Eliciting and Representing Joint Distributions From Experts

Quantification of a Human Performance Model for Risk Analysis

Master Thesis

by

Wiebke Jäger

Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

Student no: 4125576
Program: Applied Mathematics
Specialization: Probability, Risk, Statistics
Graduation date:

Examination committee: Prof. dr. F. H. J. Redig, Applied Probability, Chairman
Dr. A. M. Hanea, Applied Probability, First supervisor
Dr. ir. C. van Gulijk, Safety Science Group, Second Supervisor
Dr. ir. O. Morales Nápoles, Applied Probability, Additional Member
Acknowledgments

I am grateful to my committee members, Frank Redig, Anca Hanea, Oswaldo Morales Nápoles and Coen van Gulijk for administering my thesis and sharing their knowledge with me. I would especially like to thank Anca Hanea, not only for giving me valuable advice and support in the writing of this thesis, but also for encouraging me in all ambitions I had, and Oswaldo Morales Nápoles for taking the time to explain his research to me and for the insightful discussions on my results. Thanks also to Daniela Hanea for introducing me to the research problem.
Abstract

In potentially hazardous industries, human factors are widely recognized as more fundamental causes to accidents and incidents than random technical errors. Therefore, human performance modeling for probabilistic risk analysis aims to give insight in the way human factors relate to the human error probability and how they influence the overall risk of engineering systems.

Nonparametric Bayesian belief nets (BBNs) are a promising tool to integrate human and technical influences on the system risk in a single homogeneous model. They are a type of probabilistic graphical model consisting of nodes (variables) and arcs (influences). Their main use is to make inferences about uncertain states when information is limited and they are able to realistically reflect inherent variabilities of a system. This is important for human performance models, because, unlike pieces of machinery, humans are likely to behave differently from each other if confronted with the same situation. The subject of this thesis is a human performance model on seven variables for the oil and gas industry.

Data is available in the form of expert opinion. Structural expert judgment is, when performed rigorously, a powerful method for obtaining uncertainty distributions. While eliciting marginal distributions and pooling them with performance based weights is a time-tested approach, the elicitation and combination of dependence estimates, more precisely rank correlations, is still a very active research topic. The objective of this thesis is to further investigate the challenges of eliciting joint distributions for nonparametric BBNs from experts based on the present data set.

We have found that most of the experts provided inconsistent dependence estimates. This points up how important it is that experts receive adequate training in assessing probabilities and correlations and how crucial it is that the model structure is explained in-depth and its implications are clear to the experts before the actual elicitation. Moreover, the data suggests that not all experts who perform well in eliciting marginal distributions are skilled dependence elicitors as well. Unfortunately, guidelines on how to evaluate the quality of experts’ dependence estimates and to devise weights for their combination are still missing. For this thesis different models have been generated based on different assumptions and compared to each other. This also shows that expert judgment is not a mathematical theory leading to a unique model, but that the choices of the analyst play a crucial role for the resulting model.

With regard to the human performance model, we recommend to validate one of the models presented in this thesis with the help of the experts that participated in the elicitation. However, the preferable option is to gather new and more accurate data. The main suggestion for ongoing research is to further develop a method to rate the experts’ performance of assessing rank correlations and a corresponding weighting scheme.
Contents

List of Figures vii

List of Tables ix

I Introduction 1

1.1 Context and Scope of the Thesis ........................................... 1
1.2 Risk, Uncertainty and Probability ........................................ 2
1.3 Research Objective ....................................................... 3
1.4 Research Questions and Reading Guide .................................. 4

I Nonparametric BBNs as a Probabilistic Risk Analysis Tool 7

2 Preliminary Concepts and Definitions 9

2.1 Product Moment Correlation and Rank Correlation ..................... 9
2.2 Copulas ................................................................. 10
  2.2.1 The normal copula .................................................... 11
  2.2.2 The t-copula ........................................................ 12
2.3 Vines ................................................................. 13
2.4 Bayesian Belief Nets ..................................................... 15
  2.4.1 Discrete BBNs ........................................................ 17
  2.4.2 Non-parametric continuous BBNs ................................. 18

3 The Zero Independence Property and Nonparametric BBNs 23

  3.1 Representation of Independence: Graph Structure versus Quantification .... 24
  3.2 A Kernel-Based Conditional Independence Test ......................... 27
  3.3 Simulation Study ..................................................... 29
  3.4 Extension of the nonparametric BBN Theorem ........................... 31
  3.5 Conclusion ............................................................ 35

II Quantifying a HPM with Structured Expert Judgment 37

4 Introduction to the Operator Model 39

5 Structured Expert Judgment .................................................. 41
  5.1 The Classical Model for Eliciting Marginal Distributions .................. 41
  5.2 Eliciting Dependencies .................................................. 44
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2.1 Assessing Ratios of Rank Correlations</td>
<td>44</td>
</tr>
<tr>
<td>5.2.2 Combining Dependence Estimates</td>
<td>47</td>
</tr>
<tr>
<td>5.3 Conclusion</td>
<td>49</td>
</tr>
<tr>
<td>6 Elicitation and Results</td>
<td>51</td>
</tr>
<tr>
<td>6.1 Marginal Distributions</td>
<td>51</td>
</tr>
<tr>
<td>6.2 Rank Correlations</td>
<td>53</td>
</tr>
<tr>
<td>6.3 Weights</td>
<td>61</td>
</tr>
<tr>
<td>6.4 Conclusion</td>
<td>62</td>
</tr>
<tr>
<td>7 Combining Experts</td>
<td>65</td>
</tr>
<tr>
<td>7.1 Combining Marginal Distributions</td>
<td>65</td>
</tr>
<tr>
<td>7.2 Combining Rank Correlations</td>
<td>66</td>
</tr>
<tr>
<td>7.2.1 Model O</td>
<td>66</td>
</tr>
<tr>
<td>7.2.2 New Models</td>
<td>67</td>
</tr>
<tr>
<td>7.2.3 Resulting Models</td>
<td>75</td>
</tr>
<tr>
<td>7.3 Conclusion</td>
<td>80</td>
</tr>
<tr>
<td>8 Assumption of Dependent Parents</td>
<td>83</td>
</tr>
<tr>
<td>8.1 Conclusion</td>
<td>90</td>
</tr>
<tr>
<td>9 Model Validation</td>
<td>91</td>
</tr>
<tr>
<td>10 Conclusions</td>
<td>93</td>
</tr>
<tr>
<td>10.1 Main Findings</td>
<td>93</td>
</tr>
<tr>
<td>10.2 Recommendations</td>
<td>95</td>
</tr>
<tr>
<td>10.3 Suggestions for Further Research</td>
<td>96</td>
</tr>
<tr>
<td>References</td>
<td>102</td>
</tr>
<tr>
<td>III Appendices</td>
<td>103</td>
</tr>
<tr>
<td>A Protocol for the Elicitation</td>
<td>105</td>
</tr>
<tr>
<td>B Additional KCI-test results</td>
<td>107</td>
</tr>
<tr>
<td>B.1 Example Run 1</td>
<td>107</td>
</tr>
<tr>
<td>B.2 Example Run 2</td>
<td>108</td>
</tr>
<tr>
<td>C Presentation</td>
<td>110</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Part of the technical model for the petrochemical application (source: Hanea et al., 2012)</td>
<td>2</td>
</tr>
<tr>
<td>2.1</td>
<td>A D-vine on 5 elements (source: Kurowicka and Cooke, 2005)</td>
<td>15</td>
</tr>
<tr>
<td>2.2</td>
<td>A simplistic BBN</td>
<td>15</td>
</tr>
<tr>
<td>2.3</td>
<td>Examples of the possible conditional independence relations induced by a BBN</td>
<td>16</td>
</tr>
<tr>
<td>2.4</td>
<td>Parents and child of a BBN on 4 variables (source: Hanea, 2008)</td>
<td>18</td>
</tr>
<tr>
<td>2.5</td>
<td>Conditional Probability Table for $X_4$ (source: Hanea, 2008)</td>
<td>18</td>
</tr>
<tr>
<td>2.6</td>
<td>BBN with one child node and six parent nodes</td>
<td>18</td>
</tr>
<tr>
<td>2.7</td>
<td>Vine corresponding to the BBN in figure 2.6</td>
<td>21</td>
</tr>
<tr>
<td>3.1</td>
<td>D-vine representing the independence of $X_1$ and $X_2$ (source: Hanea, 2010)</td>
<td>24</td>
</tr>
<tr>
<td>3.2</td>
<td>Different graph structures and specifications leading to the same CI statements for normal and t-copula</td>
<td>25</td>
</tr>
<tr>
<td>3.3</td>
<td>Different graph structures and specifications leading to different CI statements for normal and t-copula</td>
<td>26</td>
</tr>
<tr>
<td>3.4</td>
<td>Example BBN on three nodes</td>
<td>32</td>
</tr>
<tr>
<td>3.5</td>
<td>Vine on 3 variables</td>
<td>33</td>
</tr>
<tr>
<td>4.1</td>
<td>Human performance model to be quantified</td>
<td>40</td>
</tr>
<tr>
<td>5.1</td>
<td>BBN with one child node and seven parent nodes</td>
<td>41</td>
</tr>
<tr>
<td>5.2</td>
<td>Relationship between $P{X_7 \geq x_{7</td>
<td>e_5} \mid X_1 \geq x_{1</td>
</tr>
<tr>
<td>5.3</td>
<td>Possible values of $r_{2,7</td>
<td>1}$ given that $r_{2,7</td>
</tr>
<tr>
<td>5.4</td>
<td>Two BBNs of three nodes</td>
<td>47</td>
</tr>
<tr>
<td>5.5</td>
<td>Marginal distributions of the DM an experts $e_5$ and $e_6$</td>
<td>49</td>
</tr>
<tr>
<td>6.1</td>
<td>Distributions of experts $e_5$ and $e_6$ for $V_7$</td>
<td>52</td>
</tr>
<tr>
<td>6.2</td>
<td>Marginal distributions of experts $e_4$, $e_5$, $e_6$ and $e_9$ for the child node $V_7$</td>
<td>62</td>
</tr>
<tr>
<td>6.3</td>
<td>Marginal distributions of experts $e_4$, $e_5$, $e_6$ and $e_9$ for the parent nodes $V_1 - V_6$</td>
<td>62</td>
</tr>
<tr>
<td>7.1</td>
<td>Distributions of experts $e_5$, $e_6$ and the DM for $V_7$</td>
<td>65</td>
</tr>
<tr>
<td>7.2</td>
<td>Differences in the marginal distributions of the experts and decision maker A</td>
<td>72</td>
</tr>
<tr>
<td>7.3</td>
<td></td>
<td>74</td>
</tr>
<tr>
<td>7.4</td>
<td>Model O: Human performance model as quantified previously</td>
<td>76</td>
</tr>
<tr>
<td>7.5</td>
<td>Human performance model A: global weights $w_5$ and $w_6$</td>
<td>77</td>
</tr>
<tr>
<td>7.6</td>
<td>Human performance model B: equal weights $w_6$ and $w_5$</td>
<td>78</td>
</tr>
</tbody>
</table>
7.7 Human performance model O conditioned on \( V \gamma = 0.7699 \) ........................................ 79
7.8 Human performance model A conditioned on \( V \gamma = 1 \) ........................................ 79
7.9 Human performance model B conditioned on \( V \gamma = 1 \) ........................................ 79

8.1 ................................................................. 83
8.2 Relationship between the bounds for \( \rho_{27} \) and \( \rho_{12} \) for various precedent assessment of \( \rho_{17} \) ......................................................... 84
8.3 Relationship between the bounds for \( \rho_{27} \) and \( \rho_{12} \) for a precedent assessment of \( \rho_{17} = 0.9 \) based on equation (8.4) ........................................ 85
8.4 Relationship between the bounds for \( \rho_{12} \) and \( \rho_{271} \) for \( \rho_{17} = 0.9 \) and \( \rho_{27} = 0.8 \) ........................................ 86
8.5 \( \rho_{17} = 0.9 \) and \( \rho_{27} = 0.95 \) (Within \(-0.7 \leq \rho_{271} \leq 0.7 \) the values for \( \rho_{12} \) have imaginary parts.) ........................................ 86
8.6 ................................................................. 86
8.7 The relationship between \( \rho_{27} \), \( \rho_{12} \) and \( \rho_{271} \) for \( \rho_{17} = 0.9 \) ........................................ 87
8.8 Views along the x-axis and the y-axis of figure 8.7 ........................................ 87
8.9 The relationship between \( \rho_{27} \), \( \rho_{12} \) and \( \rho_{271} \) for \( \rho_{17} = -0.3 \) ........................................ 87
8.10 Views along the x-axis and the y-axis of figure 8.9 ........................................ 88

9.1 Flowchart illustrating possible steps towards model validation ..................... 91
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Overview of common copulas and the zero independence property (for more details see e.g. [Schmidt, 2006][Kurowicka and Cooke, 2006][Gatz, 2007])</td>
<td>24</td>
</tr>
<tr>
<td>3.2</td>
<td>Average p-values for the KCI-test computed from 100 tests</td>
<td>31</td>
</tr>
<tr>
<td>4.1</td>
<td>Definitions of the Variables in the BBN</td>
<td>40</td>
</tr>
<tr>
<td>5.1</td>
<td>Experts’ quantiles</td>
<td>49</td>
</tr>
<tr>
<td>6.1</td>
<td>Excalibur Output: optimized global and item weights with significance level $\alpha = 0.6876$</td>
<td>53</td>
</tr>
<tr>
<td>6.2</td>
<td>Experts’ ranking of the parent nodes according to highest absolute rank correlation of the child node $V_7$</td>
<td>54</td>
</tr>
<tr>
<td>6.3</td>
<td>Results of the dependence elicitation. For computations in Matlab the (conditional) rank correlations have been discretized as $-0.99 : 0.01 : 0.99$</td>
<td>55</td>
</tr>
<tr>
<td>6.4</td>
<td>Global weights without optimization and significance level $\alpha = 0$</td>
<td>62</td>
</tr>
<tr>
<td>7.1</td>
<td>Overview of the increase of the mean $\mu$, $\Delta \mu(V)$, when conditioning on $V_7 = 1$ for model A, B and C and on $V_7 = 0.7699$ for model O</td>
<td>79</td>
</tr>
<tr>
<td>7.2</td>
<td>Comparison of the effect on $V_7$ of conditioning all parent nodes individually in model A and model B</td>
<td>80</td>
</tr>
<tr>
<td>8.1</td>
<td>Values of $\rho_{12,+}$ and $\rho_{12,-}$ for $\rho_{27</td>
<td>1} = 1$ and $\rho_{27</td>
</tr>
<tr>
<td>8.2</td>
<td>Lower and upper bounds for the correlations between $X_1^{+}$ and $X_2^{+}$</td>
<td>89</td>
</tr>
<tr>
<td>B.1</td>
<td>P-values for the KCI-test</td>
<td>108</td>
</tr>
<tr>
<td>B.2</td>
<td>P-values for the KCI-test</td>
<td>109</td>
</tr>
<tr>
<td>C.1</td>
<td>Excalibur Output: optimized global and item weights with significance level $\alpha = 0.6876$</td>
<td>110</td>
</tr>
</tbody>
</table>
1. Introduction

High-hazard industries, such as nuclear or oil and gas, are complex socio-technical systems in which humans have been identified as the "weakest links" [Turner and Pidgeon, 1978]. Human errors are often more fundamental causes of major accidents than random technical errors. They can occur in various tasks, such as regulation, administration, management, design, installation, maintenance, inspection or operation of the engineering system.

Some well-known examples of accidents that were largely caused by human error are the Three Mile Island accident (1979), the Bhopal gas tragedy in India (1984), the space shuttle Challenger disaster (1986) or, more recently, the deepwater horizon blow-out in the gulf of Mexico (2010) and the Fukushima Daiichi nuclear disaster in Japan (2011). Because the consequences are often fatal, risk management of such industries quests for "no-accident-is-tolerable" strategies [Rasmussen, 1997]. In spite of many years of research, satisfactory models for human performance are still lacking [Bedford and Cooke, 2001a], low-probability high-consequence accidents do not appear to occur less frequently and human factors keep being cited as major contributors. Advancing approaches to human performance modeling is therefore of great social relevance.

1.1 Context and Scope of the Thesis

For quantitative modeling the human error probability (HEP), the probability that a task is carried out erroneously, is a useful measure for human performance. Naturally, the HEP is influenced by the circumstances in which a person works, e.g. time pressure, workload, adequacy of training, regulation and procedures, complexity of task or experience. These factors are called human factors (HF) or performance shaping factors (PSF), a term first introduced by Swain and Guttmann, 1983.

Human performance modeling attempts to give insight in the way human factors relate to the human error probability and how that, in turn, affects the overall 'system risk'. It is usually part of a larger probabilistic risk analysis (PRA) and applied after a comprehensive system analysis. PRA is a systematic methodology to evaluate risks associated with complex engineering systems that aims to answer three basic questions [U.S.NRC, 2012]:

1. What can go wrong?
INTRODUCTION

2. How likely is it?

3. What are the consequences?

The safety science group at the Delft University of Technology has made significant efforts to conduct a wider-reaching PRA by integrating human and technical influences on the system risk in a single homogeneous model\(^1\). A recent application is CATS, a model for air transport safety [Ale et al., 2006a, Ale et al., 2006b, Ale et al., 2009, Lin, 2011]. Ongoing efforts [Lin et al., 2012, Ale et al., 2012, Hanea et al., 2012, Jäger, 2013] concentrate on an application for a petrochemical plant and attempt to include also dynamic relations.

Typically, these kind of risk models contain hundreds of variables and relationships. Figure 1.1 depicts part of the technical risk model for the petrochemical plant. This model is a so-called nonparametric Bayesian belief net (BBN), a probabilistic graphical model representing variables and their relations as nodes and arcs. BBNs belong to the class of directed acyclic graphs. A cause-effect relationship is indicated by the direction of an arc. In this manner they can display very complicated relationships between a multitude of random variables in an intuitive manner. For nonparametric BBNs the nodes are quantified as marginal probability distributions and the arcs are quantified as (conditional) rank correlations, which are realized by a copula for which zero rank correlation implies independence. A BBN reflects all the possible states of the world in form of the joint distribution of all variables in the model.

\(^1\)As opposed to several different models or hybrid models
Different parts of a BBN can be built separately and connected at a later stage. The various distinct clusters of nodes and arcs in figure 1.1 suggest that they have been individual sub-models. This thesis is engaged with the quantification of such a sub-model: a human performance model for a specific group of operators, i.e. those people who directly interact with the technical interface. Later on, the sub-model will be, at least it is intended to be, connected to the main BBN. We will refer to this model as operator model.

Because the notion of risk as well as of uncertainty and their relation to probability are the foundation of PRA, we briefly describe what we mean by those terms in the following section, before describing the research problem and objectives in detail.

1.2 Risk, Uncertainty and Probability

Although the term risk is used in many different ways throughout the literature, (probably) all definitions include the two dimensions extent of consequence and probability of occurrence for a risk event. In the most basic version it is the product of the two. Risk management is the process of keeping hazards under control, i.e. reducing or eliminating the probability for events with unacceptable consequences.

The notion of uncertainty is tight-knit with the notion of risk. Uncertainty “is that what disappears when we become certain” [Bedford and Cooke, 2001a, p.19]. In a practical context, certainty is achieved through observation or, said differently, by learning about the system. However, not all kinds of uncertainty can be observed; there is a distinction between aleatory and epistemic uncertainties. Aleatory uncertainties refer to the natural variability of the system; they cannot be learned. In contrast, epistemic uncertainties refer to lack of knowledge of the system. A quantitative measure of uncertainty is probability. Having absolutely no clue about the outcome of an event can be interpreted as a uniform distribution. Knowledge about the event is reflected by shape and range of the distribution. [Bedford and Cooke, 2001a]

Human factors include both types of uncertainty. At least for the purpose of modeling human factors, the following categorization is useful. Predicting the performance of a random individual in a random situation is impossible. Indeed, a significant part of human behavior is cryptic to bystanders: genetic endowments and previous experiences of life are probably the main determinants. This is the aleatory uncertainty. However, we could learn about certain preconditions of the individual that affect his/her performance, e.g. knowledge, stress level or fatigue, and in this way reduce the epistemic uncertain.

Because a great deal of uncertainty is always involved when modeling humans, and especially an entire "population" of humans, we quantify the HEP as a, parametric or nonparametric, probability distribution on \([0, 1]\).

\(^2\text{cf. [Bedford and Cooke, 2001a, section 2.6] for an example on how the same uncertainty may be classified differently for different models with different goals.}\)
1.3 Research Objective

The standard interpretation of probability is the frequency interpretation. It defines the probability of an event as the limit of its relative frequency in a large number of trials, i.e. if $S$ is a finite set of outcomes of an experiment that can be repeated indefinitely under identical conditions and let $A \in S$, then $P(A) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I_{A_i}$, where $I_{A_i}$ is the indicator function of the event that the outcome of the $i$th experiment is in $A$. In practice this probability is approximated with the relative frequency in many trials. [Bedford and Cooke, 2001a]

However, this interpretation is for two reasons not optimal for the human error probabilities. First, the error event has to be well specified, which seems difficult for humans, because they, unlike pieces of machinery, can act and fail in infinite and a priori unimaginable ways. Second, empirical data appropriate for estimating probabilities needs to be present. Since probabilistic risk analysis deals with failure events that almost never happen, it is often impossible to obtain sufficiently large amount of observations. For low probability - high impact events historical data is rarely available and experiments are generally infeasible, because the consequences would be unacceptable, the costs would be too high, it would be technically too difficult and/or the situation to be modeled would be too unique. An additional complication for human factors is that it is difficult to operationalize them such that they are measurable. How would one for example go about measuring "insufficient knowledge to execute a particular task"?

Under these conditions, expert opinion seems to be a suitable source of data. The classical model of expert judgment, proposed by [Cooke, 1991], assumes the experts to be rational individuals who provide their subjective probabilities for the quantities of interest. The subjective interpretation of probability is based on Savage’s theory of rational decision. Let $S$ be a finite or infinite set of possible states of the world (or a set of possible worlds), $s^*$ denote the actual real state and $A \in S$, then $P(A)$ is a degree of belief of a rational individual that $s^*$ is an element of $A$. When working with subjective probabilities of different experts it is important to find a rational consensus. The classical model achieves this by combining expert distribution with performance based weights. These weights are obtained from so-called calibration questions for which the answers are known to the analyst but not to the experts.

Quantifying a nonparametric BBN involves determining the marginal probability distributions as well as the (conditional) rank correlations of all arcs. Marginal distributions have been elicited and combined numerous times (see [Goossens et al., 2008] for examples), while very little experiences have been made with the elicitation and combination of (conditional) rank correlations [Morales Nápoles et al., 2008]. Morales-Nápoles et al., 2013]. Prior to this thesis project, the variables and the structure of the operator model have been defined and the expert elicitation has been completed. Also first steps to build a BBN from the data have been taken. However, it is suspected that both the elicitation as well as combination procedure have shortcomings. The thesis further investigates the challenges of combining experts’ assessment based on the present data and suggests ways to deal with them. It aims to deliver a set of considerations for both a good practice of eliciting experts and treating the resulting data.
1.4 Research Questions and Reading Guide

The research question guiding this thesis work may be formulated as

\[
\textit{How can the joint distribution of a nonparametric BBN be inferred from structured expert judgment?}
\]

and sub-divided into the following four sub-questions.

1. Why are nonparametric BBNs beneficial to represent a joint distributions elicited from experts?
2. What went wrong in the dependence elicitation and previous combination?
3. How serious are the shortcomings of the conducted dependence elicitation and previous combination?
4. What makes the combination of dependence estimates particularly challenging?

The thesis is organized in two parts. The first sub-question is treated in the first part. It concerns itself with the methodology of nonparametric Bayesian belief nets. Chapter 2 lays the theoretical foundation for the thesis. It reviews the theory of nonparametric BBNs as well as the necessary background knowledge, including correlations, copulas and vines. Currently these BBNs are restricted to copula families for which zero rank correlation implies independence. However, there has been doubt whether this restriction is necessary. For this reason chapter 3 contains a small simulation study, which compares the normal and the t-copula. Finally the main proof is reformulated as to accommodate all kinds of copulas.

The second part deals with the quantification of nonparametric BBNs based on expert opinion, that is the last three sub-questions. Chapter 4 introduces the operator model. Chapter 5 describes the classical model of structured expert judgment and describes the hitherto made efforts to combine dependences for other applications. Chapter 6 is dedicated to the conducted expert judgment elicitation and the results, while chapter 7 is concerned with the combination of the experts’ joint distributions to a decision maker distribution. Since, the experts’ estimates are found to be very inconsistent, we question in chapter 8 whether the model structure of the operator actually captures the relations between the variables that the experts belief in. Chapter 9 touches upon a possible validation of the models and finally, conclusions of both parts are given in chapter 10.
Part I

Nonparametric BBNs as a Probabilistic Risk Analysis Tool
2. Preliminary Concepts and Definitions

This chapter presents the theoretical background needed for the following chapters. In particular, it introduces the notion of correlation as well as copulas laying the foundation for two probabilistic graphical models, vines and Bayesian belief nets, which are described afterwards. Definitions, propositions and theorems are taken from [Kurowicka and Cooke, 2006] unless stated otherwise.

2.1 Product Moment Correlation and Rank Correlation

The product moment correlation and the rank correlation are very common dependence measures. They are defined below.

Definition 2.1 (Product moment correlation / linear correlation). The product moment correlation of two random variables $X$ and $Y$ with finite expectations and nonzero finite variances is

$$
\rho_{XY} = \frac{E(XY) - E(X)E(Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.
$$

(2.1)

The product moment correlation is a measure of linear dependence. If $X = aY + b$ for some $a > 0, b \in \mathbb{R}$, then $\rho(X, Y) = 1$, and the converse holds. In other cases $-1 < \rho(X, Y) < 1$. Moreover, for $\alpha, \gamma \in \mathbb{R} \setminus \{0\}, \beta, \gamma \in \mathbb{R}$ it has the property that

$$
\rho(\alpha X + \beta, \gamma Y + \delta) = \text{sign}(\alpha \gamma)\rho(X, Y).
$$

Hence, product moment correlation is invariant under linear transformations. If two variables are independent they have correlation zero, but the converse is not true.

Definition 2.2 (Rank correlation). The rank correlation of two random variables $X$ and $Y$ with cumulative distribution functions (cdfs) $F_X$ and $F_Y$ is

$$
r_{XY} = \rho_{F_X(X)F_Y(Y)}.
$$

(2.2)

In contrast to the product moment correlation, the rank correlation is invariant under non-linear strictly monotonic transformations. Both types of correlation yield a scalar measurement
for the dependence between two random variables. For certain distributions, e.g. normal and student t, rank correlations can be transformed to product moment correlations and vice-versa using *Pearson’s formula*:

\[ \rho = 2 \sin \left( \frac{\pi}{6} r \right) \]  

(2.3)

In this thesis conditional rank correlation are of special importance.

**Definition 2.3 (Conditional correlation).** The conditional correlation of \( Y \) and \( Z \) given \( X \)

\[ \rho_{YZ \mid X} = \frac{\text{E}(YZ \mid X) - \text{E}(Y \mid X)\text{E}(Z \mid X)}{\sqrt{\text{Var}(Y \mid X)\text{Var}(Z \mid X)}} \]

is the product moment correlation computed with the conditional distribution of \( Y \) and \( Z \) given \( X \).

A useful property of elliptical distributions is that partial correlations is equal to conditional correlation [Baba et al., 2004].

**Definition 2.4 (Partial correlation).** Consider variables \( X_i \) with zero mean and standard deviations \( \sigma_i = 1 \), \( i = 1, \ldots, n \) and let the numbers \( b_{12,3,\ldots,n}, \ldots, b_{1n;2,\ldots,n-1} \) minimize

\[ \text{E}[(X_1 - b_{12,3,\ldots,n} X_2 - \ldots - b_{1n;2,\ldots,n-1} X_n)^2]. \]

The partial correlation is defined as

\[ \rho_{12;3,\ldots,n} = \text{sgn}(b_{12,3,\ldots,n}) \sqrt{b_{12,3,\ldots,n} \cdot b_{21,3,\ldots,n}}. \]

The recursive formula (2.4) allows to calculate partial correlations from correlations.

\[ \rho_{12;3,\ldots,n} = \frac{\rho_{12,4\ldots,n} - \rho_{13;4\ldots,n} \cdot \rho_{23;4\ldots,n}}{\sqrt{(1 - \rho_{13;4\ldots,n}^2) \cdot (1 - \rho_{23;4\ldots,n}^2)}} \]  

(2.4)

This enables us, for elliptical distributions, to compute the correlation matrix from conditional rank correlations using recursion (2.4) and to sample joint distributions directly from the correlation matrix rather than following the more complicated algorithm for sampling so-called D-vines (cf. [Kurowicka and Cooke, 2006 section 6.4.2]). D-vines are treated in the section 2.3.

2.2 Copulas

In principle, a joint distribution is characterized by the marginal distributions of the individual variables as well as their dependence structure. The copula approach allows to isolate the description of the dependence structure from the marginals. [McNeil et al., 2010]
2.2. COPULAS

Definition 2.5 (Copula [McNeil et al., 2010]). A \(d\)-dimensional copula is a distribution function on \([0,1]^d\) with standard uniform marginal distributions.

The theoretical foundation for the application of copulas is laid by Sklar’s theorem:

Theorem 2.1 (Sklar 1959 [McNeil et al., 2010]). Let \(H_{1...d}\) be a \(d\)-dimensional joint distribution function with margins \(F_1, ..., F_d\). Then there exists a copula \(C: [0,1]^d \rightarrow [0,1]\) such that, for all \(x_1, ..., x_d\) in \(\mathbb{R} = [-\infty, \infty]\),

\[
H_{1...d}(x_1, ..., x_d) = C_{1...d}(F_1(x_1), ..., F_d(x_d)).
\] (2.5)

If the margins are continuous, then \(C\) is unique; otherwise \(C\) is uniquely determined on \(\text{Ran} F_1 \times \text{Ran} F_2 \times \cdots \times \text{Ran} F_d\), where \(\text{Ran} F_i = F_i(\mathbb{R})\) denotes the range of \(F_i\). Conversely, if \(C\) is a copula and \(F_1, ..., F_d\) are univariate distribution functions, the function \(H_{1...d}\) defined in (2.5) is a joint distribution function with margins \(F_1, ..., F_d\).

In fact, if an \(n\)-dimensional distribution function has continuous marginals \(F_1, ..., F_n\) the copula in (2.5) has the following expression

\[
C_{1...n}(u_1, ..., u_n) = H_{1...n}\{F_1^{-1}(u_1), ..., F_n^{-1}(u_n)\},
\] (2.6)

where \(F_i^{-1}(t) = \inf\{x \in \mathbb{R} \mid F_i(x) \geq t\}\) for all \(t\) in \([0,1]\) is the generalized inverse of \(F_i\).

Also a joint density can be represented with a uniquely defined \(n\)-variate copula density:

\[
f_{1...n}(x_1, ..., x_n) = c_{1...n}\{F_1(x_1), ..., F_n(x_n)\} \cdot f_1(x_1) \cdots f_n(x_n).
\] (2.7)

There is a variety of commonly used copulas, but for this thesis we restrict ourselves to the independent copula, the normal copula and to the t-copula. The independent copula is defined as

\[
C_{1...n}(u_1, ..., u_n) = u_1 \cdots u_n,
\]

which is equivalent to the fact that two random variables are independent if and only if \(F_{1...n}(X_1, ..., X_n) = F_1(X_1) \cdots F_n(X_n)\). Both the normal copula and the t-copula, which belong to the class of elliptical distributions, are described in an individual subsection.

2.2.1 The normal copula

The multivariate normal distribution of a \(d\)-dimensional vector \(X = [X_1, ..., X_d]\) is fully characterized by its mean vector, \(\mu = \text{E}[X]\), and covariance matrix, \(\text{Cov}(X) = \Sigma\). If \(\Sigma\) is positive-definite, the joint density is given by

\[
\phi_{1...d}(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)\}, \quad x \in \mathbb{R}^d,
\] (2.8)

where \(|\Sigma|\) denotes the determinant of \(\Sigma\).
Definition 2.6 (Normal copula). The multivariate normal copula is defined as
\[
C_{1\ldots n_{\Sigma}}(u_1, \ldots, u_n) = \Phi_{1\ldots n_{\Sigma}}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)), \quad u_1, \ldots, u_n \in [0, 1], 
\]  
(2.9)
where \( \Phi_{1\ldots n_{\Sigma}} \) is the multivariate normal distribution function with covariance matrix \( \Sigma \) and \( \Phi^{-1} \) is the inverse of the standard univariate distribution function.

Note that the product moment correlation, more precisely the covariance matrix, is a parameter of the copula. To sample the joint distribution we can use the following algorithm.

Algorithm 2.1 (Random variate generation from the normal copula \( C_{1\ldots n_{\Sigma}}^{Ga} \) [Embrechts et al., 2001]).

- Find the Cholesky decomposition \( A \) of \( \Sigma \).
- Simulate \( n \) independent random variates \( z_1, \ldots, z_n \) from \( N(0, 1) \).
- Set \( x = Az \).
- Set \( u_i = \Phi(x_i) \), \( i = 1, \ldots, n \).
- \( (u_1, \ldots, u_n)^T \sim C_{1\ldots n_{\Sigma}}^{Ga} \).

The family of normal copulas has the zero independence property, i.e. if the bivariate copula "has zero correlation, then it is the independent copula" [Kurowicka and Cooke, 2006, p.39], which is relevant for vines and nonparametric BBNs.

