Copula函数

1.什么是Copula函数

- 来自于拉丁语，意为连接，由Sklar提出
- lar认为，对于N个随机变量的联合分布，可以将其分解为这N个变量各自的边缘分布和一个Copula函数，从而将变量的随机性和耦合性分离开来。其中，随机变量各自的随机性由边缘分布进行描述，随机变量之间的耦合特性由Copula函数进行描述。
- **Copula理论的数学表达**

  假设 $X_1, X_2, \ldots, X_N$ 是 N 个随机变量，它们各自的边缘分布分别为 $F_1(x_1), F_2(x_2), \ldots, F_N(x_N)$，它们的联合分布为 $H(x_1, x_2, \ldots, x_N)$，则存在一个将边缘分布和联合分布"连接"起来的函数 $C(\cdot)$，使得：

  $$H(x_1, x_2, \ldots, x_N) = C(F_1(x_1), F_2(x_2), \ldots, F_N(x_N))$$

- 考察两个随机变量X和Y，其连续累积分布函数分别为F_X和F_Y。通过分别在两个随机变量上应用概率积分变换，得到$U = F_X(X)$和$V = F_Y(Y)$。U和V都是服从$[0,1]$上均匀分布的变量，它们的相关性取决于X和Y是否相关：如果X和Y是不相关的，那么U和V也是不相关的，因为这个转换是可逆的，**定义X和Y之间的相关性等价于U和V之间的相关性**。由于U和V是均匀分布的随机变量，所以问题被简化为定义一个在两个均匀分布之上的联合分布，这就是 Copula。Copula 函数的基本思想就是，**通过把边缘变量转化为均匀分布变量而不再需要考察很多不同的边缘分布以简化问题，然后再把相关性定义为一个在均匀分布之上的联合分布。**

2.为什么要引入Copula函数

- 当多个随机变量之间不独立，对于联合建模比较困难。但当已知各随机变量之间的边缘分布时，Copula函数便可以很好地对其相关性进行建模。

3.Copula函数分类

- 1.椭圆Copula函数簇
  - t Copula函数
  - Gaussian Copula函数
- 2.阿基米德Copula函数簇

Name	Note
Clayton	下尾相关，对分布的下尾部的变化比较敏感，能够捕捉到下尾相关的变化
Gumbel-Hougaard（GH）	上尾相关，对分布的上尾部的变化比较敏感，能够捕捉到上尾相关的变化
Ali-Mikhail-Haq（AMH）	尾相关
Frank	无尾相关，对称尾部，无法捕捉到随机变量之间的非对称的尾部相关
Joe	上尾相关

  - 1.2.1 Clayton Copula函数

    二维 Clayton Copula 函数仅能够适用于变量存在正相关的情况，其分布函数如下：

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

    概率密度函数如下：

    $$c(u,v) = \frac{(1+\theta)u^{-1-\theta}v^{-1-\theta}}{(u^{-\theta} + v^{-\theta} - 1)^{\frac{1+2\theta}{\theta}}} \mathrm{m}, \theta \in [0, \infty)$$

    式中，u、v均为边缘分布函数；θ为Copula函数参数。
  - 1.2.2 Gumbel–Hougaard（GH）Copula函数

    二维 Gumbel-Hougaard（GH）Copula 函数仅适用于变量存在正相关的情况，其分布函数如下：

    $$C(u,v) = \exp\{-[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{1/\theta}\} \in [-1, \infty)$$

    概率密度函数如下：

    $$c(u,v) = \frac{(-\ln u)^{\theta-1}(-\ln v)^{\theta-1}[\theta - 1 + (-\ln u)^{\theta} + (-\ln v)^{\theta}]^{1/\theta}}{uve^{[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{1/\theta}}}, \theta \in [-1, \infty)$$

    式中，u、v均为边缘分布函数；θ为Copula函数参数。
  - 1.2.3 Ali–Mikhail–Haq（AMH）Copula函数

    二维 Ali-Mikhail-Haq（AMH）Copula 函数对相关关系的正负不作要求，但只适用于相关关系较弱的随机变量，其分布函数如下：

    $$C(u,v) = uv / [1 - \theta(1-u)(1-v)], \theta \in [-1, 1)$$

    概率密度函数如下：

    $$c(u,v) = \frac{-1 + \theta^2(-1 + u + v - uv) - \theta(-2 + u + v - uv)}{[-1 + \theta(1-u)(1-v)]^3}, \theta \in [-1, 1)$$

