Functional Data Analysis (Lecture 3)

Zoltán Szabó

October 18, 2016
Reminder, contents

- Last time: $PEN_L$-regularized least squares.
- Today:
  1. smoothing with constraints,
  2. positivity: daily precipitation, counts of errors, ... 
  3. monotonicity: growth curves (height, length); in registration!
  4. probability density function.
Reminder, contents

- Last time: $PEN_L$-regularized least squares.
- Today:
  1. smoothing with constraints,
     - positivity: daily precipitation, counts of errors, ... 
     - monotonicity: growth curves (height, length); in registration!
     - probability density function.
  2. curve registration:
     - shift-, feature-, continuous registration.
Smoothing with constraints
Idea: parameterize $\log[x(t)]$, $\log := \ln$. 

Notes: $J$: nonquadratic in $c$ \(\Rightarrow\) iterative solvers, typically: $c_0 = 0$, fast convergence.
Idea: parameterize log[$x(t)$], log := ln.

Objective:

$$x(t) = e^{W(t)}, \ W(t) = c^T \phi(t),$$

$$J(c) = \left( y - e^{W(t)} \right)^T W \left( y - e^{W(t)} \right) + \lambda \| LW \|^2 \rightarrow \min_{c \in \mathbb{R}^B}.$$
Idea: parameterize $\log[x(t)]$, $\log := \ln$.

Objective:

$$x(t) = e^{W(t)}, \quad W(t) = c^T \phi(t),$$

$$J(c) = \left[ y - e^{W(t)} \right]^T W \left[ y - e^{W(t)} \right] + \lambda \| LW \|^2 \rightarrow \min_{c \in \mathbb{R}^B}.$$

Notes:

- $J$: nonquadratic in $c \Rightarrow$ iterative solvers,
- typically: $c_0 = 0$, fast convergence.
Motivation: \( x(t) = e^{wt} \iff Dx(t) = wx(t) \Rightarrow \text{Let } Dx(t) = w(t)x(t). \)
Motivation: $x(t) = e^{wt} \leftrightarrow Dx(t) = wx(t) \Rightarrow$ Let $Dx(t) = w(t)x(t)$.

Note: $w(t) > 0$ more rapid increase as $x(t)$ grows.
Motivation: \( x(t) = e^{wt} \iff Dx(t) = wx(t) \Rightarrow \text{Let } Dx(t) = w(t)x(t). \)

Note: \( w(t) > 0 \) more rapid increase as \( x(t) \) grows.

Solution:

\[
\begin{align*}
    x(t) &= x(t_0) e^{\int_{t_0}^{t} w(u) du} \\
    &= C e^{\log(C) \int_{t_0}^{t} w(u) du} \\
    &= e^{\log(C) \int_{t_0}^{t} w(u) du =: W(t)}
\end{align*}
\]

\((*)\): if \( C = x(t_0) > 0 \). Else: take \( -x(t) \).
Smoothing with monotonicity: explicit way

Idea:

\[ x: \text{strictly increasing} \iff Dx: \text{positive}. \]
**Smoothing with monotonicity: explicit way**

- **Idea:**
  \[ x: \text{strictly increasing} \Leftrightarrow Dx: \text{positive}. \]

- **Solution of** \( Dx(t) = e^{W(t)} \):
  \[
  x(t) = C \int_{t_0}^{t} e^{W(u)} \, du.
  \]
  \[= x(t_0) \]
Smoothing with monotonicity: differential equation

Idea: $D(Dx) = w(Dx)$.

Note: solving it & suitable $W(t)$ choice gives again

$$x(t) = C \int_{t_0}^{t} e^{W(u)} du.$$
Smoothing with pdf constraint

- Density estimation: given $t_1, \ldots, t_n \overset{i.i.d.}{\sim} p$; task: $\hat{p}$. 
Smoothing with pdf constraint

- Density estimation: given \( t_1, \ldots, t_n \stackrel{i.i.d.}{\sim} p \); task: \( \hat{p} \).

