BEHR (Bayesian Estimation of Hardness Ratios): Computing Hardness Ratios with Poissonian Errors

David A. van Dyk (dvd@ics.uci.edu, Department of Statistics, University of California at Irvine),
Taeyoung Park (tpark@stat.harvard.edu, Department of Statistics, Harvard University),
Vinay L. Kashyap, and Andreas Zezas (Harvard-Smithsonian Center for Astrophysics).

CHASC (California-Harvard AstroStatistics Collaboration)
www.ics.uci.edu/~dvd/astrostat.html

Summary

- Hardness ratios are commonly used to characterize the spectrum of an X-ray source when spectral fitting is not possible.
- The classical method is based on the net number of counts and fails to account for the asymmetric nature of the Poisson counts. This is a problem with low counts, especially when the counts are zero or cannot statistically be distinguished from zero.
- The errors bars associated with the classical method are based on Gaussian assumptions and do not provide realistic confidence limits.
- In this poster, we present a statistically coherent scheme for computing hardness ratios and their associated errors.
- In this scheme, we model the detected photons as independent Poisson variables and calculate hardness ratios using a sophisticated Bayesian approach.
- Finally, we present a simulation study comparing a new Bayesian method with the classical method, which demonstrates the new method provide more reliable results especially for low count data.
- **BEHR (Bayesian Estimation of Hardness Ratios)** that uses the new Bayesian method is free statistical software and will soon be available on the CIAO contributed software page.

The Classical Method

A Hardness Ratio:

- Given observed counts in the soft band (S) and the hard band (H), a hardness ratio can be computed as a summary of a spectrum:
  1. Simple counts ratio, $\mathrm{R} = \dfrac{S}{H}$,
  2. X-ray color, $\mathrm{C} = \log_{10}\dfrac{S}{H}$, and
  3. Fractional difference hardness ratio, $\mathrm{HR} = \dfrac{H-S}{H+S}$.
- In the presence of background where B_S and B_H are collected in an area of c times the source region, the above is generalized to
  1. $\mathrm{R} = \dfrac{S - B_S/c}{H - B_H/c}$,
  2. $\mathrm{C} = \log_{10}\left(\dfrac{S - B_S/c}{H - B_H/c}\right)$, and
  3. $\mathrm{HR} = \dfrac{(H - B_H/c) - (S - B_S/c)}{(H - B_H/c) + (S - B_S/c)}$,

  and their errors are computed **under Gaussian assumptions**:
  1. $\sigma_{\mathrm{R}} = \frac{S}{H}\sqrt{\frac{\sigma_S^2+\sigma_{B_S}^2/c^2}{(S-B_S/c)^2} + \frac{\sigma_H^2+\sigma_{B_H}^2/c^2}{(H-B_H/c)^2}}$,
  2. $\sigma_{\mathrm{C}} = \frac{1}{\ln(10)}\sqrt{\frac{\sigma_S^2+\sigma_{B_S}^2/c^2}{(S-B_S/c)^2} + \frac{\sigma_H^2+\sigma_{B_H}^2/c^2}{(H-B_H/c)^2}}$, and
  3. $\sigma_{\mathrm{HR}} = \frac{2\sqrt{(H-B_H/c)^2\left(\sigma_S^2+\sigma_{B_S}^2/c^2\right)+(S-B_S/c)^2\left(\sigma_H^2+\sigma_{B_H}^2/c^2\right)}}{\left[(H-B_H/c)+(S-B_S/c)\right]^2}$,

  where each σ is approximated, e.g., $\sigma_S \approx \sqrt{S+0.75}+1$.

Modeling the Hardness Ratios

- The typical Gaussian assumptions are inappropriate for low counts.
- Instead, we directly model photons from a source (η) and photons from background (β) as **independent Poisson variables**:

– $S = \eta_S + \beta_S \sim \mathrm{Poisson}(\lambda_S + \xi_S)$, $H = \eta_H + \beta_H \sim \mathrm{Poisson}(\lambda_H + \xi_H)$,

– $B_S \sim \mathrm{Poisson}(c\,\xi_S)$, and $B_H \sim \mathrm{Poisson}(c\,\xi_H)$.

where λ and ξ denote the expected source and background counts in the source region.

