Stable Ratios and Robust PCA and ICA for heavy tailed distributions

Tuncay Alparslan, Richard Davis, John Nolan,
Sid Resnick, Jingjing Zou

DOE, American U, Columbia U, Cornell U

MURI Meeting, CFEM, NYC
12 May 2017

Outline

1. Ratios of stable random variables
2. Traditional PCA
3. Robust PCA
4. Robust ICA

Moving averages of heavy tailed innovations

Let Z_i, $i = 1, 2, 3, ...$ be i.i.d. heavy tailed and regular varying:

$$\begin{aligned} P(|Z_j| > x) &= x^{-\alpha} L(x) \\ \frac{P(Z_j > x)}{P(|Z_j| > x)} &\to p \\ \frac{P(Z_j < -x)}{P(|Z_j| > x)} &\to q = 1 - p \end{aligned}$$

where $L(\cdot)$ is slowly varying and $p \in [0, 1]$.

Moving averages of heavy tailed innovations

Let Z_i, $i = 1, 2, 3, \ldots$ be i.i.d. heavy tailed and regular varying:

$$\begin{aligned} P(|Z_j| > x) &= x^{-\alpha} L(x) \\ \frac{P(Z_j > x)}{P(|Z_j| > x)} &\to p \\ \frac{P(Z_j < -x)}{P(|Z_j| > x)} &\to q = 1 - p \end{aligned}$$

where $L(\cdot)$ is slowly varying and $p \in [0, 1]$.
Consider the moving average

$$X_t = \sum_{j=-\infty}^{\infty} c_j Z_j.$$

Cline (1983) showed this exists if $\sum_{j=-\infty}^{\infty} |c_j|^\delta < 0$ for some $0 < \delta < \alpha$, $\delta \leq 1$.

Davis and Resnick (1986)

For simplicity, assume $1 < \alpha < 2$ and $EZ_j = 0$. Correlations of lag h

$$\rho(h) = \frac{\sum_j c_j c_{j+h}}{\sum_j c_j^2}$$
$$\widehat{\rho}(h) = \frac{\sum_j X_j X_{j+h}}{\sum_j X_j^2}$$

Then

$$\frac{a_n^2}{\widetilde{a}_n} (\widehat{\rho}(1), \ldots, \widehat{\rho}(d)) \xrightarrow{d} \left(\frac{S_1}{S_0}, \ldots, \frac{S_d}{S_0} \right) \quad n \to \infty,$$

where S_j's are independent, S_0 positive $\alpha/2$ stable and $S_1, \ldots, S_d$) are i.i.d. α stable.

Testing for independence

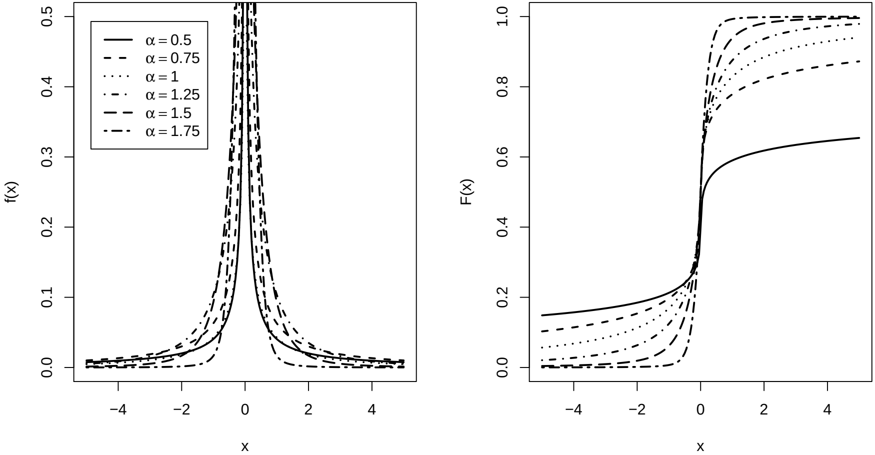
To test $\rho(1) = 0$ with correct significance level, need to compute quantiles of $R = S_1/S_0$, the ratio of two independent stable terms. The cdf and pdf of R are given by

$$F_R(x) = P(R \leq x) = \int_{-\infty}^{\infty} F_1(tx) f_0(t) dt,$$

$$f_R(x) = F_R'(x) = \int_{-\infty}^{\infty} t\, f_1(tx) f_0(t) dt,$$

where F_i is the cdf of S_i and f_i is the pdf of S_i.

