科技部補助專題研究計畫成果報告
期末報告

一種多維度關聯結構的建構方式

計 畫 類 別 ： 個別型計畫
計 畫 編 號 ： MOST 103-2118-M-004-001-
執 行 期 間 ： 103 年 08 月 01 日至 104 年 07 月 31 日
執 行 單 位 ： 國立政治大學統計學系

計 畫 主 持 人 ： 黃子銘

計畫參與人員： 碩士班研究生-兼任助理人員：賴品霖
碩士班研究生-兼任助理人員：游博安
博士後研究：鄭宇翔

處 理 方 式 ：
1.公開資訊：本計畫涉及專利或其他智慧財產權，1 年後可公開查詢
2.「本研究」是否已有嚴重損及公共利益之發現：否
3.「本報告」是否建議提供政府單位施政參考：否

中 華 民 國 104 年 07 月 31 日

中 文 摘 要 ： 本研究提出一種基於 tensor product B-splines 之估計多維度關聯結構的方式. 這種關聯結構估計具有一致性, 而且容許在估計時加入已知的二維邊際關聯結構的資訊.

中文關鍵詞： 關聯結構, 多維度

英 文 摘 要 ： In this study, a nonparametric copula estimator based on tensor product B-splines is proposed. The proposed estimator allows for including information of bivariate marginals. The consistency of the proposed estimator is established.

英文關鍵詞： copula, multivariate, tensor product B-splines

Nonparametric Estimation for Multivariate Copulas Based on Tensor Product B-splines

July 31, 2015

1 Introduction

Copula is a function that fully describes the dependence among random variables. For a random vector $X = (X_1, \ldots, X_d)^\tau$ with distribution function F, the Sklar's Theorem (Sklar [10]; Schweizer and Sklar [9]) assures the existence of a distribution function C on $[0,1]^d$ such that

$$F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots F_d(x_d)\right), \tag{1}$$

where F_i is the distribution function for X_i for $i = 1, \ldots, d$. The function C is called the copula associated with X. (1) implies that the copula C gives the relationship among the components of X when the marginals are standardized. In fact, for the bivariate case, some common measures of dependence, such as Kendall's tau or Spearman's rho, can be explicitly expressed in terms of the copula. A comprehensive overview of copulas and their mathematical properties can be found in Nelsen [7].

Copulas can be estimated using nonparametric methods. However, in Genest, Gerber, Goovaerts, and Laeven [3], it is pointed out that parametric copulas are preferred for certain purposes such as prediction, yet the choices for parametric copulas are rather limited when $d > 2$. In order to fulfil the need of constructing of multivariate copulas, treelike constructions have been proposed, which allow for specifying a distribution with given marginals. See, for example, the pair-copula construction (PCC) demonstrated in Aas, Czado, Frigessi and Bakken [1] based on construction in Bedford and Cooke [2]. Earlier works by Joe ([5], [6]) also involve treelike constructions.

When implementing treelike constructions such as PCC, there are some problems to be solved. Choosing the conditional set of variables, and estimating the parameters in the specified copulas. The parameter estimation can be challenging as the number of parameters grow quickly with d (Hobæk Haff [4]).

In this project, a nonparametic estimator for multivariate copula density based on tensor product B-splines is proposed, which allows for including knowledge of some of the bivariate marginals. Tensor product B-splines are known to be flexible in modelling smooth functions ([8], [11]). As a result, the consistency of the proposed copula estimator is established. The details of the proposed estimator and the consistency property are presented in the following sections.

