

Analysis of Multiple Myeloma Life Expectancy Using Copula

Seung-Hwan Lee¹, Phillip Yulin Deng¹ & Eun-Joo Lee²

¹ Department of Mathematics, Illinois Wesleyan University, Bloomington, USA

² Department of Mathematics, Millikin University, Decatur, USA

Correspondence: Seung-Hwan Lee, Department of Mathematics, Illinois Wesleyan University, Bloomington, IL 61702, USA. Tel: 1-309-556-3421. E-mail: slee2@iwu.edu

Received: September 10, 2012 Accepted: January 7, 2013 Online Published: January 16, 2013

doi:10.5539/ijsp.v2n1p44 URL: <http://dx.doi.org/10.5539/ijsp.v2n1p44>

Abstract

Multiple myeloma is a blood cancer that develops in the bone marrow. It is assumed that in most cases multiple myeloma develops in association with several medical factors acting together, although the leading cause of the disease has not yet been identified. In this paper, we investigate the relationship between the factors to measure multiple myeloma patients' survival time. For this, we employ a copula that provides a convenient way to construct statistical models for multivariate dependence. Through an approach via copulas, we find the most influential medical factors that affect the survival time. Some goodness-of-fit tests are also performed to check the adequacy of the copula chosen for the best combination of the survival time and the medical factors. Using the Monte Carlo simulation technique with the copula, we re-sample survival times from which the anticipated life span of a patient with the disease is calculated.

Keywords: elliptical copula, multiple myeloma, tail dependence

1. Introduction

Multiple myeloma is a blood cancer caused by the accumulation of abnormal plasma cells in the bone marrow. It is the second most common blood cancer, after non-Hodgkin's lymphoma, and represents approximately 1 percent of all cancers and 2 percent of all cancer deaths. The American Cancer Society estimates that 20,180 people were diagnosed with multiple myeloma during 2010 (Multiple Myeloma: Disease Overview, Multiple Myeloma Research Foundation, 2010). The prognosis of the disease is often unpredictable and overall survival is ranged from a few months to more than 10 years (Kyle & Rajkumar, 2008). It has been known that multiple myeloma may be the result of several medical risk factors that act together. Therefore, it is important to understand the relationship between survival time of patients and the risk factors in the bone marrow for the disease. So this work investigates possible associations in multiple myeloma data, carried out at the Medical Centre at the University of West Virginia, with a particular focus on potential survival time (denoted by ST hereafter) of a patient over the treatment period in association with the following four medical factors: level of blood urea nitrogen (BUN), serum calcium (CA), hemoglobin (HB), and the percentage of plasma cells (PC). See Collect (1999) and Krall et al. (1975) for the details of the data.

When analyzing dependence of several variables, such as the survival time and associated disease factors, an often used measure is Pearson's linear correlation coefficient. However, the linear correlation coefficient does not detect non-linear behaviors of the variables considered, is strongly affected by extreme values and not invariant under non-linear transformations (Embrechts, McNeil, & Straumann, 2002; Frees & Valdez, 1998; Schweizer & Wolff, 1981). To overcome these problems, we employ a statistical function called a copula which links multivariate distributions to marginal distributions. A copula captures both the linear and non-linear dependence that may exist between the variables. It is also invariant under monotone transformations. Especially, the use of a copula has gained importance as a simple tool to measure the amount of dependence in the tails. For example, the treatment of a disease can either exceed or fall below a given level at either the early or late stage of the disease. The use of a copula has been studied in several disciplines such as survival analysis (Zheng & Klein, 1995), risk management and financial applications (Breyman et al., 2003; Embrechts et al., 2002; 2003). Excellent reviews on copulas can be found in Nelson (1999).

Copulas have varying amounts of tail dependence. So an appropriate copula that fits the data should be used. A poorly chosen copula may lead to undesirable results. This issue has been studied by many authors, including

Melchiori (2003), Durrelaman (2000), Kumar and Shoukri (2008), Frees and Valdez (1998), and Genest and Rivest (1993). Similar to the approach employed in the literature, based on the distance of copula and its empirical version, we choose a copula that best fits data. We then assess the adequacy of the copula chosen. The procedures are based on the Monte Carlo simulation technique by which the distribution of the distance can be seamlessly approximated under the null hypothesis of the no model misspecification. As a numerical measure of the model adequacy, the empirical - value, obtained from the simulated distance process, is used.

