Tech Note #16

Updated: 10 Dec 2008

Multivariate Normal (MultiNormal) Distributions

Overview

A model may have pairs or sets of distributions that should be positively or negatively correlated when sampled during Monte Carlo simulation. The new MultiNormal distribution type in TreeAge Pro 2009 allows modelers to represent a multivariate normal distribution, based on a set of correlations or variances/covariances.

Table of Contents

Distribution Structure

A single MultiNormal distribution will sample multiple normally-distributed values. References to the MultiNormal distribution's individual, correlated sample values are included in variable definitions and other tree expression in the same way as references to the multiple sample values generated by a Dirichlet distribution.

For example, assume that a MultiNormal distribution with the index #1 is created based on a table with four value columns, corresponding to four correlated Normal variates. The four variates can be referenced using the *Dist(index; variate)* function syntax:

```
Var1 = Dist(1; 1)  # A function argument, such as variate, can be a variable.
Var2 = Dist(1; 2)
Var3 = Dist(1; 3)
Var4 = Dist(1; 4)
```

The Cholesky decomposition table

A single table is used to store the inputs for a MultiNormal distribution. The table must contain the Cholesky decomposition of the correlation matrix (in the diagonal and the lower triangular matrix). An example Cholesky decomposition for 4 correlated standard normal distributions is shown below:

8	Index	Value 1	Value 2	Value 3	Value 4
9	1	1	0	0	0
10	2	0.19	0.981784	0	0
11	3	0.05	-0.0606	0.996909	0
12	4	-0.83	-0.51162	-0.19009	0.114934

If four correlated Normal variates are required, for example, the table must include four value columns. The table would also have four rows. (You can optionally include two additional rows, containing the mean and standard deviation for each variate, as a shortcut for converting from the standard normal.)

Setting Up the Correlations

To aid in setting up the Cholesky table, TreeAge includes a new function for performing Cholesky decomposition on a similarly structured table containing pairwise correlations. To demonstrate this, a "source" table representing correlations among the distributions is shown below, in an Excel worksheet:

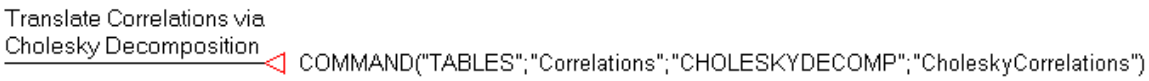
	A	B	C	D	E	F
1	Table Nam	Correlations				
2	File Name	Correlations		Note: table defined in package		
3	Comment					
4	Missing R	interpolate				
5	Off Edge is	no				
6	Default Co	1				
7						
8	Index	V1	v2	v3	v4	
9	1	1	0.19	0.05	-0.83	
10	2	0.19	1	-0.05	-0.66	
11	3	0.05	-0.05	1	-0.2	
12	4	-0.83	-0.66	-0.2	1	

The correlation table for *N* correlated Normal variates is an N x N symmetric, positive definite matrix, with a "unit" diagonal. Correlations are entered repetitively for each pair of variates. For example, the correlation between variates 1 and 2 is entered into cells B10 and C9, the correlation between variates 2 and 3 is entered into cells C11 and D10, and so on.

Cholesky Decomposition

This table should not be used directly by the MultiNormal distribution, however. The matrix must be transformed via Cholesky decomposition. TreeAge Pro provides a command function to convert the correlation table into the required Cholesky decomposition table.

The Cholesky decomposition function is illustrated in the following trivial model:



Calculating the tree (e.g., via rollback) performs Cholesky decomposition on the table named “Correlations” and generates the require table “CholeskyCorrelations”. Any table names can be used.

The resulting CholeskyCorrelations table is presented below.

	A	B	C	D	E	F
1	Table Nam	CholeskyCorrelations				
2	File Name	CholeskyCorrelations	Note: table defined in package			
3	Comment					
4	Missing Rc	interpolate				
5	Off Edge is	no				
6	Default Co	1				
7						
8	Index	Value 1	Value 2	Value 3	Value 4	
9	1	1	0	0	0	
10	2	0.19	0.981784	0	0	
11	3	0.05	-0.0606	0.996909	0	
12	4	-0.83	-0.51162	-0.19009	0.114934	

The converted table is then ready for use by the distribution.

Alternatively, you can create this table manually if you already have the Cholesky decomposition result matrix; there is no requirement to use the Command() function shown above.

Note on Means and Standard Deviations:

The example above assumes standard Normal distributions, with mean 0 and standard deviation 1.0.

Note that you can have TreeAge Pro apply means and standard deviations for the Normal variates after they are correlated. To do this, include two extra rows in the table(s): a row of mean values, followed by a row of standard deviations.

