Efficient X-ray Spectral Fitting with Narrow Emission Lines

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).

Abstract

From a statistical point of view, spectral analysis is the modeling of the distribution of photon energies. This distribution can be formulated as a finite mixture of two photon groups, a continuum term and a set of emission lines. While the continuum describes a general shape of a spectrum, each emission line represents a positive aberration from the continuum in a narrow band of energies. Here, we focus on a single emission line that can be modeled with a Gaussian distribution or a delta function. Spectral data are contaminated by several non-trivial physical processes including non-homogeneous stochastic censoring, blurring of photon energies, and background contamination. To account for these processes, we consider a hierarchical structure of missing data under a Bayesian perspective. To fit the resulting highly structured multilevel spectral models, we devise efficient Gibbs sampling strategies. As an illustration, we apply our strategies to the X-ray spectrum of the high redshift quasar, PG1634+706.

Modeling the X-ray Spectrum

X-ray Spectra:

- The Gaussian assumptions that are inherent in traditional χ^2 fitting are inappropriate for low count data in each bin of the X-ray spectrum.
- 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) a continuum term and (2) a set of 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,
  3. blurring of photon energies (stochastic redistribution), and
  4. background contamination.
- We design a highly structured multilevel spectral model with components for the data degradation processes (van Dyk et al., 2001).

Efficient X-ray Spectral Fitting

Hierarchical Structure of Missing Data:

- $\mathbf{Y}_{\text{mis1}}$: variables that describe the absorption, effective area, blurring, and background contamination.
- $\mathbf{Y}_{\text{mis2}}$: a mixture component indicator variable that indicates which photons originated from the line and which from the continuum.

Partitions of Model Parameters:

- ψ : parameters for the continuum.
- μ : a parameter for the line location.
- ν : a parameter for the line width; for a delta function, ν is set to 0.

Difficulty with Identifying Narrow Emission Lines:

- When an emission line is modeled with a delta function, an ordinary Gibbs sampler to fit the finite mixture distribution breaks down because the mixture indicator variables attribute the same subset of photons to the line from iteration to iteration. The line location cannot move from its starting value.
- We suggest fitting the line location without conditioning upon the mixture components indicator variables.
- This difficulty persists when the location and width of a narrow Gaussian emission line are simultaneously fitted.

Partially Marginalized Gibbs Samplers:

- Park and van Dyk (2006) devised several partially marginalized Gibbs samplers for a spectral model with a delta function or Gaussian line.
- Our goal is to construct quickly converging samplers with the target posterior distribution as their stationary distribution.

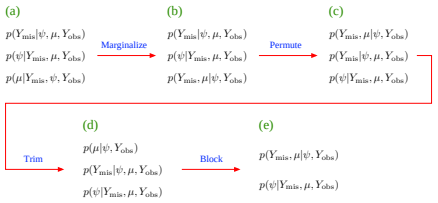

Figure 2: Illustration of the derivation of the partially marginalized Gibbs sampler when we partially marginalize all of the missing data out of the sampling step for the delta function line location.

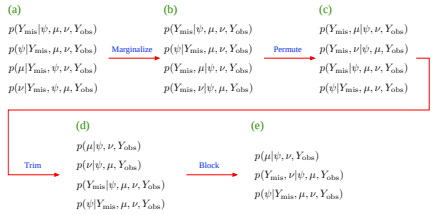

Figure 3: Illustration of the derivation of the partially marginalized Gibbs sampler when we partially marginalize all of the missing data out of the sampling steps for the Gaussian line location and width.

Simulation Study

Four Different Cases: Data are simulated under the following cases.

- **Case 1**: No emission line.
- **Case 2**: One narrow and weak line at 2.85 keV with width 0.04 keV.
- **Case 3**: One wide and strong line at 3.4 keV with width 0.207 keV.
- **Case 4**: One narrow and strong line at 3.4 keV with width 0.04 keV.

The Posterior Distribution of the Line Location:

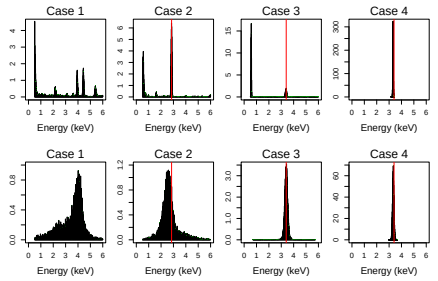

Figure 4: The top row corresponds to the posterior distribution of the **delta function line location**; the bottom row that of the **Gaussian line location**. The red solid line represents the true line location.

- **Case 2** : With narrow lines, delta functions are better suited for fitting the data.
- **Case 3** : With wide lines, Gaussian lines are better suited.
- **Case 4** : With narrow strong lines, both models work well.

