Bayesian Estimate of logN-logS (BENS)

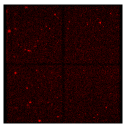
Self-consistent Modeling of the logN-logS in the Poisson Limit

Nondas Sourlas (sourlas@bu.edu, Boston University, CHQOER),
Vinay L. Kashyap and Andreas Zezas (Smithsonian Astrophysical Observatory),
David A. van Dyk (Department of Statistics, University of California at Irvine).
CHASC (California-Harvard AstroStatistics Collaboration) www.ics.uci.edu/~dvd/astrostat.html

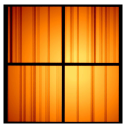
Summary

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 demonstate the power of our algorithm by applying it to a set of Chandra observations. We find that the sources detected in the Chandra Deep Field (North) ObsID 2234 follow a broken power law ditribution with indices -0.9 and -1.52..

Data



Chandra Deep Field: ObsID 2232



Exposure Map: ObsID 2232

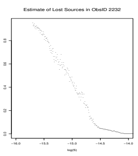
The Data consist of:

- 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 $\pi(S_j, L_j, B(L_j))$, for a source of intrinsic flux S_j, at location L_j, and a local background $B(L_j)$.

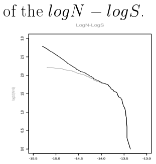
1. Statistical Biases

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 an artificial turnover of the $logN - logS$.



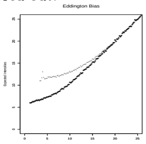
Lost Sources. We show the probability of not detecting a source in Chandra ObsID 2232 as a function of log(Flux); the fainter the source the more likely it will remain undetected.



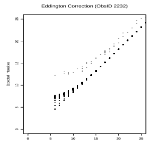
Effect on LogN-LogS: Estimate missed sources, part of the logN-logS solution, accounting for detection probabilities. Here, lower curve reflects only observed sources; upper includes the missing ones.

1.2 Eddington bias

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



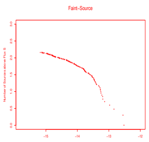
Eddington Bias: This simulation, which shows the true vs. the predicted source intensities, illustrates the bias for two different thresholds. The lower curve using a threshold of 5, and the upper one a threshold of 10.



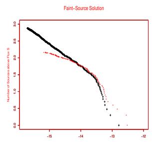
Bias reduction of our algorithm. The lower dots are sets of expected counts (drawn at each iteration in analyzing ObsID 2232) ploted against observed counts. The grey, upper dots, show the bias in absence of the BENS correction.

1.3 Faint 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 deflected into higher flux regimes, causing steepening of the inferred slope.



Part of the bias in this logN-logS of ObsID 2232 observed data, is attributable to the curvature we see in higher fluxes. Note that these fluxes are observed w/ probab. 1 and but overestimated.



Ploted in red are the fluxes of the observed sources at the first iteration of our algorithm, and in black is the last iteration including missed sources. Note the red line above the black at high fluxes.

2. Method

- Generate a simulated sample of source intensities over the observed field of view from an adopted model of the $logN(> S)$ distribution.
- Account for observational biases by folding the source numbers through the detection probabilities.
- The simulated source intensities, locations, and the observed background at these locations are used to determine the numbers and distributions of missed sources.
- The model parameters are updated by comparing the simulated dataset modified as above, with the observed sources.

2.1 A Model for the logN-logS

1. Slope:

The cummulative distribution of S, $F(> S)$, is modeled using a *Power law* model

$$F(> S) = \alpha_{pl} * S^{-\beta_{\rm PL}}; \qquad S \geq S_{min}, \tag{1}$$

or in some cases a *Broken power law*

$$F(> S) = \begin{cases} \alpha_{pl} S^{-\beta_{pl}} & if \quad S \geq S_\star \\ \alpha_{pl} S^{-\beta_{pl} - \gamma_{pl}} S_\star^{\gamma_{pl}} & if \quad S_{min} \leq S < S_\star \end{cases} \tag{2}$$

