Supervise Learning With Copulas

Xiaoping Shen[1], Robert L. Ewing[2], and Jia Li[3]

[1]Department of Mathematics, Ohio University, Athens, OH 45701
[2]Sensors Directorate, Air Force Research Lab, WPAFB, OH 45433
[3]Dept. of Electrical and Computer Engineering, Oakland University, Rochester, MI 48309

***Abstract*—The naïve Bayes classifier plays an important role among the classifiers based on supervised learning, although it requires strong condition on the feature independence assumptions. A measurement for the independency checking in the data preprocessing is necessary to guarantee the effectiveness of the classifier. Copula Theory is a mathematical tool in dependency modeling. In this paper, we recall elements of copulas and introduce a new algorithm to construct multiscale copula estimators which can be used for the independency testing to improve the accuracy of the Naïve Bayes classifier.**

Keywords — Copula, nonlinear independency modeling, Naïve Bayes classifier, probabilistic classifier, non-parametric copula estimator.

I. Introduction

Copula was first adopted by Abe Sklar in 1959 to describe statistical models [1]. It is a function connects joint distributions and their margins. As mentioned in [2], the name "copula" was chosen to emphasize the manner in which a copula "couples" a joint distribution with its univariate margins. It separates the joint distribution into two contributions, the marginal distributions of the individual variables and the interdependency of the probabilities. A detailed description of the copula theory can be found in [2, 3]. Copulas were introduced to model statistical dependency in the mid-nineties. Since then, numerous papers have been published to advance the theory and application of copulas. The most active areas of applications are found in finance, economics (see [4], and the survey paper [5]), risk management ([6]-[10]) and related areas such as insurance and actuarial science. Its connection to the area of machine learning was overlooked until recent years [11]. Copulas have been used by the data preprocessing for the probability classifier - Naïve Bayesian Classifier ([12]-[19]) to improve the efficiency of the classifiers, for example. These applications motivate the study of the analytical properties and numerical construction of copulas. In particular, the multiscale copula based on wavelets, which has the potential to measure the independency at different levels.

This paper served as an introduction to the multiscale copula and its connection to supervised learning. It is organized as follows, following this introduction, elements of copula theory will be reviewed in section 2. Section 3 discuss the algorithm used to construct bivariate copula. Numerical examples are shown in section 4. The article is concluded with a summary and areas of future studies.

II. Background

A. Copulas

There are a great variety of families of copulas that can used for different application purposes. Different copulas have their own characteristics and may fit different types of data. Commonly used properties that a "good" parametric family of multivariate copulas should have to be considered "interesting" in statistical applications are

- Interpretability
- Flexible and wide range of dependence.
- Easy-to-handle

The foundation of the theory of copulas are base on the following definition by Sklar.

Definition (Copula in n-dimension). Let F be a n-dimensional cumulative distribution function(cdf) with margins $F_1,\ldots, F_n$, Then there exists a n-copula C such that for all $(y_1,\ldots, y_n)$:

$$\begin{aligned} F(y_1,\cdots,y_n) &= P(Y_1 \le y_1,\cdots,Y_n \le y_n) \\ &= C(F_1(y_1),\cdots,F_n(y_n)) \\ &= C(u_1,\cdots,u_n) \end{aligned}$$

C is a n-d operator, it is a distribution function with margins $F_1,\ldots, F_n$ and satisfied the following conditions,

(1) $C(1,\cdots,1,u_i,1,\cdots,1) = u_i$ for all $i \le n$ and all u_i in [0,1];

(2) $C(u_1,\cdots,u_i,\cdots,u_n) = 0$ if $u_i = 0$ for all $i \le n$;

(3) C is n-increasing.

The bivariate copulas (n=2), attracts most of the attention in the research community and they can be used in bi-objective feature selection for discriminant analysis in two-class classification.

The 2-copula has the following properties:

For every *u* and *v* in [0,1], we have

1. Uniformity of margins.
   C(u; 0) = 0 = C(0; v), C(u; 1) = u and C(1; v) = v;
2. 2- increasing property.

For every $u_1; u_2; v_1; v_2$ in [0,1] and $u_1 \le u_2$ and $v_1 \le v_2$, then

$$C(u_2; v_2) - C(u_2; v_1) - C(u_1; v_2) + C(u_1; v_1) \ge 0.$$

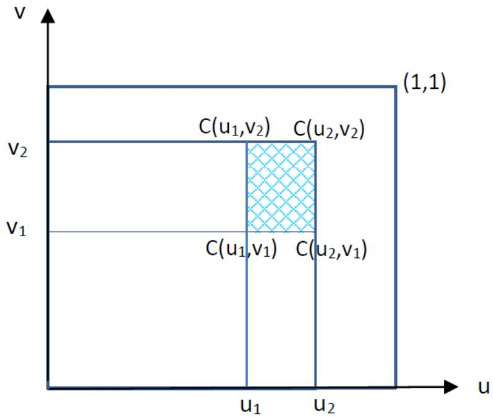

Fig. 1. Graphical illustration of 2-increasing property (highlighted area represent the left side of the inequality.

