Frailty model and dependence structure for bivariate survival data

Nesrine Idiou [1,*], Fatah Benatia [2] & Mounir Mesbah [3]

[1] Salah Boubnider University of Constantine 3, Department of Process Engineering, Algeria
[2] Mohamed Khider University of Biskra, Department of Mathematics, Algeria
[3] University of Sorbonne, Department of Theoretical and Applied Statistics Laboratory, France

ARTICLE TYPE
Original scientific paper

ARTICLE INFO
Received: June 24, 2025
Accepted: September 9, 2025
DOI: 10.62366/crebss.2025.2.001
JEL: C14, C41, C51

SUMMARY

Copulas and their uses in statistics, namely biostatistics, are a relatively new field of research. When modeling the dependence structure between a vector random variable's joint distribution and marginal distributions, copulas are crucial. In this article, we discuss recent research on this theory and how it can potentially used to analyze multivariate survival data. Lastly, an example of such a method is shown using an application on bivariate survival data from research that was examined using the SAS software's Proc Copula approach.

KEYWORDS

Archimedian copulas, dependence, frailty model, survival analysis

1. Introduction and research problem

Recently, the problem of inference on copulas has attracted an abundance of attention. The Latin word "copŭlae", which means connection, relationship, or union, is whence the word "copula" originated. Among the most significant statistical works in copula theory are those by Hoeffding (1994a) and Hoeffding (1994b), who utilized copulas to study nonparametric measures of association (Spearman's rho, see Section 2). He derived optimal inequalities, providing upper and lower bounds for particular cases of copulas, referred to in the theorem known as the Fréchet's bounds (Fréchet, 1960). The monographs of Deheuvels (1979), Cook and Johnson (1981), Cherubini et al. (2004), Nelsen (2006), Joe (1997), and Genest and Rivest (1993) summarize the results in this field to some extent. Broadly speaking, a copula function is a function that links or couples a multivariate distribution function to its univariate marginal distribution functions, as expressed by Sklar (1959) in the theorem bearing his name. He demonstrated that, under certain conditions, there exists a unique copula function C such that:

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)) \tag{1}$$

*Corresponding author

where C is the associated copula function and $F_1, ..., F_d$ are the margins. This function plays a crucial role in modeling dependence in various fields of statistics (finance, actuarial science, and more recently, biology and health, among others). In this context, and especially in the health domain, copulas are used to study the relationships between the occurrence dates of a disease, possible recovery dates, relapses, and deaths. However, in this aspect, an Archimedean copula is presented by:

$$C(u_1, ..., u_d) = \varphi^{-1}(\varphi(u_1) + ... + \varphi(u_d)), \tag{2}$$

where, φ is a continuous, convex and decreasing function called the generator of C, defined on $I = [0,1] \rightarrow [0,\infty]$ and verifies $\varphi(1) = 0$. Several multivariate survival models accounting for the dependence between random variables are based on the concept of copulas, often without explicitly referencing them (Idiou, 2022). The nature of these problems in survival analysis leads to constructing families of multivariate survival functions from univariate marginal survival functions. The study of survival data it generally refers to the time elapsed between a start date and the occurrence of events, which typically correspond to the onset of a disease, death, relapse, etc. The probability that an individual is alive or free from a condition beyond time t is given by the survival function. When multiple events are considered simultaneously, it is referred to as multivariate survival. A multivariate distribution can be constructed using copulas in the context of survival analysis. Let T be a survival time, i.e., a non-negative random variable (also called a variable of interest) with distribution function F and density f. The survival function, denoted S, at time t, is defined by:

$$S(t) = P[T > t] = 1 - F(t), \quad t \geqslant 0. \tag{3}$$

In the modeling of the survival function, a fundamental concept is that of the hazard rate or hazard function $h(t)$ at a fixed time t, defined by:

$$h(t) = \lim_{\Delta \to 0} \frac{P[t \leq T \leq t + \Delta \mid T \geqslant t]}{\Delta}. \tag{4}$$

