Lineare Korrelation

Definition 9 *Sei* $(X_1, X_2)^T$ *ein Zufallsvektor mit* $var(X_1), var(X_2) \in (0, \infty)$. *Der lineare Korrelationskoeffizient für* $(X_1, X_2)^T$ *ist*

$$\rho_L(X_1, X_2) = \frac{cov(X_1, X_2)}{\sqrt{var(X_1)var(X_2)}}, \text{ wobei } cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2).$$

Eigenschaften der linearen Korrelationen bzw. Kovarianzen:

- $\rho_L(X1, X_2) \in [-1, 1]$.
- $|\rho_L(X_1, X_2)| = 1 \Longleftrightarrow \exists a \in \mathbb{R} \setminus \{0\}$ und $\exists b \in \mathbb{R}$, sodass $X_2 = aX_1 + b$ fast sicher und $sgn(\rho_L(X_1, X_2)) = sgn(a)$.
- Sei $\alpha, \gamma \in \mathbb{R} \setminus \{0\}$ und $\beta, \delta \in \mathbb{R}$. Dann gilt: $\rho_L(\alpha X_1 + \beta, \gamma X_2 + \delta) = sign(\alpha\gamma)\rho_L(X_1, X_2)$. D.h. die lineare Korrelation ist eine invariante bzgl. streng monoton steigende lineare Transformationen.
- Seien $A, B \in \mathbb{R}^{p \times d}$, $a, b \in \mathbb{R}^p$ und seien X, Y zwei d-dimensionale Zufallsvektoren. Dann gilt: $cov(AX + a, BY + b) = Acov(X, Y)B^T$.
- Sei $\alpha \in \mathbb{R}^d$ und X ein d-dimensionaler Zufallsvektor. Es gilt: $var(\alpha^T X) = \alpha^T cov(X, X)\alpha$.

Theorem 13 *Sei $(X_1, X_2)^T$ ein Zufallsvektor mit Randverteilungsfunktionen F_1, F_2 und einer nicht spezifizierten Abhängigkeitsstruktur. Angenommen $var(X_1), var(X_2) \in (0, \infty)$. Dann gilt:*

1. *Die Menge der möglichen linearen Korrelationen von X_1 und X_2 ist ein Interval $[\rho_{L,min}; \rho_{L,max}]$ mit $0 \in (\rho_{L,min}; \rho_{L,max})$.*

2. *Die minimale lineare Korrelation wird dann und nur dann erreicht wenn X_1 und X_2 anti-monoton sind. Die maximale lineare Korrelation wird dann und nur dann erreicht wenn X_1 und X_2 co-monoton sind.*

Beispiel 11 *Sei $X_1 \sim Lognormal(0, 1)$ und $X_2 \sim Lognormal(0, \sigma^2)$, $\sigma > 0$. Bestimmen Sie $\rho_{L,min}(X_1, X_2)$ und $\rho_{L,max}(X_1, X_2)$.*

Beispiel 12 *Betrachten Sie zwei ZV, die die Verluste zweier Portfolii darstellen. Annahme: beide Zufallsvariablen sind standard normal verteilt und ihre lineare Korrelation ist gleich 0. Zeigen Sie, dass es zwei Zufallsvektoren $(X_1, X_2)^T$ und $(Y_1, Y_2)^T$ gibt, sodass*
(a) die jeweiligen Komponentenpaare die obigen Annahmen erfüllen, und
(b) die zwei Zufallsvektoren haben unterschiedliche Gesamtverteilungsfunktionen.

Bestimmen Sie die Quantile $F^{\leftarrow}_{X_1+X_2}(\alpha)$ und $F^{\leftarrow}_{Y_1+Y_2}(\alpha)$ der Gesamtverluste und vergleichen Sie diese miteinander.

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. 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.

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

Definition 11 *Sei $(X_1, X_2)^T$ ein Zufallsvektor. 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.

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

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

Theorem 14 *Sei $(X_1, X_2)^T$ ein Zufallsvektor mit stetigen Randverteilungen und eindeutige 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* $(X_1, X2)^T$ *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" colupa 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 *Es gilt* $\widehat{C}(1 - u_1, 1 - u_2) = 1 - u_1 - u_2 + C(u_1, u_2)$*.*

Theorem 15 *Sei* $(X_1, X_2)^T$ *ein Zufallsvektor mit stetigen Randverteilungsfunktionen und Copula* C*. Es gilt*

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

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

vorausgesetzt die Limes existieren.

Beispiel 13 *Die Gumbel Copulas:*

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

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

Beispiel 14 *Die Clayton Copulas:*

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

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 charakteristische erzeugende Funktion (oder Generator) von X. Für $d = 1$ stimmen die elliptischen Verteilungen mit den symmetrischen Verteilungen überein.

Theorem 16 *(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 $A \in \mathbb{R}^{d \times k}$, $A^T A = \Sigma$, 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 $F_1, F_2, \ldots, F_d$. Die eindeutige Copula C von F, $C(u) = F(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d))$, heißt elliptische Copula.*

Beispiel 15 *Gauss'sche Copulas*
Sei C_R^{Ga} die Copula einer d-dimensionalen standard 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 standard Normalverteilung ist und ϕ^{-1} die Inverse der Verteilungsfunktion einer univariaten standard Normalverteilung ist. C_R^{Ga} heißt Gauss'sche 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* $X \stackrel{d}{=} \mu + \frac{\sqrt{\nu}}{\sqrt{S}} Z$, *wobei* $\mu \in \mathbb{R}^d$, $\nu \in \mathbb{N}$, $\nu > 1$, $S \sim \chi^2_\nu$ *und* $Z \sim N_d(0, \Sigma)$, *und* S *und* Z *unabhängig sind. Es heißt,* X *hat eine* d*-dimensionaler* t*-Verteilung mit Mittelwert* μ *(für* $\nu > 1$*) und Kovarianzmatrix* $Cov(X) = \frac{\nu}{\nu-2}\Sigma$ *(für* $\nu > 2$*).* $Cov(X)$ *existiert nicht für* $\nu \leq 2$*.*

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

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

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

$t^d_{\nu,R}$ *ist die Verteilungsfunktion von* $\frac{\sqrt{\nu}}{\sqrt{S}} Y$, *wobei* $S \sim \chi^2_\nu$ *und* $Y \sim N_d(0, R)$ *unabhängig sind, und* t_ν *sind die Randverteilungen von* $t^d_{\nu,R}$.

Bivariater Fall ($d = 2$):

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

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