Extending the $\log N - \log S$ by incorporating the uncertainty of γ

Vasileios Stampoulis
Supervisor : Prof. David van Dyk

Department of Mathematics
Imperial College London

Boston, July 2016

Outline

1. The Scientific Problem
2. Estimating $p(\gamma)$
   - Parameters of $p(\gamma | a_\gamma, b_\gamma)$
3. Incorporating in the $\log N - \log S$
4. CDFS dataset
   - Single Power Law Model
   - Broken Power Law Model

Outline

1. The Scientific Problem
2. Estimating $p(\gamma)$
   - Parameters of $p(\gamma|a_\gamma, b_\gamma)$
3. Incorporating in the $\log N - \log S$
4. CDFS dataset
   - Single Power Law Model
   - Broken Power Law Model

Introduction

- During our discussion about the flux to count rate conversion, we introduced a factor γ:

$$Y_i^{src} | S_i, E_i, \gamma_i \overset{\text{ind}}{\sim} Poisson\left(\lambda(S_i, E_i, \gamma_i) = \frac{S_i E_i}{\gamma_i} = \lambda_i \right) \quad (1)$$

- This parameter depends on the spectral model that is assumed and the energy band of the source.
- In the previous analysis this factor was considered to be constant for all the sources and set to the value of $\gamma = 2.679e - 9$.

Introduction

- Due to recent extended spectroscopic surveys, we have available the spectrum for most of the sources we observe.
- By choosing a spectral model and using relevant software (Sherpa), we can extract a distribution of the γ for each source.
- More specifically, we get MCMC draws from the joint distribution of the 3 parameters of the spectral model we assume, i.e. N_h, Γ, A_m.
- Using again Sherpa, we can get the posterior distribution of γ_{Sherpa} for the observed sources.

Introduction

- We will only have the distributions of γ_i's for the observed sources, $\gamma_{obs,i}$.
- Assume that the individual characteristics of the spectrum of each source that affect the distribution of γ are independent of the missing data mechanism (the MCAR notion).
- Then we can assume that the distributions of γ_{mis} for the missing sources would not differ from that of the observed sources.

Outline

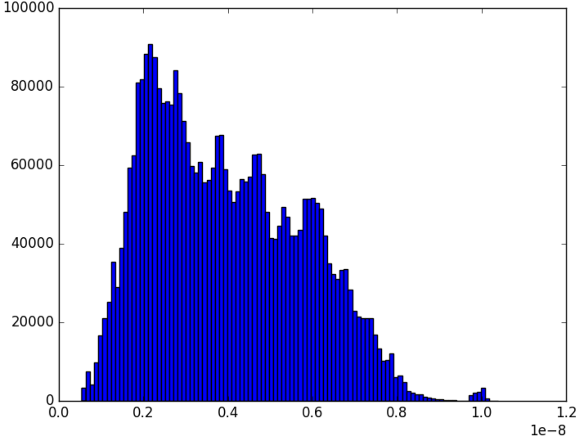
Modelling $p(\gamma)$

- For the observed source population we use Sherpa to draw samples from the posterior distribution $p(\gamma|SD)$, where SD describes the spectral data.
- We postulate that

$$\gamma \sim \text{Gamma}(a_\gamma, b_\gamma) \tag{2}$$

- The assumption for the Gamma distribution comes from the shape of the distribution of all the samples from γ obtained from Sherpa.

Modelling $p(\gamma)$



Modelling $p(\gamma)$

- The distribution in Equation (2) can be viewed as a hierarchical prior on $\gamma = (\gamma_1, ..., \gamma_N)$ for the complete source population.
- Hierarchical because it is specified in terms of parameters that are themselves fit to the data.
- Denote this hierarchical prior distribution by $p(\gamma|a_\gamma, b_\gamma)$.

Modelling $p(\gamma)$

- The prior distribution used in Sherpa however does not coincide with that one described in Equation (2).
- Sherpa assumes flat priors on the parameters of the assumed spectral model N_H, Γ, A_m.
- This translates to a non standard prior distribution $p_{\text{Sherpa}}(\gamma)$ that can be computed numerically using a non parametric density estimator.

