Reporting point estimates manufactures false confidence. Better to report probabilistic parameter estimates via full posterior distributions.

Hierarchical Bayes:

Probabilistic modeling of populations or groups.

Why Hierarchical Bayesian Modeling?

- The uncertainty unique to each observation can be propagated into the estimation of the population parameters.
- **Selection effects** can be modeled and incorporated into the inference. (e.g., groups of observations of the same objects from different instruments)
- Information about each object in the population is used to make inferences about one another. Objects "pool and muster strength" (a.k.a. "**shrinkage**").
- Can handle large measurement uncertainty, censored data (upper limits), and **missing data**.
- Naturally handles **over-fitting**

What kinds of things might we want to learn by using HBM in Astronomy?

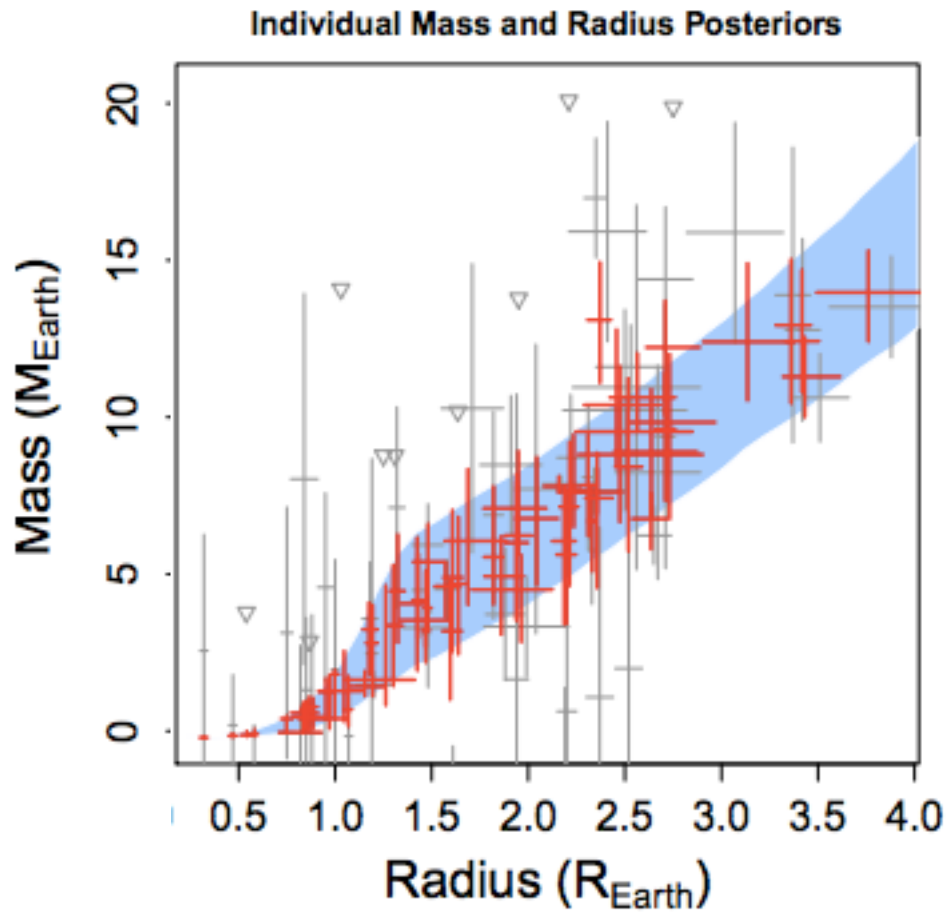
- If the distribution of interest is correlated with other properties it may shed light on interesting physics.
- What is the typical scale of the population?
- Is there evidence for multiple populations?
- Do we have many uncertain measurements of the same object?

A few examples in the literature:

- Evidence for two Hot Jupiter Formation Paths **Nelson et al**. 2017
- *HBM simulated RV exoplanet eccentricities* **Hogg et al**. 2010
- *Most 1.6 Earth-radius Planets are Not Rocky,* **Rogers** 2015
- *HBM for exoplanet mass-radius relationship* **Wolfgang, Rogers, & Ford** 2016
- *The eccentricity distribution of Hot Jupiters from Kepler* **Shabram et al.** 2016
- *Multi-level model for luminosity distribution of gamma ray bursts* **Loredo & Wasserman**, 1998
- *HBM for Supernovae* **Mandel** et al. 2009
- *HBM for galaxy clusters* **Andreon & Hurn**, 2010
- *HBM for Milky Way satellites* **Martinez** 2015

