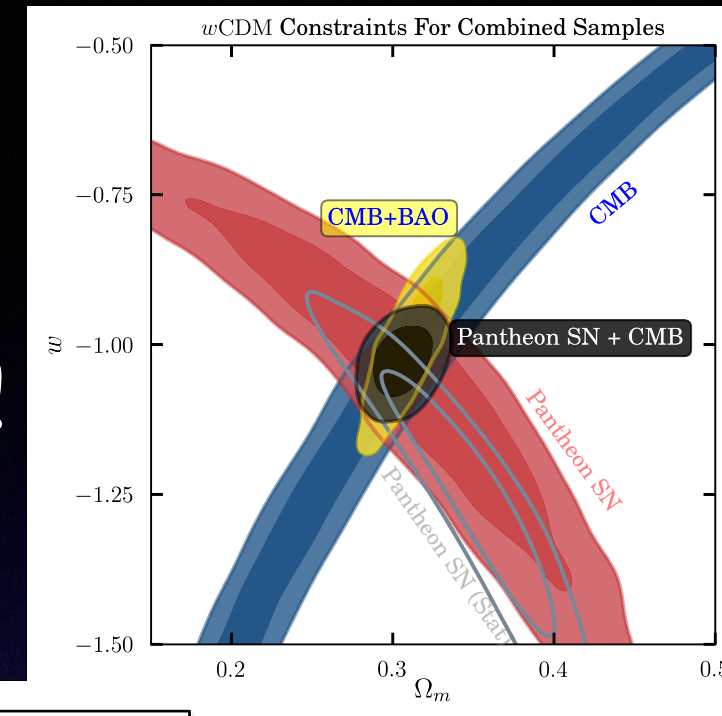
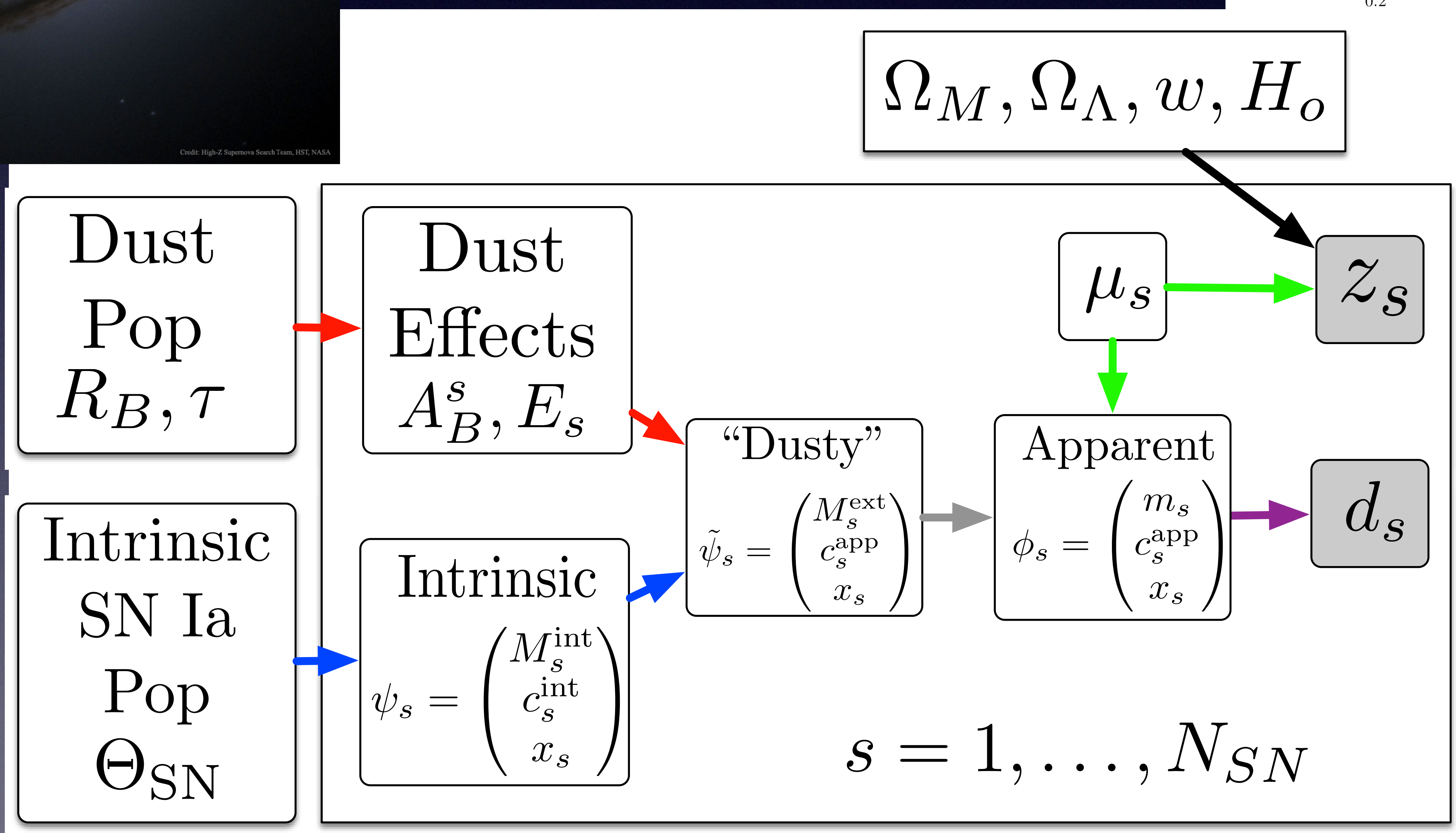
Astrostatistics: 09 Mar 2020

https://github.com/CambridgeAstroStat/PartIII-Astrostatistics-2020

- Example Class 3: Fri 13 Mar, 12pm, MR5
- Today: Probabilistic Graphical Models (Bishop, Ch 8) & Hierarchical Bayes

