Independent Component Analysis

From a Functional Data Analysis Viewpoint

J. S. Marron

Department of Statistics

University of North Carolina

Functional Data Analysis, Background

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

- curves
- images
- 3-d shapes

Functional Data Analysis, Background (cont.)

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

Independent Component Analysis

Idea: Find “directions that maximize independence”

Motivating Context: Signal Processing

In particular: “Blind Source Separation”

References:

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

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

ICA, motivating example

"Cocktail party problem":

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

Model for "conversations": time series:

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

show ICAeg1p1d1Ori.ps

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):

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

Show ICAeg1p1d1Mix.ps

ICA, motivating example (cont.)

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

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:

Show ICAeg1p1d1PCAdecomp.ps

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

Approach 2: ICA:

Show ICAeg1p1d1ICAdecomp.ps

"Independent Component" directions do

ICA, motivating example (cont.)

Relation to FDA: recall "data matrix"

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

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

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

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

ICA, motivating example (cont.)

Scatterplot View: plot

- signals $\{(s_1(t), s_2(t)) : t = 1, \ldots, n\}$

Show ICAeg1p1d1Ori.ps and ICAeg1p1d1OriSP.ps

- data $\{(x_1(t), x_2(t)) : t = 1, \ldots, n\}$

Show ICAeg1p1d1Mix.ps and ICAeg1p1d1MixSP.ps

- 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"

show ICAeg1p1d1MixPCA.ps

- which is wrong direction for "signal separation"

show ICAeg1p1d1PCAdecomp.ps

ICA, Algorithm

ICA Step 1:

- "sphere the data"
- i.e. find linear transf'n to make mean = $\underset{\rightarrow}{0}$, cov = I
- i.e. work with $Z = \hat{\Sigma}^{-1/2}\left(X - \hat{\mu}\right)$
- 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

again show ICAeg1p1d1OriSP.ps and ICAeg1p1d1MixSP.ps

ICA, Algorithm (cont.)

ICA Step 2:

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

- Worst case: Gaussian – sph'd data is 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 ($EX^4 - 3\left(EZ^2\right)^2 = EX^4 - 3$, 4th order cumulant)
- negative entropy
- mutual information
- nonparametric maximum likelihood
- “infomax” in neural networks
- $\exists$ interesting connections between these

ICA, Algorithm (cont.)

Matlab Algorithm (optimizing any of above):   “FastICA”

- numerical gradient search method
- can find directions “iteratively”
- or by “simultaneous optimization”
- appears fast, with good defaults

show ICAeg1p1d1ICAdecomp.ps and again show ICAeg1p1d1MixICA.ps

ICA, Toy Examples

More Toy examples:

1. 2 sine waves, original and “mixed”

show ICAeg1p1d2Ori.ps and ICAeg1p1d2Mix.ps (everything on this page is combined in ICAeg1p1d2Combine.pdf)

- Scatterplots show “time series structure”(not “random”)

  show ICAeg1p1d2OriSP.ps and ICAeg1p1d2MixSP.ps

- PCA finds wrong direction

  show ICAeg1p1d2MixPCA.ps and ICAeg1p1d2PCAdecomp.ps

- Sphering is enough to solve this (“orthogonal to PCA”)

  Again show ICAeg1p1d2MixSP.ps

- So ICA is good (note: “flip”, and “constant signal power”)

  show ICAeg1p1d2MixICA.ps and ICAeg1p1d2ICAdecomp.ps

ICA, Toy Examples (cont.)

2. Sine wave and noise

Show ICAeg1p1d4Ori.ps, ICAeg1p1d4OriSP.ps, ICAeg1p1d4Mix.ps and ICAeg1p1d4MixSP.ps
(everything on this page is combined in ICAeg1p1d4Combine.pdf)

- PCA finds “diagonal of parallelogram”

Show ICAeg1p1d4MixPCA.ps and ICAeg1p1d4PCAdecomp.ps

- Sine is all in one, but still “wiggles” (noise still present)

- ICA gets it right (but note noise magnified)

Show ICAeg1p1d4MixICA.ps and ICAeg1p1d4PCAdecomp.ps

ICA, Toy Examples (cont.)

