The Astrophysical Journal, 584:1016–1020, 2003 February 20

A NEW TECHNIQUE FOR DETERMINING THE NUMBER OF X-RAY SOURCES PER FLUX DENSITY INTERVAL

A. T. Kenter and S. S. Murray
Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138
Received 2002 August 5; accepted 2002 October 28

ABSTRACT

We present a technique for determining the number of X-ray sources per flux density interval, the log N–log S relationship, that is mathematically analogous to spectral fitting. This technique is ideal for X-ray source counts obtained using the *Chandra X-ray Observatory* since the telescope and the focal plane instruments have been well modeled. This technique is of general applicability. In this paper, we apply it to a wavelet source-detect analysis of a *Chandra* Advanced CCD Imaging Spectrometer (ACIS-I) mosaic image of a 1.35 deg^2 survey of the Lockman Hole. We verify the technique via Monte Carlo simulations.

Subject heading: methods: miscellaneous — surveys — X-rays: galaxies

1. INTRODUCTION

Detecting the number of sources per flux density interval, $N(S)$, is a typical problem in astronomy. If the sources are all of the same class and are located in the same distant galaxy or cluster, $N(S)$ can be used to study the source-class luminosity function. For extragalactic X-ray sources at different distances, the relationship between N and S is typically well approximated by a single or composite power-law distribution (Hasinger et al. 1993). The normalization and index of the power-law distribution are of great interest as they have cosmological implications; for sources uniformly distributed in a Euclidian nonevolving universe, the $N(S)$ relationship is a power law of index 5/2. The power-law nature of $N(S)$ or of its integral form $N(>S)$ is often presented in the form of a log N versus log S plot.

A typical recipe for determining an X-ray $N(S)$ is to detect the sources, determine the X-ray counts for each source, and correct for effective area and exposure time. The X-ray counts for each source are then converted to physical flux units (e.g., ergs cm^{-2} s^{-1}) based on a spectral model. The measured $N(S)$ is then the number of sources at a specific flux density, and in a specific flux density interval. The sky coverage of the survey is then incorporated to give $N(S)$ per unit solid angle. There are a number of biases and selection effects present when determining $N(S)$, several of which are listed below:

1. *Sensitivity*.—Source detection algorithms detect higher flux density sources with a greater probability than lower flux density sources.

2. *"Poisson migration" or Eddington bias* (Eddington 1940).—Sources with a flux density near the detection threshold will be preferentially detected when they have upward fluctuations. The net consequence of this is that the flux density of these sources is overestimated.

3. *Source confusion*.—Sources with angular separation smaller than the telescope point spread function (PSF) can be mistakenly identified as a single source. Similarly, the extended wings of the PSF of a bright source may contribute to the counts of adjacent sources.

4. *Background events*.—Fluctuations in the background may add (or subtract) counts from the source.

These and other biases and selection effects, as well as various techniques for correcting for them, have been discussed elsewhere (Hasinger et al. 1993; Vikhlinin et al. 1995; Murdoch, Crawford, & Jauncey 1973; Schmitt & Maccacaro 1986; Moretti et al. 2002).

Techniques for accommodating these biases can be analytical or numerical and are often investigated with extensive Monte Carlo simulations. The simplest techniques involve limiting source selection to those brightest sources whose count uncertainty is significantly less than the bias level. One can further correct for sensitivity by dividing the differential $N(S)$ by the probability $P(S)$ of detecting a source of flux density S. In addition to statistical biases, implementations of source detection algorithms such as the sliding cell or wavelet deconvolution may have their own inherent selection effects or nonlinearities.

In this paper, we present a technique for determining $N(S)$ that is mathematically equivalent to spectral fitting. As will be discussed, many of the issues and solutions of spectral fitting have direct analogs in our method. The technique inherently takes into account statistical biases and instrumental selection effects, and it inherently calibrates the photometry scale for the source detection algorithm. This method does not fully take into account source confusion, in the same way that most spectral fitting algorithms do not fully take into account photon pileup. As will be discussed below, the probability of source confusion where j multiple sources are mistakenly grouped as a single source is proportional to $[(N/\Omega) \times \langle \delta\Omega_{\rm psf} \rangle]^j$, where N/Ω is the number of sources per solid angle and $\langle \delta\Omega_{\rm psf} \rangle$ is the effective solid angle of the point spread function averaged over the observation. Since the *Chandra* point spread function is small, source confusion is negligible as long as $[(N/\Omega) \times \langle \delta\Omega_{\rm psf} \rangle] \ll 1$.

