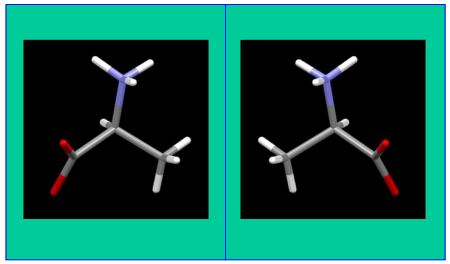
Absolute Structure Determination

Simon Parsons
The University of Edinburgh

Absolute Structure & The Flack Parameter

$$I_{model}(\mathbf{h}) = (1-x)\,|F_{single}(\mathbf{h})|^2 + x\,|F_{single}(\bar{\mathbf{h}})|^2$$



Refinement of *x* in SHELXL

```
TITL alanine K2 in P2(1)2(1)2(1) 36 runs
CELL 1.54178   5.791   5.944  12.269  90.000  90.000  90.000
ZERR    4.00   0.002   0.002   0.002   0.000   0.000   0.000
LATT -1
SYMM 0.5-X, -Y, 0.5+Z
SYMM -X, 0.5+Y, 0.5-Z
SYMM 0.5+X, 0.5-Y, -Z
SFAC C  H  N  O
UNIT 12 28 4 8
L.S. 8
TWIN -1 0 0 0 -1 0 0 0 -1
BASF 0.1
ACTA
FMAP 2
PLAN 5
TEMP -173
WGHT    0.043900    0.037700
EXTI    0.008815
FVAR       4.86499
C1    1    0.645032    0.967483    0.338819    11.00000    0.01086    0.01519
        0.01265   -0.00143   -0.00034   -0.00055
H1    2    0.660159    0.928796    0.263538    11.00000    0.01094
C2    1    0.399489    1.055044    0.358987    11.00000    0.01303    0.01455
        0.00904    0.00193    0.00051    0.00047
```

TWIN -1 0 0 0 -1 0 0 0 -1
BASF 0.1

N	value	esd	shift/esd	parameter
1	4.86497	0.02065	0.000	OSF
2	**-0.03716**	**0.26951**	**0.000**	**BASF 1**
3	0.00881	0.00260	0.000	EXTI

The Problem

- A precise determination of *x* requires large values of *f* ” relative to f_0+*f* ’.
- But even with Cu-radiation *f*’(O) etc. are small and conventional methods give u(x) > 0.1

f” values	Mo	Cu
C	0.002	0.009
N	0.003	0.018
O	0.006	0.032
S	0.124	0.558

Friedif

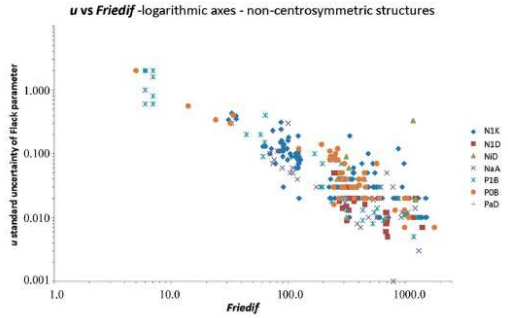


Figure 1
Plot of *u versus Friedif* on logarithmic axes for non-centrosymmetric structures. These are measured with high and low Friedel coverage and may have a centrosymmetric substructure.

Flack & Bernardinelli. Acta Cryst. (2008) A64, 484

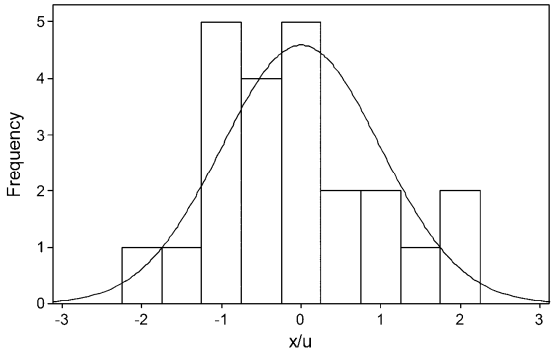
Friedif and refined *x*

Code	Formula	Friedif	*x* (Normal Ref't)
L-Alanine	$C_3H_7NO_2$	34	-0.04(27)
GKO02	$C_{25}H_{31}NO_5$	32	0.01(15)
R-CYCLO	$C_{19}H_{26}N_6O$	21	-0.02(27)
TWA16A	$C_{16}H_{18}N_2$	13	0.00(69)
Cholestane	$C_{27}H_{48}$	9	-0.01(77)

