Advanced Physics Laboratory
PHY 326/327 and PHY 426/427/428/429

Lectures on Error Analysis in Scientific Measurement:
Part 1: The meaning and calculation of uncertainty
Part 2: The least-squares method of fitting data

J. H. Thywissen

September 12, 2006.

Outline:

References:

J. R. Taylor, *An Introduction to Error Analysis* (University Science Books, 1982)

P. R. Bevington & D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill)

J. Orear, *Notes on Statistics for Physicists, revised* (unpublished, 1982) online:
http://nedwww.ipac.caltech.edu/level5/Sept01/Orear/Orear.html

Part 1: The meaning and calculation of uncertainty

Outline:

- 1.1 Illegitimate Errors
- 1.2 Accuracy versus Precision
- 1.3 Recording Error
- 1.4 What is meant by statistical uncertainty?
- 1.5 Error propagation
- 1.6 Examples of error propagation
- 1.7 The Normal distribution
- 1.8 Weighted average & standard deviation of the mean

Lecture 1. Uncertainty & Error in Scientific Measurement

1.0 What is error? [Taylor, Ch. 1]

"Error" does not carry the usual connotation of 'mistake' or 'blunder', but means the inevitable uncertainty that ~~attends~~ is associated with all measurements.

1.1 Illegitimate errors

Errors that originate from mistakes or blunders in measurement or computation, are not considered in scientific uncertainty. These can be eliminated by carefully repeating observations, procedures, or calculation. Scientific errors are not mistakes.

1.2 Accuracy versus Precision [Bevington Ch. 1]

Accuracy: how close experimental result is to the true value.

Precision: how well a value has been determined without reference to true value.

②

1.2, cont

Imagine the Egyptian who didn't know π, measuring a bunch of circles (Q: what is ratio of circ. to diameter?)

Example: Assume π is correct answer.

3 ± 1 = accurate, not precise

2.93846 ± 0.00001 = precise, but not accurate

3.14159 ± 0.00001 = accurate & precise

10 ± 3 = neither.

Typically, our quoted uncertainties reflect precision. These are often the types of uncertainty that can be improved with ~~more~~ repeating the same measurements, and are well described with statistics. Precision is also known as "statistical error" for that reason.

Accuracy is harder to pin down. If, for instance, your metal ruler is callibrated for STP & you use it in a hot room, ~~you will measure the same thing ev~~ all your measurements will be off by ($\approx 10^{-5}\,K^{-1}$) A major part of planning & understanding an experiment has to do with reducing systematic errors.

1.3 Recording Error [Taylor, Ch. 2]

Typically, we write

$$\underbrace{3.14}_{\text{value}} \pm \underbrace{0.02}_{\text{statistical uncertainty OR precision}}$$

How many digits? As many as meaningful, due to uncertainty.

$$3.14\underbrace{394}_{\text{too many digits}} \pm 0.02$$

You may write ~~uncertainty with~~ two digits, especially when a single digit would be "1":

$$3.1416 \pm 0.0034$$

Rules: Write uncertainty with one or two digits
write value to same ~~# of digits~~ significant digits.

Because of this rule, 3.1416(34) unambiguous.
[Alternate notation.]

1.3, cont.

Could also report systematic error:

$$\underbrace{3.19 \pm 0.01}_{\text{statistical}} \underbrace{\pm 0.03}_{\text{systematic}}$$

Difficulty in estimating systematic errors mean this is often only reported in conclusion of experiment.

1.4 What is meant by statistical uncertainty?

It is difficult (often poorly defined) to give the precision of a single measurement. Instead, we typically repeat measurements many times, to eliminate (or measure) any random influences, and get a good idea of how *repeatable* a measurement is.

1.4, cont.

Given a series or list of measurements

$x_1 = 3.3$, $x_2 = 3.8$, $x_3 = 3.0$, $x_4 = 3.2$, $x_5 = 3.3$, etc.

We can calculate the mean (or average)

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

where N is the number of measurements we made, and the standard deviation σ, where

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$ population standard deviation

$$\boxed{s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$ sample standard deviation

How does this relate to our problem?

The st. dev. is the uncertainty of a single measurement, and the mean is the best guess of the "true" value μ.

[We'll use σ and s interchangably in these notes. Typically, since we don't know μ, the "sample" definition is the correct one to use.]

1.5 Error propagation

Let's say we want to determine a ~~function q~~ quantity q in our experiment, which is a function of ~~many~~ two variables:

$$q = q(x, y)$$

we measure x and y in several runs of the experiment, and want to know what the uncertainty of q is.

Answer: $$\boxed{\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2 + 2 \frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\sigma_{xy}}$$

where • $\frac{\partial q}{\partial x}$ & $\frac{\partial q}{\partial y}$ are partial derivatives evaluated at $\bar{y}$ & $\bar{x}$, ~~respectively~~

• σ_x is std dev (or uncertainty) in x:
[ie, $\overset{m}{x} = \bar{x} \pm \sigma_x$ from previous calculation]

• σ_{xy} is correlation of x & y, defined as $\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$, for N measurements, but 0 if not measured simultaneously.

1.5, cont'd

Proof: $\sigma_q^2 = \frac{1}{N-1}\sum\left[q_i-\bar{q}\right]^2$, std deviation.

Take linear expansion (Taylor):

$$q_i \approx q(\bar{x},\bar{y}) + \left.\frac{\partial q}{\partial x}\right|_{\substack{y=\bar{y}\\ x=\bar{x}}}(x_i-\bar{x}) + \left.\frac{\partial q}{\partial y}\right|_{\bar{x},\bar{y}}(y_i-\bar{y})$$

Then:

$$\sigma_q^2 = \frac{1}{N-1}\sum\left[\underbrace{q(x,\bar{y}) - \bar{q}} + \frac{\partial q}{\partial x}(x_i-\bar{x}) + \frac{\partial q}{\partial y}(y_i-\bar{y})\right]^2$$

Assume $\bar{q} = q(\bar{x},\bar{y})$, so 1st 2 terms cancel.

Now:

$$\sigma_q^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\frac{\partial q}{\partial x}(x_i-\bar{x}) + \frac{\partial q}{\partial y}(y_i-\bar{y})\right]^2$$

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

$$+ \left(\frac{\partial q}{\partial y}\right)^2\frac{1}{N-1}\sum(y_i-\bar{y})^2$$

$$+ 2\left(\frac{\partial q}{\partial x}\right)\left(\frac{\partial q}{\partial y}\right)\underbrace{\frac{1}{N-1}\sum(x_i-\bar{x})(y_i-\bar{y})}_{=\sigma_{xy}}$$

QED.

Note: σ_{xy} has units of xy

Sometimes called σ_{xy}^2, but might be <0!

1.5, cont.

case I.

If variables x & y are not correlated, then $\sigma_{xy} = 0$, &

$$\boxed{\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2}$$

which probably looks familiar. "Uncorrelated errors are added in quadrature."

Note that if $q = x \pm y$, then

$$\left(\frac{\partial q}{\partial x}\right)^2 = 1 \quad \text{and} \quad \left(\frac{\partial q}{\partial y}\right)^2 = 1,$$

so $\boxed{\sigma_q^2 = \sigma_x^2 + \sigma_y^2}$ "for addition & subtraction."

case II. More simple formulae? Taylor Ch. 3

In the absence of knowledge about correlations, we might assume the worst case: From the Schwarz Inequality, one can show that

$$|\sigma_{xy}| \leq \sigma_x \sigma_y .$$

1.5, case II, cont.

because of S. Inequality

So $\left|\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\right|\sigma_x\sigma_y \geq \left|\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\right||\sigma_{xy}|$

$$\geq \left(\frac{\partial q}{\partial x}\right)\left(\frac{\partial q}{\partial y}\right)\sigma_{xy}$$

because the absolute value of a number is greater than or equal to the original Number.

Thus: $\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2\sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2\sigma_y^2 + 2\left(\frac{\partial q}{\partial x}\right)\left(\frac{\partial q}{\partial y}\right)\sigma_{xy}$

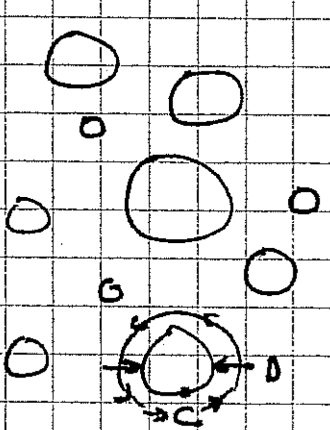
$$\leq \left(\frac{\partial q}{\partial x}\right)^2\sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2\sigma_y^2 + 2\left|\frac{\partial q}{\partial x}\cdot\frac{\partial q}{\partial y}\right|\sigma_x\sigma_y$$

$$= \left[\left|\frac{\partial q}{\partial x}\right|\sigma_x + \left|\frac{\partial q}{\partial y}\right|\sigma_y\right]^2$$

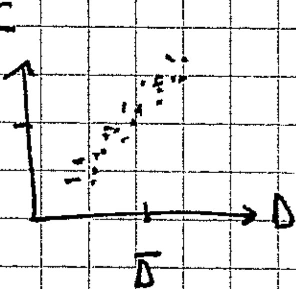
$$\longrightarrow \boxed{\sigma_q \leq \left|\frac{\partial q}{\partial x}\right|\sigma_x + \left|\frac{\partial q}{\partial y}\right|\sigma_y}$$

This is the "worst case", where all errors fluctuate together. In the absence of other info, you can use this as a "safe" error estimate.

