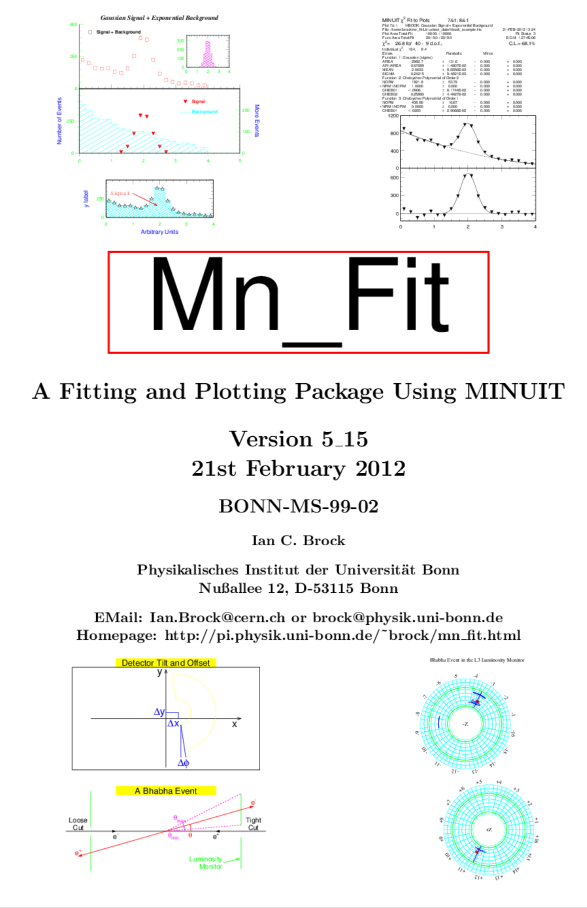
Statistical Methods in Particle Physics

Heidelberg+LHCb Workshop
Neckarzimmern

Ian C. Brock
22nd February 2012

Part 1


Mn_Fit

A Fitting and Plotting Package Using MINUIT

Version 5_15
21st February 2012

BONN-MS-99-02

Ian C. Brock

Physikalisches Institut der Universität Bonn
Nußallee 12, D-53115 Bonn

EMail: Ian.Brock@cern.ch or brock@physik.uni-bonn.de
Homepage: http://pi.physik.uni-bonn.de/~brock/mn_fit.html

CV of Statistics for Data Analysis

- TASSO PhD – unbinned likelihood fit to muon pair forward-backward asymmetry
- Crystal Ball – mostly hardware
- CLEO – development of Mn_Fit for fitting $\Upsilon(3S)$ decays to $\pi\pi$ and invariant mass spectrum – visualisation of MINUIT fit results; ended up as competition for PAW
  - Over 20 years later, Mn_Fit still used in CLEO
- L3 – LEP EWWG heavy flavour combinations
- ZEUS – looking for new physics
- Lecturing – Statistical Methods of Data Analysis
- CLEO-c – fitting series of D decays
- ATLAS – how to combine measurements
- So far mostly Fortran and Mn_Fit, learning C++ and Root

Overview

- Part 1:
  - Tools and literature
  - Measurements and presentation of data
  - Distributions
  - Central limit theorem
  - Error propagation
- Part 2:
  - Systematic errors
  - Estimation
    - Likelihood
      - Maximum likelihood examples
    - Least squares
      - Straight line fit
    - Bayesian statistics
    - Confidence levels
    - (Hypothesis testing)

Tools and Literature

- Spreadsheets
- Gnuplot etc.
- PAW, Mn_Fit
  - Fortran based
  - “Intuitive” commands – can be abbreviated
- Root
  - Full power and complexity of C++
- Origin etc.

- R. Barlow: Statistics
- A. Frodeson et al: Probability and Statistics in Particle Physics.
- G. Cowan: Statistical Data Analysis
- S. Brandt: Data Analysis
- W. Verkerke: Data Analysis BND 2004
  www.slac.stanford.edu/~verkerke/bnd2004/data_analysis_2004_v17.ppt
- I. Brock - Statistical Methods of Data Analysis
  pi.physik.uni-bonn.de/~brock/teaching/stat_ss11
- MINUIT manual
- Mn_Fit manual
  pi.physik.uni-bonn.de/~brock/mn_fit.html

Measurements

- Experimental measurement:
  value ± error
  - Calculate result by combining data
  - Calculate errors
  - Compare with expectations

Measure
$g = 9.70 \pm 0.15\ \mathrm{m\ s^{-2}}$
Expect
$g = 9.81\ \mathrm{m\ s^{-2}}$

- Questions:
  - How to get to result?
  - How to **ESTIMATE** the errors?
  - Sources of error?
    - Statistical (random fluctuations)
    - Systematic (apparatus, procedure, ...)
      - Random (e.g. intercalibration)
      - Bias (e.g. energy scale)
      - Forgotten/unknown effects

How accurately can you estimate the error?
<10% is doing well ⇒ ≤2 significant digits for error!
I never want to see $g = 9.7034 \pm 0.1545\ \mathrm{m\ s^{-2}}$

Interpretation of Error

Measure $g = 9.70 \pm 0.15\ \mathrm{m\ s^{-2}}$

- Engineer:
  - $9.55 \leq g \leq 9.85$
  - i.e. error indicates tolerance or range of allowed values
- Physicist:
  - Repeat experiment many times
  - $9.55 \leq g \leq 9.85$ 68% of the time
  - $9.40 \leq g \leq 10.00$ 95% of the time (assuming Gaussian errors)

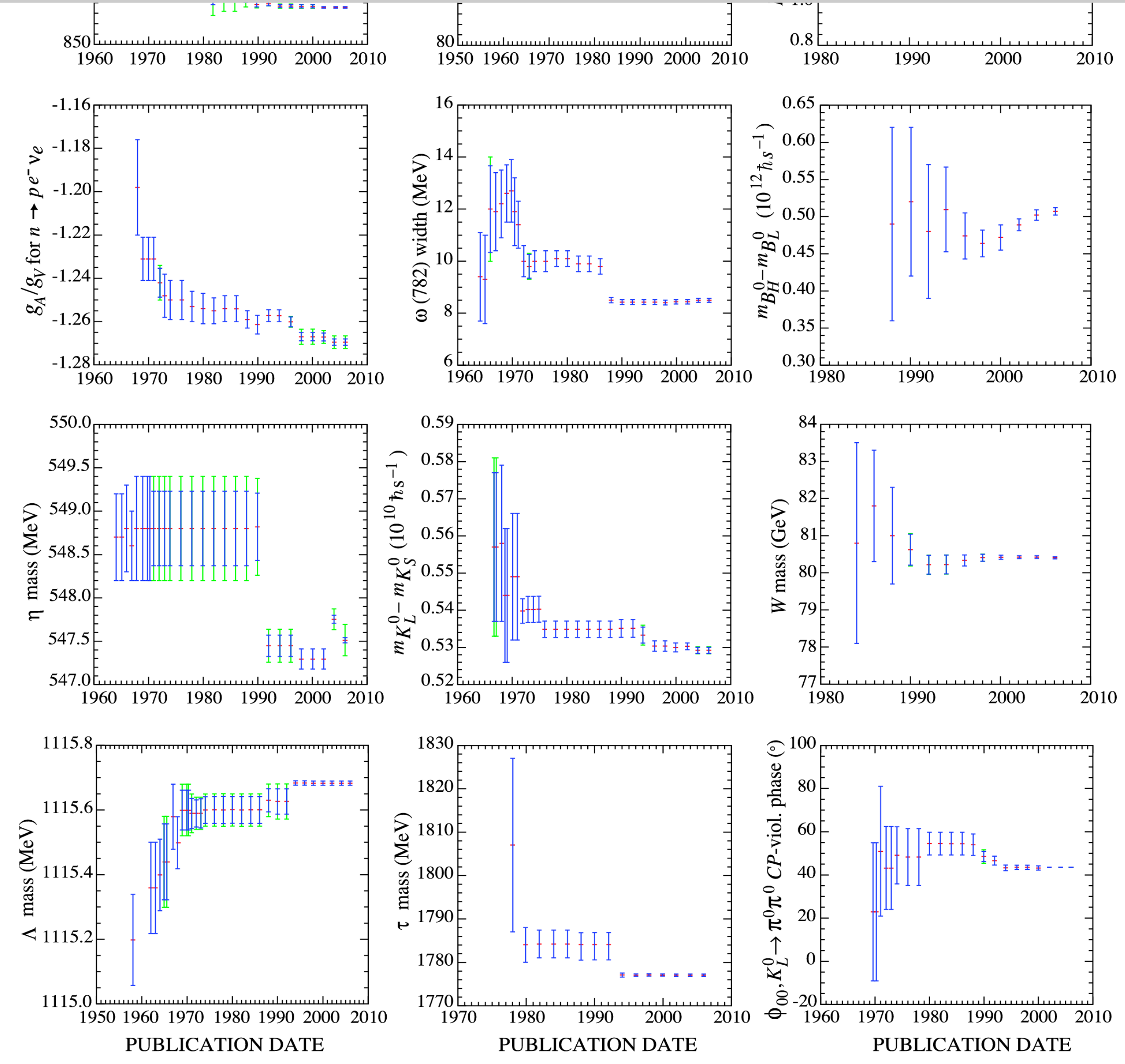
Warning: When evaluating systematic errors, tendency is often to treat them as a tolerance and not as a Gaussian error!!

Possible Conclusions

- Good experiment
  - Agrees with expectations
- New discovery
  - Book ticket to Stockholm
- Measurement not good enough
  - How can we improve it?
- Which reaction is the correct one?
- Measurement does not agree with expectations.
- Why?
  - Experiment wrong
  - Errors underestimated
  - New discovery
- Such questions often only asked in this case!
- Even worse:
  - Which error sources move result in "right" direction?

Warning: You are supposed to be *objective!*
Do not let *subjective* prejudices influence your considerations

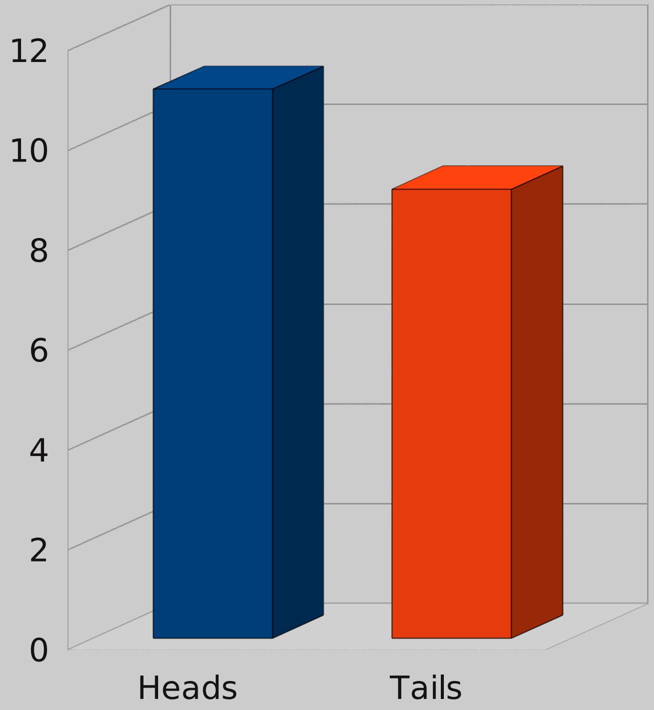
PDG Experience



Just in case you thought professional physicists were completely objective and scientific in their approach!

