Nonparametric Estimation of Copula Based Stress-strength Models

Mohammadi, M.[1], **Emadi, M.**[2] **and Naderi, H.**[3]

[1] Department of Statistics, University of Zabol, Zabol, Iran

[2] Department of Statistics, Ferdowsi University of Mashhad Mashhad, Iran

[3] Department of Mathematical Sciences, University of Sistan and Baluchestan, Zahedan, Iran

Abstract: The purpose of this paper is to provide a nonparametric method for the estimation of copula-based stress-strength models. These method is based on improved probit transformation method for copula density estimation. This nonparametric method is a novel application based on an existing bivariate kernel method combined with Monte Carlo estimation without specification of the copula or the marginal distributions. Simulation results suggests that the nonparametric estimation method has better performance than the empirical esimation method.

Keywords: Copula, Probit transformation, Stress-strength.

1 Introduction

In reliability analysis, the stress-strength model describes the reliability of an individual which has a random strength X and is subject to a random stress Y. The individual fails if the strength cannot resist on the stress. Hence, $R = P(Y < X)$ represents the reliability of the individual. the stress-strength models have been widely discussed in the statistical and reliability literature. There are

[1]Mohammadi, M.: *mo.mohammadi@uoz.ac.ir*

many literatures have investigated the stress-strength models under different distributions. It is usually assumed that the stress and strength variables are independent, then based on this assumption to analysis the characteristics of the stress-strength models. However, in many cases, the stress and strength variables are dependent in some way. Nevertheless, a bivariate distribution model needs the marginal distribution are the same type. To overcome this limitation, a copula-based approach, which admits the margins are any type and not necessarily belonging to the same family, was considered by some researchers. The Farlie-Gumbel-Morgenstern copula is used by [3] to analyze the dependence in stress-strength models. Recently, parametric and nonparametric inference for the reliability of copula-based stress-strength models is discussed by [2].

This paper provides a general framework for estimating the reliability in copula-based stress-strength models with an emphasis on model-robust inference. These method is based on improved probit transformation method for copula density estimation. We have not focused on any specific family of distributions (margins) or copulas.

The rest of the paper is arranged as follows. In Section 2, the preliminaries for copulas are described. The estimation of the copula density function using local likelihood probit transformation method is provided in Section 3. In Section 4, the nonparametric method for the estimation of copula-based stress-strength is presented. The simulation results are provided in Section 5 and concluding remarks are given in Section 6.

2 Copulas

Some definitions related to a copula functions will be briefly reviewed based on [5]. Let (X,Y) be a continuous random variable with joint cumulative distribution function (cdf) F, then copula C corresponding to F defined as:

$$F(x,y) = C(F_X(x), F_Y(y)), \qquad (x,y) \in R^2, \tag{1}$$

where F_X and F_Y are the marginal distributions of X and Y, respectively. A bivariate copula function C is a cumulative distribution function of random vector (U,V), defined on the unit square $[0,1]^2$, with uniform marginal distributions as $U = F_X(X)$ and $V = F_Y(Y)$.

The authors shall write $C(u,v;\theta)$ for a family of copulas indexed by the parameter θ. If $C(u,v;\theta)$ is an absolutely continuous copula distribution on $[0,1]^2$, then its density function is $c(u,v;\theta) = \frac{\partial^2 C(u,v;\theta)}{\partial u \partial v}$. As a result, the relationship between the copula density function (c) and the joint density function $f_{X,Y}(\cdot,\cdot)$ of (X,Y) according to equation (1) can be represented as

$$f_{X,Y}(x,y) = c(F_X(x),F_Y(y);\theta) f_X(x) f_Y(y), \qquad (x,y) \in R^2, \tag{2}$$

where $f_X(\cdot)$ and $f_Y(\cdot)$ are the marginal density functions of X and Y, respectively.

