International Journal of Applied Mathematics
& Statistical Sciences (IJAMSS)
ISSN(P): 2319-3972; ISSN(E): 2319-3980
Vol. 5, Issue 6, Oct - Nov 2016; 61-70

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

MULTIVARIATE LIFETIME DISTRIBUTIONS

SAMIA A. ADHAM

Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

ABSTRACT

The aim of this paper is to introduce and investigate a novel method for constructing multivariate lifetime distributions. The idea is based on the combined use of copula and mixtures. Both have been used on their own for constructing multivariate lifetime distributions, but with only moderate success. Our aim is to show that their joint use possesses some distinct advantages.

KEYWORDS: Bayesian Inference, Dependence, Frank's Copula, Gaussian copula, Gompertz Distribution, Mixture

1. INTRODUCTION

The construction of multivariate lifetime distributions is an important problem and there are a number of methods that have been used successfully in their construction. Some of these methods are discussed in the book of bivariate continuous distributions by Hutchinson and Lai (1990). Two ideas, mixtures (Hougaard, 1986; Crowder, 1989) and copulas (Joe, 1997), have been used in particular. The mixture idea is easy to use, but may not provide adequate dependence structures (see Walker and Stephens, 1999). On the other hand, the copula methods allow flexible models but are hard to study. Song (2000) presents a class of multivariate dispersion models generated from the Gaussian copula and studied some of its properties. Adham and Walker (2001) applied the mixture representation of the Gompertz distribution in order to motivate a new family of distributions which extends naturally to the multivariate case. The goal of this paper is to combine the mixture and copula ideas in order to construct multivariate lifetime distributions which are easy to analyze and allow full dependence structures.

It is well known that any lifetime density function $f_T(t)$ of a non-negative continuous random variable *T*can be written as

$$f_T(t) = h_T(t) S_T(t),$$

Where $h_T(t)$ is the hazard function, $S_T(t)$ is the survival function, which can itself be written as

$$S_T(t) = \exp\{-H_T(t)\},$$

Where

$$H_T(t) = \int_0^t h_T(w)\, dw, \tag{1}$$

Is the cumulative hazard function That is,

$$f_T(t) = h_T(t) \exp\{-H_T(t)\}.$$

The mixture idea of Walker and Stephens (1999) is to introduce a non-negative latent variable *U* which follows a gamma distribution with shape parameter 2 and scale parameter 1, denoted by *G (2, 1)*. The density function can be written as a mixture:

$$f_{T|U}(t|u) = \frac{h_T(t)}{u} I\{u > H_T(t)\},$$
$$f_U(u) = u \exp(-u), \quad u \geq 0.$$

The mixture representation can be written in a more general form follows

$$f_{T|U}(t|u) = \frac{g(t)}{u} I\{u > A(t)\},$$

$U \sim G(2, 1).$

Table 1 provides the functions $g(t)$ and $A(t)$ which can be used to present a mixture from the most used lifetime distributions. The functions $g(t)$ and $A(t)$ are not unique; but one can take $g(t)$ as the hazard function and $A(t)$ as the cumulative hazard function when known in closed form.

Table 1: $g(t)$ and $A(t)$ Functions of the Most Used Lifetime Distributions

Distribution	$g(t)$	$A(t)$
$Exp(c)$	c	$c\,t$
$Weibull(a, c)$	$a\,c\,t^{a-1}$	$c\,t^{a}$
$Gompertz(a, c)$	$a\,c\exp(at)$	$c\{\exp(at)-1\}$
$G(a, c)$	$\frac{c^a}{\Gamma(a)} t^{a-1}$	$c\,t$
$LN(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi}\,\sigma\,t}$	$\frac{1}{2}\left(\frac{\log t-\mu}{\sigma}\right)^2$

Intuitively, it can be seen from Table 1, for exponential, Weibull and Gompertz distributions, the functions $g(t)$ and $A(t)$ are their hazard and cumulative hazard functions, respectively. Here we concentrate on the Gompertz distributions with parameters *a* and *c*, denoted by *Gompertz* (*a*, *c*), where the parameters are assumed to be positive.

