Astronomy
&
Astrophysics

Maximum likelihood estimation of single X-ray point-source parameters in ROSAT data

F. G. Boese and S. Doebereiner

Max–Planck–Institut für extraterrestrische Physik, 85741 Garching, Germany

Abstract. The point source parameter estimation via Maximum Likelihood Estimation in ROSAT data is documented. The infall of X-ray photons is treated as a Poisson process. Based on the Poisson model, the log-likelihood function for unbinned data is derived. The adopted source and background models are explained. In the chosen parameterization, the Maximum Likelihood equations and their numerical solution are expounded. The performance of the method is illustrated by an example and tests on simulated observations.

Key words. X-ray: general, space vehicles: instrumentation: detectors, techniques: image processing

1. Introduction

In a class of contemporary X-ray observatories, the observing instruments consist of one or more sets of nested telescopes and one or more detectors in the focal plane. The detectors are often designed to register single X-ray photons. Several geometrical and physical quantities of individual X-ray photons are recorded, see Trümper (1983). In the case of the ROSAT observatory, these are (a) the time of photon arrival, (b) the impact location in the detector, (c) the amplitude of the detector signal, and (d) the infalling direction[1]. Depending on the ROSAT observation mode, the instrument's optical axis is, relative to the observed target, nominally fixed or moving in time. During an observation time interval, a number of photons is collected. After data preparation, the observed sample forms the data set to be analyzed.

The majority of all observed celestial sources are point sources. The capacity of analyzing point sources with respect to (a) X-ray flux, (b) celestial direction and (c) extent (if larger than but akin to point sources) thus makes it possible to treat a large percentage of all observed sources routinely.

In this work, the technical details of the statistical inference from data to point-like and circular disk shaped X-ray sources are investigated. To X-ray flux, celestial direction, and source extent correspond point source parameters. Estimation of these parameters using Maximum Likelihood Estimation (MLE) will be described in general

and also some of the peculiarities of the implementation within the ROSAT data analysis system EXSAS. The latter software system assists in analyzing ROSAT observational data, see Zimmermann et al. (1998). It is distributed to more than 300 observers in 75 institutes worldwide.

Up to December 2000, about 4 500 scientific papers based on ROSAT data appeared in journals. Among the many ROSAT catalogues are the ROSAT All-Sky Survey Bright Source Catalogue with 18 811 sources and the Source Catalogue from Pointed PSPC Observations with 74 301 sources, see Voges et al. (1999). A considerable percentage of all related data analysis was carried out with the SASS or EXSAS analysis systems – an imperative reason to document the underlying method. Apart from a preliminary conference report by Cruddace et al. (1987), no technical description exists, internally or publically.

In Sect. 2, the observed ROSAT sample is presented. In Sect. 3, the source model is introduced and the logarithmic likelihood functional for unbinned data is derived from basic principles in considerable generality. Background model and point source model are set up in Sect. 4. The maximum likelihood equations are compiled in Sect. 5. Statistical errors for the estimated source parameters will be discussed in Sect. 6. The iterative solution of the maximum likelihood equations is treated in Sect. 7. A point source estimation example is contributed and performance tests on simulated observations are presented in Sect. 8. The finishing Sect. 9 contains the discussion followed by the conclusions in Sect. 10. An Appendix, Sect. 11, explains EXSAS commands related to the MLE of point sources.

Send offprint requests to: F. G. Boese, e-mail: gub@mpe.mpg.de

[1] This direction is not recorded per photon but evaluated per photon. The attitude of the optical axis is used thereby.

2. The observed sample

The ROSAT satellite was launched on July 1, 1990 into a circular orbit with an inclination of $53°$ and an altitude of 580 km. The last observation was made on December 18, 1998. The ROSAT detectors were (a) the Positional Sensitive Proportional Counter (PSPC), (b) the High Resolution Imager (HRI), and (c) the Wide Field Camera (WFC). Two telescopes were in operation: the larger one for the PSPC and HRI, the smaller for the WFC.

2.1. The photon representation in ROSAT data

During an observation interval[2] $\boldsymbol{T} := [\underline{T}, \bar{T}]$ of length $T := \bar{T} - \underline{T}$ starting at time $t = \underline{T}$ and ending at time $t = \bar{T} := \underline{T} + T$, the instrument consisting of a telescope and a detector registers a number $M := M(\underline{T}, \bar{T})$ of photons[3]. The mth photon, $\wp_m$, $m = 1, 2, ..., M$, comes from the celestial direction (X_m, Y_m), arrives at time $t = t_m$, and hits the instrument's focal plane at the detector location (x_m, y_m). The photon causes a detector signal amplitude A_m.

Upon completion of the observation and data preprocessing, we are given a 6-variate time-ordered photon series[4]

$$\mathcal{P} := \{\wp_k\}_{k=1}^{M},$$
$$\wp_k := (t_k, x_k, y_k, X_k, Y_k, A_k), \quad t_1 < t_2 < ... < t_M. \quad (1)$$

Let S be an estimate for the number of sources in the sample $\mathcal{P}$. Then, the full sample will be partitioned into $S + 1$ subsamples in Sect. 4.1,

$$\mathcal{P} =: \mathcal{P}_0 \cup \mathcal{P}_1 \cup ... \cup \mathcal{P}_S. \quad (2)$$

To the first source belongs subsample $\mathcal{P}_1$, to the second source belongs subsample $\mathcal{P}_2$, and so forth. The remainder subsample, $\mathcal{P}_0$, comprises all photons which do not belong to any subsample $\mathcal{P}_1$ to $\mathcal{P}_S$. The latter background corrected subsamples together with a background distribution derived from $\mathcal{P}_0$ are the input data for MLE. While the size M of $\mathcal{P}$ may be in the order of millions, the sizes of $\mathcal{P}_1$ to $\mathcal{P}_S$ may go down to dozens. The processing time for the whole sample $\mathcal{P}$ is a function of the size M.

2.2. The amplitude space

Source spectra can be formed when the detector discriminates between photon energies. Often, detector amplitudes are the primary measured quantities and not the photon

---

[2] The expressions "$A := B$" or "$B =: A$" mean: A is defined by B.

[3] Generally, the incoming photon time series is different from the registered time series due to the counter characteristics like dead time or absorption. Here, we assume that all possible corrections are done, the detector is operated within specifications so that no significant differences between the two processes remain.

[4] In EXSAS terminology, $\mathcal{P}$ is held in the *photon event file*.

Fig. 1. The 256×729 PSPC detector response matrix $p_{c|k}$ from (5). For $c = 1$ to 7, the entries are set to zero $p(c \mid k) = 0$ (to avoid noisy data). Dark = high values $p_{c|k} \in [0, 0.0826]$

energies. Then, spectra are estimated by statistical inference in the detector's amplitude space $\boldsymbol{A}$ to be presented here.

Let the instrument be designed for the *photon energy space* $\boldsymbol{E} := [\underline{E}, \bar{E}]$ of width $\Delta E := \bar{E} - \underline{E}$. A single infalling photon of energy $E \in \boldsymbol{E}$ causes an electrical impulse of amplitude A containing a superimposed (small) random amplitude in the instrument's *amplitude space* $\boldsymbol{A} := [\underline{A}, \bar{A}]$ of width $\Delta A := \bar{A} - \underline{A}$. The response distribution to a Dirac delta input single line photon spectrum $\delta(E - E')$, $E, E' \in \boldsymbol{E}$, of energy E has the probability density

$$p(A|E), \qquad \int_{\boldsymbol{A}} p(A|E) \mathrm{d}A = 1. \quad (3)$$

Due to absorption effects at both ends of $\boldsymbol{E}$, the normalization in (3) is there only approximative. Neglecting this, $p(A|E)\mathrm{d}A$ is the fraction of all photons of given energy E causing detector amplitudes lying in an interval of infinitesimal length $\mathrm{d}A$ around the amplitude A. We partition the amplitude space $\boldsymbol{A}$ into C amplitude channels,

$$\boldsymbol{A} =: \boldsymbol{A}_1 \cup ... \cup \boldsymbol{A}_C, \quad \boldsymbol{A}_c := [A_c, A_{c+1}],$$
$$\Delta A_c := A_{c+1} - A_c, \quad c = 1, ..., C. \quad (4)$$

The index c is called *channel number*, a reminiscence of the involved electronics. The discrete version of $p(A \mid E)$ for the channel numbers c from (4),

$$p_{c|k} := \frac{1}{\Delta E_k} \int_{\boldsymbol{A}_c \times \boldsymbol{E}_k} p(A \mid E) \mathrm{d}A \mathrm{d}E, \quad 1 \le k \le K, \quad (5)$$

is called (energetical) *detector response matrix*. Figure 1 shows the last update of this measured matrix for the ROSAT PSPC detector. It gives a good overview on the development of the mode (= peak) position and spread of

$p(A \mid E)$ with increasing energy E. Here $C = 256$ amplitude channels of equal width partition $\boldsymbol{A} := [1, 256]$. The energy band $\boldsymbol{E} := [0, 3]\,\mathrm{keV}$ is partitioned by $K = 729$ energy bins of unequal widths ΔE_k. To hold the impulse height distribution density $p(A \mid E)$ internally in table form $p_{c|k}$ allows us a faster access than function evaluation.

2.3. The photon space

According to (1), a photon $\wp := (t, x, y, X, Y, A)$ in ROSAT data has six coordinates t to A. Four of them, t, X, Y, A, are for MLE of prime significance and span the *photon space* $\boldsymbol{P}$. This is the space in which all calculations are done, where the sought distributions are accommodated. To understand this space thoroughly, we follow the photon's optical path through the telescope to the detector plane. Its relation to the (X, Y) celestial directions space will be explained.

2.3.1. The astronomical directions space

The equatorial system with mean epoch 2000 is taken in ROSAT as the principal astronomical coordinate system.

Disregarding all distances from Earth's centre to celestial objects leads to the unit sphere $\boldsymbol{S}'^2$ in 3-space referred to equatorial coordinates $\{\mathcal{O}'; \alpha, \delta\}$. The choice of the distance unit is unimportant. The null direction $\mathcal{O}'$ is the vernal equinox, the angle $\alpha \in [0, 2\pi)$ is the *right ascension* and the angle $\delta \in [-\pi/2, \pi/2]$ the *declination*. Seen by a ROSAT detector, there is a photon count rate distribution $\lambda'(t, \alpha, \delta, A)$ representing the celestial directional distribution in X-rays at time t, in photon detector amplitude A, and coming from direction (α, δ). ROSAT took samples from this celestial distribution $\lambda'(t, \alpha, \delta, A)$. Count rates will be formally introduced in (18).

ROSAT's pointing direction $p'(t) := (\alpha_p(t), \delta_p(t)) \in \boldsymbol{S}'^2$ moves in time t due to (a) a possible survey motion, (b) a normally applied wobbling motion, and (c) an unavoidable stochastical motion caused by the attitude control loop dynamics and, at that, excited by external stochastical forces. The wobbling motion will be explained in Sect. 2.3.3.

2.3.2. The tangential space

The direct approach to calculate with data on the sphere $\boldsymbol{S}'^2$ was not taken. Instead, spherical data are mapped to a tangential plane referred to Cartesian coordinates. In this plane, all directional calculations are actually effectuated. At the end of the analysis, the calculated results are mapped back to the sphere, i.e. expressed in equatorial coordinates[5]. We shall encounter the mapping details

by explaining the geometry and kinematics of the data acquisition.

Consider the pointing direction $p' := (\alpha_p, \delta_p) \in \boldsymbol{S}'^2$ as it evolves in time t. For any fixed time, a tangential Cartesian coordinate system $\mathcal{T} := \{p'; X, Y, Z\}$ with origin at[6] p' is erected. The Z axis goes through the origin of the unit sphere and the point p'. The X and Y axes lie in the tangential plane $\boldsymbol{T}(p') =: \boldsymbol{S}$ with tangency point p'. The plane $\boldsymbol{S}$ is termed *sky space*. It is thus adapted to the individual pointed observation. To explain the axes' directions, consider the two oriented, mutually orthogonal circles of constant right ascension $\mathcal{C}(\alpha_p)$ and constant declination $\mathcal{C}^\perp(\delta_p)$ through p',

$$\mathcal{C}(\alpha_p) := \left\{ (\alpha, \delta) \in \boldsymbol{S}'^2 : \ \alpha = \alpha_p \right\},$$
$$\mathcal{C}^\perp(\delta_p) := \left\{ (\alpha, \delta) \in \boldsymbol{S}'^2 : \ \delta = \delta_p \right\}. \tag{6}$$

$\mathcal{C}(\delta_p)$ is oriented with increasing right ascension α and $\mathcal{C}^\perp(\alpha_p)$ with increasing declination δ. Denote by $t_\alpha(p')$ the tangential vector along $\mathcal{C}^\perp(\delta_p)$ at p' and by $t_\delta(p')$ the tangential vector along $\mathcal{C}(\alpha_p)$. The requirements

$t_\alpha(p')$ is the direction of the X axis,

$t_\delta(p')$ is the direction of the Y axis $\qquad$ (7)

define the Cartesian coordinate system $\mathcal{T}$ uniquely. Now, we carry out a rotation $S(t)$ of the celestial sphere $\boldsymbol{S}'^2$ to obtain $\boldsymbol{S}''^2$ which will be referred to the *relative spherical coordinate system* $\{\mathcal{O}''; \alpha', \delta'\}$. The new null direction $\mathcal{O}''$ becomes the pointing direction p' and the new equatorial plane is the (X, Z) plane. The Y axis points to the new north pole $NP' := (\alpha_p(t), \delta_p(t) + \pi/2)$. In the reverse direction lies the new south pole $SP' := (\alpha_p(t), \delta_p(t) - \pi/2)$. In the new spherical coordinate system, the general celestial direction $p := (\alpha, \delta) \in \boldsymbol{S}'^2$ has the *relative longitude* α' and *relative latitude* δ'. These can be expressed by the tangential coordinates $X_\mathrm{T}, Y_\mathrm{T}, Z_\mathrm{T}$ in $\mathcal{T}$ by

$$\alpha' = \alpha'[\alpha, \delta, \alpha_p(t), \delta_p(t)] := \arctan\left[\frac{X_\mathrm{T}}{1 + Z_\mathrm{T}} \right],$$
$$\delta' = \delta'[\alpha, \delta, \alpha_p(t), \delta_p(t)] := \arcsin\left[Y_\mathrm{T} \right],$$
$$X_\mathrm{T} := \sin[\alpha - \alpha_p(t)] \cos(\delta),$$
$$Y_\mathrm{T} := \sin(\delta)\cos(\delta_p) - \cos[\alpha - \alpha_p(t)]\cos(\delta)\sin(\delta_p),$$
$$1 + Z_\mathrm{T} := \cos[\alpha - \alpha_p(t)]\cos(\delta)\cos(\delta_p) + \sin(\delta)\sin(\delta_p). \tag{8}$$

For $\alpha_p = \delta_p = 0$, no rotation is performed, consequently, (8) becomes the identity, $\alpha' = \alpha$, $\delta' = \delta$. The projection $\pi_{\boldsymbol{S}}$ from the rotated sphere $\boldsymbol{S}''^2$ to the tangential plane $\boldsymbol{S}$ is of cylindrical type with respect to both spherical coordinates,

$$X = \alpha', \quad Y = \delta', \quad -\pi \le X < \pi, \ -\pi/2 \le Y \le \pi/2. \tag{9}$$

The points on $X = -\pi$ and $X = \pi$ are to be pointwise identified making the domain (9) a cylinder. The new poles

---

[5] When appropriate for the object under study, further transformation to other astronomical coordinate systems, i.e. ecliptic or galactic, may follow.

[6] Points on the unit sphere and directions are denoted by the same symbol. The actual meaning becomes clear from the context.

NP', SP' are mapped to the line segments $Y = \pm\pi/2$. This means that large length distortions must emerge near these lines. Rolling the cylinder (9) first along the new equator of $\boldsymbol{S''}^2$ and then along the new meridional direction to reach the final point (α', δ') realizes the mapping (9). We call the composed mapping $A(t) := S(t) \circ \pi_{\boldsymbol{S}}$: $\boldsymbol{S'}^2 \to \boldsymbol{S}$ attitude-sky mapping. Since the sphere rotation $S(t)$ is invertible and obviously (9) too, the attitude mapping $A(t)$ is invertible $A(t)^{-1} := \pi_{\boldsymbol{S}}^{-1} S(t)^{-1}$.

The two coordinates in the sky space $\boldsymbol{S}$ are termed *sky longitude X* and *sky latitude Y*. Both coordinates are measured in angular units, in units of $sky\_pix := 0.5$ arcsec. This unit is appropriate for the spatial resolutions of the three ROSAT detectors.

A fixed celestial direction $p := (\alpha, \delta)$ is mapped by $A(t)$ at time t to the sky point $(X(t), Y(t))$. All ROSAT detectors have fields of view in the order $|\alpha'|, |\delta'| \approx 1°$. Evaluating $A(t)$ for small fields of view, i.e. for $\max\{|\alpha - \alpha_p|, |\delta - \delta_p|\} \to 0$ yields[7]

$$X = \alpha - \alpha_p + \mathcal{O}\left[\max\{|\alpha - \alpha_p|^2, |\delta - \delta_p|^2\}\right],$$
$$Y = \delta - \delta_p + \mathcal{O}[\max\{|\alpha - \alpha_p|^2, |\delta - \delta_p|^2\}]. \quad (10)$$

Figure 2 shows the image curves of the 12 circles[8] $\mathcal{C}(\alpha)$, $\alpha = 0°(30°)330°$, and of the 11 circles $\mathcal{C}^\perp(\delta)$, $\delta = -75°(15°)75°$, for the pointing direction $p' := (30°, 45°)$. As mentioned, ROSAT observed the source count rate distribution $\lambda'(\alpha, \delta) := \lambda'(t, \alpha, \delta, A) \geq 0$ on the celestial sphere $\boldsymbol{S'}^2$. The attitude mapping $A(t)$ is not area-preserving. We discuss the consequence for our photon counting experiment.
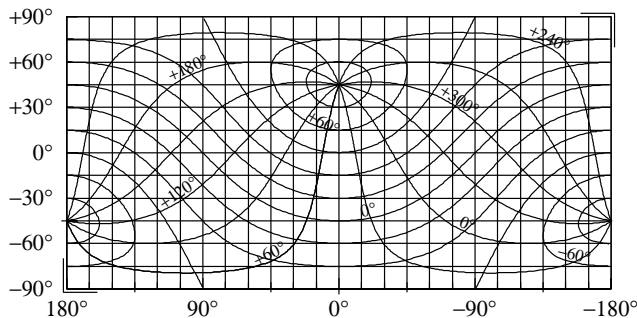
Fig. 2. The images of the circles $\mathcal{C}(\alpha)$ and $\mathcal{C}(\delta)$ with $\alpha = 0°(30°)330°$ and $\delta = -75°(15°)75°$ under the attitude map $A(t)$ (8),(9) for the pointing direction $p' := (30°, 45°)$

The one-to-one image $\lambda(X, Y)$ of $\lambda'(\alpha, \delta)$ under the attitude mapping $A(t)$ is distorted in shape by the aereal distortion factor $d[\alpha, \delta, \alpha_p(t), \delta_p(t)]$ appearing in

$$\lambda(X, Y)\mathrm{d}X\mathrm{d}Y := \lambda'(\alpha, \delta)d[\alpha, \delta, \alpha_p(t), \delta_p(t)]\mathrm{d}\alpha\mathrm{d}\delta,$$

$$d[\alpha, \delta, \alpha_p(t), \delta_p(t)] := \frac{D[\alpha, \delta, \alpha_p(t), \delta_p(t)]}{\sqrt{1 - Y_T^2} \cdot [(1 + Z_T)^2 + X_T^2]},$$

[7] From now on, $\mathcal{O}$ stands for Landau's order symbol and no longer for the origin of a coordinate frame.