Data and Their Presentation

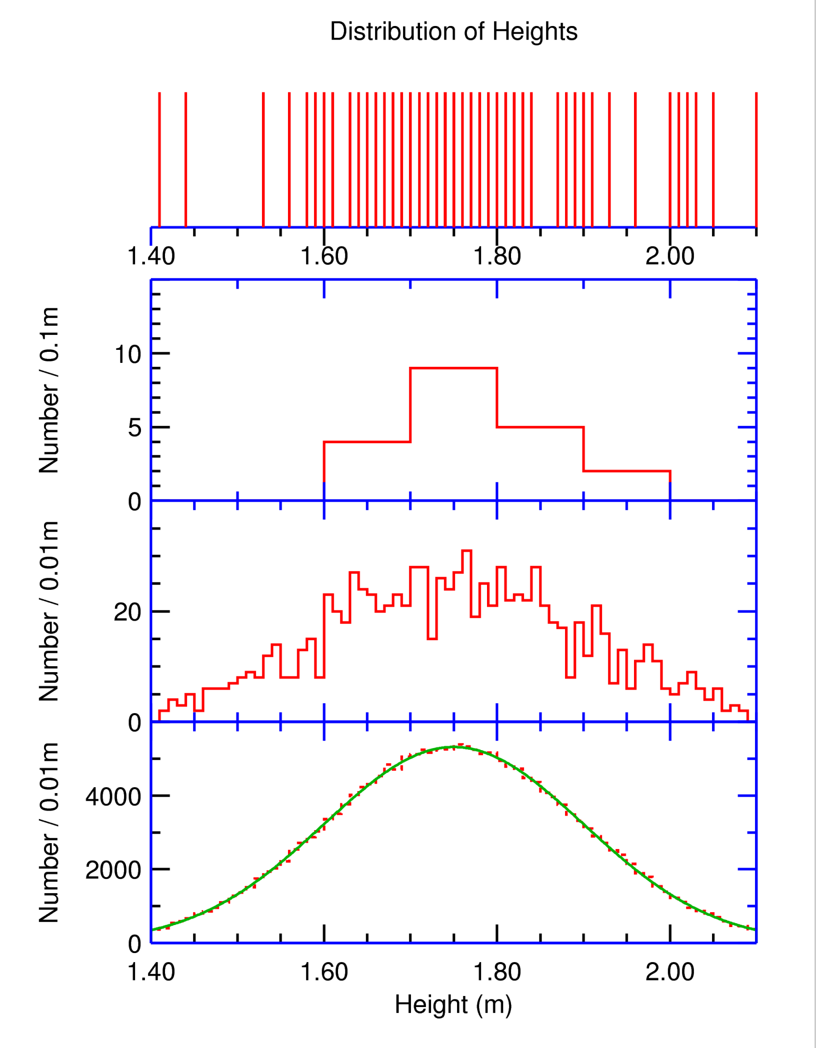
- Data can be qualitative or quantitative
- Only discuss quantitative data here
- Discrete (integers, head/tail)
  - Use list, set or bar chart:
    - HHHTTHHHTHTTTHHTTHHT
    - {11 heads, 9 tails}
    - Bar chart



- Continuous (reals/floats)
  - Histogram common
  - Height ∝ number of entries? Usual way to fill histograms
  - Area ∝ number of entries? Appropriate for cross-sections

Continuous Data

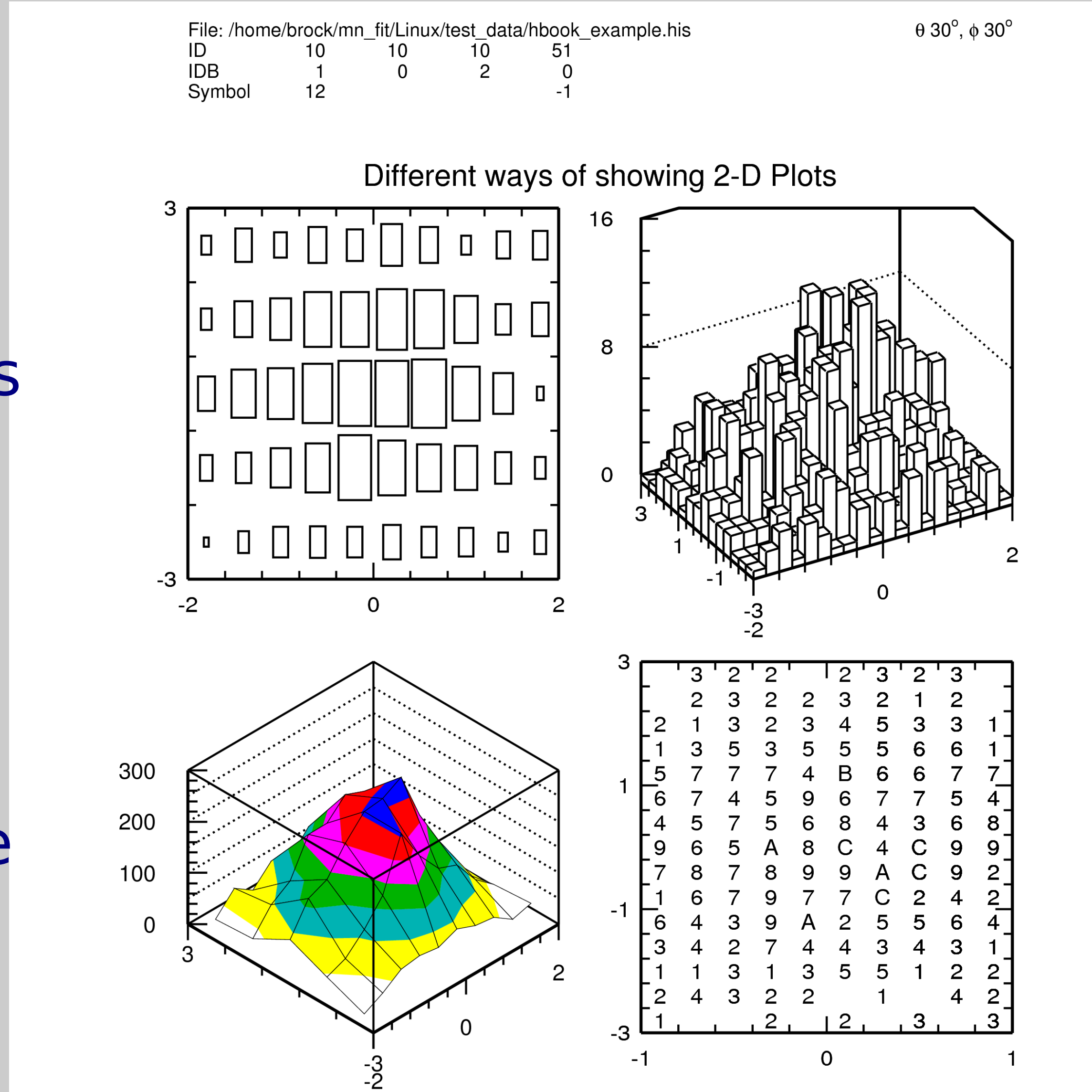
- Histogram by far most common way of showing data
- Bin width?
  - Appropriate for statistics
  - Similar to resolution (measurement accuracy)
  - Enough bins to see structure
- Don't forget to label the axes
- Make sure scale and labels are large enough
- Pie charts are a good way to split data in different categories



Two Dimensions

Area ∝ Entries

Lego

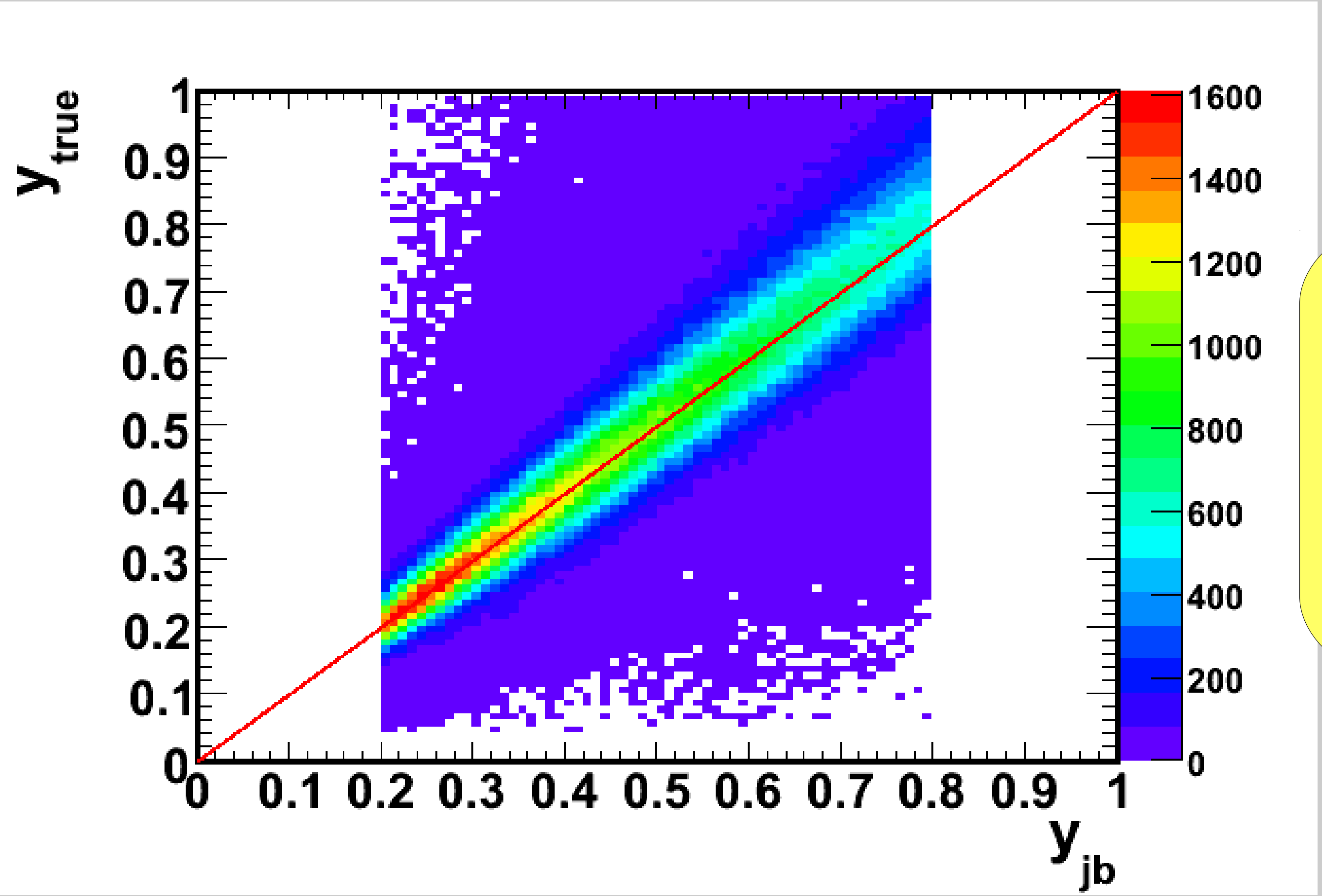


Surface

Table

"Scatter plot" with # dots ∝ bin entries also popular

Colour and Two Dimensions



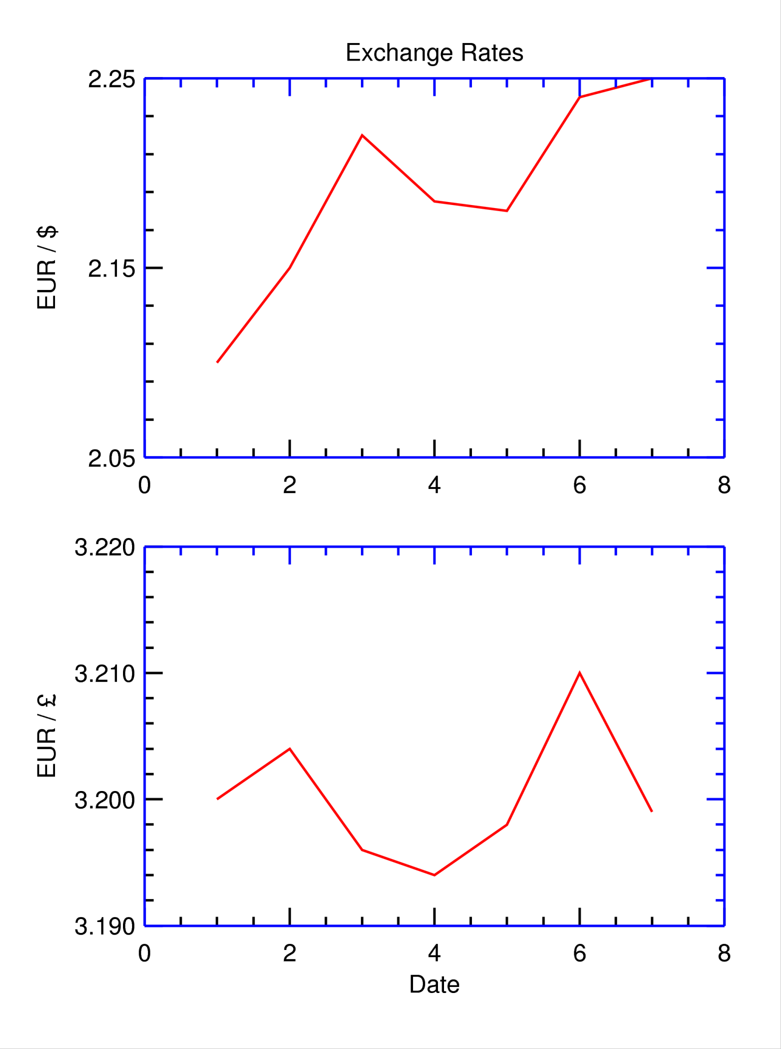
Colour can be very helpful, but don't forget most journals are in black & white!

