

Poisson vs. Gaussian statistics for sparse X-ray data: Application to the soft X-ray spectrometer

Shinya YAMADA,^{1,*} Magnus AXELSSON,^{1,2,*} Yoshitaka ISHISAKI,¹
Saori KONAMI,¹ Nozomi TAKEMURA,¹ Richard L. KELLEY,³
Caroline A. KILBOURNE ,³ Maurice A. LEUTENEGGER,³
F. Scott PORTER,³ Megan E. ECKART,⁴ and Andrew SZYMKOWIAK⁵

¹Department of Physics, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji, Tokyo 192-0397, Japan

²Department of Physics, KTH Royal Institute of Technology, 106 91 Stockholm, Sweden

³NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt, MD 20771, USA

⁴Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA

⁵Department of Physics, Yale University, PO Box 208120, New Haven, CT 06520-8120, USA

*E-mail: syamada@tmu.ac.jp, magnusa@astro.su.se

Received 2015 May 12; Accepted 2019 April 11

Abstract

Reliable results when fitting X-ray data require proper consideration of the statistics involved. We probe the impact of Gaussian versus Poisson statistics at low count levels using both the standard χ^2 method and maximum likelihood based on Poisson (C) statistics. The difference is studied and quantified through simulated spectra with known properties. We then test the results through analysis of Mn $K\alpha$ calibration data taken with the flight spare microcalorimeter for the Hitomi soft X-ray spectrometer. Through comparison with simulations, our results show that the χ^2 method tends to give overly optimistic estimates of the detector energy resolution, in particular when there are few counts. Given an energy resolution of ~ 5 eV and a line with about 100 photons, the line width becomes $\sim 10\%$ lower in the χ^2 method than in Poisson statistics. This is a consequence of the uncertainties being dominated by counting statistics, and therefore highlights the need to choose the appropriate fit statistic.

Key words: instrumentation: detectors — methods: data analysis — methods: statistical

1 Introduction

Finding a spectral model to describe measured data is a nearly ubiquitous problem in X-ray astronomy. Using a given theoretical model, comparison between measured and theoretical spectra is made by performing a spectral fit to the data. The (known) theoretical spectrum of the model is convolved with the instrument response, and the parameters adjusted to achieve a good match to the observed data. However, in order for the test to provide accurate results it

is vital to understand the underlying statistical distribution of the data. Photon counts follow Poisson statistics, yet the fitting techniques are often based on least squares, which is derived from Gaussian statistics. Using such techniques is known to cause potential bias in the recovered parameter values, both in applications within X-ray astrophysics—such as the comparison of χ^2 and C statistics in Nousek and Shue (1989), or the discussion on maximum likelihood methods in Leccardi and Molendi (2007)—and in

other fields (e.g., Laurence & Chromy 2010). Further difficulties are encountered when comparing and choosing between different models (see, for example, Protassov et al. 2002).

This paper describes a comparison between the commonly used χ^2 statistic (based on Gaussian statistics) and the C statistic (based on Poisson statistics; Cash 1979) in the regime of sparse X-ray data (i.e. low photon statistics). In particular, we exemplify the effect on data used to evaluate instrument performance, taken from operational tests made to verify the performance of the pulse-shape processor (PSP) for the soft X-ray spectrometer (SXS) detector on the Hitomi mission. This provides a cleaner case than typical applications as the input spectrum is known.

The remainder of this section is dedicated to a presentation of the SXS detector. In section 2 we describe the results when using simulated data. A comparison with real data is made in section 3, and the results discussed in terms of the expectations from a statistical point of view in section 4. Finally, the main conclusions are summarized in section 5.

1.1 The SXS instrument

Hitomi is the sixth Japanese X-ray satellite, and was launched in 2016 (Takahashi et al. 2010). The development was led by the Japan Aerospace Exploration Agency (JAXA) in collaboration with the National Aeronautics and Space Administration (NASA), the European Space Agency (ESA), and international institutes. The onboard X-ray microcalorimeter, a soft X-ray spectrometer, is one of two soft X-ray telescope focal plane instruments (Kelley et al. 2016). SXS uses a 6×6 microcalorimeter array to measure the heat input from incident X-ray photons. The low-heat-capacity detector operates at a temperature of 50 mK and is designed to achieve an energy resolution of 7 eV (full-width half-maximum: FWHM) or better in the energy range 0.3–12.0 keV. The SXS demonstrated an energy resolution of better than 4.5 eV at 6 keV in orbit (Kilbourne et al. 2016).

The SXS comprises three main parts: the detector assembly, X-ray box (Xbox) electronics units (Porter et al. 2016), and the PSP (Ishisaki et al. 2016). When an incident X-ray is absorbed by the detector, it causes a local increase in temperature. The heat input is converted to an electronic signal which is filtered, amplified, and digitized by the Xbox front-end electronics unit. The digital waveform is then analyzed by the PSP via an optimized filter method (Szymkowiak et al. 1993; Boyce et al. 1999) to determine the energy and arrival time of the incident X-ray signal. The PSP identifies X-ray events from the continuous waveform produced by the Xbox, and evaluates the pulse heights corresponding to the energy of input X-ray photons. Full descriptions of the SXS and its detector subsystems both on

ground and in orbit are given by Mitsuda et al. (2010) and Porter et al. (2010), Tsujimoto et al. (2016), and Kilbourne et al. (2016), respectively.