[8] The expression "$v = a(b)c$" means: variable v runs from a to c in steps of b.

$$D[\alpha, \delta, \alpha_p(t), \delta_p(t)] := (1 + Z_T) \cdot (X_{T,\alpha} Y_{T,\delta} - X_{T,\delta} Y_{T,\alpha})$$
$$- X_T \cdot (Z_{T,\alpha} Y_{T,\delta} - Z_{T,\delta} Y_{T,\alpha}). \quad (11)$$

In the count rate transformation rule (11), (X, Y) and (α, δ) are mutual image points. The suffixes α, δ on X_T, Y_T, Z_T denote partial derivatives with respect to suffix variables. The factor $|D| \leq 2$ is free of singularities and coprime with the denominator factors. Hence, the points on the line $Z_T = -1$, $X_T = 0$ intersecting $\boldsymbol{S''}^2$ are pole singularities of d. The latter are located at the new pole pair NP', SP' having as image under $A(t)$ the pair of horizontal line segments $Y = \pm\pi/2$ in Fig. 2. The weaker singularity of d in the plane pair $Y_T = \pm 1$, the tangential plain pair at the poles NP', SP' specifies thus again the new pole pair.

From (8), it follows near the origin $(X, Y) = (0, 0)$ in the sky space $\boldsymbol{S}$, i.e. for $|\alpha - \alpha_p(t)|, |\delta - \delta_p(t)| \to 0$,

$$d[\alpha, \delta; \alpha_p(t), \delta_p(t)] = 1 + \mathcal{O}(\max\{|\alpha - \alpha_p(t)|, |\delta - \delta_p(t)|\}). \quad (12)$$

The attitude mapping $A(t)$ is thus almost area-preserving near the pointing direction. A distorted source count density distribution $\lambda(X, Y)$ may give the non-expert inspector a misleading visual impression and may cause the software to generate artificial sources. The use of $A(t)$ only near the origin $(X, Y) = (0, 0)$ together with the local almost area-preserving property (12) excludes this risk greatly. No correction factor d for transforming count rates from the (X, Y) directional space to equatorial space $\boldsymbol{S'}^2$ is calculated in EXSAS. For a thorough survey article on celestial coordinates, see Calabretta & Greisen (2000).

2.3.3. The detector space

Photon impact locations in the detector plane $\boldsymbol{R}^2$ are measured in the plane Cartesian detector coordinate system $\{\mathcal{O}; x, y\}$. The origin, $\mathcal{O}$, is the piercing point of the instrument's optical axis in the detector plane. The *detector field $\boldsymbol{D}$* contains the *field of view $\boldsymbol{F}$* and is parallel to the axes in $\boldsymbol{R}^2$, i.e. $\boldsymbol{F} \subset \boldsymbol{D} \subset \boldsymbol{R}^2$. Figure 3 shows the field of view (black) of the ROSAT PSPC detector. The detector field $\boldsymbol{D}$ is a bit larger than the smallest square parallel to the axes containing $\boldsymbol{F}$. The pointing direction $p'(t)$ and the optical axis direction coincide[9]. Consequently, the detector plane $\boldsymbol{R}^2$ must be a rotated sky space plane $\boldsymbol{R}^2$. The time-dependent rotation angle $\rho(t)$ is called *roll angle*. It is defined to be the angle between the detector x-axis and sky X-axis measured counterclockwise in the sky coordinate system. The normal direction to the surface of the ROSAT solar paddles has to point to the Sun. This requirement defines the roll angle $\rho(t)$ at time t.

Taking into account the different units in both spaces leads to the invertible linear mapping between sky space

[9] The measured misalignment between both directions gives rise to a "boresight correction".
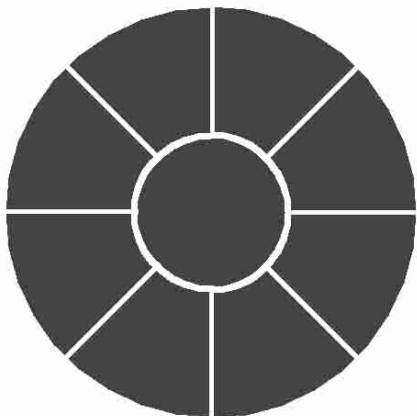
Fig. 3. The PSPC field of view $\boldsymbol{F}$ and detector field $\boldsymbol{D}$. Photons are censored by the support ribs, the hub (white), and outside the black disk

and detector space

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = R(t) \begin{pmatrix} X \\ Y \end{pmatrix},$$

$$R(t) := c \cdot \begin{pmatrix} \cos[\rho(t)], & \sin[\rho(t)] \\ -\sin[\rho(t)], & \cos[\rho(t)], \end{pmatrix},$$

$$c := \frac{\text{det\_pix}}{\text{sky\_pix}}. \tag{13}$$

The pointing direction $p'(t)$ supplemented by the roll angle $\rho(t)$ is called *satellite attitude* $a(t) := (p'(t), \rho(t))$. We call the composed invertible mapping from astronomical directions to detector points $D(t) := R(t) \circ A(t)$: $\boldsymbol{S'}^2 \to \boldsymbol{R}^2$ *attitude-detector mapping*. The knowledge of the instrument's attitude $a(t)$ at any time $t \in \boldsymbol{T}$ together with the relation between detector positions and photon infall directions (ray-trace relation) permits us to map detector positions to celestial directions by the inverse $D(t)^{-1}$.

Assume a time-constant nominal pointing direction p'_0 and imagine a source which falls on a rib in Fig. 3. Such a source is totally censored. In order to lessen the detrimental censoring by the ribs, a periodic wobbling motion $w(t)$ in saw-tooth form and with a time-constant tangential vector W is superimposed to the effect that the said photons are now distributed along a line segment crossing the rib (when the rib and W are not parallel),

$$p'(t) := p'_0 + W \cdot w(t), \quad w(t + T_w) = w(t), \quad W \in \boldsymbol{S}. \tag{14}$$

The nominal wobbling period was $T_w := 403\,\text{s}$ and the nominal wobbling amplitude $|W| = 3\,\text{arcmin}$. As Fig. 8 witnesses, the ribs from Fig. 3 are almost made to disappear due to wobbling motion.

2.3.4. The ROSAT photon space

The sample $\mathcal{P}$ from (1) lies in the instrument's photon space $\boldsymbol{P}$. This is the Cartesian product of the observation interval $\boldsymbol{T}$, the detector field $\boldsymbol{D}$, the sky space $\boldsymbol{S}$, and the detector's amplitude space $\boldsymbol{A}$,

$$\boldsymbol{P} := \boldsymbol{T} \times \boldsymbol{D} \times \boldsymbol{S} \times \boldsymbol{A}. \tag{15}$$

In our setting, the task is to infer the hypothesized parent source distribution $\lambda(t, X, Y, A)$ in $\boldsymbol{P}$ from which the sample $\mathcal{P} \subset \boldsymbol{P}$ stems.

3. The source model and the log-likelihood function

As mentioned in Sect. 1, some state-of-the art detectors register the X-ray photons one after the other in time, resolve them with respect to energy and infall direction.

A given instrument has its individual photon space. Its dimension may be lower, equal, or larger than that of (15). The constitutive factor spaces may be different from the factors $\boldsymbol{T}, \boldsymbol{D}, \boldsymbol{S}, \boldsymbol{A}$ in (15). In this section, $\boldsymbol{P}$ is the photon space of any astronomical photon counting instrument. However, to be not overly abstract in this section, a paradigmatic choice is made: photons possess an arrival time t on a certain (X, Y) target space and transport an energy E. Let $\mathcal{P} \subset \boldsymbol{P}$ be a sample of size M in this section.

It is appropriate to view the infall of photons as a stochastic *counting process* in the instrument's photon space $\boldsymbol{P}$.

3.1. The Poisson source model

As X-ray photons are observed one by one (by introductory assumption), a stochastical point of view is adopted in the pertinent parts of this article. A *Poisson emitter* in photon space $\boldsymbol{P}$ is the stochastic X-ray source model adopted for ROSAT data and defined by (P1), (P2) in (16) below.

Let $\boldsymbol{Q} \subset \boldsymbol{P}$ be an arbitrary set in photon space and $N(\boldsymbol{Q})$ be the random number of photons falling into $\boldsymbol{Q}$. According to the definition, of a Poisson process, see Parzen (1962, p. 118), two properties of the process $\{N(\boldsymbol{Q}), \boldsymbol{Q} \subset \boldsymbol{P}\}$ are implied:

(P1) $N(\boldsymbol{Q})$ has independent increments: whenever $\boldsymbol{Q}_1, ..., \boldsymbol{Q}_n$ are disjoint sets in $\boldsymbol{P}$, the photon counts $N(\boldsymbol{Q}_1), ..., N(\boldsymbol{Q}_n)$ are independent random variables.

(P2) For any $\boldsymbol{Q} \subset \boldsymbol{P}$, $N(\boldsymbol{Q})$ is Poisson distributed,

$$\text{Prob}[N(\boldsymbol{Q}) = n] := \frac{e^{-\Lambda(\boldsymbol{Q})}\Lambda(\boldsymbol{Q})^n}{n!}, \quad n = 0, 1, ... \tag{16}$$

The relation $\boldsymbol{E}[N(\boldsymbol{Q})] = \Lambda(\boldsymbol{Q})$ provides the statistical interpretation of the intensity parameter $\Lambda(\boldsymbol{Q})$ as the expected counts from $\boldsymbol{Q}$. The functional $\boldsymbol{E}$ takes the expectation of a random variate. In terms of the *local mean count rate* λ, we have the representation

$$\Lambda(\boldsymbol{Q}) := \int_{\boldsymbol{Q}} \lambda \, \mathrm{d}\boldsymbol{Q}. \tag{17}$$

Basically, in an actual application, it remains to be statistically tested whether (16) is compatible with the data $\mathcal{P}$ or not. However, particle counting experiments belong

since the days of Lord Rutherford to the stock of canonical applications of Poisson processes.

The *local mean count rate*, λ, introduced in (17) depends on all four photon parameters spanning $\boldsymbol{P}$ and is assumed to be bounded,

$$\lambda := \lambda(t, X, Y, E). \tag{18}$$

Consequently, the infinitesimal increment $\lambda \mathrm{d}\boldsymbol{Q}$ becomes

$$\mathrm{d}\Lambda(t, X, Y, E) := \lambda(t, X, Y, E)\, \mathrm{d}t\mathrm{d}X\mathrm{d}Y\mathrm{d}E, \tag{19}$$

the local mean number of photon counts expected in the infinitesimal box $\mathrm{d}\boldsymbol{Q}$ (parallel to the axes in $\boldsymbol{P}$) around the central point $(t, X, Y, E) \in \boldsymbol{P}$ with the lengths elements $\mathrm{d}t$, $\mathrm{d}X$, $\mathrm{d}Y$, $\mathrm{d}E$ and the 4-volume element $\mathrm{d}V := \mathrm{d}t\mathrm{d}X\mathrm{d}Y\mathrm{d}E$. Notice that to the local mean count rate λ from (18) belongs the unit

$$\frac{(\text{photons})}{(\text{unit time}) \cdot (\text{unit area}) \cdot (\text{unit energy})}. \tag{20}$$

The implied unit of the 4-volume is given by the denominator in (20).

In possession of the sample $\mathcal{P} \subset \boldsymbol{P}$ and having specified the parent source model, we have to decide by which statistical estimation procedure we wish to infer $\lambda(t, X, Y, E)$ from $\mathcal{P}$.

Among several standard estimation procedures at disposal, estimation via MLE has proven to be appropriate in astronomical application, see Cash (1979), Strong (1985) and also in X-ray astronomy with ROSAT, see Cruddace et al. (1987).

In X-ray observations, one has often to cope with weak sources. Due to observation time restrictions, not more than a handful of photons are collected from weak sources. Any loss of data information due to binning is unacceptable. Technically spoken, many bins without counts are the rule. As one might expect from (P2) in (16) and as we shall see in (22), the statistics associated with MLE for photon counting is based on the Poisson distribution – a distribution for rare events. It is applicable for small occupation numbers. So, the choice of MLE is almost mandatory for ROSAT data under small source sample sizes and/or high precision requirements.

3.2. Derivation of the likelihood function

Within the MLE we have to establish the *log-likelihood function*, L, for our photon counting experiment. While the likelihood e^L is in our context the prime probability quantity, it is technically preferable to assign this role to the exponent $L \leq 0$. The L, seen as a functional of the hypothetical count rate, is proportional to the logarithm of the probability that the observation emerged from the assumed count rate, given the observation data.

To start with the derivation of L, imagine a partition Δ of the instrument's photon space $\boldsymbol{P}$ into finitely many bins $\boldsymbol{P}_{i,j,k,l}$. The independent increment property (P1) in (16) allows us to partition the photon space $\boldsymbol{P}$ arbitrarily into

finitely many parts of positive 4-volume and to have identical independent distribution forms in each bin $\boldsymbol{P}_{i,j,k,l}$. We are free to design bins in form of Cartesian products so that the partition Δ has the bins

$$\boldsymbol{P}_{i,j,k,l} := \boldsymbol{X}_i \times \boldsymbol{Y}_j \times \boldsymbol{E}_k \times \boldsymbol{T}_l,$$
$$\boldsymbol{X}_i := [X_i, X_{i+1}), \qquad X_{i+1} - X_i =: \Delta X_i > 0,$$
$$\boldsymbol{Y}_j := [Y_j, Y_{j+1}), \qquad Y_{j+1} - Y_j =: \Delta Y_j > 0,$$
$$\boldsymbol{E}_k := [E_k, E_{k+1}), \qquad E_{k+1} - E_k =: \Delta E_k > 0,$$
$$\boldsymbol{T}_l := [T_l, T_{l+1}), \qquad T_{l+1} - T_l =: \Delta T_l > 0,$$
$$1 \leq i \leq I,\ 1 \leq j \leq J,\ 1 \leq k \leq K,\ 1 \leq l \leq L'. \tag{21}$$

Let $n_{i,j,k,l}$ be the number of counts[10] observed in bin $\boldsymbol{P}_{i,j,k,l}$. By (P1), these counts are independent random variates. Hence, their joint probability is the product of the involved individual probabilities. Assuming a continuous count rate λ, the Poisson distribution property (P2) from (16) gives the probability $P_\Delta(n_{1,1,1,1}, ..., n_{I,J,K,L'})$ to observe the count numbers $n_{1,1,1,1}$ to $n_{I,J,K,L'}$ as

$$P_\Delta(n_{1,1,1,1}, ..., n_{I,J,K,L'}) :=$$
$$\prod_{i,j,k,l} \frac{\mathrm{e}^{-\lambda_{i,j,k,l}\Delta V_{i,j,k,l}} \left(\lambda_{i,j,k,l}\Delta V_{i,j,k,l}\right)^{n_{i,j,k,l}}}{n_{i,j,k,l}!},$$
$$\lambda_{i,j,k,l} := \frac{1}{\Delta V_{i,j,k,l}} \int_{\boldsymbol{P}_{i,j,k,l}} \lambda(t, X, Y, E)\, \mathrm{d}t\mathrm{d}X\mathrm{d}Y\mathrm{d}E,$$
$$\Delta V_{i,j,k,l} := \Delta X_i \Delta Y_j \Delta E_k \Delta T_l. \tag{22}$$

The product in (22) extends over all bin multi-indices (i, j, k, l) from (21). The quantities $\lambda_{i,j,k,l}$ are called *bin-averaged count rates*.

We wish to get rid of the loss of information incurred by the finite binning (21). To this end, we consider an infinite sequence $\{\Delta^{(1)}, \Delta^{(2)}, ...\}$ of binnings of the photon space $\boldsymbol{P}$ with the infinite resolution property

$$\Delta X_i \to 0, \quad \Delta Y_j \to 0, \quad \Delta E_k \to 0, \quad \Delta T_l \to 0 \tag{23}$$

for all the $I \cdot J \cdot K \cdot L'$ bins $\boldsymbol{P}_{i,j,k,l}$ of $\Delta := \Delta^{(n)}$ for $n \to \infty$. We will write simpler $\Delta \to 0$ for the passages to the limits in (23) for $n \to \infty$.

The same observation appears then in infinitely many binnings. We wish to investigate the likelihood $L_\Delta := \ln[P_\Delta]$ in the passage $\Delta \to 0$.

We split the probability $P_\Delta =: P_\Delta[\lambda]$ in (22) into three factors. The first one, $P_{\Delta,\mathrm{e}}$, comprises all normalization factors and the remaining two are the contributions from the void and occupied bins,

$$P_\Delta := P_{\Delta,\mathrm{e}} \cdot P_{\Delta,0} \cdot P_{\Delta,>0},$$
$$P_{\Delta,\mathrm{e}} := \prod_{i,j,k,l} \mathrm{e}^{-\lambda_{i,j,k,l}\Delta V_{i,j,k,l}},$$
$$P_{\Delta,0} := \prod_{n_{i,j,k,l}=0} \frac{\left(\lambda_{i,j,k,l}\Delta V_{i,j,k,l}\right)^{n_{i,j,k,l}}}{n_{i,j,k,l}!},$$

---

[10] The terms *count numbers* and *occupation numbers* will be used synonymously though slightly differing in meaning.

$$P_{\Delta,>0} := \prod_{n_{i,j,k,l}>0} \frac{(\lambda_{i,j,k,l}\Delta V_{i,j,k,l})^{n_{i,j,k,l}}}{n_{i,j,k,l}!} =: P_{\Delta,\lambda}\cdot P_{\Delta,V},$$

$$P_{\Delta,\lambda} := \prod_{n_{i,j,k,l}>0} (\lambda_{i,j,k,l})^{n_{i,j,k,l}},$$

$$P_{\Delta,V} := \prod_{n_{i,j,k,l}>0} \frac{(\Delta V_{i,j,k,l})^{n_{i,j,k,l}}}{n_{i,j,k,l}!} =: V_\Delta. \qquad (24)$$

We evaluate the above four products $P_{\Delta,e}$ to $P_{\Delta,V}$.

1. The normalization factor. Taking the logarithm and summing over all bins leads with the total number, Λ, of expected counts from (17) immediately to

$$P_{\Delta,e} = \mathrm{e}^{-\Lambda}, \qquad \Lambda := \int_{\boldsymbol{P}} \lambda(t,X,Y,E)\mathrm{d}t\mathrm{d}X\mathrm{d}Y\mathrm{d}E. \quad (25)$$

We notice that Λ is independent of the binning Δ;

2. The contribution of the void bins. By definition,

$$P_{\Delta,0} = 1, \qquad (26)$$

so that these bins do not contribute to P_Δ, irrespective of the binning;

3. The contribution due to the count rate. By hypothesis, λ is continuous. Applying the mean value theorem to the integral expressions $\lambda_{i,j,k,l}$ in (22) entails in terms of the general point (t,X,Y,E) in the photon space $\boldsymbol{P}$ and the observed photon positions

$$\lim_{(t,X,Y,E)\in \boldsymbol{P}_{i,j,k,l},\,\Delta\to 0} \lambda_{i,j,k,l} = \lambda(t,X,Y,E),$$

$$\lim_{\Delta\to 0} P_{\Delta,\lambda} = \prod_{m=1}^{M} \lambda(t_m,X_m,Y_m,E_m). \qquad (27)$$

Remind; the point $(t,X,Y,E)\in \boldsymbol{P}$ determines uniquely the bin $\boldsymbol{P}_{i,j,k,l}$ containing it;

4. The contribution due to the bin volumes. The product V_Δ depends on the occupation numbers, on the binning Δ, but not on the count rate λ. Obviously, for any positive number M of total observed counts holds

$$V_\Delta \to 0 \quad \text{for} \quad \Delta \to 0. \qquad (28)$$

In order to avoid the divergence to zero, we take V_Δ, as renormalization quantity, called in this use *occupation volume product*, and redefine the log-likelihood by

$$L_\Delta := \ln[P_\Delta/V_\Delta]. \qquad (29)$$

A probabilistic interpretation of V_Δ is provided by the relation $V_\Delta = \mathrm{e}^{V(\boldsymbol{P})}P_\Delta[1]$ where $V(\boldsymbol{P})$ is the volume of $\boldsymbol{P}$. Thus, P_Δ/V_Δ can be seen as proportional to a likelihood ratio (The hypothesis of a unit count rate $\lambda = 1$ may seem far-fetched in the light of the observed occupation numbers).

