# Making high-order asymptotics practical: correcting goodness-of-fit test for astronomical count data

Xiaoli Li

Joint work with M. Bonamente, Y. Chen, D. van Dyk, V. Kashyap and X. Meng

Department of Statistics, The University of Chicago

# Outline

In X-ray astronomy, what is of interest is whether the data collected by the observatory follow a parent spectrum model.

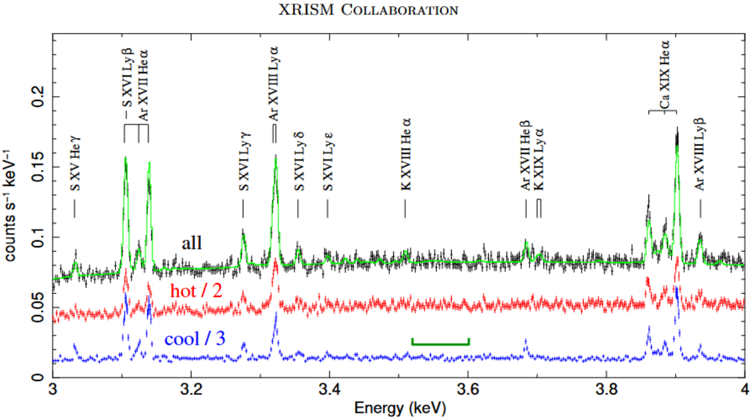



Figure 1: The stacked XRISM spectra in the rest-frame 3-4 keV. (Taken from XRISM Collaboration et al. [2025]).

# Spectral Model

Let

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

where $s_i(\boldsymbol{\theta})$ is the spectral model and is typically regarded as a *known* function of unknown parameters $\boldsymbol{\theta}$.

In the astronomy community, the $\chi^2$-statistic,

$$\chi_n^2(\hat{\boldsymbol{\theta}}) := \sum_{i=1}^{n} \frac{(N_i - s_i(\hat{\boldsymbol{\theta}}))^2}{{\sigma_i}^2}, \tag{2}$$

is typically adopted for both model fitting and goodness-of-fit assessment[Kaastra, 2017, Bonamente, 2019], where $\sigma_i^2 = N_i \vee 1$ or $s_i(\hat{\boldsymbol{\theta}})$, which corresponds to the Neyman $\chi^2$ statistic [Neyman, 1949] and the Pearson statistic [Pearson, 1900] respectively.

# Spectral Model

The spectral model in a given bin, labeled by the index $i$, can be defined as

$$s_i(\boldsymbol{\theta}) = T\left(\int_{\underline{X}}^{\overline{X}} \mathrm{R}(X,i)A(X)g(X,\boldsymbol{\theta})\,\mathrm{d}X + B_i\right), \tag{3}$$

which in practice is approximated by the sum

$$\hat{s}_i(\boldsymbol{\theta}) = \sum_{j=1}^{J} R(\tilde{X}_j, i)A(\tilde{X}_j)g(\tilde{X}_j, \boldsymbol{\theta})\,[X_{j+1} - X_j] + B_i. \tag{4}$$

Here a common choice is the power-law model, written as

$$g(X, \boldsymbol{\theta} = \{K, \Gamma\}) = K \cdot (X/X_0)^{-\Gamma}, \tag{5}$$

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

$$H_0 : s_i(\boldsymbol{\theta}) = \sum_{j=1}^{J+1} c_{ij}\tilde{g}_j(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \quad \text{vs} \quad H_1 : \text{every } s_i \text{ is free} \tag{6}$$

# C Statistics: Definitions

First consider the simple null $\Theta = \{\boldsymbol{\theta}\}$. The C statistics under the simple null $C_{\text{true}}$ is defined as the likelihood ratio statistic

$$C_{\text{true}} = C_n(\boldsymbol{\theta}) = 2\sum_{i=1}^{n}\left[s_i(\boldsymbol{\theta}) - N_i \log s_i(\boldsymbol{\theta}) - N_i + N_i \log N_i\right]. \tag{7}$$

And for the composite null $\Theta \subset \mathbb{R}^d$, $C_{\min}$ is obtained by plugging the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$, i.e.

