Accounting for Calibration Uncertainty in High Energy Astrophysics

David A. van Dyk[1]
Vinay Kashyap[2] Taeyoung Park[3] Jin Xu[4]
Imperial-California-Harvard AstroStatistics Collaboration

[1]Statistics Section, Imperial College London
[2]Smithsonian Astrophysical Observatory
[3]Yonsei University
[4]Department of Statistics, University of California, Irvine

Royal Society Meeting on Signal Processing & Physical Science
Buckinghamshire, UK, March 2012

Imperial College London

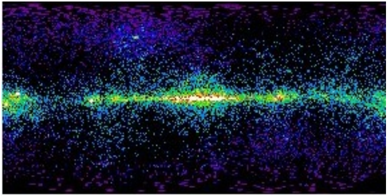
Outline

Outline

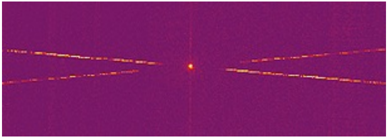
Imperial College London

High-Energy Astrophysics

- Produced by multi-million degree matter, e.g., magnetic fields, extreme gravity, or explosive forces.
- Provide understanding into the hot turbulent regions of the universe.
- X-ray and γ-ray detectors typically count a *small number of photons* in each of a *large number of pixels*.



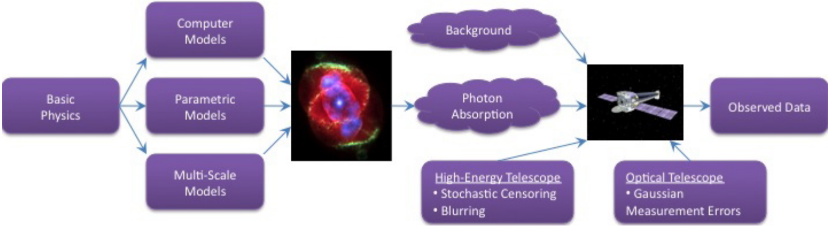
EGERT γ-ray counts >1GeV (entire sky and mission life).



Dispersion grating spectrum of an Active Galactic Nucleus; emission from matter accreting onto a massive Black Hole.

Imperial College London

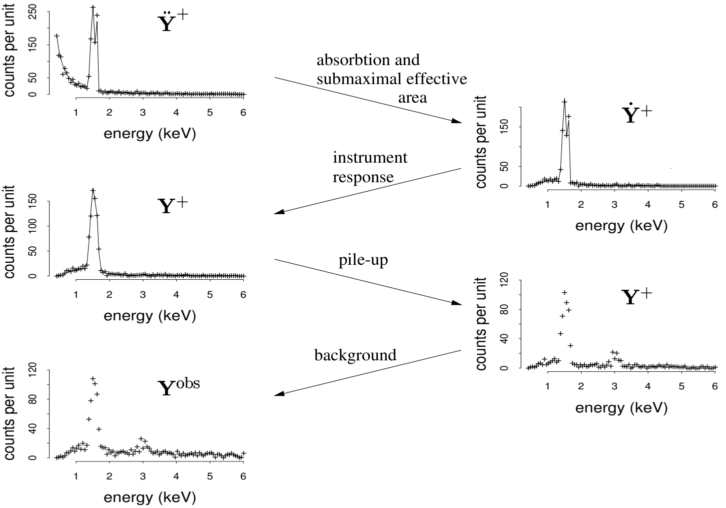
The Basic Statistical Model



- Aim to formulate models in terms of specific questions of scientific interest.
- Must account for complexities of data generation.
- Embed complex physics-based and/or instrumental models into multi-level statistical models.
- State of the art data and computational techniques enable us to fit the resulting complex model.

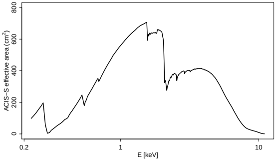
Imperial College London

Degradation of the Photon Counts

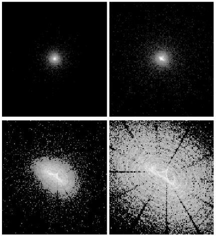


Calibration Products

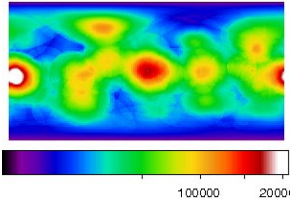
- Analysis is highly dependent on *Calibration Products:*
  - Effective area records sensitivity as a function of energy
  - Energy redistribution matrix can vary with energy/location
  - Point Spread Functions can vary with energy and location
  - Exposure Map shows how effective area varies in an image
- In this talk we focus on uncertainty in the effective area.