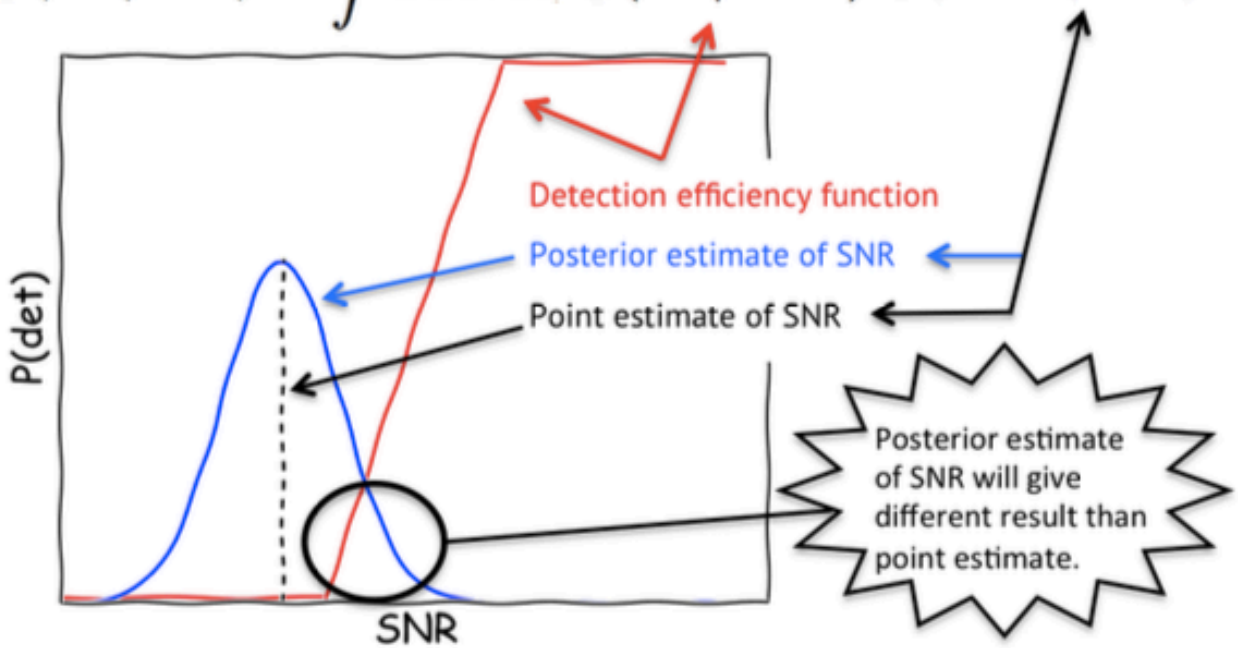
HBM in action: The mass-radius relation

- Individual estimates are pulled toward the population mean and overall RMS error is lower“ – example in **Wolfgang et al. 2015**, Figure 4, arXiv: 1504.07557.



HBM in action: *Kepler* Occurrence Rates

$$p(det|data) \propto \int d\,SNR \;\; p(det|SNR) \;\; p(SNR|data)$$



Depicting Hierarchical Bayesian Models

- Mathematical equations
- Graphical models
- Pseudo-code

~ distributed as
<- defined as (data transformations)

e.g., hyperparameters for a mixture model:

$$\boldsymbol{f}_a \sim Dirichlet(\alpha = 1)\; for\; a = 1...N_{comp,a}$$

$$\boldsymbol{\sigma}_a \sim Uniform(0,1)\; for\; a = 1...N_{comp,a}$$

$$\boldsymbol{f}_b \sim Dirichlet(\alpha = 1)\; for\; b = 1...N_{comp,b}$$

$$\boldsymbol{\sigma}_b \sim Uniform(0,1)\; for\; b = 1...N_{comp,b}$$

HBM equation generalized

$$\underbrace{p(\theta|d)}_{\text{posterior}} \propto \underbrace{p(\theta)}_{\text{prior}}\underbrace{p(d|\theta)}_{\text{likelihood}}$$

$$\underbrace{p(\phi,\theta|d)}_{\text{posterior}} \propto \underbrace{p(\phi)}_{\text{prior}}\underbrace{p(\theta|\phi)p(d|\theta,\phi)}_{\text{likelihood}}$$

$$\underbrace{p(\phi|d)}_{\text{posterior}} \propto \underbrace{p(\phi)}_{\text{prior}}\underbrace{\int d\theta p(\theta|\phi)p(d|\theta)}_{\text{likelihood}}$$

$$p(\phi|d) \propto p(\phi) \int d\theta p(\theta|\phi) p(d|\theta)$$

Integrating over the latent variables (population constituents) gives marginal posteriors for the hyperparameters

$$p(\phi|d) \propto p(\phi) \int d\theta p(\theta|\phi) \prod_{j=1}^{N_{\mathrm{targ}}} p(d_j|\theta)$$

If each measurable is independent from another, they are considered "separable" (product)

HBM equation generalized

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{d}, M) = \frac{p(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi}, M)p(\boldsymbol{\theta}|\boldsymbol{\phi}, M)p(\boldsymbol{\phi}, M)}{\int d\boldsymbol{\phi}\, p(\mathbf{d}|\boldsymbol{\theta}, \boldsymbol{\phi}, M)p(\boldsymbol{\theta}, \boldsymbol{\phi}|M)}$$

