

Testing Independence Between Angular and Radial Components and ICA for Heavy-Tails

Richard A. Davis

Columbia University

October 7, 2014

Part I (joint with Phyllis)

Offshoots from network models (Samorodnitsky, Resnick, Towsley, Davis, Willis, Wan)

1. Estimation of parameters in the angular distribution
 - ▶ Motivated by Krapivsky model
 - ▶ Use *likelihood* of Θ when R is large

Part I (joint with Phyllis)

Offshoots from network models (Samorodnitsky, Resnick, Towsley, Davis, Willis, Wan)

1. Estimation of parameters in the angular distribution
 - ▶ Motivated by Krapivsky model
 - ▶ Use *likelihood* of Θ when R is large
2. Choice of threshold
 - ▶ Use Hill??
 - ▶ Something else: when are Θ and R independent?

Part I (joint with Phyllis)

Offshoots from network models (Samorodnitsky, Resnick, Towsley, Davis, Willis, Wan)

1. Estimation of parameters in the angular distribution
 - ▶ Motivated by Krapivsky model
 - ▶ Use *likelihood* of Θ when R is large
2. Choice of threshold
 - ▶ Use Hill??
 - ▶ Something else: when are Θ and R independent?
3. Testing independence of Θ and R
 - ▶ Distance correlation
 - ▶ Proof of concept

Part I (joint with Phyllis)

Offshoots from network models (Samorodnitsky, Resnick, Towsley, Davis, Willis, Wan)

1. Estimation of parameters in the angular distribution
 - ▶ Motivated by Krapivsky model
 - ▶ Use *likelihood* of Θ when R is large
2. Choice of threshold
 - ▶ Use Hill??
 - ▶ Something else: when are Θ and R independent?
3. Testing independence of Θ and R
 - ▶ Distance correlation
 - ▶ Proof of concept
4. To-do
 - ▶ Limit theory for distance correlation
 - ▶ Bootstrapping distance correlation
 - ▶ Application to network models
 - ▶ Sequential à la Nguyen & Samorodnitsky (2012)

General Framework

- ▶ Set-up: X_1, X_2, \dots iid $RV(\alpha)$, $R_i = \|X_i\|$, $\Theta = X_i/\|X_i\|$.
- ▶ Recall:

$$P(R_i > x) = L(x)x^{-\alpha},$$

with $L(x)$ slowly varying, and:

$$P(\Theta \in \cdot | R > r) \rightarrow P_\eta(\Theta \in \cdot), r \rightarrow \infty.$$

Likelihood Estimation

- ▶ Example: Suppose $P_\eta(\Theta \in \cdot)$ has pdf $f(\theta|\eta)$. Estimate MLE from

$$L(\eta) = \sum_{i=1}^n \log f(\theta_i|\eta) \mathbf{1}_{[R_i > r]}.$$

General Framework

- ▶ Set-up: X_1, X_2, \dots iid $RV(\alpha)$, $R_i = \|X_i\|$, $\Theta = X_i/\|X_i\|$.
- ▶ Recall:

$$P(R_i > x) = L(x)x^{-\alpha},$$

with $L(x)$ slowly varying, and:

$$P(\Theta \in \cdot | R > r) \rightarrow P_\eta(\Theta \in \cdot), r \rightarrow \infty.$$

Likelihood Estimation

- ▶ Example: Suppose $P_\eta(\Theta \in \cdot)$ has pdf $f(\theta|\eta)$. Estimate MLE from

$$L(\eta) = \sum_{i=1}^n \log f(\theta_i|\eta) \mathbf{1}_{[R_i > r]}.$$

- ▶ Choice of Threshold: r (large?) s.t. $(\Theta_i \mathbf{1}_{[R_i > r]}, R_i \mathbf{1}_{[R_i > r]})$ are approximately independent.
 - ▶ Hill estimation
 - ▶ or something else...

Need approximate independence of Θ_i and R_i when R_i is large.

Strategy

- ▶ For each r , test the independence of $(\Theta \mathbf{1}_{[R>r]}, R \mathbf{1}_{[R>r]})$ from the data until the result becomes significant.
 - ▶ Use distance correlation for independence testing.

Distance Correlation (Székely et al., 2007)

- ▶ Define a measure of dependence

$$\mathcal{V}^2(X, Y; w) = \|\phi_{X,Y}(t, s) - \phi_X(t)\phi_Y(s)\|_w$$

- ▶ ϕ denotes the characteristic function
- ▶ $\|\cdot\|_w$ -norm is defined as

$$\|\gamma(t, s)\|_w^2 = \int_{\mathbb{R}^{p+q}} \gamma(t, s) \overline{\gamma(t, s)} w(t, s) dt ds$$

Strategy

- ▶ For each r , test the independence of $(\Theta \mathbf{1}_{[R>r]}, R \mathbf{1}_{[R>r]})$ from the data until the result becomes significant.
 - ▶ Use distance correlation for independence testing.

Distance Correlation (Székely et al., 2007)

- ▶ Define a measure of dependence

$$\mathcal{V}^2(X, Y; w) = \|\phi_{X,Y}(t, s) - \phi_X(t)\phi_Y(s)\|_w$$

- ▶ ϕ denotes the characteristic function
- ▶ $\|\cdot\|_w$ -norm is defined as

$$\|\gamma(t, s)\|_w^2 = \int_{\mathbb{R}^{p+q}} \gamma(t, s) \overline{\gamma(t, s)} w(t, s) dt ds$$

- ▶ Distance covariance:

$$w(t, s) = (c_p c_q |t|_p^{1+p} |s|_q^{1+q})^{-1}.$$

Strategy

- ▶ For each r , test the independence of $(\Theta \mathbf{1}_{[R>r]}, R \mathbf{1}_{[R>r]})$ from the data until the result becomes significant.
 - ▶ Use distance correlation for independence testing.

