Statistics for High-Energy Astronomy

Vinay Kashyap
CHASC/CXC/CfA
Outline
Statistics is more than just means and standard deviations

1. Photon Counts and the Poisson distribution
2. Gaussian
   1. Likelihood and $\chi^2$
   2. Poisson vs Gaussian
   3. Error propagation
3. Fitting
   1. Best fit
   2. goodness of fit
   3. cstat
4. CIAO/Sherpa
1. Counts

- ACIS and HRC are photon counting detectors. Events are recorded as they arrive, usually sloooowly
- What does this imply?
Light curve of steady source HZ 43 binned at 1 sec

Notice: asymmetry, scatter around the mean
Poisson likelihood

Notice:
- asymmetry
- +ve integers
- distribution
1. Poisson Likelihood

- $p(k|\theta) = \frac{1}{k!} \theta^k e^{-\theta}$
- The probability of seeing $k$ events when $\theta$ are expected
- e.g., $\theta = \text{count rate} \times \text{time interval} \equiv r \cdot \Delta t$
- mean, $\mu = \sum_k k \cdot p(k|\theta) = \theta$
- variance, $\sigma^2 = \overline{k}^2 - \overline{k}^2 = \theta$
$p(k|\theta)$ for different $\theta$
2. Gaussian

- A Gaussian distribution is convenient
- Symmetric, ubiquitous (because of the Central Limit Theorem), easy to handle uncertainties
- \[ N(x; \mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
2.1 likelihood

- Probability of obtaining observed data given the model
  \[ p(x|\theta,\sigma) \, dx = N(x; \theta,\sigma^2) \, dx \]

- When you have several data points
  \[
  p\left(\{x_i\}|\theta_i\right) = (2\pi)^{-N/2} \, \prod_k \sigma_k^{-1} \, e^{-\left(x_k-\mu_k\right)^2/2\sigma_k^2} \\
  = (2\pi)^{-N/2} \left(\prod_k \sigma_k^{-1}\right) \exp[-\sum_k (x_k-\mu_k)^2/2\sigma_k^2]
  \]

- log Likelihood \( \propto -\sum_k (x_k-\mu_k)^2 / 2\sigma_k^2 \)
2.2 Poisson $\rightarrow$ Gaussian

- Variance of Poisson is $= \text{mean}$
- As $\theta \uparrow$
  \[
  \text{Pois}(k \mid \theta) \rightarrow N(k; \theta, (\sqrt{\theta})^2)
  \]
- Convenient!
Poisson(k|θ) → N(θ, θ) as θ↑
2.3 Error Propagation

- How to propagate uncertainty from one stage to another — if \( g = f(x) \), and \( \sigma_x \) is known, what is \( \sigma_g = f(\sigma_x) \)?

- Simple case: if everything is distributed as a Gaussian, and has well-defined means and standard deviations,

\[
g = g(a_i) \Rightarrow \sigma^2_g = \sum_i (\partial g / \partial a_i)^2 \sigma^2_{a_i}
\]
2.3 Error Propagation

\[ \begin{align*}
g &= C \cdot a \\
\Rightarrow \sigma_g &= C \cdot \sigma_a \\
\text{uncertainties scale} \\
g &= \ln(a) \\
\Rightarrow \sigma_g &= \frac{\sigma_a}{a} \\
\text{converts to fractional error} \\
g &= \frac{1}{a} \\
\Rightarrow \sigma_g &= \left(\frac{1}{a^2}\right) \sigma_a \equiv \left(\frac{g}{a}\right) \sigma_a \\
\Rightarrow \frac{\sigma_g}{g} &= \frac{\sigma_a}{a} \\
\text{fractional errors stay as they are} \\
g &= a + b \\
\Rightarrow \sigma_g^2 &= \sigma_a^2 + \sigma_b^2 \\
\text{errors square-add} \\
\end{align*} \]
3.1 Best-fit

- The best fit is one that maximizes the likelihood
- e.g., linear regression — \( y_i = \alpha + \beta x_i + \epsilon \)

solve by finding extremum of log likelihood

\[
\ln L \propto \sum_k (y_k - \alpha - \beta x_k)^2
\]

\[
\frac{\partial \ln L}{\partial \alpha} = \frac{\partial \ln L}{\partial \beta} = 0
\]

\( \Rightarrow \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \) and \( \hat{\beta} = \text{Cov}(x,y)/\text{Var}(x) \)

Notice notation:

\( \bar{\text{bar}} \) and \( \hat{\text{hat}} \) to indicate sample averages and best-fit values

\( \Gamma \varepsilon \kappa \) letters for model quantities, Roman for data quantities
3.2 Goodness-of-fit

❖ How good is the model as a description of your data?

❖ How can you tell when you do have a “good” fit?

❖ Recall the log Likelihood — its -ve is called the chi-square,
  
  \[ \chi^2 = \sum_k (x_k - \mu_k)^2 / 2\sigma_k^2 \]

❖ and its distribution describes the probability of getting \((x_k, y_k)\) to match “similarly” for several bins

❖ When observed \(\chi^2 \sim \text{dof} \pm \sqrt{2\sqrt{\text{dof}}}\), model is doing excellent job of matching the data. The farther it is from this range, the less likely it is that the model is a good description of the data

❖ But always use your judgement, because this is a probabilistic rule!

❖ Watch out for how \(\sigma^2\) is defined (model variance is best)
3.3 cstat

- Poisson log Likelihood: $-\ln \Gamma (k+1) + k \cdot \ln \theta - \theta$
- Apply Stirling’s approximation, $\ln \Gamma (k+1)=k \ln k-k$
  - $\ln \text{PoissonLikelihood} = k \cdot (\ln \theta - \ln k) + (k - \theta)$
- Just as $\chi^2$ is $-2\ln\text{Likelihood}$,
  - $\text{cstat} = 2 \sum_i (M_i - D_i + D_i \cdot (\ln D_i - \ln M_i))$
  - where $D_i$ are observed counts, and $M_i$ are model predicted counts in bin $i$
- unbiased for low counts than $\chi^2$, asymptotically $\chi^2$, rudimentary goodness-of-fit exists (Kaastra 2017, A&A 605, A51)
- Watch out: only asymptotically $\chi^2$, not quite the Poisson likelihood, 0s are thrown away, background must be explicitly modeled
4. Statistical Tools in CIAO/Sherpa

- **fit**: non-linear minimization fitting
- **conf/covar/projection/int_proj/reg_proj**: uncertainty intervals and error bars
- **sample_flux**: parametric bootstrap to get model fluxes
- **get_draws**: MCMC engine pyBLoCXS (Bayesian Low-Counts X-ray Spectral analysis; van Dyk et al. 2001, ApJ 548, 224)
- **calc_ftest**: model comparison via F-test
- **celldetect/wavdetect/vtpdetect/mkvtpbkg**: source detection in images
- the python interpreter in Sherpa gives access to python libraries, and can be used to call upon packages and libraries in R, which are written by statisticians for statisticians