A note about McStas errorbars

A. Vickery, P. K. Willendrup and E. Knudsen

January 31, 2014

Contents

1 Introduction

In a typical simulation, one of the results consists of a count of neutrons histories ("rays") with different weights w. The sum of these weights, I_i, is an estimate of the mean number of neutrons hitting bin number i in the monitor (or detector) per second in a "real" experiment,

$$I_i = \sum_j w_{i,j} \,, \tag{1}$$

where $w_{i,j}$ is the weight of ray number j arriving in bin number i. If we increase the number of rays (by increasing the value assigned to the `--ncount` parameter), the statistical mean value of I_i will remain unchanged, but the *precision* (in the meaning of repeatability) of the simulation will be enhanced. Another result of the simulation is an estimate of the statistical errors of the intensities, i.e. a measure of the simulation *precision*. Sec. 4.2.1 in the *User and programmers guide to the neutron ray-tracing package McStas, version 2.0*[1] explains how these statistical errors are estimated within McStas, and the standard deviation σ is approximated by

$$\sigma^2(I_i) = \sum_j w_{i,j}^2 \,. \tag{2}$$

The paper[2] by K. Lefmann et al describes in detail how simulated data may be scaled to absolute measurement times.

2 Estimation of counting times and derivation of virtual data from simulations

2.1 Real vs simulated data

A dataset from a "real" experiment is often analyzed using Poisson statistics. When the number of counts n_i per bin exceeds ~20, the Normal Approximation to the Poisson distribution is adequate. Here the standard deviation σ_i for each point is

$$\sigma_i^2 = n_i \,, \tag{3}$$

and the total number of counts is

$$N = \sum_i n_i \tag{4}$$

with the standard deviation σ

$$\sigma^2 = \sum_i \sigma_i^2 = N \,. \tag{5}$$

The approximation improves as n_i increases. As is illustrated in Sec. 2.4 in general the "counts" per bin in a simulated data set rarely demonstrates this relation.

[1] http://mcstas.org/documentation/manual/mcstas-2.0-manual.pdf

[2] Journal of Neutron Research, Vol. 16, Nos. 3–4, September–December 2008, 97–111

2.2 Counting time and data quality

With a simulated data set $(I_i, I_{\mathrm{err},i})$ and a counting time t we will have $n_i = tI_i$ counts in bin number i, with the standard deviation $t\,I_{\mathrm{err},i}$. Simulations can (also) be useful to estimate a necessary counting time to get data of a specified quality. To make a reliable estimate of this counting time, first of all the simulated instrument model must be realistic. Further, the noise content of the simulated data set must be *less* than it would be in a real experiment, so we require that

$$t\,I_{\mathrm{err},i} \leq \sqrt{n_i} = \sqrt{t\,I_i} \tag{6}$$

from which we get

$$t \leq \frac{I_i}{I_{\mathrm{err},i}^2}\,. \tag{7}$$

Eq. 6 must hold for every individual bin in the data set. As is illustrated in Sec. 2.4, in general Eq. 6 is not fulfilled. Either the noise content of the simulated data bins is lower or higher than it would be in a real experiment. If Eq. 6 is fulfilled, the noise content of the simulated data is less than it is in a measured data set, and the simulated data set can then be used as the fundament at which base our estimate of a "real" data set.

2.3 Estimating the appearance of a "real" data set

Assume that we have a simulated data set (I, I_{err}) which, for a certain counting time t fulfill Eq. 6, so we have a fundament at which we can base the estimate of a "real" data set. Assume that we have $n_i = tI_i$ counts in bin number i and the standard deviation of that bin is $tI_{\mathrm{err},i}$. We then require that the noise content of a "real" data set is

$$\sigma_i = \sqrt{n_i} > tI_{\mathrm{err},i}\,, \tag{8}$$

so we must *add noise* to the simulated data in order to derive a more realistic set of data. The amount of noise which must be added to *each bin* is estimated from the simulated standard deviations. For each bin the noise could be added as a random number from a normal distribution with standard deviation $E_{+,i}$ and mean parameter 0. The magnitude of $E_{+,i}$ is

$$E_{+,i}^2 = n_i^2 - (tI_{\mathrm{err},i})^2\,. \tag{9}$$

Sec. 2.4 gives an example to illustrate the use of Eq. 9.

2.4 Example

Fig. 1**A** shows 2 data sets simulated with very different statistics, i.e. 1e7 and 1e11 neutron rays. The circles indicate the simulated intensities whereas the standard deviations are shown with dots. The relative error of the two data sets is about 0.5% and 50% respectively. As expected, the intensities simulated with the fewer neutron rays has a higher content of noise, which is also visualized by a comparison of the standard deviations belonging to the two data sets: the difference in magnitude amounts to a factor $\sqrt{10^4}$, which is consistent with the errorbars scaling with the $1/\sqrt{\texttt{--ncount}}$.

