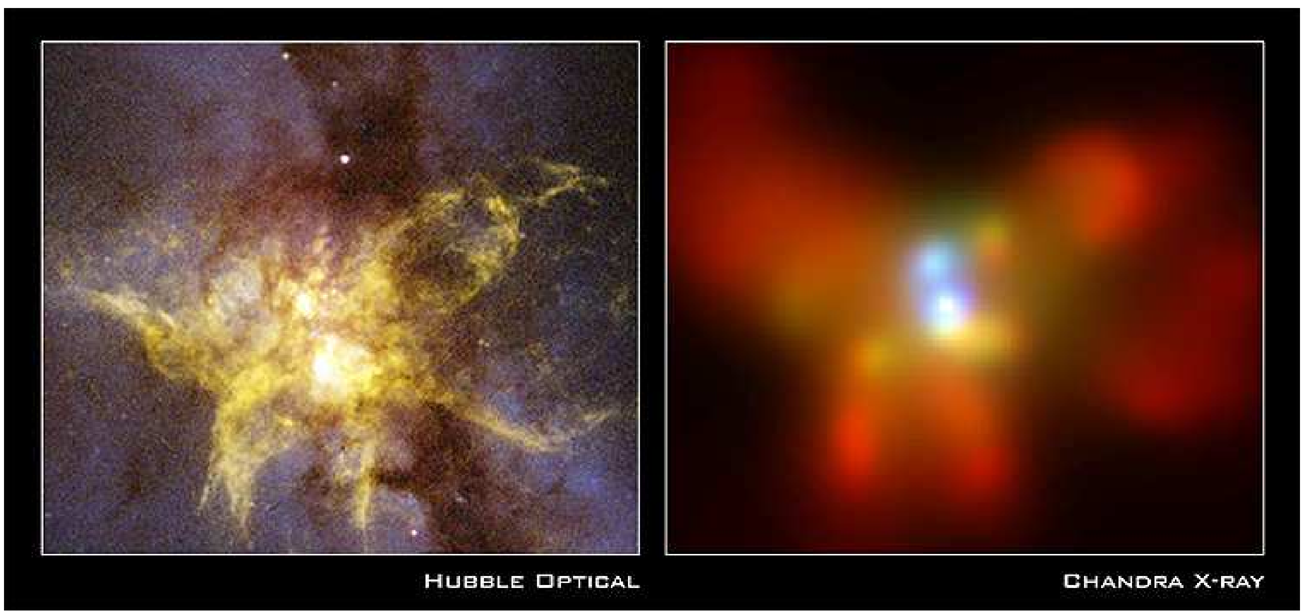
Highly Structured Models in High Energy Astrophysics

David van Dyk

Department of Statistics
University of California, Irvine

Joint work with
The California-Harvard Astrostatistics Collaboration



Outline of Presentation

This talk has three components:

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

- Astrostatistics:

  Complex Sources, Complex Instruments, and Complex Questions

  Key: All three are the domain of Astrostatistics
- Model-Based Statistical Solutions
- Monte Carlo-Based Bayesian Analysis

B. Examples

1. The EMC2 package for Image Analysis (A detailed example.)
2. The BRoaDEM package for DEM Reconstruction
3. The BLoCXS package for Spectral Analysis
4. The BEHR package for computing Hardness Ratios

C. Using Incompatible Conditional Distributions in Gibbs Samplers

Primary Collaborators

EMC2

Alanna Connors, David Esch, Margarita Karovska, and David van Dyk

BRoaDEM

Alanna Connors, Hosung Kang, Vinay L. Kashyap, and David van Dyk

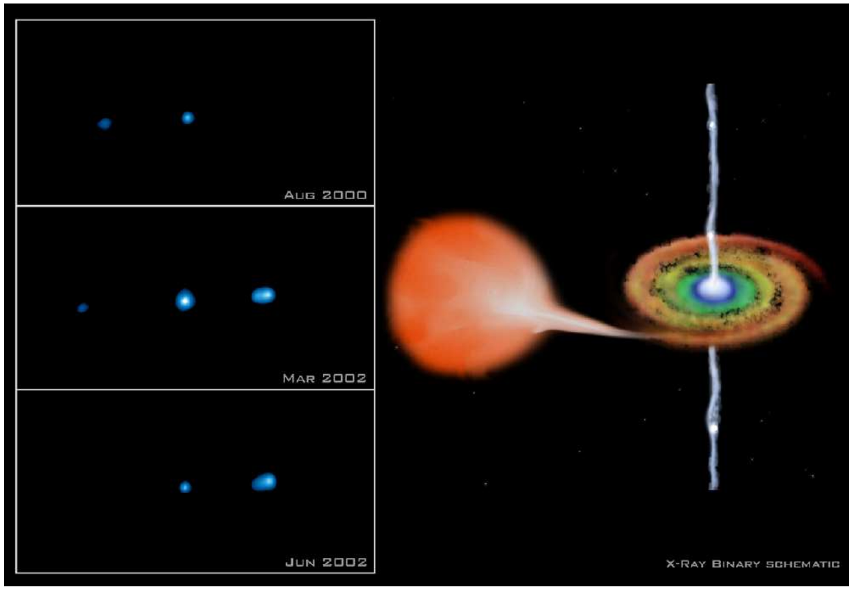
BLoCXS

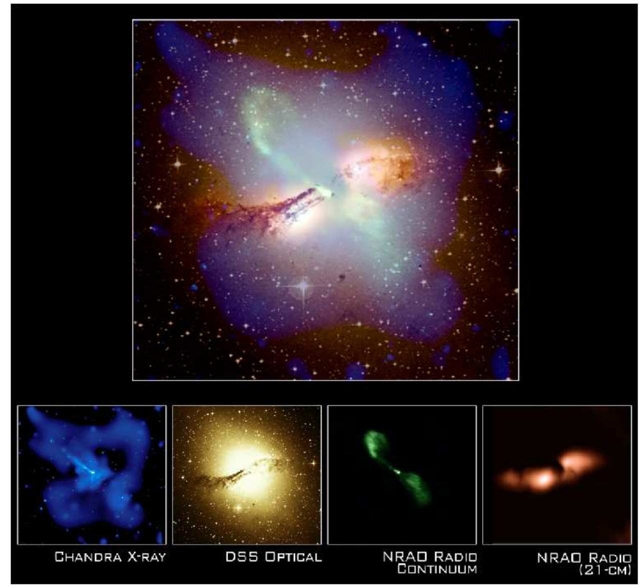
Alanna Connors, Peter Freeman, Christopher Hans, Vinay L. Kashyap, Taeyoung Park, Rostislav Protassov, Aneta Siemiginowska, David van Dyk, Yaming Yu, and Andreas Zezas,

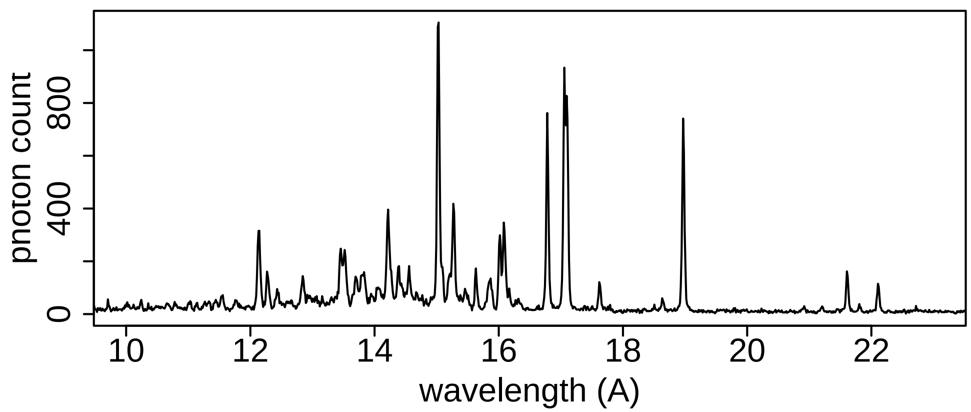
BEHR

Vinay L. Kashyap, Taeyoung Park, David van Dyk, and Andreas Zezas,

Complex Astronomical Sources

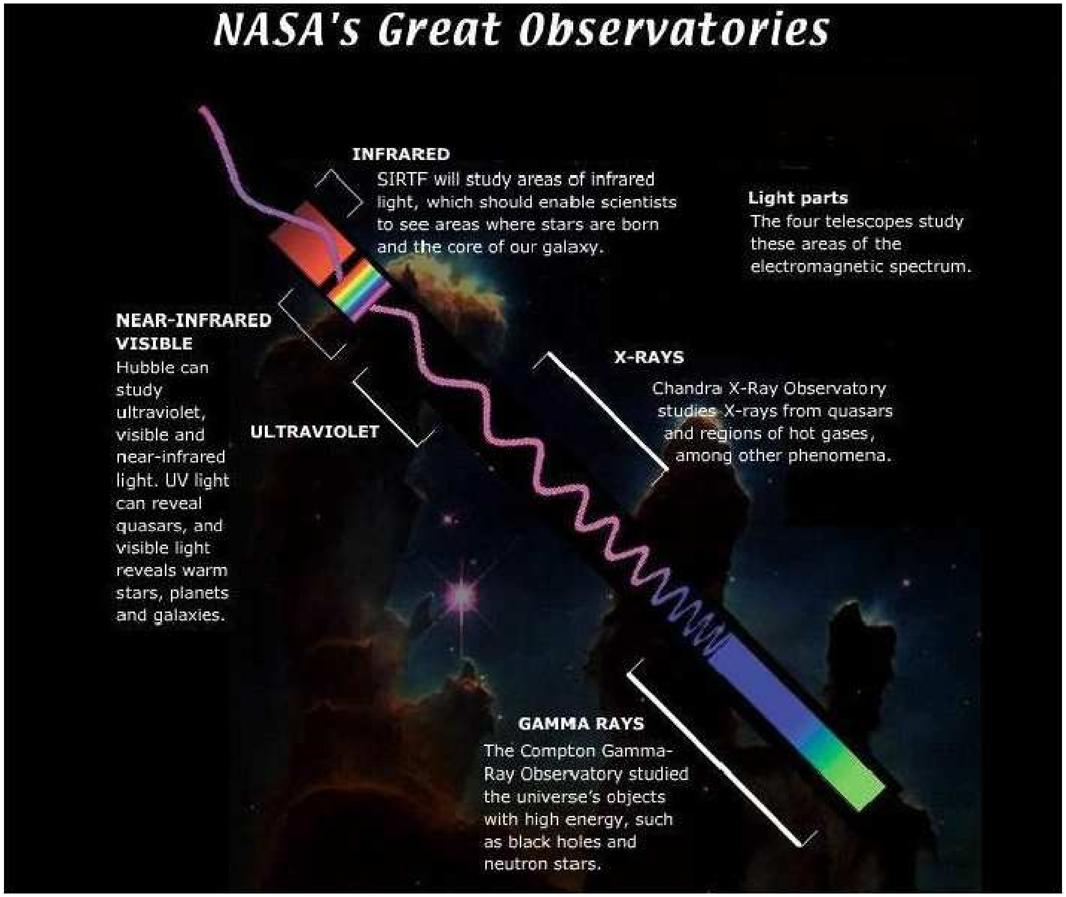






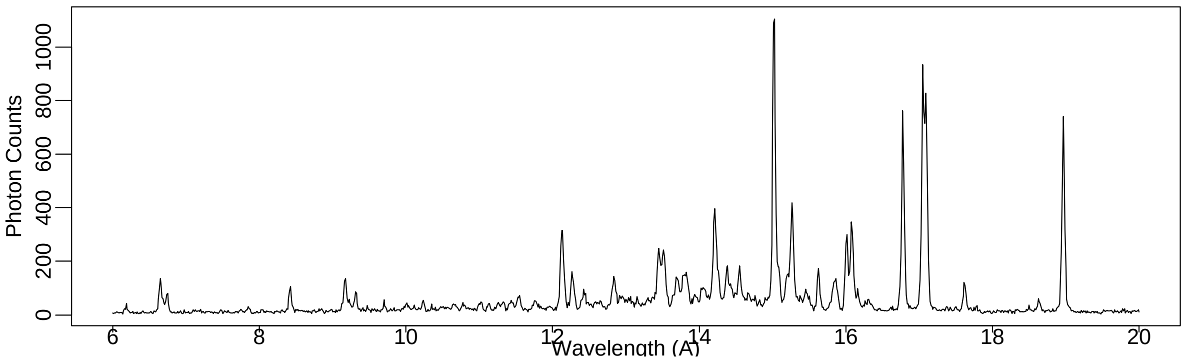
Images may exhibit Spectral, Temporal, and Spatial Characteristics.

