submitted to the Astrophysical Journal

Bayesian Estimation of Hardness Ratios: Modeling and Computations

Taeyoung Park[1], Vinay L. Kashyap[2], Aneta Siemiginowska[2], David A. van Dyk[3], Andreas Zezas[2], Craig Heinke[4], & Bradford J. Wargelin[2]

[1]*Department of Statistics*
Harvard University
One Oxford Street, Cambridge, MA 02138

tpark@stat.harvard.edu

[2]*Smithsonian Astrophysical Observatory,*
60 Garden Street, Cambridge, MA 02138

kashyap@head.cfa.harvard.edu
aneta@head.cfa.harvard.edu
azezas@head.cfa.harvard.edu
bwargelin@head.cfa.harvard.edu

[3]*Department of Statistics*
University of California,
364 ICS Bldg One, Irvine, CA 92697-1250

dvd@ics.uci.edu

[4]*Northwestern University,*
2131 Tech Drive, Evanston, IL 60208

cheinke@northwestern.edu

ABSTRACT

A commonly used measure to summarize the nature of a photon spectrum is the so-called Hardness Ratio, which compares the number of counts observed in different passbands. The hardness ratio is especially useful to distinguish between and categorize weak sources as a proxy for detailed spectral fitting. However, in this regime classical methods of error propagation fail, and the estimates of spectral hardness become unreliable. Here we develop a rigorous statistical treatment of hardness ratios that properly deals with detected photons as independent Poisson random variables and correctly deals with the non-Gaussian nature of the error propagation. The method is Bayesian in nature, and thus can be generalized to carry out a multitude of source-population–based analyses. We verify our method with simulation studies, and compare it with the classical method. We apply this method to real world examples, such as the identification of candidate quiescent Low-mass X-ray binaries in globular clusters, and tracking the time evolution of a flare on a low-mass star.

Subject headings: methods: statistical – stars: flare – X-rays: binaries

1. Introduction

The shape of the spectrum of an astronomical source is highly informative as to the physical processes at the source. But often detailed spectral fitting is not feasible due to various constraints, such as the need to analyze a large and homogeneous sample for population studies, or there being insufficient counts to carry out a useful fit, etc. In such cases, a hardness ratio, which requires the measurement of accumulated counts in two or more broad passbands, becomes a useful measure to quantify and characterize the source spectrum.[a] A hardness ratio is defined as either the ratio of the counts in two bands called the *soft* and *hard* bands, or a monotonic function of this ratio. We consider three types of hardness ratios,

$$\begin{aligned} \text{Simple Ratio,} \qquad & \mathcal{R} \equiv \frac{S}{H} \\ \text{Color,} \qquad & \mathrm{C} \equiv \log_{10}\left(\frac{S}{H}\right) \\ \text{Fractional Difference,} \qquad & \mathcal{HR} \equiv \frac{H-S}{H+S} \end{aligned} \tag{1}$$

where S and H are the source counts in the two bands, called the *soft* and *hard* passbands. The simple formulae above are modified for the existence of background counts and instrumental effective areas.

---

[a]In the context of *Chandra* data, the passbands often used are called *soft* (0.3-0.9 keV), *medium* (0.9-2.5 keV), and *hard* (2.5-8.0 keV) bands. Sometimes the *soft* and *medium* bands are merged into a single band, also called the *soft* band (e.g., 0.5-2 keV). In all cases, note that the energies refer to the nominal detector PHA or PI channel boundaries and not to the intrinsic photon energies. The choice of energy or PI channel range is flexible and situational.

Spectral colors in optical astronomy, defined by the standard optical filters (e.g., UBVRIJK, U-B, R-I, etc), are well known and constitute the main observables in astronomy. They have been widely applied to characterize populations of sources and their evolution. The hardness ratio concept was adopted in X-ray astronomy for early X-ray detectors (mainly proportional counter detectors) which had only a limited energy resolution. The first application of the X-ray hardness ratio was presented in the X-ray variability studies with SAS-3 observations by Bradt et al. (1976). Zamorani et al. (1981) investigated the X-ray hardness ratio for a sample of quasars observed with the *Einstein* X-ray Observatory. They calculated each source intensity in two energy bands (1.2-3 keV and 0.5-1.2 keV) and discussed a possible evolution of the hardness ratio (the ratio of both intensities) with redshift for their X-ray sample of 27 quasars. Similar ratios have been used in surveys of stars (Vaiana et al. 1981), galaxies (e.g., Kim, Fabbiano, & Trinchieri 1992) and X-ray binaries (Tuohy et al. 1978) studied with the *Einstein* and earlier observatories. In the case of X-ray binaries in particular, they were used to define two different classes of sources (Z-track and atoll sources; Hasinger & van der Klis, 1989) depending on their time evolution on $\mathcal{HR}$ v/s $\mathcal{HR}$ diagrams. Since then the concept of X-ray hardness ratio has been developed and broadly applied in a number of cases, most recently in the analysis of large data samples from the *Chandra* X-ray Observatory (Weisskopf et al. 2000) and XMM-*Newton* (Jansen et al. 2001).

Advanced X-ray missions such as *Chandra* and XMM-*Newton* allow for the detection of many very faint sources in deep X-ray surveys (see review by Brandt & Hasinger 2005). For example, in a typical observation of a galaxy, several tens and in some cases even hundreds of sources are detected, most of which have fewer than 50 counts, and many have only a few counts. Similar types of sources are detected in the ChaMP serendipitous survey (Kim et al. 2005). They provide the best quality data to date for studying source populations and their evolution. The most interesting sources are the sources with the smallest number of detected counts, because they have never been observed before. Further, the number of faint sources increases with improved sensitivity limits, i.e., there are more faint sources in deeper observations. Because these faint sources have only a few counts, hardness ratios are commonly used to study properties of their X-ray emission and absorption (e.g., Alexander et al. 2001, Brandt et al. 2001a, Brandt et al. 2001b, Giacconi et al. 2001, Silverman et al. 2005). With long observations the background counts increase, so the background contribution becomes significant and background subtraction fails to work. Still the background contribution must be taken into account to evaluate the source counts.

The sensitivity of X-ray telescopes is usually energy dependent, and thus the classical hardness ratio method can only be applied to observations made with the same instrument. Usually the measurements carry both detection errors and background contamination. In a typical hardness ratio calculation the background is subtracted from the data and only net counts are used. This background subtraction is not a good solution, especially in the low counts regime (see van Dyk et al. 2001). In general the Gaussian assumption present in the classical method (see §§ 1.1 and 3) is not appropriate for faint sources in the presence of a significant background. Therefore, the classical approach to calculating the hardness ratio and the errors is inappropriate for low counts. Instead, adopting a Poisson distribution as we do here (see below), hardness ratios can be reliably computed for both low and high counts cases. The Bayesian approach allows us to include the information about the background, difference in collecting area, effective areas and exposure times between the source and background regions.

Our Bayesian model-based approach follows a pattern of ever more sophisticated statistical methods that are being developed and applied to solve outstanding quantitative challenges in empirical Astronomy and Astrophysics. Bayesian model-based methods for high-energy high-resolution spectral analysis can be found for example in Kashyap & Drake (1998), van Dyk et al. (2001), van Dyk and Hans (2002), Protassov et al. (2002), van Dyk and Kang (2004), Gillessen & Harney (2005), and Park et al. (2006, in preparation). More generally, the monographs on Statistical Challenges in Modern Astronomy (Feigelson and Babu, 1992, 2003, Babu and Feigelson, 1997) and the special issue of *Statistical Science* devoted to Astronomy (May 2004) illustrate the wide breadth of statistical methods applied to an equal diversity of problems in astronomy and astrophysics.

Here, we discuss the classical method and present the fully Bayesian method for calculating the hardness ratio for counts data.[b]

1.1. The Classical Method

A conventional hardness ratio is calculated as set out in Equations 1, where S and H are the "soft" and "hard" counts accumulated in two non-overlapping passbands. In general, an observation will be contaminated by background counts, and this must be accounted for in the estimate of the hardness ratios. The background is usually estimated from an annular region surrounding the source of interest, or from a suitable representative region on the detector that is reliably devoid of sources. The difference in the exposure time and aperture area of source and background observations are summarized by a known constant r for which the expected background counts in the source exposure area are adjusted. With the background counts in the soft band (B_S) and the hard band (B_H) collected in an area

[b] A Fortran and C-based program which calculates the hardness ratios following the methods described in this paper is available for download from `http://hea-www.harvard.edu/AstroStat/BEHR/`

of r times the source region, the hardness ratio is generalized to

$$
\begin{aligned}
\mathcal{R} &= \frac{S - B_S/r}{H - B_H/r} \\
\mathrm{C} &= \log_{10}\left(\frac{S - B_S/r}{H - B_H/r}\right) \\
\mathcal{HR} &= \frac{(H - B_H/r) - (S - B_S/r)}{(H - B_H/r) + (S - B_S/r)}
\end{aligned} \tag{2}
$$

The adjusted counts in the background exposure area are directly subtracted from those in the source exposure area. The above equations can be further modified to account for variations of the detector effective areas by including them in the constant r, in which case the constants for the two bands will be different.[c] Errors are propagated assuming a Gaussian regime, i.e.,

$$
\begin{aligned}
\sigma_{\mathcal{R}} &= \frac{S - B_S/r}{H - B_H/r}\sqrt{\frac{\sigma_S^2 + \sigma_{B_S}^2/r^2}{(S - B_S/r)^2} + \frac{\sigma_H^2 + \sigma_{B_H}^2/r^2}{(H - B_H/r)^2}} \\
\sigma_{\mathrm{C}} &= \frac{1}{\ln(10)}\sqrt{\frac{\sigma_S^2 + \sigma_{B_S}^2/r^2}{(S - B_S/r)^2} + \frac{\sigma_H^2 + \sigma_{B_H}^2/r^2}{(H - B_H/r)^2}} \\
\sigma_{\mathcal{HR}} &= \frac{2\sqrt{(H - B_H/r)^2(\sigma_S^2 + \sigma_{B_S}^2/r^2) + (S - B_S/r)^2(\sigma_H^2 + \sigma_{B_H}^2/r^2)}}{\left[(H - B_H/r) + (S - B_S/r)\right]^2}
\end{aligned} \tag{3}
$$

where σ_{X_S}, σ_{X_H}, σ_{B_S}, and σ_{B_H} are typically approximated with the Gehrels prescription (Gehrels 1986)

$$
\sigma_X \approx \sqrt{X + 0.75} + 1\,, \tag{4}
$$

where X are the observed counts, and the deviation from $\sqrt{X}$ is due to identifying the 68% Gaussian (1σ) deviation with the $16^{\mathrm{th}} - 84^{\mathrm{th}}$ percentile range of the Poisson distribution. In addition to the approximation of a Poisson with a faux Gaussian distribution, classical error-propagation also fails on a fundamental level to properly describe the variance in the hardness ratios. Because $\mathcal{R}$ is strictly positive, its probability distribution is skewed and its width cannot be described simply by $\sigma_{\mathcal{R}}$. The fractional difference $\mathcal{HR}$ is better behaved, but the range limits of $[-1, +1]$ cause the coverage rates (see §3) to be incorrect. The color C is the best behaved, but the error range will be asymmetrical when the errors are large.

[c]Gross differences in detector sensitivity, e.g., between different observations, or for sources at different off-axis angles, can be accounted for with additional multiplicative factors modifying the background-subtracted counts. This is however not statistically meaningful and can lead to unreliable estimates and error bars in the small counts regime. Instead, we apply such scaling factors directly to the source intensities (see Equation 5).

Moreover, it has been demonstrated (van Dyk et al. 2001) that using background subtracted source counts (which is the case in the method outlined above) gives a biased estimate of the source intensity, especially for very faint sources. Furthermore, although we can account for gross differences in the detector sensitivity between observations, this is not done in a statistically correct manner, but instead simply by rescaling the background-subtracted counts in each band using additional pre-computed correction factors. Finally, the classical hardness ratio method cannot be applied when a source is not detected in one of the two bands (which is quite common since CCD detectors are more sensitive in the soft band).

An alternative method based on the median (or some quantile) of the distribution of the energies of the detected counts in a source was suggested by Hong et al. (2004). This method can be very powerful for classifying faint sources, but is not suited for comparisons of spectral properties between observations with different instruments. In this method, because the spectral parameter grids are highly non-linear and multi-valued in quantile-quantile space, it is also necessary that the spectral shape of the source being analyzed be known in order to interpret the numerical values.

We have therefore developed fully Bayesian approaches to appropriately compute hardness ratios and their errors, by properly accounting for the Poissonian nature of the observations. The value of our method is heightened in several astrophysically meaningful cases. In §2, we describe the model structure we adopt. In §3, we carry out simulations to compare our method with the classical method. In §4, we outline various applications of our method; we demonstrate its use in identifying candidate quiescent Low-mass X-ray binary sources in a globular cluster and in studying the evolution of a stellar X-ray flare. In §5, we discuss some nuances of our method such as the adopted prior distributions, as well as the advantages and limitations. A summary of the method and its advantages is presented in §6. A detailed mathematical derivation of the posterior probability distributions of the hardness ratios and the computational methods we use to calculate them are described in Appendix A, and a detailed comparison of the effects of priors is described in Appendix B.

2. Modeling the Hardness Ratios

2.1. The Poisson Case

Hardness ratios are often used to describe faint sources with very few counts; it is not uncommon for either or both of the soft and hard counts to be small enough that the source is undetected in that band. The classical method (§1.1) generally fails to produce reliable estimates (or upper and lower limits) in such cases, and indeed, it is sometimes not even possible to apply it, such as when the background is large and simple subtraction results in a negative estimate for the source counts. Thus, instead of the Gaussian assumptions on the detected counts, we directly model them as an inhomogeneous Poisson process.

Since the observed counts are the sum of counts due to the source and the background, and because the sum of two independent Poisson random variables is a Poisson random variable with the sum of two Poisson intensities, we may write[d]

$$S \sim \text{Poisson}(e_S \cdot (\lambda_S + \xi_S)) \quad \text{and} \quad H \sim \text{Poisson}(e_H \cdot (\lambda_H + \xi_H)), \tag{5}$$

where S and H are independent data points, λ_S and λ_H are the expected soft and hard source counts intensities, ξ_S and ξ_H are the expected soft and hard background counts intensities, and e_S and e_H are correction factors that take into account variations in effective areas, exposure times, and other instrumental effects.[e] The observed background counts are also modeled as independent Poisson random variables,

$$B_S \sim \text{Poisson}(r \cdot e_S \cdot \xi_S) \quad \text{and} \quad B_H \sim \text{Poisson}(r \cdot e_H \cdot \xi_H) \tag{6}$$

where ξ is scaled by a known correction factor r, to account for differences in source and background areas and sensitivities. We implicitly assume throughout that the observed data are discrete photon events, or counts. The intensities λ and ξ may have other units, determined by how the instrument efficiency is supplied. Thus, if e_S is given in [ct s cm^2 ph^{-1}], then λ_S will have units [ph s^{-1} cm^{-2}]. Alternately, e_S could describe the exposure time in [s], in which case λ_S will have units of [ct s^{-1}]. In the case where λ_S and λ_H themselves have the units [ct], then e_S and e_H are dimensionless, and usually taken to be unity. This allows us to compute hardness ratios of fluxes and count rates as well as counts. We place no restriction on the units of the model intensities (other than that λ and ξ have the same units); they are entirely determined by the problem at hand. Thus, our analysis is valid for any instrumental configuration or combination, and the only requirement is that the observed data be described by a Poisson distribution.