Table 1 presents summary information of some well-known bivariate copulas such as the parameter space and Kendall's tau (τ) of them. In this table, Clayton, Gumbel, and Frank copulas belong to the class of Archimedean copulas and Gaussian and T copulas belong to the class of Elliptical copulas. The copula-based Kendall's tau association for continuous variables X and Y with copula C is given by $\tau = 4\int_{[0,1]^2} C(u,v)dC(u,v) - 1$.

Table 1: Some well-known bivariate copulas

Copula	$C(u,v;\theta)$	Parameter Space	Kendall's τ
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$\theta \in [0,+\infty)$	$\frac{\theta}{\theta+2}$
Gumbel	$\exp\left\{-\left[(-\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{1/\theta}\right\}$	$\theta \in [1,+\infty)$	$\frac{\theta-1}{\theta}$
Gaussian[1]	$\Phi_2(\Phi^{-1}(u),\Phi^{-1}(v);\theta)$	$\theta \in [-1,+1]$	$\frac{2}{\pi} arcsin(\theta)$

3 Local likelihood probit transformation estimation

Transformation method to kernel copula density estimation was introduced by [1]. The simple idea is to transform the data so that it is supported on the full R^2. On this transformed domain, standard kernel techniques can be used to estimate the density. An adequate back-transformation then yields an estimate of the copula density.

Let $(U_i, V_i)_{i=1,\ldots,n}$ are independent and identically distributed observations from the bivariate copula C and the purpose is to estimate the corresponding copula density function. Denote Φ as the standard Gaussian distribution and ϕ as its first order derivative. Then $(S_i, T_i) = (\Phi^{-1}(U_i), \Phi^{-1}(V_i))$ is a random vector with Gaussian margins and copula C. According to (2), the corresponding density function can be written as $f(s,t) = c(\Phi(s), \Phi(t))\phi(s)\phi(t)$. Thus, an estimation of the copula density function can be given by

$$\hat{c}_n^{(\mathscr{P}\mathscr{T})}(u,v) = \frac{\hat{f}_n(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}, \qquad (u,v) \in (0,1)^2. \tag{3}$$

As the (U_i, V_i) are unavailable and one has to use the pseudo-transformed sample $(\hat{S}_i, \hat{T}_i) = (\Phi^{-1}(\hat{U}_i), \Phi^{-1}(\hat{V}_i))$, instead. As a first natural idea, the standard kernel density estimator for $\hat{f}_n$ in (3) can be considered as follows:

$$\hat{f}_n(s,t) = \frac{1}{n|\mathbf{H}_{ST}|^{\frac{1}{2}}} \sum_{i=1}^{n} \mathbf{K}\left(\mathbf{H}_{ST}^{-\frac{1}{2}} \begin{pmatrix} s - \hat{S}_i \\ t - \hat{T}_i \end{pmatrix}\right),$$

where $\mathbf{K} : R^2 \to R$ is a kernel function, and $\mathbf{H}_{ST} = b_n I$ is a bandwidth matrix.

This kernel estimator has asymptotic problems at the edges of the distribution support. To remedy this problem, local likelihood probit transformation ($\mathscr{L}\mathscr{L}\mathscr{P}\mathscr{T}$) method was recently suggested by [4]. Instead of applying the standard kernel estimator, they locally fit a polynomial to the log-density of the transformed sample. This method can fix the boundary issues in a natural way and able to cope with unbounded copula densities. Recently, [6] with a comprehensive simulation study has shown that $\mathscr{L}\mathscr{L}\mathscr{P}\mathscr{T}$ method for copula density estimation yields very good.