Astrostatistics: Complex Data Collection



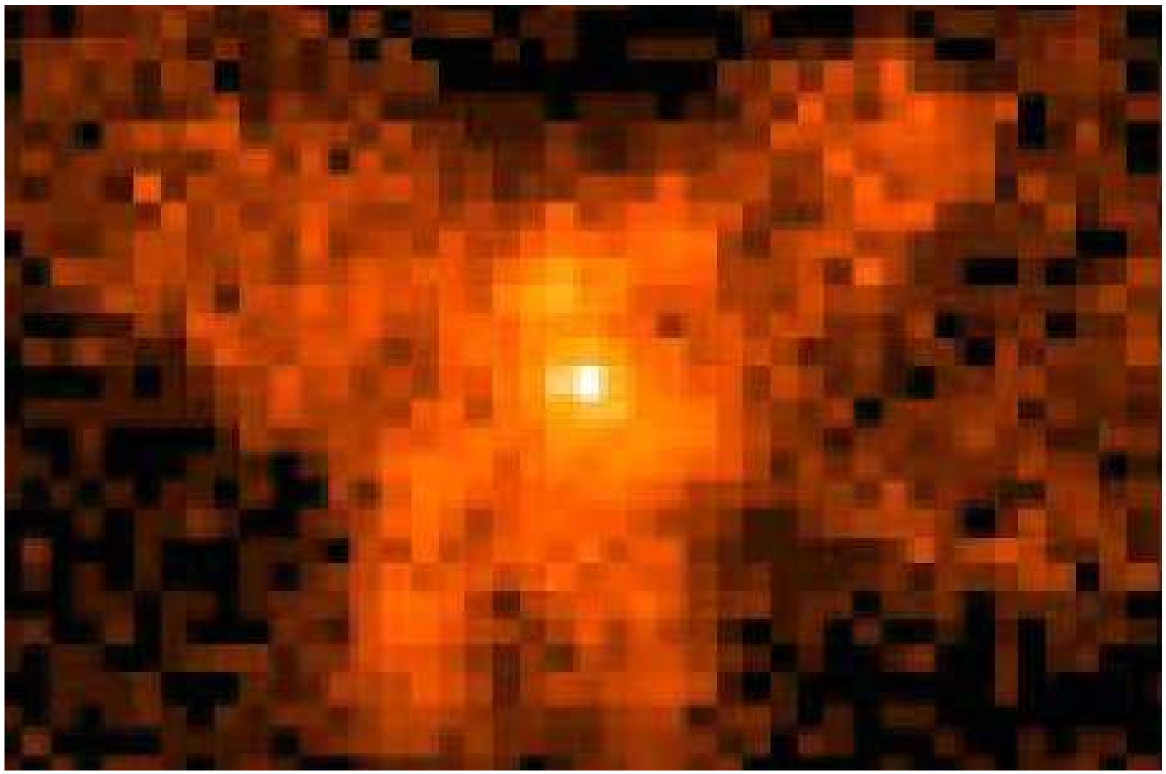
- A very small sample of instruments
- Earth-based, survey, interferometry, etc.
- X-ray alone: at least four planned missions
- Instruments have different data-collection mechanism

Astrostatistics: Complex Questions



- What is the composition and temperature structure?

Astrostatistics: Complex Questions



- Are the loops of hot gas *real*?

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)

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

NGC 6240

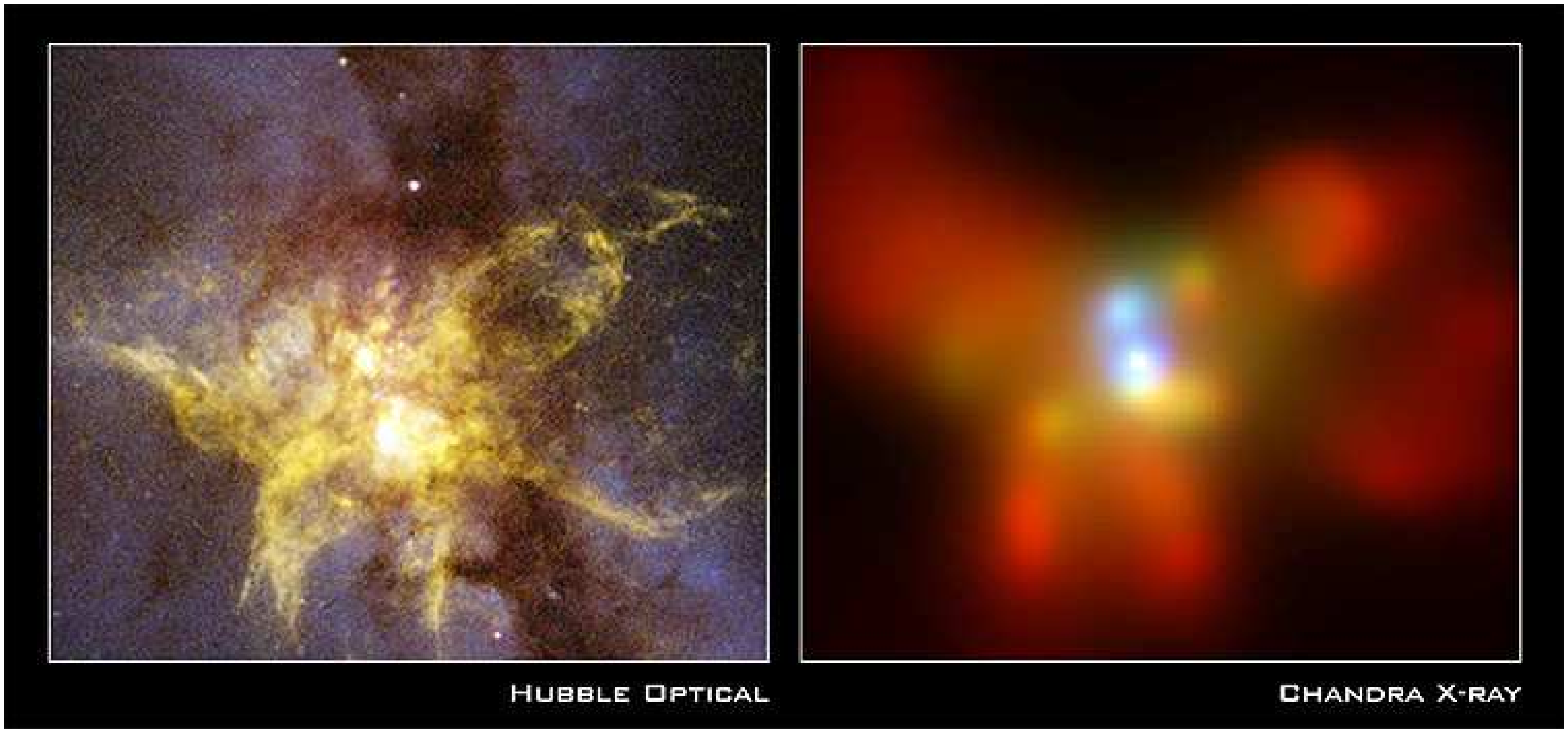
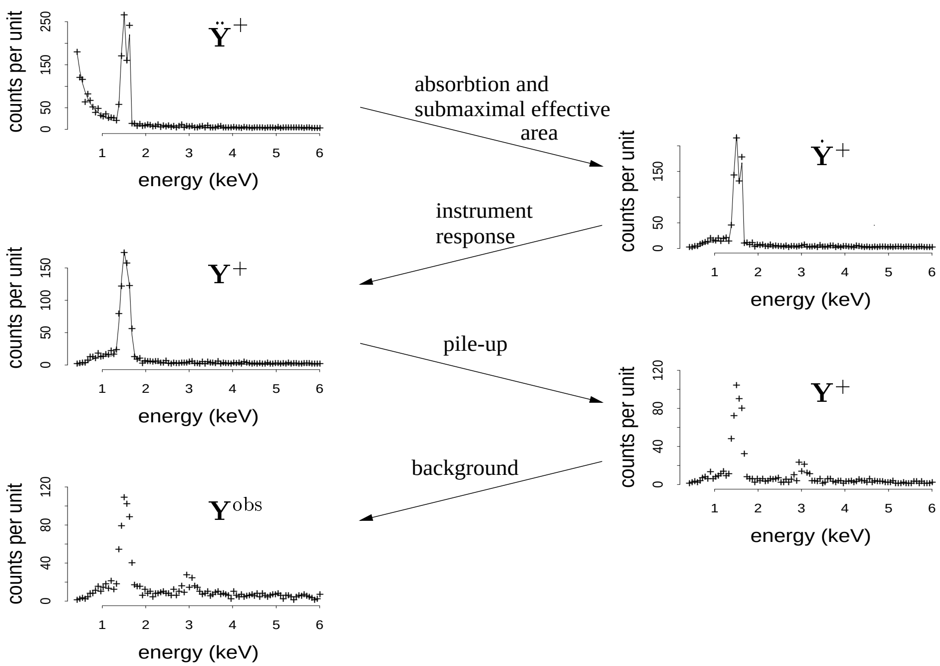


Image Credits.
X-ray: NASA/CXC/MPE/, Komossa et al. (2003, ApJL, 582, L15);
Optical: NASA/STScI/R.P.van der Marel & J.Gerssen.

Highly Structured Models

Modelling 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>

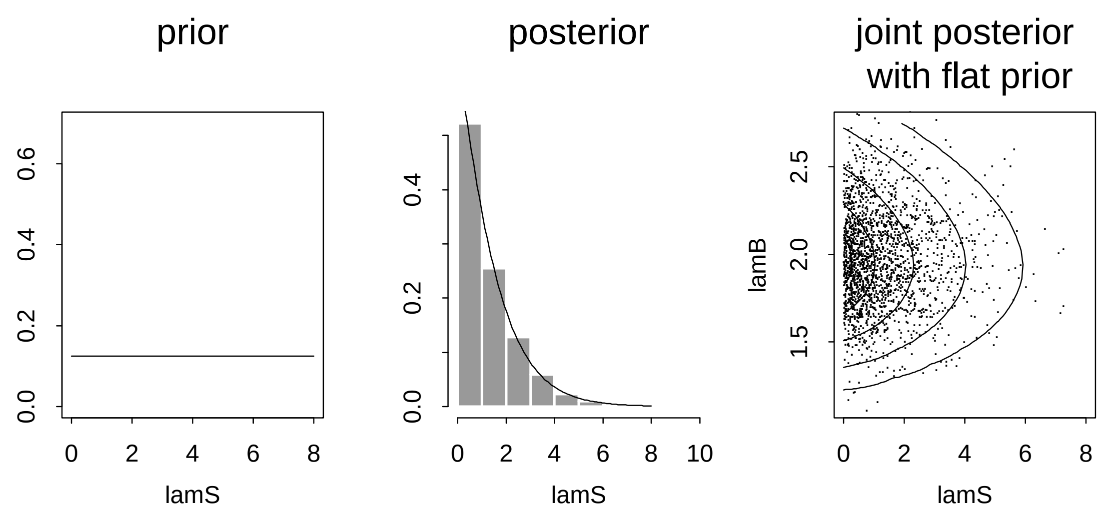
1. Priors can be used
   - to incorporate information from outside the data, or
   - to impose structure.
2. Priors offer a principled compromise between "fixing" a parameter & letting it "float free".
3. Setting `min` and `max` limits in XSPEC amounts to using a flat prior over a specified range.

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: $p(\psi|Y) \propto p(Y|\psi)p(\psi)$

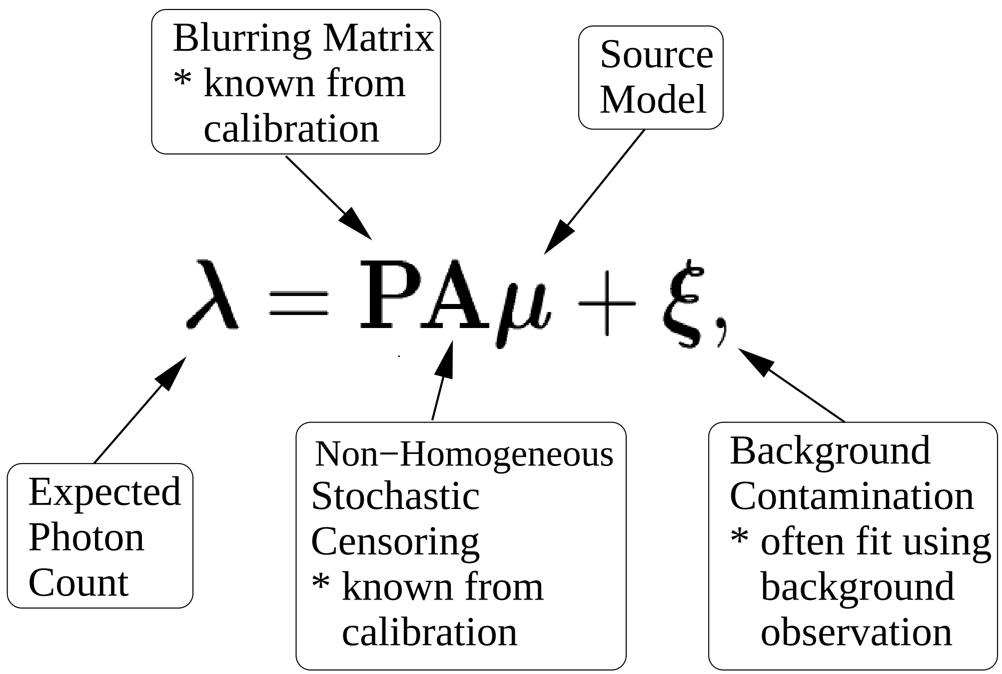
Inference Using a Monte Carlo Sample:



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

Bayesian Deconvolution

- The Data Collection Mechanism



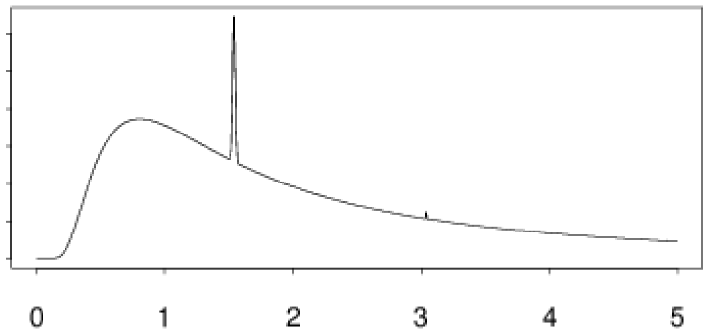
$$\boldsymbol{\lambda} = \mathbf{P}\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\xi},$$