- Idea:
  1. \( p \): positive function \( \Rightarrow p(t) = Ce^{W(t)} \),
  2. with a \( \int p(t) dt = 1 \) constraint \( \Rightarrow C = \frac{1}{\int e^{W(t)} dt} \).
Smoothing with pdf constraint

- Density estimation: given \( t_1, \ldots, t_n \sim i.i.d. \ p \); task: \( \hat{p} \).
- Idea:
  1. \( p \): positive function \( \Rightarrow p(t) = C e^{\int W(t) dt} \),
  2. with a \( \int p(t) dt = 1 \) constraint \( \Rightarrow C = \frac{1}{\int e^{\int W(t) dt}} \).
- Use ML estimation, \( W(t) = c^T \phi(t) \):

\[
\max_{W/c} \log [p(t; W)]
\]
Density estimation: given $t_1, \ldots, t_n \overset{i.i.d.}{\sim} p$; task: $\hat{p}$.

Idea:
1. $p$: positive function $\Rightarrow p(t) = C e^{W(t)}$,
2. with a $\int p(t)dt = 1$ constraint $\Rightarrow C = \frac{1}{\int e^{W(t)}dt}$.

Use ML estimation, $W(t) = c^T \phi(t)$:

$$\max_{W/c} \log \left[ p(t; W) \right] = \sum_{i=1}^{n} \log p(t_i; W)$$
Density estimation: given \( t_1, \ldots, t_n \overset{i.i.d.}{\sim} p \); task: \( \hat{p} \).

Idea:

1. \( p \): positive function \( \Rightarrow p(t) = Ce^{W(t)} \),
2. with a \( \int p(t)dt = 1 \) constraint \( \Rightarrow C = \frac{1}{\int e^{W(t)}dt} \).

Use ML estimation, \( W(t) = c^T\phi(t) \):

\[
\max_{W/c} \log \left[ p(t; W) \right] = \sum_{i=1}^{n} \log p(t_i; W)
\]

\[
= \sum_{i=1}^{n} W(t_i) + \log(C)
\]
Smoothing with pdf constraint

Density estimation: given \( t_1, \ldots, t_n \sim p \); task: \( \hat{p} \).

Idea:

1. \( p \): positive function \( \Rightarrow p(t) = Ce^{W(t)} \),
2. with a \( \int p(t)dt = 1 \) constraint \( \Rightarrow C = \frac{1}{\int e^{W(t)}dt} \).

Use ML estimation, \( W(t) = c^T \phi(t) \):

\[
\begin{align*}
\max_{W/c} \log [p(t; W)] &= \sum_{i=1}^{n} \log p(t_i; W) \\
&= \sum_{i=1}^{n} W(t_i) + \log(C) = \sum_{i=1}^{n} c^T \phi(t_i) + n \log(C).
\end{align*}
\]
Density estimation: given \( t_1, \ldots, t_n \overset{i.i.d.}{\sim} p \); task: \( \hat{p} \).

Idea:

1. \( p \): positive function \( \Rightarrow p(t) = Ce^{W(t)} \),
2. with a \( \int p(t) dt = 1 \) constraint \( \Rightarrow C = \frac{1}{\int e^{W(t)} dt} \).

Use ML estimation, \( W(t) = c^T \phi(t) \):

\[
\max_{W/c} \log [p(t; W)] = \sum_{i=1}^{n} \log p(t_i; W) \\
= \sum_{i=1}^{n} W(t_i) + \log(C) = \sum_{i=1}^{n} c^T \phi(t_i) + n \log(C).
\]

Penalty: \( L = D^3 \Rightarrow W \approx \text{quadratic} \leftrightarrow p \approx \text{Gaussian.} \)
Smoothing with pdf constraint

- Density estimation: given $t_1, \ldots, t_n \overset{i.i.d.}{\sim} p$; task: $\hat{p}$.
- Idea:
  1. $p$: positive function $\Rightarrow p(t) = Ce^{W(t)}$,
  2. with a $\int p(t) dt = 1$ constraint $\Rightarrow C = \frac{1}{\int e^{W(t)} dt}$.
- Use ML estimation, $W(t) = c^T \phi(t)$:

$$
\max_{W/c} \log [p(t; W)] = \sum_{i=1}^{n} \log p(t_i; W)
= \sum_{i=1}^{n} W(t_i) + \log(C) = \sum_{i=1}^{n} c^T \phi(t_i) + n \log(C).
$$

