TECHNISCHE UNIVERSITÄT MÜNCHEN,
Ingenieurfakultät Bau Geo Umwelt

Reliability analysis with Bayesian networks

Kilian Martin Zwirglmaier

Vollständiger Abdruck der von der
Ingenieurfakultät Bau Geo Umwelt
der Technischen Universität München zur Erlangung des akademischen Grades
eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigten Dissertation.

Vorsitzende/-r: Univ.-Prof. Dr.-Ing. André Borrmann

Prüfende/-r der Dissertation:

1. Univ.-Prof. Dr. sc. techn. Daniel Straub

2. Prof. Dr. Paolo Franchin (Sapienza Università di Roma)

3. Prof. Dr. Iris Tien (Georgia Institute of Technology)

Die Dissertation wurde am 13.09.2016 bei der Technischen Universität München
eingereicht und durch die
Ingenieurfakultät Bau Geo Umwelt am 06.12.2016 angenommen.

Abstract

Ensuring the reliability of anthropogenic systems is an important responsibility of engineers. Engineers therefore have to deal with uncertainties associated with the loads and capacities, and uncertainties on human factors. To deal with these uncertainties, engineers can build probabilistic models. A probabilistic modeling framework with large potential for reliability engineering and engineering risk analysis is the Bayesian network (BN). While BNs have been successfully applied to reliability engineering, there are remaining issues, some of which are addressed in this thesis. Reliability analysis with BNs can be considered a three step procedure: elicitation of the qualitative model structure, quantification of the model and inference in the model. The focus in this thesis is on the first and the last step. Elicitation of the BN structure, which is in practice done often in a rather ad-hoc manner, is addressed by proposing a classification of structure elicitation approaches. Inference is addressed by proposing two methods that are applicable in the context of rare events, as encountered in reliability analysis.

The directed acyclic graph (DAG) of a BN represents the qualitative part of the model i.e. the (causal) dependence structure between the random variables in the model. The DAG thus is a pivotal part of a model's traceability, which is important to ensure the validity of a probabilistic model. Elicitation of the DAG is therefore an essential part of the whole modeling process. A classification of model elicitation approaches that are used in reliability and engineering risk analysis is proposed. Along with this, a review of the respective literature with focus on structure elicitation is provided. This classification distinguishes between four approaches, namely structure elicitation based on (1) other probabilistic models, (2) general models, (3) structure learning from data and (4) structure elicitation based on domain expert knowledge. The typical challenges associated with these approaches are discussed. On this basis a framework for deriving traceable BN structures for human reliability analysis (HRA) directly from the findings from psychological literature is proposed. The framework is applied to a crew failure mode (CFM) from the integrated decision-tree human event analysis system (IDHEAS) HRA method, which is being developed by the U.S. Nuclear Regulatory Commission (NRC). An approach for quantification of the structures based on both experts and data is presented.

Inference algorithms are applied to infer the conditional probability distribution of a random variable of interest, possibly conditional on having information on other random variables

in the BN. Exact inference is however only possible for discrete BNs and a number of hybrid/continuous special cases. A straightforward approach for inference in continuous or hybrid BNs is to discretize the continuous random variables. In reliability analysis, one is typically interested in probabilities of rare (failure) events. In BNs representing such problems, discretization is therefore critical. An efficient discretization procedure for reliability problems, where the performance of the system is described through a physical model, is developed. This procedure is based on finding the optimal discretization for an approximate reliability problem, which is linear in standard normal space, through optimization. Based on this a heuristic is derived. The proposed discretization procedure is applied to an example from the field of civil aviation. In particular a prototype of a warning system that can be used to prevent runway overrun accidents of landing aircrafts is developed.
Sampling based approximate inference is an alternative to discretization of continuous nodes in BNs. Estimating probabilities of rare events through standard Monte Carlo sampling is however computationally costly. Thus advanced sampling techniques, such as subset simulation (SuS) have been developed in the field of structural reliability. In this thesis the applicability of SuS to (Gibbs-) sampling based inference in BNs is briefly investigated. The first results obtained are promising.

Zusammenfassung

Zu den wichtigsten Aufgaben von Ingenieuren gehört es, die Zuverlässigkeit menschengemachter Systeme zu gewährleisten. Dabei gilt es, Unsicherheiten über die auf das System einwirkenden Lasten, über die Widerstandsfähigkeit des Systems sowie über die relevanten menschlichen Einflussfaktoren zu berücksichtigen. Probabilistische Modelle stellen eine Möglichkeit dar, um mit diesen Unsicherheiten besser umgehen zu können. Bayes'sche Netze bilden ein Modellierungswerkzeug mit großem Potential für die Zuverlässigkeits- und Risikoanalyse. Obschon Bayes'sche Netze bereits erfolgreich im Rahmen von Zuverlässigkeitsanalysen im Ingenieurwesen eingesetzt wurden, bestehen nach wie vor offene Fragestellungen und Probleme im Bezug auf deren generelle Anwendbarkeit. Viele dieser Fragen werden im Rahmen dieser Arbeit behandelt.

Generell kann die Zuverlässigkeitsanalyse mit Bayes'schen Netzen in drei Schritte aufgeteilt werden. Diese sind: Erstellung des qualitativen Modells, Quantifizierung des Modells und zuletzt Inferenz im Modell. Der Fokus in dieser Arbeit liegt auf dem ersten sowie dem letzten dieser Schritte. Zunächst wird eine allgemein Klassifizierung von Ansätzen zur qualitativen Modellierung von Bayes'schen Netzen vorgeschlagen. Da dies in der Ingenieurpraxis oftmals nur für den Einzelfall betrachtet wird. Zum Thema Inferenz in Baye'schen Netzen werden zwei Methoden vorgeschlagen, die zur Berechnung von Wahrscheinlichkeiten seltener Ereignisse im Rahmen von Bayes'schen Netzen dienen.

Im Bayes'schen Netz repräsentieren gerichtete a-zyklische Graphen die (kausalen) qualitativen Abhängigkeiten zwischen den Zufallsvariablen des Modells. Solch eine graphische Darstellung der Abhängigkeiten spielt eine wichtige Rolle dabei ein Modell nachvollziehbar zu machen, und somit auch seine Gültigkeit sicherzustellen. Aus diesem Grund ist die Herleitung des qualitativen Modells ein wichtiger Bestandteil des gesamten Modellierungsprozesses. Im Rahmen der Arbeit werden verschiedene Ansätze, die im Bereich der Risiko- und Zuverlässigkeitsanalyse zur Herleitung des qualitativen Modells Anwendung finden, kategorisiert. Für die Kategorisierung werden relevante Publikationen aus dem Gebiet der ingenieurmäßigen Risiko und Zuverlässigkeitsanalyse berücksichtigt. Vier Ansätze werden hier unterschieden, nämlich: (1) Ableitung der Struktur ausgehend von anderen probabilistischen Modellen. (2) Herleitung ausgehend von allgemeinen Modellen. (3) Automatisches Lernen der Struktur basierend auf Daten. (4) Ableitung der Struktur aus Fachexpertenwissen. Es werden die Schwierigkeiten, die im Zusammenhang mit den Ansätzen üblicherweise auftreten, disku-

tiert. Darauf aufbauend wird ein Ansatz zur Herleitung von Modellen für die menschliche Zuverlässigkeitsanalyse vorgeschlagen. Dieser Ansatz basiert darauf, Bayes'sche Netzwerkstrukturen direkt aus der relevanten Literatur aus dem Gebiet der Psychologie abzuleiten. Beispielhaft wird der Ansatz auf einen Crew Fehlertypen aus der IDHEAS Methode zur menschlichen Zuverlässigkeitsanalyse angewandt. Die IDHEAS Methode wird derzeit von der U.S. Nuclear Regulatory Commission (NRC) entwickelt. Darüber hinaus wird ein Ansatz zur Quantifizierung des Bayes'schen Netzes, basierend sowohl auf Daten als auch auf Expertenschätzungen, vorgestellt.

Inferenz-Algorithmen werden dazu verwendet, die marginale/bedingte Wahrscheinlichkeitsverteilung einer Zufallsvariable, möglicherweise bedingt auf Beobachtungen anderer Zufallsvariablen im Bayes'schen Netz, zu ermitteln. Die exakte Verteilung kann dabei nur für Bayes'sche Netze, deren Variablen diskret sind, sowie für einige kontinuierliche/hybride Spezialfälle, berechnet werden. Ein naheliegender Ansatz zur approximativen Inferenz in kontinuierlichen/hybriden Bayes'schen Netzen ist daher, die kontinuierlichen Variablen zu diskretisieren. Da man im Bereich der Zuverlässigkeitsanalyse üblicherweise an der Bestimmung kleiner (Versagens-)wahrscheinlichkeiten interessiert ist, ist Diskretisierung von Bayes'schen Netzen in diesem Kontext kritisch. Im Rahmen dieser Arbeit wird ein effizienter Diskretisierungsansatz für Zuverlässigkeitsprobleme bei denen das Systemverhalten durch ein physikalisches Modell beschrieben ist, entwickelt. Dieser beruht darauf, dass für angenäherte Probleme, die im Standardnormalraum linear (und damit analytisch lösbar) sind, optimale Diskretisierungsschemata ermittelt werden. Aus den Optimierungsergebnissen, wird eine Heuristik abgeleitet. Der vorgeschlagene Diskretisierungsansatz wird auf ein Beispiel aus dem Bereich der Luftfahrt angewandt. Es wird der Prototyp eines Warnsystems entwickelt, das dazu dient, Unfälle zu vermeiden, die vom Überschießen der Landebahn bei der Landung herrühren.

Approximative Inferenz mithilfe von Samplingmethoden stellt eine Alternative zur Diskretisierung von kontinuierlichen Variablen im Bayes'schen Netz dar. Die Bestimmung kleiner Wahrscheinlichkeiten mit Hilfe einer standardmäßigen Monte-Carlo simulation ist jedoch (oftmals zu) rechenintensiv. Aus diesem Grund wurden im Bereich der Strukturzuverlässigkeit neue Samplingmethoden entwickelt. Ein Beispiel dafür ist die sogenannte Subset Simulation. In dieser Arbeit wird die Integrierbarkeit von Subset Simulation in die Sampling basierte Inferenz in Bayes'schen Netzen untersucht. Die ersten Resultate sind dabei vielversprechend.

Acknowledgment

This thesis would not have been possible, if it had not been for a number of individuals.
I would therefore like to express my deepest gratitude to them.
Above all I would like to thank my advisor Prof. Dr. Daniel Straub not only for giving me the opportunity to do a PhD but also for providing me with his support and invaluable advise throughout the last 3 1/2 years.
I would furthermore like to thank my colleagues at ERA group for making my working environment so enjoyable. To Dr. Katrina Groth I owe my gratitude for giving me the opportunity to spend 3 month working at Sandia National Laboratories. Not only was she a great advisor but she also made my stay in Albuquerque a great experience.
Prof. Dr. Florian Holzapfel I would like to acknowledge for agreeing to be my mentor during the time of my thesis. Furthermore I would like to thank Prof. Dr. Iris Tien and Prof. Dr. Paolo Franchin for agreeing to be examiners of this thesis.
Moreover I would like to express my gratitude to my friends and especially my parents and my family for their support, and encouragement not only during my PhD. Finally I would like to thank my girlfriend Carina for her constant love and support.

Kilian Zwirglmaier,
February 12, 2017

Contents

1 Introduction **1**

1.1 Reliability analysis 1

1.1.1 Purpose 1

1.1.2 Engineering systems and the need for probabilistic modeling 2

1.1.3 Probabilistic modeling of complex engineering systems 3

1.2 Bayesian networks 4

1.3 Scope of the thesis 6

1.4 Outline of the Thesis 7

2 Theoretical background **9**

2.1 Bayesian networks 9

2.1.1 The DAG 9

2.1.2 d-separation criteria 9

2.1.3 Local conditional probability distributions and the chain rule 10

2.2 Reliability engineering 11

2.2.1 Structural reliability 11

2.2.2 Human reliability analysis (HRA) 15

I MODEL ELICITATION **19**

3 Structure elicitation **21**

3.1 BN structures derived from other probabilistic models 22

3.1.1 Fault trees and general system reliability models 22

3.1.2 Event trees 24

3.1.3 Random processes, random fields and joint probability distributions . 25

3.1.4 Other models 27

3.1.5 Application example 28

3.2 Structure elicitation based on general models 29

3.2.1 Representation of a general model in a BN 30

3.2.2 Simplification of a BN through node removal 30

3.2.3 Parent divorcing 31

3.2.4 Causal independence representation 31
3.2.5 Application example 32
3.3 Data based structure learning 32
3.3.1 Implementation 34
3.3.2 Application example 35
3.4 BN structures based on domain expert knowledge 36
3.4.1 Modeling approach 36
3.4.2 Relation to other structure elicitation approaches 38
3.4.3 Application example 39
3.5 Combination of the approaches 39
3.6 Discussion 41

4 Quantification **43**
4.1 Quantification based on data 43
4.2 Quantification based on experts 44
4.3 Combination of expert estimates and data 44
4.3.1 Bayesian updating of parameters 45
4.3.2 Elicitation of prior beliefs for binary nodes 47
4.4 Nodes representing engineering models 48

II INFERENCE **51**

5 Introduction to inference **53**
5.1 Exact inference algorithms for discrete BNs 54
5.1.1 Variable elimination algorithm 54
5.1.2 Junction tree algorithm 55
5.2 Computational limitations of exact inference 58

6 Efficient discretization for rare events **59**
6.1 Methodology 59
6.1.1 Treatment of a reliability problem in a BN 59
6.1.2 Discretization of basic random variables 60
6.1.3 Efficient discretization 63
6.2 Development of an efficient discretization procedure 66
6.2.1 Optimization of the first order reliability method (FORM) approximation 66
6.2.2 Parametric function of optimal discretion frame 69
6.2.3 Summary of the proposed procedure 70
6.3 Verification examples 70
6.3.1 Verification example I 70
6.3.2 Verification example II 72

6.4 Discussion 74

7 Sampling based inference **79**
7.1 Forward sampling, rejection sampling and likelihood weighting 79
7.2 Markov chain Monte Carlo (MCMC) sampling 81
7.2.1 The Metropolis-Hastings algorithm 82
7.2.2 Gibbs sampling 82
7.2.3 Challenges of MCMC 83
7.3 BN sampling for rare events 85
7.3.1 SuS in the context of BNs 85
7.3.2 A simple verification example in OpenBUGS 86
7.3.3 Discussion of the sampling approach for rare events 87

III APPLICATIONS **89**

8 Reliability analysis for runway overrun **91**
8.1 Physical model of runway overrun (RWO) 92
8.2 BN model 94
8.3 Results 97
8.4 Discussion 98

9 Elicitation of BNs for HRA **99**
9.1 Development of a BN structure for the CFM 99
9.1.1 Summary of approach and models 101
9.1.2 BN model of original IDHEAS decision tree (DT) 101
9.1.3 BN model with full causal details 102
9.1.4 BN model reduction to facilitate IDHEAS-like quantification 104
9.1.5 Discussion/ implications of the models 104
9.2 CFM BN quantification 106
9.2.1 Quantifying the BN model based on the original DT 106
9.2.2 Quantification of BN model of DT with performance influencing factor (PIF) specification nodes 107
9.3 Updating the parameters of the BN with data 110
9.4 Example results with the “critical data misperceived” BN 113
9.5 Discussion 114

IV CONCLUSION **117**

10 Concluding remarks and outlook **119**
10.1 Concluding remarks 119

10.2 Main contributions of this thesis 122
10.3 Outlook 123
10.3.1 Promoting applications of BNs in industry 123
10.3.2 Exact inference algorithms for large system reliability problems 123
10.3.3 Sampling based inference 124
10.3.4 Combination of exact and approximate inference algorithms 124

V APPENDICES 125

A conditional probability tables (CPTs) of the BN in Fig. 3.12a 127

B Questions defining the PIFs 129

C Quantification of PIF nodes 131
C.1 Deterministic rule quantifying the node HSI/environment: 131
C.2 Deterministic rule quantifying the node training: 131
C.3 Deterministic rule quantifying the node workload: 132

D Survey for eliciting probabilities of PIF specification nodes 133

List of Figures 137

List of Tables 141

ABC approximate Bayesian computation

AIC Akaike information criterion

BIC Bayesian information criterion

BN Bayesian network

CDF cumulative distribution function

CFM crew failure mode

CLG conditional linear Gaussian model

CPD conditional probability distribution

CPT conditional probability table

CRT crew response tree

DAG directed acyclic graph

DT decision tree

EM expectation maximization

ET event tree

FORM first order reliability method

FT fault tree

FMEA failure mode and effects analysis

hBN hybrid BN

HFE human failure event

HEP human error probability

HRA human reliability analysis

HSI human system interface

IATA international air transport association

ID influence diagram

IDHEAS integrated decision-tree human event analysis system

LSF limit state function

MCMC Markov chain Monte Carlo

MCS Monte Carlo simulation

MoTBF Mixtures of truncated basis functions

MTE mixture of truncated exponentials

NBC naïve Bayesian classifier

NRC Nuclear Regulatory Commission

PDF probability density function

PIF performance influencing factor

PMF probability mass function

PRA probabilistic risk assessment

RBD reliability block diagram

RWO runway overrun

SACADA scenario authoring, characterization, and debriefing application

SADT structured analysis and design technique

SORM second order reliability method

SuS subset simulation

TAN tree augmented naïve Bayes

Chapter 1

Introduction

1.1 Reliability analysis

1.1.1 Purpose

As early as the 18th century BC an ancient Babylonian law (the codex Hammurabi) stated: *"If a builder build[s] a house for someone, and does not construct it properly, and the house, which he built fall[s] in and kill[s] its owner, then that builder shall be put to death."* (King, 2008). Already at that time, the reliability was an important requirement for anthropogenic structures. And it was the responsibility of the builder (respectively the engineer) to ensure reliability.

