Modelling and Estimation of Stochastic Dependence

Uwe Schmock

Based on joint work with Dr. Barbara Dengler

Financial and Actuarial Mathematics and
Christian Doppler Laboratory for Portfolio Risk Management
Vienna University of Technology, Austria
www.fam.tuwien.ac.at

101. Annual Meeting of the Swiss Association of Actuaries
September 10./11., 2010, St. Gallen

Research areas for stochastic dependence (1)

- Modelling and estimation of dependent credit rating transitions
  $\rightarrow$ Ph. D. thesis of Verena Goldammer (2010)
- Market and credit risk aggregation: a bottom-up approach
  $\rightarrow$ Ph. D. thesis of Robert Schöftner (2010)
- Adapted dependence
  $\rightarrow$ Ph. D. project of Karin Hirhager
  - Relaxing the independence of biometric and financial market risks when estimating the risk of unit-linked life insurance contracts
  - Modelling consumer behaviour dependent on financial market development (related to American option)
    $\rightarrow$ Variable annuities

Research areas for stochastic dependence (2)

- Generalization of Panjer's recursion for dependent claim numbers (collective model, CreditRisk^{+})
  $\rightarrow$ Ph. D. project of Cordelia Rudolph
- Joint term-structure models for credit spreads and risk-free interest rates
  $\rightarrow$ Ph. D. project of Sühan Altay
  We aim for
  - Non-negative interest rates and credit spreads,
  - Negative covariation between them,
  - Zero-coupon bond prices easy to calculate.
- Asymptotic variance of estimators of dependence (linear correlation, Kendall's tau)
  $\rightarrow$ mainly the Ph. D. thesis of Barbara Dengler (2010)

Outline

Linear correlation coefficient

Definition

The linear correlation coefficient for a random vector (X, Y) with non-zero finite variances is defined as

$$\varrho = \frac{\mathbb{C}\mathrm{ov}[X, Y]}{\sqrt{\mathbb{V}\mathrm{ar}[X]}\sqrt{\mathbb{V}\mathrm{ar}[Y]}}.$$

Estimator

The standard estimator for a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ is

$$\hat{\varrho}_n = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2}\sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}}$$

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$.

Asymptotic behaviour of the standard estimator

Theorem (Asymptotic normality, e.g. Witting/Müller-Funk '95, p. 108)

For an i. i. d. sequence of non-degenerate real-valued random variables (X_j, Y_j), $j \in \mathbb{N}$, with $\mathbb{E}[X^4] < \infty$ and $\mathbb{E}[Y^4] < \infty$, the standard estimators $\hat{\varrho}_n$, normalized with $\sqrt{n}$, are asymptotically normal,

$$\sqrt{n}\,(\hat{\varrho}_n - \varrho) \xrightarrow{\text{d}} \mathcal{N}(0, \sigma_\varrho^2), \quad n \to \infty .$$

The asymptotic variance is

$$\sigma_\varrho^2 = \Bigl(1 + \frac{\varrho^2}{2}\Bigr)\frac{\sigma_{22}}{\sigma_{20}\,\sigma_{02}} + \frac{\varrho^2}{4}\Bigl(\frac{\sigma_{40}}{\sigma_{20}^2} + \frac{\sigma_{04}}{\sigma_{02}^2} - \frac{4\,\sigma_{31}}{\sigma_{11}\,\sigma_{20}} - \frac{4\,\sigma_{13}}{\sigma_{11}\,\sigma_{02}}\Bigr),$$

where $\sigma_{kl} := \mathbb{E}[(X - \mu_X)^k (Y - \mu_Y)^l]$, $\mu_X := \mathbb{E}[X]$, $\mu_Y := \mathbb{E}[Y]$.

Kendall's tau

Definition

Kendall's tau for a random vector (X, Y) is defined as

$$\tau = \mathbb{P}[\underbrace{(X - \widetilde{X})(Y - \widetilde{Y}) > 0}_{\text{concordance}}] - \mathbb{P}[\underbrace{(X - \widetilde{X})(Y - \widetilde{Y}) < 0}_{\text{discordance}}]$$
$$= \mathbb{E}[\,\mathrm{sgn}(X - \widetilde{X})\,\mathrm{sgn}(Y - \widetilde{Y})\,],$$

where $(\widetilde{X}, \widetilde{Y})$ is an independent copy of (X, Y).

