimate realistically all uncertainties
b) sum the systematic uncertainties in quadrature
c) try to reduce total systematic uncertainty ≤ statistical.

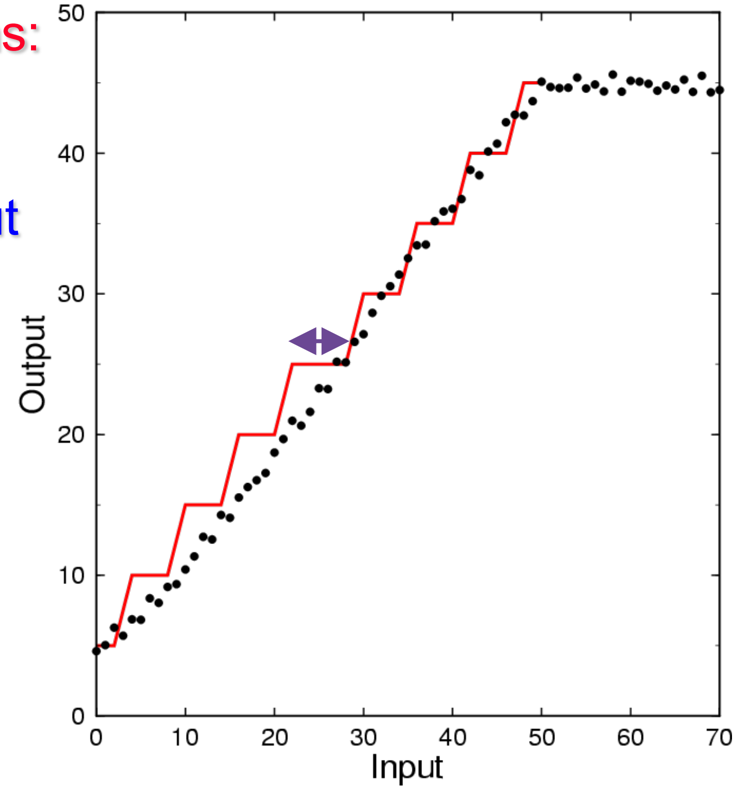
Resolution (erotuskyky)

Most common definitions:

largest change of input without change of output

• given as per cent of input range

• common example: digitalization resolution (*digitointiresoluutio*) (given as bits: 32 bits, 64 bits etc..)

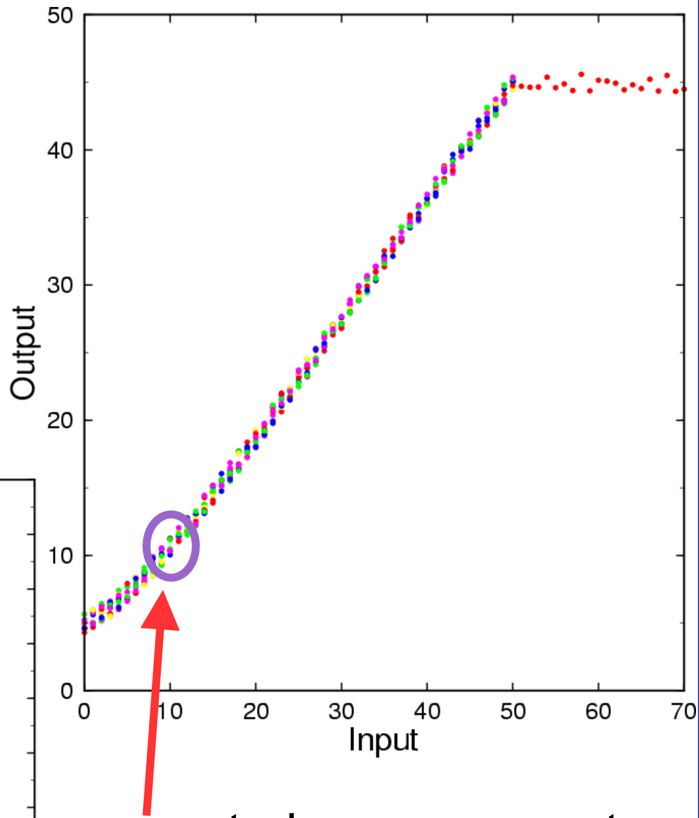


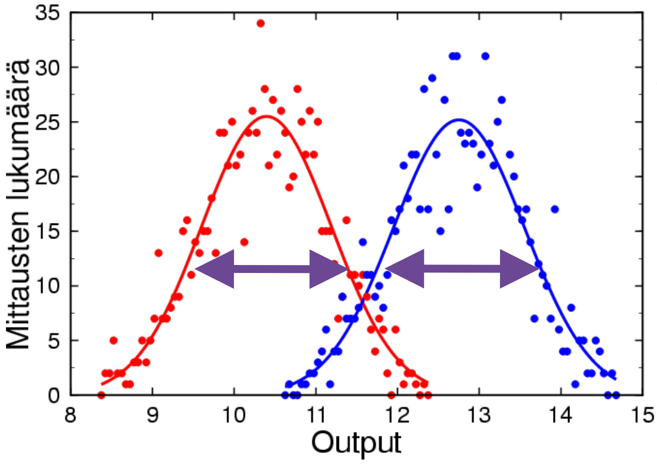
resolution of output

• full width at half maximum (**FWHM**, puoliarvoleveys) ≈ 2.35 × Gaussian σ

• related to statistical uncertainty (or precision)

• given in units of output





repeated measurements at input values 10 &12

Physics terminology convention:

- **Precision:** (sisäinen tarkkuus) – ”statistical” uncertainty
  how well repeated measurements agree
- **Accuracy**: (ulkoinen tarkkuus) – ”bias” / ”systematic”
  how well result reflects true value

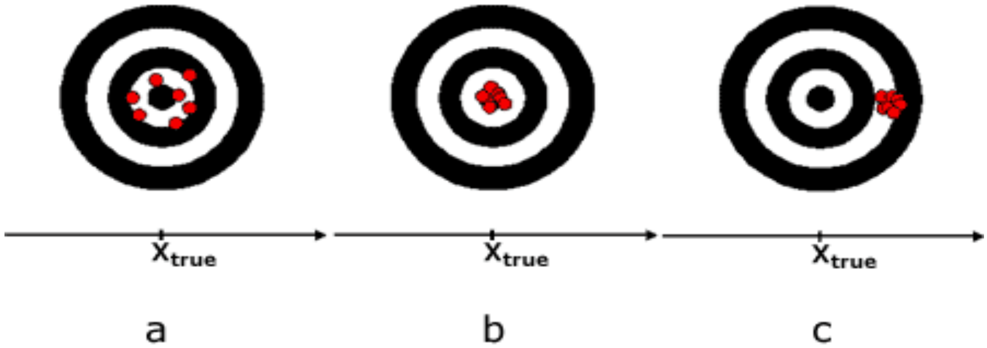
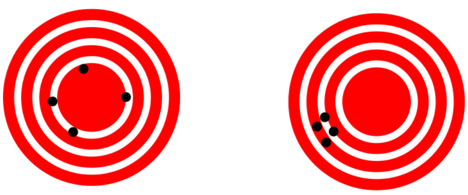


Figure a) Good accuracy (ulkoinen tarkkuus) but bad precision (sisäinen tarkkuus); b) both good accuracy and precision; c) good precision but bad accuracy.