At that time, approaches to ensure reliability of engineering systems were rather empirical and it was not until the 20th century that reliability was described in a formal and mathematical way. In the 20th century, more formal approaches to reliability were promoted by large-scale production of engineering systems, increasingly complex systems and higher requirements to safety. These approaches were not least motivated by limited resources, which had to be allocated efficiently. Important milestones in this respect include: The use of probability theory for reliability analysis of the Bell telephone system in the 1920s (Zachmann, 2014), efforts to increase the reliability of military equipment in the 1940s and 50s (Cruse, 1997), and Rasmussen's (nuclear) reactor safety study WASH–1400 in 1975 that extensively applied probabilities (NRC, 1975).

While it would be desirable to have engineering systems that do not fail with absolute certainty, this is not possible in practice, since systems are always subjected to (possibly large) uncertainties. In general neither loads acting on the system (e.g. future wind load, earthquakes or users acting on the system) nor its exact properties are known with certainty (e.g. material properties of a engineering structure, behavior of operators). Probability theory provides the means for describing these uncertainties in a quantitative manner. As a result, the reliability of a system has to be expressed probabilistically as well. The reliability of an engineering system is defined as the probability of it not failing i.e. $R_Z = 1 - \Pr(F)$. Failure (F) is the event of a system being unable to fulfill some "acceptable level of performance" (Stewart

and Melchers, 1998). Reliability analysis is thus concerned with computing the probability of failure of a system. Motivations for performing a reliability analysis include:

- Assessing whether the reliability of a system is acceptable (e.g. does the probability of a total loss comply with the goals set by the regulatory body?). In such cases the quantity of interest is the reliability (or probability of failure) itself.
- Reliability analysis is part of a risk analysis, whose objective is to calculate the expected adverse consequences (e.g. to calculate the risk of a nuclear power plant, its failure probability needs to be estimated).
- Reliability analysis is part of a decision analysis, where the objective is to determine the optimal decision alternative according to the basic principle of utility theory (Raiffa and Schlaifer, 1961) (e.g. failure probabilities need to be calculated to determine the optimal design among a number of design alternatives or to determine the optimal inspection policy for a system).

Assumptions are inevitable when it comes to reliability engineering. In particular the engineering model, based on which the reliability analysis is performed, can only be an approximation of the real system. Also the stochastic models used to model the uncertain factors associated to the system are always based on assumptions. For a reliability analysis, it is important that these assumptions are made with care.
Independent of the context in which a reliability analysis is performed, a sensitivity analysis (Saltelli et al., 2000) should accompany a reliability analysis. The objective of a sensitivity analysis is to determine the sensitivity of the outcome of the reliability analysis with respect to changes in the assumptions.

1.1.2 Engineering systems and the need for probabilistic modeling

The reliability requirements for critical engineering systems are very high. Due to the small failure probabilities of modern engineering systems, the task of assessing their reliability is difficult. Examples of typical failure probabilities for engineering systems are in the range (Bedford and Cooke, 2001):

- $< 10^{-9}$ per flight hour: Probability of a catastrophic failure according to safety goals in the aviation industry
- 10^{-3} to 10^{-6}: Failure probabilities of marine structures for different accident types
- $< 10^{-4}$ per year: Probability of a large scale core melt in a nuclear power plant

If the compliance of the aviation industry with the above safety goal was to be verified purely on observed frequencies a large number of flights would need to be observed. For instance to observe one catastrophic failure one needs to wait on average 10^9 flight hours. To get an

estimate of the probability of failure, which allows for a certain level of confidence, a large number of such catastrophic failures needed to be encountered. Due to the large number of flights that are operated on a daily basis, it may be actually possible to estimate the reliability level for the complete aviation industry purely based on observed failure frequencies. In practice, one may however be interested not only in the safety standards of the aviation industry but in the safety standards of a specific airline or in the probability of failure for a specific route. Based on such information, an airline could decide whether to introduce new safety measures or whether to abandon flying to certain destinations. Estimating such probabilities through observed frequencies is typically hindered by data scarcity.

It is easy to find similar examples from different engineering disciplines. For instance data scarcity typically hinders estimating the probability of failure of a specific bridge (with specific design, and specific loads acting on it), while estimating the probability of failure of a bridge in general may be possible through observed frequencies. To nevertheless be able to estimate e.g. the probability of a total loss for a specific route or the probability of failure of a specific bridge, a different approach is required. To this end engineers can develop probabilistic models.

1.1.3 Probabilistic modeling of complex engineering systems

Probabilistic models of engineering systems require a definition of the performance of the system combined with a description of the joint probability distribution of the random variables that describe the uncertainties associated to the engineering system. The means by which this is achieved are quite diverse and depend largely on the considered field. A probabilistic model is always an abstraction of the real engineering system. The necessary degree of abstraction of a model depends on the purpose of the model but also on the complexity of the real system. For complex systems a large number of assumptions may be necessary to make probabilistic modeling feasible. Many modern engineering systems are highly complex, meaning that they consist of a large number of components, which are highly interdependent. The components that need to be accounted for may be hardware or structural components but also software and human (Zio, 2009).

In particular, human elements can add a large degree of complexity to engineering systems. Considering human error events in a reliability analysis is essential for many systems, since around 80% of all industrial accidents are due to human error (e.g. McCafferty, 1995; Rothblum, 2000; Shappell and Wiegmann, 2012). The field of human reliability analysis (HRA) is concerned with estimating human error probabilities. Large uncertainties are associated to human error, since the respective cognitive processes are complex and not well understood. In consequence the models that are used in HRA are oversimplifications of reality (Bedford and Cooke, 2001) and the quantitative estimates obtained from them should be handled with care (French et al., 2011). So far HRA methods are mainly developed and applied in the context of nuclear power plants (Kirwan, 1994) and a few other industries, such as aviation (Shappell

and Wiegmann, 2012) and chemical (Schneider, 2010).

Modeling complex systems requires appropriate tools. Fault trees (FTs) represent an example of a popular, traditional probabilistic modeling tool for reliability engineering. Using FTs, an engineering system needs to be decomposed into its components by means of boolean logic. Capturing the interdependencies in complex engineering systems in this way sufficiently well is often difficult. Other traditional probabilistic modeling tools have similar shortcomings and limitations e.g. one event tree (ET) can represent only one initiating event and dependencies between different ETs or initiating events are difficult to model (Bedford and Cooke, 2001). More flexible modeling tools are thus needed to adequately represent complex engineering systems. BNs represent such a modeling tool, which has a number of advantages over the traditional tools.

1.2 Bayesian networks

BNs have become increasingly popular in the last decades and have been successfully applied as a tool for reasoning under uncertainty in various fields including machine learning (Bishop, 2006), medicine (Lucas, 2001), law (Fenton and Neil, 2000) and environmental modeling and management (Aguilera et al., 2011; Barton et al., 2012). BNs have also been applied as a modeling tool for reliability engineering (e.g. Torres-Toledano and Sucar, 1998; Friis-Hansen, 2000; Langseth and Portinale, 2007; Straub and Der Kiureghian, 2010b).

BN enable an efficient representation of the joint probability distribution of a number of random variables. In a BN, the overall dependence structure between the variables is represented through a directed acyclic graph (DAG). The conditional and the marginal distributions of the random variables in the BN are quantified by means of local conditional probability distributions. As an example, consider the BN in Fig. 1.1. The DAG shows that a storm event can cause both high wind speeds and rain. Depending on the windspeed and the roof condition, the roof may be damaged by the wind. Damage to the roof and rain both contribute to the damage to a house.

To quantify the BN, a local probability distribution is assigned to each node. These local probability distributions are marginal distributions for nodes that do not have parents (*Storm event* and *Roof condition*) and probability distributions that are conditional on the parents for nodes that do have parents.

The DAG of a BN represents (conditional) independence properties. For example, in Fig. 1.1, the *roof condition* is independent of *wind speed* as long as no evidence on the variables *roof damage* or *damage to house* is available. These independence properties enable an efficient representation of the joint distribution of all variables in the BN. Without making use of any independence properties the joint probability distribution of the variables,

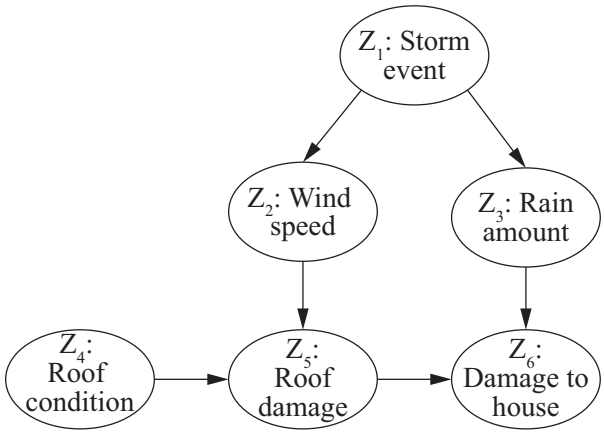


Figure 1.1: Example of a BN modeling the damage of a house due to a storm event.

$\mathbf{Z} = [Z_1, Z_2, \ldots, Z_6]$, from the BN in Fig. 1.1 can be written by means of the chain rule as:

$$p(\mathbf{Z}) = p(Z_1) \cdot p(Z_2|Z_1) \cdot p(Z_3|Z_1, Z_2) \cdot p(Z_4|Z_1, Z_2, Z_3) \cdot p(Z_5|Z_1, Z_2, Z_3, Z_4) \cdot p(Z_6|Z_1, Z_2, Z_3, Z_4, Z_5) \tag{1.1}$$

By making use of independence properties from the DAG, the joint probability distribution can be represented more efficiently as:

$$p(\mathbf{Z}) = p(Z_1) \cdot p(Z_2|Z_1) \cdot p(Z_3|Z_1) \cdot p(Z_4) \cdot p(Z_5|Z_2, Z_4) \cdot (Z_6|Z_3, Z_5) \tag{1.2}$$

This has positive effects both on the ease of model elicitation as well as the computational cost for inference. The features that make BNs a powerful tool in the context of a reliability analysis are:

Graphical representation

As stated above, probabilistic modeling typically requires a large number of assumptions. In addition to the assumptions about the stochastic models used to quantify the model, assumptions must be made on the general model structure. It is pivotal to present these assumptions to stakeholders and other domain experts, so that they can challenge them and understand the limitations of the models. As already stated, in a BN the general model structure is represented through a DAG. Although not required by BN theory, it is typically desirable for DAGs to represent causality. Since causal DAGs are well in line with human reasoning, the overall dependence structure is traceable even for people who are not experts in the area of probabilistic modeling (Pearl, 1988).

Modular quantification

The local probability distributions represent the quantitative part of the BN. Since quantitative data about engineering systems is typically scarce, it is desirable for a probabilistic

modeling tool to only require a limited amount of data. In the context of BNs, the number of parameters that need to be elicited and thus also the amount of quantitative information is reduced by making use of independence properties. For example, assuming that all nodes are binary, for the BN in Fig. 1.1, elicitation of 14 parameters is required (cf. Eq. 1.2) whereas quantification of the same model without independence assumptions (Eq. 1.1) requires 63 parameters.
A consequence of the scarcity of quantitative information is that all sources of information need to be combined to arrive at sound probability estimates. Because in a BN each node is quantified separately through a local (conditional) probability distribution, the BN lends itself to such a combination of information sources. Moreover Bayesian updating can be applied to combine different sources of information, when quantifying a single node.

Bayesian updating

Often observations on some of the variables in the model become available. These observations potentially change one's belief about other variables in the model. Bayesian updating can be used to compute probability distributions of a variable conditional on observing a set of other variables. Inference algorithms that enable efficient Bayesian updating exist for BNs. While for BNs with only discrete random variables and for a couple of continuous/hybrid special cases inference can be performed exactly, for general BNs inference can only be performed approximately e.g. through sampling.

1.3 Scope of the thesis

The features stated above motivate the use of BNs for probabilistic modeling in the scope of reliability engineering. Nevertheless there are challenges with respect to the application of BNs to reliability engineering. The objective of this thesis is to address some of these. The first part of this thesis is dedicated to the elicitation of the BN model. Typically the process of eliciting the qualitative BN structure (the DAG) is done in a rather ad-hoc manner. To overcome this, a classification of structure elicitation approaches is proposed in chapter 3. It is shown that in many instances, the BN elicitation can be standardized to the same extend as modeling with traditional tools. For more complex models, for which the flexibility of BNs is key, model elicitation requires a larger degree of modeling expertise.
As stated in the previous section, BNs enable straightforward Bayesian updating with new information. While this is true for discrete BNs that are not excessively large, inference can be challenging for general BNs. Thus, the second part of this thesis deals with inference in BNs. Many models encountered in reliability engineering are continuous or hybrid rather than purely discrete. For these models, exact inference is typically not possible. An approach to approximate inference in such cases is to discretize the continuous/hybrid model and perform exact inference for the discretized BN. Since exact inference is also limited by computational

power, discretization should be performed efficiently i.e. the discretization error should be minimized with only a minimum number of intervals. An efficient discretization procedure for reliability problems, where the performance of the system is described through an engineering model, is proposed in this thesis.
As an alternative to discretization, sampling methods can be used to approximate inference in general BNs. A short presentation of the most important sampling based BN inference algorithms is provided in this thesis. Rare events are challenging for all presented sampling based inference algorithms. The reason being that the number of samples required to estimate small probabilities is large, and the computational effort for generating these samples can quickly become unfeasible. In order to overcome this issues, efficient sampling techniques, that require a smaller number of samples to estimate small probabilities, have been developed in the field of structural reliability. One representative of these methods is subset simulation (SuS). The possibilities of combining SuS with Gibbs sampling, a popular sampling based inference algorithm, are briefly discussed in this thesis. While first convergence studies are presented, convergence should be further investigated in the future.
The methods presented in this thesis are demonstrated for a number of small examples and two larger applications.

