Intellectual impairment and blood lead

Example I (courtesy of Rich Canfield, Nutrition, Cornell)

- blood lead and intelligence measured on children
- Question: how do low doses of lead affect IQ?
  - important since doses are decreasing with lead now out of gasoline
- several IQ measurements per child
  - so longitudinal
- nine “confounders”
  - e. g., maternal IQ
  - need to adjust for them
- effect of lead appears nonlinear
  - important conclusion
Dose-response curve

Thanks to Rich Canfield for data and estimates

Spinal bone mineral density example

Example II (in Ruppert, Wand, Carroll (2003), *Semiparametric Regression*)

- age and spinal bone mineral density measured on girls and young women
- several measurements on each subject
- increasing but nonlinear curves
Spinal bone mineral density data

What is needed to accommodate these examples

We need a model with

- potentially many variables
- possibility of nonlinear effects
- random subject-specific effects

The model should be one that can be fit with readily available software such as SAS, Splus, or R.
Underlying philosophy

- minimalist statistics
  - keep it as simple as possible
- build on classical parametric statistics
- modular methodology
  - so we can add components to accommodate special features in data sets

Outline of the approach

- Start with linear mixed model
  - allows random subject-specific effects
  - fine for variables that enter linearly
- Expand the basis for those variables that have nonlinear effects
  - we will use a spline basis
  - treat the spline coefficients as random effects to induce empirical Bayes shrinkage = smoothing
- End result
  - linear mixed model from a software perspective, but nonlinear from a modeling perspective

(Much like polynomial regression, but without the drawbacks of polynomials.)
Multiple linear regression

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \epsilon_i \]

**Examples of predictor variables:**
- \( X_{i1} = \) blood lead concentration of \( i\)th child
- \( X_{i2} = X_{i1}^2 \)
- \( X_{i3} = 1 \) if \( i\)th child lives with both parents (is 0 otherwise)

In the standard linear model:
- \( \epsilon_1, \ldots, \epsilon_n \) are independent with a constant variance

Polynomial regression

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \cdots + \beta_p X_{i1}^p + \text{other variables} + \epsilon_i \]

- This is an example of **basis expansion**
- But polynomials are not nearly as good as splines at approximating other nonlinear functions
Example: pig weights (random effects)

**Example III** (from Ruppert, Wand, and Carroll (2003))

\[ Y_{ij} = (\beta_0 + b_{0i}) + \beta_1 \text{week}_j \]

- \( Y_{ij} \) = weight of \( i \)th pig at the \( j \)th week
- \( \beta_0 \) is the average intercept for pigs
- \( b_{0i} \) is an offset for \( i \)th pig
- So \( (\beta_0 + b_{0i}) \) is the intercept for the \( i \)th pig
Are random intercepts enough?

Example III

\[ Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) \text{week}_j \]

- \( \beta_1 \) is the average slope
- \( b_{ii} \) is an adjustment to slope of the \( i \)th pig
- So \( (\beta_1 + b_{1i}) \) is the slope for the \( i \)th pig
- \( b_{0i} \) and \( b_{1i} \) seem positively correlated
  - makes sense: faster growing pigs should be larger at the start of data collection
General form of linear mixed model

- \( \mathbf{X}_i = (X_{i1}, \ldots, X_{ip}) \) and \( \mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iq}) \) are vectors of predictor variables
- \( \boldsymbol{\beta} = (\beta_1, \ldots, \beta_p) \) is a vector of fixed effects
- \( \mathbf{b} = (b_1, \ldots, b_q) \) is a vector of random effects
  - \( \mathbf{b} \sim \text{MVN}(0, \Sigma(\theta)) \)
  - \( \theta \) is a vector of variance components
- Model is:
  \[
  Y_i = \mathbf{X}^T_i \boldsymbol{\beta} + \mathbf{Z}^T_i \mathbf{b} + \epsilon_i
  \]
- Note use of inner product notation:
  \[
  \mathbf{X}^T_i \boldsymbol{\beta} = \sum_{j=1}^{p} X_{ij} \beta_j \text{ and } \mathbf{Z}^T_i \mathbf{b} = \sum_{j=1}^{q} Z_{ij} b_j
  \]

Estimation in linear mixed models

- \( \boldsymbol{\beta} \) and \( \theta \) are the parameter vectors
  - estimated by
    - ML (maximum likelihood), or
    - REML (maximum likelihood with degrees of freedom correction)
- \( \mathbf{b} \) is a vector of random variables
  - predicted by a BLUP (Best linear unbiased predictor)
  - BLUP is shrunk towards zero (mean of \( \mathbf{b} \))
  - amount of shrinkage depends on \( \hat{\theta} \)
Estimation in linear mixed models, cont.

