Fitting Narrow Emission Lines in X-ray Spectra

Taeyoung Park

Department of Statistics, University of Pittsburgh

October 11, 2007
This talk has three components:

A. Motivation
- Looking for narrow emission lines in high energy spectra.
- Spectral analysis of the high redshift quasar, PG1634+706.

B. Goals
- Building and fitting highly structured spectral models.
- Model-based statistical inference for line locations.
- Posterior predictive checks for a model component.

C. Conclusion
Astronomical Source
Chandra X-ray Observatory

1. Field of View 2.5 Deg
2. Focal Surface
3. 4 Nested Hyperboloids
4. Doubly Reflected X-rays
5. 4 Nested Paraboloids
6. X-rays
7. 10 meters
8. Mirror elements are 0.8 m long and from 0.6 m to 1.2 m diameter

Taeyoung Park <tpark@pitt.edu>
An observed spectrum consists of photon counts in a number of energy bins.

Use non-homogenous Poisson process.

The photons are originated from two groups: (1) the continuum and (2) several emission lines.
Observed spectrum of the high redshift quasar, PG1634+706:

1. Estimate parameters that describe a spectrum.
2. Investigate the uncertainty of the estimates.
3. Test the evidence for the emission lines.
The Basic Source Model

- A simplified Poisson process for the scientific model,

\[ X_i \sim \text{Poisson} \left( \Lambda_i = \alpha E_i^{-\beta} + \lambda \pi_i(\mu, \sigma^2) \right) \]

where \( i = 1, 2, \ldots, n \) index the energy bins.

- Construct a narrow emission line model using the delta function so that

1. the emission line is contained entirely in one energy bin, but
2. we do not know which bin.

- That is, \( \{\pi_i\} \) can be parameterized in terms of a single location parameter \( \mu \), and \( \sigma^2 \) is set equal to zero.
The Basic Source Model (Cont’d)

- The Poisson process: \( X_i \sim \text{Poisson} \left( \Lambda_i = \alpha E_i^{-\beta} + \lambda \pi_i(\mu) \right) \).

- Using *Data Augmentation* to fit this finite mixture model:

  \[ Z_{i\ell} = \begin{cases} 
  1 & \text{(indicator that photon } \ell \text{ in bin } i \text{ corresponds to the emission line)} \\
  0 & \text{otherwise} 
  \end{cases} \]

- With the target posterior distribution \( p(Z, \psi, \mu | X) \) where \( \psi = (\alpha, \beta, \lambda) \), the Gibbs sampler iterates among

1. Draw \( Z \sim p(Z | \psi, \mu, X) \), via \( Z_{i\ell} \sim \text{Bernoulli} \left( \frac{\lambda \pi_i(\mu)}{\alpha E_i^{-\beta} + \lambda \pi_i(\mu)} \right) \),

2. Draw \( \psi \sim p(\psi | Z, \mu, X) \), and

3. Draw \( \mu \sim p(\mu | Z, \psi, X) \).

*In This Case, the Gibbs Sampler Breaks Down!*
The Basic Source Model (Cont’d)

- The Poisson process: \( X_i \sim \text{Poisson} \left( \Lambda_i = \alpha E_i^{-\beta} + \lambda \pi_i(\mu) \right) \).

- Using \textit{Data Augmentation} to fit this finite mixture model:
  \[
  Z_{i\ell} = \begin{cases} 
  \text{indicator that photon } \ell \text{ in bin } i \quad & \text{corresponds to the emission line} 
  \end{cases}.
  \]

- With the target posterior distribution \( p(Z, \psi, \mu | X) \) where \( \psi = (\alpha, \beta, \lambda) \), the Gibbs sampler iterates among

  1. Draw \( Z \sim p(Z | \psi, \mu, X) \), via \( Z_{i\ell} \sim \text{Bernoulli} \left( \frac{\lambda \pi_i(\mu)}{\alpha E_i^{-\beta} + \lambda \pi_i(\mu)} \right) \),