This technique has been developed, in part, for the purpose of analyzing a set of Lockman Hole (LH) data collected as part of the HRC GTO program. Although we have used the LH data set to determine various input parameters to our procedure, the technique, with minor modifications, is of general applicability. Further in-depth discussion and analysis of the LH data set will be presented in future publications.

In the following sections of this paper we describe our technique. We then apply the technique to our LH data set. Finally, we apply the technique to Monte Carlo simulations. For the sake of simplicity, we shall keep flux density units, S, in terms of the equivalent number of counts, n, that would have been detected in the observation time for a given energy spectrum; this sampling of flux density at specific increments is done so that the discussion and results are independent of any specific source spectral model. Keeping flux density in terms of equivalent counts is ideally suited to our LH data, where we have approximately uniform sky coverage over the survey.

2. THE TECHNIQUE

A detected number–flux density relationship, $N(S')$, can be related to the true number–flux density relationship, $N(S)$, by a redistribution matrix function

$$N(S') = \int R(S, S')N(S)dS \, , \tag{1}$$

or by sampling the input flux density appropriately, we can write this in terms of discrete counts:

$$N(m) = \sum R(m, n)N(n) \, . \tag{2}$$

The redistribution matrix, $R(m, n)$, describes how a "true" n-count source through statistical and nonstatistical biases may be detected as an m-count source. It is possible to populate the elements of the $R(m, n)$ matrix by using what is equivalently a series of discrete delta function input distributions of the form $N(n) = \delta_{j,n}$, where $j = \{1, 2, 3, \ldots\}$. In practice, this involves using a Monte Carlo simulation that generates sources with flux densities sampled so as to give approximately integer number of counts if the sources were on axis. With added complexity, observation specific factors such as spatially varying background and real source positions with respect to detector features (such as inter-chip gaps) can be incorporated. Having determined the $R(m, n)$ matrix, a candidate $N(n)$ distribution with free parameters is tested by convolving it with $R(m, n)$. The free parameters are varied, and the $N(n)$ $R(m, n)$ convolution is fitted to the detected $N(m)$ distribution.

The general technique is independent of the source detection algorithm and of the particular telescope simulation, as long as one is consistent throughout the whole process. To populate the $R(m, n)$ matrix we used a Monte Carlo simulation encompassing the MARX (Modeling *AXAF* Response to X-rays) *Chandra* modeling software package. MARX is the presently accepted standard simulation package of the *Chandra* telescope and is discussed in detail elsewhere (Wise et al. 1999). We have generated multiple simulated fields, each containing multiple n-count sources. This process was repeated with $n = \{1, 2, 3, \ldots, 50\}$. To make this appropriate to our LH observations, the source locations were distributed randomly over an ACIS-I field of view that was determined by an actual observation exposure map. Based on the LH data set, appropriate numbers of background events were added to these simulated source fields. These background events were taken from archival ACIS-I background and the nature of these archival background data is discussed elsewhere (Markevitch 2001[1]). Although the nature of the energy spectrum is not important to our technique, the energy spectrum of the simulated point sources was set to a single power law with a photon index of 1.7; this was done so that in the end, we could check and convert from counts to physical flux density units for our LH data. The conversion between on-axis source flux density and ACIS-I counts was checked against the latest models of the telescope and instrument using the PIMMS software package.

