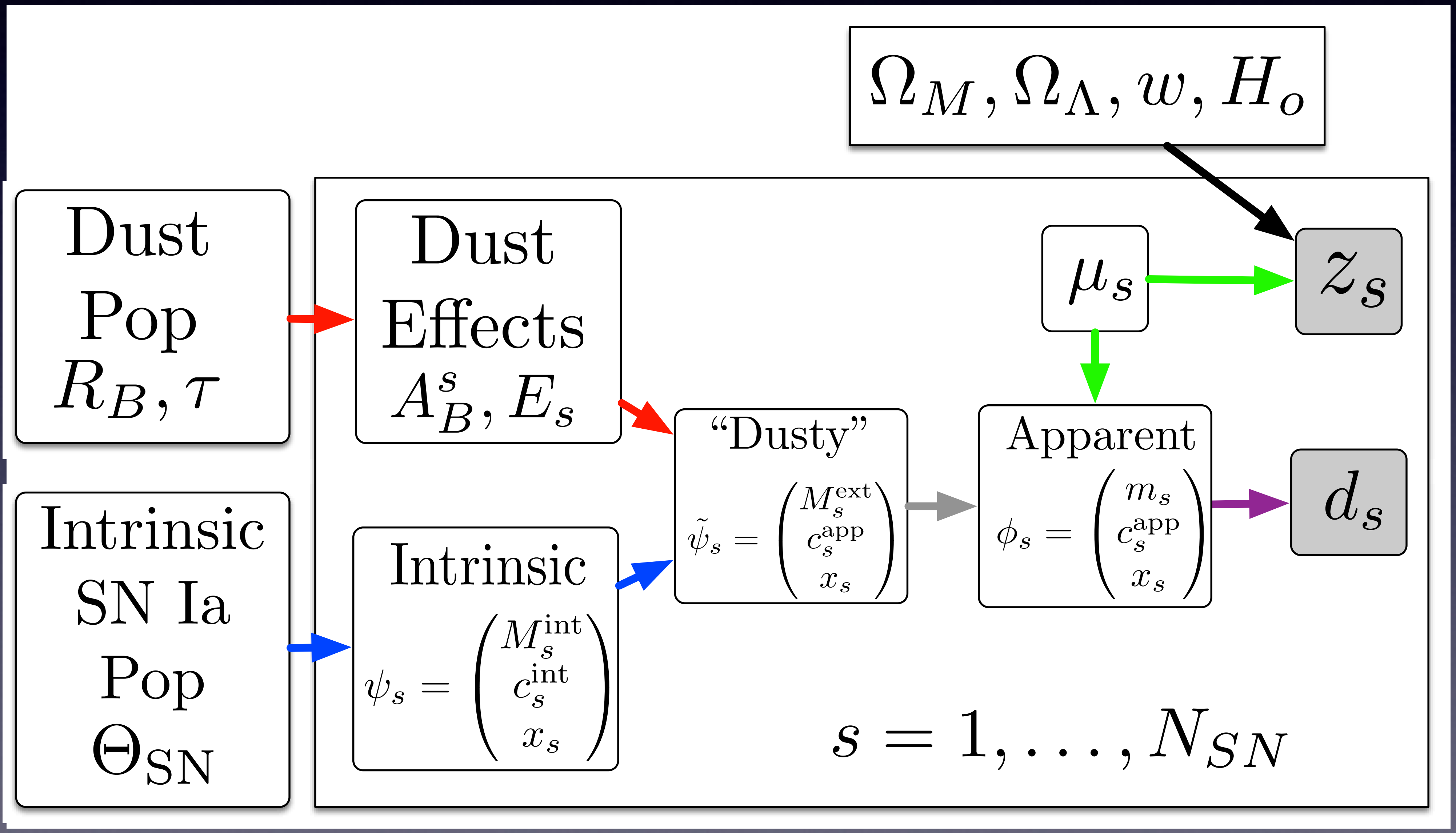
Astrostatistics: Thu 08 Mar 2017

https://github.com/CambridgeAstroStat/PartIII-Astrostatistics

- Example Class 3: **Mon, Mar 12, 15:30~17:30pm MR14**
    - Quasar Time Delay Problem brings together Bayesian inference, MCMC, and Gaussian Processes
- Example Class 4 & Revision Class in Easter Term
- Email List & Course Questionaires
- Last time: Gaussian Processes
- Today: Jeffrey's Prior & Start Probabilistic Graphical Models & Hierarchical Bayes
    - Gelman BDA, Chapter 5
    - Bishop, Pattern Recognition & Machine Learning, Chapter 8

Hierarchical Bayes, Huh?
What is it good for?
ABSOLUTELY EVERYTHING!



Mandel et al. 2017

Common Problems in Astronomy

- Want to learn about a population of objects from a finite sample of individuals, each measured with error
- Observed Data is actually a combination of uncertain astrophysical & instrumental & selection effects. Need to model them to infer the "intrinsic" properties of the object or population of objects ("deconvolve")
- Research examples from the Class?

What is Hierarchical Bayes?

Simple Bayes: $\mathcal{D}|\,\theta \sim \text{Model}(\theta)$

Posterior (Bayes' Theorem): $P(\theta|\mathcal{D}) \propto P(\mathcal{D}|\,\theta)P(\theta)$

Hierarchical Bayes: $\theta_i:$ Parameter of Individual

α, β: Hyperparameter of Population

$$\mathcal{D}_i|\,\theta_i \sim \text{Model}(\theta_i)$$

$$\theta_i|\,\alpha, \beta \sim \text{PopModel}(\alpha, \beta)$$

Joint Posterior:

$$P(\{\theta_i\}, \alpha, \beta|\{\mathcal{D}_i\}) \propto \left[\prod_{i=1}^{N} P(\mathcal{D}_i|\,\theta_i)P(\theta_i|\alpha, \beta)\right]P(\alpha, \beta)$$

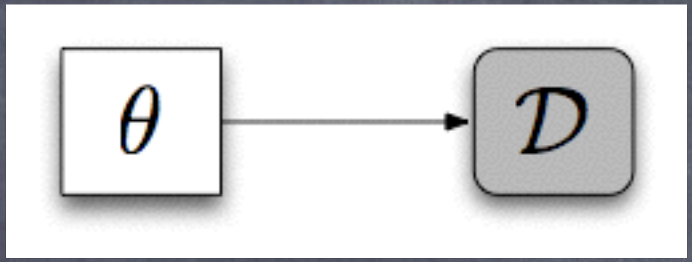
Build up complexity by layering conditional probabilities

Probabilistic Graphical Models:

a visual way to understand complex statistical models

Simple Bayes:

$$\mathcal{D}|\,\theta \sim \mathrm{Model}(\theta)$$

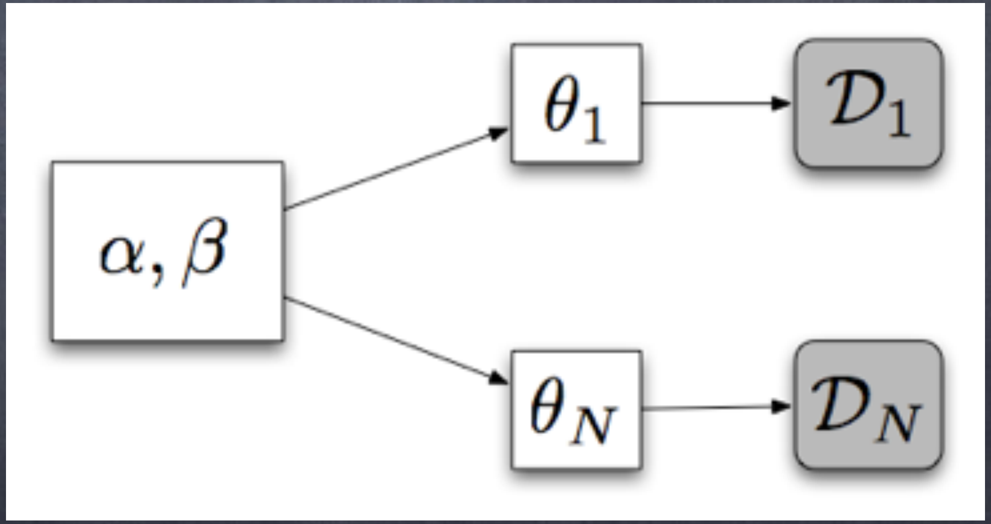


$$P(\theta|\,\mathcal{D}) \propto P(\mathcal{D}|\,\theta)P(\theta)$$

Hierarchical Bayes:

$$\mathcal{D}_i|\,\theta_i \sim \mathrm{Model}(\theta_i)$$

$$\theta_i|\,\alpha,\beta \sim \mathrm{PopModel}(\alpha,\beta)$$



$$P(\{\theta_i\},\alpha,\beta|\{\mathcal{D}_i\}) \propto \left[\prod_{i=1}^{N} P(\mathcal{D}_i|\,\theta_i)P(\theta_i|\alpha,\beta)\right]P(\alpha,\beta)$$

Build up complexity by layering conditional probabilities

Probabilistic Graphical Models

Forward Model:

$$\theta_i | \, \alpha, \beta \sim \mathrm{PopModel}(\alpha, \beta)$$

$$\mathcal{D}_i | \, \theta_i \sim \mathrm{Model}(\theta_i)$$

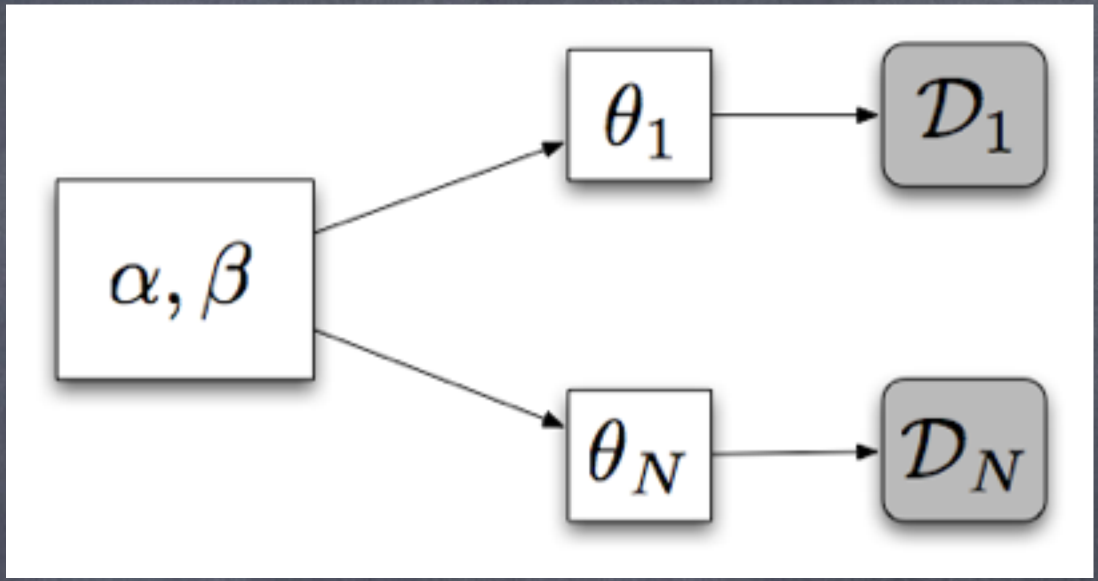
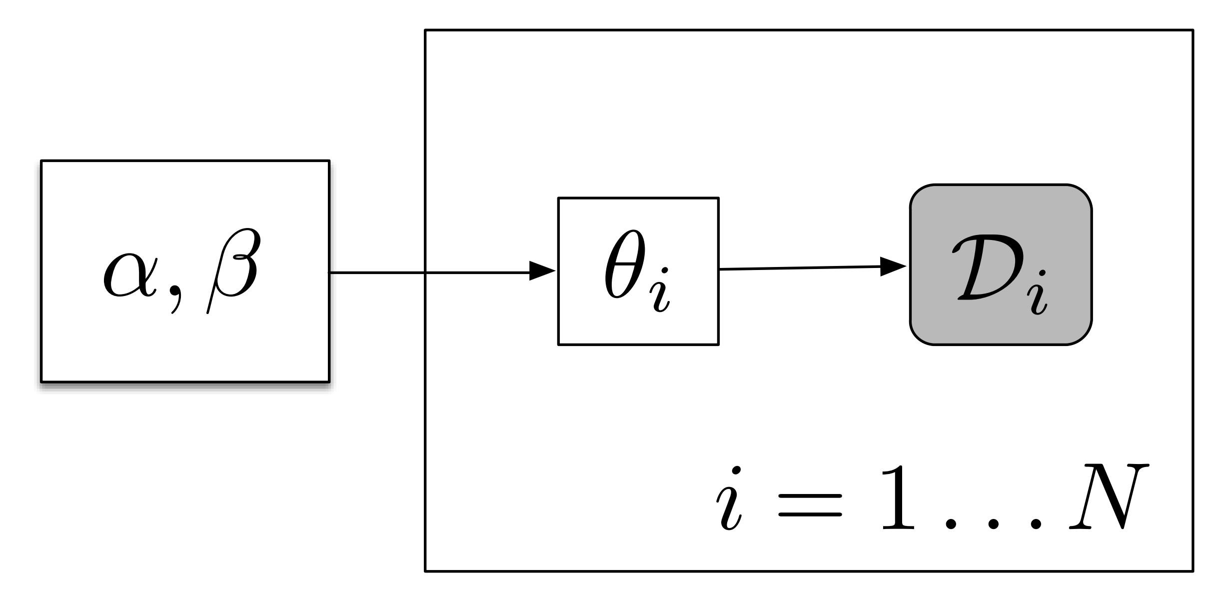


