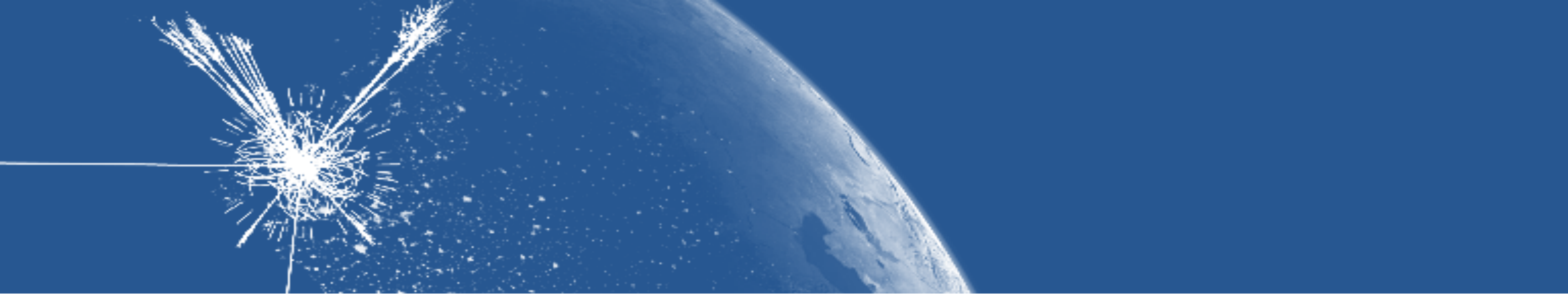
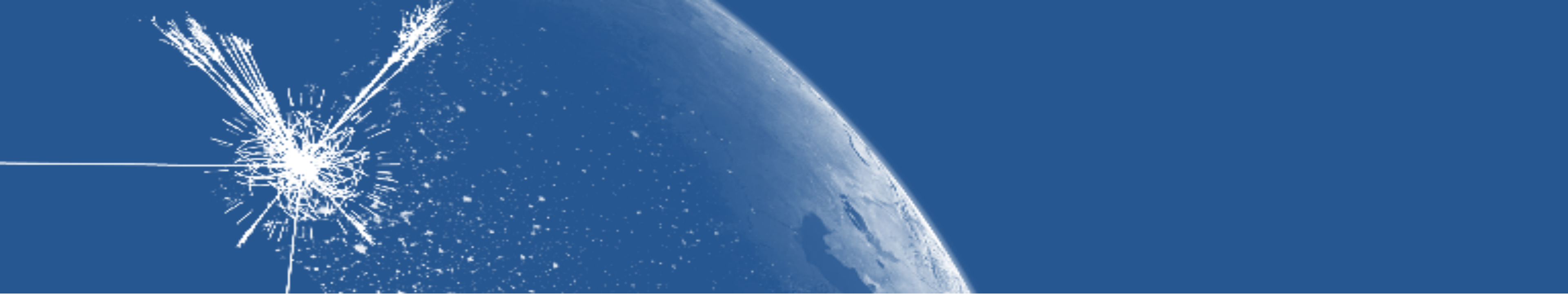


Introduction to probability and statistics (1)

Andreas Hoecker (CERN)

CERN Summer Student Lecture, 18–21 July 2016



Outline (4 lectures)

1st lecture:
- Introduction
- Probability

2nd lecture:
- Probability axioms and hypothesis testing
- Parameter estimation
- Confidence levels

3rd lecture:
- Maximum likelihood fits
- Monte Carlo methods
- Data unfolding

4th lecture:
- Multivariate techniques and machine learning

Acknowledgements and some further reading

- I am particularly grateful to Helge Voss, whose comprehensive statistics lectures serve as model for the present lectures. His latest lectures on multivariate analysis and neural networks at the School of Statistics, 2016: https://indico.in2p3.fr/event/12667/other-view?view=standard

- Glen Cowan's book "Statistical data analysis" represents a great introduction: http://www.pp.rhul.ac.uk/~cowan/sda

- PDG reviews by G. Cowan on probability (http://pdg.lbl.gov/2013/reviews/rpp2013-rev-probability.pdf) and statistics (http://pdg.lbl.gov/2013/reviews/rpp2013-rev-statistics.pdf) :

- Luca Lista's lectures "Practical Statistics for Particle Physicists", at the CERN-JINR school 2016: https://indico.cern.ch/event/467465/other-view?daysPerRow=5&view=nicecompact

- Kyle Cranmer's article "Practical Statistics for the LHC", arXiv:1503.07622

- Harrisson Prospers 2015 CERN Academic Training lecture "Practical Statistics for LHC Physicists": https://indico.cern.ch/event/358542/

- Machine learning introduction: T. Hastie, R. Tibshirani, J. Friedman, "The Elements of Statistical Learning", Springer 2001; C.M. Bishop, "Pattern Recognition and Machine Learning", Springer 2006

- Very incomplete list. Some more references are given throughout the lecture.

Why we need probability in the particle world

Since Laplace's times (1749–1827) the universe's fate was deterministic and, in spite of technical difficulties, was considered predictible if the complete equation of state were known.

Challenged by Heisenberg's uncertainty principle (1927), Albert Einstein proclaimed "*Gott würfelt nicht*" ("God does not play dice"), but *hidden variables* to bring back determinism through the back door into the quantum world were never found.

In quantum mechanics, particles are represented by wave functions. The size of the wave function gives the probability that the particle will be found in a given position. The rate, at which the wave function varies from point to point, gives the speed of the particle.

Quantum phenomena like particle reactions occur according to certain probabilities. Quantum field theory allows us to compute cross-sections of particle production in scattering processes, and decays of particles. It cannot, however, predict how a single event will come out. We use probabilistic "Monte Carlo" techniques to simulate event-by-event realisations of quantum probabilities.

Pierre-Simon de Laplace

Albert Einstein

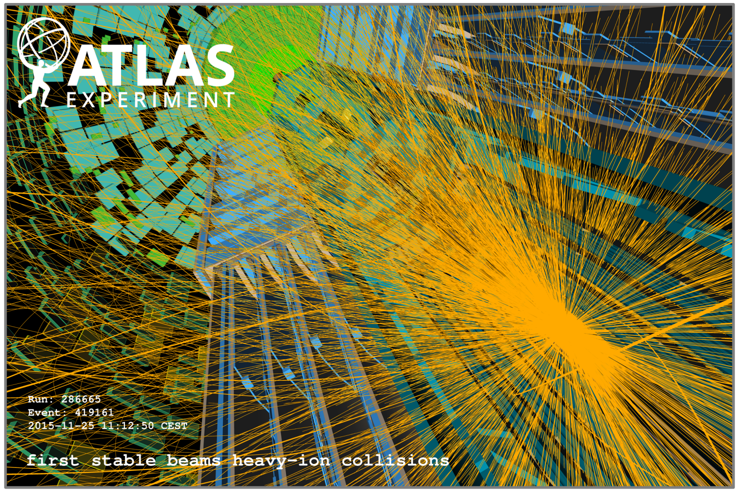
Werner Heisenberg

Statistics of large systems

Statistical physics uses probability theory and statistics to make statements about the approximate physics of large populations of stochastic nature, neglecting individuals.

Heavy-ion collisions at the LHC are modelled using hydrodynamics (strongly interacting medium behaves like perfect fluid)

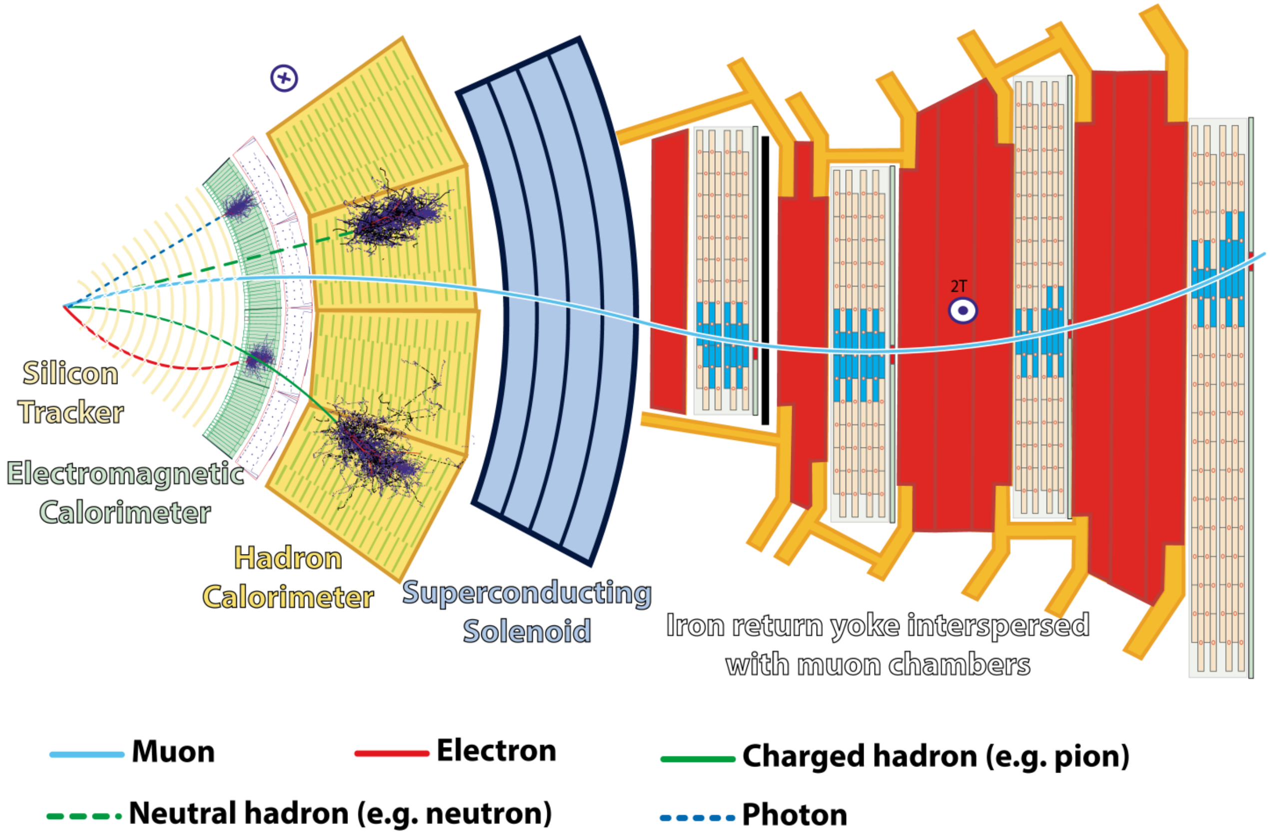
Statistical mechanics provides a framework for relating the microscopic properties of individual atoms and molecules to the macroscopic properties of materials that can be observed in everyday life, therefore explaining thermodynamics as a natural result of statistics, classical mechanics, and quantum mechanics at the microscopic level.



