# Bayesian Estimation of $\log N - \log S$

Paul D. Baines

Department of Statistics
University of California, Davis

March 6th, 2012

# Introduction

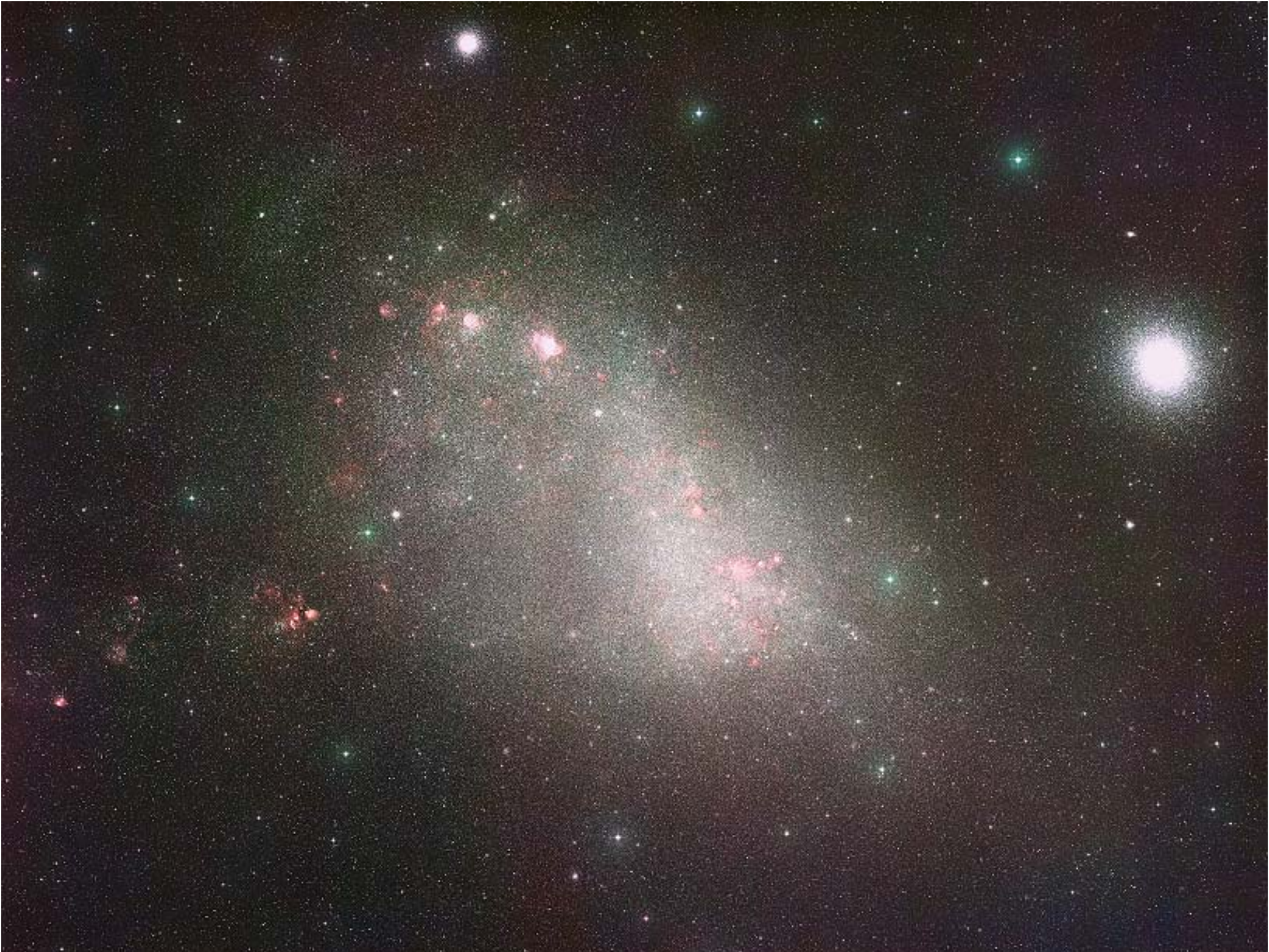
## Project Goals

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

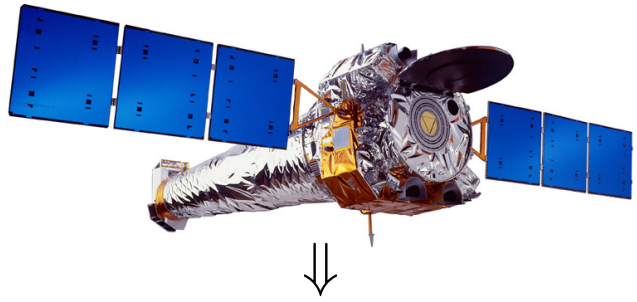
*More generally, to also infer luminosity functions for source populations.*

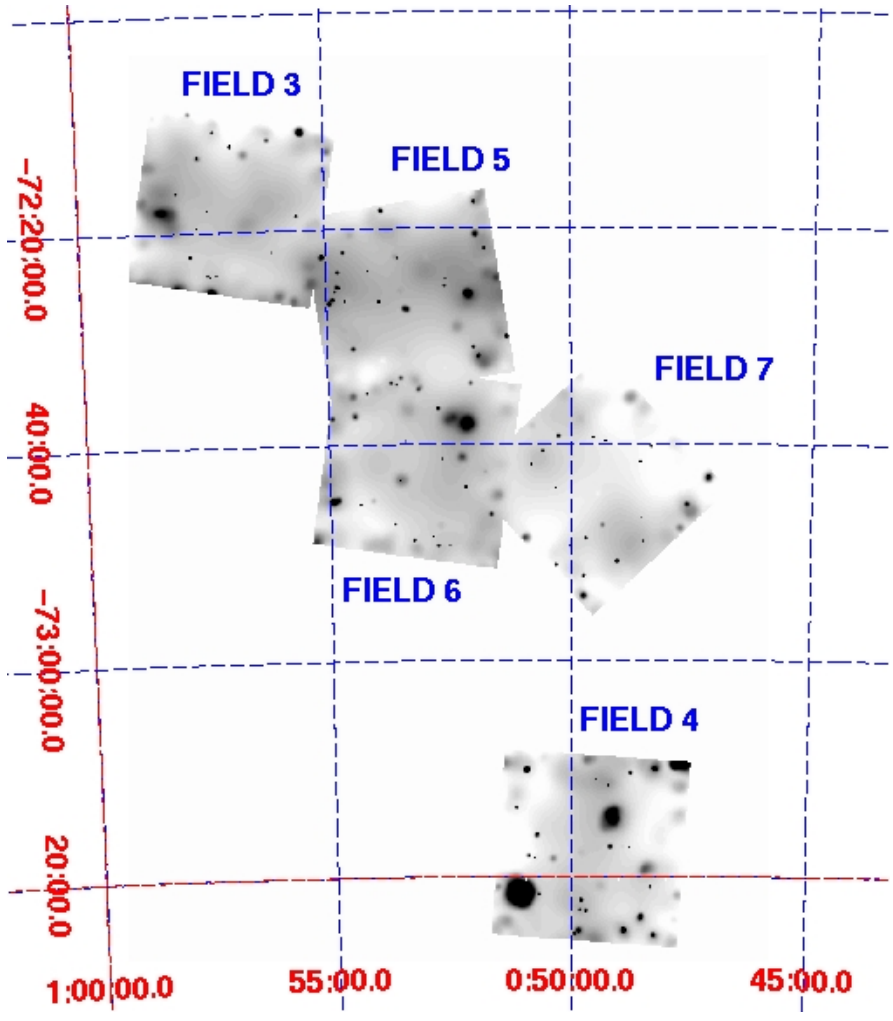
---

Collaborators: Irina Udaltsova (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

# Estimating Flux Distributions

Our goal is to estimate the distribution of fluxes for the source population.

Toy example: Uniformly distributed source population, same intrinsic luminosity $L_0$, then for telescopic sensitivity $S$, sources will be detectable to:

$$d = \sqrt{\frac{L_0}{4\pi S}}$$

The number of sources within this distance is then:

$$N(< d) = N(> S) = n_0 \left( \frac{4\pi}{3} d^3 \right) \propto S^{-3/2}$$

Therefore, the convention is to plot the log (base 10) of the cumulative number of sources as a function of log (base 10) flux.

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

In the simple case we have:

$$\log_{10}\left(N(>S)\right) = \beta_0 + \beta_1 \log_{10}(S),$$

Cosmology complicates the anticipated linearity somewhat, but in many cases the relationship is approximately linear.

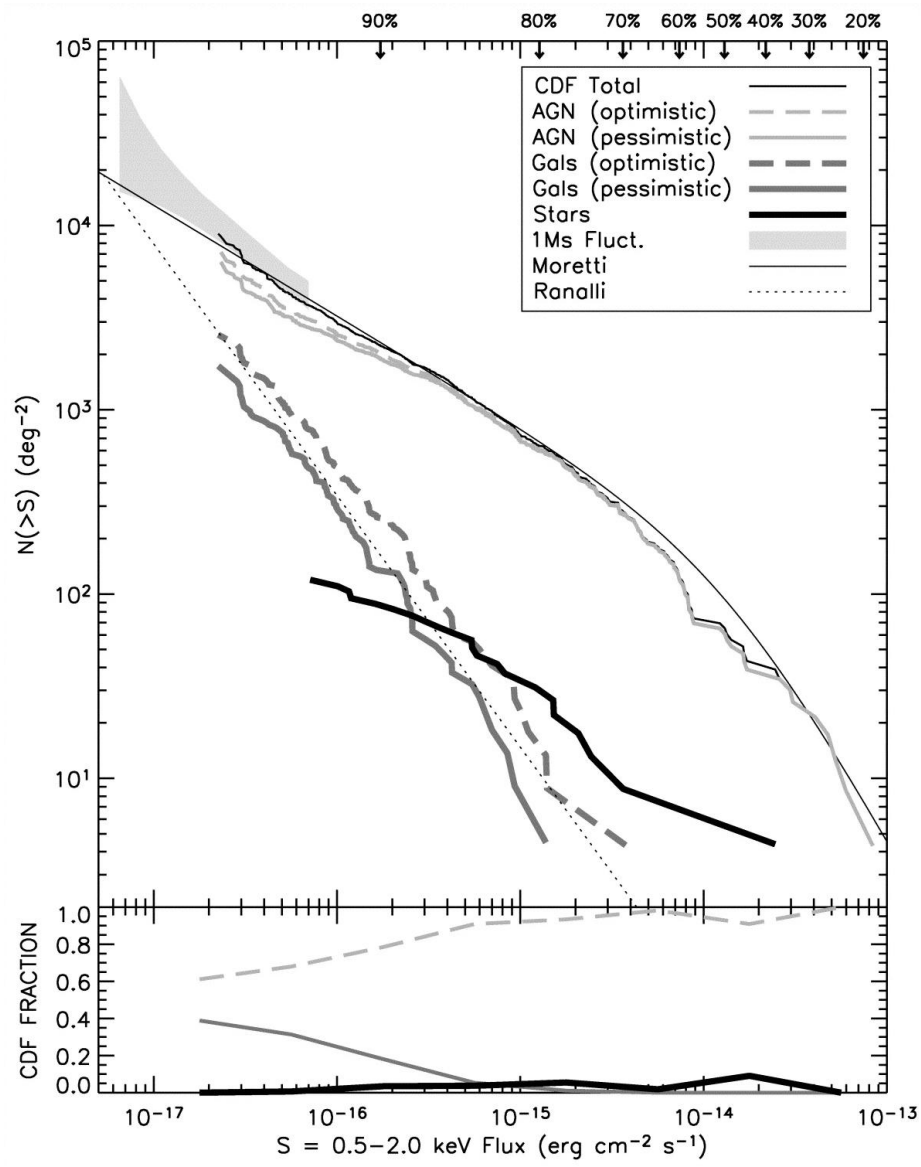
More generally, knowing the specific relationship for different objects (e.g., stars, galaxies, pulsars) gives a lot of information about the underlying physics (e.g., the mass of galaxies).

---

Primary Goal: Estimate $\beta_1$, 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 $s$).