  2. Draw \( \psi \sim p(\psi | Z, \mu, X) \), and

  3. Draw \( \mu \sim p(\mu | Z, \psi, X) \).

\textit{In This Case, the Gibbs Sampler Breaks Down!}
Why the Gibbs Sampler Fails

\[
\begin{align*}
\mu^{(0)} & = \bullet \\
\pi_i(\mu^{(0)}) & = 0, 0, 1, 0, 0
\end{align*}
\]

\[
Z_i \overset{iid}{\sim} \text{Ber}\left(\frac{\lambda \pi_i(\mu)}{\alpha E_i^{-\beta} + \lambda \pi_i(\mu)}\right)
\]

\[
\rightarrow Z = 0 \\
\rightarrow Z = 1
\]

---

Taeyoung Park <tpark@pitt.edu>

Fitting Narrow Emission Lines in X-ray Spectra
Why the Gibbs Sampler Fails

Why the Gibbs Sampler Fails

Total counts per bin

<table>
<thead>
<tr>
<th>Total counts, $X$</th>
<th>10</th>
<th>5</th>
<th>6</th>
<th>3</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu^{(0)}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\pi_i(\mu^{(0)})$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$\downarrow$

Continuum counts

<table>
<thead>
<tr>
<th>Continuum counts</th>
<th>10</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line counts</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$\downarrow$

$\mu^{(1)}$

$\mathbf{Z}_i \overset{iid}{\sim} \text{Ber} \left( \frac{\lambda \pi_i(\mu)}{\alpha E_i^{-\beta} + \lambda \pi_i(\mu)} \right)$

$\rightarrow Z = 0$

$\rightarrow Z = 1$
Improving the Convergence of the Gibbs Sampler

- The Gibbs sampler breaks down because the line location $\mu$ cannot move from its starting value.
- Because of the binning, we are left with a finite number of possible line locations.
- Thus, we can evaluate the observed posterior probability of $\mu$ at the midpoint of each bin.
- The observed posterior distribution for $\mu$ is given by

$$p(\mu|\psi, X) = \text{Multinomial}(1; \{p(\mu|\psi, X)|_{\mu=E_i}\}).$$

- To improve convergence, can we capitalize on $p(\mu|\psi, X)$?
- Note that $p(\mu|\psi, X) \propto \int p(Z, \psi, \mu|X) dZ$!
Introducing Incompatibility into the Gibbs Sampler

- The target posterior distribution is given by \( p(Z, \psi, \mu | X) \).
- Iteratively sample from complete conditional distributions:
  
  **Step 1**: Draw \( Z \sim p(Z | \psi, \mu, X) \),
  
  **Step 2**: Draw \( \psi \sim p(\psi | Z, \mu, X) \), and
  
  **Step 3**: Draw \( \mu \sim p(\mu | Z, \psi, X) \).

- Form a Markov chain with stationary distribution \( p(Z, \psi, \mu | X) \).
- **Basic Question**: When \( p(\mu | \psi, X) \) is available, use it or not?

  And the answer is... **USE WITH CAUTION**!
The target posterior distribution is given by \( p(Z, \psi, \mu | X) \).

Iteratively sample from complete conditional distributions:

1. Step 1: Draw \( Z \sim p(Z | \psi, \mu, X) \),
2. Step 2: Draw \( \psi \sim p(\psi | Z, \mu, X) \), and
3. Step 3: Draw \( \mu \sim p(\mu | Z, \psi, X) \).

Form a Markov chain with stationary distribution \( p(Z, \psi, \mu | X) \).

**Basic Question**: When \( p(\mu | \psi, X) \) is available, use it or not?

*And the answer is... USE WITH CAUTION!*
Construction of Partially Collapsed Gibbs Samplers

### Steps of Construction

- \( p(Z|\psi, \mu, X) \)
- \( p(\psi|Z, \mu, X) \rightarrow p(\psi|Z, \mu, X) \)
- \( p(\mu|Z, \psi, X) \)

### General Principle

- \( p(Z|\psi, \mu, X) \)
- \( p(\psi|Z, \mu, X) \)
- \( p(\mu|Z, \psi, X) \)

We move \( Z \) to the left of the conditioning sign in Step 3. This does not alter the stationary distribution, but improves the rate of convergence.

