

In search of an unbiased temperature estimator for statistically poor X-ray spectra

A. Leccardi^{1,2} and S. Molendi²

¹ Università degli Studi di Milano, Dip. di Fisica, via Celoria 16, 20133 Milano, Italy
e-mail: leccardi@iasf-milano.inaf.it

² INAF-IASF Milano, via Bassini 15, 20133 Milano, Italy

Received 13 February 2007 / Accepted 29 May 2007

ABSTRACT

Although commonly employed by X-ray astronomers, maximum likelihood estimators are known to be biased. In this paper we investigate the bias associated with the measure of the temperature from an X-ray thermal spectrum. We show that, in the case of low surface brightness regions, commonly adopted estimators, such as those based on χ^2 and Cash statistics, return strongly biased results. We stress that this can have strong implications when measuring the temperature of cluster outer regions with current experiments. We consider various approaches to overcoming this problem and the most effective is a technique that allows us to correct the bias a posteriori. Extensive Monte Carlo simulations show that our correction returns excellent results under different conditions.

Key words. methods: statistical – X-rays: galaxies: clusters

1. Introduction

With the advent of *XMM-Newton* and *Chandra* it has to a certain extent become possible to explore the physical properties of the intra-cluster medium (ICM) in the outer regions of galaxy clusters. In dealing with these regions, there are both statistical and systematic issues that need to be addressed; typically, the spectra have poor statistics (i.e. few counts/bin) and a high background, especially at high energies, where the instrumental background dominates other components, also due to the sharp decrease of the effective area of the experiments. In this paper we employ simulations to examine how best to analyze this kind of spectra. Here we focus on the treatment of statistical errors and neglect systematic ones, which will be discussed in a forthcoming paper (Leccardi & Molendi 2007). More specifically the question we wish to address is the following: “*What are the effects of pure statistical uncertainties in determining interesting parameters of highly non linear models (e.g. the temperature of the ICM), when we analyze spectra accumulated from low surface brightness regions using current X-ray experiments?*” To deal with this, we perform a set of simulations. First, we choose the input values for model parameters and produce the expected spectrum; then, we generate a large number of perturbed spectra representing a large set of measurements; finally, we analyze them with different techniques based on the method of maximum-likelihood (hereafter ML) and compare the results. Our choices of simulation parameters (e.g. spectral model, energy band, fixed parameters, etc.) are justified by our practical issue, i.e. determination of the temperature in the outer regions of massive galaxy clusters. Our analysis is mainly focused on *XMM-Newton*; however, most of our results are valid in all cases when analyzing low-count Poisson-distributed data.

From a more general perspective, ours may be viewed as an attempt to quantify the significance of the bias of ML estimators

commonly adopted by X-ray astronomers to determine spectral parameters. As we shall see, the most common ML estimators, indeed all those available within XSPEC¹, are characterized by a substantial bias when applied to our specific case. A long-term solution to the problem requires an unbiased, or perhaps a less biased, estimator to be found and implemented within standard fitting packages (i.e. XSPEC). Another, faster, solution involves correcting the bias a posteriori by making use of extensive Monte Carlo simulations.

The outline of the paper is the following. In Sect. 2 we consider the idealized source-only case. In Sect. 3 we include the background considering two cases: in the first the source contribution is much more important than the background one, in the second the opposite is true. In Sect. 4 we try to correct the bias: in Sect. 4.1 we search for a long-term solution of the problem using different estimators; in Sects. 4.2 and 4.3 we correct the bias a posteriori; in particular, in Sect. 4.3 we explain how to use our ad-hoc recipe (i.e. the so-called “triplet” method). In Sect. 5 we summarize our main results.

Errors are quoted at one sigma for one interesting parameter, unless otherwise stated.

2. The source-only case

In this section we deal with the idealized source-only case. We represent the source with an absorbed thermal model (WABS*MEKAL in XSPEC). The parameter values are the following: the equivalent hydrogen column density along the line of sight, N_{H} , is $2.5 \times 10^{20} \text{ cm}^{-2}$; the metallicity, Z , and the redshift, z , are respectively 0.25 solar and 0.2; the temperature, kT , is 7 keV and the normalization, N_{S} , is 3.5×10^{-3} in XSPEC units.

¹ We used XSPEC 11.3 (<http://heasarc.nasa.gov/docs/xanadu/xspec/xspec11/index.html>)

Our redistribution matrix (RMF) and effective area (ARF) have been produced from the observation number 0093030101 of the galaxy cluster Abell 1689 with the EPIC-MOS1 instrument; the angular size of the region for which we accumulate the simulated spectra is about 4 arcmin², which corresponds to the ring between 1.0' and 1.5' centered on the cluster emission peak. The exposure times considered are 5 ks, 10 ks, 100 ks, and 1000 ks and the total counts in the 2.0–10.0 keV band are respectively about 600, 1200, 12 000, and 120 000. For each channel we perturb the number of counts with a Poisson distribution centered on the expected value. We repeat this step N_{meas} times (with N_{meas} very large) to obtain N_{meas} spectra, which simulate N_{meas} independent measurements of the source.