It is interpreted as the instantaneous death rate and is expressed as follows: $h(t) = \frac{f(t)}{S(t)}$. The cumulative hazard function $\Lambda(t)$ (Andersen et al., 1993), or integrated hazard function (Hougaard, 1999) is the integral: $\Lambda(t) = \int_0^t h(s)\, ds$. The link between Λ and S is given by: $S(t) = \exp(-\Lambda(t))$, which allows us to write:

$$f(t) = h(t) \exp(-\Lambda(t)) \tag{5}$$

Another important concept in survival modeling is the baseline hazard function $h_0(t)$ (Frees and Valdez, 1998). It is particularly used in the widely employed Cox model (Cox, 1972). The Cox model involves modeling the hazard function $h(t)$ as $h(t) = \exp(X\beta^T)h_0(t)$, where X is a row vector of covariates, and β is the transpose of a vector of parameters associated with these covariates. This regression model allows for the analysis of the effects of covariates on the distribution of lifetime duration. It is a proportional hazards model. It belongs to a broader family of models given by $h(t) = g(X\beta^T)h_0(t)$, where $g(.)$ is some function, which in the Cox model is the exponential function.

One of the advantages of this model is the ease of interpretation of the parameters. Consider the example of a covariate X_j, which can take two values: 0 if the individual receives

treatment A and 1 if they receive treatment B. The coefficient β_j, or rather $exp(\beta_j)$, is the relative instantaneous risk of death for treatment B compared to the reference treatment A.

In this paper, we assume that the survival times are continuous and take values in $\mathbb{R}_+$. Typically, the cumulative distribution function is defined by: $F(t) = \Pr\{T \leq t\}$, which explains the definition of the survival function here as $S(t) = 1 - F(t) = \Pr(T > t)$. However, we can also adopt the definition $S(t) = \Pr(T \geq t)$. The relationship between S and Λ cannot be simply formulated as in the univariate case. For example, in the bivariate case, we get:

$$S(t_1, t_2) = S_1(t_1)S_2(t_2)e^{-\Lambda(t_1,t_2)}. \tag{6}$$

In most applications, we are interested in the lifetime of individuals in a given population, which gives special importance to this copula. Here we define a specific copula that is associated with this survival concept, and we focus primarily on the bivariate case (Figure 1).

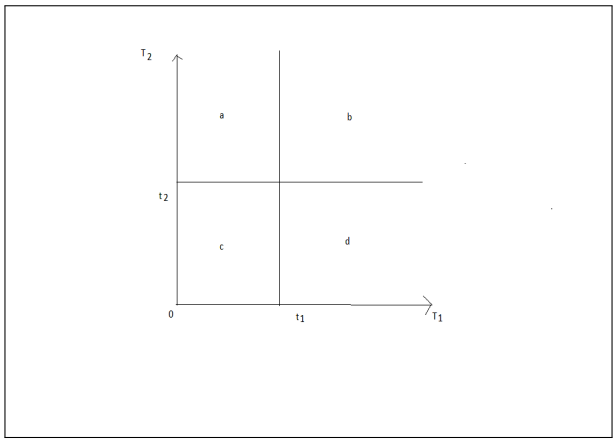

Figure 1. $F_1(t_1) = a + c; F_2(t_2) = c + d; S(t_1, t_2) = b; F(t_1, t_2) = c$

In the univariate case, the survival probability is defined by:

$$S_{X_1}(x_1) = P(X_1 > x_1) = 1 - F_1(x_1), \tag{7}$$

where F_1 is the distribution function of the random variable X_1.

If $F = P(X_1 \leq x_1 \text{ and } X_2 \leq x_2)$ is the joint distribution function of the random pair (X_1, X_2) with marginal distributions F_1 and F_2, and C_{X_1,X_2} is the copula of (X_1, X_2), then the survival function of the pair (X_1, X_2) is given by:

$$\begin{aligned} S(x_1, x_2) &= P(X_1 > x_1, X_2 > x_2) \\ &= S_1(x_1) + S_2(x_2) - 1 + C_{X_1,X_2}(1 - S_1(x_1), 1 - S_2(x_2)), \end{aligned} \tag{8}$$

where S_1 and S_2 are the marginal survival functions of X_1 and X_2, respectively.

