

# Unbinned maximum-likelihood estimators for low-count data

## Applications to faint X-ray spectra in the Taurus molecular cloud

K. Arzner<sup>1</sup>, M. Güdel<sup>1</sup>, K. Briggs<sup>1</sup>, A. Telleschi<sup>1</sup>, M. Schmidt<sup>2</sup>, M. Audard<sup>3</sup>, L. Scelsi<sup>4</sup>, and E. Franciosini<sup>4</sup>

<sup>1</sup> Paul Scherrer Institut, 5232 Villigen, Switzerland  
e-mail: arzner@astro.phys.ethz.ch

<sup>2</sup> Institute for Data Analysis and Process Design, Zurich University of Applied Sciences, Postfach 805, 8401 Winterthur, Switzerland

<sup>3</sup> Columbia Astrophysical Laboratory, 550 West 120th St, MC 5247, New York, NY 10027, USA

<sup>4</sup> Dipartimento di Scienze Fisiche e Astronomiche, Piazza del Parlamento 1, 90134 Palermo, Italy

<sup>5</sup> INAF-Osservatorio Astronomico di Palermo, Piazza del Parlamento 1, 90134 Palermo, Italy

Received 27 January 2006 / Accepted 6 September 2006

### ABSTRACT

Traditional binned statistics such as  $\chi^2$  suffer from information loss and arbitrariness of the binning procedure, which is especially important at low count rates as encountered in the XMM-Newton Extended Survey of the Taurus Molecular Cloud (XEST). We point out that the underlying statistical quantity (the log likelihood  $L$ ) does not require any binning beyond the one implied by instrumental readout channels, and we propose to use it for low-count data. The performance of  $L$  in the model classification and point estimation problems is explored by Monte-Carlo simulations of Chandra and XMM-Newton X-ray spectra, and is compared to the performances of the binned Poisson statistic ( $C$ ), Pearson's  $\chi^2$  and Neyman's  $\chi^2_N$ , the Kolmogorov-Smirnov, and Kuiper's statistics. It is found that the unbinned log likelihood  $L$  performs best with regard to the expected chi-square distance between true and estimated spectra, the chance of a successful identification among discrete candidate models, the area under the receiver-operator curve of reduced (two-model) binary classification problems, and generally also with regard to the mean square errors of individual spectrum parameters. The  $\chi^2$  ( $\chi^2_N$ ) statistics should only be used if more than 10 (15) predicted counts per bin are available. From the practical point of view, the computational cost of evaluating  $L$  is smaller than for any of the alternative methods if the forward model is specified in terms of a Poisson intensity and normalization is a free parameter. The maximum- $L$  method is applied to 14 XEST observations, and confidence regions are discussed. The unbinned results are compared to binned XSPEC results, and found to generally agree, with exceptions explained by instability under re-binning and by background fine structures. In particular, HO Tau is found by the unbinned method to be rather cool ( $kT \sim 0.2$  keV), which may be a sign of shock emission. The maximum- $L$  method has no lower limit on the available counts, and allows to treat weak sources which are beyond the means of binned methods.

**Key words.** methods: statistical – X-rays: stars – stars: formation

## 1. Introduction

Energy histograms are often used as spectrum estimators. While this procedure has the advantage of simplicity, the grouping of counts into a histogram is also associated with information loss. This becomes especially important if only few counts are available, so that the spectral fine structures are sparsely sampled by the observed counts. Such a situation arises in the XMM-Newton Extended Survey of the Taurus Molecular Cloud (XEST; Güdel et al. 2007a), thus prompting the search for alternative unbinned methods.

Another motivation for histogram formation is to use  $\chi^2$  as a simple and simply coded measure of agreement between theory and observation. The role of  $\chi^2$  is thus to provide a plausibility ordering of alternative models, and to assess their absolute credibility. Since the histograms follow a multinomial distribution, which becomes gaussian only in the limit of infinite sample size, the use of  $\chi^2$  represents an approximation and corrections must be applied for small  $n$  (e.g., Kendall & Stuart 1958; Wachter 1979; Nousek 1989; Mighell 1999; Arzner 2004). Alternatively,

one may use the binned Poisson ( $C$ ) statistics. There exists, however, a simpler solution.

Namely, the relevant statistical quantity, the likelihood function, can be defined for an unlimited instrumental resolution without any reference to binning. Using this unbinned expression avoids arbitrariness of histogram formation, and thus a potential source of discrepant results. It has been successfully applied to ROSAT (Boese & Doebereiner 2001) and EGRET (Digel 2000) observations. Related approaches invoke piecewise-constant intensity models (Scargle 1998; see Stelzer et al. 2007, for an application to the XEST) or transformations to uniform null hypotheses (Kinoshita 2002), and applications in other fields than astronomy include particle physics (Baker & Cousin 1984) and medicine (Miller et al. 2002). While the  $C$  statistic is often used in astronomical applications (e.g., Dolphin 2002, for star formation statistics; Babu & Feigelson 1996, for a general overview), the unbinned (exact) Poisson likelihood is less often used. Alternative unbinned methods such as the Kolmogorov-Smirnov or Kuiper's statistics have the advantage of being (asymptotically) model-independent, which,

however, also entails sub-optimal performance if the modeling was correct.

In this article we revisit the issue by Monte-Carlo simulations of *XMM-Newton* and *Chandra* CCD spectra of the Taurus Molecular Cloud (TMC). By simulating counts from a known spectrum, and applying different statistics in order to find the best-fit model, we assess the performances of the various statistics, and thereby gain a more differentiated picture. The present study represents an extension of the work of Nousek (1989) and Wachter et al. (1979) to more complex spectral models and a broader range of statistics.

The article is organized as follows. Section 2 introduces the models for sources and background spectra. Section 3 discusses the counting statistics. Section 4 describes measures of agreement between models, and between models and data, which can be used to assess the performance of various statistics. Section 5 describes the Monte-Carlo simulations and their numerical results. Real-data applications from the TMC are presented in Sect. 6, where also the issue of confidence regions is briefly touched, and the unbinned estimates are compared to the binned estimates from the XSPEC software. Section 7 contains a summary and conclusions.

## 2. Spectral models

We start with introducing the spectral models  $f(E)$  used in this study, which represent the expected number of counts per unit energy.

In modern observations,  $E$  can usually not assume continuous values but is restricted to a discrete set of instrumental output channels. For example, the PN detector of *XMM-Newton* has by default  $N_c = 4096$  such channels; *Chandra*/ACIS has  $N_c = 1024$ . In both cases, the channel separation  $\delta E$  is much smaller than the instrumental resolution as given by the spectral response matrix, so that  $f(E)$  is fully resolved. In order to stress the quasi-continuous nature of the channel coverage, we shall write  $f(E) dE$  rather than  $f_j \delta E_j$  ( $j = 1 \dots N_c$ ) unless stated otherwise. The channels are also much finer than the bins typically used when computing the  $C$  or  $\chi^2$  statistics. In this sense, methods which directly use the energy channels will be referred to as “unbinned”, whereas methods which group the channels first into larger “bins” will be referred to as “binned”.

In what follows we exclusively work with counts and focus on the problem of finding the best spectral model for a given set of counts. Accordingly, we do not consider here the problem of spectral deconvolution, but absorb the instrumental response in the forward model  $f(E)$ . The symbol  $E$  thus denotes the observed (channel) energy rather than the incoming photon energy.

### 2.1. Source

Our source models assume absorbed collisional ionization equilibrium (APEC; Smith et al. 2001, as implemented in XSPEC; Arnaud 1996), convolved with the instrumental response. The templates are defined on the  $N_c$  instrumental channels, and depend on only two physical parameters: the (electron) temperature  $kT$  in units of keV, and the hydrogen column density  $N_H$  in units of  $10^{22} \text{ cm}^{-2}$ . This simple model is adapted to the faint sources addressed by the  $L$  statistic. The parameters ( $kT$ ,  $N_H$ ) are specified on a double logarithmic lattice with sufficiently dense coverage that intermediate spectra may be interpolated with error  $d_{\chi^2} < 10^{-5}$  (see Sect. 4.1). This gridding is used for simulations only. The total number of expected counts  $N_{\text{src}}$  is an

additional parameter. The normalized templates are displayed in Fig. 1 (top panel, selection). For better physical clarity, the  $x$ -axis refers to channel energy rather than to channel number. In the Monte-Carlo simulations, atomic lines are included assuming a fixed metallicity of 0.2 times the solar abundances of Anders & Grevesse (1989), and we use the *Chandra* rather than the *XMM-Newton* instrumental response because of its smaller number of instrumental channels, which accelerates the simulations. This choice applies to the simulations only, and we return to the *XMM-Newton* instrumental response and to a more refined abundance model when dealing with real XEST data. The function  $f(E)$  contains a background which is interpolated from calibration observations (see Sect. 2.2). The astrophysical relevance of the parameters ( $kT$ ,  $N_H$ ,  $N$ ) is elucidated in the accompanying article of Güdel et al. (2007a). From an empirical point of view, the effect of large  $N_H$  is to suppress (absorb) the spectrum from at low energies, and the effect of large  $kT$  is to enhance it mostly at high energies.

### 2.2. Background

Although *Chandra*, owing to its high spatial resolution, has little background, the present simulations include an additive background for the sake of generality and applicability to *XMM-Newton* data. The background is included in the forward model according to

$$f(E) = f_{\text{src}}(E) + f_{\text{bg}}(E) \quad (1)$$

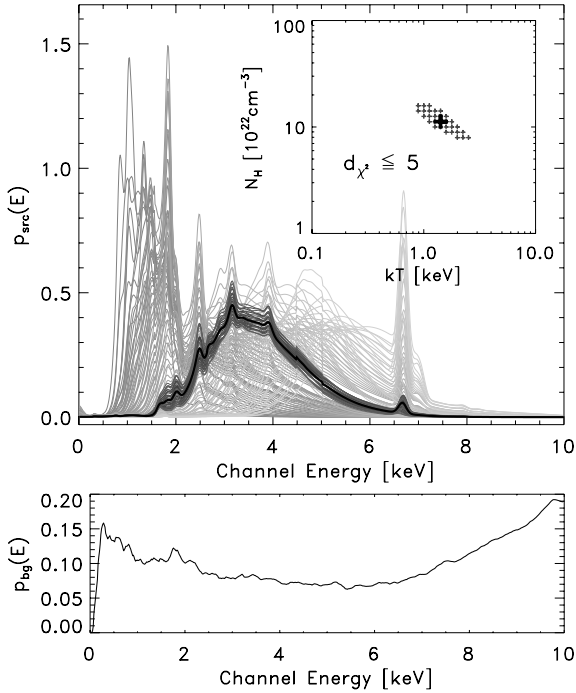
$$N = N_{\text{src}} + N_{\text{bg}}. \quad (2)$$

The background model  $f_{\text{bg}}(E)$  is obtained from an extended source-free extraction region. Since the background estimate is typically much better than the source estimate, the background is considered as known; this represents a simplifying assumption. The background density<sup>1</sup>  $f_{\text{bg},j}$  is determined from the observed background histogram  $H_{\text{bg},j}$  by minimizing the goal function  $\sum_{j=1}^{N_c} [f_{\text{bg},j} - H_{\text{bg},j}]^2 + \sum_{j=1}^{N_c-1} (\sigma(E_j)/\delta E_j)^2 [f_{\text{bg},j+1} - f_{\text{bg},j}]^2$ , where  $\sigma(E) = \kappa \sqrt{E/\text{keV}}$  represents the approximate instrumental resolution. The coefficient  $\kappa$  is 0.07 for *Chandra*/ACIS, and 0.05 for *XMM-Newton*/EPIC, and is chosen such that  $\sigma(E)$  gives the full width half maximum energy resolution of the respective instruments. The choice of the above goal function amounts to first-order regularization and yields a tridiagonal system which is easily solved numerically. It ensures<sup>2</sup> that  $f_{\text{bg},j}$  is unbiased in the (weak) sense that  $\sum_j f_{\text{bg},j} = \sum_j H_{\text{bg},j}$ , and that  $f_{\text{bg},j}$  has the approximately correct energy resolution in the sense that the response  $f_{\text{bg},j}^*$  to a unit pulse  $H_{\text{bg},j} = \delta_{j,k}$  has the mean-square width  $\sum_j (E_j - E_k)^2 f_{\text{bg},j}^* / \sum_j f_{\text{bg},j}^* \approx 2\sigma^2(E_k)$ .