Root:

```
// Set colour palette
Int_t *colors = 0;
gStyle->SetPalette(1, colors);
```

Morals

- Presentation should be:
  - Simple
  - Clear
  - True
- Do
  - Indicate suppressed 0
  - Label the axes
  - Give units and binning
  - Make sure scale and labels are large enough (usually not!)
- Do not use style of almost all stock market plots!



Don't forget: There are lies, damn lies and statistics

Statistics

- Average values
- Spread
- Covariance and correlations
- Common distributions
  - Binomial, Poisson, Gauss
- Central limit theorem
- Weighted mean
- PDG averaging
- Combining errors
- Error propagation
- Systematic errors

Mean

- Set of unbinned data (measurements)

  $\{ x_1, x_2, \ldots, x_N \}$

  Mean is: f is any function of x

  e.g. $f = ax$, $f = x^2$

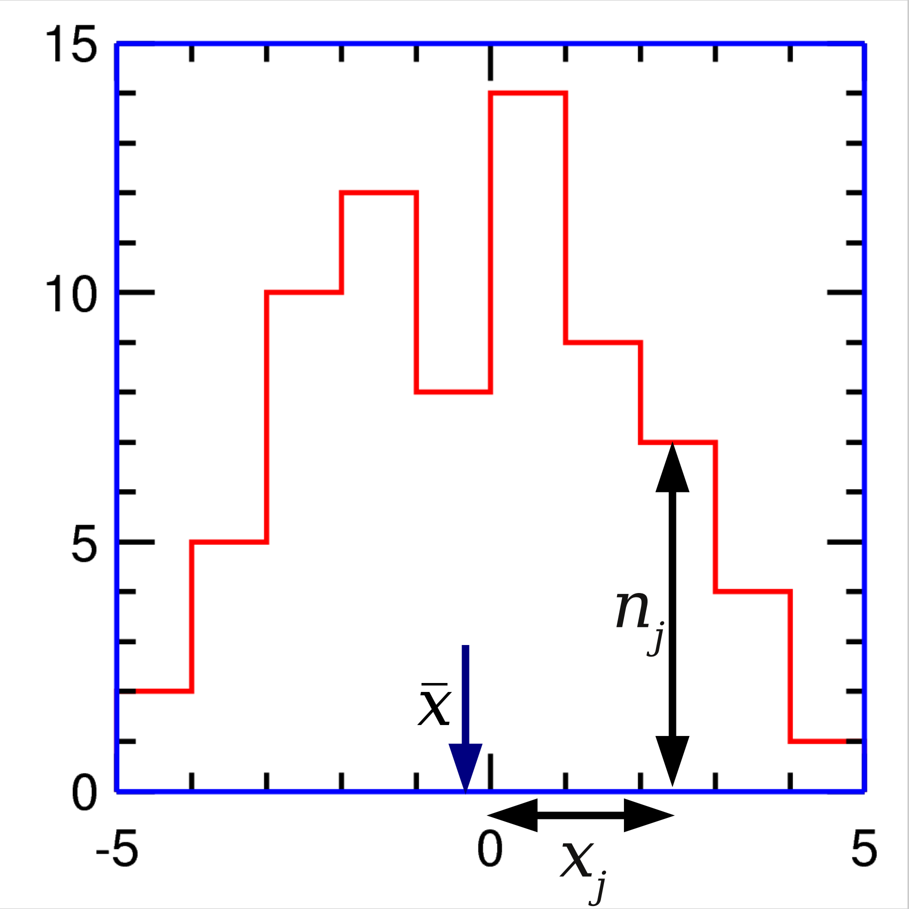
$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\bar{f} = \frac{1}{N} \sum_{i=1}^{N} f(x_i)$$

- For binned data

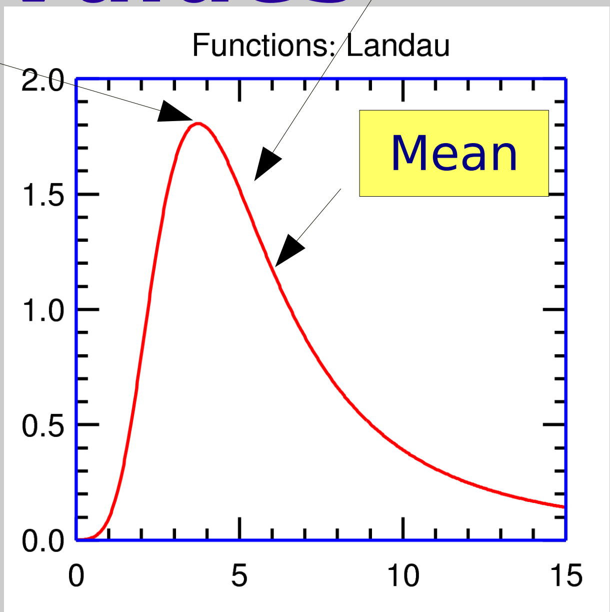
$$\bar{x} = \frac{1}{N} \sum_{j=1}^{Nbin} n_j x_j \qquad \bar{f} = \frac{1}{N} \sum_{j=1}^{Nbin} n_j f(x_j)$$

  - n_j entries in bin
  - x_j bin centre



universitätbonn

Alternative Average Values



- Mode - most probable value
  - Fine for high statistics;
    fluctuations large for low statistics
  - Well suited for skew distributions,
    e.g. Landau
    (charge deposited in thin material)
- Median - half data points are below and half above
  - Odd # measurements? – middle value
  - Even # measurements? – arithmetic mean of central 2 values
  - Binned data? - centre of bin
    for which < ½ data below and < ½ data above
- Geometric mean
- Harmonic mean

Rarely used in particle physics

Characterising the Spread

- Variance:

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

Iterative formula for evaluating spread

$$V_N = \frac{(N-1)}{N}V_{N-1} + \frac{(x - \bar{x}_{N-1})^2}{N}$$

- Standard Deviation

$$\sigma \equiv \sqrt{V(x)} = \sqrt{\overline{x^2} - \bar{x}^2}$$

Be careful of computer precision. Often good to take a rough estimate of the mean

$$\overline{(x - x_0)^2} = \overline{(x - \bar{x})^2} + (\bar{x} - x_0)^2$$

R.M.S. and FWHM + Higher Orders

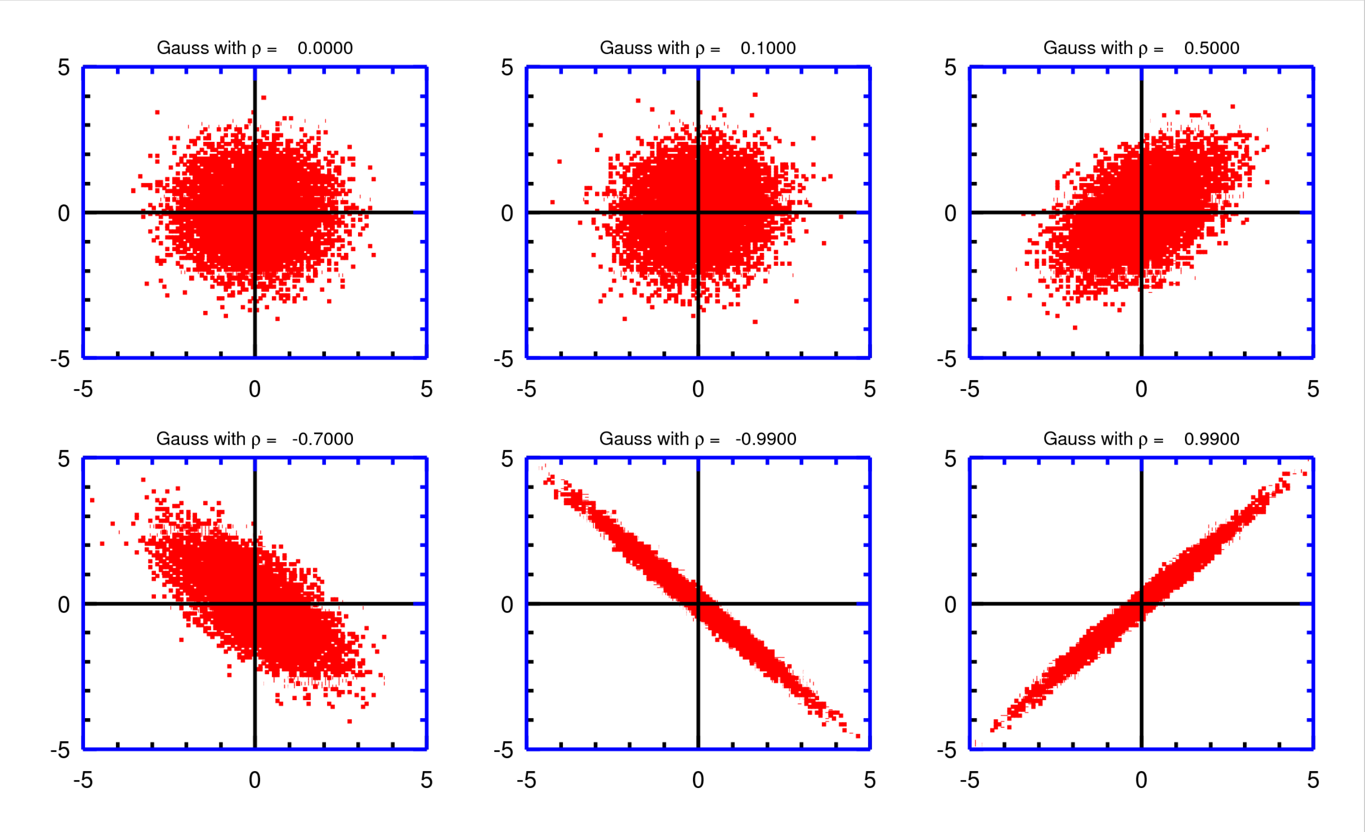
- R.M.S. = Standard deviation for above definition
- FWHM (full width half maximum)
  - Useful for asymmetric distributions or ones with long tails
  - Problems with fluctuations with low statistics
  - For a Gaussian: FWHM = 2.35 σ
- Skewness (tests for asymmetry)

$$\gamma(x) = \frac{1}{N\sigma^3}\sum_i \left(x_i - \bar{x}\right)^3 = \frac{1}{\sigma^3}\overline{\left(x - \bar{x}\right)^3}$$

- Moments:
  - rth moment $\frac{1}{N}\sum_i x_i^r$ rth central moment $\frac{1}{N}\sum_i (x_i - \bar{x})^r$

Several Variables

- Know how to plot 2 variables
- Mean and spread of each can also be calculated
- How do we characterise dependence on each other?