The hazard function is

$$H_T(t) = a\,c\exp(at), \quad t > 0, \tag{2}$$

A literature search suggests many bivariate and multivariate lifetime distributions. However some of them are difficult to study. Furthermore, the dependence properties might be unknown or no extension to the multivariate case seems possible. For example, Frees et al. (1996) applied a bivariate Gompertz distribution with Frank's copula. However, there seems to be no available extension of Frank's copula to dimensions higher than 2, except for discrete distributions (Joe, 1993). This is not the story for the Gaussian copula. Carrier (2000) studied a data set from a life annuity portfolio on six different bivariate Gompertz distributions, each of which is based on a specific copula model. Precisely, the Gompertz marginal distributions are coupled with the Frank's, frailty, normal generalized Frank's, linear mixing frailty and correlated frailty copulas. Carrier (2000) found the maximum likelihood estimates for the parameters in the bivariate models. Song (2000) presented a class of multivariate dispersion models generated from the Gaussian copula and studied some of its properties. Adham and Walker (2001) applied the mixture representation of the Gompertz distribution in order to motivate a new family of distributions which extends naturally t the multivariate case. Some other studies were devoted to using copulas in studying multivariate non-normal distributions such as Abd EL Lteef(2005), AL-Hussaini and Ateya (2006), AL-Dayian et al. (2008), Gupta et al. (2010),Meisch et al. (2014), Alai et al. (2015), Alai et al. (2016) among others.

Here we introduce our ideas for the constructions of multivariate lifetime distributions. Specifically, we assume that $T=(T_1,\ldots,T_p)'$ to be a*p*-dimensional non-negative random vector which are conditionally independent given $U=(U_1,\ldots,U_p)'$ i.e.

$$f_{T|U}(t|u)=\prod_{j=1}^{p} f_{T_j|U_j}(t_j|u_j). \tag{3}$$

If we take the U_j's to be independent *G* (*2*, *1*) then the T_j's will be independent. To generate a dependence structure we construct a multivariate distribution for *U* by using a copula, specifically Frank's copula and the Gaussian copula. That is,

$$f_U(u_1,\ldots,u_p)=\frac{\partial^p C(v_1,\ldots,v_p)}{\partial v_1\ldots\partial v_p}\prod_{j=1}^{p} f_{U_j}(u_j), \tag{4}$$

Where, for $j=1,\ldots,p$,$v_j=F_{U_j}(u_j)$ and f_{U_j}, F_{U_j} are density and distribution functions connected with the gamma distribution with shape parameter 2 and scale parameter 1. Here *C* is the copula.

This way we avoid the problems of placing a copula directly onto an unknown distribution with unknown parameters, and instead are placing it onto the distribution with fixed marginals. There is also a nice interpretation: the top level sorts out the marginal aspects. While, the lower level concerns with the dependence structure and the two levelscan be dealt with separately.

Describing the layout of the paper, Section 2 illustrates the construction of the multivariate Gompertz distribution based on Frank's copula, while Section 3 deals with the multivariate Gompertz distribution based on the Gaussian copula. Section 4 studies dependence properties of the new multivariate models.Finally, an illustrative example is presented in Section 5.

2. FRANK'S COPULA

This section provides a multivariate Gompertz distribution based on Frank's copula. Since there seems to be no available extension of Frank's copula to dimension higher than 2, except for discrete distributions (Joe, 1993), we concentrate on $p=2$.

Frank's copula for the distribution functions $\nu_j = F_{U_j}(u_j), \quad j = 1,2,$ is given by

$$C(\nu_1, \nu_2) = \frac{1}{\alpha} \log\left\{1 + \frac{(e^{\alpha\nu_1} - 1)(e^{\alpha\nu_2} - 1)}{e^{\alpha} - 1}\right\}, \quad \alpha \neq 0. \tag{5}$$

The joint distribution function of the mixing (latent) variables U_1 and U_2 can be written as

$$F_U(u_1, u_2) = C(\nu_1, \nu_2),$$

Where $C(\nu_1, \nu_2)$ is given by (5). Given that C and $F_{U_j}(u_j)$ are continuous and differentiable, the bivariate gamma density function of the pair (U_1, U_2) is then given by

$$f_U(u_1, u_2) = f_{U_1}(u_1)\, f_{U_2}(u_2)\, C'(\nu_1, \nu_2), \tag{6}$$

Where $f_{U_j}(u_j)$ is *G* (*2*, *1*) density function and

$$C'(\nu_1, \nu_2) = \frac{\partial^2 C(\nu_1, \nu_2)}{\partial \nu_1 \partial \nu_2} = \frac{\alpha \exp\{\alpha(\nu_1 + \nu_2)\}(e^{\alpha} - 1)}{\{(e^{\alpha\nu_1} - 1)(e^{\alpha\nu_2} - 1) + e^{\alpha} - 1\}^2}, \quad \alpha \neq 0. \tag{7}$$

