Multivariate Location and Scatter
Lecture 10
Scatter Matrix based ICA

Pauliina Ilmonen

Outline

IC Model

IC Model

In the independent component (IC) model it is assumed that the p-variate random vector

$$x = \Omega z + \mu, \tag{1}$$

where μ is a location vector, Ω is a full rank $p \times p$ mixing matrix, and z is a p-variate vector with mutually independent components with common median zero.

Independent Component Analysis

In the independent component analysis (ICA) the aim is to find an estimate of an unmixing matrix Γ such that Γx has independent components.

Applications

Applications

ICA is an important and timely research area. The field of applications of ICA is wide, varying from biomedical image data applications to signal processing, and economics.

Standardization

Standardization

Standardization of the IC model

The mixing matrix Ω in Model (1) is clearly not uniquely defined: for any $p \times p$ permutation matrix P and any full-rank diagonal matrix D, one can indeed always write

$$x = [\Omega P D]\,[(PD)^{-1} z] + \mu = \tilde{\Omega}\,\tilde{z} + \mu, \qquad (2)$$

where $\tilde{z}$ still has independent components with median zero.

Solving this identifiability problem requires either standardizing z or standardizing the mixing matrix Ω.

Independence Property

Location and Scatter Functionals

If a scatter matrix functional $S(F_x)$ is a diagonal matrix for all x having independent components, it is said to posses the independence property.

Location and Scatter Functionals

The regular covariance matrix is a scatter matrix with the independence property. Another example of a scatter matrix with the independence property is the matrix based on fourth moments.

Location and Scatter Functionals

Most scatter functionals do posses the independence property only if all the components (or all the components except for one) are symmetric. However, every scatter/shape matrix functional $S(F_x)$ can be symmetrized by setting

$$S_{sym}(F_x) = S(F_{x_1 - x_2}),$$

where x_1 and x_2 are independent random vectors having the same cumulative distribution function F_x. The resulting symmetrized scatter matrix does always have the independence property

Back to Standardization of the IC model

Standardization of the IC model

Vector z in Model (1) can be standardized using two different location functionals and two different scatter matrix functionals.

Standardization of the IC model

The marginal distributions of z in Model (1) can be standardized using two different location functionals T_1 and T_2 and two different scatter functionals S_1 and S_2, possessing the independence property, by setting

$$\begin{aligned} T_1(F_z) &= 0, \; S_1(F_z) = I_p, \\ T_2(F_z) &= \delta \; \text{ and } \; S_2(F_z) = D, \end{aligned}$$

where δ is a p-vector with all components $\delta_i \geq 0$, $i = 1, ..., p$, and D is a diagonal matrix with diagonal elements $d_1 \geq ... \geq d_p > 0$. If now $\delta_i > 0$, $i = 1, ..., p$, and if the diagonal elements of D are distinct, then the mixing matrix Ω is uniquely defined.

IC functionals Based on the Use of Two Scatter Matrices

Approach based on the use of two scatter matrices

Let $S_1(F_x)$ and $S_2(F_x)$ denote two different scatter functionals with the independence property. The IC functional $\Gamma(F_x)$ based on the scatter matrix functionals $S_1(F_x)$ and $S_2(F_x)$ is defined as a solution of the equations

$$\Gamma S_1(F_x)\Gamma^T = I_p \text{ and } \Gamma S_2(F_x)\Gamma^T = \Lambda,$$

where $\Lambda = \Lambda(F_x)$ is a diagonal matrix with diagonal elements $\lambda_1 \geq ... \geq \lambda_p > 0$.

Approach Based on the Use of Two Scatter Matrices

One of the first solutions for the ICA problem, the fourth order blind identification (FOBI) functional is obtained if the scatter functionals $S_1(F_x)$ and $S_2(F_x)$ are the scatter matrices based on the second and fourth moments, respectively.

