A PROCEDURE FOR SIMULATION WITH CONSTRUCTED COPULAS

DONALD F. BEHAN AND SAMUEL H. COX

ABSTRACT. We describe a simple way to construct a bivariate copula with specified marginals and partially specified dependence. For example, our method allows one to model tail behavior scenarios while holding the core of the joint distribution fixed. We provide an *Excel* workbook to illustrate the method, including its use in Monte Carlo simulation.

1. INTRODUCTION

Any multivariate distribution may be constructed from its marginal distributions and a copula. This is described in general terms by Nelsen (1999), which is a good introduction to copulas. Frees, Carriere and Valdez (1996) use copulas to model dependent lifetimes, an early actuarial application of copulas. Venter (2002) surveys copulas for actuarial applications, especially in the non–life fields. McNeil, Frey and Embrechts (2005) describe copulas and how they arise in risk management. Venter et al. (2007) discuss the possibilities for modeling more than two dependent variables with applications to financial models.

While copulas are fundamental to modeling dependence, covariance is more often used in discussion of risk. For example, Gorski, Longley-Cook and Reiskytl (2002) discuss the importance of covariance in risk-based capital models, but they do not mention copulas explicitly. The American Academy of Actuaries reports similarly use the term covariance, but probably do not intend to exclude other kinds of dependence[1]. The covariance concept cannot capture all aspects of dependence, but it seems that the term covariance is often used in a general, non-technical way to describe dependence in general.

The key concept is that the dependence structure of any multivariate distribution is represented by a *copula*. In the case of continuous distributions the copula is unique, however the uniqueness does not play an important role in the simulation process. The representation of a multivariate distribution by a copula and

[1]American Academy of Actuaries (Org.) American Academy of Actuaries Report to NAIC, Correlations for Period 1978-1996. December 1997.

its marginal distributions is totally general, with no restrictions on the marginal distributions or the dependence structure.

A 2-dimensional copula is a distribution function $C(u, v)$ on $\mathbf{I} \times \mathbf{I}$ with standard uniform marginal distributions where $\mathbf{I}$ is the closed unit interval $[0, 1]$. Nelsen (1999, page 8) gives this simple equivalent definition:

A 2-dimensional copula is is a function $C : \mathbf{I} \times \mathbf{I} \longrightarrow \mathbf{I}$ with the following properties:

(i) For all $x, y \in \mathbf{I}$,

$$C(u, 0) = 0 = C(0, v)$$

and

$$C(u, 1) = u \text{ and } C(1, v) = v.$$

(ii) For $u_1, u_2, v_1, v_2 \in \mathbf{I}$ with $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

The condition (ii) above is equivalent to the non-negativity of the probability associated with $[u_1, u_2] \times [v_1, v_2]$. In high dimensions, the analogous condition is much more complicated.

We will also use the fundamental result in the theory of copulas, Sklar's theorem (McNeil, Frey and Embrechts, 2005, page 186) or (Nelsen, 1999, page 15):

If $F(x, y)$ is a joint distribution function with marginals $F_1(x)$ and $F_2(y)$, then there is a copula $C(u, v)$ such that

$$F(x, y) = C(F_1(x), F_2(y)).$$

2. *Excel* EXAMPLE

We have prepared an *Excel* workbook to be used as an educational tool for actuaries who would like to explore the use of copulas for simulation of extreme events. This section provides brief instructions on the use of the workbook. The workbook example is highly simplified in the sense that the levels of probability are limited to a small set of values. In this way the conceptual issues are illustrated while keeping the level of detail within reasonable bounds. The example can be scaled up to any degree without introducing any new conceptual issues. In preparing this workbook our principal reference on copulas was *An Introduction to Copulas* by Nelsen (1999).

The dependence structure of empirical probability distributions is difficult to establish for the extremes of the distribution, because of the rarity of data that could be used for empirical determination of the dependence relationship. Actuaries

encounter situations in which the distributions of important individual variables are known, as well as the dependence structure between the variables for common events, but the dependence structure in the tails is either unknown or based on very limited data. The workbook discussed in this paper allows distributions to be established to use this known information, but to allow exploration of hypothetical dependence relationships that satisfy the criteria for a valid probability distribution.

