Lyon PhD Course Actuarial Science

Chapter 7 - Dependence measurement, copulas and estimation

Professor : Etienne Marceau

May 16, 2017

Contents

1 Measuring Dependence

The objective of the next sections is to measure the strength of the dependence relation between two random variables.

We consider two types of dependence measures :

- linear correlation
- rank correlation

Then, we discuss briefly about estimation procedures.

2 Concordance Measures

2.1 Desirables properties

Let (X_1, X_2) be a pair of continuous rvs.

Let $\pi(X_1, X_2)$ be a concordance measure.

Axiom 1 *The desirable properties for the dependence measure are :*

1. *Symetry:* $\pi(X_1, X_2) = \pi(X_2, X_1)$;
2. *Normalization:* $-1 \leq \pi(X_1, X_2) \leq 1$;
3. *Comonotonicity:* $\pi(X_1, X_2) = 1$ *if and if* X_1 *and* X_2 *sont comonotonic ;*
4. *Countermonotonicity:* $\pi(X_1, X_2) = -1$ *if and if* X_1 *and* X_2 *sont countermonotonic ;*

5. *Invariance: for every stricty monotone function* $\phi : \mathbb{R} \to \mathbb{R}$, *we have*

$$\pi(\phi(X_1), X_2) = \begin{cases} \pi(X_1, X_2), \text{ if } \phi \text{ is increasing} \\ -\pi(X_1, X_2), \text{ if } \phi \text{ is decreasing} \end{cases}.$$

2.2 Pearson's linear coefficient

Let (X_1, X_2) be a pair of continuous rvs such that the expectation and the variance exist.

Pearson's linear coefficient is defined by

$$\rho_P(X_1, X_2) = \frac{\mathsf{Cov}(X_1, X_2)}{\sqrt{\mathsf{Var}(X_1)\,\mathsf{Var}(X_2)}}.$$

Properties #1, #2 are satisfied.

Properties #3, #4, #5 are not satisfied.

The marginals influence the value of $\rho_P(X_1, X_2)$.

It means that $\rho_P(X_1, X_2)$ does not only measure the dependence relation between X_1 and X_2.

2.3 Measures of rank correlation and concordance

Let (X_1, X_2) be a pair of continuous rvs

Two main measures of rank correlation :

- Spearman's rho
- Kendall's tau

Since (X_1, X_2) is a couple of continuous rvs, there is exists a unique copula such that

$$F_{X_1,X_2}(x_1, x_2) = C\left(F_{X_1}(x_1), F_{X_2}(x_2)\right).$$

Nelsen (2009) :

- the term "correlation measure" is more often used for Pearson's linear correlation coefficient ;
- the terms "association measure" or "concordance measure" are used for Spearman's rho and Kendall's tau.

We do not treat measures of rank correlation for pair of discrete rvs.

3 Spearman's Rho

3.1 Definition

Spearman's rho is a rank measure.

It is also consider as a concordance meaure;

(Historical note : Charles Spearman was a colleague of Karl Pearson.)

Definition 2 *Let (X_1, X_2) be a pair of continuous rvs. Spearman's rho is defined by*

$$\rho_S(X_1, X_2) = \rho_P\left(F_{X_1}(X_1), F_{X_2}(X_2)\right).$$

Spearman's rho measures the linear correlation between the marginal cdfs of the rvs X_1 and X_2.

Recall that

$$\begin{aligned} U_1 &= F_{X_1}(X_1) \\ U_2 &= F_{X_2}(X_2), \end{aligned}$$

Then, we have

$$\begin{aligned} \rho_S(X_1, X_2) &= \rho_P\left(F_{X_1}(X_1), F_{X_2}(X_2)\right) \\ &= \frac{E[U_1U_2] - E[U_1]E[U_2]}{\sqrt{\mathrm{Var}(U_1)\mathrm{Var}(U_2)}} \end{aligned} \tag{1}$$

Since

$$\sqrt{\mathrm{Var}\left(U_1\right)\mathrm{Var}\left(U_2\right)} = \frac{1}{12}$$

and

$$E\left[U_1\right]E\left[U_2\right] = \frac{1}{4},$$

we have

$$\rho_S\left(X_1, X_2\right) = 12\left(E\left[U_1U_2\right] - \frac{1}{4}\right).$$

Note that, if C is absolutely continuous with pdf c, then

$$E\left[U_1U_2\right] = \int_0^1\int_0^1 u_1u_2c\left(u_1, u_2\right)\mathrm{d}u_1\mathrm{d}u_2.$$

Also, from the initial definition of the Spearman's rho and if F_{X_1,X_2} is absolutely continuod with pdf f_{X_1,X_2}, we mention that

$$\begin{aligned} E\left[U_1U_2\right] &= E\left[F_{X_1}\left(X_1\right)F_{X_2}\left(X_2\right)\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_1}\left(x_1\right)F_{X_2}\left(x_2\right)f_{X_1,X_2}\left(x_1,x_2\right)\mathrm{d}x_1\mathrm{d}x_2. \end{aligned}$$

3.2 Spearman's rho and desirable properties

Spearman's rho satisfies the five desirable properties :

- Symetry
- Normalization
- Comonotonicity
- Countermonotonicity
- Invariance

It implies that Spearman's rho is a measure of concordance.