2.2.2 The t-copula

The multivariate student t distribution exists in many forms. We focus on the most common and natural form, the so-called canonical student t distribution. The \( p \)-dimensional random vector \( X = (X_1, \ldots, X_p) \) is said to have the \( p \)-variate t distribution with \( \nu \) degrees of freedom, mean vector \( \mu \) and positive definite matrix \( \Sigma \), which is denoted as \( X \sim t_p(\nu, \mu, \Sigma) \), if its joint probability distribution function (pdf) is given by [Kotz and Nadarajah, 2004]

\[
f_{1\ldots n_{\nu,\Sigma}}(t) = \frac{\Gamma\left(\frac{\nu + p}{2}\right)}{(\pi \nu)^{\frac{p}{2}} \Gamma\left(\frac{\nu}{2}\right) |\Sigma|^{\frac{1}{2}}} \left[ 1 + \frac{1}{\nu} (t - \mu)^T \Sigma^{-1} (t - \mu) \right]^{-\frac{\nu + p}{2}}. \tag{2.10}
\]

Note that \( \frac{\nu}{\nu - 2} \Sigma \) is the covariance matrix and is defined only if \( \nu > 2 \). If \( \nu \leq 2 \) then \( \text{Cov}(X) \) is not defined and \( \Sigma \) is interpreted as the shape parameter of the distribution of \( X \). If \( p = 1 \), \( \mu = 0 \) and \( \Sigma = 1 \) then (2.10) reduces to the univariate pdf

\[
f_{1\ldots n_{\nu}}(t) = \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\sqrt{\pi \nu} \Gamma\left(\frac{\nu}{2}\right)} \left[ 1 + \frac{t^2}{\nu} \right]^{-\frac{\nu + 1}{2}}. \tag{2.11}
\]

It is well known that \( X \) has the stochastic representation

\[
X = \mu + \frac{\sqrt{\nu}}{\sqrt{S}} Z, \tag{2.12}
\]
where $\mu \in \mathbb{R}^p$, $S \sim \chi_\nu^2$ and $Z \sim \mathcal{N}(0, \Sigma)$ are independent. At least two other representations exist (cf. [Kotz and Nadarajah, 2004]), but we concentrate on this one, because it enables a simple sampling procedure.

The multivariate t copula can be written as

$$C_{1,\ldots,n,\Sigma}^t = t_{i_{n,\Sigma}}^n(t_{i_{1,\Sigma}}^{-1}(u_1),...,t_{i_{n,\Sigma}}^{-1}(u_n)), \quad (2.13)$$

where we use the notation $t_{i_{n,\Sigma}}$ for the distribution function (df) of $\sqrt{\nu} \sqrt{S} \tilde{Z}$, where $\tilde{Z} \sim \mathcal{N}(0, \Sigma)$ and $\Sigma$ is the correlation matrix. Further, $t_{i_{\nu}}$ and $t_{i_{\nu}}^{-1}$ denote the df and the quantile function of a standard univariate $t_{\nu}$ distribution.

Unlike the normal copula, the t-copula does not enjoy the zero independence property. We can easily see that in the example of the bivariate t-distribution

$$f_{n,\rho}(t_1, t_2) = \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi \nu \sqrt{1 - \rho^2}} \left[1 + \frac{(t_1 - \mu_1)^2 + (t_2 - \mu_2)^2 - 2 \rho(t_1 - \mu_1)(t_2 - \mu_2)}{\nu(1 - \rho^2)}\right]^{-\frac{\nu+2}{2}}. \quad (2.14)$$

Taking $\mu = [0, 0]^T$ and $\rho = 0$ yields

$$f_{n,\rho}(t_1, t_2) = \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi \nu} \left[1 + \frac{1}{\nu} \left(t_1^2 + t_2^2\right)\right]^{-\frac{\nu+2}{2}}, \quad (2.15)$$

which is clearly not the product of two univariate t-distributions.

A simple sampling algorithm for random variates from the t-copula is based on representation (2.12).

**Algorithm 2.2** (Random variate generation from the t-copula $C_{1,\ldots,n,\Sigma}^t$ [Embrechts et al., 2001]).

- Find the Cholesky decomposition $A$ of $\Sigma$.
- Simulate $n$ independent random variates $z_1, \ldots, z_n$ from $\mathcal{N}(0, 1)$.
- Set $y = Az$.
- Simulate a random variate $s$ from $\chi_{\nu}^2$ independent of $z_1, \ldots, z_n$.
- Set $x = \frac{\sqrt{\nu}}{\sqrt{s}} y$.
- Set $u_i = t_{i_{\nu}}(x_i)$, $i = 1, \ldots, n$.
- $(u_1, \ldots, u_n)^T \sim C_{1,\ldots,n,\Sigma}^t$.

### 2.3 Vines

Vines are a graphical model introduced by [Bedford and Cooke, 2002]. They are a generalization of Markov trees and therefore fall into the category of undirected graphs.
Definition 2.7 (Vine, regular vine). \( \mathcal{V} \) is a vine on \( n \) elements, if

1. \( \mathcal{V} = (T_1, \ldots, T_{n-1}) \).
2. \( T_1 \) is a connected tree with nodes \( N_1 = \{1, \ldots, n\} \), and a set of edges denoted \( E_1 \); for \( i = 2, \ldots, n-1 \) \( T_i \) is a connected tree with nodes \( N_i = E_{i-1} \).

\( \mathcal{V} \) is a regular vine, if additionally

3. (proximity) For \( i = 2, \ldots, n-2 \), if \( \{a, b\} \in E_i \), then \( \#a \triangle b = 2 \), where \( \triangle \) denotes the symmetric difference. In other words, if \( a = \{a_1, a_2\} \) and \( b = \{b_1, b_2\} \), then exactly one of the \( a_i \) equals one of the \( b_i \).

The edges of the vine are associated with a constraint, conditioning and a conditioned set, which are defined as follows.

Definition 2.8 ([Cooke et al., 2011]).

1. For \( e \in E_i, \ i \leq n-1 \), the constraint set associated with \( e \) is the complete union \( U^*_e \) of \( e \), that is, the subset of \( 1, \ldots, n \) reachable from \( e \) by the membership relations.
2. For \( i = 1, \ldots, n-1 \), \( e \in E_i \), if \( e = \{j, k\} \) then the conditioning set associated with \( e \) is
   \[ D_e = U^*_j \cap U^*_k \]
   and the conditioned set associated with \( e \) is
   \[ \{C_{e,j}, C_{e,k}\} = \{U^*_j \setminus D_e, U^*_k \setminus D_e\} \].

Regular vines can be specified by associating bivariate and conditional bivariate copulas. The copulas can be chosen independent of each other. A bivariate copula vine specification uniquely determines a joint density in the form given by the following theorem:

Theorem 2.2 ([Bedford and Cooke, 2001b]). Let \( \mathcal{V} = (T_1, \ldots, T_{n-1}) \) be a regular vine on \( n \) elements. For an edge \( e \in E(\mathcal{V}) \) with conditioned elements \( e_1, e_2 \) and conditioning set \( D_e \), let the conditional copula and copula density be \( C_{e_1, e_2|D_e} \) and \( c_{e_1, e_2|D_e} \), respectively. Let the marginals distributions \( F_i \) with densities \( f_i, i = 1, \ldots, n \) be given. Then the vine-dependent distribution is uniquely determined, as has a density given by

\[
f_{1\ldots n} = f_1 \cdots f_n \prod_{e \in E(\mathcal{V})} c_{e_1, e_2|D_e} \{F_{e_1|D_e}, F_{e_2|D_e}\}. \tag{2.16}
\]

Two main families of vines are have been treated in the literature: the C-vine or canonical vine and the D-vine. For this thesis, only the family of D-vines is relevant and we will restrict our attention to this family.

Definition 2.9 (D-vine). A regular vine is called a D-vine if each node in \( T_1 \) has at most degree 2.
An example of a D-vine on 5 elements is presented in figure 2.1. The edges of the vine are associated with (conditional) rank correlations between $[-1, 1]$. They can be realized by a copula and a consistent joint distribution that satisfies the copula-vine specification can be constructed. If a copula with the zero independence property is used, a zero rank correlation assigned to the edge between two variables $X_1$ and $X_2$ means that they are independent. Alternatively, each edge can be associated with a partial correlation. In the case of the multivariate normal distribution partial correlations are equal to conditional product moment correlations, so the assignments are the same. This is called a complete partial correlation vine specification and is useful to construct a joint normal distribution as the following theorem shows.

**Theorem 2.3 (Bedford and Cooke, 2002).** Given any complete partial correlation vine specification for standard normal variables $X_1, \ldots, X_n$, there is a unique joint normal distribution for $X_1, \ldots, X_n$ satisfying all partial correlations.

A BBN is a directed acyclic graph (DAG) whose nodes represent random variables and whose arcs indicate a probabilistic influence from the parent node to the child node. It reflects all the possible "states" of the system in form of the joint probability. A simplistic BBN in which a variable $X$ influences a variable $Y$ is shown in figure 2.2. $X$ is called the parent or direct predecessor of the child or direct descendant $Y$. For this BBN, the joint distribution is specified by the marginal distribution of $X$ and the conditional distribution of $Y$ given $X$ as

$$f_{xy}(x, y) = f_x(x)f_{y|x}(y|x).$$

(2.17)
More generally, for \( n \) variables, the chain rule allows us to decompose the joint density using only conditional probabilities:

\[
f_{1...n}(x_1, ..., x_n) = f_1(x_1) \prod_{i=2}^{n} f_{i|1...i-1}(x_i|x_1...x_{i-1}).
\] (2.18)

This expression is usually not very compact, since it requires specifying values of an \( n \)-dimensional function. Under conditional independence assumptions however, the expression can be simplified. BBNs represent the joint probability more compactly by specifying a set of (conditional) independence assumptions in the form of a DAG. The graphical criterion that characterizes all independence assumptions is called \textit{d-separation} criterion.

**Definition 2.10** (d-separation [Hanea, 2008]). If \( A, B \) and \( C \) are three disjoint subsets of nodes in a BBN, then \( C \) \textit{d-separates} \( A \) from \( B \) if there is no path between a node in \( A \) and a node in \( B \) along which the following conditions hold:

- every node with converging arrows is in \( C \) or has descendants in \( C \) and
- every other node is outside \( C \).

From this criterion follows that each variable is conditionally independent of all non-descendants given its parents.

Figure 2.3 illustrates the (conditional) independence relations that can be induced by a BBN. In figure 2.3a \( X_1 \) and \( X_2 \) are predecessors of \( X_3 \). Since \( X_2 \) is direct predecessor, \( X_3 \) is independent of \( X_1 \) given \( X_2 \), i.e. \( X_1 \perp X_3 \mid X_2 \). In figure 2.3b \( X_2 \) is parent of \( X_1 \) and \( X_3 \). Again \( X_1 \perp X_3 \mid X_2 \). In figure 2.3c \( X_2 \) is the child of \( X_1 \) and \( X_3 \). In contrast to the previous two cases \( X_1 \perp X_3 \mid X_2 \) is not implied by the graph, but \( X_1 \perp X_3 \), because their parent set is empty.

\footnote{We say Z has converging arrows if on the acyclic path between X and Y two arrows point to Z.}

Figure 2.3: Examples of the possible conditional independence relations induced by a BBN
The conditional independence property allows us to simplify the conditional probability functions, so that we can write

\[ f_{i|1...i-1}(x_i|x_1...x_{i-1}) = f_{1...pa(i)}(x_i|x_{pa(i)}) \]  \hspace{1cm} (2.19)

for each \( i \), where \( pa(i) \) denotes the parent set for \( x_i \). If \( x_i \) does not have predecessors it is called a source node (a node without children is called a sink node) and \( f_{i|pa(i)}(x_i|x_{pa(i)}) = f_i(x_i) \) \cite{Hanea2008}. This yields

\[ f_{1...3}(x_1, x_2, ..., x_3) = f_1(x_1) \prod_{i=2}^{n} f_{i|pa(i)}(x_i|x_{pa(i)}). \]  \hspace{1cm} (2.20)

Note that we only have to specify functions with dimension not greater than the maximal number of parents of any node.

The main use of BBNs is updating: Once new evidence on one or more variables is obtained, its effect can be propagated through the network using Bayes’ theorem. Evidence can be propagated both, forward and backward, which allows for predictive as well as diagnostic reasoning. Predictive reasoning updates the probability for a child node, based on new evidence of a parent node. The other way around, diagnostic reasoning updates the probability of a parent node, based on new evidence for a child node.

BBNs exist in various forms. Well known are e.g. discrete BBNs, discrete-normal continuous BBNs and nonparametric continuous BBNs. The following sections describe the first and the last type; the first, because it is the most popular type of BBNs and the last because it is used for this thesis. For discrete-normal BNNs the reader is referred to \cite[section 5.2.2]{KurowickaCooke2006}.

### 2.4.1 Discrete BBNs

Discrete BBNs consist of discrete random variables. Each variable can be in one of a number of different states. The states should be exclusive (it is not possible for more than state at once to hold) and exhaustive (all possible states are specified) \cite{BedfordCooke2001a}. If we want to model continuous factors, they have to be discretized. The joint distribution is specified by the marginal distributions of the source nodes and the conditional distributions of all child nodes in the form of a conditional probability table (CPT). These distributions can be obtained from data, if available, or alternatively, from expert judgment.

As stated before, the main use of BBNs is to update distributions after observations. If some variables have been observed, we want to infer the probability distributions of the other variables, which have not yet been observed. Updating is based on Bayes Theorem and quite complex. An algorithm which allows fast updating for large BBNs has been proposed by \cite{LauritzenSpiegelhalter1988}. Commercially available software that does all the calculations are, for example, Netica or Hugin.

If we consider the BBN from figure 2.4 to have discrete random variables the marginal distributions of \( X_1 \), \( X_2 \) and \( X_3 \) and the conditional distribution of \( X_4 \) given \( X_1 \), \( X_2 \) and
X₃ have to be specified. Table 2.5 shows the CPT for variable X₄ assuming each node can take k values, denoted xᵢ^j, i = 1, ..., 4, j = 1, ..., k. The resultant CPT contains k⁴ probability entries. For four binary variables, e.g. with values yes and no, 16 values need to be specified. For four variables with 10 states each 10000 values need to be found. This exponentially increasing assessment burden is one of the weakest points of discrete BBNs. If a model contains random variables of continuous nature, these have to be discretized. Considering the very high assessment burden, only a very limited number of discrete values can be allowed or the model has to be simplified drastically in order to make the quantification of the values feasible. A second serious shortcoming is the maintainability of discrete BBNs. They are very inflexible with respect to changes in modeling. For instance, if one parent node is added, all the previous quantification for the children of this node have to be redone.

Such disadvantages motivate the use of continuous BBN. A promising type of continuous BBN, the non-parametric continuous BBN, is described in the following section.

![Figure 2.4: Parents and child of a BBN on 4 variables (source: Hanea, 2008)](image)

| X₁ | X₂ | X₃ | P(X₄ = x₁^1 | X₁, X₂, X₃) | P(X₄ = x₁^2 | X₁, X₂, X₃) | ... | P(X₄ = x₁^k | X₁, X₂, X₃) |
|----|----|----|----------------|----------------|----|----------------|
| x₁ | x₂ | x₃ | ? | ? | ... | ? |
| x₁ | x₂ | x₃ | ? | ? | ... | ? |
| ... | ... | ... | ... | ... | ... | ... |
| x₁ | x₂ | x₃ | ? | ? | ... | ? |

Figure 2.5: Conditional Probability Table for X₄ (source: Hanea, 2008)

### 2.4.2 Non-parametric continuous BBNs

Shortcomings with regard to discretization, maintainability and flexibility give reasons to use non-parametric continuous Bayesian belief nets (NPBBNs) to model risk. They have been introduced by [Kurowicka and Cooke, 2005] as an approach to continuous BBNs using vines together with copulae. In a NPBBN nodes are associated with continuous univariate random variables and arcs with (conditional) rank correlations realized by a chosen copula. Any copula may be chosen, as long as it has invertible conditional distribution functions and possesses the zero independence property. Theorem 2.4 shows that these assignments together with the graph structure uniquely determine the joint distribution.

[Kurowicka and Cooke, 2005] present a protocol according to which the (conditional) rank
2.4. BAYESIAN BELIEF NETS

![Figure 2.6: BBN with one child node and six parent nodes](diagram)

Correlations should be assigned to the arcs. The protocol is illustrated using the BBN in figure 2.6, which consists of one child node and seven parent nodes:

1. A sampling order is constructed, such that all predecessors of node \( i \) appear before \( i \) in the ordering and the variables are indexed according to the sampling order \( 1, \ldots, 7 \). This has already been done for the BBN in figure 2.6. Of course, the sampling order is generally not unique.

2. The joint distribution is factorized following the sampling order

\[
\begin{align*}
&f_1(1) \cdot f_{2|1}(2 \mid 1) \cdot f_{3|21}(3 \mid 21) \cdot f_{4|312}(4 \mid 321) \cdot f_{5|4321}(5 \mid 4321) \cdot f_{6|54321}(6 \mid 54321) \\
&\quad \cdot f_{7|654321}(7 \mid 654321)
\end{align*}
\] (2.21)

3. Those variables that are not parents of the conditioned variable are underlined.

\[
\begin{align*}
&f_1(1) \cdot f_{2|1}(2 \mid 1) \cdot f_{3|21}(3 \mid 21) \cdot f_{4|312}(4 \mid 321) \cdot f_{5|4321}(5 \mid 4321) \cdot f_{6|54321}(6 \mid 54321) \\
&\quad \cdot f_{7|654321}(7 \mid 654321)
\end{align*}
\] (2.21)

According to (2.20) these variables are not necessary in sampling the conditioned variable and can be omitted, which yields

\[
\begin{align*}
&f_1(1) \cdot f_2(2) \cdot f_3(3) \cdot f_4(4) \cdot f_5(5) \cdot f_6(6) \cdot f_{7|654321}(7 \mid 654321)
\end{align*}
\] (2.22)

In general for \( n \) random variables \( X_1, \ldots, X_n \), this so-called characteristic factorization is given by

\[
f_1(X_1) \prod_{i=2}^{n} f_{i|pa(i)}(X_i \mid X_{pa(i)})
\] (2.23)

4. For each term \( i \) with parents \( i_1, \ldots, i_{p(i)} \) in (2.22), the arc \( i_{p(i)−k} \rightarrow i \) is associated with the conditional rank correlation

\[
\begin{align*}
&\begin{cases}
  r(i, i_{p(i)}) , & k = 0 \\
  r(i, i_{p(i)−k} \mid i_{p(i)} , \ldots, i_{p(i)−k+1}) , & 1 \leq k \leq p(i) − 1
\end{cases}
\end{align*}
\] (2.24)

For this example the rank correlations that need to be assigned to the arcs are thus \( r_{17} \), \( r_{27|1} \), \( r_{37|12} \), \( r_{47|123} \), \( r_{57|1234} \) and \( r_{67|12345} \).
The following theorem shows that these assignments are algebraically independent and together with the graph structure, the margins and a copula with the zero independence property uniquely determine the joint distribution. The proof of this main result for nonparametric BBNs is based on the theory of vines.

**Theorem 2.4** ([Hanea et al., 2006]). Given

1. a directed acyclic graph with \( n \) nodes specifying conditional independence relationships in a BBN,
2. \( n \) variables, assigned to the nodes, with invertible distribution functions,
3. the specification (2.24), \( i = 1, \ldots, n \), of conditional rank correlations on the arcs of the BBN,
4. a copula realizing all correlations \([-1,1]\) for which zero correlation entails independence;

the joint distribution of the \( n \) variables is uniquely determined. This joint distribution satisfies the characteristic factorization (2.23) and the conditional rank correlations in (2.24) are algebraically independent.

The proof by [Hanea et al., 2006] is given below, because we shall come back to it in the following chapter and shall be looking into it more thoroughly.

**Proof.** Given that all univariate distributions are known, continuous, invertible functions, one can use them to transform each variable to uniform on \((0,1)\). Hence, we can assume, without any loss of generality, that all univariate distributions are uniform distributions on \((0,1)\).

The first term in (2.24) is assigned vacuously. We assume the joint distribution for \( \{1, \ldots, i - 1\} \) has been determined. The \( i \)th term of the factorization (2.21) involves \( i - 1 \) conditional variables, of which \( \{i_{p(i)} + 1, \ldots, i - 1\} \) are conditionally independent of \( i \) given \( \{i_1, \ldots, i_{p(i)}\} \). We assign

\[
r(i, i_j | i_1, \ldots, i_{p(i)}) = 0, \quad i_{p(i)} < i_j \leq i - 1
\]

(2.25)

Then the conditional rank correlations (2.24) and (2.25) are exactly those on a D-vine with i-variables, \( \mathcal{D} \) involving variable \( i \). The other conditional bivariate distributions on \( \mathcal{D} \) are already determined. It follows that the distribution on \( 1, \ldots, i \) is uniquely determined. Since zero conditional rank correlation implies conditional independence,

\[
f_{1 \ldots i}(1, \ldots, i) = f_{i_{[1 \ldots i-1]}(i | 1, \ldots, i - 1)} \cdot f_{1 \ldots i-1}(1, \ldots, i - 1) = f_{i_{[1 \ldots i_{p(i)}]}(i | i_1, \ldots, i_{p(i)})} \cdot f_{1 \ldots i-1}(1, \ldots, i - 1)
\]

from which it follows that the factorization (2.23) holds. \( \square \)

The last condition implicates that not all copulas can be used to realize the joint distribution, e.g. not the t-copula. However, there is doubt among the authors whether this condition
is needed; especially in the light of the recently introduced "pair copula constructions for non-Gaussian DAG models" \cite{Bauer2012}, a very similar approach that does not require the zero independence property. To address this doubt the following chapter contains a small simulation study by means of which the consequences of using the t-copula instead of the normal copula to realize the joint distribution are investigated.

The joint distribution is obtained through sampling making use of the theory of D-vines. Sampling of D-vines is described in e.g. \cite[section 6.4]{KurowickaCooke2005}. A freely available software for nonparametric BBNs based on the normal copula is Uninet \footnote{available from http://www.lighttwist.net/wp/uninet}.

The following sampling protocol for the normal copula-vine approach has been proposed by \cite{Hanea2006}. Let $X_1, ..., X_n$ be random variables of a BBN with continuous, invertible distribution functions $F_1, ..., F_n$ and with rank correlations according to (2.24).

1. $X_1, ..., X_n$ are transformed to standard normal variables $Y_1, ..., Y_n$ using $Y_i = \Phi^{-1}(F_i(X_i))$, $i = 1, ..., n$, where $\Phi$ denotes the cumulative distribution function of the standard normal distribution.

2. The vine for the standard normal variables $Y_1, ..., Y_n$ is constructed using the (conditional) rank correlations that are assigned to the BBN (see figure 2.7). They also correspond to the edges of this vine, because the transformations $Y_i = \Phi^{-1}(F_i(X_i))$ are strictly increasing.

3. The rank correlations are transformed to product moment correlations using Pearson’s formula (2.3). This yields a complete partial correlation vine specification for $Y_1, ..., Y_n$, since conditional and partial correlation are equal for normal variables. According to theorem 2.3 there is a unique joint distribution for $Y_1, ..., Y_n$ satisfying all partial correlations.

4. The correlation matrix $R$ is computed using the recursive formula (2.4).

5. The joint normal distribution with correlation matrix $R$ is sampled.

6. For each sample the standard normal margins are transformed back to the original distributions using $F_i^{-1}(\Phi(y'_i))$, where $y'_i$ is the $j$th sample from the previous step.

The main advantage of this approach are that the assigned (conditional) rank correlations need not be revised if the marginal distribution of one node is change. Furthermore, if a node is added or removed, then the previously assessed (conditional) rank correlation need not be reassessed. In general, the assessment burden is much lower than for discrete BBNs.

If new evidence becomes available, the joint distribution can be updated by conditioning on the variables for which the new information has been obtained. With the described normal-vine-copula approach updating can be performed analytically on the joint normal vine, because the conditional distribution of one normal vector given another normal vector will also be normal \cite[Proposition 3.13]{EatonEaton1983}. The conditional distribution of the original
variable can be found by transforming back using its inverse distribution function, as the following proposition shows.

**Proposition 2.1** (Hanea et al., 2006). Let $X_1$ and $X_2$ be random variables with continuous, invertible distribution functions $F_1$ and $F_2$. Let $Y_1$ and $Y_2$ be the transformation of $X_1$ and $X_2$ to standard normal variables. Then the conditional distribution $X_1 \mid X_2$ can be calculated as $F_1^{-1}(\Phi(Y_1 \mid Y_2))$, where $\Phi$ is the cumulative distribution function of the standard normal distribution.

The theory on nonparametric BBNs presented here can be extended to arbitrary copulas, as long as they possess the zero independence property. The consequences of using a copula for which zero correlation does not imply independence are not quite known as yet. As mentioned before, a paper by Bauer et al., 2012 suggests that, at least, an akin modeling approach can be based on arbitrary copulas. To encourage a scientific discussion on this subject the following chapter contains a small simulation study comparing the outcome of the normal copula and the t-copula for the same BBN.

Subsequently, the second part of the thesis demonstrates how a BBN can be quantified in the absence of data by eliciting experts.
3. The Zero Independence Property and Nonparametric BBNs

Recently [Bauer et al., 2012] introduced so-called "pair copula constructions for non-Gaussian DAG models". Their main theorem reformulates, and possibly advances, the vine-based approach for nonparametric BBNs introduced in section 2.4 using graph theoretical considerations only. This new approach is not limited to copula families with the zero independence property, while this property is a key assumption in theorem 2.4 the main result for nonparametric BBNs.

Because most of the commonly used copulas possess the property (see table 3.1), the assumption is not too restrictive: only the popular t-copula and the elliptical copula do not possess it [Kurowicka and Cooke, 2006, Gatz, 2007]. Nonetheless, there is doubt within the Risk Analysis group if this is a necessary assumption for theorem 2.4. For this reason this chapter investigates by means of a small simulation study the consequences of using the t-copula to sample the joint distribution of a nonparametric BBN. In particular the effect on conditional independence (CI) relationships is examined. To this end the joint distribution of a demonstration BBN is sampled with both the normal and the t-copula and tested for CI relationships.

The first section introduces the demonstration BBN and explains which CI statements can be represented by the graph structure and which follow or do not follow from the quantification when using the normal copula or t-copula, respectively. The second section describes a kernel-based CI test and the third sections discusses the simulation study.

\footnote{as introduced by [Kurowicka et al., 2000]. Note that this is another usage of the name "elliptical" and does not relate to the class of elliptical copulas.}
Table 3.1: Overview of common copulas and the zero independence property (for more details see e.g. [Schmidt, 2006], [Kurowicka and Cooke, 2006], [Gatz, 2007])

<table>
<thead>
<tr>
<th>Copula Type</th>
<th>Zero Independence Property</th>
<th>Relation between partial and conditional correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elliptical copulas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal copula</td>
<td>yes</td>
<td>$\rho_{X,Y</td>
</tr>
<tr>
<td>t-copula</td>
<td>no</td>
<td>$\rho_{X,Y</td>
</tr>
<tr>
<td>Archimedean copulas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frank copula</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Gumbel copula</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Clayton copula</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Other copulas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagonal band copula</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Elliptical copula</td>
<td>no</td>
<td>$\rho_{X,Y</td>
</tr>
<tr>
<td>Minimum information copula</td>
<td>yes</td>
<td></td>
</tr>
</tbody>
</table>

### 3.1 Representation of Independence: Graph Structure versus Quantification

Nonparametric BBNs use vines and copulas that represent (conditional) independence as zero (conditional) rank correlation. Because vines are fully connected, (conditional) dependences and independences can only be represented through quantification. Edges of a regular vine can be associated with a (conditional) rank correlation. If the copula realizing the correlation has the zero independence property, independence of two variables can be represented by an edge between them with a zero rank correlation. For instance, for the vine in figure 3.1 $X_1 \perp X_2$. Of course, independence will always yield a zero correlation for the corresponding edge. But if the used copula does not have the zero independence property, the vine specification with rank correlations does not give any information on whether the variables are independent or just uncorrelated.

![Figure 3.1: D-vine representing the independence of $X_1$ and $X_2$? (source: Hanea, 2010)](image)

Differently, DAGs represent (conditional) dependences and independences through their structure. Consider the simple BBN in figure 3.2a as an example. This DAG postulates the following relations: $X_1 \perp X_2$, $X_1 \perp X_3$ and $X_2 \perp X_3$. Considering that all arcs are quantified with nonzero correlation we also have that $X_1 \not\perp X_4$, $X_2 \not\perp X_4$ and $X_3 \not\perp X_4$ as well as $X_1 \not\perp X_2 \mid X_4$, $X_1 \not\perp X_3 \mid X_4$ and $X_2 \not\perp X_3 \mid X_4$. 
3.1. REPRESENTATION OF INDEPENDENCE: GRAPH STRUCTURE VERSUS QUANTIFICATION

When using a copula with the *zero independence property* additional (conditional) independence statements are obtained by assigning correlation zero to an arc. In other words, there are (conditional) independence statements implied by the graphical structure and (conditional) independence statements implied by the specification. For the BBN depicted in figure 3.2b, depending on the ordering of the variables, the arc between $X_3$ and $X_4$ can be associated with $r_{34}$, $r_{34|1}$ or $r_{34|12}$ and, thus, quantifying it as zero correlation can imply either $X_3 \perp X_4$, $X_3 \perp X_4 | X_1$, $X_3 \perp X_1 | X_2$ or $X_3 \perp X_1 | X_1, X_2$. This is different to nonexistence of this arc; the BBN in figure 3.2c postulates all four (conditional) independence statements.

![Figure 3.2: Different graph structures and specifications leading to the same CI statements for normal and t-copula](image)

For the normal copula it follows for the graphical structure in figure 3.2b and from recursion (2.4) that

$$r_{34} = 0 \iff r_{34|1} = 0 \iff r_{34|2} = 0 \iff r_{34|12} = 0.$$  (3.1)

In this case either specification of the arc between $X_3$ and $X_4$ implies all four (conditional) independence statements and the graph in figure 3.2b is equivalent to the one in figure 3.2c.

The fact that a BBN with a zero correlation arc is equivalent to a BBN without that particular arc is an attribute of this particular example and not true in general. Consider for instance the graphs from figure 3.2b and figure 3.2c but with an extra arc between $X_1$ and $X_3$ as displayed in figure 3.3a and figure 3.3b:

- The BBN in figure 3.3a implies that $X_1 \perp X_2$, $X_2 \perp X_3$, $X_2 \perp X_3 | X_1$, $X_3 \perp X_4 | X_1$ and $X_3 \perp X_4 | X_1, X_2$.

- If the arc between $X_3$ and $X_4$ of the BBN in figure 3.3b is specified as $r_{34} = 0$, it does **not** follow that $r_{34|1} = 0$ and neither that $r_{34|12} = 0$. Hence this BBN would imply $X_1 \perp X_2$, $X_2 \perp X_3$, $X_2 \perp X_3 | X_1$ and $X_3 \perp X_4$.

If instead the arc between $X_3$ and $X_4$ is quantified as $r_{34|1} = 0$, $r_{34} \neq 0$ and also $r_{34|12} \neq 0$. Hence, this BBN implies only that $X_1 \perp X_2$, $X_2 \perp X_3$, $X_2 \perp X_3 | X_1$ and $X_3 \perp X_4 | X_1$.

In either case the independence statements differ from the ones postulated by the graph structure in figure 3.2c.

Returning to the original example, the joint distribution of the BBN in figure 3.2b is identical to the one in figure 3.2c for the normal copula. It is specified by the marginal
THE ZERO INDEPENDENCE PROPERTY AND NONPARAMETRIC BBNS

(a) Graph structure missing an arc
(b) Graph structure with an arc assigned correlation zero

Figure 3.3: Different graph structures and specifications leading to different CI statements for normal and t-copula distributions and by the covariance matrix

\[
R_{Ga} = \begin{pmatrix}
1 & 0 & 0 & \rho_{14} \\
0 & 1 & 0 & \rho_{24} \\
0 & 0 & 1 & 0 \\
\rho_{14} & \rho_{24} & 0 & 1
\end{pmatrix}
\]

(3.2)

and can easily be sampled with the algorithm given in section 2.2.

If, instead, we were to use the t copula to sample a BBN, extra (conditional) independence statements cannot be implied by the specification, but only by the structure of the DAG, because it lacks the zero independence property. More specifically, \( X_3 \not\perp X_4, X_3 \not\perp X_4 \mid X_1, X_3 \not\perp X_4 \mid X_2 \) or \( X_3 \not\perp X_4 \mid X_1, X_2 \) for the BBN in figure 3.2b no matter how the arc between \( X_3 \) and \( X_4 \) is specified. However, because partial correlation equals conditional correlation also for the t-copula

\[
r_{34} = 0 \iff r_{34|1} = 0 \iff r_{34|2} = 0 \iff r_{34|12} = 0
\]

(3.3)
yielding

\[
R^t = \begin{pmatrix}
1 & 0 & 0 & \rho_{14} \\
0 & 1 & 0 & \rho_{24} \\
0 & 0 & 1 & 0 \\
\rho_{14} & \rho_{24} & 0 & 1
\end{pmatrix}
\]

(3.4)

In contrast, the BBN in figure 3.2c implies, also for the t-copula, \( X_3 \perp X_4, X_3 \perp X_4 \mid X_1, X_3 \perp X_4 \mid X_2 \) or \( X_3 \perp X_4 \mid X_1, X_2 \). Naturally (3.3) and correlation matrix (3.4) follow. Hence, the joint distribution obtained with the t-copula for the graph in figure 3.2b is equal to the one for the graph in figure 3.2c, albeit they represent different (conditional) independence statements. In other words, the correlation matrix gives no information on whether two variables are uncorrelated or independent. This information can only be represented by the DAG.