Sample from 23 data sets

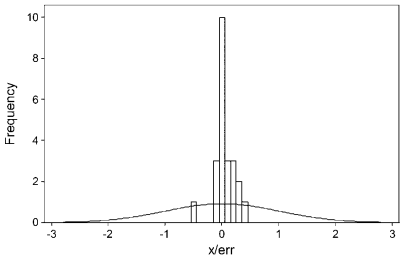
Parsons, Flack & Wagner *Acta Cryst* (2013) **B69**, 249

Expected distribution of 23 values of *x/u*



Refinement of *x*

Code	*x* (Normal Ref't)
L-Alanine	-0.04(27)
GKO02	0.01(15)
R-CYCLO	-0.02(27)
TWA16A	0.00(69)
Cholestane	-0.01(77)



Distribution of x/u compared to a unit Gaussian for 23 Structures

$Chi^2 = 0.03$

The 'Quotient' Method in ShelxL-2014

- Systematic errors like absorption may drown-out anomalous differences.
- Measure Friedel opposites in such as way that absorption errors are the same for both.
- Stoe - measure

  $I(\mathbf{h})$ at 2θ, ω, χ and ϕ

  $I(-\mathbf{h})$ at -2θ, -ω, χ and ϕ
- The quotient $I(\mathbf{h})/I(-\mathbf{h})$ is free from absorption and extinction errors. Also scale-free.

Le Page, Gabe & Gainsford. J. Appl. Cryst. (1990), **23**, 406

Quotients

$$\frac{I(\mathbf{h})-I(-\mathbf{h})}{I(\mathbf{h})+I(-\mathbf{h})}=\frac{F_o^2(\mathbf{h})-F_o^2(-\mathbf{h})}{F_o^2(\mathbf{h})+F_o^2(-\mathbf{h})}=(1-2x)\frac{F_{\text{single}}^2(\mathbf{h})-F_{\text{single}}^2(-\mathbf{h})}{F_{\text{single}}^2(\mathbf{h})+F_{\text{single}}^2(-\mathbf{h})}$$

Parsons, Flack & Wagner Acta Cryst (2013) **B69**, 249

Quotients

$$\frac{F_o^2(\mathbf{h})-F_o^2(-\mathbf{h})}{F_o^2(\mathbf{h})+F_o^2(-\mathbf{h})}=(1-2x)\frac{F_{\text{single}}^2(\mathbf{h})-F_{\text{single}}^2(-\mathbf{h})}{F_{\text{single}}^2(\mathbf{h})+F_{\text{single}}^2(-\mathbf{h})}$$

This can be calculated from your data set.

This can be calculated (F_c^2 for a model refined with $x = 0$).

$$\frac{F_o^2(\mathbf{h})-F_o^2(-\mathbf{h})}{F_o^2(\mathbf{h})+F_o^2(-\mathbf{h})}=(1-2x)\frac{F_{\text{single}}^2(\mathbf{h})-F_{\text{single}}^2(-\mathbf{h})}{F_{\text{single}}^2(\mathbf{h})+F_{\text{single}}^2(-\mathbf{h})}$$

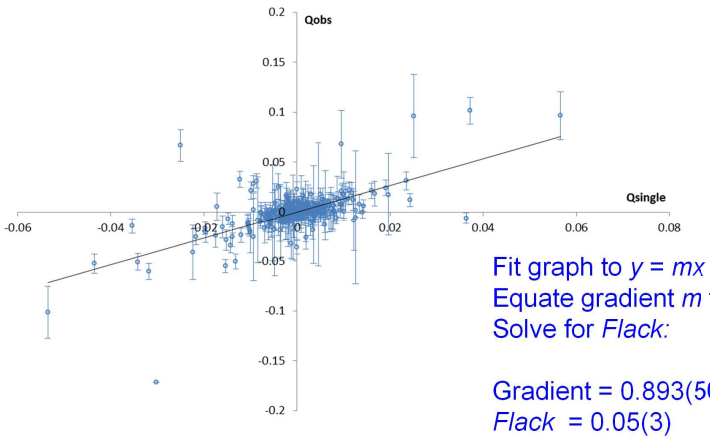
$$Q_o(\mathbf{h})=(1-2x)\,Q_{\text{single}}(\mathbf{h})$$