International Organisation of Standarization (ISO) convention (used e.g. in metrology):

- **Precision:** repeatability/reproducibility
  closeness of agreement among a set of results
- **Trueness**: calibration
  closeness of mean of a set of measurement results to the actual (true) value
- **Accuracy:** (kokonaistarkkuus)
  overall closeness of a measurement to the true value

Low accuracy due to a low precision



Low accuracy due to a poor trueness

Uncertainty propagation

Consider f to be a function of a random variable x. For small deviations can do a Taylor expansion around x_0:

$$f(x) \approx f(x_0) + (x - x_0)(\partial f/\partial x)\big|_{x=x_0} \quad \Rightarrow$$

$$V(f) \approx (\partial f/\partial x)^2 V(x) \quad \& \quad \sigma_f \approx |(\partial f/\partial x)|\sigma_x$$

Approximation works well for small deviations as long as the 1st differential doesn't change significantly over a scale of few σ's.

$$f(x,y) \Rightarrow V(f) = \left(\frac{\partial f}{\partial x}\right)^2 V(x) + \left(\frac{\partial f}{\partial y}\right)^2 V(y) + 2\left(\frac{\partial f}{\partial x}\right)\left(\frac{\partial f}{\partial y}\right)\mathrm{cov}(x,y)$$

$$\text{where} \quad V(x) = \sigma_x^2 \quad V(y) = \sigma_y^2 \quad \mathrm{cov}(x,y) = \rho\sigma_x\sigma_y$$

The law of combination of uncertainties: consider n uncorrelated variables x_i (virheitten kasautumislaki)

$$f(x_1,\ldots,x_n) \quad \Rightarrow \quad \sigma_f \approx \sqrt{\sum_i (\partial f/\partial x_i)^2 \sigma_{x_i}^2}$$

valid only for uncorrelated variables i.e. $\mathrm{cov}(x_i,x_j) = 0$ for all $i \neq j$

If not all $\mathrm{cov}(x_i, x_j) = 0$, general formula have to be used:

$$\sigma_f \approx \sqrt{\sum_i \sum_j (\partial f/\partial x_i)(\partial f/\partial x_j)\mathrm{cov}(x_i,x_j)}$$

Can be generalized to arbitrary # of functions (m) that are functions of arbitrary # of variables (n). Uncertainty matrix:

$$\mathbf{U}_f = \mathbf{G}\,\mathbf{V}_x\,\mathbf{G}^T \text{ , } where\ G_{ki} = \left(\frac{\partial f_k}{\partial x_i}\right)$$

$\mathbf{U}_f$ ($\mathbf{V}_x$) symmetric & square matrices size $\mathrm{m} \times \mathrm{m}$ ($\mathrm{n} \times \mathrm{n}$).
$\mathbf{G}$ a rectangular $\mathrm{m} \times \mathrm{n}$ matrix. Next a simple example.

Typical radiation detectors that measure charged particles in a magnetic field perpendicular to the transverse plane have cylindrical shape & particle position measured in cylindrical coordinates (r, ϕ, z). r mostly known quite precisely, σ_r small. ϕ & z measured with uncorrelated uncertainties σ_ϕ & σ_z.

$$\text{Then } \mathbf{G} = \begin{pmatrix} \partial x/\partial r & \partial x/\partial\phi & \partial x/\partial z \\ \partial y/\partial r & \partial y/\partial\phi & \partial y/\partial z \\ \partial z/\partial r & \partial z/\partial\phi & \partial z/\partial z \end{pmatrix} = \begin{pmatrix} \cos\phi & -r\sin\phi & 0 \\ \sin\phi & r\cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\Rightarrow \mathbf{U} = \begin{pmatrix} \sigma_x^2 & U_{xy} & U_{xz} \\ U_{yx} & \sigma_y^2 & U_{yz} \\ U_{zx} & U_{zy} & \sigma_z^2 \end{pmatrix} = \mathbf{G}\begin{pmatrix} \sigma_r^2 & 0 & 0 \\ 0 & \sigma_\phi^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}\mathbf{G}^\mathbf{T}$$

$$= \begin{pmatrix} x^2\sigma_r^2/r^2 + y^2\sigma_\phi^2 & xy\sigma_r^2/r^2 - xy\sigma_\phi^2 & 0 \\ xy\sigma_r^2/r^2 - xy\sigma_\phi^2 & y^2\sigma_r^2/r^2 + x^2\sigma_\phi^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}$$

x uncertainty depend on the y value as well & vice-versa. Also xy covariances $\neq$ 0. z uncertainties & covariances not affected.

How to avoid correlation between variables?
transform to orthogonal variables $x' \Rightarrow \text{cov}[x_i', x_j'] = 0$ for all $i \neq j$ (see e.g. G. Cowen: Statistical data analysis 1.7)

Example: straight line fit $y = a'(x - \bar{x}) + b'$ & not $y = ax + b$
Coefficients a' & b' uncorrelated, not the case for a & b.

Nice in theory but often not so useful in reality since physical meaning of new variables often ambigious & they usually cannot be directly measured from experimental data anyway.

Correlations (derived using Copliot)

if x = r sin phi and y = r cos phi, how are the uncertainties of x and y depending on the uncertainty on phi and r? Derive also the dependence of the uncertainty of x and y on the exact x and y value.

Copilot

Great question! Let's go step by step to derive how the uncertainties in x and y depend on the uncertainties in r and ϕ, and also how they depend on the actual values of x and y.

1. Definitions

Given:

- $x = r \sin\phi$
- $y = r \cos\phi$

Let:

- σ_r = uncertainty in r
- σ_ϕ = uncertainty in ϕ
- σ_x, σ_y = uncertainties in x, y

Assume r and ϕ are **uncorrelated**, so covariance terms vanish.

2. Error Propagation

Using the general formula for error propagation:

For $x = r\sin\phi$:

$$\sigma_x^2 = \left(\frac{\partial x}{\partial r}\right)^2 \sigma_r^2 + \left(\frac{\partial x}{\partial \phi}\right)^2 \sigma_\phi^2 = (\sin\phi)^2\sigma_r^2 + (r\cos\phi)^2\sigma_\phi^2$$

For $y = r\cos\phi$:

$$\sigma_y^2 = \left(\frac{\partial y}{\partial r}\right)^2 \sigma_r^2 + \left(\frac{\partial y}{\partial \phi}\right)^2 \sigma_\phi^2 = (\cos\phi)^2\sigma_r^2 + (r\sin\phi)^2\sigma_\phi^2$$

We want to express the uncertainties σ_x and σ_y in terms of the actual values of x and y.

From the definitions:

- $x = r\sin\phi$
- $y = r\cos\phi$

So:

- $r = \sqrt{x^2+y^2}$
- $\sin\phi = \frac{x}{r}$
- $\cos\phi = \frac{y}{r}$

Substitute into the uncertainty formulas:

For σ_x^2:

$$\sigma_x^2 = \left(\frac{x}{r}\right)^2\sigma_r^2 + \left(r\cdot\frac{y}{r}\right)^2\sigma_\phi^2 = \frac{x^2}{r^2}\sigma_r^2 + y^2\sigma_\phi^2$$

For σ_y^2:

$$\sigma_y^2 = \left(\frac{y}{r}\right)^2\sigma_r^2 + \left(r\cdot\frac{x}{r}\right)^2\sigma_\phi^2 = \frac{y^2}{r^2}\sigma_r^2 + x^2\sigma_\phi^2$$

4. Final Expressions

Let $r = \sqrt{x^2+y^2}$, then:

$$\sigma_x^2 = \frac{x^2}{x^2+y^2}\sigma_r^2 + y^2\sigma_\phi^2$$

$$\sigma_y^2 = \frac{y^2}{x^2+y^2}\sigma_r^2 + x^2\sigma_\phi^2$$

Interpretation

- The uncertainty in x depends more on σ_r when x is large compared to y, and more on σ_ϕ when y is large.
- Similarly, the uncertainty in y depends more on σ_r when y is large, and more on σ_ϕ when x is large.
- This shows a **cross-dependence**: uncertainty in angle affects the orthogonal coordinate more strongly.