M <- additional information implicit in the model

Graphical models for HBM

- Hyperparameters/ hyperpriors
- Latent variables
- Measurables

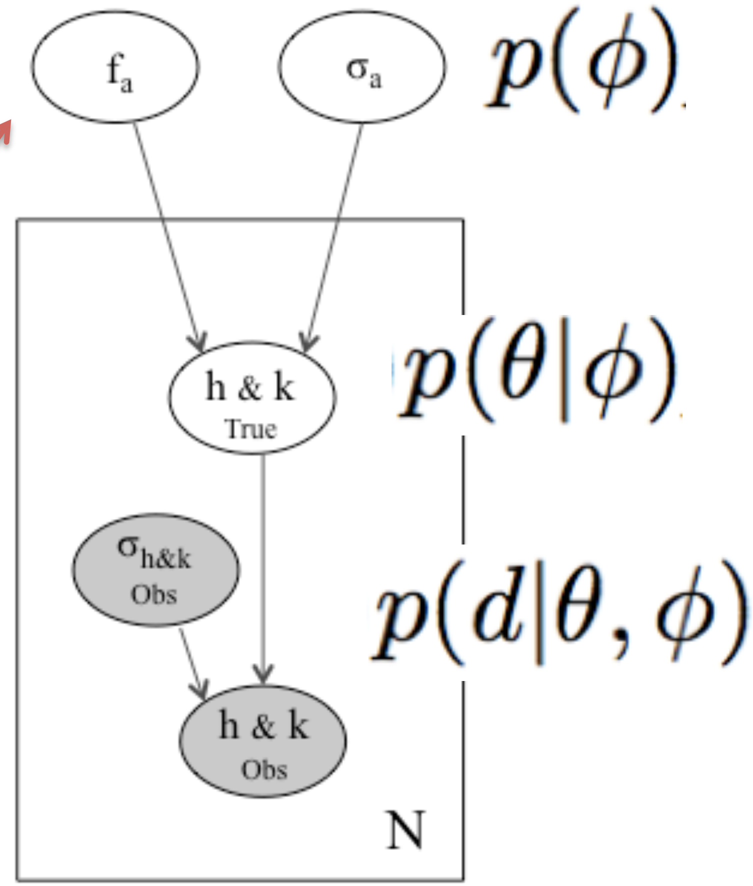
$\boldsymbol{f_a} \sim Dirichlet(\alpha = 1)\ for\ a = 1...N_{comp,a}$

$\boldsymbol{\sigma_a} \sim Uniform(0, 1)\ for\ a = 1...N_{comp,a}$

$\boldsymbol{f_b} \sim Dirichlet(\alpha = 1)\ for\ b = 1...N_{comp,b}$

$\boldsymbol{\sigma_b} \sim Uniform(0, 1)\ for\ b = 1...N_{comp,b}$

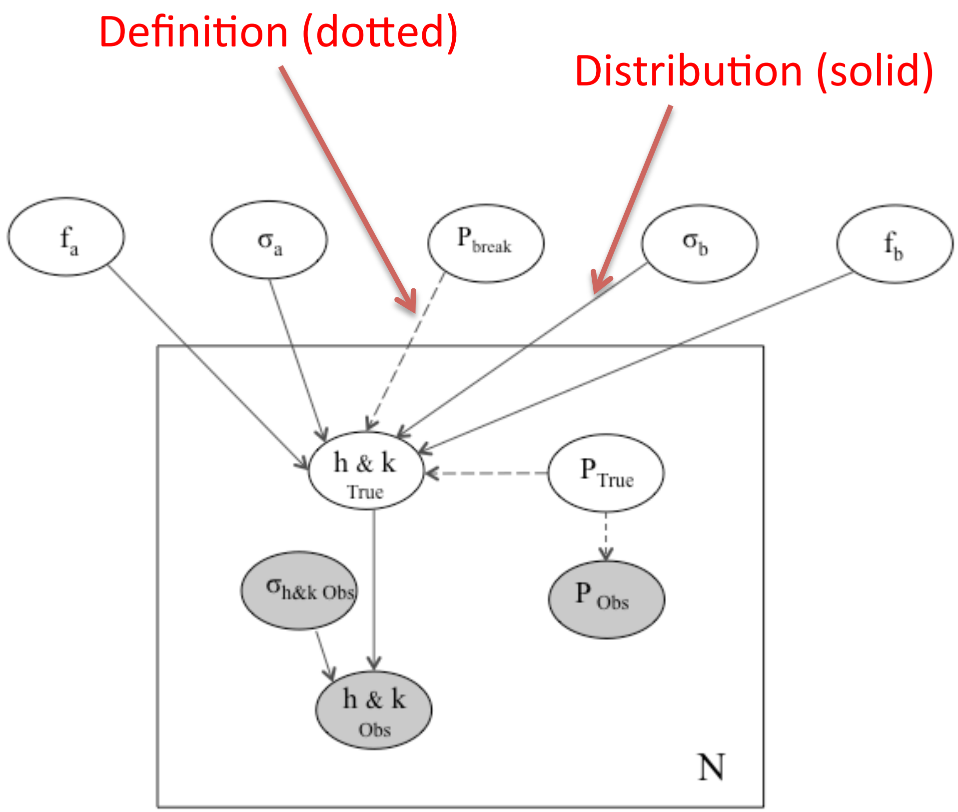
- Physical models vs. probabilistic models



