Introduction to Error Analysis

Part 1: the Basics

Andrei Gritsan

based on lectures by Petar Maksimović

February 1, 2010

Overview

- **Definitions**
- **Reporting results and rounding**
- **Accuracy vs precision – systematic vs statistical errors**
- **Parent distribution**
- **Mean and standard deviation**
- **Gaussian probability distribution**
- **What a "1σ error" means**

Definitions

- μ_{true}: **'true' value of the quantity x we measure**
- x_i: **observed value**
- **error on μ: difference between the observed and 'true' value, $\equiv x_i - \mu_{\mathrm{true}}$**

  All measurement have errors $\Longrightarrow$ 'true' value is unattainable
- **seek best estimate of 'true' value, μ**
- **seek best estimate of 'true' error $\equiv x_i - \mu$**

One view on reporting measurements (from the book)

- keep only one digit of precision on the error – everything else is noise

  Example: $410.5163819 \longrightarrow 4 \times 10^2$

- exception: when the first digit is 1, keep two:

  Example: $17538 \longrightarrow 1.7 \times 10^4$

- round off the final value of the measurement up to the significant digits of the errors

  Example: 87654 ± 345 kg $\longrightarrow (876 \pm 3) \times 10^2$ kg

- rounding rules:
  - 6 and above $\longrightarrow$ round up
  - 4 and below $\longrightarrow$ round down
  - 5: if the digit to the right is even round down, else round up

    (reason: reduces systematic erorrs in rounding)

A different view on rounding

From Particle Data Group (authority in particle physics):

http://pdg.lbl.gov/2009/reviews/rpp2009-rev-rpp-intro.pdf

- **between 100 and 354, we round to two significant digits**

  Example: 87654 ± 345 **kg** $\longrightarrow (876.5 \pm 3.5) \times 10^2$ **kg**

- **between 355 and 949, we round to one significant digit**

  Example: 87654 ± 365 **kg** $\longrightarrow (877 \pm 4) \times 10^2$ **kg**

- **lie between 950 and 999, we round up to 1000 and keep two significant digits**

  Example: 87654 ± 950 **kg** $\longrightarrow (87.7 \pm 1.0) \times 10^3$ **kg**

Bottom line:

Use consistent approach to rounding which is sound and accepted in the field of study, use common sense after all

Accuracy vs precision

- **Accuracy:** how close to 'true' value
- **Precision:** how well the result is determined (regardless of true value); a measure of reproducibility
- **Example:** $\mu = 30$

  - $x = 23 \pm 2$ **precise, but inacurate**
    $\Rightarrow$ $\exists$ **uncorrected biases**
    (large systematic error)

  - $x = 28 \pm 7$ **acurate, but imprecise**
    $\Rightarrow$ **subsequent measurements will scatter around** $\mu = 30$ **but cover the true value in most cases**
    (large statistical (random) error)

$\Rightarrow$ **an experiment should be both acurate and precise**

Statistical *vs.* systematic errors

- Statistical (random) errors:
  - describes by how much subsequent measurements scatter the common average value
  - if limited by instrumental error, use a better apparatus
  - if limited by statistical fluctuations, make more measurements
- Systematic errors:
  - all measurements biased in a common way
  - harder to detect:
    - faulty calibrations
    - wrong model
    - bias by observer
  - also hard to determine (no unique recipe)
  - estimated from analysis of experimental conditions and techniques
  - may be *correlated*

Parent distribution

(assume no systematic errors for now)

- **parent distribution: the probability distribution of results if the number of measurements $N \longrightarrow \infty$**
- **however, only a limited number of measurements: we observe only a sample of parent dist., a sample distribution**

$\Rightarrow$ **prob. distribution of our measurements only approaches parent dist. with $N \longrightarrow \infty$**

$\Rightarrow$ **use observed distribution to infer the parameters from the parent distribution, *e.g.*, $\mu \longrightarrow \mu_{\mathrm{true}}$ when $N \longrightarrow \infty$**

Notation

Greek: parameters of the parent distribution

Roman: experimental estimates of params of parent dist.

Mean, median, mode

- **Mean:** of experimental (sample) dist:

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

. . . of the parent dist

$$\mu \equiv \lim_{N \to \infty} \left(\frac{1}{N} \sum x_i \right)$$

mean $\equiv$ centroid $\equiv$ average

- **Median:** splits the sample in two equal parts

- **Mode:** most likely value (highest prob.density)

Variance

- **Deviation:** $d_i \equiv x_i - \mu$, **for single measurement**

- **Average deviation:**
  $\langle x_i - \mu \rangle = 0$ **by definition**
  $\alpha \equiv \langle |x_i - \mu| \rangle$
  but, absolute values are hard to deal with analytically

- **Variance:** **instead, use mean of the deviations squared:**

$$\sigma^2 \equiv \langle (x_i - \mu)^2 \rangle = \langle x^2 \rangle - \mu^2$$

$$\sigma^2 = \lim_{N \to \infty} \left(\frac{1}{N} \sum x_i^2 \right) - \mu^2$$

("mean of the squares minus the square of the mean")

Standard deviation

- **Standard deviation:** **root mean square of deviations:**

$$\sigma \equiv \sqrt{\sigma^2} = \sqrt{\langle x^2 \rangle - \mu^2}$$

associated with the 2nd moment of x_i distribution

- **Sample variance:** **replace μ by $\overline{x}$**

$$s^2 \equiv \frac{1}{N-1} \sum (x_i - \overline{x})^2$$

$N-1$ instead of N because $\overline{x}$ is obtained from the same data sample and not independently

So what are we after?