The primary objective of this paper is to estimate the expected life span of a patient with the disease by identifying the dependence between the variables considered. To this end, a systematic study in search for the most influential medical factors that affects the survival time is performed through copulas. Based on a large number of simulated data obtained from a copula chosen for a best combination of the survival time and medical factors, we calculate maximum extension possible for a life with reference to the factors that influence the survival time. This work is organized as follows. An overview of copulas is given in section 2. Section 3 discusses the procedures of finding an appropriate copula for multiple myeloma data. Numerical results of the anticipated life span of a patient with the disease are presented in section 4. Section 5 concludes the paper.

2. Copula

If X_1, \dots, X_n are random variables, the multivariate distribution function is defined as

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n),$$

which completely describes the dependence between the random variables. A copula provides a useful way to construct such a multivariate distribution of two or more random variables. The essential idea of the copula approach is that a joint distribution is factored into two components: the marginal distributions and a dependence function called a copula, as described in the following theorem (Sklar, 1959).

Theorem *If F is a multivariate distribution function with marginal distribution functions F_1, \dots, F_n , then there is a copula C such that*

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)). \quad (1)$$

Theorem states that a copula function defines a joint distribution, evaluated at x_1, \dots, x_n , with marginal distributions F_1, \dots, F_n . Letting f be the probability density function of F , $u_i = F_i(x_i)$, $i = 1, \dots, n$, and c the density function of a copula C , it can be easily shown that

$$f(x_1, \dots, x_n) = c(u_1, \dots, u_n) \times \prod_i f_i(x_i).$$

This indicates that a multivariate probability density function $f(x_1, \dots, x_n)$ can be split into the univariate marginal probability density functions $f_i(x_i)$'s and the copula density function $c(u_1, \dots, u_n)$ that determines a dependence structure. Hence, it is possible to separately specify marginal density functions and the dependence relation determined by the copula density function. A copula itself is in fact a multivariate distribution with standard uniform marginal distributions due to the fact that $F_i(X_i)$'s are uniformly distributed over the interval $[0,1]$. So it maps points on the n -dimensional unit square to values between 0 and 1. Specifically, with the values u_1, \dots, u_n of standard uniform random variables U_1, \dots, U_n , the copula $C(u_1, \dots, u_n)$ of the multivariate distribution function F is

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)), \quad (2)$$

where $F_1^{-1}, \dots, F_n^{-1}$ denote the quantile functions of the marginal distributions F_1, \dots, F_n . Note that from (1) and (2), a specified copula determines the dependence structure of data, while the marginal distribution does not. This is due to the fact that a copula links univariate marginal distributions to their multivariate distribution and is independent of marginal distributions. Therefore, we need to consider several copulas, leading to different dependence structures, from which an appropriate copula for data is chosen. In this work, we look at elliptical copulas the most commonly used copulas.

Elliptical copulas are the copulas of elliptical distributions. Examples include the Gaussian and t copulas (Embrecht et al., 2002; Demnarta & McNeil, 2005). The key advantages of elliptical copulas are that they are suitable in modeling dependence structures with multi-dimensions and specify different levels of correlations between the marginal distribution functions. As seen from the expressions in (1) and (2), any marginal distributions can be imposed over an elliptical copula, provided that marginal distributions are known and can be consistently estimated

from data (Nelson, 1999; Embrechts et al., 2002). A copula with given marginals is called meta-elliptical copula. However, for simplicity and without loss of generality, elliptical copula indicates meta-elliptical copula hereafter.