Approach Based on the Use of Two Scatter Matrices

The functionals and corresponding sample statistics $G(X)$ and $L(X)$ are affine equivariant and invariant in the sense that

$$G(AX + b1_n^T) = G(X)A^{-1} \;\text{ and }\; L(AX + b1_n^T) = L(X)$$

for all $A \in \mathcal{M}$ and $b \in \mathbb{R}^p$. For the asymptotics, it is therefore not a restriction to assume that X is a random sample from a distribution F_x with $S(F_x) = I$ and $S_2(F_x) = \Lambda$, where the diagonal elements of Λ are $\lambda_1 \geq ... \geq \lambda_p > 0$.

Approach Based on the Use of Two Scatter Matrices

Assume that

$$\sqrt{n}(S_1(X) - I) = O_p(1) \text{ and } \sqrt{n}(S_2(X) - \Lambda) = O_p(1),$$

with $\lambda_1 > ... > \lambda_p > 0$, and assume that the diagonal elements of $G(X)$ are set to be positive. Then

$$\begin{aligned}
\sqrt{n}(G(X)_{ii} - 1) &= -\frac{1}{2}\sqrt{n}(S_1(X)_{ii} - 1) + o_p(1), \\
(\lambda_i - \lambda_j)\sqrt{n}G(X)_{ij} &= \sqrt{n}S_2(X)_{ij} - \lambda_i\sqrt{n}S_1(X)_{ij} + o_p(1), \ i \neq j, \text{ and} \\
\sqrt{n}(L(X)_{ii} - \lambda_i) &= \sqrt{n}(S_2(X)_{ii} - \lambda_i) - \lambda_i\sqrt{n}(S_1(X)_{ii} - 1) + o_p(1).
\end{aligned}$$

Approach Based on the Use of Two Scatter Matrices

It is interesting to note that the asymptotic behavior of the diagonal elements of $G(X)$ does not depend on $S_2(X)$ at all. The three equations above are in fact true if λ_i is distinct from all the other eigenvalues λ_j, $j \neq i$. The limiting joint distributions of the sample eigenvectors and sample eigenvalues for a subset with distinct population eigenvalues can then be derived from the limiting distributions of $S_1(X)$ and $S_2(X)$.

Complex valued Time Series ICA

Complex Valued Time Series IC Model

Let $X = (x_t)$, be a complex valued p variate is stochastic process. Assume that

$$x_t = \Omega z_t,$$

where Ω is a full-rank $p \times p$ complex valued mixing matrix, and where the components of $Z = (z_t)$ are assumed to be mutually independent with a common mean $E[z_t] = 0$.

Solution

Let a $p \times p$ matrix valued functional $G = G(X)$, and a $p \times p$ diagonal matrix valued functional $\Lambda = diag(\lambda_1, \cdots, \lambda_p) = \Lambda(X)$ be defined such that, if $z_t = G(X)x_t$, then

$$Cov(Z) = I_p, \text{ and } S_{\tau,symm}(Z) = \Lambda,$$

where $|\lambda_1| \geq ... \geq |\lambda_p|$.

The matrix $S_{\tau,symm}$ above is the symmetrized autocovariance matrix with lag τ.

Properties

Asymptotics under short and long range dependence...

Robust versions...

Extensions

Other latent variable models...

Tensor valued observations...

References I

J-F. Cardoso, Source separation using higher moments, Proceedings of IEEE international conference on acustics, speech and signal processing (1989), 2109–2112.

A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons, New York (2001).

P. Ilmonen, J. Nevalainen and H. Oja, Characteristics of multivariate distributions and the invariant coordinate system, Statistics and Probability Letters 80(23-24) (2010), 1844–1853.

P. Ilmonen, D. Paindaveine, Semiparametrically efficient inference based on signed ranks in symmetric independent component models. The Annals of Statistics 39(5) (2011), 2448–2476.

References II

H. Oja, S. Sirkiä, J. Eriksson, Scatter matrices and independent component analysis, Austrian Journal of Statistics 35 (2006), 175–189.

E. Ollila, H. Oja, V. Koivunen, Complex-valued ICA based on a pair of generalized covariance matrices, Computational Statistics & Data Analysis 52 (2008), 3789–3805.