Pulses identified by the PSP are graded depending on the time to neighboring pulses. The cleanest class is the high primary, where only one pulse is detected in the search window, 1024 samples of $80 \mu\text{s}$ ($= 81.92 \text{ ms}$). Two pulses in the same window are classified as either medium primary (first pulse) and secondary (second pulse) or low primary (first pulse) and secondary (second pulse), depending on the time between them. High- and medium-grade pulses are processed in the real-time operating system on the CPU board with an optimal filtering technique using a template of 1024 and 256 samples, respectively. For low-grade events a simple pulse height is calculated by subtracting a baseline value from the maximum value of the waveform. The analysis allows the incident photon energy and arrival time to be determined. Details of the analysis procedure and its in-orbit performance can be found in Shimoda et al. (2012) and Ishisaki et al. (2016), respectively.

2 Simulations

In a first simple comparison of the two fit statistics, simulations were performed using *xspec*. A simple model using a single Gaussian line was defined. Sample spectra were then generated from this model, using the pre-flight SXS response matrix for the Hitomi Collaboration. It assumes a constant energy resolution over the whole energy range, and includes quantum efficiencies of the absorber and filter transmission of the detector based on the preliminary measurements of the flight models. In order to compare the two underlying statistical distributions, Gaussian and Poisson, each dataset was fitted using both χ^2 and C statistics (see appendix 1).

Initially, the parameters of the line were kept fixed ($\mu = 5 \text{ keV}$, $\sigma_{\text{FWHM}} = 5 \text{ eV}$), while the energy binning was varied. The procedure was repeated 20 times for two different exposure times, 1000 s and 10 ks. The normalization of the Gaussian was chosen to give an average of 500 and 5000 total counts, respectively. The results are shown in figure 1.

While there is a clear correlation between the results, some trends can also be seen: there is a systematic shift toward higher values of normalization and energy resolution in the fits using C statistics. Although the difference is in most cases within errors, it is consistent for all points. The test statistic indicates that both methods are finding equally acceptable fits. The same trend is seen for both exposure times, but the shift is most evident for the shorter exposure (1000 s) and smaller bin sizes.

The fact that the shift increases for short exposure times and smaller bins indicates that it arises when the number of

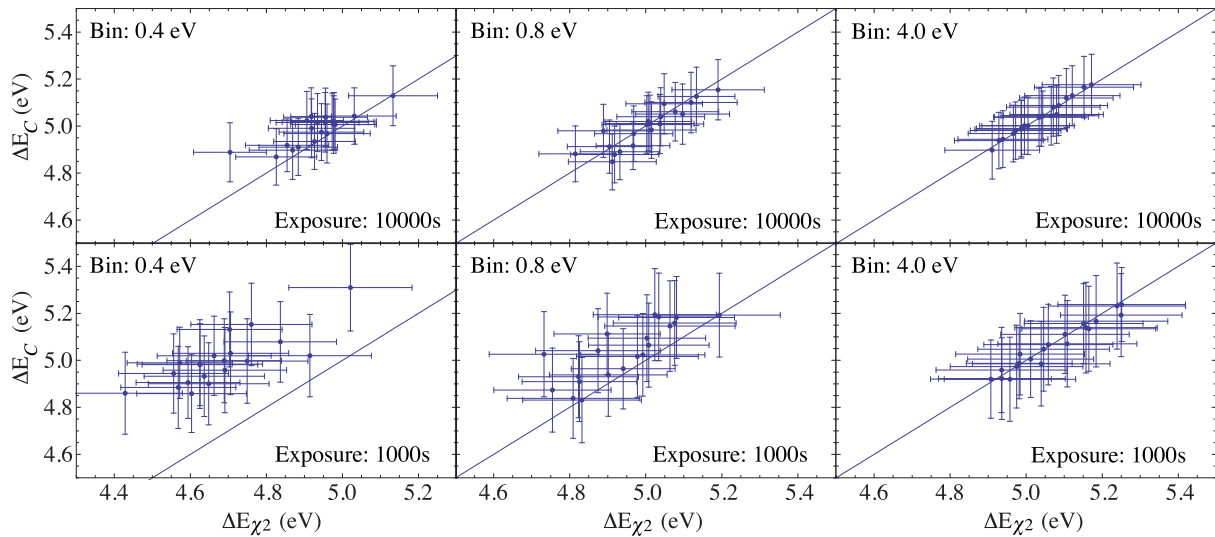


Fig. 1. Comparison of the estimated resolution ΔE from the two fitting techniques when using simulated data. The solid line indicates where the two values are equal. The upper panels have an exposure time of 10 ks, the lower ones 1 ks. From left to right the energy bin size is 0.4 eV, 0.8 eV, and 4 eV. (Color online)