Estimator (Representation as U-statistic)

The tau-estimator for a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ is

$$\hat{\tau}_n = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \mathrm{sgn}(X_i - X_j)\,\mathrm{sgn}(Y_i - Y_j)\,.$$

U-statistics

Definition

Fix $m \in \mathbb{N}$. For $n \geq m$ let $Z_1, \ldots, Z_n$ be random variables taking values in the measurable space $(\mathcal{Z}, \mathfrak{Z})$ and let $\kappa : \mathcal{Z}^m \to \mathbb{R}$ be a symmetric measurable function. The U-statistic $\hat{U}_n(\kappa)$ belonging to the kernel κ of degree m is defined as

$$\hat{U}_n(\kappa) := \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \cdots < i_m \leq n} \kappa(Z_{i_1}, \ldots, Z_{i_m}) \,.$$

The tau-estimator is a U-statistic with kernel κ_τ of degree 2:

$$\kappa_\tau : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R} \,,$$
$$\kappa_\tau\big((x, y), (x', y')\big) = \operatorname{sgn}(x - x') \operatorname{sgn}(y - y') \,.$$

Properties of the tau-estimator

If the observations are i. i. d., then $\hat{\tau}_n$ is an unbiased estimate of τ.

Theorem (Asymptotic normality, e.g. Borovskikh '96)

For an i. i. d. sequence of $\mathbb{R}^2$-valued random vectors, the tau-estimators $\hat{\tau}_n$, normalized with $\sqrt{n}$, are asymptotically normal,

$$\sqrt{n}\,(\hat{\tau}_n - \tau) \xrightarrow{\mathrm{d}} \mathcal{N}\big(0, \sigma_\tau^2\big), \quad n \to \infty\,.$$

The asymptotic variance is

$$\sigma_\tau^2 = 4\,\mathbb{V}\mathrm{ar}\big[\mathbb{E}[\operatorname{sgn}(X - \widetilde{X})\operatorname{sgn}(Y - \widetilde{Y})\,|\,X, Y\,]\big]\,,$$

where $(\widetilde{X}, \widetilde{Y})$ is an independent copy of (X, Y).

Applications of asymptotic variance

- Asymptotic normality leads to asymptotic confidence intervals of the form

$$\left[\hat{\tau}_n - \frac{\sigma_\tau}{\sqrt{n}}\, u_{\frac{1+\alpha}{2}},\ \hat{\tau}_n + \frac{\sigma_\tau}{\sqrt{n}}\, u_{\frac{1+\alpha}{2}}\right]$$

  for given confidence level $\alpha \in (0, 1)$, where $u_{\frac{1+\alpha}{2}}$ is the corresponding quantile of the standard normal distribution.

- This allows in particular to test for dependence.
- Estimators can be evaluated by their asymptotic variance and different ways of estimation can be compared, e.g. for elliptical distributions.

Definition of a copula and Sklar's theorem

Definition

A two-dimensional copula C is a distribution function on $[0,1]^2$ with uniform marginal distributions.

Let (X, Y) be an $\mathbb{R}^2$-valued random vector with marginal distribution functions F and G. Then, by Sklar's theorem, there exists a copula C such that

$$\mathbb{P}[X \le x, Y \le y] = C\big(F(x), G(y)\big), \quad x, y \in \mathbb{R}.$$

If the marginal distribution functions F and G are continuous, then Sklar's theorem also gives uniqueness of the copula C.

Kendall's tau and asymptotic variance for copulas

Assume that X and Y have continuous distribution functions. Then

$$U := F(X) \quad \text{and} \quad V := G(Y)$$

are uniformly distributed on $[0, 1]$ and Kendall's tau becomes

$$\tau = 4\,\mathbb{E}[\,C(U, V)] - 1\,.$$

Theorem (Dengler/Schmock)

The asymptotic variance for the tau-estimators is

$$\sigma_\tau^2 = 16\,\mathbb{V}\mathrm{ar}[2\,C(U, V) - U - V]\,.$$

Note: Both quantities depend only on the copula C.

Examples of copulas for calculating the asymptotic variance for the tau-estimators

- Archimedean copulas
  - Product (independence) copula
  - Clayton copula
  - Ali–Mikhail–Haq copula
- Non-Archimedean copulas
  - Farlie–Gumbel–Morgenstern copula
  - Marshall–Olkin copula

Archimedean copulas

- An Archimedean copula is defined by a generator, i.e., by a continuous, strictly decreasing and convex function $\varphi : [0,1] \to [0,\infty]$ with $\varphi(1) = 0$.
- The pseudo-inverse $\varphi^{[-1]}$ of φ is given by

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) & \text{for } t \in [0, \varphi(0)]\,, \\ 0 & \text{for } t \in (\varphi(0), \infty]\,. \end{cases}$$

- The copula is defined as

$$C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))\,, \quad u, v \in [0,1]\,.$$

- If $\varphi(0) = \infty$, then the generator φ and its copula C are called strict.

