Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 26 August 2010
Accepted 28 September 2010

Is there a fundamental upper limit for the significance $I/\sigma(I)$ of observations from X-ray and neutron diffraction experiments?

Julian Henn[a]* and Kathrin Meindl[b]

[a]Laboratory of Crystallography, University of Bayreuth, Universitätsstrasse 30, 95447 Bayreuth, Germany, and [b]Georg-August University of Göttingen, Tammannstrasse 4, 37077 Göttingen, Germany. Correspondence e-mail: julian.henn@uni-bayreuth.de

The answer is yes. A fundamental limit exists, which is not strongly applicable to individual reflections but to a sufficiently large set of reflections such as any set for structure determination. The limit originates from Poisson statistics which gives a minimum (average) error. The proposed limit of significance and a proposed decrease in significance due to data processing are also tested by monitoring $W = \langle I_{hkl}^{1/2}\rangle / \langle \sigma(I_{hkl})\rangle$ for raw and for Bragg data. Since Poisson statistics are the lower limit for the experimental standard uncertainties, it is expected that $W < 1$ for raw and Bragg data, and that W decreases upon data processing. W also gives a measure of systematic errors in the experimental data as $W \simeq 1$ characterizes pure Poisson data, $W \gg 1$ is physically impossible for sufficiently large data sets of unmerged reflections and $W < 1$ describes the contamination with systematic errors. Systematic differences depending on the software used to process the data were found. Also, the frequency distributions in particular of $\sigma(I)$ values change considerably depending on the data-processing software used. We have no explanation for these differences in the distributions of $\sigma(I)$, which lead to distinct changes in the frequency distribution of the significances $I/\sigma(I)$ compared with the raw data. Another consequence of Poisson statistics is that lower limits also exist for the agreement factors, the internal agreement factor and the goodness of fit. These limits depend on the moments $\langle 1/I_o\rangle$, $\langle I_o\rangle$, $\langle I_o^2\rangle$ and $\langle I_o^{1/2}\rangle$ of the observed set of intensities I_o about the origin. These agreement factors are theoretically attainable when no systematic sources of error apply. They may be used in future to construct further measures of systematic error in experimental data.

1. Introduction

Electromagnetic radiation comes in bunches and exposure times are always limited; this has important consequences for the signals from X-ray experiments: a fundamental upper limit exists for the significance of a large set of reflections. This limit is induced by Poisson statistics. Poisson statistics also apply to the variances of neutron diffraction data (Blessing, 1987), therefore these limits also apply to neutron diffraction experiments (Jeffrey, 1992). Poisson statistics place measurement errors and intensities on an absolute scale. Individual reflections respect this limit only when measured with high redundancy. Owing to stochastic fluctuations a low redundancy may lead to too significant or too insignificant reflections, such that for large data sets this should average out, unless weighting schemes systematically suppress part of the data. It is well known from the literature that the calculation of variances from small samples of Poisson-distributed numbers is difficult (see *e.g.* Haight, 1967, and references therein). The problem of obtaining a large range of values for standard uncertainties in repeated measurements of the same reflection is also a consequence of the Poisson statistics and it becomes increasingly important the lower the expected mean intensity value is. This has important consequences for weak and zero-intensity observations. We propose that the significance of a reflection is a property of the reflection itself (when the intensity of the incoming beam and exposure times are fixed) and should be invariant under data processing, *i.e.* changes in individual reflections that can effectively be described by an individual scaling factor should also be applied to the standard uncertainty of the reflection. To be more precise: the invariance under such scaling transformations is given when the scaling factors are known to have arbitrary numerical accuracy. If the scaling factors are affected by an error, the transformation must lead to decreasing significances. Data-processing steps involving subtraction as in the correction for the background signal also decrease the significance or leave it constant. We test experimental data

with respect to the question whether these are in conflict with a Poisson distribution and with respect to the question of conserved or decreasing significance.

2. Data processing

In order to obtain Bragg data from the integrated raw intensities, several processing steps have to be made. First of all, the background intensity I_{bg} is usually subtracted from I_{raw}, leading to a reduced net intensity I_{net},

$$I_{net} = I_{raw} - I_{bg}. \tag{1}$$

As this increases[1] the variance of the net intensity,

$$\sigma^2(I_{net}) = \sigma^2(I_{raw}) + \sigma^2(I_{bg}) \geq \sigma^2(I_{raw}), \tag{2}$$

it reduces the significance $I/\sigma(I)$ of the reflections,

$$\frac{I_{raw}}{\sigma(I_{raw})} \geq \frac{I_{net}}{\sigma(I_{net})}. \tag{3}$$

The last equation is valid for individual reflections. The equal sign only holds for a vanishing background. The net intensity is subject to further transformations, such as Lorentz and polarization corrections, and of course it must be weighted with the measurement time. Corrections owing to scattering by air may also be taken into account. We describe all the corrections and weightings together in one effective positive scaling factor k, which may be different for each individual reflection,

$$I_{Bragg} = kI_{net}. \tag{4}$$

How does this affect the variances? With a little algebra (see supplementary material, §3) one obtains from the k-scaled intensities

$$\sigma^2(I_{Bragg}) = k^2\sigma^2(I_{net}), \tag{5}$$

i.e. the standard uncertainties σ scale like the intensities. As a consequence, *the significance I/σ(I) of a reflection is invariant under scaling with an exactly known scaling factor*. Therefore, equation (2) can be extended,

$$\frac{I_{raw}}{\sigma(I_{raw})} \geq \frac{I_{net}}{\sigma(I_{net})} = \frac{I_{Bragg}}{\sigma(I_{Bragg})}. \tag{6}$$

Please note that the inverse is also true: when standard uncertainties are scaled, the intensities have to be scaled with the same factor. When the reflection is scaled to a time interval, this introduces additional variances owing to the uncertainty in measuring the time interval. Equations (5) and (6) are, in view of these (and more) additional variances, still too optimistic and should read better as

$$\sigma^2(I_{Bragg}) = k^2\sigma^2(I_{net}) + \sigma_x^2, \tag{7}$$

where σ_x^2 represents a positive number in which all these effects are collected. This leads to the following statement: *the individual significance of observations decreases during data processing or remains constant (at best), but it never increases*:

$$\frac{I_{raw}}{\sigma(I_{raw})} \geq \frac{I_{net}}{\sigma(I_{net})} \geq \frac{I_{Bragg}}{\sigma(I_{Bragg})}. \tag{8}$$

This equation holds strongly only when a Bragg intensity is derived from each individual raw intensity.