Graphical models for HBM

- Hyperparameters/ hyperpriors
- Latent variables
- Measurables

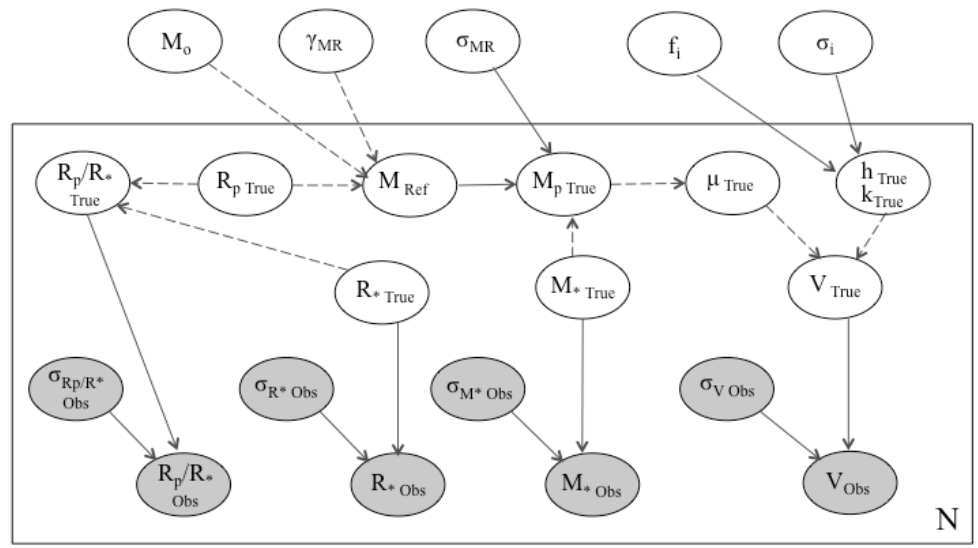
- Physical models vs. probabilistic models