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

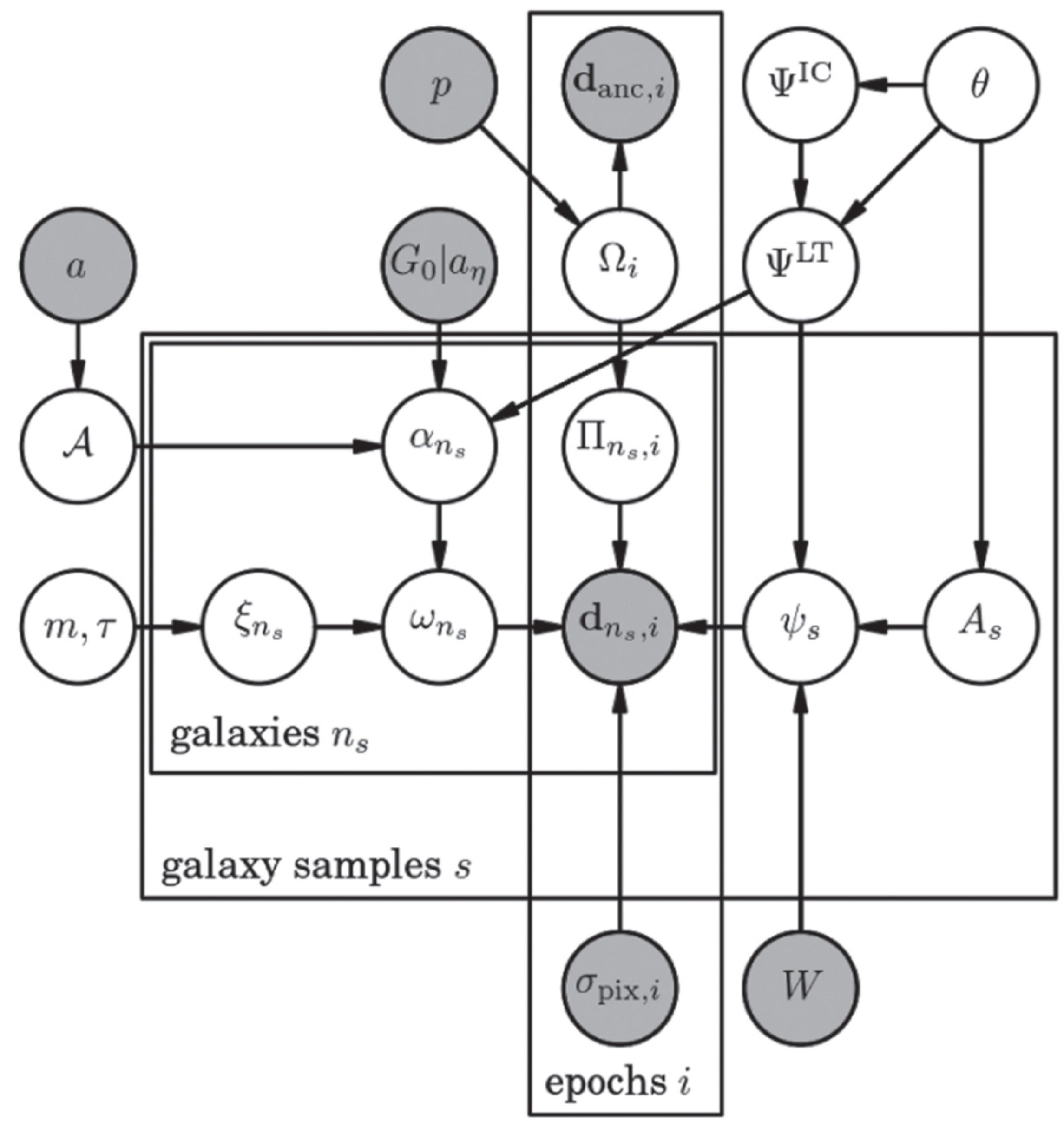




Mandel et al. 2017

Hierarchical Models in Astrophysics

- M. Schneider et al. 2015, ApJ 807, 87, “Hierarchical Probabilistic Inference of Cosmic Shear”. (weak gravitational lensing)
- “Hierarchical Probabilistic” = “Hierarchical Bayesian” = “Bayesian Hierarchical” = “Multilevel Model (MLM)”



Probabilistic Graphical Model

Generative Model for the Sky: Regier et al. “Learning an Astronomical Catalog of the Visible Universe through Scalable Bayesian Inference”