1.6 Examples



Suppose a variety of circles are given to the experimenter, who measures their diameter and circumference.



Data is taken, and both mean and standard deviation are calculated for each quantity: $\bar{D} = \cancel{4.7}\ \text{cm}$ 4.80 $\quad \sigma_D = \cancel{1.3\ \text{cm}}$ 1.74

(Sorry about the mess — see next page.) $\bar{C} = \cancel{14.3}\ \text{cm}$ 14.92 $\quad \sigma_c = \cancel{4.0}\ \text{cm}$ 5.10

The plot shows these variables are correlated, but ~~ignoring~~ forgetting the correlations one would calculate

$$\pi = \frac{C}{D} \rightarrow \frac{\partial \pi}{\partial C} = \frac{1}{D}, \quad \frac{\partial \pi}{\partial D} = -\frac{C}{D^2},$$

So

$$\sigma_\pi^2 = \left(\frac{1}{\bar{D}}\right)^2 \sigma_c^2 + \left(\frac{\bar{C}}{\bar{D}^2}\right)^2 \sigma_D^2$$

$$\sigma_\pi = \sqrt{\left(\frac{\cancel{1.3}\,4.0\ \text{cm}}{4.7\ \text{cm}}\right)^2 + \left(\frac{14.3\ \text{cm}\cdot \cancel{4.0}\,1.3\ \text{cm}}{(4.7\ \text{cm})^2}\right)^2} = \cancel{2.6}\ \cancel{1.2}\ 1.55$$

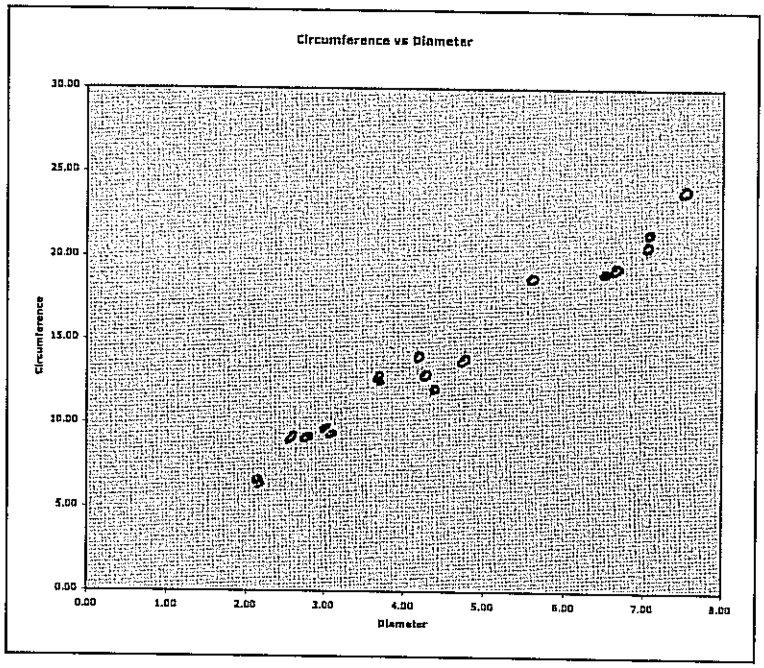
$\bar{\pi} = 3.11 \longrightarrow \cancel{3.11 \pm 1.55}$

3.1 ± 1.5 Awful.

D C $\pi \equiv \frac{C}{D}$

16 measurements.

diameter	circ	ratio
4.23	13.13	3.10
7.05	20.61	2.92
2.78	9.20	3.31
4.35	12.10	2.78
2.14	6.69	3.13
6.62	19.42	2.93
7.50	24.19	3.23
6.54	19.20	2.94
4.18	14.14	3.38
4.76	14.01	2.94
5.58	18.79	3.37
4.77	14.11	2.96
7.03	21.36	3.04
2.59	9.36	3.61
3.67	12.69	3.46
3.03	9.64	3.18



	diameter	circ	correlation			
mean	4.80	14.91				
stdev	1.74	5.10	8.75	ratio		
					3.11	
			uncertainty (naive)		1.55	-0.02
			uncertainty (correct)		0.21	-0.2 sigma off

See section 1.8

If use st dev mean:

	diameter	circ	ratio	unc	
mean	4.801	14.915	3.106	0.053	-0.7 sigma off
stdev	0.436	1.276			

1.6, cont.

With correlations, calculated to be,
one instead finds

$$\sigma_\pi^2 = \left(\frac{1}{\bar{D}}\right)^2 \sigma_C^2 + \left(\frac{\bar{C}}{\bar{D}^2}\right)^2 \sigma_D^2 + 2\left(\frac{1}{\bar{D}}\right)\left(\frac{-\bar{C}}{\bar{D}^2}\right) \sigma_{CD}$$

8.75

$$\longrightarrow \quad \sigma_\pi = 0.21$$

$$\longrightarrow \quad \pi = \bar{\pi} \pm \sigma_\pi$$

$$= 3.11 \pm 0.21$$ more reasonable uncertainty.

Example 2: Of course in many cases we don't take paired measurements, so harder to define these quantities. ~~Here~~
If the diameter of a single circle was measured to be $D = 3.03 \pm 0.08$, and circumference $C = 9.64 \pm 0.09$, then don't expect any correlation because the measurements were independent.

Thus $\sigma_\pi^2 = \dfrac{\sigma_C^2}{\bar{D}^2} + \dfrac{\bar{C}^2}{\bar{D}^4}\sigma_D^2 \longrightarrow \pi = 3.18 \pm 0.09$

1.7 Normal distribution

Ch. 5 Taylor
Ch. 1,2 Bevington

Different types of measurements have different distributions characterizing the spread in results. A "limiting distribution" or "parent distribution" is the probability distribution approached in the limit of an infinite number of measurements. This is to be contrasted with the "sample distribution", which is your data.

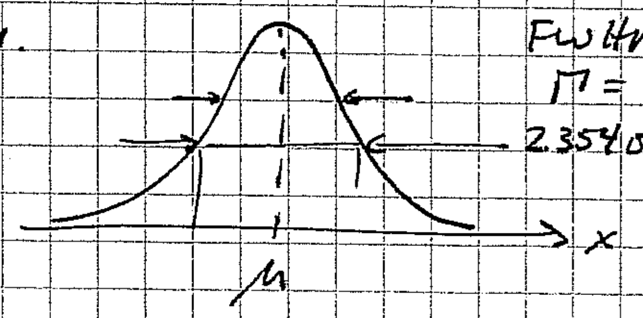
The Central Limit Theorem states that in the presence of many small sources of random error, the probability of measuring a value x between x and x+dx is

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

where μ = "true" mean,

σ = "true" deviation.

This curve is also called a "Gaussian."



1.7, cont.

So, what does σ mean?

P is defined such that $\int_{-\infty}^{\infty} P(x)\,dx = 1$,
so the ~~sum of all measurements~~ probability of making a measurement at any value is 100%. (Logical!)

The probability of making a measurement within σ of μ is

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{\mu-\sigma}^{\mu+\sigma} dx \, \exp\left[-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right] = 0.68$$

meaning 68% of all measurements are within σ of μ.

This is an important ~~meaning of~~ point about statistical error: there is nothing "wrong" with a single measurement outside the rms error! In fact, there would be something wrong if none of your measurements were outside the quoted σ_x: About 1/3 ought to be. Look back @ example from §1.6

1.8 Weighted Average and Standard Deviation of the Mean.

Ch. 7 Taylor

Let's say you've made several measurements of the same quantity, with varying amounts of data and precision:

try 1: $x_1 \pm \sigma_1$

try 2: $x_2 \pm \sigma_2$

etc., for N tries.

How do you combine these, to choose the best value (μ), and what is the uncertainty with which you have determined the "best" value?

Approach: Maximize the probability of the observations having occurred. Say (assume) each measurement describes a Gaussian distribution. Then $P = P_1(x_1)\,P_2(x_2) \cdots P(x_N)$,

where $P_1(x_1) = \dfrac{1}{\sigma_1\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\dfrac{x_1-\mu}{\sigma_1}\right)^2\right\}$, etc.

thus,

$$P = (2\pi)^{-N/2} (\sigma_1 \cdots \sigma_N) \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma_i}\right)^2\right\}$$

To maximize P, we set its derivative to zero, so $\frac{dP}{d\mu}=0 \rightarrow \frac{d}{d\mu}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma_i}\right)^2=0$

So $-2\sum_{i=1}^{N}\frac{1}{\sigma_i^2}(x_i-\mu)=0$

Can drop -2 & split terms:

$$\sum\frac{x_i}{\sigma_i^2}=\mu\sum\frac{1}{\sigma_i^2} \rightarrow \boxed{\mu_{best}=\frac{\sum\frac{1}{\sigma_i^2}\cdot x_i}{\sum\frac{1}{\sigma_i^2}}}$$

In other words, the best guess for x is a weighted sum, where each measurement is weighted by $\frac{1}{\sigma^2}$.