Plate Notation:
(loop over individuals in sample)



$$P(\{\theta_i\}, \alpha, \beta | \{\mathcal{D}_i\}) \propto \left[\prod_{i=1}^{N} P(\mathcal{D}_i | \, \theta_i) P(\theta_i | \alpha, \beta) \right] P(\alpha, \beta)$$

Build up complexity by layering conditional probabilities

Advantages of Hierarchical Bayesian Models

- Common Problem in Astronomy: Infer properties of population from finite sample of individuals with noisy measurements
- Incorporate multiple sources of randomness & uncertainty as "latent variables" with distributions underlying the data
- Express structured probability models adapted to data-generating process ("forward model")
- Bayesian: Full (non-gaussian) probability distribution = Global, coherent quantification of uncertainties
- Completely Explore & Marginalize Posterior trade-offs/ degeneracies between parameters/hyperparameters

Simplest Hierarchical Bayesian / Multi-level Model: "Normal-Normal" (Gelman BDA, Sec 5.4)

Level 1: Population Distribution of Latent Variables (Absolute Mags)

$$M_s \sim N(\tilde{M}, \tau^2)$$

Latent Variables

Hyperparameters (Pop Mean & Variance)

Level 2:
Measurement Error Process

$$\hat{M}_s | M_s \sim N(M_s, \sigma_s^2)$$

Measurements (Data)

Heteroskedastic Meas. Error Variance (known)

