Bayesian Methods for Epidemiologic Data Analysis:
Hierarchical Models

Rich MacLehose

National Institute of Environmental Health Sciences

June 18, 2007

Outline

Outline

Outline

Outline

1 Introduction

2 Hierarchical Regression Models

3 Missing Data

What makes a model hierarchical?

Definition

The joint distribution, (such as $f(y, \beta, \mu, \tau^2)$) can be written as a series of conditional distributions

Basic Bayesian regression

$f(y, \beta, \mu, \tau^2) = f(y \mid \beta) f(\beta \mid \mu, \tau^2)$

What makes a model hierarchical?

- The parameters of the model are arranged in levels
- Parameters at one level are dependent on parameters of higher levels
- In the basic Bayes logistic regression example
  - y_i depends on β
  - β depends on μ, τ^2
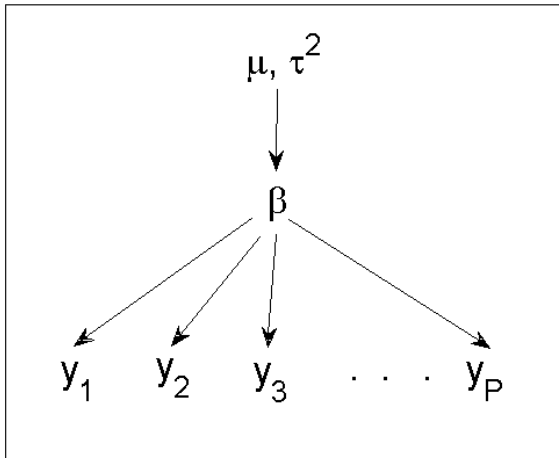
Breast Cancer in North Carolina

Hierarchical models are a natural way to formulate many
applied problems

Breast Cancer in North Carolina

Hierarchical models are a natural way to formulate many applied problems

- Want to estimate county-specific breast cancer rates in North Carolina in 2004
- y_i is the number of cancer cases in the ith county
- N_i is the population of each county in 2004
- Standard maximum likelihood techniques provide a simple answer $\hat{\theta}_i = y_i/N_i$

Breast Cancer in North Carolina

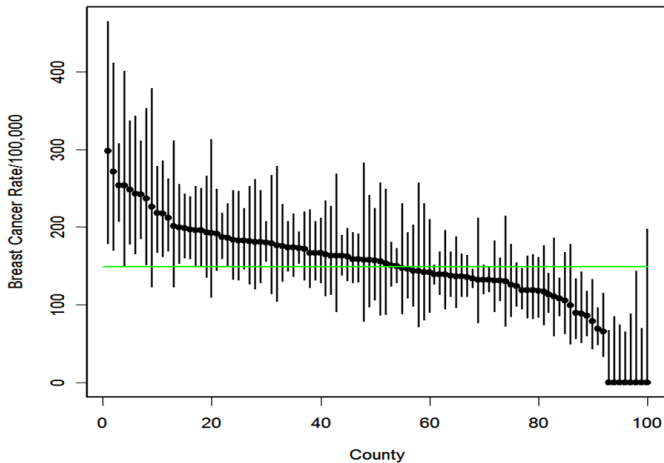Hierarchical models are a natural way to formulate many applied problems

- Want to estimate county-specific breast cancer rates in North Carolina in 2004
- y_i is the number of cancer cases in the ith county
- N_i is the population of each county in 2004
- Standard maximum likelihood techniques provide a simple answer $\hat{\theta}_i = y_i/N_i$
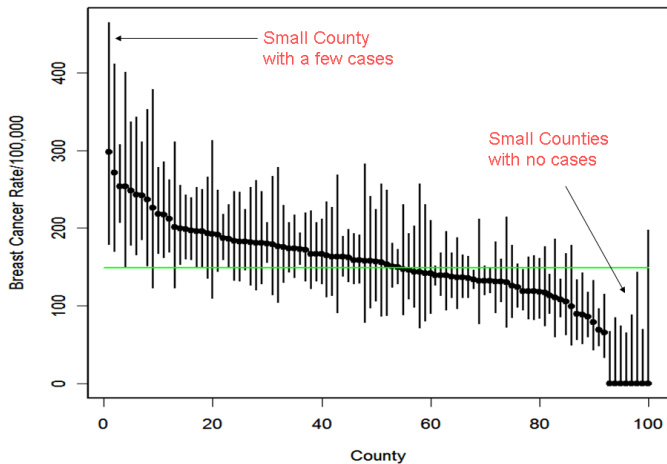
But is it a reasonable answer?