- **We want μ.**
- **Best estimate of μ is sample mean, $\overline{x} \equiv \langle x \rangle$**
- **Best estimate of the error on $\overline{x}$ (and thus on μ is square root of sample variance, $s \equiv \sqrt{s^2}$**

Weighted averages

- **$P(x_i)$ – discreete probability distribution**
- **replace $\sum x_i$ with $\sum P(x_i) x_i$ and $\sum x_i^2$ by $\sum P(x_i) x_i^2$**
- **by definition, the formulae using $\langle\rangle$ are unchanged**

Gaussian probability distribution

- **unquestionably the most useful in statistical analysis**
- **a limiting case of Binomial and Poisson distributions** *(which are more fundamental; see next week)*
- **seems to describe distributions of random observations for a large number of physical measurements**
- **so pervasive that all results of measurements are always classified as 'gaussian' or 'non-gaussian'** *(even on Wall Street)*

Meet the Gaussian

- **probability density function:**
  - **random variable** x
  - **parameters 'center'** μ **and 'width'** σ**:**

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

- **Differential probability:**

probability to observe a value in $[x, x+dx]$ **is**

$$dP_G(x;\mu,\sigma) = p_G(x;\mu,\sigma)dx$$

- **Standard Gaussian Distribution:**

  replace $(x-\mu)/\sigma$ with a new variable z:

$$p_G(z)dz = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$$

$\Rightarrow$ **got a Gaussian centered at 0 with a width of 1.**

All computers calculate Standard Gaussian first, and then 'stretch' it and shift it to make $p_G(x;\mu,\sigma)$

- **mean and standard deviation**

  By straight application of definitions:

  - **mean = μ (the 'center')**
  - **standard deviation = σ (the 'width')**

$\Rightarrow$ **This makes Gaussian so convenient!**

Interpretation of Gaussian errors

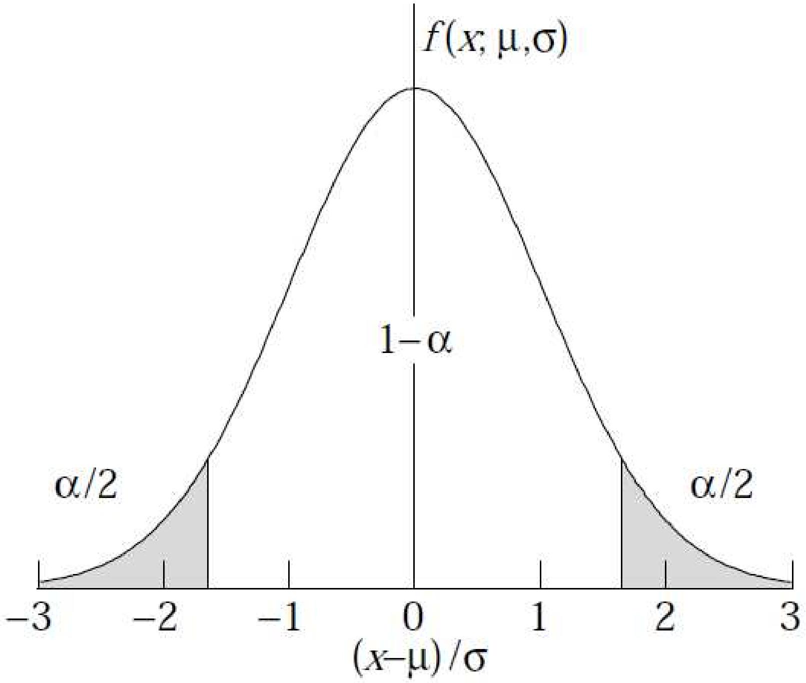
- **we measured $x = x_0 \pm \sigma_0$; what does that tell us?**
- **Standard Gaussian covers 0.683 from -1.0 to $+1.0$**

$\Rightarrow$ **the true value of x is contained by the interval $[x_0 - \sigma_0, x_0 + \sigma_0]$ 68.3% of the time!**

The Gaussian distribution coverage

Table 32.1: Area of the tails α outside $\pm\delta$ from the mean of a Gaussian distribution.

α	δ	α	δ
0.3173	1σ	0.2	1.28σ
4.55×10^{-2}	2σ	0.1	1.64σ
2.7×10^{-3}	3σ	0.05	1.96σ
6.3×10^{-5}	4σ	0.01	2.58σ
5.7×10^{-7}	5σ	0.001	3.29σ
2.0×10^{-9}	6σ	10^{-4}	3.89σ



Introduction to Error Analysis

Part 2: Fitting

Overview

- principle of Maximum Likelihood
- minimizing χ^2
- linear regression
- fitting of an arbitrery curve

Likelihood

- observed N data points, from parent population
- assume Gaussian parent distribution (mean μ, std. deviation σ)
- probability to observe x_i, given true μ, σ

$$P_i(x_i|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]$$

- probability to have measured μ' in this single measurement, given observed x_i and σ_i, is called **likelihood**:

$$P_i(\mu'|x_i,\sigma_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]$$

- for N observations, total likelihood is

$$L \equiv P(\mu') = \Pi_{i=1}^{N} P_i(\mu')$$

Principle of Maximum Likelihood

- **maximizing $P(\mu')$ gives μ' as the best estimate of μ**
  ("the most likely population from which data might have come is assumed to be the correct one")

- **for Gaussian individual probability distributions ($\sigma_i = \text{const} = \sigma$)**

$$L = P(\mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left[-\frac{1}{2}\sum\left(\frac{x_i - \mu}{\sigma}\right)^2\right]$$

- **maximizing likelihood $\Rightarrow$ minimizing argumen of Exp.**

$$\chi^2 \equiv \sum\left(\frac{x_i - \mu}{\sigma}\right)^2$$

Example: calculating mean

- cross-checking...

$$\frac{d\chi^2}{d\mu'} = \frac{d}{d\mu'} \sum \left(\frac{x_i - \mu'}{\sigma} \right)^2 = 0$$

- derivative is linear

$\Rightarrow$

$$\sum \left(\frac{x_i - \mu'}{\sigma} \right) = 0$$

$\Rightarrow$

$$\mu' = \overline{x} \equiv \frac{1}{N} \sum x_i$$

The mean really is the best estimate of the measured quantity.