1.4 Outline of the Thesis

Chapter 2 provides a short introduction to the principles from BN theory and reliability engineering that are required as a background for the topics treated in this thesis. The remainder of the thesis consists of four parts. Part I deals with BN elicitation, part II with inference in BNs and in part III two applications are presented. In the last part, overall concluding remarks on the topics covered in this thesis are provided together with a short outlook.
Part I:
In chapter 3, a classification of BN structure elicitation approaches is proposed. This classification is based on a review of existing literature from the field of engineering risk analysis and reliability. Four different approaches are distinguished, these are structure elicitation based on existing probabilistic models, structure elicitation based on general models, data-based structure learning and structure elicitation based on domain expert knowledge. Relations between these structure elicitation approaches are pointed out and typically challenges associated to the different approaches are discussed.
Quantification of the BNs is discussed in chapter 4. Addressed are quantification based on data and based on expert elicitation as well as a combination of them. For nodes, for which engineering models are available, quantification based on these models is discussed.
Part II:
In chapter 5, exact inference for discrete BNs is discussed. Two of the most important representatives of exact inference algorithms for discrete BNs are introduced in this context,

namely the variable elimination and the junction tree algorithm. While these algorithms can be applied for exact inference in discrete BNs, by discretizing the continuous random variables they can also be applied for approximate inference in hybrid/continuous BNs. The discretization scheme used determines the approximation error. Therefore discretization is critical. To this end, an efficient discretization procedure for reliability problems, where the performance is described through an engineering model, is proposed in chapter 6. Sampling based algorithms represent an alternative to discretization for approximate inference. In chapter 7, the most important sampling based inference algorithms are introduced. Estimating probabilities of rare events by means of these algorithms is computationally costly. An approach for overcoming this issue by combining sampling based inference with SuS is presented.

Part III:

In chapter 8, a BN based prototype of a runway overrun early warning system is proposed. The BN is developed from an existing physical model. The continuous basic random variables of the structural reliability problem are discretized using the the discretization procedure from chapter 6. Chapter 9 presents a framework for structure elicitation and quantification of BNs for HRA. This framework is based on the theory discussed in chapters 3 and 4. An example from the IDHEAS HRA method is used to demonstrate the approach.

Part IV:

In chapter 10 an overall discussion of the topics that are covered in this thesis is provided together with an outlook.

Chapter 2

Theoretical background

This chapter provides a brief introduction to the most important theories that form the basis of the remainder of this thesis.

2.1 Bayesian networks

The description of BN theory here is limited to the parts that are necessary as a background for the remainder of this thesis. For a deeper introduction the interested reader is referred to the relevant textbooks. These include (Jensen and Nielsen, 2007; Koller and Friedman, 2009; Kjærulff and Madsen, 2013).

2.1.1 The DAG

The qualitative dependence structure of a BN is encoded in a directed acyclic graph (DAG) G that consists of a set of vertices (nodes) V and a set of edges (links) E, i.e. $G = (V, E)$[1]. The vertices represent variables and the edges dependencies among these variables. An example of a DAG with vertices $V = \{X_1, X_2, \ldots, X_5\}$ and edges
$E = \{[X_1, X_3], [X_1, X_4], [X_2, X_5], [X_3, X_5], [X_4, X_5]\}$ is shown in Fig. 2.1. A directed graph fulfills the requirement for acyclicity, if for non of the vertices $X_i \in V$ there exists a directed path, such that $X_i \to \ldots \to X_i$.
In the context of BNs family terms are used to describe relationships among nodes. For example, in the DAG of Fig. 2.1, X_5 is a child of X_2, X_3 and X_4, which in turn are its parents, $pa(X_5) = \{X_2, X_3, X_4\}$; X_1 to X_4 are ancestors of X_5, $ac(X_5) = \{X_1, X_2, X_3, X_4\}$ and X_3, X_4, X_5 are descendants of the node X_1, $dc(X_1) = \{X_3, X_4, X_5\}$.

2.1.2 d-separation criteria

The efficiency of BNs is determined by its independence structure. Independence statements are incorporated in the DAG through d-separation properties. In particular, if according

[1]Note: The terms vertices and nodes are used interchangeably in this thesis. So are the terms edges and links.

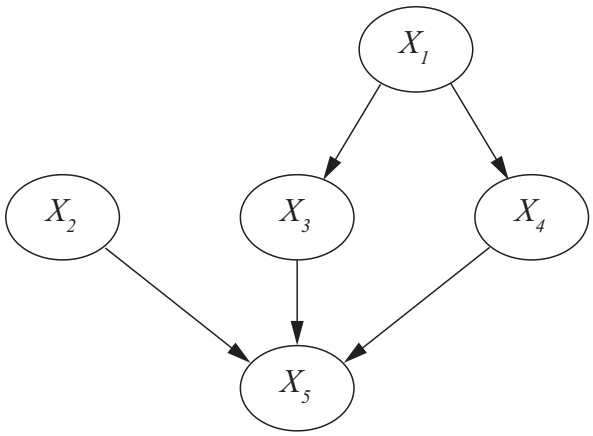


Figure 2.1: Example BN.

to their joint probability distribution, two random variables X_i and X_j are independent conditional on knowing the states of some variables $\mathbf{X^e} = \mathbf{x^e}$, then they are d-separated in the corresponding BN. There are three basic connection types in BNs, namely serial connections (e.g. $X_1 \rightarrow X_3 \rightarrow X_5$ in Fig. 2.1), diverging connections ($X_3 \leftarrow X_1 \rightarrow X_4$ in Fig. 2.1) and converging connections (e.g. $X_2 \rightarrow X_5 \leftarrow X_3$ in Fig. 2.1).
While nodes that are connected through a serial connection or a diverging connection are d-separate if the state of the intermediate node is known with certainty, nodes that are connected in converging connections are d-separated if neither the intermediate node nor any of its descendants have received any evidence.
From the d-separation properties, the concept of Markov blankets can be derived. In a BN over the nodes $\mathbf{X}$, the Markov blanket of a node X_i consists of its parents, its children and the other parents of its children. It is the minimal set of nodes $MB(X_i) \in \mathbf{X}$, such that if the states of all nodes in $MB(X_i)$ are known, X_i is d-separated from the rest of the network. For a BN, whose purpose it is to predict the state of X_i, the consequence of this is that if all random variables in X_i's Markov blanket are observable at all times, then it is enough to include $\{X_i, MB(X_i)\}$ in the model.

2.1.3 Local conditional probability distributions and the chain rule

A local conditional probability distribution (CPD) is assigned to each node in the network that represents a random variable. For a node representing a random variable X_i, with parents $pa(X_i)$, we denote this CPD as $p(x_i|pa(x_i))$. In a discrete BN, the CPD is typically represented in the form of a conditional probability table (CPT). In analogy to Eqs. 1.1 and 1.2, the well-known chain rule to represent the joint distribution $p(\mathbf{x}) = p(x_1, \ldots, x_n)$ in Eq. 2.1 can be represented more efficiently by exploiting the independence properties of the BN (Eq. 2.2).

$$p(\mathbf{x}) = p(x_n|x_1, \ldots, x_{n-1}) \cdot p(x_{n-1}|x_1, \ldots, x_{n-2}) \cdot \ldots \cdot p(x_2|x_1) \cdot p(x_1) \tag{2.1}$$

$$p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i|pa(x_i)) \tag{2.2}$$

For example, the chain rule for the BN structure in Fig. 2.1 is:

$$p(\mathbf{x}) = p(x_1) \cdot p(x_2) \cdot p(x_3|x_1) \cdot p(x_4|x_1) \cdot p(x_5|x_2, x_3, x_4) \tag{2.3}$$

2.2 Reliability engineering

While reliability is an inherent requirement for engineering systems, structured approaches for reliability assessment of engineering systems have emerged only in the 20th century.
Reliability of an engineering system is defined as the probability of it not failing (or not failing in its service life), i.e. it can be written as one minus the probability of failure. Reliability engineering can be differentiated into hardware reliability (Rausand and Høyland, 2004), software reliability (Lyu et al., 1996) and human reliability (HRA) (Swain and Guttmann, 1983; Kirwan, 1994).
For quantitative reliability analysis, engineering components or systems are often described through engineering models. This is often the case for engineering structures. Therefore the methods for estimating the reliability (respectively failure probability) based on such physical models have been mainly developed in the field of structural reliability (Melchers, 1999). An introduction to the field of structural reliability is provided in the following. It should however be noted that the methods from this field are not limited to structures as it will also become evident from the application in chapter 8. After that an introduction to the field of HRA is provided.

2.2.1 Structural reliability

[2]Since the 1970s, structural reliability methods have been developed and applied in the engineering community to estimate failure probabilities $\Pr(F)$ of components or systems, based on physical or empirical models. The performance of engineering components is described by a limit state function (LSF) $g(\mathbf{x})$, where $\mathbf{X} = [X_1; \ldots; X_n]$ is a vector of basic random variables influencing the performance of the component. By definition, failure corresponds to $g(\mathbf{x})$ taking non-positive values, i.e. the failure event is $F = \{g(\mathbf{x}) \leq 0\}$. $g(\mathbf{x})$ includes the physical or engineering model, which is often computationally demanding. The probability of failure is calculated by integrating the probability density function (PDF) of $\mathbf{X}$, $f_{\mathbf{X}}(\mathbf{x})$, over the failure domain:

$$Pr(F) = \int_{g(x)\leq 0} f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{2.4}$$

[2]The description of structural reliability and FORM analysis in this section is adapted from (Zwirglmaier and Straub, 2016a).

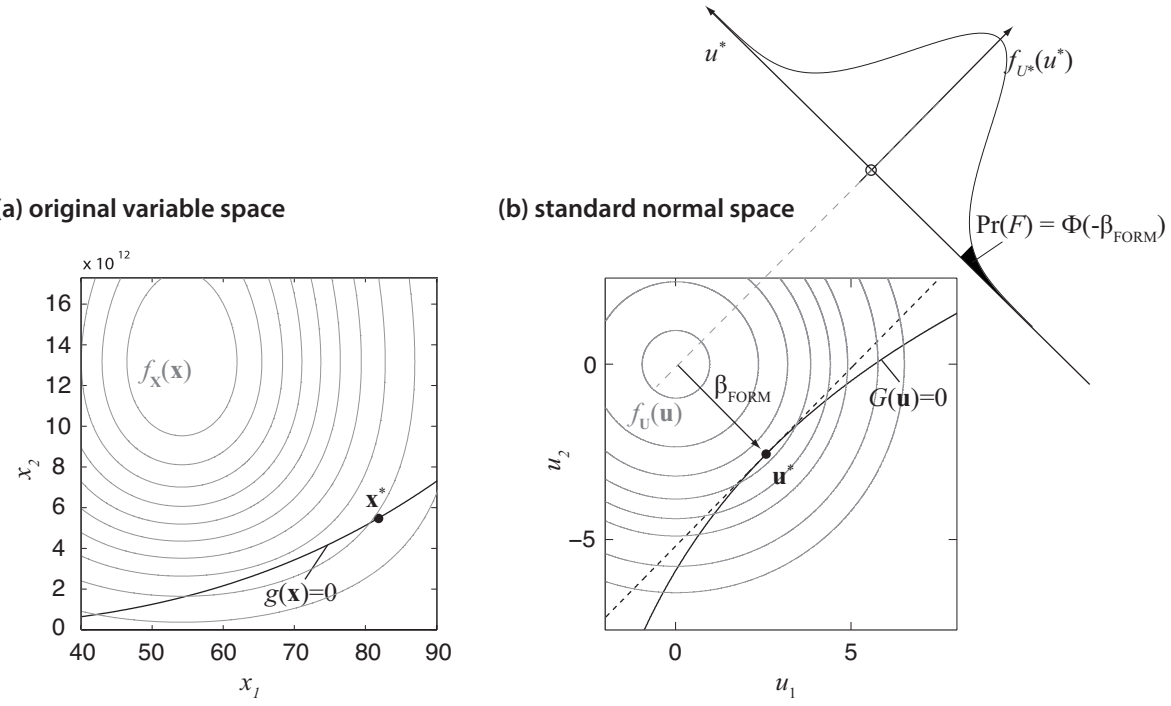


Figure 2.2: Design point and linear approximation of the limit state surface. Left side: original random variable space; right side: standard normal space (from (Straub, 2014b)).

In general the integral in Eq. 2.4 cannot be solved analytically. To this end a number of methods have been developed to approximately calculate it, these include first order reliability method (FORM) and second order reliability method (SORM) as well as a number of sampling techniques. FORM and the sampling technique SuS are introduced in the following.

The first order reliability method (FORM)

To obtain an approximation of the probability of failure through FORM, the LSF $g(\mathbf{X})$ is transformed to an equivalent LSF $G(\mathbf{U})$ in the space of uncorrelated standard normal random variables $\mathbf{U} = [U_1; \ldots; U_n]$ (Fig. 2.2). The transformation is probability conserving, so that $\Pr[g(\mathbf{X}) \leq 0] = \Pr[G(\mathbf{U}) \leq 0] = \Pr(F)$. A suitable transformation for this purpose, which is consistent with the BN, is the Rosenblatt transformation (Hohenbichler and Rackwitz, 1981). In case all basic random variables are independent, this transformation reduces to the marginal transformations: $U_i = \Phi^{-1}[F_{X_i}(x_i)]$, with Φ^{-1} being the inverse standard normal cumulative distribution function (CDF) and F_{X_i} being the CDF of X_i. In cases where conditional distributions are not readily available the Nataf transformation can be applied as an alternative.

The FORM approximation of $\Pr(F)$ is obtained by substituting the LSF in U-space $G(\mathbf{U})$ by a linear function $G_L(\mathbf{U})$, i.e. a first-order Taylor expansion of $G(\mathbf{U})$. The key idea of FORM is to choose as the expansion point the so-called design point $\mathbf{u}^*$, which is the point that minimizes $\|\mathbf{u}^*\|$ subject to $G_L(\mathbf{U}) \leq 0$. It is also known as the most likely failure point,

as it is the point in the failure domain with the highest probability density. Since all marginal distributions of the standard uncorrelated multinormal distribution are standard normal, it can be shown that the FORM probability of failure $\Pr\left[G_L\left(\mathbf{U}\right) \leq 0\right]$ is:

$$\Pr\left[G_L\left(\mathbf{U}\right) \leq 0\right] = \Phi\left(-\beta_{FORM}\right) \tag{2.5}$$

where Φ is the standard normal CDF and β_{FORM} is the distance from the origin to the design point, i.e. $\beta_{FORM} = \|\mathbf{u}^*\|$. The problem thus reduces to finding the design point $\mathbf{u}^*$. If $G\left(\mathbf{U}\right)$ is linear, the FORM solution of the probability of failure is exact, otherwise it is an approximation, which however is sufficiently accurate in most practical applications with limited numbers of random variables (Rackwitz, 2001). The linearized LSF $G_L\left(\mathbf{U}\right)$ can be written as:

$$G_L\left(\mathbf{U}\right) = \beta_{FORM} - \boldsymbol{\alpha}^T\mathbf{U} \tag{2.6}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]$ is the vector of FORM importance measures. These importance measures are defined as:

$$\alpha_i = \frac{u_i^*}{\beta_{FORM}} \tag{2.7}$$

where u_i^* is the i-th component of the design point coordinates. The α_i's take values between -1 and 1, and it is $\|\boldsymbol{\alpha}\| = 1$. α_i is 0, if the uncertainty on U_i has no influence on $\Pr\left(G_L\left(\mathbf{U}\right) \leq 0\right)$, and it is 1 or -1, if U_i is the only random variable affecting $\Pr\left(G_L\left(\mathbf{U}\right) \leq 0\right)$. When the original random variables X_i are mutually independent, the α_i's are readily applicable also in the original space, otherwise the α_i's can be transformed as described in (Der Kiureghian, 2005).