Around $(s,t) \in R^2$ and (s', t') close to (s,t), the local log-quadratic likelihood estimation of $\log f(s,t)$ from the pseudo-transformed sample is defined as:

$$log f(s', t') = a_{2,0}(s,t) + a_{2,1}(s,t)(s' - s) + a_{2,2}(s,t)(t' - t)$$

[1] Φ^{-1} is the inverse of the standardized univariate Gaussian distribution and Φ_2 is the standardized bivariate Gaussian distribution with correlation parameter θ.

$$+a_{2,3}(s,t)(s'-s)^2+a_{2,4}(s,t)(t'-t)^2+a_{2,5}(s,t)(s'-s)(t'-t)$$
$$\equiv P_{a_2}(s'-s,t'-t).$$

The vector $a_2(s,t) \equiv (a_{2,0}(s,t),\cdots,a_{2,5}(s,t))$ is then estimated by

$$\hat{a}_2(s,t) = arg\ \max_{a_2}\Big\{\sum_{i=1}^{n}\mathbf{K}\Big(\mathbf{H}_{ST}^{-\frac{1}{2}}\binom{s-\hat{S}_i}{t-\hat{T}_i}\Big)P_{a_2}(\hat{S}_i-s,\hat{T}_i-t)$$
$$-n\int_{R^2}\mathbf{K}\Big(\mathbf{H}_{ST}^{-\frac{1}{2}}\binom{s-s'}{t-t'}\Big)exp(P_{a_2}(s'-s,t'-t))ds'dt'\Big\}.$$

Therefore, the estimation of $f(s,t)$ is $\tilde{f}^p(s,t) = \exp\{\hat{a}_2(s,t)\}$ and thus $\mathscr{LLPT}$ estimator of a copula density is

$$\hat{c}_n^{(\mathscr{LLPT})}(u,v) = \frac{\tilde{f}^p(\Phi^{-1}(u),\Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}, \qquad (u,v)\in[0,1]^2. \tag{4}$$

When the underlying density is on $[0,1]^2$, the performance of the kernel estimator depends on the choice of the kernel function and the bandwidth (smoothing parameter). For bandwidth choice, a practical approach is to consider the minimization of the AMISE on the level of the transformed data. In this article, the bandwidth choice based on nearest-neighbor method; see [4].

4 Estimation of copula based stress-strength

It is natural to consider nonparametric methods especially when the data analyst is unsure about the specification of margins and copula. In this section, we propose a combination of Monte Carlo and bivariate kernel copula density estimation to obtain a nonparametric estimate of the reliability.

Let $\{(X_i,Y_i)\}_{i=1,\cdots,n}$ be a random sample of size n from dependent variabels X and Y. The empirical estimator of R based on these observations is

$$\hat{R} = \frac{1}{n}\sum_{i=1}^{n} I_{[Y_i<X_i]}. \tag{5}$$

The reliability for dependent X and Y can be rewritten based on copula density as

$$R = P(Y < X) = \int_{-\infty}^{+\infty} \int_{-\infty}^{x} f_{X,Y}(x,y) dy dx$$
$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{x} c(F_X(x), F_Y(y); \theta) f_X(x) f_Y(y) dy dx.$$

By considering $U = F_X(X)$ and $V = F_Y(V)$,

$$R = \int_{0}^{1} \int_{0}^{F_X^{-1}(u)} c(u,v;\theta) dv du. \tag{6}$$

Thus, the Kernel estimation of R can be presented as

$$\tilde{R} = \int_{0}^{1} \int_{0}^{u} \hat{c}_n^{(\mathscr{L}\mathscr{L}\mathscr{P}\mathscr{T})}(u,v) dv du, \tag{7}$$

where $\hat{c}_n^{(\mathscr{L}\mathscr{L}\mathscr{P}\mathscr{T})}(\cdot,\cdot)$ is the local likelihood probit transformation estimation of copula density in equation (4).

We summarize the steps for the construction of our proposed nonparametric estimator $\tilde{R}$:

1. Given the data (X_i, Y_i) obtain the pseudo observations $(\tilde{U}_i, \tilde{V}_i)$, where $\tilde{U}_i = R_i/(n+1)$, $\tilde{V}_i = S_i/(n+1)$ for $i = 1, \cdots, n$, and R_i and S_i are the ranks of the observation X_i and Y_i, respectively.

