Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Supplementary Material to: 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

1. Frequency distributions of I, $\sigma(I)$ and $I/\sigma(I)$ for raw and processed data of data sets 2–9

The frequency distributions for intensities (left column), standard uncertainties (middle column) and significances (right column) for raw data (first line), data processed with SORTAV (middle line) and data processed with SADABS (third line) for the in-house data sets 2-10.

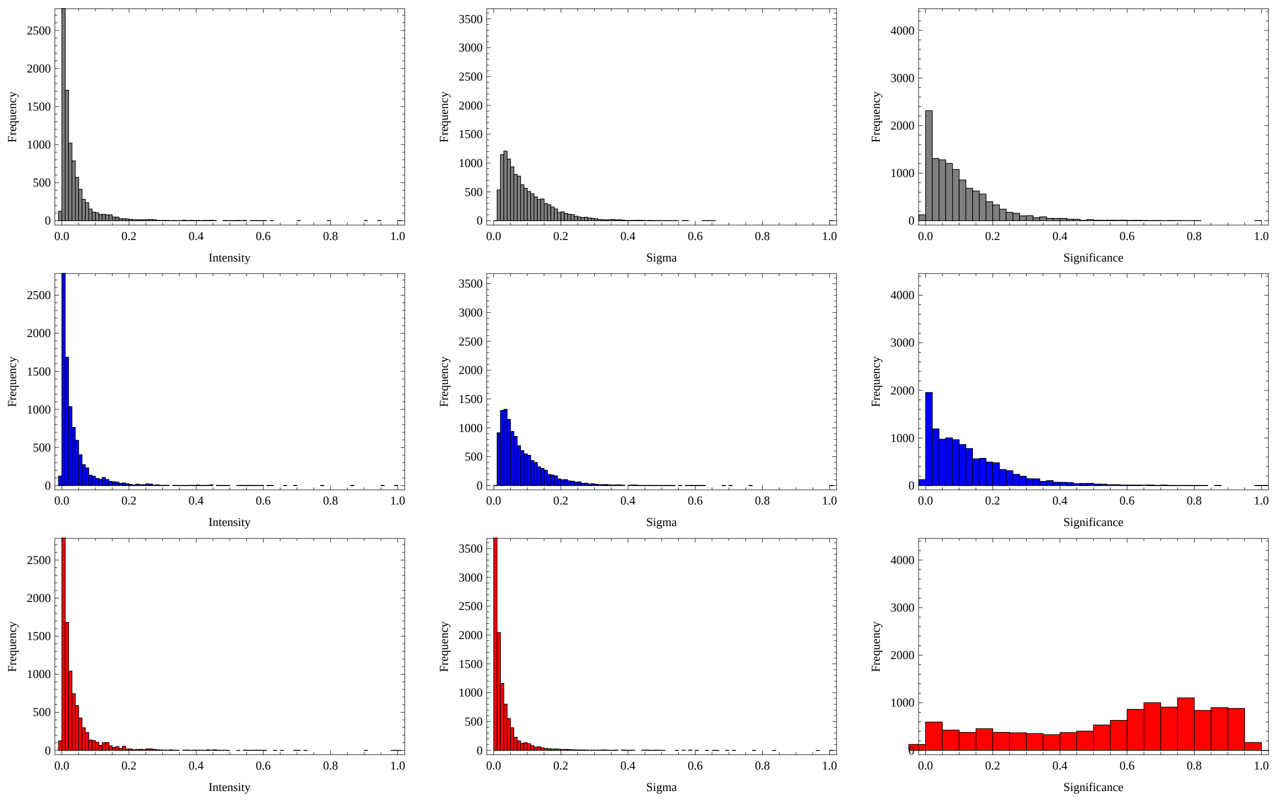

Figure 1
Frequency distributions of the intensities (1st column), the standard uncertainties (2nd column) and the significances (3rd column) for raw data (1st line), Bragg data after processing with SORTAV (2nd line) and after processing with SADABS (3rd line) for data set 2. The distributions were normalized to their maximum value.

