

基于 Copula 函数的联合概率法在海洋工程中的应用

秦振江, 孙广华, 闫同新, 宋 华, 张 婧

(中南大学数学科学与计算技术学院 湖南 长沙 410075)

摘 要: 本文介绍了 Copula 函数的定义、属性和几种特殊的 Copula 函数. 将 Copula 函数应用到近海工程中, 用其建立了最大有效波高和最大风速的联合分布函数并检验, 结果表明: Copula 函数能够比较好的模拟实际的联合概率分布. 通过 Copula 方法, 可以由边缘分布和一个连接它们的 Copula 函数来得出联合分布函数, 发挥其描述相关性尤其是尾部相关性的优势. 该方法在近海工程的工程可靠度、设计标准和失效概率的计算中具有广泛的应用前景.

关键词: Copula; 海洋工程; 波高; 风速; 联合分布

中图分类号: P731 **文献标识码:** A **文章编号:** 1003 - 0239 (2007) 2 - 0083 - 08

1 综述

海洋环境设计标准的选定是决定工程结构安全和工程造价的主要因素, 多年来, 在近海及海岸工程中, 人们习惯采用较简单一元极值理论, 仅考虑影响最明显的因素, 根据重现期估计重现水平, 从而确定工程结构需承受的设计荷载^[1].

而事实上, 多数工程结构失效乃至破坏往往不取决于某一项环境荷载超值, 而是同时作用的几种环境整体荷载达到或超过某一临界值, 如海洋石油平台倾覆往往是风、浪、流共同作用的结果, 水库安全则与汇水区域内各点的降雨量及其加权总和有关. 因此, 为了保证工程安全和避免浪费, 则应考虑多个变量及其极值之间的相关关系.

20 世纪 90 年代以后, 针对这一问题, 专家们提出了几种方法^[1], 它们各有优缺点, 其中用联合概率法估算多变量分布是较直观的方法. 但是运用联合概率法的假设是苛刻的, 所采用的边缘分布函数一般是正态分布, 因为只有各变量均服从正态分布或对数正态分布时, 才容易求出其联合概率密度函数的解析表达式. 而对于多维 Weibull 分布一类的非线性、非对称、非正态相关的多维随机变量的联合分布是很难求其解析表达式的, 由于变量之间存在相关关系, 不可能完全独立, 这样联合密度函数也就不可能通过将边缘密度函数简单相乘而得到, 变量间的相依关系分析成了难点. 而用 Copula 函数来构造多变量极值事件的相关结构和联合分布函数便可以很好的解决了这一问题. Copula 方法可以采用不同的边缘分布函数来推求联合分布函数, 因此该方法应用范围广阔. 从 1998 年开始, 出现了较多关于 Copula 函数在金融、保险等领域的应用研究^[2~4]. 近几年国外工程界也已经逐渐重视采用 Copula 函数来建立多变量联合分布函数, 较早的有 Favre,

Anne-Catherine (2004), 把 Copula 用于水文频率分析^[5], Salvatore Grimaldi (2005) 等人对更高维的情况进行了研究^[6], Zhang L (2006) 的研究还与传统的研究方法作了优劣比较^[7]。在区域降雨分布方面 Grimaldi, Salvatore (2005) 等人用 3 维 Copula 函数建立联合分布函数^[8], 而 Shiau J T (2006) 用 Copula 建立了旱灾的预测模型^[9]。在近海工程方面, Wist Hanne T (2004) 用二元正态 Copula 建立了波高和波周期的联合分布^[10], 并成功地应用于日本海域的海洋工程。国内学者有熊立华 (2005) 把 Copula 函数应用在多变量水文频率分析中^[11]。

本文主要是用 Copula 方法解决近海工程中多元概率分析方面的问题。其内容有:

- (1) 介绍 Copula 联结函数的定义、属性和几种特殊的 Copula 函数;
- (2) 采用 Copula 函数来推求最大有效波高和最大风速的联合分布函数, 建立模型并检验;
- (3) 提出了一些可继续研究的问题。