Consequently, the density function of the bivariate Gompertz distribution with Frank's copula can be written as

$$f_T(t_1, t_2) = \int_{K_1}^{\infty} \int_{K_2}^{\infty} \left(\prod_{j=1}^{2} \frac{a_j\, c_j\, e^{a_j t_j}}{u_j} \right) C'\{F_{U_1}(u_1), F_{U_2}(u_2)\} du_1\, du_2,$$

Where $K_j = c_j \left(e^{a_j t_j} - 1\right), \quad j = 1,2$.

3. GAUSSIAN COPULA

In this section we present a multivariate Gompertz distribution based on the Gaussian copula. Similar to the illustrated multivariate Gompertz distribution based on Frank's copula, provided in the previous section, we use the mixture representation of the univariate Gompertz distribution. By extending the mixing distribution, *G*(*2*, *1*), to multivariate form, but here it is based on the Gaussian copula with joint density function

$$f_U(u) = \left\{ \prod_{j=1}^{p} f_{U_j}(u_j) \right\} C'(\nu_1, \ldots, \nu_p), \tag{8}$$

Where $f_{U_j}(u_j)$ is the density function of *G* (2,1) distribution, $\nu_j = F_{U_j}(u_j)$, and $C'(\nu_1, \ldots, \nu_p)$ is the *p*-dimensional Gaussian copula density function given by

$$C'(\nu_1, \ldots, \nu_p) = |R|^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\, y'(R^{-1} - I) y\right\}, \tag{9}$$

Where R^{-1} is the inverse of the $p \times p$ correlation matrix, *I* is $p \times p$ identity matrix and $y = (y_1, \ldots, y_p)'$ is a vector such that $y_j = \Phi^{-1}(v_j)$. The joint density function of $T = (T_1, \ldots, T_p)$ is given by

$$f_T(t) = \int_{H_{T_p}(t_p)}^{\infty} \ldots \int_{H_{T1}(t_1)}^{\infty} f_{T|U}(t|u) f_U(u) du_1 \ldots du_p, \tag{10}$$

Where $f_{T|U}(t|u)$ and $f_U(u)$ are given by (3) and (8), respectively, and $H_{T_j}(t_j)$, $j = 1, \ldots, p$, is given by (1) after indexing by *j*.

4. DEPENDENCE PROPERTIES

One reason for using Frank's and the Gaussian copulas, rather than others, is that they provide full range of dependence. For Frank's copula, Nelsen (2010) evaluated Kendall's (K_τ) and Sperman's rho (S_ρ) coefficients of correlation to be, respectively

$$K_\tau = 1 - \frac{4}{\alpha}\{D_1(-\alpha) - 1\},$$

And

$$S_\rho = 1 - \frac{12}{\alpha}\{D_2(-\alpha) - D_1(-\alpha)\},$$

Where for $k = 1,2$, $D_k(\cdot)$ is "Debye" function. One can find out that:

- The correlation parameter $\alpha \to 0$, $C(v_1, v_2)$ given by (5), approaches $v_1 v_2$ implying independence.
- When $\alpha \to \infty$, the positive correlation increases such that the correlation measure approaches $+1$.
- When $\alpha \to -\infty$, the correlation measure approaches -1.

On the other hand it is known that the Gaussian copula family provides a full range of dependence between the joint variables. That is

$$-1 \le Corr_U(U_1, U_2) \le +1.$$

For simplicity in studying dependence properties of the multivariate model based on mixtures and Gaussian copula we provide the multivariate exponential distribution as a special case of both multivariate Weibull and multivariate Gompertz distributions. For the bivariate exponential distribution with marginal distributions $Exp(c_j)$, $j = 1,2$, one can easily find that the conditional expectation

$$E_{T|U}(T_1 T_2 | U_1 = u_1, U_2 = u_2) = \frac{u_1 u_2}{4 c_1 c_2}.$$

Consequently, the correlation between T_1 and T_2 is given by

$$Corr_T(T_1,T_2)=\frac{1}{4}E_U(U_1U_2)-1, \tag{11}$$

And the correlation between U_1 and U_2 is given by

$$Corr_U(U_1,U_2)=\frac{1}{2}E_U(U_1U_2)-2. \tag{12}$$

From equations (11) and (12) one can show that the dependence of the bivariate exponential besed onmixtures and the Gaussian copula is

$$Corr_T(T_1,T_2)=\frac{1}{2}Corr_U(U_1,U_2).$$

That is, our idea of combining both mixture and copula provides a wide range of both positive and negative dependence.