3.3 Spearman's rho and comonotonicity

Remark 3 *If X_1 and X_2 sont comonotonic, then $\rho_S(X_1, X_2) = 1$.*

Let X_1 and X_2 be two continuous comonotonic rvs.

Let $U \sim Unif(0, 1)$.

Then, we have

$$F_{X_1}(X_1) = U_1 = U \quad \text{and} \quad F_{X_2}(X_2) = U_2 = U.$$

It follows that

$$E[U_1 U_2] = E\left[U^2\right] = \frac{1}{3}$$

and

$$\rho_S\left(X_1, X_2\right) = 12\left(\frac{1}{3} - \frac{1}{4}\right) = 1.$$

3.4 Spearman's rho and countermonotonicity

Remark 4 *If X_1 and X_2 sont countermonotonic, then $\rho_S(X_1, X_2) = -1$.*

Let X_1 and X_2 be two continuous countermonotonic rvs.

Let $U \sim Unif(0,1)$.

Then, we have

$$F_{X_1}(X_1) = U_1 = U \quad \text{and} \quad F_{X_2}(X_2) = U_2 = 1 - U.$$

It follows that

$$E[U_1 U_2] = E[U] - E\left[U^2\right] = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

and

$$\rho_S(X_1, X_2) = 12\left(\frac{1}{6} - \frac{1}{4}\right) = -1.$$

3.5 Spearman's rho and independence

Remark 5 *If X_1 and X_2 sont independent, then $\rho_S(X_1, X_2) = 0$.*

Let X_1 and X_2 be two continuous independent rvs.

Let $U_1 \sim Unif(0,1)$ and $U_2 \sim Unif(0,1)$ be independent rvs.

Then, we have

$$F_{X_1}(X_1) = U_1 \quad \text{and} \quad F_{X_2}(X_2) = U_2$$

It follows that

$$E[U_1 U_2] = E[U_1]\, E[U_2] = \frac{1}{4}$$

and

$$\rho_S(X_1, X_2) = 12\left(\frac{1}{4} - \frac{1}{4}\right) = 0.$$

3.6 Spearman's rho and copula

Using (**??**), the expression for Spearman's rho de Spearman can be written as follows :

$$\rho_S(X_1, X_2) = 12\left(\int_0^1 \int_0^1 C(u_1, u_2)\,\mathrm{d}u_1\mathrm{d}u_2 - \frac{1}{4}\right).$$

How do we get this result ?

To simplify the presentation, assume that the copula C is absolutely continuous with cdf c.

Recall that

$$E[U_1U_2] = \int_0^1 \int_0^1 u_1u_2c(u_1, u_2)\,\mathrm{d}u_1\mathrm{d}u_2.$$

With integration by parts, we have

$$E\left[U_1U_2\right] = \int_0^1\int_0^1 C\left(u_1,u_2\right)\mathrm{d}u_1\mathrm{d}u_2.$$

Then, we get the desired result with

$$\begin{aligned}\rho_S\left(X_1,X_2\right) &= 12\left(\int_0^1\int_0^1 u_1u_2c\left(u_1,u_2\right)\mathrm{d}u_1\mathrm{d}u_2-\frac{1}{4}\right)\\ &= 12\left(\int_0^1\int_0^1 C\left(u_1,u_2\right)\mathrm{d}u_1\mathrm{d}u_2-\frac{1}{4}\right)\end{aligned}$$

3.7 Spearman's rho and EFGM copula

Let $(U_1, U_2) \sim C_\alpha$ where

$$C_\alpha (u_1, u_2) = u_1 u_2 + \alpha u_1 u_2 (1 - u_1) (1 - u_2)$$

with

$$c (u_1, u_2) = 1 + \alpha (1 - 2u_1) (1 - 2u_2) .$$

Then

$$
\begin{aligned}
E\left[U_1U_2\right] &= \int_0^1\int_0^1 u_1u_2c\left(u_1,u_2\right)\mathrm{d}u_1\mathrm{d}u_2 \\
&= \int_0^1\int_0^1 u_1u_2\left(1+\alpha\left(1-2u_1\right)\left(1-2u_2\right)\right)\mathrm{d}u_1\mathrm{d}u_2 \\
&= \int_0^1\int_0^1 u_1u_2\left(1+\alpha\right)\mathrm{d}u_1\mathrm{d}u_2 \\
&\quad -\alpha\int_0^1\int_0^1 2u_1^2u_2\mathrm{d}u_1\mathrm{d}u_2 \\
&\quad -\alpha\int_0^1\int_0^1 u_12u_2^2\mathrm{d}u_1\mathrm{d}u_2 \\
&\quad +\alpha\int_0^1\int_0^1 2u_1^22u_2^2\mathrm{d}u_1\mathrm{d}u_2 \\
&= \left(1+\alpha\right)\left(\frac{1}{2}\times\frac{1}{2}\right)-\alpha\left(\frac{2}{3}\times\frac{1}{2}\right)-\alpha\left(\frac{1}{2}\times\frac{2}{3}\right)+\alpha\left(\frac{2}{3}\times\frac{2}{3}\right).
\end{aligned}
$$

