

多目标决策研究:基于藤 copula 和 SMAA 方法^{*}

邓维,叶五一[†],杨锋
(中国科学技术大学管理学院 管理科学系,合肥 230026)
(2019 年 12 月 31 日收稿;2020 年 4 月 27 日收修改稿)

Deng W, Ye W Y, Yang F. Using copula method with SMAA in decision-making analysis[J]. Journal of University of Chinese Academy of Sciences, 2022, 39(4): 449-462. DOI: 10. 7523/j.ucas. 2020. 0061.

摘 要 运用藤 copula 构造模型分析多组决策变量之间的相互依赖以及不确定性,创新地将 SMAA 方法与 copula 相依性分析结合在一起,得到更优良更有效的随机多准则可接受性分析方法。介绍基本的藤 copula 建模分析和 SMAA 的方法,展示两种方法结合使用的完整步骤,并在数据模拟分析实践中,通过对比新方法与原有的 SMAA 方法,得到在不同相互依赖结构下的表现结果,证明了新方法更普适、更准确的优点。
关键词 决策分析;随机多标准可接受度分析;相依分析;连接函数;藤结构
中图分类号:C934 **文献标志码:**A **DOI:**10. 7523/j.ucas. 2020. 0061

Using copula method with SMAA in decision-making analysis

DENG Wei, YE Wuyi, YANG Feng
(Department of Management Science, School of Management, University of Science and Technology of China, Hefei 230026, China)

Abstract Stochastic multi-criteria acceptability analysis (SMAA) is a series of methods for multi-criteria decision making. It is used to give decision opinions on the problem of uncertain or missing preferences of decision makers, the uncertainty of decision variables and incomplete or missing preference information can be represented by probability distribution. This method improves the past methods by considering inversely what specific parameters each decision outcome is determined by. However, the existing SMAA methods does not take into account the impact of the interdependence between these variables on the analysis results, or simply using a simple model such as a Gaussian distribution to describe these interdependencies, so that the analysis results do not have sufficient scientific basis. In order to fix this fault, this paper create an innovation by combining copula dependency analysis with stochastic multi-criteria acceptability analysis method, and uses vine copula modeling method to describe the uncertainty of decision variables and their interdependence, making the SMAA method complete and more effective. This paper introduces the basic methods of SMAA and vine copula separately, and gives the specific steps of combining the two methods. In simulation experiments, a comparative analysis is conducted between the original SMAA method and the method given in the paper to show their advantages and disadvantages under different

^{*} 国家自然科学基金面上项目(71973133)资助

[†] 通信作者, E-mail: wyue@ustc.edu.cn

dependency structures.

Keywords decision-making analysis; stochastic multi-criteria acceptability analysis (SMAA); dependency; copula; vines

现实生活中有很多决策问题,如商铺选址和公交地铁系统设计站点等,需要考虑很多影响因素之间的互相关联和限制的问题。这类选择问题的共同处在于可选项是有限的,而最后决策的优劣性的评估标准也是有限的。通过这些评估标准比较各个选择,能够得到最优选,其次可以从整体上得到各个选择的评级和分类。这类对多种标准的决策问题进行评价定级、分类的行为称作多标准决策分析(decision analysis by multiple criteria, MCDA)。MCDA 的选择及其评估标准是有限的,这些评估标准可以是有序数字,那么分析后将获得不同选择的评级和排序;评估标准也可以是基础数值,即选择被量化后的具体值。通过累计这些基数或序数的标准,并考虑决策者们(decision makers, DM)对不同标准的重视和偏好程度,对每一个选择作出整体性的评估,是 MCDA 的最终目的。由此可见,MCDA 不仅依靠对标准值的量化规则,还依靠不同决策者主观上的偏好,以及评估中用到的模型和模型中的各项设定参数。如果可以通过各种 MCDA 方法对决策问题进行分析以获得决策建议,那么将这些方法应用于决策是有用的。因此,决策者需要熟悉使用各种多准则决策分析技术的先决条件以及灵活应用不同方法的优缺点。

MCDA 研究是一个较新的领域,很多方法仅用于处理实际生活中某些特定领域的特定多目标决策问题,大致分为两类。一类是 Keeney 等^[1]发展出的多属性效用方法(multiattribute utility theory, MAUT),是广泛应用至今的主要方法之一,而另一大类被广泛采用的多标准决策分析方法是高分法(outranking methods),如 Roy^[2]提出的 ELECTRE 方法,Brans 和 Mareschal^[3]提出的 PROMETHEE 方法,以及 Xu^[4]给出的 SIR 方法等。MAUT 理论的构建为当今 MCDA 的发展打下了根基,相对于高分法,其应用了更多相对复杂的数学模型;而高分法,比如 ELECTRE 方法,能够从更多不同的角度对偏好信息建模,也具有一定优势。这两类方法都得到了广大决策者的青睐和应用。

MAUT 基于效用函数,是目前 MCDA 中应用

时间最长的方法,并广受好评。这类方法首先建立效用函数,然后根据标准的值得出效用评分,最后根据得出的效用评分评估不同选择的优劣。近年来,效用函数方法广泛应用,但同时也出现了越来越多无法得到模型所需具体参数的情况,例如,决策者不愿意透露其偏好信息,导致缺少必要的建模参数,或是评估这些选择的各个标准的值本身就是不确定或者不准确的。原始的 MAUT 和高分法在这一方面皆有不足,因此,应对模型参数的不确定性或不准确性,需要对以往的效用理论方法进行改进。逆向推断进行决策分析是一个可行的改进思路,也就是说,与其去咨询或寻找具体决策问题所用模型中的具体参数,不如反过来考虑每一种决策结果是由哪些具体参数所决定。于是,随机多标准可接受度分析(stochastic multi-criteria acceptability analysis, SMAA)方法应运而生并在近些年飞速发展。它就是这样一种逆向分析的方法,允许标准值和偏好信息的不确定性,从而解决了很多其他方法所遇到的困难。

不同的 SMAA 方法可以用来处理多标准决策中的主要 3 种情形:选优、评级和分类。这类方法包含了一系列多标准决策方法,其优势在于扩展了决策分析的角度,即反向分析所有可能的参数空间中的取值,并且允许标准值和偏好信息的缺失,甚至在没有任何偏好信息的极端情况下也可以使用。这里所说的标准值和偏好信息的缺失包含 3 种可能的情形,即不完整、不准确和不确定。不完整意味着部分信息未获得,不准确意味着信息虽获得但准确性不足,不确定意味着虽然决策问题的相关观察者给出了他们完整且准确的信息,但这些信息并不可靠。偏好信息可量化为不同标准所占的权重,SMAA 方法基于探索权重向量空间来得到使每个备选方案成为最优方案的偏好信息,或者为特定选择方案提供其取得不同评级的概率。例如,相比于给出一个整体评级,SMAA-2 方法提供了各个选择取得特定评级的概率。

Lahdelma 等^[5]首次提出 SMAA 方法。Lahdelma 和 Salminen^[6]将秩可接受度的概念引入 SMAA 方法中,发展出 SMAA-2 方法。该方法分析得到的

主要结果是秩可接受度指数、中心权重向量以及在中心权重向量下的置信因子。秩可接受度指数描述了使得选择方案位于特定等级(秩)的不同偏好的所占比例,中心权重向量表示使每个选择成为最优的偏好的典型代表,而置信因子衡量这些秩可接受度是否足够可信,从而做出明智的决定。Lahdelma 等^[7]将 SMAA 方法进一步扩展,发展出 SMAA-O 方法,这一方法将序数作为标准值,即决策者对每一项标准进行评级并得到序数,再进行 SMAA 分析。上述几种 SMAA 方法都是针对选择或评级之类的决策问题,而 Yu^[8]提出的 ELECTRE TRI 方法是一类针对分类问题的决策方法,在现实生活中有广泛的应用,Tervonen 等^[9]将其扩展为 SMAA-TRI 方法,其目的在于允许参数的缺失和不确定,并研究 ELECTRE TRI 方法分析结果的稳健性,其分析结果展现了哪些参数取值将一种选择归为特定的分类中。Tervonen 等^[10]还将类似的扩展应用到另一类常用的评级方法——ELECTRE III 方法中,给出了 SMAA-III 方法,同样解决了参数不确定或不准确所带来的难题,并且可以很方便地进行稳健性分析。数据包络分析(data envelopment analysis, DEA)也是基于效用理论的多标准决策方法,它将不同的选择判定为有效的或无效的选择,和 SMAA 类似,它也用一个总效用函数来描述每一种选择,其中几个标准的不同重要性则由权重来体现。Lahdelma 和 Salminen^[11]考虑到将 DEA 和 SMAA-2 方法结合使用,提出 SMAA-D 方法,它可被视为是将 DEA 类型的效用函数应用到 SMAA-2 中,用来提供随机的有效性的度量。前景理论(prospect theory)是经济学研究中广泛应用的思想,但很少被应用到决策分析中,Lahdelma 和 Salminen^[12]首次将前景理论和 SMAA 方法相结合,提出了 SMAA-P 方法,该方法以前景理论所认为的损失重于得到的观点为出发点进行 SMAA 分析。由此可见,SMAA 方法的核心思想可以与以往各类多标准决策方法相结合。除上述这些方法外,Corrente 等^[13]提出的 SMAA-PROMETHEE 方法,则是将 SMAA 与另一种常用的多标准决策方法——PROMETHEE 方法相结合,进而探索与决策者提供的某些偏好信息一致相容的所有权重。根据应用场景的不同,SMAA 方法会涉及不同的关于概率分布的积分,并且可能非常复杂。通常,积分也具有很高维数。基于离散化每个维度分布

