**CORRELATION: COPULAS AND CONDITIONING**

This note reviews two methods of simulating correlated variates: copula methods and conditional distributions, and the relationships between them. Particular emphasis will be on how to select among the different copulas, especially by using the relationships they produce in the tails of distributions.

1. **Copulas**

A joint distribution function \( F(x,y) \) can be analyzed as \( F(x,y) = C(F_X(x), F_Y(y)) \), where \( F_X(x) \) and \( F_Y(y) \) are the distribution functions for \( X \) and \( Y \). There has to be a function \( C \) defined on the unit square that makes this work, because \( F_X(x) \) and \( F_Y(y) \) are order preserving maps of the real line or some segment of it to the unit interval. To find \( C(u,v) \), just take \( F(F_X^{-1}(u), F_Y^{-1}(v)) \). Then \( C(F_X(x), F_Y(y)) = F(F_X^{-1}(F_X(x)), F_Y^{-1}(F_Y(y))) = F(x,y) \). The function \( C(u,v) \) is called a copula. For many bivariate distributions, the copula form is the easiest way to express and generate the joint probabilities.

Copulas work in the multi-variate context as well, but initially let us focus on joint distributions of two variates. A copula is a joint distribution of two uniform random variates \( C(u,v) = \Pr(U \leq u, V \leq v) \). Let \( c(u,v) \) denote the corresponding probability density. The simplest copula is the uniform density for independent draws, i.e., \( c(u,v) = 1 \). Drawing from this produces two independent \([0,1] \) variates, and they can be inverted by \( F^{-1}(u), G^{-1}(v) \) to produce two independent draws from the variates of interest.

A more interesting example is the normal copula. Two correlated standard normal deviates can be drawn from the joint normal distribution (how later), and their probabilities computed. These probabilities are the uniform variates that are then inverted to simulate the correlated target variates. Letting \( \Phi \) denote the standard normal distribution function, the copula is computed at a pair of points \((u,v)\) by inverting them with \( \Phi \) and then calculating the joint normal probabilities for the inverted pair, i.e., \( C(u,v) = H(\Phi^{-1}(u), \Phi^{-1}(v)) \), where \( H \) denotes the joint normal distribution function. In simulation the order of these steps is reversed.

Here is an example. First, 600 observations are simulated from the joint standard normal distribution with correlation of 0.8. A realization is graphed below.

![Normal Pairs](image-url)
Next these are mapped via the marginal probabilities to the unit square. For any point \((x,y)\) this is done by computing the mapped point \((\Phi(x), \Phi(y))\), illustrated below.

The bunching in the corners is due to the change of scale in going from the normal distribution to the uniform. The entire infinite tails are mapped into the areas near \((0,0)\) and \((1,1)\). These pairs can be used to simulate other distributions by computing \(F_X^{-1}(u), F_Y^{-1}(v)\) at each \(u,v\) pair. For heavy tailed distributions, the corner points will be mapped back out over a large region in the simulation. However, for distributions with support on the positive reals only, high correlation can occur near \((0,0)\).

As an example, the (shifted) Pareto \(F(x) = 1 - (1 + x)^{-1.5}\), with inverse \(x = (1 - u)^{-2/3} - 1\) was simulated. This has mean 2 and infinite variance. The simulated points are shown below.
The correlation is not so evident in the simulated Pareto distribution, and in fact it is not as high. (Correlation coefficients are not preserved by copula simulation.) However, there is still a relationship between the two Paretos, which can be seen on the log scale graph below.

In repeated simulations of this sample of 600, the correlation of the logs of the generated Pareto pairs was consistently around 80%, while the unlogged Pareto pairs had inconsistent correlation, generally about 50% but sometimes below 20%. This illustrates that non-linear transformations that retain order can significantly affect correlation.

It may help in visualizing this process to see the joint densities involved. First, below is the joint normal density, which in this case is plotted with axes on a normal scale.
After transforming to the copula, a lot of density is pushed into the corners, due to the non-linear change of scale. The copula density is graphed below, truncated at the corners.

Finally, inversion produces the joint Pareto density (axes rotated).

High correlation is evident for small losses with reduced correlation for large losses. This is in part due to the heavy tail of the Pareto, which washes out correlation with the high variance. However,
low tail correlation is actually a general feature of the normal copula. Looking back at the original graph of the normal pairs shows little relationship among the observations greater than two standard deviations from the mean. Other copulas provide greater tail correlation. The next example is a copula in which the tail correlation can be controlled directly.