We fit simulated spectra using the χ^2 and the Cash statistics, where the latter is more suitable for analyzing spectra with a few counts per channel (Cash 1979; Nousek & Shue 1989; Mighell 1999; Arzner et al. 2006). We recall that each measurement can be represented by the number of counts, O_i , observed in each channel i ($i = 1, \dots, N$ where N is the number of channels). The probability, Q , of obtaining this particular measurement (i.e. this particular spectrum) is the product of Poisson distributions and can be expressed as a function of the expected counts, E_i , which depend² on the particular set of model parameters, α (e.g. in this case $\alpha = (N_{\text{H}}, kT, Z, z, N_{\text{S}})$):

$$Q(\alpha) = \prod_{i=1}^N \frac{E_i^{O_i} e^{-E_i}}{O_i!}. \quad (1)$$

The associated log-ML function C (Cash 1979) is defined as

$$C(\alpha) = -2 \ln Q(\alpha) = -2 \sum_{i=1}^N (O_i \ln E_i - E_i - \ln O_i!). \quad (2)$$

The best set of parameters is determined by maximizing Q (i.e. minimizing C) with respect to α . Conversely, the χ^2 statistic is based on the hypothesis that each spectral bin contains a sufficient number of counts for the deviations of the O_i from the E_i to have a Gaussian distribution. This hypothesis is satisfied for large O_i , when Q can be approximated by a product of Gaussian distributions, and the associated log-ML function χ^2 is defined as

$$\chi^2(\alpha) = \sum_{i=1}^N \frac{(O_i - E_i)^2}{\sigma_i^2}, \quad (3)$$

where σ_i is usually the uncertainty in the i th bin ($\sigma_i = O_i^{1/2}$). The larger O_i , the better the approximation of the Gaussian regime. Channel grouping is a widely used strategy that allows the bias introduced by this approximation to be reduced. We group channels in order to have at least 25 counts per bin, which is a commonly adopted compromise. Conversely, when using the Cash statistic we perform a minimal grouping to avoid channels with no counts; i.e. the spectrum is substantially unbinned and no spectral information is lost. Each spectrum is fitted between 2.0 and 10.0 keV (the energy band we are interested in) with the absorbed thermal model mentioned above. The N_{H} is fixed to the input value (typical values of N_{H} for cluster observations have negligible effects above 2 keV), z is allowed to vary between 0.186 and 0.214 ($\pm 7\%$ of the input value), while kT , Z , and N_{S} are free. We determine best-fit values and one sigma uncertainties for all parameters.

² In the following equations the dependency of E_i from α is omitted for clarity.

Table 1. Weighted averages of temperature best fit values compared to the input value and relative differences $\Delta T/T_0$, using different exposure times and statistics.

Exp. ^a	kT_0 ^b	kT ^c	χ^2	Cash	
				kT ^c	$\Delta T/T_0$ ^d
1000	7.00	6.89 ± 0.01	-1.6%	7.00 ± 0.01	+0.0%
100	7.00	6.83 ± 0.01	-2.4%	7.03 ± 0.01	+0.4%
10	7.00	6.76 ± 0.03	-3.4%	6.91 ± 0.02	-1.3%
5	7.00	6.59 ± 0.04	-5.9%	6.81 ± 0.03	-2.7%

Notes: ^a exposure time in kiloseconds; ^b input temperature in keV; ^c measured temperature in keV; ^d relative difference.

Table 2. Weighted averages of temperature best-fit values compared to the input value and relative differences $\Delta T/T_0$, using different channel groupings.

N_{bin} ^a	kT_0 ^b	kT ^c	$\Delta T/T_0$ ^d
400	7.00	6.99 ± 0.01	-0.1%
100	7.00	6.95 ± 0.01	-0.7%
25	7.00	6.89 ± 0.01	-1.6%

Notes: ^a counts per bin; ^b input temperature in keV; ^c measured temperature in keV; ^d relative difference.

In Table 1 we compare the weighted average of the N_{meas} measured temperatures to the input value, $kT_0 = 7$ keV, for different exposure times and statistics. We choose N_{meas} in order to have similar uncertainties on the average ($N_{\text{meas}} = 1200$ for 5 and 10 ks, $N_{\text{meas}} = 300$ for 100 and 1000 ks). In almost all cases, the true temperature is underestimated by a few percent and the effect becomes more evident for shorter exposure times. We recall that both χ^2 and Cash statistics are based on the ML method. Although X-ray astronomers make extensive use of ML estimators, it is well known from the literature (e.g. Cowan 1998) that: 1) ML estimators may be biased, i.e. the expectation value may be different from the true value of the quantity to estimate; 2) ML estimators are usually Gaussian and unbiased only in the asymptotic limit. In the case at hand, the asymptotic limit is approached when the total number of counts becomes large. The results reported in Table 1 show that: 1) both ML estimators are biased; 2) both estimators are asymptotically unbiased; 3) the Cash estimator tends to the true value more quickly than the χ^2 one.

As we have just pointed out, the χ^2 is significantly more biased than the Cash estimator (i.e. the difference between the expected and the true value is greater). This is because the approximation of the Gaussian regime fails for few counts per bin. The obvious implication is that, to improve the precision of the χ^2 estimates, we need to increase the number of counts in each bin, N_{bin} . In Table 2 we compare the results obtained using the χ^2 with different channel groupings (note that this is not necessary when using the Cash statistic). The input temperature is 7 keV, the exposure time is 1000 ks, and the number of measurements is 300. As expected, we find that the greater the number of counts in each bin, N_{bin} , the smaller the bias. However, in practice, grouping of a large number of channels is not desirable, because it causes loss of spectral information; 25 counts per bin is a commonly adopted compromise. We have to mention the existence of an alternative way of reducing the bias that affects the χ^2 estimator for few counts per bin. Some authors (e.g. Churazov et al. 1996; Gehrels 1986; Kearns et al. 1995) choose different statistic weights (σ_i in Eq. (3)) instead of the standard $O_i^{1/2}$. We re-analyzed our spectra using all the alternative weights

implemented in XSPEC and we obtain results somewhat similar to those already discussed for the Cash statistic.