In the situation described in the paragraph above, the use of a copula allows the focus of the modeling process to be placed on the unknown elements of dependence, while maintaining the known information on marginal distributions and dependence for common events. The concept of a copula is not limited to two dimensions. The process outlined in this paper could be extended to any number of dependent marginal distributions with a copula of the corresponding dimensionality. The practical implementation of such an approach involves, of course, an exponential increase in the work involved.

The *Excel* workbook that we have prepared can be used with software such as *Simtools*[2] or *@Risk*[3] to simulate events with the known marginal distributions and dependence relationships. The use of a copula allows fully general evaluation of dependence, subject to the known information, while maintaining the necessary characteristics of a probability distribution. To illustrate the procedure we have constructed copulas with a bivariate normal distribution for common events (defined in this case as a cumulative probability between 0.05 and 0.95) and an arbitrary relative density for other (*i.e.*, relatively uncommon) events. The use of the normal copula for this purpose was purely a matter of convenience any copula could be substituted for the normal copula without any change in the rest of the process. The correlation coefficient for the bivariate normal distribution may be specified, and an estimate of Spearman's rho is produced for the combined numerical copula. This numerical copula may then be used for simulation.

3. BASIC STRUCTURE OF THE EXCEL WORKBOOK

The workbook `CopulaExample.xls` consists of eight spreadsheets. The user can focus on the first four of these spreadsheets, as the remaining sheets are only used to provide values for the normal copula (described in Appendix A), which did not seem to be readily available in prepared form. The spreadsheets of interest to the user are shown in the following table.

---

[2]This is an *Excel* add-in developed by Roger Myerson of the University of Chicago. It is available to be downloaded for free at `http://home.uchicago.edu/~rmyerson/addins.htm`

[3]This is a popular commercial *Excel* simulation add-in. More information here: `http://www.palisade.com/risk/`

Worksheet	Purpose
Constructed Copula	Allows user to build copula with known "core"
Marginals & Simulation	Using marginals and constructed copula simulates values
Simulation Output	Results of applying *Simtools* to "Marginals & Simulation"
Copula Examples	Seven examples of known copulas with various properties

The starting point for use of the workbook is the first spreadsheet, `Constructed Copula`. This is set up for the creation of a copula where the dependence structure for common events is based on the normal copula. In other words, if the marginal distributions were normal distributions, this example would result in the bivariate normal distribution for common events. The normal copula is a one-parameter family of copulas where the parameter is the linear correlation coefficient for the bivariate normal distribution. The spreadsheet initially has this value specified as 0.6 in cell `C2` of the spreadsheet, but this value can be specified by the user as any value in the open interval $(-1, 1)$.

The spreadsheet consists of a series of tables with headings that represent the cumulative probabilities of the two marginal distributions. Our convention for presenting density values is to present the value for a cell with the headings representing the upper right corner of the cell. For example, the default value of 0.00393 shown in cell `E20` with an X value of 0.2 and a Y value of 0.95 is actually the total density in the cell $[0.05, 0.2] \times [0.8, 0.95]$.

The starting point for the construction of the copula example consists of the two tables in `C17:J25` and `C34:J40`. In `C17:J25` the density values of the normal copula are entered for common events, which we have arbitrarily defined as events with a cumulative probability between 0.05 and 0.95. These values of the copula are assumed to be based on known dependence relationships obtained from empirical data, and are fixed throughout the rest of the construction.

The values in cells `C34:J40` are intended to be changed by the user to construct a copula with desired properties. Initially these are filled in with numerical values from the normal copula. (Note: If the correlation coefficient is changed, the values in cells `M19:S25` will have to be copied to `M34:S40` using "paste special," "values." These values are then copied to the appropriate cells within `C34:I40`.) These values are entered initially simply to provide a valid starting point for the construction.