The renormalization quantity V_Δ was purposely defined so that we have with the number of expected counts Λ from (25)

$$P[\lambda] := \lim_{\Delta\to 0} \frac{P_\Delta}{V_\Delta} = \mathrm{e}^{-\Lambda}\cdot \prod_{m=1}^{M} \lambda(t_m,X_m,Y_m,E_m),$$

$$= \mathrm{e}^{-\int_{\boldsymbol{P}} \lambda(t,X,Y,E)\,\mathrm{d}t\mathrm{d}X\mathrm{d}Y\mathrm{d}E + \sum_{m=1}^{M} \ln[\lambda(t_m,X_m,Y_m,E_m)]}. \qquad (30)$$

This is the probability for the data sample $\mathcal{P}\subset \boldsymbol{P}$ to occur assigned by the hypothetical count rate λ under the applied renormalization. The log-likelihood functional, $L[\lambda]$, defined by $P[\lambda] =: \mathrm{e}^{L[\lambda]}$ is read off directly from (30),

$$L[\lambda] := -\int_{\boldsymbol{P}} \lambda(t,X,Y,E)\,\mathrm{d}t\mathrm{d}X\mathrm{d}Y\mathrm{d}E$$
$$+ \sum_{m=1}^{M} \ln[\lambda(t_m,X_m,Y_m,E_m)]. \qquad (31)$$

Interpreted as a likelihood, $\mathrm{e}^{L[\lambda]}$ is proportional[11] to the likelihood of the count rate λ to have caused the sample $\mathcal{P}$.

The *Maximum likelihood principle* directs us to choose the $L[\lambda]$ maximizing count rate, $\bar{\lambda}$, as estimate for λ. Generally, the maximizer $\bar{\lambda}$ need neither to be unique nor existent – an unhappy situation with unpleasant numerical consequences, when really given.

The $L[\lambda]$ from (31) is the general log-likelihood functional for a Poisson counting process in the paradigmatic instrument photon space $\boldsymbol{P}$ for the hypothetical count rate $\lambda(t,X,Y,E)$ from (18), (20), given the photon data sample $\mathcal{P}$ from (1).

As explained, the integral in (31) is the total (not necessarily integer) number, Λ, of the expected photons falling into the entire photon space $\boldsymbol{P}$ and thus the model counter part of M, the number of observed photon counts,

$$\Lambda := \Lambda(\boldsymbol{P}) := \int_{\boldsymbol{P}} \lambda(t,X,Y,E)\mathrm{d}t\mathrm{d}X\mathrm{d}Y\mathrm{d}E. \qquad (32)$$

Moreover, $\lambda(t_m,X_m,Y_m,E_m)$ is the hypothetical count rate evaluated at the location of the mth observed photon in the photon space $\boldsymbol{P}$.

In some cases, the model count rate depends in a special way on all parameters. Let the full model parameter vector $p =: p_1 \oplus p_2$ be split into two groups p_1, p_2 spanning $\boldsymbol{P}_1, \boldsymbol{P}_2$ such that the count rate λ factorizes,

$$\lambda(p) := \lambda_1(p_1)\cdot \lambda_2(p_2). \qquad (33)$$

For example, an X-ray source may vary rapidly in time but may possess a time-constant spectrum at the scale of the observation period.

The likelihood function $L[\lambda]$ for λ from (31) becomes in terms of the likelihood functions of the two count rate factors, $L[\lambda_1]$, $L[\lambda_2]$ with the associated expected counts $\Lambda_j := \Lambda(\boldsymbol{P}_j)$, $j = 1,2$ from (32)

$$L[\lambda_1 \cdot \lambda_2] = \Lambda_1 + \Lambda_2 - \Lambda_1 \cdot \Lambda_2 + L[\lambda_1] + L[\lambda_2]. \qquad (34)$$

The likelihood from (34) is cheaper to evaluate and alleviates the solution of the maximum likelihood equations presented in Sect. 5.

The parameterization of λ is still open. It is desirable to have the number of expected photons, Λ, among the parameters as the counter part for the number of observed photons M. This can be achieved upon the factorization

---

[11] The equality does not hold because of the division of P_Δ by the volume V_Δ.

$\lambda =: \Lambda \cdot \lambda'$. For notational convenience the prime on λ' is dropped again. In this parameterization and notation, we arrive at the representation

$$L[\lambda] := M \ln(\Lambda) - \Lambda + \sum_{m=1}^{M} \ln \left[\lambda \left(t_m, X_m, Y_m, E_m \right) \right],$$

$$1 =: \int_{\boldsymbol{P}} \lambda(t, X, Y, E) \, \mathrm{d}t \mathrm{d}X \mathrm{d}Y \mathrm{d}E. \tag{35}$$

Notice, λ in (35) is a normalized count rate distribution: one count is distributed over the entire photon space $\boldsymbol{P}$. The maximum of $L[\lambda]$ with respect to Λ is reached at the ML estimate $\hat{\Lambda} = M$ for the number of expected photons Λ. In words: given that M photons are observed, the expected number of photons is equated to the number of observed photons.

Often, if not always, the count rate is conceived as an additive mixture of several sources. In the simplest case of a binary mixture, one discerns the *source count rate* of proper interest, λ_{S}, and the remaining count rate, called then *background count rate* λ_{B}. We split the count rate λ accordingly with (unknown) mixture proportions $p_{\mathrm{B}}, p_{\mathrm{S}}$

$$\lambda := p_{\mathrm{B}} \lambda_{\mathrm{B}} + p_{\mathrm{S}} \lambda_{\mathrm{S}}, \qquad p_{\mathrm{B}} + p_{\mathrm{S}} = 1, \qquad p_{\mathrm{B}}, p_{\mathrm{S}} \geq 0. \tag{36}$$

With the splitting (36), the likelihood function for the source count rate λ_{S}, contaminated by a background source count rate λ_{B} and the expected number of photons Λ from both sources assumes the shape

$$L[\lambda] := M \ln(\Lambda) - \Lambda + \sum_{m=1}^{M} \ln[p_{\mathrm{B}} \lambda_{\mathrm{B}} \left(t_m, X_m, Y_m, E_m \right)$$
$$+ p_{\mathrm{S}} \lambda_{\mathrm{S}} \left(t_m, X_m, Y_m, E_m \right)],$$

$$1 =: \int_{\boldsymbol{P}} \lambda_{\mathrm{B}}(t, X, Y, E) \, \mathrm{d}t \mathrm{d}X \mathrm{d}Y \mathrm{d}E,$$

$$1 =: \int_{\boldsymbol{P}} \lambda_{\mathrm{S}}(t, X, Y, E) \, \mathrm{d}t \mathrm{d}X \mathrm{d}Y \mathrm{d}E. \tag{37}$$

Since background is ubiquitous, the signal–plus–noise situation (37) is fairly general and placed in this section. In the next step of model refinement, the instrument's point spread function, will be taken into account in Sect. 4.

The derivation of L for unbinned photon data is basically known, see Cash (1979) for a less elaborate but excellent treatment. Some details are worked out here.

3.3. Finite binning of the photon space and simplified likelihood evaluation

In practice, the resolution limit is given by the detector binning Δ_{D}. The limiting likelihood $L[\lambda]$ from (31), (35) for unbinned data can only be reached up to a certain binning error $\delta L[\lambda]$ which we now wish to assess.

We do this by observing first the photons with the fictitious detector of unrestricted resolution. Then the ideal observation is compared with the real observation in the finite binning Δ used in the data analysis. In the latter, only the bin centre $(t_{\mathrm{c}}, X_{\mathrm{c}}, Y_{\mathrm{c}}, E_{\mathrm{c}})$ of the bin into which

the individual photon falls is retained leading to uncertain photon positions. Hence, a finite photon binning entails unavoidably the first error of the list

(1) Photon localization error,

(2) Numerical integration error. $\tag{38}$

The second error source in (38) is tolerated in order to save program execution time. The most primitive (but cheapest) numerical integration rule is applied in order to evaluate the bin-averaged count rates $\lambda_{i,j,k,l}$ from (22). We investigate the likelihood error

$$\delta L[\lambda] := L_{\Delta}[\lambda] - L[\lambda] \tag{39}$$

with $L_{\Delta}[\lambda]$ from (29) in presence of the two error sources from (38).

Let Δ be a binning of $\boldsymbol{P}$ not finer than Δ_{D} with bin numbers I, J, K, L'. We approximate the bin-averaged count rates,

$$\lambda_{i,j,k,l} \approx \lambda(t_{\mathrm{c}}, X_{\mathrm{c}}, Y_{\mathrm{c}}, E_{\mathrm{c}}) \cdot \Delta V_{i,j,k,l}, \tag{40}$$

by bin-central count rates $\lambda(t_{\mathrm{c}}, X_{\mathrm{c}}, Y_{\mathrm{c}}, E_{\mathrm{c}})$ of the same bin. With an unknown mean position $\tilde{p} := (\tilde{t}, \tilde{X}, \tilde{Y}, \tilde{E}) \in \boldsymbol{P}_{i,j,k,l}$ for $p := (t, X, Y, E)$ and an unknown fraction $\theta \in (0, 1)$ we arrive at the exact error representation in terms of the gradient and the Hessian of the count rate λ:

$$\delta L[\lambda] := \sum_{i,j,k,l} \epsilon_{i,j,k,l} \Delta V_{i,j,k,l} + \sum_{m=1}^{M} \epsilon_m,$$

$$\epsilon_{i,j,k,l} := \frac{1}{\Delta V_{i,j,k,l}} \int_{\boldsymbol{P}_{i,j,k,l}} \lambda(t, X, Y, E) \, \mathrm{d}t \mathrm{d}X \mathrm{d}Y \mathrm{d}E$$
$$- \lambda(t_{\mathrm{c}}, X_{\mathrm{c}}, Y_{\mathrm{c}}, E_{\mathrm{c}}),$$

$$= \frac{1}{2} \cdot \sum_{\alpha, \beta = 1}^{4} \frac{\partial^2 \lambda(\theta \tilde{p})}{\partial p_\alpha \partial p_\beta} \cdot \tilde{p}_\alpha \tilde{p}_\beta,$$

$$\epsilon_m := \ln \left[\lambda(t_{\mathrm{c}}, X_{\mathrm{c}}, Y_{\mathrm{c}}, E_{\mathrm{c}}) / \lambda(t_m, X_m, Y_m, E_m) \right],$$

$$= \sum_{\alpha=1}^{4} \frac{1}{\lambda(\theta \tilde{p})} \frac{\partial \lambda(\theta \tilde{p})}{\partial p_\alpha} \cdot \tilde{p}_\alpha. \tag{41}$$

In (41), the mean value positions $\tilde{p}$ and the fractions θ are supplied by the multivariate Taylor expansion with remainder and differ in both occurrences. Associated to the mth photon is the bin index (i, j, k, l) of the bin into which it falls.

For a given binning Δ and a chosen tolerance ϵ, the third and fifth line in (41) to establish observation feasibility margins on the variability of the count rate λ in time, space, and energy in terms of first and second partial derivatives of the count rate λ.

3.3.1. The photon space for directional estimation

The replacement of bin-averaged count rates $\lambda_{i,j,k,l}$ by the right-hand side of (40), as done in EXSAS, yields an enormous gain in calculation speed. A further gain in speed

is brought about by the adaption of the photon space to the task. *Joint* light curve, spectral, and directional source estimations were possible with (31) but were not included in EXSAS. Spectral and light curve estimation is delegated to other EXSAS tasks. Therefore, the photon space is in time and detector signal amplitudes A (replacing the energy E from Sect. 3.1) trivially binned, $L' = 1$ and $K = 1$. In other words, photons of all arrival times and of all detector signal amplitudes falling into selectable intervals can be included,

$$T_1 := [\underline{T}, \bar{T}], \qquad A_1 := [\underline{A}, \bar{A}]. \tag{42}$$

For instance, the user has the choice to make soft or hard band images of an observation by choosing the margins of A_1 appropriately. This means (P1), (P2) in (16) is only required with respect to the spatial binning in sky space S.

4. Background model and point source model

The point source detection sensitivity depends on the imaging properties of the instrument and the background against which a point source is to be detected. For the ROSAT detectors, four background count sources are relevant: (a) instrumental background, (b) charged particles induced background, (c) scattered solar radiation, and (d) diffuse celestial X-ray background. The contributions from these background sources for the three ROSAT detectors are compiled in the instrument description, see Trümper (1991a) and the detector specific papers. For instance, the ROSAT PSPC detector has no instrumental background. The particle induced background contributes $2.5 \ 10^{-5}$ counts/(s arcmin^2) over 95% of the observation time and is uniformly distributed over the detector field D.

As in the general case (36), two separate types/classes of photons are distinguished, *background photons* and *source photons*. The first ones are those which are defined not to come from the observation target. The definition itself will be made in Sect. 4.1.3.

4.1. The background count rate distribution

Apart from the parameterization in (36) we assume in (18) an additive two-component-mixture of a local mean background count rate $\lambda_B \geq 0$ and a local mean source count rate component $\lambda_S \geq 0$ in the count rate vector $\lambda := (\lambda_B, \lambda_S)$,

$$\lambda(t, X, Y, A) =: \lambda_B(t, X, Y, A) + \lambda_S(t, X, Y, A). \tag{43}$$

Within this model, the estimation of λ becomes the joint estimation of both components λ_B and λ_S. This joint estimation can be carried out algorithmically by a series of repeated iterated estimation steps for one background component until the change between the last λ and the penultimate estimation λ' is small enough,

(1) choose an increment threshold $\Delta\lambda > 0$,

(2) estimate λ_B,

(3) estimate λ_S, given λ_B,

(4) repeat steps (2), (3) until $|\lambda - \lambda'| \leq \Delta\lambda$. (44)

The distance $|\cdot|$ is to be defined in the (λ_B, λ_S) quadrant. The convergence to a limit vector $\lambda := (\lambda_B, \lambda_S)$ in (44) for an increasing number of repetitions is presumed. In EXSAS, a multi-step background estimation procedure embodies (2) in (44). This is the subject of the next three subsections. Step (3) is the principal and last step of the MLE. No further corrections for the initial λ_B as indicated in (44) are calculated, i.e. the repetition (4) is not performed.

4.1.1. The background count distribution in EXSAS

According to the ROSAT photon space (15), we parameterize the three λ in (43) by amplitude A instead of energy E. The transition from count rates to counts follows. With the trivial time and amplitude binning from (42), and the fine binning (21) of the sky space we have with $C \in \{B, S\}$ the hypothetical bin counts

$$\Lambda_{C,i,j} := \int_{T_1 \times X_i \times Y_j \times A_1} \lambda_C(t, X, Y, A) \mathrm{d}t \mathrm{d}X \mathrm{d}Y \mathrm{d}A. \tag{45}$$

The aim now is to obtain estimates $\hat{\Lambda}_{B,i,j}$ for $\Lambda_{B,i,j}$. The sample $\mathcal{P}$ is subjected to the same binning (45) yielding the observed photon histogram $\mathcal{N}$ with spatial bin occupation numbers $n_{i,j}$,

$$\mathcal{N} := \left\{ n_{i,j} : (i,j) \in Z_{I,J}^2 \right\}, \quad Z_{I,J}^2 := [1, I] \times [1, J]. \tag{46}$$

Departing from $\mathcal{N}$ with equal sizes $I = J = 512$, a smooth estimate $\hat{\Lambda}_B(X, Y)$ over D with the approximate count preservation requirement

$$\Lambda_{B,i,j} \approx \frac{1}{\Delta X_i \Delta Y_j} \int_{X_i \times Y_j} \hat{\Lambda}_B(X, Y) \mathrm{d}X \mathrm{d}Y \tag{47}$$

is calculated in three steps.

4.1.2. Crude source number and location estimation

To find an initial estimate on source number and location is the first challenging task within MLE.

Point sources appear in spatial count space as local count enhancements in a form given by the PSF in effect. In a brute-force approach, source regions of a form and size compatible with the PSF are hypothesized and all their translations and dilatations in a specified search region $\mathcal{S}$ are tested at a selectable significance level to contain count variations. Testing all the hypothetical source regions guarantees that no one can be overlooked. The form of the region, however, is kept fixed.

First, specify the region of interest $\mathcal{S} \subset Z_{I,J}^2$. Small hypothetical source squares $S_{i,j}(s)$ parallel to the axes, called *windows* in some publications, of linear half-width $s \in \mathbf{N}$ around a central pixel position (i, j) together with a background pixel neighborhood $B_{i,j}(s)$ around the source

square are chosen. Denote by $T_{\boldsymbol{S},i,j}(s), T_{\boldsymbol{B},i,j}(s)$ the total exposure times for the respective suffix regions. The number of counts[12], $N_{\boldsymbol{S},i,j}(s)$, falling into $\boldsymbol{S}_{i,j}(s)$ and the number of counts, $N_{\boldsymbol{B},i,j}(s)$, falling into $\boldsymbol{B}_{i,j}(s)$ are determined,

$$N_{\boldsymbol{S},i,j}(s) := \sum_{(k,l)\in\boldsymbol{S}_{i,j}(s)} n_{k,l}, \quad (i,j)\in\mathcal{S},$$

$$N_{\boldsymbol{B},i,j}(s) := \sum_{(k,l)\in\boldsymbol{B}_{i,j}(s)} n_{k,l},$$

$$T_{\boldsymbol{S},i,j}(s) := \sum_{(k,l)\in\boldsymbol{S}_{i,j}(s)} T_{k,l},$$

$$T_{\boldsymbol{B},i,j}(s) := \sum_{(k,l)\in\boldsymbol{B}_{i,j}(s)} T_{k,l},$$

$$\boldsymbol{S}_{i,j}(s) := \left\{(k,l)\boldsymbol{Z}^2_{I,J}: |k-i|\le s, |l-j|\le s\right\},$$
$$\boldsymbol{B}_{i,j}(s) := \boldsymbol{S}_{i,j}([(1+5s)/3])\setminus\boldsymbol{S}_{i,j}(s). \quad (48)$$

In (48), $[\cdot]$ stands for rounding to the nearest integer. For $s \bmod 3 = 1$ no rounding takes place. Figure 4 shows a hypothetical source region $\boldsymbol{S}_{i,j}(1)$ surrounded by the related background region $\boldsymbol{B}_{i,j}(1)$ from (48).
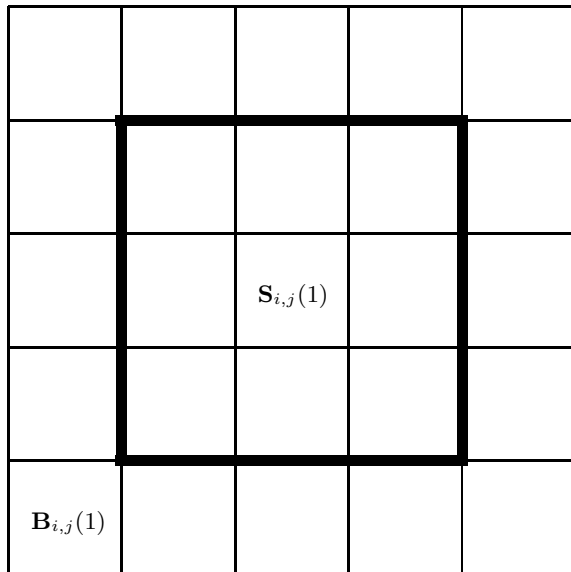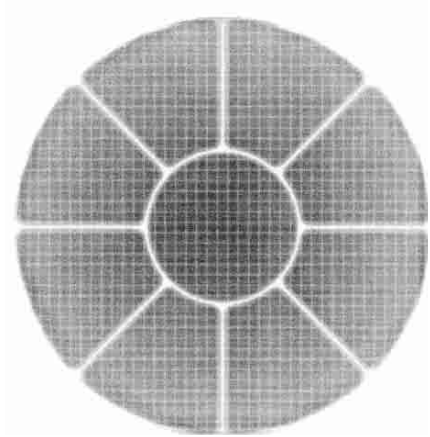
Fig. 4. The 3×3 pixel source region $\boldsymbol{S}_{i,j}(1)$ surrounded by the 16 pixel background region $\boldsymbol{B}_{i,j}(1)$. The central pixel is located at (i,j)

The exposure time histogram $\mathcal{T}$ has the binning from (46). Figure 5 shows an example of this histogram for the PSPC detector. To account for extended sources, the dilatations $s := 2^d$, $d = 0(1)D$, in (48) are investigated in a multi-grid implementation from fine to coarse. The highest dilatation exponent D can be chosen appropriately. This means: only translations (ks, ls) with natural k, l are possible[13].

