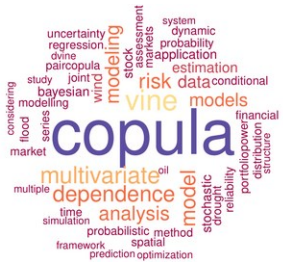
Department of Mathematics
TUM School of Computation, Information and Technology
Technical University of Munich

Workshop: Dependence models, Vines, and their Applications

Institute for Advanced Study
Lichtenbergstraße 2 a
85748 Garching

July 22, 20224 – July 24, 2024



Time	Session	Day 1 (July 22, 2024)	Speaker(s)
09:00 - 09:45		Tutorial Session 1	
09:45 - 10:00		Welcome	
10:00 - 10:45		**Keynote**: Scientometrics with vine regression	Roger Cooke
10:45 - 11:15		Break	
11:15 - 11:40	Copula-based modeling and inference	Quantile mediation analysis via copulas	Peter Song
11:40 - 12:05		Copula models for covariate distributions in the life sciences	Niklas Hartung
12:05 - 12:30		Virtual patient simulation in pharmacometric modeling using copulas	Laura Zwep
12:30 - 13:30		Lunch Break	
13:30 - 14:15	Methodological Innovations	**Keynote**: X-Vine Models for Multivariate Extremes	Anna Kiriliouk
14:15 - 14:40		Optimizing effective numbers of tests by vine copula modeling	Thorsten Dickhaus
14:40 - 15:05		An explainable vine copula-based generalization of logistic regression	Ingrid Hobæk Haff
15:10 - 15:40		Coffee Break	
15:40 - 16:05	Applications of Copula Models	D-Vine Copula based Probabilistic Weather Forecasting	Annette Möller
16:05 - 16:30		Gradient-Boosted Generalized Linear Models for Conditional Vine Copulas	David Jobst
16:30 - 16:55		An application of Y-vine copulas to meteorological data in Germany	Ferdinand Buchner
17:00 - 18:30		Poster Session & Drinks	

Time	Session	Day 2 (July 23, 2024)	Speaker(s)
09:00 - 09:45		Tutorial Session 2	
10:00 - 10:45		**Keynote**: Vine regression with discrete explanatory or response variables	Harry Joe
10:45 - 11:15		Break	
11:15 - 11:40	Specialized Copula Families & Dependence Measures	Testing Truncation Dependence: The Gumbel-Barnett Copula	Anne-Marie Toparkus
11:40 - 12:05		On copulas with a trapezoid support	Piotr Jaworski
12:05 - 12:30		Measuring Dependence between Events	Jan-Lukas Wermuth
12:30 - 13:30		Lunch Break	
13:30 - 14:15	Copulas in Finance & Insurance	**Keynote**: Pair Copula Construction with Insurance Applications	Peng Shi
14:15 - 14:40		Vine copulas for stochastic volatility	Alexander McNeil
14:40 - 15:05		Moving Aggregate Modified Autoregressive Copula-Based Time Series Models (MAGMAR-Copulas) Without Markov Restriction	Sven Pappert
15:10 - 15:40		Coffee Break	
15:40 - 16:05	Computational advances	(py|r)vinecopulib: an overview	Thibault Vatter
16:05 - 16:30		Vine copula models with zero inflated marginals	Thomas Nagler
16:30 - 16:55		A multivariate bias correction for partly zero-inflated margins based on vine copulas	Henri Funk
16:55 - 17:20		Computational aspects of the structural reliability assessment of submerged floating tunnels using vine copulas	Marcel 't Hart
18:30 - open end		Conference dinner	

Time	Session	Day3 (July 24, 2024)	Speaker(s)
10:00 - 10:45		**Keynote**: Exploring Copulas in Deep Learning	Natasa Tagasovska
10:45 - 11:15		Break	
11:15 - 11:40	Copulas in Machine Learning	Vine copula-based classifiers with applications	Özge Şahin
11:40 - 12:05		Vine copula regression with application to flight data	Claudia Czado
12:05 - 13:00		Lunch Break	
13:00 - 13:45	Miscellaneous with Pair-Copula Constructions	**Keynote**: Restrictions of PCBNs for integration-free computations	Dorota Kurowicka & Alexis Derumigny
13:45 - 14:10		Conditional Frugal Parameterisations	Daniel Manela
14:10 - 14:35		Comparison results for Markov tree distribution	Ansari Jonathan
14:35 - 15:00		Coffee Break	
15:00 - 15:25	Theoretical aspects	Directional Tail-weighted Dependence Measures for Multivariate Copulas	Xiaoting Li
15:25 - 15:50		Observability properties of copula-based state space models	Ariane Hanebeck
15:50 - 16:15		Parameter estimation in high-dimensional vine copula models	Jana Gauss
16:15 - 16:40		How simplifying and flexible is the simplifying assumption in pair-copula constructions	Sebastian Fuchs
16:40 - 16:50		Final remarks	

Day 1 (July 22, 2024)

