Statistical Models

So far we have focussed on *exploratory data analysis*

- Smoothing
- Functional covariance
- Functional PCA

Now we wish to examine predictive relationships $\rightarrow$ generalization of linear models.

$$y_i = \alpha + \sum \beta_j x_{ij} + \epsilon_i$$

Functional Linear Regression

$$y_i = \alpha + \mathbf{x}_i \beta + \epsilon_i$$

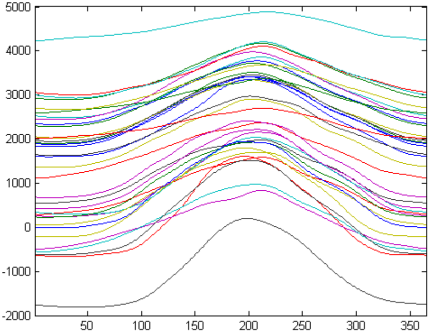
Three different scenarios

- Functional covariate, scalar response
- Scalar covariate, functional response
- Functional covariate, functional response

We will deal with each in turn.

Models for Scalar Responses

Temperature curves shifted by total annual precipitation:



We want to relate annual precipitation to the *shape* of the temperature profile (see handout or text).

A First Idea

We observe $y_i, x_i(t)$

Choose $t_1, \dots, t_k$

Then we set

$$
\begin{aligned}
y_i &= \alpha + \sum \beta_j x_i(t_j) + \epsilon_i \\
&= \alpha + \mathbf{x}_i \beta + \epsilon
\end{aligned}
$$

And do linear regression.

But how many $t_1, \dots, t_k$ and which ones?

In the Limit

If we let $t_1, \ldots$ get increasingly dense

$$y_i = \alpha + \sum \beta_j x_i(t_j) + \epsilon_i = \alpha + \mathbf{x}_i \beta + \epsilon$$

becomes

$$y_i = \alpha + \int \beta(t) x_i(t) dt + \epsilon_i$$

Minimize squared error:

$$\beta(t) = \operatorname{argmin} \sum \left(y_i - \alpha - \int \beta(t) x_i(t) dt \right)^2$$

Identification

Problem:

- In linear regression, we must have fewer covariates than observations.
- if I have $y_i, x_i(t)$, there are *infinitely* many covariates.

$$y_i = \alpha + \int \beta(t) x_i(t) dt + \epsilon_i$$

I can always make the $\epsilon_i = 0$

Smoothing

We want to insist that $\beta(t)$ is smooth.

Fit by penalized squared error

$$\text{PENSSE}_\lambda(\beta) = \sum_{i=1}^{n} \left(y_i - \alpha - \int \beta(t) x_i(t) dt \right)^2 + \lambda \int [L\beta(t)]^2 \, dt$$

Very much like smoothing.

Still need to represent $\beta(t)$ – use a basis expansion

$$\beta(t) = \sum c_i \phi_i(t)$$

Choosing A Basis

Smoothing problem

$$\text{PENSSE}_\lambda(\beta) = \sum_{i=1}^{n} \left(y_i - \alpha - \int \beta(t) x_i(t) dt \right)^2 + \lambda \int [L\beta(t)]^2 \, dt$$

has a unique minimizer over all functions for which $\text{PENSSE}_\lambda(\beta) < \infty$.

- No explicit representation is available.
- However, using a rich enough basis will approximate it well.
- Usually, "rich enough" isn't too rich.
- If $x_i(t)$ are represented by a basis, using the same basis often works well.

Calculation

$$\begin{aligned} y_i &= \alpha + \int \beta(t)x_i(t)dt + \epsilon_i = \alpha + \left[\int \Phi(t)x_i(t)dt\right]\mathbf{c} + \epsilon_i \\ &= \alpha + \mathbf{x}_i\mathbf{c} + \epsilon_i \end{aligned}$$

so

$$\mathbf{y} = Z\begin{bmatrix} \alpha \\ \mathbf{c} \end{bmatrix} + \epsilon$$

and

$$[\hat{\alpha}\ \hat{\mathbf{c}}^T]^T = \left(Z^TZ + \lambda R\right)^{-1} Z^T\mathbf{y}$$

Then

$$\hat{\mathbf{y}} = \int \hat{\beta}(t)x_i(t)dt = Z\begin{bmatrix} \hat{\alpha} \\ \hat{\mathbf{c}} \end{bmatrix} = S\mathbf{y}$$

Confidence Intervals

Following from smoothing methods we have that

$$\mathrm{Var}\begin{bmatrix} \hat{\alpha} \\ \hat{\mathbf{c}} \end{bmatrix} = \sigma_e^2 \left(Z^T Z + \lambda R\right)^{-1} Z^T Z \left(Z^T Z + \lambda R\right)^{-1}$$

Assuming independent

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

Estimate

$$\hat{\sigma}_e^2 = SSE/(n - df),\ df = \mathrm{trace}(S)$$

And confidence intervals are

$$\Phi(t)\hat{\mathbf{c}} \pm 2\sqrt{\Phi(t)^T \mathrm{Var}[\hat{\mathbf{c}}]\Phi(t)}$$

Multivariate Functional Linear Regression

What if there are scalar covariates $\mathbf{z}$ and multiple functional covariates $x_1(t), \ldots, x_K(t)$?

$$y_i = \alpha + \mathbf{z}_i\gamma + \sum_{j=1}^{k} \int \beta_j(t)x_{ij}(t)dt + \epsilon_i$$

Then the penalized sum of squares is

$$\sum_{i=1}^{n} \left(y_i - \alpha - \mathbf{z}_i\gamma + \sum_{j=1}^{k} \int \beta_j(t)x_{ij}(t)dt \right)^2 + \sum_{j=1}^{K} \lambda_j \int \left[L_j\beta_j(t)\right]^2 dt$$

Summary

- Functional linear regression: move from summation to integration
- Identifiability – use a smoothing penalty (remarkable similarity to smoothing)
- Cross validation for smoothing penalty choice
- Confidence intervals based on residual errors
- Can generalized to multiple and mixed covariates