Hierarchical Modeling of Astronomical Images and Uncertainty in Truncated Data Sets

Brandon Kelly

Harvard-Smithsonian Center for Astrophysics

Overview

- Deriving physical parameters from astronomical images with two-level error structure
- Understanding the correlations among the physical parameters and their spatial distribution
- Deriving the distribution of astrophysical parameters from a truncated data set
- Uncertainty in the selection probability creates unstable likelihood functions and/or posterior distributions, how to account for this?

Dust Images: Scientific Motivation

Cold Clouds of gas and dust become unstable, collapse

↓

Eventually, some regions become dense and hot enough to start fusion

↓

Star is formed

But there's a lot to this process we don't understand…

Image Courtesy:
NASA, STScI,
N. Evans

Far-IR Images of Starless Cores

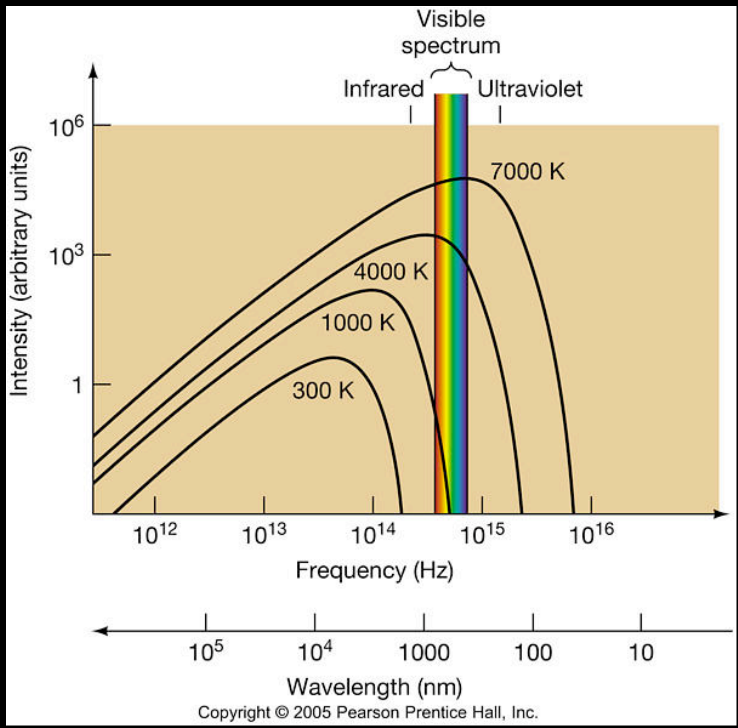
- Provide insight into physical properties of starless cores
  - E.g., models predict Temperature decreases toward core of cloud
- Hopefully will lead to a better understanding of star formation
- Most of emission in images due to cold dust, so analysis of images will lead to a better understanding of astrophysical dust as well
  - E.g., does dust opacity depend on its temperature?

Modeling Dust Emission

- Model dust brightness as a modified ‘black-body’:

$$f(\nu) = C\nu^{\beta} B_{\nu}(T),\ B_{\nu}(T) \propto \nu^{3}\left[\exp\left(\frac{h\nu}{kT}\right) - 1\right]^{-1}$$

- Parameters are the dust temperature, T, and the power-law modification index, β
- β is expected to depend on the properties of the dust particles, describes their ‘opacity’



Mode and Normalization of BB increase as temperature increases

Model for Measurement Process of Images

- Each pixel is assumed to have an additive normally-distributed measurement error with known standard deviation
- Each of J images is also assumed to have a multiplicative log-normally distributed calibration error with known standard deviation. This error is the same for all n pixels in the image