2. Build the estimated copula density with local likelihood probit transformation ($\mathscr{L}\mathscr{L}\mathscr{P}\mathscr{T}$) method.

3. Estimate the reliability by Monte Carlo method according to (7).

5 Simulation study

In this section, a Monto Carlo simulation is presented to illustrate the estimation methods which are described. We demonstrate that the suggested nonoparametric estimator based on local likelihood probit transformation method is efficient than the empirical estimator of R.

Consider, the dependent data (U,V) come from the Clayton, Gumbel and Gaussian copulas with Kendall's tau 0.2, 0.5, and 0.8 that are presented in

Table 1. These copulas cover different dependence structures. Gaussian copula exhibit symmetric and no tail dependence in both lower and upper tails. The Clayton copula exhibits strong left tail dependence and the Gumbel copula has strong right tail dependence. Moreover, 1000 Monte Carlo samples of sizes $n = 100$ and 500 are generated from each type of copulas with maginals normal and exponential (by rate 1 and 2) distributions. The estimators obtained are compared via the Bias and root mean square error (RMSE).

Table 2: Estimated Bias and RMSE of the empirical and kernel estimations for clayton copula

Margianls (X & Y)	n	τ	Empirical esimation		Kernel estimation	
			Bias	RMSE	Bias	RMSE
Normal & Normal	100	0.2	0.5170	0.5269	-0.0150	0.0126
		0.5	0.5229	0.5334	-0.0198	0.0158
		0.8	0.5346	0.5445	-0.0229	0.0186
	500	0.2	0.5032	0.5052	-0.0112	0.0114
		0.5	0.5117	0.5137	-0.0134	0.0103
		0.8	0.5286	0.5303	-0.0182	0.0093
Normal & Exp(2)	100	0.2	0.3079	0.3241	0.1915	0.1134
		0.5	0.3580	0.3710	0.2333	0.1735
		0.8	0.3885	0.4528	0.2704	0.2504
	500	0.2	0.2643	0.2872	0.1623	0.0923
		0.5	0.2946	0.3270	0.2050	0.1350
		0.8	0.3169	0.3997	0.2326	0.1823
Exp(1) & Exp(2)	100	0.2	0.7064	0.7118	-0.2932	0.2012
		0.5	0.8113	0.8151	-0.3602	0.3090
		0.8	0.8610	0.8640	-0.4123	0.3639
	500	0.2	0.6287	0.6399	-0.2016	0.1824
		0.5	0.7051	0.7258	-0.3076	0.2648
		0.8	0.7602	0.7507	-0.3620	0.3285

Results of the simulation study are presented in Tables 2, 3, and 4. These tables present the Bias and RMSE relative to the two estimators of the respective copulas for different values of sample sizes and Kendall's tau and different marginal distributions. The simulation procedure was performed for the positive and negative values of Kendall's tau and according to the symmetry of the obtained results, the results have been reported only for positive values of Kendall's tau. As the results for the sample sizes greater than 500 were in

Table 3: Estimated Bias and RMSE of the empirical and kernel estimations for gumbel copula

Margianls (X & Y)	n	τ	Empirical esimation		Kernel estimation	
			Bias	RMSE	Bias	RMSE
Normal & Normal	100	0.2	0.4315	0.4117	0.0182	0.0120
		0.5	0.4911	0.5029	0.0236	0.0193
		0.8	0.5496	0.5684	0.0293	0.0226
	500	0.2	0.4015	0.3638	0.0142	0.0101
		0.5	0.4273	0.4393	0.0180	0.0163
		0.8	0.4989	0.5023	0.0211	0.0192
Normal & Exp(2)	100	0.2	0.4063	0.3201	0.1947	0.1154
		0.5	0.4655	0.3808	0.2380	0.1862
		0.8	0.5370	0.4516	0.2929	0.2561
	500	0.2	0.3016	0.2845	0.1641	0.0832
		0.5	0.3666	0.3499	0.2176	0.1129
		0.8	0.4487	0.4123	0.2532	0.1985
Exp(1) & Exp(2)	100	0.2	0.7130	0.5189	-0.2263	0.1812
		0.5	0.7933	0.5970	-0.2955	0.2364
		0.8	0.8280	0.6312	-0.3443	0.2978
	500	0.2	0.6561	0.4072	-0.2069	0.1445
		0.5	0.7046	0.4654	-0.2559	0.2012
		0.8	0.7687	0.5294	-0.3026	0.2495

line with our expectation that the increase in sample size will improve the parameter estimation, the corresponding results were omitted from the tables for brevity.

