Treatment of Errors

Simon Parsons
The University of Edinburgh

Random & Systematic Errors
Distributions
The Normal Distribution
[Tutorials]
Calculating Averages
Making Comparisons
Weights in Least Squares
[Tutorials]

Errors

- Whenever we measure some quantity the measurement is always subject to some sort of error.
- It is important to be aware of this when interpreting data - either raw intensities or derived parameters.
- Errors are classed either as
  – RANDOM
  or
  – SYSTEMATIC
- Treatment of an error depends on which type it is.

Random Errors

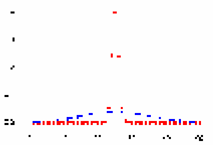
- These come from random fluctuations in the conditions of measurement.
- These can never be eliminated, but they can be minimised by careful experimental design.
- There are usually lots of sources (IMPORTANT- REMEMBER THIS.)
- Examples
  – Temperature of a CCD chip;
  – Fluctuations in tube output;
  – Fluctuations in temperature of the crystal.

Systematic Errors

- Random errors determine the distribution of observations about some true value.
- Systematic errors produce systematic shifts in the results.
- We may or may not be aware of their presence.
- May come from the experiment or the refinement model
- Examples
  – Absorption
  – Wrong crystal-detector distance.
  – Mis-centred crystal.
  – Wrong scattering factors.
  – Wrong atoms.
  – Mistyped cell dimensions.

Precision and Accuracy

- Accuracy
  – A measure of how close an experimental measurement is to its true value.



- Precision
  – A measure of how reproducible a result is.
  – It refers to the scatter of observations.
  – Measured by a standard uncertainty.

Standard Uncertainties

- Also referred to as estimated standard deviations or e.s.d.'s.
- In crystallography we use a bracket notation.
- 1.590(4) Å means a distance of 1.590 Å with an s.u. of 0.004 Å.
- 1.59(4) Å means 1.59 Å with an s.u. of 0.04 Å. *10x less precise*.

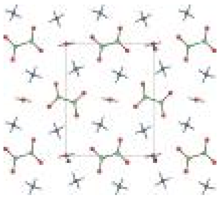
Standard Uncertainties

- Also referred to as estimated standard deviations or e.s.d.'s.
- In crystallography we use a bracket notation.
- Older texts use the ± notation, but this appears to specify a strict range for the measurement.
- 1.590(4) Å means a distance of 1.590 Å with an s.u. of 0.004 Å.
- 1.59(4) Å means 1.59 Å with an s.u. of 0.04 Å. *10x less precise*.
- 1.59(4) = 1.59 ± 0.04.

Random and Systematic Errors

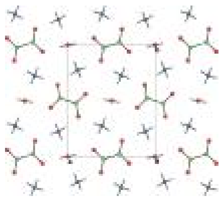
- Random Errors
  - Effect the spread of results of an experiment.
  - High scatter = low precision = high s.u.
- Systematic errors
  - shift the measurements
  - They MAY not affect precision at all.
  - Others may shift different observations by different amounts, so decreasing precision.
  - Usually get a bit of both.

Systematic Errors



b	***R***	**C-C**
8.0265	**2.7**	**1.5644(6)**
8.5265	**2.8**	**1.6532(7)**

Systematic Errors



b	***R***	**C-C**
8.0265	**2.7**	**1.5644(6)**
no abs	**2.8**	**1.5633(9)**

Random and Systematic Errors

- Random Errors
  - Treated using statistical distributions.
  - Normal (or Gaussian).
  - Poisson.
- Systematic Errors
  - Can not be treated with a general theory.
  - Corrections for the physical effects causing the error should either be made to the data or included in the refinement model.

Random & Systematic Errors

Distributions

The Normal Distribution

[Tutorials]

Calculating Averages

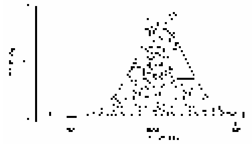
Making Comparisons

Weights in Least Squares

[Tutorials]