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

The Source Models

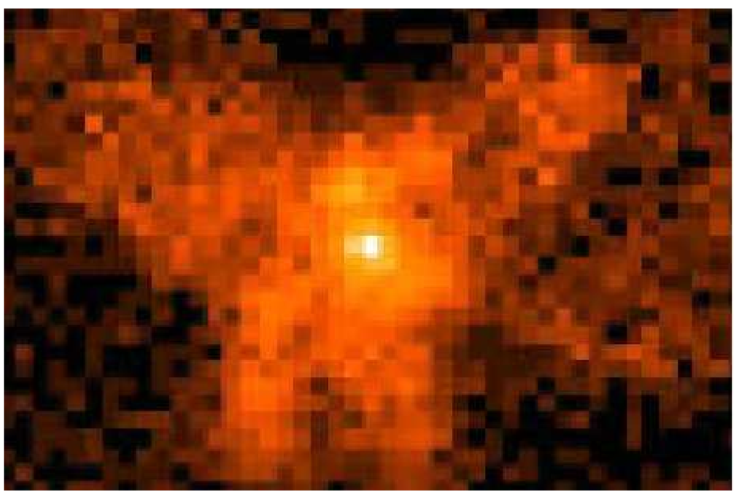
Parameterized finite mixture models

(source models w/ several components)



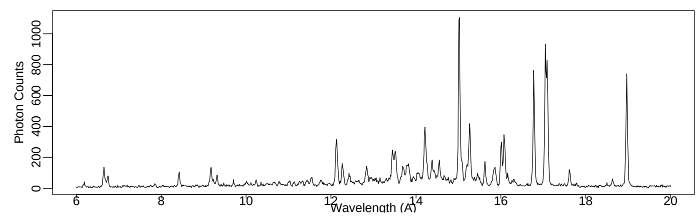
Smoothing prior distributions

(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 three components:

A. Highly Structured Models in High-Energy Astrophysics

- Astrostatistics:

  Complex Sources, Complex Instruments, and Complex Questions

  Key: All three are the domain of Astrostatistics

- Model-Based Statistical Solutions
- Monte Carlo-Based Bayesian Analysis

B. **Examples**

1. The EMC2 package for Image Analysis (A detailed example.)
2. The BRoaDEM package for DEM Reconstruction
3. The BLoCXS package for Spectral Analysis
4. The BEHR package for computing Hardness Ratios

C. Using Incompatible Conditional Distributions in Gibbs Samplers

Example 1: The EMC2 package for Image Analysis

The Source Model

- A Poisson Process for the *missing ideal counts.*

$$Z_i \sim \text{Poisson}(\mu_i)$$

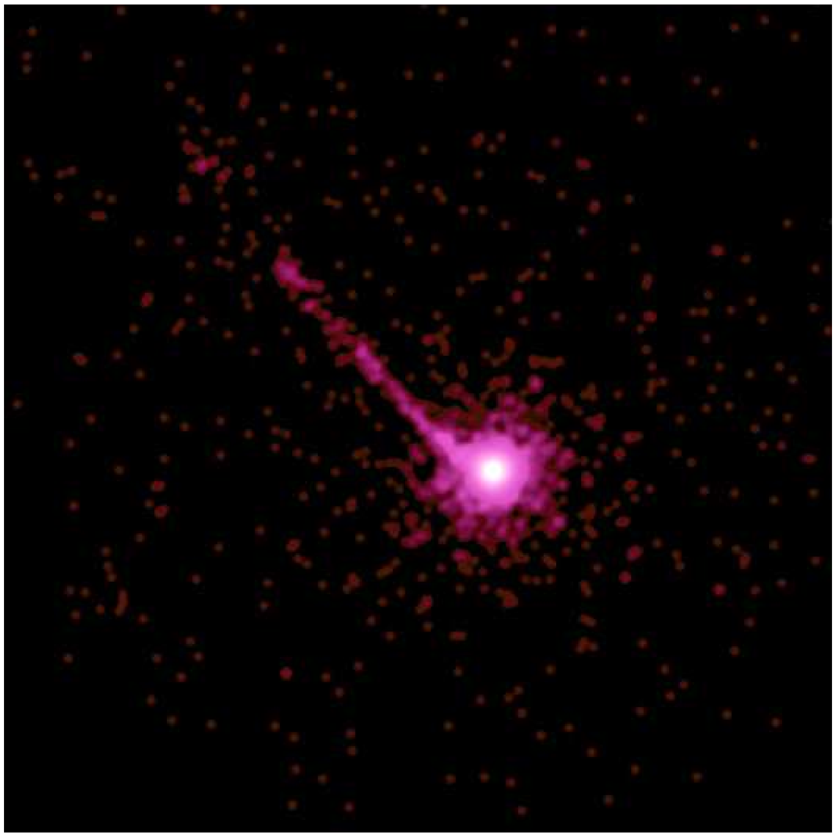
- A useful source model must allow for
  1. extended diffuse nebula with irregular and unpredictable structure
  2. one or more concentrated X-ray emitters.

$$\mu_i = \mu_i^{\text{ES}} + \sum_{k=1}^{K} \mu_k^{\text{PS}} p_{ik}$$

The point sources can be modeled as delta functions, Gaussians or Lorentzians.

Additional Model Components

We can add additional model components

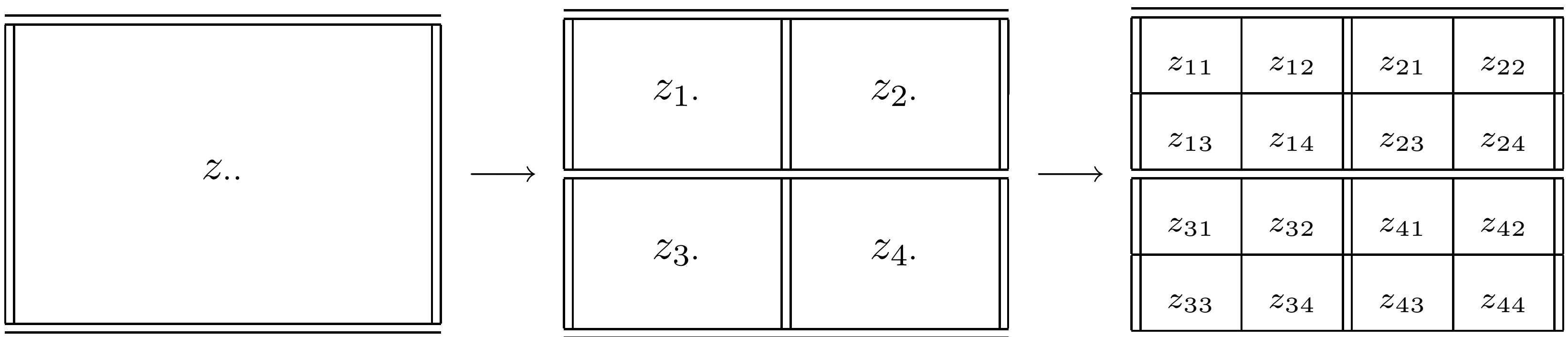


A jet can be modeled as a string of elongated Gaussian distributions.

A Smoothing Prior for the Extended Source

The Nowak-Kolaczyk Multiscale Model:

Low Resolution *High Resolution*



$z_{\cdot\cdot} \sim \text{Poisson}(\mu)$

$\mu \sim \text{Gamma}\{(\alpha_0, \beta_1)\}$

$\boldsymbol{z}_{,\cdot} | z_{\cdot\cdot} \sim \text{Multinomial}(\boldsymbol{p}_1)$

$\boldsymbol{p}_1 \sim \text{Dirich.}\{(\alpha_1, \alpha_1, \alpha_1, \alpha_1)\}$

$\boldsymbol{z}_{i,} | z_{i\cdot} \sim \text{Multinomial}(\boldsymbol{p}_{2i})$

$\boldsymbol{p}_{2i} \sim \text{Dirich.}\{(\alpha_2, \alpha_2, \alpha_2, \alpha_2)\}$

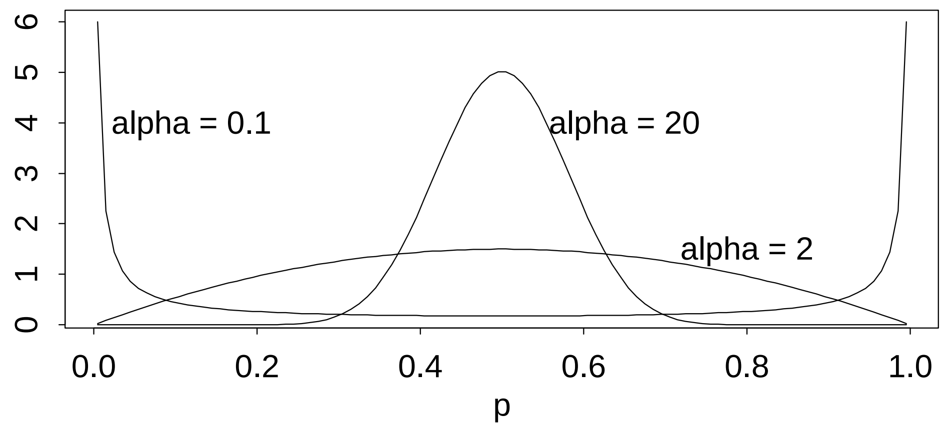
Wavelet like model in a fully Bayesian analysis.

Setting the Smoothing Parameters

The Multiscale prior distribution is specified in terms of a number of Dirichlet smoothing parameters $(\alpha_1, \alpha_2, \ldots)$.

- There is one parameter at each level of resolution.
- Larger values of each α encourage more smoothing

With binary splits



- Some researchers suggest parameterizing the α_j, e.g., setting $\alpha_j = ak^j$.
- Based on statistical properties of the model, e.g., correlation functions and posterior concavity (Nowak and Kolaczyk; Bouman, Dukic, and Meng).

Instead, we propose a strategy that fits the smoothing parameters to the data.

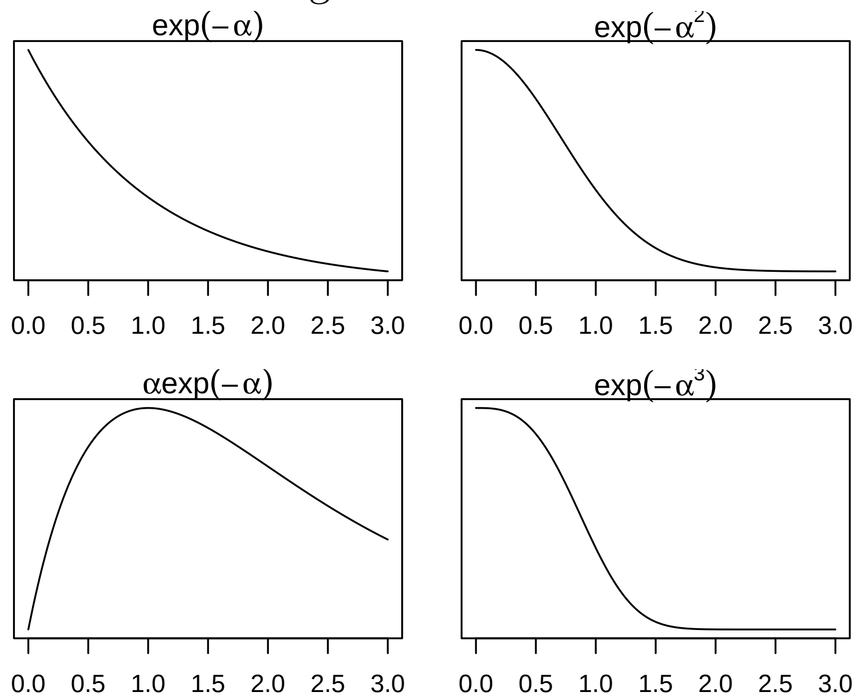
Fitting the Smoothing Parameters

We may fit the smoothing parameters (α_k) if we regularize their values.

- We use a common prior
  - Too much mass near zero leads to numerical instability. (Priors that put all mass in 1 quadrant.)
  - Too much mass far from zero results in too much smoothing.
- A compromise:

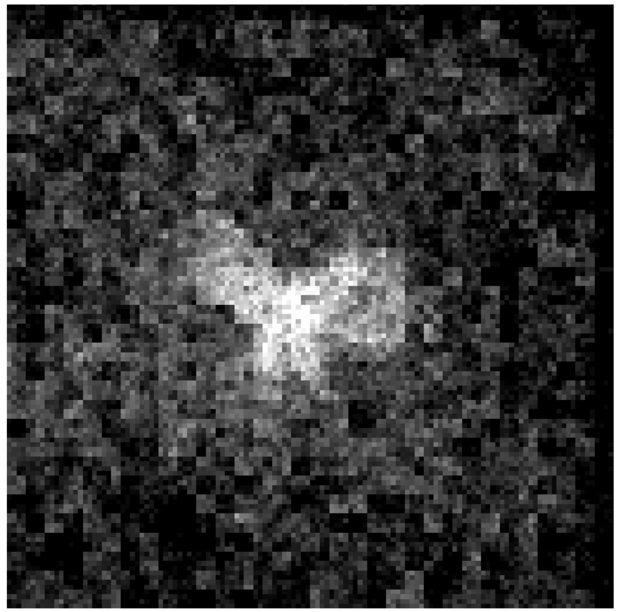
$$\alpha_k \sim \exp(-\delta\alpha^3/3)$$

The exact shape of the prior matters less than its general features.



These priors can be viewed as a smoother way of setting the "range" of the smoothing parameters, with δ specifying the range.

Prior Correlation Structure



A mixture prior distribution

- Our prior specification depends on the choice of coordinates.
- For each choice there is a corresponding multiscale prior distributions.
- We propose using an equally weighted mixture of each of these priors.
- Removes the checker-board pattern in the results.

This "cycle-spinning" strategy is analogous to what is done with wavelets.