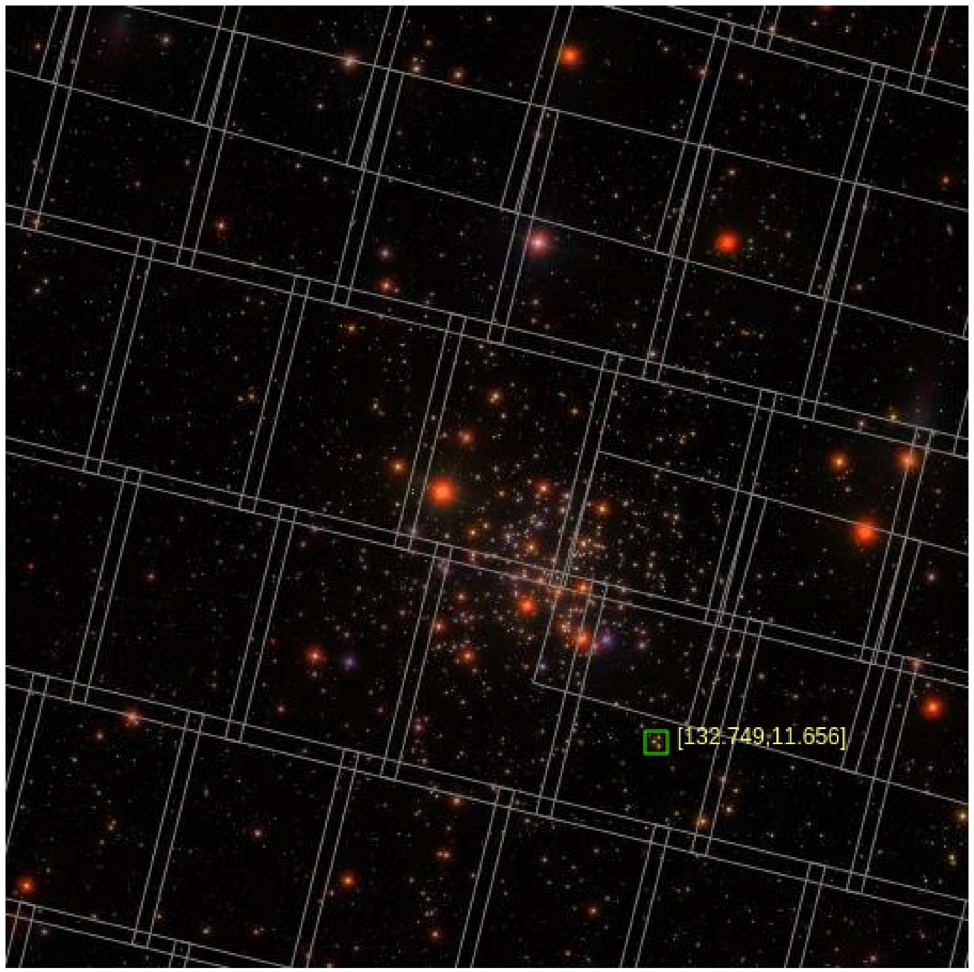


Figure 1: SDSS image boundaries. Some images overlap substantially. Some light sources appear in multiple images that do not overlap. Celeste uses all relevant data to locate and characterize each light source whereas heuristics typically ignore all but one image in regions with overlap. credit: SDSS DR10 Sky Navigate Tool.

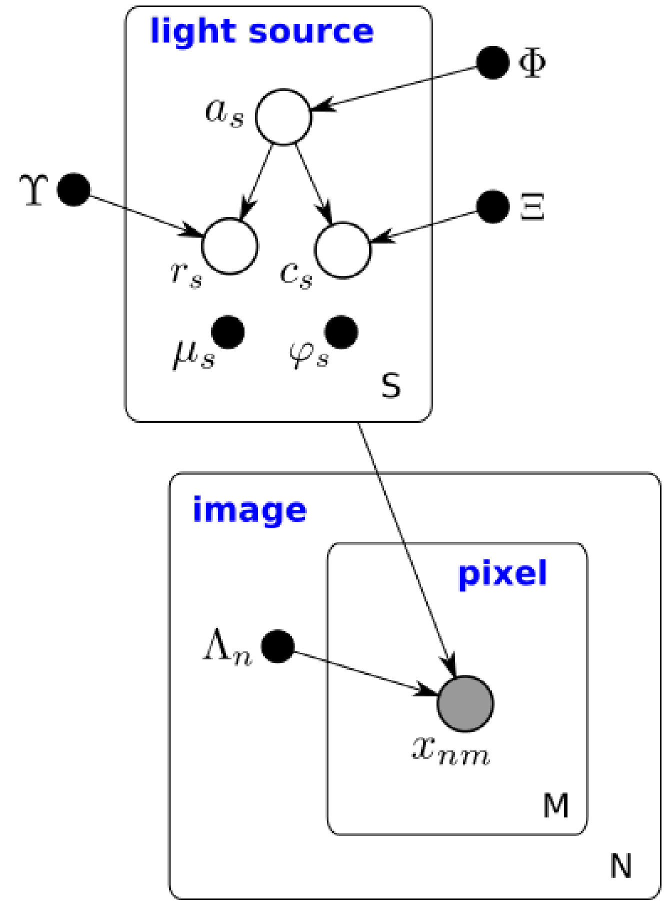


Figure 2: The Celeste graphical model. Shaded vertices represent observed random variables. Empty vertices represent latent random variables. Black dots represent constants. Constants denoted by uppercase Greek characters are fixed throughout our procedure. Constants denoted by lowercase Greek letters are inferred, along with the posterior distribution of the latent random variables. Edges signify permitted conditional dependencies. Plates (the boxes) represent independent replication.

Probabilistic Graphical Model

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”)

What is Hierarchical Bayes?

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

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

Hierarchical Bayes: $\theta_i:$ Parameter of Individual

$\alpha, \beta:$ Hyperparameter of Population

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

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

Joint Posterior:

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

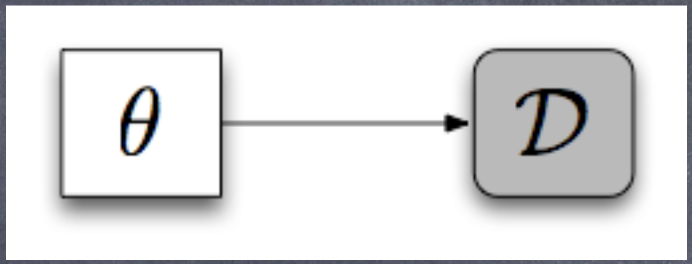
Build up complexity by layering conditional probabilities

Probabilistic Graphical Models:

a visual way to understand complex statistical models

Simple Bayes:

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

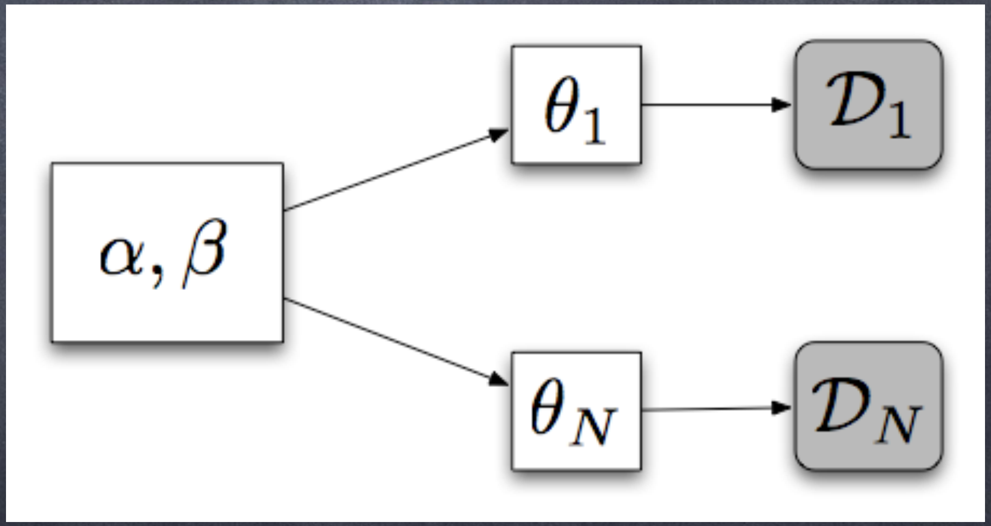


$$P(\theta|\,\mathcal{D}) \propto P(\mathcal{D}|\,\theta)P(\theta)$$

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:

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

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

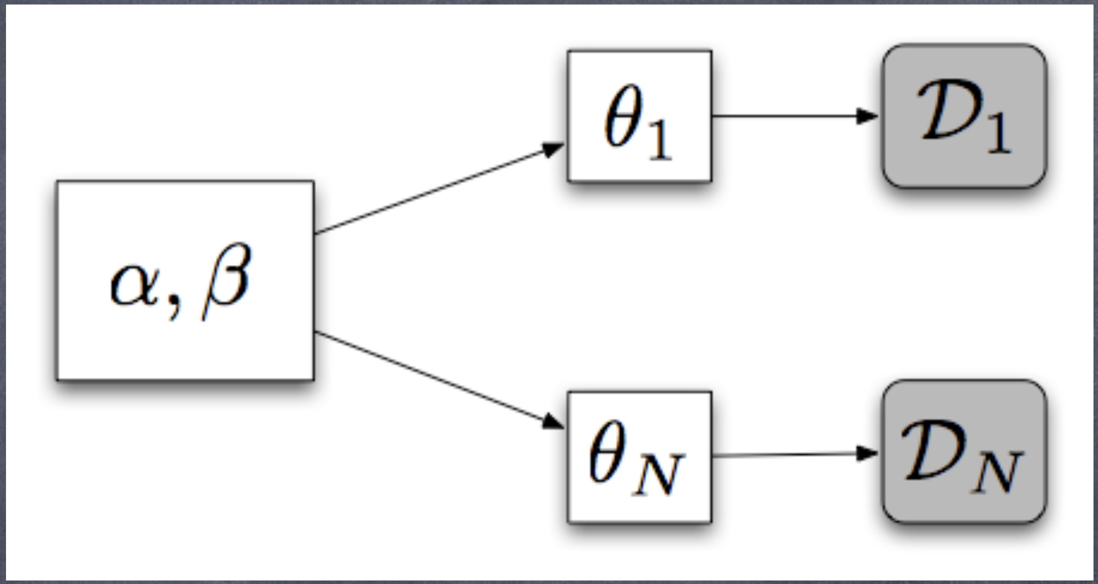
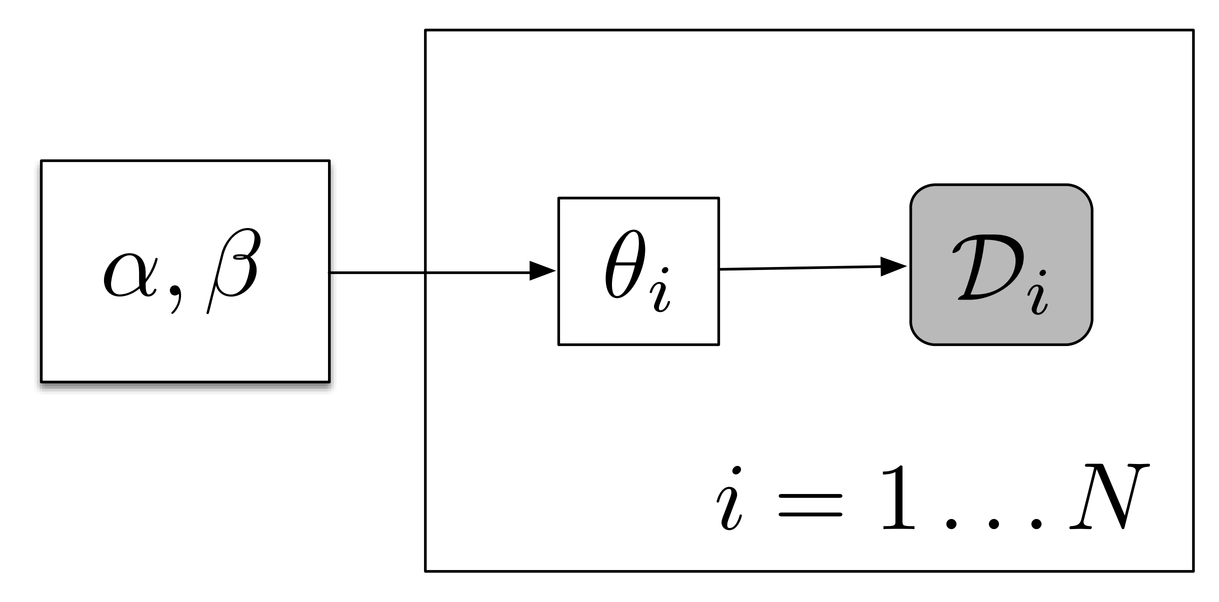