Thus, if we define a function $\widetilde{C}$ from $I^2 \to I$ as:

$$\tilde{C}(u, v) = u + v - 1 + C(1 - u, 1 - v), \tag{9}$$

then $S(x_1, x_2) = \widetilde{C}(S_1(x_1), S_2(x_2))$. Note that $\widetilde{C}$ is itself a copula. In the frailty model, the multiplicative conditional hazard is present. The term indicates that, in the context of medical care, one patient may be more fragile than another, putting them at a different risk of passing away or their condition getting worse.

A random parameter (called frailty), potentially shared by a group of patients (group effect), is introduced. The joint survival function is obtained by integrating the conditional survival distribution with respect to the density of this random frailty parameter. The joint survival functions for these models take the form of an Archimedean copula. Based on this observation, it is often stated that the frailty model corresponds to a specific model of Archimedean copulas (Manatunga and Oakes, 1999; Viswanathan and Manatunga, 2001; Andersen, 2005). It is somewhat surprising that these two very different modeling approaches have led to this convergence.

The aim of this paper is to present the "copula" approach in modeling multivariate survival. This approach appears implicitly in Clayton (1978), one of the first to propose a bivariate association model for survival analysis, and subsequently in Marshall and Olkin (1988). In the epidemiological context, frailty models are models that involve an individual–specific parameter. These epidemiological models assume that some subjects may be more frail than others and thus face a higher risk of death or other pathological events. Frailty models use a proportional hazard model.

Vaupel et al. (1979) suggested the application of the gamma model in combination with another frailty model. Gamma distributions are used due to their computational attractiveness, being well known with simple densities. The application of these models in survival analysis is contemporary (Clayton, 1978). The main idea is to introduce dependence between survival times $T_1, ..., T_d$ by using an unobserved random variable W.

The remainder of the paper is structured as follows. In Section 2, we consider the approach of multivariate survival models with frailty variables presented the copulas associated with this type of model. As well as, Section 3, the copula function for the Clayton–Oakes model whiche is the joint survival function of a frailty model and whose Laplace transform is that of a gamma–distributed random variable is presented as an example. However, in Section 4 we apply frailty models to examine hemodialysis data reported by McGilchrist and Aisbett (1991) in application using the Proc Copula procedure in SAS software to analyze real data on recurrent durations in hemodialysis. Our paper ends with some discussions in Section 5. The Appendix interacts with codes in SAS.

2. Main results

The main idea of this section is to introduce the dependence between the survival times $T_1, \ldots, T_d$ using an unobserved random variable W, called frailty. Alternatively stated, it is a latent variable survival model. This model assumes that conditionally on frailty W, with distribution G, the survival times are independent. Thus, the conditional survival function is given by:

$$\begin{aligned} S(t_1, \ldots, t_d|w) =& P[T_1 > t_1, \ldots, T_d > t_d|W = w] \\ =& \prod_{j=1}^{d} P[T_j > t_j|W = w] \\ =& \prod_{j=1}^{d} S_j(t_j|W = w). \end{aligned} \tag{10}$$

The unconditional survival function is given by:

$$S(t_1,\ldots,t_d) = E\left(E(S(t_1,\ldots,t_d|W))\right) = \int S(t_1,\ldots,t_d|w)\, dG(w). \tag{11}$$

To provide an interesting representation of frailty models, we need the theorem of Marshall and Olkin (1988) as follows:

Theorem 2.1 (Marshall and Olkin (1988))**.** *Let $F_1,\ldots,F_d$ be univariate distribution functions, and G be a distribution function in d variables such that $G(0,\ldots,0) = 0$, with univariate marginal functions $G_j, j = 1,\ldots,d$. Let the Laplace transforms of G, denoted φ, and of the marginal distributions G_j, denoted φ_j, be given. Let C be a joint distribution function with all univariate marginal distributions uniform on $[0,1]$. Then, if $H_j(x) = \exp(-\varphi_j^{-1}(F_j(x)))$, we have:*

$$F(x_1,\ldots,x_d) = \int C\left[H_1(x_1)^{w_1},\ldots,H_d(x_d)^{w_d}\right] dG(w_1,\ldots,w_d) \tag{12}$$

which is a distribution function in d variables with marginals $F_1,\ldots,F_d$.

