Tools for Physicists: Statistics

Introduction

Wolfgang Gradl

Institut für Kernphysik

Summer semester 2025

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

The scientific method: how we create ‘knowledge’

Theory / model

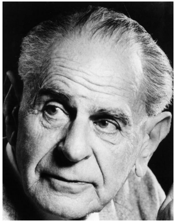
- usually mathematical
- self-consistent
- simple explanations, few (arbitrary) parameters
- testable predictions / hypotheses

Experiment

- modify or even reject theory in case of disagreement with data
- if theory requires too many adjustments it becomes unattractive
- generate surprises

Advance of scientific knowledge is *evolutionary* process with occasional revolutions

Statistical methods are important part of this process
in particular in quantitative sciences like physics



Karl Popper
(1902–1994)

Statistics in science

Statistics is needed to:

- characterise and summarise experimental results (impractical to always deal with raw data)
- quantify uncertainty of a measurement
- assess whether two measurements of the same quantity are compatible, combine measurements
- estimate parameters of an underlying model or theory
- test hypotheses:
  determine whether a model is compatible with data
- …

Aims of this mini-series

- Understand statistical concepts
  - Ability to understand physics papers
  - Know some methods / standard statistical toolbox
- Statistical inference: from data to knowledge
  - Should we believe a physics claim?
  - Develop intuition
  - Know (some) pitfalls: avoid making mistakes others have already made
- Use tools
  - Hands-on part with Python / Jupyter
  - Application to your own work? You decide!

Practical information

Two sessions:

1. Basics, introduction, statistical distributions
2. Parameter estimation, confidence intervals, hypothesis testing

About 60–90 minutes of lecture, hands-on tutorial (continued in your own time?)

I hope this will be useful for you,
but keep in mind that there is much more
to statistics than can be covered
in a few brief hours.

Useful reading material

Books:

- G. Cowan, Statistical Data Analysis
- R. Barlow, Statistics: A guide to the use of statistical methods in the physical sciences
- L. Lyons, Statistics for Nuclear and Particle Physicists
- A. J. Bevan, Statistical data analysis for the physical sciences
- G. Bohm, G. Zech, Introduction to Statistics and Data Analysis for Physicists (available online)

Lectures on the web:

- G. Cowan, Royal Holloway University London: Statistical Data Analysis
- K. Reygers, U Heidelberg, Stat. Methods in Particle Physics

Dealing with uncertainty

- Underlying theory is probabilistic (quantum mechanics / QFT)
  source of **true** randomness
- Limited knowledge about measurement process
  even without QM
  random measurement errors
- Things we could know in principle, but don't
  e.g. from limitations of cost, time, …

Quantify uncertainty using tools and concepts from **probability**

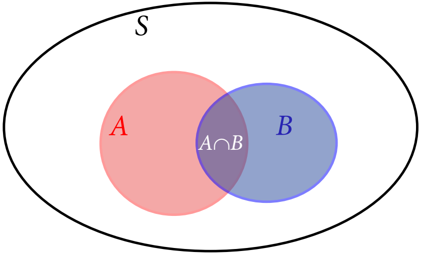
Mathematical definition of probability

Kolmogorov axioms:

Consider a set S (the **sample space**) with subsets A, B, …(**events**).

Define a function on the power set of S, $P : \mathfrak{P}(S) \mapsto [0, 1]$ with

1. $P(A) \geq 0$ for all $A \subset S$
2. $P(S) = 1$
3. $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$,
   i.e. when A and B are exclusive



From these we can derive further properties:

- $P(\bar{A}) = 1 - P(A)$
- $P(A \cup \bar{A}) = 1$
- $P(\emptyset) = 0$
- If $A \subset B$, then $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

for the mathematically inclined: proper treatment will use *measure theory*

Interpretation — intuition about probability

- Classical definition
  - Assign equal probabilities based on symmetry of problem, e.g. rolling ideal dice: $P(6) = 1/6$
  - difficult to generalise, sounds somewhat circular

- Frequentist: relative frequency, proportion of outcomes
  - $A, B, \ldots$ outcomes of a repeatable experiment

$$P(A) = \lim_{n\to\infty} \frac{\text{times outcome is } A \text{ in } n \text{ repetitions}}{n}$$

- Bayesian: subjective probability, degree of belief
  - $A, B, \ldots$ are hypotheses (statements that are either true or false)

$$P(A) = \text{degree of belief that } A \text{ is true}$$

…all three definitions consistent with Kolmogorov's axioms

Conditional probability, independent events

Conditional probability for two events A and B:

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

"probability of A given B"

Example: rolling dice

$$P(n < 3|n \text{ even}) = \frac{P((n < 3) \cap (n \text{ even}))}{P(n \text{ even})} = \frac{1/6}{1/2} = 1/3$$

Events A and B independent $\iff P(A \cap B) = P(A) \cdot P(B)$

A is independent of B if $P(A|B) = P(A)$

JG|U

Bayes' theorem

Use definition of conditional probability:

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

But obviously $P(A \cap B) = P(B \cap A)$, so:

Theorem

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

Allows to 'invert' statements about probability:
of great interest to us. Want to infer $P(\text{theory}|\text{data})$ from $P(\text{data}|\text{theory})$

Often these two are confused, knowingly or unknowingly
(advertising, political campaigns, …)

Bayes' theorem: degree of belief in a theory

likelihood

prior (before seeing the data, subjective)

$$P(\text{theory}|\text{data}) = \frac{P(\text{data}|\text{theory})P(\text{theory})}{P(\text{data})}$$

posterior probability, i.e., after seeing the data

normalization

Example for Bayes' theorem: Rare disease

Base probability (for anyone) to have a disease D:

$$P(D) = 0.0001$$
$$P(\text{no } D) = 0.9999$$

Example for Bayes' theorem: Rare disease

Base probability (for anyone) to have a disease D:

$$P(D) = 0.0001$$
$$P(\text{no } D) = 0.9999$$

Consider a test for D: result is positive or negative (+ or –):

$$P(+|D) = 0.98 \qquad P(+|\text{no } D) = 0.03$$
$$P(-|D) = 0.02 \qquad P(-|\text{no } D) = 0.97$$

Example for Bayes' theorem: Rare disease

Base probability (for anyone) to have a disease D:

$$P(D) = 0.0001$$
$$P(\text{no } D) = 0.9999$$

Consider a test for D: result is positive or negative (+ or –):

$$P(+|D) = 0.98 \qquad P(+|\text{no } D) = 0.03$$
$$P(-|D) = 0.02 \qquad P(-|\text{no } D) = 0.97$$

Suppose your result is +; should you be worried?

Example for Bayes' theorem: Rare disease

Base probability (for anyone) to have a disease D:

$$P(D) = 0.0001$$
$$P(\text{no } D) = 0.9999$$

Consider a test for D: result is positive or negative (+ or –):

$$P(+|D) = 0.98 \qquad P(+|\text{no } D) = 0.03$$
$$P(-|D) = 0.02 \qquad P(-|\text{no } D) = 0.97$$

Suppose your result is +; should you be worried?

$$\begin{aligned} P(D|+) &= \frac{P(+|D)\,P(D)}{P(+|D)\,P(D) + P(+|\text{no } D)\,P(\text{no } D)} \\ &= \frac{0.98 \times 0.0001}{0.98 \times 0.0001 + 0.03 \times 0.9999} = 0.0033 \end{aligned}$$

Probability that you have disease is 0.33%, i.e. you're probably ok

Digression: what if prevalence is (much) higher?

Assume 100× higher prevalence in population:

$$P(D) = 0.01$$
$$P(\text{no } D) = 0.99$$

Then,

$$P(D|+) = \frac{P(+|D)\,P(D)}{P(+|D)P(D) + P(+|\text{no } D)P(\text{no } D)}$$
$$= \frac{0.98 \times 0.01}{0.98 \times 0.01 + 0.03 \times 0.99} = 0.248$$

should you be worried? This can't be answered by statistics, of course …

At least take another (independent) test …

Classification

Population P that either carries (P) or does not carry (N) a specific marker

D or no *D*, signal candidate or background event, …

Classifier ("test"): predict positive (PP) or negative (PN) outcome

$+$ or $-$

Confusion matrix

		predicted	
		predicted pos.	**predicted neg.**
actual	**positive**	true positive	false negative
	negative	false positive	true negative

Statisticians call these errors:

Type I error: false positive

Type II error: false negative

Classification

$$\text{sensitivity} = P(+|D) = \frac{\text{true positives}}{\text{actual positives}} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

Higher sensitivity: lower type II error rate

$$\text{specificity} = P(-|\text{no } D) = \frac{\text{true negatives}}{\text{actual negatives}} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$$

Higher specificity: lower type I error rate

Problem e.g. in machine learning:

Given a concrete classifier, how can we pick the 'best' threshold?

Frequentists vs. Bayesians

- Criticisms of the frequentist interpretation
  - $n \to \infty$ can never be achieved in practice. When is n large enough?
  - Want to talk about probabilities of events that are not repeatable
    - $P(\text{rain tomorrow})$ — but there's only one tomorrow
    - $P(\text{Universe started with a big bang})$ — only one universe available
  - P is not an intrinsic property of A, but depends on how the ensemble of possible outcomes was constructed
    - $P(\text{person I talk to is a physicist})$ strongly depends on whether I am at a conference or at the beach

