Copula的统计估计

杨静平

北京大学数学科学学院金融数学系

2019年10月

Outline

1. Copula的统计估计与模拟
    - Copula的统计推断
    - 精确最大似然估计方法
    - IFM 方法
    - CML 方法
    - 非参数估计
    - Kernel copula

Outline

Copula的统计推断

数据为n维向量列$\{\mathbf{Y}_t, t = 1, 2, \cdots, T\}$。可以代表金融资产的收益，或者股票的板块指数等。记

$$\mathbf{Y}_t = (Y_{1t}, Y_{2t}, \cdots, Y_{nt})'$$

的分布和密度分别为$F(\mathbf{y})$和$f(\mathbf{y})$。其中，

$$\mathbf{y} = (y_1, \cdots, y_n),$$

边缘分布分别记为$f_j(y_j), F_j(y_j)$. 下面我们考虑边缘分布是连续的情况。考虑n维的数据。

Outline

精确最大似然估计方法

对于分布函数

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

有密度函数

$$f(x_1,x_2,\cdots,x_n)=c(F_1(x_1),\cdots,F_n(x_n))\prod_{j=1}^{n}f_j(x_j)$$

其中，

$$c(u_1,u_2,\cdots,u_n)=\frac{\partial C(u_1,u_2,\ldots,u_n)}{\partial u_1\ldots\partial u_n}$$

令 $\chi_t = (x_{1t}, x_{2t}, \cdots, x_{nt}), t = 1, \cdots, T$ 为样本点序列。 则有似然函数

$$l(\theta) = \sum_{t=1}^{T} lnc(F_1(x_{1t}), F_2(x_{2t}), \cdots, F_n(x_{nt})) + \sum_{t=1}^{T}\sum_{j=1}^{n} lnf_j(x_{jt})$$

其中，θ是所有边缘分布和copula的参数集。
需要给出所有的边缘分布的估计， 以及定义相应的copula函数。

希望考虑最大似然估计

$$\hat{\theta}_{MLE} = \max_{\theta \in \Theta} l(\theta).$$

在给定的条件下，有

$$\sqrt{T}(\hat{\theta}_{MLE} - \theta_0) \to N(0, \mathfrak{F}^{-1}(\theta_0))$$

其中，$\mathfrak{F}(\theta_0)$为Fisher's 信息矩阵，θ_0是真值。

例子：多维正态分布。假设R是一个对称的正定的矩阵， 对角元素都为1， Φ_R是斜方差阵为R的标准正态分布,

$$C(u_1,u_2,\cdots,u_n)=\Phi_R(\Phi^{-1}(u_1),\Phi^{-1}(u_2),\cdots,\Phi^{-1}(u_n))$$

可计算似然函数

$$l(\theta)=-\frac{T}{2}\ln(|R|)-\frac{1}{2}\sum_{t=1}^{T}\epsilon_t'(R^{-1}-I)\epsilon_t$$

其中,

$$\epsilon_t'=(\Phi^{-1}(u_{1t}),\Phi^{-1}(u_{2t}),\cdots,\Phi^{-1}(u_{nt}))$$

Outline

IFM 方法

利用最大似然估计计算量大。因此， 可以采用另外的方法：
第一步： 根据边缘的数据估计边缘分布的参数θ_1,

$$\hat{\theta}_1 = ArgMax_{\theta_1} \sum_{t=1}^{T} \sum_{j=1}^{n} ln f_j(x_{jt}; \theta_1).$$

第二步： 根据得到的$\hat{\theta}_1$， 估计copula的参数θ_2,

$$\hat{\theta}_2 = ArgMax_{\theta_2} \sum_{t=1}^{T} \ln c(F_1(x_1), F_2(x_2), \cdots, F_n(x_n); \theta_2, \hat{\theta}_1).$$

这种方法称为inference for the margins (IFM). IFM的估计量定义为向量

$$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)'.$$

考虑l是整体的似然函数，l_j是第j个边缘的似然函数, l_C是copula C的似然函数。因此IFM的解满足

$$\left(\frac{\partial l_1}{\partial \theta_{11}}, \frac{\partial l_2}{\partial \theta_{12}}, \cdots, \frac{\partial l_n}{\partial \theta_{1n}}, \frac{\partial l_c}{\partial \theta_2}\right)' = 0$$

而最大似然函数满足

$$\left(\frac{\partial l}{\partial \theta_{11}}, \frac{\partial l}{\partial \theta_{12}}, \cdots, \frac{\partial l}{\partial \theta_{1n}}, \frac{\partial l}{\partial \theta_2}\right)' = 0.$$

Outline

CML 方法

将样本数据$\{x_{1t}, x_{2t}, \cdots, x_{nt}\}$转换为均匀变量数据$\{u_{1t}, u_{2t}, \cdots, u_{nt}\}$，然后估计copula参数。

