Journal of Modern Applied Statistical Methods

Volume 11 | Issue 2 Article 22

11-1-2012

Weighted Cook-Johnson Copula and their Characterizations: Application to Probably Modeling of the Hot Spring Eruptions

Hakim Bekrizadeh
Shahid Chamran University, Ahvas, Iran

Gholam Ali Parham
Shahid Chamran University, Ahvas, Iran

Mohamd Reza Zadkarmi
Shahid Chamran University, Ahvas, Iran

Recommended Citation
Bekrizadeh, Hakim; Parham, Gholam Ali; and Zadkarmi, Mohamd Reza (2012) "Weighted Cook-Johnson Copula and their Characterizations: Application to Probably Modeling of the Hot Spring Eruptions," *Journal of Modern Applied Statistical Methods*: Vol. 11 : Iss. 2 , Article 22.
DOI: 10.22237/jmasm/1351743660

Journal of Modern Applied Statistical Methods
November 2012, Vol. 11, No. 2, 517-523

1538 – 9472/12/$95.00

Weighted Cook-Johnson Copula and their Characterizations: Application to Probable Modeling of the Hot Spring Eruptions

Hakim Bekrizadeh Gholam Ali Parham Mohamd Reza Zadkarmi
Shahid Chamran University,
Ahvaz, Iran

Copulas have emerged as a practical method for multivariate modeling. A limited amount of work has been conducted regarding the application of copula-based modeling in context analysis. This study generalizes the Cook-Johnson copula under the appropriate weighted function and provides examples and the properties of the generalized Cook-Johnson copula. Results show that the generalized Cook-Johnson copula is suitable for probable modeling of hot spring eruption.

Key words: Cook-Johnson copula, weighted distribution functions, measures of dependence, hot spring eruption.

Introduction

Research has shown that it is important to consider dependence among variables in studies as opposed to ignoring the dependence structure solely for mathematical simplicity. As De Michele, et al. (2005) showed, among other consequences, failure to take dependence between variables into account may lead to either over- and under- estimation of the parameters of a model. Copulas have emerged as a practical and efficient method for multivariate event analysis (Joe, 1997; Nelsen, 2006). Application of copulas has increased in the hydrological field as evidenced by Genest and Favre (2007), Gebremichael and Krajewski (2007) and others in a special issue of the *Journal of Hydrological Engineering*. An advantage of using copulas is that marginal behaviors and dependence structure can be studied separately. Most copula applications are in bivariate analysis, for example, De Michele and Salvadori (2003; 2004 a, b) applied copula modeling to examine the dependence between storm intensity and duration. In addition, Zhang and Singh (2007) and Kao and Govindaraju (2007) modeled the dependence between peak storm intensity and depth, depth and duration and peak intensity and duration.

One of the most popular parametric families of copulas is the Cook-Johnson (1981) copula family, defined by

$$C^{CJ}(u,v) = [(1-u)^{-\frac{1}{\beta}} + (1-v)^{-\frac{1}{\beta}} - 1]^{-\beta}, \quad \beta > 0. \tag{1}$$

Inverting this copula results in the joint function:

$$H(x,y) = [x+y-1]^{-\beta}, \quad x \ge 1,\ y \ge 1,\ \beta > 0, \tag{2}$$

where X and Y are identical type I Pareto distributions $F(x) = 1 - x^{-\beta}$ and $F(y) = 1 - y^{-\beta}$, respectively (Genest & Rivest, 1993).

This study extends the copula family of Cook-Johnson by considering a weighted function and examines values of dependence. In addition, the new family is described and examples of copulas taken in this family are

Hakim Bekrizadeh, Gholam Ali Parham and Mohamd Reza Zadkarmi are in the Department of Statistics. Email them at: H_bekri@yahoo.com, Parham_g@scu.ac.ir and Zadkarmi_m@scu.ac.ir.

provided. The associated Kendall's τ_k and tail dependence coefficient are also considered and an application of the generalized Cook-Johnson copula in analysis of probable modeling of a hot spring eruption is presented.