In this section we have analyzed the ideal case of a thermal source without a background. The results are summarized as follows. The standard χ^2 statistic works well only in the Gaussian regime, which is reached when performing a strong channel grouping (see Table 2). When using a realistic grouping (e.g. 25 counts per bin) the measured temperature, kT , is lower than the true temperature, kT_0 (see Table 1). The Cash statistic (Cash 1979) works better, because it is based on the ML function for Poisson processes; however, when the total number of spectral counts is small, kT is lower than kT_0 by a few percent (see Table 1). This means that the Cash ML estimator is only asymptotically unbiased (for a review about parameter estimation and ML concepts see Cowan 1998). Many efforts (e.g. Cash 1979; Wachter et al. 1979; Baker & Cousins 1984; Gehrels 1986; Nousek & Shue 1989; Kearns et al. 1995; Churazov et al. 1996; Jading & Riisager 1996; Mighell 1999; Hauschild & Jentschel 2001; Bergmann & Riisager 2002; Arzner et al. 2006) have been devoted to extending the standard theories about curve fitting to the case of low-count spectra (best fit parameters and confidence intervals estimation, goodness-of-fit test, etc.). However no definitive solution has been found.

3. The source-plus-background case

In this section we consider a more realistic situation by introducing a simplified instrumental background. We model it with a power law (PEGPWRLW/b within XSPEC), which is convolved with the RMF but not multiplied by the ARF. The power-law slope, Γ_B , is fixed to 0.25; the normalization, N_B , is calculated at the center of the energy band to minimize the correlation with Γ_B .

There are two ways of analyzing spectra with background: subtract it using a spectrum from blank field observations or model it. When modeling the background, one can use the whole energy band (2.0–11.3 keV rather than 2.0–10.0 keV as when using the subtraction) to increase the statistic. Indeed, due to the high-energy sharp decrease of the ARF of EPIC-MOS1, the thermal component becomes negligible beyond 10 keV. The background subtraction using the χ^2 statistic is a widely used technique; however, in the previous section (see Sect. 2) we showed that, for low-count spectra, the Cash statistic is more suitable than the χ^2 with reasonable grouping. Since the Cash statistic requires the number of counts in each channel to be greater than zero, the background has to be modeled (Cash 1979). We analyze simulated spectra in both ways and compare the results. Hereafter we call “sub- χ^2 ” the standard analysis technique and “mod-C” the analysis using the Cash statistic and the background modeling.

We proceed as for the source-only case, considering the Abell 1689 observation mentioned in Sect. 2 as a guideline. We produce a simulated spectrum choosing realistic input values for an absorbed thermal (see Sect. 2) plus background (see above) model (WABS*MEKAL+PEGPWRLW/b in XSPEC) and produce N_{meas} different measurements with a Poissonian perturbation of the number of counts in each channel. In the mod-C case, each spectrum is associated to the RMF and the ARF and is fitted between 2.0 and 11.3 keV with the WABS*MEKAL+PEGPWRLW/b model. The N_H and Γ_B are fixed to the input values, z is allowed to vary between 0.186 and 0.214 ($\pm 7\%$ of the input value), kT , Z , N_S , and N_B are free. We determine best-fit values and one sigma uncertainties for all parameters. Finally, we compute the weighted average of all N_{meas} values of each parameter using

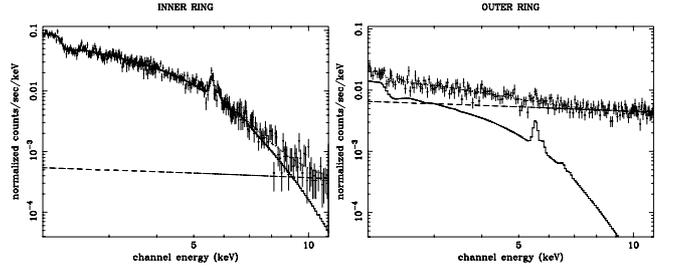


Fig. 1. Simulated spectra accumulated in an inner ring (left panel, Fig. 1a) and in an outer ring (right panel, Fig. 1b). The solid line is the source contribution, the dashed line the background, and the dotted the sum. In the outer ring, beyond 3 keV, background counts dominate source counts. See text for further details and for model parameters.

one sigma uncertainties, and we compare it with the input value. In the sub- χ^2 case, we simulate a background-only spectrum with a long exposure time. We consider a PEGPWRLW/b model (slope and normalization are equal to those of the power law in the source observation mentioned above) and we perturb the expected spectrum as explained above. The adopted background spectrum is the same for all N_{meas} measurements, and its exposure time is 1000 ks. We group the channels of each of the N_{meas} source spectra to have at least 25 counts per bin and associate the background spectrum, the RMF, and the ARF to the binned spectrum. We fit the net spectrum with a thermal model only (WABS*MEKAL in XSPEC) in the 2.0–10.0 keV band to determine the best-fit values, then compute the weighted averages and compare them with the input values.