We permute the order of the steps. This can have minor effects on the rate of convergence, but does not affect the stationary distribution.

We remove \( Z \) from the draws in Step 1, since the transition kernel does not depend on this quantity.

We refer to these three steps as Marginalization, Permutation, and Trimming. They form a general strategy for constructing partially collapsed Gibbs (PCG) samplers.
Construction of Partially Collapsed Gibbs Samplers

**Steps of Construction**

\[
\begin{align*}
p(Z|\psi, \mu, X) &\quad p(Z|\psi, \mu, X) \\
p(\psi|Z, \mu, X) &\quad p(\psi|Z, \mu, X) \\
p(\mu|Z, \psi, X) &\quad p(Z, \mu|\psi, X)
\end{align*}
\]

**General Principle**

- We move \( Z \) to the left of the conditioning sign in Step 3. This does not alter the stationary distribution, but **improves the rate of convergence**.

\[
\begin{align*}
p(\psi|Z, \mu, X) &\quad p(Z, \mu|\psi, X) \\
\rightarrow &\quad p(Z|\psi, \mu, X) \\
p(\psi|Z, \mu, X) &\quad p(Z|\psi, \mu, X)
\end{align*}
\]

- We permute the order of the steps. This can have minor effects on the rate of convergence, but does not affect the stationary distribution.

\[
\begin{align*}
p(Z, \mu|\psi, X) &\quad p(Z|\psi, \mu, X) \\
\rightarrow &\quad p(Z|\psi, \mu, X) \\
p(\psi|Z, \mu, X) &\quad p(\psi|Z, \mu, X)
\end{align*}
\]

- We remove \( Z \) from the draws in Step 1, since the transition kernel does not depend on this quantity.

*We refer to these three steps as Marginalization, Permutation, and Trimming. They form a general strategy for constructing partially collapsed Gibbs (PCG) samplers.*
Improved Convergence Characteristics

**Theorem** (Park and van Dyk, 2007)

*Sampling more model components in any single step of a Gibbs sampler reduces the resulting maximal autocorrelation.*

- Directly assessing the convergence rate is difficult.
- The spectral radius of a forward operator
  - generally governs the convergence (Liu et al, 1995), and
  - is bounded above by the maximal autocorrelation.
- By reducing conditioning in any step (i.e., partial marginalization) we reduce the maximal autocorrelation, the upper bound of the spectral radius.
- We thus expect partially collapsed Gibbs samplers to have better convergence than ordinary Gibbs samplers.
Actually we do not observe the latent Poisson process,
\[ X_i \sim \text{Poisson}\left( \Lambda_i = \alpha E_i^{-\beta} + \lambda \pi_i(\mu) \right). \]

Rather, we observe
\[ Y_j \sim \text{Poisson}\left( \gamma_j \sum_i M_{ij} \Lambda_i + \theta_j^B \right), \]
because the photons are subject to a series of non-trivial physical processes:
- \( \gamma_j \): non-homogeneous stochastic censoring (absorption),
- \( M_{ij} \): stochastic redistribution (blurring), and
- \( \theta_j^B \): background contamination.
### Data Augmentation

We consider a hierarchical missing data structure:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal data separated by mixture indicators</td>
<td>( Z )</td>
</tr>
<tr>
<td>Mixed ideal data</td>
<td></td>
</tr>
<tr>
<td>Mixed ideal data after absorption</td>
<td>( X = {X_i} )</td>
</tr>
<tr>
<td>Mixed ideal data after absorption and blurring</td>
<td></td>
</tr>
<tr>
<td>Mixed ideal data after absorption, blurring, and background contamination, i.e., observed data</td>
<td>( Y_{\text{obs}} = {Y_j} )</td>
</tr>
</tbody>
</table>
The Gibbs Sampler

- The target posterior distribution is \( p(X, Z, \psi, \mu | Y^{\text{obs}}) \).
- The Gibbs sampler now iterates among
  
  \begin{align*}
  \text{Step 1} : & \quad \text{Draw } (X, Z) \sim p(X, Z | \psi, \mu, Y^{\text{obs}}), \\
  \text{Step 2} : & \quad \text{Draw } \psi \sim p(\psi | X, Z, \mu, Y^{\text{obs}}), \text{ and} \\
  \text{Step 3} : & \quad \text{Draw } \mu \sim p(\mu | X, Z, \psi, Y^{\text{obs}}),
  \end{align*}

  where \( \psi \) represents all model parameters but \( \mu \).

