Kendall's Tau und Spearman's Rho

Seien $(x,y)^T$ und $(\tilde{x},\tilde{y})^T$ zwei Beobachtungen von einem Zufallsvektor $(X,Y)^T$. $(x,y)^T$ und $(\tilde{x},\tilde{y})^T$ heißen *übereinstimmend* falls $(x-\tilde{x})(y-\tilde{y}) > 0$ und *nicht übereinstimmend* falls $(x-\tilde{x})(y-\tilde{y}) < 0$.

Definition 10 *Sei $(X_1,X_2)^T$ ein Zufallsvektor mit stetigen Randverteilungen. Der Kendall's Tau ist für $(X_1,X_2)^T$ folgendermaßen definiert:*

$$\rho_\tau(X_1,X_2) = P((X_1-X_1')(X_2-X_2') > 0) - P((X_1-X_1')(X_2-X_2') < 0),$$

wobei $(X_1',X_2')^T$ is eine unabhängige Kopie von $(X_1,X_2)^T$.

Äquivalent: $\rho_\tau(X_1,X_2) = E(sign[(X_1-X_1')(X_2-X_2')])$.

Im d-dimensionalen Fall $X \in \mathbb{R}^d$: $\rho_\tau(X) = cov(sign(X-X'))$, wobei $X' \in \mathbb{R}^D$ eine unabhängige Kopie von $X \in \mathbb{R}^d$ ist.

Der Kendall's Tau der Stichprobe:

Sei $\{(x_1,y_1)^T,(x_2,y_2)^T,\ldots,(x_n,y_n)^T\}$ eine Stichprobe von n Beobachtungen des Zufallsvektors $(X,Y)^T$ dessen Randverteilungen stetig sind. Sei c die Anzahl der übereinstimmenden Paare und d die Anzahl der nicht übereinstimmenden Paare aus der Stichprobe.

$$\tilde{\rho}_\tau(X,Y) = \frac{c-d}{c+d} \stackrel{a.s.}{=} \frac{c-d}{n(n-1)/2}$$

Definition 11 *Sei $(X_1, X_2)^T$ ein Zufallsvektor mit stetigen Randverteilungen. Der Spearman's Rho ist für $(X_1, X_2)^T$ folgendermaßen definiert:*

$$\rho_S(X_1, X_2) = 3(P((X_1 - X_1')(X_2 - X_2'') > 0) - P((X_1 - X_1')(X_2 - X_2'') < 0)),$$

wobei $(X_1', X_2')^T$, $(X_1'', X_2'')^T$ unabhängige Kopien von $(X_1, X_2)^T$ sind.

Äquivalente Definition (ohne Beweis):
Seien F_1 und F_2 die stetigen Randverteilungen von $(X_1, X_2)^T$. Es gilt $\rho_S(X_1, X_2) = \rho_L(F_1(X_1), F_2(X_2))$, d.h. der Spearman's Rho ist die lineare Korrelation der eindeutigen Copula von $(X_1, X_2)^T$.

Im d-dimensionalen Fall $X \in \mathbb{R}^d$:
$\rho_S(X) = \rho(F_1(X_1), F_2(X_2), \ldots, F_d(X_d))$ ist die Korrelationsmatrix der eindeutigen Copula von X, wobei F_1,F_2,...,F_d die stetigen Randverteilungen von X sind.

Theorem 16 *Sei $(X_1, X_2)^T$ ein Zufallsvektor mit stetigen Randverteilungen und eindeutiger Copula C. Für die Rankkorrelationen $\rho_\tau(X_1, X_2)$ und $\rho_S(X_1, X_2)$ gilt:*

$$\rho_\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1$$

$$\rho_S(X_1, X_2) = 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) du_1 du_2 = 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3$$

Eigenschaften von ρ_τ und ρ_S.

- ρ_τ und ρ_S sind symmetrische Abhängigkeitsmaße mit Wertebereich $[-1, 1]$.