Since the t-copula does not possess the zero independence property, it is not obvious whether the joint distribution will exhibit (conditional) independence relations for the zero
3.2 A Kernel-Based Conditional Independence Test

Testing for conditional independence of continuous variables is much more difficult than testing for unconditional dependence. Traditional methods often make the simplifying assumption that the variables are linearly dependent, in which case \( X \perp Y \mid Z \) reduces to testing zero conditional correlation, \( \rho_{XY \mid Z} \), or zero partial correlation, \( \rho_{XY ; Z} \). For instance, the Fisher Z test for conditional independence of the Bayes Net Toolbox (BNT) for Matlab\(^2\) makes this assumption. However, for this simulation study, this kind of assumption would defeat the purpose of the test, since the point is to investigate the effect of a copula has on the CI statements of the BBN, if zero correlation does not imply independence.

A so-called kernel-based conditional independence test (KCI-test) that does not make this assumption and moreover does not require an explicit estimation of any joint or conditional densities or discretization of the variables has recently been proposed by Zhang et al., 2012.\(^3\) The Matlab source code for this test has been provided by the authors\(^3\).

As the name suggests the test statistic for this CI test is calculated from the kernel matrices associated with \( X, Y \) and \( Z \). The basic idea is that CI can be characterized by uncorrelatedness between functions in certain spaces. In that sense, this characterization can be regarded as a generalization of the partial correlation based characterization of CI for Gaussian variables. The test statistic is then defined in terms of kernel matrices under the condition of uncorrelatedness in such spaces. The following paragraphs introduce the main aspects of the KCI-test, for more detailed theory and proofs we refer to the original paper.

In nonparametric statistics a kernel is defined as any smooth function \( K \) such that \( K(x) \geq 0 \) and

\[
\int K(x)dx = 1, \quad \int xK(x)dx = 0 \quad \text{and} \quad \sigma_K^2 = \int x^2K(x)dx > 0 \tag{3.5}
\]

[Wasserman, 2006]. Let \( K_X \) be the kernel matrix of the sample \( x = (x_1, ..., x_1, ..., x_n) \). By default, the CI test uses the Gaussian kernel, i.e. the \((i,j)\)th entry of \( K_X \) is

\[
k(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||^2}{2\sigma_X^2}\right),
\]

where \( \sigma_X^2 \) is the kernel width. The centralized kernel matrix corresponding to \( K_X \) is \( \tilde{K}_X = HK_XH \), where \( H = I - \frac{1}{n}11^T \) with \( I \) being the \( n \times n \) identity matrix and \( 1 \) being the vector of 1’s. Furthermore, suppose \( \tilde{K}_X = V_X\Lambda_XV_X^T \) is the eigenvalue decomposition (EVD) of the centralized kernel matrix, where \( \Lambda_X \) is the diagonal matrix that contains the non-negative eigenvalues \( \lambda_{x,i} \) sorted in descending order, i.e. \( \lambda_{x,1} \geq \lambda_{x,2} \geq \ldots \geq \lambda_{x,n} \geq 0 \). Ana-

\(^2\)available at https://code.google.com/p/bnt/

\(^3\)available at http://people.tuebingen.mpg.de/kzhang/KCI-test.zip
logue notations are used for \( Y \) and \( Z \). Finally, define \( \psi_x = [\psi_1(x), ..., \psi_n(x)] \triangleq V_x \Lambda_x^{\frac{1}{2}} \) and \( \phi_y = [\phi_1(y), ..., \phi_n(y)] \triangleq V_y \Lambda_y^{\frac{1}{2}} \).

A test statistic for the unconditional independence test is given by the following theorem and provides the foundation for the KCI-test.

**Theorem 3.1 (Independence test).** Under the null hypothesis that \( X \) and \( Y \) are statistically independent, the statistic

\[
T_{UI} \triangleq \frac{1}{n} \text{Tr}(\bar{K}_X \bar{K}_Y)
\]

has the same asymptotic distribution as

\[
\bar{T}_{UI} \triangleq \frac{1}{n} \sum_{i,j=1}^{n} \lambda_{x,i} \lambda_{y,j} z_{ik}^2,
\]

i.e., \( T_{UI} \overset{d}{\rightarrow} \bar{T}_{UI} \) as \( n \rightarrow \infty \).

The unconditional independence test is performed with the following steps:

1. Calculate \( \bar{K}_X \) and \( \bar{K}_Y \) as well as their eigenvalues and eigenvectors
2. Evaluate the statistic \( T_{UI} \) according to equation (3.6).
3. Generate the null distribution \( \bar{T}_{UI} \) with Monte Carlo simulation according to (3.7).

For the KCI-test we need to introduce a few more variables. First, denote \( \bar{X} \triangleq (X, Y) \), the corresponding centralized kernel matrix as \( \bar{K}_X \) and \( \varphi_x = [\varphi_1(\bar{x}), ..., \varphi_n(\bar{x})] \triangleq V_x \Lambda_x^{\frac{1}{2}} \).

The centralized kernel matrix for \( \bar{X} \vert Z \) can be calculated as \( \bar{K}_{\bar{X}|Z} = \bar{\sigma}_x \bar{\sigma}_y^T = R_Z \bar{K}_{\bar{X}} R_Z \) and the one for \( Y \vert Z \) as \( \bar{K}_{\bar{Y}|Z} = R_Z \bar{K}_Y R_Z \), where \( R_Z = (\bar{K}_Z + \epsilon I)^{-1} \) and \( \epsilon \) is a small positive regularization parameter. Again, define \( \psi_{\bar{x}|z} = [\psi_{\bar{x}|z,1}(\bar{x}), ..., \psi_{\bar{x}|z,n}(\bar{x})] \triangleq V_{\bar{x}|z} \Lambda_{\bar{x}|z}^{\frac{1}{2}} \) and \( \phi_{\bar{y}|z} = [\psi_{\bar{y}|z,1}(\bar{y}), ..., \psi_{\bar{y}|z,n}(\bar{y})] \triangleq V_{\bar{y}|z} \Lambda_{\bar{y}|z}^{\frac{1}{2}} \) based on the EVDs \( \bar{K}_{\bar{X}|Z} = V_{\bar{x}|z} \Lambda_{\bar{x}|z} V_{\bar{x}|z}^T \) and \( \bar{K}_{\bar{Y}|Z} = V_{\bar{y}|z} \Lambda_{\bar{y}|z} V_{\bar{y}|z}^T \).

The following proposition provides the test statistic of the KCI-test.

**Proposition 3.1 (Conditional independence test).** Under the null hypothesis that \( X \) and \( Y \) are conditionally independent given \( Z \), the statistic

\[
T_{CI} \triangleq \frac{1}{n} \text{Tr}(\bar{K}_{\bar{X}|Z} \bar{K}_{\bar{Y}|Z})
\]

has the same distribution asymptotic distribution as

\[
\bar{T}_{CI} \triangleq \frac{1}{n} \sum_{k=1}^{n} \hat{\lambda}_k \cdot z_k^2,
\]

where \( \hat{\lambda}_k \) are eigenvalues of \( \bar{w} \bar{w}^T \) and \( \bar{w} = [\bar{w}_1, ..., \bar{w}_n] \), with the vector \( \bar{w}_t \) obtained by stacking \( M_t = [\psi_{\bar{x}|z,1}(\bar{x}_t), ..., \psi_{\bar{x}|z,n}(\bar{x}_t)]^T \cdot [\phi_{\bar{y}|z,1}(\bar{y}_t), ..., \phi_{\bar{y}|z,n}(\bar{y}_t)] \), and \( z_k^2 \) are i.i.d. \( \chi_1^2 \) samples.
Performing the KCI-test involves the following steps:

1. Calculate $\tilde{K}_{X|Z}$ and $\tilde{K}_{Y|Z}$ as well as their eigenvalues and eigenvectors.
2. Evaluate the statistic $T_{CI}$ according to equation (3.8).
3. Generate the null distribution $\tilde{T}_{CI}$ with Monte Carlo simulation according to (3.9).

An alternative to the Monte Carlo simulation in both tests is to approximate the null distribution with a Gamma distribution. For more details on the corresponding parameters as well as for the (optimal) choice of hyperparameters, $\sigma$ and $\epsilon$, we refer again to the original paper by Zhang et al., 2012.

### 3.3 Simulation Study

Because the simulation study investigates whether a copula without the zero independence property might be suitable to approximate the joint distribution of the variables in the graph, we sample two joint distributions for the same graph specification: the BBN in figure 3.2b. As a matter of fact it could also be the BBN in figure 3.2c since the correlation matrices are identical for both copulas. The first sampling distribution is generated by the normal copula and the second by the t-copula with three degrees of freedom, i.e. $\nu = 3$, according to algorithm 2.1 and algorithm 2.2. The sample size is $n = 1000$.

The arcs are specified as $r_{14} = 0.5$, $r_{24|1} = 0.5$ and $r_{34|12} = 0$. Since all parents are independent $r_{12} = 0$, $r_{13} = 0$ and $r_{23} = 0$. The corresponding product moment correlation matrix for both copulas is

$$
\Sigma = \begin{pmatrix}
1 & 0 & 0 & 0.5176 \\
0 & 1 & 0 & 0.4429 \\
0 & 0 & 1 & 0 \\
0.5176 & 0.4420 & 0 & 1
\end{pmatrix}. \tag{3.10}
$$

Since the first three steps of both algorithms are identical, they only need to be performed once. This also means that the two joint distributions are computed from the same sample $z$ of i.i.d. standard normal variables and ensures that differences in the joint distributions can be attributed to the two copulas and are not caused by the initial sample. Of course, the outcome of $s$, the $\chi^2_{3}$ distributed variable in the algorithm for the t-copula, has some influence.

In total 100 independent simulations have been performed. For each of the simulations, the empirical correlation matrices differ slightly. For example, an empirical correlation matrix for the joint distribution obtained with the normal copula is

$$
\Sigma_{C_{Ga}^{R}}^{n} = \begin{pmatrix}
1 & -0.0463 & 0.0489 & 0.5053 \\
-0.0463 & 1 & -0.0363 & 0.3713 \\
0.0489 & -0.0363 & 1 & -0.0245 \\
0.5053 & 0.3713 & -0.0245 & 1
\end{pmatrix}. \tag{3.11}
$$
and the corresponding one for the t-copula is

\[
\Sigma_{C_t^{3,R}} = \begin{pmatrix}
1 & -0.0481 & 0.0493 & 0.5079 \\
-0.0481 & 1 & -0.0368 & 0.3730 \\
0.0493 & -0.0368 & 1 & -0.0247 \\
0.5079 & 0.3730 & -0.0247 & 1
\end{pmatrix}.
\] (3.12)

It can be noted, that the empirical correlation matrices differ more from the original one [3.10] than they differ from each other. As a matter of fact the differences between \( R_{C_G^a} \) and \( R_{C_t^{3,R}} \) are rather negligible. This is reasonable, as the t-distribution as well known as an approximation to the normal distribution.

The KCI-test and its unconditional equivalent is performed to test for a number of (conditional) independence statements, which are listed in the first column of table 3.2. Not all simulation runs exhibit the same KCI-test results. The p-values vary remarkably and sometimes not all independence relations of the graph are recognized. However, the results are almost identical for normal and t-copula and usually if an independence statement is not recognized, it is neither recognized for the sampling distribution of the normal copula nor for the one of the t-distribution (cf. appendix B for a couple of examples). This suggests, that the cause lies within the KCI-test itself. The average p-values of 100 tests are given in the second and third column for the normal and t-copula, respectively. The fourth and firth column indicate whether the CI statement from the first column is supported by data or not.

The p-values of the normal copula and the t-copula are very similar, which is not surprising considering that \( R_{C_G^a} \) and \( R_{C_t^{3,R}} \) are almost the same for each simulation.

Table 3.2: Average p-values for the KCI-test computed from 100 tests

<table>
<thead>
<tr>
<th>( H_0 )</th>
<th>( p )-value</th>
<th>( C_G^a )</th>
<th>( C_t )</th>
<th>( H_0 ) accepted?</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_1 \perp X_2 )</td>
<td>0.4549 0.4703</td>
<td>yes yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( X_1 \perp X_3 )</td>
<td>0.4488 0.4547</td>
<td>yes yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( X_2 \perp X_3 )</td>
<td>0.5293 0.5221</td>
<td>yes yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( X_1 \perp X_4 )</td>
<td>0.4732 0.4958</td>
<td>yes yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( X_2 \perp X_4 )</td>
<td>0.5293 0.5221</td>
<td>yes yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( X_3 \perp X_4</td>
<td>X_1 )</td>
<td>0.4785 0.4923</td>
<td>yes yes</td>
<td></td>
</tr>
<tr>
<td>( X_3 \perp X_4</td>
<td>X_2 )</td>
<td>0.4659 0.4896</td>
<td>yes yes</td>
<td></td>
</tr>
<tr>
<td>( X_3 \perp X_4</td>
<td>X_1,X_2 )</td>
<td>0.4333 0.4537</td>
<td>yes yes</td>
<td></td>
</tr>
<tr>
<td>( X_1 \perp X_2</td>
<td>X_4 )</td>
<td>0.5170 0.5164</td>
<td>yes yes</td>
<td></td>
</tr>
<tr>
<td>( X_1 \perp X_3</td>
<td>X_4 )</td>
<td>0.4542 0.4535</td>
<td>yes yes</td>
<td></td>
</tr>
<tr>
<td>( X_2 \perp X_3</td>
<td>X_4 )</td>
<td>0.4785 0.4923</td>
<td>yes yes</td>
<td></td>
</tr>
</tbody>
</table>

On average the KCI-test indicates on a significance level of 0.05 for both copulas that

- the parent nodes are independent as postulated by the graph: \( X_1 \perp X_2 \), \( X_1 \perp X_3 \) and \( X_2 \perp X_3 \).
• parents and child are dependent of each other, if connected with a nonzero arc and
independent otherwise: \(X_1 \not\perp X_4, \quad X_2 \not\perp X_4\) and \(X_3 \perp X_4\).

• specifying the arc between \(X_3\) and \(X_4\) as \(r_{34|12} = 0\) implies independence of the parent
and the child node as well as conditional independence given the other parent nodes:
\(X_3 \perp X_4, \quad X_3 \perp X_4 | X_1, \quad X_3 \perp X_4 | X_2\) and \(X_3 \perp X_4 | X_1, X_2\).

• parents are conditionally dependent of each other given the child, if connected with a zero
arc, and otherwise they are conditionally independent:
\(X_1 \not\perp X_2 | X_4, \quad X_1 \perp X_3 | X_4\) and \(X_2 \perp X_3 | X_4\).

For the normal copula the results have been expected. For the t-copula they are not
as obvious, since zero (conditional) correlation generally does not imply (conditional) independence. The reason all zero correlated variables are independent lies within the sampling
procedure (see algorithm 2.2). By applying \(y = Az\), where \(A\) is the Cholesky decomposition
of the product moment correlation matrix \(\Sigma\), we can find a linear combination of standard normal variables that realizes \(\Sigma\). For these standard normal variables zero correlation implies independence. The student t-sample is then obtained by setting \(x = \frac{\sqrt{s}}{\sqrt{n}}y\), where \(s\) is a random variate from \(\chi^2\) independent of \(z_1, ..., z_n\). In this operation no dependence is introduced to the sample and also for t-copula all zero correlation variables are independent. Hence, the t-copula will not violate any CI statements imposed by the graph and represent the same CI statements as the normal copula.

More generally, this means that a sampling algorithm solely based on the correlation matrix
cannot construct a joint distribution of uncorrelated random variables that is distinguished
from the one for independent variables. Actually, in sampling practice the t-copula exhibits
the zero independence property. So far there is no apparent reason why it should not be used
to sample nonparametric BBNs. A more theoretical approach to the question whether the t-copula can be used in nonparametric BBNs is given in the next section.

### 3.4 Extension of the nonparametric BBN Theorem

An essential idea of BBN model is that the initial factorization of the joint density, for instance
for the BBN in figure 3.2a
\[
f_1(x_1) \cdot f_{2|1}(x_2 | x_1) \cdot f_{3|21}(x_3 | x_2, x_1) \cdot f_{4|312}(x_4 | x_3, x_2, x_1)
\]
(3.13)
can be simplified to
\[
f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_{4|21}(x_4 | x_2, x_1)
\]
(3.14)
by taking advantage of the conditional independence statements implied by the graph structure. According to theorem 2.4, a joint distribution that satisfies this so-called characteristic factorization can be sampled if the following (conditional) rank correlations are associated with each arc \(i_{p(i)−k} \rightarrow i\) for each node \(i\) with parents \(i_1, ..., i_{p(i)}\).

\[
\begin{cases}
r(i, i_{p(i)}), & k = 0 \\
r(i, i_{p(i)−k} | i_{p(i)}, \ldots, i_{p(i)−k+1}), & 1 \leq k \leq p(i) − 1
\end{cases}
\]
(3.15)
Because the theorem suggests to specify the conditional independences also as zero correlations it demands that a copula, for which zero correlation implies independence is used to realized the rank correlations. In the previous section we have given an example using the t-copula, which does not possess this zero independence property, but for which the sampled distribution exhibits all the conditional independence statements implied by the graph. As explained above the reason lies within the sampling procedure of the t-copula.

While using this copula might work well in practice, it does not have any theoretical foundation. Therefore, this section proposes a slight modification of theorem 2.4 as to allow for various types of copulas. If we would not specify the conditional independence statements as zero rank correlation, but with the (conditional) independent copula, then the (conditional) rank correlations associated with the arcs could be realized by any copula, as long as it is able to realize all correlations $[-1, 1]$. For instance, the BBN in the previous section would be quantified with a mixture of the (conditional) independent copula and the t-copula. This approach has been used by [Aas et al., 2009, Bauer et al., 2012, Bauer and Czado, 2012] in their so-called pair-copula constructions.

The following example illustrates the idea. The joint density of three random variables $X_1$, $X_2$ and $X_3$ can be factorized as

$$f_{123}(x_1, x_2, x_3) = f_1(x_1) \cdot f_{2\mid 1}(x_2 \mid x_1) \cdot f_{3\mid 21}(x_3 \mid x_2, x_1).$$  \hspace{1cm} (3.16)

For the BBN in figure 3.4 this reduces to

$$f_1(x_1) \cdot f_{2\mid 1}(x_2 \mid x_1) \cdot f_{3\mid 1}(x_3 \mid x_1),$$  \hspace{1cm} (3.17)

because the directed acyclic graph postulates $X_3 \perp X_2 \mid X_1$.  

![Figure 3.4: Example BBN on three nodes](image)

We know that a bivariate density can be written as in terms of bivariate copula densities and marginals

$$f_{12}(x_1, x_2) = c_{12}\{F_1(x_1), F_2(x_2)\} \cdot f_1(x_1) \cdot f_2(x_2)$$  \hspace{1cm} (3.18)

and it follows easily that

$$f_{2\mid 1}(x_2 \mid x_1) = c_{12}\{F_1(x_1), F_2(x_2)\} \cdot f_2(x_2).$$  \hspace{1cm} (3.19)

Similarly, for three random variables, a possible decomposition is

$$f_{3\mid 21}(x_3 \mid x_2, x_1) = c_{23\mid 1}\{F_{2\mid 1}(x_2 \mid x_1), F_{3\mid 1}(x_3 \mid x_1)\} \cdot f_{3\mid 1}(x_3 \mid x_1)$$

$$= c_{23\mid 1}\{F_{2\mid 1}(x_2 \mid x_1), F_{3\mid 1}(x_3 \mid x_1)\} \cdot c_{13}\{F_1(x_1), F_3(x_3)\} \cdot f_3(x_3).$$  \hspace{1cm} (3.20)
Hence, the joint density can be written as
\[ f_{123}(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{12}\{F_1(x_1), F_2(x_2)\} \cdot c_{13}\{F_1(x_1), F_3(x_3)\} \cdot c_{23|1}\{F_2(x_2 | x_1), F_3(x_3 | x_1)\}. \] (3.21)

In fact, this decomposition corresponds to the bivariate vine-copula specification of the vine in figure 3.5, which is exactly the vine underlying the nonparametric BBN in figure 3.4. Now, \( X_2 \perp X_3 | X_1 \) if and only if \( c_{23|1}\{F_2(x_2 | x_1), F_3(x_3 | x_1)\} = 1 \), i.e. they are joint by the conditional independent copula. This simplifies (3.21) to
\[ f_{123}(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{12}\{F_1(x_1), F_2(x_2)\} \cdot c_{13}\{F_1(x_1), F_3(x_3)\} \] (3.22)

and can be written in terms of (conditional) densities:
\[ f_1(x_1) \cdot f_{2|1}(x_2 | x_1) \cdot f_{3|1}(x_3 | x_1). \] (3.23)

We have obtained the desired characteristic factorization (3.17). No assumptions have been made for \( c_{12} \) and \( c_{13} \); they could be any copula. The conditional independence statement has been realized by the conditional independent copula.

The following theorem is a slight modification of theorem 2.4 and demands that the (conditional) independent copula is used to realize (conditional) independence statements implied by the graph. On the other hand any copula can be used to realize the (conditional) correlations assigned to the arcs, as long as it can realize all correlations in \([-1, 1]\).

**Theorem 3.2.** Given

1. a directed acyclic graph with \( n \) nodes specifying conditional independence relationships in a BBN,
2. \( n \) variables \( X_1, ..., X_n \), assigned to the nodes, with invertible distribution functions \( F_1, ..., F_n \),
3. the specification (2.24), \( i = 1, ..., n \), of (conditional) rank correlations on the arcs of the BBN,
4. any copula realizing all correlations \([-1, 1]\),
5. the (conditional) independent copula realizing all (conditional) independence relationships specified by the graph structure of the BBN;

...
the joint distribution of the \( n \) variables is uniquely determined. This joint distribution satisfies the characteristic factorization (2.23) and the conditional rank correlations in (2.24) are algebraically independent.

**Proof.** Given that all univariate distributions are known, continuous, invertible functions, one can use them to transform each variable to uniform on \((0,1)\). Hence, we can assume, without any loss of generality, that all univariate distributions are uniform distributions on \((0,1)\).

The first term in (2.24) is assigned vacuously. We assume the joint distribution for \( \{1,\ldots,i-1\} \) has been determined. The \( \text{th} \) term of the factorization (2.21) involves \( i-1 \) conditional variables, of which \( \{i_{p(i)+1},\ldots,i_{i-1}\} \) are conditionally independent of \( i \) given \( \{i_1,\ldots,i_{p(i)}\} \). We assign

\[
c_{i,i_{p(i)}|i_1,\ldots,i_{p(i)}}(X_i \mid X_{i_1},\ldots,X_{i_{p(i)}}) = 1
\]

for \( i_{p(i)} < j_i \leq i-1 \). Then these copulas together with the ones realizing the conditional rank correlations (2.24) are exactly those on a D-vine with \( i \) variables, \( \mathcal{D}^i \), involving variable \( i \). The other conditional bivariate distributions on \( \mathcal{D}^i \) are already defined. It follows from theorem 2.2 that the distribution on \( \{1,\ldots,i\} \) is uniquely determined. Since the conditional independent copula is used to represent all conditional independence statements of the graph,

\[
f_{1\ldots i}(1,\ldots,i) = f_{i_1\ldots i-1}(1,\ldots,i-1) \cdot f_{1\ldots i-1}(1,\ldots,i-1)
\]

from which it follows that the factorization (2.23) holds. \( \Box \)

The only difference of this proof compared to the one presented in chapter 2 is the use of the (conditional) independent copula (3.24) instead of zero (conditional) rank correlations (2.25).

Applying this theorem to the above example, we associate the rank correlations \( r_{12} \) and \( r_{13} \) with the two arcs. We denote the bivariate copula density of \( X_1 \) and \( X_2 \) realizing \( r_{12} \) as \( c_{r_{12}}\{F_1(X_1),F_2(X_2)\} \) and the one of \( X_1 \) and \( X_3 \) realizing \( r_{13} \) as \( c_{r_{13}}\{F_1(X_1),F_3(X_3)\} \).

Because \( X_2 \perp X_3 \mid X_1 \), we set \( c_{r_{23}}\{F_{21}(X_2 \mid X_1),F_{31}(X_3 \mid X_1)\} = 1 \). Of course, these are the copulas of the vine in figure 3.5 and the unique vine-dependent joint density is, according to theorem 2.2, given by

\[
f_{123}(x_1,x_2,x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{r_{12}}\{F_1(x_1),F_2(x_2)\} \cdot c_{r_{13}}\{F_1(x_1),F_2(x_3)\}.
\]

This is the same characteristic factorization that we obtained through the decomposition of the joint density in bivariate and conditional bivariate copulas. Note that there are no restrictions on the type of copula for \( c_{r_{12}}\{F_1(x_1),F_2(x_2)\} \) and \( c_{r_{13}}\{F_1(x_1),F_2(x_3)\} \) as long as they can realize all rank correlations in \([-1,1]\).

Of course this reformulated theorem still accommodates the use of a copula with the zero independence property for all rank correlations and (conditional) independence statements, since such a copula reduces to the independent copula for correlation zero. For the normal
3.5. Conclusion

Bayesian belief nets are a tool to simplify a joint density function by representing it as a directed acyclic graph (DAG) which implies certain conditional independence statements. Needless to say, the sampled density has to exhibit these conditional independence relations.

Up to now two different theorems have been available stating that the joint density of the DAG is uniquely determined by its marginal distributions and its (conditional) bivariate copulas. While the proof of [Bauer and Czado, 2012] relies on graph theoretic considerations only, the proof of [Hanea et al., 2006] is based on the specification of the BBN. The first theorem states that the margins and copulas can be freely chosen, the latter proposes a protocol on how to specify the arcs: as conditional rank correlations and realize them with a copula for which zero (conditional) correlation implies (conditional) independence. However, knowing that based on graph theory any copula can be used, this chapter has investigated, if and how the t-copula can be used to sample the joint distribution while following the specification guidelines of theorem 2.4.

Two findings have emerged. First, in the simulation study the sampling algorithm 2.2 of the t-copula has generated (conditional) independence relations whenever the correlation was zero. It is doubtful whether it is actually possible to sample a joint t-distribution in which variables are dependent, but uncorrelated. This means that in practice using the t-copula to realize all rank correlations, despite its not possessing of the zero independence property, should not be problematic. Second, a minor modification of theorem 2.4, namely realizing all (conditional) independence statements with the (conditional) independent copula, enables us to use any copula to realize the rank correlations. The specification protocol for the rank correlations on the arcs remains the same. If this modification would be implemented in a software like Uninet, the software user will be able to specify marginals and rank correlation as before. Additionally he/she may choose which copula should be used for the rank correlations on the arcs. He/she will not have to think about the fact that the conditional independence statements are not based on correlation zero, but on the (conditional) independence copula. This change will be "hidden" in the underlying code.

At last remark should be made on the KCI-test by [Zhang et al., 2012], which has been used to test the sample distributions for (conditional) independence. The p-values for the various null hypothesis, that is the independence statements, have varied strongly from simulation run to simulation. On a $\alpha = 5\%$ level, quite a few independence statements have not been
"recognized". However, it seemed the test gives a p-value equal to zero, if the variables are not (conditionally) independent and a p-value greater than zero, although sometimes very close to zero, otherwise. In order to obtain reliable results, also for other applications, it may be necessary to repeat the test for a high number of simulations.

The second part of the thesis, which is dedicated to quantifying a nonparametric BBNs based on expert opinion, restricts itself to the normal copula and follows the original formulation of the theorem, that is theorem 2.4.
Part II

Quantifying a HPM with Structured Expert Judgment
4. Introduction to the Operator Model

Starting point of this case study is a human performance model for an oil and gas company, the operator model described in the introduction. Its parameters have been elicited using structured expert judgment. The Safety Science group of the Delft University of Technology has defined the structure of the model and decided on the number of variables included as well as their operational definitions. The group has also conducted the expert judgment exercise and has started to analyze the obtained data and to quantify the model. However, due to time constraints for the project the dependence estimates of the various experts have not been combined using the theory presented in the previous chapter, but by directly averaging correlation matrices. In this second part of the thesis, we redo the combination of experts in a mathematically sound manner. A new model with the same weights, but the proper combination method, is built and compared to the old model. Additionally, the consequences of other choices for weights on the model output are investigated.

The model depicted in figure 4.1 is a sub-model of a holistic risk model for the oil and gas industry that incorporates human performance and management influences in addition to technical failures. It aims to represent the influences on human performance in an intuitive manner and to show which management actions should be taken in order to decrease the likelihood of human errors. For example, the model could predict, that improving procedures may be more effective in reducing human errors than improving man-machine interfaces.

The factors included in the model are deemed the most relevant contributors to human performance in this application. They have been defined based on literature and discussions with managers. The definitions of the factors are given in table 4.1. It can be noted that all variables are defined as likelihood of something. For example, $V_7$ denotes the contractor performance. $V_7 = 1$ means that a task is performed with at least one or more significant errors, while $V_7 = 0$ means that a task is performed correctly. Hence, the larger the value for $V_7$, the higher the likelihood of an error occurring, the lower the contractor performance. The fact, that high values are associated with low performance may be confusing and care should be taken when interpreting the model results. This is similar for all variables.

The objective of the following chapters is to analyze the data obtained from the expert

---

\footnote{This table can also be found in the elicitation protocol in appendix A, with one big difference: the variables are described as the opposite, e.g. \textit{without any errors} instead of \textit{with one or more errors}. However, it appears that the variables have actually been used in the way presented here (cf. the table below question Q8 of the elicitation protocol in appendix A).}
judgment exercise, to critically review how it has been used in order to build a nonparametric BBN and to propose improvements. The next chapter introduces the concepts of structured expert judgment. It will become clear that eliciting a joint distribution from experts involves several tasks which can be executed sequentially. These tasks are described one by one in the chapters thereafter and the previous efforts are directly contrasted with suggestions for improvement. Chapter 6 describes how the marginal distributions and rank correlation are computed from the experts statements. Additionally it discusses possible choices of weights. Chapter 7 concerns itself with the pooling of experts to a decision maker and presents the resulting joint distributions as a nonparametric Bayesian belief net. Thereafter, chapter 8 questions if the presumed model structure actually corresponds to the experts’ beliefs and chapter 9 proposes steps towards a validation of the model.
<table>
<thead>
<tr>
<th>Factor</th>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contractor performance</td>
<td>V7</td>
<td>Likelihood that a task is performed <strong>with one or more errors</strong> which in the end hamper the operation</td>
</tr>
<tr>
<td>Competence</td>
<td>V1</td>
<td>Likelihood that the contractor have <strong>insufficient</strong> knowledge/skill/experience to execute a particular task</td>
</tr>
<tr>
<td>Job specific communica-</td>
<td>V2</td>
<td>Likelihood that during permit to work/shift/on the job communication, the required information (job hazard, safety precaution) is <strong>not</strong> transferred to the contractors</td>
</tr>
<tr>
<td>Commitment</td>
<td>V3</td>
<td>Likelihood that during critical activities the individual <strong>does not</strong> carry out the required behavior at the right moment, with the right care and attention to detail in order to control the risk</td>
</tr>
<tr>
<td>Quality of procedure</td>
<td>V4</td>
<td>The likelihood that work instruction is <strong>not</strong> available, accurate, safe, clear, up-to date and easy to read and understand</td>
</tr>
<tr>
<td>Human factors in de-</td>
<td>V5</td>
<td>Likelihood that equipment and work stations are <strong>not</strong> designed for easy access and easy operation and maintenance</td>
</tr>
<tr>
<td>Capacity to work</td>
<td>V6</td>
<td>Likelihood that the contractor shows the sign of <strong>in-availability</strong> to work (including fitness to work, alertness, physical state, and psychological state)</td>
</tr>
</tbody>
</table>
5. Structured Expert Judgment

Structured expert judgment is a recognized scientific method to quantify the uncertainty of an unknown parameter in the absence of data. A group of experts from the field of the desired parameter is selected and each expert gives his opinion for each uncertain quantity in form of a subjective probability. The experts’ different assessments of each quantity are combined to a single probability of the so-called decision maker through weighted averaging. Generally, weights based on performance rules yield better estimates for the unknown quantities than equal weights for each expert. Furthermore, the method can be extended to elicit dependencies between variables.

The method is explained with the example of the generic nonparametric BBN in figure 5.1 which has the same structure as the operator model introduced in the previous chapter. Hence, the exact same steps as described here are taken to quantify the operator model. Seven marginal distributions have to be elicited as well as one unconditional rank correlation, $r_{17}$, and five conditional rank correlations, $r_{27|1}$, $r_{37|12}$, $r_{47|123}$, $r_{57|1234}$ and $r_{67|12345}$. Since the normal copula is assumed, it is also possible to elicit the rank correlation matrix (5.1) and to compute the conditional rank correlations using Pearson’s transform (2.3) and the recursion for partial correlations (2.4).