Sensitivity Analysis

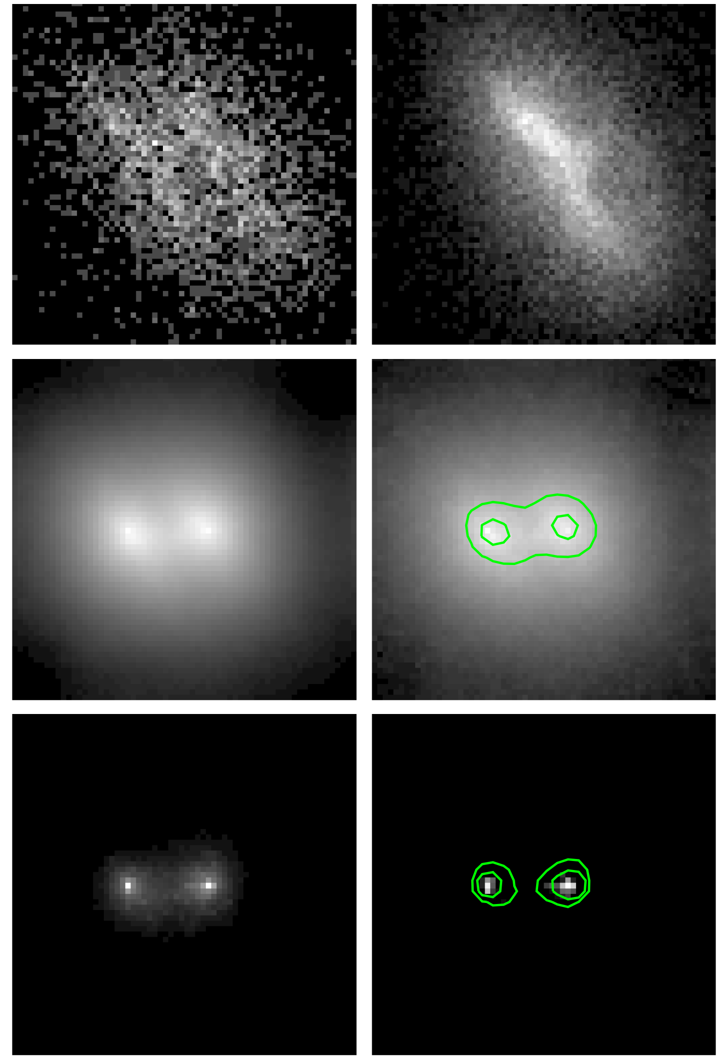
- A data set was simulated using a binary source.
- Fit using 2 priors.
- Significance maps plot posterior mean over posterior std dev.
- Contours are at levels 3 and 10.

simulated data and PSF

results under

$p(\alpha) \propto \exp(-\frac{1000}{3}\alpha^3)$

$p(\alpha) \propto \exp(-\frac{10}{3}\alpha^3)$



In the spirit of significance testing, if we are looking for evidence of an extended source, we pick a prior distribution that favors a point source.

Statistical Computation

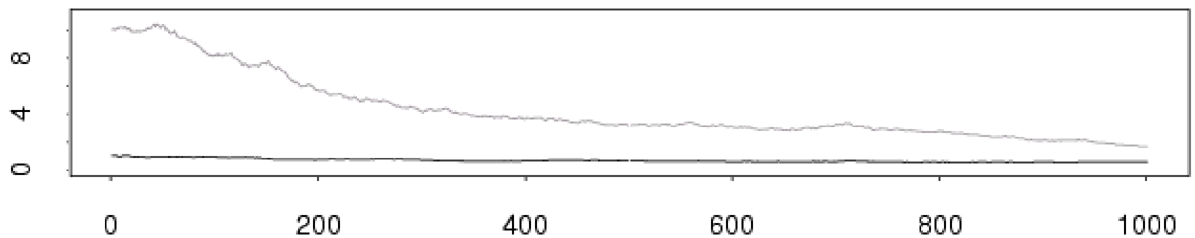
- We use a three-step Gibbs sampler to construct a Markov chain with stationary distribution equal to the target posterior distribution:

  Step 1: Sample $\boldsymbol{Z}$ given $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, and $\boldsymbol{Y}$

  Step 2: Sample $\boldsymbol{\mu}$ given $\boldsymbol{Z}$, $\boldsymbol{\alpha}$, and $\boldsymbol{Y}$.

  Step 3: Sample $\boldsymbol{\alpha}$ given $\boldsymbol{Z}$, $\boldsymbol{\mu}$, and $\boldsymbol{Y}$.

  Here, $\boldsymbol{Z}$ is the ideal counts, $\boldsymbol{\mu}$ is the image $\boldsymbol{Y}$ is the data, and $\boldsymbol{\alpha}$ is the smoothing parameters.



1000 draws of a smoothing parameter using two starting values.

Poor Mixing!

The Advantage of Blocking

- Original Sampler:

  Step 1: Sample $\boldsymbol{Z}$ given $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, and $\boldsymbol{Y}$

  Step 2: Sample $\boldsymbol{\mu}$ given $\boldsymbol{Z}$, $\boldsymbol{\alpha}$, and $\boldsymbol{Y}$.

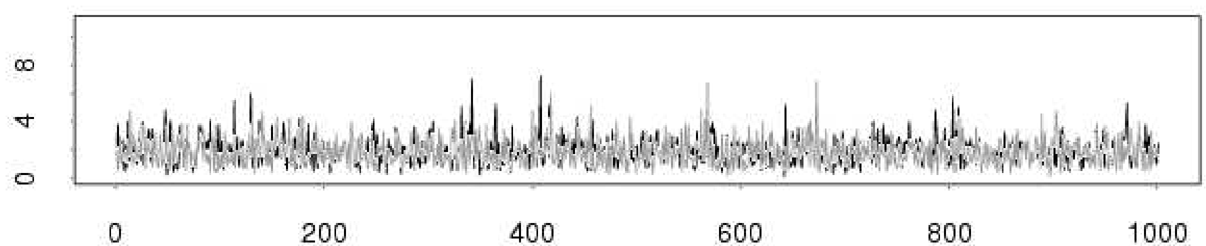
  Step 3: Sample $\boldsymbol{\alpha}$ given $\boldsymbol{Z}$, $\boldsymbol{\mu}$, and $\boldsymbol{Y}$.

- A simple change:

  Step 1: Sample $\boldsymbol{Z}$ given $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, and $\boldsymbol{Y}$

  Step 3: Sample $\boldsymbol{\alpha}$ given $\boldsymbol{Z}$ and $\boldsymbol{Y}$.

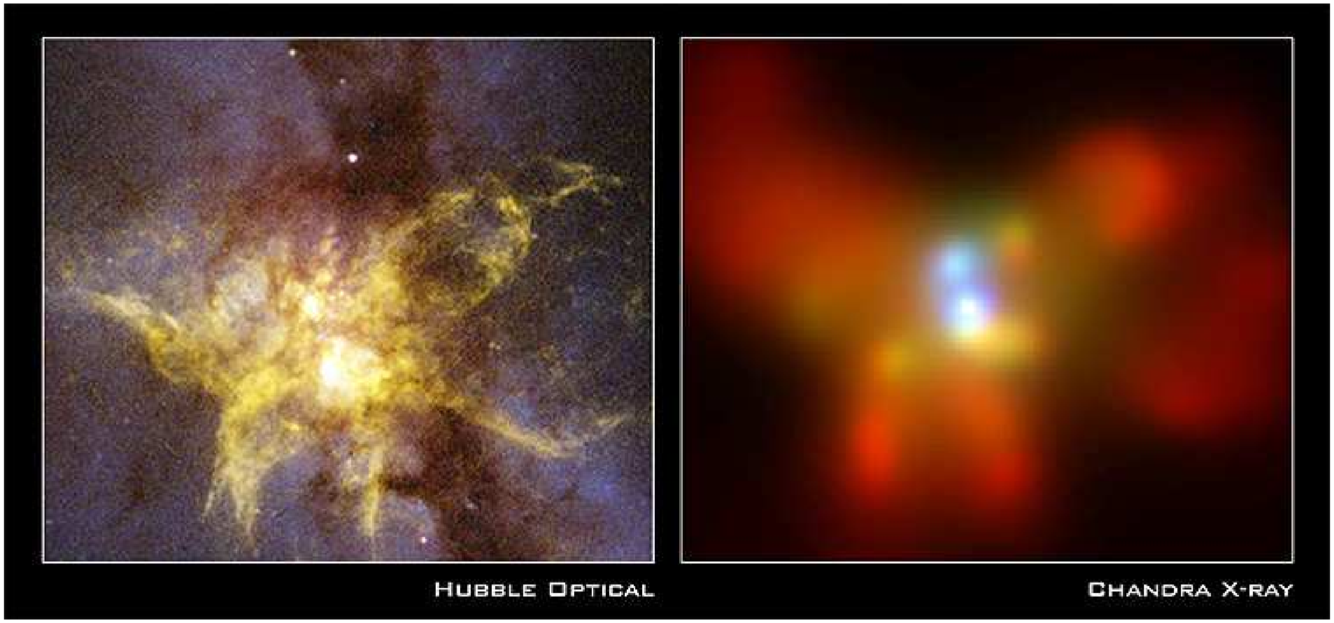
  Step 2: Sample $\boldsymbol{\mu}$ given $\boldsymbol{Z}$, $\boldsymbol{\alpha}$, and $\boldsymbol{Y}$.



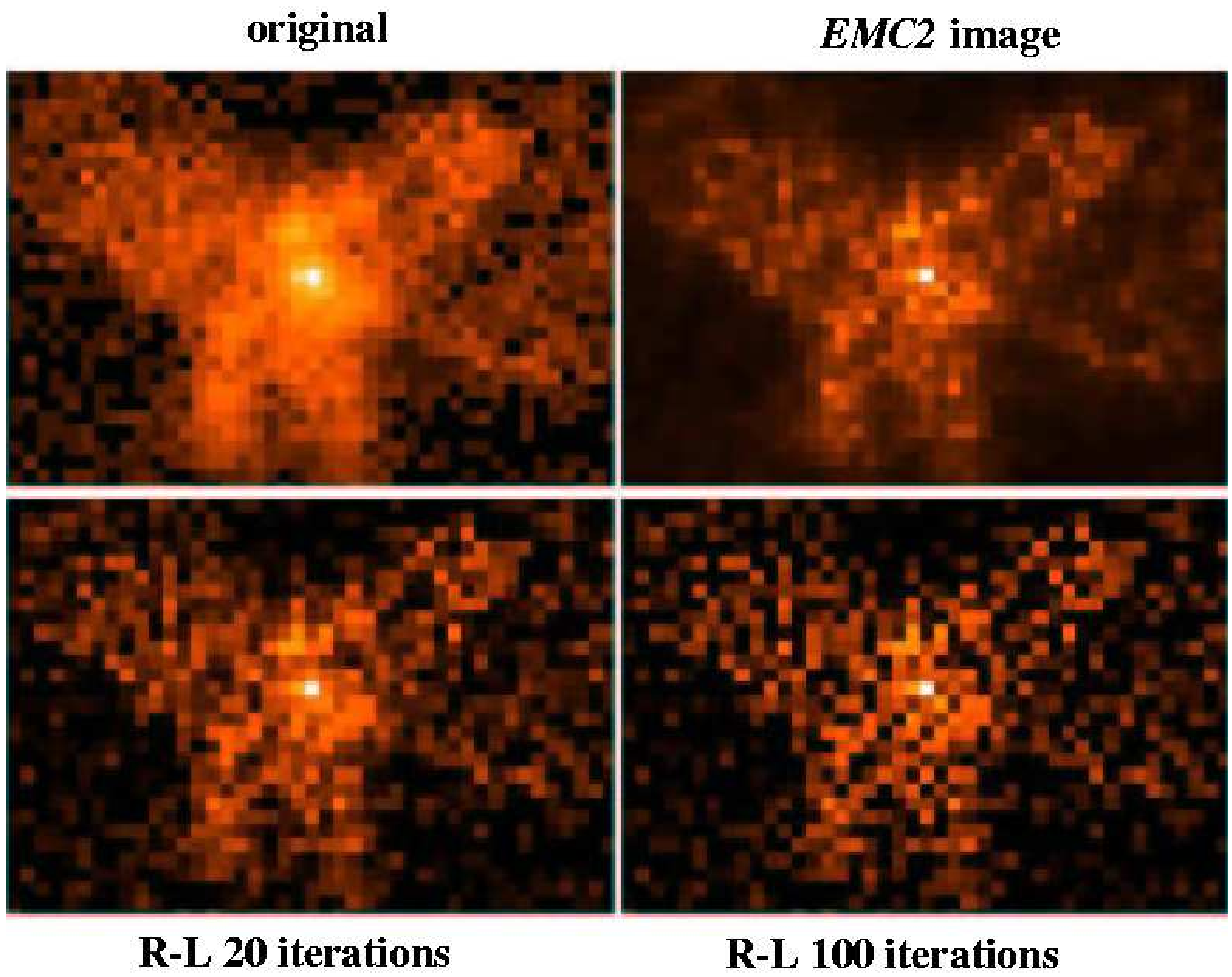
1000 draws of a smoothing parameter using two starting values.

Much Better Mixing!