10:00 - 10:45

Roger Cooke
Resources for the Future, Washington DC
Title: Scientometrics with vine regression.
Abstract: The College van Bestuur at the Delft University of Technology urged all Delft researchers to cite their Delft colleagues in their publications. Scientometrics, the "quantitative study of science, communication in science, and science policy", has a growing impact on science policy and citations are a key measure of academic performance. We use scientometric data of 10,725 research articles in Web of Science published between 2010 and 2014. At least one author of each publication is affiliated to Delft University of Technology (TU Delft). Citations up until 2017 have been recorded and the average over the number of years since publication provides the response variable: the mean citations score (mcs) of the publication. Explanatory covariates are the numbers of authors (authors) and references (refs) for each publication, the journal citation score (js), the greatest number of years since first publication of a co-author (ageMax), the largest number of publications among co-authors (pMax), the highest average citation count (mcsMax) and the highest proportion of highly cited publications (pptopMax). This data is analyzed with vine regression and linear regression. Vine regression models the joint distribution as a regular vine and computes the conditional expectation of the response variable given each vector of covariates in the data. This is the complete empirical regression function. Results are compared with standard linear regression. For example, vine regression explains 37% of the variance in msc whereas linear regression explains 28%. The effects of policy interventions are estimated by increasing the values of 'decision variables' and noting the change in the response variable's expectation. This assumes of course a causal relation underlying the correlations. In linear regression conditional expectation is linear in the covariates, resulting in absurd predictions for very large interventions. Vine regression offers a more plausible alternative. The data also raises interesting questions about correlation versus causality. A large number of references certainly couldn't cause a large number of citations.... or could it? What about a large number of co-authors? A notion of endogenous causality is introduced to sort this out.
Joint work with Tina Nane.

11:15 - 11:40

Peter Song
University of Michigan, Ann Arbor, MI
Title: Quantile mediation analysis via copulas.
Abstract: Mediation analysis is widely used in practice to examine if, and if so how, an intermediate variable mediates an exposure effect on an outcome. In some biomedical studies such as biological aging and obesity, quantiles of the outcome, instead of its mean, are of central interest. We develop a new methodological framework of quantile mediation analysis through copula models to facilitate identifying, estimating, and testing quantile mediation effects under a class of directed acyclic graphs, in which resulting key estimates take some closed form expressions. Since the null hypothesis of no mediation effect is composite, traditional tests such as the Sobel test or MaxP test appear conservative and underpowered. To improve statistical power, we adopt an adaptive bootstrap (AB) method to construct a calibrated test statistic that enables to properly control type I error under a composite null hypothesis. We establish theoretical guarantees for the proposed AB test method. We examine numerically both type I error and power of our AB test method through extensive simulation experiments and illustrate it in assessing mediation effects of lipid metabolites on the association between exposure to phthalates and childhood obesity.

11:40 - 12:05

Niklas Hartung
University of Potsdam, DEU
Title: Copula models for covariate distributions in the life sciences.
Abstract: Many dynamic models used in the life sciences depend on covariates, which often represent physiological or anatomical characteristics. Such covariates typically have non-Gaussian dependency structures, calling for non-Gaussian frameworks such as copula-based models. In this work, we thoroughly evaluate copula-based covariate distribution models in terms of Kullback-Leibler (KL) divergence, a scale-invariant information-theoretic goodness-of-fit criterion for distributions. In data sets of different sizes and dimensionalities containing both continuous and discrete covariates, copula-based models showed consistent improvements in KL divergence compared to simpler Gaussian or scale transform approximations. Furthermore, KL divergence estimates were robust to inclusion of latent variables, large fractions of missing values, no trend for overfitting could be seen, and they were able to translate correlation structures between populations. Parametric copula models were found to scale much better with the dimension of the dataset than nonparametric copula models. These findings corroborate the potential of copula-based models for modelling realistic life science covariate distributions.

12:05 - 12:30

Laura Zwep
Leiden University, NLD
Title: Virtual patient simulation in pharmacometric modeling using copulas.
Abstract: Virtual patient simulation is increasingly performed to support model-based optimization of clinical trial designs or individualized dosing strategies. Quantitative pharmacological models typically incorporate individual-level patient characteristics, or covariates, which enable the generation of virtual patient cohorts. The individual-level patient characteristics, or covariates, used as input for such simulations should accurately reflect the values seen in real patient populations. Current methods often make unrealistic assumptions about the correlation between patient's covariates or require direct access to actual data sets with individual-level patient data, which may often be limited by data sharing limitations. We propose and evaluate the use of copulas to address current shortcomings in simulation of patient-associated covariates for virtual patient simulations for model-based dose and trial optimization in clinical pharmacology. Copulas are multivariate distribution functions that can capture joint distributions, including the correlation, of covariate sets. We compare the performance of copulas to alternative simulation strategies, and we demonstrate their utility in several case studies. Our work demonstrates that copulas can reproduce realistic patient characteristics, both in terms of individual covariates and the dependence structure between different covariates, outperforming alternative methods, in particular when aiming to reproduce higher dimensional covariate sets. In conclusion, copulas represent a versatile and generalizable approach for virtual patient simulation which preserve relationships between covariates, and offer an open science strategy to facilitate re-use of patient data sets.
Zwep, L.B., Guo, T., Nagler, T., Knibbe, C.A.J., Meulman, J.J. and van Hasselt, J.G.C. (2024), Virtual Patient Simulation Using Copula Modeling. *Clinical Pharmacology & Therapeutics*, 115: 795-804. https://doi.org/10.1002/cpt.3099

