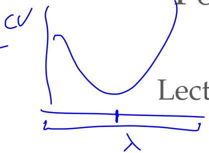
Linear Regression Models
P8111

• HW 4 notes
- w shaped CV
- outlier Removal
-



Lecture 23

Jeff Goldsmith

April 19, 2016

THE DEPARTMENT OF BIOSTATISTICS
Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Today's Lecture

- Multilevel models
  - Hierarchical / nested models
  - Crossed designs
- Bayesian methods

Longitudinal data

- We observe data $y_{ij}, \boldsymbol{x}_{ij}$ for subjects $i = 1, \ldots I$ at visits $j = 1, \ldots, J_i$
- Overall, we pose the model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2 V$ and

$V(\rho)$

$$V = \begin{bmatrix} \boxed{V_1} & 0 & \ldots & 0 \\ 0 & \boxed{V_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \boxed{V_I} \end{bmatrix}$$

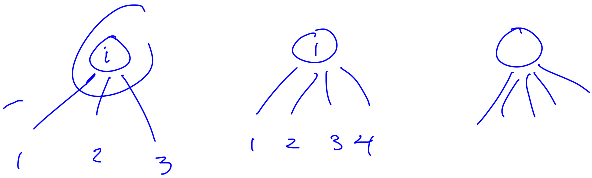
Longitudinal data

- Extended cross-sectional models to allow repeated subject observations
- Repeated observations had a time element
- One basic approach was random effects

Multilevel models

- Multilevel models are a "more general" class of models
- Repeated observations don't necessarily have to be taken in time
- Examples of two-level models include students in a class, members in a family, patients in a hospital, etc

y_{ij}



Two-level model

The repeated observations structure we developed for longitudinal data helps for two-level models.
Specifically for repeated observations j within clusters i, we could write

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_i + \epsilon_{ij}$$

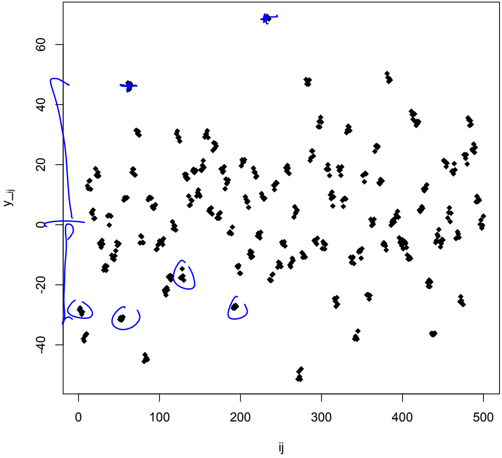
with

- $b_i \sim \mathrm{N}\left[0, \tau^2\right]$
- $\epsilon \sim \mathrm{N}\left[0, \nu^2\right]$

Intuition, estimation, induced correlation, interpretation – all of these were established for LDA and transfer here

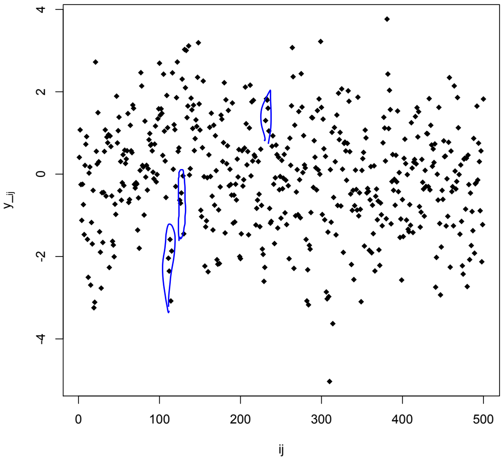
Example I

$\tau^2 >> \nu^2$ / $ICC \approx 1$



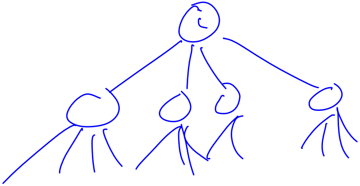
Example II

$\tau^2 \ll \nu^2$ / $ICC \simeq 0$



Three level model

- Sometimes, the data have a more complex nested structure
- Each cluster is part of a larger cluster
- Examples include students in classes in universities, members in families in towns, patients in hospitals in regions