Linear regression

- **simplest case: linear functional dependence**
- **measurements y_i, model (prediction) $y = f(x) = a + bx$**
- **in each point, $\mu \equiv y(x_i) = a + bx_i$**
  (special case $f(x_i) = \text{const} = a = \mu$)
- **minimize**

$$\chi^2(a,b) = \sum \left(\frac{y_i - f(x_i)}{\sigma} \right)^2$$

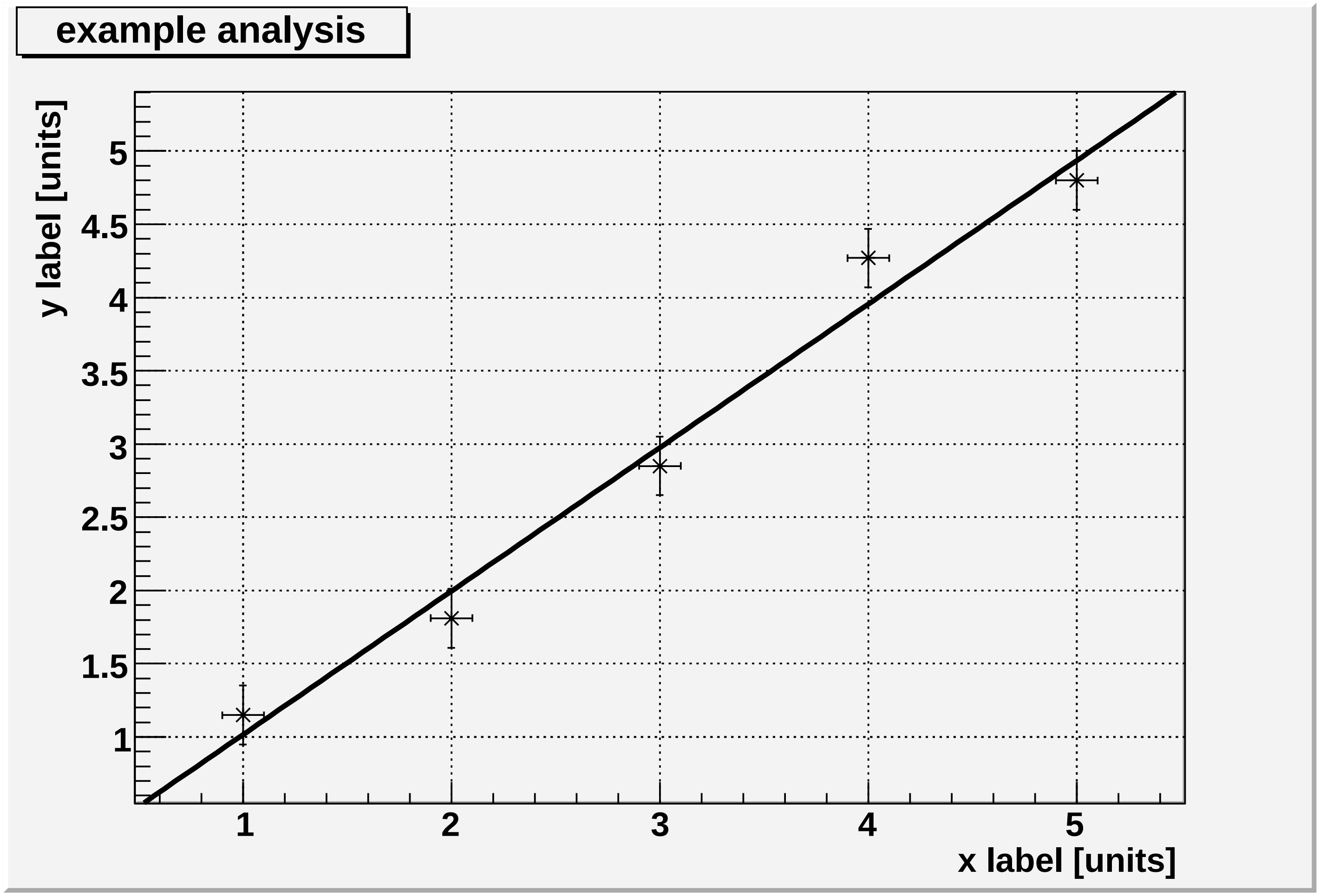
- **conditions for a minimum in 2-dim parameter space:**

$$\frac{\partial}{\partial a}\chi^2(a,b) = 0 \qquad \frac{\partial}{\partial b}\chi^2(a,b) = 0$$

- **can be solved analytically, but don't do it in real life**
  (e.g. see p.105 in Bevington)

Familiar example from day one: linear fit

- A program (e.g. ROOT) will do minimization of χ^2 for you



- This program will give you the answer for a and b ($= \mu \pm \sigma$)

Fitting with an arbitrary curve

- **a set of measurement pairs** $(x_i, y_i \pm \sigma_i)$
  (note no errors on x_i!)
- **theoretical model (prediction) may depend on several parameters $\{a_i\}$ and doesn't have to be linear**

$$y = f(x; a_0, a_1, a_2, ...)$$

- **identical approach: minimize total χ^2**

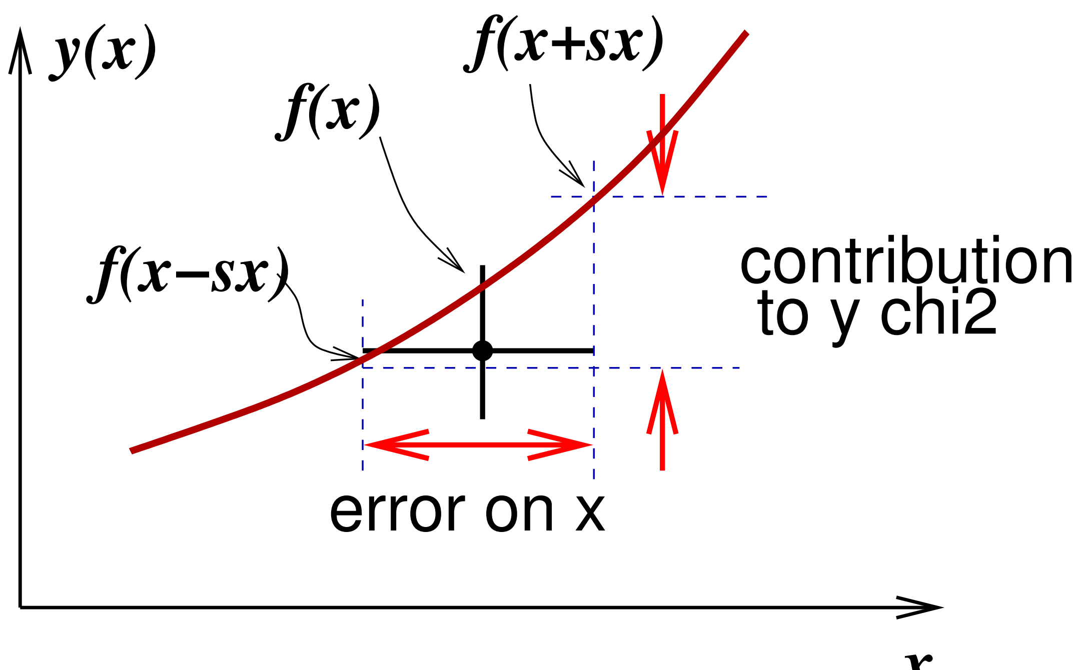
$$\chi^2(\{a_i\}) = \sum \left(\frac{y_i - f(x_i; a_0, a_1, a_2, ...)}{\sigma_i} \right)^2$$

