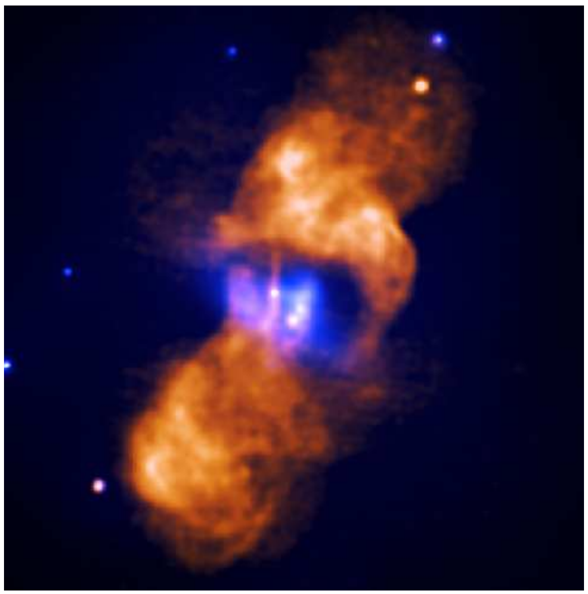
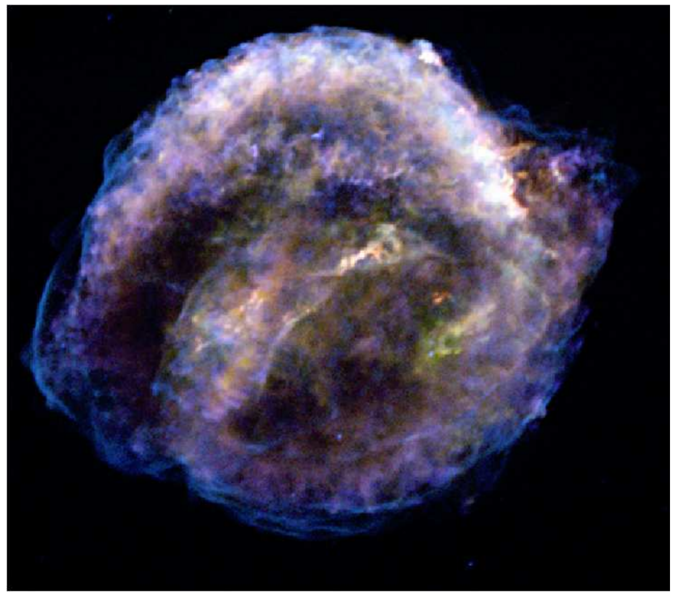
Spectral Analysis of Faint Astronomical Objects: Bayesian Modeling, Computation, and Inference

David A. van Dyk

Department of Statistics
University of California, Irvine

Joint work with
Taeyoung Park, Aneta Siemiginowska, and CHASC





Outline of Presentation

This talk has two components:

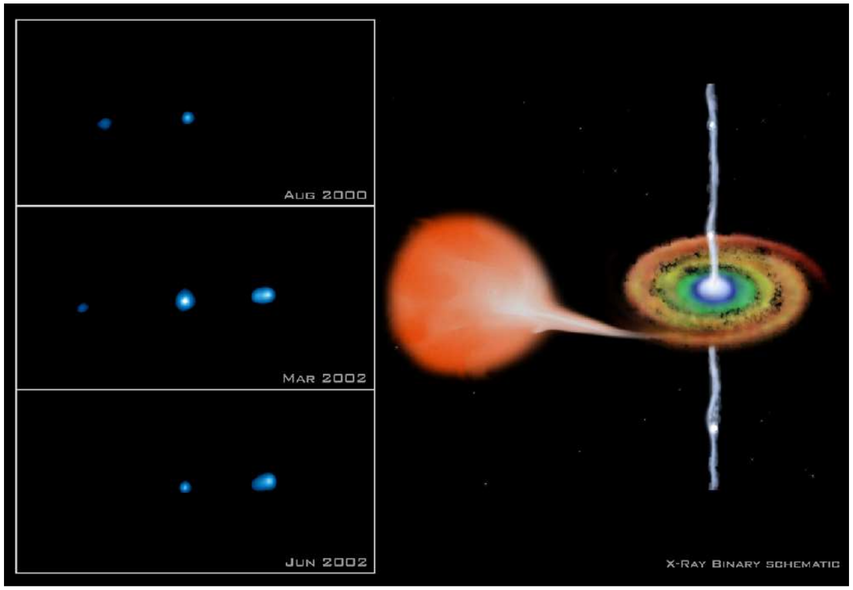
A. **Highly Structured Models in High-Energy Astrophysics**

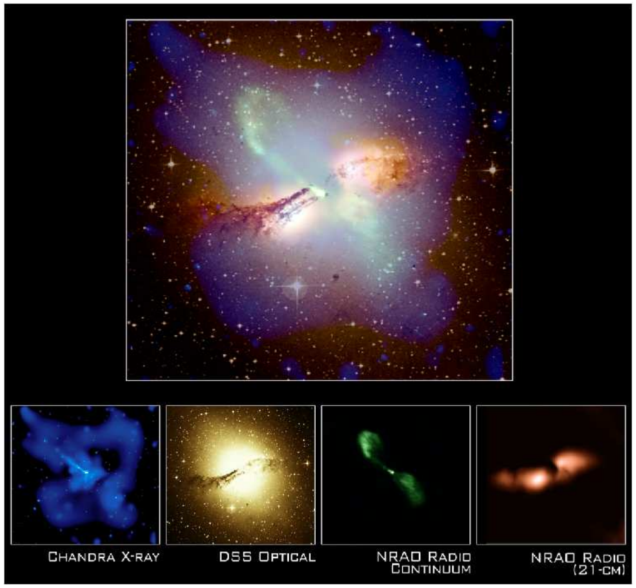
1. **Astrostatistics:**
   Complex Sources, Complex Instruments, and Complex Questions
   Key: All three are the domain of Astrostatistics
2. Spectral Analysis:
   Model-Based Statistical Solutions
   Monte Carlo-Based Bayesian Analysis

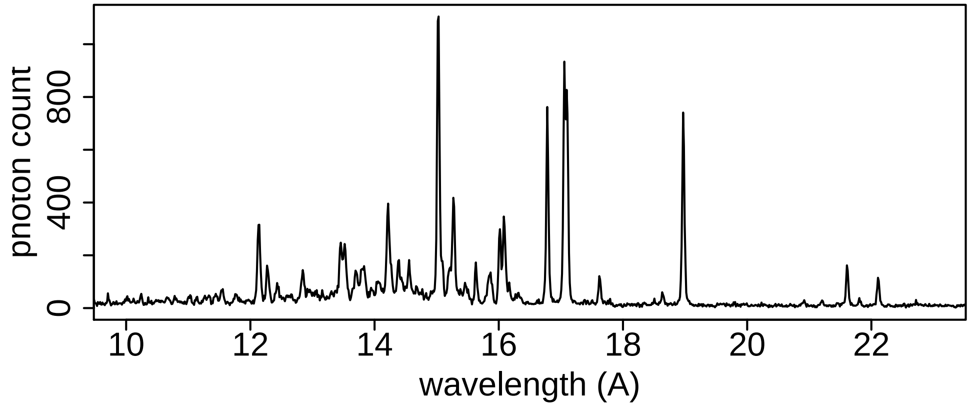
B. Looking for Narrow Emission Lines in High-Energy Spectra

1. Highly-Multimodal Likelihood
2. An Advantage of Model Misspecification?
3. Formal Tests
4. Statistical Computational: From EM to Incompatible Gibbs Samplers

Complex Astronomical Sources

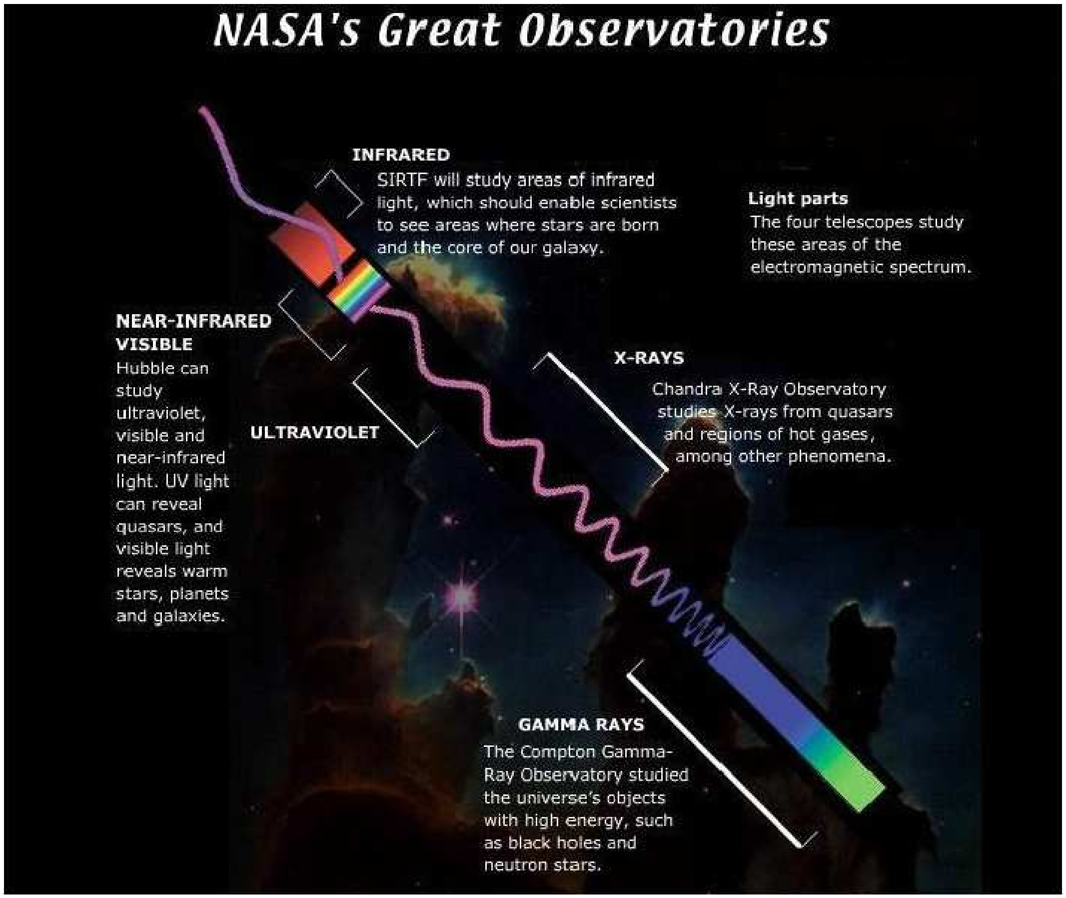






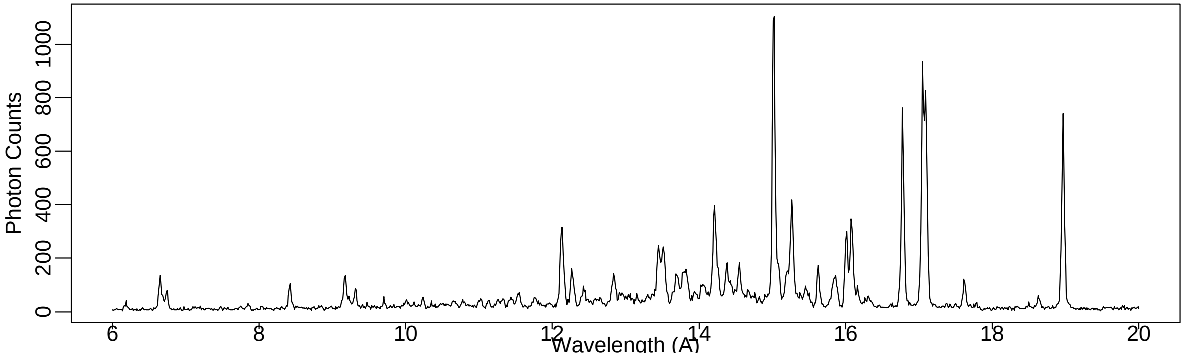
Images may exhibit Spectral, Temporal, and Spatial Characteristics.

