Statistical Methods in Particle Physics / WS 13

Lecture III

PDFs, Error Propagation, Numbers on a Computer

Niklaus Berger

Physics Institute, University of Heidelberg

Emmy Noether-Programm

Deutsche Forschungsgemeinschaft

DFG

Recap of Lecture II: Probability Density Function

Suppose outcome of an experiment is a continuous value x:

- P(x in [x + dx]) = f(x)dx
- with f(x) the probability density function
- and $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (Normalization)
- f(x) > 0
- f(x) is not a probability; it has dimension 1/x

Recap of Lecture II: PDF Properties

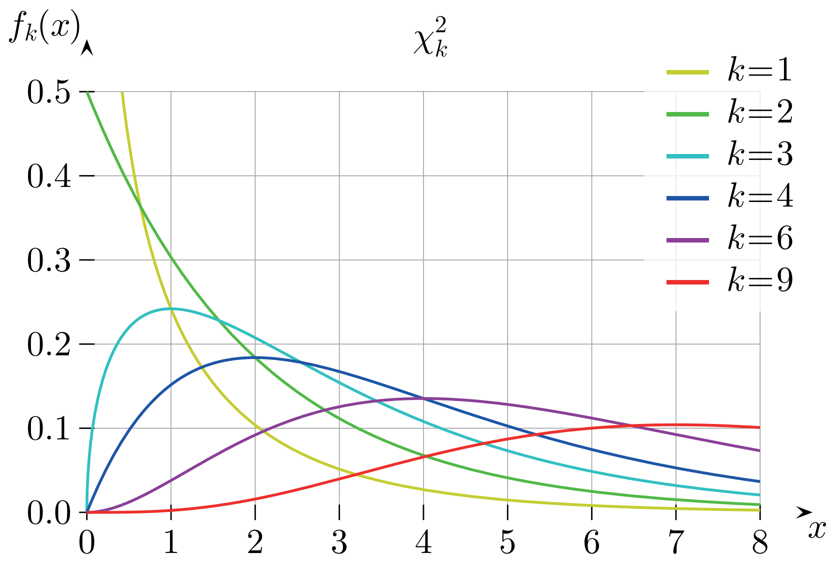
- Continuous random variable x with pdf f(x)
- Expectation value or mean:

$$E[x] = \int x f(x) dx = \mu$$

- Variance:

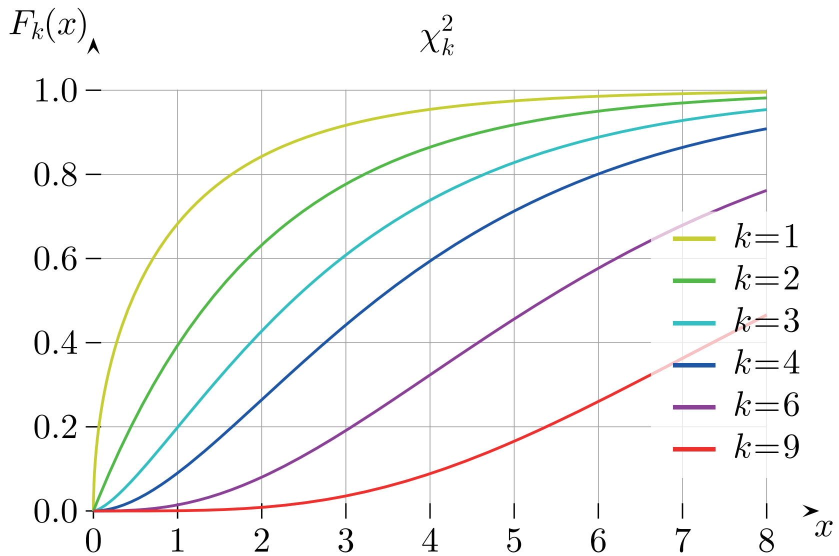
$$V[x] = \int (x-y)^2 f(x) dx = \sigma^2$$