The simulated data sets were run through the identical point source detection algorithms and exposure map corrections that were used for the real LH data. For each series of fixed input $N(n) = \delta_{j,n}$ $(j = \{1, 2, 3..\})$ count simulations, our detection process generated a column in the $R(m, n)$ matrix. The normalization of each matrix column was set equal to the fraction of the n-count sources that were detected; this fraction is the probability, $P(n)$, of detecting an n-count source

$$P(n) = \sum_m R(m, n) \, . \tag{3}$$

The quantity $P(n)$, when converted to physical flux density units, is the sensitivity, $P(S)$, and when scaled by the solid angle, is the "completeness"; it is analogous to the ancillary response function (ARF) in spectral fitting. Our determined $P(n)$, as well as the $n = 5$ and $n = 9$ columns of the $R(m, n)$ matrix, are presented in Figures 1 and 2.

The effect of Poisson migration can be examined by looking at the ratio of mean detected counts to input counts. This quantity is inherent in the $R(m, n)$ matrix

$$M(m, n) = \frac{1}{n}\left[\frac{\sum_m mR(m, n)}{\sum_m R(m, n)}\right] \, . \tag{4}$$

[1] http://cxc.harvard.edu/contrib/maxim/bg

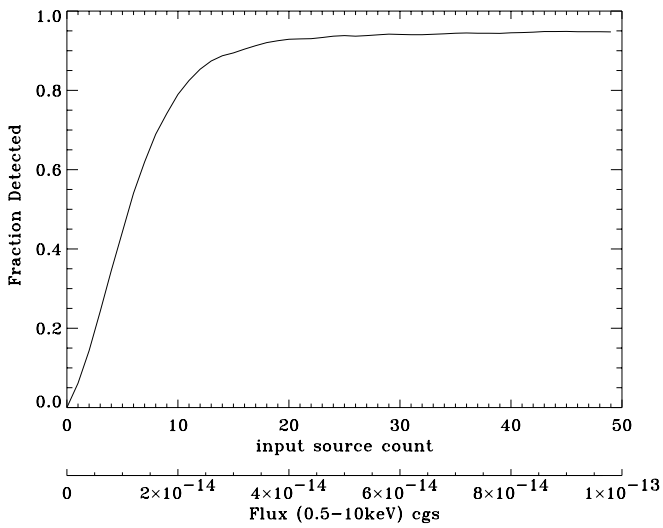

FIG. 1.—Probability of detecting an n-count source is analogous to the ancillary response function (ARF) in spectral fitting. It is embedded in the $R(m, n)$ matrix as the normalization of each column, $P(n) = \sum_m R(m, n)$. When divided by the solid angle of the survey it is the "completeness" of the survey. The cgs scale is appropriate to our Lockman Hole (LH) data set.

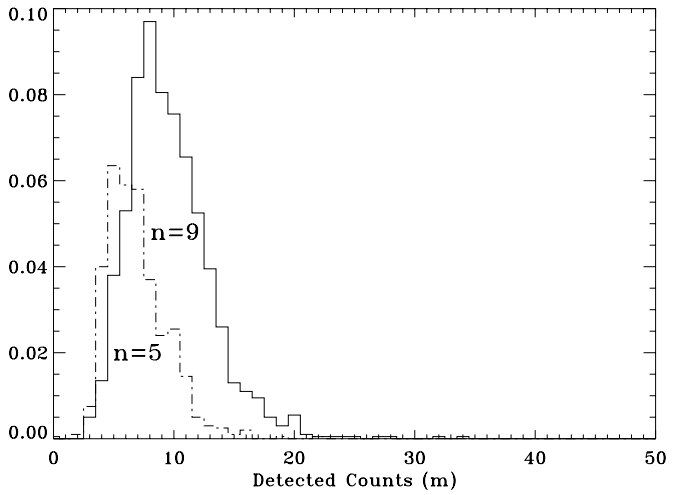

FIG. 2.—The two distributions shown are the $n = 5$ and $n = 9$ columns of the $R(m, n)$ matrix; they show how input $n = 5$-count and $n = 9$-count sources would be redistributed by statistics, by the telescope and instrument, and by detection biases and systematics.