$$\frac{F_o^2(\mathbf{h}) - F_o^2(-\mathbf{h})}{F_o^2(\mathbf{h}) + F_o^2(-\mathbf{h})} = (1-2x)\frac{F_{single}^2(\mathbf{h}) - F_{single}^2(-\mathbf{h})}{F_{single}^2(\mathbf{h}) + F_{single}^2(-\mathbf{h})}$$

$$Q_o(\mathbf{h}) = (1-2x)\,Q_{single}(\mathbf{h})$$

$$y = mx$$

$$Q_o(\mathbf{h}) = (1-2x)\,Q_{single}(\mathbf{h})$$



Fit graph to $y = mx$
Equate gradient m to $(1-2x)$
Solve for *Flack:*

Gradient = 0.893(50) = 1 – 2*x*
Flack = 0.05(3)

Implemented in XPREP and Shelxl-2014.
No TWIN/BASF instructions!

D:\My Documents\ppt\iucr_2014\alanine.ins

```
TITL alanine K2 in P2(1)2(1)2(1) 36 runs
CELL 1.54178   5.791   5.944  12.269  90.000  90.000  90.000
ZERR    4.00   0.002   0.002   0.002   0.000   0.000   0.000
LATT -1
SYMM 0.5-X, -Y, 0.5+Z
SYMM -X, 0.5+Y, 0.5-Z
SYMM 0.5+X, 0.5-Y, -Z
SFAC C  H  N  O
UNIT 12 28 4 8
L.S. 8
    TWIN -1 0 0 0 -1 0 0 0 -1
    BASF 0.5
ACTA
FMAP 2
PLAN 5
TEMP -173
WGHT    0.043900    0.037700
EXTI    0.008815
FVAR       4.86499
C1    1    0.645032    0.967483    0.338819    11.00000    0.01086    0.01519 =
         0.01265   -0.00143   -0.00034   -0.00055
H1    2    0.660159    0.928796    0.263538    11.00000    0.01094
C2    1    0.399489    1.055044    0.358987    11.00000    0.01303    0.01455 =
         0.00904    0.00193    0.00051    0.00047
O3    4    0.374967    1.227248    0.416223    11.00000    0.01546    0.01726 =
         0.01574   -0.00257    0.00099    0.00073
O4    4    0.238898    0.942509    0.315752    11.00000    0.01155    0.01878 =
         0.01722   -0.00206   -0.00143   -0.00144
N5    3    0.816578    1.148299    0.362397    11.00000    0.01200    0.01468 =
         0.01343   -0.00115   -0.00012   -0.00011
H5A   2    0.969576    1.088139    0.353472    11.00000    0.02911
H5B   2    0.809463    1.194604    0.433027    11.00000    0.01505
H5C   2    0.798009    1.268559    0.316083    11.00000    0.02045
C6    1    0.696715    0.762300    0.409269    11.00000    0.01416    0.01602 =
         0.01975    0.00029   -0.00055    0.00051
H6A   2    0.848362    0.702334    0.390663    11.00000    0.02019
H6B   2    0.582482    0.648754    0.393412    11.00000    0.02144
H6C   2    0.700471    0.806018    0.486583    11.00000    0.02403
HKLF 4
```