The Partial Perfect Correlation Copula of Kreps

The idea here is to draw two perfectly correlated deviates in some cases and two uncorrelated deviates otherwise. More specifically, let \( h(u,v) \) be a symmetric function of \( u \) and \( v \). To implement the simulation, draw three unit random deviates \( u, v, \) and \( w \). If \( h(u,v) < w \), simulate \( x \) and \( y \) as \( F_X^{-1}(u) \) and \( F_Y^{-1}(v) \) respectively. Otherwise take the same \( x \) but let \( y = F_Y^{-1}(u) \). Thus some draws are independent and some are perfectly correlated. The choice of the \( h \) function provides a lot of control over how often pairs will be correlated and what parts of the distributions are correlated.

For instance, \( h \) can be set to 0 or 1 in some interval like \( j < u,v < k \) to provide independence or perfect correlation in that interval, or it could be set to a constant \( p \) to provide correlation in 100p% of the cases in that interval. Another choice is \( h(u,v) = (uv)^q \). This creates more correlation for larger values of \( u \) and \( v \), with \( q \) controlling how much more. The graphs below illustrate the case where \( h(u,v) = (uv)^{0.3} \) and both \( X \) and \( Y \) are distributed Pareto with \( F(x) = 1 - (1 + x)^{-4} \). The correlated and uncorrelated instances clearly show up separately, in either the log or regular scale.
The partial perfect correlation copula thus provides a good deal of flexibility and control over how much correlation is incorporated and where in the distribution it occurs. The exact form of the simulated distributions may seem unlikely, but for most applications this should not be a problem.

The Gumbel Copula

Another interesting copula is given by \( C(u,v) = \exp \left\{ \left[ (\ln u)^{1/a} + (\ln v)^{1/a}\right]^{a} \right\} \) for \( 0 < a \leq 1 \). This is called the logistic or Gumbel copula. Embrechts, McNeil and Strauman in the XXX ASTIN Colloquium give an example of two gamma variates \( X \) and \( Y \) with correlation = 70% simulated from the normal copula and the Gumbel copula. They compute the probability that \( X \) exceeds the 99th percentile given that \( Y \) does. For the normal copula this turns out to be 33%, vs. 75% for the Gumbel copula. More generally they cite sources that show that the normal copula is asymptotically independent, but the Gumbel copula is not. In effect, the Gumbel copula puts more weight in the corners than does the normal copula, especially in the region near \([1,1]\). The two graphs above illustrate this by showing the two copula densities on a log scale.

To simulate uniform deviates from the Gumbel copula with parameter \( a \), first simulate two independent uniform deviates \( u \) and \( v \). Next solve numerically for \( s > 0 \) with \( u^{s} = 1 + a s \). Then the pair \( [\exp(-sv^{a}), \exp(-s(1-v)^{a})] \) will have the Gumbel copula distribution. (See Embrechts et al.)

Conditioning with Copulas

The conditional distribution can be defined using copulas. Letting \( C_{1}(u,v) \) denote the first partial derivative of \( C(u,v) \). When the joint distribution of \( X \) and \( Y \) is given by \( F(x,y) = C(F_{X}(x),F_{Y}(y)) \), then the conditional distribution of \( Y \mid X=x \) is given by:

\[
F_{Y\mid X}(y) = C_{1}(F_{X}(x),F_{Y}(y))
\]

If \( F_{Y} \) is of closed form and \( C_{1} \) is simple enough to invert algebraically, then the marginal distributions can be maintained with simulation done using the derived conditional distributions. The next copula illustrates this.

Frank’s Copula Applied to Loss Adjustment Expense

Klugman and Parsa in *Fitting bivariate loss distributions with copulas*, Insurance: Mathematics and Economics 24 (1999) p. 139, estimate the bivariate severity distribution for loss and allocated loss adjustment expense for an unidentified population of liability claims using copula methods. They arbitrarily chose to use Frank’s copula, which is:
\[ C(u,v) = -a^{-1} \ln \left[ 1 + \left( e^{au} - 1 \right) \left( e^{av} - 1 \right) / (e^a - 1) \right] \]

This has first partial (wrt \( u \)):

\[ C_1(u,v) = \frac{g_u g_v + g_v}{g_u g_v + 1} \]

where \( g_x = e^{ax} - 1 \).