的数值积分方法是不可行的,因为所需的计算量取决于维数。因此,相比于尝试获得积分的精确值,应用蒙特卡罗模拟来获得足够准确的近似值成为了常用的手段。Tervonen 和 Lahdelma^[14]给出了有关模拟算法的完整说明,并研究了计算的准确性以及复杂性问题。

多种多样的 SMAA 方法已经在各种现实生活中的决策问题中有所应用,Hokkanen 等^[15]研究如何确定总体计划的实施顺序的问题,并且解决了港口引证的问题^[16],还研究了选择污染土壤的清除剂的决策^[17];Lahdelma 等^[18]研究废物处理设施引证的问题;Tervonen 等^[19]设计了电梯规划方案,还曾设计了溢油应急响应框架^[9];Kangas 等^[20]给出了森林规划的决策建议。在国内近些年的研究中,SMAA 方法也在各个领域发挥出了作用。Xia 等^[21]用 SMAA 方法为物流服务供应商提供了针对物流配送中心的选址方面的建议;Zhu 等^[22]研究了水库洪水防治的操作策略。

改革开放后中国经济飞速发展,时局瞬息万变,决策者们在营销策划、产品更替、项目投资等等领域中,面临选择时需要考虑越来越多的因素和标准,而这些决策因素和标准相互之间联系紧密,在数据结构上存在着复杂的相依关系,在应用 SMAA 方法研究这些问题时,可以通过概率分布来表示部分或全部缺失的偏好信息以及标准值的不确定性,然而这些相依性则很难被准确地量化。尽管 SMAA 方法已经发展成熟,出现了很多形式和变化,但现有方法大都不能很有效地对这些相依结构的信息进行量化和分析。Lahdelma 等^[23]提出用独立分布来处理高维数据是不合理的,必须要充分考虑这些变量之间相依性的信息,考虑到高斯分布被广泛应用到各个领域,他们首次提出将高斯分布(多元正态模型)运用到 SMAA 方法中。虽然独立分布并不可取,但是,文献[23]并未充分论述采用高斯分布的合理性。实际上,高斯分布局限于变量之间是线性的相依关系的情形,对于非线性结构,就不太准确和适用。弥补这方面不足是本文研究的出发点。考虑到 copula 函数是分析变量间相依关系最有利的工具,擅长处理多元变量间各种类型的相依关系,已经在金融市场、风险传递等领域广泛应用,因而本文从这一点出发,将 copula 相依分析与 SMAA 方法首创性地相结合来改善以往的 SMAA 方法。一些 copula 理论研究的前沿文献为本文提供了相应的

理论基础。Genest 等^[24-25]总结了关于二元 copula 函数的拟合优度检验和参数估计的方法。以 Joe^[26]提出的关于给定边际分布的多元分布函数的构造理论为基础, Bedford 和 Cooke^[27]提出了藤结构的概念, 通过 pair-copula 模型, 首次解决了多维变量之间相依关系的问题。Aas 等^[28]首次用统计方法推断出藤 copula 的计算公式。相比较 C 族藤和 D 族藤的 copula 函数, R 藤 copula 有着更加灵活多样的相依结构。随着相关研究的进展, R 藤模型越来越多地应用于各个领域, 来处理多维变量的相依结构。

本文首先对 SMAA-2 的方法进行回顾, 介绍 R 藤结构的构造算法和 copula 函数的拟合方法, 总结 SMAA-2 和 R 藤 copula 方法相结合的应用步骤。然后进行模拟试验分析, 模拟出一万个样本, 这些样本都是包含 2 个评估标准和 3 个选择方案的 6 维决策变量。应用藤 copula 拟合方法, 拟合出 6 维决策变量的联合概率分布, 来量化其相依结构和不确定性, 并结合使用 SMAA-2 方法, 根据联合概率分布计算出备选方案的各等级的秩可接受度。最后, 在两种不同相依结构的决策变量下, 分别对采用多元正态模型处理相依关系的方法以及本文提出的新方法进行对比分析, 比较出在不同情况下, 这些方法是否能够适用, 以及它们各自的优缺点。

1 SMAA 基础理论

SMAA 是一种多标准决策方法。在应用 SMAA 方法时, 决策者不需要将他们的偏好信息明确地表达出来, 相反, 基于对权重空间的探索, 该方法可以描述出每个备选方案担任最优解的概率以及计算的可靠性。决策变量的不确定或不准确的值由概率分布来表示。本文应用到的 SMAA-2 方法, 通过计算各个选择的所有排序的接受度, 也称为秩可接受度, 综合地判断每个选择的优劣, 最早由 Lahdelma 和 Salminen^[6]提出该方法。

SMAA-2 应用对权重空间分析和划分来描述每种选择方案, 即将总权重空间划分为使某种选择成为最优选择的子权重空间, 以及将其置于其他各个等级的子权重空间。一般地, 决策问题表示为一组 m 个选择 $\{x_1, x_2, \dots, x_m\}$ 根据 n 个标准进行评估, 其评估的优劣用各自定义的效用函数来表示, 进而研究如何使得加权后总效用最优。

权重向量即代表了 DM 的偏好结构, 第 i 个选择的加权效用由实值效用函数 $u(x_i, \mathbf{w})$ 表示。这个总效用函数通过使用权重向量 \mathbf{w} 来量化 DM 的主观偏好, 从而将不同的选择方案映射到一个实际的值。SMAA-2 方法适用于决策变量和权重均未精确了解的情况。不确定的决策变量由随机变量 ξ_{ij} 表示, 它是空间 $X \subseteq \mathbb{R}^{m \times n}$ 中的随机变量, 其联合密度函数记为 $f_X(\boldsymbol{\xi})$ 。决策者的未知或部分已知的偏好信息由在总权重空间 W 中的权重向量 \mathbf{w} 来表示, 记 \mathbf{w} 的联合密度函数为 $f_W(\mathbf{w})$ 。在偏好信息可获取时, 偏好信息相当于权重空间 W 上的一个约束条件。在偏好信息完全缺失时, 则权重采用均匀分布, 即 $f_W(\mathbf{w}) = \frac{1}{\text{vol}(W)}$ 。权重空间也可以根据需要来定义, 但通常权重是非负的和标准化的, 即

$$W = \{\mathbf{w} \in \mathbb{R}^n : w_j \geq 0, \sum_{j=1}^n w_j = 1\}. \quad (1)$$

效用函数用于将随机决策变量和权重分布映射到效用分布 $u(\boldsymbol{\xi}_i, \mathbf{w})$ 。基于效用分布, 可以通过排序函数定义每个选择的等级, 其取值从最佳等级 (1) 到最差等级 (m):

$$\text{rank}(i, \boldsymbol{\xi}, \mathbf{w}) = 1 + \sum_{k=1}^m \rho(u(\boldsymbol{\xi}_k, \mathbf{w}) > u(\boldsymbol{\xi}_i, \mathbf{w})), \quad (2)$$

其中: $\rho(\text{true}) = 1, \rho(\text{false}) = 0$ 。SMAA-2 是以分析最优等级 (秩) 的随机权重集合为基础的, 其秩为 r 的随机权重子空间定义为

$$W_r^i(\boldsymbol{\xi}) = \{\mathbf{w} \in W : \text{rank}(i, \boldsymbol{\xi}, \mathbf{w}) = r\}, \quad (3)$$

这个 W_r^i 权重空间里的任意权重都使选择 x_i 的被排在第 r 个等级。

SMAA-2 的第一个描述性指标是秩可接受度 b_i^r , 其计算方法为

$$b_i^r = \int_{\boldsymbol{\xi} \in X} f_X(\boldsymbol{\xi}) \int_{\mathbf{w} \in W_r^i} f_W(\mathbf{w}) d\mathbf{w} d\boldsymbol{\xi}, \quad (4)$$

式中对 \mathbf{w} 的积分表示 W_r^i 权重空间占总权重空间的比例, 之后对 $\boldsymbol{\xi}$ 进行积分得到它的期望值, 直观地理解也就是使选择 x_i 排序为 r 的权重空间占总权重空间的比例的期望值。秩可接受度可以直接用于多标准选择的评估, 当秩取 $r = 1$ 时, 简称为可接受度, 记为 a_i 。一般地, 认为可接受度越大, 那么对应的选择越可靠。然而在一些特殊情况下, 例如存在两种非常相近的选择时, 它们各自的可接受度会被彼此“剥削”而各得一半, 使得每个