A reasonable starting point for using the workbook would be for the user to perform a simulation on the basis of the workbook as provided, before making any changes to the spreadsheets. As we noted earlier, Roger Myerson of the University of Chicago offers an *Excel* simulation add-in, *Simtools* for free at `http://home.uchicago.edu/~rmyerson/addins.htm`. If this add-in tool is used, a simulation can be performed by highlighting cells A1:C501 in the "Simulation Output" spreadsheet, and then running the "simulation table" function of *Simtools*. This will fill in the simulated values, and update the graph included in this spreadsheet.

A function on $[0, 1] \times [0, 1]$ is a copula if and only if it is a cumulative probability distribution with uniform marginal distributions. These conditions are equivalent to the set of formulas in cells `C74:J81` of the "Constructed Copula" worksheet, together with the condition that all of the weights in cells `C74:J81` and `C34:I40` are non-negative (assuming that the non-negative values in `E20:H23` are used). The cell computations are automatic as values in `C34:I40` are changed. The non-negativity is tested and reported in cell `K30`. The user may alter values in `C34:I40` as desired, and as long as `K30` shows "OK," the final copula will adhere to the required conditions.

Cell `L59` of the "Constructed Copula" spreadsheet provides an approximation for Spearman's rho, a measure of dependence frequently used for copulas (as discussed below). A simple second step in using the program would be to modify one of the cells in `C34:I40` and observe the change in Spearman's rho, as well as carry out a simulation with the modified value. With 81% of the values determined by the fixed central region of the copula, the degree of change in Spearman's rho is limited, but still illustrates the process for constructing a copula with desired properties. For example, replace the entries in cells `F34` and `I37` with 0.003 and `G34` and `I36` with 0.01. Cell `K30` indicates that these changes do not violate the non-negativity constraints. The value of Spearman's rho increases from 0.51802 to 0.51873. While this is a very modest increase, when we performed one example of a simulation of 500 observations, the number of cases in which both variables were either below the 5th percentile or above the 95th percentile increased from 9 cases to 13 cases, a significant increase if these cases represented occurrences associated with financial losses.

4. Construction of Modified Copula

In the situations that we envision as applications of these techniques the marginal distributions can be approximated reasonably well, and the dependence between the variables is known to a reasonable approximation for commonly encountered values. We assume that linear correlation is a reasonable approximation to the dependence pattern for commonly encountered values, but that the dependence relationship for extreme values is not well known.

The calculations presented in the spreadsheet are for purposes of illustration of the concepts. As such, the granularity of the computation grid was purposely limited to a coarse grid with few coordinate points, to simplify review of the model. The values of the constructed copula are specified on a rectangular grid of coordinates. The coordinates are chosen so that the grid points include all of the points with x and/or y equal to 0 or 1. Our method involves stipulating the probability associated with each sub-rectangle of the grid. The conditions for a copula are satisfied if these probabilities are non-negative, and the sums of the probabilities in each row or column are equal to the difference between respective cumulative marginal probabilities. The copula will satisfy the conditions $C(x,0) = 0$, $C(0,y) = 0$, $C(x,1) = x$ and $C(1,y) = y$. A real-world model would follow these procedures, but would likely have a much more detailed computation grid.

Our procedure creates a copula that matches the probability densities of the bivariate normal with a correlation coefficient specified by the user. The marginal probabilities for the bivariate normal range between 0.05 and 0.95. These values were chosen arbitrarily to mark the transition from common to rare events, and could be adjusted by a modification of the Excel workbook. The structure outside of this range is created separately, and combined with the normal copula values within the common range. The constructed portion of the copula is controlled to meet the conditions required of a copula. It is necessary and sufficient that the values be equal to those of a cumulative probability distribution on $[0,1] \times [0,1]$ with uniform marginals. The normal copula automatically satisfies these conditions, so it is known that at least one feasible solution exists for the constructed cells. The constructed portion of the copula must be determined so that the final product continues to satisfy these constraints.

