1

Statistics

Aneta Siemiginowska

a chapter for "X-ray Astronomy Handbook"

October 2008

1.1 Introduction

Why do we need statistic? Wall and Jenkins (2003) give a good description of the scientific analysis and answer this question. Statistic allows us to take decisions in Science, evaluate observations, models, formulate questions and proceed forward with investigations. The statistic is needed at every step of our scientific analysis. Statistic is a quantity that summarized the data (mean, averages etc.) and astronomers cannot avoid statistics.

Here is a question asked by an X-ray astronomy school student:

"I wanted to know how many counts would be needed to get a good fit for a CIE plasma model with every parameter (save redshift) free. I discussed this topic with Randall Smith once who told me it took 500-1000 counts to get a decent fit, but I couldn't remember if this assumed that metallicity is fixed. Can someone get a good fit for metallicity with low counts (e.g. 500-1000)?"

However, what does it mean "a good fit" or 'a 'decent fit", and what does constitute "low counts" data? These expressions carry a certain meaning, but taken out of context are not precise enough. Are we asking whether a total number of counts in the spectrum is "low" or a number of counts per resolution element is "low"? In the question about a total number of counts 500-1000 seems "high", but it is considered "low" if we take a number of total counts and divide by a resolution element (for example there are 1024 independent detector channels in a *Chandra* ACIS spectrum). Also "high" is relative to a scientific question we pose and a type of the considered model. For example a simple power law model of a continuum could be well constrained by a *Chandra* spectrum with 500-1000 counts, but probably not a plasma model.

1.2 Probability Distributions

Probability is a quantity that describes a fraction of favorable events and it is a numerical measure of our belief. Laplace principle of indifference states that all events have equal probability. The Kolmogorov axioms give a base for the probability theory:

- Any random event, A, has a probability $0 < P(A) < 1$
- A probability of a sure event is equal to 1, $P(A) = 1$
- If A and B are exclusive events then P(AB)= P(A)+P(B)

The probability distribution is a function describing a probability of an event given a total number of events. If x is a continuous random variable then $f(x)$ is its probability density function or simply *probability distribution* when:

(1) $\text{prob}(a < x < b) = \int_a^b f(x)dx$
(2) $\int_{-\infty}^{\infty} f(x)dx = 1$
(3) $f(x)$ is a single non-negative number for all real x.

The two most widely used in X-rays probability distributions are Poisson and Gaussian distributions.

- Poisson probability:

$$\mathcal{P}(n;\mu) = \frac{e^{-\mu}\mu^n}{n!} \tag{1.1}$$

where μ is the mean of the distribution, e.g. a "count rate" - an average of number of photons received from a source per unit time (out of total number of emitted photons), $\mathcal{P}(n;\mu)$ describes then the probability of receiving n photons in a given exposure time.

- Gaussian probability, normal distribution:

$$\mathcal{N}(n;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[\frac{-(n-\mu)^2}{2\sigma^2}\right] \tag{1.2}$$

where μ is the mean of the distribution, σ is the variance, $\mathcal{N}(n;\mu,\sigma)$ describes the probability of receiving n photons.

1.3 Bayesian and Classical Methods

There are two main schools in Statistics: Bayesian and Classical. They differ in the basic assumptions, treatment of probabilities and philosophy. Classical or *frequentist* methods calculate a probability of the data given a model. Bayesian methods calculate a probability of the model given the data and use prior knowledge about the source, experiments etc. to assess the probability.

1.4 What do we really do in the analysis of X-ray data?

Section ??† describes some main concepts of spectral fitting in X-ray astronomy. The instruments' characteristics stored in calibration files are explicitly included in the X-ray data analysis. To evaluate the expected model counts a physical model of an X-ray emitting source, $m(E)$ is 'convolved' with the calibration data (most commonly defined by a redistribution matrix, RMF file, $r(E, i)$ and an effective area of the X-ray telescope, ARF file, $a(E)$).

$$M_i = \int r(E, i)\, a(E)\, m(E) dE \tag{1.3}$$

The predicted model counts M_i in a detector channel i are then compared to the observed counts. In order to understand how well the physical model describes the data we need *statistic* - a variable characterizing the property of the model in respect to the observed data. Typically χ^2 statistic is used, because it gives a 'goodness-of-fit' and because the X-ray data are usually binned (see below for methods in unbinned data). Note that χ^2 statistic is a random variable with the χ^2 distribution.

$$\chi^2 = \sum_i \frac{(D_i - M_i)^2}{\sigma_i^2} \tag{1.4}$$

where σ_i^2 is the variance. In Pearson's (1900) classical paper σ_i^2 is defined in respect to the model predicted counts M_i and it is assumed that the observed data come from the normal distribution. However, with a limited number of observed counts χ^2 applications can be restricted or even biased as we show in an example below. Variety of weighting of this basic formula has been defined to accommodate X-ray low counts observations (i.e. Gehrels, Churazov, Primini).

