UNIVERSITÀ DEGLI STUDI DI UDINE
hic sunt futura
Department of Economics and Statistics

Discussion on **Dependencies in higher dimensions and in complex data structures** by Irène Gijbels

Paolo Vidoni

Department of Economics and Statistics
University of Udine - Italy

paolo.vidoni@uniud.it

Statistical methods and models for complex data
800 years of research to understand a complex world

Department of Statistical Sciences, University of Padova
21-23 September 2022

Introduction

The the main topic of the presentation is the **measurement of the association between the components of a random vector**, in particular when

- the dimension of the random vector is large (and, potentially, growing to infinity)
- dependencies and data structures are complex

The **copula-based approach** is considered, so that the bivariate association measures can be generalized to suitable multivariate association measures.

- axiomatic perspective
- non-parametric solutions
- multivariate association measures for the increasing dimension case

An alternative approach based on divergences

The idea is to measure the stochastic dependence between the components of a random vector by considering the **(pseudo-)distance** between the (estimated) joint distribution and the (estimated) distribution in case of independence.

Let us consider:

- a continuous random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$, with $d \geq 1$, following an **unknown density function** $f(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbf{R}^d$
- the marginal density $f_i(x_i)$ of X_i, $i = 1, \ldots, d$, so that $f_0(\boldsymbol{x}) = \prod_{i=1}^d f_i(x_i)$ is the density representing the independence case

A **(multivariate) association measure** can be defined as

$$D(\widehat{f}, \widehat{f_0})$$

where $D(\cdot, \cdot)$ is a suitable divergence and $\widehat{f} = \widehat{f}(\boldsymbol{x})$ and $\widehat{f_0} = \widehat{f_0}(\boldsymbol{x})$ are appropriate estimates for $f(\boldsymbol{x})$ and $f_0(\boldsymbol{x})$, based on the observed sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, $n \geq 1$.

UNIVERSITÀ DEGLI STUDI DI UDINE
hic sunt futura
Department of Economics and Statistics

Some difficulties

The true data generating model f is unknown and to specify a useful statistical model can be difficult: data may exhibit complex structures and dependencies.

An approach, alternative to copulas, is based on the idea of **model combination**.

$f(\boldsymbol{x})$ can be surrogated by a combination of basic density functions $p_j(\boldsymbol{x};\theta_j)$, $j = 1,\ldots,J$, which are simple probability models for $\boldsymbol{X}$, describing particular features of the interest random phenomenon.

One possibility is to specify a **multiplicative model combination**

$$f_p(\boldsymbol{x}; w, \theta) = c(w,\theta)^{-1} \prod_{j=1}^{J} p_j(\boldsymbol{x};\theta_j)^{w_j}$$

with $c(w,\theta)$ the **normalizing constant** (supposed to be finite), $w = (w_1,\ldots,w_J)$ a vector of non-negative weights and $\theta = (\theta_1,\ldots,\theta_J)$.

Some difficulties

Estimating the surrogate model $f_p(\boldsymbol{x}; w, \theta)$ may not be easy.

The computation of the normalizing constant $c(w, \theta)$ could be demanding and the dimension of the vector of the unknown parameters (w, θ) could be large.

To face these problems, it could be useful

- to set an inferential procedure based on a convenient scoring rule (e.g. the Hyvärinen score)
- to consider a suitable boosting algorithm, useful in case of high-dimensional optimization problems, when also regularization is required

The choice of the divergence $D(\cdot, \cdot)$

There are many alternative choices (e.g. the ϕ-divergences), that highlight different aspects related to the models under consideration.

UNIVERSITÀ DEGLI STUDI DI UDINE
hic sunt futura
Department of Economics and Statistics

A SIMPLE MOTIVATING APPLICATION

Let us assume that $\boldsymbol{X}$ follows a d-**dimensional Gaussian distribution** with a null mean vector $\mu = 0$ and a non-singular covariance matrix $\Sigma = (\sigma_{rs})$.