A *CHANDRA* effective area.



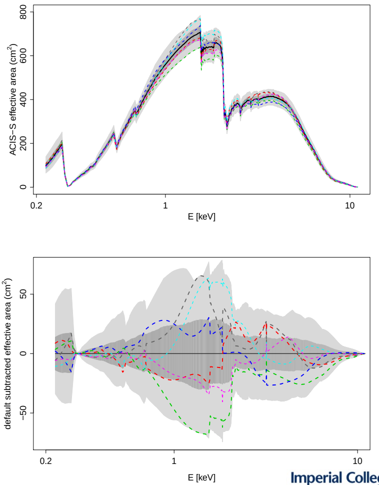
Sample Chandra psf's
(Karovska et al., ADASS X)



EGERT exposure map
(area $\times$ time)

Derivation of Calibration Products

- Prelaunch ground-based and post-launch space-based empirical assessments.
- Aim to capture deterioration of detectors over time.
- Complex computer models of subassembly components.
- Calibration scientists provide a sample representing uncertainty
- *Calibration Sample* is typically of size $\approx$ 1000.



Imperial College London

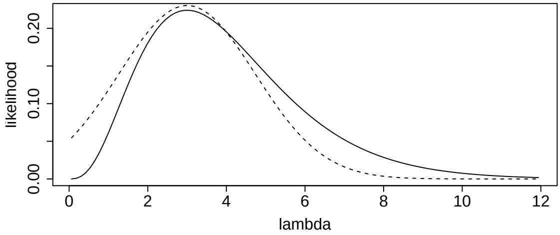
Outline

Bayesian Statistical Analyses: Likelihood

Likelihood Functions: The distribution of the data given the model parameters. E.g., $Y \sim \mathrm{Poisson}(\lambda_S)$:

$$\text{likelihood}(\lambda_S) = e^{-\lambda_S}\lambda_S^Y / Y!$$

Maximum Likelihood Estimation: Suppose $Y = 3$



The likelihood and its normal approximation.

Can estimate λ_S and its error bars.

Imperial College London

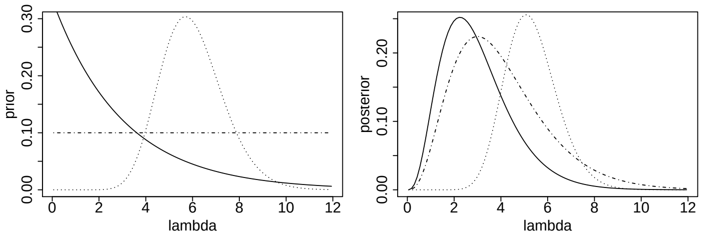
Bayesian Analyses: Prior and Posterior Dist'ns

Prior Distribution: Knowledge obtained *prior* to current data.

Bayes Theorem and Posteror Distribution:

$$\text{posterior}(\lambda) \propto \text{likelihood}(\lambda)\text{prior}(\lambda)$$

Combine past and current information:



Bayesian analyses allows us to incorporate external information via the prior distribuiton.

Imperial College London

The Standard Approach

In high energy astrophysics the effective area curve is invariably assumed known:

$$p(\theta|A, Y) \propto p(Y|\theta, A)p(\theta|A).$$

- θ: Model parameters, of primary scientific interest.
- A: Effective area curve, typically assumed known.
- Y: Observed data–bin counts.

Treating A as known is a ***very*** *strong prior!!*

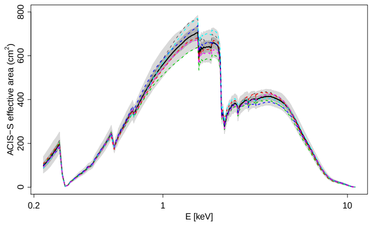
Our Approach

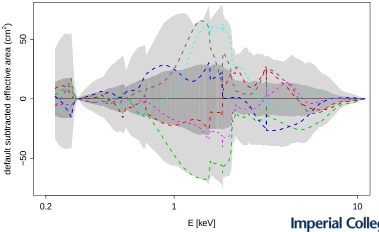
We

- introduce a Bayesian approach to **reduce** prior assumptions,
- propose to the use the calibration sample to represent the prior distribution for A, and
- base analysis on:

$$p(\theta, A|Y) \propto p(Y|\theta, A)p(\theta|A)p(A).$$

$$p(\theta|Y) = \int p(\theta, A|Y)dA.$$





Imperial College London

Two Possible Target Distributions

We consider inference under:

A PRAGMATIC BAYESIAN TARGET: $\pi_0(A, \theta) = p(A)p(\theta|A, Y)$.