Suppose \( x \) has been simulated and we want to simulate \( y \) by the conditional distribution \( F_{Y|X}(y) = C_1(F_X(x),F_Y(y)) \). Say we found that \( F_X(x) = u \), so we want to simulate a value of \( y \) from \( C_1(u,F_Y(y)) \). If we can invert \( F_Y \) either in closed form or numerically, all we need to simulate is \( v = F_Y(y) \) from \( C_1(u,v) \). So simulate a \( w \) which is uniform \((0,1)\) and set \( C_1(u,v) = w \). Solving for \( g_v \) gives:

\[ g_v = \frac{wg_u}{g_u - wg_u + 1} \]

\[ v = -a^{-1} \ln (1 + g_v) \]

so

\[ y = F_Y^{-1}(v). \]

For their data set, Klugman and Parsa use maximum likelihood to estimate the parameters of the joint distribution, which consists of a combination of the loss and expense severity distributions and the \( a \) parameter of Frank’s copula. After trying many alternatives, they use an inverse Burr distribution for ALAE and an inverse paralogistic distribution for loss. These have the forms and fitted parameters as follows:

- **ALAE**: \( F_Y(y) = \left[ 1 + \left( y/b \right)^p \right]^{-q} \quad b = 10,100; \quad p = 1.5766; \quad q = 0.57353 \)
- **Loss**: \( F_X(x) = \left[ 1 + \left( x/d \right)^r \right]^{-s} \quad d = 11,578; \quad r = 1.0460 \)

Frank’s \( a \) was estimated as \( a = 3.0744 \).

The inversion of \( F_Y \) is \( y = b(v^{-q} - 1)^{-1/p} \).

This copula, with this \( a \), but in general, does not generate a lot of correlation in the tails. Unlike the normal and Gumbel copulas, it has relatively small values at \([0,0]\) and \([1,1]\). This may or may not be a problem for loss adjustment expense, where you can have big loss expense helping to keep losses small. But it would be inappropriate in cases where strong tail correlation is expected. The copula density is plotted below. This is on a numerical scale, and would look almost flat on a log scale.
Simulated values are shown below.

There is no apparent relationship among loss and ALAE for the large values. The next graph is in logs.

The relationship looks significant but still somewhat weak here as well, which it in fact is.
Another analysis of loss and expense distributions was provided by Frees and Valdez in *Understanding Relationships Using Copulas*, North American Actuarial Journal, vol 2, number 1. They use a similar data source to that used by Klugman and Parsa, but they pay more attention to the choice of copula. They adopt a procedure, developed by Genest and Rivest in *Statistical Inference Procedures for Bivariate Archimedean Copulas*, Journal of the American Statistical Association 88, that is able to distinguish among some copulas. They are able to show in particular that the Gumbel copula works better than Frank’s for their data.

The procedure is to take two statistical measures of association among the data pairs, and test the relationship between the two against what it is mathematically implied to be for each copula.

Suppose there are \( n \) data pairs, numbered 1 to \( n \). The first measure \( T \) is an estimate of a statistic of a copula called the Kendall correlation coefficient. Let:

\[
T = 2\sum_{i<j} \text{sign}[(X_{1i} - X_{1j})(X_{2i} - X_{2j})]/n(n-1)
\]

The second measure starts by defining, for the \( i \)th pair \((X_{1i}, X_{2i})\), \( Z_i = \) number of other pairs \((X_{1j}, X_{2j})\) with \( X_{1i} < X_{1j} \) and \( X_{2i} < X_{2j} \) \((n-1)\). Then define as the second measure the function:

\[
K(z) = \text{proportion of } Z_i \text{'s less than or equal to } z.
\]

For the Gumbel and Frank copula the relationship between \( K \) and \( T \) has been quantified. For Gumbel’s copula this is given by:

\[
K(z) = z(1 - \ln z)/(1 - T).
\]

For Frank’s copula it is a little less simple:

\[
K(z) = z + a^{-1}(1 - e^{az})\ln[(e^{az} - 1)/(e^a - 1)], \quad \text{where } a \text{ is the solution of:}
\]

\[
(1 - T)/4 = a^{-1} + a^{-2}\int_0^a w/(1 - e^w) \, dw.
\]

Frees and Valdez graph the empirical \( K \) function against the functions implied by \( T \) for each copula and find that Gumbel’s gives a very good fit and Frank’s less so.

