CHAPTER 4
Bivariate Copulas Estimation Using Semiparametric Methods

Rahmah Mohd Lokoman, Fadhilah Yusof, and Zulkifli Yusop

4.1 INTRODUCTION

A copula function is the product of integrating two or more uniform marginal distributions into a single joint distribution function and it was pioneered by Sklar (1959). Since this method allows one to apply a certain distribution to the marginal variables and select a certain copula to create the dependency distribution for the variables, it can resolve the limitations of the conventional approach. The copula method can be used to construct joint distribution of bivariate with different marginal distributions for each variable and this was studied by Zhang and Singh (2007). There are numerous copula families, such as Archimedean and Elliptical families, that have been proposed and created to cover a broad range of dependency structures. The parameter of the copula function must first be estimated and defined in order to assess a given copula model. Many methods have been developed for estimating a copula dependency parameter, such as parametric and semiparametric.

In parametric approach, marginal distribution must fit a parametric distribution. The marginal parameters and copula dependency parameters are the parameters of interest. The fact that parametric methods are widely used is because the methods estimate the parameters with a precise estimate. Lokoman and Yusof (2019) did a comparative study between three copula

estimation methods in parametric approach: adaptive maximization by parts (AMBP), inference function of margins (IFM) and maximum likelihood (ML) estimation. The results show that the AMBP has the best estimation performance and IFM and AMBP can have relatively same performance as the number of data and the correlation value increase. However, they found that a misassumption of the parametric marginal distribution model could lead to a significant estimation error. Therefore, to solve the misassumption problems in parametric approach, semiparametric approach is proposed and developed.

The name semiparametric approach is given because the marginal distributions are computed nonparametrically while the joint distribution is computed parametrically. Marginal distributions are nonparametrically estimated in order to overcome the misassumption problem. Under this approach, four estimation methods have been developed, maximum pseudo likelihood (MPL), maximum likelihood on kernel density (ML-KD), inversion of Kendall's tau (iTAU) and inversion of Spearman's rho (iRHO). MPL estimation method was proposed by Genest et al. (1995). Marginal distributions can be constructed using univariate empirical cumulative distribution functions since there are no parametric assumptions for the margins.

Bouyé et al. (2001) and Cherubini et al. (2004) named this MPL method, canonical maximum likelihood (CML). Tsukahara (2005), Kim et al. (2007), Kojadinovic and Yan (2010) found that MPL has the best performance in estimating the dependency estimator in several comparative studies of copula parameter estimation methods. According to Kim et al. (2007), MPL overcomes the disadvantage of ML estimation and IFM, because of the misspecification of the non-robustness toward the marginal distribution.

A method known as ML-KD is another estimation method with a structure close to MPL (Charpentier et al., 2007). In terms of unknown marginal distributions, they are identical. Kernel density estimation (KDE), also known as the kernel method, is used in ML-KD to estimate marginal distribution functions. Rosenblatt (1956) and Parzen (1962) developed the kernel method, which is a nonparametric method to estimate a random variable x's Probability Density Function (PDF), $f(x)$. Epanechnikov, uniform and Gaussian are only a few of the kernel functions available in KDE. According to Sharma et al. (1998), Kim and Heo (2002), Kim et al. (2006), and Jou et al. (2008), the Gaussian kernel function is the most common option in hydrology analyses. Since the parameter is used to smooth the shape of the approximate PDF, bandwidth selection is crucial for precise kernel smoothing. According to Jaki and West (2008), choosing a bandwidth that is too small will result in a KDE with significant errors, whereas choosing a bandwidth that is too large will result in an over smoothed KDE. The ML-KD method is rarely used for estimating the copula dependency parameter in hydrologic applications, even though the kernel method is commonly used to estimate the marginal density of hydrological variables.

iTAU and iRHO methods use correlation coefficient of Kendall's tau and Spearman's rho respectively to estimate the copula parameter. Kendall's tau and Spearman's rho are two of the most well-known and widely used nonparametric methods for measuring correlations between random variables (Fredricks & Nelson, 2007; Vandenberghe et al., 2010; Chen et al., 2015). Correlation analyses using these two measures are commonly used to assess the strength of a relationship between two variables based on data ranks. However, these two measures