Stopping and Looking I: On Error Bars in Time Series,

Guillaume Belanger, ESAC

In X-ray Astronomy, the error bar or uncertainty associated with the value of a bin in a time series is usually calculated using the square root of the number of events in that bin. This standard practice is erroneous both on theoretical and on practical grounds. On theoretical grounds, it fails to conform to the fact that the uncertainty on a measurement depends upon the accuracy with which the measurement is made, not on the value measured. On practical grounds, it fails to give the correct value of the mean count rate when the error bars are taken into account. Therefore, this practice should be abandoned in favour of an alternative where the error bar is a more appropriate representation of the uncertainty on the measurement, which does not depend upon the number of events in the given bin.

1 CONTEXT

In high-energy astronomy, each detected photon's arrival time is recorded, and its energy is reconstructed. These quantities for many such photons yields an event list.

A collection of ordered photon arrival times, binned to estimate some average property over a given time scale, or not, can be termed a time series. We thus make the distinction between binned and un-binned time series.

It is customary to construct binned time series (light curves) from event lists, and express the result as count rates as a function of time. It is also highly customary to calculate the uncertainty (error bar) on each bin's intensity by taking the square root of the number of events in it ($\sqrt{n}$). If the intensity is presented as a rate, then $\sqrt{n}$ is divided by the bin time.

We argue that this practice is should be abandoned on theoretical grounds, and show that it always leads to an under-estimate of the mean count rate—the most fundamental of statistical quantities.

2 UNCERTAINTY ON A MEASUREMENT

For any measurement, that of the length of a pencil, for example, the uncertainty can be, and usually is, established before the actual measurement. This accuracy obviously does not depend on the length of the pencil. If we were to use the same ruler to measure a much smaller object, the uncertainty on our measurement would remain the same, because it depends on the instrument, not on the value of the measurement.

This reasoning applies to all measurements, no matter what instrument is used, no matter what is measured, and no matter what is the value of the measurement. Why then do we assign to the estimate of the count rate in each bin an uncertainty derived from the number of events in the bin? This inevitably—and wrongly—gives more statistical weight to bins with fewer events.

3 WHAT ARE WE MEASURING?

Counting the number of events that have occurred in a given time interval, we are working with Poisson statistics. The Poisson distribution is characterised by a single parameter: the variance of the Poisson distribution with a mean of ν is ν, and its standard deviation is thus $\sqrt{\nu}$. But this bears no direct relation to the estimation of an uncertainty on a measured count rate over a time interval, nor to the underlying process at the source.

It is absolutely essential to distinguish the nature of the process of photon detection—a Poisson process, from that giving rise to the photons at the source—not a Poisson process in general.

A time-domain observation of an astrophysical source can be considered as measuring a unique observable—the mean count rate. The event list as a whole is thus used to estimate the average number of events per unit time by dividing the total number of events by the duration of the observation.

Such an observation can also be considered a means to measure a different random variable—the instantaneous count rate. For this, it is necessary to bin the events, but choosing a bin time is a delicate matter. This fact is appreciated by considering the two limiting cases. The first in which each bin contains at most a single event. Here, it is impossible to make a statement about the instantaneous count rate other than that it varies between 0 and 1 event per bin time interval. The second, in which all events are contained within a single bin, in which case we have a single estimate, the mean count rate, as in the previous case.

The consideration of the bin width is indeed very important, for it immediately leads to the notion of a measure of confidence (uncertainty) in the estimate of the instantaneous count rate.

4 WITHOUT ERROR BARS

The fact is that an event is either detected or not. Thus, the number of events in a bin, whether it is 0, 1, 25 or 473, is simply the number of events detected during the given time interval. This number does not have an uncertainty.

