Multivariate Archimedische Copulas

Definition 23 *Eine Funktion $g\colon [0,\infty) \to [0,\infty)$ heißt vollständig monoton wenn alle höheren Ableitungen von g existieren und folgende Ungleichungen gelten für $k = 0, 1, 2, \ldots$:*

$$(-1)^k \left(\frac{d^k}{ds^k} g(s) \right) \bigg|_{s=t} \geq 0, \forall t \in (0,\infty)$$

Theorem 25 *(Kimberling 1974)*
Sei $\phi\colon [0,1] \to [0,\infty]$ eine stetige und streng monoton fallende Funktion, sodass $\phi(0) = \infty$ und $\phi(1) = 0$. Die Funktion $C\colon [0,1]^d \to [0,1]$, $C(u) = \phi^{-1}(\phi(u_1) + \phi(u_2) + \ldots + \phi(u_d))$ ist eine Copula für $d \geq 2$ dann und nur dann wenn ϕ^{-1} vollständig monoton in $[0,\infty)$ ist.

Lemma 6 *Eine Funktion $\psi\colon [0,\infty) \to [0,\infty)$ ist die Laplace-Stieltjes Transformation einer Verteilungsfunktion G in $[0,\infty)$ ($\psi(s) = \int_0^\infty e^{-sx} dG(x)$, $s \geq 0$) dann und nur dann wenn ψ vollständig monoton und $\psi(0) = 1$.*

Theorem 26 *Sei G eine Verteilungsfunktion in $[0,\infty)$, sodass $G(0) = 0$ und sei ψ die Laplace-Stieltjes Transformation von G*

$$\psi(s) = \int_0^\infty e^{-sx} dG(x), \text{ für } s \geq 0.$$

Sei X eine ZV mit Verteilungsfunktion G und seien $U_1, U_2, \ldots, U_d$ bedingt unabhängige Zufallsvariablen in $[0,1]$ für ein gegebenes $X = x$ mit folgender bedingter Verteilungsfunktion:

$$F_{U_k|X=x}(u) = \exp(-x\psi^{-1}(u)) \text{ für } u \in [0,1].$$

Es gilt dann

$$P(U_1 \leq u_1, U_2 \leq u_2, \ldots, U_d \leq u_d) = \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2) + \ldots + \psi^{-1}(u_d)).$$

und die Verteilungsfunktion von $U = (U_1, U_2, \ldots, U_d)$ ist eine Archimedische Copula mit Generator ψ^{-1}.

Vorteile und Nachteile Archimedischer Copulas:

- Modellierung einer breiteren Klasse von Abhängigkeitsstrukturen
- Darstellung in geschloßener Form möglich
- Wenige freie Parameter vorhanden
- Technische Voraussetzungen für die Generator Funktionen multivariater Archimedische Copulas.

Simulation von Gauss'schen Copulas und t-Copulas

Sei $R \in \mathbb{R}^{d\times d}$ eine symmetrische positiv definite Matrix. Sei $AA^T = R$ die Cholesky Zerlegung von R ($A \in \mathbb{R}^{d\times d}$). Falls $Z_1, Z_2, \ldots, Z_d \sim N(0,1)$ unabhängig dann gilt $\mu + AZ \sim N_d(\mu, R)$.

Algorithmus 1 *zur Erzeugung eines Zufallsvektors $U = (U_1, U_2, \ldots, U_d)$ dessen Verteilungsfunktion die Copula C_R^{Ga} ist.*

- *Berechne die Cholesly Zerlegung A von R: $R = AA^T$.*
- *Simuliere d unabhängige standard normal verteilte ZV $Z_1, Z_2, \ldots, Z_d \sim N(0,1)$*
- *Setze $X = AZ$*
- *Setze $U_k = \phi(X_k)$ für $k = 1, 2, \ldots, d$, wobei ϕ die Verteilungsfunktion der standard Normalverteilung ist.*
- *$U = (U_1, U_2, \ldots, U_d)$ hat Verteilungsfunktion C_R^{Ga}.*

Algorithmus 2 *zur Erzeugung eines Zufallsvektors* $U = (U_1, U_2, \ldots, U_d)$ *dessen Verteilungsfunktion die Copula* $C^t_{\nu,R}$ *ist.*

- *Berechne die Cholesly Zerlegung* A *von* R*:* $R = AA^T$*.*
- *Simuliere* d *unabhängige standard normal verteilte ZV* $Z_1, Z_2, \ldots, Z_d \sim N(0,1)$
- *Simuliere eine ZV* $S \sim \chi^2_\nu$ *unabhängig von* $Z_1, \ldots, Z_d$*.*
- *Setze* $Y = AZ$
- *Setze* $X = \frac{\sqrt{\nu}}{\sqrt{S}} Y$
- *Setze* $U_k = t_\nu(X_k)$ *für* $k = 1, 2, \ldots, d$*, wobei* t_ν *die Verteilungsfunktion einer standard* t*-Verteilung mit* ν *Freiheitsgraden ist.*
- $U = (U_1, U_2, \ldots, U_d)$ *hat Verteilungsfunktion* $C^t_{\nu,R}$*.*

Simulationen der Gumbel und Clayton Copulas

Aus Satz 26 lässt sich ein Algorithmus zur Erzeugung dieser Copulas konstruieren.