Distance Correlation (Székely et al., 2007)

- ▶ Define a measure of dependence

$$\mathcal{V}^2(X, Y; w) = \|\phi_{X,Y}(t, s) - \phi_X(t)\phi_Y(s)\|_w$$

- ▶ ϕ denotes the characteristic function
- ▶ $\|\cdot\|_w$ -norm is defined as

$$\|\gamma(t, s)\|_w^2 = \int_{\mathbb{R}^{p+q}} \gamma(t, s) \overline{\gamma(t, s)} w(t, s) dt ds$$

- ▶ Distance covariance:

$$w(t, s) = (c_p c_q |t|_p^{1+p} |s|_q^{1+q})^{-1}.$$

- ▶ Distance correlation:

$$\mathcal{R}^2(X, Y) = \frac{\mathcal{V}^2(X, Y)}{\sqrt{\mathcal{V}^2(X, X)\mathcal{V}^2(Y, Y)}}.$$

Empirical Distance Correlation

- Given $\{(X_k, Y_K), k = 1, \dots, n\}$, define

$$a_{kl} = |X_k - X_l|_p, \bar{a}_{k\cdot} = \frac{1}{n} \sum_{l=1}^n a_{kl}, \bar{a}_{\cdot l} = \frac{1}{n} \sum_{k=1}^n a_{kl}, \bar{a}_{..} = \frac{1}{n^2} \sum_{k,l=1}^n a_{kl}$$

$$A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{..}$$

Define B_{kl} similarly with $b_{kl} = |Y_k - Y_l|_q$'s. The empirical distance covariance $\mathcal{V}_n(X, Y)$ is given by

$$\mathcal{V}_n^2(X, Y) = \frac{1}{n^2} \sum_{k,l=1}^n A_{kl} B_{kl}$$

Empirical Distance Correlation

- Given $\{(X_k, Y_k), k = 1, \dots, n\}$, define

$$a_{kl} = |X_k - X_l|_p, \bar{a}_{k\cdot} = \frac{1}{n} \sum_{l=1}^n a_{kl}, \bar{a}_{\cdot l} = \frac{1}{n} \sum_{k=1}^n a_{kl}, \bar{a}_{..} = \frac{1}{n^2} \sum_{k,l=1}^n a_{kl}$$

$$A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{..}$$

Define B_{kl} similarly with $b_{kl} = |Y_k - Y_l|_q$'s. The empirical distance covariance $\mathcal{V}_n(X, Y)$ is given by

$$\mathcal{V}_n^2(X, Y) = \frac{1}{n^2} \sum_{k,l=1}^n A_{kl} B_{kl}$$

Empirical Distance Correlation

- If $E|X|_p + E|Y|_q < \infty$, the empirical distance correlation

$$\mathcal{V}_n(X, Y) = \|f_{X,Y}^n(t, s) - f_X^n(t)f_Y^n(s)\|^2,$$

where $f_X^n, f_Y^n, f_{X,Y}^n$ are the corresponding empirical characteristic functions.

Empirical Distance Correlation

- The empirical distance correlation satisfies

$$\mathcal{R}_n^2(X, Y) = \frac{\mathcal{V}_n^2(X, Y)}{\sqrt{\mathcal{V}_n^2(X, X)\mathcal{V}_n^2(Y, Y)}} \rightarrow_{a.s.} \mathcal{R}^2(X, Y)$$

Empirical Distance Correlation

- The empirical distance correlation satisfies

$$\mathcal{R}_n^2(X, Y) = \frac{\mathcal{V}_n^2(X, Y)}{\sqrt{\mathcal{V}_n^2(X, X)\mathcal{V}_n^2(Y, Y)}} \rightarrow_{a.s.} \mathcal{R}^2(X, Y)$$

Independence Testing

- Define

$$S_2 = \frac{1}{n^2} \sum_{k,l=1}^n |X_k - X_l|_p \frac{1}{n^2} \sum_{k,l=1}^n |Y_k - Y_l|_q$$

Then given $E|X|_p + E|Y|_q < \infty$, we have

$$n\mathcal{V}_n^2/S_2 \rightarrow_D Q$$

when X and Y are independent, where Q is a nonnegative quadratic form of centered Gaussian random variables $\sum_{j=1}^{\infty} \lambda_j Z_j$ and $E[Q] = 1$.

Empirical Distance Correlation

- The empirical distance correlation satisfies

$$\mathcal{R}_n^2(X, Y) = \frac{\mathcal{V}_n^2(X, Y)}{\sqrt{\mathcal{V}_n^2(X, X)\mathcal{V}_n^2(Y, Y)}} \rightarrow_{a.s.} \mathcal{R}^2(X, Y)$$

Independence Testing

- Define

$$S_2 = \frac{1}{n^2} \sum_{k,l=1}^n |X_k - X_l|_p \frac{1}{n^2} \sum_{k,l=1}^n |Y_k - Y_l|_q$$

Then given $E|X|_p + E|Y|_q < \infty$, we have

$$n\mathcal{V}_n^2/S_2 \rightarrow_D Q$$

when X and Y are independent, where Q is a nonnegative quadratic form of centered Gaussian random variables $\sum_{j=1}^{\infty} \lambda_j Z_j$ and $E[Q] = 1$.

