Parameterization of copula functions for bivariate survival data in the `surrosurv` package. Modelling and simulation

Federico Rotolo

July 13, 2017

Let define the joint survival function of S and T via a copula function:

$$S(s,t) = P(S > s, T > t) = C(u,v)|_{u=S_S(s),v=S_T(t)}\,, \tag{1}$$

where $S_S(\cdot) = P(S > s)$ and $S_T(\cdot) = P(T > t)$ are the marginal survival functions of S and T.

Modelling

In the case of possibly right-censored data, the individual contribution to the likelihood is

- $S(s,t) = C(u,v)|_{S_S(s),S_T(t)}$ if S is censored at time s and T is censored at time t,
- $-\frac{\partial}{\partial t}S(s,t) = \frac{\partial}{\partial v}C(u,v)\big|_{S_S(s),S_T(t)} f_T(t)$ if S is censored at time s and $T = t$,
- $-\frac{\partial}{\partial s}S(s,t) = \frac{\partial}{\partial v}\,C(u,v)|_{S_S(s),S_T(t)} f_S(s)$ if $S = s$ and T is censored at time t,
- $\frac{\partial^2}{\partial s \partial t}S(s,t) = \frac{\partial^2}{\partial u \partial v}C(u,v)\Big|_{S_S(s),S_T(t)} f_S(s)f_t(t)$ if $S = s$ and $T = t$.

Clayton copula

The bivariate Clayton [1978] copula is defined as

$$C(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}, \qquad \theta > 0. \tag{2}$$

The first derivative with respect to u is

$$\begin{aligned}\frac{\partial}{\partial u}C(u,v) &= \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1+\theta}{\theta}} u^{-(1+\theta)} \\ &= \left[\frac{C(u,v)}{u}\right]^{1+\theta}.\end{aligned} \tag{3}$$

The second derivative with respect to u and v is

$$\frac{\partial^2}{\partial u \partial v}C(u,v) = (1+\theta)\frac{C(u,v)^{1+2\theta}}{(uv)^{1+\theta}}. \tag{4}$$

The Kendall [1938]'s tau for the Clayton copula is

$$\tau = \frac{\theta}{\theta+2}. \tag{5}$$

Plackett copula

The bivariate Plackett [1965] copula is defined as

$$C(u,v) = \frac{\left[Q - R^{1/2}\right]}{2(\theta-1)}, \qquad \theta > 0, \tag{6}$$

with

$$\begin{aligned} Q &= 1 + (\theta-1)(u+v), \\ R &= Q^2 - 4\theta(\theta-1)uv. \end{aligned} \tag{7}$$

Given that

$$\frac{\partial}{\partial u} Q = \theta - 1, \tag{8}$$

$$\begin{aligned} \frac{\partial}{\partial u} R &= 2(\theta-1)\Big(1 - (\theta+1)v + (\theta-1)u\Big) \\ &= 2(\theta-1)(Q - 2\theta v), \end{aligned} \tag{9}$$

the first derivative of $C(u,v)$ with respect to u is

$$\begin{aligned} \frac{\partial}{\partial u} C(u,v) &= \frac{1}{2}\left[1 - \frac{1-(\theta+1)v+(\theta-1)u}{R^{1/2}}\right] \\ &= \frac{1}{2}\left[1 - \frac{Q-2\theta v}{R^{1/2}}\right]. \end{aligned} \tag{10}$$

By defining

$$f = Q - 2\theta v, \tag{11}$$

$$g = R^{1/2} \tag{12}$$

and given that

$$f' = \frac{\partial}{\partial v} f = -(\theta+1), \tag{13}$$

$$\begin{aligned} g' = \frac{\partial}{\partial v} g &= \frac{\theta-1}{R^{1/2}}\Big(1 - (\theta+1)u + (\theta-1)v\Big) \\ &= \frac{\theta-1}{R^{1/2}}\Big(Q - 2\theta u\Big), \end{aligned} \tag{14}$$

then, the second derivative with respect to u and v is

$$\begin{aligned} \frac{\partial^2}{\partial u \partial v} C(u,v) &= -\frac{f'g - fg'}{2g^2} \\ &= \frac{\theta}{R^{3/2}}\Big[1 + (\theta-1)(u+v-2uv)\Big] \\ &= \frac{\theta}{R^{3/2}}\Big[Q - 2(\theta-1)uv\Big]. \end{aligned} \tag{15}$$

The Kendall's tau for the Plackett copula cannot be computed analyticaly and is obtained numerically.