In the case of merged reflections the inequality (8) holds for the arithmetic mean. From everything said so far an upper limit for the significance can be deduced.

On average, the inequality

$$\sigma(I_{Bragg}) > I_{Bragg}^{1/2} \tag{9}$$

should hold. This is indeed a conservative estimate, as the true standard uncertainty additionally contains the background variance and systematic sources of error. For a whole data set this implies

$$\sum_{hkl} \sigma(I_{hkl}) > \sum_{hkl} I_{hkl}^{1/2}. \tag{10}$$

The inequality is also valid for the arithmetic averages

$$\langle\sigma(I_{hkl})\rangle > \langle I_{hkl}^{1/2}\rangle, \tag{11}$$

or, with the abbreviation W,

$$W = \frac{\langle I_{hkl}^{1/2}\rangle}{\langle\sigma(I_{hkl})\rangle} < 1. \tag{12}$$

In words: *owing to the Poisson process of photon/neutron arrivals, the ratio between the mean square root of the intensities and the mean standard uncertainty should be smaller than 1*. This is because the square root is the minimum (average) error and other errors enter additionally. Every error entering contributes another variance to the total sum of variances. Obviously, a similar expression,

$$W_2 = \frac{\langle I_{hkl}\rangle}{\langle\sigma^2(I_{hkl})\rangle} < 1, \tag{13}$$

could be derived in an analogous way. For small data sets with a large mean intensity, these W values may scatter in a small region around 1, but for large data sets it is expected that the W values decrease with decreasing total mean intensity. They express a statistical relationship, not a causal one. An optimal value for W and W_2 would be close to 1. That equations (10)–(13) do not hold for individual reflections is also a consequence of the Poisson process. This process does not only exhibit a large absolute variance, but when the variance is calculated from a small number of observations it scatters itself widely. Individual values from small samples may therefore severely underestimate the true variance (and others will overestimate it). Equation (13) also provides a consistency check for the experimental variance like equation (12). The contents of these two equations, however, are not independent and the following discussions refer to the analysis of equation (12). Equation (12) [and (13)] should not only hold for Bragg data but generally, *i.e.* also for the raw intensities and their standard uncertainties and each intermediate step in data processing. With W_{Bragg} abbreviating the calcu-

[1] It does not matter whether addition or subtraction takes place, the total variance always increases as it is a measure of the total uncertainty composed of individual additive uncertainties.

lation of W according to equation (12) with Bragg intensities $I_{\rm Bragg}$ and Bragg standard uncertainties $\sigma_{\rm Bragg}(I_{\rm Bragg})$ and $W_{\rm raw}$ abbreviating the calculation of W with raw intensities $I_{\rm raw}$ and raw standard uncertainties $\sigma_{\rm raw}(I_{\rm raw})$, the results can be summarized as

$$1 \geq W_{\rm raw} > W_{\rm Bragg} \geq 0, \tag{14}$$

which will be tested in §3. The requirement of W being smaller or equal to one stems from assuming $\sigma(I)$ to be at least equal to the value from Poisson statistics for the raw data, whereas the first inequality is a consequence of equations (1) and (2).

2.1. Consequences of the Poisson distribution on the agreement factors and on the goodness of fit

Under the optimistic assumption that the variance of the reflections is given only by counting statistics, one can derive limiting values for the agreement factors and the goodness of fit. Usually, to calculate R values a model is refined against the data; model-derived entities are used such as the calculated structure amplitudes $F_{\rm calc}$ (or calculated structure factors $F^2_{\rm calc} = I_{\rm calc}$) for a comparison with the experimental data. For the purpose of this article we may assume that we have the 'true' model with no systematic errors and perfect measurements and machines. All the remaining deviations from the expected intensity values are then solely due to the Poisson statistics. Under these conditions the calculated intensities are identical to the true intensities and it is possible to calculate expectation values for the agreement factors by using Poisson statistics. The R factors calculated in this way give an impression of what can be achieved at best experimentally. They also give a measure of what fraction of each R value is due to Poisson statistics. The remaining part must then be due to other sources of error. Now suppose there exists a perfect model and perfect data. The perfect data conform completely to the applied scattering theory. The perfect model is completely satisfactory and adequate in all electron-density and thermal-motion respects. Therefore, the data are unbiased on the true intensities (which are also the calculated intensities),

$$\lim_{N\to\infty} \langle I_{\rm o} \rangle = I_{\rm c} \tag{15}$$

with a variance

$$\sigma^2(I_{\rm o}) = I_{\rm o}. \tag{16}$$

In this case, the assumption

$$\big||F_{\rm o}| - |F_{\rm c}|\big| \simeq \sigma(F_{\rm o}) \tag{17}$$

is justified in a statistical sense for a large data set, *i.e.* it is valid on average $[\sum_{i=1}^{N} \big||F_{{\rm o},i}| - |F_{{\rm c},i}|\big| \simeq \sum_{i=1}^{N} \sigma(F_{{\rm o},i})]$. The agreement factor R_1 is defined according to

$$R_1 = \frac{\sum \big||F_{\rm o}| - |F_{\rm c}|\big|}{\sum |F_{\rm o}|}. \tag{18}$$