Display of ATLAS Run-2 Heavy-Ion collision

Probability and statistics are fundamental ingredients & tools in all modern sciences

Statistics in measurement processes

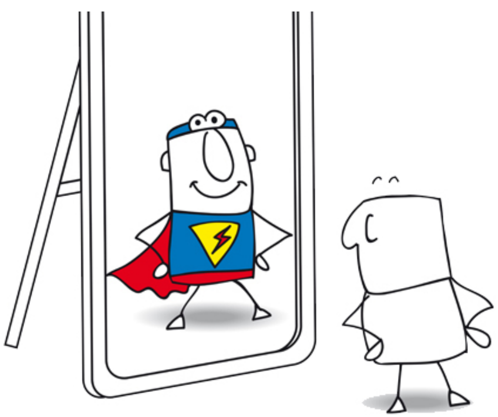


In addition to the intrinsic probabilistic character of particle reactions, the measurement process through the interaction of particles with active detector materials contributes statistical degrees of freedom leading to measurement errors and to genuine systematic effects (eg, detector misalignment), that need to be considered in the statistical analysis

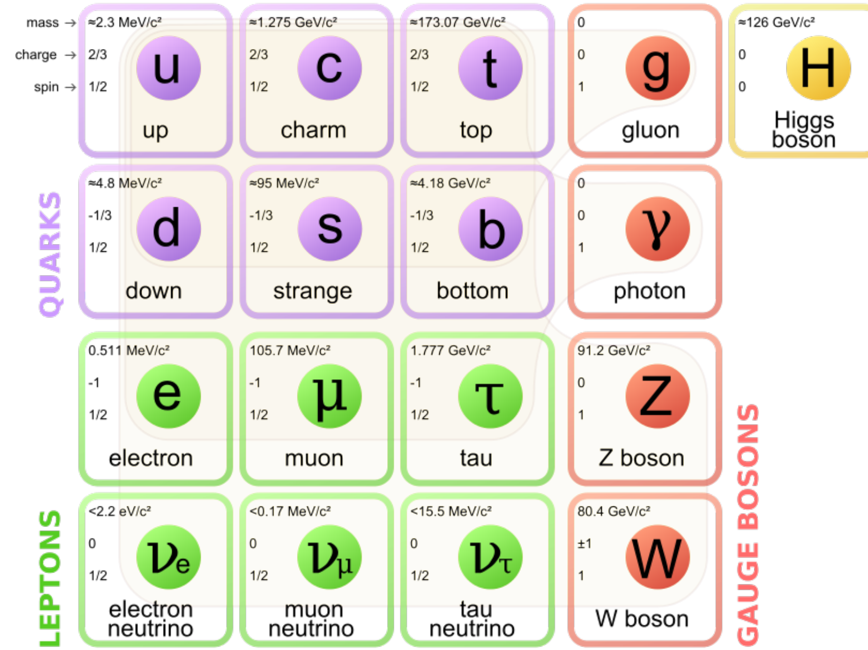
Measurements and hypothesis testing

From a measured data sample, we want to determine parameters of a known model (eg, the top-quark mass in the Standard Model), we want to discover and measure missing pieces of the model (eg, the Higgs boson, neutrino masses), *and* we want to watch out for the unknown (test the data versus the predictions of a known model), or exclude parameters of suggested new physics models

Supersymmetry ?



The Standard Model



Ingredients

Due to the intrinsic randomness of the data, probability theory is required to extract the information that addresses our questions. Statistics is used for the actual data analysis.

The interpretation of probability depends on the statement we want to make.

- For repeatable experiments, probability can be a measure of how frequently a statement is true
- In a more subjective approach, one could expresses a degree of belief of a statement

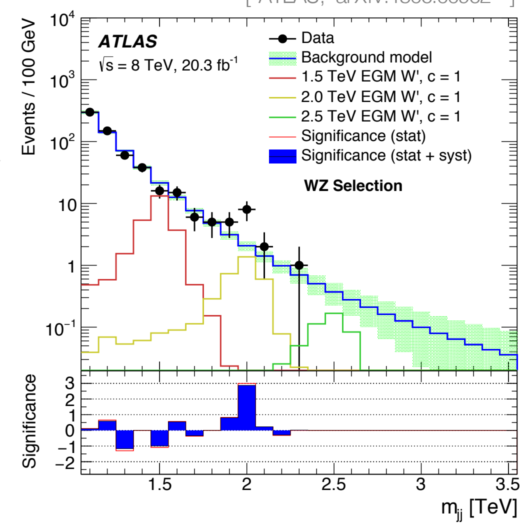
Repeatable experiments are, eg:

- Playing a dice and finding 6
- Fluctuations in a background distribution and finding a peak of some size or more (note that the contrary: the probability that *the* peak is due to a background fluctuation, is *not* repeatable)

Non-repeatable statements are, eg:

- The probability that dark matter is made of axions
- The probability that the new 125 GeV boson is the Higgs boson

[ATLAS, arXiv:1506.00962]



Statistical distributions

Measurement results typically follow some “distribution”, ie, the data do not appear at fixed values, but are “spread out” in a characteristic way

Which type of distribution it follows depends on the particular case

- It is important to know the occurring distributions to be able to pick the correct one when interpreting the data (example: *Poisson* vs. *Compound Poisson*)
- …and it is important to know their characteristics to extract the correct information

Note: in statistical context, instead of “data” that follow a distribution, one often (typically) speaks of a “random variable”

Probability distribution / density of a random variable

Random variable $\boldsymbol{k}$ (discrete) or $\boldsymbol{x}$ (continuous): quantity or point in sample space

Discrete variable

$$P(k_i) = p_i$$

Continuous variable

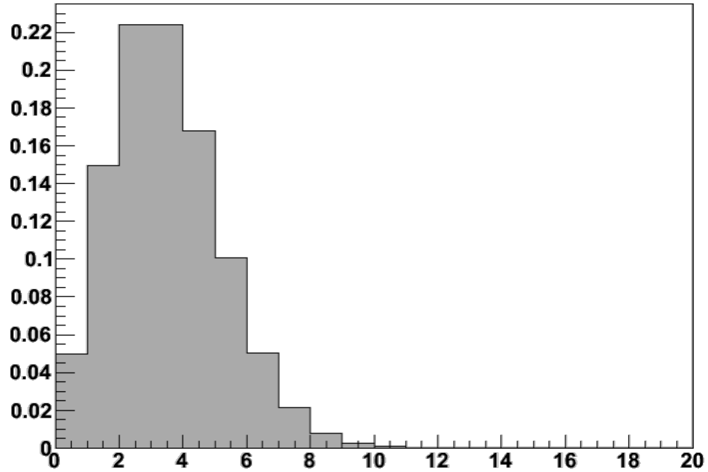
$$P(x \in [x, x + dx]) = p_x(x)dx$$

Normalisation (your parameter/event space covers all possibilities - *unitarity*)

$$\sum_i P(k_i) = 1$$

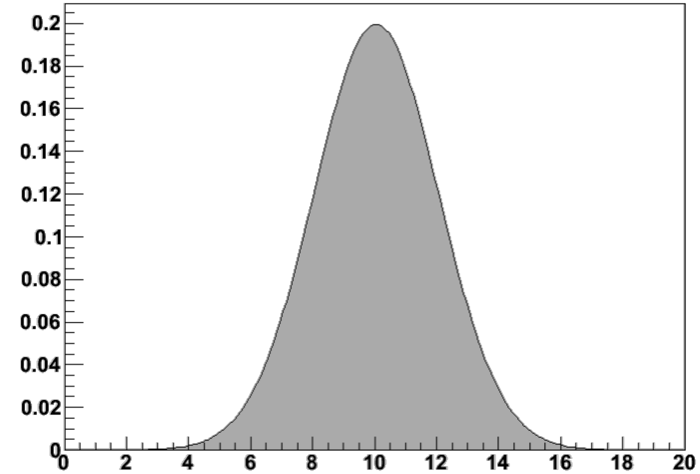
$$\int_{-\infty}^{\infty} p_x(x)dx = 1$$

Poisson distribution



$k_i = N_{\text{observed}}$

Gaussian (or *Normal)* distribution



$x = m_{\text{observed}}$

Cumulative distribution

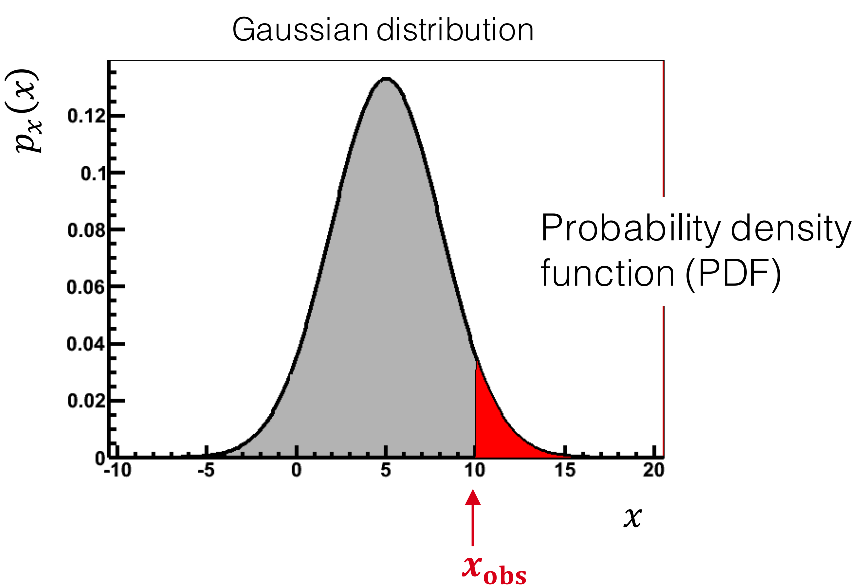
$\boldsymbol{p_x(x)}$: probability density distribution for some "measurement" $\boldsymbol{x}$ under the assumption of some model and its parameters

