Detecting Variations in Source Intensity for Poisson Data

Y. Yu, XL. Meng, D. van Dyk, V. Kashyap, and A. Zezas

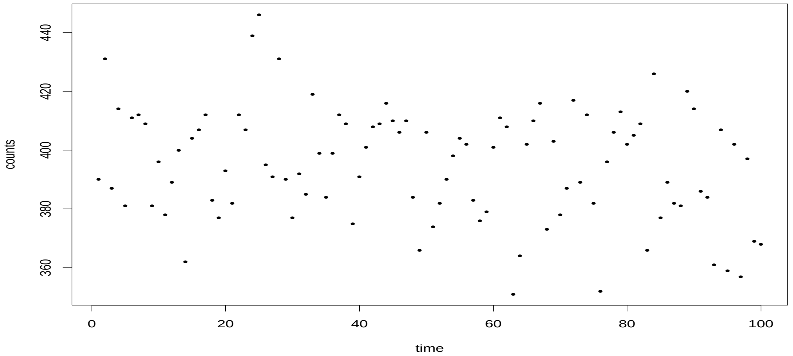
Abstract

The analysis of light curves is an especially challenging statistical task for low-count Poisson data: Variation in the source intensity may be shadowed by the Poisson variation of the counts. Here we discuss a class of statistical models that are designed to capture trends and autocorrelations in Poisson data that are collected over time. We expect such models to be useful not only in the analysis of sources that vary in their intensities but also in joint spectral-temporal modeling tasks such as analyzing the variability in spectral power law parameters or hardness ratios over time. To tackle these problems, we propose a Bayesian statistical model that takes into account both the variation in source intensity and the Poissonian character of the count data. Parameter estimation and error bars are computed using sophisticated MCMC (Markov chain Monte Carlo) methods. We illustrate our methods using several X-ray sources.

yu@stat.harvard.edu

Is There Variation in Source Intensity?

- If not, then counts in equal time intervals should follow independent Poisson distribution with the same parameter.
- Use *Posterior Predictive P-value* (ppp), a Bayesian hypothesis test method:
  - Simulate replicate datasets using parameters drawn from the posterior distribution.
  - Select a statistic T. Compute T_{obs}, the observed T for the original dataset.
  - For each replication, compute T. Compare T_{obs} with the distribution of T. The ppp is the proportion of time T_{obs} exceeds T.
- Example: A downward trend?



- Use ppp. Choose T as the correlation between counts and time index.
- The trend is statistically significant:

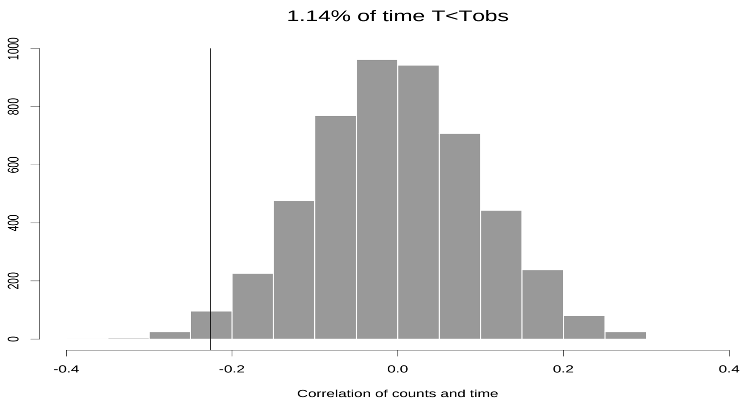


Fig. The posterior predictive distribution of T, vertical line being T_{obs}.

- Advantage: No theoretical/analytical calculations needed; choice of T is open.

Statistical Modeling for the Variation in Source Intensity

- We can design statistical models that account for two levels of variation:
  - the variation in source intensity, and
  - the Poisson variation of photon arrivals.
- Model formulation: Let Y_t be the counts in bin t, $t = 1, \cdots, T$.

$$Y_t \sim Poisson(d_t e^{\alpha+\mu t} \epsilon_t),$$

where d_t is the width (e.g., in seconds) of bin t;

$$\epsilon_t = e^{\xi_t - \tau^2/2}, \ \xi_t = \rho \xi_{t-1} + Z_t,$$

where $Z_t \sim N(0, (1-\rho^2)\tau^2)$, so that $\xi_t \sim N(0, \tau^2)$ is a stationary autoregressive process.

- Note: ϵ_t has mean 1 and variance $\sigma^2 = e^{\tau^2} - 1$. For small τ^2, $\sigma^2 \approx \tau^2$.