We use equations (16)–(18) and, from the literature, (19) (Massa, 1996),

$$\sigma^2(F) = \frac{\sigma^2(I)}{4F^2}, \tag{19}$$

to derive after a short calculation

$$R_1 = \frac{1}{2}\frac{1}{\langle I_{\rm o}^{1/2} \rangle}, \tag{20}$$

with the mean of the square root of the intensity $\langle I_{\rm o}^{1/2} \rangle$. In real data sets the variance of the structure factors will be on average larger. As a consequence, the value for R_1 obtained from equation (20) is a lower limit. wR_1 is defined according to

$$wR_1 = \left(\frac{\sum w \big||F_{\rm o}| - |F_{\rm c}|\big|^2}{\sum w |F_{\rm o}|^2} \right)^{1/2}. \tag{21}$$

Using again equations (16), (17) and (19) gives

$$wR_1 = \frac{1}{2}\left(\frac{\sum w}{\sum w I_{\rm o}} \right)^{1/2}, \tag{22}$$

which can be further simplified to

$$wR_1 = \frac{1}{2}\frac{1}{\langle I_{\rm o} \rangle^{1/2}} \tag{23}$$

for $w = 1$, or for w being any constant. For weights $w = 1/\sigma^2(I)$ one arrives at

$$wR_1 = \frac{1}{2}\langle 1/I_{\rm o} \rangle^{1/2}, \tag{24}$$

which is different from equation (23), as in equation (23) the intensities are averaged first and then the reciprocal is taken. This leads to $wR_1|_{w=1/\sigma^2(I)} \geq wR_1|_{w=\rm const.}$ by the Cauchy–Schwarz inequality [see *Appendix A* for more details; Bronstein *et al.*, 2008: equation (1.114*b*), p. 31]. The appearance of a zero intensity or a close-to-zero intensity poses no problem in equation (23), it just lowers the average value, whereas in equation (24) the average of the reciprocal intensities is calculated. A single close-to-zero-intensity observation might significantly increase the total result and zero-intensity observations must be excluded from the evaluation of equation (24).

wR_2 follows from the definition

$$wR_2 = \left(\frac{\sum w (F_{\rm o}^2 - F_{\rm c}^2)^2}{\sum w (F_{\rm o}^2)^2} \right)^{1/2} \tag{25}$$

after setting $w = 1$ or to any other constant,

$$\begin{aligned} wR_2|_{w=1} &= \left[\frac{\langle (\Delta I_{\rm o})^2 \rangle}{\langle I_{\rm o}^2 \rangle} \right]^{1/2} \\ &= \left[\frac{\langle \sigma^2(I_{\rm o}) \rangle}{\langle I_{\rm o}^2 \rangle} \right]^{1/2} \\ &= \left[\frac{\langle I_{\rm o} \rangle}{\langle I_{\rm o}^2 \rangle} \right]^{1/2}, \end{aligned} \tag{26}$$

i.e. $wR_2|_{w=1}$ is determined by the square root of the ratio of the first and the second moment of the intensity distribution about zero. From equation (25) with $w = 1/\sigma^2(I_{\rm o})$ it follows that

$$wR_2|_{w=1/\sigma^2(I_o)} = \frac{1}{\langle I_o\rangle^{1/2}}. \quad (27)$$

From the definition $R_{\rm int} = \sum\left|I_o - \langle I_o\rangle\right|/\sum I_o$ with $\left|I_o - \langle I_o\rangle\right| \simeq \sigma(I_o)$ and Poisson statistics it follows that

$$R_{\rm int} = \frac{\langle I_o^{1/2}\rangle}{\langle I_o\rangle}. \quad (28)$$

Equations (20), (23), (24), (26), (27) and (28) are to be understood as limiting values, so all real data should result in values well above these, which are only theoretically attainable. The equations are valid only under the optimistic assumption expressed by equation (16) (which corresponds to a Poisson distribution) and they state that even with the absolutely correct electron-density and thermal-motion models, with a perfect imperfect crystal, and with data not affected by systematic errors, background signals and so on, the R values should be larger than zero. The equations also say that it is reasonable to expect the R values to become smaller for data sets with low resolution in $\sin\theta/\lambda$, heavily diffracting atoms, and long exposure times and low temperatures, as all of these will increase $\langle I_o\rangle$. Inversely, the inclusion of a large amount of weak data leads to increasing R values as these diverge: $\lim_{\langle I_o\rangle\to 0} R = \infty$.

A similar expression can be derived for the goodness of fit, S,

$$S = \left[\frac{\sum w(F_o^2 - F_c^2)^2}{n-p}\right]^{1/2} \quad (29)$$

with weights w, number of reflections n and number of model parameters p. We set $p = 0$ to approach the limit in which deviations from the mean value are caused solely by Poisson statistics. If again equation (16) is valid and we choose to use unit weights $w = 1$, then from the definition follows

$$S = \langle\sigma^2(I_o)\rangle^{1/2} = \langle I_o\rangle^{1/2}, \quad (30)$$

which states that the expected value for the goodness of fit is the square root of the mean intensity. This can also be used as a test as to whether or not the data are consistent with a Poisson distribution since in real data $S \gg \langle I_o\rangle^{1/2}$ is expected, and if it is found that $S < \langle I_o\rangle^{1/2}$ then something is most likely wrong. It can also be seen from equation (30) that the resolution shell increment of S with unit weights tends to decrease with the resolution shell as the mean intensity tends to decrease with increasing resolution in $\sin\theta/\lambda$.

If $w = 1/\sigma^2(I_o)$ is chosen in equation (29), it follows that $S \approx 1$. In this case S should also be equal to 1 for sufficiently large resolution shells, however, a larger deviation from unity is not only acceptable but expected for the inner resolution shells. Also, larger deviations are expected the smaller the sample that S is calculated from, as equation (16) is a statistical relation that need not hold for a small sample.

