Functional regression analysis using \texttt{R}

Christian Ritz

Statistics Group
Faculty of Life Sciences (LIFE)
University of Copenhagen, Denmark

Dortmund, August 13 2008
Examples

What are functional data?

- Activity and disease patterns
  *(eg. monitoring birds, children or insects over time)*

- Animal and human growth curves
  *(eg. weight gain in pigs and dietary studies)*

- Fluorescence curves
  *(eg. photosynthesis processes over time (Ritz and Streibig, 2008))*

- Reproduction histories
  *(eg. longevity of medflies (Chiou et al, 2003))*
Examples

What are functional data?

- Activity and disease patterns
  (eg. monitoring birds, children or insects over time)

- Animal and human growth curves
  (eg. weight gain in pigs and dietary studies)

- Fluorescence curves
  (eg. photosynthesis processes over time (Ritz and Streibig, 2008))

- Reproduction histories
  (eg. longevity of medflies (Chiou et al, 2003))
More about fluorescence curves

- **Experiment:**
  - dark-adapted leaves exposed to light
    
    \textit{(only the first seconds of this process is recorded!)}

- **Functional response:**
  - proportion of light not used in the photosynthesis

- **High throughput measurements:**
  - fast and non-invasive
  - informative long before visual effects

- Curve trajectory changes with species and stress level
Observed fluorescence curves

Three *replicates*
More about functional data

Common features:

- *repeated measurements* on the same subject or unit
- basic observation: *smooth function*
  (in practice observed discretely on a grid)

Use of functional data:

- classification/clustering
- ANOVA- and regression-like models
- prediction

Smoothness being exploited in various ways
More about functional data

Common features:

- *repeated measurements* on the same subject or unit
- basic observation: *smooth function*
  (in practice observed discretely on a grid)

Use of functional data:

- classification/clustering
- ANOVA- and regression-like models
- prediction

Smoothness being exploited in various ways
Functional regression

How to relate functional responses to scalar, explanatory variables?

Available functional regressions models:

- **Semi-parametric approaches:**
  - additive effects models (Ramsay & Silverman, 2005) ([R package fda](https://cran.r-project.org/web/packages/fda/index.html) on CRAN and R-Forge)
  - multiplicative effects models (Chiou *et al*., 2003) ([R package fmer](https://cran.r-project.org/web/packages/fmer/index.html) soon on CRAN)

- ...
Functional regression

How to relate functional responses to scalar, explanatory variables?

Available functional regressions models:

- Semi-parametric approaches:
  - additive effects models (Ramsay & Silverman, 2005) (R package fda on CRAN and R-Forge)
  - multiplicative effects models (Chiou et al., 2003) (R package fmer soon on CRAN)
  - ...

Christian Ritz  (Statistics at LIFE)
Functional multiplicative effects models

A little notation:

- $y_i : T \mapsto \mathbb{R}$ is a function ($i = 1, \ldots, N$)
- $T \subseteq \mathbb{R}$ is the interval
- Observed at points $t_1, \ldots, t_K$ ($K$ large)

Multiplicative effects regression model:

$$E(y_i(t) | z_i) = \psi(t, z_i)\mu(t)$$

Right-hand side:

- $\mu$: capturing the overall average trend
- $\psi$: multiplicative effects: low-degree polynomials in $t$ with coefficients depending on explanatory variable $z_i$
Estimation – in two steps

1. Non-parametric estimation:
   - $\mu$: smoothing based on all curves (R package `KernSmooth`)
   - coefficients in $\psi$: obtained using least squares

2. Parametric or semi-parametric estimation for coefficients:
   1. choose GLM (`glm()`) or quasi-likelihood model
   2. iterative estimation: (IWLS+smoothing)
      - link and/or variance functions (not in GLM case)
      - parameters in linear predictor
Using R

library(fmer)

bo.m1 <- fmerm(fluo2 ~ log(time), id2, id0, data = barleyOat, quad = TRUE)

Arguments to fmerm:

- **fluo2**: function values
- **log(time)**: grid values
- **id2**: curve id (54 curves in total)
- **id0**: treatment factor
- **quad**: $\psi$ quadratic in $t$
Model fit components

- Estimated overall mean
- Estimated regression curves (use \texttt{plot} method)
- For each coefficient in $\psi$:
  - estimated link and variance functions
  - estimated parameters (use \texttt{summary} method)
  - fitted values and residuals (use \texttt{fitted and residuals})
Fitted fluorescence curve

Using the `plot` method:
Pros and cons

- **Advantages:**
  - non-parametric modelling of *the form* of the curves (separating the time effect from other effects)
  - parametric regression models for *the differences* between curves
  - graphical model check available (*ratioPlot*)

- **Drawbacks:**
  - automatic bandwidth selection needed (used repeatedly)
  - two-step estimation procedure (some variation lost)
Pros and cons

- **Advantages:**
  - non-parametric modelling of *the form* of the curves (separating the time effect from other effects)
  - parametric regression models for *the differences* between curves
  - graphical model check available (*ratioPlot*)

- **Drawbacks:**
  - automatic bandwidth selection needed (used repeatedly)
  - two-step estimation procedure (some variation lost)
Future R work

- Testing on more datasets!!!

- Setting up a modular structure for model fitting:
  - one function per step in estimation procedure
  - plug-ins for different smoothing methods
  - choice between bandwidth selection methods
  - more flexible model specification

- Constructing extractors for various fit components
Future theoretical work

- Joint estimation
- Extended modelling including the residual process
- Model checking diagnostics
References