THE FULLY BAYESIAN POSTERIOR: $\pi(A, \theta) = p(A|Y)p(\theta|A, Y)$.

Concerns:

Statistical Fully Bayesian target is "correct".

Cultural Astronomers have concerns about letting the current data influence calibration products.

Computational Both targets pose challenges, but pragmatic Bayesian is easier to fit.

Practical How different are $p(A)$ and $p(A|Y)$?

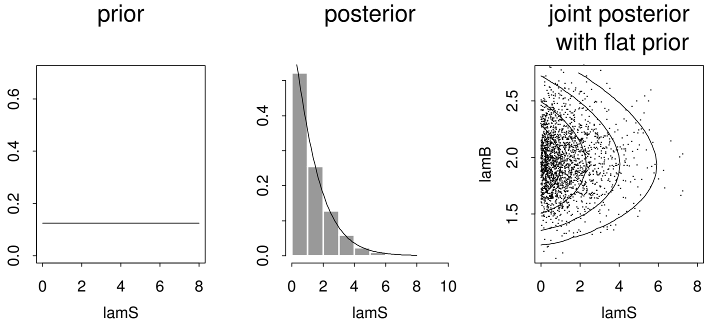
Model Fitting via Monte Carlo

Consider a contaminated Poisson count:

Source: $Y \sim \text{Poisson}(\lambda_s + \lambda_b)$

Background: $X \sim \text{Poisson}(c\lambda_b)$

Exploring the posterior distribution via Monte Carlo:



Imperial College London

A Gibbs Sampler for Calibration Uncertainty

A Markov chain Monte Carlo sampler iterates:

Step 1: Sample the effective area curve given the current value of model parameters.

Step 2: Sample the model parameters given the current value of the effective area.

- Step 2 samples under standard approach.
- Step 1 swaps in a different effective area at each iteration.
- Fully Bayes: Step 1 samples $\pi(A|\theta)$.
- Pragmatic Bayes: $\pi_0(A)$ is easier to sample than $\pi_0(A|\theta)$.
- Both effectively average over calibration uncertainty.

MH within Partially Collapsed Gibbs Samplers

MCMC for Pragmatic Bayes

Step 1: Sample the effective area curve from $\pi_0(A)$.

Step 2: Sample the model parameters from $\pi_0(\theta|A)$.
This requires an MH update.

A *naive* Sampler:

STEP 1: $\psi_1 \sim p(\psi_1)$

STEP 2: $\psi_2 \sim \mathcal{M}(\psi_2|\psi_1)$ via MH with limiting dist. $p(\psi_2|\psi_1)$

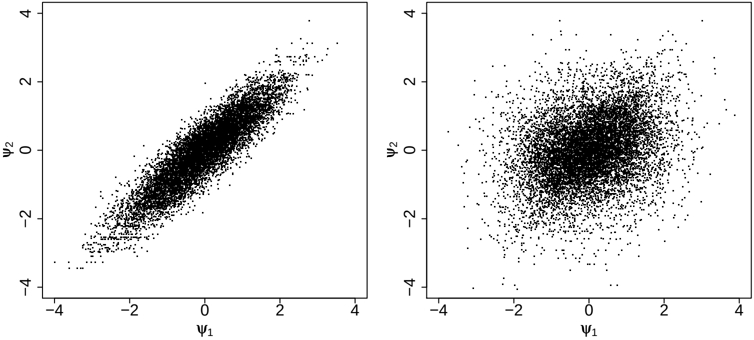
Simulation Study:

- Suppose $\binom{\psi_1}{\psi_2} \sim \mathrm{N}_2 \left[\binom{0}{0}, \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix} \right]$
- MH: a Gaussian jumping rule centered at previous draw.

Imperial College London

Be Careful When Combining MH and PCG Sampling

MH within Gibbs Sampler | The *naive* Sampler



What Goes Wrong

The *naive* Sampler:

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

STEP 2: $\psi_2^{(t)} \sim \mathcal{M}(\psi_2|\psi_1^{(t)}, \psi_2^{(t-1)})$ via Metropolis Hastings

The update of ψ_2 depends on both $\psi_1^{(t)}$ and $\psi_2^{(t-1)}$:

- The limiting distribution of the MH step is $p(\psi_2|\psi_1^{(t)})$.
- If the proposal is rejected, ψ_2 is set to $\psi_2^{(t-1)}$.