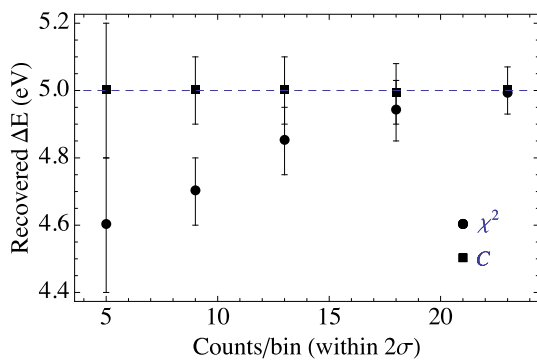


Fig. 2. Results from the simulation showing the recovered value of the resolution ΔE from the two fitting techniques as a function of average counts per bin within 2σ of the simulated peak. The dashed line indicates the true value (5.0 eV). C statistics recover the correct value even at a few counts per bin, although the errors grow larger. On the contrary, χ^2 statistics give an incorrect resolution when there are low counts.

counts per bin is low. As the “true” value of the line width is known for the simulations, we can use the simulations to test which method gives the correct result. Focusing on the region around the peak, we now repeat the simulations, gradually increasing the integration time. Figure 2 shows the recovered value of ΔE for the two statistics as a function of the number of average counts per bin within two standard deviations from the peak.

As the number of counts decreases, the recovered value of ΔE decreases when using χ^2 statistics, even though no change has been made to the model spectrum. Conversely, the value recovered by C statistics remains constant, and recovers the correct value even at low counts per bin, albeit with large errors.

Table 1. Values of $\Delta E_C / \Delta E_{\chi^2}$ for different combinations of line width and total counts.*

Line width [†]	Counts (4900–5100 eV)					
	130	225	400	700	1300	2250
5 eV	1.11	1.08	1.06	1.04	1.01	1.00
20 eV	— [‡]	1.10	1.09	1.06	1.02	1.01
50 eV	— [‡]	— [‡]	1.15	1.12	1.05	1.02

*The fits are made only to the region 4900–5100 eV, and the counts are for this energy range only. The line center is in all cases fixed to 5000 eV.

[†]The FWHM width.

[‡]No reliable fit could be found.

2.1 Using the SXS instrument response

We also performed the simulations to more precisely determine the effect on SXS data. In addition to a Gaussian line at 5000 eV, these simulations included a constant (with Gaussian errors) background component and were limited to the energy range 4900–5100 eV. We also used the same pre-flight SXS response matrix to generate and fit the simulated data. Each spectrum was fitted with both χ^2 and *pgstat* (since the background was Gaussian) statistics. The procedure was repeated for three different values of line width and varying line strength, with the background kept constant at 100 counts. Table 1 shows the values of $\Delta E_C / \Delta E_{\chi^2}$, and thus gives a measure of the “error” introduced by using χ^2 statistics.

Although we tested different ratios between the strength of the line and the background, we found that it is the total number of counts (i.e., line + background) which determines how different the results from the two methods will be. This is expected if the difference is purely due to statistical

reasons. Clearly, there is a limit in that the line must be strong enough to be detected, as can be seen in the left columns of table 1: at low counts no reliable fit can be found for the broad lines. In general, a narrow line will give more reliable results for a given number of counts, as the counts in the line will be more concentrated (leading to higher numbers of counts in the energy bins covered by the line). As a rule of thumb, given an energy resolution of ~ 5 eV and a line with about 100 photons, the line width becomes $\sim 10\%$ lower with the χ^2 method than with Poisson statistics.

3 Practical application: PSP data

The excellent energy resolution of the SXS was one of the most important requirements for the Hitomi mission. In order to verify and calibrate the resolution, tests were performed by irradiating the detector with X-rays from a calibration source with known energy and spectrum. By comparing the measured spectrum to the natural line shape of the source, the energy resolution can be determined. Using such calibration data, we can perform the same comparison as for the simulations.

3.1 Data analysis

The data were gathered during a PSP function test performed at NASA/GSFC on 2014 July 31. The test was performed on the flight spare calorimeter sensor together with an engineering model PSP. A PSP-generated template was used, and the raw packet telemetry was converted into the first fits file by `mksxsfff` version 2014-07-23.¹ A total of 48 ks of data were collected. The data were analyzed using two integration times, 1 ks and 10 ks, and the standard bin size used was 0.8 eV. Long-term variations in instrument gain have been corrected for using a linear gain stretch and a reference fiducial line.

During the irradiation tests, two sources of X-rays were used. The main detector was illuminated using an X-ray generator with a Mn target, exciting the $K\alpha$ transitions. However, the calibration pixel was illuminated by a ^{55}Fe source, which decays to an excited state of ^{55}Mn , thus producing $K\alpha$ emission.

