Brandon C. Kelly, Harvard-Smithsonian Center for Astrophysics

Statistics II: More Advanced Issues

Outline

- Previous lecture focused on classical results
- Now we focus on more advanced techniques:
  - Data analysis with low numbers of counts
  - Introduction to Monte Carlo Methods
  - Feature Detection and Upper or Lower Limits
  - Bayesian Methods
- Many of these techniques are the subject of ongoing research

Review of MLE for Gaussian Data

- Recall that for Gaussian Data:

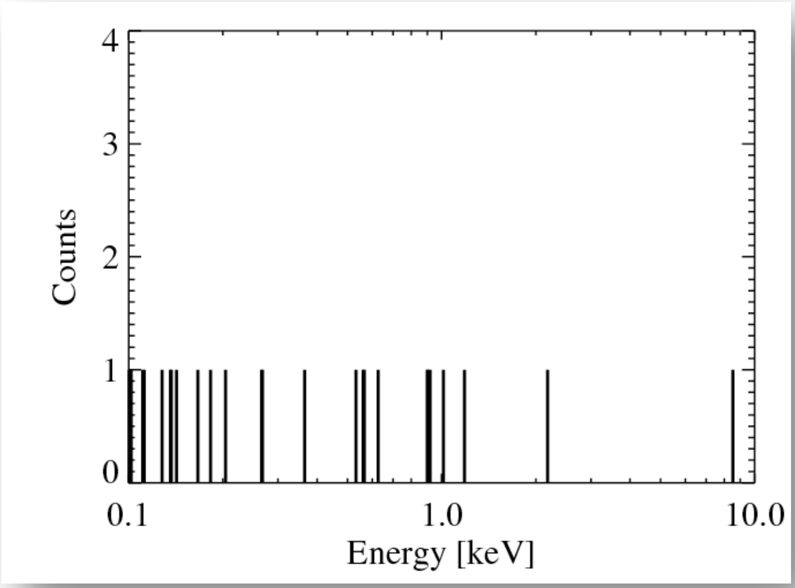
$$\ln p(y \mid \theta) = -\tfrac{1}{2}\chi^2 + Const$$

$$E(\theta_{\chi^2}) \xrightarrow[n\to\infty]{} \text{True } \theta, \quad Var(\theta_{\chi^2}) \xrightarrow[n\to\infty]{} 2\left(\left. \frac{d^2\chi^2}{d\theta^2} \right|_{\theta_{\chi^2}} \right)^{-1}$$

- For X-ray data, if the data are binned (grouped) with at least ~10 counts / bin, then assuming the Gaussian limit is usually good enough

Low Counts

- What if we don't have enough counts to bin the data, e.g., 25 photons?
- Instead, we just have a list of photon arrival times and energies
- Need to work directly with the Poisson likelihood!



Review of the Poisson Distribution

- Gives the probability of detecting k photons when the average is λ:

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

- In the limit of large λ, the Poisson distribution converges to a Gaussian distribution with mean and variance equal to λ

Using the Poisson Distribution for MLE

- We model the count rate in an energy bin as $\lambda(E,\theta)$, where θ parameterizes our model (e.g., $\lambda(E,\theta)$ could be a power-law)

- If the i^{th} energy bin has k_i counts, the likelihood is

$$p(k_i \mid \theta) = (k_i!)^{-1} \lambda(E_i,\theta)^{k_i} e^{-\lambda(E_i,\theta)}$$

- So, the likelihood function of the data is:

$$p(k_1,\ldots,k_n \mid \theta) = \prod_{i=1}^{n} p(k_i \mid \theta) = \prod_{i=1}^{n} (k_i!)^{-1} \lambda(E_i,\theta)^{k_i} e^{-\lambda(E_i,\theta)}$$

Example: Power-law Spectrum

- Assume a power-law spectrum:

$$\lambda(E_i, k_0, \Gamma) = k_0 E_i^{-\Gamma}$$

- Plug this into the likelihood function:

$$p(k_1, \ldots, k_n \mid \theta) = \prod_{i=1}^{n} (k_i!)^{-1} k_0^{\,k_i} E_i^{-k_i \Gamma} \exp(-k_0 E_i^{-\Gamma})$$