BUT: $\psi_1^{(t)} \sim p(\psi_1)$—independent of $\psi_2^{(t-1)}$ *at every iteration*.

STEP 2 *will never produce samples from* $p(\psi_2|\psi_1)$.

Imperial College London

Two Simple Solutions

Two possible samplers

1. A PCG (Simple Collapsed) Gibbs Sampler:

   STEP 1: $A^{(t)} \sim p(A)$

   STEP 2: Sample $\theta^{(t-1+\ell/L)} \sim \mathcal{M}(\theta|A^{(t)}, \theta^{(t-1)})$
   L times via MH to obtain $\theta^{(t)} \sim p(\theta|A^{(t)})$.

2. A pure MH Sampler:
   Jumping Rule: $(A^{\star}, \theta^{\star}) \sim p(A^{\star})\mathcal{M}(\theta^{\star}|A^{\star}, \theta^{(t-1)})$.

Tradeoff: MH is faster, PCG gives independent draws.

PCG has larger expected acceptance probability and lower empirical autocorrelation
(compared with L iterations of pure MH).

Imperial College London

Multiple Imputation

A simpler solution involves Multiple Imputation:

- Treat m effective areas from calibration sample as "imputations".
- Fit the model m times in standard way, once with each imputation.
- Compute estimates & errors with *Multiple Imputation Combining Rules*.

$$\hat{\theta} = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_m.$$

$$W = \frac{1}{M} \sum_{m=1}^{M} \mathrm{Var}(\hat{\theta}_m), \quad B = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\theta}_m - \hat{\theta})(\hat{\theta}_m - \hat{\theta})^\top.$$

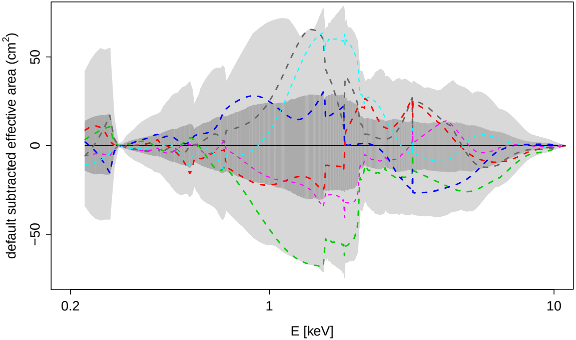
$$T = W + \left(1 + \frac{1}{M}\right) B,$$

Approximate Pragmatic Bayes: Replicates of $A \sim p(A)$.

Imperial College London

Sticking Point

- We only have a sample from $p(A)$.
- How do we incorporate this sample into our analysis?
- We do not want to store the entire calibration sample.



Imperial College London

Simple Emulation of Complex Variability

We use *Principal Component Analysis* to formulate a degenerate Gaussian approximation to the calibration sample:

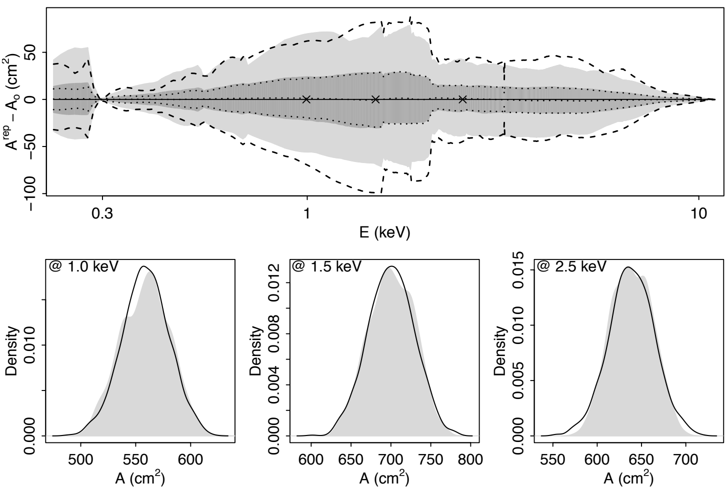
$$A \sim A_0 + \bar{\delta} + \sum_{j=1}^{m} e_j r_j \boldsymbol{v}_j,$$

A_0: default effective area,
$\bar{\delta}$: mean deviation from A_0,
r_j and $\boldsymbol{v}_j$: first m principle component eigenvalues & vectors,
e_j: independent standard normal deviations.