Weighted Cook-Johnson Copulas

A new family of generalized Cook-Johnson copula is proposed using a type II Pareto weighted distribution function. The weighted distributions occur when a random sample from an entire population of interest cannot be obtained (as in the tails) or when a random sample is not desired (as in the selection models).

Let X be a random variable with density $\mathrm{f(x)}$ and $\mathrm{w(x)}$ be a non-negative real value function such that $\mathrm{E[w(X)]<\infty}$. The weighted random variable X, for example $\mathrm{X^w}$, then has weighted probability density function (pdf) of

$$\mathrm{f_{X^w}(x)=\frac{w(x)f_X(x)}{E[w(X)]}} \tag{3}$$

(Patil & Rao, 1978; Mahfoud & Patil, 1982).

If X is a type I Pareto distribution and $w(x)=x^{-\alpha}$, $\alpha>0$, is a non-negative real valued function, then the weighted pdf of X is

$$f_{X^w}(x)=(\alpha+\beta)x^{-(\alpha+\beta)-1}, \quad x>1,\ \alpha,\beta>0, \tag{4}$$

and the survival weighted function is

$$F_{X^w}(x)=1-x^{-(\alpha+\beta)}, \quad x>1,\ \alpha,\ \beta>0 \tag{5}$$

Cook-Johnson Copula Theorem

Let (X,Y) be a type I joint Pareto distribution. Under the weight function $W(x,y)=[x+y-1]^{-\alpha}$ for $\alpha>0$, the weighted copula is

$$C_\alpha^{CJ}(u,v)=[(1-u)^{-\frac{1}{\beta+\alpha}}+(1-v)^{-\frac{1}{\beta+\alpha}}-1]^{-(\beta+\alpha)}, \quad 0<\mathrm{u},\ \mathrm{v}<1,\ \alpha,\ \beta>0, \tag{6}$$

and the type I Pareto joint weighted function is

$$H(x,y)=[x+y-1]^{-(\beta+\alpha)}, \quad x,\ y>1,\ \alpha,\ \beta>0 \tag{7}$$

and $\mathrm{X^w}$, $\mathrm{Y^w}$ are two identical weighted type I Pareto random variables with survival weighted functions

$$F_{X^w}(x)=1-x^{-(\alpha+\beta)}, \quad x>1,\ \alpha,\ \beta>0$$

and

$$F_{Y^w}(y)=1-y^{-(\alpha+\beta)}, \quad y>1,\ \alpha,\ \beta>0$$

respectively.

Cook-Johnson Copula Proof

For any pair of random variables (X,Y), it can be shown from (6) that the joint pdf for $\mathrm{(X,Y)}$ is

$$f_{(X,Y)}(x,y)=\beta(\beta+1)(x+y-1)^{-(\beta+2)}, \quad x,\ y>1,\ \beta>0. \tag{8}$$

Under the weight function $\mathrm{W(x,y)=[x+y+1]^{-\alpha}}$ for $\alpha>0$, a type I joint Pareto weighted density function of $\mathrm{(X,Y)^w}$

$$h^w(x,y)=\frac{w(x,y)h(x,y)}{E[w(X,Y)]} \tag{9}$$

is

$$h^w(x,y)=(\alpha+\beta)(\alpha+\beta+1)(x+y-1)^{-(\alpha+\beta)+2}, \quad x,\ y>1,\ \beta,\ \alpha>0 \tag{10}$$

with marginal weighted density, such as (5), for X^w, Y^w and type I joint weighted function

$$H^w(x,y)=[x+y-1]^{-(\beta+\alpha)},\quad x,\ y>1,\ \beta>\alpha>0. \tag{11}$$

The marginal weighted functions F_{X^w} and G_{Y^w} are

$$F_{X^w}(x)=1-x^{-(\beta+\alpha)},\quad x>1$$

and

$$G_{Y^w}(y)=1-y^{-(\beta+\alpha)},\quad y>1.$$

Inverting the weighted functions and employing the version of Corollary 2.3.7 (Nelson, 2006) yields the weighted copula