A plot showing $M(m, n)$ is presented in Figure 3. It is clear that for low-count sources there is Poisson migration in detected source counts. This ratio of detected counts to input counts is >2 for low-count sources. Figure 3 also shows that there is a systematic bias in the photometry scale; the ratio of detected source counts to input counts for even the highest-count sources is $\sim 3\%$ low. We have investigated this systematic bias and attribute approximately half of this $\sim 3\%$ to events lost during the simulated CCD frame transfer. The remaining $\sim 1.5\%$ is due to the extent of the wings of the simulated PSF exceeding the wavelet photometry region.

It should be noted that regardless of cause, if the Monte Carlo simulation correctly describes the actual telescope,

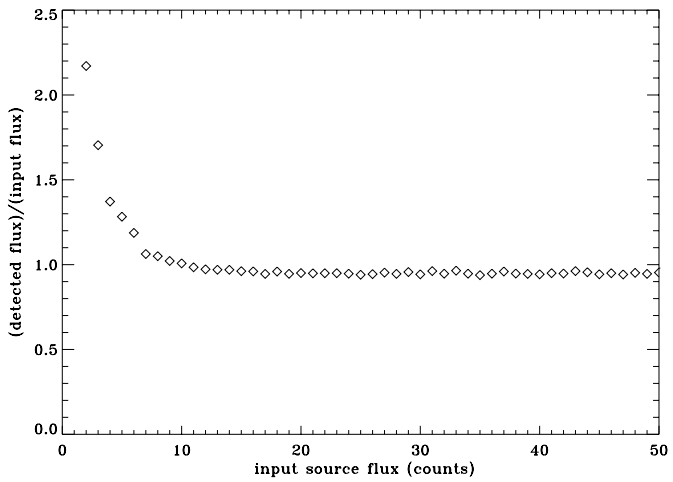

FIG. 3.—Ratio of mean detected source counts to input source counts. This plot shows the effect of "Poisson migration" (Eddington bias). Sources with counts near the detection threshold will be preferentially detected when they have upward fluctuations. The net consequence of this is that the flux density of these sources is overestimated. These data also show that the recovered flux using the wavelet source detection algorithm (wavdetect) is still $\sim 3\%$ low. Approximately half of this $\sim 3\%$ has been attributed to events lost during the simulated frame transfer. The rest is in the wings of the PSF, which are outside the wavelet photometry region.

these biases are embedded in the $R(m, n)$ matrix and are corrected for in the convolution and fitting process of equations (1) and (2). For this reason, the accuracy of our Monte Carlo simulation depends critically on how well MARX represents the response of the *Chandra* telescope and instruments. These simulations used MARX version 3.0. The accuracy of MARX has been tested and is well documented (Wise, Huenemoerder, & Davis 1997; Wise et al. 1999).

This technique does not fully take into account source confusion. Given a random distribution of N sources in a field of view of solid angle Ω, the probability of detecting j sources within the same effective PSF, $\langle \delta\Omega_{\rm psf} \rangle$, can be calculated assuming Poisson statistics to be

$$P(j) = \frac{1}{j!} \exp\{-(N/\Omega)\langle \delta\Omega_{\rm psf} \rangle\} \times \left(\frac{N}{\Omega} \langle \delta\Omega_{\rm psf} \rangle \right)^j \quad (5)$$

For $j = 2$, in the limit of $[(N/\Omega) \times \langle \delta\Omega_{\rm psf} \rangle] \ll 1$ this becomes $(1/2!)[(N/\Omega)\langle \delta\Omega_{\rm psf} \rangle)^2]$. Using our LH data set (~ 400 point sources in 1.35 deg^2), we have calculated $\langle \delta\Omega_{\rm psf} \rangle$ and find $[(N/\Omega) \times \langle \delta\Omega_{\rm psf} \rangle] < 10^{-3}$. The probability of source confusion is negligible given our typical N/Ω and the small *Chandra* PSF.

Source confusion in our technique is analogous to photon pileup in spectral fitting; the nature of the correction is dependent on the $N(n)$ distribution just as the correction for pileup in spectral fitting is dependent on the energy-spectrum. For cases where source confusion is non-negligible, post facto corrections to our technique can be calculated and applied. We have not applied any source confusion correction to our data set.