- **Random intercepts example:**
  \[ Y_{ij} = (\beta_0 + b_{0i}) + \beta_1 \text{week}_j \]

  - **high variability** among the intercepts \( b_{0i} \) towards 0
    - extreme case: intercepts are fixed effects
  - **low variability** among the intercepts \( b_{0i} \) \( \Rightarrow \) more shrinkage
    - extreme case: common intercept (another fixed effects model)

Comparison between fixed and random effects modeling

- fixed effects models allow only the two extremes:
  - no shrinkage
  - maximal shrinkage to a common intercept
- mixed effects modeling allows all possibilities between these extremes
Splines

- polynomials are **excellent** for **local** approximation of functions
- in practice, polynomials are relatively **poor** at **global** approximation
- a spline is made by joining polynomials together
  - takes advantage of polynomials strengths without inheriting their weaknesses
- splines have "maximal smoothness"

Splines have "maximal smoothness"

**Is this a linear spline?**

![Graph 1](image1.png)

**NO: has a jump**

![Graph 2](image2.png)

**YES: kink is okay**
Splines have "maximal smoothness," cont.

**Is this a quadratic spline?**

<table>
<thead>
<tr>
<th>NO: has a jump</th>
<th>NO: has a kink</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

**YES: change in concavity is okay**

<table>
<thead>
<tr>
<th>NO: has a jump</th>
<th>NO: has a kink</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

---

**Piecewise linear spline model**

"Positive part" notation:

\[
x_+ = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases}
\]

\( (1) \)

Linear spline:

\[
m(x) = \{\beta_0 + \beta_1 x\} + \{b_1(x - \kappa_1)_+ + \cdots + b_K(x - \kappa_K)_+\}
\]

- \( \kappa_1, \ldots, \kappa_K \) are "knots"
- \( b_1, \ldots, b_K \) are the spline coefficients
Linear “plus” function with $\kappa = 1$

$$m(x) = \beta_0 + \beta_1 x + b_1(x - \kappa_1)_+ + \cdots + b_K(x - \kappa_K)_+$$

- slope jumps by $b_k$ at $\kappa_k$, $k = 1, \ldots, K$
Fitting LIDAR data with plus functions

Generalization: higher degree splines

\[ m(x) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + b_1 (x - \kappa_1)^p_+ + \cdots + b_K (x - \kappa_K)^p_+ \]

- \( p \)th derivative jumps by \( p! b_k \) at \( \kappa_k \)
- first \( p - 1 \) derivatives are continuous
Quadratic “plus” function

LIDAR data: ordinary Least Squares
LIDAR data: penalized least-squares

- Use matrix notation:

\[ m(X_i) = \beta_0 + \beta_1 X_i + \cdots + \beta_p X_i^p \]

\[ + b_1 (X_i - \kappa_1)^p + \cdots + b_K (X_i - \kappa_K)^p \]

\[ = X_i^T \beta X + B^T(X_i)b \]

- Minimize

\[ \sum_{i=1}^{n} \left\{ Y_i - (X_i^T \beta X + B^T(X_i)b) \right\}^2 + \lambda b^TDb. \]

Penalized least-squares, cont.

- From previous slide: minimize

\[ \sum_{i=1}^{n} \left\{ Y_i - (X_i^T \beta X + B^T(X_i)b) \right\}^2 + \lambda b^TDb. \]

- \( \lambda b^TDb \) is a penalty that prevents overfitting
- \( D \) is a positive semidefinite matrix
  - so the penalty is non-negative
  - Example:
    \[ D = I \]
- \( \lambda \) controls the amount of penalization
- the choice of \( \lambda \) is crucial
Penalized Least Squares

Choice of $\lambda$ is crucial
How can \( \lambda \) be chosen?

