Continuous/Discrete Non Parametric Bayesian Belief Nets with UNICORN and UNINET

R.M. Cooke[1], D. Kurowicka,
A. M. Hanea, O. Morales,
D. A. Ababei
Department of Mathematics
Delft University of Technology
Mekelweg 4, 2628 CD Delft
The Netherlands
[1]r.m.cooke@ewi.tudelft.nl

B. Ale
Department of Technical
Administrative Science
Delft University of Technology
The Netherlands

A. Roelen
National Aeronautics
Laboratory
Amsterdam
The Netherlands

Abstract

Hanea et al. (2006) presented a method for quantifying and computing continuous/discrete non parametric Bayesian Belief Nets (BBN). Influences are represented as conditional rank correlations, and the joint normal copula enables rapid sampling and conditionalization. Further mathematical background is in Kurowicka and Cooke (2007). This article sketches the current stage of development. The driving application currently involves 133 continuous and discrete probabilistic nodes, and 330 functional nodes. Boolean functions enable fault trees to be fully represented as functional nodes in a BBN. Repeated nodes are easily handled with the identity function. Current perspectives and challenges conclude the paper.

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 modeling for high dimensional distributions. Random variables can be coupled using a number of dependence structures. These can be either:

- Empirical multivariate distribution,
- Markov Trees,
- Vines or,
- Bayesian Belief Nets (BBNs).

Markov Trees and Vines can be used with a user specified copula. BBNs can also be used with arbitrary copulae (see Hanea et al. 2007), however the current implementation works only with the normal copula.

UNINET is a continuous and discrete non parametric Bayesian belief net system, functioning as module of UNICORN. It was developed to support a Causal Model of Air Transport Safety (CATS) under contract with the Dutch Ministry of Transport and Waterworks (Ale et al 2005, 2006, 2007). UNICORN is available free from http://dutiosc.twi.tudelft.nl/~risk, together with supporting scientific documentation. This article briefly describes the driving application, and outlines the design philosophy and main program features.

2. Driving Application

The application driving the development of continuous non parametric BBNs is a causal model for air transport safety. The model so far covers the flight phases take-off, en route and landing. It involves probabilistic nodes whose marginal distributions are, in most cases retrieved from field data. In a few cases structured expert judgment is applied (Cooke, 1991). The influences between probabilistic nodes are quantified by expert judgment (Morales et al appearing) In addition to probabilistic nodes, there are functional nodes that capture fault tree modeling via Boolean functions. The model is still under development, but a substantial portion has been completed and quantified. This portion involves 133 discrete and continuous probabilistic nodes and 330 functional nodes. The model is pictured below; neither the graphic resolution, nor the purposes of this article permits a detailed picture of the individual nodes. The probabilistic nodes are measurable variables which influence human error probabilities. Expected values of these probabilities are fed into the fault trees, which are shown at the upper layers of Figure 1:
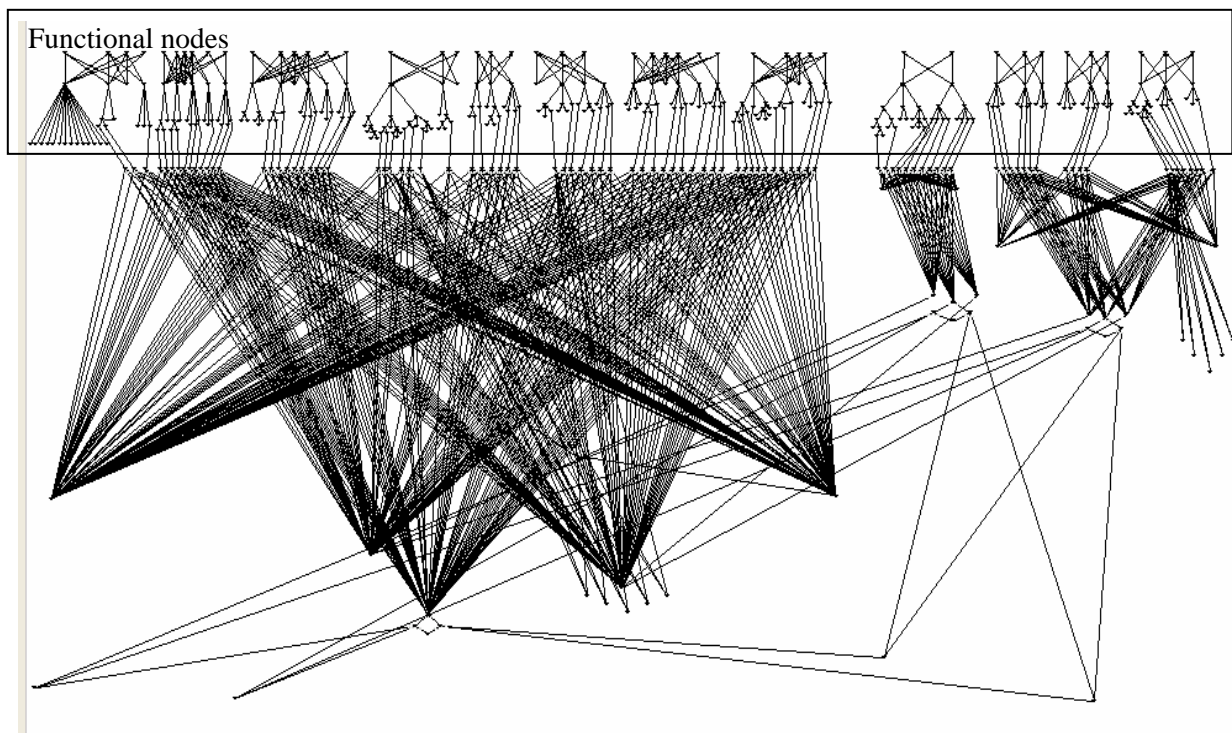

