M2 Physics and Instrumentation
S3M Doctoral School Training course

Data acquisition and analysis

Module DataFit: Data analysis and modelling

Lectures – Ian Sims
Tutorials – Chinmai Sai Jureddy

Website: https://perso.univ-rennes1.fr/ian.sims/DataFit/

Institut de Physique de Rennes

Administration

15 hours lectures – Ian Sims (ian.sims@univ-rennes.fr)

15 hours tutorials – Chinmai Sai Jureddy

(chinmai-sai.jureddy@univ-rennes.fr)

All sessions run from 09.00-12.00 in rooms at the PNRB (Pôle Numérique Rennes Beaulieu), building 9b on the Beaulieu campus

Thursday 11/09/2025 TD5	Monday 06/10/2025 TD5
Friday 19/09/2025 TD3	Friday 10/10/2025 TD3
Friday 26/09/2025 TD3	Monday 13/10/2025 TD5
Monday 29/09/2025 TD5	Friday 17/10/2025 TD3
Friday 03/10/2025 TD3	Friday 24/10/2025 TD3

Connection from remote sites by Zoom

Website: https://perso.univ-rennes1.fr/ian.sims/DataFit/

(for lecture notes, tutorial sheets)

Assessment will be by two individual projects

References

Philip R. Bevington and D. Keith Robinson
Data Reduction and Error Analysis for the Physical Sciences, 3rd ed.
McGraw-Hill, 2002, ISBN 0072472278

John R. Taylor
An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, 2nd ed.
University Science Books, U.S., 1997, ISBN 093570275X

(French translation available in Library ISBN 2100043072)

Or more recent editions.

3

Part 1 – Error Estimation and Part 2 – Data Fitting Course Contents

General Introduction: Data fitting and error estimation in the physical sciences.

Part 1 – Error estimation

Error estimation and statistical description of data; Introduction: Uncertainties in measurement (accuracy and precision); Distributions and averages; Central limit theorem; Error analysis – internal and external errors; Simple error estimation – propagation of errors; Rejection of outliers – Chauvenet's criterion; Weighted means and weighted errors.

Part 2 – Data fitting

Linear Least-squares data fitting: χ^2 minimisation; Straight line fit; Confidence limits; Testing the fit; Student's *t*-distribution; General linear least squares.

Non-linear least squares data fitting: Introduction; Examples of non-linear functions common in nature; Exponential decay; Methods of minimising χ^2; The Marquardt algorithm.

Other methods of data fitting: Least absolute deviation; Maximum likelihood method; Robust estimation; Data smoothing

4

1.1 Introduction: The importance of error estimation

As physical scientists, much of what we do involves measurement.

A result without an associated error estimation is of little use.

Take, for example, these three results for the rate coefficient between CN and O_2 at room temperature:

- $k = 2.3018 \times 10^{-11}$ cm^3 molecule^{-1} s^{-1}
- $k = (2.3018 \pm 0.05) \times 10^{-11}$ cm^3 molecule^{-1} s^{-1}
- $k = (2.3 \pm 0.1^{a}) \times 10^{-11}$ cm^3 molecule^{-1} s^{-1}

  [a]errors quoted correspond to $\pm 2\sigma$ statistical error only.

It is essential that

- we have some understanding of errors, and
- we are able to estimate the error associated with any particular measurement.

Important also to minimise errors! Involves repeating measurements, and data analysis/reduction.

5

1.2.1 Error Estimation: Uncertainties in measurement

Let's have some definitions: what is error?

Error: "the difference between an observed or calculated value and the true value"

We assume there is a 'true value' to the quantity we are trying to measure. Is the difference between our quoted value and this 'true value' the error?

If I *knew* the error, I'd know the 'true value' and quote this instead! So, the best I can do is quote my best estimate of both the measured value and the *probable* error.

This isn't a course on probability and statistics. Instead, I'll recap some of the main results as and when they are needed.

6

1.2.2 Accuracy and Precision

Accuracy

what's the difference?

Precision

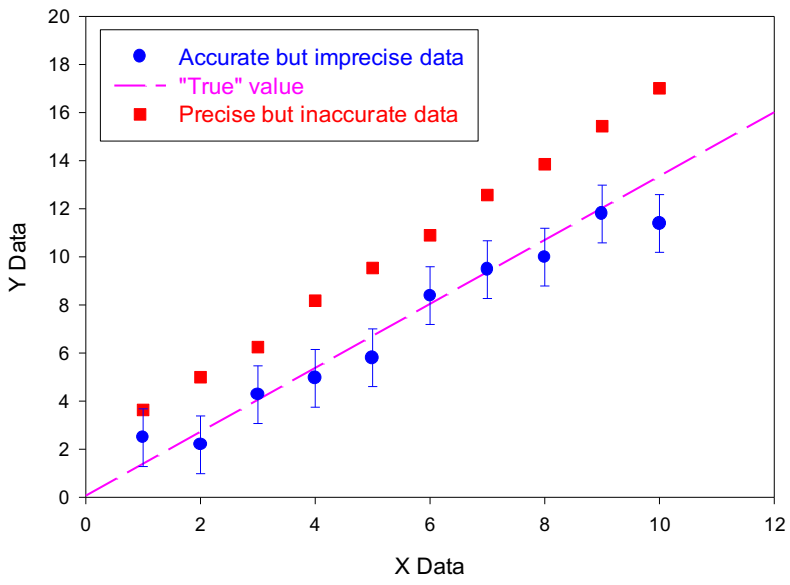
Accuracy :

- a measure of how close the result of the experiment is to its true value

Precision :

- a measure of how well the result has been determined, without reference to its agreement with the true value

7



8

1.2.3 Types of error

What are the different types of error?