Plate Notation:
(loop over individuals in sample)



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

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

Switch to PRML Slides on properties of Graphical Models

PGM Slides:
lecture21-22_prml_slides-8_abridged.pdf

Slides on: Bayesian Networks, Curve Fitting, Generative Models, Conditional Independence, D-Separation

Simplest Hierarchical Bayesian / Multi-level Model: "Normal-Normal" for Standard CandleMagnitudes

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

$$M_s \sim N(M_0, \tau^2)$$

Latent Variables

Hyperparameters (Pop Mean & Variance)

Level 2: Measurement Error Process

$$D_s | M_s \sim N(M_s, \sigma_s^2)$$

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)

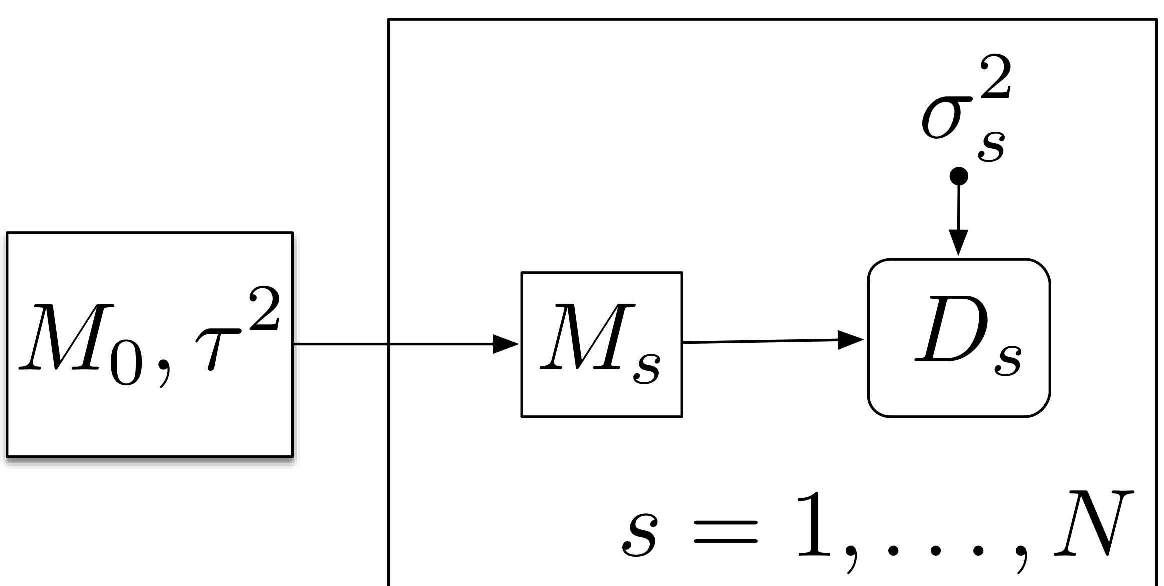
$$M_s \sim N(M_0, \tau^2) \quad s = 1 \dots N$$

Level 2 : Measurement Error Process

$$D_s | M_s \sim N(M_s, \sigma_s^2) \quad s = 1 \dots N$$

Joint Probability Density of Data, Latent Variables, **H**yperparameters

(Derive on board) $P(\{D_s\}, \{M_s\}, \boldsymbol{H} = M_0, \tau^2)$



Joint factors into Conditional and Marginal densities based on Model Assumption

Hierarchical Bayesian "Normal-Normal" Model

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

$$M_s \sim N(M_0, \tau^2)$$

Population Dist'n / Prior

Level 2 : Measurement Error Process

$$D_s | M_s \sim N(M_s, \sigma_s^2)$$

Measurement Likelihood

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

$$P(\{D_s\}, \{M_s\}, \boldsymbol{H}) = \Bigg[\prod_{s=1}^{N} \underbrace{P(D_s | M_s)}_{\text{Measurement Likelihood}} \underbrace{P(M_s | M_0, \tau^2)}_{\text{Population Dist'n / Prior}} \Bigg] \times \underbrace{P(\boldsymbol{H})}_{\text{Hyperprior}}$$

(Derive on board)

Putting the Bayesian in Hierarchical Bayesian

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

$$P(\{D_s\}, \{M_s\}, \boldsymbol{H}) = \left[\prod_{s=1}^{N} P(D_s|M_s)P(M_s|M_0, \tau^2)\right] \times P(\boldsymbol{H})$$

Data Unknowns

Joint Posterior of all unknowns given the data

$$P(\{M_s\}, \boldsymbol{H}|\,\{D_s\}) = \frac{P(\{M_s\}, \boldsymbol{H}|\,\{D_s\})}{P(\{D_s\})}$$

(Normalisation Constant)

Putting the Bayesian in Hierarchical Bayesian

Joint Posterior of all unknowns given the data

$$P(\{M_s\}, \boldsymbol{H} \,|\, \{D_s\}) = \frac{P(\{M_s\}, \boldsymbol{H} \,|\, \{D_s\})}{P(\{D_s\})}$$

Unknowns

Data

(Ignorable Normalisation Constant)

(Posterior on N + 2 dimensional parameter space)

$$P(\{M_s\}, \boldsymbol{H} | \{D_s\}) \propto \left[\prod_{s=1}^{N} P(D_s | M_s) P(M_s | M_0, \tau^2) \right] \times P(\boldsymbol{H})$$

