Two Statistical Problems in X-ray Astronomy

Alexander W Blocker

October 21, 2008

Outline

Introduction

- Recent projects have focused on two areas:
  - Analysis of faint (low-count) x-ray data with Bayesian models
  - Analysis of events in time series
- Each has presented a unique set of challenges

General analysis of faint x-ray sources

General analysis of faint x-ray sources

- In multiwavelength x-ray studies, astronomers identify potential sources using catalogs in one waveband (typically optical or infrared) and observe the selected sources in x-rays.

General analysis of faint x-ray sources

- In multiwavelength x-ray studies, astronomers identify potential sources using catalogs in one waveband (typically optical or infrared) and observe the selected sources in x-rays.
- This frequently leads to a sample containing many faint, undetected sources.

General analysis of faint x-ray sources

- In multiwavelength x-ray studies, astronomers identify potential sources using catalogs in one waveband (typically optical or infrared) and observe the selected sources in x-rays.
- This frequently leads to a sample containing many faint, undetected sources.
- We want to combine information from these undetected sources to make inferences about our selected sample.

Current method: stacking

- Based on background subtraction
- For source i, observe $c_{s,i}$ counts in source aperture and $c_{b,i}$ counts in background aperture.
- Calculate net counts as $c_{n,i} = c_{s,i} - \frac{A_{s,i}}{A_{b,i}} c_{b,i}$, where $A_{s,i}$ and $A_{b,i}$ are the effective areas for the source and background regions (taking into account exposures), respectively.

Current method: stacking

- Based on background subtraction
- For source i, observe $c_{s,i}$ counts in source aperture and $c_{b,i}$ counts in background aperture.
- Calculate net counts as $c_{n,i} = c_{s,i} - \frac{A_{s,i}}{A_{b,i}} c_{b,i}$, where $A_{s,i}$ and $A_{b,i}$ are the effective areas for the source and background regions (taking into account exposures), respectively.
- Calculate stacked flux as $\bar{f}_x = \frac{\overline{\mathrm{ECF}}}{\sum_i A_{s,i}} \sum_i c_{n,i}$, where $\overline{\mathrm{ECF}}$ is the mean energy conversion factor.

Current method: stacking

- Based on background subtraction
- For source i, observe $c_{s,i}$ counts in source aperture and $c_{b,i}$ counts in background aperture.
- Calculate net counts as $c_{n,i} = c_{s,i} - \frac{A_{s,i}}{A_{b,i}} c_{b,i}$, where $A_{s,i}$ and $A_{b,i}$ are the effective areas for the source and background regions (taking into account exposures), respectively.
- Calculate stacked flux as $\bar{f}_x = \frac{\overline{\mathrm{ECF}}}{\sum_i A_{s,i}} \sum_i c_{n,i}$, where $\overline{\mathrm{ECF}}$ is the mean energy conversion factor.
- Calculate stacked luminosity as $\bar{L}_x = \frac{1}{N} \sum_i \mathrm{LCF}_i c_{n,i}$, where LCF_i is the luminosity conversion factor for source i
  - $\mathrm{LCF}_i = \frac{4\pi d_{\ell,i}^2 \mathrm{ECF}_i \times A_{corr,i} \times K_{corr,i}}{A_{s,i}}$

Problems with conventional stacking

Problems with conventional stacking

- Use of background subtraction $\Rightarrow$ Gaussian assumption; clearly inappropriate here.
- Above manifests as negative net counts; for sufficiently faint samples, can lead to negative stacked fluxes and luminosities.

Problems with conventional stacking

- Use of background subtraction $\Rightarrow$ Gaussian assumption; clearly inappropriate here.
- Above manifests as negative net counts; for sufficiently faint samples, can lead to negative stacked fluxes and luminosities.
- No clean measure of uncertainties on luminosities.

Problems with conventional stacking

- Use of background subtraction $\Rightarrow$ Gaussian assumption; clearly inappropriate here.
- Above manifests as negative net counts; for sufficiently faint samples, can lead to negative stacked fluxes and luminosities.
- No clean measure of uncertainties on luminosities.
- Solution: model data as Poisson

A hierarchical Bayesian model for "stacking"

Observation Model

- For source i, we assume that $c_{n,i} \sim \text{Pois}(\lambda_{n,i})$
- Also assume $c_{b,i} \sim \text{Pois}(\lambda_{b,i} \frac{A_{b,i}}{A_{s,i}})$
- Finally, $c_{s,i} - c_{n,i} \sim \text{Pois}(\lambda_{b,i})$