To compute the **covariance between the uncertainties in x and y**, we need to apply the **error pro formula for covariance** between two functions of correlated variables.

Given:

- $x = r\sin\phi$
- $y = r\cos\phi$

Let:

- σ_r, σ_ϕ be the uncertainties in r and ϕ
- $\mathrm{Cov}(x,y)$ be the covariance between x and y

Assuming r and ϕ are **uncorrelated**, i.e., $\mathrm{Cov}(r,\phi) = 0$, the covariance between x and y is:

Covariance Formula

$$\mathrm{Cov}(x,y) = \frac{\partial x}{\partial r}\frac{\partial y}{\partial r}\sigma_r^2 + \frac{\partial x}{\partial \phi}\frac{\partial y}{\partial \phi}\sigma_\phi^2$$

Compute the partial derivatives:

- $\frac{\partial x}{\partial r} = \sin\phi$
- $\frac{\partial y}{\partial r} = \cos\phi$
- $\frac{\partial x}{\partial \phi} = r\cos\phi$
- $\frac{\partial y}{\partial \phi} = -r\sin\phi$

Substitute:

$$\mathrm{Cov}(x,y) = \sin\phi\cos\phi\,\sigma_r^2 + (r\cos\phi)(-r\sin\phi)\,\sigma_\phi^2$$

$$\mathrm{Cov}(x,y) = \sin\phi\cos\phi\,\sigma_r^2 - r^2\sin\phi\cos\phi\,\sigma_\phi^2$$

Factor out $\sin\phi\cos\phi$:

$$\mathrm{Cov}(x,y) = \sin\phi\cos\phi\left(\sigma_r^2 - r^2\sigma_\phi^2\right)$$

covariance expressed using x and y values

Copilot

To express the **covariance between** $x = r\sin\phi$ and $y = r\cos\phi$ **in terms of the actual values of x and y**, we start from the derived expression:

$$\mathrm{Cov}(x,y) = \frac{xy}{r^2}\sigma_r^2 - xy\sigma_\phi^2$$

Now, since $r^2 = x^2 + y^2$, we substitute:

$$\boxed{\mathrm{Cov}(x,y) = \frac{xy}{x^2+y^2}\sigma_r^2 - xy\sigma_\phi^2}$$

✅ **Interpretation:**

- The **first term** $\frac{xy}{x^2+y^2}\sigma_r^2$ reflects how the uncertainty in the radial distance r contributes to the covariance.
- The **second term** $-xy\sigma_\phi^2$ reflects how the angular uncertainty ϕ contributes.
- If either $x = 0$ or $y = 0$, the covariance is zero.
- If $\sigma_r^2 = r^2\sigma_\phi^2$, the two terms cancel out and the covariance is zero.

Would you like to visualize how this covariance behaves for different values of x, y, σ_r, and σ_ϕ?

Binomial distribution

N independent trials (or experiments):
outcome of each trial either "success" or "failure",
the probability of success on any given trial p

Define discrete random variable n = # of successes ($0 \le n \le N$).
Probability of a specific outcome (in a specific order) e.g. *ssfsf*

$$pp(1-p)p(1-p) \Rightarrow n \text{ success, } N-n \text{ failures } = p^n(1-p)^{N-n}$$

but order irrelevant; $\dfrac{N!}{n!(N-n)!}$ ways to get n successes in N trials

So the total probability to have n successes in N events

$$f(n;N,p) = \frac{N!}{n!(N-n)!}p^n(1-p)^{N-n}$$

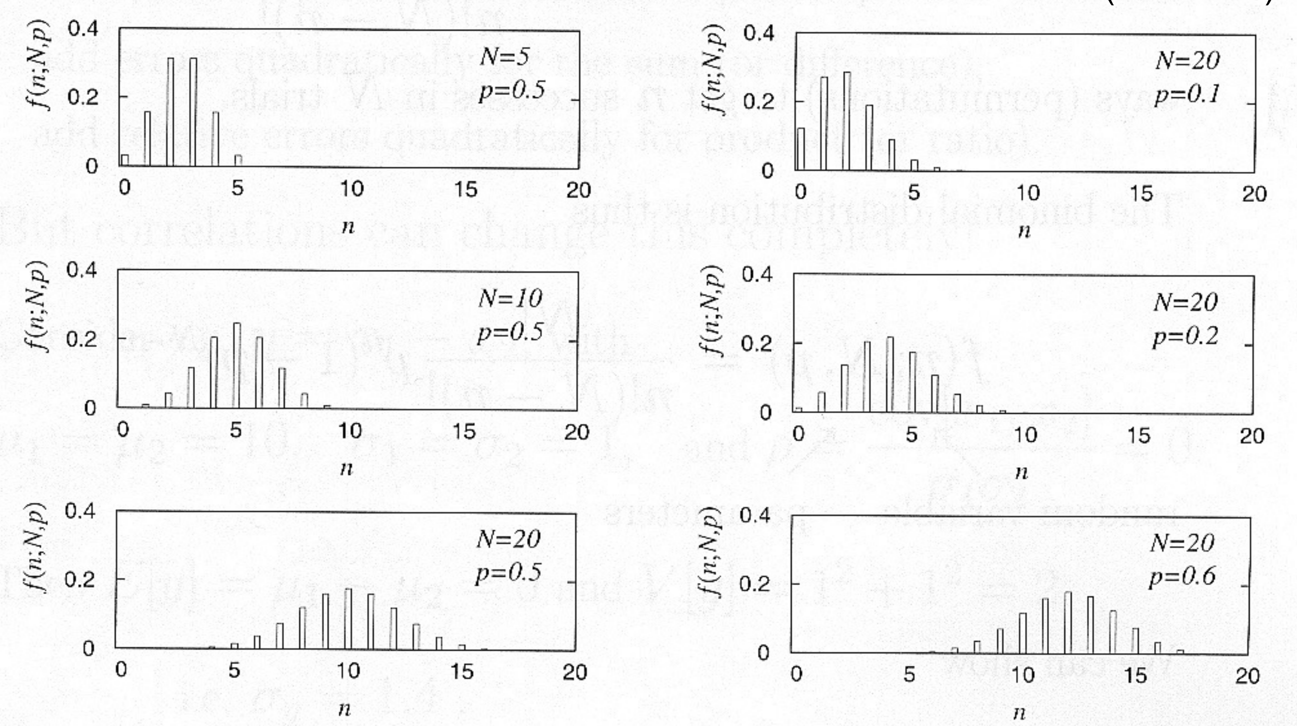
n – random variable
N, p – parameters

such a distribution **binomial.** The normalisation

$$\sum_{n=0}^{N}\frac{N!}{n!(N-n)!}p^n(1-p)^{N-n} = 1$$

OK by default.

