Bayesian Modeling of $\log N - \log S$

Irina S Udaltsova
Vasileios Stampoulis

Department of Statistics
University of California, Davis

July 26th, 2016

Introduction

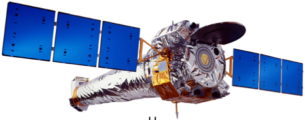
Scientific Objectives:
Develop a comprehensive method to infer (properties of) the distribution of source fluxes for a wide variety source populations.

Statistical Objectives:

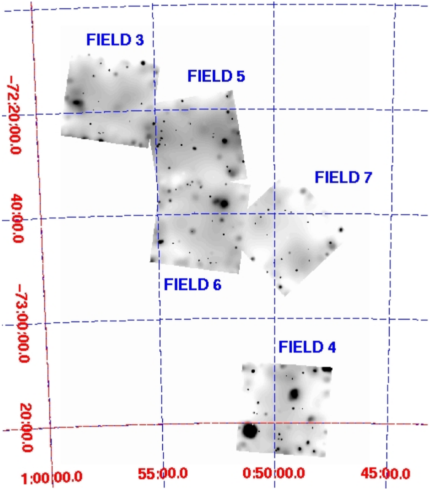
- Inference: Account for non-ignorable missing data
- Model Checking: Evaluate the adequacy of a given model
- Model Selection: Select the best model for a given dataset

Collaborators: Paul D. Baines (UCD), Andreas Zezas (University of Crete & CfA), Vinay Kashyap (CfA).

CHANDRA:



⇓


FIELD 3
FIELD 5
FIELD 7
FIELD 6
FIELD 4
-72:20:00.0
40:00.0
-73:00:00.0
20:00.0
1:00:00.0
55:00.0
0:50:00.0
45:00.0

INTRODUCTION: 'log $N - \log S$'

- Cumulative number of sources detectable at a given sensitivity:

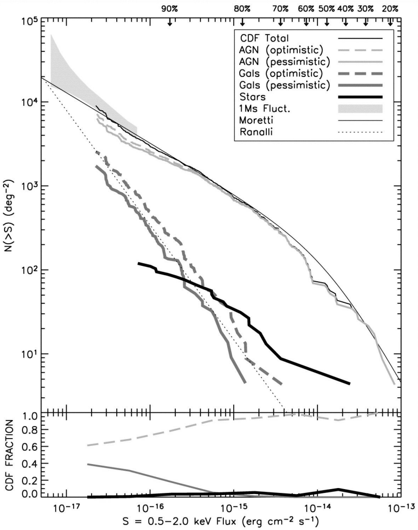
$$N(>S) = \sum_i I_{\{S_i > S\}}$$

  or the number of source fluxes brighter than a threshold, S.
- The name 'log $N - \log S$' refers to the relationship between (or plot of) $\log_{10} N(>S)$ and $\log_{10} S$.
- In many cosmological applications there is strong theory that expects the $\log N - \log S$ to obey a *Power law*:

$$N(>S) \;=\; \sum_{i=1}^{N} I_{\{S_i > S\}} \;\approx\; \alpha S^{-\theta}, \quad S > S_{min}$$

  Taking the logarithm gives the linear $\log(N) - \log(S)$ relationship.

The Rationale for $\log N - \log S$ Fitting



Knowing the specific relationship for different objects (e.g., stars, galaxies, pulsars) gives a lot of information about the underlying physics (e.g., the mass of galaxies). Helps in tests for cosmological parameters, constrains evolutionary models, etc.

Primary Goal: Estimate θ, the power law slope, while properly accounting for detector uncertainties and biases.

Note: There is uncertainty on both x- and y-axes (i.e., N and S).

Flux Distribution

Probabilistic Connection: Under independent sampling, linearity on the $\log N - \log S$ scale is *equivalent* to the flux distribution being a Pareto distribution.

$$S_i|\tau_1, \theta \overset{iid}{\sim} Pareto\,(\theta, \tau_1)\,, \qquad i = 1, \ldots, N.$$

Piecewise-linear power-law extends to broken power-law model for flux distribution, subject to continuity constraint at breakpoint τ_2:

$$\log_{10}(1 - F_G(s)) = \begin{cases} \alpha_0 - \theta_0 \log_{10}(s) & \tau_1 < s \leq \tau_2 \\ \alpha_1 - \theta_1 \log_{10}(s) & s > \tau_2 \end{cases}$$

$$Y \sim I \cdot X_1 + (1 - I) \cdot X_2, \quad \text{where:} \quad I \sim \mathrm{Bin}\left(1, \left[1 - \left(\frac{\tau_1}{\tau_2}\right)^{-\theta_1}\right]\right)$$

$$X_1 \sim \text{Truncated-Pareto}\,(\tau_1, \theta_1, \tau_2)\,, \qquad X_2 \sim \text{Pareto}\,(\tau_2, \theta_2)\,.$$

Can be extended to a multiple broken power-law with arbitrary number, $m - 1$, of break-points.

Missing Data

There are many potential causes of uncertainty and missing data in astronomical data:

- Low-count sources (below detection threshold)
- Detector schedules (source not within detector range)
- Background contamination (e.g., total=source+background)
- Foreground contamination (other objects between the source and detector)
- etc.

Important: Whether a source is observed is a function of its source count (intensity), which is unobserved for unobserved sources – missing data mechanism is non-ignorable, so it needs to be carefully accounted for in the analysis.

The Data

- Flux is not measured directly; instead, the data are a list of photon counts, with some extra information about the background and detector properties.