# Process/Analyze Raw Image to obtain:



# Inferential Process

To infer the $\log N - \log S$ relationship there are a few steps:

1. Collect raw data images
2. Run a detection algorithm to extract 'sources' from the image
3. Produce a dataset describing the photon counts of all 'sources' (and uncertainty about them, background etc.)
4. Infer physical properties about the source population (e.g., the $\log N - \log S$ distribution) from this dataset

Our analysis is focused on the final step – accounting for some (but not all) of the detector-induced uncertainties...

Adding further layers to the analysis to start with raw images is possible but that is for a later time...

# The Data

The data is essentially just a list of photon counts – with some extra information about the background and detector properties.

| Src_ID | Bgr_intensity | counts | Src_area | Off_axis | Effective_area |
|---|---|---|---|---|---|
| 1 | 3.16 | 285 | 1720 | 4.98 | 307.306 |
| 2 | 0.19 | 133 | 96 | 5.72 | 143.110 |
| 3 | 0.61 | 37 | 396 | 6.17 | 40.4481 |
| 4 | 0.22 | 15 | 128 | 6.34 | 15.9011 |
| 5 | 0.96 | 12 | 604 | 4.51 | 13.3569 |

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

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

# Naïve Fitting

One can obtain simple estimates for the fluxes source-by-source (e.g., MOM). Then do a least squares ($\chi^2$) fit for the slope.

Slightly more sophisticated, we could fit using maximum likelihood methods, but these generally fail to properly account for incompleteness and other detector-based uncertainties.

We will approach the problem with the goal with building a framework that can account for the many sources of uncertainty.

# Probabilistic Connection

The standard approaches make it difficult to coherently incorporate detector effects and uncertainties (more later).

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

(Follows from log-linearity of the survival function)

The probabilistic representation for the flux distribution now allows for more rigorous analysis by embedding within a hierarchical model.

# Beyond the Pareto

With Pareto flux distribution we obtain a linear relationship on the $log\, N - \log(N > S)$ scale. In general, with complete-data flux distribution $G$, we have:

$$S_i \overset{iid}{\sim} G \qquad \Rightarrow \qquad \log_{10}(1 - F_G(s)) := H(\log_{10}(s)). \quad (1)$$

The function $H$ is linear if and only if $G$ is the Pareto distribution. Our framework will allow for flexible specification of the (parametrized) flux distribution.

Since linearity has both theoretical and empirical support, a commonly used generalization is a broken power-law:

$$\log_{10}(1 - F_G(s)) = \begin{cases} \alpha_0 - \theta_0 \log_{10}(s) & s \leq K \\ \alpha_1 - \theta_1 \log_{10}(s) & s > K \end{cases}, \quad (2)$$

subject to a continuity constraint.

# Broken Power-Law Modeling

It is natural to ask then: 'What, if any, probability distribution does the broken power-law in (2) correspond to?'.

$$Y \sim I \cdot X_0 + (1 - I) \cdot X_1,$$

where:

$$I \sim \text{Bin}\left(1, \left[1 - \left(\frac{K}{S_{min}}\right)^{-\theta_0}\right]\right)$$

$$X_0 \sim \text{Truncated-Pareto}\left(S_{min}, \theta_0, K\right), \qquad X_1 \sim \text{Pareto}\left(K, \theta_1\right).$$

The result is also an 'if and only if' result:

## Theorem

*Any distribution whose* $\log N - \log S$ *relationship is a broken power law, with* $M$ *breakpoints, can be represented as a mixture of* $M$ *truncated Pareto distributions and another (untruncated) Pareto distribution.*

# Multiple Broken Power-Law Model

The broken power-law model can be generalized to a piece-wise linear relationship with arbitrary number, $m$, of break-points:

$$F_G(s) = \begin{cases} 1 - \alpha_0^* s^{-\theta_0} & K_0 \leq s < K_1 \\ \vdots & \vdots \\ 1 - \alpha_{m-1}^* s^{-\theta_{m-1}} & K_{m-1} \leq s < K_m \\ 1 - \alpha_m^* s^{-\theta_m} & s \geq K_m \end{cases}.$$

Similarly to the single broken power-law setting we can show:

$$F_j(s) = \frac{1}{p_j} \left\{ 1 - \sum_{i=0}^{j-1} p_i \right\} \left[ 1 - \left( \frac{s}{K_j} \right)^{-\theta_j} \right], \qquad j = 0, \ldots, m. \tag{3}$$

Constraints lead to a recursive relationship for mixture probabilities:

$$p_j = \left[ 1 - \left( \frac{K_{j+1}}{K_j} \right)^{-\theta_j} \right] \left( 1 - \sum_{i=0}^{j-1} p_i \right), \qquad j = 0, 1, \ldots, m. \tag{4}$$

i.e., the multiple broken power-law assumption corresponds to:

$$Y \sim I_0 X_0 + I_1 X_1 + \cdots + I_m X_m$$

$$I_j \sim \text{Multinomial}\left(1; p_0, p_1, \ldots, p_m\right), \quad X_j \sim \text{Truncated-Pareto}\left(K_j, \theta_j, K_{j+1}\right).$$

# Physically Motivated Fitting

The insight from the probabilistic setting reveals that the broken power-law model has a number of unphysical properties (to be expected).

Notably, it requires an 'initial source population' to have a sharp cut-off, before yielding to a secondary source population present only above the threshold.

More physically realistic descriptions are also more natural statistically e.g., a mixture of populations:

$$S \sim \sum_{j=1}^{m} I_j X_j, \quad I_j \overset{iid}{\sim} \text{Multinomial}(1; p_1, \ldots, p_m), \quad X_j \sim \text{Pareto}\,(K_j, \theta_j)\,.$$

The goal may be to detect possible 'subpopulations', or just to fit a more flexible model. Note: the resulting $\log N - \log S$ plot will now be curved!

# Mind the Gap

The previous discussion centered around the flux distribution.

Now things get interesting:

- We only observe photon counts from the source with intensity proportional to the flux
- There is background contamination for all sources
- Different sensitivities across the detector
- Some sources will not be observed to detector limitations
- We do not know how many sources there actually are
- Some 'sources' extracted from the image may not actually be sources

▸ To Model

# Missing Data Overview

There are many potential causes of 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.

Some are more problematic than others. . .

# Missing Data Mechanisms

In the nicest possible case, if the particular data that is missing does not depend on any unobserved values then we can essentially ignore the missing data.

In this context, whether a source is observed is a function of its source count (intensity) – which is unobserved for unobserved sources. This missing data mechanism is non-ignorable, and needs to be carefully accounted for in the analysis.

Heuristically: Whether or not a source is missing provides information about the parameters we are trying to estimate.

# Non-Ignorable Missingness

Let $Y_{com} = (Y_{obs}, Y_{mis})$, and the missing data indicators be $M$. The missing data mechanism is defined to be $p(M|Y_{com}, \psi)$.

The observed data likelihood is based on:

$$p(Y_{obs}|\theta) = \int p(Y_{com}|\theta) dY_{mis}. \tag{5}$$

The complete data likelihood is based on:

$$p(Y_{obs}, M|\theta, \psi) = \int p(Y_{com}|\theta) p(M|Y_{com}, \psi) dY_{mis}. \tag{6}$$

Inference based on (5) is valid only if inference about $\theta$ agrees with that from (6). In these cases, the missing data mechanism is called ignorable.

Main condition for ignorability is the data be missing at random (MAR):

$$p(M|Y_{com}, \psi) = p(M|Y_{obs}, \psi) \qquad \forall \quad Y_{mis}, \psi. \tag{7}$$

Here this is not true! Missingness depends on the unobserved flux.

# The Model

Standard power-law flux distribution:

$$S_i | S_{min}, \theta \overset{iid}{\sim} \text{Pareto}\left(\theta, S_{min}\right), \qquad i = 1, \ldots, N.$$

Source and background photon counts:

$$Y_i^{tot} | S_i, L_i, E_i \overset{\perp\!\!\!\perp}{\sim} Pois\left(\lambda(S_i, L_i, E_i) + k(L_i, E_i)\right), \qquad i = 1, \ldots, N,$$

Incompleteness, missing data indicators:

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

Prior distributions:

$$N \sim NegBinom\left(\alpha, \beta\right),$$
$$\theta \sim \text{Gamma}(a, b).$$

# Model Overview

Some important notes:

- The dimension of the missing data is unknown (care must be taken with conditioning)
- Incompleteness function $g$ can take any form and is problem-specific
- The lower limit, $S_{min}$, is fixed: inferential sensitivity?
- Prior parameters can be science-based or 'weakly informative'

# Broken Power-Law Model

Broken power-law flux distribution (known break-points $\vec{C}$):

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

Source and background photon counts:

$$Y_i^{tot} | S_i, L_i, E_i \overset{\perp\!\!\!\perp}{\sim} \mathit{Pois}\left(\lambda(S_i, L_i, E_i) + k(L_i, E_i)\right), \qquad i = 1, \ldots, N,$$

Incompleteness, missing data indicators:

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

Prior distributions:

$$N \sim \mathit{NegBinom}\left(\alpha, \beta\right),$$

$$\theta_j \overset{\perp\!\!\!\perp}{\sim} \text{Gamma}(a_j, b_j), \qquad j = 1, \ldots, M.$$

# Mixture Pareto Model

Mixture Pareto flux distribution (fixed number $m$ mixture components and known mixture-points $\vec{K} = (K_1, \ldots, K_m)$):

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

i.e.,

$$S_i|\vec{K}, \theta \sim \sum_{j=1}^{m} I_j X_j, \quad I_j \sim \text{Multinomial}(1; p_1, \ldots, p_m), \quad X_j \sim \text{Pareto}\left(K_j, \theta_j\right).$$

Source and background photon counts:

$$Y_i^{tot}|S_i, L_i, E_i \overset{\perp\!\!\!\perp}{\sim} Pois\left(\lambda(S_i, L_i, E_i) + k(L_i, E_i)\right), \qquad i = 1, \ldots, N,$$

Incompleteness, missing data indicators:

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

Prior distributions:

$$N \sim NegBinom\left(\alpha, \beta\right),$$

$$\theta_j \overset{\perp\!\!\!\perp}{\sim} \text{Gamma}(a_j, b_j), \qquad j = 1, \ldots, M.$$

# Posterior Inference

For all of these versions of the model, inference about $\theta$, $N$ and $S$ is based on the observed data posterior distribution. Care must be taken with the variable dimension marginalization over the unobserved fluxes.

The posterior can be shown to be:

$$p\left(N, \theta, S_{obs}, S_{mis}, Y_{obs}^{src} | n, Y_{obs}^{tot}\right) \propto$$
$$p\left(N\right) \cdot p\left(\theta | N\right) \cdot p\left(n | N, \theta\right) \cdot p\left(S_{obs} | n, N, \theta\right) \cdot$$
$$p\left(S_{mis} | n, N, \theta\right) \cdot p\left(Y_{obs}^{tot} | n, N, S_{obs}\right) \cdot p\left(Y_{obs}^{src} | n, N, Y_{obs}^{tot}, S_{obs}\right)$$

Computation is performed by Gibbs sampling.

▸ To Model Checking

# Computational Details

The Gibbs sampler consists of four steps:

$$[N|n,\theta]\,, \quad [\theta|n,N,S_{obs},S_{mis}]\,, \quad \left[Y_{obs}^{src}|n,Y_{obs}^{tot},\theta,S_{obs},S_{mis}\right],$$
$$\left[S_{obs}|n,N,Y_{obs}^{tot},Y_{obs}^{src},\theta\right], \quad [S_{mis}|n,N,\theta]\,.$$

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

$$p(N|n,\theta) \propto \frac{\Gamma(N+\alpha)}{\Gamma(N-n+1)}\left(\frac{1-\pi(\theta)}{\beta+1}\right)^{N} \mathbb{I}\{N \geq n\}$$

  Note: The (prior) marginal detection probability $\pi(\theta)$ is pre-computed via the numerical integration.