Astrostatistics: Complex Data Collection



- A very small sample of instruments
- Earth-based, survey, interferometry, etc.
- X-ray alone: at least four planned missions
- Instruments have different data-collection mechanism

Astrostatistics: Complex Questions



- What is the composition and temperature structure?

Scientific Context

The Chandra X-Ray Observatory

- Chandra produces images at least thirty times sharper then any previous X-ray telescope.
- X-rays are produced by multi-millions degree matter, e.g., by high magnetic fields, extreme gravity, or explosive forces.
- Images provide understand into the hot and turbulent regions of the universe.

Unlocking this information requires subtle analysis:

The California Harvard AstroStatistics Collaboration (CHASC)

- van Dyk, et al. (*The Astrophysical Journal*, 2001)
- Protassov, et al. (*The Astrophysical Journal*, 2002)
- van Dyk and Kang (*Statistical Science*, 2004)
- Esch, Connors, van Dyk, and Karovska (*The Astrophysical Journal*, 2004)
- van Dyk et al. (*Bayesian Analysis*, 2006)
- Park et al. (*The Astrophysical Journal*, 2006 & under review)

Data Collection

Data is collected for each arriving photon:

- the (two-dimensional) sky coordinates,
- the energy, and
- the time of arrival.

All variables are discrete:

- High resolution $\longrightarrow$ finer discretization.
  e.g., 4096×4096 spatial and 1024 spectral bins

The four-way table of photon counts:

- Spectral analysis models the one-way energy table;
- Spatial analysis models the two-way table of sky coordinates; and
- Timing analysis models the one-way arrival time table

The Image: A moving 'colored' picture

Kepler's Supernova Reminant

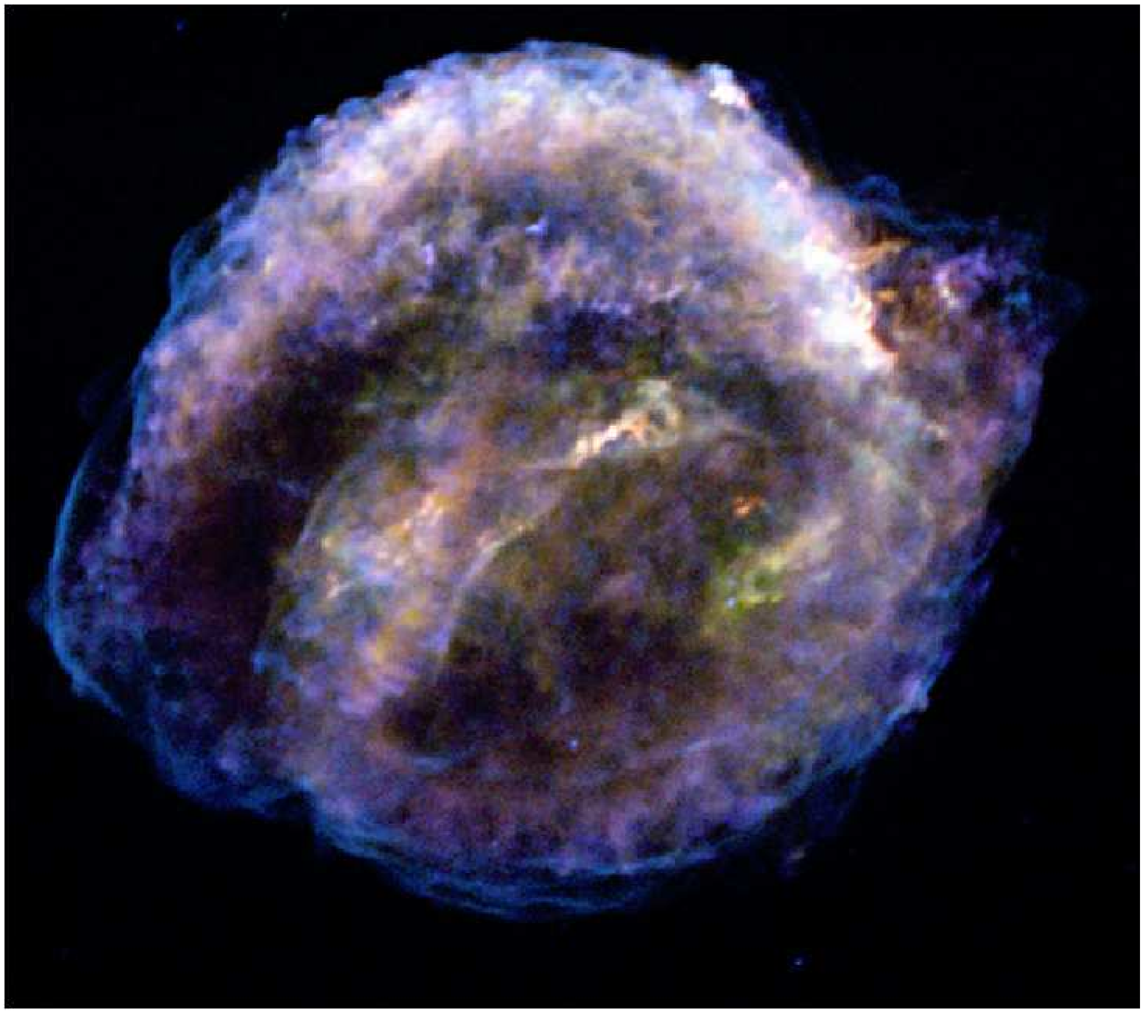


Image Credits" NASA/CXC/NCSU/S.Reynolds et al.

Outline of Presentation

This talk has two components:

A. Highly Structured Models in High-Energy Astrophysics

1. Astrostatistics:

   Complex Sources, Complex Instruments, and Complex Questions

   Key: All three are the domain of Astrostatistics

2. **Spectral Analysis:**

   Model-Based Statistical Solutions

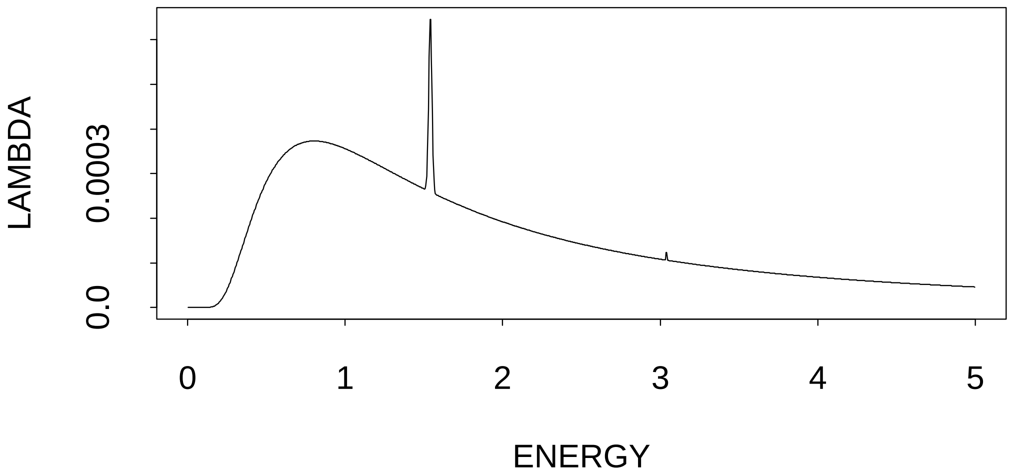
   Monte Carlo-Based Bayesian Analysis

B. Looking for Narrow Emission Lines in High-Energy Spectra

1. Highly-Multimodal Likelihood
2. An Advantage of Model Misspecification?
3. Formal Tests
4. Statistical Computational: From EM to Incompatible Gibbs Samplers

The Basic Spectral Models

- Photon counts modeled with Poisson process.
- The Poisson parameter is a function of energy, with two basic components:
  1. The *continuum,* a GLM for the baseline spectrum (e.g., $\alpha E^{-\beta}$),
  2. Several *emission lines*, a mixture of Gaussians added to the continuum.
  3. Several *absorption lines* multiply by the continuum.
  4. The continuum indicates the temperature of the source while the emission and absorption lines gives clues as to the relative abundances of elements