![Figure 5.1: BBN with one child node and seven parent nodes](image-url)
5.1 The Classical Model for Eliciting Marginal Distributions

The classical model proposed by [Cooke, 1991] is a performance based weighted average model to elicit and combine marginal distributions and has been widely discussed in the literature [Cooke, 1991, Cooke and Goossens, 2008]. It is a practical tool based on first principles, not a mathematical theory. This means that the resulting mathematical model is not unique, but influenced by ad hoc choices of the analyst. The classical model constructs the weights for each expert based on strictly proper scoring rules that reward high calibration and high information. *(Strictly) proper* means that "a subject receives his maximal expected score if (and only if) his stated assessment corresponds to his true opinion" [Cooke, 1991, p. 136]. Calibration and information score are computed from a number of so-called seed (or calibration) variables whose realizations are known to the analyst, but not to the experts. Both scores can be used to elicit discrete as well as continuous variables, but for many applications (practically) continuous variables are more natural to work with and are therefore used in the majority of studies [Hanea, 2011]. To assess continuous uncertain quantities \(X_1, \ldots, X_N\) experts give pre-defined quantiles \(x_{q_1}, \ldots, x_{q_R}\) such that \(q_0 := 0 < q_1 < \ldots < q_R \leq 1 := q_{R+1}\) of their subjective probability distributions. Typically \(q_1 = 5\%, q_2 = 50\%\) and \(q_3 = 95\%\) are chosen.

The freely available software Excalibur\(^1\) can be used to construct marginal distribution from the quantiles, to compute weights and to calculate the decision maker distribution.

The **relative information score** \(I(e)\) represents the degree to which a distribution is concentrated. The wider bounds an expert states, the more uncertainty he includes and the less informative he is. Informativeness can only be measured with respect to a background measure, which is chosen by the analyst. Commonly, and in this thesis, the uniform background measure is used. To compute the score a cumulative minimal information distribution \(F_{jei}\) is associated with each quantile assessment \(x_{q,j,ei}\) of expert \(e_i, i = 1, \ldots, E\), for variable \(j, j = 1, \ldots, N\), such that \(F_{jei}(x_{q,j,ei}) = q_r\). For the uniform background measure the associated density \(f_{jei}\) is constant between quantiles and the mass between the quantiles agrees with \(p_r = q_r - q_{r-1}\) for \(r = 1, \ldots, R + 1\), or equivalently, with the inter-quantile vector

\[
p = (p_1, p_2, p_3, p_4) = (0.05, 0.45, 0.45, 0.05)
\]

when using the 5%--, 50%- and 95%-quantile. Further, the uniform background measure requires an intrinsic range which can be created by the \(k\%\) overshoot rule. Typically, a simple 10%-overshoot is used above and below the smallest interval containing the assessed quantiles.

\[\Sigma_{rank} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & r_{17} \\
0 & 1 & 0 & 0 & 0 & r_{27} \\
0 & 0 & 1 & 0 & 0 & r_{37} \\
0 & 0 & 0 & 1 & 0 & r_{47} \\
0 & 0 & 0 & 0 & 1 & r_{57} \\
r_{17} & r_{27} & r_{37} & r_{47} & r_{57} & 1
\end{pmatrix}
\]

\[ (5.1) \]

\[^1\text{available from http://risk2.ewi.tudelft.nl/oursoftware/6-excalibur} \]
of all experts as well as the realization (if known). Hence, the cut-off points are global for all experts. Let $x_{j1}$ denote the smallest 5% quantile stated by any expert and $x_{jR}$ denote the largest 95% quantile stated by any expert. Moreover, $x_j$ denotes the realization of random variable $X_j$ and, if unknown, it is assumed that $x_{j1} \leq x_j \leq x_{jR}$. Then the two cut-off points of intrinsic range $[x_{j0}, x_{jR}+1]$ for variable $X_j$ can be computed by:

$$x_{j0} = \min(x_{j1}, x_j) - k \frac{\max(x_{jR}, x_j) - \min(x_{j1}, x_j)}{100},$$

and

$$x_{jR+1} = \max(x_{jR}, x_j) + k \frac{\max(x_{jR}, x_j) - \min(x_{j1}, x_j)}{100}.$$

Finally, the relative information for expert $e_i$ is defined as

$$I(e_i) := \frac{1}{N} \sum_{j=1}^{N} \left[ \ln(x_{jR+1} - x_{j0}) + \sum_{r=1}^{R+1} p_r \ln \frac{p_r}{x_{q_r,je_i} - x_{q_{r-1}e_i}} \right].$$

This score is proportional to the relative informativeness of the expert’s joint distribution for all assessed variables under the assumption those variables are independent. An expert reaches a high information score if he chooses the quantiles very close to each other and the score is not influenced by whether or not an expert has captured the realization with his assessment. [Hanea, 2011; Cooke, 1991]

The calibration score $C(e_i)$ for each expert measures how accurate his/her probability statements can be considered based on the realizations of the seed questions. If the true values for the seed variables fall, in the majority of the cases, outside 90% confidence bands given by the expert, it is sensible to assume that the expert either defines his bands to narrow and/or that he grossly dislocates them. Low scores, close to zero, indicate that the expert’s statements are unlikely to be correct. For each unknown variable, the experts divide its range into inter-quantile intervals with known probability values corresponding to entries of vector $p$ from equation (5.2). The calibration score for expert $e_i$ is the p-value of the null hypothesis $H_0 = \"the inter-quantile interval \{given by expert e\} containing the true value for each variable is drawn independently from probability vector p\" [Hanea, 2011 p.77] and can be computed with a $\chi^2$ test. From the realizations $x_1,...,x_N$ the sample distribution of the inter-quantile intervals can be computed for each expert. Let $n_r(e_i)$ be the number of variables expert $e_i$ assigned to the interval $I_r = [q_r, q_{r-1}]$, then $s_r(e_i) = \frac{n_r}{N}$ is the sample probability mass of $I_r$. When using the 5%- 50%- and 95%-quantile the sample distribution of expert $e_i$ is

$$s(e) = (s_1, s_2, s_3, s_4) = \left(\frac{n_1}{N}, \frac{n_2}{N}, \frac{n_3}{N}, \frac{n_4}{N}\right).$$

Because

$$P\{2NI(s(e_i), p) \leq x\} \to \chi^2_{n-1} \quad \text{as } N \to \infty,$$

where the relative information of $s$ with respect to $p$ of expert $e_i$ is defined as

$$I(s(e_i), p) := \sum_{j=1}^{n} s_j(e_i) \ln \frac{s_j(e_i)}{p_j},$$

(5.8)
the null hypothesis can easily be tested using the test statistic $2NI(s(e_i), p)$. The calibration score for expert $e_i$, or p-value for $H_0$, is

$$C(e_i) = 1 - \chi^2_R[2NI(s(e_i), p)].$$

As for the information score, all variables are assumed to be independent. [Hanea, 2011][Cooke, 1991]

A good expert has high calibration and high informativeness. Calibration, being a much faster function, dominates over informativeness. Thus, informativeness acts as modulator between more or less equally calibrated experts. The weights $w_{e_i}$ are globally\footnote{For item weights cf. [Cooke, 1991]} determined for each expert $e_i$ and satisfy $w_{e_i} \geq 0$ and $\sum w_{e_i} = 1$ assuming that $w_{e_i} > 0$ for at least one expert. The unnormalized weights are computed as a product of calibration score, information score and an indicator function

$$w^*_{e_i} = C(e_i) \cdot I(e_i) \cdot I_\alpha(C(e_i)),$$

(5.10)

where $I_\alpha(x) = 0$ if $x < \alpha$ and $I_\alpha(x) = 1$ otherwise. The indicator function ensures that the score is strictly proper, but there is no constraint for the value of $\alpha$. Because, the choice of $\alpha$ influences the resulting distribution of the decision-maker, it is chosen such that it maximizes the decision maker’s virtual weight. The normalized weights are given by

$$w_{e_i} = \frac{w^*_{e_i}}{\sum_{i=1}^E w^*_{e_i}} \quad \text{assuming} \quad \sum_{i=1}^E w^*_{e_i} > 0.$$

(5.11)

Finally, the marginal distribution of the decision maker for variable $j$ is computed by

$$f_{\alpha,DM}(j) = \sum_{i=1}^E w_{\alpha(e_i)} * f_{e_i,j}.$$  

(5.12)

5.2 Eliciting Dependencies

Differently than the elicitation of marginal distributions, the elicitation of dependencies is known less well and still under development [Morales Nápoles et al., 2013]. Two different methods to quantify (conditional) rank correlations through structured expert judgment can be found in the literature: (1) Asking the expert for probabilistic statements, typically conditional probabilities and (2) asking the expert for ratios of rank correlations. For the operator model, which is the subject-matter of our case study, the experts have been elicited prior to this thesis according to the latter approach. This so-called direct approach is described below following [Morales Nápoles et al., 2008][Morales Nápoles, 2010].
5.2. Eliciting Dependencies

5.2.1 Assessing Ratios of Rank Correlations

For the direct approach, experts rank the parent nodes according to the highest absolute rank correlation to the child node $X_7$. The parent nodes are denoted as $X_1,\ldots,X_6$ such that $|r_{17}| \geq |r_{27}| \geq |r_{37}| \geq |r_{47}| \geq |r_{57}| \geq |r_{67}|$. Generally this ranking differs for all experts.

To quantify the highest rank correlation, $r_{17}$, the experts assess the conditional probability statement of equation (5.13). This equation reads: "Suppose that variable $X_1$ was observed to be above its median. What is the probability that $X_7$ is above its median as well?"

$$R_{1i}^e = P\{X_7 \geq x_{r_{i,0.50}}^{e_i} \mid X_1 \geq x_{r_{1,0.50}}^{e_i}\} \rightarrow r_{1,7}^{e_i} \quad (5.13)$$

$$R_{2i}^e = \frac{r_{27}^{e_i}}{r_{17}^{e_i}} \rightarrow r_{27,1}^{e_i} \quad (5.14)$$

$$R_{3i}^e = \frac{r_{37}^{e_i}}{r_{17}^{e_i}} \rightarrow r_{37,12}^{e_i} \quad (5.15)$$

$$R_{4i}^e = \frac{r_{47}^{e_i}}{r_{17}^{e_i}} \rightarrow r_{47,123}^{e_i} \quad (5.16)$$

$$R_{5i}^e = \frac{r_{57}^{e_i}}{r_{17}^{e_i}} \rightarrow r_{57,1234}^{e_i} \quad (5.17)$$

$$R_{6i}^e = \frac{r_{67}^{e_i}}{r_{17}^{e_i}} \rightarrow r_{67,12345}^{e_i} \quad (5.18)$$

Assuming that $X_1$ and $X_7$ can be joint by the normal copula, which is the copula used later on to built a BBN in Uninet, a relation between the expert’s conditional probability statement $P_{1i}^{e_i}$ and the product moment correlation $\rho_{17}^{e_i}$ can be found:

$$R_{1i}^e = P\{X_7 \geq x_{r_{i,0.50}}^{e_i} \mid X_1 \geq x_{r_{1,0.50}}^{e_i}\}$$

$$= P\{F_{X_7}^{e_i}(X_7) \geq \frac{1}{2} \mid F_{X_1}^{e_i}(X_1) \geq \frac{1}{2}\}$$

$$= P\{F_{X_7}^{e_i}(X_7) \geq \frac{1}{2}, F_{X_1}^{e_i}(X_1) \geq \frac{1}{2}\} \frac{P\{F_{X_1}^{e_i}(X_1) \geq \frac{1}{2}\}}{P\{F_{X_1}^{e_i}(X_1) \geq \frac{1}{2}\}}$$

$$= \int_0^\infty \int_0^\infty \phi_{\Sigma_1^{e_i}}(\tilde{x}_1, \tilde{x}_7)d\tilde{x}_1 d\tilde{x}_7$$

$$= 2 \cdot \int_0^\infty \int_0^\infty \phi_{\Sigma_1^{e_i}}(\tilde{x}_1)d\tilde{x}_1 d\tilde{x}_7, \quad (5.19)$$

\[3\] In principle other copulas can be used as well, e.g. Frank’s copula was used by [Morales Nápoles, 2010]. However, sampling may be needed to find a relation between the conditional probability statement and the rank correlation.
where $\Sigma_1^i = 1$ is the variance of the standard normal distribution and $\Sigma_{17} = \begin{pmatrix} 1 & \rho_{17}^i \\ \rho_{17}^i & 1 \end{pmatrix}$ is the covariance matrix of the bivariate normal distribution of $X_1$ and $X_7$. The relationship between $P\{X_7 \geq x_{7, q_{50}}^i \mid X_1 \geq x_{1, q_{50}}^i \}$ and $\rho_{17}$ under the normal copula assumption is plotted in figure 5.2. Note that higher absolute values of $\rho_{17}$ cause larger differences in $P\{X_7 \geq x_{7, q_{50}}^i \mid X_1 \geq x_{1, q_{50}}^i \}$ than lower absolute values. To obtain $\rho_{17}$ Pearson’s transformation (2.3) is applied to $\rho_{17}$.

The remaining conditional rank correlations are assessed through the ratios (5.14) - (5.18). All assessments depend on the previous answers of the expert and the graph structure for which the rank correlations are assessed. For example, $r_{2,7|1}$ is computed in the following way. First $r_{1,7}$ and $r_{2,7}$, obtained from $P_{1}^{i}$ (5.13) and $P_{2}^{i}$ (5.14), are transformed to $\rho_{17}$ and $\rho_{27}$ using Pearson’s formula (2.3). Then the recursive formula 2.4 can be used to calculate $\rho_{27|1}$, because conditional and partial correlations are equal under the normal copula assumption. Since the graph structure (cf. figure 5.1) stipulates that $X_1$ and $X_2$ are independent, i.e. $\rho_{12} = 0$, we have

$$\rho_{27|1} = \frac{\rho_{27}^i - \rho_{12} \cdot \rho_{17}^i}{\sqrt{1 - (\rho_{12}^i)^2} \cdot (1 - (\rho_{17}^i)^2)} \quad (5.20)$$

and $r_{2,7|1}$ follows again from Pearson’s formula 2.3.

However, the ratios that each expert can provide are bounded by his previous estimates. Considering that the parents for this BBN are independent, $X_2$ can only explain the part of $X_7$ that has not yet been explained by $X_1$, $X_3$ can only explain the part of $X_7$ that has not yet been explained by $X_1$ and $X_2$, and so on. The bounds can be computed using the fact that all conditional rank correlations associated with the arcs of a BBN are algebraically independent and, therefore, can take any number in $(-1, 1)$. For the BBN in figure 5.1 $r_{17}$,
5.2. ELICITING DEPENDENCIES

Figure 5.3: Possible values or \( r_{2,7}^{e_1} \) given that \( r_{2,7|1} \) can take any value in \((-1, 1)\) and the expert stated \( P_{e_1} = 0.8 \), which results in \( r_{1,7}^{e_1} = 0.7953 \)

\( r_{27|1}, r_{37|12}, r_{47|123}, r_{57|1234} \) and \( r_{67|12345} \) are algebraically independent. Of course, if a different parent ordering could have been chosen and a different set of algebraically independent rank correlations would have resulted. The bounds for \( r_{27}^{e_i} \) given the expert’s estimate for \( r_{17}^{e_i} \) can be obtained using the inverse of (5.21):

\[
\rho_{27|1}^{e_i}_{\text{min/max}} = \rho_{27|1}_{\text{min/max}} \sqrt{1 - (\rho_{17}^{e_i})^2}, \tag{5.22}
\]

where

\[
\begin{align*}
\rho_{27|1}_{\text{min}} &= 2 \sin\left(\frac{\pi}{6} \cdot (-1)\right) = -1 \\
\rho_{27|1}_{\text{max}} &= 2 \sin\left(\frac{\pi}{6} \cdot 1\right) = 1.
\end{align*} \tag{5.23}
\]

Applying Pearson’s transformation yields the bounds for the corresponding rank correlations.

The following example illustrates the procedure. Let’s assume expert \( e_1 \) states that \( P_{1}^{e_1} = 0.8 \), which results in \( r_{1,7}^{e_1} = 0.7953 \). Figure 5.3 depicts the relationship of \( r_{27|1} \) and \( r_{27} \) for \( r_{1,7}^{e_1} = 0.7953 \) and shows that this requires \( r_{2,7} \in [-0.5644, 0.5644] \) or equivalently \( R_{2}^{e_1} \in [-0.7097, 0.7097] \). Estimates outside this interval would lead to a conditional correlation value \( |r_{2,7|1}| > 1 \). Higher order conditional rank correlations and bounds are determined in the same manner. The above equation can be easily extended to higher order conditional correlations.

Generally, the bounds for the ratios of rank correlations should be computed in real time so that the experts’ estimates can be discussed and revised if necessary.

5.2.2 Combining Dependence Estimates

Combining the dependence estimates poses particular challenges. One reason pointed out by [Morales Nápoles, 2010] is that a convex combination of rank correlation matrices will
generally not satisfy the conditional independence statements implied by the graph as the following example shows.

**Example 5.1.** Consider the graph in figure 5.4, which stipulates \( X_1 \perp X_3 \mid X_2 \). Assuming that all variables are standard normal, \( \rho_{13;2} = 0 \) if and only if \( X_1 \perp X_3 \mid X_2 \). According to recursion (2.4) \( \rho_{13;2} = 0 \) is equivalent to \( \rho_{13} = \rho_{12} \cdot \rho_{23} \).

Let

\[
\Sigma^A = \begin{pmatrix}
1 & 0.6 & 0.12 \\
0.6 & 1 & 0.2 \\
0.12 & 0.2 & 1
\end{pmatrix}
\]

and

\[
\Sigma^B = \begin{pmatrix}
1 & 0.5 & 0.25 \\
0.5 & 1 & 0.5 \\
0.25 & 0.5 & 1
\end{pmatrix}
\]

be the covariance matrices of two experts. We can easily check that they are positive definite and represent the conditional independence statement. A direct combination, e.g. \( \Sigma^C = 0.5 \cdot \Sigma^A + 0.5 \cdot \Sigma^B \), yields

\[
\Sigma^C = \begin{pmatrix}
1 & 0.55 & 0.185 \\
0.55 & 1 & 0.35 \\
0.185 & 0.35 & 1
\end{pmatrix}
\]

\( \Sigma^C \) is a correlation matrix, but \( \Sigma^C (1,2) \cdot \Sigma^C (2,3) = 0.1925 \neq 0.185 = \Sigma^C (1,3) \). Hence, a direct combination generally does not satisfy the conditional independence statements implied by the BBN.

![Figure 5.4: Two BBNs of three nodes](image)

One may argue that the 0.1925 and 0.185 are not very different, especially not if the data is based on expert judgment. In fact, there may even be cases where a direct combination preserves the conditional independence statements of the graph, e.g. if the experts state the same correlation values. However, if the experts believe in different marginal distributions when assessing rank correlations it is still not possible to directly average rank correlations. We will demonstrate this using example 5.2 at the end of this chapter. Before we get to the example we motivate and describe the alternative approach to combine rank correlations proposed by Morales Nápoles, 2010.

Transforming the rank correlations to exceedance probabilities based on each experts marginal distributions and combining them via linear pooling also involves complications. Because the experts elicited the marginal distributions as well the medians they assume to be true for each variable \( X_1 \) to \( X_7 \) differ from expert to expert, which means that a weighted average would be taken over different events. In order to preserve conditional independence
5.2. ELICITING DEPENDENCIES

statements the following protocol has to be followed. The main idea is to compute the exceedance probabilities for each expert from his/her rank correlations that he/she "would have stated, if he/she had been asked probabilistic statements regarding the median of the DM".

1. For each variable the marginals of the experts are combined to the marginals of the DM and the respective DM’s medians are determined.
2. The rank correlation estimates, \( r_{17}^{ei}, r_{27}^{ei}, r_{37}^{ei}, r_{47}^{ei}, r_{57}^{ei} \) and \( r_{67}^{ei} \) of each expert are converted to exceedance probabilities with respect to the DM’s median.
3. These exceedance probabilities are combined to the DM’s exceedance probabilities using weights.
4. The DM’s exceedance probabilities are translated back to rank correlations.

The steps 2 - 4 are illustrated for \( r_{17} \) in (5.24):

\[
\begin{align*}
    r_{17}^{e1} & \rightarrow P_{1*}^{e1} \\
              & \vdots \\
    r_{17}^{eE} & \rightarrow P_{1*}^{eE}
\end{align*}
\]

\[
P_{1*}^{DM} = \sum_{i=1}^{E} w_\alpha(e_i) P_{1*}^{ei} \rightarrow r_{17}^{DM},
\]

where

\[
P_{1*}^{e1} = P\{F_{X_7}^{e1}(X_7) \leq F_{X_7}^{DM}(x_{1,q_{50}}) \mid F_{X_1}^{e1}(X_1) \leq F_{X_1}^{DM}(x_{1,q_{50}})\}
\]

\[
\vdots
\]

\[
P_{1*}^{eE} = P\{F_{X_7}^{eE}(X_7) \leq F_{X_7}^{DM}(x_{1,q_{50}}) \mid F_{X_1}^{eE}(X_1) \leq F_{X_1}^{DM}(x_{1,q_{50}})\}
\]

for all experts \( e_i, i = 1, ..., E \).

The remaining rank correlations \( r_{27}^{DM}, r_{37}^{DM}, r_{47}^{DM}, r_{57}^{DM} \) and \( r_{67}^{DM} \) are combined analogously. As a final step, the (conditional) rank correlations are computed in the usual way using recursion (2.4).

**Example 5.2.** Let \( X_1 \) and \( X_2 \) be random variables whose joint distribution is to be quantified by two experts, \( e_1 \) and \( e_2 \). Suppose the experts stated the three quantiles of their subjective probabilities for each variable presented in the table below and suppose that \( e_1 \)’s and expert \( e_2 \)’s estimate for \( r_{12} \) be identical: \( r_{12}^{e1} = r_{12}^{e2} = 0.8 \). We assume the normal copula to realize the joint distribution. The experts are combined with equal weights \( w_1 = w_2 = 0.5 \). The marginal distributions and the resulting decision maker distribution are shown in figure 5.5. A direct averaging of the rank correlations obviously leads to \( r_{12}^{DM} = 0.8 \). However, if we use the approach described above to combine the rank correlations we obtain \( r_{12}^{DM} = 0.8276 \). The

---

4. In section 7.2 similar calculations are described in more detail.
Table 5.1: Experts’ quantiles

<table>
<thead>
<tr>
<th></th>
<th>$e_1$</th>
<th>$e_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{1,q5}$</td>
<td>0.01</td>
<td>0.005</td>
</tr>
<tr>
<td>$x_{1,q50}$</td>
<td>0.1</td>
<td>0.025</td>
</tr>
<tr>
<td>$x_{1,q95}$</td>
<td>0.3</td>
<td>0.05</td>
</tr>
<tr>
<td>$x_{2,q5}$</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>$x_{2,q50}$</td>
<td>0.1</td>
<td>0.05</td>
</tr>
<tr>
<td>$x_{2,q95}$</td>
<td>0.25</td>
<td>0.1</td>
</tr>
</tbody>
</table>

difference in this example is not very high, but it demonstrates that it is not possible to combine rank correlation directly. We can easily imagine a case with more variables and experts stating identical rank correlations. The directly combined DM correlation matrix is identical to either of the experts’ matrices and thus still complies with the conditional independence statements, but it is incorrect.

figure 5.5: Marginal distributions of the DM an experts $e_5$ and $e_6$

5.3 Conclusion

This chapter has shown that methods are available to obtain both the marginals as well as the dependence structure from experts and, thus, that it is possible to quantify a nonparametric BBN solely based on expert judgment.

While the classical model is a time-tested approach for eliciting marginal distributions, the elicitation of the dependence structure is relatively new and less tried. The classical model treats experts as classical statistical hypothesis and offers a framework to derive strictly proper scoring rules and weights according to the scoring. This enables the analyst to verify for
himself whether an expert’s answers make sense and to justify the marginals of the decision maker that he constructs with the weights. For the dependence elicitation, on the other hand, such a framework is as yet not available. A proper scoring scheme based on seed questions representing, to the same degree, the expert’s skill to elicit marginals as well as his/her skill to elicit dependence is needed. At the present time one can make the questionable assumption that experts perform equally in estimating marginals as in estimating rank correlations and use the weights from the classical model, but other than that there is no indication for how to determine the weights. As a consequence, there is less ground to defend a chosen dependence structure than the marginal distributions.
6. Elicitation and Results

Since no information about the variables was available, both the marginals and the (conditional) rank correlations had to be elicited with structured expert judgment. 14 employees of an oil and gas company participated as experts in the study. The elicitation protocol (Appendix A) consisted of three different parts: questions on probability distributions and correlations, calibration questions and open questions about management actions. The latter are not of interest for this thesis.

For the elicitation of the marginal distributions, the variables have been interpreted as relative frequencies, e.g. contractor performance ($V_7$) was operationalized as the number of erroneously executed tasks out of 1000 safety critical tasks. Other examples of operationalization can be found in Appendix. For each variable the experts stated the 5%-,$50%$- and 95%-quantiles of their subjective probabilities. The unconditional and conditional rank correlations were elicited with the procedure described in section 5.2.1. First the experts were asked to order the parent nodes according to the highest absolute rank correlation to the child node, $V_7$. Then they estimated the exceedance probability $R_1$ given by (5.13) and the five ratios of rank correlations, $R_2-R_6$ given by (5.14) - (5.18).

The protocol contained 12 calibration questions, which again asked the experts for the 5%-,$50%$- and 95%-quantiles of their subjective probabilities of uncertain quantities. They are designed to compute the performance based weights for the marginals in Excalibur. Three of these questions covered human factors, such as supervision, availability of information and adequate communications. The remaining questions targeted more general knowledge about fatalities in the oil and gas industry and onshore as well as offshore operation.

6.1 Marginal Distributions

When the experts’ quantiles for all distribution questions and calibration questions are entered in Excalibur, the software computes the corresponding minimum information distribution which can then be exported. The distributions of expert $e_5$ and $e_6$ for variable $V_7$ are presented in figure 6.1a. There is two issues with these distributions. First, the probabilities for all values below the 5%-quantile are negative. This seems to be a bug of the current version of the program. Second, Excalibur uses the 10% overshoot rule to determine the intrinsic range of the variables. As a result, $V_7$ only has probability mass on the interval $[0,0.77]$. Taking
into account that the variables are defined as likelihood (cf. table 4.1) and operationalized as relative frequencies for the questions in the elicitation protocol, it seems more meaningful to use the interval \([0,1]\). Especially, if one considers that also non-mathematicians will use the model. They may have difficulties to grasp the concept of attributing probability 0 to the event that an error occurs, but having probability \(1\) that an error occurs with less than probability \(0.77\), that is, \(P\{V_7 \leq 0.77\} = 1\). Furthermore, it may be relevant for the holistic risk model that a human error can happen. It may desirable to condition on the event "task has been executed with an error" and explore how the many other variables react to that and what the consequences for the overall system risk are. For this reason, the empirical cumulative distributions for each variable and each expert have been computed in Matlab for all quantiles from 0.005 to 0.995 with steps of 0.005. The distribution for \(V_7\) is shown in figure 6.1b.

![Figure 6.1: Distributions of experts \(e_5\) and \(e_6\) for \(V_7\)](image)

Changing the intrinsic range influences of course the relative informativeness of the experts. However, it changes it for all experts in the same way. Therefore, the scores and weights calculated by Excalibur can be still used. The Excalibur results for the global weight decision maker (DM) with DM optimization is presented in table 6.1. A significance level of \(\alpha = 0.6876\) (supposedly) maximizes the virtual weight of the decision maker. The second column shows the calibration score, the third column the information score for the seed variables and the fourth column the unnormalized weights using \(\alpha = 0.6876\). Column six and seven contain the normalized weights with and without the virtual weight of the decision maker, respectively. For this choice of \(\alpha\) only expert 5 and expert 6 receive a nonzero weight. The last row contains the item weight decision maker. He has the same significance level \(\alpha = 0.6876\) and the same calibration score as the global weight decision maker, but a slightly higher information score, \(I(DM_{item}) = 0.6805\) versus \(I(DM_{global}) = 0.6614\). Because he performs more or less the same as the global weight DM and has not been used in the previous analysis, we do not discuss item weights in more detail. According to Excalibur the optimal global weights are \(w_5 = 0.5093\) and \(w_6 = 0.4907\). These weights have been used in the previous analysis. The DM's estimate for the marginals of each variable have been computed by Excalibur based on equation (5.12) for \(j = 1, \ldots, 7\).
Surprisingly, the calibration and information scores of the DM are lower than the ones of the experts. Expert 5 and expert 6 have the same calibration score $C(e_5) = C(e_6) = 0.6876$, but expert $e_5$ has a slightly higher informativeness, $I(e_5) = 0.8534$, than expert $e_6$, $I(e_6) = 0.8222$. As [Hora, 2004] showed, it is impossible to linearly combine the probability distributions of two well-calibrated experts and to produce an at least equally well calibrated DM, unless they provide the same values for each quantile. Hence, combining $e_5$ and $e_6$ results in a DM with a worse calibration than each of the experts, $C(DM) = 0.6638$. Since the unnormalized weights are calculated as $w_{e_i}^* = C(e_i) \cdot I(e_i) \cdot I_\alpha(C(e_i))$, all experts with a calibration $C(e_i) \geq \alpha_{optimal}$ receive a weight. Experts $e_5$ and $e_6$ have an identical calibration score, thus both receive a weight. In this particular and rare case the method fails to find a DM that performs at least equally well as the individual experts. The Excalibur output, $w_5 = 0.5093$ and $w_6 = 0.4907$, does not correspond to the optimal global weights. The correct optimal DM is based on the global weights $w_{e_5} = 1$ and $w_{e_i} = 0$, for $i = 1, .., 14$, $i \neq 5$.

Table 6.1: Excalibur Output: optimized global and item weights with significance level $\alpha = 0.6876$

<table>
<thead>
<tr>
<th></th>
<th>Calibration</th>
<th>Information</th>
<th>Unnormalized weights</th>
<th>Normalized weights</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>without DM</td>
<td>with DM</td>
</tr>
<tr>
<td>$e_1$</td>
<td>0.5338</td>
<td>0.4866</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_2$</td>
<td>0.00348</td>
<td>0.4561</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_3$</td>
<td>0.00348</td>
<td>1.016</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_4$</td>
<td>0.02651</td>
<td>0.9024</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_5$</td>
<td>0.6876</td>
<td>0.8534</td>
<td>0.5868</td>
<td>0.5093</td>
</tr>
<tr>
<td>$e_6$</td>
<td>0.6876</td>
<td>0.8222</td>
<td>0.5654</td>
<td>0.4907</td>
</tr>
<tr>
<td>$e_7$</td>
<td>0.5405</td>
<td>0.6878</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_8$</td>
<td>0.08763</td>
<td>1.089</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_9$</td>
<td>0.2357</td>
<td>0.6681</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_{10}$</td>
<td>0.00348</td>
<td>0.8325</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_{11}$</td>
<td>0.01566</td>
<td>0.7833</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_{12}$</td>
<td>0.6638</td>
<td>0.8263</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_{13}$</td>
<td>0.00007086</td>
<td>1.02</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_{14}$</td>
<td>0.5338</td>
<td>0.9099</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$DM_{global}$</td>
<td>0.6638</td>
<td>0.6614</td>
<td>0.439</td>
<td>0.2759</td>
</tr>
<tr>
<td>$DM_{item}$</td>
<td>0.6638</td>
<td>0.6805</td>
<td>0.4517</td>
<td>0.2816</td>
</tr>
</tbody>
</table>

For BBN that has been built prior to this thesis project, we will refer to this BBN as model $O$, the distributions have been exported from Excalibur and the weights $w_5 = 0.5093$ and $w_6 = 0.4907$ have been used. The new models will use the distributions obtained from Matlab. Weights for the new models are discussed after the following section.

\footnote{for a slightly different measure of calibration}
6.2 Rank Correlations

Because the experts ranked the variables in different orders, they assessed different rank correlations. The node V7 is now denoted as $X_7$ and the parent nodes are denoted as $X_1, ..., X_6$ such that $|r_{17}| \geq |r_{27}| \geq |r_{37}| \geq |r_{47}| \geq |r_{57}| \geq |r_{67}|$. Hence, $X_4$ may denote a different variable for each expert. An overview of the experts’ ranking is given in table 6.2. The last row contains the decision maker (DM), whose ordering does not satisfy $|r_{17}| \geq |r_{27}| \geq |r_{37}| \geq |r_{47}| \geq |r_{57}| \geq |r_{67}|$, but has been chosen by the analyst.

The experts’ assessments are presented in table 6.3. The first six rows correspond to expert $e_1$, the second six rows to expert $e_2$ and so on. The respective first rows of each expert correspond to $R_1$ and the result that can be computed from it, $r_{17}$. Likewise, the respective second rows correspond to $R_2$, $r_{27}$ and $r_{27|1}$. The same applies to the remaining four rows. Now we explain the columns. The third column contains the experts’ estimates of $R_j$, $j = 1, ..., 6$, (5.13 - 5.18). The seventh column contains the rank correlation $r_{j7}$ that corresponds to this estimate. Column four and five and column eight and nine, contain the bounds for $R_j$ and $r_{j7}$, respectively, considering the experts’ previous assessments or previous bounds. If a previous assessment was greater than its bound, that bound has been used to calculate the bound for the following assessment. Lets clarify this with the example of expert $e_1$; the remaining columns later on are discussed later on in this section. Expert $e_1$ stated that $P\{X_7 \geq x_{7,q0}^{e_1} \mid X_1 \geq x_{1,q0}^{e_1}\} = 0.8$. The corresponding rank correlation is $r_{17}^{e_1} = 0.8$. It is the correlation between quality of procedures, $V_4$, and contractor performance, $V_7$, as this expert ranked $V_4$ as most important factor. He then stated $r_{27}^{e_1}/r_{17}^{e_1} = 0.95$ which gives $r_{27}^{e_1} = 0.76$, the correlation between competence, $V_1$, and contractor performance, $V_7$. However, the valid interval for $r_{27}^{e_1}$ for the given model structure would have been $[-0.56, 0.56]$.