Capture 95% of variability with $m = 6 - 9$.

Imperial College London

Accounting for Uncertainty



The Two Possible Target Distributions

We consider inference under:

A PRAGMATIC BAYESIAN TARGET: $\pi_0(A, \theta) = p(A)p(\theta|A, Y)$.

THE FULLY BAYESIAN POSTERIOR: $\pi(A, \theta) = p(A|Y)p(\theta|A, Y)$.

- MCMC can be used with either target distribution.
- Fully Bayesian computation is more challenging.
- Multiple Imputation gives valid inference under the Pragmatic Bayesian distribution.
- Compare results using simulation studies & data analyses.

Outline

The Simulation Studies

Simulated Spectra

- Spectra were sampled using an absorbed power law,

$$f(E_j) = \alpha e^{-N_H x(E_j)} E_j^{-\Gamma},$$

  accounting for instrumental effects; E_j is the energy of bin j.
- Parameters (Γ and N_H) and sample size/exposure times:

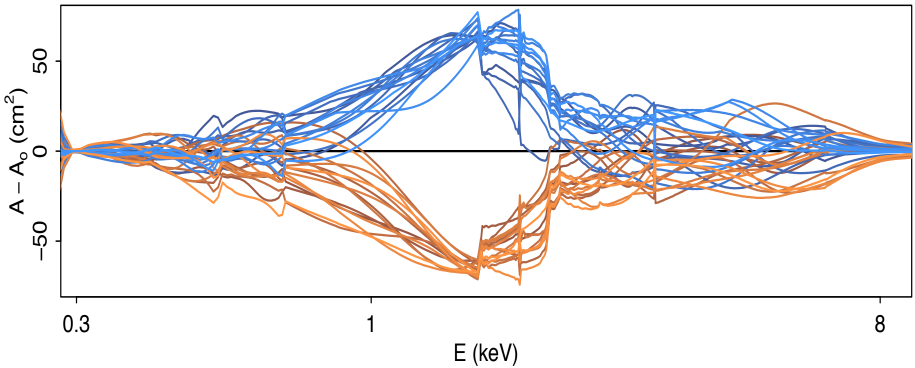
	Effective Area		Nominal Counts		Spectal Model	
	Default	Extreme	10^5	10^4	Hard†	Soft‡
SIM 1	X		X		X	
SIM 2	X		X			X
SIM 3	X			X	X	

†An absorbed powerlaw with $\Gamma = 2$, $N_{\rm H} = 10^{23}/\mathrm{cm}^2$

‡An absorbed powerlaw with $\Gamma = 1$, $N_{\rm H} = 10^{21}/\mathrm{cm}^2$

Imperial College London

30 Most Extreme Effective Areas in Calibration Sample

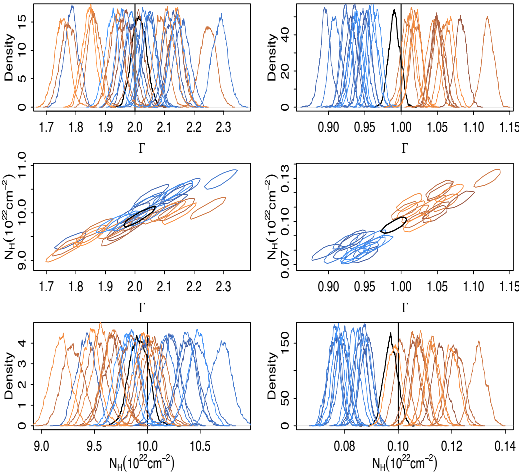


15 largest and 15 smallest determined by maximum value

Imperial College London

The Effect of Calibration Uncertainty

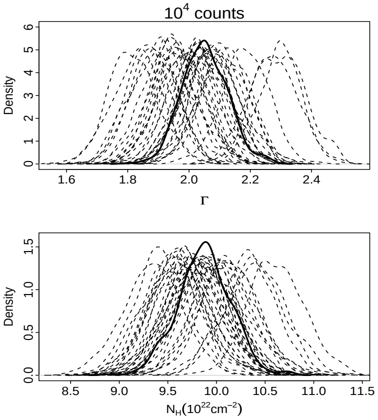
SIMULATION 1 SIMULATION 2



- Columns represent two simulated spectra.
- True parameters are horizontal lines.
- Posterior under default calibration is plotted in black.
- The posterior is highly sensitive to the choice of effective area!

Imperial College London

The Effect of Sample Size



The effect of Calibration Uncertainty is more pronounced with larger sample sizes.