- Penalty: $L = D^3 \Rightarrow W \approx \text{quadratic} \leftrightarrow p \approx \text{Gaussian}$.
- Objective:

$$
J(c) = -\sum_{i=1}^{n} c^T \phi(t_i) + \lambda \int [LW(t)]^2 dt \rightarrow \min_{c \in \mathbb{R}^B}.
$$
Curve registration
Curve registration: motivation

Examples:

1. ∀ child grows at his/her own pace
2. weather: winter: may started at different time, ...
Curve registration: amplitude/phase variability

Amplitude variation

Phase variation

Age
Shift registration

- Given: \( \{x_i\}_{i=1}^{N} \) curves (\( \Leftarrow \) smoothing). Goal: choose \( \{\delta_i\}_{i=1}^{N} \) s.t.

\[
x_i^*(t) = x_i(t + \delta_i)
\]

-s are aligned.
Shift registration

- Given: \( \{x_i\}_{i=1}^N \) curves (\( \Leftrightarrow \) smoothing). Goal: choose \( \{\delta_i\}_{i=1}^N \) s.t.

\[
x_i^*(t) = x_i(t + \delta_i)
\]
- s are aligned.

- Registration (Procrustes method):

1. \( x_1, \ldots, x_N \xrightarrow{\text{average/smoothing}} x_0: \text{’mean’ curve.} \)
Shift registration

- Given: \( \{x_i\}_{i=1}^{N} \) curves (↔ smoothing). Goal: choose \( \{\delta_i\}_{i=1}^{N} \) s.t.
  \[
x_i^*(t) = x_i(t + \delta_i)
  \]
  -s are aligned.

- Registration (Procrustes method):
  1.\( x_1, \ldots, x_N \xrightarrow{\text{average/smoothing}} x_0: \text{‘mean’ curve.} \)
  2.\( J(\delta) = \sum_{i=1}^{N} \int [x_i(t + \delta_i) - x_0(t)]^2 dt \rightarrow \min_{\delta \in \mathbb{R}^N}. \)
Given: \( \{x_i\}_{i=1}^N \) curves (\( \Leftarrow \) smoothing). Goal: choose \( \{\delta_i\}_{i=1}^N \) s.t.

\[ x_i^*(t) = x_i(t + \delta_i) \] -s are aligned.

Registration (Procrustes method): in iteration

1. \( x_1, \ldots, x_N \xrightarrow{\text{average/smoothing}} x_0: \text{'mean' curve.} \)
2. \( J(\delta) = \sum_{i=1}^N \int [x_i(t + \delta_i) - x_0(t)]^2 dt \xrightarrow{\min_{\delta \in \mathbb{R}^N}} \)
3. curves:= registered ones, i.e. \( x_i(t) := x_i \left( t + \hat{\delta}_i \right) \forall i. \)
Algorithm ($\alpha > 0$, Newton method: $\alpha = 1$): step 2-3 in iteration

1. Input: $\{\delta_i\}_{i=1}^N$.
2. Mean curve: $x_1, \ldots, x_N \rightarrow x_0$.
3. Update the shifts:

$$
\delta_i := \delta_i - \alpha \frac{\partial J}{\partial \delta_i} \frac{\partial^2 J}{\partial \delta_i^2}, \quad (\forall i).
$$
Algorithm ($\alpha > 0$, Newton method: $\alpha = 1$): step 2-3 in iteration

1. Input: $\{\delta_i\}_{i=1}^{N}$.
2. Mean curve: $x_1, \ldots, x_N \rightarrow x_0$.
3. Update the shifts:

$$\delta_i := \delta_i - \alpha \frac{\partial J}{\partial \delta_i}, \quad (\forall i).$$

Derivatives: $J(\delta) = \sum_{i=1}^{N} \int [x_i(t + \delta_i) - x_0(t)]^2 dt \Rightarrow$

$$\frac{\partial J}{\partial \delta_i} = 2 \int [x_i(t + \delta_i) - x_0(t)] Dx_i(t + \delta_i) dt,$$

$$\frac{\partial^2 J}{\partial \delta_i^2} = 2 \int [Dx_i(t + \delta_i)]^2 + [x_i(t + \delta_i) - x_0(t)] D^2 x_i(t + \delta_i) dt,$$
Feature or landmark registration

- Idea: align only curve features.
- Assumption: features are visible on all curves. Acceleration:
Task: find \( \{h_i\}_{i=1}^{N} \) such that

\[
x_i^* = x_i \circ h_i \quad (\forall i)
\]

-s are aligned (in terms of the curve features).
Task: find \( \{h_i\}_{i=1}^N \) such that

\[
x_i^* = x_i \circ h_i \quad (\forall i)
\]

- are aligned (in terms of the curve features).