We consider two spatial regions: the ring between 1.0' and 1.5' centered on the cluster emission peak, where the source dominates the background (see Fig. 1a) and the ring between 4.5' and 6.0', where the background dominates (see Fig. 1b). Input values for the normalizations of both components are the best-fit values measured in the two rings of the Abell 1689 observation mentioned in Sect. 2. More specifically, in the inner ring $N_S = 3.5 \times 10^{-3}$ and $N_B = 1.5$, and in the outer ring $N_S = 7.0 \times 10^{-4}$ and $N_B = 17.5$ (XSPEC units). For each ring we consider three input temperatures (namely 5, 7, and 9 keV) and two exposure times (10 and 100 ks).

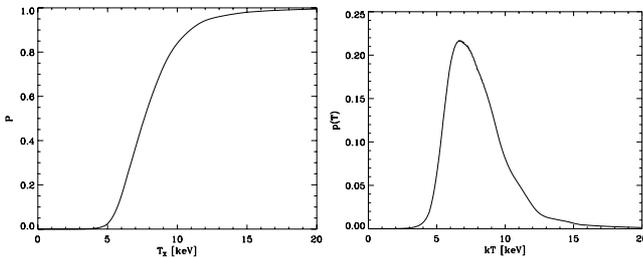
In Table 3 we show the comparison between the two different data analysis techniques described above (i.e. sub- χ^2 and mod-C). First we consider the inner ring, where the source dominates the background. The results are very similar to the case without background (see Sect. 2, Table 1). For the 100 ks case, mod-C returns the correct temperature and sub- χ^2 slightly underestimates it. For shorter exposure times, both techniques return a slightly biased value (bias $\approx 3\%$). No significant trend with the input temperature, kT_0 , is found. When considering the outer ring, where the background dominates, we find different results, because the true temperature is strongly underestimated in all cases. There is a clear trend with the input temperature: the higher kT_0 , the stronger the bias. For long exposure times, mod-C (bias $\approx 10\%$) works better than sub- χ^2 (bias $\approx 30\%$). Conversely, for short exposure times both techniques underestimate the true temperature by a factor of about 2. These results are qualitatively similar to those found for the source-only case (see Table 1), but the bias is much stronger.

We have repeated the same analysis described above for a particular set of simulated spectra (namely in the outer ring, with exposure time of 10 ks and $kT_0 = 7$ keV) modeling the source with a bremsstrahlung rather than a MEKAL. The bremsstrahlung model is simpler and can be expressed as an analytic function

Table 3. Comparison between the results obtained using the sub- χ^2 and the mod-C data analysis techniques.

Ring	Exp. ^a	kT_0 ^b	sub- χ^2		mod-C	
			kT ^c	$\Delta T/T_0$ ^d	kT ^c	$\Delta T/T_0$ ^d
1.0'–1.5'	100	5.00	4.84 ± 0.01	–3.2%	4.96 ± 0.01	–0.8%
1.0'–1.5'	100	7.00	6.78 ± 0.02	–3.1%	6.97 ± 0.02	–0.4%
1.0'–1.5'	100	9.00	8.69 ± 0.02	–3.4%	8.97 ± 0.03	–0.3%
1.0'–1.5'	10	5.00	4.81 ± 0.03	–3.8%	4.82 ± 0.03	–3.6%
1.0'–1.5'	10	7.00	6.78 ± 0.05	–3.1%	6.79 ± 0.05	–3.0%
1.0'–1.5'	10	9.00	8.68 ± 0.11	–3.6%	8.62 ± 0.08	–4.2%
4.5'–6.0'	100	5.00	3.95 ± 0.01	–21.0%	4.71 ± 0.02	–5.8%
4.5'–6.0'	100	7.00	5.24 ± 0.02	–25.1%	6.45 ± 0.03	–7.9%
4.5'–6.0'	100	9.00	6.43 ± 0.02	–28.6%	8.09 ± 0.04	–10.1%
4.5'–6.0'	10	5.00	3.02 ± 0.03	–39.6%	3.20 ± 0.03	–36.0%
4.5'–6.0'	10	7.00	3.68 ± 0.04	–47.4%	3.94 ± 0.04	–43.7%
4.5'–6.0'	10	9.00	4.11 ± 0.05	–54.3%	4.52 ± 0.06	–49.8%

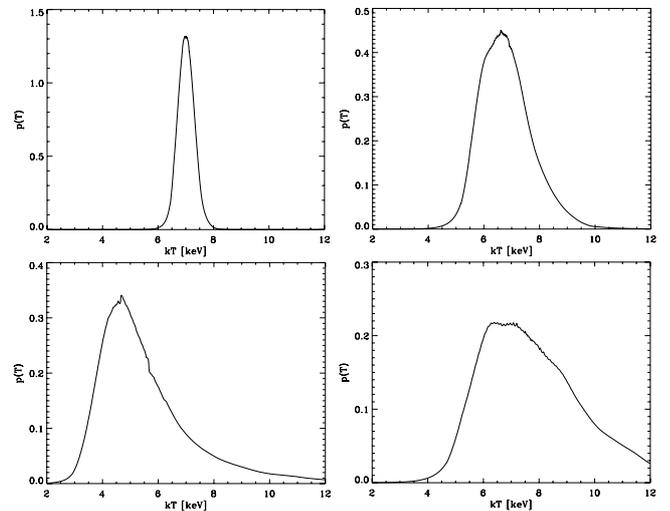
Notes: ^a exposure time in kiloseconds; ^b input temperature in keV; ^c measured temperature in keV; ^d relative difference.