(G. Cowan)



Binomial & multinomial distribution

Expectation value & variance:

$$E[n] = \sum_{n=0}^{N} n\, f(n; N, p) = Np$$

$E[n]$ & $V[n]$ not functions of random number n, but constants of parameters N & p

$$V[n] = E[n^2] - \left(E[n]\right)^2 = Np(1-p)$$

$$\sqrt{V[n]}/E[n] = \sqrt{(1-p)/p}/\sqrt{N}$$

NB! $\boxed{\sigma_p = \sqrt{\mathrm{V}[p]} = \sqrt{V[n]}/N = \sqrt{(1-p)p}/\sqrt{N}}$

NB! When p ≈ 0/1, σ_p is not accurate (too small!), recommend to use exact estimate (Clopper-Pearson).

Examples: Monte Carlo based efficiencies; decay fractions; histograms; # of decays $X \to YZ$, n, a binomial random variable, p = $X \to YZ$ branching fraction, N = # of X decays.

Multinomial distribution:

like binomial but now m possible outcomes instead of two, the probabilities now $\bar{p} = (p_1, \ldots, p_m)$ with $\sum_{i=1}^{m} p_i = 1$

For N trials, want to know probability to obtain: n_1 times of outcome 1, n_2 times of outcome 2, ..., n_m times of outcome m ⇒ the **multinomial distribution** for $\bar{n} = (n_1, \ldots, n_m)$:

$$f(\bar{n}; N, \bar{p}) = \frac{N!}{n_1! n_2! \ldots n_m!} p_1^{n_1} p_2^{n_2} \ldots p_m^{n_m}$$

Consider outcome i as "success", other outcomes "failure" ⇒ all individual n_i binomially distributed with parameters N, p_i.

$E[n_i] = Np_i$, $V[n_i] = Np_i(1-p_i)$ for all i. One can also define covariance $V_{ij} = E[(n_i - E[n_i])(n_j - E[n_j])] = -Np_i p_j$ for $i \neq j$.

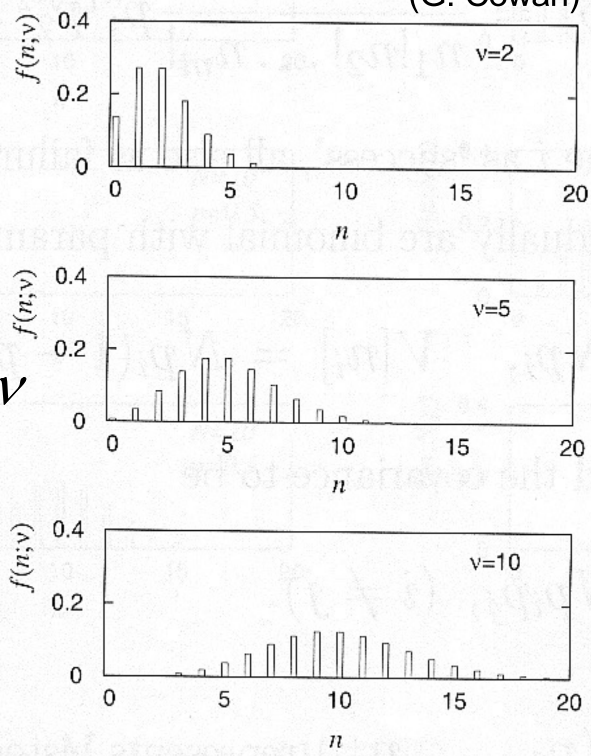
Example: probability to obtain a particular histogram from N observations (= entries). $\bar{n} = (n_1, \ldots, n_n)$ equals bin content & $\bar{p} = (p_1, \ldots, p_n)$ bin probability. Negative V_{ij} ⇒ # of entries in any 2 bins correlated. If larger # entries than expected in bin i ($n_i > Np_i$), probability increased for $n_j < Np_j$ & vice–versa.

Poisson distribution

'Events in a continuum' e.g. # of lightning during a thunder storm. Binomial random variable n in the limit: $N \to \infty$, $p \to 0$, $E[n] = Np \to \nu$ (i.e. a finite value), n follows a **Poisson distribution.**

$$f(n;\nu) = \frac{\nu^n}{n!} e^{-\nu} \quad (0 \leq n \leq \infty)$$ pdf has only 1 parameter ν

(G. Cowan)



Expectation value & variance of Poisson random variable n

$$E[n] = \sum_{n=0}^{\infty} n \frac{\nu^n}{n!} e^{-\nu} = \nu$$

$$V[n] = \sum_{n=0}^{\infty} (n-\nu)^2 \frac{\nu^n}{n!} e^{-\nu} = \nu$$

$$\sqrt{V[n]} / E[n] = 1/\sqrt{\nu}$$

Examples: # of decays in a fixed time interval from a radioactive source, # of scattering events n occuring for a fixed integrated luminosity $\mathcal{L}$ ($\approx$ particle flux) due to some process with cross section σ ($\nu = \sigma \int \mathcal{L}\, dt$).

Poisson vs binomial: hitchhiking; cars passing by according to a Poisson distribution with mean frequency 1 / minute. Probability of individual car giving lift is 1 %. Calculate probability that hitch–hiker still waits for a lift (i) after 60 cars passed (ii) after 1 hour.

(i) $(1-p)^N = 0.99^{60} = 0.5472$ (# of trials, binomial)

(ii) $e^{-Np} = 0.5488$ (time span ≈ random, Poisson)

avoids many worries

Assume 2 Poisson with means ν_1 & ν_2. The combined sample e.g. two types of decays of a particle ⇒ get a convolution: $P(n) = \sum f(n',\nu_1) f(n-n',\nu_2) = f(n, \nu_1+\nu_2)$, a new Poissonian with $\nu = \nu_1+\nu_2$

Uniform distribution, simplest continuous distribution. Have continuous random variable x. Uniform pdf definition

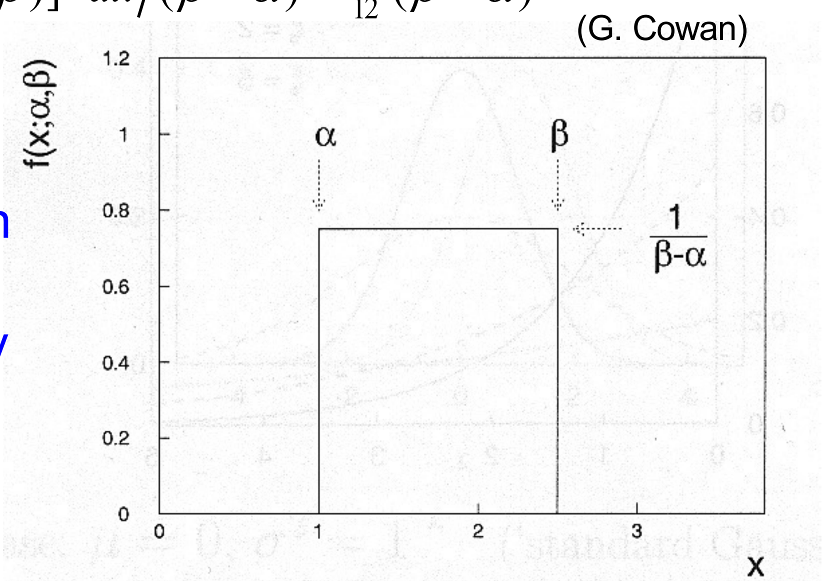
$$f(x;\alpha,\beta)=\begin{cases}\dfrac{1}{(\beta-\alpha)} & \alpha\le x\le\beta\\ 0 & \text{otherwise}\end{cases}$$

Expectation value & the variance of a uniform pdf:

$$E[x]=\int_\alpha^\beta x\,dx\big/(\beta-\alpha)=\tfrac{1}{2}(\alpha+\beta)$$

$$V[x]=\int_\alpha^\beta [x-\tfrac{1}{2}(\alpha+\beta)]^2\,dx\big/(\beta-\alpha)=\tfrac{1}{12}(\beta-\alpha)^2$$

Uniform distribution often our ”naive” guess but seldomly correct.