2.1 Gaussian and t Copulas

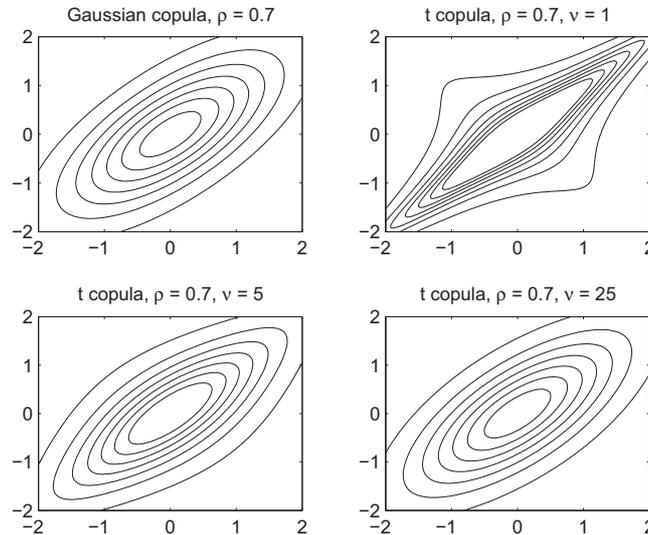


Figure 1. Contour plots of Gaussian and t copulas, $\rho = 0.7$

The Gaussian copula is the copula of multivariate normal distribution. From (2), it is defined as

$$C_G(u_1, \dots, u_n) = \Phi_\rho(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)),$$

where Φ_ρ denotes the standard multivariate normal distribution function with correlation coefficient matrix ρ and Φ^{-1} the inverse of the standard univariate normal distribution function. As ρ approaches -1 and 1, the Gaussian copula captures stronger positive and negative linear relationship between random variables, respectively. See Figure 1 for these phenomena. The linear correlation coefficients depend on the marginal distributions, measuring the overall strength of a linear relationship, but give no information about how that varies across the distribution. So it is not preserved by copulas. It is important to note that the linear correlation coefficient has such shortcomings, but it is still needed to parameterize the Gaussian copula. Rank based correlations describe the global association of variables and are invariant under any monotonic transformations. In this work, the linear correlation coefficient (ρ) calculated by the Kendall's rank correlation (τ) was used such that $\rho = \sin(\tau\pi/2)$. We plug the marginal distributions into a Gaussian copula function to obtain a multivariate distribution. For example, using the copula $C_G(u_1, \dots, u_n)$ and the marginal distribution functions F_1, \dots, F_n , we have the multivariate distribution

$$F(x_1, \dots, x_n) = C_G(F_1(x_1), \dots, F_n(x_n)).$$

Note that the multivariate normal distribution has thin tails. Thus, the Gaussian copula is not appropriate in the analysis of tail dependence of variables that have heavy-tailed distributions (Embrechts et al., 2002). To repair this problem, we consider the t copula which can capture dependence between variables in the tails of the distribution. The t copula is a copula of the multivariate t distribution. From (2) it is defined as

$$C_t(u_1, \dots, u_n) = t_{\rho, \nu}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_n)),$$

where $t_{\rho, \nu}$ denotes the standard multivariate t distribution function with correlation coefficient matrix ρ and the degrees of freedom ν , and t_ν^{-1} is the inverse of the standard univariate t distribution function. Similar to the Gaussian copula, ρ is approximated by Kendall's τ . The second parameter ν controls the thickness of the tails of the distribution, exhibiting co-movement of the variables in the tails as seen in Figure 1.

Note that the Gaussian copula is a limiting case of the t copula as $\nu \rightarrow \infty$, and the t copula with $\nu = 1$ is often called the Cauchy copula. This implies that with increasing degrees of freedom, the t copula tends to look more

like the Gaussian copula. We insert the marginal distributions into the t copula function to obtain a multivariate distribution, as follows.

$$F(x_1, \dots, x_n) = C_t(F_1(x_1), \dots, F_n(x_n)).$$

Table 1. Coefficient of Tail Dependence of the t copula (Embrechts et al., 2002)

$\nu \setminus \rho$	-0.5	0	0.5	0.9	1
2	0.06	0.18	0.39	0.72	1
4	0.01	0.08	0.25	0.63	1
10	0	0.01	0.08	0.46	1
∞	0	0	0	0	1