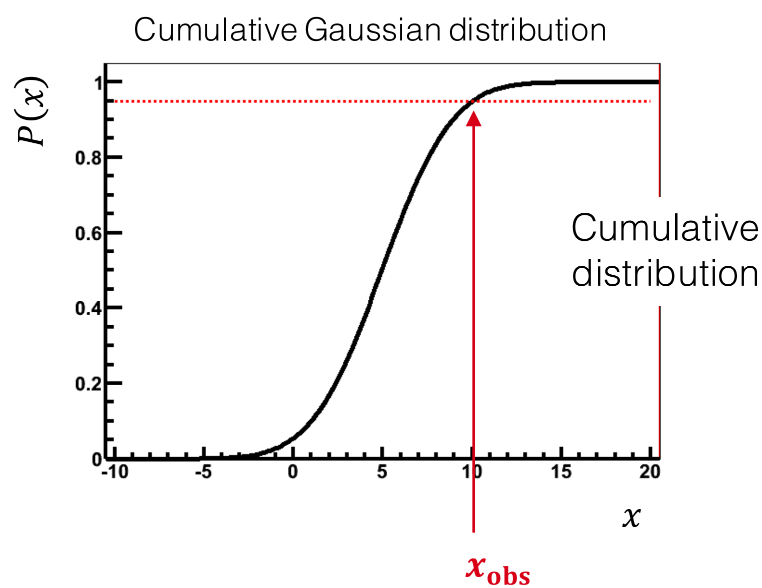
The cumulative distribution $\boldsymbol{P(x)}$ is *the probability to observe a random value* $\boldsymbol{x}$ *smaller than the one observed*, $\boldsymbol{x_{\mathrm{obs}}}$

→ Examples for cumulative distributions: χ^2, p-values, confidence limits (will come back to this)

$$p_x(x) = dP(x)/dx$$

$$\int_{-\infty}^{x} p_x(x')dx' \equiv P(x)$$





Selected probability (density) distributions

Imagine a monkey discovered a huge bag of alphabet noodles. She blindly draws noodles out of the bag and places them in a row before her. The text reads:
“TO BE OR NOT TO BE”

The probability for this to happen is about 10^{-22}

Infinite monkey theorem: provided enough time, the monkey will type Shakespeare's Hamlet

Bernoulli distribution

Experiment with two possible discrete outcomes: ***k* = 1 / *k* = 0** (or **yes / no** or **head / tail**, etc)

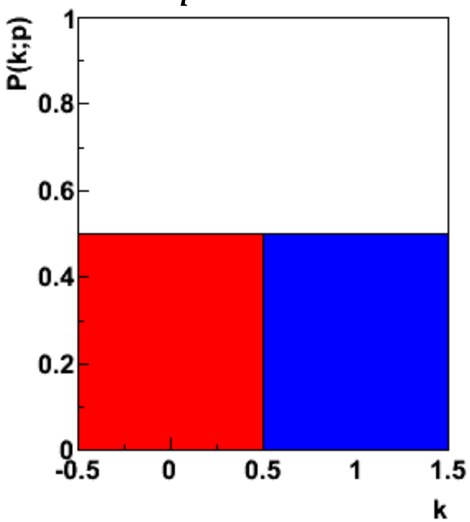
What is the probability of one or the other ?

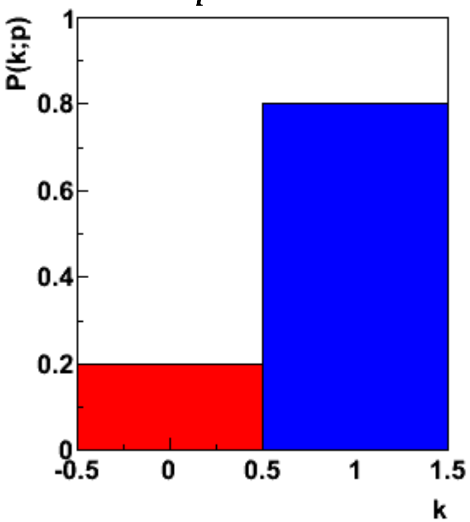
$$P(\text{head}) = p \ (\text{where } 0 \leq p \leq 1), \ \ P(\text{tail}) = 1 - P(\text{head}) = 1 - p$$

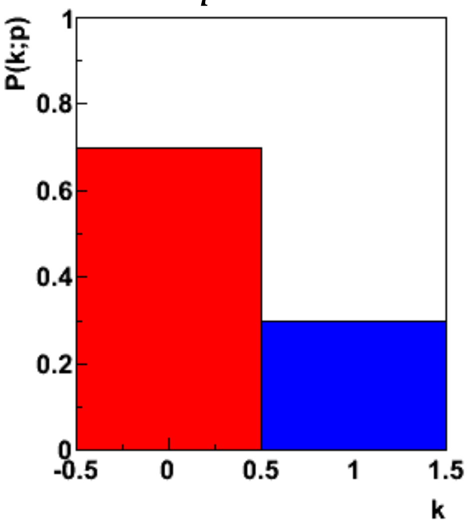
$$\Longrightarrow \ \ P(k;p) = p^k \, (1-p)^{1-k} \ \ \text{for} \ \ k \in \{0,1\}$$

$p = {}^1/_2$

$p = 0.5$ $\qquad$ $p = 0.8$ $\qquad$ $p = 0.3$







Binomial distribution (very important!)

Now let's get more complex: throw $\boldsymbol{N}$ coins (or similar binary choices)

How often (likely) is $\boldsymbol{k}\times\mathbf{head}$ and $(\boldsymbol{N}-\boldsymbol{k})\times\mathbf{tail}$?

- Each coin: $P(\text{head}) = p, P(\text{tail}) = 1 - p$
- Pick $\boldsymbol{k}$ particular coins → the probability of all having $\mathbf{head}$ is:

$$P(k\times\text{head}) = P(\text{head})\cdot P(\text{head})\cdot\dots\cdot P(\text{head}) = P(\text{head})^k = p^k$$

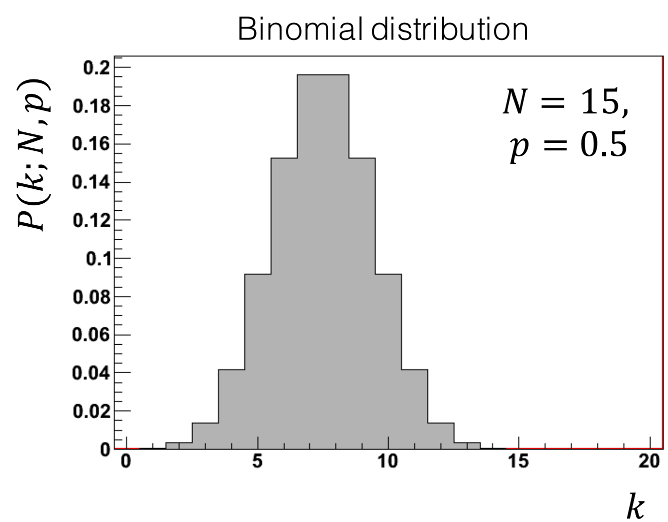
- Multiply this by the probability that all remaining $\boldsymbol{N\text{–}k}$ coins land on $\mathbf{tail}$:

$$P(\text{head})^k \cdot P(\text{tail})^{N-k} = p^k\,(1-p)^{N-k}$$

- This was for a particular choice of $\boldsymbol{k}$ coins
- Now include all $\binom{N}{k}$ permutations for *any* $\boldsymbol{k}$ coins

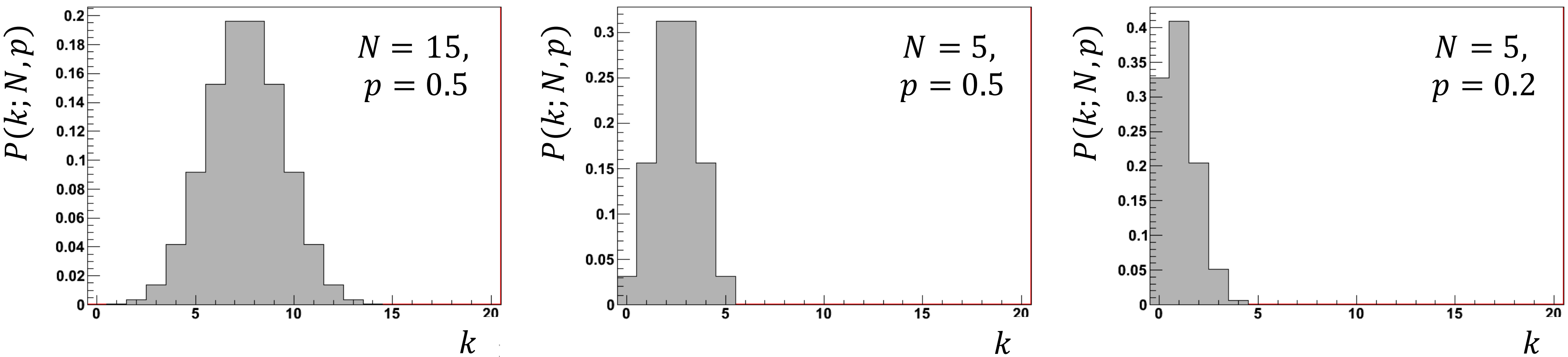
$$P(k;N,p) = p^k\,(1-p)^{N-k}\binom{N}{k}$$

where $\binom{N}{k} = \frac{k!}{k!(N-k!)}$ is the binomial coefficient



Binomial distribution (continued)