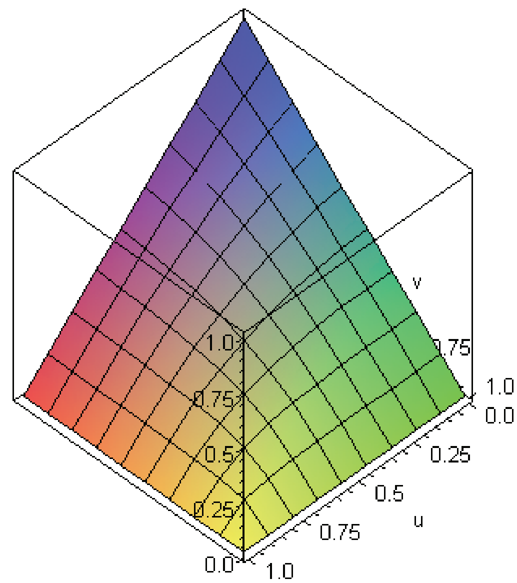
$$C_\alpha^{CJ}(u,v)=[(1-u)^{-\frac{1}{\beta+\alpha}}+(1-v)^{-\frac{1}{\beta+\alpha}}-1]^{-(\beta+\alpha)},\quad 0<u,\ v<1,\ \alpha,\ \beta>0 \tag{12}$$

This relation may be called the weighted Cook-Johnson Copula. The copula density is given by:

$$c_\alpha^{CJ}(u,v)=\frac{\partial^2 C_\alpha^{CJ}(u,v)}{\partial u\partial v}=\left[\frac{\beta+\alpha+1}{\beta+\alpha}\begin{pmatrix}(1-u)^{-\frac{1}{\beta+\alpha}+1}\\ \times(1-v)^{-\frac{1}{\beta+\alpha}+1}\end{pmatrix}\right]\begin{bmatrix}(1-u)^{-\frac{1}{\beta+\alpha}}\\ +(1-v)^{-\frac{1}{\beta+\alpha}}-1\end{bmatrix}^{-(\beta+\alpha)-2}.$$

Figure 1 illustrates the bivariate plots for the extended Cook-Johnson copula; the figure shows a surface bounded within a unit cube that is tied up along the two axes in the first quadrant. After placing different β's and α's, the graph of a generalized copula in terms of different β's and α's did not change. Figure 1 also shows upper correlations at the right in large quantities.

Figure 1: Bivariate Plots of the Extended Cook-Johnson Copulas



Cook-Johnson Copula Corollary

Under the assumption of the Cook-Johnson Copula theorem, the generalized weighted Cook-Johnson Copula for a d-dimensional is

$$C_\alpha^{CJ}(u_1,u_2,\ldots,u_d)=\left(\sum_{i=1}^{d}(1-u_i)^{-\frac{1}{\beta+\alpha}}-d+1\right)^{-(\beta+\alpha)},$$

and for a sub-dimensional it is

$$C_\alpha^{CJ}(u_{k+1},u_{k+2},\ldots,u_d)=\left(\sum_{i=k+1}^{d}(1-u_i)^{-\frac{1}{\beta+\alpha}}-(d-k-1)\right)^{-(\beta+\alpha)}.$$

The generalized copula can be used to make bivariate distributions.

Generalized Weighted Bivariate Beta Distribution Example

If $X \sim beta(\theta,1)$ and $Y \sim beta(\theta,1)$, then $u = x^{\theta}, v = y^{\theta}$ and the generalized weighted bivariate beta distribution is

$$H(x,y) = [(1-x^{\theta})^{-\frac{1}{\alpha+\beta}} + (1-y^{\theta})^{-\frac{1}{\alpha+\beta}} - 1]^{-(\alpha+\beta)},$$
$$0 < x,\ y < 1,\ \alpha,\ \beta,\ \theta > 0.$$