Example: relativistic kinematics: decay angles of a 2-body particle decay (spin = 0) in its rest frame, e.g. azimuthal angle φ uniform in $[0, 2\pi]$ & polar angle θ uniform in $[-\pi, \pi]$.

Monte Carlo methods start from an uniform distribution. N.B. for random variable x with cumulative distribution, $F(x)$, always uniformly distributed in [0,1] $\Rightarrow$ used for MC.

Exponential distribution:

Exponential pdf for a continuous random variable x

$$f(x;\xi) = \frac{1}{\xi} e^{-x/\xi} \quad (x \geq 0)$$

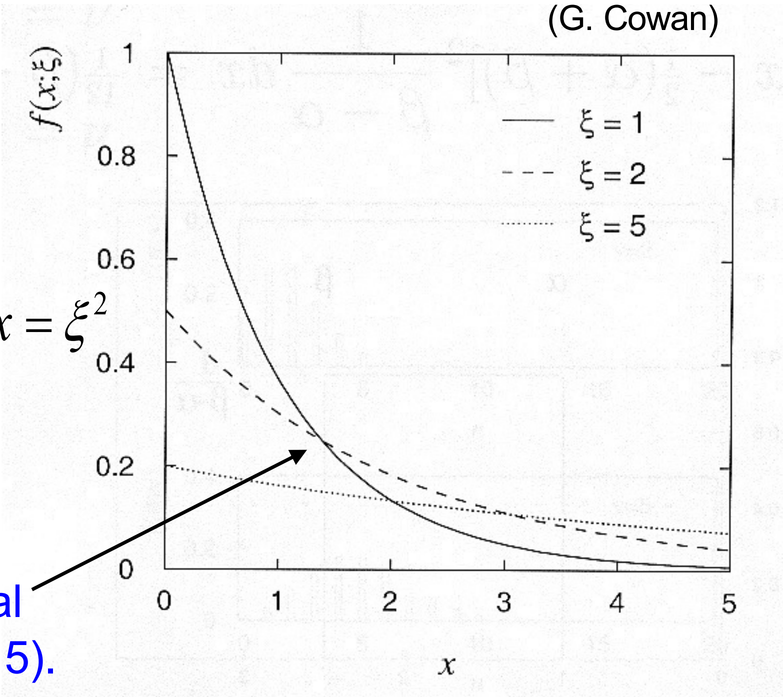
Exponential pdf characterized by only one parameter ξ.
Expectation value & variance of exponential distribution

$$E[x] = \int_0^\infty x \frac{1}{\xi} e^{-x/\xi} dx = \xi$$

$$V[x] = \int_0^\infty (x-\xi)^2 \frac{1}{\xi} e^{-x/\xi} dx = \xi^2$$

$$\sqrt{V[x]} / E[x] = 1$$

(G. Cowan)



Examples of exponential distributions (ξ = 1, 2 & 5).

Example: proper decay time of an unstable particle/state

$$f(t;\tau) = e^{-t/\tau} / \tau \quad (t \geq 0)$$

τ = mean life time

NB! unique feature of exponential pdf – ”lack of memory”

$$f(t - t_0 \,|\, t \geq t_0) = f(t)$$

absolute starting (& end) point (”zero”) irrelevant

Very convenient for any lifetime measurement. Also: attenuation of light/particles in medium, radioactive decay.

Gaussian (or normal) distribution:

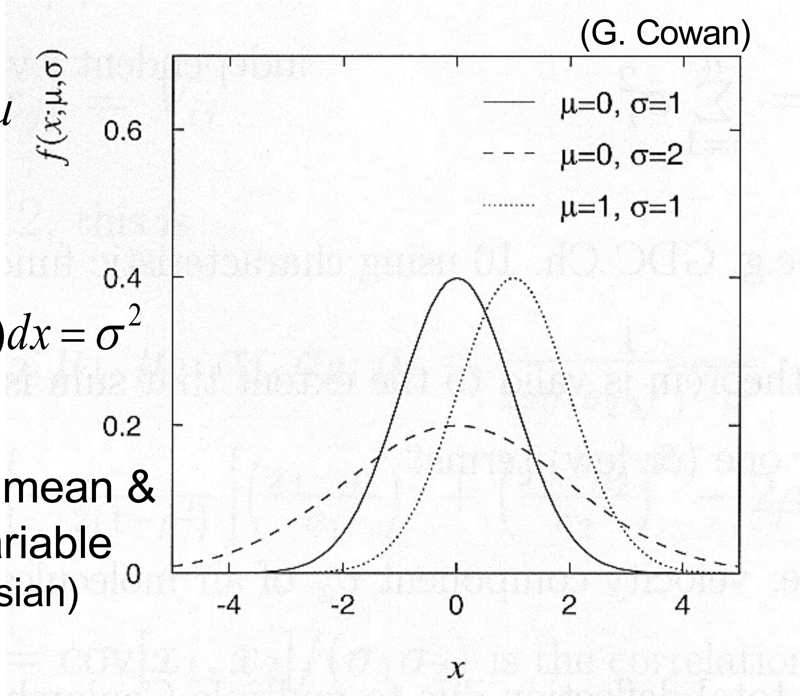
Gaussian pdf for a continuous random variable x

$$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \quad (-\infty < x < \infty)$$

Gaussian pdf characterized by two parameters: μ & σ.
Expectation value & variance of gaussian distribution:

$$E[x]=\int_{-\infty}^{\infty} x\, f(x;\mu,\sigma)dx=\mu$$

$$V[x]=\int_{-\infty}^{\infty}(x-\mu)^2 f(x;\mu,\sigma)dx=\sigma^2$$



NB! μ & σ often used for mean & spread of any random variable (that not necessarily a Gaussian)

Special case: μ = 0, σ = 1 ("standard Gaussian")

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-x^2/2\right), \qquad \Phi(x)=\int_{-\infty}^{x}\varphi(x')dx'$$

if y Gaussian with μ & σ, then $x = (y-\mu)/\sigma$ follows $\varphi(x)$ & $F(y) \leftrightarrow \Phi(x)$. No analytic expression of cumulative distribution $\Phi(x)$. Numerical evaluations of $\Phi(x)$ as well as α–points $x_\alpha = \Phi^{-1}(\alpha)$ tabulated and/or available in program libraries. E.g. 68.3 % within 1σ, 90 % within 1.645σ, 95% within 1.960σ, 99.7 % within 3σ… (2-tailed Gaussians).

Why uncertainties often are Gaussian?

