Statistical Methods

- Tilastolliset menetelmät

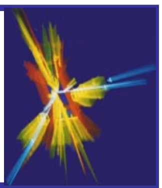
A 5 ECTS credit course autumn 2026

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 Exactum CK111
exercise sessions Fri 14-15 in Physicum D112
lectures: weeks 37-39, 41-42, 44-49
exercise sessions: starting week 39



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

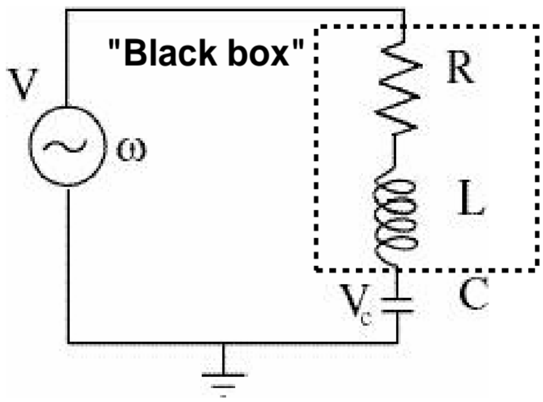
e.g. a simple electric circuit. Assume you built the circuit below ("Black box") & need to find out its resistance R & inductance L. Have to study response to the applied potential V_{AC} as function of the frequency ω. Specially uncertainties on ω, case (b) below, make things more complicated but needs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

4. An unknown electronics circuit ("Black box", see figure 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:

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

when $y \equiv \cot\theta$, $\alpha_1 \equiv \omega_0 L/R$, $\alpha_2 \equiv 1/(\omega_0 RC)$, $x \equiv \omega/\omega_0$ 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
6.87±1.90	26390±540

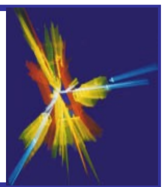


(i) Determine the L and R values (and their errors) for the "Black box" neglecting uncertainties in x. What is the P-value of your solution? [3.5p]
(ii) Determine the L and R values (and their errors) for the "Black box" taking now into account both uncertainties in x and y. Hint: Non-linear problem, solution must be found numerically. What is the P-value now? [4p]

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:

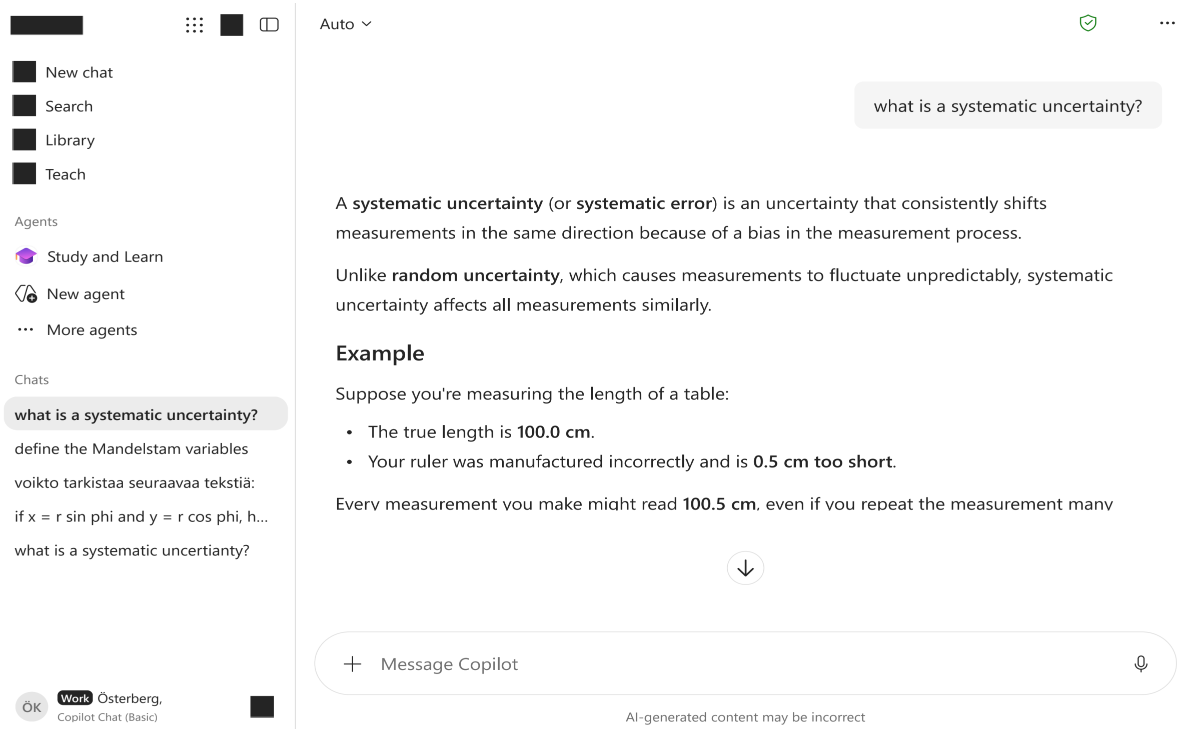
- ✓ Fundamental concepts: experimental uncertainties & their correct interpretation, frequentist & Bayesian interpretation of probability and common distributions.
- ✓ MC methods & statistical tests: Monte Carlo methods, the concept of hypothesis & test statistic, rejection of a hypothesis, discriminant 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)

• 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 ($m_p = (1.672\,621\,923\,69 \pm 0.000\,000\,000\,51) \cdot 10^{-27}$ 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 provides you with methods & tools to determine a parameter & its uncertainty correctly.