13:30 - 14:15

Anna Kiriliouk
University of Namur, BEL
Title: X-Vine Models for Multivariate Extremes.
Abstract: Regular vine sequences permit the organization of variables in a random vector along a sequence of trees. Regular vine models have become greatly popular in dependence modelling as a way to combine arbitrary bivariate copulas into higher-dimensional ones, offering flexibility, parsimony, and tractability. We use regular vine structures to decompose and construct the exponent measure density of a multivariate extreme value distribution, or, equivalently, the tail copula density. Although these densities pose theoretical challenges due to their infinite mass, their homogeneity property offers simplifications. The theory sheds new light on existing parametric families and facilitates the construction of new ones, called X-vines. Computations proceed via recursive formulas in terms of bivariate model components. We develop simulation algorithms for X-vine multivariate Pareto distributions as well as methods for parameter estimation and model selection based on threshold exceedances. The methods are illustrated by a case study on US flight delay data.
This is joint work with Jeongjin Lee (University of Namur) and Johan Segers (Catholic University of Louvain).

14:15 - 14:40 **Thorsten Dickhaus**
University of Bremen, DEU
Title: Optimizing effective numbers of tests by vine copula modeling.
Abstract: In the multiple testing context, we utilize vine copula for optimizing the effective number of tests. It is well known that for the calibration of multiple tests for control of the family-wise error rate the dependencies between the marginal tests are of utmost importance. It has been shown in previous work, that positive dependencies between the marginal tests can be exploited in order to derive a relaxed Šidák-type multiplicity correction. This correction can conveniently be expressed by calculating the corresponding "effective number of tests" for a given (global) significance level. This methodology can also be applied to blocks of test statistics so that the effective number of tests can be calculated by the sum of the effective numbers of tests for each block. In this presentation, we demonstrate how the power of the multiple test can be optimized by taking blocks with high inner-block dependencies. The determination of those blocks will be performed by means of an estimated vine copula model. An algorithm is presented which uses the information of the estimated vine copula to make a data-driven choice of appropriate blocks in terms of (estimated) dependencies. Numerical experiments demonstrate the usefulness of the proposed approach.
This is joint work with Nico Steffen, and the presentation is based on [1]
[1] Steffen, N, Dickhaus, T, Optimizing effective numbers of tests by vine copula modeling, *Depend. Model*. 8, 172-185, 2020.

14:40 - 15:05 **Ingrid Hobæk Haff**
University of Oslo, NOR
Title: An explainable vine copula-based generalization of logistic regression.
Abstract: We propose a generalization of the logistic regression model to account for non-linear main effects and interaction effects in order to improve the prediction performance, but in such a way that the inherent interpretability of the logistic regression model is kept. We do this by using a specification of the model on generative form. More specifically, we set up a model for each of the two classes as a combination of certain marginal distributions and a vine copula, accounting for the dependence. The resulting model for the log odds becomes a sum of the logistic regression model with linear main effects and non-linear terms that involve two covariates or more. However, the parameters are estimated as in the discriminative framework and dependence between certain covariates is only included when it differs sufficiently in the two classes. Hence, the purpose is not to model the distribution of the covariates in the two classes as well as possible, but to make the best possible model to discriminate between the two classes. Further, a scheme for model selection and estimation is presented. In order to assess the performance of our model, we have run an extensive simulation study. The results from the study, as well as from a couple of examples on real data, show that our model performs at least as well as natural competitors, especially in the presence of non-linearities and complex interactions, even when the number of observations is not that large compared to the number of covariates.
Joint work with Simon Boge Brant and Haakon Bakka.

15:40 - 16:05 **Annette Möller**
Bielefeld University, DEU
Title: D-Vine Copula based Probabilistic Weather Forecasting.
Abstract: Current practice in predicting future weather is the use of numerical weather prediction (NWP) models, which consist of a set of differential equations describing the physics of the atmosphere. These models are run multiple times with different initial conditions and/or model formulations to obtain an ensemble of forecasts that represents model and forecast uncertainty. Despite of enormous improvements in NWP models over the last decades, the resulting ensemble forecasts still tend to exhibit forecast bias and dispersion errors and consequently lack calibration. Therefore, these forecasts should be improved by statistical postprocessing models.
In this work we propose a D-vine copula-based quantile regression (DVQR) approach for postprocessing of ensemble weather forecasts. The DVQR is a powerful tool for above application field, as it incorporates important predictor variables from a large set of potentially relevant predictors by a sequential forward selection procedure and is able to model complex nonlinear relationships among them. The DVQR approach is highly data driven and allows to adopt more general dependence structures as the state-of-the-art Ensemble model output statistic (EMOS) postprocessing model. However, the current DVQR does not explicitly allow to account for additional covariate effects, e.g. temporal or spatio-temporal information.
Consequently, we additionally introduce an extension of the DVQR, where we parametrize the bivariate copulas in the D-vine copula through Kendall's tau, which can then be linked to additional covariates. The parametrization of the correlation parameter allows generalized additive models (GAMs) and spline smoothing to detect potentially hidden covariate effects. The new method is called GAM-DVQR, and its performance is illustrated in a case study for postprocessing of 2m surface temperature forecasts. We investigate a constant as well as a time-dependent Kendall's tau. The GAM-DVQR models are compared to the benchmark methods Ensemble Model Output Statistics (EMOS), its gradient-boosted extension (EMOS-GB) and basic DVQR. The results indicate that the GAM-DVQR models indeed capture time-dependent correlations. They significantly outperform the state-of-the-art methods EMOS and EMOS-GB. Furthermore, the introduced parameterization allows using a static training period for GAM-DVQR, yielding a more efficient and stable model estimation in comparison to state-of-the-art methods as well as DVQR using a sliding training window.
Joint work with David Jobst and Jürgen Groß.

