

Weighted Clayton Copulas and their Characterizations: Application to Probable Modeling of the Hydrology Data

Hakim Bekrizadeh*, Gholam Ali Parham and Mohammad Reza Zadkarmi
Shahid Chamran University

Abstract: Copulas have recently emerged as practical methods for multivariate modeling. To our knowledge, only a limited amount of work has been done to apply copula-based modeling in context analysis. In this study, we generalized Clayton copula under the appropriate weighted function. In some examples, bivariate distributions by using the weighted Clayton copula are generalized. Also the properties of generalized Clayton copula are provided. The Clayton copula and weighted Clayton model cannot be used for negative dependence. These have been used to study left tail dependence. This property is stronger in weighted Clayton model with respect to ordinary Clayton copula. It will also be shown that the generalized Clayton copula is suitable for the probable modeling of the hydrology data.

Key words: Clayton copula, measures of dependence, the hydrology data, weighted distribution functions.

1. Introduction

There exists plenty of evidence for dependence among variables, ignoring the dependence structure solely for mathematical simplicity. As shown by De Michele *et al.* (2005), among other consequences, failure to take into account the dependence between variables may lead to overestimation and underestimation of the parameters of the model.

Copulas have recently emerged as practical and efficient methods for analysis of multivariate events (e.g., Joe, 1997; Nelsen, 2006). Hydrological application of copulas has surged (see, e.g., Genest and Favre, 2007; Gebremichael and Krajewski, 2007 and others in a special issue of Journal of Hydrological Engineering). An important advantage of using copulas is that the marginal behavior and the dependence structure can be studied separately. Copula modeling has been applied in design storm analysis recently, with most applications in bivariate analysis. De

*Corresponding author.

Michele and Salvadori, in a sequence of papers (De Michele and Salvadori, 2005; Salvadori and De Michele, 2004a, b), modeled the dependence between storm intensity and duration.

Zhang and Singh (2007) and Kao and Govindaraju (2007) modeled the dependence between peak intensity and depth, depth and duration, and peak intensity and duration. One of the most popular parametric families of copulas is the Clayton copula family, defined as

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

Clayton (1978) examines this family in detail. Inverting this survival copula, we obtain a joint survival function

$$\bar{H}(x, y) = (x + y + 1)^{-\beta}, \quad x \geq 0, \quad y \geq 0, \quad \beta > 0, \quad (2)$$

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

This assay aims at extending family copula of Clayton by considering the weighted function. The weighted distributions arise when a random sample from the entire population of interest cannot be obtained (as in the tails) or it is not desired (as in the selection models). An important advantage of using weighted copulas is that the marginal behavior and the dependence structure can be studied separately. This study applies the weighted copula modeling in design hydrology analysis, with most applications in bivariate analysis, for example, the dependence modeled between storm intensity and duration, peak intensity and depth, depth and duration, and peak intensity and duration. It also attempts to study values of dependence. Then it describes the new family and provides some examples of copulas taken in this family, in Section 2. The associated Kendall's τ_k and tail dependence coefficient are studied in Section 3. The application of the generalized Clayton copula in the analysis of the probable modeling of the hydrology data will be studied in Section 4. And finally, the results of this research will be carried out in Section 5.

2. Weighted Clayton Copula

The density $f(x)$ of a potential observation can be distorted so that it may be multiplied by some non-negative weight function $W(x)$ involving parameters (Patil and Rao, 1978; Mahfoud and Patil, 1982). Thus the observed data is a random sample from the following weighted version of $f(x)$. Now; we propose a new family of the generalized Clayton copula by using weighted distribution function of type II Pareto distribution. The weighted distributions arise when a random sample from the entire population of interest cannot be obtained (as in

the tails) or it is not desired (as in the selection models). Let X be a random variable with density $f(x)$ and $w(x)$ be non-negative real valued function so that $E[w(X)] < \infty$. Then weighted random variable X , say X^w , will have a weighted pdf of