Data consistent (or not) with prediction/model? Course provides methods for sorting that out.

Probability

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

$$\text{For all } A \subset S, \quad P(A) \geq 0$$

$$\text{If } A \cap B = \emptyset \text{ (mutually exclusive)}, P(A \cup B) = P(A) + P(B)$$

$$P(S) = 1$$

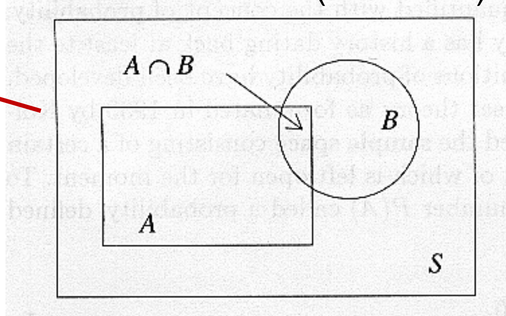
"random variable" has a specific value for each element of S.

"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 a hypothetical 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 Bayesian used e.g. for "one-off" 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 $m_H = 125.13 \pm 0.11$ GeV ?
(m_H is the mass of the Higgs boson)

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

i.e. $P(125.02 \leq m_H \leq 125.24$ GeV$) = 0.68$ (Bayesian) means:

Probability that $125.02 \leq m_H \leq 125.24$ GeV same as probability to draw a white ball out of a container with 100 balls of which 68 are white & rest black (cf. G. D'Agostini, CERN Yellow Report 99-03, 1999).
→ **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 at 2027 ice hockey world championships) = 68 %
$P(0.305 \leq \Omega_M \leq 0.325)$ = 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 2027) = ???

Frequentist interpretation: can $P(125.02 \leq m_H \leq 125.24$ GeV) = 68 % mean, consider many universes with different values for m_H; 68 % of them will have m_H-values in [125.02,125.24] GeV range?
$\Rightarrow$ usual stumbling block for the interpretation of a measurement

$P(m_H^{true}|m_H^{meas}) \propto e^{-(m_H^{true}-m_H^{meas})^2/2(\sigma_H^{meas})^2}$ for a certain m_H^{meas}, not possible to give any definitive statement about m_H^{true}

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

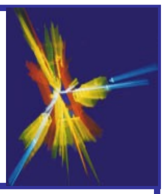
$$P(\text{theory}\,|\,\text{result}) \propto \frac{P(\text{result}\,|\,\text{theory})}{P(\text{result})} \cdot P(\text{theory})$$

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. $P(m_H)$ = 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.



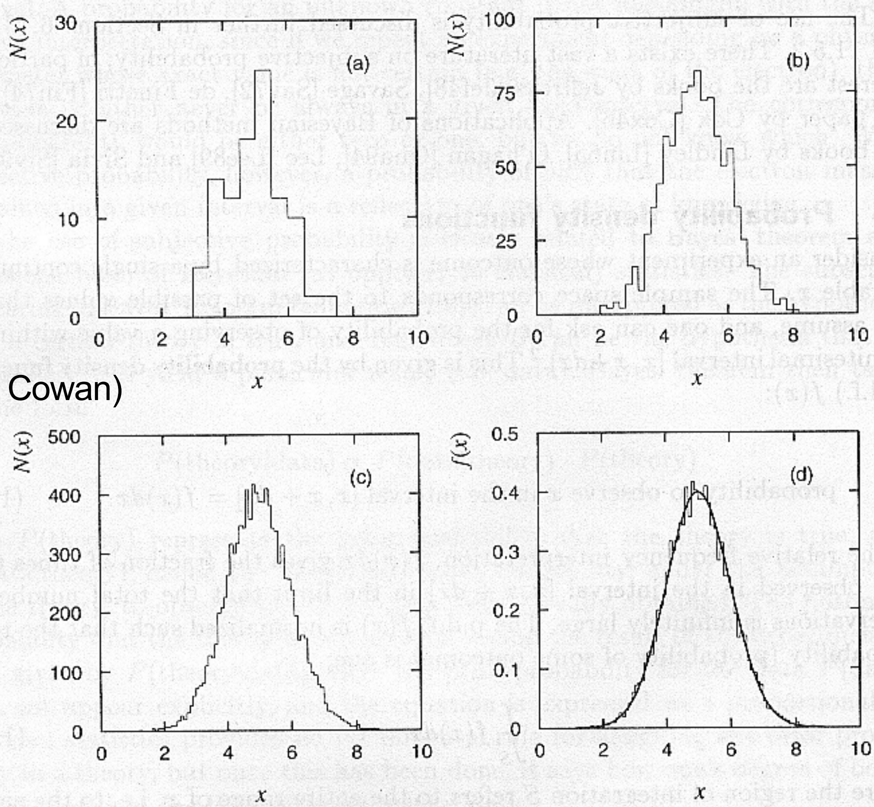
Probability density function

Let's examine meaning of measurement uncertainty σ. However let's first define the normalised distribution of repeated measurements of variable x, <u>the probability density function (pdf)</u> $f(x)$ (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 $x_i = f_i$

$$\sum_i f_i = 1$$

Probability density function

Random variables:

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

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

$\Rightarrow f(x) =$ **probability density function** (pdf) (todennäköisyys-tiheysfunktio)

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

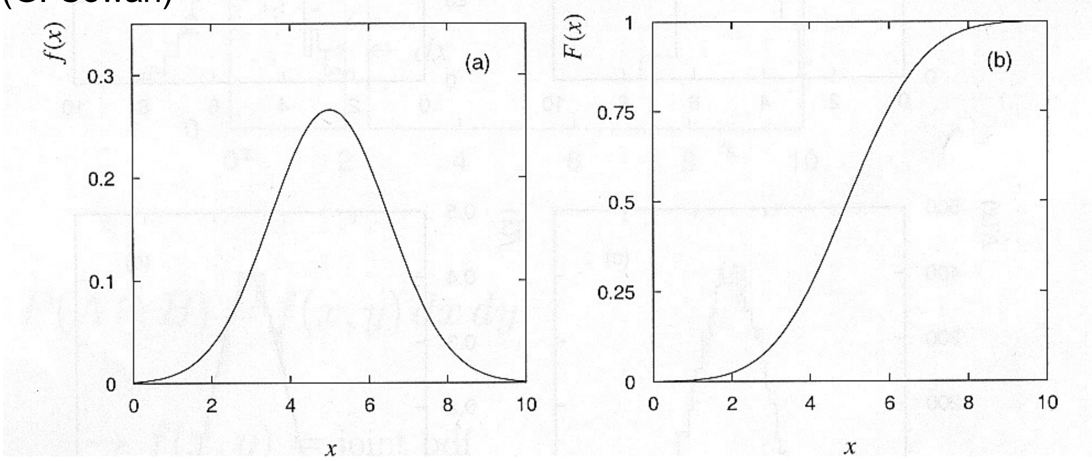
normalization: $$\int_S f(x)dx = 1 \quad (\text{e.g. } S = [-\infty, \infty]),$$

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):

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

Expectation values

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

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

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

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

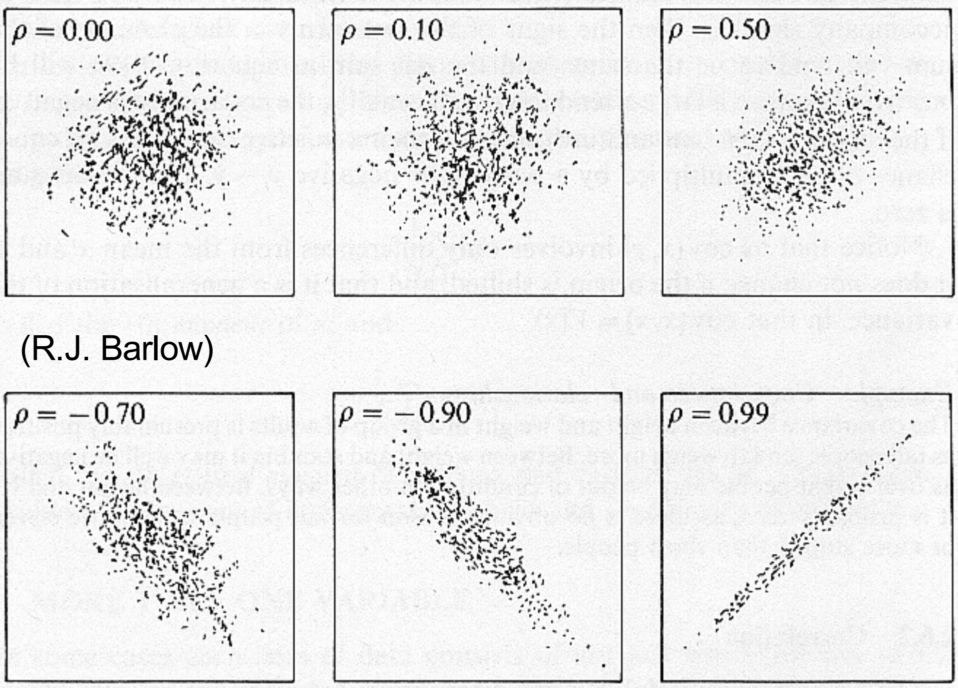
i.e. x & y 'uncorrelated'   NB! not always true the other way.

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.

Experimental uncertainties

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, calibra-tion, environment etc ... influencing result (un)knowingly to experimentalist – (usually) not statistics dependent. N.B. usually capability to determine systematic uncertainties 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.
N.B. return to subject when parameter determination is discussed

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