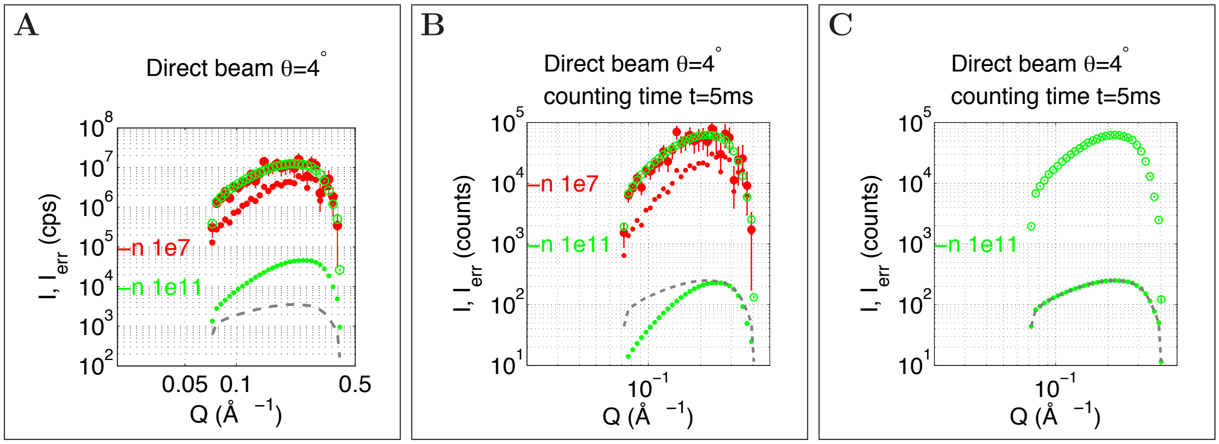

Figure 1: Data simulated using the FREIA instrument file with different statistics, 1e7 rays and 1e11 rays. The data is the direct beam incident with a grazing angle $\theta = 4°$ at a 4×4 cm^2 horizontal time-of-flight sensitive monitor at the sample position. The data is binned to constant $\delta Q/Q = 6\%$, where $Q = 4\pi \sin\theta/\lambda$ is the wavevector transfer. The wavelengths λ are calculated from the time-of-flight and grazing angle θ. θ is defined by a pair of collimating slits at 2.28 m and 0.28 m upstream the sample position.
A: The circles show the simulated intensities and the dots show the associated standard deviations for 2 simulations of the direct beam. The simulations have been conducted with identical instrument parameters but different number of rays (1e7 and 1e11 respectively). The grey dashed line show the squareroot of the intensity simulated with 1e11 rays.
B: The data from **A** but scaled with a counting time of $t = 5$ ms. The circles show the number of counts tI, the dots are the simulated standard deviations tI_{err} and the grey dashed line show the squareroot of tI.
C: The data from **B** (1e11 rays only) with added noise according to Eq. 9.

Now imagine that the simulated intensities shown in Fig. 1**A** were data from a real experiment with counting time 1 second. Then the y-axis label would be "counts" and the standard deviations would lie at the dashed grey line which indicate the squareroot of the green 'counts'. Due to the very high intensity of the direct beam, the relative error of a 1 s measurement would be extremely low, of the order of $\sqrt{10^7}/10^7 \approx 0.03\%$. Even when simulating with 10^{11} neutron rays, the standard deviations of the simulated data points are about one order of magnitude above the grey dashed line. This indicates that the content of noise in the actual simulation is much higher than it would be in a real measurement with counting time $t = 1$ s. Therefore we conclude that the quality of the simulated data set is not good enough to estimate the appearance of a real data set with counting times of the order of seconds or longer. Fig. 1**B** shows the same simulated data sets, but now both of them have been scaled by a counting time of $t = 5$ ms. That is the circles now indicate the estimated number of counts $t\,I$ and the dots indicate the standard deviations $t\,I_{\mathrm{err}}$. The grey dashed line shows the squareroot of the number of counts $\sqrt{t\,I}$, that is the magnitude of the standard deviation if it was a real measurement. With this short counting time $t = 5$ ms, the noise content of the good statistics data set is less than or equal to $\sqrt{t\,I}$. That is the bins with the highest standard deviations corresponds to a real measurement with counting time $t = 5$ ms. After scaling with a counting time of 5 ms the majority of the simulated data points now contains *less* noise than a real meaurement with the same counting time.

If the purpose of the simulation is to get a realistic picture of the appearance of a "real" 5 ms measurement, we must add an amount of noise to the simulated data. The amount of noise is derived from the simulated standard deviation and is given by Eq. 9. Fig. 1**C** shows the estimate of the "real" 5 ms measurement and the standard deviations of the points which now coincide with the dashed grey line indicating $\sqrt{t\,I}$.