Posterior Predictive Check:

- We compare two models to quantify the evidence in the data for the emission line via $T(y^{\text{rep}}) = \log\{\sup_{\theta\in\Theta_1} L(\theta|y^{\text{rep}})/\sup_{\theta\in\Theta_0} L(\theta|y^{\text{rep}})\}$ where Θ_k represents the parameter spaces under Model k and y^{rep} denotes replicated data sets under Model 0:
  - Model 0 : There is no emission line in the spectrum.
  - Model 1 : There is one emission line in the spectrum.

Simulated data set	Delta function line		Gaussian line	
	$T(Y_{\text{obs}})$	ppp-value	$T(Y_{\text{obs}})$	ppp-value
Case 1	2.51	0.541	4.94	0.471
Case 2	4.57	**0.101**	5.40	0.318
Case 3	5.96	**0.025**	11.58	**0.002**
Case 4	18.75	**0.000**	24.43	**0.000**

Table 1: We test the statistical evidence for the line by using pp-check. A low value of ppp-value (e.g., <0.1) indicates Model 1 is more likely than Model 0, which gives evidence for the line.

Spectral Analysis of PG1634+706

The Quasar, PG1634+706:

- PG1634+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.
- The fluorescent Fe-K-α emission line has been observed in the quasar rest frame of near **6.4 keV**, which corresponds to **2.74 keV** in the observed frame of PG1634+706.
- The **location** of the line indicates the ionization state of iron in the emitting plasma; the **width** of the line the velocity of the plasma.

Fitting a spectral model to the *Chandra* data of PG1634+706:

- For each of the six *Chandra* data sets, we fit the spectral model with a delta function or Gaussian emission line.
- After the fitting, we compare the Fe-K-α emission line (**2.74 keV**) observed in the other similar sources with the fitted line location in the X-ray spectrum of PG1634+706.

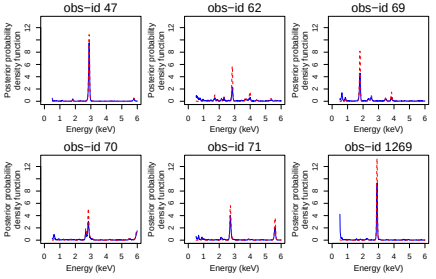

Figure 5: The posterior distribution of the **delta function line location** resulting from the spectral analysis of PG1634+706 data. The blue line of each panel represents the posterior distribution; the red dotted line the profile posterior distribution.

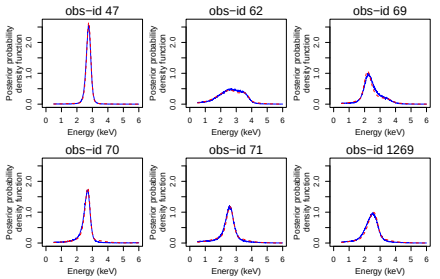

Figure 6: The posterior distribution of the **Gaussian line location** resulting from the spectral analysis of PG1634+706 data. The blue line of each panel represents the posterior distribution; the red dotted line the profile posterior distribution.

Posterior Predictive Check for *Chandra* Data of PG1634+706:

Observed data set	Delta function line		Gaussian line	
	$T(Y_{\text{obs}})$	ppp-value	$T(Y_{\text{obs}})$	ppp-value
obs-id 47	5.53	**0.036**	8.70	**0.019**
obs-id 62	2.42	0.638	5.10	0.430
obs-id 69	3.54	0.270	4.76	0.567
obs-id 70	3.13	0.366	6.68	**0.104**
obs-id 71	3.42	0.274	5.69	0.265
obs-id 1269	4.96	**0.076**	5.73	0.229

Table 2: We test the statistical evidence for the line in the spectrum of PG1634+706 by using pp-check. The evidence for the inclusion of the line is shown in obs-id 47, obs-id 70, and obs-id 1269.

Conclusions

- Combining all the Chandra data sets for PG1634+706, we detect the delta function line at **2.865 keV** (posterior mode) and the Gaussian line at **2.650 keV** (posterior mean) in the observed frame.
- These observed lines are redshifted into **6.69 keV** (ionization state) and **6.19 keV** (neutral iron) in the quasar rest frame, respectively.

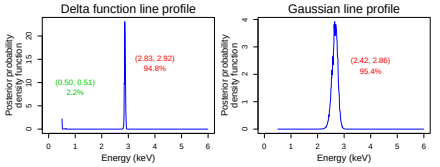

Figure 7: The posterior distribution of the line location given all of the data sets gives the strong evidence for the line near the Fe-K-α emission line (**2.74 keV**) observed in the other similar sources.

Acknowledgments

The authors gratefully acknowledge funding for this project partially provided by NSF grant 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, www.ics.uci.edu/~dvd/astrostat.html) whose members include J. Chiang, A. Connors, D. van Dyk, V. L. Kashyap, X.-L. Meng, J. Scargle, A. Siemiginowska, E. Sourlas, T. Park, Y. Yu, and A. Zezas.