Figure 1: BBN for air safety, 330 functional nodes, 133 probabilistic nodes.

3. Design Philosophy

The basic design philosophy was introduced in Kurowicka and Cooke (2005) and Hanea et al. (2006). Influence is represented as conditional rank correlation.

The conditional rank correlations between parent and child are shown in Figure 2. The order of conditioning can be changed by the user. This representation of influence was chosen because (Kurowicka and Cooke 2005):

- The numerical values of the conditional rank correlations are algebraically independent,

- Together with the conditional independence statements implied by the graph, the univarate marginal distributions, and a copula realizing the correlations, they uniquely determine the joint distribution. If zero correlation corresponds to the independent copula, then the conditional probability statements implied by the graph are satisfied.
- Influences can be added without changing the values already chosen (unlike partial regression coefficients),
- Conditioning must be performed by simulation or by the 'Hybrid'method of Hanea et al 2006, except in the case of the joint normal copula, where conditioning on probabilistic nodes can be done analytically.
- Tested protocols for eliciting these values from experts exist, and do not depend on the expert's marginal distributions (see below).
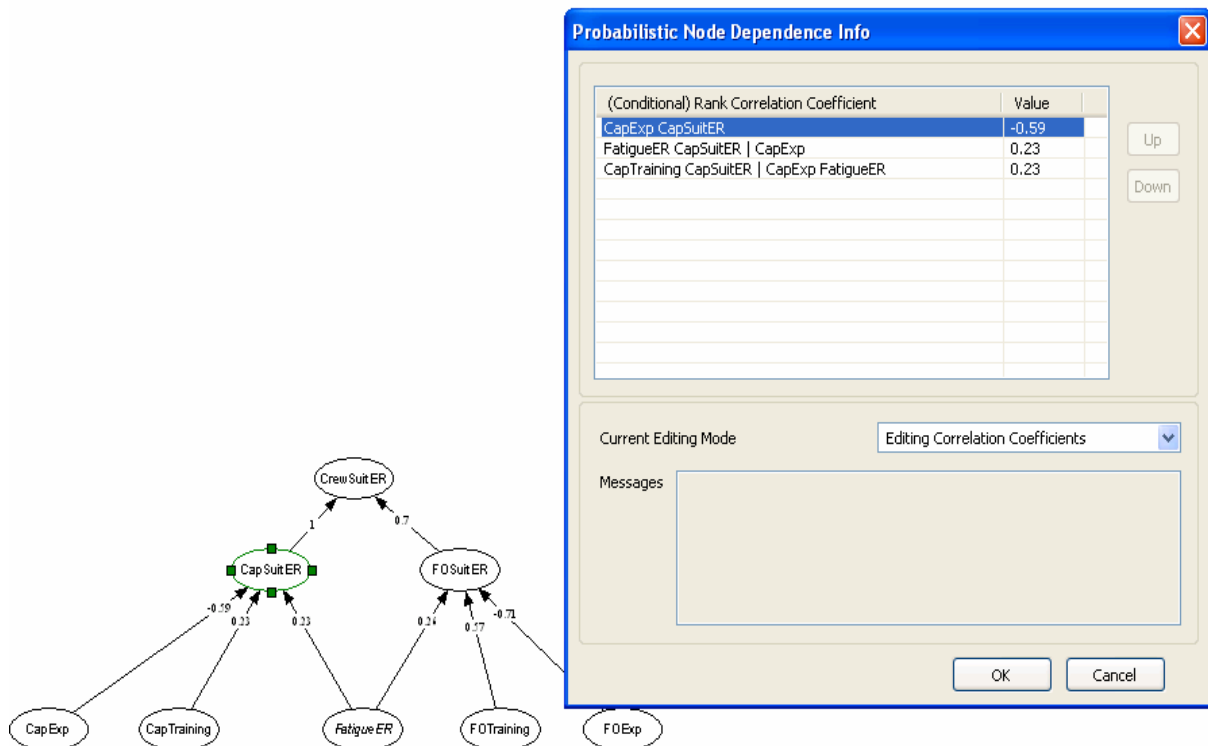