To summarize, we use a smooth density estimator for the background which interpolates the observed background counts, preserves their total number, and has the (approximately) correct instrumental resolution. The normalized background

<sup>1</sup> For numerical conciseness, we use here channel-indexed ( $f_j \delta E_j$ ) rather than continuous ( $f(E) dE$ ) to notation.

<sup>2</sup> These properties are most easily seen from a continuous version where  $\int_B [(f_{\text{bg}}(E) - H(E))^2 + \sigma^2(E) f_{\text{bg}}'(E)^2] dE$  is minimized with respect to  $f_{\text{bg}}(E)$ . Carrying out the variation (with  $f_{\text{bg}}(E)$  kept fixed at the boundaries) one obtains  $f_{\text{bg}}(E) - (\sigma^2(E) f_{\text{bg}}'(E))' = H(E)$ . This may be integrated to  $\int f_{\text{bg}}(E) dE = \int H(E) dE$ , assuming that  $f'(E)$  vanishes at the boundaries. Multiplying the above differential equation by  $(E - E_k)^2$ , taking  $H(E) = \delta(E - E_k)$ , and assuming that  $\sigma^2(E)$  varies slower than  $f_{\text{bg}}(E)$ , one finds  $\int f_{\text{bg}}(E) (E - E_k)^2 dE \approx 2\sigma^2(E_k) \int f_{\text{bg}}(E) dE$  by integration by parts, q.e.d.



**Fig. 1.** *Top:* normalized source models used in this study. An example model (black) and its neighbours (dark gray) are highlighted for better clarity. Inlet: the example model and its neighbours in parameter space ( $kT$ ,  $N_H$ ). *Bottom:* normalized background model.

spectrum of a typical *Chandra*/ACIS observation is shown in Fig. 1 (bottom panel). The peak at 1.8 keV is due to silicon in the CCD.

### 3. Counting statistics

Photon counting experiments such as *XMM-Newton* and *Chandra* follow the Poisson statistics, which is briefly recalled here. More detailed discussions may be found e.g. in Feller (1968), Eadie et al. (1971), Santaló (1976), Wachter (1979), Reiss (1993), and Protassov et al. (2002).

#### 3.1. The Poisson process

We consider throughout a non-homogeneous Poisson process in an interval  $B = (E_{\min}, E_{\max})$  of the real axis with finite intensity  $f(E)$ . Generalizations to (pointwise) infinite  $f(E)$  and higher dimensions can be found e.g. in Reiss (1993). The Poisson process is then characterized by the following two properties: (i) the probability of finding  $n_j$  counts in a sub-interval  $\Delta_j \subset B$  is

$$\text{Prob}(n_j) = \frac{e^{-\lambda_j} \lambda_j^{n_j}}{n_j!} \quad \text{where} \quad \lambda_j = \int_{\Delta_j} f(E) dE, \quad (3)$$

and (ii) the numbers of counts in disjoint  $\Delta_j$ 's are statistically independent of each other. The first equation in (3) defines the Poisson distribution, and  $\lambda_j$  is called the parameter of the Poisson distribution.

When counts from different intensity functions  $f_k(E)$  are added, the resulting process is again Poissonian with intensity  $\sum_k f_k(E)$ . This fact may be used to include an observational background, and to construct a Poisson process of arbitrary overall expectation

$$\mathcal{N} = \int_B f(E) dE \quad (4)$$

by first drawing the number  $n$  of counts from a Poisson distribution (Eq. (3)) with parameter  $\mathcal{N}$ , and then drawing each count  $E_i$ ,  $i = 1 \dots n$  from the probability density

$$p(E) = f(E)/\mathcal{N} \quad (5)$$

(see Reiss 1993, Theorem 1.2.1). The above construction relies on the factorization of the number of counts from their position on the energy axis. The integral in Eq. (4) is numerically approximated as a sum over instrumental channels.

#### 3.2. Likelihood function and asymptotic forms

The likelihood function is defined as the probability of finding a certain observation given the true model from which the observation derives. Given the observed sample size  $n$ , the likelihood of the observation  $\{E_1, \dots, E_n\}$  is

$$P(E_1, \dots, E_n | kT, N_H) = p(E_1) \cdot \dots \cdot p(E_n). \quad (6)$$

Equation (6) represents a probability density at  $(E_1, \dots, E_n)$ . The shape of the function  $p(E)$  is determined by the parameters  $kT$  and  $N_H$ .

As outlined above, the normalization  $\mathcal{N}$  factorizes out, and can be estimated from the total number of observed counts alone. We shall do this using the maximum-likelihood estimator (i.e., that  $\mathcal{N}$  which maximizes  $e^{-\mathcal{N}} \mathcal{N}^n / n!$ ). Since the background contribution  $\mathcal{N}_{\text{bg}}$  is known (Sect. 2.2), only the source contribution  $\mathcal{N}_{\text{src}}$  needs to be estimated; this will be done using the maximum-likelihood estimator

$$\mathcal{N}_{\text{src}} = \max(0, n - \mathcal{N}_{\text{bg}}). \quad (7)$$

The exact likelihood (Eq. (6)) can be recast in alternative and asymptotic forms which are often used in astrophysics. When binned into finite bins  $\Delta_j$  of predicted content  $p_j = \int_{\Delta_j} p(E) dE$  containing  $n_j$  observed counts, the likelihood (given the total number of observed counts  $\sum_{j=1}^{N_b} n_j = n$ ) becomes the binomial distribution

$$P(n_1, \dots, n_{N_b} | kT, N_H) = \frac{n!}{n_1! \cdot \dots \cdot n_{N_b}!} p_1^{n_1} \cdot \dots \cdot p_{N_b}^{n_{N_b}}. \quad (8)$$

In the asymptotic limit ( $n_j \rightarrow \infty$ ), the standardized variables  $(n_j - np_j) / \sqrt{np_j}$  become normal and the quantity  $X^2 = \sum_j x_j^2$  becomes  $\chi^2$  distributed with  $(n - 1)$  degrees of freedom (Kendall 1958). Accordingly, the log likelihood (logarithm of Eq. (8)) becomes chi square distributed with  $(n - 1)$  degrees of freedom, and we may use the chi square statistic. The approximation of the logarithm of a multinomial by a chi square distribution is especially good when all the  $np_j$  are (approximately) equal; in this case (and only in this case), the  $np_j$  need not be large by themselves (Wise 1963). There are different forms of the chi square statistics, and we shall return to these – and to alternative statistics – in Sect. 4.2.

#### 3.3. Binning

The choice of optimal bins (Schott 1992) is an important issue which affects the outcome of the binned methods. Histograms can have uniform or non-uniform bins. While uniform (equal-size) bins have the advantage of being independent of the data, they are also less adapted and may contain very few counts. Non-uniform bins are usually chosen to contain the same (predicted) probability mass or observed counts.

For either (uniform or non-uniform) type of bins, one needs to decide on the number of bins or the average number of counts per bin. A well-known method to choose the number  $N_b$  of uniform bins is Sturges' rule (Sturges 1926)

$$N_b = 1 + \log_2 n \quad (9)$$

where  $n$  is the total number of observed counts. Sturges' rule is based on the assumption that the counts  $E_i$  follow a normal distribution, and tends to under-estimate the number of bins needed to resolve more complicated forms such as the  $f(E)$  considered here. An alternative approach, which is more adapted to the actual shape of  $p(E)$ , minimizes the asymptotic mean square integrated error (AMISE). This implies a trade-off between the integrated variance (due to finite counts per bin) and the integrated squared bias (due to variation of  $p(E)$  across the bins). Expanding  $p(E)$  to first order in each histogram bin, the minimum-AMISE rule yields the expression

$$h_* = (6/R(p'))^{1/3} n^{-1/3} \quad (10)$$

for the optimal bin width, where  $R(p') = \int_B p'(E)^2 dE$  measures the roughness of the spectrum (Scott 1992, theorem 3.3).

In our simulations, the optimal number  $n^*$  of counts per bin is either taken according to Eq. (9) as  $n^* = n/(1 + \log_2 n)$ , or according to Eq. (10) as  $n^* = n^{2/3} B^{-1} (6/\langle R(p') \rangle)^{1/3}$  where the (unknown) true  $R(p')$  has been replaced by an average  $\langle R(p') \rangle$  over all models under consideration. The choices based on Eqs. (9) and (10) will be referred to as Sturges' and AMISE methods. Alternatively,  $n^*$  is fixed at a given value, similar as in the standard *XMM-Newton* and *Chandra* data analysis software. This will be referred to as fixed- $n^*$  method; a typical choice is 10 counts per bin. (In XSPEC, this becomes  $\geq 10$  cts/bin for high-intensity flares which are not considered here.) For the problem considered here ( $\langle R(p') \rangle = 2.9$  cts/keV<sup>3</sup>;  $N$  ranging from a few 10 to a few 100), the AMISE method gives less than 10 counts per bins, whereas Sturges' method gives a few to a few ten counts per bin.

Once the number of counts per bin  $n^*$  is chosen, the binning itself is performed using either uniform or non-uniform bins. For the case of non-uniform bins we use by default bins with equal numbers of observed counts, with the  $k$ -th bin running from  $E_{kn^*}$  to  $E_{(k+1)n^*}$ . This simple method is not optimal but adopted for consistency with the standard analysis software. Alternative methods have also been considered, which are based on equal predicted bin content, and found to yield similar results.

## 4. Distance measures

Next we define measures of agreement between different models, and between models and observations. These will allow to estimate models from the data, and to assess the discrepancy between true and estimated models in the subsequent Monte-Carlo simulations.

### 4.1. Chi-square distance between models

Different sets of parameters ( $kT$ ,  $N_H$ ,  $N$ ) result in different model spectra, whose discrepancy can be measured by a suitable distance in the space of predicted spectra ( $f$ -space). It is important to realize that it is the distance in  $f$ -space which is relevant for the observational discrimination between alternative models, and not the (say, Euclidean) distance in parameter space, because the latter may degenerate solely due to ambiguous parameterization. Recalling that  $f(E)$  is the intensity of a Poisson process,

and assuming that  $f_1(E)$  is the true and  $f_2(E)$  is the estimated spectrum, we use here the chi square distance (e.g., Gibbs 2002)

$$d_{\chi^2}(f_1, f_2) = \int_B \frac{(f_1 - f_2)^2}{f_1} dE. \quad (11)$$

An illustration of  $d_{\chi^2}$  is given in Fig. 1, where the dark gray spectra deviate by  $d_{\chi^2} \leq 5$  from the reference spectrum (black), assuming that  $N = 100$ . The parameters  $kT$  and  $N_H$  are not independent, as indicated by the banana-shaped region in Fig. 1 (inlet). It should be pointed out that it is *not* this kind of degeneracy which is addressed in the present paper, but rather the principal capability of various statistics to distinguish between models which differ in the sense of Eq. (11). Note that  $d_{\chi^2}(f_1, f_2)$  is not a metric since it is not symmetric in its arguments, and also violates the triangle inequality with respect to the first argument  $f_1$  (but not with respect to the second argument  $f_2$ ). Alternative probabilistic distances may be found e.g. in Basseville (1989), Rachev (1991), Reiss 1993, Müller (1995), O'Sullivan et al. (1998), Robinson et al. (2000), Johnson & Sinanovic (2001), and Gibbs & Su (2002). We have chosen here Eq. (11) because it does not require probabilistic normalization of  $f$ , and because it is not explicitly adapted to the  $L$  statistics, the performance of which is to be demonstrated.