- mistakes or blunders in measurement or computation
  - we will not consider these further: repeated measurement and careful checking needed to identify and avoid
- random error
- systematic error

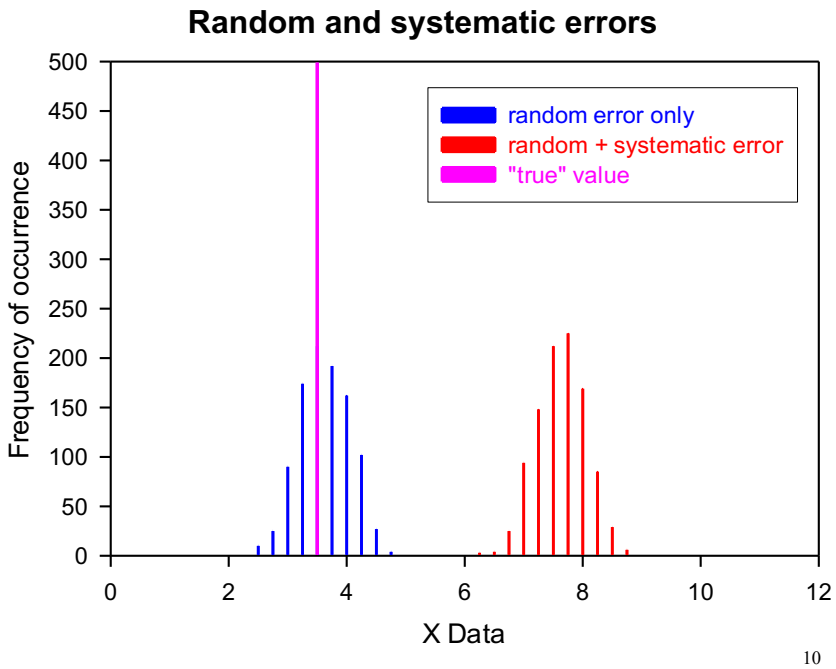
Random error:

- fluctuations in observations yielding results that differ from experiment to experiment, need repeated measurements to give precise results.

Systematic error:

- a 'bias' which makes results differ from the 'true' values by a reproducible discrepancy.
  - hard to detect and not easily studied by statistical analysis.

9



10

1.2.3 Types of error: significant figures

Both types of error can be reduced by paying careful attention to experiment, but

systematic errors may go undetected as we often lack the ability to estimate them.

Significant figures

Precision of experimental result implied by the number of digits recorded – though uncertainty should be quoted

Quote one more sig. fig. than dictated by precision, to reduce rounding errors

e.g., measure a time of $t = 1.203$ s, but we know that our real precision is only $\pm$ 0.1 s

so, quote $t = (1.20 \pm 0.1)$ s

11

1.2.4 Parent and Sample Distributions

Parent distribution

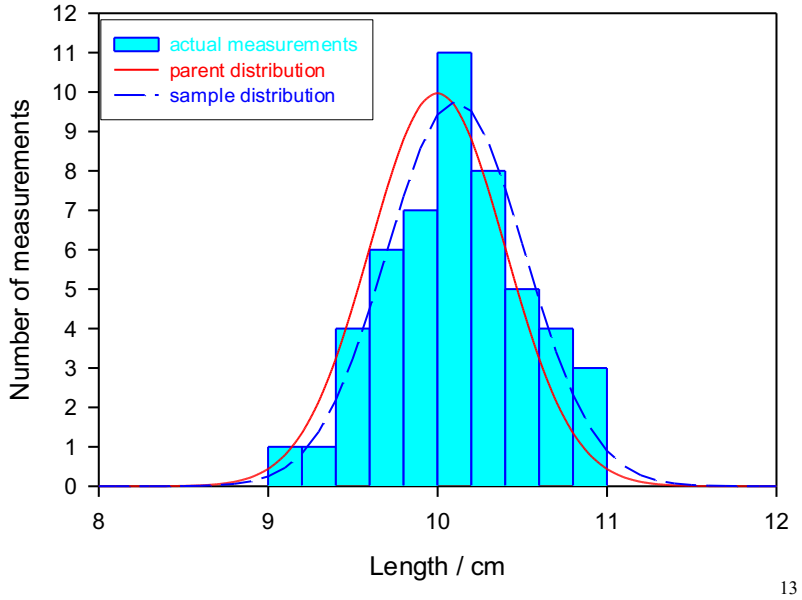
- a hypothetical distribution which determines the probability of getting any particular observation.

Sample distribution

- hypothesise that the measurements are samples from the parent distribution, and they form the sample distribution.
- In the limit of infinite measurements, the sample distribution becomes the parent distribution.

12

1.2.4 Parent and Sample Distributions (contd)



13

1.2.5 Mean, Median and Mode

If we make N measurements x_1, x_2, x_3,… and so on, up to a final measurement x_N, we write the sum of these measurements as follows:

$$\sum x_i \equiv \sum_{i=1}^{N} x_i$$

The mean of our experimental sample distribution is taken as

$$\bar{x} = \frac{1}{N}\sum x_i$$

while the mean, μ, of the parent population is defined as

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

14

1.2.5 Mean, Median and Mode (contd)

The **median** of the parent population, $\mu_{½}$, is defined as that value for which, in an infinite number of determinations x_i, half the determinations will be less than the median and half will be greater:

$$P(x_i < \mu_{½}) = P(x_i \geq \mu_{½}) = \frac{1}{2}$$

Not often used as a statistical parameter.

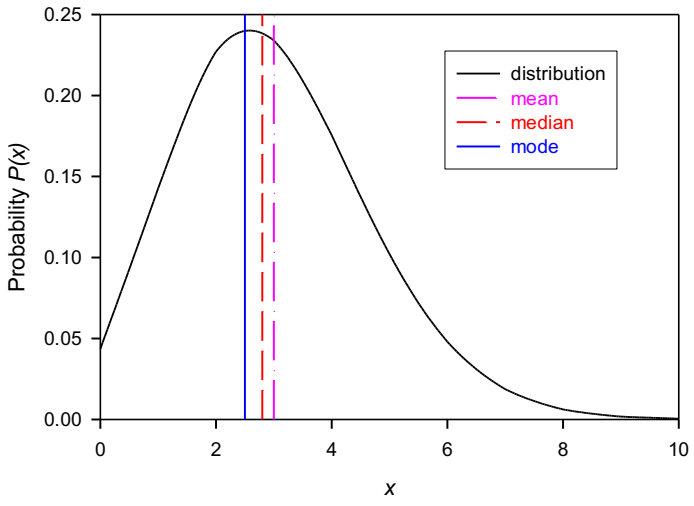
The **mode**, or most probable value, μ_{max}, of the parent population is defined as that value for which the parent distribution has the greatest value. In the limit of a large number of observations, this value will probably occur most often:

$$P(\mu_{max}) \geq P(x \neq \mu_{max})$$

For a symmetrical distribution these three quantities will be identical. For an asymmetrical distribution they will differ as shown.

15

1.2.5 Mean, Median and Mode (contd)



16

1.2.6 Deviations from the Mean

Average deviation, α

$$\alpha \equiv \lim_{N\to\infty}\left[\frac{1}{N}\sum|x_i-\mu|\right]$$

Inconvenient for statistical analysis. Better is standard deviation, σ, or the variance σ^2.

