A New Generalized Gumbel Copula for Multivariate Distributions

Chandra R. Bhat*

The University of Texas at Austin
Department of Civil, Architectural & Environmental Engineering
1 University Station, C1761, Austin, TX 78712-0278
Phone: (512) 471-4535, Fax: (512) 475-8744
Email: bhat@mail.utexas.edu

*corresponding author

August 2009

ABSTRACT

This paper constructs a new generalized multivariate version of the Gumbel copula that, to our knowledge, has not appeared in the statistical or mathematical literature.

A copula is a device or function that generates a stochastic dependence relationship (*i.e.*, a multivariate distribution) among random variables with pre-specified marginal distributions. In essence, the copula approach separates the marginal distributions from the dependence structure, so that the dependence structure is entirely unaffected by the marginal distributions assumed. This provides substantial flexibility in developing dependence among random variables (see Bhat and Eluru, 2009; Trivedi and Zimmer, 2007).

The precise definition of a copula is that it is a multivariate distribution function defined over the unit cube linking uniformly distributed marginals. Let C be a K-dimensional copula of uniformly distributed random variables U_1, U_2, U_3, ..., U_K with support contained in $[0,1]^K$. Then,

$$C_\theta(u_1, u_2, \ldots, u_K) = \Pr(U_1 < u_1, U_2 < u_2, \ldots, U_K < u_K), \tag{1}$$

where θ is a parameter vector of the copula commonly referred to as the dependence parameter vector. A copula, once developed, allows the generation of joint multivariate distribution functions with given marginals. Consider K random variables Y_1, Y_2, Y_3, ..., Y_K, each with univariate continuous marginal distribution functions $F_k(y_k) = \Pr(Y_k < y_k)$, $k = 1, 2, 3, \ldots, K$. Then, by Sklar's (1973) theorem, a joint K-dimensional distribution function of the random variables with the continuous marginal distribution functions $F_k(y_k)$ can be generated as follows:

$$F(y_1, y_2, \ldots, y_K) = \Pr(Y_1 < y_1, Y_2 < y_2, \ldots, Y_K < y_K) = \Pr(U_1 < F_1(y_1), U_2 < F_2(y_2), \ldots, U_K < F_K(y_K))$$
$$= C_\theta(u_1 = F_1(y_1), u_2 = F_2(y_2), \ldots, u_K = F_K(y_K)). \tag{2}$$

Conversely, by Sklar's theorem, for any multivariate distribution function with continuous marginal distribution functions, a unique copula can be defined that satisfies the condition in Equation (2). Thus, given a known multivariate distribution $F(y_1, y_2, \ldots, y_K)$ with continuous margins $F_k(y_k)$, the inversion method may be used to obtain a copula using Equation (2) (see Nelsen, 2006):

$$C_\theta(u_1, u_2, \ldots, u_K) = \Pr(U_1 < u_1, U_2 < u_2, \ldots, U_K < u_K)$$
$$= \Pr(Y_1 < F^{-1}{}_1(u_1), Y_2 < F^{-1}{}_2(u_2), \ldots, Y_3 < F^{-1}{}_3(u_3)) \tag{3}$$
$$= F(y_1 = F^{-1}{}_1(u_1), y_2 = F^{-1}{}_2(u_2), \ldots, y_K = F^{-1}{}_k(u_k)).$$

Once the copula is developed, one can revert to Equation (2) to develop new multivariate distributions with arbitrary univariate margins.

A rich set of bivariate copula types have been generated using inversion and other methods, including the Gaussian copula, the Farlie-Gumbel-Morgenstern (FGM) copula, and the Archimedean

class of copulas (including the Clayton, Gumbel, Frank, and Joe copulas). Of these, the Gaussian and FGM copulas can be extended to more than two dimensions in a straightforward manner, allowing for differential dependence patterns among pairs of variables. In fact, the multivariate normal distribution used in the spatial probit model corresponds to the Gaussian copula with univariate normal distributions. Recently, Bhat and Sener (2008) proposed the use of the FGM copula with univariate logistic distributions for spatial modeling in a binary choice context, but point out that the maximal correlation allowable between pairs of variables is 0.303. In any case, the Gaussian and FGM copulas assume the property of asymptotic independence. That is, in these copulas, a positive (negative) correlation gets manifested as clustering along the southwest-to-northeast (northwest-to southeast) plane close to the center point of the joint distribution. However, toward extreme tails, there is scattering or dependence reduction (see Bhat and Eluru, 2009 for a graphical illustration of this point). Further, the dependence structure is radially symmetric about the center point in the Gaussian and FGM copulas. That is, for a given correlation, the level of dependence is equal in the upper and lower tails. On the other hand, extreme tail dependence and asymmetric tail dependence may be important characteristics in spatial data (even after conditioning each marginal variable in terms of observed covariates). For instance, closely located neighborhoods may simultaneously experience high crime rates or high trip rates, but not necessarily low crime rates or low trip rates. This is the case of strong right tail dependence (strong correlation at high values) but weak left tail dependence (weak correlation at low values).