IUCr macros version 2.1β1: 2007/05/15

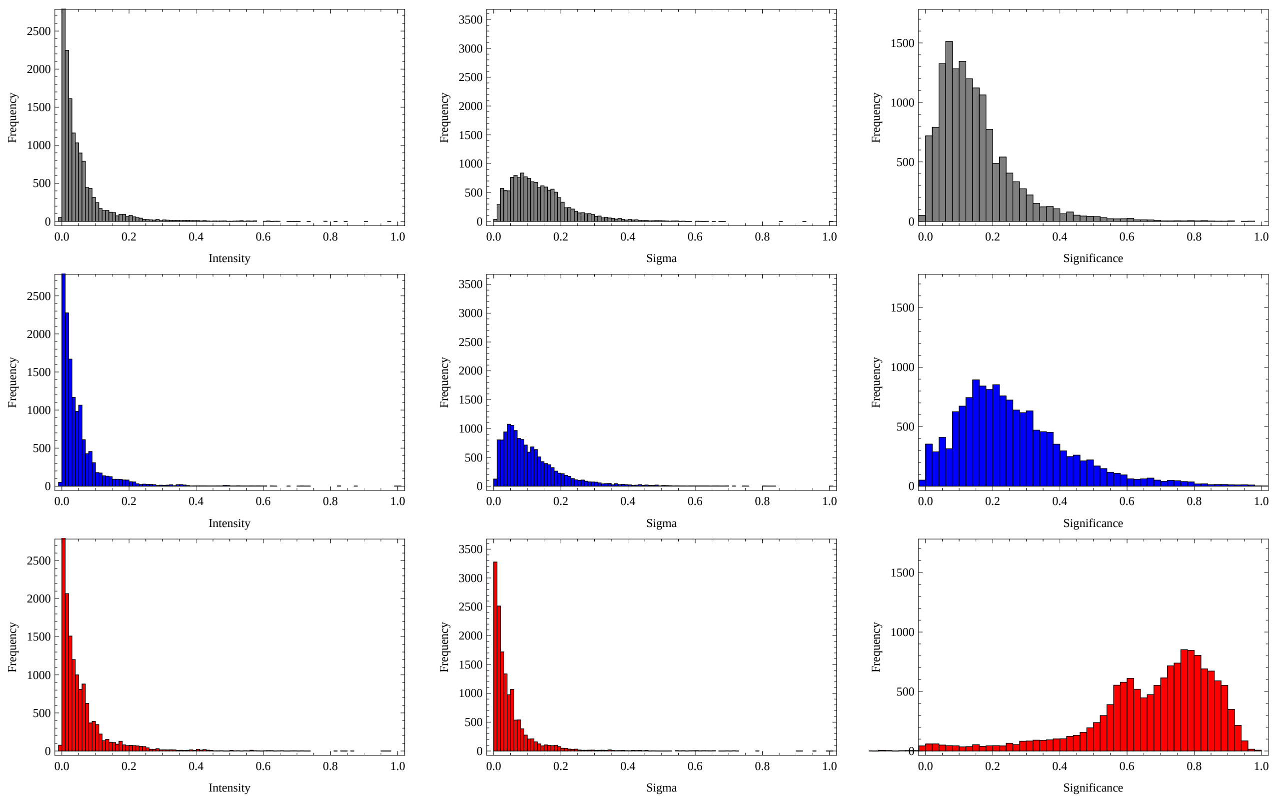

Figure 2
Frequency distributions of the intensities (1st column), the standard uncertainties (2nd column) and the significances (3rd column) for raw data (1st line), Bragg data after processing with SORTAV (2nd line) and after processing with SADABS (3rd line) for data set 3. The distributions were normalized to their maximum value.