A consequence **of Central Limit Theorem (CLT).** Look at behaviour of variable equal to sum of several others. Irrespective of distribution of orginal variables, if one takes sum X of n independent variables x_i, i = 1,...,n, each taken from a distribution with mean μ_i & variance V_i, distribution for X has expectation value & variance & becomes Gaussian $n \to \infty$.

$$\langle X \rangle = \sum_i \mu_i \quad V(X) = \sum_i V_i$$

note $V(X)$ equation above holds only for independent variables, formal proof of CLT tedious so we'll give a MC ”proof” instead:

1 random number $\langle X \rangle$ = 0.5 & $V(X)$ = 1/12

2 random numbers $\langle X \rangle$ = 1.0

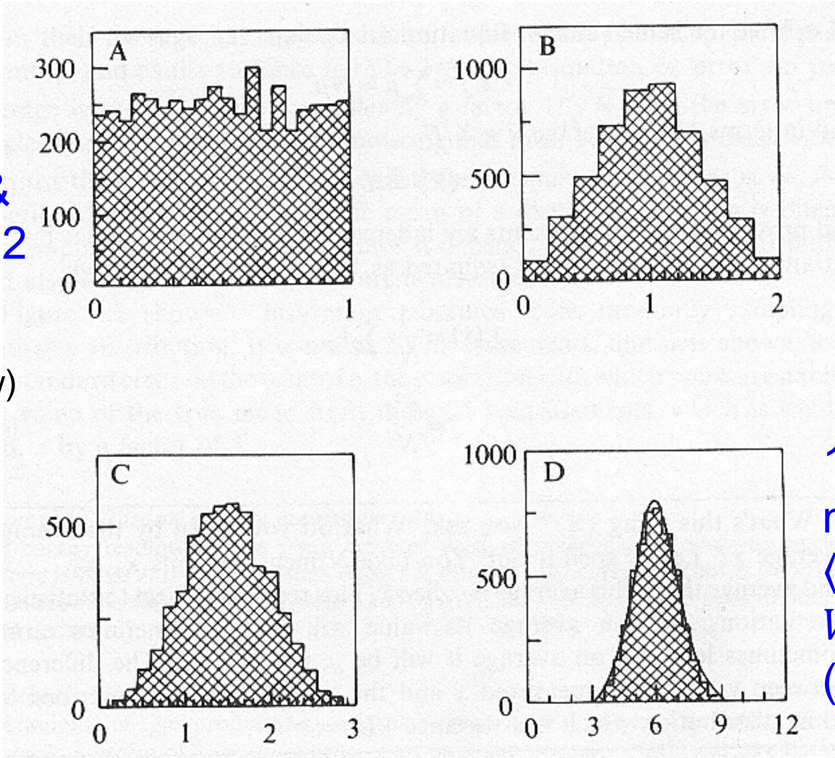


Fig. 4.1. The CLT at work.

(R.J. Barlow)

3 random numbers $\langle X \rangle$ = 1.5

12 random numbers $\langle X \rangle$ = 6.0 & $V(X)$ = 1.0 (~Gaussian)

Already after summing ~12 uniformly distributed random numbers in [0,1], one obtains a Gaussian like distribution.

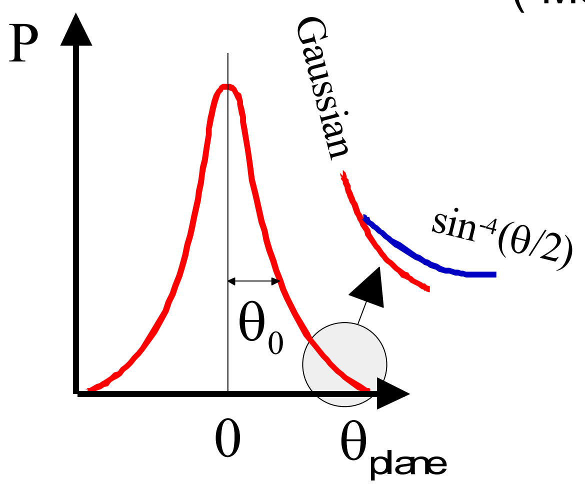
Central limit theorem

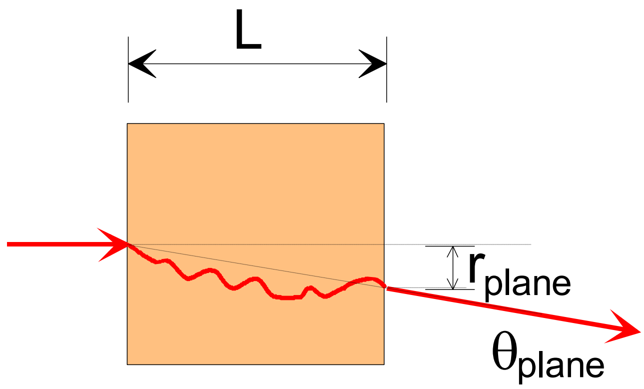
In practice, important to know how CLT works for n uncertainty contributions. Approximately true if large # of small contributions. Discrepancies arise when individual terms have long "tails" so that occasional large value will give large contribution to the sum

$\Rightarrow$ Gaussian approximation works usually well for central part of distribution, but "tails" (or wings) might be very non-Gaussian.

Below two examples where CLT (at least partially) breaks down:

• charged particle scattering in material. Average scattering angle $\langle\theta\rangle = 0$. If sufficiently thick material, particle scatter many times, **multiple scattering.** Sometimes θ large $\Rightarrow$ non-gaussian tails ("Moliere tails" or "Moliere scattering")





• total # of electron-hole (electron-ion) pairs created by a charged particle traversing a (thin) solid state (gaseous) detector. Described by the **Landau distribution** having a long tail extending to large values. High ionization events ("delta electrons") make up a significant fraction. A Gaussian approximation generally not valid in real systems.

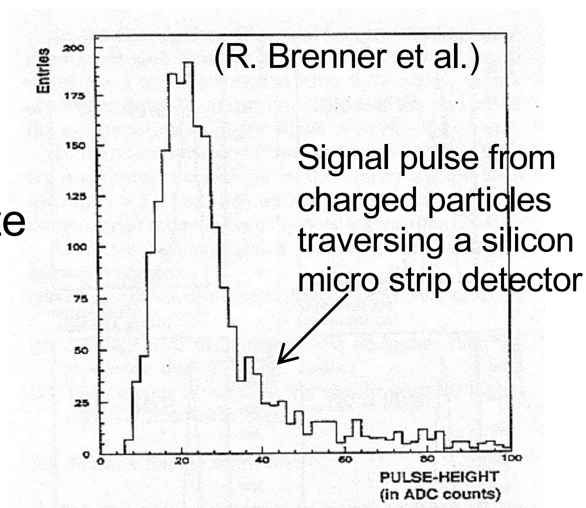


Fig. 10. VTT detector n-side. Data taken at 50°. Pulse-height distributions of all the strips in the cluster except the head and the tail. The shape of the distribution is determined by fluctuations in energy loss.

A binomial or a Poisson distribution can be satisfactorily approximated with a Gaussian when Np or ν large.

A binomial can be approximated by a Gaussian with $\mu = Np$ & $\sigma = [Np(1-p)]^{1/2}$ when Np large (> ~10). Approximation better for $p \approx 0.5$ than for small or large p–values (that requires $Np \gg 10$).