- Some of the point estimates are absurdly high or low
- Some of the confidence intervals are huge

How can we improve on this?

- We have some prior knowledge about BrCa rates
- Lets incorporate it through a hierarchical model

# Hierarchical Modeling of Cancer Rates

Model

$$y_i \sim \text{Poisson}(N_i \exp(\theta_i))$$
$$\theta_i \sim \text{Normal}(\mu, \tau^2)$$

- Poisson model with different intercepts for each county
- $\exp(\theta_i)$ is the county-specific rate
- N_i is commonly called the offset term

counties.txt contains the data for this example and counties.R has the code

A quick change in software package

- SAS has really nice Bayes capabilities and will probably get even better shortly
- But it's new, and you can reach its limitations quickly
- The BUGS family - winBUGS, openBUGS, etc - are very adaptable to most epidemiology applications (and its free)
- Using BUGS directly is a pain, but you can call it through R easily
- I won't bore you with an introduction to this software
  - I've posted a brief tutorial online, with code and data for all the examples
  - http://dir.niehs.nih.gov/dirbb/serbayes/
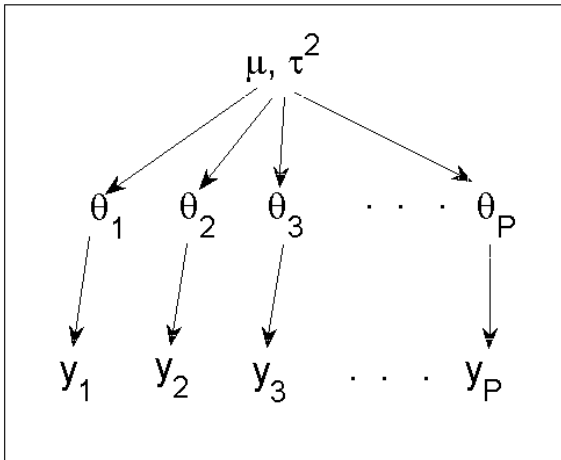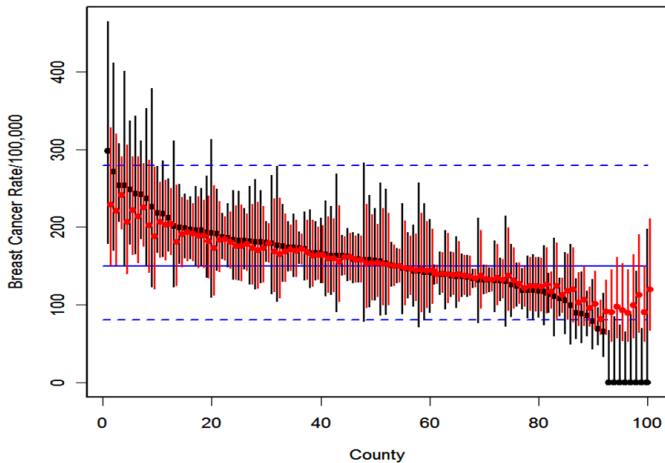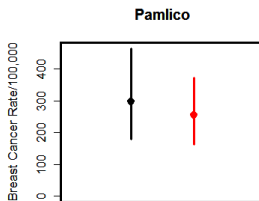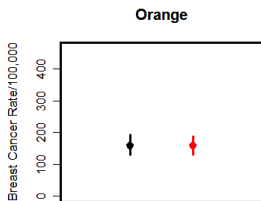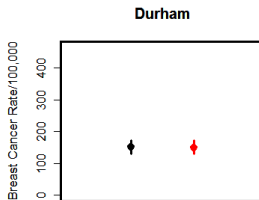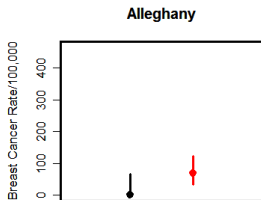
Key Properties of Hierarchical Models