Gaussians all have same σ

Covariance and Correlation

- Data sample with pairs of variables:
  $\{ (x_1, y_1), (x_2, y_2), .., (x_N, y_N) \}$
- Calculate $\bar{x}, \bar{y}, V(x), V(y)$ as before
- Covariance tells you dependence on each other:

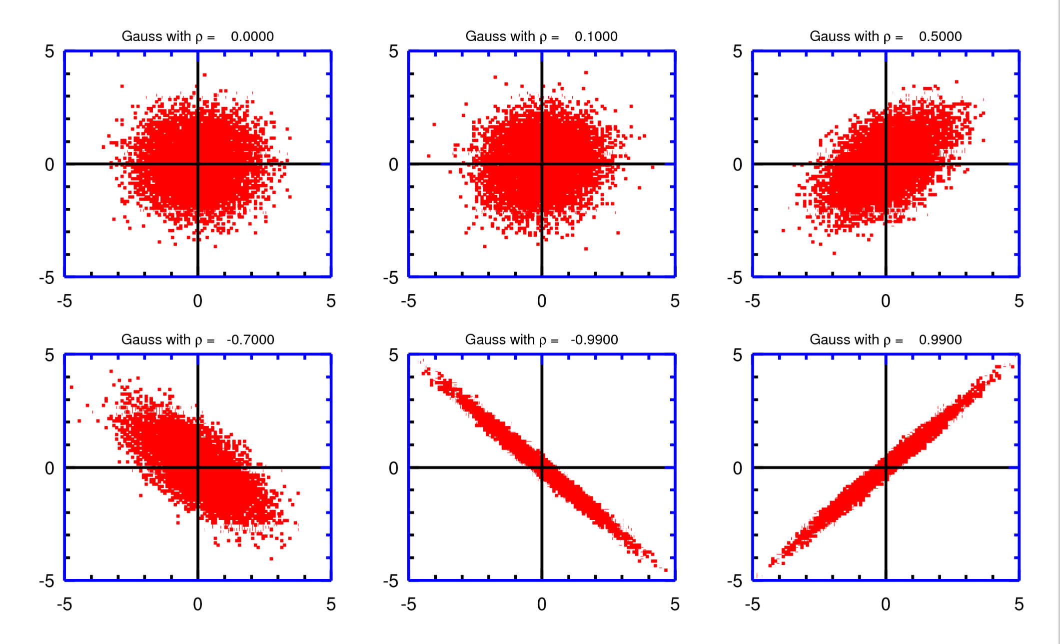
$$
\begin{aligned}
cov(x,y) &= \frac{1}{N}\sum_i (x_i - \bar{x})(y_i - \bar{y}) \\
&= \frac{1}{N}\sum_i x_i y_i - \frac{\sum x_i}{N}\frac{\sum y_i}{N} \\
&= \overline{x\,y} - \bar{x}\cdot\bar{y}
\end{aligned}
$$

- If $\overline{x\,y} = \bar{x}\cdot\bar{y}$ variables are independent

Covariance and Correlation

$$\rho = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y}$$

- Covariance is generalisation of variance
- Carries dimensions
  - Scale it by standard deviation
- Correlation coefficient, ρ
  - $-1 \le \rho \le 1$
    - -1: 100% anticorrelated,
      e.g. weight/stamina,
      BR for 2 decay channe
    - 0: uncorrelated
      e.g. height/IQ
    - +1: 100% correlated
      e.g. height/weight
  - Independent of scale &
    a shift in zero point



If x and y correlated
their variances are also affected

Covariance Matrix

- One measurement has n elements with values:

  $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$

- Define covariance between each pair:

  $$cov\left(x_{(i)}, x_{(j)}\right) = \overline{x_{(i)}, x_{(j)}} - \overline{x_{(i)}} \cdot \overline{x_{(j)}}$$

- These form elements of an n x n symmetric matrix:

  $$V_{ij} = cov\left(x_{(i)}, x_{(j)}\right)$$

- Correlation matrix is dimensionless form of covariance matrix:

  $$V_{ij} = \frac{cov\left(x_{(i)}, x_{(j)}\right)}{\sigma_i \sigma_j}$$

Deriving a Distribution

- Throw a coin in air 4 times, probability to get $4,3,2,1,0$ heads:
- Expected probability distribution:
- Do exercise by hand or on a computer (see next slide)
- As number of measurements increases observation closer and closer to expectation
- Law of large numbers:

$$P(4) = \frac{1}{16}$$
$$P(3) = \frac{1}{4}$$
$$P(2) = \frac{3}{8}$$
$$P(1) = \frac{1}{4}$$
$$P(0) = \frac{1}{16}$$

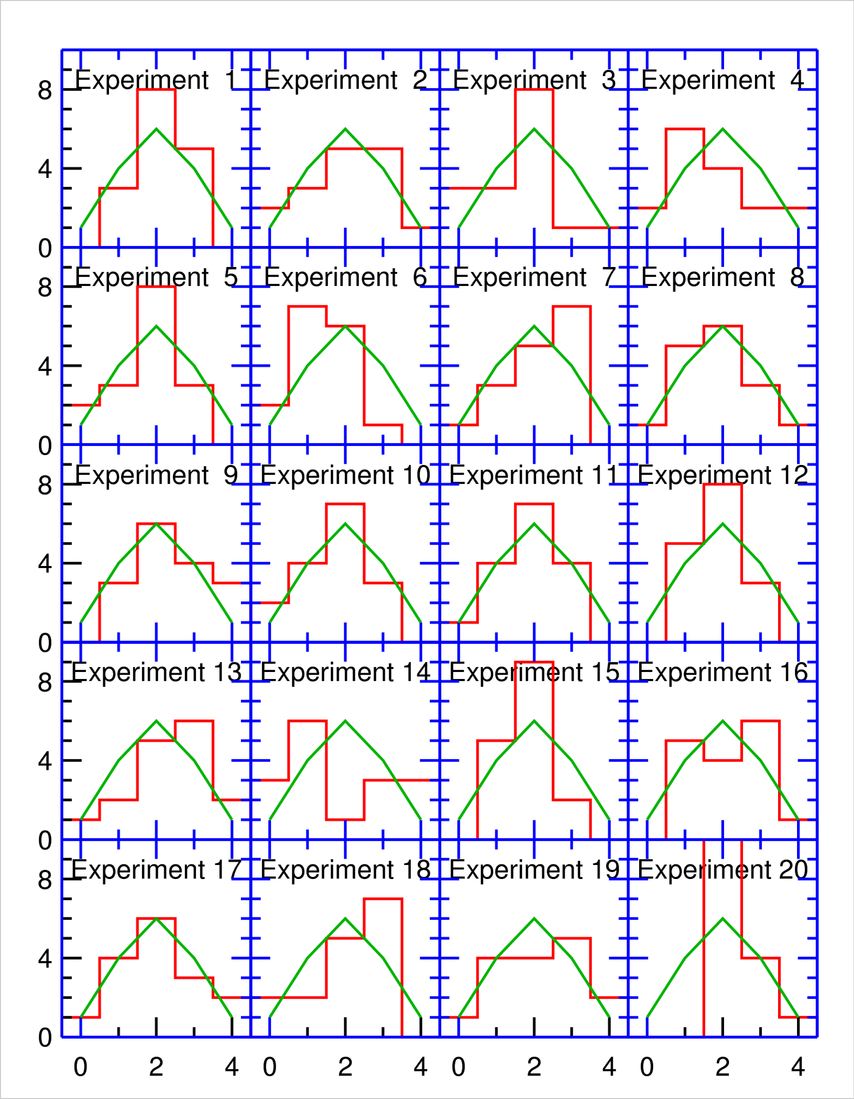
$$\lim_{N\to\infty} N(r) = N \cdot P(r)$$

Observed frequency distribution

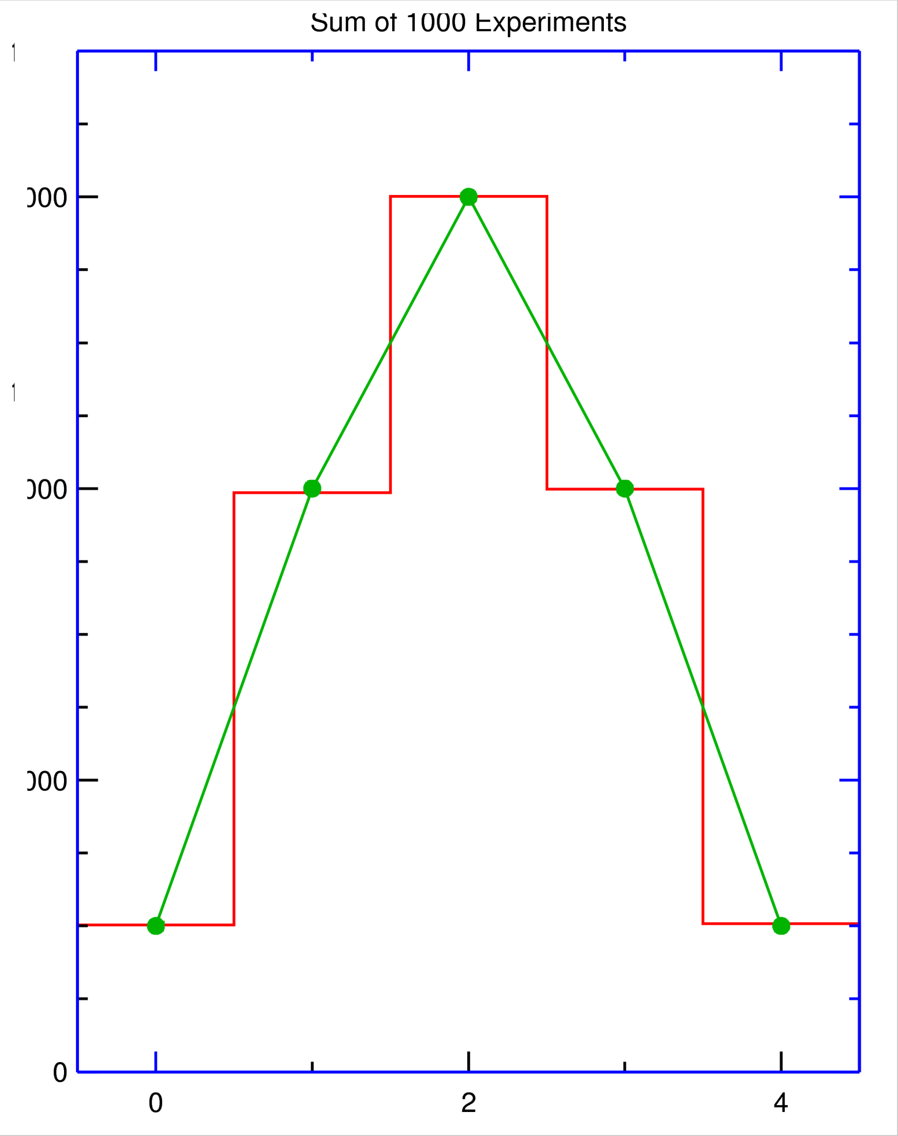
Sample

Expected frequency distribution

Frequency Distributions



Number of Heads



universitätbonn

Expectation Value

- Expectation value (as for histogram)

$$\langle r \rangle = \sum_r r P(r)$$

- Law of large numbers:

$$\lim_{N \to \infty} \bar{f} = \langle f \rangle$$

- Properties:
  - Add:

$$\langle f+g \rangle = \sum (f+g) P(r) = \sum f P(r) + \sum g P(r) = \langle f \rangle + \langle g \rangle$$

  - Multiply:

$$\langle f g \rangle \neq \langle f \rangle \langle g \rangle$$

(unless f and g are independent)

Probability Density Function

- Continuous variables
  - $P(x)\,dx$ is probability to get a value between x and $(x + dx)$
- Total probability must be 1 $$\int_{-\infty}^{+\infty} P(x)\,dx = 1$$
- For continuous distributions $$\langle x \rangle = \int_{-\infty}^{+\infty} x\,P(x)\,dx$$ $$\langle f \rangle = \int_{-\infty}^{+\infty} f(x)\,P(x)\,dx$$

Measurement	→ Mean	$\bar{x}$
Theoretical distribution	→ Expectation value	$\langle r \rangle$