Given the expected source counts intensities (λ_S and λ_H), it is legitimate for the hardness ratio to be rewritten as

$$\text{Simple Ratio,} \quad \mathcal{R} = \frac{\lambda_S}{\lambda_H},$$

[d]Note that throughout this paper, Greek letters indicate unobserved quantities and Roman letters denote observed data values (except for $\mathcal{R}$, C, and $\mathcal{HR}$, which are used to represent the hardness ratio model parameters in the Poisson case).

[e]Note that in general the effective area changes within an adopted passband, so the chosen values of the correction factor are averages that are calculated after making some prior assumptions about the underlying source spectrum. Their values may change depending on the analysis results.

$$\text{Color,}\quad \mathrm{C} = \log_{10}\left(\frac{\lambda_S}{\lambda_H}\right),$$
$$\text{Fractional Difference,}\quad \mathcal{HR} = \frac{\lambda_H - \lambda_S}{\lambda_H + \lambda_S}. \tag{7}$$

These quantities are characteristics of the spectrum in question, rather than of the observed data, like the quantities in Equation 2. As such, the rewritten hardness ratio is of more direct scientific interest. Note also that while these quantities are defined entirely in terms of expected source intensities in the different passbands, the background is implicitly taken into account via direct modeling, as in Equations 5 and 6.

2.2. Bayesian Approach

As a new method of evaluating hardness ratios, we adopt a Bayesian approach. Bayesian analysis is explained in detail in numerous articles in the astronomical literature (e.g., Gregory & Loredo 1992, Kashyap & Drake 1998, van Dyk et al. 2001, Kashyap et al. 2002, Protassov et al. 2002), and here we describe it only in rudimentary terms. In this approach, the prior knowledge for a parameter (the *prior probability distribution*, e.g., $p(\lambda_S, \xi_S)$) is combined with the information in the observed data (the *likelihood*, e.g., $p(S, B_S|\lambda_S, \xi_S)$) to derive the probability distribution of the model parameters (the *posterior distribution*, e.g., $p(\lambda_S, \xi_S|S, B_S)$) via Bayes' theorem,[f] i.e.,

$$p(\lambda_S, \xi_S|S, B_S) = \frac{p(\lambda_S, \xi_S)p(S, B_S|\lambda_S, \xi_S)}{\iint p(\lambda_S, \xi_S)p(S, B_S|\lambda_S, \xi_S)d\xi_S d\lambda_S}, \tag{8}$$

and similarly for $p(\lambda_H, \xi_H|H, B_H)$. Because the posterior distribution in Equation 8 needs only be computed up to a normalizing constant, we often deal with an unnormalized posterior density, i.e., $p(\lambda_S, \xi_S|S, B_S) \propto p(\lambda_S, \xi_S)p(S, B_S|\lambda_S, \xi_S)$. Bayesian statistical inferences for a parameter are based on the posterior distribution, so that we consider its various summary statistics; see §2.2.1 for details.

Because the underlying likelihood functions are Poisson distributions, we assign so-called conjugate γ-prior distributions for both source and background intensities (van Dyk et al. 2001); a conjugate prior distribution ensures that a posterior distribution follows the same parametric form as the prior distribution, e.g., a γ-distribution is a conjugate prior for a

---

[f]The notation $p(A|B)$ is to be read as "the probability that A is true *given* that B is true." Thus, except for prior distributions, all the probability distributions that we use are conditional distributions.

Poisson likelihood.[g] That is, we assign independent γ-prior distributions for the source and background counts λ and ξ,

$$\lambda_S \sim \gamma(\psi_{S_1}, \psi_{S_2}) \quad \text{and} \quad \lambda_H \sim \gamma(\psi_{H_1}, \psi_{H_2})\,, \tag{9}$$

$$\xi_S \sim \gamma(\psi_{S_3}, \psi_{S_4}) \quad \text{and} \quad \xi_H \sim \gamma(\psi_{H_3}, \psi_{H_4})\,, \tag{10}$$

where the values of ψ are calibrated according to our prior knowledge, or lack thereof, about the parameters; see §5.1 for our discussion on choosing a prior distribution.

We describe the application of Bayes' Theorem and the derivation of the posterior probability distribution functions for the interesting parameters in Appendix A. Here, we note that since λ_S and λ_H are independent of each other, their joint posterior probability distribution is the product of the individual marginal posterior distributions,[h] written as (see Equation A8):

$$p(\lambda_S, \lambda_H | S, H, B_S, B_H) = p(\lambda_S|S, B_S) p(\lambda_H|H, B_H)\,. \tag{11}$$

Then the posterior distribution of each type of hardness ratio can be computed by transforming λ_S and λ_H to the appropriate variables and marginalizing over the resulting nuisance variable, as follows:

$$p(\mathcal{R}|S, H, B_S, B_H)\, d\mathcal{R} = d\mathcal{R} \int_{\lambda_H} d\lambda_H\ \lambda_H\ p(\mathcal{R}\lambda_H, \lambda_H|S, H, B_S, B_H)\,, \tag{12}$$

$$p(\mathrm{C}|S, H, B_S, B_H)\, d\mathrm{C} = d\mathrm{C} \int_{\lambda_H} d\lambda_H\ \lambda_H\ 10^{\mathrm{C}} \ln(10)\ p(10^{\mathrm{C}}\lambda_H, \lambda_H|S, H, B_S, B_H)\,, \tag{13}$$

$$p(\mathcal{HR}|S, H, B_S, B_H)\, d\mathcal{HR} = d\mathcal{HR} \int_{\omega} d\omega\ \frac{\omega}{2}\ p\Big(\frac{(1-\mathcal{HR})\omega}{2}, \frac{(1+\mathcal{HR})\omega}{2}\Big| S, H, B_S, B_H\Big)\,, \tag{14}$$

with $\omega = \lambda_S + \lambda_H$. The integrals can be computed using either Monte Carlo integration or Gaussian quadrature (throughout this paper, we use the term "quadrature" to refer

---

[g]Formally, these are *semi-conjugate* prior distributions in that they are conjugate to a particular conditional distribution. For example, the γ-prior distribution on λ_S is conjugate to the Poisson distribution of the source counts among the observed soft counts, denoted η_S in Appendix A1. The $\gamma(\alpha, \beta)$ distribution is a continuous distribution on the positive real line with probability density function

$$p(X) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} X^{\alpha-1} e^{-\beta X}\,,$$

and has a mean of α/β, and a variance of α/β^2 for $\alpha, \beta > 0$.

[h]The marginal posterior distribution of interest is computed by integrating out the so-called nuisance parameters out of their joint posterior distributions. For example, $p(\lambda_S|S, B_S) = \int p(\lambda_S, \xi_S|S, B_S)\, d\xi_S$.

exclusively to the latter). We use both methods of integration because neither has the clear advantage over the other in our case; see §5.4 and Appendix A.

2.2.1. Finding Point Estimates and Error Bars

Bayesian statistical inferences are made in terms of probability statements, and thus we consider various summary statistics for a posterior distribution. Here we briefly describe the standard summaries of location and methods of computing error bars; a fuller description along with detailed examples can be found in Park et al. (2006, in preparation). Commonly used summaries of location are the mean, median, and mode(s) of the distribution: The posterior mean is the posterior expectation of the parameter, the posterior median is the point that divides the posterior probability evenly such that 50% of the probability resides below and above its value, and the posterior mode is the most likely value of the parameter given the data.

When a posterior simulation of size N is obtained with a Monte Carlo sampler (see §A.1), the posterior mean is the simple average of the sample draws and the posterior median is the $[N/2]^{\text{th}}$ draw after sorting the draws in increasing order. To compute the posterior mode, we repeatedly bisect the Monte Carlo draws and choose the half with more draws until the range of the chosen half is sufficiently narrow. The midpoint of the resulting narrow bin approximates the posterior mode.

With quadrature (see §A.2), we can obtain equally spaced abscissas and the corresponding posterior probabilities. Thus, the posterior mean is computed as the sum of the product of an abscissa with their probabilities (i.e., the dot product of the vector of abscissa with the corresponding probability vector). The posterior median is computed by summing the probabilities corresponding to the ordered abscissa one-by-one until a cumulative probability of 50% is reached. The posterior mode is simply the point among the abscissa with the largest probability.

Unlike with point estimates above, there is no unique or preferred way to summarize the variation of a parameter. Any interval that encompasses a suitable fraction α of the area under the probability distribution qualifies as an estimate of the variation. Of these, two provide useful measures of the uncertainty: the equal-tail posterior interval, which is the central interval that corresponds to the range of values above and below which lies a fraction of exactly $(\alpha/2)$ of the posterior probability, is a good measure of the width of the posterior distribution; and the highest posterior density (HPD) interval, which is the range of values that contain a fraction $(1-\alpha)$ of the posterior probability, and within which the probability density is never lower than that outside the interval. The HPD-interval always contains the mode, and thus serves as an error bar on it.

For a symmetric, unimodal posterior distribution, these two posterior intervals are identical. The equal-tail interval is invariant to one-to-one transformations and is usually easier to compute. However, the HPD-interval always guarantees the interval with the smallest length among the intervals with the same posterior probability. Neither of these intervals is necessarily symmetric around the point estimate, i.e., the upper and lower bounds may be of different lengths. For consistency, we refer to such intervals as *posterior intervals*; others also refer to them as confidence intervals[i] or credible intervals.

For a Monte Carlo simulation of size N, we compute either the equal-tail interval or an interval that approximates the HPD interval. (Unless otherwise stated, here we always quote the equal-tail interval for the Monte Carlo method and the HPD-interval for the quadrature.) The equal-tail posterior interval in this case is computed by choosing the $[(\alpha/2)N]^{\rm th}$ and the $[(1-\alpha/2)N]^{\rm th}$ draws as the boundaries. An approximate HPD-interval is derived by comparing all intervals that consist of the $[X]^{\rm th}$ and the $[X+(1-\alpha)N]^{\rm th}$ draws and choosing that X which gives the shortest length among them.

When the posterior density is computed by the quadrature, we split parameter space into a number of bins and evaluate the posterior probability at the midpoint of each bin. In this case, a $100(1-\alpha)\%$ HPD-interval can be computed by beginning with the bin with the largest posterior probability and adding additional bins down to a given value of the probability density until the resulting region contains at least a fraction $(1-\alpha)$ of the posterior probability.

3. Verification: Simulation Study

In order to compare the classical method with our Bayesian method, we carried out a simulation study to calculate coverage rates of the classical and posterior intervals. Given pre-specified values of the parameters, source and background counts were generated and then used to construct 95% classical and posterior intervals of each hardness ratio using the methods discussed in this article. From the simulated data we calculated the proportion of the computed intervals that contain the true value of the corresponding hardness ratio. (This is the coverage rate of the classical and probability intervals.) In the ideal case, 95% of the

---

[i]Technically, confidence intervals are defined in terms of how frequently they contain the "true" value of the estimand, rather than in terms of a posterior probability distribution.

simulated intervals would contain the true value of each hardness ratio. Besides the coverage rate, the average length of the intervals and the mean square error of point estimates were also computed and compared. The mean square error of the point estimate $\hat{\theta}$ of θ is defined as the sum of the variance and squared bias for an estimator, i.e., $\mathrm{MSE}(\hat{\theta}) = \mathrm{E}[(\hat{\theta} - \theta)^2] = \mathrm{Var}(\hat{\theta}) + [\mathrm{E}(\hat{\theta}) - \theta]^2$ A method that constructs shorter intervals with the same coverage rate and produces a point estimate with a lower mean square error is generally preferred. The entire simulation was repeated with different magnitudes of the source intensities, λ_S and λ_H. Intrinsically, we are interested in the following two prototypical cases:

Case I : hardness ratios for high counts, i.e.,

$$\lambda_S = \lambda_H = 30, \ \xi_S = \xi_H = 0.1; \tag{15a}$$

Case II : hardness ratios for low counts, i.e.,

$$\lambda_S = \lambda_H = 3, \ \xi_S = \xi_H = 0.1. \tag{15b}$$

In both cases we adopt a background-area–to–source-area ratio of $r = 100$ (see Equation 6), i.e., we take the observed counts in the background region to be 10. Note that these are written with reference to the counts observed in the "source region", i.e., the units of $\lambda_S, \lambda_H, \xi_S, \xi_H$ are all [ct (source area)$^{-1}$], and that we have set $e_S = e_H = 1$ here. The actual extent of the source area is irrelevant to this calculation.

This simulation study illustrates two typical cases, i.e., high counts and low counts sources: Case I represents high counts sources for which Poisson assumptions tend to agree with Gaussian assumptions; Case II represents low counts sources where the Gaussian assumptions are inappropriate. In Case I, the posterior distributions of the hardness ratios agree with the corresponding Gaussian approximation of the classical method, but in Case II, the Gaussian assumptions made in the classical method fail. This is illustrated in Figure 1, where we compare the two methods for a specific and particular simulation: we assume $S = H = 30$ in Case I and $S = H = 3$ in Case II, compute the resulting posterior distributions of hardness ratios using the Bayesian method, and compare it to a Gaussian distribution with mean and standard deviation equal to the classical estimates. The right panels in Figure 1 show that there is a clear discrepancy between the two methods.

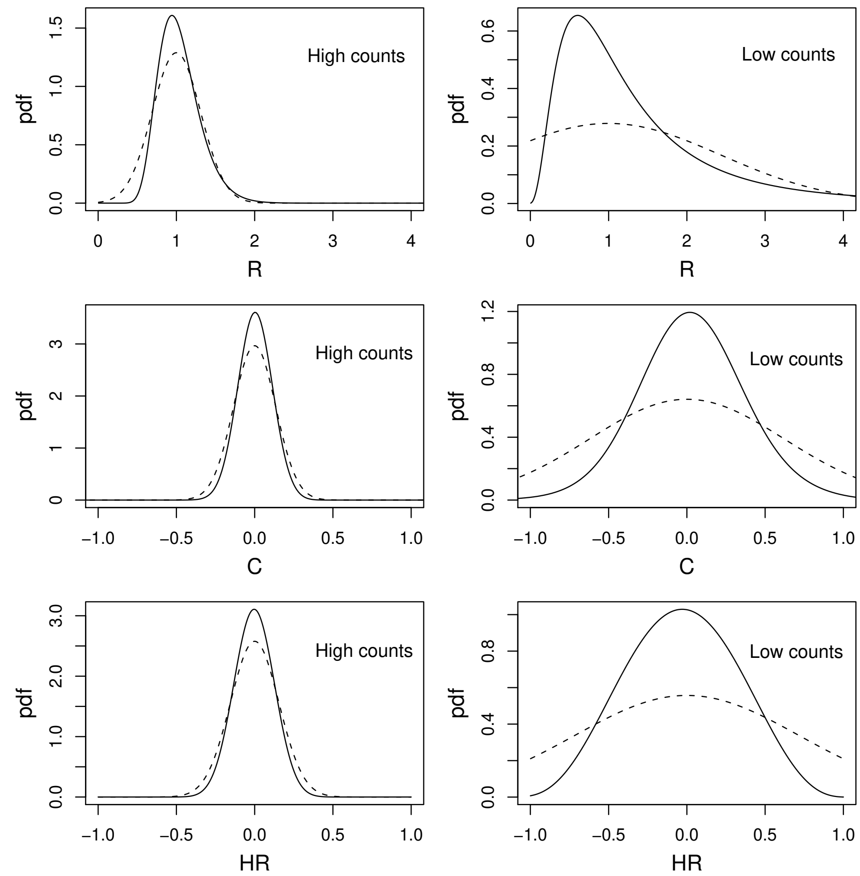

Fig. 1.— Comparison of the shape of posterior distribution for hardness ratios $\mathcal{R}$, C, and $\mathcal{HR}$ with the classical Gaussian approximation for high counts (Case I; left column of the figure) and for low counts (Case II; right column of the figure). The solid lines represent the posterior distributions of the hardness ratios and the dashed lines a Gaussian distribution with mean and standard deviation equal to the classical estimates. A prior distribution that is flat on the real line ($\phi = 1$, see §5.1) is adopted for the Bayesian calculations. Note that as the number of counts increase, the two distributions approach each other. At low counts the two distributions differ radically, with the Classical distributions exhibiting generally undesirable properties (broad, and extending into unphysical regimes).