These can be evaluated numerically using existing algorithms for $F_1(\cdot)$ and $f_i(*\cdot)$.



Plots of the density $f_R(x)$ on the left and distribution function $F_R(x)$ on the right of the ratio $R = S_1/S_0$, where $S_1 \sim S(\alpha, 0, 1, 0)$ and $S_0 \sim S(\alpha/2, 1, 1, 0)$.

Quantiles $F_R^{-1}(p)$ of $R = S_1/S_0$ when $S_1 \sim S(\alpha, 0, 1, 0)$ and $S_0 \sim S(\alpha/2, 1, 1, 0)$.

	p			
α	0.9	0.95	0.975	0.99
1.0	2.491	6.187	13.987	43.810
1.1	1.867	4.111	8.181	19.968
1.2	1.452	2.914	5.326	11.541
1.3	1.142	2.116	3.595	7.069
1.4	0.901	1.557	2.483	4.541
1.5	0.704	1.143	1.723	2.921
1.6	0.537	0.825	1.178	1.866
1.7	0.391	0.570	0.774	1.141
1.8	0.258	0.359	0.465	0.634
1.9	0.131	0.175	0.217	0.274

To test simultaneously that d correlations are 0, need to know the distribution of

$$M = \max(S_1/S_0, S_0/S_0, \ldots, S_d/S_0),$$

where $S_0, S_1, \ldots, S_d$ are independent stable and $S_1, S_0, \ldots, S_d$ i.i.d. Using conditional independence given S_0, M has cdf and pdf

$$\begin{aligned} F_M(t) &= P(M \leq t) = P(S_j/S_0 \leq t, \forall j) = \int_0^\infty P(S_j \leq ts, \forall j) f_0(s) ds \\ &= \int_0^\infty [F_1(ts)]^d f_0(s) ds, \qquad (1) \\ f_M(t) &= F_M'(t) = d \int_0^\infty [F_1(ts)]^{d-1} f_1(ts) s f_0(s) ds. \end{aligned}$$

To test simultaneously that d correlations are 0, need to know the distribution of

$$M = \max(S_1/S_0, S_0/S_0, \ldots, S_d/S_0),$$

where $S_0, S_1, \ldots, S_d$ are independent stable and $S_1, S_0, \ldots, S_d$ i.i.d. Using conditional independence given S_0, M has cdf and pdf

$$\begin{aligned} F_M(t) &= P(M \le t) = P(S_j/S_0 \le t, \forall j) = \int_0^\infty P(S_j \le ts, \forall j) f_0(s) ds \\ &= \int_0^\infty [F_1(ts)]^d f_0(s) ds, \qquad (1) \\ f_M(t) &= F_M'(t) = d \int_0^\infty [F_1(ts)]^{d-1} f_1(ts) s f_0(s) ds. \end{aligned}$$

Also have code to numerically compute the cdf and pdf of the product $P = X_1 X_2$ of independent stable terms.

Robust PCA and ICA

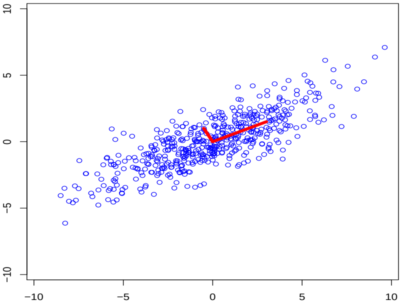
Principal components analysis (PCA) is a popular technique for analyzing data, tries to extract the directions with maximum dispersion. Traditional PCA can behave poorly when the data is heavy tailed; we propose a robust PCA.

When there is no elliptical structure, we propose using an modified version of Independent Component Analysis (ICA).

Traditional PCA

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a d-dim. sample
Compute the sample covariance matrix S.