Modelling $p(\gamma)$

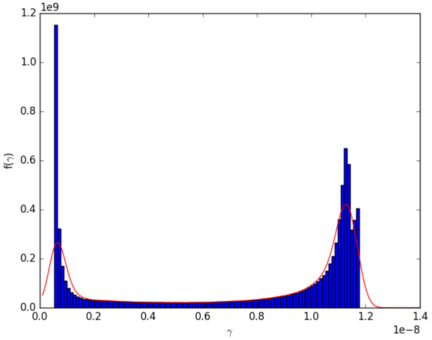


Figure: Histogram of the non standard prior distribution $p_{\text{Sherpa}}(\gamma)$. The red line is the probability density of the distribution computed using a non parametric kernel density estimation with a Gaussian kernel.

Modelling $p(\gamma)$

Sherpa produces a posterior sample from

$$p(\gamma|\mathrm{SD})_{\mathrm{Sherpa}} = \frac{p(\mathrm{SD}|\gamma) \cdot p_{\mathrm{Sherpa}}(\gamma)}{p_{\mathrm{Sherpa}}(\mathrm{SD})} \tag{3}$$

The target posterior distribution $p(\gamma|a_\gamma, b_\gamma, \mathrm{SD})$ is:

$$p(\gamma|a_\gamma, b_\gamma, \mathrm{SD}) = \frac{p(\mathrm{SD}|\gamma) \cdot p(\gamma|a_\gamma, b_\gamma)}{p(\mathrm{SD})} \tag{4}$$

Modelling $p(\gamma)$

The difference between $p(\gamma|a_\gamma, b_\gamma)$ and $p_{\text{Sherpa}}(\gamma)$ indicates that Sherpa produces a Monte Carlo sample from

$$p(\gamma|\text{SD})_{\text{Sherpa}} = \frac{p(\gamma|a_\gamma, b_\gamma, \text{SD}) \cdot p_{\text{Sherpa}}(\gamma)}{p(\gamma|a_\gamma, b_\gamma)} \quad (5)$$

Modelling $p(\gamma)$

The **proposed strategy** in order to draw samples from the target posterior distribution $p(\gamma|a_\gamma, b_\gamma, \mathrm{SD})$ is to use samples from the Sherpa posterior as a proposal rule in a Metropolis Hastings update.

Sampling Algorithm
Step 1: Run Sherpa for all the observed sources to obtain posterior samples.
Step 2: Set $\gamma^{(0)}$ to a randomly selected value from the Sherpa Monte Carlo sample of γ. Using standard Bayesian methods, fit the $\gamma^{(0)}$ to the model in Equation (2) to obtain $a_\gamma^{(0)}, b_\gamma^{(0)}$.

Modelling $p(\gamma)$

Step 3: For $t = 1, \ldots, T$
Step 3a: Select randomly a proposal γ^{prop} from the Sherpa Monte Carlo sample of γ.
Step 3b: Compute the n Metropolis Hastings acceptance probabilities,

$$r_i = \frac{p(\gamma_i^{\text{prop}} \mid a_\gamma^{(t-1)}, b_\gamma^{(t-1)}, \text{SD}) p_{\text{Sherpa}}(\gamma_i^{(t-1)} \mid \text{SD})}{p(\gamma_i^{(t-1)} \mid a_\gamma^{(t-1)}, b_\gamma^{(t-1)}, \text{SD}) p_{\text{Sherpa}}(\gamma_i^{\text{prop}} \mid \text{SD})} = \frac{p(\gamma_i^{\text{prop}} \mid a_\gamma^{(t-1)}, b_\gamma^{(t-1)}) \cdot p_{\text{Sherpa}}(\gamma_i^{(t-1)})}{p(\gamma_i^{(t-1)} \mid a_\gamma^{(t-1)}, b_\gamma^{(t-1)}) \cdot p_{\text{Sherpa}}(\gamma_i^{\text{prop}})} \tag{6}$$

for $i = 1, .., n$.

Modelling $p(\gamma)$

Step 3c: For $i = 1, ..., n$ set

$$\gamma_i^{(t)} = \begin{cases} \gamma_i^{prop} & \text{with probability } \min(1, r_i) \\ \gamma_i^{(t-1)} & \text{otherwise} \end{cases} \quad (7)$$

Step 3d: Sample $a_\gamma^{(t)}, b_\gamma^{(t)}$ using standard Bayesian methods.

