Uncertainty Analysis with UNICORN

D.A.Ababei*   D.Kurowicka   R.M.Cooke

D.A.Ababei@ewi.tudelft.nl   D.Kurowicka@ewi.tudelft.nl   R.M.Cooke@ewi.tudelft.nl

Delft Institute for Applied Mathematics
Delft University of Technology
The Netherlands

1   Introduction

UNICORN is a standalone uncertainty analysis software package. The name of the package stands for "UNcertainty analysis wIth CORrelatioNs" and its main focus is dependence modelling for high dimensional distributions. Random variables can be coupled using a number of dependence structures (such as dependence trees, vines, Bayesian belief nets). An extended formula parser is available for defining the model output, as are a number of post-processing options such as report generation, various graphical interpretations, conditioning, sensitivity analysis and probabilistic inversion.

We intend to explore some of the more important facets of the software by means of a simple example. The theoretical background as well as more detailed examples and Unicorn exercises can be found in Kurowicka and Cooke (2006).

2   Example: Investment

We invest \$1000 for five years. The yearly interest (Vi; i=1,...,5) is uniform on the interval $[0.05, 0.15]$. Successive yearly interests have a rank correlation of 0.70. We will try to model this scenario in Unicorn. The random variables necessary as input for the model are created (Section 2.1) and the model output is defined, as a function of the input variables (Section 2.2). We try to find a dependence structure across the random variables that best suits the problem (Section 2.3). Once the model is thus defined, it is sampled (Section 2.4), and we explore a few post-processing techniques available for analysing the results (Section 2.5). Bayesian belief net functionality is described in Section 2.6.

2.1   Random Variables

The creation of input random variables is the first step in a Unicorn project, and this takes place in the *Random Variable View*. Each variable can be assigned any of the

---

*Correspondence to: D.A.Ababei, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands, telephone: +31 1527 87262.

following distributions (each with its own set of parameters): constant, uniform, log-uniform, normal, log-normal, exponential, Gamma, Weibull, Beta, triangular or discrete. Alternatively a random variable's distribution can be imported from a distribution file created elsewhere.

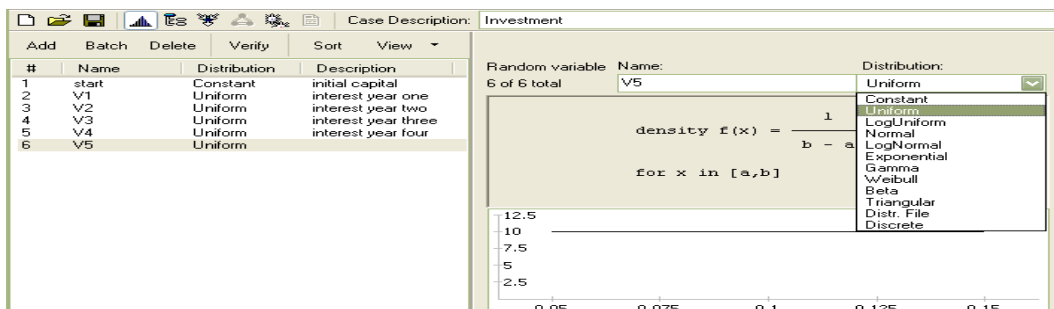We create six random variables, one constant (1000) and five uniforms on $[0.05, 0.15]$.

Figure 2.1: Random Variable View

2.2 Formulas (User Defined Functions)

We proceed to add a formula, in the *Formula/User Defined Function View*, a function of the input random variables that defines the model output. We create a formula named $5yrReturn$ with the expression $start \cdot (1 + V1) \cdot (1 + V2) \cdot (1 + V3) \cdot (1 + V4) \cdot (1 + V5)$.
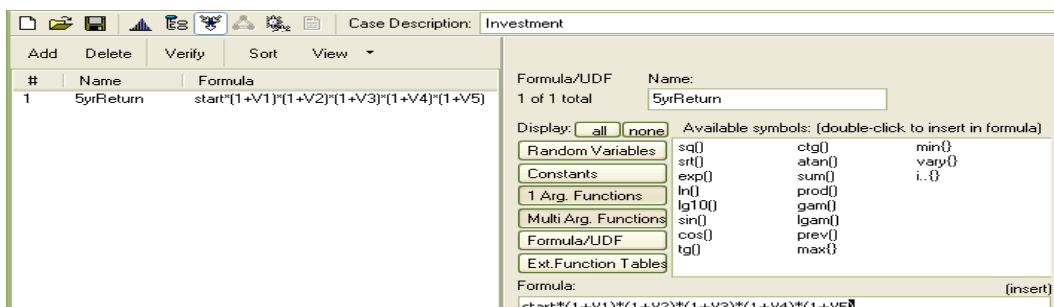

Figure 2.2: Formula/User Defined Function View

If the random variables Vi were independent, this would suffice for our model. However, input variables very rarely are. Our exercise is no exception: its random variables are dependent, and we must incorporate the dependence information in our model.