    式中，u、v均为边缘分布函数；θ为Copula函数参数。
  - 1.2.4 Frank Copula函数

    二维 Frank Copula 函数要求最低，对相关性正负和程度均无限制，其分布函数如下：

    $$C(u,v) = -\frac{1}{\theta}\ln[1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{(\exp(-\theta) - 1)^2}], \theta \in R$$

    概率密度函数如下：

    $$c(u,v) = \frac{\theta \exp(\theta(1+u+v))(\exp(\theta) - 1)}{\exp(\theta) - \exp(\theta(1+u)) - \exp(\theta(1+v)) + \exp(\theta(u+v))}, \theta \in R$$

    式中，u、v均为边缘分布函数；θ为Copula函数参数。
- 3.二次型Copula函数

藤Copula函数

- 为何引入：常规Copula能很好描述二维变量的相关性，但是高维Copula函数面临维数灾，藤Copula将高维分解为多个二维的乘积

Copula函数构造过程

- 建步骤如下：

  ①确定随机变量的边缘分布；

  ②根据随机变量相关性特点，选取合适的 Copula 函数，以便能较好地描述随机变量之间的相关关系。常用的 Copula 函数[13]主要有正态 Copula 函数、t-Copula 函数、Gumbel Copula 函数、Clayton Copula 函数和 Frank Copula 函数等；

  ③根据已选择的 Copula 函数，估计 Copula 模型中的未知参数，本文采用极大似然方法对 Copula 模型进行参数估计。
- 基于 Copula 函数产生场景的基本思想是：首先，产生随机变量的边缘分布函数，由此构建多元 Copula 函数；其次，离散化 Copula 函数，将所得到的 Copula 函数离散场景，通过边缘分布函数的逆运算，得到原联合分布函数场景。具体步骤为：

  1）产生满足 Copula 函数分布的 $N \times M$ 维数据样本，N 为样本总数，M 为随机变量维数；

  2）确定场景数目 S，采用 K-均值聚类[15]方法将 $N \times M$ 阶数据样本分为 S 类，将各类中心(该类中所有样本的均值)$\boldsymbol{u}^s = [u_1^s, u_2^s, \cdots, u_d^s]$作为场景的分位点；统计落在该类中的样本占样本总数的比例，将其作为各类的概率值 $p_s(s = 1,2,\cdots,S)$，从而保证各类中心点的估计相对于整个样本空间是无偏的；

  3）采用式 $x_i^s = F_i^{-1}(u_i^s)$ 将 $\boldsymbol{u}^s = [u_1^s, u_2^s, \cdots, u_d^s]$ $(s = 1,2,\cdots,S)$转换为原联合分布函数场景，即可获得所需场景的分位点，各分位点对应的概率为 p_s $(s = 1,2,\cdots,S)$。
- **本文运用 Copula 理论对雨型特征参数总雨量和峰值雨强两变量联合概率分布进行求解时，选取 Gumbel、Clayton 和 Frank 三种函数构建二维联合分布函数模型，并优选出拟合效果最好的参数作为最终 Copula 函数，优选的具体步骤如下：**

  （1）应用水文频率分析方法，选用相应随机变量概率分布函数模拟总雨量和峰值雨强的边缘分布函数，并运用非参数 Kolmogorov-Smimov 检验方法对两变量边缘分布函数进行拟合检验；

  （2）选取 Pearson 线性相关系数 r、Kendall 秩相关系数 τ、和 Spearman 秩相关系数 ρ 三种相关性度量指标分析总雨量和峰值雨强之间的相关性；

  （3）采用相关性指标法分别估计二维 Gumbel、Clayton 和 Frank 函数的未知参数，并运用非参数 Kolmogorov-Smimov 检验方法对每个联合概率分布函数进行拟合检验；

  （4）运用 RMSE 准则法、AIC 信息准则法和 BIC 准则法对二维 Gumbel、Clayton 和 Frank 函数进行拟合优度评价，优选出拟合效果最好的 Copula 函数模型作为总雨量和峰值雨强的联合概率分布模型。