A hierarchical Bayesian model for "stacking"

Observation Model

- For source i, we assume that $c_{n,i} \sim \text{Pois}(\lambda_{n,i})$
- Also assume $c_{b,i} \sim \text{Pois}(\lambda_{b,i} \frac{A_{b,i}}{A_{s,i}})$
- Finally, $c_{s,i} - c_{n,i} \sim \text{Pois}(\lambda_{b,i})$

Intensity Model

- If redshifts are known, can model luminosities directly & assume $L_i \sim \text{Lognormal}(\mu_L, \sigma_L)$ (or $L_i \sim \Gamma(\alpha_L, \beta_L)$)

A hierarchical Bayesian model for "stacking"

Observation Model

- For source i, we assume that $c_{n,i} \sim \text{Pois}(\lambda_{n,i})$
- Also assume $c_{b,i} \sim \text{Pois}(\lambda_{b,i} \frac{A_{b,i}}{A_{s,i}})$
- Finally, $c_{s,i} - c_{n,i} \sim \text{Pois}(\lambda_{b,i})$

Intensity Model

- If redshifts are known, can model luminosities directly & assume $L_i \sim \text{Lognormal}(\mu_L, \sigma_L)$ (or $L_i \sim \Gamma(\alpha_L, \beta_L)$)
- Otherwise, can apply analogous framework to flux f_i.

A hierarchical Bayesian model for "stacking"

Observation Model

- For source i, we assume that $c_{n,i} \sim \text{Pois}(\lambda_{n,i})$
- Also assume $c_{b,i} \sim \text{Pois}(\lambda_{b,i} \frac{A_{b,i}}{A_{s,i}})$
- Finally, $c_{s,i} - c_{n,i} \sim \text{Pois}(\lambda_{b,i})$

Intensity Model

- If redshifts are known, can model luminosities directly & assume $L_i \sim \text{Lognormal}(\mu_L, \sigma_L)$ (or $L_i \sim \Gamma(\alpha_L, \beta_L)$)
- Otherwise, can apply analogous framework to flux f_i.
- Generally assume $\lambda_{b,i} \sim \Gamma(\alpha_b, \beta_b)$

A hierarchical Bayesian model for "stacking"

Observation Model

- For source i, we assume that $c_{n,i} \sim \text{Pois}(\lambda_{n,i})$
- Also assume $c_{b,i} \sim \text{Pois}(\lambda_{b,i} \frac{A_{b,i}}{A_{s,i}})$
- Finally, $c_{s,i} - c_{n,i} \sim \text{Pois}(\lambda_{b,i})$

Intensity Model

- If redshifts are known, can model luminosities directly & assume $L_i \sim \text{Lognormal}(\mu_L, \sigma_L)$ (or $L_i \sim \Gamma(\alpha_L, \beta_L)$)
- Otherwise, can apply analogous framework to flux f_i.
- Generally assume $\lambda_{b,i} \sim \Gamma(\alpha_b, \beta_b)$
- Using noninformative priors on hyperparameters (Jefferys)

A hierarchical Bayesian model for “stacking”, continued

Key assumptions

A hierarchical Bayesian model for “stacking”, continued

Key assumptions

- For luminosity-based inference, assuming that redshifts are known
  - Relatively plausible for spectroscopic; not as much for photometric

A hierarchical Bayesian model for “stacking”, continued

Key assumptions

- For luminosity-based inference, assuming that redshifts are known
  - Relatively plausible for spectroscopic; not as much for photometric
- Assuming the spectra of sources are know & identical
  - Typically assume power law with photon index ≈ 1.7

A hierarchical Bayesian model for “stacking”, continued

Key assumptions

- For luminosity-based inference, assuming that redshifts are known
  - Relatively plausible for spectroscopic; not as much for photometric
- Assuming the spectra of sources are know & identical
  - Typically assume power law with photon index ≈ 1.7
- Attempting to make inferences only on selected sample, for now; not dealing with selection effects, etc.