Subset Simulation

[3]The classical Monte Carlo simulation (MCS) approach provides an unbiased estimate $\hat{\Pr}\left(F\right)$ of Eq. 2.4 by generating n_S samples from the joint distribution $f\left(\mathbf{x}\right)$ and evaluating $g\left(\mathbf{x}\right)$ for each of the samples $\mathbf{x}^i$. The estimated probability of failure is then:

$$\hat{\Pr}\left(F\right) = \frac{1}{n_S}\sum_{i=1}^{n_S} I_{g(\mathbf{x})\leq 0}\left(\mathbf{x}^i\right) \tag{2.8}$$

where $I_{g(\mathbf{x})\leq 0}\left(\mathbf{x}^i\right)$ is an indicator function, which is 1 for $g\left(\mathbf{x}^i\right) \leq 0$ and 0 otherwise. This MCS approach becomes computationally unfeasible for small probabilities of failure. For a desired coefficient of variation of the estimate $\rho_{\hat{\Pr}(F)}$ the required number of samples is:

$$n_S = \frac{1 - \Pr\left(F\right)}{\rho^2_{\hat{P}_F}\Pr(F)} \tag{2.9}$$

For a small probability of failure of $\Pr\left(F\right) = 10^{-8}$ and a desired coefficient of variation of

[3]The description of Subset Simulation is adapted from (Zwirglmaier et al., 2014).

$\rho_{\hat{P}_F} = 0.1$ approximately 10^{10} LSF evaluations would be necessary. SuS proposed by (Au and Beck, 2001) overcomes this problem by expressing the failure event F as a sequence of nested events F_i.

$$F_1 \subset F_2 \subset \ldots \subset F_n = F \tag{2.10}$$

With these nested events the probability of failure can be rewritten as:

$$\Pr(F) = \Pr(F_1) \prod_{i=2}^{n} \Pr(F_i|F_{i-1}) \tag{2.11}$$

The probability of the first intermediate event, $\Pr(F_1)$, is not conditional on a previous intermediate event and can therefore be estimated with a standard MCS. All other probabilities $\Pr(F_i|F_{i-1})$ in Eq. 2.11 are conditional on a previous intermediate event, these conditional probabilities are estimated using MCMC simulation procedures. In Fig. 2.3 the MCMC approach is schematically shown for drawing samples conditional on the first failure domain F_1 to approximately compute $\Pr(F_2|F_1)$. The grey points represent the samples generated from the bivariate standard normal distribution $f(\mathbf{u})$ in the initial MCS step. Those samples $\mathbf{u}^i$, which are in Fig. 2.3 above the intermediate limit state surface (shown here as a black line) are said to be in the failure domain F_1, i.e. for them $\mathbf{u}^i \in F_1$. From each of those m intermediate failure samples, a Markov Chain of length n_S/m is generated (black samples). In these chains a new candidate state is generated conditional on the previous state of the chain. Furthermore if the newly generated candidate state $\mathbf{u}'$ fulfills the condition $\mathbf{u}' \in F_1$ it becomes the new state of the chain otherwise the previous state is repeated.

SuS is performed in the uncorrelated standard normal space (U-Space). The component wise Metropolis Hastings algorithm, which was proposed by (Au and Beck, 2001) as a MCMC algorithm for SuS, makes use of this by generating the candidate states independently for each dimension. The performance of SuS thus becomes independent of the number of dimensions of the problem. An alternative MCMC algorithm for use in SuS is proposed by (Papaioannou et al., 2015). Like the algorithm of (Au and Beck, 2001) this algorithm works in a component-wise fashion and its performance is therefore also independent of the dimensionality of the problem. However, the generation of the pre-candidate states is done, such that every pre-candidate state is accepted. Typically the intermediate failure events F_i are chosen adaptively, such that $\Pr(F_1) = \Pr(F_2|F_1) = \ldots = \Pr(F_{n-1}|F_{n-2}) = p_0$. Often a value of 0.1 is chosen for p_0. The final failure event $F_n = F$ is fixed and therefore its conditional probability cannot be chosen a priori. Eq. 2.11 is then:

$$\Pr(F) = p_0^{n-1} \Pr(F|F_{n-1}) \tag{2.12}$$

Like FORM, SuS is usually applied in U-space, therefore the random variables $\mathbf{X}$ have to be transformed to U-space.

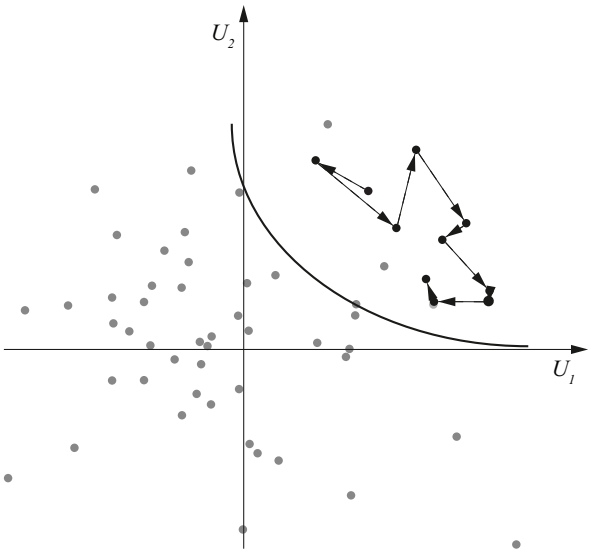


Figure 2.3: MCMC algorithm for SuS.

2.2.2 Human reliability analysis (HRA)

[4]A comprehensive probabilistic risk assessment (PRA) is an essential element of safety and reliability assurance for many complex engineering systems. The aim of the PRA is to understand the possible failure scenarios, the corresponding adverse consequences, and the failure scenarios' probabilities. Most engineering systems can be characterized as human-machine systems, in which the human operator and the technical system are interacting. For that reason it is essential for a PRA to consider not only failures of technical components but also the effect of human actions and human inaction. HRA models human elements as part of PRAs; in general through identification and quantification of human failure events (HFEs) in PRA models. A variety of methods have been developed and applied in this field to determine human error probabilitys (HEPs) corresponding to HFEs. Among the most important representatives are THERP (Swain and Guttmann, 1983), SPAR-H (Gertman et al., 2005) and ATHEANA (Cooper et al., 1996). The limitations of existing HRA methods have been widely discussed in the literature (Woods, 1990; Hollnagel, 2000; Mosleh and Chang, 2004; Sträter, 2004; Boring et al., 2007; French et al., 2011; Groth and Swiler, 2013). Two interrelated shortcomings in existing HRA methods are the limited scientific basis used to develop those methods and the use of simplified modeling techniques, which lack causal structure and quantitative traceability.

Ongoing research into human performance is addressing the first shortcoming. The scientific foundations for human reliability have been explored and documented in the work by (Whaley et al., 2012) on the psychological basis of HRA. In particular, they introduce a set of psychological failure mechanisms and proximate causes, which can lead to human failure events. Furthermore, they provide detailed insight into the factors that affect human performance

[4]The description of HRA is adapted from (Zwirglmaier et al., 2017).

(PIFs), the dependency between those factors, and the causal pathways from those factors to human errors. International data collection activities offer insight into human performance in complex engineered systems (Park and Jung, 2007; CSNI, 2012; Chang et al., 2014), which provide new opportunities to improve the quantitative basis of HRA.
The second shortcoming, the lack of causal structure and quantitative traceability, is being addressed through advanced modeling efforts. BN models have become increasingly popular within HRA as a means for addressing these shortcomings because of their ability to explicitly model cause and effect combined with the ability to incorporate information from different sources (Baraldi et al., 2015; Mkrtchyan et al., 2015). Ongoing international research has demonstrated the ability of BNs both to capture the causal relationships among PIFs and to facilitate quantification of those relationships (Groth and Mosleh, 2012; Sundaramurthi and Smidts, 2013; Musharraf et al., 2014; Podofillini and Dang, 2014). The psychological foundation has been leveraged in the development of two new HRA Methods, the IDHEAS (Integrated Decision-Tree Human Event Analysis System) method (Xing et al., 2013) and the PHOENIX method (Ekanem and Mosleh, 2014; Ekanem et al., 2016). Both IDHEAS and PHOENIX introduce the concept of CFMs, a characterization of ways that a human failure event can occur during a crew interaction with the system. Both methods include a quantitative model relating PIFs to CFMs. However, the quantitative models in IDHEAS fall short of both causal and quantitative traceability; e.g. the motivation for the exclusion of cognitive mechanisms and PIFs from the method remains unclear (Stetkar, 2014). The PHOENIX method uses a BN model for quantification, but there are no directed links from one PIF to another, and thus the causal paths from the cognitive literature are not fully captured.

Crew failure modes

CFMs are used as modules, for representing HFEs. In this thesis we propose a framework for modeling CFMs probabilistically by means of BNs. Two new HRA methods incorporate the concept of crew failure modes: the IDHEAS method developed by the U.S. NRC, and the PHOENIX method developed at the University of Maryland (Ekanem and Mosleh, 2014; Ekanem et al., 2016). PHOENIX and IDHEAS follow a similar modeling approach combining both qualitative and quantitative steps:

- Performing a qualitative task analysis and documenting crew failure paths in a crew response tree (CRT).
- Selecting applicable CFMs for each event in a CRT.
- Quantifying the individual CFMs (via Decision Trees for IDHEAS, and via BN for PHOENIX) and combining probabilities of the relevant CFMs to calculate the HEPs for each event.
- Analyzing HFE dependencies and possible recovery actions.

In both methods, the CFMs are a crucial element, which translates the concept of human errors from a psychological perspective (which may or may not have an impact on a system) into crew errors, which could lead to an HFE. IDHEAS and PHOENIX each derive their CFMs from the psychological failure mechanisms (Whaley et al., 2012). In both methods, PIFs are used to characterize the context of the task. The two methods differ in the number of CFMs used, as well as the quantification approach. IDHEAS considers 14 CFMs representing failures that are typical for human performance in nuclear power plant control rooms. The CFMs in IDHEAS are summarized in Tab. 2.1. The PHOENIX method considers 19 different CFMs. These are summarized in Tab. 2.2. In IDHEAS each CFM is quantified using a DT[5] (c.f. chapter 9). Each PIF is represented as a branch point in the DT. For simplicity, the IDHEAS developers chose to limit the number of PIFs in each DT to four.

.

Table 2.1: Crew Failure Modes used in the IDHEAS method (Xing et al., 2013)

Phase of	Plant status assessment	Response planning	Execution
Crew Failure Mode CFM	Key alarm not attended to †	Delay implementation†	Fail to initiate execution
	Data misleading or not available	Misinterpret procedure†	Fail to execute response correctly
	Premature termination of critical data collection	Choose inappropriate strategy	
	Critical data misperceived		
	Wrong data source attended to†		
	Critical data not checked with appropriate frequency		
	Critical data dismissed/discounted†		
		Misread or skip step in procedure*†	
	Critical data miscommunicated**†		

† CFM, for which data was collected.

* May occur in either 'Response Planning' or 'Execution' phases.

** May occur in any of the three phases.

In PHOENIX, the CFMs are quantified using one BN that maps relationships between PIFs and all nineteen CFMs. PHOENIX considers nine "primary" PIFs , which all directly influence each of the CFMs. PHOENIX also includes an expanded qualitative BN model, which includes

[5]Note: The applied models are referred to as decision trees in the IDHEAS report (Xing et al., 2013). However, since there are no decisions involved, the tool should be more appropriately termed event tree in a PRA sense. To be in line with the original IDHEAS report, the thesis adopts its terminology in the parts addressing IDHEAS.

approximately 20 additional PIFs that have been collapsed into the nine primary PIFs . The BN model used in PHOENIX does not directly model interdependency between the PIFs.

Table 2.2: CFMs used in the PHOENIX method.

Phase of	Information Processing	Diagnosis/Decision Making	Action Taking
Crew Failure Mode CFM	Key alarm not responded to (intentional & unintentional)	Plant/system state misdiagnosed	Incorrect timing of action
	Data not obtained (intentional)	Procedure misinterpreted	Incorrect operation of component/ object
	Data discounted	Failure to adapt procedure to the situation	Action on wrong component/ object
	Decision to stop gathering data	Procedure step omitted (intentional)	
	Data incorrectly processed	Deviation from procedure	
	Reading error	Decision to delay action	
	Information miscommunicated	Inappropriate strategy chosen	
	Wrong data source attended to		
	Data not Checked with appropriate frequency		

Part I

MODEL ELICITATION

Chapter 3

Structure elicitation

[6]In engineering risk and reliability analysis, domain experts from a variety of disciplines and experts in probability and risk assessment have to work closely together. They have to jointly develop probabilistic models, which are often multi-disciplinary. Ensuring consistency among and within the models is crucial. A common model framework is key in this process.
The BN is an increasingly popular choice for such a framework (cf. chapter 1). Its graphical structure supports the model building process among different experts and facilitates the communication of the model to stakeholders. It helps ensuring that the model is traceable, which is a basic requirement to a solid risk assessment (Bedford and Cooke, 2001). Its modular nature allows for large models to be built and evaluated, thereby incorporating existing models. The derivation of the general dependence structure requires a profound understanding of the problem as well as an understanding of the BN semantics. Once the graphical structure is developed, one can deduce from the DAG, which CPDs need to be quantified from data, expert estimates, alternative models or combinations thereof.
In the field of engineering risk analysis and reliability, the structure elicitation process, i.e. the construction of the DAG, has received little attention and is seldom presented in a systematic and formal manner. Exceptions include (Langseth and Portinale, 2007), who explain the quantitative and qualitative BN model building and inference process for reliability applications. They provide suggestions on BN elicitation based on domain experts and describe the derivation of BNs from fault trees (FTs). Furthermore (Hanea et al., 2006) discuss the elicitation of BN model from expert knowledge. In a more general context (Fenton and Neil, 2012) propose the use of generic BN substructures (or idioms) that represent common situations in BN elicitation. These can be used to elicit BN structures in a modular way. (Kjærulff and Madsen, 2013) define classes of variables in a BN and discuss how these classes typically interdepend. The resulting high-level dependence structures can support BN structure elicitation. (Conrady and Jouffe, 2015) discuss BN elicitation and distinguish data and theory as possible basis for BN model elicitation. Additionally, (Chen and Pollino, 2012) propose guidelines for BN modeling of environmental systems.

[6]This chapter is adapted from (Zwirglmaier and Straub, 2016b)

This chapter attempts to structure the qualitative model elicitation process. Based on an extensive literature review, four approaches for BN structure elicitation are distinguished, namely: (1) transformation from existing probabilistic models such as fault trees, event trees; (2) derivation from existing generic (physical or empirical) models; (3) structure elicitation based on data; (4) structure elicitation based on domain expert knowledge. These four approaches are reviewed in the following sections together with small applications from the field of engineering risk analysis and reliability. Analogies and common challenges among these approaches are highlighted. Combinations of these approaches are discussed as well as typical challenges associated with the different structure elicitation approaches.

3.1 BN structures derived from other probabilistic models

FTs and ETs are examples of probabilistic modeling tools that are popular in industry due to their simplicity and the possibility for standardized elicitation. Because BNs are more flexible, they are increasingly replacing or enhancing these models. As described in this section, it is often straightforward to replace an existing probabilistic model with a BN, but the BN may be inefficient in some cases.