### 4.2. Distance between models and measurements

After having defined a measure of agreement between different models, we shall specify measures of agreement between models and observations. To this end, we consider the unbinned likelihood (Eq. (6)) together with a selection of some of the statistics which are often used in an astrophysical and astronomical context (e.g., Gosset 1987; Wachter et al. 1979; Nousek & Shue 1989; Babu & Feigelson 1996; Mighell 1999; Metchev 2002; Paltani 2004):

$$L = \sum_{i=1}^n \ln p(E_i) \quad (12)$$

$$C = \sum_{j=1}^{N_b} (n_j \ln \lambda_j - \lambda_j) \quad \text{with} \quad \lambda_j = \int_{\Delta_j} dE f(E) \quad (13)$$

$$\chi^2 = \sum_{j=1}^{N_b} \frac{(n_j - \lambda_j)^2}{\lambda_j} \quad (14)$$

$$\chi_N^2 = \sum_{j=1}^{N_b} \frac{(n_j - \lambda_j)^2}{n_j} \quad (n_j \neq 0) \quad (15)$$

$$D = \max_E |P(E) - S_N(E)| \quad \text{with} \quad P(E) = \int_{E_{\min}}^E p(E') dE' \quad (16)$$

$$V = \max_E (P(E) - S_N(E)) + \max_E (S_N(E) - P(E)). \quad (17)$$

Above,  $L$  is the unbinned log likelihood (logarithm of Eq. (6)),  $C$  is the  $C$  statistic (Cash 1979), and  $D$  is the Kolmogorov-Smirnov (or uniform) distance between two probability densities, of which the Kuiper statistic  $V$  is a variant with more balanced sensitivity across  $[E_{\min}, E_{\max}]$  (Kuiper 1962).  $P(E)$  is the predicted cumulative distribution, and  $S_N(E)$  is its observed counterpart, i.e.,  $S_N(E)$  is the number of counts with energy smaller than  $E$ , normalized by the total number  $n$  of counts (thus  $0 \leq S_N(E) \leq 1$ ).  $N_b$  denotes the number of bins of the binned methods (Sect. 3.3). The subscript "N" in Eq. (15) stands for

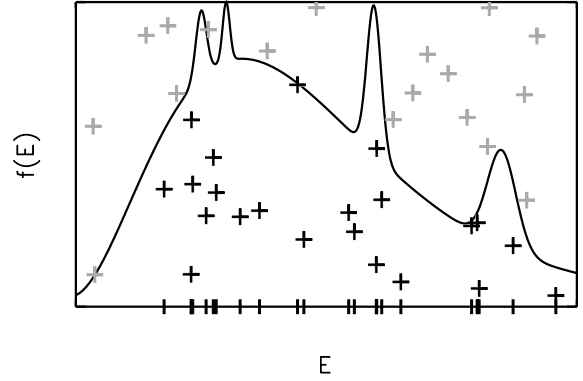
Neyman; it is undefined for bins with  $n_j = 0$  which are therefore excluded. (The reference to Neyman is for his detailed analysis; the replacement of theoretical variances by observed ones has already been proposed by Pearson – see Hald 1998). Note that evaluation of  $L$  is computationally less expensive than evaluation of the binned statistics when there are fewer counts than channels and if the model provides  $p(E)$  rather than  $P(E)$ .

It should be pointed out that the selection of statistics (Eqs. (12)–(17)) is aimed at our astrophysical application of estimating X-ray spectra, and is not free of personal bias. The  $L$  statistic is the one which is primarily addressed by the present article. The  $C$  and  $\chi_N^2$  statistics are implemented in the standard XSPEC software package; they are included for benchmarking and comparison of observational results. The  $\chi^2$  statistic is an obvious variant of  $\chi_N^2$ , which is obtained as the asymptotic limit of the binned Poisson log likelihood (cf. Sect. 3.2). Among the unbinned statistics, the Kolmogorov-Smirnov  $D$  is probably the most fundamental one used in astrophysics. The Kuipers statistic  $V$  has been included as an example of an improved version of  $D$ . Others (such as the Anderson-Darling statistic) could have been included as well, but we decided to restrict the list in order not to overload the diagnostics.

In order to treat all statistics (12)–(17) on the same footing, the source normalization  $\mathcal{N}_{\text{src}}$  is always estimated from Eq. (7), and only the shape parameters ( $kT$ ,  $N_{\text{H}}$ ) are estimated differently for each statistic. The case  $\mathcal{N}_{\text{src}} = 0$  is rarely met in practice.

### 4.3. Receiver-operator characteristics

Equation (11) can be used to measure the discrepancy between true and estimated models, and thus to assess the performance of the statistics (Eqs. (12)–(17)). Alternatively, their performance can also be characterized in terms of the classical receiver-operator characteristics (ROC; Peterson et al. 1954; Hanley & McNeill 1982; Michel 2003). This diagnostics applies to the binary classification problem, and visualizes the trade-off between the reliability and completeness of detections. In its standard formulation, the binary classification problem considers two types of objects (healthy and ill patients, for example) with a real-valued attribute. The attribute is a random variate, the probability density of which depends on the object type. The task is to classify the objects by their attributes, which is most simply done by applying a threshold. The classification is only unique if the probability densities of the attribute do not overlap. There are four possible outcomes of the classification procedure, with frequencies indicated in brackets: a type 1 object may be correctly classified as type 1 ( $n_{11}$ ) or erroneously as type 2 ( $n_{12}$ ), and a type 2 object may be correctly classified as type 2 ( $n_{22}$ ) or erroneously as type 1 ( $n_{21}$ ). From these frequencies we define the sensitivity  $n_{11}/(n_{11} + n_{12})$  and the specificity  $n_{22}/(n_{21} + n_{22})$ . (The asymmetric naming stems from associating type 1 and 2 with asymmetric entities like healthy and ill patients.) Let us assume that a low attribute is typical for type 1, and a high attribute is typical for type 2. If the threshold is high, only few objects are accepted, but most of them are correctly classified as type 2. However, the bulk of type 2 objects is missed since it interferes with type 1. Thus the specificity is high but the sensitivity is small. Conversely, a low threshold detects all type 2 objects, but at the expense of type 1 contamination. In this case, the sensitivity is high but the specificity is low. As the threshold is varied, it traces out a trade-off between specificity and sensitivity that is traditionally displayed in a (1–specificity) versus sensitivity diagram, equivalent to false positive rate versus true positive rate. This graph is called the receiver-operator characteristics (ROC).



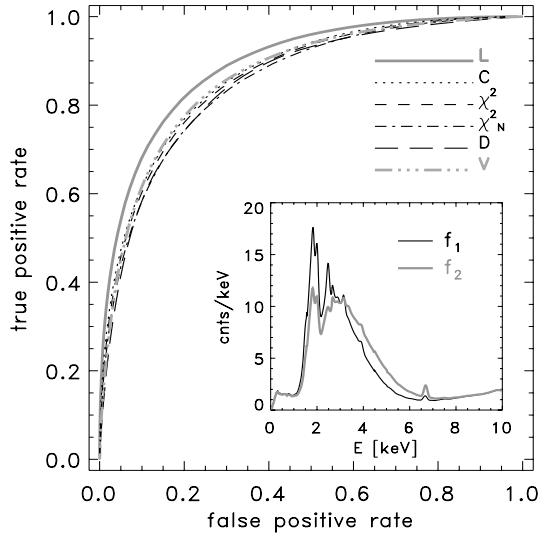
**Fig. 2.** Schematic construction of a continuous non-homogeneous Poisson process by the rejection method: out of  $N_p$  uniform random points in the rectangle  $B \times [0, \max(f)]$  only those beneath the curve  $f(E)$  are accepted; their abscissae constitute the event list (ticks). The number  $N_p$  itself is Poisson distributed with expectation value  $|B| \times \max(f)$ .

In order to apply the ROC procedure to the *XMM-Newton* spectrum estimation problem, we restrict the latter to only two possible models  $f_1$  and  $f_2$ , and use the log likelihood ratio  $L_2 - L_1$  as an attribute, and similarly  $C_2 - C_1$ ,  $\chi_1^2 - \chi_2^2$ ,  $\chi_{N,1}^2 - \chi_{N,2}^2$ ,  $D_1 - D_2$ , and  $V_1 - V_2$ . By convention, a low attribute thus hints to model 1 and a high attribute to model 2. We then choose at random one of the two models, generate an eventlist, and compute the above attributes. This procedure is repeated some  $10^5$  ( $= \sum_{ij} n_{ij}$ ) times, and the sensitivity and specificity are computed for varying thresholds. The resulting ROC curves are shown in Fig. 3. The two models  $f_1$  and  $f_2$  have the same hydrogen column density  $N_{\text{H}} = 4.5$  and normalization  $\mathcal{N} = 30$  but their temperatures differ by a factor two,  $kT_1 = 1.4$  keV and  $kT_2 = 2.5$  keV. Such a difference is within astrophysical expectations. The chi square distance between the two models is  $d_{\chi^2}(f_1, f_2) = 3.2$ . Different curves represent different statistics. Ideally (100% sensitivity and 100% specificity), the ROC curve would go through the upper left corner (0,1) of Fig. 3, whereas a completely non-discriminating test would yield a straight line along the diagonal from (0,0) to (1,1). The actually realized curves are between the two extremes, with the exact likelihood coming closest to the ideal point (0,1).

The ROC curve can be used to assess the performance of the statistics (Eqs. (12)–(17)) in the binary classification problem. A commonly used performance measure is the area under the ROC curve (AUC). Like the chi square distance, the AUC gives a measure of observable discrepancy between two models. For Fig. 3 we obtain the AUC's 0.761 ( $L$ ), 0.705 ( $C$ ), 0.679 ( $\chi^2$ ), 0.655 ( $\chi_N^2$ ), 0.699 ( $D$ ), and 0.727 ( $V$ ), thus establishing the ordering  $L > V > C \gtrsim D > \chi^2 > \chi_N^2$ .

## 5. Monte-Carlo simulations

In order to explore the performance of the unbinned likelihood and compare it to the alternative statistics we have conducted extended Monte-Carlo simulations. The simulations involve two general steps. In the first step, a model is chosen at random and an event list is generated by either the inversion or rejection method (Devroye 1986). The inversion method is described in Appendix A, and the rejection method is illustrated in Fig. 2. The inversion method is faster and thus used in most simulations; both methods have been checked against the analytical Poisson probabilities. In the second step, we compare the event list with models which are close enough to the true model in a  $d_{\chi^2}$  sense



**Fig. 3.** Receiver-operator characteristics for the binary classification problem with only two models (inlet).

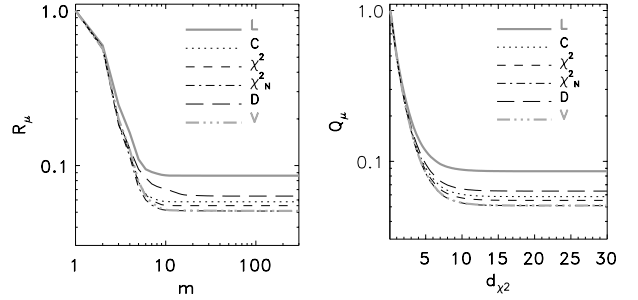
in order to be potentially successful candidates. This is done using all statistics listed in Eqs. (12)–(17); the decision rule is to accept the model with largest ( $L, C$ ) or smallest ( $\chi^2, \chi^2_N, D, V$ ). The above two-step procedure is repeated for many realizations of the event list, and several diagnostics are applied. When the parameter  $N_{\text{src}}$  is varied in the simulations, this is done by taking it uniformly distributed out of the indicated interval. The background  $N_{\text{bg}}$  is kept fixed. Several runs addressing different aspects have been performed.

### 5.1. Safe discrimination

In a first family of simulations, we investigate the minimum chi-square distance between two models which allows a safe distinction on grounds of the observed event list.

To this end, we work with the discrete set of models of Sect. 2.1 and call an estimate a “successful identification” if the estimated model equals the true model. We are interested in the chance of a successful identification, and proceed as follows. For each realization  $\mathcal{E}$  of the eventlist drawn from a true model  $f_1$ , the candidate models  $\{f_2\}$  are ordered by increasing  $d_{\chi^2}(f_1, f_2)$ , and examined sequentially on grounds of the statistics (12)–(17). The first candidate is always taken as the true model. As the number  $m$  of candidates increases, the initial (successful) estimate may be abandoned in favour of a wrong estimate. Let us assume that this happens, under statistics  $\mu$  and for the realization  $\mathcal{E}$ , at the  $k$ -th candidate and define  $\theta_\mu^\mathcal{E}(m)$  to be one for  $m = 1 \dots k$  and zero for  $m > k$ . When the simulation is repeated for a large number of realizations  $\mathcal{E}$  (using different  $f_1$ ), the quantity  $R_\mu(m) = S_\mu(m)/S_\mu(1)$  with  $S_\mu(m) = \sum_{\mathcal{E}} \theta_\mu^\mathcal{E}(m)$  gives the probability of a successful identification in a search over  $m$   $d_{\chi^2}$ -ordered candidates using statistics  $\mu$ .

An example of  $R_\mu(m)$  is shown in Fig. 4 (left), with different curves referring to different statistics  $\mu$ . The binned methods invoked Sturges’ rule to determine the number of counts per (equal-size) bin. As increasingly unlikely candidates become included, the  $R_\mu(m)$  converge to a constant values. This holds true for all statistics; however, there is a clear performance ordering  $R_L > R_D > R_C > R_{\chi^2} > R_V \approx R_{\chi^2_N}$ . The exact likelihood performs best, followed by the Kolmogorov-Smirnov distance.