Perform an eigenvalue decomposition of S: eigenvalue λ_j with associated eigenvector $\mathbf{v}_j$, $j = 1, \ldots, d$. Assume eigenvalues are ranked: $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_d \geq 0$. Then $\mathbf{v}_1$ is the first principal component, $\mathbf{v}_2$ is the second, etc.



One application - dimension reduction

In regression, principal component regression is used to reduce the number of variables. In engineering and computer science, high dimensional data use PCA to find a small number of directions that explain most of the variability.

Example: image processing. V. Hlavac at Czech Technical University in Prague applied PCA to images with $321 \times 261 = 83781$ pixels. Stack the columns of a picture to get a 83781 dimensional vector. Started with 32 photos of a boy with different facial expressions. So data matrix has $n = 32$ samples in a $d = 83781$ dimensional space.

He found the first four (!?) principal components and then 'reconstructed' each of the 83781 dim. photos as a linear combination: image $= q_1\mathbf{b}_1 + q_2\mathbf{b}_2 + q_3\mathbf{b}_3 + q_4\mathbf{b}_4$, where $\mathbf{b}_1, \ldots, \mathbf{b}_4$ are the eigenvectors associated with the 4 largest eigenvalues.

What if we have 32 instances of images?

Approximation by 4 principal components only

- Reconstruction of the image from four basis vectors $\mathbf{b}_i$, $i = 1, \ldots, 4$ which can be displayed as images.
- The linear combination was computed as $q_1\mathbf{b}_1 + q_2\mathbf{b}_2 + q_3\mathbf{b}_3 + q_4\mathbf{b}_4 = 0.078\,\mathbf{b}_1 + 0.062\,\mathbf{b}_2 - 0.182\,\mathbf{b}_3 + 0.179\,\mathbf{b}_4$.

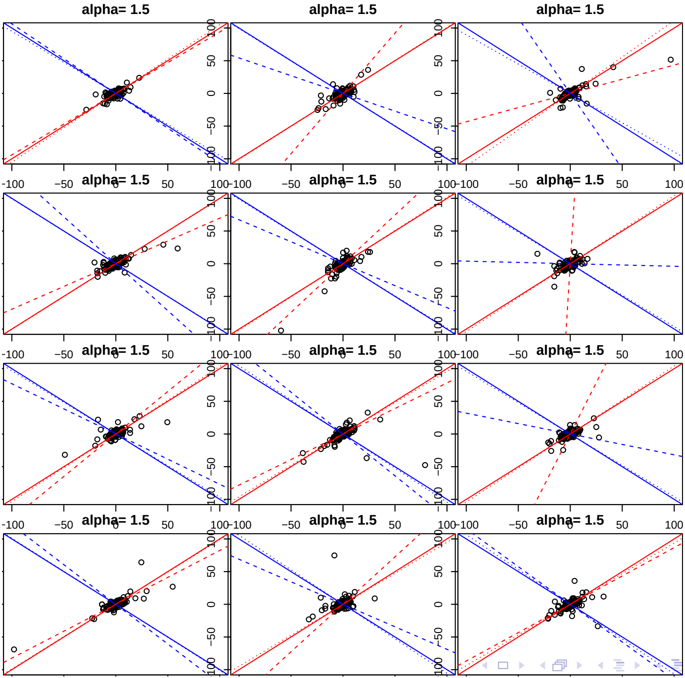
Note: most change is around the eyes and mouth.

Reconstruction fidelity, 4 components

What happens when data is heavy tailed?

Computing the sample covariance requires calculating the means of all the components and second moments. The mean is heavily influenced by outliers, and second moments are even more heavily influenced by outliers.