The data analyzed come from both the calibration pixel (number 12) and a standard pixel (number 23). The total count rates of pixels 12 and 23 in an energy range of 5.8–6 keV were $3.79 \text{ counts s}^{-1}$ and $0.95 \text{ counts s}^{-1}$, respectively. When only high primary events were selected, they became $2.1 \text{ counts s}^{-1}$ and $0.82 \text{ counts s}^{-1}$, respectively. Throughout the paper, only high primary events are used. This is the cleanest class, and thus provides an estimate for

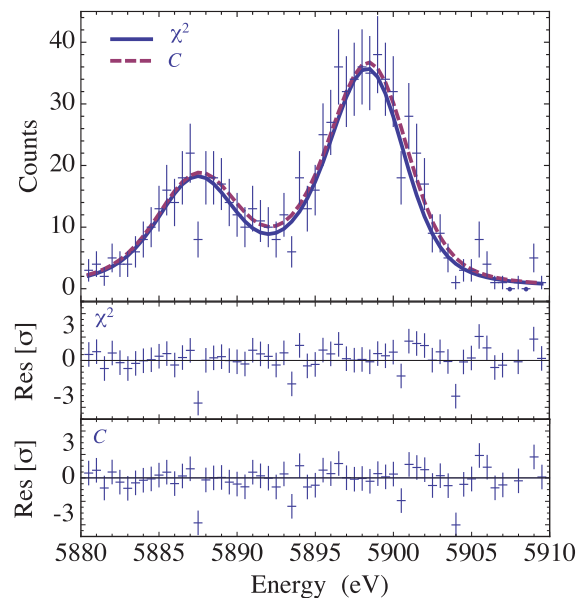


Fig. 3. Example fit of the same spectrum (pixel 12, 1000 s) using χ^2 and C statistics. Only the region 5880–5910 eV is considered in the fit. The two methods give very similar, and equally acceptable, fits, as shown by the residuals (lower panels). (Color online)

the greatest energy resolution achieved with the SXS. The events were further cleaned based on rise time and slope after the peak (e.g., removing events where a rapid jump after the peak may indicate an unresolved second hit immediately after the first), though such events are rare in high primary events for Mn K lines.

The model used was a combination of Voigt functions, known to describe the Mn $K\alpha$ complex (Hölzer et al. 1997). The parameters of the individual peaks were fixed, and are shown in appendix 2. The theoretical spectrum was then “smeared” by convolution with the instrument response containing three free parameters: the normalization (N), gain (g), and energy resolution (ΔE). The fit was performed in `root`² over the energy range 5880–5910 eV, covering the two main peaks of the source spectrum.

3.2 Fit results

Examples of fit results from the two methods are shown in figure 3. As can be seen from the figure, the two statistics give similar fit results and fit quality. The empirically derived model does not fully capture the shape of the spectrum; in particular, there were signs of an excess above 5910 eV in the data. Rather than include additional components in the model, the fit was limited to the range 5880–5910 eV. While this does leave some residuals in the wings, it is the spectral peaks that dominate the determination of

¹ This is a prototype version of the tool, developed by Y. Ishisaki.

² (<https://root.cern.ch/>).

Table 2. Results of fitting the calibration source spectrum for each tested pixel using different integration times.

Integration time	Pixel	Normalization* (N)		Energy resolution (ΔE) [eV]		Gain (g)	
		χ^2	C-stat	χ^2	C-stat	χ^2	C-stat
1000 s	12	1647 ± 17	1682 ± 17	4.66 ± 0.096	4.74 ± 0.097	$1 \pm 7 \times 10^{-6}$	$1 \pm 7 \times 10^{-6}$
	23	654 ± 11	680 ± 11	4.94 ± 0.16	5.15 ± 0.16	$1 \pm 2 \times 10^{-5}$	$1 \pm 2 \times 10^{-5}$
10000 s	12	16380 ± 53	16447 ± 53	4.92 ± 0.031	4.96 ± 0.032	$1 \pm 2 \times 10^{-6}$	$1 \pm 2 \times 10^{-6}$
	23	6372 ± 33	6407 ± 33	4.98 ± 0.05	5.01 ± 0.05	$1 \pm 4 \times 10^{-6}$	$1 \pm 4 \times 10^{-6}$

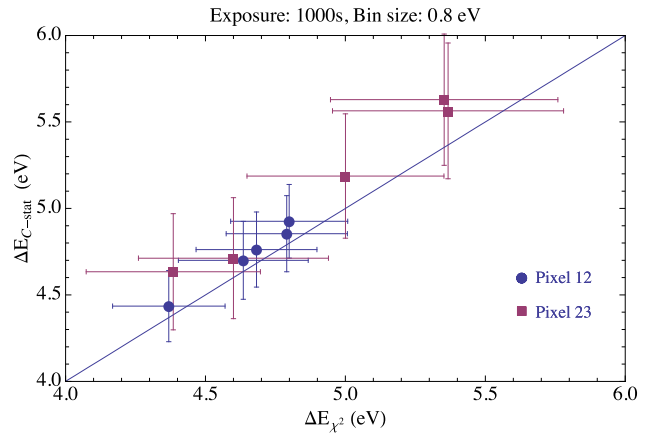
*Normalization is in units of photons $\text{cm}^{-2} \text{s}^{-1}$.

the energy resolution. Limiting the energy range improves the fit quality, but the results presented below are not dependent on the energy range considered.