Distributions

```
INTENSITIES OF THE 114 REFLECTION.  N=67
1684.78 1787.27 1794.81 1807.33 1819.65 1825.30 1853.30
1743.72 1788.16 1796.12 1807.53 1819.81 1826.18 1854.28
1756.32 1788.23 1798.56 1807.54 1819.88 1827.00 1856.05
1761.98 1788.50 1801.34 1808.86 1820.28 1830.38 1867.75
1767.55 1789.60 1802.79 1812.50 1821.31 1830.85 1872.35
1767.86 1789.69 1804.08 1813.05 1821.57 1832.63 1881.82
1772.06 1793.45 1804.38 1813.05 1822.44 1834.59 1902.13
1772.38 1793.93 1804.49 1813.54 1823.11 1836.25
1784.30 1794.50 1804.54 1814.43 1823.32 1837.49
1784.60 1794.52 1804.75 1819.36 1823.51 1841.55
```

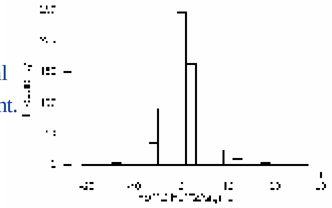


Distributions

- Values of

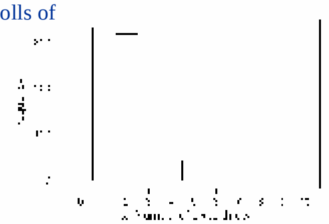
$$\frac{F_o^2 - F_c^2}{u(F_o^2)}$$

taken after a crystal structure refinement.



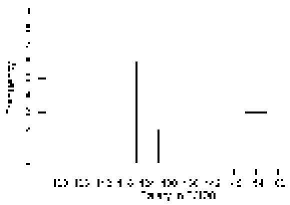
Distributions

- Number of sixes thrown in 1000 rolls of 10 dice.



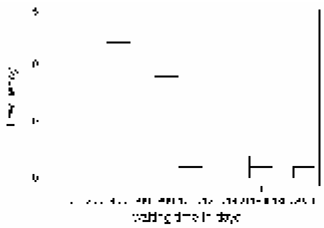
Distributions

- Salaries of 30 workers in an American factory.

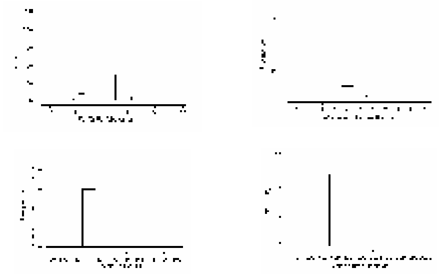


Distributions

- Waiting times between large earthquakes between 1901 and 1977.

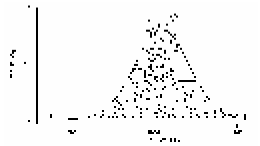


Modelling Data



Probability Distributions

- Functions which describe the shape of the graphs are called probability density functions or pdfs.
- Different pdfs are *indexed* using different quantities.



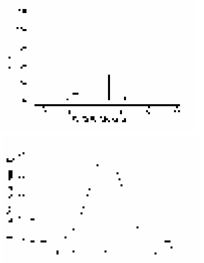
$$P_N(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Discrete Distributions

- Values of data may be limited in some way.
- E.g. if throwing dice it is impossible to score a fractional number of sixes.
- Distributions like this are called *discrete*.

Continuous Distributions

- By contrast if a variable can adopt any value, the corresponding distribution is called *continuous*.



Some PDFs

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

$$P_P(x;\mu)=\frac{\mu^x}{x!}e^{-\mu}$$

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

Binomial Distribution

- Dice problem: Roll 10 dice 1000 times. How often do we score 0 sixes, 1 six, 2 sixes...?

$$\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}=\left(\frac{5}{6}\right)^{10}=0.1615$$

$$\frac{1}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}=\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{9}=0.0323$$

$$P_B(x;10,\frac{1}{6})=\frac{10!}{x!(10-x)!}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{10-x}$$

Normal and Binomial Distributions

- The normal distribution closely resembles the binomial distribution.
- Which is used depends on whether its easier to model your data in terms of *n* and *p* or μ and σ^2, and on whether the distribution is discrete or continuous.
- Consider $P_B(x;16,0.5)$. μ is 8 and σ is 2.

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

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

Characterising Data

- Mean

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

- Median: Put the observations in order; the median is the middle value.

- Standard deviation

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

s = standard deviation of our sample.
Sometimes σ is used instead.

The VARIANCE is the square of the standard deviation.

Characterising Distributions

- Mean

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

- Standard deviation

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

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

Both of these equations are calculated on the basis of the data that have been collected. It is possible that we could have made an infinite number of measurements, but usually we only have *sampled* a *parent distribution*.