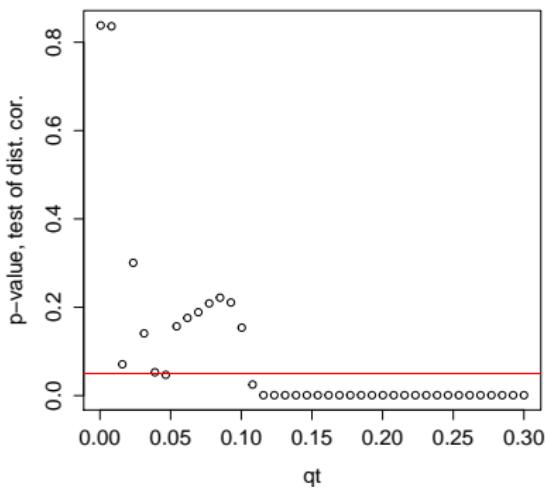
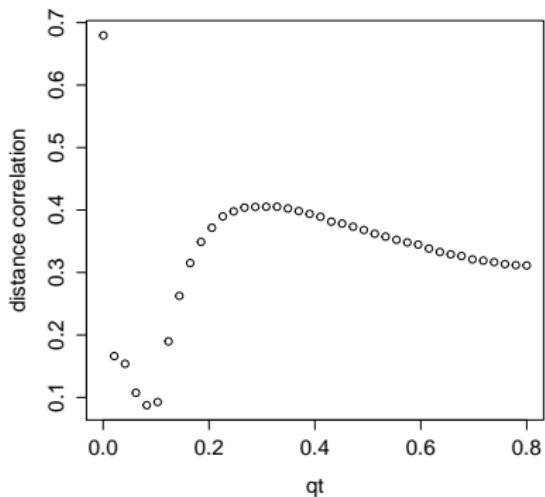
- For actual computation, we can use bootstrapping to approximate the distribution of Q .

Choice of Threshold

- ▶ For a range of values in r , compute the distance correlation of $(\Theta \mathbf{1}_{[R>r]}, \log(R) \mathbf{1}_{[R>r]})$ from the data and the p-value from the corresponding independence test. Choose the threshold r when the p-value exceeds a designated significance level.
- ▶ To demonstrate, we plot the p-value vs. threshold plot for different sets of data. Here we represent the cut-off value by the cut-off quantile, where r_q is the upper q -quantile from the data.

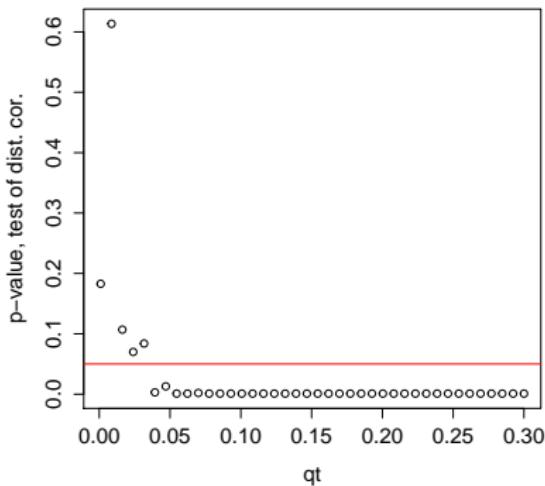
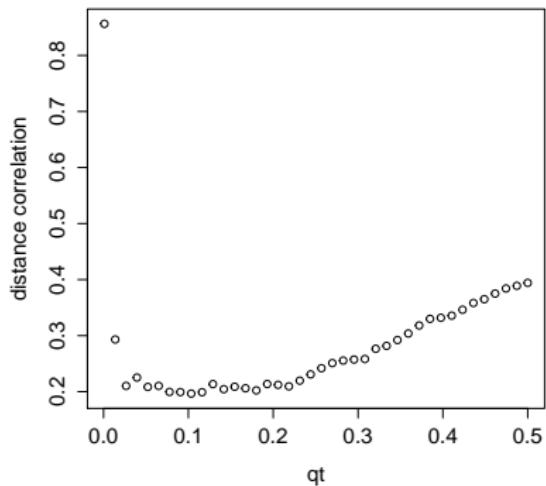
Demonstration 1

- ▶ $n = 5000$, $R \sim_{iid} t(2)$.
- ▶ $\Theta|R \sim U(0, \pi/2)$ when R is larger than the upper 10%-quantile.
- ▶ $\Theta|R = \arctan(R + A_R) + B_R \pmod{\pi/2}$, where $A_R \sim U(0, R)$,
 $B_R \sim U(0, \pi/4)$, for the rest of R .



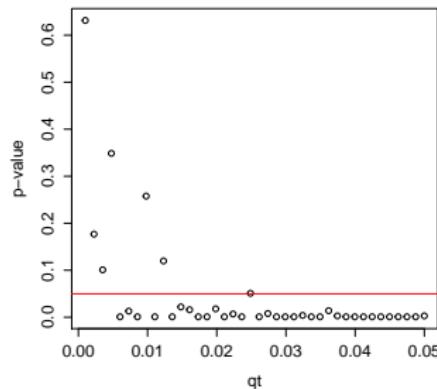
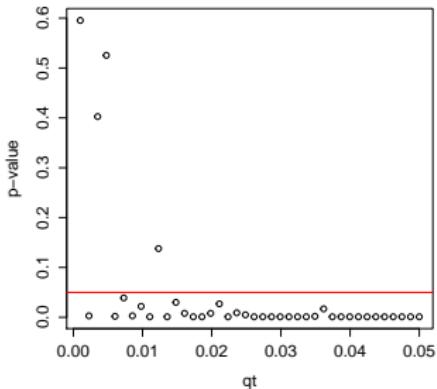
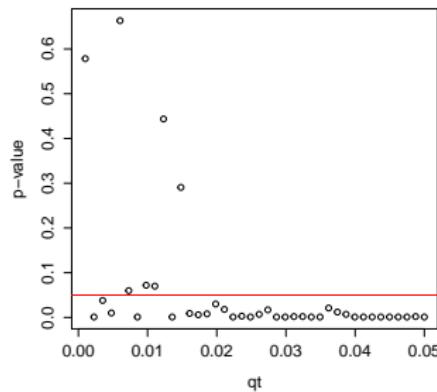
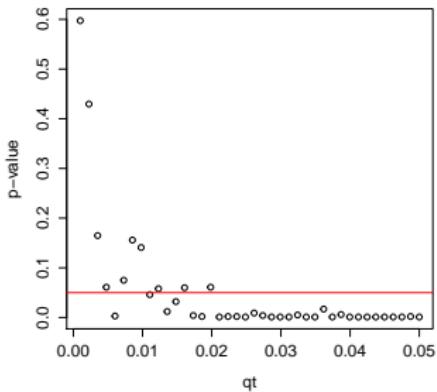
Demonstration 2