2 Copula 函数的定义与性质

作为相关性分析和多元统计分析的工具, Copula 函数可以看成连接一元边缘分布和其多元联合分布的函数。Copula 方法最早由 Sklar (1959) 提出, 但直到 20 世纪 90 年代末, 这一方法作为相关性分析和多元统计分析的工具, 才开始广泛地应用到金融领域, 2000 年后才引入工程领域。通过 Copula 方法, 可以从边缘分布和一个连接它们的 Copula 函数得出其联合分布, 从而大大的简化了联合分布的计算问题。

Nelsen^[12] 对 Copula 函数的含义和性质做了全面地总结和详细地介绍。这里介绍最基本的 Sklar 定理:

对于一个具有一元边缘分布 F_1, \dots, F_N 的联合分布函数, 一定存在 Copula 函数 C , 满足:

$$F(x_1, \dots, x_n, \dots, x_N) = C(F(x_1), \dots, F(x_n), \dots, F(x_N)) \quad (1)$$

若 F_1, \dots, F_N 连续, 则 C 惟一确定; 且若 F_1, \dots, F_N 为随机变量的边缘分布, 那么由上式定义的函数 F 是 F_1, \dots, F_N 的联合分布函数。

显然, 若令 $u_n = F_n(x_n)$ 为一随机变量, $u_n \in [0, 1]$, $n=1, 2, \dots, N$, 则 $C(u_1, \dots, u_n, \dots, u_N)$ 是一个在 N 维 $[0, 1]$ 空间上具有 $[0, 1]$ 均匀边缘分布的多元分布函数, 亦即 Copula 函数。通过 Copula 函数的密度函数 C 和边缘密度函数, 可以求出 N 元分布函数的密度函数 f 即:

$$f(u_1, \dots, u_n, \dots, u_N) = c(F_1(x_1), \dots, F_n(x_n), \dots, F_N(x_N)) \prod_{n=1}^N f_n(x_n) \quad (2)$$

其中 $c(u_1, \dots, u_n, \dots, u_N) = \frac{\partial c(u_1, \dots, u_n, \dots, u_N)}{\partial u_1, \dots, \partial u_n, \dots, \partial u_N}$, f_n 是边缘分布 F_n 的密度函数。在这里, N 个边缘函数 $F(x_1), \dots, F(x_n), \dots, F(x_N)$ 可属于不同的分布类型, 这是采用 Copula 函

数来构造多变量联合分布函数的一大优点。Copula 联结函数一个重要特性就是它对于随机变量的严格单调增变换是不变的,设随机变量 X 、 Y 的 Copula 函数是 $C(u, v)$, $T(X)$, $T(Y)$ 为严格递增的连续变换,则 (X, Y) 与 $(T(X), T(Y))$ 有相同的 Copula。利用这一性质,可以得出两个非常有意义的推论:一是 Copula 函数反映了随机变量之间的相关结构;这个相关结构与各个随机变量的度量单位无关;二是由 Copula 函数导出的相关性指标,比人们常用的 Pearson 相关系数更加符合人们的要求,这是因为 Pearson 相关系数是线性变换下不变的一种相关性指标,如果涉及到非线性函数的相关性,它就会导出错误的结论。

3 几类 Copula 函数介绍

Copula 函数的构造方法比较多,比较常见的有:椭圆型 (Elliptical Copula) 和阿基米得型 (Archimedean Copula)。椭圆型 Copula 函数则有:

(1) 正态 Copula 函数:

$$C_R^N(u, v) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi(1-R^2)^{1/2}} \exp\left\{-\frac{s^2-2Rst+t^2}{2(1-R^2)}\right\} ds dt \quad (3)$$

其中 R 为相关系数,这个 Copula 函数的两个边缘分布都为正态分布,积分上限 $\phi^{-1}(x)$ 为标准正态分布函数的逆。

(2) Student-t Copula 函数:

$$C_{v,R}^t(u, v) = \int_{-\infty}^{\tau_v^{-1}(u)} \int_{-\infty}^{\tau_v^{-1}(v)} \frac{1}{2\pi(1-R^2)^{1/2}} \left\{1 + \frac{s^2-2Rst+t^2}{v(1-R^2)}\right\}^{-(v+2)/2} ds dt \quad (4)$$