2.2 Tail Dependence

The relationship between random variables, especially in the tails, is controlled by the choice of copulas. It can be described by the coefficient of asymptotic tail dependence of a copula. For a pair of random variables, X and Y , it quantifies the probability to observe a large Y given that X is large. Specifically, the coefficient of upper tail dependence is

$$\lambda_U = \lim_{q \rightarrow 1^-} P(Y > F_2^{-1}(q) | X > F_1^{-1}(q)),$$

provided the limit exists, and the coefficient of lower tail dependence is

$$\lambda_L = \lim_{q \rightarrow 0^+} P(Y \leq F_2^{-1}(q) | X \leq F_1^{-1}(q)),$$

provided the limit exists. See Embrechts et al. (2002) for details on these expressions. Note that $\lambda_U = \lambda_L$ for the Gaussian and t copulas, due to the symmetric property of the elliptical distributions. Let $\lambda_U = \lambda_L = \lambda$. Then the following formula is often used for computational purposes:

$$\lambda = 2t_{\nu+1}(-\sqrt{\nu+1} \sqrt{1-\rho} / \sqrt{1+\rho}), \quad (3)$$

where t_ν is the standard univariate t distribution with ν degrees of freedom. Table 1 shows the results of the tail dependence coefficients for a few representative values of ρ and ν (Embrechts et al., 2002) calculated by the formula in (3).

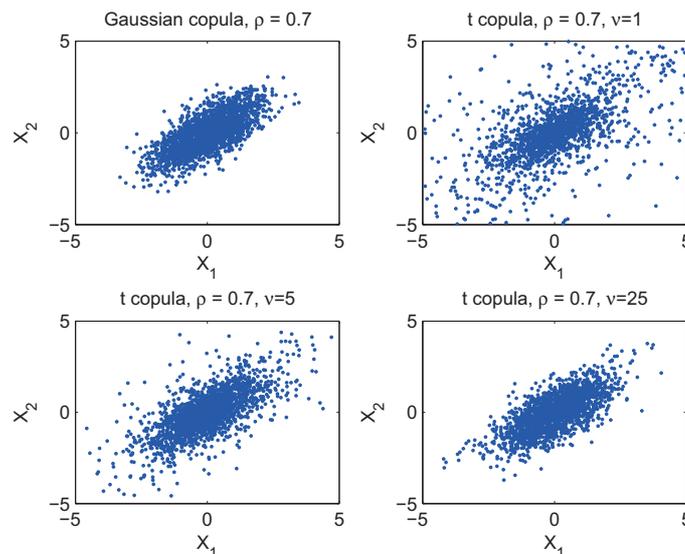


Figure 2. Scatter plots of 2000 simulated data points for Gaussian and t copulas

The tail dependence coefficient λ tends to zero as the degrees of freedom ν tends to infinity for $\rho < 1$. Since the Gaussian copula is a limiting case of the t copula as $\nu \rightarrow \infty$, the value of λ for the Gaussian copula is 0. So the

Gaussian copula exhibits no tail dependence. On the contrary, the t copula has non-negative values of λ for all values of ρ , and so the association of extreme values is captured by the t copula, with different amounts depending on ν at a fixed value of ρ . As seen in Table 1, λ increases as ν decreases at a fixed value of ρ . This indicates that the t copula has tail dependence that increases, whether it is upper tail dependence or lower tail dependence, with decreasing parameter ν . So it is useful when dependence of extreme values is observed in data.

Note that although the t copulas generate different dependence structures, they may still have the same marginal distribution functions and the same correlation. To illustrate this, we generate 2000 simulated data points using the Gaussian and t copulas with 1, 5, 25 degrees of freedom for $\rho=0.7$. The scatter plots of those simulated values are presented in Figure 2, displaying that the tail dependence is not remarkable in the t copula with $\nu=25$ when compared to either $\nu=1$ or $\nu=5$. It is evident from this figure that the tail dependence becomes stronger as the degrees of freedom decreases, and this indicates that an increase in ν results in fewer occurrences of joint extremes. The plots also demonstrate that the dependence structure of multivariate distributions may not be perfectly identified only by their marginal distributions and correlations. In summary, a copula fully captures real dependence structure among random variables and hence provides a model that reflects on more detailed information about data.

3. Procedures

Multiple myeloma is a malignant disease caused by the accumulation of abnormal plasma cells in the bone marrow. Pain emanating from bone tissue and cancerous destruction of bone tissue may occur as a result. This section analyzes the effect of the medical factors on the survival time of patients with the disease carried out at the Medical Centre of the University of West Virginia. Survival time of patients (ST), the level of blood urea nitrogen (BUN), serum calcium (CA), hemoglobin (HB), and the percentage of plasma cells (PC) in the bone marrow are considered. Data can be found in Krall et al. (1975). For simplicity, complete data points for males in the data set are used, where the sample size is 22.