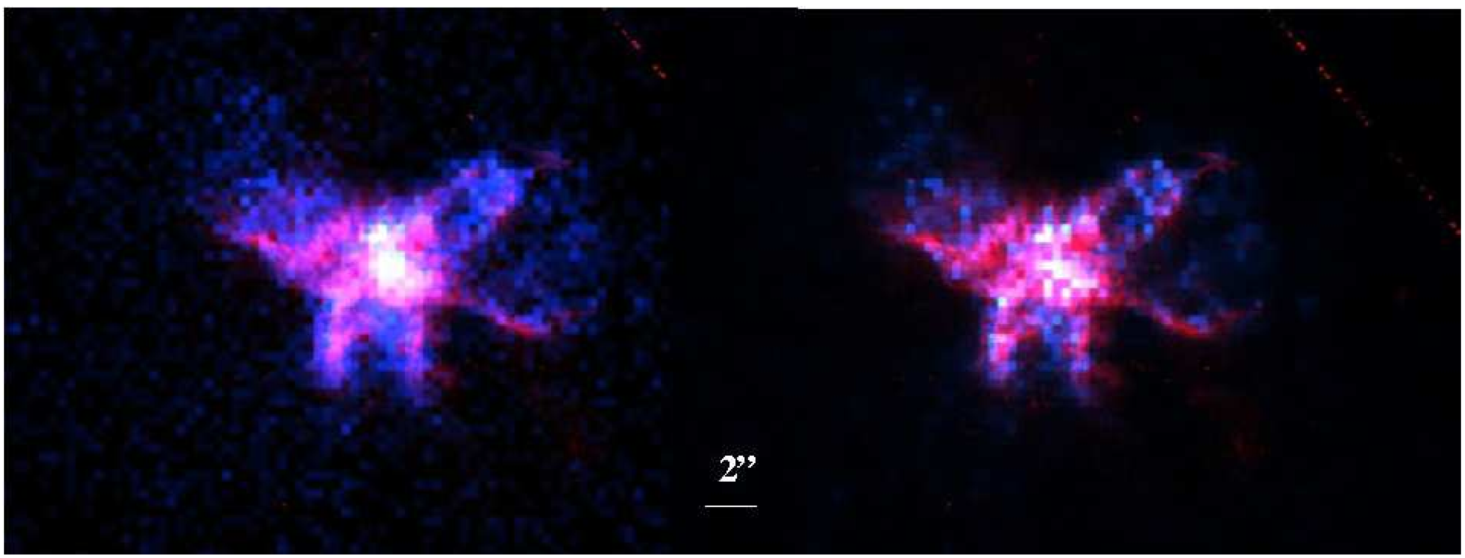
NGC 6240



The Effect of Nowak-Kolaczyk Multiscale Smoothing Prior

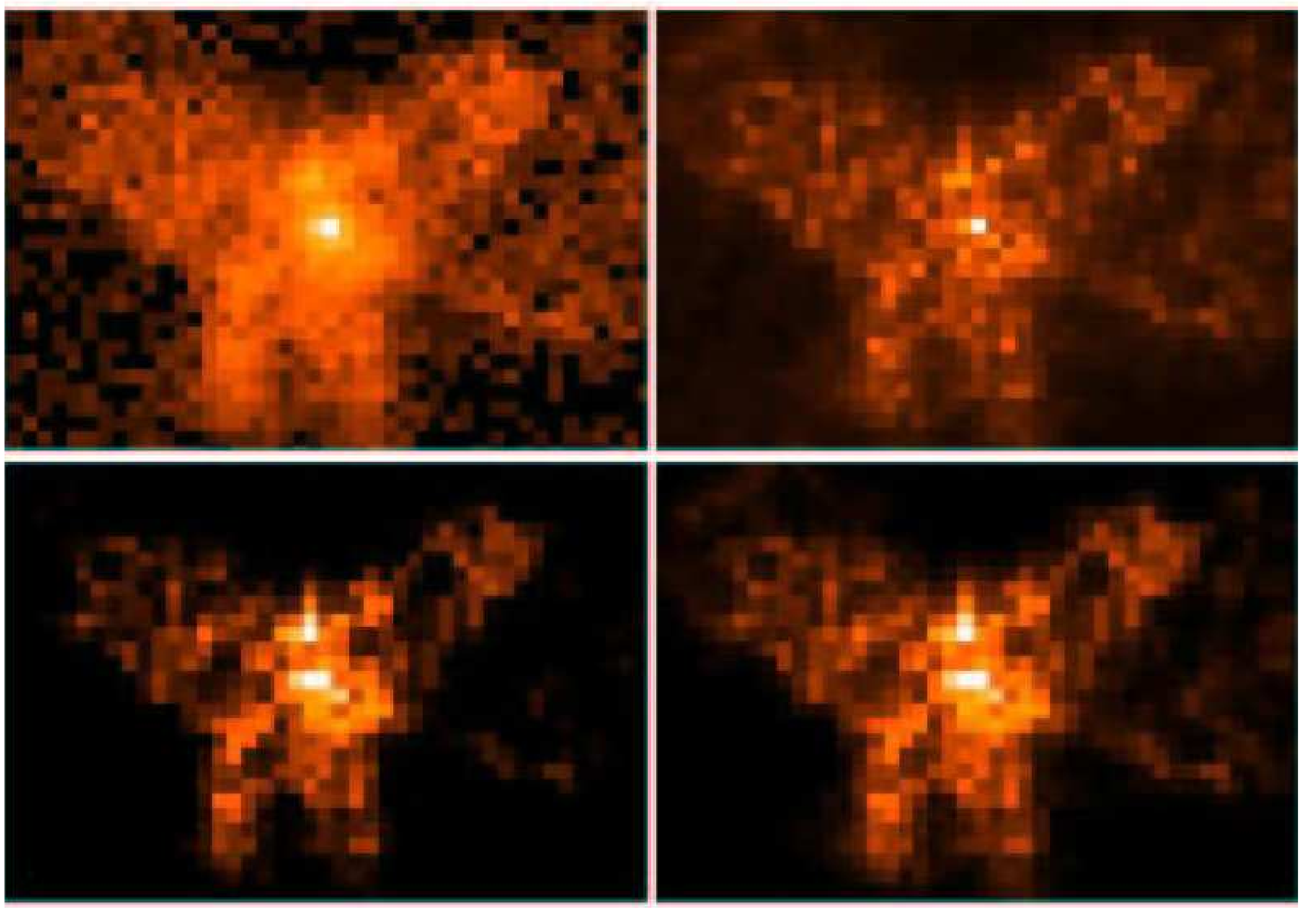


Evaluating the Fit



The Significance Maps

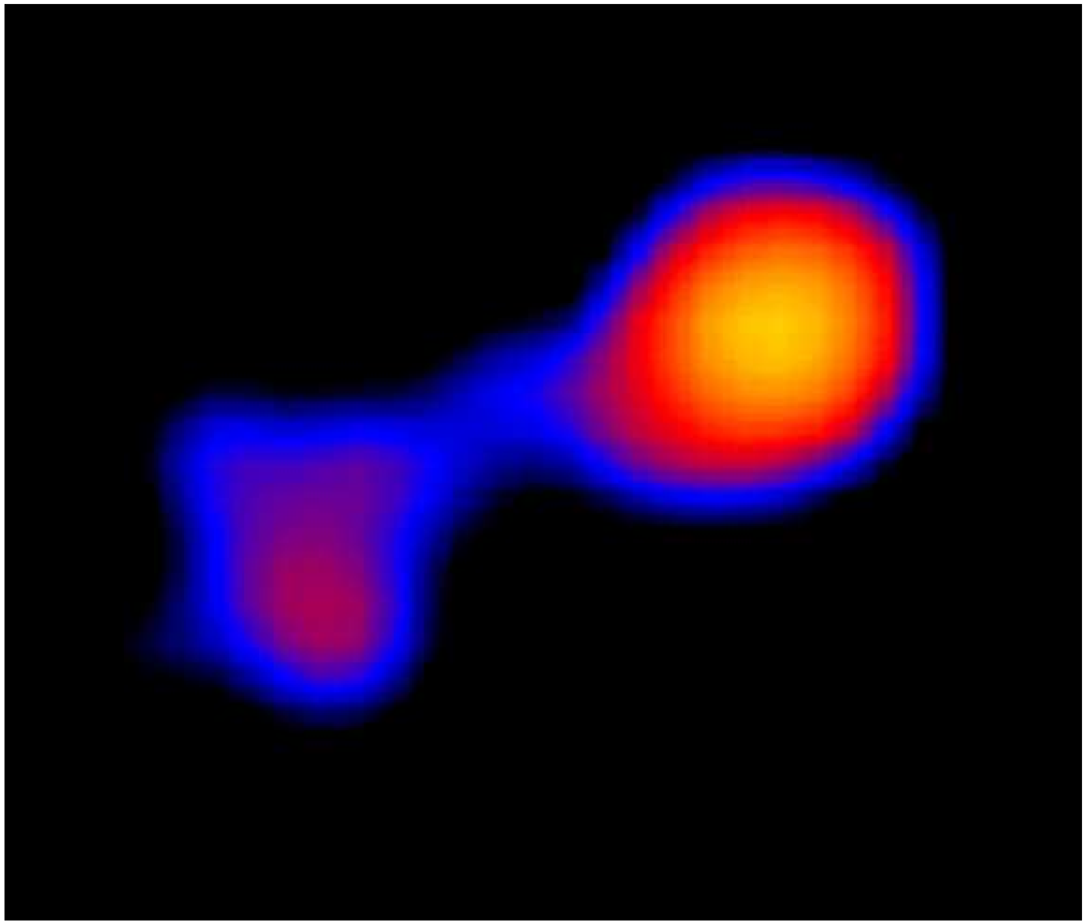
original — ***EMC2* image**



EMC2 significance map: ***3 sigma*** — ***EMC2*** significance map: ***1 sigma***

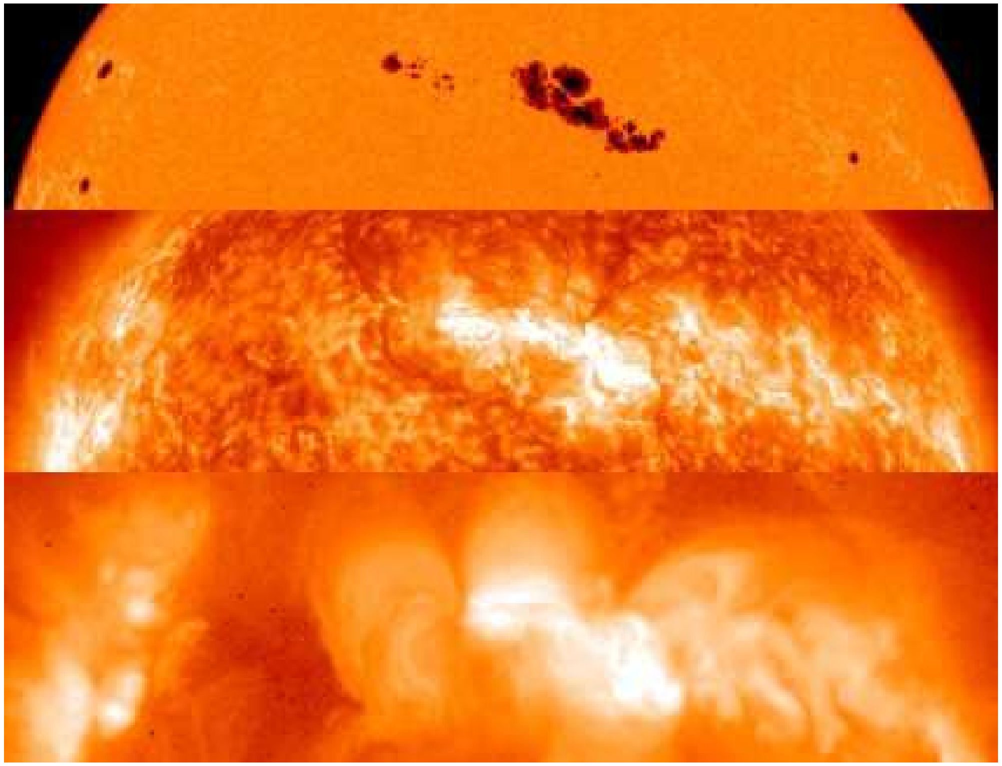
Mira: The Wonderful Star

An EMC2 image on "Astronomy Picture of the Day" (May 5, 2005)



Credit: X-ray Image: M. Karovska (Harvard-Smithsonian CfA) et al., CXC / NASA

Exam. II: The BRoaDEM package for DEM Reconstruction

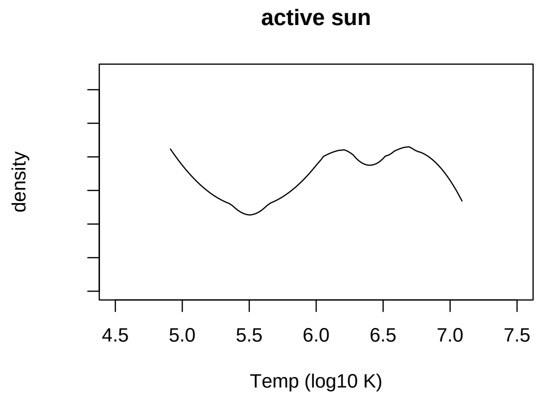


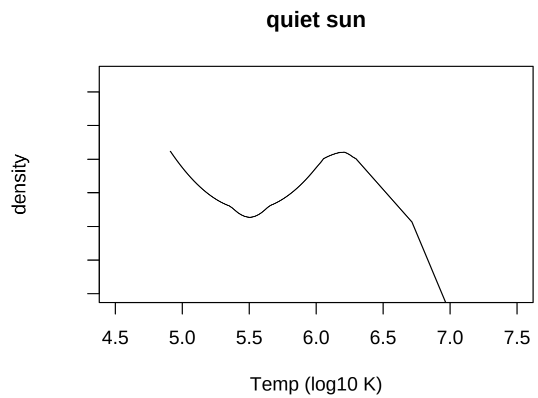
March 2001: Largest Sunspot Group in a Decade.

- Optical, Extreme UV, and X-ray Images
- Reveal different layers of atmosphere
- Higher Energy Emission
  $\rightarrow$ Hotter source
  $\rightarrow$ Extended atmosphere
- X-ray: Hot plasma arching high above the solar surface inside the loops of magnetic fields.

The complex structure in the X-ray emission across the solar corona is a tracer of the temperature and density of the plasma.