Measurement Likelihood

Population Dist'n / Prior

Hyperprior

Hierarchical vs Regular Bayes

- Could regard as just a general Bayesian inference problem in a very high dimensional parameter space, e.g.

$$\boldsymbol{\theta} = \{M_1, \ldots, M_N; M_0, \tau^2\} = \{\boldsymbol{M}; M_0, \tau^2\}$$

$$P(\boldsymbol{\theta}|\boldsymbol{D}) \propto P(\boldsymbol{D}|\boldsymbol{\theta})P(\boldsymbol{\theta})$$

$$P(\boldsymbol{\theta}|\boldsymbol{D}) \propto P(\boldsymbol{D}|\boldsymbol{M})P(\boldsymbol{M}|M_0, \tau^2)P(M_0, \tau^2)$$

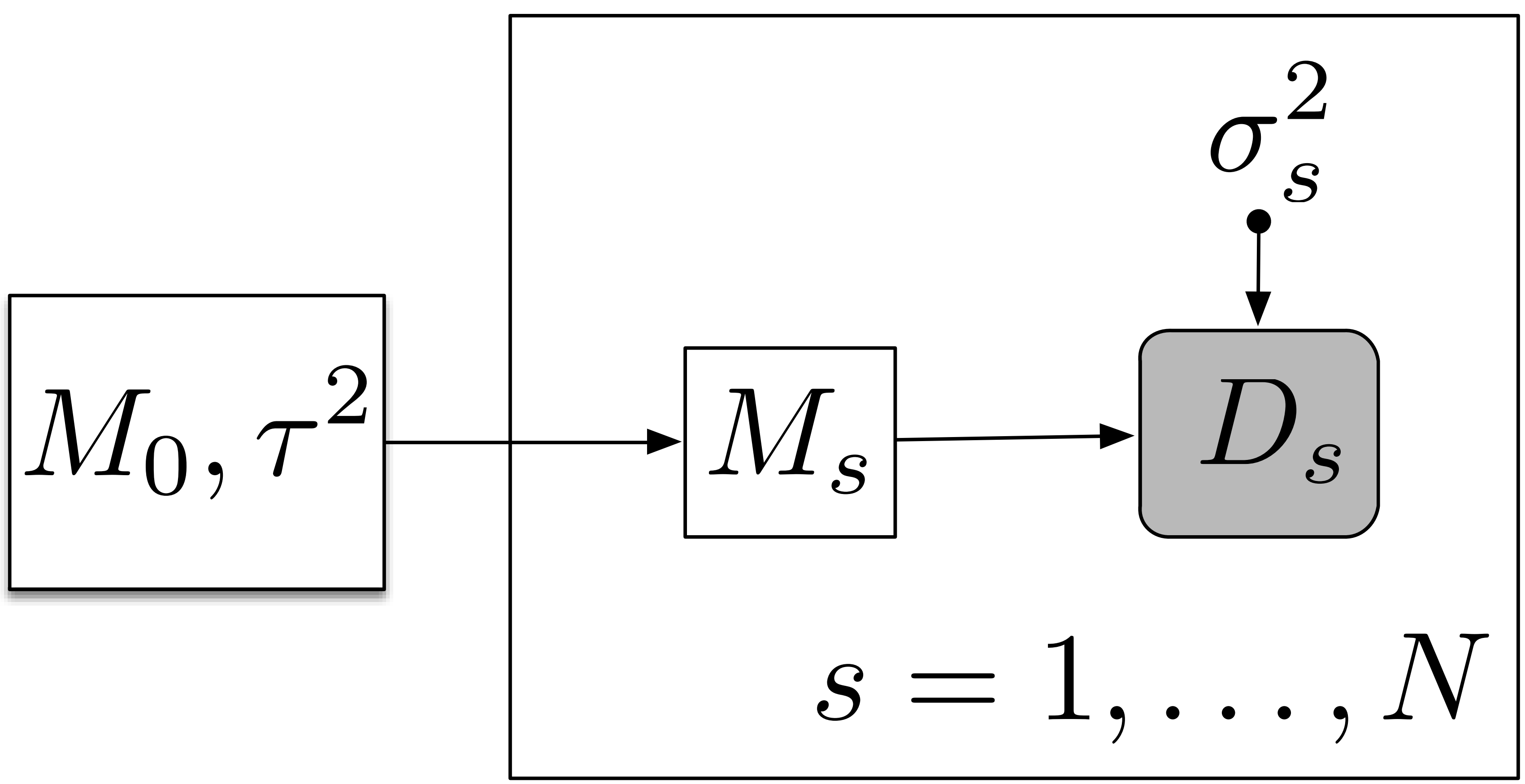
$$P(\boldsymbol{\theta}|\boldsymbol{D}) \propto \left[\prod_{s=1}^{N} P(D_s|M_s)P(M_s|M_0, \tau^2)\right] P(M_0, \tau^2)$$

- However, special hierarchical structure is useful for modelling, estimation, and computation

- For large N, wouldn't want to do an N+2 dimensional Metropolis MCMC!