2 The proposed estimator based on tensor product B-splines

Suppose that $X_1, \ldots, X_n$ are IID observations with copula density $c \in L^2([0,1]^d)$. Let W_h be the space of tensor product linear B-splines with equally spaced knots, where $h = (h_1, \ldots, h_d)$ and h_j be the distance between two adjacent knots for the j-th dimension for $j = 1, \ldots, d$. Let $B_1, \ldots, B_k$ denote the tensor product B-spline basis functions for W_h. Then, we consider approximating c using

$$\alpha_1 B_1 + \cdots + \alpha_k B_k = \alpha^T B,$$

where $\alpha = (\alpha_1, \ldots, \alpha_k)^T$ and $B = (B_1, \ldots, B_k)^T$. The coefficient vector α will be estimated using a generalized method of moments approach with constraints that guarantee that $\alpha^T B$ is a copula density allowing for a penalty term that penalizes the L^2 distance between the bivariate marginals of c and the bivariate marginals of $\alpha^T B$, if the bivariate marginals of c are known. To describe the estimator for α, we introduce some notations. Let

$$P = \begin{pmatrix} \int_{[0,1]^d} B_1(u)B_1(u)du & \cdots & \int_{[0,1]^d} B_1(u)B_k(u)du \\ \vdots & \ddots & \vdots \\ \int_{[0,1]^d} B_k(u)B_1(u)du & \cdots & \int_{[0,1]^d} B_k(u)B_k(u)du \end{pmatrix},$$

and

$$M = \Big(\int_{[0,1]^d} c(u)B_1(u)du, \ldots, \int_{[0,1]^d} c(u)B_k(u)du \Big)^T.$$

Define $H_{ij} : L^2([0,1]^d) \to L^2([0,1]^2)$ to be the linear mapping such that

$$(H_{ij}f)(x_i, x_j) = \int_0^1 \cdots \int_0^1 f(x_1, \ldots, x_d)dx_1 \cdots dx_{i-1}dx_{i+1} \cdots dx_{j-1}dx_{j+1} \cdots dx_d,$$

and

$$Pen(c,\alpha) = \sum_{1\le i<j\le d} \int_0^1 \int_0^1 \left|(H_{ij}c)(u_i,u_j) - (H_{ij}\sum_{t=1}^k \alpha_t B_t)(u_i,u_j)\right|^2 du_i du_j.$$

When we have information of the bivariate marginals $H_{ij}c$: $1 \le i < j \le d$, $Pen(c,\alpha)$ measures the deviation of the estimated bivariate marginals from the true known marginals $H_{ij}c$s.

Suppose that the i-th observation $X_i = (X_{i,1}, \ldots, X_{,d})$. Let $\hat{F}_j$ be the empirical CDF based on $X_{1,j}$, …, $X_{n,j}$. That is,

$$\hat{F}_j(x) = \frac{1}{n}\sum_{i=1}^n I(X_{i,j} \le x),$$

where $I(X_{i,j} \le x)$ is the indicator function that is 1 when $X_{i,j} \le x$ and is 0 otherwise. Let

$$\hat{M}_i = \frac{1}{n}\sum_{j=1}^n B_i\big(\hat{F}_1(X_{j1}), \ldots, \hat{F}_d(X_{jd})\big)$$

be a moment estimator for M_i. The estimator for α is the minimizer of

$$(P\alpha - \hat{M})^T(P\alpha - \hat{M}) + \lambda Pen(c,\alpha)$$

under the constraints that $\alpha^T B$ is nonnegative and the marginals of $\alpha^T B$ are uniform density on $[0,1]$. Since $\alpha^T B$ is in the space of tensor product of linear B-splines and the marginals of $\alpha^T B$ are linear B-splines, the constraints of nonnegativity and uniform marginals is equalivalent to that $\alpha^T B$ is nonnegative at points in $\{(u_1, \ldots, u_d) : u_j \in S_j\}$ and the j-th marginal is equal to 1 on S_j for $1 \le j \le d$. Thus the above constrained minimization problem can be solved using quadratic programming.

3 Consistency of the proposed estimator

Let $\hat{\alpha}$ be the estimator of α mentioned in Section 2. In this section, we will show that $\hat{\alpha}$ is consistent under proper conditions.

