Combining Statistical and Systematic Uncertainties for High Count Spectra

H.Lee, V.Kashyap, A.Siemiginowska, A.Connors, J.Drake, P.Ratzlaff, D.van Dyk

Abstract

Instrument effective areas affect spectral fitting results, and the derived parameter values will depend on the particular effective area chosen for the analysis. This is an issue for high-count, high-quality data more than low-counts spectra because systematic uncertainties may produce a larger variance in the parameters than statistical uncertainties, whereas statistical uncertainties dominate for low counts spectra.

With plausible effective areas that span the systematic variations, we show that their effect can be easily computed using existing tools like Sherpa. Here we describe that the variances due to statistical and systematic uncertainties can be combined with the well-known multiple imputation combining rule. We demonstrate its effectiveness with simulation examples using typical variations expected of ACIS-S effective areas. We show that approximately 20 full-fledged spectral fits, carried out with different effective areas, are needed to capture the full extent of calibration uncertainty in the parameter error estimates. We apply this technique to a set of observed AGN spectra and directly demonstrate the relevance of such calculations.

Effective Area Uncertainty

Best fit parameters and error bars depend on the choice of a calibration file. A set of effective areas can be synthesized by calibration scientists and its information must be aggregated into spectral fitting analysis. Fig. 1 illustrates a set of 100 effective area curves (A) showing the systematic uncertainty in the nominal ACIS-S effective areas (Drake et al. 2006; see also Drake et al. 2009 [C.11]). The top panel shows the set of effective areas, the middle panel shows the deviations from the nominal, and the bottom panel shows the same set reconstructed from a small subset of its Principal Components. Some illustrative curves are shown in the middle panel. The dark grey areas cover 68% of the variations.

Calibration Uncertainty & Inference

Without considering calibration uncertainty, best fit parameters are subject to be biased and error bars are underestimated. Fig. 5 shows the behavior of posterior probability densities (pdf) conditioning on different effective area from fitting a powerlaw*wabs faked spectrum (E = 2, N$_{90}$ = 10$^{27}$phot/s, 10$^5$ counts). The center and width of each pdf indicates best fit and statistical error. The ellipses are 90% contours. Colors identify 30 arbitrary selected effective areas and matching best fits and errors. These systematic variations due to calibration uncertainty are difficult to be quantified independent of model and data.

Importance of How to combine Errors

Given a set of effective areas, (A$_i$, i = 1, ..., m). Sherpa can compute a series of best fit parameters (θ$_i$) and associated statistical error bars (σ$_{stat}$). The average of (θ$_i$) is an estimate of the statistical error σ$_{stat}$. The spread in (θ$_i$) represents the variations due to systematic error. These two can be added in quadrature to get the combined error on the best fit.

Using the multiple imputation combining rule (Schafer 1999) to achieve this. The average statistical error, also the within variance W$_{stat}$, is

W$_{stat}$ = $\sqrt{\sum_{i=1}^{m} \sigma_i^2}$

and the variance in the best fits, also called the between variance B$_{stat}$, is

B$_{stat}$ = $\frac{1}{m} \sum_{i=1}^{m} \bar{\theta}_i - \bar{\theta}^2$

The total variance T$_{stat}$ is

T$_{stat} = B_{stat} + (1 + 1/m)B_{stat}$

The best parameter fit θ is the average of (θ$_i$) and the combined error σ$_{combined}$ is $\sqrt{T_{stat}}$. Note that (1 + 1/m)B$_{stat}$ estimates the increase in variance due to the effective area uncertainty. This combining rule approximates the sampling distribution of θ by t distribution

T$_{stat}$/m$^{1/2} (t_{B} - \bar{\theta}) \sim E$

with the degree of freedom ν = (m - 1)(1 + 1/m)B$_{stat}$. The confidence interval from this t distribution results in an error bar of statistical and systematic uncertainties combined.

How many calibration files?

Fig. 4 shows m = 20 to fit the same powerlaw*wabs model of Fig. 3 with calibration uncertainty by choosing m random A to compute T$_{stat}$ for 100 times. The average and sample std. dev. of 100 T$_{stat}$ is marked by dots and vertical lines.

Example

The same QSO spectra in Fig. 2 are fit to compare uncertainties from BLOCS and Sherpa/combining rule (m = 20). Simply applying the combining rule within existing tools like Sherpa can handle calibration uncertainties in spectral fitting analysis.

Procedure

- spectrum, ARF, RMF, information of effective area uncertainty.
- fit: get confidence intervals
- for i = 1, ..., m
- and statistical error σ$_{stat}$
- apply the combining rule Eq (1) on (θ$_i$) and (σ$_{stat}$) for θ and σ$_{combined}$

Summary

- Calibration uncertainties inflate error bars in best fit parameters of high count spectra.
- Eq (1) provides a simple rule of combining systematic and statistical uncertainties.
- This combining rule can be adapted to other types of calibration uncertainties as long as the sample or the summary covers the full span of variations.
- Plausible samples of calibration uncertainties will be provided through chain/Sherpa distribution to be incorporated into spectral analysis as described in Procedure.

References

Drake, J. et al. (2009) [C.10]
Schafer, J.L. (1999) SMMR, 8, p.3

Acknowledgment

This work was supported by NASA/ASRP grant NNG06GF17G (HLAC) and by CXC NASA contract NAS8-30973 (VLK, AS, JJD, PR).