Src_ID	Counts	Bgr_intensity	Src_area	Off_axis_L	Effective_area
1	1093	47.38195	466	6.18	383.609
2	927	16.40961	180	5.75	392.709
3	31	12.66816	126	4.43	396.570
4	5	1.155294	12	0.48	278.892
5	286	17.50082	190	5.82	345.492
6	469	44.74188	436	5.36	358.845

- ...and an incompleteness function (selection function), specifying the probability of source detection under a range of conditions:

$$\mathbb{P}(\text{ Detecting a source with flux S, background intensity B, location L and effective area E }) \equiv g(S, B, L, E)$$

Single Power-Law Model

Standard power-law flux distribution:

$$S_i|\tau, \theta \overset{iid}{\sim} \mathrm{Pareto}\left(\theta, \tau\right), i = 1, \ldots, N.$$

Source and background photon counts:

$$Y_i^{tot}|S_i, B_i, L_i, E_i \overset{indep}{\sim} \mathrm{Pois}\left(\lambda(S_i, B_i, L_i, E_i) + k(B_i, L_i, E_i)\right), i = 1, \ldots, N.$$

Incompleteness, missing data indicators:

$$I_i \sim \mathrm{Bernoulli}\left(g\left(S_i, B_i, L_i, E_i\right)\right).$$

Prior distributions:

$$\begin{aligned} N &\sim \mathrm{Neg\text{-}Bin}\left(a_N, b_N\right), \\ \theta &\sim \mathrm{Gamma}(a, b), \\ \tau &\sim \mathrm{Gamma}(a_m, b_m). \end{aligned}$$

Broken Power-Law Model

Broken power-law flux distribution:

$$S_i | \vec{\tau}, \vec{\theta} \overset{iid}{\sim} \text{Broken-Pareto}\left(\vec{\theta}, \vec{\tau}\right), i = 1, \ldots, N.$$

Source and background photon counts:

$$Y_i^{tot} | S_i, B_i, L_i, E_i \overset{indep}{\sim} \text{Pois}\left(\lambda(S_i, B_i, L_i, E_i) + k(B_i, L_i, E_i)\right), i = 1, \ldots, N.$$

Incompleteness, missing data indicators:

$$I_i \sim \text{Bernoulli}\left(g\left(S_i, B_i, L_i, E_i\right)\right).$$

Prior distributions:

$$N \sim \text{Neg-Bin}\left(a_N, b_N\right),$$

$$\theta_j \overset{indep}{\sim} \text{Gamma}(a_j, b_j), j = 1, \ldots, m,$$

$$\tau_1 \sim \text{Gamma}(a_m, b_m)$$

$$\tau_j = \tau_1 + \sum_{k=2}^{j} e^{\eta_k}, \ \eta_j \overset{indep}{\sim} \text{Normal}(\mu_j, c_j), j = 2, \ldots, m.$$

Model Overview

Unusual points and important notes:

- The dimension of the missing data is unknown (care must be taken with conditioning)
- The incompleteness function g must be known well; can take any form; is problem-specific
- The number of flux populations in sample, m, must be specified
- The flux lower limit and break-points, $\tau_j, j = 1, \ldots, m$, can be estimated
- Prior parameters can be science-based, i.e., 'weakly informative'

Posterior Inference (Single Pareto)

Inference about θ, N, S, τ is based on the *observed data* posterior distribution. Care must be taken with the variable dimension marginalization over the unobserved fluxes.

The single power-law posterior can be shown to be:

$$p\left(N, \theta, \tau, S_{obs}, Y_{obs}^{src} | n, Y_{obs}^{tot}, B_{obs}, L_{obs}, E_{obs}\right)$$

$$\propto \binom{N}{n} \mathbb{I}_{\{n \leq N\}} \cdot (1 - \pi(\theta, \tau))^{(N-n)} \cdot \binom{N + a_N - 1}{a_N - 1} \left(\frac{1}{1 + b_N}\right)^N \left(\frac{b_N}{1 + b_N}\right)^{a_N} \mathbb{I}_{\{N \in \mathbb{Z}^+\}} \cdot$$
$$\cdot \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \mathbb{I}_{\{\theta > 0\}} \cdot \frac{b_m^{a_m}}{\Gamma(a_m)} \tau^{a_m - 1} e^{-b_m \tau} \mathbb{I}_{\{\tau > 0\}} \cdot \left[\prod_{i=1}^{n} p\left(B_i, L_i, E_i\right) \cdot \theta \tau^{\theta} S_i^{-(\theta+1)} \mathbb{I}_{\{\tau < S_i\}} \cdot\right.$$
$$\cdot g(S_i, B_i, L_i, E_i) \cdot \frac{(\lambda_i + k_i)^{Y_i^{tot}}}{Y_i^{tot}!} e^{(\lambda_i + k_i)} \mathbb{I}_{\{Y_i^{tot} \in \mathbb{Z}^+\}}$$
$$\left.\cdot \binom{Y_i^{tot}}{Y_i^{src}} \left(\frac{\lambda_i}{\lambda_i + k_i}\right)^{Y_i^{src}} \left(1 - \frac{\lambda_i}{\lambda_i + k_i}\right)^{Y_i^{tot} - Y_i^{src}} \mathbb{I}_{\left\{Y_i^{src} \in \{0, 1, \ldots, Y_i^{tot}\}\right\}}\right]$$

with $\lambda_i \equiv \lambda(S_i, B_i, L_i, E_i)$ and $k_i \equiv k(B_i, L_i, E_i)$.

Posterior Inference (Broken-Pareto)