- **minimization proceeds numerically**

Fitting data points with errors on both x and y

- $x_i \pm \sigma_i^x, y_i \pm \sigma_i^y$
- **Each term in χ^2 sum gets a correction from the σ_i^x contribution:**

$$\sum \left(\frac{y_i - f(x_i)}{\sigma_i^y} \right)^2 \rightarrow \sum \frac{(y_i - f(x_i))^2}{(\sigma_i^y)^2 + \left(\frac{f(x_i + \sigma_i^x) - f(x_i - \sigma_i^x)}{2} \right)^2}$$



Behavior of χ^2 function near minimum

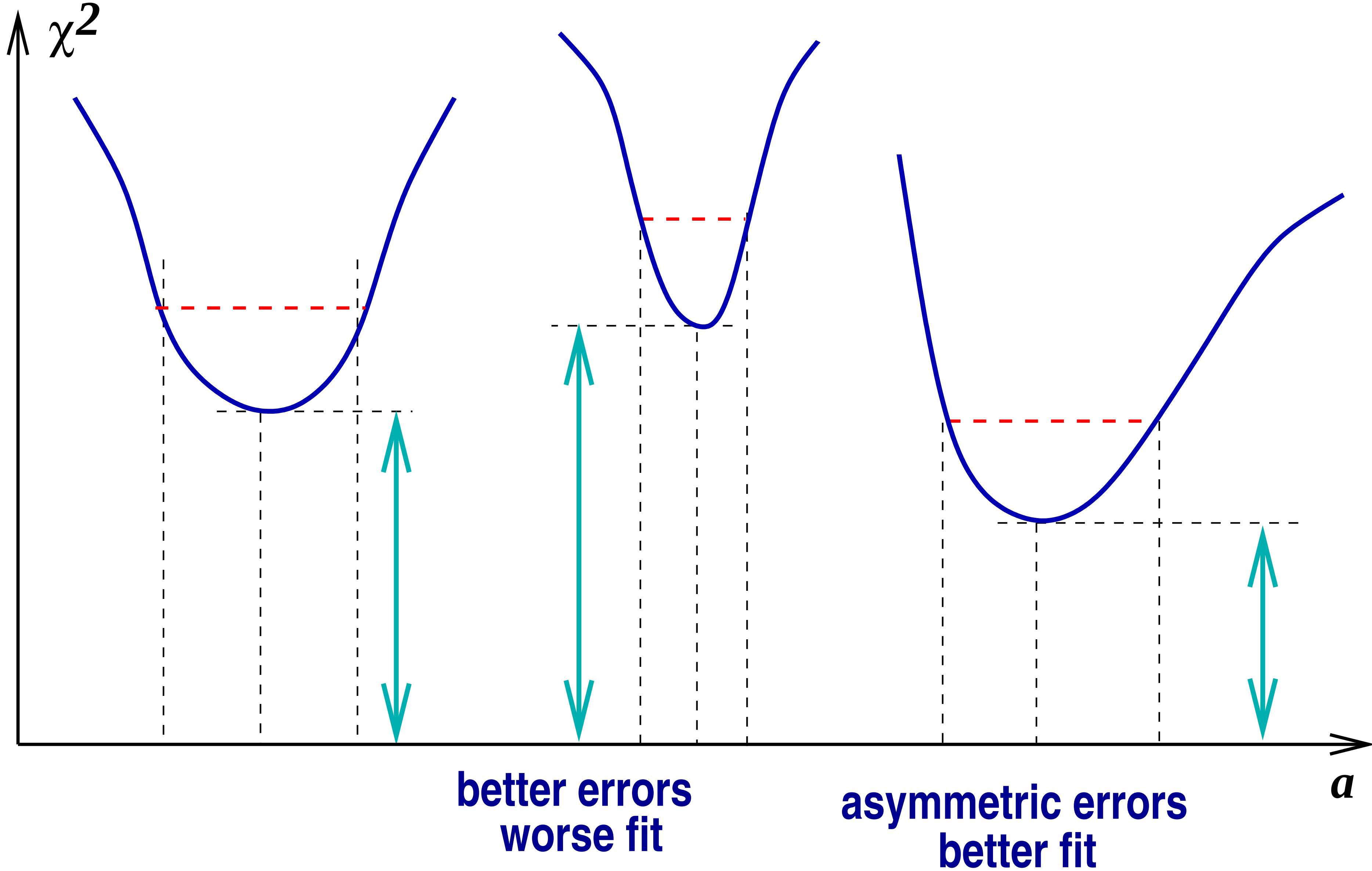
- when N is large, $\chi^2(a_0, a_1, a_2, \ldots)$ becomes quadratic in each parameter near minimum

$$\chi^2 = \frac{(a_j - a'_j)^2}{\sigma_j^2} + C$$

- known as parabollic approximation
- C tells us about goodness of the overall fit (function of all uncertaintines + other $\{a_k\}$ for $k \neq j$
- $\Delta a_j = \sigma_j \quad \Rightarrow \quad \Delta\chi^2 = 1$
  valid in all cases!
- parabollic error is the curvature at the minimum

$$\frac{\partial^2 \chi^2}{\partial a_j^2} = \frac{2}{\sigma_j^2}$$

χ^2 shapes near minimum: examples



Two methods for obtaining the error

1.

$$\sigma_j^2 = \frac{2}{\frac{\partial^2 \chi^2}{\partial a_j^2}}$$

2. **scan each parameter around minimum while others are fixed until $\Delta\chi^2 = 1$ is reached**

- **method #1 is much faster to calculate**
- **method #2 is more generic and works even when the shape of χ^2 near minimum is not exactly parabollic**
- **the scan of $\Delta\chi^2 = 1$ defines a so-called *one-sigma contour*. It contains the *'truth'* with 68.3% probability**