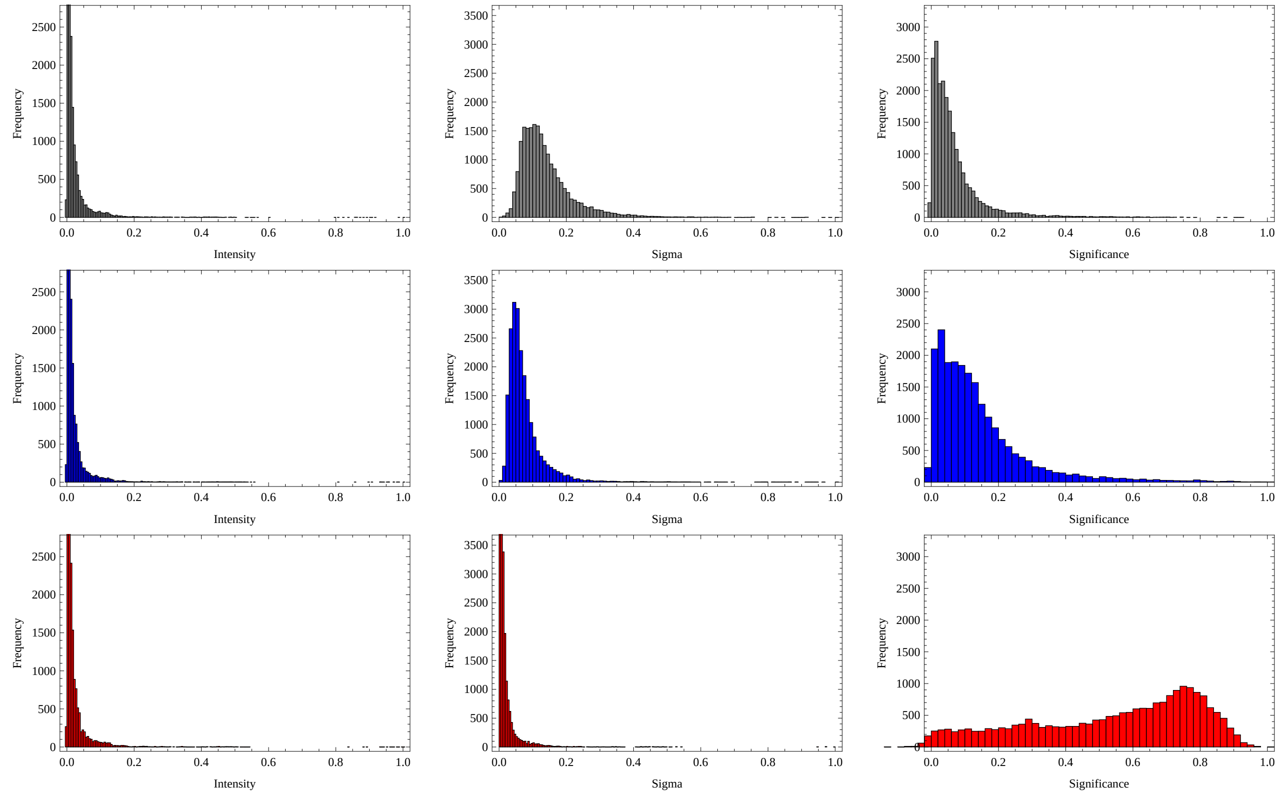

Figure 3
Frequency distributions of the intensities (1st column), the standard uncertainties (2nd column) and the significances (3rd column) for raw data (1st line), Bragg data after processing with SORTAV (2nd line) and after processing with SADABS (3rd line) for data set 4. The distributions were normalized to their maximum value.