2.3 Dependence Structures

We know the marginal distribution of the interest variables, and that successive variables are correlated with a rank correlation of 0.7. This does not fully specify the joint distribution, and we explore a few different dependence structures (the *Dependence View*).

First we construct a symmetric structure, where all interests are correlated to eachother at 0.70: a tree where all yearly interests are correlated to a latent variable, with a rank correlation equal to the square root of the desired correlation (Figure 2.3). We first
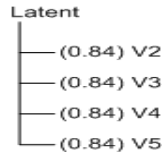
2

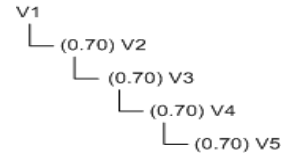Figure 2.3: Tree for Investment with latent variable

Figure 2.4: Tree for Investment

add the latent random variable to the list of random variables (we can go back to the *Random Variables View* and specify new variables at any time).

It does however seem unlikely that the strength of the dependence between all yearly interests would be the same. We construct the tree in Figure 2.4, which is a model where the correlation between two random variables decreases with the distance between them in the tree (e.g. the correlation between $V1$ and $V5$, equal to 0.24, is a lot smaller than that between $V1$ and $V2$) while fulfilling our condition of a 0.70 rank correlation between successive yearly interests.

Further we may reflect that, if the interest is high in one year, it is unlikely to be high in both the year before and the year after it. This relation between two yearly interests, conditional on a third, cannot be described in a probability tree, but it can be captured with a D-vine with a negative conditional rank correlation between Vi and $Vi+2$, given $Vi+1$, $i = 1, 2, 3$ (we take -0.7 as this correlation).
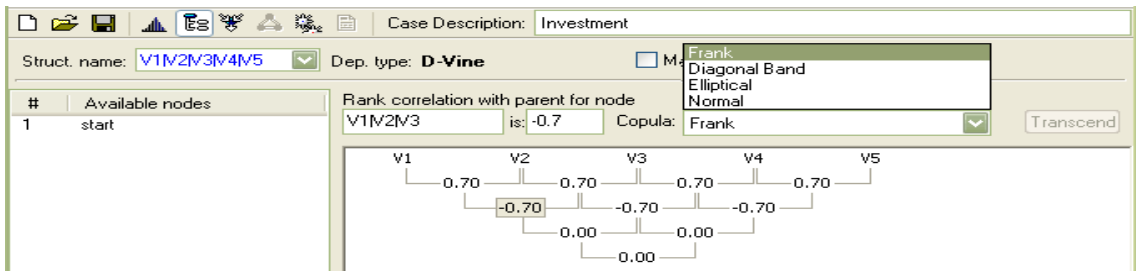

Figure 2.5: D-vine for Investment, in the *Dependence View*

A glance at the distribution of the output variable, $5yrReturn$, reveals that it is different for each of the three dependence structures described above. The means and standard deviations only give us an idea of the differences (Table 1).

Dependence Structure	*5yrReturn* mean	*5yrReturn* standard deviation
Tree with latent variable	1630	185.7
Tree with 0.7 correlations	1620	167.6
D-vine	1610	130.8

Table 1: *5yrReturn* distribution details for different dependence structures

It is also possible to create a hybrid structure, in which one or more of a tree's nodes are vines themselves (Kurowicka and Cooke, 2006, Ch.6). Other dependence structures can be specified. Canonical vines (C-vines) are also available, as are Bayesian belief nets (which we will treat separately at the end of this paper). Whatever the structure, though, the choice of copula for realising the specified rank correlations is important, and in the