Next slide shows 2-dimensional simulations with elliptical stable data and $\alpha = 1.5$. Solid lines show exact "principal components", dashed lines show estimates from traditional PCA.


alpha= 1.5
alpha= 1.5
alpha= 1.5
alpha= 1.5
alpha= 1.5
alpha= 1.5
alpha= 1.5
alpha= 1.5
alpha= 1.5
alpha= 1.5
alpha= 1.5
alpha= 1.5
-100
−50
0
50
100
-100
−50
0
50
100

Robust PCA in the elliptical stable case

When **X** is elliptical stable, there is a corresponding shape matrix R, a $d \times d$ positive definite matrix that determines the shape. Two representations:

$$\mathbf{X} = R^{-1/2}\mathbf{Z} + \delta,$$

where **Z** is isotropic/radially symmetric α-stable.

$$\mathbf{X} = A^{1/2}\mathbf{G} + \delta,$$

where $A > 0$ is a positive $(\alpha/2)$-stable univariate stable r.v. and $\mathbf{G} \sim$N$(0, R)$.

Robust PCA: estimate center δ and shape matrix R by a method that takes into account the heavy tails in the data. Then do an eigen-decomposition on R (instead of sample covariance) to get principal values from the

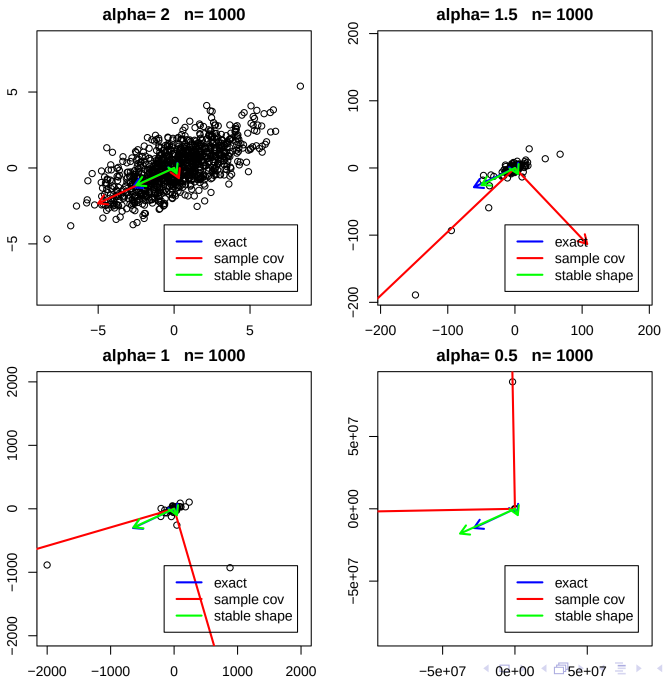
Projection approach

If $\mathbf{X}$ is multivariate α-stable, then for any vector (direction) $\mathbf{u}$, the inner product $\langle \mathbf{u}, \mathbf{X} \rangle$ is univariate stable, with parameters $\mathbf{S}(\alpha, \beta()\mathbf{u}), \gamma(\mathbf{u}), \delta(\mathbf{u})$.

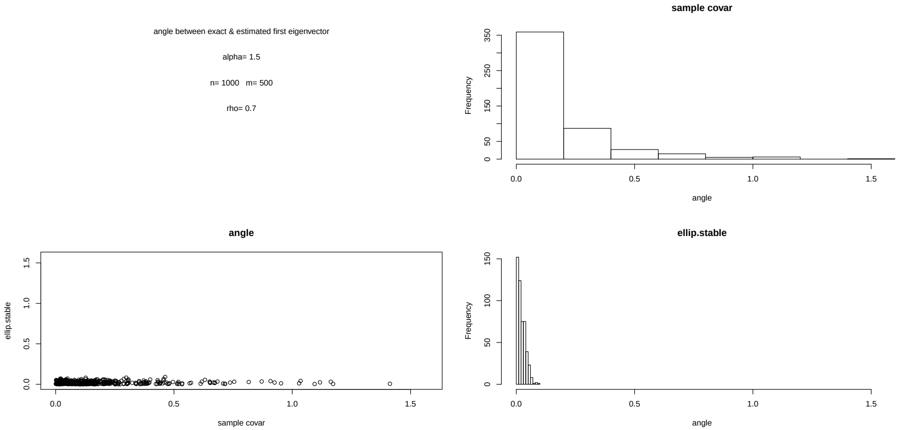
For PCA, we will assume elliptical symmetry, in which case, $\langle \mathbf{u}, \mathbf{X} \rangle \sim \mathbf{S}(\alpha, 0, \gamma(\mathbf{u}), 0))$, where $\gamma(\mathbf{u}) = (\mathbf{u}^T R \mathbf{u})^{1/2}$ completely determines the joint distribution.