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

Now let X be type II Pareto distribution and $w(x) = (1 + x)^{-\alpha}$, $\alpha \geq 0$, be non-negative real valued function. Then the weighted pdf of X is

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

and weighted distribution function is

$$\bar{F}_{X^w}(x) = (1 + x)^{-(\alpha+\beta)}, \quad x \geq 0, \alpha \geq 0, \beta > 0. \tag{5}$$

Theorem. Let (X, Y) be a type II joint Pareto survival distribution. Then under weight function $W(x, y) = [x + y + 1]^{-\alpha}$ for $\beta > 0$, $\alpha \geq 0$, the weighted Clayton copula is

$$C_w^C(u, v) = [u^{-(\beta+\alpha)} + v^{-(\beta+\alpha)} - 1]^{-\frac{1}{\beta+\alpha}}, \quad \alpha \geq 0, \beta > 0, \tag{6}$$

and weighted type II Pareto joint survival function is

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

where X^w, Y^w are two identical weighted type II Pareto random variables with survival weighted function $\bar{F}_{X^w}(x) = (1 + x)^{-(\alpha+\beta)}$, $x \geq 0$, $\alpha \geq 0$, $\beta > 0$ and $\bar{F}_{Y^w}(y) = (1 + y)^{-(\alpha+\beta)}$, $y \geq 0$, $\alpha \geq 0$, $\beta > 0$, respectively.

Proof. For any pair random variables (X, Y) , from (2) it is easy to show that joint pdf for (X, Y) is

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

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

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

is

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

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

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

Then the marginal weighted survival functions \bar{F}_{X^w} and \bar{G}_{Y^w} are $\bar{F}_{X^w}(x) = (1 + x)^{-(\beta+\alpha)}$, $x \geq 0$ and $\bar{G}_{Y^w}(y) = (1 + y)^{-(\beta+\alpha)}$, $y \geq 0$ respectively. Let $\varphi(t) = t^{-\beta} - 1$ for t in $[0, 1]$ and $\theta > 0$. Because $\varphi(0) = \infty$, φ is strict. Thus $\varphi^{[-1]}(t) = (1 + t)^{-1/\beta}$, now by using Lemma 4.1.2 (Nelsen, 2006), we have

$$C_w^C(u, v) = [u^{-(\beta+\alpha)} + v^{-(\beta+\alpha)} - 1]^{-\frac{1}{\beta+\alpha}}. \quad (12)$$

Relation (12) is called the weighted Clayton copula.

The copula density (12) is provided by

$$\begin{aligned} c_W^C(u, v) &= \frac{\partial^2}{\partial u \partial v} C_W^C(u, v) \\ &= (\beta + \alpha + 1) u^{-(\beta+\alpha-1)} v^{-(\beta+\alpha-1)} [u^{-(\beta+\alpha)} + v^{-(\beta+\alpha)} - 1]^{-\frac{1}{\beta+\alpha}-2}. \end{aligned}$$

Corollary. The generalized weighted Clayton copula for d -dimensional is

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

and for a sub-dimensional of the above relation we have

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

Proof. Let $\varphi_{\alpha+\beta}(t) = t^{-(\alpha+\beta)} - 1$ for $\alpha + \beta > 0$, which generates a subfamily of the bivariate weighted Clayton family, the subfamily whose generators are strict. Here $\varphi_{\alpha+\beta}^{[-1]}(t) = (1 + t)^{-1/(\alpha+\beta)}$, this is easily shown to be complete monotone on $[0, \infty)$. Thus we can generalize the weighted Clayton family of 2-copulas to a family of d -copulas for $\alpha + \beta > 0$ and any $d \geq 2$.

One of the advantages of the weighted Clayton model is that we can use this model, and generalize bivariate weighted distributions. Some examples are mentioned in the following sections.

Example 2.1. (Generalized weighted bivariate beta distribution).

