Copula Based Independent Component Analysis

CUNY October 2008

Kobi Abayomi [+]

[+]Asst. Professor
Industrial Engineering - Statistics Group
Georgia Institute of Technology

October 2008

Outline

A copula is a distribution function...

Take $X_1 \sim F_{X_1}, \; X_2 \sim F_{X_2}$

...on the marginal distributions of random variables

Set $U = F_{X_1}$ and $V = F_{X_2}$.

- The pair (U, V) are the '*grades*' of (X_1, X_2)
- i.e. the mapping of (X_1, X_2) in F_{X_1}, F_{X_2} space.

...on the marginal distributions of random variables

Set $U = F_{X_1}$ and $V = F_{X_2}$.

- The pair (U, V) are the '*grades*' of (X_1, X_2)
- i.e. the mapping of (X_1, X_2) in F_{X_1}, F_{X_2} space.

Copula specification

A copula is a function that takes the 'grades' as arguments and returns a joint distribution function, with marginals F_{X_1}, F_{X_2}.

$$C(U, V) = F_{X_1, X_2}$$

Copula generation

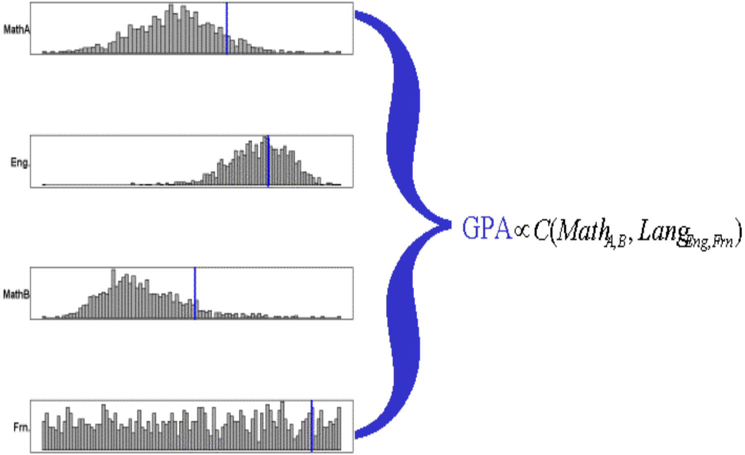
Any multivariate distribution function can yield a copula function.

$$F_{X_1,X_2}(F_{X_1}^{-1}(U), F_{X_2}^{-1}(V)) = C'(U, V)$$

Heuristically speaking

The correspondence which assigns the value of the joint distribution function to each ordered pair of values (F_{X_1}, F_{X_2}) for each X_1, X_2 is a distribution function called a Copula.

GPA



Simple Example: Bivariate Distribution

$$H_\theta(x, y) = 1 - e^{-x} - e^{-y} + e^{-(x+y+\theta xy)}$$

for $x, y \in \mathbb{R}^+$. $H = 0$ otherwise.

Simple Example: Bivariate Distribution

$$H_X^{(-1)}(u) = -\ln(1-u); H_Y^{(-1)}(v) = -\ln(1-v)$$

Simple Example: Bivariate Distribution

$$C_\theta(u,v) = H(-ln(1-u), -ln(1-v)) =$$

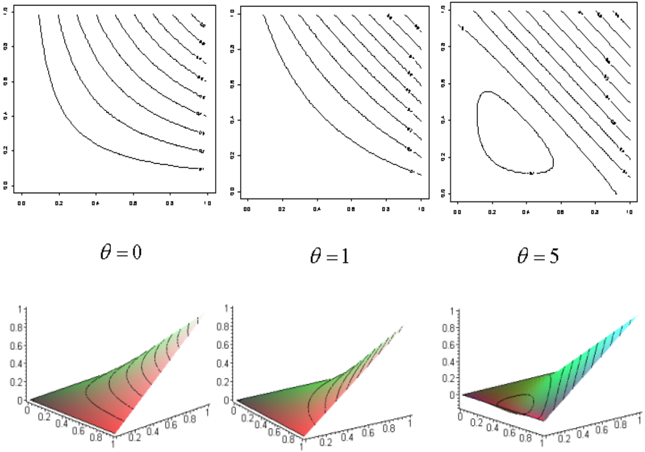
$$= (u+v-1) + (1-u)(1-v) * e^{-\theta \ln(1-u)\ln(1-v)}$$

Notice if $\theta = 0$

$$C_\theta(u,v) = uv$$

...the independence copula.

Bivariate Copula $\theta = 0, 1, 5$



Why use a Copula

Copulas are useful when modelling:

- Multivariate settings where a different family is needed for each marginal distribution.
- Parametric estimates/versions of measures of association.

Why use a Copula

Copulas are useful when modelling:

- Multivariate settings where a different family is needed for each marginal distribution.
- Parametric estimates/versions of measures of association.

Copulas dependent measures of association

Many measures of association can be expressed as solely copula dependent. Kendalls' Tau and, for example, Spearman's Rho with $U = F_X, V = F_Y$:

$$\rho_{(X,Y)} = 12 \int \int C(u,v) du dv - 3$$

$$\rho_{(U,V)} = 12E(C(u,v)) - 3$$

Outline

Copula Density

With $\mathbf{u} = (F_{X_1}, ..., F_{X_k})$, the copula density $dC(\mathbf{u})$ is

$$dC(\mathbf{u}) = \frac{dF_{\mathbf{x}}(\mathbf{x})}{\prod dF_{x_i}(x_i)}$$

Mutual Information

It turns out the mutual information

$$MI(\mathbf{x}) \equiv \int_{\mathbb{R}^k} dF_{\mathbf{X}} log(\frac{dF_{\mathbf{X}}}{\prod dF_{X_i}})$$

is solely copula dependent...

Mutual Information is Copula based

Since, with $\mathbf{u} = (F_{X_1}, ..., F_{X_k})$

$$MI(\mathbf{x}) \equiv \int_{\mathbb{I}^k} dC(\mathbf{u}) log(dC(\mathbf{u}))$$

Thus the MI is a copula based measure of association.

Mutual Information

- MI is often used as a proxy for independence in general (i.e. non gaussian) settings.
- MI is the Kullback-Liebler divergence ('distance') between dependence and independence.
- MI$= 0$ implies independence.

Mutual Information

- MI is often used as a proxy for independence in general (i.e. non gaussian) settings.
- MI is the Kullback-Liebler divergence ('distance') between dependence and independence.
- MI$=0$ implies independence.

