Binary Independent Component Analysis : A Non-stationarity-based Approach

Hyttinen, Antti

2022

Hyttinen , A , Barin-Pacela , V & Hyvärinen , A 2022 , ' Binary Independent Component Analysis : A Non-stationarity-based Approach ' , 38th Conference on Uncertainty in Artificial Intelligence, UAI 2022 , Eindhoven , Netherlands , 01/08/2022 - 05/08/2022 pp. 874-884 . < https://www.cs.helsinki.fi/u/ajhyttin/codes/hyttinen_232-poster.pdf >

http://hdl.handle.net/10138/355812

unspecified
publishedVersion

Binary Independent Component Analysis: A Non-stationarity-based Approach

Antti Hyttinen, Vitória Barin-Pacela, Aapo Hyvärinen

Department of Computer Science and HIIT, University of Helsinki, Finland
Mila, Université de Montréal, Montréal, Canada

Summary

- Binary ICA is much less developed than continuous ICA, despite abundant binary data.
- Linear mixing model + continuous latent variables + binary observed variables.
- Employ **non-stationarity** of the sources (binarization would destroy non-Gaussianity).
- We develop a closed form likelihood via the Gaussian CDF and an approximate MLE.
- **Identifiable** only for certain numbers of observed variables, latent sources and segments.

Background

- ICA decomposes multivariate data into underlying factors.
- Non-stationarity is a well-known "path for ICA".
- Data is divided into segments, which express the non-stationarity.
- Examples of additionally observed variable: non-stationary time series, experimental condition, class label.
- Identifiability of ICA from binary data has not been theoretically proven.
- Potential applications in paleontological data, medical diagnosis, primary user separation in cognitive radio networks.

Binary ICA Model

The sources are non-stationary, in each segment u:

$$\mathbf{z}^u \sim \mathcal{N}(\boldsymbol{\mu}^u_{\mathbf{z}}, \boldsymbol{\Sigma}^u_{\mathbf{z}}),$$

- $\boldsymbol{\Sigma}^u_{\mathbf{z}}$ is diagonal
- Thus, the sources are independent given observed segment index u.

Linear mixing model:

$$\mathbf{y}^u = \mathbf{A}\mathbf{z}^u,$$

- $\mathbf{y}^u$ are continuous mixtures that would be observed in continuous ICA
- $\mathbf{A}$ is the mixing matrix.

Binary observations $\mathbf{x}^u$ through linking function:

$$P(x^u_i = 1) = \Phi(\sqrt{\frac{\pi}{8}} y^u_i | 0, 1)$$

- Φ is the Gaussian CDF
- matches closely the sigmoid, allows the evaluation of Gaussian integrals.

Closed-Form Likelihood

Probability of observing an assignment of ones:

$$P(\mathbf{x}^u = \mathbf{1}) = \int \Phi(\sqrt{\frac{\pi}{8}}\mathbf{y}^u|\mathbf{0},\mathbf{I})\mathcal{N}(\mathbf{y}^u|\mathbf{A}\boldsymbol{\mu}^u_{\mathbf{z}}, \mathbf{A}\boldsymbol{\Sigma}^u_{\mathbf{z}}\mathbf{A}^\intercal)d\mathbf{y}^u = P(\underbrace{\mathbf{n}}_{\sim\mathcal{N}(\mathbf{0},\mathbf{I})} < \sqrt{\frac{\pi}{8}}\mathbf{y}^u)$$

$$= P(\underbrace{\mathbf{n} - \sqrt{\frac{\pi}{8}}\mathbf{y}^u}_{\mathbf{q}^u} < \mathbf{0}) = \Phi(\mathbf{0}| \underbrace{-\sqrt{\frac{\pi}{8}}\mathbf{A}\boldsymbol{\mu}^u_{\mathbf{z}}}_{\boldsymbol{\mu}^u_{\mathbf{q}}}, \underbrace{\mathbf{I} + \frac{\pi}{8}\mathbf{A}\boldsymbol{\Sigma}^u_{\mathbf{z}}\mathbf{A}^\intercal}_{\boldsymbol{\Sigma}^u_{\mathbf{q}}}),$$

can be computed via the multivariate Gaussian CDF!

Likelihood:

$$l = \sum_u \sum_{\mathbf{x}^u} c(\mathbf{x}^u) \log \Phi(l(\mathbf{x}^u), u(\mathbf{x}^u)|\boldsymbol{\mu}^u_{\mathbf{q}}, \boldsymbol{\Sigma}^u_{\mathbf{q}}),$$

where the multivariate Gaussian PDF integrated from/to:

$$l(\mathbf{x}^u)_i = \begin{cases} -\infty \text{ if } x^u_i = 1 \\ 0 \text{ otherwise} \end{cases} \quad u(\mathbf{x}^u)_i = \begin{cases} 0 \text{ if } x^u_i = 1 \\ \infty \text{ otherwise} \end{cases}$$

- $c(\mathbf{x}^u)$ is the number of assignments $\mathbf{x}^u$ in segment u
- In 2D: assignment probability is the mass of a Gaussian in a quadrant.