The broken power-law posterior can be shown to be:

$$p\left(N, \theta, \tau, S_{obs}, Y_{obs}^{src} | n, Y_{obs}^{tot}, B_{obs}, L_{obs}, E_{obs}\right)$$

$$= \frac{1}{p(n, Y_{obs}^{tot}, B_{obs}, L_{obs}, E_{obs})} \cdot \left[\binom{N}{n} \mathbb{I}_{\{n \leq N\}} \cdot (1 - \pi(\theta, \tau))^{(N-n)}\right]$$
$$\cdot \left[\binom{N + a_N - 1}{a_N - 1} \left(\frac{1}{1 + b_N}\right)^N \left(\frac{b_N}{1 + b_N}\right)^{a_N} \mathbb{I}_{\{N \in \mathbb{Z}^+\}}\right] \cdot \left[\prod_{j=1}^{m} \frac{b_j^{a_j}}{\Gamma(a_j)} \theta_j^{a_j - 1} e^{-b_j \theta_j} \mathbb{I}_{\{\theta_j > 0\}}\right]$$
$$\cdot\, p\left(\tau_1, \ldots, \tau_m\right) \mathbb{I}_{\{0 < \tau_1 < \tau_1 < \cdots < \tau_m\}} \cdot \left[\prod_{i=1}^{n} p\left(B_i, L_i, E_i\right) \cdot g(S_i, B_i, L_i, E_i)\right.$$
$$\cdot \sum_{j=1}^{m} \left\{\prod_{l=1}^{j-1} \left(\frac{\tau_{l+1}}{\tau_l}\right)^{-\theta_l}\right\} \left(\frac{\theta_j}{\tau_j}\right) \left(\frac{S_i}{\tau_j}\right)^{-(\theta_j + 1)} \cdot \mathbb{I}_{\left\{\tau_j \leq S_i < \tau_{j+1}\right\}}$$
$$\cdot \frac{(\lambda_i + k_i)^{Y_i^{tot}}}{Y_i^{tot}!} e^{(\lambda_i + k_i)} \mathbb{I}_{\{Y_i^{tot} \in \mathbb{Z}^+\}}$$
$$\left. \cdot \binom{Y_i^{tot}}{Y_i^{src}} \left(\frac{\lambda_i}{\lambda_i + k_i}\right)^{Y_i^{src}} \left(1 - \frac{\lambda_i}{\lambda_i + k_i}\right)^{Y_i^{tot} - Y_i^{src}} \mathbb{I}_{\left\{Y_i^{src} \in \{0, 1, \ldots, Y_i^{tot}\}\right\}}\right]$$

with $\tau_{m+1} = +\infty$, $\lambda_i \equiv \lambda(S_i, B_i, L_i, E_i)$, $k_i \equiv k(B_i, L_i, E_i)$, and $\prod_{i=1}^{0} \left(\frac{\tau_{i+1}}{\tau_i}\right)^{-\theta_i} = 1$.

Computation Strategy: Gibbs Sampler

The Gibbs sampler consists of five steps:

$$[N|n,\theta], \quad [\theta|n,N,S_{obs},\tau], \quad [\tau|n,N,\theta,S_{obs},B_{obs},L_{obs},E_{obs}],$$
$$\left[S_{obs}|N,\theta,\tau,I_{obs},Y^{tot}_{obs},Y^{src}_{obs},B_{obs},L_{obs},E_{obs}\right],$$
$$\left[Y^{src}_{obs}|Y^{tot}_{obs},B_{obs},L_{obs},E_{obs},I_{obs},S_{obs}\right]$$

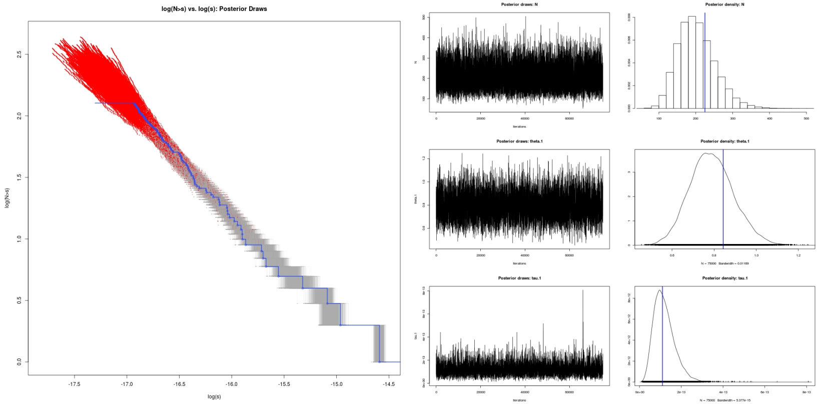
where each parameter block is sampled via various numerical and probabilistic routines, including Metropolis Hastings.

▸ To Model Checking

▸ To Model Validation

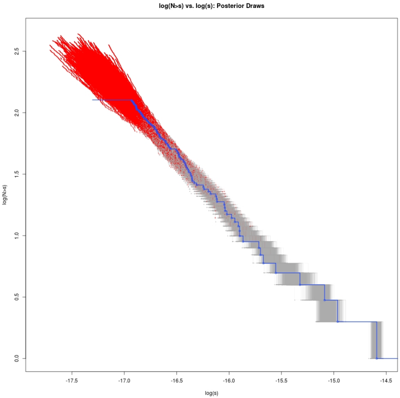
▸ To Computational Details

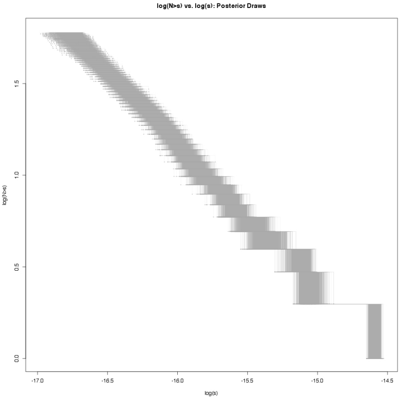
Simulation Example MCMC Output



(L) Posterior logN-logS (red: missing, gray: observed), truth (blue).
(R) Posterior distributions for N, θ, τ

Non-Ignorable Missingness





(L) Nonignorable (full) analysis:	$\hat{\theta} = 0.990$, $(0.803, 1.192)$
(R) Ignorable analysis:	$\hat{\theta} = 0.784$, $(0.520, 0.978)$
Truth:	$\theta = 0.986$

Posterior Predictive Checking

Consider testing the hypothesis:

$$\mathcal{H}_0: \text{ The model is correctly specified, } \textit{vs.},$$
$$\mathcal{H}_1: \text{ The model is not correctly specified.}$$

MCMC draws allow to check the adequacy of the model fit for Bayesian models using the posterior predictive distribution:

$$p(y^*|y) = \int p(y^*, \theta|y) d\theta = \int p(y^*|\theta) \cdot p(\theta|y) d\theta$$

Idea: (Assuming conditional independence) We expect the predictive distribution of new data to look 'similar' to the empirical distribution of the observed data.

