What is AstroStatistics for?

Obtain estimates and uncertainties on quantities useful for astrophysical inference,
while taking into account instrument sensitivities, statistical fluctuations, and circumstances of observation,
and avoid the pitfalls of making incorrect inferences.
Outline

❖ Properties of X-ray data
❖ Making peace with jargon
❖ Statistical concepts
  1. Error Propagation
  2. Bootstrap
  3. Distributions
  4. $p$-values and Hypothesis Tests
  5. Bayesian analysis
  6. MCMC
  7. Model Fitting
  8. Things to be afraid of
❖ Tools at our disposal
X-ray data is not like optical data

- A list of events \{x,y,t,E\} → marked Poisson process
- Calibration
- Poisson likelihood:
  \[
  \Pr(\text{k counts when intensity is } \theta) = \frac{\theta^k e^{-\theta}}{\Gamma(k+1)}
  \]
- Gaussian approximation is widely used: \( \mu = \sigma^2 = k \)
Jargon

- Probability, $p(\cdot)$ — *frequency of occurrence or degree of belief*
- Likelihood, $L \equiv p(D \mid \theta)$ — probability of obtaining observed data assuming a particular model
- Fitting
  - $\chi^2$ — measure of closeness, also goodness of fit $\equiv -2 \ln(\text{Gaussian likelihood})$
  - $cstat/cash \equiv -2 \ln(\text{Poisson Likelihood})$
- $p$-values / Null Hypothesis Significance Testing
- Tests of dissimilarity: Kolmogorov-Smirnoff, F-test
1. Error Propagation

- How to propagate the uncertainty from one stage to another
- Simple case: assume everything is distributed as a Gaussian, and has well-defined means and standard deviations
- $g = g(a_i)$
  
  $\Rightarrow \sigma^2(g) = \sum_i (\frac{\partial g}{\partial a_i})^2 \sigma^2(a_i)$
1. square adding

\[ g = C \cdot a \]
\[ \rightarrow \sigma_g = C \cdot \sigma_a \]

\[ g = 1/a \]
\[ \rightarrow \sigma_g/g = \sigma_a/a \]

\[ g = \ln(a) \]
\[ \rightarrow \sigma_g = \sigma_a/a \]

\[ g = a + b \]
\[ \rightarrow \sigma^2_g = \sigma^2_a + \sigma^2_b \]
2. Bootstrap

❖ How to estimate the uncertainty within almost any set of measurements

❖ Steps:
  ❖ 1: construct summary statistic
  ❖ 2: extract random sample of same size from original dataset and recompute summary statistic from Step 1
  ❖ 3: repeat Step 2 a large number of times and compute mean and variance of summary statistic

❖ Quick and easy

❖ Accurate, if sample in hand is a good representation of population (e.g., don’t try this with power-laws)
3. Distributions

- **Binomial** — one or the other, with probability $\rho$
  
  $k$ of one out of a total of $N$, $p(k|N,\rho) = \binom{N}{k} \rho^k (1-\rho)^{N-k}$

- **Poisson** — events occur randomly
  
  $p(k|\theta) = \frac{1}{k!} \theta^k e^{-\theta}$

- **Gaussian** (aka Normal) — all summary statistics that have a sufficiently large sample
  
  $f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$ 

- **Gamma** — continuous variable conjugate to Poisson
  
  $p(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} e^{-\beta x}$, $x \geq 0$, $\alpha \geq 0$, $\beta \geq 0$; Poisson for $\beta=1$ and $\alpha=k+1$

- **$\chi^2$** — measure of similarity and distance between samples
  
  $p(\chi^2|n) = \Gamma\left(n/2\right) \left(\frac{2}{n/2}\right)! \left(\chi^2\right)^{(n/2-1)/2} \exp\left[-\chi^2/2\right] \propto \left(\chi^2\right)^{(n/2-1)/2} \exp\left[-\chi^2/2\right] \equiv \Gamma\text{amma}(\chi^2;n/2,-1/2)$
4. $p$-values and Hypothesis Tests

- Compare distributions by setting up competing hypotheses
- Null hypothesis $H_0$ is that both samples are drawn from the same distribution
- Calculate a statistic from the data and compare to the expected distribution of the statistic. If calculated value exceeds a critical threshold, reject the null hypothesis.
5. Basics of Bayesian Analysis

- Axioms
- Bayes’ Theorem
- Example — hardness ratio
5.1 Axioms of Probability Theory

1. \( p(A \text{ or not } A) = p(A) + p(\text{not } A) = 1 \)

2. \( p(A \text{ and } B) = p(B) \ p(A \text{ given } B) \equiv p(A) \ p(B \text{ given } A) \)
5.1 Axioms of Probability Theory

1. \( p(A \text{ or not } A) = p(A) + p(\text{not } A) = 1 \)
   \[
p(A + \bar{A}) = p(A) + p(\bar{A}) = 1
   \]

2. \( p(A \text{ and } B) = p(B) \ p(A \text{ given } B) \equiv p(A) \ p(B \text{ given } A) \)
   \[
p(A \ B) = p(B) \ p(A \mid B) \equiv p(A) \ p(B \mid A)
   \]
5.2 Bayes’ Theorem

\[ p(AB|C) = p(A|BC) \cdot p(B|C) = p(B|AC) \cdot p(A|C) \]

\[ p(A|BC) = \frac{p(B|AC) \cdot p(A|C)}{p(B|C)} \]

\[ p(\theta|D I) = \frac{p(D|\theta I) \cdot p(\theta|I)}{p(D|I)} \]
5.2 prior, likelihood, posterior

\[ p(\theta|D I) = p(D|\theta I) \cdot p(\theta|I) / p(D|I) \]

