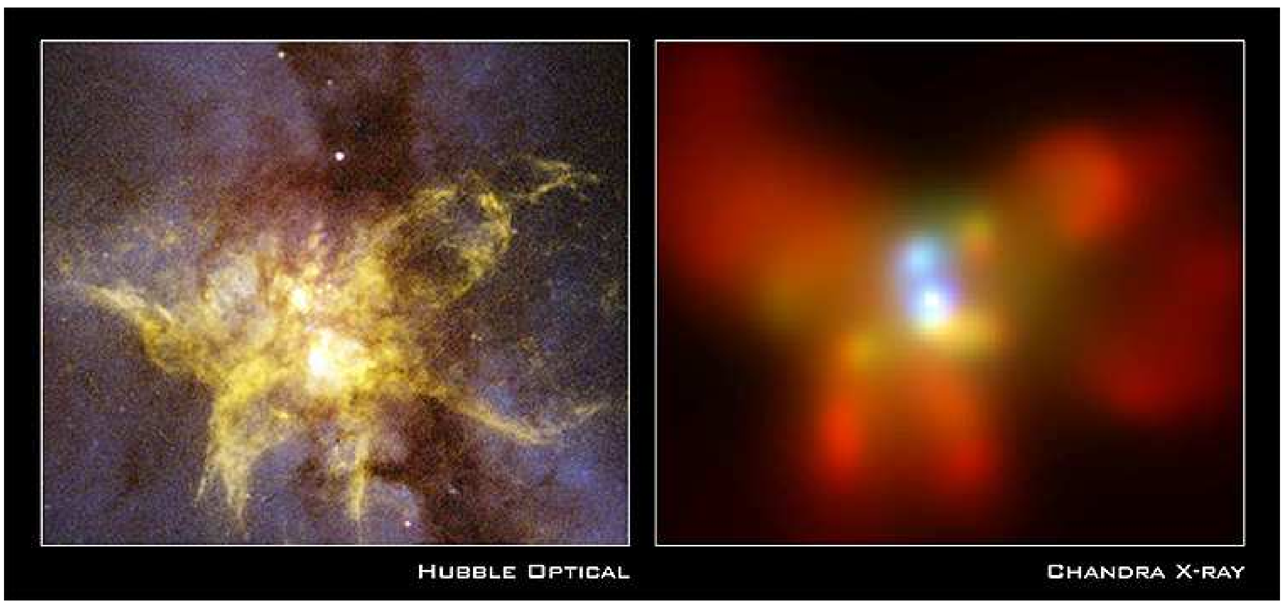
Going Beyond Compatibility: The Future of the Gibbs Sampler?

David van Dyk

Department of Statistics
University of California, Irvine

Joint work with
Xiao-Li Meng and Taeyoung Park



Outline of Presentation

This talk has three components:

A. **Working Parameters and Marginal Augmentation**

- Improves the convergence of Gibbs Samplers *(no details here)*.
- Show how this takes advantage of *incompatible conditional distributions.*
- Two joint distributions with the same margins, the target posterior distribution and a joint distribution with less correlation.

B. Can we use this idea to do *better than i.i.d. sampling?*

C. Extension: Reduce conditioning is selected steps of a Gibbs Sampler.

- This improves convergence but may introduce incompatibility.
- Extended example from Astronomy.

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....

Working Parameters

(Posterior) distribution of interest: $p(\psi)$.

Introducing a *working parameter* α:

$$p(\psi, \alpha) = p(\psi)p(\alpha).$$

Using a transformation $\tilde{\psi} = D_\alpha(\psi)$, we construct a sampler using conditional distributions of

$$p(\tilde{\psi}, \alpha).$$

Meng and van Dyk (*Biometrika*, 1999) showed that for *two-step* samplers with $\psi = (\psi_1, \psi_2)$ and $\tilde{\psi} = (\tilde{\psi}_1, \psi_2)$ the rate of convergence of $\{\psi_2^{(t)}, t = 1, 2, \ldots\}$:

STEP 1: $(\tilde{\psi}_1, \alpha) \sim p(\tilde{\psi}_1, \alpha | \psi_2)$

STEP 2: $(\psi_2, \alpha) \sim p(\psi_2, \alpha | \tilde{\psi}_1)$

is no worse than that of the standard two step sampler based on $p(\psi_1, \psi_2)$.

"Marginal Augmentation" can be very fast in practice

(van Dyk and Meng,*JCGS*, 2001).

A Simple Gaussian Example

Suppose $\psi \sim N_2(\mu, \Sigma)$.

Two Samplers:

1. We can construct a simple Gibbs sampler:

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

   - Lag-one autocorrelation for ψ_2 is ρ^2.

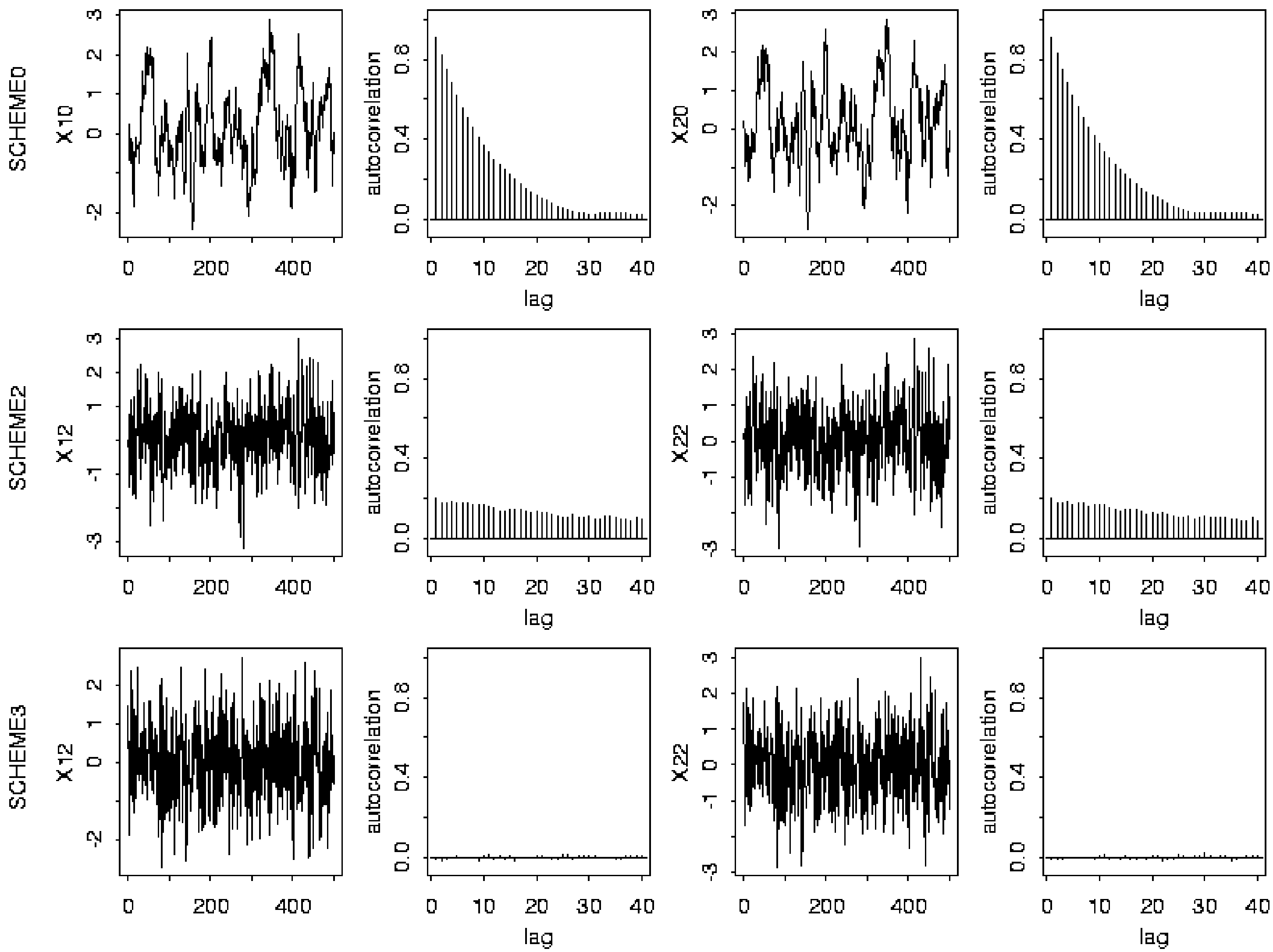
2. Introduce a *working parameter*: $\alpha \sim N(0, \omega^2\sigma_1^2)$.