(Draw PGM / DAG on Chalkboard) $s = 1 \dots N$

Hierarchical Bayesian "Normal-Normal" Model

Level 1: Population Distribution of Latent Variables (Absolute Mags)

$$M_s \sim N(\tilde{M}, \tau^2)$$

Level 2 : Measurement Error Process

$$\hat{M}_s | M_s \sim N(M_s, \sigma_s^2)$$

Joint Probability Density of Data, Latent Variables, Hyperparameters

$$P(\{\hat{M}_s\}, \{M_s\}, \tilde{M}, \tau^2)$$

Factor into Conditional and Marginal distributions based on Model

$$P(\{\hat{M}_s\}, \{M_s\}, \tilde{M}, \tau^2) = P(\{\hat{M}_s\}, \{M_s\} | \tilde{M}, \tau^2) \times P(\tilde{M}, \tau^2)$$

Hierarchical Bayesian "Normal-Normal" Model

Level 1: Population Distribution of Latent Variables (Absolute Mags)

$$M_s \sim N(\tilde{M}, \tau^2)$$

Level 2 : Measurement Error Process

$$\hat{M}_s | M_s \sim N(M_s, \sigma_s^2)$$

Joint Probability Density of Data, Latent Variables, Hyperparameters

$$P(\{\hat{M}_s\}, \{M_s\}, \tilde{M}, \tau^2) = P(\{\hat{M}_s\}, \{M_s\} | \tilde{M}, \tau^2) \times P(\tilde{M}, \tau^2)$$

$$P(\{\hat{M}_s\}, \{M_s\} | \tilde{M}, \tau^2) = P(\{\hat{M}_s\} | \{M_s\}; \tilde{M}, \tau^2) \times P(\{M_s\} | \tilde{M}, \tau^2)$$

(Likelihood)

$$P(\{\hat{M}_s\} | \{M_s\}; \tilde{M}, \tau^2) = P(\{\hat{M}_s\} | \{M_s\})$$

(Conditional Independence)

Hierarchical Bayesian "Normal-Normal" Model

Level 1: Population Distribution of Latent Variables (Absolute Mags)

$$M_s \sim N(\tilde{M}, \tau^2)$$

Level 2 : Measurement Error Process

$$\hat{M}_s | M_s \sim N(M_s, \sigma_s^2)$$

Likelihood function:

$$P(\{\hat{M}_s\}|\{M_s\}) = \prod_{s=1}^{N} P(\hat{M}_s|\{M_i\}) = \prod_{s=1}^{N} P(\hat{M}_s|M_s)$$

(Independent Measurements) (Dist'n of Measurement s only depends on latent Mag s)

$$= \prod_{s=1}^{N} N(\hat{M}_s|\, M_s, \sigma_s^2)$$

Hierarchical Bayesian "Normal-Normal" Model

Level 1: Population Distribution of Latent Variables (Absolute Mags)

$$M_s \sim N(\tilde{M}, \tau^2)$$

Level 2 : Measurement Error Process

$$\hat{M}_s | M_s \sim N(M_s, \sigma_s^2)$$

Joint Probability Density of Data, Latent Variables, Hyperparameters

$$P(\{\hat{M}_s\}, \{M_s\}, \tilde{M}, \tau^2) = P(\{\hat{M}_s\}, \{M_s\} | \tilde{M}, \tau^2) \times P(\tilde{M}, \tau^2)$$

$$P(\{\hat{M}_s\}, \{M_s\} | \tilde{M}, \tau^2) = P(\{\hat{M}_s\} | \{M_s\}; \tilde{M}, \tau^2) \times P(\{M_s\} | \tilde{M}, \tau^2)$$

(Population Distribution)

$$P(\{\hat{M}_s\} | \{M_s\}; \tilde{M}, \tau^2) = P(\{\hat{M}_s\} | \{M_s\})$$

(Conditional Independence)

