Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Operations Research & Information Engineering, Cornell University

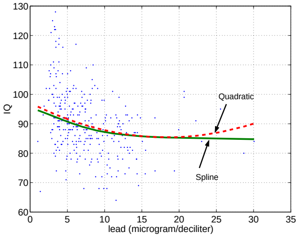
Oct 13, 2008

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

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples

Introduction

Linear and mixed linear models

Univariate spline models

Back to examples

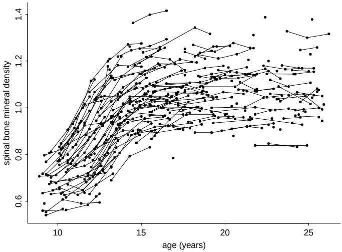
More general spline models

Summary

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

Linear Statistical Models to Mixed Models to Semiparametric Regression
David Ruppert
Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

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

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

Outline of the approach

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

- 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 ith child
- $X_{i2} = X_{i1}^2$
- $X_{i3} = 1$ if ith 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

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples

Introduction

Linear and mixed linear models

Univariate spline models

Back to examples

More general spline models

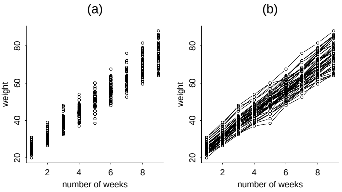
Summary

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

- 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))



Random intercept model

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

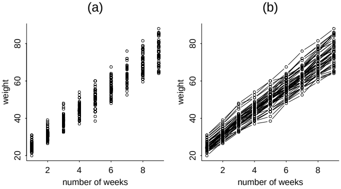
Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

$$Y_{ij} = (\beta_0 + b_{0i}) + \beta_1 \texttt{week}_j$$

- Y_{ij} = weight of ith pig at the jth week
- β_0 is the average intercept for pigs
- b_{0i} is an offset for ith pig
- So $(\beta_0 + b_{0i})$ is the intercept for the ith pig

Are random intercepts enough?

Example III



Random lines model

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})\,\texttt{week}_j$$

- β_1 is the average slope
- b_{ii} is an adjustment to slope of the ith pig
- So $(\beta_1 + b_{1i})$ is the slope for the ith 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
  - $\mathbf{b} \sim MVN\{0, \Sigma(\theta)\}$
  - θ is a vector of variance components
- Model is:

$$Y_i = \mathbf{X}_i^{\mathsf{T}} \boldsymbol{\beta} + \mathbf{Z}_i^{\mathsf{T}} \mathbf{b} + \epsilon_i$$

- Note use of inner product notation:

$$\mathbf{X}_i^{\mathsf{T}} \boldsymbol{\beta} = \sum_{j=1}^{p} X_{ij}\beta_j \text{ and } \mathbf{Z}_i^{\mathsf{T}} \mathbf{b} = \sum_{j=1}^{q} Z_{ij} b_j$$

Estimation in linear mixed models

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples

Introduction

Linear and mixed linear models

Univariate spline models

Back to examples

More general spline models

Summary

- $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ 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 $\widehat{\boldsymbol{\theta}}$

Estimation in linear mixed models, cont.

- Random intercepts example:

$$Y_{ij} = (\beta_0 + b_{0i}) + \beta_1 \texttt{week}_j$$

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

Comparison between fixed and random effects modeling

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples

Introduction

Linear and mixed linear models

Univariate spline models

Back to examples

More general spline models

Summary

- 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

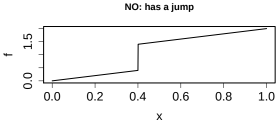
Splines

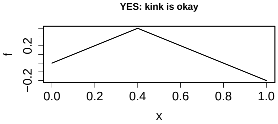
- polynomials are excellent for local approximation of functions
- in practice, polynomials are relatively poor at global approximation
- a spline is made by joining polynomials together
  - takes advantage of polynomials strengths without inheriting their weaknesses
- splines have "maximal smoothness"

Splines have "maximal smoothness"

Linear Statistical Models to Mixed Models to Semiparametric Regression
David Ruppert
Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

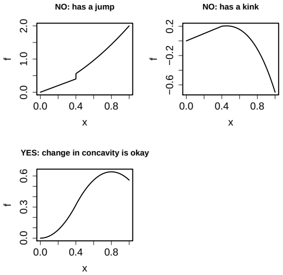
Is this a linear spline?





Splines have "maximal smoothness," cont.

Is this a quadratic spline?