3. Tests on experimental data

We used an empirical approach by analyzing a variety of raw data files and processed reflection data files. The data available from a raw file from *SAINT* (Bruker, 2006) correspond to $I_{\rm raw}$ (PKSUM), $I_{\rm net}$ (FI) and $\sigma(I_{\rm net})$ (SI). Unfortunately, there is no $\sigma(I_{\rm raw})$ listed, which would allow for a check in consistency of the raw data with Poisson statistics. Therefore, we have to postulate that this consistency is given. However, the entries 'FI', which is identified with $I_{\rm net}$, and 'SI', which is identified with $\sigma(I_{\rm net})$, give the opportunity for a check by calculation of the mean values of the sums from equation (12). The results should indicate consistency with Poisson statistics. A ratio larger than 1 indicates a conflict with Poisson statistics. According to the preservation of significances, the ratio of the sums should remain constant or decrease, but not increase, when proceeding from the net data to the Bragg data. Keeping the decrease as small as possible is important in order to extract as much information as possible from the experimental data. The data were not merged, such that the initial distribution of σ values stems from the integration software and such that the individual σ values are kept separate for multiply observed structure factors. Details of data processing for the investigated data sets can be taken from the cited references.

We start the discussion of Table 1 with column 4, $\langle I_{\rm net}/I_{\rm net}^{1/2}\rangle$. This entry serves two purposes: first it gives the mean value of the ratio between the net intensity and its square root for the raw data. This can be interpreted as an upper limit for the best possible average significance, provided the intensities are on an absolute scale. Second, the entry also characterizes the data, $\langle I_{\rm net}/I_{\rm net}^{1/2}\rangle = \langle I_{\rm net}^{1/2}\rangle$, a number giving (indirectly) control over scaling. The numbers in this column vary between 12.097 (data set 6) and 300.936 (data set 3). These should be compared with those in column 7 with the heading $\langle I_{\rm Bragg}/I_{\rm Bragg}^{1/2}\rangle$. These are the equivalent values for the processed (Bragg) data as in column 4 for the raw data. The data were processed with *SADABS* (upper line) and *SORTAV* (lower line), however, the raw data were the same in each case. It is obvious that the scale of the data processed with *SADABS* changes heavily, for example, for data set 1 the change is from 93.656 for the raw data to 7.681 for data processed with *SADABS* and to 102.540 for data processed with *SORTAV*, whereas data processed with *SORTAV* generally have little changes in the scale. However, this is just a side observation as a change of scale is perfectly valid, when the corresponding $\sigma(I)$ values are changed accordingly. This brings us to the next pair of columns.

Column 2 gives the ratio between the average square root of the intensity and the average standard uncertainties for the raw data. This column also serves more than one purpose. First of all it gives a numerical value which should, for theoretical reasons, be well below 1. Recall that the variance of a Poisson distributed number is equal to the λ parameter of the Poisson distribution. From this it follows that the value in column 2 should be close to 1 if only counting statistics apply. On the other hand, if there are systematic sources of error, the variance of the Poisson distribution gives a lower limit to the

Table 1
Mean intensities and mean standard uncertainties, and their ratio for pre-processed unmerged data after integration with *SAINT* (columns 2, 3 and 4) and processed unmerged data after *SADABS* (Bruker, 2008) (columns 5, 6 and 7; first line of each data set) and after *SORTAV* (Blessing, 1995) (columns 5, 6 and 7; second line of each data set).

$W_{raw} = \langle I_{net}^{1/2}\rangle/\langle\sigma(I_{net})\rangle$ and $W_{Bragg} = \langle I_{Bragg}^{1/2}\rangle/\langle\sigma(I_{Bragg})\rangle$. I_{net} and $\sigma(I_{net})$ have been taken from the raw files. The resolution is approximately 0.8–0.9 Å for all the data sets. Reflections with $\sigma = 0$ have been excluded from the calculation of the average values in columns 3 and 6, as have observations with $I < 0$ in columns 5 and 6, and $I \leq 0$ in column 7.

No.	W_{raw}	$\langle I_{net}/\sigma(I_{net})\rangle$	$\langle I_{net}/I_{net}^{1/2}\rangle$	W_{Bragg}	$\langle I_{Bragg}/\sigma(I_{Bragg})\rangle$	$\langle I_{Bragg}/I_{Bragg}^{1/2}\rangle$	Reference	CCDC No.
1	0.274	27.822	93.656	2.187	20.744	7.681	Widjaja *et al.* (2008)	668189
				0.238	26.001	102.540		
2	0.183	51.272	265.654	1.429	25.373	13.341	Widjaja *et al.* (2008)	668187
				0.180	50.260	265.480		
3	0.203	64.361	300.936	0.934	24.974	20.131	Liang *et al.* (2005)	252113
				0.144	51.168	330.688		
4	0.252	29.036	103.123	0.955	13.603	11.006	Kurahashi *et al.* (2007)	642256
				0.197	24.971	116.829		
5	0.939	25.337	21.306	1.322	8.518	4.723	Stasch *et al.* (2009)	726162
				0.969	25.023	20.586		
6	0.765	13.890	12.097	0.776	5.505	5.316	Stasch *et al.* (2009)	726161
				0.759	13.741	12.101		
7	0.171	36.040	200.096	4.295	17.552	2.983	Unpublished	
				0.114	26.513	214.107		
8	0.451	17.376	32.943	1.365	13.135	7.661	Lösgen *et al.* (2007)	628365
				0.450	17.293	32.951		
9	0.193	30.715	143.050	1.839	18.301	8.202	Liang *et al.* (2005)	252114
				0.184	30.199	150.199		

true variance. In this case the value in column 2 will drop well below 1, which is the case for almost all data sets. Secondly, column 2 can be compared with the corresponding values for the processed data. These are in column 5. It is expected that data processing tends to increase $\langle\sigma(I)\rangle$, which should bring the ratio further down closer to zero. This is indeed the case for all data processed with *SORTAV* apart from data set 5, where there is a slight increase from 0.939 for the raw data to 0.969 for the processed data. For data set 1 the value drops from 0.274 for the raw data to 0.238 for the processed data and for data set 2 from 0.183 to 0.180. For data processed with *SADABS*, the ratio increases. For some data sets this increase is to values far above 1. An increase of W could indicate a downscaling of $\sigma(I)$, which would lead to an increase in significances.