Computation

Computation

- Using data augmentation algorithm with $\overrightarrow{c_n}$ as missing data

Computation

- Using data augmentation algorithm with $\overrightarrow{c_n}$ as missing data
- For Γ hyperdistributions, using Metropolis-Hastings step within Gibbs sampler to draw α & β

Computation

- Using data augmentation algorithm with $\overrightarrow{c_n}$ as missing data
- For Γ hyperdistributions, using Metropolis-Hastings step within Gibbs sampler to draw α & β
- For Lognormal hyperdistribution, using Gibbs step to draw μ_L & σ_L; Metropolis-Hastings step used to draw $\overrightarrow{\lambda_n}$

Computation

- Using data augmentation algorithm with $\overrightarrow{c_n}$ as missing data
- For Γ hyperdistributions, using Metropolis-Hastings step within Gibbs sampler to draw α & β
- For Lognormal hyperdistribution, using Gibbs step to draw μ_L & σ_L; Metropolis-Hastings step used to draw $\overrightarrow{\lambda_n}$
  - MH step here is very efficient; using Haley's method to identify posterior modes in parallel and tune proposal distribution.

Computation

- Using data augmentation algorithm with $\overrightarrow{c_n}$ as missing data
- For Γ hyperdistributions, using Metropolis-Hastings step within Gibbs sampler to draw α & β
- For Lognormal hyperdistribution, using Gibbs step to draw μ_L & σ_L; Metropolis-Hastings step used to draw $\overrightarrow{\lambda_n}$
  - MH step here is very efficient; using Haley's method to identify posterior modes in parallel and tune proposal distribution.
- From posterior simulations, can retain posterior mean & standard deviation of each source flux (and luminosity, if available) in addition to hyperparameter samples.
- This provides a great deal of information that is not available with conventional stacking in addition to estimates of sample properties with uncertainties.

Potential directions for further work

Potential directions for further work

- Currently have a very fast method that requires no more data than conventional stacking (and makes few additional assumptions).

Potential directions for further work

- Currently have a very fast method that requires no more data than conventional stacking (and makes few additional assumptions).
- Room for improvement in some areas:
  - Explicit handling of the PSF

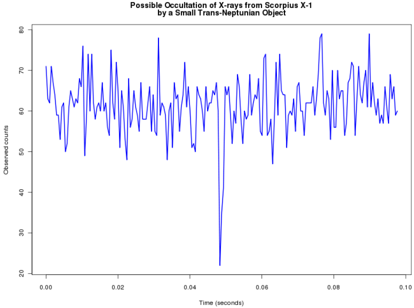
Potential directions for further work

- Currently have a very fast method that requires no more data than conventional stacking (and makes few additional assumptions).
- Room for improvement in some areas:
  - Explicit handling of the PSF
  - Incorporation of spectral uncertainties

Potential directions for further work

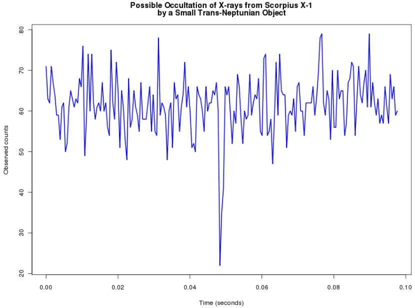
- Currently have a very fast method that requires no more data than conventional stacking (and makes few additional assumptions).
- Room for improvement in some areas:
  - Explicit handling of the PSF
  - Incorporation of spectral uncertainties
  - Incorporation of photometric redshift uncertainties

Testing time symmetry for astronomical events



- We have a set of x-ray light curves like the above, each of which is believed to contain an event (in this case, an occultation).

Testing time symmetry for astronomical events



- We have a set of x-ray light curves like the above, each of which is believed to contain an event (in this case, an occultation).
- Interested in testing if the event (a dimming, in this case) is time-symmetric.

Testing time symmetry for astronomical events

- Even for the Gaussian case, this is not entirely straightforward.

Testing time symmetry for astronomical events

- Even for the Gaussian case, this is not entirely straightforward.
  - Question of how much structure to place on shape of event.
  - Taking maximum over possible centers of event for less structured approach $\Rightarrow$ complex distribution of test statistic.

Testing time symmetry for astronomical events

- Even for the Gaussian case, this is not entirely straightforward.
  - Question of how much structure to place on shape of event.
  - Taking maximum over possible centers of event for less structured approach $\Rightarrow$ complex distribution of test statistic.
- With Poisson data, we really need a structured model.