Product copula

$$C^{\perp} : [0,1]^2 \to [0,1]$$
$$C^{\perp}(u,v) = u\,v$$

- Copula for two independent random variables, $\tau^{\perp} = 0$.
- The product copula is a strict Archimedean copula with generator $\varphi(t) = -\log t$ for $t \in [0,1]$.
- Asymptotic variance of the tau-estimator:

$$(\sigma_{\tau}^{\perp})^2 = \frac{4}{9}$$

Clayton copula with parameter $\theta \in (0, \infty)$

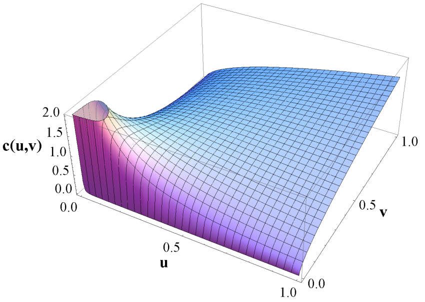
$$C^{\mathrm{Cl},\theta}(u,v) = \begin{cases} \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta} & \text{for } u, v \in (0,1]\,, \\ 0 & \text{otherwise} \end{cases}$$

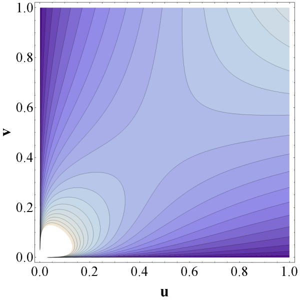
- The Clayton copula is a strict Archimedean copula with generator $\varphi(t) = \frac{1}{\theta}\left(t^{-\theta} - 1\right)$ for $t \in [0,1]$.
- Kendall's tau is $\tau^{\mathrm{Cl},\theta} = \frac{\theta}{\theta+2} \in (0,1)$.
- Asymptotic variance of the tau-estimator for $\theta \in \{1,2\}$:

$$\left(\sigma_\tau^{\mathrm{Cl},1}\right)^2 = \frac{16}{9}\left(6\pi^2 - 59\right) \approx 0.387$$

$$\left(\sigma_\tau^{\mathrm{Cl},2}\right)^2 = \frac{337}{15} - 32\log(2) \approx 0.286$$

Clayton copula, density and results





$$\tau = \frac{2}{9}, \quad \theta = \frac{2\tau}{1-\tau} = \frac{4}{7}, \quad \left(\sigma_\tau^{\mathrm{Cl},\theta}\right)^2 \approx 0.430$$

Note: An estimate for τ gives an estimate for the parameter θ.

Ali–Mikhail–Haq copula with parameter $\theta \in [-1, 1)$

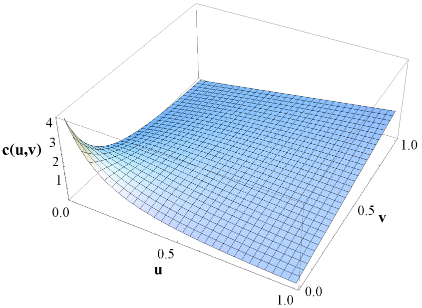
$$C^{\mathrm{AMH},\theta}(u,v) = \frac{u\,v}{1-\theta\,(1-u)\,(1-v)}\,, \quad u,v \in [0,1]$$

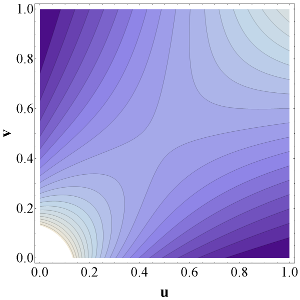
- The AMH copula is a strict Archimedean copula with generator $\varphi(t) = \log\bigl(\frac{1-\theta\,(1-t)}{t}\bigr)$ for $t \in [0,1]$.
- Product copula corresponds to $\theta = 0$.
- Results for $\theta \neq 0$ (with Li_2 denoting the dilogarithm):

$$\tau^{\mathrm{AMH},\theta} = \frac{3\theta-2}{3\theta} - 2\frac{(1-\theta)^2}{3\theta^2}\log(1-\theta)$$

$$\bigl(\sigma_\tau^{\mathrm{AMH},\theta}\bigr)^2 = -\frac{100}{9} - 8\frac{4-(\theta^2+9\theta+2)\,\tau^{\mathrm{AMH},\theta}}{\theta(1-\theta)} + 4\bigl(\tau^{\mathrm{AMH},\theta}\bigr)^2 + 32\frac{\theta+1}{\theta^2}\,\mathrm{Li}_2(\theta)$$

Ali–Mikhail–Haq copula, density and results





$$\tau = \frac{2}{9}, \quad \theta \approx 0.77152, \quad \left(\sigma_\tau^{\mathrm{AMH},\theta}\right)^2 \approx 0.399$$