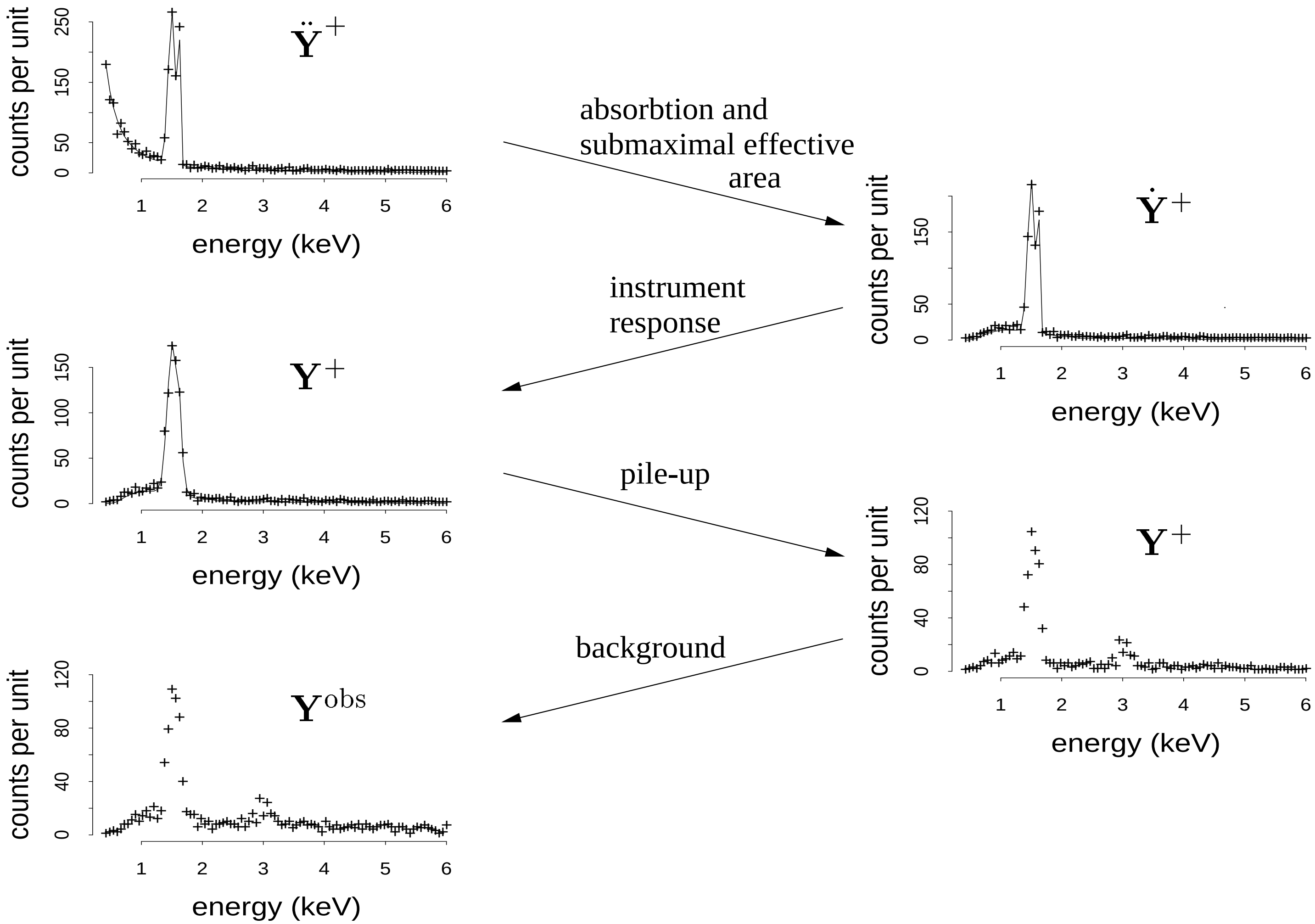
Physics of a Hot Plasma

- Many astronomical sources are made up of very hot plasma ($> 10^6$K).
- Ions are in an excited state: The electrons populate higher energy states.
- An (inelastic) collision of two ions:
  - The ions slow down;
  - In order to preserve energy, electrons jump to higher energy states;
  - Ions spontaneously decay to a lower more stable energy state; and
  - The difference in energy between the two states is emitted in the form of a photon.
- The energy difference is unique to the state transition of a particular ion.
- The frequency of a particular state transition is informative as to the temperature and density of the source.

A line can be identified with a particular ion, and thus we obtain information on the environment of the astronomical source.

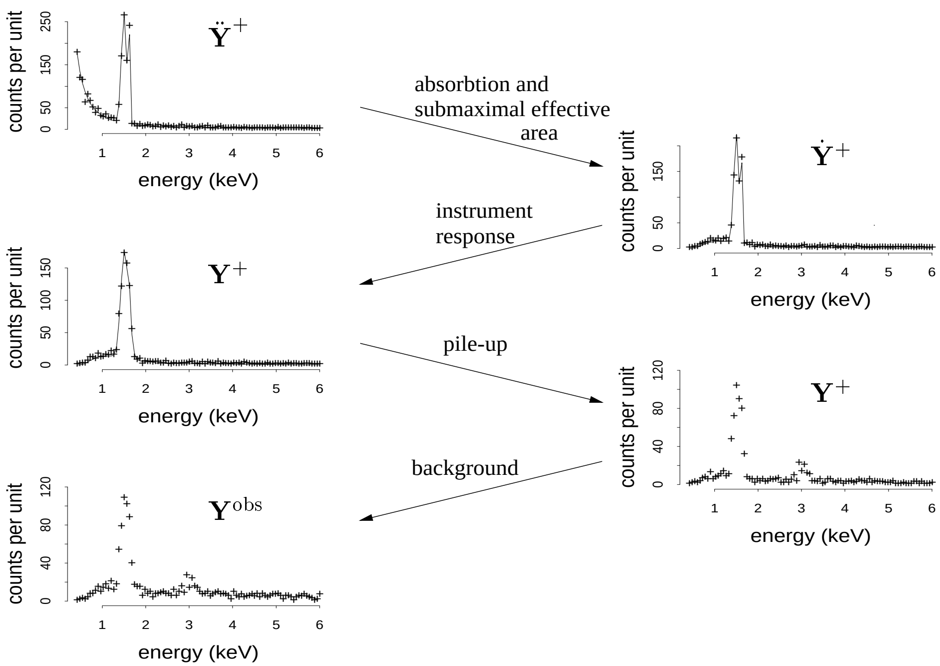
Highly Structured Models

Modeling the *Chandra* data collection mechanism.



Highly Structured Models

Modeling the *Chandra* data collection mechanism.



- The method of Data Augmentation: EM algorithms and Gibbs samplers.
- We can separate a complex problem into a sequence of problems, each of which is easy to solve.

We wish to directly model the sources and data collection mechanism and use statistical procedures to fit the resulting highly-structured models and address the substantive scientific questions.

A Model-Based Statistical Paradigm

1. Model Building
   - Model source spectra, image, and/or time series
   - Model the data collection process
     - background contamination
     - instrument response
     - effective area and absorption
     - pile up
   - Results in a highly structured hierarchical model
2. Model-Based Statistical Inference
   - Bayesian posterior distribution
   - Maximum likelihood estimation
3. Sophisticated Statistical Computation Methods Are Required
   - Goals: computational stability and easy implementation
   - Emphasize natural link with models: *The Method of Data Augmentation*

<u>What are Prior distributions?</u>

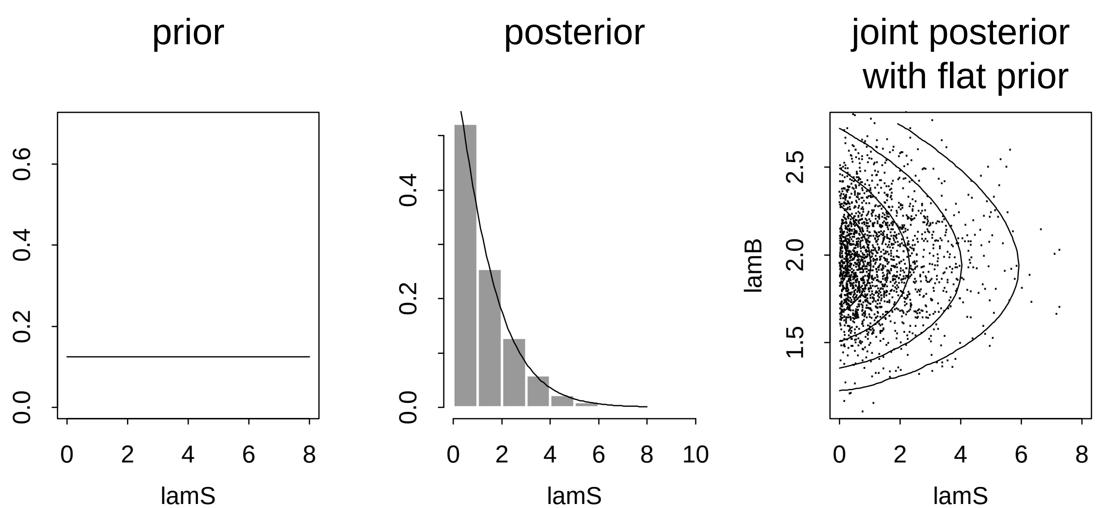
1. Priors can be used
   - to incorporate information from outside the data, or
   - to impose structure.
2. Priors offer a principled compromise between "fixing" a parameter & letting it "float free".
3. Setting `min` and `max` limits in XSPEC amounts to using a flat prior over a specified range.

Bayesian Inference Using Monte Carlo

The Building Block of Bayesian Analysis

1. The sampling distribution: $p(Y|\psi)$.
2. The prior distribution: $p(\psi)$.
3. Bayes theorem and the posterior distribution: $p(\psi|Y) \propto p(Y|\psi)p(\psi)$

Inference Using a Monte Carlo Sample:



We use *MCMC* (e.g., the *Gibbs Sampler*) to obtain the Monte Carlo sample.

Bayesian Deconvolution

- The Data Collection Mechanism

Blurring Matrix
* known from calibration

Source Model

$$\boldsymbol{\lambda} = \mathbf{P}\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\xi},$$

Expected Photon Count

Non−Homogeneous Stochastic Censoring
* known from calibration

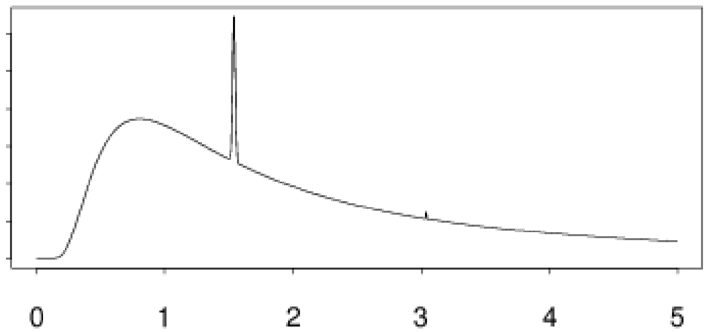
Background Contamination
* often fit using background observation

The observed counts are modeled as independent Poisson variables with means given by $\boldsymbol{\lambda}$.

The Source Models

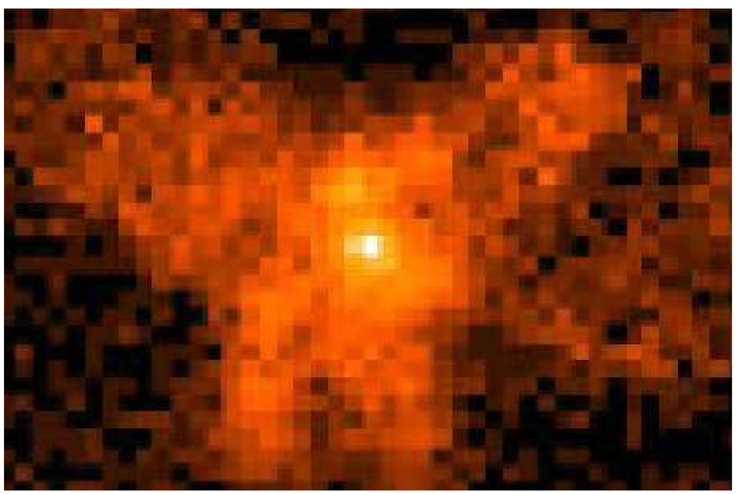
Parameterized finite mixture models

(source models w/ several components)



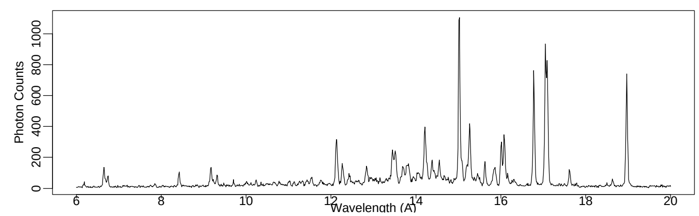
Smoothing prior distributions

(Multiscale models for diffuse emission)



Compound deconvolution models

(simultaneous instrumental & physical "deconvolution" of complex sources)



$$\boldsymbol{\lambda} = \mathbf{P_1 A_1} \left(\mathbf{P_2 A_2} \boldsymbol{\mu}_T \right) + \boldsymbol{\xi}$$

Outline of Presentation

This talk has two components:

A. Highly Structured Models in High-Energy Astrophysics

1. Astrostatistics:

   Complex Sources, Complex Instruments, and Complex Questions

   Key: All three are the domain of Astrostatistics

2. Spectral Analysis:

   Model-Based Statistical Solutions

   Monte Carlo-Based Bayesian Analysis

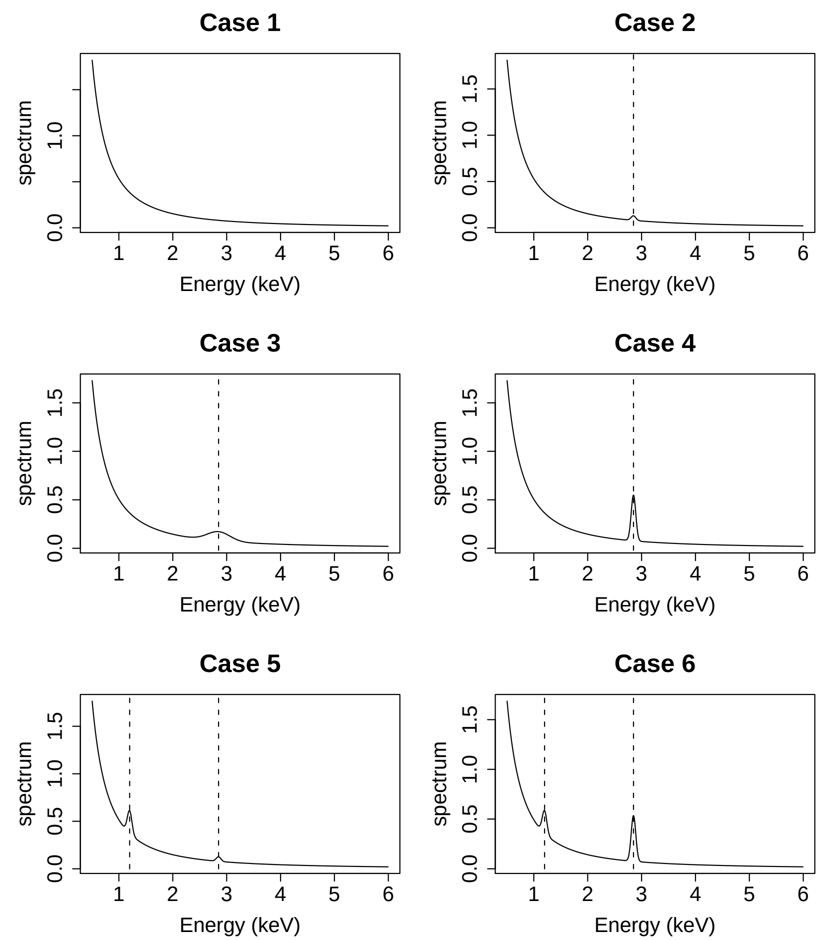
B. **Looking for Narrow Emission Lines in High-Energy Spectra**

1. **Highly-Multimodal Likelihood**
2. An Advantage of Model Misspecification?
3. Formal Tests
4. Statistical Computational: From EM to Incompatible Gibbs Samplers

A Simulation Study

To illustrate the statistical properties of our fitted spectra, we ran a simulation.

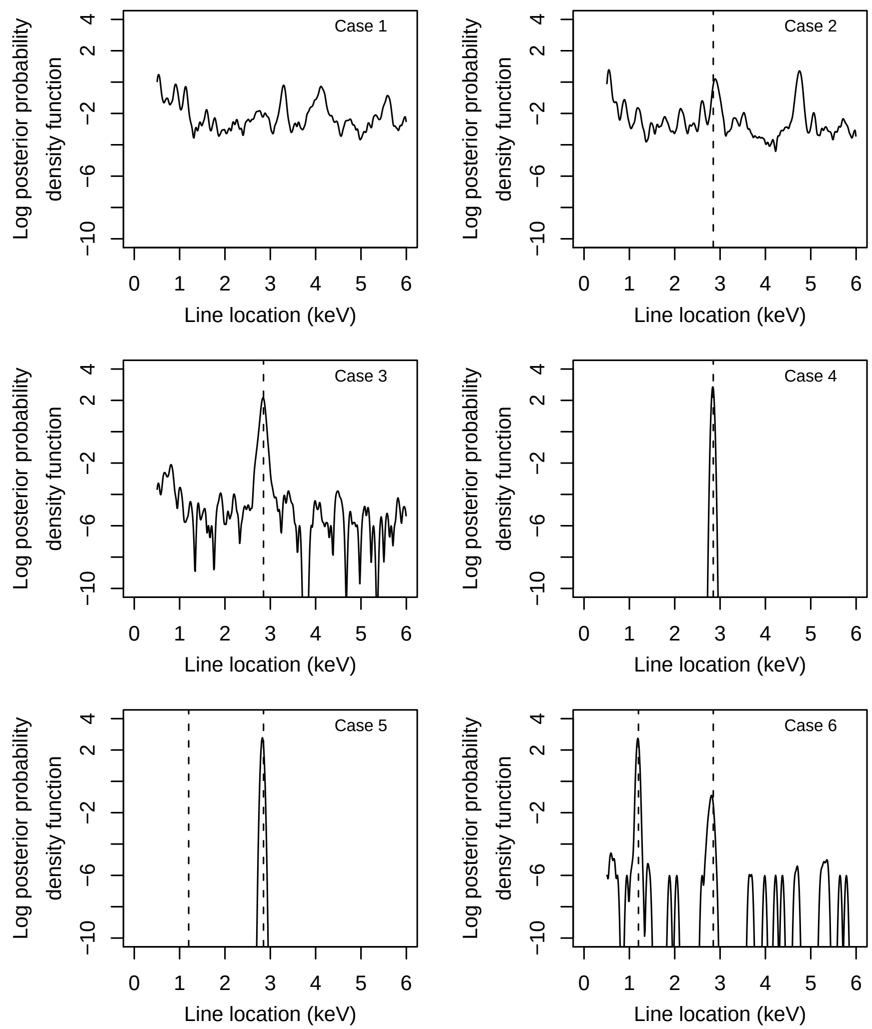
- We generated 10 data sets (1500 cnts) from 6 spectra.
- We use a typical continuum, effective area, & instrument response function.
- There were 0, 1, or 2 lines.
- Each line was either narrow or wide and weak or strong.
- Fitted models included one or two emission lines.
- We used both Gaussian lines (fit location, width, and height) and delta functions (fit location and height).



Results: Highly Multi-Modal Likelihoods

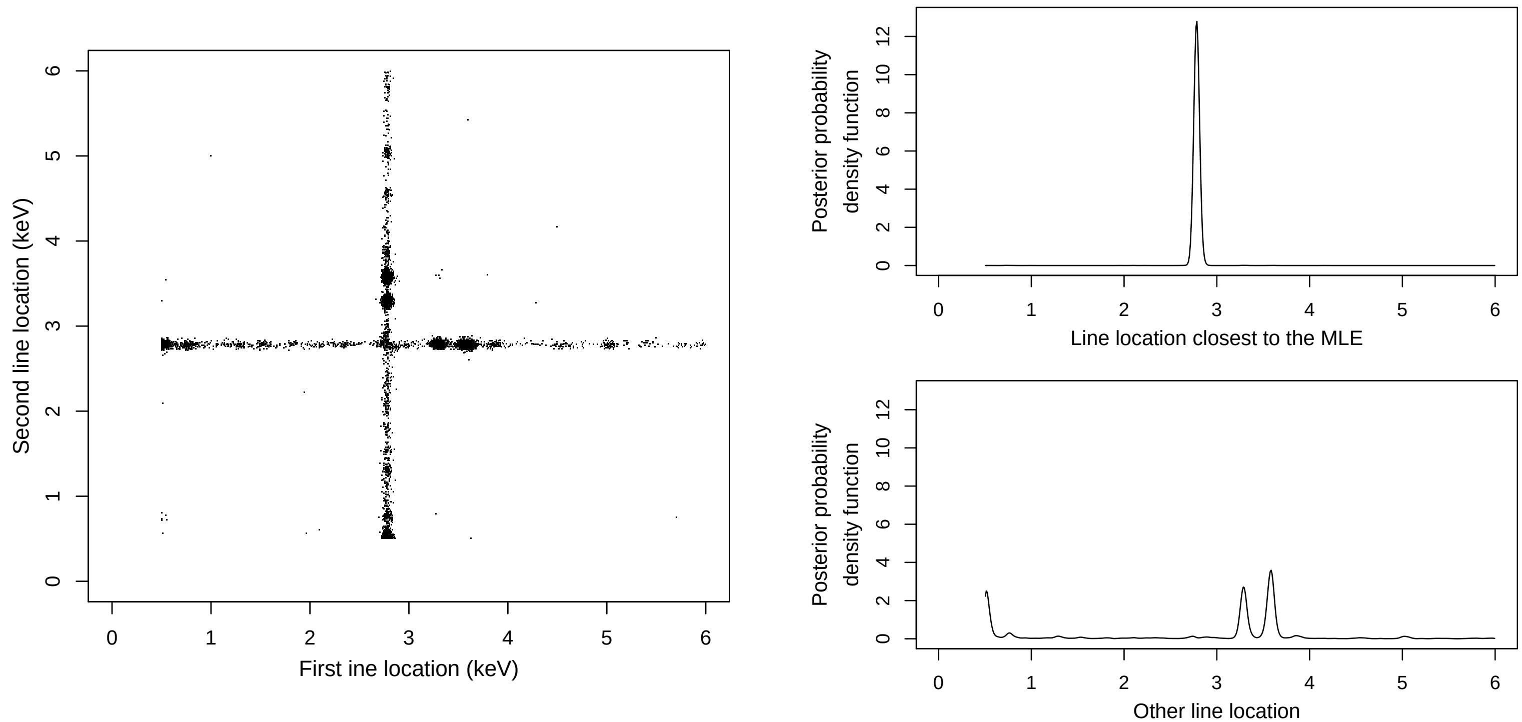
- We ran MCMC samplers for model fitting.
- Results with one delta function line in the fitted model.
- The marginal posterior density is estimated using Gaussian Kernel Smoothing (band width=0.02).

The likelihood function is highly multimodal.



Fitting Two Emission Lines

The joint posterior distribution of two line locations with data generated under Case 2 (one narrow weak line at 2.85 keV).