- With a Delta Function Line Model, the Gibbs Sampler Fails!
- The observed posterior distribution for \( \mu \) is computed as
  \[
  p(\mu | \psi, Y^{\text{obs}}) = \text{Multinomial}(1; \left\{ p(\mu | \psi, Y^{\text{obs}}) | \mu = \mu_j, \forall j \right\}).
  \]
- To improve convergence, we capitalize on \( p(\mu | \psi, Y^{\text{obs}}) \).
PCG Sampler I

- \( p(\mu|\psi, Y^{\text{obs}}) \) is available.
- Partially marginalize over all of the missing data, \((X, Z)\).

\[
\begin{align*}
p(X, Z|\psi, \mu, Y^{\text{obs}}) & \quad \text{Marginalize} \quad p(X, Z|\psi, \mu, Y^{\text{obs}}) \\
p(\psi|X, Z, \mu, Y^{\text{obs}}) & \\
p(\mu|X, Z, \psi, Y^{\text{obs}}) & \\
\end{align*}
\]

\[
\begin{align*}
p(X, Z, \mu|\psi, Y^{\text{obs}}) & \quad \text{Trim} \quad p(\mu|\psi, Y^{\text{obs}}) \\
p(X, Z|\psi, \mu, Y^{\text{obs}}) & \\
p(\psi|X, Z, \mu, Y^{\text{obs}}) & \\
\end{align*}
\]
PCG Sampler I

- $p(\mu|\psi, Y^{\text{obs}})$ is available.
- Partially marginalize over all of the missing data, $(X, Z)$.

$$
\begin{align*}
    \text{Marginalize} & \quad p(X, Z|\psi, \mu, Y^{\text{obs}}) \\
    \text{Marginalize} & \quad p(\psi|X, Z, \mu, Y^{\text{obs}}) \\
    \text{Marginalize} & \quad p(\mu|X, Z, \psi, Y^{\text{obs}})
\end{align*}
$$

$$
\begin{align*}
    \text{Trim} & \quad p(X, Z, \mu|\psi, Y^{\text{obs}}) \\
    \text{Trim} & \quad p(\psi|X, Z, \mu, Y^{\text{obs}})
\end{align*}
$$

The resulting sampler corresponds to a **blocked** version of the original Gibbs sampler!
\( p(\mu | X, \psi, Y^{obs}) \) is also available and quick to compute.

Partially marginalize over part of the missing data, \( Z \).

\[
\begin{align*}
p(X, Z | \psi, \mu, Y^{obs}) & \quad \text{Marginalize} \quad p(X, Z | \psi, \mu, Y^{obs}) \\
p(\psi | X, Z, \mu, Y^{obs}) & \quad \quad \quad p(\psi | X, Z, \mu, Y^{obs}) \\
p(\mu | X, Z, \psi, Y^{obs}) & \quad \quad \quad p(Z, \mu | X, \psi, Y^{obs})
\end{align*}
\]

Permute
\[
\begin{align*}
p(Z, \mu | X, \psi, Y^{obs}) & \quad \text{Trim} \quad p(\mu | X, \psi, Y^{obs}) \\
p(X, Z | \psi, \mu, Y^{obs}) & \quad \quad \quad p(X, Z | \psi, \mu, Y^{obs}) \\
p(\psi | X, Z, \mu, Y^{obs}) & \quad \quad \quad p(\psi | X, Z, \mu, Y^{obs})
\end{align*}
\]

The resulting sampler cannot be blocked! It is the partially collapsed Gibbs sampler.
$p(\mu | X, \psi, Y^{obs})$ is also available and quick to compute.