In the parameterized approach, the model parameters are varied during

† Keith on X-ray fitting

the 'fit' process (minimization, optimization) to obtain the best fit parameters, e.g. the parameters that minimize the χ^2 statistics and best describe the observations. These parameters give also the minimum difference between the model predicted counts and the observed number of counts.

The model parameter space is an N-dimensional space with a number of parameters defining its dimension. Note that for a different model expression the parameter space is different and could be completely disjoined from the other one. Note, however the importance of the *nested* models where a simpler model is nested within a more complex model

The best fit parameters define just one point in that model parameter space. Is this the only physical model describing the observed source? We can only answer this question in terms of probabilities. Because X-ray data are Poisson distributed there is a "Poisson" noise inherently present in the data. How this noise affects our determination of the best fit parameters? The uncertainty associated with the statistical noise propagates into the way we determine the parameters and the best fit parameters have a range of acceptable values. There are many physical models that can explain observations. The statistics helps to constrain the ranges of the acceptable parameter space. However, there is no single answer and we can only obtain the probability quantifying a chance of the physical model that can describe the observations.

1.5 Maximum Likelihood

Probability density function $f(X, \Theta)$, where X describes the X-ray data and Θ model parameters, is the Poisson probability because of the Poisson nature of the X-ray data. We want to estimate Θ. If $X_1, X_2..., X_N$ are the X-ray data, independent and drawn from the Poisson distribution $\mathcal{P}$ then the likelihood function is:

$$\mathcal{L}(X_1, X_2, ...X_N) = \mathcal{P}(X_1, X_2, ...X_N|\Theta) \tag{1.5}$$

$$= \mathcal{P}(X_1|\Theta)\mathcal{P}(X_2|\Theta)....\mathcal{P}(X_N|\Theta) \tag{1.6}$$

$$= \prod^{N} \mathcal{P}(X_i \mid \Theta) \tag{1.7}$$

Finding the maximum likelihood means finding the parameter Θ_0 that maximizes the likelihood function.

Below, I use an example to illustrate the maximum likelihood method in the X-ray spectral analysis. An X-ray spectrum is often modeled as a

power law function $m(E) = AE^{-\Gamma}$, where A is the normalization in photons cm^2 sec keV, E is photon energy and Γ is the photon index. Then the predicted number of counts M_i in i detector channel is defined by

$$M_i = \int r(E,i)\, a(E)\, AE^{-\Gamma} dE \tag{1.8}$$

This integral can be solved in XSPEC or *Sherpa* to calculate model predicted counts M_i for a given values of the parameters (A,Γ). Assuming *Chandra* ACIS calibration data and the power law parameters A = 0.001 photons/cm^2/sec/keV and $\Gamma = 2$ the model predicted counts in the ACIS channels $i = (10, 100, 200)$ are estimated to be $M_i = (10.7, 508.9, 75.5)$.

The Poisson likelihood for the observed counts X_i=(15,520,74) is then:

$$\mathcal{L}(X_i) = \prod^{N} \mathcal{P}(X_i \mid M_i(A,\Gamma)) \tag{1.9}$$

$$= \mathcal{P}(15 \mid 10.7)\mathcal{P}(520 \mid 508.9)\mathcal{P}(74 \mid 75.5) \tag{1.10}$$

$$= 0.116 \tag{1.11}$$

here we used the Incomplete Gamma function $\Gamma(X_i, M_i)$ to calculate individual Poisson probabilities given the observed data in these three channels.

An observed X-ray spectrum has many channels and the model needs to be evaluated in all the detector channels. Finding the maximum likelihood means finding the best set of parameters (A,Γ) that maximize the Poisson likelihood function for the assumed model. For an assumed distribution of parameters one can "sample" the parameters (A_j, Γ_j) many times, calculate the likelihood and choose the parameters that give the maximum value of $\mathcal{L}$. Many numerical methods have been developed to make such iterative process the most efficient. Van Dyk et al (2001) describe Monte Carlo algorithms specific to X-rays spectra.

