Inference on the Second Moment Structure of High-Dimensional Sensor-Type Data in a K - Sample Setting

Nils Mause [1]

In this talk we consider K high-dimensional vector time series $\mathbf{Y}_{T,1}, \ldots, \mathbf{Y}_{T,K}$ generated by independently working stations $j = 1, \ldots, K$ of sensors within a time interval $[0, T]$ for some $T > 0$. Here the sensors located in different stations may collect and transmit data at distinct sampling frequencies $\omega_j \in [0, 1]$, such that the resulting sample sizes given by $N_j = \lfloor \omega_j T \rfloor$ may also be different.
Projections $\mathbf{w}_T' \mathbf{Y}_{T,j}$ based on such high-dimensional time series with some appropriate weighting vector $\mathbf{w}_T$ appear naturally in many statistical procedures and applications, such as the Principal Component Analysis and Portfolio Optimization, and are a common method of dealing with high-dimensional data sets. Taking the data of all K stations into account, quadratic forms (and more generally bilinear forms) based on the pooled sample variance-covariance matrix need to be analyzed in order to draw inference on the variance of these projections.
Within the high-dimensional framework where not only the time horizon T shall go to infinity but also the dimension d_T of the data is allowed to grow with the time horizon we establish a new strong approximation result for the càdlàg process

$$\mathcal{D}_T(t_1, \ldots, t_K) = \frac{1}{\sum_{l=1}^K N_l} \boldsymbol{v}_T' \left(S_{T, \lfloor t_1 N_1 \rfloor, \ldots, \lfloor t_K N_K \rfloor} - E(S_{T, \lfloor t_1 N_1 \rfloor, \ldots, \lfloor t_K N_K \rfloor}) \right) \boldsymbol{w}_T,$$

for $t_1, \ldots, t_K \in [0, 1]$, where $S_{T,k_1,\ldots,k_K} = \sum_{j=1}^K \sum_{i=1}^{k_j} \boldsymbol{Y}_{T,j} \boldsymbol{Y}_{T,j}', k_j \leq N_j, j = 1, \ldots, K$, and $\boldsymbol{v}_T, \boldsymbol{w}_T$ are uniformly ℓ_1 - bounded weighting vectors. In particular, without any constraints on the ratio of dimension and sample sizes, we receive the strong approximation

$$\sup_{t_1, \ldots, t_K \in [0,1]} \left| \mathcal{D}_T(t_1, \ldots, t_K) - \sum_{j=1}^K \alpha_{T,j} B_T^{(j)} \left(\frac{\lfloor t_j N_j \rfloor}{\sum_l N_l} \right) \right| = o(1), \quad T \to \infty,$$

by the scaled sum of independent Brownian motions $\{B_T^{(j)}(t_j) : t_j \in [0, 1]\}$ with variances $\alpha_{T,j}^2, j = 1, \ldots, K$. Moreover, we will show that a similar result also holds for an increasing (and possibly infinite) amount, L_T, of bilinear forms based on the pooled sample variance-covariance matrix. These approximation results are therefore also applicable in situations where the dimension d_T grows much faster

[1] Institute of Statistics, RWTH Aachen University, Germany, E-mail: mause@stochastik.rwth-aachen.de

than the sample sizes N_j and the time horizon T.
As an application of the strong approximation results we will deal with a change in variance problem, where, under the null hypothesis of no change, the data $\mathbf{Y}_{T,j,1}, \ldots, \mathbf{Y}_{T,j,N_j}$ is supposed to form a stationary d_T - dimensional vector time series with mean zero and variance - covariance matrix $\mathbf{\Sigma}_{T,0}^{(j)}$, which can either be known or unknown.
Lastly, we conduct a simulation study in order to illustrate the finite sample performance of our change-point test statistic.

Acknowledgement:
This is joint work with Ansgar Steland from the Institute of Statistics, RWTH Aachen University, Germany. It was supported by the German Research Foundation / Deutsche Forschungsgemeinschaft (DFG) [grant number STE 1034/11-1].

References

[1] Steland A., von Sachs R. *Large sample approximations for variance-covariance matrices of high-dimensional time series.* Bernoulli, 2017, v. 23(4A), p. 2299–2329.

[2] Steland A., von Sachs R. *Asymptotics for high-dimensional covariance matrices and quadratic forms with applications to the trace functional and shrinkage.* Stochastic Process. Appl., to appear.