Marshall and Olkin later studied a particularly interesting and simple case of (1) (see Manatunga and Oakes (1999), page 15), where all the marginal functions are identical. The expression (1) then becomes:

$$F(x_1,\ldots,x_d) = \varphi_1\left(\varphi_1^{-1}(F_1(x_1)) + \cdots + \varphi_1^{-1}(F_d(x_d))\right) \tag{13}$$

Thus, it is an Archimedean copula whose generator is the inverse of the Laplace transform. We can now state the definition of survival functions in a frailty model.

Definition 2.2. A survival function is said to be with frailty if it can be written as:

$$S(t_1,\ldots,t_d) = \tilde{C}(S_1(t_1),\ldots,S_d(t_d)) \tag{14}$$

where $\tilde{C}$ is an Archimedean copula whose generator is the inverse of the Laplace transform of the distribution of the frailty variable W. More generally, the generator is the inverse of a Laplace transform.

The following part discusses modeling survival data using two variables. Consider two survival times (T_1, T_2), which might correspond, for example, to two durations for obtaining a diagnosis performed on the same individual using two different techniques, such as T_1 for radiography and T_2 for ultrasound. Let $S_1(t)$ and $S_2(t)$ be the survival functions for each of the methods (Goethals et al., 2008). A frailty model is given by:

$$h_{ij}(t) = w_i h_{j,w_i}(t), \tag{15}$$

where

- $h_{ij}(t)$ is the hazard function at time t for individual $i = 1,\ldots,n$, with diagnostic technique $j = 1, 2$,
- $h_{j,w}(t)$ is the hazard function at time t for an individual with frailty equal to w and diagnostic technique j, and
- w_i is the frailty term for individual i.

To define copula models and frailty models, we need a particular family of Archimedean copulas where the generator φ is the inverse of a Laplace transform. The copula is given by:

$$C(u,v) = \varphi\left\{\varphi^{-1}(u) + \varphi^{-1}(v)\right\}, \quad \text{with} \quad \varphi(0) = 1. \tag{16}$$

Thus, we only need a family of functions $\varphi(.)$. Let $g_W(.)$ be the density of the frailty random variable, defined on the support $[0,\infty)$, and $\varphi_W(s)$ its Laplace transform, i.e.,

$$\varphi_W(s) = E\left\{\exp(-sw)\right\} = \int_0^\infty \exp(-sw) g_W(w)\, dw. \tag{17}$$

Thus, the conditional joint survival function given W is written as:

$$\begin{aligned} S_W(t_1,t_2) =& \tilde{C}(S_{1,W}(t_1), S_{2,W}(t_2)) \\ =& \varphi_W\left\{\varphi_W^{-1}(S_{1,W}(t_1)) + \varphi_W^{-1}(S_{2,W}(t_2))\right\}. \end{aligned} \tag{18}$$

For the frailty model, the conditional survival function is given by:

$$S_{W_i}(t_1,t_2) = \exp\left[-w_i\left\{\Lambda_{1,w_i}(t_1) + \Lambda_{2,w_i}(t_2)\right\}\right], \quad i = 1,\ldots,n, \tag{19}$$

with $\Lambda_{j,w_i}(t) = \int_0^t h_{j,w_i}(s)\, ds$, $j = 1,2$, the cumulative hazard function. Then, integrating with respect to the frailty density, the joint survival function of the model is:

$$\begin{aligned} S(t_1,t_2) =& \int_0^\infty S_W(t_1,t_2) g_W(w)\, dw \\ =& E\left[\exp\left\{-W\left(\Lambda_{1,w}(t_1) + \Lambda_{2,w}(t_2)\right)\right\}\right] \\ =& \varphi\left\{\Lambda_{1,w}(t_1) + \Lambda_{2,w}(t_2)\right\}. \end{aligned} \tag{20}$$

Since the marginal survival function can be written as:

$$S_j(t) = \varphi\left\{\Lambda_{j,W}(t)\right\} \quad \Rightarrow \quad \Lambda_{j,w}(t) = \varphi^{-1}(S_j(t)), \tag{21}$$

we can substitute Equation 21 into Equation 20, and the joint survival function of the frailty model becomes:

$$S(t_1,t_2) = \varphi\left\{\varphi^{-1}(S_1(t)) + \varphi^{-1}(S_2(t))\right\}. \tag{22}$$

From Equation 18 and Equation 22, we observe that the two models are of a different nature because the copula used in the joint survival functions in Equation 18 and Equation 22 is the same, but the marginal survival functions are not the same (Idiou, 2022).