5. Measures of Dependence

The linear correlation coefficient of a joint probability distribution depends not only on the dependence structure of the distribution, but also on the specifics of the marginal distributions. For example, consider a joint probability distribution in which X follows a uniform distribution on $[0,1]$, but Y follows the exponential distribution $f(y) = \mathrm{e}^{-y}$ for $0 < y < \infty$. Suppose also that X and Y are completely

order dependent, *i.e.* that if X is given then $Y = -\log(1-X)$. In this case the correlation of X and Y is given by

$$\frac{1}{\sigma_X \sigma_Y} \int_0^1 (x - 1/2)(-\log(1-x) - 1)\, \mathrm{d}x = \frac{\sqrt{3}}{2}.$$

Thus, even though there is a perfect association between X and Y, the correlation coefficient is less than 1. Of course if the marginal distributions were identical, the correlation coefficient in this case would be 1. This shows that the dependence structure does not determine the correlation coefficient, so the copula does not determine a unique correlation coefficient. To measure association in a way that is determined by the copula we turn to order statistics. An example of a measure of the degree of association between variables using order statistics is rank correlation. Such a measure is Spearman's rho, which is given for a copula C by

$$\rho_S = 12 \int_0^1 \int_0^1 (C(x,y) - xy)\, \mathrm{d}x\, \mathrm{d}y.$$

We have approximated this value numerically in the *Excel* workbook by using average values over each cell of the grid. We tested this numerical approximation on the basis of the bivariate normal copula, for which Spearman's rho, ρ_S, is given by

$$\rho_S = \frac{6}{\pi} \arcsin \frac{\rho}{2}$$

where the symbol ρ on the right side of the equation represents the linear correlation coefficient (McNeil, Frey and Embrechts, 2005, page 215). Tests using correlation coefficients from plus or minus 0.1 to plus or minus 0.9 showed that the approximation was accurate to within one percent of the true value on the basis of the grid spacing used in this workbook. The user can observe the changes to Spearman's rho, shown in cell L59 of the "Constructed Copula" spreadsheet as changes are introduced in cells C34:I40.

6. MARGINAL DISTRIBUTIONS

The marginal distributions are specified in the "Marginals & Simulation" spreadsheet of the workbook. The selection of the copula and of the marginal distributions are totally independent of each other – the choice of marginal distributions is not restricted in any way. We have chosen to illustrate the process with X based on a normal distribution and Y based on a lognormal distribution. These are specified in the "Marginals & Simulation" spreadsheet in rows 33 to 48. The upper portion of that spreadsheet demonstrates the simulation of uniform random variables with the dependence structure specified by the bivariate copula.

The process of selecting dependent uniform variables is straightforward, but appears complex because of the workaround needed to look up the boundaries of a particular grid cell using the `=vlookup()` function in *Excel*. The simulation of X and Y starts with the selection of a uniform pseudo-random number u_1. A second uniform pseudo-random number u_2 is then used with the copula to approximate c_2 such that for the bivariate uniform probability distribution associated with C the cumulative probability of c_2 given u_1 is u_2. To perform this calculation we look at the copula as a cumulative probability distribution, and find c_2 such that the cumulative probability of c_2 given u_1 is u_2. The two uniform random variables with dependence based on the copula are then used to compute X and Y with these values as, respectively, the cumulative probability of X and the cumulative probability of Y given X.

The user may modify the marginal distributions as desired without the necessity of making any other changes to the workbook. This may be of interest as a way to observe the separate effects of the copula and the marginal distributions on the simulated results.