Figure 2: Assignment of conditional rank correlations to nodes, for Crew Suitability En Route (Crew suit ER)

Switching to the distribution view, we can observe the effects of conditionalization. Figure 3 shows the unconditional distributions for Crew Suitability En Route. The horizontal histogram for FatigueER indicates that this is a discrete distribution. Dealing with rank correlations for discrete variables is discussed in Hanea and Kurowicka (2007). The expectations and standard deviations are shown below each distribution. Crew suitability is influenced by the captain's suitability and also that of the first officer. The captain's suitability is defined as the number of captains, out of 10,000, who would fail the next proficiency check test. The negative correlation between the Captain's experience in flight hours (CapExp) and CapSuitER reflects the fact that those with more experience are *less* likely to fail the "prof. check test". Captain's training (CapTraining) is the time since the last training course.

In Figure 4 we conditionalize on the value CapExp = 10,000 flight hours. The expected number of captains, out of 10,000 failing the next prof check test, given that they have 10,000 hours flight experience, drops from 234 to 171. The standard deviation drops from 235 to 166. Conditionalizing the entire distribution in this way takes about 5 sec. on a fast PC.
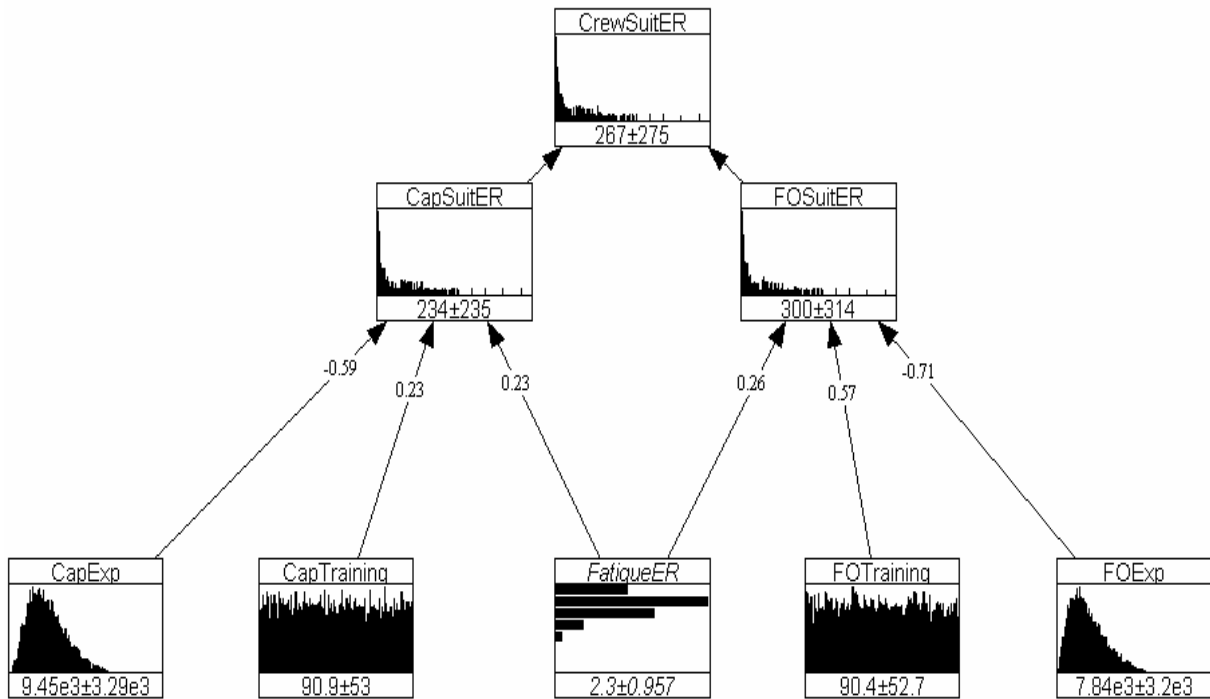