   - Define $\tilde{\psi}_1 = \psi_1 + \alpha$.
   - A new (working parameter) sampler:

$$\tilde{\psi}_1 \sim p(\tilde{\psi}_1|\psi_2, \alpha) \quad \text{and} \quad (\psi_2, \alpha) \sim p(\psi_2, \alpha|\tilde{\psi}_1)$$

   - Lag-one autocorrelation: $\rho^2/(1 + \omega^2)$

Computational Gain



Why Include a Working Parameter?

FAST algorithms that are EASY to Implement!

Now many examples: t-models, probit regression, mixed-effects models, Poisson imaging, generalized linear mixed models, multinomial probit model, etc.

EM-type Algorithms: van Dyk and Meng (JRSS-B, 1997, 1998); van Dyk (*JCGS*, 2000; *Stat. Sinica*, 2000); van Dyk and Tang (*Comp. Statist.*, 2003).

Samplers: van Dyk and Meng (*Biometrika*, 1999; *JCGS*, 2001) Imai and van Dyk (*J of Econometrics*, 2005), van Dyk, Meng, and Kang (in prep).

The choice of the working prior distribution:

- More diffuse prior distributions tends to be better.
- Improper working prior distribution? $p(\psi, \alpha) = p(\psi)p(\alpha)$
- The resulting Markov chain in NOT positive recurrent.
- With careful choice of the working prior distribution:

1. $\{\psi_2^{(t)}, t = 1, \ldots\}$ is Markovian.
2. The stationary distribution of this chain is $p(\psi_2)$.

The Transition Kernel for ψ_2

Gaussian Example with Improper Working Prior Distribution:

STEP 1: $\tilde{\psi}_1 | \psi_2', \alpha' \sim \mathrm{N}(\rho\psi_2' + \alpha', 1 - \rho^2)$

STEP 2: $\alpha | \tilde{\psi}_1 \sim \mathrm{N}(\tilde{\psi}_1, 1)$ and $\psi_2 | \tilde{\psi}_1, \alpha \sim \mathrm{N}(-\rho(\alpha - \tilde{\psi}_1), 1 - \rho^2)$

Equivalently,

STEP 2*: $\psi_2 | \tilde{\psi}_1 \sim \mathrm{N}(0, 1)$.

Since Step 2* doesn't depend on α', we set $\alpha' = 0$ in Step 1.

Thus, by using α with $\omega \to \infty$, the Gibbs Sampler:

STEP 1: $\psi_1 | \psi_2' \sim \mathrm{N}(\rho\psi_2', 1 - \rho^2)$

STEP 2: $\psi_2 | \psi_1 \sim \mathrm{N}(\rho\psi_1, 1 - \rho^2)$

evolves into:

STEP 1: $\psi_1 | \psi_2' \sim \mathrm{N}(\rho\psi_2', 1 - \rho^2)$

STEP 2*: $\psi_2 | \psi_1 \sim \mathrm{N}(0, 1)$

The Result: Incompatible Distributions

Gaussian Example with Improper Working Prior Distribution:

Thus, by using α with $\omega \to \infty$, the Gibbs Sampler:

STEP 1: $\psi_1|\psi_2' \sim \mathrm{N}(\rho\psi_2', 1-\rho^2)$

STEP 2: $\psi_2|\psi_1 \sim \mathrm{N}(\rho\psi_1, 1-\rho^2)$

evolves into:

STEP 1: $\psi_1|\psi_2' \sim \mathrm{N}(\rho\psi_2', 1-\rho^2)$

STEP 2*: $\psi_2|\psi_1 \sim \mathrm{N}(0, 1)$

Notice:

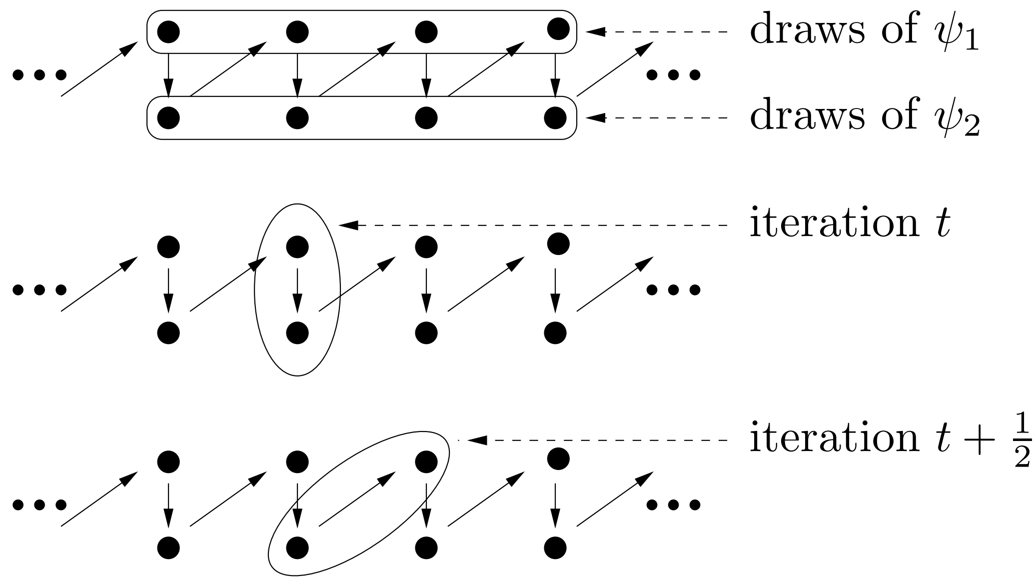
- *Conditional* $p(\psi_2|\psi_1)$ "mutated" into *marginal* $p(\psi_2)$!
- *Worst* choice for compatibility: Under Step 2*, $\psi_1 \perp \psi_2$.
- *Best* for convergence: $\psi_2^{(t)}$ converges in one iteration!
- We are not usually this lucky...

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,

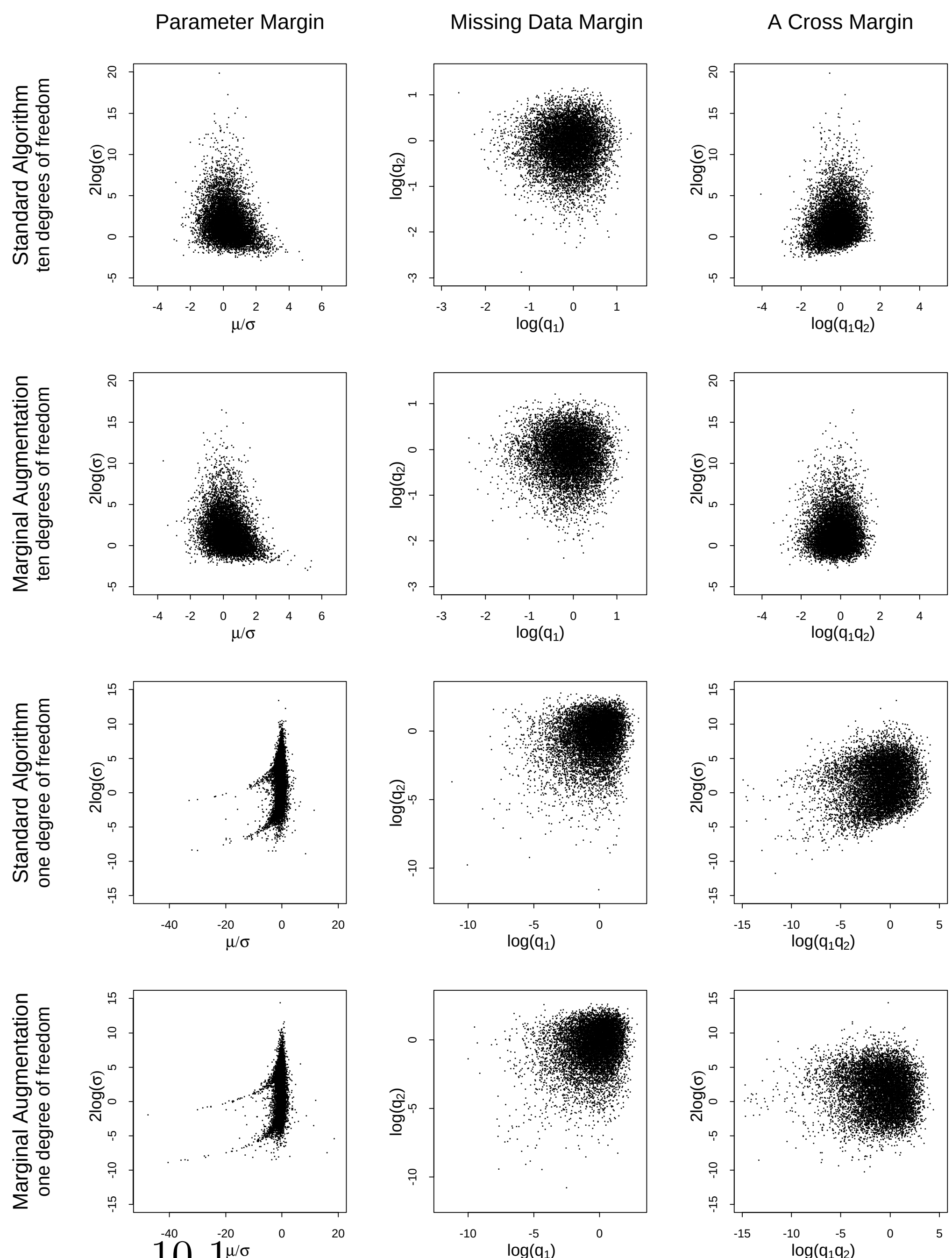
We are mixing the conditional distributions form two different joint distributions with the same marginals. These conditional distributions are incompatible.