If the pdf describing the parent distribution is known, then the mean (μ) and standard deviation (σ) can be calculated directly from the function describing the pdf (see later).

Estimation of μ and σ

$$m = \bar{x}$$

$$s^2 = \frac{N}{N-1}s^2$$

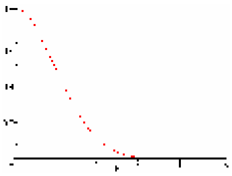
$$< f(x) >= \int_{-\infty}^{\infty} f(x)P(x)dx$$

$$< x >= m = \int_{-\infty}^{\infty} xP(x)dx$$

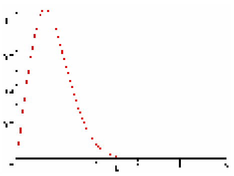
$$s^2 = \int_{-\infty}^{\infty} (x-m)^2 P(x)dx$$

$$< x >= m = \sum_{i=1}^{n} x_i P(x_i)$$

Intensity Statistics



$$P_{-1}(|E|) = \sqrt{\frac{2}{p}}\exp\left(\frac{-|E|^2}{2}\right)$$



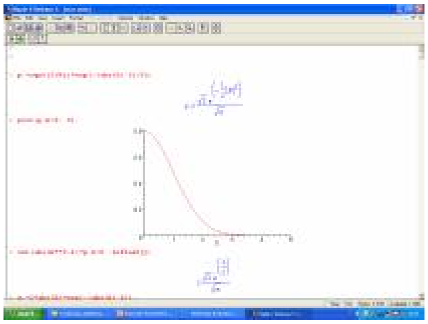
$$P_1(|E|) = 2|E|\exp\left(-|E|^2\right)$$

$$<|E^2-1|>=\sqrt{\frac{2}{p}}\int_0^\infty|E^2-1|\exp\left(\frac{-|E|^2}{2}\right)dE=2\sqrt{\frac{2}{p}}\exp\left(\frac{-1}{2}\right)=0.968$$

$$<|E^2-1|>=2\int_0^\infty|E^2-1||E|\exp\left(-|E|^2\right)dE=\frac{2}{e}=0.736$$

$$<|E^2-1|>=\sqrt{\frac{2}{p}}\int_0^\infty|E^2-1|\exp\left(\frac{-|E|^2}{2}\right)dE=2\sqrt{\frac{2}{p}}\exp\left(\frac{-1}{2}\right)=0.968$$

$$<|E^2-1|>=2\int_0^\infty|E^2-1||E|\exp\left(-|E|^2\right)dE=\frac{2}{e}=0.736$$



Random & Systematic Errors

Distributions

The Normal Distribution

[Tutorials]

Calculating Averages

Making Comparisons

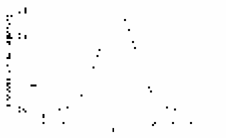
Weights in Least Squares

[Tutorials]

The Normal Distribution

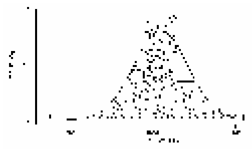
- This is the most important distribution in science.
- It is also referred to as the Gaussian distribution.
- It is indexed on the mean and variance of the data set.

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



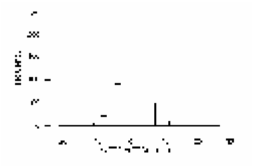
Some Data

```
INTENSITIES OF THE 114 REFLECTION.  N=67
1684.78 1787.27 1794.81 1807.33 1819.65 1825.30 1853.30
1743.72 1788.16 1796.12 1807.53 1819.81 1826.18 1854.28
1756.32 1788.23 1798.56 1807.54 1819.88 1827.00 1856.05
1761.98 1788.50 1801.34 1808.86 1820.28 1830.38 1867.75
1767.55 1789.60 1802.79 1812.50 1821.31 1830.85 1872.35
1767.86 1789.69 1804.08 1813.05 1821.57 1832.63 1881.82
1772.06 1793.45 1804.38 1813.05 1822.44 1834.59 1902.13
1772.38 1793.93 1804.49 1813.54 1823.11 1836.25
1784.30 1794.50 1804.54 1814.43 1823.32 1837.49
1784.60 1794.52 1804.75 1819.36 1823.51 1841.55
```



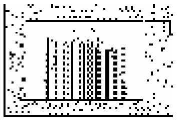
Serine Data

- In crystal structure determinations the errors in our measurements follow a Normal distribution.