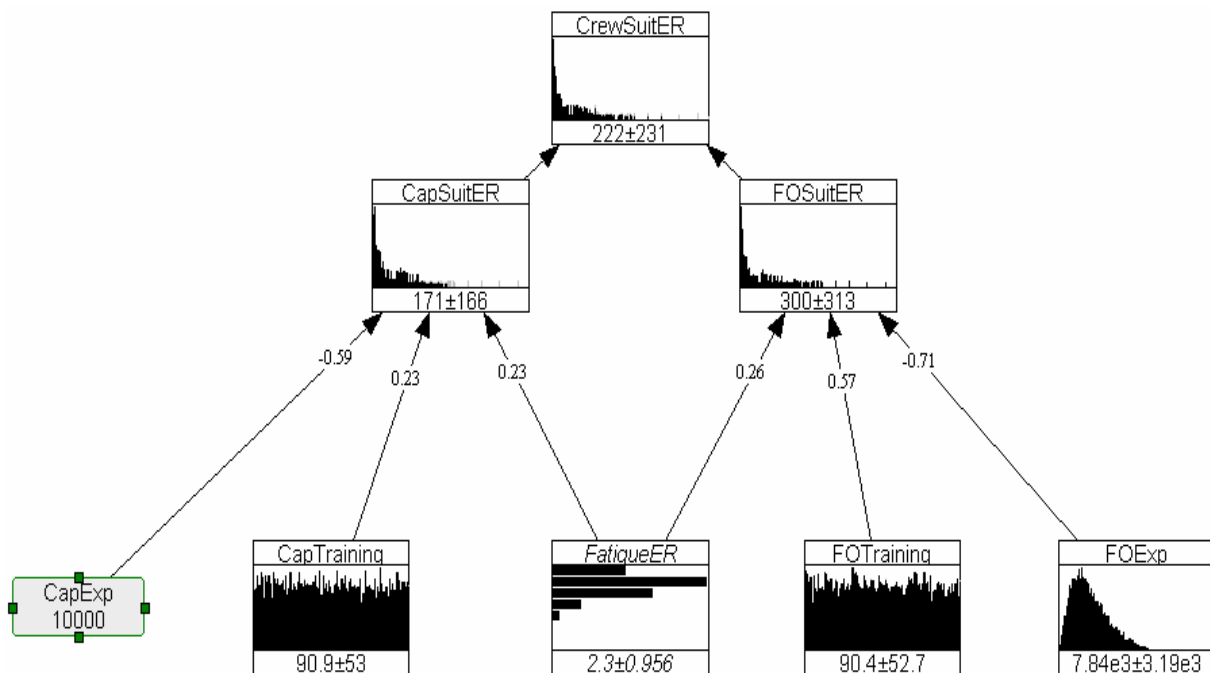Figure 3: Unconditional distributions for crewsuitER

