A bivariate copula model for the failure of electrical cable insulation

Mariana Craiu and Romeo Bercia

Abstract. In this paper we propose a joint distribution function for modeling a process of electrical treeing using a Farlie-Gumbel Morgenstern copula.

M.S.C. 2000: 62H12.
Key words: copula, empirical distributions, bivariate distribution, semi-parametric method, electrical treeing.

1 Introduction

The theory of copulas is used to build multivariate distribution laws. It is a strong emphasis on random-effects models for multivariate lifetime data and a close connection between random-effects models and copula models. The books by Kotz (2000) and Joe (1997) present extensive coverage of multivariate distributions and concepts of association. Statistical methods for multivariate distributions are discussed by Clayton, Hougaard and Joe.

The dependence properties of a continuous multivariate distribution is defined by a copula that is associated with the distribution. One attractive feature of a copula model is that margins do not depend on the choice of the dependency structure.

Definition. ([1]). A two dimensional copula C is a function with the properties:
1) $C : [0,1] \times [0,1] \to [0,1]$
2) $C(u,1) = C(1,u) = u$, $C(u,0) = C(0,u) = 0$, $\forall u \in [0,1]$
3) $C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$, $\forall u_1 \leq u_2$, $\forall v_1 \leq v_2$.
The connection between a copula and a bivariate distribution is given by Sklar [1].

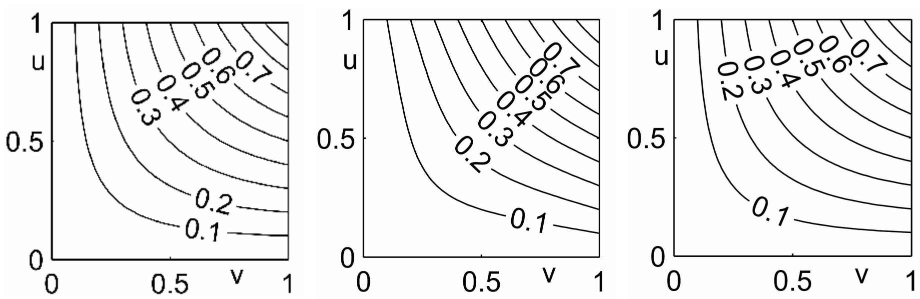
Theorem. *If C is a copula and F, G are univariate distribution functions, then the function $H_{(X,Y)}(x,y) = C(F_X(x), G_Y(y))$ is a joint distribution function with margins F and G. Otherwise if $H_{(X,Y)}$ is a joint distribution function with margins F and G, there exists a copula C such that $H(x,y) = C(F(x), G(y))$ for all $x, y \in (-\infty, \infty)$.*

In this paper we want to determine the joint distribution function for a process of electrical treeing. For this we'll choose a model using the *best* copula between: Farlie-Gumbel Morgenstern copulas (1.1) that contains copulas with both horizontal and

Proceedings of The 4-th International Colloquium "Mathematics in Engineering and Numerical Physics" October 6-8 , 2006, Bucharest, Romania, pp. 52-57.

vertical quadratic sections, a particular case of Farlie's family (1.2) and the copula (1.3) of Lu and Bhattacharyya (1990)

$$\begin{aligned}
&(1.1) \qquad C_\theta(u,v) &=& \quad uv + \theta uv(1-u)(1-v), \forall\theta \in [-1,1]. \\
&(1.2) \qquad C_\alpha(u,v) &=& \quad uv + \alpha uv(1-3u+2u^2)(1-3v+2v^2), \forall\alpha \in [-1,1]. \\
&(1.3) \qquad C_\delta(u,v) &=& \quad u+v-1+(1-u)(1-v)\mathrm{e}^{-\delta uv}, \forall\delta \in [-1,1].
\end{aligned}$$



a) b) c)

Figure 1: Contour plots for copulas (1.1)-(1.3): a) C_θ for $\theta = 0.8334$, b) C_α for $\alpha = 0.9793$, c) C_δ for $\delta = -0.6583$.

2 Some dependence measures

The dependence measures between the variables T_1 and T_2 are defined by the copulas.

a) σ is a measure of the *average distance* between the distributions of T_1 and T_2 and independence among T_1 and T_2

$$\sigma = \int_{[0,1]^2} |C(u,v) - uv| \mathrm{d}u\mathrm{d}v \tag{2.1}$$

b) The Kendall-tau τ is a measure for the probability of concordance minus the probability of discordance for a pair of observations from (T_1, T_2).

$$\tau = 4 \int_{[0,1]^2} C(u,v)\mathrm{d}C(u,v) - 1 \tag{2.2}$$

c) Spearman ρ-*distance* is the distance between the distributions of T_1 and T_2 as represented by their copula C and independence represented by uv.

$$\rho = 12 \int_{[0,1]^2} C(u,v)\mathrm{d}u\mathrm{d}v - 3 \tag{2.3}$$

3 Estimation of Joint Bivariate Distribution

Copula based models are usefull when the association between the variables is important. To estimate the dependence parameter, two strategies can be envisaged:
1) The first is by a two stage parametric estimation when it is assumed that the functional forms of the margins are known and there are a finite number of unknown parameters. This is a parametric estimation procedure. First the parameters of marginals are estimated by the method of likelihood function for the data and at the second stage the copula's parameter is estimated by the same method.
2) The second procedure, developed by Genest [6], is a two stage semi-parametric one. Now, at first stage the marginal distribution functions are estimated by empirical distribution functions and at the second stage the copula's parameters are estimated by the likelihood method. Models in a semi-parametric case are based on parametric copulas but nonparametric marginal distributions. In this case the inference about the dependence parameter is margin-free. This is sensible in applications where the focus of the analysis is on the dependence structure.