3. APPLICATION AND VERIFICATION OF THE TECHNIQUE

We have applied our technique to our LH data and on simulated data fields. The simulated data fields were made using the LH data as a template. The $R(m, n)$ matrix that was generated used uniform background levels appropriate to the actual LH observations. We have used the standard CIAO "wavdetect" and "dmtools" software packages (v. 2.2.1) throughout this analysis (Freeman et al. 2002; CIAO users manual 2001[2]).

3.1. *The Lockman Hole (LH) Data*

The LH data set consists of a 1.35 deg^2 mosaic of 21 ACIS-I observations each of ~ 5000 s. The data underwent standard *Chandra* event processing and filtering. The data were then analyzed separately in the 0.5–2 keV, 0.5–7 keV, and 2–7 keV bands. For the purpose of this paper we will limit our discussion to the sources found in the 0.5–7 keV band. Further discussion and analysis of the entire LH data set will be presented in future publications.

The point source population of the LH observations was detected using the CIAO "wavdetect" software. The wavelet scales searched were $\{1, 2, 4, 8\} \times 1\farcs96$ with a probability threshold of 1×10^{-6} for identifying a pixel as belonging to a source. In ~ 5000 s the number of background events in any putative source is negligible; the detections are flux limited. For the total 0.5–7 keV band, 403 sources were detected and the S/N ($\sqrt{n}$) for the faintest sources was ≈ 2

[2] http://cxc.harvard.edu/ciao/download/doc/detect_html_manual

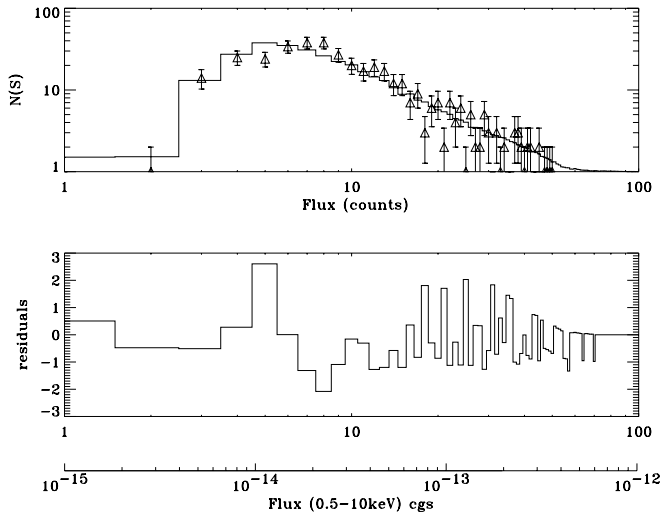

FIG. 4.—Best fit power law $N(n) = A \times n^{-\Gamma}$ convolved with $R(m, n)$ and fitted to detected Lockman Hole $N(m)$ distribution. We have also converted to cgs units based on an average power-law energy spectrum of photon index = 1.7 spanning 0.5–10 keV appropriate for the LH data set. The reduced $\chi^2 = 1.1$ for $\nu = 48$.

σ. The wavelet-determined counts of each detected source were corrected for effective area and exposure time.

The form of $N(S)$ is often modeled as a single or broken power law. To determine the functional form of $N(S)$, we have convolved a power law $N(n) = A \times n^{-\Gamma}$ with $R(m, n)$ and fitted to the LH $N(m)$ data. The results of this fitting are shown in Figure 4.

We find that the underlying $N(n)$ distribution is well described by a single power law with the following form:

$$N(n) = 2122^{+2981}_{-1471} n^{-2.09^{+2.17}_{-1.87}}$$

(Fit reduced $\chi^2 = 1.1$ for $\nu = 48$). The 90% confidence level limits for the parameters are given. We have also converted our results to physical flux units based on a single average power-law energy spectrum of photon index 1.7 and the angular coverage of our survey. In integral form, for the LH data 0.5–10 keV band (extrapolated from 0.5–7 keV), we get

$$N(> S_{14}) = 293^{+411}_{-203} S_{14}^{-1.09^{+1.17}_{-0.87}} \deg^{-2} ,$$

where S_{14} indicates that flux is in units of 10^{-14} ergs cm^{-2} s^{-1}. We have assumed a single average spectrum for the sources; the true conversion of n to S varies from source to source, and this would contribute to our uncertainties.