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

Based on the likelihood ratio test (LRT), we reject $H_0$ at level $\alpha$ if $C_{\min} \geq \Delta_\alpha C_{\min}$, where $\Delta_\alpha C_{\min}$ is the critical value at level $\alpha$.

# C statistics: Wilks' Theorem

**Lemma (Wilks' Theorem for C-statistics)**

*Under mild conditions and $H_0$,*

$$\Gamma_n := C_n(\boldsymbol{\theta}^\star) - C_n(\hat{\boldsymbol{\theta}}) \to \chi_d^2 \quad \text{as} \quad n \to \infty, \tag{9}$$

where $d$ is the number of adjustable/free parameters of the null model and $\boldsymbol{\theta}^*$ is the (unknown) true parameter under $H_0$. Note that here the dimension of the full space is not fixed, which is actually $n$. Thus the Wilks'theorem does not hold for $C_{\text{min}}$.

Roughly speaking, this lemma indicates that when $d \ll n$, the distribution of $C_{\text{min}}$ is close enough to the distribution of $C_{\text{true}}$. Thus asymptotically, we only need to analyze the distribution of $C_{\text{true}} = C_n(\boldsymbol{\theta}) = C(N_1, \ldots, N_n; \boldsymbol{\theta})$.

# Hypothesis Testing: Easy Case

Define $\mu(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[C_{\text{true}}]$ and $\sigma^2(\boldsymbol{\theta}) = \text{Var}_{\boldsymbol{\theta}}(C_{\text{true}})$. Because $C_{\text{true}}$ is sum of independent random variables, using Central Limit Theorem, we have

$$\frac{C_{\text{true}} - \mu(\boldsymbol{\theta}^*)}{\sigma(\boldsymbol{\theta}^*)} \to N(0,1). \tag{10}$$

When the true parameter $\boldsymbol{\theta}^*$ is known, with this asymptotic normality, the critical value at level $\alpha$ can be calculated as

$$\Delta_\alpha C = \mu(\boldsymbol{\theta}^*) + z_{1-\alpha}\sigma(\boldsymbol{\theta}^*), \tag{11}$$

where $z_\alpha$ is the $\alpha$-quantile of standard normal distribution.

However, for the common composite null $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$, the true parameter $\boldsymbol{\theta}^*$ is unknown.

# Hypothesis Testing: Challenging Case

To fix this issue, several methods based on empirical approximation are proposed [Kaastra, 2017, Bonamente, 2019]. These methods are asymptotically equivalent to the following naive method.

1: Fit the model to get the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$,
2: Calculate $\mu(\hat{\boldsymbol{\theta}})$ and $\sigma(\hat{\boldsymbol{\theta}})$ by assuming $\boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}$,
3: Calculate $\widehat{\Delta_\alpha C} = \mu(\hat{\boldsymbol{\theta}}) + z_{1-\alpha}\sigma(\hat{\boldsymbol{\theta}})$.

Although $\mu(\hat{\boldsymbol{\theta}})$ and $\sigma(\hat{\boldsymbol{\theta}})$ are consistent estimators, when $n$ is large, we have

$$\frac{C_{\text{true}} - \mu(\hat{\boldsymbol{\theta}})}{\sigma(\hat{\boldsymbol{\theta}})} \approx \frac{C_{\text{true}} - \mu(\boldsymbol{\theta}^*)}{\sigma(\boldsymbol{\theta}^*)} - \frac{\mu(\hat{\boldsymbol{\theta}}) - \mu(\boldsymbol{\theta}^*)}{\sigma(\boldsymbol{\theta}^*)}, \tag{12}$$

where the second term is $O_p(1)$ with zero mean. Thus, the limiting distribution is zero-mean with non-identical variance, which is not $N(0,1)$ in most cases.

# Hypothesis Testing: Background Summary

We want to test whether the data follow a parent spectrum model, i.e. $H_0 : s_i = s_i(\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}$.