  (assuming Gaussian errors)

What to remember

- In the end the fit will be done for you by the program

  you supply the data, e.g. (x_i, y_i)
  and the fit model, e.g. $y = f(x; a_1, a_2, \ldots)$

  the program returns $a_1 = \mu_1 \pm \sigma_1, a_2 = \mu_2 \pm \sigma_2$,...
  and the plot with the model line through the points

- You need to understand what is done
- In more complex cases you may need to go deep into code

Introduction to Error Analysis

Part 3: Combining measurements

Overview

- propagation of errors
- covariance
- weighted average and its error
- error on sample mean and sample standard deviation

Propagation of Errors

- x is a known function of u, $v \ldots$

$$x = f(u, v, \ldots)$$

- assume that most probable value for x is

$$\bar{x} = f(\bar{u}, \bar{v}, \ldots)$$

$\bar{x}$ is the mean of $x_i = f(u_i, v_i, \ldots)$

- by definition of variance

$$\sigma_x = \lim_{N \to \infty} \left[\frac{1}{N} \sum (x_i - \bar{x})^2 \right]$$

- expand $(x_i - \bar{x})$ in Taylor series:

$$x_i - \bar{x} \approx (u_i - \bar{u}) \left(\frac{\partial x}{\partial u} \right) + (v_i - \bar{v}) \left(\frac{\partial x}{\partial v} \right) + \cdots$$

Variance of x

$$\sigma_x^2 \approx \lim_{N\to\infty} \frac{1}{N} \sum \left[(u_i - \bar{u}) \left(\frac{\partial x}{\partial u}\right) + (v_i - \bar{v}) \left(\frac{\partial x}{\partial v}\right) + \cdots \right]^2$$

$$\approx \lim_{N\to\infty} \frac{1}{N} \sum \left[(u_i - \bar{u})^2 \left(\frac{\partial x}{\partial u}\right)^2 + (v_i - \bar{v})^2 \left(\frac{\partial x}{\partial v}\right)^2 \right.$$

$$\left. + 2(u_i - \bar{u})(v_i - \bar{v}) \left(\frac{\partial x}{\partial u}\right) \left(\frac{\partial x}{\partial v}\right) + \cdots \right]$$

$$\approx \sigma_u^2 \left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right)^2 + 2\sigma_{uv}^2 \left(\frac{\partial x}{\partial u}\right) \left(\frac{\partial x}{\partial v}\right) + \cdots$$

This is the *error propagation equation*.

σ_{uv} is COvariance. Describes correlation between errors on u and v.

For uncorrelated errors $\sigma_{uv} \to 0$

Examples

- $x = u + a$

  where $a =$const. Thus $\partial x / \partial u = 1$

  $\Rightarrow \sigma_x = \sigma_u$

- $x = au + bv$

  where $a, b =$const.

  $\Rightarrow \sigma_x^2 = a^2 \sigma_u^2 + b^2 \sigma_v^2 + 2ab\sigma_{uv}^2$

  - correlation can be negative, i.e. $\sigma_{uv}^2 < 0$
  - if an error on u counterballanced by a proportional error on v, σ_x can get very small!

More examples

- $x = auv$

$$\left(\frac{\partial x}{\partial u}\right) = av \qquad \left(\frac{\partial x}{\partial v}\right) = au$$

$$\Rightarrow \sigma_x^2 = (av\sigma_u)^2 + (au\sigma_v)^2 + 2a^2uv\sigma_{uv}^2$$

$$\Rightarrow \quad \frac{\sigma_x^2}{x^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} + 2\frac{\sigma_{uv}^2}{uv^2}$$

- $x = a\frac{u}{v}$

$$\Rightarrow \quad \frac{\sigma_x^2}{x^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} - 2\frac{\sigma_{uv}^2}{uv^2}$$

- **etc., etc.**

Weighted average

From part #2: calculation of the mean

- **minimizing**

$$\chi^2 = \sum \left(\frac{x_i - \mu'}{\sigma_i} \right)^2$$

- **minimum at $d\chi^2/d\mu' = 0$, but now $\sigma_i \neq$ const.**

$$0 = \sum \left(\frac{x_i - \mu'}{\sigma_i^2} \right) = \sum \left(\frac{x_i}{\sigma_i^2} \right) - \mu' \sum \left(\frac{1}{\sigma_i^2} \right)$$

$\Rightarrow$ **so-called *weighted average* is**

$$\mu' = \frac{\sum \left(\frac{x_i}{\sigma_i^2} \right)}{\sum \left(\frac{1}{\sigma_i^2} \right)}$$

- **each measurement is *weighted* by $1/\sigma_i^2$!**

Error on weighted average

- N points contribute to a weighted average μ'
- straight application of the error propagation equation:

$$\sigma_\mu^2 = \sum \sigma_i^2 \left(\frac{\partial \mu'}{\partial x_i}\right)^2$$

$$\left(\frac{\partial \mu'}{\partial x_i}\right) = \frac{\partial}{\partial x_i} \frac{\sum (x_j/\sigma_j^2)}{\sum (1/\sigma_k^2)} = \frac{1/\sigma_i^2}{\sum (1/\sigma_k^2)}$$

- putting both together

$$\frac{1}{\sigma_\mu^2} = \left[\sum \sigma_i^2 \left(\frac{1/\sigma_i^2}{\sum (1/\sigma_k^2)}\right)^2\right]^{-1} = \sum \frac{1}{\sigma_k^2}$$

Example of weighted average

- $x_1 = 25.0 \pm 1.0$
- $x_2 = 20.0 \pm 5.0$
- error

$$\sigma^2 = \frac{1}{1/1 + 1/5^2} = 25/26 \approx 0.96 \approx 1.0$$

- weighted average

$$\bar{x} = \sigma^2 \left(\frac{25}{1^2} + \frac{20}{5^2} \right) = \frac{25}{26} \times \frac{25 \times 25 + 20 \times 1}{25} = 24.8$$