**Fig. 4.** The probability of successful identifications versus the number of  $d_{\chi^2}$ -ordered candidate models (left), and versus the maximal distance of the candidates from the true model (right). See text. The simulation has  $10 < N_{\text{src}} < 100$ ,  $N_{\text{bg}} = 0$ , and the binned methods use equal-size bins based on Sturges’ rule, resulting on average in  $\langle n^* \rangle = 9.1$  counts per bin.

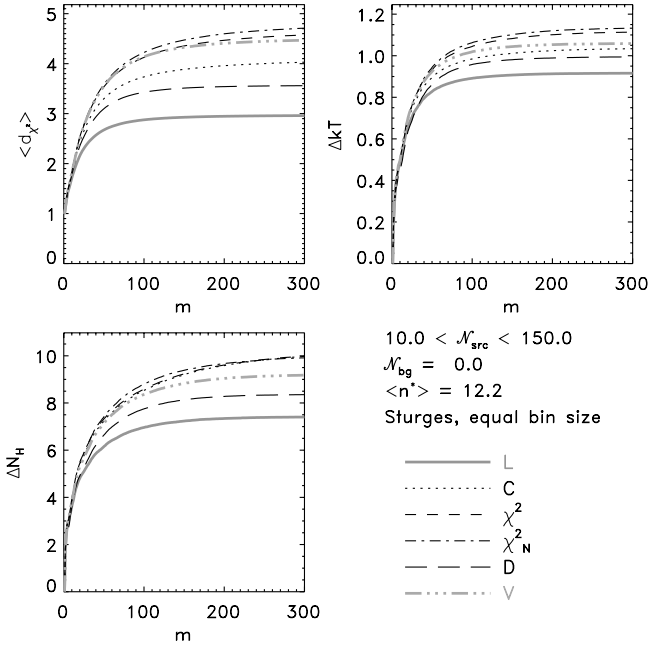
While the quantity  $R_\mu(m)$  directly relates to the simulation procedure, the number  $m$  by itself is not of interest and should be replaced by  $d_{\chi^2}$  in order to obtain a more meaningful characteristic. This is achieved by replacing the order indicator  $\theta_\mu^\mathcal{E}(m)$  by a distance indicator  $\theta_\mu^\mathcal{E}(d_{\chi^2})$  which switches from unity to zero at the distance of the first erroneously accepted candidate, and proceeding in the same way as for  $R_\mu(m)$ . As a result we obtain the probability  $Q_\mu(d_{\chi^2})$  of a successful identification in a  $d_{\chi^2}$ -ordered search among (discrete) candidates with maximal distance  $d_{\chi^2}$  from the true model. The graphs of  $Q_\mu(d_{\chi^2})$  are shown in the right panel of Fig. 4, referring to the same simulation as Fig. 4 (left). Like  $R_\mu(m)$ ,  $Q_\mu(d_{\chi^2})$  converges at large  $d_{\chi^2}$  to a constant value which depends only on the statistics used. The performance ordering derived from  $Q_\mu(d_{\chi^2})$  is the same as for  $R_\mu(m)$ .

Aside from the success rates one may ask for the error in parameter- and  $f$ -space if the identification fails. The closer the estimated solution is to the true one, the better the method. Figure 5 displays the residuals in  $f$ - and parameter space for the same situation as in Fig. 4 (left); the quantities  $(\Delta kT, \Delta N_H)$  are the standard deviations between true and estimated model parameters. They give the typical accuracy of best-fit parameters, averaged over the whole parameter space. As can be seen from Fig. 5, the exact likelihood yields the smallest  $\langle d_{\chi^2} \rangle$  (top left) and also the smallest parameter errors  $(\Delta kT, \Delta N_H)$ . The error in the normalization estimate (Eq. (7)) is simply  $\Delta N = \sqrt{N}$  and is not shown.

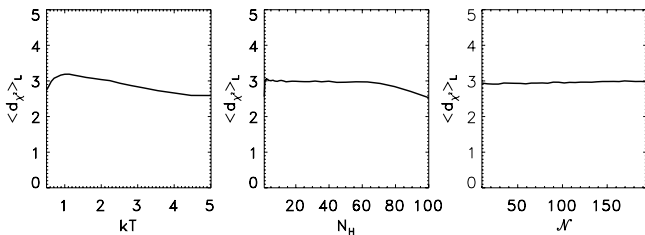
The stabilization of the curves observed in Figs. 4 (right) and 5 (top left) can be used to define a maximum chi-square distance  $d_{\chi^2}^{\text{max}}$  between true and estimated models above which mis-estimation becomes highly unlikely. From Fig. 5 (top left) we find that  $\langle d_{\chi^2} \rangle_L$  stabilizes around 3, whereas Fig. 4 (right) indicates that erroneous estimates almost never occur for  $d_{\chi^2} \gtrsim 15$ . This behaviour is generic and holds true for a wide range of model parameters (Fig. 6), suggesting that the chi square distance is a useful parameter-free measure of distance between two Poisson intensities. We thus select

$$d_{\chi^2}^{\text{max}} = 50 \quad (18)$$

as a safe cutoff for potentially successful candidate models. Candidate models which differ by more than  $d_{\chi^2}^{\text{max}}$  from the true model are not considered. The error introduced by this neglect is small; from the decay of the histogram of best-fit distances  $d_{\chi^2}$  we have estimated that less than  $10^{-6}$  of all realizations are concerned.



**Fig. 5.** Average  $f$ -space (top left) and parameter (other panels) deviations between true and estimated models. The number  $m$  of candidate models delineates the tested volume in parameter space.



**Fig. 6.** Average chi-square distance between true and estimated models as a function of the true model parameters (marginal distributions).

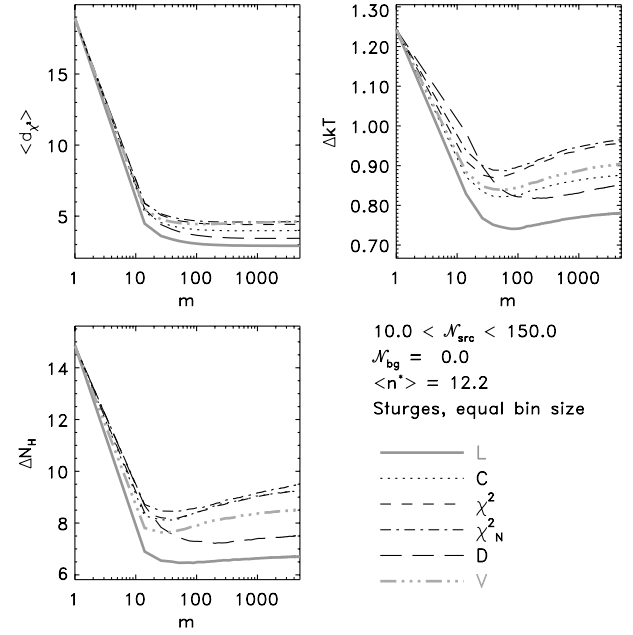
## 5.2. Resolution

In a second family of simulations, the restriction to the discrete models of Sect. 2.1 is relaxed by interpolating the models in  $(kT, N_H)$ -space. Equation (18) is then used to limit the search for best-fit candidates. The number  $m$  of candidate models thus delineates the density of models in parameter space, and as  $m$  is increased, the best-fit candidates converge to the best-fit solutions. Figure 7 shows the result of this density exploration. Note that the  $\langle d_{\chi^2} \rangle$  curves converge at large  $m$  to the same values as in Fig. 5, confirming that  $d_{\chi^2}^{\text{max}}$  and the maximal  $m$  are ample and do not effect the outcome of the simulation.

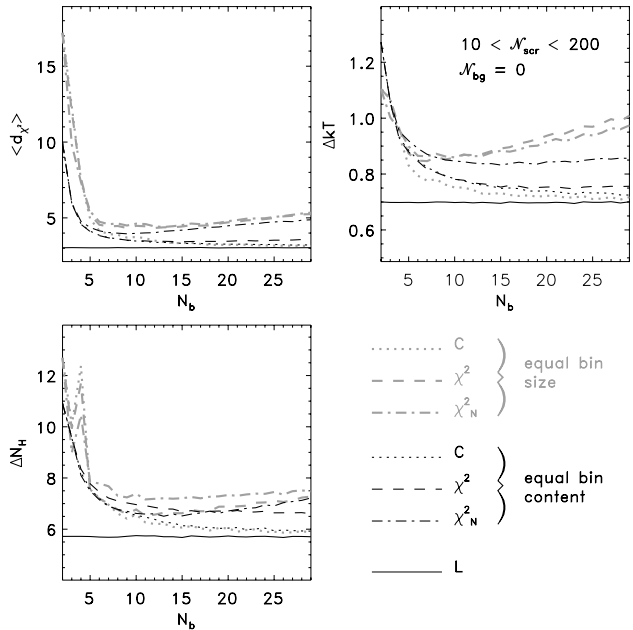
## 5.3. Effect of binning, count rate, and background

Up to here the binned methods invoked only Sturges' rule (Eq. (9)) for better comparability. In this Section, the effect of different binning methods and of the bin size is explored. The case of a single bin must be excluded, since the predicted bin content is then insensitive to the parameters  $kT$  and  $N_H$ .

In order to demonstrate the effect of the bin size we arbitrarily vary the number of bins  $N_b$  (Fig. 8). Black curves refer to equal-content bins, and gray refer to equal-size bins. The curves represent averages over  $(kT, N_H)$  and  $10 < \mathcal{N} < 200$ . At small  $N_b$  (many counts/bin), the  $(C, \chi^2, \chi^2_N)$  curves collapse for both equal-content and equal-size cases. At large  $N_b$  (few

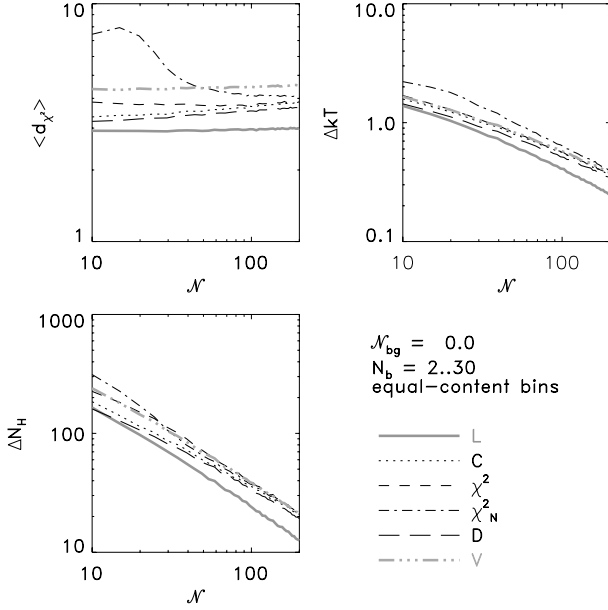


**Fig. 7.** Similar to Fig. 5 but with  $m$  proportional to the density of candidate models in parameter space.



**Fig. 8.** The effect of binning on the performances of  $(C, \chi^2, \chi^2_N)$ . The corresponding unbinned  $L$  values are shown for comparison.

counts/bin), the performance of  $C$  monotonically approaches the  $L$  limit (black solid line). For large  $N_b$ , the  $\chi^2$  and  $\chi^2_N$  statistics are not applicable due to low count rates. As a consequence, their performance does not approach the  $C$  and  $L$  limit with increasing  $N_b$ . This is especially pronounced for the equal-sized bins, which generally perform worse than equal-content bins. From Fig. 8 one may deduce the minimum number of counts per bin at which the chi square statistics apply: the increase of the  $(\chi^2, \chi^2_N)$  curves at  $N_b \gtrsim 10$  indicates that at least some 5–20 counts per bin should be present. By narrowing and shifting the  $\mathcal{N}$ -cut one may explore the  $\mathcal{N}$ -dependence of the turnaround, and thereby find that  $\chi^2$  and  $\chi^2_N$  require about 10 and 15 counts per bin, respectively.