Example binomial distributions:



- Expectation value: sum over all possible outcomes and *average*

$$\mathrm{E}[k] = \sum_k kP(k) = Np$$

- *Variance:* (see next slide for definition)

$$V[k] = Np(1-p)$$

Characteristic quantities of distributions

Quantity	Discrete variable	Continuous variable
Expectation (mean) value $\boldsymbol{E}$	$E[k] = \langle k \rangle = \sum_k kP(k)$	$E[x] = \langle x \rangle = \int x \cdot p_x(x)\, dx$
Variance (spread) $\boldsymbol{V} = \sigma^2$	$E[(k - \langle k \rangle)^2] = E[k^2] - (E[k])^2$	same with $k \to x$
Higher moments: skew	$E[(k - \langle k \rangle)^3]$	same with $k \to x$

Note that "expectation value" and "variance" are properties of the full data population. Unbiased estimates can be derived from $\boldsymbol{N}$ samples extracted from the population:

Sample variance (unbiased)	$\frac{1}{N-1}\sum_{i=1}^{N}(k_i - \langle k \rangle)^2$	same with $k \to x$

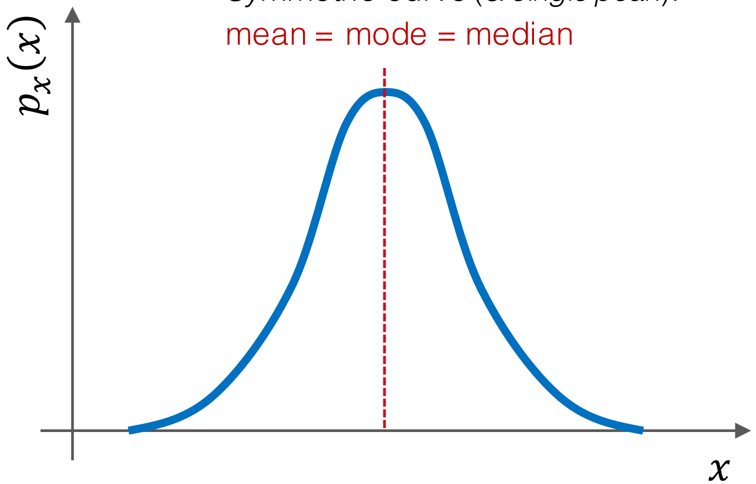
Characteristic quantities of distributions (continued)

Mean, Mode, Median:

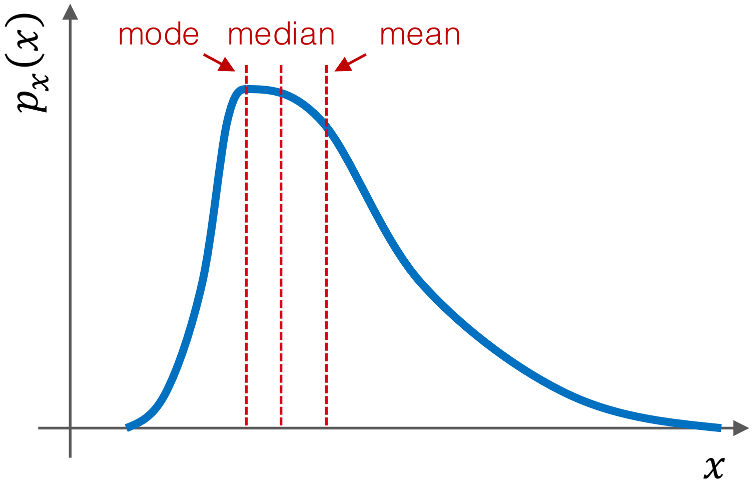
- **Mean:** $\langle x \rangle$ — defined before
- **Mode**: most probable value $x_{\mathbf{mode}}$: $p_x(x_{\mathbf{mode}}) \geq p_x(x), \; \forall x$
- **Median**: *2-quantile*: 50% of x values are larger than $x_{\mathbf{median}}$, 50% are smaller

  Can generalise *k-quantile* : points at regular intervals of the cumulative distribution.
  Boundaries of binning chosen such that each bin contains the 1/*k*-th of total integral of distribution.

Symmetric curve (& *single peak*):



Skewed curve:



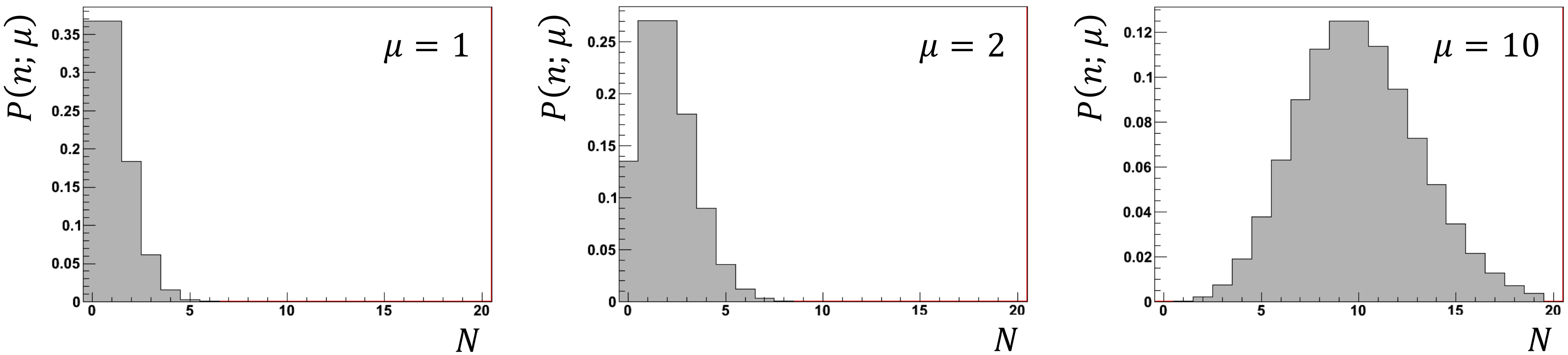
Poisson distribution

Recall: individual events each with two possible outcomes → Binomial distribution

How about: number of counts in radioactive decay experiment during given time interval $\boldsymbol{\Delta t}$?

- Events happen "randomly" but there is no such 2nd outcome. $\boldsymbol{\Delta t}$ is continuous, no discrete trials
- $\boldsymbol{\mu}$: average number of counts in $\boldsymbol{\Delta t}$. What is the probability of observing $\boldsymbol{N}$ counts?
- Limit of Binomial distribution for $N \to \infty$ & $p \to 0$ so that $Np \to \mu$.

→ $\boxed{\text{Poisson distribution:} \quad P(N;\mu) = \frac{\mu^N}{N!} e^{-\mu}}$



Expectation value: $E[N] = \sum_N N \cdot P(N) = \mu$, Variance: $V[N] = \mu$

Poisson is good approximation for Binomial distribution for $N \gg \mu \; (= Np)$

Gaussian (also: "Normal") distribution

In limit of large $\boldsymbol{\mu}$ a Poisson distribution approaches a symmetric Gaussian distribution

- This is the case not only for the Poisson distributions, but for almost any sufficiently large sum of samples with different sub-properties (mean & variance) → **Central Limit Theorem** (will discuss later)
- Gaussian distribution is of utter use, and luckily has simple properties

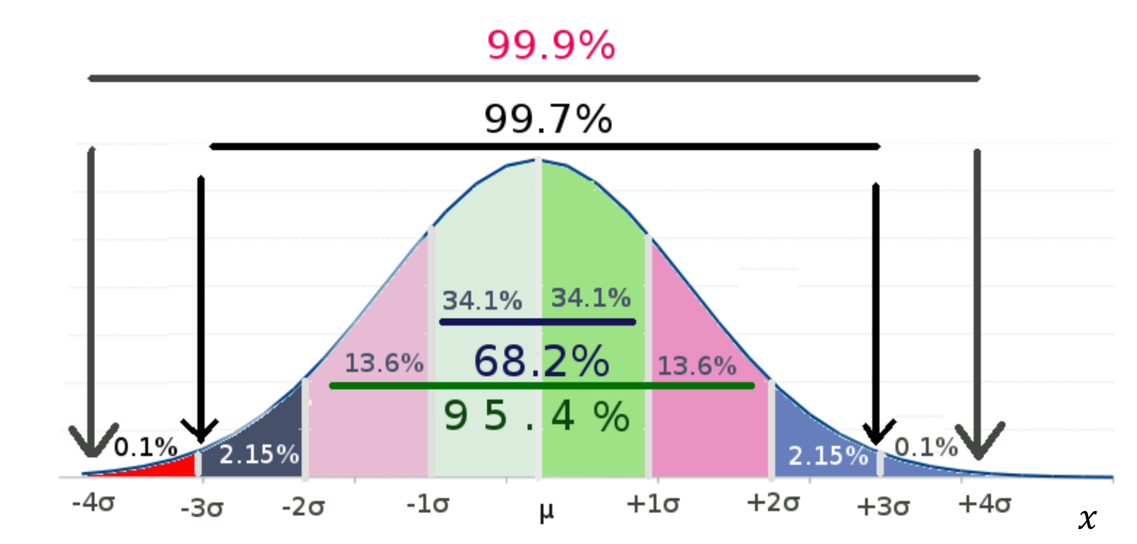
→ Gauss distribution: $P(x;\mu,\sigma) = \dfrac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Symmetric distribution:

- Expectation value: $E[x] = \mu$
- Variance: $V[x] = \sigma^2$
- Probability content:

$$\int_{-\sigma}^{+\sigma} P(x;\mu,\sigma)dx = 68.2\%$$

$$\int_{-2\sigma}^{+2\sigma} P(x;\mu,\sigma)dx = 95.4\%$$



Some other distributions

Uniform (“flat”) distribution

Exponential distribution

- Particle decay density versus time
  (in the particle’s rest frame!)