A Poission distribution can be approximated by a Gaussian when ν large (> ~10). Then $\mu = \nu$ & $\sigma = \sqrt{\nu}$

Multivariate Gaussian distribution:

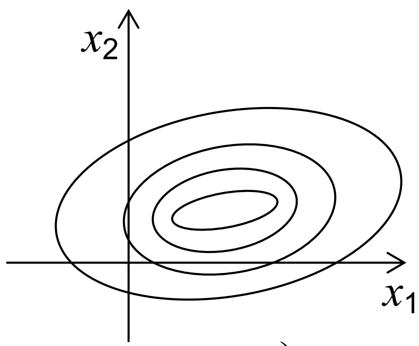
consider a vector $\bar{x} = (x_1,\ldots,x_m)$ all Gaussian distributed variables

$$f(\bar{x};\bar{\mu},\mathbf{V}) = \frac{1}{(2\pi)^{m/2}|\mathbf{V}|^{1/2}} \exp\left(-\frac{1}{2}(\bar{x}-\bar{\mu})^T \mathbf{V}^{-1}(\bar{x}-\bar{\mu})\right)$$

where $\bar{\mu} = (\mu_1,\ldots,\mu_m)$ the vector containing the means & $\mathbf{V}$ a symmetric $m \times m$ matrix containing $m(m+1)/2$ free parameters. The expectation values & (co)variances can be computed to be

$$E[x_i] = \mu_i \qquad V[x_i] = V_{ii} \qquad \mathrm{cov}[x_i,x_j] = V_{ij}$$

For example in two dimensions the pdf is



$$f(x_1,x_2;\mu_1,\mu_2,\sigma_1,\sigma_2,\rho_{12}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho_{12}^2)}} \times$$

$$\exp\left(-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right),$$

where $\rho_{12} = \mathrm{cov}[x_1,x_2]/(\sigma_1\sigma_2)$ the correlation coefficient. The pdf can be expressed by contour lines of equal probability that are ellipses in the x–y plane centered at ($x_1 = \mu_1$, $x_2 = \mu_2$).

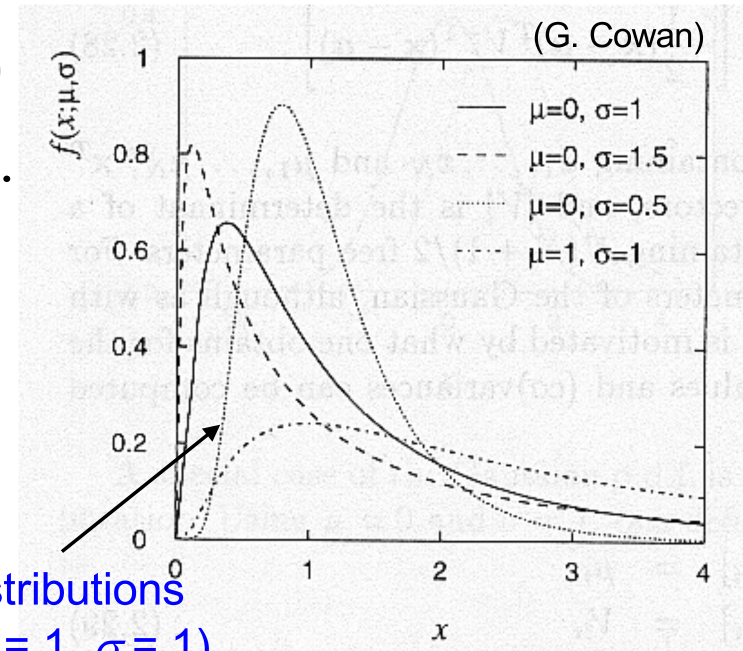
Log-normal distribution:

Assume continuous variable y to be gaussian with mean μ & variance σ^2, then $x = e^y$ follows **the log-normal distribution**:

$$f(x;\mu,\sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\ln x - \mu)}{2\sigma^2}\right) \qquad (x \geq 0)$$

The log-normal pdf characterized by same parameters as corresponding Gaussian of y, i.e. μ & σ. N.B. now μ & σ has nothing to do with the mean & the standard deviation of x.
Expectation value & variance of log-normal distribution:

$$E[x] = \exp\left(\mu + \tfrac{1}{2}\sigma^2\right)$$
$$V[x] = \exp\left(2\mu + \sigma^2\right) \cdot [\exp(\sigma^2) - 1]$$
$$\sqrt{V[x]} / E[x] = \sqrt{[\exp(\sigma^2) - 1]}$$



Examples of log-normal distributions ($\mu = 0$, $\sigma = 0.5$, 1 & 1.5 + $\mu = 1$, $\sigma = 1$).

Recall from CLT: if random variable y, a sum of a large number of small contributions, will be distributed according to a Gaussian $\Rightarrow$ if random variable x, a product of many small factors, will be distributed according to log-normal distribution. Used in modeling random uncertainties, that change result by a multiplicative factor.

Examples: describes the size of the clusters of silver atoms in photographic emulsion and weight & blood pressure of humans.

Chi-square (χ^2) distribution:

Chi-square pdf for a continuous random variable z

$$f(z;n)=\frac{1}{2^{n/2}\Gamma(n/2)}z^{(n/2)-1}e^{-z/2} \quad (z\geq 0, n=1,2,...)$$

where parameter n = "**number degrees of freedom**" (ndof)

& gamma function $\Gamma(x)$ defined as $\Gamma(x)=\int_0^\infty e^{-t}t^{x-1}dt$

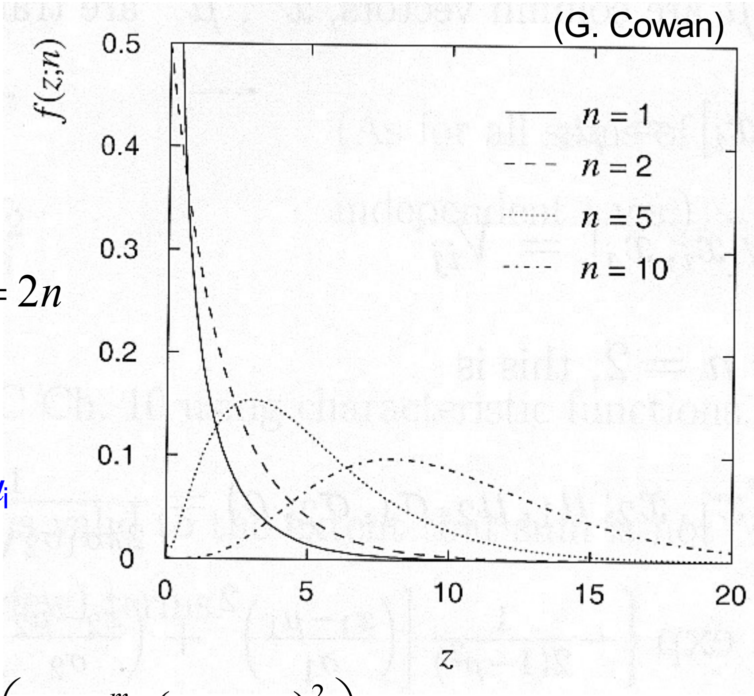
For our purposes only need to know following features of $\Gamma(x)$

$$\Gamma(n)=(n-1)! \text{ for integer } n \qquad \Gamma(x+1)=x\Gamma(x) \qquad \Gamma(1/2)=\sqrt{\pi}$$