According to this model an un-recovered loss of control from spatial disorientation in the en-route flight phase might happen if a flight crew member is spatially disoriented **AND** the flight crew fails to maintain control. Nodes in Figure 5 are Boolean functions of probabilistic nodes. The probability of the "top" event in the unconditional joint distribution is 9.56e-8.
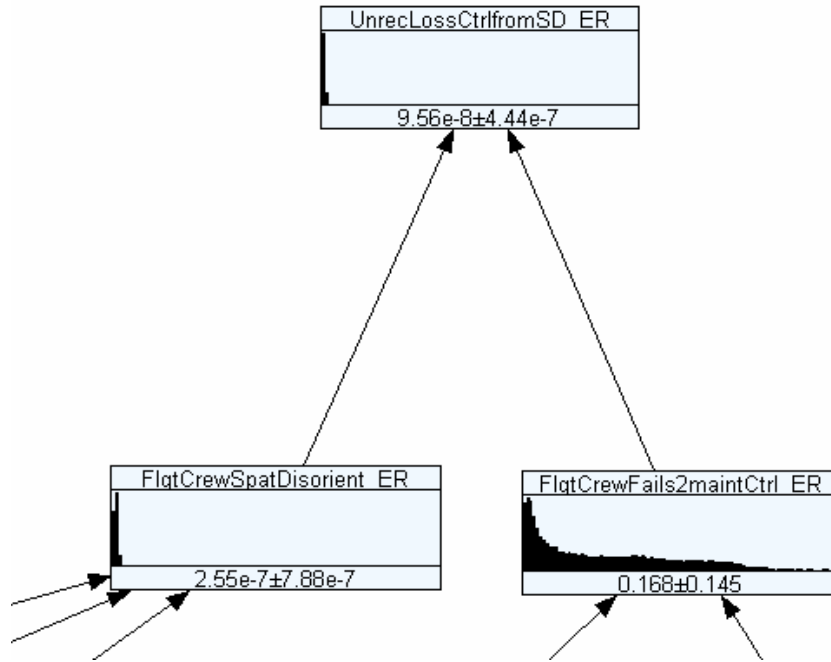

Figure 5: Un-recovered loss of control after spatial disorientation.
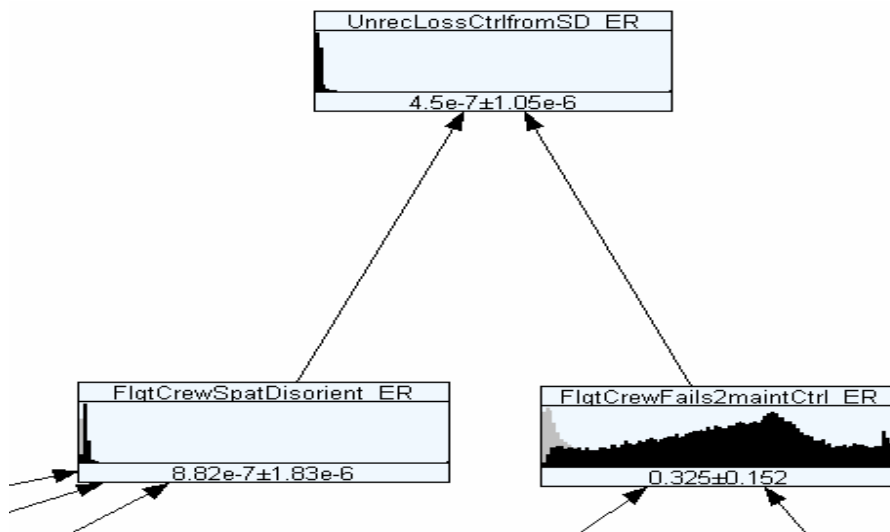

Figure 6: Un-recovered loss of control after spatial disorientation

Suppose that we condition on the expected number of captains, out of 10,000 failing the next prof check test being equal to 1,000. After updating the joint distribution, the probability of an un-recovered loss of control after spatial disorientation becomes approximately 4.7 times larger than in the unconditional distribution. Observe that both the probability of a flight crew member being spatially disoriented and the flight crew failing to maintain control have increased to 8.82e-7 and 0.325 respectively. The unconditional (gray) and conditional (black) distributions are both visible in the FlgtcrewFail2maintctrl ER box (in Figure 4 the overlap is hardly visible).