The results show that estimated Bias and RMSE of estimations decrease as sample size increases and estimations improve. The accuracy of the estimations decrease with increasing Kendall's tau. Based on Bias and RMSE, the results show that the empirical esimation ($\hat{R}$) has better performance than the kernel estimation ($\tilde{R}$). Finally, it is necessary to note that although the time required to compute the kernel estimation method is longer than the empirical esimation method, but the kernel estimation method has accurate and acceptable results especially for normal marginal distributions.

Table 4: Estimated Bias and RMSE of the empirical and kernel estimations for gaussian copula

Margianls (X & Y)	n	τ	Empirical esimation		Kernel estimation	
			Bias	RMSE	Bias	RMSE
Normal & Normal	100	0.2	0.6056	0.5166	0.0093	0.0114
		0.5	0.6402	0.5423	0.0131	0.0184
		0.8	0.6968	0.6060	0.0181	0.0223
	500	0.2	0.5445	0.4366	0.0061	0.0084
		0.5	0.5858	0.4673	0.01134	0.0124
		0.8	0.6231	0.5052	0.01572	0.0194
Normal & Exp(2)	100	0.2	0.3044	0.3177	0.1930	0.1831
		0.5	0.3505	0.3643	0.2364	0.2134
		0.8	0.4360	0.4536	0.2834	0.2535
	500	0.2	0.2070	0.2396	0.1629	0.1525
		0.5	0.2607	0.2938	0.2168	0.1961
		0.8	0.3450	0.3476	0.2534	0.2233
Exp(1) & Exp(2)	100	0.2	0.6966	0.7025	-0.2281	0.1923
		0.5	0.7926	0.7967	-0.2842	0.2375
		0.8	0.8339	0.8307	-0.3337	0.2943
	500	0.2	0.6194	0.6407	-0.1980	0.1632
		0.5	0.7162	0.6872	-0.2441	0.2045
		0.8	0.7817	0.7133	-0.3032	0.2564

6 Conclusions

In this paper, we studied the nonparametric estimation of the stress-strength reliability for dependent stress and strength variables based on copulas. The simulation results suggests that the nonparametric estimation of copula based stress-strength models via local likelihood probit transformation method has better performance than the empirical esimation method especially for normal marginal distributions.

References

[1] Charpentier, A., Fermanian, J., and Scaillet, O. (2006), Copulas: From theory to application in finance, chapter The Estimation of Copulas: Theory and Practice, *Risk Books*, Torquay, UK, **1**, 35-62.

[2] de Andrade, B. B., do Nascimento, A. R., and Rathie, P. N. (2020), Parametric and nonparametric inference for the reliability of copula-based stress-strength models, *Quality and Reliability Engineering International*, **36**, 2249-2267.

[3] Domma, F., and Giordano, S. (2013), A copula-based approach to account for dependence in stress-strength models, *Statistical Papers*, **54**, 807-826.

[4] Geenens, G., Charpentier, A., and Paindaveine, D. (2017), Probit transformation for nonparametric kernel estimation of the copula density, *Bernoulli*, **23**, 1848-1873.

[5] Joe, H. (2014). *Dependence modeling with copulas*. Chapman and Hall/CRC.

[6] Nagler, T. (2018), kdecopula: An R Package for the Kernel Estimation of Bivariate Copula Densities, *Journal of Statistical Software*, **84**, 1-22.