Procedures are based on a copula chosen by the Kolmogorov-Smirnov type distance. We select a best copula within the class of the t copulas from the Kolmogorov-Smirnov type distance criterion and then check its adequacy using some goodness of fit tests. Finally, we explore the anticipated longest life span of a patient with the disease based on a large number of simulated data points generated by the copula. We start with finding the marginal distribution of the medical factors.

3.1 Marginal Distribution

Appropriate marginal distributions should be plugged in the t copula. Toward an optimal distribution for the data given, numerical model checking methods, such as the Kolmogorov-Smirnov and Anderson-Darling tests, were used to check if the distribution chosen is appropriate. The Kolmogorov-Smirnov test uses maximum difference between the empirical distribution and the theoretical distribution, defined as $\sup_x \{|\hat{F}(x) - F(x)|\}$ where $\hat{F}(x)$ is the empirical distribution. The Anderson-Darling test is the test that is more sensitive in the tails of the distribution than the Kolmogorov-Smirnov test. Best fitting marginal distributions for data are selected such that the distance function is minimized. Judging from those two criteria, we arrive at the distributions summarized in Table 2. Table 3 shows the correlations coefficients for the data.

Table 2. Distribution and parameter estimate for the multiple myeloma data

	Distribution	Parameter Estimate	Mean	Std. Dev.	Skewness	Kurtosis
ST	Lognormal	$\mu=2.47, \sigma=1.29$	22.77	24.42	1.44	1.26
BUN	Lognormal	$\mu=3.28, \sigma=0.66$	34.0	32.44	3.37	13.50
CA	Lognormal	$\mu=9.59, \sigma=1.26$	10.32	1.62	1.37	2.14
HB	Lognormal	$\mu=14.79, \sigma=0.73$	10.43	2.77	-0.14	-0.89
PC	Lognormal	$\mu=3.55, \sigma=0.65$	42.59	27.45	0.94	0.01

Table 3. Correlation coefficient for the multiple myeloma data

	S	BUN	CA	HB	PC
S	1	-0.3454	-0.0196	0.3435	-0.2899
BUN	-0.3454	1	0.1219	-0.3624	-0.0599
CA	-0.0196	0.1219	1	0.2175	-0.0831
HB	0.3435	-0.3624	0.2175	1	-0.2985
PC	-0.2899	-0.0599	-0.0831	-0.2985	1

The distribution function gives the probability that a random variable is less than a given value, describing the parent distribution. The empirical distribution function is similar, the difference being that the empirical distribution is calculated by data, resembling the theoretical distribution that fits data. So the plot of \hat{F} versus F should be close to each other if the distribution considered is legitimate for data. For example, Figure 3 shows that the empirical distribution for the survival time is fairly close to the specified distribution, the lognormal distribution. Hence the lognormal distribution appears to be an appropriate distributional model. This graphical method confirms the result from the numerical methods.

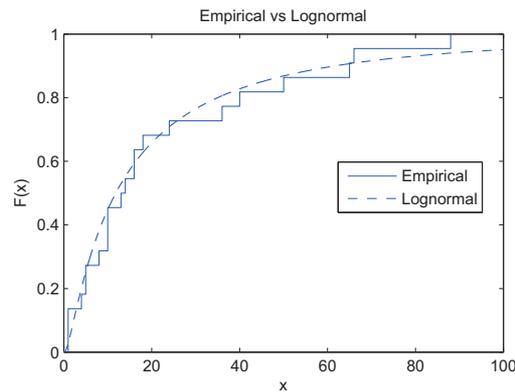


Figure 3. Plot of empirical vs lognormal for survival time

3.2 Copula Selection and Its Adequacy

Depending on the degrees of freedom, the t copulas specify different amounts of structural information, especially on tail behavior of distributions. Even with identical correlations, the dependence structure of distributions may be different depending on the choice of copulas. The t copula with low degrees of freedom captures the non-linear trends well, while the Gaussian copula or the t copula with high degrees of freedom fit well when the dependence is mostly linear. So a copula that well describes the structure of data should be used in practice. This section finds such a suitable copula to use and checks its statistical significance. The procedures are based on the empirical copula, \hat{C} , introduced in Deheuvels (1978), defined as