Relativistic Breit-Wigner distribution

- Distribution of resonance of unstable particle as function of centre-of-mass energy in which the resonance is produced (originates from the propagator of an unstable particle)

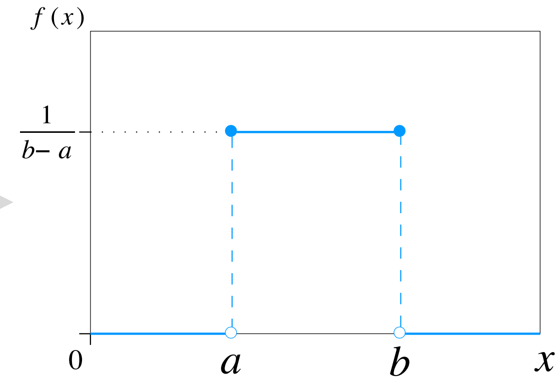
Chi-squared (χ^2) distribution

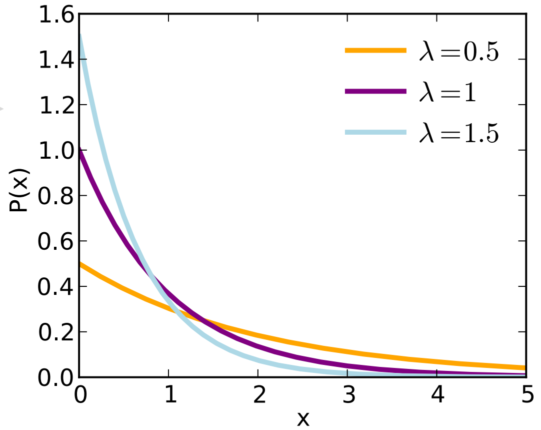
- Sum of squares of Gaussian distributed variables; used to derive goodness of a fit to describe data

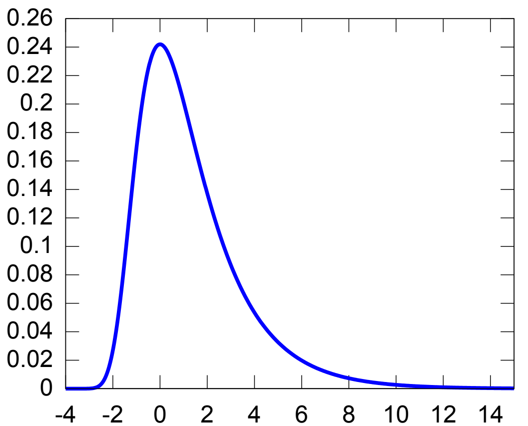
Landau distribution

- Fluctuation of energy loss by ionization of charged particle in thin matter (eg, charge deposition in silicon detector)

Many more, see http://pdg.lbl.gov/2015/reviews/rpp2015-rev-probability.pdf for definitions and properties.





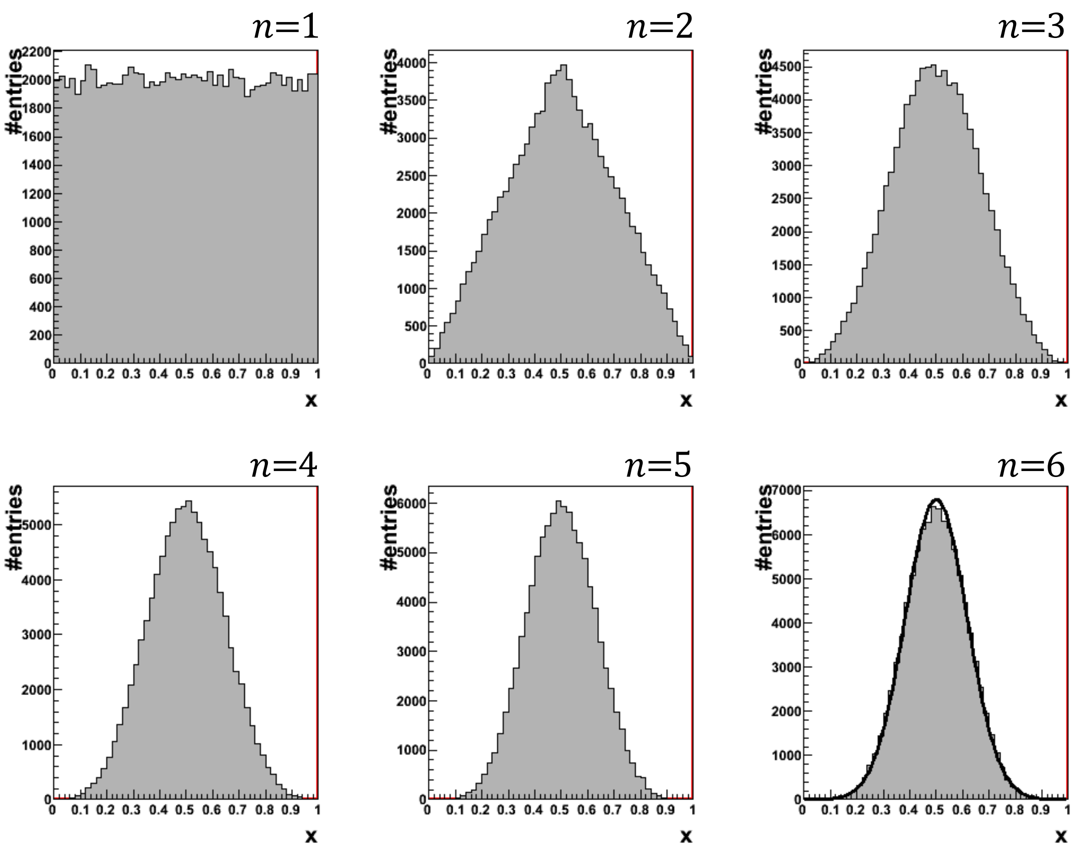


Central limit theorem (CLT)

CLT: the sum of $\boldsymbol{n}$ independent samples $\boldsymbol{x_i}$ $(i = 1, \ldots, n)$ drawn from any PDF $\boldsymbol{D(x_i)}$ with well defined expectation value and variance is Gaussian distributed in the limit $\boldsymbol{n} \to \infty$

$$\boldsymbol{D}\text{: } E_D[x_i] = \mu;\ V_D[x_i] = \sigma_D^2,\ \text{and: } y = \sum_{i=1}^{n} x_i \implies E_{\text{Gauss}}[y] = \mu;\ V_{\text{Gauss}}[y] = \frac{\sigma_D^2}{n}$$

Averaging reduces the variance

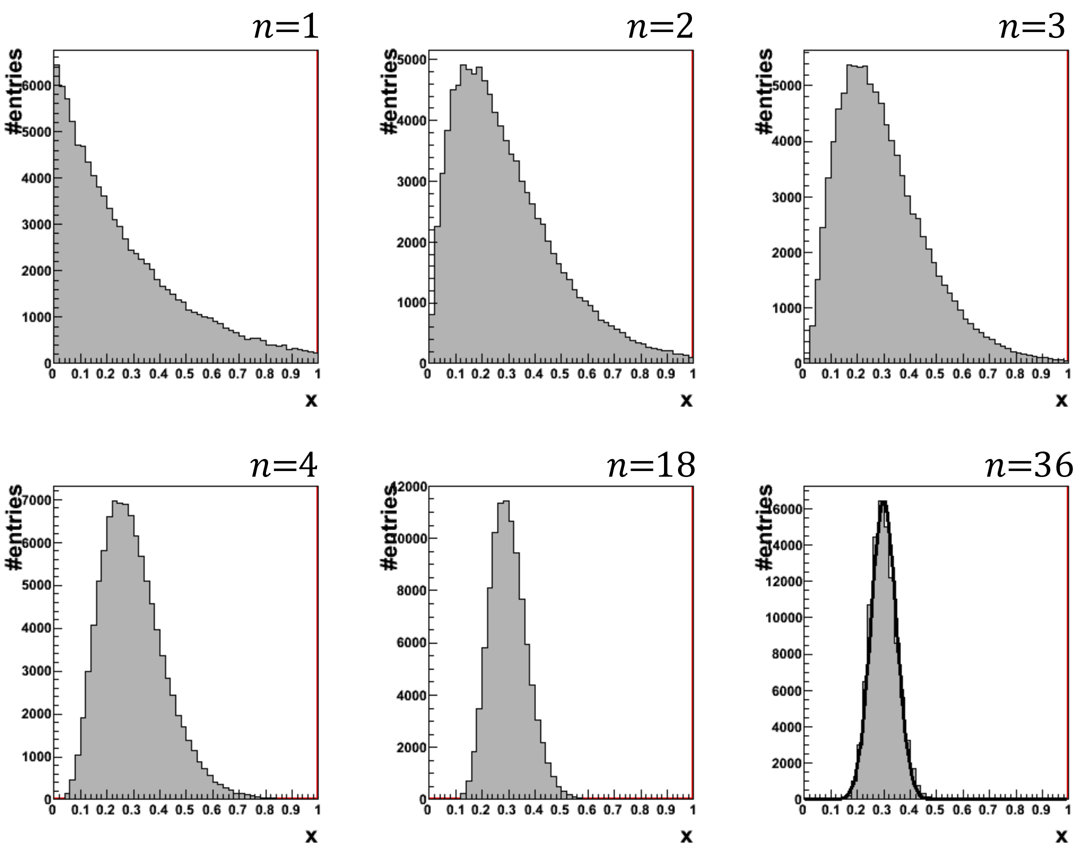


Example: summing up ensembles uniformly distributed within [0,1]

Central limit theorem (CLT)