Hierarchical Bayesian "Normal-Normal" Model

Level 1: Population Distribution of Latent Variables (Absolute Mags)

$$M_s \sim N(\tilde{M}, \tau^2)$$

Level 2 : Measurement Error Process

$$\hat{M}_s | M_s \sim N(M_s, \sigma_s^2)$$

Population Distribution (Parameterised Prior)

$$P(\{M_s\} | \tilde{M}, \tau^2) = \prod_{s=1}^{N} P(M_s | \tilde{M}, \tau^2)$$

(Each latent s is independent, conditional on pop hyperparameters)

$$= \prod_{s=1}^{N} N(M_s | \tilde{M}, \tau^2)$$

Hierarchical Bayesian "Normal-Normal" Model

Level 1: Population Distribution of Latent Variables (Absolute Mags)

$$M_s \sim N(\tilde{M}, \tau^2)$$

Measurement Likelihood

Level 2 : Measurement Error Process

$$\hat{M}_s | M_s \sim N(M_s, \sigma_s^2)$$

Population Dist'n / Prior

Joint Probability Density of ALL THE THINGS:
Data, Latent Variables, Hyperparameters

$$P(\{\hat{M}_s\}, \{M_s\}, \tilde{M}, \tau^2) = \prod_{s=1}^{N} [P(\hat{M}_s | M_s) P(M_s | \tilde{M}, \tau^2)] \times P(\tilde{M}, \tau^2)$$

Measurement Likelihood

Population Dist'n / Prior

Hyperprior

Putting the Bayesian in Hierarchical Bayesian

Joint Probability Density of ALL THE THINGS:
Data, Latent Variables, Hyperparameters

$$P(\{\hat{M}_s\}, \{M_s\}, \tilde{M}, \tau^2) = \prod_{s=1}^{N} [P(\hat{M}_s | M_s) P(M_s | \tilde{M}, \tau^2)] \times P(\tilde{M}, \tau^2)$$

Data   Unknowns

Joint Posterior of all unknowns given the data

$$P(\{M_s\}; \tilde{M}, \tau^2 | \{\hat{M}_s\}) = \frac{P(\{\hat{M}_s\}, \{M_s\}, \tilde{M}, \tau^2)}{P(\hat{M}_s)}$$

(Normalisation Constant)

Putting the Bayesian in Hierarchical Bayesian

Joint Posterior of all unknowns given the data

$$P(\{M_s\}; \tilde{M}, \tau^2 | \{\hat{M}_s\}) = \frac{P(\{\hat{M}_s\}, \{M_s\}, \tilde{M}, \tau^2)}{P(\hat{M}_s)}$$

Unknowns

Data

(Ignorable Normalisation Constant)

(Posterior on N + 2 dimensional parameter space)

$$P(\{M_s\}; \tilde{M}, \tau^2 | \{\hat{M}_s\}) \propto \left[\prod_{s=1}^{N} P(\hat{M}_s | M_s) P(M_s | \tilde{M}, \tau^2) \right] \times P(\tilde{M}, \tau^2)$$

Measurement Likelihood

Population Dist'n / Prior

Hyperprior

Gibbs sampling & Hierarchical Bayes

Utilises Conditional Independence structure of PGM to derive conditional posterior densities

$$P(\{M_s\}; \tilde{M}, \tau^2 | \{\hat{M}_s\}) \propto$$
$$\left[\prod_{s=1}^{N} P(\hat{M}_s | M_s) P(M_s | \tilde{M}, \tau^2)\right] \times P(\tilde{M}, \tau^2)$$

1. Sample Latent Variables from Conditional: $s = 1 \ldots N$

$$P(M_s | \tilde{M}, \tau^2; \{\hat{M}_s\}) \propto P(\hat{M}_s | M_s) P(M_s | \tilde{M}, \tau^2)$$

2. Sample Hyperparameters from Conditional:

$$P(\tilde{M}, \tau^2 | \{M_s\}; \{\hat{M}_s\}) = P(\tilde{M}, \tau^2 | \{M_s\})$$ (Conditional Independence)

$$\propto \left[\prod_{s=1}^{N} P(M_s | \tilde{M}, \tau^2)\right] \times P(\tilde{M}, \tau^2)$$