- ▶ $(X_1, X_2, X_3) \sim$ multivariate logistic distribution (Gumbel, 1960) with parameter 0.7;
- ▶ $n = 5000$



Demonstration 3

- ▶ Standardized in- and out-degree data from Krapivsky model
- ▶ $n = 10^6$
- ▶ The computational capacity for distance correlation is large for data with large sample size. Therefore, when the truncation level is low, we calculate the p-value from a random 1000 sample from the full data.



Part II (joint with Jingjing, Nolan, Resnick, Alparslan)

The Model

$$X = A \cdot S$$

- ▶ d -dimensional response $X = (x_1, \dots, x_d)^T$
- ▶ d -dimensional independent components $S = (S_1, \dots, S_d)^T$
- ▶ Full rank $d \times d$ transformation matrix A
- ▶ $S = W \cdot X$ with unmixing matrix $W = (w_1, \dots, w_d)^T = A^{-1}$
- ▶ The Independent Component Analysis (ICA) model of the distribution of X

$$P(B) = \prod_{j=1}^d P_j(w_j^T B), \quad \forall B \in \mathcal{B}_d$$

- ▶ The goal is to recover the unmixing matrix W and $S = W \cdot X$

A Strategy: Project to the Space of Log-Concave Densities

- ▶ P_d : space of d -dimensional distributions satisfying non-singularity conditions
- ▶ \mathcal{F}_d : space of d -dimensional log-concave densities
- ▶ Log-concave: exponential of piece-wise linear densities, normal, Laplace
- ▶ Not log-concave: t, stable, Pareto
- ▶ Projection $\Psi^*(P) : P_d \rightarrow \mathcal{F}_d$

$$\Psi^*(P) := \operatorname{argmax}_{f \in \mathcal{F}_d} \int_{\mathbb{R}^d} \log(f) dP$$

Projection to $\mathcal{F}_d^{\text{ICA}}$

Define $\mathcal{F}_d^{\text{ICA}}$ to be

$$\left\{ f \in \mathcal{F}_d : f(x) = |\det W| \prod_{j=1}^d f_j(w_j^T x), f_1, \dots, f_d \in \mathcal{F}_1 \right\}$$

Theorem (Samworth and Yuan (2012))

If distribution P has density $f(x) = |\det W| \prod_{j=1}^d f_j(w_j^T x)$, then

$\Psi^*(P) = \Psi^{**}(P) := \operatorname{argmax}_{f \in \mathcal{F}_d^{\text{ICA}}} \int_{\mathbb{R}^d} \log(f) dP$, and it equals to

$$f^{**}(x) = |\det W| \prod_{j=1}^d f_j^*(w_j^T x),$$

where $f_j^* = \Psi^*(f_j)$.

Estimation Procedure

- ▶ Start from an arbitrary initial value of W
- ▶ Step 1: Find log-concave projection \hat{f}_j^* of the distribution of $w_j^T X$
- ▶ Step 2: With \hat{f}_j^* , update W to maximize the log-likelihood

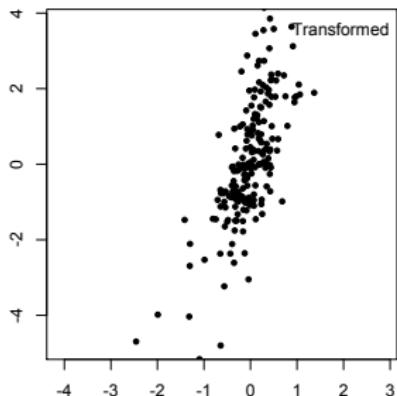
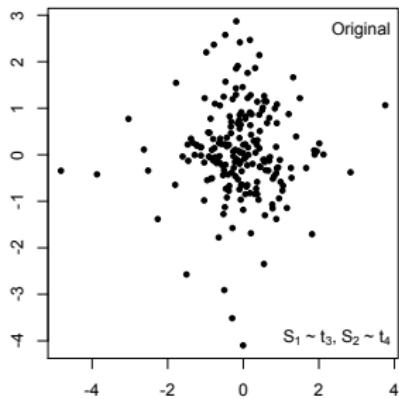
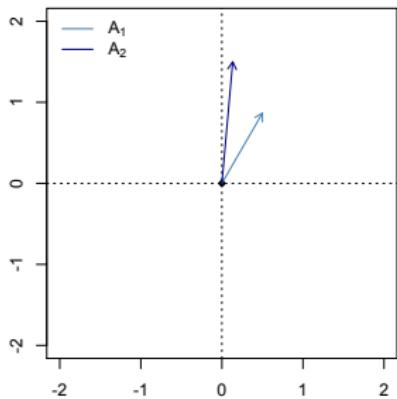
$$\log |\det W| + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^d \log \hat{f}_j^*(w_j^T x_i)$$