- Sample the power-law slope, $\theta$, (Rejection Sampling):

$$\theta|n,N,S_{obs},S_{mis} \sim \mathrm{Bin}\left(n;N,\pi(\theta)\right) \cdot \mathrm{Gamma}\left(\theta;a+N,b+\sum_{i=1}^{n}\log\left(\frac{S_i^{(obs)}}{S_{min}}\right)+\sum_{j=1}^{N-n}\log\left(\frac{S_j^{(mis)}}{S_{min}}\right)\right)$$

# Computational Details cont. . .

- Sample the observed photon counts:

$$Y^{src}_{obs,i}|n, Y^{tot}_{obs,i}, S_{obs,i} \sim \text{Bin}\left(Y^{tot}_{obs,i}, \frac{\lambda(S_{obs,i}, L_{obs,i}, E_{obs,i})}{\lambda(S_{obs,i}, L_{obs,i}, E_{obs,i}) + k}\right),$$

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

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

$$\begin{aligned} S_{obs,i}|n, N, Y^{tot}_{obs,i}, Y^{src}_{obs,i}, \theta & \\ \sim \text{Pareto}(S_i; \theta, S_{min}) g(S_i, E_i, L_i) & \\ \cdot \text{Pois}(Y^{tot}_{obs,i}; \lambda(S_{obs,i}, E_{obs,i}, L_{obs,i}) + k(E_{obs,i}, L_{obs,i})) & \\ \cdot \text{Bin}\left(Y^{src}_{obs,i}; Y^{tot}_{obs,i}, \frac{\lambda(S_{obs,i}, E_{obs,i}, L_{obs,i})}{\lambda(S_{obs,i}, E_{obs,i}, L_{obs,i}) + k(E_{obs,i}, L_{obs,i})}\right) & \end{aligned} \tag{8}$$

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

$$S_{mis,i}|n, N, \theta \sim \text{Pareto}(S_i; \theta, S_{min})(1 - g(S_i, E_i, L_i))$$

# Computational Notes

Some important things to note:

- For single power-law models computation is fast (secs), 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
- Generalized mixtures of Pareto's (or other forms) require only minor modifications of general scheme

# Model Checking

The model we have presented contains a lot of assumptions. Obviously, where possible, we would like to check those assumptions.

Bayesian model checking is primarily built around the posterior predictive distribution, and can be done in a few different ways...

# Posterior Predictive $p-$values

The posterior predictive $p-$value, first proposed by Rubin (1984), is a valuable tool for assessing the adequacy of the model fit for Bayesian models.

It is based on the posterior predictive distribution:

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

where the second identity follows only if the predictive distribution of $y^*$ depends only on $\theta$ (this is usually the case).

# Posterior Predictive Checking

The idea:

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

More generally:

(Assuming conditional independence) we expect the predictive distribution of functions (e.g., test statistics) of the new data to look 'similar' to the empirical distribution of functions of the observed data.

In fact, the idea extends beyond test statistics (functions of the data) to 'generalized test variables' (functions of both the data and parameters) (Meng, 1994).

# Posterior Predictive $p-$values

Consider testing the hypothesis:

$$\begin{aligned} \mathcal{H}_0 &: \text{ The model is correctly specified , } vs., \\ \mathcal{H}_1 &: \text{ The model is not correctly specified .} \end{aligned}$$

Select a test statistic $T(x)$ to perform the test, then we define the posterior predictive $p-$value to be:

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

In practice we use:

$$p_b^* = 2 \cdot \min\left\{ p_b, 1 - p_b \right\}.$$

# Posterior Predictive $p-$values

Some things to note:

- This is easy to compute since $p(y^*|y)$ is easy to sample from:
  1. Sample $\theta$ from the posterior distribution $p(\theta|y)$
  2. Given $\theta$, sample $y^*$ from $p(y^*|\theta)$

  Use MC to compute the posterior predictive $p-$values.

- Minimal extra work once posterior samples have been obtained

- Choice of test statistic: mean? median? max? min? other?

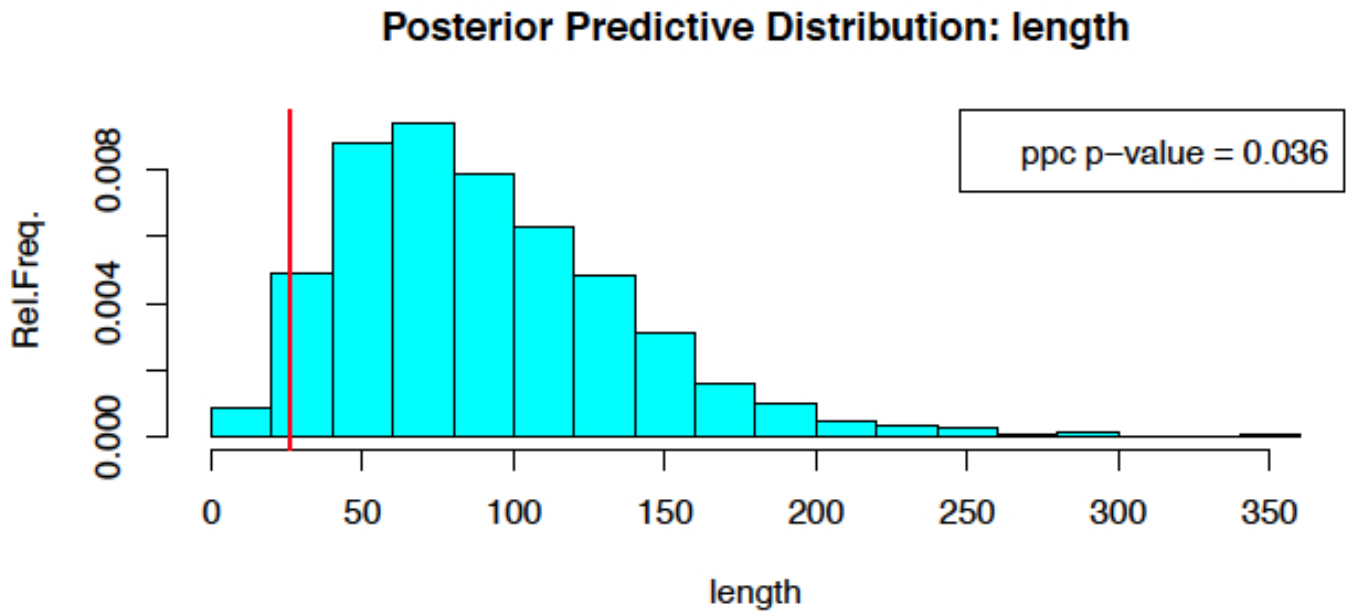
- Test statistic must be a function of photon counts

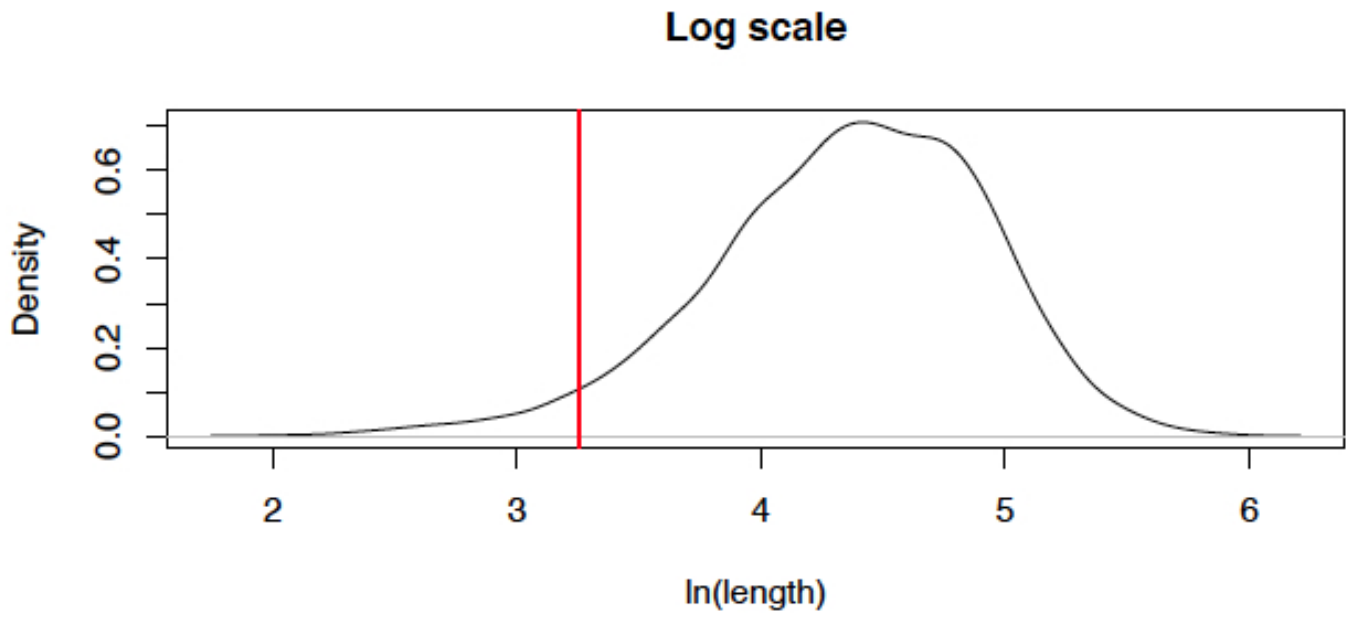
# Posterior Predictive Check: Example

For example, in the logN-logS example our predictive data are the photon counts for 'replicate datasets'.

We could take the number of observed sources in the replicate dataset as our statistic.

This will essentially test part of the assumptions about the missing data mechanism...


Posterior Predictive Distribution: length
ppc p−value = 0.036
Rel.Freq.
0.000
0.004
0.008
0
50
100
150
200
250
300
350
length


Log scale
Density
0.0
0.2
0.4
0.6
2
3
4
5
6
ln(length)

# Extended Posterior Predictive $p-$values

The idea behind posterior predictive $p-$values extends to multivariate test statistics, First, lets reformulate:

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

as the tail probability:

$$p_b = \int I_{\{T(y^*)\geq T(y)\}} p\left(y^*|y, \mathcal{H}_0\right) dy^*$$

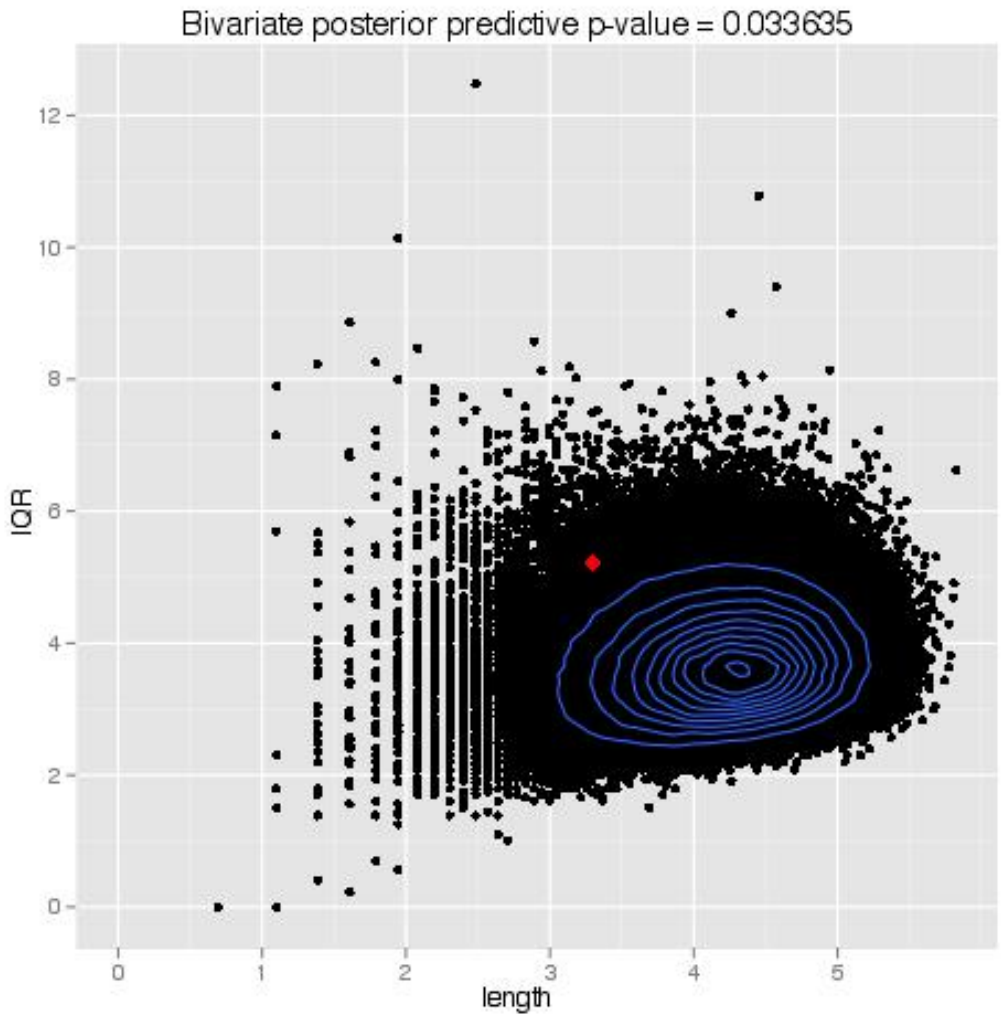
Then the multivariate generalization is immediate.