**Fig. 9.** Dependence of the estimation errors on the true expected counts  $N$ .

Complementary to Fig. 8, Fig. 9 shows the effect of  $N$ , averaged over  $N_b$ . Again,  $N_b$  is independent of  $N$  for demonstration purposes. The bins have equal content. The  $f$ -space errors (top left) are found to be weakly dependent on  $N$  except for the  $\chi^2_N$  statistics, where the neglect of empty bins has a significant effect at low count rates ( $N \lesssim 40$ ). The errors  $\Delta kT$  and  $\Delta N_H$  scale like  $N^{-1/2}$ , which is inherited from the normalization error  $\Delta N \sim N^{1/2}$  (not shown).

We dismiss now the ad-hoc choice of  $N_b$  used in Figs. 8 and 9, and return to the better adapted binning methods of Sect. 3.3. The number of counts per bin is thus determined from the Sturges’ or AMISE rules, or fixed at  $n^* = 8$ ; the bins have either equal size or equal (observed) content. The result of this simulation family is summarized in Table 1, showing the performance ranking based on the mean chi square distance between true and estimated models. All simulations involve  $10^5$  realizations, and the true source normalization is either  $N_{src} = 20$  or  $N_{src} = 150$ , while the parameters  $(kT, N_H)$  are continuous (interpolated). A background of  $N_{bg} = 50$  has been included for the sake of generality, so that both background-dominated and source-dominated situations are addressed. From Table 1 it is seen that the exact likelihood ( $L$ ) performs best and the Kolmogorov-Smirnov distances ( $D$ ) performs worst. The relatively bad performance of  $D$  compared to Sect. 5.1 is found to be caused by the background. Pearson’s  $\chi^2$  generally performs better than Neyman’s  $\chi^2_N$ ; Kuiper’s  $V$  performs better than Kolmogorov’s  $D$ . A detailed comparison of pairs of simulations differing only by the count rate reveals that the chi square distance between true and estimated models is always smaller for  $N = 150$  than for  $N = 20$ , which is expected since more counts allow a sharper distinction of the models. Equal-content bins generally perform better than equal-size bins.

When the area under the operator-receiver curve (AUC; Sect. 4.3) is used as an alternative performance measure instead of the average chi square distance between true and estimated models, we obtain the results of Table 2. This is similar to Table 1 but with  $A > B$  indicating that  $\langle AUC \rangle_A < 0.99 \langle AUC \rangle_B$  and  $A \gtrsim B$  indicating  $0.99 \langle AUC \rangle_B < \langle AUC \rangle_A < \langle AUC \rangle_B$  (like small  $\langle d_{\chi^2} \rangle$ , small  $\langle AUC \rangle$  indicates close agreement of true and

**Table 1.** Performance ordering of the statistics (Eqs. (12)–(17)), based on the average chi square distance between true and estimated models. The notation  $A > B$  indicates here that  $\langle d_{\chi^2} \rangle_A < 0.95 \langle d_{\chi^2} \rangle_B$ , and  $A \gtrsim B$  indicates that  $0.95 \langle d_{\chi^2} \rangle_B < \langle d_{\chi^2} \rangle_A < \langle d_{\chi^2} \rangle_B$ . The labels “e.s.” and “e.c.” refer to “equal size” and “equal count” bins, respectively.  $N_{bg} = 50$ .

method	binning	$N_{src}$	
Sturges	e.s.	20.0	$L > \chi \gtrsim C \gtrsim \chi_N \gtrsim V > D$
Sturges	e.s.	150.0	$L > V \gtrsim \chi \gtrsim C > \chi_N \gtrsim D$
Sturges	e.c.	20.0	$L > C \gtrsim \chi \gtrsim \chi_N > V > D$
Sturges	e.c.	150.0	$L > C \gtrsim \chi_N \gtrsim \chi > V > D$
AMISE	e.s.	20.0	$L \gtrsim \chi \gtrsim C > V \gtrsim \chi_N > D$
AMISE	e.s.	150.0	$L \gtrsim C > \chi > V > \chi_N > D$
AMISE	e.c.	20.0	$L \gtrsim C > \chi \gtrsim \chi_N > V > D$
AMISE	e.c.	150.0	$L \gtrsim C > \chi \gtrsim \chi_N > V > D$
$n^*=8.0$	e.s.	20.0	$L > \chi \gtrsim C > \chi_N > V > D$
$n^*=8.0$	e.s.	150.0	$L > C \gtrsim \chi > V \gtrsim \chi_N > D$
$n^*=8.0$	e.c.	20.0	$L > C \gtrsim \chi \gtrsim \chi_N > V > D$
$n^*=8.0$	e.c.	150.0	$L \gtrsim C \gtrsim \chi \gtrsim \chi_N > V > D$

**Table 2.** Similar to Table 1, but using the area under the receiver-operator curve (Fig. 3) as a measure of performance.  $N_{bg} = 50$ .

method	binning	$N_{src}$	
Sturges	e.s.	20.0	$L > V \gtrsim C \gtrsim \chi > \chi_N > D$
Sturges	e.s.	150.0	$L > V \gtrsim C \gtrsim \chi > D \gtrsim \chi_N$
Sturges	e.c.	20.0	$L > C \gtrsim \chi > \chi_N \gtrsim V > D$
Sturges	e.c.	150.0	$L > C \gtrsim \chi \gtrsim V \gtrsim \chi_N > D$
AMISE	e.s.	20.0	$L > C > \chi > V > \chi_N > D$
AMISE	e.s.	150.0	$L \gtrsim C > \chi > V > D \gtrsim \chi_N$
AMISE	e.c.	20.0	$L > C > \chi > V \gtrsim \chi_N > D$
AMISE	e.c.	150.0	$L \gtrsim C > \chi > V \gtrsim \chi_N > D$
$n^*=8.0$	e.s.	20.0	$L > C > \chi > V > \chi_N > D$
$n^*=8.0$	e.s.	150.0	$L \gtrsim C > \chi > V > D \gtrsim \chi_N$
$n^*=8.0$	e.c.	20.0	$L > C > \chi > \chi_N > V > D$
$n^*=8.0$	e.c.	150.0	$L \gtrsim C > \chi > \chi_N \gtrsim V > D$

estimated models). The average is over  $10^4$  pairs of models covering the (continuous) parameter space  $(kT, N_H)$ , with each pair of models being probed by  $2 \times 10^4$  realizations of Poisson data. Although the AUC criterion is quite different in spirit from the chi square distance criterion (binary classifier versus estimation problems), the performance orderings obtained by the two methods are surprisingly similar. In particular, the unbinned likelihood ( $L$ ) performs always best, while the internal performance ordering of  $(C, \chi^2, \chi^2_N, D, V)$  partially changes. It was verified that  $L$  performs best not only on the average, but also for all model pairs individually.

Figures 4–9 did not include background ( $N_{bg} = 0$ ). In order to investigate the effect of the background we have repeated the simulations with variable background ratio  $N_{bg}/N$ . It turns out that the presence of a background does not affect superior performance of the unbinned likelihood  $L$ . There are, however, differences in the performance ranking of the other statistics; in particular, the unbinned  $D$  and  $V$  statistics are found to be degraded by the background. We shall not show here the full diagnostic applied to the zero-background case, but restrict ourselves to an exemplary result. Figure 10 shows the number of successful identifications versus the number of models at choice for  $N_{bg} = 50$ , and is to be compared to Fig. 4. The lead of the  $L$  statistic is even more pronounced in the presence of background; the performance ordering derived from Fig. 10 is  $L > V \gtrsim C \gtrsim \chi^2 \gtrsim \chi^2_N > D$ .



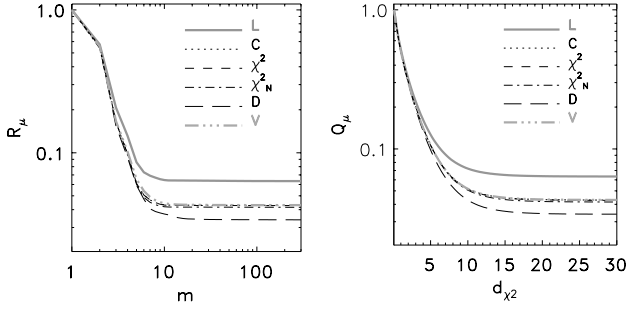


Fig. 10. Similar to Fig. 4, but including background ( $N_{\text{bg}} = 50$ ).

## 6. Real-data application

### 6.1. Observations and forward model

After establishing the superior performance of the unbinned likelihood by Monte-Carlo simulations, we apply it to actual *XMM-Newton* observations of the TMC. The data set (Table 3) comprises 14 objects, and has been selected for low numbers of observed counts ( $N_{\text{cnt}}$ ) and simplicity of the spectra so that parameterisation by  $(kT, N_{\text{H}}, N)$  is adequate. Also, we have as far as possible avoided cases where the background is not well known, and where the maximum-likelihood parameters lie on the boundary of the XSPEC parameter space, which would require a more careful analysis (Protassov et al. 2002). “Bad” time intervals with increased background were omitted.

The source models are similar as in Fig. 1 (top) but adapted for the *XMM-Newton* instrumental response and background. The background spectrum is estimated according to the procedure of Sect. 2.2, setting  $\kappa = 0.05$  to characterize the approximate *XMM-Newton*/EPIC resolution. Only the PN detector of the *XMM-Newton*/EPIC instrument is used. The abundances, relative to solar values (Anders & Grevesse 1989), are representative for highly active young stars with inverse FIP effect (Telleschi et al. 2005, Scelsi et al. 2005, Garcia-Alvarez et al. 2005): He:1.0 – C: 0.45 – N: 0.788 – O: 0.426 – Ne: 0.832 – Mg: 0.263 – Al: 0.50 – Si: 0.309 – S: 0.417 – Ar: 0.55 – Ca: 0.195 – Fe: 0.195 – Ni: 0.195. The total exposure time  $T_{\text{exp}}$  used for the analysis is in the order of a few 10 kiloseconds, and is given in Table 3. The best-fit fluxes are converted into luminosities  $L_X$  using XSPEC and assuming a distance of 140 pc to the TMC.

### 6.2. Best-fit models

An example observation is shown in Fig. 11, using data of Haro 6–13. The source extraction region contains 128 counts between 0.3 and 7.5 keV after elimination of bad time intervals, whereas the corresponding (scaled) background contribution is  $N_{\text{bg}} = 28.8$  counts. The observed counts are marked by ticks in Fig. 11, while the background spectrum  $f_{\text{bkg}}(E)$  is shown by the dashed line. Energies below 0.3 keV are discarded. The maximum-likelihood estimator for the source normalization is  $N_{\text{src}} = 99.2$  counts. The unbinned log likelihood is computed inside the parameter range of Fig. 11 (inlet), and attains its maximum at  $kT = 3.03$  keV,  $N_{\text{H}} = 0.48 \times 10^{22} \text{ cm}^{-2}$ . These values are marked by crosses in the inlets of Fig. 11, and the corresponding best-fit model  $f_{\text{src}}(E)$  is shown by solid line (main panel). For comparison, the corresponding minimum- $\chi^2$  parameters, using 10 equal-content bins, are given by  $kT = 2.55$  keV,  $N_{\text{H}} = 0.48 \times 10^{22} \text{ cm}^{-2}$ . The minimum  $\chi^2$ - and maximum  $L$ -estimates thus agree within errors (see below), but are not

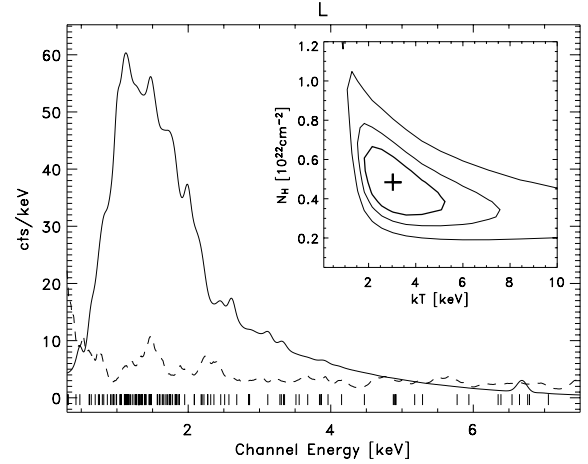


Fig. 11. *XMM-Newton*/EPIC observation of Haro 6–13, comprising 128 counts. Solid line: maximum- $L$  solution  $f_{\text{src}}(E)$ ; dashed: background spectrum  $f_{\text{bkg}}(E)$ ; ticks: observed counts. Inlets: best-fit parameters (crosses) and likelihood profiles (contours) at nominal confidence levels 68% (boldface), 90% and 99%. See Sect. 6.3.

equal. The parameters  $(kT, N_{\text{H}})$  do not follow an equally simple trend, and are more sensitive to the statistics used. The other best-fit parameters listed in Table 3 have been obtained in a similar way as for Haro 6–13.

