Fitting Narrow Emission Lines in X-ray Spectra

Taeyoung Park (tpark@stat.harvard.edu, Department of Statistics, Harvard University),
David A. van Dyk (Department of Statistics, University of California at Irvine), and
Aneta Siemiginowska (Harvard-Smithsonian Center for Astrophysics).

CHASC (California-Harvard AstroStatistics Collaboration)

www.ics.uci.edu/~dvd/astrostat.html

Abstract

Spectral emission lines are local spectral features that represent extra emissions of photons in a narrow band of energy. Given the physical constraints, it is often appropriate to model the emission lines with a very small width in a spectral model.

The counts collected by *Chandra* are subject to various data degradation mechanisms such as absorption, effective area, instrument response, and background contamination; this necessitates accounting for the data generation mechanism within the spectral model. Bayesian methods offer a straightforward way of handling the complexity of *Chandra* data. When a Gaussian function is used to model the spectral lines, standard Bayesian statistical algorithms are readily available to solve the problem; see van Dyk et al. (2001) for details. As the width of the line tends to zero, however, the algorithms simply break down requiring the use of state-of-the-art statistical algorithms. Here, we illustrate how to detect the narrow spectral lines in X-ray spectra and statistically test the significance of the lines with the *Chandra* data sets for the energy spectrum of the high redshift quasar, PG 1634+706.

The Quasar, PG 1634+706

- PG 1634+706 is a redshift z=1.334 radio quiet and optically bright quasar. The source was observed with the *Chandra X-ray Observatory* as a calibration target six times on March 23 and 24, 2000.
- Fluorescent Fe-K-α emission line near **6.4 keV** has been observed in quasars. We look for this line in the Chandra spectrum of PG1634+706.

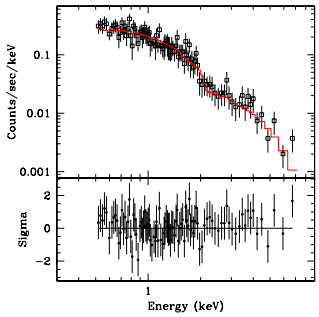

Figure 1: *Chandra* ACIS-S spectrum of PG1634+706 (obs-id 62) fit with the absorbed power law model in Sherpa. The data are grouped so the S/N> 3σ in each bin. The top panel shows the data (black points) and the best fitted model (red), and the bottom panel residuals.

Modeling the X-ray Spectrum

X-ray Spectra:

- Low counts are expected in each bin of the X-ray spectrum.
- The Gaussian assumptions that are inherent in traditional χ^2 fitting are inappropriate for such low count data.
- Instead, we explicitly model photon arrivals as an inhomogeneous Poisson process (van Dyk et al., 2001).

The Basic Spectral Model: The X-ray spectrum can be separated into (1) continuum and (2) emission lines.

- The continuum is described by a parametric form (e.g., a power law).
- The emission lines are statistically described by adding a narrow Gaussian, a narrow Lorentzian, or a delta function to the continuum.

Data Degradation:

- The photons are subject to various physical processes which significantly degrade the source model. Namely,
  1. absorption,
  2. the effective area (e.g., ARF),
  3. blurring of photons' energy (e.g., RMF),
  4. photon pile up, and
  5. background contamination.
- We design a highly structured multilevel spectral model with components for the data degradation processes (van Dyk et al., 2001).

Fitting a Spectral Model

New Statistical Algorithms (Park et al., 2004):

- **Maximum A Posteriori Fitting**: For each *Chandra* data set, we search for probable line locations with 56 different starting values equally spaced between 0.5 keV and 6.0 keV.

observed data set	mode (keV)	starting values (keV) converged to each mode	log posterior probability
obs-id 47	**2.885**	0.5–6.0	2026.66
obs-id 62	**2.845**	0.5–6.0	1594.60
obs-id 69	1.805	0.5–6.0	1568.41
obs-id 70	**2.835**	0.5–6.0	1913.54
obs-id 71	**2.715**	0.5–3.7, 3.9–4.6, and 4.9–5.8	1355.31
	5.605	3.8, 4.7–4.8, and 5.9–6.0	1354.90
obs-id 1269	**2.905**	0.5–6.0	3341.49

Table 1: Identified Modes for the Line Location.

- **The Posterior Distribution of the Line Location**: For each *Chandra* data set, we simulate the line location starting at the identified probable locations to investigate their uncertainty.

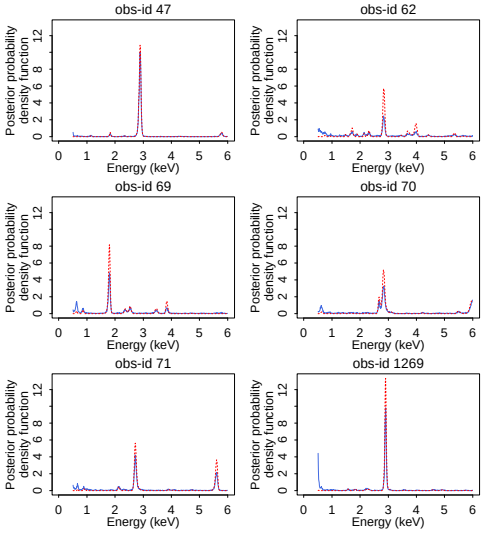

Figure 2: The Posterior Distribution of the Line Location. The blue solid line of each panel represents the posterior distribution of the line location, and the orange dotted line the profile posterior distribution.