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

σ is associated with the second moment of the data about the mean. We can only estimate the parameters of the parent population by measuring the standard deviation of the sample population, *s*:

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

17

1.2.7 Probability Distributions

Three main types of probability distribution:

- binomial distribution
- Poisson distribution
- Gaussian distribution

- **binomial distribution** applied to experiments where result is one of a small number of possible final states, e.g. 'heads' or 'tails'. The other two distributions considered as limiting cases of binomial distribution.
- **Poisson distribution** appropriate for counting experiments e.g., radioactive decay
- **Gaussian or normal error distribution** is most important as seems to describe the distribution of random observations for most experiments.

18

1.2.7 Probability Distributions: the Binomial Distribution

If

- probability of observing 'heads up' any coin is p and
- tails is q (= 1 – p),

then the probability for observing each combination of x heads and n-x tails with n coins is $p^x q^{n-x}$

The probability $P_B(x; n, p)$ of observing x of the n items to be in a state with probability p is given by the binomial distribution

$$P_B(x;n,p) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

The mean of the binomial distribution is simply

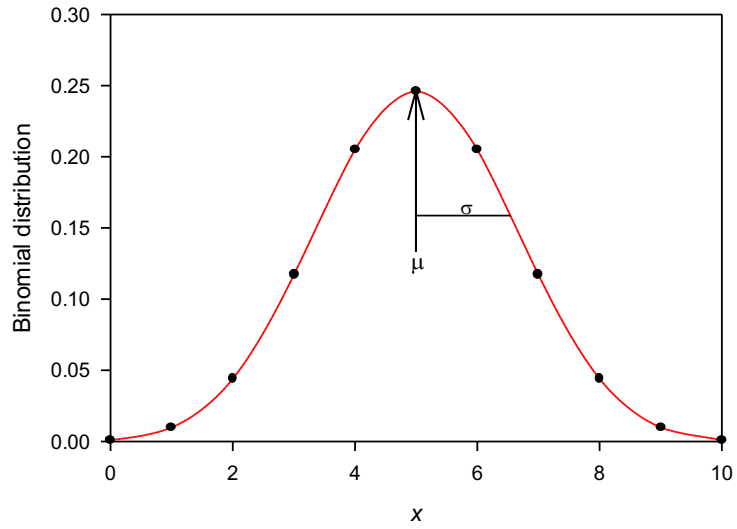
$\mu = np$

and the variance is given by

$\sigma^2 = np(1 - p)$

19

1.2.7 Probability Distributions: the Binomial Distribution



20

1.2.7 Probability Distributions: the Poisson Distribution

Approximation to binomial distribution for special case where average number of successes is much smaller than possible number, i.e.

$$\mu << n \text{ and } p << 1$$

$$\lim_{p\to 0} P_B(x;n,p) = P_P(x;\mu) \equiv \frac{\mu^x}{x!} e^{-\mu}$$

The mean is simply given by the parameter μ and the standard deviation is equal to the square root of the mean.

21

1.2.7 Probability Distributions: the Gaussian Distribution

Approximation to binomial distribution where $np >> 1$,
also limiting case for Poisson distribution when μ becomes large.
It is defined as

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

or as the standard Gaussian distribution with $z = (x - \mu)/\sigma$:

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

Mean and standard deviation correspond to the parameters μ and σ
Probabilities are about 68% and 96% that a given measurement will fall within one or two standard deviations of the mean.

22

1.2.8 The Central Limit Theorem

This is a very useful theorem, the proof of which justifies our use of the Gaussian distribution. It states:

If you take the sum X of N independent variables, x_i, where i = 1,2,3,…,N, each taken from a distribution of mean μ_I and variance σ_i^2, the distribution for X

a) has an expectation value $\langle X \rangle = \Sigma\, \mu_I$

b) has variance $V(X) = \Sigma\, \mu_i^2$

c) becomes Gaussian as $N \rightarrow \infty$

CLT applies only when the variables are independent

c) is the reason that the Gaussian is so important.

Proofs of the CLT may be found in many text books.

23

1.2.8 Central Limit Theorem: Repeated measurements

Suppose we measure the same quantity many times. We can use the CLT in a simple form since all the μ_I have the same value – call it μ – and all σ_i have the same value σ.

From the CLT we have

$$\langle X \rangle = \Sigma\, \mu = N\mu$$

and, in terms of the average

$$\bar{x} = \frac{X}{N} \qquad \langle \bar{x} \rangle = \mu$$

and provided the measurements are independent the variance of the average is just the variance of X, divided by N^2

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

thus, **the standard deviation of this average falls as $N^{-½}$**

24

1.2.8 Repeated measurements (cont)

Reminder: for a measurement of a single parameter with normally distributed errors, we have the standard deviation *s* of the sample population (i.e., our measurements) given by

$$s = \left\{ \frac{1}{N-1} \sum \left(x_i - \bar{x} \right)^2 \right\}^{\frac{1}{2}}$$

We are usually relaxed about writing σ instead of *s*.

However, what we really want to know are the **confidence limits** for our measurement.

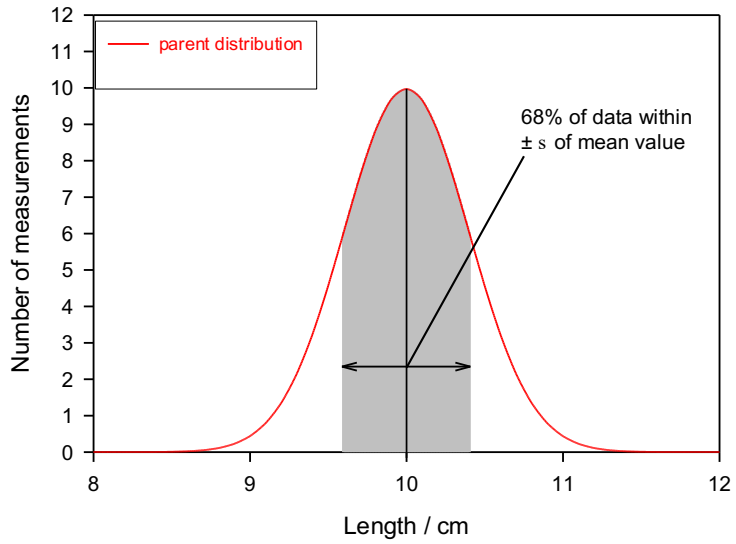
For this, we first need the standard deviation of the distribution of our sample mean about the parent mean μ. This is called the standard error, σ_{std}, and is related to the standard deviation of our sample as follows:

$$\sigma_{std} = \frac{\sigma \text{ (}s\text{ strictly)}}{\sqrt{\nu}} = \frac{\sigma}{\sqrt{N-1}} \text{ for single parameter determination}$$

If *N* is large (or we think of the 'parent population'), 68% of our data fall within ± σ of the mean, and 96% within ±2 σ.

