Comparison of Goodness–of–fit Assessment Methods with C statistics in Astronomy

Yang Chen
Joint work with X. Li, X. Meng, V. Kashyap and M. Bonamente

Department of Statistics, University of Michigan

Outline

1. Spectral Model
2. C Statistics
3. Hypothesis Testing
4. Numerical Performance

Outline

1. Spectral Model
2. C Statistics
3. Hypothesis Testing
4. Numerical Performance

Spectral Model

Let

$$N_i|\boldsymbol{\theta} \overset{\text{indep.}}{\sim} \text{Poisson}(s_i(\boldsymbol{\theta})); \quad i = 1, \ldots, n. \tag{1}$$

The spectral model in each energy bin i is $s_i(\boldsymbol{\theta})$, which can be regarded as a *known* function of unknown parameters $\boldsymbol{\theta}$.

In the astronomy community, the chi-squared statistics defined in Equation (2) is typically adopted for performing both model fitting and goodness-of-fit assessment [Kaastra, 2017; Bonamente, 2019]:

$$\chi^2(\boldsymbol{\theta}) := \sum_{i=1}^{n} \frac{(N_i - s_i(\boldsymbol{\theta}))^2}{s_i(\boldsymbol{\theta})} \quad \text{or} \quad \sum_{i=1}^{n} \frac{(N_i - s_i(\boldsymbol{\theta}))^2}{N_i}. \tag{2}$$

Spectral Model

The spectral model in a given bin, labeled by the index i, is defined as

$$s_i(\theta) = \int_{\underline{E}}^{\overline{E}} \mathrm{R}(E, i) A(E) f(E, \theta)\; dE + B_i, \tag{3}$$

where the spectral model $f(E, \theta)$ can be

- three-parameter absorbed power–law model [Rybicki and Lightman, 1979]

$$f(E, \theta = \{K, N_H, \Gamma\}) = \mathrm{K} \cdot e^{-N_H \cdot \sigma(E)} \cdot E^{-\Gamma}, \tag{4}$$

- absorbed thermal spectrum [Allen et al., 2004] defined as

$$f(E, \theta = \{K, N_H, T, A\}) = K \cdot e^{-N_H \cdot \sigma(E)} \cdot \epsilon(T, A), \tag{5}$$

We are interested in the goodness-of-fit assessment of

$$H_0 : s_i = f_i(\boldsymbol{\theta}) = \sum_{j=1}^{m} c_{ij} g_j(\boldsymbol{\theta}) \quad \text{versus} \quad H_1 : \text{every } s_i \text{ is free}, \tag{6}$$

where c_{ij} are constants and g_i are smooth functions that may depend on i and m.

Outline

C Statistics: Definitions

Define a null hypothesis and the C statistics *function* as

$$H_0 : \mathcal{S} = \{s_i(\boldsymbol{\theta}), 1 \leq i \leq n\} \subset \mathbb{R}^n. \tag{7}$$

$$C_n(\boldsymbol{\theta}) = 2 \sum_{i=1}^{n} \left[s_i(\boldsymbol{\theta}) - N_i \log s_i(\boldsymbol{\theta}) - N_i + N_i \log N_i \right]. \tag{8}$$

And $C_{\min}$ is obtained by plugging the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$, i.e.

$$C_{\min} = C_n(\hat{\boldsymbol{\theta}}) = 2 \sum_{i=1}^{n} [s_i(\hat{\boldsymbol{\theta}}) - N_i \log s_i(\hat{\boldsymbol{\theta}}) - N_i + N_i \log N_i] = -2 \log \Lambda_n(\hat{\boldsymbol{\theta}}).$$

Notion of Asymptotics

Lemma

If $\sum_{i=1}^{n} s_i(\boldsymbol{\theta}^\star) \to \infty$, *then there exists* $\{m_1, \ldots, m_n\}$ *such that (1)* $\sum_{i=1}^{n} m_i \to \infty$, *(2)* $m_i = 1$ *when* $s_i(\boldsymbol{\theta}^\star) \leq 1$, *(3)* $0.5 < s_i(\boldsymbol{\theta}^\star)/m_i < 1$ *when* $s_i(\boldsymbol{\theta}^\star) > 1$, *(4) the likelihood is equivalent to the likelihood of the following model*

$$\tilde{N}_{ij} \overset{\text{indep.}}{\sim} \text{Poisson}\left(\frac{s_i(\boldsymbol{\theta}^*)}{m_i}\right), \quad \sum_{j=1}^{m_i} \tilde{N}_{ij} = N_i. \tag{9}$$