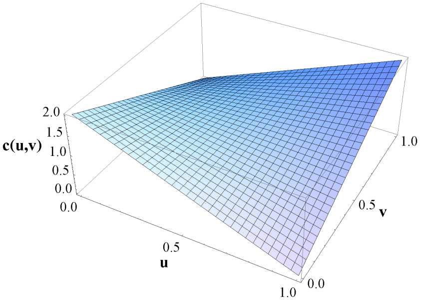
Farlie–Gumbel–Morgenstern copula with $\theta \in [-1, 1]$

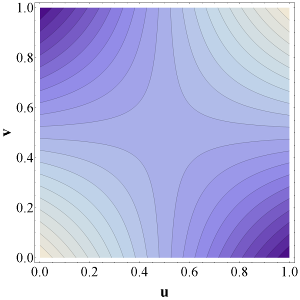
$$C^{\mathrm{FGM},\theta}(u, v) = u\,v + \theta\,u\,v\,(1 - u)\,(1 - v)\,, \quad u, v \in [0, 1]$$

- Kendall's tau is $\tau^{\mathrm{FGM},\theta} = \frac{2\theta}{9} \in [-\frac{2}{9}, \frac{2}{9}]$.
- Asymptotic variance of the tau-estimator:

$$\left(\sigma_\tau^{\mathrm{FGM},\theta}\right)^2 = \frac{4}{9} - \frac{46}{25}\left(\tau^{\mathrm{FGM},\theta}\right)^2$$

Farlie–Gumbel–Morgenstern copula, density and results





$$\tau = \frac{2}{9}, \quad \theta = \frac{9}{2}\tau = 1, \quad \left(\sigma_\tau^{\mathrm{FGM},\theta}\right)^2 = \frac{716}{2025} \approx 0.354$$

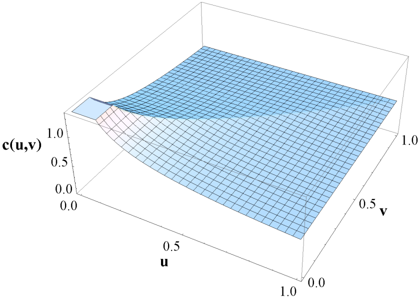
Marshall–Olkin copula with parameters $\alpha, \beta \in (0,1)$

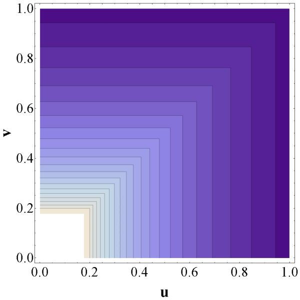
$$C^{\mathrm{MO}}_{\alpha,\beta}(u,v) = \min\{u^{1-\alpha}\, v, u\, v^{1-\beta}\}\,, \quad u,v \in [0,1]$$

- Kendall's tau is $\tau^{\mathrm{MO}}_{\alpha,\beta} = \frac{\alpha\beta}{\alpha+\beta-\alpha\beta} \in (0,1)$.
- Asymptotic variance of the tau-estimator:

$$\left(\sigma_\tau^{\mathrm{MO},\alpha,\beta}\right)^2 = \frac{64\,(\alpha+\beta+\alpha\beta)}{9\,(\alpha+\beta-\alpha\beta)} - \frac{32\,(2\alpha+3\beta+\alpha\beta)}{3\,(2\alpha+3\beta-2\alpha\beta)} - \frac{32\,(3\alpha+2\beta+\alpha\beta)}{3\,(3\alpha+2\beta-2\alpha\beta)} + \frac{16\,(\alpha+\beta)}{(2\alpha+2\beta-\alpha\beta)} + \frac{8\,\alpha\beta}{\alpha+\beta-\alpha\beta} - \frac{4\,\alpha^2\beta^2}{(\alpha+\beta-\alpha\beta)^2} + \frac{20}{3}$$

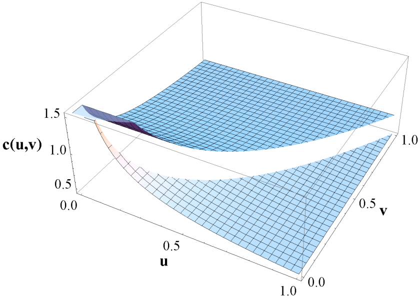
Marshall–Olkin copula, density and results (1)

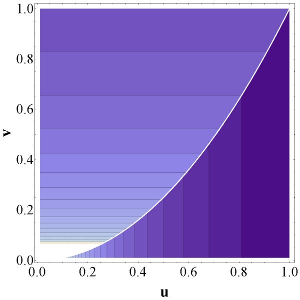




$$\tau = \frac{2}{9}, \quad \alpha = \beta = \frac{4}{11}, \quad \left(\sigma_\tau^{\mathrm{MO},\alpha,\beta}\right)^2 \approx 0.538$$