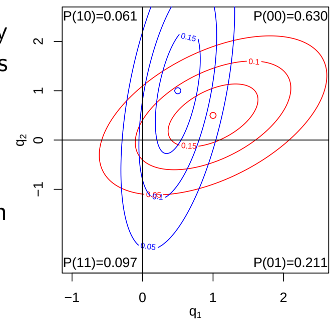
On Identifiability

Column scale and order Indeterminacy. Scale and order indeterminacies in the columns of the mixing matrix and sources, as in Linear ICA.

Binarization Indeterminacy. The probability of an assignment stays the same even if $\mathbf{q}^u$ is multiplied by a diagonal positive matrix $\mathbf{Q}^u$:

$$P(\mathbf{q}^u < \mathbf{0}) = P(\mathbf{Q}^u\mathbf{q}^u < \mathbf{0}).$$

- Figure: identical probabilities in each quadrant, but different Gaussians
- Scale information is lost in binarization.



Row Order Indeterminacy in 2D.

- Row order of 2-by-2 mixing matrix can be reversed without affecting the observed binary distributions.
- Thus, LiNGAM or ANM style causal orientation between two variables is not possible here.

Correlation identifiability. Correlation matrix of $\mathbf{q}^u$ is identifiable from binary data. (But covariance matrix (of $\mathbf{q}^u$) directly is not.)

Estimation Methods

Full MLE. Optimize the likelihood given with L-BFGS. Likelihood is computationally expensive to evaluate in large dimensions.

BLICA.

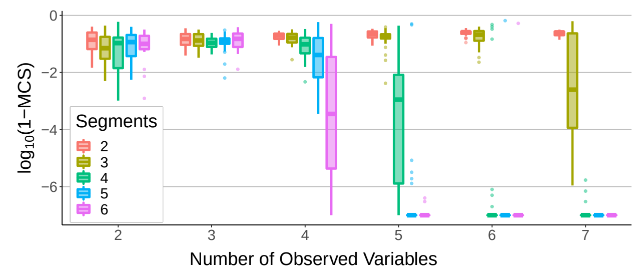
1. Estimate correlations of $\mathbf{q}^u$ via pairwise MLE.
2. Regularize correlation matrices.
3. Match correlation matrices to model parameters (continuous Gaussian MLE with L-BFGS, also $\mathbf{Q}^u$s as parameters – binarization indeterminacy).

Linear iVAE. Apply linear iVAE, optimization with L-BFGS.

Experiments

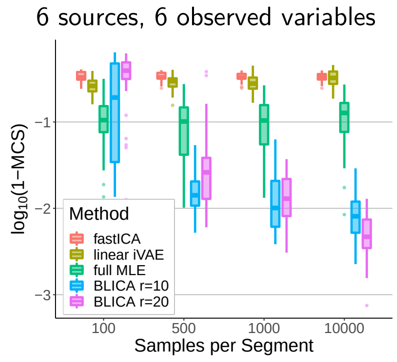
MCS. Mean cosine similarity (MCS) of the columns of $\mathbf{A}$. Does not penalize column scale and order.

Identifiability. Empirical identifiability when number of observed variables and segments is large enough.

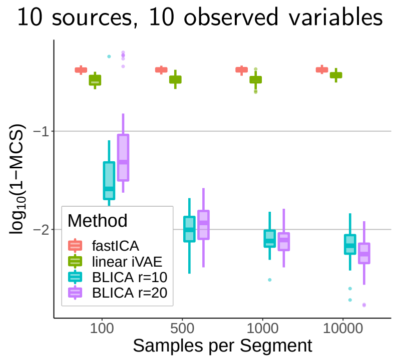


Finite sample performance. 40-segment datasets from the Binary ICA model.

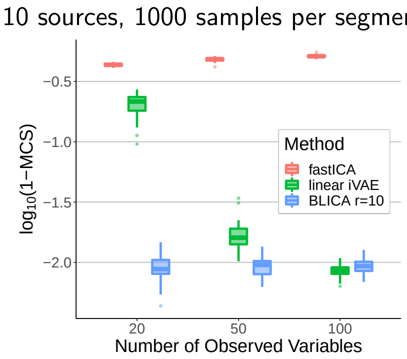
6 sources, 6 observed variables



10 sources, 10 observed variables



10 sources, 1000 samples per segment



Running time of BLICA