5. ILLUSTRATIVE EXAMPLE

The data set of this example, fracture toughness, is taken from Walker and Stephens (2000). They developed probability models for the analysis of the fracture toughness data in the ductile to brittle temperature transition range. Assuming that the observed crack length (in mm) is presented by a random variable T_1 and random variable T_2 is the fraction toughness (in KJ).In this study we concentrate on the case of ductile fracture occurrence, both T_1 and T_2 are positive. Specifically, we deal with a subset of the comprises 23 data pairs. Table 2, presents the considered data set.

Applying WinBUGS, three bivariate distributions are fitted, separately and twice, to the data set. One time for Frank's copula and the other for the Gaussian copula The three bivariate distributions are bivariate exponential, bivariate Weibull and bivariate Gompertz. The six model parameters were estimated. Figures 1, 2 and 3 show the marginal posterior distributions of the parameters of the bivariate models with Frank's copula. And Figures 4, 5 and 6, show the marginal posterior distributions of the parameters of the bivariate models with Gaussian copula.

Table 2: Fracture Toughness Data Set. (Source: Walker and Stephens, 2000)

T_1	T_2
0.14	101.9
0.85	309.0
0.311	219.0
0.611	270.0
1.17	309.8
1.0	318.0
0.2	169.0
0.17	177.0
1.922	291.9
0.474	270.6
2.260	390.5
0.190	256.9
0.430	293.9
0.680	328.5
0.760	333.4

1.170	394.9
0.860	337.9
0.890	339.6
0.870	340.9
0.430	286.7
0.670	278.0
0.377	197.6
0.294	220.1

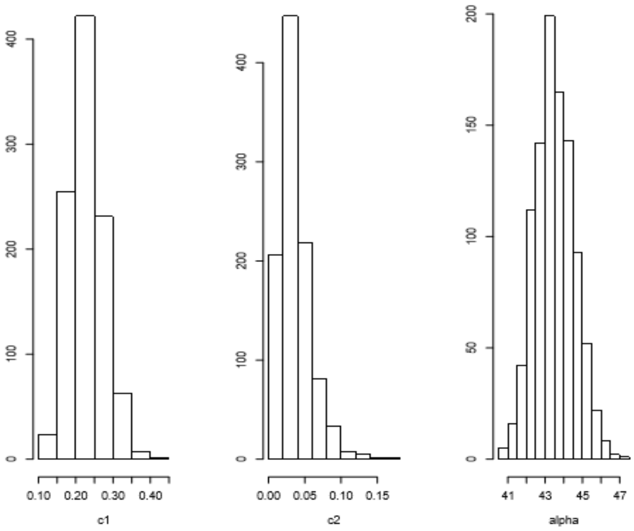

Figure 1: Marginal Posterior Distributions of the Correlation Parameters of the Bivariate Exponential Distribution Based on Frank's Copula

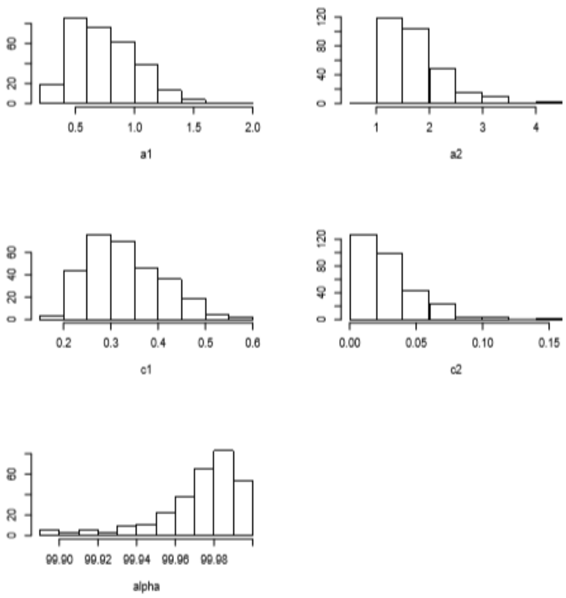

Figure 2: Marginal Posterior Distributions of the Parameters of the Bivariate Weibull Distribution Based on Frank's Copula