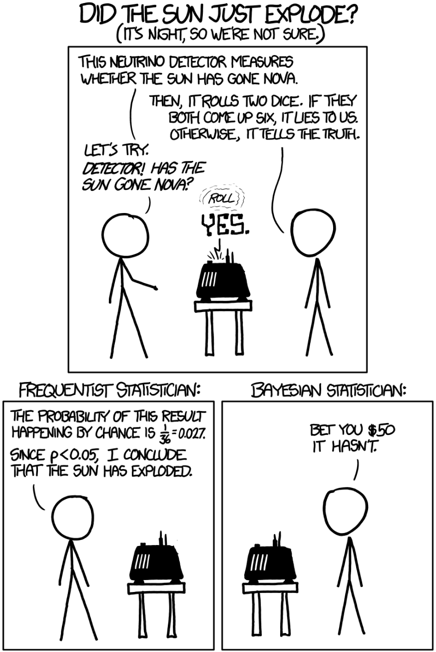
Frequentists vs. Bayesians

- Criticisms of the frequentist interpretation
  - $n \to \infty$ can never be achieved in practice. When is n large enough?
  - Want to talk about probabilities of events that are not repeatable
    - $P(\text{rain tomorrow})$ — but there's only one tomorrow
    - $P(\text{Universe started with a big bang})$ — only one universe available
  - P is not an intrinsic property of A, but depends on how the ensemble of possible outcomes was constructed
    - $P(\text{person I talk to is a physicist})$ strongly depends on whether I am at a conference or at the beach
- Criticisms of the subjective interpretation
  - 'Subjective' estimate has no place in science
  - How can quantify the prior state of our knowledge?

Frequentists vs. Bayesians

- Criticisms of the frequentist interpretation
  - $n \to \infty$ can never be achieved in practice. When is n large enough?
  - Want to talk about probabilities of events that are not repeatable
    - P(rain tomorrow) — but there's only one tomorrow
    - P(Universe started with a big bang) — only one universe available
  - P is not an intrinsic property of A, but depends on how the ensemble of possible outcomes was constructed
    - P(person I talk to is a physicist) strongly depends on whether I am at a conference or at the beach
- Criticisms of the subjective interpretation
  - 'Subjective' estimate has no place in science
  - How can quantify the prior state of our knowledge?

'Bayesians address the questions everyone is interested in by using assumptions that no one believes, while Frequentists use impeccable logic to deal with an issue that is of no interest to anyone'
— Louis Lyons



https://xkcd.com/1132/

Describing data

Random variables and probability density functions

Random variable:

- Variable whose possible values are numerical outcomes of a random phenomenon

Probability density function (pdf) of a continuous variable:

$$P(X \text{ found in } [x, x + \mathrm{d}x]) = p(x)\mathrm{d}x$$

Normalisation:

$$\int_{-\infty}^{+\infty} p(x)\mathrm{d}x = 1 \qquad x \text{ must be somewhere}$$

JG|U

Visualisation: Histograms

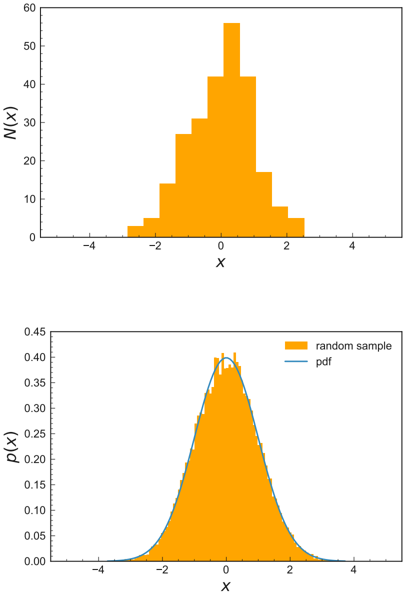
Histogram

- representation of the frequencies of numerical outcome of a random phenomenon

pdf $\simeq$ histogram for

- infinite data sample
- zero bin width
- normalised to unit area

$$p(x) = \lim_{\Delta x \to 0} \frac{N(x)}{N \Delta x}$$



Summary statistics: median, mean, and mode

Arithmetic **mean** of a data sample ('sample mean'):

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

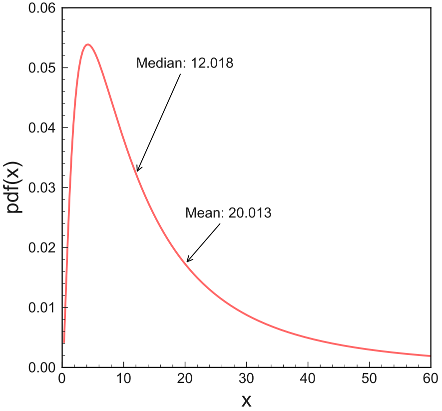
Mean of a pdf:

$$\mu \equiv \langle x \rangle \equiv \int x\, p(x) \mathrm{d}x$$

$$\equiv \text{ expectation value } E[x]$$

Median:
point with 50% probability above and 50% prob. below

Mode:
most likely value



not necessarily the same, for skewed distributions

Variance, standard deviation

Variance of a **distribution** (pdf):

$$V(x) = \int \mathrm{d}x\, p(x)\,(x-\mu)^2 = E[(x-\mu)^2]$$

Variance of a **data sample**

$$V(x) = \frac{1}{N}\sum_i (x_i-\mu)^2 = \overline{x^2} - \mu^2$$

Requires knowledge of *true* mean μ.