- Estimate k_o and Γ by maximizing the log-likelihood

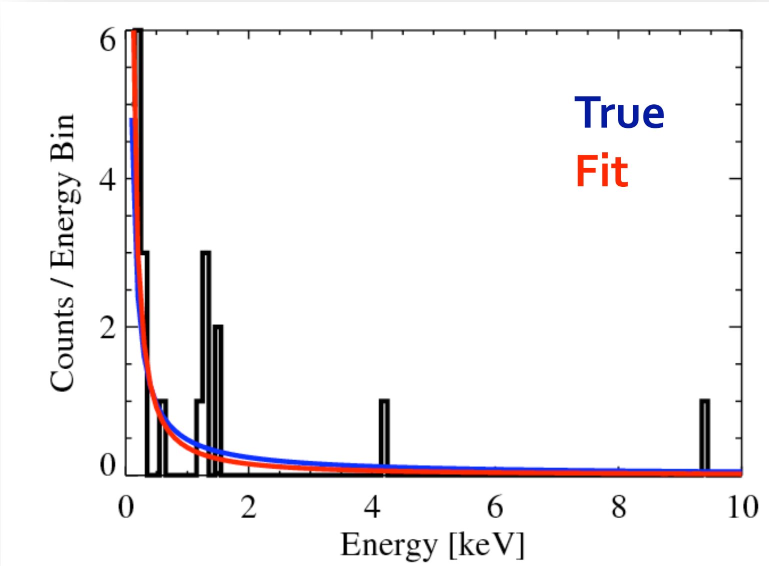
Estimating Uncertainties for low counts

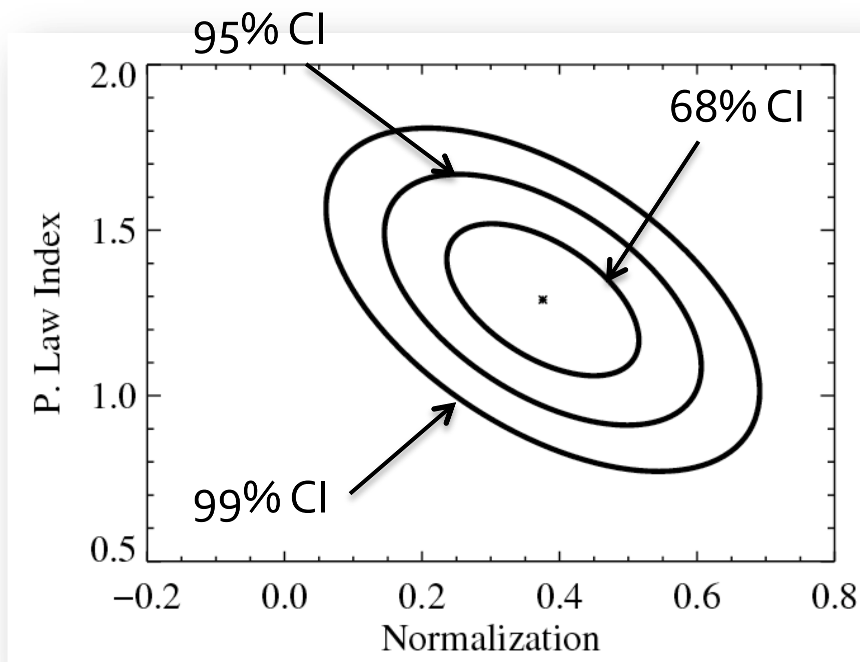
- Can use the asymptotic results for MLE:
  - The MLE is Gaussian with mean and variance:

$$E(\theta_{MLE}) \xrightarrow[n\to\infty]{} \text{True } \theta, \quad Var(\theta_{MLE}) \xrightarrow[n\to\infty]{} -\left(\frac{d^2}{d\theta^2} \ln p(y \mid \theta) \Big|_{\theta_{MLE}} \right)^{-1}$$

- But, we are in the low count regime, so be careful with using results applicable for large sample sizes
- More accurate results obtained using monte carlo or Bayesian methods

Power-law Example





Approximate!!