3.1.1 Fault trees and general system reliability models

In system reliability, the failure (survival) of a system is generally described by a structure function ϕ (Rausand and Høyland, 2004). ϕ can be represented by FTs, reliability block diagrams, minimal cut sets, minimal link sets or other models. The structure function can be transformed into a BN, as first outlined in (Torres-Toledano and Sucar, 1998). In industrial applications, the structure function is often constructed and represented by means of fault trees (Stewart and Melchers, 1998; Rausand and Høyland, 2004). The transformation of a FT into a BN was first proposed in (Bobbio et al., 2001; Mahadevan et al., 2001) and has been described in multiple publications (e.g. Langseth and Portinale, 2007; Marquez et al., 2010; Kusz et al., 2011; Khakzad et al., 2011). To transform a FT to a BN, events from the former need to be represented through binary nodes in the latter. The states of the nodes correspond to the failure and the survival event. Component nodes are connected to the system node through a series of converging connections, which mirror the gates of the FT (cf. Figs. 3.1 and 3.3). The CPTs of the system node differ according to the gate type used in the FT representation. Fig. 3.1 shows the transformation of an AND/OR gate from a FT (respectively a parallel/series subsystem from a reliability block diagram (RBD)) in a BN. Figs. 3.2 and 3.3 illustrate the transformation of a simple example FT into its corresponding BN. Tab. 3.1 shows the CPT representing the 2-out-of-3 gate. CPTs representing FT gates are deterministic, i.e. conditional probabilities take only value 0 or 1. Whereas a basic event may appear multiple times in the FT (Tsunami in Fig. 3.3), it is represented by a single node in the BN. In this way, dependences in the system are more evident in the BN and computation

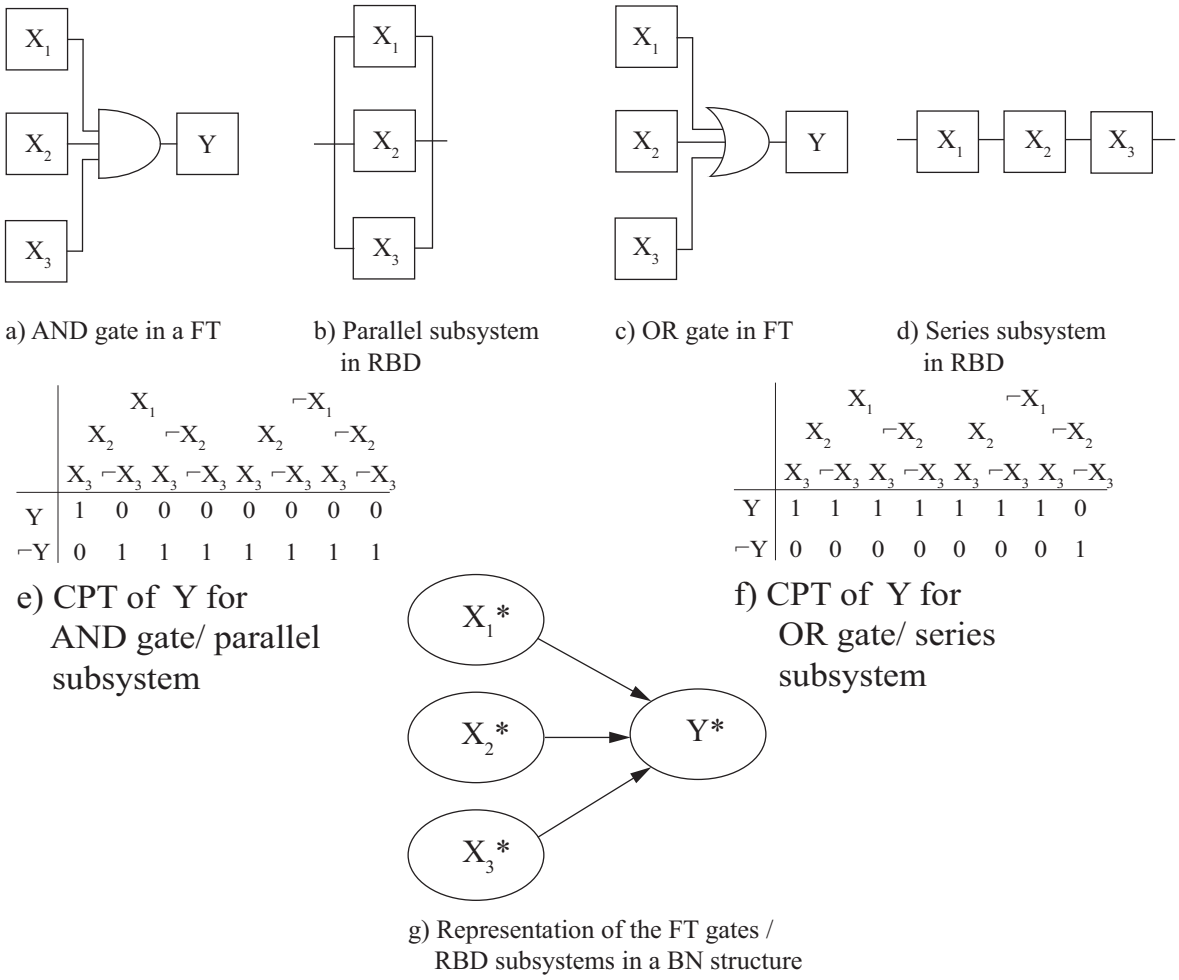


Figure 3.1: Representation of AND/OR gate from a FT respectively a series/parallel subsystem from a RBD in a BN. The events X_i and Y are represented through binary nodes X_i^* and Y^* in the BN.

is straightforward. The flexibility of the BN also allows to include additional features that are difficult to consider in the FT, such as noisy (non-deterministic) gates, sequential failures and multi-state variables (Marquez et al., 2010). An approach for representing dynamic FTs (Dugan et al., 1992) in BNs is proposed by (Boudali and Dugan, 2005). (Montani et al., 2008) represent dynamic FTs through dynamic BNs and provide a software tool (RYDYBAN) to automate this transformation. Applications of transformations of FTs into BNs include accident modeling in railways (Marsh and Bearfield, 2007) and aviation (Ale et al., 2009; Morales, 2010); a hydro-power fault diagnosis system (Jong and Leu, 2013); a model of a regasification system of liquefied natural gas on board of a floating, storage and regasification unit (Martins et al., 2014). A general representation of system reliability problems in BNs is obtained by linking all component nodes directly to the system node, which represent the structure function ϕ. This is the so-called naïve representation (cf. section 3.2), which is severely limited by the size of CPT of the system node. A more efficient alternative is provided

by transforming the minimal cut sets/ minimal link sets representation of the system into a BN, following the systematic procedure of (Bensi et al., 2013). The approach does however require optimization to identify efficient BN structures, which is also limited by system size.

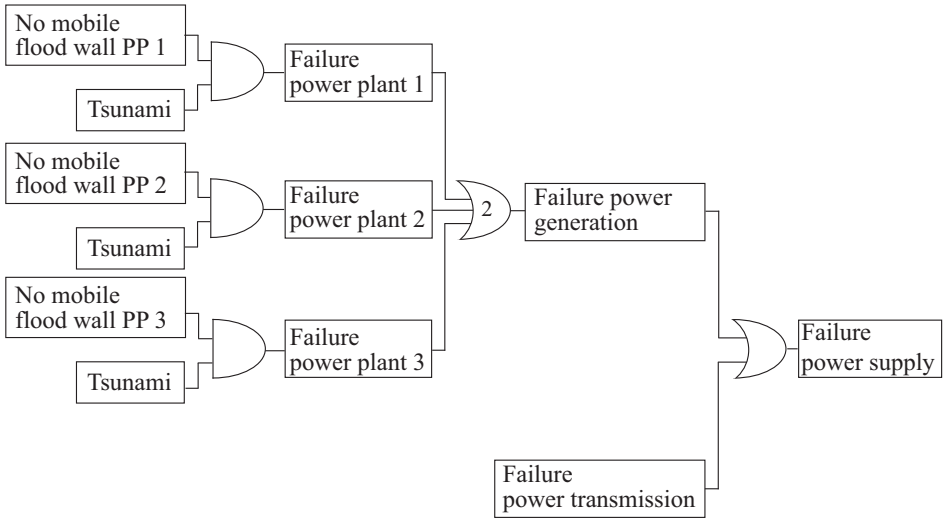


Figure 3.2: Example FT modeling power supply. Power can be supplied, if enough power is generated and the power can be transmitted. Enough power is generated if at least two power plants are working. A power plant fails, if a tsunami occurs and no mobile flood walls are erected.

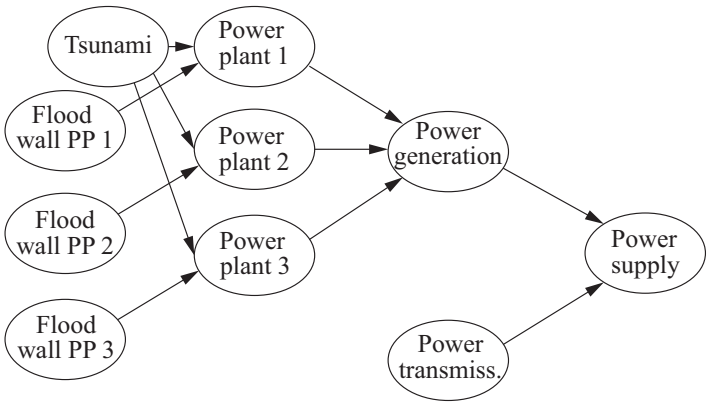


Figure 3.3: BN corresponding to the FT from Fig. 3.2.

3.1.2 Event trees

The event tree ET is a basic probabilistic modeling tool to evaluate the probability of different scenarios arising from an initial (top) event. It also enables the computation of the expected consequences associated with these scenarios. The transformation of ETs to BNs is described in (Bearfield and Marsh, 2005; Friis-Hansen, 2000).

A general ET is shown in Fig. 3.4a. The ordering of random variables $X_1, \dots X_n, Y$ from left to right typically follows the causal direction, which is here implied in the correspond-

Table 3.1: CPT of the node *Power generation* from the BN in Fig. 3.3, which represents the 2-out-of-3 gate from the FT in 3.2.

Power plant 1	Failure				Survival			
Power plant 2	Failure		Survival		Failure		Survival	
Power plant 3	Failure	Surv.	Failure	Surv.	Failure	Surv.	Failure	Surv.
Sufficient	0	0	0	1	0	1	1	1
Insufficient	1	1	1	0	1	0	0	0

ing BN shown in Fig. 3.4b. Whether a link between any two nodes in the BN is actually present can be inferred from the ET. A link from X_i to X_j is required if, and only if, $p\left(X_j|ac\left(X_j\right)\right) \neq p\left(X_j|ac\left(X_j\right) \setminus X_i\right)$, for at least one combination of X_i and X_j. I.e., the link is required if the two corresponding random variables are dependent conditional on knowing their (potential) joint ancestors and their (potential) intermediate nodes and not having received any evidence on their (potential) descendants. However, if not all branches of the ET are explicitly quantified (because some branches have no effect on Y), this test cannot be carried out. In the absence of other information, the link from X_i to X_j is omitted. This is further discussed in the application example of subsection 3.1.5. The expected consequences associated with a top event are determined from the consequence node Y, by evaluating its marginal probability distribution $p\left(y\right)$ and then computing the expected value $E\left[Y\right]$. Alternatively, utility nodes can be added to the BN. Such utility nodes are available in influence diagrams (IDs), which are an extension of the BN to enable the analysis of decisions based on the expected utility theory (Raiffa and Schlaifer, 1961).

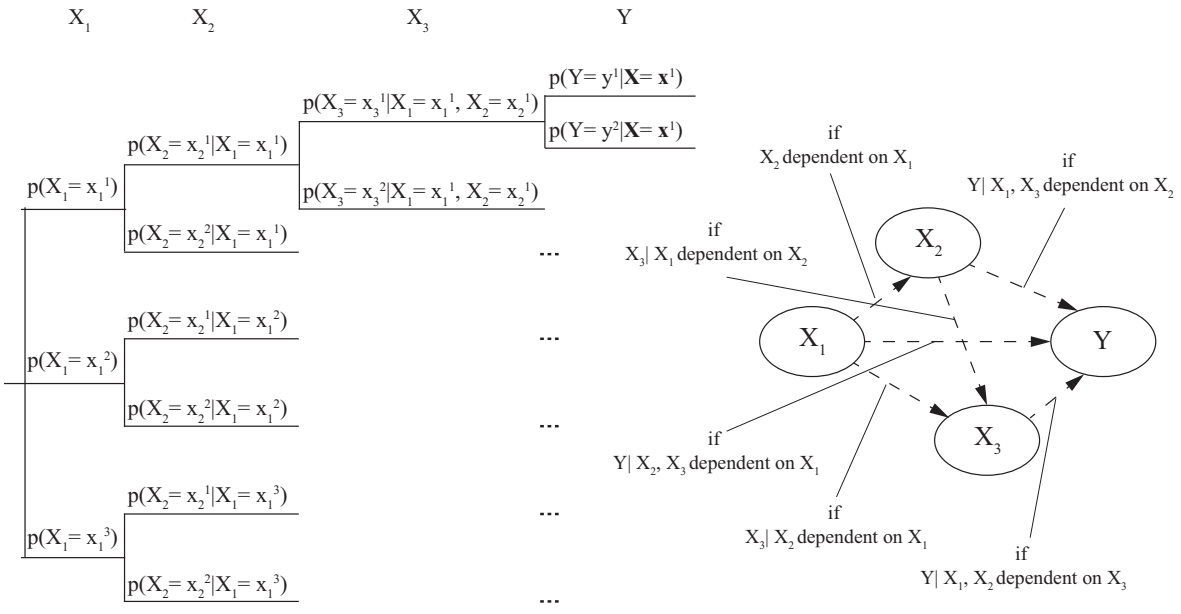
Often FTs and ETs are combined into so-called bow tie models, where the top event from the FT represents the initiating event of the ET (Andersen et al., 2004). While FTs model the occurrence of a top (failure) event starting from several initiating events, ETs model the paths from the top event (failure) to the possible final states (different consequences associated with the failure event). Transformation of bow ties to BNs is described by (Duijm, 2009; Khakzad et al., 2013).

3.1.3 Random processes, random fields and joint probability distributions

Among others natural phenomena are often described through random processes, $X\left(t\right)$, where t may be either continuous or discrete. Let the probability distribution $f_{X(t_1)}\left(x\left(t_1\right)\right)$ represent the probability distribution of the random variables X at point in time/space $t = t_1$. A discrete random process respectively a continuous random process, which was discretized in the time/space domain t, can be represented in a BN. An appropriate framework for representing such random processes is the dynamic BN (Fig. 3.5).

A discrete/discretized random process is said to have the Markov property if $X\left(t_i\right)$ at time slice t_i does only depend on $X\left(t_{i-1}\right)$ at time slice t_{i-1}, i.e.

$f_{X(t_i)|X(t_{i-1}),\ldots,X(t_0)}\left(x\left(t_i\right)|x\left(t_{i-1}\right),\ldots,x\left(t_0\right)\right) = f_{X(t_i)|X(t_{i-1})}\left(x\left(t_i\right)|x\left(t_{i-1}\right)\right)$.



a) Event tree with three intermediate random variables $\mathbf{X}=\{X_1, X_2, X_3\}$ and a target random variable Y

b) BN representation of the event tree

Figure 3.4: Representation of an ET in a BN. It should be checked carefully that the sequence in the ET follows the causal direction. From the conditional probabilities in the ET it can be assessed, whether a link is required or not.

Such a process is shown in (Fig. 3.5) together with its BN representation. An example of a BN modeling a stochastic process is given in (Straub, 2009).