What is the uncertainty?

$$\sigma_q^2=\sum_i\left(\frac{\partial q}{\partial x_i}\right)^2\sigma_{x_i}^2$$ if uncorrelated, so

$$\frac{\partial\mu_{best}}{\partial x_i}=\frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2} \quad \ldots \quad \boxed{\sigma_{\mu_{best}}=\frac{1}{\sqrt{\sum_i 1/\sigma_i^2}}}$$

1.8, cont.

If all σ_i the same, then $\sum_i 1/\sigma_i^2 = N/\sigma^2$

and $\mu_{best} = \frac{\sum \frac{1}{\sigma^2} x_i}{\sum \frac{1}{\sigma^2}} = \frac{\frac{1}{\sigma^2} \sum_i x_i}{N \cdot \frac{1}{\sigma^2}} = \frac{1}{N} \sum x_i$

In other words, $\boxed{\mu_{best} = \overline{x}}$

The uncertainty is then

$$\sigma_{\mu_{best}} = \frac{1}{\sqrt{N/\sigma^2}} = \boxed{\frac{\sigma}{\sqrt{N}}}$$

Hmm.. this is not what we said before. Instead of σ as the error ~~in~~ our measurements, we have $\sigma/\sqrt{N}$, where N is the number of measurements made. This is called the "standard error of the mean", and is the best estimation of the uncertainty in a calculated average, so long as your errors are truly statistical!

$$\boxed{S_{\overline{x}}^2 = \frac{1}{N(N-1)} \sum_{i=1}^{N} (x_i - \overline{x})^2}$$

Part 2: The least-squares method of fitting data

Outline:

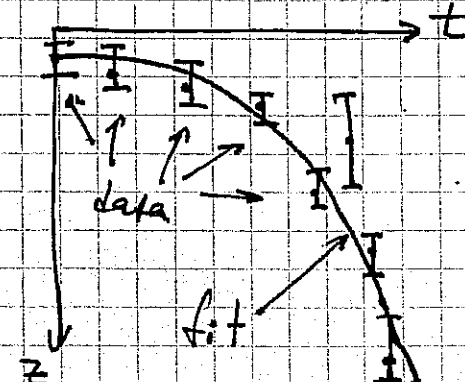
- 2.1 Fitting a function to data
- 2.2 Definition of χ^2
- 2.3 Finding the best fit
- 2.4 Uncertainty in fit parameters
- 2.5 Linear and polynomial fit solutions
- 2.6 Including x uncertainty
- 2.7 Goodness of fit
- 2.8 Summary

2.1 Fitting a function to data

As often as possible we try to make direct measurements of the quantities we are interested in. For instance, measuring the length of an object, we use a ruler, and no fitting or analysis may be required.

Unfortunately, not all quantities can be measured directly. Instead, we may have to use a model to infer the value we are interested in.

For instance, to figure out gravitational acceleration, we might measure the height of a falling object at various times, after release. We expect that



$$z = -\frac{1}{2} g t^2$$

and have measured many data points

$$\boxed{\begin{matrix} t_1 & z_1 & \sigma_{z_1} \\ t_2 & z_2 & \sigma_{z_2} \\ \vdots & \vdots & \vdots \end{matrix}}$$

2.1, cont

In this lecture we will try to answer the questions:

- What is the value of g indicated by our measurements?
- What is the uncertainty with which we have determined g?
- Is the model consistent with the data, or have we neglected something (eg, drag, initial velocity, etc.)

For most of these notes we will ignore the uncertainty in the independent variable assuming it can be chosen very accurately. We will extend our methods to include "x" uncertainty in §2.6.

We'll typically refer to our data set as
$\{x_i, y_i, \sigma_i\}$
and the fit function as $f(x)$ or $y(x)$, interchangeably.

2.2 Definition of χ^2

Given a set of data and a fit, how do we quantify how "close" or "far" they are from one another?

At each x_i, we measured $y_i \pm \sigma_i$, and calculate that we expect $f(x_i)$. Since σ_i tells us about the expected (average) deviation, the distance $\Delta y = y_i - f(x_i)$ can be given in terms of σ_i. Summing over all points, we get a single number

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2 \qquad \text{"Chi Squared"}$$

pronounced "ki"

Note that χ^2 is dimensionless.

There are N data points in our sum. (We won't always write the start & end index, but it will typically be from 1 to N.)

2.3 Finding the best fit

Using the maximum likelihood method, we choose the "best fit" parameters to maximize the probability that our particular set of observations occurred.

If you believe that $y(x_i)$ is the 'true' value of the dependent variable at x_i, then the probability of measuring y_i was

$$P_i = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(y_i - y(x_i))^2}{\sigma_i^2}\right]$$

The joint probability of all measurements having been made is then

$$P = P_1 P_2 \cdots P_N = \prod_{i=1}^{N} P_i$$

$$= \left(\prod_i \frac{1}{\sigma_i \sqrt{2\pi}}\right) \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i - y(x_i)}{\sigma_i}\right)^2\right]$$

Maximizing this probability means finding its stationary points with respect to the fit parameters. We've been writing $y(x)$ for convenience, but in fact the function also depends on the parameters $a_1, a_2, a_3, a_4, \ldots$:

$$y(x) = f(x, a_1, a_2, a_3 \ldots a_m)$$

where m is the number of fit parameters.

2.3, cont.

Now we can write the condition for maximizing P as:

$$\frac{\partial P}{\partial a_j} = 0 \quad \text{for} \quad j = 1 \text{ to } m$$

Looking back, we see that

$$P = \underbrace{\text{constant}}_{\substack{\text{doesn't depend} \\ \text{on } \{a_j\}}} \times \exp\left[-\frac{1}{2}\chi^2\right]$$

does depend on $\{a_j\}$ because $f(x, \{a_j\})$ inside

& thus

$$\frac{\partial P}{\partial a_j} = \text{const.} \cdot \frac{\partial}{\partial a_j} \exp\left[-\frac{1}{2}\chi^2\right]$$

$$= P \cdot \left(-\frac{1}{2}\right) \cdot \frac{\partial}{\partial a_j}\chi^2 = 0 \quad @ \text{ max}$$

$\longrightarrow$ $\boxed{\dfrac{\partial \chi^2}{\partial a_j} = 0 \quad \text{for all } j = 1 \text{ to } m \quad \text{at best fit.}}$

m simultaneous equations to solve.

In fact, this turns out to be minimizing χ^2.

2.4 Uncertainty in fit parameters

For a sufficiently large data set, we can apply the central limit theorem, and describe the distribution of fit parameters by a Gaussian distribution about the best fit. Let's consider first the example of a one-parameter function

$$y(x_i) = f(x_i, a)$$

Now $P(a) = \text{const} \times \exp\left[-\frac{1}{2}\frac{(a-a^*)^2}{\sigma_a^2}\right]$,

where a^* is the best fit value and σ_a is its uncertainty.

We can also write

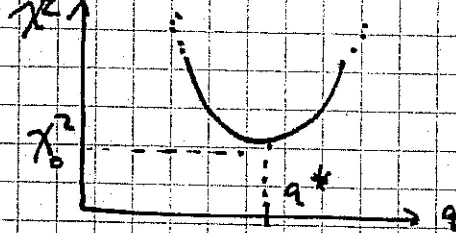
$$\ln P(a) = \ln(\text{const}) + \left(-\frac{1}{2}\right)\left(\frac{a-a^*}{\sigma_a}\right)^2$$

Comparing this to P from §2.3,

$$\ln P = \ln\left(\begin{matrix}\text{another}\\ \text{constant}\end{matrix}\right) + \left(-\frac{1}{2}\right)\chi^2$$

we see that

$$\chi^2 = \frac{(a-a^*)^2}{\sigma_a^2} + C$$



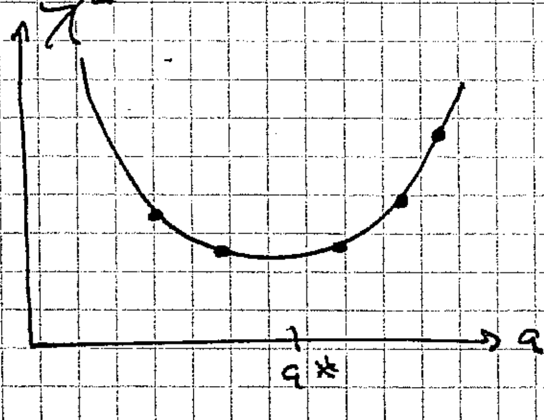
where C is yet another constant. Actually, we know what C must be: at $a = a^*$, χ is at its best fit value χ_0^2. More or less, this tells us that χ^2 varies quadratically about $a = a^*$, with curvature

2.4, continued

$$\frac{d^2 \chi}{da^2} = \frac{2}{\sigma_a^2}$$

which tells us that $\boxed{\sigma_a^2 = 2\left(\frac{d^2\chi}{da^2}\right)^{-1}}$

for a one-parameter fit.