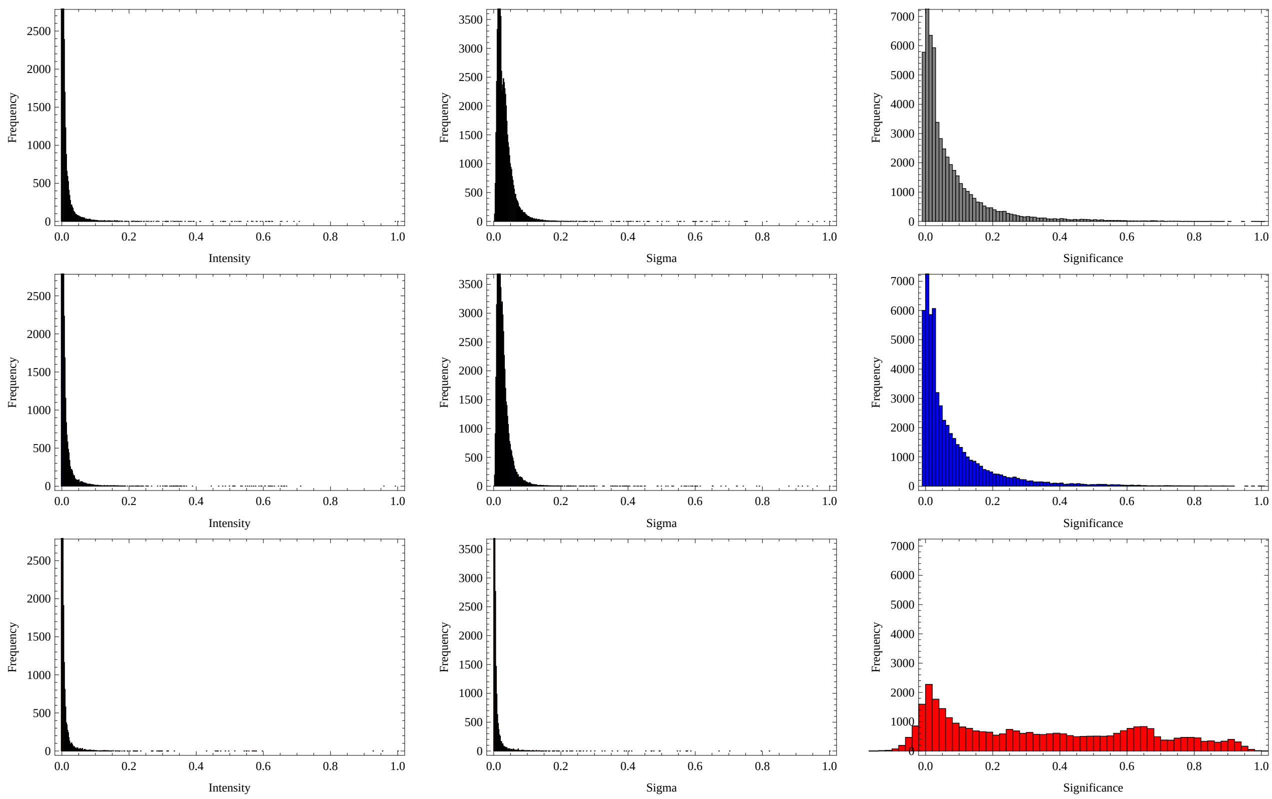

Figure 4
Frequency distributions of the intensities (1st column), the standard uncertainties (2nd column) and the significances (3rd column) for raw data (1st line), Bragg data after processing with SORTAV (2nd line) and after processing with SADABS (3rd line) for data set 5. The distributions were normalized to their maximum value.

