Mixed high-dimensional copulas for multivariate time series in public health

Date of submission: October 13, 2021

Supervisor:

Bruno Rémillard: Honorary Professor, Department of decision sciences, HEC Montréal, and Adjunct Prof. Department of Mathematics, U. de Sherbrooke.
Web page: Bruno Rémillard
Email: bruno.remillard@hec.ca

Co-supervisor:

Bouchra Nasri: Assistant Professor, Département de médecine sociale et préventive, U. de Montréal.
Web page: Bouchra Nasri
Email: bouchra.nasri@umontreal.ca

Collaborator:

Pavel Krupskiy: Lecturer, School of Mathematics and Statistics, U of Melbourne, Australia.
Web page: Pavel Krupskiy
Email: pavel.krupskiy@unimelb.edu.au

Abstract

The main objective of this research project is to model and analyze multivariate time series with mixed distributions, i.e., discrete or continuous, encountered in health data and environmental data provided by OQACC (Québec Observatory for Adaptation to Climate Change) and INSPQ (Québec's National Institute of Public Health). First, univariate discrete or continuous time series models will be fitted to the data. In the discrete case, randomized conditional Rosenblatt's transforms will be used to define pseudo-observations that are approximately iid uniform, while in the continuous case conditional Rosenblatt's transforms will be used. The goodness-of-fit tests will be constructed from the empirical processes of the pseudo-observations. Next, the pseudo-observations will be used for the inference of time-varying non-central squared copulas, which are new copula families exhibiting sometimes non-monotonic dependence between the components. Also, new time-varying factor copulas families will be constructed, using also the non-central squared copulas. Since the data are high-dimensional, pairwise composite likelihood and recursive methods will be used for the estimation, while pairwise composite goodness-of-fit tests will be developed.

Interdisciplinary/applied experience

This project is motivated by important problems in public health related to climate change issues. Through Dr Bustinza at INSPQ we will have access to important health data of the SUPREME project (Système de surveillance et de prévention des impacts sanitaires des événements météorologiques extrêmes). Furthermore, P. Valois, director of OQACC, will provide data on adaptation of populations to climate change. We will organize monthly meetings with these two collaborators. Through these interactions, the PDF will become acquainted with statistical methods used in public health and research at organizations like INSPQ. So in addition to theoretical training in copulas, time series models, and mixed data, the PDF will also be exposed to unique data and applications as well as challenges of collaboration between researchers from academia and public sector.

Teaching/training/education

Schedule for year 1: Mentoring of students for the SSC case study competition or participating in a CANSSI summer training program.

Schedule for year 2: Teaching a one-semester course at HEC Montréal. The chair of the Department of Decisions Sciences wrote a letter confirming that the PDF could teach. At HEC Montréal there are courses in English and French so language does not matter. Note that ESPUM has few courses in statistics (all in French) and they are taught by faculty members so the teaching must be performed at HEC Montréal.

In addition, depending on the PDF qualifications, she could also be involved in mentoring graduate students of the mentors. Having an experience in HQP supervision could help the PDF in job applications.

Mentoring of the postdoctoral fellow

The mentors will design a training plan adapted to the PDF situation, based on her research potential, with a list of short-term and long-term goals to achieve, a detailed schedule of the important milestones, a formation plan if needed, e.g., learning hidden Markov models, discrete time series models, bootstrapping methods for computing P-values, or computation methods such as parallel computing, interfacing with C++, together with a dissemination plan. As mentors, we will schedule weekly meetings so that the PDF keep us informed of the progress and we can make quick adjustments in case of difficulties. We will encourage the PDF to attend seminars in statistical science as well as seminars in public health related issues.
We will also encourage the PDF to make presentations at scientific meetings in statistics, biometrics and biostatistics. As said before, the PDF will meet with our collaborators P. Valois and R. Bustinza, so she will be able to establish a network of collaborators in public health. The PDF will become familiar with interdisciplinary research, which is important to develop and demonstrate leadership. Finally, the PDF will have access to career planning training offered by HEC Montréal and U. de Montréal.
The PDF will follow the training module on unconscious biases of Canada Research Chairs. During meetings, we will also ensure that cultural, dietary, family, and religious constraints are respected. Finally, we will encourage the PDF to strengthen or acquire skill in both official languages, in order to increase her chances of finding a position. Finally, during the second year, the mentors will help the PDF prepare for job applications.

Proposed schedule for postdoctoral fellowship

Research schedule:

- First 6 months: estimation and goodness-of-fit tests for univariate time series with mixed distributions (P1, 1 publication, 1 CRAN package)

- Next 9 months: inference methods for time-varying non-central squared copulas (1 publication, 1 CRAN package)

- Last 9 months: inference methods for time-varying factor copulas (1 publication, 1 CRAN package)

Additional schedule for year 1: Mentoring of students for the SSC case study competition or participating in a CANSSI summer training program. Frequent meetings with P. Valois and Dr Bustinza.

Additional schedule for year 2: Teaching a one-semester course at HEC Montréal. Frequent meetings with P. Valois and Dr Bustinza. Preparation for job applications.

A list of qualifications of suitable candidates

The candidate should have expertise in copulas and goodness-of-fit tests based on empirical processes. Programming skills in R are essential. Finally, a good knowledge of time series modelling and health data are an asset.