Mutual Information

- MI is often used as a proxy for independence in general (i.e. non gaussian) settings.
- MI is the Kullback-Liebler divergence ('distance') between dependence and independence.
- MI$= 0$ implies independence.

Outline

The PCA model

- Given multivariate data $\mathbf{x}_k$ with scatter matrix Σ
  PCA program: Find B such that $\mathbf{y} = B\mathbf{x}$ yields uncorrelated y_i and y_j.

The PCA model

- Given multivariate data $\mathbf{x}_k$ with scatter matrix Σ
  PCA program: Find B such that $\mathbf{y} = B\mathbf{x}$ yields uncorrelated y_i and y_j.

The PCA model

Well known result: Singular Value Decomposition (SVD)

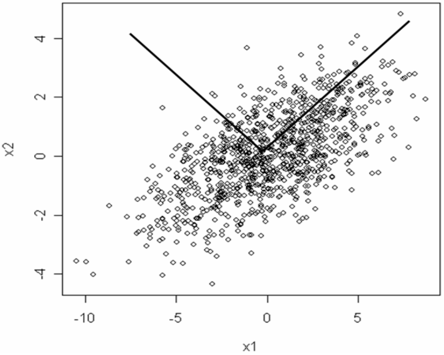
$$\Sigma = \mathbf{e}^t \Lambda \mathbf{e}$$

Yielding:

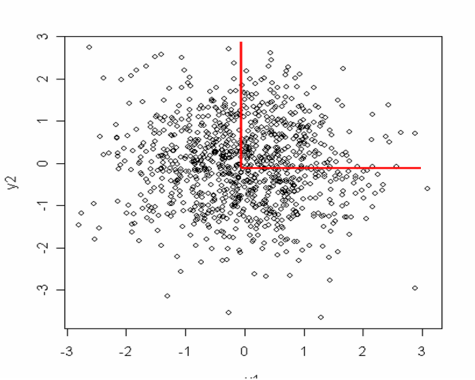
$$y_i = \mathbf{e}^t \mathbf{x}$$

with $Cov(y_i, y_j) = 0,\ i \neq j$.

The PCA model - mixed data



The PCA model - rotated data



Independent Component Analysis (ICA)

ICA is the PCA program under the more general assumption of statistical independence

- Given $\mathbf{x}_k$
- ICA program: Find $\mathbf{y} = B\mathbf{x}$ such that $y_i = b_i\mathbf{x}$ are independent of $y_j = b_j\mathbf{x}$.

Independent Component Analysis (ICA)

ICA is the PCA program under the more general assumption of statistical independence

- Given $\mathbf{x}_k$
- ICA program: Find $\mathbf{y} = B\mathbf{x}$ such that $y_i = b_i\mathbf{x}$ are independent of $y_j = b_j\mathbf{x}$.

Independent Component Analysis (ICA)

- In ICA: $\mathbf{x} = A\mathbf{s}$ — the observed data — are the mixed outputs, B, of independent sources $\mathbf{s}$.
- The $\mathbf{y}$ are the estimates ($\hat{\mathbf{s}}$) of these independent components, or signals.
- B is an estimate of A^{-1}; $B = \widehat{A^{-1}}$.

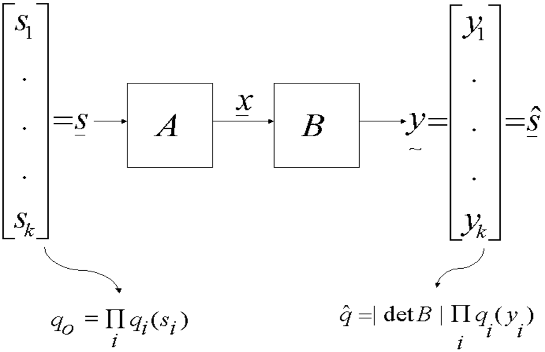
Independent Component Analysis (ICA)

- In ICA: $\mathbf{x} = A\mathbf{s}$ — the observed data — are the mixed outputs, B, of independent sources $\mathbf{s}$.
- The $\mathbf{y}$ are the estimates ($\hat{\mathbf{s}}$) of these independent components, or signals.
- B is an estimate of A^{-1}; $B = \hat{A}^{-1}$.

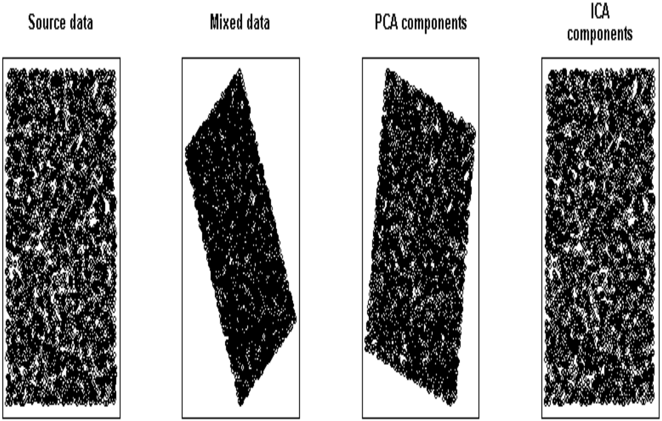
Independent Component Analysis (ICA)

- In ICA: $\mathbf{x} = A\mathbf{s}$ — the observed data — are the mixed outputs, B, of independent sources $\mathbf{s}$.
- The $\mathbf{y}$ are the estimates ($\hat{\mathbf{s}}$) of these independent components, or signals.
- B is an estimate of A^{-1}; $B = \hat{A}^{-1}$.

The ICA model - illustration



The ICA model



Comments on ICA

- In both PCA and ICA the objective is the recovery of the mixing A of the independent signals $\mathbf{x}$. The difference is the characterization of statistical independence.
- True statistical independence requires factorization of probability densities.
- ICA procedures often use high order moments, or empirical mutual information as independence proxies.
- Most are non-parametric approaches to estimating statistical independence.

Comments on ICA

- In both PCA and ICA the objective is the recovery of the mixing A of the independent signals $\mathbf{x}$. The difference is the characterization of statistical independence.
- True statistical independence requires factorization of probability densities.
- ICA procedures often use high order moments, or empirical mutual information as independence proxies.
- Most are non-parametric approaches to estimating statistical independence.