Imperial College London

Expanded Simulation for Pragmatic Bayes

Simulated Spectra

- Spectra were sampled using an absorbed power law,

$$f(E_j) = \alpha e^{-N_H x(E_j)} E_j^{-\Gamma},$$

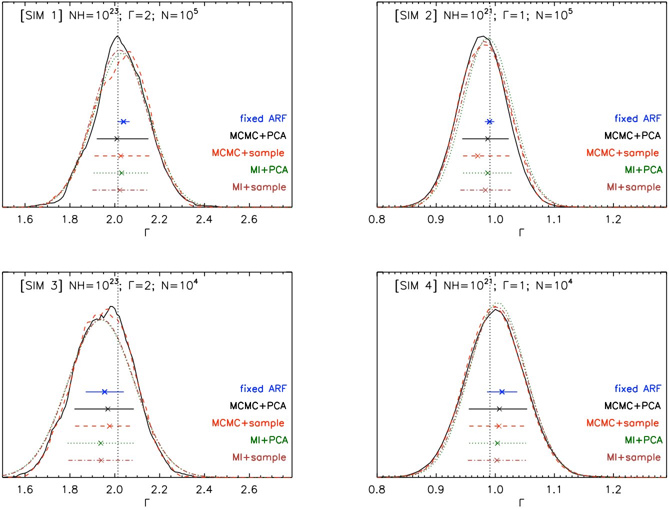
	Effective Area		Nominal Counts		Spectal Model	
	Default	Extreme	10^5	10^4	Hard†	Soft‡
SIMULATION 1	X		X		X	
SIMULATION 2	X		X			X
SIMULATION 3	X			X	X	
SIMULATION 4	X			X		X
SIMULATION 5		X	X		X	
SIMULATION 6		X	X			X
SIMULATION 7		X		X	X	
SIMULATION 8		X		X		X

† An absorbed powerlaw with $\Gamma = 2$, $N_{\rm H} = 10^{23}/{\rm cm}^2$

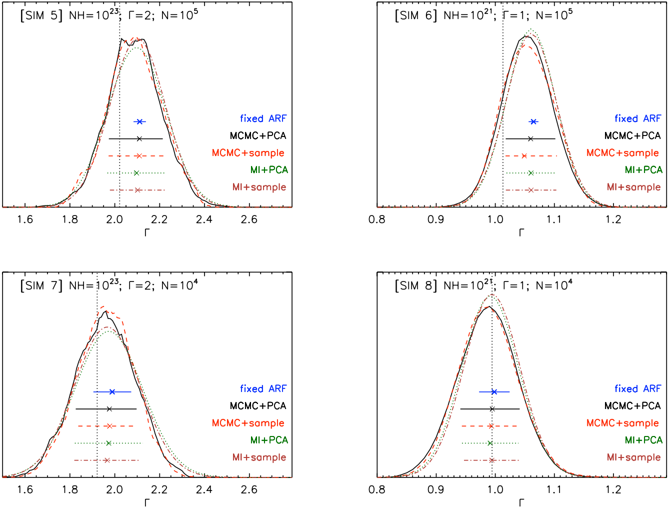
‡ An absorbed powerlaw with $\Gamma = 1$, $N_{\rm H} = 10^{21}/{\rm cm}^2$

Imperial College London

Pragmatic Bayes: Higher Variance Than Default



Pragmatic Bayes: Better Coverage Than Default



Imperial College London

A Simple Simulation for the Fully Bayesian Sampler

A Simple Simulation.

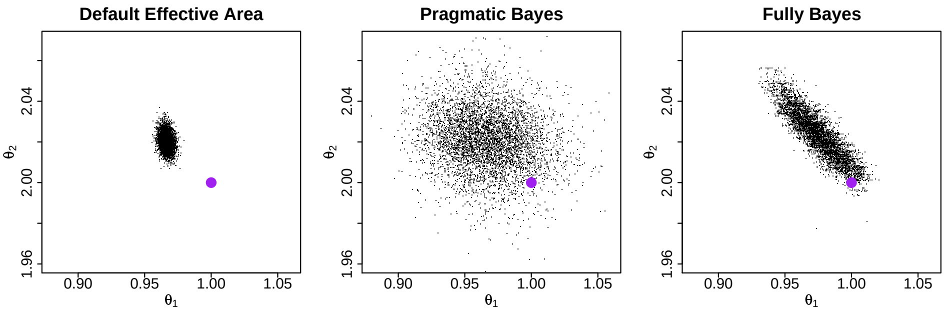
- Sampled 10^5 counts from a power law spectrum:

$$f(E_j) = \theta_1 e^{-\theta_3 x(E_j)} E_j^{-\theta_2}$$

- No energy blurring or backgraound contamination.
- Effective area used in the simulation differed from default:

  A_{true} is 1.5σ from the center of the calibration sample.

