GLOBAL DEPENDENCE MEASURES

Daniel Reis & Nguyen Ngoc

Department of Statistics, University of São Paulo

July 7, 2017

Introduction

- The copula captures the "nonparametric", "distribution-free" or "scale-invariant" nature of the association between random variables.
- Copulas provide a natural way to study and measure dependence between random variables.
- Copula properties are invariant under strictly increasing transformations of the underlying random variables.
- Pearson's correlation is most frequently used in practice as a measure of dependence. However, it can often be quite misleading and should not be taken as the canonical dependence measure.

Linear Correlation (Pearson's correlation)

Definition

Let (X, Y) be a vector of random variables with nonzero finite variances. The linear correlation coefficient for (X, Y) is

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}},$$

where $Cov(X, Y) = E(XY) - EX \cdot EY$ is the covariance of (X, Y), and $Var(X)$ and $Var(Y)$ are the variances of X and Y.

Properties

- Linear correlation is a measure of linear dependence. In the case of perfect linear dependence, i.e., $Y = aX + b$ almost surely for $a \in \mathbb{R} \setminus \{0\}$, we have $|Corr(X, Y)| = 1$. More important is that the converse also holds
- Linear correlation has the property that

$$Corr(\alpha X + \beta, \gamma Y + \delta) = sign(\alpha\gamma) Corr(X, Y)$$

  Hence, linear correlation is invariant under strictly increasing linear transformations

Comments

Problems with the Pearson correlation coefficient

- The Pearson correlation is a measure for linear dependence only.
- The linear correlation is not invariant under nonlinear strictly increasing transformations.
- Linear correlation only measures the degree of dependence but does not clearly discover the structure of dependence.
- The linear correlation coefficient does not completely determine the joint distribution.
- The linear correlation associated with $(f(X), X)$ is in general lower than 1, where f non-linear deterministic function.

Examples

- **Example 1:** Let $X \sim \mathbf{N}(0,1)$ and $Y = X^2$. Then

$$Cov(X,Y) = E(X^3) - E(X)E(X^2) = 0$$

  That is the correlation is 0. However given X, one can predict Y.

- **Example 2:** Consider the bivariate distribution with uniform margins. For $(u_1, u_2) \in [0,1]$,

$$C(u_1,u_2) = u_1 u_2 + \alpha[u_1(u_1-1)(2u_1-1)][u_2(u_2-1)(2u_2-1)]$$

  with $\alpha \in [-1, 2]$.

  If the margins F_1 and F_2 are continuous and symmetric, the Pearson correlation is zero, but for $\alpha \neq 0$ the random variables are not independent.

- **Example 3:** Let U_1 and U_2 be two $U(0,1)$ random variables with joint distribution

$$C(u_1,u_2)=\begin{cases} u_1, & 0\leq u_1\leq u_2/2\leq 1/2,\\ u_2/2, & 0\leq u_2/2\leq u_1\leq 1-u_2/2,\\ u_1+u_2-1, & 1/2\leq 1-u_2/2\leq u_1\leq 1. \end{cases}$$

We have $Cov(U_1,U_2)=0$, but $P(U_2=1-|2U_1-1|)=1$. That is, the two r.v.s are uncorrelated but one can be perfectly predicted from the other.

Concordance

- The most widely known scale-invariant measures of association are the population versions of Kendall's tau and Spearman's rho. Both measure a form of dependence known as concordance.

Definition

Two observations (x_1, y_1) and (x_2, y_2) of a pair (X, Y) of continuous random variables are concordant if $x_1 > x_2$ and $y_1 > y_2$ or if $x_1 < x_2$ and $y_1 < y_2$, i.e., if $(x_1 - x_2)(y_1 - y_2) > 0$; and discordant if $x_1 > x_2$ and $y_1 < y_2$ or if $x_1 < x_2$ and $y_1 > y_2$, i.e., if $(x_1 - x_2)(y_1 - y_2) < 0$.

- Geometrically, two distinct points (x_1, y_1) and (x_2, y_2) in the plane are concordant if the line segment connecting them has positive slope, and discordant if the line segment has negative slope.

- The sample version of the measure Kendall's tau is defined in terms of concordance as follows: Suppose that we have a random sample from a vector (X, Y) of continuous random variables. Kendall's tau is given by

$$\frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\text{total number of pairs}}$$

- Analogous to the sample version, we let $(X_1, Y_1), (X_2, Y_2)$ be independent random vectors with a common joint distribution. The population version of Kendall's tau is

$$\tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Let (X_1, Y_1) and (X_2, Y_2) be independent vectors (of continuous r.v.s.) with joint distribution functions H_1 and H_2 and copulas C_1 and C_2, respectively. Further, X_1 and X_2 have the same distribution F and Y_1 and Y_2 have the common distribution G.