- Given the expected source counts, the hardness ratio is rewritten as:
  1. Simple counts ratio, $\mathrm{R} = \dfrac{\lambda_S}{\lambda_H}$,
  2. X-ray color, $\mathrm{C} = \log_{10}\dfrac{\lambda_S}{\lambda_H}$, and
  3. Fractional difference hardness ratio, $\mathrm{HR} = \dfrac{\lambda_H - \lambda_S}{\lambda_H + \lambda_S}$.

New Bayesian Method

Bayesian Approach:

- Bayesian inferences for a parameter are based on a *posterior distribution* [e.g., $p(\lambda_S, \xi_S | S, B_S)$] which combines a *prior distribution* [e.g., $p(\lambda_S, \xi_S)$] with the likelihood [e.g., $p(S, B_S | \lambda_S, \xi_S)$] via **Bayes' theorem**,

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

Computing Posterior Distributions of Hardness Ratios:

- The posterior distribution of a hardness ratio is computed from the joint posterior distributions of λ_S and λ_H:
  1. the posterior distribution of R is computed from

  $$\begin{aligned} & p(\mathrm{R}, \lambda_H | S, H, B_S, B_H)\, d\mathrm{R}\, d\lambda_H \\ &= p(\lambda_S, \lambda_H | S, H, B_S, B_H) \left| \frac{\partial(\lambda_S, \lambda_H)}{\partial(\mathrm{R}, \lambda_H)} \right| d\lambda_S\, d\lambda_H \\ &= p(\mathrm{R}\lambda_H, \lambda_H | S, H, B_S, B_H)\lambda_H\, d\mathrm{R}\, d\lambda_H, \end{aligned}$$

  where we integrate out λ_H;

  2. the posterior distribution of C is computed from

  $$\begin{aligned} & p(\mathrm{C}, \lambda_H | S, H, B_S, B_H)\, d\mathrm{C}\, d\lambda_H \\ &= p(\lambda_S, \lambda_H | S, H, B_S, B_H) \left| \frac{\partial(\lambda_S, \lambda_H)}{\partial(\mathrm{C}, \lambda_H)} \right| d\lambda_S\, d\lambda_H \\ &= p(10^{\mathrm{C}}\lambda_H, \lambda_H | S, H, B_S, B_H) 10^{\mathrm{C}} \ln(10) \lambda_H\, d\mathrm{C}\, d\lambda_H, \end{aligned}$$

  where we integrate out λ_H; and

  3. the posterior distribution of HR is computed from

  $$\begin{aligned} & p(\mathrm{HR}, \omega | S, H, B_S, B_H)\, d\mathrm{HR}\, d\omega \\ &= p(\lambda_S, \lambda_H | S, H, B_S, B_H) \left| \frac{\partial(\lambda_S, \lambda_H)}{\partial(\mathrm{HR}, \omega)} \right| d\lambda_S\, d\lambda_H \\ &= p\left(\frac{(1-\mathrm{HR})\omega}{2}, \frac{(1+\mathrm{HR})\omega}{2} \middle| S, H, B_S, B_H \right) \frac{\omega}{2}\, d\mathrm{HR}\, d\omega, \end{aligned}$$

  where we integrate out $\omega = \lambda_S + \lambda_H$.

- The Bayes' theorem analytically computes a high dimensional joint posterior distribution of all unknown quantities.
- To integrate out everything but λ_S and λ_H of the joint posterior distribution, we use either **Monte Carlo integration** or **efficient numerical integration**.
- We use both methods of integration because neither has the advantage over the other in our case.