16:05 - 16:30 **David Jobst**
University of Hildesheim, DEU
Title: Gradient-Boosted Generalized Linear Models for Conditional Vine Copulas.
Abstract: Nowadays, gradient-boosting is frequently employed within the framework of distributional regression models to estimate model coefficients, and to provide by construction an intrinsic variable selection. Recently, more and more research focus on extending the gradient-boosting algorithm to multivariate distributional regression models. In our work, we contribute to the estimation of conditional vine copulas in a gradient-boosting setting. We employ conditional bivariate copulas to build a conditional vine copula, where the parameter of each bivariate copula depends on covariates. The covariates are linked via a generalized linear model (GLM) to the Kendall's τ correlation coefficient, from which the parameter for the corresponding bivariate copula can be derived. The coefficients in the linear predictor of the GLM are estimated via gradient-boosting followed by an additional covariate deselection procedure. This two-step estimation procedure is used for each conditional bivariate copula sequentially tree-by-tree.
In a simulation study for low- and high-dimensional covariate settings, we separately assess the performance of the two-step estimation procedure for conditional bivariate copulas and conditional vine copulas. We analyze aspects such as the accuracy of coefficient estimates, variable selection, and copula family selection.
In a case study, we apply conditional vine copulas estimated via gradient boosting to the multivariate postprocessing of ensemble weather forecasts in a lowdimensional covariate setting. The results indicate that conditional vine copulas help to restore the temporal dependence among the ensemble weather forecasts better than a multivariate Gaussian copula with constant parameters.
Joint work with Annette Möller and Jürgen Groß

16:30 - 16:55 **Ferdinand Buchner**
Technical University of Munich, DEU
Title: An application of Y-vine copulas to meteorological data in Germany.
Abstract: Ensembles of weather forecasts usually suffer from biases and dispersion errors. We propose using Y-vine copulas for the bivariate postprocessing of zonal and meridional wind vectors. Y-vines allow for a symmetric modelling of both wind vectors and their dependence using the forecast ensembles of multiple weather variables. We apply this approach to ensemble forecasts of wind vectors over Germany and compare the results to univariate methods for both wind vectors, like D-vine quantile regression and ensemble model output statistics.

17:00 – 18:00 **Aleksey Min**
Technical University of Munich, DEU
Title: Stationary vine copula models for multivariate time series.
Joint work with Thomas Nagler and Danier Krüger.

Han Li
The University of Melbourne, AUS
Title: Modeling cold-related excess deaths via stationary vine copulas.
Joint work with Thomas Nagler and Claudia Czado.

Wieger Schipper
Delft University of Technology, NLD
Title: Consistent resampling schemes for bootstrap hypothesis tests, with application to copulas.

Ichiro Nishi
Graduate University for Advanced Studies, Tachikawa, Tokyo, JPN
Title: Inter-dependencies of Japanese stock market modelled by D-vine copula-based approach.
Joint work with Yoshinori Kawasaki.

Paria Jahansa:
University of Oldenburg, DEU
Title: Using Copula Functions for Modeling Dependency in Cognitive Psychology.

Rieke Santjer:
Delft University of Technology, NLD
Title: Analyzing Mooring Line Tensions for Offshore Aquaculture: A Gaussian Copula-based Bayesian Network Approach.

Antoine Faul
University of Bern, CHE
Title: Easy conditioning way beyond the Gaussian case.
Joint work with D. Ginsbourger.

Day 2 (July 23, 2024)

10:00 - 10:45 **Harry Joe**
University of British Columbia, Vancouver, BC, CAN
Title: Vine regression with discrete explanatory or response variables.
Abstract: The vine pair-copula construction works for a mix of continuous and discrete variables via sequential mixing of conditional distributions.
Suppose Y is a response variable and $\boldsymbol{X}$ is a vector of explanatory variables, and the data are $(\boldsymbol{x}_i, y_i)$: $i = 1, \ldots, n$ from a random sample. A vine-based joint distribution $F_{\boldsymbol{X},Y}$ can be constructed with Y as a leaf node in each vine tree, and then prediction intervals can be based on the fitted $F_{Y|\boldsymbol{X}}(\cdot\,|x)$
In this context, several topics will be covered.
(a) Conditions on pair-copulas so that the vine-based $F_{Y|\boldsymbol{X}}$ matches homoscedastic Gaussian regression;
(b) conditions on when it is useful to use binary dummy variables in place of an ordinal explanatory variable;
(c) development of new algorithms with binary dummy variables, because existing vine regression algorithms cannot be directly used;
(d) in data applications, comparisons of estimated $F_{Y|\boldsymbol{X}}$ based on types of pair-copulas that can be used.