Clearly, using other copulas to characterize extreme and asymmetric tail dependence would be useful. Unfortunately, beyond the Gaussian and FGM copulas, extending other copulas to a multivariate setting with differential dependence between pairs of variables is not straightforward. The Archimedean copulas can be extended to multiple dimensions, but such extensions, for the most part, do not allow differential dependence between pairs of variables (in fact, the straightforward way to extend Archimedean copulas restricts the dependence parameter between all pairs of variables to be identical, a restriction that is clearly not appropriate for spatial dependence situations; see Cherubini, 2004). But one can construct a generalized version of the Gumbel copula that allows different dependence parameters between each variable pair using the inversion method. To derive such a copula, consider the following cumulative multivariate extreme-value distribution (see Bhat and Guo, 2004):

$$F(y_1, y_2, \ldots, y_q, \ldots, y_Q) = \exp\left\{ -\sum_{q=1}^{Q-1} \sum_{k=q+1}^{Q} \left[(\alpha_{q,qk} e^{-y_q})^{1/\rho} + (\alpha_{k,qk} e^{-y_k})^{1/\rho} \right]^{\rho} \right\} \tag{4}$$

where $0 \leq \alpha_{q,qk} \leq 1$ for all q and k, $0 < \rho \leq 1$, and $\sum_k \alpha_{q,qk} = 1$ for all q ($\alpha_{q,qq} = 0$ for all q by convention). The marginal distribution of each element y_q is univariate extreme value as follows:

$$\begin{aligned} F_q(y_q) &= \exp\left\{ -\sum_{k \neq q} \alpha_{q,qk} e^{-y_q} \right\} \\ &= \exp\left\{ -e^{-y_q} \right\}. \end{aligned} \tag{5}$$

From this, we get:

$$F_q^{-1}(u_q) = -\ln\{-\ln u_q\}. \tag{6}$$

Inserting this in Equation (3), we get the following copula:

$$C(u_1, u_2, \ldots, u_q, \ldots, u_Q) = \exp\left\{ -\sum_{q=1}^{Q-1} \sum_{k=q+1}^{Q} \left[(-\alpha_{q,qk} \ln u_q)^{1/\rho} + (-\alpha_{k,qk} \ln u_k)^{1/\rho} \right]^{\rho} \right\}, \tag{7}$$

$0 \leq \alpha_{q,qk} \leq 1$, $\sum_k \alpha_{q,qk} = 1$ for all q, $0 < \rho \leq 1$.

The reader will note that the copula just derived is a generalized version of the Gumbel copula (1960). The authors are not aware of any earlier derivation and use of such a copula in the statistics/econometrics literature. Note also that the bivariate margin of the GG copula is as follows:

$$C_\theta(u_1, u_2) = \exp[-\{[(-\alpha_{1,12} \ln u_1)^{1/\rho} + (-\alpha_{2,12} \ln u_2)^{1/\rho}] - (1-\alpha_{1,12}) \ln u_1 - (1-\alpha_{2,12}) \ln u_2\}]$$

Bhat, C.R., and N. Eluru (2009) A Copula-Based Approach to Accommodate Residential Self-Selection in Travel Behavior Modeling. *Transportation Research Part B*, 43(7), 749-765.

Bhat, C.R., and I.N. Sener (2008) A Copula-Based Closed-Form Binary Logit Choice Model for Accommodating Spatial Correlation Across Observational Units. Forthcoming, *Journal of Geographical Systems*.

Trivedi, P.K. and D.M. Zimmer (2007) Copula Modeling: An Introduction for Practitioners. *Foundations and Trends in Econometrics*, 1(1), Now Publishers.