Average parameter values from five fits using integration times of 1 ks and 10 ks are shown in table 2. The table displays the results from both pixels. From the values of the normalization, which is proportional to the total number of detected counts, it can be seen that pixel 12 received a larger number of counts than pixel 23. Despite this, the two pixels give comparable results for the energy resolution. The gain is in all cases very close to 1, indicating that the gain correction has been successful. The test data thus show that the PSP is able to provide an energy resolution of ≤ 5 eV at energies around 6 keV. This result is found with both types of statistic and is consistent with expectations from the FM sensor test (Porter et al. 2010).

Table 2 also shows that the results are similar when using the two statistics. However, closer inspection reveals that the discrepancies observed in the simulations also appear in the data. It can first be seen that the fits return values with similar variations as those of the simulations. This indicates that these variations are purely statistical, as there is no extra source of noise in the simulations. The real data further show the same trend as the simulations (the recovered energy resolution is systematically lower when using χ^2 statistics). Figure 4 shows the comparison for the energy resolution using all fits made with 1000 s integration time; just like in the simulation there is a systematic shift between the results. The figure additionally shows that the estimate of the energy resolution also varies within a given pixel, roughly spanning the range 4.5–5.5 eV. The reason is statistical fluctuations and non-stationary noise, which would not be the same effect among the pixels since the physical, electrical, and thermal properties are not completely identical.

In the simulations the shift was reduced for longer integration times (comparing the upper and lower rows in figure 1), and clearly affected by changes to the bin size (from left to right in figure 1). To search for this effect in the data, the bin size was changed to 0.4 eV and 1.6 eV for the integration time of 1 ks and the median ratio between

**Fig. 4.** Comparison of the estimated resolution ΔE from the two fitting techniques for pixels 12 (circles) and 23 (squares). The solid line indicates where the two values are equal. There is a clear systematic shift between the two methods. (Color online)**Table 3.** Comparison between recovered ΔE for C and χ^2 statistics.*

Simulation		Data (pixel 12)		Data (pixel 23)	
Bins	Ratio	Bins	Ratio	Bins	Ratio
0.4 eV	1.06	0.4 eV	1.05	0.4 eV	1.07
0.8 eV	1.02	0.8 eV	1.02	0.8 eV	1.04
4 eV	1.00	1.6 eV†	1.02	1.6 eV†	1.02

*The ratio is calculated as $\Delta E_C / \Delta E_{\chi^2}$. The data are taken from both pixels and the integration time is 1000 s.

†The largest bin size that could be used for the real data; for wider bins the fit does not converge.

the two techniques ($\Delta E_C / \Delta E_{\chi^2}$) measured in each case. The results are shown in table 3 together with the corresponding results from the simulations. The same trend is clear in both cases: the difference between the two methods decreases with increasing bin size. For the data the bins could not be increased above 1.6 eV, as the fits did not converge for larger bins. Presumably this is due to the more complex shape of the Mn K α spectrum compared to the simple peak used for the simulations.

4 Discussion

The systematic difference seen between the two fitting techniques is a clear indication that the underlying statistics play an important role. Theoretically the difference should be greatest when there are few counts, and become smaller as the number of counts increases. The numbers of counts are lowest for short integration times, and small energy bins. Indeed, it is precisely for these conditions that we see the greatest differences, both in the simulations and the data. The fact that the two methods converge at high counts is thus a sign that counting (i.e., Poisson) statistics dominate. This confirms that the SXS follows the standard behavior for normal photon-counting detectors.

The difference of irradiating sources for the two pixels is not likely to affect our results. Although there may be minor differences in the spectral features (Schnopper 1967), these will be too small to impact our fits. This is supported by the fact that we see the same difference for both pixels – the systematic shift is between the two fitting methods, and not between the two pixels.

Based on figure 2, χ^2 statistics should only be used when bins have ~ 20 counts or more. This applies when counting statistics dominate. In cases where the data are known to follow Gaussian statistics, χ^2 likely provides a better description. The tests show that the behavior of the data closely matches that of the simulations, and thus confirms that counting statistics are the dominant source of error when fitting the PSP spectra. In the example of calibration through measurements of Mn K α , using χ^2 to fit the data gives too optimistic values for the energy resolution. Fitting high-resolution data, where counts per bin are low, requires Poisson statistics to give reliable results.