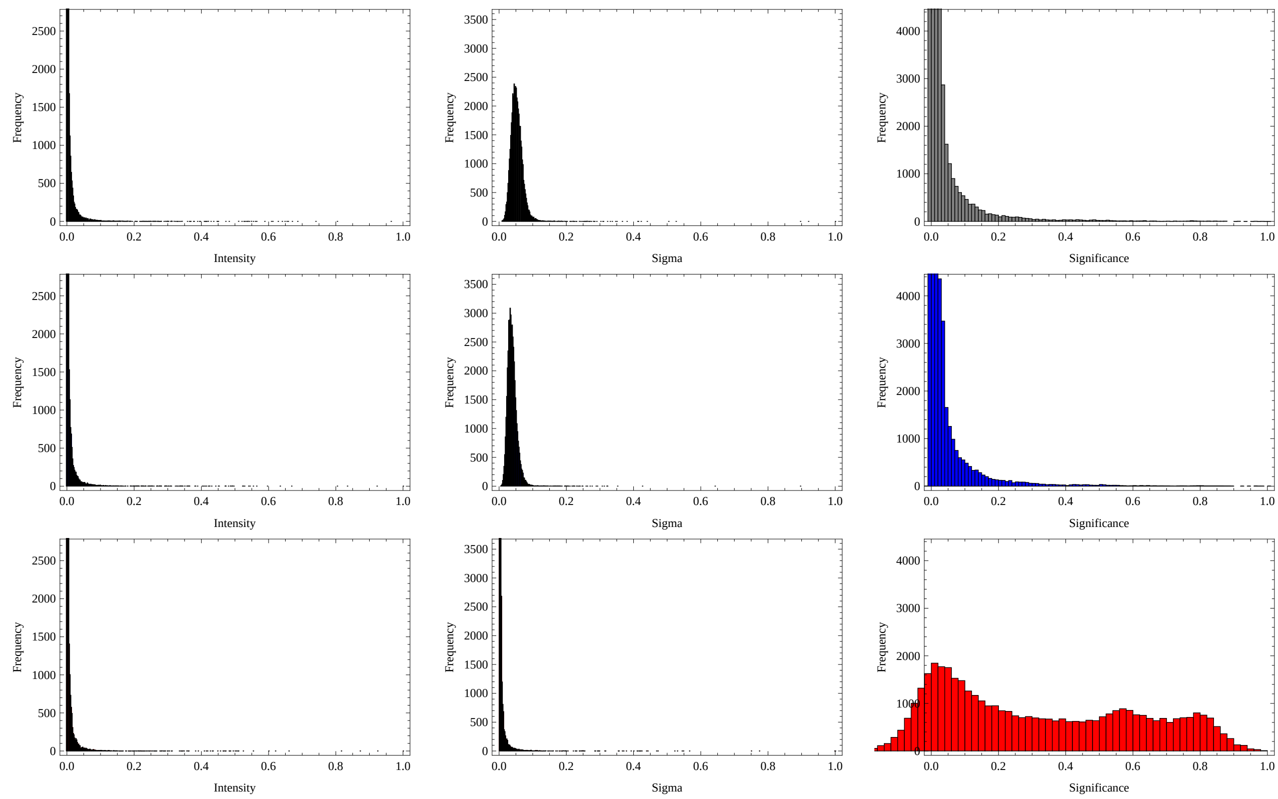

Figure 5
Frequency distributions of the intensities (1st column), the standard uncertainties (2nd column) and the significances (3rd column) for raw data (1st line), Bragg data after processing with SORTAV (2nd line) and after processing with SADABS (3rd line) for data set 6. The distributions were normalized to their maximum value.

