Statistical Methods in Particle Physics

1. Basics Concepts

Prof. Dr. Klaus Reygers (lectures)
Dr. Sebastian Neubert (tutorials)

Heidelberg University
WS 2017/18

Introduction

Aims of this Course

- Statistical inference: from data to knowledge
  - Should a believe a physics claim?
  - Develop intuition
  - Know pitfalls: avoid mistakes already made by others
- Understand statistical concepts
  - Ability to understand physics papers
  - Know methods / the standard statistical toolbox
- Use tools
  - Learn to use root
  - Get ready for your own data analysis

How Knowledge is Created?

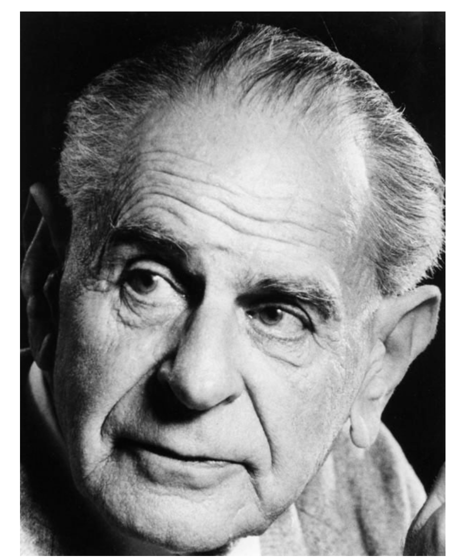
Guess theory/model

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

Perform experiment

- reject / modify theory in case of disagreement with data
- if theory requires too many adjustments it becomes unattractive

The advance of scientific knowledge is an *evolutionary* process

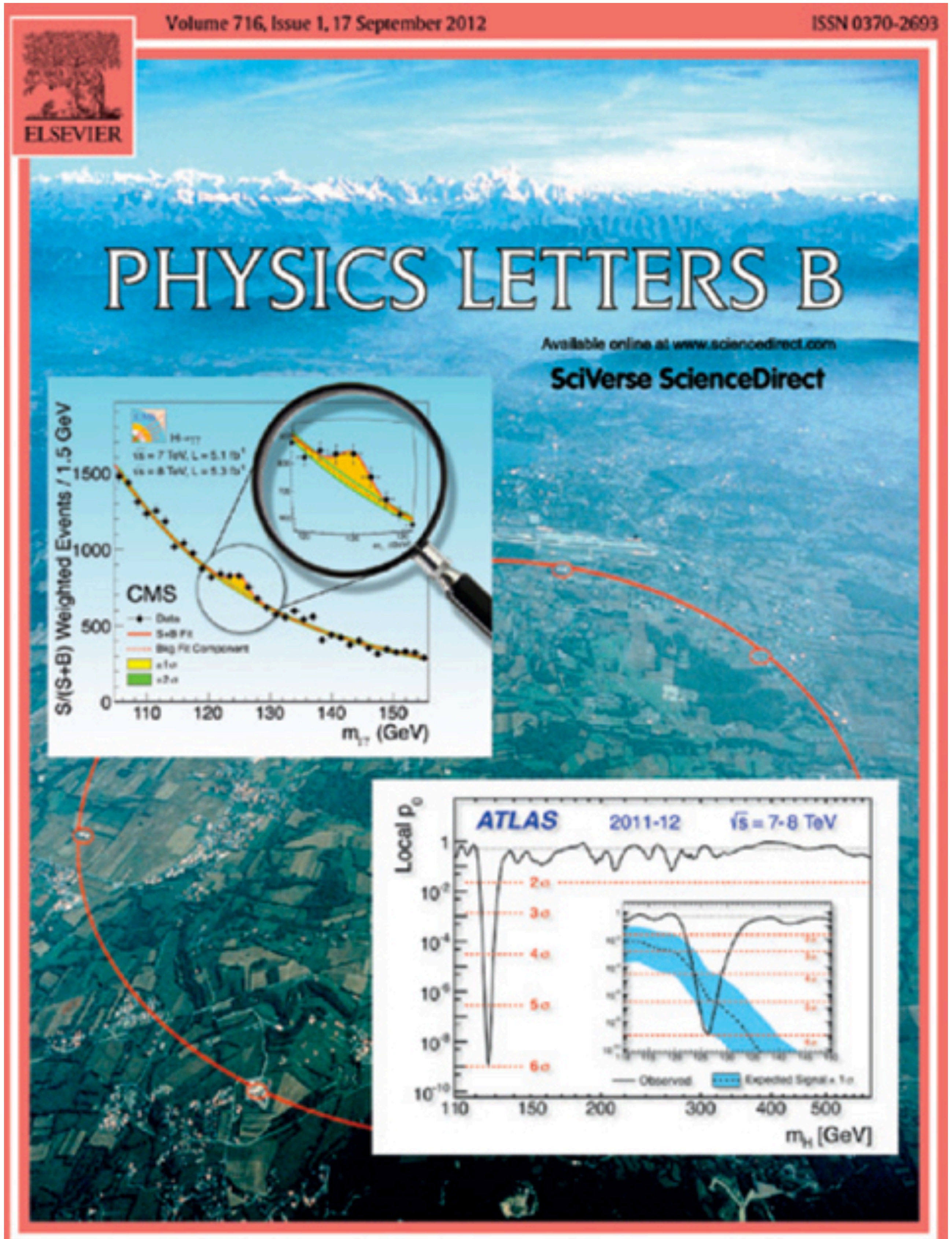


source: Wikipedia

Karl Popper (1902–1994)

Statistical methods are an important part of this process

Understanding Particle Physics Papers



Physics Letters B

Volume 716, Issue 1, 17 September 2012, Pages 1–29

Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC ☆

This paper is dedicated to the memory of our ATLAS colleagues who did not live to see the full impact and significance of their contributions to the experiment.

ATLAS Collaboration *

G. Aad[48], T. Abajyan[21], B. Abbott[111], J. Abdallah[12], S. Abdel Khalek[115], A.A. Abdelalim[49], O. Abdinov[11], R. Aben[105], B. Abi[112], M. Abolins[88], O.S. AbouZeid[158], H. Abramowicz[153], H. Abreu[136], B.S. Acharya[164a, 164b], L. Adamczyk[38], D.L. Adams[25], T.N. Addy[56], J. Adelman[176], S. Adomeit[98], P. Adragna[75], T. Adye[129], S. Aefsky[23], J.A. Aguilar-Saavedra[124b, a], M. Agustoni[17], M. Aharrouche[81], S.P. Ahlen[22], F. Ahles[48], A. Ahmad[148], M. Ahsan[41], G. Aielli[133a, 133b], T. Akdogan[19a],