Shrinkage

- Estimates are pulled away from their unbiased MLE and toward the prior mean
  - Bayesian: natural consequence of combining prior with data
  - Frequentist: introduce bias to reduce MSE (biased but more precise)
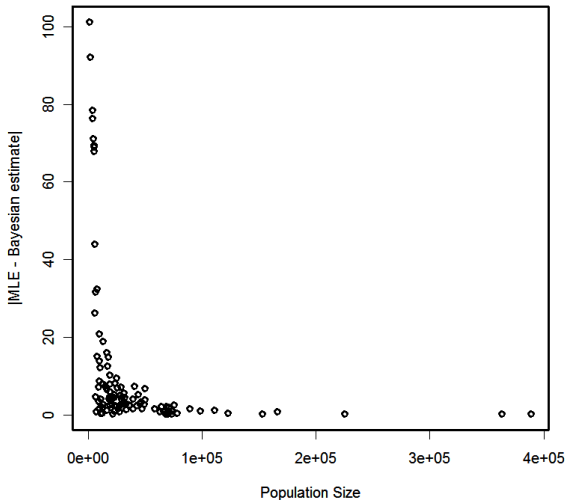
Pro's and Con's

Pro's

- Works really well - bias/variance trade off
- Incorporates prior knowledge
- Certain problems that are very difficult for frequentist inference are easily solved in Bayesian hierarchical models

    - Non-parametrics, change point problems, spatial data

Con's

- More difficult to program
- SAS doesn't have a proc hierarchical (yet)

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Outline

1 Introduction

2 Hierarchical Regression Models

3 Missing Data

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Hierarchical Regression

- Incorporating prior knowledge and shrinking estimates is very appealing
- Want to incorporate variable effects in hierarchical regression models
    - Logistic regression, though linear or poisson are easy extensions
- Motivating example of DBP's and SAB
    - Simulated data set and code provided

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Spontaneous Abortion (SAB)

- Pregnancy loss prior to 20 weeks gestation
- Very common ($> 30\%$ of all pregnancies)
- Relatively little known about its causes
  - maternal age, smoking, prior pregnancy loss, occupational exposures, caffeine

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Spontaneous Abortion (SAB)

- Pregnancy loss prior to 20 weeks gestation
- Very common ($> 30\%$ of all pregnancies)
- Relatively little known about its causes
  - maternal age, smoking, prior pregnancy loss, occupational exposures, caffeine
  - disinfection by-products (DBPs)?

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Disinfection By-Products

- A vast array of DBPs are formed in the disinfection process
- We focus on 2 main types:

Trihalomethanes (THMs)

$CHCl_3$, $CHBr_3$, $CHCl_2Br$, $CHClBr_2$

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Disinfection By-Products

- A vast array of DBPs are formed in the disinfection process
- We focus on 2 main types:

Trihalomethanes (THMs)

$CHCl_3$, $CHBr_3$, $CHCl_2Br$, $CHClBr_2$

Haloacetic Acids (HAAs)

ClAA, Cl_2AA, Cl_3AA, BrAA, Br_2AA, Br_3AA, BrClAA, Br_2ClAA, $BrCl_2AA$

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

DBPs and SABs

Early Studies

- Noted an increased risk of SAB with increased tap-water consumption

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

DBPs and SABs

Early Studies

- Noted an increased risk of SAB with increased tap-water consumption

More Recent Studies

- Increased risk of SAB with exposure to THMs
- Notably, $CHBrCl_2$ in Waller et al (1998)
  - OR=2.0 (1.2, 3.5)

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Specific Aim

To estimate the effect of each of the 13 constituent DBPs (4 THMs and 9 HAAs) on SAB

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

RFTS - briefly

- 2507 enrolled in three metropolitan areas in U.S.
- Years: 2001-2004
- Recruitment
  - Prenatal care practices (52%)
  - Health departments (32%)
  - Promotional mailings (3%)
  - Drug stores, referrals, etc. (13%)

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Data Collection

Baseline Interview

demographic information, medical history, other confounders

Pregnancy loss

self report or chart abstraction

DBP concentration

- Disinfecting utilities
- Weekly samples at two sites with high DBPs
- Every other week at third site with low DBPs