# Multivariate PP $p-$values

The multivariate posterior predictive $p-$values can yield extra insight when the dimensions of the test statistic are highly correlated (i.e., have a strong dependence structure).

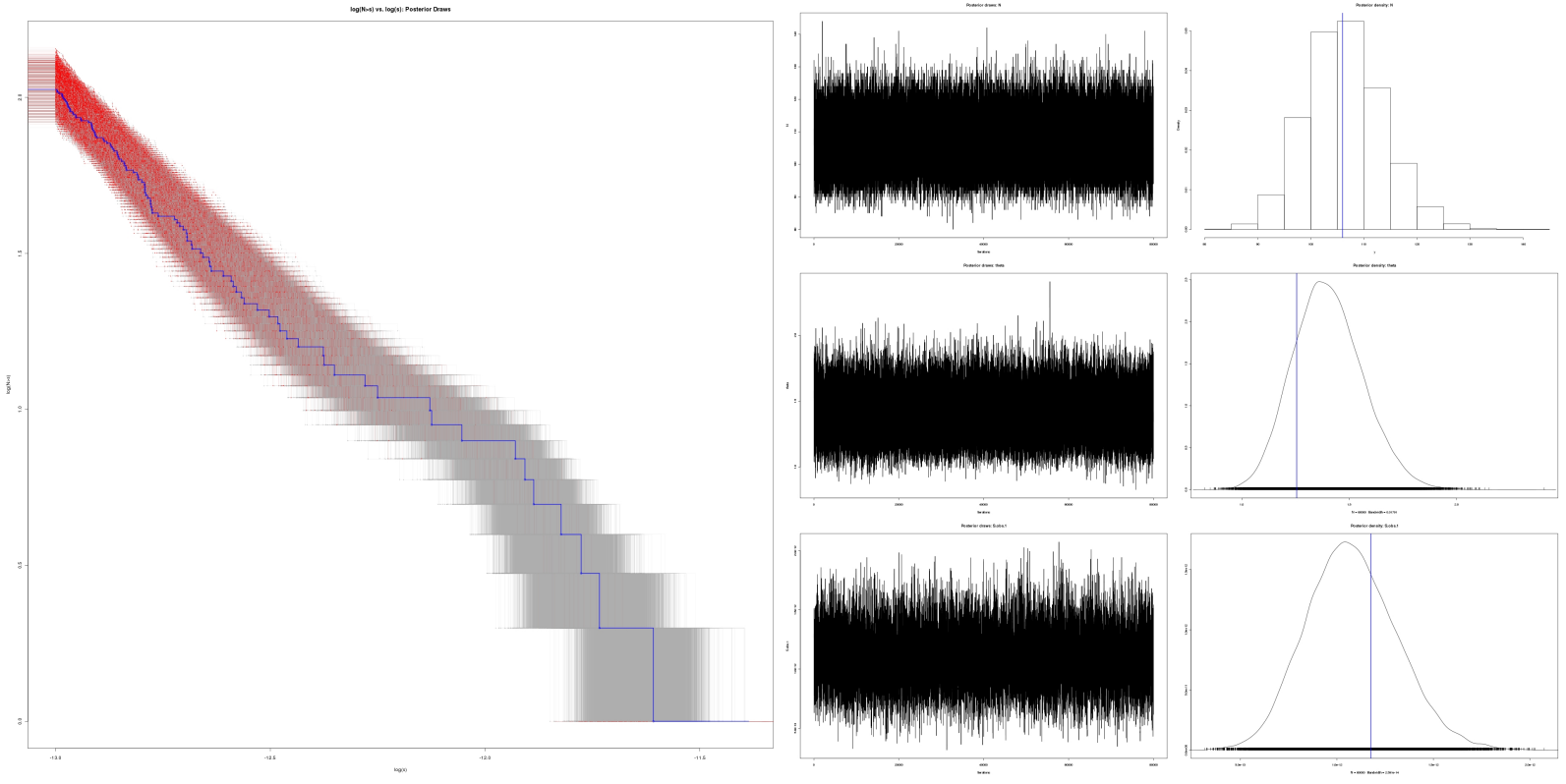
For example, let:

$$T\left(Y^*\right) = \begin{pmatrix} \bar{Y}^* \\ \min\left\{Y^*\right\} \\ \max\left\{Y^*\right\} \end{pmatrix}.$$


Bivariate posterior predictive p-value = 0.033635
IQR
length
0
2
4
6
8
10
12
0
1
2
3
4
5

▸ To Example

# Example MCMC Output




(L) Posterior logN-logS (red: missing, gray: observed), truth (blue).

(R) Posterior distributions for $N$, $\theta$, $S_{obs,i}$

# Validating Bayesian Computation

Given the complexity of the model and computation, it is important to validate that everything works.

Bayesian methods lend themselves to a straightforward self-consistency check:

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 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.

# Validation Details

Parameter specifications as follows:

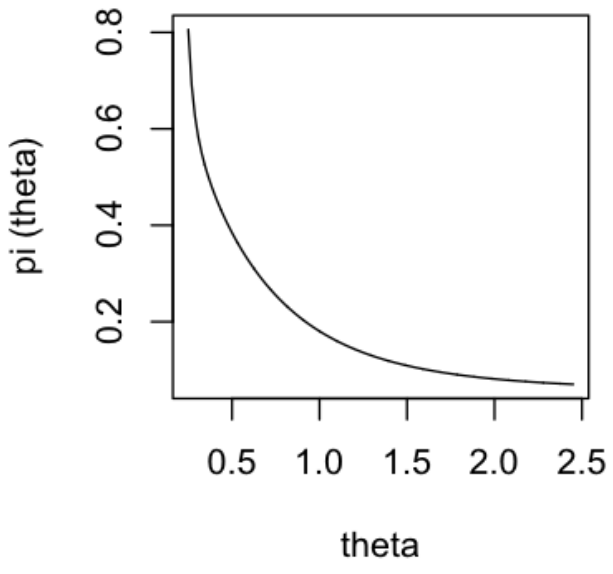
- $N \sim \mathrm{NegBinom}(\alpha, \beta)$, where $\alpha = 200 =$ shape, $\beta = 2 =$ scale
- $\theta \sim \mathrm{Gamma}(a, b)$, where $a = 20 =$ shape, $b = 1/20 =$ scale
- $S_i|\theta \sim Pareto(\theta, S_{min})$, where $S_{min} = 10^{-13}$
- $Y_i^{src}|S_i, L_i, E_i \sim Pois(\lambda(S_i, L_i, E_i))$
- $Y_i^{bkg}|S_i, L_i, E_i \sim Pois(k(L_i, E_i))$
- $\lambda = \frac{S_i \cdot E_i}{\gamma}$, where effective area $E_i \in (1,000, 100,000)$, and the energy per photon $\gamma = 1.6 \times 10^{-9}$
- $k_i = z \cdot E_i$, where the rate of background photon count intensity per million seconds $z = 0.0005$
- $n_{iter} = 250000$, Burnin $= 50000$

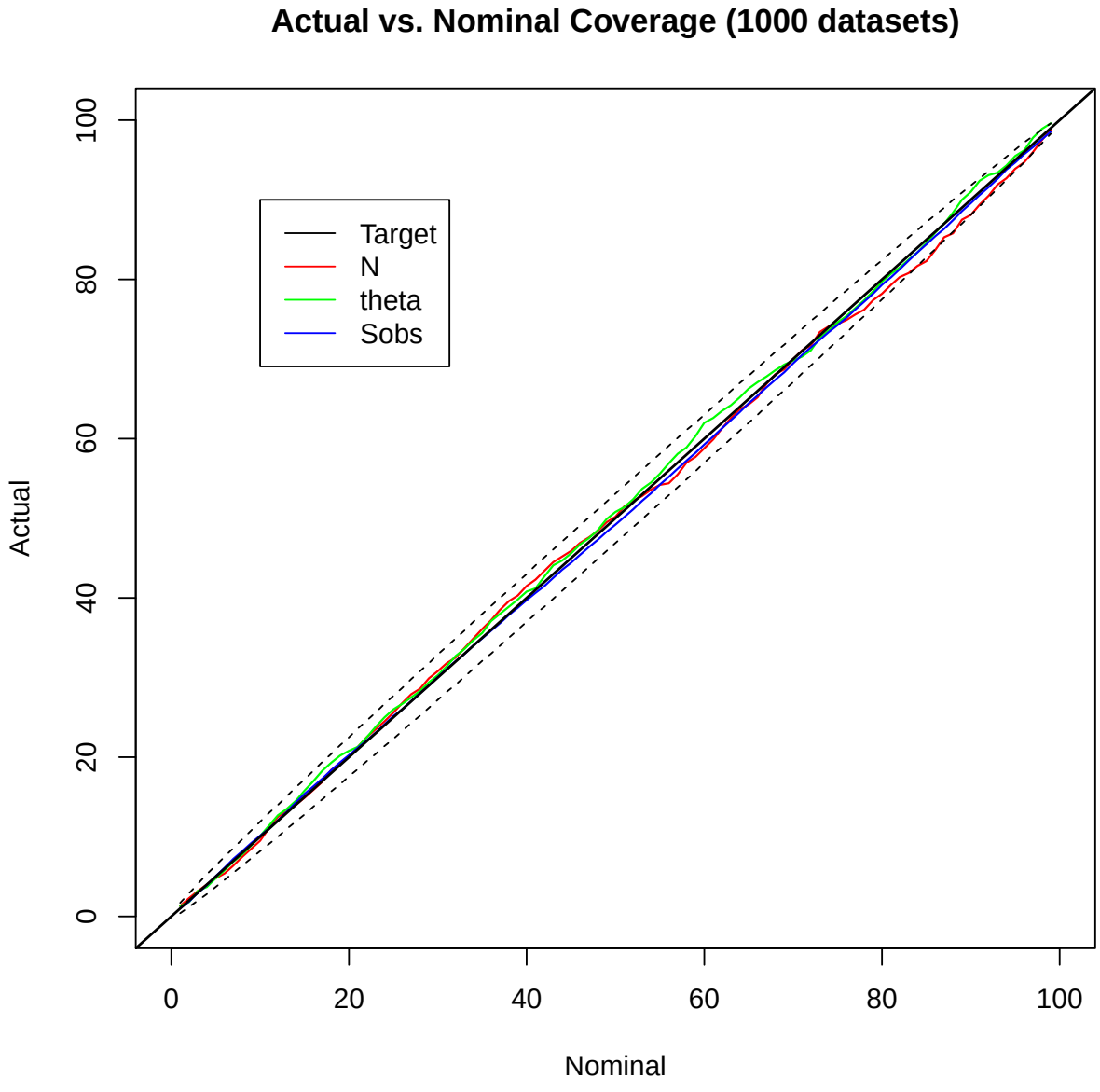
# Simulated Example cont. . .