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.



Incompatible Gibbs Samplers

A Markov chain constructed using Marginal DA with improper working prior satisfies the conditions of

Theorem: If $\{\psi_2^{(t)}, t = 0, 1, \ldots\}$ has stationary distribution $p(\psi_2)$, $\mathcal{K}(\psi_1|\psi_2) = p(\psi_1|\psi_2)$, but $\mathcal{K}(\psi_2|\psi_1) \neq p(\psi_2|\psi_1)$

(i) $\{\psi_1^{(t)}, t = 0, 1, \ldots\}$ has stationary distribution $p(\psi_1)$;

(ii) $\widetilde{\mathcal{M}} = \{(\psi_1^{(t+1)}, \psi_2^{(t)}), t = 0, 1, \ldots\}$ has the stationary distribution $p(\psi_1, \psi_2)$;

(iii) $\mathcal{M}$ has stationary distribution

$$p(\psi_1)\mathcal{K}(\psi_2|\psi_1) \neq p(\psi_1, \psi_2);$$

(iv) $\mathcal{K}(\psi_1|\psi_2)$ and $\mathcal{K}(\psi_2|\psi_1)$ are not compatible.

This fully describes the behavior of the marginal (two-step) Gibbs sampler:

1. Although $\mathcal{M}$ does not have $p(\psi_1, \psi_2)$ as its stationary distribution, $\widetilde{\mathcal{M}}$ does.
2. The conditional distributions used in the sampler are not compatible, at least with two-step samplers.

Outline of Presentation

This talk has three components:

A. Working Parameters and Marginal Augmentation

- Improves the convergence of Gibbs Samplers *(no details here)*.
- Show how this takes advantage of *incompatible conditional distributions*.
- Two joint distributions with the same margins, the target posterior distribution and a joint distribution with less correlation.

B. **Can we use this idea to do *better than i.i.d. sampling?***

C. Extension: Reduce conditioning is selected steps of a Gibbs Sampler.

- This improves convergence but may introduce incompatibility.
- Extended example from Astronomy.

A Gaussian Dessert

Consider sampling from $N(0,1)$, as a margin of either of

$$\begin{bmatrix}\psi_1\\ \psi_2\end{bmatrix} \sim N\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}1 & \rho_i\\ \rho_i & 1\end{bmatrix}\right), \; i=1,2$$

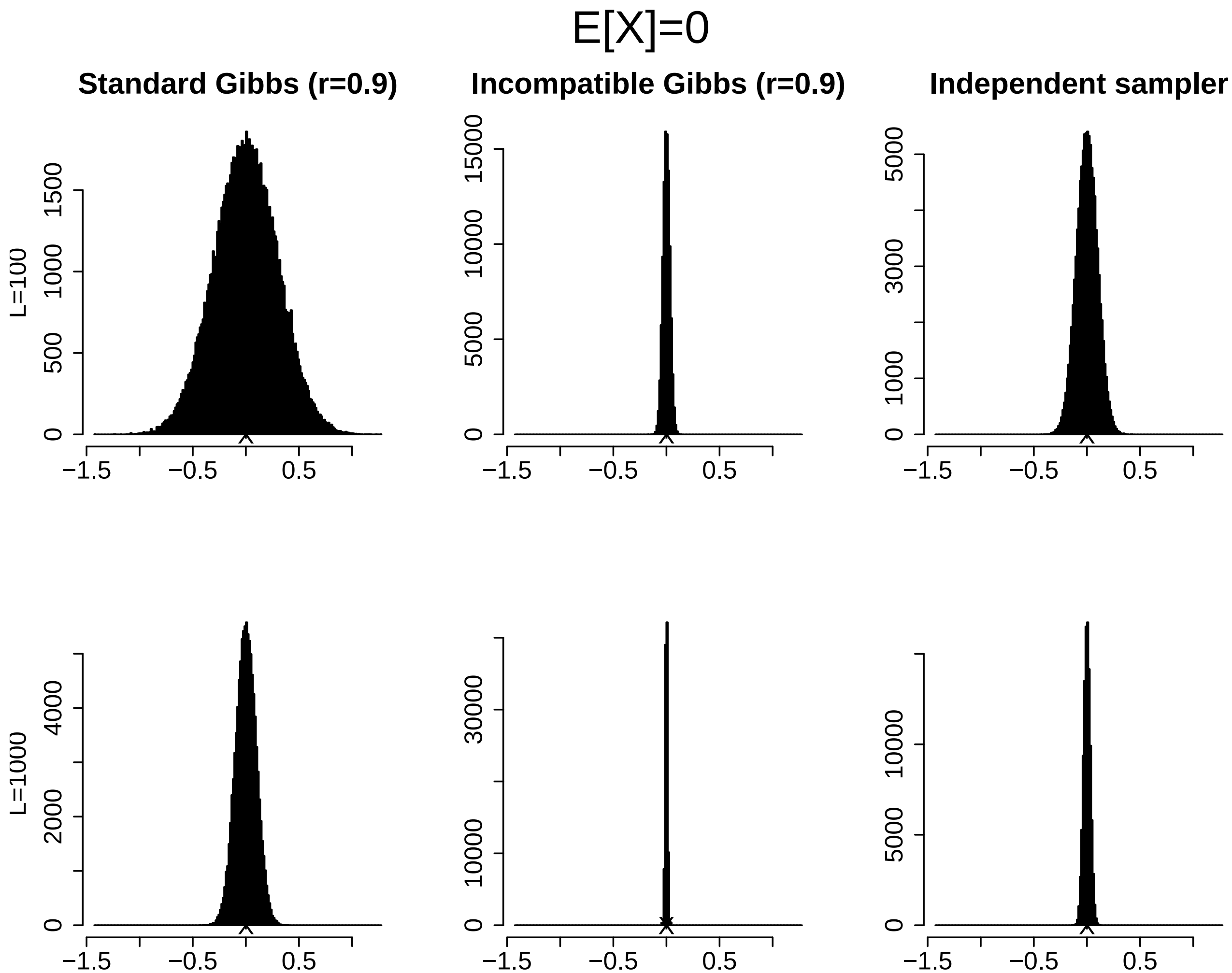
With standard Gibbs Sampler: $\text{Cor}(\psi_1^{(t+1)}, \psi_1^{(t)}) = \rho_i^2$.

An Incompatible Alternative:

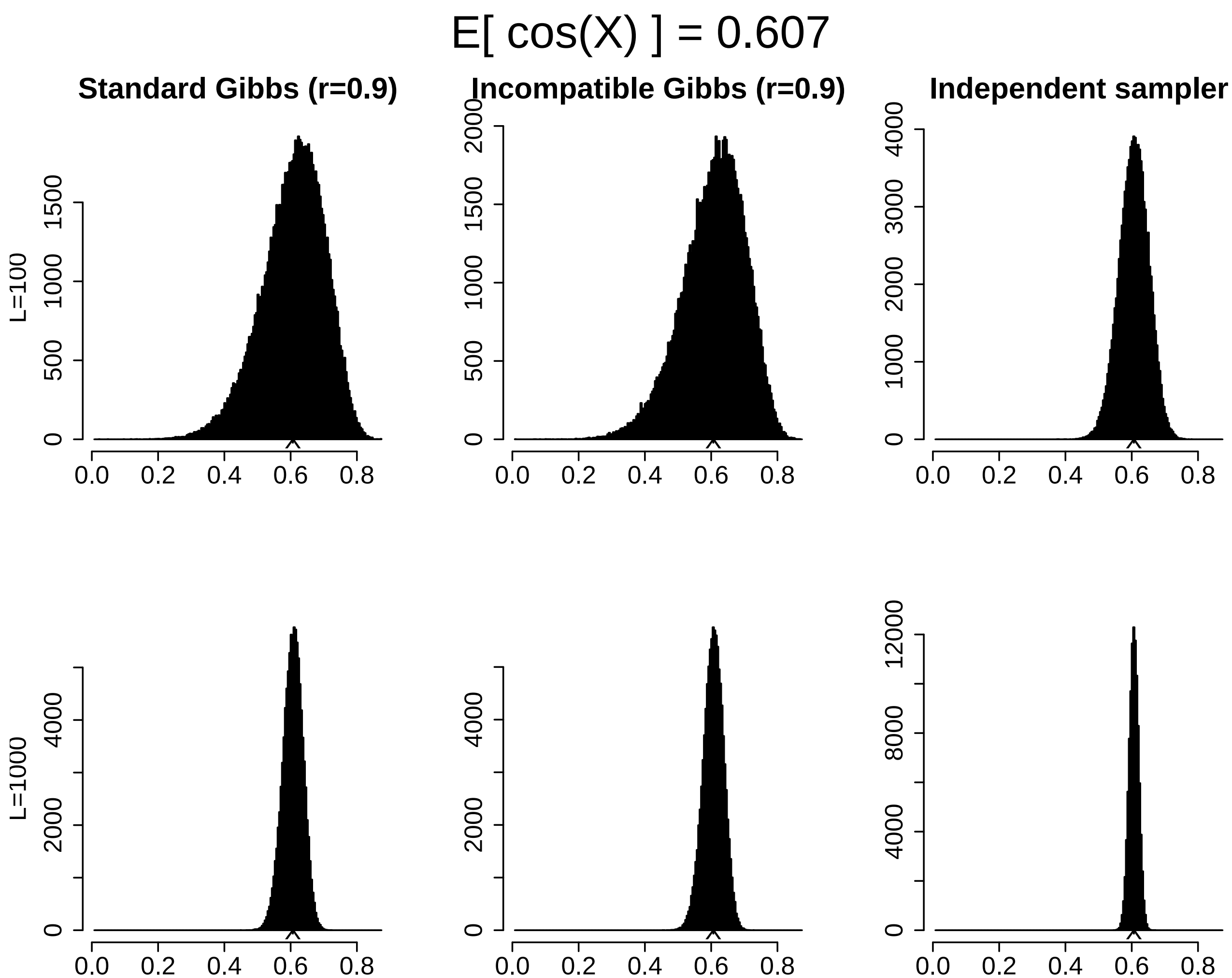
$$\begin{aligned}\psi_1|\psi_2 &\sim \mathrm{N}(\rho_1\psi_2, 1-\rho_1^2)\\ \psi_2|\psi_1 &\sim \mathrm{N}(\rho_2\psi_1, 1-\rho_2^2)\end{aligned} \Longrightarrow \text{Cor}(\psi_1^{(t+1)}, \psi_1^{(t)}) = \rho_1\rho_2.$$

1. *The two conditionals are incompatible when* $\rho_1 \neq \rho_2$*, and the joint chain does not converge.*
2. *But the marginal chain* $\{\psi_1^{(t)}, t=1,\ldots\}$ *is positive recurrent with* $N(0,1)$ *as its limiting distribution.*
3. *Furthermore, if we take* $\rho_1 = -\rho_2 = \rho$*, then* $\text{Cor}(\psi_1^{(t+1)}, \psi_1^{(t)}) = -\rho^2 < 0$*, not possible with a compatible Gibbs sampler.*

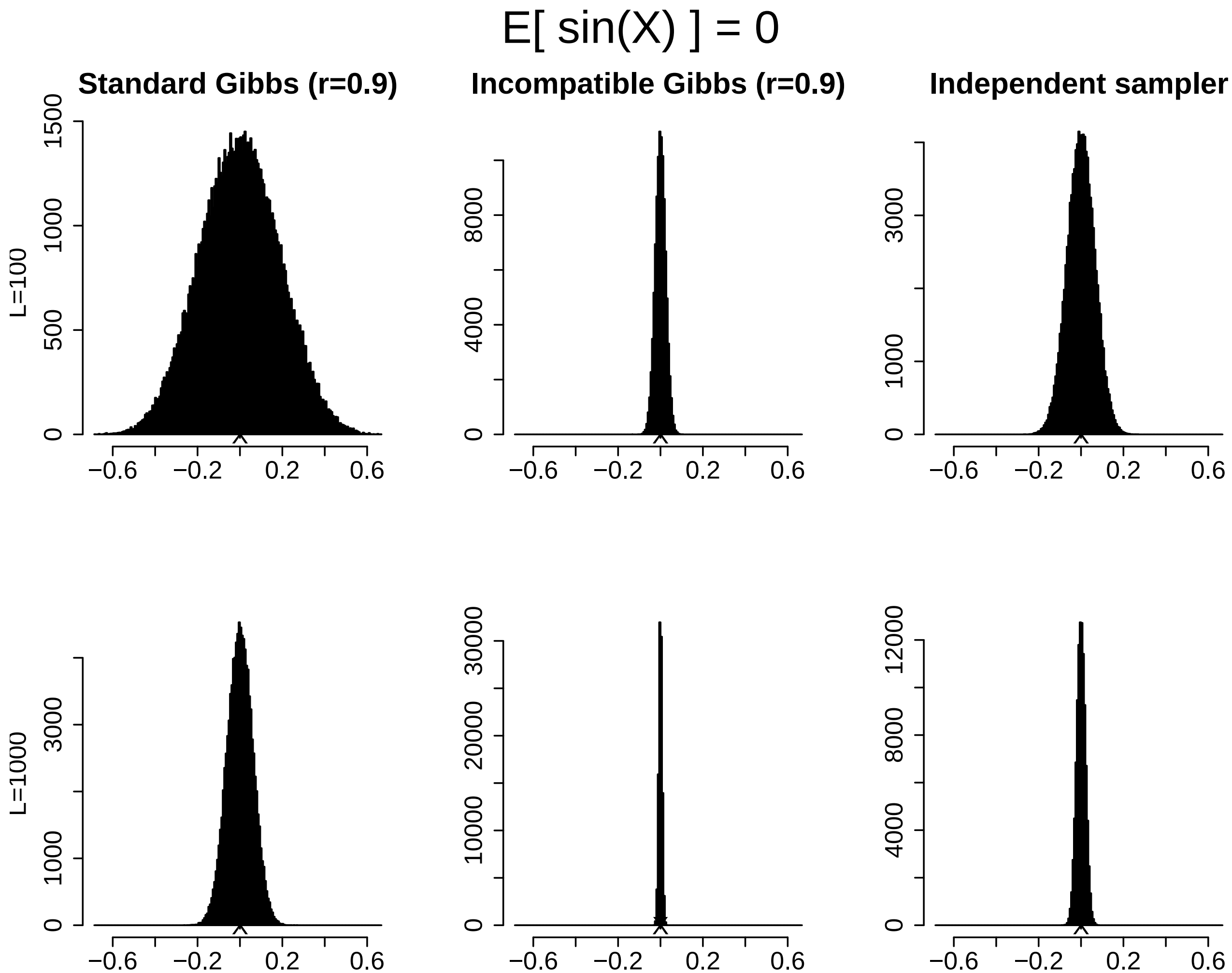
Can Gibbs Sampling be Better than Perfect Simulation??



It Depends on Your Objective!



It Depends on Your Objective!



Outline of Presentation

This talk has three components:

A. Working Parameters and Marginal Augmentation

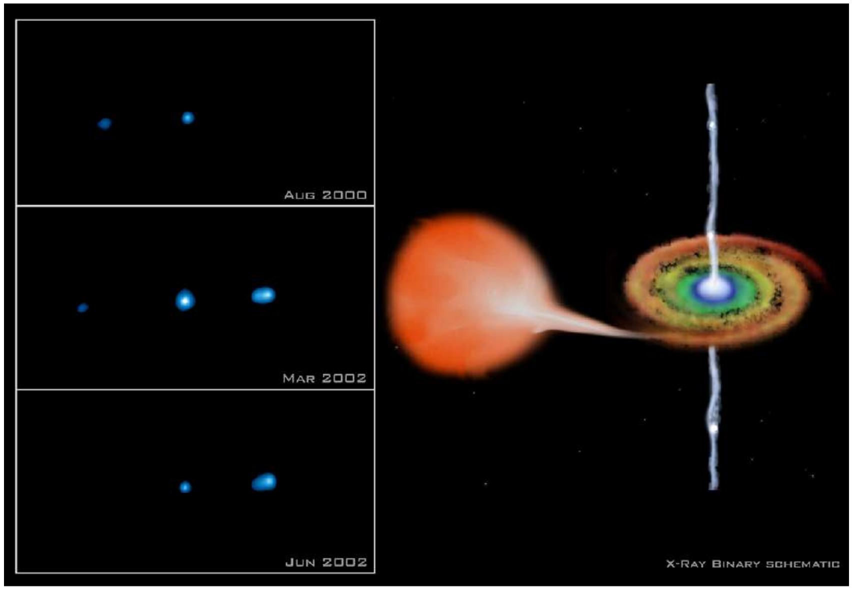
- Improves the convergence of Gibbs Samplers *(no details here)*.
- Show how this takes advantage of *incompatible conditional distributions.*
- Two joint distributions with the same margins, the target posterior distribution and a joint distribution with less correlation.