第一步： 根据边缘的数据估计边缘分布的经验分布,$\hat{F}_i(x_{it})$, $i = 1, \cdots, n$. 使用非参数方法。

第二步： 通过最大似然估计来估计copula的参数：

$$\hat{\theta}_2 = ArGMax_{\theta_2} \sum_{t=1}^{T} lnc(\hat{F}_1(x_{it}), \hat{F}_1(x_{1t}), \cdots, \hat{F}_n(x_{nt}); \theta_2).$$

这种方法称为Canonical Maximum Likelihood方法 (CML)。

Outline

非参数估计

令

$$X_t = (X_{1t}, X_{2t}, \ldots, X_{nt}), \quad t = 1, \cdots, T$$

为i.i.d. 序列，分布为F， 以及连续的边缘分布F_j。令$\{x_1^{(t)}, \ldots, x_n^{(t)}\}$为次序统计量，$\{r_1^{(t)}, \ldots, r_n^{(t)}\}$为样本的Rank统计量。

Deheuvels' empirical copula. 经验copula被定义在格子点上，满足

$$\mathcal{L} = \{(\frac{t_1}{T}, \frac{t_2}{T}, \cdots, \frac{t_n}{T}) : 1 \le j \le n, t_j = 0, 1, \cdots, T\}$$

则经验copula采用如下的数据来估计：

$$\hat{C}(\frac{t_1}{T}, \frac{t_2}{T}, \cdots, \frac{t_n}{T}) = \frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{n} I(r_j^{(t)} \le t_j).$$

经验copula频率，定义为

$$\hat{c}\left(\frac{t_1}{T}, \frac{t_2}{T}, \cdots, \frac{t_n}{T}\right) = \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}\cdots\sum_{i_n=1}^{2}(-1)^{\sum_{j=1}^{n} i_j} \times \hat{C}\left(\frac{t_1-i_1+1}{T}, \frac{t_2-i_2+1}{T}, \cdots, \frac{t_n-i_n+1}{T}\right)$$

Bernstein copula：定义为

$$B_T(C)(u_1, u_2, \cdots, u_n) = \sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\cdots\sum_{t_n=1}^{n} B_{t_1,T}(u_1)B_{t_2,T}(u_2)\cdots B_{t_n,T}(u_n)\hat{C}\left(\frac{t_1}{T}, \frac{t_2}{T}, \cdots, \frac{t_n}{T}\right).$$

其中，

$$B_{i,n}(x) = \binom{n}{i}x^i(1-x)^{n-i}.$$

Outline

1. Copula的统计估计与模拟
   - Copula的统计推断
   - 精确最大似然估计方法
   - IFM 方法
   - CML 方法
   - 非参数估计
   - Kernel copula

Kernel copula

Kernel意味着一种函数形式，通常用于平滑目的。核函数k_{ij}满足

$$\int_R k_{ij}(x)dx = 1, i = 1, 2, \cdots, m; j = 1, 2, \cdots, n.$$

以及

$$K_i(x; h) = \prod_{j=1}^{n} k_{ij}(\frac{x_j}{h_j}), i = 1, 2, \cdots, m.$$

其中带宽h_j是T的正的函数，满足

$$|h| + \frac{1}{T|h|} \to 0, T \to \infty.$$

Y_{jt} 的密度函数$f_j(y_{ij})$需要通过如下的方法来估计：

$$\hat{f}_j(y_{ij}) = \frac{1}{Th_j}\sum_{t=1}^{T} k_{ij}(\frac{y_{ij}-Y_{it}}{h_j})$$

联合密度函数

$$\hat{f}(y_{i1},\cdots,y_{in}) = \frac{1}{T|h|}\sum_{t=1}^{T}\prod_{j=1}^{n} k_{ij}(\frac{y_{ij}-Y_{it}}{h_j}).$$

因此可以得到Y_{jt}的边缘分布函数

$$\hat{F}_j(y_{ij}) = \int_{-\infty}^{y_{ij}} \hat{f}_j(x)dx$$

以及$\mathbf{Y}_t$的分布函数

$$\hat{F}(y_i) = \int_{-\infty}^{y_{i1}} \cdots \int_{-\infty}^{y_{in}} \hat{f}(x)dx.$$

实际中，Gaussian 核函数是一种选择：

$$k_{ij}(x) = \psi(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2}).$$

可以通过

$$\hat{C}(u_i) = \hat{F}(\hat{\xi}_i)$$

得到copula 函数。其中，

$$\hat{\xi}_i = (\hat{\xi}_{i1}, \cdots, \hat{\xi}_{in})'$$

$$\hat{\xi}_{ij} = \inf_{y \in R} \{y : \hat{F}_j(y) \geq u_{ij}\}.$$

如果使用Gaussian Kernel 函数

$$k_{ij}(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2).$$

则可以得到

$$\hat{F}_j(y_{ij}) = \frac{1}{Th_j} \sum_{t=1}^{T} \Phi(\frac{y_{ij} - Y_{jt}}{h_j})$$

以及

$$\hat{F}(y_{i1}, \cdots, y_{in}) = \frac{1}{T|h|} \sum_{t=1}^{T} \prod_{j=1}^{n} \Phi(\frac{y_{ij} - Y_{jt}}{h_j}).$$