4.   Elicitation

Methods for eliciting conditional rank correlations from experts can also be applied to infer these correlations from data, if such data are available. The method employed in this application is based on the fact that the normal copula is used for all distributions. To assess the correlations between CapsuitER and its parents, experts are first asked:

Suppose for a given captian, the experience is known to lie above the median value, what is now your probability that CapsuitER is also above its median value?

If CapExp has no influence onCapsuitER, the answer will be ½; if there is a strong positive influence, the answer will be near one, with strong negative influence it will be near zero.

After answering this question, the expert is asked:

Suppose for a given captian, the experience is known to lie above the median value AND ALSO the time to last training is above its median value, what is now your probability that CapsuitER is also above its median value?

The possible values are now constrained by the previous answer and by the choice of copula. An elicitation tool was created to compute the feasible bounds for the current value given the preceding information. A screen shot from this tool is shown below in Figure 7. The first conditional exceedance probability, 0.7, led to a conditional rank correlation of 0.57. The second exceedance probability is constrained to the interval [0.4, 1].

The experts in this case were commercial airline pilots, air traffic controllers, and air safety professionals. After initial hesitation, they become quickly comfortable with these assessment tasks, and learned to appreciate the meaning of the dependence relations. The elicitation tool proved very helpful in this regard. Details of the elicitation procedure are described in Morales et al. (appearing).

5.   Conclusion

The problems identified in Hanea et al. (2006) and Kurowicka and Cooke (2005) regarding continuous and discrete BBNs have been largely overcome, along the lines indicated in those publications. The examples in this paper show that the conditional rank correlation representation of influence, with the joint normal copula, can handle large problems without excessive assessment burden, without excessive computational burden, and without reverting to partial regression coefficients in 'normal units' different from the physical units of the variables involved. Elicitation protocols are easy and intuitive, and have been successfully applied with domain experts without special training in the elicitation tasks. Combining BBN's with fault trees, or indeed with any other functional nodes poses

no special problems. Of course we do not obtain a list of minimal cut sets. Repeated basic events in a fault tree are handled simply by making them functionally dependent on one another.
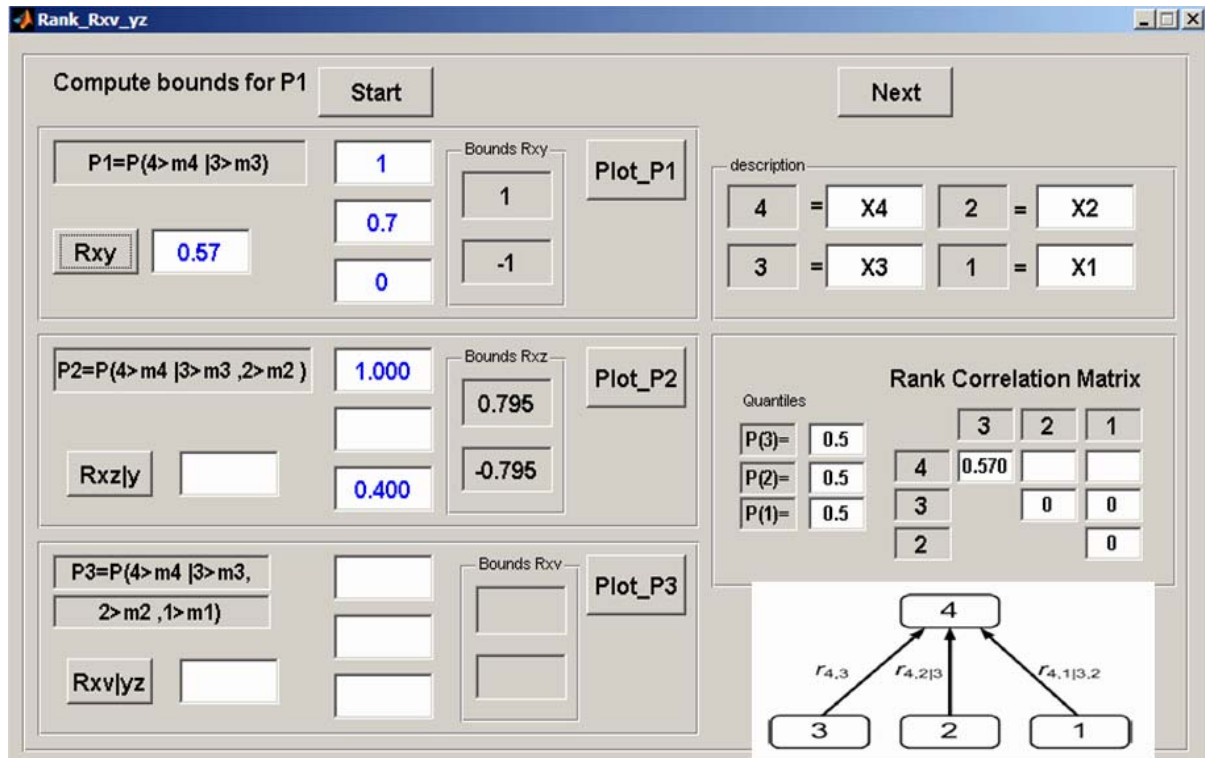