11:15 - 11:40 **Anne-Marie Toparkus**
Universität Rostock, DEU
Title: Testing Truncation Dependence: The Gumbel-Barnett Copula.
Abstract: In studies on lifetimes, occasionally, the population contains statistical units that are born before the data collection has started. Left-truncated are units that deceased before this start. For all other units, the age at the study start often is recorded and we aim at testing whether this second measurement is independent of the genuine measure of interest, the lifetime. Our basic model of dependence is the one-parameter Gumbel-Barnett copula. For simplicity, the marginal distribution of the lifetime is assumed to be Exponential and for the age-at-study-start, namely the distribution of birth dates, we assume a Uniform. Also, for simplicity, and to fit our application, we assume that units that die later than our study period, are also truncated. As a result, from point process theory, we can approximate the truncated sample by a Poisson process and thereby derive its likelihood. Identification, consistency and asymptotic distribution of the maximum-likelihood estimator are derived. Testing for positive truncation dependence must include the hypothetical independence which coincides with the boundary of the copula's parameter space. By non-standard theory, the maximum likelihood estimator of the exponential and the copula parameter is distributed as a mixture of a two- and a one-dimensional normal distribution. For the proof, the third parameter, the unobservable sample size, is profiled out.
An interesting result is, that it differs to view the data as truncated sample, or, as simple sample from the truncated population, but not by much. The application are 55 thousand double-truncated lifetimes of German businesses that closed down over the period 2014 to 2016. The likelihood has its maximum for the copula parameter at the parameter space boundary so that the p-value of test is 0.5. The life expectancy does not increase relative to the year of foundation. Using a Farlie-Gumbel-Morgenstern copula, which models positive and negative dependence, finds that life expectancy of German enterprises even decreases significantly over time. A simulation under the condition of the application suggests that the tests retain the nominal level and have good power.
Joint work with Raphael Weißbach.

11:40 - 12:05 **Piotr Jaworski**
University of Warsaw, POL
Title: On copulas with a trapezoid support.
Abstract: The goal of the talk is to present a new family of bivariate copulas, which are supported not on the whole unit square $[0,1]^2$, but on its subset, a trapezoid with vertices $(1/2, 1)$, $(0, 1)$, $(1, 0)$, $(0, 0)$.
The family is parametrized by generators F, the symmetric cumulative distribution functions, with full support, which are convex on the negative half line

$$C_F(u; v) = F\left(2F^{-1}\left(\frac{v}{2}\right) + F^{-1}(u)\right), \text{ for } v + 2u \leq 2.$$

The basic properties and necessary conditions for absolute continuity of C will be discussed. Several examples will be provided with different tail behavior of the generator F.
The interest in such copulas has aroused from the study of self-similar stochastic processes and the dependence between such process and its running minima or maxima processes.

12:05 - 12:30 **Jan-Lukas Wermuth**
Goethe University Frankfurt, DEU
Title: Measuring Dependence between Events.
Abstract: Measuring dependence between two events, or equivalently between two binary random variables, amounts to expressing the dependence structure inherent in a $\mathbf{2} \times \mathbf{2}$ contingency table in a real number between $-\mathbf{1}$ and $\mathbf{1}$. Countless such dependence measures exist, but there is little theoretical guidance on how they compare and, on their advantages, and shortcomings. Thus, practitioners might be overwhelmed by the problem of choosing a suitable measure. We provide a set of natural desirable properties that a proper dependence measure should fulfill. We show that Yule's Q and the little-known Cole coefficient are proper, while the most widely-used measures, the phi coefficient and all contingency coefficients, are improper. They have a severe attainability problem, that is, even under perfect dependence they can be very far away from $-\mathbf{1}$ and $\mathbf{1}$, and often differ substantially from the proper measures in that they understate strength of dependence. The structural reason is that these are measures for equality of events rather than of dependence. We derive the (in some instances non-standard) limiting distributions of the measures and illustrate how asymptotically valid confidence intervals can be constructed. In a case study on drug consumption we demonstrate how misleading conclusions may arise from the use of improper dependence measures.
Joint work with Marc-Oliver Pohle and Timo Dimitriadis.

13:30 - 14:15 **Peng Shi**
University of Wisconsin, Madison, WI
Title: Pair Copula Construction with Insurance Applications.
Abstract: Pair copula construction has emerged as a powerful tool in dependence modeling, particularly in high dimensional setting. In this presentation, we explore two low-dimensional applications of this technique in nonlife insurance context. The first concerns deductible pricing. We employ pair copula construction to jointly model the policyholder's deductible, number of claims, and individual claim amounts within compound distributions. The proposed method provides insurers an empirical tool to uncover the underlying risk distribution of the potential customers. The second is regarding individual loss reserving. We introduce a copula-based point process framework to model the recurrent events of payment transactions from an insurance claim, where the longitudinal payment amounts and the time-to-settlement outcome are formulated as the marks and the terminal event of the counting process, respectively. The model allows insurers to dynamically forecast unpaid losses by leveraging the granular transaction data on individual claims.

14:15 - 14:40 **Alexander McNeil**
University of York, UK
Title: Vine copulas for stochastic volatility.
Abstract: Models from the GARCH class have proved to be extremely useful models for forecasting volatility and measuring risk in financial time series. However, they are something of a black box with respect to their serial dependence structure and they may not be the best models for all-time series exhibiting stochastic volatility.
To shed more light on how GARCH models work, we examine the bivariate copulas that describe their serial dependencies and higher-order partial serial dependencies. We show how these copulas can be approximated using a combination of standard bivariate copulas and uniformity-preserving transformations known as v-transforms. The insights help us to construct stationary d-vine models to rival and often surpass the performance of GARCH processes in modelling volatile financial return series.