3.2. *Verification*

The only proper way to verify our technique is to know a priori the underlying $N(S)$ distribution and then to use the technique to reconstruct it; this can be done via Monte Carlo simulation. We have generated simulated observations with source populations that have distributions $N(S)$ with similar statistics to those encountered in our LH data set. The simulated source population $N(n)$ distribution is of the form $N(n) = 2200n^{-2.0}$, where N is the number of sources that would give us n counts in the given observation time.

We used MARX to generate output event files to which we added appropriate amounts of archival ACIS-I background events consistent with our real data. Images were generated and analyzed with the identical source detection algorithms and probability thresholds as were used to analyze the real data. The resultant source counts were corrected by the appropriate exposure map. The detected $N(m)$ distribution was fitted to a power law convolved with our $R(m, n)$ matrix. The data, fit, and residuals are presented in Figure 5. The resultant best fit distribution was

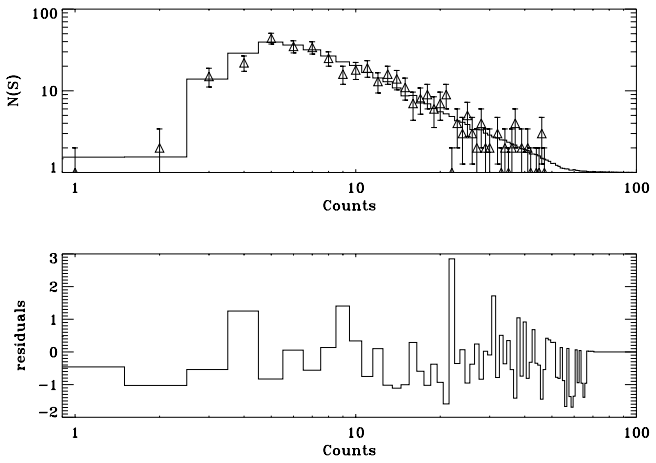

FIG. 5.—Power law $N(n) = A \times n^{-\Gamma}$ convolved with $R(m, n)$ and fitted to detected Monte Carlo simulated $N(m)$ distribution. The reduced $\chi^2 = 0.7$ for $\nu = 48$.

$$N(n) = 2400^{+3500}_{-1640} n^{-2.09^{+2.23}_{-1.88}}$$

with a $\chi^2/\nu = 0.7$ for $\nu = 48$. The 90% confidence level limits for the parameters are given. We have also tested for false detections using just background data. We find that we can expect $\sim 5 \pm 1$ spurious detections in our real data set, given our detection thresholds. Figure 6 shows the input $N(n)$ distribution, the best fit power law to the *input* distribution, and the power law that was recovered via the $R(m, n)$ convolution technique.

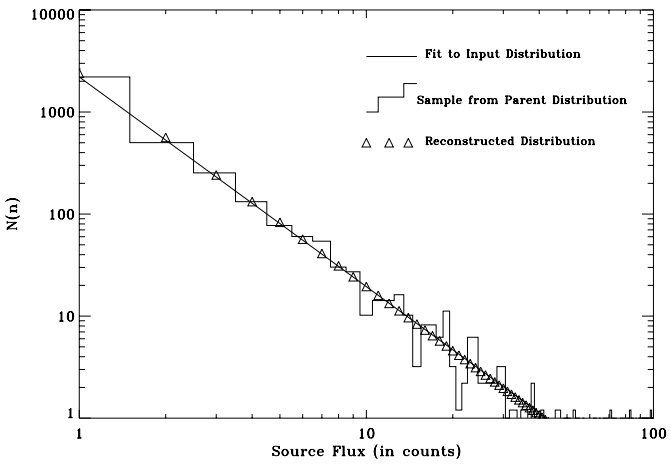

FIG. 6.—Results of Monte Carlo simulation to recover the underlying $N(n)$ distribution. Histogram is the underlying *input* power law sample $N(n)$. The solid line is the fit to the input sample. The triangle data points are the result of recovering the distribution via the $R(m, n)$ convolution and fitting to $N(m)$.