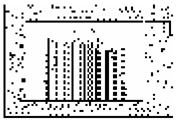
Central Limit Theorem

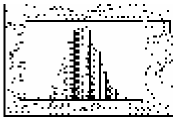
- Consider a quantity *y* which is the sum of many other quantities *x*.
- The *x*’s all have their own pdfs.
- What is the pdf of *y*?
- 5000 random numbers taken from a uniform distribution.
- Sum of 12 random numbers?



Central Limit Theorem

- Consider a quantity *y* which is the sum of many other quantities *x*.
- The *x*’s all have their own pdfs.
- What is the pdf of *y*?
- 5000 random numbers taken from a uniform distribution.
- Sum of 12 random numbers?



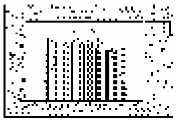


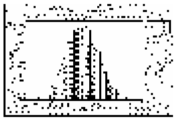
Central Limit Theorem

$$y = \sum_{i=1}^{N} x_i$$

$$m_y = \sum_{i=1}^{N} m_i$$

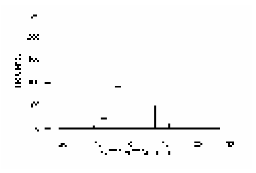
$$s_y^2 = \sum_{i=1}^{N} s_i^2$$





Central Limit Theorem

- In a crystal structure determination our measurements may be subject to lots of small errors, but their cumulative effect still makes the measurement errors follow a normal distribution.
- This has important consequences for refinement (Q3).



Repeated Measurements

- This figure shows the spread of measurements obtained for a particular reflection intensity.
- Suppose we did another set of measurements of this reflection, $F^2(114)$, and calculated the mean for the new set of measurements.
- And then another, and another....
- What would the distribution of the means look like?

- Intuitively we would expect the distribution of the means to have a narrower distribution.

Standard Error on the Mean

- A consequence of the CLT is that:

$$s^2(\bar{x}) = \frac{s^2(x)}{N}$$

- Note that if we measure something twice as many times the standard deviation will decrease by a factor 1/√2.
- This idea applies to intensity measurements, precision of least squares parameters etc.
- Particularly evident in standard deviations of unit cell dimensions!

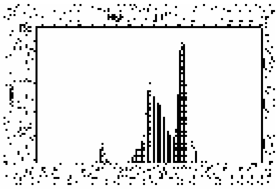
Are we really making repeated measurements?

- If we have a set of data we think that they may represent the same quantity. But do they really?
- Is there some systematic factor that has been neglected?
- You may not know that you're neglecting something!
- Example: All CN bonds in the CSD.

N	153987
Mean	1.3981
Standard dev.	0.081
Standard Error	0.0002

Are we really making repeated measurements?

- Repeated measurements of the same quantity should follow a normal distribution.

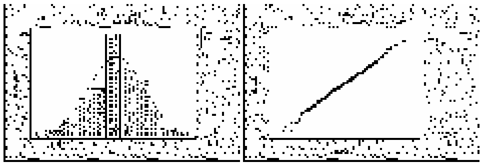


Are we really making repeated measurements?: Normal Probability Plots

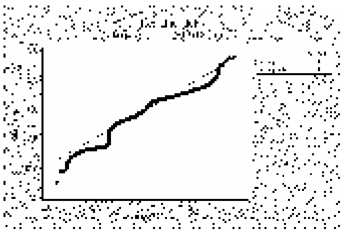
- If we have *j* measurements in a dataset we can predict what the distribution of delta/sig should be.
- When plotted against obs. values of del/sig the result should be:
  - a straight line
  - passing through the origin
  - with a gradient of 1.0

$$\frac{j-2i+1}{j} = \frac{1}{\sqrt{2p}}\int_{-a}^{a} \exp\left(\frac{-x^2}{2}\right)dx$$