The Wilks' Theorem indicates that when $d \ll n$, asymptotically it is equivalent to test $H_0 : s_i = s_i(\boldsymbol{\theta}^*)$, where $\boldsymbol{\theta}^*$ is the (unknown) true parameter under the original $H_0$.

When $\boldsymbol{\theta}^*$ is unknown, because of approximation error, the naive method, which essentially tests $H_0 : s_i = s_i(\hat{\boldsymbol{\theta}})$, i.e. assuming $\boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}$, fails.

Question: Can we either have a better approximation, or consider other distributions?

# Corrected Method I

Therefore, we try to modify the test statistics so that its limiting distribution is standard normal.

**Theorem (Computable corrected limiting distribution of C statistics)**

*Under $H_0$ and mild conditions, as $n \to \infty$, we have*

$$T_1 = \frac{C_{\text{true}} - \mu(\hat{\boldsymbol{\theta}})}{\sqrt{\sigma^2(\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}})}} \xrightarrow{D} N(0,1), \tag{13}$$

*where* $Q(\boldsymbol{\theta}) = \boldsymbol{c}^\top(\boldsymbol{\theta}) I_n^{-1}(\boldsymbol{\theta}) \boldsymbol{c}(\boldsymbol{\theta})$, *and* $\boldsymbol{c}(\boldsymbol{\theta}) = \mathrm{Cov}_{\boldsymbol{\theta}}\left\{C_n(\boldsymbol{\theta}), \boldsymbol{D}\ell(\boldsymbol{\theta})\right\}$.

Here $Q(\boldsymbol{\theta})$ will decrease to 0 dramatically when $s_i(\boldsymbol{\theta}) \to \infty$ uniformly. In practice, $Q(\boldsymbol{\theta})$ is negligible when $s_i(\boldsymbol{\theta}) \geq 1$ uniformly, which means that the naive method works well in this case (moderate count regime).

Note that here all the quantities only depend on $\hat{\boldsymbol{\theta}}$, which means they are computable from the data.

# Failure of Naive Method – Approximation View

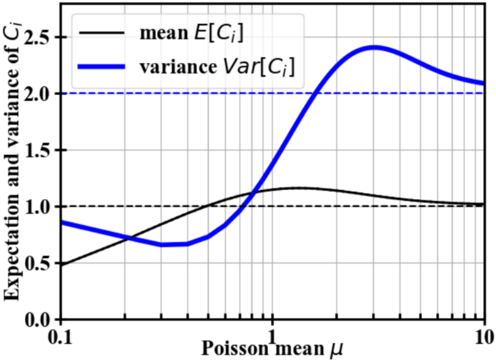



Figure 2: Expectation and variance of each term of C statistic (taken from Bonamente [2019]).

When $\mu > 1$, the local fluctuation of expectation is small, making $Q(\boldsymbol{\theta})$ negligible.

# Failure of Naive Method – Post-Selection View

Intuitively, the naive method will fail because we use the data twice. First to get MLE $\hat{\boldsymbol{\theta}}$. Then again for hypothesis testing.

Example 1. Consider a simple case $\Theta = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$. Assume $\boldsymbol{\theta}^* = \boldsymbol{\theta}_1$. Then when $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_1$, the estimated moments are perfect. However, when $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_2$, the estimation can be very poor.

Example 2. When the null is $H_0 : s_i(\theta) = \mu \in \mathbb{R}$, the MLE is $\hat{\theta} = \bar{N} = \sum_{i=1}^{n} N_i/n$ and the conditional distribution of $N_i|\hat{\theta} \sim \text{Multinomial}(n\bar{N}, 1/n)$. In this case, we can calculate the exact conditional distribution and do inference. In contrast, the naive method uses a mis-specified distribution $N_i|\hat{\theta} \sim \text{Pois}(\bar{N})$.

# Corrected Method II

Inspired by the examples above, we also analyze the conditional distribution of $C_{\min} = C_n(\hat{\boldsymbol{\theta}})$.