Given a data set, estimate α and pick a sequence of directions $\mathbf{u}_1, \ldots, \mathbf{u}_m$ and estimate scale function $\widehat{\gamma}(\mathbf{u}_j)$, $j = 1, \ldots, m$. (Can use any univariate estimation method: quantile, fractional moments, characteristic function method, maximum likelihood, etc.) Use these estimated scale functions to estimate shape matrix R. (Extra step: guarantee that R is non-negative definite.)

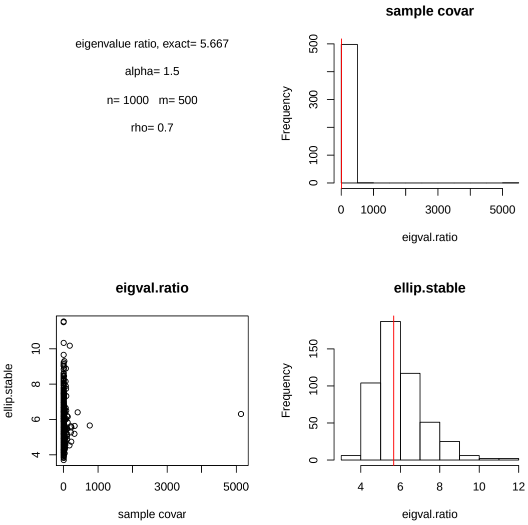
$d = 2$ dimensional examples with varying α.


alpha= 2 n= 1000
alpha= 1.5 n= 1000
alpha= 1 n= 1000
alpha= 0.5 n= 1000
exact
sample cov
stable shape
−5
0
5
−200
−100
100
200
−2000
−1000
1000
2000
−5e+07
0e+00
5e+07

Estimating direction of first eigenvector



Estimating eigenvalues



Higher dimensions

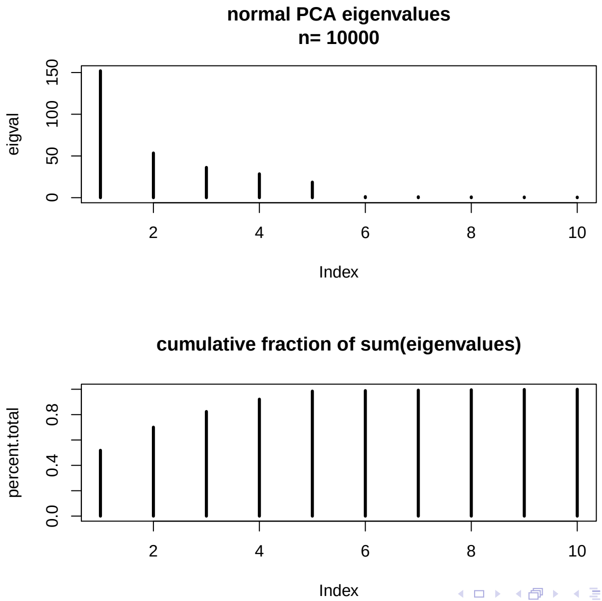
Method works for dimension d up to 100.

To assess how well the methods works in moderate dimensions, we consider a model with $d = 10$ dimensional elliptical stable model with $\alpha = 1.5$ and shape matrix R a diagonal matrix with diagonal (3,3,3,3,3,0.1,0.1,0.1,0.1,0.1). This is essentially a 5 dimensional model embedded in 10 dimensional space. We examine two issues:

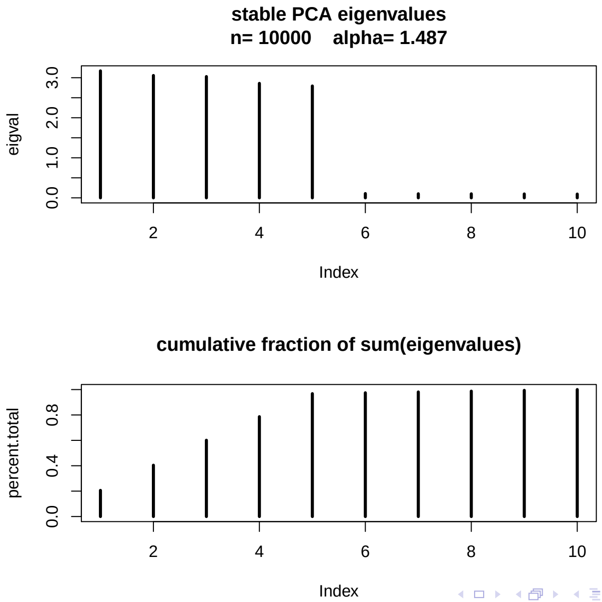
- how well normal PCA estimates the eigenvalues vs how well stable PCA estimates them
- how well the two methods estimate the span of the eigenvalues

For the first question, we simulated a large data set ($n = 10000$) and estimated the principal components using both methods. Following figures shows that normal PCA gives extremely large estimates of the first 5 eigenvalues, and they vary noticeably. In contrast, stable PCA (second page) gives a correct estimate of the first 5 eigenvalues and a clear, abrupt change for the remaining 5 eigenvalues.

Eigenvalues for normal PCA

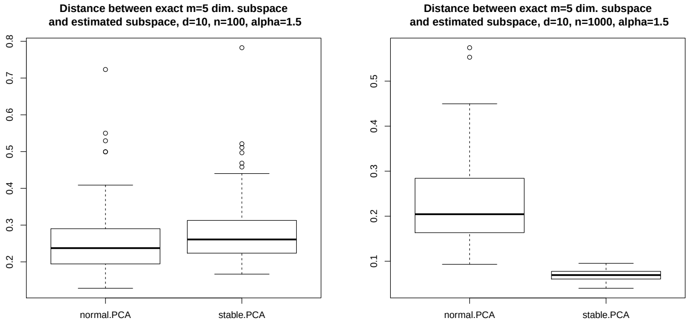


Eigenvalues for stable PCA



Span of eigenvalues

Crone and Crosby (1995) define a distance between subspaces, which we use to measure the distance between exact and estimated span of $k = 5$ principal components for same simulation ($\alpha = 1.5$, $d = 10$, most of dispersion in first 5 components) for sample sizes $n = 100$ and $= 1000$:



Non-elliptical stable case - Independent Component Analysis (ICA)

Here it can be meaningless to use PCA, even the robust PCA described above: there are very non-elliptical dependence structures. One interesting case is when there are independent components:

$$\mathbf{X} = A\mathbf{Z},$$

where A is a $d \times m$ matrix of coefficients and $\mathbf{Z} = (Z_1, \ldots, Z_m)$ are i.i.d. stable. (Equivalently, the spectral measure of $\mathbf{X}$ is discrete.)

Want an robust ICA.

Non-elliptical stable case - Independent Component Analysis (ICA)

Here it can be meaningless to use PCA, even the robust PCA described above: there are very non-elliptical dependence structures. One interesting case is when there are independent components:

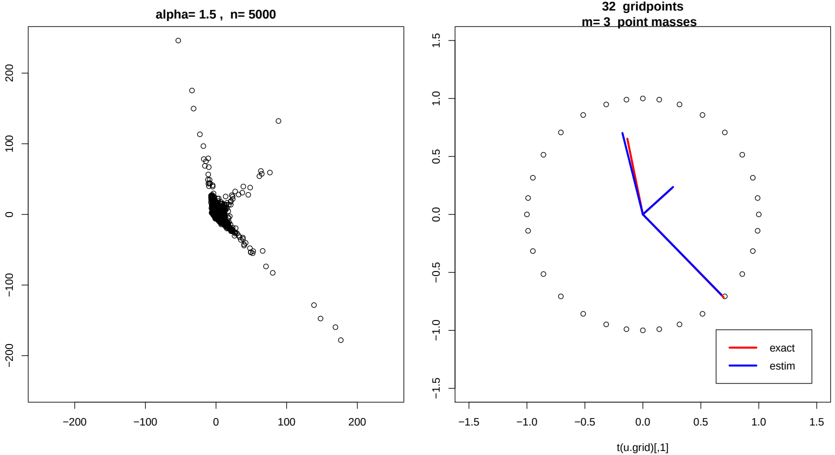
$$\mathbf{X} = A\mathbf{Z},$$

where A is a $d \times m$ matrix of coefficients and $\mathbf{Z} = (Z_1, \ldots, Z_m)$ are i.i.d. stable. (Equivalently, the spectral measure of $\mathbf{X}$ is discrete.)