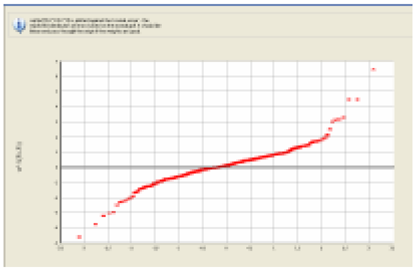
Normal Probability Plots



Normal Probability Plots



Normal Probability Plots



Are we really making repeated measurements?: Chi-Squared

- Calculate χ^2_{red}.
- If a dataset follows a normal distribution this should =1.
- In this case χ^2_{red} ~ 166000 if the standard error is used for sigma!

$$c^2 = \frac{\sum (x_i - \bar{x})^2}{s^2}$$

$$c^2_{red} = \frac{c^2}{N - P}$$

Tutorials

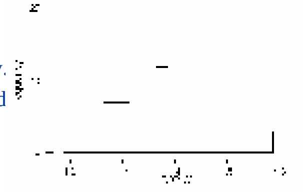
Chapter 16 Q1
Chapter 17 Q2
Chapter 16 Q2b
Chapter 16 Q3

Random & Systematic Errors
Distributions
The Normal Distribution
[Tutorials]
Calculating Averages
Making Comparisons
Weights in Least Squares
[Tutorials]

Averages

- Before taking an average from a set of data it is important to ask whether the average is a meaningful quantity.
- This can be assessed quickly using a histogram.

Taylor & Kennard, *Acta*, (1983), **B39,** 517-525.



Averages

- Remember that different techniques measure different quantities.
- Neutron diffraction measures inter-nuclear distances, X-rays measure distances between maxima in electron density,
- Vibrations affect bond distances.
- It may be necessary to carry out a 'riding' correction if thermal ellipsoids are large.

Weighted and Unweighted Averages

- It is possible to define weighted and unweighted means.
- Weights can be derived from the s.u.s on the parameters being averaged.
- But when should these be used?
- Note that if all the weights are the same the formula are equivalent.

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

$$\bar{x}_w = \sum_{i=1}^{n} w_i x_i \Big/ \sum_{i=1}^{n} w_i$$

$$w_i = \frac{1}{s^2(x_i)}$$

Experimental and Environmental Effects

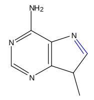
- If environmental effects are negligible then

$$c^2 = \sum w_i (x_i - \bar{x}_w)^2$$

should follow a chi-squared distribution.

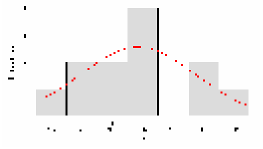
- Reduced chi-squared should be about 1.

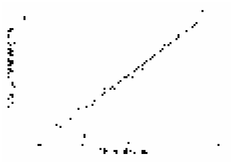
x_i	σ(x_i)
1.315	0.003
1.311	0.003
1.322	0.012
1.329	0.012
1.347	0.021
1.301	0.0225
1.378	0.0285
1.325	0.030
1.314	0.030
1.333	0.0315
1.294	0.045
1.315	0.045



Taylor & Kennard, *Acta*, (1983), **B39,** 517-525.

Adenine Data





Adenine Data

x_i	σ(x_i)
1.315	0.003
1.311	0.003
1.322	0.012
1.329	0.012
1.347	0.021
1.301	0.0225
1.378	0.0285
1.325	0.030
1.314	0.030
1.333	0.0315
1.294	0.045
1.315	0.045

$$c^2 = 11.66$$

$$DOF = 12 - 1 = 11$$

$$c^2_{red} = 1.06$$

Under these conditions use the weighted mean with a standard deviation $\sqrt{\frac{1}{\sum w}}$

Weighted Mean = 1.314(2) Å

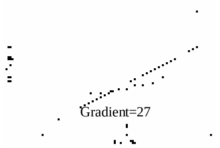
H-bond Data

x_i	σ(x_i)
1.814	0.0015
1.844	0.003
1.728	0.003
1.832	0.003
2.121	0.003
1.997	0.0075
1.808	0.0075
1.833	0.009
1.739	0.009
1.772	0.009
1.742	0.0105
1.877	0.012
1.948	0.012

$$c^2 = 11245$$

$$DOF = 13 - 1 = 12$$

$$c^2_{red} = 937$$



Under these conditions use the unweighted average. The recommended variance of the mean is

$$s^2(m) = s^2(sample) - \overline{s^2(x_i)}$$

but in practice this is little different from σ^2(sample): 1.85(11) Å.