2.7. χ^2 Distribution



$$P_\chi(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

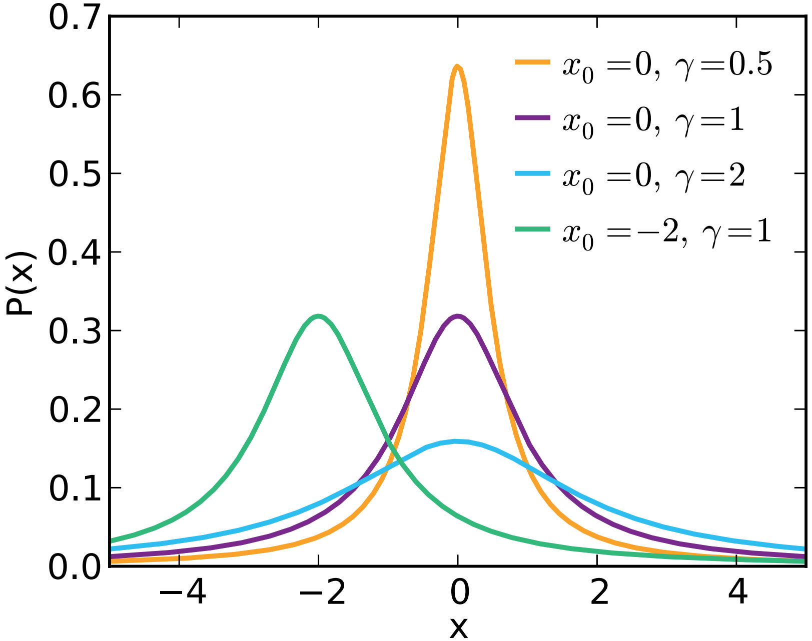
- Mean: $E[x] = k$
- Variance: $V[x] = 2k$

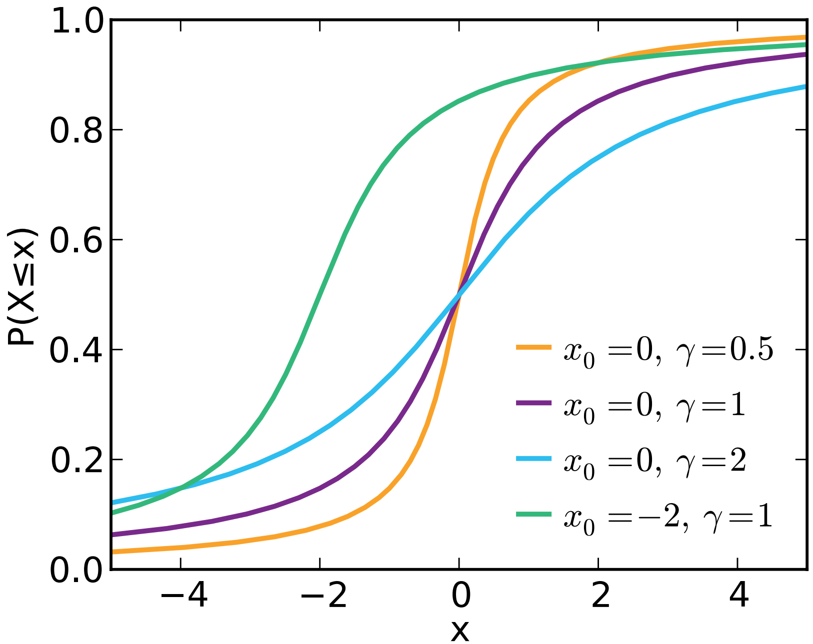


- Sum of squares of a standard normal distributed variable follows χ^2 distribution
- Application in goodness of fit tests (especially least squares) - details soon
- χ^2 distribution for 2 degrees of freedom is an exponential