Figure 7: Elicitation of conditional rank correlations via conditional exceedance probabilities

Challenges remain, however. The requisite computational speed can only be obtained with the joint normal copula. It is hoped that other copula could be used in the future. Another challenge is posed by conditioning on functional nodes or conditioning on intervals of probabilistic variables. In the current framework, such conditionalization can only be preformed on samples generated by the BBN. More sophisticated sampling and computational methods might offer other possibilities.

References

Cooke, R.M. (1991) Experts in Uncertainty, Osford University Press.

Hanea, A.M., Kurowicka, D. and Cooke, R.M. (2006). Hybrid Method for Quantifying and Analyzing Bayesian Belief Nets, *Quality and Reliability Engineering International* 22, pp 709-729.

Kurowicka, D. and Cooke, R.M. (2005). Distribution - Free Continuous Bayesian Belief Nets. In A. Wilson, N. Limnios, S. Keller-McNulty and Y. Armijo, (Eds.), *Modern Statistical and Mathematical Mathematical Methods in Reliability.* pp 309-323

Morales, O., Kurowicka. D. and Roelen, A. (appearing). Eliciting Conditional and Unconditional Rank Correlations from Conditional Probabilities, *Reliability engineering and System Safety*.

Hanea, A. and Kurowicka, D. (2007) Mixed Non-Parametric Continuous and Discrete Bayesian Belief Nets , *MMR*, Glasgow.

Kurowicka, D. and Cooke, R.M. (2007). Sampling algorithms for generating joint uniform distributions using the vine-copula method. *Computational Statistics & Data Analysis*, 51(6), pp 2889-2906 .

Ale, B.J.M., L.J. Bellamy, R.M. Cooke, L.H.J. Goossens, A.R. Hale, D. Kurowicka, A.L.C. Roelen, E.Smith (2005). Development of a causal model for air transport safety. In Kolowrocki (Eds)S, *Advances in Safety and Reliability (ESREL),* Taylor and Francis Group, London, ISBN, 0415 383404.

Ale,B.J.M., L.J. Bellamy, R.M. Cooke, L.H.J.Goossens, A.R. Hale, A.L.C.Roelen, E. Smith, (2006). Towards a causal model for air transport safety – an ongoing research project, *SAFETY SCIENCE*, 44(8), pp 657–673.

B.J.M. Ale , L.J. Bellamy*,* R. van der Boom J., R.M. Cooke, L.H.J. Goossens, .R. Hale, D. Kurowicka and O. Morales, A.L.C. Roelen , J. Spouge (2007). Further development of a Causal model for Air Transport Safety (CATS); building the mathematical heart *Reliability and Societal Safety – Aven & Vinnem (eds) Taylor & Francis Group, London, ISBN 978-0-415-44786-7*