Averages

- In many cases in structural chemistry there really are genuine differences in parameters because of intermolecular interactions.
- In general the unweighted average is used in practice.
- In many cases an average is a completely meaningless number!
- Think about quoting a range instead.
- E.g. Metal - ligand distances of 1.925(2) and 1.985(2) Å.

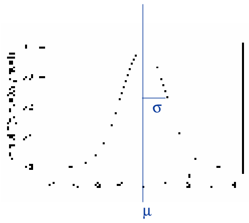
Random & Systematic Errors
Distributions
The Normal Distribution
[Tutorials]
Calculating Averages
Making Comparisons
Weights in Least Squares
[Tutorials]

Two Examples

- Comparing Bond Lengths
- Absolute Structure

Using the Normal Distribution

- The height of the curve is reduced to exp(-0.5) of its max. height at $x = \mu + \sigma$.
- Only 68.3% of the area under the curve lies within 1 σ of the mean.
- But 99.7% lies within 3σ of the mean.
- This is where the 3σ rule come from.



When are Two Bond Lengths Different?

- Is the difference in bond lengths greater than 3x its su?

$$f = x - y$$

$$s^2(f) = \left(\frac{\partial f}{\partial x}\right)^2 s^2(x) + \left(\frac{\partial f}{\partial y}\right)^2 s^2(y)$$

$$s^2(f) = s^2(x) + s^2(y)$$

Examples

- 1.523(3) and 1.540(4) are (just) significantly different.

$$s^2(f) = s^2(x) + s^2(y)$$

$$su = \sqrt{0.003^2 + 0.004^2} = 0.005$$

$$1.540 - 1.523 = 0.017$$

$$difference = \frac{0.017}{0.005} = 3.4s$$

- 2.003(4) and 2.010(2) are not.

$$su = \sqrt{0.002^2 + 0.004^2} = 0.004$$

$$2.010 - 2.003 = 0.007$$

$$difference = \frac{0.007}{0.004} = 1.8s$$

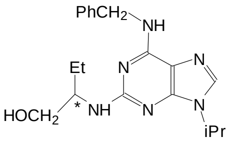
Standard Uncertainties from Least Squares

- The s.u.s from least squares take no account of systematic errors.
- It is probable that they are too small by a factor of about 1.3-2.0.
- More realistic estimate might be 1.5x value from least squares.

- Evidence
  - Repeated determinations of crystal structures.
  - Normal probability plots used to compare structures.

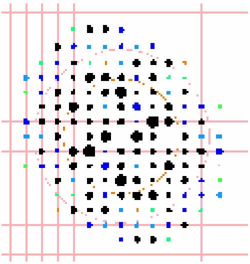
2: Absolute Structure

- X-ray crystallography is so popular because it directly produces images of molecules. No other technique does this.
- One problem with the technique is that characterisation of the absolute configuration of a molecule is difficult.
- This information is often vitally important though.



The Problem

- All diffraction patterns are centrosymmetric according to Friedel's law.



The Problem

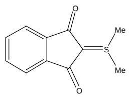
- *F*(*hkl*) = *F*(-*h*-*k*-*l*)
- If Friedel's law held exactly absolute structure determination by X-ray crystallography would be impossible.
- But anomalous scattering introduces deviations which carry absolute structure information.
- Magnitudes depend on
  - Elements present
  - Wavelength of X-rays used.

The Problem

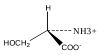
- The number of interest is the anomalous scattering factor *f*".
- If *f*" is large, the deviations from Friedel's law are large and absolute structure determination is no problem.
- Generally *f*" increases with *Z*. This is why making a heavy atom derivative can assist in absolute structure determination.

	Mo		Cu	
	f'	f"	f'	f"
C	0.003	0.002	0.018	0.009
N	0.006	0.003	0.031	0.018
O	0.011	0.006	0.049	0.032
S	0.125	0.124	0.333	0.558

Some Data



```
-2 -3 -1  I = 1.56±0.05
 2  3  1  I = 1.85±0.05
```



```
-2 -9 -3  I = 80±3
 2  9  3  I = 74±3
```

Note that achiral molecules may give a chiral crystal structure! Absolute structure needs to be considered in ALL non-centro. space groups.