The Monte Carlo Method

- In general, analytical formula for probability distributions can’t be derived
- But, we have computers, so we do the next best thing: simulate!
- This is a very versatile technique known as monte carlo, and is not limited to statistics

Example: Estimating a distribution

- Consider a toy statistical model for variability of an accretion disk:

$$N \sim Poisson(\lambda)$$

$$\log E_1,\ldots,\log E_N \sim Gaussian(\mu,\sigma^2)$$

$$E_{Tot} = \sum_{i=1}^{N} E_i$$

E_{tot} is the total energy released by the flares in a given time interval

- Use simulation to estimate the distribution of E_{Tot}

Can use MC to estimate the expected value of a model

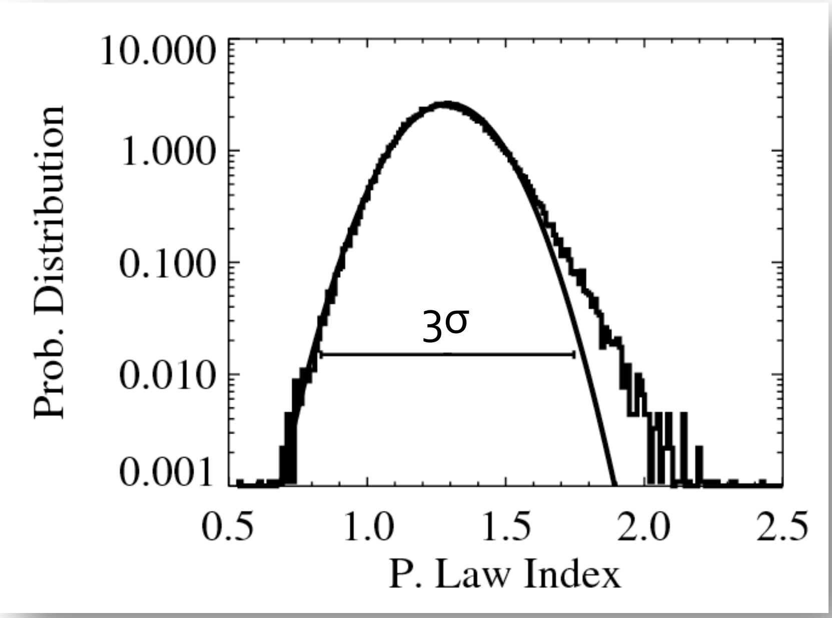
- Sometimes we can't derive the expected value of our data based on our statistical model, so no analytical expression for χ^2
- Can use MC simulations to estimate the average (expected) value of the data for our model
- Good example of this: estimating power spectra of X-ray lightcurves of accreting black holes (Uttley, McHardy, & Papadakis 2002, see later lectures on timing analysis)

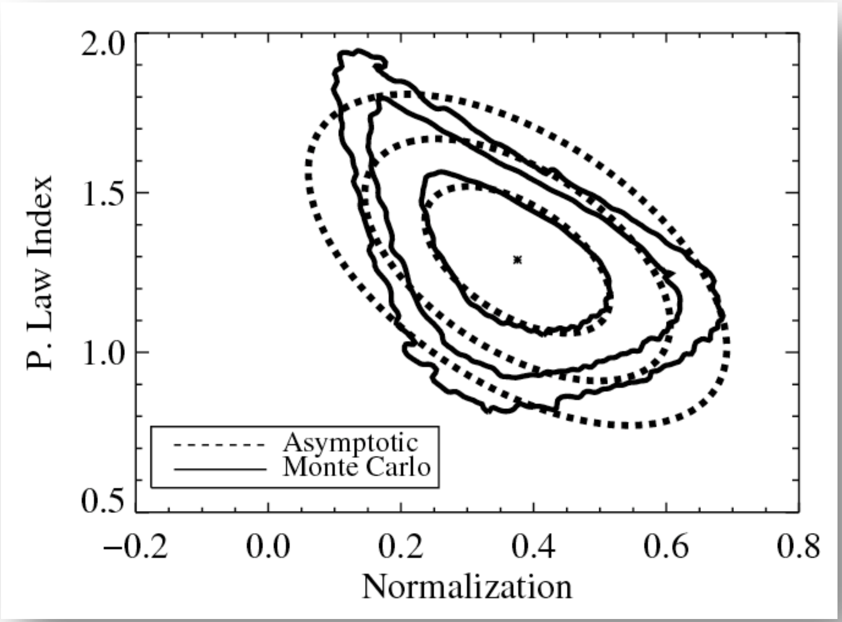
Monte Carlo Estimate of Errors

- Use MC to simulate the data generation/ collection/analysis process :
  - Assume the best fit value of your model parameters
  - Simulate an observed data set based on your best fit parameters
  - Fit the simulated data using the same method you used to get your best fit value
  - Repeat
- Result is an **estimate** of the sampling distribution of your parameters

Power-law Example Revisited

- Fit power-law to ~25 counts using MLE
- Simulated 10^5 random X-ray event lists assuming the MLE values, and refit them using MLE:





Error Propagation

- How do we estimate the uncertainty on a parameter (e.g., flux) we derive from the parameters we estimated from the data?
- Classical approach: do error propagation:

$$\sigma^2_{f(\theta)} \approx \sum_{i=1}^{p} Var(\theta_i) \frac{\partial^2 f}{\partial \theta_i^2} + \sum_{i=1}^{p} \sum_{j=1, j \neq i}^{p} Cov(\theta_i, \theta_j) \frac{\partial^2 f}{\partial \theta_i \partial \theta_j}$$

- Better and easier method: just use output from MC simulation

Using MC for Hypothesis Testing

- Recall that we can use the likelihood ratio or F test to assess the evidence for additional components
- Unfortunately, often the distribution of the LR or F statistic under the null hypothesis is not known
- So, we use MC techniques to estimate the sampling distribution of the LR under the null hypothesis (i.e., assuming the simpler model)

Example: Spectral Line revisited

- Recall that before we used the F-test to assess the evidence for a spectral line over a power-law continuum
- We could do this because we fixed the line centroid and width, and allowed it to be an emission or absorption line
- But what happens when we test for an emission line, and/or treat the center and width as free parameters?
- **Can't use F-test (or LRT)!**

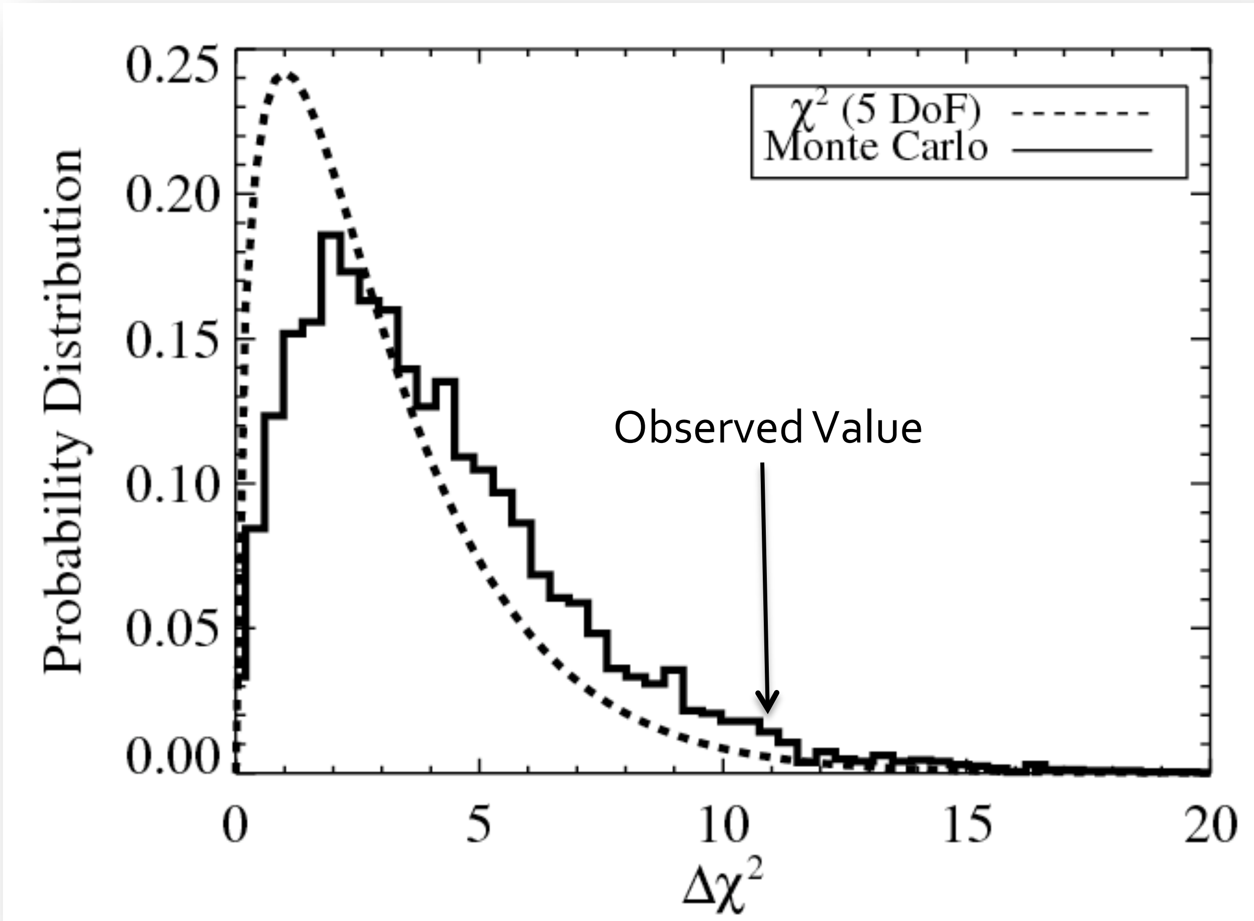
Monte Carlo test for an emission line

- We can't use the F-test since we don't know the sampling distribution under the null hypothesis that are spectrum is a power-law:
  - The models are not nested: the line centroid and width are undefined under the simpler power-law model
  - Since we are testing for an emission line, the line flux is on the boundary of the parameter space under the simpler model
- So, instead, use simulation to estimate the sampling distribution (e.g., see Protassov et al. 2002)

MC Algorithm for testing for a emission line

- For Gaussian data, calculate the MLE under both models. Compute the difference in χ^2
- Simulate a new data set assuming the best-fit values for the simpler power-law only model
- Fit the simulated data for both models. Compute the difference in χ^2
- Repeat numerous times.
- Compare the actual observed value of χ^2 with the simulated distribution of χ^2. Is it consistent?