We first establish an upper bound for the error $\|\hat{\alpha}^T B - c\|^2$. Suppose that there exists α^* such that

$$\|(\alpha^*)^T B - c\| \le \Delta_1$$

and α^* satisfies the restrictions that make $(\alpha^*)^T B$ a copula density. From the construction of $\hat{\alpha}$, we have

$$\begin{aligned} &(P\hat{\alpha} - \hat{M})^T(P\hat{\alpha} - \hat{M}) + \lambda Pen(c,\hat{\alpha}) \\ \le\ &(P\alpha^* - \hat{M})^T(P\alpha^* - \hat{M}) + \lambda Pen(c,\alpha^*). \end{aligned}$$

Let

$$\begin{aligned} I_0 &= (P\hat{\alpha} - \hat{M})^T(P\hat{\alpha} - \hat{M}) + \lambda Pen(c, \hat{\alpha}) \\ &\quad -[(P\alpha^* - \hat{M})^T(P\alpha^* - \hat{M}) + \lambda Pen(c, \alpha^*)] \\ &\quad -[(P\hat{\alpha} - M)^T(P\hat{\alpha} - M) + \lambda Pen(c, \hat{\alpha})] \\ &\quad +(P\alpha^* - M)^T(P\alpha^* - M) + \lambda Pen(c, \alpha^*), \end{aligned}$$

then

$$\begin{aligned} & (P\hat{\alpha} - M)^T(P\hat{\alpha} - M) + \lambda Pen(c, \hat{\alpha}) \\ \le\ & (P\alpha^* - M)^T(P\alpha^* - M) + \lambda Pen(c, \alpha^*) + |I_0|. \end{aligned}$$

Therefore,

$$\begin{aligned} & \min(eigen(P))^2(\hat{\alpha} - \alpha^*)^T(\hat{\alpha} - \alpha^*) \\ & \le (\hat{\alpha} - \alpha^*)^T P^T P(\hat{\alpha} - \alpha^*) \\ & \le 2(P\hat{\alpha} - M)^T(P\hat{\alpha} - M) + 2(P\alpha^* - M)^T(P\alpha^* - M) \\ & \le 4(P\alpha^* - M)^T(P\alpha^* - M) + 2\lambda Pen(c, \alpha^*) + 2|I_0| \\ & \le (4 + \lambda(d^2 - d))\Delta_1^2 + 2|I_0|, \end{aligned}$$

where $eigen(P)$ is the vector of eigen values of P. As a result, we have

$$\begin{aligned} \|\hat{\alpha}^T B - c\|^2 &\le 2\|(\hat{\alpha} - \alpha^*)^T B\|^2 + 2\|(\alpha^*)^T B - c\|^2 \\ &\le 2(\hat{\alpha} - \alpha^*)^T P(\hat{\alpha} - \alpha^*) + 2\Delta_1^2 \\ &\le 2\big((4 + \lambda(d^2 - d))\Delta_1^2 + 2|I_0|\big)/\min(eigen(P))^2 + 2\Delta_1^2. \end{aligned}$$

Let $\delta_M = (\hat{M} - M)^T(\hat{M} - M)$, then $|I_0|$ can be controlled using Lemmas 1 and 2:

Lemma 1

$$|I_0| = O(1)(\delta_M/\kappa_1^*)^{1/2}\Big(1 + \delta_M^{1/2}\Big),$$

where κ_1^* is the smallest eigenvalue of $P^T P$.

Lemma 2 *For $t = 1, \ldots, k$, $\hat{M}_t = \frac{1}{n}\sum_{i=1}^{n} B_t\big(\hat{F}_1(X_{i,1}), \ldots, \hat{F}_d(X_{i,d})\big)$. Let $h_* = \min(h_1, \ldots, h_d)$. Then $E\left(\sum_{t=1}^{k} |\hat{M}_t - M_t|^2\right) = O\left(\frac{k}{n}\right)\left(1 + \left(\frac{d}{h_*}\right)^2\right)$.*

The control of Δ_1 is based on Theorem 1.