其中 R 为相关系数, $\tau_v^{-1}(x)$ 为自由度为 v 的标准 t 分布函数的逆。

而函数类有一个共同的性质是他们都可以由一个严格单调递增的凸函数 $\phi(u)$ 产生, $\phi(u)$ 满足 $\lim_{\varepsilon \rightarrow 0} \phi(\varepsilon) = \infty$ 及 $\phi(1) = 0$ 。当 $\phi(u)$ 给定时,可以产生的一个 Archimedean Copula 形,如: $C(u, v) = \phi^{-1}(\phi(u) + \phi(v))$ 。Archimedean Copula 可以分为 Gumbel Copula 和 Clayton Copula 等。Gumbel Copula 形式有:

$$C(u, v) = \exp\left(-[(-\ln(u))^\theta + (-\ln(v))^\theta]^{1/\theta}\right) \quad (5)$$

其中, u, v 为 $[0, 1]$ 上的均匀分布变量, θ 为描述两个变量间相依性关系的参数。它由 Gumbel^[13] 提出。Clayton Copula 形式有:

$$C(u, v) = \max([u^{-\theta} + v^{-\theta} - 1]^{1/\theta}, 0) \quad (6)$$

其中, u, v 为 $[0, 1]$ 上的均匀分布变量, θ 为描述两个变量间相依性关系的参数。它由 Clayton 等^[14] 人的研究提出。

4 应用示例

表 1 的数据来源是我国南海某海域 20 次的台风资料^[15]。为了确定海洋石油平台的设计标准,我们想知道这一观测点上百年一遇或千年一遇的波高或风速同时发生的概率有多大,或者在给定其中一个事件发生概率的条件下推求另外一个事件的发生概率。这样就有必要求出联合分布函数。记观测点最大波高的分布函数为 $F(x)$,最大风速的分布函数为 $F(y)$,则观测点的最大波高和风速的联合分布函数用 Copula 联结函数表示为: $F(x, y)=C(F(x), F(y))$ 。

表 1 我国南海某海域 20 次的台风资料

台风代号	最大有效波高 (m)	最大风速 (m/s)	台风代号	最大有效波高 (m)	最大风速 (m/s)
JHN5317. SUS	8.7	39.9	JHN6817. SHI	6.8	23.8
JHN5413. IDA	12.8	40.3	JHN6821. WEN	7.4	23.6
JHN5427. PAM	9.9	41.7	JHN6907. UIO	12.6	38.2
JHN5429. RUB	10.3	28.5	JHN7119. LUC	12.2	45.8
JHN5719. GLO	7.5	26.4	JHN7524. ELS	8.7	34.2
JHN6219. WAN	12.3	47.1	JHN7911. HOP	13.2	31.2
JHN6318. EAY	10.3	40.5	JHN8309. ELL	11.8	35.2
JHN6412. IDA	7.5	22.9	JHN8507. HAL	11.0	41.2
JHN6419. RUB	9.3	40.5	JHN8616. WAY	10.5	41.5
JHN6420. SAL	12.7	42.2	JHN8621. ELL	4.6	22.4

4.1 确定边缘分布模型

一般风、浪、流的分布服从 Gumbel、Weibull、Log-Normal、 Γ 其中的一种类型,在给出历史数据后,分布参数可以用极大似然估计法得出。根据文献[15],设最大波高、最大风速的分布都服从 Weibull 分布,则其密度函数为:

$$y=f(x\ a, b)=abx^{b-1}e^{-ax^b}I_{0,\ \infty}(x) \tag{7}$$

用极大似然估计法得:置信度为95%的最大波高密度函数的参数为 $a=0.00000405306728$, $b=5.19623578005639$,最大风速的密度分布的参数为 $a=0.00000000190386$, $b=5.50184681414617$ 。

