Non-Linear Factor Selection and Copula of Copulas

Ulrich Müller-Funk, joint work with Kay Hildebrand

Introduction

Approaches

Tools

- Analyst's attitude depending on the number of
  - observations at hand, say N
  - factors available, say M
  - problem complexity

  For M small, $M \ll N$ and moderate model complexity:

  „all-in"–approach – no dimension reduction

- Phases: Factor selection is possibly part of
  - **model finding process**
  - model selection – via regularization,
  - model evaluation – via tests or information criteria.

- Aspects
  - Relevance i.e. no classification bias
  - Minimality i.e. little noise only
- Quality („risk“) assessment by means of some real-valued index
  e.g. MIS (classification) or ESS (clustering)
- Constraint optimization problem—but not formulated that way
- In what follows: non-parametric setting, data mining

- Characteristics:
  - *N* large, *M* possibly too
  - Attributes measured on different scales
  - Data base heterogeneous
  - Data mostly results from business operations
- Consequences:
  - Modeling by means of general finite mixture distributions
  - Linear / likelihood-based methods largely obsolete
  - **Necessity of a process model for factor selection**
    $\hookrightarrow$ cost considerations $\hookrightarrow$ (semi-) automized algorithm
  - Importance of data-preprocessing (data quality)

Introduction

Approaches

Tools

- Factor selection equals measuring dependencies
  - Unconditionally: $X \leftrightarrow Y$, $X_1 \leftrightarrow X_2, \ldots$
  - Conditionally: $Y \leftrightarrow X_1 | X_2$, $X_1 \rightarrow X_2 | X_3, \ldots$

  where Y is a target variable (classification) and
  X_1, X_2, X_3 are disjoint blocks of explanatory/predictive variables

- Recognition and measurement
  - Complete: Copulas and their partial derivatives
  - Numerically: Indices e.g. **correlation**, association, PRE

- Statistics
  - Thinning by means of correlation indices, simultaneous testing
  - Aggregation: attribute clustering
  - Orthogonal projection: projection pursuit and exploratory PLS or factor analysis $\hookrightarrow$ artificial constructs
  - Matching: MDS, homogeneity analysis, correspondence analysis $\hookrightarrow$ latent constructs
- Machine Learning (for nominal classification problems)
  - Filters / Wrappers $\hookrightarrow$ Boolean analysis
  - PRE measures (in statistical parlance)

  Cf. Hall (1999) for an overview

- Originally designed for special situations, controlled experiments in medicine, agriculture, ...
- Applicable to linear, stationary and Gaussian models and methods
- Main techniques: PCA, CCA, PLS (-regression), partial correlation—all based on covariances respectively Pearson correlations
- Wanted: Corresponding techniques for non-linear analysis

Introduction

Approaches

Tools

- Hoeffding–Sklar–Factorization of a df F via copula C_F:

$$F(x_1, \dots, x_M) = C_F(F_1(x_1), \dots, F_M(x_M))$$

  where F_i are the univariate marginal df

- Frechet Bounds: $C_- \leq C_F \leq C_+$
- Conditional copulas:
  $\mathcal{L}(X_1, X_2 | X_3)$ has copula $D_3 C_F(t_1, t_2, t_3), F_3(X_3)$
  where D_3 is/are partial derivative/s
- Extension of the following analysis from the unconditional to the conditional case

Multiple Correlation Coefficients

- Functional form: L suitable „loss function".

$$\gamma(F) = \text{const} * \text{dist}(C_F, C_R), \quad C_R \text{ reference copula}$$
$$= \text{const} * \int_{[0,1]^M} L(C_F - C_R)\,dC_A, \quad C_A \text{ averaging copula}$$

- Examples (cf. Schmid et al. 2010):
  - Spearman, Fechner-Kendall, Blomqvist (signed, $C_R = C_0$)
  - Gini, Spearman's Footrule (signed, $C_R = \frac{1}{2}(C_+ + C_-)$)
  - Hoeffding, Schweizer-Wolff (unsigned, $C_R = C_0$)
- For $M > 2$, C_0 is „no longer the midpoint" of C_+ and C_-
  Cf. Wolff (1980)

- Idea: Choose C_+ as a reference
- First version:
  - $c_M = (M+1)/(1-\frac{1}{M!})$
  - $\delta_+(F) = c_M * \int\limits_{[0,1]^M} (C_+(u) - C_F(u))du \in [0,1]$
  - For $C_F = C_0$ we get $v_M = \gamma_+(C_0) = c_M(\frac{1}{M+1} - \frac{1}{2^M})$
  - Choose any $q_M : [0,1] \to [-1,1]$ that is antitonic, concave and satisfies $q_M(0) = 1$, $q_M(v_M) = 0$, $q_M(1) = -1$
- Define $\gamma_+(F) = q_M(\delta_+(F))$

Empirical Analysis: Convergence

- based on an M-dimensional i.i.d. sample with parent df F, emp. df $\hat{F}_N$, empirical copula $\hat{C}_N$
- Empirical measures $\hat{\gamma}_N = \gamma(\hat{F}_N)$. Validity of
  - SLLN: $\hat{\gamma}_N \to \gamma(\hat{F}_N)$ F a. s.,
  - CLT: $\mathcal{L}(\sqrt{N}(\hat{\gamma}_N - \gamma(F)) = AN(0, \sigma^2(C, q_M)))$
- Basic functional limit theorem, cf. Rüschendorf (1976)

  $$\sqrt{N}(\hat{C}_N(u) - C_F(u)) \longrightarrow_w B_C(u) - \sum_{m=1}^{M} D_m C_F(u) B_C^{(m)}(u_m),$$