In this paper we'll choose the second procedure.

If $C_p(u,v)$ is a class of parametric copulas with unknown parameter p, then an estimator of p is $\hat{p} = \arg\max \frac{1}{n} L_n(p)$, where

$$L_n(p) = \sum_{j=1}^{n} \ln c_p \left(\hat{F}_1(x_{1j}), \hat{F}_2(x_{2j}) \right), \tag{3.1}$$

$c_p(u,v) = \frac{\partial^2 C_p}{\partial u \partial v}(u,v)$ and $\hat{F}_i(x) = \frac{1}{n} \sum\limits_{j=1}^{n} I(X_{ij} \le x)$, $i = 1,2$, are the empirical distributions of marginals. In 1995 Goudi and Rivest have established the asymptotic normality of $\hat{p}$.

Using copulas defined by (1.1)-(1.3) one obtain the following equations

$$\sum_{j=1}^{n} \frac{(1-2u_j)(1-2v_j)}{1+\theta(1-2u_j)(1-2v_j)} = 0, \tag{3.2}$$

$$\sum_{j=1}^{n} \frac{(1-6u_j+6u_j^2)(1-6v_j+v_j^2)}{1+\alpha(1-6u_j+6u_j^2)(1-6v_j+v_j^2)} = 0, \tag{3.3}$$

$$\sum_{j=1}^{n} u_j v_j = \sum_{j=1}^{n} \frac{-1+2u_j+2v_j-3u_jv_j+2\delta u_jv_j(1-u_j)(1-v_j)}{1+\delta(-1+2u_j+2v_j-3u_jv_j)+\delta^2 u_jv_j(1-u_j)(1-v_j)}, \tag{3.4}$$

where $u_j = \hat{F}_1(t_{1j})$ and $v_j = \hat{F}_2(t_{2j})$.

4 Model selection

The various copula models are compared using both information criteria and empirical distance measures.

T_1	228	106	246	700	473	155	414	1374	128
T_2	30	8	66	72	25	7	30	90	4
T_1	1227	254	435	1155	195	117	724	300	
T_2	39	46	85	85	27	27	21	96	

Table 1: Stone's data

$\hat{F}_1(t_{1j})$	1/3	1/18	7/18	13/18	11/18	1/6	5/9	1	1/9
$\hat{F}_2(t_{2j})$	1/2	1/6	11/18	2/3	5/18	1/9	1/2	1/6	1/18
$\hat{F}_1(t_{1j})$	17/18	4/9	11/18	8/9	5/18	1/9	7/9	1/2	
$\hat{F}_2(t_{2j})$	5/9	11/18	5/6	5/6	7/18	7/18	2/9	1	

Table 2: Empirical distribution functions

The Akaike Information Criterion is used because of its optimality properties established by many researchers like Dias and Embrechts in 2004. The expression of this criterion is $AIC = 2r - 2\ln L_n(p)$, where n is the sample size and r the number of estimated parameters. The Bayes Information Criterion $BIC = r\ln n - 2\ln L_n(p)$ penalizes more strictly for over fitting a model. The best model is selected as one that minimizes AIC or BIC.

In [4] the authors propose a pseudo-likelihood ratio test for selecting semiparametric copula models in wich the marginal distributions are unspecified and the copula function is parametrized. By this test, the model with q will be selected if:

$$\sum_{j=1}^{n} \ln \frac{c_p(u_j, v_j)}{c_q(u_j, v_j)} < 0 \tag{4.1}$$

A such test is based on the Akaike Information Criterion.

We apply all these results to analyse a phenomenon known as **electrical treeing**. In the process known as electrical treeing there is an inception of a failure process that at some point eventually determines the failure of the insulation. We shall analize data obtained by Stone [5] from an experiment on the failure of epoxy electrical cable-insulation specimens under conditions of constant voltage stress. We are interested to establish the dependence among these two moments of time T_1 -the moment to detect of inception and T_2 the moment of failure. We model the association of bivariate failure times by copula functions because in a such model the margins do not depend on the choice of the dependency structure.

The Stone's data (see table 1) are from an experiment with 20 items subjected to a 55-voltage stress. The empirical distribution functions are given in table 2. We apply the second method, that is the semi-parametric one, for establish the association between these two variables.

Using equations (3.2)-(3.4) results: $\theta = 0.8334$, $\alpha = 0.9793$ and $\delta = -0.6583$. By the Akaike Information Criterion $AIC(C_\theta) = 0.3874$, $AIC(C_\alpha) = 1.092$, $AIC(C_\delta) = 0.638$, and so the model of C_θ is better than the others. One obtain the same conclusion using the BIC criterion. The other criterion (4.1) based also on the Akaike Information Criterion gives the same result.