Whereas Klugman and Parsa just assumed a copula and tested several severity distributions for loss and loss expense, Frees and Valdez do just the opposite: after testing copulas they just assume Pareto distributions for the severities. Thus a definitive study of both the marginals and the copula for loss and loss expense is yet to be done. The parameters from Frees and Valdez in the notation used here are:

ALAE: \( F_Y(y) = 1 - [1 + (y/b)]^{-q} \quad b = 14,219; \quad q = 2.118 \)

Loss: \( F_X(x) = 1 - [1 + (x/d)]^{-r} \quad d = 14,036; \quad r = 1.122 \)

Gumbel \( a = 0.688 \)

The following graphs show simulated values from this joint distribution on a regular and log scale. The large and small tail values appear a little more related than in the Klugman and Parsa fit, although the overall relationship seems otherwise similar.
Gumbel Simulated Loss and ALAE

Gumbel Simulated Loss and ALAE on Log Scale
A Heavy Right Tail Copula

Consider the copula given by $C(u,v) = u + v - 1 + [(1 - u)^{1/a} + (1 - v)^{1/a} - 1]^{-a}$, $a>0$. This has very little correlation in the left tail, but high correlation in the right tail, corresponding to large losses. This is graphed below for $a = 1$.

The conditional distribution given by the derivative $C_1(u,v)$ can be solved in closed form for $v$, so simulation can be done by conditional distributions as in Frank’s copula. The derivative is given by $C_1(u,v) = 1 - [(1 - u)^{1/a} + (1 - v)^{1/a} - 1]^{-a-1}(1 - u)^{-1/a}$. If this is simulated to take value $w$, then for a given $u$, $v$ can be solved for in terms of $u$ and $w$.

Frees and Valdez show how this copula can arise in the production of joint Pareto distributions through a common mixture process. Generalizing this slightly, a simple joint Burr distribution is produced when the $a$ parameter of both is the same as that of the heavy right tail copula. That is, let $F(x) = 1 - (1 + (x/b)^p)^{-a}$ and $G(y) = 1 - (1 + (y/d)^q)^{-a}$. Then the joint distribution is:

$$F(x,y) = 1 - (1 + (x/b)^p)^{-a} - (1 + (y/d)^q)^{-a} + [1 + (x/b)^p + (y/d)^q]^{-a}$$

The conditional distribution of $y | X = x$ is also Burr:

$$F(y | X = x) = 1 - \{1 + [y/(1 + x/b)^{p/q}]^a\}^{-a+1}.$$
Quantifying Correlation

It is not generally possible to predict the correlation coefficient produced by copula methods. However an alternative measure of correlation, called rank correlation, is actually a property of the copula, and is constant for all distributions simulated from the copula. The rank correlation may be defined as \( R(X,Y) = 12E[(F_X(x) - .5)(F_Y(y) - .5)] \). Since only the relationship among the probabilities is needed in this definition, all joint distributions with the same copula will have the same rank correlation.

It turns out that R can be estimated by the usual linear correlation coefficient applied to the ranks of the deviates, which gives the name. The ranks are given by replacing each observation with its numerical order for that variate. The ranks are invariant under any order-preserving transformation, so will be the same for any distributions produced from a given copula.

For the standard multi-variate normal distribution, the linear correlation can be calculated from the rank correlation as \( 2 \sin(\pi R/6) \). Thus for a given R the normal copula with that R can be readily determined. However, as in the example above, a highly non-linear transformation can make the rank correlation only loosely productive of a relationship among the variates. In addition, the lack of correlation in the tail for correlated normal variates limits the utility of this procedure in many applications.

For the partial perfect copula, R can be determined numerically for any h, which will thus determine the R for any simulations from this copula.

It is sometimes asserted that the correlation coefficient determines the joint distribution only in the case of normal distributions. This is somewhat misleading. Two normal marginal distributions simulated by the partial perfect correlation copula will have a different joint distribution than bivariate normals with the same correlation. It is the multivariate normal which is uniquely determined by its correlation matrix and marginal distributions. It should be possible to define other joint distribution families having this property.

More than Two Variates

The Frank, heavy right tail, and Gumbel copulas can be extended to the multivariate case in a limited way; the rank correlations among all the variates will be the same. The normal copula extends more generally. Any multi-dimensional rank correlation matrix can be transformed by \( 2 \sin(\pi R/6) \) to the standard normal correlation matrix. That can then be used to simulate deviates, which can be mapped by the standard normal distribution function to make a multi-dimensional normal copula. This provides a unique degree of flexibility for generating several correlated variates, but is subject to the limitations on normal copulas discussed above. The Kreps copula can be extended to a true multivariate case as well. See Rodney Kreps’ article *A Partially Comonotonic Algorithm for Loss Generation* on pages 165 ff of the Proceedings of the XXXIst International ASTIN Colloquium.