- result: $\bar{x} = 24.8 \pm 1.0$

$\Rightarrow$ **morale:** x_2 **practically doesn't matter!**

Error on the mean

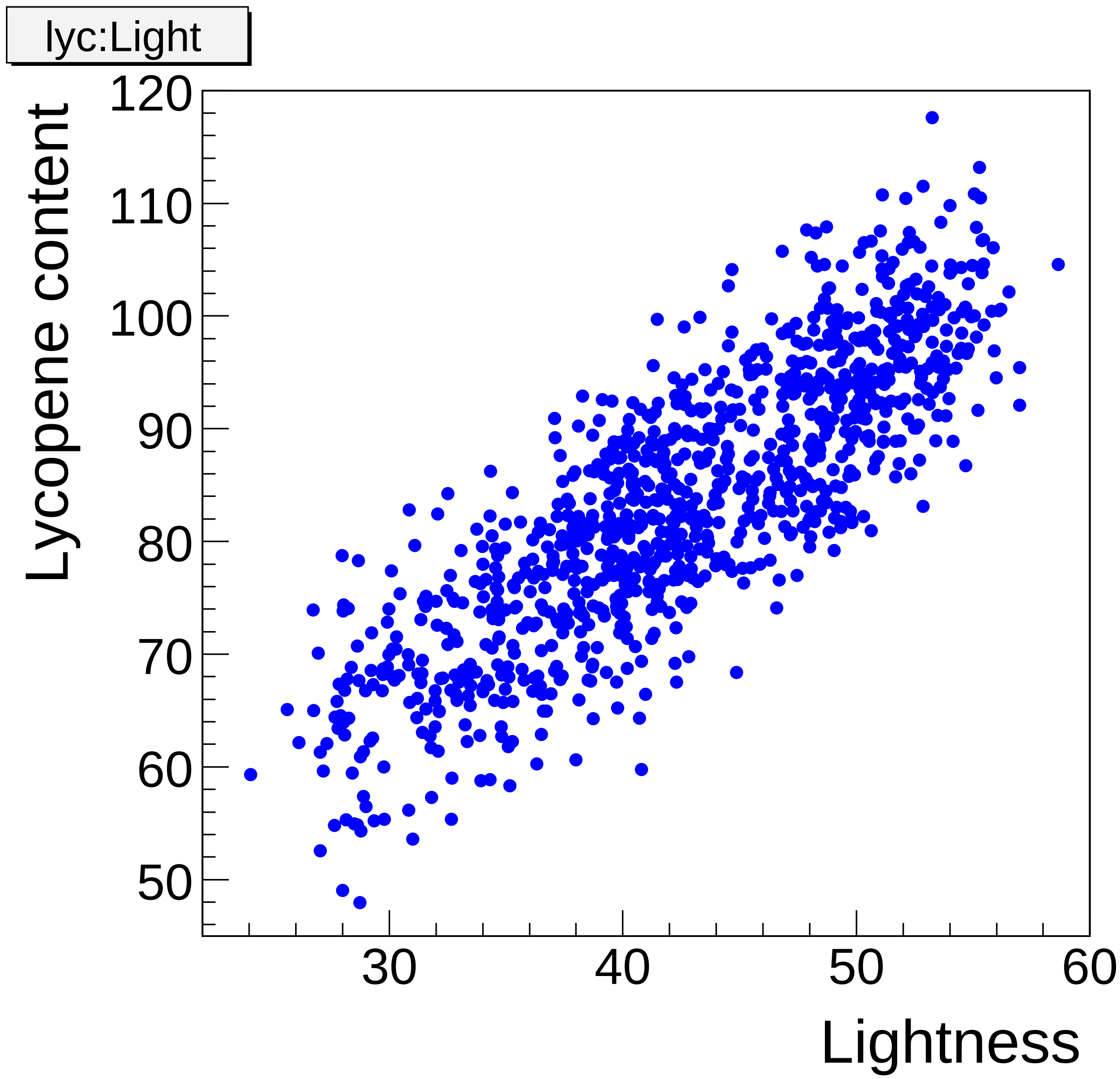
- N measurements from the same parent population (μ, σ)
- from part #1: sample mean μ' and sample standard deviation are best estimators of the parent population
- but: more measurements still gives same σ:
  - our knowledge of shape of parent population improves
  - and thus of original true error on each point
  - but how well do we know the true value? (i.e. μ?)
- if N points from same population with σ:

$$\frac{1}{\sigma_\mu^2} = \sum \frac{1}{\sigma^2} = \frac{N}{\sigma^2}$$

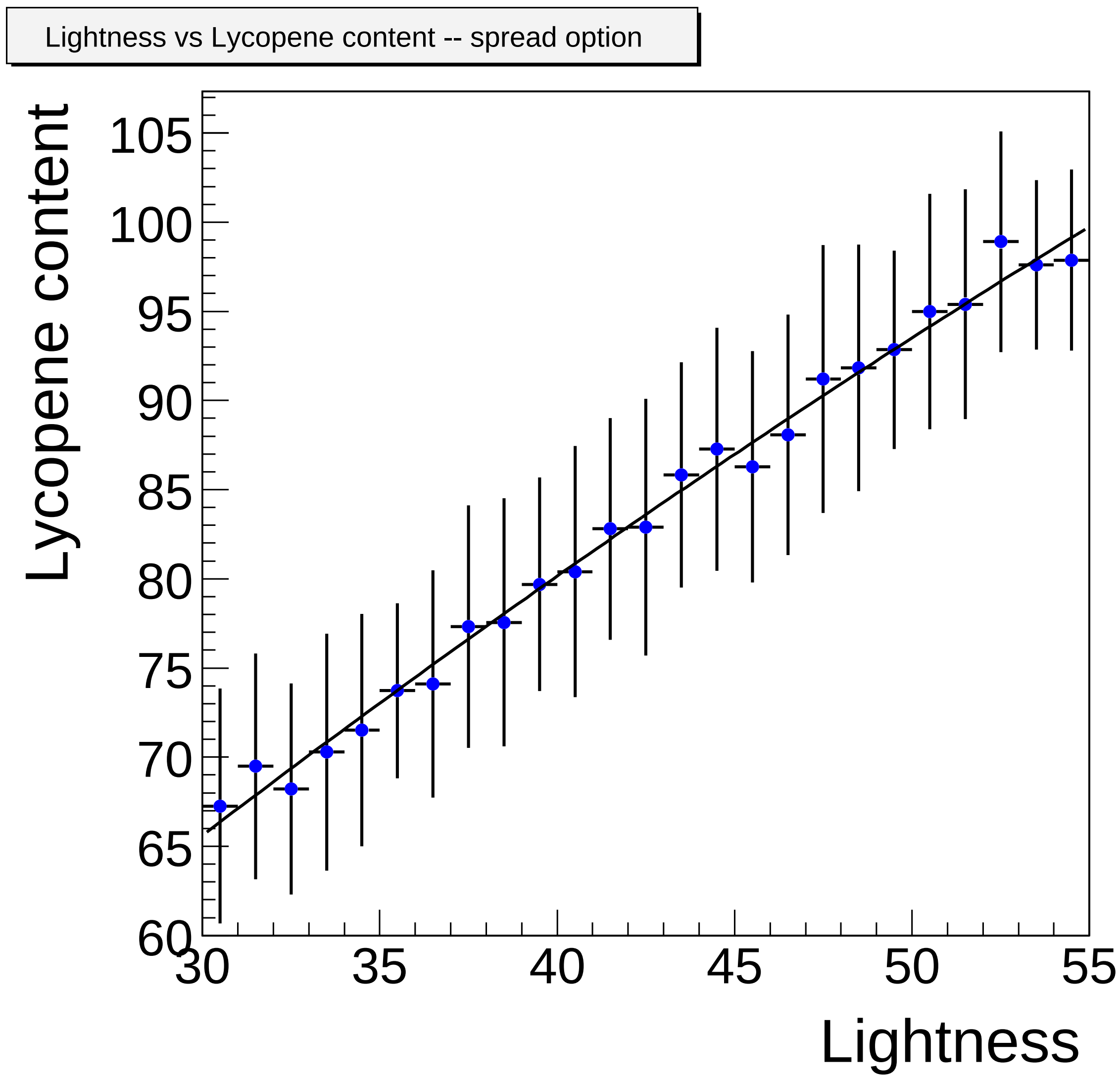
$$\Rightarrow \quad \sigma_\mu = \frac{\sigma}{\sqrt{N}} \approx \frac{s}{\sqrt{N}}$$