Detection probability:

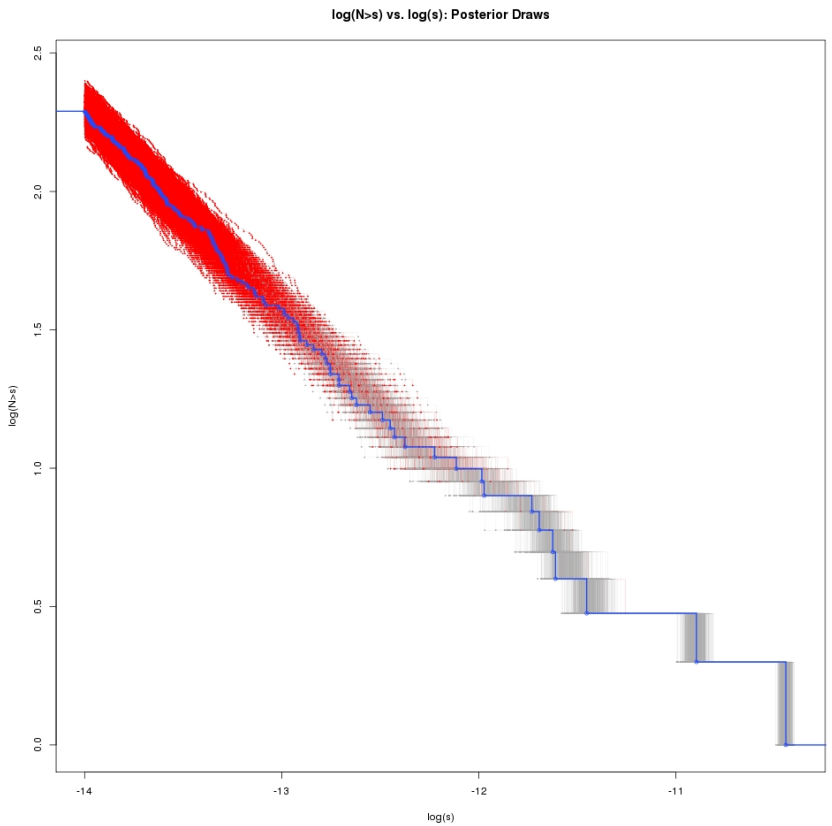
- $g(\lambda, k) = 1.0 - a_0 \cdot (\lambda + k)^{a_1} \cdot e^{a_2 \cdot (\lambda + k)}$, where $a_0 = 11.12, a_1 = -0.83, a_2 = -0.43$

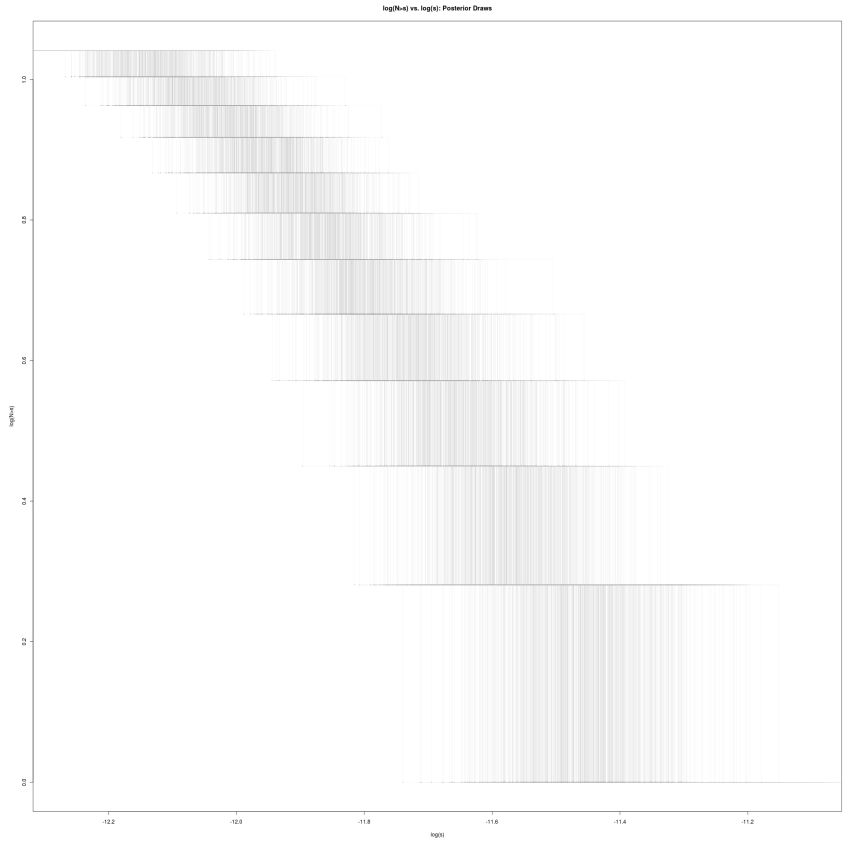
Marginal detection probability:




Actual vs. Nominal Coverage (1000 datasets)
Target
N
theta
Sobs
Actual
Nominal
0
20
40
60
80
100

# Non-Ignorable Missingness




**(L) Nonignorable (full) analysis:** $\hat{\theta} = 1.113,\ (0.957, 1.286)$

**(R) Ignorable analysis:** $\hat{\theta} = 0.465,\ (0.391, 0.586)$

**Truth:** $\theta = 1.070$

# Example

Application: X-ray pulsars in the Small Magellanic Cloud (SMC).

- The SMC is the second nearest galaxy ( 200k light years away)
- It has 1/100 th of the mass of our Galaxy
- It is interacting with our Galaxy and the LMC.
- This interaction triggered the formation of a large population of stars about 40 million years ago.
- This population of stars gives the X-ray pulsars we observe in X-ray observations.
- Close proximity allow the study of low luminosity sources that are hard to observe in other galaxies.

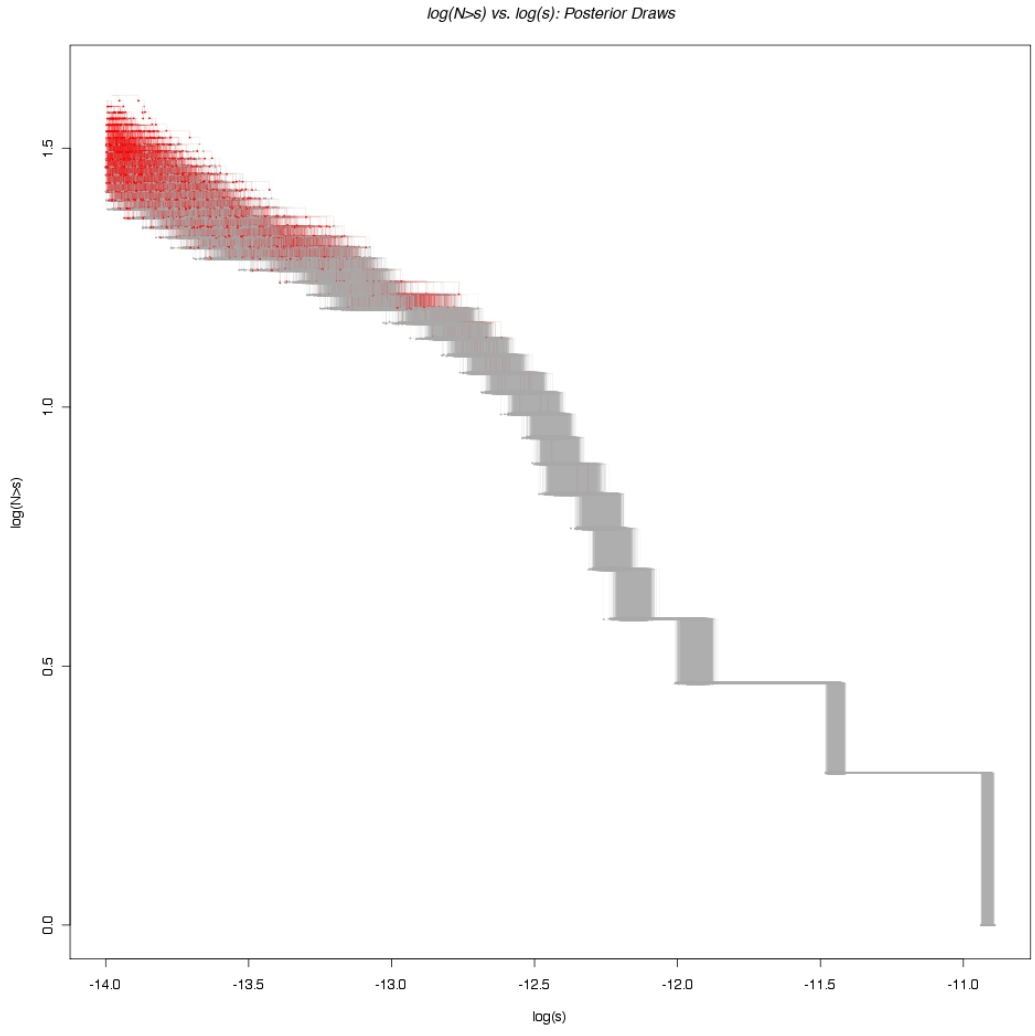
---

Dataset consists of 26 observed sources.

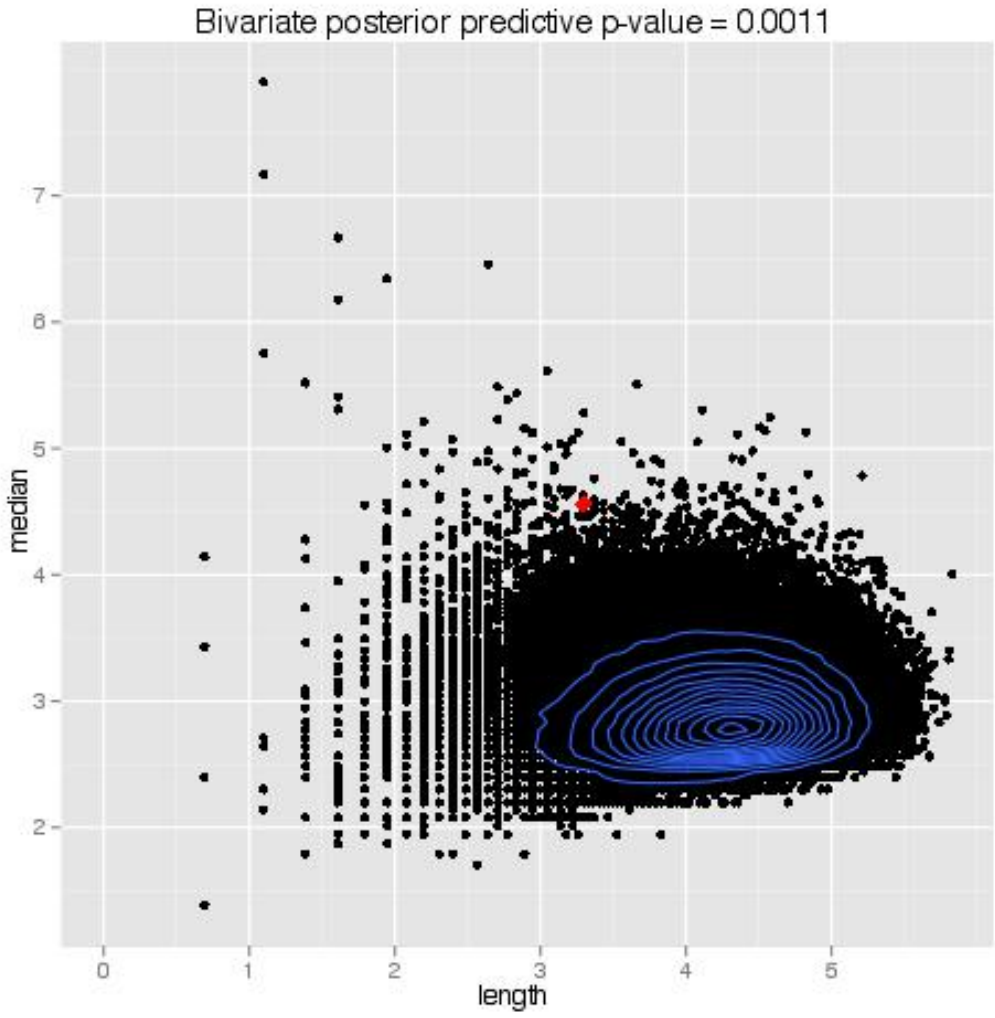
Priors: $a = b = 4$, $\mathbb{E}[N] = 100$, $\mathrm{Var}(N) = 50^2$.