The Environment of the Solar Corona





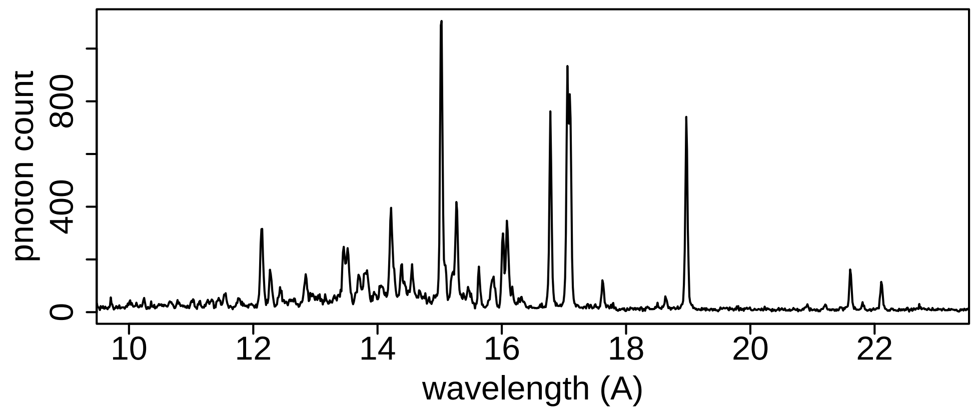
Temp density of coronal plasma
(DEM: Diffuse Emission Measure)

element	abundance (%/%H)
H	1.00000
He	0.07943
C	0.00039
N	0.00010
O	0.00077
Ne	0.00012
Mg	0.00014
Al	0.00001
Si	0.00013
S	0.00002
Fe	0.00013

There is MUCH less information available for stellar corona.

Data for a Stellar Corona

- No star except the sun can be imaged.
- Ultra-high resolution spectral data is available from the *Chandra X-ray Observatory*.



- *Chanda* counts photons in a large number of narrow spectral bins.

Unlocking the information in this forest of spectral lines requires subtle statistical analysis.

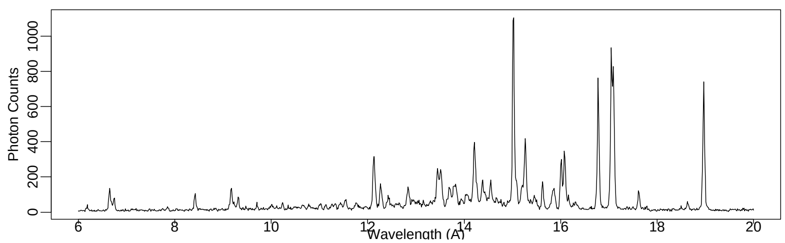
Physics of a Stellar Corona

- A stellar corona is 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;
  - 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.

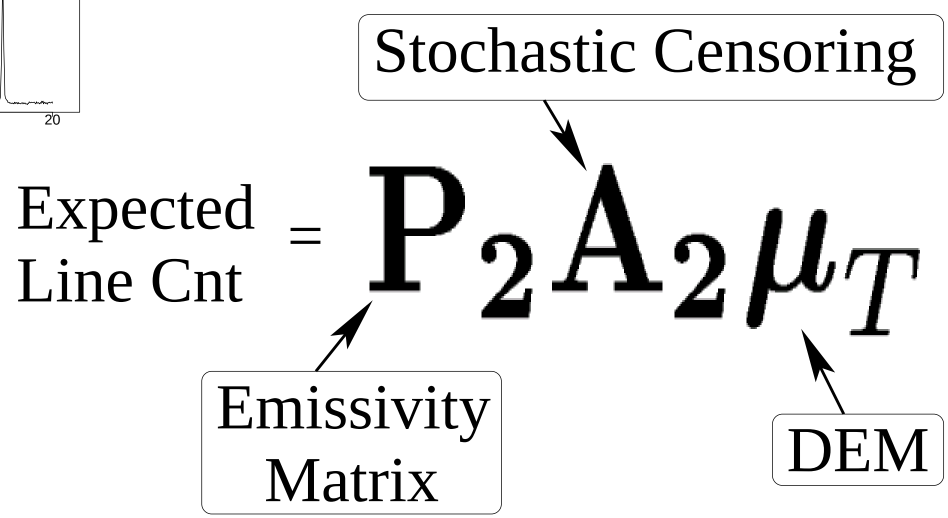
Each line in the forest can be identified with a particular ion, and thus we obtain information on the environment in a stellar corona.

Reconstructing the DEM

- DEM = marginal distribution of temperature
- A given ion at a given temperature emits a known distribution of X-rays
- X-rays appear as a forest of lines, representing a mixture of ions & temps

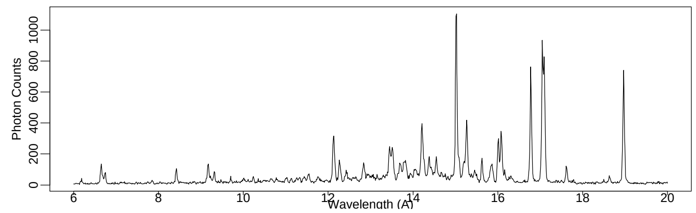


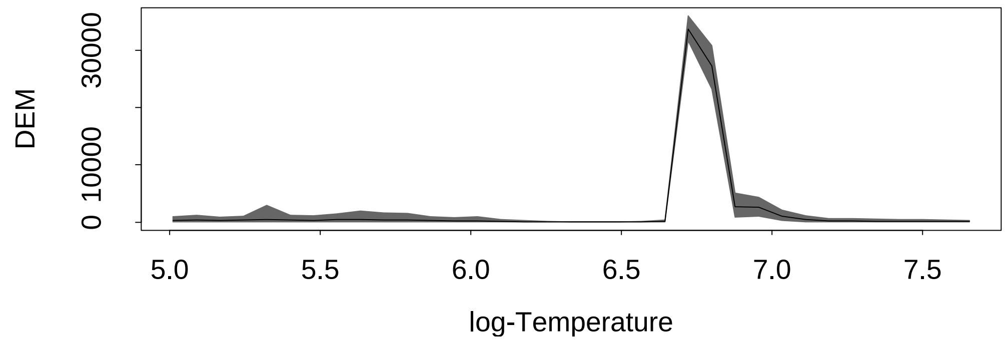
$$\boldsymbol{\lambda} = \mathbf{P_1 A_1}\left(\boldsymbol{\mu}_C + \mathbf{P_2 A_2}\boldsymbol{\mu}_T\right) + \boldsymbol{\xi}$$



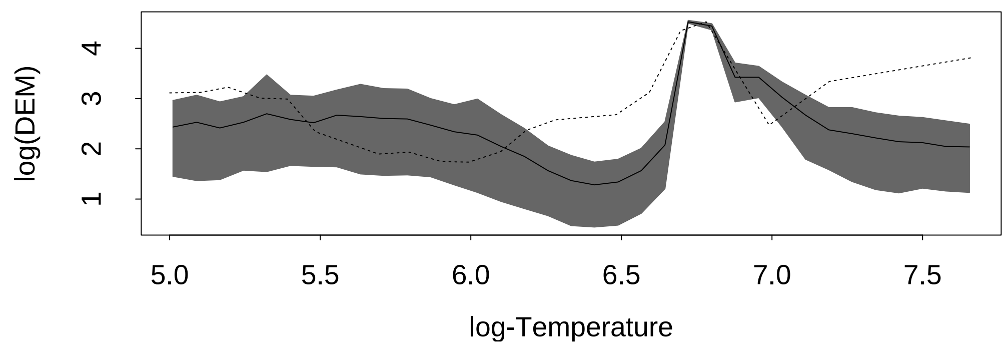
Reconstructing Capella's DEM

Capella is an X-ray bright star.





Reconstruction:



Examples 3 and 4: Spectral Analysis and Hardness Ratios

High-Resolution Spectra

- High resolution detectors such as those aboard *Chandra* herald a quantum leap forward for empirical high-energy astronomy
- Unfortunately, standard methods (e.g., χ^2 fitting) rely on Gaussian assumptions and thus require a minimum count per bin.
- Ad-hoc procedures that group bins are wasteful and sacrifice the desirable high-resolution inherent in the data.

Hardness-Ratios

- A rough summary of a spectrum is a comparison of the expected hard and expected soft counts.
- This is the lowest resolution spectral analysis, but can be useful for classifying faint sources.
- Again, the validity of standard methods depends on Gaussian assumptions.
- For faint sources either the hard or soft counts can be very small.

Solution: Poisson Statistics

- Rather than basing statistical techniques on Gaussian assumptions, we can use the Poisson Distribution as a statistical model for low-count data.
- Specifically, we replace the Gaussian likelihood with a Poisson likelihood:

$$\text{Gaussian Likelihood:} \qquad -\sum_{\text{bins}} \sigma_i - \sum_{\text{bins}} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$

$$\text{Poisson Likelihood:} \qquad -\sum_{\text{bins}} \mu_i + \sum_{\text{bins}} x_i \log \mu_i$$

- Bayesian Methods combine the likelihood with a prior distribution that can
  - Model the dist'n of spectral characteristics in a population of sources.
  - Include information from outside the data as to the spectral shape.
  - Smooth the reconstructed spectrum.

Requires Sophisticated Statistical Computing.

BLoCXS

B*ayesian fitting of* **Hi***gh* **R***esolution* **X***-ray* **S***pectra.*

BLoCXS Functionality

- Uses Poisson models and no Gaussian assumptions. Thus, BLoCXS has no trouble with low count data.
- Corrects for instrument response as quantified by `.rmf` or `.rsp` files.
- Corrects for effective area using `.arf` files.
- Uses a Poisson model-based strategy to correct for background contamination. There is no background subtraction and no negative counts.
- Can fit absorption due to the ISM or IGM.
- Allows for (broken) powerlaw, bremsstrahlung, and blackbody continuums.
- Can include Gaussian, Lorentzian, and delta function emission lines.
- Can compute principled p-values to test for emission lines.
- An extension that will allow for pile-up correction is under development.

BLoCXS Availability: Scheduled for release in the next version of CIAO.

Principled P-values to Test for a Model Component

Fallible F-tests

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

- The LRT is valid for comparing nested models. But the smaller model's parameter must be in the interior of the larger model's parameter space.
- *This is not the case when testing for a model component in a spectral model. The F-test is not properly calibrated for this problem.*
- 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 paper has already been cited 44 times.**

BEHR

B*ayesian* **E***stimation of* **H***ardness* **R***atios.*

BEHR Functionality

- BEHR uses Poisson models background contaminated soft and hard counts. Thus, BEHR has no trouble with low count data.
- BHER computes hardness ratio estimates and intervals with reliable frequency properties. (See simulation study.)

BEHR Availability

- BEHR will soon be available on the CXC contributed software page (`cxc.harvard.edu/cont-soft/soft-exchange.html`).

BEHR Examples and References

van Dyk, D. A. et al. (2005). Deconvolution in High-Energy Astrophysics: Science, Instrumentation, and Methods. *Bayesian Analysis*, to appear.

Park, T., van Dyk, D. A., Kashyap, V. L., & Zezas, A. (2004). Computing Hardness Ratios with Poissonian Errors. *CHASC Technical Report*.

Verifying BEHR

Simulation Study

- $S = H = 3$; each with expected background contamination = 0.1.
- Background exposure is 100 times longer.
- $R = S/H, \qquad HR = (H - S)/(H + S), \qquad C = \log_{10}(R)$

Table 1: Coverage of Bayesian and Standard Methods.

Method	Hardness Ratio	True Value	Coverage Rate	Mean Length	Mean Square Error by mode	Mean Square Error by mean
BEHR	R	1	95.0%	7.30	0.59	12.34
	HR	0	91.5%	1.23	0.53	0.42
	C	0	98.0%	1.53	0.42	0.46
Standard Method	R	1	96.5%	138.29	73.58	
	HR	0	99.5%	3.44	0.63	
	C	0	100.0%	7.26	5.58	

Outline of Presentation

This talk has three components:

A. Highly Structured Models in High-Energy Astrophysics

- Astrostatistics:

  Complex Sources, Complex Instruments, and Complex Questions

  Key: All three are the domain of Astrostatistics

- Model-Based Statistical Solutions
- Monte Carlo-Based Bayesian Analysis

B. Examples

1. The EMC2 package for Image Analysis (A detailed example.)
2. The BRoaDEM package for DEM Reconstruction
3. The BLoCXS package for Spectral Analysis
4. The BEHR package for computing Hardness Ratios

C. **Using Incompatible Conditional Distributions in Gibbs Samplers**

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:

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

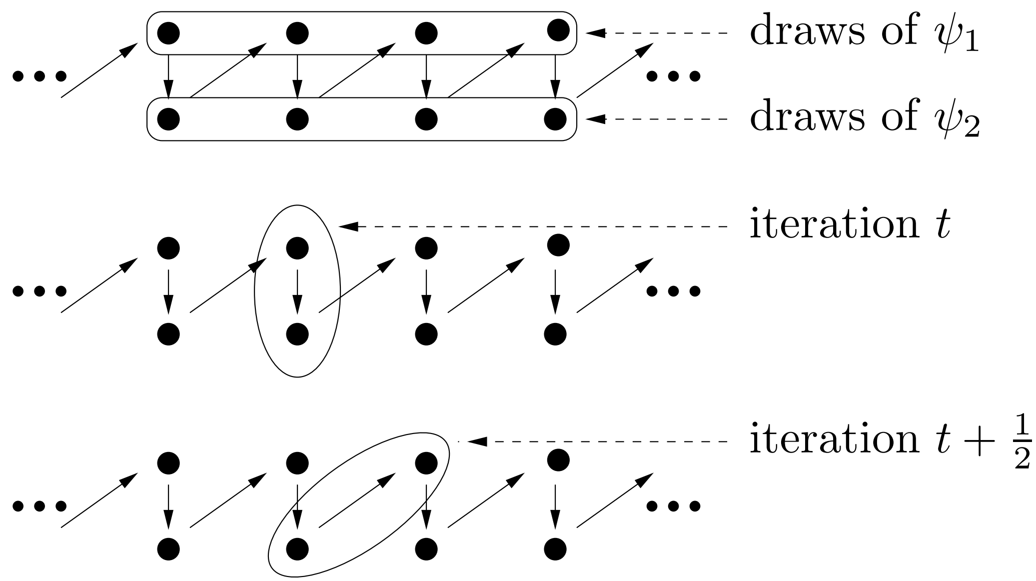
- Questions:
  1. Does the resulting Markov chain have a stationary distribution?
  2. If so, what is it?
  3. Why use such a chain?
- I cannot fully answer these questions, but can offer tantalizing examples....