Algorithmus 3 *zur Erzeugung eines Zufallsvektors* $U = (U_1, U_2, \ldots, U_d)$ *dessen Verteilungsfunktion die Archimedische Copula* $C(u) = \phi^{-1}(\phi(u_1) + \phi(u_2) + \ldots + \phi(u_d))$ *mit Generator* ϕ *ist.*

- *Simuliere eine Variable* X *mit Verteilungsfunktion* G*, sodass die Laplace-Stieltjes Transformation* ψ *von* G *die inverse Funktion des Generators der gesuchten Copula ist,* $\psi = \phi^{-1}$*.*

- *Simuliere unabhängige gleichverteilte Zufallsvariablen* V_1,V_2,..., V_d *in* $[0, 1]$.

- *Setze* $U = (\psi(-\ln(V_1)/X), \psi(-\ln(V_2)/X), \ldots, \psi(-\ln(V_d)/X))$. U *hat Verteilungsfunktion* $C(u)$*.*

Der Generator $\phi(t) = (t^{-\theta} - 1)/\theta$, $\theta > 0$ erzeugt die Clayton Copula C_θ^{Cl}. Aber auch $\tilde{\phi}(t) = t^{-\theta} - 1$ ist ein Generator der Clayton Copula.

Für $X \sim Gamma(1/\theta, 1)$ d.h. $f_X(x) = x^{1/\theta - 1} e^{-x} / \Gamma(1/\theta)$ gilt:

$$E(e^{-sX}) = \int_0^\infty e^{-sx} \frac{1}{\Gamma(1/\theta)} x^{1/\theta-1} e^{-x} dx = (s+1)^{-1/\theta} = \tilde{\phi}^{-1}(s).$$

Algorithmus 4 *zur Erzeugung eines Zufallsvektors $U = (U_1, U_2, \ldots, U_d)$ dessen Verteilungsfunktion die Clayton Copula C_θ^{Cl} ist.*

- *Simuliere $X \sim Gamma(1/\theta, 1)$.*
- *Simuliere unabhängige gleichverteilte Zufallsvariablen V_1,V_2,..., V_d in $[0,1]$.*
- *die Verteilungsfunktion des Vektors*

  $$U = (\psi(-\ln(V_1)/X), \psi(-\ln(V_2)/X), \ldots, \psi(-\ln(V_d)/X))$$

  ist die Clayton Copula C_θ^{Cl}.

Die Simulation von Gumbel Copulas C_θ^{Gu} :

Sei X eine positive stabile ZV, $X \sim St(1/\theta, 1, \gamma, 0)$
mit $\gamma = (\cos(\pi/(2\theta)))^\theta$, $\theta > 1$.

Die Laplace-Stieltjes Transformation von F_X ist $\phi(t) = \exp(-t^{1/\theta})$.

Simulation von $Z \sim ST(\alpha, \beta, 1, 0)$: siehe Nolan 2002.

Für $\gamma \neq 1$ gilt: $X = \delta + \gamma Z \sim St(\alpha, \beta, \gamma, \delta)$.

Alternativer Ansatz:
Sei $\theta \geq 1$ und $\bar{F}(x) = 1 - F(x) = \exp(-x^{1/\theta})$ für $x \geq 0$.
Sei $V \sim U(0,1)$.
Sei S eine von V unabhängige ZV. mit Dichtefunktion

$$h(s) = (1 - 1/\theta + s/\theta)\exp(-s)$$

Sei $(Z_1, Z_2)^T = (VS^\theta, (1-V)S^\theta)$.
Die Verteilungsfunktion von $(\bar{F}(Z_1), \bar{F}(Z_2))^T$ ist C_θ^{Gu}.
Überzeugen Sie sich (Hausübung)!

Algorithmus 5 *zur Erzeugung eines Zufallsvektors $U = (U_1, U_2, \ldots, U_d)$ dessen Verteilungsfunktion die Gumbel Copula C_θ^{Cu} ist.*

- *Simuliere zwei unabhängige ZV. $V_1, V_2 \sim U(0,1)$.*
- *Simuliere zwei unabhängige ZV. $W_1 \sim \Gamma(1,1)$, $W_2 \sim \Gamma(2,1)$*
- *Setze $S = I_{V_2 \leq 1/\theta} W_1 + I_{V_2 > 1/\theta} W_2$.*
- *Setze $(Z_1, Z_2) = (V_1 S^\theta, (1-V_1)S^\theta)$.*
- *Die Verteilungsfunktion von $U = (\exp(-Z_1^{1/\theta}), \exp(-Z_2^{1/\theta})^T$ ist C_θ^{Gu}.*

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.