Astrostatistics: 11 Mar 2019

https://github.com/CambridgeAstroStat/PartIII-Astrostatistics-2019

- Make-Up Lecture, Thu 14 Mar, 11am-12: MR 9 ? 14 ? 15
- Example Class, Fri 15 Mar 1pm MR 12 ?
- Today:
- Continue Probabilistic Graphical Models (Bishop, Ch 8)
- Hierarchical Bayes & Gibbs Sampling

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

Forward Model:

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

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

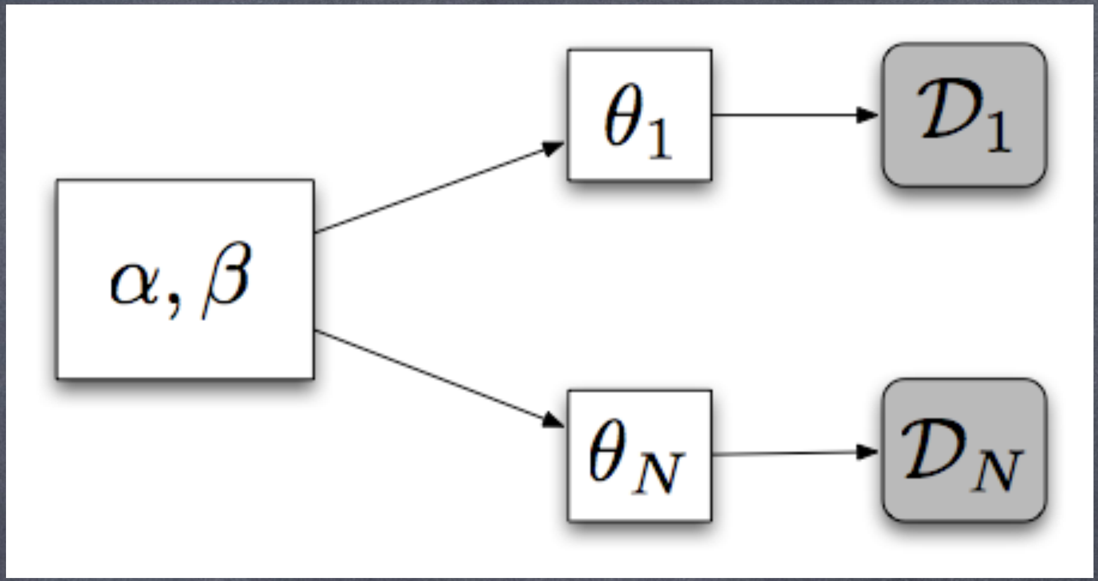
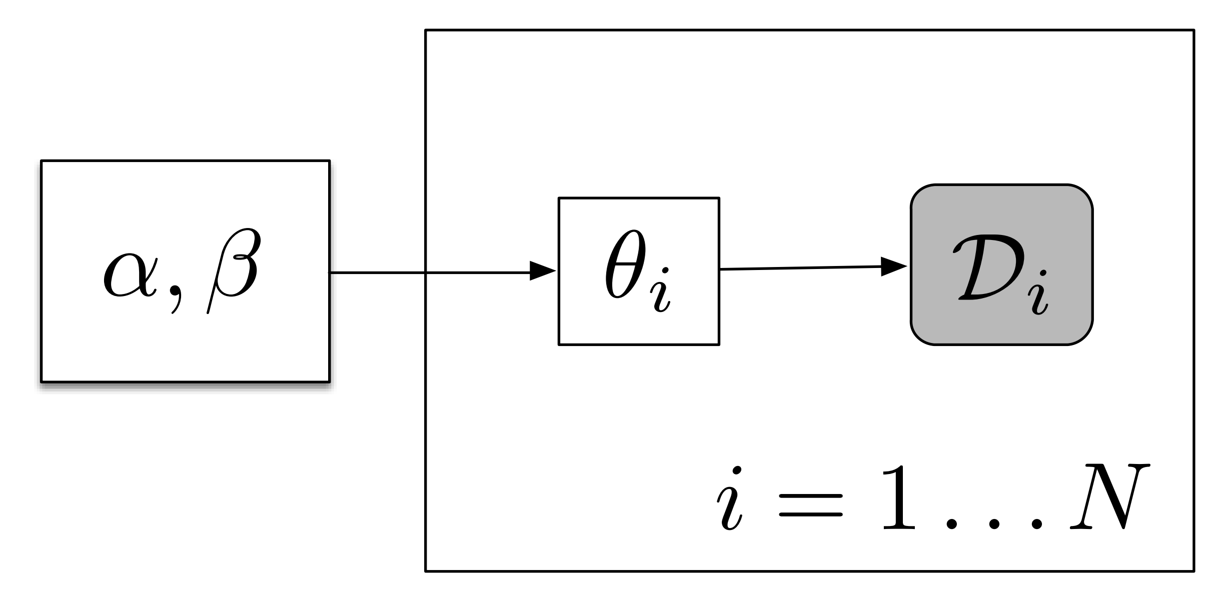