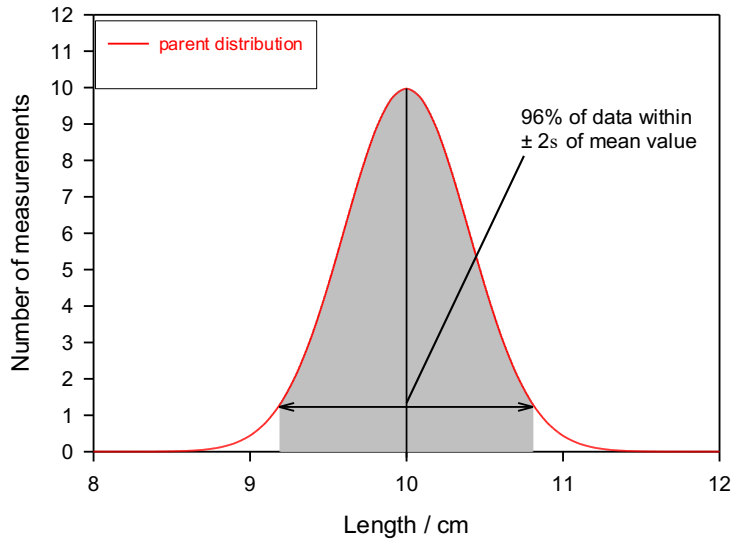
25

1.2.8 Repeated measurements (cont)



26

1.2.8 Repeated measurements (cont)



27

1.2.8 Repeated measurements (cont)

For small *N*, we must use the so-called Student's *t*-factor to define our confidence limits.

This is usually defined at the 95% confidence limit, so our confidence range can be defined as $\pm (t_{0.95} \times \sigma_{std})$

Values of *t* can be found in many statistics textbooks, and they depend on the degrees of freedom, v (= *N* – no. of parameters ≡ *N* – 1 in this case)

28

1.2.8 Repeated measurements (cont)

v (= N – 1 in this case)	$t_{0.95}$
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
6	2.447
7	2.365
8	2.306
9	2.262
10	2.228
11	2.201
12	2.179
20	2.086
30	2.042
60	2.000
∞	1.960

29

1.2.9 Error analysis

Instrumental and Statistical Uncertainties

Instrumental uncertainties arise from imprecision in the measuring instrument

Two ways of estimating the uncertainty:

The **external method** of considering the equipment and the experiment itself, e.g. precision of the measuring scale

The **internal method**, calculate the standard deviation from the spread of measurements.

Statistical uncertainties arise from statistical fluctuations in the collections of finite numbers of counts over finite periods of time, e.g., photon counting experiments.

We can directly estimate the standard deviation in this case, according to the Poisson distribution:

$\sigma = \sqrt{\mu}$

30

1.2.9 Error analysis: Propagation of errors

For $x = f(u, v, \ldots)$, the error propagation equation is:

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

where the covariance, σ_{uv}^2, is defined as

$$\sigma_{uv}^2 \equiv \lim_{N\to\infty}\left[\frac{1}{N}\sum\left[(u_i - \overline{u_i})(v_i - \overline{v_i})\right]\right]$$

Fortunately, fluctuations in measurements of u and v are often uncorrelated, and this term vanishes:

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

31

1.2.9 Error analysis: Propagation of errors (example)

Derive one error propagation formula as example. If $x = ae^{\pm bu}$ then:

$$\frac{\partial x}{\partial u} = \pm abe^{\pm bu} = \pm bx$$

and the relative uncertainty becomes

$$\frac{\sigma_x}{x} = \pm b\sigma_u$$

or if $x = a^{\pm bu}$ then

$$\frac{\sigma_x}{x} = \pm(b\ln a)\sigma_u$$

32

1.2.9 Error Analysis: Specific error propagation formulae

Derivations for the following can be found in most text books on errors:

$x^2 = au \pm bv$	$\sigma_x^2 = a^2\sigma_u^2 + b^2\sigma_v^2 \pm ab\sigma_{uv}^2$
$x = auv$	$\frac{\sigma_x^2}{x^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} + 2\frac{\sigma_{uv}^2}{uv}$
$x = \frac{au}{v}$	$\frac{\sigma_x^2}{x^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} - 2\frac{\sigma_{uv}^2}{uv}$
$x = au^{\pm b}$	$\frac{\sigma_x}{x} = \pm b\frac{\sigma_u}{u}$
$x = ae^{\pm bu}$	$\frac{\sigma_x}{x} = \pm b\sigma_u$
$x = a^{\pm bu}$	$\frac{\sigma_x}{x} = \pm(b\ln a)\sigma_u$
$x = a\ln(\pm bu)$	$\sigma_x = a\frac{\sigma_u}{u}$

33

1.2.10 Chauvenet's criterion: elimination of data points

Example: 100 measurements of the length of an object (parent mean 10 cm, standard deviation 0.5 cm

- one measurement is recorded as 98.2 cm: blunder

But what if we saw a measurement of, say, 12.1 cm? 4 standard deviations away from mean – 0.06% probability, so in 100 measurements only expect to collect 0.06 such events.

Chauvenet's criterion states

- discard a data point if we expect less than half an event to be further from the mean than the suspect point.

This will have a bigger effect on the standard deviation than on the mean, so beware of removing further points!

'So unexpected was the hole that for several years computers analysing ozone data had systematically thrown out the readings that should have pointed to its growth.'

New Scientist, 31 March 1988

34

1.2.10 Elimination of data points: practical rules

1) Calculate mean value and *average deviation* d_{av} of points *excluding* the suspect point

$$d_{av} = \frac{1}{N}\sum_{i=1}^{N}|d_i| \qquad d_i = x_i - \bar{x}$$

2) Discard a point if it lies more than $4d_{av}$ from this mean value

3) Do not discard more than one point in 5

4) Do not discard two data values if they are the same.

35

1.3.1 Weighting data: weighted means

Problem: some data points might be measured with better or worse precision than others.

Assume parent distributions with the same mean μ but different standard deviations σ_I.