$$\hat{C}(u_1, \dots, u_n) = \hat{H}(\hat{F}_1(x_1), \dots, \hat{F}_n(x_n)), \quad (4)$$

where

$$\hat{H}(u_1, \dots, u_n) = \frac{1}{n} \sum_{i=1}^m I\{U_{1,i} \leq u_1, \dots, U_{n,i} \leq u_n\}$$

and \hat{F}_i 's are the usual empirical distributions that correspond to univariate marginal distributions. Similar to the empirical distribution that estimates an unknown distribution, the empirical copula obtained by data estimates a theoretical copula. The empirical copula in (4) actually provides approximate probabilities of the number of pairs (u_1, \dots, u_n) such that $u_1 \leq u_{1(i)}, \dots, u_n \leq u_{n(i)}$, where $u_{(i)}$ denotes the order statistic. A best copula within a class of the t copulas is chosen such that the distance of the empirical copula \hat{C} and the theoretical copula C is minimized (Durrleman et al., 2000). However, this does not mean that a copula chosen is a fitted model for data in an absolute sense. Statistical testing procedures should be conducted to decide whether the copula chosen is adequate for data. To this end, based on the Cramr-von Mises type statistic (Genest & Rmillard, 2008; Genest et al., 1993; Genest et al., 2009), define the process in x

$$D(x) = n \int_0^x \{\hat{C} - C\}^2 d\hat{C},$$

for $0 < x \leq 1$. From this, define the simulated process,

$$D^*(x) = n \int_0^x \{\hat{C} - C^*\}^2 d\hat{C}, \quad (5)$$

where C^* is the simulated copula of C from which data are re-sampled. Under the null hypothesis that the copula C is valid for data, the empirical copula resembles the assumed copula. Therefore, comparison of D to a large number

of simulated realizations from the process D^* in (5) will lead to goodness-of-fit tests for the model. Specifically, since the process $D(x)$ randomly fluctuates around zero under the null hypothesis, a distinguishably large value of D compared to the values of D^* would indicate model misspecification. Thus, $S = \sup_{0 < x \leq 1} |D(x)|$ can be used as a numerical measure for the assessment of the model adequacy. A large value of S will lead to rejection of the null hypothesis. Let $S^* = \sup_{0 < x \leq 1} |D^*(x)|$. Then, the p -value defined as $P(S \geq s)$, where s is the observed value of S , can be used as a measure of strength of the model adequacy, and this p -value is approximated by $P(S^* \geq s)$ which can be empirically estimated by the Monte Carlo simulation method, similar to Lee et al. (2008). We use this p -value as a numerical measure of how well a copula model chosen fits the data. Lower the p -value, the less likely the goodness of fit is.

4. Numerical Results

The goal of this study is to estimate the expected life span of a patient with the disease by finding a combination of the survival time and the medical factors that best summarizes it. We considered various cases, each of which contains the survival time denoted by ST. Specifically, we investigate the following four classes: 1. (ST, HB), (ST, CA), (ST, BUN), (ST, PC), 2. (ST, BUN, CA), (ST, BUN, HB), (ST, BUN, PC), (ST, CA, HB), (ST, CA, PC), (ST, HB, PC), 3. (ST, BUN, CA, HB), (ST, BUN, CA, PC), (ST, BUN, HB, PC), (ST, CA, HB, PC), and 4. (ST, BUN, CA, HB, PC). Classes 1, 2, 3, and 4 consider the association of two, three, four and all the factors, respectively. To find the best copula from the four classes, in the first step we performed the copula selection process mentioned in Section 3.2. Based on this, we arrived at the following cases as superior within their own class: (ST, HB), (ST, CA, HB), (ST, BUN, CA, HB), and (ST, BUN, CA, HB, PC). Results are summarized in Table 4. Note that to find the best t copula for each case, the degrees of freedom ranging from 1 to 30 was considered. The t copula with low degrees of freedom would be selected if the dependence of extreme values is detected.

