ORIE 779: Functional Data Analysis

From last meeting

SiZer Background

Central Question (in application of smoothing methods):

Which features are “really there”?

- Solution, Part 1: Scale space
- Solution, Part 2: SiZer

SiZer Background (cont.)

SiZer:

Significance of Zero crossings, of the derivative, in scale space

Combines:

- needed statistical inference
- novel visualization

To get: a powerful exploratory data analysis method

Main reference:

Chaudhuri, P. and Marron, J. S. (1999) SiZer for exploration of structure in curves, *Journal of the American Statistical Association*, 94, 807-823.

SiZer Background (cont.)

Basic idea: a “bump” is characterized by:

an increase, followed by a decrease

Generalization: many “features of interest” captured by

sign of the slope of the smooth

Foundation of SiZer:

Statistical inference on slopes, *over scale space*

SiZer Background (cont.)

SiZer Visual presentation:

Color map over scale space:

- Blue: slope significantly upwards (deriv . CI above 0)
- Red: slope significantly downwards (der. CI below 0)
- Purple: slope insignificant (deriv. CI contains 0)

SiZer Background (cont.)

SiZer analysis of Fossils data:

Upper Left: Scatterplot, family of smooths, 1 highlighted

Upper Right: Scale space rep’n of family, with SiZer colors

Lower Left: SiZer map, more easy to view

Lower Right: SiCon map – replace "slope" by "curvature"

Slider (in movie viewer) highlights *different smoothing levels*

SiZer Background (cont.)

SiZer analysis of Fossils data (cont.)

Oversmoothed (top of SiZer map):

- Decreases at left, not on right

Medium smoothed (middle of SiZer map):

- Main valley significant, and left most increase
- smaller valley not statistically significant

Undersmoothed (bottom of SiZer map):

- “noise wiggles” not significant

Additional SiZer color: gray - not enough data for inference

SiZer Background (cont.)

SiZer analysis of Fossils data (cont.)

Common Question: which is “right”?

- decreases on left, then flat (top of SiZer map)
- up, then down, then up again (middle of SiZer map)
- no significant features (bottom of SiZer map)

Answer: *All* are “right”, just different “scales of view”,

i.e. “levels of resolution of data”

SiZer Background (cont.)

Simulated example 1: Marron - Wand Trimodal, #9

n=100: only one mode "significant"

n=1000: two modes now "appear from background noise"

n=10,000: finally all 3 modes are "really there"

Simulated example 2: Marron - Wand Discrete Comb, #15

- similar lessons to above
- someday: "draw" local bandwidth on SiZer map

SiZer Background (cont.)

Finance "tick data":   (time, price) of single stock transactions

Idea:

"on line" version of SiZer

for viewing and understanding trends

Notes:

- "trends" depend heavily on "scale"
- "double points" and more
- "background color" transition (flop over at top)

SiZer Background (cont.)

Internet traffic data analysis:

SiZer analysis of time series of packet times at internet hub

- across very wide range of scales
- needs more pixels than screen allows
- thus do zooming view (zoom in over time)
- zoom in to yellow bd'ry in next frame
- readjust vertical axis

SiZer Background (cont.)

Internet traffic data analysis (cont.):

Insights from SiZer analysis:

- Coarse scales: amazing amount of "significant structure"
- evidence of “self-similar fractal” type process?
- fewer significant features at small scales
- but they exist, so not Poisson process
- Poisson approximation OK at small scale???
- smooths (top part) "stable" at large scales?

SiZer Background (cont.)

Summary: Usefulness of SiZer in exploratory data analysis:

- Smoothing experts: saves time
- Smoothing beginners: avoids terrible mistakes:
  - don't find things that "aren't there"
  - do find important features
- Directly targets critical scientific question:

  Is a deeper analysis worthwhile?

SiZer Background (cont.)

Would you like to try a SiZer analysis?

Matlab software:

http://www.unc.edu/depts/statistics/postscript/papers/marron/Matlab6Software/Smoothing/

JAVA version (demo, beta): Follow the SiZer link from the Wagner Associates home page:

http://www.wagner.com/www.wagner.com/SiZer/

More details, examples and discussions:

http://www.stat.unc.edu/faculty/marron/DataAnalyses/SiZer_Intro.html

SiZer Extensions

(won't put in more class time, but ask if interested)

- 2 dimensions (main challenge: visualization)
- censored data
- hazard estimation
- length-biased estimation
- jump (change point) detection
- smoothing spline version
- time series (dependent data)

Big Challenge: what is “trend” vs. “dependence artifact”?