Gumbel–Hougaard copula

The bivariate Gumbel [1960]–Hougaard [1986] copula is defined as

$$C(u,v) = \exp\Big(-Q^{\theta}\Big), \qquad \theta \in (0,1), \tag{16}$$

with

$$Q = (-\ln u)^{1/\theta} + (-\ln v)^{1/\theta}. \tag{17}$$

Given that

$$\frac{\partial}{\partial u} Q = -\frac{(-\ln u)^{1/\theta - 1}}{\theta u}, \tag{18}$$

then, the first derivative with respect to u is

$$\frac{\partial}{\partial u} C(u, v) = \frac{(-\ln u)^{1/\theta - 1}}{u} C(u, v) Q^{\theta - 1} \tag{19}$$

and the second derivative with respect to u and v is

$$\frac{\partial^2}{\partial u \partial v} C(u, v) = \frac{\Big[(-\ln u)(-\ln v)\Big]^{1/\theta - 1}}{uv} C(u, v) Q^{\theta - 2} \left[\frac{1}{\theta} - 1 + Q^{\theta}\right]. \tag{20}$$

The Kendall's tau for the Gumbel–Hougaard copula is

$$\tau = 1 - \theta. \tag{21}$$

Simulation

Clayton copula

The function `simData.cc()` generates data from a Clayton copula model. First, the time value for the surrogate endpoint S is generated from its (exponential) marginal survival function:

$$S = -\log(U_S/\lambda_S), \qquad \text{with } U_S := S_S(S) \sim U(0, 1). \tag{22}$$

Then, the time value for the true endpoint T is generated conditionally on the value s of S. The conditional survival function of $T \mid S$ is

$$S_{T|S}(t \mid s) = \frac{-\frac{\partial}{\partial s} S(s, t)}{-\frac{\partial}{\partial s} S(s, 0)} = \frac{\frac{\partial}{\partial u} C(u, v)}{\frac{\partial}{\partial u} C(u, 1)} \tag{23}$$

As the Clayton copula is used, we get (see Equation 3)

$$\begin{aligned} S_{T|S}(t \mid s) &= \left[\frac{C(S_S(s), S_T(t))}{C(S_S(s), 1)}\right]^{1+\theta} = \left[\frac{U_S^{-\theta} + S_T(t)^{-\theta} - 1}{U_S^{-\theta}}\right]^{-\frac{1+\theta}{\theta}} \\ &= \Big[1 + U_S^{\theta}(S_T(t)^{-\theta} - 1)\Big]^{-\frac{1+\theta}{\theta}} \end{aligned} \tag{24}$$

By generating uniform random values for $U_T := S_{T|S}(T \mid s) \sim U(0, 1)$, the values for $T \mid S$ are obtained as follows:

$$\begin{aligned} U_T &= \Big[1 + U_S^{\theta}(S_T(T)^{-\theta} - 1)\Big]^{-\frac{1+\theta}{\theta}} \\ S_T(T) &= \left[\left(U_T^{-\frac{\theta}{1+\theta}} - 1\right) U_S^{-\theta} + 1\right]^{-1/\theta} \\ T &= -\log(S_T(T)/\lambda_T). \end{aligned} \tag{25}$$

Gumbel–Hougaard copula

The function `simData.gh()` generates data from a Gumbel-Hougaard copula model. First, the time value for the surrogate endpoint S is generated from its (exponential) marginal survival function:

$$S = -\log(U_S/\lambda_S), \qquad \text{with } U_S := S_S(S) \sim U(0, 1). \tag{26}$$

The conditional survival function of $T \mid S$ is (see Equation 19)

$$\begin{aligned} S_{T|S}(t \mid s) &= \exp\left(Q(S_S(s),1)^\theta - Q(S_S(s),S_T(t))^\theta\right)\left[\frac{Q(S_S(s),S_T(t))}{Q(S_S(s),1)}\right]^{\theta-1} \\ &= \exp\left(-\log U_S - \left[(-\log U_S)^{\frac{1}{\theta}} + (-\log S_T(t))^{\frac{1}{\theta}}\right]^\theta\right)\left[1+\left(\frac{\log S_T(t)}{\log U_S}\right)^{\frac{1}{\theta}}\right]^{\theta-1} \end{aligned} \tag{27}$$

By generating uniform random values for $U_T := S_{T|S}(T \mid s) \sim U(0,1)$, the values of $S_T(T)$ are obtained by numerically solving

$$U_T - \exp\left(-\log U_S - \left[(-\log U_S)^{\frac{1}{\theta}} + (-\log S_T(T))^{\frac{1}{\theta}}\right]^\theta\right)\left[1+\left(\frac{\log S_T(T)}{\log U_S}\right)^{\frac{1}{\theta}}\right]^{\theta-1} = 0 \tag{28}$$

and then the times $T \mid S$ are

$$T = -\log(S_T(T)/\lambda_T). \tag{29}$$

References

D G Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151, 1978. doi:10.1093/biomet/65.1.141.

E J Gumbel. Distributions des valeurs extrêmes en plusieurs dimensions. *Publ Inst Statist Univ Paris*, 9:171–3, 1960.

P Hougaard. A class of multivariate failure time distributions. *Biometrika*, 73:671–678, 1986. doi:10.1093/biomet/73.3.671.

M G Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938. URL `http://www.jstor.org/stable/2332226`.

R B Nelsen. *An introduction to copulas*. Springer, New York, NY, 2006. doi:10.1007/0-387-28678-0.

R L Plackett. A class of bivariate distributions. *Journal of the American Statistical Association*, 60:516–522, 1965. doi:10.1080/01621459.1965.10480807.