3. Illustrative example

As an example, the copula function for the Clayton–Oakes model is the joint survival function of a frailty model whose Laplace transform is that of a gamma–distributed random variable:

$$\varphi_\theta(s) = (1+\theta s)^{-1/\theta}. \tag{23}$$

Consider again the example of durations corresponding to two diagnostic techniques $j = 1,2$. The inverse Laplace transform of a gamma distribution with a single parameter θ is given by:

$$\varphi_\theta^{-1}(s) = \frac{(s^{-\theta} - 1)}{\theta}, \quad \theta \geq 0. \tag{24}$$

Directly applying Equation 18 for the joint survival function of the Clayton–Oakes copula gives:

$$\begin{aligned} S_c(t_1,t_2) =& C_\theta(S_{1,c}(t_1), S_{2,c}(t_2)) \\ =& \left[\{S_{1,c}(t_1)\}^{-\theta} + \{S_{2,c}(t_2)\}^{-\theta} - 1\right]^{-1/\theta}. \end{aligned} \tag{25}$$

The joint survival function for this frailty model becomes:

$$S_m(t_1,t_2) = \left[1 + \theta\left\{\Lambda_{1,u}(t_1) + \Lambda_{2,u}(t_2)\right\}\right]^{-1/\theta}, \tag{26}$$

which allows us to write:

$$\begin{aligned} S_m(t_1,t_2) =& \left\{1 + \left[(S_{1,m}(t_1))^{-\theta} - 1\right] + \left[(S_{2,m}(t_2))^{-\theta} - 1\right]\right\}^{-1/\theta} \\ =& \left\{(S_{1,m}(t_1))^{-\theta} + (S_{2,m}(t_2))^{-\theta} - 1\right\}^{-1/\theta}. \end{aligned} \tag{27}$$

This expression resembles the form of the copula represented earlier, but $S_{j,m}(t) \neq S_{j,c}(t)$ for all $j = 1, 2$.

4. Application to hemodialysis data

In this part, we apply frailty models to examine hemodialysis data reported by McGilchrist and Aisbett (1991). The observation of recurring infections in hemodialysis patients with kidney failure that is, the purification of blood toxins produced by the body and their removal by an artificial filter was the driving force for this investigation. The catheter site, a hollow plastic tube placed into a vein for hemodialysis, is susceptible to infection. The catheter is taken out and the infection is treated if an infection is noticed.

For the subsequent hemodialysis, a fresh catheter is placed, and so on. The intervals between each catheter insertion and the ensuing infection are noted for every patient.For each patient, the durations between each catheter insertion date and the subsequent infection are observed. Only two observations per patient are considered. There may also be censoring if, during the study period, one or both infections did not occur.

We will take into account all durations as observed and will not discuss the filtered case here. McGilchrist and Aisbett (1991) use a frailty model to address the censored case. In order to achieve this, we estimate the association parameter using the new SAS method (Proc Copula) and select the copula that best fits the data. The analysis of these bivariate data using a copula model, in the presence of censoring, will be the subject of future work.

The histograms in Figure 2, are those of the marginal durations of the two recurrence times. The continuous curves represent the fitted theoretical densities, assumed to be Weibull. A good fit is observed.

Figure 3, shows the empirical marginal distribution functions corresponding to recurrences 1 and 2. The continuous lines correspond to the Weibull distribution functions fitted from the estimated parameters.

The estimates of the Pearson, Spearman, and Kendall (tau) correlation coefficients are 0.07522 (with a significance level p=0.6535), 0.01040 (p=0.9506), and 0.01004 (p=0.9298), respectively. The parameter θ for the Clayton copula is estimated to be 1.054×10^{-7}. Figure 4, shows the scatter plots ((t_1, t_2) and (t_2, t_1)) (a) as well as the level curves of the bivariate distribution (b) and the surface corresponding to this distribution (c), estimated using a Clayton copula. Frank and Gumbel copulas can be obtained in the same way.