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

2. Normalization:

The total number of objects N consists of the number of detected objects N_{obs} and undetected objects N_{mis}, and is assumed to follow a Poisson distribution,

$$N|\delta \sim Po(\delta) \qquad \delta > 0, \tag{3}$$

$$N_{mis} \sim Po(\delta[1 - \Pi]), \tag{4}$$

where δ, the Poisson parameter, is assumed to follow a noninformative Gamma distribution and Π is the marginal probability of observing a source,

$$1 - \Pi = \int P(S_j|\alpha, \beta) \int\int (1 - \pi(S_j, L_j, B(L_j))) P(L_j) dL_j dB(L_j) dS_j. \tag{5}$$

2.2 Procedure

With starting values, $\alpha_0, \beta_0, \gamma_0$, and fixed break point, $S^\star$, we

1. Calculate the probability that a source is not detected $(1 - \Pi)$ as described in (5)
2. Sample N^{mis} conditional on N^{obs}, and Π,
$$N^{mis}|N^{obs}, a, b, \pi \sim NegBinom\left(N^{obs} + a, \frac{b + \pi}{1 - \pi}\right) \qquad a = b = 0. \tag{6}$$
where a, b are the parameters of the uninformative gamma prior
3. Account for background by modeling $Y_j^{obs} \sim Po(\lambda(S_j) + \lambda_j^B)$, and sampling the source count
$$Y_j^{src}|Y_j^{obs}, S_j, B_j \sim Binom\left(Y_j^{obs}, \frac{\lambda(S_j)}{\lambda(S_j) + \lambda(B_j)}\right), \tag{7}$$
where $\lambda(S_j)$ the expected source count for a source with flux S_j, and $\lambda(B_j)$, the expected background count adjusted for effective area.

   We use a counts-to-energy conversion factor of $1.168 * 10^8 \ \ ph/erg$, $0.1 - 10 \ \ keV$ based on a power-law spectrum with $\Gamma = 1.7$ and $N_H = 180cm^{-2}$.
4. Sample the fluxes S_j for observed (a) and unobserved (b) sources
   (a) $i = 1, \ldots, N^{obs}$ from $p(S_j|M_j = 0, Y_j^{src}, L_j, B_j, E_j, \alpha, \beta, \gamma, N)$
   (b) $i = 1, \ldots, N^{mis}$, from $p(S_j|M_j = 1, L_j, B_j, E_j, \alpha, \beta, \gamma, N)$
5. Sample $\alpha, \beta, \gamma|S_1, \ldots, S_N$
   (a) Fit a log-linear model to obtain maximum likelihood estimates for (α, β, γ).
   (b) Use the MLE obtained and the unscaled covariance matrix to be the parameters of a t_4 distribution and sample from the posterior of (α, β, γ).
6. Repeat steps 1. - 4. to simulate the posterior distributions. This is accomplished using MCMC and data augmentation algorithms. Results are based on the cumulative draws from three Markov chains, of size 100 each, after discarding the first 50 draws, for a total of 150 draws. Posterior inference is based on these 150 simulated values.

3. Results

ObsID	Model	Soft (0.3-2 keV)			Broad (0.3-8 keV)		
		γ_{MLE}	β_{BENS}	γ_{BENS}	γ_{MLE}	β_{BENS}	γ_{BENS}
2232	SPL	-1.26 (0.22)	-1.24 (0.05)		-1.24 (0.22)	-1.23 (0.045)	
	BPL		-0.9 (0.03)	-1.25 (1.1)		-0.826 (0.035)	-1.20 (0.13)
2234	SPL	-1.44 (0.26)	-1.11 (0.04)		-1.45 (0.27)	-1.12 (0.03)	
	BPL		-0.892 (0.027)	-1.52 (0.15)		-0.832 (0.03)	-1.55 (0.15)

Table 2: Posterior Means obtained for BENS compared to Maximum Likelihood Estimates according to Crawford et.al (1970), for the slope parameters of the logN-logS. The MLE value is for the slope to the right of the brake point. For the Broken Power law (BPL) β is the slope to the left of the brake point and γ to the right.