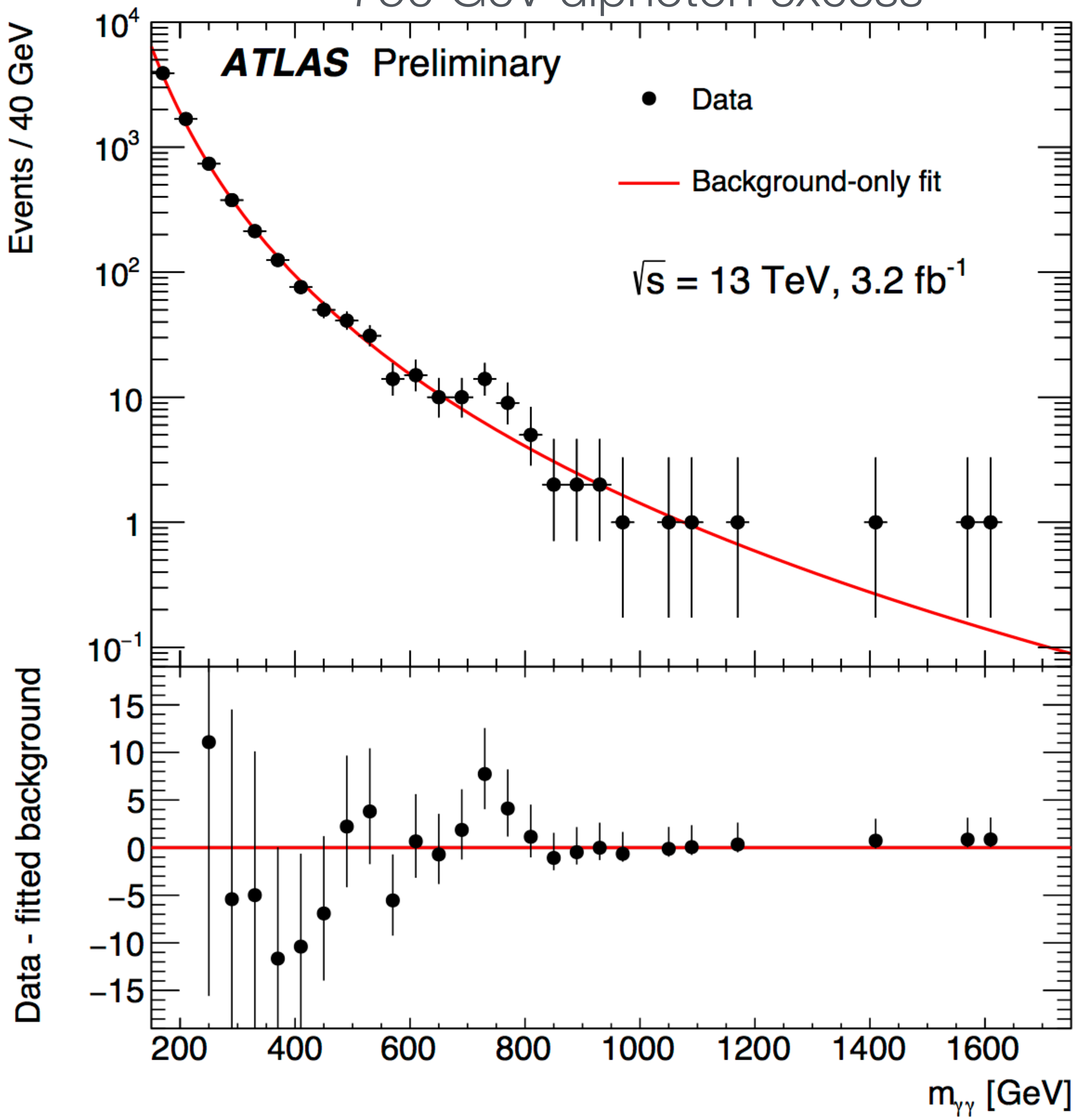
Show more

doi:10.1016/j.physletb.2012.08.020

Get rights and content

A Heavy Higgs Boson?

"750 GeV diphoton excess"



- Two-photon invariant mass spectrum
- New particle with mass $m \approx 750$ GeV?
- Local significance: 3.6σ

Peak disappeared with more data … [link]

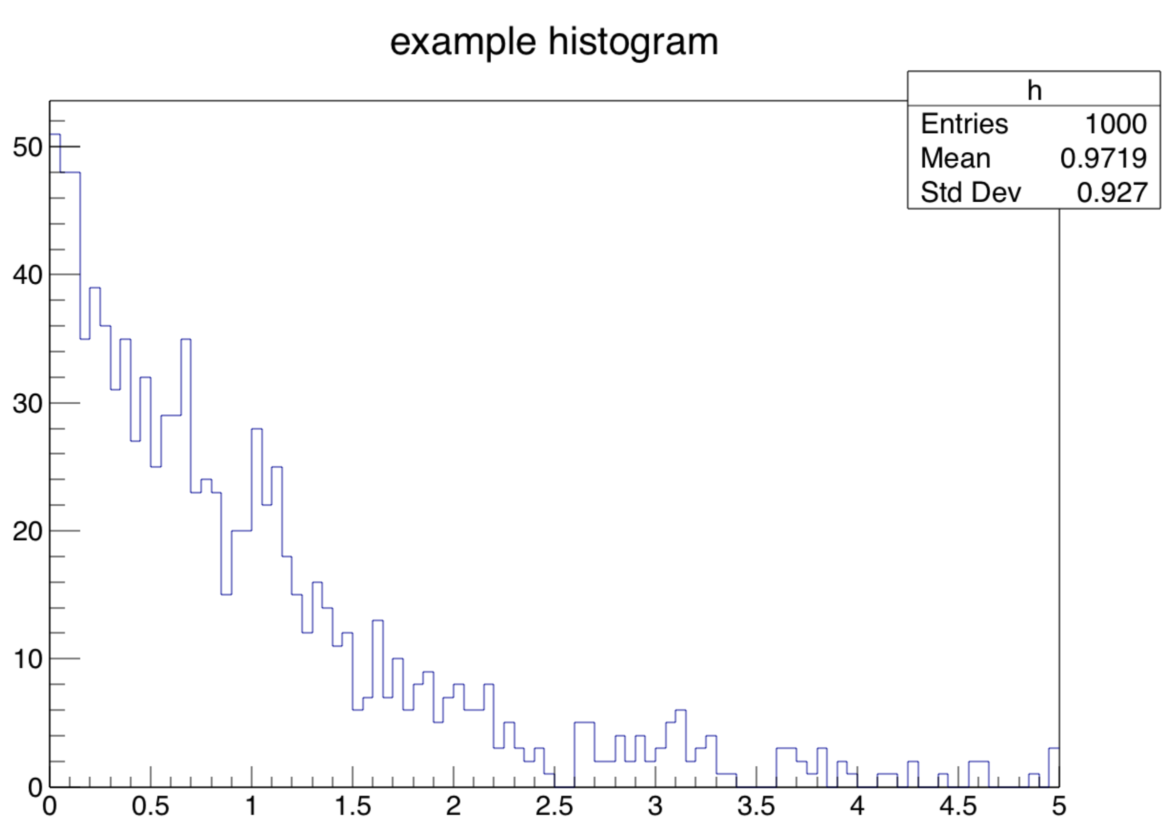
Presentations by CMS and ATLAS, December 2015: https://indico.cern.ch/event/442432/

This is an Applied Course

https://root.cern.ch/

ROOT
Data Analysis Framework

```
root [1] TH1F h("h","example histogram",100,0.,5.);
root [2] ifstream inp; double x;
root [3] inp.open("expo.dat");
root [4] while (inp >> x) { h.Fill(x); }
root [5] h.Draw();
root [6] inp.close();
```



- We will use lots of examples from “real life” particle physics
- We will sometimes talk about implementation on a computer
- You should ask questions, discuss
- You will write code (C++), the tutorials will provide a step-by-step introduction to root

Topics

1. Basics concepts
   - Probability
   - Mean, median, mode
   - Covariance and correlation
2. Examples of probability distributions
3. Uncertainty
   - Statistical and systematic uncertainties
   - Propagation of uncertainties
   - Combination of uncorrelated measurements
4. Monte Carlo and numerical methods
   - Generation of random numbers
   - Monte Carlo integration
   - Applications in HEP
5. Parameter estimation
   - Basics: consistency, bias, efficiency
   - Maximum likelihood method
   - The method of least squares
6. Hypothesis testing
   - χ^2 test
   - Significance
   - Neyman-Pearson construction
7. Confidence limits and intervals
8. Multivariate analysis
9. Unfolding

https://uebungen.physik.uni-heidelberg.de/vorlesung/20172/smipp

Practical Information (I)

- Slides of the lecture will be provided on the lecture web site
  - https://uebungen.physik.uni-heidelberg.de/vorlesung/20172/smipp
  - Goal: slides available a couple of days before the lecture
- Weekday/time of the lecture
  - Mondays, 14:15–15:45, KIP SR 3.404
  - There were requests to change the week, but this turned out to be difficult
- Tutorials
  - Mondays, 16:00–17:30
  - CIP pool of the Physikalisches Institut, **not in KIP CIP pool**
  - Information on CIP pool:
    http://www.physi.uni-heidelberg.de/Einrichtungen/CIP
  - Homework problems will be made available on lecture website
  - Solutions to be handed in by Wednesday, 12:00, of the following week
  - Groups of two students can (actually should!) hand in homework together
  - First homework sheet is available,
    to be handed in by Wednesday, October 25, 2017, 12:00

Practical Information (II)

- Exam
  - There will be a written exam at the end of the semester
  - 60% of the points of the homework sheets required to be eligible to write the exam
  - Date to be fixed
- Successful participating requires to pass the written exam
- Final grade
  - 2/3 of the points of the homework assignments
  - 1/3 written exam

Useful Reading Material

Books:

- G. Cowan, Statistical Data Analysis
- L. Lista, Statistical Methods for Data Analysis in Particle Physics
- Behnke, Kroeninger, Schott, Schoerner-Sadenius: Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods
- R. Barlow, Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences
- Bohm, Zech, Introduction to Statistics and Data Analysis for Physicist [available online]
- Blobel, Lohrmann: Statistische Methoden der Datenanalyse (in German), [free ebook]
- Lyons: Statistics for Nuclear and Particle Physicists (Cambridge University Press)
- F. James, Statistical Methods in Experimental physics

Further Material

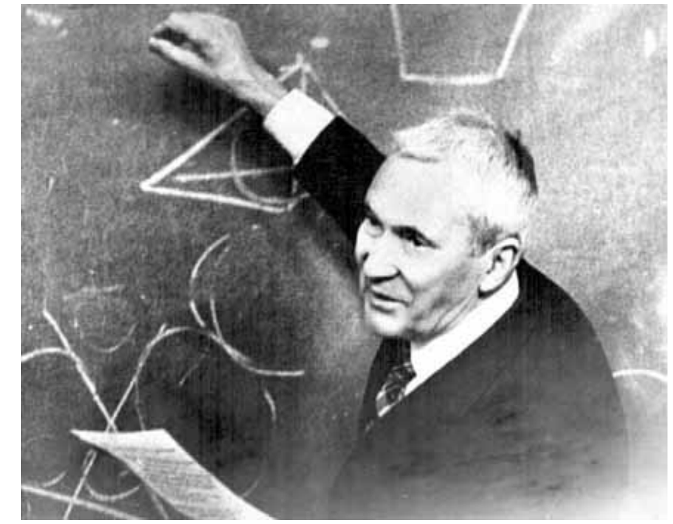
- Lot's of material from previous lectures by Oleg Brandt at Heidelberg University and others
  - Many thanks!
- Glen Cowan: http://www.pp.rhul.ac.uk/~cowan/stat_course.html
- Scott Oser: http://www.phas.ubc.ca/~oser/p509/
- Particle Data Group reviews on Probability and Statistics [link]

Sources of Uncertainty

- Underlying theory (quantum mechanics) is probabilistic
  - true randomness
- Limited knowledge about the measurement process
  - present even without quantum mechanics

We quantify uncertainty using **probability**

Mathematical Definition of Probability



Kolmogorov, 1933

Let A be an event. Then probability is a number obeying three conditions, the *Kolmogorov axioms*:

1. $P(A) \geq 0$
2. $P(S) = 1$, where S is the set of all A, the sample space
3. $P(A \cup B) = P(A) + P(B)$ for any A, B which are exclusive, i.e., $A \cap B = 0$

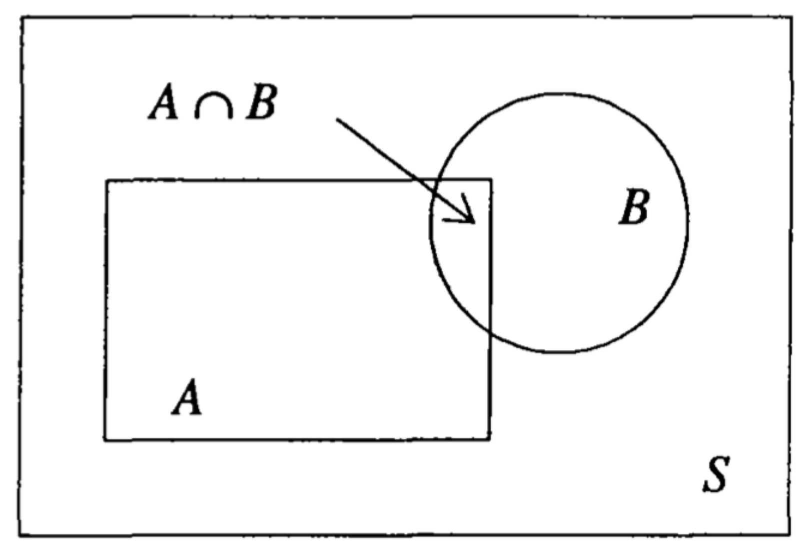
From these axioms further properties can be derived, e.g.:

$P(\bar{A}) = 1 - P(A)$

$P(\varnothing) = 0$

if $A \subset B$ then $P(A) \leq P(B)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$



But what does *P* mean?

Interpretations of Probability

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

- Frequentist definition
  - Let A, B, … be outcomes of an repeatable experiment:

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

- Bayesian definition (subjective probability)
  - A, B, … are hypotheses (statements that are true or false)

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

All three definitions are consistent with Kolmogorov's axioms

Criticisms of the Probability Interpretations

- Criticisms of the frequency interpretation
  - $n \to \infty$ can never be achieved in practice. When is n large enough?
  - We want to talk about the probability of events that are not repeatable
    - Example 1: P(it will rain tomorrow), but there is only one tomorrow
    - Example 2: P(Universe started with a Big Bang), but only one universe
  - P is not an intrinsic property of A, it depends on the how the ensemble of possible outcomes was constructed
    - Example: P(person I talk to is a physicist) depends on whether I am in a football stadium or at a scientific conference

- Criticisms of the subjective interpretation
  - “Subjective” estimates have no place in science
  - How to quantify the prior state of our knowledge upon which we base our probability estimate?

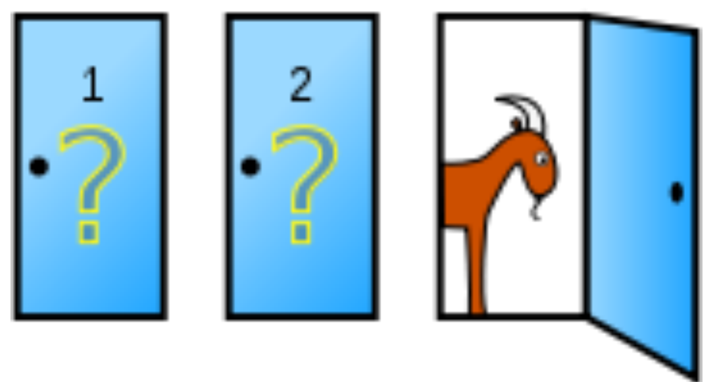
"Bayesians address the questions everyone is interested in by using assumptions that no one believes. Frequentist use impeccable logic to deal with an issue that is of no interest to anyone.” – Louis Lyons

Fun With Probabilities

https://en.wikipedia.org/wiki/Monty_Hall_problem

Monty Hall problem ("Ziegenproblem")

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?



Standard assumptions

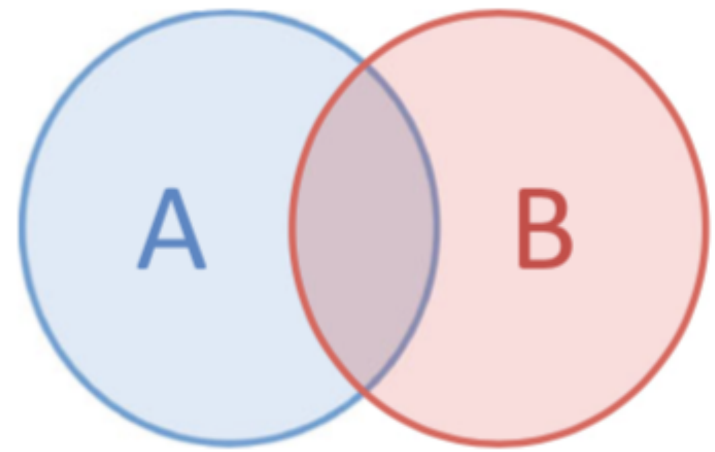
- The host must always open a door that was not picked by the contestant
- The host must always open a door to reveal a goat and never the car.
- The host must always offer the chance to switch between the originally chosen door and the remaining closed door.

Under these assumptions you should switch your choice!

Conditional Probability and Independent Events

For two events A and B, the conditional probability is defined as

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



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

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

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

Bayes' Theorem

Definition of conditional probability:

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

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

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

[doubtful whether the portrait actually shows Bayes]

First published (posthumously) by the Reverend Thomas Bayes (1702–1761)

First modern formulation by Pierre-Simon Laplace in 1812

Example of Using Bayes' Theorem: Test for a Rare Disease

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

$$P(\mathrm{D}) = 0.001$$
$$P(\mathrm{no\ D}) = 0.999$$

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

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

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

$$P(\mathrm{D}|+) = \frac{P(+|\mathrm{D})P(\mathrm{D})}{P(+|\mathrm{D})P(\mathrm{D}) + P(+|\mathrm{no\ D})P(\mathrm{no\ D})}$$
$$= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.03 \times 0.999} = 0.032$$

Probability for you to have the disease is 3.2%, i.e., you're probably ok.