While stochastic processes vary in a one-dimensional time or space domain, a random field $X(\mathbf{z})$ varies in a multidimensional domain $\mathbf{z}$ (Vanmarcke, 2010). Analogues to a stochastic processes, a random field can be discretized in the domain $\mathbf{z}$. A random field can be represented in a BN through a number of selected points $\mathbf{z} = \mathbf{z}_i$. Since $X(\mathbf{z})$ is auto-correlated for all pairs of points $\mathbf{z}_i$ and $\mathbf{z}_j$ the naïve approach for modeling $X(\mathbf{z})$ in a BN would be to add links between all pairs of points. Even for a medium number of points $\mathbf{z}_i$, this becomes computationally unfeasible. To overcome this issue, (Bensi et al., 2011) propose a BN representation of the random field by exploiting decomposition methods such as the Karhunen-Loéve expansion (Spanos and Ghanem, 1989). Applying decomposition methods, a vector $\mathbf{X}$ corresponding to the random field $X(\mathbf{z})$ evaluated at n selected points $\mathbf{z}_i, \ldots, \mathbf{z}_n$ can be approximated as a function of m independent standard normal random variables $\mathbf{U} = [U_1, \ldots, U_m]^T$:

$$\mathbf{X} \approx \mathbf{T}\mathbf{U} \tag{3.1}$$

The transformation matrix $\mathbf{T}$ can be obtained by a decomposition method. The approximation of Eq. 3.1 is exact if $m = n$ independent standard normal random variables are used. Elicitation of BNs, when an analytical model, like the one in Eq. 3.1 is given, will also be

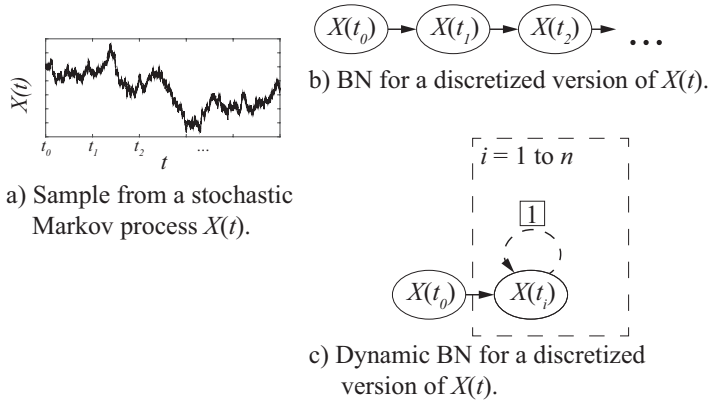


Figure 3.5: Dynamic BN model of a stochastic Markov process.

discussed in the following section. For the transformation to a BN the same approach as discussed there can be applied. The framework is applied in (Bensi et al., 2014) in the context of post-earthquake risk assessment.

Since a BN essentially represents the joint distribution between its random variables, it is possible, if a joint distribution is given (e.g. through a random process or a random field model), to derive all local, conditional probability distributions required for quantifying a BN representation of it. A structure can be derived from the joint probability distribution in a similar way to structure learning from data. The difference between this approach and data based structure learning (section 3.3) is that in this case one has perfect information (i.e. this would correspond to an infinitely large dataset). However just as it is the case for data based structure learning, there is no guarantee for the derived structure to represent causality.

3.1.4 Other models

Further probabilistic models that are well suited for transformation to BNs include safety barrier models, which are often used instead of FTs or ETs. Transformation of such a safety barrier model to a BN is described in (Léger et al., 2008). Probabilistic HRA models, which are used for the prediction of human errors in varying operational contexts, have been transformed into BNs (e.g. Kim et al., 2006; Groth and Swiler, 2013). (Hanea et al., 2012) investigate the dynamic BN as an alternative to the ensemble Kalman filter for reservoir engineering.

In principle, any probabilistic model can be transformed into a BN, because they all represent a joint probability distribution over a set or random variables (subsection 3.1.3). However, the BN is not necessarily an optimal representation in all cases. To understand, whether or not a BN is an effective representation of an existing probabilistic model, one should look into whether or not the probabilistic model is faithful to a DAG (Spirtes et al., 2001). A joint probability distribution is faithful to a DAG, if a DAG exists, which exactly represents the conditional independences and dependences found in the joint probability distribution

through the d-separation properties encoded in the graph structure. The Markov network, shown in Fig. 3.6a is an example of a probabilistic model that is not faithful to a DAG. In a Markov network, two nodes are d-separated if all undirected paths between them are blocked (Koller et al., 2007). Example conditional independence properties described by Fig. 3.6a are $X_1 \perp X_4|X_2, X_3$ (i.e. X_1 is independent of X_4 given X_2 and X_3) and $X_2 \perp X_3|X_1, X_4$. Example conditional dependences are $X_2 \not\perp X_4|X_1$ (i.e. X_2 is dependent on X_4 given X_1) and $X_3 \not\perp X_4|X_1$. It is impossible to represent all these four conditions simultaneously in a DAG. For example, in the BN from Fig. 3.6b the first independence condition is fulfilled but not the second while in the BN from Fig. 3.6c the second is fulfilled but not the first.

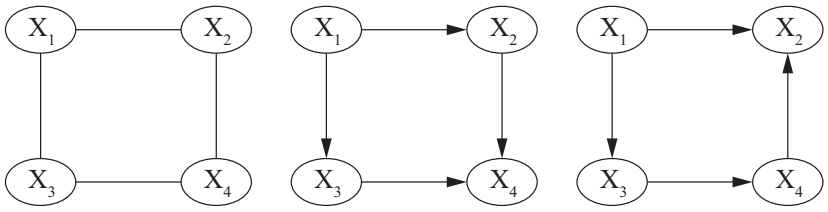


Figure 3.6: In a) a Markov network is shown, for which it is impossible to represent all its (conditional) (in-)dependence properties simultaneously in a BN. Both BN I and BN II fail at doing so.

3.1.5 Application example

An ET modeling the consequences of a ship collision accident, conditional on the event ship being on a collision course, is shown in Fig. 3.7a (Nývlt and Ferkl, 2013). The end state is dependent on whether the ship fails to detect the collision course (event $A = 0$), whether the platform fails to detect the collision course (event $B = 0$) and whether the ship fails to change the course in time (event $C = 0$). The occurrences of the events $A = 0$ and $B = 0$ due to lower-level events is modeled using FTs (Fig. 3.7b). Fig. 3.7c represents a BN model of both the ET and the FTs. The ET in Fig. 3.7a does not contain the conditional probabilities needed to quantify the corresponding, complete joint distribution e.g. it does not contain $p\,(B = 0|A = 1, \text{coll. course})$ and $p\,(C = 0|A = 0, B = 0, \text{coll. course})$. (Bearfield and Marsh, 2005) refer to these as "don't care" conditions, since these probabilities are irrelevant for the prediction of the end state. This may be either because one might know with certainty that e.g. $B = 0$ does (not) occur if $A = 1$ or because the state of B may be irrelevant for C or the End state if $A = 1$. Following (Bearfield and Marsh, 2005), if the objective of the BN is to only predict the probability of the End state, the parameters can be chosen freely, and optimally in a way that they represent causal independence. E.g. one can choose $p\,(B = 0|A = 1, \text{coll. course}) = 1 - p\,(B = 1|A = 1, \text{coll. course}) = p\,(B = 0|A = 0, \text{coll. course})$, thus there is no link between A and B. One can furthermore choose $p\,(C = 0|A = 0, B = 0, \text{coll. course}) = 1 - p\,(C = 1|A = 0, B = 0, \text{coll. course})$

$= p\,(C = 0|A = 0, B = 1, \text{coll. course})$ thus there is no link between B and C. Since $p\,(C = 0|A = 0, B = 0, \text{coll. course}) \neq p\,(C = 0|A = 1, B = 0, \text{coll. course})$ there is however a link from A to C. Freely choosing the parameters for "don't care" conditions is however only possible, if the BN is only used to determine the probability distribution of the End state (potentially given evidence for the other random variables).

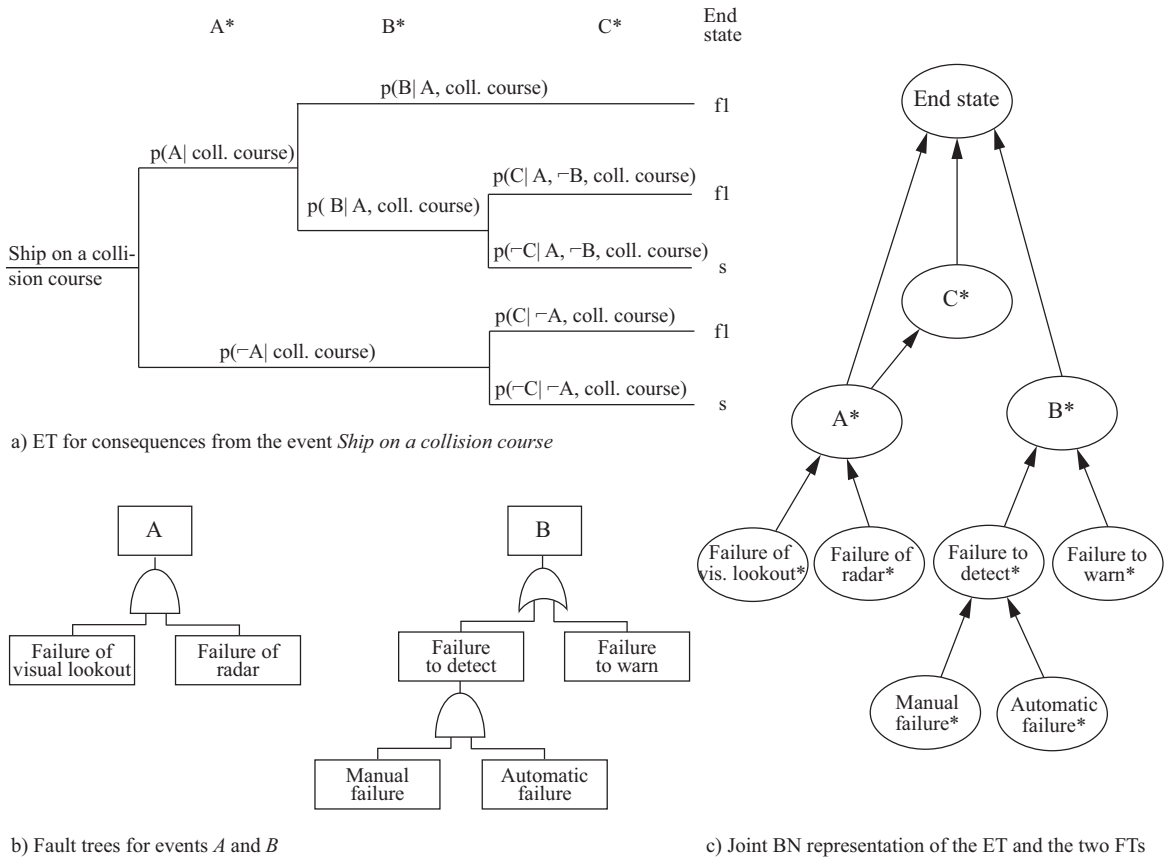


Figure 3.7: An ET modeling the consequences of a ship collision starting from the initiating event *ship being on a collision course* (a). The events A and B from the ET are modeled through FTs (b). The corresponding BN representing both the ET and the two FTs is shown in (c).

3.2 Structure elicitation based on general models

The behavior of anthropogenic or environmental systems can be described through physical or empirical models. In combination with a stochastic model of its input parameters, such models can be used in a probabilistic risk or reliability analysis. To extend such models or include them in a larger model, it can be beneficial to transform the model in a BN (Straub and Der Kiureghian, 2010a). The transformation has also been proposed for the purpose of Bayesian analysis in near-real-time decision support systems, because of the potentially fast inference algorithms available for BNs. Examples include the transformation of empirically-based mod-

els for earthquake risk assessment into BNs (Bayraktarli and Faber, 2011; Bensi et al., 2014; Franchin et al., 2016); physical models of upheaval buckling of pipelines (Friis-Hansen, 2000); physically and empirically based models for levee reliability (Roscoe and Hanea, 2015) and costal storm impact (Jäger et al., 2015); models of deterioration processes through dynamic BNs (Straub, 2009; Nielsen and Sørensen, 2010; Luque and Straub, 2015; Zhu and Collette, 2015).

3.2.1 Representation of a general model in a BN

Consider a general model g with scalar output Y and input $\mathbf{X}$; it is $Y = g(\mathbf{X})$. When the input $X = [X_1, \ldots, X_n]$ consists of independent random variables, the model can be generically represented by the BN shown in Fig. 3.8. Dependence between the random variables in $\mathbf{X}$ can be represented by including directed links among them or through common parents, following section 3.1, but this does not affect the representation of g. The model in Fig. 3.8 is the naïve BN representation of a general model (cf. section 3.1). The issue with this approach is the computational limitations associated with the converging connections to Y. For discrete BNs, the size of Y's CPT increases exponentially with the number of inputs, n. Computational efficiency is often a bottleneck in BN modeling. In particular this may be the case in system reliability problems with many components; physical models with many basic random variables or also expert elicited structures, where the experts assume a node to have many parents. Various approaches for making BN structures more efficient have been proposed. Some of the most important ones are presented in the following.

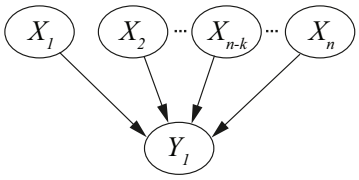


Figure 3.8: Naïve BN representation of a general model $Y = g(\mathbf{X})$. To model additional dependencies among the variables $\mathbf{X}$, links between the variables X_i or common parents can be added.

3.2.2 Simplification of a BN through node removal

[7]Removing random variables from a BN is one possibility to reduce the computational effort associated with a model. A formal approach for removing nodes from a BN is described in (Straub and Der Kiureghian, 2010b). In order to decide, which nodes to remove from the BN, the following questions should be considered (cf. section 3.4):

- Which random variables are relevant for prediction?
- Which random variables can potentially be observed?

[7]This subsection is adapted from (Zwirglmaier and Straub, 2016a).

- Which random variables simplify the modeling of dependencies?
- For which random variables is it desirable to explicitly show their influence on Y?

If a random variable does not belong to any of these categories, the corresponding node in the BN can be removed. Since the computational efficiency of the model is governed by the size of the CPT of the node representing Y, the primary interest is in removing parents of Y (basic random variables) from the network. As a measure for the relevance of a basic random variable X_i, importance measures α_i from a FORM analysis may be used. To better understand the relation between α_i and X_i's relevance for prediction consider a linearized LSF $G_L(\mathbf{U})$. Following (Der Kiureghian, 2005) the variance of $G_L(\mathbf{U})$ can be decomposed as:

$$\sigma^2_{G_L} = \|\nabla G\|^2(\alpha_1^2 + \alpha_2^2 + \ldots + \alpha_n^2) \tag{3.2}$$

where ∇G denotes the gradient vector of the non-linearized LSF $G(\mathbf{U})$. From Eq. 3.2 it is seen that a random variable X_i with corresponding α_i accounts for $\alpha_i^2 \cdot 100\%$ of the variance $\sigma^2_{G_L}$. Therefore, observing a random variable X_i with $\alpha_i = 0.1$ will reduce the variance $\sigma^2_{G_L}$ by 1%, whereas the fixing of X_j with $\alpha_j = 0.5$ will reduce $\sigma^2_{G_L}$ by 25%.

3.2.3 Parent divorcing

The number of parameters in the CPT of a variable Y is a function of the number of states of Y, $|Y|$ and the number of sates of the parents, $pa(Y)$. If it is possible to combine the effect of a number of parent nodes $\{X_1, \ldots, X_k\} \in pa(Y)$ on the node Y, in one mediating node I, where $|I| < \prod_{i=1}^{k} |X_i|$, the size of the CPT of Y can be reduced. This is shown in Fig. 3.9 where the effect of X_1 and X_2 is combined in the node I.

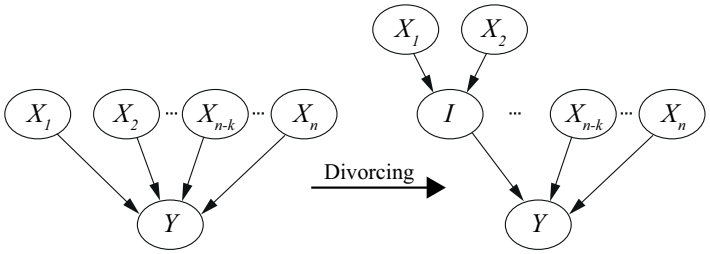


Figure 3.9: Parent divorcing.

3.2.4 Causal independence representation

The causal independence representation (Fig. 3.10) can make elicitation of Y's CPD easier and inference more efficient. This representation was introduced for temporal conditional independence statements in (Heckerman, 1993). In (Heckerman and Breese, 1994) an a-temporal definition of causal independence is introduced and empirical results showing its

superiority with respect to inference are presented. For the causal independence representation to be applicable, it is required that the order in which the causes (the X_i's) are introduced is irrelevant and that the effect of a variable X_i on the variable Y is independent of the effect of all other variables $\{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$ on Y.