Finally, we get

$$\begin{aligned} E\left[U_1U_2\right] &= \frac{1}{4}+\alpha\left(\frac{1}{4}-\frac{1}{3}-\frac{1}{3}+\frac{4}{9}\right) \\ &= \frac{1}{4}+\alpha\left(\frac{1}{36}\right). \end{aligned}$$

We conclude that

$$\begin{aligned} \rho_S\left(U_1,U_2\right) &= 12\left(\int_0^1\int_0^1 u_1u_2c\left(u_1,u_2\right)\mathrm{d}u_1\mathrm{d}u_2-\frac{1}{4}\right) \\ &= 12\left(\frac{1}{4}+\alpha\frac{1}{36}-\frac{1}{4}\right) \\ &= \frac{\alpha}{3}. \end{aligned}$$

Then, for this copula, we observe

$$-\frac{1}{3}\leq\rho_S\left(U_1,U_2\right)\leq\frac{1}{3}.$$

Let (X_1, X_2) be a pair of continuous rvs with

$$F_{X_1,X_2}(x_1, x_2) = C\left(F_{X_1}(x_1), F_{X_2}(x_2)\right),$$

where C is the EFGM copula.

Then, for any marginals distribution for X_1 and X_2,

$$\rho_S(X_1, X_2) = \frac{\alpha}{3}$$

and

$$-\frac{1}{3} \leq \rho_S(X_1, X_2) \leq \frac{1}{3}.$$

3.8 Copula and explicit expression for Spearman's rho

Copula	$\rho_S(X_1, X_2)$
Clayton	–
Normal	$\frac{6}{\pi}\arcsin(\alpha)$
Gumbel	–
Fréchet	$(\alpha - \beta)$
EFGM	$\frac{\alpha}{3}$
Marshall-Olkin	$\frac{3\alpha\beta}{2\alpha+\beta-2\alpha\beta}$

3.9 Spearman's rho and alternative representation

Recall that

$$\begin{aligned} E\left[U_1 U_2\right] &= E\left[F_{X_1}\left(X_1\right) F_{X_2}\left(X_2\right)\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_1}\left(x_1\right) F_{X_2}\left(x_2\right) f_{X_1,X_2}\left(x_1, x_2\right) \mathrm{d}x_1 \mathrm{d}x_2. \end{aligned}$$

Integration by part leads to

$$\begin{aligned} E\left[U_1 U_2\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_1}\left(x_1\right) F_{X_2}\left(x_2\right) f_{X_1,X_2}\left(x_1, x_2\right) \mathrm{d}x_1 \mathrm{d}x_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_1,X_2}\left(x_1, x_2\right) f_{X_1}\left(x_1\right) f_{X_2}\left(x_2\right) \mathrm{d}x_1 \mathrm{d}x_2. \end{aligned}$$

Then, we have the following alternative representation for $\rho_S(X_1, X_2)$:

$$\begin{aligned}\rho_S(X_1, X_2) &= 12\left(E\left[F_{X_1}(X_1)\, F_{X2}(X_2)\right] - E\left[F_{X_1}(X_1)\right] E\left[F_{X_2}(X_2)\right]\right) \\ &= 12\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_1,X_2}(x_1, x_2)\, f_{X_1}(x_1)\, f_{X_2}(x_2)\, \mathsf{d}x_1 \mathsf{d}x_2 - \left(\int_{-\infty}^{\infty}\right.\right. \\ &= 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(F_{X_1,X_2}(x_1, x_2) - F_{X_1}(x_1)\, F_{X_2}(x_2)\right) f_{X_1}(x_1)\, f_{X_2}\end{aligned}$$

Frechet's upper and lower bounds :

- Let $\left(X_1^+, X_2^+\right)$ be a pair of continuous comonotonic rvs with $X_1^+ \sim X_1$ and $X_2^+ \sim X_2$ i.e.

$$F_{X_1^+,X_2^+}(x_1, x_2) = \min\left(F_{X_1}(x_1), F_{X_2}(x_2)\right)$$

- Let $\left(X_1^-, X_2^-\right)$ be a pair of continuous countermonotonic rvs with $X_1^- \sim X_1$ and $X_2^- \sim X_2$ i.e.

$$F_{X_1^-,X_2^-}(x_1,x_2) = \max\left(F_{X_1}(x_1) + F_{X_2}(x_2) - 1; 0\right)$$

- We know that

$$F_{X_1^-,X_2^-}(x_1,x_2) \leq F_{X_1,X_2}(x_1,x_2) \leq F_{X_1^+,X_2^+}(x_1,x_2)$$

for all $F_{X_1,X_2} \in \Gamma\left(F_{X_1}, F_{X_2}\right)$.