Fig. 5. The 512×512 PSPC exposure time histogram $\mathcal{T}$

Does the central pixel (i,j) belong to a source region? In order to lend this question a statistically testable form, we formulate a null hypothesis in terms of the spatial count rates $\lambda_{\boldsymbol{S},i,j}, \lambda_{\boldsymbol{B},i,j}$ from (45) but without the spatial integration in the suffix regions from (48),

H_0: The spatial count rates are equal, $\lambda_{\boldsymbol{S},i,j} = \lambda_{\boldsymbol{B},i,j}$. (49)

Consider any region $\boldsymbol{Q}$ in sky space. Partition it into a hypothetical source region $\boldsymbol{S}$ and a hypothetical background region $\boldsymbol{B}$.

Let $\Lambda(\boldsymbol{S}), \Lambda(\boldsymbol{B})$ be the related mean counts from (17), (45). According to (P1), (P2) from (16), the probability $p(n_{\boldsymbol{S}} \mid n, H_0)$ to observe $n_{\boldsymbol{S}}$ counts in the region $\boldsymbol{S}$ conditional upon $n := n_{\boldsymbol{S}} + n_{\boldsymbol{B}}$ being fixed is binomial. This probability and the cumulative probability, $P(n_{\boldsymbol{S}} \mid n, H_0)$, to observe $n_{\boldsymbol{S}}$ or more source counts are given by

$$p(n_{\boldsymbol{S}} \mid n, H_0) := \binom{n}{n_{\boldsymbol{S}}} p_{\boldsymbol{S}}^{n_{\boldsymbol{S}}} \cdot p_{\boldsymbol{B}}^{n-n_{\boldsymbol{S}}},$$

$$P(n_{\boldsymbol{S}} \mid n, H_0) := \sum_{k=n_{\boldsymbol{S}}}^{n} p(k \mid n, H_0),$$

$$p_{\boldsymbol{S}} := \frac{\Lambda(\boldsymbol{S})}{\Lambda(\boldsymbol{S}) + \Lambda(\boldsymbol{B})},$$

$$p_{\boldsymbol{B}} := \frac{\Lambda(\boldsymbol{B})}{\Lambda(\boldsymbol{S}) + \Lambda(\boldsymbol{B})}. \quad (50)$$

With (17), the probabilities from (50) applied to the pair of regions $\boldsymbol{S} := \boldsymbol{S}_{i,j}(s)$, $\boldsymbol{B} := \boldsymbol{B}_{i,j}(s)$ become when dropping three arguments for notational convenience,

$$P_{\boldsymbol{S},i,j}(s) := P(N_{\boldsymbol{S},i,j}(s) \mid N_{i,j}(s), H_0)_{|S=S_{i,j}, B=B_{i,j}},$$

$$p_{\boldsymbol{S},i,j}(s) := \frac{T_{\boldsymbol{S},i,j}(s)}{T_{i,j}(s)}, \quad N_{i,j}(s) := N_{\boldsymbol{S},i,j}(s) + N_{\boldsymbol{B},i,j}(s),$$

$$p_{\boldsymbol{B},i,j}(s) := \frac{T_{\boldsymbol{B},i,j}(s)}{T_{i,j}(s)}, \quad T_{i,j}(s) := T_{\boldsymbol{S},i,j}(s) + T_{\boldsymbol{B},i,j}(s).$$

$$(51)$$

---

[12] Double subscripts are avoided for readability.
[13] Of practical significance is that for constant exposure times the ratio $T_{\boldsymbol{S},i,j}(s) : T_{\boldsymbol{B},i,j}(s) = 9 : 16$ is maintained. This

permits the tabulation of the likelihoods $L_{i,j}$ from (56). For ratios different from $16 : 9$, an ad-hoc correction is made.

We select an appropriate error probability $P_0 := \exp[-L_0]$ parameterized by[14] $L_0 > 0$. If the counts $N_{\boldsymbol{S},i,j}$ are large enough so that the probability, $P_{\boldsymbol{S},i,j}(s)$, to observe $N_{\boldsymbol{S},i,j}$ or more counts in the source square $\boldsymbol{S}_{i,j}(s)$ satisfies

$$P_{\boldsymbol{S},i,j}(s) \leq P_0 \tag{52}$$

then the central pixel (i,j) is loosely referred to as source pixel at *detection likelihood* L_0. This is the technical term for the quantitative degree of source existence in the final source list used in many ROSAT publications, see Voges et al. (1999) for instance. The likelihood sequence increases strictly with the number $n_{\boldsymbol{S}}$ of observed source counts and decreases with increasing source probability $p_{\boldsymbol{S}}$,

$$0 := L_0 < L_1 < ... < L_{n-1} < L_n := -n\ln(p_{\boldsymbol{S}}),$$
$$L_k := -\ln[P(k \mid n, H_0)], \qquad 0 \leq k \leq n. \tag{53}$$

Figure 6 shows for the source probabilities $p_{\boldsymbol{S}} = 0.1(0.1)0.9$ the likelihood sequences for $n = 1000$ total counts. For less source counts than expected, i.e. for $n_{\boldsymbol{S}} < np_{\boldsymbol{S}}$, the likelihoods L_k stay about in the interval $[0, \ln(2)]$. Of practical interest is the likelihood interval of say $[5, 20]$.
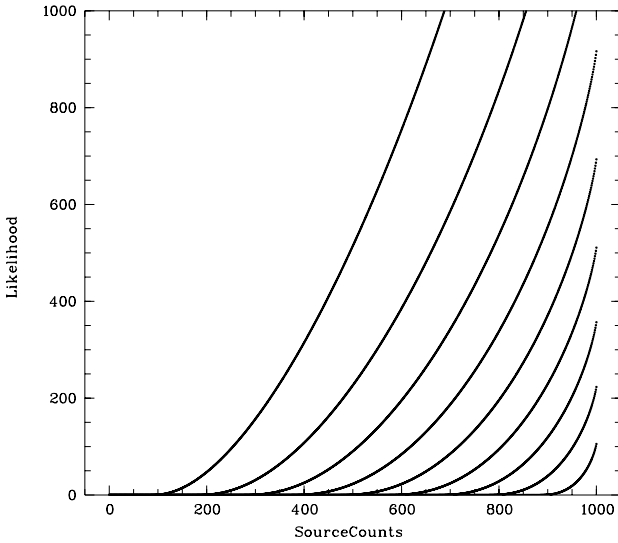
Fig. 6. The likelihood sequences L_k for $p_{\boldsymbol{S}} = 0.1(0.1)0.9$ (from left to right) for $n = 1000$ total counts in the likelihood interval $[0, 1000]$

The evaluation of the binomial cumulative probability $P(k, \mid n, H_0)$ from (53) becomes impractical for large count numbers n and k. The univariate version of the Central Limit Theorem supplies us with the classical asymptotical approximation for $n \to \infty$ (with the notorious slow convergence speed $1/\sqrt{n}$) in terms of the univariate normal density. Under the change $t^2 \to t$ of the independent variable t, the approximation can be rewritten in terms of the density $\chi_1(t)$, of the χ^2 distribution with one degree of freedom. This approximation is

[14] The standard choice is $L_0 = 8$, i.e. $P_0 := 0.0003$.

used in the central region $0 \leq n_{\boldsymbol{S}} - np_{\boldsymbol{S}} \leq 3\sqrt{np_{\boldsymbol{S}}p_{\boldsymbol{B}}}$. Additionally, a lower bound sequence $\underline{L}_k$ of excellent approximation quality is used. Figure 7 gives the error sequence $\Delta L_k := L_k - l_k$ with l_k being the said approximation.

For high counts in $\mathcal{C}_{\mathrm{H}}$ (see (57)), the likelihoods $L_{i,j}$ are approximated by the tail probability of the χ_1^2 distribution with one degree of freedom,

$$L_{i,j} := -\ln\left[\int_{X_{i,j}^2}^{\infty} \chi_1(t)\mathrm{d}t\right], \quad \chi_1(t) := \frac{\mathrm{e}^{-t/2}}{\sqrt{2\pi t}},$$
$$X_{i,j} := \frac{N_{\boldsymbol{S},i,j}T_{i,j} - N_{i,j}T_{\boldsymbol{S},i,j}}{\sqrt{N_{i,j}T_{\boldsymbol{S},i,j}T_{\boldsymbol{B},i,j}}} \geq 0. \tag{54}$$

The function $L(X^2) := -\ln\left[\int_{X_{i,j}^2}^{\infty} \chi_1(t)\mathrm{d}t\right]$ is tabulated in EXSAS for $X^2 = 0(0.1)X_0^2$. Beyond the upper bound $X_0^2 := 39.9$, the lower bound $L_1(X^2)$ in

$$L_1\left(X^2\right) \leq L\left(X^2\right) \leq L_2\left(X^2\right), \quad X^2 > 0,$$
$$L_1(X^2) := \frac{1}{2}\left(X^2 + \ln\left(\frac{\pi}{2}X^2\right)\right),$$
$$L_2(X^2) := \frac{1}{2}\left(X^2 + \ln\left(\frac{\pi}{2}\left(X + \frac{1}{X}\right)^2\right)\right). \tag{55}$$

is used. The error is for $X^2 \geq 40$ less than 0.0007.
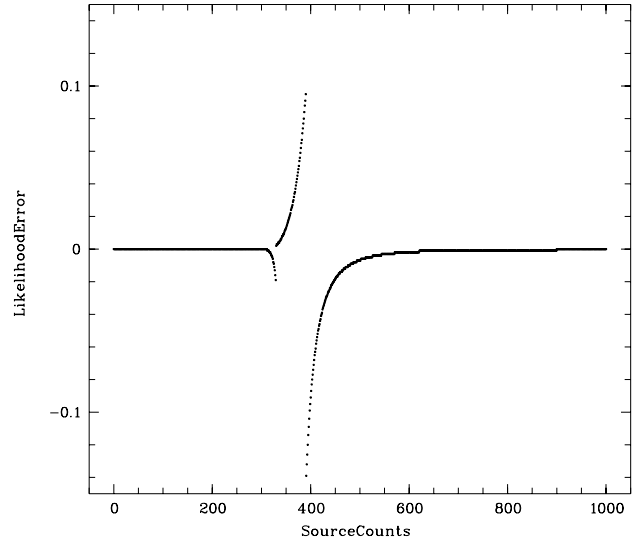
Fig. 7. The likelihood error sequence ΔL_k for $p_{\boldsymbol{S}} = 0.36$ and $n = 1000$ counts

Each pixel position (i,j) satisfying (52) becomes a point of the support of the (first) likelihood histogram

$$\mathcal{L}_1(L_0) := \{\lambda_{i,j}: \ (i,j) \in \mathcal{S}\},$$
$$\lambda_{i,j} := \begin{cases} L_{i,j}, & L_{i,j} \geq L_0, \\ 0, & L_{i,j} < L_0, \end{cases}$$
$$L_{i,j} := -\ln\left[P_{\boldsymbol{B},i,j}(s)\right]. \tag{56}$$

In order to adjust the approximation error properly later, select $N_0 \in \boldsymbol{N}$ large enough. The observed histogram

count space $\mathcal{C}$ is partitioned into a low count region $\mathcal{C}_{\mathrm{L}}$ and the complementary high count region $\mathcal{C}_{\mathrm{H}}$ by

$$
\begin{aligned}
\mathcal{C}_{\mathrm{L}} &:= \{(n_{\boldsymbol{S}}, n_{\boldsymbol{B}}) \in \mathcal{C}: \ 0 \leq n_{\boldsymbol{S}} + n_{\boldsymbol{B}} < N_0\}, \\
\mathcal{C}_{\mathrm{H}} &:= \{(n_{\boldsymbol{S}}, n_{\boldsymbol{B}}) \in \mathcal{C}: \ n_{\boldsymbol{S}} + n_{\boldsymbol{B}} \geq N_0\}, \\
\mathcal{C} &:= \{(n_{\boldsymbol{S}}, n_{\boldsymbol{B}}) \boldsymbol{Z}^2: \ n_{\boldsymbol{S}}, n_{\boldsymbol{B}} \geq 0\}.
\end{aligned}
\tag{57}
$$

In EXSAS, $N_0 = 100$ was chosen. In $\mathcal{C}_{\mathrm{L}}$, the likelihoods $L_{i,j}$ from (56) are evaluated. The histogram $\mathcal{L}_1(L_0)$ is algorithmically decomposed into a union of connected components $\mathcal{L}_{\mathrm{s}}(L_0)$. For each component, the maximizer pixel position $(i_{\mathrm{s}}, j_{\mathrm{s}})$ is found. The center of gravity in a small neighborhood, $\mathcal{L}'_{\mathrm{s}}(L_0)$, of the maximizer is the estimate of the source position,

$$
\begin{aligned}
\mathcal{L}_1(L_0) &:= \cup_{s=1}^{S_1} \mathcal{L}_{1,\mathrm{s}}(L_0), \\
(i_{\mathrm{s}}, j_{\mathrm{s}}) &:= \arg\left[\max_{(i,j) \in \mathcal{L}_{1,\mathrm{s}}(L_0)} \lambda_{i,j}\right], \\
X_{\mathrm{s}} &:= \frac{1}{\lambda'_{\mathrm{s}}} \sum_{(i,j) \in \mathcal{L}'_{1,\mathrm{s}}} i\lambda_{i,j}, \quad Y_{\mathrm{s}} := \frac{1}{\lambda'_{\mathrm{s}}} \sum_{(i,j) \in \mathcal{L}'_{1,\mathrm{s}}} j\lambda_{i,j}, \\
\lambda'_{\mathrm{s}} &:= \sum_{(i,j) \in \mathcal{L}'_{1,\mathrm{s}}} \lambda_{i,j}.
\end{aligned}
\tag{58}
$$

This way, a first source list, $\mathcal{S}_1(L_0)$, of S_1 source candidate positions in sky space at detection likelihood L_0 is found,

$$
\mathcal{S}_1(L_0) := \{(X_{\mathrm{s}}, Y_{\mathrm{s}})\}_{s=1}^{S_1}.
\tag{59}
$$

Besides positions and detection likelihoods $L_{i,j}$, 13 additional source characteristics are incorporated in this table – the intermediate result after the first step. See Voges et al. (1999) for a specimen of this list.

4.1.3. Background domain definition

In the second step, the background estimate $\hat{\Lambda}_{\mathrm{B}}(X, Y)$ is found. $S := S_1$ circular disks around the source positions with radii R_{s} and a union $\boldsymbol{U}$ thereof serve to define the background domain $\boldsymbol{B}$,

$$
\begin{aligned}
\boldsymbol{D}(X_{\mathrm{s}}, Y_{\mathrm{s}}, R_{\mathrm{s}}) &:= \left\{(X - X_{\mathrm{s}})^2 + (Y - Y_{\mathrm{s}})^2 \leq R_{\mathrm{s}}^2\right\}, \\
R_s &:= \rho FWHM(\epsilon, A), \\
\boldsymbol{U} &:= \cup_{s=1}^{S} \boldsymbol{D}(X_{\mathrm{s}}, Y_{\mathrm{s}}, R_{\mathrm{s}}), \\
\boldsymbol{B} &:= \boldsymbol{X} \times \boldsymbol{Y} \setminus \boldsymbol{U}.
\end{aligned}
\tag{60}
$$

For the radii $\rho := 2.5$ is standardly chosen. The *FWHM* function, explained in Sect. 4.2, is evaluated at the disk centre off-axis angle ϵ and photon amplitude A. The photons falling within the background domain $\boldsymbol{B}$ constitute the subsample $\mathcal{P}_0$ and are defined to be background photons. Clearly, the error probabilities that a source photon stems from the background and vice versa are expected highest near the disk boundaries. A user-defined trade-off is possible by adjusting the radii R_{s}.

Figure 8 shows an example for the background photons (black dots) corresponding to $\mathcal{P}_0$, introduced in (2) in binned form $\mathcal{N}_0$.
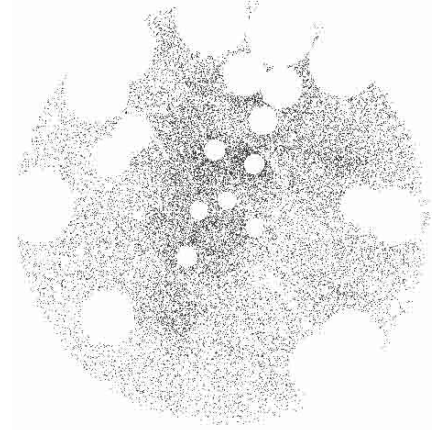
Fig. 8. Censored 512×512 background. Disks around source positions are cut out

4.1.4. Background count distribution

A bivariate least square cubic spline surface $\hat{\Lambda}_{\mathrm{B}}(X, Y)$ over $\boldsymbol{D}$ solving the optimization problem,

$$
\sum_{(i,j) \in \boldsymbol{B}} w_{i,j} \left[\Lambda_B(X_i, Y_j) - n_{i,j}\right]^2 \to \min, \quad w_{i,j} := T_{i,j}
\tag{61}
$$

is fitted to the observed count histogram $\mathcal{N}$ restricted to the background domain $\boldsymbol{B}$. This is the proper estimation step. The pixel centres (X_i, Y_j) used in (61) are rarefied, a coarser binning is applied. This is adequate for the background component postulated to exhibit no abrupt variations.

The spline surface $\hat{\Lambda}_{\mathrm{B}}(X, Y)$ is the adopted count distribution estimate over the sky field. So, we assume the $\Lambda_{\mathrm{B},i,j}$ in (45) known.