Sampling From the Full Posterior

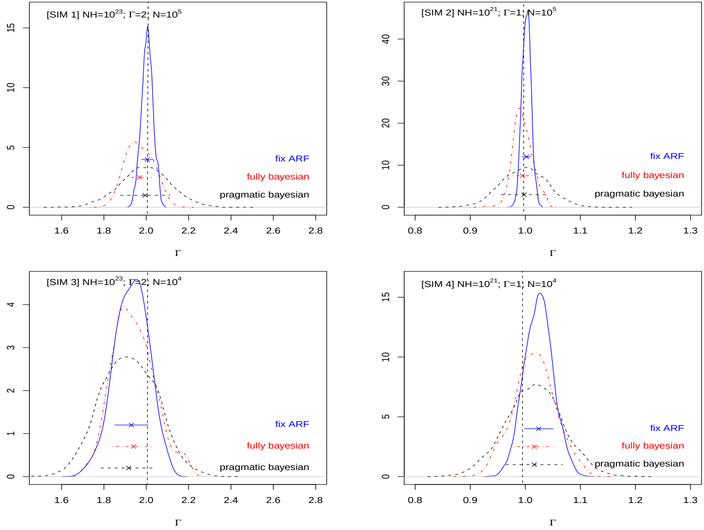


θ_1 = normalization, θ_2 = power law parameter
purple bullet = truth

Pragmatic Bayes is clearly better than current practice, but a Fully Bayesian Method is the ultimate goal.

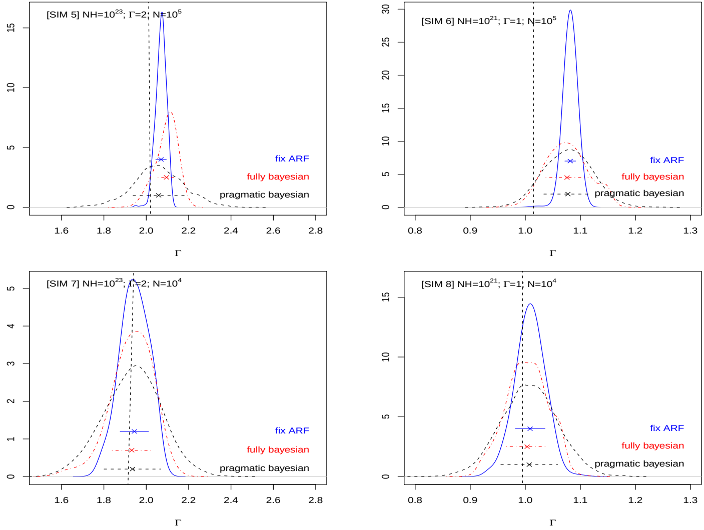
Imperial College London

Fully Bayesian: Less Variance that Pragmatic



Imperial College London

Fully Bayesian: Better Coverage than Default

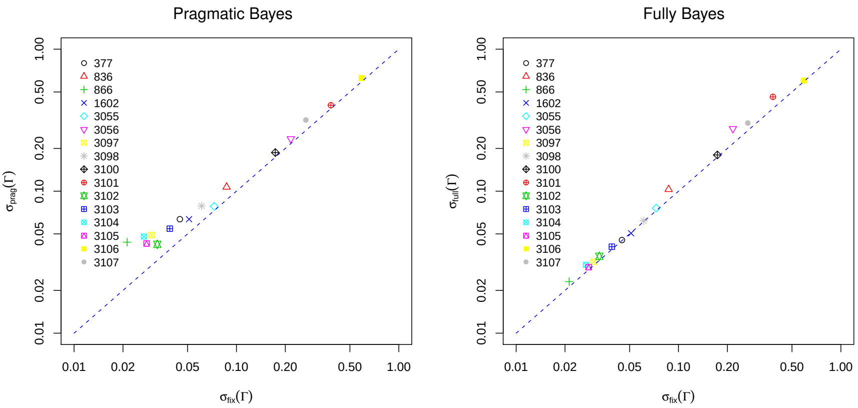


The Effect of Sample Size Redux

A Set of Radio Loud Quasar Spectra

- Pragmatic and Fully Bayesian Methods were applied to a set of Quasars.
- Quasars are among the most distant distinguishable astronomical objects.
- The sixteen Quasar observations varied is size from 20 to over 10,000 photon counts.