Table 4. Copula selected and its p -value

Case	Copula	Distance/ n	p -value
(ST, HB)	t copula, $\nu=6$	0.0038	0.8580
(ST, CA, HB)	t copula, $\nu=16$	0.0065	0.7250
(ST, BUN, CA, HB)	t copula, $\nu=6$	0.0082	0.4590
(ST, BUN, CA, HB, PC)	t copula, $\nu=5$	0.0192	0.3640

In the next step, we obtained the p -values using the goodness-of-fit test procedures described in Section 3.2 to check the adequacy of the copulas chosen. The estimated p -values obtained for (ST, HB), (ST, CA, HB), (ST, BUN, CA, HB), and (ST, BUN, CA, HB, PC) are presented in Table 4. Results are based on 2000 simulated realizations. Judging from the results, we conclude that the relationship between survival time and hemoglobin, (ST, HB), which is modeled by the t copula with the degrees of freedom 6 would most likely be distinguishable among others. Multiple myeloma has been staged by the method developed by Durie and Salmon (1975). From the staging method by Durie and Salmon (1975), it was found that the level of hemoglobin in the blood of a multiple myeloma patient is strongly associated with the tumor mass and thus is a strong indicator of the disease progress (Durie & Salmon, 1975; Kyle & Rajkumar, 2008). Our results are in agreement with this study.

Lastly, we examine the anticipated life expectancy of a patient with multiple myeloma from diagnosis until death by generating a large number of simulated survival times from the copula chosen. Specifically, based on the aforementioned results, in association with the marginal distributions described in Table 2, we re-sampled survival times of size 200,000 from the t copula ($\nu=6$) with reference to hemoglobin only. Table 5 presents the estimated survival times at various percent confidence levels, from 5% (shortest) survival time to 95% (longest) survival time, in months, from diagnosis until death from multiple myeloma. Table 5 also shows the corresponding estimated hemoglobin levels. For example, the 95th percentile of the survival times indicates that survival times of a patient under treatment could extend to 99.13 months (approximately 8.3 years) and its corresponding hemoglobin level is 14.68. Figure 4 shows 2000 simulated realizations under the t copula with degrees of freedom 6. The positive effect of hemoglobin on survival time seems somewhat stronger than the negative effect in the scatter plot.

Table 5. Multiple myeloma life expectancy for (ST, HB), 200,000 samples

Percentile	5%	10%	30%	50%	70%	90%	95%
Estimated life expectancy	1.41	2.26	6.04	11.84	23.25	62.02	99.13
Estimated hemoglobin	7.52	7.53	6.67	7.89	9.51	10.65	14.68

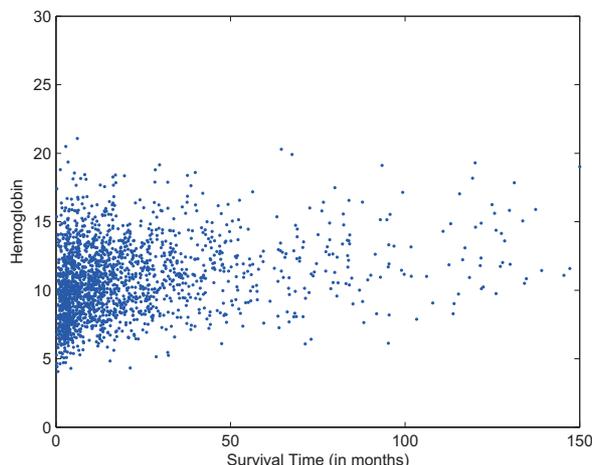


Figure 4. Survival time vs Hemoglobin, t copula with $\nu=6$, 2000 simulated samples

5. Concluding Remarks

Multiple myeloma is a malignant disease that develops in the bone marrow. In most cases this disease develops in association with several medical factors that act together. A copula provides a convenient tool in the analysis of multivariate data that represent such medical factors. It was found that the relationship between survival time and hemoglobin which is modeled by the t copula with the degrees of freedom 6 would be most distinguishable. Some goodness-of-fit tests based on the simulated process were then performed to check the adequacy of the copula model. Based on the copula, we simulated a large number of simulated realizations of the survival time from which the anticipated life span of a patient with the disease was calculated.