Note that, when $\theta = 1$, the generalized weighted bivariate uniform distribution becomes:

$$H(x,y) = [(1-x)^{-\frac{1}{\alpha+\beta}} + (1-y)^{-\frac{1}{\alpha+\beta}} - 1]^{-(\alpha+\beta)},$$
$$0 < x,\ y < 1,\ \alpha,\ \beta > 0.$$

Generalized Weighted Bivariate Weibull Distribution Example

If $X \sim \mathrm{W}(\theta,\lambda)$ and $Y \sim \mathrm{W}(\theta,\lambda)$, then $u = 1 - e^{-\lambda x^{\theta}}$, $v = 1 - e^{-\lambda y^{\theta}}$ and the generalized weighted bivariate Weibull distribution is

$$H(x,y) = [(1-e^{-\lambda x^{\theta}})^{-\frac{1}{\alpha+\beta}} + (1-e^{-\lambda y^{\theta}})^{-\frac{1}{\alpha+\beta}} - 1]^{-(\alpha+\beta)},$$
$$0 < x,\ y,\ \alpha,\ \beta,\ \theta,\ \lambda.$$

It should be noted that, when $\theta = 1$, the generalized weighted bivariate exponential distribution is obtained as

$$H(x,y) = [(1-e^{-\lambda x})^{-\frac{1}{\alpha+\beta}} + (1-e^{-\lambda y})^{-\frac{1}{\alpha+\beta}} - 1]^{-(\alpha+\beta)},$$
$$0 < x,\ y,\ \alpha,\ \beta,\ \lambda;$$

And, when $\theta = 2$, by replacing $\frac{1}{2\lambda}$ as opposed to λ, the generalized weighted bivariate Rayleigh distribution is obtained as

$$H(x,y) = [(1-e^{-\frac{x^2}{2\lambda}})^{-\frac{1}{\alpha+\beta}} + (1-e^{-\frac{y^2}{2\lambda}})^{-\frac{1}{\alpha+\beta}} - 1]^{-(\alpha+\beta)},$$
$$0 < x,\ y,\ \alpha,\ \beta,\ \lambda.$$

Calculating Measures of Dependence

Copulas can be used in the study of dependence between random variables in many different ways. For a historical review of association measures and concepts of independence, see Gebremichael and Krajewski (2007), Genest and Rivest (1993) and Hutchinson and Lai (1990).

Kendall's τ_k

If X and Y are continuous random variables with copula C, then the population version of Kendall's tau for X and Y (denoted by τ_k) is given by

$$\tau_k = 4\int_0^1\int_0^1 C(u,v)\,dC(u,v) - 1, \tag{13}$$

where C is the copula associated to (X,Y). Thus, the Kendall's τ_k of the generalized Cook-Johnson copula is given by:

$$\tau_k = 1 - \frac{2}{2(\beta+\alpha)+1}.$$

Tail Dependence

The concept of tail dependence relates to the amount of dependence in the upper-right quadrant tail or lower-left-quadrant tail of a bivariate distribution (Farlie, 1960). It is a concept that is relevant for the study of dependence between extreme values. Tail dependence between two continuous random variables *X* and *Y* is a copula property, hence, the amount of tail dependence is invariant under strictly increasing transformations of *X* and *Y*.

Definition 1

If a bivariate copula C is such that

$$L_U = \lim_{u\to 1}\frac{1-2u+C(u,u)}{1-u} \tag{14}$$

exists, then C has upper tail dependence if $L_U \in (0,1]$ and upper tail independence if $L_U = 0$. This measure is used extensively in extreme value theory; it is the probability that one variable is extreme given that the other is extreme, that is, $L_U = P(U > u | V > u)$. Thus L_U may be viewed as a quantile dependent measure of dependence (Coles, Currie & Tawn, 1999).

Definition 2

The concept of lower tail dependence can be defined in a similar way. If the limit

$$L_L = \lim_{u \to 0} \frac{C(u,u)}{u} \tag{15}$$

exists, then C has lower tail dependence if $L_L \in (0,1]$ and lower tail independence if $L_L = 0$. Similarly, lower tail dependence is defined as $L_L = P(U < u | V < u)$. For copulas without a simple closed form an alternative formula for L_L is more useful; thus, the upper tail dependence of the generalized Cook-Johnson copula using (14) is $L_U = 2 - 2^{-\frac{1}{\beta+\alpha}}$, and the lower tail dependence of the generalized Cook-Johnson copula using (15) is $L_L = 0$.