Assign to each data point x_i, its own standard deviation σ_I, obtain:

$$P(\mu') = \prod_{i=1}^{n}\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)\exp\left[-\frac{1}{2}\sum\left(\frac{x_i-\mu'}{\sigma_i}\right)^2\right]$$

Method of maximum likelihood states that most probable value for μ' is the one that gives maximum value to $P(\mu')$, so minimise the argument in the exponential:

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

36

1.3.1 Weighting data: weighted means and errors

This gives the most probable value as the weighted average of the data points:

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

and the general formula for the uncertainty in the weighted mean is:

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

37

2.1.1 Linear least-squares data fitting

Data sample $\{(x_i, y_i)\}$, x_i known exactly, y_i have been measured, each with some known resolution σ_i.

y is function f of x, depends on a parameter a

Invoking the CLT, the probability of a particular y_i, for a given x_i, is

$$P(y_i;a) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{y_i - f(x_i;a)}{\sigma_i}\right)^2\right]$$

and the probability for the complete data set is then given by

$$P(y;a) = \prod_{i=1}^{N}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{y_i - f(x_i;a)}{\sigma_i}\right)^2\right]$$

$$P(y;a) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N\exp\left[-\frac{1}{2}\sum\left(\frac{y_i - f(x_i;a)}{\sigma_i}\right)^2\right]$$

38

2.1.1 Linear least-squares data fitting : χ^2 minimisation

To maximise the likelihood, minimise the quantity

$$\sum_{i=1}^{N}\left(\frac{y_i - f(x_i;a)}{\sigma_i}\right)^2$$

i.e., make the weighted sum of the squared differences as small as possible – the method of least squares

This weighted sum is known as χ^2

Least squares seems a reasonable estimator, appears to work in practice

Other estimators, e.g., least absolute deviation, might, in some circumstances, be better.

39

2.1.2 Least squares fit to a straight line

If $y_0 = a_0 + b_0 x$ then χ^2, the sum to be minimised is

$$\sum_{i=1}^{N}\left(\frac{y_i - a - bx_i}{\sigma_i}\right)^2$$

To find the values of a and b which yield minimum of χ^2, set to zero the partial derivatives of χ^2 with respect to each of the parameters.

These equations can then be rearranged as a pair of linear simultaneous equations in the unknown parameters a and b, solution (preferably by the method of determinants) yields the following:

$$a = \frac{1}{\Delta}\left(\sum\frac{x_i^2}{\sigma_i^2}\sum\frac{y_i}{\sigma_i^2} - \sum\frac{x_i}{\sigma_i^2}\sum\frac{x_i y_i}{\sigma_i^2}\right)$$

$$b = \frac{1}{\Delta}\left(\sum\frac{1}{\sigma_i^2}\sum\frac{x_i y_i}{\sigma_i^2} - \sum\frac{x_i}{\sigma_i^2}\sum\frac{y_i}{\sigma_i^2}\right)$$

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

40

2.1.3 Least squares fit to a straight line: confidence limits

For a parameter z we can write the variance σ_z^2, using the error propagation method

$$\sigma_z^2 = \sum \left[\sigma_i^2 \left(\frac{\partial z}{\partial y_i} \right)^2 \right]$$

For the straight line case, we obtain

$$\sigma_a^2 = \frac{1}{\Delta} \sum \frac{x_i^2}{\sigma_i^2}$$

$$\sigma_b^2 = \frac{1}{\Delta} \sum \frac{1}{\sigma_i^2}$$

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

41

2.1.3 Straight line confidence limits: problems

What if we don't know the individual σ_i's of the data points?

- set all of individual σ_i's to 1

OK for parameter estimates, but error estimates will be **incorrect**, as they have been derived in part from the individual σ_i's (set to 1).

For normally distributed errors, expect the value of the reduced χ^2 to approach 1. So, without any justification we simply assume that this is the case for our fit.

Multiply initial estimates of σ_a and σ_b by $(\chi^2 / (N - 2))^{½}$, using the value of χ^2 computed using the fitted parameters a and b

This gives the best estimate for the probable uncertainties.

If we define σ^2 as

$$\sigma^2 = \sum_{i=1}^{N} (y_i - a - bx_i)^2 \qquad \text{then}$$

42

2.1.3 Straight line confidence limits: unweighted fit

$$\sigma^2 = \sum_{i=1}^{N} (y_i - a - bx_i)^2$$

$$\Delta = N\sum x_i^2 - \left(\sum x_i\right)^2$$

$$a = \frac{1}{\Delta}\left(\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i\right)$$

$$b = \frac{1}{\Delta}\left(N\sum x_i y_i - \sum x_i \sum y_i\right)$$

$$\sigma_a{}^2 = \frac{\sigma^2}{\Delta}\sum x_i^2$$

$$\sigma_b{}^2 = N\frac{\sigma^2}{\Delta}$$

43

2.1.4 Least squares fit to a straight line: testing the fit

Linear correlation coefficient: measures how well correlated y is with x
It has a value of 0 for no correlation, and ± 1 for complete correlation.
Better are the following tests:

a) look at the fit! Or perhaps more usefully, examine the residuals $((y_i - f(x_i))$ versus $x_i)$: are they evenly distributed about 0?

b) if you have estimates of the individual σ_i's then use a program which can calculate a 'goodness-of-fit probability', *Q*.

Q is the probability that a value of χ^2 as *poor* (*great)* as the best fit value should occur by chance. If

- Q > 0.1 – goodness of fit is believable.
- Q > 0.001 – may be acceptable if errors are nonnormal or moderately underestimated.
- Q < 0.001 – model and/or estimation procedure suspect. Best then to use unweighted fit.

44

2.1.5 Student's *t*-distribution (again)

Say you want to measure the rate coefficient k of the reaction
$OH + HCl \rightarrow H_2O + Cl$

You operate under pseudo-first-order conditions (an excess of HCl) and obtain 6 values of k_{1st} vs [HCl] A linear plot of k_{1st} vs [HCl] gives

$k = 4.102 \times 10^{-10}$ cm^3 molecule^{-1} s^{-1} and $\sigma = 0.080 \times 10^{-10}$ cm^3 molecule^{-1} s^{-1}

Your friend who does atmospheric modelling wants the value with 95% error limits. So, do you just quote $k \pm 2\sigma$? Why not?

The value of σ obtained does take into account the number of data points: for small samples, errors are not distributed according to a Gaussian distribution, but rather

Student's *t*-distribution, which

varies according to the number of degrees of freedom, v, and,

especially for small v, has larger tails at the side.

Presented as table of t for differing confidence limits and v

6 points, less 2 parameters gives v = 4, and $t_{0.95} = 2.776$, so quote:

$k = (4.10 \pm 0.22^a) \times 10^{-10}$ cm^3 molecule^{-1} s^{-1}

[a]errors quoted are $\pm t\sigma$ random error where t is the appropriate value of Student's t-distribution for the 95% point.