Table 1: Statistical Properties of Bayesian and Classical Methods.

	Method	Hardness Ratio	Coverage Rate	Mean Length	Mean Square Error of mode	Mean Square Error of mean
CASE I (high counts)	Classical method	$\mathcal{R}$	95.0%	1.24	0.065	
		C	99.0%	0.54	0.013	
		$\mathcal{HR}$	98.5%	0.61	0.016	
	Monte Carlo method	$\mathcal{R}$	96.0%	1.07	**0.056**	0.069
		C	96.0%	0.44	**0.012**	0.012
		$\mathcal{HR}$	96.0%	0.49	0.016	**0.015**
	Quadrature	$\mathcal{R}$	94.5%	1.03	**0.057**	0.069
		C	96.0%	0.43	**0.012**	0.012
		$\mathcal{HR}$	94.5%	0.49	0.016	**0.015**
CASE II (low counts)	Classical method	$\mathcal{R}$	97.5%	192.44	93.27	
		C	100.0%	6.02	0.27	
		$\mathcal{HR}$	100.0%	3.70	0.21	
	Monte Carlo method	$\mathcal{R}$	98.0%	15.77	**0.328**	85.482
		C	98.0%	1.54	**0.078**	0.113
		$\mathcal{HR}$	98.0%	1.26	0.181	**0.083**
	Quadrature	$\mathcal{R}$	97.0%	8.18	**0.394**	20.338
		C	99.5%	1.51	**0.074**	0.112
		$\mathcal{HR}$	95.0%	1.23	0.187	**0.083**

In order to compare the statistical properties of estimates based on the Bayesian and classical methods, we simulate test data under both CASE I and CASE II. Table 1 presents the results of 200 simulated sources for each of the two cases. The data are used to compute point estimates and 95% error bars by using the classical method, Monte Carlo sampler, and numerical integration by quadrature. The Bayesian methods use non-informative prior distributions on λ and ξ; in particular, $\lambda \sim \gamma(1, 0)$ and $\xi \sim \gamma(0.5, 0)$, see §5.1. The results of CASE I indicate that all three methods are comparable although the quadrature yields slightly shorter intervals and point estimates with a smaller mean square error. Figure 2 illustrates that the resulting intervals from the methods are almost identical. On the other hand, the results of CASE II appear in Figure 3 and confirm that the Bayesian methods outperform the classical method because the mean length of the classical intervals are wider than the posterior intervals, and the classical estimates exhibit much larger mean square error. For example, the hardness ratio $\mathcal{HR}$ by definition is limited to the range $[-1, +1]$, so that the maximum length of the intervals should be two. However, the mean length of simulated

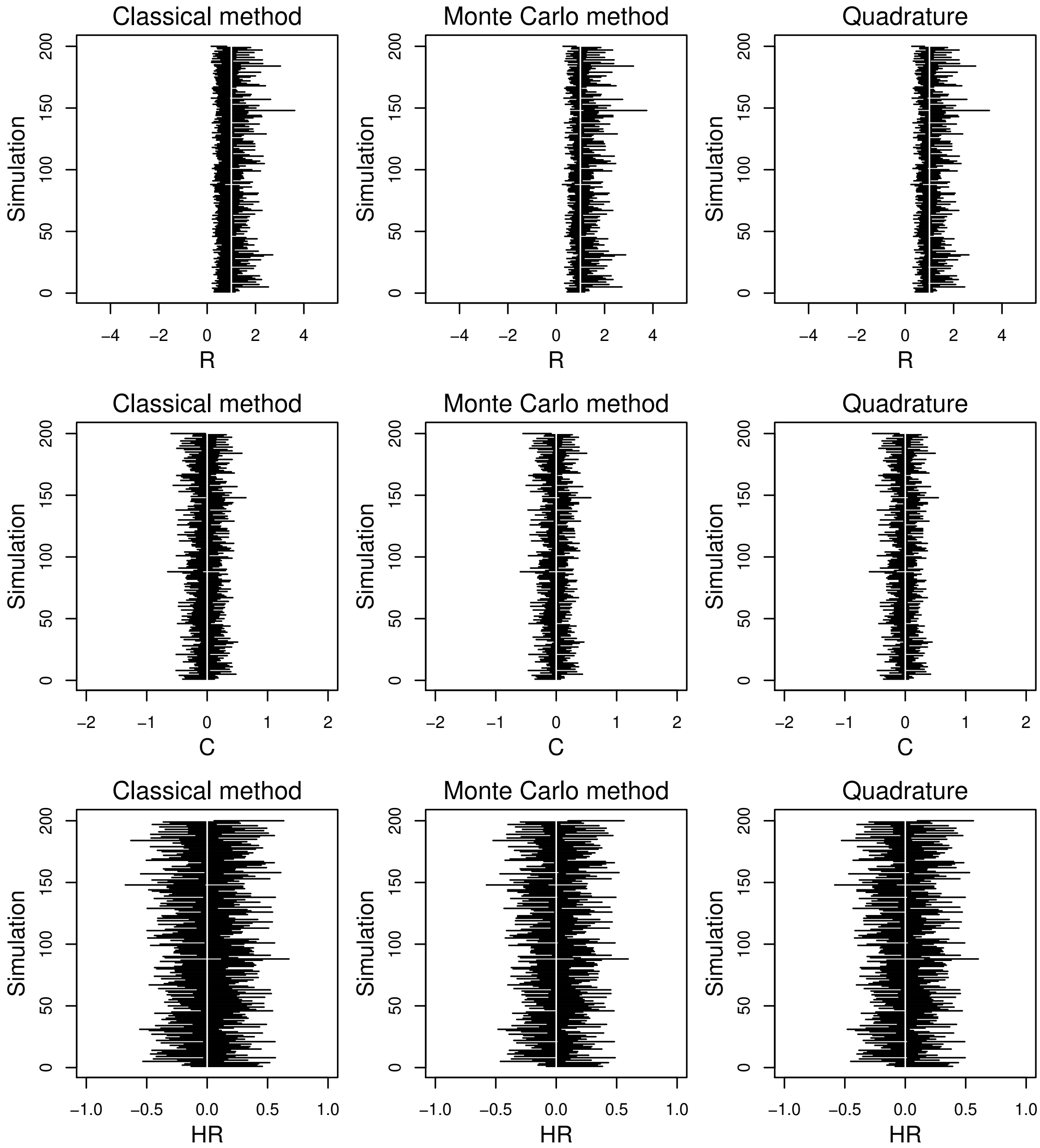

Fig. 2.— Simulated range of hardness ratios calculated using three different methods in the high counts case (Case I; see Equation 15a) – the classical method (*left*), Monte Carlo method (*middle*), and quadrature (*right*) – and for the three types of hardness ratios – $\mathcal{R}$ (*top*), C (*middle*), and $\mathcal{HR}$ (*bottom*). The horizontal lines are the 95% intervals computed for each simulated set of counts, and the vertical white line in each panel represents the true value of the hardness ratio. Note that all the different methods exhibit similar performance in this case.

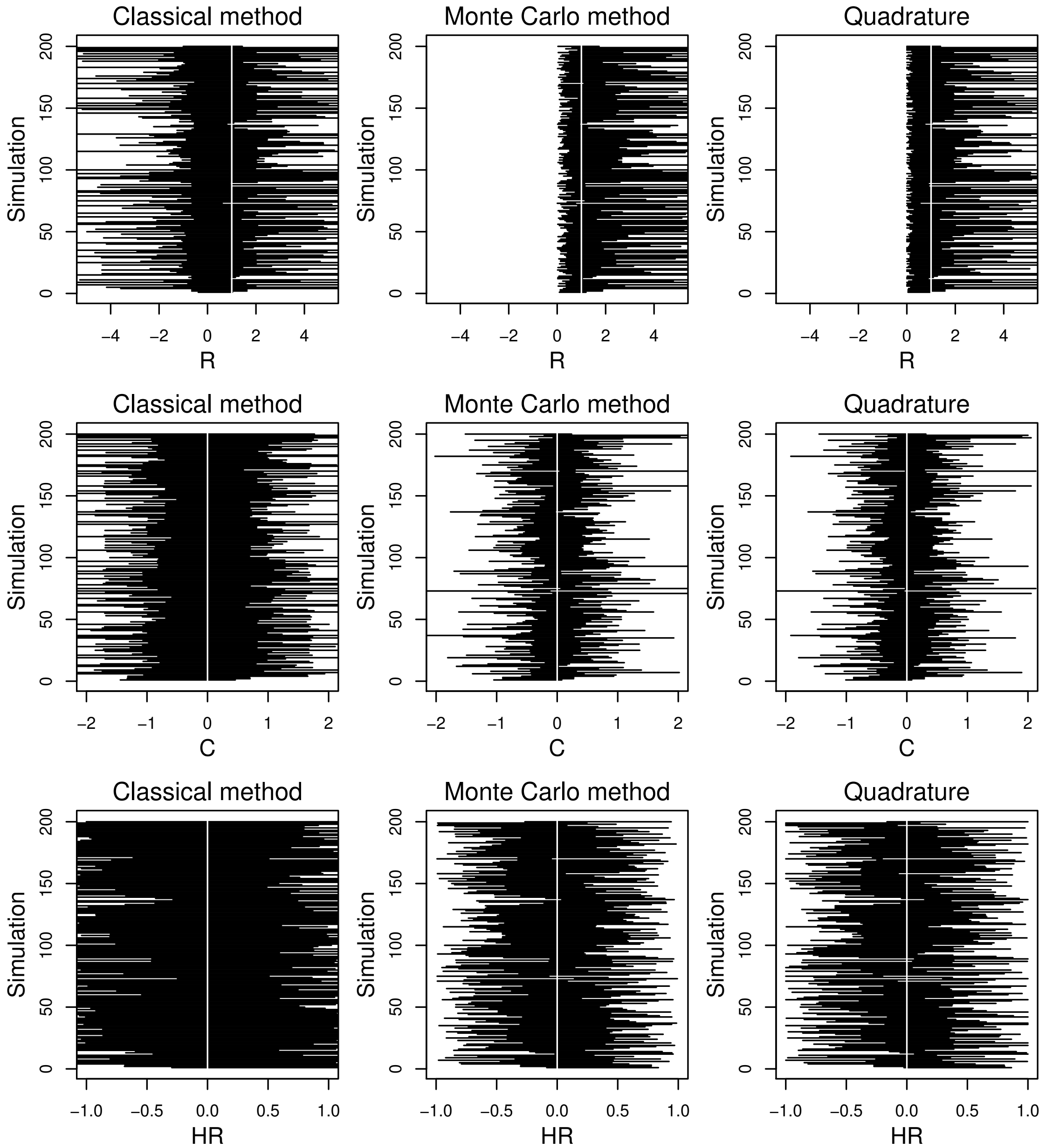

Fig. 3.— As in Figure 2, for simulated range of hardness ratios in the low counts case (Case II; Equation 15b). In this, the low counts case, the the Bayesian methods dramatically outperform the classical method.

classical intervals is 3.7, which is clearly larger than needed. Since quadrature tends to yield estimates with smaller mean square errors and shorter posterior intervals while maintaining high coverage rates, it is generally preferred to the Monte Carlo method. However, this comes at a price: because the summation inside the posterior density is over the detected source counts, quadrature is computationally more expensive as the source counts increase; on the other hand, the Monte Carlo method is much faster regardless of the number of source counts. Thus, we recommend using quadrature with relatively low counts data (e.g., $S < 20$ or $H < 20$) and the Monte Carlo method with higher counts.

Because the Bayesian inference for each hardness ratio is based on its posterior distribution, we consider these two point estimates, the posterior mode and the posterior mean. Table 1 presents a comparison of the two proposed estimates in terms of their mean square errors. In the case of $\mathcal{HR}$, the posterior mean seems to be more robust to the shape of a posterior distribution than a posterior mode because both left and right limits of $\mathcal{HR}$ are bounded. For $\mathcal{R}$ and C, on the other hand, the posterior mode appears to perform better. Thus, the posterior mode is preferable for both $\mathcal{R}$ and C and the posterior mean for $\mathcal{HR}$, as shown in Table 1.

4. Applications

4.1. Quiescent Low-Mass X-ray Binaries

Low-mass X-ray binaries (LMXBs) are binary systems composed of a neutron star or black-hole and a low mass ($< 1-2\ M_\odot$) donor. These binaries are generally transient sources and spend most of their time in quiescence (qLMXB's). However, even when not active they emit X-rays due to thermal emission from the surface of the neutron star, and sometimes due to an additional hard component associated with residual low-level accretion (e.g., Campana et al. 1998, Brown, Bildsten, & Rutledge 1998). Identification and study of these quiescent systems is very important since they provide a direct census of the overall LMXB population (and therefore the population of neutron stars), and also constrain their duty cycles. Although only the identification of their optical counterparts can unambiguously classify an X-ray source as a qLMXB, soft X-ray spectra or hardness ratios consistent with those expected for thermal emission by neutron stars are very useful for identifying candidate qLMXB's (Heinke et al. 2003).