*a priori* distribution: \( p(\theta|I) \)

likelihood: \( p(D|\theta I) \)

*a posteriori* distribution: \( p(\theta|D I) \)
5.3 Example: Hardness Ratios

- Measure counts in Soft ($S$: lower energies, longer wavelengths; e.g., 1/2-2 keV) and Hard ($H$: higher energies, shorter wavelengths; e.g., 2-8 keV) passbands
- \( HR := \frac{(H-S)}{(H+S)} \), \( R := \frac{S}{H} \), \( C := \log(S/H) \)
- Problem: Gaussian error propagation fails for low counts, or when \( HR \) is close to ±1, or because ratios are not distributed as Gaussians
- Need to compute \( p(hr|H,S) \), \( p(r|H,S) \), \( p(c|H,S) \)
5.3 Example: Hardness Ratios

- For all the details, see Park et al. 2006 (ApJ 652, 610)
- \( p(H \mid h) \) and \( p(S \mid s) \) are Poisson likelihoods, \( p(s) \) and \( p(h) \) are usually chosen as Gamma priors
- \( p(h,s \mid H,S) \propto p(H,S \mid h,s) \ p(h,s) = p(H \mid h) \ p(S \mid s) \ p(h) \ p(s) \)
  \[ hr = (h-s)/(h+s), \quad w = (h+s) \Rightarrow h,s = (1 \pm hr) \cdot w / 2 \]
  \[ J(h,s;hr,w) = \left| \frac{\partial(h,s)}{\partial(hr,w)} \right| = w / 2 \]
  \[ p(h,s \mid ...) \ \text{d}h \ \text{d}s = p((1+hr) \cdot w / 2, (1-hr) \cdot w / 2 \mid ...) \ J(h,s;hr,w) \ \text{d}hr \ \text{d}w \]
  \[ p(hr) \ \text{d}hr = \int \text{d}w \ p(hr,w) \]
6. Markov Chain Monte Carlo

❖ What is it?
  ❖ A method to quickly explore high-dimensional parameter spaces and obtain representative measures of parameter values and uncertainties

❖ Why do it?
  ❖ Robust, insensitive to starting conditions, easy to code

❖ How does it work?
  ❖ Compute the likelihood for given parameter values, get a new, randomly drawn value, and compare the new likelihood to the old one
  ❖ If it improves the likelihood, accept the new value and repeat the cycle
  ❖ If it does not improve the likelihood, accept with a probability equal to the ratio, else reject and get a new value
7. Fitting

- Non-linear metric minimization
  - $\chi^2$ (any of several varieties) — $\sum_i (D_i - M_i)^2/\sigma_i^2$
  - fit is good if $\chi^2$/dof $\sim 1 \pm \sqrt{2}$/dof
- cstat — $2 \sum_i (M_i - D_i + D_i \cdot (\ln D_i - \ln M_i))$
  - asymptotically $\chi^2$ — otherwise use parametric bootstrap to determine goodness of fit
7.1 Model Comparison

- Model comparison
  - use F-test iff simpler ("null") model is fully contained within complex ("alternate") model
    - simulate fake datasets from best-fit parameters of null model
    - fit with both null and alternate model
    - compute distributions of ratios of the best-fit statistic and compare against the ratio for actual data
    - if ratio from observed data is far in the tail of the simulated distribution, then it is unlikely that the null model is a good descriptor of the data
8. Danger Danger

❖ asymptotic validity — be aware of the assumptions made to get easy analytical results (e.g., \( p \)-value for F-test, \( \chi^2 \) as measure of goodness)

❖ convergence, stopping rules, effect of priors — always do sensitivity tests

❖ overfitting — to avoid fitting fluctuations in the data, balance bias against variance

❖ \( p \)-values — measure of how far in the tail of a distribution the current observation is, not a proof of the validity of an alternative hypothesis, nor of the falsity of the null hypothesis

❖ Type I, Type II, Type S, Type M errors — false positive, false negatives, sign errors on weak effects, Eddington Bias
8. Types of Bias

- **Type I** — false positives, when you claim a detection over a background because of a fluctuation above some threshold.

- **Type II** — false negatives, when you fail to detect an event because its response fell below the detection threshold.

- **Type M** — an incorrect estimation of the *size* of the effect because large fluctuations are preferentially detected (cf. Eddington bias).

- **Type S** — an incorrect estimation of the *sign* of a weak effect because of fluctuations in the wrong direction.
8. Type I and Type II Errors

Type I error
\[ \alpha \leq 0.10 \]

Type II error
\[ 1 - \beta \]

\[ \beta = 0.70 \]
8. Type S Error

\[ f(x | \lambda) \]

- \( \lambda = 90 \)
- \( \lambda = 99 \)
- \( \lambda = 101 \)
- \( \lambda = 110 \)
8. Eddington Bias

Statistical Tools in CIAO/Sherpa

- fit/conf/projection: non-linear minimization fitting and uncertainty intervals
- calc_ftest: model comparison via F-test
- celldetect/wavdetect/vtpdetect: source detection in images
- the python interpreter in Sherpa gives access to python libraries, and can be used to call upon libraries in R, which are written by statisticians for statisticians