Extension: We expect function summaries of interesting features (e.g., test statistics) of the new data to look 'similar' to the empirical distribution of functions of the observed data.

Posterior Predictive Checking

- Procedure:
  1. Sample θ from the posterior distribution $p(\theta|y)$
  2. Given θ, sample y^* (replicate photon counts) from $p(y^*|\theta)$
  3. Compute desired statistic $T(y^*)$
  4. Compute the posterior predictive $p-$value $=$ proportion of the replicate cases in which $T(y^*)$ exceeds $T(y)$:

$$p_b = \mathbb{P}\left(T(y^*) \geq T(y)|y, \mathcal{H}_0\right).$$

- Choice of test statistic $T(y)$ to summarize the desired features of distribution of the photon counts?

  $n = \dim(y_1, ..., y_n)$
  $Max = \max(y_1, ..., y_n)$
  $Med = Q_2(y_1, ..., y_n)$
  $IQR = Q_3(y_1, ..., y_n) - Q_1(y_1, ..., y_n)$
  $Skewness = \text{skewness}(y_1, ..., y_n)$
  $R^2_{S_{crude}}$ = coefficient of determination between $\log N - \log S_{crude}$, where $S_{crude} = (y - bg)\gamma/E$

Posterior Predictive Check: Examples

If the model is true, the posterior predictive *p*-value will almost certainly be very close to 0.5.

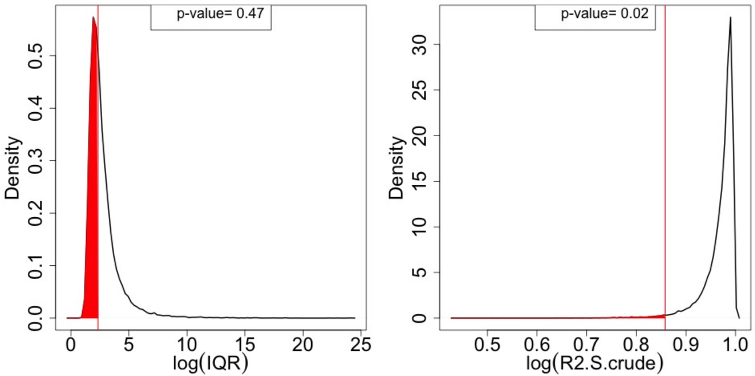


Figure: (Left) If the model is in fact true, the sample *IQR* of new sets of *n* observations will look similar to that of observed *n* data points. (Right) The sample value of $R^2_{S_{crude}}$ of new sets of *n* observations does not look similar to that of observed *n* data points, so the data is unlikely to have been produced from current model.

Bivariate PP $p-$values

The bivariate posterior predictive $p-$value $=$ area under slice of bivariate posterior predictive distr. of two test statistics

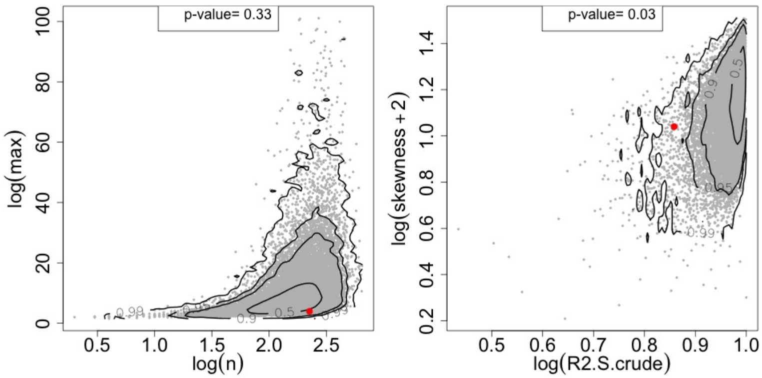


Figure: (Left) Model correctly specified, (Right) Model incorrectly specified

Model Selection

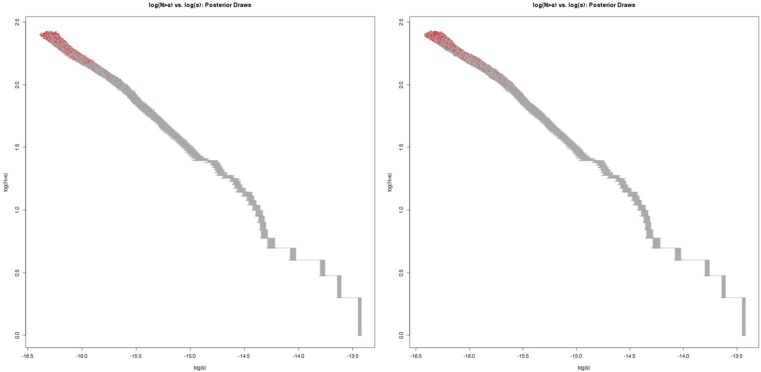


Figure: (Left) Regular Pareto, (Right) Broken Pareto

Note: similar posterior $\log N - \log S$, *but very different predictive and inferential properties*

Model Selection

Model selection is a challenging problem in Bayesian applications. Popular methods are: Bayes Factor, DIC, and WAIC.