7. MAXIMIZATION OR MINIMIZATION OF SPEARMAN'S RHO

The constraints resulting from the construction of a copula on an M by M grid with a fixed N by N sub-grid specified by the densities of the normal copula can be reduced to a system of $2M - 1$ linear equations in $M^2 - N^2$ variables with $M^2 - N^2$ non-negativity constraints on the respective variables. Assume a grid spacing given by

$$x_0 = 0, x_1, x_2, \ldots, x_{M-1}, x_M = 1$$
$$y_0 = 0, y_1, y_2, \ldots, y_{M-1}, y_M = 1$$

Let the density in the grid sector $[x_{i-1}, x_i] \times [y_{j-1}, y_j]$ be denoted by d_{ij}. The following system of equations and constraints is equivalent to the conditions required of a copula:

$$d_{iM} = y_{i+1} - y_i - \sum_{j=1}^{M-1} d_{ij}$$

$$d_{Mj} = x_{j+1} - x_j - \sum_{i=1}^{M-1} d_{ij}$$

$$d_{MM} = 1 - \sum_{i=1}^{M-1}\sum_{j=1}^{M-1} d_{ij} - \sum_{i=1}^{M-1} d_{iM} - \sum_{j=1}^{M-1} d_{Mj}$$

$$d_{ij} \geq 0 \qquad i, j = 1, 2, \ldots, M$$

The existence of the normal copula demonstrates that solutions exist for this system of equations and constraints. The set of values satisfying these constraints is a simplex.

The calculation to estimate Spearman's rho is linear in the copula values. Therefore the problem of maximizing or minimizing Spearman's rho subject to the above constraints is a problem in linear programming. The simplex method of linear programming can be used to solve this problem in a finite number of steps. In this way a dependence structure can be created that maximizes or minimizes the rank correlation, subject to the known core of the copula. We have not implemented the simplex method within the *Excel* workbook, but software for the simplex method is readily available. We have provided an example of the adjusted cells of a copula in which we approximated the result that would be obtained by using the simplex method to maximize rho. The values are included in cells C88:I94 of the "Constructed Copula" spreadsheet. The resulting value of Spearman's rho is 0.54598. The specific simplex problem used to find this solution is presented in Appendix B.

8. CONCLUSION

This workbook is intended as an example of how a constructed copula could be used in simulation when the marginals of the variables are known. Use of this approach in a practical implementation would involve extending the grid to a more refined set of values, but would not involve any conceptual changes. We hope this example will help users to apply the copula methodology in risk management.

REFERENCES

American Academy of Actuaries (Org.): American Academy of Actuaries Report to NAIC, Correlations for Period 1978-1996. December 1997

Frees, Edward W., Carriere, Jacques and Valdez, Emiliano: Annuity Valuation with Dependent Mortality. Journal of Risk and Insurance, 63 1996, Nr. 2, 229–261

Gorski, Larry M., Longley-Cook, Alastair G. and Reiskytl, James F.: Risk-Based Capital Update. In **Cook, Alastair G. longley, editor:** Record of the Society of Actuaries. Volume 28, 2002, 1–28

McNeil, Alexander J., Frey, Rüdiger and Embrechts, Paul: Quantitative Risk Management. Princeton University Press, 2005

Nelsen, Roger B.: An Introduction to Copulas. Springer-Verlag New York. Inc, 1999, Lecture Nnotes in Statistics 139

Venter, Gary G.: Tails of Copulas. Proceedings of the Casualty Actuarial Society, LXXXIX 2002, 68 – 113

Venter, Gary G. et al.: Multivariate Copulas for Financial Modeling. Variance 1 2007, Nr. 1

APPENDIX A. CALCULATION OF BIVARIATE NORMAL DISTRIBUTION WITH *Excel*

To compute the cumulative bivariate normal distribution function we prepared an *Excel* spreadsheet to compute densities for the bivariate standard normal distribution function $F_N(x, y)$ with a variable correlation coefficient ρ with x and y varying from -4 to 4 with a grid spacing equal to the lesser of 0.025 or one fourth the difference between $F_N^{-1}(0.05n)$ and the nearest multiple of 0.1. In other words, our grid spacing was 0.025 unless we needed to shrink the spacing to achieve an exact marginal probability equal to a multiple of 0.05.