- Then, we have

$$\begin{aligned}\rho_S(X_1,X_2) &= 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(F_{X_1,X_2}(x_1,x_2) - F_{X_1}(x_1)F_{X_2}(x_2)\right) f_{X_1}(x_1) \\ &\leq 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(F_{X_1^+,X_2^+}(x_1,x_2) - F_{X_1}(x_1)F_{X_2}(x_2)\right) f_{X_1}(x \\ &= \rho_S\left(X_1^+, X_2^+\right) = 1\end{aligned}$$

- Also, we have

$$\begin{aligned}\rho_S(X_1, X_2) &= 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(F_{X_1,X_2}(x_1, x_2) - F_{X_1}(x_1)\,F_{X_2}(x_2)\right) f_{X_1}(x_1)\\ &\geq 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(F_{X_1^-,X_2^-}(x_1, x_2) - F_{X_1}(x_1)\,F_{X_2}(x_2)\right) f_{X_1}(x\\ &= \rho_S\left(X_1^-, X_2^-\right) = -1.\end{aligned}$$

- We conclude

$$-1 = \rho_S\left(X_1^-, X_2^-\right) \leq \rho_S(X_1, X_2) \leq \rho_S\left(X_1^+, X_2^+\right) = 1,$$

for all $F_{X_1,X_2} \in \Gamma\left(F_{X_1}, F_{X_2}\right)$.

The transformation in (1) allows to take into account the ranks by eliminating the effect of the marginals (marginal cdfs).

- If the X_1 tends to increase when the rv X_2 increases, then Spearman's tho will take a positive value.

- If the X_1 tends to decrease when the rv X_2 increases, then Spearman's tho will take a negative value.

3.10 Additional remark

If ϕ_1, ϕ_2 are strictly monotone increasing functions, then

$$\rho_S\left(\phi_1\left(X_1\right), \phi_2\left(X_2\right)\right) = \rho_S\left(X_1, X_2\right).$$

If ϕ_1, ϕ_2 are strictly monotone decreasing functions, then

$$\rho_S\left(\phi_1\left(X_1\right), \phi_2\left(X_2\right)\right) = \rho_S\left(X_1, X_2\right).$$

3.11 Estimation of Spearman's rho

Let $\underline{X} = (X_1, X_2)$ be a pair of continuous rvs with cdf F_{X_1,X_2}.

Let

$$\left\{\left(X_{1,j}, X_{2,j}\right), j = 1, 2, ..., n\right\}.$$

be a finite sequence of iid pairs of rvs where $\underline{X}_j = \left(X_{1,j}, X_{2,j}\right) \sim \underline{X} = (X_1, X_2)$.

Let

$$\left\{\left(x_{1,j}, x_{2,j}\right), j = 1, 2, ..., n\right\}$$

be the empirical observations of

$$\left\{\left(X_{1,j}, X_{2,j}\right), j = 1, 2, ..., n\right\}.$$

Notation : $\underline{x}_j = \left(x_{1,j}, x_{2,j}\right)$

Definition: For a fixed $i \in \{1, 2\}$, the rank of the observation $x_{i,j}$, denoted $rank\left(x_{i,j}\right)$, corresponds to its position ("rank") in the observations $\left\{x_{i,1}, ..., x_{i,n}\right\}$.

Empirical cdf:

$$F_{1,n}\left(x_{1,j}\right) = \frac{1}{n+1}\sum_{l=1}^{n} 1_{\left\{x_{1,j} \leq x_{1,j}\right\}} = \frac{rank\left(x_{1,j}\right)}{n+1}$$

and

$$F_{2,n}\left(x_{2,j}\right) = \frac{1}{n+1}\sum_{l=1}^{n} 1_{\left\{x_{2,j} \leq x_{2,j}\right\}} = \frac{rank\left(x_{2,j}\right)}{n+1}.$$

Then, we have the sequence of pairs of pseudo-observations :

$$\left\{\left(F_{1,n}\left(x_{1,j}\right), F_{1,n}\left(x_{2,j}\right)\right), j = 1, 2, ..., n\right\}.$$

For $i = 1, 2$, the empirical means are

$$\begin{aligned}
\widetilde{F}_1 &= \frac{1}{n}\sum_{l=1}^{n} F_{1,n}\left(x_{1,j}\right) \\
&= \frac{1}{n}\frac{1}{n+1}(1 + ... + n) \\
&= \frac{1}{n}\frac{1}{n+1}\frac{(n \times (n+1))}{2} = \frac{1}{2}
\end{aligned}$$

et

$$
\begin{aligned}
\widetilde{F}_2 &= \frac{1}{n}\sum_{l=1}^{n} F_{2,n}\left(x_{2,j}\right) \\
&= \frac{1}{n}\frac{1}{n+1}\frac{(n\times(n+1))}{2} = \frac{1}{2}.
\end{aligned}
$$

The empirical variance is

$$
\frac{1}{n}\sum_{j=1}^{n}\left(F_{1,n}\left(x_{1,j}\right)-\widetilde{F}_1\right)^2
$$

and

$$
\frac{1}{n}\sum_{j=1}^{n}\left(F_{2,n}\left(x_{1,j}\right)-\widetilde{F}_2\right)^2.
$$

The empirical estimator of $\rho_S(X_1, X_2)$ is defined by $\widehat{\rho}_S(X_1, X_2)$ where

$$\widehat{\rho}_S(X_1, X_2) = \widehat{\rho}_P\left(F_{1,n}(X_1), F_{2,n}(X_2)\right).$$