4.1.5. Refined source number and location estimation

Knowing the background estimate $\hat{\Lambda}_{\mathrm{B}}(X, Y)$, the source determination can be redone. This is the third step. From $\hat{\Lambda}_{\mathrm{B}}(X, Y)$ the background histogram $\mathcal{N}_{\mathrm{B}}$ of the same format as $\mathcal{N}$ from (46) is formed. The background regions and background counts are redefined,

$$
\begin{aligned}
\mathcal{N}_{\mathrm{B}} &:= \left\{n_{\boldsymbol{B},i,j}: \ (i,j) \boldsymbol{Z}_{I,J}^2\right\}, \\
N_{\boldsymbol{B},i,j}(s) &:= \sum_{(k,l) \in \boldsymbol{B}_{i,j}(s)} n_{\boldsymbol{B},k,l}, \\
\boldsymbol{B}_{i,j}(s) &:= \boldsymbol{S}_{i,j}(s).
\end{aligned}
\tag{62}
$$

With the remaining quantities from (48), the source position estimation as described in Sect. 4.1.2 is repeated. The outcome is a second list, $\mathcal{L}_2(L_0)$, of additional S_2 source candidate positions in sky space. The second list is merged with the first list $\mathcal{L}_1(L_0)$ from (58) into a list $\mathcal{L}(L_0)$ of $S := S_1 + S_2$ sources in total

$$
\begin{aligned}
\mathcal{L}_2(L_0) &:= \{(X_{\mathrm{s}}, Y_{\mathrm{s}})\}_{s=1}^{S_2}, \\
\mathcal{L}(L_0) &:= \mathcal{L}_1(L_0) \cup \mathcal{L}_2(L_0).
\end{aligned}
\tag{63}
$$

With finally S source positions from (63), the corresponding disks $\boldsymbol{D}$ from (59) are determined. The photons in

these disks constitute the subsamples $\mathcal{P}_s$, $s = 1, ..., S$, mentioned in (15); they are taken as source photons – the input data for MLE in the narrow sense.

4.2. The point source model

The image of a celestial *point source in a fixed celestial direction at infinite distance*, is, by definition, the *point spread function* (PSF). In the case of the ROSAT detectors, the PSFs are parameterized by the (a) *off-axis angle* ϵ and (b) the photon detector amplitude A. The off-axis angle ϵ is the angle which subtends the PSF centre with the instrument's optical axis.

As a compromise between practicability (in terms of computing time) of the resulting MLE procedure and model precision, a bivariate normal circular symmetric distribution density, p, defined over the whole sky space $\boldsymbol{S}$ is used to represent the measured point spread function in EXSAS for the MLE of single X-ray point source parameters. In statistical interpretation, the probability dP to find a photon in sky space in a box around the central point (X, Y) with infinitesimal side lengths dX and dY when a photon arrived at the point (x, y) in the detector plane causing a detector amplitude A is $dP := p(X, Y \mid X_S, Y_S, \nu_S, A)\, dX dY$. The spread parameter ν_S appearing in (64) will be explained in the sequel.

More precise point spread function models of the ROSAT instruments are used in EXSAS for other analysis purposes, see Boese (2000). For instance, the distribution tail in the measured PSF is higher than that of (64). The celestial *local mean source count rate* λ_S^c in (64) entails the count rate λ_S in sky space

$$
\begin{aligned}
\lambda_S(t, X, Y, A) &:= \lambda_S^c(t, X_S, Y_S, A) \cdot V[A, \epsilon(x,y)] \\
&\quad \times p(X, Y \mid X_S, Y_S, \nu_S, A), \\
p(X, Y \mid X_S, Y_S, \nu_S, A) &:= \frac{1}{2\pi\nu(x, y, A)} \\
&\quad \times e^{-[(X-X_S)^2+(Y-Y_S)^2]/\nu(x,y,A)}, \\
\nu(x, y, A) &:= 2\sigma^2[A, \epsilon(x,y)] + \nu_S, \quad \nu_S \geq 0, \\
\epsilon(x, y) &:= \sqrt{x^2 + y^2}.
\end{aligned} \tag{64}
$$

Notice the joint presence of sky coordinates (X, Y) and detector coordinates (x, y) in (64). The dependence of p on the detector coordinates is not denoted. The vignetting factor V will be explained in the next subsection.

To account for extended sources, the spread parameter $\nu(x, y, A)$, in fact a double variance parameter, containing additively the nonnegative *source extent parameter* $\nu_S \geq 0$, has been introduced. Though the source parameter ν_S does not estimate the full morphology[15] of extended sources, it offers the *azimuthally averaged part* of the morphology as a parameter of source extendedness. In operational terms, a radial two-dimensional Gaussian photon distribution is fitted to an extended source photon distribution.

---

[15] This notion means here: the family of contours around an extended source at all count rate levels.

4.3. The vignetting

Apart from the form of the distribution density $p(X, Y \mid X_S, Y_S, \nu_S, A)$, two further functions characterize the instrument. One of them is the instrument's *vignetting function*, $V[A, \epsilon]$, fulfilling

$$
0 \leq V[A, \epsilon] \leq V[A, 0] := 1. \tag{65}
$$

The vignetting function describes the instrument's relative collection efficiency as a function of photon detector amplitude A and off-axis angle ϵ. The reference value is the value of highest efficiency, found in on-axis position, see (65). In a stochastical interpretation, it describes the loss of photon counts with increasing angular off-axis distance $\epsilon(x, y)$ from (64) for a given photon detector amplitude A.

A prominent property of the Poisson process is that it is preserved under random sampling. We view the (non-)vignetting as a realization of random sampling, and the infalling local mean source count rate $\lambda(t, X, Y, E)$ diminishes under vignetting from the on-axis maximal count rate $\lambda^0(t, X, Y, A)$ to the lower *vignetted count rate* $\lambda^V(t, X, Y, A)$,

$$
\lambda_S^V(t, X, Y, A) := \lambda_S^0(t, X, Y, A) \cdot V[A, \epsilon(x,y)]. \tag{66}
$$

The second function, $\sigma^2(A, \epsilon) > 0$, appearing in (64) is the *variance* of distribution density p. It characterizes the spread of the point spread function in dependence of the source photon amplitude A and the off-axis angle ϵ. More precisely, $r_{1/2} := \sigma(A, \epsilon) \cdot \sqrt{\ln(4)} = \sigma(A, \epsilon) \cdot 1.177...$ is the *median radius*, i.e. the radius of the circular disk around the maximizer (X_S, Y_S) of p in which 50% of all photons fall. Into the larger disk with radius 3σ and the same centre fall 98.9% of all photons.

According to the source model (64), we have four single–source parameters $X_S, Y_S, \lambda_S^V, \nu_S$, representing in turn: the source position (X_S, Y_S), vignetted source count rate, and source extent parameter. Depending on the context, we regard either λ_S^V or λ_S^0 from (66) as the basic count rate parameter. Only the first one is observable, and the one of eventuel physical interest is λ_S^c.

4.4. The time-dependent point spread function

The optical axis moves across the Sky in time. Correspondingly, a celestial X-ray source in a fixed celestial direction traces in the detector plane a source trajectory $\mathcal{D}$ during the observation interval $\boldsymbol{T}$. The motion $\mathcal{D}$ in time-parameterized form and parameterized by the off-axis-angle ϵ, viewed as a random variate with distribution $T(\epsilon)$, $0 \leq \epsilon \leq \bar{\epsilon}$, for the individual trajectory $\mathcal{D}$ are

$$
\begin{aligned}
\mathcal{D}&: \ (x_S(t), y_S(t)) \in \boldsymbol{F}, \qquad t \in \boldsymbol{T}, \\
\mathcal{D}&: \ T(\epsilon) := |\{t \in \boldsymbol{T}: \ x_S^2(t) + y_S^2(t) \leq \epsilon^2\}|/T.
\end{aligned} \tag{67}
$$

By definition (67), $T(\epsilon)$ is the fraction of the observation time T the trajectory $\mathcal{D}$ stays in or on the circle of radius ϵ. The upper limit $\bar{\epsilon}$ of the range of ϵ is the radius of the

field of view $\boldsymbol{F}$. Let $t(\epsilon) := T'(\epsilon)$ be the density of the distribution $T(\epsilon)$.

As mentioned in Sect. 2.3.3, in pointing mode, the motion (67) is nominally constant apart from a standardly applied deliberate small wobbling motion of the optical axis and superimposed by a still smaller stochastic motion due to stochastic forces. So, $t(\epsilon)$ is expected to be unimodal with a mode position near $(x_{\mathrm{S}}, y_{\mathrm{S}})$, the source position $(X_{\mathrm{S}}, Y_{\mathrm{S}})$ in detector coordinates. In survey mode, the source under consideration executes (under the ROSAT survey scanning scheme) a series of neighbouring parallel tracks in the field of view $\boldsymbol{F}$. Hence, a density $t(\epsilon)$ close to the uniform density $t_{\mathrm{u}}(\epsilon) := 1/|\boldsymbol{F}|$ is expected. To save execution time within MLE in EXSAS, no off-axis histogram approximating $t(\epsilon)$ has been constructed and used. Equally, the individual density $t(\epsilon)$ was not explicitly replaced by more general and thus less close approximations to $t(\epsilon)$. The underlying rationale will be discussed.

Substituting the decomposition (43) into (31) gives the likelihood functional L from (31) in the parameterized form

$$L(X_{\mathrm{S}}, Y_{\mathrm{S}}, \lambda_{\mathrm{S}}, \nu_{\mathrm{S}}) := -\Lambda_{\mathrm{B}} - \Lambda_{\mathrm{S}}$$
$$+ \sum_{m=1}^{M} \ln[\lambda_{\mathrm{B}}(t_m, X_m, Y_m, A_m)$$
$$+ \lambda_{\mathrm{S}}(t_m, X_m, Y_m, A_m)],$$
$$\Lambda_{\mathrm{B}} := \int_{\boldsymbol{P}} \lambda_{\mathrm{B}}(t, X, Y, A)\, \mathrm{d}t \mathrm{d}X \mathrm{d}Y \mathrm{d}A,$$
$$\Lambda_{\mathrm{S}} := \int_{\boldsymbol{P}} \lambda_{\mathrm{S}}(t, X, Y, A)\, \mathrm{d}t \mathrm{d}X \mathrm{d}Y \mathrm{d}A. \tag{68}$$

In (68), the detector coordinates $(x, y) := (x_{\mathrm{S}}(t), y_{\mathrm{S}}(t))$ are taken along the source trajectory $\mathcal{D}$ from (67). This leads to a time-dependent off-axis angle $\epsilon(t) := \sqrt{x_{\mathrm{S}}^2(t) + y_{\mathrm{S}}^2(t)}$ and thus to a time-dependent PSF density p in (64) explaining the first argument t in $\lambda_{\mathrm{S}}(t, X, Y, A)$. The integration with respect to time t in (68) can be replaced by the integration with respect to the off-axis angle, $\mathrm{d}t = t(\epsilon)\mathrm{d}\epsilon$ over its range $[0, \bar{\epsilon}]$.

Conceptually, the number Λ_{B} of expected background photons in (68) is constant with respect to all four source parameters; to source position $(X_{\mathrm{S}}, Y_{\mathrm{S}})$, mean local source count rate λ_{S}, and to the source extent parameter $\nu_{\mathrm{S}} \geq 0$. This is also in practice so, save for cases in which the background region $\boldsymbol{B}$ from (60) is small compared to the field of view $\boldsymbol{F}$. The (real) number Λ_{S} of expected source photons in (68) depends in source positions near the boundary $\partial \boldsymbol{F}$ of the field of view $\boldsymbol{F}$ remarkably on the source parameters – as the following arguments clarify.

Assuming a continuous local mean count rate λ_{S} with respect to all four arguments, the mean value theorem applied to the integral Λ_{S} gives with an average mean local source count rate $\tilde{\lambda}_{\mathrm{S}}$ (to be determined) and a mean off-axis angle $\tilde{\epsilon}$ (to be determined) the exact representation

$$\Lambda_{\mathrm{S}} =: \tilde{\lambda}_{\mathrm{S}} \cdot |\boldsymbol{F}| \cdot \Delta A \cdot T, \qquad \epsilon = \tilde{\epsilon}. \tag{69}$$

In (69), $|\boldsymbol{F}|$ is the area of the field of view and $\Delta A := \bar{A} - \underline{A}$ the width of the amplitude space used. In the case of strong sources, both T and ΔA, can be minimized by selecting photons with arrival times and amplitudes close together with the consequence $\tilde{\lambda}_{\mathrm{S}} \approx \lambda_{\mathrm{S}}$. For weak sources, an averaging in (68) is in effect.

Under the trival binning (42) with respect to time and amplitude, no resolution with respect to these variables is achieved and only the *time- and amplitude-averaged parts* $\lambda_{\mathrm{S}}(X, Y)$ of a source count distribution $\lambda_{\mathrm{S}}(t, X, Y, A)$ can be estimated – a limitation without detrimental consequences when one aims at celestial directions and fluxes only.

We reverse the argument when we state: for

(S1) stationary sources, $\lambda_{\mathrm{S}}(t, X, Y, A) := \lambda_{\mathrm{S}}(X, Y, A)$,

(S2) sources with uniformly distributed amplitudes,
      $\lambda_{\mathrm{S}}(t, X, Y, A) := \lambda_{\mathrm{S}}(t, X, Y)$,

(S3) celestial point sources of (69) with λ_{S} from (64),
$$\tag{70}$$

Eq. (69) holds with the spatially averaged count rate

$$\tilde{\lambda}_{\mathrm{S}} := \frac{1}{|\boldsymbol{F}|} \int_{\boldsymbol{F}} \lambda_{\mathrm{S}}(X, Y)\mathrm{d}X\mathrm{d}Y. \tag{71}$$

The constancy of λ_{S} in (S1), (S2) with respect to time t and amplitude A leads immediately to $\tilde{\lambda}_{\mathrm{S}}$ in (71). The normalized form (35) is less favourable in cases when several components of the total count rate λ in (43) are discerned.

For a source possessing all three properties (S1) to (S3), i.e. for *stationary amplitude-uniform point sources*, one arrives with an unknown mean off-axis angle $\tilde{\epsilon}$ and a constant on-axis source count rate λ_{S}^0 at

$$\Lambda_{\mathrm{S}} := \lambda_{\mathrm{S}}^0 \cdot T \int_{\boldsymbol{F}} V[A, \tilde{\epsilon}] \cdot p(X, Y \mid X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}}, A)\, \mathrm{d}X\mathrm{d}Y,$$
$$= \lambda_{\mathrm{S}}^0 \cdot T \cdot V[A, \tilde{\epsilon}] \cdot [1 - \omega_{\boldsymbol{F}}(X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}}, A)]. \tag{72}$$

The approximation $\tilde{\epsilon} := \epsilon(x_{\mathrm{S}}, y_{\mathrm{S}})$ in (72) is deemed to be reasonably accurate in both observation modes, survey and pointing.

It is justified to ignore censoring by the field of view $\boldsymbol{F}$ when the source position $(X_{\mathrm{S}}, Y_{\mathrm{S}})$ lies in the central part of the field of view $\boldsymbol{F}$. The fraction of photons not falling into the field of view,

$$\omega_{\boldsymbol{F}}(X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}}, A) :=$$
$$\int_{\boldsymbol{R}^2 \backslash \boldsymbol{F}} p(X, Y \mid X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}}, A)\, \mathrm{d}X\mathrm{d}Y \approx 0 \tag{73}$$

is disregarded. Or, seen as a requirement for the point spread function and source positions, source positions must be bounded away from the boundary $\partial \boldsymbol{F}$, say farther than 3σ from the boundary $\partial \boldsymbol{F}$ of the field of view.

Under the approximations (69), (73) for stationary amplitude-uniform point sources, it follows the

log-likelihood representation

$$L(X_S, Y_S, \lambda_S, \nu_S) := -\Lambda_B - \lambda_S \cdot T \cdot V[A, \epsilon(x_S, y_S)]$$

$$+ \sum_{m=1}^{M} \ln[\lambda_B(t_m, X_m, Y_m, A_m)$$

$$+ \lambda_S \cdot V[A_m, \epsilon(x_m, y_m)] \cdot p(X_m, Y_m \mid X_S, Y_S, \nu_S, A)]. (74)$$

In (74), the on-axis source count rate λ_S^0 from (66) is again denoted by λ_S for notational convenience[16]. In a more parsimonious notation and evaluating the vignetting factor at a mean amplitude $\tilde{A}$, (74) reads in terms of a vignetted observation time T^V

$$L(X_S, Y_S, \lambda_S, \nu_S) := -\Lambda_B - \lambda_S T^V + \sum_{m=1}^{M} \ln[\lambda_{B,m} + \lambda_S$$

$$\times V[A_m, \epsilon(x_m, y_m)] \cdot p_m(X_S, Y_S, \nu_S)],$$

$$T^V := T \cdot V[\tilde{A}, \tilde{\epsilon}],$$

$$\lambda_{B,m} := \lambda_B(t_m, X_m, Y_m, A_m),$$

$$p_m(X_S, Y_S, \nu_S) := p^*(X_m, Y_m \mid X_S, Y_S, \nu_S, A_m). \quad (75)$$

In (75), p^* means that p is to be evaluated at the off-axis angle $\epsilon(x_m, y_m)$ and not at $\epsilon(x_S, y_S)$ as demanded by (64). Under this convention, ν is a function of ν_S but not of X_S, Y_S.

The mean off-axis angle $\tilde{\epsilon}$ is taken constant with respect to all four source parameters. All that follows will be based on this form of the log-likelihood function L. The form of L in (75) or (74) is the one used in EXSAS.

5. The maximum likelihood equations

R. A. Fisher coined in 1921 the term "likelihood" and in 1922 the "maximum likelihood estimate". Readers interested in the historical development are referred to Aldrich (1977), Edwards (1997), Hald (1999), and the literature cited therein.

Having specified the distributional form of the underlying source count distribution density in (75) up to the source parameter vector

$$\boldsymbol{s} := (X_S, Y_S, \lambda_S, \nu_S) =: (s_1, s_2, s_3, s_4), \quad (76)$$

it remains to achieve an estimate $\hat{\boldsymbol{s}} := \left(\hat{X}_S, \hat{Y}_S, \hat{\lambda}_S, \hat{\nu}_S\right)$ for the unknown parent counterpart $\boldsymbol{s}$ from (76) given the photon time series $\mathcal{P}$ from (1). As already said, for the purposes of this section, we assume a stationary monoenergetic point source.

Among several point estimation procedures, we chose estimation via the *Maximum Likelihood Principle*. The latter directs us to choose $\hat{\boldsymbol{s}}$ so that with L from (75) holds

$$L\left(\hat{X}_S, \hat{Y}_S, \hat{\lambda}_S, \hat{\nu}_S\right) = \max_{\boldsymbol{s} \in \boldsymbol{S}} L(X_S, Y_S, \lambda_S, \nu_S). \quad (77)$$

The *source parameter space* $\boldsymbol{S}$ in (77) is defined by

$$\boldsymbol{S} := \{(X_S, Y_S, \lambda_S, \nu_S) \boldsymbol{R}^4 : (X_S, Y_S) \in \boldsymbol{F}, \lambda_S \geq 0,$$

$$\nu_S \geq 0\}. \quad (78)$$

[16] The symbol λ_S^0 will be redefined in (109).

Two cases are possible: either, the maximizer of L lies in the interior of $\boldsymbol{S}$ or on its boundary $\partial\boldsymbol{S}$. The first case is viewed as the generic one and assumed further on.