**Theorem (Computable conditional limiting distribution of C stat)**

*Under $H_0$ and mild conditions, as $n \to \infty$, conditional on $\hat{\boldsymbol{\theta}}$, we have*

$$T_2 = \frac{C_n(\hat{\boldsymbol{\theta}}) - \mathbb{E}_{\boldsymbol{\theta}^*}[C_n(\hat{\boldsymbol{\theta}}) \mid \hat{\boldsymbol{\theta}}]}{\sqrt{\mathrm{Var}_{\boldsymbol{\theta}^*}[C_n(\hat{\boldsymbol{\theta}}) \mid \hat{\boldsymbol{\theta}}]}} \xrightarrow{D} N(0,1), \tag{14}$$

*and*

$$\mathbb{E}_{\boldsymbol{\theta}^*}(C_n(\hat{\boldsymbol{\theta}}) \mid \hat{\boldsymbol{\theta}}) = \hat{\kappa}_1^{(\cdot)} + O(1) = \mu(\hat{\boldsymbol{\theta}}) + O(1), \tag{15}$$

$$\mathrm{Var}_{\boldsymbol{\theta}^*}(C_n(\hat{\boldsymbol{\theta}}) \mid \hat{\boldsymbol{\theta}}) = \hat{\kappa}_2^{(\cdot)} - \hat{\kappa}_{11}^\top \hat{X}(\hat{X}^\top \hat{V}^{-1}\hat{X})^{-1}\hat{X}^\top \hat{\kappa}_{11} + O(1) \tag{16}$$

$$= \sigma^2(\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}) + O(1). \tag{17}$$

Counts in $n = 100$ channels are generated under a power-law model with and without an added emission or absorption line

$$\begin{aligned} s_i(\boldsymbol{\theta}) &= KE_i^{-\Gamma}, \quad \text{for } 1 \le i \le m_1,\ m_2 < i \le n, \quad \text{and} \\ s_i(\boldsymbol{\theta}) &= \Psi, \qquad\quad \text{for } m_1 < i \le m_2, \end{aligned} \tag{18}$$

where $K > 0$ and $\Gamma \in \mathbb{R}$ are fitted parameters, $\{E_i = 1 + i/n;\ i = 1, \ldots, n\}$ define energy bins over the range $(1, 2)$, and $1 \le m_1 \le m_2 \le n$ define the location and width of the line. We test the power-law model with no line $(m_1 = m_2)$ as the null $H_0$.

| Algorithm Number (Name) | Method |
|---|---|
| Alg.1 (LR-$\chi^2$) | $\chi^2$ approximation |
| Alg.2 (Naive Z-test) | Z-test with naive moments estimation |
| Alg,3 (Corrected Z-test) | Z-test with corrected moments estimation |
| Alg.4 (Param. Bootstrap) | Parametrically resample $\{N_i\}_{i=1}^n$ B times |

Table 1: List of algorithms considered in numerical studies.

# Simulation Experiments – Setting

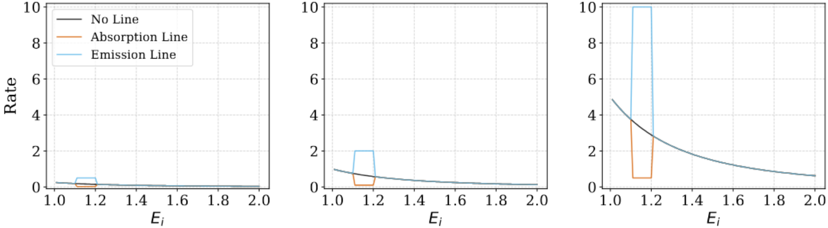
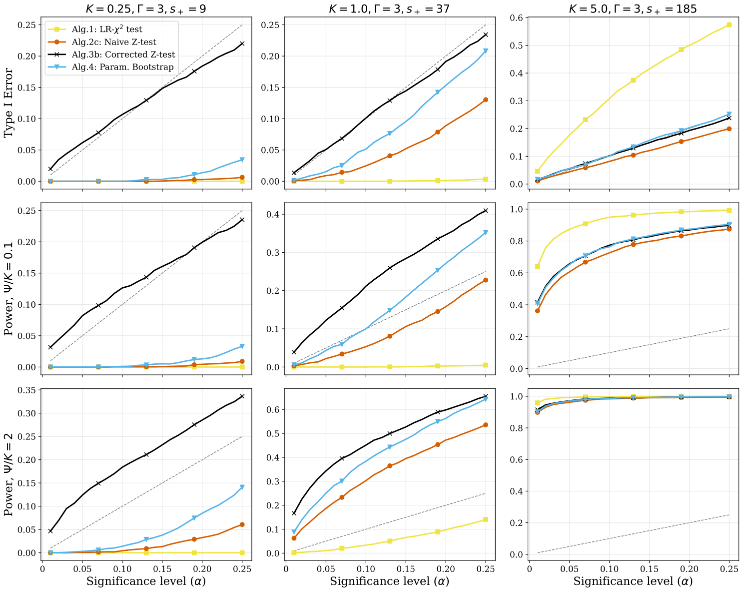