**Fig. 2.** A cumulative distribution function (at left) and the associated probability density function (at right).

of its two free parameters (i.e. the temperature and the normalization). Conversely, the MEKAL model has two further parameters (metallicity and redshift), and its complicated dependency on the parameters is not expressed in an analytic form: the expected values are tabulated on a finite grid as a function of all parameters. For this particular set of spectra, the bias for MEKAL and bremsstrahlung models is the same. This suggests that the bias is not related to the approximation of a finite grid of values.

Some insight into the origin of the bias can be gained by inspecting the probability density function (p.d.f.) of the parameter of interest (in this case the temperature). Here we deal with the Cash statistic, and similar considerations apply to the χ^2 . For each measurement, we define as C_{\min} the absolute minimum value of the function $C(\alpha)$. As in the previous section (see Sect. 2), we minimize $C(\alpha)$ (see Eq. (2)) to determine our best estimate, α_{best} , of the parameter set ($C_{\min} \equiv C(\alpha_{\text{best}})$). Cash (1979) showed that the function ΔC (i.e. $C - C_{\min}$) follows a χ^2 distribution, therefore the confidence intervals can be generated in a standard way (e.g. using the XSPEC command ERROR). With the XSPEC command STEPPAR we produce the function C for each free parameter (here we consider only the temperature). We calculate $\Delta C(T)$ and, since it is χ^2 distributed, we can associate to each temperature, T_X , the probability that the true value is less or equal to T_X , i.e. the cumulative distribution function (c.d.f.) of the temperature, $P(T_X)$. Given that $P(T_X) = \int_0^{T_X} p(T) dT$, we can derive the p.d.f. of the temperature, $p(T)$, for each single measurement (see Fig. 2).

In Fig. 3 we compare representative $p(T)$ for single measurements in different conditions. For each case, the $p(T)$ is chosen randomly from the N_{meas} different measurements; therefore, the attention should be focused on the shapes of the p.d.f., rather

**Fig. 3.** A comparison between representative $p(T)$ for single measurements extracted randomly in different conditions. The attention should be focused on the shapes of the p.d.f., rather than on temperature values. *Top panels:* the inner ring, where the source dominates over the background. *Bottom panels:* the outer ring, where the background is dominant. For left panels the exposure time is 100 ks, for right panels 10 ks. The input temperature is always 7 keV. Note that the scales in ordinate are different. Clearly the curves become less symmetric and less similar to Gaussians, as the exposure time decreases and the background contribution increases.

than on temperature values. Clearly the curves become less symmetric and less similar to Gaussians, as the exposure time decreases and the background contribution increases. The input temperature also plays a role: the higher is the temperature, the less symmetric is the curve. Summarizing, the poorer the statistical quality of the data, the more asymmetric the $p(T)$ and the stronger the bias.

The way measurements are combined does not change the result. We have experimented with two different methods: the weighted average of individual measurements and the product of individual p.d.f. The weighted average roughly approximates the $p(T)$ to a Gaussian function and implies the neglect of the contribution of high temperature tails. A more appropriate way to join information from different and independent measurements is to multiply single p.d.f. The best value for the parameter corresponds to the maximum of the joined p.d.f. We

Table 4. Comparison between the results obtained using different estimators of the temperature. The input temperature is 7 keV.

Est. ^a	100 ks		20 ks		10 ks	
	kT ^b	$\Delta T/T_0$ ^c	kT ^b	$\Delta T/T_0$ ^c	kT ^b	$\Delta T/T_0$ ^c
T	6.44 ± 0.03	-8.0%	4.96 ± 0.05	-29.1%	4.04 ± 0.07	-42.3%
T^{-1}	6.96 ± 0.03	-0.6%	7.46 ± 0.08	+6.6%	8.83 ± 0.14	+26.1%
$T^{-1/2}$	6.88 ± 0.03	-1.7%	6.59 ± 0.06	-5.9%	6.36 ± 0.09	-9.1%
$T^{-1/4}$	6.80 ± 0.03	-2.9%	6.24 ± 0.06	-10.9%	5.72 ± 0.08	-18.3%

Notes: ^a temperature estimator; ^b measured temperature in keV; ^c relative difference.

multiply all N_{meas} p.d.f., computed as explained above and still find a discrepancy between the best fit and true values. The bias is only slightly weaker than when computing a weighted average. We have also tried computing the $p(T)$ in a different way, i.e. using the parametric bootstrap technique (Press et al. 1992), which consists in creating and analyzing a large number of fake datasets starting from model best-fit values. We essentially obtain the same results.

In Table 3 we have showed that the strength of the bias mainly depends on the total number of counts and on the background contribution. One possibility for increasing the total counts is to extend the band to lower energies. We have explored it by analyzing one of our set of simulated spectra (namely in the outer ring, with an exposure time of 10 ks and $kT_0 = 7$ keV) between 0.5 and 11.3 keV. In this energy band the correlation between kT and N_S is weaker and the uncertainty on both parameters for a single measurement is smaller. Using the mod-C technique, the bias in the broad band is smaller ($\approx 10\%$ vs. $\approx 40\%$) than in the narrow band; this suggests that also the parameter degeneracy could play an important role when fitting in the 2.0–11.3 keV band. However, in practice, it is not useful to enlarge the band to lower energies. The main reasons are the imperfect calibration of EPIC instruments and the presence of the galactic X-ray background (negligible beyond 2 keV), which introduce systematic effects that are hard to take into account. Moreover, broadband spectra are substantially more contaminated by emission from low temperature components located on the same line of sight as the dominant component (see Mazzotta et al. 2004; Vikhlinin et al. 2005).