a. Recurrence 1 b. Recurrence 2

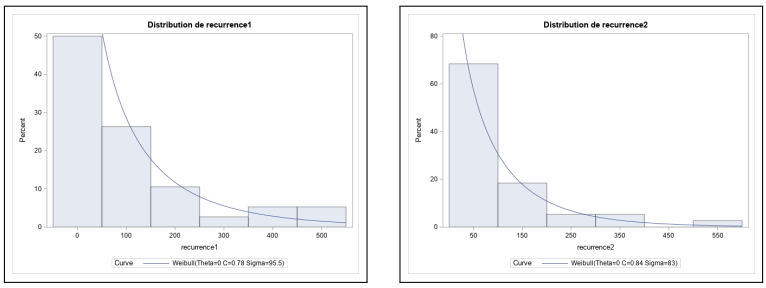

Figure 2. *Empirical histograms and fitted densities (Weibull) for the two recurrence times*

a. Recurrence 1 b. Recurrence 2

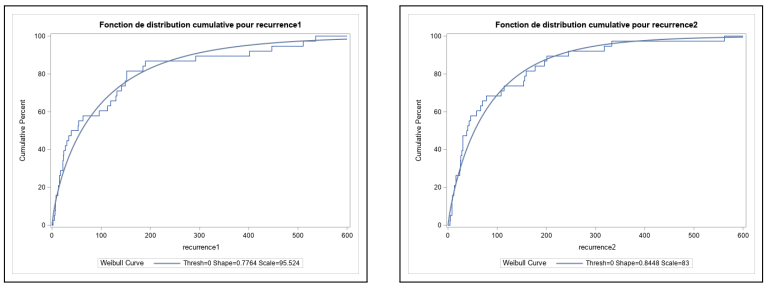

Figure 3. *Empirical distribution functions and their fitted estimates (Weibull) for the two recurrence times*

a. Scatter plot b. Clayton copula: contours c. Clayton copula: surface

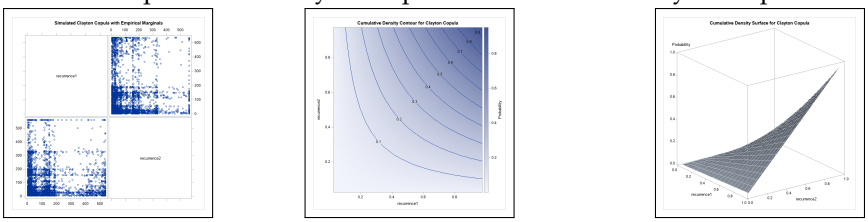

Figure 4. *Empirical bivariate distribution and associated graphs for the Clayton copula*

Table 1 summarizes the results for the estimated association parameter θ under different copula model choices (Clayton, Frank, and Gumbel). The Akaike Information Criterion (AIC) is used to select the model that best fits the data (the smallest AIC value). The choice of the Gumbel copula appears to be the best.

The two recurrence times are weakly correlated. These bivariate data could be analyzed as univariate data. Here, we have presented a brief analysis of the association between the two recurrence times. We have not analyzed the effects of covariates. It would be interesting to compare the analysis of the effects of these covariates by considering the data as univariate versus accounting for the bivariate nature modeled by a Gumbel copula, chosen here by the AIC. It would also be interesting to compare the effects of covariates with those obtained from a Cox regression model with frailty. The SAS codes used are provided in the Appendix.

Table 1. *Estimation of the association parameter for three copulas under Akaike's information criteria*

Copula	θ	Std. Error	AIC
Clayton	1.054×10^{-7}	0.000000	2.00000
Frank	0.150408	1.031073	1.97873
Gumbel	1.039736	0.127503	1.89611

5. Conclusion and perspectives

An additional tool often used for modeling multivariate survival data is the introduction of individual random parameters, often interpreted as frailty parameters. In this work, we used this model for bivariate survival data by considering Archimedean copulas. We specifically focused on the cases of Clayton–Oakes copulas and the gamma–type frailty model. For each of these two models, the copulas used for the bivariate survival functions are the same. However, the marginal survival functions are modeled differently.

We then turned to applications for health–related survival data. Survival data can also originate from reliability studies in the industrial field. A significant challenge in analyzing this type of data, including the univariate case, is the presence of censoring. In the bivariate case, this question remains largely unresolved. The analysis of bivariate and censored survival data using copulas is the subject of ongoing work.