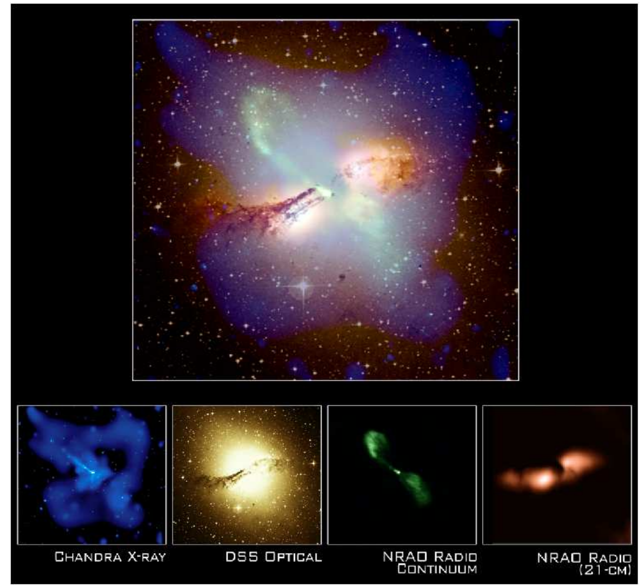
B. Can we use this idea to do *better than i.i.d. sampling?*

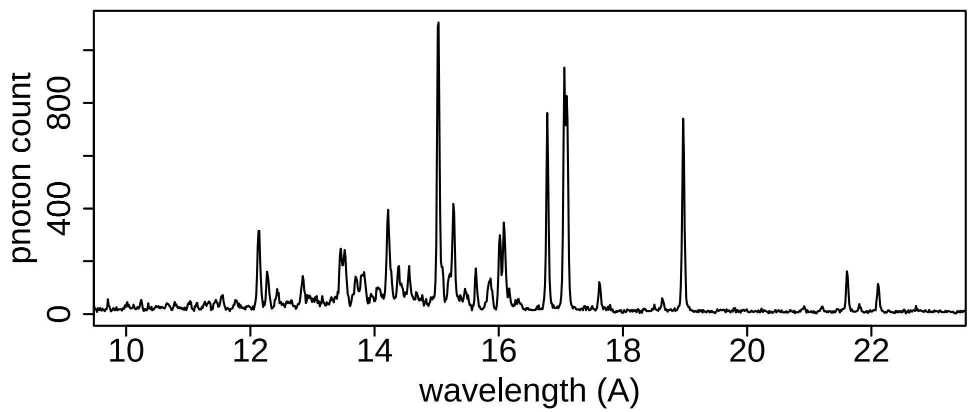
C. **Extension: Reduce conditioning is selected Gibbs steps.**

- This improves convergence but may introduce incompatibility.
- **Extended example from Astronomy.**

Complex Astronomical Sources

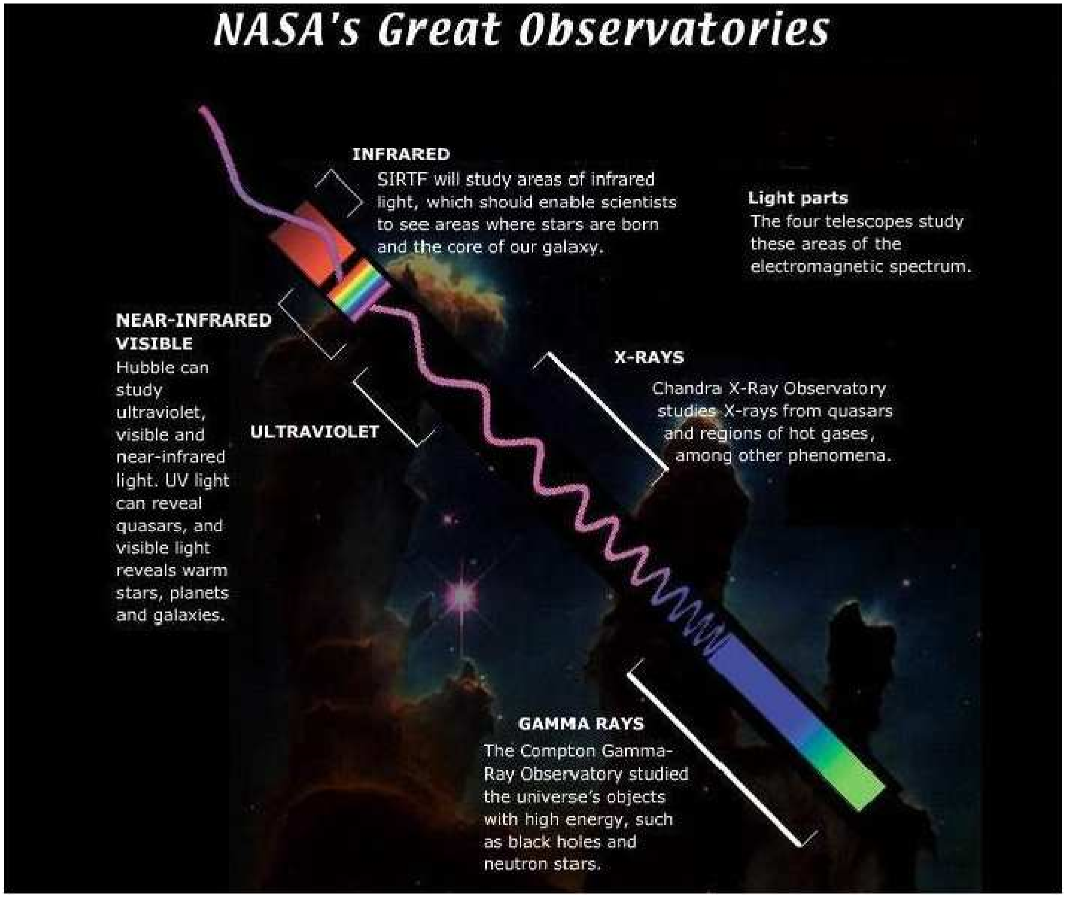






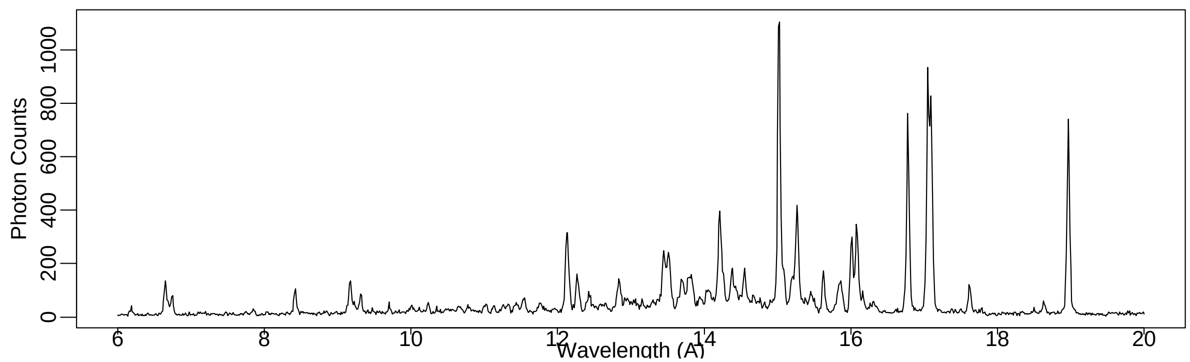
Images may exhibit Spectral, Temporal, and Spatial Characteristics.

