Lecture 13
Fundamentals of Bayesian Inference

Dennis Sun
Stats 253

August 11, 2014

Outline of Lecture

1. Bayesian Models
2. Modeling Correlations Using Bayes
3. The Universal Algorithm
4. BUGS
5. Wrapping Up

Where are we?

Frequentist Models

- The perspective taken in this class has been **frequentist**:
  - Model has some parameters $\boldsymbol{\theta}$, which are unknown but fixed.
  - We estimate them by maximum likelihood, i.e., choosing $\boldsymbol{\theta}$ to maximize $p(\boldsymbol{y}|\boldsymbol{\theta})$.
- Examples
  - AR processes: $y_i - \mu_i = \phi \sum_j w_{ij}(y_j - \mu_j) + \epsilon_i$.
  - Kriging: variogram model $\gamma_{\boldsymbol{\theta}}(h)$

Bayesian Models

In the **Bayesian** perspective, $\boldsymbol{\theta}$ is random.

- Now it makes sense to talk about a *prior* $p(\boldsymbol{\theta})$ and *posterior* $p(\boldsymbol{\theta}|\boldsymbol{y})$.
- $p(\boldsymbol{\theta}|\boldsymbol{y}) = \dfrac{p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{y})} = \dfrac{p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})}{\int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$. (Bayes' rule)
- This tells us how probable all values of $\boldsymbol{\theta}$ are!
- If we must reduce to a single summary:
  - maximum *a posteriori* (MAP): choose $\boldsymbol{\theta}$ that maximizes $p(\boldsymbol{\theta}|\boldsymbol{y})$.
  - posterior mean: calculate $\mathrm{E}(\boldsymbol{\theta}|\boldsymbol{y})$. (This is the MMSE estimator; it minimizes $\mathrm{E}||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}||^2$.)

Why Bayes?

1. Provides an intuitive way to model correlations in data.
2. There is a "universal" algorithm for calculating the posterior.

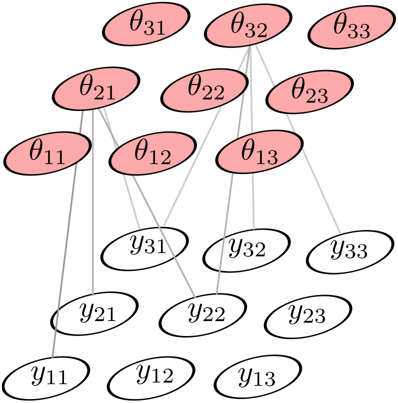
Bayesian methods are taking spatio-temporal modeling by storm!

Why Not Bayes?

1. Computationally much slower than frequentist approaches.
2. Requires specification of prior $p(\boldsymbol{\theta})$. (The problem is not so much that it's subjective as that it could be terribly wrong.)

Where are we?

Bayesian Lattice



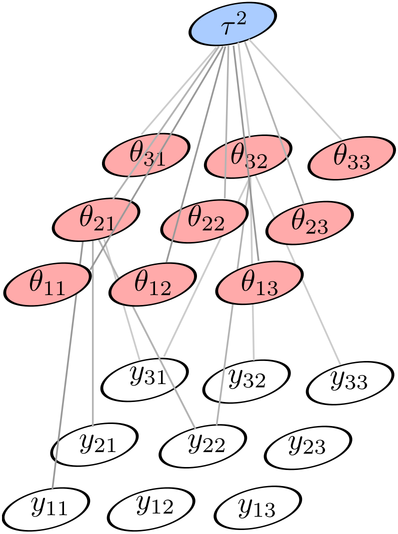
Consider the model

$$y_{ij} = \sum_{(k,\ell)\in N(i,j)} \theta_{k\ell} + \epsilon_{ij}.$$

with $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$.

- Suppose θ_{ij} fixed. Is there dependence between y_{ij} and $y_{k\ell}$?
- What if $\theta_{ij} \sim N(0, \tau^2)$?

Hierarchical Bayesian Models



$$\theta_{ij} \sim N(0, \tau^2)$$

- What if we don't know τ^2?
- Put a prior on it, e.g.,

$$\tau^2 \sim \text{Gamma}(\alpha, \beta)$$

- Inferences will be less sensitive to choice of α, β than to τ.
- Can also put priors on α, β, etc.
- At some level, have to make arbitrary choice of prior.

A Real (Non-Spatial) Example

Rater models: y_{ij} is the rating user i gives to item j:

$$y_{ij} = \theta_i + \alpha_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

e.g., users rating movies, students grading peers, etc.