Results of $p(\gamma)$

After 12,000 iterations of the sampling algorithm, the posterior estimates for the parameters of interest, a_γ and b_γ, are:

Table: Posterior estimates of a_γ and b_γ after 12,000 iterations of the sampling algorithm (we neglect the first 2,000 iterations as burn in).

	Mean	2.5%	97.5%
a_γ	5.58	4.72	6.48
b_γ	7.07e-10	6.04e-10	8.38e-10

For the observed source populations, we also get from the sampling algorithm the $p(\gamma_i|a_\gamma, b_\gamma, \text{SD})$ for $i = 1, ..., n$.

Results of $p(\gamma)$

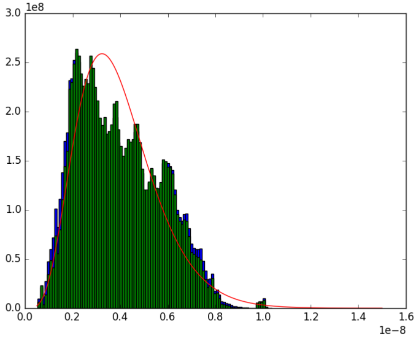


Figure: The blue histogram is the histogram of all the samples of γ obtained from Sherpa. The green histogram shows the distribution of all the samples of γ after correcting for the difference in the prior. The red line is the pdf of the prior Gamma distribution with parameters the posterior means of a_γ, b_γ

Outline

1. The Scientific Problem
2. Estimating $p(\gamma)$
   - Parameters of $p(\gamma|a_\gamma, b_\gamma)$
3. Incorporating in the $\log N - \log S$
4. CDFS dataset
   - Single Power Law Model
   - Broken Power Law Model

Posterior Distribution

The new marginalised posterior distribution is

$$
\begin{aligned}
&p(\gamma_{\text{obs}}, N, \theta, \tau, S_{\text{obs}}, Y^{\text{src}}_{\text{obs}} \mid n, Y^{\text{tot}}_{\text{obs}}, B_{\text{obs}}, L_{\text{obs}}, E_{\text{obs}}, A_{\text{obs}}, \text{SD}_{\text{obs}}, a_{\gamma}, b_{\gamma}) \\
&= \frac{1}{p(n, Y^{\text{tot}}_{\text{obs}}, B_{\text{obs}}, L_{\text{obs}}, E_{\text{obs}}, A_{\text{obs}}, \text{SD}_{\text{obs}}, a_{\gamma}, b_{\gamma})} \binom{N}{n} (1 - \pi(\theta, \tau))^{N-n} \\
&\quad \cdot p(N) \cdot p(\theta) \cdot p(\tau) \cdot p(B_{\text{obs}}, L_{\text{obs}}, E_{\text{obs}} | N, \theta, \tau) \\
&\quad \cdot p(\gamma_{\text{obs}} \mid \text{SD}_{\text{obs}}, a_{\gamma}, b_{\gamma}) \cdot p(S_{\text{obs}} | \theta, \tau) \\
&\quad \cdot p(I_{\text{obs}} \mid \gamma_{\text{obs}}, S_{\text{obs}}, B_{\text{obs}}, L_{\text{obs}}, E_{\text{obs}}) \\
&\quad \cdot p(Y^{\text{tot}}_{\text{obs}} \mid I_{\text{obs}}, S_{\text{obs}}, B_{\text{obs}}, L_{\text{obs}}, E_{\text{obs}}, \gamma_{\text{obs}}) \\
&\quad \cdot p(Y^{\text{src}}_{\text{obs}} \mid Y^{\text{tot}}_{\text{obs}}, I_{\text{obs}}, S_{\text{obs}}, B_{\text{obs}}, L_{\text{obs}}, E_{\text{obs}}, \gamma_{\text{obs}})
\end{aligned}
$$