14:40 - 15:05 **Sven Pappert**
Technical University of Dortmund, DEU
Title: Moving Aggregate Modified Autoregressive Copula-Based Time Series Models (MAGMAR-Copulas) Without Markov Restriction.
Abstract: Copula-based time series models implicitly assume a finite Markov order. In reality a time series may not follow the Markov property. We modify the copula-based time series models by introducing a moving aggregate (MAG) part into the model updating equation. The functional form of the MAG-part is given as the inverse of a conditional copula. The resulting MAG-modified Autoregressive Copula-Based Time Series model (MAGMAR-Copula) is discussed in detail and distributional properties are derived in a D-vine framework. The model nests the classical ARMA model as well as the copula-based time series model. The modeling performance is compared with the model from McNeil and Bladt, modeling US inflation. Our model is competitive in terms of information criteria. It is a generalization of ARMA and also copula-based time series models and is in spirit similar to other moving average time series models such as ARMA and GARCH.

15:40 - 16:05 **Thibault Vatter**
University of Applied Sciences Western, Carouge, CHE
Title: (py|r)vinecopulib: an overview.
Abstract: Vine copulas have been applied to a wide range of scientific problems, and their success has led to continual advances in related methodology. Powering most of this research is open source software, which has seen significant improvement over the years. In this presentation, we discuss the vinecopulib ecosystem, heir apparent to the popular VineCopula R package.
As a standalone C++ library, vinecopulib provides high-performance implementations of the core features of VineCopula. In addition to shorter runtimes and lower memory consumption, vinecopulib's advantages include a modular and extensible implementation with R and Python interfaces, support for both parametric and nonparametric models, partially specified structures, handling of discrete variables, and much more.

16:05 - 16:30

Thomas Nagler
Ludwig-Maximilians-Universität München, DEU
Title: Vine copula models with zero inflated marginals.
Abstract: Real-world datasets often contain zero-inflated variables, which traditional vine copula models cannot handle. In this talk, I will introduce novel extensions of vine copula models specifically designed to manage zero-inflated data. These models are based on a density decomposition that generalizes the foundational result of Bedford and Cooke (2001). Our new result accommodates random variables that combine a continuous component with a countable number of point masses.
I will also discuss the nonparametric estimation of simplified vine models that involve such variables, using ideas from density estimation under interval censoring. Following this presentation, Henri Funk will demonstrate the application of these tools in the environmental sciences.

16:30 - 16:55

Henri Funk
Ludwig-Maximilians-Universität München, DEU
Title: A multivariate bias correction for partly zero-inflated margins based on vine copulas.
Abstract: Climate model large ensembles are an important research tool to analyze and quantify natural climate variability and provide reliable information for rare extreme events. However, the models produce instances of reality via simulation and therefore contain biases. The research field of bias adjustment includes methods that correct these errors in model data using information from observed data while preserving the specific pattern of the model.
For a reliable correction, the marginal univariate climate variables and their inter-variable dependencies in the data require to be modelled and corrected jointly [1, 2]. The choice of method and design will determine the quality of the correction [3]. For example, univariate correction methods cannot account for multivariate relationships. We introduce vineBC, a novel bias-correction method, that models and transfers multivariate dependence structures in the data from its error-prone model domain to a realistic observed domain. vineBC estimates the model and the observed distribution using vine copulas [4] and transfers the model distribution via (inverse) Rosenblatt transformation from the model to the observed domain. We developed a vine copula model for partially inflated data. This novel approach to vine copulas allows more accurate modelling of zero-inflated climate data like precipitation or radiation and their joint dependencies than in other approaches, where the KDE estimation is simply truncated at an arbitrary value above zero. By shifting the data to a more realistic distribution via Rosenblatt transformation, the models' weather patterns and internal variability are preserved naturally. Application results of vine BC show that the novel correction method shows an improvement both in the strength of correction and in the preservation of variability when being compared to state-of-the-art approaches.
Joint work with Helmut Küchenhoff, Ralf Ludwig, Thomas Nagler.
[1] Cannon, A. J. Multivariate quantile mapping bias correction: An N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate dynamics*, 50, 31-49 (2018).
[2] Vrac, M. Multivariate bias adjustment of high-dimensional climate simulations: The Rank Resampling for Distributions and Dependences (R2D2) bias correction. *Hydrology and Earth System Sciences*, 22(6), 3175-3196 (2018).
[3] François, B., Vrac, M., Cannon, A. J., Robin, Y., & Allard, D. Multivariate bias corrections of climate simulations: which benefits for which losses? *Earth System Dynamics*, 11(2), 537-562 (2020).
[4] Czado, C., & Nagler, T. Vine copula-based modeling. *Annual Review of Statistics and Its Application*, 9, 453-477 (2022).