Bayes' Theorem: Degree of Believe in a Theory Given a Certain Set of Data

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

- likelihood
- prior (before seeing the data, subjective)
- posterior probability, i.e., after seeing the data
- normalization

Bayesian Inference: Jeffreys' Prior

How to model complete ignorance about the value of a parameter θ?

- Uniform distribution in θ, exp θ, ln θ, 1/θ, …?
- Example: Lifetime τ of a particle, uniform distribution in τ or particle's width Γ = 1/τ ?

Jeffreys' prior (non-informative prior) for a model $L(\vec{x}|\vec{\theta})$ of the measurement:

$$\pi(\vec{\theta}) \propto \sqrt{I(\vec{\theta})} \qquad I(\vec{\theta}) = \det \left[\left\langle \frac{\partial \ln L(\vec{x}|\vec{\theta})}{\partial \theta_i} \frac{\partial \ln L(\vec{x}|\vec{\theta})}{\partial \theta_j} \right\rangle \right]$$

invariant under re-parameterization

determinant of the Fisher information matrix

expectation value evaluated by $\vec{x}$ integrating over all possible results

Examples:

PDF parameter	**Jeffreys' prior**
Poissonian mean μ	**p(μ) ∝ 1/√μ**
Gaussian mean μ	**p(μ) ∝ 1**

Jeffreys' Prior: Example

Gaussian distribution with mean parameter:

$$L(x \mid \mu) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \qquad (\sigma \text{ fixed})$$

Jeffreys' prior:

$$\pi(\mu) \propto \sqrt{I(\mu)} = \sqrt{\mathrm{E}\left[\left(\frac{d}{d\mu}\ln L(x \mid \mu)\right)^2\right]} = \sqrt{\mathrm{E}\left[\left(\frac{x-\mu}{\sigma^2}\right)^2\right]}$$

$$= \sqrt{\int_{-\infty}^{+\infty} L(x \mid \mu)\left(\frac{x-\mu}{\sigma^2}\right)^2 dx} = \sqrt{\frac{\sigma^2}{\sigma^4}} \propto 1.$$

independent of μ

Frequentist Inference

Typical example:
μ = true value, measurement process modeled by a Gaussian distribution:

$$\text{Measurement} \triangleq \text{drawing random number from } G(x; \mu, \sigma)$$

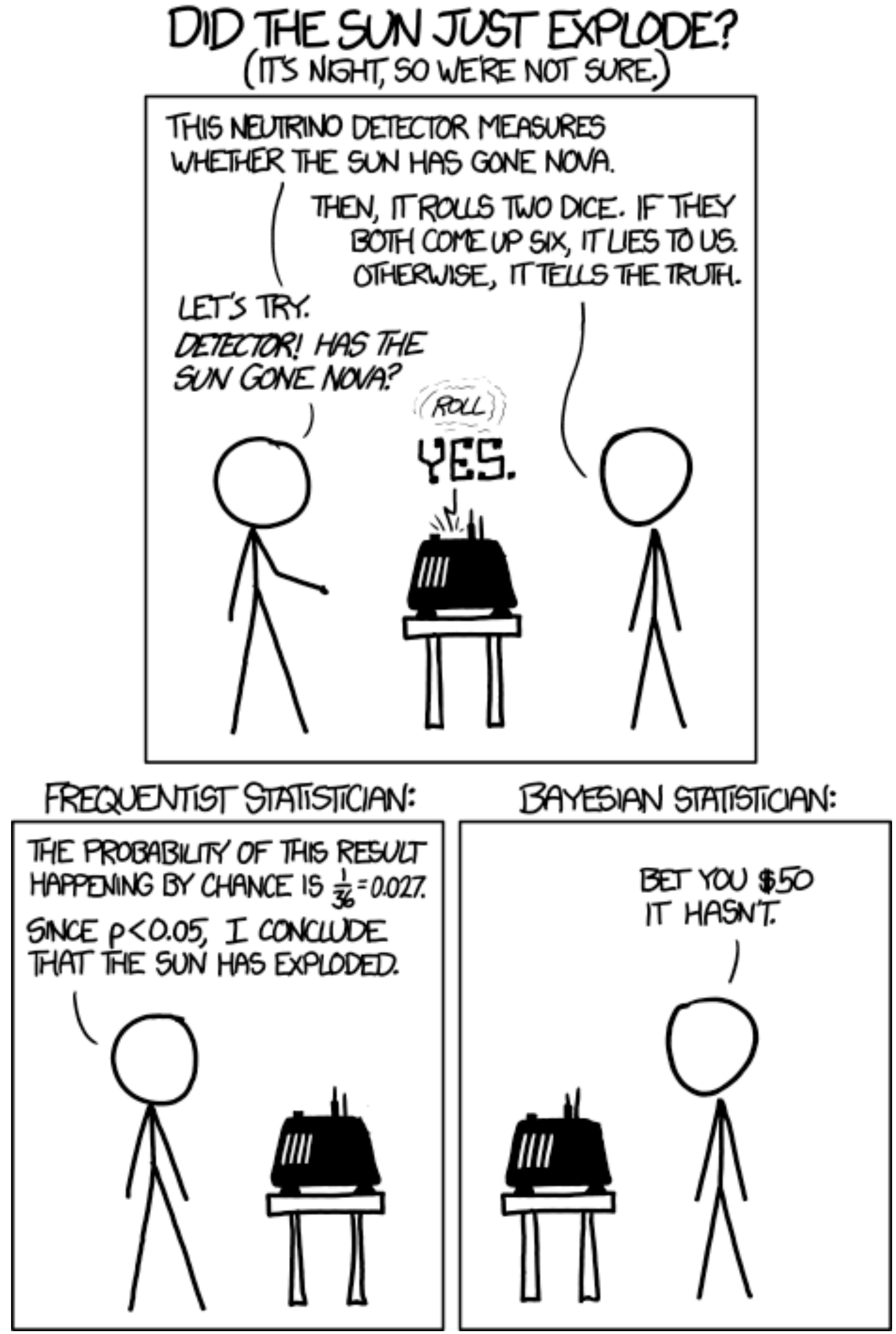
Measurement is reported as $x \pm \sigma$.

This is a statement about the interval $[x-\sigma, x+\sigma]$. For a large number of hypothetically repeated experiments the interval would contain the true value in 68% of the cases. In the frequentist approach, one cannot make a probabilistic statement about the true value (the true value is what it is).

In other words, the frequentist rather makes a statement about statements:
The statement "μ lies in $[x-\sigma, x+\sigma]$" has a probability of 68% of being true.

Both the frequentist and the Bayesian approach require a statistical model of the measurement process.

A Recurrent Theme: Frequentist vs. Bayesian Statistics



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
- Can be discrete or continuous

Probability density function (pdf) of a continuous variable:

probability density function

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

Normalization:

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

"x must be somewhere"

Histograms

Histogram:

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

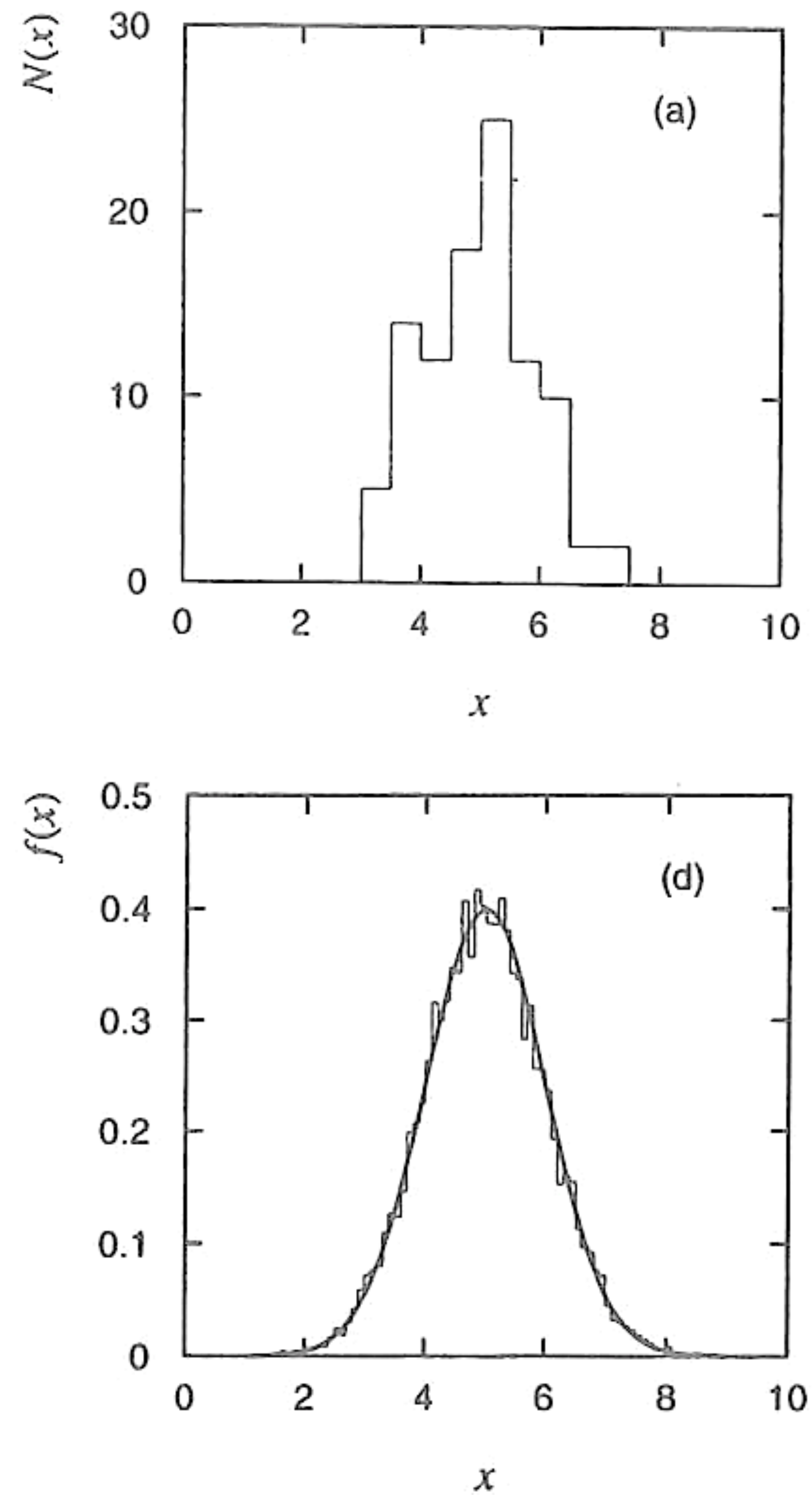
pdf = histogram for

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

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

$$n = \text{total number of entries}$$

$$\Delta x = \text{bin width}$$



Mean, Median, and Mode

Mean of a **data sample**:

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

"sample mean"

Mean of a pdf:

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

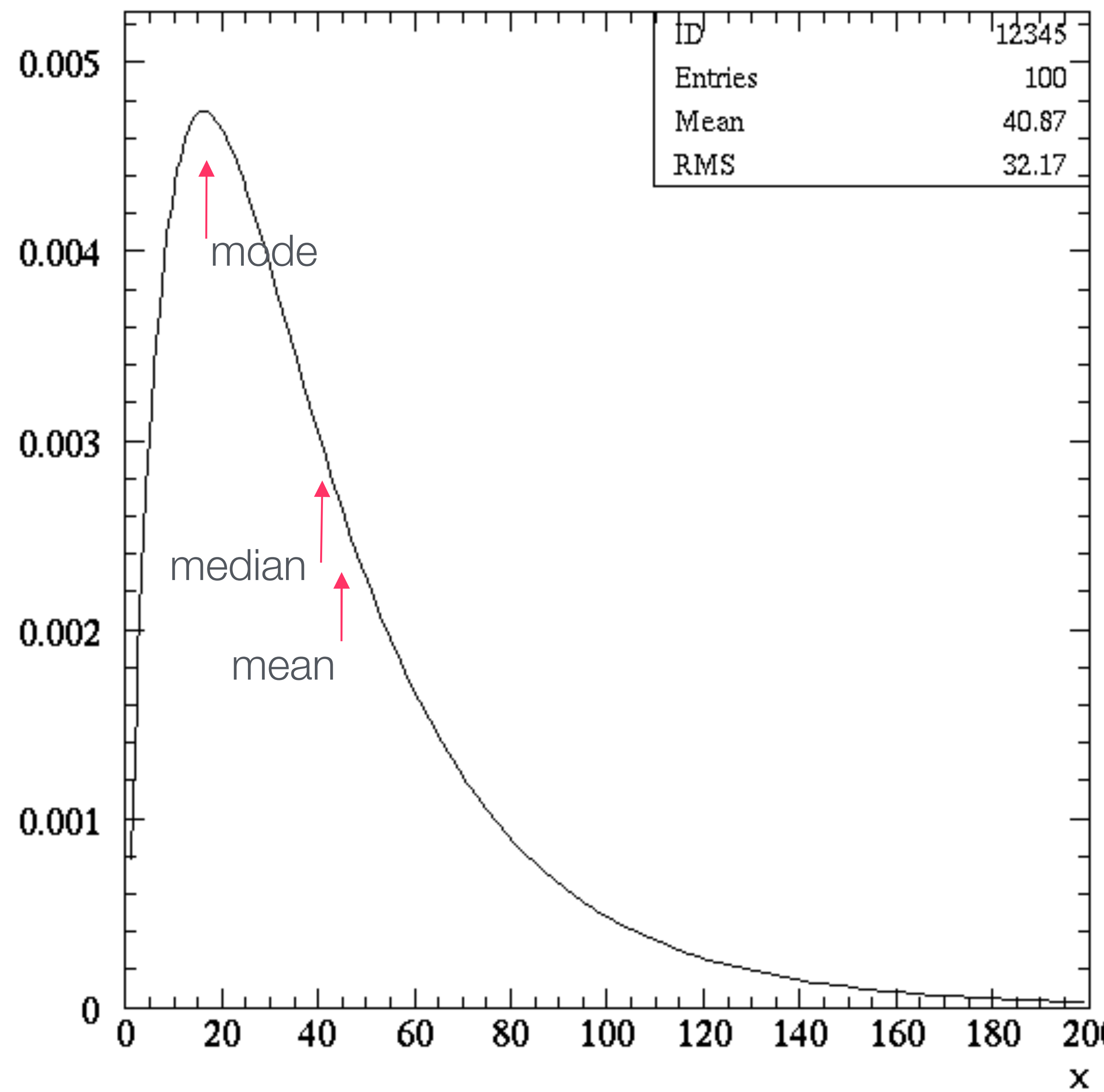
$\equiv$ expectation value $E[x]$

Median:

point with 50% probability above and 50% probability below

Mode:

the most likely value



Variance and Standard Deviation

expectation value

Variance of a distribution: $$V(x) = \int \mathrm{d}x\, P(x)(x-\mu)^2 = E[(x-\mu)^2]$$

$$V(x) = \int \mathrm{d}x\, P(x) x^2 - 2\mu \int \mathrm{d}x\, P(x) x + \mu^2 \int \mathrm{d}x\, P(x) = \langle x^2 \rangle - \mu^2 = \langle x^2 \rangle - \langle x \rangle^2$$

Variance of a **data sample**: $$V(x) = \frac{1}{N} \sum_i (x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2$$

This formula underestimates the variance of underlying distribution as it used the mean calculated from data!