B. Examples of Copulas

Example 1. Gaussian copula. The most popular copula is the Gaussian copula. Gaussian Copula. X and Y are random variables following Gaussian distribution,

$$C_\rho(u,v) = F_\rho(F_X^{-1}(x), F_Y^{-1}(y))$$

Where, ρ is the correlation coefficient of the random variables X and Y. A Gaussian copula is show in Fig. 2.

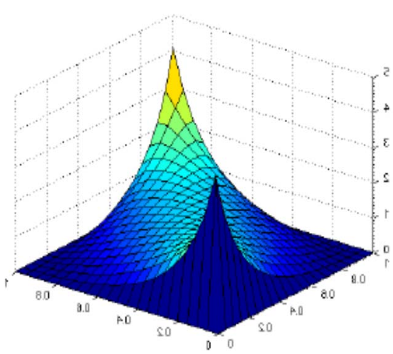

Fig. 2. Gaussian copula.

Fig.2 shows the cumulative probability of the Clayton copula. We can easily verify these properties. More examples of copulas are show in Table 1.

Example 2. Clayton copula. The Clayton copula is asymmetry. Symmetric density and thus unable to catch the skewness (3rd moment) associated with many risky, real world events. Asymmetric copulas perform better in modeling simultaneous extreme events in either or both tails.

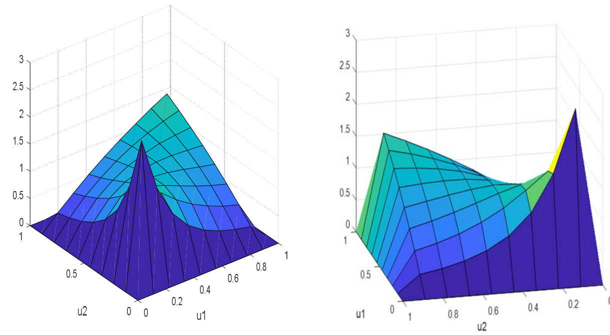

Fig. 3. Clayton copula (from different viewpoints).

Example 3. Independence copula. The independence copula is a generalization of the Gaussian copula when correlation matrix of the two random variables are identity matrix,

$$C_\theta(u,v) = uv + \theta uv(1-u)(1-v),$$

where parameter $\theta \in [0,1]$.

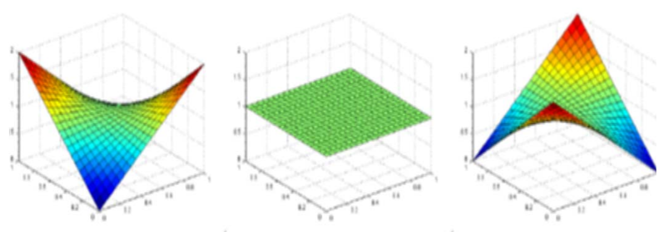

Fig. 4. Independence copula $\theta = -1,\ 0,\ \text{and}\ 1$.

Independence copula is used to define dependence classifier with θ as visual distance (as defined in optics).

III. CONSTRUCT COUPLAS

Using definition to construct copulas presents challenges at most of the cases. The following theorem plays a significant role of construct a class of copulas - Archimedean copulas.

Theorem [2, p. 111] Let $\varphi: \mathrm{I} \to [0, \infty]$ be a continuous, strictly decreasing function from I to [0;1] such that $\varphi(1) = 0$, and let $\varphi^{[-1]}$ be the pseudo-inverse of φ defined by

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) & , \text{if } 0 \le t \le \varphi(0), \\ 0 & , \text{if } \varphi(0) \le t \le \infty. \end{cases}$$

Then the function C from I^2 to I given by

$$C(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2)),$$

is a copula if and only if φ is convex.

Function φ is called the generator of the copula. Table 1 shows some well known copulas and their generators.

TABLE I. EXAMPLES OF COPULA

Name	Generator $\varphi(t)$	Generator Inverse $\varphi^{-1}(t)$	Parameter
Independent	$-\ln t$	e^{-t}	
Clayton	$t^{-\theta}-1$	$(1+t)^{-1/\theta}$	$\theta \in (0,\infty)$
Frank	$-\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$-\frac{\ln(1-(1-e^{-\theta})e^{-t})}{\theta}$	$\theta \in \mathbb{R}\backslash\{0\}$
Gumbel	$(-\ln t)^{\theta}$	$e^{-t^{1/\theta}}$	$\theta \in [1,\infty)$

[20]-[22] have some discussions for copulas in higher dimensions.

In practice, parametric and nonparametric methods are used to estimate copula. Fig. 5 shows a multiscale Frank copula we constructed using raised cosine wavelets, where $f_m(x,y)$ is typical wavelet approximation of Frank copula at level m.