- Falls X_1, X_2 unabhängig, dann $\rho_\tau(X_1, X_2) = \rho_S(X_1, X_2) = 0$. Die Umkehrung gilt i.a. nicht.

- X_1, X_2 co-monoton dann und nur dann wenn $\rho_\tau(X_1, X_2) = \rho_S(X_1, X_2) = 1$. X_1, X_2 anti-monoton dann und nur dann wenn $\rho_\tau(X_1, X_2) = \rho_S(X_1, X_2) = -1$.

- Seien F_1, F_2 die stetigen Randverteilungen von $(X_1, X_2)^T$ und T_1, T_2 zwei streng monotone Funktionen in $[-\infty, \infty]$. Dann gilt $\rho_\tau(X_1, X_2) = \rho_\tau(T_1(X_1), T_2(X_2))$ und $\rho_S(X_1, X_2) = \rho_S(T_1(X_1), T_2(X_2))$.

(Siehe Embrechts et al., 2002).

Tail Abhängigkeit

Definition 12 *Sei* $(X_1, X_2)^T$ *ein Zufallsvektor mit Randverteilungen* F_1 *und* F_2*. Der Koeffizient der oberen Tail-Abhängigkeit von* $(X_1, X_2)^T$ *wird folgendermaßen definiert:*

$$\lambda_U(X_1, X_2) = \lim_{u \to 1^-} P(X_2 > F_2^{\leftarrow}(u) | X_1 > F_1^{\leftarrow}(u))$$

vorausgesetzt der Limes existiert.

Der Koeffizient der unteren Tail-Abhängigkeit von $(X_1, X_2)^T$ *wird folgendermaßen definiert:*

$$\lambda_L(X_1, X_2) = \lim_{u \to 0^+} P(X_2 \leq F_2^{\leftarrow}(u) | X_1 \leq F_1^{\leftarrow}(u))$$

vorausgesetzt der Limes existiert.

Wenn $\lambda_U > 0$ $(\lambda_L > 0)$ *heißt es,* $(X_1, X_2)^T$ *hat eine obere (untere) Tail-Abhängigkeit.*

Definition 13 *Sei Copula C die Verteilungsfunktion von $(U_1, U_2, \ldots, U_d)$ mit $U_i \sim U[0,1]$, $i = 1, 2, \ldots, d$. Die Verteilungsfunktion von $(1-U_1, 1-U_2, \ldots, 1-U_d)$ heißt "Survival Copula" von C und wird mit $\widehat{C}$ bezeichnet.*

Lemma 4 *Sei X ein Zufallsvektor mit multivariater Tail-Funktion $\bar{F}$ $(\bar{F}(x_1, x_2, \ldots, x_d) = Prob(X_1 > x_1, X_2 > x_2, \ldots, X_d > x_d))$ und Randverteilungsfunktionen F_i, $i = 1, 2, \ldots, d$. Sei $\bar{F}_i = 1 - F_i$, $i = 1, 2, \ldots, d$. Es gilt*

$$\bar{F}(x_1, x_2, \ldots, x_d) = \widehat{C}(\bar{F}_1(x_1), \bar{F}_2(x_2), \ldots, \bar{F}_d(x_d))$$

Lemma 5 *Für jede Copula C gilt*

$$\widehat{C}(1-u_1, 1-u_2) = 1 - u_1 - u_2 + C(u_1, u_2)\,,$$

wobei $\widehat{C}$ die Survival-Copula von C ist.

Theorem 17 *Sei $(X_1, X_2)^T$ ein Zufallsvektor mit stetigen Randverteilungsfunktionen und Copula C. Es gilt*

$$\lambda_U(X_1, X_2) = \lim_{u \to 1^-} \frac{1 - 2u + C(u,u)}{1-u} \quad \textit{und}$$

$$\lambda_L(X_1, X_2) = \lim_{u \to 0^+} \frac{C(u,u)}{u}$$

vorausgesetzt die Limes existieren.