The smoothing parameter \( \lambda \) can be chosen automatically using mixed model software.

Ridge Regression

From earlier slide:

\[
\sum_{i=1}^{n} \left\{ Y_{i} - (X_{i}^{T} \beta_{X} + B^{T}(X_{i})b) \right\}^2 + \lambda b^{T}Db.
\]

Let \( \mathcal{X} \) have row \( \begin{pmatrix} X_{i}^{T} & B^{T}(X_{i}) \end{pmatrix} \). Then

\[
\begin{pmatrix} \hat{\beta}_{X} \cr \hat{b} \end{pmatrix} = \left\{ \mathcal{X}^{T} \mathcal{X} + \lambda \text{ blockdiag}(0, D) \right\}^{-1} \mathcal{X}^{T}Y.
\]

- This is a ridge regression estimator
- Also, as we will see, it is a BLUP in a mixed model and an empirical Bayes estimator
Linear Mixed Models

- Assume the linear mixed model:
  \[ Y = X\beta + Zb + \varepsilon \]

  - where
    - \( b \) is \( N(0, \sigma_b^2 \Sigma_b) \)
    - \( \varepsilon \) is \( N(0, \sigma^2 I) \)
    - \( X\beta \) are the “fixed effects”
    - \( Zb \) are the “random effects”

  - **Henderson’s equations.**

  \[
  \begin{pmatrix}
  \hat{\beta} \\
  \hat{b}
  \end{pmatrix} = \left( \begin{pmatrix}
  X^T X & X^T Z \\
  Z^T X & Z^T Z + \lambda \Sigma_b^{-1}
  \end{pmatrix} \right)^{-1} \begin{pmatrix}
  X^T Y \\
  Z^T Y
  \end{pmatrix}.
  \]

  \[ \lambda = \frac{\sigma^2}{\sigma_b^2}. \]

---

From previous slides:
Ridge regression: Let \( \mathcal{X} \) have row \( (X_i^T, B^T(X_i)) \). Then

\[
\begin{pmatrix}
\hat{\beta}_X \\
\hat{b}
\end{pmatrix} = \left\{ \mathcal{X}^T \mathcal{X} + \lambda \text{ blockdiag}(0, 0, D) \right\}^{-1} \mathcal{X}^T Y.
\]

Linear mixed model:

\[
\begin{pmatrix}
\hat{\beta} \\
\hat{b}
\end{pmatrix} = \left( \begin{array}{cc}
X^T X & X^T Z \\
Z^T X & Z^T Z + \lambda \Sigma_b^{-1}
\end{array} \right)^{-1} \begin{pmatrix}
X^T Y \\
Z^T Y
\end{pmatrix}
\]

\[ = \left\{ (X Z)^T (X Z) + \lambda \text{ blockdiag}(0, \Sigma_b^{-1}) \right\}^{-1} (X Z)^T Y \]
Selecting $\lambda$

To choose $\lambda$ use:
- one of several model selection criteria:
  - cross-validation (CV)
  - generalized cross-validation (GCV)
  - AIC
  - $C_P$
- ML or REML in mixed model framework

Modeling the blood lead and IQ data

For the $j$th measurements on the $i$th subject:

$$IQ_{ij} = b_i + m(lead_{ij}) + \beta_1 X_{ij}^1 + \cdots + \beta_L X_{ij}^L + \epsilon_{ij}$$

- $m(\cdot)$ is a spline
  - include the population average intercept
- $b_i$ is a random subject-specific intercept
  - $E(b_i) = 0$
  - model assumes parallel curves
- $X_{ij}^\ell$ is the value of the $\ell$th confounder, $\ell = 1, \ldots, L$
Return to spinal bone mineral density study

\[ SBMD_{i,j} = U_i + m(\text{age}_{i,j}) + \epsilon_{i,j}, \]
\[ i = 1, \ldots, m = 230, \quad j = i, \ldots, n_i. \]