Binomial Distribution

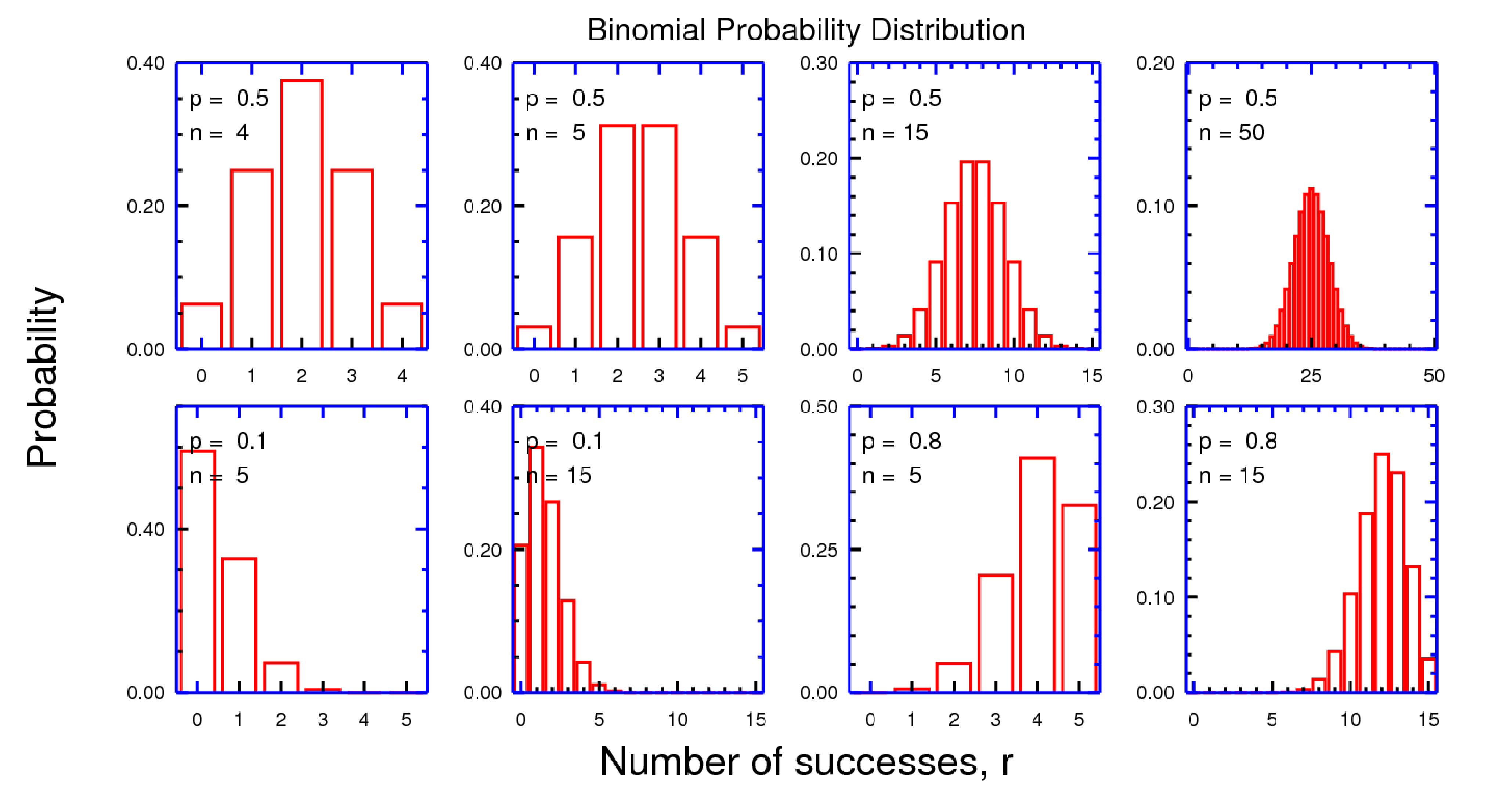
- Processes where result can be one of two values, e.g. tossing coin, detector channel fires or not
- Prob. success = p, Prob. Failure = $(1-p) = q$
- n trials, prob. for r successes and $(n-r)$ failures?
  - Number of ways to select r from n is: $n! / r!\,(n-r)!$
  - r successes with prob. p;
    $(n-r)$ failures with prob. $(1-p)$
    $\Rightarrow$ total prob. $p^r(1-p)^{(n-r)}$
- Properties:

$$P(r;p,n) = p^r (1-p)^{(n-r)} \frac{n!}{r!(n-r)!}$$

$$\begin{aligned} \langle r \rangle &= np \\ V &= np(1-p) \\ \sigma &= \sqrt{np(1-p)} \end{aligned}$$

Convention:
Before ";" variable of interest
After ";" dependencies

Examples


Binomial Probability Distribution
Probability
Number of successes, r
p = 0.5
n = 4
p = 0.5
n = 5
p = 0.5
n = 15
p = 0.5
n = 50
p = 0.1
n = 5
p = 0.1
n = 15
p = 0.8
n = 5
p = 0.8
n = 15

Detector Efficiency

- Single layer efficiency in design 95%
- Need 3 points to reconstruct track
- Track reconstruction prob. with:
  - **3 layers**: $P(3;0.90,3)$: $(0.95)^3 =$ **0.857**
  - **4 layers**: $P(3;0.90,4) +$ $P(4;0.90,4)$ $4(0.95)^3(0.05)^1 + (0.95)^4 =$ **0.986**
  - **5 layers**: $P(3;0.90,5) + \ldots$ $10(0.95)^3(0.05)^2 +$ $5(0.95)^4(0.05)^1 + (0.95)^5 =$ **0.999**

- What if real detector efficiency is only 90%?
  - 3 layers: 0.729
  - 4 layers: 0.948
  - 5 layers: 0.991
- Moral:
  - Redundancy is very important!
  - Always consider effect of lower than expected efficiency + dead channels when designing detector

Measuring Detector Efficiency

- Prob. to get a hit in all 5 detector layers is: $P(5;\, \mathrm{p},5) = (p)^5$
  - Use this to measure and extract p
- Now calculate $P(4;\, p,5),\, P(3;\, p,5),\ldots$
  - Compare with measurements
  - If agree all layers are equally efficient and your measured value is OK
- Suppose detector has a crack covering 5% solid angle. What would we measure?

$$P(0;p,5) = 0.05$$
$$P(1,2,3,4;p,5) = 0.0$$
$$P(5;p,5) = 0.95$$

- In order to have a binomial distribution inefficiency must be randomly distributed

Warning: Do not blindly use a distribution
Think about assumptions made for a distribution to be valid

Poisson Distribution

- Discrete events, but number of trials unknown
- Derive as limit of binomial with $n \to \infty$

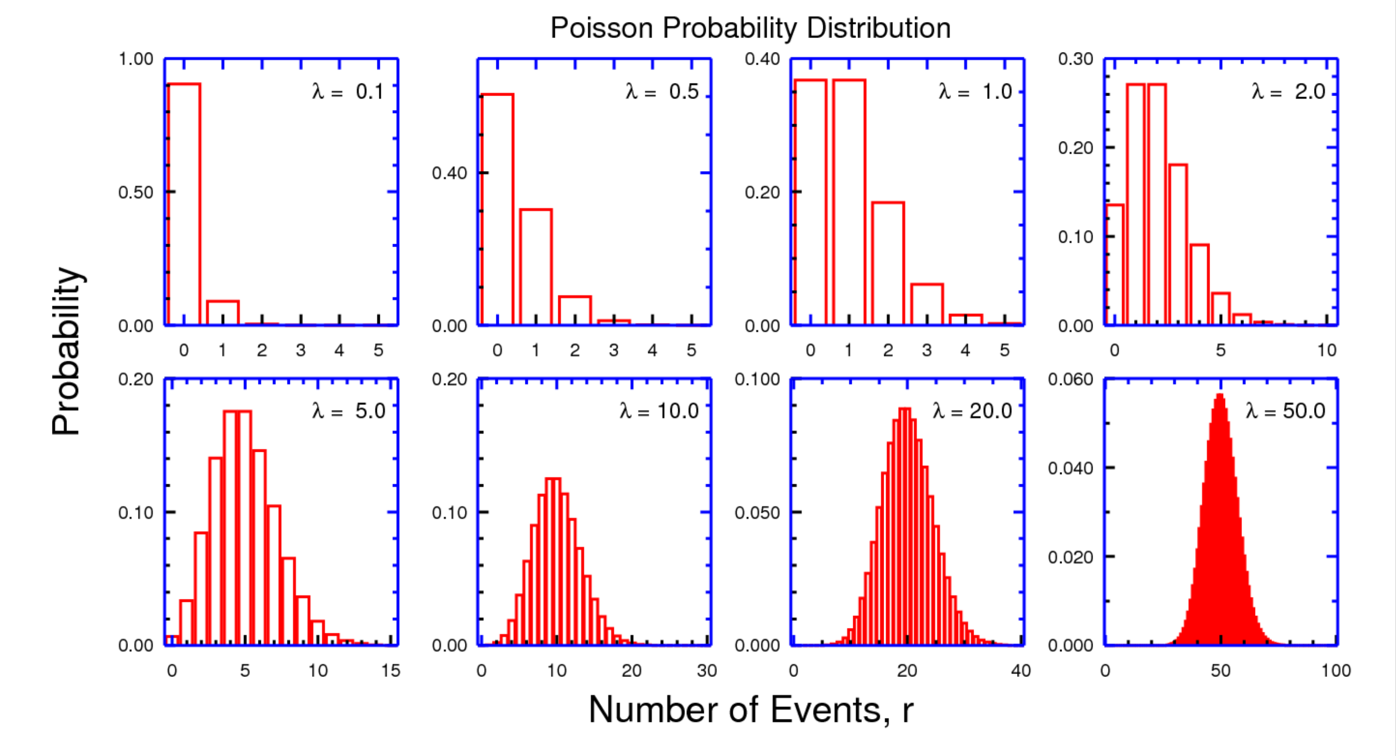
$$P(r;\lambda) = \frac{e^{-\lambda}\lambda^r}{r!}$$

- Properties:

$$\begin{aligned} \sum_{r=0}^{\infty} P(r;\lambda) &= 1 \\ \langle r \rangle &= \lambda \\ V &= \lambda \\ \sigma &= \sqrt{\lambda} \end{aligned}$$

- 2 classes of event each follow a Poisson, sum is also Poisson distributed with expectation value: $\lambda = \lambda_A + \lambda_B$

Poisson Characteristics

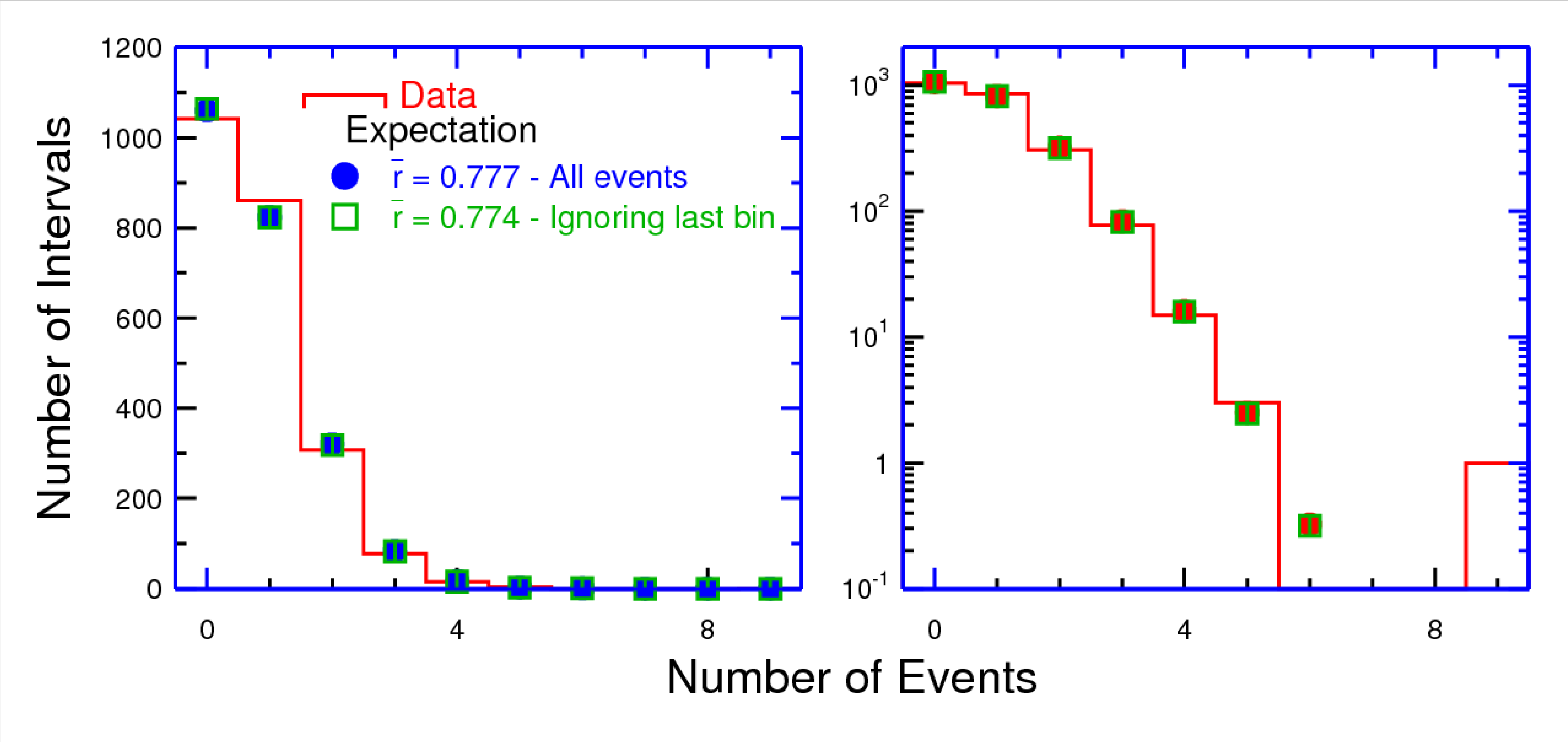


- For λ < 1, most probable value is 0!
- For λ integer, λ and (λ-1) are equally likely
  - λ is mean, but not mode!
- Poisson is wider than binomial
- Long tail to +ve values for small λ
- Shape changes significantly as λ increases