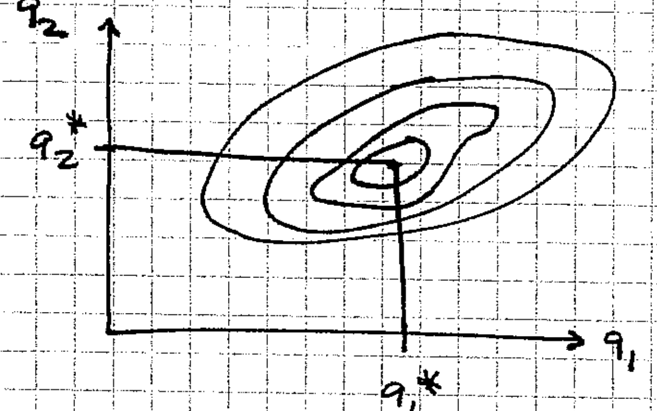
This immediately suggests a method of evaluating fit uncertainty σ_a in a program: calculate various values of $\chi^2(a)$ around a^*, and do another fit to find the curvature!

What about when χ^2 depends on ~~many~~ more than one parameter? Instead of taking ~~m~~ the inverse of the curvature of χ^2, you need to take the inverse of a matrix whose elements ~~have~~ are curvatures & gradients of χ^2. This is written out in the next few pages, and proved in Orear (1982).

Variation of χ^2 about its minimum.

In general the variation of χ^2 about its best fit value (its minimum) can be evaluated numerically, and may look something like:

(Shown: contours of equal χ^2)



where q_1 & q_2 are fit parameters, and q_1^* and q_2^* are minimal values for χ^2.

We can expand χ^2 around this minimum:

Since the solutions $q_i = q_i^*$ are given by finding the stationary point of χ^2, ie

$$\frac{\partial \chi^2}{\partial q_i} = 0 \quad \text{for all } q_i$$

we can expect that χ^2 will be a quadratic function of parameters. A Taylor expansion is given in general by

$$y(x_i) = y(x_i^*) + \sum_i \left.\frac{\partial y}{\partial x_i}\right|_{x_i^*} (x_i - x_i^*) + \frac{1}{2}\sum_i \sum_j \frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}(x_i - x_i^*) \times (x_j - x_j^*)$$

$$+ \text{ higher order.}$$

2-4, cont.

We know $\frac{\partial \chi^2}{\partial a_i} = 0$ @ $a_i = a_i^*$, so

$$\chi^2 = \chi^2(a_1^*, a_2^*, \ldots a_n^*) + \frac{1}{2}\sum_i \sum_j \left.\frac{\partial \chi^2}{\partial x_i}\right|_{x_i^*} \left.\frac{\partial \chi^2}{\partial x_j}\right|_{x_j^*} (x_i - x_i^*)(x_j - x_j^*)$$

Defining the curvature matrix

$$\boxed{\alpha_{ij} = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial a_i \partial a_j}}$$

One can show that [Orear, 1982, unpublished] the error in the fit parameters is given by

$$\boxed{\sigma_{a_i}^2 = \epsilon_{ii}}$$

and the correlation in error between a_i and a_j is

$$\boxed{\sigma_{a_i a_j} = \epsilon_{ij}}$$

where ϵ is the error matrix, defined as

$$\boxed{\epsilon \equiv \alpha^{-1}} \quad \text{(matrix inverse)}$$

$$\left(\text{or} \quad \sum_k \epsilon_{ik}\alpha_{kj} = \delta_{ij}\right)$$

(10)

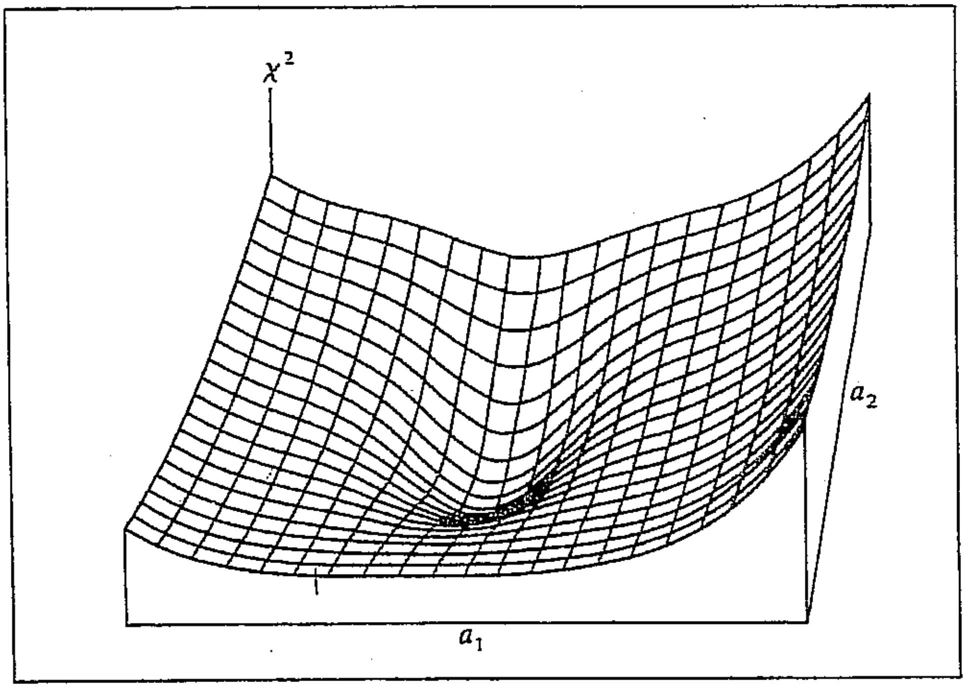


FIGURE 8.2
Chi-square hypersurface as a function of two parameters.

Source: Bevington.

2.5 Linear Fit Solution

For simple functions, one can calculate the best fit directly from the data set, without the need of a nonlinear minimization routine.

The following excerpts from Bevington show the derivation of best fit parameters for a linear fit

$$y = a + bx$$

to a data set including only y errors:

$$\{x_i, y_i, \sigma_{y_i}\}$$

$(6.12) \longrightarrow$ best fit $a = a^*$, $b = b^*$

$(6.21) \longrightarrow \sigma_a^2$

$(6.22) \longrightarrow \sigma_b^2$

Now you can write your own fitting program!

(12)

6.3 MINIMIZING χ^2

To find the values of the coefficients a and b that yield the minimum value for χ^2, we set to zero the partial derivatives of χ^2 with respect to each of the parameters

$$\begin{aligned}
\frac{\partial}{\partial a}\chi^2 &= \frac{\partial}{\partial a}\sum\left[\frac{1}{\sigma_i^2}(y_i - a - bx)^2\right] \\
&= -2\sum\left[\frac{1}{\sigma_i^2}(y_i - a - bx_i)\right] = 0 \\
\frac{\partial}{\partial b}\chi^2 &= \frac{\partial}{\partial b}\sum\left[\frac{1}{\sigma_i^2}(y_i - a - bx)^2\right] \\
&= -2\sum\left[\frac{x_i}{\sigma_i^2}(y_i - a - bx_i)\right] = 0
\end{aligned} \tag{6.10}$$

These equations can be rearranged as a pair of linear simultaneous equations in the unknown parameters a and b:

$$\begin{aligned}
\sum\frac{y_i}{\sigma_i^2} &= a\sum\frac{1}{\sigma_i^2} + b\sum\frac{x_i}{\sigma_i^2} \\
\sum\frac{x_i y_i}{\sigma_i^2} &= a\sum\frac{x_i}{\sigma_i^2} + b\sum\frac{x_i^2}{\sigma_i^2}
\end{aligned} \tag{6.11}$$

The solutions can be found in any one of a number of different ways, but, for generality we shall use the method of determinants. (See Appendix B.) The solutions are

$$\begin{aligned}
a &= \frac{1}{\Delta}\begin{vmatrix} \sum\frac{y_i}{\sigma_i^2} & \sum\frac{x_i}{\sigma_i^2} \\ \sum\frac{x_i y_i}{\sigma_i^2} & \sum\frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta}\left(\sum\frac{x_i^2}{\sigma_i^2}\sum\frac{y_i}{\sigma_i^2} - \sum\frac{x_i}{\sigma_i^2}\sum\frac{x_i y_i}{\sigma_i^2}\right) \\
b &= \frac{1}{\Delta}\begin{vmatrix} \sum\frac{1}{\sigma_i^2} & \sum\frac{y_i}{\sigma_i^2} \\ \sum\frac{x_i}{\sigma_i^2} & \sum\frac{x_i y_i}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta}\left(\sum\frac{1}{\sigma_i^2}\sum\frac{x_i y_i}{\sigma_i^2} - \sum\frac{x_i}{\sigma_i^2}\sum\frac{y_i}{\sigma_i^2}\right) \\
\Delta &= \begin{vmatrix} \sum\frac{1}{\sigma_i^2} & \sum\frac{x_i}{\sigma_i^2} \\ \sum\frac{x_i}{\sigma_i^2} & \sum\frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \sum\frac{1}{\sigma_i^2}\sum\frac{x_i^2}{\sigma_i^2} - \left(\sum\frac{x_i}{\sigma_i^2}\right)^2
\end{aligned} \tag{6.12}$$

Bevington, pp. 103-104

For the special case in which all the uncertainties are equal ($\sigma = \sigma_i$), they cancel and the solutions may be written

$$\begin{aligned}
a &= \frac{1}{\Delta'}\begin{vmatrix} \Sigma y_i & \Sigma x_i \\ \Sigma x_i y_i & \Sigma x_i^2 \end{vmatrix} = \frac{1}{\Delta'}(\Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma x_i y_i) \\
b &= \frac{1}{\Delta'}\begin{vmatrix} N & \Sigma y_i \\ \Sigma x_i & \Sigma x_i y_i \end{vmatrix} = \frac{1}{\Delta'}(N\Sigma x_i y_i - \Sigma x_i \Sigma y_i) \\
\Delta' &= \begin{vmatrix} N & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{vmatrix} = N\Sigma x_i^2 - (\Sigma x_i)^2
\end{aligned} \tag{6.13}$$