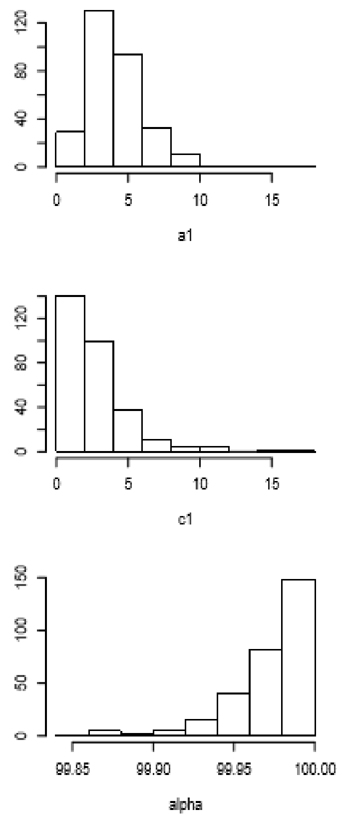

Figure 3: Marginal Posterior Distributions of the Parameters of the Bivariate Gompertz Distribution Based on Frank's Copula

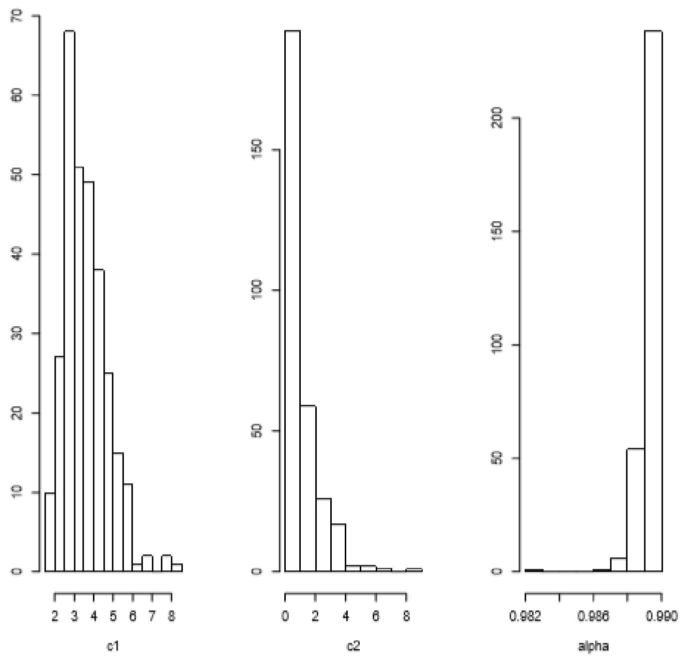

Figure 4: Marginal Posterior Distributions of the Parameters of the Bivariate Exponential Distribution Based on the Gaussian Copula

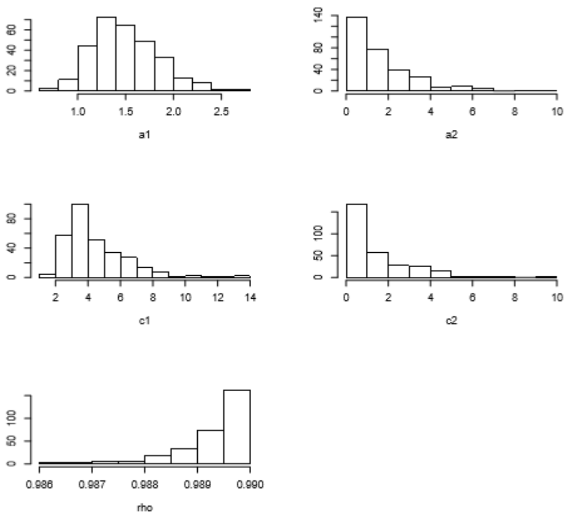

Figure 5: Marginal Posterior Distributions of the Parameters of the Bivariate Weibull Distribution Based on the Gaussian Copula

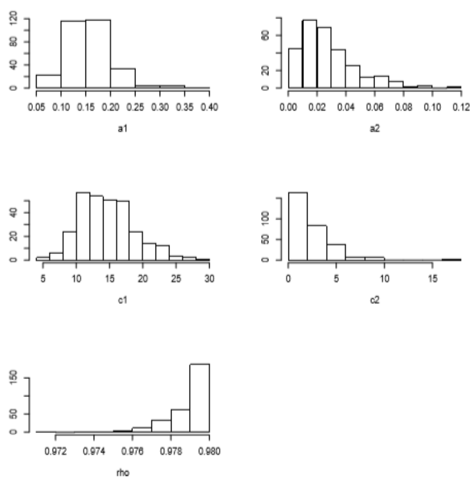

Figure 6: Marginal Posterior Distributions of the Parameters of the Bivariate Gompertz Distribution Based on the Gaussian Copula

ACKNOWLEDGEMENT*S*

Thank to Professor Stephen G. Walker for his helpful advises in preparation of this paper.