Figure 3: The spectra used in each simulation setting, i.e., the Power-law model with/without a line. Left: Challenging low-count case, where $K = 0.25$ on average only $s_+ \approx 9$ counts are expected to be observed over $n = 100$ bins. Middle: Difficult low-count case, where $K = 1.0$ and $s_+ \approx 40$. Right: Easy moderate-count case, where $K = 5.0$ and $s_+ \approx 200$.

# Simulation Experiments – Performance



# Simulation Experiments – Computational Time

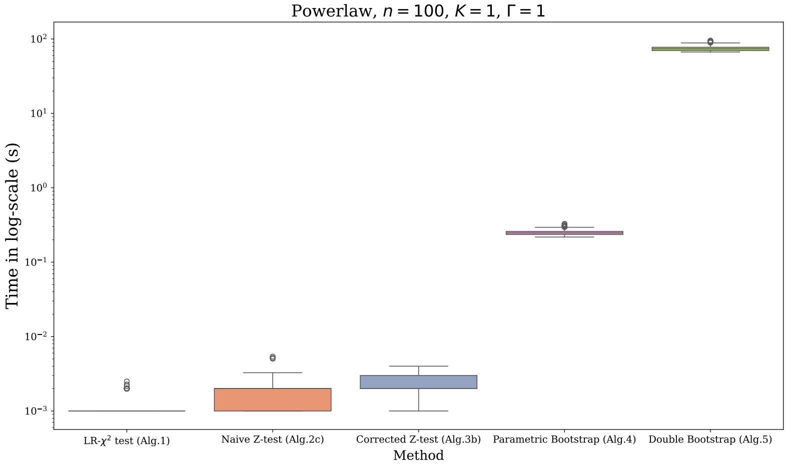



Figure 4: Comparison on the computational time of the algorithms.

# Conclusions

Conclusions

- In large-count settings, all algorithms perform similarly well.
- In small-count settings, the corrected Z-test based on high-order asymptotics give superior performances compared to other methods including the bootstrap.
- We publish a Python package to implement our proposed method for conducting goodness-of-fit tests with Poisson count data, which is called *High-Energy Astronomy Goodness-of-Fit (HEAGOF)*.

Questions

- In extreme cases, the distribution of C-stat will degenerate to a point, makeing the test powerless. C-stat itself is also divisible, which is not a good test statistic as pointed by Algeri and Khmaladze [2025].
- The methods we propose above are restricted to Poisson count data. Can we extend them to over-dispersed data such as negative binomial count data?

# References I


S. Algeri and E. V. Khmaladze. On the statistical analysis of grouped data: when pearson $\chi^2$ and other divisible statistics are not goodness-of-fit tests. *Under Review*, 2025.