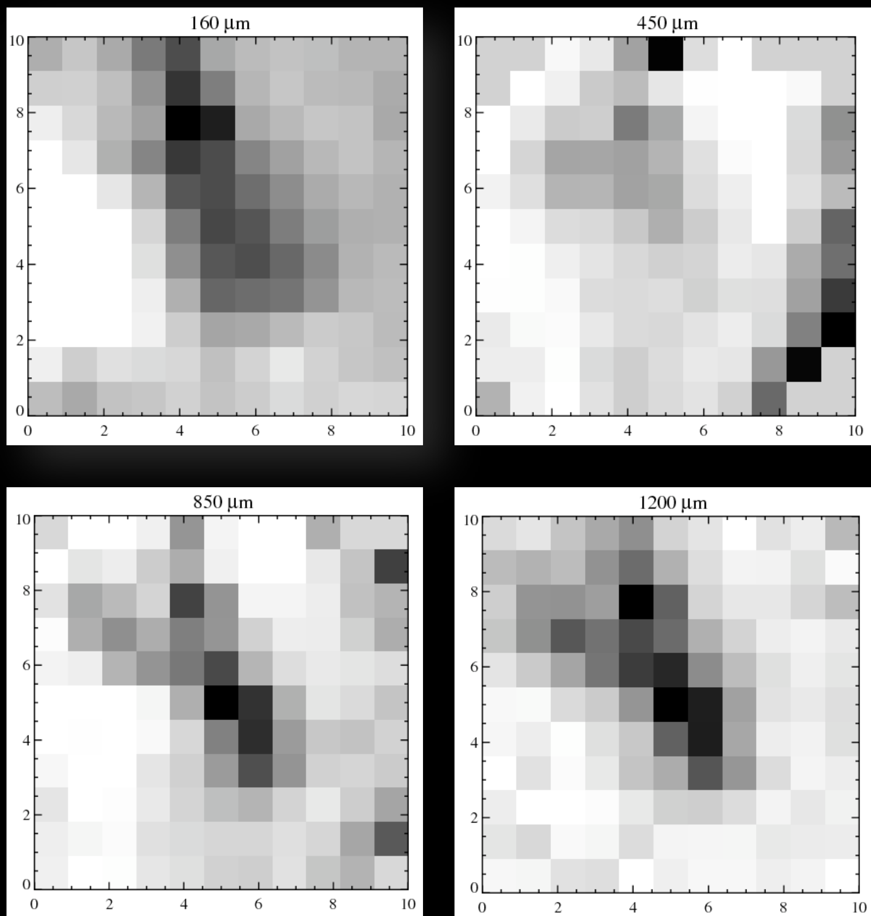
$$\hat{f}_{ij} = \delta_j (f_{ij} + \varepsilon_{ij}), \; f_{ij} = C_i \nu_j^{\beta_i} B(\nu_j, T_i)$$

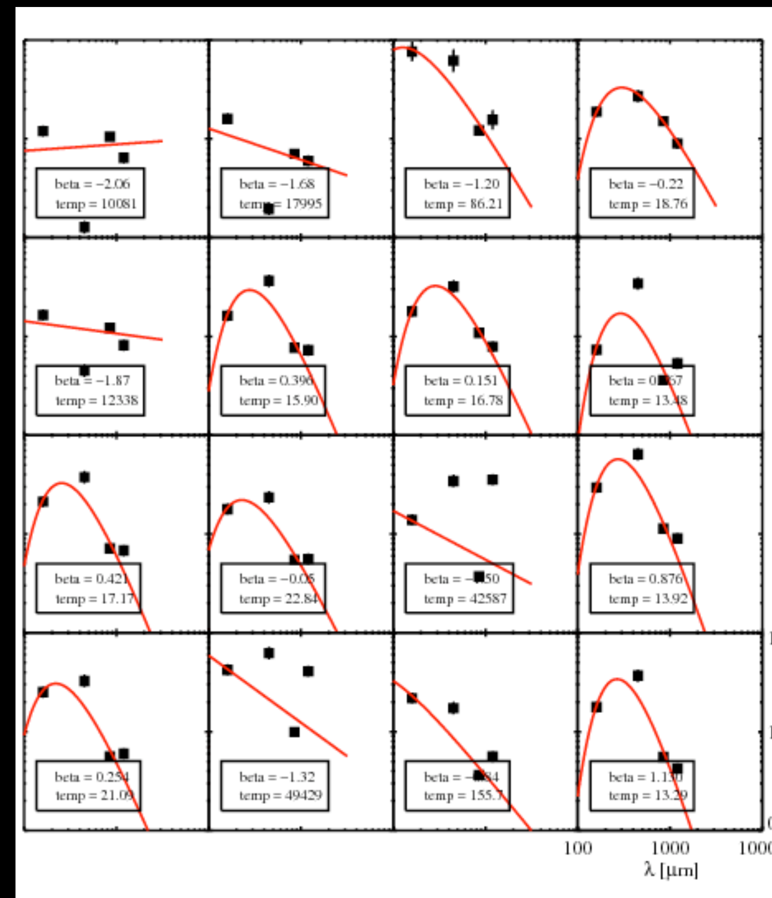
$$\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$$