Plot source: Wikipedia

2.8. Cauchy (Breit-Wigner) Distribution





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

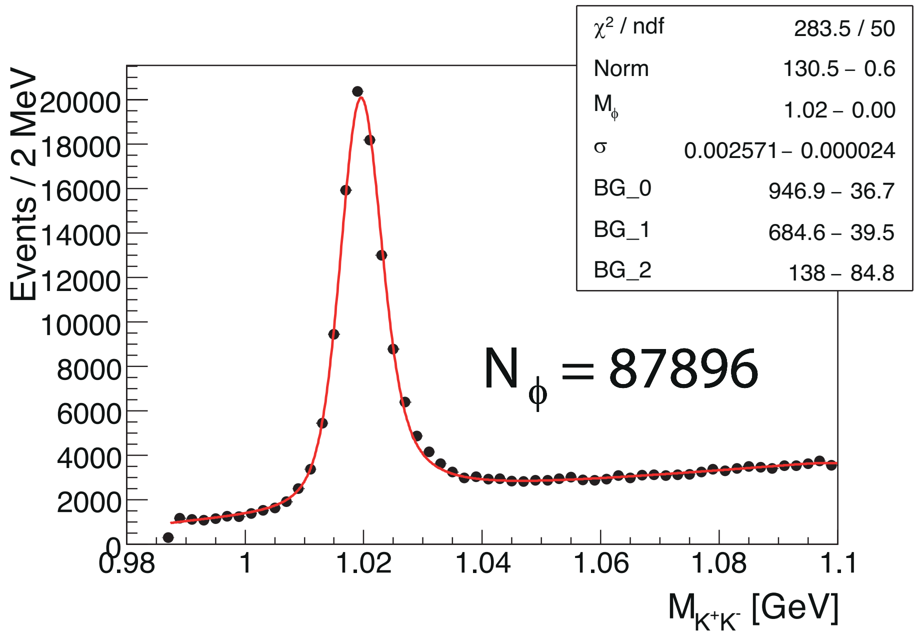
- Mean: $E[x]$ = undefined
- Variance: $V[x]$ = undefined
- Mode: x_0

Plot source: Wikipedia

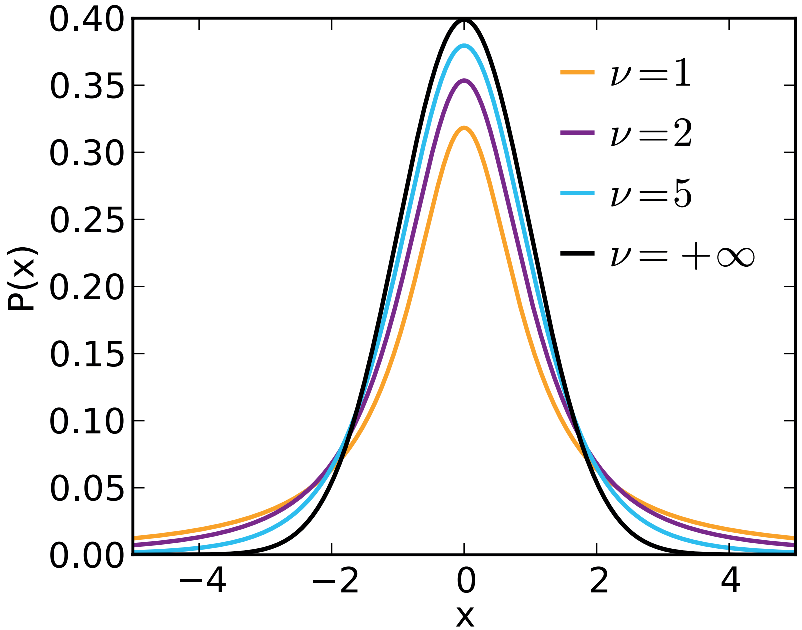
Breit-Wigner Distribution: Applications

- Damped resonance
- Decay of unstable particles: Here $\phi \rightarrow K^+K^-$
- Note that infinite variance is not really a problem, as there are edges to the phase space