- ▶ Iterate steps 1 and 2, until convergence of the log-likelihood

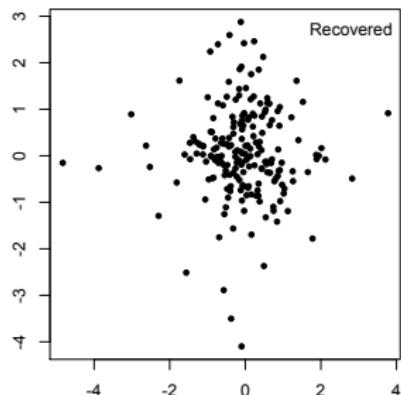
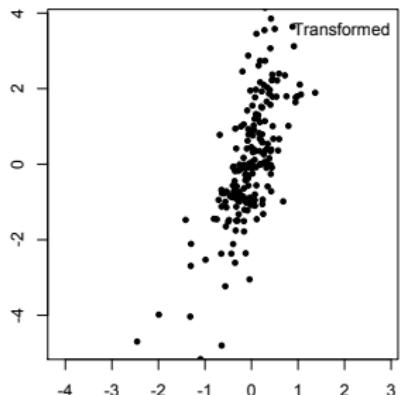
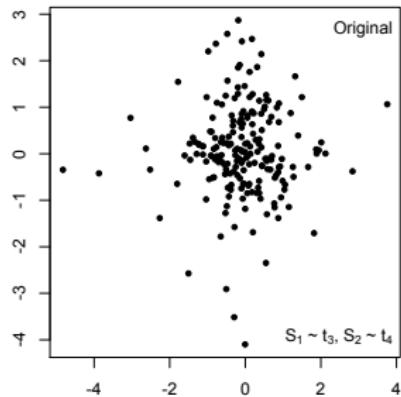
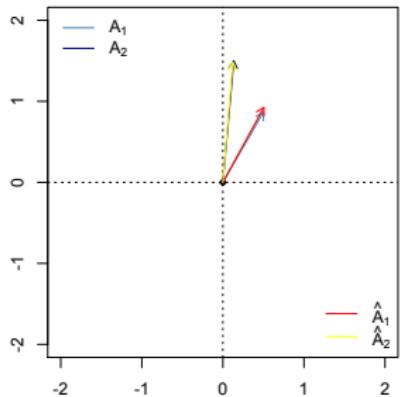
Pre-Whitening

- ▶ Assume each component of S has finite variance (can relax, c.f. Chen and Bickel (2005))
- ▶ Let $\Sigma = \text{cov}(X)$ and $Z = \Sigma^{-1/2}X$
- ▶ $S = O \cdot Z$, where $O = W \cdot \Sigma^{-1/2}$ is an orthogonal matrix
- ▶ Number of unknown parameters is reduced from d^2 to $d(d - 1)/2$

Non-Orthogonal Transformation, $S_1 \sim t_3$, $S_2 \sim t_4$

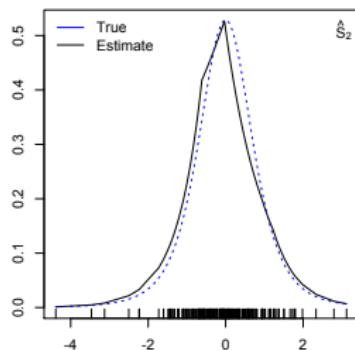
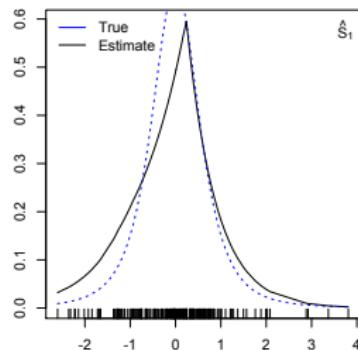
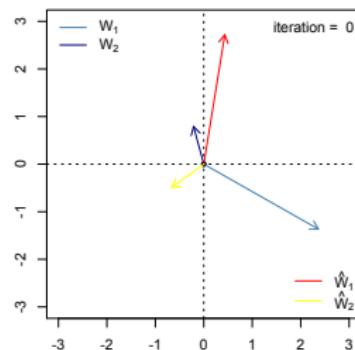


Non-Orthogonal Transformation, $S_1 \sim t_3$, $S_2 \sim t_4$



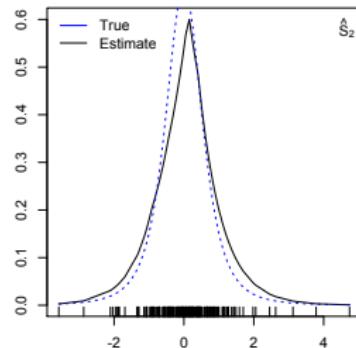
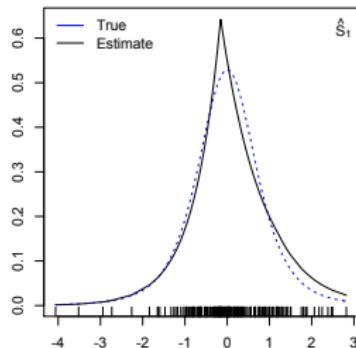
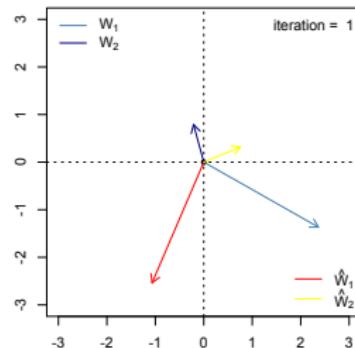
Convergence of Estimation