CLT: the sum of $\boldsymbol{n}$ independent samples $\boldsymbol{x_i}$ $(i = 1, \dots, n)$ drawn from any PDF $\boldsymbol{D(x_i)}$ with well defined expectation value and variance is Gaussian distributed in the limit $\boldsymbol{n} \to \infty$

$$\boldsymbol{D}\text{: } E_D[x_i] = \mu;\ V_D[x_i] = \sigma_D^2, \text{ and: } y = \sum_{i=1}^{n} x_i \implies E_{\text{Gauss}}[y] = \mu;\ V_{\text{Gauss}}[y] = \frac{\sigma_D^2}{n}$$



Example: summing up exponential distributions

Central Gaussian limit works even if $\boldsymbol{D}$ doesn't look Gaussian at all

Multidimensional random variables

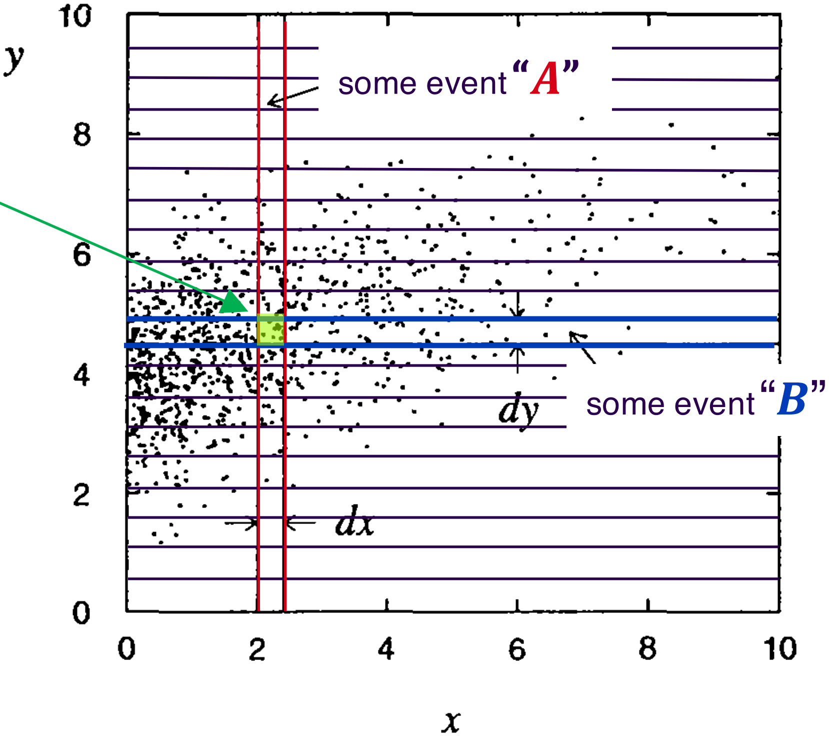
What if a measurement consists of two variables?

Let:

$\boldsymbol{A}$ = measurement $\boldsymbol{x}$ in $[\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{dx}]$

$\boldsymbol{B}$ = measurement $\boldsymbol{y}$ in $[\boldsymbol{y}, \boldsymbol{y} + \boldsymbol{dy}]$

From: Glen Cowan, Statistical data analysis



Joint probability: $P(A \cap B) = p_{xy}(x,y)\,dx\,dy$

(where $p_{xy}(x,y)$ is joint PDF)

If the two variables are independent:

$$P(A,B) = P(A) \cdot P(B)$$

$$p_{xy}(x,y) = p_x(x) \cdot p_y(y)$$

Marginal PDF: if one is not interested in dependence on $\boldsymbol{y}$ (or cannot measure it),

→ integrate out ("marginalise") $\boldsymbol{y}$, ie, project onto $\boldsymbol{x}$

→ resulting one-dimensional PDF: $p_x(x) = \int p_{xy}(x,y)\,dy$

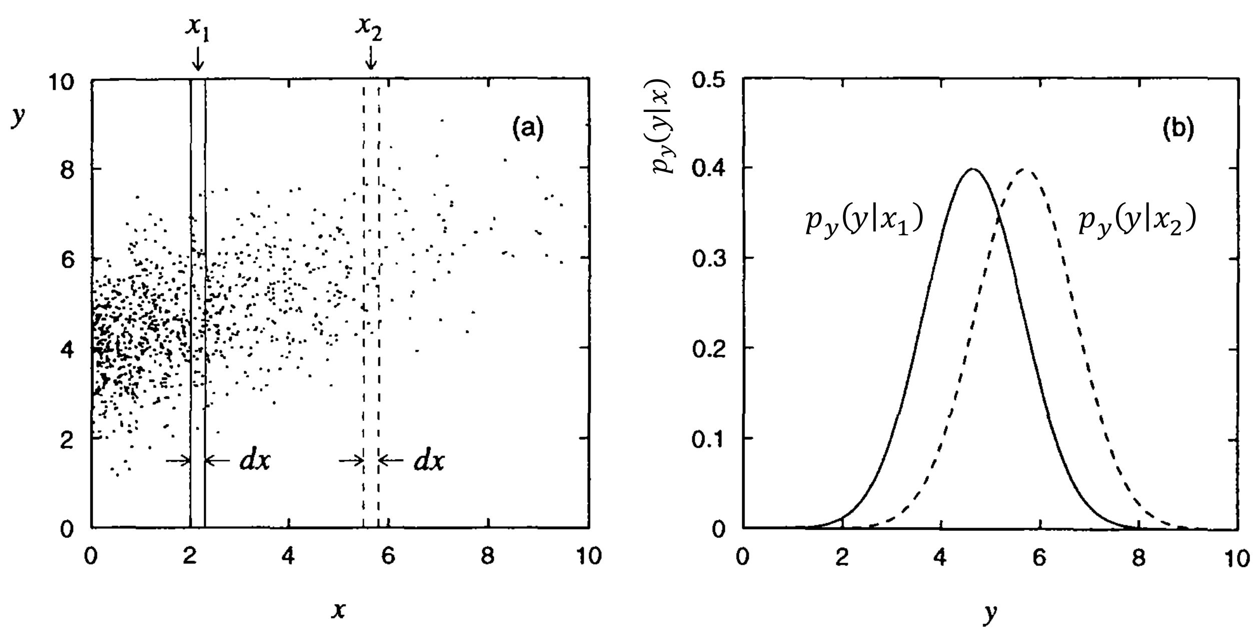
Conditioning versus marginalisation

Conditional probability $\boldsymbol{P(A|B)}$: [read: $P(A|B)$ = "probability of A *given* B"]

$$\boldsymbol{P(A|B)} = \frac{P(A \cap B)}{P(B)} = \frac{p_{xy}(x,y)dxdy}{p_y(y)dx}$$

Rather than integrating over the whole y region (marginalisation),
look at one-dimensional (1D) slices of the two-dimensional (2D) PDF $p_{xy}(x,y)$:

$$p_y(y|x_1) = p_{xy}(x = \text{const} = x_1, y)$$



From: Glen Cowan, *Statistical data analysis*

Covariance and correlation

Recall, for 1D PDF $\boldsymbol{p_x(x)}$ we had: $E[x] = \mu_x; \;\; V[x] = \sigma_x^2$

For a 2D PDF $\boldsymbol{p_{xy}(x, y)}$, one correspondingly has: μ_x, μ_y, σ_x, σ_y

How do $\boldsymbol{x}$ and $\boldsymbol{y}$ co-vary ? → $\mathrm{C}_{xy} = \mathrm{covariance}_{xy} = E\big[(x - \mu_x)\big(y - \mu_y\big)\big] = E[xy] - \mu_x\mu_y$

Or the scale / dimension invariant *correlation coefficient*:

$$\rho_{xy} = \frac{\mathrm{C}_{xy}}{\sigma_x \sigma_y}$$, where $\rho_{xy} \subset [-1, +1]$

- If x, y are independent: $\rho_{xy} = 0$, ie, they are *uncorrelated* (or they *factorise*)

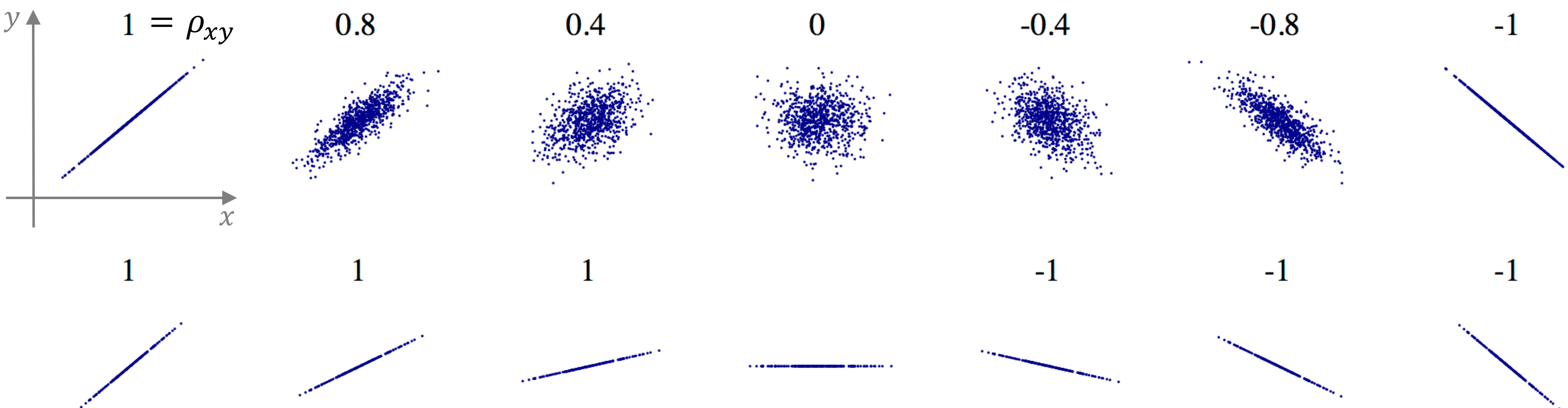
  Proof: $E[xy] = \iint xy \cdot p_{xy}\,(x, y) dx dy = \int x \cdot p_x(x) dx \cdot \int y \cdot p_y(y) dy = \mu_x \mu_y$