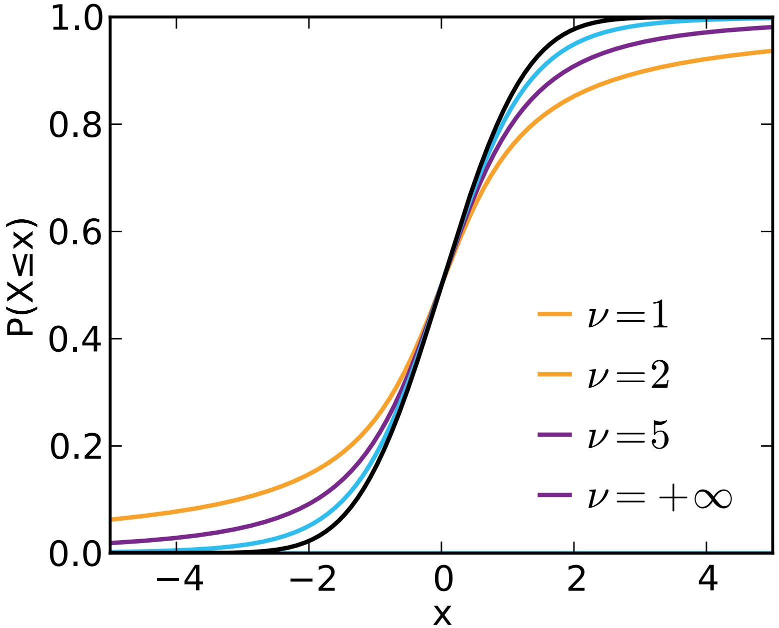
- Do not fit two Breit-Wigners in the same mass distribution; there is always interference



2.8. Student's t Distribution



$$P_S(x;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\;\Gamma(\nu/2)}\left(1+\frac{x^2}{\nu}\right)^{\left(\frac{\nu+1}{2}\right)}$$