Non-orthogonal transformation, $S_1 \sim t_3$, $S_2 \sim t_4$



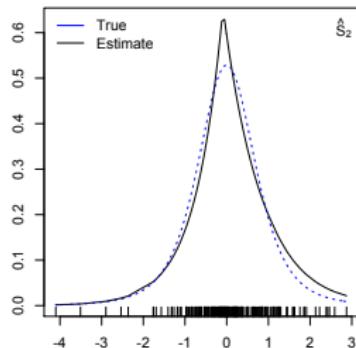
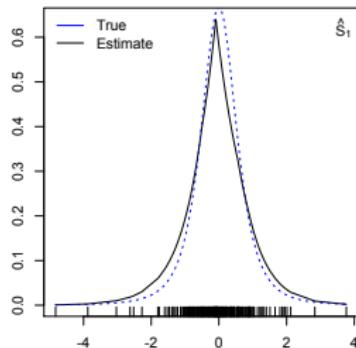
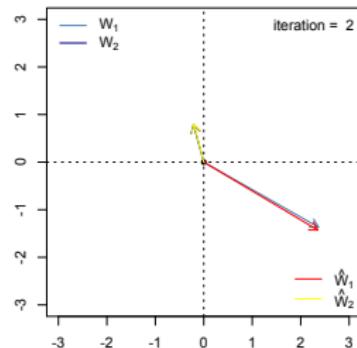
Convergence of Estimation

Non-orthogonal transformation, $S_1 \sim t_3$, $S_2 \sim t_4$

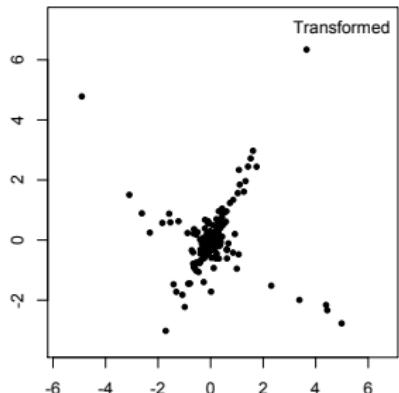
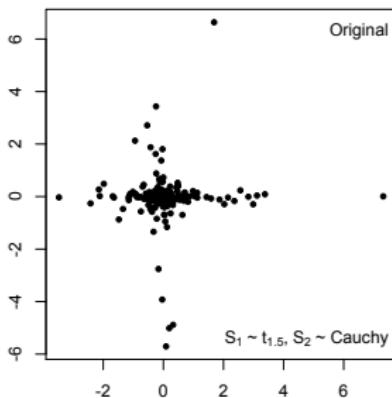
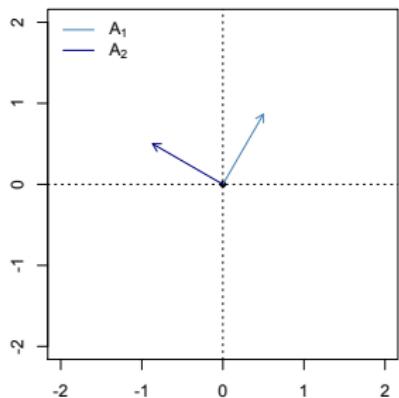


Convergence of Estimation

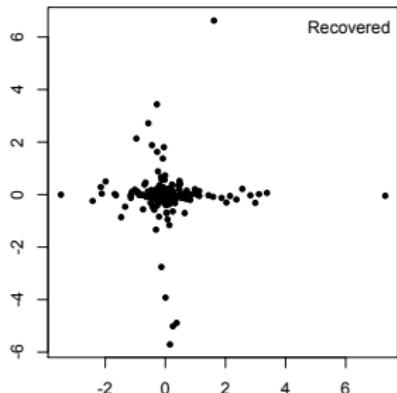
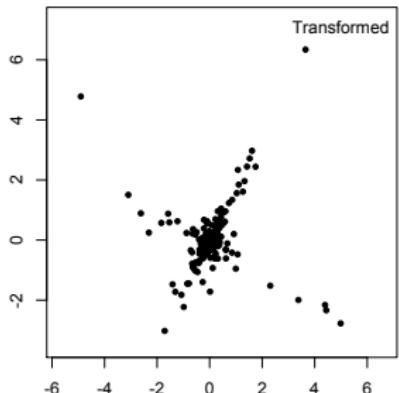
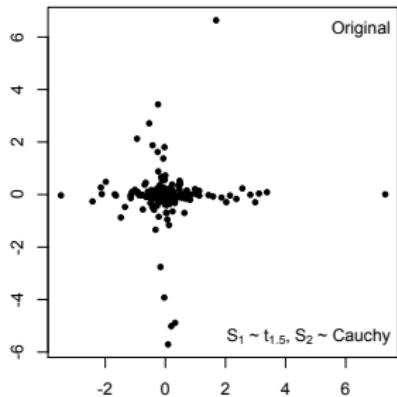
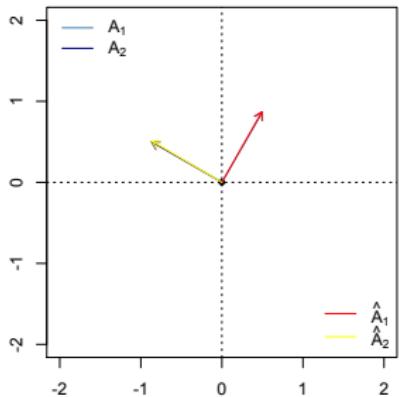
Non-orthogonal transformation, $S_1 \sim t_3$, $S_2 \sim t_4$



Rotation, $S_1 \sim t_{1.5}$, $S_2 \sim \text{Cauchy}$

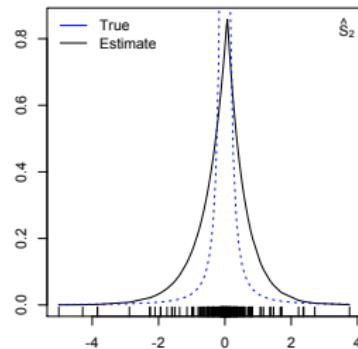
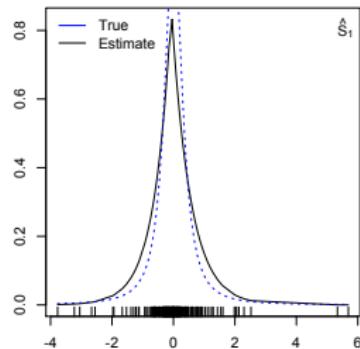
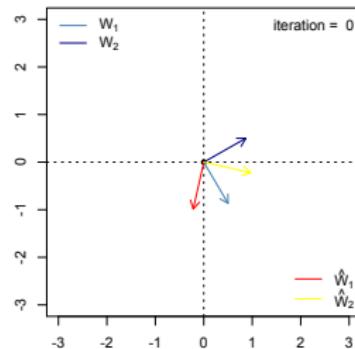