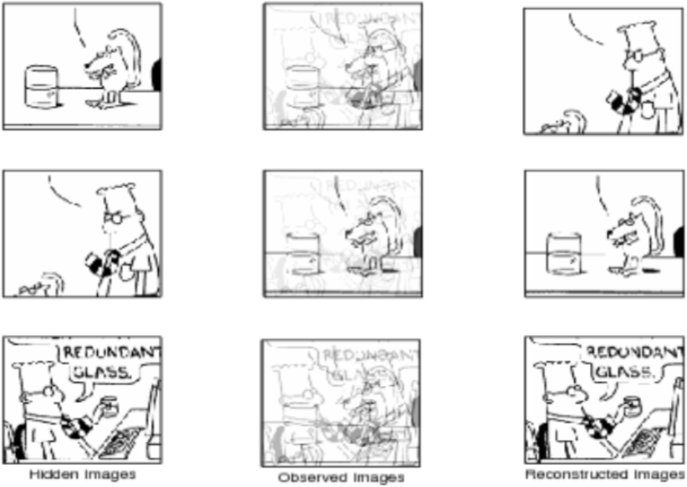
Comments on ICA

- In both PCA and ICA the objective is the recovery of the mixing A of the independent signals $\mathbf{x}$. The difference is the characterization of statistical independence.
- True statistical independence requires factorization of probability densities.
- ICA procedures often use high order moments, or empirical mutual information as independence proxies.
- Most are non-parametric approaches to estimating statistical independence.

Comments on ICA

- In both PCA and ICA the objective is the recovery of the mixing A of the independent signals $\mathbf{x}$. The difference is the characterization of statistical independence.
- True statistical independence requires factorization of probability densities.
- ICA procedures often use high order moments, or empirical mutual information as independence proxies.
- Most are non-parametric approaches to estimating statistical independence.

The ICA model



Copula Based Independent Component Analysis (CICA)

General Approach

- Replace non-parametric measures of dependence-independence with parametric copula families
- Appeal to the information theoretic 'distance' - K-L divergence
- Exploit the role of the copula.

Copula Based Independent Component Analysis (CICA)

General Approach

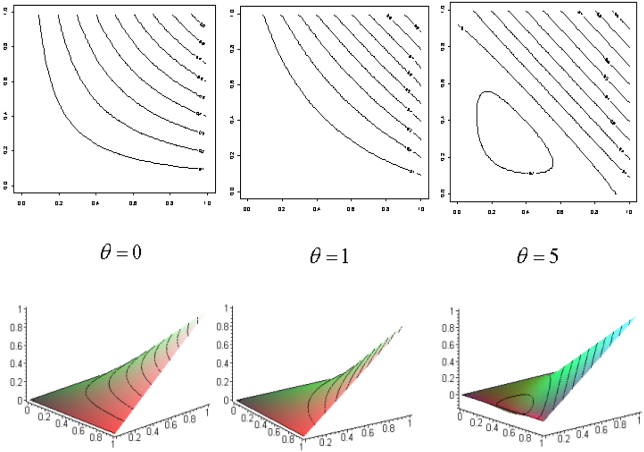
- Replace non-parametric measures of dependence-independence with parametric copula families
- Appeal to the information theoretic 'distance' - K-L divergence
- Exploit the role of the copula.

Copula Based Independent Component Analysis (CICA)

General Approach

- Replace non-parametric measures of dependence-independence with parametric copula families
- Appeal to the information theoretic ‘distance’ - K-L divergence
- Exploit the role of the copula.

Bivariate Copula $\theta = 0, 1, 5$



K-L divergence

The Kullback-Liebler divergence between two probability density functions $f(\mathbf{t})$ and $g(\mathbf{t})$ is

$$\mathbb{K}(f,g) = \int_{\mathbf{t}} f(\mathbf{t}) log(\frac{f(\mathbf{t})}{g(\mathbf{t})}) \tag{1}$$

I can (ab)use the notation $\mathbb{K}(\mathbf{w},\mathbf{z})$ for the divergence between the distribution of two random vectors **w** and **z**. $\mathbb{K} \geq 0$ with equality if and only if **w** and **z** have the same distribution.

Divergence decomposition

A classic property (Kullback [1968], others) of (1) is

$$\mathbb{K}(\mathbf{y}, \mathbf{s}) = \mathbb{K}(\mathbf{y}, \mathbf{y}^*) + \mathbb{K}(\mathbf{y}^*, \mathbf{s}) \tag{2}$$

with $\mathbf{y}^*$ a random vector with independent entries and margins distributed as $\mathbf{y}$; $\mathbf{s}$ is an independent vector.

$$\begin{pmatrix}\text{Total}\\ \text{Mismatch}\end{pmatrix} = \begin{pmatrix}\text{Deviation from}\\ \text{Independence}\end{pmatrix} + \begin{pmatrix}\text{Marginal}\\ \text{Mismatch}\end{pmatrix}$$

Divergence decomposition

- The LHS — $\mathbb{K}(\mathbf{y}, \mathbf{s})$ — the divergence of the ICA outputs (estimates) from the hypothesized inputs (sources).
- $\mathbb{K}(\mathbf{y}, \mathbf{y}^*)$ is the independence-dependence of the outputs
- $\mathbb{K}(\mathbf{y}^*, \mathbf{s})$ is the mismatch of the margins of the estimates from the margins of the sources.

Divergence decomposition

- The LHS — $\mathbb{K}(\mathbf{y}, \mathbf{s})$ — the divergence of the ICA outputs (estimates) from the hypothesized inputs (sources).
- $\mathbb{K}(\mathbf{y}, \mathbf{y}^*)$ is the independence-dependence of the outputs
- $\mathbb{K}(\mathbf{y}^*, \mathbf{s})$ is the mismatch of the margins of the estimates from the margins of the sources.

Divergence decomposition

- The LHS — $\mathbb{K}(\mathbf{y}, \mathbf{s})$ — the divergence of the ICA outputs (estimates) from the hypothesized inputs (sources).
- $\mathbb{K}(\mathbf{y}, \mathbf{y}^*)$ is the independence-dependence of the outputs
- $\mathbb{K}(\mathbf{y}^*, \mathbf{s})$ is the mismatch of the margins of the estimates from the margins of the sources.

L maximization implies KL minimization

Model **x** as generated from $A\mathbf{s} = \mathbf{x}$.

- Log likelihood for n independent samples of **x**, under our estimate $\hat{q}$ of the true source distribution is:

$$L_A = n^{-1} \sum log(\hat{q}(\mathbf{A}^{-1}\mathbf{x})) - log(|det\mathbf{A}|)$$

- Which converges to

$$\int q(\cdot) log(\hat{q}(\cdot)) + cst$$

.