universitätbonn

Neutrinos from supernovae

No of events	0	1	2	3	4	5	6	7	8	9
No. of intervals	1042	860	307	78	15	3	0	0	0	1
Prediction	1064	823	318	82	16	2	0.3	0.03	0.003	0.0003



Gaussian Distribution

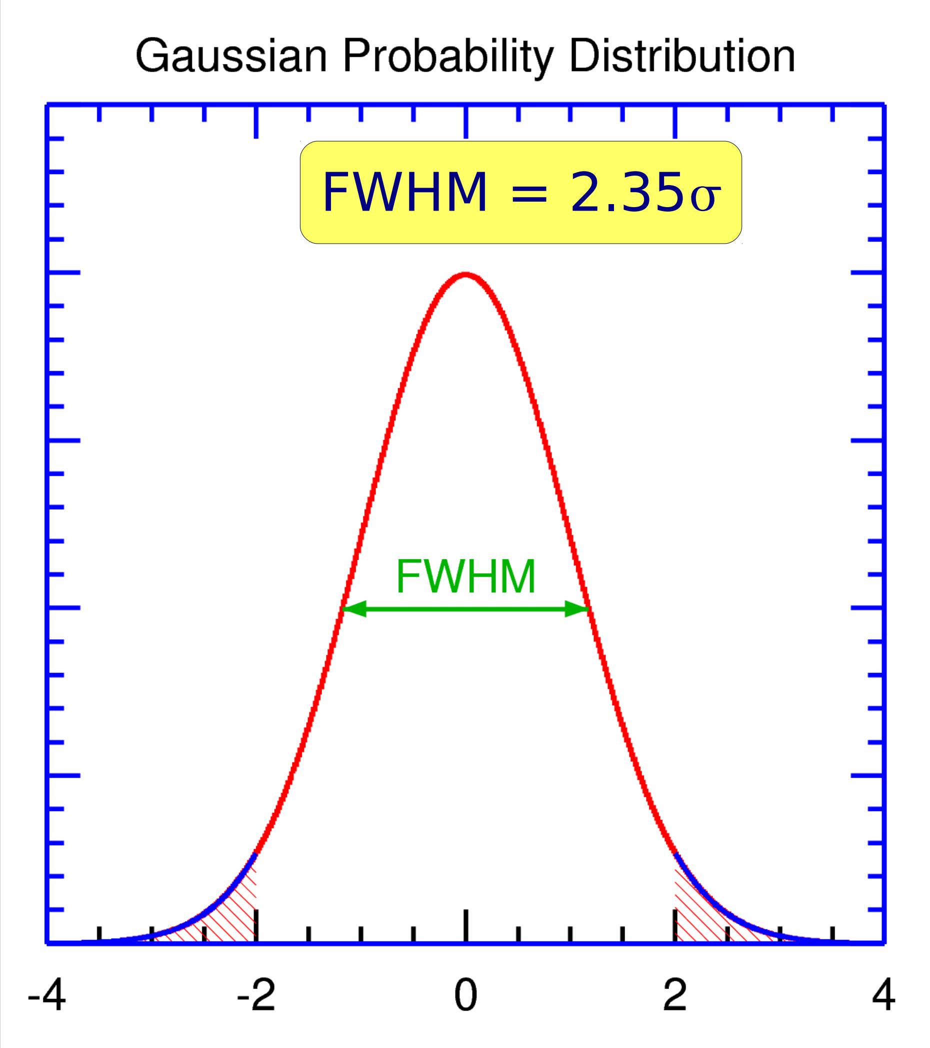
- Most common, useful and used distribution in statistics:

$$P(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

- Can always express it in a standard form:

$$P(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

where $z = (x-\mu)/\sigma$



Properties of Gaussian

- Can calculate mean and variance analytically

$$\int_{-\infty}^{+\infty} P(x;\mu,\sigma)\,dx = 1$$

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

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

- For large mean Poisson and Gaussian similar
  - How big?
  - Some say $\lambda = 5$;
    I prefer $\lambda = 10$

- Definite integrals from tables or numerical integration

68.3%	of area in range	$\mu \pm \sigma$
95.5%	of area in range	$\mu \pm 2\sigma$
99.7%	of area in range	$\mu \pm 3\sigma$
90%	of area in range	$\mu \pm 1.65\sigma$
95%	of area in range	$\mu \pm 1.96\sigma$

- Get one-sided errors from two-sided value for which double the area lies outside

Uniform Distribution

- Express as probability density:

$$P(x) = \frac{1}{(b-a)} \quad \text{for } a \leqslant x \leqslant b$$
$$= 0 \quad \text{elsewhere}$$

- Evaluate variance

$$V(x) = \int_{-\infty}^{+\infty} (x-\bar{x})^2 P(x)\,dx$$
$$= \frac{1}{12}(b-a)^2$$

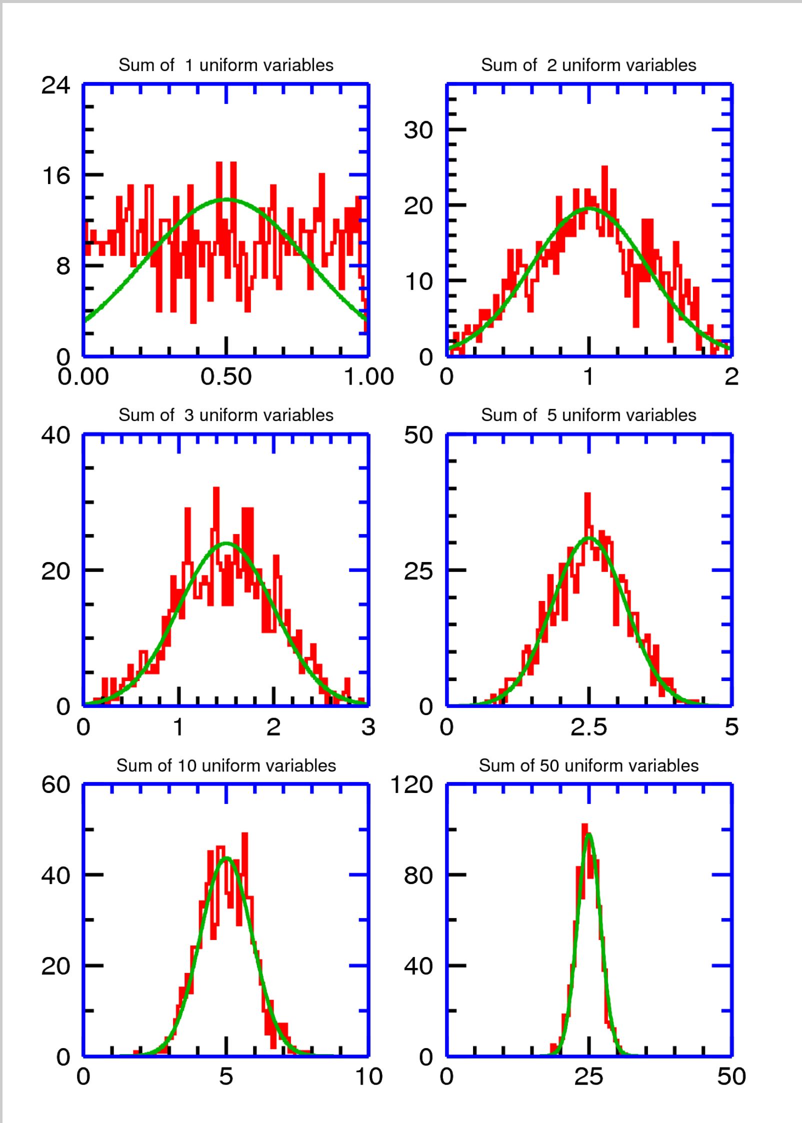
- Often occur in detectors.
  - Particle went through, but you do not know where.
  - Resolution is size of detector / √12

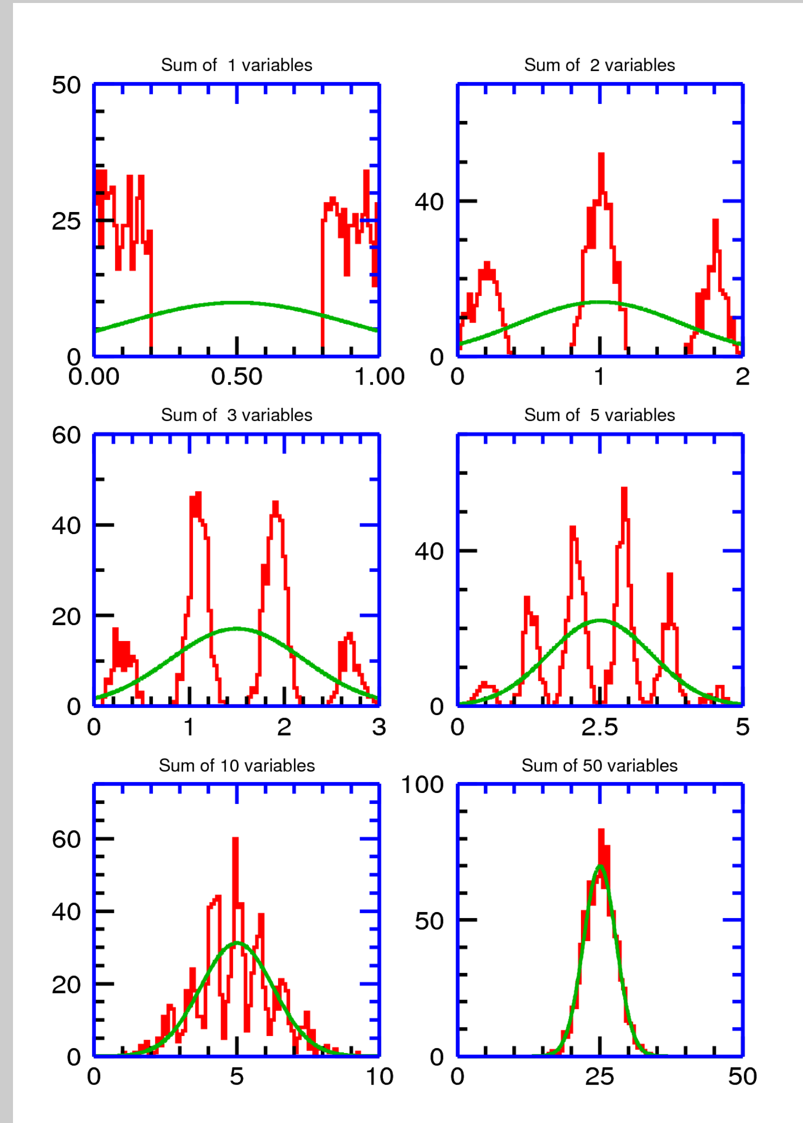
Central Limit Theorem

- Take the sum, X, of N *independent* variables x_i, where $i=1,2,3,\ldots,n$
- Each x_i is taken from a distribution with mean μ_i and variance V_i (or σ_i^2)
- Distribution of X has following characteristics:
  - Expectation value: $\langle X \rangle = \sum_{i=1}^{N} \mu_i$
  - Variance: $\langle V(X) \rangle = \sum_{i=1}^{N} V_i = \sum_{i=1}^{N} \sigma_i^2$
  - Tends to a Gaussian as $N \to \infty$

