A Parameter-Free Cauchy-Schwartz Information Measure for Independent Component Analysis

Lei Sun[†,⁎], Badong Chen[‡], Kar-Ann Toh[⋆], Zhiping Lin[⁎]

[†] School of Information and Electronics, Beijing Institute of Technology,
Beijing 100081, PR China. (e-mail: sunlei@bit.edu.cn)
[†] Xi'an Jiaotong University, Xi'an 710049, PR China.
[⋆] Yonsei University, Seoul 120-749, Republic of Korea.
[⁎] Nanyang Technological University,Singapore 639798, Singapore.

Mar., 2016

Outline

- Parameter-free survival distribution estimation
- SCS-MI is a valid Statistical Independence measure
- Empirical SCS-MI estimator for multiple variables
- Proposed ICA algorithm based on SCS-MI estimator
- Experiment and conclusion

Parameter-free survival distribution estimation

Survival distribution

$$\bar{F}_X(x) = \int_x^{\infty} f_X(u)du$$

a.k.a. the cumulative residual distribution or the tail distribution.

and its empirical estimation

$$\bar{F}_N(x) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}(u_n > x)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function ($\mathbb{I}(A)$ is 1 if event A occurs and is 0 otherwise).

Parameter-free survival distribution estimation

Survival distribution

$$\bar{F}_X(x) = \int_x^{\infty} f_X(u)du$$

a.k.a. the cumulative residual distribution or the tail distribution.

and its empirical estimation

$$\bar{F}_N(x) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}(u_n > x)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function ($\mathbb{I}(A)$ is 1 if event A occurs and is 0 otherwise).

Survival estimator is parameter-free

PDF estimator (Parzen window method)

$$\hat{p}_X(x) = \frac{1}{N}\sum_{n=1}^{N} k(x, u_n)$$

Survival estimator

$$\bar{F}_N(x) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}(u_n > x)$$

Survival estimator is parameter-free

PDF estimator (Parzen window method)

$$\hat{p}_X(x) = \frac{1}{N}\sum_{n=1}^{N} k(x, u_n)$$

Survival estimator

$$\bar{F}_N(x) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}(u_n > x)$$

Definition of CSIP

Cross Survival Information Potential (CSIP)

For two random variables X and Y (both in $\mathbb{R}_+$),

$$S_c(X, Y) = \int_{\mathbb{R}_+} \bar{F}_X(x)\bar{F}_Y(x)dx.$$

Joint and marginal CSIPs

Let (X, Y) and XY denote $\bar{F}_{(X,Y)}(X, Y)$ and $\bar{F}_X(X)\bar{F}_Y(Y)$, respectively.

$$S_c((X,Y),(X,Y)) = \int_{\mathbb{R}_+} \bar{F}_{(X,Y)}(x, y)\bar{F}_{(X,Y)}(x, y)dxdy,$$

$$S_c((X, Y), XY) = \int_{\mathbb{R}_+} \bar{F}_{(X,Y)}(x, y)\bar{F}_X(x)\bar{F}_Y(y)dxdy,$$

$$S_c(XY, XY) = \int_{\mathbb{R}_+} \bar{F}_X(x)\bar{F}_Y(y) \times \bar{F}_X(x)\bar{F}_Y(y)dxdy.$$

Definition of CSIP

Cross Survival Information Potential (CSIP)

For two random variables X and Y (both in $\mathbb{R}_+$),

$$S_c(X, Y) = \int_{\mathbb{R}_+} \bar{F}_X(x)\bar{F}_Y(x)dx.$$

Joint and marginal CSIPs

Let (X, Y) and XY denote $\bar{F}_{(X,Y)}(X, Y)$ and $\bar{F}_X(X)\bar{F}_Y(Y)$, respectively.

$$S_c((X,Y),(X,Y)) = \int_{\mathbb{R}_+} \bar{F}_{(X,Y)}(x, y)\bar{F}_{(X,Y)}(x, y)dxdy,$$

$$S_c((X, Y), XY) = \int_{\mathbb{R}_+} \bar{F}_{(X,Y)}(x, y)\bar{F}_X(x)\bar{F}_Y(y)dxdy,$$

$$S_c(XY, XY) = \int_{\mathbb{R}_+} \bar{F}_X(x)\bar{F}_Y(y) \times \bar{F}_X(x)\bar{F}_Y(y)dxdy.$$