We apply the methods described in §2 on a long (40 ks) *Chandra* ACIS-S observation of the globular cluster Terzan-5 (ObsID 3798; PI: R.Wijnands). A total of 49 sources were detected within the half-mass radius of the cluster (for details on the observation, reduction, and analysis, see Heinke et al. 2006, in preparation). We performed the source photometry in the 0.5-2.0 keV (*soft*) and 2.0-6.0 keV (*hard*) bands. In Figure 4 we present a color-luminosity diagram of all the detected sources (note that the color here is scaled by a factor 2.5 to be similar to the definition of an optical colors such as $B-V$). The posterior modes of the color for all detected sources, computed here using the Jeffrey's prior ($\phi = 1/2$), along with the 95% highest posterior density intervals, are shown. Those sources which are not amenable to analysis via the classical method (weak sources with background-subtracted source counts in at least one band being < 5) are marked with open circles, while the remainder are marked with filled circles. Thermally emitting neutron stars are expected to have X-ray colors in the range $[0.5, 2.5]$ (indicated by dashed lines in the figure). The softer limit (dashed curve to the right; color ≈ 2) corresponds to a purely thermal emission spectrum, and the harder limit (dashed curve to the left; color ≈ 0.5) corresponds to a 20% contribution from a power-law of index 1.5. From this figure it is clear that the Bayesian method, apart from providing more accurate confidence intervals for the source colors, allows us to estimate hardness ratios for the sources which were undetected in one of the two bands. We find 15 sources whose error bars overlap this region and are thus candidate qLMXB's. Note that 8 of these candidates could not have been so classified using the classical method because there are too few source counts. Conversely, many sources that lie at the edge of the region of interest would have been mistakenly classified as candidate qLMXB's with the classical method if the overly large error bars it produces had been used (see discussion of coverage rates in Table 1).

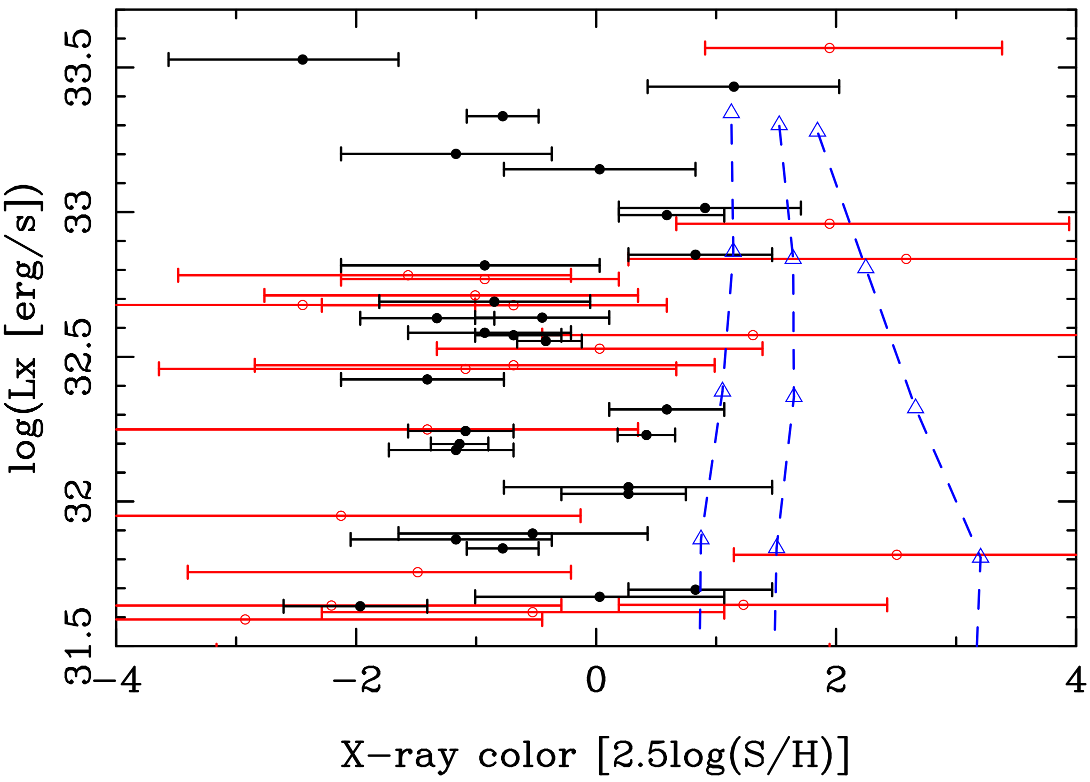

Fig. 4.— The X-ray "color-magnitude" diagram for globular cluster Terzan-5. For each of the detected sources, a form of color C = $\log_{10}(\lambda_S/\lambda_H)$, scaled by a factor 2.5 for similarity with optical colors, is plotted along the abscissa, and the $\log_{10}$(Luminosity [ergs s^{-1}]) in the 0.5-6 keV band, along the ordinate. The modes of the posterior probability functions (computed using quadrature) for all sources are shown, along with the 95% confidence intervals (the error bars on the luminosity are not plotted). The sources which would result in unreliable estimates if the color were computed with the classical method (background subtracted source counts < 5 in at least one band) are marked with open circles (red), and the remainder are marked with filled circles. The dashed (blue) lines represent the color-luminosity tracks expected for neutron stars with black-body spectra of different temperatures (the open triangles correspond to temperatures of $\log_{10}(T[\mathrm{K}]) = 6.3$, 6.2, 6.1 and 6.0 from top to bottom), and varying contribution from a power-law component (slope $\Gamma = 1.5$; pure black-body for the rightmost curve, 10% and 20% contribution from a power-law component in the 0.5-6 keV band luminosity for the middle and left curves respectively). For more details on these spectral models see Heinke et al. (2006).

4.2. Evolution of a Flare

In the previous section (§4.1), we demonstrated the usefulness of using the full Bayesian machinery in classifying weak sources in comparison to the classical method. Here, we shall go further and show that new avenues of analysis and interpretation are opened up when we take advantage of the full probabilistic description available to us. Such analysis methods are difficult to implement correctly and generally cannot be justified under the classical method. As a specific illustrative example, we consider the time evolution of a large X-ray flare on a low-mass dwarf, Ross 154 (Wargelin et al. 2006).

Stellar coronae consist of magnetically confined hot plasma ($T \sim 1-10$ MK) in the outer atmospheres of late-type stars, and emit optically thin thermal emission composed of Bremsstrahlung continuum and atomic line emission components. Detailed analysis of these line-rich spectra yield valuable clues regarding the temperature and emission structure on the star, its composition, and the processes that drive coronal heating (see e.g., Rosner, Golub, & Vaiana, 1985). Of special importance are flares, which result from impulsive energy input into the corona by the reconnection of highly strained magnetic flux tubes, and provide additional information on the dynamics of the corona. Analysis of flare spectra during decay yields constraints on the physical site and size of the flare loops and on the heating process (see, e.g., van den Oord & Mewe 1989, Pallavicini, Tagliaferri, & Stella 1990, Serio et al. 1991, Schmitt & Favata 1999).

A useful technique developed explicitly to analyze hydrodynamically evolving loops is to track their evolution in density-temperature space (Reale, Serio, & Peres 1993, Reale et al. 1997, Reale et al. 2004). This makes possible the modeling of flares to derive physically meaningful parameters such as the length of the loop, the heating function, etc. Unfortunately, a comprehensive analysis along these lines requires that complex model spectra be fit to the data at each stage of the flare evolution, and because fitting requires a large number of counts (generally > 500) to produce reliable results, the resolution at which the flare evolution can be studied is strongly limited. This is especially so during the rising phase of flares, where the physically most interesting processes are of short duration and the observed counts are also small. However, hardness ratios can be constructed at relatively finer time binning resolution and can be used as a proxy for the spectral fit parameters. In the following, we apply the Bayesian method described above (see §2) to *Chandra* data of a stellar X-ray flare, and demonstrate its usefulness.

A large flare was observed on Ross 154, a nearby (2.93 pc) active M dwarf (M4.5Ve), during a 57 ks observation with the *Chandra*/ACIS-S detector (ObsID 2561; PI: B.Wargelin). During the flare, the counting rate increased by over a factor of 100. Because of strong pileup in the core of the PSF, only those data within an annular region surrounding the center can be used. (Details of the reduction and analysis are given in Wargelin et al. 2006.) The light-curve of the source during the flare is shown in Figure 5, along with light-curves for smaller passbands (*soft*, 0.25-1 keV; *medium*, 1-2 keV; *hard*, 2-8 keV). Note that the *hard* band light-curve peaks at an earlier time than the softer bands, an impression that is borne out by the evolution of the color, C (lower panel of Figure 5).

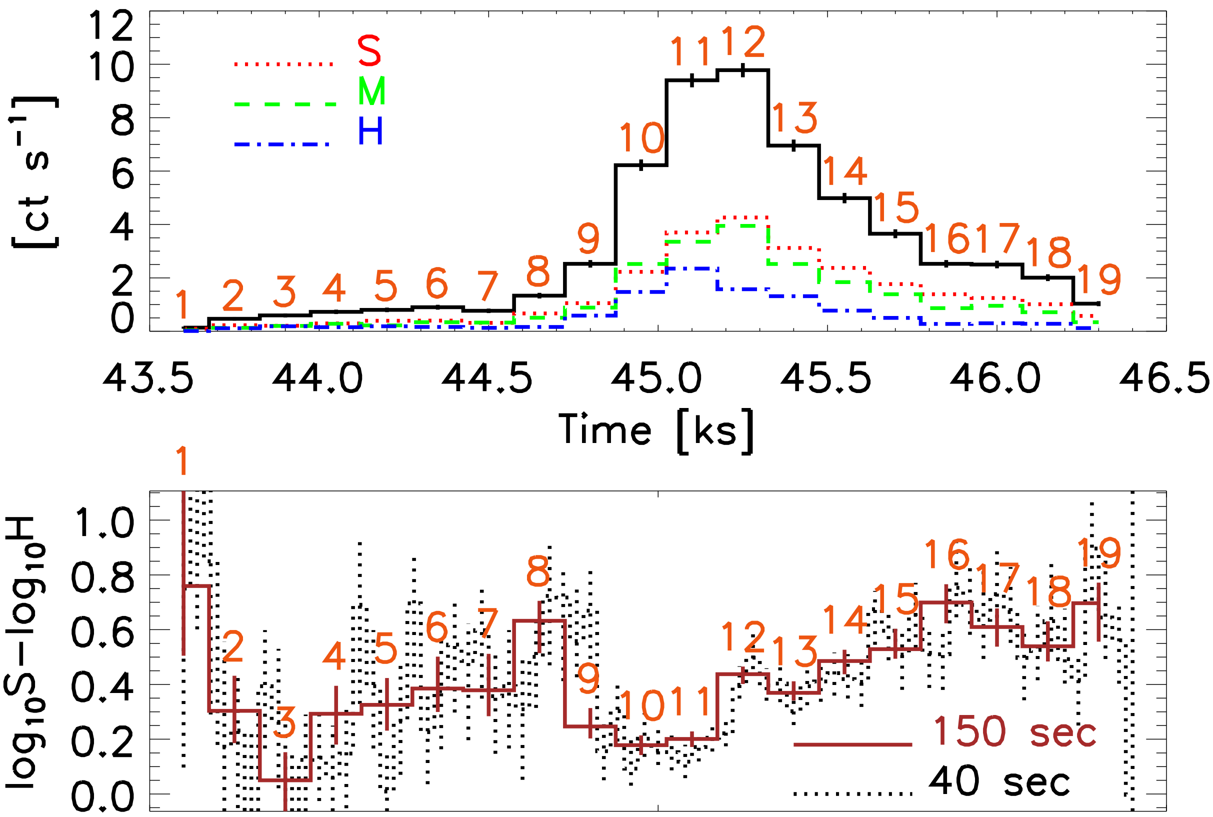

Fig. 5.— *Upper Panel:* Light-curves of Ross 154. The observed count rate in the source is shown as a function of time since the beginning of the observation, for counts in the *soft* (dotted histogram), *medium* (dashed histogram), *hard* (dash-dotted histogram), and the total (solid histogram). A bin size of 150 s was chosen for convenience, and the sequence of bins are marked along the light curve for future reference. *Lower Panel:* Hardness ratio evolution of Ross 154. The color, C = $\log_{10}\frac{S}{H}$ is plotted for a bin size of δt = 150 s (solid histogram) and δt = 40 s (dotted histogram), along with the associated error bars. Larger values of C indicate a softer spectrum, and vice versa.

The same data can also be represented in a color-intensity diagram, as in Figure 6, which shows scatter plots of the color and intensity at each time bin along with their computed errors (connected by a thin line to indicate the time sequence). The advantage of doing so is that color is inversely correlated with plasma temperature, and the intensity approximately tracks the plasma density.[j] A hysteresis-like evolution is discernible in the color-intensity track, even at a time resolution as small as 40 s. However, the error bars on the data points are large enough to obscure this structure at small time bin sizes, whereas at larger bin sizes, the number of data bins are smaller and it becomes difficult to discern the totality of the evolutionary track, even though the pattern is persistent as bin size and starting times are varied. Mainly, the choices of the bin size and the phase (i.e., the starting point) is completely arbitrary and the choice of no single value can be properly justified. However, Bayesian analysis provides a simple method to work around this difficulty, by simply considering the bin size as a nuisance parameter of the problem and to marginalize over it.

Thus, in order to ameliorate the effects of choosing a single time bin size and a starting time, we have carried out a series of calculations using the Monte Carlo method: we obtain different light-curves by varying the phase of the binning (i.e, the starting time), and for each light-curve thus obtained, we obtain 2000 samples of (λ_S, λ_H) for each data point. We hold the time bin size fixed during this process. After cycling through all the phases, the samples of (λ_S, λ_H) are all combined to form a single image (the shaded images in Figure 6). This procedure, also called "cycle-spinning" (Esch 2003), allows us to discern the pattern of the evolutionary track clearly. The longer the source spends at any part of the color-intensity diagram, the darker the shade in the images, and the statistical error in the calculation is reflected by the gradient of the shading. Further, all the cycle-spun images thus constructed, for bin sizes ranging from 40 s to 400 s, were then averaged to produce a single coherent image, thus effectively marginalizing over the bin sizes; such an image includes all the errors associated with the analysis (Figure 7).

Note that this type of analysis is not possible using the classical method. For instance, in the Bayesian case, the samples are drawn from the posterior probability distributions of S and H. No such sampling technique is available in the classical case,[k] and naively using

---

[j]X-ray luminosity, $L_X \propto EM \cdot P(T)$, where $EM = n_e^2 V$ is the emission measure of coronal plasma with an electron density n_e occupying a volume V, and $P(T)$ is the power emitted by the plasma at temperature T. $P(T)$ is a weak function of T over the temperature range 6-20 MK expected in this source, and assuming that the volume of emission remains constant, the observed intensity is proportional to the square of the electron density. For a more detailed modeling of the $L_X - T$ conversion, see Wargelin et al. (2006).

[k]New counts realizations can be obtained by bootstrapping from the observed counts, but such sampling will be quantized and will render the shaded images unusable at lower counts intensities.