Flack's Method

- Example: we wish to determine whether a chiral centre in a molecule is R or S.
- Solve the structure in the usual way, assuming that the chiral centre is R.
- Then treat the molecule as an inversion twin: part of the crystal (1-*x* mole fraction) is in the R-form, the rest is in the S-form (*x* mole fraction).
- Refine the twin scale factor (*x*); *x* is called the *Flack parameter.*

Flack’s Method

- This is equivalent to refining the R-isomer competitively against the S-isomer.
- If $x = 0$ our initial assumption was correct, and the molecule is the R-isomer.
- If $x = 1$ then we should invert the model.
- Intermediate values of x imply inversion or *racemic* twinning.

Refinement in ShelxL

```
TWIN -1 0 0 0 -1 0 0 0 -1
BASF 0.5
```

```
TWIN
BASF 0.5
```

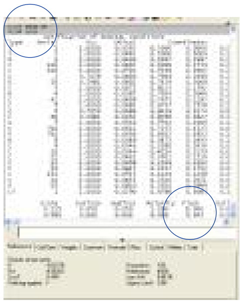
Refinement of Absolute Structure

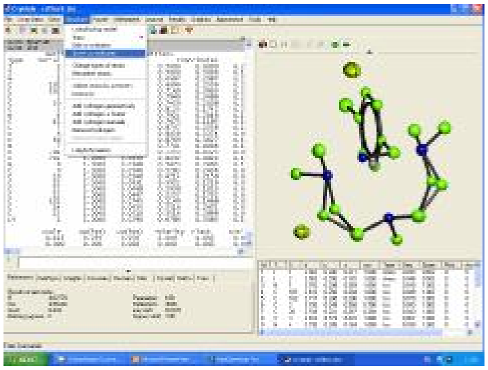
- Effectively we refine one ‘hand’ competitively against the other.
- Flack Parameter = fraction of the structure in the alternative hand.

If Flack = 0 the model is correct.
If Flack = 1 the model needs inverting.

Example

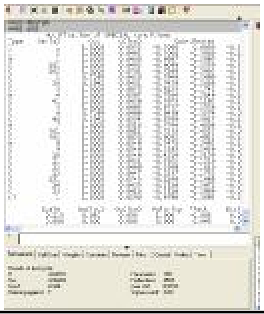
- In this case we happen to have the structure in the wrong ‘hand’.
- Therefore it should be inverted.





Example

- Refinement of the inverted structure yields a Flack parameter of -0.00(4).
- The high precision reflects the presence of iodine (a heavy atom).



When the Flack Method Doesn't Work…

- Crystallography only has a problem distinguishing enantiomers, not diastereomers.
  - Derivatise with a group of known chirality.
- Absolute structure is hardest for light atoms (C,H,N,O) structures.
  - Incorporate a heavy atom (see later).

The Precision of *x*

- *x* is a least-squares parameter, and its refined value has a standard uncertainty, *u*.
- Flack & Bernardinelli: *J. Appl. Cryst.* (2000), **33**, 1143-1148.
- How small should *u* be if an absolute structure can be said to be reliably determined?
- *x* = 0.3(8) is clearly no good.

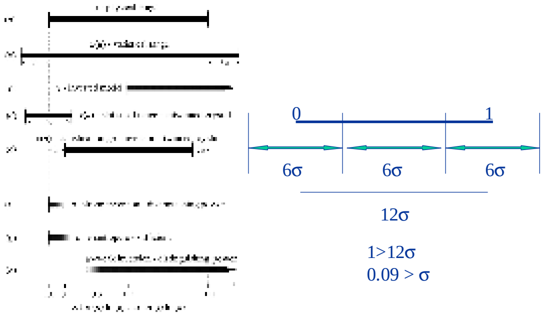
The Precision of *x*

- But what about -0.4(3) for one 'hand' versus +1.3(3) for the alternative?

The Precision of *x*

- But what about -0.4(3) for one 'hand' versus +1.3(3) for the alternative?
- *The authors have clearly not understood how to interpret the Flack parameter and are unaware that their experiment has not been able to determine the absolute configuration of the compound studied.*

The Precision of *x*



1>12σ
0.09 > σ

Precision of *x*

- Flack showed that even if a compound is *known (?)* to be chirally pure, *u* must be less than **0.1** before any conclusion should be drawn.
- See questions 6 and 7

Random & Systematic Errors

Distributions

The Normal Distribution

[Tutorials]

Calculating Averages

Making Comparisons

Weights in Least Squares

[Tutorials]

Aims of Refinement

h	k	l	$\|F_{calc}\|^2$	$\|F_{obs}\|^2$	$\sigma(\|F_{obs}\|^2)$
1	0	0	295.88	318.11	8.25
2	0	0	72119.66	68116.74	707.70
3	0	0	4909.10	5006.32	58.25
4	0	0	5423.07	5145.43	63.35
5	0	0	1710.00	1730.55	27.37
6	0	0	5212.70	5024.92	185.27
7	0	0	2951.36	2975.44	47.25

...