Definition of SCS-MI

Survival Cauchy-Schwartz mutual information (SCS-MI) for two random variables is defined to evaluate the Cauchy-Schwartz divergence between the joint survival function $\bar{F}_{(X,Y)}(X, Y)$ and the product of the marginal survival functions $\bar{F}_X(X)\ \bar{F}_Y(Y)$:

SCS-MI

$$\mathcal{I}_{SCS}(X;Y) \stackrel{\text{def}}{=} -\log \frac{S_c((X,Y), XY)}{\sqrt{S_c((X,Y),(X,Y))}\sqrt{S_c(XY', XY')}}.$$

Definition of SCS-MI

Survival Cauchy-Schwartz mutual information (SCS-MI) for two random variables is defined to evaluate the Cauchy-Schwartz divergence between the joint survival function $\bar{F}_{(X,Y)}(X, Y)$ and the product of the marginal survival functions $\bar{F}_X(X)\ \bar{F}_Y(Y)$:

SCS-MI

$$\mathcal{I}_{\text{SCS}}(X;Y) \stackrel{\text{def}}{=} -\log \frac{S_c((X,Y),XY)}{\sqrt{S_c((X,Y),(X,Y))}\sqrt{S_c(XY,XY)}},$$

CSIPs for multiple variables

Let $(X^1, \ldots, X^D)$ denote the joint survival distribution $\bar{F}_{(X^1,\ldots,X^D)}(X^1, \ldots, X^D)$ and $X^1 \cdots X^D$ denote the product of marginal survival distributions $\bar{F}_{X^1}(X^1) \cdots \bar{F}_{X^D}(X^D)$.

$$S_c\left((x^1,\ldots,x^D),(x^1,\ldots,x^D)\right) = \int \left[\bar{F}_{(X^1,\ldots,X^D)}(x^1,\ldots,x^D)\right]^2 dx^1 \cdots dx^D$$

$$S_c\left(x^1 \cdots x^D, x^1 \cdots x^D\right) = \int \left[\bar{F}_{X^1}(x^1) \times \cdots \times \bar{F}_{X^D}(x^D)\right]^2 dx^1 \cdots dx^D$$

$$S_c\left((x^1,\ldots,x^D), x^1 \cdots x^D\right) =$$

$$\int \left[\bar{F}_{(X^1,\ldots,X^D)}(x^1,\ldots,x^D)\right] \times \left[\bar{F}_{X^1}(x^1) \times \cdots \times \bar{F}_{X^D}(x^D)\right] dx^1 \cdots dx^D,$$

SCS-MIs for multiple variables

Let $(X^1, \ldots, X^D)$ denote the joint survival distribution $\bar{F}_{(X^1,\ldots,X^D)}(X^1, \ldots, X^D)$ and $X^1 \cdots X^D$ denote the product of marginal survival distributions $\bar{F}_{X^1}(X^1) \cdots \bar{F}_{X^D}(X^D)$.

SCSM-MI

$$\mathcal{I}_{\text{SCSM}}\left(X^1; \ldots; X^D\right) \stackrel{\text{def}}{=}$$

$$-\log \frac{s_c\left((X^1,\ldots,X^D), X^1 \cdots X^D\right)}{\sqrt{s_c\left((X^1,\ldots,X^D),(X^1,\ldots,X^D)\right)}\sqrt{s_c(X^1 \cdots X^D, X^1 \cdots X^D)}},$$

SCS-MI is a valid Statistical Independence measure

Proposition 1

$\mathcal{I}_{\mathsf{SCS}}(X;Y) \geq 0$ and the equality holds if and only if X and Y are mutually independent.

Proposition 2

$\mathcal{I}_{\mathsf{SCSM}}\left(X^1;\ldots;X^D\right) \geq 0$ and the equality holds if and only if $X^1,\ldots,X^D$ are mutually independent.