In view of the differentiability of L with respect to $\boldsymbol{s}$, the equations for $\hat{\boldsymbol{s}}$

$$\frac{\partial L(\hat{\boldsymbol{s}})}{\partial X_S} = 0, \quad \frac{\partial L(\hat{\boldsymbol{s}})}{\partial Y_S} = 0, \quad \frac{\partial L(\hat{\boldsymbol{s}})}{\partial \lambda_S} = 0, \quad \frac{\partial L(\hat{\boldsymbol{s}})}{\partial \nu_S} = 0,$$

$$H(\hat{\boldsymbol{s}}) < 0. \quad (79)$$

are necessary. When the estimate $\hat{\boldsymbol{s}}$ satisfies (79), then $\boldsymbol{s}$ in (77) runs only through all zeros of the gradient vector $\partial L/\partial\boldsymbol{s}$ with components as given in the first line of (79) and obeying the negative definiteness in (79). The Hessian 4×4 matrix $H(\boldsymbol{s})$ of the log-likelihood function L in (79) has the elements

$$H_{j,k}(\boldsymbol{s}) := \frac{\partial^2 L(\boldsymbol{s})}{\partial s_j \partial s_k}, \qquad j, k = 1, 2, 3, 4. \quad (80)$$

As usual, $H(\hat{\boldsymbol{s}}) < 0$ means that $H(\boldsymbol{s})$ is negative definite at the point $\boldsymbol{s} := \hat{\boldsymbol{s}}$ in source parameter space $\boldsymbol{S}$. The four equations along with the negative definiteness in (79) will be called *ML equations*.

By Sylvester's famous algebraic criterion, we know that a real 4×4 matrix is negative definite if and only if all its four principal minors d_1 to d_4 are negative,

$$d_j(\hat{s}) < 0,$$

$$d_j(\hat{s}) := \det\left[H^{(j)}(\hat{s})\right], \qquad j = 1, 2, 3, 4. \quad (81)$$

In (81), $H^{(j)}(\hat{s})$ is the principal matrix of order j of $H(\boldsymbol{s})$ evaluated at the maximizer $\boldsymbol{s} = \hat{\boldsymbol{s}}$,

$$d_1 := H_{1,1},$$

$$d_2 := \det\begin{bmatrix} H_{1,1}, & H_{1,2} \\ H_{2,1}, & H_{2,2} \end{bmatrix},$$

$$d_3 := \det\begin{bmatrix} H_{1,1}, & H_{1,2}, & H_{1,3} \\ H_{2,1}, & H_{2,2}, & H_{2,3} \\ H_{3,1}, & H_{3,2}, & H_{3,3} \end{bmatrix},$$

$$d_4 := \det[H]. \quad (82)$$

5.1. Maximizer at the boundary

We come briefly to the non-generic case. There may be special cases where the observation dictates a dedicated analysis effort. When $\hat{\boldsymbol{s}}$ is a boundary point of $\boldsymbol{S}$ at least one component of $\hat{\boldsymbol{s}}$ assumes a boundary value. The ML equations are to be altered accordingly,

$$\frac{\partial L}{\partial \nu_S} < 0, \qquad \hat{\nu}_S = 0,$$

$$\frac{\partial L}{\partial \lambda_S} < 0, \qquad \hat{\lambda}_S = 0,$$

$$\frac{\partial L}{\partial r_S} < 0, \qquad \frac{\partial L}{\partial t_S} = 0, \qquad (\hat{X}_S, \hat{Y}_S) \in \partial\boldsymbol{F}. \quad (83)$$

In (83), $\partial L/\partial r_S$ means the inwardly directed radial direction derivative and $\partial L/\partial t_S$ is the derivative in

tangential direction. The latter direction is that of the boundary curve $\partial \boldsymbol{F}$ at the boundary point $(\hat{X}_{\mathrm{S}}, \hat{Y}_{\mathrm{S}})^{17}$.

Sources of interest lie seldom near the boundary of the field of view $\boldsymbol{F}$. However, a vanishing source extent parameter $\hat{\nu}_{\mathrm{S}}$ is not uncommon. The same applies to a vanishing estimate for the source count estimate $\hat{\lambda}_{\mathrm{S}}$ in weak source count enhancements, cf. Sect. 6.1. The second condition in (83) is equivalent to $\partial L(0)/\partial \lambda_{\mathrm{S}} < 0$ and is fulfilled for a high background or small number of source counts within the observation time.

In both cases, the corresponding inequalities in (83) are to be satisfied. Estimations for sources near the boundary $\partial \boldsymbol{F}$ of the field of view become difficult because of data censoring and tend to be unreliable and are not pursued any further here.

5.2. Gradient and Hessian of the log-likelihood

To solve the ML Eqs. (79) and to assign error bars, all partial derivatives of first and second order are needed. We consider $\tilde{T} := T \cdot V[A, \tilde{\epsilon}]$ as a constant with respect to the point source parameters $X_{\mathrm{S}}, Y_{\mathrm{S}}, \lambda_{\mathrm{S}}, \nu_{\mathrm{S}}$, and arrive with the denotations of (75) and $\nu_m := \nu(x_m, y_m, A_m)$ as well as $V_m := V[A_m, \epsilon(x_m, y_m)]$ at the representation for the four first partial derivatives of L,

$$\frac{\partial L}{\partial X_{\mathrm{S}}} := 2\lambda_{\mathrm{S}} \sum_{m=1}^{M} \frac{V_m p_m(X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}})}{\lambda_{\mathrm{B},m} + \lambda_{\mathrm{S}} V_m p_m(X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}})} \frac{X_m - X_{\mathrm{S}}}{\nu_m},$$

$$\frac{\partial L}{\partial Y_{\mathrm{S}}} := 2\lambda_{\mathrm{S}} \sum_{m=1}^{M} \frac{V_m p_m(X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}})}{\lambda_{\mathrm{B},m} + \lambda_{\mathrm{S}} V_m p_m(X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}})} \frac{Y_m - Y_{\mathrm{S}}}{\nu_m},$$

$$\frac{\partial L}{\partial \lambda_{\mathrm{S}}} := -\tilde{T} + \sum_{m=1}^{M} \frac{V_m p_m(X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}})}{\lambda_{\mathrm{B},m} + \lambda_{\mathrm{S}} V_m p_m(X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}})},$$

$$\frac{\partial L}{\partial \nu_{\mathrm{S}}} := \lambda_{\mathrm{S}} \cdot \sum_{m=1}^{M} \frac{V_m p_m(X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}})}{\lambda_{\mathrm{B},m} + \lambda_{\mathrm{S}} V_m p_m(X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}})}$$
$$\times \frac{(X_m - X_{\mathrm{S}})^2 + (Y_m - Y_{\mathrm{S}})^2 - \nu_m}{\nu_m^2}. \quad (84)$$

In order to represent the second order derivatives in the separated formula block (93) concisely, we use a yet shorter notation with V_m, ν_m as before and

$$p_m := p_m(X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}}),$$
$$\lambda_m := \lambda_{\mathrm{B},m} + \lambda_{\mathrm{S}} V_m p_m(X_{\mathrm{S}}, Y_{\mathrm{S}}, \nu_{\mathrm{S}}),$$
$$r_m^2 := (X_m - X_{\mathrm{S}})^2 + (Y_m - Y_{\mathrm{S}})^2. \quad (85)$$

In this notation we have for the 10 different elements $H_{j,k}$, $j = 1, 2, 3, 4$, $j \leq k$, of the symmetric Hessian of L the expressions in (93). Later in Sect. 6, formula (108) we shall need also the higher partial derivatives $\partial^k L/\partial \lambda_{\mathrm{S}}^k$, $k > 2$. From the entry for $\partial^2 L/\partial \lambda_{\mathrm{S}}^2$ in the eighth line of (93), it follows in the notation of (85),

$$\frac{\partial^k L}{\partial \lambda_{\mathrm{S}}^k} := (-1)^{k-1}(k-1)! \cdot \sum_{m=1}^{M} \left(\frac{V_m p_m}{\lambda_m}\right)^k, \quad k > 1. (86)$$

---

[17] If the boundary curve $\partial \boldsymbol{F}$ has a corner at $(\hat{X}_{\mathrm{S}}, \hat{Y}_{\mathrm{S}})$, the two one-sided derivatives must satisfy $\partial L/\partial t_{\mathrm{S}} < 0$.

6. Statistical errors for the source parameter estimates

Two questions are fundamental for the physical interpretation of the source parameter estimates:

(a) Does the count enhancement really stem from a celestial X-ray source or could it stem from a haphazard count fluctuation?

(b) When (a) is answered in favour of a real source and its source parameters are estimated: how large are the error bars?   (87)

In a statistical setting, (a) from (87) can be formulated as a pair of exclusive hypotheses in our adopted source model:

H_0: A celestial X-ray source caused the count enhancement, $\lambda_{\mathrm{S}} > 0$,

H_1: No celestial X-ray source caused the count enhancement, $\lambda_{\mathrm{S}} = 0$.   (88)

In the form of (88), problem (a) of (87) turned into a problem of statistical hypothesis testing. Using the associated Likelihood Ratio (appearing in the form of the difference of two log-likelihoods), we wish to decide the alternative (88). Within this method we have to evaluate two log-likelihoods associated to H_0, H_1,

$$\hat{L}_0 := L\left(\hat{X}_{\mathrm{S}}, \hat{Y}_{\mathrm{S}}, \hat{\lambda}_{\mathrm{S}}, \hat{\nu}_{\mathrm{S}}\right),$$
$$\hat{\hat{L}}_1 := L\left(\hat{X}_{\mathrm{S}}, \hat{Y}_{\mathrm{S}}, 0, \hat{\hat{\nu}}_{\mathrm{S}}\right). \quad (89)$$

In (89), $\hat{\boldsymbol{s}} := \left(\hat{X}_{\mathrm{S}}, \hat{Y}_{\mathrm{S}}, \hat{\lambda}_{\mathrm{S}}, \hat{\nu}_{\mathrm{S}}\right)$ is the unconditional maximizer as in (77) and $\hat{\hat{\boldsymbol{s}}} := \left(\hat{X}_{\mathrm{S}}, \hat{Y}_{\mathrm{S}}, 0, \hat{\hat{\nu}}_{\mathrm{S}}\right)$ is the conditional maximizer over the restricted source space $\boldsymbol{S}_1 := \{\boldsymbol{s} \in \boldsymbol{S} : \lambda_{\mathrm{S}} = 0\}$ of H_1. Clearly, $\hat{L}_0 \geq \hat{\hat{L}}_1$. Hence, twice the log-likelihood difference, $\hat{l}$, is nonnegative,

$$\hat{l} := 2 \cdot \left(\hat{L}_0 - \hat{\hat{L}}_1\right) \geq 0. \quad (90)$$

With the factor 2 in (90), the *likelihood statistics l* follows asymptotically a standard distribution, see (94). In terms of the parent statistics l, the number $\hat{l}_{\mathrm{S}} := \hat{l}/2$ is the *likelihood of source existence* of the sample returned by MLE and mentioned in many publications on ROSAT data.

Let $\alpha \in (0, 1)$ be the requested significance level of the test and let $g_M(l)$ be the sample frequency function of the test statistics l from (90). The α-quantile, $l_M(\alpha)$, is the solution of the equation for $l_M(\alpha)$

$$\int_0^{l_M(\alpha)} g_M(l)\,\mathrm{d}l = \alpha. \quad (91)$$

With $l_\alpha := l_M(\alpha)$ the decision rule for the alternative (88) is in terms of l from (90) and (89)

$$\hat{l} \begin{cases} > l_\alpha, & H_0 \text{ is accepted}, \\ \leq l_\alpha, & H_0 \text{ is rejected}. \end{cases} \quad (92)$$

$$\frac{\partial^2 L}{\partial X_{\mathrm{S}}^2} := 2\lambda_{\mathrm{S}} \sum_{m=1}^{M} \frac{V_m p_m}{\nu_m \lambda_m} \left[\frac{2\lambda_{\mathrm{B},m}}{\nu_m \lambda_m} \cdot (X_m - X_{\mathrm{S}})^2 - 1 \right],$$

$$\frac{\partial^2 L}{\partial X_{\mathrm{S}} \partial Y_{\mathrm{S}}} := 4\lambda_{\mathrm{S}} \sum_{m=1}^{M} \frac{\lambda_{\mathrm{B},m} V_m p_m}{\nu_m^2 \lambda_m^2} \cdot (X_m - X_{\mathrm{S}})(Y_m - Y_{\mathrm{S}}),$$

$$\frac{\partial^2 L}{\partial X_{\mathrm{S}} \partial \lambda_{\mathrm{S}}} := -2 \sum_{m=1}^{M} \frac{\lambda_{\mathrm{B},m} V_m p_m}{\nu_m \lambda_m^2} \cdot (X_m - X_{\mathrm{S}}),$$

$$\frac{\partial^2 L}{\partial X_{\mathrm{S}} \partial \nu_{\mathrm{S}}} := 2\lambda_{\mathrm{S}} \sum_{m=1}^{M} \frac{V_m p_m}{\nu_m^2 \lambda_m} \cdot (X_m - X_{\mathrm{S}})$$
$$\times \left[\frac{\lambda_{\mathrm{B},m}(r_m^2 - \nu_m)}{\nu_m \lambda_m} - 1 \right],$$

$$\frac{\partial^2 L}{\partial Y_{\mathrm{S}}^2} := 2\lambda_{\mathrm{S}} \sum_{m=1}^{M} \frac{V_m p_m}{\nu_m \lambda_m} \left[\frac{2\lambda_{\mathrm{B},m}}{\nu_m \lambda_m} \cdot (Y_m - Y_{\mathrm{S}})^2 - 1 \right],$$

$$\frac{\partial^2 L}{\partial Y_{\mathrm{S}} \partial \lambda_{\mathrm{S}}} := -2 \sum_{m=1}^{M} \frac{\lambda_{\mathrm{B},m} V_m p_m}{\nu_m \lambda_m^2} \cdot (Y_m - Y_{\mathrm{S}}),$$

$$\frac{\partial^2 L}{\partial Y_{\mathrm{S}} \partial \nu_{\mathrm{S}}} := 2\lambda_{\mathrm{S}} \sum_{m=1}^{M} \frac{V_m p_m}{\nu_m^2 \lambda_m} \cdot (Y_m - Y_{\mathrm{S}})$$
$$\times \left[\frac{\lambda_{\mathrm{B},m}(r_m^2 - \nu_m)}{\nu_m \lambda_m} - 1 \right],$$

$$\frac{\partial^2 L}{\partial \lambda_{\mathrm{S}}^2} := - \sum_{m=1}^{M} \left(\frac{V_m p_m}{\lambda_m} \right)^2,$$

$$\frac{\partial^2 L}{\partial \lambda_{\mathrm{S}} \partial \nu_{\mathrm{S}}} := \sum_{m=1}^{M} \frac{\lambda_{\mathrm{B},m} V_m p_m}{\nu_m^2 \lambda_m^2} \cdot \left(r_m^2 - \nu_m \right),$$

$$\frac{\partial^2 L}{\partial \nu_{\mathrm{S}}^2} := \lambda_{\mathrm{S}} \sum_{m=1}^{M} \frac{V_m p_m}{\nu_m^4 \lambda_m}$$
$$\times \left[\frac{\lambda_{\mathrm{B},m}}{\lambda_m} (r_m - \nu_m)^2 - \left(2r_m^2 - \nu_m \right) \nu_m \right]. \quad (93)$$

After (92), the tasks are (a) the determination of the sample frequency distribution, g_M, of l and (b) the solution of (91) for l_α when the sample size M is given. The latter poses no numerical difficulties since the integrand g_M is positive.

To avoid the appearance of the sample frequeny function from (91), we resort again to large sample sizes.

According to Doob (1934), Wilks (1938), for sample size $M \to \infty$ the distribution $g_M(l)$ tends to the χ_ν^2 distribution with $\nu > 0$ degrees of freedom with $t \in [0, \infty)$

$$g_M(t)\mathrm{d}t \to \frac{1}{\Gamma(\nu/2)} \mathrm{e}^{-t/2} \left(\frac{t}{2} \right)^{\nu/2-1} \mathrm{d}\left(\frac{t}{2} \right) =: \chi_\nu^2(t)\mathrm{d}t. (94)$$

The number of restrictions in H_1 equals $\nu = 1$ in (94). Given l by (90), for the large sample sizes, the significance level

$$\alpha(l) := \int_0^l \chi_1^2(t)\mathrm{d}t \quad (95)$$

serves equally well to judge the likelihood of source existence. It should always be borne in mind that a large sample size is presupposed when replacing $g_M(t)$ by $\chi_1^2(t)$.

More delicate are the error bars from (b) in (87). The defining maximum relation (77) for the ML source estimate vector $\hat{\mathbf{s}}$ defines a random vector when the photon sample $\mathcal{P}$ is viewed as a set of random vectors. As such, $\mathbf{s}$ possesses a sampling distribution, say $\mathbf{P}_{\mathbf{S}}^{(M)}$ given by (91) with l as upper integration limit. Then, one can calculate the mean vector $\bar{\mathbf{s}}$, its sample counterpart $\hat{\bar{\mathbf{s}}}$, and the elements $C_{j,k}$ of the covariance matrix $\mathbf{Cov}[\mathbf{s}]$,

$$C_{j,k} := \mathbf{E}_{\mathbf{P}_{\mathbf{S}}^M}[(s_j - \bar{s}_j)(s_k - \bar{s}_k)], \quad j, k = 1, 2, 3, 4,$$
$$\bar{s}_j := \mathbf{E}_{\mathbf{P}_{\mathbf{S}}^M}[s_j]. \quad (96)$$

The subscript $\mathbf{P}_{\mathbf{S}}^M$ in (96) means: expectation $\mathbf{E}$ is to be taken with respect to the distribution $\mathbf{P}_{\mathbf{S}}^{(M)}$. The $\bar{s}_j, C_{j,k}$ are assumed to exist for all sample sizes $M > 1$. If the distribution $\mathbf{P}_{\mathbf{S}}^{(M)}$ is of the type of a normal distribution, i.e. at least unimodal, and the statistical interdependence among the components of $\mathbf{s}$ is low, it is justified to define the quantities

$$\sigma_k := \sqrt{\mathbf{Cov}[\mathbf{s}]_{k,k}}, \qquad k = 1, 2, 3, 4, \quad (97)$$

as standard errors. Unfortunately, the sample distribution $\mathbf{P}_{\mathbf{S}}^{(M)}$ is analytically nearly inaccessible. Also, our log-likelihood function $L(X_{\mathrm{S}}, Y_{\mathrm{S}}, \lambda_{\mathrm{S}}, \nu_{\mathrm{S}})$ does not allow for a set of four jointly sufficient statistics for the four source parameters. The distributions in the "sufficiency" class are explicitly known, see Stuart & Ord (1991), and our L lies outside. In this situation, other possibilities to define statistical errors are to be drawn into consideration.