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

XRISM Collaboration, Marc Audard, Hisamitsu Awaki, Ralf Ballhausen, Aya Bamba, Ehud Behar, Rozenn Boissay-Malaquin, Laura Brenneman, Gregory V. Brown, Lia Corrales, Elisa Costantini, Renata Cumbee, Maria Diaz Trigo, Chris Done, Tadayasu Dotani, Ken Ebisawa, Megan E. Eckart, Dominique Eckert, Satoshi Eguchi, Teruaki Enoto, Yuichiro Ezoe, Adam Foster, Ryuichi Fujimoto, Yutaka Fujita, Yasushi Fukazawa, Kotaro Fukushima, Akihiro Furuzawa, Luigi Gallo, Javier A. García, Liyi Gu, Matteo Guainazzi, Kouichi Hagino, Kenji Hamaguchi, Isamu Hatsukade, Katsuhiro Hayashi, Takayuki Hayashi, Natalie Hell, Edmund Hodges-Kluck, Ann Hornschemeier, Yuto Ichinohe, Daiki Ishi, Manabu Ishida, Kumi Ishikawa, Yoshitaka Ishisaki, Jelle Kaastra, Timothy Kallman, Erin Kara, Satoru Katsuda, Yoshiaki Kanemaru, Richard Kelley, Caroline Kilbourne, Shunji Kitamoto, Shogo Kobayashi, Takayoshi Kohmura, Aya Kubota, Maurice Leutenegger, Michael Loewenstein, Yoshitomo Maeda, Maxim Markevitch, Hironori Matsumoto, Kyoko Matsushita, Dan McCammon, Brian McNamara, François Mernier, Eric D. Miller, Jon M. Miller, Ikuyuki Mitsuishi, Misaki Mizumoto, Tsunefumi Mizuno, Koji Mori, Koji Mukai, Hiroshi Murakami, Richard Mushotzky, Hiroshi Nakajima, Kazuhiro Nakazawa, Jan-Uwe Ness, Kumiko Nobukawa, Masayoshi Nobukawa, Hirofumi Noda, Hirokazu Odaka, Shoji Ogawa, Anna Ogorzałek, Takashi Okajima, Naomi Ota, Stephane Paltani, Robert Petre, Paul Plucinsky, Frederick S. Porter, Katja Pottschmidt, Kosuke Sato, Toshiki Sato, Makoto Sawada, Hiromi Seta, Megumi Shidatsu, Aurora Simionescu, Randall Smith, Hiromasa Suzuki, Andrew Szymkowiak, Hiromitsu Takahashi, Mai Takeo, Toru Tamagawa, Keisuke Tamura, Takaaki Tanaka, Atsushi Tanimoto, Makoto Tashiro, Yukikatsu Terada, Yuichi Terashima, Yohko Tsuboi, Masahiro Tsujimoto, Hiroshi Tsunemi, Takeshi Tsuru, Aysegül Tümer, Hiroyuki Uchida, Nagomi Uchida, Yuusuke Uchida, Hideki Uchiyama, Shutaro Ueda, Yoshihiro Ueda, Shinichiro Uno, Jacco Vink, Shin Watanabe, Brian J. Williams, Satoshi Yamada, Shinya Yamada, Hiroya Yamaguchi, Kazutaka Yamaoka, Noriko Yamasaki, Makoto Yamauchi, Shigeo Yamauchi, Tahir Yaqoob, Tomokage Yoneyama, Tessei Yoshida, Mihoko Yukita, Irina Zhuravleva, Jean-Paul Breuer, Priyanka Chakraborty, Stefano Ettori, Andrew Fabian, Annie Heinrich, Marie Kondo, Julie HLavacek-Larrondo, Hannah McCall, Paul Nulsen, Tom Rose, Helen Russell, Arnab Sarkar, Evan Scannapieco, Kazunori Suda, Ming Sun, Prathamesh Tamhane, Nhut Truong, Norbert Werner, and Congyao Zhang. Xrism constraints on unidentified x-ray emission lines, including the 3.5 kev line, in the stacked spectrum of ten galaxy clusters, 2025. URL https://arxiv.org/abs/2510.24560.

J. S. Kaastra. On the use of C-stat in testing models for X-ray spectra. *Astronomy and Astrophysics*, 605:A51, 2017. doi: 10.1051/0004-6361/201629319.

J. Neyman. *Contribution to the theory of the chi-square test*, pages 139–273. University of California Press, Berkeley, 1949. ISBN 9780520327016. doi: doi:10.1525/9780520327016-030. URL https://doi.org/10.1525/9780520327016-030.

K. Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900. doi: 10.1080/14786440009463897.