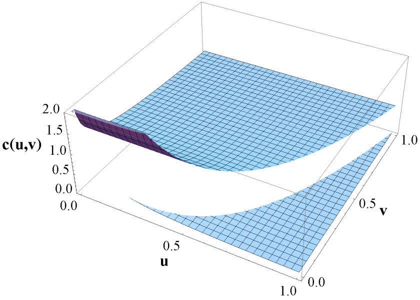
Marshall–Olkin copula, density and results (2)

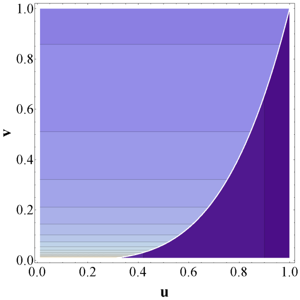




$$\tau = \frac{2}{9}, \quad \alpha = \frac{6}{11}, \quad \beta = \frac{\alpha}{2} = \frac{3}{11}, \quad \left(\sigma_\tau^{\text{MO},\alpha,\beta}\right)^2 \approx 0.505$$

Marshall–Olkin copula, density and results (3)





$$\tau = \frac{2}{9}, \quad \alpha = \frac{10}{11}, \quad \beta = \frac{\alpha}{4} = \frac{5}{22}, \quad \left(\sigma_\tau^{\mathrm{MO},\alpha,\beta}\right)^2 \approx 0.429$$

Spherical distributions

Definition

$X = (X_1, \dots, X_d)^\top$ is spherically distributed if it has the stochastic representation

$$X \overset{\mathrm{d}}{=} RS\,,$$

where

1. S is uniformly distributed on the $(d-1)$-dimensional unit sphere $\mathcal{S}^{d-1} = \{s \in \mathbb{R}^d : s^\top s = 1\}$, and
2. $R \geq 0$ is a radial random variable, independent of S.

Note: A spherical distribution is invariant under orthogonal transformations.

Elliptical distributions

Definition

$X = (X_1, \dots, X_d)^\top$ is elliptically distributed with location vector μ and dispersion matrix Σ, if there exist $k \in \mathbb{N}$, a matrix $A \in \mathbb{R}^{d \times k}$ with $AA^\top = \Sigma$, and random variables R, S satisfying

$$X \stackrel{\mathrm{d}}{=} \mu + RAS\,,$$

where

1. S is uniformly distributed on the unit sphere $\mathcal{S}^{k-1} = \{s \in \mathbb{R}^k : s^\top s = 1\}$, and
2. $R \geq 0$ is a radial random variable, independent of S.

Note: An elliptical distribution is an affine transformation of a spherical distribution.

Linear correlation and standard estimator for non-degenerate elliptical distributions

The (generalized) linear correlation coefficient is defined by

$$\varrho = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\,\Sigma_{22}}}\,.$$

Theorem (Dengler/Schmock)

For elliptical distributions the asymptotic variance of the standard estimator simplifies to

$$\sigma_\varrho^2 = \frac{\mathbb{E}[R^4]}{2\,\mathbb{E}[R^2]^2}\,\left(\varrho^2 - 1\right)^2,$$

provided the radial variable R satisfies $0 < \mathbb{E}[R^4] < \infty$.

Connection between the linear correlation coefficient and Kendall's tau for elliptical distributions

Theorem (Lindskog/McNeil/Schmock, 2003)

Let $(X,Y)^\top$ be elliptically distributed with non-degenerate components. Define

$$a_X = \sum_{x\in\mathbb{R}} \bigl(\mathbb{P}[X = x]\bigr)^2,$$

where the sum extends over all atoms of the distribution of X. Then

$$\tau = \frac{2(1-a_X)}{\pi} \arcsin \varrho\,.$$

Transformation of Kendall's tau into an alternative linear correlation estimator

- Define the transformed tau-estimator by

$$\hat{\varrho}_{\tau,n} = \sin\left(\frac{\pi}{2(1-a_X)}\,\hat{\tau}_n\right).$$

If the random variables are non-degenerate, then $\hat{\varrho}_{\tau,n}$ is an estimator for the (generalized) linear correlation ϱ.