Fixed effects

\[
X = \begin{bmatrix}
1 & \text{age}_{11} \\
\vdots & \vdots \\
1 & \text{age}_{1n_1} \\
\vdots & \vdots \\
1 & \text{age}_{m1} \\
\vdots & \vdots \\
1 & \text{age}_{mn_m}
\end{bmatrix}
\]
Random effects

$$Z = \begin{bmatrix}
1 & \cdots & 0 & (\text{age}_{11} - \kappa_1)^+ & \cdots & (\text{age}_{11} - \kappa_K)^+\\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & \cdots & 0 & (\text{age}_{1n_1} - \kappa_1)^+ & \cdots & (\text{age}_{1n_1} - \kappa_K)^+\\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 1 & (\text{age}_{m1} - \kappa_1)^+ & \cdots & (\text{age}_{m1} - \kappa_K)^+\\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 1 & (\text{age}_{mn_m} - \kappa_1)^+ & \cdots & (\text{age}_{mn_m} - \kappa_K)^+
\end{bmatrix}$$

$$u = \begin{bmatrix}
U_1 \\
\vdots \\
U_m \\
b_1 \\
\vdots \\
b_K
\end{bmatrix}$$
Variability bars on $\hat{m}$ and estimated density of $U_i$
Model with ethnicity effects

$$SBMD_{ij} = U_i + m(age_{ij}) + \beta_1 black_i + \beta_2 hispanic_i + \beta_3 white_i + \epsilon_{ij}, \quad 1 \leq j \leq n_i, \quad 1 \leq i \leq m.$$  

Asian is the reference group.

Only requires an expansion of the fixed effects by adding the columns

$$\begin{bmatrix}
black_1 & hispanic_1 & white_1 \\
\vdots & \vdots & \vdots \\
black_1 & hispanic_1 & white_1 \\
\vdots & \vdots & \vdots \\
black_m & hispanic_m & white_m \\
\vdots & \vdots & \vdots \\
black_m & hispanic_m & white_m
\end{bmatrix}$$
Ethnicity effects

Possible enrichment of the model

- In this model, the age effects curve for the four ethnic groups are **parallel**.
- Could we model them as non-parallel?
- Might be problematic in this example because of the small values of the $n_i$.
- But the methodology should be useful in other contexts.
Penalized Splines and Additive Models

Bivariate Additive model:

\[ Y_i = m_1(X_i) + m_2(Z_i) + \epsilon_i \]

- Generalizes easily to more than two predictors
- No interactions: so easy to interpret

Bivariate additive spline model

\[
\begin{align*}
Y_i &= \beta_0 \\
&+ \beta_{x,1}X_i + b_{x,1}(X_i - \kappa_{x,1})_+ + \cdots + b_{x,K}(X_i - \kappa_{x,K})_+ \\
&+ \beta_{z,1}Z_i + b_{z,1}(Z_i - \kappa_{z,1})_+ + \cdots + b_{z,K}(Z_i - \kappa_{z,K})_+ \\
&+ \epsilon_i
\end{align*}
\]
Penalized Splines and Additive Models

- no need for backfitting
- computation very rapid
- no identifiability issues
- inference is simple

Milan study of mortality and air pollution

Data:
- daily mortality
- daily weather variables
- TSP = total suspended particulate matter

Additive Model:

$$\sqrt{\text{mortality}_t} = \beta_0 + \beta \text{TSP}_t + f_1(t) + f_2(\text{temperature}_t) + f_3(\text{humidity}_t) + \varepsilon_t$$
Milan study: results

Other models that fit in this framework

- generalized regression
  - response is not Gaussian
  - e.g., logistic regression for a binary response
- variance functions
  - for nonconstant response variance
- measurement error
  - when $X$ is measured with error
- bivariate smoothing
  - e.g., for spatial data
- spatially adaptive smoothing
  - where there are regions of high and of low curvature
Summary

- Mixed models allow subject-specific effects to be similar but not the same
- Splines are excellent at approximating nonlinear functions
- Splines can be embedded in mixed models by treating the spline coefficients as random effects
- The amount of smoothing can be determined automatically by REML
- Modular statistical methodology is essential in practice