Concordance function:

$$K = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Then

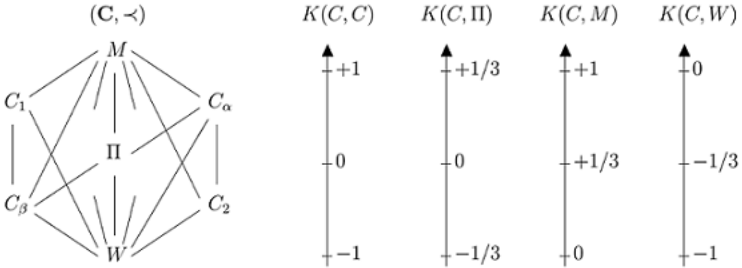
$$K = K(C_1, C_2) = 4 \int_0^1 \int_0^1 C_2(u, v) dC_1(u, v) - 1,$$

i.e. K depends only on the copulas.

Properties of Kandall's tau

Some properties of K are as follows:

- K is symmetric in its arguments: $K(C_1, C_2) = K(C_2, C_1)$;
- K is nondecreasing in each argument $C_1(u,v) \leq C_1'(u,v)$ and $C_2(u,v) \leq C_2'(u,v)$ for all $(u,v) \in I^2$ implies $K(C_1, C_2) \leq K(C_1', C_2')$.
- $K(M,M) = 1$, $K(W,W) = -1$, $K(\Pi,\Pi) = 0$, $K(M,\Pi) = 1/3$, $K(W,\Pi) = 1/3$, $K(M,W) = 0$,
  where $M(u,v) = \min(u,v), W(u,v) = \max(u+v-1, 0), \Pi = u \cdot v$.



Kendall's tau

Definition

If X and Y are continuous random variables with copula C, then the population version of Kendall's tau has a "simple" expression in terms of K:

$$\tau(X, Y) = \tau_C = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 = K(C, C).$$

Example: Let $C = C_\theta$ be a member of the Farlie-Gumbel-Morgenstern (FGM) family:

$$C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v), \quad \theta \in [-1, 1].$$

Then $\tau_C = 2\theta/9$. Since $\tau_C \in [-2/9, 2/9]$, FGM copulas can only model relatively weak dependence.

Two different expressions of Kendall's tau formula and their applications.

1) C is singular, or possesses a singular component, the form for τ_C is not amenable to computation. For many such copulas, the flowing expression is useful

$$\tau_C = 1 - 4\int_0^1\int_0^1 \frac{\partial}{\partial u}C(u,v)\frac{\partial}{\partial v}C(u,v)dudv.$$

Example: Let $C_{\alpha,\beta}$ be a member of the Cuadras-Augé family of copulas

$$C_{\alpha,\beta} = min(u^{1-\alpha}v, uv^{1-\beta}), \quad \alpha,\beta \in (0,1].$$

When $\alpha,\beta \in (0,1]$, there is a singular component on the curve $u^\alpha = v^\beta$. However, the partial derivatives of $C_{\alpha,\beta}$ are easily evaluated. Using above formula, we have $\tau_{\alpha,\beta} = \alpha\beta/(\alpha - \alpha\beta + \beta)$.

2) The integral which appears in formula of τ_C can be interpreted as the expected value of the function $C(U, V)$ of uniform $(0, 1)$ random variables U and V whose joint distribution function is the copula C:

$$\tau_C = 4E(C(U, V)) - 1 = 4\int_0^1 t dF_C(t) - 1 = 3 - 4\int_0^1 F_C(t)dt,$$

where F_C denotes the distribution function of the random variable $C(U, V)$.

Example: When C is an Archimedean copula with additive generator ϕ, $F_C(t) = t - \phi(t)/\phi'(t^+)$, and thus

$$\tau_C = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)} dt.$$

Spearman's rho

Spearman's rho: Let $(X_1, Y_1), (X_2, Y_2)$, and (X_3, Y_3) be independent random vectors with a common joint distribution function H (whose margins are F and G), and with copula C. Then the population version of Spearman's rho is defined as the difference between probabilities of concordance and discordance of the vectors (X_1, Y_1) and (X_2, Y_3)

$$\rho = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]).$$

Since the copula of (X_1, Y_1) is C and the copula of (X_2, Y_3) is Π, we have

$$\rho(X, Y) = 3K(C, \Pi) = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3,$$

where $\Pi(u, v) = uv$.