- Table 2 presents the 95% HPD (Highest Posterior Density) region of the line location for each data set. A HPD region consists of a number of intervals; we show only the intervals whose probabilities are greater than 10%.

observed data set	mode (keV)	HPD region for μ	posterior probability
obs-id 47	**2.885**	(2.67, 3.07)	83.74%
obs-id 62	0.545	(0.50, 1.99)	35.21%
	2.845	(2.45, 2.99)	23.78%
	3.985	(3.36, 4.10)	15.65%
obs-id 69	0.625	(0.50, 1.21)	23.64%
	1.805	(1.57, 1.97)	39.51%
	2.535	(2.20, 2.82)	13.33%
obs-id 70	0.605	(0.50, 0.99)	13.14%
	2.845	(2.51, 3.19)	50.82%
	5.995	(5.73, 6.00)	14.36%
obs-id 71	0.665	(0.50, 1.19)	15.41%
	2.715	(2.50, 3.07)	41.78%
	5.595	(5.36, 5.72)	21.99%
obs-id 1269	0.505	(0.50, 1.21)	18.25%
	2.905	(2.75, 3.12)	65.11%

Table 2: 95% HPD Regions for the Line Location.

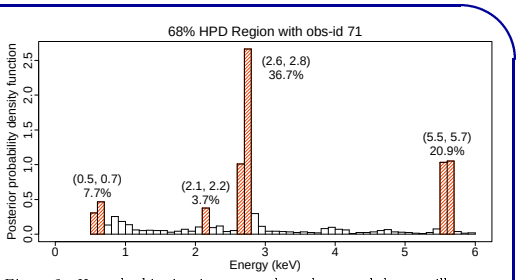

Figure 3: Here the binning is coarser than the actual data to illustrate how we construct the HPD region. The 68% HPD region contains the bins with the most probability and has total probability of at least 68%.

Testing the Evidence for the Line

Posterior Predictive Check: We compare three models to quantify the evidence in the data for the emission line:

Model 0: There is no emission line.

Model 1: There is an emission line with fixed location in the spectrum.

Model 2: There is an emission line with unknown location and intensity.

observed data set	Model 0 vs. Model 1			Model 0 vs. Model 2	
	fixed location	$T_1(Y_{obs})$	ppp-value	$T_2(Y_{obs})$	ppp-value
obs-id 47	2.885	5.53	**0.001**	5.53	**0.036**
obs-id 62	2.845	2.42	**0.015**	2.42	0.638
obs-id 69	2.865	0.00	1.000	3.54	0.270
obs-id 70	2.845	3.10	**0.007**	3.13	0.366
obs-id 71	2.715	3.41	**0.004**	3.42	0.274
obs-id 1269	2.905	4.96	**0.002**	4.96	**0.076**

Table 3: We test the statistical evidence for the line by using pp-check.

- A test statistic, $T_k(Y_{obs})$, is the likelihood ratio of Model k to Model 0.
- Model 1 fixes the line locations identified near **2.74 keV** (i.e., **6.4 keV** in the rest frame) from the posterior distributions, while Model 2 assumes no such prior information as to the line.
- The ppp-value is interpreted as how likely the observed test statistic is as compared to the posterior predictive simulated test statistics.
- A low value of ppp-value **(e.g., < 0.1)** indicates Model 1 or 2 is much more likely than Model 0, which gives the evidence for the line.

Conclusions

- Combining all of the data sets, we detect the line at 2.865 keV.

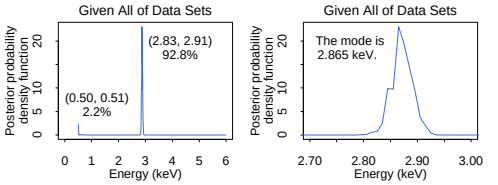

Figure 4: Given all of the data sets, the interval **(2.83,2.91)** has the most posterior probability in the 95% HPD region, which gives the strong evidence for the line at **2.865 keV** (i.e., **6.68 keV** in the rest frame).

References

Park, T., van Dyk, D. A., and Siemiginowska, A. (2004). Spectral analysis with delta functions emission lines. *A CHASC Technical Report* .

van Dyk, D. A., Connors, A., Kashyap, V., and Siemiginowska, A. (2001). Analysis of energy spectra with low photon counts via Bayesian posterior simulation. *The Astrophysical Journal* **548**, 224–243.

Acknowledgments

We gratefully acknowledge funding for this partially provided by NSF Grants DMS-01-04129, DMS-04-38240, and DMS-04-06085 and by NASA Contracts NAS8-39073 and NAS8-03060(CXC). This work is a product of joint work with the California-Harvard astrostatistics collaboration (CHASC) whose members include J. Chiang, A. Connors, D. van Dyk, D. Esch, P. Freeman, H. Kang, V. L. Kashyap, X.-L. Meng, A. Siemiginowska, E. Sourlas, T. Park, Y. Yu, and A. Zezas.