For large sample sizes, see Wald (1943), the distribution $\mathbf{P}_{\mathbf{S}}^{(M)}$ tends to a multivariate normal distribution with zero mean and covariance matrix with elements according to (96) at $\mathbf{s} = \hat{\mathbf{s}}$,

$$\mathbf{Cov}_{j,k}[\hat{\mathbf{s}}] := -\mathbf{E}\left[\frac{\partial^2 L(\hat{\mathbf{s}})}{\partial s_j \partial s_k} \right] = \mathbf{E}\left[\frac{\partial L(\hat{\mathbf{s}})}{\partial s_j} \cdot \frac{\partial L(\hat{\mathbf{s}})}{\partial s_k} \right]$$
$$= \int_{\mathbf{R}^2} \mathrm{e}^{L\left(X,Y \mid \hat{X}_{\mathrm{S}}, \hat{Y}_{\mathrm{S}}, \hat{\lambda}_{\mathrm{S}}, \hat{\nu}_{\mathrm{S}}\right)} \cdot \frac{\partial L(\hat{\mathbf{s}})}{\partial s_j} \frac{\partial L(\hat{\mathbf{s}})}{\partial s_k} \mathrm{d}X\mathrm{d}Y. (98)$$

To circumvent the integration in (98), we replace the parent product moments $C_{j,k}$ by the sample counterparts and obtain

$$\hat{C}_{j,k} := \mathbf{Cov}_{j,k}[\hat{\mathbf{s}}] := \frac{1}{M} \sum_{m=1}^{M} \mathrm{e}^{L(X_k, Y_k \mid \hat{\mathbf{s}})}$$
$$\times \frac{\partial L(X_k, Y_k \mid \hat{\mathbf{s}})}{\partial s_j} \cdot \frac{\partial L(X_k, Y_k \mid \hat{\mathbf{s}})}{\partial s_k},$$
$$\hat{\sigma}_k := \sqrt{\hat{\mathbf{Cov}}_{k,k}[\hat{\mathbf{s}}]}. \quad (99)$$

In EXSAS, a less costly approximation is accepted. For a large number of source photons, the frequency distribution e^L in (98) tends to a Dirac distribution supported at the ML estimate $\hat{\mathbf{s}}$. Thus, the approximation $\mathbf{E}[H] \approx H(\hat{s})$ of the expected matrix by the information matrix $H(\hat{s})$ lies near and is used.

Let $c \neq 0$ be a nonvanishing (column) p-vector and $H = H^{\mathrm{T}} > 0$ a real, symmetric, positive definite, simple

$p \times p$ matrix. Then for any positive number $h > 0$ the following equality holds:

$$m(c) := \max_{h = x^T H x} c^T x = \sqrt{h \cdot c^T H^{-1} c}. \qquad (100)$$

To calculate the positive number $c^T H^{-1} c$, somewhat less than the inverse H^{-1} needs to be known. Let H be given through its diagonalized form $\Lambda := \text{diag}(\lambda_1, ..., \lambda_p)$ by

$$H = E^T \Lambda E, \quad E^T E = I, \quad E := (e_1, ..., e_p). \qquad (101)$$

In terms of the eigenvectors e_1 to e_p and positive eigenvalues[18] λ_1 to λ_p of H, the maximum $m(c)$ receives the representation

$$m(c) = \sqrt{h \cdot (Ec)^T \Lambda^{-1} (Ec)}. \qquad (102)$$

We wish to find the smallest box with bounding hyperplanes perpendicular to the axes in x-space which contains the ellipsoid $h = x^T H x$. In other words, the distances of the said hyperplanes from the origin are the quantities sought for. To find them, we choose for $c = \pm\delta_k := \pm(0, ..., 0, 1, 0, ..., 0)$ the kth standard base vector, $k = 1, ..., p$, and obtain (102) upon the replacement of h by the α-quantile, $\chi^2_{1,\alpha}$, of the χ^2 distribution with one degree of freedom as

$$\hat{\sigma}_k(\alpha) := m(\pm\delta_k) = \sqrt{\chi^2_{1,\alpha} \cdot \left(\frac{e^2_{k,1}}{\lambda_1} + ... + \frac{e^2_{k,p}}{\lambda_p} \right)}, \qquad (103)$$

where $\alpha \in (0, 1)$ is the used significance level[19].

In EXSAS, the 1σ-error bars are calculated. As usual, 1σ codifies the significance level $\alpha = 0.682...$ This level corresponds exactly to the quantile value $\chi^2_{1,\alpha} = 1$. The statistical interpretation of the EXSAS error bars, $\hat{\sigma}_k$, $k = 1, 2, 3, 4$, for the unknown true point source parameter s_k is then: in 68.3% of all MLEs, the containment relation $s_k \in [\hat{s}_k - \hat{\sigma}_k, \hat{s}_k + \hat{\sigma}_k]$ holds good – under the caveat of a large number of source photons.

6.1. Upper confidence limits for the source count rate

All X-ray sources are variable on source-specific time scales. A certain source found in one observation may not show up in another observation. The task is then to prove quantitatively the non-detectability of a source in a given direction. We wish to assert: based on the observation, with a selectable probability β, a hypothetical source cannot have a count rate λ_S exceeding an upper limit count rate $\hat{\lambda}^+_{S,\beta}$.

Assume a (hypothetical) source position (X^0_S, Y^0_S) with source extent parameter ν^0_S is given. For such a source with

these parameters and an unknown source count rate λ_S, EXSAS offers the (approximate) calculation of the upper limit of a confidence interval, $[\hat{\lambda}^-_{S,\beta}, \hat{\lambda}^+_{S,\beta}]$, containing the unknown true source count rate parameter λ_S with a selectable confidence level, $\beta \in (0, 1)$, see the EXSAS command (125). The statistical interpretation is: the unknown true source count rate λ_S obeys in the fraction β of all MLEs the inequality

$$\hat{\lambda}^-_{S,\beta} \leq \lambda_S \leq \hat{\lambda}^+_{S,\beta},$$
$$\hat{\lambda}^+_{S,\beta} := \hat{\mathcal{L}}^{-1}_+ \left[\hat{\mathcal{L}}\left(\hat{\lambda}_S \right) - \chi^2_{1,\beta}/2 \right],$$
$$\hat{\lambda}^-_{S,\beta} := \hat{\mathcal{L}}^{-1}_- \left[\max\left\{ \hat{\mathcal{L}}(0), \hat{\mathcal{L}}\left(\hat{\lambda}_S \right) - \chi^2_{1,\beta}/2 \right\} \right],$$
$$\hat{\mathcal{L}}(\lambda_S) := L(X^0_S, Y^0_S, \lambda_S, \nu^0_S). \qquad (104)$$

Inspection of $\partial L/\partial\lambda_S$ in (84) and $\partial^2 L/\partial\lambda^2_S$ in (108) shows that $\hat{\mathcal{L}}(\lambda_S)$ is a concave function with a unique maximizer $\bar{\lambda}_S(= \hat{\lambda}_S)$, see Lemma 1 in Sect. 7 for details. Thus the inverse function $\hat{\mathcal{L}}^{-1}$ has for $\hat{\mathcal{L}}'(0) > 0$ a lower (= weaker) branch $\hat{\mathcal{L}}^{-1}_-$ with domain of definition $[\hat{\mathcal{L}}(0), \hat{\mathcal{L}}(\bar{\lambda}_S)]$ and an upper (= stronger) count rate branch $\hat{\mathcal{L}}^{-1}_+$ with domain $(-\infty, \hat{\mathcal{L}}(\bar{\lambda}_S)]$. The upper confidence limit $\hat{\lambda}^+_{S,\beta}$ exists for all significance levels β in contrast to the lower limit. There is a critical confidence level $\hat{\beta}_c$ defined by the equation $\chi^2_{1,\beta} = \hat{\mathcal{L}}(\hat{\lambda}_S) - \hat{\mathcal{L}}(0)$ for β such that the lower limit becomes trivial, $\hat{\lambda}^-_{S,\beta} = 0$, for $\beta \geq \hat{\beta}_c$. The achievable confidence level at the trivial lower bound, $\hat{\beta}_0$, stays then below the requested one, $\hat{\beta}_0 < \beta$.

The subject of estimator variances in astronomical context is a recurrent theme. The works of Avni (1976), Cash (1978), Strong (1985) belong to a larger body of literature.

7. Numerical solution of the ML equations

Given an initial estimate, $\hat{s}^{(0)}$, for the source parameter vector s, we wish to set up a numerical iterative algorithm that generates a sequence of refined estimates

$$(\hat{s}^{(k)})_{k \in N_0} := \left(\hat{X}^{(k)}_S, \hat{Y}^{(k)}_S, \hat{\lambda}^{(k)}_S, \hat{\nu}^{(k)}_S \right)_{k \in N_0} \qquad (105)$$

such that the sequence (105) has a limit, $\hat{s}^{(*)}$, for $k \to \infty$ and the necessary maximum conditions (79) are satisfied in the limit for $s = \hat{s}^{(*)}$. If the initial estimate $\hat{s}^{(0)}$ is near the true maximizer of L, the convergence of (105) to the latter is likely.

The first step towards this goal consists in writing three of the four equations in (79) in fixed-point form. The third equation thereof, $\partial L/\partial\lambda_S = 0$, the *local mean count rate equation*, will be treated differently. As mentioned in Sect. 5.1, $\partial L(0)/\partial\lambda_S \leq 0$ entails the vanishing maximizer $\hat{\Lambda}_S = 0$. In the remaining case $\partial L(0)/\partial\lambda_S 0$ a positive local mean count rate $\lambda_S > 0$ maximizes. In this case, we obtain from (79) the system of four equations,

---

[18] In terms of eigenvalues, negative/positive definiteness is equivalent to the negativity/positivity of the eigenvalues. The superscript T means matrix transposition.

[19] The quantity $\chi^2_{1,\alpha}$, is the upper integration limit in (95) where α is given. In Strong (1985), our (103) appears as (9) there and our (102) as (11) there in a less terse denotation.

three of them in fixed-point form,

$$X_S = X(X_S, Y_S, \lambda_S, \nu_S),$$
$$Y_S = Y(X_S, Y_S, \lambda_S, \nu_S),$$
$$\tilde{T} = T(X_S, Y_S, \lambda_S, \nu_S),$$
$$\nu_S = N(X_S, Y_S, \lambda_S, \nu_S),$$

$$X(X_S, Y_S, \lambda_S, \nu_S) := \frac{\sum_{m=1}^{M} W_m(X_S, Y_S, \lambda_S, \nu_S) \cdot \frac{X_m}{\nu_m}}{\sum_{m=1}^{M} W_m(X_S, Y_S, \lambda_S, \nu_S) \cdot \frac{1}{\nu_m}},$$

$$Y(X_S, Y_S, \lambda_S, \nu_S) := \frac{\sum_{m=1}^{M} W_m(X_S, Y_S, \lambda_S, \nu_S) \cdot \frac{Y_m}{\nu_m}}{\sum_{m=1}^{M} W_m(X_S, Y_S, \lambda_S, \nu_S) \cdot \frac{1}{\nu_m}},$$

$$T(X_S, Y_S, \lambda_S, \nu_S) := \sum_{m=1}^{M} W_m(X_S, Y_S, \lambda_S, \nu_S),$$

$$N(X_S, Y_S, \lambda_S, \nu_S) := \frac{\sum_{m=1}^{M} W_m(X_S, Y_S, \lambda_S, \nu_S) \cdot q_m}{\sum_{m=1}^{M} W_m(X_S, Y_S, \lambda_S, \nu_S) \cdot \frac{1}{\nu_m^2}},$$

$$q_m := q_m(X_S, Y_S) := \frac{(X_S - X_m)^2 + (Y_S - Y_m)^2 - \nu_m^0}{\nu_m^2},$$

$$W_m(X_S, Y_S, \lambda_S, \nu_S) := \frac{V_m p_m(X_S, Y_S, \nu_S)}{\lambda_{B,m} + \lambda_S p_m(X_S, Y_S, \nu_S)},$$

$$\nu_m^0 := 2\sigma^2 [A_m, \epsilon(x_m, y_m)]. \tag{106}$$

The third equation in (106), the local mean count rate equation, is not in fixed-point form. Fortunately, the function $T(X_S, Y_S, \lambda_S, \nu_S)$ is the simplest of all right-hand sides in (106). We shall benefit from the fact that T is a convex decreasing function in λ_S as inspection of (106) and (86) for $k = 3$ reveals.

The local mean count rate equation will be treated separately. Given X_S, Y_S, ν_S, we consider $\tilde{T} = T(X_S, Y_S, \lambda_S, \nu_S)$ as an equation for λ_S alone and denote by

$$\lambda_S := \lambda_S\left(X_S, Y_S, \nu_S, \tilde{T}\right) \tag{107}$$

its solution. A result on solution existence and uniqueness of the count rate λ_S will be stated below in Lemma 1. We begin with some preparations.

The monotonicity property

$$(-1)^k \cdot \frac{\partial^k T(X_S, Y_S, \lambda_S, \nu_S)}{\partial \lambda^k} > 0, \quad k \in \mathbf{N}, \quad \lambda_S \geq 0. \tag{108}$$

derives from (86). The monotony relations (108) permit us to construct a lower bound, $\underline{T}$, and an upper bound, $\bar{T}$, to T. Let $\lambda_S^0 \geq 0$ be any fixed local mean source count rate. Under this choice of λ_S^0 holds

$$\underline{T}(X_S, Y_S, \lambda_S, \nu_S) < T(X_S, Y_S, \lambda_S, \nu_S),$$
$$T(X_S, Y_S, \lambda_S, \nu_S) < \bar{T}(X_S, Y_S, \lambda_S, \nu_S),$$
$$\bar{T}(X_S, Y_S, \lambda_S, \nu_S) := \min\{\bar{T}_1(X_S, Y_S, \lambda_S, \nu_S),$$
$$\bar{T}_2(X_S, Y_S, \lambda_S, \nu_S)\},$$

$$\bar{T}_1(X_S, Y_S, \lambda_S, \nu_S) := \sum_{m=1}^{M} W_m(X_S, Y_S, \lambda_S, \nu_S)\Big|_{\lambda_{B,m}=0}$$
$$=: \frac{W(X_S, Y_S, \nu_S)}{\lambda_S},$$

$$\bar{T}_2(X_S, Y_S, \lambda_S, \nu_S) := T(X_S, Y_S, \lambda_S^0, \nu_S)$$
$$+ \frac{\partial T(X_S, Y_S, \lambda_S^0, \nu_S)}{\partial \lambda_S}(\lambda_S - \lambda_S^0)$$
$$+ \frac{1}{2}\frac{\partial^2 T(X_S, Y_S, \lambda_S^0, \nu_S)}{\partial \lambda_S^2}(\lambda_S - \lambda_S^0)^2,$$

$$\underline{T}(X_S, Y_S, \lambda_S, \nu_S) := T(X_S, Y_S, \lambda_S^0, \nu_S)$$
$$+ \frac{\partial T(X_S, Y_S, \lambda_S^0, \nu_S)}{\partial \lambda_S}(\lambda_S - \lambda_S^0). \tag{109}$$

The functions $\underline{T}$ and $\bar{T}_2$ can be formed for all $\lambda_S^0 \geq 0$. We choose $\lambda_S^0 := 0$. Inverting the inequalities $\underline{T} \leq T \leq \bar{T}$ for λ_S yields the following result concerning the local mean source count rate (MCR) equation.

Lemma 1 (MCR existence, uniqueness, enclosure).
The local mean count rate equation

$$\tilde{T} = T(X_S, Y_S, \lambda_S, \nu_S) \tag{110}$$

has a unique solution $\lambda_S := \lambda_S(X_S, Y_S, \nu_S, \tilde{T})$ iff

$$0 < \tilde{T} \leq T(X_S, Y_S, 0, \nu_S) =: \bar{T}(X_S, Y_S, \nu_S). \tag{111}$$

The solution λ_S is enclosed by

$$\underline{\lambda}_S(X_S, Y_S, \nu_S) \leq \lambda_S \leq \bar{\lambda}_S(X_S, Y_S, \nu_S) \tag{112}$$

where with T from (106) and W from (109)

$$\underline{\lambda}_S(X_S, Y_S, \nu_S, \tilde{T}) := (\tilde{T} - T_0)/T_1,$$
$$\bar{\lambda}_S(X_S, Y_S, \nu_S, \tilde{T}) := \min\{\bar{\lambda}_1(X_S, Y_S, \nu_S, \tilde{T}),$$
$$\bar{\lambda}_2(X_S, Y_S, \nu_S, \tilde{T})\},$$
$$\bar{\lambda}_1(X_S, Y_S, \nu_S, \tilde{T}) := \frac{W(X_S, Y_S, \nu_S)}{\tilde{T}},$$
$$\bar{\lambda}_2(X_S, Y_S, \nu_S, \tilde{T}) := \begin{cases} \tau, & \underline{T} \leq \tilde{T} \leq \bar{T}, \\ +\infty, & \tilde{T} < \underline{T}, \end{cases}$$
$$\tau := -T_1 - \sqrt{T_1^2 - 2T_2(\tilde{T} - T_0)},$$
$$T_0 := T(X_S, Y_S, 0, \nu_S),$$
$$T_1 := \frac{\partial T(X_S, Y_S, 0, \nu_S)}{\partial \lambda_S},$$
$$T_2 := \frac{\partial^2 T(X_S, Y_S, 0, \nu_S)}{\partial \lambda_S^2},$$
$$\underline{T} := T_0 - \frac{T_1^2}{2T_2}. \tag{113}$$

The upper bound $\bar{\lambda}_S$ stays always finite (and may have a jump).

Once an enclosing interval $[\underline{\lambda}_S, \bar{\lambda}_S]$ for the source count rate λ_S is available, the length of the enclosing interval can be shrunk to any given length $\epsilon_{\lambda_S} > 0$ by bisection or otherwise. Using $\hat{X}_S^{(0)}, \hat{Y}_S^{(0)}, \hat{\nu}_S^{(0)}$ in the third equation, $\tilde{T} = T$, of (106), its approximative solution to precision ϵ_{λ_0} is denoted by $\hat{\lambda}_S^{(0)}$. Now, given an initial estimate $\hat{s}^{(0)}$ formed with that $\hat{\lambda}_S^{(0)}$, its successor estimate $\hat{s}^{(k+1)}$ is calculated

by performing the iteration step from the kth iterate to the $(k+1)$th,

$$X_{\rm S}^{(k+1)} := X\left(X_{\rm S}^{(k)}, Y_{\rm S}^{(k)}, \lambda_{\rm S}^{(k)}, \nu_{\rm S}^{(k)}\right),$$

$$Y_{\rm S}^{(k+1)} := Y\left(X_{\rm S}^{(k+1)}, Y_{\rm S}^{(k)}, \lambda_{\rm S}^{(k)}, \nu_{\rm S}^{(k)}\right),$$

$$\nu_{\rm S}^{(k+1)} := N\left(X_{\rm S}^{(k+1)}, Y_{\rm S}^{(k+1)}, \lambda_{\rm S}^{(k)}, \nu_{\rm S}^{(k)}\right),$$

$$\tilde{T} := T\left(X_{\rm S}^{(k+1)}, Y_{\rm S}^{(k+1)}, \lambda_{\rm S}^{(k+1)}, \nu_{\rm S}^{(k+1)}\right). \tag{114}$$

The last equation in (114) defines $\lambda_{\rm S}^{(k+1)}$ implicitly and it is solved iteratively as described above. With appropriately chosen tolerances for position and extend components,

$$\epsilon_{X_{\rm S}} > 0, \quad \epsilon_{Y_{\rm S}} > 0, \quad \epsilon_{\nu_{\rm S}} > 0, \tag{115}$$

the iteration step (114) is carried out as long as the termination criterion,

$$\left|X_{\rm S}^{(k+1)} - X_{\rm S}^{(k)}\right| < \epsilon_{X_{\rm S}},$$

$$\left|Y_{\rm S}^{(k+1)} - Y_{\rm S}^{(k)}\right| < \epsilon_{Y_{\rm S}},$$

$$\left|\nu_{\rm S}^{(k+1)} - \nu_{\rm S}^{(k)}\right| < \epsilon_{\nu_{\rm S}}, \tag{116}$$

is not met. During the iteration the likelihood, $L^{(k)}$, belonging to the kth iteration step and the relative change in likelihood, $\delta L^{(k)}$,

$$L^{(k)} := L\left(X_{\rm S}^{(k)}, Y_{\rm S}^{(k)}, \lambda_{\rm S}^{(k)}, \nu_{\rm S}^{(k)}\right),$$

$$\delta L^{(k)} := \frac{L^{(k)} - L^{(k-1)}}{L^{(k)}} \tag{117}$$

are superintended. With a threshold ϵ_L in the order of 10^{-5} the violation of

$$\delta L^{(k)} > \epsilon_L \tag{118}$$

serves, in addition to (116), as a termination criterion for the iteration (114) in EXSAS.