### 6.3. Confidence regions

A rough estimate on the errors of the best-fit parameters can be obtained from the quantity  $2\Delta L \doteq 2(L_{\text{max}} - L)$ , which is asymptotically chi square distributed with the number of degrees of freedom equal to the number of model parameters. Here,  $L_{\text{max}} = \max_f L$  is the maximum achievable log likelihood, which is associated with the full parameter space of Wilks theorem (Wilk, 1938). By thresholding  $2\Delta L$  at the  $\alpha$ -quantiles of a chi square distribution with 2 degrees of freedom, accounting for the parameters  $(kT, N_{\text{H}})$ , and 1 degree of freedom, accounting for the parameter  $N_{\text{src}}$ , approximate confidence domains in parameter space are obtained. They are shown by solid contours in Fig. 11, representing cuts through the likelihood surfaces at confidence levels 68% (boldface), 90%, and 99%. The first errors given in Table 3 (outside brackets) represent projections of the 68% confidence surfaces obtained in this way. Note that the absolute credibility of the best-fit solution (i.e. the value of  $L_{\text{max}}$ ) is not addressed by the statistic  $2\Delta L$ , which decouples the goodness-of-fit problem from the confidence domain problem.

The parameter errors can also be predicted from Monte-Carlo simulations alone, similar as in an instrumental design study. To this end, the true and best-fit parameters (under  $L$  statistic) of more than  $10^6$  samples are classified according to the best-fit parameters. For each class (containing about 100 samples), the means ( $b$ ) and standard deviations ( $\sigma$ ) of the differences between true and best-fit parameters are evaluated, and taken as a proxy for the biases and errors of the best-fit parameters. The simulation is repeated for each *XMM-Newton* observation, so that the correct background and instrumental response are taken into account. The resulting error bounds  $b \pm \sigma$  are given in parentheses in Table 3, and are to be compared with the 68% confidence limits from the likelihood thresholding method. As can be seen, the Monte-Carlo errors are of the same order as those obtained from the likelihood thresholding, and the biases are generally in consistent direction.

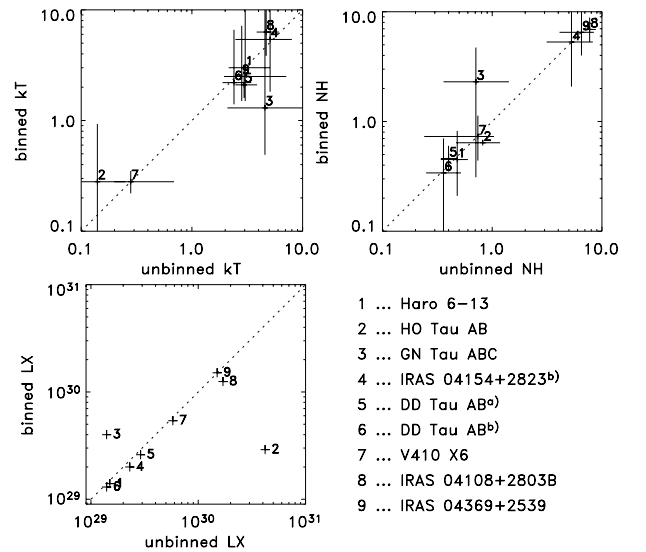
**Table 3.** Maximum- $L$  parameters deduced from *XMM-Newton* observations of the TMC. Superscripts a) and b) refer to different observations of the same object;  $T_{\text{exp}}$  is the exposure time, and  $N_{\text{cnt}}$  is the observed number of counts (corresponding to  $N_{\text{src}} + N_{\text{bg}}$ ). Errors outside brackets represent projections of the 68% likelihood surfaces (Fig. 11 inlets); errors in brackets are obtained from Monte-Carlo simulations (Sect. 6.3). Luminosities  $L_X$  refer to  $N_{\text{H}} = 0$  and the energy band from 0.3 to 10 keV.

Object	$T_{\text{exp}}$ [ks]	$N_{\text{cnt}}$	$kT$ [keV]	$N_{\text{H}}$ [ $10^{22}\text{cm}^{-3}$ ]	$N_{\text{src}}$ [counts]	$L_X$ [erg/s]
Haro 6-13	14.05	128	$3.03^{+2.02(+2.26)}_{-0.88(-1.01)}$	$0.48^{+0.12(+0.13)}_{-0.14(-0.15)}$	$99.2^{+11.5(+11.4)}_{-10.7(-11.4)}$	$1.5 \cdot 10^{29}$
GN Tau ABC	14.25	79	$4.57^{+5.43(+1.92)}_{-2.48(-1.91)}$	$0.71^{+0.71(+0.35)}_{-0.35(-0.24)}$	$58.6^{+9.1(+8.0)}_{-8.4(-9.6)}$	$1.4 \cdot 10^{29}$
TMC 1A	14.25	39	$1.59^{+5.82(+4.87)}_{-0.70(-1.49)}$	$44.41^{+44.65(+15.44)}_{-30.94(-24.52)}$	$24.0^{+6.5(+6.2)}_{-5.9(-6.2)}$	$4.7 \cdot 10^{30}$
IRAS 04369+2539	14.25	148	$2.82^{+4.30(+3.84)}_{-0.87(-1.54)}$	$6.51^{+1.95(+1.11)}_{-2.40(-1.75)}$	$120.4^{+12.0(+12.8)}_{-11.6(-11.3)}$	$1.5 \cdot 10^{30}$
HO Tau AB	17.28	55	$0.14^{+0.11(+0.32)}_{-0.09(-0.12)}$	$0.82^{+0.36(+0.20)}_{-0.35(-0.63)}$	$42.8^{+7.7(+7.7)}_{-7.0(-7.3)}$	$4.2 \cdot 10^{30}$
IRAS 04325+2402	25.72	36	$2.70^{+5.30(+2.41)}_{-1.70(-1.09)}$	$8.35^{+11.65(+3.54)}_{-5.27(-3.32)}$	$21.6^{+6.2(+4.7)}_{-5.6(-7.3)}$	$2.1 \cdot 10^{29}$
IRAS 04108+2803B	26.76	467	$4.69^{+1.41(+2.01)}_{-0.84(-1.10)}$	$7.73^{+1.09(+1.00)}_{-0.95(-1.03)}$	$380.8^{+21.3(+21.6)}_{-19.8(-22.4)}$	$1.7 \cdot 10^{30}$
CIDA 1	26.76	47	$0.78^{+0.23(+0.23)}_{-0.57(-0.23)}$	$0.19^{+0.68(+0.18)}_{-0.17(-0.27)}$	$34.2^{+7.1(+5.9)}_{-6.4(-7.9)}$	$1.5 \cdot 10^{28}$
V410 A13	37.27	26	$0.22^{+0.56(+0.44)}_{-0.17(-0.14)}$	$1.14^{+1.27(+0.37)}_{-0.89(-0.95)}$	$12.7^{+5.4(+3.5)}_{-4.7(-6.7)}$	$5.1 \cdot 10^{29}$
DD Tau AB <sup>a)</sup>	37.26	395	$2.96^{+0.92(+1.10)}_{-0.49(-0.56)}$	$0.40^{+0.08(+0.06)}_{-0.06(-0.08)}$	$325.8^{+19.6(+20.2)}_{-18.2(-19.5)}$	$2.9 \cdot 10^{29}$
DD Tau AB <sup>b)</sup>	21.90	141	$2.40^{+1.04(+1.81)}_{-0.51(-0.71)}$	$0.36^{+0.16(+0.09)}_{-0.11(-0.14)}$	$103.3^{+12.0(+11.9)}_{-11.2(-11.7)}$	$1.4 \cdot 10^{29}$
V410 X6	37.23	158	$0.28^{+0.41(+0.18)}_{-0.08(-0.09)}$	$0.74^{+0.30(+0.15)}_{-0.50(-0.27)}$	$122.2^{+12.7(+13.6)}_{-11.7(-11.6)}$	$5.8 \cdot 10^{29}$
IRAS 04154+2823 <sup>a)</sup>	21.89	24	$2.66^{+2.34(+0.99)}_{-1.66(-1.01)}$	$5.19^{+2.81(+1.00)}_{-4.19(-3.19)}$	$11.1^{+5.2(+4.5)}_{-4.5(-5.8)}$	$7.5 \cdot 10^{28}$
IRAS 04154+2823 <sup>b)</sup>	37.27	145	$5.09^{+2.91(+1.01)}_{-2.62(-2.32)}$	$5.28^{+2.98(+2.05)}_{-2.15(-1.06)}$	$89.5^{+12.2(+10.6)}_{-11.5(-13.6)}$	$2.3 \cdot 10^{29}$

In summary, we have used here two complementary characterizations of parameter errors, based on the log likelihood ratio  $2\Delta L$  and on Monte-Carlo simulations. The errors derived from the Monte-Carlo simulations do not invoke the asymptotic assumption underlying Wilks theorem, and are therefore better adapted to low count rates. But then, they do not involve the actual observation at all, and rely entirely on the assumption that the true spectrum is correctly modeled by the template spectra and the background model. Any systematic error contribution is thus neglected. Therefore, the errors from the Monte-Carlo simulation tend to under-estimate the true errors and should be considered as lower bounds. As a further test, we have checked whether the minimum- $\chi^2$  estimates are within the errors of the maximum- $L$  solution, and found that this is so for all cases when  $2\Delta L$  is used, and holds true in half of the cases when the Monte-Carlo error bounds are invoked (both methods referring to 68% confidence level). We shall not pursue here a deeper discussion of confidence regions, but terminate with a hint to the literature where more refined constructions may be found in Eadie et al. (1971), Wachter et al. (1979), Cousins (1995), Porter (1996), Feldman & Cousins (1998), and Giunti (1999).

#### 6.4. Comparison with binned estimates

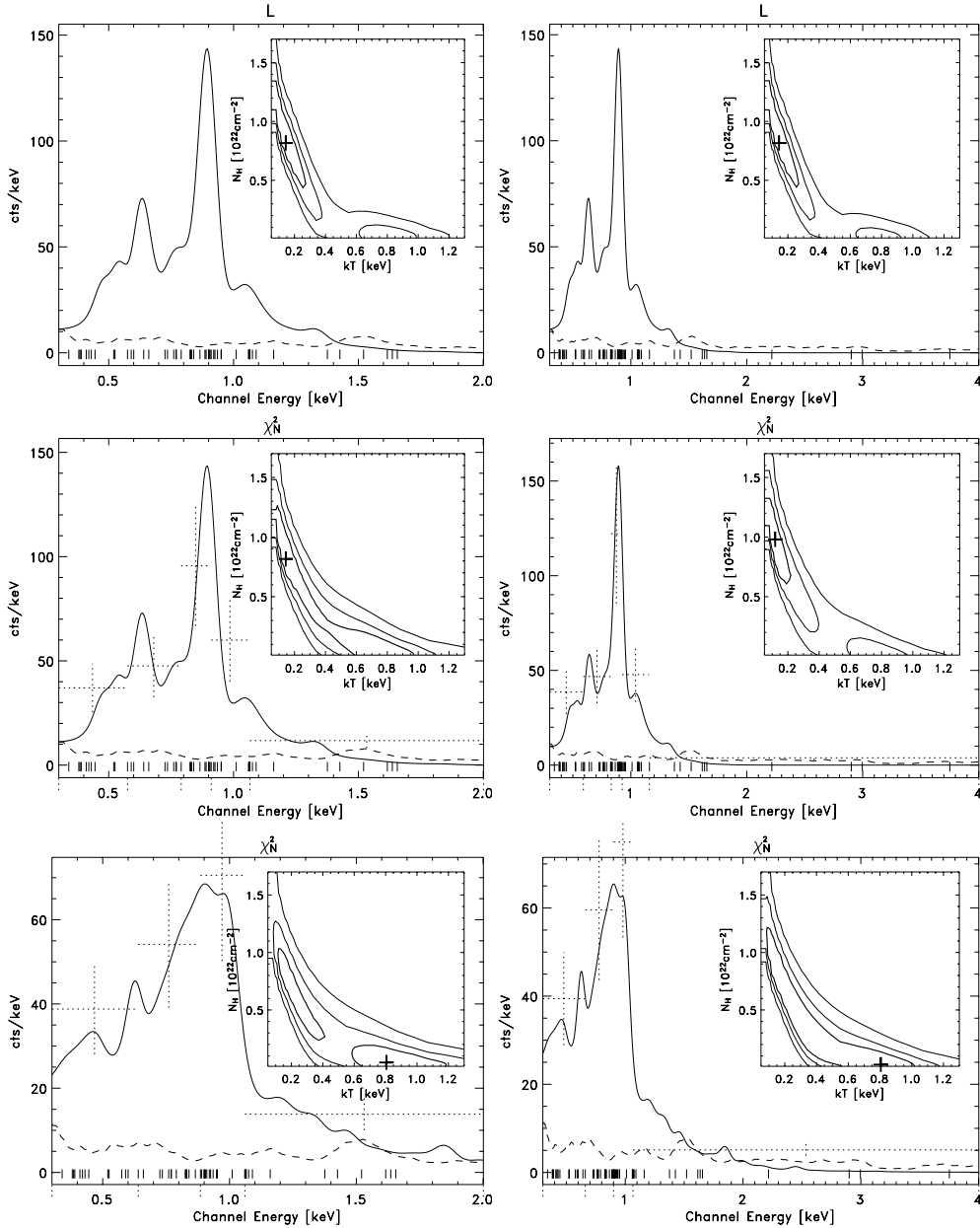
In order to compare the unbinned estimates with the standard XSPEC results, the XSPEC iterative fitting package was applied to background-subtracted observations with more than 50 counts, using bins of (at least) 10 counts and default ( $\chi^2_{\text{N}}$ ) statistics. The spectral model is identical to the one used for unbinned estimation (single temperature,  $N_{\text{H}}$ , abundances). The results are summarized in Fig. 12, displaying unbinned versus binned results. Numerical values for the binned results are tabulated Güdel et al. (2007a). The crosses are centered at the best-fit solutions and cover formal  $1\text{-}\sigma$  errors. For  $L_X$ , which is a