- Note that the contrary is not always true: non-linear correlations can lead to $\rho_{xy} = 0$,
  → see next page

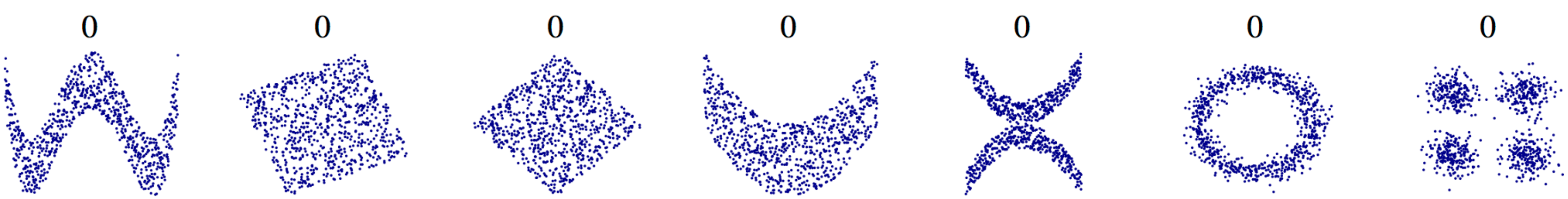
Correlations

Figure from: https://en.wikipedia.org/wiki/Correlation_and_dependence

The correlation coefficient measures the noisiness and direction of a linear relationship:



…it does not measure the slope ρ_{xy} (see above figures)



…and non-linear correlation patterns are not or only approximately captured by ρ_{xy} (see above figures)

Correlations

Non-linear correlation can be captured by the "*mutual information*" quantity $\boldsymbol{I_{xy}}$:

$$I_{xy} = \iint p_{xy}(x,y) \cdot \ln\left(\frac{p_{xy}(x,y)}{p_x(x)p_y(y)}\right) dxdy$$

where I_{xy} =0 only if $\boldsymbol{x}, \boldsymbol{y}$ are fully statistically independent

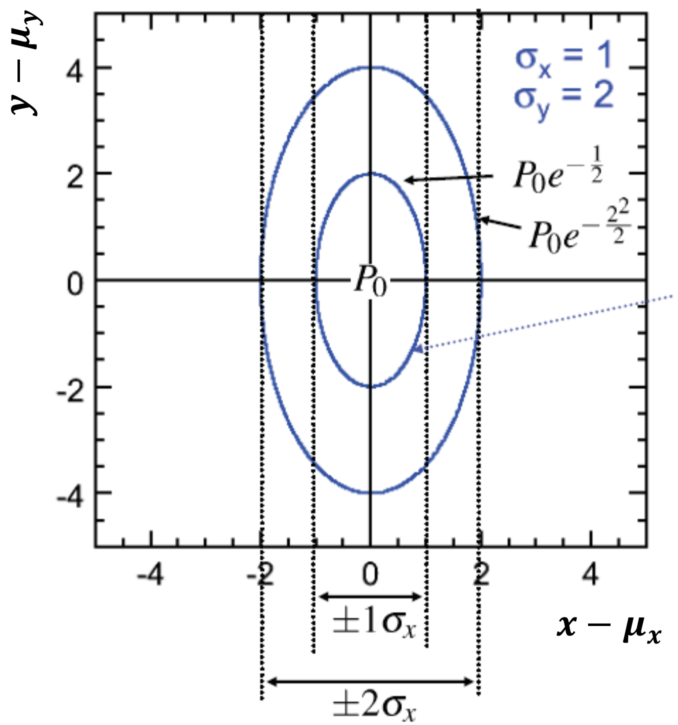
Proof: if independent, then $p_{xy}(x,y) = p_x(x)p_y(y) \Rightarrow \ln(\dots) = 0$

NB: $I_{xy} = H_x - H_x(y) = H_y - H_y(x)$,
where $H_x = -\int p_x(x) \cdot \ln\big(p_x(x)\big)\, dx$ is *entropy*, $H_x(y)$ is *conditional entropy*

2D Gaussian (uncorrelated)

Two variable $\boldsymbol{x}, \boldsymbol{y}$ are independent: [$p_{xy}(x, y) = p_x(x) \cdot p_y(y)$]

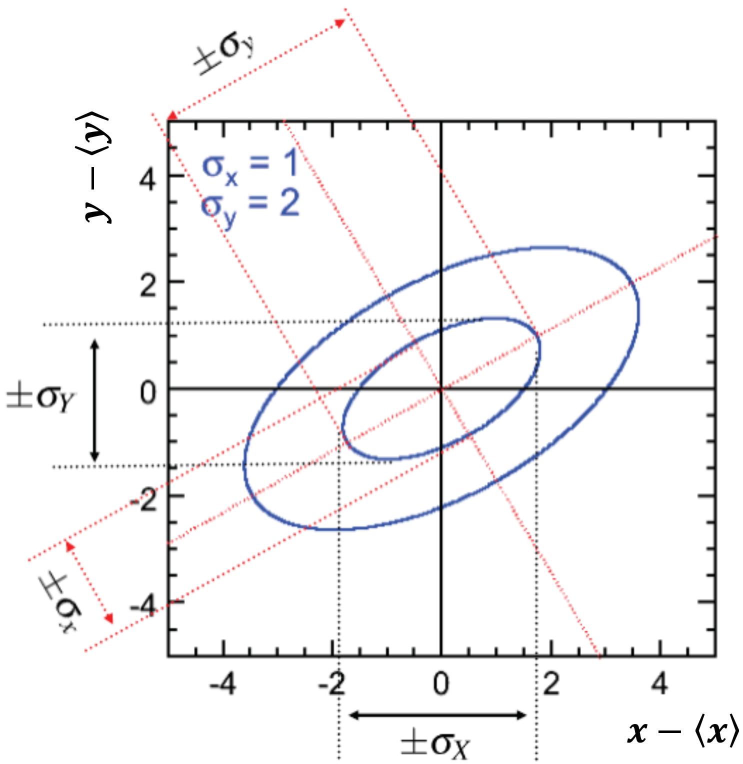
$$P(x, y) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}$$



2D Gaussian (correlated)

Two variable $\boldsymbol{x}, \boldsymbol{y}$ are *not* independent: [$p_{xy}(x,y) \neq p_x(x) \cdot p_y(y)$]

$$P(\vec{x}) = \frac{1}{2\pi\sqrt{\det(C)}} \cdot \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T C^{-1} (\vec{x}-\vec{\mu})\right)$$



where (in 2D case):

$$C = \begin{pmatrix} \langle x^2\rangle - \langle x\rangle^2 & \langle xy\rangle - \langle x\rangle\langle y\rangle \\ \langle xy\rangle - \langle x\rangle\langle y\rangle & \langle y^2\rangle - \langle y\rangle^2 \end{pmatrix}$$

is the (symmetric) *covariance matrix*

Corresponding correlation matrix elements:

$$\rho_{ij} = \rho_{ji} = \frac{C_{ij}}{\sqrt{C_{ii} \cdot C_{jj}}}$$

SQRT decorrelation

Find variable transformation that diagonalises a covariance matrix C

Determine "square-root" C' of C (such that: $C = C' \cdot C'$) by first diagonalising C

$$D = S^T \cdot C \cdot S \quad \Leftrightarrow \quad C' = S \cdot \sqrt{D} \cdot S^T$$

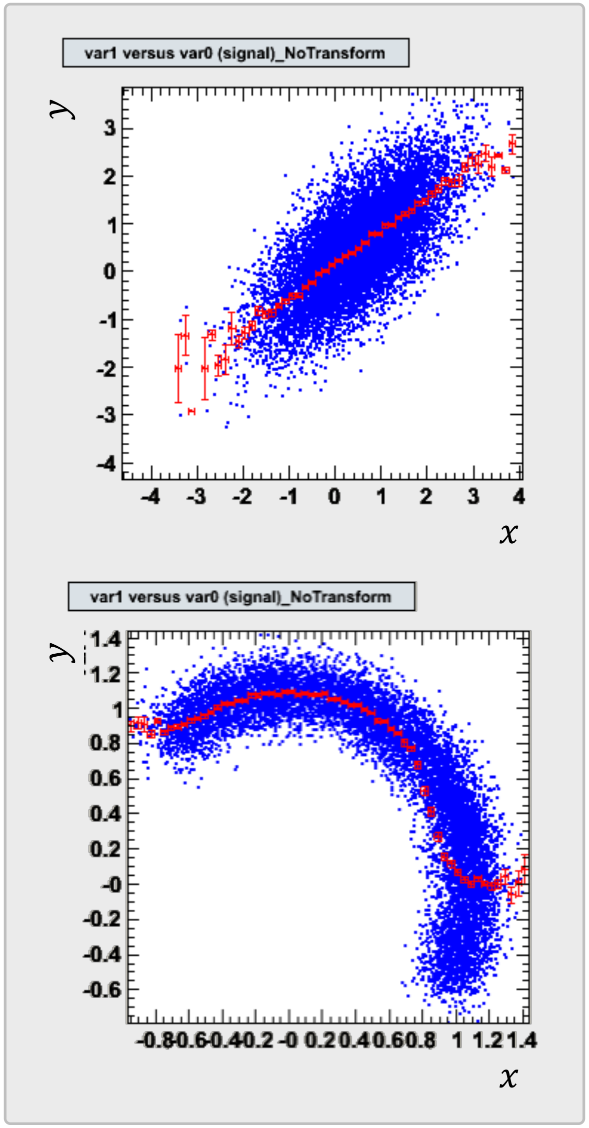
where D is diagonal, $\sqrt{D} = \{\sqrt{d_{11}}, \dots, \sqrt{d_{nn}}\}$, and S an orthogonal matrix