可接受度都不高,因此仅仅考虑可接受度 a_i 是不够的,对于这些特殊情况,将 b_i^r 考虑进来,综合考虑使得每种选择排序靠前的可能性,将得到更好的分析结果,这也是 SMAA-2 方法相比于 SMAA 方法所改进的地方。对于大样本问题,Lahdelma 和 Salminen^[6] 建议用迭代的形式,定义其 k 最佳迭代秩可接受度为

$$a_i^k = \sum_{r=1}^n b_i^r. \tag{5}$$

在用秩可接受度来比较选择方案的优劣性时,除了直接比较上式中 $a_i^r(r$ 的值由决策者选择)的大小,还可以采用一种互补的方法,对秩可接受度求加权和,从而得到每个选择的整体可接受性指数 a_i^h ,即

$$a_i^h = \sum_r \alpha^r b_i^r. \tag{6}$$

这里,如何选择每个秩的权重,也需要有所考量,通常排在前面的权重应大于秩靠后的权重。

中心权重向量 w_i^c 是使得选择方案成为最优选择的权重向量空间的期望重心(质心)。中心权重向量表示支持此决策方案的决策者的“典型”偏好。可以向决策者呈现不同选择方案的中心权重,以帮助他们理解不同权重如何与决策分析结果的不同最优选择相对应。中心权重向量 w_i^c 在数值上的计算公式如下

$$w_i^c = \int_{\xi \in X} f_X(\xi) \int_{w \in W_i^1} f_W(w) w dw d\xi / b_i^1. \tag{7}$$

置信因子 p_i^c 是当使用代表某个选择的中心权重向量作为偏好信息时,该选择确实成为最优选择的概率。置信因子计算公式如下

$$p_i^c = \int_{\xi \in X, \text{rank}(i, \xi, w) = 1} f_X(\xi) d\xi. \tag{8}$$

事实上,对于任何给定的权重向量,都可以类似地计算置信因子。置信因子可以衡量秩可接受度能否足够准确地辨别选择方案是否有效。如果一个选择的可接受度很高,但置信因子较低的话,那么则需要综合考虑其各个秩可接受度再做分析。相反的,如果一个选择可接受度非常低,而置信因子又非常高的话,那么则有充分的理由排除这个选择。

2 藤 copula 方法

2.1 多元 copula 的 PCC 构造

copula 函数是研究随机变量间相依关系的有力工具,通过它可以将边缘分布和变量间的相依结构分开处理。copula 的概念最早由 Sklar^[29] 提

出,它用来描述多维分布函数和低维边缘分布之间的关系。Sklar 给出的定理成为这一领域理论研究的基础,也是概率空间理论中的重要定理之一。为方便后文的理解,这里将 Sklar 定理复述如下:

定理 2.1 令 F 为联合分布函数,且边缘分布为 F_1, \dots, F_n , 则存在 copula 函数 $C: [0, 1]^n \rightarrow [0, 1]$, 使得对所有 $x_1, \dots, x_n \in \mathbb{R} = [-\infty, \infty]$, 有

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)).$$

如果边缘分布都是连续的,那么函数 C 是唯一的。 C 称为分布 F 的(唯一) copula 函数。

由 copula 函数可定义联合分布的密度函数如下

$$f(x_1, \dots, x_n) = c_{12 \dots n}(F_1(x_1), \dots, F_n(x_n)) \prod_{i=1}^n f_n(x_i), \tag{9}$$

其中 $c_{12 \dots n} = \frac{\partial^n}{\partial u_1 \dots \partial u_n} C(u_1, \dots, u_n)$ 称为 copula 密度函数。

构造高维情形下的 copula 通常是件很困难的事情。已有的具有显示表达式的多参数多元 copula 函数并不多,并且特定的函数形式未必能够很好地拟合现实生活中的数据。以 Joe^[26] 的方法为基础,Bedford 和 Cooke^[27] 给出一种能灵活地利用二元 copula 函数构造高维联合分布的方法,将高维数据的处理转化为对多组二元数据的处理。而对于用 copula 函数拟合二元数据,则有相当完备的方法,大致分为 3 步:第一是对模型参数的估计,第二是对估计结果进行假设检验,第三是对不同的备选模型进行比较。

对模型参数的估计通常采用秩相关系数^[30], 或者采用极大似然法。其对数似然函数可以表示为

$$l(\theta) = \sum_{i=1}^n \log c_{\theta}(F(X_i), G(Y_i)), \tag{10}$$

其中 c_{θ} 是 copula 密度函数,这里的极大似然性是对式(10)中全部参数计算的,包括边缘分布 F, G 的未知参数。在实际应用中,为简化运算,可以利用经验分布来代替边缘分布,经验分布采用如下形式

$$F_n(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{I}(X_i \leq x),$$
$$G_n(y) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{I}(Y_i \leq y),$$

其中采用 $n+1$ 是为了保证分布函数值在区间

$[0, 1)$ 内部, \mathbb{I} 是示性函数, 当括号内不等式成立时取 1, 否则取 0. 用 $(F_n(X_i), G_n(Y_i))$ 替代似然函数 (10) 中的 $F(X_i), G(Y_i)$, 再对其最大化计算出极大似然估计, 省去了对边缘分布的参数估计, 这也称为伪似然法。

对于模型的假设检验, 通常采用一类 goodness-of-fitting 检验方法, 它们可以用来拒绝不好的拟合模型, 已经被广泛应用于各类统计研究中。在采用多种参数型 copula 和多种参数估计的方法得到备选二元 copula 函数模型后, 需要进行比较, 选取拟合结果最好的作为最终模型。AIC 准则 (Akaike information criterion) 是使用最为广泛的对模型进行选择的方法, 对于二元 copula 的建模, 其 AIC 定义为

$$\text{AIC} = -2 \sum_{i=1}^n \ln[c_\theta(U_{i1}, U_{i2})] + 2k. \quad (11)$$

根据 AIC 准则将选取 AIC 最小的模型作为最优模型。

以二元 copula 为基础, 可将多元随机变量转换为多个二元 copula 函数的乘积形式, 进而用二元的方法来处理高维情形的问题。对于多元随机变量, 其联合密度函数可以分解为条件边缘密度的乘积的形式

$$f(x_1, \dots, x_n) = f(x_1) \cdot f(x_2 | x_1) \cdot f(x_3 | x_1, x_2) \cdots \cdot f(x_n | x_1, \dots, x_{n-1}), \quad (12)$$

其中等式右边的第 2 项可改写为

$$f(x_2 | x_1) = \frac{f(x_1, x_2)}{f(x_1)} = c(F_1(x_1), F_2(x_2)) \cdot f(x_2).$$

类似地, 式 (12) 中等号右边的第 3 项可以展开为

$$\begin{aligned} f(x_3 | x_1, x_2) &= \frac{f(x_1, x_2, x_3)}{f(x_1, x_2)} = \frac{f(x_2, x_3 | x_1)}{f(x_2 | x_1)} \\ &= c_{23|1}(F_{2|1}(x_2 | x_1), F_{3|1}(x_3 | x_1)) \cdot \\ &\quad f(x_3 | x_1) = \frac{f(x_1, x_3 | x_2)}{f(x_1 | x_2)} \\ &= c_{13|2}(F_{1|2}(x_1 | x_2), F_{3|2}(x_3 | x_2)) \cdot \\ &\quad f(x_3 | x_2). \end{aligned}$$

其中 $c_{23|1}$ 和 $c_{13|2}$ 分别表示连接条件分布 $F_{2|1}$ 和 $F_{3|1}$, 以及连接条件分布 $F_{1|2}$ 和 $F_{3|2}$ 的 copula 函数的密度。一般地, 关于条件密度的分解有下述公式

$$f(x_i \mathbf{v}) = c_{x_i, x_j | \mathbf{v}_{-j}}(F(x_i | \mathbf{v}_{-j}), F(x_j | \mathbf{v}_{-j})) \cdot f(x_i | \mathbf{v}_{-j}), \quad (13)$$

其中: $i, j = 1, 2, \dots, n$; \mathbf{v} 是集合 $\{x_1, x_2, \dots, x_n\}$ 的任意子集且包含 x_j 但不包含 x_i ; 而 \mathbf{v}_{-j} 即表示从 \mathbf{v}

中除去 x_j 所得的集合。不难看出, 利用式 (13) 将式 (12) 中的条件密度每做一次分解, 将产生一个新 copula 密度, 而条件集的元素则减少一个。不断地重复这样的分解步骤, 即可将联合密度 $f(x_1, \dots, x_n)$ 展开为两部分, 一部分为边缘密度的乘积, 即 $\prod_{i=1}^n f_i(x_i)$, 而另一部分则是 $n(n-1)/2$ 个二元 copula 密度函数的乘积。这样就构造出了一个多元联合密度函数的二元 copula 密度分解 (pair copula construction, PCC)。注意到式 (13) 中, 对 j 的不同取值, 可得出不同的分解形式, 同样, 式 (12) 中条件密度的分解也不是唯一的, 因此可以得出多种不同的 PCC。