References

- Breymann, W., Dias, A., & Embrechts, P. (2003). Dependence Structures for Multivariate High-Frequency Data in Finance. *Quantitative Finance*, 3(1), 1-16. <http://dx.doi.org/10.1080/713666155>
- Collect, D. (1999). *Modeling Survival Data in Medical Research*. Chapman.
- Deheuvels, P. (1978). Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes. *Publications de l'Institut de Statistique de l'Université de Paris*, 23, 1-36.
- Durie, B. G., & Salmon, S. E. (1975). A clinical staging system for multiple myeloma. Correlation of measured myeloma cell mass with presenting clinical features, response to treatment, and survival. *Cancer*, 36(3), 842-854. [http://dx.doi.org/10.1002/1097-0142\(197509\)36:3<842::AID-CNCR2820360303>3.0.CO;2-U](http://dx.doi.org/10.1002/1097-0142(197509)36:3<842::AID-CNCR2820360303>3.0.CO;2-U)
- Durrelman, V., Nikeghbail, A., & Roncalli, T. (2000). Which copula is the right one? *Credit Lyonnais*. Retrieved from <http://ssrn.com/abstract=1032545>
- Embrechts, P., Lindskog, F., & McNeil, A. (2003). Modelling Dependence with Copulas and Applications to Risk Management. In S. Rachev (Ed.), *Handbook of Heavy Tailed Distributions in Finance* (Chapter 8, 329-384), Elsevier. <http://dx.doi.org/10.1016/B978-044450896-6.50010-8>
- Embrechts, P., McNeil, A., & Straumann, D. (2002). Correlation and Dependency in Risk Management: Properties and Pitfall. In M. Dempster (Ed.), *In Risk Management: Value at Risk and Beyond* (pp. 176-223). Cambridge: Cambridge University Press.
- Frees, M. J., & Valdez, E. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2, 1-25.
- Genest, C., & Rivest, L. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of American Statistical Association*, 88, 1034-1043. <http://dx.doi.org/10.1080/01621459.1993.10476372>
- Genest, C., & Rmillard, B. (2008). Validity of the Parametric Bootstrap for Goodness-of-fit Testing in Semiparametric Models. *Annales l'Institut Henri-Poincaré*, 44, 1096-1127.
- Genest, C., Rmillard, B., & Beaudoin, D. (2009). Goodness-of-fit Tests for Copulas: A Review and a Power Study.

- Insurance: Mathematics and Economics*, 44, 199-213. <http://dx.doi.org/10.1016/j.insmatheco.2007.10.005>
- Krall, J. M, Uthoff, V. A., & Harley, J. B. (1975). A step-up procedure for selecting variables associated with survival. *Biometrics*, 31, 49-51. <http://dx.doi.org/10.2307/2529709>
- Kumar, P., & Shoukri, M. M. (2008). Evaluating Aortic Stenosis using the Archimedean copula methodology. *Journal of Data Science*, 6, 173-187.
- Kyle, R. A., & Rajkumar, S. V. (2008). Multiple myeloma. *Blood*, 111(6), 2962-2972. <http://dx.doi.org/10.1182/blood-2007-10-078022>
- Lee, S., Lee, E.-J., & Omolo, B. O. (2008). Using integrated weighted survival difference for the two-sample censored data problem. *Computational Statistics and Data Analysis*, 52, 4410-4416. <http://dx.doi.org/10.1016/j.csda.2008.02.022>
- Melchiori, M. R. (2003). Which Archimedean copula is the right one? *Yield Curve*, 37, 1-20.
- Multiple Myeloma: Disease Overview. (2010). Multiple Myeloma Research Foundation. Retrieved from www.themmr.org
- Nelsen, R. B. (1999). *An introduction to copulas*. Springer. <http://dx.doi.org/10.1007/978-1-4757-3076-0>
- Schweizer, B., & Wolff, E. F. (1981). On the Nonparametric Measures of Dependence for Random Variables. *Annals of Statistics*, 9, 879-885. <http://dx.doi.org/10.1214/aos/1176345528>
- Sklar, A. (1959). *Functions de Repartition a n Dimensions et leurs Merges*. Publication of the Institute of Statistics, University of Paris 8, 229-231.
- Zheng, M., & Klein, J. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, 82, 127-138. <http://dx.doi.org/10.1093/biomet/82.1.127>