It is convenient to consider the *precision matrix* $Q = \Sigma^{-1} = (q_{rs})$ (define the conditional dependence structure of a Gaussian Markov random field).

If $\Sigma_0 = \text{diag}(\sigma_{11}, \ldots, \sigma_{dd})$ is the covariance matrix in case of *independence*, an **association measure** (based on the Kullback-Leibler divergence) can be defined as

$$D(f, f_0) = \frac{1}{2} \log(|\Sigma_0 Q|)$$

Given the observed sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, the aim is to estimate Σ_0 and Q (or, equivalently, Σ).

The maximum likelihood estimator for Σ is not satisfactory, in particular for a large dimension d.

UNIVERSITÀ DEGLI STUDI DI UDINE hic sunt futura
Department of Economics and Statistics

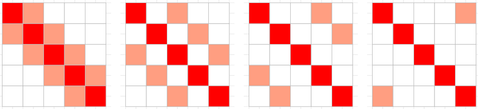
A simple motivating application

Multiplicative combination of d**-variate Gaussian densities** with a suitable (simple) symmetric precision matrix Q_j, $j = 1, \ldots, J$: $f_p(\boldsymbol{x}; w, \theta)$ is a Gaussian density **with precision matrix** $Q_p = \sum_{j=1}^{J} w_j Q_j$.

The *choice of the system of matrices* Q_j, $j = 1, \ldots, J$, is crucial for the effectiveness of the inferential procedure.

If we assume that the **true** Q **is a band matrix** (AR-type dependence), we may consider $J = d - 1$ *component matrices having the main diagonal and only two equal non-null, symmetric diagonals*.

For example, with $d = 5$,

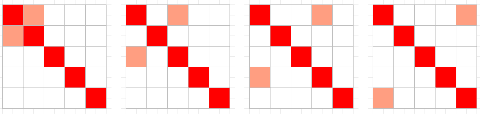


UNIVERSITÀ DEGLI STUDI DI UDINE
hic sunt futura
Department of Economics and Statistics

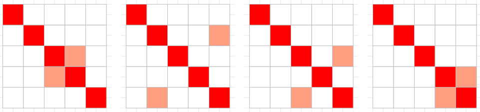
A simple motivating application

If the precision matrix does not follow a band structure: alternative more general *system of* **component pairwise precision matrices**.

For example, with $d = 5$, we may specify an extremely general system of component precision matrices, such as



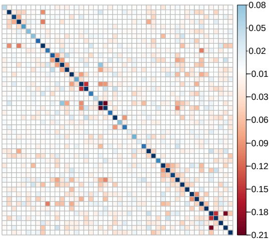
$\ldots \quad \ldots$



Example: returns of US industries portfolios

Monthly returns of 48 US industries portfolios from August 2009 to July 2019 (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/): $d = 48$, $n = 120$.

Estimate of the 48×48 precision matrix using a boosting algorithm (based on the Hyvärinen score) applied to a general multiplicative model combination.



The association measure based on the KL divergence corresponds to -37.48.

UNIVERSITÀ DEGLI STUDI DI UDINE hic sunt futura

Department of Economics and Statistics

References

- Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, **29**, 1189 – 1232.
- Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, **164**, 130 – 141.
- Hinton, G.E. (2002). Training products of experts by minimizing contrastive divergence. *Neural Computation*, **14**, 1771 – 1800.
- Hyvärinen, A. (2007). Some extensions of score matching. *Computational Statistics and Data Analysis*, **51**, 2499 – 2512.
- Micheas, A.C. and Zografos, K. (2006).Measuring stochastic dependence using ϕ-divergence. *Journal of Multivariate Analysis*, **97**, 765 – 784.
- Vidoni, P. (2021). Boosting multiplicative model combination. *Scandinavian Journal of Statistics*, **48**, 761 – 789.