3

case of trees and vines we can choose between the diagonal band, Frank's and elliptical copulae. Two other copulae are additionally available, namely the minimum information copula for trees and the normal copula for vines.
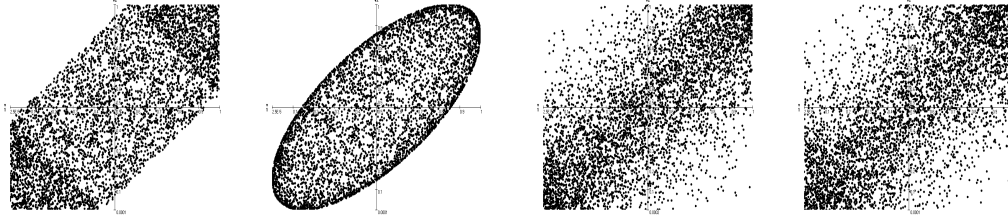

Figure 2.6: The diagonal band, elliptical, Frank's and minimum information copulae with correlation 0.7 (scatter plot rendered by Unicorn's graphical interpretation tool)

2.4 Sampling

In order to evaluate our model, we will sample the joint distribution described by the random variables' marginal distributions and the dependence structure that we have chosen. We have control over the number of samples to be generated, as well as the format(s) of the output sample file(s).

2.5 Post Processing

After the joint distribution has been sampled, we have a few options open to us for analysing the result.

Reporting summarises information about the input and output data. Various properties of the output distributions are described (percentiles, moments, etc) as is the correlation matrix (rank or product moment) of the entire joint distribution.

The *graphical interpretation* tool allows a visual exploration of the joint distribution, with scatter plots (Figure 2.6), density plots, histograms and cobweb plots available. In a cobweb plot, variables are represented as vertical lines (the percentile, natural or logarithmic scale can be chosen for the variables). Each sample realizes one value of each variable. Connecting these values by line segments, one sample is represented as a jagged line intersecting all the vertical lines (Figure 2.7).
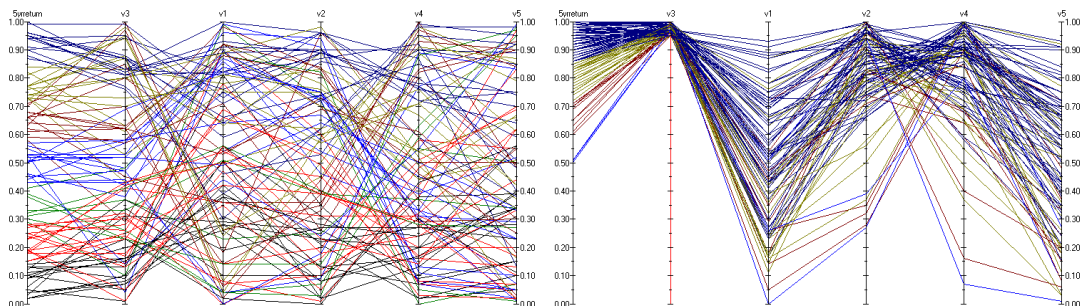

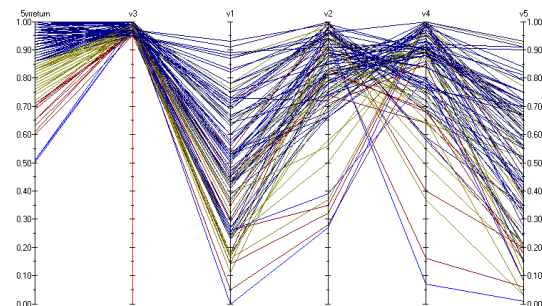Figure 2.7: Cobweb plot for the D-vine for Investment

Figure 2.8: Cobweb plot for the D-vine for Investment, conditionalised on the upper 5% of $V1$'s distribution

4

An interesting feature of the cobweb plot representation is that it is interactive, in the sense that it allows us to *conditionalise* on certain ranges. Suppose that the interest for the third year, $V3$, is in the upper 5% of its distribution. Figure 2.8 shows the result of this conditionalising in the D-vine. Note that the very high values of $V3$ imply high values for $5yrReturn$. $V3$ was chosen for a reason: after examining the five cobweb plots conditionalised on the upper 5% of the distribution of each of the interest variables, we have noticed that $V3$ yields the highest values for $5yrReturn$. In this sense, the graphical interpretation tool can also be seen as providing local sensitivity analysis.