Linear Statistical Models to Mixed Models to Semiparametric Regression
David Ruppert
Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

Piecewise linear spline model

“Positive part” notation:

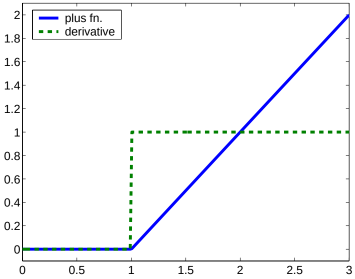
$$x_+ = x, \text{ if } x > 0 \quad (1)$$
$$= 0, \text{ if } x \leq 0 \quad (2)$$

Linear spline:

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

- $\kappa_1, \ldots, \kappa_K$ are “knots”
- $b_1, \ldots, b_K$ are the spline coefficients

Linear "plus" function with $\kappa = 1$



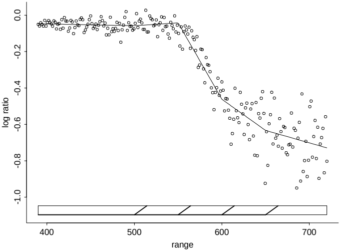
Linear spline

Linear Statistical Models to Mixed Models to Semiparametric Regression
David Ruppert
Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

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

- slope jumps by b_k at κ_k, $k = 1, \ldots, K$

Fitting LIDAR data with plus functions

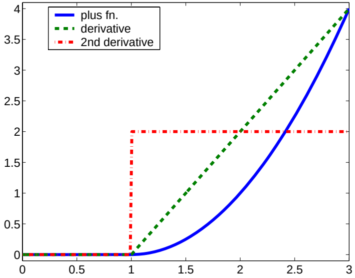


Generalization: higher degree splines

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

- pth derivative jumps by $p!\, b_k$ at κ_k
- first $p - 1$ derivatives are continuous

Quadratic "plus" function



Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples

Introduction

Linear and mixed linear models

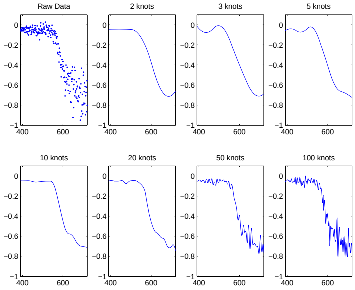
Univariate spline models

Back to examples

More general spline models

Summary

LIDAR data: ordinary Least Squares



LIDAR data: penalized least-squares

- Use matrix notation:

$$m(X_i) = \beta_0 + \beta_1 X_i + \cdots + \beta_p X_i^p$$
$$+ b_1 (X_i - \kappa_1)_+^p + \cdots + b_K (X_i - \kappa_K)_+^p$$
$$= \mathbf{X}_i^{\mathsf{T}} \boldsymbol{\beta}_X + \mathbf{B}^{\mathsf{T}}(X_i)\mathbf{b}$$

- Minimize

$$\sum_{i=1}^{n} \left\{ Y_i - (\mathbf{X}_i^{\mathsf{T}} \boldsymbol{\beta}_X + \mathbf{B}^{\mathsf{T}}(X_i)\mathbf{b}) \right\}^2 + \lambda\, \mathbf{b}^{\mathsf{T}} \mathbf{D} \mathbf{b}.$$

Penalized least-squares, cont.

Linear Statistical Models to Mixed Models to Semiparametric Regression
David Ruppert
Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

- From previous slide: minimize

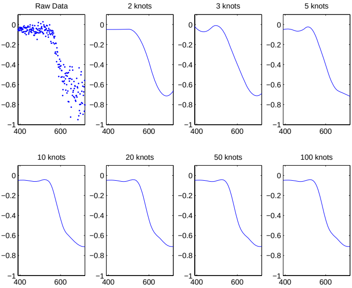
$$\sum_{i=1}^{n} \left\{ Y_i - (\mathbf{X}_i^{\mathsf{T}} \boldsymbol{\beta}_X + \mathbf{B}^{\mathsf{T}}(X_i)\mathbf{b}) \right\}^2 + \lambda\, \mathbf{b}^{\mathsf{T}} \mathbf{D} \mathbf{b}.$$

- $\lambda\, \mathbf{b}^{\mathsf{T}} \mathbf{D} \mathbf{b}$ is a penalty that prevents overfitting
- $\mathbf{D}$ is a positive semidefinite matrix
  - so the penalty is non-negative
  - Example:

$$\mathbf{D} = \mathbf{I}$$

- λ controls that amount of penalization
- the choice of λ is crucial

Penalized Least Squares



Choice of λ is crucial

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples

Introduction

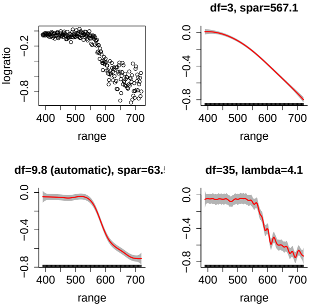
Linear and mixed linear models

Univariate spline models

Back to examples

More general spline models

Summary



How can λ be chosen?