Intensity model

- Define λ_t to be the intensity (count-rate) of our source at time t

Intensity model

- Define λ_t to be the intensity (count-rate) of our source at time t
- We model λ_t as:

$$\lambda_t = c - \alpha g(t; \tau, \theta)$$

where $\lim_{t \to \infty} g(t; \tau, \theta) = \lim_{t \to -\infty} g(t; \tau, \theta) = 0$ and $\sup_{\mathbb{R}} g(t; \tau, \theta) = g(\tau; \tau, \theta) = 1$

Intensity model

- Define λ_t to be the intensity (count-rate) of our source at time t
- We model λ_t as:

$$\lambda_t = c - \alpha g(t; \tau, \theta)$$

where $\lim_{t \to \infty} g(t; \tau, \theta) = \lim_{t \to -\infty} g(t; \tau, \theta) = 0$ and $\sup_{\mathbb{R}} g(t; \tau, \theta) = g(\tau; \tau, \theta) = 1$

- Thus, c characterizes our baseline source intensity, α characterizes the extent of the deviation from this baseline during the event, and $g(t; \tau, \theta)$ characterizes the shape of the event itself.

Observation model

- Given our series of intensities λ_t, we then model the observed counts at time t as:

$$n_t \sim \text{Pois}\left(\lambda_t\right)$$

Observation model

- Given our series of intensities λ_t, we then model the observed counts at time t as:

$$n_t \sim \mathrm{Pois}\left(\lambda_t\right)$$

- This approach generalizes easily to the high count regime with only minor modifications.

Testing

- We can then test the hypothesis of time symmetry by placing the appropriate restrictions on θ and calculating a likelihood-ratio test statistic.
- The challenge is then to find a parsimonious yet flexible form for the "event profile" $g(t; \tau, \theta)$.

Testing

- We can then test the hypothesis of time symmetry by placing the appropriate restrictions on θ and calculating a likelihood-ratio test statistic.
- The challenge is then to find a parsimonious yet flexible form for the "event profile" $g(t;\tau,\theta)$.
- One possibility: a "bilogistic" event profile

$$g(t;\tau,h_1,h_2,k_1,k_2)=\frac{1+e^{\frac{-h_t}{k_t}}}{1+e^{\frac{|t-\tau|-h_t}{k_t}}}$$

$$h_t=\begin{cases} h_1 & t<\tau \\ h_2 & t\geq\tau \end{cases}$$

$$k_t=\begin{cases} k_1 & t<\tau \\ k_2 & t\geq\tau \end{cases}$$

Testing

- We can then test the hypothesis of time symmetry by placing the appropriate restrictions on θ and calculating a likelihood-ratio test statistic.
- The challenge is then to find a parsimonious yet flexible form for the "event profile" $g(t;\tau,\theta)$.
- One possibility: a "bilogistic" event profile

$$g(t;\tau,h_1,h_2,k_1,k_2)=\frac{1+e^{\frac{-h_t}{k_t}}}{1+e^{\frac{|t-\tau|-h_t}{k_t}}}$$

$$h_t=\begin{cases} h_1 & t<\tau \\ h_2 & t\geq\tau \end{cases}$$

$$k_t=\begin{cases} k_1 & t<\tau \\ k_2 & t\geq\tau \end{cases}$$

Testing, continued

- Can also use Gaussian profile for event; tradeoff between degrees of freedom to characterize event and computational requirements.

Testing, continued

- Can also use Gaussian profile for event; tradeoff between degrees of freedom to characterize event and computational requirements.
- Because data is non-Gaussian, still need to simulate under null hypothesis to obtain actual distribution of test statistic (cannot necessarily rely on χ^2 approximation).

Maximizing the likelihood

- Another challenge: maximizing the likelihood for this model

Maximizing the likelihood

- Another challenge: maximizing the likelihood for this model
- It is very multimodal (lots of small, annoying, local maxima)

Maximizing the likelihood

- Another challenge: maximizing the likelihood for this model
- It is very multimodal (lots of small, annoying, local maxima)
- The good news: only the location parameter τ is truly troublesome

Maximizing the likelihood

- Another challenge: maximizing the likelihood for this model
- It is very multimodal (lots of small, annoying, local maxima)
- The good news: only the location parameter τ is truly troublesome
- A solution:
  1. Randomly draw a set of starting values for τ (possibly based on scan statistics or another simple method).
  2. For each starting value, run a fast, local optimization algorithm (such as Gauss-Newton) until convergence.
  3. Take the maximum of the values given by the local algorithms.

Maximizing the likelihood

- Another challenge: maximizing the likelihood for this model
- It is very multimodal (lots of small, annoying, local maxima)
- The good news: only the location parameter τ is truly troublesome
- A solution:
  1. Randomly draw a set of starting values for τ (possibly based on scan statistics or another simple method).
  2. For each starting value, run a fast, local optimization algorithm (such as Gauss-Newton) until convergence.
  3. Take the maximum of the values given by the local algorithms.
- This approach parallelizes extremely well, making it ideal for use in a cluster environment (such as Odyssey).