This work also does not address statistical inference issues, which are often complex, especially in the presence of censoring. Recent research in the field of copulas is more often statistical in nature than theoretical. Manatunga and Oakes (1999), Deheuvels (1979), Genest (1987), Genest and Rivest (1993) are excellent references on this subject.

Survival data, generally of a continuous quantitative nature, are often used in the medical field for diagnostic or prognostic purposes. Qualitative and discrete data are also frequently encountered (Frees and Valdez, 1998). This paper opens up several perspectives for future work and the development of copula–based models for this type of data.

In the field of copulas, research has also focused on computational aspects, with the publication of programs in the R language and, more recently, SAS with Proc Copula. We used Proc Copula to process hemodialysis data. Unfortunately, this procedure does not support censored data. A useful future project would be to write a SAS macro addressing this limitation.

The frailty variables considered here are latent, unobserved, and one dimensional. In the presented example, this variable characterized the individual's effect on recurrence time. These individuals could originate from several hospital centers. The differential, unobserved effect of these centers would then be a latent variable. A perspective for future research would be the modeling and analysis of multivariate lifetimes with latent variables that are themselves multivariate using copulas. The difficulty in processing such data would primarily be computational.

Acknowledgment

The authors thank Prof. R. Senoussi for his careful reading of this article. His remarks, and the multiple discussions which followed, made it possible to significantly improve a previous version of this item.

Statement on the use of artificial intelligence

No artificial intelligence (AI) tools were used in the preparation of this paper, and the authors are fully responsible for the content, analyses, interpretation of results, and conclusions.

References

Andersen, E. W. (2005). Two–stage estimation in copula models used in family studies. *Lifetime Data Analysis*, 11(3), 333–350. doi:10.1007/s10985-005-2966-7

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993). *Statistical models based on counting processes*. Springer Series in Statistics, Springer: New York. doi: 10.1007/978-1-4612-4348-9

Cherubini, U., Luciano, E. & Vecchiato, W. (2004). *Copula methods in finance*. Wiley Finance Series, John Wiley and Sons: Chichester. doi: 10.1002/9781118673331

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1), 141–151. doi:10.1093/biomet/65.1.141

Cook, R. D. & Johnson, M. E. (1981). A family of distributions for modelling non–elliptically symmetric multivariate data. *Journal of the Royal Statistical Society*, Series B, 43(2), 210–218. doi:10.1111/j.2517-6161.1981.tb01173.x

Cox, D. R. (1972). Regression models and life–tables (with discussion). *Journal of the Royal Statistical Society*, Series B, 34(2), 187–220. doi: 10.1111/j.2517-6161.1972.tb00899.x

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Bulletin de la Classe des Sciences de l'Académie royale de Belgique*, 65(6), 274–292. doi: 10.3406/barb.1979.58521

Frees, E. W. & Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2(1), 1–25. doi: 10.1080/10920277.1998.10595667

Fréchet, M. (1960). Sur les tableaux dont les marges et des bornes sont données. *Review of the International Statistical Institute*, 28(1/2), 10–32. doi: 10.2307/1401846

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika*, 74(3), 549–555. doi: 10.1093/biomet/74.3.549

Genest, C. & Rivest, L.–P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88(423), 1034–1043. doi: 10.1080/01621459.1993.10476372

Goethals, K., Janssen, P. & Duchateau, L. (2008). Frailty models and copulas: Similarities and differences. *Journal of Applied Statistics*, 35(9), 1071–1079. doi: 10.1080/02664760802271389

Hoeffding, W. (1994). *Scale–invariant correlation theory*. In: Fisher, N. I. & Sen, P. K. (Eds.), The Collected Works of Wassily Hoeffding, Springer Series in Statistics, Springer: New York, 57–107. doi: 10.1007/978-1-4612-0865-5_4

Hoeffding, W. (1994). *Scale–invariant correlation measures for discontinuous distributions*. In: Fisher, N. I. & Sen, P. K. (Eds.), The Collected Works of Wassily Hoeffding, Springer Series in Statistics, Springer: New York, 109–133. doi: 10.1007/978-1-4612-0865-5_5