Replacing μ by sample mean $\bar{x}$ results in underestimated variance!

Instead, use this:

$$\hat{V}(x) = \frac{1}{N-1}\sum_i (x_i-\bar{x})^2$$

Standard deviation:

$$\sigma = \sqrt{V(x)}$$

Variance, standard deviation

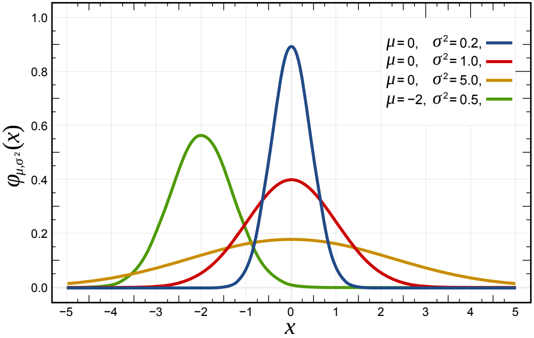
Why this definition of ‘spread’ of a distribution?

Gaussian pdf

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

Mean: $E[x] = \mu$

Variance: $V[x] = \sigma^2$



Standard normal distribution: $\mu = 0, \sigma = 1$

Cumulative distribution related to error function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz = \frac{1}{2}\left[\text{erf}\left(\frac{x}{\sqrt{2}}\right) + 1\right]$$

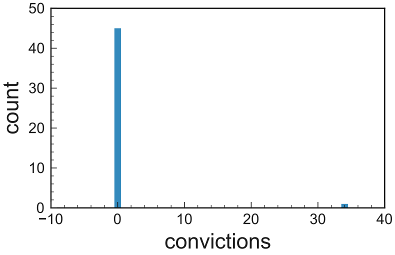
In Python: `scipy.stats.norm(loc, scale)`

Robustness?

Beware of distributions with large outliers:

Sample mean and variance as defined above not very good ('robust') estimators for the shape of the bulk of the distribution, can be grossly misleading!

Robust statistics deals with methods how to handle this — for a short writeup and pointers to literature, see *e.g.* **`https://www.stats.ox.ac.uk/~ripley/StatMethods/Robust.pdf`**



As of 31st May 2024, the average US president has been convicted of 0.74 felonies

Multivariate distributions

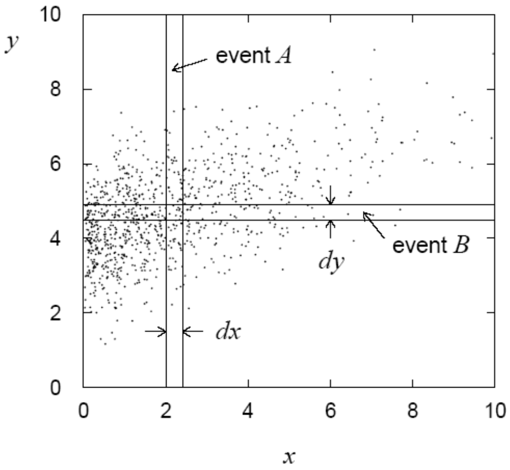
Outcome of an experiment characterised by tuple $(x_1, \ldots, x_n)$

$$P(A \cap B) = f(x, y) \mathrm{d}x \, \mathrm{d}y$$

with $f(x, y)$ the 'joint pdf'

Normalisation

$$\int \cdots \int f(x_1, \ldots, x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n = 1$$



Sometimes, only the pdf of one component is wanted:

$$f_1(x_1) = \int \cdots \int f(x_1, \ldots, x_n) \mathrm{d}x_2 \cdots \mathrm{d}x_n$$

≈ projection of joint pdf onto individual axis: marginalised pdf

Covariance and correlation

Covariance:

$$\mathrm{cov}[x,y] = E[(x-\mu_x)(y-\mu_y)]$$

Correlation coefficient:

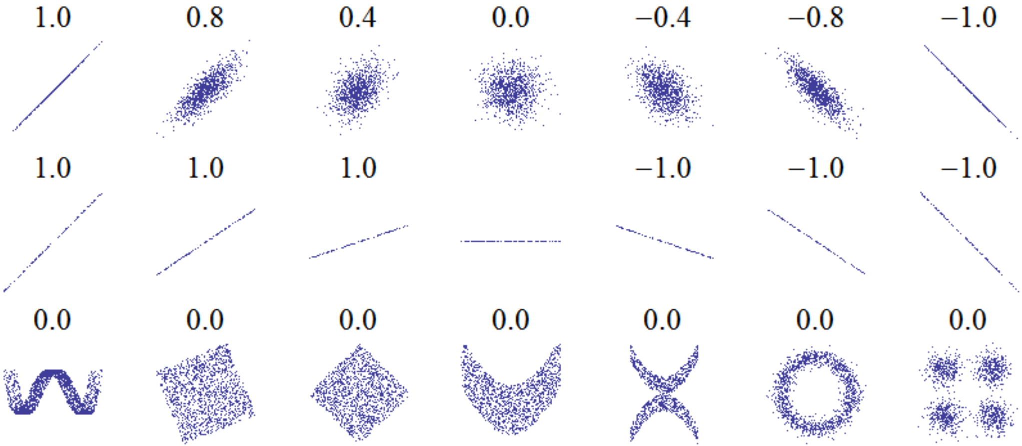
$$\rho_{xy} = \frac{\mathrm{cov}[x,y]}{\sigma_x\,\sigma_y}$$

If x, y independent:
pdf factorises, *i.e.* $f(x,y) = f_x(x)\,f_y(y)$,
and covariance becomes

$$E[(x-\mu_x)(y-\mu_y)] = \int (x-\mu_x) f_x(x)\mathrm{d}x \int (y-\mu_y) f_y(y)\mathrm{d}y = 0$$

Note: converse not necessarily true

Covariance and correlation



Same (linear) correlation coefficient, but very different 2D shapes!

Always visualise your data!

```python
import pandas as pd ···
```

```python
dataset = pd.read_csv('./ds.csv', header=None, names=['x', 'y'])
dataset
```

[2] ✓ 0.7s Python

	x	y
0	55.3846	97.1795
1	51.5385	96.0256
2	46.1538	94.4872
3	42.8205	91.4103
4	40.7692	88.3333
...	...	...
137	39.4872	25.3846
138	91.2821	41.5385
139	50.0000	95.7692
140	47.9487	95.0000
141	44.1026	92.6923

142 rows × 2 columns

Always visualise your data!

```python
dataset.describe()
```

[12] ✓ 0.7s Python

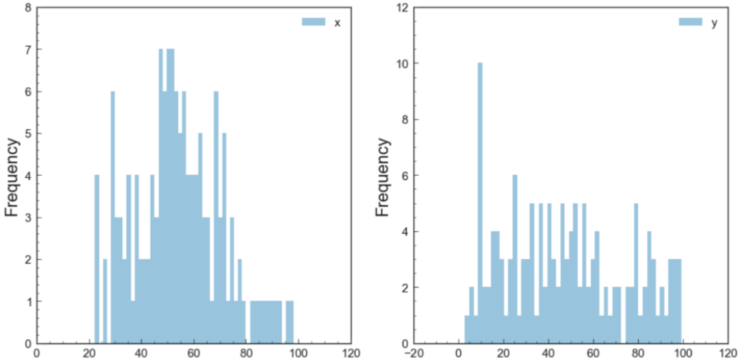
...

	x	y
count	142.000000	142.000000
mean	54.263273	47.832253
std	16.765142	26.935403
min	22.307700	2.948700
25%	44.102600	25.288450
50%	53.333300	46.025600
75%	64.743600	68.525675
max	98.205100	99.487200

Always visualise your data!

```python
fig,axs = plt.subplots(1,2,figsize=(16,8))
dataset.plot('x', kind='hist', bins=50, alpha=0.5, ax=axs[1])
dataset.plot('y', kind='hist', bins=50, alpha=0.5, ax=axs[0]);
```