Results

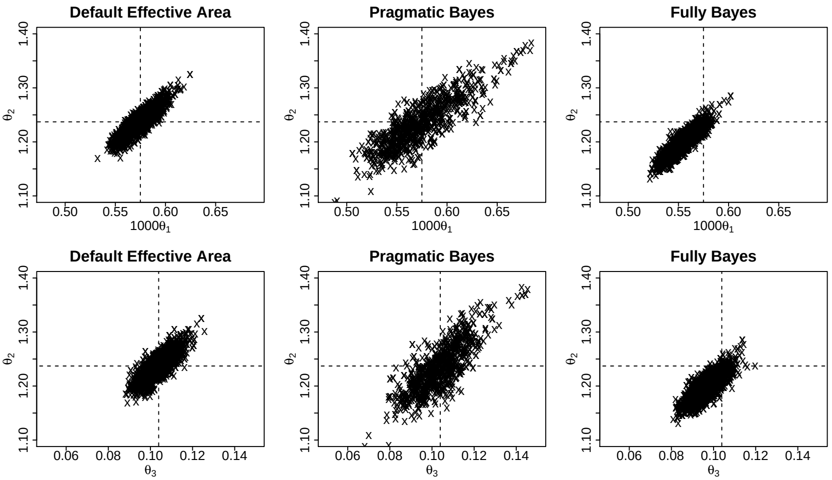


For large spectra calibration uncertainty swamps statistical error.

In large spectra fully Bayes identifies A and reduces uncertainty.

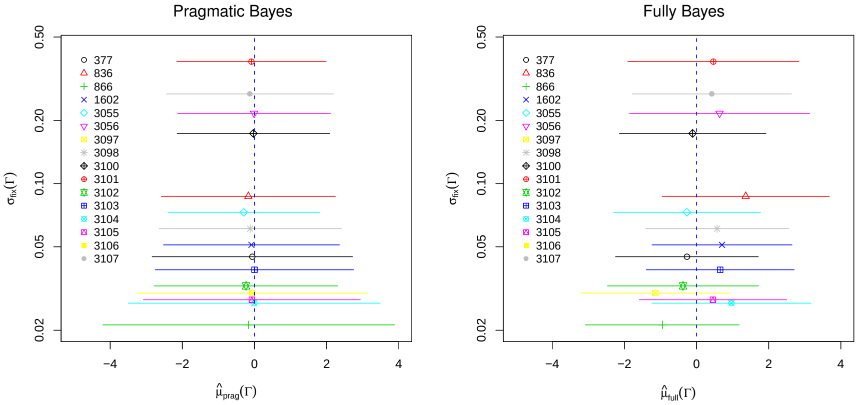
Imperial College London

Quasar 866



Fully Bayes Shifts Posterior Without Increasing SD.

Results: 95% Intervals Standardized by Standard Fit



For large spectra calibration uncertainty swamps statistical error.

In large spectra fully Bayes identifies A and shifts interval.

Imperial College London

For Further Reading I

Lee, H., Kashyap, V., van Dyk, D., Connors, A., Drake, J., Izem, R., Min, S., Park, T., Ratzlaff, P., Siemiginowska, A., and Zezas, A.
Accounting for Calibration Uncertainties in X-ray Analysis: Effective Area in Spectral Fitting.
The Astrophysical Journal, **731**, 126–144, 2011.

van Dyk, D. A. and Park, T.
Partially Collapsed Gibbs Sampling & Path-Adaptive Metropolis-Hastings in
High-Energy Astrophysics. In *Handbook of Markov Chain Monte Carlo* (Editors: S. Brooks, A. Gelman, G. Jones and X.-L. Meng), Chapman & Hall/CRC Press, 383-399, 2011.

Park, T. and van Dyk, D. A.
Partially Collapsed Gibbs Samplers: Illustrations and Applications.
Journal of Computational and Graphical Statistics, **18**, 283-305, 2009.

van Dyk, D. A. and Park, T.
Partially Collapsed Gibbs Samplers: Theory and Methods.
Journal of the American Statistical Association, **103**, 790–796, 2008.

Park, T., van Dyk, D. A., and Siemiginowska, A.
Searching for Narrow Emission Lines in X-ray Spectra: Computation and Methods.
The Astrophysical Journal, **688**, 807–825, 2008.

Imperial College London