- Which can be rewritten:

$$\int q(\cdot) log(q(\cdot)) - \int q(\cdot) log(\frac{q(\cdot)}{\hat{q}(\cdot)}) = H(\mathbf{y}) - K(q(\cdot), \hat{q}(\cdot))$$

L maximization implies KL minimization

Model **x** as generated from $A\mathbf{s} = \mathbf{x}$.

- Log likelihood for n independent samples of **x**, under our estimate $\hat{q}$ of the true source distribution is:

$$L_A = n^{-1} \sum log(\hat{q}(\mathbf{A}^{-1}\mathbf{x})) - log(|det\mathbf{A}|)$$

- Which converges to

$$\int q(\cdot)log(\hat{q}(\cdot)) + cst$$

.

- Which can be rewritten:

$$\int q(\cdot)log(q(\cdot)) - \int q(\cdot)log(\frac{q(\cdot)}{\hat{q}(\cdot)}) = H(\mathbf{y}) - K(q(\cdot), \hat{q}(\cdot))$$

L maximization implies KL minimization

Model **x** as generated from $A\mathbf{s} = \mathbf{x}$.

- Log likelihood for n independent samples of **x**, under our estimate $\hat{q}$ of the true source distribution is:

$$L_A = n^{-1} \sum log(\hat{q}(\mathbf{A}^{-1}\mathbf{x})) - log(|det\mathbf{A}|)$$

- Which converges to

$$\int q(\cdot) log(\hat{q}(\cdot)) + cst$$

.

- Which can be rewritten:

$$\int q(\cdot) log(q(\cdot)) - \int q(\cdot) log(\frac{q(\cdot)}{\hat{q}(\cdot)}) = H(\mathbf{y}) - K(q(\cdot), \hat{q}(\cdot))$$

ML maximization implies KL minimization

$$= H(\mathbf{y}) - \mathbb{K}(\mathbf{y}, \mathbf{s}) = H(\mathbf{x}) - \mathbb{K}(\mathbf{y}, \mathbf{s})$$

The RHS first term is the entropy of the inputs; the RHS second term is the distance between (true) sources and their estimates. ML maximization is equivalent to minimization of the KL ‘distance’ between the outputs and the sources.

Copula Based Independent Component Analysis (CICA)

The procedure is to recast (2) via the copula:

- $\mathbb{K}(\mathbf{y}, \mathbf{y}^*) = \mathbf{MI}(\mathbf{y}) = -H(dC(\mathbf{u}))$ if $\mathbf{u}$ are the marginal distribution functions $u_i = F_i(y_i)$.
- $\mathbb{K}(\mathbf{y}^*, \mathbf{s}) = \sum \mathbb{K}({y^*}_i, s_i)$, since both $\mathbf{y}^*, \mathbf{s}$ have independent entries.

Copula Based Independent Component Analysis (CICA)

The procedure is to recast (2) via the copula:

- $\mathbb{K}(\mathbf{y}, \mathbf{y}^*) = \mathbf{MI}(\mathbf{y}) = -H(dC(\mathbf{u}))$ if $\mathbf{u}$ are the marginal distribution functions $u_i = F_i(y_i)$.
- $\mathbb{K}(\mathbf{y}^*, \mathbf{s}) = \sum \mathbb{K}({y^*}_i, s_i)$, since both $\mathbf{y}^*, \mathbf{s}$ have independent entries.

Copula Based Independent Component Analysis (CICA)

Under fixed assumptions about the distribution of the sources, I have to minimize two terms: the true objective, the mutual information, expressed via the copula; the mismatch of the marginal distributions to the assumed distributions.

Copula Based Component Analysis - Summary

To cast the K-L partitioning in terms of the copula I write the independence term as

$$\min_B \mathbf{MI}(\mathbf{y}; B) = \min_B \mathbb{E}_q(log(dC_\Theta(\mathbf{u})))$$

and the marginal fit term as

$$\min_\Theta [C_\Theta(\mathbf{u}) - \prod_{i=1}^{k}(u_i)].$$

That is, I minimize the mutual information via the copula via rotation $B = \hat{A}^{-1}$ after minimizing the distance between parametric copula and independent marginals.

Copula Based Component Analysis - Overview

Setting $\mathbf{u}^* = G(\mathbf{y}^*)$ where $\mathbf{y}^*$ is still a random, mutually independent vector with margins distributed equivalently with $\mathbf{y}$. Thus, $\mathbf{u}^*$ is independent with margins distributed as $\mathbf{y}$.

$$\mathbb{K}(\hat{\mathbf{u}}, \mathbf{u}) = \mathbb{K}(\hat{\mathbf{u}}, \mathbf{u}^*) + \mathbb{K}(\mathbf{u}^*, \hat{\mathbf{u}})$$

with $\hat{\mathbf{u}}$ the estimate of the true sources output from a copula based procedure and $\mathbf{u}$ the true distribution of the sources. The K-L distance between the outputs and the sources is then: (1) the fit of the outputs to independence $\mathbb{K}(\hat{\mathbf{u}}, \mathbf{u}^*)$; and (2) the fit of the marginals of the outputs to the true source distributions.

Copula Based Component Analysis - Full Model

Minimizing the mutual information of the outputs... (with the distributional assumption either fixed or parameterized) in the copula representation:

$$\min_B \mathbf{MI}(\mathbf{y}; B) = \min_B \mathbb{E}_{\hat{q}}(log(dC_\Theta(\mathbf{u})))$$

Copula Based Component Analysis - Full Model

...is equivalent to maximizing the score

$$\frac{\partial L}{\partial B} = -\frac{\partial}{\partial B}\mathbb{K}(q(\cdot), \hat{q}(\cdot, B))$$

via the marginal distributions

$$\frac{\partial L}{\partial B} = -\frac{\partial}{\partial B}\mathbb{K}(\hat{\mathbf{u}}, \mathbf{u})$$

using the copula model.

Copula Based Component Analysis - Full Model

The mixing matrix estimate B can be recovered via gradient ascent/descent, for example.
Versions where the joint dependence is captured in a single parameter (multivariate or scalar) may not be applicable for the ICA problem. If $\Theta^{\perp}$ is the copula parameter at independence then $\lim_{\Theta \to \Theta^{\perp}} C_{\Theta}(\mathbf{u}) = \prod_i^k u_i$ and the mixing matrix at $\Theta^{\perp}$ is unidentifiable.

Copula Based Component Analysis - Partite Model

Another approach is to set $\mathbf{y} = RW\mathbf{x}$, with W a 'whitening' matrix - the product of PCA - and R the ICA rotation. This allows orthogonalization of a mutual information matrix via well known procedures like Singular Value Decomposition.