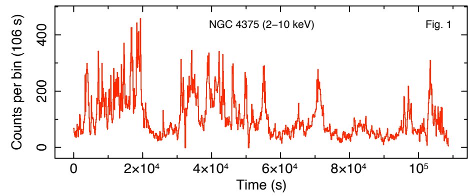
Hence, we work directly with the un-binned time series as much as possible, and otherwise, with the number of events per bin, without error bars. Subtracting a 'background' time series inevitably causes the loss of valuable information about the distribution in time of the detected events.

The probability of observing a number of events during a time interval, depends entirely on the expected number of events, which in turn depends on the underlying probability distribution. For a Poisson process, the probability of detecting n events while expecting ν is given by the integral of $(\nu^n/n!)\, e^{-\nu}$ from n to ∞. But what if it is not a Poisson process at the source?

5 UNJUSTIFIABLE ASSUMPTIONS

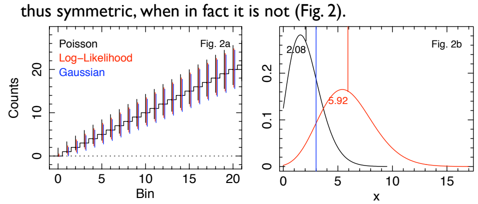
Using $\sqrt{n}$ for the error bar relies on unjustifiable assumptions:

• That the process giving rise to the photons we detect is a Poisson process, and thus that the number of events in a bin is independent from that of every other bin, even though it only very rarely can reasonably be considered as such, especially in astrophysics where we observe complex, physically extended systems, consisting of many interacting components in which everything that happens is related to everything else, even if how is not known (Fig. 1).



• That we can treat the number of events in a bin as if it were a Poisson distribution whose mean is given by the number of events, and therefore, whose standard deviation is the square root of that. But in fact, the events in a bin are simply a number of events, not a distribution.

• That the number of events is large enough for the Poisson distribution associated with a bin to be close to Gaussian and thus symmetric, when in fact it is not (Fig. 2).



If we were to consider the time interval that covers each bin as an independent observation of the source—which it is not, and if the process could be considered to be a Poisson process—which in practice it almost never is, then we could consider each bin as independent measurements of the *same* mean count rate, estimated over the duration of the bin. We could then draw the distribution of these 'independent' estimates of the count rate. This, we assume, will give us a single well-defined Poisson distribution. In general, however, this is not be the case (Fig. 3).

6 ZERO EVENTS

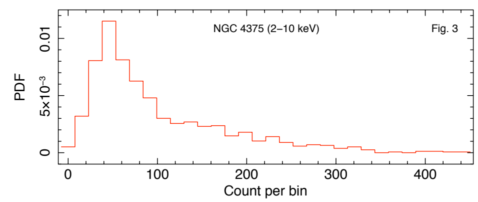
There is no better example to illustrate the inadequacy of using $\sqrt{n}$ than the case of detecting zero events during the bin time interval. It is simply non-sensical to ascribe zero uncertainty to the estimate of the count rate derived from a bin that spans a period during which no events were detected, because this effectively gives it *infinite statistical weight*. And yet, this is done. Alternatively, noticing the problem of infinite statistical weight, the zero-event bins are simply ignored in the analysis, yielding unreliable results.

In the case of zero detected events, even under the assumptions discussed above, the number of events is clearly not large enough to assume that the distribution is symmetric. How then do we calculate the error bars when the number of events is nil?

One way to address this problem is to ask the question: what is the mean of the Poisson distribution for which zero is at the edge of the region that contains to 68/2% of the distribution from the mean? It turns out that the answer is 2 (Fig. 2a).

We could also ask: what is the 1σ bound of the maximum likelihood function for the Poisson distribution in the case of zero counts per seconds? In this case, the answer is 0.5 (Fig. 2a).

Although each of these methods offers a solution to the problems associated with the use of $\sqrt{n}$ in the case of zero-event bins, because they are neither given infinite statistical weight nor ignored. However, since they rely on the other assumptions, they both fail to give what we consider to be a representative measure of the uncertainty on the estimates of instantaneous count rate: both methods yield error bars of variables sizes, proportional to the number of events in the bin.