Now an explicit check on the average significance of the data sets is appropriate. As the calculation of the significance requires division by $\sigma(I)$, the average values may be dominated by those reflections which have small standard uncertainties by means of stochastic fluctuations. Therefore, these average values should be taken with caution. Column 3 gives the average significance for the raw data. It varies between 13.890 (data set 6) and 64.361 (data set 3). The fact that systematic sources of error are at work was already seen from column 2 and it can additionally be deduced from a comparison of the values in columns 3 and 4, which give the mean significance and the upper limit for the attainable mean significance given the intensities and provided the data are on an absolute scale. For all data sets but 5 and 6, the values in column 4 are larger than those in column 3, as expected. For data sets 5 and 6 the ratio decreases, which may indicate that in these data sets the value $\langle I_{net}/\sigma(I_{net})\rangle$ is dominated by a few reflections with exceptionally small values for $\sigma(I_{net})$. Such characteristics of the raw data deserve more attention and systematic treatment, however, this goes beyond the scope of this work.

In any case, when comparing column 4 with the average significance of the processed data in column 6 it should drop or remain constant at best. For all data processed with *SORTAV* this is indeed the case. The average significance always decreases, if sometimes also only slightly. For most data sets this decrease in the average significance is even larger for data processed with *SADABS*. This indicates a loss of information by data processing with *SADABS*. There are seemingly conflicting indicators for these data. An increasing W ratio might be taken as a hint toward an increased significance. This would indeed be the case if all σ values were downscaled by the same factor (or very similar factors). However, when the average significance is monitored, it decreases despite the increase of W. It must be concluded that the *distributions* either of the I values or of the σ values (or both) have changed upon data processing.

To compare the distributions, histograms of the intensities, standard uncertainties and resulting significances are plotted in Fig. 1 for data set 1. Similar figures for data sets 2–9 can be found in the supplementary material.[2] As these properties are on different scales, they were normalized to unity, *i.e.* the maximum value was set to one in order to be able to compare these.

The first column of Fig. 1 shows the intensity distributions for raw and processed data. Each data set was rescaled such that the maximum intensity equals 1. Therefore, the x axis

[2] Supplementary data for this paper are available from the IUCr electronic archives (Reference: SH5117). Services for accessing these data are described at the back of the journal.

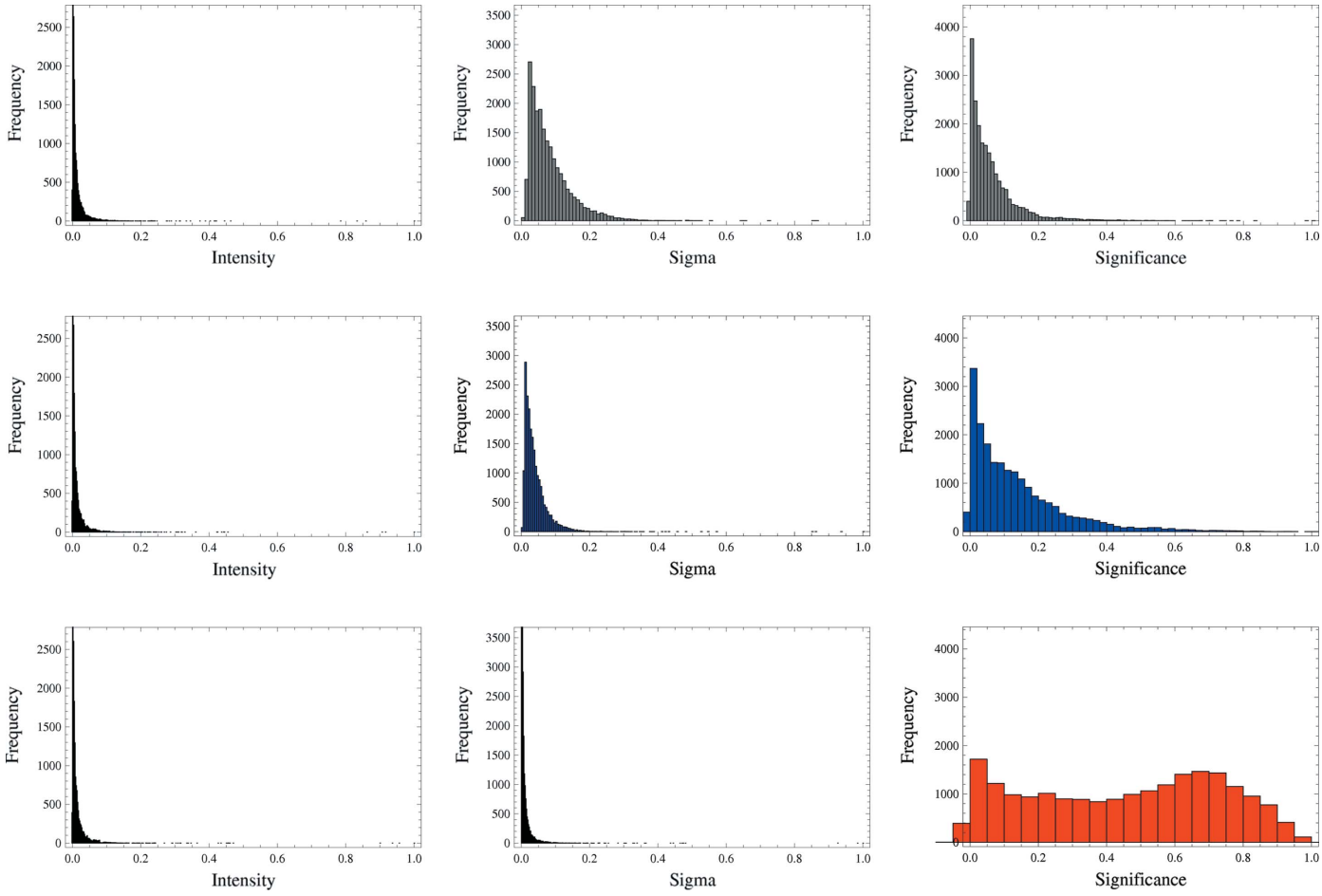

Figure 1
Frequency distributions of the intensities (first column), the standard uncertainties (second column) and the significances (third column) for raw data (first line), Bragg data after processing with *SORTAV* (second line) and after processing with *SADABS* (third line) for data set 1. The distributions were normalized to their maximum value.