In this section we have analyzed the realistic case of a thermal source with a background. In such conditions a stronger bias is expected (Eadie et al. 1971; Bergmann & Riisager 2002). As in the source-only case (see Sect. 2), we find that 1) the χ^2 and the Cash estimators are strongly biased, and that 2) the Cash estimator is less biased than the χ^2 one, especially for long exposure times. The strength of the bias depends mainly on two factors: the total number of counts and the background contribution.

4. Different attempts to correct the bias

Having established that neither sub- χ^2 nor mod-C return acceptable results, we are faced with two alternative ways to proceed. A long-term solution to the problem requires that an unbiased, or perhaps a less biased, ML estimator be found and implemented within standard fitting packages (i.e. XSPEC). Another, faster, solution involves correcting the bias a posteriori, making use of extensive Monte Carlo simulations. In the following sections we show our main results obtained by exploring both approaches.

4.1. Using different estimators

From the literature (e.g. Cowan 1998), we know that even if \hat{X} is an unbiased estimator of X , $f(\hat{X})$ is not necessarily an unbiased

estimator of $f(X)$. Reversing the argument one can argue that if \hat{T} is a biased estimator of T , a transformation, f , may exist such that $f(\hat{T})$ is an unbiased (or at least less biased) estimator of $f(T)$. To test this idea we define in XSPEC (using the MDEFINE command) an analytic model similar to bremsstrahlung, which we dub BREM2:

$$S_T(E) = N_S E^{-4/3} T^{-1/2} \exp\left(-\frac{E}{T}\right), \quad (4)$$

where the energy, E , is expressed in keV; the normalization, N_S , is chosen to reproduce the same flux as a MEKAL with no metals.

We simulate 3 sets of 3000 thermal plus background (BREM2+PEGPWRLW/b) spectra with the following input parameters: $N_S = 7.2 \times 10^{-4}$, $N_B = 17.5$, and $T = 7$ keV. These parameters correspond to those adopted in the case of the outer region (see Sect. 3). Each set has a different exposure time: 10, 20, and 100 ks. We define 3 different estimators of the temperature,

$$\begin{aligned} A &= T^{-1}, \\ B &= T^{-1/2}, \\ C &= T^{-1/4}, \end{aligned}$$

and their respective models:

$$\begin{aligned} S_A(E) &= N_S E^{-4/3} A^{1/2} \exp(-A E), \\ S_B(E) &= N_S E^{-4/3} B \exp(-B^2 E), \\ S_C(E) &= N_S E^{-4/3} C^2 \exp(-C^4 E). \end{aligned}$$

For simplicity we have considered only power laws as different $f(T)$. We fit each set of spectra with these models and measure the best estimate of $f(T)$. We compute the weighted average of the 3000 $f(T)$ and calculate T using the inverse function, f^{-1} . In Table 4 we report the results of this analysis for different exposure times and estimators. The choice of the estimator strongly affects the bias. When using T as estimator, we obtain very similar results to those obtained with a MEKAL. This is expected, because the model BREM2 is very similar to a bremsstrahlung (see Eq. (4)); note also that in Sect. 3 we showed that the bias is roughly the same when considering a MEKAL or a bremsstrahlung. When considering the bias as a function of the power-law index, we find a minimum corresponding to $T^{-1/2}$, which is the best estimator among those considered. For short exposure times (i.e. 10 ks) the use of $T^{-1/2}$ instead of T reduces the bias by a factor of 4. We suggest that this could be related to the degree of complexity of the derivative of $S(E)$ with respect to the estimator. Note also that, when slightly increasing the statistic (e.g. when considering 20 ks of exposure time), the bias associated to the $T^{-1/2}$ estimator is almost negligible if compared to typical statistic uncertainties. These results encourage exploration of this approach (i.e. to consider different estimators) in order to find a rigorously derived unbiased estimator of the temperature.

Table 5. Results obtained fitting $p(T)$ with a log-normal distribution.

Ring	Exp. ^a	kT_0 ^b	kT ^c	$\Delta T/T_0$ ^d
1.0'-1.5'	100	5.00	4.96 ± 0.01	-0.8%
1.0'-1.5'	100	7.00	6.97 ± 0.01	-0.4%
1.0'-1.5'	100	9.00	8.97 ± 0.01	-0.3%
1.0'-1.5'	10	5.00	4.88 ± 0.02	-2.4%
1.0'-1.5'	10	7.00	6.90 ± 0.05	-1.4%
1.0'-1.5'	10	9.00	8.81 ± 0.13	-2.1%
4.5'-6.0'	100	5.00	4.90 ± 0.02	-2.0%
4.5'-6.0'	100	7.00	6.77 ± 0.04	-3.3%
4.5'-6.0'	100	9.00	8.51 ± 0.09	-5.4%
4.5'-6.0'	10	5.00	4.68 ± 0.13	-6.4%
4.5'-6.0'	10	7.00	5.90 ± 0.24	-15.7%
4.5'-6.0'	10	9.00	7.67 ± 0.51	-14.8%

Notes: ^a exposure time in kiloseconds; ^b input temperature in keV; ^c measured temperature in keV; ^d relative difference.