Conditional Posterior Distribution of γ

By independence of the observed sources, we can sample component-wise for $i = 1, ..., n$ as:

$$p(\gamma_i|\cdot) \propto p(\gamma_i|SD_i, a_\gamma, b_\gamma) \cdot g(S_i, B_i, L_i, E_i, \gamma_i)$$
$$\cdot\, Poisson(Y_i^{tot}; \frac{S_i E_i}{\gamma_i} + B_i A_i)$$
$$\cdot\, Binomial\left(Y_i^{src}; Y_i^{tot}, \frac{\frac{S_i E_i}{\gamma_i}}{\frac{S_i E_i}{\gamma_i} + B_i A_i}\right).$$

Conditional Posterior Distribution of γ

We use the Metropolis- Hastings Algorithm in order to sample each γ_i, using $p(\gamma_i|SD_i, a_\gamma, b_\gamma)$ as proposal distribution. The algorithm would be as follows:

Step 1 Sample a proposal state randomly from $p(\gamma_i|SD_i, a_\gamma, b_\gamma)$, i.e. $\gamma_i^{prop} \sim p(\gamma_i|SD_i, a_\gamma, b_\gamma)$.

Step 2 Compute the Metropolis Hastings ratio :

$$\alpha = \frac{p(\gamma_i^{prop}|\cdot) \cdot p(\gamma_i^{curr}|SD_i, a_\gamma, b_\gamma)}{p(\gamma_i^{curr}|\cdot) \cdot p(\gamma_i^{prop}|SD_i, a_\gamma, b_\gamma)}$$

$$= \frac{g(\gamma_i^{prop}, \cdot) \cdot Pois(Y_i^{tot}; \frac{S_iE_i}{\gamma_i^{prop}} + B_iA_i) \cdot Bin\left(Y_i^{src}; Y_i^{tot}, \frac{\frac{S_iE_i}{\gamma_i^{prop}}}{\frac{S_iE_i}{\gamma_i^{prop}} + B_iA_i}\right)}{g(\gamma_i^{curr}, \cdot) \cdot Pois(Y_i^{tot}; \frac{S_iE_i}{\gamma_i^{curr}} + B_iA_i) \cdot Bin\left(Y_i^{src}; Y_i^{tot}, \frac{\frac{S_iE_i}{\gamma_i^{curr}}}{\frac{S_iE_i}{\gamma_i^{curr}} + B_iA_i}\right)}$$

Marginal posterior probability of detection

The marginal posterior probability of detection is

$$\pi(\theta,\tau) = \int P(I = 1 \mid S, B, L, E, \gamma) \cdot p(S, B, L, E, \gamma|\theta, \tau, a_\gamma, b_\gamma) \, \mathrm{d}S \, \mathrm{d}B \, \mathrm{d}E \, \mathrm{d}L \, \mathrm{d}\gamma$$

$$= \int g(S, B, L, E, \gamma) \cdot p(S|\theta, \tau) \cdot p(B, L, E) \cdot p(\gamma \mid a_\gamma, b_\gamma) \, \mathrm{d}S \, \mathrm{d}B \, \mathrm{d}E \, \mathrm{d}L \, \mathrm{d}\gamma \qquad (8)$$

Outline

1. The Scientific Problem
2. Estimating $p(\gamma)$
   - Parameters of $p(\gamma|a_\gamma, b_\gamma)$
3. Incorporating in the log N – log S
4. CDFS dataset
   - Single Power Law Model
   - Broken Power Law Model

Introduction

- I use 355 out of the 358 sources (no spectral data for 3 of them).
- I use the incompleteness function from Wright et al., 2015.

Single Power Law Model

Table: The posterior estimates for log N – log S assuming γ = constant.

	Mean	Median	SD	2.5%	97.5%
θ	0.825	0.836	0.03	0.750	0.858
N	1943.5	1940	158.5	1643	2333
τ	1.19e-17	1.21e-17	8.46e-19	8.56e-18	1.25e-17

Table: The posterior estimates for log N – log S with γ uncertainty.

	Mean	Median	SD	2.5%	97.5%
θ	0.711	0.709	0.05	0.620	0.806
N	2246	2197	330.2	1739	3060
τ	9.20e-18	9.58e-18	1.62e-18	5.60e-18	1.13e-17

Single Power Law Model

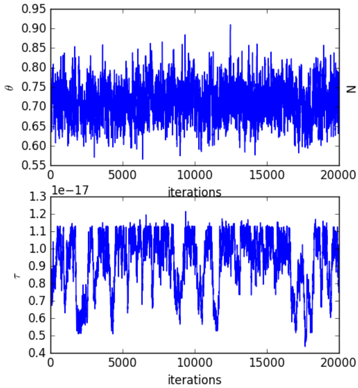


Figure: Trace plots of θ, N, τ.