Remove these two lines

C:\Windows\system32\cmd.exe

```
Max. shift = 0.003 A for H6C      Max. dU = 0.001 for H5B
wR2 = 0.0597 before cycle   3 for      776 data and      84 /     84 parameters
GooF = S =     1.125;      Restrained GooF =     1.125 for       0 restraints
Mean shift/esd =   0.043  Maximum =    -0.189 for  U12 C1          at 15:10:27
Max. shift = 0.002 A for H5C      Max. dU = 0.000 for H6B
wR2 = 0.0597 before cycle   4 for      776 data and      84 /     84 parameters
GooF = S =     1.125;      Restrained GooF =     1.125 for       0 restraints
Mean shift/esd =   0.004  Maximum =    -0.017 for  U23 C2          at 15:10:27
Max. shift = 0.000 A for H5C      Max. dU = 0.000 for H5C
wR2 = 0.0597 before cycle   5 for      776 data and      84 /     84 parameters
GooF = S =     1.126;      Restrained GooF =     1.126 for       0 restraints
Mean shift/esd =   0.001  Maximum =     0.007 for  U11 H5C         at 15:10:27
Max. shift = 0.000 A for H5C      Max. dU = 0.000 for H5C
wR2 = 0.0597 before cycle   6 for      776 data and      84 /     84 parameters
GooF = S =     1.125;      Restrained GooF =     1.125 for       0 restraints
Mean shift/esd =   0.000  Maximum =     0.002 for   y  H5C         at 15:10:27
Max. shift = 0.000 A for H5C      Max. dU = 0.000 for H5C
wR2 = 0.0597 before cycle   7 for      776 data and      84 /     84 parameters
GooF = S =     1.125;      Restrained GooF =     1.125 for       0 restraints
Mean shift/esd =   0.000  Maximum =     0.000 for  U11 H5C         at 15:10:27
Max. shift = 0.000 A for H5C      Max. dU = 0.000 for H5C
wR2 = 0.0597 before cycle   8 for      776 data and      84 /     84 parameters
GooF = S =     1.125;      Restrained GooF =     1.125 for       0 restraints
Mean shift/esd =   0.000  Maximum =     0.000 for   y  O3          at 15:10:27
Max. shift = 0.000 A for H5C      Max. dU = 0.000 for H5A
wR2 = 0.0597 before cycle   9 for      776 data and       2 /     84 parameters
GooF = S =     1.125;      Restrained GooF =     1.125 for       0 restraints
R1 =  0.0215 for      771 Fo > 4sig(Fo)  and  0.0217 for all     776 data
wR2 =  0.0597,  GooF = S =   1.125,  Restrained GooF =    1.125  for all data
Flack x =     0.023(35) from 290 selected quotients (Parsons' method)
     0  atoms may be split and      0  atoms NPD
R1 =  0.0228 for      485 unique reflections after merging for Fourier
Highest peak    0.15  at  0.0245  0.4855  0.6539  [  0.76 A from C1 ]
Deepest hole   -0.16  at  0.1703  0.5363  0.0959  [  0.84 A from N5 ]

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
+  alanine       finished at 15:10:27   Total elapsed time:      0.19 secs  +
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

D:\My Documents\ppt\iucr_2014>_
```

Code	Friedif	*x* (Normal Ref't)	*x*(QUOT)
L-Alanine	34	-0.04(27)	0.01(4)
GKO02	32	0.01(15)	0.02(3)
R-CYCLO	21	-0.02(27)	0.00(4)
TWA16A	13	0.00(69)	0.18(8)
Cholestane	9	-0.01(77)	-0.01(13)

Results are the same as Hooft (PLATON) method with a Gaussian PDF

Good results are also obtained for CHNO materials with Mo-Kα radiation provided data go to high resolution.

Results are also fairly insensitive to disorder modelling: you get a pretty good indication of the value of x early in refinement.

Hooft, Straver & Spek. J. Appl. Cryst. (2008), **41**, 96.
Escudero-Adán, Benet-Buchholz & Ballester. Acta Cryst. (2014), **B70**, 660