4.2 确定 Copula 函数

考虑构造简单和方便使用的原因,本文采用阿基米得型的 Clayton Copula。在该函数中生成的参数可以由下式来估算:

$$\theta=2\tau/(1-\tau),\ \theta\in[-1,\ \infty]\backslash 0 \tag{8}$$

对二元的情况,设 (X_1, Y_1) 和 (X_2, Y_2) 是二元随机变量,一致性相关系数 τ (Kendall 系数)的定义为:

$$\tau=Pr\{(X_1-X_2)(Y_1-Y_2)>0\}-Pr\{(X_1-X_2)(Y_1-Y_2)<0\} \tag{9}$$

容易看出, τ 反应了随机变量 X, Y 变化一致与否的程度, 根据观测资料可以首先求得 $\tau=0.31428$, 从而 $\theta=0.91666$ 。这样, 联合分布函数就完全确定如下:

$$F(x, y)=C(F(x), F(y))=\max([F(x)^{-0.91666}+F(y)^{-0.91666}-1]^{1/0.91666}, 0) \quad (10)$$

为了直观地表示 Copula 描述变量间的相关结构效果, 下面给出 (见图 1 ~ 3) τ 取给定值时的联合密度函数图和密度函数俯视图。

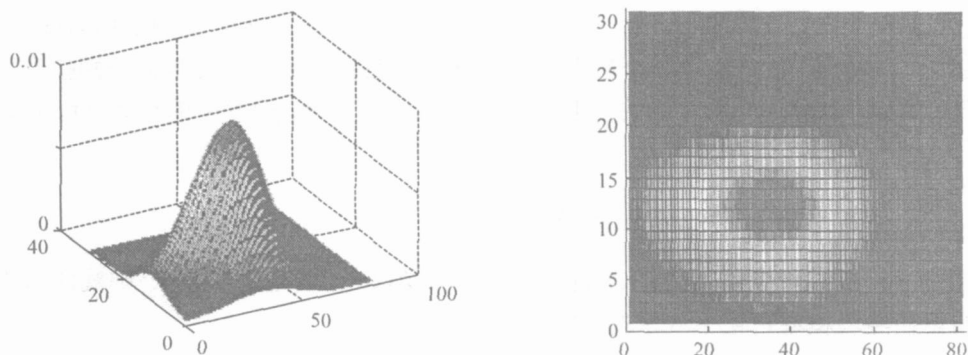


图 1 当两个变量独立即 $\tau=0$ 时的联合密度函数图、密度函数俯视图

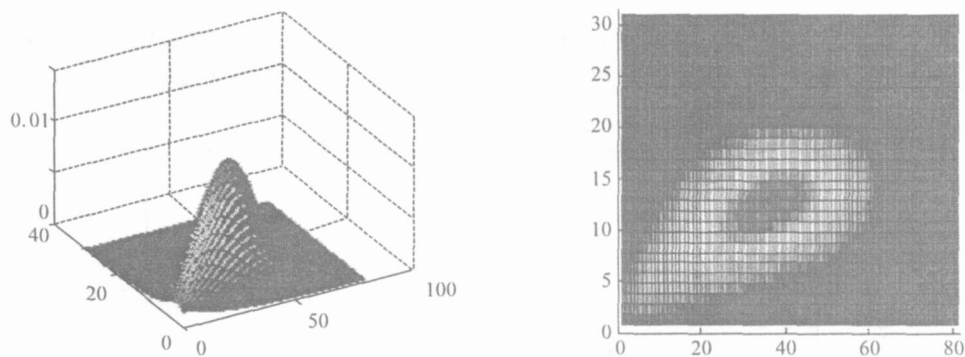


图 2 当 $\tau=0.31428$, $\theta=0.91666$ 时的联合密度函数图、密度函数俯视图

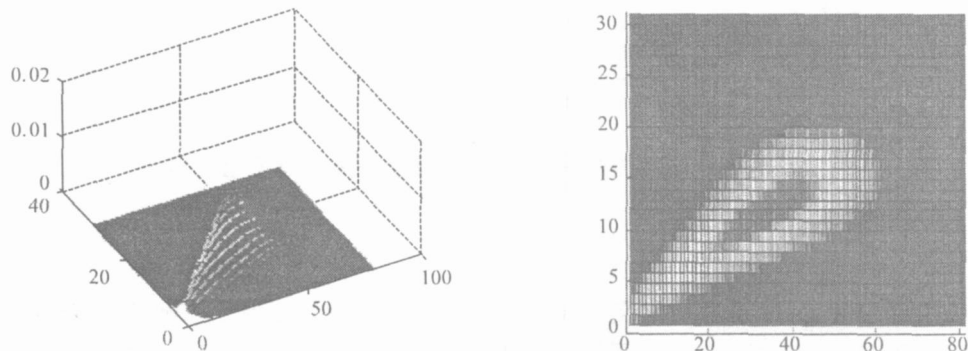


图 3 当 $\tau=0.5$, $\theta=2$ 时的联合密度函数图、密度函数俯视图

由图象可知，当 τ 越靠近1，密度函数表明两变量变化越一致，这正说明了 Copula 函数有效地描述了变量之间的相关关系。