Outline

Hypothesis Testing with Wilk's Theorem

Lemma (Wilk's Theorem for C statistics)

Under the regularity conditions (B1) and assuming that the null hypothesis H_0 given in (7) *is true, then*

$$\Gamma_n := C_{\rm true} - C_{\min} \to \chi^2_d \quad \text{as} \quad n \to \infty, \tag{10}$$

where d is the number of adjustable/free parameters of the null model. Note that here the dimension of the full space is not fixed, which is actually n, thus the Wilk's theorem of the likelihood ratio test does not hold for the $C_{\min}$ statistics itself, i.e., $C_{\min} = C_n(\hat{\boldsymbol{\theta}})$ does not converge to a chi-squared distribution with degree of freedom $n - d$ as $n \to \infty$.

Asymptotic Normality of C Statistic

Under H_0 and regularity conditions (B), we have

$$\text{(i) Not computable} \quad \frac{C_n(\hat{\boldsymbol{\theta}}) - \mathbb{E}[C_n(\hat{\boldsymbol{\theta}})]}{\sqrt{\mathrm{Var}(C_n(\hat{\boldsymbol{\theta}}))}} \to N(0,1), \tag{11}$$

$$\text{(ii) Bootstrap failure} \quad T = \frac{C_n(\hat{\boldsymbol{\theta}}) - \hat{\mu}(\boldsymbol{\theta})}{\hat{\sigma}(\boldsymbol{\theta})} \to N(0,1), \tag{12}$$

and conditionally on $\hat{\boldsymbol{\theta}}$

$$\text{(iii) Computable} \quad \frac{C_n(\hat{\boldsymbol{\theta}}) - \mathbb{E}[C_n(\hat{\boldsymbol{\theta}})|\hat{\boldsymbol{\theta}}]}{\sqrt{\mathrm{Var}(C_n(\hat{\boldsymbol{\theta}})|\hat{\boldsymbol{\theta}})}} \to N(0,1), \tag{13}$$

where $\mu(\boldsymbol{\theta}) = \mathbb{E}[C_n(\boldsymbol{\theta})]$, $\sigma^2(\boldsymbol{\theta}) = \mathrm{Var}(C_n(\boldsymbol{\theta})) - Q(\boldsymbol{\theta})$, $Q(\boldsymbol{\theta}) = \mathbf{c}^\top(\boldsymbol{\theta})\mathbf{I}^{-1}(\boldsymbol{\theta})\mathbf{c}(\boldsymbol{\theta})$ and $\mathbf{c}(\boldsymbol{\theta}) = \mathrm{Cov}\left\{C_n(\boldsymbol{\theta}), \mathbf{D}\ell(\boldsymbol{\theta})\right\} = O_p(n)$.

Algorithms of Goodness-of-fit

Algorithm Number (Name)	Method
Algorithm 1 (LR-χ^2)	Likelihood ratio with χ^2 statistics
Algorithm 2a (K-B-Expansion)	Z-test with Polynomial approximation
Algorithm 2b (Bootstrap-Gaussian)	Z-test with bootstrap mean & variance
Algorithm 3a (High-Order-Marginal)	Z-test based on (ii) in Theorem
Algorithm 3b (High-Order-Conditional)	Z-test based on (iii) in Theorem
Algorithm 4a (Parametric-Bootstrap)	Parametric bootstrap with estimated p-value
Algorithm 4b (B-C-Bootstrap)	Parametric bootstrap with bias correction
Algorithm 4c (Double-Bootstrap)	Double bootstrap with adjusted p-value

Table 1: List of algorithms considered in numerical studies.

Alg.1	Alg.2a	Alg.2b	Alg.3	Alg.4a	Alg.4b	Alg.4c
$O(1)$	$O(n^2)$	$O(Bn^2)$	$O(n^2)$	$O(Bn^2)$	$O((B_1 + B_2)n^2)$	$O(B_1 B_2 n^2)$

Table 2: Computational complexity of the different algorithms when the link function of s involves summing over all channels.