- Bayes Factor

$$\frac{p(M_1|y)}{p(M_2|y)} = \frac{p(y|M_1)}{p(y|M_2)}\frac{p(M_1)}{p(M_2)} = \mathbf{BF_{12}}\frac{p(M_1)}{p(M_2)},$$

- Has good model selection performance in simple settings

Problem: Requires evaluating marginal probability of the data given the model M_k

$$p(y|M_k) = p(y) = \int_\Omega p(y|\theta)p(\theta)d\theta.$$

- DIC, Deviance Information Criteria (Spiegelhalter, 2002)

$$DIC = -2 \log p(y|\tilde{\theta}_{Bayes}) + 2p_{DIC}$$
$$p_{DIC} = 2\left\{\log p(y|\tilde{\theta}_{Bayes}) - E[\log p(y|\theta)|y]\right\}$$

- WAIC, Widely Applicable Information Criteria (Watanabe, 2010)

$$WAIC = -2\sum_{i=1}^{N} \log E[p(y_i|\theta)|y] + 2p_{WAIC}$$
$$p_{WAIC} = 2\sum_{i=1}^{N} \{\log E[p(y_i|\theta)|y] - E[\log p(y_i|\theta)|y]\}$$

Problem: In complex hierarchical problems, DIC and WAIC can have poor model selection performance by over-fitting. Alternatives?

Adaptive Fence Method (AF)

Original method of adaptive fence (Jiang et.al., 2008).
For a collection of nested candidate models $M_1, M_2, ..., \tilde{M}_k$, with the full model $\tilde{M}_k$,

- Select a measure of goodness of fit ($Q = -\log p(y|\hat{\theta}_{MLE})$)
- Select optimality criterion (minimal dimension)
- Construct a fence

$$Q(M_j) - Q(\tilde{M}) \leq c$$

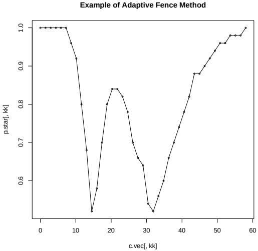
  where c is a constant, and necessarily $Q(M_j) > Q(\tilde{M})$ for all j
- Label model M_c if it is in the Fence and satisfied optimality criterion
- The model to maximize

$$p^* = P(M_c = M_{opt})$$

  is chosen to be the correct model
- The probability P is evaluated under model $\tilde{M}$

Adaptive Fence Method (AF)

Recent article (Jiang, 2014) presents various fence methods.



- Adaptive Fence (AF) selects the cut-off c by maximizing the empirical probability of selection
- Parametric bootstrap is used to evaluate the maximum empirical probability P given parameter MLE of model $\tilde{M}$

Problem: MLE are not easily available for $\log(N) - \log(S)$ application with missing data

Problem: AF has adequate performance when true model is (closest to) the boundary model of the nested candidate models

Problem: Non-nested candidate models, for which no full model can be easily identified?

Bayesian Adaptive Fence Method (BAFM)

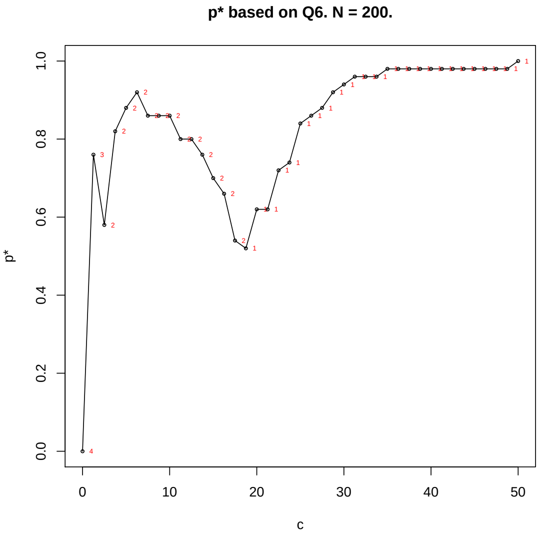
We extend AF method for Bayesian applications and propose

- Use non-parametric bootstrap
- Use alternate (Bayesian version) measures of goodness-of-fit,

$$
\begin{aligned}
Q_1 &= -\sum_{i=1}^{N} \log \left[\frac{1}{L} \sum_{\ell=1}^{L} p(y_i | \beta_{MCMC}^{(\ell)}) \right], \\
Q_2 &= -\log p(y | \tilde{\beta}_{PostMean}), \\
Q_3 &= -\log p(y | \tilde{\beta}_{PostMedian}), \\
Q_4 &= -\log p(y | \tilde{\beta}_{PostMode}), \\
Q_5 &= -\frac{1}{L} \sum_{\ell=1}^{L} \log p(y | \beta_{MCMC}^{(\ell)}), \\
Q_6 &= -Median \left\{ \log p(y | \beta_{MCMC}^{(\ell)}), \ell = 1, \ldots, L \right\}.
\end{aligned}
$$

  where $l = 1, \ldots, L$ draws of MCMC parameters $\theta^{(l)}$ are collected.
- Introduce two phantom boundary models of best and worst fit $Q(\tilde{M}_{k+1}) = 0$ and $Q(M_0) > \max\{Q(M_1), \ldots, Q(M_k)\}$.

Example: p^* vs c plot (BAFM)



Simulation of Model Selection Performance

Simulated true model

SimID	Truth	τ_1, τ_2, τ_3	$\theta_1, \theta_2, \theta_3$
1	bp0	1×10^{-17}, –, –	0.5, –, –
2	bp1	$1 \times 10^{-17}, 5 \times 10^{-17}$, –	0.5, 0.7, –
3	bp1	$1 \times 10^{-17}, 5 \times 10^{-17}$, –	0.5, 1.5, –
4	bp2	$1 \times 10^{-17}, 8 \times 10^{-17}, 1.8 \times 10^{-16}$	0.3, 1.0, 3.0
5	bp2	$1 \times 10^{-17}, 8 \times 10^{-17}, 1.8 \times 10^{-16}$	0.3, 0.7, 0.9

Results

SimID	Truth	DIC_{mean}	DIC_{median}	DIC_{mode}	BF	BAFM
1	bp0	0.51	0.43	0.41	0.23	1.00 (Q_4)
2	bp1	0.69	0.83	0.84	0.29	0.70 (Q_4)
3	bp1	0.93	0.96	0.99	0.49	0.70 (Q_4)
4	bp2	0.90	0.94	0.99	0.53	1.00 (Q_4)
5	bp2	0.67	0.78	0.85	0.34	0.80 (Q_2)

Table: Proportion of correctly selecting true model based on DIC at posterior mean, median, mode, BF with harmonic mean approximation, and Bayesian Adaptive Fence Method for best performing measure Q with automatic peak detection.