Three level model

For a model with three levels (repeated observations k within clusters j, within super-clusters i), we can write

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk} + b_i + b_{ij} + \epsilon_{ijk}$$

with

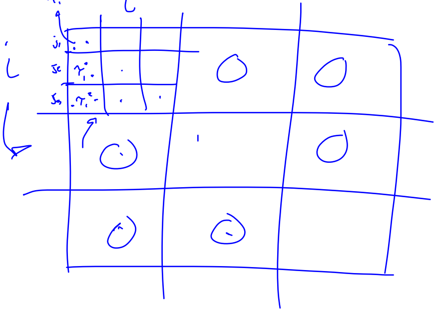
- $b_i \sim \mathrm{N}\left[0, \tau^2_{(1)}\right]$
- $b_{ij} \sim \mathrm{N}\left[0, \tau^2_{(2)}\right]$
- $\epsilon \sim \mathrm{N}\left[0, \nu^2\right]$

ICCs

This model gives two levels of correlation (observations within clusters, clusters within super clusters), and therefore a couple of ICCs:

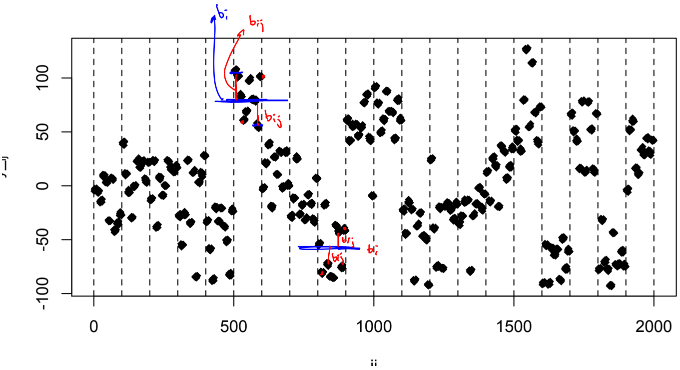
- $cov(y_{ijk}, y_{ijk'}) = Cov(\beta_0 + \beta_1 x_{ijk} + b_i + b_{ij} + \epsilon_{ijk}, \; \beta_0 + \beta_1 x_{ijk'} + b_i + b_{ij} + \epsilon_{ijk'})$

  $= \tau_1^2 + \tau_2^2$



- $cov(y_{ijk}, y_{ij'k}) = \tau_1^2$

Example



Example

$y_{ijk} = \beta_0 + b_{i0} + b_{i1} x^1_{ijk}$

$+ b_{ij0} + b_{ij1} x^2_{ijk}$

$(1 + x_1 | L_1) + (1 + x_2 | L_2)$

vs

$(1 | L_1) + (1 + x_1 + x_2 | L_2)$

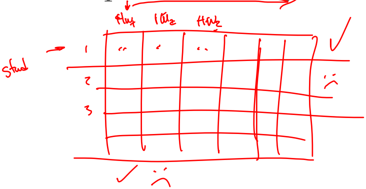
Nested model

$y_{ijk} = \beta_0 + b_i + b_{ij} + \epsilon_{ijk}$

```
> nested.mod = lmer(yij ~ (1 | L1) + (1 | L2))
> summary(nested.mod)
Linear mixed model fit by REML ['lmerMod']
Formula: yij ~ (1 | L1) + (1 | L2)

REML criterion at convergence: 7464.337

Random effects:
 Groups   Name        Variance Std.Dev.
 L2       (Intercept)  527.003 22.957
 L1       (Intercept) 2137.453 46.233
 Residual                1.004  1.002
Number of obs: 2000, groups: L2, 200; L1, 20

Fixed effects:
            Estimate Std. Error t value
(Intercept)   0.7712    10.4646   0.074
```

Crossed designs

- Alternatively to nested (hierarchical) models, sometimes there is a crossed design
- Each subject is observed under multiple "treatments", so there are both subject and treatment effects
- For example, each student is graded in multiple classes; each patient is assayed for multiple genes



y_{ij}

Crossed designs

For a crossed model (with subjects i and treatments j), we can write

$$\underline{y_{ij}} = \beta_0 + \beta_1 x_{ij} + \underline{b_i} + \underline{b_j} + \epsilon_{ij}$$

with

- $b_i \sim \mathrm{N}\left[0, \tau^2_{(1)}\right]$
- $b_j \sim \mathrm{N}\left[0, \tau^2_{(2)}\right]$
- $\epsilon \sim \mathrm{N}\left[0, \nu^2\right]$

Here there is covariance within subjects across treatments, and within treatments across subjects.