Beispiel 13 *Die Gumbel Familie von Copulas:*

$$C_\theta^{\mathsf{Gu}}(u_1,u_2) = \exp\left(-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{1/\theta}\right),\ \ \theta \geq 1$$

Es gilt $\lambda_U = 2 - 2^{1/\theta}$, $\lambda_L = 0$.

Beispiel 14 *Die Clayton Familie von Copulas:*

$$C_\theta^{\mathsf{Cl}}(u_1,u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{1/\theta},\ \ \theta > 0$$

Es gilt $\lambda_U = 0$, $\lambda_L = 2^{-1/\theta}$.

Elliptische Copulas

Definition 14 *Sei X ein d-dimensionaler Zufallsvektor, seien $\mu \in \mathbb{R}^d$ und $\Sigma \in \mathbb{R}^{d\times d}$ zwei Konstanten, und sei $\psi\colon [0,\infty) \to \mathbb{R}$ eine Funktion. Wenn $\phi_{X-\mu} = \psi(t^T\Sigma t)$ gilt, wobei $\phi_{X-\mu}$ die charakteristische Funktion von $X-\mu$ ist, dann ist X eine elliptisch verteilter Zufallsvektor mit Parameter μ, Σ, ψ: $X \sim E_d(\mu,\Sigma,\psi)$.*

ψ heißt erzeugende Funktion (oder Generator) von X.

Für $d = 1$ stimmen die elliptischen Verteilungen mit den symmetrischen Verteilungen überein.
Überzeugen Sie sich! Verwenden Sie die stochastische Darstellung einer elliptischen Verteilung.

Theorem 18 *(Stochastische Darstellung)*
Ein d-dimensionaler Zufallsvektor X ist elliptisch verteilt, $X \sim E_d(\mu,\Sigma,\psi)$ und $rang(\Sigma) = k$, dann und nur dann wenn es eine Matrix $A \in \mathbb{R}^{d\times k}$, $A^TA = \Sigma$, sowie eine nicht negative Zufallsvariable R und einen k-dimensionalen auf der Einheitskugel $\mathcal{S}^{k-1} = \{z \in \mathbb{R}^k\colon z^z z = 1\}$ gleichverteilten Zufallsvektor U gibt, sodass R und U unabhängig sind und $X \stackrel{d}{=} \mu + RAU$.

Anmerkung: Eine elliptische Verteilung X ist *radial symmetrisch*: $X - \mu \stackrel{d}{=} \mu - X$.

Definition 15 *Sei* $X \sim E_d(\mu, \Sigma, \psi)$ *mit Verteilungsfunktion* F *und stetigen Randverteilungen* $F_1, F_2, \ldots, F_d$*. Dann wird die eindeutige Copula* C *von* F*,* $C(u) = F(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d))$*, elliptische Copula genannt.*

Beispiel 15 *Gauss'sche Copulas sind elliptische Copulas*
Sei C_R^{Ga} *die Copula einer* d*-dimensionalen Normalverteilung mit Korrelationsmatrix* R*:*

$$C_R^{Ga}(u) = \phi_R^d(\phi^{-1}(u_1), \ldots, \phi^{-1}(u_d)),$$

wobei ϕ_R^d *die Gesamtverteilungsfunktion einer* d*-dimensionalen Normalverteilung mit Erwartungsvektor* 0 *und Korrelationsmatrix* R *und* ϕ^{-1} *die Inverse der Verteilungsfunktion einer univariaten standard Normalverteilung ist. Da die Normalverteilung eine elliptische Verteilung ist, ist dieGauss'sche Copula* C_R^{Ga} *eine elliptische Copula.*