The values of dependence measures in these cases are: The sample correlation coefficient is $R = 0.5038$ and so, these two variables are not independent. More information about the dependence is given by $\sigma_\theta = 0.2778 = \rho_\theta$ and $\sigma_\alpha = 0.043$, $\tau_\theta = 0.1852$ and $\tau_\alpha = 0 = \rho_\alpha$. All these values prove that the model given by $C_\theta(u, v)$ is the best.

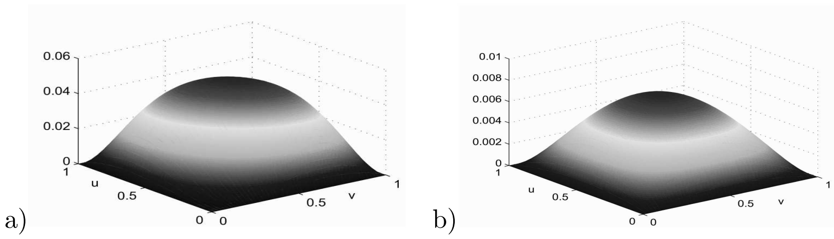


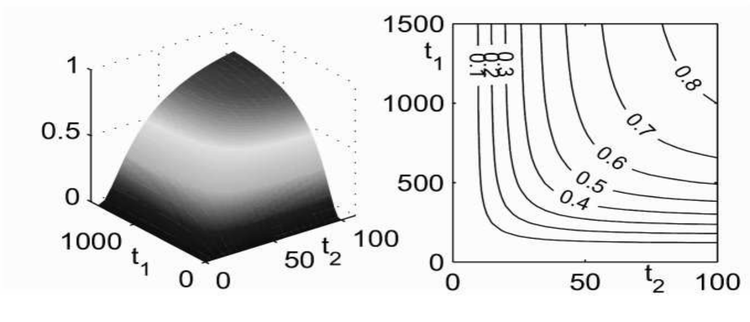
Figure 2: Surface plot of the difference between copulas: a) $C_\theta - C_\alpha$, b) $C_\theta - C_\delta$.

5 Conclusions

From this analysis we can accept that $C_\theta(u, v)$ is a good model for this process, and so, the bivariate distribution function for (T_1, T_2) is

$$H(t_1, t_2) = F(t_1)G(t_2)\left(1 + 0.8334(1 - F(t_1))(1 - G(t_2))\right), \tag{5.1}$$

where F and G are the univariate distributions of T_1 and T_2.



a) b)

Figure 3: The bivariate distribution function H: a) surface plot, b) contour plot.

By the Kolmogorov-Smirnov test we can accept that T_1 and T_2 have lognormal distributions with $\hat{\mu}_1 = 5.8562$, $\hat{\sigma}_1^2 = 0.6952$ and $\hat{\mu}_2 = 3.4692$, $\hat{\sigma}_2^2 = 0.8998$. In Table 3 there are calculated some values for $H(t_1, t_2) = P(T_1 \leq t_1, T_2 \leq t_2)$.

	$t_2 = 20$	$t_2 = 40$	$t_2 = 60$	$t_2 = 80$	$t_2 = 100$
$t_1 = 150$	0.071	0.118	0.136	0.144	0.148
$t_1 = 300$	0.176	0.302	0.357	0.384	0.399
$t_1 = 600$	0.263	0.477	0.583	0.639	0.672
$t_1 = 900$	0.289	0.538	0.667	0.738	0.781
$t_1 = 1200$	0.299	0.563	0.704	0.782	0.829
$t_1 = 1500$	0.303	0.576	0.721	0.803	0.852

Table 3: Some values of the joint distribution function

References

[1] Roger B. Nelsen, *An Introduction to Copulas*, Springer-Verlag, 1998

[2] H. Joe, *Asymptotic efficiency of the two-stage estimation method for Copula-based models*, J. Multivariate Analysis, 2005

[3] C. Amblard, S. Girard, *Symetry and dependence properties within a semiparametric family of bivariate copulas*, CRM-269, Sept.2000, Univ.Montreal

[4] X. Chen, Y. Fan, *Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection*, The Canadian Journal of Statistics, 2005(33), Nr.2,

[5] J. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons, 2003

[6] C. Genest, K. Choudi, L.P. Rivet, *A semiparametric estimation of dependence parameters in multivariate families of distributions*, Biometrika, 1995(82), 3, 543-552

[7] H. Joe, *Multivariate Models and Dependence Concepts*, Chapman and Hall, London, 1997

[8] J.-C. Lu, G.K. Bhattacharyya, *Some new considerations of bivariate Weibull models*, Ann. Inst. Stat. Math, 1990(42), 543-559

Authors' address:

Mariana Craiu and Romeo Bercia
Faculty of Applied Sciences, University Politehnica of Bucharest
313, Splaiul Independentei, 060042 Bucharest, Romania
email: craiuus@yahoo.com, r_bercia@yahoo.com