Copula Summary

In the bivariate case there are a number of copulas to choose from. They differ particularly in where the correlation is strongest – in one or both tails or somewhere in the middle. In many property-casualty applications losses might be correlated in the right tails. For instance catastrophes can affect numerous lines of business. The heavy right tail and Gumbel copulas thus may be most appropriate. The test connecting two statistics of the Gumbel copula might be useful in checking if this copula is appropriate for the data at hand. Another way to decide between copulas may be to fit some and test using information criteria such as the Akaike information criterion. With single parameter copulas these would come down to comparing the likelihood functions of the fits.
2. Conditioning

Conditional distributions can be computed from copulas, and when the conditional distribution can be inverted, this can be used to simulate the copula. With the rarity of flexible multi-dimensional copulas it may be necessary in the multivariate case to try to define the conditional distributions directly, without reference to a copula.

What allows the normal copula to extend readily to multi-dimensions, and how can this be applied to simulating other distributions? The key is looking at the conditional distributions rather than the marginal. The multi-variate normal is unique in that not just the marginal distributions, but all the conditional distributions for one variate given any collection of the others, are all normal. Furthermore, all the conditional distributions are determined from the correlation matrix. This allows the sequential simulation of the joint distribution, as follows.

To simulate an instance of the multi-variate normal, first the regression coefficients are calculated that specify the conditional distributions. The second variate is expressed as a regression on the first, with a given residual variance. The third is a regression on the first two, with its residual variance, etc. Then the simulation proceeds by taking independent random draws for all the residual (i.e., conditional) distributions. Then each variate is simulated in turn from the regression on the previous variates plus its own residual. Mathematically this is all worked out in vector notation as a matrix product of a vector of standard normal draws with a transform of the correlation matrix, called the Choleski decomposition, which contains all the regression coefficients.

Formally, the Choleski matrix B is lower triangular, i.e., \( b_{ij} = 0 \) for \( i < j \), and elsewhere is defined recursively by:

\[
b_{ij} = \frac{r_{ij} - \sum_{s<j} b_{is} b_{js}}{\left[1 - \sum_{s<j} b_{js}^2\right]^{1/2}}
\]

Then given a column vector \( Y \) of independent standard normals, \( BY \) is a vector of standard normals with correlation matrix \( \{r_{ij}\} \). This formal procedure, though, has at its core the sequential simulation of conditional distributions.

How can this be generalized? Unfortunately there are no other distributions that work as well as the normal. But in practice it might be acceptable in many cases to give up the exact form of the marginal distributions and just work with conditional distributions to simulate correlated variates.

Suppose for example that an insurer wants to simulate winter storm losses for its homeowners, personal auto, commercial auto, small commercial property, large commercial property, and inland marine lines. Say homeowners is the largest exposure, and a Pareto distribution in (1000, 2) has been determined to be a reasonable severity distribution. A study could then be done on the distribution of personal auto losses given the homeowners loss for a storm. Say the homeowners loss is \( x \). It may be reasonable to say the personal auto loss \( y \) is lognormal with mean \( x/2 \) and standard deviation \( x/10 \). Then the commercial auto loss \( z \) may be gamma with mean \( x/20 + y/3 \) and standard deviation 500, etc. The various correlated losses can all be simulated from the conditional distributions. The study would look at both the conditional and resulting marginal distributions to see if the estimates are consistent with the data.

This method would give a good deal of control over the relationships among the variates simulated. For instance, if correlation were found to be greater for larger losses, the conditional distribution could be specified to have less variability for larger losses, etc.
Conclusions

Correlated variates can be simulated using copulas and marginal distributions or directly using conditional distributions. The choice of copula determines not the degree of co-dependency, but rather where in the distribution the co-dependency occurs. The normal copula provides weak correlation in the tails, but Frank’s copula has even less. The Gumbel copula has strong right tail correlation. Tail relationships can be controlled explicitly with the partial perfect copula, but this does not give many near correlated cases. The parameters of each copula will determine the rank correlation of the copula and any distributions simulated from it, but this in itself is not strongly indicative of the relationship between two variates.

In the multi-dimensional case, the only copulas that gives the complete flexibility of being able to specify different relationships among more than two variates are the normal and Kreps copulas. The problem again with the normal is that it has weak tail correlation. The Kreps procedure gives a lot of control over both the degree of correlation and where in the distribution it is strongest, but not all data will be fit well with a combination of perfectly correlated and uncorrelated cases. The best alternative in some cases may be estimating and simulating the conditional distributions directly.