Aim to get as close an agreement as possible between obs.

and calc. data.

Aim to get as close an agreement as possible between obs.

and calc. data. This is achieved by LEAST SQUARES:

F-Refinement (usually includes a σ cut-off):

$$\sum_{h,k,l} w_{hkl}(|F_{o,hkl}|-|F_{c,hkl}|)^2 \qquad w_{hkl}=\frac{1}{s\,(|F_{o,hkl}|)^2}$$

*F*²-Refinement (usually uses ALL data):

$$\sum_{h,k,l} w_{hkl}(|F_{o,hkl}|^2-|F_{c,hkl}|^2)^2 \qquad w_{hkl}=\frac{1}{s\,(|F_{o,hkl}|^2)^2}$$

These are *Minimisation Functions.*

Tutorial Question 3

If our measurement errors follow a normal distribution we can use least squares weighted using the inverse variances of our observations:

$$\sum_{h,k,l} w_{hkl}(|F_{o,hkl}|-|F_{c,hkl}|)^2 \qquad w_{hkl}=\frac{1}{s\,(|F_{o,hkl}|)^2}$$

$$F = scale\sum f\{[Cos2p\,(hx+ky+lz)]^2+Sin2p\,(hx+ky+lz)]^2\}^{\frac{1}{2}}e^{(8p^2USin^2q/l^2)}$$

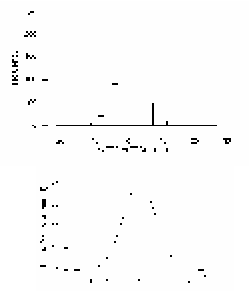
Poisson Distribution

- Counting statistics follow a *Poisson Distribution.*
- The important property is that the mean and variance are equal.
- If we measure an intensity *I* the variance is also *I*, and so σ(*I*) = √*I*.
- This is the form taken by σ in Rietveld refinement. In single crystal work this is one contribution to σ.

$$P_P(x;\mu)=\frac{\mu^x}{x!}e^{-\mu}$$

Weights in Refinement

- If our measurements follow a normal distribution then we are justified in using the 1/ σ²'s as weights.
- But notice the long tails in the distribution.

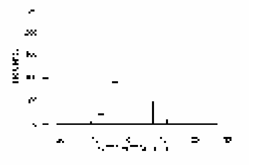


Tails

- Long tails are often observed in crystallographic refinements.
- Measurement errors cease to be normal at the tails of the distribution.
- Uncorrected systematic errors.
  - absorption
  - extinction
  - spherical scattering factors.

Solution

- Fiddle the σ's so that the distribution is normal!
- Most of the largest errors are associated with the strong data.
- Strong data will be down-weighted.



$$w = \frac{1}{s^2(|F_o^2|) + a|F_o^2|}$$

Weights

- Wilson showed that this scheme induced bias into the refined parameters.
- One obvious solution is to use F_c instead.

$$w = \frac{1}{s^2(|F_o^2|) + a|F_o^2|}$$

$$w = \frac{1}{s^2(|F_o^2|) + a|F_c^2|}$$

Weights

- But this also induced bias in the refined parameters.
- BUT the bias was in the opposite sense to when F_o was used and about ½ as great.
- The value of *a* is optimised to give a flat analysis of variance.

$$w = \frac{1}{s^2(|F_o^2|) + a|F_o^2|}$$

$$w = \frac{1}{s^2(|F_o^2|) + a|F_c^2|}$$

$$w = \frac{1}{s^2(|F_o^2|) + aP}$$

$$P = \frac{1}{3}(|F_o^2| + 2|F_c^2|)$$

Random & Systematic Errors
Distributions
The Normal Distribution
[Tutorials]
Calculating Averages
Making Comparisons
Weights in Least Squares
[Tutorials]

Tutorials

Ch 16 Q6, 7
Ch 17 Q1