Correlation

Typically dependency between nodes in a DynRisk model is constructed by letting the output value of a node be calculated as a combination of its global input and its local value. If you e.g., choose "Single" or "Double" as the algorithm for a node, the dependency is represented as a simple functional relation between the local value of the node, its global input values, and its output value. As a side effect, the probability distribution of the output values will typically not be the same as the distribution of the local values.

Sometimes, however, you may wish to make sure that the distribution of the output values of the node is as close as possible to the distribution you specified for the local values of the node. At the same time you want the node to be dependent on other nodes. In such cases the "Correlation" operator is useful.

When using the correlation operator, you also have to specify a correlation coefficient. Simply stated the correlation coefficient is a number between -1 and +1 describing the degree of linear dependence between two random variables.

If the correlation coefficient is positive, then large values of the first random variable typically occur in situations when the second variable is large as well, and vice versa. If the correlation coefficient is negative, then large values of the first random variable typically occur in situations when the second variable is small, and vice versa. A coefficient equal to either -1 or +1 means that one of the random variables is simply a linear function of the other. If on the other hand the coefficient is 0, there is no linear dependence at all. [The two random variables may still be dependent, but the dependence has a more complex non-linear structure.]

For each node you can only specify a single correlation coefficient. If, say Node 1 only has one input node, the correlation coefficient describes the dependence between the input node and Node 1. If Node 1 has several input nodes, the correlation coefficient describes the dependence between a certain linear combination of the input nodes and Node 1. This linear combination is determined automatically by DynRisk such that all the input nodes in a sense contribute equally to the uncertainty. More specifically, the input value, which we denote by G here, is calculated according to the following rule:

$$G = a_0 + a_1 X_1 + a_2 X_2 + \ldots$$

where $X_1, X_2,\ldots$ denotes the output values of the input nodes, and $a_0, a_1, a_2,$

… are constants determined by DynRisk such that:

E(G) = 0 and SD(G) = 1

and:

SD($a_1 X_1$) = SD($a_2 X_2$) = …

where E denotes expected value, and SD denotes standard deviation.

Note that to get correct results all the input nodes must be independent of each other. Otherwise DynRisk has no way of determining the right constants, a_0, a_1, a_2, …

[You ensure that the input nodes of a correlation node are independent by avoiding undirected cycles among the correlation nodes and their respective predecessors.]

Unfortunately, for mathematical reasons it is not always possible to preserve all aspects of the local distribution when correlation is introduced. However, the way this is handled by DynRisk, ensures that the distribution of the output values at least has the right expected value and standard deviation. Specifically, the output value of a node with "Correlation" as algorithm is computed according to the following rule:

Output value = a G + b L + c

where G denotes the input value calculated as above, L is the local value, and a, b and c are constants determined by DynRisk such that the output value gets the correct expected value and standard deviation.

To make all this work, there are certain restrictions that apply:

1. You must include stochastic values in a correlation node.

2. The local factor of a correlation node must be non-zero.

3. You must include stochastic values in all input nodes to a correlation node.

4. The local factors of all input nodes to a correlation node must be non-zero.

5. All input nodes to a correlation node must be either correlation nodes or independent nodes.

6. You cannot use edge functions in edges pointing into a correlation node.

If you violate any of these rules, DynRisk will warn you by displaying an error message.