Single Power Law Model

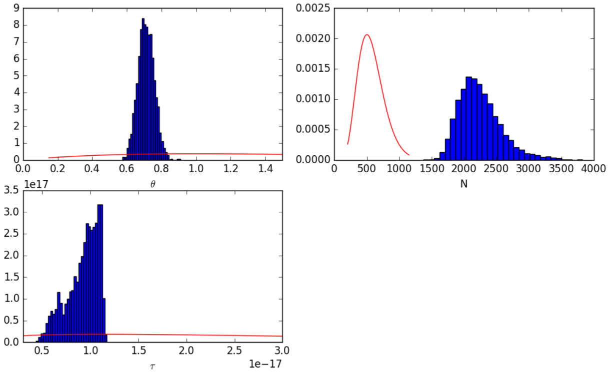


Figure: Histograms of the marginal posteriors of θ, N, τ with the priors over plotted.

Single Power Law Model

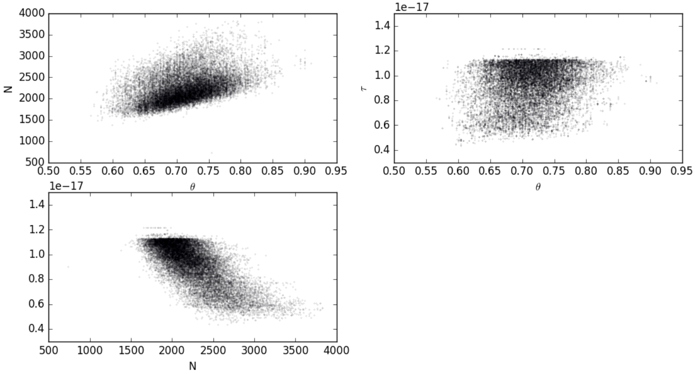


Figure: Scatter plots of θ, N, τ.

Single Power Law Model

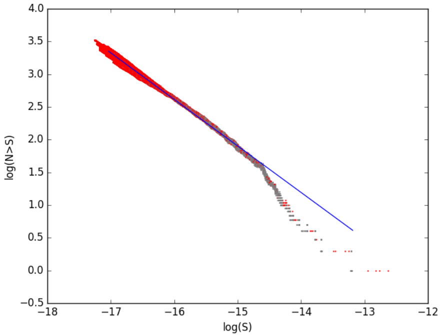


Figure: The posterior distribution of the $\log(N) - \log(S)$ plot for the CDFS data. Each line in the plot corresponds to a set of fluxes for the complete source population sampled from a single iteration of MCMC scheme with observed sources shown in grey and missing sources in red. Current plot exhibits sample of 50 flux sets. The blue line is the estimated $\log(N) - \log(S)$ using the posterior means of θ, N, τ.

Broken Power Law Model

Table: The posterior estimates for log N – log S assuming γ = constant.

	Mean	Mode	SD	2.5%	97.5%
θ_1	0.764	0.763	0.05	0.660	0.864
θ_2	1.21	1.18	0.36	0.651	2.07
N	2021	1977	295	1575	2722
τ_1	9.96e-18	1.01e-17	1.68e-18	5.74e-18	1.24e-17
τ_2	2.19e-15	1.45e-15	1.98e-15	2.43e-16	7.62e-15

Table: The posterior estimates for log N – log S with γ uncertainty.

