Simple fitting of subject-specific curves for longitudinal data

Maria Durbán y Matt Wand

Dept. Estadística y Econometría, Universidad Carlos III de Madrid

Department of Statistics University of New South Wales

IX Conferencia Española de Biometría, A Coruña, Mayo 2003
What is this talk about?
What is this talk about?

- Introduction
  - The data
  - P-splines
  - Mixed model representation of P-splines
What is this talk about?

- Introduction
  - The data
  - P-splines
  - Mixed model representation of P-splines

- P-spline models for longitudinal data.
  - Subject-specific Curves
What is this talk about?

- Introduction
  - The data
  - P-splines
  - Mixed model representation of P-splines

- P-spline models for longitudinal data.
  - Subject-specific Curves

- Analysis of dogs data.
The data

Objective: Determine the effect of 4 surgical treatments on coronary sinus potassium in dogs

- 36 dogs
- 4 treatments
- 7 measurements per dog
Smoothing

- For each dog, potassium concentration varies smoothly along time
Smoothing

- For each dog, potassium concentration varies smoothly along time

- A suitable model for these data could be:

\[ y = f(t) + \epsilon \]

where \( t \) is the covariate (Time) and \( f \) is a smooth function of \( t \) which depends on \( \lambda = \text{smoothing parameter} \)
**Smoothing**

- For each dog, potassium concentration varies smoothly along time

- A suitable model for these data could be:

\[
y = f(t) + \epsilon
\]

where \( t \) is the covariate (Time) and \( f \) is a smooth function of \( t \) which depends on \( \lambda = \)smoothing parameter

- Smoothing methods fall into two groups:
  - Specified by the fitting procedure: **Kernels**
  - Solution of a minimisation problem: **Splines**
A graph showing the change in potassium over time.
B-splines

- We write $f = B a$, $B$ is a B-spline basis and $a$ are coefficients
- B-spline: bell-shaped like Gauss curve
- Used as the basis for the regression
- Polynomial pieces smoothly joining at the knots
**B-splines**

- We write $f = B\alpha$, $B$ is a B-spline basis and $\alpha$ are coefficients.
- B-spline: bell-shaped like Gauss curve.
- Used as the basis for the regression.
- Polynomial pieces smoothly joining at the knots.
Penalized splines (P-splines)
Penalized splines (P-splines)

- Regress on $k$ B-splines
- Difference penalty on the regression coefficients
Penalized splines (P-splines)

- Regress on $k$ B-splines
- Difference penalty on the regression coefficients
- Minimise penalised sum of squares

$$y = f(x) + \epsilon \quad f(x) \approx Ba \quad S = (y - Ba)'(y - Ba) + \lambda a'D'Da$$

$$\hat{a} = (B'B + \lambda D'D)^{-1}B'y$$
Penalized splines (P-splines)

- Regress on \( k \) B-splines
- Difference penalty on the regression coefficients
- Minimise penalised sum of squares

\[
y = f(x) + \epsilon \quad f(x) \approx Ba \quad S = (y - Ba)'(y - Ba) + \lambda a'D'Da
\]

\[
\hat{a} = (B'B + \lambda D'D)^{-1}B'y
\]

- Moderate dimension (\( k \) between 10 and 40)
- Computationally faster than smoothing splines
Mixed model representation of P-splines
Mixed model representation of P-splines

\[ y = B\alpha + \epsilon \]

\[ \downarrow \quad \alpha \text{ treated as mixed effects} \]

\[ y = X\alpha + Zu + \epsilon \quad u \sim N(0, \sigma_u^2) \]
Mixed model representation of P-splines

\[ y = B\alpha + \epsilon \]

\[ \Downarrow \quad \alpha \text{ treated as mixed effects} \]

\[ y = X\alpha + Zu + \epsilon \quad u \sim N(0, \sigma_u^2) \]

\[ \lambda = \sigma^2 / \sigma_u^2 \]
Mixed model representation of P-splines

\[ y = Ba + \epsilon \]

\[ \downarrow \quad \alpha \text{ treated as mixed effects} \]

\[ y = X\alpha + Zu + \epsilon \quad u \sim N(0, \sigma_u^2) \]

\[ \lambda = \sigma^2 / \sigma_u^2 \]

Best Linear Unbiased Predictor = Penalized likelihood fit
Advantages
Advantages

- No need for backfitting in additive models
Advantages

- No need for backfitting in additive models
- Fast with large datasets, not posible with smoothing splines
Advantages

- No need for backfitting in additive models
- Fast with large datasets, not possible with smoothing splines
- Easily implemented in Splus and R

```
lme(y~X-1,random=pdIdent(~Z-1))
```
Advantages

• No need for backfitting in additive models

• Fast with large datasets, not possible with smoothing splines

• Easily implemented in Splus and R

  \texttt{lme(}y\sim X-1,\texttt{random=pdIdent(}\sim Z-1)\texttt{)}

• GAM $\Rightarrow$ GLMM
P-splines with truncated lines basis

\[ y_{ij} = f(t_{ij}) + \epsilon_{ij} \quad 1 \leq j \leq 7 \]

\[ f(t_{ij}) = \alpha_0 + \alpha_1 t_{ij} + \sum_{k=1}^{K} u_k (t_{ij} - \kappa_k)_{+} \quad u_k \sim N(0, \sigma_u^2) \]

\[ \downarrow \]

\[ y = X\alpha + Zu + \epsilon \quad Cov \begin{bmatrix} u \\ \epsilon \end{bmatrix} = \begin{bmatrix} \sigma_u^2 I & 0 \\ 0 & \sigma^2 I \end{bmatrix} \]

\[ Z = \begin{bmatrix} (t_{i1} - \kappa_1)_{+} & \cdots & (t_{i1} - \kappa_K)_{+} \\ \vdots & \ddots & \vdots \\ (t_{i7} - \kappa_1)_{+} & \cdots & (t_{i7} - \kappa_K)_{+} \end{bmatrix} \]
P-spline models for longitudinal data