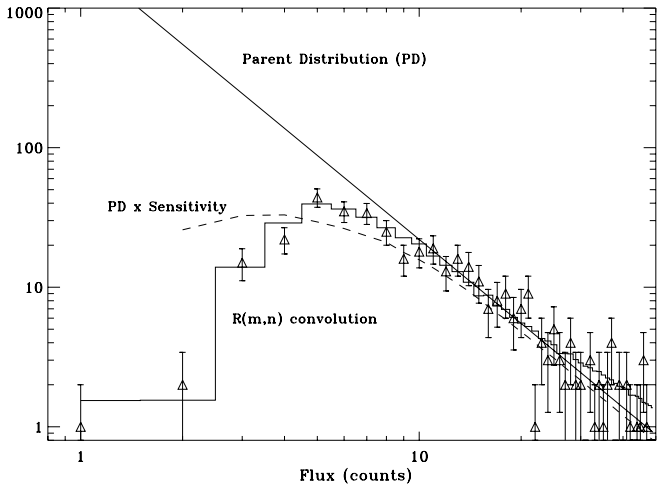

FIG. 7.—This figure shows how the raw detected number flux relationship, $N(m)$ (*data points with error bars*), is successively represented by the parent power law (*straight line*), parent power law times sensitivity (*dashed curve*), and finally by the parent power law convolved with the $R(m, n)$ matrix (*histogram*). It is clear that the convolution technique more accurately represents the data down to sources with ~2 detected counts.

In Figure 7 we present results of Monte Carlo simulations that compare a detected $N(m)$ distribution with the underlying parent $N(n)$ distribution, the parent distribution taking into account sensitivity $[N(n) \times P(n)]$, and finally $N(n)$ convolved with $R(m, n)$. For sources with n higher than ~16 the effect of the various biases are negligible and $N(m) \approx N(n)$. Taking into account sensitivity allows us to get below $m \approx 10$, while the full convolution technique allows us predict $N(m)$ to sources with 2 detected counts.

4. CONCLUSIONS

We have presented a technique for determining the number of X-ray sources per flux density interval that is mathematically identical to spectral fitting. The technique, though of general applicability, was applied to a set of Lockman Hole *Chandra* ACIS-I observations and further verified via Monte Carlo simulations. The benefits of this technique are that it inherently embodies and accounts for statistical and systematic biases, provided that the Monte Carlo is of sufficient fidelity. A series of improvements to the technique is presently being investigated, including more simulations to improve statistics (particularly for the lowest-count sources), improved fitting algorithms that use different or more appropriate fit statistics, and a post facto correction for cases where source confusion is a dominant problem.

Funding for this research was provided by NASA contract NAS8-38248 and NAS8-01130.

REFERENCES

Eddington, A. S. 1940, MNRAS, 100, 354
Freeman, P. E., Kashyap, V., Rosner, R., & Lamb, D. Q. 2002, ApJS, 138, 185
Hasinger, G., Burg, R., Giacconi, R., Hartner, G., Schmidt, M., Trumper, J., & Zamorani, G. 1993, A&A, 275, 1
Moretti, A., Lazzati, D., Campana, S., & Tagliaferri, G. 2002, ApJ, 570, 502
Murdoch, H. S., Crawford, F. F., & Jauncey D. L. 1973, ApJ, 183, 1
Schmitt, J. H. M. M., & Maccacaro, T. 1986, ApJ, 310, 334
Vikhlinin, A., Forman, W., Jones, C., & Murray, S. 1995, ApJ, 451, 542
Wise, M. W., Davis, J. E., Huenemoerder, D. P., Houck, J. C., & Dewey, D. 1999, MARX 3.0 Technical Manual (Cambridge: CXC)
Wise, M. W., Huenemoerder, D. P., & Davis, J. E. 1997, in A. S. P. Conf. Ser. 125, Astronomical Data Analysis Software and Systems VI, ed. G Hunt & H. E. Payne (San Francisco: ASP), 477