Im bivariaten Fall gilt:

$$C_R^{Ga}(u_1, u_2) = \int_{-\infty}^{\phi^{-1}(u_1)} \int_{-\infty}^{\phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{\frac{-(x_1^2 - 2\rho x_1 x_2 + x_2^2)}{2(1-\rho^2)}\right\} dx_1 dx_2,$$

wobei $\rho \in (-1, 1)$*.*

t-Copula: ein weiteres Beispiel elliptischer Copulas

Definition 16 *Sei* $X \stackrel{d}{=} \mu + \frac{\sqrt{\alpha}}{\sqrt{S}} AZ \sim t_d(\alpha, \mu, \Sigma)$, *wobei* $\mu \in \mathbb{R}^d$, $\alpha \in \mathbb{N}$, $\alpha > 1$, $S \sim \chi^2_\alpha$, $A \in \mathbb{R}^{d\times k}$ *mit* $AA^t = \Sigma$ *und* $Z \sim N_k(0, I_k)$, *und* S *und* Z *unabhängig sind. Es heißt,* X *hat eine* d*-dimensionale* t*-Verteilung mit Mittelwert* μ *(für* $\alpha > 1$*) und Kovarianzmatrix* $Cov(X) = \frac{\alpha}{\alpha-2}\Sigma$ *(für* $\alpha > 2$*).* $Cov(X)$ *existiert nicht für* $\alpha \leq 2$.

Definition 17 *Die Copula* $C^t_{\alpha,R}$ *von* X *heißt* t*-Copula. Für die* t*-Copula gilt:*

$$C^t_{\alpha,R}(u) = t^d_{\alpha,R}(t^{-1}_\alpha(u_1), \ldots, t^{-1}_\alpha(u_d)).$$

$R_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$, $i, j = 1, 2 \ldots, d$, *ist die Korrelationsmatrix von* Z,

$t^d_{\alpha,R}$ *ist die Verteilungsfunktion von* $\frac{\sqrt{\alpha}}{\sqrt{S}} RZ$, *wobei* $S \sim \chi^2_\alpha$ *und* $Z \sim N_k(0, I_k)$ *unabhängig sind, und* t_α *sind die Randverteilungen von* $t^d_{\alpha,R}$.

Bivariater Fall ($d = 2$):

$$C^t_{\alpha,R}(u_1, u_2) = \int_{-\infty}^{t^{-1}_\alpha(u_1)} \int_{-\infty}^{t^{-1}_\alpha(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \left\{1 + \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{\alpha(1-\rho^2)}\right\}^{-(\alpha+2)/2} dx_1 dx_2,$$

für $\rho \in (-1, 1)$. R_{12} ist der lineare Korrelationskoeffizient der dazugehörigen bivariaten t_α-Verteilung für $\alpha > 2$.

Copulas: Weitere Eigenschaften

Definition 18 *(Radiale Symmetrie oder Kugel-Symmetrie)*
Ein Zufallsvektor X (oder eine Verteilungsfunktion) heißt radial symmetrisch (oder kugel-symmetrisch) um den Punkt a wenn $X - a \stackrel{d}{=} a - X$.

Beispiel: Ein elliptisch-verteilter Zufallsvektor $X \sim E_d(\mu, \Sigma, \psi) \in \mathbb{R}^d$ ist radial-symmetrisch um μ.

Definition 19 *(Radiale Symmetrie von Copulas)*
Eine Copula C heißt radial-symmetrisch wenn

$$(U_1 - 0.5, \ldots, U_d - 0.5) \stackrel{d}{=} (0.5 - U_1, \ldots, 0.5 - U_d) \Longleftrightarrow U \stackrel{d}{=} 1 - U,$$

wobei $(U_1, U_2, \ldots, U_d)$ ein Zufallsvektor mit Verteilungsfunktion C ist.

Für eine radial symmetrische Copula gilt $C = \widehat{C}$.

Beispiel: Elliptische Copulas sind radial symmetrisch.
Die Gumbel und Clayton Copulas sind es nicht. Überzeugen Sie sich!

Die Dichtefunktion einer Copula

Copulas haben nicht immer eine Dichtefunktion. Z.B. die Co-Monotonie Copula M bzw. die Anti-Monotonie Copula W haben keine Dichtefunktion.