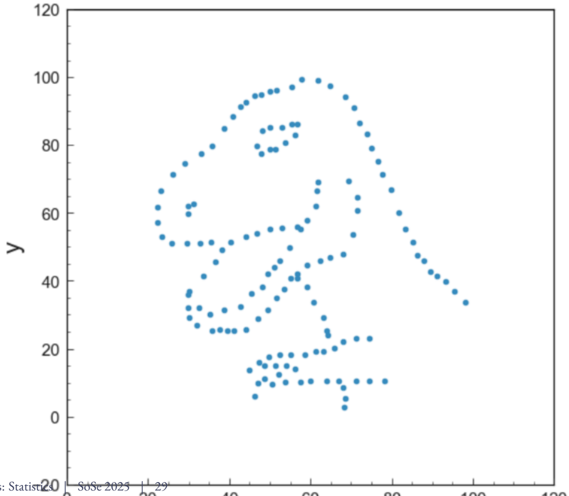
[13] 1.1s Python



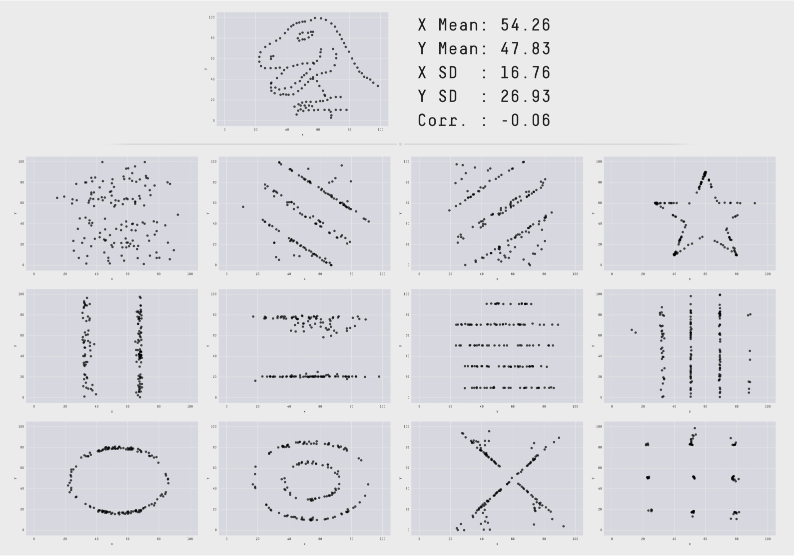
Always visualise your data!

```python
fig, ax = plt.subplots(1,1,figsize=(8,8))
dataset.plot.scatter(x='x', y='y', ax=ax);
```

[16] ✓ 0.3s Python



Always visualise your data!



https://www.autodesk.com/research/publications/same-stats-different-graphs

Linear combinations of random variables

Consider two random variables x and y with known covariance $\textbf{cov}[x, y]$

$$\begin{aligned}\langle x + y\rangle &= \langle x\rangle + \langle y\rangle \\ \langle ax\rangle &= a\,\langle x\rangle \\ V[ax] &= a^2 V[x] \\ V[x + y] &= V[x] + V[y] + 2\,\textbf{cov}[x, y]\end{aligned}$$

For uncorrelated variables, simply add variances.

How about combination of N independent measurements (estimates) of a quantity, $x_i \pm \sigma$, all drawn from the same underlying distribution?

$$\begin{aligned}\bar{x} &= \frac{1}{N}\sum x_i \quad \text{best estimate} \\ V[N\bar{x}] &= N^2\sigma \\ \sigma_{\bar{x}} &= \frac{1}{\sqrt{N}}\sigma\end{aligned}$$

Combination of measurements: weighted mean

Suppose we have N independent measurements of the same quantity, but each with a different uncertainty: $x_i \pm \delta_i$

Weighted sum:

$$x = w_1 x_1 + w_2 x_2$$
$$\delta^2 = w_1^2 \delta_1^2 + w_2^2 \delta_2^2$$

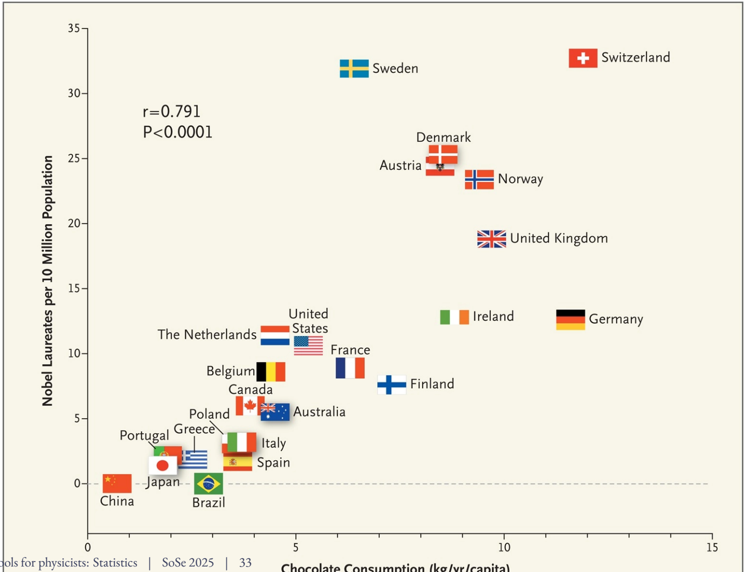
Determine weights w_1, w_2 under constraint $w_1 + w_2 = 1$ such that δ^2 is minimised:

$$w_i = \frac{1/\delta_i^2}{1/\delta_1^2 + 1/\delta_2^2}$$

If original raw data of the two measurements are available, can improve this estimate by combining raw data

alternatively, use log-likelihood curves to combine measurements

Correlation ≠ causation



Correlation coefficient: 0.791

significant correlation
($p < 0.0001$)

0.4 kg/year/capita to produce one additional Nobel laureate

improved cognitive function associated with regular intake of dietary flavonoids?

Tools

Usable and useful tools (e.g. for your analysis) depend on environment / external constraints and other factors

- within working group
- international collaboration
- personal preferences
- …

Don’t underestimate the cost of choosing a different approach than everyone else around you!

It may be worth it, though; just be aware of the implications!

For example: R vs python vs ROOT? Well-maintained or niche packages in python?