- The following theorem is relationship between Spearman's rho and Pearson's correlation.

Theorem

Let (X, Y) be a vector of continuous random variables, $X \sim F(x), Y \sim G(y), U =^d F(X) \sim U(0,1)$ and $V =^d G(Y) \sim U(0,1)$. Then

$$\rho(X, Y) = \frac{Cov(U, V)}{\sqrt{Var(U)Var(V)}} = Corr(F(X), G(Y)).$$

Theorem

Let (X, Y) be a vector of continuous random variables with copula $C(u, v)$. Then the measure Spearman's rho for (X, Y) is given by

$$\rho(X, Y) = 12 \int_0^1 \int_0^1 [C(u, v) - uv] dudv.$$

- This result provides a geometric interpretation for the coefficient $\rho(X, Y)$: it is proportional to the volume between the surfaces of copula $C(u, v)$ and independence copula $\Pi(u, v) = uv$.

Gini's coefficient of cograduation

- The population version of Gini's measure, for random variables X and Y with copula C, is given by

$$\gamma = 2\int_0^1\int_0^1 (|u+v-1| - |u-v|)dC(u,v).$$

- This measure, like Kendall's tau and Spearman's rho, can also be expressed in terms of the concordance function K:

$$\gamma_{X,Y} = \gamma(C) = K(C,M) + K(C,W).$$

Comments

The measures of Kendall, Spearman and Pearson

- The classical non-parametric Kendall's tau and Spearman's rho are **preferable** dependence measures than $Corr(X, Y)$, since they are **invariant under increasing variable** transformations (since the corresponding copula is invariant);
- If X and Y are continuous random variables with copula $C(u, v)$, then

$$C(u, v) = M(u, v) = \min(u, v) \Leftrightarrow \tau(X, Y) = \rho(X, Y) = 1;$$

$$C(u, v) = W(u, v) = \max(u + v - 1, 0) \Leftrightarrow \tau(X, Y) = \rho(X, Y) = -1$$

Tail dependence

- The concept of tail dependence relates to the amount of dependence in the upper-right-quadrant tail or lower-left-quadrant tail of a bivariate distribution.
- The tail dependence coefficient is roughly speaking the probability that a random variable exceeds a certain threshold given that another random variable has already exceeded that threshold.

Definition

Let (X, Y) be a two dimensional random vector with marginal distribution functions F_X and G_Y. The coefficient of upper tail dependence of (X, Y) is defined as:

$$\lambda_U = \lim_{v \nearrow 1} P\{X > F_X^{-1}(v) | Y > G_Y^{-1}(v)\}$$

Definition

Analogously, the coefficient of lower tail dependence of (X, Y) is defined as:

$$\lambda_L = \lim_{v \searrow 0} P\{X \leq F_X^{-1}(v) | Y \leq G_Y^{-1}(v)\}$$

Definition

We say that (X, Y) is upper (lower) tail dependent if and only if $\lambda_U > 0 \;\; (\lambda_L > 0)$. If $\lambda_U = 0 \;\; (\lambda_L = 0)$ we say that (X, Y) is upper (lower) tail-independent.

Theorem

The coefficient of upper tail dependence can be written in terms of copula:

$$\lambda_U = \lim_{v \nearrow 1} \frac{1 - 2v + C(v, v)}{1 - v}$$

where C is the copula of (X, Y). Analogously, we have

$$\lambda_L = \lim_{v \searrow 0} \frac{C(v, v)}{v}$$

Example 1: For the Gumbel copula, we have

$$C(v, v) = \exp[-\{(-\log v)^\theta + (-\log v)^\theta\}^{\frac{1}{\theta}}] = \exp(2^{\frac{1}{\theta}} \log v) = v^{2^{\frac{1}{\theta}}},$$

and therefore,

$$\lambda_U = 2 - 2^{\frac{1}{\theta}}, \quad \lambda_L = 0.$$

Example 2: For the Clayton's copula, we have

$$C(v, v) = (v^{-\theta} + v^{-\theta} - 1)^{-\theta^{-1}} = (2v^{-\theta} - 1)^{-\theta^{-1}},$$

and therefore,

$$\lambda_U = 0, \quad \lambda_L = 2^{-\frac{1}{\theta}}.$$

This means that a Gumbel copula is able to model upper, whereas a Clayton copula can model lower tail dependence.

References

Roger B. Nelsen (Second edition)
An introduction to copulas
Springer, 2006.

Thank You