13

Uncertainties in the Parameters

In order to find the uncertainty in the estimation of the coefficients a and b in our fitting procedure, we use the error propagation method discussed in Chapter 3. Each of our data points y_i has been used in the determination of the parameters and each has contributed some fraction of its own uncertainty to the uncertainty in our final determination. Ignoring systematic errors, which would introduce correlations between uncertainties, the variance σ_z^2 of the parameter z is given by Equation (3.10) as the sum of the squares of the products of the standard deviations σ_i of the data points with the effects that the data points have on the determination of z:

$$\sigma_z^2 = \sum\left[\sigma_i^2\left(\frac{\partial z}{\partial y_i}\right)^2\right] \tag{6.19}$$

Thus, to determine the uncertainties in the parameters a and b, we take the partial derivatives of Equation (6.12):

$$\begin{aligned}\frac{\partial a}{\partial y_j} &= \frac{1}{\Delta}\left(\frac{1}{\sigma_j^2}\sum\frac{x_i^2}{\sigma_i^2} - \frac{x_j}{\sigma_j^2}\sum\frac{x_i}{\sigma_i^2}\right)\\ \frac{\partial b}{\partial y_j} &= \frac{1}{\Delta}\left(\frac{x_j}{\sigma_j^2}\sum\frac{1}{\sigma_i^2} - \frac{1}{\sigma_j^2}\sum\frac{x_i}{\sigma_i^2}\right)\end{aligned} \tag{6.20}$$

We note that the derivatives are functions only of the variances and of the independent variables x_i. Combining these equations with the general expression of Equation (6.19) and squaring, we obtain for σ^2,

$$\begin{aligned}\sigma_a^2 &\simeq \sum_{j=1}^{N}\frac{\sigma_j^2}{\Delta^2}\left[\frac{1}{\sigma_j^4}\left(\sum\frac{x_i^2}{\sigma_i^2}\right)^2 - \frac{2x_j}{\sigma_j^4}\sum\frac{x_i^2}{\sigma_i^2}\sum\frac{x_i}{\sigma_i^2} + \frac{x_j^2}{\sigma_j^4}\left(\sum\frac{x_i}{\sigma_i^2}\right)^2\right]\\ &= \frac{1}{\Delta}\left[\sum\frac{1}{\sigma_j^2}\left(\sum\frac{x_i^2}{\sigma_i^2}\right)^2 - 2\sum\frac{x_j}{\sigma_j^2}\sum\frac{x_i^2}{\sigma_i^2}\sum\frac{x_i}{\sigma_i^2} + \sum\frac{x_j^2}{\sigma_j^2}\left(\sum\frac{x_i}{\sigma_i^2}\right)^2\right]\\ &= \frac{1}{\Delta^2}\left(\sum\frac{x_i^2}{\sigma_i^2}\right)\left[\sum\frac{1}{\sigma_j^2}\sum\frac{x_i^2}{\sigma_i^2} - \left(\sum\frac{x_i}{\sigma_i^2}\right)^2\right]\\ &= \frac{1}{\Delta}\sum\frac{x_i^2}{\sigma_i^2}\end{aligned} \tag{6.21}$$

14

Bevington, pp. 108-109

and for σ_b^2,

$$\begin{aligned}\sigma_b^2 &\simeq \sum_{j=1}^{N}\frac{\sigma_j^2}{\Delta^2}\left[\frac{1}{\sigma_j^4}\left(\sum\frac{1}{\sigma_i^2}\right)^2 - \frac{2x_j}{\sigma_j^4}\sum\frac{1}{\sigma_i^2}\sum\frac{x_i}{\sigma_i^2} + \frac{1}{\sigma_j^4}\left(\sum\frac{x_i}{\sigma_i^2}\right)^2\right]\\ &= \frac{1}{\Delta^2}\left[\sum\frac{1}{\sigma_j^2}\left(\sum\frac{1}{\sigma_i^2}\right)^2 - 2\sum\frac{x_j}{\sigma_j^2}\sum\frac{1}{\sigma_i^2}\sum\frac{x_i}{\sigma_i^2} + \sum\frac{1}{\sigma_j^2}\left(\sum\frac{x_i}{\sigma_i^2}\right)^2\right]\\ &= \frac{1}{\Delta^2}\left(\sum\frac{1}{\sigma_i^2}\right)\left[\sum\frac{1}{\sigma_j^2}\sum\frac{1}{\sigma_i^2} - \left(\sum\frac{x_i}{\sigma_i^2}\right)^2\right]\\ &= \frac{1}{\Delta}\sum\frac{1}{\sigma_i^2}\end{aligned} \tag{6.22}$$

For the special case of common uncertainties in y_i, $\sigma_i = \sigma$, these equations reduce to

$$\sigma_a^2 = \frac{\sigma^2}{\Delta'}\Sigma x_i^2 \quad\text{and}\quad \sigma_b^2 = N\frac{\sigma^2}{\Delta'} \tag{6.23}$$

with σ given by Equation (6.15) and Δ' given by Equation (6.13).

The uncertainties in the parameters σ_a and σ_b, calculated from the original error estimates, are listed in Tables 6.1 and 6.2. For Example 6.1, revised uncertainties σ_a' and σ_b', based on the revised common data uncertainty calculated from Equation (6.18), are also listed.

2.5 Polynomial fit solution

It turns out than any function which is linear in its parameters $\{a_j\}$ can be fit analytically, even when nonlinear in the measured values $\{x_i\}$. For instance, the following pages 121-123 from Bevington fit to

$$y(x) = \sum_{k=1}^{m} a_k f_k(x)$$

where $f_k(x)$ must not depend on any of the $\{a_j\}$. Basically this is polynomial fitting:

$$\text{e.g.,} \quad y(x) = a_1 + a_2 x + a_3 x^2$$

(7.19) → best fit values
(7.25) → uncertainty in fit values.

Note that the error matrix ϵ_{ij} is shown to give the parameter uncertainty here, which we stated without proof previously.

In the example given, note than simply using the inverse curvature $\frac{1}{\alpha_{11}}$ as the error on a_1 underestimates $\sigma_{a_1}^2$ by about a factor of 3. (cf ϵ_{11}, the correct $\sigma_{a_1}^2$)

Also note how (7.27) uses the covariances given by the error matrix ϵ_{ij} !!!

(7)

7.1, we obtain for an estimate of the variance,

$$\sigma'^2 = \sigma^2 \times \chi^2/(N-n) = 0.05 \times 26.6/18 = 0.06°\text{C}$$

suggesting, perhaps, that the student slightly underestimated the uncertainty in her measurements of V.

7.2 MATRIX SOLUTION

The techniques of least-squares fitting fall under the general name of regression analysis. Because we have been considering only problems in which the fitting function

$$y(x_i) = \sum_{k=1}^{m} a_k f_k(x_i) \tag{7.12}$$

$f_k(x) = x^{k-1}$, for instance.

is linear in the *parameters* a_k, we are considering only linear regression or multiple linear regression, usually shortened to multiple regression. In Chapter 8 we shall deal with techniques for handling problems with fitting functions that are not linear in the parameters.

Matrix Equations

We have not yet determined the uncertainties in the three coefficients we obtained when we fitted the second-order equation to the data of Example 7.1. We could fit the uncertainties by extending the method used for the linear fits of Examples 6.1 and 6.2. However, the algebra becomes even more tedious as the number of terms in the fitted equation increases, and in fact, our method only yielded estimates of the variances σ_k^2 and not of the covariances σ_{kl}^2, which are often important for fitted parameters. Rather than pursue the determinant method, we shall discuss immediately the more elegant and general matrix method for solving the multiple regression problem. Some of the properties of matrices are discussed in Appendix B.

Equations (7.7) can be expressed in matrix form as the equivalence between a row matrix $\boldsymbol{\beta}$ and the product of a row matrix $\mathbf{a}$ with a symmetric matrix $\boldsymbol{\alpha}$, all of order m:

$$\boldsymbol{\beta} = \mathbf{a}\boldsymbol{\alpha} \tag{7.13}$$

The elements of the row matrix $\boldsymbol{\beta}$ are defined by

$$\beta_k \equiv \sum \left[\frac{1}{\sigma_i^2} y_i f_k(x_i)\right] \tag{7.14}$$

those of the symmetric matrix $\boldsymbol{\alpha}$ by

$$\alpha_{lk} \equiv \sum \left[\frac{1}{\sigma_i^2} f_l(x_i) f_k(x_i)\right] \tag{7.15}$$

and the elements of the row matrix $\mathbf{a}$ are the parameters of the fit. For $m = 3$,

(18)

the matrices may be written as

$$\boldsymbol{\beta} = [\beta_1 \quad \beta_2 \quad \beta_3] \qquad \mathbf{a} = [a_1 \quad a_2 \quad a_3] \tag{7.16}$$

and

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix} \tag{7.17}$$

To solve for the parameter matrix $\mathbf{a}$ we multiply both sides of Equation (7.13) on the right by the inverse $\boldsymbol{\epsilon}$ of the matrix $\boldsymbol{\alpha}$, defined such that $\boldsymbol{\alpha\epsilon} = \boldsymbol{\alpha\alpha}^{-1} = 1$, the unity matrix. We obtain

$$\boldsymbol{\beta\epsilon} = \mathbf{a}\boldsymbol{\alpha\epsilon} = \mathbf{a} \tag{7.18}$$

which gives

$$\mathbf{a} = \boldsymbol{\beta\epsilon} = \boldsymbol{\beta\alpha}^{-1} \tag{7.19}$$

Equation (7.19) can also be expressed as

$$a_l = \sum_{k=1}^{m} (\beta_k \epsilon_{kl}) = \sum_{k=1}^{m} \left\{\epsilon_{kl} \sum \left[\frac{1}{\sigma_i^2} y_i f_k(x_i)\right]\right\} \tag{7.20}$$

where the β_k's are given by Equation (7.16).