45

2.1.6 General linear least-squares

Linear least-squares applicable to any function linear in its parameters.

An example would be the polynomial

$y(x) = a_1 + a_2x + a_3x^2 + \ldots + a_Mx^{M-1}$

could include sines and cosines etc., so long as the overall function is linear in its parameters a_k:

$$y(x) = \sum_{k=1}^{M} [a_k f_k(x)]$$

Routines available for generalised linear least squares optimisation, relying on matrix methods.

With the advent of fast computers, largely superseded by generalised **non-linear** routines.

46

2.2.1 Non-linear least-squares fitting: Introduction

Standard linear least squares methods restricted to fitting functions that are linear in parameters a_k:

$$y(x) = \sum_{k=1}^{M} [a_k f_k(x)]$$

Minimising χ^2 only yields coupled equations linear in M unknown parameters if fitting functions $y(x)$ linear in the parameters

Cannot obtain an analytic solution for minimising χ^2 in non-linear case

Non-linear functions are very common in nature, for example, any process whose rate depends upon the magnitude of a population, as is the case for first-order chemical reactions, will behave in this way:

$A \rightarrow$ products

$$\text{rate of reaction} = \frac{-d[A]}{dt} = k[A]$$

$$[A]_t = [A]_0 \exp(-kt)$$

47

2.2.2 Non-linear least-squares fitting: Exponential decay

For some cases, for example the single exponential decay with no background, can *linearise* function, and then use linear least squares:

$$[A]_t = [A]_0 \exp(-kt)$$

$$\ln\left(\frac{[A]_t}{[A]_0}\right) = -kt$$

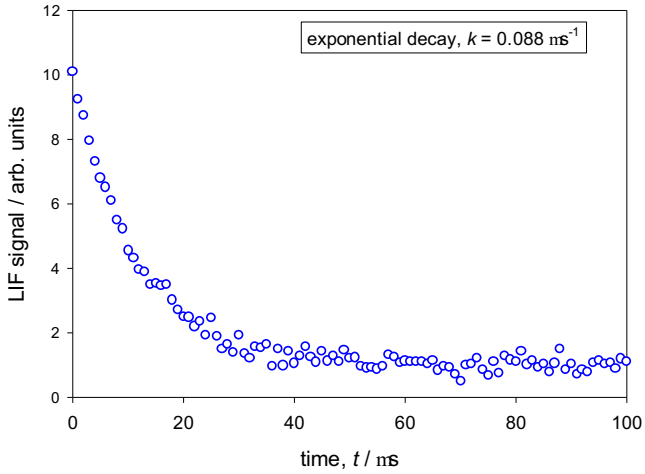
Warning: this will place undue weight on data at low signal levels

Example: often, we measure rates of reactions under pseudo-first-order conditions, and measure a laser-induced fluorescence (LIF) signal which is proportional to the concentration of the decaying species:

i.e. LIF signal $\propto [A]$

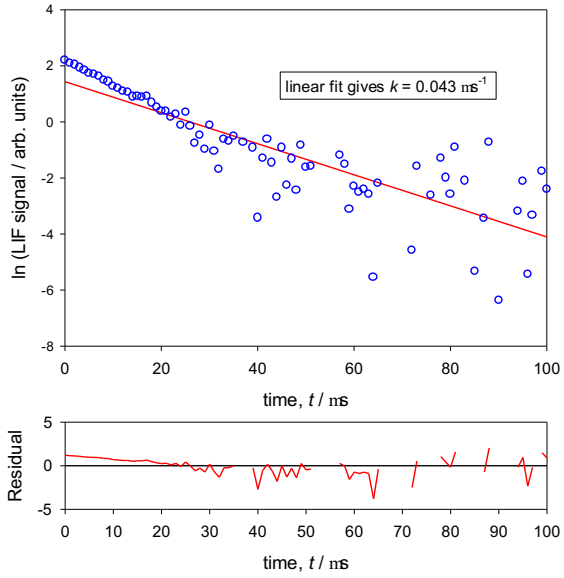
48

2.2.2 Non-linear least-squares fitting: Exponential decay



49

2.2.2 Non-linear least-squares fitting: Exponential decay



50

2.2.3 NLLSQ: methods of minimising χ^2

Consider χ^2 surface, or hypersurface, as a $f(a_k)$. Various numerical techniques are available for finding the minimum:

grid search: calculate grid of χ^2 values by varying each a_k in turn to find minimum value

- robust, but extremely inefficient

gradient search: vary all parameters simultaneously, and go down direction of 'steepest descent'

- more efficient at first than grid search, but very inefficient near the minimum where gradient → 0

expansion methods: expand the χ^2 surface as an analytic function, e.g., a parabola, then calculate minimum directly

- works well close to the minimum, but if started far away then results will be in error, and may even tend towards maximum

Marquardt method: see on

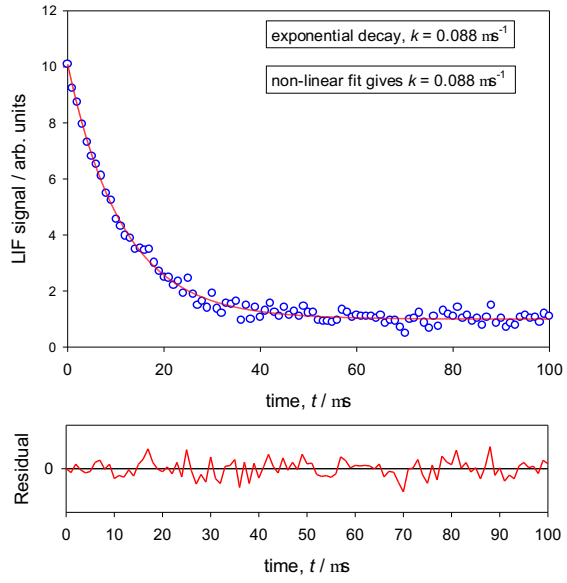
51

2.2.4 NLLSQ: The Marquardt method

- combines both gradient search and expansion methods, using an adjustable parameter λ to switch between an initial gradient search far from the minimum to a parabolic expansion near the minimum
- though implementation is most complex of methods discussed, it is clearly best, used in most packages (e.g., *Igor Pro, Origin*)
- Marquardt's algorithm uses matrix algebra, with λ initially set (at 0.001) to take advantage of both gradient and expansion methods:
  - an initial value for χ^2 is calculated, and then a step is taken according to the combined gradient/expansion algorithm.
  - If χ^2 increases, λ is increased by a factor 10, putting more emphasis on the line of steepest descent.
  - If χ^2 decreases, λ is decreased by a factor 10, tending towards the expansion method.
  - repeated until χ^2 ceases to vary (by more than a set tolerance).
  - minimum will have been calculated using a parabolic expansion, curvature yields the covariance matrix and hence the errors.