Table 6.2: Experts’ ranking of the parent nodes according to highest absolute rank correlation of the child node V7

<table>
<thead>
<tr>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>$X_2$</td>
<td>$X_5$</td>
<td>$X_4$</td>
<td>$X_1$</td>
<td>$X_3$</td>
<td>$X_6$</td>
</tr>
<tr>
<td>$e_2$</td>
<td>$X_1$</td>
<td>$X_4$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_6$</td>
<td>$X_5$</td>
</tr>
<tr>
<td>$e_3$</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_5$</td>
<td>$X_4$</td>
<td>$X_6$</td>
</tr>
<tr>
<td>$e_4$</td>
<td>$X_3$</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_5$</td>
<td>$X_4$</td>
<td>$X_6$</td>
</tr>
<tr>
<td>$e_5$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_1$</td>
<td>$X_5$</td>
<td>$X_6$</td>
<td>$X_4$</td>
</tr>
<tr>
<td>$e_6$</td>
<td>$X_1$</td>
<td>$X_4$</td>
<td>$X_3$</td>
<td>$X_2$</td>
<td>$X_5$</td>
<td>$X_6$</td>
</tr>
<tr>
<td>$e_7$</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_4$</td>
<td>$X_3$</td>
<td>$X_6$</td>
<td>$X_5$</td>
</tr>
<tr>
<td>$e_8$</td>
<td>$X_1$</td>
<td>$X_3$</td>
<td>$X_4$</td>
<td>$X_2$</td>
<td>$X_5$</td>
<td>$X_6$</td>
</tr>
<tr>
<td>$e_9$</td>
<td>$X_2$</td>
<td>$X_4$</td>
<td>$X_3$</td>
<td>$X_1$</td>
<td>$X_5$</td>
<td>$X_6$</td>
</tr>
<tr>
<td>$e_{10}$</td>
<td>$X_4$</td>
<td>$X_1$</td>
<td>$X_5$</td>
<td>$X_6$</td>
<td>$X_2$</td>
<td>$X_3$</td>
</tr>
<tr>
<td>$e_{11}$</td>
<td>$X_2$</td>
<td>$X_1$</td>
<td>$X_4$</td>
<td>$X_3$</td>
<td>$X_5$</td>
<td>$X_6$</td>
</tr>
<tr>
<td>$e_{12}$</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_4$</td>
<td>$X_5$</td>
<td>$X_6$</td>
</tr>
<tr>
<td>$e_{13}$</td>
<td>$X_1$</td>
<td>$X_3$</td>
<td>$X_4$</td>
<td>$X_5$</td>
<td>$X_6$</td>
<td>$X_2$</td>
</tr>
<tr>
<td>$e_{14}$</td>
<td>$X_1$</td>
<td>$X_3$</td>
<td>$X_2$</td>
<td>$X_4$</td>
<td>$X_5$</td>
<td>$X_6$</td>
</tr>
<tr>
<td>DM</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$X_4$</td>
<td>$X_5$</td>
<td>$X_6$</td>
</tr>
</tbody>
</table>
Table 6.3: Results of the dependence elicitation. For computations in Matlab the (conditional) rank correlations have been discretized as $-0.99 : 0.01 : 0.99$.

<table>
<thead>
<tr>
<th>e</th>
<th>$R$</th>
<th>$R_{\text{min}}$</th>
<th>$R_{\text{max}}$</th>
<th>$\hat{r}$</th>
<th>$r_{\text{min}}$</th>
<th>$r_{\text{max}}$</th>
<th>$\text{err}$</th>
<th>$\text{err}_{\text{relative}}$</th>
<th>$\sum \text{err}_{\text{relative}}$</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>e1</td>
<td>$R_1$</td>
<td>0.8</td>
<td> </td>
<td> </td>
<td>$r_{17}$</td>
<td>0.8</td>
<td>-0.99</td>
<td>0.99</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_2$</td>
<td>0.95</td>
<td>-0.71</td>
<td>0.71</td>
<td>$r_{27}$</td>
<td>0.76</td>
<td>-0.56</td>
<td>0.56</td>
<td>0.19</td>
<td>0.34</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_3$</td>
<td>0.7</td>
<td>-0.09</td>
<td>0.09</td>
<td>$r_{37}$</td>
<td>0.56</td>
<td>-0.07</td>
<td>0.07</td>
<td>0.48</td>
<td>6.44</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_4$</td>
<td>0.68</td>
<td>-0.01</td>
<td>0.01</td>
<td>$r_{47}$</td>
<td>0.54</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.53</td>
<td>52.77</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_5$</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>$r_{57}$</td>
<td>0.32</td>
<td>0</td>
<td>0</td>
<td>0.32</td>
<td>234.2</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_6$</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>$r_{67}$</td>
<td>0.24</td>
<td>0</td>
<td>0</td>
<td>0.24</td>
<td>1310.82</td>
</tr>
<tr>
<td>e2</td>
<td>$R_1$</td>
<td>0.9</td>
<td> </td>
<td> </td>
<td>$r_{17}$</td>
<td>0.95</td>
<td>-0.99</td>
<td>0.99</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_2$</td>
<td>0.95</td>
<td>-0.31</td>
<td>0.31</td>
<td>$r_{27}$</td>
<td>0.9</td>
<td>-0.29</td>
<td>0.29</td>
<td>0.61</td>
<td>2.06</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_3$</td>
<td>0.9</td>
<td>-0.04</td>
<td>0.04</td>
<td>$r_{37}$</td>
<td>0.85</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.81</td>
<td>20.64</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_4$</td>
<td>0.7</td>
<td>-0.01</td>
<td>0.01</td>
<td>$r_{47}$</td>
<td>0.66</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.66</td>
<td>124.16</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_5$</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>$r_{57}$</td>
<td>0.47</td>
<td>0</td>
<td>0</td>
<td>0.47</td>
<td>663.85</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_6$</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>$r_{67}$</td>
<td>0.24</td>
<td>0</td>
<td>0</td>
<td>0.24</td>
<td>2471.05</td>
</tr>
<tr>
<td>e3</td>
<td>$R_1$</td>
<td>0.8</td>
<td> </td>
<td> </td>
<td>$r_{17}$</td>
<td>0.8</td>
<td>-0.99</td>
<td>0.99</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_2$</td>
<td>0.8</td>
<td>-0.71</td>
<td>0.71</td>
<td>$r_{27}$</td>
<td>0.64</td>
<td>-0.56</td>
<td>0.56</td>
<td>0.07</td>
<td>0.13</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_3$</td>
<td>0.7</td>
<td>-0.09</td>
<td>0.09</td>
<td>$r_{37}$</td>
<td>0.56</td>
<td>-0.07</td>
<td>0.07</td>
<td>0.48</td>
<td>6.44</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_4$</td>
<td>0.4</td>
<td>-0.01</td>
<td>0.01</td>
<td>$r_{47}$</td>
<td>0.32</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.31</td>
<td>30.63</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_5$</td>
<td>0.35</td>
<td>0</td>
<td>0</td>
<td>$r_{57}$</td>
<td>0.28</td>
<td>0</td>
<td>0</td>
<td>0.28</td>
<td>204.8</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_6$</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>$r_{67}$</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>1092.18</td>
</tr>
<tr>
<td>e4</td>
<td>$R_1$</td>
<td>0.75</td>
<td> </td>
<td> </td>
<td>$r_{17}$</td>
<td>0.69</td>
<td>-0.99</td>
<td>0.99</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_2$</td>
<td>0.9</td>
<td>-0.99</td>
<td>0.99</td>
<td>$r_{27}$</td>
<td>0.62</td>
<td>-0.68</td>
<td>0.68</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_3$</td>
<td>0.75</td>
<td>-0.42</td>
<td>0.42</td>
<td>$r_{37}$</td>
<td>0.52</td>
<td>-0.29</td>
<td>0.29</td>
<td>0.23</td>
<td>0.8</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_4$</td>
<td>0.6</td>
<td>-0.06</td>
<td>0.06</td>
<td>$r_{47}$</td>
<td>0.41</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.38</td>
<td>9.75</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_5$</td>
<td>0.55</td>
<td>-0.01</td>
<td>0.01</td>
<td>$r_{57}$</td>
<td>0.38</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.37</td>
<td>72.29</td>
</tr>
<tr>
<td>&amp;</td>
<td>$R_6$</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>$r_{67}$</td>
<td>0.35</td>
<td>0</td>
<td>0</td>
<td>0.34</td>
<td>494.49</td>
</tr>
<tr>
<td>e5</td>
<td>$R_1$</td>
<td>0.8</td>
<td> </td>
<td> </td>
<td>$r_{17}$</td>
<td>0.8</td>
<td>-0.99</td>
<td>0.99</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$R_2$</td>
<td>0.9 -0.71 0.71</td>
<td>$r_{27}$</td>
<td>0.72 -0.56 0.56</td>
<td>$r_{271}$</td>
<td>0.99</td>
<td>0.15</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_3$</td>
<td>0.9 -0.09 0.09</td>
<td>$r_{37}$</td>
<td>0.72 -0.07 0.07</td>
<td>$r_{3712}$</td>
<td>0.99</td>
<td>0.64</td>
<td>8.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_4$</td>
<td>0.75 -0.01 0.01</td>
<td>$r_{47}$</td>
<td>0.6 -0.01 0.01</td>
<td>$r_{47123}$</td>
<td>0.99</td>
<td>0.59</td>
<td>58.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_5$</td>
<td>0.6 0 0</td>
<td>$r_{57}$</td>
<td>0.48 0 0</td>
<td>$r_{571234}$</td>
<td>0.99</td>
<td>0.48</td>
<td>351.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_6$</td>
<td>0.5 0 0</td>
<td>$r_{67}$</td>
<td>0.4 0 0</td>
<td>$r_{6712345}$</td>
<td>0.99</td>
<td>0.4</td>
<td>2185.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_6$</td>
<td>$R_1$</td>
<td>0.8</td>
<td></td>
<td>$r_{17}$</td>
<td>0.8 -0.99 0.99</td>
<td>$r_{17}$</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>2103.22</td>
</tr>
<tr>
<td>$R_2$</td>
<td>0.9 -0.71 0.71</td>
<td>$r_{27}$</td>
<td>0.72 -0.56 0.56</td>
<td>$r_{271}$</td>
<td>0.99</td>
<td>0.15</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_3$</td>
<td>0.8 -0.09 0.09</td>
<td>$r_{37}$</td>
<td>0.64 -0.07 0.07</td>
<td>$r_{3712}$</td>
<td>0.99</td>
<td>0.56</td>
<td>7.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_4$</td>
<td>0.7 -0.01 0.01</td>
<td>$r_{47}$</td>
<td>0.56 -0.01 0.01</td>
<td>$r_{47123}$</td>
<td>0.99</td>
<td>0.55</td>
<td>54.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_5$</td>
<td>0.5 0 0</td>
<td>$r_{57}$</td>
<td>0.4 0 0</td>
<td>$r_{571234}$</td>
<td>0.99</td>
<td>0.4</td>
<td>293.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_6$</td>
<td>0.4 0 0</td>
<td>$r_{67}$</td>
<td>0.32 0 0</td>
<td>$r_{6712345}$</td>
<td>0.99</td>
<td>0.32</td>
<td>1748.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_7$</td>
<td>$R_1$</td>
<td>0.9</td>
<td></td>
<td>$r_{17}$</td>
<td>0.95 -0.99 0.99</td>
<td>$r_{17}$</td>
<td>0.95</td>
<td>0</td>
<td>0</td>
<td>7026.29</td>
</tr>
<tr>
<td>$R_2$</td>
<td>0.9 -0.31 0.31</td>
<td>$r_{27}$</td>
<td>0.85 -0.29 0.29</td>
<td>$r_{271}$</td>
<td>0.99</td>
<td>0.56</td>
<td>1.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_3$</td>
<td>0.9 -0.04 0.04</td>
<td>$r_{37}$</td>
<td>0.85 -0.04 0.04</td>
<td>$r_{3712}$</td>
<td>0.99</td>
<td>0.81</td>
<td>20.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_4$</td>
<td>0.8 -0.01 0.01</td>
<td>$r_{47}$</td>
<td>0.76 -0.01 0.01</td>
<td>$r_{47123}$</td>
<td>0.99</td>
<td>0.75</td>
<td>142.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_5$</td>
<td>0.7 0 0</td>
<td>$r_{57}$</td>
<td>0.66 0 0</td>
<td>$r_{571234}$</td>
<td>0.99</td>
<td>0.66</td>
<td>929.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_6$</td>
<td>0.6 0 0</td>
<td>$r_{67}$</td>
<td>0.57 0 0</td>
<td>$r_{6712345}$</td>
<td>0.99</td>
<td>0.57</td>
<td>5931.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_8$</td>
<td>$R_1$</td>
<td>0.9</td>
<td></td>
<td>$r_{17}$</td>
<td>0.95 -0.99 0.99</td>
<td>$r_{17}$</td>
<td>0.95</td>
<td>0</td>
<td>0</td>
<td>1468.2</td>
</tr>
<tr>
<td>$R_2$</td>
<td>0.6 -0.31 0.31</td>
<td>$r_{27}$</td>
<td>0.57 -0.29 0.29</td>
<td>$r_{271}$</td>
<td>0.99</td>
<td>0.27</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_3$</td>
<td>0.5 -0.04 0.04</td>
<td>$r_{37}$</td>
<td>0.47 -0.04 0.04</td>
<td>$r_{3712}$</td>
<td>0.99</td>
<td>0.43</td>
<td>11.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_4$</td>
<td>0.4 -0.01 0.01</td>
<td>$r_{47}$</td>
<td>0.38 -0.01 0.01</td>
<td>$r_{47123}$</td>
<td>0.99</td>
<td>0.37</td>
<td>70.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_5$</td>
<td>0.3 0 0</td>
<td>$r_{57}$</td>
<td>0.28 0 0</td>
<td>$r_{571234}$</td>
<td>0.99</td>
<td>0.28</td>
<td>397.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_6$</td>
<td>0.1 0 0</td>
<td>$r_{67}$</td>
<td>0.09 0 0</td>
<td>$r_{6712345}$</td>
<td>0.99</td>
<td>0.09</td>
<td>987.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_9$</td>
<td>$R_1$</td>
<td>0.7</td>
<td></td>
<td>$r_{17}$</td>
<td>0.57 -0.99 0.99</td>
<td>$r_{17}$</td>
<td>0.57</td>
<td>0</td>
<td>0</td>
<td>14.14</td>
</tr>
<tr>
<td>$R_2$</td>
<td>0.8 -1.38 1.38</td>
<td>$r_{27}$</td>
<td>0.46 -0.79 0.79</td>
<td>$r_{271}$</td>
<td>0.57</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_3$</td>
<td>0.8 -1.11 1.11</td>
<td>$r_{37}$</td>
<td>0.46 -0.63 0.63</td>
<td>$r_{3712}$</td>
<td>0.7</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_4$</td>
<td>0.7 -0.76 0.76</td>
<td>$r_{47}$</td>
<td>0.4 -0.43 0.43</td>
<td>$r_{47123}$</td>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_5$</td>
<td>0.6 -0.31 0.31</td>
<td>$r_{57}$</td>
<td>0.34 -0.18 0.18</td>
<td>$r_{571234}$</td>
<td>0.99</td>
<td>0.16</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_6$</td>
<td>0.6 -0.04 0.04</td>
<td>$r_{67}$</td>
<td>0.34 -0.02 0.02</td>
<td>$r_{6712345}$</td>
<td>0.99</td>
<td>0.32</td>
<td>13.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_{10}$</td>
<td>$R_1$</td>
<td>0.9</td>
<td></td>
<td>$r_{17}$</td>
<td>0.95 -0.99 0.99</td>
<td>$r_{17}$</td>
<td>0.95</td>
<td>0</td>
<td>0</td>
<td>4613.98</td>
</tr>
<tr>
<td>$R_i$</td>
<td>$R_j$</td>
<td>$r_{ij}$</td>
<td>$r_{ij}$</td>
<td>$r_{ij}$</td>
<td>$r_{ij}$</td>
<td>$r_{ij}$</td>
<td>$r_{ij}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.2. RANK CORRELATIONS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_1$</td>
<td>0.9</td>
<td>-0.31</td>
<td>0.31</td>
<td>0.95</td>
<td>-0.99</td>
<td>0.99</td>
<td>0.95</td>
<td>0</td>
<td>0</td>
<td>669.37</td>
</tr>
<tr>
<td>$R_2$</td>
<td>0.5</td>
<td>-0.31</td>
<td>0.31</td>
<td>0.97</td>
<td>-0.29</td>
<td>0.29</td>
<td>0.97</td>
<td>0.18</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>$R_3$</td>
<td>0.85</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.8</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.8</td>
<td>0.77</td>
<td>19.44</td>
<td></td>
</tr>
<tr>
<td>$R_4$</td>
<td>0.8</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.76</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.76</td>
<td>0.75</td>
<td>142.05</td>
<td></td>
</tr>
<tr>
<td>$R_5$</td>
<td>0.75</td>
<td>0</td>
<td>0</td>
<td>0.71</td>
<td>0</td>
<td>0</td>
<td>0.71</td>
<td>0.71</td>
<td>996.27</td>
<td></td>
</tr>
<tr>
<td>$R_6$</td>
<td>0.7</td>
<td>0</td>
<td>0</td>
<td>0.66</td>
<td>0</td>
<td>0</td>
<td>0.66</td>
<td>0.66</td>
<td>6920.74</td>
<td></td>
</tr>
<tr>
<td>$R_1$</td>
<td>0.8</td>
<td>-0.71</td>
<td>0.71</td>
<td>0.56</td>
<td>-0.56</td>
<td>0.56</td>
<td>0.56</td>
<td>0.44</td>
<td>3.77</td>
<td></td>
</tr>
<tr>
<td>$R_2$</td>
<td>0.7</td>
<td>-0.15</td>
<td>0.15</td>
<td>0.56</td>
<td>-0.12</td>
<td>0.12</td>
<td>0.56</td>
<td>0.46</td>
<td>29.45</td>
<td></td>
</tr>
<tr>
<td>$R_3$</td>
<td>0.6</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.48</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.48</td>
<td>0.48</td>
<td>225.45</td>
<td></td>
</tr>
<tr>
<td>$R_4$</td>
<td>0.6</td>
<td>0</td>
<td>0</td>
<td>0.48</td>
<td>0</td>
<td>0</td>
<td>0.48</td>
<td>0.48</td>
<td>1682.96</td>
<td></td>
</tr>
<tr>
<td>$R_5$</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0.19</td>
<td>0</td>
<td>0</td>
<td>0.19</td>
<td>0.19</td>
<td>264.94</td>
<td></td>
</tr>
<tr>
<td>$R_6$</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0.19</td>
<td>0</td>
<td>0</td>
<td>0.19</td>
<td>0.19</td>
<td>1976.64</td>
<td></td>
</tr>
</tbody>
</table>

61
During the elicitation process, the bounds have not been calculated in real time and it could therefore not be ensured that the experts’ assessments are consistent with previous assessment. Indeed, not just expert $e_1$ but all experts gave estimates out of bounds. For model $O$ this situation has been “fixed” by applying the linear map

$$f : [0, 1] \rightarrow [r_{ij_{\min}}, r_{ij_{\max}}], \quad f(\hat{r}_{ij}) := (r_{ij_{\max}} - r_{ij_{\min}}) \cdot \hat{r}_{ij} + r_{ij_{\min}}, \quad \text{for } i = 2, ..., 6 \quad (6.1)$$

where $\hat{r}_{ij}$ is the expert’s estimate and $r_{ij_{\max}}$ and $r_{ij_{\min}}$ are the bounds based the expert’s previous "fixed" assessments $r_{j7} = f(\hat{r}_{j7})$, $j = 1, ..., i - 1$. For expert $e_5$ and $e_6$ the following rank correlations result

$$\begin{bmatrix}
  r_{17}^{e_5,\text{map}} \\
  r_{27}^{e_5,\text{map}} \\
  r_{37}^{e_5,\text{map}} \\
  r_{47}^{e_5,\text{map}} \\
  r_{57}^{e_5,\text{map}} \\
  r_{67}^{e_5,\text{map}} \\
\end{bmatrix} = \begin{bmatrix}
  0.8090 \\
  0.4684 \\
  0.2820 \\
  0.1071 \\
  0.0375 \\
  0 \end{bmatrix}, \quad \begin{bmatrix}
  r_{17}^{e_6,\text{map}} \\
  r_{27}^{e_6,\text{map}} \\
  r_{37}^{e_6,\text{map}} \\
  r_{47}^{e_6,\text{map}} \\
  r_{57}^{e_6,\text{map}} \\
  r_{67}^{e_6,\text{map}} \\
\end{bmatrix} = \begin{bmatrix}
  0.8090 \\
  0.4684 \\
  0.2118 \\
  0.1133 \\
  0 \\
  -0.0520 \end{bmatrix}. \quad (6.2)$$

"Fixing" the experts assessments is necessary, unless one assumes a different model structure (see chapter 5). Otherwise the resulting correlation matrix is not positive definite and the conditional rank correlations are outside $[-1, 1]$ and/or complex. The challenge is to modify the estimates such that they can be used to make a model, while adding as little information as possible. The mapping approach described above may not be the optimal approach to adjust the estimates. It has some serious drawbacks. Foremost, it is not clear why a map from $[0, 1] \rightarrow [r_{\min}, r_{\max}]$ is used and not from $[-1, 1] \rightarrow [r_{\min}, r_{\max}]$, which seems more natural. But in general, such a mapping approach involves changing all estimates, even the ones which are within the bounds, and is a strong intervention of the analyst. Consider for example expert $e_9$, who stated

$$\begin{bmatrix}
  r_{17}^{e_9} \\
  r_{27}^{e_9} \\
  r_{37}^{e_9} \\
  r_{47}^{e_9} \\
  r_{57}^{e_9} \\
  r_{67}^{e_9} \\
\end{bmatrix} = \begin{bmatrix}
  0.57 \\
  0.46 \\
  0.46 \\
  0.4 \\
  0.34 \\
  0.34 \\
\end{bmatrix}. \quad (6.3)$$

The first four estimates are consistent. The bounds for $r_{57}$ are $[-0.18, 0.18]$ and the ones for $r_{67}$ are $[-0.02, 0.02]$. Using a map from $[0, 1] \rightarrow [r_{\min}, r_{\max}]$ yields

$$\begin{bmatrix}
  r_{17}^{e_9,\text{map}} \\
  r_{27}^{e_9,\text{map}} \\
  r_{37}^{e_9,\text{map}} \\
  r_{47}^{e_9,\text{map}} \\
  r_{57}^{e_9,\text{map}} \\
  r_{67}^{e_9,\text{map}} \\
\end{bmatrix} = \begin{bmatrix}
  0.56 \\
  0.47 \\
  0.37 \\
  0.20 \\
  0.09 \\
  0.09 \\
\end{bmatrix}. \quad (6.4)$$

While the first two entries are still comparable, the differences for $r_{37} - r_{67}$ are significant.

A less invasive approach may be to assume that an expert, who is out of bounds, intended to assign the highest (or smallest) correlation possible and to set his assessment to the maximum
The map adjusts the second entry to a value that is significantly smaller than the bound, to sequent rank correlations are close to zero. As we see later on in section 6.3, experts' estimates, which is reduced to the maximum or minimum possible value, the bound for the sub-
if the expert is mostly consistent.
Comparing (6.3), (6.4) and (6.5) shows that this new approach is less invasive than the map, we also have a look at their original, the mapped and the new rank correlations and compare and

\[
\begin{pmatrix}
  r_{e_9,new}^{17} \\
  r_{e_9,new}^{27} \\
  r_{e_9,new}^{37} \\
  r_{e_9,new}^{37} \\
  r_{e_9,new}^{57} \\
  r_{e_9,new}^{67}
\end{pmatrix}
= 
\begin{pmatrix}
  0.57 \\
  0.46 \\
  0.46 \\
  0.4 \\
  0.18 \\
  0.02
\end{pmatrix}
\] (6.5)

Comparing (6.3), (6.4) and (6.5) shows that this new approach is less invasive than the map, if the expert is mostly consistent.

A disadvantage of this approach is that, as soon as an expert makes an inconsistent statement, which is reduced to the maximum or minimum possible value, the bound for the subsequent rank correlations are close to zero. As we see later on in section 6.3, experts e_4, e_5 and e_6 come, in addition to expert e_9, into consideration for a DM combination. Therefore we also have a look at their original, the mapped and the new rank correlations and compare the effect.

\[
\begin{pmatrix}
  r_{e_4,new}^{17} \\
  r_{e_4,new}^{27} \\
  r_{e_4,new}^{37} \\
  r_{e_4,new}^{37} \\
  r_{e_4,new}^{57} \\
  r_{e_4,new}^{67}
\end{pmatrix}
= 
\begin{pmatrix}
  0.69 \\
  0.52 \\
  0.41 \\
  0.38 \\
  0.35 \\
  0
\end{pmatrix}
\] (6.6)

\[
\begin{pmatrix}
  r_{e_5,new}^{17} \\
  r_{e_5,new}^{27} \\
  r_{e_5,new}^{37} \\
  r_{e_5,new}^{37} \\
  r_{e_5,new}^{57} \\
  r_{e_5,new}^{67}
\end{pmatrix}
= 
\begin{pmatrix}
  0.8 \\
  0.72 \\
  0.72 \\
  0.6 \\
  0.48 \\
  0
\end{pmatrix}
\] (6.7)

\[
\begin{pmatrix}
  r_{e_6,new}^{17} \\
  r_{e_6,new}^{27} \\
  r_{e_6,new}^{37} \\
  r_{e_6,new}^{37} \\
  r_{e_6,new}^{57} \\
  r_{e_6,new}^{67}
\end{pmatrix}
= 
\begin{pmatrix}
  0.8 \\
  0.72 \\
  0.64 \\
  0.56 \\
  0.4 \\
  0.32
\end{pmatrix}
\] (6.8)

Let us first consider expert e_5 who always exceeds the bounds and is ranked 10th out of 14. The map adjusts the second entry to a value that is significantly smaller than the bound, to 0.45 instead of 0.56. As a consequence the bounds for the following rank correlation are much
wider. While the allowed interval for \( r_{37} \) is \([-0.07, 0.07]\) given that \( r_{17} = 0.8 \) and \( r_{27} = 0.56 \), it is only \([-0.42, 0.42]\) if we have given that \( r_{17} = 0.79 \) and \( r_{27} = 0.45 \). For \( r_{37} \), which is mapped to 0.27 and the following rank correlations, \( r_{47} \) to \( r_{67} \), the mapping is actually closer to the original assessment of the expert. Now one could argue that the mapping is providing better results, but that is not necessarily true. It creates a result that is closer to the judgment of an expert who is not able to provide consistent answers. It does not necessarily create a result that better reflects the real world. Whether the new approach yields more realistic estimates is questionable.

Something similar happens to expert \( e_4 \). However, in this case the first two assessments are preserved with the new approach while they are not with the mapping. The third entry is comparable for the mapping and the new approach and significantly smaller than the expert’s estimate. The values of \( r_{47}^{e_{4, \text{map}}} \) and \( r_{47}^{e_{4, \text{new}}} \) differ more, but the correlation is rather small in general, so the effect on the model output of this difference should not be very strong. Both \( r_{57} \) and \( r_{67} \) are close to zero for both adjustment methods. In this case it is difficult to judge which of the adjustments yields more realistic results. Nonetheless, it seems easier to defend the new approach than to defend the map.

The results of expert \( e_5 \) are comparable to expert \( e_6 \), although something strange happens for \( r_{67} \). The mapping results in a negative correlation. Because the magnitude is close to zero, the effect may not even be visible in a model, but it certainly contradicts the experts estimate as it inverts the meaning of the relation the expert judged. Now the statement is that a high likelihood of the contractor showing a sign of in-availability to work decreases the likelihood that a task is performed with one or more errors, while the expert (and common sense) stated that it increases the likelihood that an error occurs. With the new approach, the rank correlations are truncated to the same values as for expert \( e_5 \). This is because both experts gave the same first estimate and have exceeded the maximum possible values for all other rank correlations.

Overall, there is no indication that the map produces more realistic results, although it may produce results closer to the original estimate in case of expert \( e_5 \). For expert \( e_6 \) it leads to a contradiction and for expert \( e_9 \) it modifies the original estimates unnecessarily strongly. Admittedly, the new approach is less than satisfying as well, however we do not have the opportunity to contact the experts again to find a better solution. Therefore, all expert assessments are adjusted with the new approach for the following analysis.

The conditional rank correlations that result from this new approach are given in column 11 of table 6.3. If all conditional rank correlations are 0.99, for example as for expert \( e_1 \), this means he always exceeded the bounds and they have been set to the maximum. The last four columns of the table attempt to show how inconsistent the various experts have been. Column 12 presents the absolute difference between the bound and the estimated rank correlation, if the estimation is out of bounds, which is termed as error, \( err \). Column 13 contains the relative error, \( err_{\text{relative}} \), and Column 14 the sum of relative errors for all six assessments of each expert to illustrate some sort of overall consistence of the expert. The last column assigns each expert a rank based on the value in column 14: 1 to the most consistent expert and 14 to the least consistent expert. Expert \( e_9 \) was ranked first. He/she was consistent for \( r_{17} - r_{47} \). His/her relative error for \( r_{37} \) is \( \approx 91\% \) and for \( r_{67} \) it is \( \approx 1323\% \). Even the best expert has exceeded
the bounds for the rank correlations by far. Expert $e_{12}$ is ranked last, whose estimates are always out of bounds. His last estimate, the one for $r_{67}$ is $\approx 692047\%$ larger than the bound. These numbers are extreme.

6.3 Weights

Having determined the marginal distributions and rank correlations for each expert, we need suitable weights to pool a decision maker. Model O has been based on $w_5 = 0.5093$ and $w_6 = 0.4907$. But since the calibration questions only asked for marginal distributions, the global weights derived from them may be appropriate to combine the marginals of the experts, but are not necessarily representative for the experts’ ability to assess dependences.

Table 6.4 presents once more the results of the elicitation of the marginals, but this time without decision maker optimization and a significance level $\alpha = 0$. The second last column contains a ranking according to the weights from the classical model and the last column contains the ranks that were already presented in table 6.3. Comparing both ranks leads to the conclusion that experts who perform well in assessing marginal distribution generally do not perform well in assessing rank correlation as well. Or they happened to disagree with the model structure strongly, a possibility discussed in chapter 8. Experts $e_5$ and $e_6$, who receive the highest weight based on their assessment of marginals, are ranked $10^{th}$ and $8^{th}$ in terms of consistency of correlations. On the other hand, experts $e_9$ and $e_4$ have been ranked $1^{st}$ and $2^{nd}$ based on their consistency in assessing correlations, but are only ranked $7^{th}$ and $9^{th}$ for their assessment of marginal distributions. Thus, it is not straightforward which weights should be used. We consider three different choices for weights. They are described and motivated below.

A: We use the weights $w_5 = 0.5093$ and $w_6 = 0.4907$, because it allows a comparison to the model built from previous analysis, model O.

B: We use equal weights for experts $e_6$ and $e_9$. The underlying idea is to combine the best expert of the marginal elicitation with the best expert of the dependence elicitation. However, we have chosen $e_6$ instead of $e_5$ for several reasons. First, they perform almost equally well on the calibration questions. They have identical calibration, just the informativeness $I(e_6) = 0.8222$ is negligibly smaller than $I(e_5) = 0.8534$. From this point of view it is practically equally preferable to combine with $e_5$ as to combine with $e_6$. However, $e_6$ is a little more consistent than $e_5$ in his/her dependence estimates. Also he/she gives almost the same ordering of variables as $e_9$, whom we trust most regarding dependence estimates. They only ranked the $V1$ and $V4$ the other way around (as first and second most important). Finally, figures 6.2 and 6.3 show that the marginal distributions of $e_6$ and $e_9$ are in most cases much closer together than the ones of $e_5$ and $e_9$. Only for variable $V2$ they are not. This means that the DM’s medians are usually closer to the medians of the experts for this combination. A consequence of the latter two aspects is that the dependence estimates of $e_9$ will be less affected by a combination with $e_6$ than with $e_5$, which we desire as the quality of their dependence estimates is
Table 6.4: Global weights without optimization and significance level $\alpha = 0$

<table>
<thead>
<tr>
<th></th>
<th>Calibration</th>
<th>Information Normalized weights Without DM</th>
<th>with DM</th>
<th>Marginals</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>0.5338</td>
<td>0.4866</td>
<td>0.08339</td>
<td>0.07743</td>
<td>6</td>
</tr>
<tr>
<td>$e_2$</td>
<td>0.00348</td>
<td>0.4561</td>
<td>0.0005095</td>
<td>0.0004731</td>
<td>13</td>
</tr>
<tr>
<td>$e_3$</td>
<td>0.00348</td>
<td>1.016</td>
<td>0.001135</td>
<td>0.001054</td>
<td>11</td>
</tr>
<tr>
<td>$e_4$</td>
<td>0.02651</td>
<td>0.9024</td>
<td>0.007681</td>
<td>0.007132</td>
<td>9</td>
</tr>
<tr>
<td>$e_5$</td>
<td>0.6876</td>
<td>0.8534</td>
<td>0.1884</td>
<td>0.1749</td>
<td>1</td>
</tr>
<tr>
<td>$e_6$</td>
<td>0.6876</td>
<td>0.8222</td>
<td>0.1815</td>
<td>0.1685</td>
<td>2</td>
</tr>
<tr>
<td>$e_7$</td>
<td>0.5405</td>
<td>0.6878</td>
<td>0.1193</td>
<td>0.1108</td>
<td>5</td>
</tr>
<tr>
<td>$e_8$</td>
<td>0.08763</td>
<td>1.089</td>
<td>0.03065</td>
<td>0.02846</td>
<td>8</td>
</tr>
<tr>
<td>$e_9$</td>
<td>0.2357</td>
<td>0.6681</td>
<td>0.05056</td>
<td>0.04694</td>
<td>7</td>
</tr>
<tr>
<td>$e_{10}$</td>
<td>0.00348</td>
<td>0.8325</td>
<td>0.0009299</td>
<td>0.0008635</td>
<td>12</td>
</tr>
<tr>
<td>$e_{11}$</td>
<td>0.01566</td>
<td>0.7833</td>
<td>0.003938</td>
<td>0.003656</td>
<td>10</td>
</tr>
<tr>
<td>$e_{12}$</td>
<td>0.6638</td>
<td>0.8263</td>
<td>0.176</td>
<td>0.1635</td>
<td>3</td>
</tr>
<tr>
<td>$e_{13}$</td>
<td>7.86E-005</td>
<td>1.02</td>
<td>2.32E-005</td>
<td>2.16E-005</td>
<td>14</td>
</tr>
<tr>
<td>$e_{14}$</td>
<td>0.5338</td>
<td>0.9099</td>
<td>0.1559</td>
<td>0.1448</td>
<td>4</td>
</tr>
<tr>
<td>DM</td>
<td>0.64</td>
<td>0.4651</td>
<td>0.07146</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

questionable. Equal weights are used, because no suitable scoring rule is available for the rank correlations.