Graphical models for HBM

- Hyperparameters/ hyperpriors
- Latent variables
- Measurables

- Physical models vs. probabilistic models

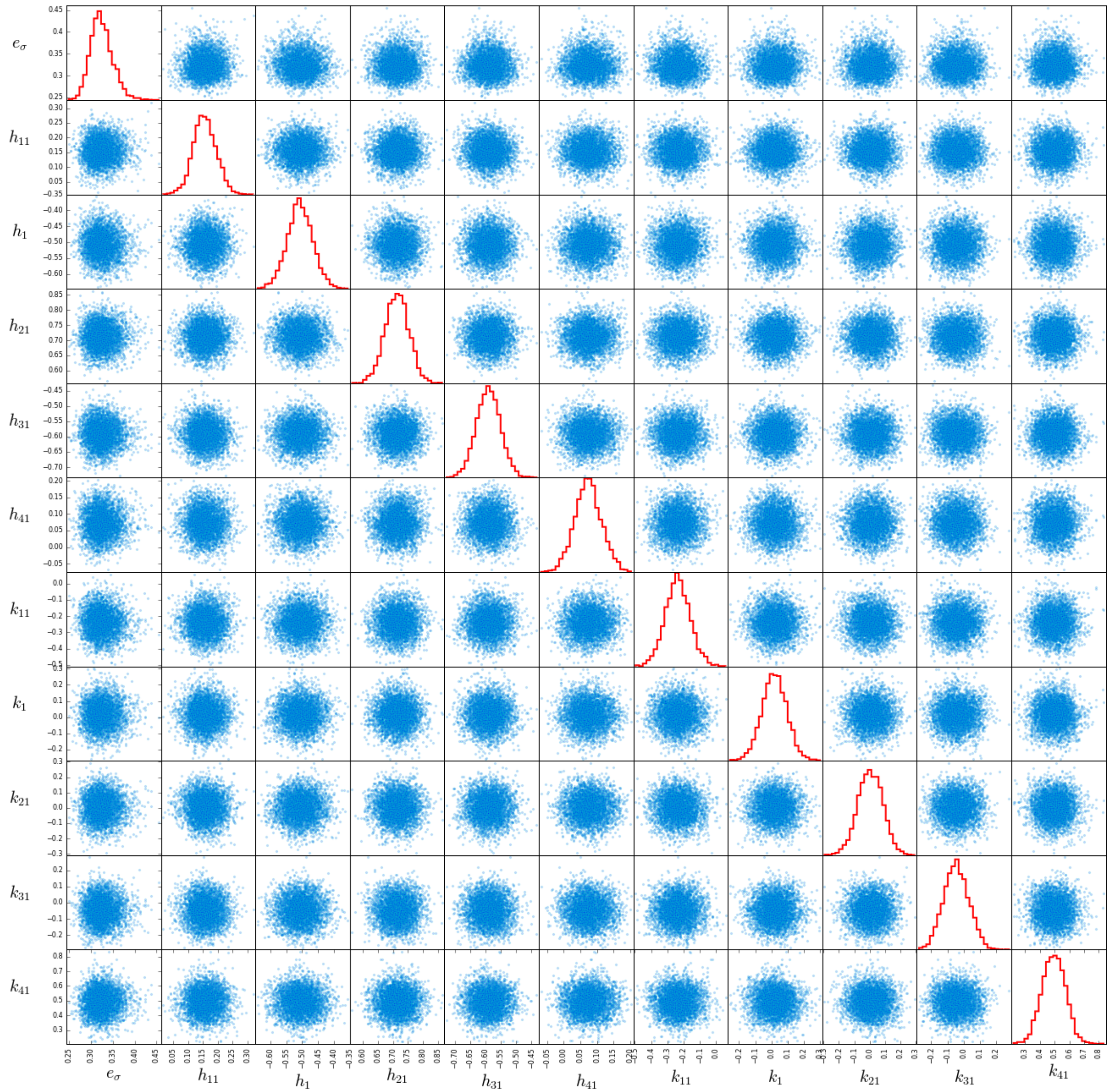


Questionable

Summarizing Posterior Distributions

- Reporting posteriors is good practice in open reproducible science.
- 1d and 2d marginal distributions
  - Integrate over all other parameters to get a 1d marginal
  - The values used in the MCMC iterations (from a “not-not” converged) numerical simulation are mathematically equivalent to the marginal distributions.
  - Get 1d marginal by plotting histogram of iterations for one parameter
  - Get 2d marginal by plotting the vector of iterations for 2 parameters against each other (scatter plot)
- KDE for 2D marginal
- 95% equal tailed credible interval: The interval containing the median, containing 95% of the probability.

Summarizing Posterior Distributions



Software for HBM

- Stan
- JAGS
- PyMC3
- TensorFlow

- Software is open source and maintained. Encourages reproducibility, and transparency of scientific research

Sampling Routine, when and why

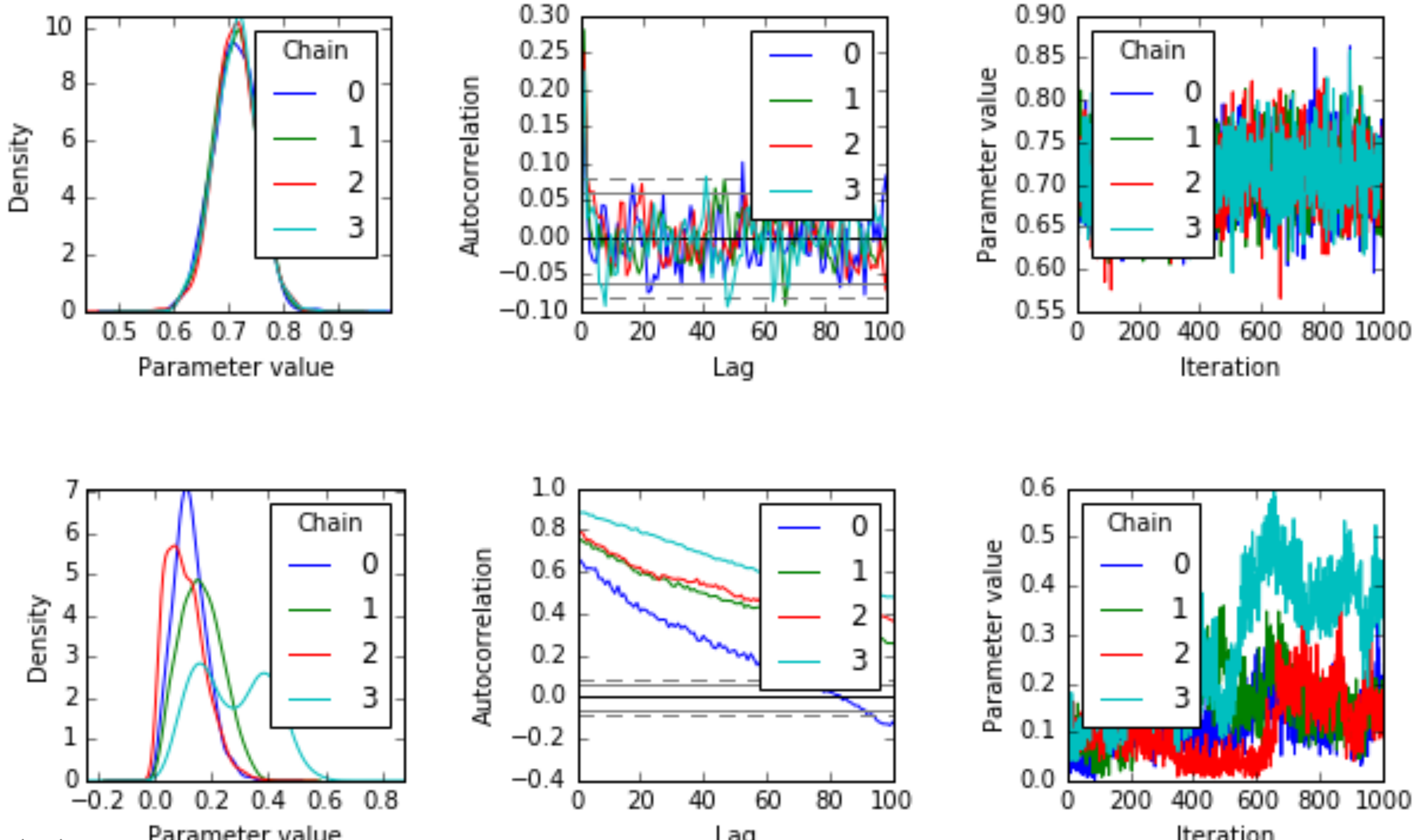
- MCMC Metropolis Hasting or Gibbs Sampling
  - Exchangeability
- HMC
  - Derivatives: continuous functions

