Latentcor: An R Package for Latent Correlation Estimation

Mingze Huang[1, 2], **Irina Gaynanova**[2], **and Christian L. Muller**[1]

1 Department of Statistics, Texas A& M University **2** Department of Economics, Texas A& M University **3** Department of Statistics, University of Munich

DOI:

Software

- Review
- Repository
- Archive

Submitted:
Published:

Summary

The R package *latentcor* provides estimation for latent correlation with mixed data types (continuous, binary, truncated and ternary). Comparing to *MixedCCA*, which estimates latent correlation for canonical correlation analysis, our new package provides a standalone version for latent correlation estimation. Also we add new functionality for latent correlation between ternary/continous, ternary/binary, ternary/truncated and ternary/ternary cases.

Compare to MixedCCA, memory footprint.

Statement of need

Currently there is no standalone package dealing with latent correlation for mixed data type like we did in *latentcor*. The R package *stats* (Team & others, 2013) have some functionality to calculate different type of correlations (Pearson, Kendall and Spearman). The R package *pcaPP* (Croux, Filzmoser, & Fritz, 2013) provides a fast calculation for Kendall's τ. The R package *MixedCCA* (Yoon, Carroll, & Gaynanova, 2020) have functionality for latent correlation estimation as an intermediate step for canonical correlation analysis on mixed data. No package deal with latent correlation across mixed data type.

Usage

	Type continuous
continuous	Liu, Lafferty, & Wasserman (2009)
binary	Fan, Liu, Ning, & Zou (2017)
truncated	Yoon, Carroll, & Gaynanova (2020)
ternary	Quan, Booth, & Wells (2018)

Definition 1 Fan et al. (2017) considered the problem of estimating Σ for the latent Gaussian copula model based on Kendall's τ. Given the observed data $(X_{1j}, X_{1k}), ..., (X_{nj}, X_{nk})$ for variables X_j and X_k, Kendall's τ is defined as

$$\hat{\tau}_{jk} = \frac{2}{n(n-1)} \sum_{1 \leq i < i' \leq n} sign(X_{ij} - X_{i'j}) sign(X_{ik} - X_{i'k})$$

Theorem 1 Let $W_1 \in \mathcal{R}\surd^{\infty}$, $W_2 \in \mathcal{R}\surd^{\in}$, $W_3 \in \mathcal{R}\surd^{\ni}$, $W_4 \in \mathcal{R}\surd^{\triangle}$ be such that $W = (W_1, W_2, W_3, W_4) \sim NPN(0, \Sigma, f)$ with $p = p_1 + p_2 + p_3 + p_4$. Let $X = (X_1, X_2, X_3, X_4) \in \mathcal{R}\surd$ satisfy $X_j = W_j$ for $j = 1, ..., p_1$, $X_j = I(W_j > c_j)$ for $j = p_1 + 1, ..., p_1 + p_2$,