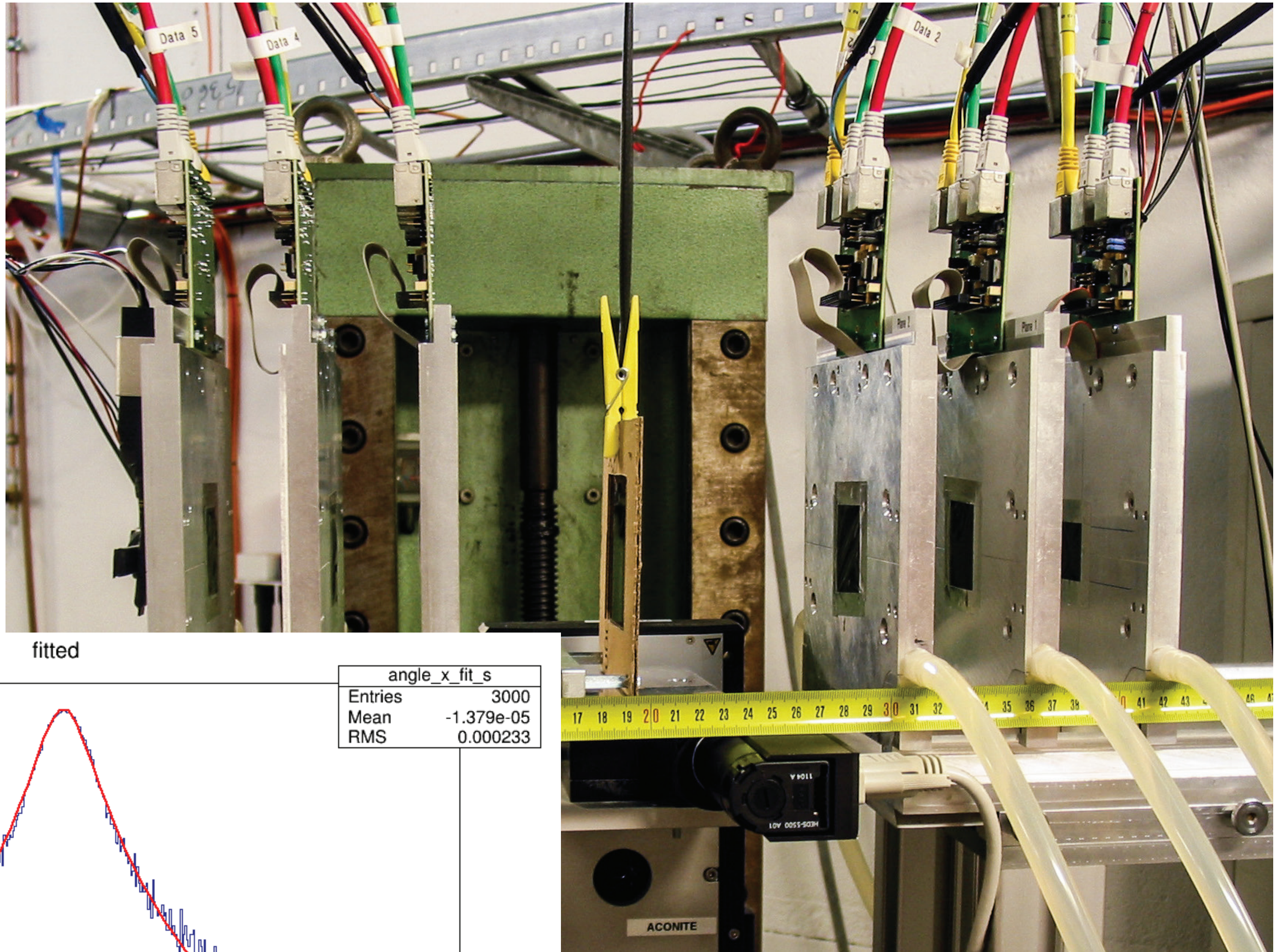


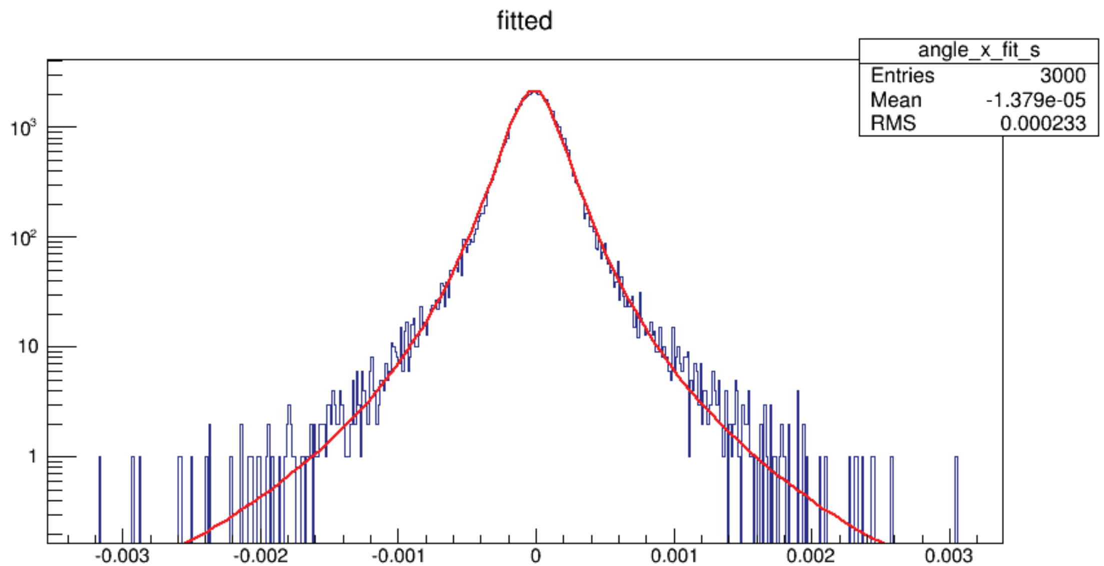
- Mean: $E[x] = 0 \qquad (\nu > 1)$
- Variance: $V[x] = \nu/(\nu-2) \quad (\nu > 2)$

