Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 104 (2017) 572 – 577

Procedia Computer Science

ICTE 2016, December 2016, Riga, Latvia

Algorithms of the Copula Fit to the Nonlinear Processes in the Utility Industry

Andrejs Matvejevs[a,*], Jegors Fjodorovs[a], Anatoliy Malyarenko[b]

[a]*Riga Technical University,Kalku 1, Riga, LV1658, Latvia*
[b]*Mälardalen University, Västerås, Sweden*

Abstract

Our research studies the construction and estimation of copula-based semi parametric Markov model for the processes, which involved in water flows in the hydro plants. As a rule analyzing the dependence structure of stationary time series regressive models defined by invariant marginal distributions and copula functions that capture the temporal dependence of the processes is considered. This permits to separate out the temporal dependence (such as tail dependence) from the marginal behavior (such as fat tails) of a time series. Dealing with utility company data we have found the best copula describing data - Gumbel copula. As a result constructed algorithm was used for an imitation of low probability events (in a hydro power industry) and predictions.

Peer-review under responsibility of organizing committee of the scientific committee of the international conference; ICTE 2016

Keywords: Copula; Diffusion processes; Time series; Semi parametric regressions

1. Introduction

Our research studies the construction and estimation of copula-based semi parametric Markov model for the processes, which involved in water flows in the hydro plants.
Copulas became popular in the finance and insurance community in the past years, where modeling and estimating the dependence structure between several univariate times series are of great interest; see Frees and Valdez[1]and Embrechts et al.[2] for reviews.

* Corresponding author. Tel.: +371 26015121.
E-mail address: Andrejs.Matvejevs@rtu.lv

Peer-review under responsibility of organizing committee of the scientific committee of the international conference; ICTE 2016
doi:10.1016/j.procs.2017.01.174

A copula function is a multivariate distribution function with standard uniform marginals. By Sklar's[3] theorem, one can always model any multivariate distribution by modeling its marginal distributions and its copula function separately, where the copula captures all the scale-free dependence in the multivariate distribution.

The central result of this theorem, which states that any continuous N-dimensional cumulative distribution function F, evaluated at point $x = (x_1, \dots, x_n)$ can be represented as

$$F(x) = C(F_1(x_1), \dots, F_n(x_n)), \tag{1}$$

where C is called a copula function and $F_i(x_i)$, $i = 1, \dots, n$ are the marginal distributions. The use of copulas therefore splits a complicated problem (finding a multivariate distribution) into two simpler tasks. The first task is to model the univariate marginal distributions and the second task is finding a copula that summarizes the dependence structure between them.

The possibility of identifying nonlinear time series using nonparametric estimates of the conditional mean and conditional variance were studied in many papers[4]. As a rule analyzing the dependence structure of stationary time series $\{x_t, t \in Z\}$ regressive models defined by invariant marginal distributions and copula functions that capture the temporal dependence of the processes. As it indicated[4] this permits to separate out the temporal dependence (such as tail dependence) from the marginal behavior (such as fat tails) of a time series. One more advantage of this type regressive approach is a possibility to apply probabilistic limit theorems for transition from deference equations to continuous time stochastic differential equations[5,6]. In our paper, we also study a class of copula-based semi parametric stationary Markov models in a form of scalar difference equation

$$t \in Z : X_t = f(X_{t-1}) + g(X_{t-1})\xi_t, \tag{1a}$$

where $\{\xi_t, t \in Z\}$ is i.i.d., N(0; 1). Regressions (1a) are high-usage equations for simulation and parameter estimation of stochastic volatility models ([2]). But, unfortunately defined by (1a) Markov chain has incompact phase space that complicates an application of probabilistic limit theorem. Copula approach helps to simplify asymptotic analysis of (1a). Let us remember that to construct a copula *C(u; v)* for pair $\{X_{t-1,} X_t\}$ from (1a) one should find a marginal invariant distribution *F(x)* for X_t and to substitute this in joint distribution function $H(x, y) = P(X_{t-1} \le x, X_t \le y)$, that is, $C(u,v) = H(F^{-1}(u), F^{-1}(v))$ and $H(x, y) = C(F(x), F(y))$. After a substitution $U_t = F(X_t)$ in equation (1) for a further diffusion approximation one can write a difference equation in a same form like (1a):

$$t \in Z : U_t = \varphi(U_{t-1}) + \psi(U_{t-1})\xi_t. \tag{2}$$

But now this equation defines Markov chain on the compact [0, 1]. This makes easier formulate construction for transition probability and further estimators of functions $\hat{f}(u)$ and $\hat{g}(u)$. After diffusion approximation of (2) one can make inverse substitution and derive stochastic differential equation as diffusion approximation for (1a).

We found that the best copula describing data is Gumbel copula. As a result constructed equation (1a) was used for low probability events imitation (hydro power industry) and predictions.

The paper is structured as follows. Section 2 describes our approach. In Section 3 we report our results for the data Section 4 concludes and discusses several possible avenues of future research.

2. Evaluation of parameters for the semi parametric regression model

Copula based semi parametric models are characterized by conditional heteroscedasticity and have been often used in modeling the variability of statistical data. The basic idea was to apply a local linear regression to the squared residuals for finding the unknown functions f and g[5,7].