In many applications a log-likelihood approach is used. Taking a natural log of the Poisson likelihood and using a simple algebra gives:

$$\ln \mathcal{L} = \sum_i (X_i \ln M_i - M_i - \ln X_i!) \tag{1.12}$$

Cash (1979) defined the C-statistics using this log-likelihood function and multiplying it by -2.

$$C = -2 \ln \mathcal{L} = \sum_i (X_i \ln M_i - M_i - \ln X_i!) \tag{1.13}$$

Because during the optimization process the difference between the values of the C-statistics is being used the term $\ln X_i!$ cancels out as it is independent on the parameters. Thus the C-statistics in the following form is minimized:

$$C = 2\sum_i^N (M_i - X_i \ln M_i) \tag{1.14}$$

Note that χ^2 statistics is also the maximum likelihood estimator. It is derived from the log-likelihood when the assumed underlying distribution is Gaussian (see Eq. XX for $\mathcal{N}(X_i|M_i)$):

$$\mathcal{L}(X_1, X_2, ...X_N) = \prod_i^N \mathcal{N}(X_i|M_i) \tag{1.15}$$

$$\ln \mathcal{L} = \sum_i^N \frac{(X_i - M_i)^2}{2\sigma_i^2} \tag{1.16}$$

1.6 Confidence Limits

Maximum likelihood estimators such as C-statistics or χ^2 are used to find the best parameters for the X-ray data. But how well do we know model parameters? This depends on the quality of the data, e.g. signal-to-noise, total number of counts etc. The data constrain the model parameters and after finding the best fit parameters one needs to estimate the confidence level for these parameters.

Avni (1976) defined the confidence region with the significance level α and a number degrees of freedom $\nu = N - 1 - j$ (N - number of channels, j - number of parameter)

$$\chi^2(\alpha) = \chi^2_{min} + \Delta(\nu, \alpha) \tag{1.17}$$

For example the significance of 90% corresponds to $\Delta = 2.71$ of the χ^2 difference above the minimum for one parameter. In practice calculating the parameter value that corresponds to this confidence level (or any other) could be numerically intensive, as it requires a parameter search around the best fit minimum.

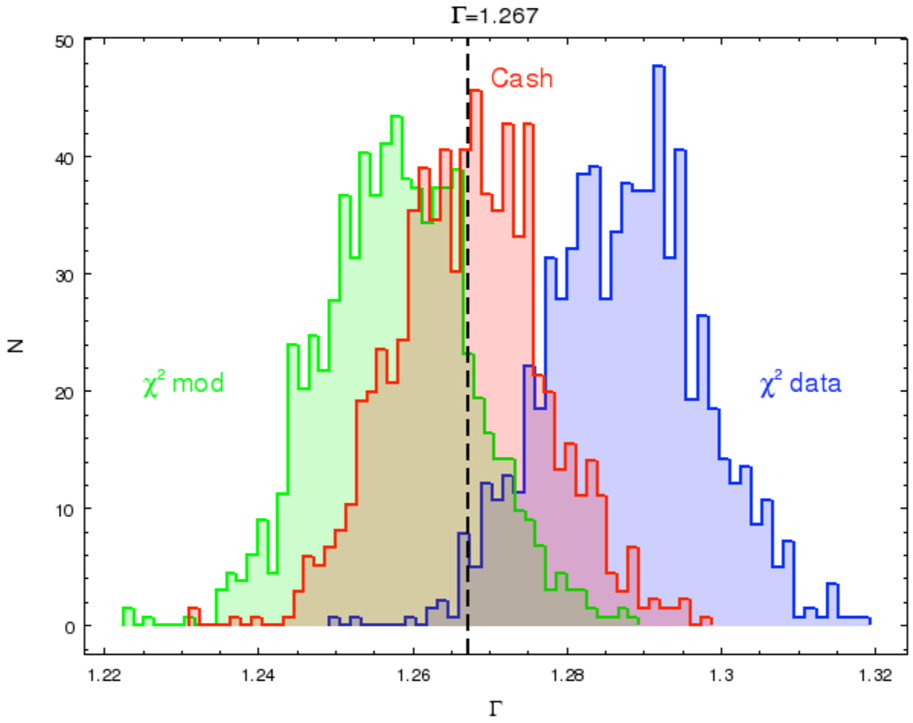

Fig. 1.1. Distributions of a photon index parame Γ obtained by fitting simulated X-ray spectra with 60000 counts and using the three different statistics: χ^2 with model variance, χ^2 with data variance and *Cash* statistics.