4.2. Fitting with a log-normal function

The shape of the $p(T)$ resembles the log-normal function, which is the p.d.f. of any random variable whose logarithm is normally distributed. If X is a random variable with a normal distribution, then $x \equiv \exp(X)$ has a log-normal distribution. The log-normal distribution has p.d.f.

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x - \mu)^2/2\sigma^2} \quad (5)$$

for $x > 0$, where μ and σ are respectively the mean and the standard deviation of the variable's logarithm. The expected value is

$$E(X) = e^{\mu + \sigma^2/2}, \quad (6)$$

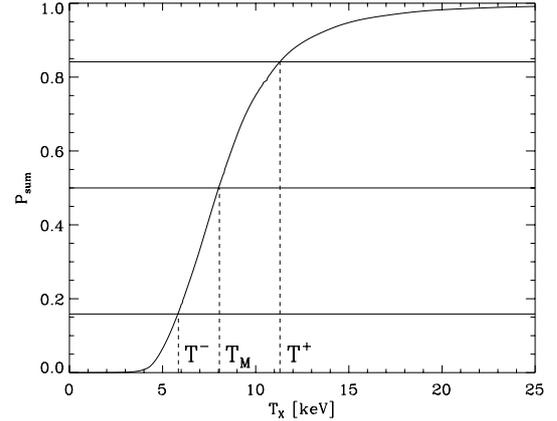
and the variance is

$$\text{var}(X) = (e^{\sigma^2} - 1) e^{2\mu + \sigma^2}. \quad (7)$$

We fit each $p(T)$ with a log-normal function, $f(x; \mu, \sigma)$ (see Eq. (5)) and calculated the best values of μ_i and σ_i . We computed a weighted average of μ_i using σ_i^{-2} as weights and calculated the expected value (see Eq. (6)) and the uncertainty, i.e. the variance (see Eq. (7)) divided by the square root of the number of measurements. The results of our test calculations are reported in Table 5. In all cases, this method provides better results than a simple weighted average (see Table 3 for a comparison). There is still a bias of a few percent, except for the case of the outer ring with 10 ks: in this case the bias is greater than 10%. Thus, when the background contribution is small, the log-normal distribution provides a good estimate, while the result is still biased when the background is dominant, especially for few total counts, but much less than when using the standard techniques.

4.3. A semi-empirical method: summing three distributions

The three EPIC instruments (MOS1, MOS2, and pn) on board *XMM-Newton* provide three simultaneous and independent measurements of the same target; therefore, when dealing with EPIC data, one needs to correctly combine these three measurements. A weighted average is the simplest procedure, however in Sect. 3 we have shown that it leads to biased results. In Sect. 3 we also showed that the strength of the bias is related to the shape of the p.d.f. and in Sect. 4.2 we showed that a fit with a log-normal function does not return sufficiently accurate results; in this section we try to proceed in a different way, emphasizing the contribution of p.d.f. tails at high temperatures. We derive N_{meas} measurements of the temperature with their corresponding p.d.f., as

**Fig. 4.** Visual representation of the definition of T^- , T_M , and T^+ from a joined cumulative distribution function.**Table 6.** Results obtained with the semi-empirical “triplet” method.

Ring	Exp. ^a	kT_0 ^b	kT ^c	$\Delta T/T_0$ ^d
1.0'-1.5'	100	5.00	4.97 ± 0.01	-0.6%
1.0'-1.5'	100	7.00	6.97 ± 0.02	-0.4%
1.0'-1.5'	100	9.00	8.98 ± 0.03	-0.2%
1.0'-1.5'	10	5.00	4.90 ± 0.04	-2.0%
1.0'-1.5'	10	7.00	6.94 ± 0.07	-0.9%
1.0'-1.5'	10	9.00	8.98 ± 0.11	-0.2%
4.5'-6.0'	100	5.00	5.00 ± 0.02	-0.0%
4.5'-6.0'	100	7.00	6.90 ± 0.04	-1.4%
4.5'-6.0'	100	9.00	8.91 ± 0.05	-1.0%
4.5'-6.0'	10	5.00	5.04 ± 0.06	+0.8%
4.5'-6.0'	10	7.00	6.97 ± 0.09	-0.4%
4.5'-6.0'	10	9.00	8.96 ± 0.13	-0.4%

Notes: ^a exposure time in kiloseconds; ^b input temperature in keV; ^c measured temperature in keV; ^d relative difference.

in the mod-C case described in Sect. 3. We divide the N_{meas} measurements into groups of three and, for each group, we consider the three p.d.f., $p_i(T)$ ($i = 1, 2, 3$), and combine them in a non-standard way by calculating the sum, rather than the product, of the single p.d.f. In practice, it is equivalent, but more useful, to sum the c.d.f., $P_i(T_X)$ directly. The sum is renormalized by dividing it by 3. We define $P_{\text{sum}}(T_X)$ as

$$P_{\text{sum}}(T_X) = \frac{1}{3} \sum_{i=1}^3 P_i(T_X). \quad (8)$$

This is a sort of joined c.d.f. of three measurements, and the associated p.d.f. is usually more symmetric than the single $p_i(T)$. We define as T^- , T_M , and T^+ the temperatures that correspond to a probability, $P_{\text{sum}}(T_X)$, of 0.1587, 0.5000, and 0.8413, respectively (see Fig. 4). We consider T_M as the best estimate for the three joined measurements, $dT^- \equiv (T_M - T^-)/\sqrt{3}$ as the lower uncertainty and $dT^+ \equiv (T^+ - T_M)/\sqrt{3}$ as the upper uncertainty. Thus we have $N_{\text{meas}}/3$ “triplet” measurements: $T_{M-dT^-}^{+dT^+}$. We compute the weighted average of the $N_{\text{meas}}/3$ “triplets”, and in Table 6 we compare the results with the input values. In almost all cases this semi-empirical method (hereafter “triplet” method) provides excellent results: the discrepancy is lower than 2% and often comparable with the statistical uncertainty.