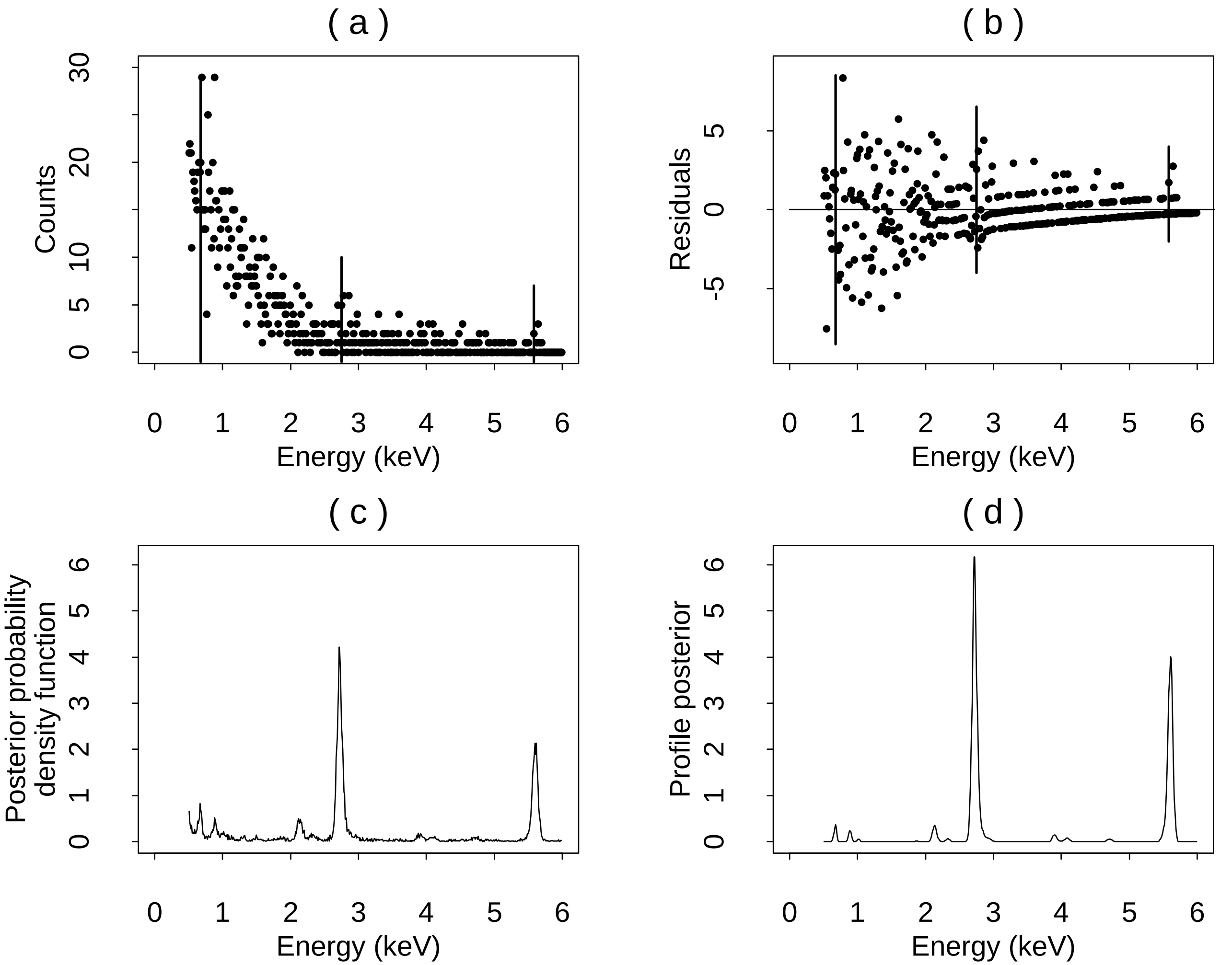
Exploratory Data Analysis

Modes are caused by excess emission in a (narrow) range of energies relative to what we expect from the (fitted) continuum alone.

Searching for Excess Emission:

- Delta functions can identify excess emission in a very narrow energy range.
- We consider using delta functions for *Exploratory Data Analysis*.
- In this stage we are simply looking for candidate line locations.
- The *Profile Posterior Distribution* is a quick to compute summary of the posterior distribution that is well suited to this purpose.

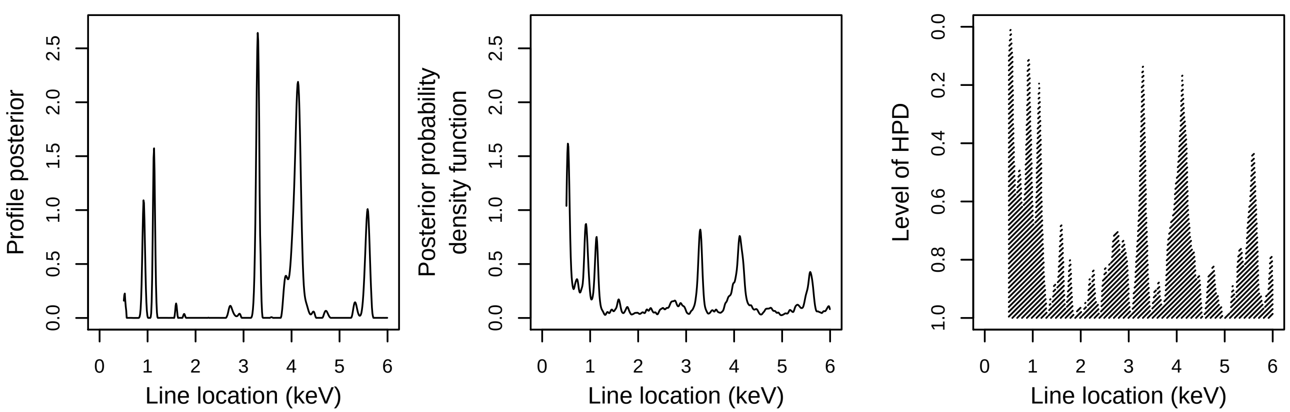
A Data Example



Posterior Regions

Highly multimodal posterior distributions cannot be summarized in standard ways familiar to astronomers (e.g., 5 ± 2 or, for asymmetric intervals, 5^{+3}_{-1}).

We use a transformation of the Posterior Density Function to visualize the HPD regions of varying probability.



Outline of Presentation

This talk has two components:

A. Highly Structured Models in High-Energy Astrophysics

1. Astrostatistics:
   Complex Sources, Complex Instruments, and Complex Questions
   Key: All three are the domain of Astrostatistics
2. Spectral Analysis:
   Model-Based Statistical Solutions
   Monte Carlo-Based Bayesian Analysis

B. Looking for Narrow Emission Lines in High-Energy Spectra

1. Highly-Multimodal Likelihood
2. **An Advantage of Model Misspecification?**
3. Formal Tests
4. Statistical Computational: From EM to Incompatible Gibbs Samplers

Coverage and Interval Length

Consider a simple Gaussian model with know variance:

$$Y \sim \mathrm{N}(\mu, \sigma^2)$$

A 95% CI for μ is given by

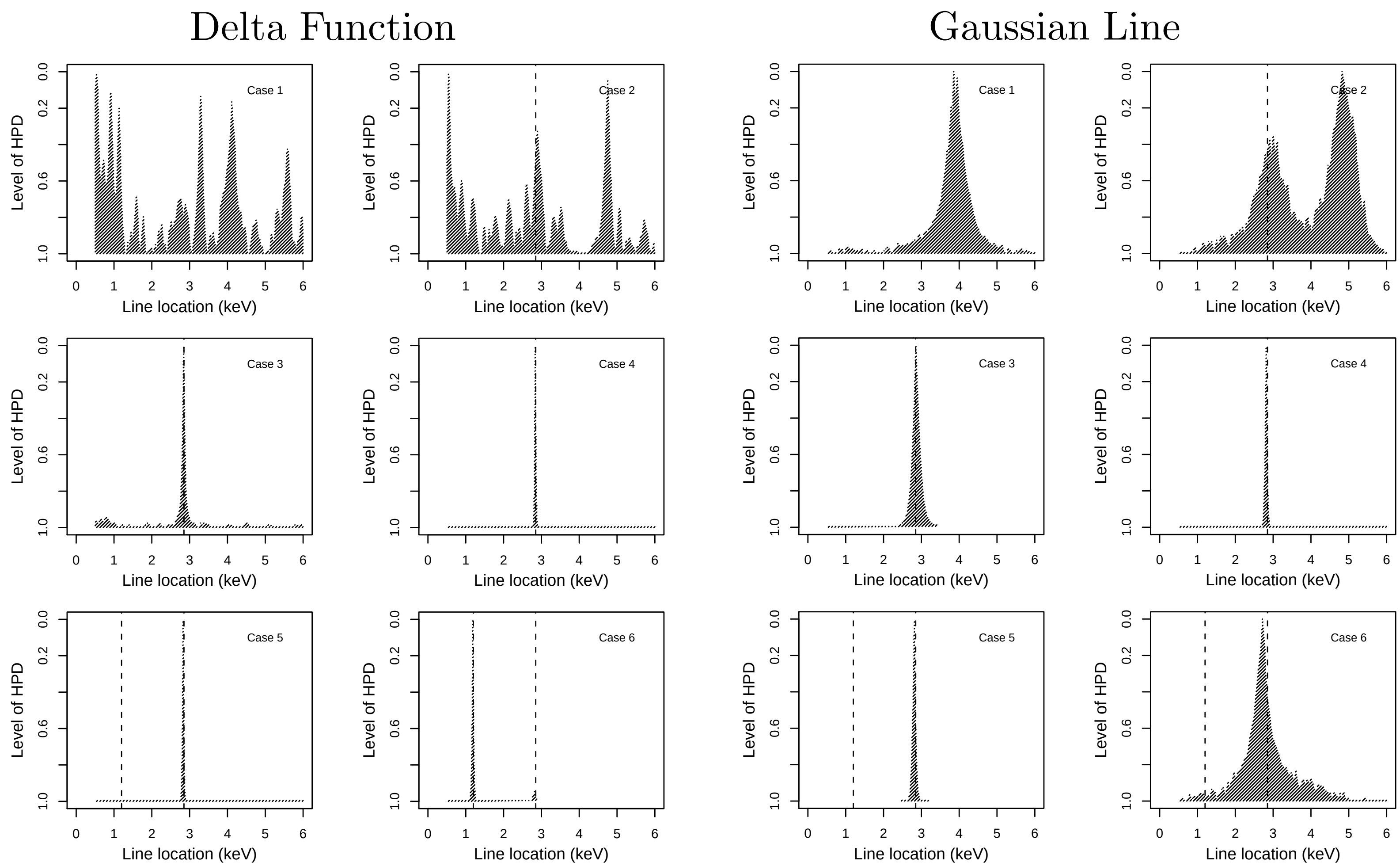
$$Y \pm 1.96 \times \sigma.$$

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

Misspecification of width of a spectral line

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

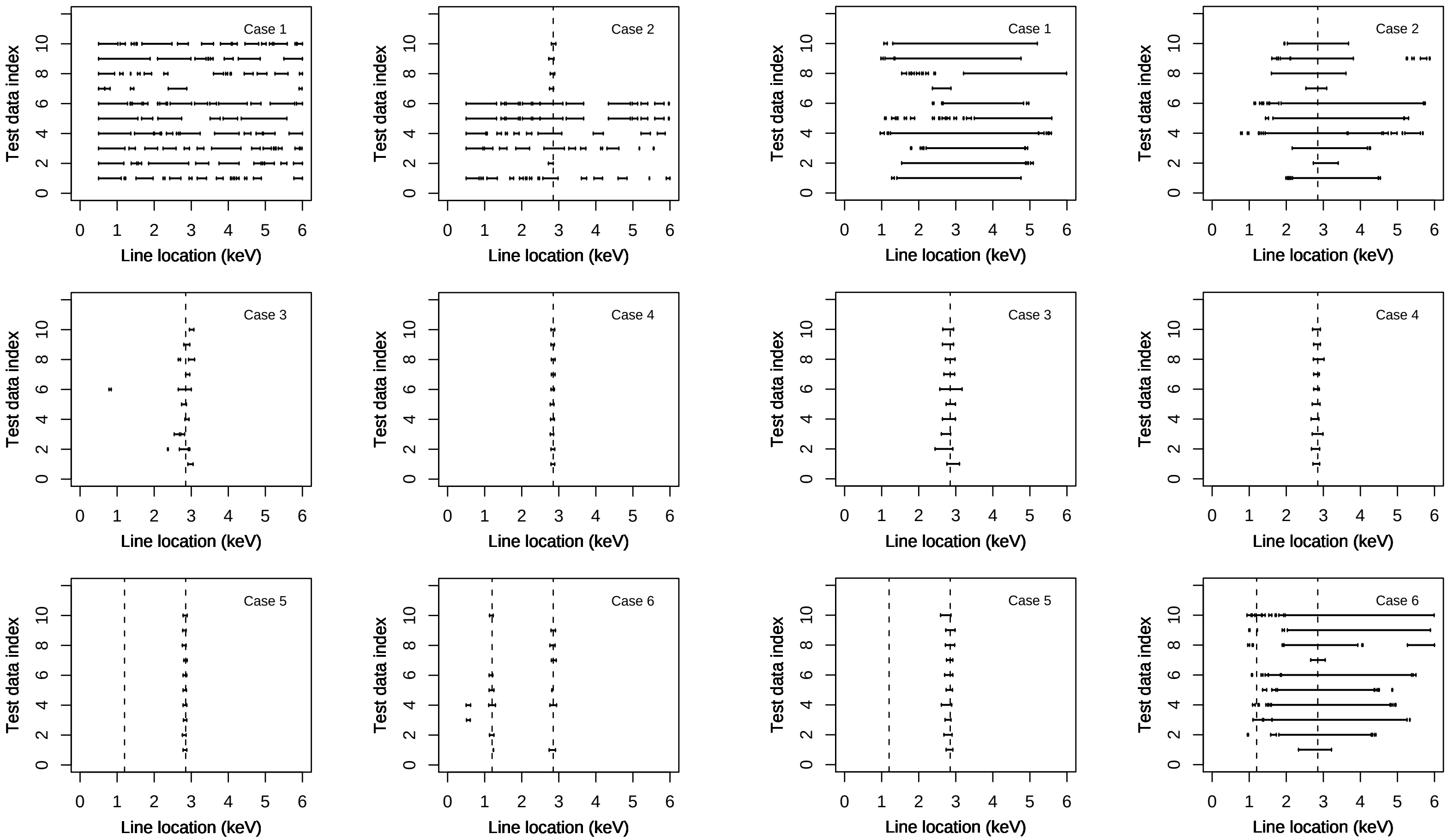
The Simulation Study



The posterior distribution is much more concentrated, but what about coverage?

More Precise Inference?

Delta Function Gaussian Line



Could Model Misspecification lead to more precise Inference?

A Closer Look

		Delta Function Line		Gaussian Line	
Case	Line Type†	Coverage⋆	Mean Width	Coverage⋆	Mean Width
1	no lines	NA	3.30	NA	3.07
2	one narrow	90%	1.56	100%	2.38
3	one wide	60%	0.19	100%	0.34
4	one narrow	100%	0.10	100%	0.21
5	two narrow	100%	0.10	100%	0.22
6	two narrow	90%	0.16	100%	3.01
total for narrow		95%		100%	

† Narrow lines are 17 bins wide (four SDs); wide lines are 85 bins wide.

⋆ Coverage: % of ten 95% HPD regions containing *at least one* true line location.

Misspecification appears to give better width 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, for inference when the true line is believed to be narrow, and show promise for use for inference with moderately wide lines.

Outline of Presentation

This talk has two components:

A. Highly Structured Models in High-Energy Astrophysics

1. Astrostatistics:

   Complex Sources, Complex Instruments, and Complex Questions

   Key: All three are the domain of Astrostatistics

2. Spectral Analysis:

   Model-Based Statistical Solutions

   Monte Carlo-Based Bayesian Analysis

B. Looking for Narrow Emission Lines in High-Energy Spectra

1. Highly-Multimodal Likelihood
2. An Advantage of Model Misspecification?
3. **Formal Tests**
4. Statistical Computational: From EM to Incompatible Gibbs Samplers

Principled P-values to Test for a Model Component

Fallible F-tests

The F-test commonly used by Astronomers to test for spectral lines is a special case (under a *Gaussian* assumption!) of the Likelihood Ratio Test.

- The LRT is not properly calibrated for this use, since the null model in on the boundary of the parameter space.
- Even more troubling, some line parameter are not defined in the null model.
- We conducted a survey of papers in ApJ, ApJL, and ApJS (1995-2001)

Type of Test	Number of Papers
Null Space on Boundary	106
Comparing Non-Nested Models	17
Other Questionable Cases	4
Seemingly Appropriate Use of Test	56

Protassov et al. develops a method based on posterior predictive p-values to properly calibrate a test. This is a parameterized bootstrap that accounts for posterior uncertainty in parameters. **This paper has been cited 103 times.**

Outline of Presentation

This talk has two components:

A. Highly Structured Models in High-Energy Astrophysics

1. Astrostatistics:

   Complex Sources, Complex Instruments, and Complex Questions

   Key: All three are the domain of Astrostatistics

2. Spectral Analysis:

   Model-Based Statistical Solutions

   Monte Carlo-Based Bayesian Analysis

B. Looking for Narrow Emission Lines in High-Energy Spectra

1. Highly-Multimodal Likelihood

2. An Advantage of Model Misspecification?

3. Formal Tests

4. **Statistical Computational: From EM to Incompatible Gibbs**

Generalizing the Gibbs Sampler

- The standard two-step sampler iterates between

$$\psi_1 \sim p(\psi_1|\psi_2) \text{ and } \psi_2 \sim p(\psi_2|\psi_1),$$

to form a Markov chain with stationary distribution

$$p(\psi_1, \psi_2).$$

- Consider a more general form using incompatible conditional distributions:

$$\psi_1 \sim \mathcal{K}(\psi_1|\psi_2) \text{ and } \psi_2 \sim \mathcal{K}(\psi_2|\psi_1)$$

- Questions:
  1. Does the resulting Markov chain have a stationary distribution?
  2. If so, what is it?
  3. Why use such a chain?
- I cannot fully answer these questions, but can offer tantalizing examples....

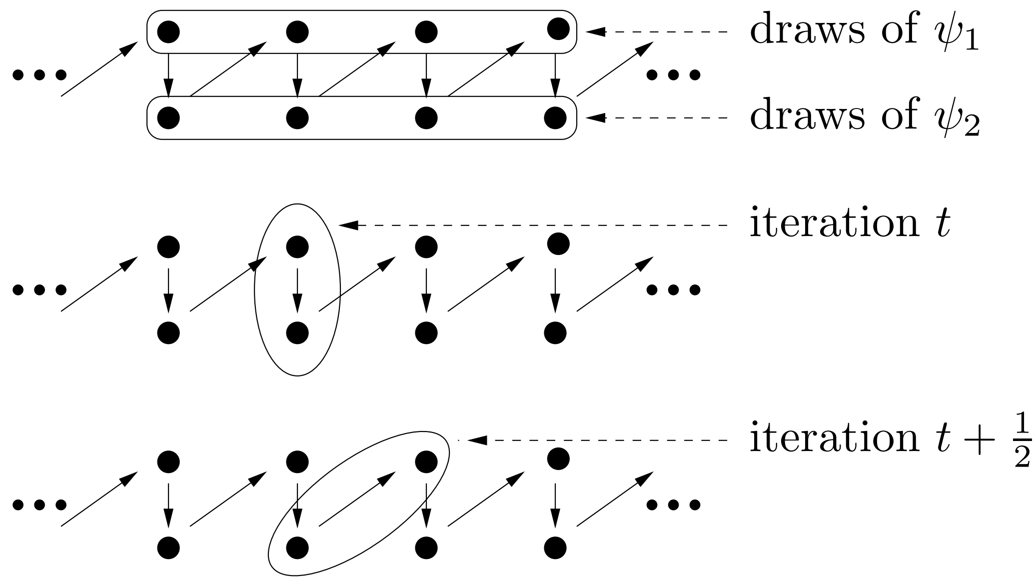
Partially Collapsed Gibbs Sampler, van Dyk & Park, *JASA to appear*

The Simplest Example

A simple 2-step sampler:

STEP 1: $\psi_1^{(t)} \sim p(\psi_1|\psi_2^{(t-1)})$

STEP 2: $\psi_2^{(t)} \sim p(\psi_2)$.



The Markov chain

$\mathcal{M} = \{(\psi_1^{(t)}, \psi_2^{(t)}), t = 0, 1, \ldots\}$

has stationary distribution $p(\psi_1)p(\psi_2)$

- with target margins but
- without the correlation of target distribution,

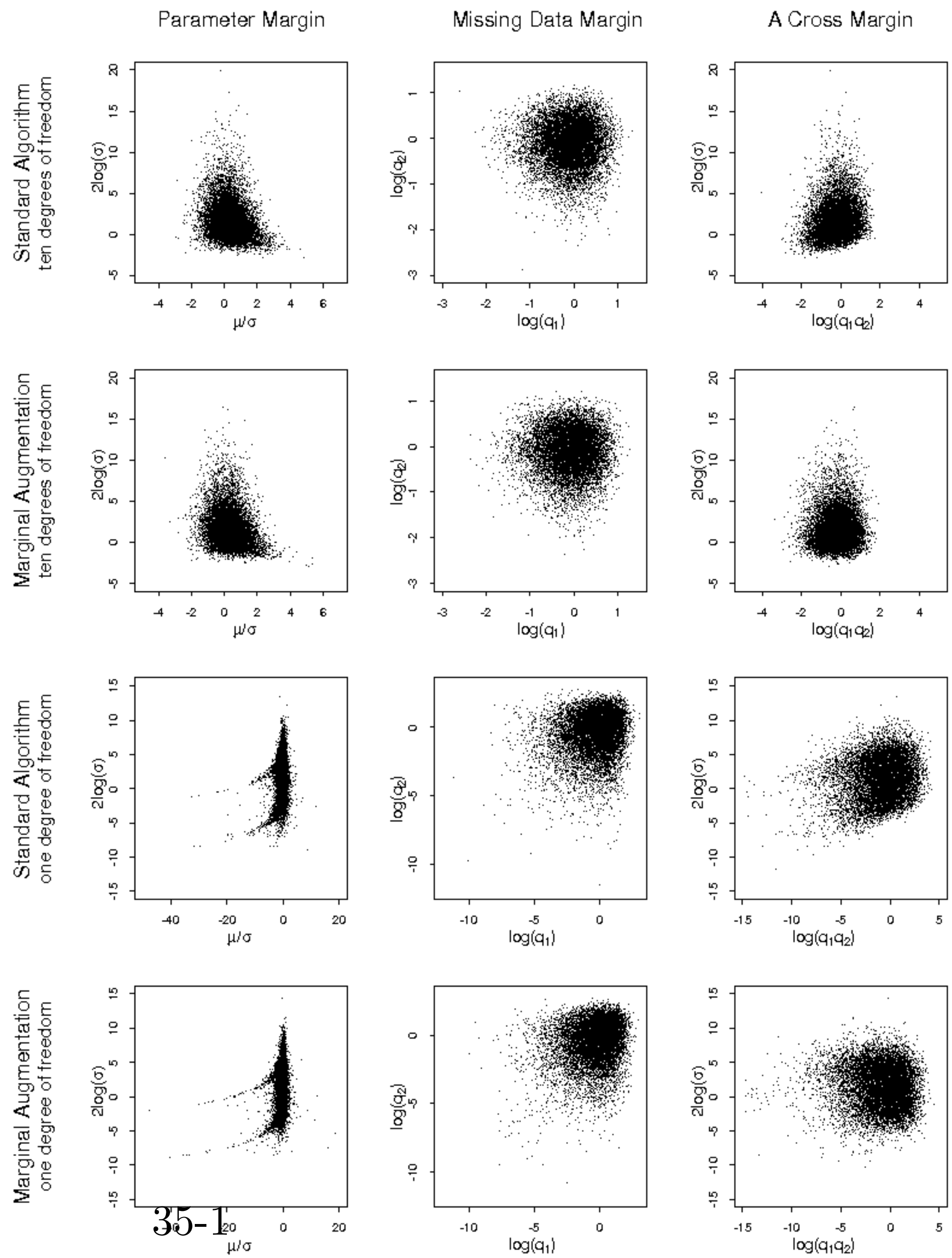
We are mixing the conditional distributions from two different joint distributions with the same marginals. These conditional distributions are incompatible.