Astrostatistics: Complex Data Collection



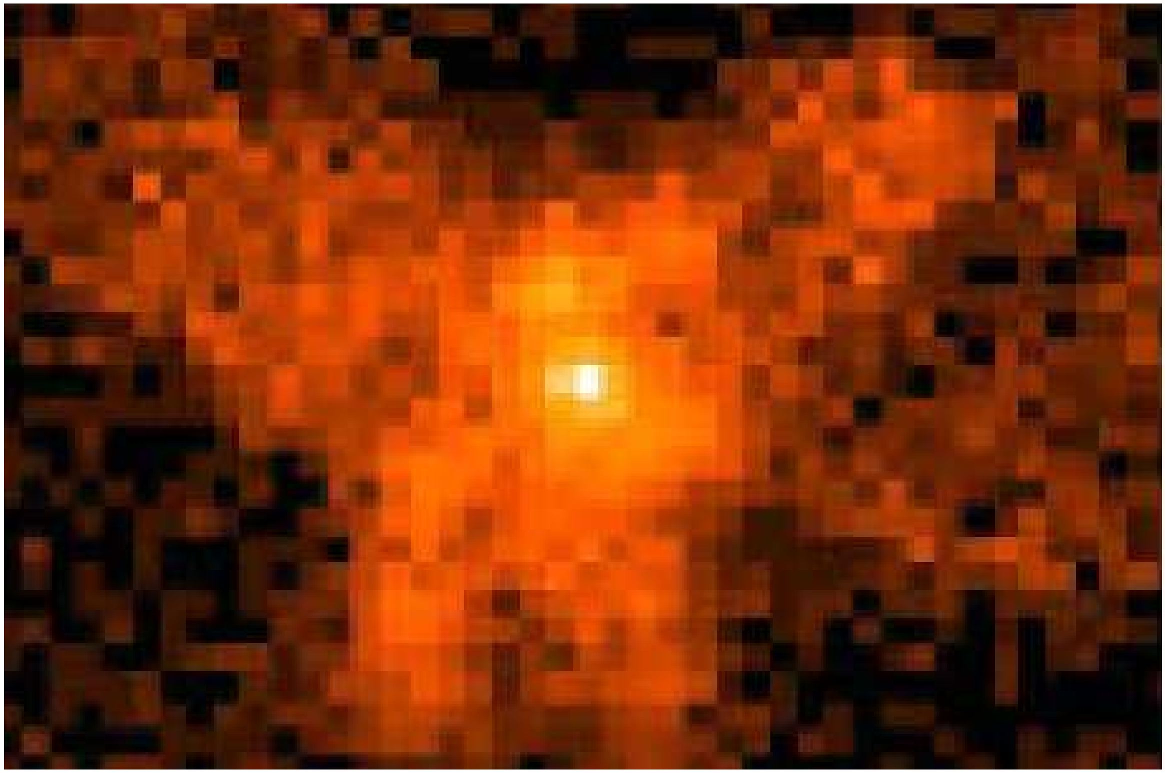
- A very small sample of instruments
- Earth-based, survey, interferometry, etc.
- X-ray alone: at least four planned missions
- Instruments have different data-collection mechanism

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

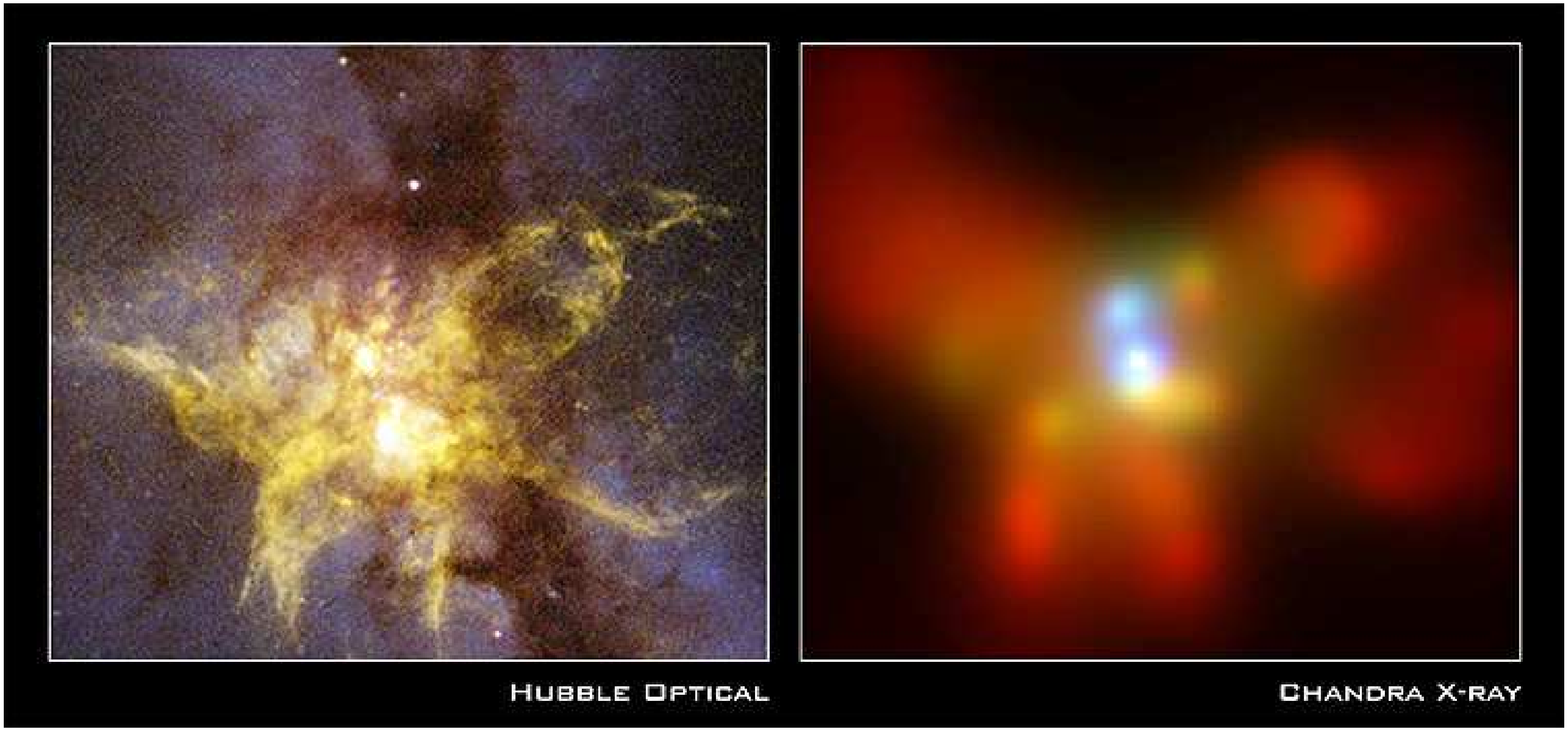
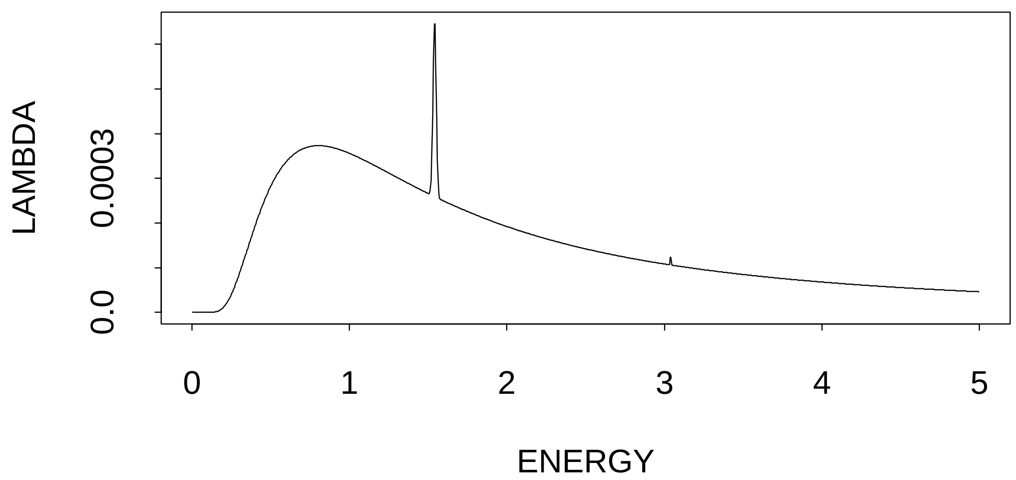


Image Credits.
X-ray: NASA/CXC/MPE/, Komossa et al. (2003, ApJL, 582, L15);
Optical: NASA/STScI/R.P.van der Marel & J.Gerssen.