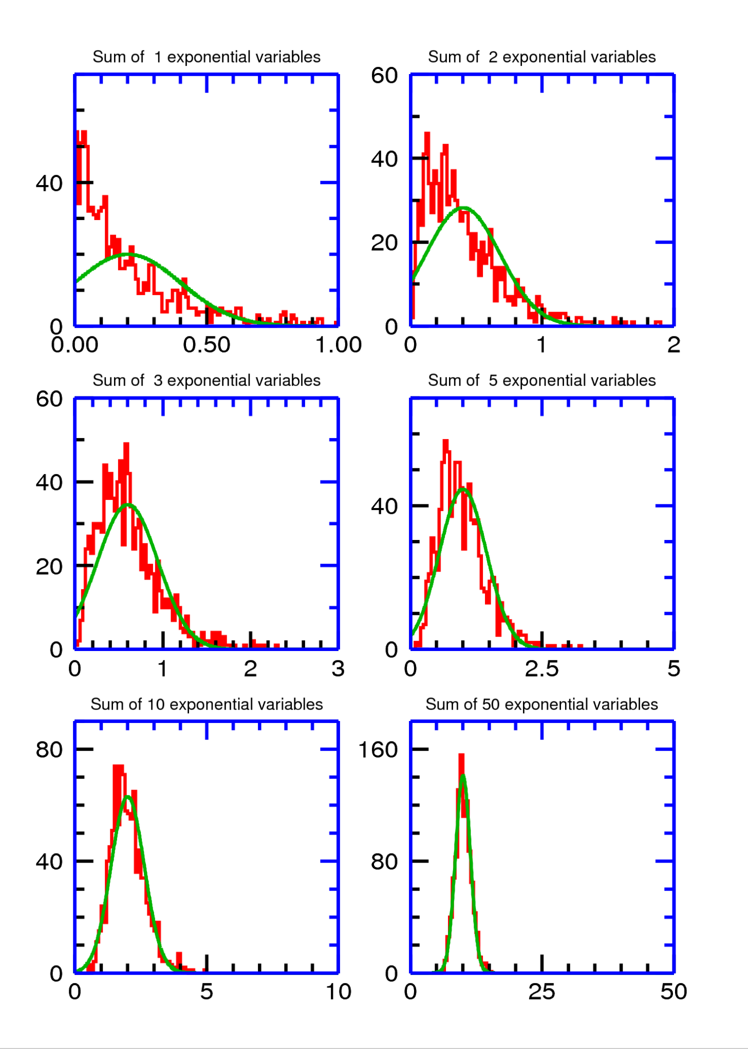
This is why Gaussian is so important!

The CLT at Work





The CLT at Work



- See that Gaussian really appears!
- For it to work important that $V \gg V_i$

Measurement with Several Error Sources

- Measured value:

$$x_i = \langle x_i \rangle + \delta x_i^{(1)} + \delta x_i^{(2)} + \delta x_i^{(3)} + \delta x_i^{(N)}$$

- For each error source

$$\langle \delta x_i^{(k)} \rangle = 0$$

where $\sigma_i^{(k)}$ is the measurement error due to source k

- CLT says that x_i is Gaussian distributed around $\langle x_i \rangle$ with variance given by sum of individual variances

Applications of CLT

- Repeated measurements:
  - Expected value always the same: μ
  - Variance always the same: σ

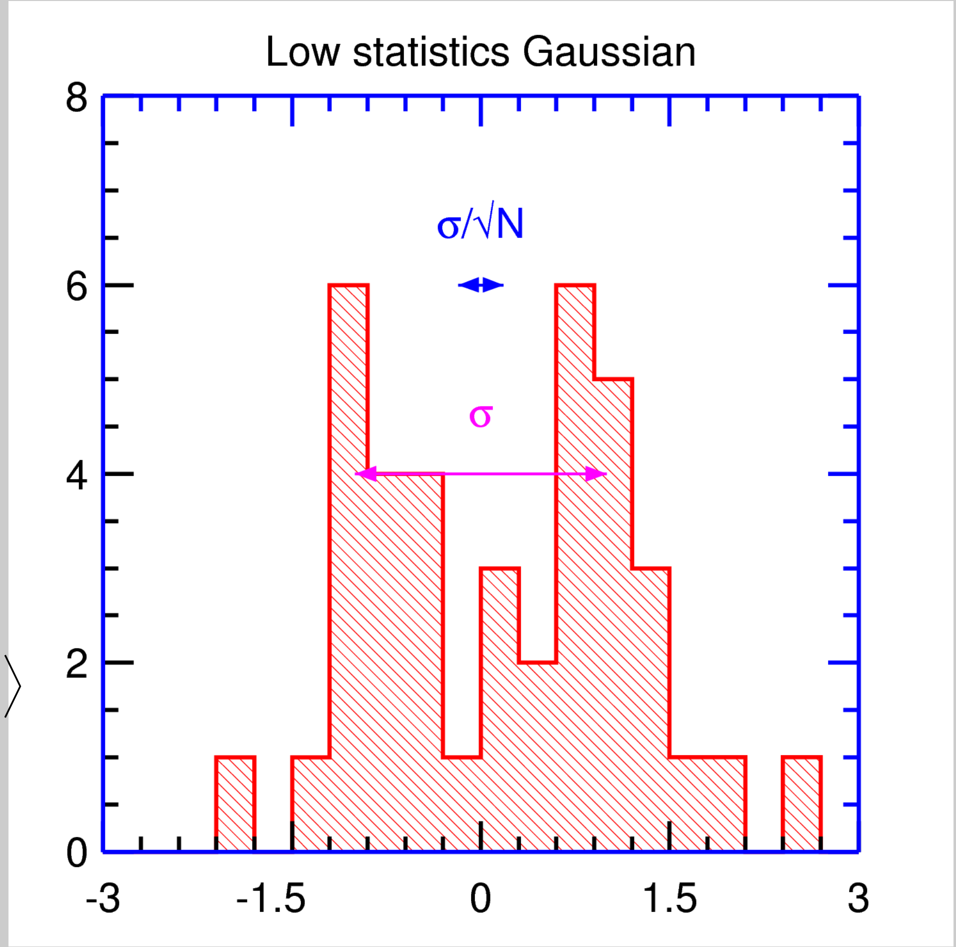
$$\langle X \rangle = \sum \mu = N\mu$$

$$\bar{x} = X/N$$

$$\langle \bar{x} \rangle = \mu$$

$$V(\bar{x}) = \frac{1}{N^2} \sum V_i = \frac{\sigma^2}{N}$$

- Expected value of average $\langle \bar{x} \rangle$
- Measurement differs from expectation by $V(\bar{x})$
- With more measurements get closer and closer to true value as a function of $1/\sqrt{N}$



universitätbonn

Weighted Mean

- How do you deal with measurements of same quantity that each have a different error?
- e.g. measure speed with 2 different radar devices
  - 4 measurements with accuracy of ±4 ms^{-1} would give an error on average of ±2 ms^{-1}
  - Give measurement with accuracy ±2 ms^{-1} 4 times the weight of measurement with accuracy ±4 ms^{-1}
- General recipe: Each measurement is given a weight:

$$\bar{x} = \frac{\sum_i x_i^2/\sigma_i^2}{\sum_i 1/\sigma_i^2}$$

$$V(\bar{x}) = \frac{1}{\sum_i 1/\sigma_i^2}$$

Are You Allowed to Average?

$$v_1 = 67 \pm 4\,\mathrm{ms}^{-1}$$
$$v_2 = 53 \pm 2\,\mathrm{ms}^{-1}$$
$$\bar{v} = 55.8 \pm 1.8\,\mathrm{ms}^{-1}$$

- Is this reasonable??
  - No, neither of the measurements is within 1σ of the mean!
- Throw out one of the results?
- Remember:
  - 2/3 measurements should be within 1σ; 1/3 should be outside
  - If more than 5% outside 2σ start getting suspicious

- Don't forget the ozone hole:

> So unexpected was the hole that for several years computers analysing ozone data had systematically thrown out readings that should have pointed to its growth
>
> New Scientist, 31 March 1988

PDG Recipe

- What is a reasonable estimate of error on mean when measurements vary by more than their errors indicate they should?
- First calculate weighted mean
- Then calculate χ^2:

$$\chi^2 = \sum_i \frac{(x_i - \bar{x})^2}{\sigma_i^2}$$

- If we expect each measurement to differ from its mean by about 1σ, expect $\chi^2 \approx N$
  - $\chi^2/(N\text{-}1) < 1$: Everything OK, use simple weighted average
  - $\chi^2/(N\text{-}1) \gg 1$: Tough. Calculate average and guess error or do not average
  - $\chi^2/(N\text{-}1) > 1$: Some or all errors underestimated?

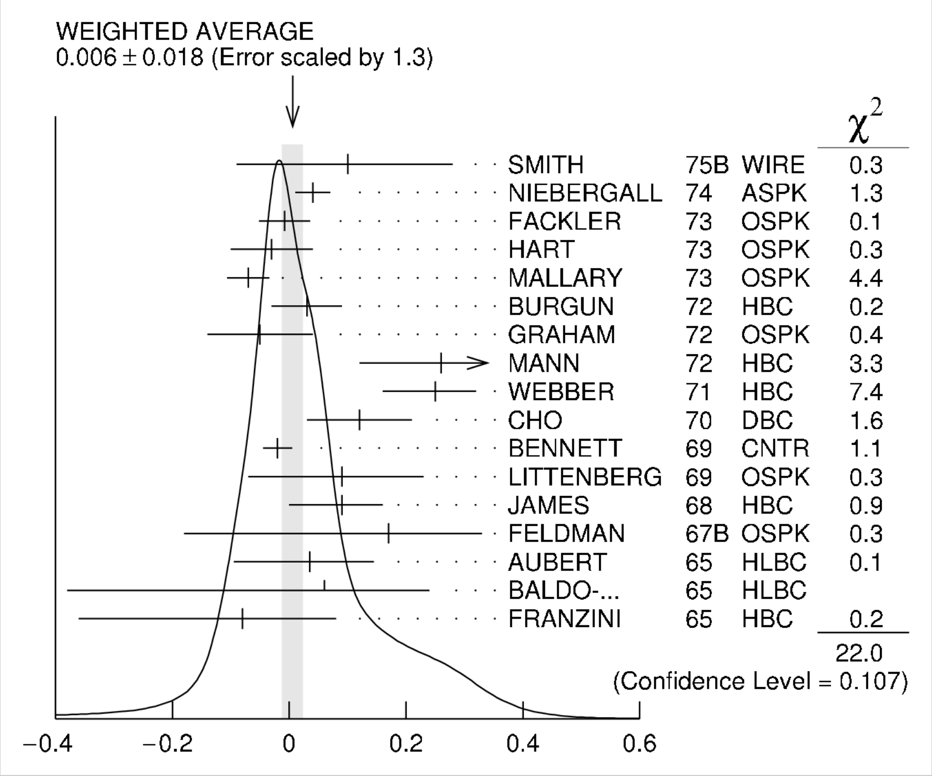
PDG Recipe for $\chi^2/(N\text{-}1) > 1$

- Scale all errors by a factor:

$$S=\sqrt{\chi^2/(N-1)}$$

- What if we have small and large errors to combine?
  - Evaluate S using only measurements with small errors
  - Only use those errors for which $\sigma_i < \sigma_0 = 3\sqrt{N}\,\sigma_{\bar{x}}$
    - Large errors do not contribute to $\bar{x}, \sigma_{\bar{x}}$ but can make significant contribution to S
- Correlations between measurements ignored here – can be taken into account