The solution of Equation (7.19) requires that the matrix $\boldsymbol{\alpha}$ be inverted. This generally is not a simple procedure, except for matrices of very low order, but computer routines are readily available. The inversion of a matrix is discussed in Appendix B.

The symmetric matrix $\boldsymbol{\alpha}$ is called the *curvature matrix* because of its relationship to the curvature of the χ^2 function in parameter space. The relationship becomes apparent when we take the second derivatives of χ^2 with respect to the parameters. From Equation (7.6), we have for the partial derivative of χ^2 with respect to any arbitrary parameter a_l,

$$\frac{\partial \chi^2}{\partial a_l} = -2\sum \left\{\frac{f_l(x_i)}{\sigma_i^2}\left[y_i - \sum_{k=1}^{m} a_k f_k(x_i)\right]\right\} \tag{7.21}$$

and the second cross-partial derivative with respect to two such parameters is

$$\frac{\partial^2 \chi^2}{\partial a_l \partial a_k} = 2\sum \left[\frac{1}{\sigma_i^2} f_l(x_i) f_k(x_i)\right] = 2\alpha_{lk} \tag{7.22}$$

Estimation of Errors

The variance of $\sigma_{a_l}^2$ for the uncertainty in the determination of any parameter a_l is the sum of the variances of each of the data points σ_i multiplied by the square of the effect that each data point has on the determination of the parameter a_l [see Equation (6.19)]. Similarly, the covariance of two parameters

(19)

a_j and a_l is given by

$$\sigma_{a_j a_l}^2 = \sum \left[\sigma_i^2 \frac{\partial a_j}{\partial y_i} \frac{\partial a_l}{\partial y_i} \right] \tag{7.23}$$

(which also gives the variance for $j = l$), where we have assumed that there are no correlations between uncertainties in the measured variables y_i. Taking the derivatives in Equation (7.23) of a_l with respect to y_i we obtain

$$\frac{\partial a_l}{\partial y_i} = \sum_{k=1}^{m} \left[\epsilon_{lk} \frac{1}{\sigma_i^2} f_k(x_i) \right] \tag{7.24}$$

and, substituting into Equation (7.23), we obtain for the weighted sum of the squares of the derivatives,

$$\begin{aligned}
\sigma_{a_j a_l}^2 &= \sum \left\{ \sigma_i^2 \sum_{k=1}^{m} \left[\epsilon_{jk} \frac{1}{\sigma_i} f_k(x_i) \right] \sum_{p=1}^{m} \left[\epsilon_{lp} \frac{1}{\sigma_i} f_p(x_i) \right] \right\} \\
&= \sum_{k=1}^{m} \left\{ \epsilon_{jk} \sum_{p=1}^{m} \left[\epsilon_{lp} \sum \left(\frac{1}{\sigma_i^2} f_p(x_i) f_k(x_i) \right) \right] \right\} \\
&= \sum_{k=1}^{m} \left\{ \epsilon_{jk} \sum_{p=1}^{m} \left[\epsilon_{lp} \cdot \alpha_{pk} \right] \right\} \\
&= \sum_{k=1}^{m} \left[\epsilon_{kj} \cdot 1_{lk} \right] = \epsilon_{jl}
\end{aligned} \tag{7.25}$$

where we have switched the order of the sums over the dummy indices i, k, and l and have used the fact that because the curvature matrix α is symmetric, its inverse ϵ must also be symmetric, so that $\epsilon_{kj} = \epsilon_{jk}$. The elements of the unity matrix, which result from the summed products of the elements of α with its inverse ϵ, are represented by 1_{lk}.

The inverse matrix $\epsilon \equiv \alpha^{-1}$ is called the error matrix or the covariance matrix because its elements are the variances and covariances of the fitted parameters $\sigma_{a_j a_l} = \epsilon_{jl}$.

Example 7.2. The matrix method is illustrated by a straight-line fit $V = a_1 + a_2 T$ to a selection of data from Example 7.1. To show clearly each step of the calculation, we have selected just six points spaced at 25° intervals between 0 and 100° and have assumed a common uncertainty in the dependent variable $\sigma_V = 0.05$ mV. The data are listed in the columns 2 and 3 of Table 7.2*a*.

We begin by calculating each of the fitting functions $f_1 = 1$ and $f_2 = x$ at each value of the independent variable T. These are listed in columns 4 and 5 of Table 7.2*a*. For each measured value of x, the values of β_k, the elements of the column matrix β, and of α_{lk}, the elements of the symmetric matrix α, are calculated according to Equations (7.14) and (7.15). The individual terms in the

(20)

TABLE 7.2
Matrix solution for linear fit to data of Example 2†
(*a*) Data and components of matrix elements

i	T	V	$f_1(x_i)$	$f_2(x_i)$	β'_1	β'_2	α'_{11}	α'_{12}	α'_{22}	V_{fit}
1	0	−0.849	1	0	−339.6	0	400	0	0	−0.947
2	20	−0.196	1	20	−78.4	−1,458	400	8,000	160,000	−0.101
3	40	0.734	1	40	293.6	11,744	400	16,000	640,000	0.745
4	60	1.541	1	60	616.4	36,984	400	24,000	1,440,000	1.590
5	80	2.456	1	80	982.4	78,592	400	32,000	2,560,000	2.436
6	100	3.318	1	100	1327.6	132,720	400	40,000	4,000,000	3.281
					2802.0	258,472	2400	120,000	8,00,000	

(*b*) Matrices

$$\alpha = \begin{bmatrix} 2{,}400 & 120{,}000 \\ 120{,}000 & 8{,}800{,}000 \end{bmatrix} \qquad \epsilon = \begin{bmatrix} 1.310 \times 10^{-03} & -1.786 \times 10^{-05} \\ -1.786 \times 10^{-05} & 3.571 \times 10^{-07} \end{bmatrix}$$

$$\beta = [2{,}802 \quad 258{,}472] \qquad a = [-0.947 \quad 0.0423]$$

†The uniform uncertainty in V was assumed to be 0.05 mV as in Example 1. The columns labelled β'_1 and α'_{11}, etc., correspond to the individual contributions by each measured coordinate pair to the summed values of β and α. The value of χ^2 for the fit was 9.1 for 4 degrees of freedom corresponding to a probability of 5.5%.

calculation of β_1 and β_2 are listed in columns 6 and 7 of Table 7.2*a* and the individual terms in the calculation of α_{lk} are listed in columns 8 through 10. (We assume symmetry in α.) The resulting matrices are displayed in Table 7.2*b*.

The symmetric matrix α is inverted to obtain the variance matrix ϵ with elements ϵ_{kl}, shown in Table 7.2*b*, and the product matrix of the fitted parameters $\alpha = \beta\epsilon$ is calculated and displayed in Table 7.2*b*. The calculated values of the fitted variable V for each value of the independent variable T are listed in the last column of Table 7.2*a*.

Program 7.1. `MultRegr` Multiple regression problems are usually solved with the help of computer programs. The program `MultRegr` calls a set of routines for fitting any function that is linear in the parameters $a_1, a_2, \ldots, a_m$ to a set of N data points. Branches in the program on the global character variable `PAE` permit selection of the fitting function for each example in this chapter, with `PAE = 'P'` for the power series in x of Example 7.1. The program uses several program units in addition to some that were referred to in Chapter 6. All routines are listed in Appendix E.

`MakeAB7` Routines to set up the arrays `alpha` and `beta`, corresponding to the matrices α and β. These routines call the function `Funct` to calculate the individual terms in the fitting function.

`Matrix` The routines `MatInv`, to invert a matrix, and `LinearBySquare`, to find the product of a linear and a square matrix. Matrix manipulation is discussed in Appendix B.

TABLE 7.3
Error matrix from a fit by the matrix method to the data of Table 7.1†

$$\begin{bmatrix} 8.907 \times 10^{-04} & -3.473 \times 10^{-05} & 2.823 \times 10^{-07} \\ -3.473 \times 10^{-05} & 1.913 \times 10^{-06} & -1.783 \times 10^{-08} \\ 2.823 \times 10^{-07} & -1.783 \times 10^{-08} & 1.783 \times 10^{-10} \end{bmatrix}$$

†The table gives the variances and covariances of the fitted parameters. The values of the parameters and of χ^2 are listed in Table 7.1.