**Fig. 12.** Comparison of the unbinned estimates with results from binned one-temperature XSPEC fits. For  $kT$  and  $N_{\text{H}}$ , the error bars refer to unbinned 68% likelihood projections and XSPEC  $\Delta\chi^2_{\text{N}} = 1$  levels (3 DOF). For  $L_X$ , only the best-fit values are indicated.

derived quantity, only the best-fit parameters are indicated. As can be seen, the agreement between binned and unbinned estimators is generally good, except for HO Tau and GN Tau.

In order to clarify the situation we have created plots of HO Tau similar to Fig. 11 for both the  $L$  and  $\chi^2_{\text{N}}$  statistics, and have systematically varied the energy band and binning. The binning procedure provides (approximately) equal number of observed counts per bin, running from low to high energies. Some exemplary results are shown in Fig. 13. The



**Fig. 13.** Exploration of the effect of energy band and binning on the HO Tau parameter estimates. Left column: energies from 0.3 to 2 keV; right column: energies from 0.3 to 4 keV. Top row: unbinned likelihood. Middle row: Neyman’s chi square with 5 bins. Bottom row: Neyman’s chi square with 4 bins. The binned data are marked by dotted line.

left column includes energies from 0.3 to 2 keV, and the right column includes energies from 0.3 to 4 keV. The top row refers to the unbinned likelihood, as quoted in Table 3 and Fig. 12. The middle and bottom rows refer to Neyman’s chi square statistic with 5 and 4 bins, respectively; containing (10,10,11,9,11), (11,11,11,11,11), (13,13,12,13), and (14,14,13,14) counts (fixed- $n^*$  method). The binned observed spectra are indicated by the dotted crosses, with errors representing  $\pm \sqrt{n}$  counts. As can be seen, the unbinned likelihood has two shallow local maxima: a “cold” one at  $kT \lesssim 0.2$  keV and  $N_H \sim 8 \times 10^{21}$  cm $^{-2}$ , and a “warm” one at  $kT \lesssim 1$  keV and  $N_H = 0$ . While the (cold) maximum- $L$  solution and the  $L$ -profiles are stable under change of the energy range, the 68%  $\chi^2_N$ -profiles (middle) undergo a transition from a single to two regions, and the minimum- $\chi^2_N$  solution flips from cold to warm as the number of bins is decreased (middle to bottom). This instability is

mostly caused by the accumulation of counts around 0.8 keV, which either fall into a single (middle right) or two (bottom) bins. The 90%  $\chi^2_N$  domain is approximately stable under re-binning, and contains the XSPEC solution ( $kT = 0.38$  keV,  $N_H = 6.4 \times 10^{21}$  cm $^{22}$ ).

Based in Fig. 13 and the instability of “warm” solutions under re-binning, one may conclude that the HO Tau data favour “cold” solutions, all the more so as the “warm” solutions lie on the border of the parameter space (Protassov et al. 2002). However, from an astrophysical point of view it is not obvious how such a cold and absorbed spectrum would arise. A possible explanation is discussed in Sect. 7. We shall therefore argue that no definitive conclusion is possible for HO Tau yet, and that further observations are needed.

For the second discrepant case, GN Tau, the binned estimates yield smaller  $kT$  and larger  $N_H$  than the unbinned ones

(the binned estimates lie in the unbinned 90% domain but outside the unbinned 68% domain). This can be traced back to 12 observed counts below 1 keV; if these are excluded then the binned and unbinned best-fit parameters agree within 20%. We propose the following interpretation. The presence of counts below 1 keV implies low absorption ( $N_{\text{H}}$ ), provided that they cannot be explained by the background. The unbinned background model has, at  $E < 1$  keV, fine-structures which do not coincide with the observed counts; hence the latter are attributed by the unbinned method to the source, entailing low  $N_{\text{H}}$  and (Fig. 1 inlet) large  $kT$ . The binned method, in contrast, assigns more [namely,  $\int_{0.3}^1 f_{\text{bg}}(E) dE$ ] counts to the background; hence,  $N_{\text{H}}$  is larger and  $kT$  is smaller. The discrepancy of GN Tau stems thus from background fine structures at energies below 1 keV.

## 7. Summary and conclusions

We have used Monte-Carlo simulations to assess the performance of the unbinned (exact) Poisson likelihood ( $L$ ), binned Poisson likelihood ( $C$ ), Pearson's  $\chi^2$  and Neyman's  $\chi^2_{\text{N}}$ , Kolmogorov-Smirnov ( $D$ ), and Kuiper's ( $V$ ) statistics in the model classification and point estimation problems of low-count *XMM-Newton* and *Chandra/ACIS* spectra parameterized by temperature ( $kT$ ), hydrogen density ( $N_{\text{H}}$ ) and normalization ( $N$ ). By “unbinned” we mean that no grouping of instrumental energy channels is involved, so that the full readout resolution is used. The  $L$  statistic equals the  $C$  statistic on the finest possible (instrumental) binning. In all cases, the source normalization  $N_{\text{src}}$  was taken from its maximum-likelihood estimator, while the shape parameters ( $kT$ ,  $N_{\text{H}}$ ) were taken such as to optimize the different statistics.

The outcome of the simulations can be summarized as follows:

- The unbinned Poisson likelihood performs best with regard to the following measures of performance: (i) the probability of a successful identification in a search over discrete  $d_{\chi^2}$ -ordered candidate models, (ii) the expected chi-square distance between continuously indexed true and estimated models, (iii) the area under the operator-receiver curves of reduced binary (two-model) classifier problems, and (iv) and generally also with respect to the mean square errors of individual parameter errors.
- Under the  $L$  statistic, two models can on average be distinguished if their chi square distance exceeds  $\sim 3$ . Similar statements hold for the other statistics ( $C, \chi^2, \chi^2_{\text{N}}, D, V$ ) and correspondingly larger distances (up to  $\sim 5$ ). The chi square distance is thus empirically found to be a useful parameter-free measure of discrepancy between Poisson intensities.
- The  $\chi^2$  statistics should not be used unless more than 10 counts per bin are available, and  $\chi^2_{\text{N}}$  should not be used for less than 15 counts.

We therefore argue that the unbinned Poisson likelihood  $L$  is beneficial in cases with fewer counts than instrumental readout channels. Using  $L$ , the narrow spectral lines and steep gradients are taken into account at their exact location, unaffected by the bin size. Even more important, any instability of results under (arbitrary) choices of bin offsets and -sizes is avoided. In our view, this is a major benefit since it eliminates user-dependent sources of discrepant results.

The  $L$  statistics has been applied to XEST data, and demonstrated to operate under real observational conditions. The major selection criteria were spectral simplicity (so that the

3-parameter model applies), low count rate, and comparably well-known background. The parameter space was rigorously explored to avoid potential difficulties with iterative optimization. The maximum- $L$  results usually agree with the minimum- $\chi^2$  and minimum- $\chi^2_{\text{N}}$  values within 68% confidence, but are not equal. There are, though, two cases (out of 14) where the unbinned and binned estimates disagree: HO Tau and GN Tau. Such disagreement, by itself, is a helpful indicator for potential problems in the forward modeling.

In particular, HO Tau is found by the  $L$  method to be rather cool ( $kT \sim 0.2$  keV). Such a cool X-ray-emitting plasma would be unusual for a T Tauri star, the majority of which have average temperatures of  $>10$  MK. In the present survey dominant plasma with similarly low temperature is only found in the low temperature component of the spectrum of the jet-driving sources DG Tau A, GV Tau A, and DP Tau, which may be due to shocks in the jet (Güdel et al. 2005; Güdel et al. 2007b). However, DG Tau A also has significant plasma at very high temperatures ( $\sim 50$  MK) and HO Tau has no known jet. A dominant low-temperature corona has been found on the nearest classical T Tauri star, TW Hya, which is thus far unique in this respect. Apparent high densities in this cool plasma suggest that collisionally-ionized shocks in columns of material accreting onto the star may generate the X-ray emission (Kastner et al. 2002; Stelzer & Schmitt 2004). TW Hya is an unusually old classical T Tauri star, approximately 10 Myr old. Using the Siess et al. (2000) stellar evolution models, HO Tau appears also to be a relatively old classical T Tauri star, at 8–9 Myr. Its mass accretion rate of  $1.3 \times 10^{-9} M_{\odot} \text{ yr}^{-1}$  (White & Ghez 2001) is several times higher than that of TW Hya. Therefore it is plausible that HO Tau may be analogous to TW Hya, in which case the bulk of its X-ray emission may be generated by accretion shocks.

On the other hand, we can check whether the derived  $N_{\text{H}}$  agrees with expectations from measured optical and near-infrared extinctions  $A_V$  and  $A_J$ . For general interstellar matter, the conversions are  $N_{\text{H}} \approx 2 \times 10^{21} A_V \text{ cm}^{-2}$  and  $N_{\text{H}} \approx 5.6 \times 10^{21} A_J \text{ cm}^{-2}$  (see Vuong et al. 2002 and references therein). The optical extinction for HO Tau is  $A_V = 1.11\text{--}1.13$  mag (Kenyon & Hartmann 1995; White & Ghez 2001), and the near-IR extinction is  $A_J = 0.32\text{--}0.46$  mag (Kenyon & Hartmann 1995; Briceño et al. 2002), thus implying  $N_{\text{H}} = (1.8\text{--}2.6) \times 10^{21} \text{ cm}^{-2}$  under the assumption of standard gas-to-dust ratios. These estimates are in better agreement with the “hot” solution (Fig. 15) although the error ranges are large. We also note that significant deviations from the standard interstellar relation between extinction and X-ray absorption are well known (e.g., in the jet-driving stars mentioned above). Such deviations may point to an “anomalous” gas-to-dust ratio for example as a consequence of dust evaporation.

We also note that both the binned and unbinned methods provide low  $kT$  and high  $N_{\text{H}}$ . The discrepancy is in  $L_X$  (see Fig. 12) which is a consequence of large uncertainties in the spectral integration because most of the soft emission from the cool plasma is strongly absorbed. Indeed, if the temperature estimate is changed from 0.14 keV (unbinned) to 0.28 keV (binned) then  $L_X$  drops by about a factor ten. The  $L_X$  of the unbinned method is indeed rather high given a stellar bolometric luminosity of this system of  $L_* = 0.17 L_{\odot} = 6.5 \times 10^{32} \text{ erg s}^{-1}$ . Adopting the usual X-ray saturation level of  $L_X/L_* = 10^{-3}$  (e.g., Vilhu & Rucinski 1983), we would expect a maximum  $L_X = 6.5 \times 10^{29} \text{ erg s}^{-1}$ , more in line with the binned solution. The low count-rate of HO Tau unfortunately precludes a high-resolution X-ray spectroscopic study to study density-sensitive triplet emission-lines from He-like ions of O, Ne and Mg, which suggest high densities

in the TW Hya plasma. As for GN Tau,  $A_j = 1.17$  mag (Luhman 2004), hence we expect  $N_H = 6.6 \times 10^{21} \text{ cm}^{-2}$ , very close to the result from the unbinned method.