	Mean	Mode	SD	2.5%	97.5%
θ_1	0.679	0.679	0.05	0.578	0.779
θ_2	1.35	1.29	0.41	0.773	2.18
N	2086	2069	264	1630	2655
τ_1	9.37e-18	9.39e-17	1.24e-18	7.04e-18	1.12e-17
τ_2	2.89e-15	1.95e-15	4.07e-15	2.36e-16	1.74e-14

Broken Power Law Model

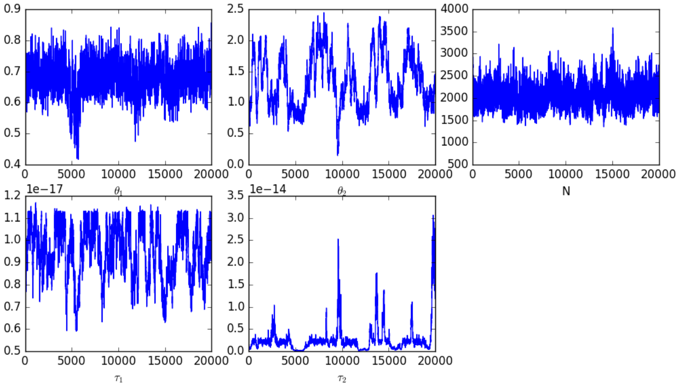


Figure: Trace plots of $\theta_1, \theta_2, N, \tau_1, \tau_2$.

Broken Power Law Model

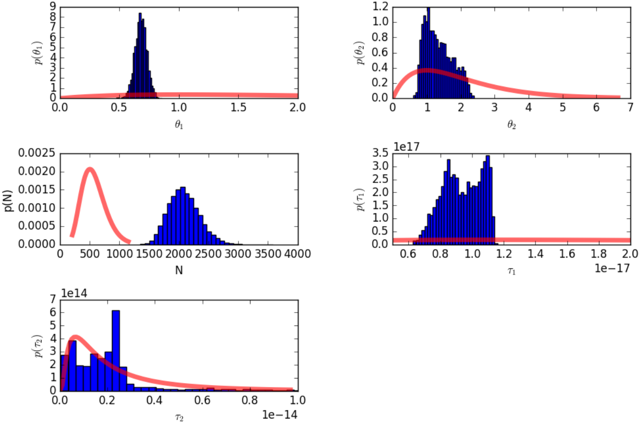


Figure: Histograms of the marginal posteriors of θ_1, θ_2, N, ... with the

Broken Power Law Model

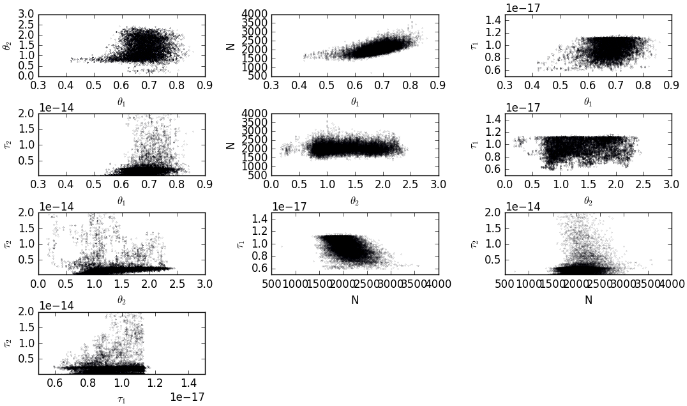


Figure: Scatter plots of $\theta_1, \theta_2, N, \tau_1, \tau_2$.

Broken Power Law Model

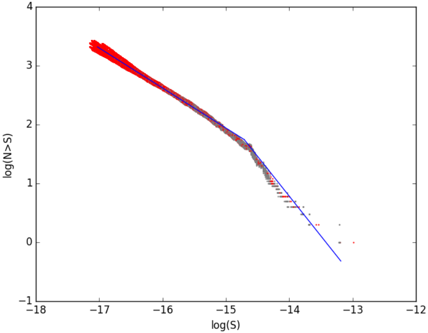


Figure: The posterior distribution of the $\log(N) - \log(S)$ plot for the CDFS data. Each line in the plot corresponds to a set of fluxes for the complete source population sampled from a single iteration of MCMC scheme with observed sources shown in grey and missing sources in red. Current plot exhibits sample of 50 flux sets. The blue line is the estimated $\log(N) - \log(S)$ using the posterior means of θ, N, τ.

Bibliography

Wright, N. J., Drake, J. J., Guarcello, M. G., Kashyap, V. L. & Zezas, A. (2015), Simulating the sensitivity to stellar point sources of chandra x-ray observations, arXiv preprint arXiv:1511.03943 .