We have

$$\begin{aligned}
\widehat{\rho}_S(X_1, X_2) &= \widehat{\rho}_P\left(F_{1,n}(X_1), F_{2,n}(X_2)\right). \\
&= \frac{\frac{1}{n}\sum_{j=1}^{n}\left(F_{1,n}\left(x_{1,j}\right) - \widetilde{F}_{,1}\right)\left(F_{2,n}\left(x_{2,j}\right) - \widetilde{F}_2\right)}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(F_{1,n}\left(x_{1,j}\right) - \widetilde{F}_1\right)^2 \frac{1}{n}\sum_{j=1}^{n}\left(F_{2,n}\left(x_{1,j}\right) - \widetilde{F}_2\right)^2}} \\
&= \frac{\sum_{j=1}^{n}\left(F_{1,n}\left(x_{1,j}\right) - \frac{1}{2}\right)\left(F_{2,n}\left(x_{2,j}\right) - \frac{1}{2}\right)}{\sqrt{\sum_{j=1}^{n}\left(F_{1,n}\left(x_{1,j}\right) - \frac{1}{2}\right)^2 \sum_{j=1}^{n}\left(F_{2,n}\left(x_{1,j}\right) - \frac{1}{2}\right)^2}} \\
&= \frac{\sum_{j=1}^{n}\left(\frac{rank(x_{1,j})}{n+1} - \frac{1}{2}\right)\left(\frac{rank(x_{1,j})}{n+1} - \frac{1}{2}\right)}{\sqrt{\sum_{j=1}^{n}\left(F_{1,n}\left(x_{1,j}\right) - \frac{1}{2}\right)^2 \sum_{j=1}^{n}\left(F_{2,n}\left(x_{1,j}\right) - \frac{1}{2}\right)^2}}
\end{aligned}$$

Rearranging the terms, we obtain

$$\widehat{\rho}_S\left(X_1, X_2\right) = \frac{12}{n\left(n+1\right) \times \left(n-1\right)} \times \sum_{j=1}^{n} rank\left(x_{1,j}\right) rank\left(x_{2,j}\right) - 3\frac{n+1}{n-1}$$

(see e.g. Favre et Genest (2007) for the details).

4 Kendall's tau

4.1 Definition

Kendall's tau is a measure of concordance between the rvs continuous X_1 and X_2.

Kendall's tau is a probability :

$$\tau(X_1, X_2) = \Pr(\text{concordance}) - \Pr(\text{discordance}).$$

Definition 6 *Let (X_1, X_2) be a pair of continuous rvs with cdf F_{X_1,X_2}. For the definition, we introduce $\left(X_1', X_2'\right)$ which is a pair of continuous rvs where*

$\left(X_1', X_2'\right)$ *and* (X_1, X_2) *are independent and* $\left(X_1', X_2'\right) \sim (X_1, X_2)$ *.Kendall's tau is defined by*

$$\tau(X_1, X_2) = \Pr\left(\left(X_1 - X_1'\right)\left(X_2 - X_2'\right) > 0\right) - \Pr\left(\left(X_1 - X_1'\right)\left(X_2 - X_2'\right) < 0\right).$$

Interpretation :

- **In class.**

Note that

$$\Pr\left(\left(X_1 - X_1'\right)\left(X_2 - X_2'\right) < 0\right) = 1 - \Pr\left(\left(X_1 - X_1'\right)\left(X_2 - X_2'\right) > 0\right).$$

The expression for $\tau(X_1, X_2)$ becomes

$$\tau\left(X_1, X_2\right) = 2 \Pr\left(\left(X_1 - X_1'\right)\left(X_2 - X_2'\right) > 0\right) - 1.$$

Also, we have

$$\Pr\left(\left(X_1 - X_1'\right)\left(X_2 - X_2'\right) > 0\right) = \Pr\left(X_1 > X_1', X_2 > X_2'\right) + \Pr\left(X_1 \leq X_1', X_2 \leq\right.$$

Since the rvs are contiuous, we have

$$\Pr\left(X_1 > X_1', X_2 > X_2'\right) = \Pr\left(U_1 > U_1', U_2 > U_2'\right)$$

and

$$\Pr\left(X_1 \leq X_1', X_2 \leq X_2'\right) = \Pr\left(U_1 \leq U_1', U_2 \leq U_2'\right).$$

We have

$$\begin{aligned}\mathsf{Pr}\left(U_1 \leq U_1', U_2 \leq U_2'\right) &= \int_0^1 \int_0^1 \mathsf{Pr}\left(U_1 \leq U_1', U_2 \leq U_2' | U_1' = u_1, U_2' = u_2\right) \mathsf{d}C\,(\\ &= \int_0^1 \int_0^1 \mathsf{Pr}\left(U_1 \leq u_1, U_2 \leq u_2\right) \mathsf{d}C\left(u_1, u_2\right) \\ &= \int_0^1 \int_0^1 C\left(u_1, u_2\right) \mathsf{d}C\left(u_1, u_2\right).\end{aligned}$$

Also, we ahve and

$$\begin{aligned}\mathsf{Pr}\left(U_1 > U_1', U_2 > U_2'\right) &= \mathsf{Pr}\left(U_1' \leq U_1, U_2' \leq U_2\right) \\ &= \int_0^1 \int_0^1 C\left(u_1, u_2\right) \mathsf{d}C\left(u_1, u_2\right)\end{aligned}$$

