Bayesian Estimate of logN-logS (BENS)
Self-consistent Modeling of the logN-logS in the Poisson Limit
Nondas Sourlas (sourlas@bu.edu, Boston University, CHQ)ER),
Vinay L. Kashyap and Andreas Zezas (Smithsonian Astrophysical Observatory),
David A. van Dyk (Department of Statistics, University of California at Irvine).

Summary

Source number counts as a function of flux (logN-logS curves), are a fundamental tool in the study of source populations. Here we present a new, powerful method using the full Poisson machinery allowing us to model the logN-logS distribution of X-ray sources in a self-consistent manner. Because we properly account for all the statistical effects and sources of bias, we can exploit the full range of the data. We use a Bayesian approach, modeling the fluxes with functional forms such as simple or broken power-laws. The photon counts are modeled conditioned on the source fluxes, the background contamination, detector sensitivity and detection probability. The built-in flexibility of the algorithm also allows simultaneous analysis of multiple datasets. We demonstrate the power of our algorithm by applying it to a set of Chandra observations. We find that the source detected in the Chandra Deep Field (North) ObsID 2232 follows a broken power law distribution with indices -0.9 and -1.52.

1. Statistical Biases

1.1 Lost sources (false negatives)

Because of statistical fluctuations of observed source intensities, a preferentially larger number of faint sources are lost below the detection threshold resulting in a true turning of the logN versus logS.

1.2 Eddington bias

For sources with intensities near the detection threshold, there is a tendency for the measured flux to be higher than the true flux because statistical fluctuations below the detection threshold will be averaged out.

1.3 False source fluctuations

When there are larger numbers of faint sources than bright sources, then statistical fluctuations result in a larger number of the fainter sources detected into higher flux regions, causing steepening of the inferred slope.

2. Method

2.1 A Model for the logN-logS

1. Slope:

The cumulative distribution of S, F(S), is modeled using a Power law model

\[ F(S) = a_S \cdot S^{-\gamma_S} \quad S \geq S_{\min} \]  

or in some cases a Broken Power law

\[ F(S) = a_S \cdot S^{-\gamma_S}\quad a_S S^{-\gamma_S}\quad a_S S^{-\gamma_S}\quad a_S S^{-\gamma_S} \quad S_{\min} \leq S < S_{\max} \]

Here, the break point \( S_0 \) is fixed and the minimum flux, \( S_{\min} \), determined by the faintest source.

2. Normalization

The total number of objects \( N \) consists of the number of detected objects \( N_d \) and undetected objects \( N_u \), and is assumed to follow a Poisson distribution

\[ N_d \sim P(\delta) \quad \delta > 0, \]

\[ N_u \sim P(1 - \Pi), \]

where \( \delta \), the Poisson parameter, is assumed to follow a noninformative Gamma distribution and \( \Pi \) is the marginal probability of observing a source.

\[ \Pi = \int P(\delta) P(S_d | \delta) d\delta \]

2.2 Procedure

With starting values, \( \alpha, \beta, \gamma, \) and fixed break point, \( S^* \), we:

1. Calculate the probability that a source is not detected (1 - \( \Pi \)) as described in (5)

2. Sample \( N_{\text{det}} \) conditioned on \( N_d \) and \( N_u \),

\[ N_{\text{det}} \sim Y^\nu \quad \alpha, \beta, \gamma, \nu \sim \text{NegBin}(\nu) \]

3. Account for background by modeling \( Y^\nu \sim P(X(S_d) + B) \)

4. Use the MLE obtained to obtain maximum likelihood estimates for \( \alpha, \beta, \gamma \)

5. Use the MLE obtained to obtain maximum likelihood estimates for \( \alpha, \beta, \gamma \)

6. Repeat steps 1 - 4, to simulate the posterior distribution.

This is accomplished using MCMC and data augmentation algorithms. Results are based on the cumulative counts from three Markov chains, of size 100 each, after discarding the first 10 draws, for a total of 105 draws. Posterior inference is based on these 105 simulated draws.

3. Results

<table>
<thead>
<tr>
<th>ObsID</th>
<th>Model (0.6-2 keV)</th>
<th>Best (0.6-2 keV)</th>
<th>Best (1.2-2 keV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2232</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPL</td>
<td>1.28</td>
<td>1.18</td>
<td>1.28</td>
</tr>
<tr>
<td>BPL</td>
<td>-0.86</td>
<td>-1.25</td>
<td>-0.86</td>
</tr>
<tr>
<td>2314</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPL</td>
<td>1.18</td>
<td>1.11</td>
<td>1.36</td>
</tr>
<tr>
<td>BPL</td>
<td>-0.86</td>
<td>-1.25</td>
<td>-0.86</td>
</tr>
</tbody>
</table>

Table 2: Posterior Means obtained for BENS compared to Maximum Likelihood Estimates according to Crawford et al. (2010), for the slope parameters of the logN-logS. The MLE value is for the slope to the right of the brace point. For the broken power law (BPL) \( \beta \) is the slope to the left of the brace point and \( \gamma \) to the right.

Data

Chandra Deep Field: ObsID 2232
Exposure Map: ObsID 2232
The data consist of:
- the measured counts and coordinates of each detected source,
- background estimate over the entire detector,
- effective area, exposure time, and vignetting,
- tables of detection probabilities \( p(S_d, B|L_d) \), for a set of intrinsic flux \( S_d \) at location \( L_d \) and a local background \( B|L_d \).