Theorem 1 *Suppose that* $W_h \cap \{f : Af = \eta\} \neq \emptyset$. *Let* V *denote the set* $\{f \in L^2([0,1]^d) : f \geq 0\}$ *and* $V^{(0)}$ *be the interior of* V. *Suppose that* $g \in \{f \in V^{(0)} : Af = \eta\}$ *and* ε *is a positive number such that* $B(g, \varepsilon) \stackrel{def}{=} \{f \in L^2([0,1]^d) : \|f - g\| < \varepsilon\} \subset V^{(0)}$ *and*

$$2d\|\bar{g}_w - g\| < \frac{\varepsilon}{2}$$

for some $\bar{g}_w \in W_h$. *Then for* $f \in V \cap \{f \in L^2([0,1]^d) : Af = \eta\}$ *and* $\bar{f}_w \in W_h$, *there exists* $f_w \in V \cap \{f \in W_h : Af = \eta\}$ *such that*

$$\|f_w - f\| \leq 2d\Big(1 + \frac{2}{\varepsilon}(\|f\| + \|g\| + \varepsilon)\Big)\|\bar{f}_w - f\|.$$

From the approximation property of tensor product B-splines ([8]), if each component of h is small enough, then there exists some $\bar{f}_w \in W_h$ such that $\|\bar{f}_w - c\|$ is small. From Theorem 1, there exists some f_w in the constrained space such that $\|f_w - c\|$ is small and Δ_1 can be made small. From the above discussion, if the components of h tend to 0 at slow rate so that k grows at a rate that is much less than n, then $\|\hat{\alpha}^T B - c\|^2$ is $o_p(1)$ and the proposed estimator is consistent.

References

[1] Kjersti Aas, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics & Economics*, 44(2):182–198, 2009.

[2] Tim Bedford and Roger M. Cooke. Vines – A new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068, 2002.

[3] Christian Genest, Hans U. Gerber, Marc J. Goovaerts, and Roger J. A. Laeven. Editorial to the special issue on modeling and measurement of multivariate risk in insurance and finance. *Insurance: Mathematics & Economics*, 44(2):143–145, 2009.

[4] Ingrid Hobæk Haff. Parameter estimation for pair-copula constructions. *Bernoulli*, 19(2):462–491, 2013.

[5] Harry Joe. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In Ludger Rüschendorf, Ludger Ruschendorf, Ludger Rueschendorf, Berthold Schweizer, and Michael D. Taylor, editors, *Distributions with Fixed Marginals and Related Topics (IMS Lecture Notes Monograph Series, Volume 28)*, pages 120–141. Institute of Mathematical Statistics, 1996.

[6] Harry Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall Ltd, 1997.

[7] Roger B. Nelson. *An Introduction to Copulas*, volume 139 of *Lecture Notes in Statistics*. Springer Verlag, New York, 1999.

[8] Larry L. Schumaker. *Spline Functions: Basic Theory*. Wiley-Interscience, 1981.

[9] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. Elsevier/North-Holland [Elsevier Science Publishing Co., New York; North-Holland Publishing Co., Amsterdam], 1983.

[10] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris, 1959.

[11] Charles J. Stone. The use of polynomial splines and their tensor products in multivariate function estimation (Disc: P171-184). *The Annals of Statistics*, 22:118–171, 1994.

科技部補助計畫衍生研發成果推廣資料表

日期:2015/07/31

<table>
<tr><td rowspan="3">科技部補助計畫</td><td>計畫名稱: 一種多維度關聯結構的建構方式</td></tr>
<tr><td>計畫主持人: 黃子銘</td></tr>
<tr><td>計畫編號: 103-2118-M-004-001- 學門領域: 統計方法</td></tr>
<tr><td colspan="2">無研發成果推廣資料</td></tr>
</table>

103年度專題研究計畫研究成果彙整表

計畫主持人：黃子銘			計畫編號：103-2118-M-004-001-				
計畫名稱：一種多維度關聯結構的建構方式							
成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	計畫成果正在改寫，準備投稿至國際期刊

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

科技部補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

　　□實驗失敗

　　□因故實驗中斷

　　□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：□已發表 □未發表之文稿 ■撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 □洽談中 ■無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

本研究提出的多維度關聯結構估計方式，可將已知低維度的資訊加入，更切合應用上的需要.