Diagnostics:
Has your numerical simulation “not-not” converged?

- Gibbs Sampling and Metropolis-Hastings MCMC,:
  - Example Modeling Language: JAGS (Just Another Gibbs Sampler)
  - Look at **trace plots** for evidence of well mixed chains.
  - **Autocorrelation**: look for a zero crossing at a small lag, and a lag that’s distributed about zero
  - **Gelman-Rubin statistic** of 1.0 or less than some small threshold close to 1.
  - Look at the truths from **simulated data** over-plotted on their marginal posteriors. Do you get back out what you put in when you know the answer?
- Hamiltonian Monte Carlo:
  - Example Modeling Language: Stan
  - Same procedure as above
  - Can also look for **divergences** and test **centered vs. non-centered** parameterizations. (Easy with R analysis tools for Stan, See Stan Case Studies).

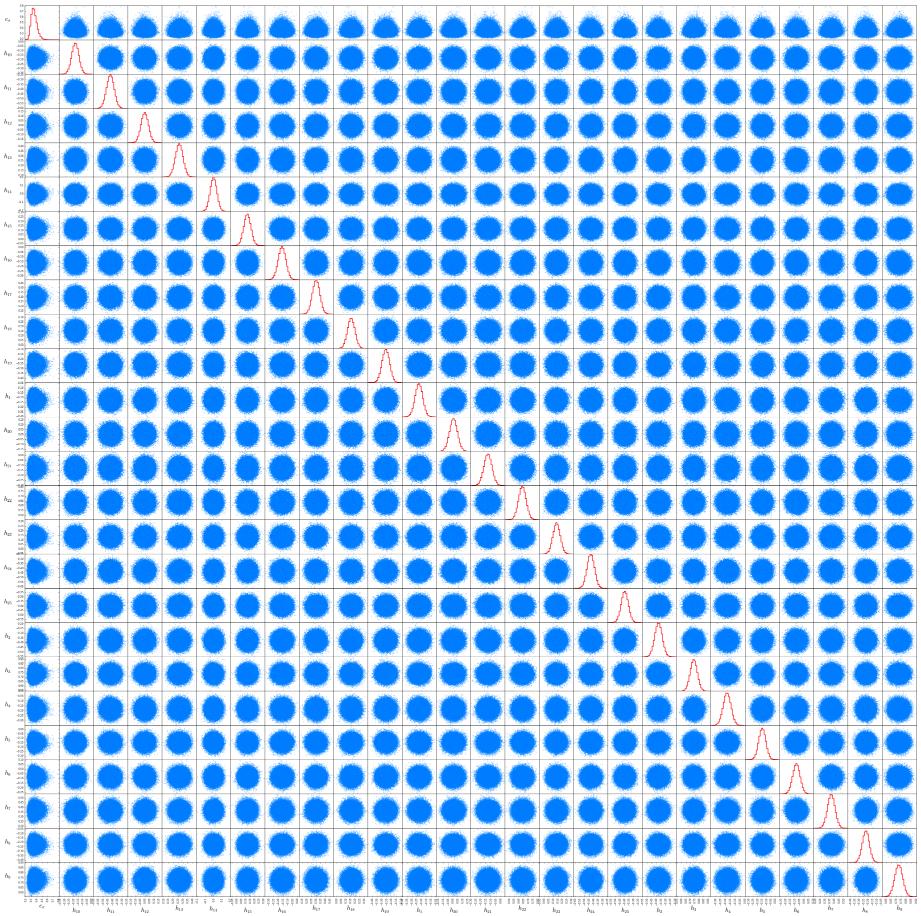
Diagnostics:

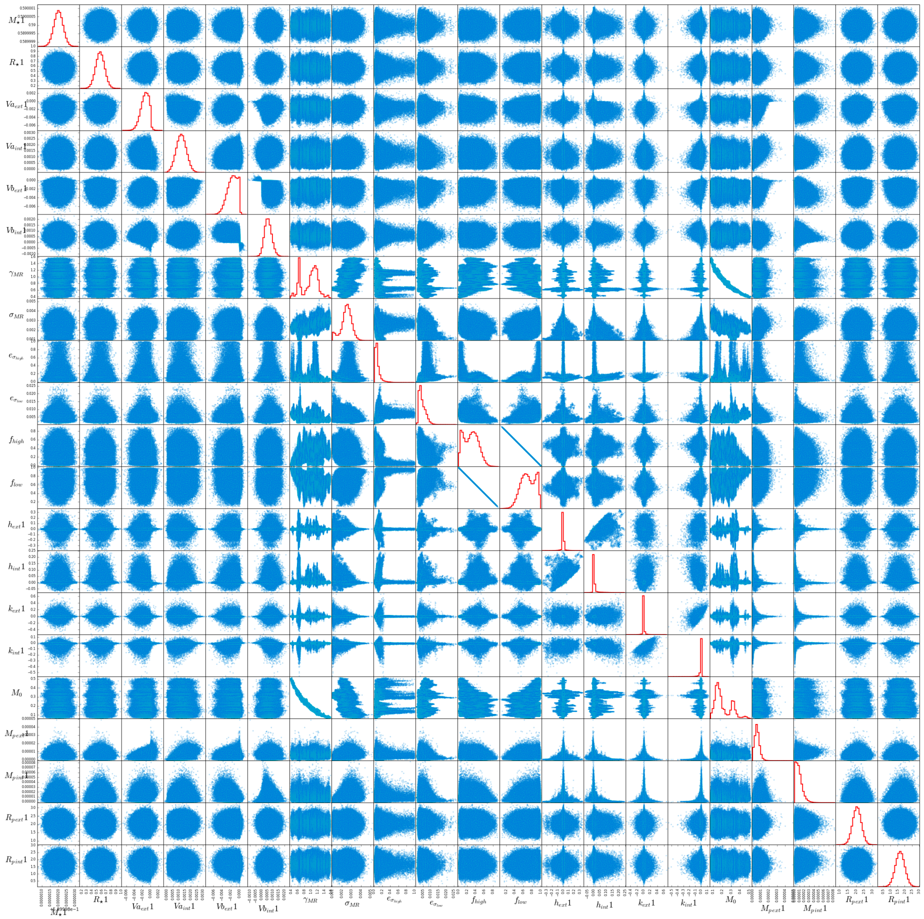
Has your numerical simulation “not-not” converged?



Diagnostics:

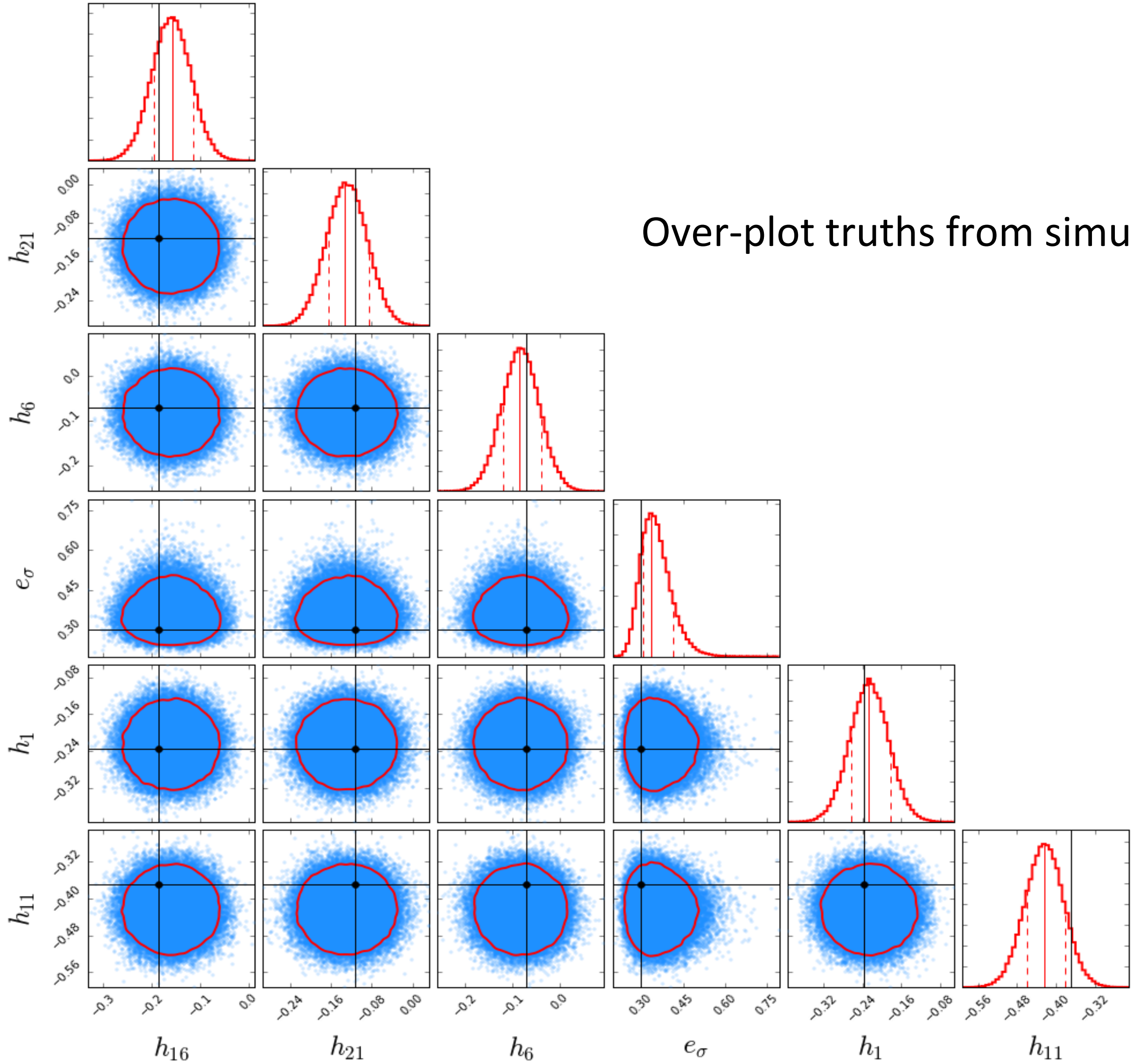
Has your numerical simulation “not-not” converged?