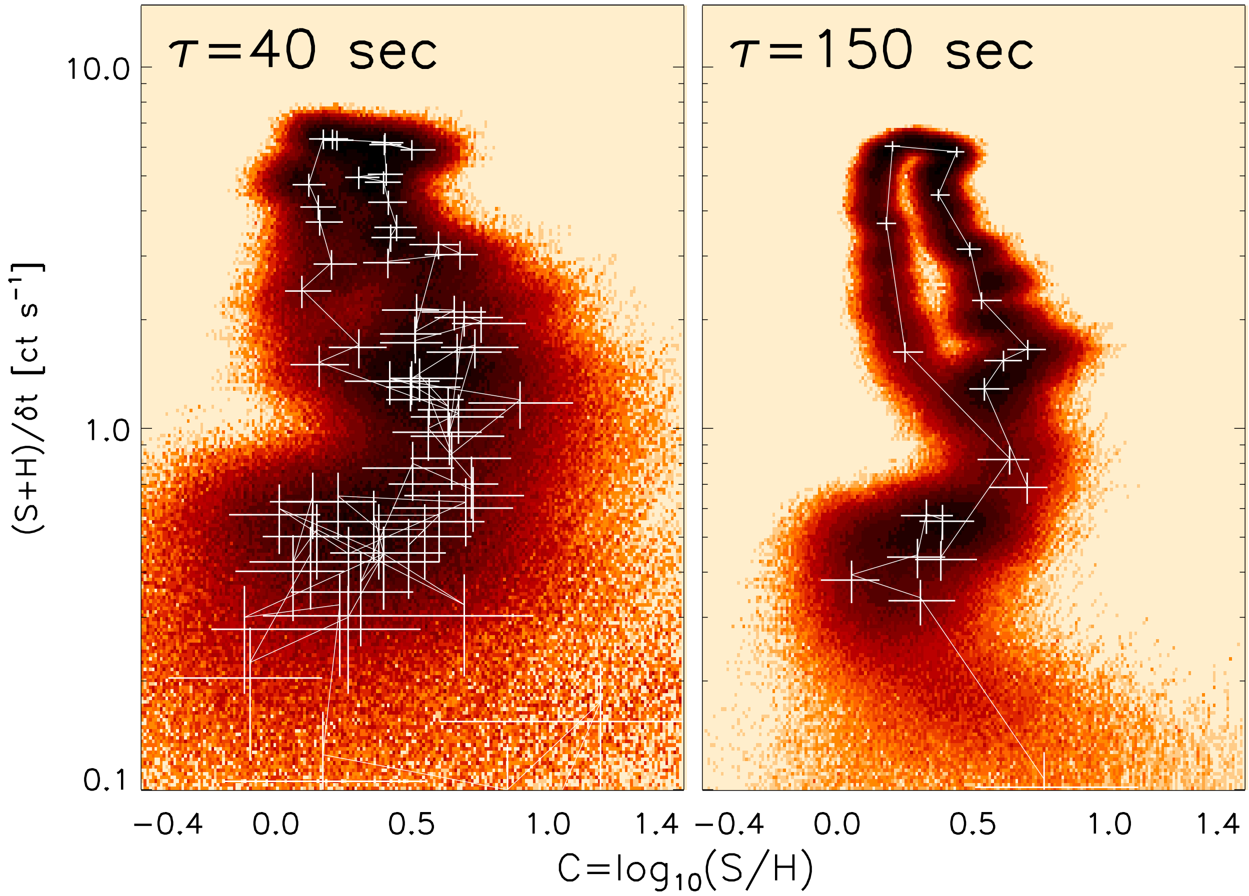

Fig. 6.— Color-intensity evolutionary tracks. Spectral hardness increases to the left and source brightness increases towards the top. The paired color and intensity points (shown separately in Figure 5) are plotted for bin sizes of 40 s (*left*) and 150 s (*right*). The error bars for each quantity are also plotted as horizontal and vertical bars. The points are connected by a thin solid line to aid in visualization of the temporal sequence (see also Figure 7). Samples of (λ_S, λ_H) obtained from their posterior probability distributions for each bin and for various starting time phases are displayed as the shaded image underneath the evolutionary tracks.

the Gaussian approximations as in Equation 4 will overestimate the errors and wash out the signal (see §3). Furthermore, the concept of marginalizing over the nuisance parameter of time bin sizes has no analog in the classical system, and cannot be justified; and cycle-spinning leads to correlations between the estimates at different bins since to the statistical independence of the data will be lost.

A sharp initial hardening of the spectrum, coincident with the onset of a small optical flare (Wargelin et al. 2006) is visible at the beginning of the time sequence (points 1-3; see Figures 5,6). A point worth noting is that the Bayesian analysis prevents overinterpretation of the results: at time point 3, the standard analysis indicates that the spectrum is at its hardest, suggesting either a non-thermal origin to the emission or significant transient absorption, possibly due to a coronal mass ejection, but the shaded image demonstrates that the statistical errors are sufficiently large that a more mundane low-density thermal expansion is quite plausible. This stage is followed by a rapid softening (points 3-8), interpreted as the thermalization of the non-thermal hard X-ray flare. Then the star, which lies in the lower central portion of the color-intensity diagram before the flare, moves to the upper left as the flare is set off (the spectrum hardens as intensity increases; points 8-11), turns right at flare peak (softens as the deposited energy cascades to lower temperatures; points 11-12), and then moves down the right flank back to its original state (decays in temperature and intensity; points 12-19). The physical consequences of this analysis are discussed in detail by Wargelin et al.

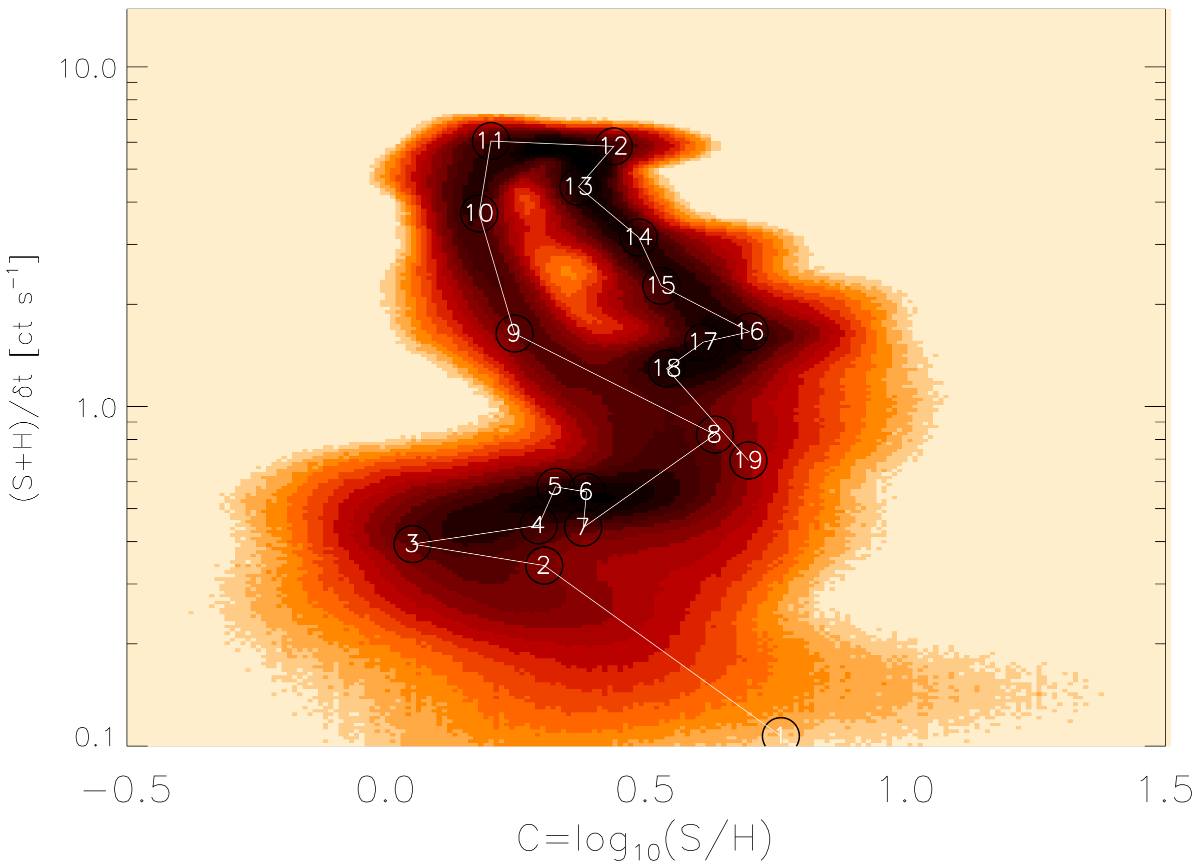

Fig. 7.— Ross 154 color-intensity track. As in Figure 6, but for an image obtained by averaging cycle-spun images at various bin sizes ranging from 40 s to 400 s. The persistence of the evolutionary track of the flare on Ross 154 is quite clear in this visualization. The points corresponding to each bin of the light-curve obtained at a bin size of 150 s are superimposed on the image, with a white solid line connecting them in temporal sequence. A clear pattern of the flaring plasma evolution can be seen, first rising in temperature and intensity, softening at the peak, and then decaying in temperature and intensity.

5. Discussion

5.1. Informative and Non-informative Prior Distributions

A major component of Bayesian analysis is the specification of a prior distribution, the probability distribution assigned *a priori* to the parameters defining the model. This is often deemed controversial because of the apparent subjectivity in the assignment: if different people assign different priors and obtain different results, how is it possible to evaluate them? The answer to this conundrum is to realize that all problems always include a bias brought to the analysis by the researcher, either in the choice of statistic, or even in the choice of analysis method, and the specification of a prior distribution codifies and makes explicit this bias. Furthermore, if indeed prior information about the value or the range of the parameter is available, the prior allows a simple and principled method by which this information can be included in the analysis. If no prior information is available, a so-called *non-informative* prior distribution must be chosen in order to not bias the result.

In our case, if there is a strong belief as to the value or range of the hardness ratio, we can incorporate this via an informative prior distribution, encoded as a γ-distribution (see van Dyk et al. 2001). In addition, the analysis of previously acquired data will produce a posterior probability distribution that can be used as an informative prior on future observations of the same source.

When no prior information is available, we normally use a non-informative prior distribution. Since a Poisson intensity is necessarily positive, three types of non-informative prior distributions immediately present themselves: when $X|\theta \sim \text{Poisson}(\theta)$,

1. a non-informative prior distribution on the original scale,

$$p(\theta) \propto 1\,, \tag{16a}$$

2. a Jeffrey's non-informative prior distribution,

$$p(\theta) \propto I_\theta^{1/2}\,, \text{ and} \tag{16b}$$

3. a non-informative prior distribution under a log transformation,

$$p(\log\theta) \propto 1 \text{ or equivalently } p(\theta) \propto \frac{1}{\theta}\,, \tag{16c}$$

where $I_\theta = \mathrm{E}[-\partial^2 \log p(X|\theta)/\partial\theta^2|\theta]$ is the expected Fisher information (Casella & Berger 2002). The first non-informative prior distribution (Equation 16a) is flat from 0 to ∞; the second (Equation 16b) is proportional to the square root of the Fisher information; and the third (Equation 16c) is flat on the whole real line under a log transformation. The functional forms of these prior distributions can be generalized into $p(\theta) \propto \theta^{\phi-1}$, i.e., $\theta \sim \gamma(\phi, 0)$ where ϕ is an index that varies from 0 to 1: the first corresponds to $\phi = 1$, the second to $\phi = 1/2$; and the third to $\phi = 0^+$ (where this notation indicates that $\phi > 0$, but arbitrarily close to zero, for reasons described below). We note that these non-informative prior distributions are *improper*, i.e., are not integrable. If an improper prior distribution causes a posterior distribution to be improper as well, then no inferences can be made using such non-integrable distributions. In our case, as long as ϕ is strictly positive, a posterior distribution remains *proper*, i.e., integrable. Hence, in the third case, we adopt values of ϕ that are strictly positive but close in value to 0, e.g., $\phi = 0.01$ or $\phi = 0.1$.

Note that while these prior distributions are non-informative, in the sense that in most cases they do not affect the calculated values of the hardness ratios (see Appendix B), they do codify some specific assumptions regarding the range of values that it is believed possible. For instance, if a large dynamic range is expected in the observations, a flat prior distribution in the log scale is more appropriate than a flat prior distribution in the original scale. The choice of the prior distribution is dictated by the analysis and must be reported along with the results.

5.2. $\mathcal{R}$ versus C versus $\mathcal{HR}$

At low counts, the posterior distribution of the counts ratio, $\mathcal{R}$, tends to be skewed because of the Poissonian nature of data; $\mathcal{R}$ only takes positive values. The color, $\mathrm{C} = \log_{10} \mathcal{R}$, is a log transformation of $\mathcal{R}$, which makes the skewed distribution more symmetric. The fractional difference hardness ratio, $\mathcal{HR} = (1-\mathcal{R})/(1+\mathcal{R})$, is a monotonically decreasing transformation of $\mathcal{R}$, such that $\mathcal{HR} \rightarrow +1$ as $\mathcal{R} \rightarrow 0$ (i.e., a source gets harder) and $\mathcal{HR} \rightarrow -1$ as $\mathcal{R} \rightarrow \infty$ (i.e., a source gets softer). The monotonic transformation results in a bounded range of $[-1, +1]$ for $\mathcal{HR}$, thereby reducing the asymmetry of the skewed posterior distribution. $\mathcal{R}$ and $\mathcal{HR}$ are bounded on one side or two sides, while C is unbounded due to the log transformation.

The posterior distribution of any hardness ratio becomes more symmetric as both soft and hard source intensities increase. Regardless of the size of the intensities, however, the color has the most symmetric posterior distribution among the popular definitions of a hardness ratio. Figure 8 illustrates the effect of the magnitude of the source intensities on the symmetry of the posterior distribution of each hardness ratio; the posterior distribution of C is confirmed to have the most symmetric posterior distribution. In the figure, we fix

$\mathcal{R} = 2$ and the soft and hard intensities are determined by beginning with $\lambda_S = 2$ and $\lambda_H = 1$ (in units of counts (source area)$^{-1}$) and increasing the intensities by a factor of 5 in each subsequent column. We assume no background contamination in each simulation.

Figure 8 also illustrates the effect of the use of different indices for the non-informative prior distribution on the resulting posterior distribution of each hardness ratio. In the case of $\mathcal{R}$, the posterior distribution becomes diffuse and more skewed toward 0 as the value of ϕ decreases; as expected, this is also the case for C. And in the case of $\mathcal{HR}$, the posterior distribution has more mass near the boundaries (i.e., ± 1) as the value of ϕ decreases, thereby exhibiting a U-shaped density in the case of very faint sources. As the intensities increase, however, the choice of a prior distribution has less effect on the posterior distribution of any of the hardness ratios.

Finally, as pointed out in §3, for $\mathcal{R}$ and C, the mode of the posterior probability distribution is a robust estimator, whereas for $\mathcal{HR}$, the mean of the distribution is a better choice. (Notice that the posterior mode of $\mathcal{HR}$ in the low count scenario is -1 when $\phi = 0^+$ in Figure 8.)

5.3. Advantages

A significant improvement that our method provides over the classical way of computing hardness ratios is that we use the correct Poisson distribution throughout and do not make any limiting (high counts) Gaussian assumptions. Thus, while the classical method works well only with high counts data and fails to give reliable results with low counts, our method is valid in all regimes. Indeed, because the observed counts are non-negative integers, it is not appropriate to model low counts with Gaussian distributions which are defined on all the real numbers.