SCS-MI is a valid Statistical Independence measure

Proposition 1

$\mathcal{I}_{\text{SCS}}\left(X;Y\right) \geq 0$ and the equality holds if and only if X and Y are mutually independent.

Proposition 2

$\mathcal{I}_{\text{SCSM}}\left(X^1;\ldots;X^D\right) \geq 0$ and the equality holds if and only if $X^1,\ldots,X^D$ are mutually independent.

The estimator of SCSM-MI is parameter-free

SCSM-MI estimator

$$\hat{\mathcal{I}}_{\text{SCSM}}\left(X^1;\ldots;X^D\right) =$$

$$-\log \frac{\sum_{n,\alpha_1,\alpha_2,\ldots,\alpha_D=1}^{N} \min(X_n^1, X_{\alpha_1}^1) \times \cdots \times \min(X_n^D, X_{\alpha_D}^D)}{\sqrt{\sum_{n,m=1}^{N}\left(\prod_{d=1}^{D} \min(X_n^d, X_m^d)\right)}\sqrt{\prod_{d=1}^{D}\left(\sum_{n,m=1}^{N} \min(X_n^d, X_m^d)\right)}}.$$

Proposed ICA algorithm based on SCS-MI estimator (1)

ICA model

Consider the estimation of D latent variables from a $N \times D$ observation matrix $\mathbf{X}$ representing a set of D variables each with N observations. The observations are assumed with linear but unknown combinations of the latent variables. The estimation goal is to find an $D \times D$ matrix $\mathbf{W}$ to recover the latent signals by

$$\hat{\mathbf{S}} = \mathbf{X}\mathbf{W},$$

where $\hat{\mathbf{S}}$ is the recovered signal matrix with each column being estimations for one of the D latent variables.

Proposed ICA algorithm based on SCS-MI estimator (2)

ICA algorithm

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} \hat{\mathcal{I}}_{\text{SCSM}}\left(\hat{S}^1; \ldots; \hat{S}^D\right)$$

where $\hat{S}^1, \ldots, \hat{S}^D$ denote the D estimated variables.

While *tol* and iteration k are within valid range

1. $\theta_u^{(k+1)} = \theta_u^{(k)} - \eta \nabla_{\theta_u} \hat{\mathcal{I}}_{\text{SCSM}} \left(\hat{S}^1; \ldots ; \hat{S}^D \right)^{\ddagger}$
2. $\mathbf{W}^{(k)} \leftarrow \prod_{i=1}^{D-1} \prod_{j=i+1}^{D} G_{ij}(\boldsymbol{\theta})$
3. $\hat{\mathbf{S}}^{(k)} \leftarrow \mathbf{X} \times \mathbf{W}^{(k)}$
4. $tol \leftarrow \hat{\mathcal{I}}_{\text{SCSM}}(\hat{\mathbf{S}}^{(k)} + t)$
5. $\theta_u^{(k)} = \theta_u^{(k+1)}$

Return $\mathbf{W}^*$, ($\hat{\mathbf{S}} = \mathbf{X}\mathbf{W}^*$)

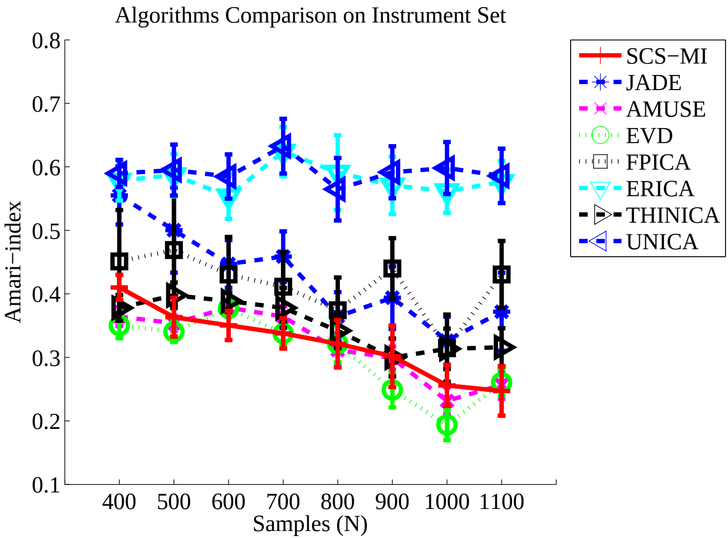
Experimental setup

Amari-index is used for de-mixing matrix quality assessment which was invariant to permutation and scaling of the columns of two compared matrices. The adopted Amari-index was defined as