通过构建 PCC 可以看出, 对于多元随机变量的建模问题被分解为对一系列二元相依关系的建模。例如对 c_{12} 的拟合, 只需考虑如何构造出边缘分布 $F_1(X_1)$ 和 $F_2(X_2)$ 的 copula 函数, 而这只需获取 X_1 和 X_2 的一系列独立样本, 即可采用多种多样的方法去实现很好的拟合。进一步, 还可以很好地拟合 $c_{23|1}$, 这只需考虑构造出条件分布 $F_{2|1}$ 和 $F_{3|1}$ 的 copula 函数, 但这一步的样本数据由上一步中的 copula 函数转换而来, 将用到如下公式

$$F(x | \mathbf{v}) = \frac{\partial C_{x, \mathbf{v}}(F(x), F(\mathbf{v}))}{\partial F(\mathbf{v})}.$$

将原始样本——代入式中, 即得到转换后的样本数据。以此类推, 由下述公式

$$F(x_i \mathbf{v}) = \frac{\partial C_{x_i, \mathbf{v}_{-j}}(F(x_i | \mathbf{v}_{-j}), F(x_j | \mathbf{v}_{-j}))}{\partial F(x_j | \mathbf{v}_{-j})}, \quad (14)$$

在得到 $F(x_i | \mathbf{v}_{-j})$ 和 $F(x_j | \mathbf{v}_{-j})$ 的样本数据后, 拟合出 $C_{x_i, \mathbf{v}_{-j}}$, 再由式 (14) 即可得到下一步拟合所需的样本数据, 重复这样的步骤直到 PCC 中的所有 copula 密度都得到拟合, 就得到了 $X = (X_1, \dots, X_n)$ 的联合概率密度。

为了后文叙述方便, 这里定义 $h(x, \mathbf{v}) = \frac{\partial C_{x, \mathbf{v}}(x, \mathbf{v})}{\partial \mathbf{v}}$, 则 $F(x | \mathbf{v}) = h(F(x), F(\mathbf{v}))$ 。由于

PCC 的多样性, 如何构造和选取 PCC, 是应用这一方法的关键。藤结构的提出完美解决了 PCC 的构造问题, 为这一方法奠定了理论基础。

2.2 藤结构

Bedford 和 Cooke^[31] 提出了用图来表示 PCC, 即将每一个 PCC 分解用藤结构来刻画, 藤结构有如下定义:

定义 2.1 记 T_1, \dots, T_{n-1} 为一列树, 其中树

T_i 的点集和边集分别为 N_i 和 E_i , 且满足

1) $N_1 = \{1, \dots, n\}$.

2) 对于 $i = 2, \dots, n - 1, N_i = E_{i-1}$.

则称 $V = (T_1, \dots, T_{n-1})$ 为一个 n 元藤结构。

若 V 还满足:

3) (正则条件) 对所有 $i = 2, \dots, n - 1$, 只要 $a, b \in N_i$ 且 $\{a, b\} \in E_i$, 那么 a, b 作为树 T_{i-1} 中的两条边, 在树 T_{i-1} 中必有一个共同的顶点。

则称 V 为一个 n 元 R 藤, 或称为正则藤。

利用 R 藤与 PCC 之间建立联系, 本质上是将是式 (13) 的分解形式用图来表示, 式中的边缘分布或条件分布与 R 藤中某一棵树的顶点对应, 而连接这两个分布的 copula 函数, 则对应树中连结两个顶点的边。具体地, 记 $E_0 = N_1$, 对 $i = 1, 2, \dots, n - 1$ 和任意 $e_i = \{a, b\} \in E_i, a, b \in E_{i-1}$, 令 $U_{e_i} = \{e \in E_0 \mid \exists e_j \in E_j, j = 1, \dots, i - 1, \text{st. } e \in e_1 \in e_2 \in \dots \in e_i\}$, 即从 e_i 出发, 返回到它所对应的 T_{i-1} 的顶点, 再返回到上一棵树中对应的顶点集合, 依次类推直到回到 T_1 , 最终所得到的 T_1 的子集。再令

$$D_{e_i} = U_a \cap U_b, C_{e_i,a} = U_a \setminus D_{e_i}, C_{e_i,b} = U_b \setminus D_{e_i}.$$

将每条边 $e = \{a, b\}$ 对应于连接了给定 X_{D_e} 时 $X_{C_{e,a}}$ 和 $X_{C_{e,b}}$ 的条件分布的二元 copula 函数, 可以证明, $|C_{e,a}| = |C_{e,b}| = 1$, 且下式给出了联合密度的一个 PCC

$$f(x_1, \dots, x_n) = f_1 \cdots f_n \prod_{i=1}^{n-1} \prod_{e \in E_i} c_{C_{e,a}, C_{e,b} | D_e}(F_{C_{e,a} | D_e}, F_{C_{e,b} | D_e}). \tag{15}$$

事实上, 对任意一种 PCC, 若根据其 $n(n - 1)/2$ 个 copula 密度, 对应构造出其图结构, 则会得到一个 R 藤, 反过来, 对于任意一个 R 藤, 可以由式 (15) 写出这个 R 藤唯一决定的 PCC。图 1 例举了 D 藤和 C 藤结构, 其中“|”的左边为变量集, 右边为条件集, 对每条边连结的顶点集合求交集, 即可得到这条边所对应的条件集 D_e , 再由对称差得出变量集 $C_{e,a} \cup C_{e,b}$ 。D 藤和 C 藤可理解为 R 藤的两个极端情形, 其定义分别为:

- 若 T_1 中顶点的最大度数是 2, 则称其为 D 藤。
- 若对 $i = 1, \dots, n - 1$, 每个 T_i 都有一个度数为 $n - i$ 的顶点, 则称其为 C 藤。

理论上可以对所有可能的 PCC 做拟合, 再比较它们的优劣, 但随着维数的增高, 这样的做法是

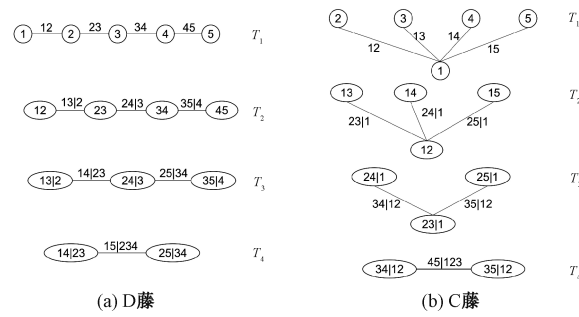


图 1 D 藤和 C 藤

Fig. 1 D-vine and C-vine

不切实际的。利用 Kendall's τ 秩相关系数及分布之间的条件独立, 可以对 R 藤的选取提供依据。若模型中的变量存在条件独立性, 则它们的条件分布对应的 copula 函数为乘积 copula: $C(u, v) = uv$, 其对应的 copula 密度为 $c(u, v) = 1$, 这样就使得模型得到了简化。通常认为, 有着较强相依性的变量之间的拟合应更加追求准确性, 本文选取的相依性度量为 Kendall's τ 秩相关系数, 基于下述理由认为, 其绝对值大的变量组合更应优先被准确地拟合:

- R 藤中第 1 棵树的拟合效果, 往往对整个模型影响较大。
 - 很多不同类型的 copula 函数都可以较好地拟合存在独立性的数据, 并且独立性相对应的参数取值相同或相近。
 - 在实际应用中, 往往可以认为数据的相依关系是由部分变量产生的, 而非全部变量, 若在 T_1 中选择相依性强的变量组合进行拟合, 则后续的转变变量之间的相依性在相比较之下更弱甚至可以认为是独立的。
- 所以, 本文结合最大生成树算法来构造 PCC。对树 N_1 生成的完全图, 定义其每条边的边值为 Kendall's τ 秩相关系数的绝对值, 由如下最大生成树算法生成 T_1 ;
- 从 N_1 中任选一点, 并从连结它的所有边中选择边值最大的作为第 1 条边。
 - 从连结了已被选取的边中任一顶点的所有边中选择边值最大的一条边, 添加到已选取的边中。
 - 重复上一步骤直到所有顶点都被已选取的边连结到为止, 由此得到的树即为最大生成树 T_1 。

2.3 藤 copula 方法与 SMAA 的结合

利用 PCC 方法需要获取模型中多元随机变量的一组独立同分布样本, 在其边缘分布已知甚

至未知的情形下,利用样本和适当的二元 copula 构造出这个多元随机变量的联合分布。但实际情况中,我们获取的数据可能是时间序列,不能作为独立同分布的样本。这样的情况,可以先采用 AMRA 和 GARCH 模型^[32]对其边缘分布进行拟合,去除自相关和异方差性,得到残差样本序列再由其经验分布转化为均匀分布的值,即可作为独立样本处理。

对于本文所提出的 SMAA 与藤 copula 相结合的方法,其具体应用步骤可归纳如下:

- 1) 获取随机变量 $(\xi_{ij})_{m \times n}$ 的一组独立同分布的样本,必要时可采用 AMRA 和 GARCH 模型将其转化为标准均匀分布。为方便起见, $(\xi_{ij})_{m \times n}$ 可重新标号并记为 (X_1, \dots, X_N) 。
- 2) 对所有变量组合计算 Kendall's τ 秩相关系数。
- 3) 根据 2.2 节中的最大生成树算法,得到最大生成树。
- 4) 结合多种方法,选择不同类型 copula 函数拟合最大生成树中每条边对应的变量组合,并用几种检验方法比较其优劣,最终确定 copula 函数类型及其参数。在拟合之前,可以先对数据做独立性检验,若认为独立性成立,则采用独立 copula 函数简化模型。
- 5) 由上一步得出的 copula 函数及其参数计算得到转换后的样本。
- 6) 根据转换后的样本,对所有可能的样本组合计算其经验分布的 Kendall's τ 秩相关系数。这里,可能的样本组合是指必须满足 R 藤的正则条件。
- 7) 重复步骤 3) 到步骤 5),直到得到完整的 mn 元 R 藤,以及所有需要拟合的 copula 函数及参数。
- 8) 因为 SMAA 所计算的积分是对上一步中得到的非常复杂的分布所做的积分,因而需要采用蒙特卡罗模拟的方法来计算积分的近似值。依据拟合出的 $(\xi_{ij})_{m \times n}$ 的联合分布,模拟出大样本,根据模拟样本来计算 SMAA 问题中的可接受度,中心权重向量以及置信度等指标。
- 9) 对上一步的计算结果进行分析,比较不同选择的优劣。

注意到,最大生成树算法是对连通图进行的,因为在构造 T_1 时,是从完全图开始的,算法条件当然满足。而在后续构造中,不难验证,定义 2.1

的正则条件确保了步骤 5) 中所有可能的样本组合构成一个连通图,即对每个已知的 T_{i-1}, T_i 的所有可能的备选边构成了集合 $E'_i = \{ \{a, b\} \in E_{i-1}^2 \mid a \cap b \mid = 1 \}$, 且 E'_i 一定是连通图,进而算法条件依然成立。为了计算 E'_i 中每条边对应的 Kendall's τ 秩相关系数,需要用到转换后的样本,其计算方法与 2.1 节中所介绍的情形是相同的。具体地,设 $e = \{a, b\} \in E'_i$, 则 e 对应于给定 X_{D_e} 条件下 $X_{C_{e,a}}$ 和 $X_{C_{e,b}}$ 的条件分布的二元 copula 函数,即下一步拟合所需的转换变量为

$$F_{C_{e,a} \mid D_e}(X_{C_{e,a}} \mid \mathbf{v}_{D_e}), F_{C_{e,b} \mid D_e}(X_{C_{e,b}} \mid \mathbf{v}_{D_e}).$$

其中 \mathbf{v}_{D_e} 表示指标集 D_e 所对应的随机变量集合。注意到定义 2.1 的正则条件保证了 a, b 有一个 T_{i-1} 中的公共元素,不妨记为 $a = \{a_1, a_2\}, b = \{a_2, b_2\}$, 则结合 U_a 的定义,不难验证 $C_{e,a} = C_{a,a_1}, D_e = D_a \cup C_{a,a_2}$, 进而由式(14)可计算转换变量如下

$$F_{C_{e,a} \mid D_e}(X_{C_{e,a}} \mid \mathbf{v}_{D_e}) = \frac{\partial C_{C_a \mid \mathbf{v}_{D_a}}(F_{C_{a,a_1} \mid D_a}(x_{C_{a,a_1}} \mid \mathbf{v}_{D_a}), F_{C_{a,a_2} \mid D_a}(x_{C_{a,a_2}} \mid \mathbf{v}_{D_a}))}{\partial F_{C_{a,a_2} \mid D_a}(x_{C_{a,a_2}} \mid \mathbf{v}_{D_a})} = h(F_{C_{a,a_1} \mid D_a}(x_{C_{a,a_1}} \mid \mathbf{v}_{D_a}), F_{C_{a,a_2} \mid D_a}(x_{C_{a,a_2}} \mid \mathbf{v}_{D_a})).$$

式中 $C_a = C_{a,a_1} \cup C_{a,a_2}$, 且转换变量 $F_{C_{a,a_1} \mid D_a}(x_{C_{a,a_1}} \mid \mathbf{v}_{D_a}), F_{C_{a,a_2} \mid D_a}(x_{C_{a,a_2}} \mid \mathbf{v}_{D_a})$ 已经在上一步的递推过程中得到。将 $X_{C_{a,a_1}}, X_{C_{a,a_2}}$ 的样本值一一代入即得到了转换样本。

尽管上述步骤看似复杂,但计算机程序可以很快捷地得到藤 copula 拟合的分布,而接下来的 SMAA-2 分析与以往方法的区别则主要是蒙特卡罗模拟的不同。Lahdelma 等^[23]建议使用高斯分布来拟合问题中不确定的标准值,而高斯分布本身具有形式简单、计算容易的特点。在进行 SMAA 分析时所计算积分,一般不需要采用数值方法或是蒙特卡罗模拟的方法,而对于本文建议的藤 copula 拟合出的分布而言,要进行后续的 SMAA-2 分析,则必须要用到蒙特卡罗模拟进行近似。虽然复杂性略有提高,但通过计算机程序可以大大简化工作量。总体上,略微提高的复杂度换来的是更加准确的拟合结果,进而使 SMAA 分析得到的决策建议更加可靠,这是本文方法的创新以及优势所在。

3 模拟分析

3.1 模拟算例分析

本文假定的研究案例有 A、B、C 3 个选择方案,而每种选择的总效用都由 2 个判别标准来量化,且这些标准的值是随机的,这些随机标准的值一共构成了 6 个决策变量,也组成了一个 6 维的随机向量。假定这些决策变量都服从标准均匀分布,之所以可以这样假定,是因为随机变量的分布函数自身是服从均匀分布的,因此对于实际数据,都可以进行一步转化,得到其分布函数的函数值,再进行数据分析。而对于效用函数,为了简化问题,本文假定每个决策变量的效用函数都为标准均匀分布的分布函数,这样便省去了对每个决策变量效用的计算。当然,在实际应用中,会根据每个决策变量的实际意义来分别定义其效用函数,并对所有决策变量都计算出效用函数值再进行下一步分析。对于模拟案例中的数据,本文用一个给定的 R 藤结构表示它们的联合分布。这个 R 藤的结构矩阵为

$$\begin{bmatrix} 6 & 0 & 0 & 0 & 0 & 0 \\ 2 & 5 & 0 & 0 & 0 & 0 \\ 3 & 2 & 2 & 0 & 0 & 0 \\ 1 & 3 & 3 & 3 & 0 & 0 \\ 4 & 1 & 4 & 4 & 4 & 0 \\ 5 & 4 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

结构矩阵与藤结构是多对一的关系^[33],它将藤结构转化为矩阵来表示,进而可方便地应用于计算机程序中,实现复杂的计算。结构矩阵为下三角矩阵,且右边的任意列中的元素都包含于它左边的列中,每个藤结构可以用多种等价的矩阵形式来表示,而每个结构矩阵反过来可得到唯一与之对应的藤结构。以本例的矩阵第 1 列为例,对角线上的变量与最后一行的变量连结了一条边,这些边构成了藤结构中的第一棵树。然后对角线上的变量与倒数第 2 行的变量在以最后一行的变量为条件下的条件变量连结了一条边,而这些边则构成了藤结构中的第 2 棵树,依次类推,这个矩阵对应的 5 棵树如图 2 所示。

按照同样的方式,可以用矩阵来记录藤结构每条边所对应的 copula 函数的函数族以及函数参数,它的 copula 函数族矩阵如下

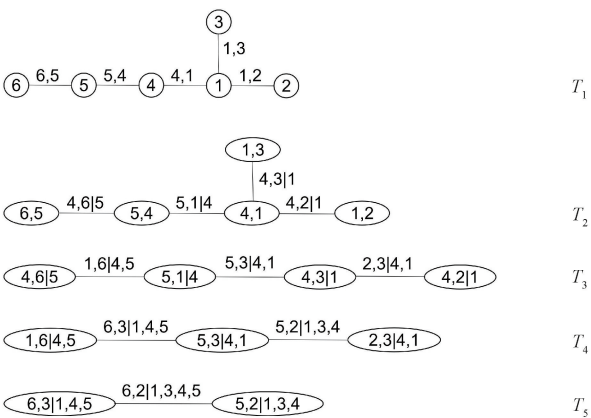


图 2 假定的 R 藤结构
Fig. 2 The assumed R-vine

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 & 0 \\ 3 & 4 & 4 & 4 & 0 & 0 \\ 3 & 4 & 1 & 1 & 3 & 0 \end{bmatrix}.$$

其中

- 1=Gaussian copula.
- 2=Student *t* copula.
- 3=Clayton copula.
- 4=Gumbel copula.

