Lecture 9
The Bayesian Paradigm

Dennis Sun
Stanford University
Stats 253

July 13, 2015

1. Frequentist Inference and Its Discontents
2. Bayesian Models
3. Bayesian Kriging
4. Bayesian CAR Models?

1. Frequentist Inference and Its Discontents

2. Bayesian Models

3. Bayesian Kriging

4. Bayesian CAR Models?

The Trouble with Frequentist Inference

- So far, the approach we've taken has been frequentist.
- The defining feature of frequentist inference is that parameters like $\boldsymbol{\beta}$, $\boldsymbol{\phi}$, and $\boldsymbol{\theta}$ are unknown, but fixed.
- Quick quiz: Suppose a 95% confidence interval for β_j is $(1.0, 3.9)$. What does this mean?
- For inference, we typically use likelihood-based methods, which work when the data is independent.
- In temporal and spatial settings, the data are not independent, so likelihood-based methods do not work. In addition, the likelihood is often intractable.

Bottom line: Inference in frequentist models is difficult, especially when you move away from Gaussian distributions.

The Bayesian Paradigm

- Bayesian inference regards the parameters as random. We put **prior distributions** on parameters, like $p(\boldsymbol{\beta})$ or $p(\boldsymbol{\theta})$.
- Inference is based on the **posterior distribution**, which can be computed using Bayes' rule:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}.$$

- There is a natural estimator of $\boldsymbol{\theta}$, the posterior mean $\mathrm{E}[\boldsymbol{\theta}|\mathbf{y}]$. We can also quantify our uncertainty using $\mathrm{Var}[\boldsymbol{\theta}|\mathbf{y}]$ and obtain intervals using $p(\boldsymbol{\theta}|\mathbf{y})$.
- Moreover, there is a universal algorithm for obtaining the posterior.

A Simple Example

- First, consider the frequentist model:

$$y_i = \theta + \epsilon_i,$$

  where ϵ_i are i.i.d. $N(0, \sigma^2)$. θ is fixed.

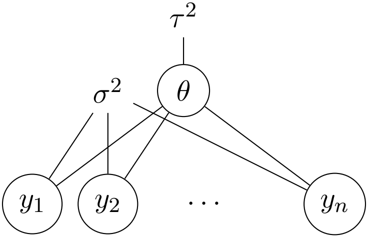
  Notice that $\mathrm{Cov}(y_i, y_j) = 0$ for $i \neq j$.

- Now what if we put a $N(0, \tau^2)$ prior on θ?

  Notice that $\mathrm{Cov}(y_i, y_j) = \mathrm{Cov}(\theta + \epsilon_i, \theta + \epsilon_j) = \tau^2$.

The prior induces a correlation among our observations.

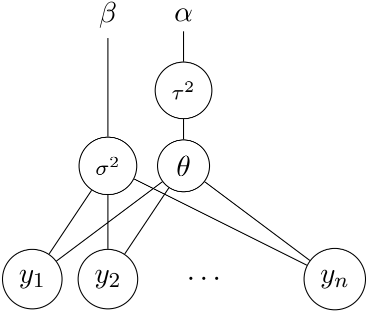
Drawing a Bayesian Model



An edge indicates that two random variables are conditionally dependent, given all the other variables in the model.

Hierarchical Bayesian Models

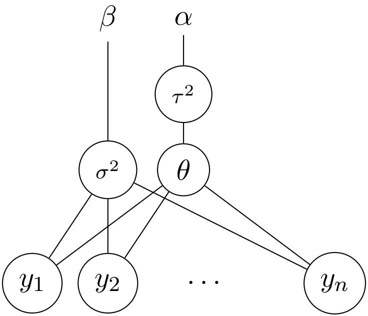
What if we don't know τ^2 or σ^2? Put priors on them too!



At some level, we have to assume that the distribution is completely known.

In general, Bayesian inferences are less sensitive to the choice of these **hyperpriors**, so add a hyperprior if you're not sure.

Bayesian Machinery



The beauty of Bayes is that there is a universal algorithm that goes from any model of this form and returns (samples from) the posterior:

$$p(\theta, \sigma^2, \tau^2 | y_1, ..., y_n).$$

JAGS Demo

There are a number of Bayesian modeling tools, including WinBUGS, OpenBUGS, JAGS, and STAN.

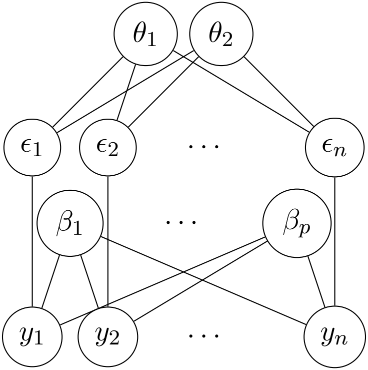
They all work in a similar way:

- Specify a model in a file.
- Call that file using a package in `R` (or another language).

I will do a demo in JAGS. This requires:

- installing JAGS (`http://mcmc-jags.sourceforge.net/`), which is a command-line utility.
- installing the `rjags` package in `R` to interface with JAGS.

Bayesian Kriging



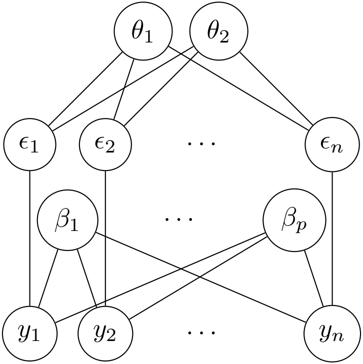
$$\boldsymbol{\epsilon}|\boldsymbol{\theta} \sim N(\mathbf{0}, \Sigma(\boldsymbol{\theta}))$$

$$\mathbf{y}|\boldsymbol{\epsilon}, \boldsymbol{\beta} \sim N(X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tau^2 I)$$

In the frequentist literature, $\boldsymbol{\epsilon}$ is called a random effect and such models are called **mixed effects models**.

Generalizing to Non-Gaussian Data

This generalizes naturally to non-Gaussian observations.



$$\boldsymbol{\epsilon} | \boldsymbol{\theta} \sim N(\mathbf{0}, \Sigma(\boldsymbol{\theta}))$$

$$\mathrm{E}[\mathbf{y} | \boldsymbol{\epsilon}, \boldsymbol{\beta}] = f(X\boldsymbol{\beta} + \boldsymbol{\epsilon})$$

where y_i are observations from an exponential family. This is an example of a **generalized linear mixed model**.

Fitting this Model

The `geoR` and `geoRglm` packages provide functions for Bayesian kriging.

- `krige.bayes` in `geoR` fits a Gaussian kriging model.
- `binom.krige` and `binom.krige.bayes` in `geoRglm` fit a binary kriging model.
- `pois.krige` and `pois.krige.bayes` in `geoRglm` fit a Poisson kriging model.

A Bayesian CAR Model?