Easy Interpretation: Trend and Variation

Source intensity has two (multiplicative) components: a deterministic term $e^{\alpha+\mu t}$ and a stochastic term ϵ_t.

- When $\tau^2 = 0$, source intensity is simply $e^{\alpha+\mu t}$; μ controls the trend. (Here we consider exponential trend. Other choices present no additional difficulty.)
- E.g., suppose $\mu = -0.1$, and counts are collected by the day, then in a week ($\log 2/0.1$ days) we expect the source intensity to decrease by half.
- σ^2, the variance of ϵ_t, controls the variation of source intensity. Note σ^2 is a monotone increasing function of τ^2.
- E.g., suppose $\sigma^2 = 0.01$, then we expect a fluctuation of about 10% ($\sqrt{0.01}$) in source intensity around the general trend.

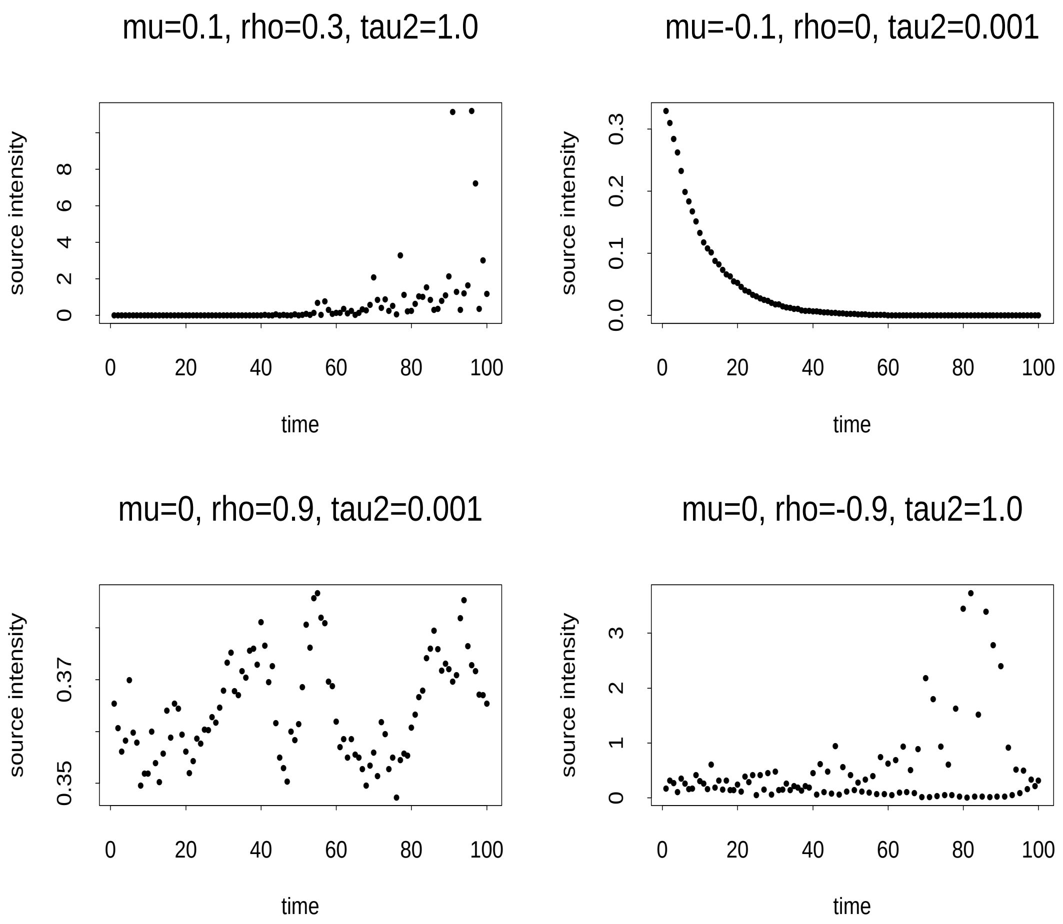
Easy Interpretation: Correlation

- ρ measures the association between ξ_t and ξ_{t-1}; ρ controls the correlation between source intensities.
- E.g., the correlation r between source intensities at t and $t-1$ (also the correlation between ϵ_t and ϵ_{t-1}):

$$r = (e^{\rho\tau^2} - 1)/\sigma^2$$

  - $r = 1$ if $\rho = 1$; $r = 0$ if $\rho = 0$; and $r = -e^{-\tau^2}$ if $\rho = -1$.
  - r is a monotone increasing function of ρ; for small τ^2, $r \approx \rho$.