We have tried joining different numbers of measurements together. Simulations show that, when considering two measurements at a time, we find the temperature to be slightly underestimated. When considering five measurements, we obtain

substantially correct results, as when using the “triplets”. This suggests that the effectiveness of our a posteriori correction depends on the number of measurements we combine. We propose using the “triplets” because three is the minimum number of measurements for which we obtain unbiased temperature estimates and because this is a natural choice when analyzing EPIC data.

We want to stress that this technique is not rigorously derived from principles of statistics, but we have shown that it is the only method that returns the expected temperature under very different situations (e.g. different background contributions and exposure times). This could be related to the fact that joined p.d.f. are usually more symmetric than the single ones.

5. Summary

It is well-known from the literature (e.g. Eadie et al. 1971; or Cowan 1998) that ML estimators are generally biased and that they are Gaussian and unbiased only in the asymptotic limit. In this paper we test the effects of statistical fluctuations in determining the temperature from a thermal spectrum. In particular, we explore a range of conditions for which ML estimators reveal their intrinsic bias.

In the source-only case we show that:

1. the estimators of the temperature based on the Cash and the χ^2 statistics are biased for short exposure times, i.e. for few counts;
2. the Cash statistic works better than the χ^2 , as pointed out by Cash (1979);
3. the χ^2 statistic works as well as the Cash, when strongly increasing channel grouping.

In the source-plus-background case we show that:

1. the standard analysis techniques (sub- χ^2 and mod-C) return heavily biased results;
2. the strength of the bias depends mainly on the total number of counts and on the background contribution;
3. the use of different estimators of the temperature, in particular $T^{-1/2}$, strongly reduces the bias.

The last point is encouraging for finding a long-term solution of the problem (i.e. a rigorously derived unbiased estimator). As an alternative and immediate solution we propose the so-called

“triplet” method, which makes use of a correction of the bias a posteriori, working on the probability distribution functions. This semi-empirical recipe returns the correct result under very different situations, even though it is not rigorously derived from statistical principles.

We point out that our results can have strong implications for the measurement of the temperature from spectra accumulated from low surface brightness regions (e.g. the outer regions of galaxy clusters) with current experiments, i.e. *XMM-Newton* and *Chandra* (see for example Piffaretti et al. 2005; Vikhlinin et al. 2005; Pratt et al. 2007). Indeed the bias seems to be related to the statistical quality of the data, which typically depends on the distance of a given region from the core. For this reason we might expect a net effect on the radial temperature profiles of galaxy clusters. In a forthcoming paper (Leccardi & Molendi 2007), we will analyze a sample of clusters to determine the mean temperature profile using our data analysis technique.

Acknowledgements. We would like to thank an anonymous referee for helpful comments. We thank A. Finoguenov for useful discussions.

References

- Arzner, K., Güdel, M., Briggs, K., et al. 2006 [arXiv:astro-ph/0609193]
 Baker, S., & Cousins, R. D. 1984, Nucl. Instr. Meth. A, 221, 437
 Bergmann, U. C., & Riisager, K. 2002, Nucl. Instr. Meth. A, 489, 444
 Cash, W. 1979, ApJ, 228, 939
 Churazov, E., Gilfanov, M., Forman, W., & Jones, C. 1996, ApJ, 471, 673
 Cowan, G. 1998, Statistical data analysis (Oxford Science Publications)
 Eadie, W. T., Drijard, D., James, F. E., Roos, M., & Sadoulet, B. 1971, Statistical Methods in Experimental Physics (Amsterdam: North-Holland Publishers)
 Gehrels, N. 1986, ApJ, 303, 336
 Hauschild, T., & Jentschel, M. 2001, Nucl. Instr. Meth. A, 457, 384
 Jading, Y., & Riisager, K. 1996, Nucl. Instr. Meth. A, 372, 289
 Kearns, K., Primini, F., & Alexander, D. 1995, in Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes, ASP Conf. Ser., 77, 331
 Leccardi, A., & Molendi, S. 2007, in preparation
 Mazzotta, P., Rasia, E., Moscardini, L., & Tormen, G. 2004, MNRAS, 354, 10
 Mighell, K. J. 1999, ApJ, 518, 380
 Nousek, J. A., & Shue, D. R. 1989, ApJ, 342, 1207
 Piffaretti, R., Jetzer, P., Kaastra, J. S., & Tamura, T. 2005, A&A, 433, 101
 Pratt, G. W., Böhringer, H., Croston, J. H., et al. 2007, A&A, 461, 71
 Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. 1992, Numerical recipes (Cambridge University Press)
 Vikhlinin, A., Markevitch, M., Murray, S. S., et al. 2005, ApJ, 628, 655
 Wachter, K., Leach, R., & Kellogg, E. 1979, ApJ, 230, 274