Independent Component Analysis

Idea: Find "directions that maximize independence"

Motivating Context: Signal Processing

"Blind Source Separation"

References:

Lee, T. W. (1998) *Independent Component Analysis: Theory and Applications*, Kluwer.

Hyvärinen and Oja (1999) *Independent Component Analysis: A Tutorial*, http://www.cis.hut.fi/projects/ica

Hyvärinen, A., Karhunen, J. and Oja, E. (2001) *Independent Component Analysis*, John Wiley & Sons.

ICA, motivating example

"Cocktail party problem":

- hear several simultaneous conversations
- would like to "separate them"

Model for "conversations": time series:

$$s_1(t) \quad \text{and} \quad s_2(t)$$

Toy Example

ICA, motivating example (cont.)

Mixed version of signals:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t)$$

And also a second mixture (e.g. from a different location):

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t)$$

Mixed version of above toy example

ICA, motivating example (cont.)

Goal: Recover "signal" $\underset{\rightarrow}{s}(t) = \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix}$ from "data" $\underset{\rightarrow}{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$

for *unknown* "mixture matrix" $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, where

$$\underset{\rightarrow}{x} = A \underset{\rightarrow}{s}, \quad \text{for all } t$$

Goal is to find "separating weights", W, so that

$$\underset{\rightarrow}{s} = W \underset{\rightarrow}{x}, \quad \text{for all } t$$

Problem: $W = A^{-1}$ would be fine, but A is *unknown*

ICA, motivating example (cont.)

"Solutions" for Cocktail Party example:

Approach 1: PCA (on "population of 2-d vectors")

"Direction of Greatest Variability" doesn't solve this problem

Approach 2: ICA (will describe method later)

"Independent Component" directions do solve the problem

(modulo "sign changes" and "identification")

ICA, motivating example (cont.)

Relation to FDA: recall "data matrix"

$$X = \left(\underset{\rightarrow}{X}_1 \quad \cdots \quad \underset{\rightarrow}{X}_n \right) = \begin{pmatrix} X_{11} & & X_{1n} \\ \vdots & \cdots & \vdots \\ X_{d1} & & X_{dn} \end{pmatrix}$$

Signal Processing: focus on rows (d time series, for $t = 1,...,n$)

Functional Data Analysis: focus on columns (n data vectors)

Note: same 2 different viewpoints as "dual problems" in PCA

ICA, motivating example (cont.)

Scatterplot View (signal processing): plot

- signals & scatterplot $\{(s_1(t), s_2(t)) : t = 1,\ldots,n\}$
- data & scatterplot $\{(x_1(t), x_2(t)) : t = 1,\ldots,n\}$
- scatterplots give hint how it is possible
- affine trans. $\underset{\rightarrow}{x} = A\underset{\rightarrow}{s}$ “stretches indep. signals into dep.”
- “inversion” is key to ICA (even when A is unknown)

ICA, motivating example (cont.)

Why not PCA?

- finds “direction of greatest variability” [PCA - scatterplot]

- which is wrong direction for “signal separation”

  [PCA decomposition]

ICA, Algorithm

ICA Step 1:

- “sphere the data” (shown on right in scatterplot view)
- i.e. find linear transf’n to make mean = $\underset{\rightarrow}{0}$, cov = I
- i.e. work with $Z = \hat{\Sigma}^{-1/2}(X - \hat{\underset{\sim}{\mu}})$
- requires X of full rank (at least $n \geq d$, i.e. no HDLSS)
  (is this critical????)
- search for “indep.” *beyond* linear and quadratic structure

ICA, Algorithm (cont.)

ICA Step 2:

- Find dir'ns that make (sph'd) data as "indep. as possible"

- Worst case: Gaussian – sph'd data are independent

Interesting "converse application" of C.L.T.:

- For S_1 and S_2 independent (& non-Gaussian)

- $X_1 = uS_1 + (1-u)S_2$ is "more Gaussian" for $u \approx \frac{1}{2}$

- so independence comes from "least Gaussian directions"

ICA, Algorithm (cont.)

Criteria for non-Gaussianity / independence:

- kurtosis ($EX^4 - 3\left(EX^2\right)^2$, 4th order cumulant)
- negative entropy
- mutual information
- nonparametric maximum likelihood
- "infomax" in neural networks
- $\exists$ interesting connections between these