Although the Monte Carlo simulations suggest that the maximum- $L$  estimates are closer to the true values when averaged over a (hypothetical) ensemble of similar observations, a few words of caution are in order. The present simulations have assumed that the true spectrum is contained in the set of candidate spectra. If this is not the case, or if the Poisson process assumption is violated, then the use of the most binding  $L$  statistics also introduces the most severe misinterpretations. A major cause of flawed models stems from the background estimation. We have assumed here that this estimation is perfect, and have not considered background errors (e.g., Conrad et al. 2003), nor the more complicated problem of estimating the background spectrum jointly with the source spectrum. An imprecise background model may be the cause for the discrepancy between binned and unbinned estimates found in GN Tau, where the most relevant low-energy background (0.3 to 1 keV) contains 200 counts only, implying about 30% statistical background uncertainty at resolution  $\Delta E \sim 0.04$  keV. Furthermore, we have focused here on the choice of likelihood functions for the point estimation and model classification problems. Accordingly, our treatment of parameterization issues and confidence regions was rather crude. In particular, we did not attempt to introduce any (Bayesian) a priori information other than implicit in the choice of the forward models.

Finally, it should be pointed out that the unbinned approach is in principle not restricted to the one-dimensional spectrum problem considered in this article, and that the energy tags might be replaced by tuples of photon energy, arrival time, and position on the detector. However, the corresponding forward models would involve many more degrees of freedom, to the point where a simple maximum-likelihood principle might no longer be sufficient to uniquely determine the solution. Additional information (such as a Bayesian prior) would then be required in order to regularize the problem. Also, the more elaborated forward models would be more vulnerable to systematic errors (i.e., from CCD boundaries). In view of the (sparse) data we shall not pursue these issues.

*Acknowledgements.* We would like to thank the International Space Science Institut (ISSI) in Bern, Switzerland, for logistic and financial support during several workshops on the TMC campaign. This research is based on observations obtained with *XMM-Newton*, an ESA science mission with instruments and contributions directly funded by ESA member states and the USA (NASA). A.T. and M.G. acknowledge support from the Swiss National Science Foundation under grant Nr. 20-66875. L.S. and E.F. acknowledge financial contribution from contract ASI-INAF I/023/05/0.

## Appendix A: Generation of non-homogeneous Poisson variates by the inversion method

This Appendix describes the inversion method (Lewis & Shedler 1979; Devroye 1986), one of the two methods used here to simulate event lists. The inversion method, in the present implementation, relies on Theorem 1.4 (Chapter 6) of Devroye (1986) stating that if  $0 < X_1 < X_2 < \dots$  is a homogeneous Poisson process with unit rate function and  $\Lambda(x)$  is a non-decreasing function with  $\Lambda(0) = 0$  then  $0 < \Lambda^{-1}(X_1) < \Lambda^{-1}(X_2) < \dots$  is a non-homogeneous Poisson process with cumulative rate function  $\Lambda(x)$ . This follows from the fact that if  $F$  is a continuous (cumulative) distribution with inverse  $F^{-1}$  and  $U$  is uniformly distributed in the unit interval  $[0,1]$ , then  $F^{-1}(U)$  has (cumulative) distribution  $F$ . Our numerical implementation makes use

of the monotony of  $\Lambda(x)$  and the fact that  $X_{i+1} - X_i$  is exponentially distributed, which allows a successive computation of  $\Lambda^{-1}(X_i)$ . It proceeds as follows. Let a rate function  $\lambda(x)$  be specified by a sufficiently resolved discrete version  $\lambda_j$  with  $j$  ranging from 0 to  $N_c - 1$ , and define the discrete cumulative rate function by  $\Lambda_0 = \tau\lambda_0$ ,  $\Lambda_j = \Lambda_{j-1} + \tau\lambda_j$ , with  $\tau = \mathcal{N}^{-1}(N_c - 1)$  and  $\mathcal{N} = \sum_{j=0}^{N_c-1} \lambda_j$ . Then, a non-decreasing bin number sequence  $n_k \in [0, 1, \dots, N_c - 1]$  of a non-homogeneous Poisson process with intensity  $\lambda_j$  is obtained by the following pseudocode,

1. Initialisation:
  - $j = 0$
  - $k = 0$
  - $X = -\tau \ln U_0$
2. Iteration:
  - while  $X < (N_c - 1)$  do
    - while  $(\Lambda_j < X)$  and  $(j < N_c)$  do
      - $j = j + 1$
      - end do
      - $n_k = j$
      - $X = X - \tau \ln U_k$
      - $k = k + 1$
    - end

where  $U_k$  are uniform random numbers in  $[0, 1]$ , so that  $-\tau \ln U_k$  is exponentially distributed. It was numerically verified that the  $n_k$  obtained in the above way are Poisson distributed with intensity  $\lambda_{n_k}$ .

## References

- Anders, E., & Grevesse, N. 1989, *Geochim. Cosmochim. Acta*, 53, 197
- Arnaud, K. A. 1996, in *Astronomical Data Software and Systems V*, ed. G. Jacoby, & J. Barnes (San Francisco: ASP), ASP Conf. Ser., 101, 17
- Arzner, K. 2004, *Inverse Problems*, 20, 1728
- Babu, G. J., & Feigelson, E. D. 1996, *Astrostatistics* (London: Chapman & Hall)
- Baker, S., & Cousins, R. D. 1984, *Nucl. Instrum. Methods*, 221, 437
- Basseville, M. 1989, *Signal Proc.*, 18, 349
- Boese, F. G., & Doebereiner, S. 2001, *A&A*, 370, 649
- Briceño, C., Luhman, K. L., Hartmann, L., Stauffer, J. R., & Kirkpatrick, J. D. 2002, *ApJ*, 580, 317
- Cash, W. 1979, *ApJ*, 228, 939
- Conrad, J., Botner, O., Hallgren, A., & de los Heros, C. P. 2003, *Phys. Rev. D*, 67, 012002
- Cousins, R. D. 1995, *Am. J. Phys.*, 63, 399
- Devroye, L. 1986, *Non-uniform random variate generation* (Springer)
- Digel, S. W. 2000, in *Proc. 5th Compton Symposium*, ed. M. McConnel, & J. Ryan, AIP Conf. Proc., 510, 601
- Dolphin, A. E. 2002, *MNRAS*, 332, 91
- Eadie, W. T., Drijard, D., James, F. E., Roos, M., & Sadoulet, B. 1971, *Statistical methods in experimental physics* (Amsterdam: Elsevier)
- Feldman, G. F., & Cousins, R. D. 1998, *Phys. Rev. D*, 57, 3873
- Feller, W. 1968, *An introduction to probability theory and its applications* (New York: John Wiley & Sons)
- García-Alvarez, D., Drake, J. J., Lin, L., Kashyap, V., & Ball, B. 2005, *ApJ*, 621, 1009
- Gibbs, A. L., & Su, F. E. 2002, *Int. Stat. Rev.*, 70, 419
- Giunti, C. 1999, *Phys. Rev. D*, 59, 053001
- Gosset, E. 1987, *A&A*, 188, 258
- Güdel, M. 2004, *A&ARv*, 12, 71
- Güdel, M., Skinner, S. L., Briggs, K. R., et al. 2005, *ApJ*, 626, L53
- Güdel, M., Briggs, K., Arzner, K., et al. 2007a, *A&A*, 468, 353
- Güdel, M., Telleschi, A., Audard, M., et al. 2007b, *A&A*, 468, 515
- Hald, A. 1998, *A History of Mathematical Statistics* (John Wiley)
- Hanley, J. A., & McNeill, B. J. 1982, *Radiology*, 143, 29
- Johnson, D. H., & Sinanovic, S. 2001, *Symmetrizing the Kullback-Leibler Distance*, submitted to *IEEE Trans. Info. Theory*, <http://citeseer.ist.psu.edu/johnson01symmetrizing.html>

- Kastner, J. H., Huenemoerder, D. P., Schulz, N. S., Canizares, C. R., & Weintraub, D. A. 2002, *ApJ*, 567, 434
- Kendall, M., & Stuart, A. 1958, *The advanced theory of Statistics* (London: Charles Griffin & Co)
- Kenyon, S. J., & Hartmann, L. 1995, *ApJS*, 101, 117
- Kinoshita, K. 2002, *Proc. Conf. on Advanced Statistical Techniques in Particle Physics*, Univ. of Durham, March 18–22, ed. M. R. Whalley, & L. Lyons
- Kuiper, N. H. 1962, *Proc. Koninklijke Nederlandse Akademie van Wetenschappen*, 63, 38
- Lewis, P. A., & Shedler, G. S. 1979, *Naval Research Logistics Quarterly*, 26, 403
- Luhman, K. L. 2004, *ApJ*, 617, 1216
- Metchev, S. A., & Grindlay, J. E. 2002, *MNRAS*, 335, 73
- Michel, O., & Ferrari, A. 2003, in *Infrared and Submillimeter Space Astronomy*, ed. C. Aime, & R. Soummer, *ESA Publ. Ser.*, Vol. 8
- Mighell, K. J. 1999, *ApJ*, 518, 380
- Miller, G., Martz, H. F., Little, T. T., & Guilmette, R. 2002, *Health Phys.*, 83, 512
- Müller, A. 1995, *Optimal selection from distributions with unknown parameters: robustness of Bayesian methods*, Technical Report, Inst. für Wirtschaftstheorie und Operations Research, Univ. Karlsruhe, Germany
- Nousek, J. A., & Shue, D. R. 1989, *ApJ*, 342, 1207
- O'Sullivan, J. A., Blahut, R. E., & Snyder, L. 1998, *IEEE Trans. Information Theory*, 44, 2094
- Paltani, S. 2004, *A&A*, 420, 789
- Porter, F. 1996, *Nucl. Inst. Meth. A*, 368, 793
- Protassov, R., van Dyk, D. A., Kashyap, V., & Siernigikova, A. 2002, *ApJ*, 571, 545
- Peterson, W., Birdsall, T., & Fox, W. 1954, *IEEE Trans. Information Theory*, 4, 171
- Rachev, S. T. 1991, *Probability metrics and the stability of stochastic models* (New York: John Wiley & Sons)
- Rao, C. 1957, *Bull. Inter. Inst. Stat.*, 35, 25
- Reiss, R.-D. 1993, *A course on Point processes*, *Springer Series in Statistics* (Springer)
- Robinson, J. W., Rung, J., Bulsara, A. R., & Inchiosa, M. E. 2000, *Phys. Rev. E*, 63, 011107
- Santaló, L. A. 1976, *Integral Geometry and Geometric Probability* (London: Addison-Wesley)
- Scargle J. D. 1998, *ApJ*, 504, 405
- Scelsi, L., Maggio, A., Peres, G., & Pallavicini, R. 2005, *A&A*, 432, 671
- Scott, D. W. 1992, *Multivariate Density Estimation* (New York: John Wiley)
- Siess, L., Dufour, E., & Forestini, M. 2000, *A&A*, 358, 593
- Smith, R. K., Brickhouse, N. S., Liedahl, D. A., & Raymond, J. C. 2001, *ApJ*, 556, L91
- Stelzer, B., & Schmitt, J. H. M. M. 2004, *A&A*, 418, 687
- Stelzer, B., Flaccomio, E., Briggs, K., et al. 2007, *A&A*, 468, 463
- Sturges, H. A. 1926, *J. Am. Statist. Assoc.*, 21, 65
- Telleschi, A., Güdel, M., Briggs, K., et al. 2005, *ApJ*, 622, 653
- Vilhu, O., & Rucinski, S. M. 1983, *A&A*, 127, 5
- Vuong, M. H., Montmerle, T., Grosso, N., et al. 2003, *A&A*, 408, 581
- Wachter, K., Leach, R., & Kellog, E. 1979, *ApJ*, 230, 274
- White, R. J., & Ghez, A. M. 2001, *ApJ*, 556, 265
- Wise, M. E. 1963, *Biometrika*, 50, 145