- Construct scatter/kernel matrix $\Gamma_{C_\Theta} = ((C_{\theta_{ij}}(\hat{F}^n_{w_i{}^T\mathbf{x}}(w_i T\mathbf{x}), \hat{F}^n_{w_j{}^T\mathbf{x}}(w_j{}^T\mathbf{x}))))_{i,j=1..k}$
- Find orthogonalization of Γ_{C_Θ}, $\lambda_1, ..., \lambda_k$
- Yield $y_k = b_k\mathbf{x}_k = r_k w_k x_k$ with $y_i \perp y_j$ via C_Θ

Copula Based Component Analysis - Partite Model

Another approach is to set $\mathbf{y} = RW\mathbf{x}$, with W a 'whitening' matrix - the product of PCA - and R the ICA rotation. This allows orthogonalization of a mutual information matrix via well known procedures like Singular Value Decomposition.

- Construct scatter/kernel matrix $\Gamma_{C_\Theta} = ((C_{\theta_{ij}}(\hat{F}^n_{w_i{}^T\mathbf{x}}(w_i T\mathbf{x}), \hat{F}^n_{w_j{}^T\mathbf{x}}(w_j{}^T\mathbf{x}))))_{i,j=1..k}$
- Find orthogonalization of Γ_{C_Θ}, $\lambda_1, ..., \lambda_k$
- Yield $y_k = b_k\mathbf{x}_k = r_k w_k x_k$ with $y_i \perp y_j$ via C_Θ

Copula Based Component Analysis - Partite Model

Another approach is to set $\mathbf{y} = RW\mathbf{x}$, with W a 'whitening' matrix - the product of PCA - and R the ICA rotation. This allows orthogonalization of a mutual information matrix via well known procedures like Singular Value Decomposition.

- Construct scatter/kernel matrix $\Gamma_{C_\Theta} = ((C_{\theta_{ij}}(\hat{F}^n_{{w_i}^T\mathbf{x}}(w_i T\mathbf{x}), \hat{F}^n_{{w_j}^T\mathbf{x}}({w_j}^T\mathbf{x}))))_{i,j=1..k}$
- Find orthogonalization of Γ_{C_Θ}, $\lambda_1, ..., \lambda_k$
- Yield $y_k = b_k\mathbf{x}_k = r_k w_k x_k$ with $y_i \perp y_j$ via C_Θ

Copula Based Component Analysis - Partite Model

- Choose exchangeable families at each bivariate pair.
- Treat the univariate distributions $u_i = F_{X_i}(x_i)$ as observed.
- Bivariate Mutual Information(s), or, $E(log(dC(u_i, v_i)))$ are the elements of 'scatter' matrix

Copula Based Component Analysis - Partite Model

- Choose exchangeable families at each bivariate pair.
- Treat the univariate distributions $u_i = F_{X_i}(x_i)$ as observed.
- Bivariate Mutual Information(s), or, $E(log(dC(u_i, v_i)))$ are the elements of 'scatter' matrix

Copula Based Component Analysis - Partite Model

- Choose exchangeable families at each bivariate pair.
- Treat the univariate distributions $u_i = F_{X_i}(x_i)$ as observed.
- Bivariate Mutual Information(s), or, $E(log(dC(u_i, v_i)))$ are the elements of ‘scatter’ matrix

Copula Based Component Analysis - Partite Model

- For well fit $C_{\hat{\theta}_{ij}} \to MI(C_{\hat{\theta}_{ij}}) \geq 0$ for all i, j. Thus
- $\mathbf{R} = ((MI(C_{\hat{\theta}_{ij}})))$ is positive semi-definite, by exchangeability.
- Singular Value Decomposition of **R** yields orthogonal basis (w.r.t MI) [Tipping and Bishop 1999].

Copula Based Component Analysis - Partite Model

- For well fit $C_{\hat{\theta}_{ij}} \rightarrow MI(C_{\hat{\theta}_{ij}}) \geq 0$ for all i, j. Thus
- $\mathbf{R} = ((MI(C_{\hat{\theta}_{ij}})))$ is positive semi-definite, by exchangeability.
- Singular Value Decomposition of **R** yields orthogonal basis (w.r.t MI) [Tipping and Bishop 1999].

Copula Based Component Analysis - Partite Model

- For well fit $C_{\hat{\theta}_{ij}} \rightarrow MI(C_{\hat{\theta}_{ij}}) \geq 0$ for all i, j. Thus
- $\mathbf{R} = ((MI(C_{\hat{\theta}_{ij}})))$ is positive semi-definite, by exchangeability.
- Singular Value Decomposition of **R** yields orthogonal basis (w.r.t MI) [Tipping and Bishop 1999].

Copula Based Component Analysis - Partite Model

- Some ICA algorithms employ “arbitrary” non-linear transforms (e.g. logistic [Bell and Sejnowski 1995], $log(cosh)$ [Teschendorf 2004]).
- Other algorithms use nonparametric estimates of MI [Stogbauer 2004] or cumulant moments [Comon 1994].

Copula Based Component Analysis - Partite Model

- Some ICA algorithms employ “arbitrary” non-linear transforms (e.g. logistic [Bell and Sejnowski 1995], $log(cosh)$ [Teschendorf 2004]).
- Other algorithms use nonparametric estimates of MI [Stogbauer 2004] or cumulant moments [Comon 1994].

Copula Based Component Analysis - following 3, Comparisons

- Copula approach allows flexible choice of non-linear **u**.
- Parametric estimators of MI ($O(n^{-1/2})$) superior to nonparametric version ($O(n^{-4/5})$).
- Parametric approach may be more stable on smaller datasets and under moment perturbation [McCullagh 1994, Everson and Roberts 2000]; and on extreme value distributions.

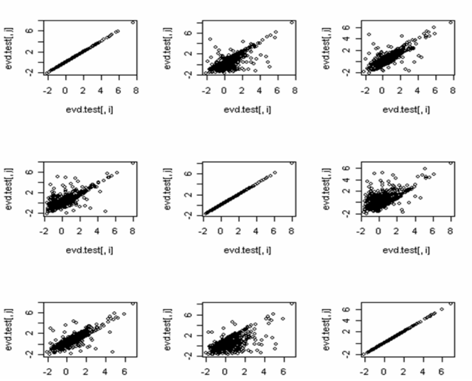
Copula Based Component Analysis - following 3, Comparisons

- Copula approach allows flexible choice of non-linear **u**.
- Parametric estimators of MI ($O(n^{-1/2})$) superior to nonparametric version ($O(n^{-4/5})$).
- Parametric approach may be more stable on smaller datasets and under moment perturbation [McCullagh 1994, Everson and Roberts 2000]; and on extreme value distributions.