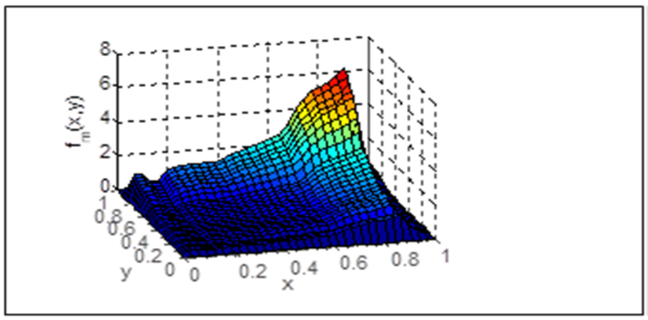

Fig. 5. A wavelet multiscal Frank copula (scale level m=3)

IV. SUMMARY AND FUTURE STUDY

In this article, we focus on brief review of the elements of copula theory and properties. More details for the algorithm to construct the multiscale copula will be reported separately. Our future study will be concentrated on two areas of applications: improving the accuracy of naïve Bayes classifier and coupling/decoupling algorithms in data integration modeling.

ACKNOWLEDGMENT

This research was supported in part by the Air Force Research Laboratory, the Sensors Directorate, through the Air Force Office of Scientific Research Summer Faculty Fellowship Program®, Contract Numbers FA8750-15-3-6003 and FA9550-15-0001.

REFERENCES

[1] A. Sklar, “Fonctions d´e repartition ´a n dimensions et leurs marges”. Publ. Inst. Stat. Univ. Paris, 8:229–231, 1959.

[2] R. Nelson, *An Introduction to Copulas, Second Edition*. Springer, 2006.

[3] J. Harry, *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability, vol.73, Chapman & Hall, London, 1997.

[4] A. J. Patton, Copula-Based Models for Financial Time Series. In T. Andersen, in Handbook of Financial Time Series. Springer Verlag, 2007.

[5] A. J. Patton, A review of copula models for economic time series, Journal of Multivariate Analysis, Vol. 110, pp 4-18 2012.

[6] X. Jin, Large portfolio risk management and optimal portfolio allocation with dynamic elliptical copulas, Dependence Modeling, 2018.

[7] P. Embrechts, F. Lindskog, and A. McNeil, “Modelling Dependence with Copulas and Applications to Risk Management”, in *Handbook of Heavy Tailed Distributions in Finance*. North-Holland, Elsevier.

[8] N. Kolev, U. dos Anjos, and B. Mendes, “Copulas: A Review and Recent Developments. Stochastic Models”, Vol 22, issue 4, 617-660.

[9] Eric Bouye, Valdo Durrleman, Ashkan Nikeghbali, Gael Riboulet & Thierry Roncalli. Copulas: an Open Field for Risk Management. Groupe de RechercheOpérationnelle, Crédit Lyonnais, Working Paper, 2001

[10] Eric Bouye, Valdo Durrleman, Ashkan Nikeghbali, Gael Riboulet & Thierry Roncalli. Copulas: an Open Field for Risk Management. Groupe de RechercheOpérationnelle, Crédit Lyonnais, Working Paper, 2001.

[11] Elidan G. (2013) Copulas in Machine Learning. In: Jaworski P., Durante F., Härdle W. (eds) Copulae in Mathematical and Quantitative Finance. Lecture Notes in Statistics, vol 213. Springer, Berlin, Heidelberg.

[12] L. Jiang, D. Wang, Z. Cai, and X. Yan, “Survey of Improving Naive Bayes for Classification”, in *Advanced Data Mining and Applications*, Lecture Notes in Computer Science book series, Vol. 4632, Springer, 2007, pp 134-145.

[13] L. Jiang, H. Zhang, and Z. Cai, “A Novel Bayes Model, Hidden Naive Bayes”, IEEE Transaction on Knowledge and Data Engineering, Vol 21, No. 10, pp. 1361 – 1371, 2009.

[14] P. Kaviani1 and S. Dhotre, “Short Survey on Naive Bayes Algorithm,” International Journal of Advance Engineering and Research Development, Vol. 4, Issue 11, 2017, pp 607-611.

[15] K. A. Vidhya, G. Aghila, “a survey of naïve Bayes machine learning approach in text document classification,” unpublished, in ArXiv 2010.

[16] A. Bauer and C. Czado, “Pair-Copula Bayesian Networks”, pp 1248-1271, 2016.

[17] K. Karra and L. Mili, “Hybrid Copula Bayesian Networks”, JMLR: Workshop and Conference Proceedings vol 52, 240-251, 2016.

[18] Md Rizwan Khan, S. K. Padhi, B. N. Sahu and S. Behera, “Non stationary signal analysis and classification using FTT transform and Naive Bayes classifier”, 2015 IEEE Power, Communication and Information Technology Conference (PCITC), 2015, pp 967 – 972.

[19] F. Autin, E. L. Pennec, and K. Tribouley, “Thresholding Methods to Estimate Copula Density”, Journal of Multivariate Analysis, 101(1), 200-222, 2010.

[20] F. Durante and C. Sempi. Copula *Theory: an Introduction. Copula Theory and Its Applications*, Lecture Notes in Statistics - Proceedings.Springer, in press, 2010.

[21] B. Choros, R. Ibragimov and E. Permiakova. *Copula Estimation. Workshop on Copula Theory and its Applications*. Springer, 2010.

[22] S. C. Chiou, Multivariate Continu.ous Time Models through Copula. Econometrics and S-tatistics Colloquium, University of Chicago, 2005.