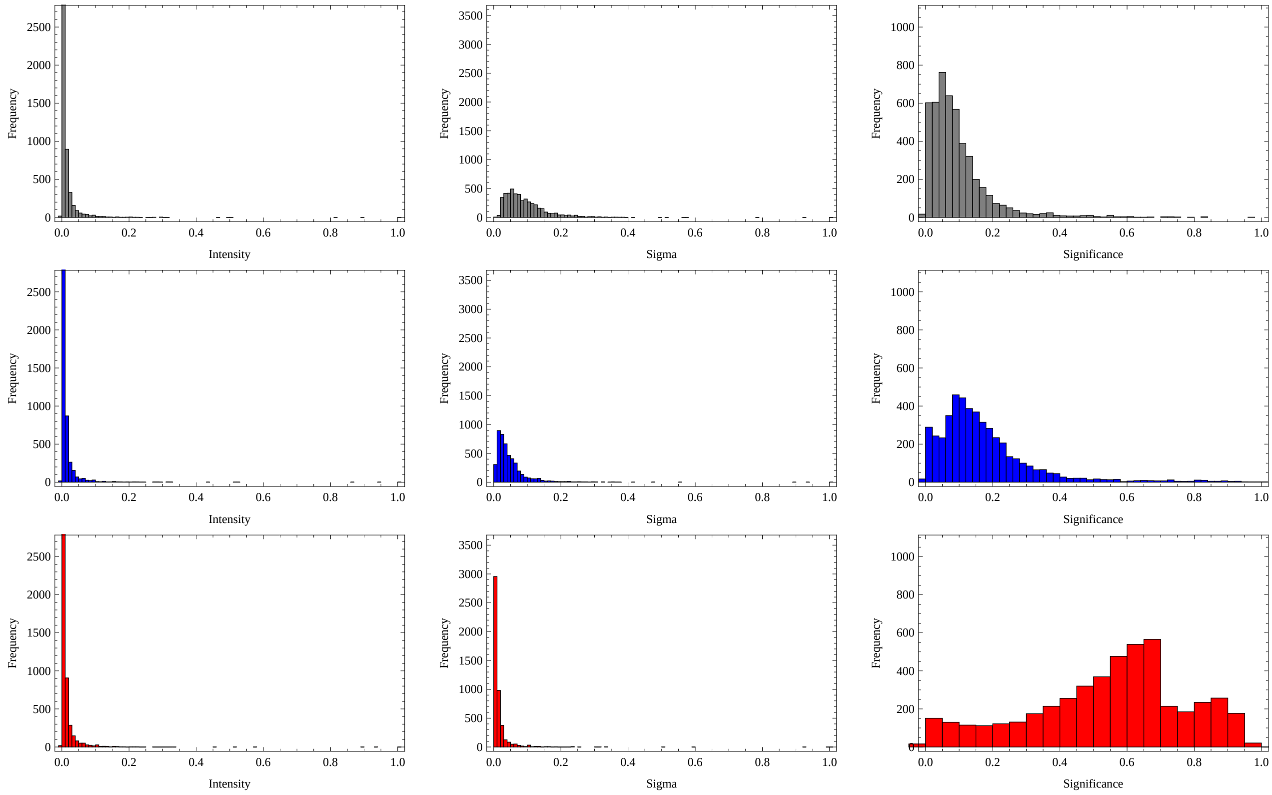

Figure 6
Frequency distributions of the intensities (1st column), the standard uncertainties (2nd column) and the significances (3rd column) for raw data (1st line), Bragg data after processing with SORTAV (2nd line) and after processing with SADABS (3rd line) for data set 7. The distributions were normalized to their maximum value.