Registration:
Task: find \( \{h_i\}_{i=1}^N \) such that

\[
x_i^* = x_i \circ h_i \quad (\forall i)
\]

-\( s \) are aligned (in terms of the curve features).

Registration:

- Feature extraction: \( x_1 \mapsto t_1, \ldots, x_N \mapsto t_N, t_n \in \mathbb{R}^F \).
Task: find \( \{h_i\}_{i=1}^N \) such that

\[
x_i^* = x_i \circ h_i \quad (\forall i)
\]

-s are aligned (in terms of the curve features).

Registration:

1. Feature extraction: \( x_1 \leftrightarrow t_1, \ldots, x_N \leftrightarrow t_N, \ t_n \in \mathbb{R}^F \).
2. Mean curve: \( x_1, \ldots, x_N \xrightarrow{\text{average}} x_0, \ x_0 \leftrightarrow t_0 \in \mathbb{R}^F \).
Task: find \( \{h_i\}_{i=1}^{N} \) such that

\[
x_i^* = x_i \circ h_i \quad (\forall i)
\]

- are aligned (in terms of the curve features).

Registration:

1. Feature extraction: \( x_1 \mapsto t_1, \ldots, x_N \mapsto t_N, \ t_n \in \mathbb{R}^F \).
2. Mean curve: \( x_1, \ldots, x_N \xrightarrow{\text{average}} x_0, \ x_0 \mapsto t_0 \in \mathbb{R}^F \).
3. Warping-functions: \( \{h_i\}_{i=1}^{N} = ?, \) solve (constrained smoothing, \( \forall i \))

\[
h_i(0) = 0, \ h_i(T) = T_i
\]
\[
h_i(t_{0f}) = t_{if} \quad f = 1, \ldots, F,
\]
\( h_i : \) strictly monotone.
Feature or landmark registration

- Task: find \( \{h_i\}_{i=1}^N \) such that
  \[
  x_i^* = x_i \circ h_i \quad (\forall i)
  \]
  -s are aligned (in terms of the curve features).

- Registration: in iteration
  1. Feature extraction: \( x_1 \mapsto t_1, \ldots, x_N \mapsto t_N, t_n \in \mathbb{R}^F \).
  2. Mean curve: \( x_1, \ldots, x_N \xrightarrow{\text{average}} x_0, x_0 \mapsto t_0 \in \mathbb{R}^F \).
  3. Warping-functions: \( \{h_i\}_{i=1}^N = ?, \) solve (constrained smoothing, \( \forall i \))
     \[
     h_i(0) = 0, h_i(T) = T_i,
     \]
     \[
     h_i(t_{0f}) = t_{if} \quad f = 1, \ldots, F,
     \]
     \[ h_i : \text{strictly monotone}. \]
  4. Update curves: \( x_i := x_i \circ h_i. \)
Continuous registration

- We do not use landmarks. We register the complete curves.
- Recall (strictly monotone functions):

\[ h(t) = C + \int_0^t e^{W(u)} du. \]

- Note:
  1. \( W(u) = 0 \): internal time \( = \) clock time.
  2. \( W(u) > 0 \): warped time grows faster than clock time.
Continuous registration: fitting criterion

Idea:

1. $x_0$ and $x^*$ only differ in amplitude $\Leftrightarrow x_0(x^*)$ is linear.
Continuous registration: fitting criterion

- Idea:
  1. $x_0$ and $x^*$ only differ in amplitude $\iff x_0(x^*)$ is linear.
  2. Take $(x_0(t), x[h(t)])$ at $n$ $t$-values; $\mathbf{X} \in \mathbb{R}^{n \times 2}$. 