Again, the success of the iteration hinges on the closeness of the initial estimate $\hat{\boldsymbol{s}}^{(0)}$ to the unknown true maximizer. Therefore, EXSAS uses an elaborate process to find positional components in the initial estimate.

8. Example of and tests on the MLE of point sources

In practice, MLE was tested innumerable times. The ROSAT All-Sky Survey (RASS) described by Voges (1992, 1999) is a monumental instance. Also, several thousands of pointed ROSAT observations were processed by MLE as described, see for example the group of contributions prefaced by Trümper (1991b).

We shall illustrate the performance of the MLE twofold. Firstly, we apply MLE to a field with many point sources. Secondly, MLE was tested on simulated data. Simulated observations give the possibility to compare estimated parameter values with true values under the conditions of the implemented MLE (and test) software.
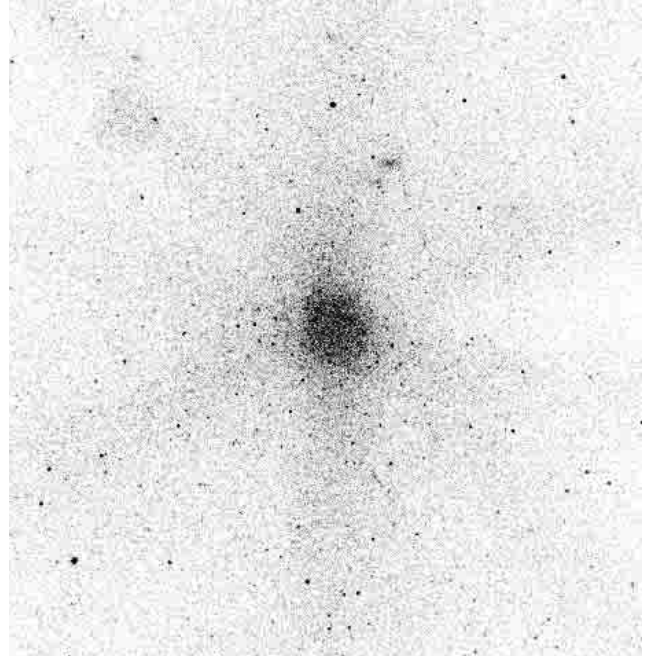
Fig. 9. $6.1° \times 6.1°$ region around the ecliptic north pole observed in the ROSAT all sky survey in the energy band $[0.1, 2.35]$ keV. The central exposure time is 15.5 ks

8.1. An example of the ML point source parameter estimation

Figure 9 shows a $6.1° \times 6.1°$ field around the ecliptic north pole observed during the ROSAT All Sky Survey in the energy band $[0.1, 2.35]$ keV. The exposure time of the image centre was 15.5 ks. Only the photons which fell into the inner 19 arcmin of the ROSAT PSPC field of view were selected resulting in the image Fig. 9. The darker strips emanating from the central yet darker region are caused by a varying exposure time due to the ROSAT survey attitude. Figure 10 shows the same region augmented by point source positions. The little dark circles indicate the 545 point source positions found by MLE.

8.2. Testing the MLE of point source parameters on simulated data

Many tests with simulated observations were performed in order to study the performance behaviour of the implemented MLE software in the high-dimensional parameter space spanned by the source photon parameters and additional hidden internal parameters. Among the varied input quantities in the test series were: source strength, source offset, background strength, and cut-off radius. Several dozens of output quantities were collected and plotted as function of the input quantity. As an typical example, we report on the source position error as a function of the source strength.

Figure 11 shows the stochastic process of the horizontal positional errors $e_X := \hat{X} - X_{\rm S}$ between the ML estimate $\hat{X}$ and the known true source position component $X_{\rm S}$ as a
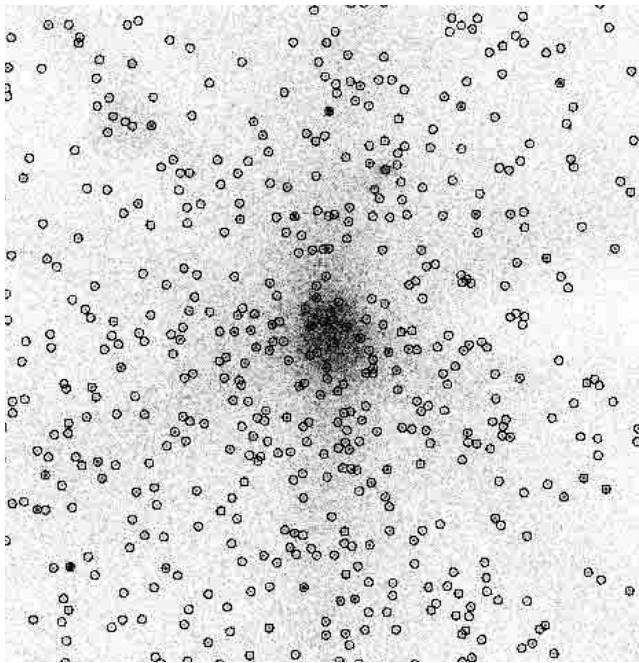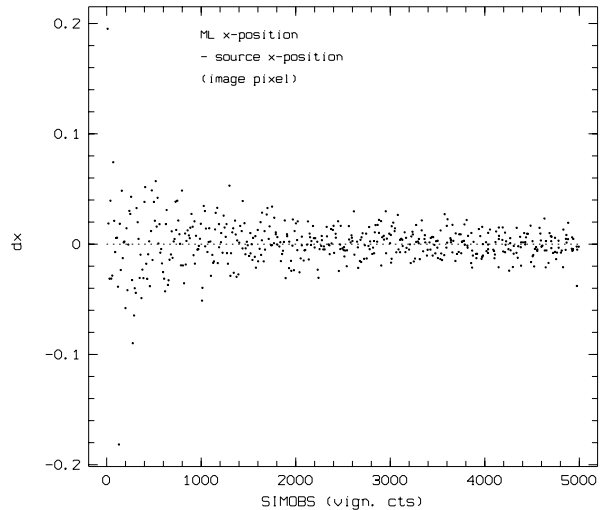
Fig. 10. The ecliptic north polar region as in Fig. 9. Circles indicate point source positions as found by MLE

Fig. 11. Horizontal positional errors $e_X := \hat{X} - X_S$ (in units of image pixels) as a function of the source photons Λ_S

Fig. 12. Vertical positional errors $e_Y := \hat{Y} - Y_S$ (in units of image pixels) in dependence of the source photons Λ_S

function of the source photon count number $\Lambda_S \in [0, 5000]$ as parameter. The plot is scaled such that all errors e_X fall within the vertical plot range. Evidently, the sample standard deviation $\sigma_{e_X}(\Lambda_S)$ becomes smaller from left to right in Fig. 11, i.e. for increasing source photon numbers Λ_S. This checks with the expected $\mathcal{O}(1/\sqrt{\Lambda_S})$ behaviour for large number of photons $\Lambda_S \to \infty$. Each of the 500 black dots means (a) the creation of a photon series $\mathcal{P}$ from (1) simulating a source with a random source position (X_S, Y_S) which is (b) afterwards subjected to a MLE analysis carried out with the help of the EXSAS commands described in the Appendix. Judged by eye, no bias around the horizontal axis $e_X = 0$ seems to be present for source photons above $\Lambda_S > 100$. An analogous picture is expected for the vertical positional error $e_Y := \hat{Y} - Y_S$ in Fig. 12. Again, all errors e_Y fall within the vertical plot range which is dilated by a factor 2 compared to Fig. 11. This stretching is due to the randomly large errors e_Y for small number of source photons, cf. $e_Y = 0.45$ image pixels for the upper left dot. A closer look reveals $\sigma_{e_X}(\Lambda_S) \approx \sigma_{e_Y}(\Lambda_S)$ as expected.
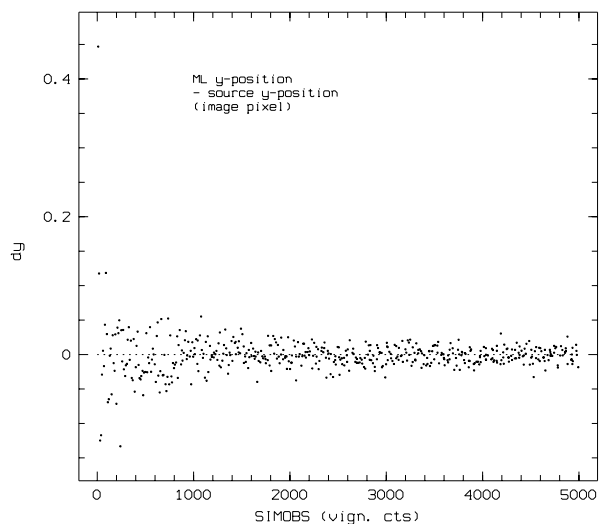
9. Discussion

A similar sliding window technique as the one described in Sect. 4.1.2 has been used in the analysis of EINSTEIN data, see Harris & Irwin (1984).

Already in the title, the application of the MLE is restricted to single sources. Thus, pairs of poorly resolved sources, less frequent triples of neighbouring sources or, generally, clusters of several sources are not admissible. In this context, distances are measured in units of the PSF width. The source parameters of source clusters must be estimated jointly. In a different EXSAS application, MLE for source groups is possible under the command name FIT/MULTI_SOURCE on image data.

At the end of Sect. 3.1, the motivations for selecting MLE were named. Besides ML estimation, source parameter estimation based on maximum entropy or wavelets are known and in use – to name only some recent methods. For the application of wavelet decomposition to ROSAT data, see Lazzati et al. (1999). A comparison of different estimation procedures is desirable, a task on its own, and far outside the scope of this article. An estimation procedure that is superior in all important respects seems not to exist.

The comparison of the hypothetical count rate distribution is made in the space of the observational data. This is the conventional strategy to circumvent the (ill-posed) inverse problem commonly referred to as deconvolution. After having estimated source parameters $\hat{X}_S, \hat{Y}_S, \hat{\lambda}_S^V, \hat{\nu}_S,$

the conversion to the source count rate λ_S and source flux is to be done. As this task is outside of MLE, details are not described in this work.

In Sect. 1, the Standard Analysis Software System (SASS) , cf. Voges et al. (1999, p. 223), was mentioned in passing. All ROSAT catalogues were produced with SASS. Meanwhile also a ROSAT all-sky survey of faint sources has been made public. For a list of all ROSAT catalogues see Voges et al. (2001). In relation to SASS, the EXSAS system is an extension of the former.

One implementation feature in EXSAS and SASS is not satisfactory. As mentioned, MLE on unbinned photon data is appropriate for weak sources and works down to, say, thrice as many photons per source as there are source parameters (i.e. 12 photons). On the other hand, the source parameter error bars need a large number of photons per source.

In practice no unrealistic error bar sizes were observed. Nevertheless, a better approximation to the small sample frequency distribution is a desideratum – a convergence proof for the estimate series (114) is another.

Error assignement in small sample situations is not a privilege of observational astronomy. Attempts to attack the problem are known. This is, however, not the place to review the current situation.

As remarked in its second paragraph, Sect. 8 provides only an illustration of performance and precision issues. A full account, however desirable, does not fit into the frame of this work.

Finally, we state two perturbation problems. The instrument's PSF in effect in observation is known to be different from the one used in data analysis, here via MLE. Also, the background component obtained from the estimation cannot be regarded the best possible. So, we have to raise the problem of stability of MLE against PSF variation and against background variation.

Let $\hat{s}_j$, $j = 1, 2$, be the ML estimates of the source parameter vector with PSF_1, PSF_2 in effect, respectively. Does the convergence in probability $\hat{s}_1 \to \hat{s}_2$ take place for $PSF_1 \to PSF_2$? If the answer is affirmative, how fast is the convergence? Practicable measures of sensitivity are to be constructed. Replacing in the problem statement "PSF" by "background" yields the problem statement on background. Both affirmative answers lie at the heart of observational astronomy, known to be utterly succesful. But a formal problem solution tells you how precise PSF and background have to be modelled.

10. Conclusions

The MLE in SASS and EXSAS grew over the years towards a stable software system. Large quantities of ROSAT data were once or several times processed in the last decade with that software. The results can be found in many catalogues and in the literature. A detailed technical description has been given here. The likelihood functional for unbinned data was derived (almost) rigorously. The underlying assumptions were laid open (see below).

Equally, the assumptions and approximations made for the EXSAS implementation were represented. An example acquaints the reader with a typical MLE application. The performance limits of the implemented MLE were shown by simulations.

For clarity, we repeat the conditions for the application of the likelihood function L from (31) or (35) for a sample of single photons as follows.

The sample must be drawn from a population which follows a Poisson distribution with respect to the parameters to be estimated. No assumption on the geometrical form of the sample in the associated photon space is made. Thus no restriction of the instrument's PSF is imposed. Equally, no (non-trivial) lower or upper bounds on the sample size and photons per source exist for MLE as method. However, the statistical quality of the parameter estimates increase with increasing source photon numbers. The observed photon positions are imprecise due to binning. In the strict sense, the resulting MLE estimates are thus at least imprecise by half a bin size (plus an implementation-dependent additive).

The MLE is qualified by method properties for high energy astronomy observations with their notorious small numbers of photons per source.

11. Appendix: EXSAS aids for the ML source parameter estimation in ROSAT data

Some commands related to MLE of the source parameters shall be explained here. The command

$$DETECT/LOCAL \tag{119}$$

comprises the operations described in Sect. 4.1.2. The input data consist of up to six photon count histograms $\mathcal{N}$ from (46) and corresponding exposure histograms $\mathcal{T}$. The output is the source lists $\mathcal{L}_1(l_0)$ from (59). File names and remaining parameters, among them the significance level, L_0, and the dilatation exponent D are collected in a steering parameter file. Parameter files also control the other commands. The command

$$CREATE/BG\_IMAGE \tag{120}$$

produces the background count estimate $\hat{\Lambda}(X, Y)$ as described in Sect. 4.1.4 in histogram form $\mathcal{N}_B$. The input consists of up to six histograms $\mathcal{N}, \mathcal{T}$ and the allied source lists $\mathcal{L}_1(l_0)$. One of the parameters is the factor ρ

from (60). Having the background available, a modification of the command (119) called

$$DETECT/MAP \qquad (121)$$

can be executed. The input files are those of the command (119) supplemented by the background histograms $\mathcal{N}_B$. The result files are the source lists $\mathcal{L}_2(L_0)$ from (63). The command

$$MERGE/SOURCE\_TABLES \qquad (122)$$

carries out the union in (63). After the execution of the foregoing sequence of commands, all provisions are made to perform the MLE in narrow sense embodied in the command

$$DETECT/MAXLIK. \qquad (123)$$

The command

$$DETECT/SOURCES \qquad (124)$$

performs the sequence of the commands (119) to (123) in one command. This command is simpler in application but allows the user less control over parameter settings.

For sources with existence likelihoods $\hat{l}_S$ below the user-selected likelihood threshold l_S^0, upper confidence limits, $\hat{\lambda}_{S,\beta}^+$, on the source count rate λ_S for a selectable confidence level $\beta \in (0,1)$ are found with the command

$$COMPUTE/UPPER\_LIMITS. \qquad (125)$$

The input consists of the photon sample $\mathcal{P}$ from (1), the background histogram $\mathcal{N}_B$ from (62) and a list of positions (X_S^0, Y_S^0), specified in equatorial coordinates, with or without the associated source extent parameters ν_S^0 according to the chosen options. The user may keep separately: source position (X_S^0, Y_S^0), source extent parameter ν_S^0 (and source count rate λ_S^0 too). Two tables with source parameters form the output. One of them contains all sources, i.e. those with source existence likelihoods $\hat{l}_S > 0$. The strong sources defined by $\hat{l}_S \geq l_S^0$ are collected in a second table.

The command

$$SIMULATE/SOURCE \qquad (126)$$

simulates point sources. The underlying PSF density is a bivariate normal density with circular symmetry, see (64). With the output sample $\mathcal{P}$, the histogram $\mathcal{N}$ from (46) and an attitude table, the input to run the above commands is provided.

References

Avni, Y. 1976, ApJ, 210, 642

Aldrich, J. 1997, Statist. Sci., 12, 162

Boese, F. G. 2000, A&AS, 141, 507

Calabretta, M., & Greisen, E. W. 2000, Representation of Celestial Coordinates in FITS, http://www.aoc.nrao.edu/~egreisen/index.html

Cash, W. 1979, ApJ, 228, 939

Cruddace, R. G., Hasinger, G. R., & Schmitt, J. H. 1987, The application of a maximum likelihood analysis to detection of sources in the ROSAT data base, in ed. F. Murtagh, & A. Heck, Astronomy from Large Databases. Scientific objectives and methodological approaches, Proceedings of the Conference 12-14 October 1987, Garching, Germany, European Southern Observatory, 177

Doob, J. L. 1934, Trans. Amer. Math. Soc., 36, 759

Edwards, A. W. F. 1997, Stat. Sci., 12, 177

Hald, A. 1999, Stat. Sci., 14, 214

Harris, D. E., & Irwin, D. 1984, Einstein Observatory: Revised User's Manual, Harvard-Smithonian Center for Astrophysics, Cambridge MA

Lazzati, D., Campana, S., Panzera, M., & Tagliaferri, G. 1999, ApJ, 524, 423

Parzen, E. 1962, Stochastic Processes (Holden-Day, San Francisco)

Strong, A. W. 1985, A&A, 150, 273

Stuart, A., & Ord, K. 1991, Kendall's Advanced Theory of Statistics, vol. 2, 5th ed. (Arnold, London)

Trümper, J. 1983, Adv. Space Res., 241

Trümper, J. 1991a, ROSAT: Call for Proposals for the second observing period, August 1991 – February 1992, Technical Appendix. Max-Planck-Institut für extraterrestrische Physik, Garching

Trümper, J. 1991b, A&A, 246, L1

Voges, W. 1992, The ROSAT all-sky X-ray survey, in ed. T. D. Guyenne, & J. J. Hunt, Environment Observation and Climate Modelling through International Space Projects, Proceeding of the European International Space Year Conference 1992, esa ISY-3, Munich, Germany, 30 March – 4 April, Space in the service of the changing Earth (ESA Publication division, ESTEC, Nordwijk, The Netherlands), vol. 3, 9

Voges, W., Aschenbach, B., Boller, Th., et al. 1999, A&A, 349, 389

Voges, W., Boller, Th., Englhauser, J., Freyberg, M. J., & Supper, R. 2001, ROSAT's all-sky X-ray view, in Mining the sky, MPA/ESO Joint Astronomy Conference 31. 7. – 4. 8. 2000, Garching, Germany (in press)

Wald, A. 1943, Trans. Amer. Math. Soc., 54, 426

Wilks, S. S. 1938, Ann. Math. Stat., 9, 60

Zimmermann, U., Boese, G., Becker, W., et al. 1998, EXSAS User's Guide: Extended Scientific Analysis System to Evaluate Data from the Astronomical X-ray Satellite ROSAT, Max-Planck-Institut für extraterrestrische Physik, Garching bei München, Germany