Result for Power-law + Spectral Line



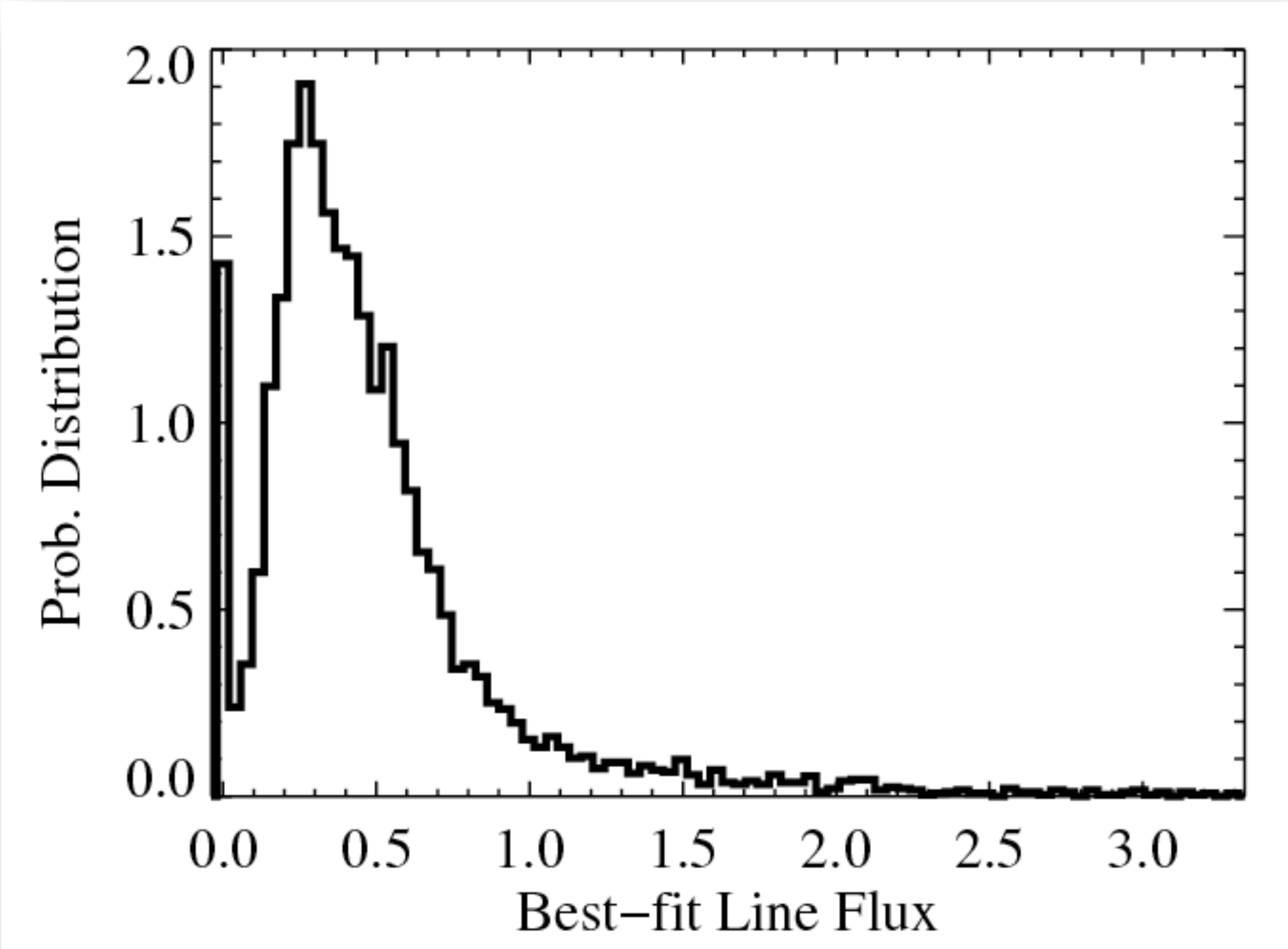
If Testing for Multiple Lines…

- If the number of additional components depends on your inspection of the residuals, you need to simulate this
- If possible, do not be an additional source of randomness in your analysis, you are very hard to simulate!

Setting an Upper Limit on the Line Emission

- Suppose we can't rule out the null hypothesis of no emission line, can we at least estimate an upper limit on the line flux?
- Can get an approximate estimate using simulation:
  - Simulate numerous data sets based on best-fit of model which includes emission line component
  - Fit each of these simulated data set
  - Find the value where, say, 99% of the values for the simulated data sets are below
- Definition still a matter of debate, see (Kashyap et al. 2010, arXiv:1006.4334)

Result for Power-law + Emission Line



Bayesian Methods

- In contrast to previous methods, Bayesian methods do not focus on obtaining an estimate of a quantity
- Instead, they want to directly calculate the probability distribution of the quantity, given the data
- So, the focus is more on how we can constrain the quantity of interest, and not in obtaining a single 'best-fit'
- In many ways, this is more similar to what we do as astronomers (businesses, on the other hand, usually want an estimate, since they need to decide a single course of action)

The posterior distribution

- The goal of all Bayesian methods is the 'posterior' probability distribution, p(θ|y)
- Recall that p(x,y) = p(x|y)p(y)
- Therefore, the posterior is

$$p(\theta \mid y) \propto p(y \mid \theta)\, p(\theta)$$

