CHAPTER 10

Nonparametric Predictive Inference with Copulas for Bivariate Data

Noryanti Muhammad, Frank Coolen,
Tahani-Coolen Maturi, and Norazlina Ismail

10.1 INTRODUCTION

Dependencies are important in many real applications. However, identifying and modelling dependencies between two or more related random quantities is a main challenge in statistics. The dependence structure in the models will be identified before any prediction or estimation can be performed toward getting the most efficient and accurate prediction and forecasting. Analyses of dependencies are of considerable importance in many sectors as an aid to better understanding the interaction of variables in a certain field of study and as an input in every aspect of our life, including engineering, health, finance, insurance, and agriculture. Statistical dependence is a relationship between two or more characteristics of units under study or review. These units may, for example, be individuals, objects, or various aspects of the environment. The dependence structure is important in knowing whether a particular model or inference might suit a given application or data set.

Several types of dependence can occur, for example, positive and negative dependence, exchangeable or flexible dependence, and dependence decreasing with lag (for data with a time index) (Joe, 1997). A popular method for modelling dependencies is using a copula (Cherubini et al., 2004; Rank, 2007). Generally, a copula is a multivariate probability distribution for which the marginal probability distribution of each variable is uniform (Joe, 1997, Nelsen, 2007). Many researchers have addressed and studied dependence using copulas (Genest et al., 1995; Embrechts et al., 2003; Scaillet & Fermanian, 2002; Tsuhakara, 2005). Often, in their studies, they estimate dependence parameter(s). Dependence is also important in prediction, where it plays a key role in decision-making processes, classifying, and other aspects involving dependence. For example, in risk of failure trajectory (e.g., the effect of random external actions like wind or unexpected reactions of the drivers), the dependence structure between vehicle criteria and safety acceptance of the models is considered to reduce road accident rate (Koita et al., 2013).

This chapter presents a new method for predictive inference considering the dependence structure. It uses Nonparametric Predictive Inference (NPI) for the marginals combined with a copula. We restricted attention to bivariate data. The important general idea in this chapter is to look at the prediction of the two random quantities. We consider the dependence structure between these two random quantities using copula, as copula gives an interesting tool for describing the dependence structures. The idea that we consider is the dependence structure between the two random quantities using parametric copula for small data sets and nonparametric copula, specifically kernel-based method for large data sets. The NPI on the marginals with the estimated copulas, presented in this chapter, is somewhat different from the usual statistical approaches based on imprecise probabilities (Augustin et al., 2014). The proposed method in this chapter uses a discretized version of the copula, which fits perfectly with the NPI method for the marginals and leads to relatively straightforward computations because there is no need to estimate the marginals and the copula simultaneously. Using the NPI for the marginals, the information shortage is most likely about the dependence structure.

This chapter is organized by first presenting nonparametric predictive inference (NPI) in Section 10.2, including the proposed method of NPI with parametric copula and semi-parametric predictive inference with real data application. Section 10.3 describes the results and discussion, and finally, Section 10.4 provides the general conclusion drawn from the chapter.

10.2 NONPARAMETRIC PREDICTIVE INFERENCE

Inference on a future observation based on past data observations is a frequentist statistical framework which has been called Nonparametric Predictive Inference (NPI) (Coolen, 2011). NPI uses lower and upper probabilities, also known as an imprecise probabilities (Augustin et al., 2014), to quantify uncertainty based on only a few assumptions.

NPI is based on the assumption $A_{(n)}$, proposed by Hill (Hill, 1968), which gives direct conditional probabilities for a future real-valued random quantity, conditional on observed values of n related random quantities (Augustin & Coolen, 2004; Coolen, 2006). Effectively, it assumes that the rank of the future observation among the observed values is equally likely to have each possible value $1,...,n+1$. Hence, this assumption is that the next observation has probability $1/(n+1)$ to be in each interval of the partition of the real line as created by the n observations. Suppose that $X_1, X_2, ..., X_n, X_{n+1}$ are continuous and exchangeable real-valued random quantities. Let the ordered observed values of $X_1, X_2, ..., X_n$