Use this if you have to estimate the mean from data (unbiased estimator):

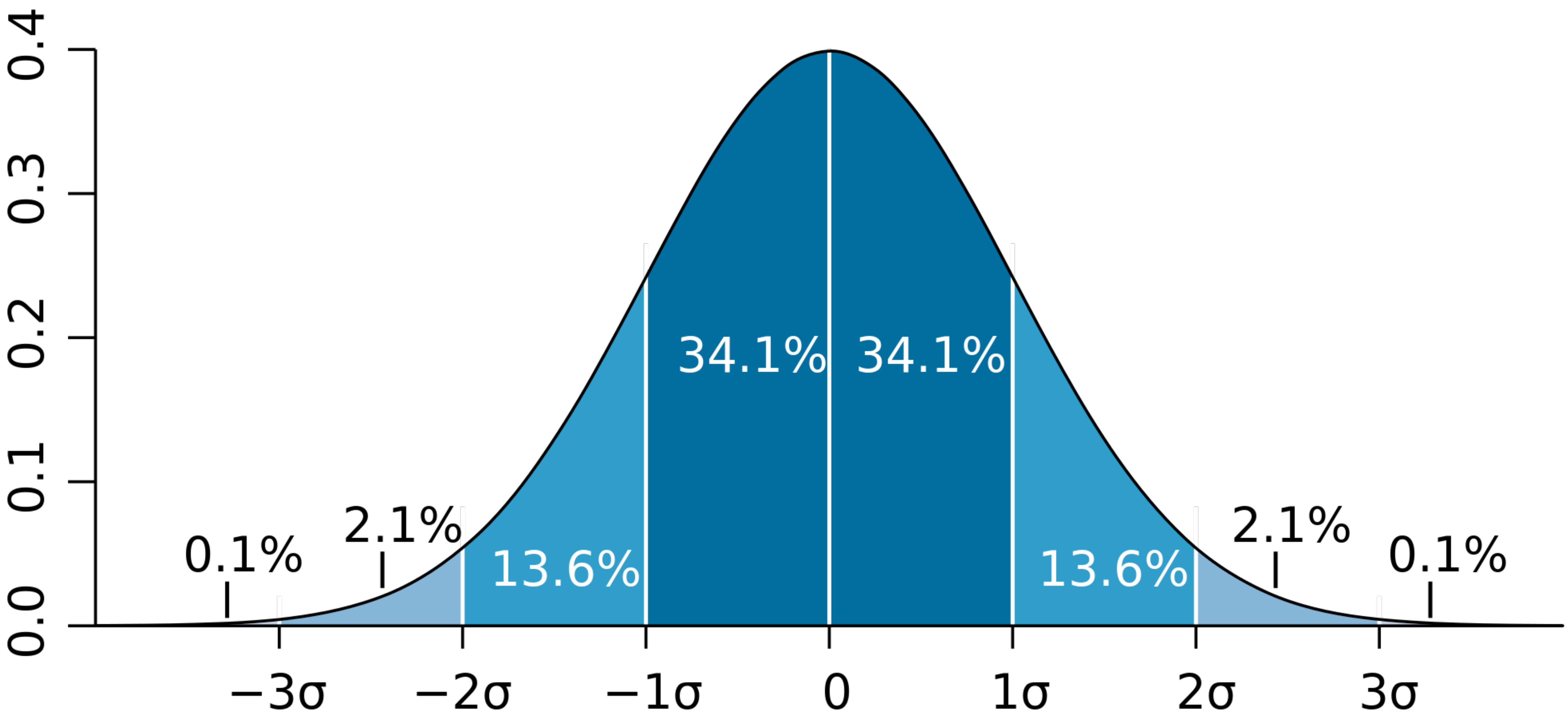
$$\hat{V}(x) = \frac{1}{N-1} \sum_i (x_i - \bar{x})^2$$

Use this if you know the true mean:

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

Standard deviation: $\sigma = \sqrt{V(x)}$

Deviation in Units of σ for a Gaussian



- about 68% of events between –σ and +σ around mean
- about 95% of events between –2σ and +2σ around mean

Multivariate Distributions

Outcome of experiment
characterized by a vector (x_1, …, x_n)

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

joint pdf

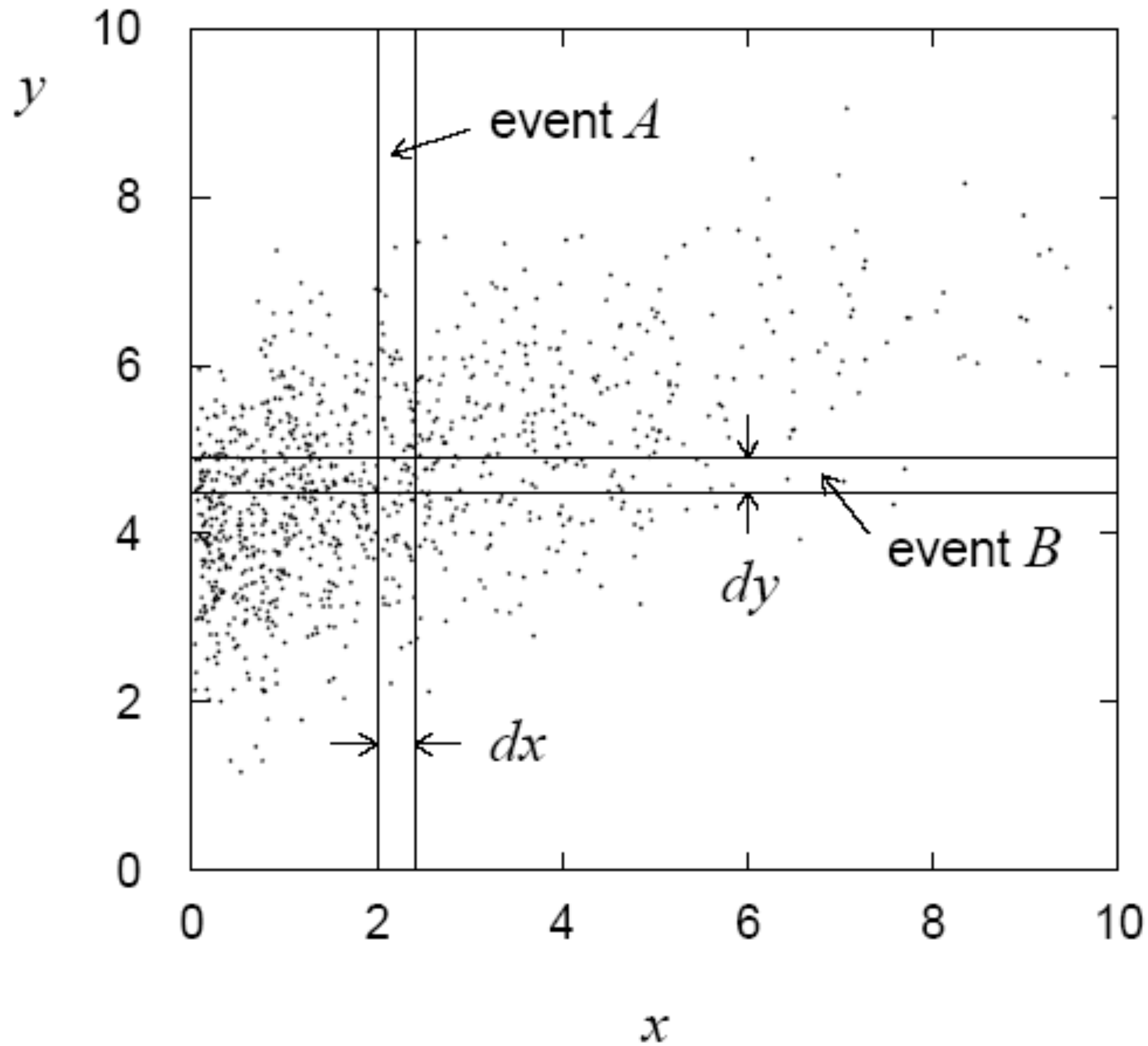
Normalization:

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

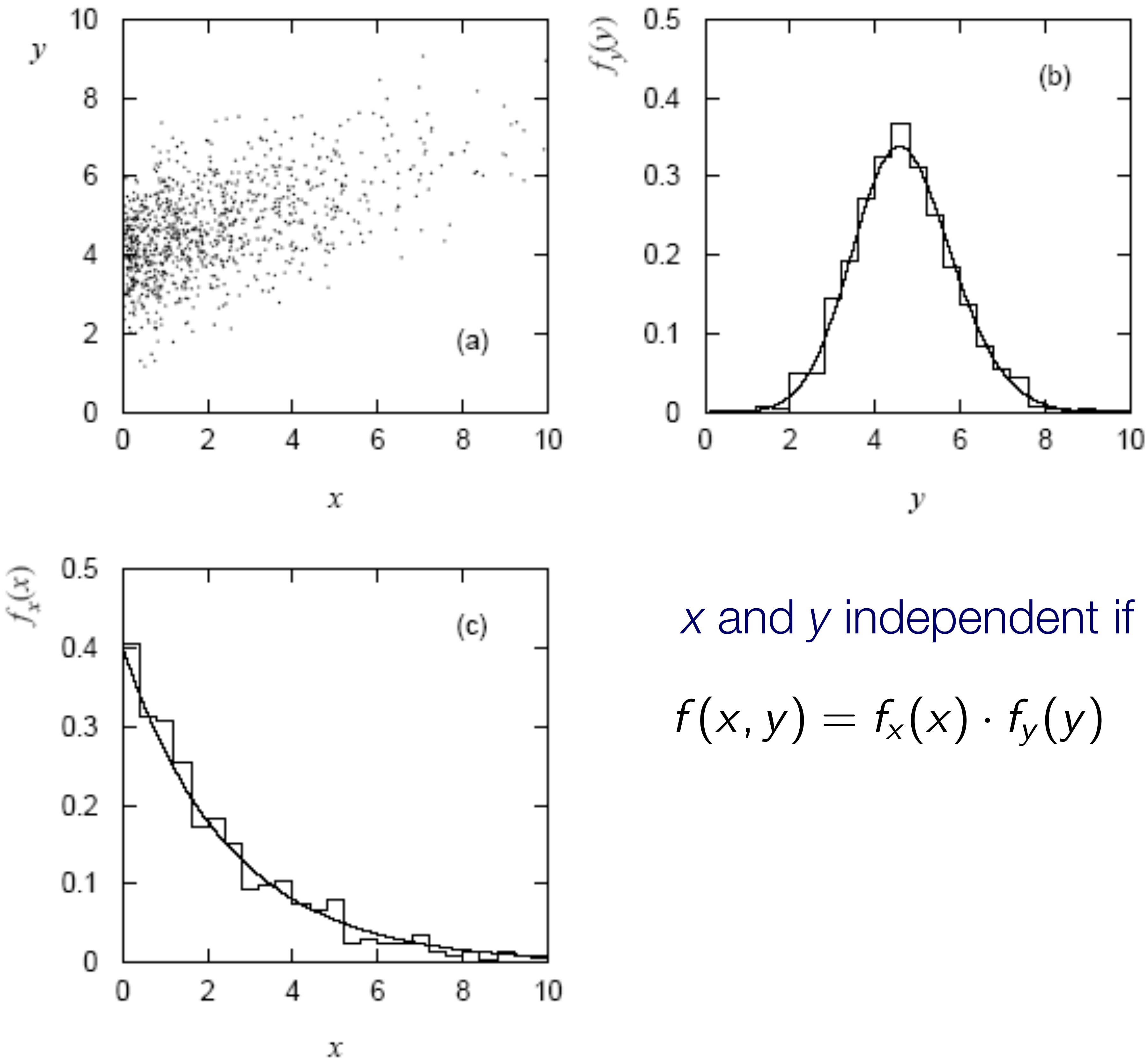
Sometimes we want only the pdf of one component:

$$f_x(x) = \int f(x, y)\,\mathrm{d}y$$

"marginal pdf"
= projection of joint pdf
onto individual axes



Marginal pdf = Projections



x and *y* independent if

$$f(x, y) = f_x(x) \cdot f_y(y)$$

Covariance and Correlation

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

Correlation coefficient (dimensionless): $$\rho_{xy} = \frac{\mathrm{cov[x, y]}}{\sigma_x \sigma_y}$$

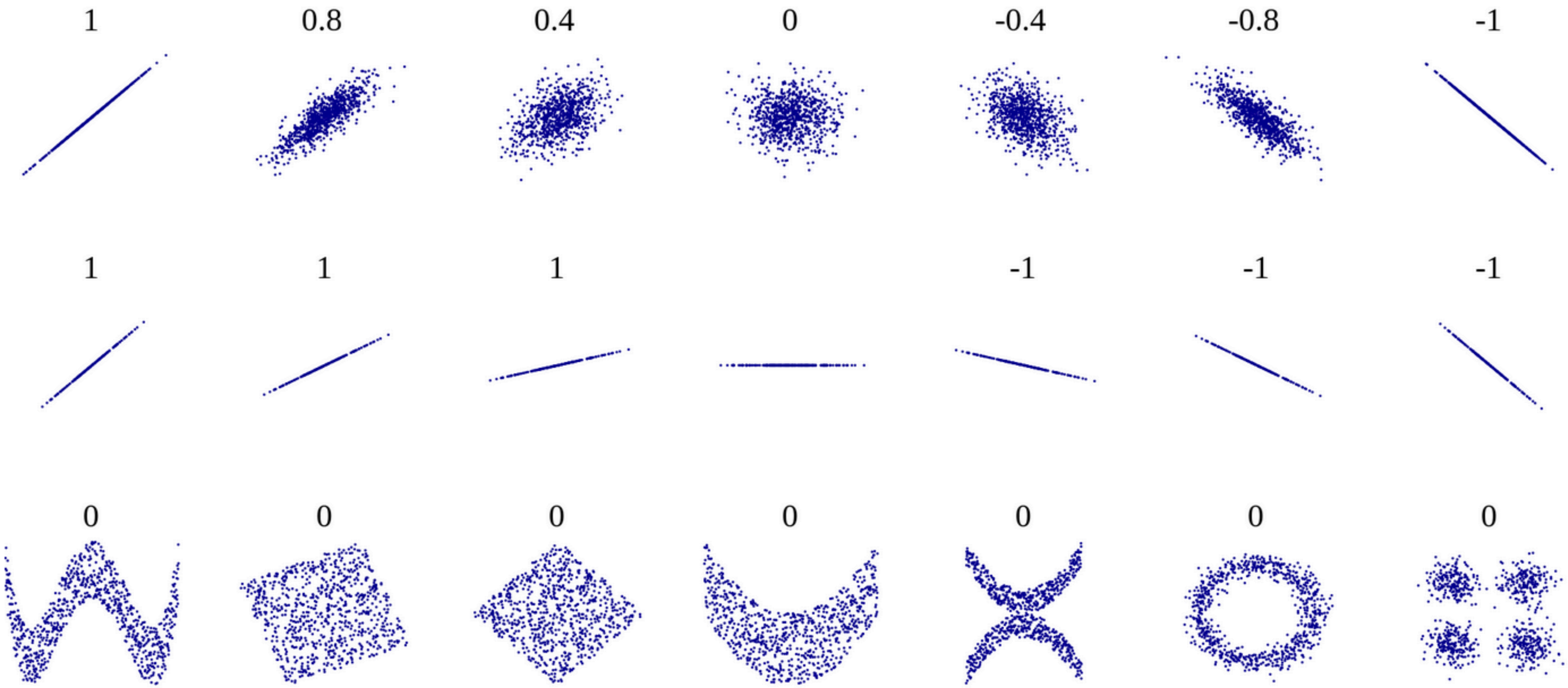
x, *y* independent:

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

$\rightarrow$ $\mathrm{cov}[x, y] = 0$

N.B. converse not always true

Correlation Coefficient



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

Linear Combinations of Random Variables

Consider two random variables with known covariance cov(x, y):

$$\langle x + y \rangle = \langle x \rangle + \langle y \rangle$$
$$\langle ax \rangle = a \langle x \rangle$$
$$V[ax] = a^2 V[x]$$
$$V[x + y] = V[x] + V[y] + 2\mathrm{cov}(x, y)$$

Check:

$$\begin{aligned} V[x+y] &= E[(x + y - \mu_x - \mu_y)^2] = E[(x - \mu_x + y - \mu_y)^2] \\ &= E[(x-\mu_x)^2 + (y-\mu_y)^2 + 2(x-\mu_x)(y-\mu_y)] \\ &= E[(x-\mu_x)^2] + E[(y-\mu_y)^2] + 2E[(x-\mu_x)(y-\mu_y)] \\ &= V[x] + V[y] + 2\mathrm{cov}(x, y) \end{aligned}$$