Zoltán Szabó  Functional Data Analysis (Lecture 3)
Continuous registration: fitting criterion

**Idea:**

1. \( x_0 \) and \( x^* \) only differ in amplitude \( \iff x_0(x^*) \) is *linear*.
2. Take \( (x_0(t), x[h(t)]) \) at \( n \) \( t \)-values; \( X \in \mathbb{R}^{n \times 2} \).
3. PCA: \( X^T X \) functional analogue

\[
T(h) = \begin{bmatrix}
\langle x_0, x_0 \rangle & \langle x_0, x \circ h \rangle \\
\langle x \circ h, x_0 \rangle & \langle x \circ h, x \circ h \rangle
\end{bmatrix} \in \mathbb{R}^{2 \times 2},
\]

where \( \langle f, g \rangle = \int f(t)g(t)dt \).

---

Zoltán Szabó  Functional Data Analysis (Lecture 3)
Continuous registration: fitting criterion

- Idea:
  1. $x_0$ and $x^*$ only differ in amplitude $\iff x_0(x^*)$ is linear.
  2. Take $(x_0(t), x[h(t)])$ at $n$ $t$-values; $X \in \mathbb{R}^{n \times 2}$.
  3. PCA: $X^T X \xrightarrow{\text{functional analogue}}$

  \[
  T(h) = \begin{bmatrix}
  \langle x_0, x_0 \rangle & \langle x_0, x \circ h \rangle \\
  \langle x \circ h, x_0 \rangle & \langle x \circ h, x \circ h \rangle
  \end{bmatrix} \in \mathbb{R}^{2 \times 2},
  \]

  where $\langle f, g \rangle = \int f(t) g(t) \, dt$.
  4. line $\iff T(h)$ has 1 non-zero eigenvalue.
Continuous registration: fitting criterion

- **Idea:**
  1. $x_0$ and $x^*$ only differ in amplitude $\iff x_0(x^*)$ is linear.
  2. Take $(x_0(t), x[h(t)])$ at $n$ $t$-values; $\mathbf{X} \in \mathbb{R}^{n \times 2}$.
  3. PCA: $\mathbf{X}^T \mathbf{X}$ \text{ functional analogue}

\[
\mathbf{T}(h) = \begin{bmatrix}
\langle x_0, x_0 \rangle & \langle x_0, x \circ h \rangle \\
\langle x \circ h, x_0 \rangle & \langle x \circ h, x \circ h \rangle
\end{bmatrix} \in \mathbb{R}^{2 \times 2},
\]

where $\langle f, g \rangle = \int f(t)g(t) \, dt$.

4. line $\iff \mathbf{T}(h)$ has 1 non-zero eigenvalue.

- **Objective**, $h = h(W)$:

\[
J(h) = \lambda_2 [\mathbf{T}(h)] + \lambda \| D^m W \|^2 \rightarrow \min_h.
\]
Continuous registration: fitting criterion

- **Idea:**
  1. $x_0$ and $x^*$ only differ in amplitude $\iff x_0(x^*)$ is *linear*.
  2. Take $(x_0(t), x[h(t)])$ at $n$ $t$-values; $\mathbf{X} \in \mathbb{R}^{n \times 2}$.
  3. **PCA:** $\mathbf{X}^T \mathbf{X} \xrightarrow{\text{functional analogue}}$

$$
\mathbf{T}(h) = \begin{bmatrix}
\langle x_0, x_0 \rangle & \langle x_0, x \circ h \rangle \\
\langle x \circ h, x_0 \rangle & \langle x \circ h, x \circ h \rangle
\end{bmatrix} \in \mathbb{R}^{2 \times 2},
$$

where $\langle f, g \rangle = \int f(t)g(t) \, dt$.

- **Objective, $h = h(W)$:**

$$J(h) = \lambda_2 [\mathbf{T}(h)] + \lambda \|D^m W\|^2 \to \min_h.$$

- $x := x_i, \forall i$. 
In practice

- Often *composition* of feature and continuous registration. ⇒
- Clearly visible landmarks: with feature registration.
- Note: dynamic programming based methods (no smoothness).
Summary

- Smoothing with constraints: positivity, monotonicity, pdf.
- Registration: shift-, feature-, continuous registration.

We covered Chapter 6-7 in [1].