The reason behind the difference between the two methods can be understood by considering the differences between the two underlying distributions, Gaussian and Poisson. For large numbers the two distributions become very similar, and the Poisson asymptotically approaches a Gaussian. However, at low counts the difference between the two distributions becomes larger and starts to affect the results of the fit. The average counts per bin for the calibration data at 0.8 eV bin size and 1 ks integration time was ~ 15 for pixel 23 and ~ 35 for pixel 12.

When estimating the total probability of the observed data given the model using equations (A4) and (A6) (see appendix 1), m_i is taken to be the mean of the model. However, the value splitting a distribution into two equally populated halves is given by the median. As the Gaussian distribution is symmetric the median and mean will be the same, allowing the model to be placed either above or below the data with equal probability. This is not the case for the Poisson distribution, as the extended tail to high

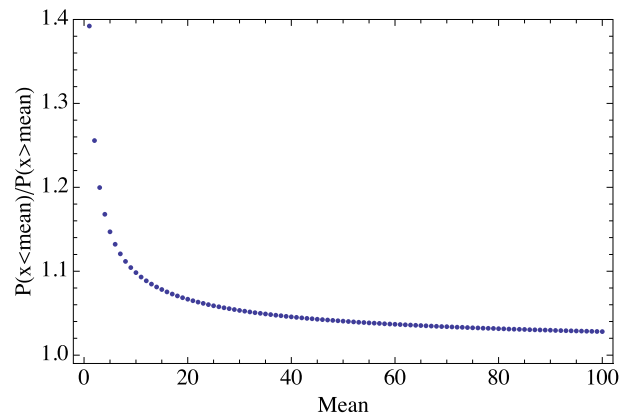


Fig. 5. Probability that a randomly selected data point in a Poisson distribution is below the mean, divided by the probability of being above the mean. The ratio is plotted as a function of the mean value for the distribution. For low values of the mean there is a very clear tendency for most samples to be below the mean and the ratio asymptotically approaches 1 as the mean increases. A Gaussian distribution is symmetric, and thereby the ratio is 1 for all values of the mean. (Color online)

values leads to the mean being higher than the median. The probability is thereby greatest that the measurement is *below* the mean, and thus using the C statistic will preferentially place the model *above* the data. This is illustrated in figure 5, which shows the relative probability that a randomly chosen value will be below the mean compared to that of being above the mean. When the mean is low there is a substantial difference from the Gaussian case, where the ratio is always 1.

Using the C statistic when fitting models with narrow features, such as emission and absorption lines, will therefore result in the model having a greater width than when using Gaussian statistics. As long as Poisson statistics dominate, using χ^2 to fit such features will lead to an incorrect (too narrow) estimate of their width. The C statistic may be used for both low and high counts and is recommended for all cases where counting statistics dominate the uncertainties. As seen in figure 5, the asymmetry is present at the level of a few percent even when there are > 100 counts bin^{-1} . Such a bias at high count levels matches the results of Humphrey, Liu, and Buote (2009), who showed that the effect can even be comparable to the statistical errors in a study using X-ray data from Chandra. The standard χ^2 method should thus only be used in cases when the uncertainties can be assumed to be Gaussian.

Unlike C statistics, the standard χ^2 statistic is a weighted statistic, and uses the variance when determining the weight of each contribution—cf. equation (A4). In the case of count data the variance is often assumed to be equal to the mean (i.e., mimicking Poisson statistics), which leads to problems in bins with no counts as the variance becomes 0.

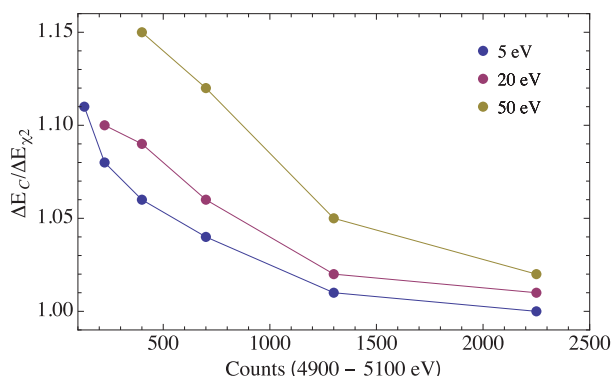


Fig. 6. Relative difference in recovered line width when fitting using Poisson and Gaussian statistics. The two statistics give a larger discrepancy when there are low counts and for broad lines. (Color online)

The problem is naturally most noticeable in cases of sparse data, and has led to various alternative weighting schemes to allow the use of a χ^2 statistic in the low count regime. A comparison of the several common variants of the χ^2 statistic is presented by Mighell (1999), who found that a modified weighting scheme may allow the χ^2 statistic to perform as well as C statistics for low counts.

We stress that the considerations above apply to cases where there is no significant source of background. In realistic situations this will, of course, not be the case, and the statistics of the background must also be considered. This can lead to a modification of the test statistic to use. Subtracting a Gaussian background—often assumed when the background is modeled from spacecraft location, attitude, etc.—will allow χ^2 statistics to remain valid in the case of a Gaussian signal, as the difference between two Gaussian distributions is also Gaussian. However, for a Poisson signal the statistic will change from the basic C statistic used here. Spectral fitting routines such as `xspec` include several test statistic options to choose from, such as `pgstat` for a Poisson signal with Gaussian background and the Castor C statistic (`cstat`; Arnaud et al. 2011) for a Poisson signal with a Poisson background.