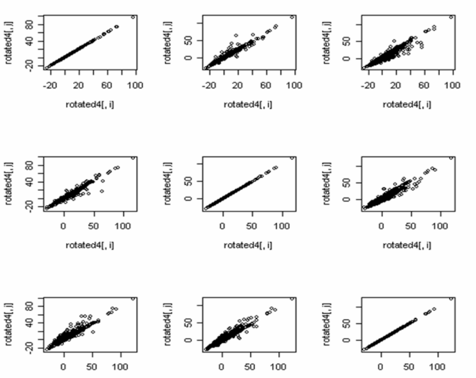
Copula Based Component Analysis - following 3, Comparisons

- Copula approach allows flexible choice of non-linear **u**.
- Parametric estimators of MI ($O(n^{-1/2})$) superior to nonparametric version ($O(n^{-4/5})$).
- Parametric approach may be more stable on smaller datasets and under moment perturbation [McCullagh 1994, Everson and Roberts 2000]; and on extreme value distributions.

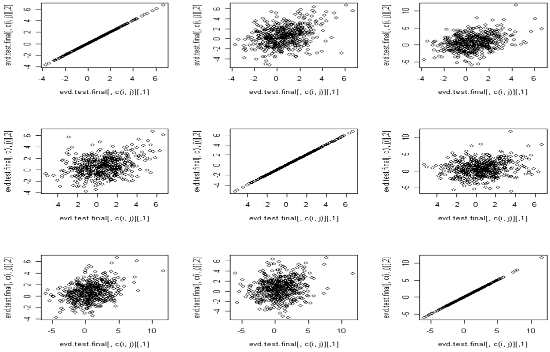
Example Results 1 - Extreme value distribution.



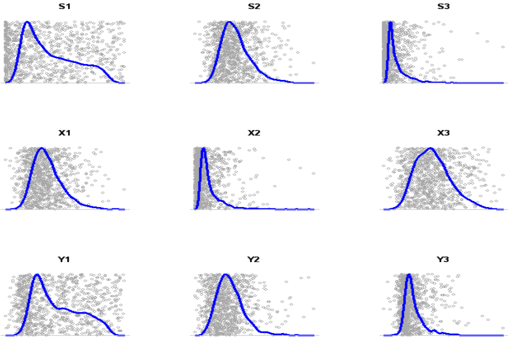
Example Results 1 - Rotated extreme value distribution.



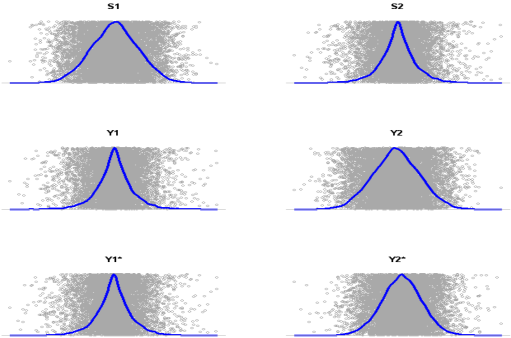
Example Results 1; Gumbel-Hougard (GH) type copula - $y_i = B_i x_i$



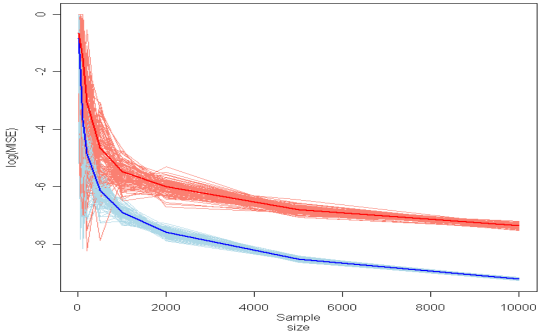
Example Results 2 - GH type dependency gradient.



Example Results 3 - CICA vs. fastICA



Sample Size Comparison, log(MISE) - CICA(GH) vs. fastICA



Partite Example

- Finding closed form MLE estimators is difficult for many models.
- Balance between tractable MLE estimation and model flexibility.

Partite Example

- Finding closed form MLE estimators is difficult for many models.
- Balance between tractable MLE estimation and model flexibility.

Two Parameter Archimedean Families

A copula family is called *archimedean* if the copula can be written:

$$C_\delta(u, v) = \phi_\delta(\phi_\delta^{-1}(u) + \phi_\delta^{-1}(v))$$

with ϕ_δ a generating function parameterized by δ.

Two Parameter Archimedean Families

Then

$$C_{\theta,\delta}(u,v) = \eta_{\theta,\delta}(\eta_{\theta,\delta}^{-1}(u) + \eta_{\theta,\delta}^{-1}(v))$$

with $\eta_{\theta,\delta}(s) = \psi_{\theta}(-log\phi_{\delta}(s))$ is a natural two-parameter extension

Application: CICA on ESI

CICA on the 2002 Environmental Sustainability Index (ESI)

- A scaled linear combination of sixty-eight metrics of environmental concern
- Traditional quantities (such as NO_x and SO_2 concentrations) are included with more expansive measures of environmental sustainability -(such as civil liberty and corruption)
- A measure of overall progress towards environmental sustainability - designed to permit systematic and quantitative comparison between nations

Application: CICA on ESI

CICA on the 2002 Environmental Sustainability Index (ESI)

- A scaled linear combination of sixty-eight metrics of environmental concern
- Traditional quantities (such as NO_x and SO_2 concentrations) are included with more expansive measures of environmental sustainability -(such as civil liberty and corruption)
- A measure of overall progress towards environmental sustainability - designed to permit systematic and quantitative comparison between nations

Application: CICA on ESI

CICA on the 2002 Environmental Sustainability Index (ESI)

- A scaled linear combination of sixty-eight metrics of environmental concern
- Traditional quantities (such as NO_x and SO_2 concentrations) are included with more expansive measures of environmental sustainability -(such as civil liberty and corruption)
- A measure of overall progress towards environmental sustainability - designed to permit systematic and quantitative comparison between nations

ESI Grouping

- **Environmental Systems** Natural stocks such as air, soil, and water.
- **Environmental Stresses** Stress on ecosystems such as pollution and deforestation.
- **Vulnerability** Basic needs such as health, nutrition, and mortality.
- **Capacity** Social and economic variables such as corruption and liberty, energy consumption, and schooling rate.
- **Stewardship** Global cooperation such as treaty participation and compliance.

