ICA Model with Log-Concave Density Estimations

Presenter: Jingjing Zou
Joint work with Richard Davis

Columbia University

March 6th, 2015
The Model

\[ X = A \cdot S \]

- \( d \)-dimensional response \( X = (x_1, \cdots, x_d)^T \)
- \( d \)-dimensional independent components \( S = (S_1, \cdots, S_d)^T \)
- Full rank \( d \times d \) transformation matrix \( A \)
- \( S = W \cdot X \) with unmixing matrix \( W = (w_1, \cdots, w_d)^T = A^{-1} \)
- 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 $\Psi^*(P) : P_d \rightarrow \mathcal{F}_d$

$$\Psi^*(P) := \arg\max_{f \in \mathcal{F}_d} \int_{\mathbb{R}^d} \log(f) \, dP$$
Projection to $\mathcal{F}_d^{ICA}$

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

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

**Theorem (Samworth and Yuan (2012))**

If distribution $P$ has density $f(x) = |\text{det} W| \prod_{j=1}^{d} f_j(w_j^T x)$, then $\Psi^*(P) = \Psi^{**}(P) := \arg\max_{f \in \mathcal{F}_d^{ICA}} \int_{\mathbb{R}^d} \log(f) \, dP$, and it equals to

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

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

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

- 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, $S_1 \sim t_{1.5}$, $S_2 \sim$ Cauchy

Original

Transformed

Recovered
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, $S_1 \sim t_{1.5}$, $S_2 \sim$ Cauchy
Overcomplete ICA

- $d$-dimensional response $X = (x_1, \cdots, x_d)^T$
- $m$-dimensional independent components $S = (S_1, \cdots, S_m)^T$
- $m > d$
- Full rank (non-degenerate) $d \times m$ transformation matrix $A = (a_1, \ldots, a_d)^T$
- $X = A \cdot S$
- The goal is to recover the transformation matrix $A$ and the independent components $S$
Overcomplete ICA: Applications

- Estimate multivariate stable distributions
  - $S = (S_1, \cdots, S_m)^T$ and each $S_j$ is univariate stable
  - $X = A \cdot S$ is multivariate stable

- Recognition tasks
  - Action recognition (Zhang et al., 2014)
  - Image feature extraction (Le et al., 2011)
Overcomplete ICA

- Recall the ICA model assumes the distribution of $X$ when $m = d$ (undercomplete) is

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

where $W = (w_1, \ldots, w_d)^T = A^{-1}$ exists when $A$ is invertible

- Therefore for each $W$ one can recover a unique estimate of $S$ and compare with the proposed $\hat{P}_j$

- In particular, we use the log-concave projection $\hat{f}_j^*$ to estimate $P_j$ and estimate $W$ and $\hat{f}_j^*$ iteratively

- The difficulty in the overcomplete case is $A$ is not invertible
Overcomplete ICA: Pre-Whitening

- Singular Value Decomposition (SVD) to reduce the number of parameters to estimate in the $d \times m$ matrix $A$
  
  $A = U\Sigma V^T$

- $U : d \times d$ orthogonal, $\Sigma : d \times d$ diagonal, $V : m \times d$ orthogonal

- As in the undercomplete case, assume each component $S_j$ has finite variance (can relax by results in Chen and Bickel, 2005) and is standardized

- $\text{cov}(X) = U\Sigma^2 \tilde{V}^T$

- Let $Y = (U\Sigma)^{-1}X$, then $Y = V^T S$

- $(U\Sigma)^{-1}$ can be estimated using the SVD of the sample covariance of $X$
Pseudo-Inverse of the Transformation

▶ For pre-whitened under-complete ICA model \( Y = V^T S \), where \( V \) is \( d \times d \) orthogonal, \( S = VY \), and \( V \) can be estimated by maximizing the log-likelihood function

\[
\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \log \hat{f}_j^*(s_{ji}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \log \hat{f}_j^*(v_j^T y_i)
\]

▶ When \( V \) is \( m \times d \) orthogonal matrix, \( \{ S : Y = V^T S \} \) is not unique

▶ A simple strategy is to use the pseudo-inverse of \( V \), which is just \( V^T \) if \( V \) is orthogonal: \( V^T V = I_d \)

▶ Consistency may fail since \( VV^T \neq I_m \) and thus \( S \neq VY \)
A Refined Strategy

- Solve for

\[
\arg\max_{\hat{f}, V} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \log \hat{f}_j^* (s_{ji}^* (V)) \right\},
\]

where

\[
s_{j}^* (V) = \arg\max_{s: Y=Vs} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \log \hat{f}_j^* (s_{ji}) \right\}
\]
Estimation Procedure for Pre-Whitened Data

- Start from an arbitrary initial value of $V$ and $S$ such that $Y = VS$
- Step 1: Find log-concave projection $\hat{f}_{j}^{*}$ of the distribution of $S_{j}$
- Step 2: With $\hat{f}_{j}^{*}$, update $V$ to maximize the log-likelihood

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \log \hat{f}_{j}^{*}(s_{ji}^{*}(V)),$$

which contains an optimization step

$$s^{*}(V) = \arg\max_{\{s: Y = VS\}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \log \hat{f}_{j}^{*}(s_{ji}) \right\}$$

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

- $m = 3, d = 2$
- $S_1 : \text{stable}(\alpha = 1.2, \beta = 0.1, \gamma = 1, \delta = 0)$
- $S_2 : \text{stable}(\alpha = 1.1, \beta = 0.7, \gamma = 1, \delta = 0)$
- $S_3 : \text{stable}(\alpha = 1.5, \beta = 0.3, \gamma = 1, \delta = 0)$
- Index parameter $\alpha$; skewness $\beta$; scale $\gamma$; and location (shift) $\delta$
- Transformation $A$: combinations of rotations with angles $(\pi/6, 2\pi/3, \pi/3)$
Data
Recovered $S$ vs. Pseudo-Inverse

- y-axis: true values of $S_j$
- Top: recovered $S$ given true $V$ and log-concave projections $\hat{f}_j$
- Bottom: With pseudo-inverse $S = V^T Y$
References