  B_C tied down Brownian sheet with intensity C, **continuous** partial derivatives $D_m C_F$

Correlations between Blocks of Factors

- Blocks of factors—allowing for different block sizes via $\gamma(F)$
- $\mathcal{I}$ and $\mathcal{J}$ disjoint sets of factors, copulas $C_F(\cdot|\mathcal{I})$, $C_F(\cdot|\mathcal{J})$. Leads to inequality:

$$\gamma(C_F(\cdot|\mathcal{I}) + C_F(\cdot|\mathcal{J}) - 1)^+) \leq \gamma(C_F(\cdot|\mathcal{I} \cup \mathcal{J})) \leq \gamma(\min(C_F(\cdot|\mathcal{I}), C_F(\cdot|\mathcal{J})))$$

- Motivated by that inequality, we put

$$\alpha(\mathcal{I}, \mathcal{J}) = \max\left(\frac{\gamma(C_F(\cdot|\mathcal{I})+C_F(\cdot|\mathcal{J})-1)^+)}{\gamma(C_F(\cdot|\mathcal{I}\cup\mathcal{J}))}, \frac{\gamma(C_F(\cdot|\mathcal{I}\cup\mathcal{J}))}{\gamma(\min(C_F(\cdot|\mathcal{I}),C_F(\cdot|\mathcal{J})))}\right)$$

- $1 - \alpha(\mathcal{I},\mathcal{J})$ is a measure of dissimilarity $\hookrightarrow$ attribute clustering.

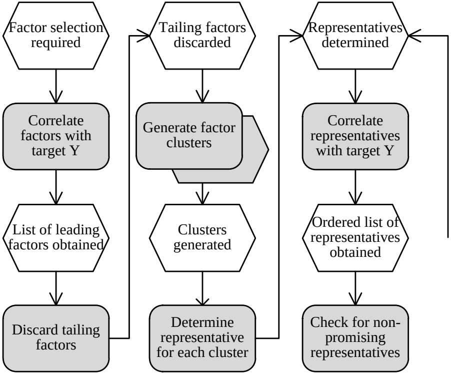
Non-Linear Factor Selection and Copula of Copulas

Ulrich Müller-Funk, joint work with Kay Hildebrand

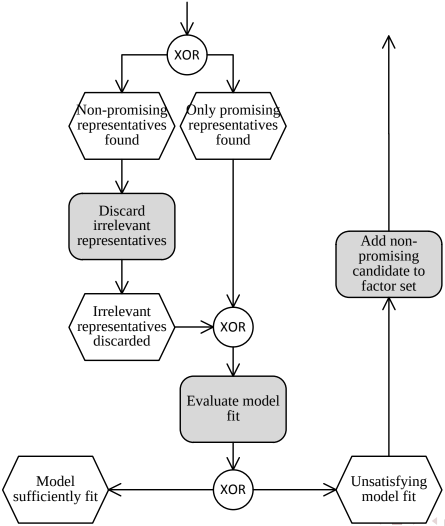
- Tools
  - PRE measures for metric variables — beyond R^2
  - Mixed scales — beyond binning
  - Variants of γ_+
- Empirical analysis concerning
  - Process model
  - Tools

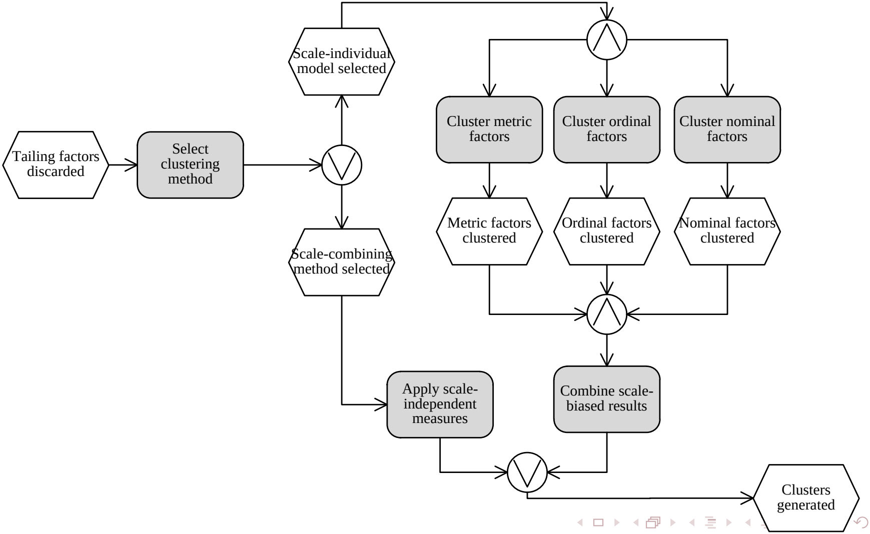
  $\hookrightarrow$ data problem
- Foundation: Variant of the basic functional limit theorem not requiring continuity of partial derivatives

Process Model – Main Process



Process Model – Main Process




Tailing factors discarded
Select clustering method
Scale-individual model selected
Scale-combining method selected
Cluster metric factors
Cluster ordinal factors
Cluster nominal factors
Metric factors clustered
Ordinal factors clustered
Nominal factors clustered
Apply scale-independent measures
Combine scale-biased results
Clusters generated