- The asymptotic distribution remains normal,

$$\sqrt{n}\,(\hat{\varrho}_{\tau,n} - \varrho) \xrightarrow{\text{d}} \mathcal{N}\big(0, \sigma^2_{\varrho(\tau)}\big), \quad n \to \infty,$$

with

$$\sigma^2_{\varrho(\tau)} = \frac{\pi^2}{4(1-a_X)^2}\,\sigma^2_\tau\,(1-\varrho^2).$$

(e.g. Lehmann/Casella '98, p. 58)

Asymptotic variance for spherical distributions

- Formula for the asymptotic variance of the tau-estimator:

$$\sigma_\tau^2 = 4\,\mathbb{V}\mathrm{ar}\big[\mathbb{E}[\operatorname{sgn}(X-\widetilde{X})\operatorname{sgn}(Y-\widetilde{Y})\,|\,X,Y]\big]\,,$$

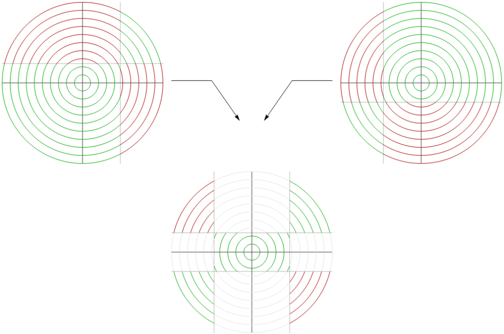
where $(\widetilde{X},\widetilde{Y})$ is an independent copy of (X,Y).

- For two random variables (X,Y) with joint spherical density f, this formula can be simplified to $(\tau=0)$

$$\sigma_\tau^2 = 4\int_{\mathbb{R}^2}\left(4\int_0^{|y|}\int_0^{|x|} f(u,v)\,du\,dv\right)^2 f(x,y)\,d(x,y)\,.$$

Formula for the asymptotic variance for spherical distributions (idea of proof)

$$\sigma_\tau^2 = 4\,\mathbb{E}\big[\,\mathbb{E}[\,\mathrm{sgn}(X-\widetilde{X})\ \mathrm{sgn}(Y-\widetilde{Y})\,|\,X,Y\,]^2\,\big]$$



$$\sigma_\tau^2 = 4\int_{\mathbb{R}^2}\Big(4\int_0^{|y|}\int_0^{|x|} f(u,v)\,du\,dv\Big)^2 f(x,y)\,d(x,y)$$

Normal variance mixture distributions

Definition

$X = (X_1, \ldots, X_d)^\top$ has a normal variance mixture distribution with location vector μ and dispersion matrix Σ, if there exist $k \in \mathbb{N}$, a matrix $A \in \mathbb{R}^{d \times k}$ with $AA^\top = \Sigma$, and random variables W, Z satisfying

$$X \stackrel{\mathrm{d}}{=} \mu + \sqrt{W} AZ,$$

with

1. Z a k-dimensional standard normally distributed random vector, and
2. $W \geq 0$, a radial random variable, independent of Z.

Asymptotic variance of the tau-estimator for standard normal variance mixture distributions

Theorem (Dengler/Schmock)

For a two-dimensional standard normal variance mixture distribution with mixing distribution function G satisfying $G(0) = 0$, the asymptotic variance of the tau-estimator simplifies to

$$\sigma_\tau^2 = \frac{16}{\pi^2} \iiint_{(0,\infty)^3} \arctan^2\Bigl(\frac{\sqrt{\upsilon\xi}}{\sqrt{\zeta}\,\sqrt{\upsilon+\xi+\zeta}}\Bigr)\, dG(\upsilon)\, dG(\xi)\, dG(\zeta)\,.$$

Standard normal distribution

The asymptotic variance of the standard estimator is slightly better than the asymptotic variance of the transformed tau-estimator:

$$\sigma_{\varrho}^2 = 1 \qquad \text{versus} \qquad \sigma_{\varrho(\tau)}^2 = \frac{\pi^2}{4}\sigma_\tau^2 = \frac{\pi^2}{9} \approx 1.097\,,$$

because $(\sigma_\tau^\perp)^2 = 4/9$ for the product copula and also

$$\sigma_\tau^2 = \frac{16}{\pi^2}\arctan^2\frac{1}{\sqrt{3}} = \frac{4}{9}$$

by the previous theorem applied to $G = 1_{[1,\infty)}$.

Student's t-distribution

Definition

A d-dim. t-distribution with location μ, dispersion matrix Σ, and $\nu > 0$ degrees of freedom is defined as the corresponding normal variance mixture distribution, where the mixing random variable W has the inverse Gamma distribution $\mathrm{Ig}(\frac{\nu}{2}, \frac{\nu}{2})$.

