Two New Statistical Tools for

Bayesian Analysis of Low-Count Spectra

David A. van Dyk

Department of Statistics
University of California, Irvine

Joint work with
The California-Harvard Astrostatistics Collaboration (CHASC)

Presented at HEAD 2004
Session 31: Astrostatistics Workshop
Friday, September 10, 2004, 6:30-8:30pm
Queen Anne Ballrom

Low-Count Spectra

High-Resolution Spectra

- High resolution detectors such as those aboard *Chandra* herald a quantum leap forward for empirical high-energy astronomy
- Unfortunately, standard methods (e.g., χ^2 fitting) rely on Gaussian assumptions and thus require a minimum count per bin.
- Ad-hoc procedures that group bins are wasteful and sacrifice the desirable high-resolution inherent in the data.

Hardness-Ratios

- A rough summary of a spectrum is a comparison of the expected hard and expected soft counts.
- This is the lowest resolution spectral analysis, but can be useful for classifying faint sources.
- Again, the validity of standard methods (e.g., method of moments solutions) depends upon Gaussian assumptions.
- For faint sources either the hard or soft counts can be very small.

Solution: Poisson Statistics

- Rather than basing statistical techniques on Gaussian assumptions, we can use the Poisson Distribution as a statistical model for low-count data.
- Specifically, we replace the Gaussian likelihood with a Poisson likelihood:

$$\text{Gaussian Likelihood:} \qquad -\sum_{\text{bins}} \sigma_i - \sum_{\text{bins}} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$

$$\text{Poisson Likelihood:} \qquad -\sum_{\text{bins}} \mu_i + \sum_{\text{bins}} x_i \log \mu_i$$

- Bayesian Methods combine the likelihood with a prior distribution that can
  - Model the dist'n of spectral characteristics in a population of sources.
  - Include information from outside the data as to the spectral shape.
  - Smooth the reconstructed spectrum.

Requires Sophisticated Statistical Computing.

BLoCXS

***B**ayesian fitting of **Lo**w **C**ount **X**-ray **S**pectra.*

BLoCXS Functionality

- Uses Poisson models and no Gaussian assumptions. Thus, BLoCXS has no trouble with low count data.
- Corrects for instrument response as quantified by `.rmf` or `.rsp` files.
- Corrects for effective area using `.arf` files.
- Uses a Poisson model-based strategy to correct for background contamination. There is no background subtraction and no negative counts.
- Can fit absorption due to the ISM or IGM.
- Allows for (broken) powerlaw, bremsstrahlung, and blackbody continuums.
- Can include Gaussian, Lorentzian, and delta function emission lines.
- Can compute principled p-values to test for emission lines.
- An extension that will allow for pile-up correction is under development. (**HEAD 2004 Poster 16.35**, Y. Yu.)

BLoCXS Documentation

BLoCXS Availability

- Scheduled for release in the next version of CIAO (`cxc.harvard.edu/ciao`).

BLoCXS Examples and References

van Dyk, D. A., Connors, A., Kashyap, V. L., & Siemiginowska, A. (2001). Analysis of Energy Spectrum with Low Photon Counts, *The Astrophysical Journal*, vol. 548, 224–243.

Protassov, R., van Dyk, D. A., Connors, A., Kashyap, V. L., & Siemiginowska, A. (2002). Statistics: Handle with Care, Detecting Multiple Model Components with the Likelihood Ratio Test, *The Astrophysical Journal*, vol. 571, 545–559.

van Dyk, D. A. & Kang, H. (2004). Highly Structured Hierarchical Models for Spectral Analysis in High Energy Astrophysics. *Statistical Science*, to appear.

Park, T., van Dyk, D. A., & Siemiginowska, A. (2004). Fitting Narrow Emission Lines in X-ray Spectra: Computation and Methods. *CHASC Technical Report*.

HEAD 2004 Poster 16.33 (T. Park).

[Related Topics **HEAD 2004 Posters 5.01** (H. Kang) and **16.32** (E. Sourlas).]