AND is "quick" to converge!

We regain the joint target distribution with a one-step shifted chain.

Empirical Illustration with a t model

- *The loss of the correlation structure is our key to success.*
- Two 'data sets' of size two are fit with 10 and 2 degrees of freedom.
- These algorithms are based on the method of *Marginal Augmentation* (Meng and van Dyk, 1999; van Dyk and Meng, 2001).
- Idea: Use both conditionals of the joint distribution with reduced correlation.



Searching for Narrow Lines

- A simplified *latent* Poisson Process for the scientific model,

$$X_i \sim \text{Poisson}\left(\Lambda_i = \alpha E_i^{-\beta} + \lambda^{\mathrm{L}}\pi_i\right).$$

- We sometimes construct a *delta function* emission line model so that
  1. the point source is contained entirely in one pixel, but
  2. we do not know which pixel.

  i.e., $\{\pi_i\}$ can be parameterized in terms of a single unknown parameter,

$$\theta^{\mathrm{L}} = \text{the location of the emission line.}$$

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

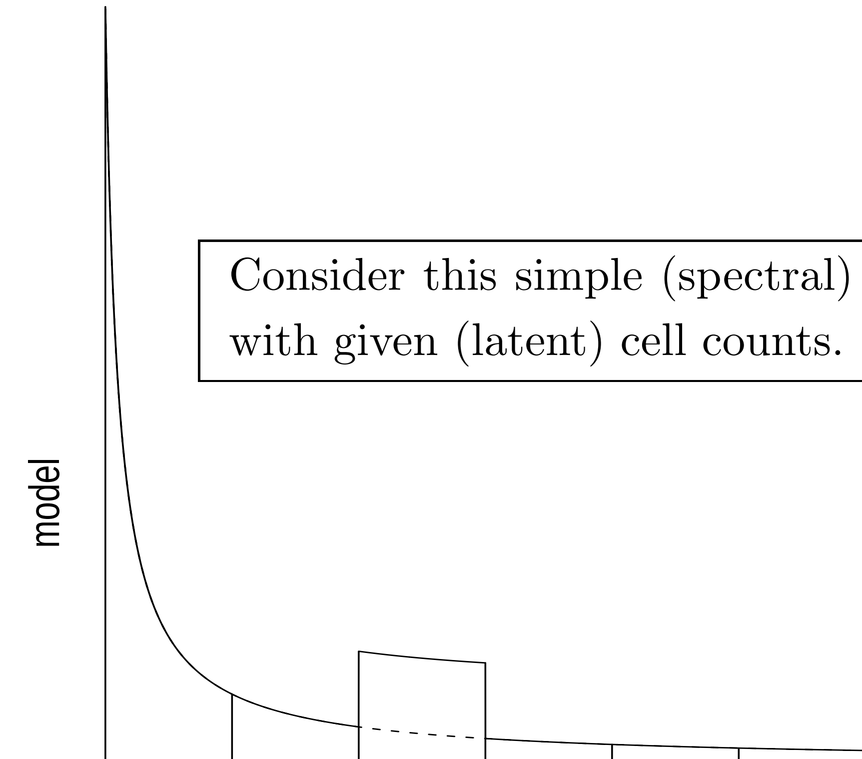
$$Z_{il} = \begin{pmatrix} \text{indicator that photon } l \text{ in bin } i \\ \text{corresponds to the emission line} \end{pmatrix}$$

  1. Given $Z = \{Z_{il}\}$ we can sample $\theta = \{\alpha, \beta, \lambda^{\mathrm{L}}, \theta^{\mathrm{L}}\}$
  2. Given θ we can sample Z, via $Z_{il} \sim \text{Ber}\left(\frac{\lambda^{\mathrm{L}}\pi_i}{\alpha E_i^{-\beta} + \lambda^{\mathrm{L}}\pi_i}\right)$

In This Case the Gibbs Sampler Fails.

Why the Gibbs Sampler Fails

Consider this simple (spectral) model with given (latent) cell counts.

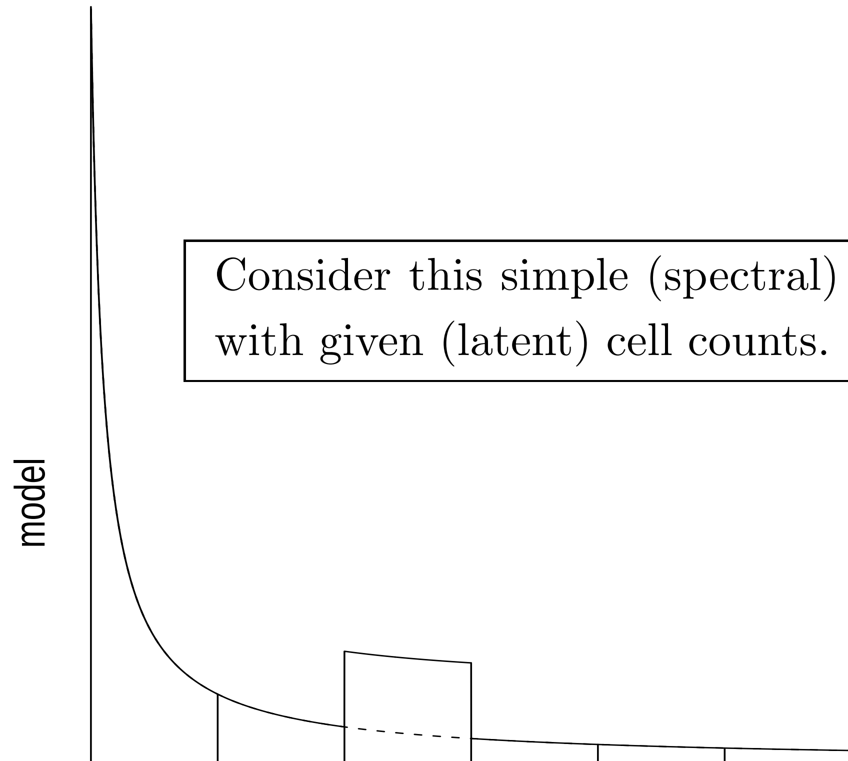


X = (latent) Cell Counts	10	4	8	1	2	0
Extend Src Counts(Z=0)						
Point Src Counts (Z=1)						

Given this Model, what is Z?

Why the Gibbs Sampler Fails

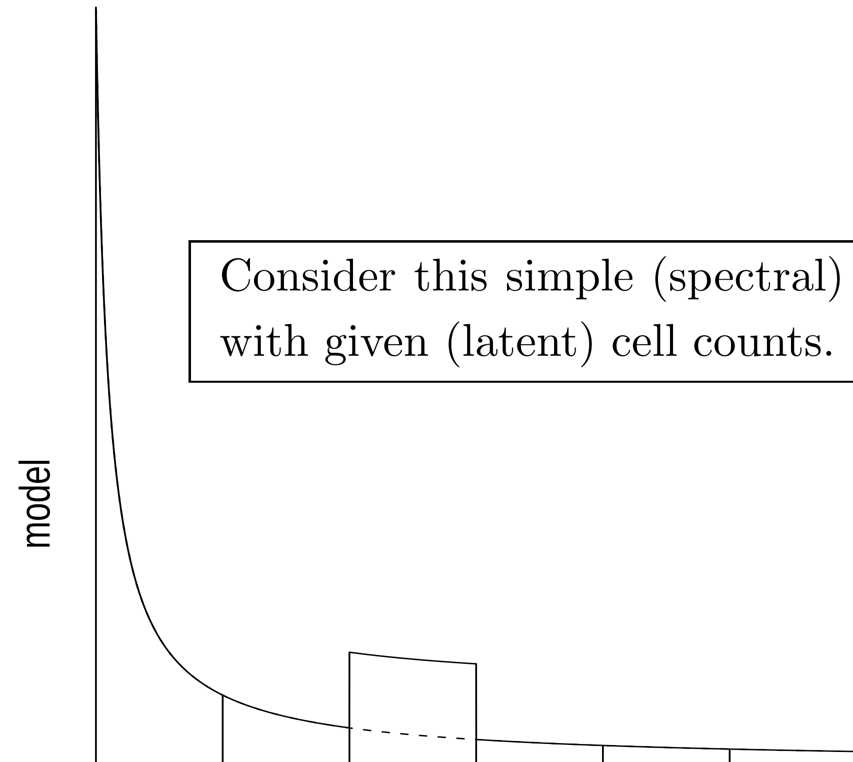
Consider this simple (spectral) model with given (latent) cell counts.



X = (latent) Cell Counts	10	4	8	1	2	0
Continuum Counts(Z=0)	10	4	~3	1	2	0
Line Counts (Z=1)	0	0	~5	0	0	0

Why the Gibbs Sampler Fails

Consider this simple (spectral) model with given (latent) cell counts.



X = (latent) Cell Counts	10	4	8	1	2	0
Continuum Counts(Z=0)	10	4	~3	1	2	0
Line Counts (Z=1)	0	0	~5	0	0	0

Given Z, what is the location of the emission line?

The Standard Gibbs Sampler

Recall we do not observe the latent Poisson Process,

$$X_i \sim \text{Poisson}\left(\Lambda_i = \alpha E_i^{-\beta} + \lambda^{\text{L}}\pi_i\right),$$

Rather we observe, $Y_j \sim \text{Poisson}\left(a_j \sum_i P_{ij}\Lambda_i + \xi_j\right)$

$$\begin{aligned}
Y_{\text{obs}} &= \{Y_j\} = \text{ obs cell cnts} \\
X &= \{X_i\} = \text{ latent cell cnts} \\
Z &= \text{emission line indicators} \\
\theta^{\text{L}} &= \text{location of emission line} \\
\theta^{\text{O}} &= \text{other model parameters}
\end{aligned}$$

The standard Gibbs sampler simulates:

1. $p(X, Z|\theta)$
2. $p(\theta|X, Z) = p(\theta^{\text{O}}|X, Z)p(\theta^{\text{L}}|X, Z)$

We tacitly condition on Y_{obs} throughout.

With a delta function point source model, this sampler fails.