REFERENCES

1. **Abd El Lteef, A.O. (2005).** *On the construction of bivariate distributions with specified marginals.* Master's thesis, Ciro University, Institute of Statistical Studies and Research, Ciro-Egypt. Unpublished thesis.

2. **Adham, S.A. and Walker, S.G. (2001).**A multivariate Gompertz-Type distribution. *Journal of Applied Statistics*, Vol. 28, No. 8, pp. 1051-1065.

3. **AL-Dayian, G.R., Adham, S.A., El Beltagy, S.H. and AbdElaal, M.K. (2008).** Bivariate half-logistic distributions based on mixtures and copula. *Academy of Business Journal, AL-Azhar University*, No. 2, pp. 94-107.

4. **Alai, Daniel H. and Landsman, Zinoviy and Sherris, Michael (2015).** Amultivariate Tweedie lifetime model: Censoring and truncation. *Insurance: Mathematics and Economics*, Vol. 64, pp. 203-213. ISSN 0167-6687. (doi:10.1016/j.insmatheco.2015.05.011)

5. **Alai, Daniel H. and Landsman, Zinoviy and Sherris, Michael (2016).** Multivariate Tweedie lifetimes: the impact of dependence. *Insurance: Scandinavian Actuarial Journal*, Vol. 2016 -Issue 8. pp. 692-712.

6. **AL-Hussaini, E.K. and Ateya, S.F. (2006).** A class of multivariate distributions and new copulas. *J. Egypt. Math. Soc.*, Vol. 14, No. 1, pp. 45-54.

7. **Carrier, J.F. (2000)**. Bivariate survival models for coupled lives. *Scandinavian Actuarial Journal*, Vol. 1, pp. 17-32.

8. **Crowder, M.J. (1989).** A multivariate distribution with Weibull connections, *Journal of the Royal Statistical Society Ser.* Part B, Vol.51, pp.93-107.

9. **Frees, E.W., Carrier, J.F., and Vales, E. (1996).** Annuity valuation with dependent mortality, *Journal of Risk and Insurance*, Vol. 63, pp. 229-261.

10. **Gupta, A.K., Zeng, W-B. And Wu, Y. (2010).** Multivariate lifetime distributions, *Probability and Statistical Models*, pp. 117-140. DOI: 10.1007/978-0-8176-4987-6_6.

11. **Hougaard, P. (1986).** A class of multivariate failure time distributions. *Biometrika*, Vol. 73, pp.671-678.

12. **Hutchinson, T.P. and Lai, C.D. (1990).** Continuous bivariate distributions, Emphasising Applications. Ramsby, Sydney, Australia.

13. **Joe, H. (1993).** Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis*, Vol. 46, pp. 262-282.

14. **Joe, H.(1997).** Testing for correlation between non- negative variates. *Biometrica*, Vol. 27, pp. 305-320.

15. **Meisch, D., Nielsen, B. F., &Leleur, S. (2014).** Multivariate phase type distributions - Applications and parameter estimation. Kgs. Lyngby: Technical University of Denmark. (DTU Compute PHD-2014; No. 331).

16. **Nelsen, R.B. (2010).** An Introduction to Copula. 2nd ed., Berlin: Springer-Verlag.

17. **Song, P.X.K. (2000).** Multivariate dispersion models generated from Gaussian copula, *Scandinavian Journal of Statistics*, Vol.27, pp. 305-320.

18. **Walker, S.G. and Stephens, D.A. (1999).** A multivariate family of distributions on $(0,\infty)^p$, *Biometrica*, Vol. 86, pp. 703-709.

19. **Walker, S.G. and Stephens, D.A. (2000).** Modelling and analysis of fracture toughness in the ductile to brittle temperature tradition range: A Bayesian competing risk approach. Unpublished manuscript.