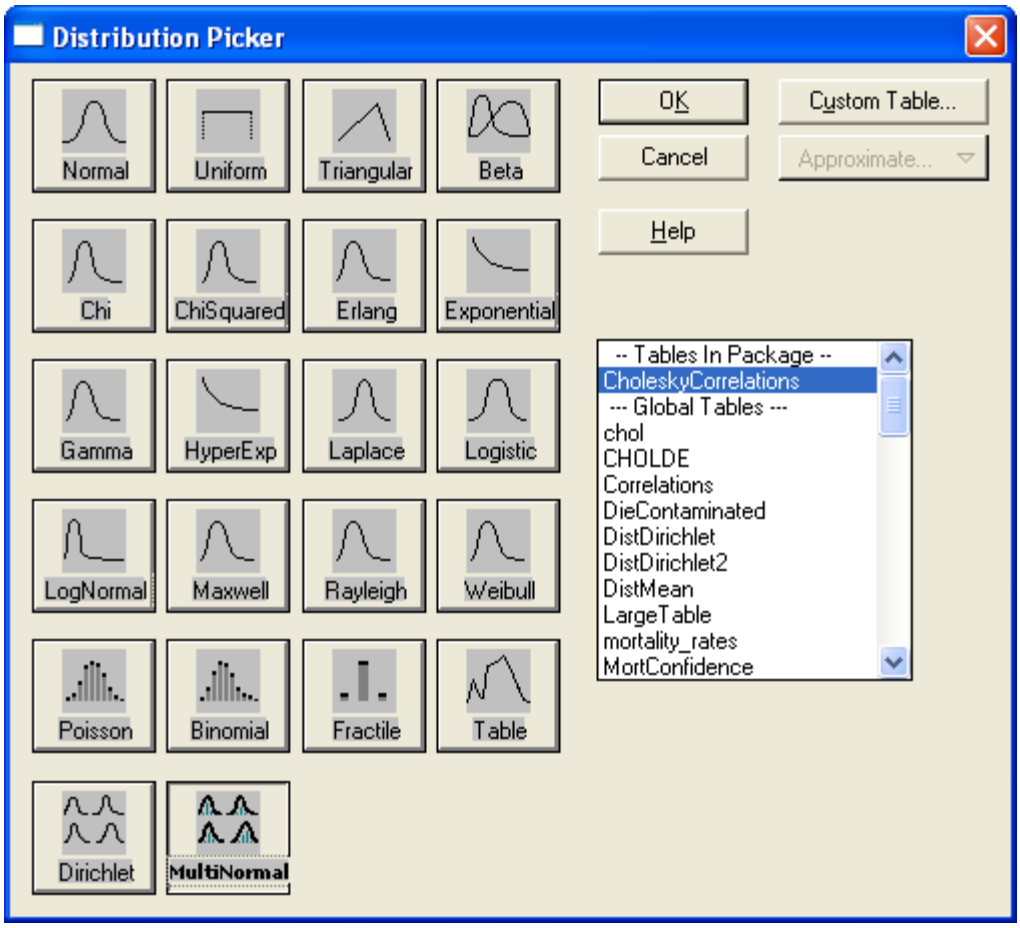
For example, the following table includes means and standard deviations for each correlated distribution in cells B13:E14.

	A	B	C	D	E	F
1	Table Nam	CholeskyCorrelations				
2	File Name	CholeskyCorrelations		Note: table defined in package		
3	Comment					
4	Missing R	interpolate				
5	Off Edge is	no				
6	Default Co	1				
7						
8	Index	Value 1	Value 2	Value 3	Value 4	
9	1	1	0	0	0	
10	2	0.19	0.981784	0	0	
11	3	0.05	-0.0606	0.996909	0	
12	4	-0.83	-0.51162	-0.19009	0.114934	
13	5	10	20	-100	-200	
14	6	2	6	-20	-50	

Creating the Distribution:

As stated earlier, a single MultiNorm distribution within TreeAge Pro generates samples from a number of correlated distributions. This section shows you how to create the MultiNorm distribution.

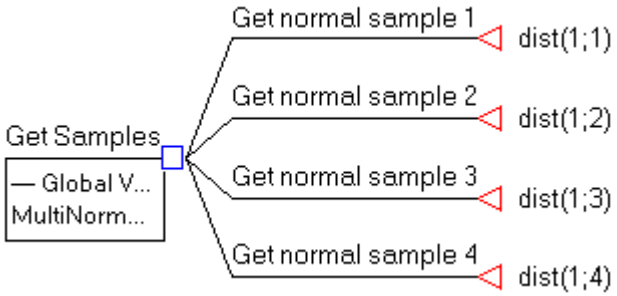
The distribution type to select is MultiNorm (note that the icon is not yet correctly showing "Normal" shaped curves):



This will then require you to select the appropriate table containing the Cholesky decomposition of the correlation matrix, as described above.