Wenn die Dichtefunktion c einer Copula C existiert dann gilt

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

Sei C die Copula einer Gesamtverteilung F mit stetigen Randverteilungsfunktionen F_1,...,F_d. Dann kann die Gleichung

$$C(u_1, \ldots, u_d) = F(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d))$$

differenziert werden um die Dichte c von C zu erhalten:

$$c(u_1, \ldots, u_d) = \frac{f(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \ldots f_d(F_d^{-1}(u_d))}$$

f ist die Gesamtdichtefunktion, f_i sind die Dichtefunktionen der Randverteilungen, $1 \leq i \leq d$, und F_i^{-1} ist die inverse Funktion von F_i.

Definition 20 *Ein Zufallsvektor X heißt vertauschbar (“exchange-able”) wenn $(X_1, \ldots, X_d) \stackrel{d}{=} (X_{\pi(1)}, \ldots, X_{\pi(d)})$ für jede Permutation $(\pi(1), \pi(2), \ldots, \pi(d))$ von $(1, 2, \ldots, d)$.*

Definition 21 *Eine Copula C heißt vertauschbar wenn sie die Gesamtverteilung eines vertauschbaren Zufallsvektors (mit Gleichverteilungen als Randverteilungen) ist.*

Für eine solche Copula gilt:

$$C(u_1, u_2, \ldots, u_d) = C(u_{\pi(1)}, u_{\pi(2)}, \ldots, u_{\pi(d)})$$

für jede Permutation $(\pi(1), \pi(2), \ldots, \pi(d))$ von $(1, 2, \ldots, d)$.

Beispiele von vertauschbaren Copulas: Gumbel, Clayton, Gauss’sche Copula C_P^{Ga}, t-Copula $C_{\nu,P}^t$ für den Fall, dass P eine Equikorrelations-matrix ist: $R = \rho J_d + (1-\rho) I_d$. $J_d \in \mathbb{R}^{d \times d}$ ist eine Matrix bestehend aus lauter Einser, und $I_d \in \mathbb{R}^{d \times d}$ ist die d-dimensionale Einheitsmatrix.

Für bivariate vertauschbare Copulas gilt:

$$P(U_2 \le u_2 | U_1 = u_1) = P(U_1 \le u_2 | U_2 = u_1).$$

Theorem 19 *Sei* $(X_1, X_2)^T$ *ein normalverteilter Zufallsvektor. Dann gilt:* $\lambda_U(X_1, X_2) = \lambda_L(X_1, X_2) = 0$.

Korollar 2 *Sei* $(X_1, X_2)^T$ *ein Zufallsvektor mit stetigen Randverteilungen und einer Gauss'schen Copula* C_ρ^{Ga}, *wobei* ρ
der Koeffizient der linearen Korrelation zwischen X_1 *und* X_2 *ist. Dann gilt:* $\lambda_U(X_1, X_2) = \lambda_L(X_1, X_2) = 0$.

Theorem 20 *Sei* $(X_1, X_2)^T$ *ein t-verteilter Zufallsvektor mit* ν *Freiheitsgraden, Mittelwert* 0 *und linearer Korrelationsmatrix* R*:* $(X_1, X_2)^T \sim t_2(0, \nu, R)$. *Für* $R_{12} > -1$ *gilt:*

$$\lambda_U(X_1, X_2) = \lambda_L(X_1, X_2) = 2\bar{t}_{\nu+1}\left(\sqrt{\nu+1}\frac{\sqrt{1-R_{12}}}{\sqrt{1+R_{12}}}\right)$$

Beweis: Ähnlich wie der Beweis von Satz 19. Hinweis:

$$X_2|X_1 = x \sim \left(\frac{\nu+1}{\nu+x^2}\right)^{1/2} \frac{X_2 - \rho x}{\sqrt{1-\rho^2}} \sim t_{\nu+1}$$

.