ESI Grouping

Environmental Systems Natural stocks such as air, soil, and water.

Environmental Stresses Stress on ecosystems such as pollution and deforestation.

Vulnerability Basic needs such as health, nutrition, and mortality.

Capacity Social and economic variables such as corruption and liberty, energy consumption, and schooling rate.

Stewardship Global cooperation such as treaty participation and compliance.

ESI Grouping

- **Environmental Systems** Natural stocks such as air, soil, and water.
- **Environmental Stresses** Stress on ecosystems such as pollution and deforestation.
- **Vulnerability** Basic needs such as health, nutrition, and mortality.
- **Capacity** Social and economic variables such as corruption and liberty, energy consumption, and schooling rate.
- **Stewardship** Global cooperation such as treaty participation and compliance.

ESI Grouping

Environmental Systems Natural stocks such as air, soil, and water.

Environmental Stresses Stress on ecosystems such as pollution and deforestation.

Vulnerability Basic needs such as health, nutrition, and mortality.

Capacity Social and economic variables such as corruption and liberty, energy consumption, and schooling rate.

Stewardship Global cooperation such as treaty participation and compliance.

ESI Grouping

Environmental Systems Natural stocks such as air, soil, and water.

Environmental Stresses Stress on ecosystems such as pollution and deforestation.

Vulnerability Basic needs such as health, nutrition, and mortality.

Capacity Social and economic variables such as corruption and liberty, energy consumption, and schooling rate.

Stewardship Global cooperation such as treaty participation and compliance.

The complete data ESI

Environmental Sustainability Index 2002

Partite Approach

The approach:

- Exploit the empirical distribution, setting $u_i = \hat{F}_i^n(x_i)$ and treating the univariate marginals as observed or unparameterized.
- Fit copulas pairwise and minimize $\mathbf{MI}(\mathbf{u})$ by diagonalization of a *mutual information matrix*.

Partite Approach

The approach:

- Exploit the empirical distribution, setting $u_i = \hat{F}_i^n(x_i)$ and treating the univariate marginals as observed or unparameterized.
- Fit copulas pairwise and minimize $\mathbf{MI}(\mathbf{u})$ by diagonalization of a *mutual information matrix*.

Implementation

- Fit bivariate copula
- The candidate copulae at two-parameter extensions of *Archimedean* type copulae.

Implementation

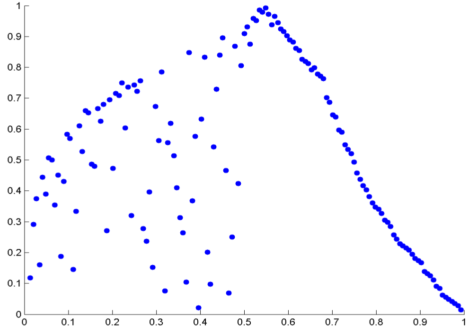
- Fit bivariate copula
- The candidate copulae at two-parameter extensions of *Archimedean* type copulae.

Implementation

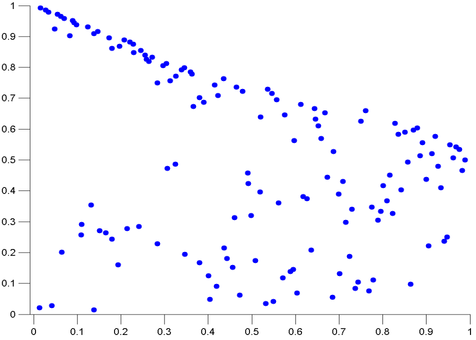
Additionally, I fit rotations of bivariate copula.
The copula are rotated by setting the argument equal to the values in the table, with $\overline{u} = 1 - u, \overline{v} = 1 - v$:

Rotation	(u, v)
0	(u, v)
90	$(\overline{v}, u)$
180	$(\overline{v}, \overline{u})$
270	$(v, \overline{u})$

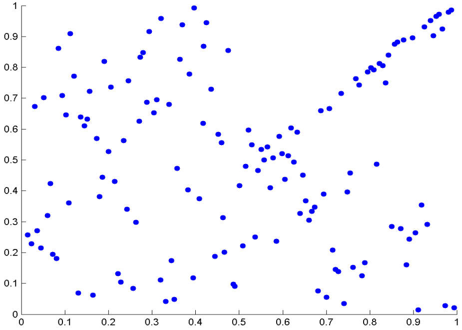
Bivariate Scatterplot - PCA_1 vs. PCA_2



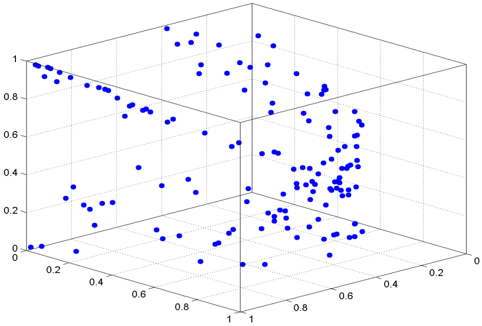
Bivariate Scatterplot - PCA_2 vs. PCA_3



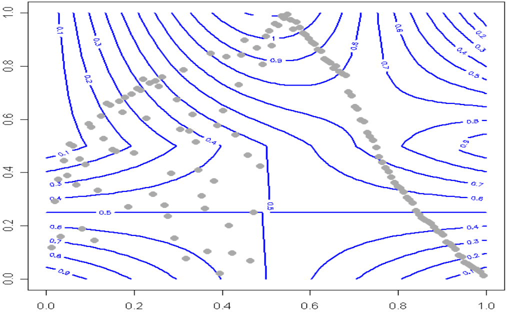
Bivariate Scatterplot - PCA_1 vs. PCA_3



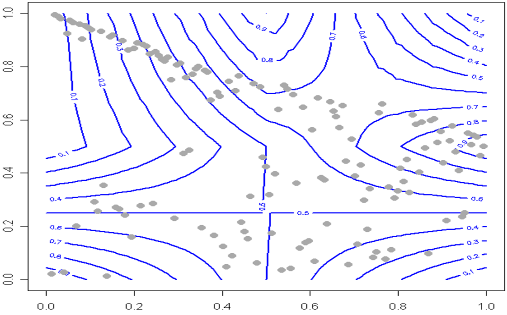
Trivariate Scatterplot - PCA_1 vs. PCA_2 vs PCA_3