Standard deviation of the mean, or *standard error*.

Example: Lightness vs Lycopene content, scatter plot

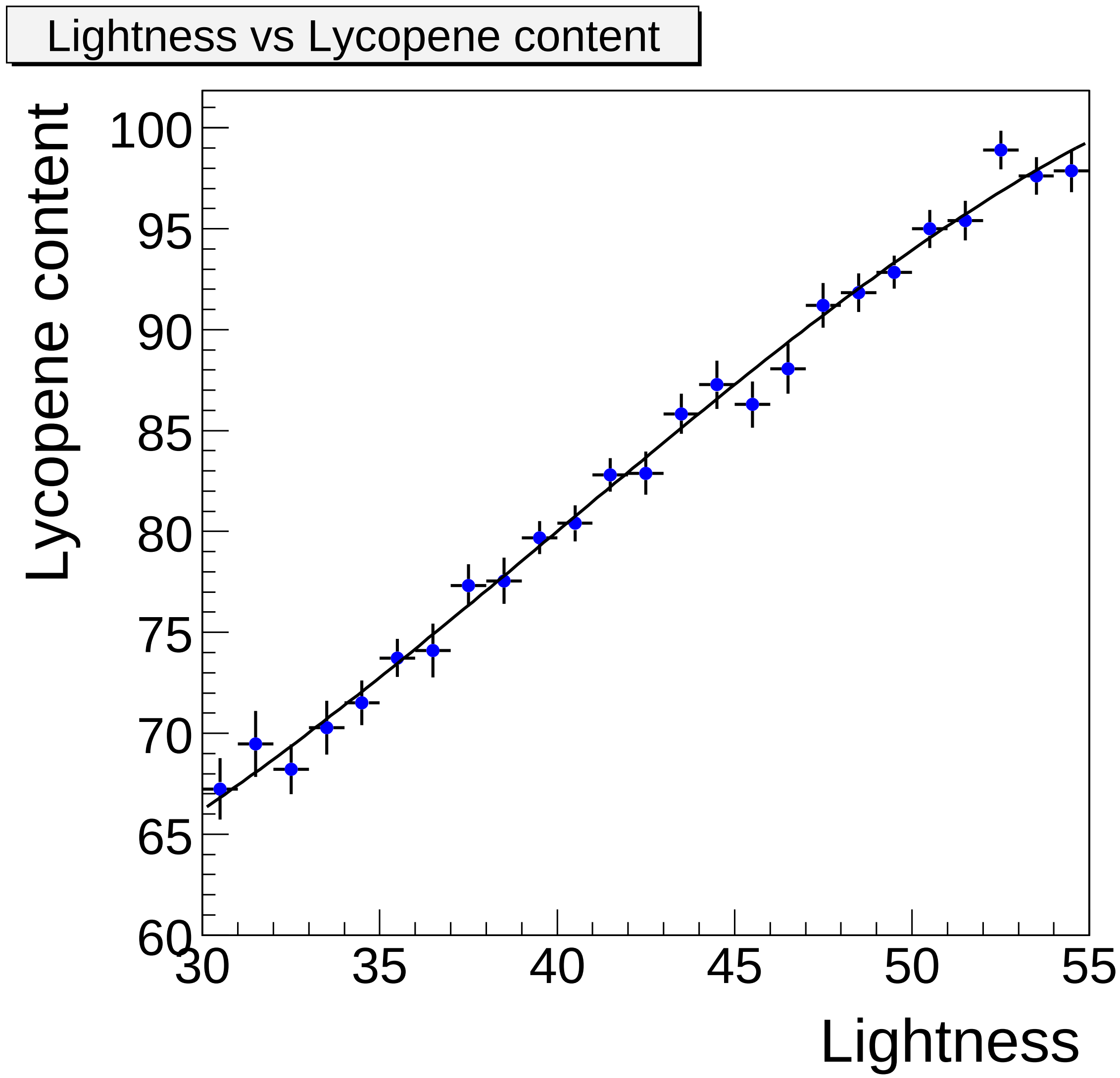


Example: Lightness vs Lycopene content: RMS as Error



Points don't scatter enough $\Rightarrow$ the error bars are too large!

Example: Lightness vs Lycopene content: Error on Mean



This looks much better!