Plot source: Wikipedia

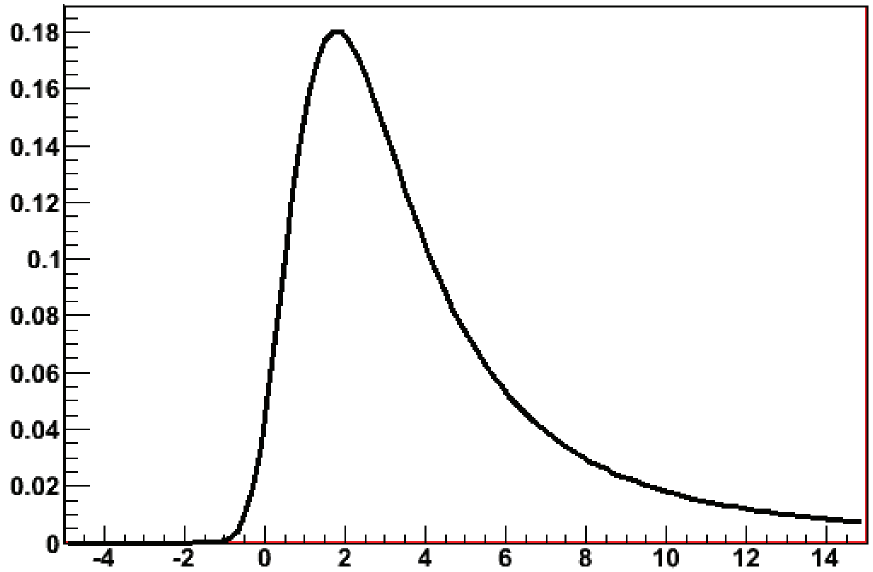
Student's t Distribution: Applications

- $\nu = 1$ gives Cauchy
- $\nu \rightarrow \infty$ gives normal distribution
- Nice for anything with non-Gaussian tails
- Example: Fit to multiple scattering distribution





2.9. Landau Distribution



$$P_L(x) = \frac{1}{\pi}\int_0^\infty e^{-t\log t - xt}\sin(\pi t)\,dt$$

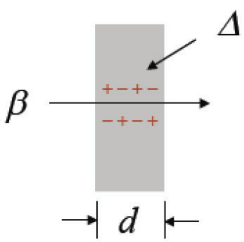
Describes energy loss of a particle in a slab of material; long tail due to hard collisions