The results of the simulations use an approximate SXS response and also include some background. They may therefore be used as a rough guide in the choice of test statistic. The discrepancy between the two statistics can be seen in figure 6, where the shift is plotted as a function of total counts in the fit range. Clearly, a narrow line is less sensitive to the choice of statistic, but even for medium count levels an inappropriate choice could introduce an error on the level of a few percent.

It is important to note that the choice of fit statistic (and weighting scheme) also affects the estimation of parameter confidence intervals. A common way to do this is to look for parameter values where the fit statistic increases from the

best fit by a given amount. However, the amount of increase corresponding to a certain confidence interval is generally calculated assuming that the difference in test statistic follows a χ^2 distribution with one degree of freedom per parameter being tested. While this is correct when using χ^2 statistics, it is only asymptotically true for C statistics. In cases where the two methods cannot be assumed to be similar (e.g., with sparse data), parameter estimation for C statistics is better performed using Monte Carlo techniques. Such techniques can in fact be used for all test statistics, but have the drawback that they can be computationally intensive.

5 Conclusions

We have compared the use of Gaussian and Poisson statistics when fitting X-ray spectra with low counts. Systematic differences between the two have been investigated using simulations, confirming the previously known bias of the χ^2 statistic. We have further applied this to data taken from an SXS detector system test, demonstrating the effects on the recovered energy resolution. The main conclusions are:

- Fitting narrow features (such as emission/absorption lines) using the χ^2 method may lead to underestimation of the width.
- The χ^2 method gives incorrect results when there are few counts per bin.
- Maximum likelihood using Poisson statistics should be employed when accuracy better than $\sim 10\%$ is needed for a line with a hundred photons for an energy resolution of ~ 5 eV.

The results can be explained by the properties of the underlying statistical distributions.

Acknowledgments

The data collection and analysis reported here was supported by JAXA, NASA, and many individual researchers, especially Kevin R. Boyce, M. P. Chiao, and T. Watanabe (NASA/GSFC); S. Takeda, M. S. Tashiro (Saitama University); and M. Tsujimoto (JAXA/ISAS). S.Y. was supported by JSPS KAKENHI Grant Numbers 16H02190, 15H05438, 18H03714, and 18H05458. Part of this work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Appendix 1. Brief statistical background

This section describes the statistical background of the χ^2 and Poisson maximum likelihood methods in the application of parameter estimation. For a more complete discussion, we refer readers to Siemiginowska (2011).

The standard approach to finding the best-fitting values of model parameters is to maximize the probability of the observed data given this model. The total probability is thus

$$P(\text{data} \mid \text{model}) = \prod_i P(y_i \mid \text{model}), \quad (\text{A1})$$

where y_i is the observed count rate in bin i . In order to estimate the probability, a distribution must be assumed for the underlying data. The most common one assumed is Gaussian, giving

$$P = \prod_i \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[-\frac{(y_i - m_i)^2}{2\sigma_i^2} \right], \quad (\text{A2})$$

where y_i and σ_i are the value and estimated uncertainty of the measured data, and m_i the predicted value from the model. Because it is easier to handle sums than products, rather than maximize the probability directly, twice the negative logarithm is instead minimized. The equation then becomes

$$-2 \ln P = \sum_i \frac{(y_i - m_i)^2}{\sigma_i^2} - 2 \ln \frac{1}{\sigma_i \sqrt{2\pi}}. \quad (\text{A3})$$

Since the last term only depends on the data (and thus will not change with different model parameters) it can be ignored, giving the common χ^2 statistic:

$$\chi^2 = \sum_i \frac{(y_i - m_i)^2}{\sigma_i^2}. \quad (\text{A4})$$

When using Poissonian statistics, the probability instead becomes

$$P = \prod_i (m_i)^{y_i} \frac{\exp(-m_i)}{y_i!}, \quad (\text{A5})$$

giving the function to be minimized as (Cash 1979)

$$C = 2 \sum_i m_i - y_i \ln m_i + \ln y_i!. \quad (\text{A6})$$

In the main text, this is referred to as the C statistic.

Appendix 2. The Mn K α complex

Both irradiation of a Mn source and the radioactive decay of ^{55}Fe give rise to excited atoms of ^{55}Mn . The atoms can then relax through emission of an Auger electron or X-rays. The strongest X-ray emission comes from the K α complex around 5.9 keV. This is a collection of multiple lines corresponding to transitions with similar energies. The strength of each contribution is determined by the probability of

Table 4. Model parameters of Mn K α emission lines.