16:55 - 17:20

Marcel 't Hart
Delft University of Technology, NLD
Title: Computational aspects of the structural reliability assessment of submerged floating tunnels using vine copulas.
Abstract: Submerged floating tunnels 0are structures which have not been built yet. Designs of these structures rely currently on codes and standards of other fields of applications such as infrastructure, offshore and (sub)marine structures. Additionally, the external impacts of loads such as waves and currents need to be considered in terms of their amplitudes, directions and different combinations. To assess the structural reliability which would overarch the different applications and standards, a mechanical model considering dynamics has been set up. For a hypothetical structure a limited number of analyses has been performed. From the input and output variables, datasets of 8 variables have been identified and used to find different Vine Copulas on these variables. In order to find the best Vine Copula, in terms of the lowest Akaike Information Criterium, a brute force approach has been applied using the Chimera atlas 0and a high-performance computer 0to process all different Regular Vine structures. With the found Vine Copulas the integrity of the system was assessed and the exceedance probabilities of different variables in the structure were calculated.
Joint work with Oswaldo Morales-Napoles
[1] ITA WG 11, An Owner's Guide to Submerged Floating Tunnels, Athens 2023, https://about.ita-aites.org/publications/wg-publications/2237/an-owner-s-guide-to-submerged-floating-tunnels
[2] Morales Napoles, Oswaldo & Rajabi-Bahaabadi, Mojtaba & Alexandra, Gina & Hart, Cornelis. (2023). Chimera: An atlas of regular vines on up to 8 nodes. *Scientific Data*. 10. DOI: 10.1038/s41597-023-02252-6.
[3] Delft High Performance Computing Centre (DHPC), DelftBlue Supercomputer (Phase 2), 2024, https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2

Day 3 (July 24, 2024)

10:00 - 10:45

Natasa Tagasovska
Genentech/Roche Group, South San Francisco, CA
Title: Exploring Copulas in Deep Learning.
Abstract: While the deep learning and copulas communities share common objectives and face similar challenging problem setups, their collaborative integration is still nascent. This presentation aims to provide an overview of the current landscape of copulas in deep learning, focusing on areas such as generative modeling and uncertainty estimation. Additionally, we will identify potential synergies and opportunities for collaboration between these two frameworks.

11:15 - 11:40

Özge Şahin
Delft University of Technology, NLD
Title: Vine copula-based classifiers with applications.
Abstract: The vine pair-copula construction can be used to fit flexible non-Gaussian multivariate distributions to a mix of continuous and discrete variables. With multiple classes, fitting univariate distributions and a vine to each class lead to posterior probabilities over classes that can be used for discriminant analysis. This is more flexible than methods with the Gaussian and/or independence assumptions, such as quadratic discriminant analysis and naive Bayes. Some variable selection methods are studied to accompany the vine copula-based classifier because unimportant variables can make discrimination worse. Simple numerical performance metrics cannot give a full picture of how well a classifier is doing. We introduce categorical prediction intervals and other summary measures to assess the difficulty of discriminating classes. Through extensive experiments on real data, we demonstrate the superior performance of our approaches compared to traditional discriminant analysis methods and random forests when features have different dependent structures for different classes.
Joint work with Harry Joe.

11:40 - 12:05

Claudia Czado
Technical University of Munich, DEU
Title: Vine copula regression with application to flight data.
Abstract: Simplified vine copulas can also be used to construct flexible classes of regression models, which can accommodate non-linear non-Gaussian dependence. This will include models for univariate and bivariate responses. For this the conditional distribution of the response given the covariates will be derived from a joint simplified vine copula model of response and covariates. This provides a distributional regression framework allowing for simple determination of conditional quantiles. The simplified vine copula-based regression models are constructed in such a way that the conditional density can be written explicitly without integration. This allows us also to develop a forward selection strategy to avoid overfitting. These approaches will be introduced and an application involving assessing risks in flight landings will be given.

13:00 - 13:45

Dorota Kurowicka & Alexis Derumigny
Delft University of Technology, NLD
Title: Restrictions of PCBNs for integration-free computations.
Abstract: The pair-copula Bayesian Networks (PCBN) are graphical models composed of a directed acyclic graph (DAG) that represents (conditional) independence in a joint distribution. The nodes of the DAG are associated with marginal densities, and arcs are assigned with bivariate (conditional) copulas following a prescribed collection of parental orders. The choice of marginal densities and copulas is unconstrained. However, the simulation and inference of a PCBN model may necessitate possibly high-dimensional integration.
We present the full characterization of DAGs that do not require integration for density evaluation or simulations. Furthermore, we propose an algorithm that can find all possible parental orders that do not lead to (expensive) integration. Finally, we show the asymptotic normality of estimators of PCBN models using stepwise estimating equations. Such estimators can be computed effectively if the PCBN does not require integration. A simulation study shows the good finite-sample properties of our estimators.

13:45 - 14:10 **Daniel Manela**
University of Oxford, UK
Title: Conditional Frugal Parameterizations.
Abstract: This paper introduces the concept of Conditional Frugal Parameterizations, which blend Pair Copula Constructions and Bayesian Networks to facilitate the accurate simulation of data in causal models. We delve into the complexities of ensuring the simulated distribution's fidelity to the desired joint distributions within the confines of frugal models. We identify and quantify the key challenges of parameterizing conditional frugal models, and show that naive attempts to parameterize the model can lead to differences between the target and sample joint for both a) the dependency structure and b) the marginal causal effect. effect the role of pair copula in preserving marginal causal effects, and the application of Bayesian Networks in modelling pre-treatment covariate distributions. We show that such frugal models can be exactly parameterized numerically. Additionally, we provide efficient approximations when this is computationally infeasible. Our findings offer significant insights into advanced statistical methods for causal inference, highlighting the potential of our approach in enhancing the accuracy and efficiency of data simulation in complex causal models. In particular, we define specific cases where frugal models can be exactly parameterized, and quantify the degree to which model misspecification affects the sampled joint. When such errors are highly undesirable, we propose using an alternative parametrization based on the non-paranormal approximation and experimentally demonstrate the benefit yielded by such an approximation.