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
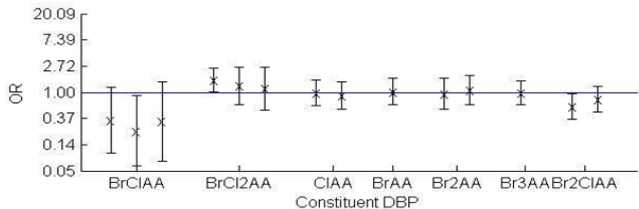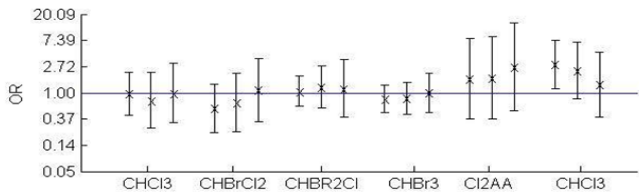Model 2 (fully Bayes)

Logistic Model

- Standard logistic analysis
- Categorize DBP variables (quartiles,tertiles)

$$\text{logit}\{\Pr(y_i = 1 \mid \cdot)\} = \alpha + \gamma_1 z_{1i} + \cdots + \gamma_p z_{pi} + \\ \beta_1 x_{1i} + \cdots + \beta_{32} x_{32i}$$

Where,

- α is the intercept
- $z_{1i} \ldots z_{pi}$ are confounders: smoking, alcohol use, ethnicity, maternal age
- x_{ki} is the concentration of the kth category of DBP for the ith individual for the time interval around conception

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Results of frequentist analysis

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Results of frequentist analysis

- Several large but imprecise effects are seen
- 4 of 32 coefficients are statistically significant
- Imprecision makes us question results

Is there a better analytic approach?

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Other common options

- Try all exposure in one model
  - Problem: unreliable estimates

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Other common options

- Try all exposure in one model
    - Problem: unreliable estimates
- Combine variables in aggregate scores
    - Problem: difficult to interpret, can mask effects

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Other common options

- Try all exposure in one model
  - Problem: unreliable estimates
- Combine variables in aggregate scores
  - Problem: difficult to interpret, can mask effects
- Analyze one variable at a time
  - Problem: uncontrolled confounding

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Why choose a Bayes hierarchical model here?

- High correlation in the data
- There is some previous knowledge to be brought in
- Interpretability

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Model 1

- A simple two-level hierarchical model popularized by Greenland
- Have seen use in nutritional, genetic, occupational, and cancer epidemiology
- Despite the name, they are Bayesian models.
  - name may refer to asymptotic methods commonly used in fitting semi-Bayes models

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Shrinkage of Regression Coefficients

- Regression coefficients are shrunk toward prior distribution as in BrCa example
- Amount of shrinkage depends on prior variance
  - Large variance $\rightarrow$ little shrinkage
  - Small variance $\rightarrow$ lots of shrinkage

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Shrinkage in model 1

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)
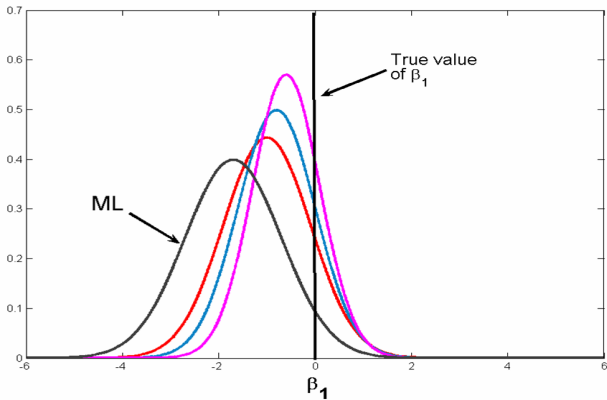
Model

$$\text{logit}\{\Pr(y_i = j \,|\cdot)\} = \alpha + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \gamma_1 z_{1i} \ldots \gamma_p z_{pi}$$
$$\beta_j \sim N(\mu, \phi^2)$$

- α = intercept
- γ's = confounder effects
- β's = exposure effects

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Prior specification

- Little prior evidence of effect: specify $\mu = 0$
- Calculate ϕ^2 from existing literature
  - Largest observed effect: OR=3.0
  - $\phi^2 = (ln(3) - ln(1/3))/(2 \times 1.96) = 0.3142$