`FitFunc7` Fitting function and χ^2 calculation. In general, every problem requires its own "FitFuncs" routine. For Example 7.1, the individual terms of the power series in x, which are required for the matrix fitting method, are calculated by the function `PowerFunc` selected through a branch on the variable `PAE` in the function `Funct`.

When we use the matrix method to fit a polynomial function to a data sample, the resulting parameters must be identical to those calculated by the determinant method, but we also obtain the full error matrix. The error matrix obtained by fitting of a second-degree polynomial to the complete data sample of Example 7.1 is listed in Table 7.3.

The error matrix can be used to estimate the uncertainty in a calculated result, including the effects of the correlations of the errors. As an example, let us suppose that we wish to find the predicted value of the voltage V and its uncertainty for a temperature of exactly 80°C. We should calculate

$$V = a_1 + a_2T + a_3T^2 \tag{7.26}$$

using the parameters determined by the fit to the data. The uncertainty in the calculated value of V, which results from the uncertainty in the parameters, is given by Equation (3.13),

$$\begin{aligned} s^2 &= \left(\frac{\partial V}{\partial a_1}\right)^2 \sigma_1^2 + \left(\frac{\partial V}{\partial a_2}\right)^2 \sigma_2^2 + \left(\frac{\partial V}{\partial a_3}\right)^2 \sigma_3^2 \\ &\quad + 2\left(\frac{\partial V}{\partial a_1}\frac{\partial V}{\partial a_2}\right)^2 \sigma_{12}^2 + 2\left(\frac{\partial V}{\partial a_1}\frac{\partial V}{\partial a_3}\right)^2 \sigma_{13}^2 + 2\left(\frac{\partial V}{\partial a_2}\frac{\partial V}{\partial a_3}\right)^2 \sigma_{23}^2 \\ &= 1 \cdot \epsilon_{11} + T^2 \cdot \epsilon_{22} + T^4 \cdot \epsilon_{33} + 2(T \cdot \epsilon_{12} + T^2 \cdot \epsilon_{13} + T^3 \cdot \epsilon_{23}) \end{aligned} \tag{7.27}$$

where ϵ_{12} and so on are the covariant terms in the symmetric error matrix. If we used only the diagonal terms in the error matrix, our result would be $V = (2.45 \pm 0.14)$ V. However, the off-diagonal terms are mainly negative, and including them reduces the uncertainty by almost a factor of 10 to 0.015, so that we should quote $V = (2.45 \pm 0.02)$ V.

2.6 Including x uncertainty

Neglecting the σ_x for all the fitting we've discussed so far is only valid when variations on the order of σ_x in x produce measurement y variations much smaller than σ_y. In other words:

$$\sigma_{x_i} \left.\frac{dy}{dx}\right|_{x_i} \ll \sigma_{y_i}$$

J. Orear, [Amer. J. Phys. Oct 1982] shows that replacing $\frac{1}{\sigma_i^2}$ in χ^2 with $\frac{1}{\delta_i^2}$, where

$$\delta_i^2 \equiv \sum_j \left(\frac{\partial f}{\partial x_j}\right)^2 \sigma_{x_{ji}}^2 + \sigma_{y_i}^2$$

variable x_j

i^{th} measurement

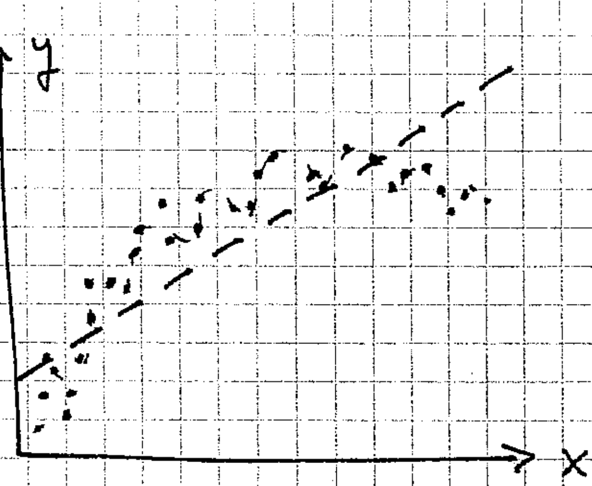
OR:

$$\delta_i^2 = \left(\left.\frac{\partial f}{\partial x}\right|_{x_i}\right)^2 \sigma_{x_i}^2 + \sigma_{y_i}^2 \quad \left\{\text{for one independent variable.}\right.$$

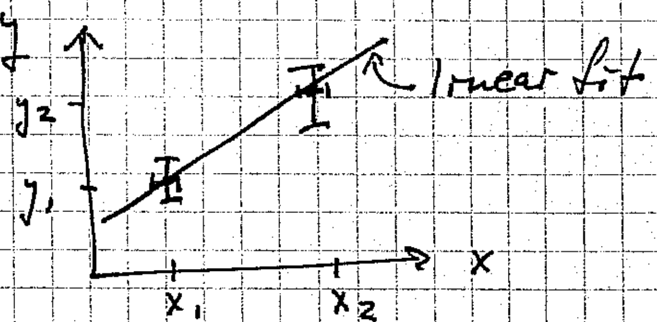
Note that we may note know $\frac{\partial f}{\partial x}$ without fitting, so typically one has to iterate: fit without including x error, then recalculate all $\sigma_i^2 \rightarrow \delta_i^2$, & fit again.

27 Goodness of fit

Let's say you did a linear fit to your data & saw the plot shown at right. This doesn't look like a good fit. But is there a way of quantifying goodness of fit?



Yes! The χ^2 value found after minimization should be about equal to the number of degrees of freedom ν. A degree of freedom is a way in which your data might have deviated from the fit. For instance, a line will always fit two points perfectly:

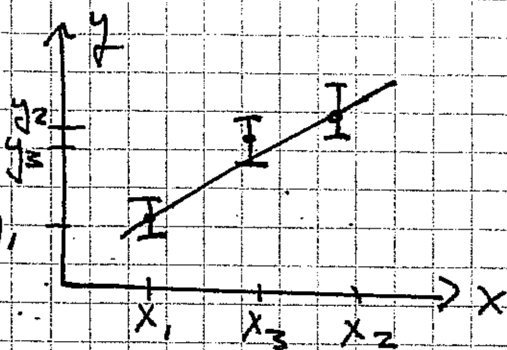


Clearly, $y(x_i) = y_i$ for both points, so

$$\chi^2 = \sum \frac{(y(x_i) - y_i)^2}{\sigma_i^2} = 0 .$$

2.7, cont.

Goodness of fit, cont.



Add one more point, and now you'd expect the deviation of the point to be (on average) comparable to its uncertainty σ_i. Thus $(y(x_3) - y_3)^2 \approx \sigma_3^2$ on average, & this contributes 1 to the total χ^2.

More generally, $\nu = N - m$

data points (N) — # fit parameters (m)

and we'll call $\chi_\nu^2 = \dfrac{\chi^2}{\nu}$ the "reduced" chi squared

The following data sheet tells you how likely your next experimental run will ~~is likely to~~ be worse than the one you are analyzing. You'd like that to be around 50%, telling you your latest run is "typical."

Example: A $\chi_\nu^2 = 2.0$ gives P_χ between 2% & 5% for $\nu = 10$.
Perhaps the fit is not good, or perhaps we were just unlucky (this should happen 1 out of 50 to 1 out of 20 times...)

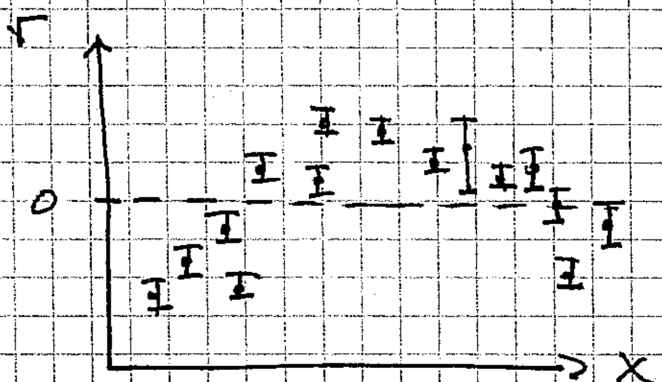
Note: If $P_\chi > 95\%$, it's likely you underestimated your data's uncertainties.

Goodness of fit, cont.

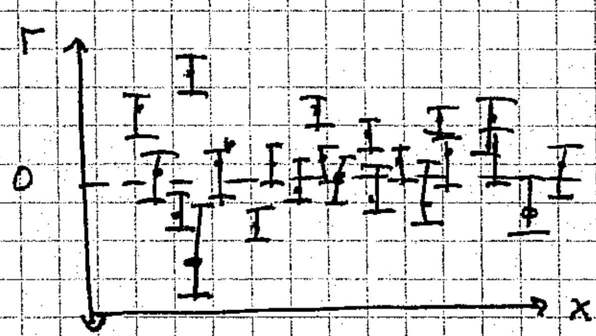
Another useful tool to evaluate goodness of fit is a <u>residual</u> plot. The residual is simply the difference between the measured value y_i & the fit value $y(x_i)$:

$$r_i \equiv y_i - y(x_i)$$

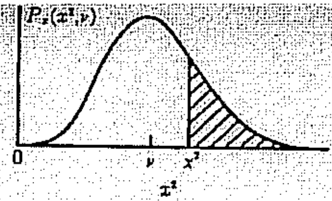
Plotting residuals, with errors σ_i, can tell you if there's a systematic problem with your fit function:



Doesn't look good! (Add a quadratic term to function)



Looks good. About 2/3 of the points have their error bars crossing the r=0 line, and no systematic trend in the data.