BAFM Issues

- Which criterion measure Q to use? Complete likelihood is unknown
- Middle peak does not always exist. Cannot select one best model uniquely
- Very computationally intensive. Need to run MCMC for each bootstrap re-sample

References

- P.D. Baines, I.S. Udaltsova, A. Zezas, V.L. Kashyap (2011) Bayesian Estimation of $\log N - \log S$, *Proc. of Statistical Challenges in Modern Astronomy V*
- P.D. Baines, I.S. Udaltsova, A. Zezas, V.L. Kashyap (2016) Hierarchical Modeling for $\log N - \log S$: Theory, Methods and Applications *Working Paper*
- J. Jiang. et. al (2008) Fence methods for mied model selection. Ann Stat, 36:1669-1692.
- J. Jiang (2014) The Fence Methods. Adv Stat, 2014:830821.
- R.J.A. Little, D.B. Rubin. (2002) *Statistical analysis with missing data*, Wiley.
- R.K.W. Wong, P. Baines, A. Aue, T.C.M. Lee, V.L. Kashyap (2014) Automatic estimation of flux distributions of astrophysical source populations. Ann Appl Stat, 8:1690?1712.
- A. Zezas et al. (2004) Chandra survey of the 'Bar' region of the SMC *Revista Mexicana de Astronomia y Astrofisica (Serie de Conferencias)* Vol. 20. IAU Colloquium 194, pp. 205-205.

Validating Bayesian Computation

Given the complexity of the hierarchical model and computation, it is important to validate that everything works correctly: a self-consistency check.

Procedure:

1. Simulate parameters from the prior, and data from the model, given those parameters
2. Fit the model to obtain posterior intervals
3. Record whether or not the 'true' value of the parameter was within the interval
4. Repeat steps 1, 2, & 3 a large number of times, and calculate the average coverage

$\Rightarrow$ The average and nominal coverages should be equal.

These validation checks are extremely important when dealing with complex procedures.

Model Validation

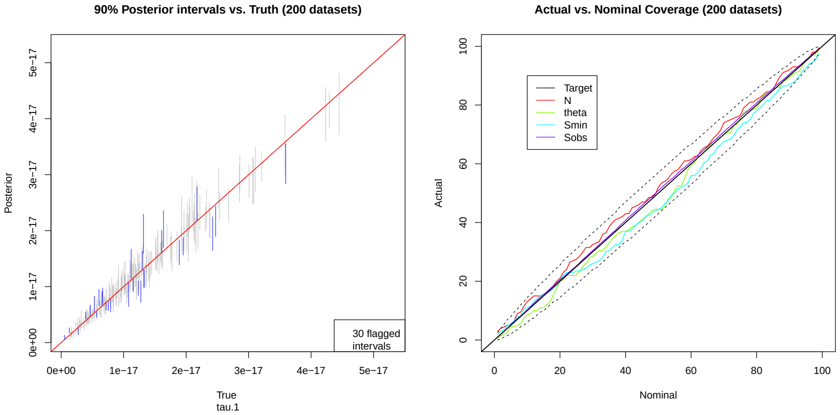


Figure: (Left) Posterior 90% intervals of $\tau_1 = S_{min}$ show reasonable coverage roughly 90% of the time, when nominal coverage is set to 90%. (Right) Average coverage probabilities for each parameter is within the expected bounds at each nominal probability.

Computational Details

The Gibbs sampler consists of five steps:

$$[N|n,\theta],\quad [\theta|n,N,S_{obs},\tau]\,,\quad [\tau|n,N,\theta,S_{obs},B_{obs},L_{obs},E_{obs}]\,,$$
$$\left[S_{obs}|N,\theta,\tau,I_{obs},Y^{tot}_{obs},Y^{src}_{obs},B_{obs},L_{obs},E_{obs}\right],$$
$$\left[Y^{src}_{obs}|Y^{tot}_{obs},B_{obs},L_{obs},E_{obs},I_{obs},S_{obs}\right].$$

- Sample the total number of sources, N, (Numerical Integration):

$$p\left(N|\cdot\right) \propto \begin{pmatrix} N \\ n \end{pmatrix} \mathbb{I}_{\{n\le N\}} \cdot \left(1-\pi(\theta,\tau)\right)^{(N-n)} \cdot p\left(N\right) \cdot p\left(S_{obs}|N,\theta,\tau\right)$$
$$\propto \frac{\Gamma(N+a_N)}{\Gamma(N-n+1)} \cdot \left(\frac{1}{b_N+1}\right)^N \cdot \left(1-\pi(\theta,\tau)\right)^{(N-n)} \mathbb{I}_{\{n\le N\}}$$

- The marginal probability of observing a source $\pi(\theta,\tau)$ is pre-computed via the numerical integration.

$$\pi(\theta,\tau) = \int g(S_i,B_i,L_i,E_i) \cdot p(S_i|\theta,\tau) \cdot p(B_i,L_i,E_i)\; dS_i\; dB_i\; dL_i\; dE_i.$$

Computational Details cont...