Partially marginalize over part of the missing data, $Z$.

\[
\begin{align*}
    p(X, Z | \psi, \mu, Y^{obs}) \\
p(\psi | X, Z, \mu, Y^{obs}) \\
p(\mu | X, Z, \psi, Y^{obs}) \\
\end{align*}
\]

\[
\begin{align*}
    \text{Marginalize} \\
    p(X, Z | \psi, \mu, Y^{obs}) \\
p(\psi | X, Z, \mu, Y^{obs}) \\
p(Z, \mu | X, \psi, Y^{obs}) \\
\end{align*}
\]

\[
\begin{align*}
    \text{Permute} \\
    p(Z, \mu | X, \psi, Y^{obs}) \\
p(X, Z | \psi, \mu, Y^{obs}) \\
p(\psi | X, Z, \mu, Y^{obs}) \\
\end{align*}
\]

\[
\begin{align*}
    \text{Trim} \\
    p(\mu | X, \psi, Y^{obs}) \\
p(X, Z | \psi, \mu, Y^{obs}) \\
p(\psi | X, Z, \mu, Y^{obs}) \\
\end{align*}
\]

The resulting sampler cannot be blocked! It is the partially collapsed Gibbs sampler.
Computational Gains

- Compare the Gibbs sampler, PCG I, and PCG II.
- The Gibbs sampler does not move from its starting value.
- PCG I has **better convergence characteristics** than PCG II.
- However, each iteration of PCG I is **more expensive**.

![Graphs comparing Gibbs sampler, PCG I, and PCG II](image-url)
To illustrate the statistical properties of fitted spectra, we run a simulation.

- We generate 10 data sets (1500 counts) from 6 spectra.
- We use a typical continuum, effective area, and instrumental response function.
- There are 0, 1, or 2 lines.
- Each line is of moderate width (SD = 4 bins) or of broad width (SD = 21 bins), and weak or strong.
- Fitted models include one or two emission lines.
Run PCG samplers for model fitting.

Results with one delta function line in the fitted model.

The marginal posterior density is estimated using Gaussian kernel smoothing (SD = 0.01).

Smooth out lower posterior probabilities while not flattening higher posterior probabilities.

*The posterior density is highly multimodal.*
The highly multimodal posterior distribution cannot be summarized in standard ways familiar to astronomers (e.g., $5 \pm 2$ or, for asymmetric intervals, $5^{+3}_{-1}$).

Use a transformation of the posterior density function to visualize the HPD regions of varying levels.

Construct a HPD region for each of the 100 different levels equally spaced between 1% and 100%.

Summarize the highly multimodal posterior distribution using a HPD distribution with 100 different levels.
The multimodal posterior distribution may suggest multiple lines.

The joint posterior distribution of two line locations with data generated under Case 2 (one narrow weak line at 2.85 keV).
The joint posterior distribution of two line locations with data generated under Case 5 (one narrow weak line at 2.85 keV and one narrow strong line at 1.2 keV).
Consider a simple Gaussian model with known SD:

$$Y \sim N(\mu, \sigma)$$

A 95% CI for $\mu$ is given by $Y \pm 1.96\sigma$.

- Misspecification of $\sigma$ with $\varsigma < \sigma$ results in a shorter interval, with lower coverage.

Misspecification of the width of a spectral line:

- Might there be an advantage of using a delta function line rather than a Gaussian line (with fitted width) if we know the spectral line is not too broad?
- The tradeoff is not as simple as in the simple Gaussian model, so we return to the simulation study.
Fitting Narrow Emission Lines in X-ray Spectra

More Precise Inference?