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

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

Note 1. When $\theta = 1$, then the generalized weighted bivariate uniform distribution becomes

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

Example 2.2. (Generalized weighted bivariate Weibull distribution). Let $X \sim W(\theta, \lambda)$, $Y \sim W(\theta, \lambda)$ then $u = 1 - e^{-\lambda x^\theta}$, $v = 1 - e^{-\lambda y^\theta}$ and the generalized weighted bivariate Weibull distribution will be

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

Note 2. When $\theta = 1$, the generalized weighted bivariate exponential distribution obtained is

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

Note 3. When $\theta = 2$ by replacing $1/2\lambda$ instead of λ , the generalized weighted bivariate Rayleigh distribution obtained will be obtained as

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

3. Calculating Measures of Dependence

In this section we will look at different ways in which copulas can be used in the study of dependence between random variables. 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).

3.1 Kendall's τ_k

Let X and Y be continuous random variables whose copula is C . Then the population version of Kendall's tau for X and Y (which we will denote by τ_k) will be 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) . So, the Kendall's τ_k of the generalized Clayton copula will be given by

$$\tau_k = \frac{\beta + \alpha}{\beta + \alpha + 2}.$$

3.2 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. It turns out that tail dependence between two continuous random variables X and Y is a copula property and hence, the amount of tail dependence is invariant under strictly increasing transformations of X and Y .

Definition 1. Given that concerning a bivariate copula C , we have

$$L_U = \lim_{u \rightarrow 1} \frac{1 - 2u + C(u, u)}{1 - u}. \quad (14)$$

Will have C has upper tail dependence if $L_U \in (0, 1]$, and upper tail independence if $L_U = 0$.

Definition 2. The concept of lower tail dependence can be defined in a similar way. If the limit,

$$L_L = \lim_{u \rightarrow 0} \frac{C(u, u)}{u}. \quad (15)$$

exists, then C has lower tail dependence if $L_L \in (0, 1]$, and lower tail independence if $L_L = 0$. For copulas without a simple closed form an alternative formula for L_L is more useful. So, the upper tail dependence of the generalized Clayton copula by using (14) is, $L_U = 0$ and the lower tail dependence of the generalized Clayton copula by using (15) is, $L_L = 2^{-1/(\beta+\alpha)}$.

It should be noted that $2^{-1/(\beta+\alpha)}$ for fixed β is increasing function in α , therefore the lower tail dependence parameter in weighted Clayton copula is greater than tail dependence in Clayton copula.

4. Application of Weighted Clayton Copula

Information is needed in analysis and management of water resources, the most important part of which is the study of features frequency of drainage area and discharge. Our data are for 22 watersheds in the Western Coastal Plain of Maryland and Virginia. The data include the drainage area (A , miles²), and discharge (Q , cfs) (McCuen Richard H., p.122, 2003). Considering the high correlation of these two features, some tools must be used to reveal the amount of relationship and impact which exists in the analysis; therefore it is necessary to determine the joint distribution of the two features, drainage area and discharge. Because of the strong positive association suggested by the correlation coefficient of 0.933, we reject independent assumption between two variables ($sig. = 0.000$).

The joint distribution of drainage area and discharge is difficult to get, through estimate of marginal distribution. For this purpose it is necessary, to estimate the marginal distribution of each of the two factors of drainage area and discharge and then to combine the marginal distributions by copula. In this paper distributions fitted to the data are Weibull, Gamma, Gamma ($3p$), Pearson 6, log-Pearson 3, log-logistic ($3p$), Pareto type II, lognormal, lognormal ($3p$), exponential, exponential ($2p$), and finally, for data of drainage area, the type II Pareto distribution with the parameters of $\mu_1 = 3.4566$ and $\sigma_1 = 46.352$, and for data of discharge the type II Pareto distribution with parameters of $\mu_2 = 4.9481$ and $\sigma_2 = 6336.2$ will better fit.

The results of fit type II Pareto distribution to data sets are illustrated in Table 1.