Want an robust ICA.

Note that we can have $m < d$, $m = d$, or $m > d$. When $m = d$ and A is invertible, we can recover the source $\mathbf{Z} = A^{-1}\mathbf{X}$. When $m < d$, can reduce to lower dimensional problem and recover $\mathbf{Z}$. When $m > d$, it is not generally possible to recover $\mathbf{Z}$ in this way. Maybe some other way? Here we are trying to discover the multivariate structure.

Simulation with $m = 3$ components in$d = 2$ dimensions



ICA in the stable case: m known

Let $\mathbf{X}_i$, $i = 1, \ldots, n$ be a sample from a multivariate stable distribution. Use the fact that linear combinations of a multivariate stable r.v. are univariate stable.

Pick a grid $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{\mathrm{ngrid}}$. For each $j = 1, \ldots,$ngrid, calculate the univariate data set: $y_{j,i} = \langle \mathbf{u}_j, \mathbf{X}_i \rangle$, $i = 1, \ldots, n$. Let $\widehat{\gamma}_j$ and $\widehat{\beta}_j$ be the univariate scale and skewness of this projection in direction $\mathbf{u}_j$.

$$A^* := \arg\min_A \left[\sum_{j=1}^{m} (\widehat{\gamma}_j^{\alpha} - \gamma_{j,A}^{\alpha})^2 + \sum_{j=1}^{m} (\widehat{\beta}_j \widehat{\gamma}_j^{\alpha} - \beta_{j,A} \gamma_{j,A}^{\alpha})^2 \right],$$

where $\gamma_{j,A}$ and $\beta_{j,A}$ are the exact scale and skewness for the projection in direction $\mathbf{u}_j$ for the ICA model given by $\mathbf{X} = A\mathbf{Z}$.
Above objective function isn't always convex - can be numerical problems minimizing.

ICA in the stable case: *m* unknown

Stepwise approach: vary *m* and look for point where larger values of *m* don't add much to the fit.

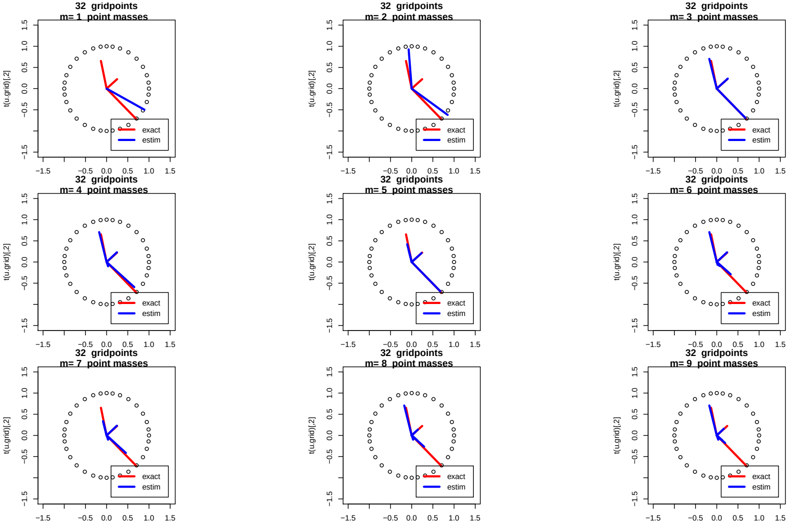
AICc seems to do a good job selecting correct *m*.

$$\mathrm{AICc} = 2m + 2\log(\mathrm{ObjFn}) + \frac{2m(m+1)}{ngrid - m - 1},$$

where ObjFn is the optimal value of the objective function on previous page.