52

2.2.4 NLLSQ: The Marquardt method



53

2.2.4 NLLSQ: The Marquardt method (cont)

As with any method of fitting, we make a basic assumption about the functional form of our data. In the same way as for straight line data we must assess how well the data fit this form:

a) look at the fit! Or perhaps more usefully, the residuals ($(y_i - f(x_i))$ versus x_i): are they evenly distributed about 0?

b) examine a goodness-of-fit criteria such as the value of χ^2

The Marquardt algorithm, while very powerful, is not fool-proof. It is still possible to find yourself stuck in a local minimum on the χ^2 surface, for example.

Essential to examine the fit and, if necessary, choose alternative starting values or step sizes.

54

2.3 Other methods of data fitting

Sometimes standard linear or non-linear least-squares data fitting may fail (as mentioned above). This may be due to (for example)

- local minima when fitting complex functions with many parameters
- undue influence of *outliers* – data points lying far from the mean – whose influence is over-emphasised in LSQ methods

This motivates the use of other so-called *robust* methods for data fitting, including so-called M-estimates (following from maximum-likelihood arguments) such as

- least absolute deviation

or other robust minimisation methods such as the

- downhill simplex algorithm

as well as techniques involving a priori knowledge of the behaviour (often time-dependent behaviour) of model parameters and their co-variances such as

- Kalman filtering

55

2.3 Other methods of data fitting

Figure from Numerical Recipes in Fortran. The Art of Scientific Computing, 2nd Edition, 1992, ISBN 0-521-43064-X.

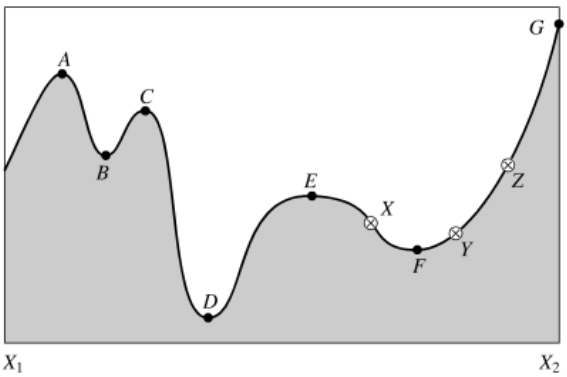


Figure 10.0.1. Extrema of a function in an interval. Points A, C, and E are local, but not global maxima. Points B and F are local, but not global minima. The global maximum occurs at G, which is on the boundary of the interval so that the derivative of the function need not vanish there. The global minimum is at D. At point E, derivatives higher than the first vanish, a situation which can cause difficulty for some algorithms. The points X, Y, and Z are said to "bracket" the minimum F, since Y is less than both X and Z.

56

2.3.1 Maximum likelihood estimation

Suppose we fit N data points (x_i, y_i) $i = 1, \ldots, N$ to a model with M adjustable parameters a_j, $j = 1, ..., M$

$f(x) = y(x) = y(x; a_1 ... a_M)$

What do we minimise to get fitted values for the parameters a_j? We have previously used least squares

$$\text{minimise over } a_1 \ldots a_M : \sum_{i=1}^{N} \left[y_i - y\left(x_i; a_1 \ldots a_M\right) \right]^2$$

But where does this come from? Leads to the idea of

- maximum likelihood estimators

Define likelihood as

- probability of the data given the parameters

and then

- fit the parameters to maximise the likelihood.

57

2.3.1 Maximum likelihood estimation (cont)

For normally distributed errors (Gaussian error distribution) the probability of the data set is the product of the probabilities of each point, as in Section 2.1.1

$$P(y; a_1, \ldots a_N) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{y_i - f(x_i; a_1, \ldots a_N)}{\sigma_i} \right)^2 \right]$$

$$P(y; a_1, \ldots a_N) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^N \exp\left[-\frac{1}{2} \sum \left(\frac{y_i - f(x_i; a_1, \ldots a_N)}{\sigma_i} \right)^2 \right]$$

and maximising this probability is the equivalent to minimising the negative of its logarithm, namely

$$\chi^2 = \sum \left(\frac{y_i - f(x_i; a_1, \ldots a_N)}{\sigma_i} \right)^2$$

which leads to the least squares methods we have been studyng. HOWEVER, this is only strictly correct with Gaussian (Normal) errors. In reality, “outliers” exist which can skew the fit, and need robust methods.

58

2.3 Other methods of data fitting

Figure from Numerical Recipes in Fortran. The Art of Scientific Computing, 2nd Edition, 1992, ISBN 0-521-43064-X.

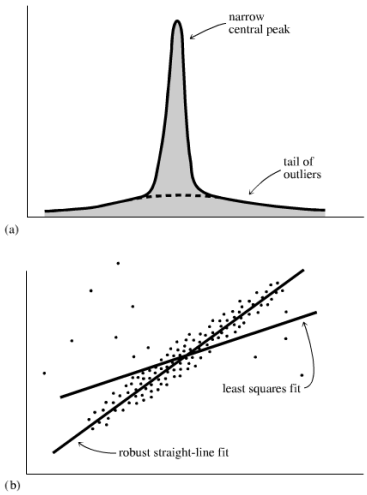


Figure 15.7.1. Examples where robust statistical methods are desirable: (a) A one-dimensional distribution with a tail of outliers; statistical fluctuations in these outliers can prevent accurate determination of the position of the central peak. (b) A distribution in two dimensions fitted to a straight line; non-robust techniques such as least-squares fitting can have undesired sensitivity to outlying points.