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,

AND is "quick" to converge!

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

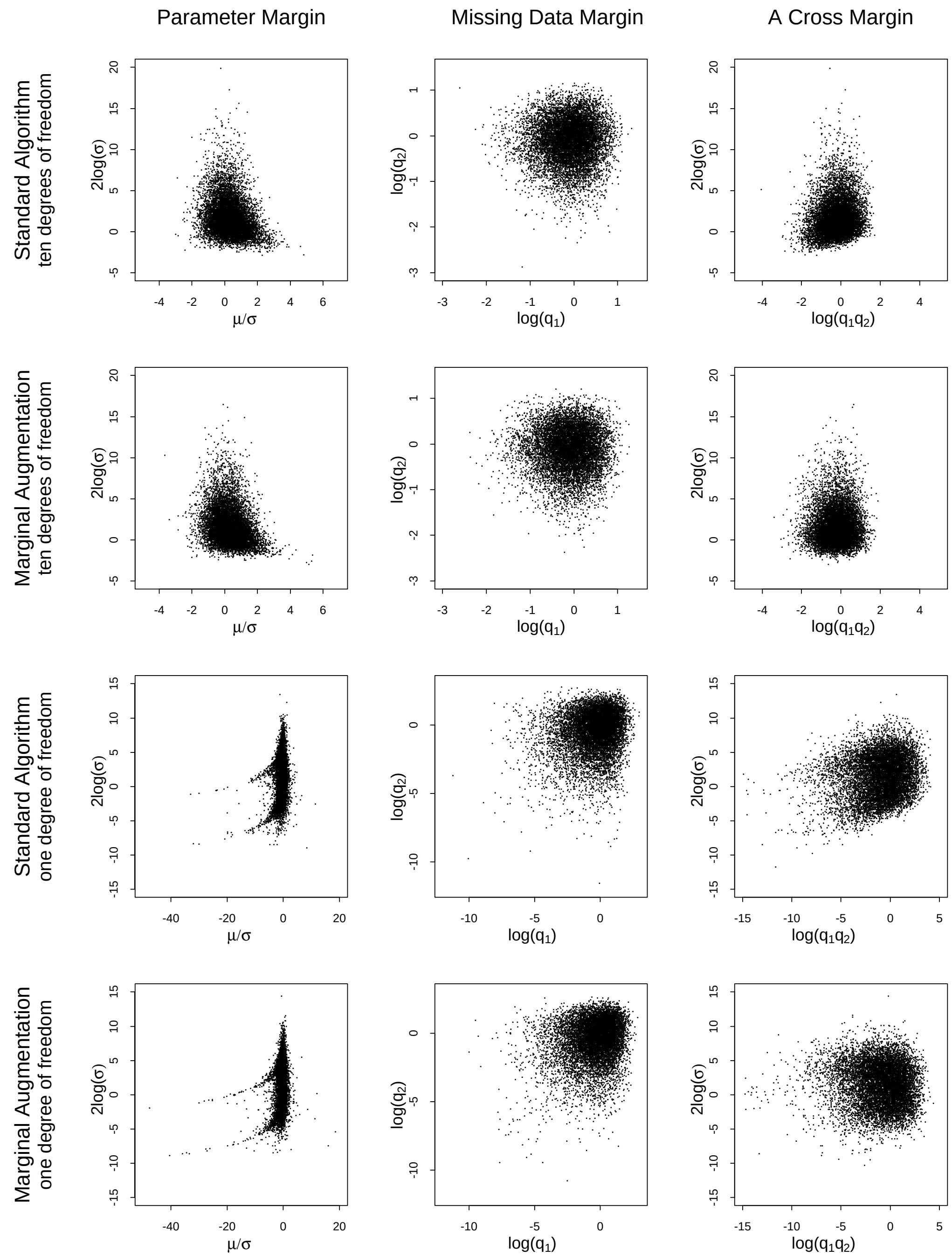
Empirical Illustration with a t model

• *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). • We omitted details and

• return to astrophysics...



Back to Astrophysics

- Recall that our (simplified) *latent* Poisson Process,

$$X_i \sim \text{Poisson}\left(\Lambda_i = \lambda_i^{\text{ES}} + \lambda^{\text{PS}} p_i\right).$$

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

$$Z_{il} = \begin{pmatrix} \text{indicator that photon } l \text{ in cell } i \\ \text{corresponds to the point source} \end{pmatrix}$$

1. Given $Z = \{Z_{il}\}$ we can sample $\theta = \{\lambda^{\text{PS}}, (\lambda_i^{\text{ES}}, p_i)\}$
2. Given θ we can sample Z, via $Z_{il} \sim \text{Ber}\left(\frac{\lambda^{\text{PS}} p_i}{\lambda_i^{\text{ES}} + \lambda^{\text{PS}} p_i}\right)$

- We sometimes construct a *delta function* point source model so as

1. the point source is contained entirely in one pixel, but
2. we do not know which pixel.

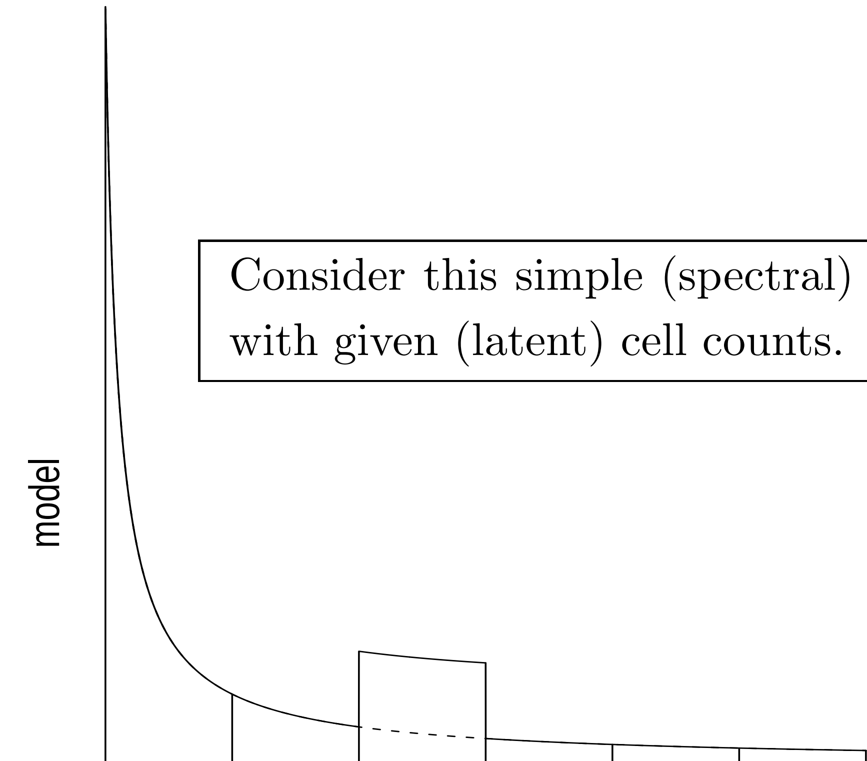
i.e., $\{p_i\}$ can be parameterized in terms of a single unknown parameter,

$$\theta^{\text{L}} = \text{the location of the point source.}$$

In This Case Data Augmentation Fails.

Why Data Augmentation Fails

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

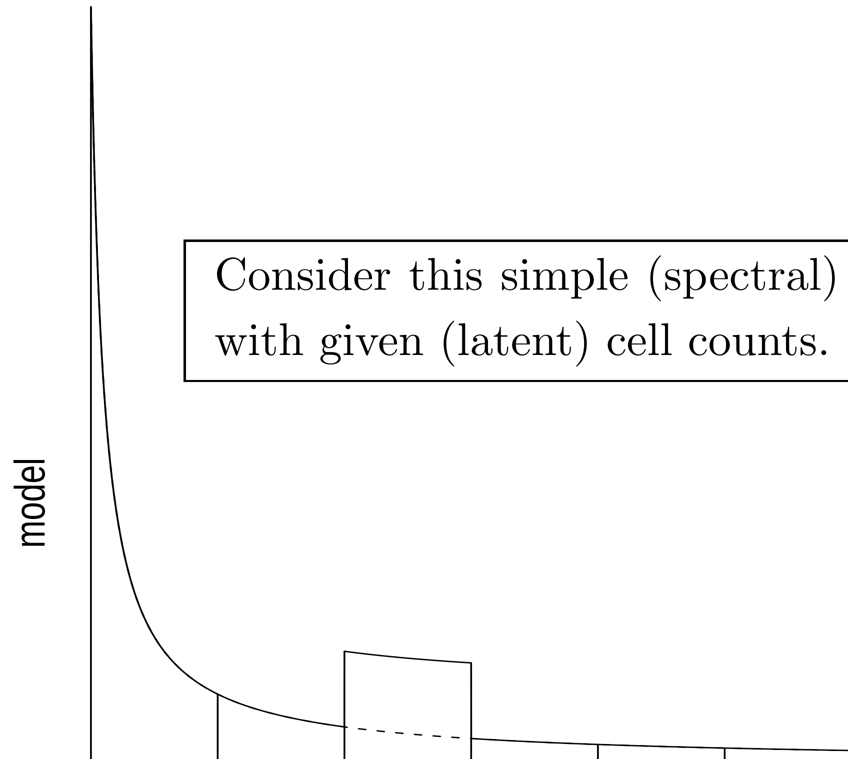


X = (latent) Cell Counts	10	4	8	1	2	0
Extend Src Counts(Z=0)						
Point Src Counts (Z=1)						

Given this Model, what is Z?

Why Data Augmentation Fails

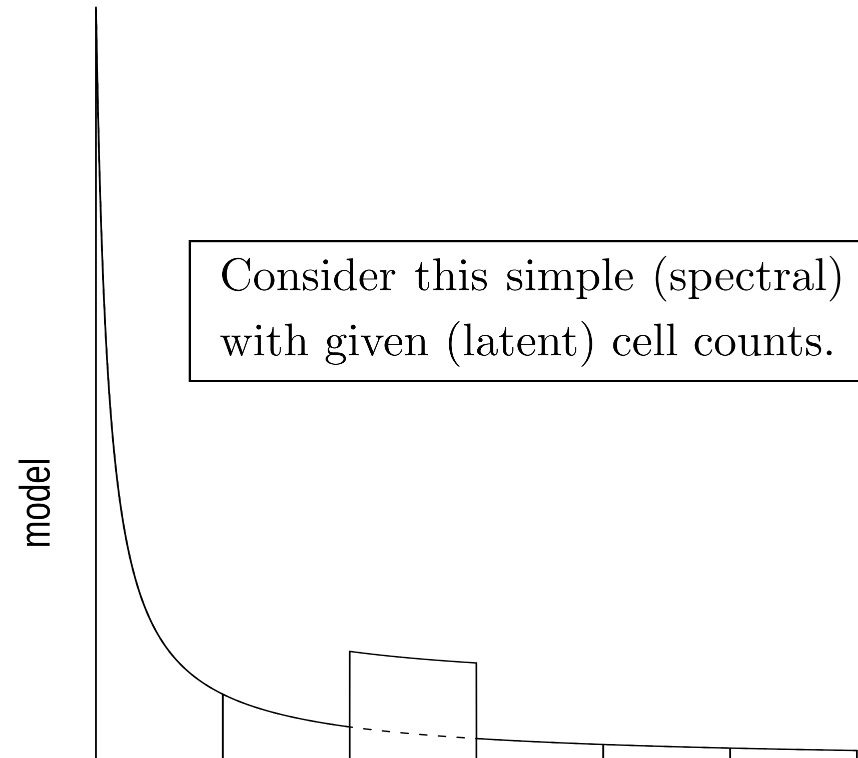
Consider this simple (spectral) model with given (latent) cell counts.



X = (latent) Cell Counts	10	4	8	1	2	0
Extend Src Counts(Z=0)	10	4	~3	1	2	0
Point Src Counts (Z=1)	0	0	~5	0	2	0

Why Data Augmentation 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)	10	4	~3	1	2	0
Point Src Counts (Z=1)	0	0	~5	0	2	0

Given Z, what is the location of the point source?