Because our methods are based on Poisson assumptions in a fully model based statistical approach, we need not rely on plug-in estimates of the hardness ratio. Instead we can evaluate the full posterior probability distribution of the hardness ratio, which provides reliable estimates and correct confidence limits even when either or both soft and hard counts are very low. In particular, our method is not limited to "detectable" counts, and can properly calculate upper and lower bounds in cases where the source is detected in only one passband. For high counts, our method gives more precise error bars for the hardness ratios than the classical method (as defined by the coverage; see Table 1 and Appendix A), although both methods yield similar results.

Further, *a priori* information can be embedded into our model and can be updated by

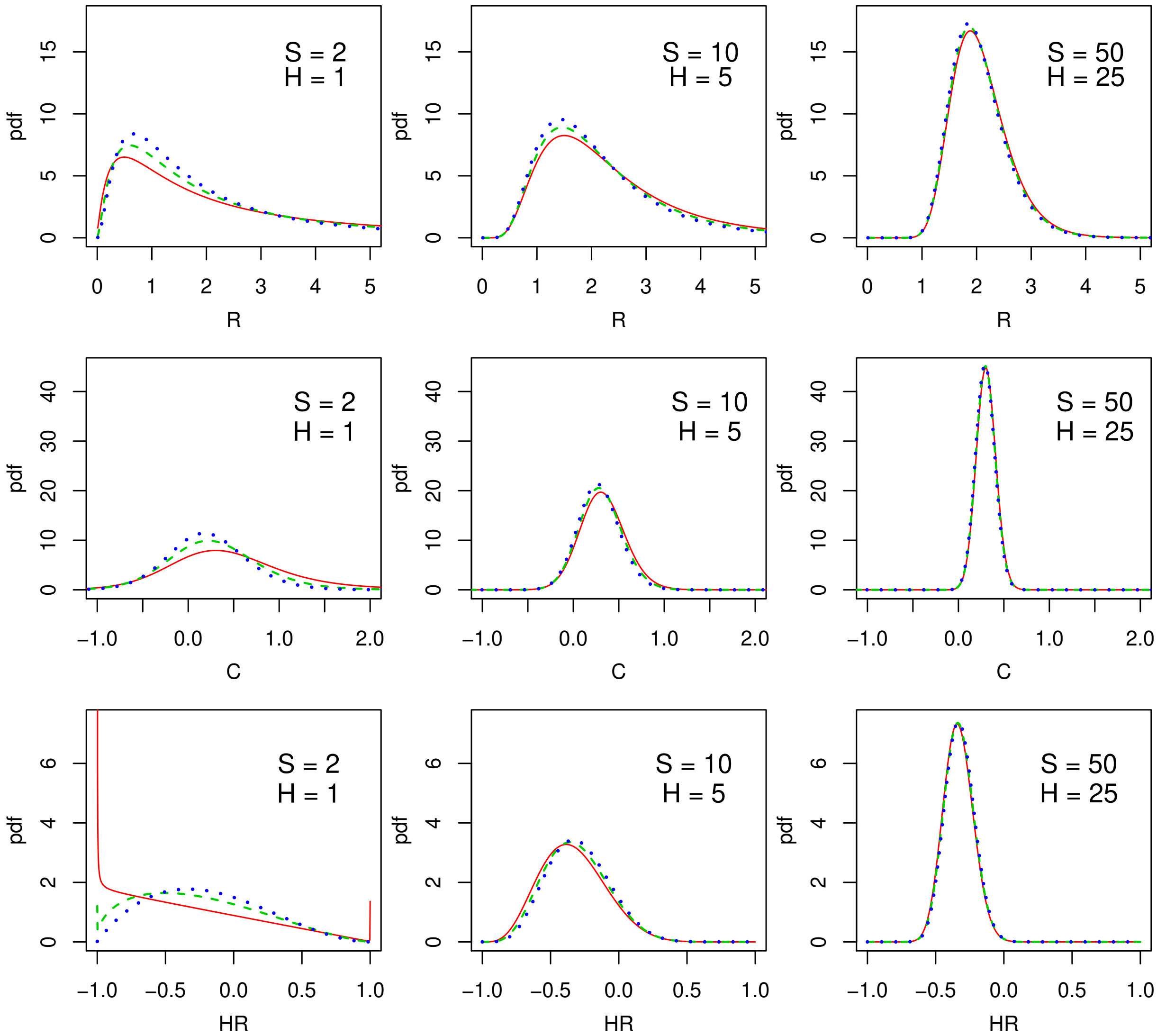

Fig. 8.— The posterior distributions of hardness ratios calculated for $\mathcal{R}$ (top row), C (middle row), and $\mathcal{HR}$ (bottom row), with different source intensities (low at left and high at right) and prior distributions. The different curves in each figure represent the posterior probability distributions computed with different non-informative prior distribution indices, i.e., $\phi = 0^+$ (solid), $\phi = 1/2$ (dashed), and $\phi = 1$ (dotted). At higher counts (large Poisson intensities; right column of figures), the posterior distributions tends to be symmetric and the effect of the prior distributions is minimal, i.e., the posterior distributions with different indices are nearly the same. At small counts (small Poisson intensities; the left column of figures), the non-symmetric shape of the posterior distribution for each hardness ratio becomes clear as does the effect of the choice of a non-informative prior distribution.

data, thereby producing more accurate results as we collect more data.

5.4. Limitations

Unlike the quantile-width method (Hong et al. 2004), the calculation of hardness ratios is limited by necessity to pre-selected non-overlapping passbands. In general, we must use different band definitions in order to achieve the maximum discriminating power for soft and hard sources. *A priori*, with no knowledge of the character of the source populations, there is no way to ensure that the chosen passbands are the most efficient for splitting the spectrum and obtaining maximum leverage to distinguish spectral model parameters. However, this restriction allows for a more uniform analysis and comparison of results across different instruments and telescopes.

Because the Bayesian method does not allow for a simple analytic solution similar to standard error propagation in the Gaussian case, the computational methods used to determine the probability distributions become important. We have implemented a Monte Carlo integration method (the Gibbs sampler, see Appendix A.1) and a method based on numerical integration (quadrature, see Appendix A.2). The Gibbs sampler is based on Monte Carlo simulation and hence causes our estimates to have simulation errors in addition to their true variability; to reduce the levels of simulation errors, the Markov chain must be run with large enough iterations. On the other hand, the quadrature precisely computes the posterior distribution as long as the number of bins is large enough; however, its computation becomes expensive as the counts become large because of the binomial expansion in Equation A5. In general, the Gibbs sampler is much faster than the method based on quadrature, but care must be taken to ensure that the number of iterations is sufficient to determine the posterior mode and the required posterior interval with good precision. We recommend using the Gibbs sampler for high counts and the quadrature for low counts.

6. Summary

We have developed a method to compute hardness ratios of counts in non-overlapping passbands that is fully Bayesian and properly takes into account

1. the Poissonian nature of the data,

2. background counts, allowing for differences in collecting areas, effective areas, and exposure times between the source and background regions, and

3. both informative and non-informative prior information.

The algorithm is implemented in two forms: a Monte Carlo based Gibbs sampler that samples from the posterior probability distributions of the counts intensities, and another that numerically integrates the joint probability distribution of the counts intensities using quadrature.

We carry out comparison tests between our method and the classical method, and show that the Bayesian method works even in the low counts regime where the classical method fails, and matches the results of the classical method at high counts where a Gaussian approximation of the errors is appropriate.

We apply the method to real world test cases of

1. determining candidate quiescent Low-Mass X-ray Binaries in the Terzan-5 cluster, where we show that it can be applied even in cases where the classical method is entirely inapplicable, such as those where the source is detected in only one passband, and

2. tracking the time evolution of a flare on the dMe star Ross 154, where we demonstrate the flexibility of the method in being able to work around the conflicting limitations imposed by time bin sizes and time resolution.

Additional applications of these methods include spectral classification of the X-ray sources detected in deep surveys and other galaxies, and spectral variability of AGNs and X-ray binaries. The physical properties of the source influence the source intensities λ_S and λ_H; given a source spectral model, the two quantities are related. Thus, variations in the hardness ratios reflect changes in the source properties (e.g., sources with different temperatures have different colors). A relation between source properties, physical model parameters and hardness ratio is a natural extension of this work and will be presented in the future.

We thank Yue Wu for assistance with code development, and Dong-Woo Kim and Peter Edmonds for useful discussions. The authors gratefully acknowledge funding for this project partially provided by NSF grants DMS-01-04129, DMS-04-38240, and DMS-04-06085 and by NASA Contracts NAS8-39073 and NAS8-03060 (CXC), and NASA grant NAG5-13056. This work is a product of joint work with the California-Harvard astrostatistics collaboration (CHASC) whose members are TP, VK, AS, DvD, and AZ, and J. Chiang, A. Connors, D. Esch, P. Freeman, H. Kang, X.-L. Meng, J. Scargle, E. Sourlas, and Y. Yu.

A. Computing the Hardness Ratios

A.1. Monte Carlo Integration: Gibbs Sampler

Monte Carlo methods can be used in Bayesian statistical inference if representative values of the model parameters can be simulated from the posterior distribution. Given a sufficient simulation size, these simulated values can be used to summarize and describe the posterior distribution and hence the Bayesian statistical inference. For example, the posterior mean can be computed by averaging these simulated values and posterior intervals are computed using quantiles of the the simulated values. Markov chain Monte Carlo (MCMC) methods (Tanner & Wong 1987) accomplish Monte Carlo simulation by constructing a Markov chain with stationary distribution equal to the target posterior distribution. The Markov chain is then run, and, upon convergence delivers the desired Monte Carlo simulation of the posterior distribution. The Gibbs sampler (Geman & Geman 1984) is a special case of MCMC that constructs the Markov chain through alternating conditional simulation of a subset of the model parameters conditioning on the others. A more detailed description of these methods can be found in van Dyk (2003) and the Appendix of van Dyk et al. (2001).

In order to construct a Gibbs sampler, we embed the statistical model described in §2.1 into a larger model. Since the observed counts are the sum of the of source counts (η) and the background counts (β), we write $S = \eta_S + \beta_S$ and $H = \eta_H + \beta_H$. That is, the source counts are modeled as independent Poisson random variables,

$$\eta_S \sim \text{Poisson}(e_S \cdot \lambda_S) \quad \text{and} \quad \eta_H \sim \text{Poisson}(e_H \cdot \lambda_H), \tag{A1}$$

and the background counts in the source exposure area as independent Poisson random variables,

$$\beta_S \sim \text{Poisson}(e_S \cdot \xi_S) \quad \text{and} \quad \beta_H \sim \text{Poisson}(e_H \cdot \xi_H), \tag{A2}$$

where λ and ξ denote the expected source and background counts in the source region. This implies Equation 5, where the detected photons in the source region are a convolution of the source and background counts, i.e., $S = \beta_S + \eta_S$ and $H = \beta_H + \eta_H$.

Using the method of data augmentation, we treat the background counts in the source exposure area (β_S and β_H) as missing data; the source counts (η_S and η_H) are fully determined when β_S and β_H are known. Intuitively, it is straightforward to estimate the Poisson intensities if detected photons are split into the source and background counts. Based on this setting, the Gibbs sampler produces Monte Carlo draws of λ_S and λ_H along with stochastically imputing the missing data (β_S and β_H). The conditional independence of λ_S and λ_H implies that a Monte Carlo simulation of each hardness ratio can be obtained by simply transforming that of λ_S and λ_H. Due to the conditional independence, both soft and hard bands also have exactly the same sampling steps, and thus we illustrate the Gibbs sampler only for the soft band. First, the joint posterior distribution of λ_S, ξ_S, and β_S is given by

$$\begin{aligned} p(\lambda_S, \xi_S, \beta_S | S, B_S) &\propto p(S|\lambda_S, \beta_S) p(B_S|\xi_S) p(\beta_S|\xi_S) p(\lambda_S) p(\xi_S) \\ &\propto \frac{1}{(S-\beta_S)!\beta_S!} \lambda^{S-\beta_S+\psi_{S_1}-1} \xi^{B+\beta_S+\psi_{S_3}-1} e^{-(e_S+\psi_{S_2})\lambda_S-(e_S+r\cdot e_S+\psi_{S_4})\xi_S}. \end{aligned} \tag{A3}$$

That is, conditional on the total soft counts (S), the unobserved background counts in the source exposure area (β_S) follows a binomial distribution: Given the current iterates of the parameters, $\lambda_S^{(t)}$ and $\xi_S^{(t)}$, Step 1 is given by

Step 1 : Draw $\beta_S^{(t+1)}$ from $\beta_S | (\lambda_S^{(t)}, \xi_S^{(t)}, S, B_S) \sim \text{Binomial}\left(S, \dfrac{\xi_S^{(t)}}{\lambda_S^{(t)} + \xi_S^{(t)}}\right)$,

for $t = 1, \ldots, T$. Next, Steps 2 and 3 draw the source and background intensities from the γ-distributions. In particular, Steps 2 and 3 find the next iterates of the intensities using

Step 2 : Draw $\lambda_S^{(t+1)}$ from $\lambda_S | (\xi_S^{(t)}, \beta_S^{(t+1)}, S, B_S) \sim \text{Gamma}(S - \beta_S^{(t+1)} + \psi_{S_1},\, e_S + \psi_{S_2})$ and

Step 3 : Draw $\xi_S^{(t+1)}$ from $\xi_S | (\lambda_S^{(t+1)}, \beta_S^{(t+1)}, S, B_S) \sim \text{Gamma}(B_S + \beta_S^{(t+1)} + \psi_{S_3},\, e_S + r \cdot e_S + \psi_{S_4})$,

for $t = 1, \ldots, T$. To accomplish one Gibbs iteration, these steps are implemented for the soft and hard bands. After iterating the Gibbs sampler T times, we collect a posterior sample $\{\lambda_S^{(t)}, \lambda_H^{(t)}, t = t_0 + 1, \ldots, T\}$ for a sufficiently long burn-in period[1] t_0. The analytical calculation to determine burn-in is far from computationally feasible in most situations. However, visual inspection of plots of the Monte Carlo output is commonly used for determining burn-in. More formal tools for determining t_0, called convergence diagnostics, have been proposed; for a recent review, see Cowles & Carlin (1996). Under a monotone transformation (see Equations 12,13,14) of the posterior samples, $(T - t_0)$ Monte Carlo draws for each hardness ratio are obtained, which enables us to find its point estimates and the corresponding error bar; see §2.2.1 for details.

[1]The term "burn-in" refers to the number of iterations necessary for the Markov chain to converge and begin to sample from the posterior probability distribution.