C: We combine the marginals and rank correlations with different sets of experts. The marginals are combined with the weights $w_5 = 0.5093$ and $w_6 = 0.4907$ and the rank correlations with $w_4 = w_9 = 0.5$. The equal weights are chosen for the same reason as in B.

![Figure 6.2: Marginal distributions of experts $e_4$, $e_5$, $e_6$ and $e_9$ for the child node $V_7$](image-url)
Figure 6.3: Marginal distributions of experts $e_4$, $e_5$, $e_6$ and $e_9$ for the parent nodes $V1$ - $V6$
6.4 Conclusion

This chapter has criticized the computation of the marginal distributions, the computation of the rank correlations and the selection of weights. It has been found that the marginal distributions will be more accurate if they are calculated with Matlab instead of being exported from Excalibur. Moreover, with Matlab the range of the variables can be set to \([0, 1]\) which is more intuitive, given the definition of variables. It also may have benefits for the model use.

The results for the rank correlations are more problematic. Because the bounds for the rank correlations or the ratios of rank correlations have not been computed and discussed with the experts during the elicitation, it is not clear, whether the experts misjudged or simply did not agree with the structure of the BBN and the bounds that are implied by this structure. At the moment we do not have the possibility to contact the experts once more. Thus, assuming the experts misjudged, we modified their rank correlation estimates. Instead of the map, a new more defendable approach has been proposed: truncating the rank correlations at the highest valid value. This implies that, for most experts, only two or three rank correlations that are not close to zero. Hence, we assume that experts believe that only two or three factors have a significant influence on the contractor error probability. Moreover, each expert would believe in different factors. This seems quite unlikely as the factors have been selected based on a discussion with managers and are generally considered to be the most important contributors to human performance.

When trying to select appropriate weights, it became apparent that a method to properly score the experts performance in the dependence elicitation is lacking and needed. According to the present data, there is no support for the assumption that the experts perform equally well in eliciting marginals as in eliciting rank correlations. However, in order to perform well in dependence elicitation, the expert needs to be able to give accurate marginals, since the dependence estimate depends on the assumed underlying distribution.

The bottom line is, the present data situation is less than optimal. For now, three different sets of weights are chosen with the goal to investigate the consequence on the model output. The decision maker combination and the resulting models A, B and C are discussed in the following chapter.
7. Combining Experts

7.1 Combining Marginal Distributions

Pooling marginal distributions is rather straightforward (though one has to be careful to combine the densities and not the quantiles) and the individual steps will not be explained here. However, there seems to be an interesting issue with the combination in Excalibur. The software offers to calculate the distribution of the decision maker for specified sets of weights. Since we calculated the marginals for models A, B and C by hand, we also had to combine them by hand, but the DM distribution for model O has been exported from Excalibur and is shown in figure 7.1a as the magenta graph. Taking a closer look on how it relates to expert $e_5$’s and expert $e_6$’s distribution, we have to realize that it is different to the one we calculated for model A. Decision maker A is combined with the same weights as decision maker O. The marginal distributions of expert $e_5$ and $e_6$ in figures 7.1a and 7.1b differ only below the 5%-quantile and above the 95%-quantile for the reasons explained in the previous chapter. However, for $0.05 \leq V7 \leq 0.95$ the DM distribution obtained from Excalibur does not look like the DM distribution in figure 7.1b.

![Distribution Comparison](image)

(a) Distributions combined in Excalibur  
(b) Distributions combined in Matlab

Figure 7.1: Distributions of experts $e_5$, $e_6$ and the DM for $V7$

In [Cooke, 1991] it is proposed that the DM distribution is calculated as $f_{DM,V7}(v_7) = w_5 \cdot f_{e_5,V7}(v_7) + w_6 \cdot f_{e_6,V7}(v_7)$. Considering that the weights are almost equal (both are really
close to 0.5) the DM distribution should be in more or less in the middle of the distributions of the two experts, e.g. $F^{DM}(0.1) = 0.5093 \cdot F^{e5}(0.1) + 0.4907 \cdot F^{e6}(0.1)$, which clearly is not the case in figure 7.1a. Apparently, the creators of Excalibur have decided to use only the coordinates $F^{DM}(x_{q5}) = 0.05$, $F^{DM}(x_{q50}) = 0.50$ and $F^{DM}(x_{q95}) = 0.95$ that result from the combination and then to sample uniformly between those points to obtain the DM distribution. The underlying thought was to avoid false precision and to prevent mode users from thinking that combining more experts is better, because the resulting distribution would be more shapely. Indeed, we see that the distribution in figure 7.1b has four breakpoints instead of three1. For models A, B and C we have decided to adhere to the approach described in [Cooke, 1991], because it seems more accurate.

### 7.2 Combining Rank Correlations

In section 5.2.2 has been shown that a convex combination of rank correlation matrices will generally not satisfy the conditional independence statements implied by the graph. Since, operator model only has conditional dependence statements, it cannot violate them, as it happens in example 5.1. Maybe for this reason one has directly averaged rank correlations for model O. Admittedly, it is less obvious that such a combination yields incorrect results and also example 5.2 suggests that it only yields to minor inaccuracies. However, we will see in the following sections that the differences may be enormous in certain situations.

#### 7.2.1 Model O

To calculate the DM’s correlation matrix for model O, the same global weights as for the marginal distributions have been used. Hence, only expert $e_5$ and expert $e_6$ have a contribution. First, the entries in their matrices have been rearranged such that they correspond to the DM ordering (cf. table 6.2), yielding

$$
\Sigma^{e5} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & \rho^{e5}_{27} \\
0 & 1 & 0 & 0 & 0 & 0 & \rho^{e5}_{37} \\
0 & 0 & 1 & 0 & 0 & 0 & \rho^{e5}_{17} \\
0 & 0 & 0 & 1 & 0 & 0 & \rho^{e5}_{67} \\
0 & 0 & 0 & 0 & 1 & 0 & \rho^{e5}_{47} \\
\rho^{e5}_{27} & \rho^{e5}_{37} & \rho^{e5}_{17} & \rho^{e5}_{67} & \rho^{e5}_{47} & 1
\end{pmatrix}
$$

(7.1)

---

1The breakpoint at $F_{V7} = 0.5$ is difficult to see in this graph.
7.2. COMBINING RANK CORRELATIONS

and

\[
\Sigma^{e_6} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & \rho_{17}^{e_6} \\
0 & 1 & 0 & 0 & 0 & \rho_{47}^{e_6} \\
0 & 0 & 1 & 0 & 0 & \rho_{37}^{e_6} \\
0 & 0 & 0 & 1 & 0 & \rho_{27}^{e_6} \\
0 & 0 & 0 & 0 & 1 & \rho_{67}^{e_6} \\
\rho_{17}^{e_6} & \rho_{47}^{e_6} & \rho_{37}^{e_6} & \rho_{27}^{e_6} & \rho_{57}^{e_6} & \rho_{67}^{e_6}
\end{pmatrix}.
\]  

(7.2)

Second, the decision makers product moment correlation matrix has been calculated as

\[
\Sigma^{DM} = w_5 \cdot \Sigma^{e_5} + w_6 \cdot \Sigma^{e_6} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0.6387 \\
0 & 1 & 0 & 0 & 0 & 0.1976 \\
0 & 0 & 1 & 0 & 0 & 0.5104 \\
0 & 0 & 0 & 1 & 0 & 0.2530 \\
0 & 0 & 0 & 0 & 1 & 0.0276 \\
0.6387 & 0.1976 & 0.5104 & 0.2530 & 0.0276 & 1
\end{pmatrix},
\]  

(7.3)

where \( w_5 = 0.5093 \) and \( w_6 = 0.4907 \). The corresponding rank correlation matrix is

\[
R^{DM} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0.6208 \\
0 & 1 & 0 & 0 & 0 & 0.1891 \\
0 & 0 & 1 & 0 & 0 & 0.4928 \\
0 & 0 & 0 & 1 & 0 & 0.2422 \\
0 & 0 & 0 & 0 & 1 & 0.0263 \\
0.6208 & 0.1891 & 0.4928 & 0.2422 & 0.0263 & 1
\end{pmatrix}.
\]  

(7.4)

Third, the conditional product moment correlations \( \rho^{DM}_{27|1} \), \( \rho^{DM}_{37|12} \), \( \rho^{DM}_{47|123} \), \( \rho^{DM}_{57|1234} \) and \( \rho^{DM}_{67|12345} \) have been computed using the recursion (2.4) and assuming that conditional correlation equals partial correlation. Finally, they have been, together with \( \rho_{17}^{DM} \) transformed back to (conditional) rank correlations.

The rank correlations that we compute using this approach differ slightly from the ones used in the model O (see figure 7.4 in the next section). We computed

\[
\begin{pmatrix}
\rho_{17}^O \\
\rho_{27|1}^O \\
\rho_{37|12}^O \\
\rho_{47|123}^O \\
\rho_{57|1234}^O \\
\rho_{67|12345}^O
\end{pmatrix} = \begin{pmatrix}
0.6208 \\
0.2460 \\
0.6690 \\
0.4508 \\
0 \\
0.0550
\end{pmatrix}.
\]  

(7.5)

While the first three entries of the conditional rank correlation vector as well as the fifth appear to be comparable, the fourth one differs by 0.05 and the last one by around 0.1. However, comparing the DM’s rank correlation matrix in (7.4) and in figure 7.4b, which are very similar, suggests that the deviation stems from rounding errors that increase as they are propagated through the recursion (2.4). Finding the exact cause proved to be very difficult, because no written documentation of the previous project exists.
7.2.2 New Models

For models A, B and C the rank correlation estimates are combined following the protocol described in section 5.2.2. First we describe some issues we encountered regarding the discretization in Matlab, then we present the resulting DM correlations and finally we illustrate, using a result as example, the calculation steps that are taken to pool rank correlations.

While testing the relevant Matlab code, we realized that the strategy of section 5.2.2 is sensitive to numerical errors when computing rank correlations that are very close to the bounds. If the combination according to scheme (5.24) is performed assigning $w = 1$ to one expert and $w = 0$ to all other experts, the rank correlation matrices $\Sigma_{ei}$ and $\Sigma_{DM}$ should be equal. Comparing

$$
\Sigma_{ei} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0.299926 \\
0 & 1 & 0 & 0 & 0 & 0 & 0.707106 \\
0 & 0 & 1 & 0 & 0 & 0 & 0.639053 \\
0 & 0 & 0 & 1 & 0 & 0 & 0.005424 \\
0 & 0 & 0 & 0 & 1 & 0 & 0.040332 \\
0 & 0 & 0 & 0 & 0 & 1 & 0.000729 \\
0.299926 & 0.707106 & 0.639053 & 0.005424 & 0.040332 & 0.000729 & 1
\end{pmatrix}
$$

and

$$
\Sigma_{DM} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0.299925 \\
0 & 1 & 0 & 0 & 0 & 0 & 0.707105 \\
0 & 0 & 1 & 0 & 0 & 0 & 0.639051 \\
0 & 0 & 0 & 1 & 0 & 0 & 0.005424 \\
0 & 0 & 0 & 0 & 1 & 0 & 0.040332 \\
0 & 0 & 0 & 0 & 0 & 1 & 0.000729 \\
0.299925 & 0.707105 & 0.639051 & 0.005424 & 0.040332 & 0.000729 & 1
\end{pmatrix}
$$

for $w_{e4} = 1$ and $w_{ei} = 0$, for $i = 1, ..., 14$, $i \neq 4$, we see that they are not exactly equal, but differ on the 6th position after decimal point. However, this small difference has a huge impact on the resulting conditional rank correlations, in this case on the last entry:

$$
\begin{pmatrix}
r_{e4}^{17} \\
r_{e4}^{27} \\
r_{e4}^{37} \\
r_{e4}^{47} \\
r_{e4}^{57} \\
r_{e4}^{67}
\end{pmatrix} = \begin{pmatrix}
0.2875 \\
0.7251 \\
0.9978 \\
0.1273 \\
0.9998 \\
0.9900
\end{pmatrix}
$$

versus

$$
\begin{pmatrix}
r_{DM}^{17} \\
r_{DM}^{27} \\
r_{DM}^{37} \\
r_{DM}^{47} \\
r_{DM}^{57} \\
r_{DM}^{67}
\end{pmatrix} = \begin{pmatrix}
0.2875 \\
0.7251 \\
0.9978 \\
0.1272 \\
0.9983 \\
0.3141
\end{pmatrix}
$$

The reason is that in order to calculate $r_{67|12345}$ with the recursion, one has to divide by $\sqrt{1 - r_{57|1234}^2}$. If $r_{57|1234}$ is very close to one, extremely small differences generate a big difference for $r_{67|12345}$. Since most of the experts gave inconsistent estimates, many of their conditional rank correlations were set to 0.99 (see table 6.3). Then, their sampling order
(SO) was changed to the one of the DM (see 6.2) and new conditional rank correlations were calculated for this SO. As a result not all values are at the maximum value of 0.99 anymore, but a few (generally the later ones) are even higher causing these enormous differences for higher order conditional rank correlations. This problem can be mediated by assuming for the discretization in Matlab that the rank correlations can only take values in \([-0.95, 0.95]\) instead of \([-0.99, 0.99]\). Using the interval \([-0.95, 0.95]\) the unconditional rank correlations of the experts \(e_4, e_5, e_6\) and \(e_9\) are given below. Note that using 0.95 instead of 0.99 has quite some impact on the results. This is why the values are so different from the ones in table 6.3. Moreover, all values have changed places, because the sampling order corresponds now to the DM and not to the individual experts which makes them change places.

\[
\begin{pmatrix}
    r_{17}^{e_4} \\
    r_{27}^{e_4} \\
    r_{37}^{e_4} \\
    r_{47}^{e_4} \\
    r_{57}^{e_4} \\
    r_{67}^{e_4}
\end{pmatrix}
= \begin{pmatrix}
    0.2768 \\
    0.6902 \\
    0.6212 \\
    0.0246 \\
    0.0824 \\
    0.0074
\end{pmatrix}
\]

(7.6)

\[
\begin{pmatrix}
    r_{17}^{e_5} \\
    r_{27}^{e_5} \\
    r_{37}^{e_5} \\
    r_{47}^{e_5} \\
    r_{57}^{e_5} \\
    r_{67}^{e_5}
\end{pmatrix}
= \begin{pmatrix}
    0.5430 \\
    0.1602 \\
    0.7953 \\
    0.0143 \\
    0.0043 \\
    0.0478
\end{pmatrix}
\]

(7.7)

\[
\begin{pmatrix}
    r_{17}^{e_6} \\
    r_{27}^{e_6} \\
    r_{37}^{e_6} \\
    r_{47}^{e_6} \\
    r_{57}^{e_6} \\
    r_{67}^{e_6}
\end{pmatrix}
= \begin{pmatrix}
    0.7953 \\
    0.0478 \\
    0.1602 \\
    0.5430 \\
    0.0143 \\
    0.0043
\end{pmatrix}
\]

(7.8)

\[
\begin{pmatrix}
    r_{17}^{e_9} \\
    r_{27}^{e_9} \\
    r_{37}^{e_9} \\
    r_{47}^{e_9} \\
    r_{57}^{e_9} \\
    r_{67}^{e_9}
\end{pmatrix}
= \begin{pmatrix}
    0.4558 \\
    0.3988 \\
    0.4550 \\
    0.5697 \\
    0.1722 \\
    0.0514
\end{pmatrix}
\]

(7.9)

The correlation vectors of expert \(e_5\) and \(e_6\) exhibit the same numbers in a different order, because their estimates are very similar. In fact they have the same estimate for \(\hat{R}_1\) and since all other values were inconsistent, they have been truncated to the same upper bounds from the second variable onwards. The different order results from the different sampling order of the two experts.
The rank correlation matrices and the conditional rank correlations of the decision maker have been calculated for different choices of weights A, B and C and are given in (7.10), (7.11), (7.12), (7.13) and (7.14), (7.15). \( R^A \) denotes \( R^{DM} \) using the weights combination A and so on.

\[
R^A = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0.7927 \\
0 & 1 & 0 & 0 & 0 & 0 & 0.0814 \\
0 & 0 & 1 & 0 & 0 & 0 & 0.3147 \\
0 & 0 & 0 & 1 & 0 & 0 & 0.2628 \\
0 & 0 & 0 & 0 & 1 & 0 & 0.0176 \\
0 & 0 & 0 & 0 & 0 & 1 & 0.0303 \\
0.7927 & 0.0814 & 0.3147 & 0.2628 & 0.0176 & 0.0303 & 1 \\
\end{pmatrix}
\] (7.10)

\[
\begin{pmatrix}
 r^A_{17} \\
r^A_{27|1} \\
r^A_{37|12} \\
r^A_{47|123} \\
r^A_{57|1234} \\
r^A_{67|12345}
\end{pmatrix}
= \begin{pmatrix}
0.7927 \\
0.1378 \\
0.5428 \\
0.5483 \\
0.0442 \\
0.0759
\end{pmatrix}
\] (7.11)

\[
R^B = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0.6234 \\
0 & 1 & 0 & 0 & 0 & 0 & 0.1280 \\
0 & 0 & 1 & 0 & 0 & 0 & 0.3753 \\
0 & 0 & 0 & 1 & 0 & 0 & 0.5470 \\
0 & 0 & 0 & 0 & 1 & 0 & 0.0755 \\
0 & 0 & 0 & 0 & 0 & 1 & 0.0371 \\
0.6234 & 0.1280 & 0.3753 & 0.5470 & 0.0755 & 0.0371 & 1 \\
\end{pmatrix}
\] (7.12)

\[
\begin{pmatrix}
 r^B_{17} \\
r^B_{27|1} \\
r^B_{37|12} \\
r^B_{47|123} \\
r^B_{57|1234} \\
r^B_{67|12345}
\end{pmatrix}
= \begin{pmatrix}
0.6234 \\
0.1669 \\
0.4991 \\
0.8632 \\
0.2404 \\
0.1219
\end{pmatrix}
\] (7.13)

\[
R^C = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.5113 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & -0.8085 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & -0.3627 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & -0.8421 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & -0.8098 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & -0.8146 \\
-0.5113 & -0.8085 & -0.3627 & -0.8421 & -0.8098 & -0.8146 & 1
\end{pmatrix}
\] (7.14)
7.2. COMBINING RANK CORRELATIONS

\[ \begin{pmatrix}
    r_{17}^C \\
r_{27}^C \\
r_{37}^C \\
r_{47}^C \\
r_{57}^C \\
r_{67}^C
\end{pmatrix} =
\begin{pmatrix}
    -0.5113 \\
    -0.9652 \\
    -2.0959 \\
    2.1359i \\
    0.8374i \\
    0.6328i
\end{pmatrix}
\]  
(7.15)

Of course, there are differences to the original correlation matrix \( \text{7.4} \) and the conditional rank correlation vector \( \text{7.5} \).

Whether \( R^O \), \( R^A \) and \( R^B \) are comparable or not is disputable. On one hand we have to keep in mind that the data is obtained from expert opinion, which means we accept some imprecision. \([\text{Morales Nápoles et al., 2013}]\) propose the Hellinger distance to measure the distance between multivariate normal distributions. It is defined as

\[ d_H(N_1, N_2) = \sqrt{1 - \eta(N_1, N_2)}, \]  
(7.16)

where \( N_1(\mu_1, \Sigma_1) \) and \( N_1(\mu_2, \Sigma_2) \) are two multivariate normal distributions with mean vectors \( \mu_1, \mu_2 \) and covariance matrices \( \Sigma_1, \Sigma_2 \) and \( \eta \) is given by

\[ \eta(N_1, N_2) = \frac{\text{det}(\Sigma_1)^{1/4} \text{det}(\Sigma_2)^{1/4}}{\text{det}(\frac{1}{2} \Sigma_1 + \frac{1}{2} \Sigma_2)^{1/2}} \cdot \exp\{-\frac{1}{8} (\mu_1 - \mu_2)^T \frac{1}{2} \Sigma_1 + \frac{1}{2} \Sigma_2 (\mu_1 - \mu_2)\}. \]  
(7.17)

\( d_H = 0 \) if and only if \( \Sigma_1 = \Sigma_2 \). Because we assume the normal copula, the marginal distributions are transformed to standard normals and the exponential term becomes zero. The Hellinger distance reaches its maximum value, \( d_H = 1 \), if \( \text{det}(\Sigma_1) = 0 \) and \( \text{det}(\Sigma_2) = 1 \), or vice versa. In this case there is perfect dependence for the variables of one distribution, while the variables of the other distribution are mutually independent. The Hellinger distances computed with the product moment correlations of the three models are \( d_H(N_O, N_A) = 0.2304, \) \( d_H(N_O, N_B) = 0.3591 \) and \( d_H(N_A, N_B) = 0.3464 \). At first glance it may be surprising that \( d_H(N_O, N_A) \) is smaller than \( d_H(N_O, N_B) \), although the values of \( R^O \) and \( R^B \) seem more similar than the values of \( R^O \) and \( R^A \). However, \( R^O \) and \( R^A \) have in common that \( r_{17} > r_{37} > r_{47} > r_{27} > r_{57} > r_{57} \), while \( R^B \) has the second most and the third most variable are interchanged: \( r_{17} > r_{47} > r_{37} > r_{27} > r_{57} > r_{67} \). The dependence structure of model O and model B are equally similar as the dependence structures of model A and model B. More generally, all three Hellinger distances are closer to 0 than to 1. Considering that the data stems from experts one may conclude that the distribution from the models are somewhat comparable. However, we will see in the following section that the model output for each of the models differs strongly.

While combination A and B yield reasonable results, combination C yields only negative correlations, even though all experts indicated positive ones. It is hard to believe that the experts would agree with \( \text{7.15} \). Especially, since it contradicts the meaning of the variables. For instance, it indicates that better communication will decrease the contractor performance. As a matter of fact, the rank correlation matrix \( \text{7.14} \) is not positive definite. Neither is the corresponding product moment correlation matrix. As a consequence the conditional rank correlation vector \( \text{7.15} \) is also not valid. The magnitude of the third entry is larger than 1.
and the last three entries are complex numbers. Of course this combination scheme is discarded and in hindsight it is more than obvious that such a scheme may not give a meaningful result. Comparing the rank correlations of DM A and B, (7.11) and (7.13), we observe that only \( r_{27}^{A} \) and \( r_{27}^{B} \), \( r_{37}^{A} \) and \( r_{37}^{B} \), and \( r_{67}^{A} \) and \( r_{67}^{B} \) have a similar order of magnitude considering that they are based on expert judgment. The others differ notably.

Before discussing the results in detail, and especially what happened in combination C, the following paragraphs illustrate what calculations have been made using the example of \( r_{17}^{A} \). Expert \( e_{5} \) assessed \( r_{17}^{e_{5}} = 0.5430 \) and expert \( e_{6} \) assessed \( r_{17}^{e_{6}} = 0.7953 \) (cf. 7.7, 7.8). The corresponding product moment correlations are \( \rho_{17}^{e_{5}} = 0.5610 \) and \( \rho_{17}^{e_{6}} = 0.8090 \). To calculate the exceedance probabilities corresponding to these correlations with respect to the DM’s median, i.e. \( P_{1}^{e_{5}} = P(F_{V_{1}}^{e_{5}}(V_{1}) \geq F_{V_{1}}^{e_{5}}(V_{1}^{DM})) \) and \( P_{1}^{e_{6}} = P(F_{V_{1}}^{e_{6}}(V_{1}) \geq F_{V_{1}}^{e_{6}}(V_{1}^{DM})) \), we have to look at the marginal distributions for \( V_{1} \) and \( V_{7} \). The relevant regions are depicted in figure 7.2. The magenta graph represents the DM’s distribution. We see that \( F_{V_{1}}^{DM}(0.042) = 0.5 \) and \( F_{V_{7}}^{DM}(0.069) = 0.5 \). Now we have to find the quantiles of the two experts that correspond to \( v_{1} = 0.042 \) and to \( v_{7} = 0.069 \). They are (cf. figure 7.2)

\[
F_{V_{1}}^{e_{5}}(0.042) = 0.2100 \quad \text{and} \quad F_{V_{7}}^{e_{5}}(0.069) = 0.3400 \\
F_{V_{1}}^{e_{6}}(0.042) = 0.8060 \quad \text{and} \quad F_{V_{7}}^{e_{6}}(0.069) = 0.6620.
\]

Admittedly, \( w_{5} F_{V_{1}}^{e_{5}}(0.042) + w_{6} F_{V_{1}}^{e_{6}}(0.042) = 0.5093 - 0.2100 + 0.4907 - 0.8060 = 0.5025 \neq 0.5 \) and, similarly, \( w_{5} F_{V_{7}}^{e_{5}}(0.069) + w_{6} F_{V_{7}}^{e_{6}}(0.069) = 0.5093 - 0.3400 + 0.4907 - 0.6620 = 0.4980 \neq 0.5 \). This is caused by the discretization of the cumulative distribution of the experts. They do not always have an entry that corresponds exactly to 0.5 and, instead, the value closest to 0.5 is used. This is a reasonable approximation, since the experts are unlikely to be accurate up to two decimal places. Moreover, the adjustment of the bounds has been much more invasive.
7.2. COMBINING RANK CORRELATIONS

Using the normal copula yields for expert \( e_5 \)

\[
P^{e_5}_{1*} = P\{ F^{e_5}_{V_7}(V_7) \geq 0.3400 \mid F^{e_5}_{V_1}(V_1) \geq 0.2100 \}
\]

\[
= \frac{\int_0^\infty \int_0^\infty \phi_{\Sigma_{17}}(\tilde{v}_1, \tilde{v}_7) d\tilde{v}_1 d\tilde{v}_7}{\Phi^{-1}(0.3400) \Phi^{-1}(0.2100)}
\]

\[
= 0.7445,
\]

(7.18)

where \( \Sigma_{17} = \begin{pmatrix} 1 & \rho_{17}^{e_5} \\ \rho_{17}^{e_5} & 1 \end{pmatrix} \) and \( \Sigma_1 = 1 \) and for expert \( e_6 \)

\[
P^{e_6}_{1*} = P\{ F^{e_6}_{V_7}(V_7) \geq 0.6620 \mid F^{e_5}_{V_1}(V_1) \geq 0.8060 \}
\]

\[
= \frac{\int_0^\infty \int_0^\infty \phi_{\Sigma_{17}}(\tilde{v}_1, \tilde{v}_7) d\tilde{v}_1 d\tilde{v}_7}{\Phi^{-1}(0.6620) \Phi^{-1}(0.8060)}
\]

\[
= 0.8548,
\]

(7.19)

where \( \Sigma_{17} = \begin{pmatrix} 1 & \rho_{17}^{e_6} \\ \rho_{17}^{e_6} & 1 \end{pmatrix} \) and \( \Sigma_1 = 1 \). The DM’s exceedance probability

\[
P^{DM}_{1} = P\{ F^{DM}_{V_7}(V_7) \geq F^{DM}_{V_1}(v_{1,0.5}) \mid F^{DM}_{V_1}(V_1) \geq F^{DM}_{V_1}(v_{1,0.5}) \}
\]

is obtained by combining the above results:

\[
P^{DM}_{1} = w_5 \cdot P^{e_5}_{1*} + w_6 \cdot P^{e_6}_{1*}
\]

\[
= 0.5093 \cdot 0.7445 + 0.4907 \cdot 0.8548
\]

\[
= 0.7986.
\]

(7.21)

Figure 5.2 indicates that this exceedance probability corresponds to a high positive correlation between \( V_1 \) and \( V_7 \). Solving

\[
0.7986 = \frac{\int_0^\infty \int_0^\infty \phi_{\Sigma_{17}^{DM}}(\tilde{v}_1, \tilde{v}_7) d\tilde{v}_1 d\tilde{v}_7}{\Phi^{-1}(0.5) \Phi^{-1}(0.5) \Phi^{-1}(0.5)}
\]

for \( \rho_{17}^{DM} \) numerically in Matlab yields 0.8064, which corresponds to \( r_{17}^{DM} = 0.7927 \) as given by (7.11). Note that this correlation is almost equal to \( r_{17}^{e_6} \). It is not obvious that it is a combination of \( r_{17}^{e_5} \) and \( r_{17}^{e_6} \). This illustrates clearly that weighted averaging of rank correlations is not possible, also if no conditional independence statements are violated, and, especially if the expert’s distributions differ from another, the resulting DM’s rank correlation may be surprising. This explains why the correlations of model O differ so much from the correlations
of model A, which have been combined with the same weights. Nonetheless, for the present data we see that if we compare the Hellinger distances the dependence structure of model O differs to the same degree from the dependence structure of model A or B as the dependence structures from model A and B differ from each other. Thus, model O turns out surprisingly well considering how it has been obtained.

Having understood the steps to obtain the decision maker’s rank correlation, it is easily traceable why the correlations resulting from combination scheme C are all negative. We use again the example of $V_1$ and $V_7$. The marginals of decision maker C are identical to the marginals of decision maker A and as depicted in figure 7.2. The exceedance probability, $P^e_{17}$, the one that is linearly combined to the exceedance probability of the DM, is determined by three values: $r_{17}^e$, $F^e_{V_1}(v_{1,q_{50}})$ and $F^e_{V_7}(v_{7,q_{50}})$. While $r_{17}^e$ is fixed for each expert, the two quantiles depend on the choice of the decision maker and the resulting decision maker distributions. Consider the following hypothetical example. Assume $r_{17}^e = 0.5$ and, for simplicity, $F^e_{V_1}(v_{1,q_{50}}) = 0.5$. We denote $q_7 \triangleq F^e_{V_7}(v_{7,q_{50}})$. The blue line in figure 7.3 shows the relation between $q$ and the exceedance probability $P^e_{17}$. If the decision maker’s distribution for $V_7$ is such that its median, $v_{7,q_{50}}$, corresponds to a quantile $q = 0.1$ in expert $e_7$’s distribution, the resulting exceedance probability, $P^e_{17}$, will be close to 1. If on the other hand the median corresponds to quantile $q = 0.9$, the resulting $P^e_{17}$, will be close to 0. Moreover, $P^e_{17} \leq 0.5$ for $q_7 \geq 0.66$ and $P^e_{17} \geq 0.5$ for $q_7 < 0.66$. If the combined exceedance probability is smaller than 0.5, a negative correlation and, otherwise, a positive correlation results. If, instead, we assume $F^e_{V_7}(v_{7,q_{50}}) = 0.5$ and denote $q_1 \triangleq F^e_{V_1}(v_{1,q_{50}})$, a different graph results. The relationship is depicted in figure 7.3 as the magenta line. In this case all attainable exceedance probability values are $\geq 0.5$. Only positive correlations can result. Of course, if neither $q_1$ nor $q_2$ are fixed, the relationship is more complicated. However, the purpose of this example is to demonstrate that the formulas above enable us to manipulate the correlations of two experts such that the resulting DM’s correlation can take a wide range of values.
The main idea of the combination process is to translate the experts rank correlation estimates to the exceedance probability that corresponds to the DM’s medians. In scheme C we use experts $e_5$ and $e_6$ to calculate the DM’s marginal distributions and experts $e_4$ and $e_9$ to calculate the DM’s rank correlations. In this case, we translate the dependence estimates of $e_4$ and $e_9$ to the marginals consented by $e_5$ and $e_6$. Actually, we translate them to a context to which neither $e_4$ nor $e_9$ have "agreed" on, to a context that has no relation with the opinions of $e_4$ and $e_9$. As illustrated in the above example, it is possible to combine $e_4$’s and $e_9$’s estimates to a negative DM’s correlation. Moreover, since there is no rational consensus between $e_4$ and $e_9$, it is not provided that the resulting correlation matrix is consistent. Indeed, $R^C$ is not positive definite. Figure 7.2 exemplifies this. While the decision maker’s distributions lie in between the distribution of $e_5$ and $e_6$, they lie below the distributions of $e_4$ and $e_9$. The quantiles of $e_4$ and $e_9$ corresponding to the DM’s median are both higher than 0.5. From figure 7.3 we know that high values of $q_7$ decrease the exceedance probability value, given the same correlation of course, while high values of $q_1$ increase the exceedance probability value. A detailed calculation in the above manner indicates that the effect of $q_7$ is stronger. It yields a negative DM correlation, $r_{17} = -0.5113$.