BEHR

B*ayesian* **E***stimation of* **H***ardness* **R***atios.*

BEHR Functionality

- BEHR uses Poisson models background contaminated soft and hard counts. Thus, BEHR has no trouble with low count data.
- BHER computes hardness ratio estimates and intervals with reliable frequency properties. (See simulation study.)

BEHR Availability

- BEHR will soon be available on the CXC contributed software page (`cxc.harvard.edu/cont-soft/soft-exchange.html`).

BEHR Examples and References

van Dyk, D. A. et al. (2004). Deconvolution in High-Energy Astrophysics: Science, Instrumentation, and Methods. *Bayesian Analysis*, to appear.

Park, T., van Dyk, D. A., Kashyap, V. L., & Zezas, A. (2004). Computing Hardness Ratios with Poissonian Errors. *CHASC Technical Report.*

HEAD 2004 Poster 16.27 (T. Park).

Verifying BEHR

Simulation Study

- $S = H = 3$; each with expected background contamination = 0.1.
- Background exposure is 100 times longer.
- $R = S/H, \qquad HR = (H - S)/(H + S), \qquad C = \log_{10}(R)$

Table 1: Coverage of Bayesian and Standard Methods.

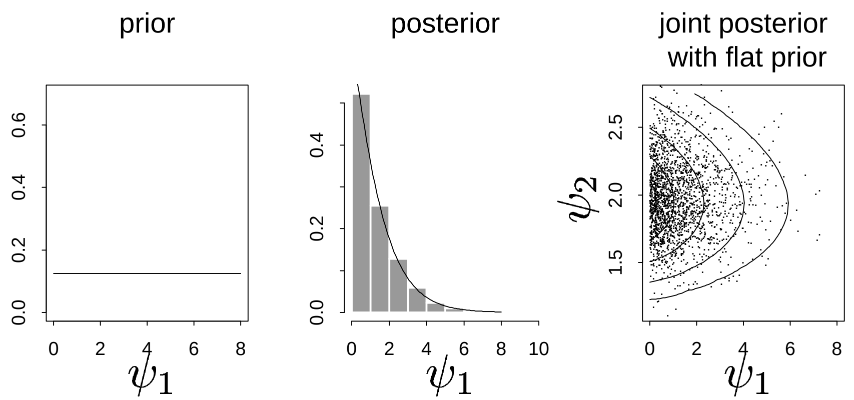
Method	Hardness Ratio	True Value	Coverage Rate	Mean Length	Mean Square Error by mode	Mean Square Error by mean
BEHR	R	1	95.0%	7.30	0.59	12.34
	HR	0	91.5%	1.23	0.53	0.42
	C	0	98.0%	1.53	0.42	0.46
Standard Method	R	1	96.5%	138.29	73.58	
	HR	0	99.5%	3.44	0.63	
	C	0	100.0%	7.26	5.58	

Bayesian Inference Using Monte Carlo

The Building Block of Bayesian Analysis

1. The sampling distribution: $p(Y|\psi)$.
2. The prior distribution: $p(\psi)$.
3. Bayes theorem and the posterior distribution: $p(\psi|Y) \propto p(Y|\psi)p(\psi)$

Inference Using a Monte Carlo Sample:



Obtaining a Monte Carlo sample using a Markov chain: MCMC

- Starting values matter, and, convergence diagnostics are critical.
- See van Dyk, Connors, Kashyap, & Siemiginowska (2001). ApJ, 548, 224.

CHASC

C*alifornia* **H***arvard* **A***stro*-**S***tatistics* **C***ollaboration.*

CHASC

The California-Harvard Astrostatistics Collaboration provides a forum for discussion of outstanding statistical problems in Astrophysics. Goals include developing papers on the interface of statistics and astronomy and incorporating state-of-the-art statistical methods into the Chandra Interactive Analysis of Observations software. Participants include faculty, researchers, graduate and undergraduate students.

The CHASC Homepage

`www.ics.uci.edu/~dvd/astrostat.html`

Participants

James Chiang, Alanna Connors, David Esch, Peter Freeman, Hosung Kang, Vinay L. Kashyap, Xiao-Li Meng, Taeyoung Park, Aneta Siemiginowska, Nondas Sourlas, David van Dyk, Yaming Yu, and Andreas Zezas,