Outline

Numerical Results: Simulation Experiments

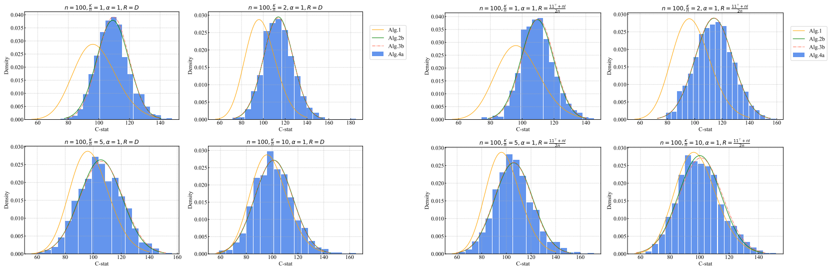

Figure 4.1: Histograms of the null distributions of $C_{\min}$ from Algorithms 1-4 under two case of RMF and a powerlaw spectrum. The bootstrap size is $B = 1000$.

n	10	30	50	75	100
Single Bootstrap ($B = 100$)	0.90	7.07	20.24	47.36	84.99
High-order	0.02	0.13	0.36	0.77	1.38

Table 3: Average run time (CPU seconds) of Algorithms 2b and 3b under Case 1.

Real Data: quasar PG 1116+215

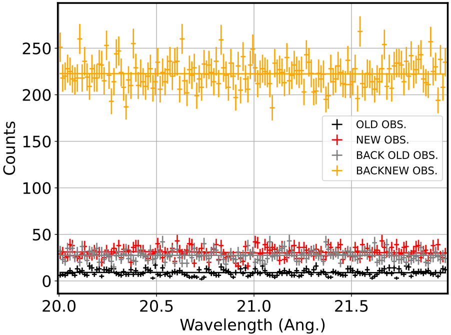

Figure 4.2: X–ray count spectra of two observations of the quasar PG 1116+215, labeled as OLD and NEW. Each observation has a total number of counts in a small source region (source plus background counts, in black and red), and the total number of counts in a larger background region that needs to be re–scaled by a deterministic factor before subtraction from the total number of counts (in grey and orange). See Bonamente et al. [2016] for details of data processing.

Spectrum	$C_{\min}$	Alg.1	Alg.2b	Alg.3b	Alg.4a
OLD	190.72	0.039*	0.060	0.054	0.060
NEW	167.67	0.284	0.324	0.312	0.313
BACK OLD	171.39	0.221	0.254	0.244	0.257
BACK NEW	153.46	0.587	0.621	0.603	0.614

Table 4: p-values of four test methods in each spectrum. The bootstrap size is $B = 1000$.

	Spectrum	Alg.1	Alg.2b	Alg.3b	Alg.4a
Average p-value	NEW	0.488	0.361	0.372	0.356
	BACK NEW	0.122	0.493	0.499	0.490
Rejections ($\alpha = 0.10$)	NEW	0	2	2	2
	BACK NEW	11	2	2	2

Table 5: Performance of four test methods in each spectrum with shorter exposure time. The bootstrap size is $B = 1000$.

References I

S. W. Allen, R. W. Schmidt, H. Ebeling, A. C. Fabian, and L. van Speybroeck. Constraints on dark energy from Chandra observations of the largest relaxed galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 353: 457–467, Sept. 2004. doi: 10.1111/j.1365-2966.2004.08080.x.

M. Bonamente. Distribution of the c statistic with applications to the sample mean of poisson data. *Journal of Applied Statistics*, pages 1–22, 2019.

M. Bonamente, J. Nevalainen, E. Tilton, J. Liivamägi, E. Tempel, P. Heinämäki, and T. Fang. A possible Chandra and Hubble Space Telescope detection of extragalactic WHIM towards PG 1116+215. *Monthly Notices of the Royal Astronomical Society*, 457:4236–4247, Apr. 2016. doi: 10.1093/mnras/stw285.

J. Kaastra. On the use of c-stat in testing models for x-ray spectra. *Astronomy & Astrophysics*, 605:A51, 2017.

G. B. Rybicki and A. P. Lightman. *Radiative Processes in Astrophysics*. New York, Wiley-Interscience, 1979. URL http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=1979rpa..book.....R&db_key=AST.