Statistics 321: Functional Data Analysis

Spring Semester 2001

Class meetings: Tuesday – Thursday 9:30 – 10:45
Room 07 Gardner Hall

Professor: J. S. Marron

Email: marron@email.unc.edu
Office: 304 New West
Phones: (Office) 962-2188 (Home) 493-2844
Personal Web Page: http://www.stat.unc.edu/faculty/marron.html

Class Web Page

http://www.stat.unc.edu/faculty/marron/321FDAhome.html

Show 321FDAhome.html

- Archive of previous material
- Will try to post upcoming material
- Could try printing and annotating
- Will announce postings on….

Class email Listserv

Name: Stat321FDA

Web Address:
http://listserv.unc.edu/cgi-bin/lyris.pl?enter=Stat321FDA

- a list of email addresses for broadcast of messages
- useful for class announcements (e.g. class cancelled),
- easily accessed from class web page
- please add yourself

Administrative Matters

Textbook: Ramsay, J. O. & Silverman, B. W. (1997) *Functional Data Analysis*, Springer, N.Y.

Grading: (no HW, no exams) 2 presentations, on any of:

a. a section from Ramsay and Silverman

b. a paper on functional data analysis

c. your work on functional data analysis

(goal: I speak one day a week, somebody else the other)

Functional Data Analysis, I

A personal view: what is the “atom” of the statistical analysis?

1st course in statistics: “atoms” are numbers

Statistical multivariate analysis: “atoms” are vectors

Functional Data: “atoms” are more complex objects

Functional Data Analysis, II

FDA: “atoms” are more complex objects, e. g.

- curves (focus of Ramsay and Silverman)

Show CurvDat\ParabsRaw.ps

- images, e.g. Cornea data (Cohen, Tripoli)

Show CorneaRobust\NORMLWR.MPG

- shapes, e.g. Corpus Callosum Data (Ho, Gerig)

Show CorpColl\CCFrawAlls3.mpg

- 3-d shapes, e.g. Vertebra (Gregg Tracton)

Show 902.1.rped-to-spinous.gif, from http://www.radonc.unc.edu/~tracton

Functional Data Analysis, III

Viewpoints: "analyzing" populations of complex objects

2 common major goals:

I. Understanding "population structure".
   - "visualization"
   - "intuition"

II. Statistical Classification, i.e. Discrimination
   - put into "known groups", based on "training data"
   - e.g. disease diagnosis

Visualization of Population Structure

More effort needed (by statisticians and scientists in general)

Central Problem:

- Human perceptual system excellent in 1, 2 and 3 dimensions
- But very weak beyond that (time, temperature)
- We “can’t visualize” higher dim’al structures
- Thus challenging to build “intuitive ideas”

Common Structure

Represent “data objects” by “feature vectors”
(vectors that summarize “shape”,
using terminology from statistical
pattern recognition)

Important duality:

Object Space $\leftrightarrow$ Feature Space

The one to one mapping between these allows conceptualization
in Object Space, with statistical analysis in Feature Space

E.g. 1: Curves

Show CurvDat\ParabsRaw.ps

Data Objects: $f_1(x),..., f_n(x)$ (conceptual model)

Digital version: $\begin{pmatrix} f_1(x_1) \\ \vdots \\ f_1(x_d) \end{pmatrix},...,\begin{pmatrix} f_n(x_1) \\ \vdots \\ f_n(x_d) \end{pmatrix}$, for a "grid" $x_1,..., x_d$

Object Space View: Overlay plots of curves (see "population")

Feature space: $\left\{\begin{pmatrix} f_i(x_1) \\ \vdots \\ f_i(x_d) \end{pmatrix} : i = 1,..., n\right\}$, e.g. dimension $d = 10$

Aside

In statistical visualization community, this is called the “parallel coordinates” or “parallel axes” plots

It was proposed by Inselberg, A. (1985) *The Visual Computer*, 1, 69-91. as a method to visualize high dim’al data (in general)

In particular: visualize each data vector $\begin{pmatrix} x_{i1} \\ \vdots \\ x_{id} \end{pmatrix}$,

by “connecting the dots” through $(1, x_{i1}), (2, x_{i2}), \ldots, (d, x_{id})$

E.g. 2: Images, Corneas

Show CorneaRobust\NORMLWR.MPG

Data Objects: map of “temperature scale radial curvature”

Feature vectors: Digitized version is “large and wasteful”

Instead use coefficients of Zernike Basis repres’n, $d = 66$

Object Space view: can’t overlay images

Instead show images sequentially
(hard to see “population structure”)

E.g. 3: shapes, Corpora Callosa

Show CorpColl\CCFrawAlls3.mpg

Data Objects: boundaries of “segmented” corpora callosa

Feature vectors: use coefficients of Fourier boundary representation, $d = 80$

Object Space view: can either overlay, or show sequentially

In either case: hard to see “population structure”

Finding and visualizing structure in populations

Powerful method: Principal Component Analysis

Our approach:

1. Focus on visualization, with only heuristic mathematics

2. Later carefully revisit mathematics & computation
   (easier with intuition and motivation in mind)