gives the percentage of the maximum intensity, whereas the y axis gives the number of reflections found in this range. All three pictures show that almost 100% of the data are found in the lowest 10% of the intensity range, as almost all of the area under the graphs is confined to regions very close to zero and smaller than 0.1 on the x axis. Column 2 shows the corresponding distributions for the standard uncertainties. In both cases (*SADABS* and *SORTAV*) data processing leads to a shift to lower relative values in the frequency distributions, *i.e.* the percentage of large σ values is reduced and that of low σ values is increased. No comment is made on the absolute values. This shift is, however, more distinct for data processed with *SADABS*. This more distinct shift leads to a drastic change in the distributions of significances, in which the absolute values of I and $\sigma(I)$ also play a role. Column 3 shows that the frequency distributions for the significances resemble each other for the raw data (column 3, top) and for the data processed with *SORTAV* (column 3, middle), but not for the frequency distribution for the data processed with *SADABS* (column 3, bottom).

This shift in the distribution of significances was indicated by the coinciding observations of an increase in W, accompanied by a decrease in $\langle I/\sigma(I)\rangle$ as given in Table 1. Already a distinct increase in W is a suspicious incident, as it either indicates an overall increase of the significances by data processing, which violates the requirement of decreasing significances, or a redistribution of either I or $\sigma(I)$ frequency values. The figures for the other data sets are similar in that the distributions of I, $\sigma(I)$ and $I/\sigma(I)$ are similar for raw data and for data processed with *SORTAV*, but not for data processed with *SADABS*, where the distribution of $\sigma(I)$ values changes more distinctively, leading to a completely different frequency distribution of $I/\sigma(I)$ values.

Please note that the change in the distribution of σ values also leads to a larger fraction of significant negative-intensity observations. This particular observation and generally the observation of changes in the frequency distribution of σ values upon data processing may be important with respect to absolute structure determination, weighting schemes, Bijvoet differences, normal probability plots, the maximum entropy method of electron-density reconstruction and the construction of experimental error models (Hooft *et al.*, 2009). For a direct comparison of the significances from raw and Bragg data, we examined the significances of the first 1000 reflections

Table 2
The mean square root of the Bragg intensity and the mean Bragg standard uncertainty and their ratio W for data taken from the literature.

Different compounds in the same publication are indicated by lower-case italic letters ($a, b, c, \ldots$) in column 1. A variety of integration software has been used. For details see the cited literature. Reflections with negative intensity observations and their standard uncertainties have been excluded from the calculations.

No.	$\langle I_{\rm Bragg}^{1/2}\rangle$	$\langle\sigma(I_{\rm Bragg})\rangle$	W	Reference	Data reduction/absorption correction
10*a*	10.1	16.5	0.613	McMahon *et al.* (2009)	*SORTAV*
10*b*	8.7	11.4	0.762		*SORTAV*
10*c*	8.8	9.1	0.962		*SORTAV*
10*d*	9.0	9.1	0.992		*SORTAV*
11*a*	11.7	5.3	2.218	von Chrzanowski *et al.* (2007)	*SADABS*
11*b*	33.3	139.7	0.239		*SORTAV*
12*a*	16.2	15.1	1.075	Coles *et al.* (2007)	*SORTAV*
12*b*	22.2	23.1	0.961		*SORTAV*
12*c*	32.0	74.1	0.432		*SORTAV*
13*a*	12.4	12.7	0.975	Siddiqui *et al.* (2008)	*SORTAV*
13*b*	14.4	21.9	0.656		*SORTAV*
13*c*	14.1	30.7	0.461		*SORTAV*
14*a*	7.5	2.8	2.684	Yasuda *et al.* (2000)	*ABSCOR95* (Higashi, 1995)
14*b*	4.2	5.6	0.754		*ABSCOR95* (Higashi, 1995)
14*c*	7.8	3.2	2.423		*ABSCOR95* (Higashi, 1995)
14*d*	5.1	2.52	2.079		*ABSCOR95* (Higashi, 1995)
15*a*	161.3	797.8	0.202	Lu *et al.* (2006)	*ABSCOR98* (Jacobson, 1998)
15*b*	176.7	970.8	0.182		*ABSCOR98* (Jacobson, 1998)
16	17.7	24.8	0.714	Jin & Zhang (2005)	*ABSCOR98* (Jacobson, 1998)
17*a*	26.0	22.1	1.180	Bucher *et al.* (2008)	numerical method (Coppens *et al.*, 1965)
17*b*	26.5	87.3	0.304		numerical method (Coppens *et al.*, 1965)
18	153.3	1359.4	0.113	Tokaychuk *et al.* (2003)	analytical method with *X-RED* (Stoe & Cie, 1999)
19	34.0	33.0	1.028	Björk *et al.* (2001)	numerical method with *JANA*98 (Petříček & Dušek, 1997)
20	31.2	67.8	0.460	Zou & Dadachov (2000)	numerical method with *X-RED* (Stoe & Cie, 1997)
21*a*	9.7	4.0	2.412	Czapik *et al.* (2010)	*CrysAlis Red* (Oxford Diffraction, 2007)
21*b*	9.8	4.6	2.140		*CrysAlis Pro* (Oxford Diffraction, 2009)
22	12.2	26.2	0.465	Howard *et al.* (2010)	*CrysAlis Pro* (Oxford Diffraction, 2009)
23*a*	17.2	41.0	0.419	Galicia Aguilar & Bernès (2009)	*CrysAlis Red* (Oxford Diffraction, 2007)
23*b*	10.1	11.8	0.906		*XSCANS* (Siemens, 1996)

of data set 1. For these, the same number of repeatedly measured reflections was available. The data sets were brought into the same order with respect to h, k, l and the significance of the processed data is plotted against the significance of the raw data in Fig. 2. It should be mentioned that this order is not unique and therefore different plots can be generated from the same data. Blue stars mark the *SORTAV versus* raw significances, whereas red crosses mark the *SADABS versus* raw significances. The black dashed line corresponds to significances not changing upon data processing. Marks under this black dashed line depict reflections with reduced significance, marks above the black dashed line depict reflections with an increased significance in comparison to the raw data. The significances from data processed with *SADABS* do not exceed a value of roughly 40, whereas the significances of data processed with *SORTAV* keep increasing with increasing significances of the raw data. The reason for this differences is the shift in the frequency distributions of σ. A scaling of the standard uncertainties with a single scaling factor for the whole set would leave its normalized frequency distributions unaffected.