# SMC Gamma Ray Pulsars

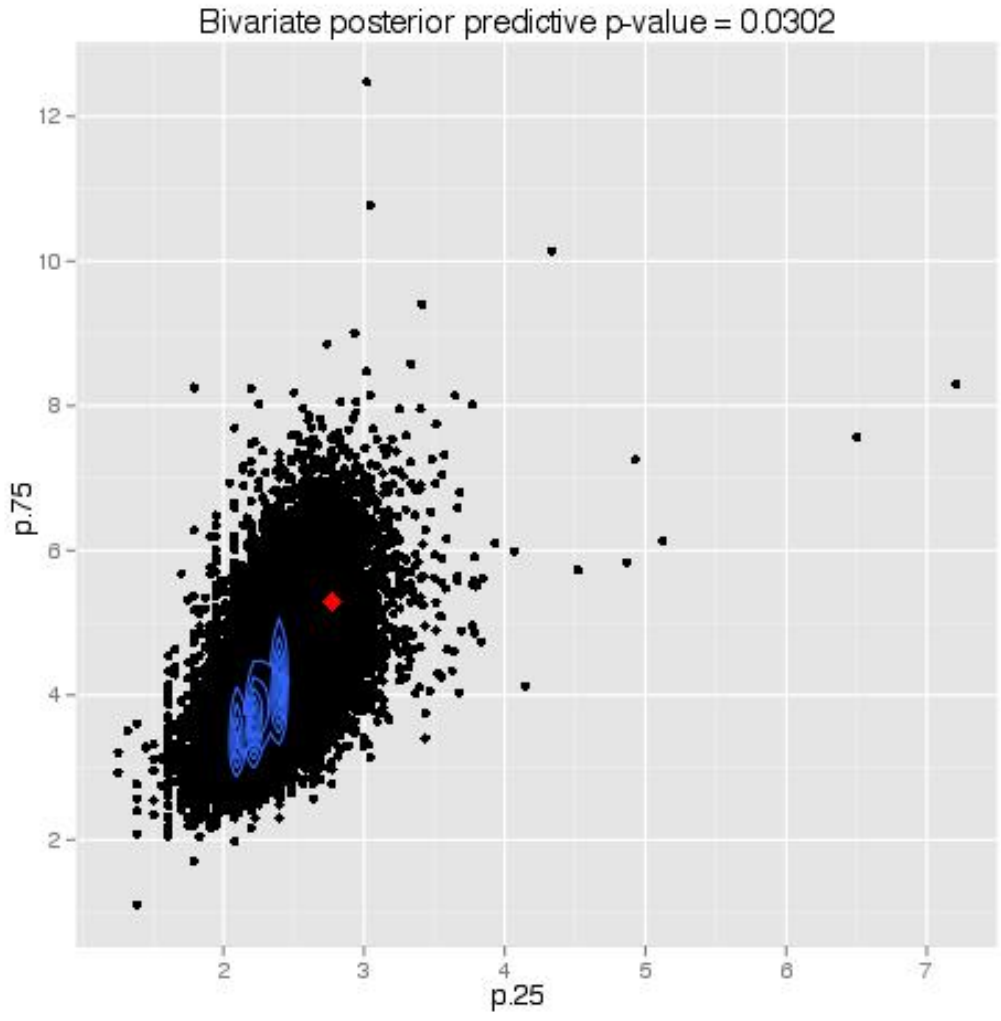
We will now take a look at a few different model fits and discuss model selection, parameter interpretation for the SMC example...



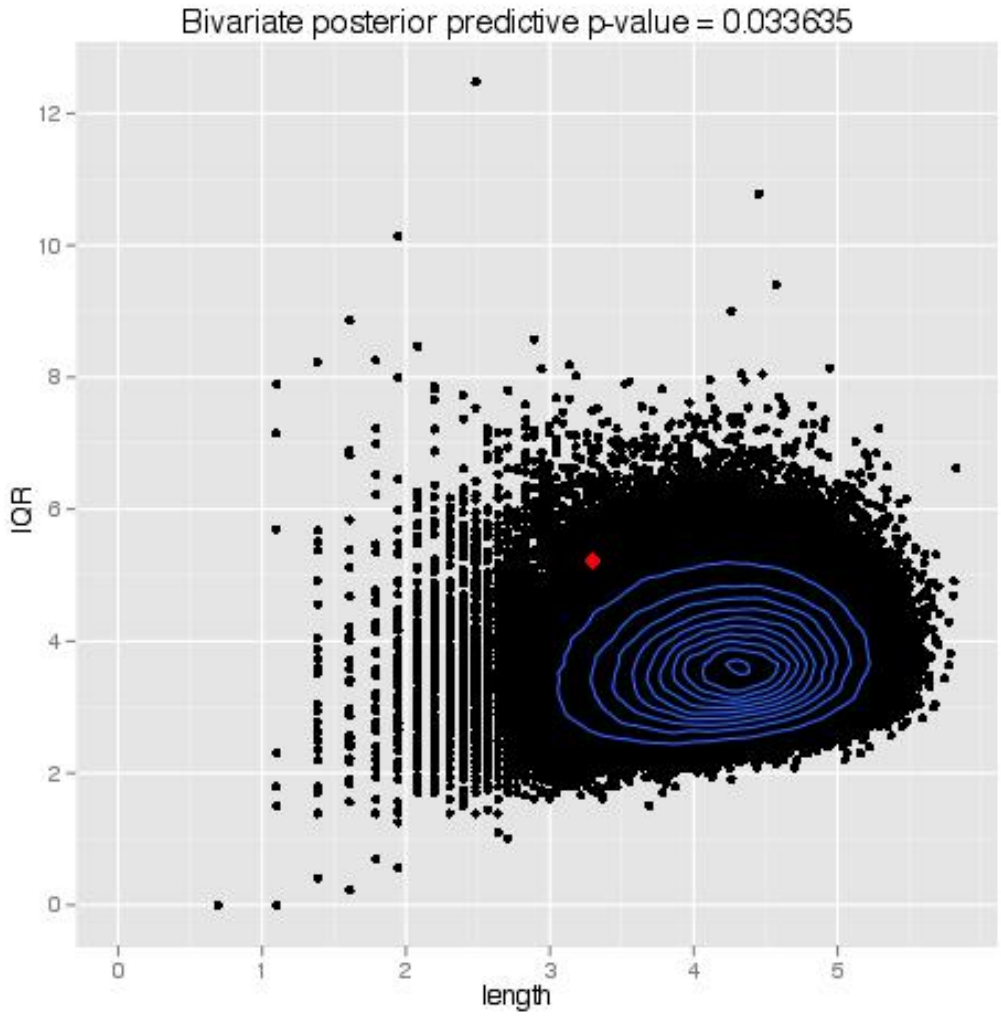

# Single Pareto log(N)-log(S)




**Single Pareto**




**Single Pareto**




**Single Pareto**

# SMC Pulsar Conclusions

- Estimated slope of 0.485 (median), 0.487 (mean)
- 95% (central) interval $= (0.312, 0.681)$
- Zezas et al. (2003) estimated a power-law slope of $\hat{\theta} = 0.45$.
- Estimated missing data fraction 7.2%, interval:$(0.0, 18.25\%)$
- 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 $(0.248, 0.511)$.

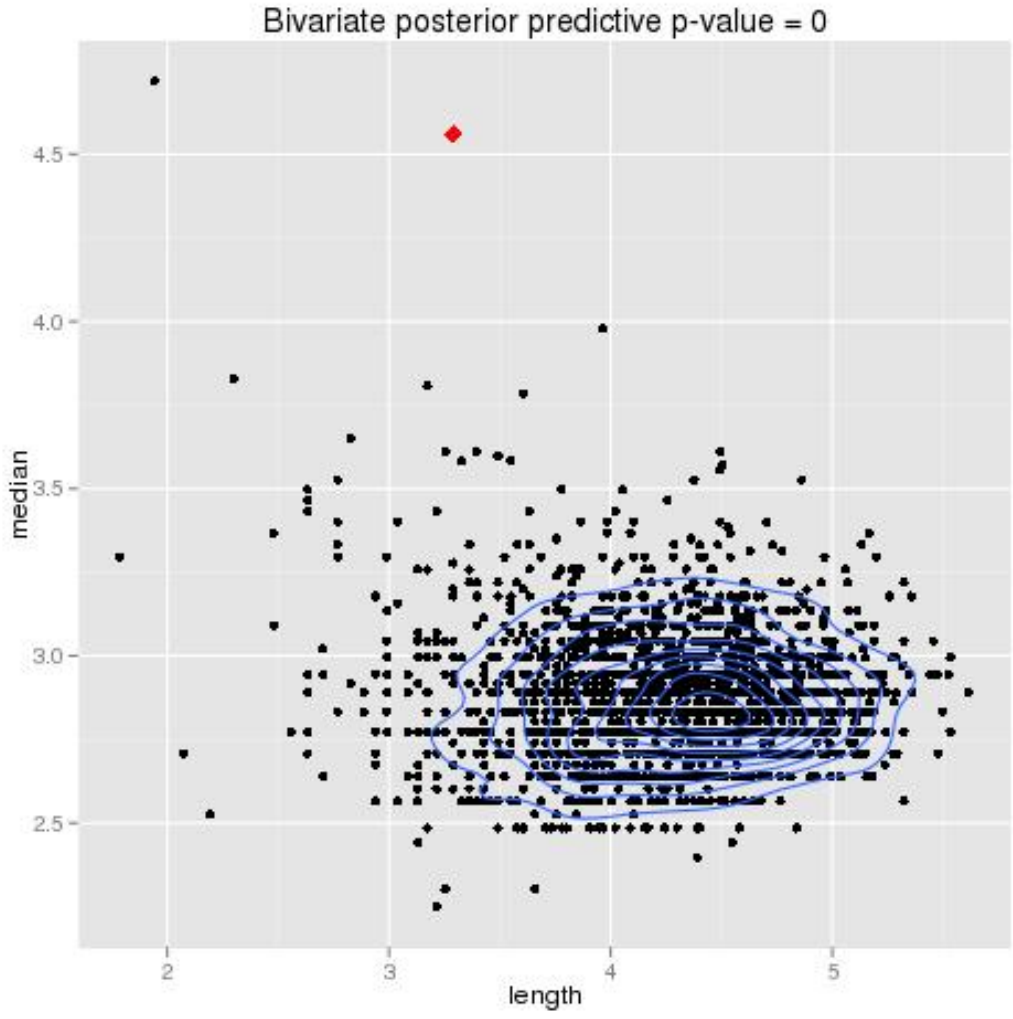
# SMC Pulsar Mixture-Pareto

We now take a look at a two component mixture-Pareto fit to the data, with fixed mixture point.