$$\log \delta_j \sim N(0, \tau_j^2)$$

$$i = 1, \ldots, n \quad and \quad j = 1, \ldots, J$$

Example: Starless Core TMC-1C

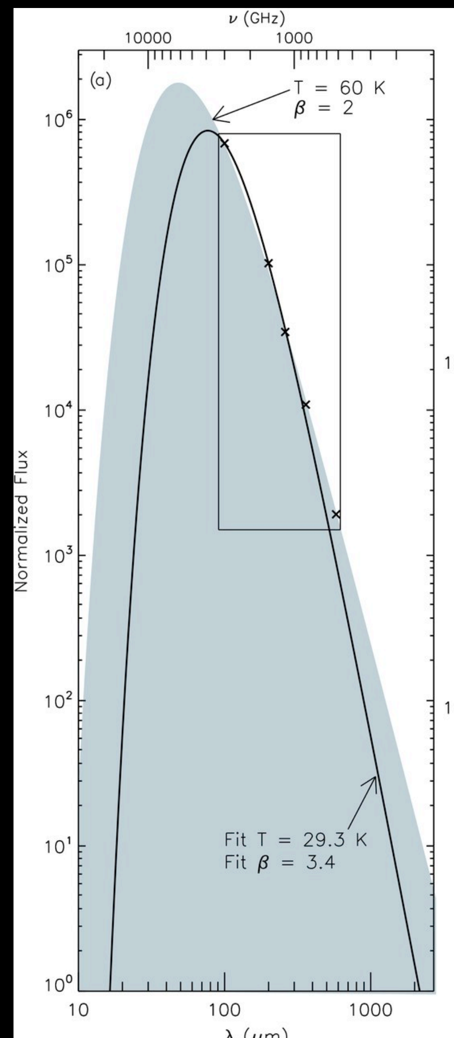




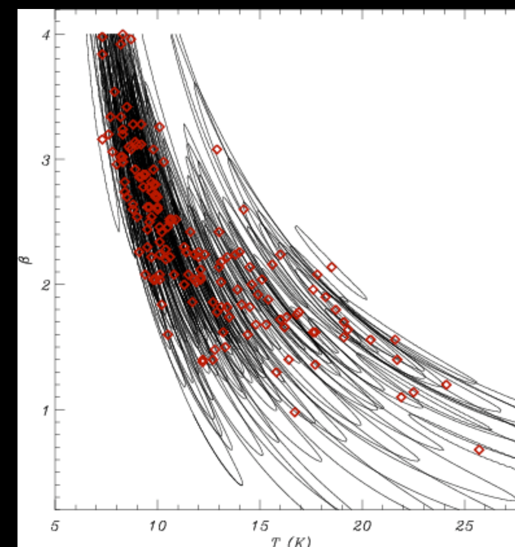
Best fit parameters typically estimated using least-squares, highly uncertain

Degeneracy between T and β

- Errors in estimated temperature and power-law index can be large and highly anti-correlated
- Biases the inferred relationship between T and β, leads to spurious anti-correlation



Shetty et al., 2009



Désert et al., 2008

Hierarchical Model for Dust Maps

- Use Bayesian hierarchical model to simultaneously model the joint distribution of T and β with the observed brightness values
- Accounts for uncertainties at all levels, pools information from all pixels
- Parameters are C, T, and β for each pixel, δ for each image, μ_C, V_C, V_β, θ, and ψ