- The term p(θ) is the 'prior' distribution. It is up to the researcher to choose a prior which quantifies our knowledge of θ, regardless of the data

Bayesians vs Frequentists (Classical Methods)

FREQUENTISTS (CLASSICAL)	BAYESIANS
▪ View the `true' parameter value as fixed, and the data random ▪ Focus on obtaining an estimate of a quantity ▪ Uncertainties are derived based on estimating the sampling distribution of the estimator ▪ Objective (sort of…)	▪ View the true parameter as random (unknown), and the data as fixed (known) ▪ Focus on calculating the probability distribution of the parameter, given the data ▪ Subjective (but not as much as frequentists claim) ▪ Can sometimes be more computationally difficult

Why use Bayesian Methods?

- As we've seen, for most realistic data analysis problems, it is very difficult to find the sampling distribution of an estimator
  - Consequently, confidence intervals are hard to estimate
- Bayesian methods avoid this: they directly calculate the probability distribution of the parameters, given the data
  - So, they give results which are exact and usually easier to interpret (and thus more trustworthy!)

The posterior probability and χ^2

- Recall that for Gaussian data, the likelihood is

$$p(y \mid \theta) \propto \exp\{-\tfrac{1}{2}\chi^2(y,\theta)\}$$

- So, the posterior probability distribution is

$$\boxed{p(\theta \mid y) \propto p(\theta)\, p(y \mid \theta) \propto p(\theta) \exp\{-\tfrac{1}{2}\chi^2(y,\theta)\}}$$

- If we use a uniform prior on our parameters, the Bayesian approach seeks to map out the χ^2 space for Gaussian data