Higher Moments

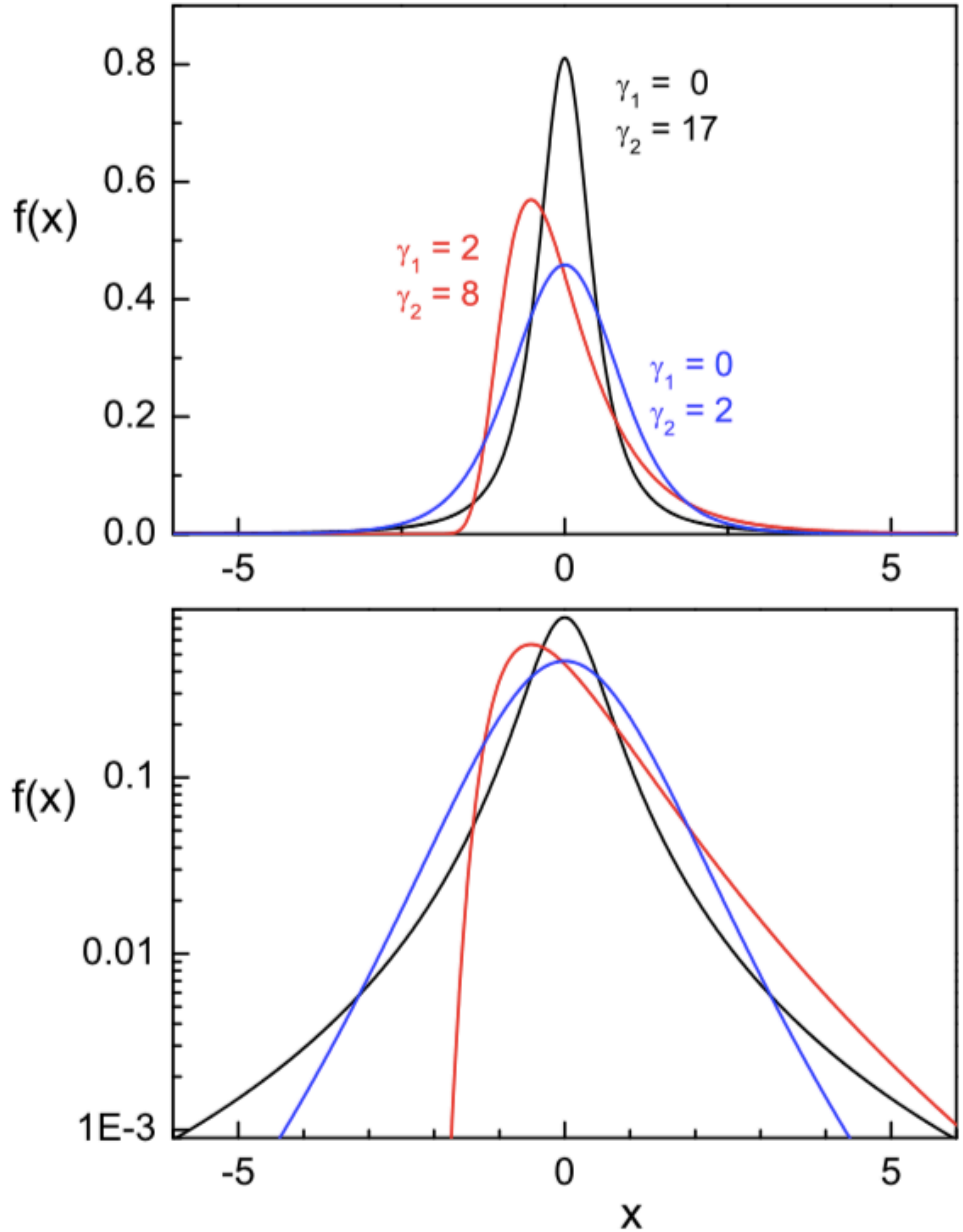
Skewness: $$\gamma_1 = \left\langle \left(\frac{x - \langle x \rangle}{\sigma} \right)^3 \right\rangle$$

symmetric distribution have skewness equal to zero

Curtosis: $$\beta_2 = \left\langle \left(\frac{x - \langle x \rangle}{\sigma} \right)^4 \right\rangle$$

$$\gamma_2 = \beta_2 - 3$$

defined such that $\gamma_2 = 0$ for the normal distribution



Correlation ≠ Causation (1)

Examples of illogically inferring causation from correlation

https://en.wikipedia.org/wiki/Correlation_does_not_imply_causation

Example 1 ("reverse causality"):

- The faster windmills are observed to rotate, the more wind is observed to be.
- Therefore wind is caused by the rotation of windmills.

Example 2 ("third factor C causes both A and B"):

- Sleeping with one's shoes on is strongly correlated with waking up with a headache.
- Therefore, sleeping with one's shoes on causes headache.

What Makes Nobel Prize Winners?

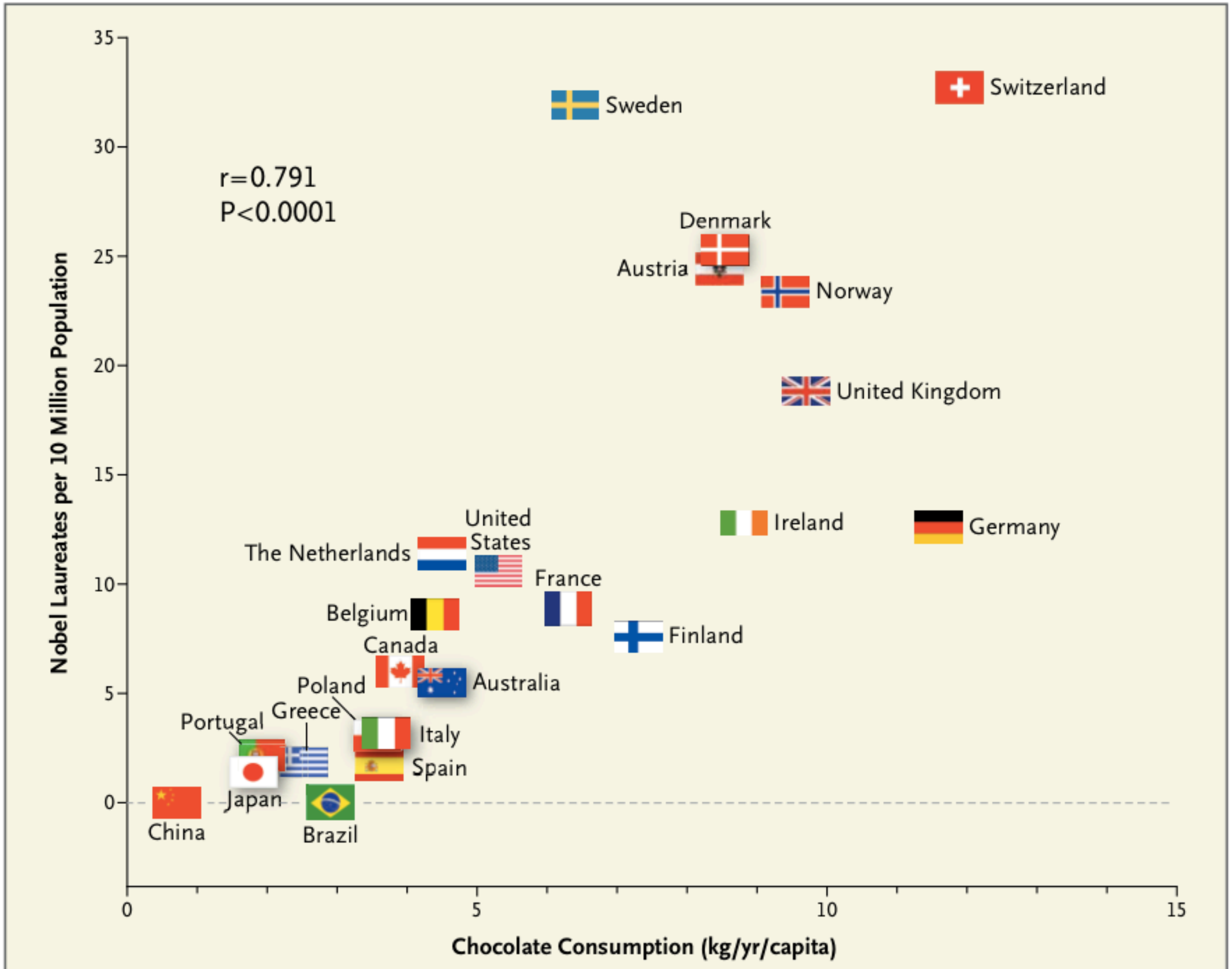


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Correlation coefficient: 0.791

Improved cognitive function associated with a regular intake of flavonoids???

Probably not …

F. Messerli, 2012, New England Journal of Medicine, 2012

Correlation ≠ Causation (2)