Ideogram can be helpful



Measurement represented by Gaussian with mean x_i and error σ_{xi}, area ∝ $1/\sigma_{xi}$

Combination of Errors

- Measure a quantity, but want a physical parameter that is a function of that quantity
- Simplest case f linear function of x (variance $V(x)$, error σ_x)

$$f = ax + b$$

$$\begin{aligned} V(f) &= \langle f^2\rangle - \langle f\rangle^2 \\ &= \langle (ax+b)^2\rangle - \langle (ax+b)\rangle^2 \\ &= a^2\langle x^2\rangle + 2\mathrm{ab}\langle x\rangle + b^2 \\ &\quad - a^2\langle x\rangle^2 - 2\mathrm{ab}\langle x\rangle - b^2 \\ &= a^2\left(\langle x^2\rangle - \langle x\rangle^2\right) \\ &= a^2 V(x) \\ \sigma_f &= |a|\sigma_x \end{aligned}$$

- If f is an arbitrary function of x, make a Taylor expansion about point x_0:

$$f(x) \approx f(x_0) + (x - x_0)\left(\frac{df}{dx}\right)_{x=x_0}$$

$$\begin{aligned} V(f) &\approx \left(\frac{df}{dx}\right)^2 V(x) \\ \sigma(f) &\approx \left|\frac{df}{dx}\right| \sigma(x) \end{aligned}$$

- Taylor expansion has to be valid, i.e. *df/dx* constant within 2-3σ of x_0

Functions of Two or More Variables

- Linear function:

$$\begin{aligned} f &= ax+by+c \\ V(f) &= a^2\left(\langle x^2\rangle-\langle x\rangle^2\right)+b^2\left(\langle y^2\rangle-\langle y\rangle^2\right)+2ab\left(\langle xy\rangle-\langle x\rangle\langle y\rangle\right) \\ &= a^2V(x)+b^2V(y)+2ab\,cov(x,y) \end{aligned}$$

- General function – Taylor expansion:

$$\begin{aligned} V(f) &= \left(\frac{df}{dx}\right)^2 V(x)+\left(\frac{df}{dy}\right)^2 V(y)+2\left(\frac{df}{dx}\right)\left(\frac{df}{dy}\right)cov(x,y) \\ \sigma_f^2 &= \left(\frac{df}{dx}\right)^2 \sigma_x^2+\left(\frac{df}{dy}\right)^2 \sigma_y^2+2\left(\frac{df}{dx}\right)\left(\frac{df}{dy}\right)\rho\sigma_x\sigma_y \end{aligned}$$

Gaussian Error Propagation

- If x and y are independent:

$$V(f) = \left(\frac{df}{dx}\right)^2 V(x) + \left(\frac{df}{dy}\right)^2 V(y)$$

$$\sigma_f^2 = \left(\frac{df}{dx}\right)^2 \sigma_x^2 + \left(\frac{df}{dy}\right)^2 \sigma_y^2 + \left(\frac{df}{dz}\right)^2 \sigma_z^2$$

- One example: A, B, θ independent measurements with errors:

$$y = A \sin\theta + B \cos\theta$$

$$\sigma_y^2 = \sin^2\theta\, \sigma_A^2 + \cos^2\theta\, \sigma_B^2 + (A\cos\theta - B\, sin\,\theta) \sigma_\theta^2$$

Standard Formulae

$$f = x \pm y$$
$$V(f) = V(x) + V(y)$$
$$\sigma_f^2 = \sigma_x^2 + \sigma_y^2$$

$$f = x^2$$
$$V(f) = (2x)^2 V(x)$$
$$\left(\frac{\sigma_f}{f}\right) = 2\left(\frac{\sigma_x}{x}\right)$$

$$f = 1/x^2$$
$$V(f) = \left(-\frac{2}{x^3}\right)^2 V(x)$$
$$\left(\frac{\sigma_f}{f}\right) = 2\left(\frac{\sigma_x}{x}\right)$$

$$f = \ln x$$
$$V(f) = \left(\frac{1}{x}\right)^2 V(x)$$
$$\sigma_f = \sigma_x / x$$

$$f = x\,y$$
$$V(f) = y^2 V(x) + x^2 V(y)$$
$$\left(\frac{\sigma_f}{f}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2$$

$$f = x/y$$
$$V(f) = \left(\frac{1}{y}\right)^2 V(x) + \left(-\frac{x}{y^2}\right)^2 V(y)$$
$$\left(\frac{\sigma_f}{f}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2$$

Efficiency

- Particle through detector
  - Signal or no signal → Binomial

$$\begin{aligned}\epsilon &= n/N \\ V(\epsilon) &= N\epsilon(1-\epsilon)/N^2 \\ \sigma_\epsilon &= \sqrt{\frac{\epsilon(1-\epsilon)}{N}}\end{aligned}$$

- With error propagation:
  - Need independent variables $N_1 = n$, $N_1 + N_2 = N$

$$\begin{aligned}\epsilon &= \frac{N_1}{N_1+N_2} \\ V(\epsilon) &= \frac{1}{(N_1+N_2)^4}\left(N_2^2 V(N_1)+N_1^2 V(N_2)\right)\end{aligned}$$

- What are $V(N_1)$, $V(N_2)$?
  - Expect $N_1 = \epsilon N$ signals
  - For error propagation to work need error on N_1 “small”, i.e. N_1 “large”
  - For large N_1 binomial → Poisson with mean ϵN and variance $\epsilon N = N_1$
  - For N_2 mean $(1-\epsilon)N$ and variance $(1-\epsilon)N = N_2$

$$\begin{aligned}V(\epsilon) &= \frac{1}{(N_1+N_2)^4}\left(N_2^2 N_1+N_1^2 N_2\right) \\ &= \frac{\epsilon(1-\epsilon)}{N}\end{aligned}$$

Efficiency

- What is the variance if the measured efficiency is 0 or 1?
  - Following formulae above set it to 0??
  - Mn_Fit follows formula originally built into program MULFIT from Paul Avery for all efficiencies:

$$V(\epsilon) = \frac{(n+1)(N-n+1)}{(N+3)(N+2)^2}$$

    - This form make some assumptions about the “prior” - see Bayesian statistics and leads to a biased efficiency estimator, but it has a nice behaviour for $n=0$ and $n=N$!
  - If $\epsilon = 0$ or $\epsilon = 1$ add unbiased estimate of n/N to error:
    - Claim is that this gives a better estimate of the confidence interval

$$\sigma(\epsilon) = \sqrt{\frac{(n+1)(N-n+1)}{(N+3)(N+2)^2} + \frac{1}{N+2}}$$

universitätbonn

Gaussian Error Propagation General Case

- n variables:

  $$x_{(1)}, x_{(2)}, x_{(3)}, \ldots, x_{(n)}$$

- Covariance for 2 variables in a sample defined as:

  $$\begin{aligned} cov(x_{(i)}, x_{(j)}) &= \overline{\left(x_{(i)} - \overline{x_{(i)}}\right)\left(x_{(j)} - \overline{x_{(j)}}\right)} \\ &= \overline{x_{(i)} x_{(j)}} - \overline{x_{(i)}}\ \overline{x_{(j)}} \end{aligned}$$

- Consider 2 variables in a joint PDF which describes probability to measure values

  $$P(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$$

- Expectation value for each variable given by

  $$\langle x_{(1)} \rangle, \langle x_{(2)} \rangle, \ldots, \langle x_{(n)} \rangle$$

  $$\mu_i \equiv \langle x_{(i)} \rangle$$

  and $\mu_{(1)}, \mu_{(2)}, \ldots, \mu_{(n)}$

General Case

- Define covariance as

$$cov(x_{(i)}, x_{(j)}) = \langle (x_{(i)} - \mu_i)(x_{(j)} - \mu_j) \rangle$$
$$= \langle x_{(i)} x_{(j)} \rangle - \mu_i \mu_j$$

- Each of these is one element of the covariance (error) matrix:

$$V_{ij} = cov(x_{(i)}, x_{(j)})$$

- Diagonal elements are variances.
- Correlation matrix is dimensionless equivalent

$$\rho_{ij} = \frac{cov(x_{(i)}, x_{(j)})}{\sigma_i \sigma_j}$$

One Step Further!

- Suppose we have m functions $f_1, f_2, \ldots, f_m$ of n variables $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$
- Each $x_{(i)}$ has an associated variance, hence also f_k.
- The f_k are correlated because they share $x_{(i)}$
- Variance on f_k: $V(f_k) = \langle f^2 \rangle - \langle f \rangle^2$
- Expand f_k as a Taylor series around expectation values for $x_{(i)}$

$$f_k \approx f_k(\mu_{1,} \mu_{2,} \ldots, \mu_n) + \left(\frac{\partial f_k}{\partial x_{(1)}}\right)(x_{(1)} - \mu_1) + \left(\frac{\partial f_k}{\partial x_{(2)}}\right)(x_{(2)} - \mu_2) + \cdots$$

- f_k are now a linear combination of the $x_{(i)}$, so can use previous variance formula:

$$V(f_k) = \sum_i \left(\frac{\partial f_k}{\partial x_{(i)}}\right)^2 V(x_{(i)}) + \sum_i \sum_{j \neq i} \left(\frac{\partial f_k}{\partial x_{(i)}}\right)\left(\frac{\partial f_k}{\partial x_{(j)}}\right) cov(x_{(i)}, x_{(j)})$$

One Step Further!

- But, f_k and f_l are also correlated, so we have to determine their covariance:

$$\langle f_k f_l \rangle - \langle f_k \rangle \langle f_l \rangle \approx \langle (x_{(1)} - \mu_1)(x_{(1)} - \mu_1) \rangle \left(\frac{\partial f_k}{\partial x_{(1)}} \right) \left(\frac{\partial f_k}{\partial x_{(1)}} \right) + \ldots + \langle (x_{(1)} - \mu_1)(x_{(2)} - \mu_2) \rangle \left(\frac{\partial f_k}{\partial x_{(1)}} \right) \left(\frac{\partial f_k}{\partial x_{(2)}} \right)$$

- Looks horrible, try a sum:

$$cov(f_k, f_l) = \sum_i \sum_j \left(\frac{\partial f_k}{\partial x_{(i)}} \right) \left(\frac{\partial f_l}{\partial x_{(j)}} \right) cov(x_{(i)}, x_{(j)})$$

- Getting better, try a matrix:

$$G_{ki} = \left(\frac{\partial f_k}{\partial x_{(i)}} \right)$$

- With $\boldsymbol{V}_x$ and $\boldsymbol{V}_f$ as error matrices for x and f

$$\boldsymbol{V}_f = \boldsymbol{G} \boldsymbol{V}_x \tilde{\boldsymbol{G}}$$

m x m

m x n

n x n

Example 1

- f is a function of 2 variables x and y, with errors σ_x, σ_y and correlation coefficient ρ

$$\boldsymbol{G} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$$

$$\boldsymbol{V_f} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \begin{pmatrix} \sigma_x^2 & \rho\,\sigma_x\sigma_y \\ \rho\,\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$$

$$V_f = \left(\frac{\partial f}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y} \right)^2 \sigma_y^2 + 2 \left(\frac{\partial f}{\partial x} \right) \left(\frac{\partial f}{\partial y} \right) \rho\,\sigma_x\sigma_y$$

Standard error propagation formula with 2 variables

universitätbonn

Example 2

- Transform track chamber measurement (r, ϕ, z) to Cartesian coordinates (x, y, z)
- $x = r\cos\phi$, $y = r\sin\phi$, no error on r (chamber construction)

$$\boldsymbol{G} = \begin{pmatrix} \cos\phi & -r\sin\phi & 0 \\ \sin\phi & +r\cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$V_{xyz} = \begin{pmatrix} \cos\phi & -r\sin\phi & 0 \\ \sin\phi & +r\cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma_\phi^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix} \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -r\sin\phi & +r\cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_\phi^2 y^2 & -\sigma_\phi^2 x y & 0 \\ -\sigma_\phi^2 x y & \sigma_\phi^2 x^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}$$

Plausible?
Close to ϕ=0 error only on y and not on x ☺
x, y correlation is -1 ☺