Rotation, $S_1 \sim t_{1.5}$, $S_2 \sim \text{Cauchy}$



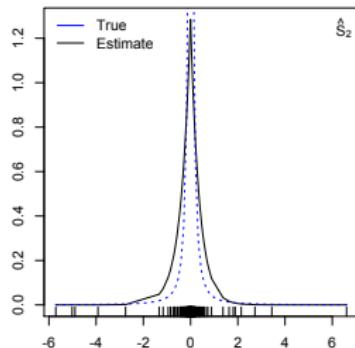
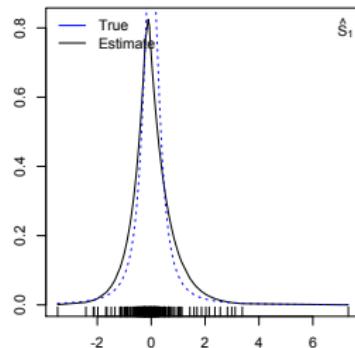
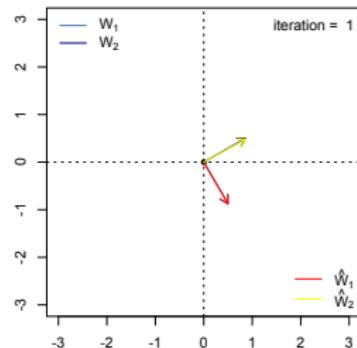
Convergence of Estimation

Non-orthogonal transformation, $S_1 \sim t_{1.5}$, $S_2 \sim \text{Cauchy}$



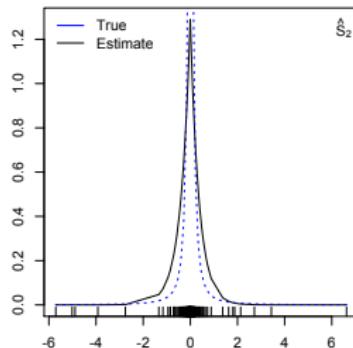
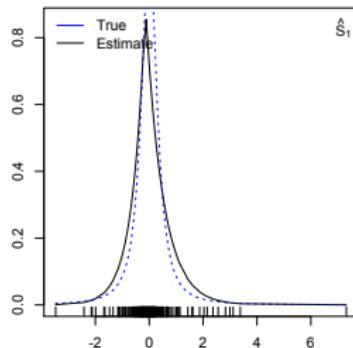
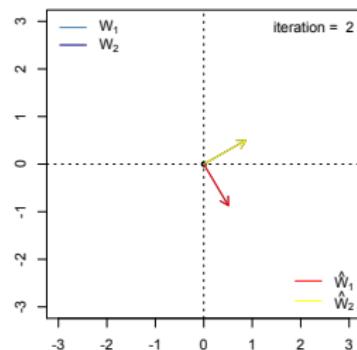
Convergence of Estimation