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Model 1 - Results

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)
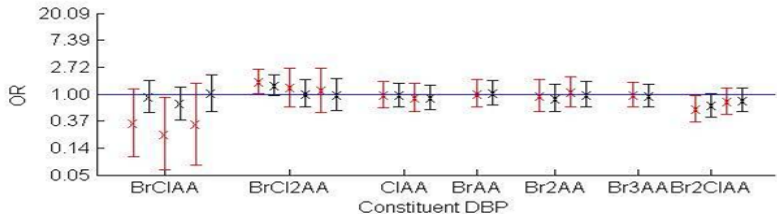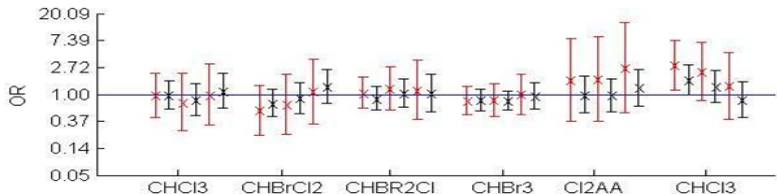
Can we improve on the semi-Bayes model?

- Assumes the prior variance, ϕ^2, is known with certainty
  - constant shrinkage of all coefficients
- Sensitivity analyses address changes to results with different prior variances
- Data contain information on prior variance

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Model 2

- Places prior distribution on ϕ^2
  - reduces dependence on prior variance

$$
\begin{aligned}
\text{logit}\{\Pr(t_i = j \,|\, \cdot)\} &= \alpha + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \gamma_{1i} z_{1i} + \cdots + \gamma_{pi} z_{pi} \\
\beta_j &\sim N(\mu, \phi^2) \\
\phi^2 &\sim IG(\alpha_1, \alpha_2)
\end{aligned}
$$

- $\mu = 0$
- ϕ^2 is random. Choose $\alpha_1 = 3.39, \alpha_2 = 1.33$
- $E(\phi^2) = 0.31$ (as in model 1)
- $V(\phi^2) = 0.07$ (at the 95th percentile of ϕ^2, 95% of β's will fall between OR=6 and OR=1/6 ... the most extreme results we believe possible)

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

What's an inverse-Gamma distribution?

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)
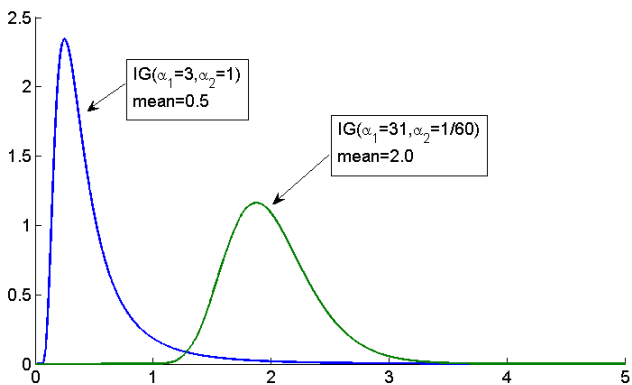
Properties of model 2

- Prior distribution on ϕ^2 allows it to be updated by the data
- As variability of estimates from prior mean increases, so does ϕ^2
- As variability of estimates from prior mean decreases, so does ϕ^2
- Adaptive shrinkage of all coefficients

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Adaptive Shrinkage of model 2

Model	Prior variance	ϕ^2, Data	Shrinkage
SB	Fixed	Constant	Constant
FB	Random	$\downarrow$	$\uparrow$

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Adaptive Shrinkage of model 2

Model	Prior variance	ϕ^2, Data	Shrinkage
SB	Fixed	Constant	Constant
FB	Random	↑	↓

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

Model 2 - Results

Introduction
Hierarchical Regression Models
Missing Data

Motivating Example: DBPs and SAB
Right From the Start
Frequentist Analysis
Hierarchical Model 1 (semi-Bayes)
Model 2 (fully Bayes)