1.7 Statistical Issues

1.7.1 Bias

The χ^2 bias can affect the results of the X-ray spectral fitting and it can be demonstrated in a simple way. The described simulations can be done in *Sherpa* or XSPEC.

We assume an absorbed power law model with the sets of 3 parameters (an absorption column, a photon index, and a normalization) to simulate Chandra X-ray spectrum given the instrument calibration files (RMF/ARF) and the Poisson noise. The resulting simulated X-ray spectrum contains the model predicted counts with the Poisson noise. This spectrum is then fit with the absorbed power law model to get the best fit parameter values for NH, photon index and normalization.

We simulated 1000 spectra and fit each of them using different statistics: χ^2 with data variance, χ^2 with model variance and Cash/C-statistics. Fig-

ure 1.1 shows the distribution of the photon index parameter obtain from the fit of the high signal to noise spectra generated for the assumed simulated value of 1.267. The χ^2 bias is evident in this analysis. While C-statistics based on the Poisson likelihood behave well χ^2 with model variance underestimates the simulated value and χ^2 with data variance overestimates this parameter.

1.7.2 Background, Source Detection and Spectral Analysis.

X-ray data are collected as individual events. Both source and background counts are contained within the assumed source region matched to the point spread function of the detector and the background contribution needs to be accounted for in the analysis. In the high counts situation with low background contribution the background is typically subtracted and the error propagated according to the standard theory.

$$S = C - \frac{A_S t_S}{A_B t_B} B$$

where S is the source counts, C - total observed counts, B - observed background counts, A_S, A_B are source and background regions respectively and t_S, t_B are source and background exposure times.

In a low counts regime a treatment of the background could be tricky and subtracting the background can lead to negative(!) counts. This is where Poisson likelihood and therefore C-statistics should be used. However, if one uses the Poisson likelihood based statistics (C-statistics/Cash) the background cannot be subtracted, but source and background models have to be fit simultaneously.

1.7.3 Rebinning

Rebinning or grouping of low counts data is sometimes used in the X-ray analysis to increase number of counts in a new group. Rebinning is not the best solution as it leads to the loss of information. For example an emission line or absorption edge could disappear when the counts are grouped. However, if there is only a handful number of and as long as a simple continuum model is being applied the grouping of data is the fastest and easiest way to proceed with the analysis.

1.7.4 Systematic Errors, Calibration and Model Uncertainties.

In addition to the statistical errors there are usually systematic errors that are present in the observations. The most common way to include those errors is to use the error propagation formulas and add the statistical and systematic errors in quadrature. However, such approach does not account for a non-linear errors such as the ones in the *Chandra* calibration files. Drake et al (2006) shows how these errors affect the confidence of the best fit parameters and that the calibration errors dominate in the error budget in a very high signal to noise observations. Then the constrain on the best fit parameters is limited by the size of these errors and as in Drake et al approach the simulations should be used to estimate the uncertainties on the parameters.

Another class of uncertainties is related to physical models. For example in the plasma emission models the line transitions are defined with certain uncertainties. Such uncertainties should be included in the analysis, however the models typically do not state their uncertainties and the standard fitting packages do not take into account such errors.

Avni, Y. 1976, Ap.J., 210, 642

Cash, W. 1979, Ap.J., 228, 939

Drake, J. J., Ratzlaff, P., Kashyap, V., Edgar, R., Izem, R., Jerius, D., Siemiginowska, A., & Vikhlinin, A. 2006, Proceedings of the SPIE, 6270,

Protassov, R., van Dyk, D. A., Connors, A., Kashyap, V. L., & Siemiginowska, A. 2002, Ap.J., 571, 545

Wall, J. V., & Jenkins, C. R. 2003, Princeton Series in Astrophysics, van Dyk et al 2001

van Dyk, D. A., Connors, A., Kashyap, V. L., & Siemiginowska, A. 2001, Ap.J., 548, 224

Web pages:

`http://hea-www.harvard.edu/AstroStat/` - it contains software link, bibliography and "Statistics Jargon for Astronomers"

`http://groundtruth.info/AstroStat/slog/` - Astrostatistics Blog