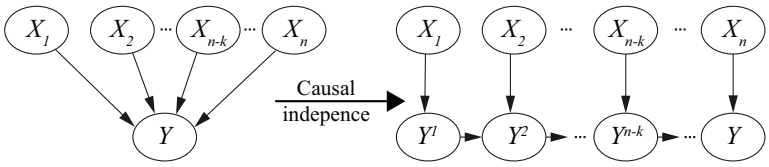


Figure 3.10: Causal independence representation.

(Spačková and Straub, 2013) apply the causal independence representation for modeling tunnel excavation processes. In (Bensi et al., 2013) the modeling technique is used for system reliability. (Gehl and D'Ayala, 2016) apply the causal independence representation to model bridge systems.

3.2.5 Application example

In this subsection a short example for the derivation of a BN structure from a generic problem is presented. The considered problem is a part of the problem discussed in chapter 8 . A numerical model g for the landing distance required by an aircraft as a function of parameters such as approach speed deviation, head wind, landing weight and multiple others has been proposed by (Drees and Holzapfel, 2012). In flight, these parameters are not known with certainty and are thus modeled by random variables $\mathbf{X}$. The model is transformed to a BN following (Zwirglmaier and Straub, 2015). The resulting BN is shown in Fig. 3.11. Only approach speed deviation, head wind, landing weight are modeled explicitly and all other random variables are treated as implicit uncertainties, since they are either not important or not observable at the required point of time. The model structure can be extended e.g. based on domain expert knowledge. This is indicated through the grey nodes airport and aircraft-type in Fig. 3.11. These nodes causally influence the basic random variables of the physical model.

3.3 Data based structure learning

The BN structure can be learned from data, which is a common approach in the field of machine learning and artificial intelligence (Friedman et al., 1997; Neapolitan, 2004; Koller et al., 2007). One can distinguish supervised and unsupervised learning. Unsupervised methods aim at finding a DAG that best represents the joint distribution of the data, either by ensuring that the d-separation properties of the DAG and the data match as closely as possible, or by maximizing the ability of the BN to predict the data with as few links as possible. Supervised methods on the other hand aim at finding a model that is best suited for predicting one

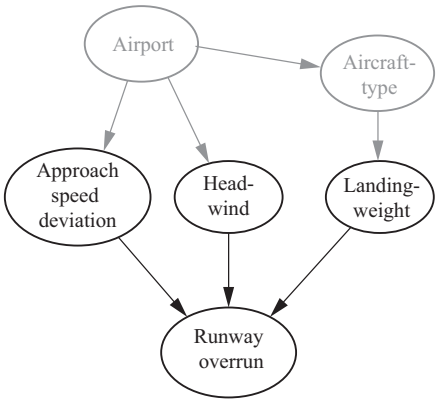


Figure 3.11: BN derived from a physical model (Zwirglmaier and Straub, 2015).

or more pre-selected target variables (problem variables). If additionally only the posterior mode of the target variable is of interest, the models are referred to as classifiers (Friedman et al., 1997). For classification tasks, often structures are not learned (or only partly learned), since predefined structures such as naïve Bayesian classifiers (NBCs) or tree augmented naïve Bayes (TAN) models have shown to be efficient (Zhang, 2004). A NBC for runway overrun, taking into account approach speed deviation, headwind and landing-weight as feature variables is shown in Fig. 3.12. However, classification is of limited interest in reliability engineering and risk analysis, where typically the posterior distribution itself is of relevance. Supervised learning can be considered an alternative to a regression analysis. A comparison between the two approaches is provided in (Kuehn, 2010).

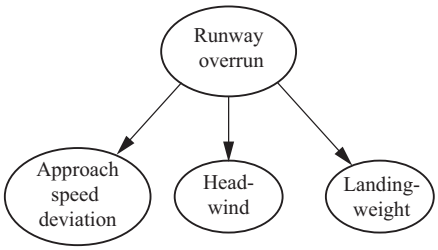


Figure 3.12: NBC for runway overrun taking into account the three parents of runway overrun from Fig. 3.11 as indicators.

In reliability engineering and engineering risk analysis, structure learning approaches are seldom used, since learning BN structures from data requires large datasets. Datasets should be especially large if rare events are of interest, if weak dependencies are to be learned or if the discrete random variables in the model have a large number of states (Kuehn et al., 2011). Reported applications of structure learning in a risk analysis context often focus on

the behavior of a system conditional on the hazard event occurring. Examples include the prediction of ground motions conditional on an earthquake event (Kuehn et al., 2011) or damage conditional on a flood event (Vogel, 2014). Whereas (Blaser et al., 2009) directly consider the hazard event tsunami. They overcome the issue of data scarcity by combining structure learning from data with expert knowledge and engineering models. (Doguc and Ramirez-Marquez, 2009) propose learning BNs for system reliability problems. However the examples investigated are limited to relatively simple systems with high probabilities of failure (i.e. > 0.1).

In general it cannot be guaranteed that DAGs that are learnt from data represent causality. A model therefore needs to be validated by checking d-separation properties rather than checking causal dependencies. Following the discussion at the beginning of this chapter this is typically more challenging. Non-causal models are furthermore difficult to handle, if they are to be combined with other models (cf. section 3.5).

3.3.1 Implementation

One distinguishes two fundamentally different approaches to structure learning based on data. The constraint-based approach attempts to find a DAG that best describes the independence properties of the joint distribution that is represented through the data. To derive the independence properties from the data, statistical independence tests are applied. Based on that a DAG is generated that represents the identified independence properties as closely as possible through d-separation properties. If the distribution, from which the data is generated, is faithful to a DAG (cf. subsection 3.1.4) , all its (in-)dependence properties can be represented. Whether all (in-)dependence properties of the underlying distribution are identified correctly depends on the size of the dataset. Examples of constrained-based algorithms are the SGS (Spirtes et al., 1989), the PC (Spirtes et al., 2001) and the NPC algorithm (Steck, 2001).

Score-based approaches attempt to find, among a set of candidate BN structures, the candidate that describes the data best (or that predicts the state of the target variable best). To decide which candidate BN describes the data best, scoring functions are used. Maximum-likelihood-based scoring functions combine the likelihood of the model given the data with a penalty term for complex BNs. Besides that, Bayesian scoring functions are possible. Optimally, one would evaluate the scoring function for each possible DAG. However, since the number of possible BN structures grows faster than exponentially with the number of nodes (Robinson, 1977), such an exhaustive search is typically unfeasible. Thus efficient search heuristics are applied to generate the candidates. Examples of scoring functions are the Akaike information criterion (AIC) by (Akaike, 1974), the closely related Bayesian information criterion (BIC) by (Schwarz, 1978), the Bayesian Dirichlet (BD) scores (Heckerman et al., 1995) or the maximum a posteriori (MAP BN) criterion (Riggelsen, 2008). Examples of search approaches are greedy search algorithms such as the repeated covered arc reversal algorithm (Castelo and Kocka, 2003), the search over equivalent BN classes (Chickering, 2002)

Table 3.2: Summary of the local (conditional) probabilities for the BN in Fig. 3.13a.

Random variable	states	probabilities
season	[spring, summer, fall, winter]	$[0.25, 0.25, 0.25, 0.25]$
vegetation	[dense trees, no dense trees]	see Annex A
slope	$[< 30°, 30° - 45°, > 45°]$	$[1/3, 1/3, 1/3]$
avalanche	[yes, no]	see Annex A

or a search over node orderings (Teyssier and Koller, 2012).

3.3.2 Application example

To illustrate the difficulty in learning even a simple BN, a small hypothetical example is presented, in which data is first generated from a reference BN, and a new BN is then learned from this data. The reference BN is presented in Fig. 3.13a. It consists of four discrete random variables with no more than four states each. The variables, their states and their CPTs are summarized in Tab. 3.2.

From the reference BN of Fig. 3.13a, 500 samples are generated. These are used in a score based structure learning approach, with a greedy search strategy and the BIC scoring function. The resulting learnt BN structure is shown in Fig. 3.13b. From these samples, the influence of season and slope on avalanche probability is identified correctly, but not the influence of vegetation. The dependence between slope and vegetation is also identified from the data, but in the non-causal direction. However the (in-)dependence criteria implied by this link direction are correct. It is not possible to identify the direction of this link based on data alone.

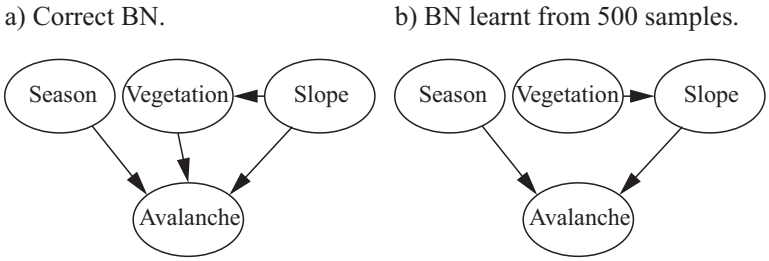


Figure 3.13: a) Original BN for predicting avalanche probability. b) BN learnt with 500 samples using a score based structure learning approach.

It is evident that the automated structure learning in this case failed to find the correct BN structure from the 500 samples. Adding more samples or combining the automated structure learning with expert knowledge can help to do so.

3.4 BN structures based on domain expert knowledge

A BN structure can be elicited directly from domain expert knowledge. This route is often followed when neither data, probabilistic nor general models exist, or existing models do not meet the modeler's requirements. Causality as a guiding principle for the construction of a DAG is not least supporting the process of eliciting the model structure from experts. In many instances, the development of the BN DAG is a rather intuitive process. However, formal approaches to the model building process are nevertheless advantageous, to ensure efficient models that can also be quantified and to avoid inconsistencies that may arise from experts thinking in a diagnostic instead of a causal mode.

3.4.1 Modeling approach

The procedure for eliciting BNs from domain expert knowledge can be structured into the following three steps:

1. Problem clarification and identification of relevant parameters
2. Modeling of relevant dependencies
3. Handling unquantifiable variables

Problem clarification and identification of relevant variables

Before building the model, the relevant variables have to be determined. Following (Kjærulff and Madsen, 2013), relevant variables can be classified into problem variables (the variables of interest), information variables (the observable variables, either background or symptom variables) and mediating variables, which are not observable but help in representing the dependencies between information variables and the problem variables. This classification gives a general idea on what type of variables are relevant for a model. To identify the variables that are relevant for a specific problem, it is necessary to get an understanding of the system. To this end it is possible to consult domain experts either directly, e.g. in the scope of organized interviews (Hanea and Ale, 2009; Cárdenas et al., 2013), or indirectly, e.g. through literature reviews (e.g. Zwirglmaier et al., 2017). In many cases it is possible to directly represent the knowledge provided by the domain experts in a BN structure. In other cases direct representation of the problem in a BN structure may not be readily possible for the modeler. In such cases, structured techniques for analyzing systems can be used as an intermediate step. A large number of structured techniques to analyze systems exist. Examples of techniques that have been used to help in BN structure elicitation include failure mode and effects analysis (FMEA) (Weber and Jouffe, 2006; Suddle, 2009; De Carlo et al., 2013) and structured analysis and design technique (SADT) (Weber and Jouffe, 2006; Trucco et al., 2008). An approach for deriving BN structures from causal maps is furthermore proposed in (Nadkarni and Shenoy, 2004).

Since nodes in a BN have to represent random variables, it is pivotal to capture the concepts used by domain experts or in other techniques through random variables that are unambiguous and clearly defined (incl. support, temporal/spatial reference).

Modeling of relevant dependencies

Directed links between the identified relevant variables define their dependencies. In general the elicitation of the dependence structure cannot be treated independently of the identification of relevant random variables. I.e. since on the one hand insight into the dependence structure may be necessary to decide, which variables are relevant, and on the other hand the methods used for identifying the relevant variables (such as interviews, literature reviews, or structured system analysis methods) will always give also some insight into the dependencies between the variables. For example in a FMEA causes and effects of a failure mode are identified. A causal link between the two would thus be directed from the former to the latter. As discussed earlier in this thesis, causality of the links is not a requirement of the BN methodology but causal links typically lead to more efficient models that are also more traceable. In general causality is a good guiding principle for BN structure elicitation, as it is in line with the human reasoning process (Pearl, 1988). However, human reasoning can also be misleading. To illustrate this, consider a result from a structural monitoring that indicates a poor condition of the structure. There are two possible ways for orienting the link between the two variables.

(I) From monitoring result to condition of the structure: This is referred to as the diagnostic direction i.e. by looking at the monitoring result one concludes that the condition of the structure is poor.

(II) From condition of the structure to monitoring result: This is referred to as the causal direction i.e. the condition of the structure determines the monitoring result.

Option I is often chosen intuitively, but option II is preferable for modeling and computation purposes. To avoid such fallacies, a number of principles that support the elicitation of the causal dependence structures exist. To get an idea about the overall dependence structure, (Kjærulff and Madsen, 2013) propose overall causal dependence structures for their categories of variables. According to (Kjærulff and Madsen, 2013), typically background variables potentially influence problem variables and/or mediating variables and/or symptom variables; problem variables potentially influence mediating and/or symptom variables and mediating variables potentially influence symptom variables. Furthermore the causal direction of a link can often be assessed from the temporal order of events, i.e. following (Pearl, 2009), “a later event can never be the cause of an earlier event”. In addition, following (Neil et al., 2000) the causal dependence structure can be inferred by projecting specific situations onto generic BN structures (idioms). (Neil et al., 2000) propose 5 idioms: the definitional/synthesis, the cause-consequence, the measurement, the induction and the reconciliation idiom. For each

of these idioms, a BN structure is provided. A modeler can make use of these idioms by identifying these situations in her problem and adapting the given generic BN substructure for that situation.

Unquantifiable variables

The resulting qualitative BN model can contain unquantifiable variables. These should be dealt with separately. Unquantifiable variables are those, for which no reasonable probability estimates can be obtained from experts or data. Note that having no data on a particular node does not necessarily imply that it cannot be quantified. The expectation maximization (EM) algorithm may under certain circumstances allow estimating the CPDs of nodes with no data (Lauritzen, 1995; Dempster et al., 1977). (Hanea et al., 2006; Hanea and Ale, 2009) propose to replace unquantifiable variables by proxy variables that are quantifiable. In cases where it is expected that data or expert knowledge for quantifying the unquantifiable variable will become available in the near future, it may be sufficient to use dummy parameters for the preliminary quantification. However, if this information is not expected to be available, these variables should be eliminated from the network to make the BN applicable in practice. In (Zwirglmaier et al., 2017) node removal algorithms, as proposed by (Shachter, 1988), are used for this purpose. (Straub and Der Kiureghian, 2010a) applied these algorithms to remove continuous variables in their enhanced BN framework.

Node reduction algorithm (Shachter, 1988):

- Barren nodes, i.e. the ones, which have no children and which have not received evidence can be removed from the network.
- To make nodes barren, arcs can be reversed. In order to revers an arc between two nodes X_i and X_j, both nodes must inherit each others parents, without making the directed graph cyclic.

The presented node reduction algorithm ensures that the (in-)dependence properties between the remaining random variables are not altered. It should be noted that the reduced structure depends on the order of node removals. Furthermore, node reduction algorithms may lead to nodes having more parents than in the original (qualitative) network.

3.4.2 Relation to other structure elicitation approaches

In the literature typically, only BN structure elicitation based on domain expert knowledge and data based structure learning are distinguished (Jensen and Nielsen, 2007; Kjærulff and Madsen, 2013). In this chapter additionally structure elicitation based on probabilistic models and structure elicitation based on generic models, were distinguished. In fact these two approaches are closely related to structure elicitation based on domain expert knowledge. The reason being that other probabilistic models or generic models can be used as an intermediate step in the latter approach to analyze the domain expert knowledge in a structured way. For

example, probabilistic modeling tools like FTs or ET can be used as a structured system analysis method to get an understanding of the system. Besides that also generic models can represent domain expert knowledge. These generic models could also be represented e.g. in technical drawings or physical flow graphs (Perelman and Ostfeld, 2012). Thus both other probabilistic models as well as generic models can be considered domain expert knowledge, but this domain expert knowledge is presented in a way, which greatly simplifies the BN elicitation.