The Problem with Model 2

- How sure are we of our parametric specification of the prior?
- Can we do better by grouping coefficients into clustering and then shrinking the cluster specific coefficients separately?
  - Amount of shrinkage varies by coefficient
- see MacLehose et al. 2007

Outline

Missing Data

- Missing data are omnipresent
- The best solution to missing data is to avoid it by good study design
- When that's not possible, you can begin to address missing data through proper analysis

Various Missing Data Techniques

The Good

Bayesian modeling, E-M algorithm, Multiple Imputation, Weighted Estimating Equations

The So-So

Hot deck imputation, Improper imputation

The Ugly

Replacing missing values with sample mean

Missing Data Terminology

Missing **C**ompletely **A**t **R**andom

- Data is missing based on a coin flip
- Lightning struck your sample
- Complete case analysis is unbiased here

Example

$$\text{Logit}\{Pr(y_i = 1)\} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2}$$
$$\text{Logit}\{Pr(r_i = 1)\} = \lambda_0$$

where r_i is an indicator for whether the x_{i1} is missing

Missing Data Terminology

Missing **A**t **R**andom

- The data *DO* contain information about why some values are missing
  - e.g., Diabetic men don't report dietary variables as often

Complete case analysis?

- If the missingness depends on the outcome, complete case analysis may be biased
- If the missingness depends only on other covariates, complete case analysis is not biased
  - However, you want to include covariates that cause missingness in the regression

Missing Data Terminology

Missing **A**t **R**andom

- The data *DO* contain information about why some values are missing
  - e.g., Diabetic men don't report dietary variables as often

Example

$$\text{Logit}\{Pr(y_i = 1)\} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2}$$
$$\text{Logit}\{Pr(r_i = 1)\} = \lambda_0 + \lambda_1 y_i + \lambda_2 x_{i2}$$

Where r_i is an indicator for whether the x_{i1} is missing

Missing Data Terminology

Not Missing at Random

- The missing data itself contains information about why some values are missing
  - e.g., Men with certain dietary habits don't report those dietary variables as often
- Complete case analysis is not OK here

Example

$$\text{Logit}\{Pr(y_i = 1)\} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2}$$
$$\text{Logit}\{Pr(r_i = 1)\} = \lambda_0 + \lambda_1 x_{i1} + \lambda_2 y_i$$

where r_i is an indicator for whether the x_{i1} is missing

Missing Outcome Data

- If the outcome is MCAR, you get unbiased estimates
- If the outcome is MAR and you're doing likelihood inference, you get unbiased estimates
    - Beware GEE in this case
- The data won't provide any information outside of what they're already telling you in the regression model
- it its not missing at random, you still need to worry

Hierarchical Analysis of Missing Data

Bayesian analysis of missing data

- Place a prior distribution on the missing covariates and proceed as usual
  - This makes Bayesian approaches to missing data relatively easy, comparatively
  - $x_{i1} \sim N(\lambda_0 + \lambda_1 y_i + \lambda_2 x_{i2}, \tau^2)$ if x_{i1} continuous
  - $x_{i1} \sim Bernoulli(p_i)$ with p_i a logistic model if x_{i1} discrete
- All we're doing is adding a few lines to our Gibbs sampling algorithm, drawing values for the missing covariates
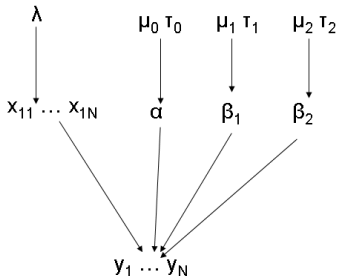
Simulated Data

- We'll simulate some data so we know what the answer should be when we analyze it
- 200 observations from a linear regression
- $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$
- $\epsilon_i \sim N(0, \sigma^2)$
- $\beta_0 = -1$, $\beta_1 = 1$, $\beta_2 = -1$, $\beta_3 = 1$, $\sigma^2 = 1$

missingdata.R contains the code for this example

Bayesian Analysis of Full Data

Model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma_y^2)$$
$$\beta_j \sim N(0, 100)$$
$$\sigma_y^2 \sim Uniform(0.00001, 100)$$

- Non-informative priors on all parameters
- Don't do this if you have prior information!
- Caution is required for the prior on σ^2