Tools

From my own experience with data analysis in HEP experiments:

- To paraphrase Willem van der Poel's 'Zero One Infinity' rule:
  The only numbers you should care about are Zero, One, and Infinity
  If you have to do something more than once, automate!

Tools

From my own experience with data analysis in HEP experiments:

- To paraphrase Willem van der Poel's 'Zero One Infinity' rule:
  The only numbers you should care about are Zero, One, and Infinity
  If you have to do something more than once, automate!
- Corollary: interactive tools are nice, but scripts are much better 'in production', especially to produce plots
  By all means explore your data using JupyterLab or other interactive tools, but then export the result as executable script

Tools

From my own experience with data analysis in HEP experiments:

- To paraphrase Willem van der Poel's 'Zero One Infinity' rule:
  The only numbers you should care about are Zero, One, and Infinity
  If you have to do something more than once, automate!
- Corollary: interactive tools are nice, but scripts are much better 'in production', especially to produce plots
  By all means explore your data using JupyterLab or other interactive tools,
  but then export the result as executable script
- Use a version control system, such as `git`, to keep track of changes in your code

Tools

From my own experience with data analysis in HEP experiments:

- To paraphrase Willem van der Poel's 'Zero One Infinity' rule:
  The only numbers you should care about are Zero, One, and Infinity
  If you have to do something more than once, automate!
- Corollary: interactive tools are nice, but scripts are much better 'in production', especially to produce plots
  By all means explore your data using JupyterLab or other interactive tools, but then export the result as executable script
- Use a version control system, such as `git`, to keep track of changes in your code
- Make use of well-maintained libraries, toolkits &c for common tasks
  Yes, you can write your own algorithms to perform function minimisation or matrix inversion, and it is very instructive to do so
  — but should you use this 'in production'?

Examples and interactive demo

For this course: `https://bit.ly/tfp-statistics-2025`

Some important distributions

Binomial distribution

N independent experiments

- Outcome of each is either ‘success’ or ’failure’
- Probability for success is p

$$f(k;N,p) = \binom{N}{k} p^k (1-p)^{N-k} \qquad E[k] = Np \qquad V[k] = Np(1-p)$$

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}$$

binomial coefficient: number of permutations to have k successes in N tries

Use binomial distribution to model processes with two outcomes

Example: detection efficiency = #(particles seen by detector) / #(all particles passing detector)

In the limit $N \to \infty, p \to 0, Np = \nu =$ const, binomial distribution can be approximated by a Poisson distribution

Poisson distribution

$$p(k;\nu) = \frac{\nu^k}{k!} e^{-\nu}$$

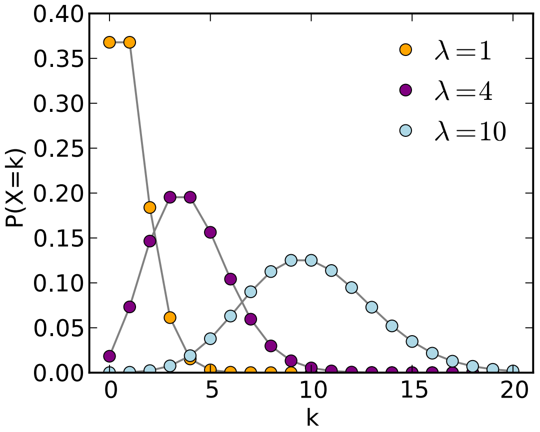
$$E[k] = \nu; \quad V[k] = \nu$$

Properties:

- If n_1, n_2 follow Poisson distribution, then also $n_1 + n_2$
- Can be approximated by Gaussian for large ν

Examples:

- Clicks of a Geiger counter in a given time interval
- Cars arriving at a traffic light in one minute



Poisson distribution

$$p(k;\nu) = \frac{\nu^k}{k!} e^{-\nu}$$

$$E[k] = \nu; \quad V[k] = \nu$$

Properties:

- If n_1, n_2 follow Poisson distribution, then also $n_1 + n_2$
- Can be approximated by Gaussian for large ν

Examples:

- Clicks of a Geiger counter in a given time interval
- Cars arriving at a traffic light in one minute

probability of k events occurring in fixed interval of time if events …

... occur with constant rate

... independently of time since last event

Poisson distribution

$$p(k;\nu) = \frac{\nu^k}{k!} e^{-\nu}$$

$$E[k] = \nu; \quad V[k] = \nu$$

Properties:

- If n_1, n_2 follow Poisson distribution, then also $n_1 + n_2$
- Can be approximated by Gaussian for large ν

Examples:

- Clicks of a Geiger counter in a given time interval
- Cars arriving at a traffic light in one minute

Rare events:

- Number of Prussian cavalrymen killed by horse-kicks
  Observe 10 army corps over 20 years:
  122 deaths due to horse kicks,
  therefore on average 0.61 deaths / (corps × year)

Number of deaths in 1 corps in 1 year	Actual number of such cases	Poisson prediction
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.1
4	1	0.6