We used Simpson's rule to carry out numerical integration of the densities to obtain the cumulative bivariate normal distribution. The standard bivariate normal density is given by

$$
\begin{aligned}
f_N(x, y) &= \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\dfrac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}} \\
&= \Pr(Y = y | X = x)\Pr(X = x) \\
&= f_N(x)f_N\left(\frac{y - \rho x}{\sqrt{1-\rho^2}}\right)
\end{aligned}
$$

We are using the fact that the conditional distribution of $Y|X = x$ is normally distributed with mean ρx and standard deviation $\sqrt{1-\rho^2}$. Integrating over the rectangle $(-\infty, x] \times (-\infty, y]$ gives the cumulative probability as an iterated integral:

$$
\begin{aligned}
F_N(x, y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_N(u, v)\, dv\, du \\
&= \int_{-\infty}^{x} g(u)\, du \\
g(x) &= f_N(x)\int_{-\infty}^{y} f_N\left(\frac{v - \rho x}{\sqrt{1-\rho^2}}\right) dv \\
&= f_N(x)F_N\left(\frac{y - \rho x}{\sqrt{1-\rho^2}}\right)
\end{aligned}
$$

We used this fact to calculate densities, using the *Excel* worksheet function `NORMDIST` to get $F_N\left(\dfrac{y - \rho x}{\sqrt{1-\rho^2}}\right)$, and then integrating $g(u)$ numerically to get $F_N(x, y)$. The

worksheet `Bivariate Normal CDF` has this numerical integration. It needs a starting value for $F_N(4, -4)$, which is determined in the worksheet `Lower Tail`using numerical integration.

Since this `NORMDIST` is highly accurate, the only source of error in the cumulative density would arise from the approximations in the numerical integration.

A check on the accuracy of the numerical integration indicated that the degree of approximation is adequate. We compared the *Excel* worksheet results to the values obtained from *Mathematica* for some key values.

	Excel	*Mathematica*
$\rho = 0.5$		
$F_N(4,4)$	0.999938	0.999937
$F_N(0,0)$	0.333334	0.333333
$F_N(1,0)$	0.468744	0.468743
$F_N(0,1)$	0.468744	0.468743
$F_N(-4,-4)$	1.12669×10^{-6}	4.87055×10^{-7}
$\rho = -0.5$		
$F_N(4,4)$	0.999938	0.999937
$F_N(0,0)$	0.166668	0.166667
$F_N(1,0)$	0.372603	0.372602
$F_N(0,1)$	0.372603	0.372602
$F_N(-4,-4)$	1.126688×10^{-6}	3.46191×10^{-17}

The spreadsheet values of the copula are shown to four decimal places, so the error in numerical integration is insignificant in relation to the values shown.

APPENDIX B. SIMPLEX PROBLEM GENERATED BY CONSTRAINTS WITH NORMAL CORE AND CORRELATION $\rho = 0.6$

Maximize $t = 0.586a + 2.665b + 7.942c + 14.978d + 20.256e + 22.334f + 11.46g$
$$+ 0.571h + 2.597i + 7.739j + 14.595k + 19.737l + 10.881m + 0.518n$$
$$+ 0.383o + 0.203p + 0.068q + 0.015r$$

subject to

$$a + b + c + d + e + f + g \le 0.022750062$$
$$f + h + i + j + k + l + m \le 0.027249938$$
$$a + h + n + o + p + q + r \le 0.027249938$$
$$n + l + e \le 0.018867098$$
$$o + k + d \le 0.015599035$$
$$p + j + c \le 0.015599035$$
$$q + i + b \le 0.018867098$$
$$2a + 2b + 2c + 2d + 2e + 2f + g + 2h + 2i + 2j + 2k + 2l + m$$
$$+ 2n + 2o + 2p + 2q + r \ge 0.123432146$$

and all variables $a, b, c, \ldots, r$ constrained to be nonnegative.

Note: The target function represents the change in Spearman's rho for a unit change in each of the variables, determined empirically by changing the respective variables. The copula was assumed to be symmetric, reducing the number of variables from 33 to 18. Variables were named alphabetically, as shown in the table below:

a	h	n	o	p	q	r
b	i					q
c	j					p
d	k					o
e	l					n
f	m	l	k	j	i	h
g	f	e	d	c	b	a