For the 2-dim. case with non-degenerate marginal distributions:

- Asymptotic variance of the standard estimator ($\nu > 4$):

$$\sigma_\varrho^2 = \left(1 + \frac{2}{\nu - 4}\right) \left(1 - \varrho^2\right)^2 .$$

- Asymptotic variance of the tau-estimator if $\varrho = 0$ ($\nu > 0$):

$$\sigma_\tau^2 = \frac{32\,\Gamma(\frac{3\nu}{2})}{\pi^2\,\Gamma^3(\frac{\nu}{2})} \int_0^\infty u^{\nu-1} \arctan^2 u \int_0^1 t^{\nu-1} \frac{(1-t)^{\nu-1}}{(u^2+t)^\nu}\, dt\, du .$$

Asymptotic variance for the uncorrelated t-distribution

Theorem (Dengler/Schmock)

For a two-dimensional uncorrelated t-distribution with $\nu \in \mathbb{N}$ degrees of freedom, the asymptotic variance of the tau-estimator has the following representation:

(i) If ν is odd, then

$$\sigma_\tau^2 = \frac{16}{\pi^2}\log^2(2) + \frac{32\,\Gamma(\frac{3\nu}{2})}{\pi\,\Gamma^3(\frac{\nu}{2})}\sum_{k=0}^{\nu-1}\frac{(-1)^{\frac{\nu-1}{2}+k}}{\nu+2k}\binom{\nu-1}{k}\binom{\nu+k-1}{k}$$
$$\times\sum_{h=1}^{\frac{\nu-1}{2}+k}\frac{1}{h}\biggl(\log(2)+\sum_{l=1}^{2h}\frac{(-1)^l}{l}\biggr);$$

(ii) If ν is even, then

$$\sigma_\tau^2 = \frac{32\,\Gamma(\frac{3\nu}{2})}{\pi^2\,\Gamma^3(\frac{\nu}{2})}\sum_{k=0}^{\nu-1}\frac{(-1)^{\frac{\nu}{2}+k-1}}{\nu+2k}\binom{\nu-1}{k}\binom{\nu+k-1}{k}$$
$$\times\sum_{l=\nu/2}^{\nu/2+k-1}\biggl(\frac{\pi^2}{4(l+1)}-\frac{1}{2l+1}\Bigl(\frac{\pi^2}{3}+\sum_{n=1}^{l}\frac{1}{n^2}\Bigr)\biggr).$$

Asymptotic variance of the transformed tau-estimators for the uncorrelated t-distribution with even ν

ν	$\sigma^2_{\varrho(\tau)} = \pi^2\sigma^2_\tau/4$
2	$\frac{8}{3} - \frac{1}{9}\pi^2$
4	$-\frac{1\,000}{27} + \frac{35}{9}\pi^2$
6	$\frac{401\,312}{675} - \frac{541}{9}\pi^2$
8	$-\frac{42\,307\,408}{3675} + \frac{10\,499}{9}\pi^2$
10	$\frac{71\,980\,077\,752}{297\,675} - \frac{220\,501}{9}\pi^2$

Asymptotic variance of the transformed tau-estimators for the uncorrelated t-distribution with odd ν

ν	$\sigma^2_{\varrho(\tau)} = \pi^2\sigma^2_\tau/4$
1	$4\log^2(2)$
3	$30 - 44\log(2) + 4\log^2(2)$
5	$-\frac{20\,221}{54} + \frac{1\,618}{3}\log(2) + 4\log^2(2)$
7	$\frac{342\,071}{50} - \frac{148\,066}{15}\log(2) + 4\log^2(2)$
9	$-\frac{1\,358\,296\,703}{9\,800} + \frac{20\,995\,691}{105}\log(2) + 4\log^2(2)$

Bounds and limits for the asymptotic variance σ_τ^2 of the tau-estimators

Theorem (Dengler/Schmock)

1. General upper bound: $\sigma_\tau^2 \le 4(1-\tau^2)$.
2. For axially symmetric distributions: $\sigma_\tau^2 \le 4/3$.
3. For uncorrelated t-distributions:

$$\lim_{\nu\to\infty} \sigma_\tau^2 = \frac{4}{9} \quad \text{and} \quad \lim_{\nu\searrow 0} \sigma_\tau^2 = \frac{4}{3},$$

hence

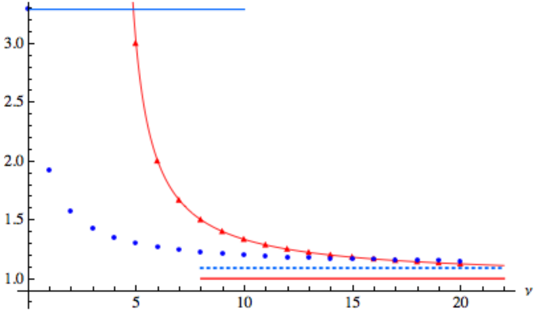
$$\sigma_{\varrho(\tau)}^2 = \frac{\pi^2}{4}\sigma_\tau^2 \to \frac{\pi^2}{3} \approx 3.290 \quad \text{as } \nu \searrow 0.$$

The upper bound in (2) is attained by (RU, RV) with independent, symmetric $\{-1,+1\}$-valued U and V, and $R \ge 0$ with density.