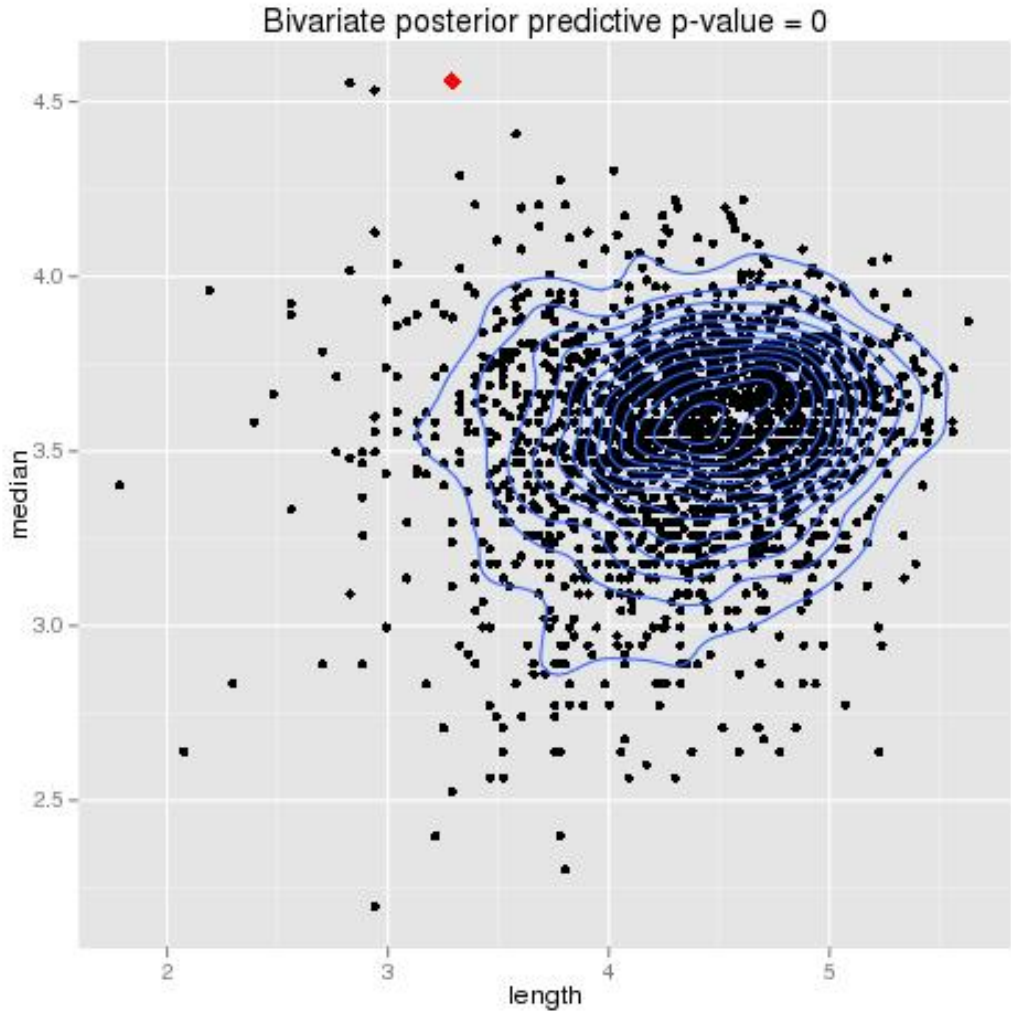
Fit the following possible mixture (break) points:

```
> round(log(possible.bps,base=10),2)
 [1] -13.70 -13.52 -13.40 -13.30 -13.22 -13.15 -13.10 -13.05 -13.00
[10] -12.96 -12.92 -12.89 -12.85 -12.82 -12.80 -12.77 -12.74 -12.72
[19] -12.70 -12.68
```

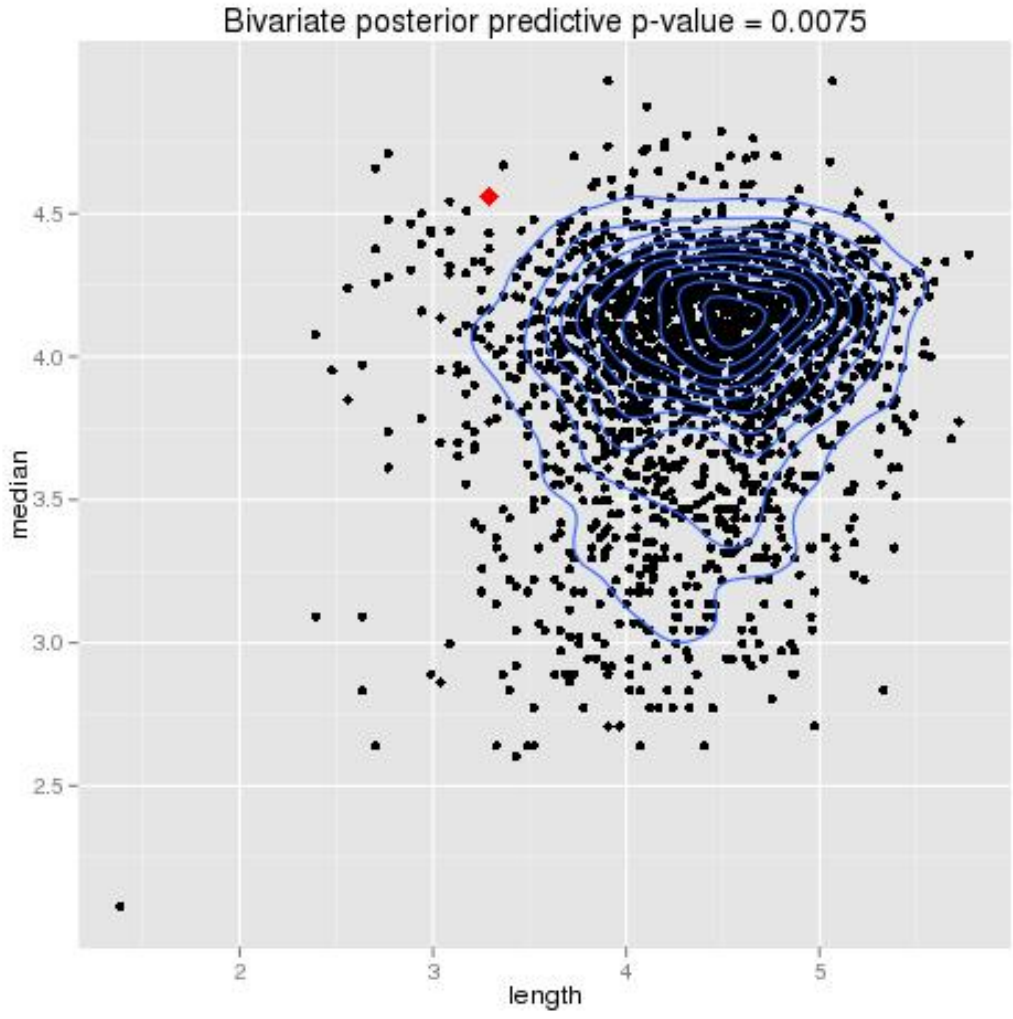
Take a look at diagnostics...