14:10 - 14:35 **Ansari Jonathan**
Paris Lodron Universität Salzburg, AUT
Title: Comparison results for Markov tree distribution.
Abstract: We develop comparison results for Markov tree distributions extending ordering results from the literature on discrete time Markov processes and recently studied ordering results for conditionally independent factor models to tree structures. Based on fairly natural positive dependence conditions, our main contribution is a comparison result with respect to the super modular order. Since this order is a pure dependence order, it has many applications in optimal transport, finance, and insurance. As an illustrative example, we consider hidden Markov models and study distributional robustness for functionals of the random walk under model uncertainty. Further, we show that, surprisingly, more general comparison results via the recently established rearrangement-based Schur order for conditional distributions, which implies an ordering of Chatterjee's rank correlation, do not carry over from star structures to trees. Several examples and a detailed discussion of the assumptions demonstrate the generality of our results and provide further insights into the behavior of multidimensional distributions.
Joint work with Moritz Ritter.

15:00 - 15:25 **Xiaoting Li**
University of British Columbia, Vancouver, BC, CAN
Title: Directional Tail-weighted Dependence Measures for Multivariate Copulas.
Abstract: Understanding the dependence of multivariate extremes is crucial but challenging, particularly in high-dimensional scenarios with sparse extreme observations. While multivariate copula models, such as vine copulas and factor copulas, offer significant flexibility for capturing tail dependence structures, the inference on extremes may be compromised if the model is fitted to the entire dataset. Therefore, there is a critical need for tail-based diagnostics to assess the fit in the tail region and guide the selection of appropriate parametric models. We introduce a novel family of directional dependence measures for multivariate copulas. The family of dependence measures is indexed by $\alpha \geq 1$. When $\alpha = 1$, they measure the strength of dependence along different paths to the joint upper or lower orthant. For large α, they become tail-weighted dependence measures that put more weight in the joint upper or lower tails of the distribution. As $\alpha \to \infty$, they converge to the multivariate tail dependence function. The distinct convergence patterns of copula families provide useful information to distinguish the tail properties. We further develop rank-based sample estimators for these dependence measures. In this talk, I will focus on discussing the uses of these sample estimators in visualizing the tail properties of data and assessing the goodness-of-fit of fitted copula models.

15:25 - 15:50 **Ariane Hanebeck**
Technical University of Munich, DEU
Title: Observability properties of copula-based state space models.
We propose a novel approach to check the observability properties of copula-based state space models (SSMs). The approach is based on deterministic density approximations, specifically modified for copula models. The approach is illustrated for different specifications of copula-based SSMs of interest. These include the novel copula-based SSM allowing for two latent variables.

15:50 - 16:15 **Jana Gauss**
Ludwig-Maximilians-Universität München, DEU
Title: Parameter estimation in high-dimensional vine copula models.
Abstract: Parameter estimation in high-dimensional vine copula models: In certain applications, the dimension of a vine copula model is large and grows with the sample size. This leads to the question under which conditions the parameters can be estimated via stepwise ML estimation. We show that under certain assumptions, the stepwise MLE is consistent and asymptotically normal if the number of parameters diverges. The results can also be applied to the generalized method of moments and can be extended to penalized estimation.

16:15 - 16:40 **Sebastian Fuchs**
Paris Lodron Universität Salzburg, AUT
Title: How simplifying and flexible is the simplifying assumption in pair-copula constructions.
Abstract: Motivated by the increasing popularity and the seemingly broad applicability of pair-copula constructions, in this contribution we tackle the question on how flexible and simplifying the commonly used 'simplifying assumption' is from an analytic perspective and provide answers to two related open questions posed by Nagler and Czado [2] in 2016. Aiming at a simplest possible setup for deriving the main results we first focus on the three-dimensional setting. We prove that the family of simplified copulas is flexible in the sense that it is dense in the set of all three-dimensional copulas with respect to the uniform metric d_∞ - considering stronger notions of convergence like the one induced by the metric D_1, by weak conditional convergence, by total variation, or by Kullback-Leibler divergence, however, the family even turn out to be nowhere dense and hence insufficient for any kind of flexible approximation. Furthermore, returning to d_∞ we show that the partial vine copula is never the optimal simplified copula approximation of a given, non-simplified copula C, and derive examples illustrating that the corresponding approximation error can be strikingly large and extend to more than 28% of the diameter of the metric space. Moreover, the mapping ψ assigning each three-dimensional copula its unique partial vine copula turns out to be discontinuous with respect to d_∞ (but continuous with respect to D_1 and to weak conditional convergence), implying a surprising sensitivity of partial vine copula approximations. The afore-mentioned main results concerning d_∞ are then extended to the general multivariate setting. Joint work with T. Mroz and W. Trutschnig
References:
[1] Mroz, T., S. Fuchs, and W. Trutschnig (2021). How simplifying and flexible is the simplifying assumption in pair-copula constructions - analytic answers in dimension three and a glimpse beyond. *Electronic Journal of Statistics* 15 (1), 1951–1992.
[2] Nagler, T. and Czado, C. (2016). Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis* 151, 69–89.