Basic Model \[ y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \beta_i + \epsilon_{ij} \quad 1 \leq j \leq 7 \quad 1 \leq i \leq 36 \]
P-spline models for longitudinal data

Basic Model \[ y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \beta_i + \epsilon_{ij} \quad 1 \leq j \leq 7 \quad 1 \leq i \leq 36 \]

\[ \Downarrow \]

Relax linearity assumption

Model A \[ y_{ij} = f_{gr(i)}(t_{ij}) + \beta_i + \epsilon_{ij} \quad 1 \leq gr(i) \leq 4 \]
P-spline models for longitudinal data

Basic Model \( y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \beta_{i0} + \epsilon_{ij} \quad 1 \leq j \leq 7 \quad 1 \leq i \leq 36 \)

\[ \Downarrow \quad \text{Relax linearity assumption} \]

Model A \( y_{ij} = f_{gr(i)}(t_{ij}) + \beta_{i0} + \epsilon_{ij} \quad 1 \leq gr(i) \leq 4 \)

\[ \Downarrow \quad \text{Add random slope + general covariance matrix} \]

Model B \( y_{ij} = f_{gr(i)}(t_{ij}) + \beta_{i0} + \beta_{i1} t_{ij} + \epsilon_{ij} \)
P-spline models for longitudinal data

Basic Model \[ y_{ij} = \alpha_0 + \alpha_1 t_{ij} + \beta_{i0} + \epsilon_{ij} \quad 1 \leq j \leq 7 \quad 1 \leq i \leq 36 \]

\[ \Downarrow \quad \text{Relax linearity assumption} \]

Model A \[ y_{ij} = f_{gr(i)}(t_{ij}) + \beta_{i0} + \epsilon_{ij} \quad 1 \leq gr(i) \leq 4 \]

\[ \Downarrow \quad \text{Add random slope + general covariance matrix} \]

Model B \[ y_{ij} = f_{gr(i)}(t_{ij}) + \beta_{i0} + \beta_{i1} t_{ij} + \epsilon_{ij} \]

\[ \Downarrow \quad \text{Subject specific curves} \]

Model C \[ y_{ij} = f_{gr(j)}(t_{ij}) + g_{i}(t_{ij}) + \epsilon_{ij} \]
The mixed model associated to Model A is:

\[ y = X\alpha + Zu + \epsilon \]

\[ \text{Cov} \begin{bmatrix} u \\ \epsilon \end{bmatrix} = \begin{bmatrix} \sigma_{\beta_0}^2 I & 0 & 0 \\ 0 & \Sigma_{gr} & 0 \\ 0 & 0 & \sigma^2 I \end{bmatrix} \]

\[ X = \begin{bmatrix} X_{\text{time}} \\ \vdots \\ X_{\text{time}} \end{bmatrix} \quad X_{\text{time}} = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_7 \end{bmatrix} \quad Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_2 \\ \vdots \\ Z_3 \\ \vdots \\ Z_4 \end{bmatrix} \]

\[ Z_{gr(i)} = \begin{bmatrix} Z_{\text{time}} \\ \vdots \\ Z_{\text{time}} \end{bmatrix} \quad \Sigma_{gr} = \begin{bmatrix} \sigma_1^2 I & \sigma_2^2 I & \sigma_3^2 I & \sigma_4^2 I \end{bmatrix} \]
The mixed model associated to Model B is:

\[ y = X\alpha + Zu + \epsilon \]

\[ \text{Cov} \begin{bmatrix} u \\ \epsilon \end{bmatrix} = \begin{bmatrix} \Sigma_{gr} & 0 \\ 0 & \text{blockdiag}(\Sigma) & 0 \\ 0 & 0 & \sigma^2 I \end{bmatrix} \]

\[ Z = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix} = \begin{bmatrix} X_{\text{time}} & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & X_{\text{time}} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X_{\text{time}} \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} \]
The mixed model associated to Model C is:

\[ y = X\alpha + Zu + \epsilon \]

\[
\text{Cov} \begin{bmatrix} u \\ \epsilon \end{bmatrix} = \begin{bmatrix} \Sigma_{gr} & 0 & 0 & 0 \\ 0 & \text{blockdiag}(\Sigma) & 0 & 0 \\ 0 & 0 & \sigma_c^2 I & 0 \\ 0 & 0 & 0 & \sigma^2 I \end{bmatrix}
\]

\[
Z = \begin{bmatrix}
Z_1 & X_{\text{time}} & 0 & \cdots & 0 & Z_{\text{time}} & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
X_{\text{time}} & 0 & \cdots & 0 & Z_{\text{time}} & 0 & \cdots & 0 \\
0 & X_{\text{time}} & \cdots & 0 & 0 & Z_{\text{time}} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & X_{\text{time}} & \cdots & 0 & 0 & Z_{\text{time}} & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & X_{\text{time}} & 0 & 0 & \cdots & Z_{\text{time}} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
Z_4 & 0 & 0 & \cdots & X_{\text{time}} & 0 & 0 & \cdots & Z_{\text{time}} \\
\end{bmatrix}
\]
Conclusions and work in progress
Conclusions and work in progress

- $P$-splines are useful tool to model longitudinal data
- $P$-splines as mixed models
- Easy to implement in standard software
- Model selection
References
References