In peer grading, student i grades response j (response of student k to question ℓ). Example model:

$$\begin{aligned} y_{ij} \mid \theta_i, \alpha_j, \sigma^2 &\sim N(\theta_i + \alpha_j, \sigma^2) \\ \theta_i \mid \mu, \tau^2 &\sim N(\mu, \tau^2) \\ \alpha_{k\ell} \mid \gamma_k, \delta_\ell, \nu^2 &\sim N(\gamma_k + \delta_\ell, \nu^2) \end{aligned}$$

- Why might an instructor want to know $\alpha_{k\ell}$?
- What would $p(\alpha_{k\ell} | \boldsymbol{y})$ tell us?

Where are we?

A Universal Algorithm

- Recall: the goal of Bayesian inference is to "calculate" the posterior $p(\boldsymbol{\theta}|\boldsymbol{y})$.
- We have seen that hierarchical models are rich and expressive.
- What makes Bayesian inference attractive is that there is a universal way to estimate the posterior in hierarchical models.

Gibbs Sampler

Recall: the Gibbs sampler is a general way to sample from $p(x_1, ..., x_n)$ by starting with $X_2, ..., X_n$ and iteratively sampling

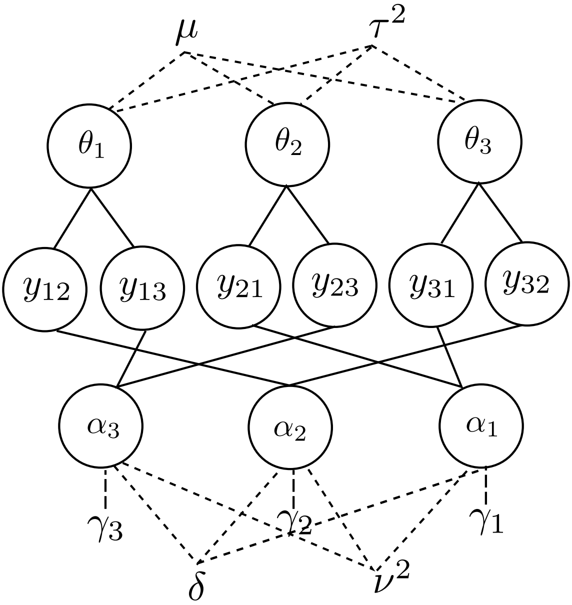
- $X_1 \sim p(x_1 | X_2, X_3, ..., X_n)$
- $X_2 \sim p(x_2 | X_1, X_3, ..., X_n)$
- ...
- $X_n \sim p(x_n | X_1, X_2, ..., X_{n-1})$.

As the number of cycles increases, the distribution of $(X_1, ..., X_n)$ approaches the desired distribution $p(x_1, ..., x_n)$.

Gibbs Sampler for Bayesian Inference

How do we use the Gibbs Sampler for Bayesian inference?

- Here, the goal is to sample from $p(\boldsymbol{\theta}|\boldsymbol{y})$.
- The Gibbs sampler says that we should sample one parameter at a time, conditional on the others and the data:
  - $\theta_1 \sim p(\theta_1|\boldsymbol{y}, \theta_2, \theta_3, ..., \theta_n)$
  - $\theta_2 \sim p(\theta_2|\boldsymbol{y}, \theta_1, \theta_3, ..., \theta_n)$
  - ...
  - $\theta_n \sim p(\theta_n|\boldsymbol{y}, \theta_1, \theta_2, ..., \theta_{n-1})$
- Why is this "easy" for hierarchical models?


μ
τ^2
θ_1
θ_2
θ_3
y_{12}
y_{13}
y_{21}
y_{23}
y_{31}
y_{32}
α_3
α_2
α_1
γ_3
γ_2
γ_1
δ
ν^2

Summary

- The Gibbs sampler allows us to only have to ever worry about "local" dependencies.
- Typically, the distribution of a parameter given its "neighbors" is simple, and it is easy to simulate from it.
- No one is really sure how long you have to run a Gibbs sampler before it converges....

Where are we?

Bayesian Software

There are several Bayesian packages that allow you to specify a hierarchical model, and it will simulate from the posterior for you!

- BUGS (Bayesian inference Using Gibbs Sampling): Windows and Linux only
- JAGS (Just Another Gibbs Sampler): all platforms
- Stan (uses Hamiltonian Monte Carlo): all platforms

All of these are integrated with R.

Demo

Let's do a demo of JAGS.

Where are we?

Recap

- Bayesian inference is a powerful alternative to frequentist inference.
- In particular, it makes hierarchical modeling easy because the Gibbs sampler provides a universal algorithm for simulating from the posterior.
- Because the Gibbs sampler is universal, it is possible to develop software that automatically simulates from the posterior for *any* hierarchical models.
- However, Bayesian methods are typically much slower than their frequentist counterparts.

Administrivia

- Please turn in Homework 4a and/or 4b.
- Please turn in any resubmits by the end of the week.
- Final projects due next Monday.