The Gibbs sampler consists of five steps:

$$[N|n,\theta], \quad [\theta|n,N,S_{obs},\tau], \quad [\tau|n,N,\theta,S_{obs},B_{obs},L_{obs},E_{obs}],$$
$$\left[S_{obs}|N,\theta,\tau,I_{obs},Y^{tot}_{obs},Y^{src}_{obs},B_{obs},L_{obs},E_{obs}\right],$$
$$\left[Y^{src}_{obs}|Y^{tot}_{obs},B_{obs},L_{obs},E_{obs},I_{obs},S_{obs}\right].$$

- Sample the power-law slope, θ, (Metropolis Hastings using a Normal-proposal):

$$p(\theta|\cdot) \propto p(\theta)\cdot p(S_{obs}|N,\theta,\tau)\cdot (1-\pi(\theta,\tau))^{(N-n)}$$
$$\propto (1-\pi(\theta,\tau))^{(N-n)}\cdot \mathrm{Gamma}\left(\theta; a+n, b+\sum_{i=1}^{n}\log\left(\frac{S_i}{\tau}\right)\right)$$

- Sample the observed photon counts $Y^{src}_{obs,i}$, $i=1,\ldots,n$:

$$p(Y^{src}_i|\cdot) \propto p(Y^{src}_i|Y^{tot}_i,S_i,B_i,L_i,E_i)$$
$$\sim \mathrm{Bin}\left(Y^{src}_i; Y^{tot}_i, \frac{\lambda(S_i,B_i,L_i,E_i)}{\lambda(S_i,B_i,L_i,E_i)+k(B_i,L_i,E_i)}\right)$$

Computational Details cont. . .

- Sample the minimum flux τ (Metropolis Hastings using a Truncated-Normal-proposal after log-transformation):

$$p(\tau|\cdot) \propto p(\tau)\cdot p(\theta)\cdot p(N)\cdot p\left(B_{obs}, L_{obs}, E_{obs}\right) \cdot p\left(n, S_{obs}, I_{obs}|N, \theta, \tau, B_{obs}, L_{obs}, E_{obs}\right) \propto \mathrm{Gamma}\left(\tau; a_m + n\theta, b_m\right)\cdot\left(1-\pi(\theta,\tau)\right)^{N-n}\cdot\mathbb{I}_{\{0<\tau<c_m\}}$$

$$p(\eta = \log(\tau)|\cdot) = e^{\eta}\cdot p(\tau = e^{\eta}|\cdot) \propto e^{\eta(n\theta + a_m + 1)}\cdot e^{-b_m e^{\eta}}\cdot\left(1-\pi(\theta,\tau = e^{\eta})\right)^{N-n}\cdot\mathbb{I}_{\{\eta<\log(c_m)\}}.$$

- Sample the fluxes $S_{obs,i}, i = 1,\ldots,n$ (Metropolis Hastings using a Normal-proposal):

$$p\left(S_i|\cdot\right) \propto p\left(S_i|N,\theta,\tau\right)\cdot p\left(I_i = 1|S_i, B_i, L_i, E_i\right)\cdot p\left(Y_i^{tot}|S_i, B_i, L_i, E_i\right)\cdot \cdot p\left(Y_i^{src}|Y_i^{tot}, S_i, B_i, L_i, E_i\right) \sim Pareto(S_i;\theta,\tau)\cdot g(S_i, B_i, L_i, E_i)\cdot Pois(Y_i^{tot};\lambda(S_i, B_i, L_i, E_i)+k(B_i, L_i, E_i)) \cdot \mathrm{Bin}\left(Y_i^{src}; Y_i^{tot}, \frac{\lambda(S_i, B_i, L_i, E_i)}{\lambda(S_i, B_i, L_i, E_i)+k(B_i, L_i, E_i)}\right)$$

Computational Details cont...

- Sample $\theta = (\theta_1, \dots, \theta_m)^T$: (Metropolis-Hastings using a Normal-proposal)

$$p(\theta | \cdot) \propto \left[(1 - \pi(\theta, \tau))^{N-n} \right] \cdot$$

$$\cdot \prod_{j=1}^{m} \mathrm{Gamma} \left(\theta_j;\ a_j + n(j) - 1,\ b_j + \mathbb{I}_{\{j \neq m\}} \log\left(\frac{\tau_{j+1}}{\tau_j}\right) \sum_{i=1}^{m} \left[n(i) \mathbb{I}_{\{i \geq j+1\}} \right] + \sum_{i \in \mathcal{I}(j)} \log\left(\frac{s_i}{\tau_j}\right) \right),$$

where $\mathcal{I}(j) = \{ i : \tau_j \leq s_i < \tau_{j+1} \}$ and $n(j)$ is the cardinality of $\mathcal{I}(j)$ i.e., $\mathcal{I}(j)$ ($n(j)$) denotes the set (number) of source indices whose flux is contained in the interval corresponding to the j-th mixture component.

- Sample the break-points $\tilde{\tau} = (\tau_2, \dots, \tau_m)^T$ (Metropolis-Hastings based on original transformed scale $\eta_j = h(\tilde{\tau} | \tau_1) = \log(\tau_j - \tau_{j-1}), j = 2, \dots, m$):

$$p(\eta | \cdot) = p(h(\tilde{\tau} | \tau_1) | \cdot) \propto \left[(1 - \pi(\theta, \tau))^{(N-n)} \right] \cdot \exp\left[-\frac{1}{2} \sum_{j=2}^{m} \{ c_j (\eta_j - \mu_j) \}^2 \right] \cdot \mathbb{I}_{\{\tau_1 < \tau_2 < \dots < \tau_m\}}$$

$$\cdot \left[\prod_{j=1}^{m} \left\{ \prod_{l=1}^{j-1} \left(\frac{\tau_{l+1}}{\tau_l} \right)^{-\theta_l} \right\}^{n(j)} \prod_{i \in \mathcal{I}(j)} \left(\frac{\theta_j}{\tau_j} \right) \left(\frac{s_i}{\tau_j} \right)^{-(\theta_j + 1)} \mathbb{I}_{\{\tau_1 < \min(s_1, \dots, s_n)\}} \right],$$

Fluxes of missing sources can (optionally) be imputed to produce posterior draws of a 'corrected' $\log N - \log S$

- Impute missing fluxes $S_{mis,i}, i = 1, \ldots, n$ (Rejection Sampling):

$$
\begin{aligned}
&(B_i, L_i, E_i) \sim p(B_i, L_i, E_i) \\
&S_i | n, N, \theta, \tau, B_i, L_i, E_i, I_i = 0 \\
&\quad \sim (1 - g(S_i, B_i, L_i, E_i)) \cdot Pareto(S_i; \theta, \tau).
\end{aligned}
$$