Comparison of the estimators for uncorrelated t-distributions with different degrees ν of freedom

ν	$\nu \downarrow 0$	1	2	3	4	5	6	7	8	9
σ_ϱ^2	n. a.	n. a.	n. a.	n. a.	n. a.	3	2	1.667	1.500	1.400
$\sigma_{\varrho(\tau)}^2$	3.290	1.922	1.570	1.423	1.345	1.296	1.263	1.240	1.222	1.208

ν	10	11	12	13	14	15	16	17	...	∞
σ_ϱ^2	1.333	1.286	1.250	1.222	1.200	1.182	1.167	1.154	...	1
$\sigma_{\varrho(\tau)}^2$	1.197	1.188	1.180	1.174	1.168	1.164	1.159	1.156	...	1.097



Results for the uncorrelated t-distribution

- For heavy-tailed t-distributions ($\nu \leq 4$), the transformed estimator is asymptotically normal with finite asymptotic variance whereas the standard estimator can not be asymptotically normal with finite variances.
- For $\nu \in \{5, 6, \ldots, 16\}$ the transformed estimator has a smaller asymptotic variance than the standard estimator and is in this sense better. Especially for small ν the difference is remarkable.
- The two estimating methods are approximately equivalent for $\nu \approx 17$, where the corresponding t-distribution is already quite similar to the normal distribution.

Asymptotic variance for the t-distribution (1)

Main steps to solve the integrals for even ν:

- Reduce $u^{\nu-1}$ to u by writing

$$u^{\nu-1} = u\,(t+u^2-t)^{\frac{\nu}{2}-1} = u \sum_{j=0}^{\frac{\nu}{2}-1} \binom{\frac{\nu}{2}-1}{j} (t+u^2)^j\,(-t)^{\frac{\nu}{2}-j-1}$$

  and dividing by $(t+u^2)^\nu$ as far as possible.
- Reduce the remaining $(t+u^2)^{\nu-j}$ to $(t+u^2)^2$ by $\nu-j-2$ integrations by parts:

$$\int_0^1 \frac{t^{\frac{3\nu}{2}-j-2}\,(1-t)^{\nu-1}}{(t+u^2)^{\nu-j}}\,dt$$

$$= \sum_{k=0}^{\nu-1} \frac{(-1)^k}{\frac{\nu}{2}+k} \binom{\nu-1}{k} \binom{\frac{3\nu}{2}-j+k-2}{\nu-j-1} \int_0^1 \frac{t^{\frac{\nu}{2}+k}}{(t+u^2)^2}\,dt$$

Asymptotic variance for the t-distribution (2)

- Reduce the $\arctan^2$ by

$$\int_0^\infty \frac{u \arctan^2 u}{(t+u^2)^2}\,du = \int_0^\infty \frac{\arctan u}{(1+u^2)\,(t+u^2)}\,du$$

- To solve the remaining integrals use

$$\frac{t^k - 1}{(1+u^2)\,(t+u^2)} = \Big(\frac{1}{1+u^2} - \frac{1}{t+u^2}\Big)\sum_{l=0}^{k-1} t^l$$

Asymptotic variance for the t-distribution (3)

Main steps to solve the integrals for odd $\nu \geq 3$:

- First steps are similar to the case of even ν.
- With $l \in \mathbb{N}$, reduce the $\arctan^2$ by

$$\int_0^1 t^l \int_0^\infty \frac{u^2 \arctan^2 u}{(t+u^2)^2}\,du\,dt = \frac{\pi^3}{24\,(2l+1)} + \frac{2l}{2l+1} \int_0^1 t^l \int_0^\infty \frac{u \arctan u}{(1+u^2)\,(t+u^2)}\,du\,dt\,.$$

- Show that

$$\int_0^\infty \frac{u \arctan u}{1+u^2} \log\left(1+\frac{1}{u^2}\right) du = \frac{\pi}{2}\left(\frac{\pi^2}{12} - \log^2(2)\right). \quad (1)$$

Some literature

- B. Dengler (2010): *On the Asymptotic Variance of the Estimator of Kendall's Tau,* Ph. D. thesis, TU Vienna.
- V. S. Koroljuk, Yu. V. Borovskich (1994): *Theory of U-statistics*, Kluwer Academic Publishers.
- Yu. V. Borovskikh (1996): *U-statistics in Banach spaces*, VSP, Utrecht.
- E. L. Lehmann, G. Casella (1998): *Theory of Point Estimation,* 2nd ed., Springer, New York.
- F. Lindskog, A. McNeil, U. Schmock (2003): *Kendall's tau for elliptical distributions*.
- H. Witting, U. Müller-Funk (1995): *Mathematische Statistik II*, B. G. Teubner, Stuttgart.