An Incompatible Gibbs Sampler

- Recall the "Simplest Example":

$$\begin{matrix} p(\psi_1|\psi_2) \\ p(\psi_2|\psi_1) \end{matrix} \longrightarrow \begin{matrix} p(\psi_1|\psi_2) \\ p(\psi_2) \end{matrix} \longrightarrow \begin{matrix} p(\psi_2) \\ p(\psi_1|\psi_2) \end{matrix} \longrightarrow p(\psi_1, \psi_2)$$

- Following this we construct:

  Sampler 1: (A Blocked Version of the Original Sampler.)

$$\begin{matrix} p(X, Z|\theta) \\ p(\theta^{\mathrm{O}}|\theta^{\mathrm{L}}, X, Z) \\ p(\theta^{\mathrm{L}}|\theta^{\mathrm{O}}, X, Z) \end{matrix} \longrightarrow \begin{matrix} p(X, Z|\theta) \\ p(\theta^{\mathrm{O}}|\theta^{\mathrm{L}}, X, Z) \\ p(\theta^{\mathrm{L}}|\theta^{\mathrm{O}}) \end{matrix} \longrightarrow \begin{matrix} p(\theta^{\mathrm{L}}|\theta^{\mathrm{O}}) \\ p(X, Z|\theta) \\ p(\theta^{\mathrm{O}}|\theta^{\mathrm{L}}, X, Z) \end{matrix} \longrightarrow \begin{matrix} p(\theta^{\mathrm{L}}, X, Z|\theta^{\mathrm{O}}) \\ p(\theta^{\mathrm{O}}|\theta^{\mathrm{L}}, X, Z) \end{matrix}$$

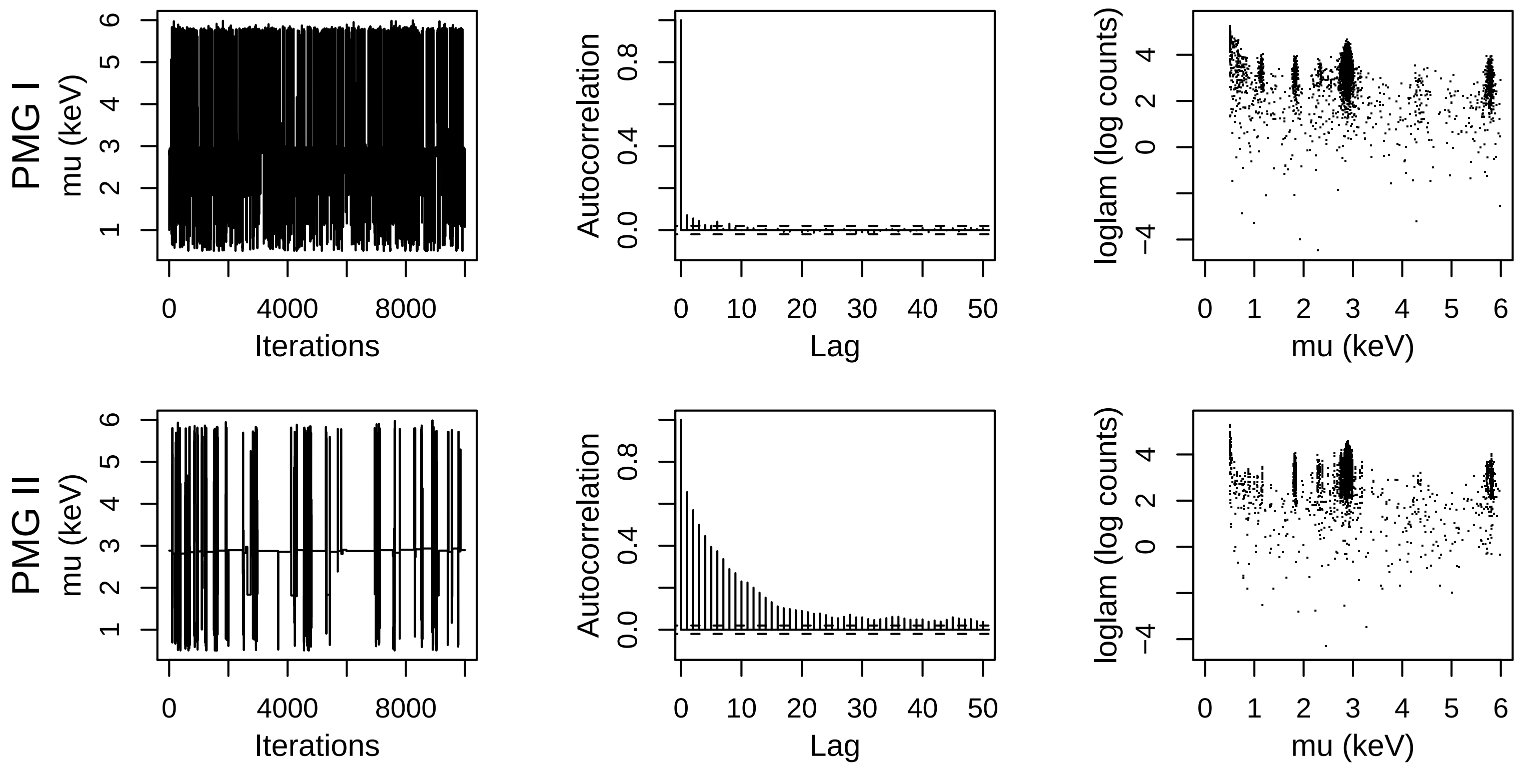
  Sampler 2: (Cannot be Blocked: An Incompatible Gibbs Sampler.)

$$\begin{matrix} p(X, Z|\theta) \\ p(\theta^{\mathrm{O}}|\theta^{\mathrm{L}}, X, Z) \\ p(\theta^{\mathrm{L}}|\theta^{\mathrm{O}}, X, Z) \end{matrix} \longrightarrow \begin{matrix} p(X, Z|\theta) \\ p(\theta^{\mathrm{O}}|\theta^{\mathrm{L}}, X, Z) \\ p(\theta^{\mathrm{L}}|\theta^{\mathrm{O}}, X) \end{matrix} \longrightarrow \begin{matrix} p(\theta^{\mathrm{L}}|\theta^{\mathrm{O}}, X) \\ p(X, Z|\theta) \\ p(\theta^{\mathrm{O}}|\theta^{\mathrm{L}}, X, Z) \end{matrix}$$

It can be shown that both samplers have the correct stationary distribution and are faster to converge than the standard sampler.

Computational Gains

- Compare Standard Sampler, Sampler 1, and Sampler 2 in a spectral analysis.
- Standard sampler doesn't move from its starting value.
- Sampler 1 has much better convergence characteristics than Sampler 2.
- However, each iteration of Sampler 1 is more expensive.



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.*

Verifying the Stationary Distribution of Sampler 2

$p(X, Z\|\theta)$ $p(\theta^{\mathrm{O}}\|\theta^{\mathrm{L}}, X, Z)$ $p(\theta^{\mathrm{L}}\|\theta^{\mathrm{O}}, X, Z)$	$\longrightarrow$	$p(X, Z\|\theta)$ $p(\theta^{\mathrm{O}}\|\theta^{\mathrm{L}}, X, Z)$ $p(\theta^{\mathrm{L}}, Z\|\theta^{\mathrm{O}}, X)$	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.
	$\longrightarrow$	$p(\theta^{\mathrm{L}}, Z\|\theta^{\mathrm{O}}, X)$ $p(X, Z\|\theta)$ $p(\theta^{\mathrm{O}}\|\theta^{\mathrm{L}}, X, Z)$	We permute the order of the steps. This can have minor effects on the rate of convergence, but does not affect the stationary distribution.
	$\longrightarrow$	$p(\theta^{\mathrm{L}}\|\theta^{\mathrm{O}}, X)$ $p(X, Z\|\theta)$ $p(\theta^{\mathrm{O}}\|\theta^{\mathrm{L}}, X, Z)$	We remove Z from the draw in Step 1, since the transition kernel does not depend on this quantitity.

We refer to these three steps tools as Marginalizing, Permuting, and Trimming. They form a general strategy for constructing incompatible Gibbs samplers.

Reducing the Conditioning

The primary computational advantage stems from the Marginalization step: We reduce the correlation in the chain by reducing the conditioning in the draws.

An Analogy with the EM algorithm

- The EM algorithm has two steps:
  1. The *complete data* (sufficient statistics) are updated given the parameters in the E-step.
  2. The parameters are updated given the complete data via the M-step

$$\theta' = \text{argmax}\ \mathrm{E}\left[\ell(\theta'|Y_{\text{com}})|\theta, Y_{\text{obs}}\right]$$

- The ECM algorithm replace the M-step by a series of K CM-steps:

$$\theta'_k = \text{argmax}\ \mathrm{E}\left[\ell(\theta'|Y_{\text{com}})|\theta, Y_{\text{obs}}\right] \text{ subject to } \theta'_j = \theta_j \text{ for } j \neq k.$$

Two and $K+1$ step Gibbs samplers can be constructed that are analogous to the EM and ECM algorithms.

The ECME and AECM Algorithms

- To improve the rate of convergence of the ECM algorithm, Liu and Rubin (1995) suggested replacing one or more of the CM steps

$$\theta'_k = \text{argmax } \mathrm{E}\left[\ell(\theta'|Y_{\text{com}})|\theta, Y_{\text{obs}}\right] \text{ subject to } \theta'_j = \theta_j \text{ for } j \neq k.$$

  with

$$\theta'_k = \text{argmax } \ell(\theta'|Y_{\text{obs}}) \text{ subject to } \theta'_j = \theta_j \text{ for } j \neq k.$$

- AECM generalizes this with

$$\theta'_k = \text{argmax } \mathrm{E}\left[\ell(\theta'|Y^{\star}_{\text{com}})|\theta, Y_{\text{obs}}\right] \text{ subject to } \theta'_j = \theta_j \text{ for } j \neq k,$$

  where $Y_{\text{obs}} \subset Y^{\star}_{\text{com}} \subset Y_{\text{com}}$.

Meng and van Dyk (1997) showed that the order of the steps can effect the celebrated monotone convergence of EM-type algorithms. We show that in the Gibbs sampler analogy, the order of the steps can effect the stationary distribution of the chain.

Reducing the Conditioning Improves the Computation.

The Advantage of Partial Collapse

An Outline of a proof:

- The dependence of consecutive iterations of the Gibbs Sampler flows through what is conditioned upon in the *first step* of each iteration.
- The maximal autocorrelation can only decrease if we reduce this conditioning.
- The Spectral Radius of the Chain
  - generally governs convergence,
  - is bounded above by the maximal autocorrelation, and
  - does not depend on which step begins the iteration, as long as the order of steps is not altered.

By reducing conditioning in any step (i.e., partial marginalization) we reduce both a bound on the spectral radius of the chain and the maximal autocorrelation for the chain that starts with that step.

Summary

I hope I have given you a taste of my strategy of utilizing
Highly Structured Statistical Models and
Sophisticated Statistical Computation
to solve outstanding substantive scientific questions
in High-Energy Astrophysics.

Selected References

The Astrophysics Spectral and Image Models

Park, T., van Dyk, D. A., and Siemiginowska A. (2008). Searching for Narrow Emission Lines in X-ray Spectra: Computation and Methods. *The Astrophysical Journal*, under revision.

van Dyk, D. A., Connors, A., Esch, D. N., Freeman, P., Kang, H., Karovska, M., and Kashyap, V. (2006). Deconvolution in High-Energy Astrophysics: Science, Instrumentation, and Methods (With Discussion). *Bayesian Analysis*, 189–248.

Esch, D. N., Connors, A., Karovska, M., and van Dyk, D. A. (2004). A Image Restoration Technique with Error Estimates. *The Astrophysical Journal,* vol. 610, 1213–1227.

Protassov, R., van Dyk, D. A., Connors, A., Kashyap, V. L., & Siemiginowska, A. (2002). Statistics: Handle with Care, Detecting Multiple Model Components with the Likelihood Ratio Test, *The Astrophysical Journal*, vol. 571, 545–559.

van Dyk, D. A., Connors, A., Kashyap, V. L., & Siemiginowska, A. (2001). Analysis of Energy Spectrum with Low Photon Counts, *The Astrophysical Journal*, vol. 548, 224–243.