

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: $\rho_\tau(X_1, X_2) = E(\text{sign}[(X_1 - X'_1)(X_2 - X'_2)])$.

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

Der Kendall's Tau der Stichprobe:

Sei $\{(x_1, y_1)^T, (x_2, y_2)^T, \dots, (x_n, y_n)^T\}$ 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.

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

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 $X \in \mathbb{R}^d$:

$\rho_S(X) = \rho(F_1(X_1), F_2(X_2), \dots, F_d(X_d))$ ist die Korrelationsmatrix der eindeutigen Copula von X , wobei F_1, F_2, \dots, 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:

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

vorausgesetzt der Limes existiert.

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

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

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, \dots, U_d) mit $U_i \sim U[0, 1]$, $i = 1, 2, \dots, d$. Die Verteilungsfunktion von $(1 - U_1, 1 - U_2, \dots, 1 - U_d)$ heißt "Survival Copula" von C und wird mit \hat{C} bezeichnet.

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

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

Lemma 5 Für jede Copula C gilt

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

wobei \hat{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

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

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

vorausgesetzt die Limes existieren.

Beispiel 13 *Die Gumbel Familie von Copulas:*

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

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

Beispiel 14 *Die Clayton Familie von Copulas:*

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

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 $\psi: [0, \infty) \rightarrow \mathbb{R}$ 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 $\mu, \Sigma, \psi: X \sim E_d(\mu, \Sigma, \psi)$.

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 $\text{rang}(\Sigma) = k$, dann und nur dann wenn es eine Matrix $A \in \mathbb{R}^{d \times k}$, $A^T A = \Sigma$, sowie eine nicht negative Zufallsvariable R und einen k -dimensionalen auf der Einheitskugel $S^{k-1} = \{z \in \mathbb{R}^k: z^T z = 1\}$ gleichverteilten Zufallsvektor U gibt, sodass R und U unabhängig sind und $X \stackrel{d}{=} \mu + RAU$.

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