Some things to bear in mind

1. Why it works
2. Validation
3. Outliers
4. What to do when it doesn't work

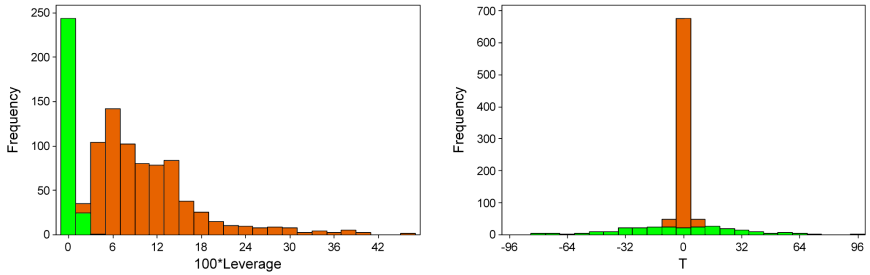
1. Why it works Differences

The original idea of using quotients was that systematic errors would cancel out.
Actually using a procedure based on differences works just as well.

$$F_o^2(\mathbf{h}) - F_o^2(-\mathbf{h}) = (1-2x)(F_{\text{single}}^2(\mathbf{h}) - F_{\text{single}}^2(-\mathbf{h}))$$
$$D_o(\mathbf{h}) = (1-2x)D_{\text{single}}(\mathbf{h})$$

Parsons, Flack & Wagner Acta Cryst (2013) **B69**, 249

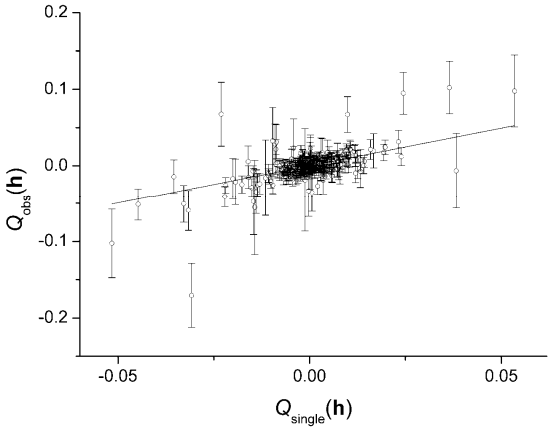
1. Why it works



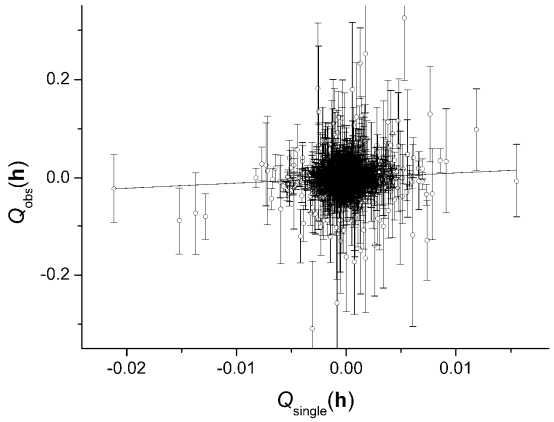
(*b*)

Figure 3
Leverage analysis for L-alanine[B]. (*a*) Leverages of $|F_{obs}(\mathbf{h})|^2$ and difference data, $D_{obs}(\mathbf{h})$. (*b*) Relative influences of $|F_{obs}(\mathbf{h})|^2$ and $D_{obs}(\mathbf{h})$ on the precision of the Flack parameter (as expressed by the quantity T). In each case $|F_{obs}(\mathbf{h})|^2$ data are shown in orange and $D_{obs}(\mathbf{h})$ data in green.

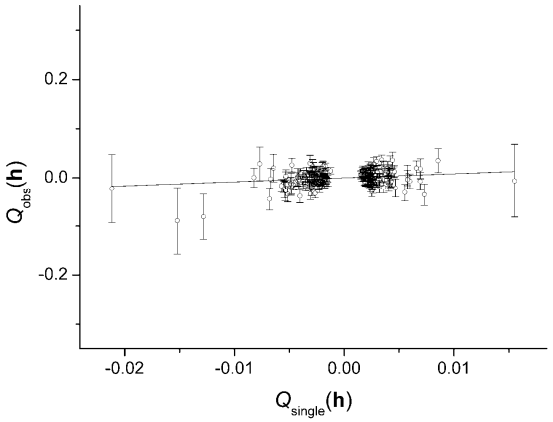
2. Validation Alanine (34)