Possible Source Intensity Patterns Produced by this Model



Computation and Application

- To learn about the parameters of interest μ, ρ, σ^2, use posterior simulation by Markov chain Monte Carlo (MCMC). Here we use a Gibbs sampler with Metropolis steps.
- A real example: The isolated neutron star/quark star candidate RX J1856.5-3754 observed by Chandra HRC-S. It doesn't seem to vary, or does it?

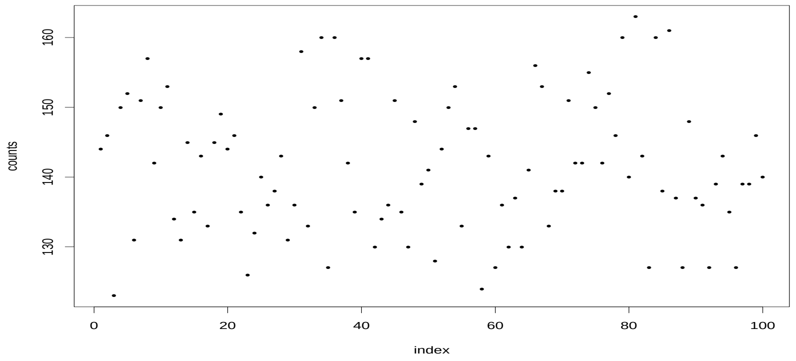


Fig. Counts Y_t (exposure time 55476 seconds, divided into 100 equal bins).

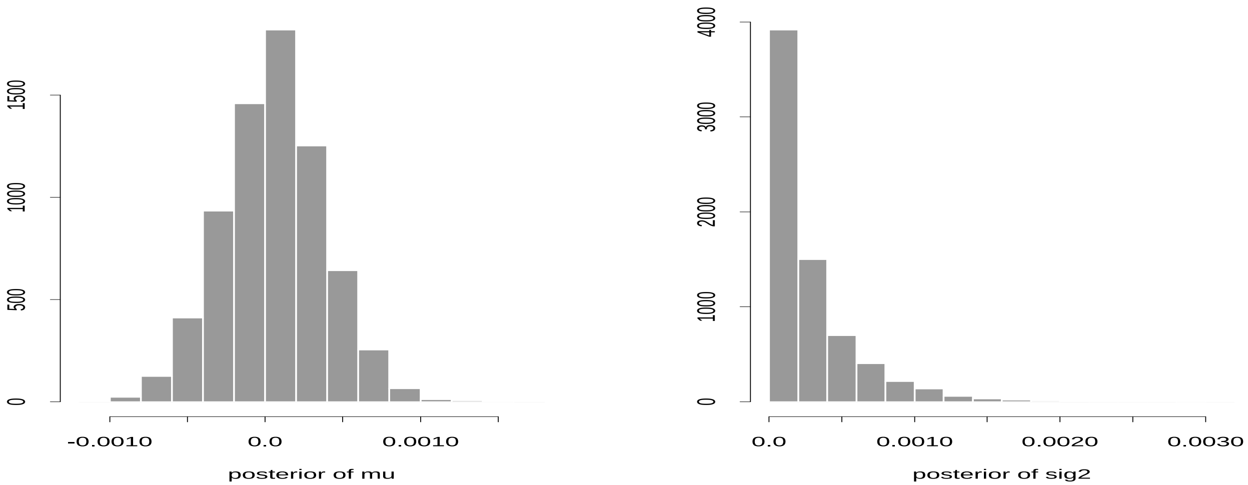


Fig. Posterior distributions of $\mu,\ \sigma^2$.

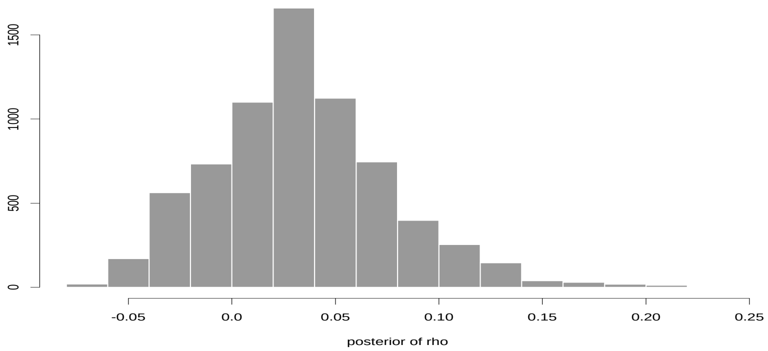


Fig. Posterior distribution of ρ.