We conclude

$$\begin{aligned}\tau(X_1, X_2) &= 4\int_0^1\int_0^1 C(u_1, u_2)\,\mathrm{d}C(u_1, u_2) - 1\\ &= 4E[C(U_1, U_2)] - 1.\end{aligned}$$

If the copula C is absolutely continuous with pdf c, we have

$$\begin{aligned}\tau(X_1, X_2) &= 4\int_0^1\int_0^1 C(u_1, u_2)\,\mathrm{d}C(u_1, u_2) - 1\\ &= 4E[C(U_1, U_2)] - 1\\ &= 4\int_0^1\int_0^1 C(u_1, u_2)\,c(u_1, u_2)\,\mathrm{d}u_1\mathrm{d}u_2 - 1.\end{aligned}$$

4.2 Kendall's tau and desirable properties

Kendall's tau satisfies the five desirable properties :

- Symetry
- Normalization
- Comonotonicity
- Countermonotonicity
- Invariance

It implies that Kendall's tau is a measure of concordance.

4.3 Kendall's tau and copulas

Copula	$\tau(X_1, X_2)$
Clayton	$\frac{\alpha}{\alpha+2}$
Normale	$\frac{2}{\pi}\arcsin(\alpha)$
Gumbel	$\frac{\alpha-1}{\alpha}$
Fréchet	$\frac{(\alpha-\beta)(\alpha+\beta+2)}{3}$
EFGM	$\frac{2\alpha}{9}$
Marshall-Olkin	$\frac{\alpha\beta}{\alpha+\beta-\alpha\beta}$

4.4 Additional remark

If ϕ_1, ϕ_2 are strictly monotone increasing functions, then

$$\tau\left(\phi_1\left(X_1\right), \phi_2\left(X_2\right)\right) = \tau\left(X_1, X_2\right)$$

If ϕ_1, ϕ_2 are strictly monotone decreasing functions, then

$$\tau\left(\phi_1\left(X_1\right), \phi_2\left(X_2\right)\right) = \tau\left(X_1, X_2\right)$$

4.5 Kendall's tau and survival copula

Kendall's tau and Spearman's rho for a specific copula are identical to those associated to the corresponding survival copulas.

Let $(U_1, U_2) \sim C$.

Then, we define (V_1, V_2) where

$$\begin{aligned} V_1 &= 1 - U_1 \\ V_2 &= 1 - U_2 \end{aligned}$$

We know that the copula for (V_1, V_2) is the survival copula $\widehat{C}$ associated to C i.e.

$$F_{V_1,V_2}(u_1, u_2) = \widehat{C}(u_1, u_2).$$

Since V_1 and V_2 correspond to the strictly decreasing transformation of U_1 and U_2, it follows

$$\rho_S\left(V_1, V_2\right) = \rho_S\left(U_1, U_2\right).$$

4.6 Estimation of Kendall's tau

We compare all $\binom{n}{2} = \frac{n(n-1)}{2}$ non-repeating combination of pairs of the following sequence of observations :

$$\left\{ \left(x_{1,j}, x_{2,j} \right), j = 1, 2, ..., n \right\}.$$

Let $\left(x_{1,j}, x_{2,j} \right)$ and $\left(x_{1,k}, x_{2,k} \right)$ be two pairs :

- The pairs are concordant if

  $$x_{1,j} > x_{1,k} \text{ et } x_{2,j} > x_{2,k}$$

  or

  $$x_{1,j} < x_{1,k} \text{ et } x_{2,j} < x_{2,k}.$$

- The pairs are discordant if

$$x_{1,j} > x_{1,k} \text{ et } x_{2,j} < x_{2,k}$$

or

$$x_{1,j} < x_{1,k} \text{ et } x_{2,j} > x_{2,k}.$$

We compare :

- $\left(x_{1,1}, x_{2,1}\right)$ with $\left(x_{1,j}, x_{2,j}\right)$, for $j = 2, ..., n$;
- $\left(x_{1,2}, x_{2,2}\right)$ with $\left(x_{1,j}, x_{2,j}\right)$, for $j = 3, ..., n$;
- $\left(x_{1,3}, x_{2,3}\right)$ with $\left(x_{1,j}, x_{2,j}\right)$, for $j = 4, ..., n$;
- ...
- $\left(x_{1,n-1}, x_{2,n-1}\right)$ with $\left(x_{1,n}, x_{2,n}\right)$.
- $\Rightarrow \binom{n}{2} =$ number of comparisons.

Definitions :

- c_n = number of concordant pairs ;
- d_n = nombre de paires concordantes.