Application of Weighted Cook-Johnson Copula

There is a need for information in the analysis and management of risk for hot spring eruption, the most important of which are the frequency of time between two eruptions and the distance of eruptions. Data used in this study was collected in Yellowstone National Park in 1978 (Weisberg, 1985). Considering the high correlation of these two features, some tools must be used to determine the amount of relationship and impact that exists in the analysis of hot spring eruption; it is necessary to determine the joint distribution of the two features, time interval between two eruptions and distance of an eruption. Because the correlation between the two factors is 0.841 and the hypothesis of independence is not significant, the joint distribution of time interval between two eruptions and distance of an eruption is difficult to obtain via an estimate of marginal distribution. For this reason it is necessary to estimate the marginal distribution of each of the two factors, time interval between two eruptions and distance of an eruption, and between the family of functions selected, the family having the conditions for modeling hot spring eruption.

In most studies in hydrology and geology, exponential, gamma and Weibull distributions are fitted to data so they are used herein and, for data of time interval between two eruptions, the exponential distribution with the parameter $\lambda_1 = 3.464152$ ($Sig. = 0.000$); for distance of an eruption, the exponential distribution with parameter $\lambda_2 = 71.13208$ $Sig. = 0.000$) is a better fit. For estimating each parameter, a copula function is inserted and the obtained function is at a maximum based on the β parameter due to the complexity of the form of density function used for the family and time of plan; therefore, to estimate the β parameter the likelihood function logarithm is defined, for drown β and by limiting the range obtain the likelihood function of maximum logarithm. All related calculations to distribution function, estimating parameter and likelihood function logarithm, are obtained using Maple software.

The parameter estimator for the Cook-Johnson copula function is $\beta = 14.435$ and the maximum value of the likelihood logarithm is −340.920279. Given that sampling devices collect data with some restrictions based on the mechanism used to record observations, it is possible that the proportional is a non-negative function of their size because the observations have a weight distribution. Thus, for data regarding time interval between two eruptions, the type I Pareto distribution with parameters $\mu_1 = 192.35$ and $\sigma_1 = 726.09$ ($Sig. = 0.000$), for distance of an eruption with a type I Pareto distribution with parameters $\mu_2 = 208.01$ and $\sigma_2 = 8940.6$ ($Sig. = 0.000$,) and by a defined weight function with parameters $\alpha = 5.623$ (see theorem), then the maximum estimated parameters for the weighted Cook-Johnson copula function are $\beta = 9.875$ and $\alpha = 5.623$.

The maximum value of the likelihood logarithm for the weighted Cook-Johnson copula is -295.012384. Better copula function is selected according to the method of likelihood maximum, thus, the weighted Cook-Johnson copula function provides more changes for observing samples related to estimated value of Cook-Johnson function. The weighted Cook-Johnson copula function with the parameters $\beta = 9.875$ and $\alpha = 5.623$ can then be used to determine the distribution combined with the time interval between two eruptions and distance of eruption. The estimated value of Kendall's tau is 0.937, the maximum likelihood estimate of the parameters of $\beta = 9.875$ and $\alpha = 5.623$, and for the fitted type I Pareto distribution estimates for the parameters are $\mu_1 = 192.35$, $\sigma_1 = 726.09$, $\mu_2 = 208.01$ and $\sigma_2 = 8940.6$.