For global sensitivity measures, the *sensitivity analysis* module is available. It assesses the importance of the input variables for the output variable(s), by calculating various statistics and sensitivity measures (such as partial correlations, correlation ratios, etc.) based on the joint distribution sample. For our exercise, the highest correlation ratio (at 0.775) is shown to be between $V3$ and $5yrReturn$, which confirms what we have noticed in the conditional cobweb plots.

We also have a *probabilistic inversion* tool at our disposal, for determining input variable distributions given specific constraints on the output distribution(s).[1]

2.6 Bayesian Belief Nets

One of the dependence structures that we can specify across the random variables in Unicorn is the Bayesian belief net (BBN). This case is treated separately because, currently, choosing this particular dependence structure allows access to functionality which is otherwise unavailable, such as *analytical conditioning*. The functionality associated with the other dependence structures will also be extended over time (the Unicorn software package is being continuously developed).

In Unicorn, the BBN nodes are associated to either continuous or discrete random variables, and the BBN arcs correspond to rank correlations and conditional rank correlations.

In the investment example we wish to include a new influence in the model: at the end of the third year the government plans to reform the tax system, which would influence interest rates. This can easily be captured, together with the previous constraints, in a BBN (Figure 2.9).

The part of the BBN containing the interest rates Vi is equivalent to the D-vine in Section 2.3. A new random variable has been added: Tax, normally distributed, with mean 20 and standard deviation 2[2]. Of course we have to determine the values for the influences of the tax on the interest for the fourth and fifth year (which are represented in the BBN by two conditional rank correlation coefficients).

After the BBN structure has been decided, evidence can be propagated through the graph, via analytical conditioning. In Unicorn, one or more of the variables can be set to a point value within their range, and the BBN can then be updated. When working with BBNs the marginal distributions of the nodes can be visualised directly in the graph, a feature which is particularly useful when conditioning the structure. Suppose we now

---

[1]More information about probabilistic inversion, some applications and examples can be found in Chapter 9 of Kurowicka and Cooke (2006).

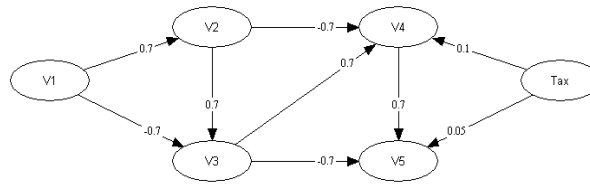[2]This distribution is arbitrarily chosen, any other distribution could have been considered.

Figure 2.9: Bayesian belief net for Investment

learn that the interests for the first and second year, $V1$ and $V2$, have a value of 0.14 and 0.1 respectively. We can see how this affects the other interest rates (Figure 2.10).
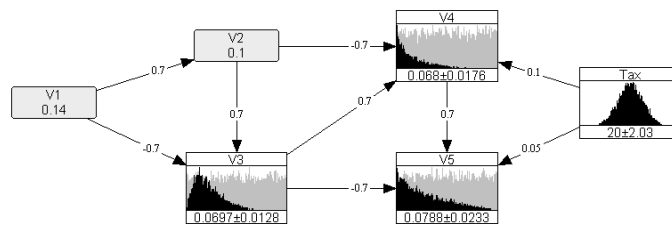

Figure 2.10: Conditionalised BBN for Investment, with $V1 = 0.14$ and $V2 = 0.1$

The grey distributions in the background are the unconditional marginal distributions, provided for comparison. The conditional means and standard deviations are displayed under the histograms.

The joint distribution described by a BBN (both unconditional and conditional) can be sampled exactly as in the case of the dependence structures discussed earlier.

The distribution of $5yr Return$ can also be analysed in both cases. We find out that its mean and standard deviation are 1610 and 130.4 for the unconditional BBN, and 1550 and 68.7 in the case of the conditional BBN. The decrease in mean is consistent with the conditional marginal distributions seen for $V3$, $V4$ and $V5$ (Figure 2.10) while the decrease in standard deviation was to be expected as evidence was introduced, and hence there is less uncertainty in the model.

3 Conclusion

The UNICORN software package is a flexible tool for dependence modelling for high dimensional distributions, incorporating a number of dependence structures which make it suitable for modelling an ample range of scenarios, as well as various post processing tools for analysing the model output. Its continuing development within Delft's Institute of Applied Mathematics ensures it stays abreast with scientific research.

References

Kurowicka, D. and Cooke, R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling.* Wiley.