Schätzung von Copulas

Gegeben sei ein Satz multi-dimensionaler Daten. Gesucht ist eine Copula und die Randverteilungen die diesem Datensatz am besten entsprechen.

1. Frage: Welche Familie von (bekannten) Copulas eignet sich am besten?
Antwort: Visueller Vergleich der graphischen Darstellungen von Daten bzw. bekannten Copulas, Berechnung der empirischen Koeffizienten der Tail-Abhängigkeit und Auswahl von dazu passenden Copula Familien.

2. Frage: Schätzung der Parameter einer vorspezifizierten Copula Familie.

Gegeben: Eine Stichprobe $\{X_1, X_2, \ldots, X_n\}$ aus eine Gesamtverteilung F mit stetigen Randverteilungen F_1, F_2,..., F_d.

Gesucht: Ein Schätzer $\widehat{\theta}$ des Parameter-Vektors θ der eindeutigen Copula C_θ, für die $F(x) = C(F_1(x_1), \ldots, F_d(x_d))$ gilt.

Die Schätzer $\widehat{\theta}$ für C_R^{Ga}, $C_{\nu,R}^t$, C_θ^{Cl} und C_θ^{Gu}

$$
\begin{array}{ll}
C_R^{Ga} = \phi_R^d(\phi^{-1}(u_1), \ldots, \phi^{-1}(u_d)) & R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2) \\
C_{\nu,R}^t = t_{\nu,R}^d(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)) & R_{ij} = \sin(\pi(\rho_\tau)_{ij}/2) \\
C_\theta^{Gu}(u) = \exp\left(-[(-\ln u_1)^\theta + \ldots + (-\ln u_d^\theta]^{1/\theta}\right) & \theta = 1/(1-(\rho_\tau)_{ij}) \\
C_\theta^{Cl}(u) = (u_1^{-\theta} + \ldots + u_d^{-\theta} - d + 1)^{-1/\theta} & \theta = 2(\rho_\tau)_{ij}/(1-(\rho_\tau)_{ij})
\end{array},
$$

wobei

$$
\begin{aligned}
(\rho_\tau)_{ij} &= \rho_\tau(X_{k,i}, X_{k,j}) \\
&= P((X_{k,i} - X_{l,i})(X_{k,j} - X_{l,j}) > 0) - P((X_{k,i} - X_{l,i})(X_{k,j} - X_{l,j}) < 0) \\
&= E(sign((X_{k,i} - X_{l,i})(X_{k,j} - X_{l,j}))).
\end{aligned}
$$

Standard Schätzer für Kendalls Tau:

$$
\widehat{\rho_\tau}_{ij} = \binom{n}{2}^{-1} \sum_{1 \le k < l \le n} sign((X_{k,i} - X_{l,i})(X_{k,j} - X_{l,j})).
$$

Schätzung von Gauss'schen Copulas und t-Copulas

Es kann passieren, dass $\widehat{R} = (\widehat{R}_{ij})$ nicht positive definit ist, wobei

$$\widehat{R}_{ij} = \sin(\pi \widehat{\rho_\tau}_{ij}/2).$$

$\widehat{R}$ wird durch eine Korrelationsmatrix R^* ersetzt, wobei R^* "unweit" von $\widehat{R}$ liegt.

Algorithmus 6 *(Eigenwert-Ansatz, siehe Rousseeuw und Molenberghs 1993)*

- *Berechne die Spektralzerlegung $\widehat{R} = \Gamma\Lambda\Gamma^T$ von $\widehat{R}$, wobei Λ eine Diagonalmatrix ist, die die Eigenwerte von $\widehat{R}$ enthält, und Γ eine orthogonale Matrix deren Spalten den Eigenvektoren von $\widehat{R}$ entsprechen.*

- *Ersetze die negativen Eigenwerte in λ durch eine kleine Zahl $\delta > 0$ um $\tilde{\Lambda}$ zu erhalten.*

- *Berechne $\tilde{R} = \Gamma\tilde{\Lambda}\Gamma^T$. $\tilde{R}$ ist symmetrisch und positive definit aber nicht unbedingt eine Korrelationsmatrix, weil die Diagonal-Elemente $\widehat{R}_{ii}$ ungleich 1 sein könnten.*

- *Setze $\widehat{R} = D\tilde{R}D$ wobei D eine diagonale Matrix mit $D_{k,k} = 1/\sqrt{\tilde{R}_{k,k}}$ ist.*

t-Copulas: Schätzung des Parameters ν der Freiheitsgrade

1. Schätzung der univariaten Randverteilungsfunktionen F_1, F_2, ..., F_d. Seien $\widehat{F}_1$, ..., $\widehat{F}_d$ die dazugehörigen Schätzer.

2. Bildung einer Pseudo-Stichprobe der Copula:

$$\widehat{U}_k = (\widehat{U}_{k,1}, \widehat{U}_{k,2}, \ldots, \widehat{U}_{k,d}) := (\widehat{F}_1(X_{k,1}), \ldots, \widehat{F}_d(X_{k,d})),$$

   für $k = 1, 2, \ldots, n$ (siehe Genest und Rivest 1993).

$\widehat{F}_k$ kann folgendermaßen erzeugt werden:

- Parametrische Schätzung: $\widehat{F}_k$ ist eine parametrische Verteilungsfunktion wobei der Parameter zB. mit einem Maximum Likelihood (ML) Ansatz geschätzt wird.

- Nicht Parametrische Schätzung: $\widehat{F}_i$ ist die empirische Verteilungsfunktion $\widehat{F}_i(x) = \frac{1}{n+1} \sum_{t=1}^{n} I_{X_{t,i} \leq x}$, $1 \leq i \leq d$.

ML-Schätzung von ν: $\nu = \arg\max_\xi \ln L(\xi; \widehat{U}_1, \widehat{U}_2, \ldots, \widehat{U}_n)$ wobei

$$L(\xi; \widehat{U}_1, \widehat{U}_2, \ldots, \widehat{U}_n) = \Pi_{k=1}^{n} c_{\xi,R}^{t}(\widehat{U}_k)$$

und $c_{\xi,R}^{t}$ die Dichte der t-Copula $C_{\xi,R}^{t}$ ist.

Daraus folgt

$$\ln L(\xi; \widehat{U}_1, \widehat{U}_2, \ldots, \widehat{U}_n) = \sum_{k=1}^{n} \ln g_{\xi,R}(t_\xi^{-1}(\widehat{U}_{k,1}), \ldots, t_\xi^{-1}(\widehat{U}_{k,d})) - \sum_{k=1}^{n} \sum_{j=1}^{d} \ln g_\xi(t_\xi^{-1}(\widehat{U}_{k,j})),$$

wobei $g_{\xi,R}$ die Gesamtdichte einer standard d-dimensionalen
t-Verteilung mit Verteilungsfunktion $t_{\xi,R}^{d}$ ist,
und
g_ξ die Dichte einer univariaten standard t-Verteilung mit
ξ Freiheitsgraden ist.