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

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

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

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



Mode depends on velocity β of particle - can be used for particle identification

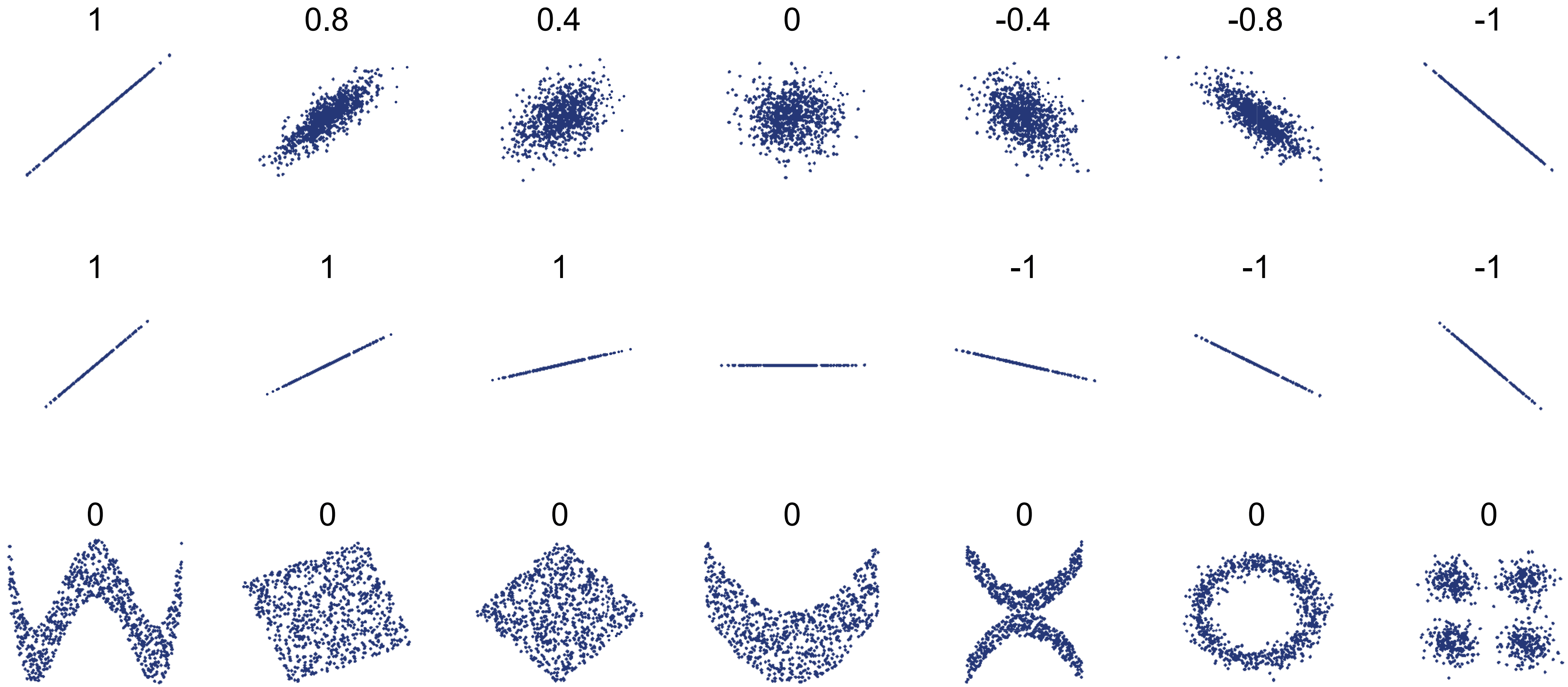
L. Landau, J. Phys. USSR **8** (1944) 201; see also
W. Allison and J. Cobb, Ann. Rev. Nucl. Part. Sci. **30** (1980) 253.

Plot source: Wikipedia

Part III:

Error Propagation and Correlations

Correlation Coefficient



Plot source: Wikipedia

Part IV:

Monte Carlo

Monte Carlo Methods

Replace analytical calculations by random sampling

Usually with a computer

4.1. Numbers on a computer - unsigned integers

Internally, numbers are always represented in binary - 1 bit as fundamental unit

C (and thus C++) does not specify data type sizes - if you want to be sure, use

`#include <stdint.h>` otherwise you might get different results on different machines

- 8 bit is a byte
  `unsigned char` or better `uint8_t` 0 to 255
- 16 bit
  `unsigned short` or better `uint16_t` 0 to 65535
- 32 bit
  `unsigned int` or better `uint32_t` 0 to 4'294'967'295
- 64 bit
  `unsigned long long int` or better `uint64_t` 0 to 18'446'744'073'709'551'615

Signed integers

Signed numbers are represented in 2's complement encoding

C (and thus C++) does not specify data type sizes - if you want to be sure, use

`#include <stdint.h>` otherwise you might get different results on different machines

- 8 bit is a byte
  `char` or better `int8_t` -128 to 127
- 16 bit
  `short` or better `int16_t` -32768 to 32767
- 32 bit
  `int` or better `int32_t` -2'147'483'648 to 2'147'483'647
- 64 bit
  `long long int` or better `int64_t` -9'223'372'036'854'775'808 to 9'223'372'036'854'775'807