The smoothing parameter λ can be chosen automatically using mixed model software

Ridge Regression

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

From earlier slide:

$$\sum_{i=1}^{n} \left\{ Y - (\mathbf{X}_i^\mathsf{T} \boldsymbol{\beta}_X + \mathbf{B}^\mathsf{T}(X_i)\mathbf{b}) \right\}^2 + \lambda\, \mathbf{b}^\mathsf{T} \mathbf{D} \mathbf{b}.$$

Let $\mathcal{X}$ have row $\left(\mathbf{X}_i^\mathsf{T} \quad \mathbf{B}^\mathsf{T}(X_i)\right)$. Then

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_X \\ \widehat{\mathbf{b}} \end{pmatrix} = \left\{ \mathcal{X}^\mathsf{T}\mathcal{X} + \lambda \ \mathrm{blockdiag}(\mathbf{0}, \mathbf{D}) \right\}^{-1} \mathcal{X}^\mathsf{T}\mathbf{Y}.$$

- 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:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}$$

  where
  - $\mathbf{b}$ is $N(0, \sigma_b^2 \boldsymbol{\Sigma}_b)$
  - $\boldsymbol{\varepsilon}$ is $N(0, \sigma_\epsilon^2 \mathbf{I})$
  - $\mathbf{X}\boldsymbol{\beta}$ are the "fixed effects"
  - $\mathbf{Z}\mathbf{b}$ are the "random effects"
- **Henderson's equations.**

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^\mathsf{T}\mathbf{X} & \mathbf{X}^\mathsf{T}\mathbf{Z} \\ \mathbf{Z}^\mathsf{T}\mathbf{X} & \mathbf{Z}^\mathsf{T}\mathbf{Z} + \lambda\boldsymbol{\Sigma}_b^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^\mathsf{T}\mathbf{Y} \\ \mathbf{Z}^\mathsf{T}\mathbf{Y} \end{pmatrix}.$$

$$\lambda = \frac{\sigma_\epsilon^2}{\sigma_b^2}.$$

Linear Mixed Models

From previous slides:

Ridge regression: Let $\mathcal{X}$ have row $\left(\mathbf{X}_i^\mathsf{T} \quad \mathbf{B}^\mathsf{T}(X_i)\right)$. Then

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_X \\ \widehat{\mathbf{b}} \end{pmatrix} = \left\{\mathcal{X}^\mathsf{T}\mathcal{X} + \lambda \text{ blockdiag}(\mathbf{0}, \mathbf{0}, \mathbf{D})\right\}^{-1} \mathcal{X}^\mathsf{T}\mathbf{Y}.$$

Linear mixed model:

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^\mathsf{T}\mathbf{X} & \mathbf{X}^\mathsf{T}\mathbf{Z} \\ \mathbf{Z}^\mathsf{T}\mathbf{X} & \mathbf{Z}^\mathsf{T}\mathbf{Z} + \lambda\boldsymbol{\Sigma}_b^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^\mathsf{T}\mathbf{Y} \\ \mathbf{Z}^\mathsf{T}\mathbf{Y} \end{pmatrix}$$

$$= \left\{(\mathbf{X} \quad \mathbf{Z})^\mathsf{T}(\mathbf{X} \quad \mathbf{Z}) + \lambda \,\text{blockdiag}(\mathbf{0}, \boldsymbol{\Sigma}_b^{-1})\right\}^{-1} (\mathbf{X} \quad \mathbf{Z})^\mathsf{T}\mathbf{Y}$$

Selecting λ

To choose λ use:

1. one of several model selection criteria:
   - cross-validation (CV)
   - generalized cross-validation (GCV)
   - AIC
   - C_P
2. ML or REML in mixed model framework

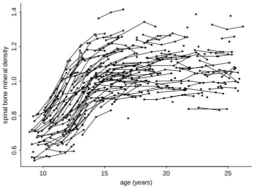
Modeling the blood lead and IQ data

For the jth measurements on the ith subject:

$$\text{IQ}_{ij} = b_i + m(\text{lead}_{ij}) + \beta_1 X_{ij}^1 + \cdots + \beta_L X_{ij}^L + \epsilon_{ij}$$

- $m(\cdot)$ 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}^ℓ is the value of the ℓth confounder, $\ell = 1, \ldots, L$

Return to spinal bone mineral density study



$$\mathtt{SBMD}_{i,j} = U_i + m(\mathtt{age}_{i,j}) + \epsilon_{i,j},$$
$$i = 1, \ldots, m = 230, \quad j = i, \ldots, n_i.$$