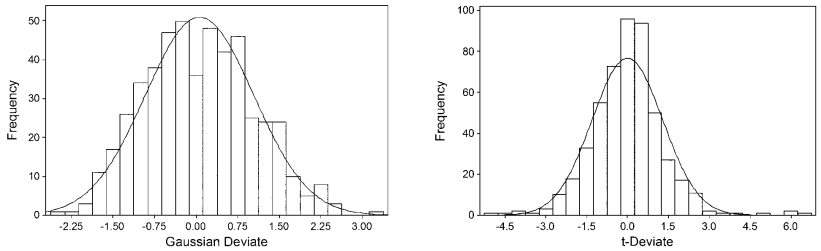


Cholestane (9)



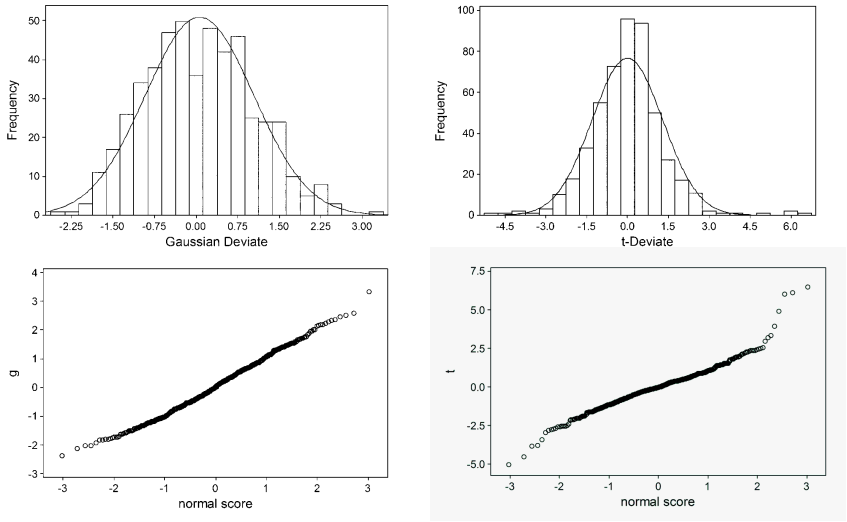
Cholestane (9)



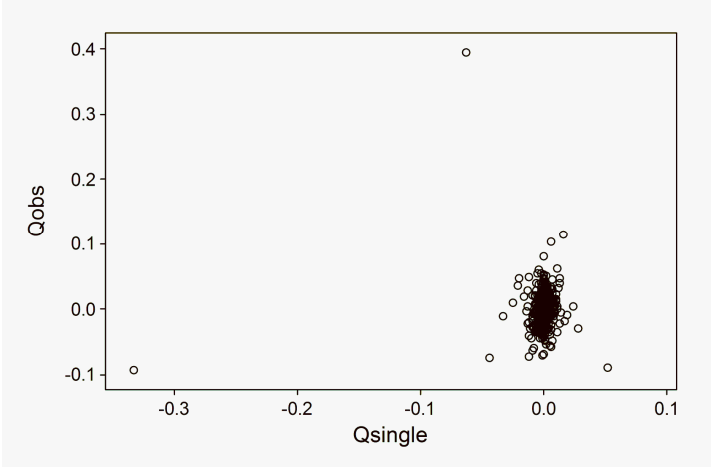


Plot a histogram of deviates:

$$\frac{[Q_{obs}(\mathbf{h}) - Q_{model}(\mathbf{h})]}{u[Q_{obs}(\mathbf{h})]}$$



3. Outlier Detection



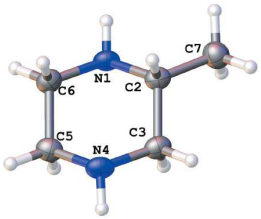
Outlier omission 0.18(8) → 0.08(8)

3. Outlier Detection

- Remove Bijvoet pairs if $D_o(\mathbf{h}) > 2D_{c,\,max}$
- For quotient calculations remove data where $F_o^2(\mathbf{h})/u(F_o^2(\mathbf{h}))$ and $F_c^2(\mathbf{h})/u(F_o^2(\mathbf{h}))$ are < 3.