3. Balanced Sine wave and noise

Show ICAeg1p1d7Combine.pdf

- Note PCA gives “even split of sine wave”

- ICA gives excellent denoising

4. 2 noise components

Show ICAeg1p1d5Combine.pdf)

- PCA finds “axis of ellipse” (happens to be “right”)

- Note even “realization” of noise is right

- ICA is “wrong” (different noise realization)

ICA, Toy Examples (cont.)

5. Long parallel points clouds

Show ICAeg1p1d6Ori.ps, ICAeg1p1d6OriSP.ps, ICAeg1p1d6Mix.ps and ICAeg1p1d6MixSP.ps

- PCA finds PC1: “noise” PC2: “signal”

Show ICAeg1p1d6MixPCA.ps and ICAeg1p1d6PCAdecomp.ps

- ICA finds signal in IC1 (most non-Gaussian), noise in IC2

Show ICAeg1p1d6MixICA.ps and ICAeg1p1d6PCAdecomp.ps

ICA, Toy Examples (cont.)

6. 2-d discrimination

show HDLSS\HDLSSod1Raw.ps

- Seek “direction” that separates red and blue projections

- PCA is poor (neither PC1, nor PC2 works)

Show HDLSS\HDLSSod1PCA.ps

- ICA is excellent (since “bimodal” = “most non-Gaussian”)

Show HDLSS\HDLSSod1ICA.ps

- <u>No class information</u> used by ICA!

- Thus “useful preprocessing” for discrimination????

- Which is “right”, spherical or original scales????

ICA, Toy Examples (cont.)

7. split X Discrimination:

Show HDLSS\HDLSSxd3Raw.ps

- PCA leaves lots of overlap

Show HDLSS\HDLSSxd3ICA.ps

- ICA gives excellent separation

Show HDLSS\HDLSSxd3ICA.ps

- IC1 has “more kurtosis”, but IC2 is best for discrimination

- Useful preprocessing for e.g. CART

ICA, CurvDat Examples

PCA for “Parabs”

Show CurvDat\ParabsCurvDat.ps

- Mean captured “parabola” shape
- PC1 is “vertical shift”
- PC2 is “tilt” (hard to see visually)
- Remaining PCs are “Gaussian noise”

ICA, CurvDat Examples (cont.)

Corresponding ICA for “Parabs”

Show ParabsCurvDatICA.ps

- mean and centered data as before
- sphered data has “no structure” (i.e. this structure is “all in covariance”, i.e. have Gaussian point cloud)
- sphered ICs choose “random non-Gaussian” directions
- sphered ICs seem to find outliers
- Original scale versions capture some “vertical shift”
- Non-orthogonality on original scale $\Rightarrow$ hard to interpret

ICA, CurvDat Examples (cont.)

PCA for “Parabs with 2 outliers”

Show CurvDat\Parabs2outCurvDat.ps

- Mean captured “parabola” shape
- PC1 is “vertical shift affected by hi-freq outlier”
- PC2 is “most of high freq.outlier”
- “low freq outlier” and “tilt” are mixed between PC3 & PC4
- hope ICA can “separate these”???

ICA, CurvDat Examples (cont.)

Corresponding ICA for “Parabs with 2 outliers”

Show Parabs2outCurvDatICA.ps

- ICA finds both outliers well (non-Gaussian direction)
- ICA still misses “shift” and “tilt”
- Since these are elliptical point cloud properties, that are ignored through sphering.
- $\exists$ analysis which keeps “both kinds of features”????

ICA, CurvDat Examples (cont.)

Recall PCA for “Parabs Up and Down” (2 clusters)

Show CurvDat\ParabsUpDnCurvDat.ps

- PC1 finds clusters
- Others find usual structure (vertical shift and tilt)

Corresponding ICA for “Parabs Up and Down”

Show ParabsUpDnCurvDatICA.ps

- Clusters not found???? (seems very “non-Gaussian”)
- sphering killed clusters????
- Problem with numerical search algorithm????

ICA, CurvDat Examples (cont.)

Attempted fix 1: Change of "nonlinear function"

Show CurvDat\ParabsUpDnCurvDatICAt5.ps

- similar results
- same happened for other choices

Attempted fix 2: use PCA directions as "starting value"

Show CurvDat\ParabsUpDnCurvDatICAt2.ps

- Gives good solution
- Is this a general problem????
- How generalizable is this solution????