In the weighted Cook-Johnson copula the distribution of time interval between two eruptions and distance of eruption is:

$$H(x,y) = \left[\left(\frac{192.35}{x} \right)^{-46.85} + \left(\frac{208.01}{y} \right)^{-576.89} - 1 \right]^{-17.498}$$

and the density is given by

$$h(x,y) = \begin{pmatrix} 1.06\left(\frac{192.35}{x} \right)^{-679.24} \\ \times \left(\frac{208.01}{y} \right)^{-8363.71} \end{pmatrix} \begin{bmatrix} \left(\frac{192.35}{x} \right)^{-46.85} \\ + \left(\frac{208.01}{y} \right)^{-576.89} - 1 \end{bmatrix}^{-17.498}$$

Using the distribution of an eruption, important information can be obtained regarding the eruption of hot spring,s for example, the probability of an eruption of a hot spring with a time of 3.5 hours between two eruptions and a distance of one eruption of 42 meters is 0.0053. Using the copula function and a marginal distribution allows probabilities and other information about eruptions of hot spring and the relationship between the time intervals and distances of eruption to be obtained. Conditional distribution can also be determined by copula to provide a basis for the probabilities of altering factors against controlled changes. This may be useful in understanding and managing the impacts of global warming on hot spring eruptions.

Conclusion

This article proposed a new family of copulas; the generalizing Cook-Johnson family generated by weighted distribution function, and obtained a generalized *d*-dimensional (multivariate) Cook-Johnson Copula. The bivariate weighted distributions and weighted marginal distribution were also generated, and the generalized bivariate distribution was obtained. The main feature of this family of copulas is to permit modeling of variables with high dependence. Moreover, it was shown that the generalizing Cook-Johnson Copula is a proper model for analyzing the problem of eruption of a hot spring. Bivariate distributions and copula functions for two parameters were analyzed. By using research methodology and the multivariate generalizing Cook-Johnson copula function, the issue of probable modeling of hot spring eruption with more variables can be studied.

References

Coles, S., Currie, J., & Tawn, J. (1999). *Dependence measures for extreme value analyses*. Working paper: Department of Mathematics and Statistics, Lancaster University.

Cook-Johnson, D. G. (1981). A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incident. *Biometrika*, *65*, 141-151.

De Michele, C., et al. (2005). Bivariate statistical approach to check adequacy of dam spillway. *Journal of Hydrologic Engineering*, *10*(*1*), 50-57.

Farlie, D. G. J. (1960).The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, *47*, 307-323.

Gebremichael, M., & Krajewski, W. F. (2007). Application of copulas to modeling temporal sampling errors in satellite-derived rainfall estimates. *Journal of Hydrologic Engineering*, *12*(*4*), 404-408.

Genest, C., & Rivest, L. (1993). Statistical Inference Procedures for bivariate Archimedean Copulas. *Journal of the American Statistical Association*, *88*, 1034-1043.

Genest, C., & Favre, A. C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrological Engineering*, **12**, 347–368.

Hutchinson, T. P., & Lai, C. D. (1990). *Continuous bivariate distributions, emphasising applications*. Adelaide: Rumsby Scientific Publishing.

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

Kao, S. C., & Govindaraju, R. S. (2007). A bivariate frequency analysis of extreme rainfall with implications for design. *Journal of Geophysical Research*, *Atmospheres*.

Mahfoud, M., & Patil, G. P. (1982). On weighted distributions. In *Statistics in theory and Practice: Essays in Honor of C. R. Rao*, G. Kallianpur, et al. (Eds.), 479-492. Amsterdam: North-Holland.

Nelsen, R. B. (2006). *An Introduction to Copulas*, *2nd Ed*. New York, NY: Springer-Verlag, Inc.

Patil, G. P., & Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, *34*,179-89.

Salvadori, G., De Michele, C. (2004a). Analytical calculation of storm volume statistics involving Pareto-like intensity-duration marginals. *Geophysical Research Letters*, *31*, L04502.1-L04502.4.

Salvadori, G., De Michele, C. (2004b). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, *40*, W12511.1-W12511.17.

Schweizer, B., & Sklar, A. (1983). Probabilistic Metric Spaces. Elsevier New York.

Weisberg S. (1985). *Applied linear regression*. New York, NY: Wiley.

Zhang, L., & Singh, V. P. (2007). Bivariate rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology*, *332*(*1-2*), 93-109.