The Binning Issue

Model interpretation depends on the binning:

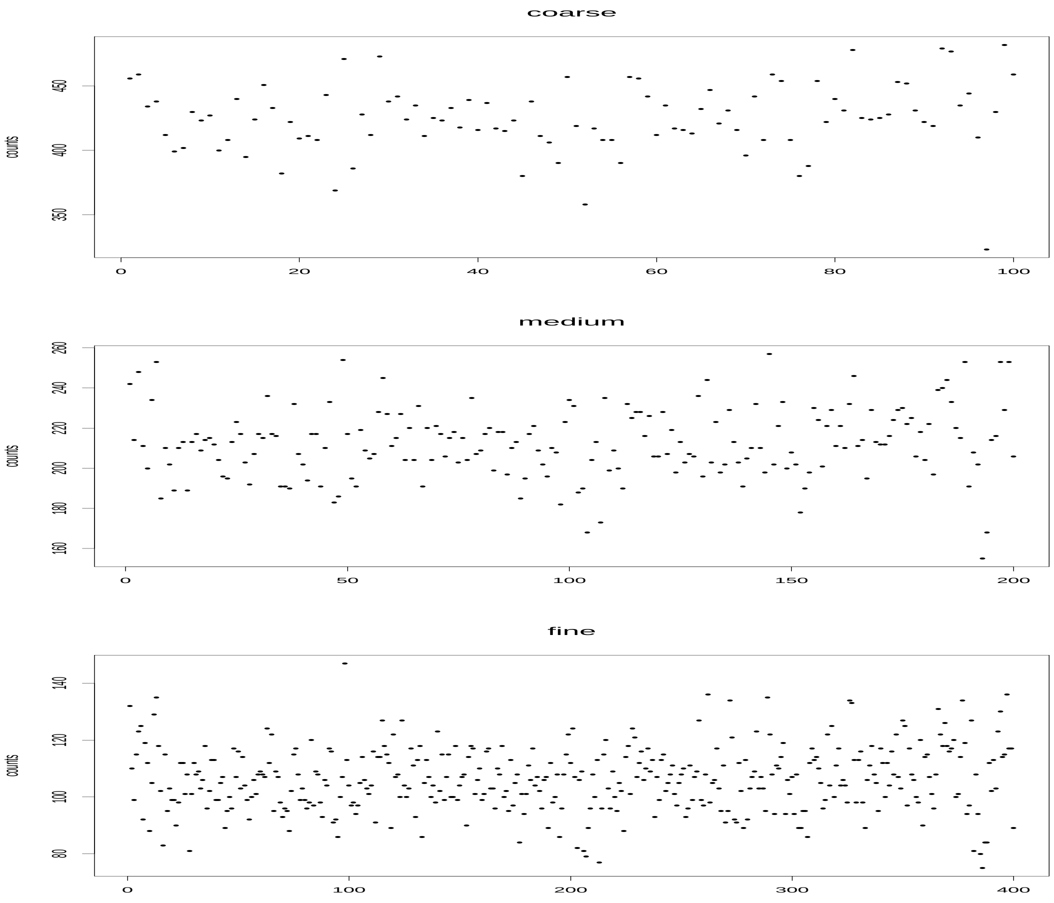


Fig. Counts data for another observation of RX J1856.5-3754, with various binning.

Parameter estimates for different binning schemes:

	μ	ρ	σ^2
coarse	0.000360 (0.000233)	0.0674 (0.0448)	0.00151 (0.000585)
medium	0.000181 (0.000103)	0.113 (0.0531)	0.00178 (0.000641)
fine	9.55e-05 (4.87e-05)	0.193 (0.0583)	0.00219 (0.000752)

Note: posterior standard deviations in parentheses.

- Estimates of μ, the trend parameter, agree, after taking into account the scale difference.
- Estimates of ρ are different. For the fine binning, the estimate shows a noticeable albeit small autocorrelation between source intensities at consecutive time points.

Work in Progress

- More efficient Markov chain algorithm. Faster convergence.
- Joint spectral-temporal modeling, e.g., the variability of power law parameters or hardness ratios over time.
- Rigorous procedures to check goodness-of-fit of the statistical model. Conventional tests such as the χ^2 test does not apply because of the stochastic component included in the source intensity formula.