因为本例中都是单参数 copula,因此只需要 1 个矩阵来记录其各个参数,如下述矩阵所示

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0.2 & 0 & 0 & 0 & 0 \\ 1.4 & 0.9 & 1.1 & 0 & 0 & 0 \\ 1.6 & 1.5 & 1.6 & 1.9 & 0 & 0 \\ 3.1 & 3.9 & 0.9 & 0.5 & 4.8 & 0 \end{bmatrix}.$$

通过这 3 个矩阵的信息,即给定了决策变量的联合分布。由这个联合分布,本文模拟出 10 000 个决策样本。对这些样本进行直接的 SMAA 分析。分析得到的结果如表 1 所示。

由表 1 可知,C 选择的最优可接受度最大,为

表 1 决策样本的秩可接受度			
Table 1 Rank acceptability of sample data			
选择	秩为 1	秩为 2	秩为 3
A	0.303 8	0.355 7	0.340 5
B	0.289 2	0.438 7	0.272 1
C	0.407 0	0.205 6	0.387 4

0.407; A 选择的最优可接受度次之,为 0.303 8; B 选择最低,为 0.289 2。可以得到结论: C 为最优选择, A 选择次之, B 选择最差。如果将秩为 2 的可接受度也考虑进来的话,不难发现 B 选择的秩可接受度最大。并且前两个秩的可接受度之和也是最大,说明 B 选择不太可能成为最差的选择,在实际应用中可以把这种情况视为一种折中策略。

用这些样本,以及第 2 节介绍的方法,可以拟合出由[3 个选择×2 个准则]组成的 6 维随机变量的联合分布,具体的计算步骤可以由 R 软件实现,计算得出的联合分布依然由多个矩阵来表示,包括藤结构矩阵, copula 函数族矩阵和对应的 1 个或 2 个 copula 参数矩阵,其中 R 藤结构矩阵为

$$\begin{bmatrix} 6 & 0 & 0 & 0 & 0 & 0 \\ 4 & 3 & 0 & 0 & 0 & 0 \\ 5 & 4 & 1 & 0 & 0 & 0 \\ 3 & 5 & 4 & 2 & 0 & 0 \\ 2 & 1 & 5 & 4 & 4 & 0 \\ 1 & 2 & 2 & 5 & 5 & 5 \end{bmatrix}.$$

$$\begin{bmatrix} 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 \\ 0.374\ 20 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 \\ 0.233\ 49 & 0.357\ 03 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 \\ 0.173\ 14 & 0.370\ 41 & 0.221\ 99 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 \\ 0.372\ 42 & -0.490\ 58 & 0.282\ 41 & 0.314\ 27 & 0.000\ 00 & 0.000\ 00 \\ 3.770\ 64 & 0.722\ 74 & 0.901\ 72 & 0.900\ 65 & 3.962\ 06 & 0.000\ 00 \end{bmatrix}.$$

由于拟合过程中考虑了 t 分布等双参数型的 copula 函数,因此还需要第 2 参数矩阵来记录其

$$\begin{bmatrix} 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 \\ 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 \\ 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 \\ 0.000\ 00 & 18.743\ 01 & 4.624\ 36 & 0.000\ 00 & 0.000\ 00 & 0.000\ 00 \\ 21.754\ 09 & 8.835\ 24 & 5.855\ 37 & 4.492\ 27 & 0.000\ 00 & 0.000\ 00 \\ 0.000\ 00 & 0.000\ 00 & 0.000\ 00 & 30.000\ 00 & 0.000\ 00 & 0.000\ 00 \end{bmatrix}$$

可以看到,拟合出的 R 藤与原假设的藤结构并不相同,但是,通过拟合出的联合分布,再次进行 SMAA 分析,分析得到结果如表 2 所示。

表 2 拟合模型的秩可接受度			
Table 2 Rank acceptability of the fitted model			
选择	秩为 1	秩为 2	秩为 3
A	0.340 1	0.328 2	0.331 7
B	0.254 0	0.480 6	0.265 4
C	0.405 9	0.191 2	0.402 9

结构矩阵对应的 5 棵树如图 3 所示。

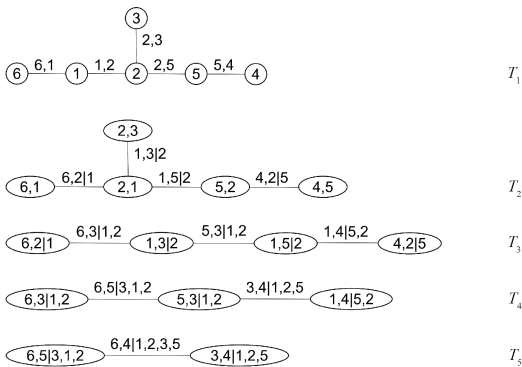


图 3 拟合的 R 藤结构
Fig. 3 The fitted R-vine

它的 copula 函数族矩阵如下

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 \\ 3 & 3 & 0 & 0 & 0 & 0 \\ 3 & 2 & 2 & 0 & 0 & 0 \\ 2 & 2 & 2 & 2 & 0 & 0 \\ 3 & 1 & 1 & 2 & 4 & 0 \end{bmatrix}.$$

它们对应的参数矩阵如下

对应的自由度。它们的自由度如下所示

由表 2 可知, C 选择的最优可接受度最大,为 0.405 9; A 选择的最优可接受度次之,为 0.340 1; B 选择最低,为 0.254 0。可以得到结论, C 为最优选择, A 选择次之, B 选择最差。这与原假设的模型分析结果完全相符。

结论: 如果不考虑 6 个变量相依结构, 即它们相互独立, 那么对于本例而言, 各个决策变量的效用彼此没有不同, 用 SMAA 分析会得到 3 个选择的秩可接受度近乎相等, 这显然与假设案例的结

果是不相符的。而考虑相依结构得到的结果与假设案例的实际情况是一致的。这说明,考虑变量间的相依关系,用藤 copula 拟合出联合分布,再进行 SMAA 分析的方法是准确的、贴合实际的,能够给决策者提供科学准确的参考依据。

3.2 对比分析

在 SMAA 方法的研究领域,已经有文献考虑了决策变量之间的相依关系,并采用常用的多元分布来取代独立分布。Lahdelma 等^[23]提出用独立分布来处理高维数据是不准确不合理的,并将高斯分布(多元正态模型)运用到 SMAA 方法中,假定随机的决策变量服从多元正态分布,再由秩可接受度来评估各个选择的优劣,其分析结果与独立假设下计算的结果有明显的区别。在金融研究领域,copula 方法被广泛应用,但应用较多的往往是形式上较为简单的函数族,例如椭圆型 copula 函数族^[34]。在椭圆型 copula 函数族中,高斯 copula 是最简单实用的一种,在实证应用研究中经常发挥重要作用。因此,本文将多元正态模型和高斯 copula 这两种函数模型加入到模拟实验中,与 R 藤模型进行对比,分析出不同模型的优劣。

这里分别用多元正态模型以及高斯 copula 模型对 3.1 节中的 10 000 个原始样本进行拟合,通过拟合出结果分别进行 SMAA 分析,然后与用样本直接计算 SMAA 以及用藤 copula 拟合计算 SMAA 的结果进行对比。结果如表 3 所示。

表 3 不同拟合模型的对比分析

Table 3 Comparison between those fitted models

选择	不同模型下的秩可接受度		
	秩为 1	秩为 2	秩为 3
原始样本	A 0.303 8	0.355 7	0.340 5
	B 0.289 2	0.438 7	0.272 1
	C 0.407 0	0.205 6	0.387 4
藤 copula	A 0.340 1	0.328 2	0.331 7
	B 0.254 0	0.480 6	0.265 4
	C 0.405 9	0.191 2	0.402 9
多元正态	A 0.341 9	0.348 2	0.309 9
	B 0.293 6	0.438 2	0.268 2
	C 0.403 2	0.200 3	0.396 5
高斯 copula	A 0.325 0	0.343 6	0.331 4
	B 0.265 9	0.466 3	0.267 8
	C 0.409 1	0.190 1	0.400 8

从表 3 中可以看出,3 种拟合方法得到的结果都是 C 选择最优,这与直接通过样本计算得到的结果是一致的。在准确性方面,观察一下 C 选择的可接受度的数值,很显然,这 3 种方法的值几

乎一致。但相对于参考样本的直接计算值,藤 copula 的误差为 0.27%;多元正态的误差为 0.93%;高斯 copula 的误差为 0.52%。因此得到以下结论:在样本数据比较常规(生成样本时选用的 copula 函数族以及参数都比较常规)时,3 种方法的计算结果都比较准确,藤 copula 方法相对误差最小,即本文提出的方法有一定的优势。

3.3 对尾部相依结构样本进行对比分析

在金融和管理领域,一些极端的数据通常容易引起广泛关注,就像数据之间的非线性关系一样,只看线性关系是没有意义的。同样地,秩相关也不能完美地描述所有的相依关系。实证研究中,需要使用尾部相依关系(二元变量中数据取值特别大或特别小时的相依关系)。为了检验本文提出的方法对实际问题的实用性,本文建立一种具有尾部相依结构的分布,并模拟生成出一组具有尾部相依性的数据,然后用藤 copula、多元正态、高斯 copula 模型分别拟合后的结果与 SMAA 方法直接分析的结果进行对于分析,与进行 SMAA 分析的结果进行对比,得出结论和建议。