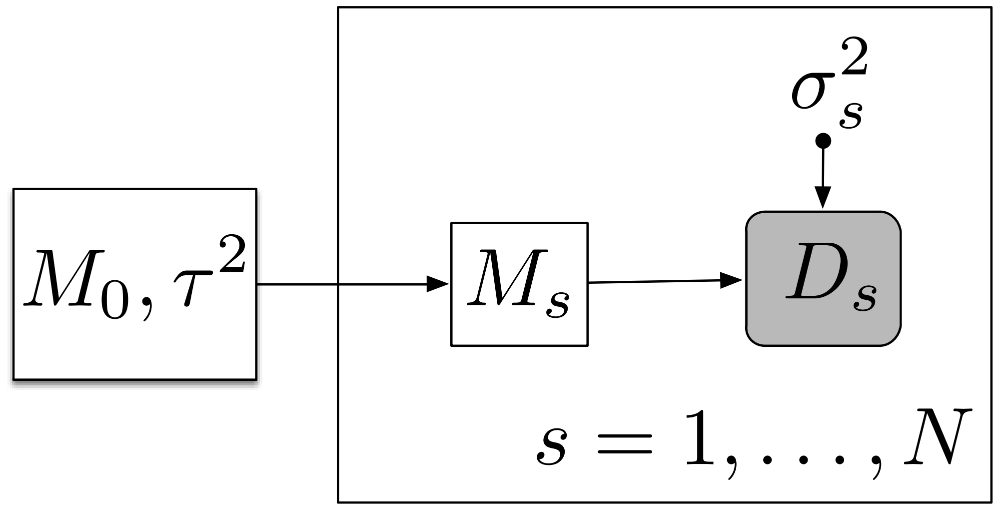
Sampling the Hierarchical Bayesian posterior

$$P(\{M_s\}, \boldsymbol{H} | \{D_s\}) \propto \left[\prod_{s=1}^{N} P(D_s | M_s) P(M_s | M_0, \tau^2) \right] \times P(\boldsymbol{H})$$

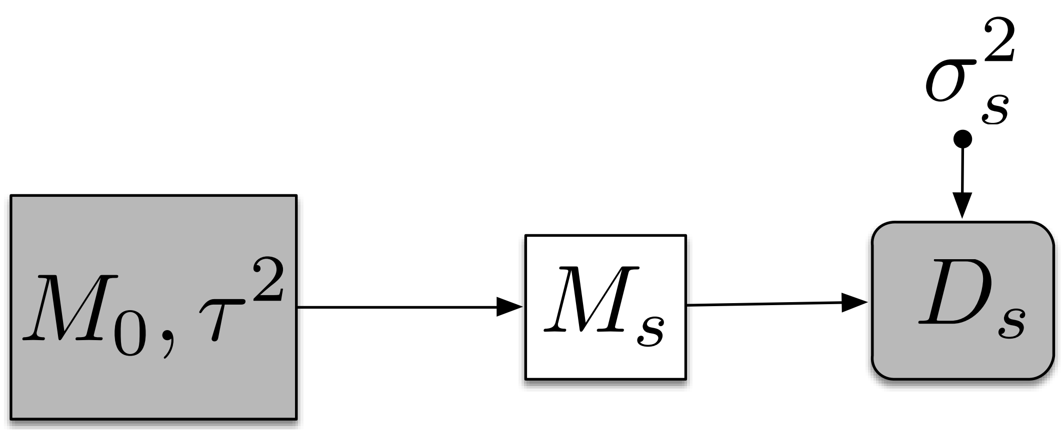


Utilise the Conditional Independence structure of PGM
to derive conditional posterior densities

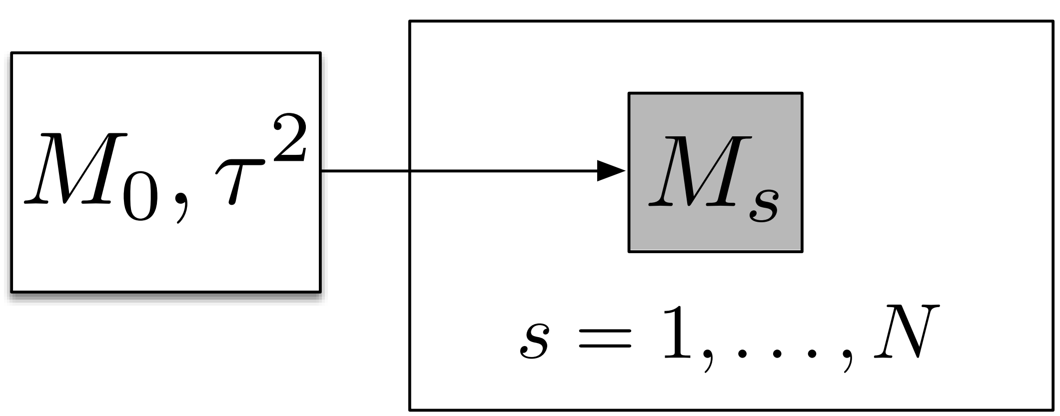
Gibbs sampling & Hierarchical Bayes



1. For s = 1 … N: Sample Latent Variables Conditional on Data and **Hyperparameters**



2. Sample Hyperparameters from Conditional on Data and Latent Variables:

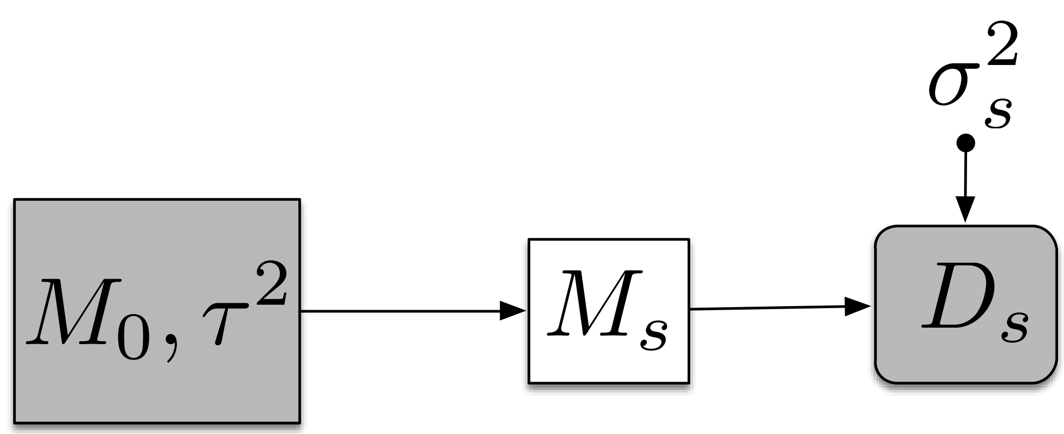


Gibbs sampling & Hierarchical Bayes

Utilises Conditional Independence structure of PGM to derive conditional posterior densities

$$P(\{M_s\}, \boldsymbol{H} | \{D_s\}) \propto \Big[\prod_{s=1}^{N} P(D_s|M_s) P(M_s|M_0, \tau^2) \Big] \times P(\boldsymbol{H})$$