The empirical estimator of Kendall's tau is

$$\begin{aligned}
\widehat{\tau}(X_1, X_2) &= \frac{\text{number of concordant pairs - number of disconcordant pairs}}{\text{total number of comparisons}} \\
&= \frac{c_n - d_n}{\binom{n}{2}}.
\end{aligned}$$

If all pairs are concordant, then

$$c_n = \binom{n}{2}$$

and

$$d_n = 0$$

which leads to

$$\widehat{\tau}\left(X_1, X_2\right) = 1.$$

If all pairs are disconcordant, then

$$c_n = 0$$

and

$$d_n = \binom{n}{2}$$

which leads to

$$\widehat{\tau}\left(X_1, X_2\right) = -1.$$

The following representation is frequently used :

$$\widehat{\tau}\left(X_1, X_2\right) = \frac{\sum_{j=1}^{n-1} \sum_{k=j+1}^{n} sign\left(\left(x_{1,j} - x_{1,k}\right)\left(x_{2,j} - x_{2,k}\right)\right)}{\binom{n}{2}},$$

where

$$sign\left(a\right) = \begin{cases} -1 & , a < 0 \\ 1 & , a > 0 \end{cases}.$$

5 Concordance measures and independence

If the continuous rvs X_1 and X_2 are independent, then $\tau(X_1, X_2) = 0$ and $\rho_S(X_1, X_2) = 0$.

However, if $\tau(X_1, X_2) = 0$ or $\rho_S(X_1, X_2) = 0$, it does not imply that the rvs X_1 and X_2 independent.

6 Estimation procedures and copulas

6.1 Introduction

Let $\underline{X} = (X_1, X_2)$ be a pair of continuous rvs with cdf

$$F_{X_1,X_2}(x_1, x_2) = C\left(F_{X_1}(x_1), F_{X_2}(x_2)\right).$$

Let

$$\left\{\left(X_{1,j}, X_{2,j}\right), j = 1, 2, ..., n\right\}.$$

be a finite sequence of iid pairs of rvs where $\underline{X}_j = \left(X_{1,j}, X_{2,j}\right) \sim \underline{X} = (X_1, X_2)$.

Let

$$\left\{\left(x_{1,j}, x_{2,j}\right), j = 1, 2, ..., n\right\}$$

be the empirical observations of

$$\left\{\left(X_{1,j}, X_{2,j}\right), j = 1, 2, ..., n\right\}.$$

Notation : $\underline{x}_j = \left(x_{1,j}, x_{2,j}\right)$

6.2 Fully Maximum Likelihood Estimation

Notations :

- $\underline{\theta}_i = \left(\theta_{i,1}, ..., \theta_{i,n_i}\right)$, $i = 1, 2$;
- $\alpha =$ dependence parameter (sometimes there are 2 ou 3 parameters, even more);
- $\underline{\theta} = (\alpha, \underline{\theta}_1, \underline{\theta}_2)$
- number of parameters $= n_1 + n_2 + 1$
- $F_{X_i}(x; \underline{\theta}_i)$
- $f_{X_i}(x; \underline{\theta}_i)$
- $F_{X_1,X_2}(x_1, x_2; \underline{\theta})$
- $f_{X_1,X_2}(x_1, x_2; \underline{\theta})$

Note:

- $F_{X_1,X_2}(x_1, x_2; \underline{\theta}) = C\left(F_{X_1}(x_1; \underline{\theta}_1), F_{X_2}(x_2; \underline{\theta}_2); \alpha\right)$
- $f_{X_1,X_2}(x_1, x_2; \underline{\theta}) = c\left(F_{X_1}(x_1; \underline{\theta}_1), F_{X_2}(x_2; \underline{\theta}_2); \alpha\right) \times f_{X_1}(x_1; \underline{\theta}_1) \times f_{X_2}(x_2; \underline{\theta}_2)$

Likelihood function :

$$L(\underline{\theta}) = \prod_{j=1}^{n} f_{X_1,X_2}\left(x_{1,j}, x_{2,j}; \underline{\theta}\right)$$

Log-likelihood function :

$$l(\underline{\theta}) = \sum_{j=1}^{n} \ln\left(f_{X_1,X_2}\left(x_{1,j}, x_{2,j}; \underline{\theta}\right)\right)$$

ML estimator of $\underline{\theta}$ is $\widehat{\underline{\theta}}^{ML}$ where

$$\widehat{\underline{\theta}}^{ML} = \arg\max\left(l(\underline{\theta})\right)$$

Remark: It can become very difficult to perform.

6.3 Variant to Maximum Likelihood Estimation : Semi-parametric method

The method relies on the first part of Sklar's theorem.

We have the set of pairs of observations :

$$\left\{\left(x_{1,j}, x_{2,j}\right), j = 1, 2, ..., n\right\}$$

Notation : $\underline{x}_j = \left(x_{1,j}, x_{2,j}\right)$

We define the sequence of pairs of pseudo-observations :

$$\left\{\left(u_{1,j}, u_{2,j}\right), j = 1, 2, ..., n\right\}$$

where

$$\left(u_{1,j}, u_{2,j}\right) = \left(F_{1,n}\left(x_{1,j}\right), F_{2,n}\left(x_{2,j}\right)\right), j = 1, 2, ..., n.$$

Step 1 : Estimate the marginal cdf of X_i, for each $i = 1, 2$, separately. (Find the appropriate cdf of X_i)

Step 2 : Use the sequence of pairs of pseudo-observations :

$$\left\{\left(u_{1,j}, u_{2,j}\right), j = 1, 2, ..., n\right\}$$

to estimate the copula itself. (find the appropriate copula)