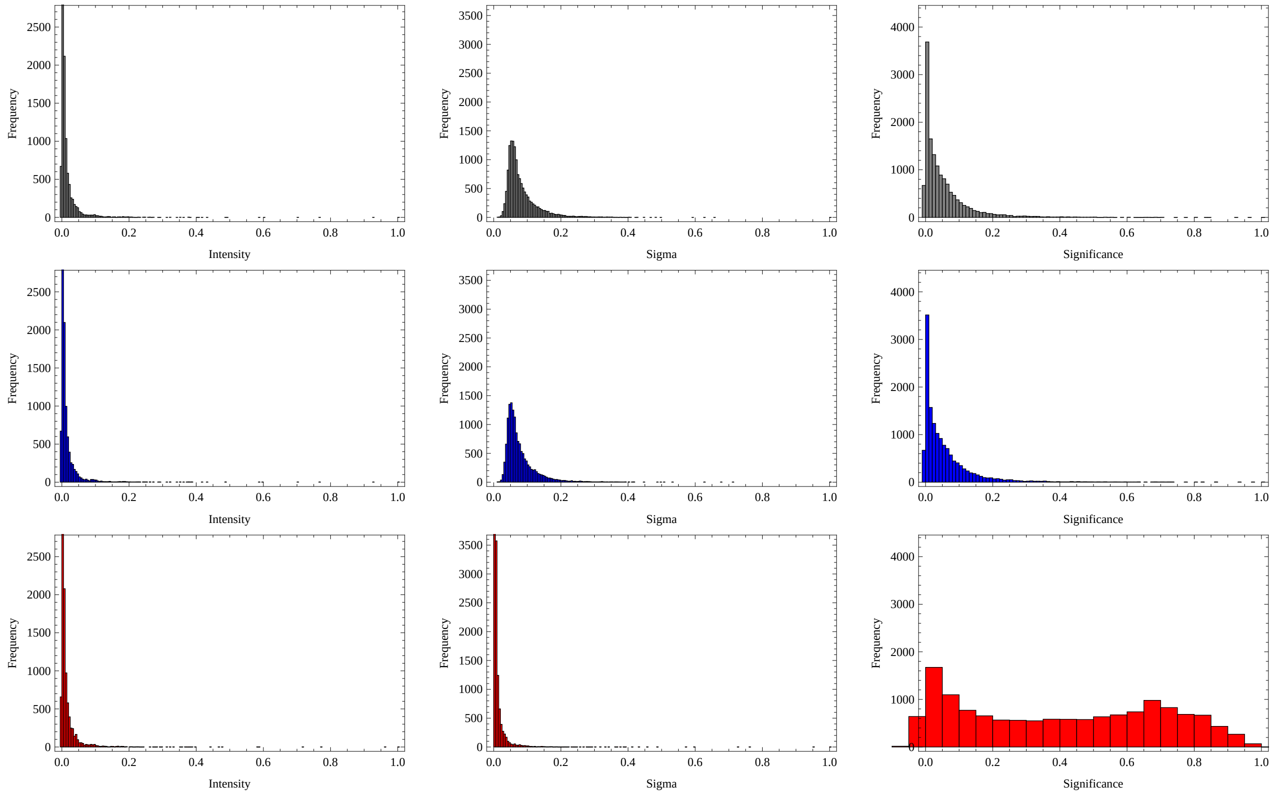

Figure 7
Frequency distributions of the intensities (1st column), the standard uncertainties (2nd column) and the significances (3rd column) for raw data (1st line), Bragg data after processing with SORTAV (2nd line) and after processing with SADABS (3rd line) for data set 8. The distributions were normalized to their maximum value.