Non-orthogonal transformation, $S_1 \sim t_{1.5}$, $S_2 \sim \text{Cauchy}$

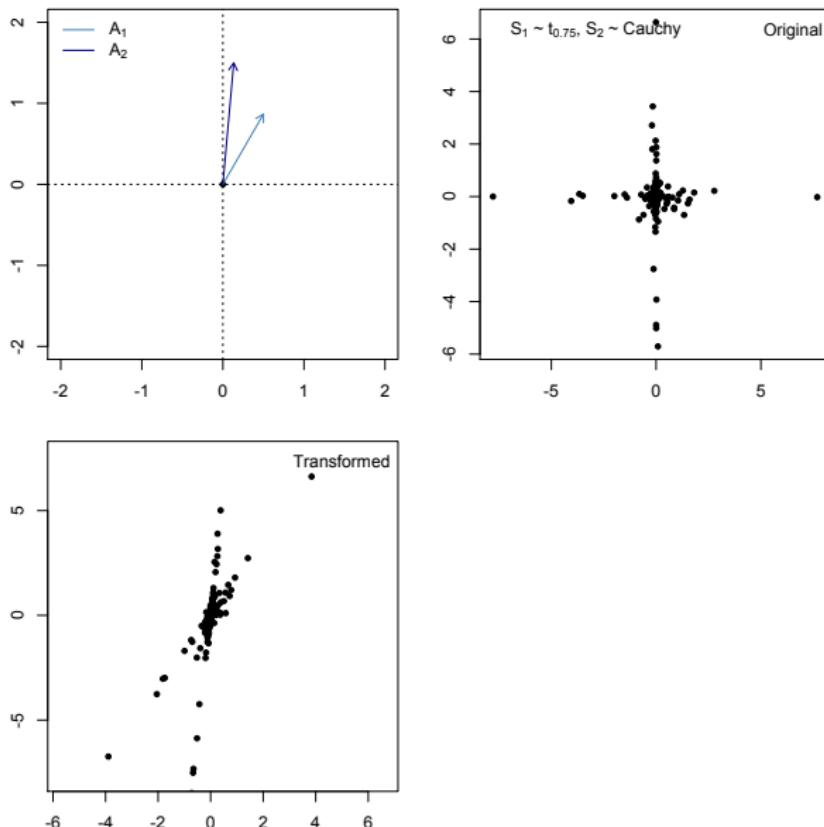


Convergence of Estimation

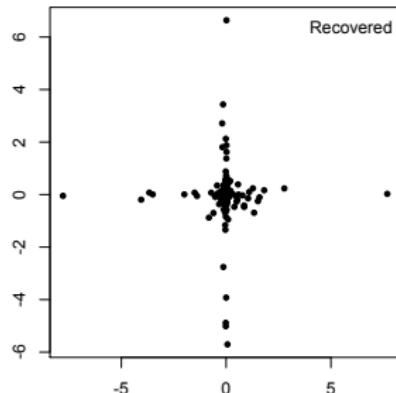
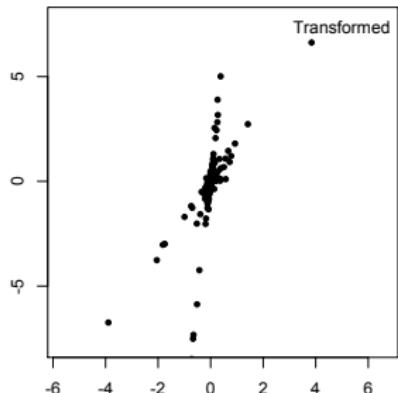
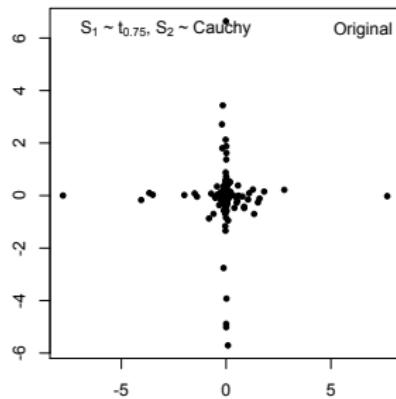
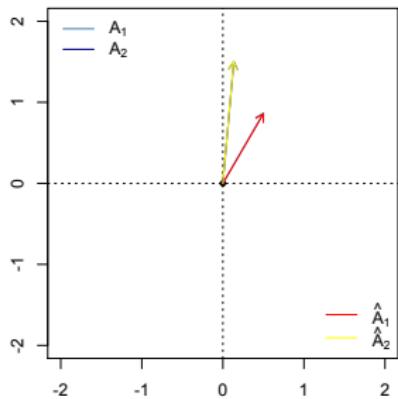
Non-orthogonal transformation, $S_1 \sim t_{1.5}$, $S_2 \sim \text{Cauchy}$



Non-Orthogonal Transformation, $S_1 \sim t_{0.75}$, $S_2 \sim \text{Cauchy}$

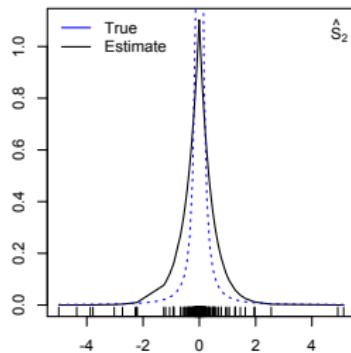
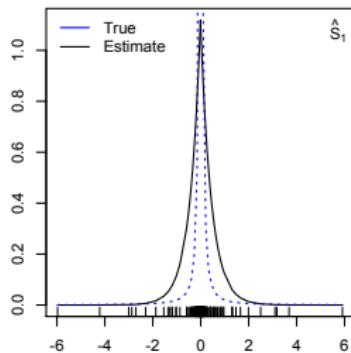
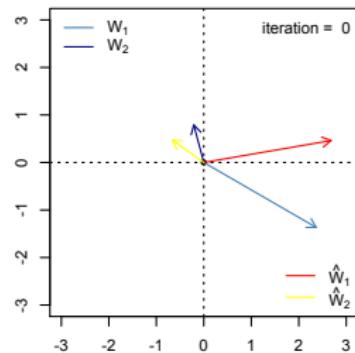


Non-Orthogonal Transformation, $S_1 \sim t_{0.75}$, $S_2 \sim \text{Cauchy}$



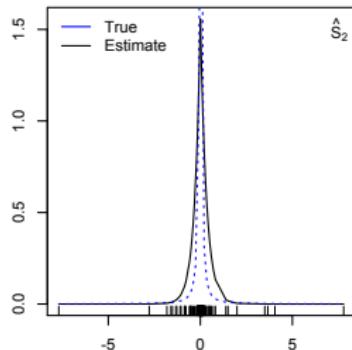
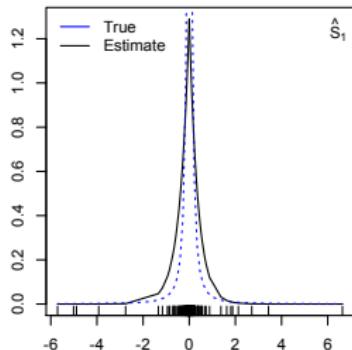
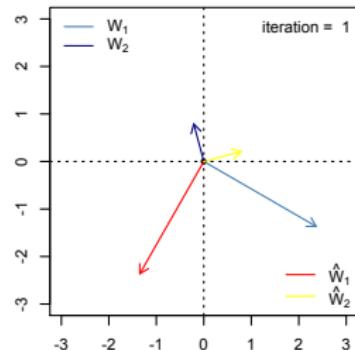
Convergence of Estimation

Non-orthogonal transformation, $S_1 \sim t_{0.75}$, $S_2 \sim \text{Cauchy}$



Convergence of Estimation

Non-orthogonal transformation, $S_1 \sim t_{0.75}$, $S_2 \sim \text{Cauchy}$



Convergence of Estimation

Non-orthogonal transformation, $S_1 \sim t_{0.75}$, $S_2 \sim \text{Cauchy}$