Notations :

- $\underline{\theta}_i = \left(\theta_{i,1}, ..., \theta_{i,n_i}\right)$, $i = 1, 2$;

- $\alpha =$ dependence parameter (sometimes there are 2 ou 3 parameters, even more);
- $\underline{\theta} = (\alpha, \underline{\theta}_1, \underline{\theta}_2)$
- number of parameters $= n_1 + n_2 + 1$
- $F_{X_i}(x; \underline{\theta}_i)$
- $f_{X_i}(x; \underline{\theta}_i)$
- $F_{X_1,X_2}(x_1, x_2; \underline{\theta})$
- $f_{X_1,X_2}(x_1, x_2; \underline{\theta})$

Note:

- $F_{X_1,X_2}(x_1, x_2; \underline{\theta}) = C\left(F_{X_1}(x_1; \underline{\theta}_1), F_{X_2}(x_2; \underline{\theta}_2); \alpha\right)$
- $f_{X_1,X_2}(x_1, x_2; \underline{\theta}) = c\left(F_{X_1}(x_1; \underline{\theta}_1), F_{X_2}(x_2; \underline{\theta}_2); \alpha\right) \times f_{X_1}(x_1; \underline{\theta}_1) \times f_{X_2}(x_2; \underline{\theta}_2)$

Likelihood function :

$$L(\alpha) = \prod_{j=1}^{n} c\left(u_{1,j}, u_{2,j}; \alpha\right)$$

Log-likelihood function :

$$l(\alpha) = \sum_{j=1}^{n} \ln\left(c\left(u_{1,j}, u_{2,j}; \alpha\right)\right)$$

ML estimator of α is $\widehat{\alpha}^{ML}$ where

$$\widehat{\alpha}^{ML} = \arg\max\left(l(\alpha)\right)$$

Remark: It can become very difficult to perform.

6.4 Variant to Maximum Likelihood Estimation : IFM Method (Joe's method)

The method relies on the first part of Sklar's theorem.

We have the set of pairs of observations :

$$\left\{\left(x_{1,j}, x_{2,j}\right), j = 1, 2, ..., n\right\}$$

Notation : $\underline{x}_j = \left(x_{1,j}, x_{2,j}\right)$

We define the sequence of pairs of pseudo-observations :

$$\left\{\left(u_{1,j}, u_{2,j}\right), j = 1, 2, ..., n\right\}$$

where

$$\left(u_{1,j}, u_{2,j}\right) = \left(F_{X_1}\left(x_{1,j}; \underline{\theta}_1^{MV}\right), F_{X_2}\left(x_{2,j}; \underline{\theta}_2^{MV}\right)\right), j = 1, 2, ..., n.$$

Step 1 : Estimate the marginal cdf of X_i, for each $i = 1, 2$, separately. (Find the appropriate cdf of X_i)

Step 2 : Use the sequence of pairs of pseudo-observations :

$$\left\{\left(u_{1,j}, u_{2,j}\right), j = 1, 2, ..., n\right\}$$

to estimate the copula itself. (find the appropriate copula)

Notations :

- $\underline{\theta}_i = \left(\theta_{i,1}, ..., \theta_{i,n_i}\right)$, $i = 1, 2$;

- $\alpha =$ dependence parameter (sometimes there are 2 ou 3 parameters, even more);
- $\underline{\theta} = (\alpha, \underline{\theta}_1, \underline{\theta}_2)$
- number of parameters $= n_1 + n_2 + 1$
- $F_{X_i}(x; \underline{\theta}_i)$
- $f_{X_i}(x; \underline{\theta}_i)$
- $F_{X_1,X_2}(x_1, x_2; \underline{\theta})$
- $f_{X_1,X_2}(x_1, x_2; \underline{\theta})$

Note:

- $F_{X_1,X_2}(x_1, x_2; \underline{\theta}) = C\left(F_{X_1}(x_1; \underline{\theta}_1), F_{X_2}(x_2; \underline{\theta}_2); \alpha\right)$
- $f_{X_1,X_2}(x_1, x_2; \underline{\theta}) = c\left(F_{X_1}(x_1; \underline{\theta}_1), F_{X_2}(x_2; \underline{\theta}_2); \alpha\right) \times f_{X_1}(x_1; \underline{\theta}_1) \times f_{X_2}(x_2; \underline{\theta}_2)$

Likelihood function :

$$L(\alpha) = \prod_{j=1}^{n} c\left(u_{1,j}, u_{2,j}; \alpha\right)$$

Log-likelihood function :

$$l(\alpha) = \sum_{j=1}^{n} \ln\left(c\left(u_{1,j}, u_{2,j}; \alpha\right)\right)$$

ML estimator of α is $\widehat{\alpha}^{ML}$ where

$$\widehat{\alpha}^{ML} = \arg\max\left(l(\alpha)\right)$$

Remark: It can become very difficult to perform.

7 Challenges for research in actuarial science

- Advanced statistical methods for copulas.
- A lot of research is going on this topic.
- Important challenge: how do we estimate the parameters of a copula when the number of rvs is large ?
- Selection of the appropriate copula is also very challenging.
- Important impact on the distribution of the sum of the rvs.