Low Counts Example, Part III

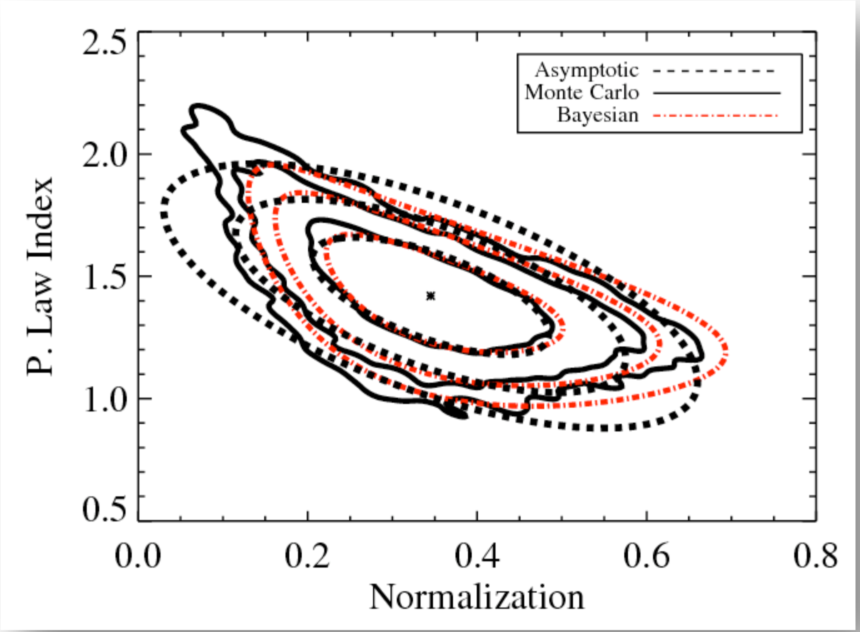
- Recall that for our power-law fit to ~ 25 counts, we used the Poisson likelihood function:

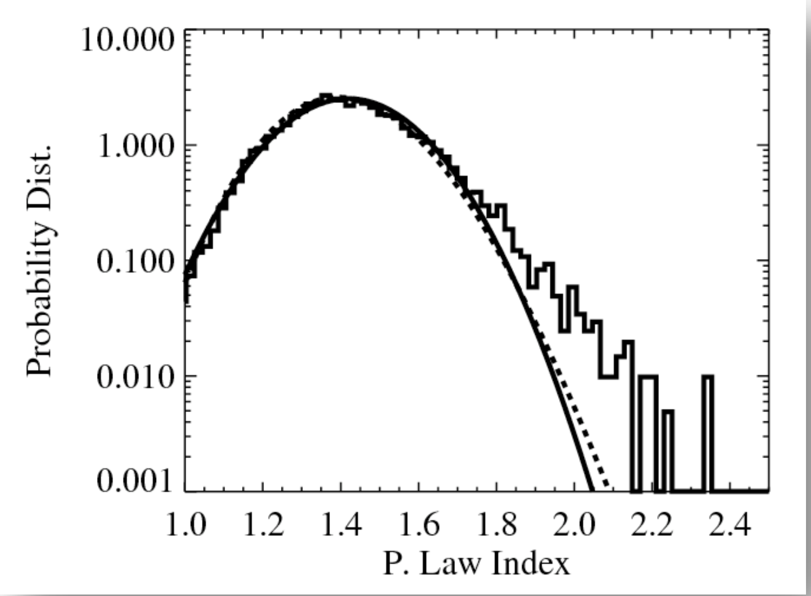
$$p(k_1,\ldots,k_n \mid \theta) = \prod_{i=1}^{n} (k_i!)^{-1} k_0^{k_i} E_i^{-k_i\Gamma} \exp(-k_0 E_i^{-\Gamma})$$

- We will assume a uniform prior on k_0 and Γ
- The posterior for k_0 and Γ is then:

$$p(\theta \mid k_1,\ldots,k_n) \propto \prod_{i=1}^{n} (k_i!)^{-1} k_0^{k_i} E_i^{-k_i\Gamma} \exp(-k_0 E_i^{-\Gamma})$$