Hougaard, P. (1999). Fundamentals of survival data. *Biometrics*, 55(1), 13–22. doi: 10.1111/j.0006-341X.1999.00013.x

Idiou, N. (2022). *Multi–parametric copula estimation based on moments method under censoring*. Doctoral thesis, University of Biskra. http://thesis.univ-biskra.dz/5685/

Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall: London. doi: 10.1201/9781482296426

Kendall, M.G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2), 81–89. doi: 10.2307/2332226

Manatunga, A. K. & Oakes, D. (1999). Parametric analysis for matched pair survival data. *Lifetime Data Analysis*, 5(4), 371–387. doi: 10.1023/A:1009692210273

Marshall, A. W. & Olkin, I. (1988). Families of multivariate distributions. *Journal of the American Statistical Association*, 83(403), 834–841. doi: 10.1080/01621459.1988.10478671

McGilchrist, C. A. & Aisbett, C.W. (1991). Regression with frailty in survival analysis. *Biometrics*, 47(2), 461–466. doi: 10.2307/2532138

Nelsen, R. B. (2006). *An introduction to copulas* (2nd ed.). Springer: New York. doi: 10.1007/0-387-28678-0

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8, 229–231. https://cir.nii.ac.jp/crid-1573387449735953792?lang=e

Shaked, M. & Shanthikumar, J. G. (1987). The multivariate hazard construction. *Stochastic Processes and their Applications*, 24(2), 241–258. doi: 10.1016/0304-4149(87)90015-9

Vaupel, J. W., Manton, K. G. & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3), 439–454. doi: 10.2307/2061224

Viswanathan, B. & Manatunga, A. K. (2001). Diagnostic plots for assessing the frailty distribution. *Lifetime Data Analysis*, 7, 143–155. doi: 10.1023/A:1011348823081

Yan, J. (2006). *Multivariate modeling with copulas and engineering applications*. In: H. Pham (Eds.), Springer Handbook of Engineering Statistics, 973–990. Springer: London. doi: 10.1007/978-1-84628-288-1_51

Appendix

SAS program: Proc Copula applied to recurrent hemodialysis data

```
proc univariate data=recurrences; var recurrence1 recurrence2;
histogram recurrence1 recurrence2/weibull;run;
proc univariate data=recurrences;var recurrence1 recurrence2;
cdfplot recurrence1 recurrence2/weibull;run;
proc corr data=recurrences kendall pearson spearman;
var recurrence1 recurrence2;run;
proc copula data=recurrences;var recurrence1 recurrence2;
fit clayton/marginals=empirical;
simulate /ndraws = 5000 seed = 12345678 marginals=empirical
plots = (distribution=cdf) out = fic1;run;
proc copula data=recurrences; var recurrence1 recurrence2;
fit frank/marginals=empirical;
simulate /ndraws = 5000 seed = 12345678 marginals=empirical
plots = (distribution=cdf) out = fic1; run;
proc copula data=recurrences; var recurrence1 recurrence2;
fit gumbel/marginals=empirical;
simulate /ndraws = 5000 seed = 12345678 marginals=empirical
plots = (distribution=cdf) out = fic1; run;
```

Model krhkosti i struktura ovisnosti za bivarijatne podatke o preživljenju

VRSTA ČLANKA
Izvorni znanstveni članak

INFORMACIJE O ČLANKU
Primljeno: 24. lipnja 2025.
Prihvaćeno: 9. rujna 2025.
DOI: 10.62366/crebss.2025.2.001
JEL: C14, C41, C51

SAŽETAK

Kopula funkcije i njihova primjena u statistici, osobito u biostatistici, predstavljaju relativno novo istraživačko područje. Pri modeliranju strukture ovisnosti između zajedničke distribucije vektora slučajnih varijabli i njihovih graničnih distribucija, kopule imaju ključnu ulogu. U ovom članku raspravlja se o recentnim istraživanjima iz tog područja te o njihovom potencijalu za analizu multivarijatnih podataka o preživljenju. Na kraju je prikazan primjer takve metode primjenom na bivarijatne podatke o preživljenju iz istraživanja koje je analizirano korištenjem Proc Copula postupka u SAS programu.

KLJUČNE RIJEČI

Arhimedove kopule, ovisnost, model krhkosti, analiza preživljenja