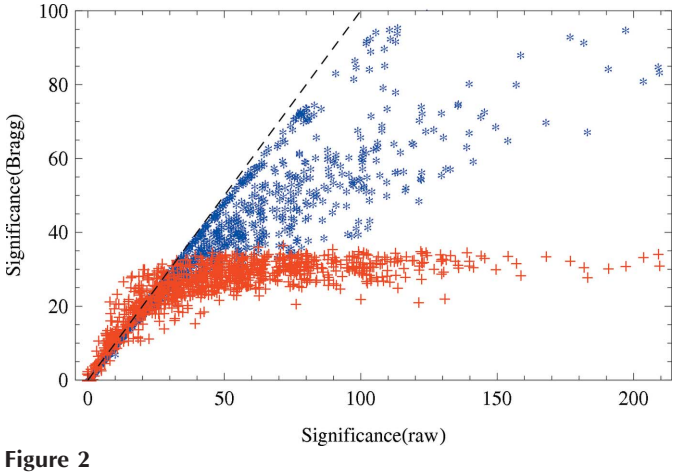

Figure 2
Significance of processed data *versus* significance of raw data from the first 1000 reflections of data set 1. Blue stars: *SORTAV versus* raw data. Red crosses: *SADABS versus* raw data. Black dashed line: significance of processed data equals the significance of raw data. The data processing started from the same raw data in both cases.

For an independent comparison we extracted reflection files from the literature (all data were downloaded from the web site for *Acta Crystallographica* Section C) and analyzed these with respect to the inequality equation (12). It was not possible to monitor the propagation of the significances as raw files were not available. The data were picked randomly without detailed prior investigation. The results are compiled in Table 2. The data differ from the data in Table 1 in the integration software used and in the data-processing software. Also, to ensure the applicability of W, the in-house data were not merged. This is not the case for the data in Table 2, such that in these cases one cannot immediately conclude from $W \gg 1$ that something must be wrong. Large values of W, however, still might indicate that an inspection of the raw data is appropriate.

The values for W scatter between 0.113 (data set 18) and 2.684 (set 14*a*). Even within one publication and using one software type, the values may scatter widely, *e.g.* for data set 14. This might indicate that the data-processing issues discussed above are not the problem of an individual software package.

4. Summary, outlook and conclusions

We investigated the propagation of the significance of reflections from the raw to the Bragg data. We explicitly formulated the laws of decreasing or constant significance during data procession and the law of the optimal ratio $W = 1$. The quality of the raw data limits the quality of the derived Bragg data. Although the requirement of decreasing significance might appear trivial to many readers, it was shown that this is not necessarily respected when handling real data sets. This holds for a variety of software, however, more comprehensive and systematic studies are required.

The inequalities used, like equations (12) or (13), may be employed in the future to characterize the quality of a data set separately for raw and Bragg data. In contrast to the Bragg intensity that theoretically obeys a Skellam distribution (Haight, 1967), *i.e.* it is derived from the difference between two theoretically Poisson distributed numbers corresponding to the peak intensity and the background intensity, the raw data are theoretically Poisson distributed, provided no systematic errors are present. The theoretical R values and goodness-of-fit value S, derived under the assumption of the absence of systematic sources of error and vanishing or negligible background intensity, may also be used as benchmark values for real data sets. Deviations from these theoretically attainable values indicate the effect of systematic sources of error, a topic which is of particular and increasing importance in charge-density studies. Therefore, the main conclusion is that much more needs to be done to characterize the quality of raw data and to investigate the reasons for and the effects of the appearance of systematic sources of error in the raw data. The present work is an attempt to establish measures of quality applicable to raw and to processed data and to find generally accepted rules of data processing, such as the rule of decreasing significances and optimal W values.

APPENDIX *A*
The Cauchy–Schwarz inequality applied to wR_1 for weights w = constant and $w = 1/\sigma^2(I)$

The Cauchy–Schwarz inequality states for N positive numbers a_i, b_i, $i = 1, ..., N$

$$(a_1b_1 + a_2b_2 + ... + a_nb_n)^2 \leq \left(a_1^2 + a_2^2 + ... + a_n^2\right) \times \left(b_1^2 + b_2^2 + ... + b_n^2\right). \quad (31)$$

Setting $a_i^2 = I_i/N$ and $b_i^2 = 1/I_iN$ leads to

$$\begin{aligned}\langle I\rangle\left\langle\frac{1}{I}\right\rangle &= \left(\sum_{i=1}^{N} a_i^2\right)\left(\sum_{i=1}^{N} b_i^2\right)\\ &\geq \left(\sum_{i=1}^{N} a_ib_i\right)^2\\ &= \left(\sum_{i=1}^{N} 1/N\right)^2\\ &= 1. \end{aligned} \quad (32)$$

Therefore

$$\left\langle\frac{1}{I}\right\rangle \geq \frac{1}{\langle I\rangle}, \quad (33)$$

and consequently

$$wR_1|_{w=1/\sigma^2(I)} \geq wR_1|_{w=\text{const.}} \quad (34)$$

by equations (23) and (24).