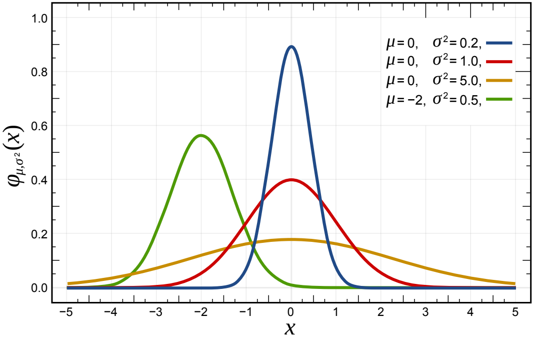
Gaussian

A.k.a. normal distribution

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

Mean: $E[x] = \mu$

Variance: $V[x] = \sigma^2$



Standard normal distribution: $\mu = 0, \sigma = 1$

Cumulative distribution related to error function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz = \frac{1}{2}\left[\mathrm{erf}\left(\frac{x}{\sqrt{2}}\right) + 1\right]$$

In Python: `scipy.stats.norm(loc, scale)`

Why are Gaussians so useful?

Central limit theorem: sum of n random variables approaches Gaussian distribution, for large n

True, if fluctuation of sum is not dominated by the fluctuation of one (or a few) terms

- Good example: velocity component v_x of air molecules
- So-so example: total deflection due to multiple Coulomb scattering.
  Rare large angle deflections give non-Gaussian tail
- Bad example: energy loss of charged particles traversing thin gas layer.
  Rare collisions make up large fraction of energy loss ➡ Landau PDF

p-value

Probability for a Gaussian distribution corresponding to $[\mu - Z\sigma, \mu + Z\sigma]$:

$$P(Z\sigma) = \frac{1}{\sqrt{2\pi}} \int_{-Z}^{+Z} e^{-\frac{x^2}{2}} = \Phi(Z) - \Phi(-Z) = \mathrm{erf}\left(\frac{Z}{\sqrt{2}}\right)$$

68.27% of area within $\pm 1\sigma$
95.45% of area within $\pm 2\sigma$
99.73% of area within $\pm 3\sigma$

90% of area within $\pm 1.645\sigma$
95% of area within $\pm 1.960\sigma$
99% of area within $\pm 2.576\sigma$

p-value:
probability that random process (fluctuation) produces a measurement at least this far from the true mean

$$p\text{-value} := 1 - P(Z\sigma)$$

Available in ROOT: `TMath::Prob(Z*Z)`
and Python: `2*stats.norm.sf(Z)`

Deviation	p-value (%)
1σ	31.73
2σ	4.55
3σ	0.270
4σ	0.006 33
5σ	0.000 057 3

χ^2 distribution

$x_1, \ldots, x_n$ be n independent standard normal ($\mu = 0, \sigma = 1$) random variables. Then the sum of their squares

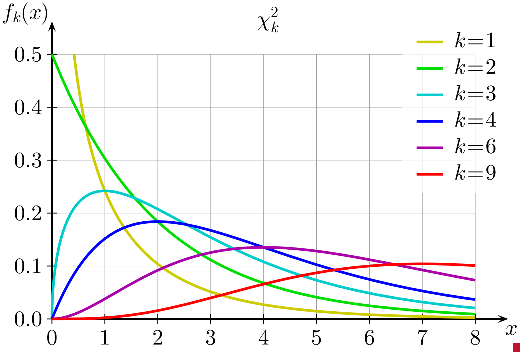
$$z = \sum_{i=1}^{n} x_i^2 = \sum_i \frac{(x' - \mu')^2}{\sigma'^2}$$

follows a χ^2 distribution with n degrees of freedom.

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

$$E[z] = n, \quad V[z] = 2n$$

Quantify goodness of fit, compatibility of measurements, …



Student's t distribution

Let $x_1, \ldots, x_n$ be distributed as $N(\mu, \sigma)$.

Sample mean and estimate of variance:
$$\bar{x} = \frac{1}{n}\sum_i x_i, \quad \hat{\sigma}^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$$

Don't know true μ, therefore have to estimate variance by $\hat{\sigma}$.

$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ follows $N(0,1)$

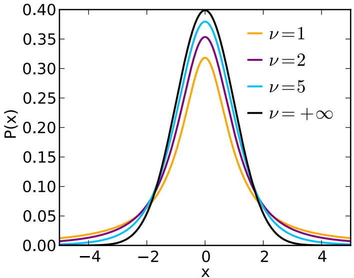
$\frac{\bar{x}-\mu}{\hat{\sigma}/\sqrt{n}}$ not Gaussian: Student's t-distribution with $n-1$ d.o.f.

$$f(t;n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

For $n \to \infty$, $f(t;n) \to N(t;0,1)$

Applications:

- Hypothesis tests: assess statistical significance between two sample means
- Set confidence intervals (more of that later)



JG|U