1. For s = 1 … N: Sample Latent Variables Conditional on Data and **Hyperparameters**



$$P(M_s|\, M_0, \tau^2, \{D_s\}) \propto P(D_s|\, M_s) \times P(M_s|M_0, \tau^2)$$

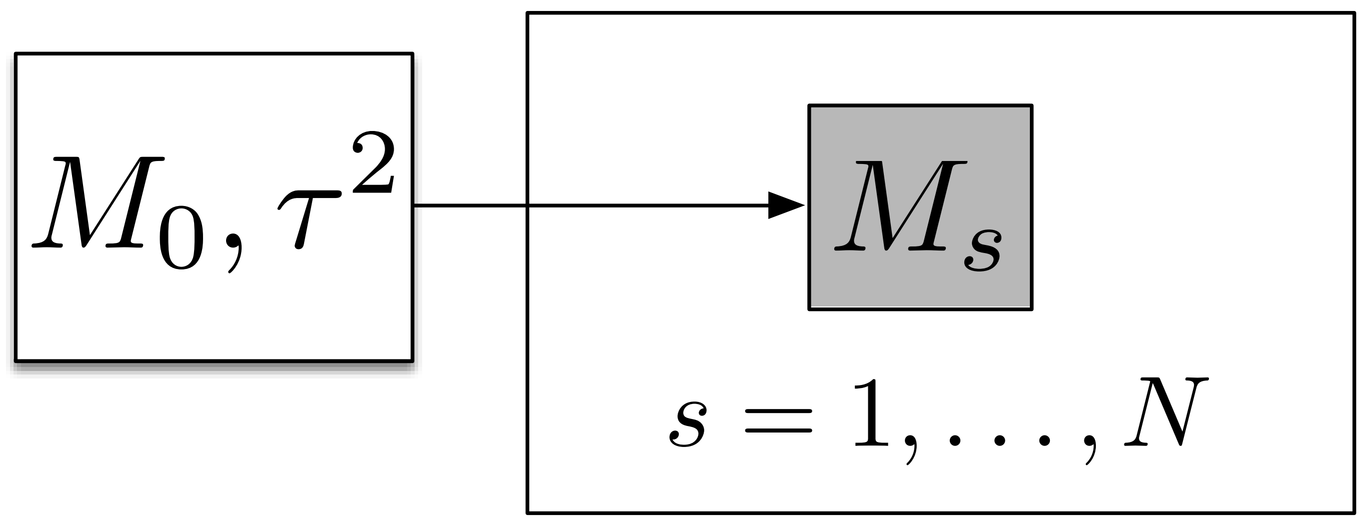
$$\propto N(D_s|M_s, \sigma_s^2) \times N(M_s|M_0, \tau^2)$$

Gibbs sampling & Hierarchical Bayes

Utilises Conditional Independence structure of PGM to derive conditional posterior densities

$$P(\{M_s\}, \boldsymbol{H} | \{D_s\}) \propto \left[\prod_{s=1}^{N} P(D_s | M_s) P(M_s | M_0, \tau^2) \right] \times P(\boldsymbol{H})$$

2. Sample Hyperparameters from Conditional on Data and Latent Variables:



$$P(M_0) \propto 1$$

Hyperprior: $P(M_0, \tau^2) = P(M_0) P(\tau^2) \propto 1$

$$P(\tau^2) \propto 1$$

$$P(M_0, \tau^2 | \{M_s\}; \{D_s\}) = P(M_0, \tau^2 | \{M_s\}) = P(M_0 | \tau^2, \{M_s\}) P(\tau^2 | \{M_s\})$$

(See Example Sheet 2, Problem 1) (Gaussian) (Inv-χ^2)

Gibbs sampling & Hierarchical Bayes

Utilises Conditional Independence structure of PGM to derive conditional posterior densities

$$P(\{M_s\}, \boldsymbol{H} | \{D_s\}) \propto \left[\prod_{s=1}^{N} P(D_s|M_s) P(M_s|M_0, \tau^2) \right] \times P(\boldsymbol{H})$$

2. Sample Hyperparameters from Conditional on Data and Latent Variables:

$$P(M_0, \tau^2 | \{M_s\}; \{D_s\}) = P(M_0, \tau^2 | \{M_s\}) = P(M_0 | \tau^2, \{M_s\}) P(\tau^2 | \{M_s\})$$

(Gaussian) (Inv-χ^2)

Hyperprior: $P(M_0, \tau^2) \propto 1$

Draw from: $$\tau^2 | \{M_s\} \sim \text{Inv-}\chi^2 \left(N-3, \frac{(N-1)}{(N-3)} s^2 \right)$$

$$M_0 | \tau^2; \{M_s\} \sim N(\bar{M}, \tau^2/N)$$

$$\bar{M} \equiv \frac{1}{N} \sum_{s=1}^{N} M_s \qquad s^2 \equiv \frac{1}{N-1} \sum_{s=1}^{N} (M_s - \bar{M})^2$$