Plate Notation: (loop over individuals in sample)



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

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 (from Cond'l independence 1)

<u>PGM Slides:</u>
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)

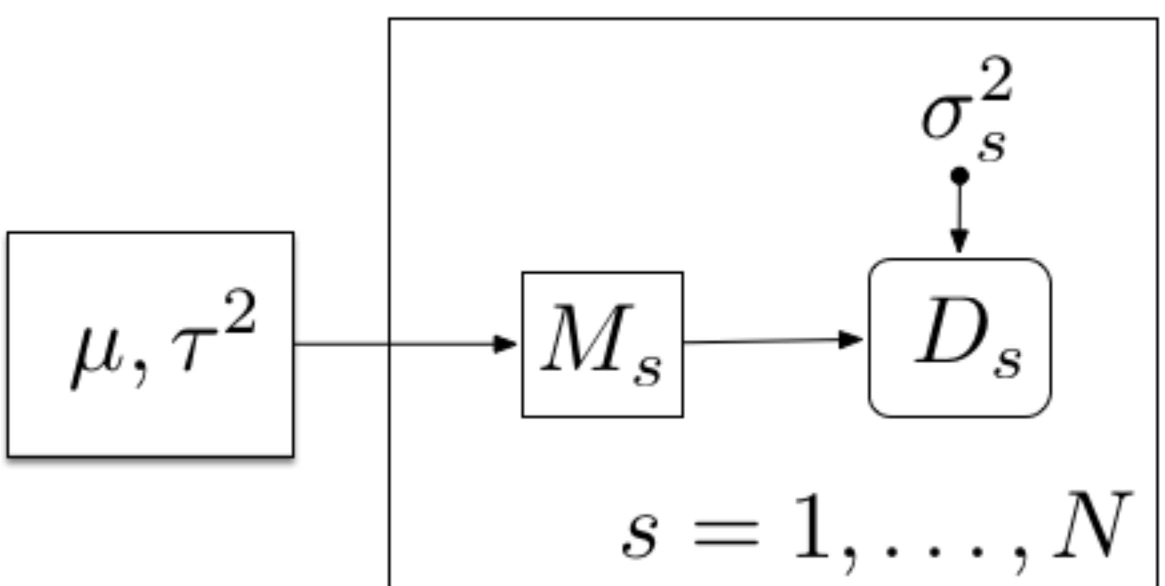
$$M_s \sim N(M_0, \tau^2) \quad s = 1 \dots N$$

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

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

(Derive on board)

Measurement Likelihood

Population Dist'n / Prior

Hyperprior

Putting the Bayesian in Hierarchical Bayesian

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

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

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

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**

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**

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

(Gaussian) (Gaussian)

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

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

(Conditional Independence) (Gaussian) (Inv-χ^2)

Reduces to the Familiar Posterior for unknown mean and variance of Gaussian Data (Example Sheet 1, Prob 4.2)

Gibbs sampling & Hierarchical Bayes

Utilises Conditional Independence structure of PGM to derive conditional posterior densities

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

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

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

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

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

Gibbs Sampling in 102 parameters: Code/Data Demo