In hindsight, it is very clear that scheme C is more of an arbitrary manipulation of numbers than a combination based on a rational consensus. Here, the resulting correlation matrix is not valid, which makes it obvious that the idea was flawed and forces us to discard this combination. However, this does not need to be the case. In general, it is not possible to combine the marginals and the dependence structure with different weights, although the resulting matrix may be positive definite. There may be applications where the results of such a combination may be "useful", but this kind of approach is certainly not robust and neither mathematically rigorous.

### 7.2.3 Resulting Models

This section presents the BBNs resulting from scheme A and B and compares them to the model O. They are depicted in figure [7.4](#), figure [7.5](#) and figure [7.6](#). For easy reference, the corresponding correlation matrices are again given below the respective figures. The model specifications differ notably from each other as well as from the previously quantified model 7.4.

All models have in common that the correlations between $V_5$ and $V_7$ as well as $V_6$ and $V_7$ are close to zero (see the correlation matrices [7.4b](#), [7.5b](#), [7.6b](#)). Namely, $r_{57}^O = 0$, $r_{57}^A = 0.0176$ and $r_{57}^B = 0.0755$ while $r_{67}^O = 0.0703$, $r_{67}^A = 0.0303$ and $r_{67}^B = 0.0371$. They also agree that $V_1$ has the most influence on $V_7$, that is $r_{17}$ has the highest value: $r_{17}^O = 0.62$, $r_{17}^A = 0.793$ and $r_{17}^B = 0.623$. Not surprisingly, model O and model A agree on the ranking of the variables. They state $r_{17} > r_{37} > r_{47} > r_{27} > r_{67} > r_{57}$. After all they have been constructed with the same weights. However, surprising is that the first three correlations of model O are much more similar to the ones of model B than the ones of model A, although model B ranks the variables according to $r_{17} > r_{47} > r_{37} > r_{27} > r_{57} > r_{67}$. This is a likely coincidence, given the way model O has been obtained.
Also the marginal distributions of model A and model B vary significantly. For example the mean of variable \( V1 \) is \( \mu^A(V1) = 0.1 \) in model A and \( \mu^B(V1) = 0.05 \) in model B, which is half of the mean in model A. For variables \( V1, V3, V5, V6 \) and \( V7 \) the means are higher in model A. For variables \( V2 \) and \( V4 \) they are lower. Because incidents and accidents happen very rarely, most of the probability mass is on very small frequencies. In all four models variable \( V4 \), i.e. a work instruction is not available, accurate, up-to-date etc., occurs an average most often: \( \mu^O(V1) = 0.222 \), \( \mu^A(V4) = 0.185 \) and \( \mu^B(V4) = 0.276 \).

![Graph with the nodes as histograms and (conditional) rank correlations in sampling order](image1)

(a) Graph with the nodes as histograms and (conditional) rank correlations in sampling order

![Correlation Matrix](image2)

(b) Corresponding correlation matrix

Figure 7.4: Model O: Human performance model as quantified previously
7.2. COMBINING RANK CORRELATIONS

(a) Graph with the nodes as histograms and (conditional) rank correlations in sampling order

(b) Corresponding correlation matrix

Figure 7.5: Human performance model A: global weights $w_5$ and $w_6$
Figure 7.6: Human performance model $B$: equal weights $w_6$ and $w_9$
Of course, using the different models will lead to different conclusions. For instance, figures 7.7, 7.8 and 7.9 show the BBNs conditioned on $V_7 = 1$, which can be interpreted as a task is being performed with one or more errors that hamper the operation. Model O cannot be conditioned on this value because the upper value of the intrinsic range is 0.7699. Therefore, it has been conditioned on this value instead. In this scenario, the on average most likely contributant to the value of $V_7$ is $V_1$ according to model O and model A. Differently, $V_4$ is the most likely contributant according to model B. This is logical considering that these variables had among of the highest correlations with $V_7$ and, in the case of model B, the mean of the variable in the unconditioned state was quite high. Table 7.1 summarizes how much the value of the variables increased on average. Naturally, the variables with the highest correlation to $V_7$ increased the most: $\mu^O(V_1)$ is 262% higher, $\mu^A(V_1)$ is 714% higher and $\mu^B(V_1)$ is 867% higher. Of course, the values for model O are only restrictively comparable to the other, because the variables cannot be conditioned on 1. Considering that a task has been performed with at least one or more errors, model A diagnoses that, on average, with a likelihood of 81.4% the contractor had insufficient knowledge or experience and model B diagnoses that this likelihood is 48.3%.

Table 7.1: Overview of the increase of the mean $\mu$, $\Delta \mu(V)$, when conditioning on $V_7 = 1$ for model A, B and C and on $V_7 = 0.7699$ for model O

<table>
<thead>
<tr>
<th></th>
<th>Model O</th>
<th>Model A</th>
<th>Model B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \mu(V_1)$</td>
<td>262%</td>
<td>714%</td>
<td>867%</td>
</tr>
<tr>
<td>$\Delta \mu(V_2)$</td>
<td>63%</td>
<td>211%</td>
<td>55%</td>
</tr>
<tr>
<td>$\Delta \mu(V_3)$</td>
<td>152%</td>
<td>28%</td>
<td>300%</td>
</tr>
<tr>
<td>$\Delta \mu(V_4)$</td>
<td>79%</td>
<td>115%</td>
<td>168%</td>
</tr>
<tr>
<td>$\Delta \mu(V_5)$</td>
<td>0%</td>
<td>5%</td>
<td>22%</td>
</tr>
<tr>
<td>$\Delta \mu(V_6)$</td>
<td>24%</td>
<td>10%</td>
<td>19%</td>
</tr>
</tbody>
</table>

Given that this human performance model is supposed to be integrated in a larger risk model with additional arcs influencing $V_1$ to $V_6$ and arcs from $V_7$ influencing other variables, the discrepancies between the models presented here will have even larger effects. Also when conditioning one of the parent nodes, for example, if $V_1 = 1$, the impact on the mean of $V_7$ varies strongly between the models. Let’s consider only model A and model B, since the predictions from model O are not comparable to the other ones, because of the different intrinsic ranges of the variables. In the unconditioned case they have $\mu^A(V_7) = 0.104$ and $\mu^C(V_7) = 0.0741$, which is somewhat comparable. However, in the conditioned case $\mu^A(V_7) = 0.829$ and $\mu^C(V_7) = 0.502$. The predicted $\mu^A(V_7)$ is around 65% higher than the predicted $\mu^C(V_7)$. Model A predicts that on average, 829 out of 1000 tasks are performed with a significant error, if executed by a contractor who has insufficient knowledge or experience. For model B it is on average 502 out of 1000 tasks. In both models it is under these circumstances more likely that an error occurs than that it does not. This result as well as the results for conditioning on other parent nodes are summarized in table 7.1. For all other parent nodes it is, on average, still less likely that an error occurs than that it does not. Model A predicts higher mean values for $V_1$ and $V_2$, while model B predicts higher mean values for $V_3$ and $V_4$. The differences for $V_5$ and $V_6$ are negligible.
Figure 7.7: Human performance model O conditioned on $V7 = 0.7699$

Figure 7.8: Human performance model A conditioned on $V7 = 1$
7.2. COMBINING RANK CORRELATIONS

Figure 7.9: Human performance model B conditioned on \( V_7 = 1 \)

Table 7.2: Comparison of the effect on \( V_7 \) of conditioning all parent nodes individually in model A and model B

<table>
<thead>
<tr>
<th></th>
<th>( \mu^A(V_7) )</th>
<th>( \mu^B(V_7) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>unconditioned</td>
<td>0.104</td>
<td>0.0741</td>
</tr>
<tr>
<td>( V_1 = 1 )</td>
<td>0.829</td>
<td>0.502</td>
</tr>
<tr>
<td>( V_2 = 1 )</td>
<td>0.253</td>
<td>0.199</td>
</tr>
<tr>
<td>( V_3 = 1 )</td>
<td>0.135</td>
<td>0.261</td>
</tr>
<tr>
<td>( V_4 = 1 )</td>
<td>0.232</td>
<td>0.437</td>
</tr>
<tr>
<td>( V_5 = 1 )</td>
<td>0.11</td>
<td>0.096</td>
</tr>
<tr>
<td>( V_6 = 1 )</td>
<td>0.114</td>
<td>0.085</td>
</tr>
</tbody>
</table>
7.3 Conclusion

For model O both the marginals and the rank correlations have not been combined in a proper way. While the pooling of the marginals could be termed as inaccurate, the pooling of rank correlations generally results in wrong numbers. For the present case study, the overall dependence structure turns out to be somewhat comparable to the one of model A and model B, but strictly speaking obtaining rank correlations does not preserve the meaning of the original statements of the experts. Model A and model B have been combined properly with the same weights for marginals and rank correlations. For this application using the same weights is indispensable as the absurd results for model C clarified.

It is difficult to say which of the models, A or B, may be more useful as part of a larger risk model. The idea behind scheme B, combining the best expert from the marginal elicitation and the best expert from the dependence elicitation, seems more reasonable than the idea behind scheme A, using optimal global weights from the marginal elicitation. To be precise, the optimal global weight would have been $w_5 = 1$ and not $w_5 = 0.5093$ and $w_4 = 0.4907$ as explained in section 6.1. However, the best performing expert for the dependence structure has been determined on the basis of his/her consistency in the estimates for the given model structure. However, taking into account that the vast majority of experts was inconsistent, we wonder whether they lack knowledge or experience in estimating dependencies or whether they actually disagree with the model structure, in which case this kind of performance measure would be meaningless.

At the moment we unfortunately do not have the opportunity to discuss this with the experts. For this reason the following chapter explores what kind of model structure the experts might have had in mind, if we assume that they disagree with the present model structure and that they gave their estimates deliberately the way they are: inconsistent with the present structure, but accurate in the sense that they exactly express their true beliefs. The chapter thereafter, chapter 9, gives some guidelines for model validation. However, they do require feedback from the experts, at least experts $e_5$, $e_6$ and $e_9$. Following these guidelines may facilitate choosing between model A or B or to come to the conclusion that neither of them is useful.
8. Assumption of Dependent Parents

We already mentioned that there are two explanations for the out-of-bounds estimates of the experts: (1) The experts were not aware of the bounds, because they had insufficient experience or knowledge of giving subjective probabilities. This has been assumed in the previous sections, where different approaches to modify the estimates in an as least invasive as possible manner have been applied and the model has been quantified in the originally intended structure. (2) The experts disagree with the model structure. Mainly, they do not believe that the parent nodes are independent of each other.

This section looks at the simplest case of explanation (2): two out of the six parent nodes are dependent. More specifically, we assume that the experts’ estimates are accurate, i.e. we fix the experts estimates for the parent-child rank correlation, and infer an interval of the rank correlation between those two parent nodes that the experts considered as having the strongest impact on the child node $V_7$, the contractor performance. This is illustrated with the blue connection between $X_1$ and $X_2$ in figure 8.1. An arc is not drawn, because we have no information on directionality of the influence. Recall that $X_1$ denotes the variable that an expert ranked most influential on the child. Hence $X_1$ denotes a different variable for each experts, e.g. it denotes $V_3$ (commitment) for expert $e_5$ or $V_4$ (quality of procedure) for expert $e_9$. Similarly, $X_2$ is the second most influential variable and so on. $X_7$ is the child and always corresponds to $V_7$.

![Figure 8.1](image-url)
In the previous section, we used the following equation to determine the bounds for $\rho_{27}$

$$\rho_{27min/max} = \rho_{27|1min/max} \sqrt{(1 - \rho_{12}^2)(1 - \rho_{17}^2)} + \rho_{12}\rho_{17}, \quad (8.1)$$

where $\rho_{17}$ is known as the expert’s previous estimate. Setting $\rho_{12} = 0$, because of the assumption that the parents are independent, yielded

$$\rho_{27min/max} = \rho_{27|1min/max} \sqrt{(1 - \rho_{17}^2)} \quad (8.2)$$

(cf. (5.22)).

The plots in figure 8.2 show the relationship between $\rho_{12}$ and $\rho_{27}$ based on equation (8.1). Plot 8.2a assumes $\rho_{17} = 0.3$ to be the expert’s previous estimate, plot 8.2b assumes $\rho_{17} = 0.5$, plot 8.2c assumes $\rho_{17} = 0.7$ and plot 8.2d assumes $\rho_{17} = 0.9$. The blue graph indicates the upper bound for $\rho_{27}$ and the magenta graph indicates the lower bound for $\rho_{27}$. They are computed with $\rho_{27|1} = 1$ and $\rho_{27|1} = -1$, respectively. Hence, the area inside the ellipse represents all the possible combinations of values of $\rho_{12}$ and $\rho_{27}$. A comparison of the four plots shows, as expected, that a high correlation between the first parent node and the child node induces strict bounds for the correlation between the second parent node and the child node. In other words, the higher $\rho_{17}$, the smaller is the allowed interval for $\rho_{27}$.

As the majority of experts gave an estimate between $\rho_{17} = 0.8$ and $\rho_{17} = 0.95$, plot 8.2d gives a good indication of the order of magnitude of the bounds in the human performance model. The accurate bounds for each expert are given in table 6.3 in the previous section. The allowed interval for $\rho_{27}$ is only symmetric for $\rho_{12} = 0$. In plot 8.2d it is $[-0.44, 0.44]$. At this point it is also largest. It diminishes and shifts towards higher values for $\rho_{12} > 0$, e.g. the interval is $[0.07, 0.83]$ for $\rho_{12} = 0.5$, and towards lower values for $\rho_{12} < 0$, e.g. $[-0.07, -0.83]$ for $\rho_{12} = -0.5$. For $\rho_{12} = 1$ and $\rho_{12} = -1$ the value for $\rho_{27}$ is prescribed, it is $\rho_{27} = \rho_{17} = 0.9$ and $\rho_{27} = -\rho_{17} = -0.9$, respectively.

The plots indicate that if the experts truly believe that $X_1$ and $X_2$ are not independent, it makes sense that they give estimates which are outside the bounds calculated with $\rho_{12} = 0$. If they believe in a positive correlation between $X_1$ and $X_2$ they will exceed the upper value of the originally calculated bounds. Now, fixing both $\rho_{17}$ and $\rho_{27}$ we can compute $\rho_{12max}$ and $\rho_{12min}$, which correspond to $\rho_{27|1} = 1$ and $\rho_{27|1} = -1$, respectively. The interval $[\rho_{12max}, \rho_{12min}]$ contains the value that a consistent expert would attribute to $\rho_{12}$. In other words, the bounds for $\rho_{12}$ give an indication for how dependent an expert considers $X_1$ and $X_2$. Solving

$$\rho_{27} = \rho_{27|1} \sqrt{(1 - \rho_{17}^2)(1 - \rho_{12}^2)} + \rho_{12}\rho_{17} \quad (8.3)$$

for $\rho_{12}$ yields two solutions

$$\rho_{12,\pm} = \frac{-\rho_{27}\rho_{17} \pm |\rho_{27|1}| \sqrt{\rho_{27|1}^2 - 2\rho_{17}^2\rho_{27|1}^2 - \rho_{27}^2 + \rho_{17}^4\rho_{27|1}^2 + \rho_{27}^4\rho_{17}^2 + \rho_{17}^4 - \rho_{17}^4}}{-\rho_{27}^2\rho_{27|1}^2 - \rho_{17}^4}, \quad (8.4)$$

Because $\rho_{27|1}$ is squared in the solving process, the information whether it was $+1$ or $-1$ is lost. Nonetheless, figure 8.3 shows that, in the region $-\rho_{17} \leq \rho_{27} \leq \rho_{17}$, the relation of $\rho_{12,+}$
and $\rho_{27}$ is the same as the relation of $\rho_{27,\text{min}}$ and $\rho_{12}$ in figure 8.2d, while the relation of $\rho_{12,\text{+}}$ and $\rho_{27}$ is the same as the relation of $\rho_{27,\text{max}}$ and $\rho_{12}$. Figure 8.4a and a small calculation give more insight. Table 8.1 contains different values for $\rho_{12,\text{+}}$ and $\rho_{12,\text{−}}$ for $\rho_{17} = 0.9$ and $\rho_{27} = 0.8$ calculated with (8.4). The results do not differ for $\rho_{27}|_1 = 1$ and $\rho_{27}|_1 = -1$. Figure 8.4a shows that $\rho_{12,\text{+}}$ and $\rho_{12,\text{−}}$ are axial symmetric around $\rho_{27}|_1 = 0$. This is because (8.4) contains the absolute value of $\rho_{27}|_1$. However, if we plug the values from table 8.1 back into (8.3) contradictions arise, because it contains $\rho_{27}|_1$ and not the absolute value $|\rho_{27}|_1$:

$$\rho_{27}(\rho_{12} = 0.4585, \rho_{27}|_1 = -1, \rho_{17} = 0.9) = 0.0253 \neq 0.8 \quad (8.5)$$
$$\rho_{27}(\rho_{12} = 0.4585, \rho_{27}|_1 = 1, \rho_{17} = 0.9) = 0.8 \quad (8.6)$$
$$\rho_{27}(\rho_{12} = 0.9815, \rho_{27}|_1 = -1, \rho_{17} = 0.9) = 0.8 \quad (8.7)$$
$$\rho_{27}(\rho_{12} = 0.9815, \rho_{27}|_1 = 1, \rho_{17} = 0.9) = 0.9668 \neq 0.8 \quad (8.8)$$

This indicates that the magenta graph is invalid left of the vertical line at $\rho_{27}|_1 = 0$, while the
Figure 8.3: Relationship between the bounds for $\rho_{12}$ and $\rho_{27}$ for a precedent assessments of $\rho_{17} = 0.9$ based on equation (8.4)

Table 8.1: Values of $\rho_{12,+}$ and $\rho_{12,-}$ for $\rho_{27|1} = 1$ and $\rho_{27|1} = -1$ calculated with $\rho_{17} = 0.9$ and $\rho_{27} = 0.8$

|                | $\rho_{27|1} = 1$ | $\rho_{27|1} = -1$ |
|----------------|-------------------|-------------------|
| $\rho_{12,+}$  | 0.4585            | 0.4585            |
| $\rho_{12,-}$  | 0.9815            | 0.9815            |

The blue graph is invalid right of it. Hence, for $-\rho_{17} \leq \rho_{27} \leq \rho_{17}$,

$$\rho_{12} = \begin{cases} 
\rho_{12,+} & \text{for } \rho_{27|1} \geq 0 \\
\rho_{12,-} & \text{for } \rho_{27|1} < 0 
\end{cases}$$

which is plotted in figure [8.4b]. The minimum $\rho_{12_{\text{min}}}$ is attained for $\rho_{27|1} = -1$, while the maximum $\rho_{12_{\text{max}}}$ is attained for $\rho_{27|1} = 1$.

Outside the region $-\rho_{17} \leq \rho_{27} \leq \rho_{17}$ we observe that $\rho_{12}$ is not a real number in $[-1, 1]$ for all values $\rho_{27|1} \in [-1, 1]$. An example is plotted in figure [8.5]. Here $\rho_{27}$ is chosen to be 0.95. This can be explained with theorem [2.4] which states that all rank correlations associated with the arcs according to protocol [2.24] are algebraically independent. Let consider only variables $X_1$, $X_2$ and $X_7$ as depicted in figure [8.6]. There are two different possible sampling orders, i.e. two different rank correlation specifications. Note that it does not matter, if the arc points from $X_1$ to $X_2$ or the opposite. The reader may choose which arc the blue edge represents. One possible specification consists of $r_{12}$, $r_{17}$ and $r_{27|1}$, the other one of $r_{12}$, $r_{27}$ and $r_{17|2}$. We have used the first so far, because it falls into line with the ordering in the previous sections. For this ordering, $r_{12}$, $r_{17}$ and $r_{27|1}$ are algebraically independent and $\rho_{12}$, $\rho_{17}$ and $\rho_{27|1}$ can take any value in $[-1, 1]$. The rank correlation $\rho_{27}$, on the other hand, cannot take any value in $[-1, 1]$, but results from the former values. It is calculated with the recursive formula [2.4]. By fixing $\rho_{17} = 0.9$, as before, since it is the first estimate of each expert, we can plot a three dimensional plane for $\rho_{27}$ as a function of $\rho_{12}$ and $\rho_{27|1}$ (figure [8.7]). A value of $\rho_{27} > \rho_{17}$, for example $\rho_{27} = 0.95$, which approximately corresponds to a line within the
Figure 8.4: Relationship between the bounds for $\rho_{12}$ and $\rho_{27|1}$ for $\rho_{17} = 0.9$ and $\rho_{27} = 0.8$

Figure 8.5: $\rho_{17} = 0.9$ and $\rho_{27} = 0.95$ (Within $-0.7 \leq \rho_{27|1} \leq 0.7$ the values for $\rho_{12}$ have imaginary parts.)

Figure 8.6: highest back stripe of the graph, can only be achieved by a very small range of values of $\rho_{12}$ and $\rho_{27|1}$. This is in agreement with figure 8.5. Hence, fixing arbitrary values for $\rho_{17}$, $\rho_{27}$ and $\rho_{27|1}$ in $[-1, 1]$, will not always result in a valid value for $\rho_{12}$. Especially, $\rho_{27}$ and $\rho_{12}$
cannot be chosen arbitrarily. Figure 8.8a, which is a view along the x-axis of figure 8.7 with the y-axis extending horizontally and the z-axis extending vertically, illustrates clearly that there is only a small range of values for $\rho_{12}$ that can realize a certain $\rho_{27}$. On the contrary we can fix arbitrary values for $\rho_{17}$ and $\rho_{27|1}$ in $[-1, 1]$ and $\rho_{27}$ in $[-|\rho_{17}|, |\rho_{17}|]$. In figure 8.8b, a view along the y-axis of figure 8.7 the highest value for $\rho_{27}$ that can be realized by all values of $\rho_{27|1}$ in $[-1, 1]$ is 0.9, while the lowest is −0.9. Similarly, if we set $\rho_{17} = -0.3$, the highest value for $\rho_{27}$ that can be realized by all values of $\rho_{27|1}$ in $[-1, 1]$ is 0.3, while the lowest is −0.3 (see figures 8.9, 8.10a, 8.10b).

Since all experts ranked the variables according to the strongest influence on the child it
Figure 8.9: The relationship between $\rho_{27}$, $\rho_{12}$ and $\rho_{27|1}$ for $\rho_{17} = -0.3$

(a) View along the x-axis with the y-axis extending horizontally and the z-axis extending vertically

(b) View along the x-axis with the y-axis extending horizontally and the z-axis extending vertically

Figure 8.10: Views along the x-axis and the y-axis of figure 8.9
holds that $0 \leq \rho_{27} \leq \rho_{17}$ and we never run into trouble when computing bounds for $r_{12}^e$ for each of the experts. $r_{12,\text{min}}^e$ is calculated from equation 8.9 and, of course, Pearson’s transformation with $r_{17}^e$, $r_{27}^e$ and $r_{27|1} = -1$, whereas $r_{12,\text{max}}^e$ is calculated with $r_{27|1} = 1$ instead. The results are given in table 8.2. The indices of the correlation refer to the DM’s ordering of the variables, for instance $r_{12}^e$ corresponds to $r_{25DM}^e$. Only experts $e_4$, $e_9$ and $e_{13}$ allow their respective $X_1$ and $X_2$ to be independent. All other experts indicate a positive dependence. And even for expert $e_9$ the interval is not symmetric around zero. Not surprisingly experts $e_4$ and $e_9$ are those who have been ranked first and second most consistent in assessing rank correlation.

<table>
<thead>
<tr>
<th>Expert</th>
<th>Relation</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>$r_{25DM}^e$</td>
<td>0.2467</td>
<td>0.9978</td>
</tr>
<tr>
<td>$e_2$</td>
<td>$r_{14DM}^e$</td>
<td>0.7195</td>
<td>0.9920</td>
</tr>
<tr>
<td>$e_3$</td>
<td>$r_{12DM}^e$</td>
<td>0.0880</td>
<td>0.9708</td>
</tr>
<tr>
<td>$e_4$</td>
<td>$r_{14DM}^e$</td>
<td>-0.0791</td>
<td>0.9952</td>
</tr>
<tr>
<td>$e_5$</td>
<td>$r_{23DM}^e$</td>
<td>0.1915</td>
<td>0.9918</td>
</tr>
<tr>
<td>$e_6$</td>
<td>$r_{16DM}^e$</td>
<td>0.1915</td>
<td>0.9918</td>
</tr>
<tr>
<td>$e_7$</td>
<td>$r_{12DM}^e$</td>
<td>0.6497</td>
<td>0.9742</td>
</tr>
<tr>
<td>$e_8$</td>
<td>$r_{13DM}^e$</td>
<td>0.2972</td>
<td>0.7927</td>
</tr>
<tr>
<td>$e_9$</td>
<td>$r_{24DM}^e$</td>
<td>-0.4114</td>
<td>0.9897</td>
</tr>
<tr>
<td>$e_{10}$</td>
<td>$r_{14DM}^e$</td>
<td>0.6497</td>
<td>0.9742</td>
</tr>
<tr>
<td>$e_{11}$</td>
<td>$r_{12DM}^e$</td>
<td>0.1916</td>
<td>0.7180</td>
</tr>
<tr>
<td>$e_{12}$</td>
<td>$r_{12DM}^e$</td>
<td>0.6497</td>
<td>0.9742</td>
</tr>
<tr>
<td>$e_{13}$</td>
<td>$r_{13DM}^e$</td>
<td>-0.0084</td>
<td>0.9402</td>
</tr>
<tr>
<td>$e_{14}$</td>
<td>$r_{15DM}^e$</td>
<td>0.6497</td>
<td>0.9742</td>
</tr>
</tbody>
</table>

Table 8.2: Lower and upper bounds for the correlations between $X_1^{e_i}$ and $X_2^{e_i}$

---

1 We use the notation $\rho_{17}^e$ for the product moment correlation between $X_1^{e_i}$, the variable that has the highest influence on $X_7$, and $X_7$. While $X_7$ corresponds to $V_7$, $X_1^{e_i}$ to $X_6^{e_i}$ denote different variables for each expert. Therefore, $\rho_{17}^e$ should not be confused with $\rho_{17DM}^e$, which denotes the correlation between $V_1$, the first parent in the ordering of the decision maker, and $V_7$, according to expert $e_i$. 
8.1 Conclusion

For the present BBN structure and assuming the normal copula, one can, by using the recursive formula for partial correlations (2.4) and by keeping in mind which correlations are algebraically independent, explore if the experts possibly believe that the first and the second parent node are dependent. Indeed, supposing that the experts statements expressed their true beliefs accurately, we find that (almost) all experts indicate this. Combining the findings of this and the last chapter, we could draw either of the two following conclusions:

1. Expert $e_4$ and $e_9$ are "best" in estimating rank correlations. All other experts estimate that the parent variables are dependent, but this does not correspond to the real world.

2. We do not know whose estimates are "better" and whose are "worse" and therefore we consider all experts' dependence estimates as equally "good". Expert $e_4$ and $e_9$ are the two experts who agree most with the model structure, as they believe in independence of $X_1$ and $X_2$. All other experts believe that all parent variables depend significantly on each other.

If we would know whether the experts truly agreed with the structure, we would draw conclusion (1). But, we do not know. As a matter of fact our measure for experts' performance in estimating rank correlations, their (in)consistency regarding the given model structure, can just as much indicate the experts' (dis)agreement with the model structure. For this reason, we might favor conclusion (2). Nonetheless, we do not have the opportunity to contact the experts again within the time frame of the thesis and the present data is not sufficient to quantify a modified structure. On one hand because the experts gave very different variable rankings. On the other hand, because we are only able to compute relatively wide bounds for $\rho_{i12}^{e_i}$. Moreover, a performance measure for the dependence estimates is missing.

To understand whether conclusion (1) or conclusion (2) is more realistic, one needs to contact the experts again. What kind of information is needed from them to establish a validated model, is described in the following chapter.
9. Model Validation

With the present data, it is not possible to validate any model structure and to indicate what kind of model would be most useful for a decision maker. Therefore, this section suggests a couple of validation steps that could be taken in the future. They are illustrated by the flowchart in figure 9.1. For the validation the experts’ feedback is indispensable.

The first step is to ask the experts explicitly whether they believe that the parent nodes $V_1 - V_6$ are mutually independent and, if so, which ones, if not all of them. If the experts indicate that $V_1 - V_6$ are independent, the analyst can proceed to asking their feedback on model A and B. If, on the other hand, $V_1 - V_6$ are deemed dependent, the analyst has two options. (1) He/she can decide to include these dependences in the model. In this case, he/she will have to design a new expert judgment elicitation protocol incorporating the lessons learned, which are described in the following conclusion chapter. (2) He/she can decide to stick to the present model structure, e.g. for the sake of visual clarity or due to constraints regarding time and resources. Even though less optimal this is the easier choice and the analyst would proceed by asking validation questions on the existing models to find out which one is more realistic.

Of course, it would be desirable to receive feedback to the validation questions from as many experts as possible. A good limited selection would be expert $e_5$ and $e_6$, because had the highest global weights, and expert $e_4$ and $e_9$, because they were most consistent in their dependence assessments. The most valuable feedback could be achieved by conducting individual interviews with the experts. A method of least effort would be to use a short questionnaire. We suggest at least the following questions in order to meaningfully validate one or the other model:

1. Please rank the variables according to their highest influence (highest absolute rank correlation coefficient) on the number of contractor errors.

Consider 1000 critical tasks from the Moerdijk site, which are performed with one or more errors, which in the end hamper the operation.

2. For how many of these tasks did a company employee not have sufficient knowledge/skill/experience about hazard and working practice?
   a) 80%
b) 50%

3. For how many of these tasks have work constructions been not available, accurate, safe, clear, up-to-date and easy to read and understand?
   a) 40%
   b) 75%

The goal of the first question is to check whether the experts are consistent and give the same, or at least a very similar, ranking than in their original assessment. This should give us an indication on how reliable the expert’s statements are. The goal of the second and third question is to find out with which model predictions the experts agree more: model A’s or model B’s? Answer 2.a) approximately corresponds to \( \mu^A(V1|V7 = 1) \) and answer 2.b) approximately corresponds to \( \mu^A(V1|V7 = 1) \). Likewise, answer 3.a) compares to \( \mu^A(V4|V7 = 1) \) and 3.b) compares to \( \mu^B(V4|V7 = 1) \).
10. Conclusions

The objective of this thesis was to investigate the challenges of combining experts' assessment based on the present data and to suggest ways to deal with them. Section 10.1 summarizes the main findings by answering the research questions. Recommendations for future elicitation practice are given in section 10.2 and potential directions for future research in the field are presented in section 10.3.

10.1 Main Findings

The main research question has been broken down into four sub-questions which guided the research process. The main findings for each sub-question are summarized below.

1. Why are nonparametric BBNs beneficial to represent a joint distributions elicited from experts?

Nonparametric BBNs are a type of probabilistic graphical model consisting of nodes (variables) and arcs (influences). They represent the joint probability distribution of the variables and can realistically reflect inherent variabilities of a system. Considering that the differences in human performance among individuals are due to many factors, e.g. genetics or private life circumstances, which we have neither influence on nor can learn about thoroughly, it is essential to reflect these uncertainties in a human performance risk model. Additionally, they can reflect how uncertain experts are about their subjective probabilities.

Another advantage of BBNs is their ability to intuitively display cause-effect relations which can facilitate communication with the experts. The context to be quantified can be visualized for the experts. Moreover, the assessment burden is not very high, while no parametric distribution needs to be adopted. Assuming the normal copula, a multivariate distribution of seven variables can be obtained by asking experts for three quantiles of their subjective probabilities of each variable and six dependence estimates. From the three quantiles a minimum information distribution can be calculated. The dependence estimates could for instance be one exceedance probability and five ratios of rank correlations. This has been used here, but one could also ask for six exceedance probabilities.
It is not required to use the normal copula for nonparametric BBNs. As we have shown, the copula that realizes the experts’ rank correlations does not even have to possess the zero independence property. However, if one uses a copula other than the normal copula, the freely available software Uninet cannot be used to calculate the joint distribution, to visualize the network and to compute updates. Instead, one would need to sample the distribution by hand according to the general copula vine approach.

2. **What went wrong in the dependence elicitation and combination?**

Since we have not been part of the elicitation exercise, we can only surmise why the results turned out the way they are. Probably, the implications of the model structure have not been communicated well to the experts. If they would have, the experts would have been more aware of the bounds and not have violated them so grossly. Or, alternatively, the experts would have protested that the parent nodes are dependent and urged to change the model. Further, the bounds have not been calculated in real-time. As a consequence the extent to which the experts exceeded the bounds may not have been immediately clear to the analysts or the experts. Finally, it is very likely that the experts did not have sufficient knowledge about probabilities and dependence.

The dependence estimates of the experts have been combined by averaging rank correlations. In chapter 7 we have demonstrated that this is not just inaccurate and violates conditional independence statements (if present), but may result in very misleading numbers. The fact that the dependence structure of model O looks reasonable may be attributed to this specific application. For other applications this may turn out differently.

3. **How serious are the shortcomings of the conducted dependence elicitation and previous combination?**

The dependence data is far from optimal. A particular problem is that we do not know whether the experts performed poorly, which means they do not have sufficient experience with estimating rank correlations under conditional independence assumptions, or whether they disagree with the model structure or both. In this thesis we have first assumed the former and tried to built the best possible models in these circumstances and after that we have assumed the latter and discovered that, under this assumption, (at least) the first and second parent node of each expert are likely to be dependent. However, without feedback from experts we do not know which assumption is more realistic.