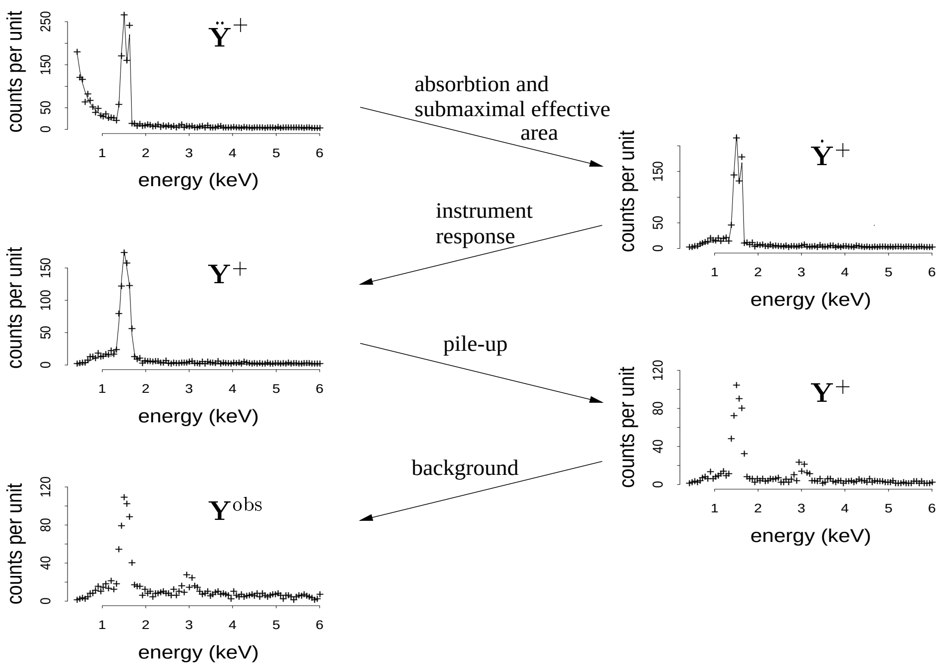
The Basic Spectral Models

- 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



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.

Searching for Narrow Lines

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

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

- We sometimes construct a *delta function* emission line model so that
  1. the emission line 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,

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

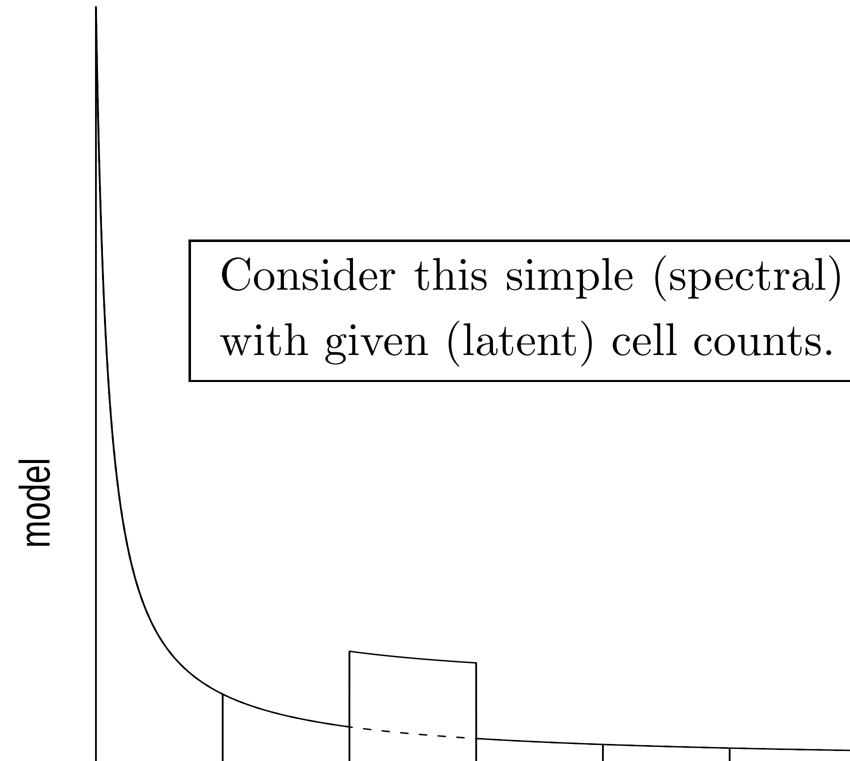
- Using *Data Augmentation* to fit this finite mixture model:

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

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

In This Case Data Augmentation Fails.

Why Data Augmentation Fails

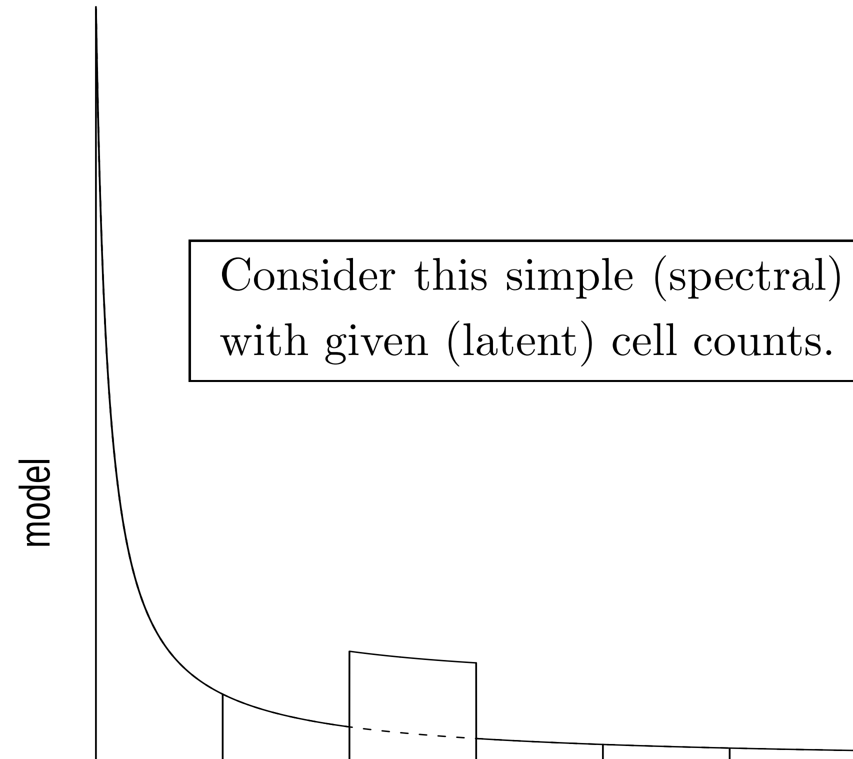


Consider this simple (spectral) model with given (latent) cell counts.

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

Given this Model, what is Z?

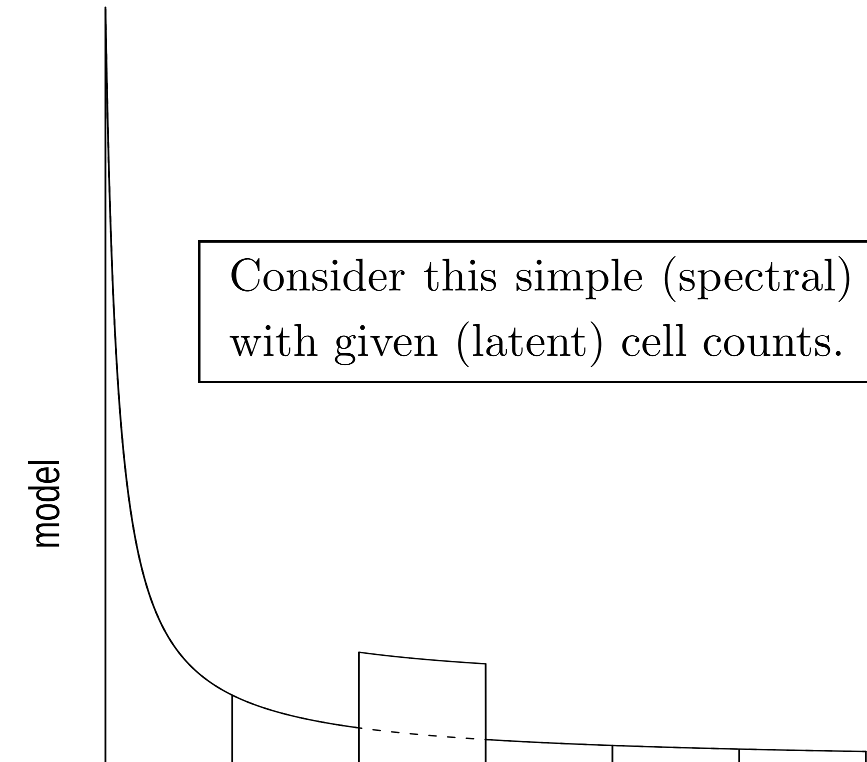
Why Data Augmentation Fails



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

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

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

The Standard Gibbs Sampler

Recall we do not observe the latent Poisson Process,

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

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

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

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.)