由图 4 可知,Clayton copula 和 Gumbel copula (在 copula 函数族矩阵中用 3 和 4 表示)具有尾部相依性,随着 τ 系数的增大而愈加明显。

从该性质出发,同样假定的研究案例是 3 个选择、2 个标准的 6 维随机变量,使用图 2 的 R 藤结构表示它们的联合分布。对于它的 copula 函数族矩阵,全部选用 Clayton copula 和 Gumbel copula,矩阵如下

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 & 0 \\ 3 & 4 & 4 & 4 & 0 & 0 \\ 3 & 4 & 4 & 4 & 3 & 0 \end{bmatrix}.$$

其中 3 代表 Clayton copula,4 代表 Gumbel copula。它们是单参数 copula 函数,为了得到更明显的尾部相依性,利用大于 5 的参数使得 τ 系数大于 0.7。得到对应的参数矩阵

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 & 0 \\ 7 & 6 & 0 & 0 & 0 & 0 \\ 19 & 17.9 & 6.1 & 0 & 0 & 0 \\ 26 & 6.5 & 5.6 & 7 & 0 & 0 \\ 8.1 & 19 & 10.9 & 26 & 8 & 0 \end{bmatrix}.$$

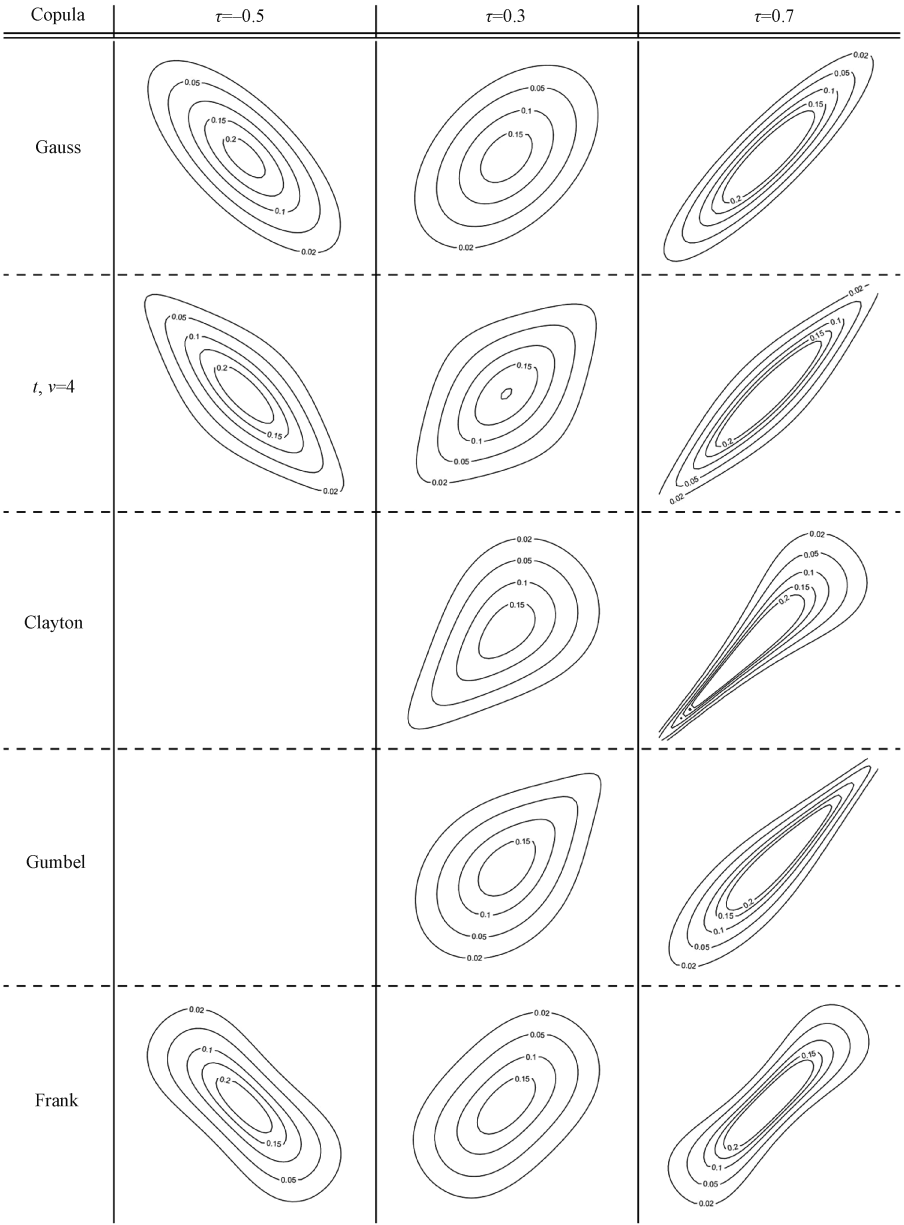


图 4 常用 copula 函数的等高线图

Fig. 4 Contour map of common copulas

通过这些参数,模拟出 10 000 个新样本。然后分别用藤 copula 模型、多元正态模型以及高斯 copula 模型对于模拟得到的 10 000 个新样本进行拟合,通过拟合出的联合分布,对于各个随机模型进行 SMAA 分析。最后对比计算结果与用新样本直接进行 SMAA 分析的结果(对比结果见表 4)。

从分析结果中可以看出,用藤 copula 拟合得到的结果是 B 选择最优,多元正态和高斯 copula 拟合得到的结果是 C 选择最优。对于新样本直接计算的实际情况而言,B 为最佳选择,这与本文提出的方法(藤 copula 拟合)结果一致。结果表

表 4 尾部相依情形下的对比分析

选择		不同模型下的秩可接受度		
		秩为 1	秩为 2	秩为 3
新样本	A	0.253 2	0.478 9	0.267 9
	B	0.406 6	0.212 7	0.380 7
	C	0.340 2	0.308 4	0.351 4
藤 copula	A	0.270 3	0.481 2	0.248 5
	B	0.389 0	0.225 3	0.385 7
	C	0.340 7	0.293 5	0.265 8
多元正态	A	0.305 7	0.445 7	0.248 6
	B	0.385 7	0.274 5	0.338 9
	C	0.413 4	0.230 6	0.356 0
高斯 copula	A	0.328 9	0.351 0	0.320 1
	B	0.272 3	0.457 2	0.270 5
	C	0.398 8	0.191 8	0.409 4

明在数据出现尾部相依性时,藤 copula 方法得到的结果与实际情况一致,而多元正态和高斯 copula 方法差距较大,无法提供精准的参考信息。因而选择方法时,建议使用藤 copula 方法拟合出联合分布,再进行 SMAA 分析,该方法可以适用于不同分布类型的数据,为决策者提供更准确和实用的参考。

4 结语

本文对于藤 copula 的原理及其相关性质进行详细介绍,该函数的适用范围较广,可应用于各种数据,具有一定的优势。通过实例研究,结合藤 copula 拟合样本数据得到样本联合分布的方法,以及应用 SMAA 方法计算各个备选方案的优劣的具体步骤。此外,与多元正态模型和高斯 copula 模型等传统方法的对比分析进一步证明了藤 copula 模型与 SMAA 方法结合的优势:准确度高、实用性强。

在实际案例中处理决策问题时,决策变量随机不确定性会导致观察值的减少。如果这些样本直接进行 SMAA 分析,由此产生的决策可能只适用于当前的具体情况,无法将该决策推广到更加全面的范围。因此,引入联合分布函数和藤系 copula 模型来拟合这些决策变量的联合分布函数,并采用 SMAA 方法进行决策分析。最后,在联合分布函数计算的基础上,评估不同备选方案的优缺点。期望值将更科学、更严谨、更能代表整体。

然而,不同的拟合方法将得到不同的联合分布,这会对 SMAA 分析的结果有所影响。Lahdelma 等^[23]将多元正态模型引入到 SMAA 方法中。而多元正态模型只能拟合考虑变量之间的线性相关条件下边际分布为正态分布的决策变量。模拟实验的研究结果表明,由于数据存在尾部相依性,对于某些模型(比如多元正态模型,高斯 copula 模型)并不适用,得到的结论甚至会有错误。但藤 copula 模型的分析计算结果并不会出现这类问题。

由于数据来源有限,尚未获得可用于数据分析的大样本实际案例数据,因此未对实际问题并没有进行数据与决策的分析,本文主要是在理论上提出了这种创新方法,通过实例模拟来验证该方法的主要步骤和优点。在模拟过程中,所有考虑的决策变量均为标准化均匀分布,效用函数定

义为变量的分布函数。对于实际数据,必须充分考虑到符合实际情况的效用函数以及每个决策变量各自的边际分布。本文提出的方法进一步提高了 SMAA 方法的应用价值。在多标准决策领域,还有其他流行的研究方法,如 DEMATEL 模型和网络分析。DEMATEL 更注重各种决策因素之间的逻辑关系和相互影响。并充分利用该领域专家的相关经验和知识建立决策模型,同时又具有这种网络分析方法的特点。

本文研究发现,与其他借助于逻辑和经验的方法相比,SMAA 方法对于权重信息缺失,或者说缺少足够的专家经验时的应用适应性更强,这种方法对数据更加重视,从大数据中挖掘决策意见,而 copula 方法加强了对于数据的重视程度,充分发挥数据利用的优势,但对实际问题的样本数据量也有一定的要求。因此,决策者可能会根据实际情况采取不同的方法。

对于未来的研究方向,除解决现实生活中的一些决策问题的实际应用研究外,本文的方法还可以扩展到研究如何将 copula 方法与精细化的随机多准则接受度分析方法相结合,例如,可以和 SMAA-O,SMAA-TRI,SMAA-D 等方法结合使用。对于一些比较复杂的实际问题,本文方法可以与 DEMATEL 模型和网络分析法等方法结合使用,或者进行对比分析,考查不同方法的优缺点,研究出更多有意义的成果。

参考文献

[1] Keeney R L, Raiffa H, Meyer R F. Decisions with multiple objectives: preferences and value trade-offs[M]. Cambridge, UK: Cambridge university press, 1993.