Running Simulations

Next, we will examine the simulation output to review the data from each sample. A trivial, 4-strategy model is used to demonstrate; during simulation, it simply reports the sample values from the MultiNormal distribution's four correlated variates. (Note that multivariate distributions in TreeAge only directly report the first variate's value in simulation output.)



If a simulation with 100 samples is performed at the decision node, we get the data like the following in the text report:

Sample	Get normal sample 1	Get normal sample 2	Get normal sample 3	Get normal sample 4	Dist(1)
1	-0.758441074	0.182989099	0.153014234	0.473990495	-0.758441074
2	0.069419493	-0.166488969	2.253342845	-0.56474663	0.069419493
3	-0.262773923	-0.171218301	0.2438458	0.201054565	-0.262773923
4	-0.092830201	0.572936362	0.991309693	-0.361842749	-0.092830201
5	1.390316255	0.058167215	-0.271137033	-1.085833272	1.390316255
6	0.378085001	0.390971999	1.352469599	-0.882331492	0.378085001
7	1.649672523	1.030234848	-0.432868496	-1.68063525	1.649672523
8	-2.291769233	-0.618967051	-0.685447084	2.140941866	-2.291769233
9	-0.167831822	-1.895345614	0.307016717	1.123673909	-0.167831822
10	1.969202569	0.561529445	-0.197359116	-1.82798638	1.969202569
11	1.323976034	0.128638171	1.085808428	-1.457131313	1.323976034
12	-0.478953824	-0.64536005	-0.247505748	0.570028556	-0.478953824
13	1.64701084	3.161676216	-0.219624305	-2.878623027	1.64701084

If we examine the data in Excel, we can calculate the sample correlations using Excel's CORREL function.

J3 =CORREL($B2:$B1001,C2:C1001)

	A	B	C	D	E	F	G	H	I	J	K	L
1	Sample	Get norma	Get norma	Get norma	Get norma	Dist(1)		Correlations from simulation results				
2	1	0.602126	0.929537	0.258905	-1.08228	0.602126		Index	V1	v2	v3	v4
3	2	-0.33958	0.766703	-1.0368	0.18892	-0.33958		1		0.201992	0.031842	-0.84309
4	3	-0.43549	-2.37602	-0.87508	1.596725	-0.43549		2			-0.07161	-0.65306
5	4	0.115133	1.094353	-0.05875	-0.75491	0.115133		3				-0.17015
6	5	-1.09512	0.965381	-1.88595	0.825444	-1.09512		4				
7	6	0.679036	1.204891	-0.74856	-0.82663	0.679036						
8	7	2.975155	-1.56893	0.818336	-1.41579	2.975155		Correlations from unconverted matrix				
9	8	-0.67135	-1.68029	0.274548	1.260836	-0.67135		Index	V1	v2	v3	v4
10	9	-0.13034	0.79696	0.312497	-0.60744	-0.13034		1	1	0.19	0.05	-0.83
11	10	0.912418	0.523506	-0.30743	-0.94111	0.912418		2	0.19	1	-0.05	-0.66
12	11	0.008194	0.605038	1.18104	-0.68825	0.008194		3	0.05	-0.05	1	-0.2
13	12	-0.66672	-0.06746	0.903058	0.163714	-0.66672		4	-0.83	-0.66	-0.2	1

After only 1,000 samples, the inter-variable correlations calculated by the CORREL function (cells J3:L6) approach the pairwise correlations in the first table (copied into cells H10:L13). This correlation would hold even if the distributions had different means and standard deviations as defined in the correlation tables.

Software and Sample Files

The documents referenced in this technical note can be downloaded from the TreeAge Software website.

As this tech note is released, the functionality for MultiNormal distributions is included in beta software only. You can access the beta software at:
http://server.treeage.com/treeagepro/support/
The functionality will be included with the regular release of TreeAge Pro 2009.

Please note that the following example models will not work without the beta version of TreeAge Pro 2008 or the released version of TreeAge Pro 2009.

Excel Workbook with Correlation Matrices and Text Report Output:
http://www.treeage.com/files/pdfs/pro2009/CorrelationTables.xls

TreeAge Pro model to convert the correlation matrix via Cholesky decomposition:
http://www.treeage.com/files/pdfs/pro2009/ConvertMatrix.pkg

TreeAge Pro model with correlated distributions:
http://www.treeage.com/files/pdfs/pro2009/MultiNormalDistribution.pkg