Convergence
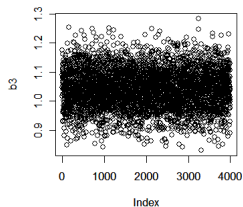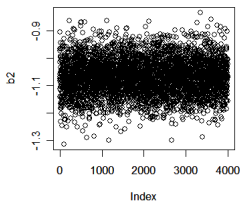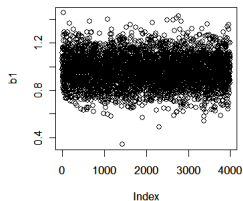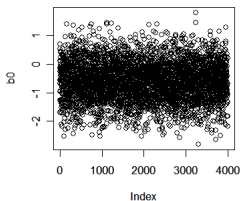
Full Data results

Coefficient	Coeff	95% CI
β_0	-0.6	(-1.9, 0.7)
β_1	1.0	(0.7, 1.2)
β_2	-1.1	(-1.2, -0.9)
β_3	1.0	(0.9, 1.2)

- $\beta_1, \beta_2, \beta_3$ almost identical to truth
- β_0 is off a bit, but the credible intervals are quite wide, and include the truth

Missing Data type 1: MCAR

MCAR Data

$$Pr(r_{i1} = 1) = 0.2$$
$$Pr(r_{i2} = 1) = 0.25$$
$$Pr(r_{i3} = 1) = 0.3$$

- The r_i's are probabilities of each variable being missing
- The probabilities are constant: MCAR
- 56.4% of the data are missing

Bayesian MCAR Analysis

Standard Bayes Model for Linear Regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

$$\beta_j \sim N(0, 100) \quad \sigma_y^2 \sim Uniform(0.00001, 100)$$

Bayesian MCAR Analysis

Standard Bayes Model for Linear Regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma_y^2)$$
$$\beta_j \sim N(0, 100) \quad \sigma_y^2 \sim Uniform(0.00001, 100)$$

...and MCAR Missing Data Model

$$x_{i1} \sim Bernoulli(p_{x_1})$$
$$\text{logit}\{p_{x1}\} = \mu_{x_1}$$
$$x_{i2} \sim N(\mu_{x_2}, \sigma_2^2) \quad x_{i3} \sim N(\mu_{x_3}, \sigma_3^2)$$
$$\sigma_2^2 \sim Uniform(0.00001, 100) \quad \sigma_3^2 \sim Uniform(0.00001, 100)$$
$$\mu_j \sim N(0, 100)$$

Some intuition on the model

- We want to estimate β in the same way as before
  - Combining prior with data
- Also want to estimate μ, the mean of the missing data
  - Under MCAR, the data provide lots of information about μ
  - For instance, the means of observed variables
  - The μ's are used to impute missing values of *x*'s

Convergence
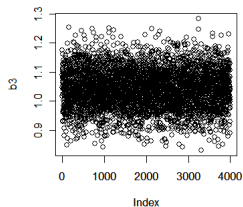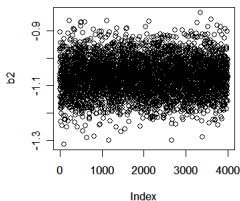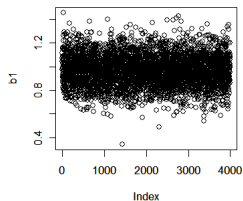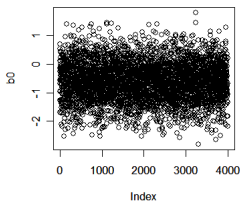
MCAR Data results

	Complete Case		Account for MCAR	
Coefficient	Coeff	95% CI	OR	95% CI
β_0	-1.6	(-3.4, 0.3)	-1.1	(-2.4, -0.6)
β_1	1.1	(0.7, 1.5)	1.0	(0.7, 1.3)
β_2	-1.2	(-1.3, -1.0)	-1.1	(-1.2, -0.9)
β_3	1.0	(0.8, 1.1)	1.0	(0.9, 1.2)

- CC and adjusted are quite similar - Not surprisingly
- Accounting for the MCAR mechanism improves precision - slightly
- This is why people usually don't bother with missing data techniques when MCAR