[2] Roy B. Decision aiding: major actors and the role of models [M]//Multicriteria methodoly for decision making. Boston, USA: Springer, 1996: 237-267.

[3] Brans J P, Mareschal B. Promethee methods[M]//Multiple criteria decision analysis. New York, USA: Springer, 2005: 163-186.

[4] Xu X Z. The SIR methods: a superiority and inferiority ranking method for multiple criteria decision making [J]. European Journal of Operational Research, 2001, 131 (3): 587-602. DOI:10. 1016/S0377-2217(00)00101-6.

[5] Lahdelma R, Hokkanen J, Salminen P. SMAA-stochastic multiobjective acceptability analysis[J]. European Journal of Operational Research, 1998, 106 (1): 137-143. DOI: 10. 1016/S0377-2217(97)00163-X.

[6] Lahdelma R, Salminen P. SMAA-2: stochastic multicriteria acceptability analysis for group decision making [J]. Operations Research, 2001, 49 (3): 444-454. DOI: 10.

- 1287/opre. 49. 3. 444. 11220.
- [7] Lahdelma R, Miettinen K, Salminen P. Ordinal criteria in stochastic multicriteria acceptability analysis (SMAA) [J]. *European Journal of Operational Research*, 2003, 147(1): 117-127. DOI:10.1016/S0377-2217(02)00267-9.
 - [8] Yu W. ELECTRE TRI: aspects méthodologiques et manuel d'utilisation [D]. Paris, France: Université de Paris-Dauphine, 1992.
 - [9] Tervonen T, Lahdelma R, Dias J A, et al. SMAA-TRI: a parameter stability analysis method for ELECTRE-TRI [C]// Kiker G A, Linkov I. *Environmental Security in Harbors and Coastal Areas*. Berlin, Germany: Springer, 2007: 217-231.
 - [10] Tervonen T, Figueira J R, Lahdelma R, et al. Towards robust ELECTRE III with simulation: theory and software of SMAA-III [J]. *European Journal of Operational Research*, under review, 2007.
 - [11] Lahdelma R, Salminen P. Stochastic multicriteria acceptability analysis using the data envelopment model [J]. *European Journal of Operational Research*, 2006, 170(1): 241-252. DOI:10.1016/j.ejor.2004.07.040.
 - [12] Lahdelma R, Salminen P. Prospect theory and stochastic multicriteria acceptability analysis (SMAA) [J]. *Omega*, 2009, 37(5): 961-971. DOI:10.1016/j.omega.2008.09.001.
 - [13] Corrente S, Figueira J R, Greco S. The SMAA-PROMETHEE method [J]. *European Journal of Operational Research*, 2014, 239(2): 514-522. DOI:10.1016/j.ejor.2014.05.026.
 - [14] Tervonen T, Lahdelma R. Implementing stochastic multicriteria acceptability analysis [J]. *European Journal of Operational Research*, 2007, 178(2): 500-513. DOI: 10.1016/j.ejor.2005.12.037.
 - [15] Hokkanen J, Lahdelma R, Miettinen K, et al. Determining the implementation order of a general plan by using a multicriteria method [J]. *Journal of Multi-Criteria Decision Analysis*, 1998, 7(5): 273-284. DOI: 10.1002/(SICI)1099-1360(199809)7: 5<273; AID-MCDA198>3.0.CO;2-1.
 - [16] Hokkanen J, Lahdelma R, Salminen P. A multiple criteria decision model for analyzing and choosing among different development patterns for the Helsinki cargo harbor [J]. *Socio Economic Planning Sciences*, 1999, 33(1): 1-23. DOI: 10.1016/S0038-0121(98)00007-X.
 - [17] Hokkanen J, Lahdelma R, Salminen P. Multicriteria decision support in a technology competition for cleaning polluted soil in Helsinki [J]. *Journal of Environmental Management*, 2000, 60(4): 339-348. DOI:10.1006/jema.2000.0389.
 - [18] Lahdelma R, Salminen P, Hokkanen J. Locating a waste treatment facility by using stochastic multicriteria acceptability analysis with ordinal criteria [J]. *European Journal of Operational Research*, 2002, 142(2): 345-356. DOI: 10.1016/S0377-2217(01)00303-4.
 - [19] Tervonen T, Hakonen H, Lahdelma R. Elevator planning with stochastic multicriteria acceptability analysis [J]. *Omega*, 2008, 36(3): 352-362. DOI:10.1016/j.omega.2006.04.017.
 - [20] Kangas J, Hokkanen J, Kangas A S, et al. Applying stochastic multicriteria acceptability analysis to forest ecosystem management with both cardinal and ordinal criteria [J]. *Forest Science*, 2003, 49(6): 928-937.
 - [21] Xia Q, Hua Z S, Song S L. Location selection of distribution centers based on SMAA [J]. *Lecture Notes in Electrical Engineering*, 2014, 273(1): 213-220. DOI:10.1007/978-3-642-40640-9-28.
 - [22] Zhu F L, Zhong P G, Wu Y N, et al. SMAA-based stochastic multi-criteria decision making for reservoir flood control operation [J]. *Stochastic Environmental Research and Risk Assessment*, 2017, 31(6): 1485-1497. DOI:10.1007/s00477-016-1253-3.
 - [23] Lahdelma R, Makkonen S, Salminen P. Multivariate gaussian criteria in SMAA [J]. *European Journal of Operational Research*, 2006, 170(3): 957-970. DOI:10.1016/j.ejor.2004.08.022.
 - [24] Genest C, Favre A C. Everything you always wanted to know about copula modeling but were afraid to ask [J]. *Journal of Hydrologic Engineering*, 2007, 12(4): 347-368. DOI: 10.1061/(asce)1084-0699(2007)12: 4(347).
 - [25] Genest C, Rémillard B, Beaudoin D. Goodness-of-fit tests for copulas: a review and a power study [J]. *Insurance Mathematics and Economics*, 2009, 44(2): 199-213. DOI: 10.1016/j.insmathco.2007.10.005.
 - [26] Joe H. Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters [J]. *Lecture Notes-Monograph Series*, 1996, 28:120-141.
 - [27] Bedford T, Cooke R M. Probability density decomposition for conditionally dependent random variables modeled by vines [J]. *Annals of Mathematics and Artificial Intelligence*, 2001, 32(1):245-268. DOI:10.1023/A: 1016725902970.
 - [28] Aas K, Czado C, Frigessi A, et al. Pair-copula constructions of multiple dependence [J]. *Insurance Mathematics and Economics*, 2009, 44(2): 182-198. DOI: 10.1016/j.insmathco.2007.02.001.
 - [29] Sklar A. Fonctions de répartition à n dimensions et leurs marges [J]. *Publications de l'Institut de Statistique de l'Université de Paris*, 1959(8): 229-231.
 - [30] Genest C, Rivest L P. Statistical inference procedures for bivariate archimedean copulas [J]. *Journal of the American Statistical Association*, 1993, 88(423): 1034-1043. DOI: 10.1080/01621459.1993.10476372.
 - [31] Bedford T, Cooke R M. Vines-a new graphical model for dependent random variables [J]. *The Annals of Statistics*, 2002, 30(4): 1031-1068. DOI: 10.1214/AOS%2F1031689016.
 - [32] 叶五一, 郭人榛, 缪柏其. 基于 R 藤 copula 变点模型的金砖四国金融传染性与稳定性检验 [J]. *中国科学技术大学学报*, 2018, 48(8): 655-666. DOI:10.3969/j.issn.0253-2778.2018.08.009.
 - [33] Dißmann J F. *Statistical inference for regular vines and application* [D]. Munich, Germany: Technische Universität München, 2010.
 - [34] Joe H, Kurowicka D. *Dependence modeling: vine copula handbook* [M]. Singapore: World Scientific, 2010. DOI: 10.1142/7699.