In either case it cannot be recommended to use the models built. If the experts misjudged the rank correlations so grossly, their judgment cannot be trusted and, consequently, the resulting models can neither. Even the "best" expert exceeded one of the upper conditional rank correlations by 1323%. If, on the other hand, the experts strongly disagree with independent parents, a model which disregards this fact may lead to wrong conclusions. For instance, risks may be systematically underestimated. Imagine this human performance model as part of a integrated risk model. An external variable may have a negative influence on the quality of communication. If the quality of communication has a positive correlation with commitment, a model which does not exhibit this relation will predict a higher contractor performance and a lower overall risk. At the very least one needs to discuss with the experts again to what extent predictions from such a model
can be useful, if one has good reasons to not include these dependencies, e.g. increasing complexity of the model.

4. **What makes the combination of dependence estimates particularly challenging?**

First of all, the approach of combining rank correlations is not as simple as combining marginal distributions. It involves the calculation of exceedance probabilities from rank correlations, the pooling of these probabilities and the translating them back to rank correlations.

However, the real issue is the lack of an adequate empirical control over the quality of dependence estimates. The present data suggests that experts generally do not perform equally well in eliciting marginal distribution and eliciting rank correlations. Guidelines on how to design calibration questions for dependence estimates and how to evaluate them are still missing. In this thesis we used the consistency of the experts to infer who performs well and who does not, but this was a method of last resort. First of all, if the elicitation would have been conducted very carefully, (almost) all of the experts should be consistent. Second, we have reason to assume that an expert who poorly estimated the marginal distributions cannot indicate good rank correlation. We have seen in chapter 7 that the notion of rank correlation is profoundly linked to the underlying joint distribution. An expert’s rank correlation is only meaningful together with the median of his marginal distributions.

### 10.2 Recommendations

Although we included some validation guidelines in this report, the preferable option is to gather new data with another expert judgment study.

Foremost, and this is of course true for all experiments, the better the preparation, the design and the more careful the execution of the elicitation, the less work is needed to analyze the data and the more reliable are the results. Sufficient time should be allocated to the expert judgment exercise and it is indispensable that the experts are trained in probability and understand the basic concepts of dependence. In particular, they need to gain experience in assessing rank correlations before the actual elicitation. Optimally they would learn, also by trial and error, with a small demonstration case that shows them plainly the consequences of their first assessment for the subsequent ones. Needless to say, the bounds should be calculated in real-time, both for the dry-run and the real elicitation, such that the experts can revise their statements if necessary.

Furthermore, we suggest to conduct the elicitation in two rounds. In the first round, the experts elicit the marginal distributions. The data is immediately analyzed and the decision maker distributions are combined with the optimal weights according to the classical model. The resulting decision maker distributions are communicated to the experts and they are given some time to contemplate them. Then, in the second round, the experts elicit the rank correlations.

\(^1\)Of course a different quantile could also be used.
correlations or possibly another dependence measure. Fortunately, the experts already agree on the distributions of the variables and, therefore, the experts’ dependence estimates can be combined with any weights. One could stick to the weights of the classical model, but if a new weighting scheme for dependence estimates is available that one could be used just as well.

This proposal for practicing expert judgment is derived from the conclusions of this thesis. However, a number of questions arise, which need to be answered before we can claim that this approach indeed improves the data obtained from the experts. This brings us to the following section on further research.

10.3 Suggestions for Further Research

While the above suggestions will lead to consistent and usable data, they do not necessarily lead to "good" data. Therefore, finding an answer to the following two questions is important:

1. Are experts able to truly change their belief of the distribution for a variable if they are confronted with the decision maker distribution?

2. If they are, would they actually give different rank correlations than if they would still believe in their own subjective distribution?

Moreover, it would be desirable to devise a method that enables the analyst to properly score the expert’s ability to assess dependences. Not all experts who perform well in eliciting marginals may be performing well in eliciting dependences. If the elicitation of marginal distributions and dependence measures is separated as proposed in the previous section, the two skills may be considered separately as well. Being good at eliciting dependences does not necessarily require being good at eliciting marginals anymore, if, of course, the conclusions to the two questions listed above is positive.

Also the suitability of different kinds of dependence measures for expert judgment can be investigated. We have used ratios of rank correlations, other studies may use exceedance probabilities. Maybe, tail dependence is important for a particular case. Is there a way to elicit the parameter of the chosen copula directly or does one better ask for exceedance probabilities with high quantiles?

Finally, it would be useful to have a training case for a dry run before the elicitation. This could even be a standardized case for all elicitations to come concerning a daily life situation, as it is supposed to make the experts familiar with the procedure and illustrate the implication of their estimates, in particular the dependence estimates. Having a standardized case would on the long run simplify the work of the analyst, since he/she does not have to design a new one for each elicitation. Furthermore, the analyst will gain a lot of experience with the case and will be able to see immediately where and why experts make mistakes without having

---

2 Morales Nápoles et al., 2013, which will shortly present their research at a conference, propose an approach using a similar principle as the classical model: a calibration score.
to consult a computational software like Matlab. Also, he/she will improve in giving useful advice to the individual experts. The bottom line is, the more experience the analyst has with a case, the better he/she can teach it to the experts, the better trained are the experts and the better are the results of the real elicitation.
References


REFERENCES

[107]


Part III

Appendices
A. Protocol for the Elicitation

This appendix provides the protocol which has been used for the case study in chapter 4. Small modifications have been made for this thesis as to not disclose confidential information.
Protocol for Elicitation of Input Parameters for Human Model in Risk Tool for an Oil and Gas Company

Part II: Quantification of Human Model for the company's employees

Safety Science Group
Delft University of Technology
Part I: Introduction

Thank you for participating in this expert judgment exercise to quantify the human performance model in a risk management tool for company operations, developed at Delft University of Technology.

The main goal of this project is to develop a model, which can help taking management decisions toward controlling short terms risks, for which the reward is immediately visible; and also the rare disasters that the individual managers are unlikely to identify in management activities and management interventions, but may have a big impact on the company as a whole. Therefore, the results of the model will incorporate not only the technical failures, but also the human performance and the management influences on both human and technology. This results in an integrated model for risk management decisions.

Within this project, a human performance sub-model is developed separately and then integrated into the large risk management model. The main goal of this sub-model is to present the factors that influence the human performance, and demonstrate what should be done in order to decrease the likelihood of human errors and improve human performance in execution of tasks.

1.1. OBJECTIVES

In this stage of the project, the focus is on the human performance model for contractors. The structure of the model is presented in Figure 1. The main objective of this exercise is to quantify this model. Currently there is no information about the variables identified in this model. Therefore, structured expert judgement method is used to quantified this model and reduce the uncertainty regarding these factors.

Each variable in the model will be explained into details in section 1.2. A brief introduction about the method used for model quantification will be presented in section 1.3.

Figure 1: Human performance model (for contractor)
1.2. THE CONTRACTOR HUMAN PERFORMANCE MODEL

Figure 1 shows the structure of the contractor performance model. It is a simplified representation of how the number of contractor errors depends on certain factors. It has been defined based on literature and discussions with managers in the company. The definition of each variable is given in Table 1.

Table 1: Definitions of factors influencing human performance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Contractor performance</td>
<td>The likelihood that a task is performed with one or more errors which in the end hamper the operation</td>
</tr>
<tr>
<td>2  Quality of procedure</td>
<td>The likelihood that work instruction is available, accurate, safe, clear, up-to date and easy to read and understand</td>
</tr>
<tr>
<td>3  Competence</td>
<td>The likelihood that the contractor have sufficient knowledge/skill/experience to execute a particular task</td>
</tr>
<tr>
<td>4  Capacity to work</td>
<td>The likelihood that the contractor shows the sign of in-availability to work (including fitness to work, alertness, physical state, and psychological state)</td>
</tr>
<tr>
<td>5  Job specific communication</td>
<td>The likelihood that during permit to work/shift/on the job communication, the required information (job hazard, safety precaution) is transferred to the contractors</td>
</tr>
<tr>
<td>6  Human Factors in design</td>
<td>The likelihood that equipment and work stations are designed for easy access and easy operation and maintenance</td>
</tr>
<tr>
<td>7  Commitment</td>
<td>The likelihood that during critical activities the individual carry out the required behaviour at the right moment, with the right care and attention to detail in order to control the risk(^1).</td>
</tr>
</tbody>
</table>

1.3. ELICITATION METHOD\(^2\)

The quantification of the human performance model means assigning probability distributions for each factor and strength of influence for each arc in Figure 1. Since this information is not available yet, experts in the field are used to assign their uncertainty regarding these quantities in a structured way.

Two kinds of information are gathered in this “structured expert judgment” exercise:

- information on the probability distributions of variables (1.3.1)
- information on the influences between variables (1.3.2)

\(^1\) Commitment to safety deals with incentives of individuals carrying out the primary business activities not to choose other criteria above safety (such as ease of working, time saving, social approval etc.) as well as the resolution of conflicts between safety and other criteria.

The method (Cooke 1991) used in this exercise has been applied successfully in various fields, such as nuclear industry, chemical and gas industry, groundwater and water pollution, aerospace and aviation, occupational sector, health, banking and volcanos, involving over 500 experts and 3600 variables, in about 67000 elicitations (Cooke and Goossens 2008).

1.3.1 INFORMATION ON THE PROBABILITY DISTRIBUTIONS OF VARIABLES

The information on the probability distributions of variables will be elicited via three numbers:

- the 50th quantile, or the median values of the distribution: if we would have 1000 samples of variable values, 500 of them should be below and 500 of them should be higher than the median value;
- the 5% quantile value: it would surprise you if more than 50 out of 1000 samples have a value lower than this value;
- the 95% quantile value: it would surprise you if more than 50 out of 1000 samples have a value higher than this value (or if more than 950 out of 1000 samples have a value lower than this value).

Here is an example of a (general) question and the answer given by one (imaginary) expert:

<table>
<thead>
<tr>
<th>Question 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consider 1000 randomly chosen students who graduated Delft University of Technology. How many of them did not pass their first exam?</td>
</tr>
<tr>
<td>5%</td>
</tr>
<tr>
<td>20</td>
</tr>
</tbody>
</table>

The expert’s answer could be interpreted as it follows:

If this experiment of randomly choosing 1000 students would be repeated 100 times, in 5 cases the number of students who did not pass their first exam is less than 20, in half of the cases, the number of students is less than 300 and in only 5 cases, the number of the students who did not pass the first exam is above 450.

It can also be interpreted as:

The expert’s best estimate is 300. The expert would be surprised if the number of the students who did not pass their first exam is below 20 or above 450.

In terms of probabilities, this can be seen in Figure 2. It can be seen that there is 5% chance that the real number of students is below 20 and 5% chance that this number is above 450. It can also be said that there is 50% chance that this number is below 300.

---

3 It has to be noticed that the three numbers given by the expert do not need to be added to 1000.
Another example of question and answer is as follows:

**Q2:** Consider the population of athletes who finish the Rotterdam Marathon. What is the average finish time?

<table>
<thead>
<tr>
<th></th>
<th>5%</th>
<th>50%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3:15</td>
<td>3:30</td>
<td>3:50</td>
</tr>
</tbody>
</table>

The expert’s answer could be interpreted as it follows:

*If we would have repeat 100 times the Rotterdam marathon, in 5 cases the average finish time is below 3:15, in 50 cases the average finish time is below 3:30 and in only 5 cases the average finish time is above 3:50.*

In other words, *it would be surprising (for that particular expert) that the average finish time will be below 3:15 in more than 5 cases and above 3:50 in more than 5 cases out of the 100 times.*

### 1.3.2 INFORMATION ON THE INFLUENCES BETWEEN VARIABLES

The purpose of obtaining the information on the influences between variables is to obtain ‘rank correlation coefficients’, which we will refer to as ‘correlation’ in the following.

Correlation tells something about the strength and “direction” of relationship between two variables. It can have any value between -1 and +1.

- If high values of a variable imply high values of another variable, there is a (strong) positive correlation, and the correlation will be close to +1 (the stronger the correlation, the closer it is to +1).
• If high values of a variable imply low values of another variable, there is a (strong) negative correlation, and the correlation will be close to -1.
• If there appears to be no such relation (the values of a variable do not imply anything on the values of another variable) the variables are called ‘independent’ and the correlation equals 0.

The information of the dependence between the variables will be elicited via 4 questions.

*Using the same example from athletes, suppose “the average finish time at Rotterdam Marathon” is influenced by three variables:
  • the average number of training hours per week
  • the average number of marathons ran before
  • the average intermediary time at half marathons.

Question 1: Ask the experts to rank the importance of influence on the average finish time at Rotterdam Marathon
  Please assign number 1 to the variable that you consider to have the strongest influence on the average finish time, 2 to the next most important and so on until the least influential variable.

Question 2: Ask the experts the directions (positive/negative) of influences on the average finish time at Rotterdam Marathon
  Please indicate which of the following statement you think is true:
  o High average number of training hours per week relate to longer finish time at Rotterdam Marathon
  o High average number of training hours per week relate to shorter finish time at Rotterdam Marathon

Question 3: For the variable that expert ranked highest, we have the following question (assume that the expert thinks “training hours per week” is the most important factor):
  Suppose that instead of randomly choosing 100 athletes who finished the Rotterdam Marathon, we choose 100 athletes who have been trained more than 20 hours per week (the median value for training time). How many of them do you expect to finish below 3h 30 min (the median number that you specified in Q1)?

Question 4: For the remaining variables, we ask the experts to indicate their influence as a portion of the influence of the variable that you ranked highest:
1.3.3 ANALYSING THE EXPERT ANSWERS

The expert is asked to assign his/her uncertainty regarding a set of variables for which the real values are known to the analyst, but unknown (at least at the moment of the interview) to the experts. These questions are called calibration or seed questions and are used to measure and validate the expert performance in uncertainty quantification. The scope of using seed questions is not to assess how good the expert is in the field, but how good the expert in quantifying the uncertainty on some variables. A “good” expert in quantifying uncertainty is one who:

- gives assignments which are statistically accurate, and
- gives assessments which are informative.

Statistical accuracy means that for each of the uncertain quantities, the interval that is given by the expert, in which he/she is certain that contains 90% of the (unknown) true values, indeed contains 90% of the true values. With other words, we will consider the expert’s answers as samples from a probability distribution. As a consequence, a “perfect calibrated expert” would be one who’s answers are as it follows: in 5% of the number of seed questions, the real value will be below the values assigned by this specific expert for the 5th quantiles, in 5% of the seed questions the real values of the variables will be above the values assigned as the 95th quantiles, in 45% of the seed questions, the real values will be between the 5th and 50th quantiles assigned be this expert, and in 45% of the number of seed questions, the real values will be between the 50th and the 95th quantiles assigned by expert. For example, suppose that 20 variables are known when they were elicited. Then the statistically accuracy would mean that: in 1 question the true value should be below the 5% percentile value; in 1 question the true value should be above the 95% percentile and in 18 questions, the true value should be between the 5% and the 95% percentile values (in 9 questions between the 5% percentile and the median value and in 9 questions between the median and the 95% percentile value).

Information or informativeness measures the degree to which the expert’s distributions are concentrated; with other words, a highly informative expert is one who catches 90% of the variables in small intervals, e.g. that the expert is 90% certain of catching the true values in small intervals.
Both performance indicators are important, but the statistical accuracy dominates the informativeness.

The calibration and information scores are used to weight expert’s answers and to combine them in order to obtain one single distribution for each variable of interest. It happens often that one expert scores higher on calibration and information, and therefore he/she receives weight 1, while the others receive weight 0. This does not mean that the answers of the other experts are not taken into account, but rather that their answers are included already in the answer of the expert with weight 1 and giving non-zero weight to their answers means only that more “noise” is added in the final distribution.

1.4 Confidentiality/ feedback
After the individual elicitation, we will report the results including the underlying argumentation and any other discussions and these will be send to you for review.

The names and qualifications of the experts will be published in the final report, as will the individual expert assessments and all information relevant, but the link between individual expert assessments and their identity will be removed in the published reports.

The questionnaire contains three sets of questions. One set related to the probability distributions for the factors presented in Figure 1 and the influences between these factors. One set is the calibration questions. The last set is the research questions which are of interest for the theoretical development of the method. In total there are 29 questions.
Part II: Elicitation of probability distributions and correlations

All questions in this section refer to *all company employees* in the Moerdijk site. When we speak about tasks made by company employees, we talk about HSSE critical tasks for process safety. The critical tasks done by company people are critical tasks such as:

- React on the high pressure alarm
- Temporarily override and reinstall safety features (barriers)
- Authorise people to do so
- Execute start-up and shut-down procedures

The errors are defined as such that are not immediately recovered and have the potential to lead to an incident/accident.

### 2.1 MARGINAL DISTRIBUTIONS

Please assign your uncertainty regarding the following quantities.

<table>
<thead>
<tr>
<th>Q1</th>
<th>Contractor performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consider 1,000 critical tasks chosen at random from Moerdijk site. How many of these tasks will the company employee execute errors that have the potential to lead to an incident/accident?</td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>50%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q2</th>
<th>Competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consider 1,000 critical tasks chosen at random from Moerdijk site. How many of these tasks will the company employee do not have sufficient knowledge/skill/experience about the hazard and working practice?</td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>50%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q3</th>
<th>Capacity to work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consider 1,000 critical tasks chosen at random from Moerdijk site. How many of these tasks will the company employee show sign of unavailability to work (including fitness to work, alertness, physical state, and psychological state)?</td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>50%</td>
</tr>
</tbody>
</table>
### Q4  Commitment

Consider 1,000 critical tasks chosen at random from Moerdijk site. How many of these tasks will the company employee not carry out with the required behavior at the right moment, with the right care and attention to detail in order to control the risk?

<table>
<thead>
<tr>
<th></th>
<th>5%</th>
<th>50%</th>
<th>95%</th>
</tr>
</thead>
</table>

### Q5  Quality of procedure

Consider 1,000 critical tasks chosen at random from Moerdijk site. How many of these tasks will the company employee have to refer to bad work instruction that is considered to be not available, not accurate, not safe, not clear, not up-to-date and not easy to read and understand?

<table>
<thead>
<tr>
<th></th>
<th>5%</th>
<th>50%</th>
<th>95%</th>
</tr>
</thead>
</table>

### Q6  Job specific communication

Consider 1,000 randomly chosen communications during permit to work/shift/on the job communication in Moerdijk site. How many of them will not clearly provide precautionary information to complete the job safely?

<table>
<thead>
<tr>
<th></th>
<th>5%</th>
<th>50%</th>
<th>95%</th>
</tr>
</thead>
</table>

### Q7  Human factor in design

Consider 1,000 critical tasks chosen at random from Moerdijk site. How many of these tasks will the company employee have to perform with equipment and work stations that are not designed for easy access and easy operation and maintenance?

<table>
<thead>
<tr>
<th></th>
<th>5%</th>
<th>50%</th>
<th>95%</th>
</tr>
</thead>
</table>
2.2 DEPENDENCE INFORMATION

In this section, we are interested in the relationship between the variable of “contractor performance” and the variables 2-7 in Table 1 (repeated below).

Q8: Please assign number 1 to the variable that you consider to have the strongest influence (highest absolute rank correlation coefficient) on the number of contractor errors, 2 to the next most important and so on until 6 to the least influential variable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of procedure: work instruction that is not available, accurate, safe, clear, up-to date and easy to read and understand</td>
<td></td>
</tr>
<tr>
<td>Competence: the contractor not have sufficient knowledge/skill/experience about the hazards and working practices necessary for executing the particular task he is working on</td>
<td></td>
</tr>
<tr>
<td>Capacity to work: contractor who shows the sign of in-availability to work (including alertness, physical state, and psychological state)</td>
<td></td>
</tr>
<tr>
<td>Job specific communication: during permit to work/shift/on the job communication, the required information (job hazard, safety precaution) is not communicated unambiguously to the contractors</td>
<td></td>
</tr>
<tr>
<td>HF in design: equipment and work stations are not designed for easy access and easy operation and maintenance</td>
<td></td>
</tr>
<tr>
<td>Commitment: contractor who does not carry out the required behaviour at the right moment, with the right care and attention to detail in order to control the risk</td>
<td></td>
</tr>
</tbody>
</table>

For the variable that you ranked highest, we have the following question:

Q9

Suppose that instead of randomly choosing 1,000 tasks from Moerdijk site, we choose a sample of 1,000 tasks for which we know that there are more than (median V1) ________ tasks with bad (V1) _________.

What is the chance that in this sample more than (median human error) ________ tasks are performed with errors?
**Q10:** For each pair below indicate which of the two effects is the most dominant?

<table>
<thead>
<tr>
<th>Effect</th>
<th>Influence as a portion of the influence of the highest ranked variable (0 – 100%)</th>
<th>Direction of the correlation (positive/negative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>When the <strong>procedure is not</strong> available, accurate, safe, clear, up-to date and easy to read and understand</td>
<td>0 higher 0 lower</td>
<td>chance of company employee making error is expected</td>
</tr>
<tr>
<td>When the contractor <strong>does not have sufficient competence</strong> (knowledge/skill/experience) about the hazards and working practices necessary for executing the particular task</td>
<td>0 higher 0 lower</td>
<td></td>
</tr>
<tr>
<td>When the contractor shows the sign of <strong>in-availability</strong></td>
<td>0 higher 0 lower</td>
<td></td>
</tr>
<tr>
<td>When the information is <strong>not communicated clearly</strong> to the contractors during safety critical communication</td>
<td>0 higher 0 lower</td>
<td></td>
</tr>
<tr>
<td>When the equipment and work stations are <strong>not designed</strong> for easy access and easy operation and maintenance</td>
<td>0 higher 0 lower</td>
<td></td>
</tr>
<tr>
<td>When the contractor is <strong>less committed</strong> to safety during work</td>
<td>0 higher 0 lower</td>
<td></td>
</tr>
</tbody>
</table>

**Q11:** For each of the remaining variables, please indicate the ratio of the influence of that variable comparing to the one that you ranked highest:

<table>
<thead>
<tr>
<th>Node</th>
<th>Rank</th>
<th>Influence as a portion of the influence of the highest ranked variable (0 – 100%)</th>
<th>Direction of the correlation (positive/negative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of procedure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capacity to work</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job specific communication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF in design</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commitment</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Q12
Suppose that instead of randomly choosing 1,000 tasks from Moerdijk site, we choose a sample of 1,000 tasks for which we know that there are more than
- (median V1) \_\_\_\_\_\_ tasks with bad (V1) \_\_\_\_\_\_.
- more than (median V2) \_\_\_\_\_\_ tasks with bad (V2) \_\_\_\_\_\_.

What is the chance that in this sample more than (median human error) \_\_\_\_\_\_ tasks are performed with errors?

### Q13
Suppose that instead of randomly choosing 1,000 tasks from Moerdijk site, we choose a sample of 1,000 tasks for which we know that there are more than
- (median V1) \_\_\_\_\_\_ tasks with bad (V1) \_\_\_\_\_\_,
- more than (median V2) \_\_\_\_\_\_ tasks with bad (V2) \_\_\_\_\_\_, and
- more than (median V3) \_\_\_\_\_\_ tasks with bad (V3) \_\_\_\_\_\_.

What is the chance that in this sample more than (median human error) \_\_\_\_\_\_ tasks are performed with errors?
Part III Calibration Questions

To capture your uncertainty, in all questions from this section, we will ask you to provide the 5%, 50% and 95% percentiles of your uncertainty distribution, which can be interpreted as that we ask for your best estimate (50%), the value which would surprise you if the real value would be lower (5%), and the value which would surprise you if the real value would be higher (95%).

All the questions refer to the OGP industry, if not otherwise specified.

<table>
<thead>
<tr>
<th>CQ 1</th>
<th>Out of the total number of fatal offshore incidents in the period 1970-2007, how many (in percentage) were in Europe?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>95%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CQ 2</th>
<th>Out of the total number of fatalities reported in the last three years (2009-2011), how many (in percentile) were during maintenance, inspection, testing?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>95%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CQ 3</th>
<th>Out of the total number of fatalities reported in the last three years (2009-2011), how many (in percentile) were due to falls from height?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>95%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CQ 4</th>
<th>Out of the total number of fatalities reported in the last three years (2009-2011), how many (in percentage) were onshore?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>95%</td>
</tr>
</tbody>
</table>
### CQ 5
Out of the total number of fatalities reported in the last two years (2010-2011), how many (in percentage) were in production operations?

| 5% | 50% | 95% |

### CQ 6
Out of the total number of fatalities reported in the last three years (2009-2011), how many were in land transportation?

| 5% | 50% | 95% |

### CQ 7
Out of the total number of lost day cases for company employers in 2010 and 2011, how many (in percentage) were in production operations?

| 5% | 50% | 95% |

### CQ 8
Out of the total number of lost day cases for contractors in 2010 and 2011, how many (in percentage) were in production operations?

| 5% | 50% | 95% |

### CQ 9
What is the percentage of worked hours reported in onshore operations?

| 5% | 50% | 95% |
### CQ 10
Out of the total number of **causal factors for the fatal incident reported in 2010 and 2011**, how many (in percentage) were **inadequate supervision**?

<table>
<thead>
<tr>
<th></th>
<th>5%</th>
<th>50%</th>
<th>95%</th>
</tr>
</thead>
</table>

### CQ 11
What is the percentage of people in **Moerdijk in 2010** who feel **well informed** about what is expected in their job?

<table>
<thead>
<tr>
<th></th>
<th>5%</th>
<th>50%</th>
<th>95%</th>
</tr>
</thead>
</table>

### CQ 12
What is the percentage of people in **Moerdijk in 2010** who think that the **communications are not handled well** when changes are made?

|   | 5% | 50% | 95% |
Part IV: Open question about management actions

Given the human model has been quantified, we are now able to estimate the influences of human factors on human errors. The next step is to identify management actions and evaluate their effectiveness on improving (or degrading) human factors. Identifying management actions and evaluating their effectiveness on improving human factors will be one of the research plans for next year.

As a preliminary study, in this section we use open questions to gather some information on the potential management actions for contractors.

We define these management factors as providing effective risk control objectives, instructions and resources to contractors in order to improve their performance and reduce their failure probabilities.

**Communication**

<table>
<thead>
<tr>
<th>Management action</th>
<th>Does this also apply for company employee?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Yes</td>
</tr>
<tr>
<td>2.</td>
<td>No</td>
</tr>
<tr>
<td>3.</td>
<td>Yes</td>
</tr>
<tr>
<td>…</td>
<td>No</td>
</tr>
</tbody>
</table>

**Commitment**

<table>
<thead>
<tr>
<th>Management action</th>
<th>Does this also apply for a company employee?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Yes</td>
</tr>
<tr>
<td>2.</td>
<td>No</td>
</tr>
<tr>
<td>3.</td>
<td>Yes</td>
</tr>
<tr>
<td>…</td>
<td>No</td>
</tr>
</tbody>
</table>


B. Additional KCI-test results

This appendix refers to section 3.3. It contains the results of two simulation runs demonstrating that according to the KCI-test the joint distributions obtained with the normal copula and the t-copula do not all inhibit (conditional) independence statements imposed by the graph. It should be noted, though, that the p-values for the normal copula and the t-copula are almost equal for each statement. For this reason, inconsistencies can be attributed to the test. If variables are not (conditionally) independent, the test seems to give a p-value equal to zero. While, if they are (conditionally) independent the p-value is greater than zero, albeit sometimes rather close to zero.

B.1 Example Run 1

The empirical correlation matrix for the joint distribution obtained with the normal copula is

$$R_{C_{G}^{R}} = \begin{pmatrix} 1 & 0.0462 & -0.0055 & 0.4973 \\ 0.0462 & 1 & -0.0200 & 0.4568 \\ -0.0055 & -0.0200 & 1 & -0.0224 \\ 0.4973 & 0.4568 & -0.0224 & 1 \end{pmatrix}$$  \hspace{1cm} (B.1)$$

and the one for the t-copula is

$$R_{C_{T,3}^{R}} = \begin{pmatrix} 1 & 0.0527 & -0.0234 & 0.4779 \\ 0.0527 & 1 & -0.0179 & 0.4333 \\ -0.0234 & -0.0179 & 1 & -0.0286 \\ 0.4779 & 0.4333 & -0.0286 & 1 \end{pmatrix}$$  \hspace{1cm} (B.2)$$
Table B.1: P-values for the KCI-test

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>$p$-value</th>
<th>$H_0$ accepted?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_{R}^{Ga}$</td>
<td>$C_{v,R}^{Ga}$</td>
</tr>
<tr>
<td>$X_1 \perp X_2$</td>
<td>0.0420 0.0640</td>
<td>no yes</td>
</tr>
<tr>
<td>$X_1 \perp X_3$</td>
<td>0.1810 0.2630</td>
<td>yes yes</td>
</tr>
<tr>
<td>$X_2 \perp X_3$</td>
<td>0.4650 0.7010</td>
<td>yes yes</td>
</tr>
<tr>
<td>$X_1 \perp X_4$</td>
<td>0 0</td>
<td>no no</td>
</tr>
<tr>
<td>$X_2 \perp X_4$</td>
<td>0 0</td>
<td>no no</td>
</tr>
<tr>
<td>$X_3 \perp X_4$</td>
<td>0.4370 0.5220</td>
<td>yes yes</td>
</tr>
<tr>
<td>$X_3 \perp X_4 \mid X_1$</td>
<td>0.8004 0.8193</td>
<td>yes yes</td>
</tr>
<tr>
<td>$X_3 \perp X_4 \mid X_2$</td>
<td>0.3356 0.4944</td>
<td>yes yes</td>
</tr>
<tr>
<td>$X_3 \perp X_4 \mid X_1, X_2$</td>
<td>yes yes</td>
<td></td>
</tr>
</tbody>
</table>

B.2 Example Run 2

The empirical correlation matrix for the joint distribution obtained with the normal copula is

$$R_{C_{R}^{Ga}} = \begin{pmatrix}
1 & 0.0175 & -0.0117 & 0.5081 \\
0.0175 & 1 & -0.0071 & 0.4430 \\
-0.0117 & -0.0071 & 1 & 0.0369 \\
0.5081 & 0.4430 & 0.0369 & 1
\end{pmatrix}$$  (B.3)

and the one for the t-copula is

$$R_{C_{3,R}^{t}} = \begin{pmatrix}
1 & 0.0191 & -0.0095 & 0.5230 \\
0.0191 & 1 & -0.0008 & 0.4591 \\
-0.0095 & -0.0008 & 1 & 0.0332 \\
0.5230 & 0.4591 & 0.0332 & 1
\end{pmatrix}$$  (B.4)
Table B.2: P-values for the KCI-test

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>p-value</th>
<th>$H_0$ accepted?</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_R^{G_a}$</td>
<td>$C_{v,R}^{t}$</td>
<td>$C_R^{G_a}$</td>
</tr>
<tr>
<td>$X_1 \perp X_2$</td>
<td>0.7690</td>
<td>0.8500</td>
</tr>
<tr>
<td>$X_1 \perp X_3$</td>
<td>0.3500</td>
<td>0.3420</td>
</tr>
<tr>
<td>$X_2 \perp X_3$</td>
<td>0.1160</td>
<td>0.1390</td>
</tr>
<tr>
<td>$X_1 \perp X_4$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$X_2 \perp X_4$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$X_3 \perp X_4$</td>
<td>0.0260</td>
<td>0.0260</td>
</tr>
<tr>
<td>$X_3 \perp X_4 \mid X_1$</td>
<td>0.0569</td>
<td>0.0363</td>
</tr>
<tr>
<td>$X_3 \perp X_4 \mid X_2$</td>
<td>0.0091</td>
<td>0.0142</td>
</tr>
<tr>
<td>$X_3 \perp X_4 \mid X_1, X_2$</td>
<td>0.0050</td>
<td>0.0093</td>
</tr>
<tr>
<td>$X_1 \perp X_2 \mid X_4$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$X_1 \perp X_3 \mid X_4$</td>
<td>0.4845</td>
<td>0.4469</td>
</tr>
<tr>
<td>$X_2 \perp X_3 \mid X_4$</td>
<td>0.0931</td>
<td>0.1420</td>
</tr>
</tbody>
</table>
### C. Presentation

Table C.1: Excalibur Output: optimized global and item weights with significance level $\alpha = 0.6876$

<table>
<thead>
<tr>
<th></th>
<th>Calibration</th>
<th>Information</th>
<th>Normalized weights without DM</th>
<th>Normalized weights with DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>0.5338</td>
<td>0.4866</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_2$</td>
<td>0.00348</td>
<td>0.4561</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_3$</td>
<td>0.00348</td>
<td>1.016</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_4$</td>
<td>0.02651</td>
<td>0.9024</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_5$</td>
<td>0.6876</td>
<td>0.8534</td>
<td>0.5093</td>
<td>0.3688</td>
</tr>
<tr>
<td>$e_6$</td>
<td>0.6876</td>
<td>0.8222</td>
<td>0.4907</td>
<td>0.3553</td>
</tr>
<tr>
<td>$e_7$</td>
<td>0.5405</td>
<td>0.6878</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_8$</td>
<td>0.08763</td>
<td>1.089</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_9$</td>
<td>0.2357</td>
<td>0.6681</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_{10}$</td>
<td>0.00348</td>
<td>0.8325</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_{11}$</td>
<td>0.01566</td>
<td>0.7833</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_{12}$</td>
<td>0.6638</td>
<td>0.8263</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_{13}$</td>
<td>7.86E-005</td>
<td>1.02</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$e_{14}$</td>
<td>0.5338</td>
<td>0.9099</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$DM_{global}$

<table>
<thead>
<tr>
<th></th>
<th>$DM_{global}$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>0.6638</td>
<td>0.6614</td>
<td>0.2759</td>
<td></td>
</tr>
<tr>
<td>$e_{12}$</td>
<td>0.6638</td>
<td>0.6805</td>
<td>0.2816</td>
<td></td>
</tr>
</tbody>
</table>