Missing Data type 2: MAR

MAR Data

$$\text{Logit}\{Pr(r_{i1} = 1)\} = 15 + 2y_i - x_{2i}$$
$$\text{Logit}\{Pr(r_{i2} = 1)\} = 3 + 0.5y_i$$
$$\text{Logit}\{Pr(r_{i3} = 1)\} = -13 - y_i - x_{1i}$$

- The missingness mechanism is different for each variable
- Depends on the rest of the observed data, so MAR
- 17% of x_1, 3% of x_2, and 37.6% of x_3 missing

Bayesian MAR Analysis

Standard Bayes Model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

$$\beta_j \sim N(0, 100) \quad \sigma_y^2 \sim Uniform(0.00001, 100)$$

Bayesian MAR Analysis

Standard Bayes Model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma_y^2)$$
$$\beta_j \sim N(0, 100) \quad \sigma_y^2 \sim Uniform(0.00001, 100)$$

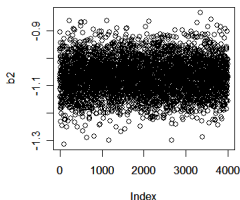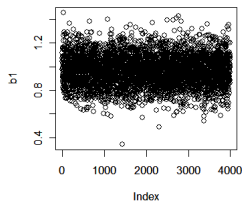*...and MAR Missing Data Model*
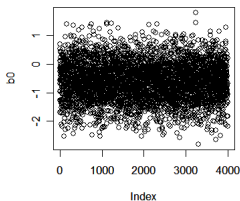
$$x_{i1} \sim Bernoulli(p_{x_1})$$
$$\text{logit}\{p_{x1}\} = \mu_{0,1} + \mu_{1,1} y_i + \mu_{2,1} x_{2i}$$
$$x_{i2} \sim N(\mu_{0,2} + \mu_{1,2} y_i, \sigma_2^2) \quad x_{i3} \sim N(\mu_{0,3} + \mu_{1,3} y_i + \mu_{2,3} x_{i2}, \sigma_3^2)$$
$$\sigma_2 \sim Uniform(0.00001, 100) \quad \sigma_3 \sim Uniform(0.00001, 100)$$
$$\mu_{j,k} \sim N(0, 100)$$

Convergence
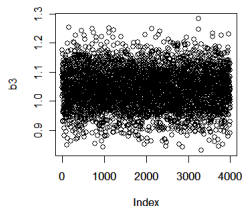
MAR Data Results

	Complete Case		Account for MAR	
Coefficient	Coeff	95% CI	OR	95% CI
β_0	-4.4	(-6.2, 2.5)	-0.2	(-1.8, 1.3)
β_1	0.4	(0.1, 0.7)	1.0	(0.6, 1.3)
β_2	-0.6	(-0.8, -0.4)	-1.1	(-1.3, -1.0)
β_3	0.6	(0.4, 0.8)	1.1	(0.9, 1.2)

- Complete Case analysis would yield incorrect answer
- Complete Case intervals don't include true effect, even!
- Bayes model accounting for MAR recovers true effects, almost exactly

Summary

- Bayesian methods for missing data can be very useful
- Generally easier to implement than other (proper) methods
- The results depend on your model for *x*
- Can't test between MCAR, MAR, NMAR
- Extensions to non-ignorable data are straightforward (see Kmetic et al, Hernan et al)

References - Hierarchical Regression

- Congdon P. *Bayesian Statistical Modelling*. 2001
- Gelman A and Hill J. Data Analysis Using Regression and Multilevel/ Hierarchical Models. Cambridge University Press. 2006
- Greenland S. Methods for epidemiologic analyses of multiple exposures. Stat Med. 12: 717-736. 1992
- Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. Environ Health Perspect. 102 (Suppl 8): 33-39. 1994
- MacLehose RF, Dunson DB, Herring AH and Hoppin JA. Bayesian methods for highly correlated data. *Epidemiology* 18(2): 199-207, 2007.

References - Missing Data

- Ibrahim JG, Chen MH, Lipsitz SR, Herring AH. Missing-data methods for generalized linear models: A comparative review. JASA 2005.
- Kmetic A, Jospeh L, Berger C, Tenenhouse A. Multiple imputation to account for missing data in a survey. Epidemiology 13:437-444. 2002
- Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004