Fixed effects

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

$$\mathbf{X} = \begin{bmatrix} 1 & \mathtt{age}_{11} \\ \vdots & \vdots \\ 1 & \mathtt{age}_{1n_1} \\ \vdots & \vdots \\ 1 & \mathtt{age}_{m1} \\ \vdots & \vdots \\ 1 & \mathtt{age}_{mn_m} \end{bmatrix}$$

Random effects

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

Random effects

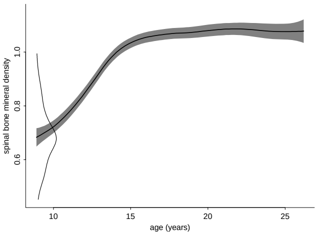
Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

$$\mathbf{u} = \begin{bmatrix} U_1 \\ \vdots \\ U_m \\ b_1 \\ \vdots \\ b_K \end{bmatrix}$$

Random effects



Variability bars on $\widehat{m}$ and estimated density of U_i

Broken down by ethnicity

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples

Introduction

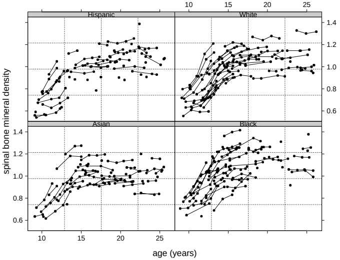
Linear and mixed linear models

Univariate spline models

Back to examples

More general spline models

Summary



Model with ethnicity effects

$$
\begin{aligned}
\mathtt{SBMD}_{ij} &= U_i + m(\mathtt{age}_{ij}) + \beta_1 \mathtt{black}_i + \beta_2 \mathtt{hispanic}_i \\
&\quad + \beta_3 \mathtt{white}_i + \varepsilon_{ij}, \quad 1 \le j \le n_i, \quad 1 \le i \le m.
\end{aligned}
$$

Asian is the reference group.

Model with ethnicity effects

Linear Statistical Models to Mixed Models to Semiparametric Regression
David Ruppert
Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

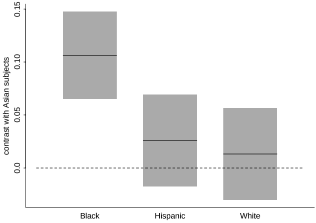
Only requires an expansion of the fixed effects by adding the columns

$$
\begin{bmatrix}
\mathtt{black}_1 & \mathtt{hispanic}_1 & \mathtt{white}_1 \\
\vdots & \vdots & \vdots \\
\mathtt{black}_1 & \mathtt{hispanic}_1 & \mathtt{white}_1 \\
\vdots & \vdots & \vdots \\
\mathtt{black}_m & \mathtt{hispanic}_m & \mathtt{white}_m \\
\vdots & \vdots & \vdots \\
\mathtt{black}_m & \mathtt{hispanic}_m & \mathtt{white}_m
\end{bmatrix}
$$

Ethnicity effects

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert



Possible enrichment of the model

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

- 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

Linear Statistical Models to Mixed Models to Semiparametric Regression
David Ruppert
Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

$$\begin{aligned} Y_i &= \beta_0 \\ &+ \beta_{x,1} X_i + b_{x,1}(X_i - \kappa_{x,1})_+ + \cdots + b_{x,K}(X_i - \kappa_{x,K_x})_+ \\ &+ \beta_{z,1} Z_i + b_{z,1}(Z_i - \kappa_{z,1})_+ + \cdots + b_{z,K}(Z_i - \kappa_{z,K_z})_+ \\ &+ \epsilon_i \end{aligned}$$

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

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples

Introduction

Linear and mixed linear models

Univariate spline models

Back to examples

More general spline models

Summary

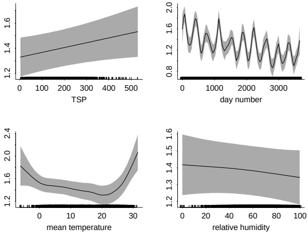
Data:

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

Additive Model:

$$\begin{aligned}\sqrt{\mathtt{mortality}_t} &= \beta_0 + \beta\,\mathrm{TSP}_t + f_1(t) + f_2(\mathtt{temperature}_t) \\ &+ f_3(\mathtt{humidity}_t) + \varepsilon_t\end{aligned}$$

Milan study: results



Other models that fit in this framework

Linear Statistical Models to Mixed Models to Semiparametric Regression

David Ruppert

Two Examples
Introduction
Linear and mixed linear models
Univariate spline models
Back to examples
More general spline models
Summary

- 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