Table 1: The results of fit type II Pareto distribution to data

Test: Type II Pareto distribution for drainage area data		
Kolmogorov-Smirnov	Parameters	
Sample Size	22	$\mu_1 = 3.4566, \sigma_1 = 46.352$
Statistic	0.12724	
p -value	0.82512	
Result: Accept Type II Pareto distribution for drainage area data		
Test: Type II Pareto distribution for discharge data		
Kolmogorov-Smirnov	Parameters	
Sample Size	22	$\mu_1 = 4.9481, \sigma_1 = 6336.2$
Statistic	0.12644	
p -value	0.83064	
Result: Accept Type II Pareto distribution for drainage area data		

Recall. We have fitted the below Pareto (second kind) distribution to two sets of data drainage area and discharge

$$F(x) = 1 - \left(\frac{\sigma}{x + \sigma} \right)^\mu, \quad 0 \leq x < \infty.$$

$\mu =$ continuous shape parameter ($\mu > 0$) and $\sigma =$ continuous scale parameter ($\sigma > 0$) (In some cases, the lower range x is the third parameter which we have not used in this paper). So, the distributions for two sets of data drainage area and discharge have two parameters μ and σ , called shape parameter and scale parameter, respectively.

For estimating parameters of Clayton copula and weighted Clayton copula the log-likelihood function is defined. All the related calculations to distribution

function, estimating parameter and logarithm likelihood function are done by Maple software. The MLE of parameter for the density of Clayton copula is $\hat{\beta}_{ML,0} = 2.9494$ and the MLE of parameters for the density of weighted Clayton copula are $\hat{\beta}_{ML,1} = 2.0215$ and $\hat{\alpha}_{ML,1} = 1.3754$ respectively. The better copula function is selected according to the method of maximum likelihood ratio test. The log-likelihood for the density of Clayton copula is -267.2240572 and the log-likelihood for the density of weighted Clayton copula is -229.1154301 . When one parameter of interest is on the boundary and no nuisance parameters are on the boundary, the asymptotic distribution of the LRT is a 50 : 50 mixture of point mass at 0 and χ^2 with 1 degree of freedom (Self and Liang, 1987), so

$$-2 \ln[L(\hat{\alpha}_{ML,0}, \hat{\beta}_{ML,0})/L(\hat{\alpha}_{ML,1}, \hat{\beta}_{ML,1})] = 76.217 \quad (sig. = 0.000),$$

where $L(\hat{\alpha}_{ML,0}, \hat{\beta}_{ML,0})$ and $L(\hat{\alpha}_{ML,1}, \hat{\beta}_{ML,1})$ are the maximum likelihood under the null-hypothesis ($H_0 : \alpha = 0$) and under the alternative hypothesis ($H_1 : \alpha > 0$), respectively.

Thus, the weighted Clayton copula function provides more changes for observing samples related to estimated value of Clayton function. So the weighted Clayton copula function with the parameter of $\hat{\beta}_{ML,1} = 2.0215$ and $\hat{\alpha}_{ML,1} = 1.3754$ can be used to determine the joint distribution between drainage area and discharge. Using marginal type II Pareto distributions and the MLE of parameters and replacing in the weighted Clayton copula, we will have the distribution of drainage area and discharge as follows:

$$H(x, y) = \left[\left(1 - \left(\frac{46.352}{x + 46.352} \right)^{3.4566} \right)^{-3.396} + \left(1 - \left(\frac{6336.2}{y + 6336.2} \right)^{4.9481} \right)^{-3.396} - 1 \right]^{-0.2945}.$$

And the density is provided by

$$h(x, y) = 4.396 \left(\left(1 - \left(\frac{46.352}{x + 46.352} \right)^{3.4566} \right) \left(1 - \left(\frac{6336.2}{y + 6336.2} \right)^{4.9481} \right) \right)^{-2.396} \\ \times \left[\left(1 - \left(\frac{46.352}{x + 46.352} \right)^{3.4566} \right)^{-3.396} + \left(1 - \left(\frac{6336.2}{y + 6336.2} \right)^{4.9481} \right)^{-3.396} - 1 \right]^{-2.0684}.$$

Using this distribution, one can obtain important information on the drainage area and discharge, for example, the probability of a drainage area less than 2.82

and a discharge less than specific 418 is 0.17213. Also, conditional distribution can be determined by copula, and then the probabilities of how to change a factor according to another controlled factor can be obtained.