Expectation value & variance of chi-square distribution

$$E[z]=\int_0^\infty z\, f(z;n)dz=n$$

$$V[z]=\int_0^\infty (z-n)^2 f(z;n)dz=2n$$



For m Gaussian random variables x_i with means μ_i & covariance matrix V_{ij} (for uncorrelated variables can use variances σ_i^2):

$$z=(\bar{x}-\bar{\mu})^T\mathbf{V}^{-1}(\bar{x}-\bar{\mu}) \qquad \left(z=\sum_{i=1}^{m}\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right)$$

follows a χ^2 distribution with m ndof

Example: variable for testing goodness-of-fit, especially with the "method of least squares" (pienemmän neliösumma menetelmä).
NB! For large m distribution for the value of $\sqrt{2\chi^2}$ can be approximated by a Gaussian distribution with $\mu=\sqrt{2m-1}$ & $\sigma=1$

Cauchy (also called Lorentzian or Breit-Wigner) **distribution:**

Cauchy (Lorentzian) pdf for a continuous random variable x

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2} \quad \left(f(x;t,s) = \frac{1}{\pi}\frac{s}{s^2+(x-t)^2} \right)$$

t & s are location & scale parameter

special case: Breit-Wigner (common in quantum mechanics)

$$f(x;\Gamma,x_0) = \frac{1}{\pi}\frac{\Gamma/2}{(\Gamma/2)^2+(x-x_0)^2}$$

where parameters x_0 & Γ mass & width of a resonant state

NB!. $\Gamma \propto 1/\tau$, where τ = mean lifetime of resonant state

The Cauchy distribution has a peculiar mathematical behaviour

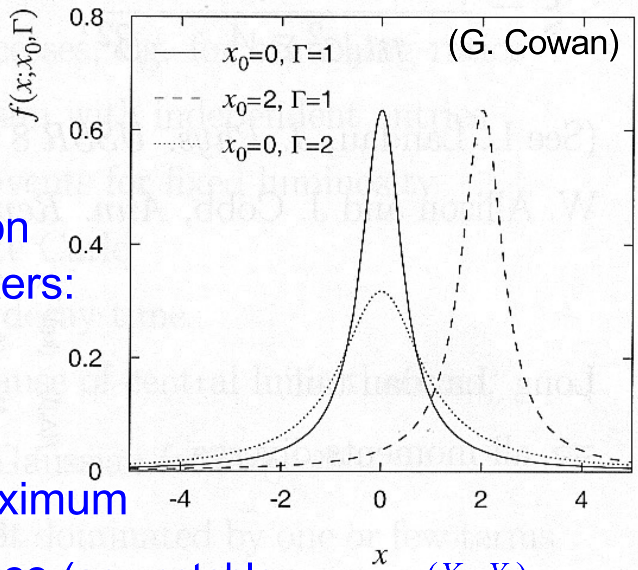
$E[x]$ = not well defined

$V[x] = \infty$

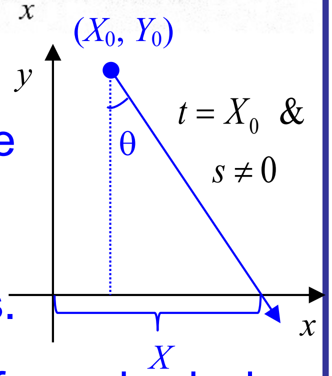
However the Cauchy distribution can be described by 2 parameters:

x_0 (or t) = peak position (i.e. mode or most probable value)

Γ (or $2 \cdot s$) = full width at half maximum



Examples: describes a resonance (an unstable particle or state) e.g. the W boson responsible for radioactive decays. Γ = decay width ($\propto$ inverse of the mean life time). Describes the distribution of horizontal distances X at which a line segment tilted at a **random** angle θ cuts the horizontal axis.



NB! in reality mean & variance mostly calculable for a physical phenomena described by a Breit-Wigner distribution since description only approximative (i.e. the resonance mass restricted by energy conservation, cannot be smaller than 0 or larger than total energy).

Landau distribution

In nuclear & particle physics one often encounters the **Landau distribution** $f(\Delta,\beta)$ for energy loss Δ of a charged particle with β = v/c traversing a layer of matter of thickness d

$$f(\Delta;\beta) = \frac{1}{\xi}\phi(\lambda), \qquad 0 \le \Delta \le \infty$$

$$\lambda = \frac{1}{\xi}\left[\Delta - \xi\left(\log\frac{\xi}{\varepsilon'} + 1 - \gamma_E\right)\right],$$

parameter ξ related to properties of material & velocity of particle whereas λ = dimensionless random variables related to both ξ & Δ

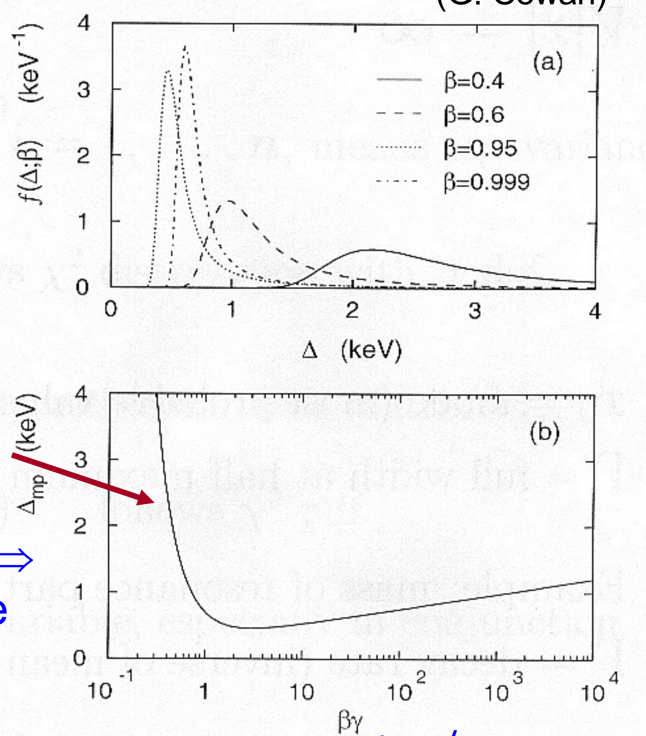
$$\xi = \frac{2\pi\, N_A e^4 z^2 \rho \sum Z}{m_e c^2 \sum A}\,\frac{d}{\beta^2}, \qquad \varepsilon' = \frac{I^2 \exp(\beta^2)}{2 m_e c^2 \beta^2 \gamma^2},$$

first written down by L. Landau in 1944

where z charge of incident particle in units of electron charge, ΣZ & ΣA sums of atomic numbers & weights of the material, $I \approx I_0 Z$ ionization energy characteristic of the material ($I_0 \approx$ 13.5 eV), $\gamma = (1-\beta^2)^{-½}$ & γ_E Eulers constant (= 0.5722...).

$$\phi(\lambda) = \frac{1}{\pi}\int_0^\infty \exp(-u\log u - \lambda u)\sin \pi u \, du,$$

(G. Cowan)



= momentum/mass

The integral above must be evaluated numerically & can be found in program libraries. Mean & variance of Landau distribution diverge due to tails. Described by 1 parameter:

Δ_{mp} = most probable value of Δ, sensitive to particle velocity β (follows "Bethe–Bloch formula") $\Rightarrow$ used for identifying particle type

for review see e.g. W. Allison & J. Cobb, *Ann. Rev. Nucl. Part. Sci.* 30 (1980) 253.