26

TABLE C.4
χ^2 distribution. Values of the reduced chi-square $\chi^2_\nu = \chi^2 / \nu$ corresponding to the probability $P_\chi(\chi^2;\nu)$ of exceeding χ^2 vs. the number of degrees of freedom ν

	P							
ν	0.99	0.98	0.95	0.90	0.80	0.70	0.60	0.50
1	0.00016	0.00063	0.00393	0.0158	0.0642	0.148	0.275	0.455
2	0.0100	0.0202	0.0515	0.105	0.223	0.357	0.511	0.693
3	0.0383	0.0617	0.117	0.195	0.335	0.475	0.623	0.789
4	0.0742	0.107	0.178	0.266	0.412	0.549	0.688	0.839
5	0.111	0.150	0.229	0.322	0.469	0.600	0.731	0.870
6	0.145	0.189	0.273	0.367	0.512	0.638	0.762	0.891
7	0.177	0.223	0.310	0.405	0.546	0.667	0.785	0.907
8	0.206	0.254	0.342	0.436	0.574	0.691	0.803	0.918
9	0.232	0.281	0.369	0.463	0.598	0.710	0.817	0.927
10	0.256	0.306	0.394	0.487	0.618	0.727	0.830	0.934
11	0.278	0.328	0.416	0.507	0.635	0.741	0.840	0.940
12	0.298	0.348	0.436	0.525	0.651	0.753	0.848	0.945
13	0.316	0.367	0.453	0.542	0.664	0.764	0.856	0.949
14	0.333	0.383	0.469	0.556	0.676	0.773	0.863	0.953
15	0.349	0.399	0.484	0.570	0.687	0.781	0.869	0.956
16	0.363	0.413	0.498	0.582	0.697	0.789	0.874	0.959
17	0.377	0.427	0.510	0.593	0.706	0.796	0.879	0.961
18	0.390	0.439	0.522	0.604	0.714	0.802	0.883	0.963
19	0.402	0.451	0.532	0.613	0.722	0.808	0.887	0.965
20	0.413	0.462	0.543	0.622	0.729	0.813	0.890	0.967
22	0.434	0.482	0.561	0.638	0.742	0.823	0.897	0.970
24	0.452	0.500	0.577	0.652	0.753	0.831	0.902	0.972
26	0.469	0.516	0.592	0.665	0.762	0.838	0.907	0.974
28	0.484	0.530	0.605	0.676	0.771	0.845	0.911	0.976
30	0.498	0.544	0.616	0.687	0.779	0.850	0.915	0.978
32	0.511	0.556	0.627	0.696	0.786	0.855	0.918	0.979
34	0.523	0.567	0.637	0.704	0.792	0.860	0.921	0.980
36	0.534	0.577	0.646	0.712	0.798	0.864	0.924	0.982
38	0.545	0.587	0.655	0.720	0.804	0.868	0.926	0.983
40	0.554	0.596	0.663	0.726	0.809	0.872	0.928	0.983
42	0.563	0.604	0.670	0.733	0.813	0.875	0.930	0.984
44	0.572	0.612	0.677	0.738	0.818	0.878	0.932	0.985
46	0.580	0.620	0.683	0.744	0.822	0.881	0.934	0.986
48	0.587	0.627	0.690	0.749	0.825	0.884	0.936	0.986
50	0.594	0.633	0.695	0.754	0.829	0.886	0.937	0.987
60	0.625	0.662	0.720	0.774	0.844	0.897	0.944	0.989
70	0.649	0.684	0.739	0.790	0.856	0.905	0.949	0.990
80	0.669	0.703	0.755	0.803	0.865	0.911	0.952	0.992
90	0.686	0.718	0.768	0.814	0.873	0.917	0.955	0.993
100	0.701	0.731	0.779	0.824	0.879	0.921	0.958	0.993
120	0.724	0.753	0.798	0.839	0.890	0.928	0.962	0.994
140	0.743	0.770	0.812	0.850	0.898	0.934	0.965	0.995
160	0.758	0.784	0.823	0.860	0.905	0.938	0.968	0.996
180	0.771	0.796	0.833	0.868	0.910	0.942	0.970	0.996
200	0.782	0.806	0.841	0.874	0.915	0.945	0.972	0.997

TABLE C.4
χ^2 distribution (*continued*)

	P							
ν	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.001
1	0.708	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	0.916	1.204	1.609	2.303	2.996	3.912	4.605	6.908
3	0.982	1.222	1.547	2.084	2.605	3.279	3.780	5.423
4	1.011	1.220	1.497	1.945	2.372	2.917	3.319	4.617
5	1.026	1.213	1.458	1.847	2.214	2.678	3.017	4.102
6	1.035	1.205	1.426	1.774	2.099	2.506	2.802	3.743
7	1.040	1.198	1.400	1.717	2.010	2.375	2.639	3.475
8	1.044	1.191	1.379	1.670	1.938	2.271	2.511	3.266
9	1.046	1.184	1.360	1.632	1.880	2.187	2.407	3.097
10	1.047	1.178	1.344	1.599	1.831	2.116	2.321	2.959
11	1.048	1.173	1.330	1.570	1.789	2.056	2.248	2.842
12	1.049	1.168	1.318	1.546	1.752	2.004	2.185	2.742
13	1.049	1.163	1.307	1.524	1.720	1.959	2.130	2.656
14	1.049	1.159	1.296	1.505	1.692	1.919	2.082	2.580
15	1.049	1.155	1.287	1.487	1.666	1.884	2.039	2.513
16	1.049	1.151	1.279	1.471	1.644	1.852	2.000	2.453
17	1.048	1.148	1.271	1.457	1.623	1.823	1.965	2.399
18	1.048	1.145	1.264	1.444	1.604	1.797	1.934	2.351
19	1.048	1.142	1.258	1.432	1.586	1.773	1.905	2.307
20	1.048	1.139	1.252	1.421	1.571	1.751	1.878	2.266
22	1.047	1.134	1.241	1.401	1.542	1.712	1.831	2.194
24	1.046	1.129	1.231	1.383	1.517	1.678	1.791	2.132
26	1.045	1.125	1.223	1.368	1.496	1.648	1.755	2.079
28	1.045	1.121	1.215	1.354	1.476	1.622	1.724	2.032
30	1.044	1.118	1.208	1.342	1.459	1.599	1.696	1.990
32	1.043	1.115	1.202	1.331	1.444	1.578	1.671	1.953
34	1.042	1.112	1.196	1.321	1.429	1.559	1.649	1.919
36	1.042	1.109	1.191	1.311	1.417	1.541	1.628	1.888
38	1.041	1.106	1.186	1.303	1.405	1.525	1.610	1.861
40	1.041	1.104	1.182	1.295	1.394	1.511	1.592	1.835
42	1.040	1.102	1.178	1.288	1.384	1.497	1.576	1.812
44	1.039	1.100	1.174	1.281	1.375	1.485	1.562	1.790
46	1.039	1.098	1.170	1.275	1.366	1.473	1.548	1.770
48	1.038	1.096	1.167	1.269	1.358	1.462	1.535	1.751
50	1.038	1.094	1.163	1.263	1.350	1.452	1.523	1.733
60	1.036	1.087	1.150	1.240	1.318	1.410	1.473	1.660
70	1.034	1.081	1.139	1.222	1.293	1.377	1.435	1.605
80	1.032	1.076	1.130	1.207	1.273	1.351	1.404	1.560
90	1.031	1.072	1.123	1.195	1.257	1.329	1.379	1.525
100	1.029	1.069	1.117	1.185	1.243	1.311	1.358	1.494
120	1.027	1.063	1.107	1.169	1.221	1.283	1.325	1.446
140	1.026	1.059	1.099	1.156	1.204	1.261	1.299	1.410
160	1.024	1.055	1.093	1.146	1.191	1.243	1.278	1.381
180	1.023	1.052	1.087	1.137	1.179	1.228	1.261	1.358
200	1.022	1.050	1.083	1.130	1.170	1.216	1.247	1.338

2.8 Summary

How do we determine $a_1 \pm \sigma_{a_1}$, where a is a parameter in a fit to the data?

Put in:

$\{x_i, y_i, \sigma_i\}$

$f(x, a_1, \dots a_m)$

maybe: $\{\sigma_{x_i}\}$

Program does:

Minimize

$$\chi^2 \equiv \sum \frac{(y_i - f(x_i, \{a\}))^2}{\sigma_i^2}$$

maybe: iterate with

$$\sigma_i^2 \rightarrow \delta_i^2 = \frac{\partial y}{\partial x}\sigma_{x_i}^2 + \sigma_{y_i}^2$$

Get out:

Minimized χ^2.

Residuals plot.

Get out:

Best parameters a_j^*

$\left.\begin{array}{l}\text{uncertainties } \sigma_j^2 \\ \text{covariances } \sigma_{jk}\end{array}\right\}$ from ϵ_{ij}

Decide: Good fit?

Yes? ok, believe values

No?

- Change function
- Re-take data
- Re-analyze uncertainties
- other...