The Standard Gibbs Sampler

Recall we do not observe the latent Poisson Process,

$$X_i \sim \text{Poisson}\left(\Lambda_i = \lambda_i^{\text{ES}} + \lambda^{\text{PS}} p_i\right),$$

Rather we observe, $Y_j \sim \text{Poisson}\left(\alpha_j \sum_i M_{ij}\Lambda_i + \theta_j^{\text{B}}\right)$

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

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":

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

- Following this we construct:

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

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

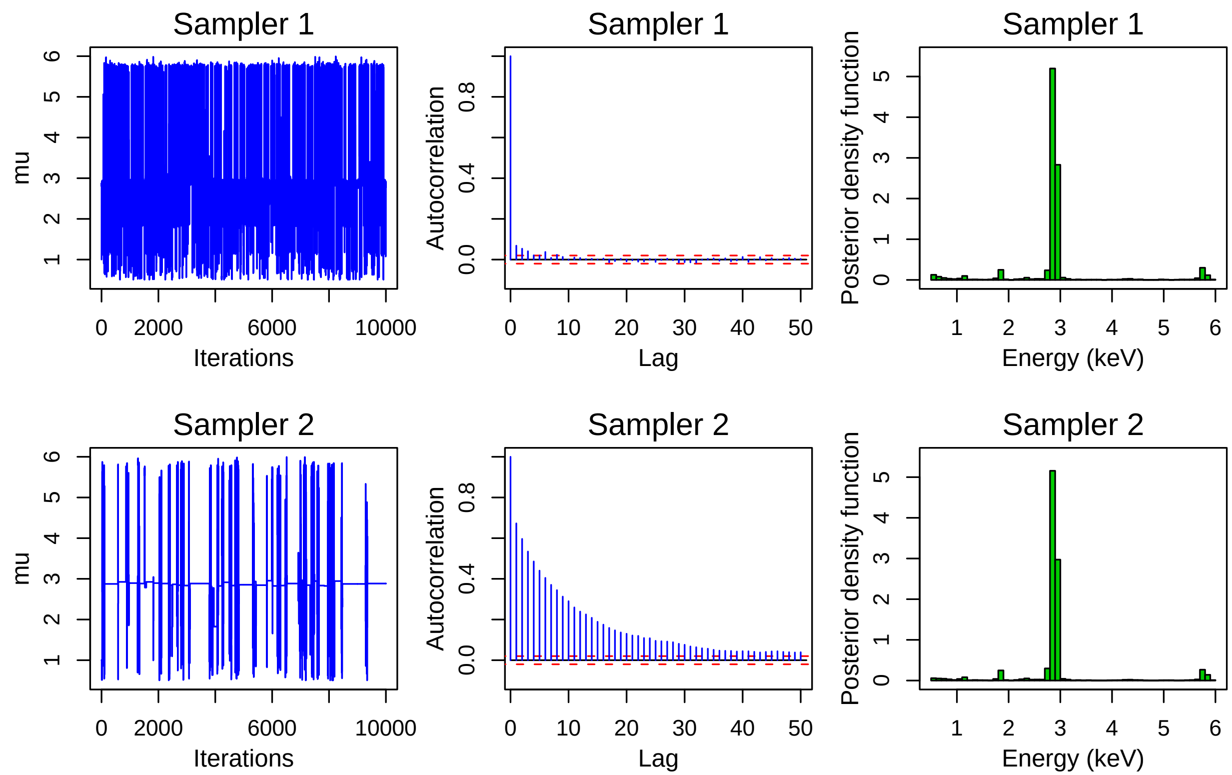
  Sampler 2: (Cannot be Blocked: An Incompatible Gibbs Sampler.)

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

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.



Verifying the Stationary Distribution of Sampler 2

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

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.

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

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

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

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.

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

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*, to appear.

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.

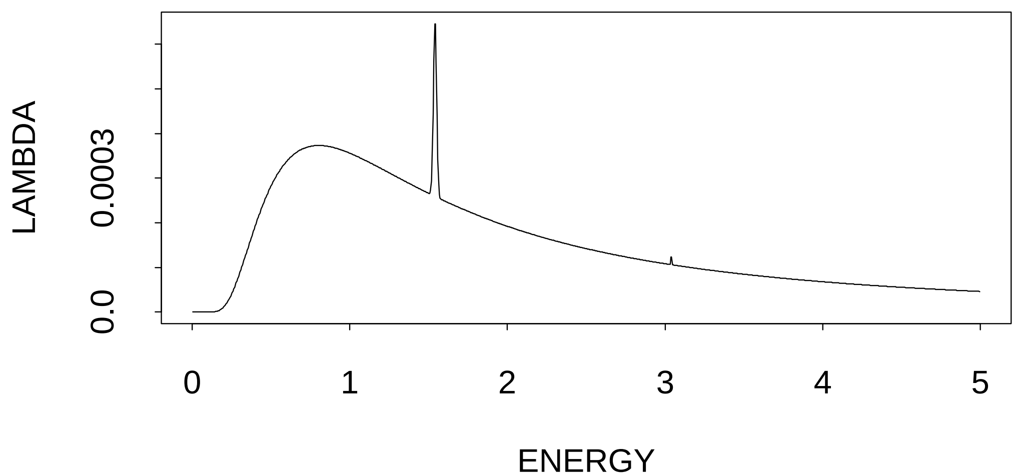
van Dyk, D. A. and Kang, H. (2004). Highly Structured Hierarchical Models for Spectral Analysis in High Energy Astrophysics. *Statist. Science*, 19, 275–293.

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.

More on Example 3: 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,
  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



A Bayesian Spectral Analysis

Quasars

- Among the most distant distinct detectable objects.
- Believed to be super massive black holes with mass a million times that of the sun.
- Give glimpse into the very distant past, perhaps 90% of the way to Big Bang.

High Red-Shift Quasar PG1637+706

- Red-shift: wavelengths elongated as object moves away: energy appears lower
- By measuring the change in energy, we can recover the recession velocity, and in a uniformly expanding universe, the velocity is a direct measure of distance.

The Spectral Model

Model

- Power Law Continuum: $f(\theta^C, E_j) = \alpha^C E_j^{-\beta^C}$
- Absorption model of Morrison and McCammon (1983) to account for the ISM and IGM.
- Power Law for Background counts: $f(\theta^B, E_j) = \alpha^B E_j^{-\beta^B}$
- Narrow Gaussian Emission Line ($\sigma = 0.125$ keV)

Three Models for the Emission Line:

MODEL 0: There is no emission line.

MODEL 1: There in an emission line with fixed location in the spectrum but unknown intensity.

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

Finding the Spectral Line

EM Algorithm

- Refit with 51 starting values for the line location between 1.0 and 6.0 keV
- ML estimate agrees with scientific expectation (between 2.74 and 2.87 keV)

Results

mode (keV)	domain of convergence (keV)	loglikelihood
1.059	1.0-1.3	2589.31
1.776	1.4-2.0	2590.37
2.369	2.1-2.3	2590.19
2.807*	2.4-3.7	2594.94
4.216	3.8-4.7	2589.57
5.031	4.8-5.2	2589.31
5.715	5.3-5.9	2589.74

Maximum loglikelihood for the model with no line: 2589.31.

Sampling the Major Mode

- A Gibbs sampler can sample the major posterior mode.
- Compute estimates, error bars, and correlations.

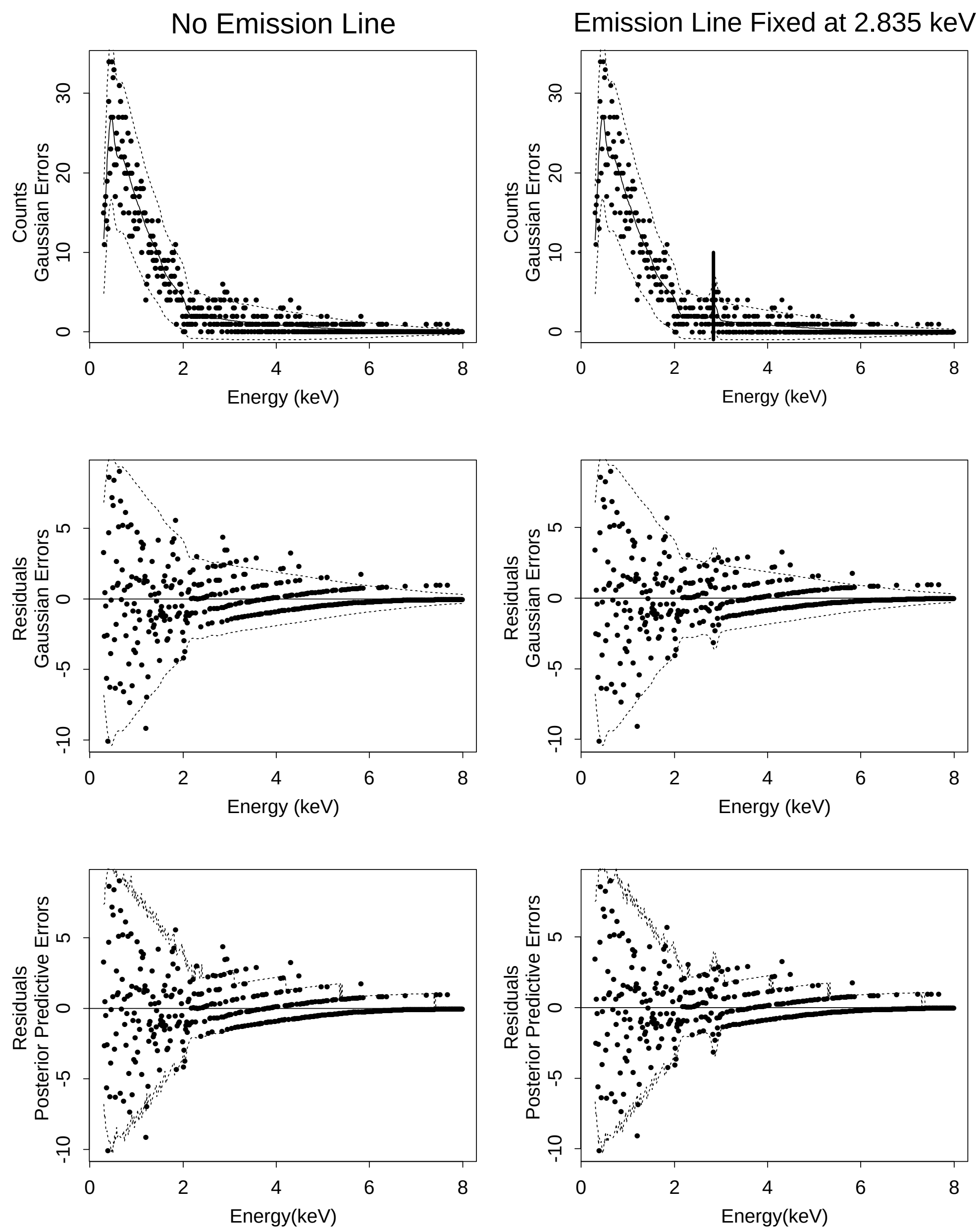
	2.5%	median	97.5%	mean
α^C	3.499e-04	3.890e-04	4.317e-04	3.895e-04
β^C	1.15683	1.34854	1.53951	1.34819
θ^A	-1.13618	-0.72117	-0.30594	-0.7213
α^B	-0.72395	-0.25793	0.14292	-0.26616
β^B	-1.32096	-0.92721	-0.52515	-0.92561
$\theta^L_{1,\lambda}$	33.9036	104.127	205.525	107.831158
$\theta^L_{1,\mu}$	2.65657	2.7948	2.9422	2.79551

GO TO NEXT SLIDE!

Model Diagnostics

Residual Plots

- Gaussian Errors
- Posterior Predictive Errors

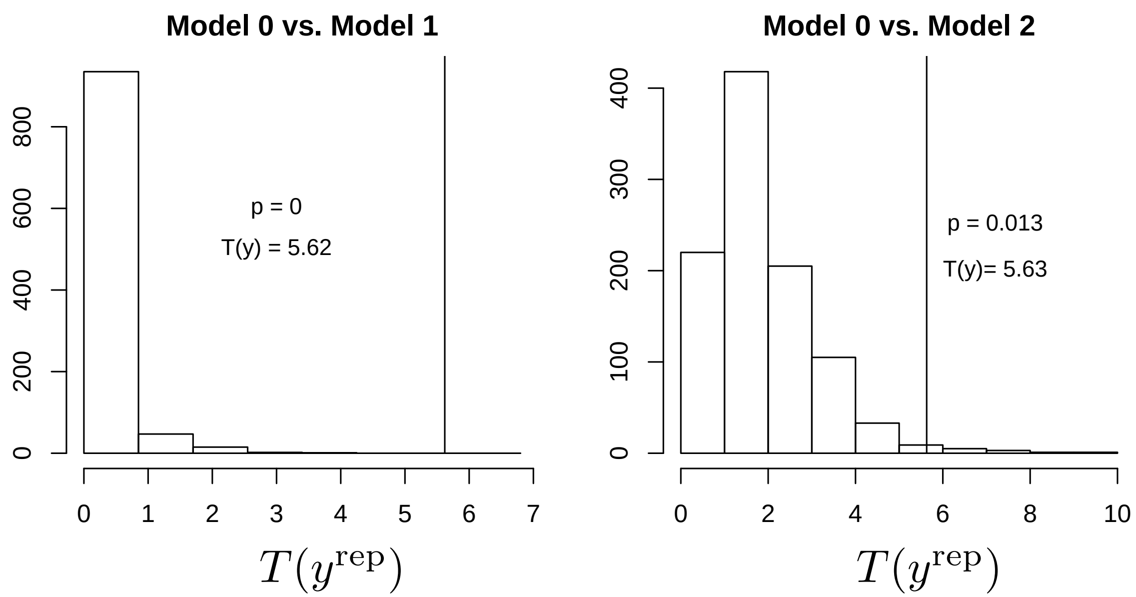


.

Model Checking

Posterior Predictive Checks

- The Likelihood Ratio Test: $T(y_{\text{rep}}) = \log \left\{ \frac{\sup_{\theta \in \Theta_i} L(\theta | y_{\text{rep}})}{\sup_{\theta \in \Theta_0} L(\theta | y_{\text{rep}})} \right\}, \quad i = 1, 2,$
- Sample y_{rep} from posterior predictive distribution under model *Model 0*.



Given the prior belief that the line is near 2.81 keV, it is legitimate to use the first ppp-value. Without such prior information, one should use the second ppp-value.

Conclusions

Motivation for Model Based Methods:

- Asymptotic approximations may not be justified
  - χ^2 fitting is not appropriate for **low counts**
- Accounting for background contamination
- Accounting for pile up

Motivation for Bayesian Methods:

- Likelihood methods also require asymptotic approximations (e.g., to compute error bars) which may not be reliable
- Testing for spectral or spatial features
- Computation for mode finders may be intractable

The Future of Data Analysis:

- Problem specific modeling and computing
- Less reliance on statistical black boxes and multi-purpose solutions