Example: Comparison of Classical, Monte Carlo, and Bayesian results





Bayesian testing for additional components in a spectrum

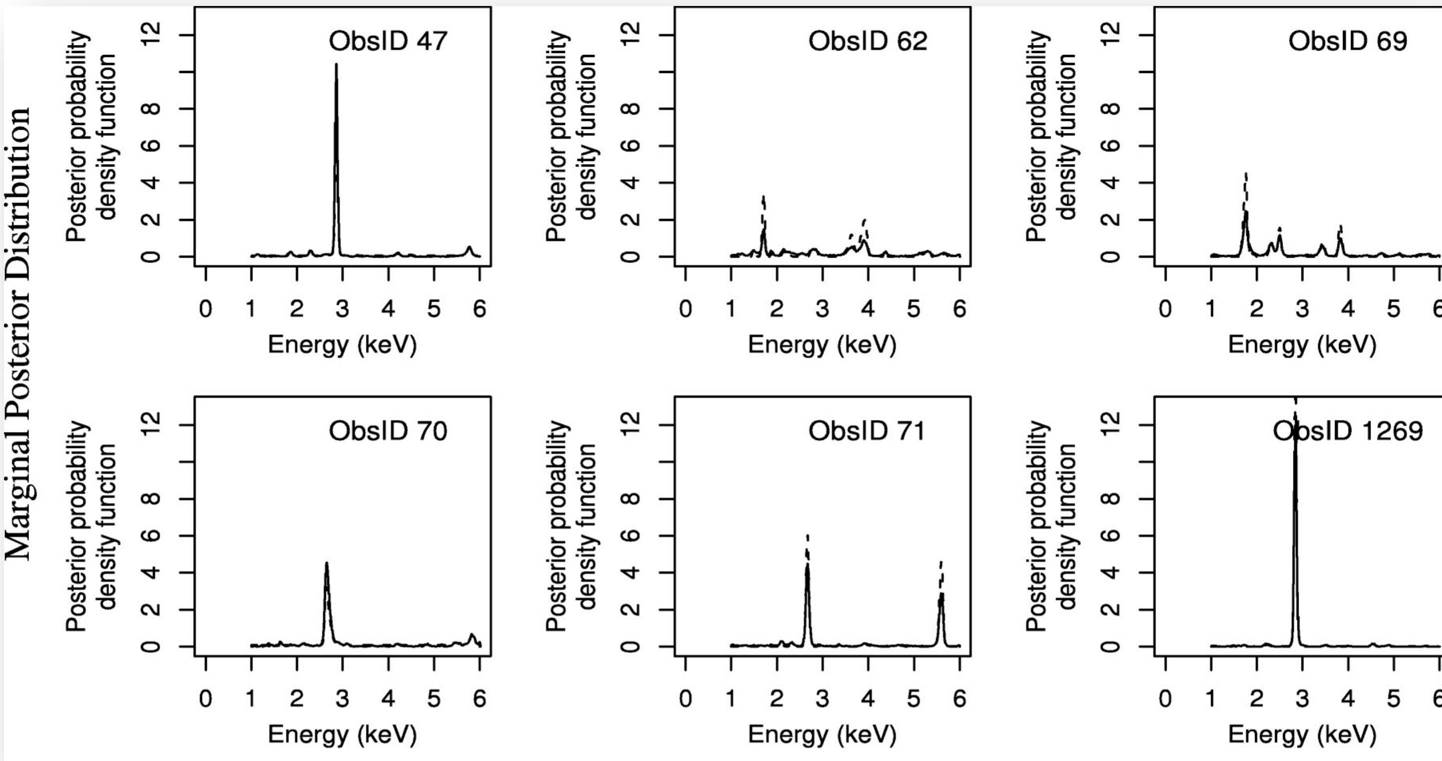
- Different methods have been proposed, but for semi-nested models I recommend this simple and straightforward approach:
  - Compute the marginal posterior for the additional parameters of the more complicated model
  - Find the probability that they have the value needed to reduce the more complicated model to the simpler one
  - If this probability is very low, the data prefer the more complicated model

Example procedure for testing for an emission line

- Suppose our model for the X-ray spectrum is a power-law + an emission line
- To assess the evidence for an emission line:
  - Compute the marginal probability distribution of the line flux, given the observed data
  - Is there a negligible amount of probability near zero line flux?
    - If yes, then you have found evidence for an emission line
    - If no, then you can't say if there's an emission line or not, but you can place an upper limit on the line flux

Example: Bayesian analysis of emission line on real data

Marginal posterior for narrow emission line location in Chandra grating data of quasar PG 1634+706



Park, van Dyke, & Siemiginowska (2008)

This is really just the tip of the iceberg

- Currently there is much active research regarding astrophysical data analysis, and much work needs to be done!
- Some outstanding problems and current issues include
  - How to test for additional components
  - How deal to incorporate calibration errors, or more generally systematics
  - What do we mean by an upper limit
  - How do we characterize and estimate the variability of an object?
- These issues only deal with what we can measure from the detector, but there's much more after we get our measurements!
  - How do we fit a straight line to data contaminated by measurement errors and containing upper limits?
  - How do we estimate the distribution of quantities subject to data truncation (e.g., luminosity functions from flux-limited samples?)
  - How can we calculate distributions of quantities which we estimate (or derive) from our data, such as black hole mass, accretion rate, and spin?