Introduction to Error Analysis

Part 4: dealing with non-Gaussian cases

Overview

- **binomial p.d.f.**
- **Poisson p.d.f.**

Binomial probability density function

- random process with exactly two outcomes (Bernoulli process)
- probability for one outcome ("success") is p
- probability for exactly r successes ($0 \le r \le N$) in N independent trials
- order of "successes" and "failures" doesn't matter
- binomial p.d.f.:

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

- mean: Np
- variance: $Np(1-p)$
- if r and s are binomially distributed with (N_r, p) and (N_s, p), then $t = r + s$ distributed with $(N_r + N_s, p)$.

Examples of binomial probability

Binomial distribution always shows up when data exhibits binary properties:

- **event passes or fails**

  efficiency (an important exp. parameter) defined as

$$\epsilon = N_{\text{pass}}/N_{\text{total}}$$

- **particles in a sample are positive or negative**

Poisson probability density function

- **probability of finding exactly n events in a given interval of x (*e.g.*, space and time)**
- **events are independent of each other and of x**
- **average rate of ν per interval**
- **Poisson p.d.f. ($\nu > 0$)**

$$f(n;\nu) = \frac{\nu^n e^{-\nu}}{n!}$$

- **mean: ν**
- **variance: ν**
- **limiting case of binomial for many events with low probability:**
  $p \longrightarrow 0$, $N \longrightarrow \infty$ while $Np = \nu$
- **Poisson approaches Gaussian for large ν**

Examples of Poisson probability

Shows up in counting measurements with small number of events

- **number of watermellons with circumference** $c \in [19.5, 20.5]\text{in.}$
- **nuclear spectroscopy – in tails of distribution**

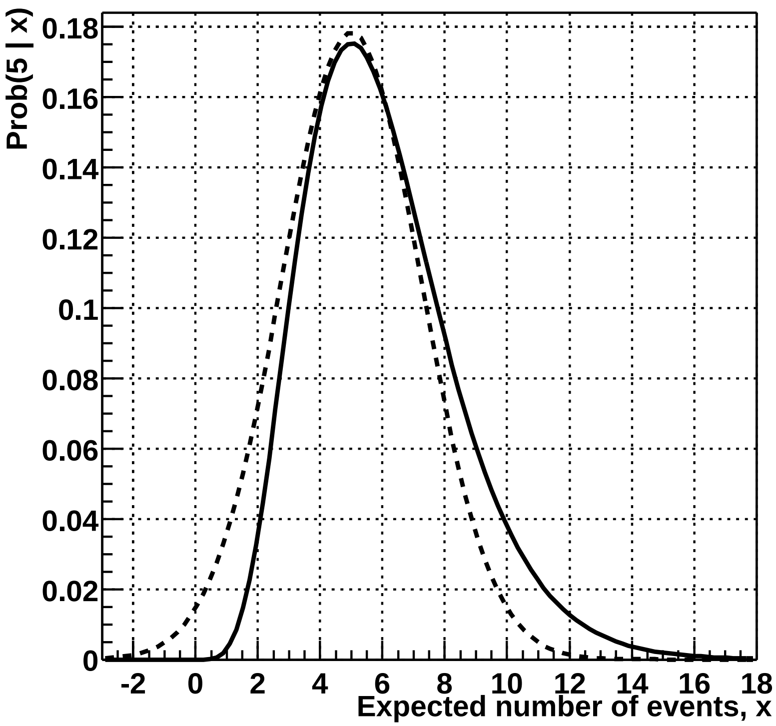
  (e.g. high channel number)
- **Rutherford experiment**

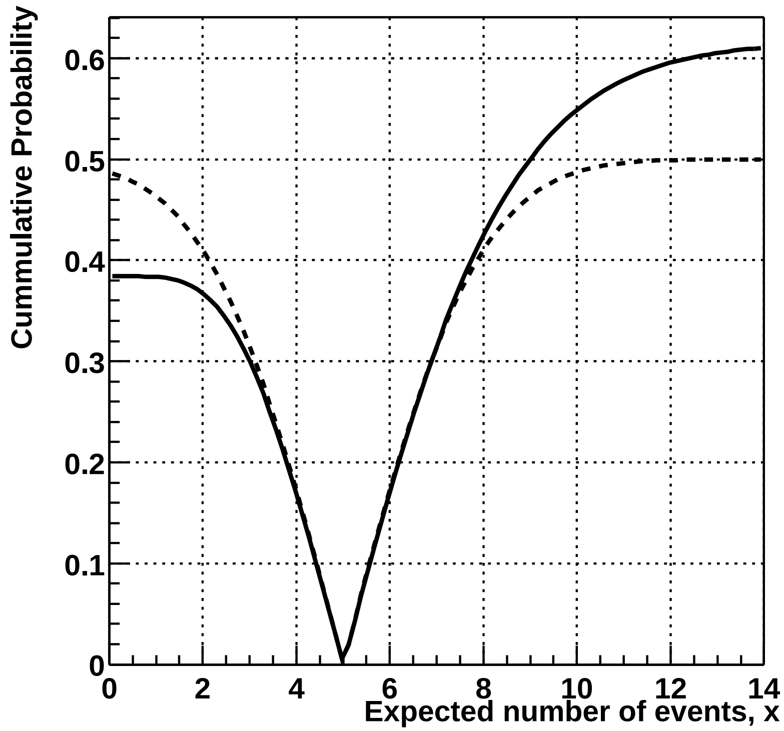
Fitting a histogram with Poisson-distributed content

Poisson data require special treatment in terms of fitting!

- **histogram, i-th channel contains n_i entries**
- **for large n_i, $P(n_i)$ is Gaussian**
  Poisson $\sigma = \sqrt{\nu}$ approximated by $\sqrt{n_i}$
  (WARNING: this is what ROOT uses by default!)
- **Gaussian p.d.f. $\longrightarrow$ symmetric errors**
  $\Rightarrow$ equal probability to fluctuate up or down
- **minimizing χ^2 $\Rightarrow$ fit (the 'true value') is** *equally likely to be above and below the data!*

Comparing Poisson and Gaussian p.d.f.





- 5 observed events
- Dashed: Gaussian at 5 with $\sigma = \sqrt{5}$
- Solid: Poisson with $\nu = 5$
- Left: prob.density functions (note: Gauss can be < 0!)
- Right: confidence integrals (p.d.f integrated from 5)