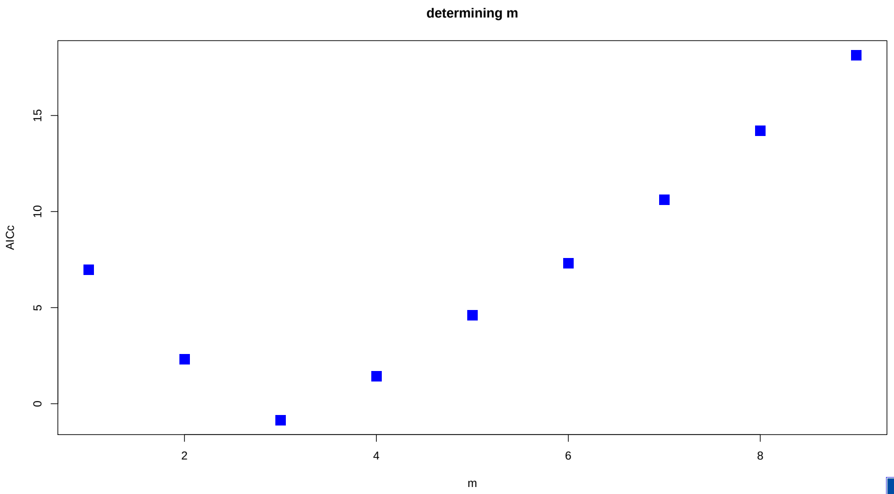
This AICc penalty uses the number of TERMS *m*, not the number of PARAMETERS= *md*. Do not know how to justify this? Richard suggested using a lasso-type penalty using lengths of new column vectors.

Varying m

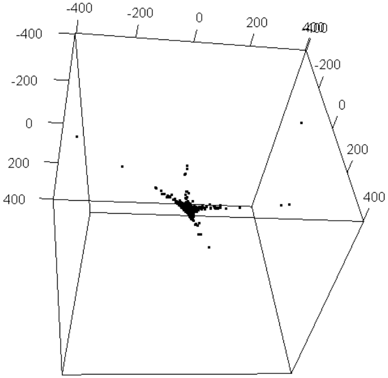


32 gridpoints
m= 1 point masses
32 gridpoints
m= 2 point masses
32 gridpoints
m= 3 point masses
32 gridpoints
m= 4 point masses
32 gridpoints
m= 5 point masses
32 gridpoints
m= 6 point masses
32 gridpoints
m= 7 point masses
32 gridpoints
m= 8 point masses
32 gridpoints
m= 9 point masses
t(u.grid)[,2]
exact
estim
−1.5 −0.5 0.0 0.5 1.0 1.5

Determining *m* using AICc

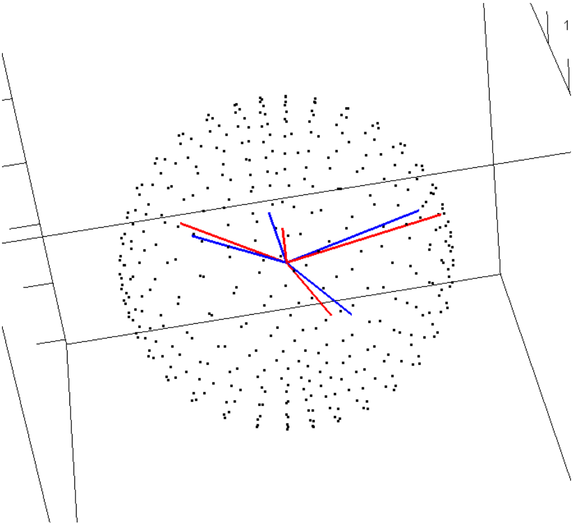


3 dim example, $m = 4$ points



alpha= 1.5 , n= 5000

Recovered vs exact point masses and locations



Red lines are exact, blue lines are estimated with stable ICA.

Future work

- **X** only in the domain of attraction, not itself stable OR elliptical, not as heavy tails, e.g. multivariate elliptical *t*-distibutions.
- ICA with different independent terms having different distributions.