A.2. Numerical Integration: Gaussian Quadrature

Instead of introducing the missing data, the Bayes' theorem can analytically compute the joint posterior distribution of λ_S and λ_H, based on which we obtain the posterior distribution of each hardness ratio through quadrature. Because the models for the hard and soft bands are independent and symmetric, we illustrate the computation only for the soft intensity λ_S. First, the joint posterior distribution of λ_S and ξ_S in Equation 8 is written out as

$$\begin{aligned} p(\lambda_S, \xi_S | S, B_S) &= \frac{p(\lambda_S)p(\xi_S)p(S|\lambda_S, \xi_S)p(B_S|\xi_S)}{\int_0^\infty\int_0^\infty p(\lambda_S)p(\xi_S)p(S|\lambda_S, \xi_S)p(B_S|\xi_S) d\xi_S\, d\lambda_S} \\ &= \frac{(\lambda_S+\xi_S)^S \lambda_S^{\psi_{S_1}-1} \xi_S^{B_S+\psi_{S_3}-1} e^{-(e_S+\psi_{S_2})\lambda_S} e^{-(e_S+r\cdot e_S+\psi_{S_4})\xi_S}}{\int_0^\infty\int_0^\infty (\lambda_S+\xi_S)^S \lambda_S^{\psi_{S_1}-1} \xi_S^{B_S+\psi_{S_3}-1} e^{-(e_S+\psi_{S_2})\lambda_S} e^{-(e_S+r\cdot e_S+\psi_{S_4})\xi_S} d\xi_S d\lambda_S}. \end{aligned} \tag{A4}$$

Then, the binomial expansion to $(\lambda_S+\xi_S)^S$, i.e.,

$$(\lambda_S+\xi_S)^S = \sum_{j=0}^{S} \frac{\Gamma(S+1)}{\Gamma(j+1)\Gamma(S-j+1)} \lambda_S^j \xi_S^{S-j}, \tag{A5}$$

enables us to analytically integrate ξ_S out of the joint distribution in Equation A4; and obtain an analytical solution to the marginal posterior distribution of λ_S. Therefore, using the binomial expansion, Equation A4 becomes

$$p(\lambda_S, \xi_S | S, B_S) = \frac{\sum_{j=0}^{S} \frac{\Gamma(S+1)}{\Gamma(j+1)\Gamma(S-j+1)} \lambda_S^{j+\psi_{S_1}-1} \xi_S^{S-j+B_S+\psi_{S_3}-1} e^{-(e_S+\psi_{S_2})\lambda_S} e^{-(e_S+r\cdot e_S+\psi_{S_4})\xi_S}}{\sum_{j=0}^{S} \frac{\Gamma(S+1)}{\Gamma(j+1)\Gamma(S-j+1)} \cdot \frac{\Gamma(S-j+B_S+\psi_{S_3})}{(e_S+r\cdot e_S+\psi_{S_4})^{S-j+B_S+\psi_{S_3}}} \cdot \frac{\Gamma(j+\psi_{S_1})}{(e_S+\psi_{S_2})^{j+\psi_{S_1}}}} \tag{A6}$$

and then ξ_S is integrated out of Equation A6, i.e.,

$$p(\lambda_S | S, B_S) = \frac{\sum_{j=0}^{S} \frac{1}{\Gamma(j+1)\Gamma(S-j+1)} \cdot \frac{\Gamma(S-j+B_S+\psi_{S_3})}{(e_S+r\cdot e_S+\psi_{S_4})^{S-j+B_S+\psi_{S_3}}} \lambda_S^{j+\psi_{S_1}-1} e^{-(e_S+\psi_{S_2})\lambda_S}}{\sum_{j=0}^{S} \frac{1}{\Gamma(j+1)\Gamma(S-j+1)} \cdot \frac{\Gamma(S-j+B_S+\psi_{S_3})}{(e_S+r\cdot e_S+\psi_{S_4})^{S-j+B_S+\psi_{S_3}}} \cdot \frac{\Gamma(j+\psi_{S_1})}{(e_S+\psi_{S_2})^{j+\psi_{S_1}}}} \tag{A7}$$

Here the prior independence of λ_S and λ_H implies that the joint posterior distribution of these two intensities is decomposed into a product of their marginal posterior distributions, i.e.,

$$p(\lambda_S, \lambda_H | S, H, B_S, B_H) = p(\lambda_S | S, B_S) p(\lambda_H | H, B_H). \tag{A8}$$

This assumption of independence between λ_S and λ_H can be relaxed, and hierarchical modeling for sources to determine clustering properties and spectral parameters can be devised (in preparation).

Using the joint posterior distribution in Equation A8, we compute the analytical solution to the posterior distribution of each hardness ratio as follows:

1. the posterior distribution of $\mathcal{R}$ is obtained after integrating λ_H out of
$$\begin{aligned} & p(\mathcal{R}, \lambda_H | S, H, B_S, B_H)\, d\mathcal{R}\, d\lambda_H \\ = \; & p(\lambda_S, \lambda_H | S, H, B_S, B_H) \Big| \frac{\partial(\lambda_S, \lambda_H)}{\partial(\mathcal{R}, \lambda_H)} \Big|\, d\lambda_S\, d\lambda_H \\ = \; & p(\mathcal{R}\lambda_H, \lambda_H | S, H, B_S, B_H) \lambda_H\, d\mathcal{R}\, d\lambda_H, \end{aligned}$$

2. the posterior distribution of C is obtained after integrating λ_H out of
$$\begin{aligned} & p(\mathrm{C}, \lambda_H | S, H, B_S, B_H)\, d\mathrm{C}\, d\lambda_H \\ = \; & p(\lambda_S, \lambda_H | S, H, B_S, B_H) \Big| \frac{\partial(\lambda_S, \lambda_H)}{\partial(\mathrm{C}, \lambda_H)} \Big|\, d\lambda_S\, d\lambda_H \\ = \; & p(10^{\mathrm{C}}\lambda_H, \lambda_H | S, H, B_S, B_H) 10^{\mathrm{C}} \ln(10) \lambda_H\, d\mathrm{C}\, d\lambda_H, \text{ and} \end{aligned}$$

3. the posterior distribution of $\mathcal{HR}$ is obtained after integrating $\omega = \lambda_S + \lambda_H$ out of
$$\begin{aligned} & p(\mathcal{HR}, \omega | S, H, B_S, B_H)\, d\mathcal{HR}\, d\omega \\ = \; & p(\lambda_S, \lambda_H | S, H, B_S, B_H) \Big| \frac{\partial(\lambda_S, \lambda_H)}{\partial(\mathcal{HR}, \omega)} \Big|\, d\lambda_S\, d\lambda_H \\ = \; & p\Big(\frac{(1-\mathcal{HR})\omega}{2}, \frac{(1+\mathcal{HR})\omega}{2} \Big| S, H, B_S, B_H \Big) \frac{\omega}{2}\, d\mathcal{HR}\, d\omega. \end{aligned}$$

To numerically integrate out a parameter in these analytical solutions, we employ quadrature, which precisely evaluates the integral of a real valued function (and has nothing to do with the Gaussian counts statistic assumptions); refer to Wichura (1989) for details of the algorithm. Due to the complex expression, the probability density for each hardness ratio is approximated for each equally-spaced abscissas equally spaced over the finite range of the hardness ratio. Then we summarize the approximate posterior density in order to calculate the inferences, as detailed in §2.2.1.

B. Effect of Non-informative Priors

In order to compare the effect of these non-informative prior distributions, we have carried out simulations for low counts and high counts cases to test the computed coverage

Table 2: Legend Key for Table 3.

<table>
<tr><td></td><td colspan="2">Hard band source intensity,
λ_H [counts (source area)$^{-1}$]</td></tr>
<tr><td rowspan="3">Soft band source intensity
λ_S [counts (source area)$^{-1}$]</td><td>Coverage rate of
intervals with $\phi = 0^+$</td><td>Average length of
intervals with $\phi = 0^+$</td></tr>
<tr><td>Coverage rate of
intervals with $\phi = 1/2$</td><td>Average length of
intervals with $\phi = 1/2$</td></tr>
<tr><td>Coverage rate of
intervals with $\phi = 1$</td><td>Average length of
intervals with $\phi = 1$</td></tr>
</table>

against the simulation: Table 3 presents the coverage rate and mean length of the 95% intervals for the color, C. In order to simplify the comparison, we assume that there is no background contamination, and simulate 1000 realizations of source counts for each case. The tables are laid out as a grid of soft and hard source counts per source area (i.e., λ_S versus λ_H, see Equation A1). For each (λ_S, λ_H) pair, we report the percentage of the simulations that result in 95% posterior intervals that actually contain (λ_S, λ_H) along with the mean length of the 95% posterior intervals from the simulations. This calculation is carried out for three choices of the prior index, $\phi = 0^+, 1/2, 1$, corresponding to the top, middle, and bottom elements in each cell of the table.

In these simulations, we theoretically expect that the computed 95% posterior intervals contain the actual value with a probability of 95%. Due to the Monte Carlo simulation errors, we expect most coverage rates to be between 0.93 and 0.97 which are three standard deviations away from 0.95; a standard deviation of the coverage probability for 95% posterior intervals is given by $\sqrt{(0.95)(0.05)/1000} = 0.0069$ under a binomial model for the Monte Carlo simulation.

Table 3 presents the coverage rate and average length of posterior intervals for small and large magnitudes of source intensities. The key to this table is given in Table 2: For each (λ_S, λ_H) pair, the posterior intervals are simulated using different prior distribution indices ($\phi = 0^+$, $1/2$, 1) and the summary statistics – the coverage rate and the mean lengths of the intervals – are displayed from top ($\phi = 0^+$) to bottom ($\phi = 1$) within each cell. The same information is shown in graphical form in Figure 10. The use of $\phi = 0^+$ tends to yield very wide posterior intervals and under-cover the true X-ray color when the source intensities are low. On the other hand, the other two non-informative prior distribution indices produce much shorter posterior intervals and maintain high coverage rates. With high counts, however, the choice of a non-informative prior distribution does not have a noticeable effect on the statistical properties of the posterior intervals. The same results are summarized in graphical form in Figure 9, where the 95% coverage rates are shown for various cases. The empirical distributions of the coverage rate computed under either low count ($\lambda_S = 0.5, 1, 2, 4$ or $\lambda_H = 0.5, 1, 2, 4$) or high count (all of the remaining) scenarios are shown, along with a comparison with the results from the classical method (last column). The distinction between the Bayesian and classical methods is very clear in this figure. Note that the posterior intervals over-cover relative to our theoretical expectation, but to a lesser degree than the classical intervals. Over-coverage is conservative and preferred to under-coverage, but should be minimal when possible. Considering the low count and high count scenarios, Figure 9 suggests using the Jeffrey's non-informative prior distribution (i.e., $\phi = 1/2$) in general. At low counts, the shape (and consequently the coverage) of the posterior distribution is dominated by our prior assumptions, as encoded in the parameter ϕ. The coverage rate will also be improved when informative priors are used.

Table 3: Coverage of the X-ray Color (C) Using the Bayesian Method with Different (see Table 2 for key) Non-Informative Prior Distribution Indices ($\phi = 0^{+}, 1/2, 1$).

		λ_H															
		0.5		1.0		2.0		4.0		8.0		16.0		32.0		64.0	
λ_S	0.5	100 %	23.99	100 %	22.67	100 %	21.19	100 %	19.02	100 %	14.07	100 %	11.32	100 %	11.14	100 %	11.11
		100 %	5.10	100 %	4.88	99.9 %	4.51	99.9 %	3.95	100 %	3.35	99.9 %	3.09	99.7 %	3.02	100 %	3.03
		100 %	2.87	100 %	2.71	99.0 %	2.50	98.4 %	2.25	98.4 %	2.02	97.2 %	1.90	97.0 %	1.82	98.4 %	1.83
	1.0	100 %	22.95	100 %	21.59	100 %	19.61	100 %	17.40	100 %	12.24	100 %	10.03	100 %	9.84	100 %	9.51
		100 %	4.91	100 %	4.69	99.7 %	4.31	100 %	3.77	99.9 %	3.04	99.8 %	2.80	99.7 %	2.74	99.8 %	2.71
		100 %	2.73	100 %	2.58	99.2 %	2.37	98.9 %	2.14	98.5 %	1.85	97.3 %	1.73	97.7 %	1.67	98.0 %	1.65
	2.0	100 %	21.42	100 %	19.57	99.8 %	17.52	99.7 %	15.04	99.6 %	10.54	99.3 %	8.45	99.9 %	7.95	99.8 %	7.82
		99.9 %	4.52	99.9 %	4.30	99.7 %	3.89	99.6 %	3.27	99.5 %	2.56	99.2 %	2.33	99.0 %	2.21	99.4 %	2.21
		99.1 %	2.51	99.5 %	2.37	99.7 %	2.14	98.8 %	1.87	97.8 %	1.60	98.1 %	1.49	97.3 %	1.40	97.7 %	1.39
	4.0	100 %	18.60	100 %	16.80	99.9 %	15.18	99.2 %	13.11	97.9 %	8.87	97.9 %	6.06	97.0 %	5.85	96.8 %	5.63
		100 %	3.92	100 %	3.68	99.6 %	3.30	99.2 %	2.64	98.5 %	1.89	98.0 %	1.60	97.9 %	1.53	96.6 %	1.49
		98.4 %	2.24	99.2 %	2.09	99.0 %	1.87	99.1 %	1.59	97.7 %	1.29	97.3 %	1.14	97.1 %	1.08	97.4 %	1.04
	8.0	100 %	13.66	100 %	12.28	99.3 %	10.93	98.3 %	8.74	99.2 %	5.13	98.2 %	2.50	97.8 %	2.26	97.4 %	2.14
		99.9 %	3.29	100 %	3.04	99.7 %	2.64	97.8 %	1.86	98.8 %	1.25	97.6 %	0.96	97.2 %	0.86	97.4 %	0.83
		97.7 %	1.99	98.7 %	1.85	98.7 %	1.63	97.5 %	1.28	98.1 %	1.01	97.3 %	0.83	96.3 %	0.75	97.1 %	0.71
	16.0	100 %	11.37	100 %	9.96	99.8 %	8.49	97.7 %	6.11	98.6 %	2.41	97.9 %	0.73	96.7 %	0.60	95.5 %	0.54
		99.9 %	3.10	99.8 %	2.79	99.2 %	2.34	97.8 %	1.60	98.0 %	0.95	96.6 %	0.65	95.6 %	0.55	94.4 %	0.50
		97.5 %	1.91	98.0 %	1.73	98.2 %	1.49	97.2 %	1.14	96.7 %	0.83	96.6 %	0.63	95.3 %	0.54	93.7 %	0.49
	32.0	100 %	10.82	100 %	9.56	99.3 %	8.00	97.7 %	5.74	97.3 %	2.33	95.7 %	0.59	97.1 %	0.43	93.8 %	0.37
		99.9 %	2.98	99.6 %	2.71	99.0 %	2.21	97.8 %	1.51	96.5 %	0.88	94.4 %	0.55	97.2 %	0.43	94.3 %	0.36
		98.2 %	1.80	97.4 %	1.65	96.0 %	1.40	97.1 %	1.06	96.0 %	0.75	94.5 %	0.54	97.2 %	0.42	94.5 %	0.36
	64.0	100 %	10.99	100 %	9.52	99.7 %	7.84	97.7 %	5.69	97.7 %	2.11	96.3 %	0.53	94.6 %	0.36	96.5 %	0.29
		99.9 %	3.02	99.6 %	2.72	99.0 %	2.23	98.1 %	1.52	97.7 %	0.83	94.8 %	0.50	94.8 %	0.36	96.5 %	0.29
		98.5 %	1.82	97.4 %	1.66	97.1 %	1.40	97.8 %	1.05	96.6 %	0.71	94.2 %	0.49	94.9 %	0.36	96.5 %	0.28