Diagnostics:

Has your numerical simulation “not-not” converged?



Over-plot truths from simulated data

Workflow Best Practices

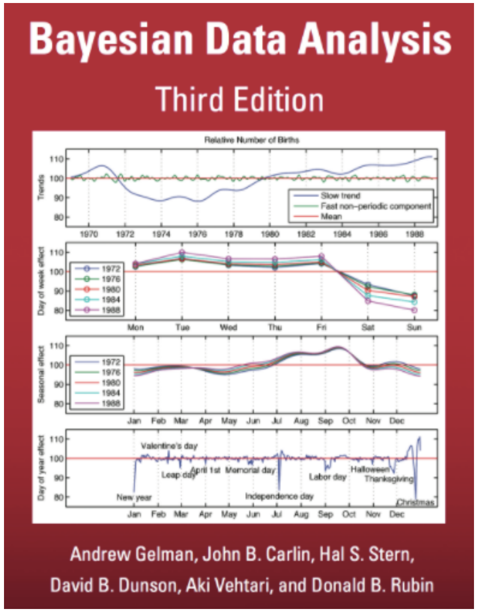
1. The model you choose depends on the science question you ask ahead of time.
   - E.g “Researcher degrees of freedom”
   - Conditional probability
   - Are parameters and hyperparameters separable?
2. The mathematical artifacts when sampling your model applied to your data are important, diagnostics are still being developed.
3. Some solutions:
   - Take out predictor variables that are highly correlated with each other when tailoring your numerical simulation
   - **Conjugate priors** for computational tractability (set of all densities to choose from (for prior) having the same functional form as likelihood)

Things to Keep in Mind:

- **Fractional Uncertainty**
- There are more parameters in your HBM than there are data points.
- Conditional probability -> weighting something by something else -> multiplying -> log Likelihood is a very large negative number (not normalized) -> Smaller large negative number has higher probability.
- **Regularization** (e.g., spread in planet mass at a given radius) naturally handles over-fitting
- Philosophy: **Prediction** or **Statistical Significance**?

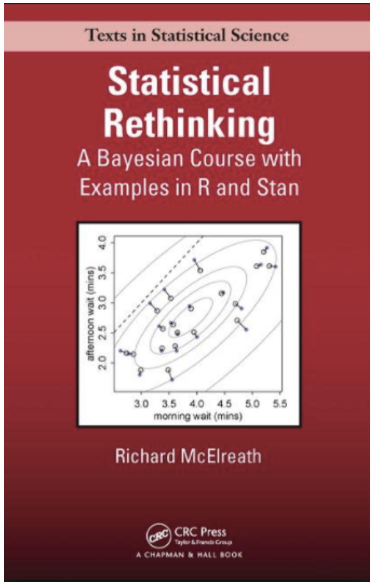
Selected Reference Material

- ***Bayesian Data Analysis***, Gelman, Carlin, Stern, & Rubin



Selected Reference Material

- ***Statistical Rethinking: A Bayesian Course with Examples in R and Stan*** (Self directed learning)



Selected Reference Material

- ***Stan – Case Studies****: Open-source Models and Methods*
  http://mc-stan.org/users/documentation/case-studies

Hierarchical Bayes and Open Reproducible Science

- **P-hacking:** "…the problem: weak statistical standards of evidence for claiming new discoveries"
  - Big names in statistics want to shake up much-mailigned P value
- **Researchers Degrees of Freedom:** "…when we look at exploratory results, we must be aware of their uncertainty and fragility."
  - The garden of forking paths: ... Andrew Gelman and Erik Loken 2013
- **There's no finality in Science:** "All data requires interpretation, which is subject to bias, and all results are preliminary."
  - Interesting opinion article in Scientific America
- HBM: Can incorporate **bias** in a quantitative manner.
- Modeling languages allow **transparency** in research workflow (e.g., JAGS, Stan, TensorFlow, and more).