$$\frac{F_o^2(\mathbf{h}) - F_o^2(-\mathbf{h})}{F_o^2(\mathbf{h}) + F_o^2(-\mathbf{h})} = (1-2x)\frac{F_{single}^2(\mathbf{h}) - F_{single}^2(-\mathbf{h})}{F_{single}^2(\mathbf{h}) + F_{single}^2(-\mathbf{h})}$$

4. When it doesn't work…



	Flack (TWIN)†	Flack (Post)‡	Parsons' Q§	Hooft-G¶	Hooft-S††
Rmpip	0.216 (871)	0.214 (832)	0.09 (23)	0.3 (3)	0.09 (8)

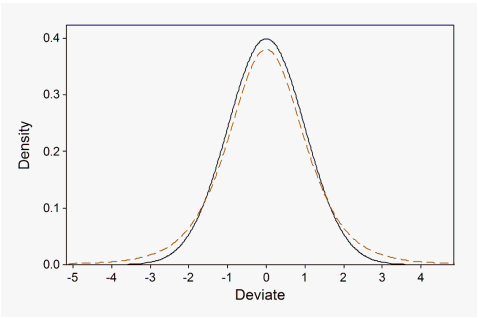
Riebenspies & Bhuvanesh *Acta Cryst*. (2013), B69, 288

4. When it doesn't work…

- The method may fail if the assumption of normally-distributed errors fails.
- In this case a different error model may produce 'better' results.
- The BIJVOET method in PLATON can use a Student-t distribution. The value of nu required is a measure of how far a distribution varies from normality.

Hooft, Straver & Spek. J. Appl. Cryst. (2008), **41**, 96.

$$p(y=0\,|\,\text{observations}) = \frac{p(\text{observations}\,|\,y=0)\,p(y=0)}{p(\text{observations})}$$

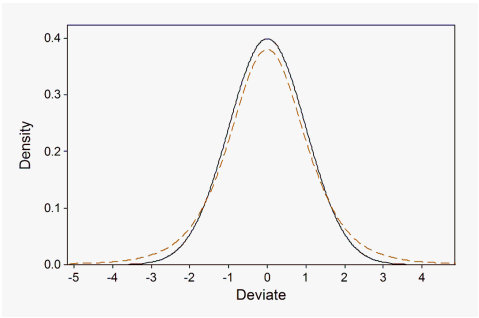


$$z_h = \frac{\Delta F^2_{\text{single}}(\mathbf{h}) - \Delta F^2_o(\mathbf{h})}{u(\Delta F^2_o(\mathbf{h}))}$$

$$p(z_h) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z_h^2}{2}\right)$$

$$p(\text{observations}\,|\,y=0) = \prod p(z_h)$$

$$p(y=0\,|\,\text{observations}) = \frac{p(\text{observations}\,|\,y=0)\,p(y=0)}{p(\text{observations})}$$



$$z_h = \frac{\Delta F^2_{\text{single}}(\mathbf{h}) - \Delta F^2_o(\mathbf{h})}{u(\Delta F^2_o(\mathbf{h}))}$$

$$p(z,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{(\nu\pi)}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

$$p(\text{observations}\,|\,y=0) = \prod p(z_h)$$

If nu~15 or more the pdf is ~normal.

Hooft, Straver & Spek. *J. Appl. Cryst.* (2010), **43**, 665

- Quotient (and difference) methods now available in ShelxL yield smaller – and more realistic – standard uncertainties than TWIN/BASF for low FRIEDIF crystals.
- Validation is important: one or two outliers affect the results.
- A utility called XNPP is available to anyone who would like it.
- Collect Friedel-complete data at LT with as much redundancy as you can.*
- Use the centrosymmetric Laue group for the multiscan absorption correction (e.g. SADABS).*
- PLATON/BIJVOET gives similar results if a Gaussian pdf is used.
- 'Poor' data may require use of a Student-t pdf, but if you use this:
  - Quote the value of nu
  - Try to explain why your errors aren't normally distributed

Acknowledgements

- Howard Flack (Geneva)
- Trixie Wagner (Novartis)
- Alan Coelho (Topas)
- Richard Cooper & David Watkin (Oxford)
- George Sheldrick (Göttingen)