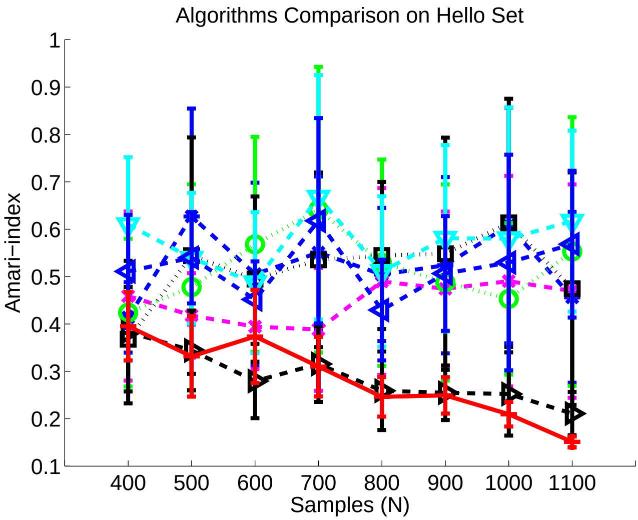
$$\text{Amari-index}(\mathbf{W}^*, \mathbf{M}) = \frac{1}{2D}\sum_{i=1}^{D}\left(\frac{\sum_{j=1}^{D}|r_{ij}|}{\max_j |r_{ij}|} - 1\right) + \frac{1}{2D}\sum_{j=1}^{D}\left(\frac{\sum_{i=1}^{D}|r_{ij}|}{\max_i |r_{ij}|} - 1\right)$$

where $r_{ij} = (\mathbf{W}^* \times \mathbf{M})_{ij}$ and $\mathbf{W}^*$ denotes the recovered de-mixing matrix. The Amari-index is equal to zero when two matrices represent the same components.

Experimental results (1)



Experimental results (2)



Conclusion

1. A probability survival distribution based Cauchy-Schwartz information measure for multiple variables is proposed
2. Empirical estimation of survival distribution is parameter-free which is inherited by the estimation of the new information measure.
3. This measure is a valid statistical independence measure and is adopted as an objective function to develop an ICA algorithm
4. This work shows promising potential regarding the use of survival distribution based information measure for ICA.

Conclusion

1. A probability survival distribution based Cauchy-Schwartz information measure for multiple variables is proposed
2. Empirical estimation of survival distribution is parameter-free which is inherited by the estimation of the new information measure.
3. This measure is a valid statistical independence measure and is adopted as an objective function to develop an ICA algorithm
4. This work shows promising potential regarding the use of survival distribution based information measure for ICA.

Conclusion

1. A probability survival distribution based Cauchy-Schwartz information measure for multiple variables is proposed
2. Empirical estimation of survival distribution is parameter-free which is inherited by the estimation of the new information measure.
3. This measure is a valid statistical independence measure and is adopted as an objective function to develop an ICA algorithm
4. This work shows promising potential regarding the use of survival distribution based information measure for ICA.

Conclusion

1. A probability survival distribution based Cauchy-Schwartz information measure for multiple variables is proposed
2. Empirical estimation of survival distribution is parameter-free which is inherited by the estimation of the new information measure.
3. This measure is a valid statistical independence measure and is adopted as an objective function to develop an ICA algorithm
4. This work shows promising potential regarding the use of survival distribution based information measure for ICA.

Acknowledgments

This is a joint work of Badong Chen from Xi'an Jiaotong Univ., China, Prof. Kar-Ann Toh from Yonsei Univ., Korea, Zhiping Lin from Nanyang Tech. Univ., Singapore, and Prof. Lei Sun is with Beijing Inst. of Tech., China. Part of this work was done when Lei Sun was a Senior Research fellow of Nanyang Tech. Univ., Singapore.

The End