Linear decorrelation of correlated vector $\boldsymbol{x}$ then obtained by

$$\boldsymbol{x}_{\text{decorr}} = (C')^{-1} \cdot \boldsymbol{x}$$

Principle component analysis (PCA) is another convenient method to achieve linear decorrelation

(PCA is linear transformation that rotates a vector such that the maximum variability is visible. It identifies most important gradients)



Example: original correlations

SQRT decorrelation

Find variable transformation that diagonalises a covariance matrix C

Determine "square-root" C' of C (such that: $C = C' \cdot C'$) by first diagonalising C

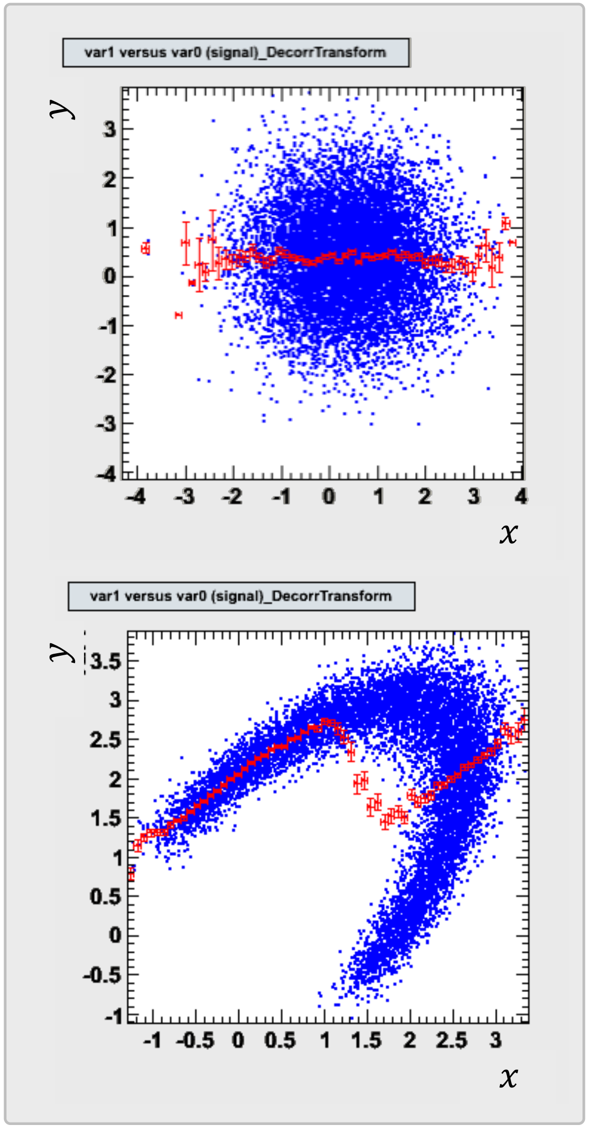
$$D = S^T \cdot C \cdot S \quad \Leftrightarrow \quad C' = S \cdot \sqrt{D} \cdot S^T$$

where D is diagonal, $\sqrt{D} = \{\sqrt{d_{11}}, \dots, \sqrt{d_{nn}}\}$, and S an orthogonal matrix

Linear decorrelation of correlated vector $\boldsymbol{x}$ then obtained by

$$\boldsymbol{x}_{\mathrm{decorr}} = (C')^{-1} \cdot \boldsymbol{x}$$

SQRT decorrelation works only for linear correlations!



Example: after SQRT decorrelation

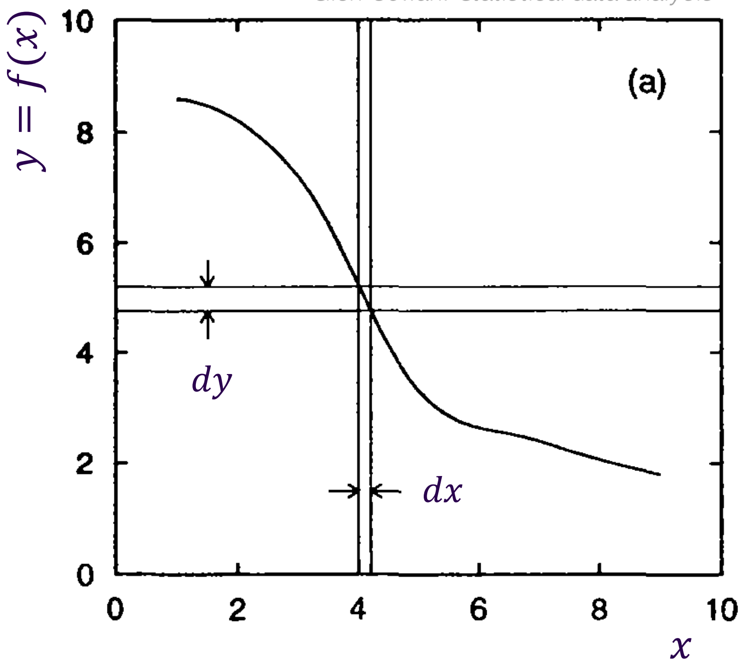
Functions of random variables

Any function of a random variable is itself a random variable

E.g., $\boldsymbol{x}$ with PDF $\boldsymbol{p_x(x)}$ becomes: $\boldsymbol{y = f(x)}$

$\boldsymbol{y}$ could be a parameter extracted from a measurement

Glen Cowan: Statistical data analysis



What is the PDF $\boldsymbol{p_y(y)}$?

- Probability conservation: $p_y(y)|dy| = p_x(x)|dx|$
- For a 1D function $f(x)$ with existing inverse:

$$dy = \frac{df(x)}{dx}dx \iff dx = \frac{df^{-1}(y)}{dy}dy$$

- Hence: $\boldsymbol{p_y(y)} = p_x(f^{-1}(y))\left|\frac{dx}{dy}\right|$

Note: this is **not** the standard error propagation but the full PDF !

Error propagation

Let's assume a measurement $\boldsymbol{x}$ with *unknown* PDF $\boldsymbol{p_x(x)}$, and a transformation $\boldsymbol{y = f(x)}$

- $\bar{x}$ and $\hat{V}$ are estimates of μ and variance σ^2 of $p_x(x)$

What are $E[y]$ and, in particular, $\boldsymbol{\sigma_y^2}$? → Taylor-expand $f(x)$ around $\bar{x}$:

- $f(x) = f(\bar{x}) + \left.\frac{df}{dx}\right|_{x=\bar{x}}(x-\bar{x}) + \cdots \quad \Rightarrow \quad E[f(x)] \simeq f(\bar{x})$ (because: $E[x-\bar{x}] = 0$!)

Now define $\bar{y} = f(\bar{x})$, and from the above follows:

$$\Leftrightarrow \quad y - \bar{y} \simeq \left.\frac{df}{dx}\right|_{x=\bar{x}}(x-\bar{x})$$

$$\Leftrightarrow \quad E[(y-\bar{y})^2] = \left(\left.\frac{df}{dx}\right|_{x=\bar{x}}\right)^2 E[(x-\bar{x})^2]$$

$$\Leftrightarrow \quad \hat{V}_y = \left(\left.\frac{df}{dx}\right|_{x=\bar{x}}\right)^2 \hat{V}_x$$

$\Leftrightarrow \quad \sigma_y = \left.\frac{df}{dx}\right|_{x=\bar{x}} \cdot \sigma_x \qquad \rightarrow \qquad$ (approximate) error propagation

Error propagation (continued)

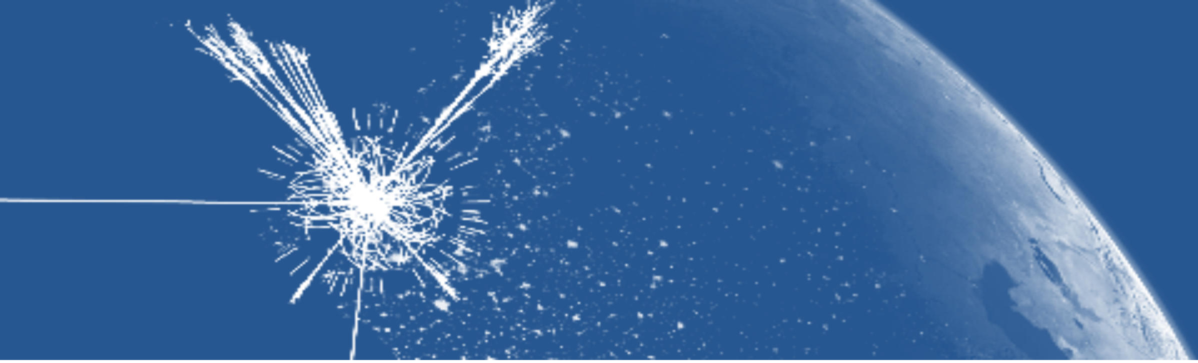
In case of several variables, compute covariance matrix and partial derivatives

- Let $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{x_1}, \dots, \boldsymbol{x_n})$ be a function of $\boldsymbol{n}$ randomly distributed variables
- $\left(\frac{df}{dx}\Big|_{x=\bar{x}}\right)^2 \hat{V}_x$ becomes: $\displaystyle\sum_{i,j=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\Bigg|_{\bar{x}} \cdot \hat{V}_{i,j}$ (where: $\bar{x} = (\bar{x}_1, \dots, \bar{x}_n)$)
- with the covariance matrix:

$$\hat{V}_{i,j} = \begin{bmatrix} \sigma_{x_1}^2 & \cdots & \sigma_{x_1 x_n} \\ \vdots & \ddots & \vdots \\ \sigma_{x_n x_1} & \cdots & \sigma_{x_n}^2 \end{bmatrix}$$

The resulting "error" (uncertainty) depends on the correlation of the input variables

- Positive correlations lead to an increase of the total error
- Negative correlations decrease the total error



Summary for today

Probability and statistics are everywhere in science, and in particular profoundly contained in particle physics and in the physics of large ensembles

Overview of some important probability density distributions given, also:

- Joint / marginal / conditional probabilities
- Covariance and correlations
- Error propagation

Next: how to use these concepts for hypothesis testing