Simulation Study

Simulated Data Sets:

- To compare our Bayesian method with the classical method, we simulate 100 data sets of S, H, B_S, and B_H for each of 100 different magnitudes of the expected source counts, λ_S and λ_H, but with the same expected background counts $\xi_S = \xi_H = 10$, the constant background area ratio $c = 100$, and the constant effective area of 1.
- We let λ_S range from 1 to 100 and $\lambda_H = \lambda_S/\mathrm{R}$ is determined by the fixed value of R.

Simulation Results:

- Figure 1 presents the estimates of hardness ratios according to total expected source counts ($\lambda_S + \lambda_H$): the **blue dots** represent the posterior modes of hardness ratios; the **red dots** represent estimates of hardness ratios based on the classical method; and the **green dotted lines** represent fixed values of hardness ratios based on which we simulate the data sets.
- The **purple dots** in the classical method indicate estimates of R that result in negative values: In the case of R and C, these estimates are reflected at zero; in the case of HR, the estimates below −1 are reflected at −1 and the estimates above 1 are reflected at 1.
- The classical method provides unreliable estimates especially for low count data, as compared to the Bayesian method.

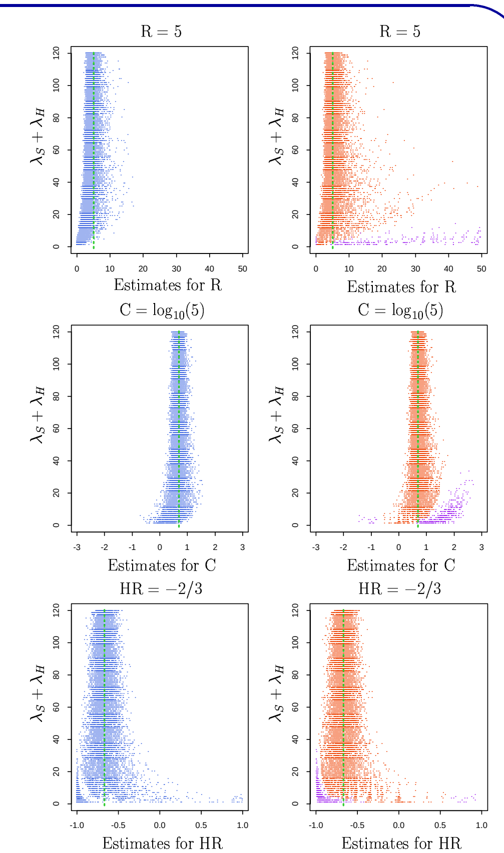

Figure 1: Simulation Study. We notice that the classical method does not provide reliable estimates for low count data, while it agrees with the Bayesian method for large count data.

A Prior Distribution

An Informative Prior Distribution:

- If there is a strong belief as to the hardness ratio (location or spread), we can incorporate the information as a prior distribution, which is called an informative prior distribution.
- The Bayesian method produces the posterior distribution, which can be used as an informative prior distribution for future observation of the same source.

A Flat Prior Distribution:

- With no prior information available, we normally use a so-called flat prior distribution. Since the Poisson intensity takes positive real values, two sorts of a flat prior are considered:
  – $p(\lambda) \propto 1$ that corresponds to $\psi = 1$ when $\lambda^{\psi-1} \propto 1$;
  – $p(\log_{10}\lambda) \propto 1$ that corresponds to $\psi = 0$ when $\lambda^{\psi-1} \propto 1$.
- With large count data, which flat prior distribution to use does not make much difference in the posterior distribution.
- When the expected counts are low, we choose the value of ψ using a simulation study. We aim to ensure that the resulting 95% intervals contain the true value at least 95% of the time.

Acknowledgments

We gratefully acknowledge funding for this 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). This work is a product of joint work with the California-Harvard astrostatistics collaboration (CHASC) whose members include J. Chiang, A. Connors, D. van Dyk, D. Esch, P. Freeman, H. Kang, V. L. Kashyap, X.-L. Meng, A. Siemiginowska, E. Sourlas, T. Park, Y. Yu, and A. Zezas.