Contour/Scatter plot: Copula on $CICA_1$ vs. $CICA_2$



Contour/Scatter plot: Copula on $CICA_2$ vs. $CICA_3$



Contour/Scatter plot: Copula on $CICA_1$ vs. $CICA_3$

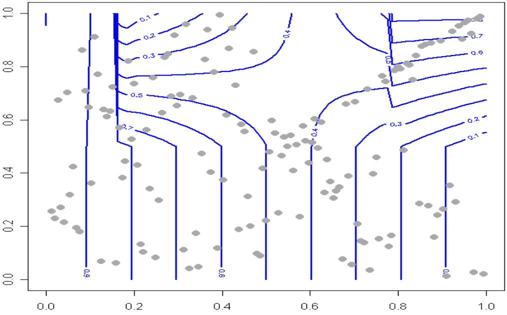
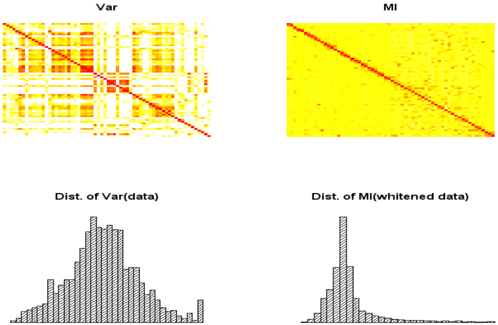
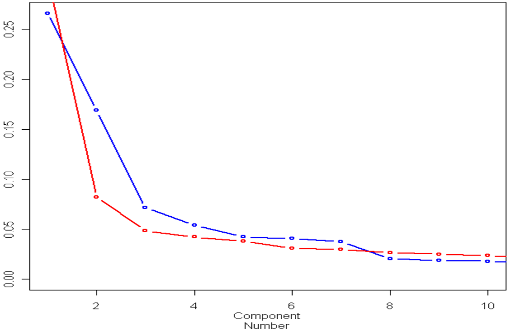


Image plots: Covariance vs. MI



Scree plot: PCA and CICA



CICA loadings

Variable Name	Component 1	Component 2	Component 3
SO2	1	33	54
NO2	2	24	42
TSP	3	16	33
ISO14	4	49	41
WATCAP	5	43	35
IUCN	6	23	25
CO2GDP	7	52	61

CICA loadings vs. PCA loadings

CICA	PCA
SO2	NUKE
NO2	BODWAT
TSP	TFR
ISO14	FSHCAT
WATCAP	PESTHA
IUCN	WATSUP
CO2GDP	GRAFT

ESI Grouping

Environmental Systems Natural stocks such as air, soil, and water.

Environmental Stresses Stress on ecosystems such as pollution and deforestation.

Vulnerability Basic needs such as health, nutrition, and mortality.

Capacity Social and economic variables such as corruption and liberty, energy consumption, and schooling rate.

Stewardship Global cooperation such as treaty participation and compliance.

ESI Grouping

Environmental Systems Natural stocks such as air, soil, and water.

Environmental Stresses Stress on ecosystems such as pollution and deforestation.

Vulnerability Basic needs such as health, nutrition, and mortality.

Capacity Social and economic variables such as corruption and liberty, energy consumption, and schooling rate.

Stewardship Global cooperation such as treaty participation and compliance.

ESI Grouping

- **Environmental Systems** Natural stocks such as air, soil, and water.
- **Environmental Stresses** Stress on ecosystems such as pollution and deforestation.
- **Vulnerability** Basic needs such as health, nutrition, and mortality.
- **Capacity** Social and economic variables such as corruption and liberty, energy consumption, and schooling rate.
- **Stewardship** Global cooperation such as treaty participation and compliance.

ESI Grouping

Environmental Systems Natural stocks such as air, soil, and water.

Environmental Stresses Stress on ecosystems such as pollution and deforestation.

Vulnerability Basic needs such as health, nutrition, and mortality.

Capacity Social and economic variables such as corruption and liberty, energy consumption, and schooling rate.

Stewardship Global cooperation such as treaty participation and compliance.

ESI Grouping

Environmental Systems Natural stocks such as air, soil, and water.

Environmental Stresses Stress on ecosystems such as pollution and deforestation.

Vulnerability Basic needs such as health, nutrition, and mortality.

Capacity Social and economic variables such as corruption and liberty, energy consumption, and schooling rate.

Stewardship Global cooperation such as treaty participation and compliance.

Outline

Comments

- Copula approach unifies PCA/ICA as likelihood based approach
- Natural first step of generalized (linear) Copula based models
- Exploits the rich family of lower dimension copulae

Comments

- Copula approach unifies PCA/ICA as likelihood based approach
- Natural first step of generalized (linear) Copula based models
- Exploits the rich family of lower dimension copulae

Comments

- Copula approach unifies PCA/ICA as likelihood based approach
- Natural first step of generalized (linear) Copula based models
- Exploits the rich family of lower dimension copulae

Next

- Bayesian/Unified approach *across* dependence families.
- 'Copula Similar' functions to recast GLMs on distributions with fixed margins.
- Closed form estimators for *k* order dependence (*within* dependence family).

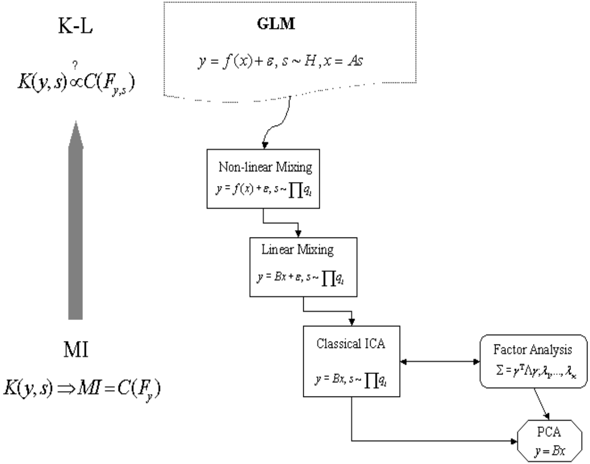
Next

- Bayesian/Unified approach *across* dependence families.
- ‘Copula Similar’ functions to recast GLMs on distributions with fixed margins.
- Closed form estimators for k order dependence (*within* dependence family).

Next

- Bayesian/Unified approach *across* dependence families.
- 'Copula Similar' functions to recast GLMs on distributions with fixed margins.
- Closed form estimators for k order dependence (*within* dependence family).

Comments, Next Steps



Comments, Next Steps