Name	Peak (eV)	Width (eV)	Amp.*	Prob.†
K α_{11}	5898.853	1.715	0.790	0.353
K α_{12}	5897.867	2.043	0.264	0.141
K α_{13}	5894.829	4.499	0.068	0.080
K α_{14}	5896.532	2.663	0.096	0.067
K α_{15}	5899.417	0.969	0.071‡	0.018
K $\alpha_{16}^\ddagger$	5902.712	1.553	0.011	0.004
K α_{21}	5887.743	2.361	0.372	0.229
K α_{22}	5886.495	4.216	0.100	0.110

*Amplitude of the profile.

†Relative probability of the line.

‡G. Hölzer, private communication to F. S. Porter.

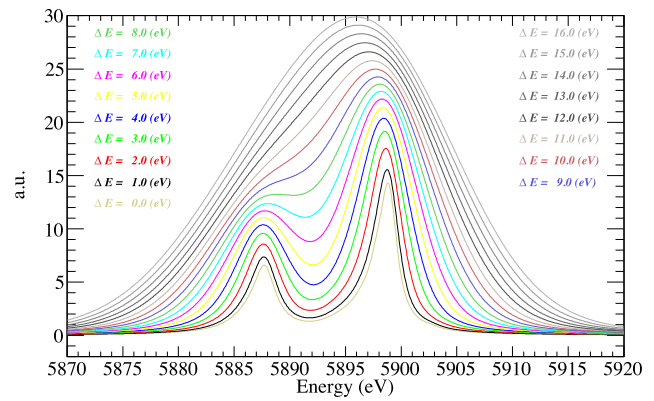


Fig. 7. Illustration of how the Mn K α complex is “smeared” by the energy resolution of the instrument. Each curve shows the total spectrum for the corresponding energy resolution. At high resolution (low values) the two peaks are distinct, whereas at values above ~ 8 eV, the peaks blend into one large feature. (Color online)

the corresponding transition. There is no complete analytical solution for this; instead, the complex is described by a phenomenological combination of Voigt profiles determined through experimental measurements. The analysis presented here is based on a combination of eight such profiles, the parameters of which are listed in table 4.

When photons are detected in the calorimeter, the recovered energy will vary from the true energy by a random amount. The energy resolution gives a measure for how great this deviation can be. Observationally, this uncertainty in energy leads to the measured spectrum being “smeared” in energy—features will become broader and smoother. The effect of this on the double peaks of the Mn K α complex is illustrated in figure 7. By comparing the amount of “smearing” needed to make the theoretical input spectrum match the observed one, the intrinsic energy resolution of the instrument can be measured. For actual observations, additional effects such as background may provide additional sources of uncertainty.

References

- Arnaud, K., Smith, R., & Siemiginowska, A. ed. 2011, *Handbook of X-ray Astronomy* (Cambridge, UK: Cambridge University Press)
- Boyce, K. R., et al. 1999, *SPIE Proc.*, 3765, 741
- Cash, W. 1979, *ApJ*, 228, 939
- Hölzer, G., Fritsch, M., Deutsch, M., Härtwig, J., & Förster, E. 1997, *Phys. Rev. A*, 56, 4554
- Humphrey, P. J., Liu, W., & Buote, D. A. 2009, *ApJ*, 693, 822
- Ishisaki, Y., et al. 2016, *SPIE*, 9905, 99053T
- Kelley, R. L., et al. 2016, *SPIE Proc.*, 9905, 99050V
- Kilbourne, C. A., et al. 2016, *SPIE Proc.*, 9905, 99053L
- Laurence, T. A., & Chromy, B. A. 2010, *Nature Methods*, 7, 338
- Leccardi, A., & Molendi, S. 2007, *A&A*, 472, 21
- Mighell, K. J. 1999, *ApJ*, 518, 380
- Mitsuda, K., et al. 2010, *SPIE Proc.*, 7732, 773211
- Nousek, J. A., & Shue, D. R. 1989, *ApJ*, 342, 1207
- Porter, F. S., et al. 2010, *SPIE Proc.*, 7732, 77323J
- Porter, F. S., et al. 2016, *SPIE Proc.*, 9905, 99050W
- Protassov, R., van Dyk, D. A., Connors, A., Kashyap, V. L., & Siemiginowska, A. 2002, *ApJ*, 571, 545
- Schnopper, H. W. 1967, *Phys. Rev.*, 154, 118
- Shimoda, Y., et al. 2012, *J. Low Temp. Phys.*, 167, 575
- Siemiginowska, A. 2011, in *Handbook of X-ray Astronomy*, ed. K. A. Arnaud et al. (Cambridge, UK: Cambridge University Press), chap. 7
- Szymkowiak, A. E., Kelley, R. L., Moseley, S. H., & Stahle, C. K. 1993, *J. Low Temp. Phys.*, 93, 281
- Takahashi, T., et al. 2010, *SPIE Proc.*, 7732, 77320Z
- Takeda, S., et al. 2014, *SPIE Proc.*, 9144, 91445B
- Tsujimoto, M., et al. 2016, *SPIE Proc.*, 9905, 99050Y