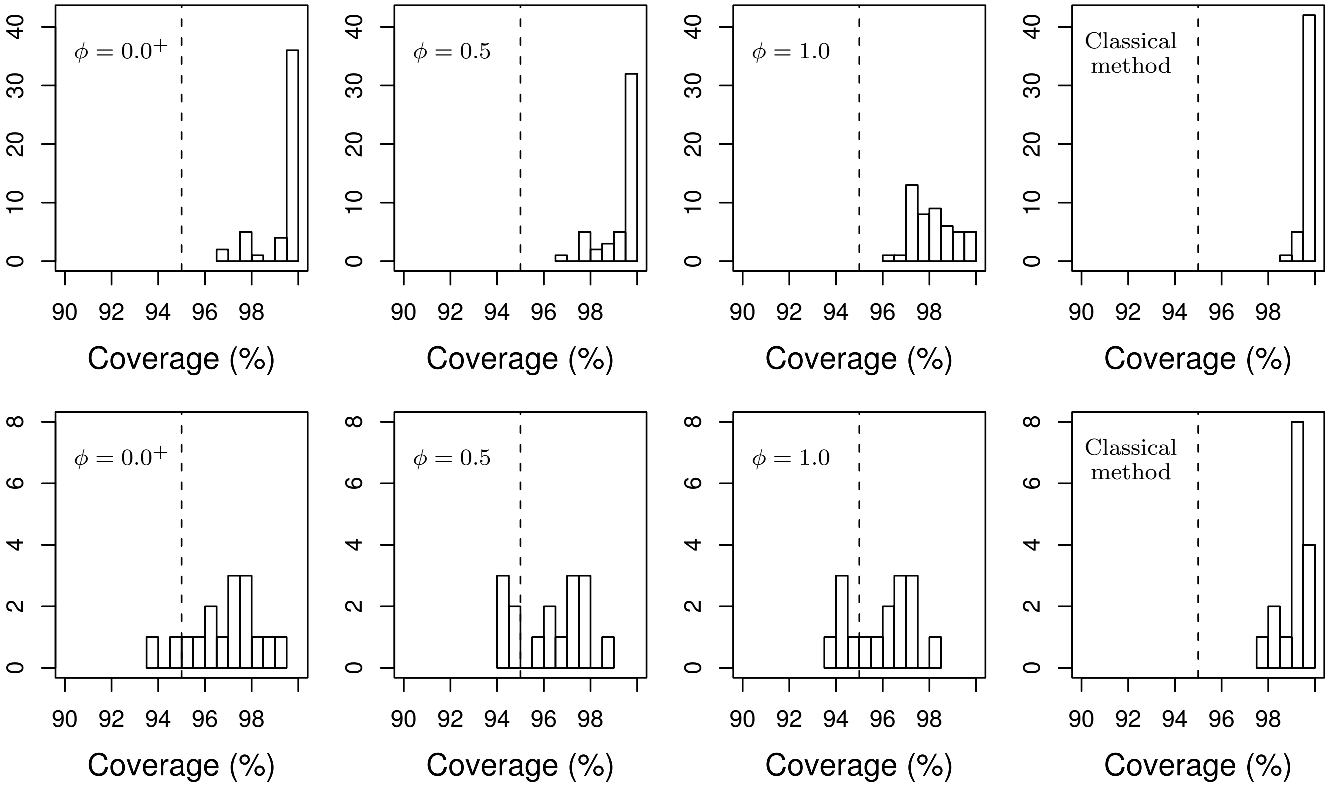

Fig. 9.— The empirical distributions of coverage rates with different prior distribution indices. The first three columns of the panels are constructed under different types of non-informative prior distributions, $\phi = 0^+$ (first), $\phi = 1/2$ (second), $\phi = 1$ (third). The fourth column depicts results from the classical method, and is presented for reference. (In calculating the coverage and the length of the intervals, We exclude simulations where $S = 0$ or $H = 0$ since C is undefined in such cases.) The top panels correspond to the low count scenarios $\lambda_S = (0.5, 1, 2, 4)$ or $\lambda_H = (0.5, 1, 2, 4)$ and the bottom panels correspond to the remaining cases. Coverage rate is the fraction of simulated intervals that contain the parameter under which the data were simulated. In this case, we compute the coverage rate based on the 95% posterior intervals for C with different combinations of values for λ_S and λ_H. The histograms represent the empirical distribution of the coverage rate. The dotted vertical lines represent the expected coverage rate of 95%.

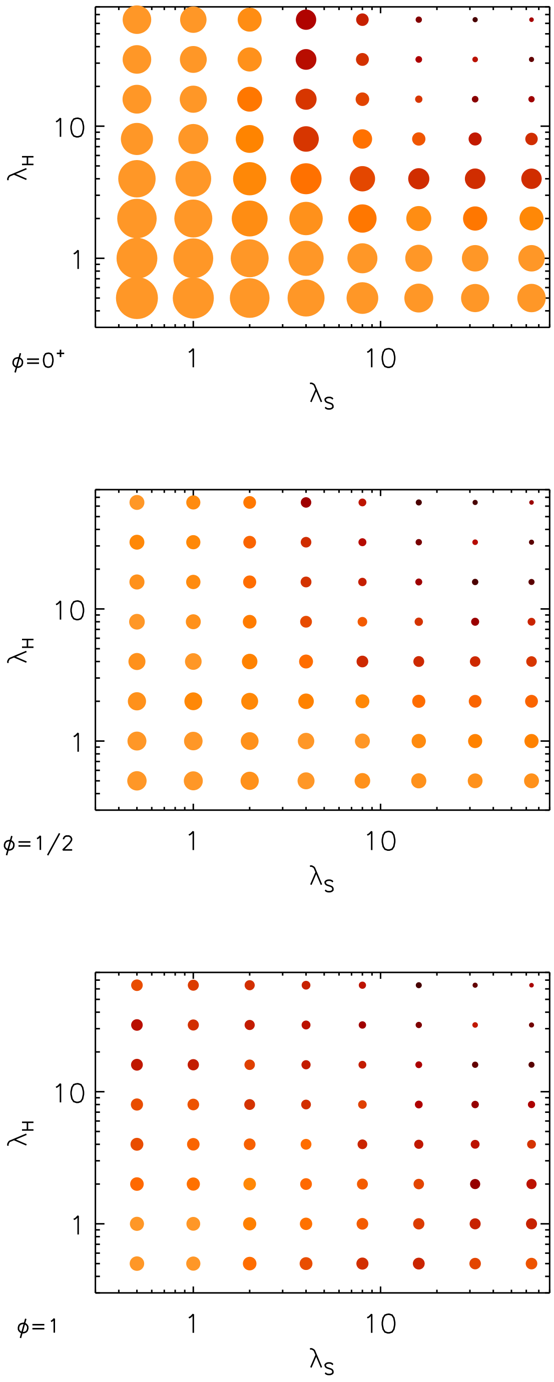

Fig. 10.— Graphical representation of the data in Table 3. The symbols are located on a grid corresponding to the appropriate λ_S (absissa) and λ_H (ordinate), for $\phi = 0^+$ (top), $\phi = 1/2$ (middle), and $\phi = 1$ (bottom). The shading of each symbol represents the coverage, with 100% being lightest and progressively getting darker as the coverage percentage decreases. The sizes of the symbols are scaled as the $\sqrt{\text{length of interval}}$. Note that for small values of ϕ, which correspond to a prior expectation of a large dynamic range in the source colors, the intervals are large when the counts are small (i.e., the choice of prior distribution has a large effect), and decrease to values similar to those at larger ϕ when there are more counts.

REFERENCES

Alexander, D.M., Brandt, W.N., Hornschemeier, A.E., Garmire, G.P., Schneider, D.P., Bauer, F.E., & Griffiths, R.E., 2001, AJ, 122, 2156

Babu, G.J., & Feigelson, E.D., 1997, *Statistical Challenges in Modern Astronomy*, Springer-Verlag:New York

Bradt, H., Mayer, W., Buff, J., Clark, G.W., Doxsey, R., Hearn, D., Jernigan, G., Joss, P.C., Laufer, B., Lewin, W., Li, F., Matilsky, T., McClintock, J., Primini, F., Rappaport, S., & Schnopper, H., 1976, ApJ, 204, L67

Brandt, W.N., Hornschemeier, A.E., Alexander, D.M., Garmire, G.P., Schneider, D.P., Broos, P.S., Townsley, L.K., Bautz, M.W., Feigelson, E.D., & Griffiths, R.E., 2001a, AJ, 122, 1

Brandt, W.N., Alexander, D.M., Hornschemeier, A.E., Garmire, G.P., Schneider, D.P., Barger, A.J., Bauer, F.E., Broos, P.S., Cowie, L.L., Townsley, L.K., Burrows, D.N., Chartas, G., Feigelson, E.D., Griffiths, R.E., Nousek, J.A., & Sargent, W.L.W., 2001b, AJ, 122, 2810

Brandt, W.N., & Hasinger, G., 2005, ARAA, 43, 827

Brown, E.F., Bildsten, L., & Rutledge, R.E., 1998, ApJ, 504, 95

Campana, S., Colpi, M., Mereghetti, S., Stella, L., & Tavani, M., 1998, A&ARv, 8, 279

Casella, G. & Berger, R.L., 2002, Statistical Inference, 2nd edition. Duxbury.

Cowles, M.K., & Carlin, B.P., 1996, J. Am. Stat. Assn., 91, 883

Esch, D.N., 2003, Ph.D. Thesis, Department of Statistics, Harvard University

Feigelson, E.D., & Babu, G.J., 1992, *Statistical Challenges in Modern Astronomy*, Springer-Verlag:Berlin Heidelberg New York

Feigelson, E.D., & Babu, G.J., 2003, *Statistical challenges in astronomy. Third Statistical Challenges in Modern Astronomy (SCMA III) Conference*, University Park, PA, USA, July 18 - 21 2001, Springer:New York

Gehrels, N., 1986, ApJ, 303, 336

Geman, S., & Geman, D., 1984, IEEE, 6, 721

Giacconi, R., Rosati, P., Tozzi, P., Nonino, M., Hasinger, G., Norman, C., Bergeron, J., Borgani, S., Gilli, R., Gilmozzi, R., & Zheng, W., 2001, ApJ, 551, 624

Gregory, P.C., & Loredo, T.J., 1992, ApJ, 398, 146

Hasinger, G., & van der Klis, M., 1989, A&A, 225, 79

Heinke, C.O., Grindlay, J.E., Edmonds, P.D., Lloyd, D.A., Murray, S.S., Cohn, H.N., & Lugger, P.M., 2003, ApJ, 598, 501

Hong, J.S., Schlegel, E.M., & Grindlay, J.E., 2004, ApJ, 614, 508

Gillessen, S., & Harney, H.L., 2005, A&A, 430, 355

Jansen, F., Lumb, D., Altieri, B., Clavel, J., Ehle, M., Erd, C., Gabriel, C., Guainazzi, M., Gondoin, P., Much, R., Munoz, R., Santos, M., Schartel, N., Texier, D., & Vacanti, G., 2001, A&A, 365, L1

Kashyap, V.L., Drake, J.J., Güdel, M., & Audard, M., 2002, ApJ, 580, 1118

Kashyap, V.L., & Drake, J.J., 1998, ApJ, 503, 450

Kim, D.-W., Barkhouse, W.A., Colmenero, E.R., Green, P.J., Kim, M., Mossman, A., Schlegel, E., Silverman, J.D., Aldcroft, T., Ivezic, Z., Anderson, C., Kashyap. V., Tananbaum, H., & Wilkes, B.J., 2005, submitted to ApJ

Kim, D.-W., Fabbiano, G., & Trinchieri, G., 1992, ApJ, 393, 134

Pallavicini, R., Tagliaferri, G., & Stella, L., 1990, A&A, 228, 403

Protassov, R., van Dyk, D.A., Connors, A., Kashyap, V.L., & Siemiginowska, A., 2002, ApJ, 571, 545

Reale, F., Betta, R., Peres, G., Serio, S., & McTiernan, J., 1997, A&A, 325, 782

Reale, F., Güdel, M., Peres, G., & Audard, M., 2004, A&A, 416, 733

Reale, F., Serio, S., & Peres, G., 1993, A&A, 272, 486

Rosner, R., Golub, L., & Vaiana, G.S., 1985, ARAA, 23, 413

Schmitt, J.H.M.M., & Favata, F., 1999, Nature, 401, 44

Serio, S., Reale, F., Jakimiec, J., Sylwester, B., & Sylwester, J. 1991, A&A, 241, 197

Silverman, J.D., Green, P.J., Barkhouse, W.A., Kim, D.-W., Aldcroft, T.L., Cameron, R.A., Wilkes, B.J., Mossman, A., Ghosh, H., Tananbaum, H., Smith, M.G., Smith, R.C., Smith, P.S., Foltz, C., Wik, D., & Jannuzi, B.T., 2005, ApJ, 618, 123

Tanner, M.A. & Wong, W.H., 1987, J. Am. Stat. Assn., 82, 528

Tuohy, I.R., Garmire, G.P., Lamb, F.K., & Mason, K.O., 1978, ApJ, 226, L17

Vaiana, G.S., Cassinelli, J.P., Fabbiano, G., Giacconi, R., Golub, L., Gorenstein, P., Haisch, B.M., Harnden, F.R.,Jr., Johnson, H.M., Linsky, J.L., Maxson, C.W., Mewe, R., Rosner, R., Seward, F., Topka, K., & Zwaan, C., 1981, ApJ, 245, 163

van den Oord, G.H.J., & Mewe, R., 1989, A&A, 213 van Dyk, D.A., 2003, in *Statistical challenges in astronomy. Third Statistical Challenges in Modern Astronomy (SCMA III) Conference*, University Park, PA, USA, July 2001, Eds. E.D.Feigelson, G.J.Babu, New York: Springer, p41

van Dyk, D.A., Connors, A., Kashyap, V.L., & Siemiginowska, A., 2001, ApJ 548, 224

van Dyk, D.A., & Hans, C.M., 2002, in Spatial Cluster Modelling, Eds. D.Denison & A.Lawson, CRC Press: London, p175

van Dyk, D.A., & Kang, H., 2004, Stat. Sci, 19, 275

Wargelin, B.J., Kashyap, V.L., Drake, J.J., García-Alvarez, D., & Ratzlaff, P.W., 2006, submitted to ApJ

Weisskopf, M.C., Tananbaum, H.D., Van Speybroeck, L.P., & O'Dell, S.L., 2000, Proc. SPIE, 4012, 2

Wichura, M.J., 1989, Technical Report No. 257, Department of Statistics, The University of Chicago

Zamorani, G., Henry, J.P., Maccacaro, T., Tananbaum, H., Soltan, A., Avni, Y., Liebert, J., Stocke, J., Strittmatter, P.A., Weymann, R.J., Smith, M.G., & Condon, J.J., 1981, ApJ, 245, 357

This preprint was prepared with the AAS LaTeX macros v5.0.