5. Conclusions

In this paper, we proposed a new family of copulas, namely, the generalizing Clayton family that is generated by weighted distribution function and then, we obtained a generalized d -dimensional (Multivariate) Clayton copula. One of the advantages of the weighted Clayton model is its usability and generalizability to bivariate weighted distributions. This family obtains generalized bivariate distributions. The main feature of this family is permitting the modeling of high dependence, while the Clayton copula and weighted Clayton model cannot be used for negative dependence. These have been used to study lower tail dependence which as a parameter in weighted Clayton copula is greater than lower tail dependence in Clayton copula. Moreover, it has also been shown that the generalizing Clayton copula is a proper model for analyzing the hydrology data problem. It should also be noted that bivariate distributions and two-parameter copula function were analyzed. Besides, using research methodology and the multivariate generalizing Clayton copula function, one can study the issue of probable modeling of hydrology data with more variables.

Acknowledgements

The authors are grateful to the editor and the referee for taking the trouble of reviewing this paper and offering many helpful comments.

References

- Clayton, D. G. (1978). 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., Salvadori, G., Canossi, M., Petaccia, A. and Rosso, R. (2005). Bivariate statistical approach to check adequacy of dam spillway. *Journal of Hydrologic Engineering* **10**, 50-57.
- Farlie, D. G. J. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* **47**, 307-323.
- Gebremichael, M. and Krajewski, W. F. (2007). Application of copulas to modeling temporal sampling errors in satellite-derived rainfall estimates. *Journal of Hydrologic Engineering* **12**, 404-408.

- Genest, C. and 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.
- Genest, C. and Rivest, L. P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association* **88**, 1034-1043.
- Hutchinson, T. P. and Lai, C. D. (1990). *Continuous Bivariate Distributions*. Emphasising Applications, Rumsby Scientific Publishing, Adelaide.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.
- Kao, S. C. and Govindaraju, R. S. (2007). A bivariate frequency analysis of extreme rainfall with implications for design. *Journal of Geophysical Research* **112**, D13119, doi:10.1029/2007JD008522.
- McCuen, R. H. (2003). *Modeling Hydrologic Change: Statistical Methods*. Lewis Publishers is an imprint of CRC Press LLC, United States of America.
- Mahfoud, M. and Patil, G. P. (1982). On weighted distributions. *Statistics in Theory and Practice*. Essays in Honor of C. R. Rao (Edited by G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh), 479-492. North-Holland Amsterdam.
- Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd edition. Springer-Verlag, New York.
- Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics* **34**, 179-189.
- Salvadori, G. and 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. and De Michele, C. (2004b). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research* **40**, W12511.1-W12511.17.
- Schweitzer, B. and Sklar, A. (1983). *Probabilistic Metric Spaces*. Elsevier North-Holland, New York.

Self, S. G. and Liang, K. Y.(1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of American Statistical Association* **82**, 605-610.

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

Received February 18, 2012; accepted November 23, 2012.

Hakim Bekrizadeh
Department of Statistics
Shahid Chamran University
Golestan Boulevard, Ahvaz, Iran
H.bekri@yahoo.com

Gholam Ali Parham
Department of Statistics
Shahid Chamran University
Golestan Boulevard, Ahvaz, Iran
Parham_g@scu.ac.ir

Mohammad Reza Zadkarmi
Department of Statistics
Shahid Chamran University
Golestan Boulevard, Ahvaz, Iran
Zadkarmi_m@scu.ac.ir