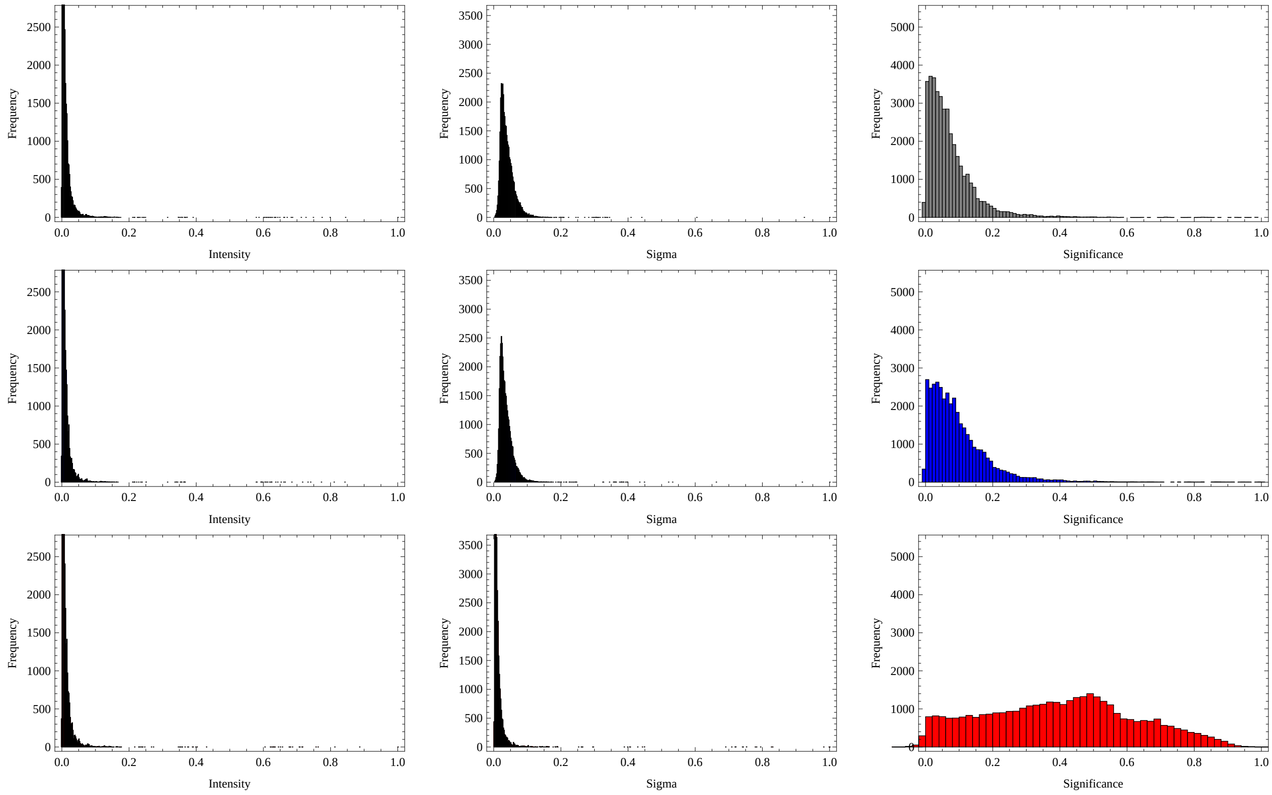

Figure 8
Frequency distributions of the intensities (1st column), the standard uncertainties (2nd column) and the significances (3rd column) for raw data (1st line), Bragg data after processing with SORTAV (2nd line) and after processing with SADABS (3rd line) for data set 9. The distributions were normalized to their maximum value.

2. Poisson statistics

If a set of discrete random numbers $\{X\}$ is distributed according to the probability distribution

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (1)$$

with $\lambda \geq 0$ it is called a Poisson distribution with the parameter λ. The mean value is given by

$$\langle X \rangle = \lambda \quad (2)$$

and the variance by

$$var(X) = \lambda. \quad (3)$$

The sum of two sets of Poisson distributed numbers $\{X\}$, $\{Y\}$ with parameters λ_1, λ_2 is again a Poisson distributed number with parameter $\lambda_3 = \lambda_1 + \lambda_2$ and variance λ_3. The mean value of the difference of two Poisson distributed numbers is given by $\lambda_1 - \lambda_2$ and their variance by $\lambda_1 + \lambda_2$.

3. Scaled variance

Let $\{X\}$ be a set of numbers with mean value $\langle X \rangle = x_{mean}$ and variance $var(X) = x_{var}$. Let $\{Y\}$ be a scaled set $\{Y\} = \{k\,X\}$.

$$\begin{aligned} var(Y) &= \frac{1}{N-1} \sum_i (Y_i - y_{mean})^2 \\ &= \frac{1}{N-1} \sum_i k^2((X_i - x_{mean})^2) \quad (4) \\ &= k^2 var(X) \end{aligned}$$

From this follows $\sigma(Y) = k\sigma(X)$, *i.e.* the standard uncertainty scales like the intensity itself and therefore re-scaling conserves the significance of an observation.

4. Deviation of equation 20 from equation 18

$$R_1 = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|} \quad (5)$$

$$= \frac{\sum \sigma(F_o)}{\sum |F_o|} \quad (6)$$

$$= \frac{\sum \frac{\sigma(I_o)}{2|F_o|}}{\sum |F_o|} \quad (7)$$

$$= \frac{N 1/2}{N\langle F_o \rangle} \quad (8)$$

$$= \frac{1}{2}\frac{1}{\langle \sqrt{I_o} \rangle} \quad (9)$$

The first line is just a definition. In the second equal sign equation (17) is used. In the next equal sign equation (19) is used. In the next equal sign it is used that $\sigma(I_o) = \sqrt{I_o}$ (from the Poisson statistics) and that $\sqrt{I_o} = F_o$. Finally each summation is replaced by N times the mean value and again $\sqrt{I_o} = F_o$ is used.