4.3 经验累积概率与理论累积概率比较

为了检验 Copula 函数对联合分布函数的拟合精度，可比较各个观测数据对 (x_i, y_i) 的经验累积概率和理论累积概率是否一致。根据 (10) 式，我们可以直接求观测点最大波高和最大风速系列所构成的各个点对 (x_i, y_i) 的理论累积概率，推求各个观测数据对 (x_i, y_i) 经验累积概率的方法，首先构造一个 $N \times N$ 的二维表（表略）， x 和 y 的观测值分别按升序排列。表中第 i 行和第 j 列所对应的元素为样本联合观测值 (x_i, y_i) 发生的概率 $p(x_i, y_i)$ ，它可由下式估计 [16]：

$$p(x_i, y_i) = Pr(X=x_i, Y=y_i) = n_{ij}/N \tag{11}$$

式中 N 表示联合观测样本数， n_{ij} 表示联合观测值 (x_i, y_i) 发生的次数。则观测数据对 (x_i, y_i) 的累积经验概率可由下式求得 [16]：

$$F_{emp}(x_i, y_i) = Pr(X \leq x_i, Y \leq y_i) = \sum_{m=1}^i \sum_{k=1}^j n_{mk} / N \tag{12}$$

以各个观测数据对 (x_i, y_i) 的理论累积概率值(F)为横坐标，经验累积概率值(F_{emp})为纵坐标打点绘在图4上，散点的线性相关系数为0.976，可见拟合效果比较满意。

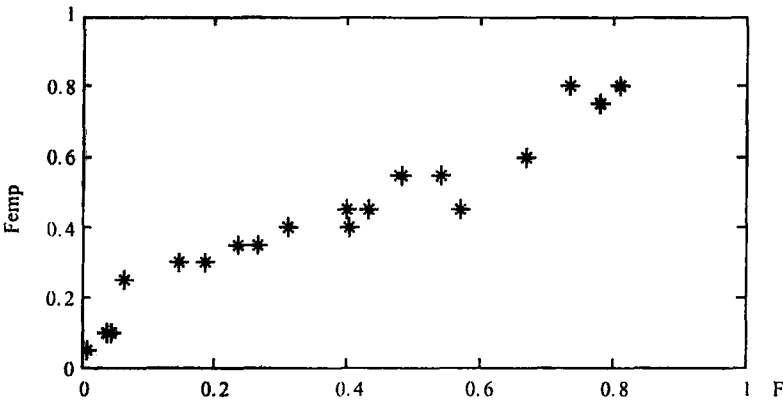


图4 观测数据对经验累积概率值和理论累积概率值比较

5 结论

本文介绍了 Copula 函数的定义、属性和类别，并且探讨了基于 Copula 函数的联合概率法在近海工程中的应用。由于采用 Copula 函数推导出来的多变量联合分布函数的方法具有很大的灵活性、且适用范围广，因此该方法值得在近海工程的实践中推广和应用。