**Delta Function Line**

**Gaussian Line**

Taeyoung Park <tpark@pitt.edu>

Fitting Narrow Emission Lines in X-ray Spectra
## A Closer Look

<table>
<thead>
<tr>
<th>Case</th>
<th>Line type¹</th>
<th>Delta function line</th>
<th>Gaussian line</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coverage²</td>
<td>Mean length</td>
<td>Coverage²</td>
</tr>
<tr>
<td>1</td>
<td>no lines</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>one narrow</td>
<td>90%</td>
<td>100%</td>
</tr>
<tr>
<td>3</td>
<td>one wide</td>
<td>50%</td>
<td>100%</td>
</tr>
<tr>
<td>4</td>
<td>one narrow</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>5</td>
<td>two narrow</td>
<td>90%</td>
<td>90%</td>
</tr>
<tr>
<td>6</td>
<td>two narrow</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>total</td>
<td>narrow</td>
<td>95%</td>
<td>97.5%</td>
</tr>
</tbody>
</table>

¹ Narrow lines are 17 bins wide (four SDs); wide lines are 85 bins wide.
² Coverage: % of ten 95% HPD regions containing at least one true line location

---

Misspecification appears to give better mean length in all cases and better coverage for narrow lines
Proceed with Caution

- Exhaustive simulations are difficult.
- Fitting involves MCMC sampling, which is slow and requires some supervision.
- Results may depend on the line location, line strength, line width, characteristics of the continuum, sample size, etc.

Delta functions emission lines are useful for exploratory data analysis and for inference when the true line is believed to be narrow, and also show promise for use for inference with moderately broad lines.
Six Observations were independently observed for PG1634+706.

\[
p(\mu | \gamma_{\text{obs}}) \propto \prod_{i=1}^6 p(\mu, \psi_i | \gamma_{\text{obs}}^i) d\psi_1 \cdots d\psi_6
\]

\[
= \prod_{i=1}^6 p(\mu, \psi_i | \gamma_{\text{obs}}^i) d\psi_i
\]

\[
= \prod_{i=1}^6 p(\mu | \gamma_{\text{obs}}^i).
\]
**Results**

*Under the delta function emission line model:*

- Given all six observations, the posterior mode of the line location is identified at **2.865 keV**.
- The nominal 95% posterior region consists of (2.83 keV, 2.92 keV) with 94.8% and (0.50 keV, 0.51 keV) with 2.2%.
- The detected line is red-shifted to **6.69 keV** in the quasar rest frame, which indicates the ionization state of iron.
Quantify the evidence of the emission line in the spectrum using ppp-values (Meng, 1994; Protassov et al, 2002).

Compare the following three models:

- **MODEL 0**: No emission line in the spectrum.
- **MODEL 1**: One line at 2.74 keV with unknown intensity.
- **MODEL 2**: One line with unknown location and intensity.

The test statistic is the sum of the log likelihood ratios:

\[
T_m(\tilde{Y}^{(\ell)}) = \sum_{i=1}^{6} \log \left\{ \frac{\sup_{\theta \in \Theta_m} L(\theta | \tilde{Y}_i^{(\ell)})}{\sup_{\theta \in \Theta_0} L(\theta | \tilde{Y}_i^{(\ell)})} \right\}, \quad m = 1 \text{ and } 2,
\]

- \(\Theta_m\) : the parameter space under **MODEL m**.
- \(\tilde{Y}^{(\ell)}\) : the simulated data sets under **MODEL 0**.
The posterior predictive method is a parameterized bootstrap that accounts for posterior uncertainty in the parameters.

Small ppp-values give strong evidence for the presence of the delta function emission line in the X-ray spectrum.

Model 0 vs. Model 1

Model 0 vs. Model 2

ppp-value = 0.006

$T_1(Y_{\text{obs}}) = 7.450$

ppp-value = 0.042

$T_2(Y_{\text{obs}}) = 23.002$
Estimating the location of narrow emission lines causes problems for standard statistical methods and algorithms.

Devise partially collapsed Gibbs (PCG) samplers that

– generalize the composition of conditional distributions,
– generalize a blocked Gibbs sampler (Liu et al., 1995),
– are viewed as a stochastic counterpart to

  * the ECME algorithm (Liu and Rubin, 1994) and
  * the AECM algorithm (Meng and van Dyk, 1997).

Summarize the highly multimodal posterior distribution using a HPD distribution with varying levels.

Use posterior predictive checks (Protassov et al., 2002) to test for the evidence of a narrow emission line.
Thank you!