Computational Notes

Some important things to note:

- For single power-law models computation is relatively fast (few hours), and insensitive to the number of missing sources
- The fluxes of the missing sources need not be imputed
- Fluxes of missing sources can (optionally) be imputed to produce posterior draws of a 'corrected' $\log N - \log S$
- Computation for the broken-power law model is slower (day)
- Generalized mixtures of Pareto's (or other forms) require only minor modifications of general scheme

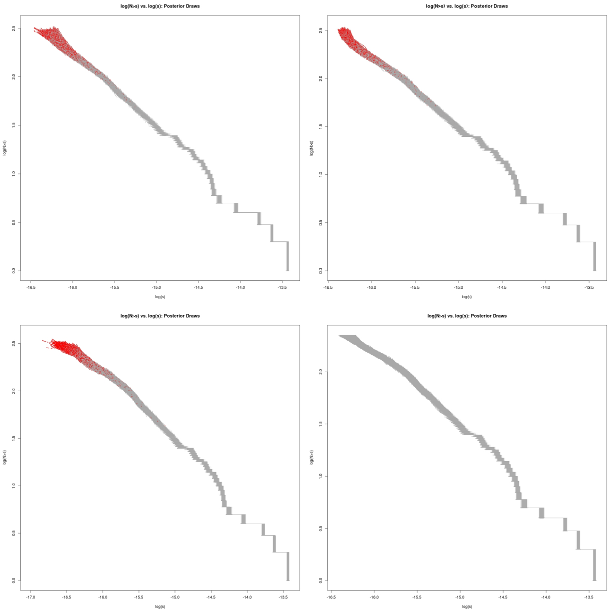
Application: X-ray emission from point sources in the *CHANDRA* Deep Field North (CDFN) survey.

- The deepest 0.5-8.0 keV band survey ever made
- 2 Ms exposure covering field of 448 sq. arcmin
- Subsample sources based on off-axis angle threshold of 8 arcmin

---

Dataset consists of 225 observed sources

Priors: $a = 10, b = 10, \mathbb{E}[N] = 300, \mathrm{Var}(N) = 100^2,$
$\mathbb{E}[\tau_1] = 1.5 \times 10^{-17}, \mathrm{Var}(\tau_1) = (9.0 \times 10^{-18})^2,$
$\mathbb{E}[\log(\tau_2 - \tau_1)] = -38, \mathrm{Var}(\log(\tau_2 - \tau_1)) = 0.7^2$



CDFN log(N)-log(S)

Method	bp0	bp1	bp2	bp0 $g=1$	BAFM	Sel.
DIC.Mean	1838.3	**1830.8**	1838.6	1873.3	Q_1	bp2
DIC.Median	1842.2	**1834.3**	1838.6	1876.9	Q_2	bp2
DIC.Mode	1839.7	**1810.1**	1827.3	1860.4	Q_3	bp2
DIC.V	1880.5	1865.2	**1851.3**	1920.4	Q_4	bp2
WAIC1	1715.9	1702.6	**1698.9**	1746.3	Q_5	bp2
WAIC1	1795.9	1781.8	**1778.8**	1826.2	Q_6	bp2

TABLE: DIC, WAIC, BAFM for CDFN analysis.

	bp0	bp1	bp2	$g = 1$
# obs src	0.34	0.08 (0.63)	0.07 (0.08)	0.23
Minimum ph ct	0.25	0.41 (0.40)	0.40 (0.41)	0.36
Maximum ph ct	0.08	0.23 (0.26)	0.22 (.24)	0.08
Median ph ct	0.17	0.05 (0.10)	0.03 (0.05)	0.40
Lower quartile of ph cts	0.11	0.17 (0.23)	0.17 (0.17)	0.28
Upper quantile of ph cts	0.14	0.08 (0.12)	0.04 (0.07)	0.32
IQR of ph cts	0.15	0.08 (0.11)	0.04 (0.07)	0.32
Crude estimate of R^2	0.10	0.32 (0.33)	0.31 (0.31)	0.15
# obs src vs. Median ph ct	0.68	0.23 (0.22)	0.15 (0.20)	0.65
Lower vs. Upper quartile	0.78	0.26 (0.26)	0.16 (0.21)	0.75
# obs sources vs. IQR	0.57	0.20 (0.20)	0.13 (0.20)	0.72
# obs sources vs. R^2	0.12	0.59 (0.59)	0.59 (0.60)	0.18

TABLE: Posterior predictive p-values for CDFN analysis. (Adjusted for low-flux*)

Parameter	Estimate Mean (Med)	95% (central) interval	(Wong et.al.,2014) Estimate	SE
θ_1	0.64 (0.59)	(0.40,1.09)	0.483	0.060
θ_2	0.62 (0.56)	(0.36,1.18)	0.854	0.224
θ_3	0.83 (0.82)	(0.67,1.03)		
$\log_{10}(\tau_1)$	-16.45 (-16.44)	(-16.60,-16.35)	-16.344	0.030
$\log_{10}(\tau_2)$	-16.00 (-16.11)	(-16.36,-15.61)	-15.657	0.271
$\log_{10}(\tau_3)$	-15.72 (-15.74)	(-16.00,-15.47)		
N	281 (280)	(259,312)		

Table: Parameter estimates for CDFN analysis based on bp2 model. Compared to competing method for estimation of the number of breakpoints and parameters via interwoven EM (Wong et.al., 2014)

CDFN Conclusions

- Estimated slope of 0.46 (median), 0.487 (mean)
- 95% (central) interval $= (0.312, 0.681)$
- Wong et al. (2014) estimated broken power-law slopes of $\hat{\theta} = (0.483, 0.854)$.
- Estimated missing data fraction 24%, interval:$(0.0, 28.25\%)$
- Evidence of a possible break in the power-law in the observed $\log N - \log S$.

Note: The ignorable analysis gives a posterior median of 0.363, mean of 0.367, and 95% interval $(0.248, 0.511)$.