The authors thank Andreas Schönleber for careful reading of the manuscript and numerous helpful suggestions for improvements, Tim Grüne for fruitful discussions on statistics, Stefan Mebs for providing data and background information as well as for prosperous discussions on data processing and the subtle philosophical consequences of the existence of statistical ensembles, and George M. Sheldrick for discussions. This work was supported by the DFG (grant HE 4573/3-1).

References

Björk, H., Dabkowska, H., Greedan, J. E., Gustafsson, T. & Thomas, J. O. (2001). *Acta Cryst.* C**57**, 331–332.
Blessing, R. H. (1987). *Crystallogr. Rev.* **1**, 3–58.
Blessing, R. H. (1995). *Acta Cryst.* A**51**, 33–38.
Bronstein, I. N., Semendjajew, K. A., Musiol, G. & Mühlig, H. (2008). *Taschenbuch der Mathematik*. 7., vollständig überarbeitete und ergänzte Auflage. Frankfurt am Main: Verlag Harri Deutsch.
Bruker AXS Inc. (2006). *SAINT*. Bruker AXS Inc., Madison, Wisconsin, USA.
Bruker AXS Inc. (2008). *SADABS*. Bruker AXS Inc., Madison, Wisconsin, USA.
Bucher, J., Blacque, O., Schmalle, H. W. & Berke, H. (2008). *Acta Cryst.* C**64**, *m*87–*m*90.
Chrzanowski, L. S. von, Lutz, M. & Spek, A. L. (2007). *Acta Cryst.* C**63**, m377–m382.
Coles, S. J., Davies, D. B., Hacıvelioğlu, F., Hursthouse, M. B., Ibişoğlu, H., Kılıç, A. & Shaw, R. A. (2007). *Acta Cryst.* C**63**, o152–o156.
Coppens, P., Leiserowitz, L. & Rabinovich, D. (1965). *Acta Cryst.* **18**, 1035–1038.
Czapik, A., Konowalska, H. & Gdaniec, M. (2010). *Acta Cryst.* C**66**, o128–o132.
Galicia Aguilar, J. A. & Bernès, S. (2009). *Acta Cryst.* C**65**, o176–o178.
Haight, F. A. (1967). *Handbook of the Poisson Distribution*. New York: John Wiley and Sons Inc.
Higashi, T. (1995). *ABSCOR*. Rigaku Corporation, Tokyo, Japan.
Hooft, R. W. W., Straver, L. H. & Spek, A. L. (2009). *Acta Cryst.* A**65**, 319–321.
Howard, R. H., Theobald, N., Bochmann, M. & Wright, J. A. (2010). *Acta Cryst.* C**66**, o310–o312.
Jacobson, R. (1998). *ABSCOR*. Rigaku Corporation, Tokyo, Japan.
Jeffrey, G. A. (1992). *Accurate Molecular Structures, Their Determination and Importance*, edited by A. Domenicano & I. Hargittay,

pp. 270–289. IUCr Monographs on Crystallography, Vol. 1. Oxford University Press.
Jin, D. & Zhang, D.-C. (2005). *Acta Cryst.* C**61**, o657–o659.
Kurahashi, T., Kozhushkov, S. I., Schill, H., Meindl, K., Rühl, S. & de Meijere, A. (2007). *Angew. Chem. Int. Ed.* **46**, 6545–6548.
Liang, S., Lee, C.-H., Kozhushkov, S. I., Yufit, D. S., Howard, J. A. K., Meindl, K., Rühl, S., Yamamoto, C., Okamoto, Y., Schreiner, P. R., Rinderspacher, B. C. & de Meijere, A. (2005). *Chem. Eur. J.* **11**, 2012–2018.
Lösgen, S., Schlörke, O., Meindl, K., Herbst-Irmer, R. & Zeeck, A. (2007). *Eur. J. Org. Chem.* pp. 2191–2196.
Lu, S.-F., Huang, Z.-X. & Huang, J.-L. (2006). *Acta Cryst.* C**62**, i73–i75.
Massa, W. (1996). *Kristallstrukturbestimmung*. Stuttgart: Teubner.
McMahon, J., Gallagher, J. F., Anderson, F. P. & Lough, A. J. (2009). *Acta Cryst.* C**65**, o345–o351.
Oxford Diffraction (2007). *CrysAlis CCD* and *CrysAlis RED*. Oxford Diffraction Ltd, Abingdon, Oxfordshire, England.
Oxford Diffraction (2009). *CrysAlis Pro*. Oxford Diffraction Ltd, Yarnton, Oxfordshire, England.
Petříček, V. & Dušek, M. (1997). *JANA*98. Institute of Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic.
Siddiqui, W. A., Ahmad, S., Siddiqui, H. L. & Parvez, M. (2008). *Acta Cryst.* C**64**, o367–o371.
Siemens (1996). *XSCANS*, Version 2.21. Siemens Analytical X-ray Instruments Inc., Madison, Wisconsin, USA.
Stasch, A., Sarish, S. P., Roesky, H. W., Meindl, K., Dall' Antonia, F., Schulz, T. & Stalke, D. (2009). *Chem. Asian J.* **4**, 1451–1457.
Stoe & Cie (1997). *X-RED*, Version 1.09/Windows, and *IPDS Software*. Stoe and Cie GmbH, Darmstadt, Germany.
Stoe & Cie (1999). *X-RED*, Version 1.19, and *IPDS Software*, Version 2.92. Stoe and Cie, Darmstadt, Germany.
Tokaychuk, Y. O., Filinchuk, Y. E., Fedorchuk, A. O. & Bodak, O. I. (2003). *Acta Cryst.* C**59**, i125–i127.
Widjaja, T., Fitjer, L., Meindl, K. & Herbst-Irmer, R. (2008). *Tetrahedron*, **64**, 4304–4312.
Yasuda, N., Uekusa, H. & Ohashi, Y. (2000). *Acta Cryst.* C**56**, 1364–1366.
Zou, X. & Dadachov, M. S. (2000). *Acta Cryst.* C**56**, 738–739.