$$\log C_i \sim N(\mu_C, V_C)$$

$$T \sim p(T \mid \psi)$$

$$\beta_i \mid T_i \sim N(\mu(\theta), V_\beta)$$

$$\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$$

$$\log \delta_j \sim N(0, \tau_j^2)$$

$$\hat{f}_{ij} = \delta_j [f_{ij}(C_i, T_i, \beta_i) + \varepsilon_{ij}]$$

Use uniform prior for hyperparameters over some regional range

Posterior Distribution

$$p(C,T,\beta,\delta,\theta,V_\beta,\psi,\mu_C,V_C \mid \hat{f}) \propto$$

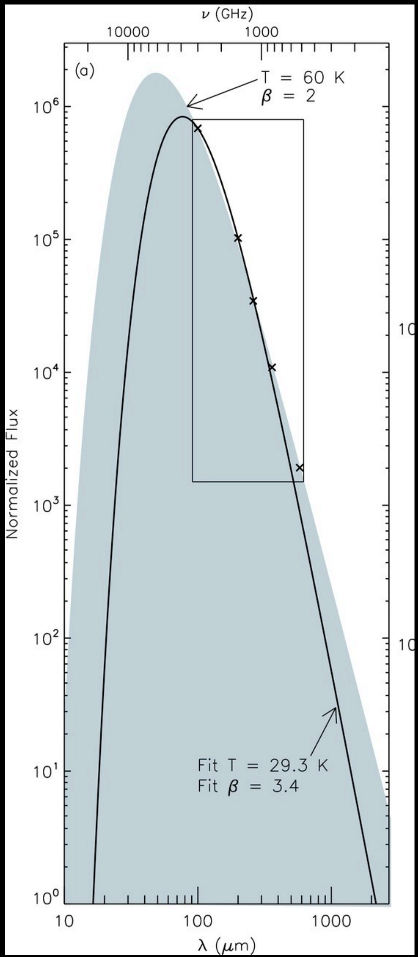
$$\prod_{i=1}^{n} N(\log C_i \mid \mu_C, V_C) N(\beta_i \mid \mu(\theta), V_\beta) p(T_i \mid \psi)$$

$$\times \prod_{j=1}^{J} N(\hat{f}_{ij} \mid \delta_j f_j(C_i,\beta_i,T_i), \delta_j^2 \sigma_{ij}^2) N(\log \delta_j \mid 0, \tau_j^2)$$

- Want to account for spatial correlation in Temperature prior
- Might be able to integrate our C, since it's a nuisance parameter
- Also, likelihood has the form of a normal variance-mean mixture, so might be able to integrate out δ

MCMC Sampler: Strong Correlations Among Parameters

- Many parameters, so use MCMC (MHA + Gibbs) to sample from the posterior
- Strong correlations exist among C,T,β, and δ, so convergence is *very* slow
- Need to come up with more clever transition kernels, or integrate out nuisance parameters (e.g., δ)



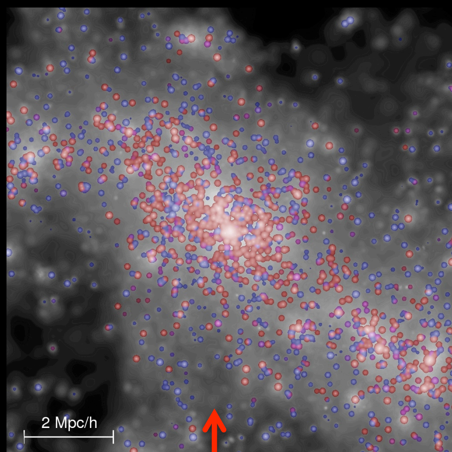
Part 2: Truncated Astronomical Surveys

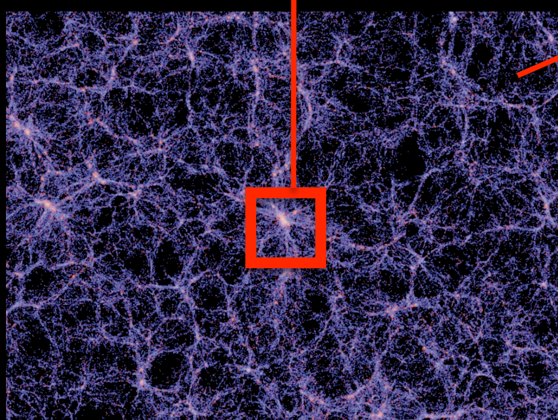
- Much astronomical research focuses on understanding populations of objects
- For extragalactic (outside of the Milky Way) sources, we also able to study how the population evolves
- The distribution of the parameters observed (e.g., luminosity) and derived is studied and compared with astrophysical models
- But, selection function (probability of a source ending up in your survey) depends on luminosity and distance (and therefore cosmic age).