ICCs

Here there is covariance within subjects across treatments, and within treatments across subjects.

- $cov(y_{ij}, y_{ij'}) =$

- $cov(y_{i'j}, y_{ij}) =$

LDA and MLM

- Estimation works basically the same for these models as for random intercept models
- Intuition is the same as well – you want to borrow strength for one subject from the population of other subjects
- Interpretation of fixed effects is *marginal*; interpretation of random effects is *conditional* $\hookrightarrow E(y) = E(E(y|b))$ $= E(y|b=0)$
- Using randomness both decreases the number of parameters and induces correlation structures

Bayesian methods

Longitudinal data analysis and multilevel models are a good place to start "thinking Bayesian"

- Even though they're frequentist, they include randomness at subject levels
- The idea of "shrinking toward a population mean" or "borrowing strength" is a pretty Bayesian concept
- Even writing down random effect distributions is reminiscent of defining prior distrubtions

Basic Bayes

LDA and MLM are fairly advanced topics, so we'll start with a simpler example

- Suppose I gather data y_i and want to learn about $E(y)$
- Suppose even more I think I already know *something* about $E(y)$
- I might write down something about what I want to learn and what I think I know

Basic Bayes

E(y)

What do I think I know?

- $\boxed{y_i|\mu \sim \mathrm{N}\left[\mu, \sigma_y^2\right]}$
- $\mu \sim \mathrm{N}\left[\mu_0, \sigma_0^2\right]$

What do I want to learn?

- $\mu|y_i \sim ???$

Basic Bayes

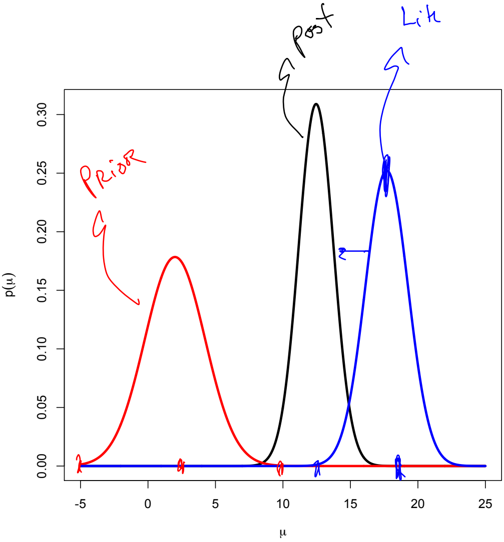
Luckily, this is all related through Bayes' formula:

$$\underbrace{p(\mu|y_i)}_{\text{posterior}} \propto \underbrace{p(y_i|\mu)}_{\text{Likelihood}}\underbrace{p(\mu)}_{\text{PRIOR}}$$

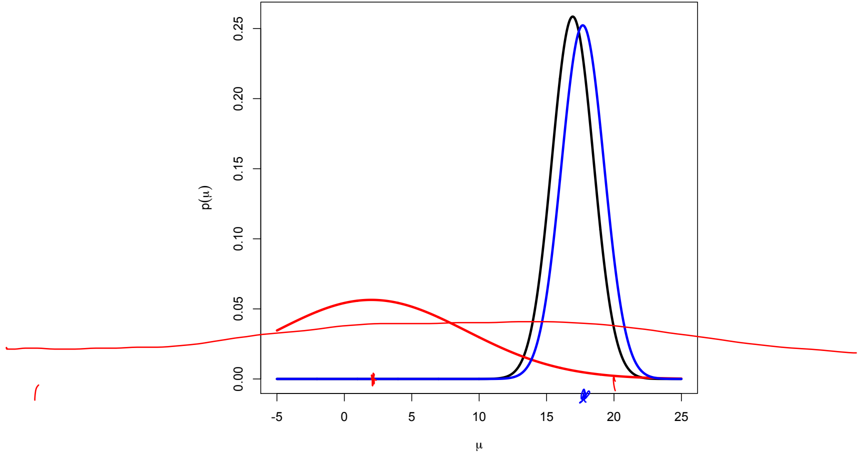
- For the Normal likelihood with a Normal prior for the mean, the posterior is also Normal:

$$\mu|y_i \sim \mathrm{N}\left[\frac{\sigma_\mu^2}{\frac{\sigma_y^2}{n}+\sigma_\mu^2}\boxed{\bar{y}}+\frac{\frac{\sigma_y^2}{n}}{\frac{\sigma_y^2}{n}+\sigma_\mu^2}\mu_0,\ \frac{\frac{\sigma_y^2}{n}\sigma_\mu^2}{\frac{\sigma_y^2}{n}+\sigma_\mu^2}\right]$$

Effect of Prior



Effect of Prior



Bayesian regression

$y | \mu$; μ

How can we pose the regression model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$$

$y | \beta \sim N(X\beta, \sigma^2 I)$

$(X^T X)^{-1} X^T y$

with $\boldsymbol{\epsilon} \sim \mathrm{N}\left[0, I_n\right]$ in a Bayesian framework?

- By making distributional assumptions about the $\boldsymbol{\beta}$
- Normal priors seemed to work well in the past ...
- Try $\boldsymbol{\beta} \sim \mathrm{N}\left[0, \sigma^2_{\boldsymbol{\beta}} I_p\right]$ where p includes the intercept

Bayesian regression

We want to obtain the posterior

$p(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}) \propto p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{X})p(\boldsymbol{\beta})$

$$\exp\left\{ -\frac{1}{2} \cdots \right\}$$

$$\exp\left\{ -\frac{1}{2} \cdots \right\}$$

Bayesian regression

Can show that

$\beta \mid y, x$

$$[\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{X}] \sim \mathrm{N}\left[\mu_p, \Sigma_p\right]$$

where

$$\Sigma_p = \left(\frac{1}{\sigma_\epsilon^2}\boldsymbol{X}^T\boldsymbol{X} + \frac{1}{\sigma_{\boldsymbol{\beta}}^2}\boldsymbol{I}\right)^{-1}$$

and

$$\mu_p = \Sigma_p\left(\frac{1}{\sigma_\epsilon^2}\boldsymbol{X}^t\boldsymbol{y}\right)$$

$$= \left(X^T X + \frac{\sigma_\epsilon^2}{\sigma_\beta^2} I\right)^{-1} X^T y$$

So, about the variances

$y \mid \beta, \sigma_\epsilon^2$

$\beta \mid y$

$\beta \mid \sigma_\beta^2$

- Throughout all of this we have implicitly conditioned on the variances σ_ϵ^2 and $\sigma_{\boldsymbol{\beta}}^2$
- Doesn't affect any of our calculations – the terms involving μ don't overlap with terms involving σ_ϵ^2 or $\sigma_{\boldsymbol{\beta}}^2$
- $\sigma_{\boldsymbol{\beta}}^2$ is often treated as fixed; σ_ϵ^2

$\beta, \sigma_\epsilon^2 \mid y$

The full posterior

- Need $[\boldsymbol{\beta}, \sigma_\epsilon^2 | \boldsymbol{y}, \boldsymbol{X}]$
- "Intractible" problem
- Just as good: sample from the posterior

Sampling from the posterior

MCMC

- aka where Bayes gets really weird
- You can draw a sample from the posterior even if you can't write down exactly what it is
- That sample is your basis for inference
  - Posterior sample average is your estimate
  - Quantiles on the posterior sample define your credible interval
- Sample describes the *joint distribution* of all model parameters

Some notes on this business

$$\hat{Var}(\hat{\beta}_{OLS}) = \hat{\sigma}^2 (X^T X)^{-1}$$

- Joint distributions are often worth the trouble
- Bayesian methods were really controversial for a long time, but are at least less controversial now
- The introduction of "prior knowledge" happens even in frequentist methods, although it is often not explicitly acknowledged

Today's big ideas

- Nested and crossed random effects models
- Bayesian stuff