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

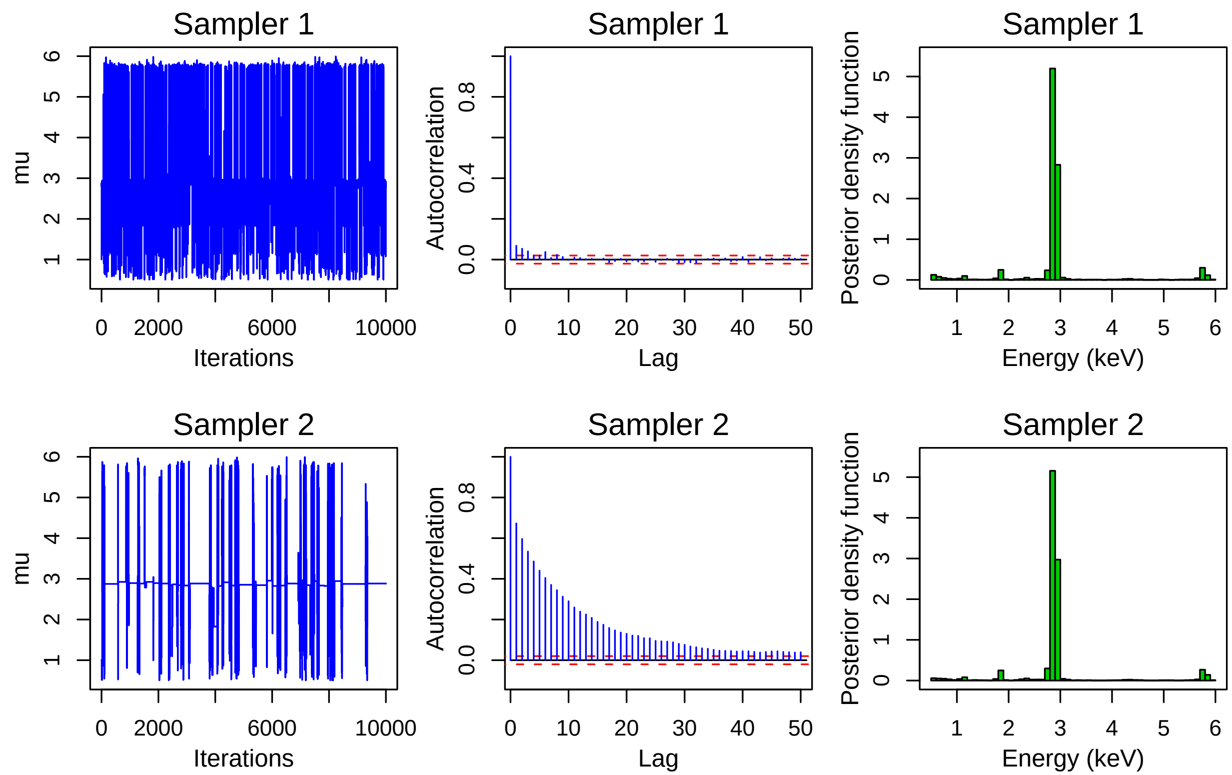
  Sampler 2: (Cannot be Blocked: An Incompatible Gibbs Sampler.)

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

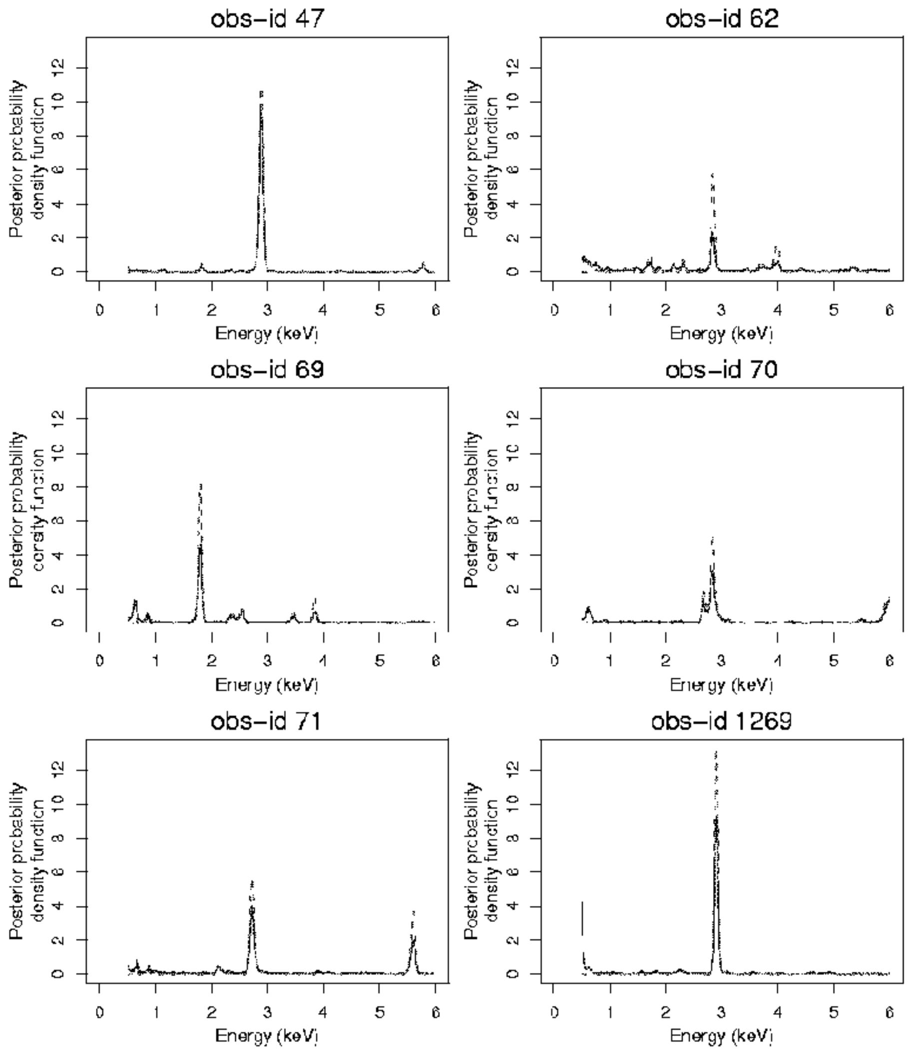
It can be shown that both samplers have the correct stationary distribution and are faster to converge than the standard sampler.

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.



Posterior Distribution of the Line Location



- Six observations were independently observed.
- Under a flat prior distribution on μ, we have

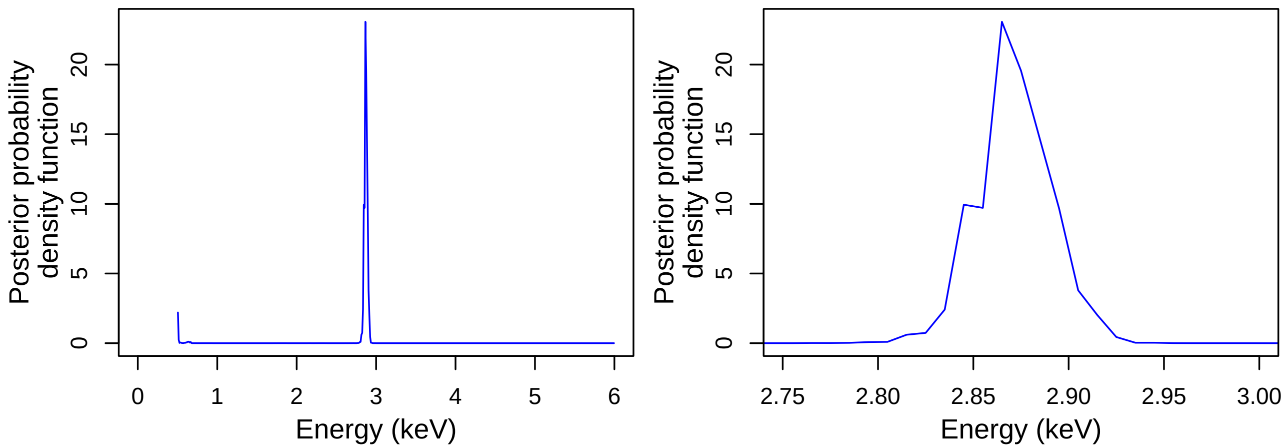
$$p(\lambda^{\mathrm{L}}|y) \propto \int \cdots \int \prod_{i=1}^{6} p(\theta^{\mathrm{L}}, \theta_i^{\mathrm{O}}|y_i) d\theta_1^{\mathrm{O}} \cdots d\theta_6^{\mathrm{O}}$$

$$= \prod_{i=1}^{6} \int p(\theta^{\mathrm{L}}, \theta_i^{\mathrm{O}}|y_i) d\theta_i^{\mathrm{O}}$$

$$= \prod_{i=1}^{6} p(\theta^{\mathrm{L}}|y_i).$$

Posterior Distribution of the Line Location (Cont'd)

- 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

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

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.

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

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

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

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.

The Advantage of Partial Marginalization

An Outline of the 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 how incompatible conditional distributions can be used in a Gibbs Sampler to improve convergence, and, how this technique can be used to fit Highly Structured Statistical Models and 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.