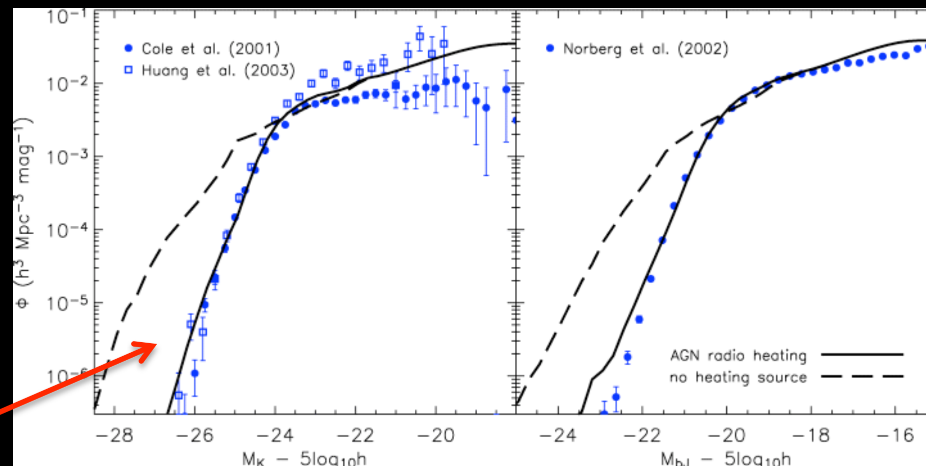
Some representative science questions

- How does the rate at which galaxies form stars depend on the galaxy's properties, and how does it change over time?
- When did supermassive black hole grow? How long were they actively growing for, on average? How do they affect the evolution of their host galaxies?
- What role did stars play in reionizing the early universe? How about active black holes?

Example: Cosmological Simulations







Comparison of actual distribution of galaxy Luminosity with that predicted from a cosmological simulation (+ additional assumptions), Croton et al.(2006)

Millennium Simulation, Springel et al. (VIRGO Consortium, 2005), and Max-Planck-Institute for Astrophysics

Likelihood Function for Truncated Data

- Denote observed data as x, distribution of x as $p(x|\Theta)$, total number of sources as N, total number of observed sources as n
- Introduce indicator variable I, where I = 1 if a source is included in the survey, and I = 0 if a source is missed. Selection function is $p(I=1|x)$ and assumed known.
- The complete data likelihood is:

$$p(I,x|\theta,N) = \binom{N}{n} \prod_{\substack{\text{Missing}\\ \text{Data}}} p(I_i = 0|x_i)\, p(x_i|\theta) \prod_{\substack{\text{Included}\\ \text{Data}}} p(I_j = 1|x_j)\, p(x_j|\theta)$$

Posterior Distribution

- The observed data likelihood is found by integrating over the missing data:

$$p(I,x_{obs}\mid\theta,N)=\binom{N}{n}\left[1-p(I=1\mid\theta)\right]^{N-n}\prod_{\substack{\text{Included}\\\text{Data}}}p(I_j=1\mid x_j)p(x_j\mid\theta),$$

$$p(I=1\mid\theta)=\int p(I=1\mid x)p(x\mid\theta)dx$$

- Assuming a prior p(N,Θ) = p(Θ) / N (i.e., uniform on log N), the marginal posterior of Θ is (e.g., Gelman et al., 2004)