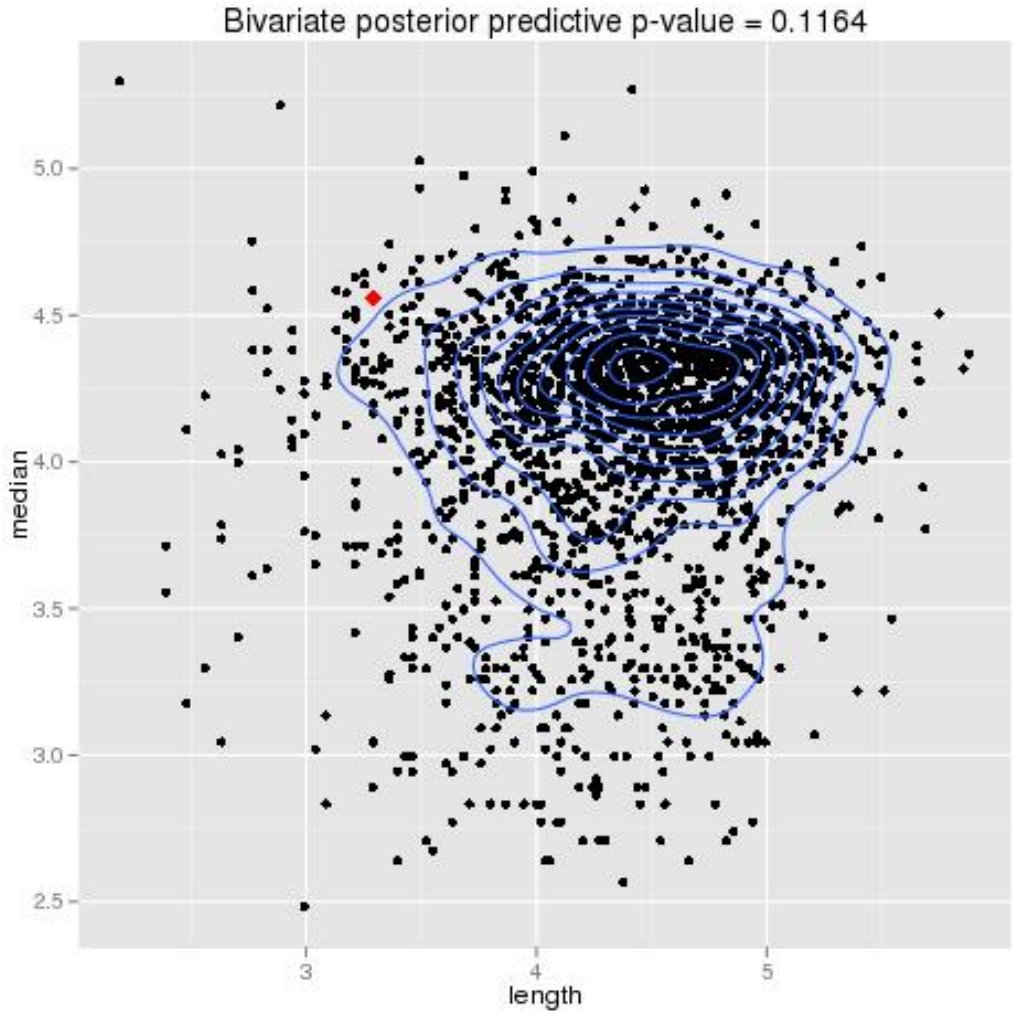

**Break-Point: -13.70**



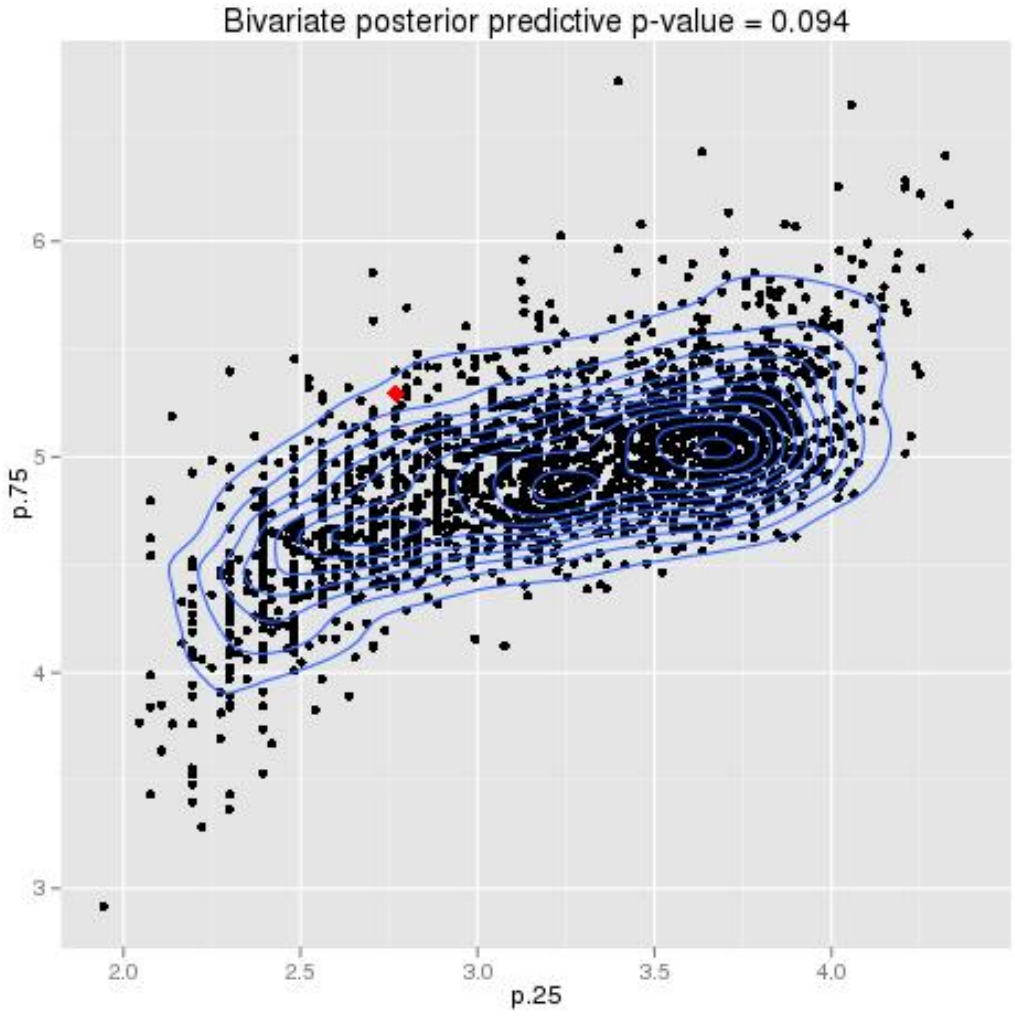

**Break-Point: -13.22**



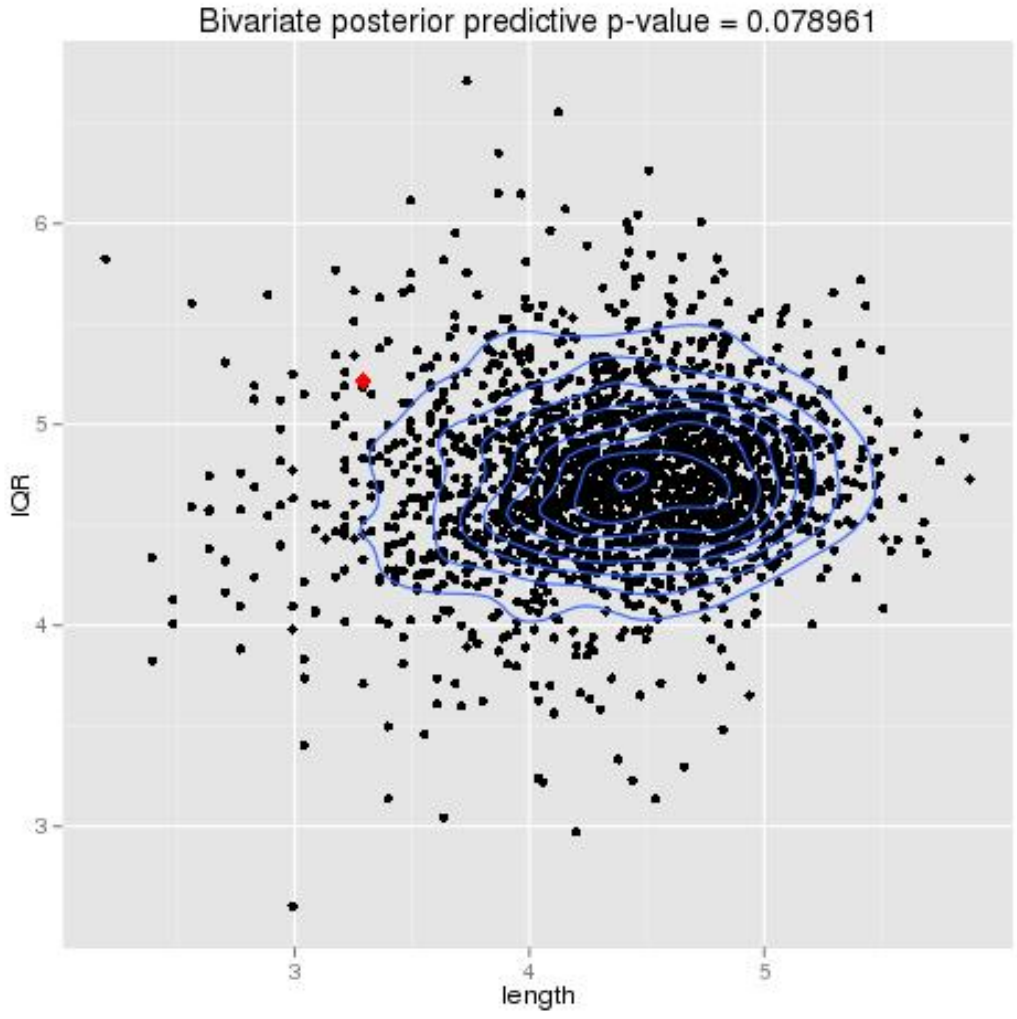

**Break-Point: -12.96**




**Break-Point: -12.89**

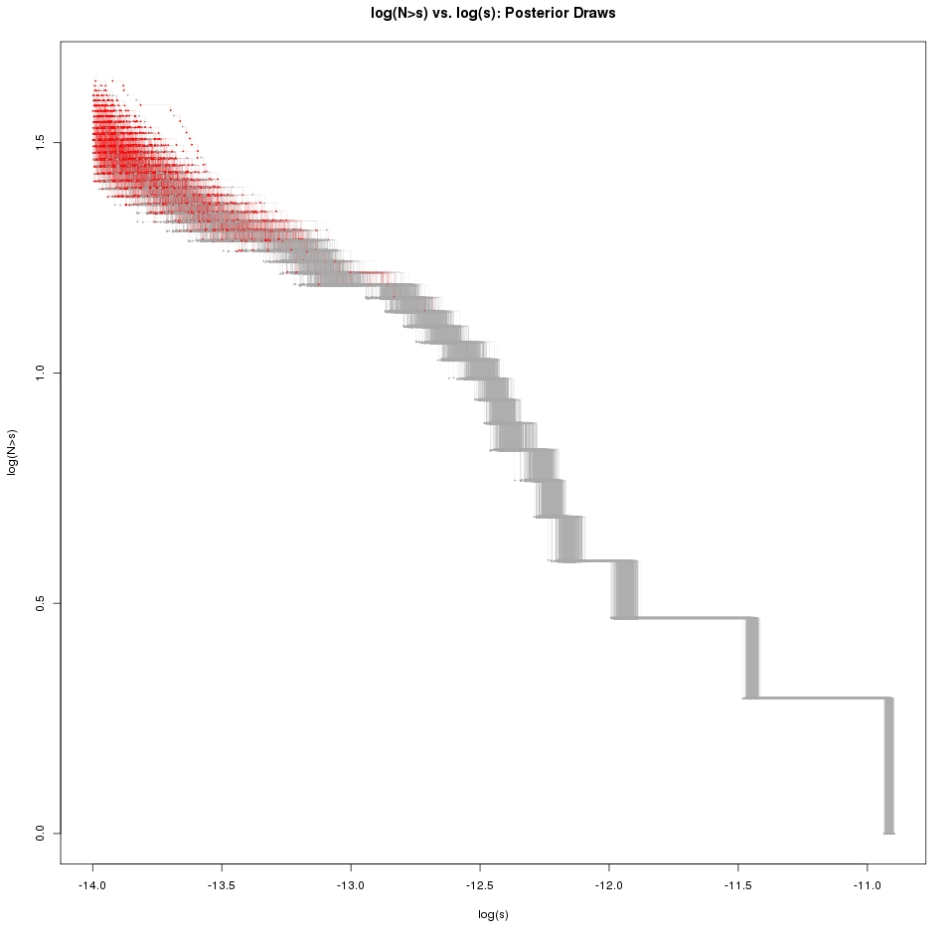



**Break-Point: -12.89**




**Break-Point: -12.89**

# SMC Pulsars: Mixture Power-Law Fit




Estimates: $\hat{\theta}_1 = 0.591(0.412, 0.809)$, $\hat{\theta}_2 = 0.853(0.661, 1.276)$.

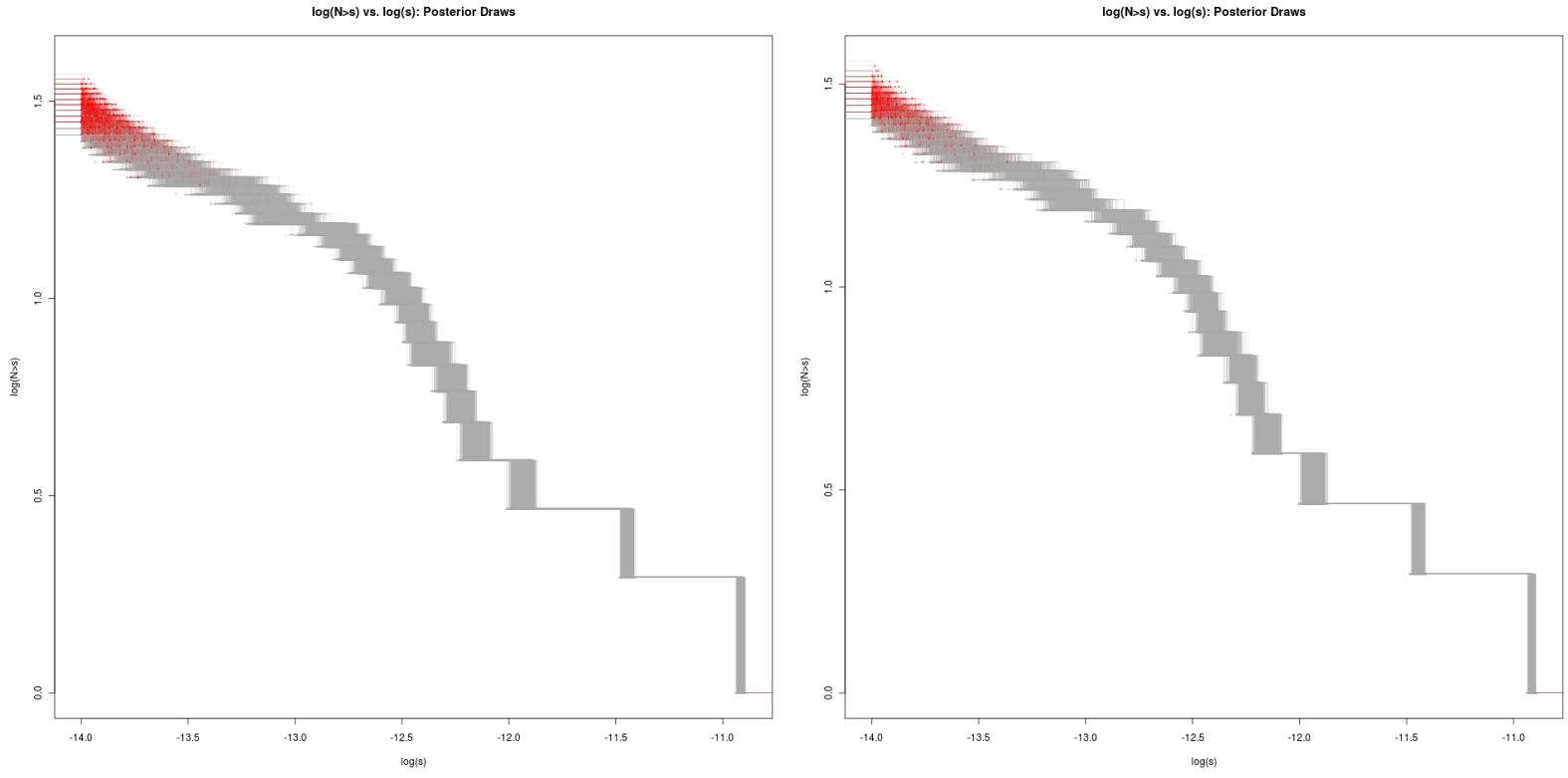
$\hat{p}_1 = 0.69(0.42, 0.93)$

# SMC Pulsars: Mixture Power-Law

Parameter interpretation in the broken and mixture power-law setting is similar but slightly more tricky.

Note that $\theta_2$ is larger than $\theta_1$, indicating the accelerated downward slope after the break-point.

By fitting multiple models it becomes possible to test for the existence of break-points, compare mixture- to broken-power laws etc.




# (L) Regular Pareto, (R) Broken Pareto

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

# Conclusions

1. Probabilistic insight allows us to build statistical procedures that correspond to more physically realistic models
2. Hierarchical modeling allows for us to account for multiple types of uncertainties
3. Importance of handling the non-ignorable missing data mechanism
4. Provides a recipe for assessing goodness-of-fit (posterior predictive checks)
5. Provides a way to include prior information
6. Allows for bias-corrected inference (Eddington, Incompleteness)
7. Flexible framework for computation (e.g., distributional assumptions for fluxes)

# Future Work

1. False sources (allowing that 'observed' sources might actually be background/artificial)
2. Field contamination (allowing a mixture of a source population with known parameters)
3. Break-point estimation for multiple power-law setting
4. Formal testing for presence/number of break-points
5. Extension to non-Poisson regimes

# 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 (2011) Hierarchical Modeling for $\log N - \log S$: Theory, Methods and Applications *Working Paper*
- R.J.A. Little, D.B. Rubin. (2002) *Statistical analysis with missing data*, Wiley.
- 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.