7 WITH ERROR BARS

In our view, the error bars on our estimates of the instantaneous count rate must fulfill three conditions: 1) they must give a realistic measure of the uncertainty on these estimates, 2) they cannot depend on the value of an individual estimate, and 3) they must be simple to calculate, and if possible, allow us to use the popular weighted mean formula.

The simplest way to fulfill the second condition, is to assign error bars of equal size for all bins (as is standard practice in ground-based infrared astronomy when using a reference star). Conveniently, this ensures that the weighted mean will always give exactly the sample mean.

Assuming systematic uncertainties can be treated independently, our error bar is a measure of the statistical uncertainty. Since it cannot depend on the value of individual bins, in the absence of a means to derive the uncertainty independently of the data, it must be derived from the data set as a whole. In addition, it cannot depend on the detailed shape of the time series since the instrument does not know if the source weak or strong (but not piled-up), constant or variable. Therefore, we propose to use as the expected statistical uncertainty on our instantaneous count rates, the square root of the average number of events per bin $\sqrt{\langle n \rangle}$ (divided by the bin time).

8 THE WEIGHTED MEAN

The weighted mean and the uncertainty on it for a set of data, x_i, each with an associated weight, w_i are defined as:

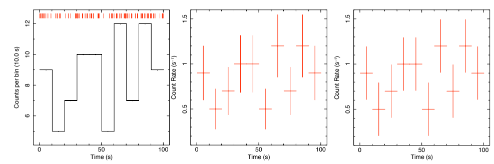
$$\mu_{\rm w} = \frac{\sum w_i \cdot x_i}{\sum w_i} \quad \text{and} \quad \sigma_{\rm w} = \frac{1}{\sqrt{\sum w_i}}$$

If the uncertainty on the value of x_i is σ_i, the weight is usually defined as $w_i = 1/\sigma_i^2$ and the equations become:

$$\mu_{\rm w} = \frac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2} \quad \text{and} \quad \sigma_{\rm w} = \frac{1}{\sqrt{\sum 1/\sigma_i^2}}$$

Let us take the simple case of a Poisson process. We have an event list with N=86 arrival times detected over T=100 s (Fig.4a; red). The average count rate is thus μ=N/T=0.86 cps.

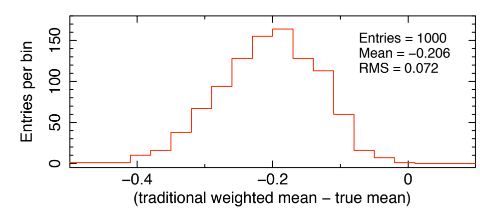
Binning in 10 s bins yields a time series that can be presented as the number of events per bin (Fig. 4a; black), or as 'instantaneous' count rates with error bars (Fig. 4b and 4c).



Computing the weighted mean, we find that using $\sqrt{n}$ error bars yields $\mu_{\rm w}$=0.78+/-0.09 cps, whereas using our method yields $\mu_{\rm w}$=0.86+/-0.09 cps. For comparison, the Maximum Likelihood Estimate of the mean within 68% is $\mu_{\rm MLE}$=0.86 +0.096 -0.089.

The negative bias appears because the uncertainty is proportional to the number of counts per bin: the bins with less counts have more weight, and therefore bias the weighted mean towards lower values.

To demonstrate this, we simulated a large number of different time series based on the same mean count rate and duration, and calculate the weighted mean using $\sqrt{n}$ error bars for each. A histogram of the difference between these so obtained weighted means and the true mean count rate, shows Gaussian distributed negative bias centred on 0.2, in this case 20% (Fig.5).



9 CONCLUDING REMARKS

Working with unbinned time series is best because all information is used. If this is not practical, (like in the case of large event lists), we can work with binned time series using the histogram of the number of events per bin without error bars. If we need to use error bars for some reason, we should never use the traditional $\sqrt{n}$ method. Rather, we should calculate the size of the error bar using the prescription described herein.