另外,在本文的分析过程中,仍然有一些值得继续深入研究的地方:

(1)对于边缘分布类型的选取和参数的估计,可以根据具体情况考虑更合适的模型,例如遇到了纵向数据,可以考虑运用广义线性模型能更准确地估计参数。

(2)可以对更复杂、更多变量、更接近实际的情况建模,这时要对 Copula 的选取作分析,因为 Copula 维数的变化,可能使复杂度的变化较大,特别是对 Archimedean Copula 来说。而在这方面 Edward W. Frees (2006)已成功的将高维的椭圆 Copula 运用于聚合损失模型。

(3)要尽量收集准确充分的原始数据,否则再好的模型也是 garbage in ,garbage out。

参考文献:

- [1] 仇学艳,王 超,秦崇仁. 多元概率分析方法在海洋工程中的应用现状 [J]. 海洋工程, (2001) 03-0091-05.
- [2] Frees W E, Valdez E A. Understanding relationships using copulas [J]. North Am. Actuarial Journal, 1998, 2 :1 ~ 25.
- [3] Romano C. Calibrating and Simulating Copula Functions : an Application to the Italian Stock Market [R]. 2002.
- [4] Embrechts P, Hoeing A, Juri A. Using copula to bound the value at Risk for functions of dependent risks [J]. Finance and Stochastics, 2003, 7 :145 ~ 167.
- [5] Anne-Catherine F, El A S, Luc P, Nathalie T, Bernard B. Multivariate hydrological frequency analysis using copulas[J]. Water Resources Research January, 2004, 40 (1): W011011 ~ W011012.
- [6] Salvatore G, Francesco S. Asymmetric copula in multivariate flood frequency analysis [J]. Advances in Water Resources, August, 2006, 29 (8):1155 ~ 1167.
- [7] Zhang L, Singh V P. Bivariate flood frequency analysis using the Copula method [J]. Journal of Hydrologic Engineering. March 2006, 150 ~ 164.
- [8] Salvatore G, Francesco S, Francesco N. Lucio U. A 3-Copula function application for design hyetograph analysis [J]. IAHS-AISH Publication, 2005, 293 : 203 ~ 211.
- [9] Shiau J T. Fitting drought duration and severity with two-dimensional copulas [J]. Water Resources Management, October, 2006, 20 (5): 795 ~ 815.
- [10] Hanne T W, Dag M, Havard R. Statistical properties of successive wave heights and successive wave periods [J]. Applied Ocean Research, May/June, 2004, 114 ~ 136.
- [11] 熊立华,郭生练,肖 义,袁汉芳. Copula 联结函数在多变量水文频率分析中的应用 [J]. 武汉大学学报(工学版), 2005.
- [12] Nelsen R B. An Introduction to Copulas [M]. New York : Springer, 1998.
- [13] Gumbel E J. Bivariate exponential distributions [J]. Journal of the American Statistical Association, 1960, (55):
- [14] Clayton D G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence [J]. Biometrika, 1978, (65):
- [15] 刘德辅,施建刚,王 锋. 海洋环境条件联合设计标准 [J]. 海洋学报, 1994, 2:
- [16] 刘次华,万建平. 概率论与数理统计 [M]. 北京:高等教育出版社, 2002.

Application of multivariate probability analysis using copulas to ocean engineering

QIN Zhen-jiang, SUN Guang-hua, YAN Tong-xin, SONG Hua, ZHANG Jing

(School of Mathematical Science and Computational Technology, Central South University, Changsha 410075 China)

Abstract : definition, properties of copula and several special copulas are introduced. Using copula method to ocean engineering, we establish the joint distribution function of the maximum wave height and the wind velocity and check it, the results show that the copula-constructed joint distribution function can well fit the observed data. Using copula and a variety of marginal distribution functions, we can establish the joint distribution function with good dependencies. It can be widely used in the computation of reliability, design standards and failure region.

Key words : copula ;ocean engineering ;wave height ;wind velocity ;multivariate joint distribution