59

2.3.2 Robust estimation – local M-estimates

If we know our measurement errors are not normally distributed then we would write for the probability of the data given the parameters

$$P(y;a_1,\ldots a_N)=\prod_{i=1}^{N}\left\{\exp\left[-\rho\left(y_i,y(x_i;a_1,\ldots a_N)\right)\right]\Delta y\right\}$$

where ρ is negative logarithm of the probability density, and we would want to minimise

$$\sum_{i=1}^{N}\rho\left(y_i,y(x_i;a_1,\ldots a_N)\right)$$

Very often, ρ depends not independently on its two arguments but only on their difference, usually scaled by some weighting factors σ_i and the **M-estimate** is said to be **local** and we can

$$\text{minimise over } a_1\ldots a_M:\ \sum_{i=1}^{N}\rho\left(\frac{y_i-y(x_i;a_1,\ldots a_N)}{\sigma_i}\right)$$

where the function $\rho(z)$ is a function of a single variable $z \equiv [y_i - y(x_i)]/\sigma_i$

If define derivative of $\rho(z)$ as a function $\psi(z)\equiv\dfrac{d\rho(z)}{dz}$

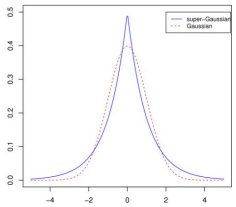
then normally distributed errors will give $\rho(z)=\dfrac{1}{2}z^2 \qquad \psi(z)\equiv z$ (normal)

60

2.3.2 Robust estimation – local M-estimates (cont)

If the errors are distributed as a double or two sided exponential then

$$\text{Prob}\left\{y_i - y\left(x_i\right)\right\} \sim \exp\left(-\left|\frac{y_i - y\left(x_i\right)}{\sigma_i}\right|\right)$$



and $\rho\left(z\right) = \left|z\right| \qquad \psi\left(z\right) \equiv \text{sgn}\left(z\right)$

(double exponential)

Maximum likelihood estimator is obtained by minimising mean absolute deviation – **least absolute deviation** method (available in Igor)

Another distribution with more extensive tails is **the Lorentzian distribution**

$$\text{Prob}\left\{y_i - y\left(x_i\right)\right\} \sim \frac{1}{1+\frac{1}{2}\left(\frac{y_i - y\left(x_i\right)}{\sigma_i}\right)^2}$$

$$\rho\left(z\right) = \log\left(1+\frac{1}{2}z^2\right) \quad \psi\left(z\right) = \frac{z}{1+\frac{1}{2}z^2} \qquad \text{(Lorentzian)}$$

ψ increases with deviation then decreases – true outliers are not counted

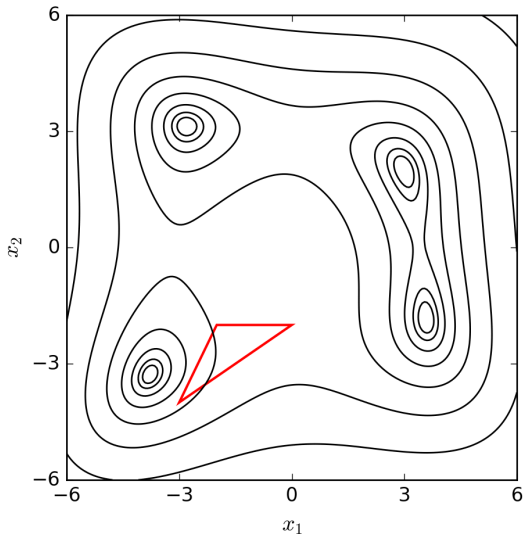
2.3.3 Robust estimation – Simplex (Nelder-Mead) algorithm

Even when errors normally distributed (and we can use χ^2 minimisation) the Marquardt algorithm can fails due to local minima and/or discontinuities on the χ^2 surface.

One (expensive) solution is the simplex algorithm for χ^2 minimisation

- a simplex is a special polytope of n + 1 vertices in n dimensions e.g. a line segment on a line or a triangle on a plane
- a simplex of n + 1 points is set up in the n-dimensional space of the variables (for example, in 2 dimensions the simplex is a triangle)
- vertex of the simplex with the largest χ^2 value is reflected in the centre of gravity of the remaining vertices and the χ^2 value at this new point is compared with the remaining χ^2 values
- depending on the outcome of this test the new point is accepted or rejected, a further expansion may be made, or a contraction.
- when no further progress can be made the sides of the simplex are reduced in length and the method is repeated until no further improvement (according to predefined tolerance)
- the method is very robust but can still be susceptible to local minima

Simplex algorithm



63

2.3.4 Robust estimation – other techniques

Sometimes we may have *a priori* knowledge about probable values and probable uncertainties of some parameters we are trying to estimate from a data set:

- neither completely freezing a parameter at a predetermined value
- nor completely leaving it to "float" (be determined by the data set)
- the formalism for this is called "use of *a priori* covariances"

Alternatively, in signal processing and control theory we may wish to "track" (maintain an estimate of) a time varying signal in presence of noise

- if parameters vary only slowly, Kalman filtering may be used to produce best parameter estimates as a function of time
- employs Bayesian inference and estimates a joint probability distribution over the variables for each timeframe.
- used in e.g. phase-locked loop (PLL) in radio receivers

We may wish to apply other techniques (e.g. filtering, deconvolution) before fitting the data. The simplest of these is smoothing.

64

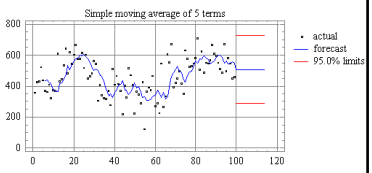
2.3.5 Data Smoothing

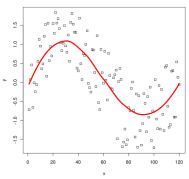
- Concept of data smoothing lies in a murky area, just beyond the fringe of these better posed and more highly recommended techniques:
  - least squares fitting to a parametric model
  - optimal filtering of a noisy signal
  - brutal honesty: show your data as it really is!
- However, it is more useful to have some techniques available which are more objective than

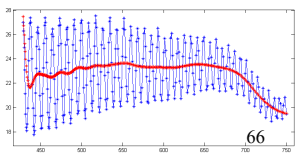
  "the smooth curve was drawn by eye through the original data"
  - through each individual data point?
  - through the forest of scattered points?
  - by a draftsman?
  - or by someone who knows the hypothesis the data are supposed to substantiate?
- Data smoothing: "art not science"
- not supposed to be tied to any particular functional form *y(x)*. However, it clearly involves some notion of averaging. Smoothing a set of values will not be the same as smoothing their logarithms. 65

2.3.5 Data Smoothing (cont)

- Some common smoothing algorithms are
- ***n*-point smoothing**
  - each point on the 'smoothed' curve is an average of *n* neighbours
  - quick, easy to program, but crude
- **Lowess smoothing**
  - locally weighted regression
  - each point on the curve is produced by a regression of data points close by, with the closest points more heavily weighted
- **Low-pass filtering**
  - removes high frequency components from signal
  - best algorithms based on Fourier transforms
  - see later for more on filtering







66