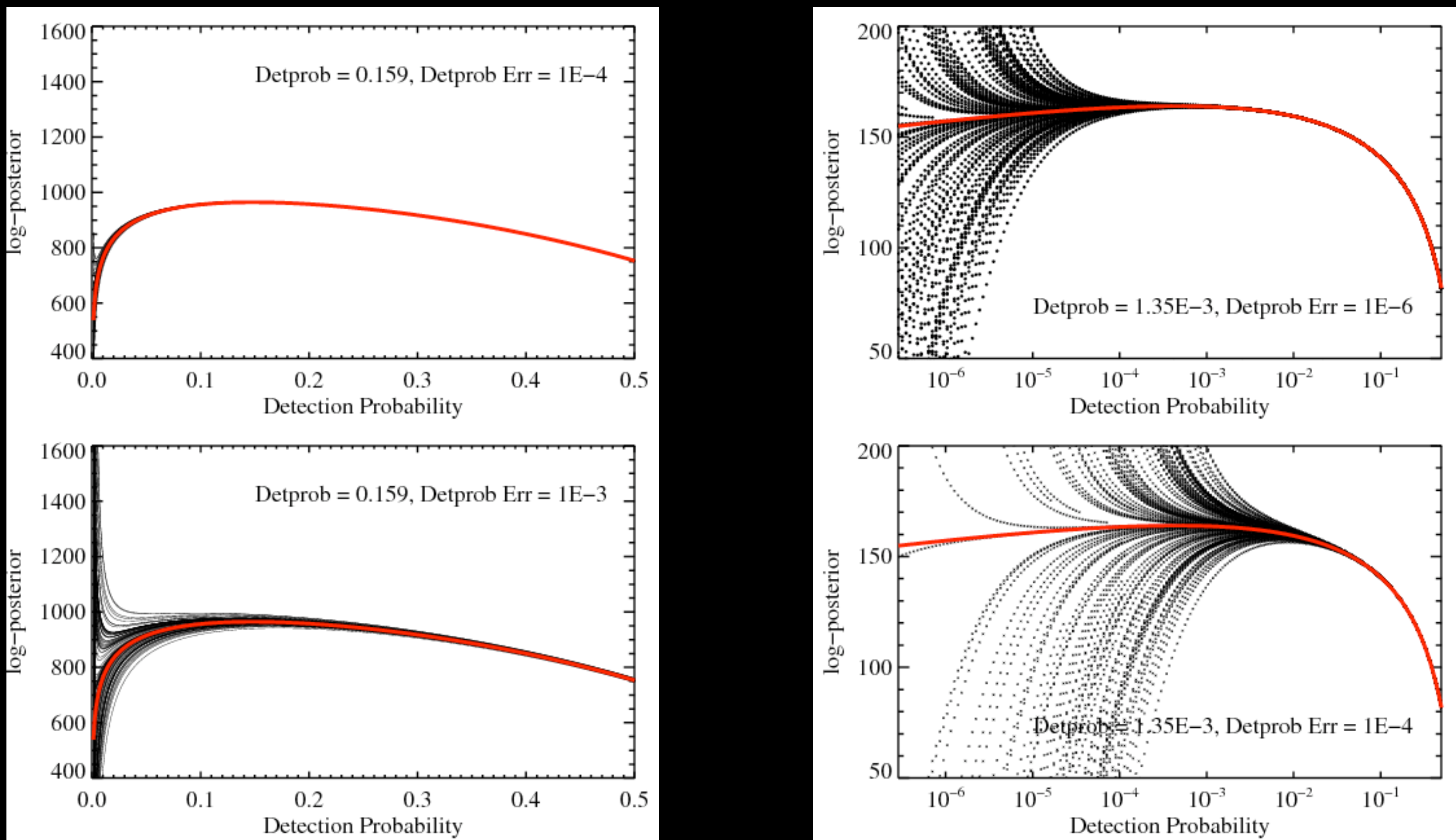
$$p(\theta\mid x_{obs},I)\propto\left[p(I=1\mid\theta)\right]^{-n}\prod_{i=1}^{n}p(x_i\mid\theta)$$

- The conditional posterior p(N|Θ,I) is a negative binomial distribution with parameters n and p(I=1|Θ)

Uncertainty in the Selection Function

- All this assumes we know $p(I=1|\Theta)$ (i.e., the selection function), or that we can calculate the integral without error
- But what happens when there is some uncertainty in the selection function?
- Alternatively, what happens when the integral cannot be calculated without error, as in stochastic integration?

Posterior/Likelihood highly unstable to errors in selection function



Posterior also unstable to stochastic integration

- Simulated a data set from standard normal with zero mean having N = 1000 sources. The mean, μ, and N are assumed unknown.
- Only kept those above x = 1
- Estimated p(I=1|μ) by simulating different numbers of data points from N(μ,1) and only keeping those above x=1.
- MCMC routine was used to obtain random draws from the posterior
- Unstable, how can we account for this????