3.4.3 Application example

A short example for the derivation of a BN structure from domain expert knowledge is presented. The considered problem is a part of the problem discussed in chapter 9: The misperception of critical data in the control room of a nuclear power plant, which can have severe consequences. Attention degradation is a concept that has been studied extensively in the field of cognitive psychology (Whaley et al., 2012). The literature from this field can be used by the BN model builder to gain an understanding of the cognitive processes. I.e. it can be deduced that attention degradation is a possible cause of misperception of critical data (Whaley et al., 2012; Xing et al., 2013). Besides that, attention degradation may come from crew-members having biased expectations or prioritizing falsely. Both issues can be influenced through missing or faulty training. Prioritization is furthermore influenced by the perception of urgency of the crew. Finally, attention degradation can be fostered by high workloads.
From these statements the relevant concepts and their interdependencies can be deduced. In a BN these concept need to be replaced by nodes that represent random variables. A representation of these random variables and their dependencies is shown in the BN of Fig. 3.14 (adapted from (Zwirglmaier et al., 2017)). Some of the parameters are not quantifiable with feasible effort. These nodes are marked in white in Fig. 3.14. Shachter's node reduction algorithm is applied to remove them from the BN. The resulting BN structure is shown in Fig. 3.14.

3.5 Combination of the approaches

The structure of a medium or large BN is typically not developed solely through one structure elicitation approach. Instead, the core of a BN structure may be elicited based on a probabilistic or a general model and the model is extended based on domain expert knowledge. Or different substructures derived using different structure elicitation approaches are combined in one larger network. As an example, (Grêt-Regamey and Straub, 2006) develop an avalanche risk model based on a (numerical) physical avalanche model, which is combined with vulnerability and consequence BN models based on domain expert knowledge. The causal model for air transport safety (CATS) (Ale et al., 2008) combines BN structures that are developed based on a transformation of FTs and event sequence diagrams with BN structure of human

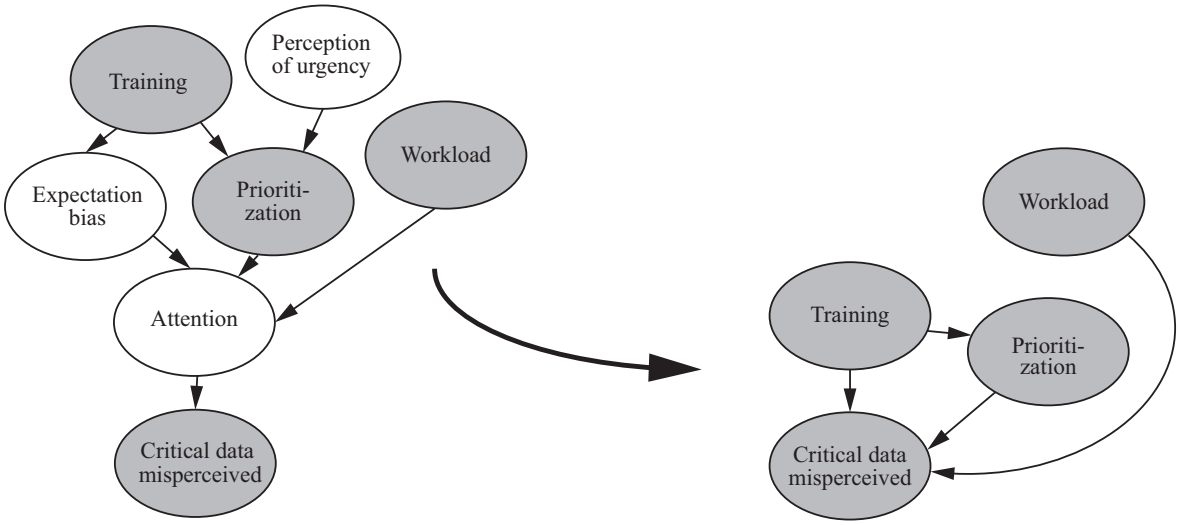


Figure 3.14: BN structure for critical data misperceived (Zwirglmaier et al., 2017). Following the node removal algorithm from (Shachter, 1988), the unquantifiable (white) nodes are removed from the network.

factors that were elicited using the domain expert knowledge. (Papakosta and Straub, 2013) include a BN that is derived from an empirical fuel moisture model in a domain expert elicited model for fire occurrence prediction.

In principle, such an extension of BN structures or combining different substructures is readily achievable using the discussed approaches, if all the links in the combined network follow the causal directions. If, however some, of the sub-networks do not follow the causal direction combination of sub-networks becomes critical. This is illustrated in Fig. 3.15. There a simple BN with the two nodes cause and consequence (black) is extended by an additional cause (grey). In Fig. 3.15a the original model is causal (the cause points to the consequence) while it is diagnostic in Fig. 3.15b and c (from the consequence one concludes about the cause). If a new possible cause is added to the BN both the new and the old cause should be dependent, if the state of consequence is known. This is neither represented by Fig. 3.15b, where the additional link is added according to its causal direction nor by Fig. 3.15c, where both links are modeled according to their diagnostic direction. Only in the model in Fig. 3.15a the (in-)dependence assumptions are modeled correctly. To model this conditional dependence an additional arc between the two causes could be added, which results in a complete network that is requires more parameters than the correct BN. It is evident that causal models are typically both simpler and easier to handle, when extending the model with additional links or combination with other BN models, since (causal) extensions of a causal model have only local effects on the BN. While typically one tries to elicit causal structures, BN structures learnt from data are in general not causal. Therefore extensions, combinations of such BN structures are critical.

Capturing the causal direction of links is often straightforward, sometimes it is challenging and sometimes it is impossible. An example of the latter situation is the simple representation

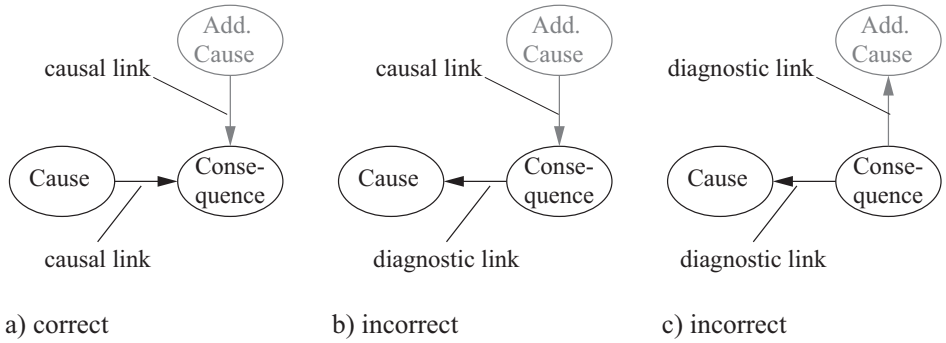


Figure 3.15: Extension of a simple causal respectively a diagnostic BN model by an additional cause.

of a random field in a complete DAG. In a random field, any two points (up to a certain distance) are correlated. Since the dependence between two points in a random field is not directed, the directions of links in the complete DAG can be chosen arbitrarily. In that case this is not critical with respect to the implied d-separations, since in a complete graph any two nodes are dependent unconditionally and conditional on any subset of other nodes in the BN. However, this can become problematic in similar situations e.g. modeling bidirectional physical flow graphs in a BN.

3.6 Discussion

Four structure elicitation approaches for BNs were distinguished. Firstly, structure elicitation based on transformation of other probabilistic models. Secondly, structure elicitation based on general models. Thirdly, structure learning from data, and finally structure elicitation based on domain expert knowledge.

In reality, the boundaries between different structure elicitation approaches are often blurred. For example a system reliability problem can be described through its system function. In this chapter system functions were discussed in the section on other probabilistic models, structure functions could however also be described in the context of general models.

Likewise, the representation of a random field in a BN can be approached directly from the probabilistic model i.e. the random field, or based on its decomposition (e.g. Karhunen-Loéve expansion). The latter one would then be approached like elicitation based on a general model. Moreover, other probabilistic models or general models may exist that are not directly suitable for transformation. In spite of this, such models can still serve as a source of domain expert knowledge for model elicitation.

While BNs represent a modeling framework that is suitable to a large variety of problems, there are a number of challenges associated to modeling with BNs. Computational issues are common to large BNs, where nodes have many parents or parents with many states. Such BN structures are often encountered in system reliability problems with many components, in reliability problems with many basic random variables or in problems with many dependencies,

e.g. random fields that are modeled through complete BNs. However such problems are also common to BN structures that are elicited by domain experts, since often domain experts tend to include too many dependencies in the BN (Langseth and Portinale, 2007). In discrete BNs, the CPT of a node increases exponentially with the number of parents. For large CPTs the computers memory space can thus easily become a bottleneck. Besides that, also eliciting the parameters of such a large CPT through domain experts or a limited datasets can easily become unfeasible.

As discussed in section 3.2, in some cases more efficient BN structures can mitigate this problem. Also compressing the CPTs as proposed in (Tien and Der Kiureghian, 2016) can help to reduce the required memory space to some extent. Besides that, inference-time can become critical for BNs with large cliques. For large system reliability problems with many components it can be the case that inference is not possible with standard BN inference algorithms, although the probability of system failure can be calculated with FT analysis methods. Again the efficient structures discussed in section 3.2 can mitigate the inference time bottleneck to some extent. Furthermore using sampling based approximate inference algorithm can help to mitigate both the memory space issues as well as the inference time issues.

Although in theory not required for BNs, modeling the links according to their causal direction is usually advisable. Especially for BN structures that are derived from domain expert knowledge, finding the causal direction of links can be challenging. But also if the BN is derived from other probabilistic models, the BN model builder should not rely on the assumption that typically these ETs etc. follow the causal direction. As discussed, approaches like modeling idioms (Neil et al., 2000) and overall dependencies between groups of random variables (Kjærulff and Madsen, 2013) can support the modeler in this.

Chapter 4

Quantification

To each node X_i in a BN, a local probability distribution, which is conditional on the (possibly empty) set of parents $pa(X_i)$, is assigned. For example, to quantify the BN structure in Fig. 2.1, the marginal/conditional distributions $p(X_1)$, $p(X_2)$, $p(X_3|X_1)$, $p(X_4|X_1)$ and $p(X_5|X_2, X_3.X_4)$ are required. In discrete BNs, these probability distributions are represented through CPTs (e.g. Tab. 4.1).

In general, efficient BN structures are key in making quantification feasible. However, even for efficient BN structures quantification is often difficult, due to scarcity of quantitative information. Because of that, all available sources of information should be used in order to come up with good quantitative models.

Typical sources of information are observed data, expert estimates or physical/empirical models. As a CPD is assigned separately to each node in the BN, it is readily possible to use a different source of information for each node.

Quantification through data is discussed in section 4.1. Quantification solely based on expert estimates is briefly touched upon in section 4.2. This presentation is followed by a discussion on how expert estimates and data can be combined (section 4.3). Finally it is discussed how sampling in combination with physical/empirical models can be used to quantify a node (section 4.4).

4.1 Quantification based on data

From a dataset containing observed data or data collected in the scope of simulator experiments, estimates for the required (conditional) probability distributions can be obtained. For continuous nodes, distribution models can be fitted to the data using the well-known approaches (Kottegoda and Rosso, 2008). These include the maximum likelihood estimator, the method of moments as well as the Bayesian estimator. For a discrete BN, the entries of the CPTs can be estimated through the frequencies in the dataset. E.g. $p(X_4 = x_4|X_1 = x_1)$

for the BN in Fig. 2.1 can be approximated as:

$$p\left(X_4 = x_4 | X_1 = x_1\right) \approx \frac{n\left(X_4 = x_4, X_1 = x_1\right)}{n\left(X_1 = x_1\right)} \tag{4.1}$$

where $n\left(X_4 = x_4, X_1 = x_1\right)$ is the number of cases in the dataset for which random variable X_4 is in state x_4 and X_1 in state x_1, and $n\left(X_1 = x_1\right)$ is the number of cases for which X_1 is in state x_1. In principle, this approach is readily applicable to nodes with multiple parents. However, the size of the dataset required to get good estimates increases with the number of parents.

Data is often missing for some of the nodes in the BN. In such cases, it is possible to estimate the most likely parameters of these nodes through the EM algorithm (Dempster et al., 1977; Lauritzen and Spiegelhalter, 1988). The EM algorithm iteratively runs through two steps, these steps are the expectation (E) step and the maximization (M) step. Starting from an initial set of parameters $\boldsymbol{\theta}$ in the E step, the algorithm computes the expected observed frequencies for the missing variables, conditional on the current parameters $\boldsymbol{\theta}$. Based on these expected observed frequencies, the parameters are recomputed. The algorithm iterates through these two steps until a convergence criterion is fulfilled. The EM algorithm can be used in situations, where a number of variables in the BN are not observed at all but also in situations, where some variables are not observed in some instances of the dataset. In the latter case it is required that the data is missing at random.

4.2 Quantification based on experts

If data is not available the local probability distributions can be elicited through experts (Bedford and Cooke, 2001). An in depth treatment of this topic is outside the scope of this thesis. The interested reader is referred to the respective literature (e.g. Cooke, 1991; O'Hagan et al., 2006). In general, when eliciting probabilities from experts, the choice of the right experts is a critical point. Also, if estimates by multiple experts are elicited, the question on how to weight differing estimates needs to be addressed. Experts are often biased. Therefore, it is important to avoid such biases in the process of expert elicitation.

4.3 Combination of expert estimates and data

If neither a sufficiently good dataset is available nor unambiguous expert estimates are readily assessable, the best available estimate can be obtained by combining experts and data. Bayesian updating is applied in such cases to enhance the experts' estimates of a parameter θ with new data. The most important aspects of Bayesian updating are revisited in this chapter; for a more in depth treatment the interested reader is referred to the relevant literature (e.g. Benjamin and Cornell, 1970; Straub and Papaioannou, 2014; Groth et al., 2014).

In the scope of this thesis Bayesian, updating of BN parameters is applied in the HRA example

from chapter 9, for the parameters of a node representing a crew failure event. In this example the node of interest is binary, thus a beta distribution is used as a prior. A simple approach for elicitation of beta priors is presented in this section, after a brief description of Bayesian updating of binary nodes. The description of Bayesian updating is extended to multistate nodes afterwards, continuous nodes are only briefly touched upon.[8]

4.3.1 Bayesian updating of parameters

Binary nodes

An example of a CPT of a binary node X_i with three binary parents $pa(X_i) = \{X_1, X_2, X_3\}$ is shown in Tab. 4.1.

Table 4.1: CPT of a binary node X_i with binary parents $\{X_1, X_2, X_3\}$

X_1	x_1^1				x_1^2			
X_2	x_2^1		x_2^2		x_2^1		x_2^2	
X_3	x_3^1	x_3^2	x_3^1	x_3^2	x_3^1	x_3^2	x_3^1	x_3^2
$X_i = 0$	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
$X_i = 1$	$1-\theta_1$	$1-\theta_2$	$1-\theta_3$	$1-\theta_4$	$1-\theta_5$	$1-\theta_6$	$1-\theta_7$	$1-\theta_8$

As evident from this CPT, for a binary node X_i with binary parents $\{X_1, X_2, X_3\}$, $n_\theta = 8$ parameters θ_i need to be elicited. In the following the parameter θ_i that is to be updated is simply denoted as θ.

The prior PDF $f_0(\theta)$ represents the belief in the state of θ before considering the new data, e.g., the probabilities based solely on expert elicitation. The data can be the result of one or more simulator experiments or operating events. Applying Bayes' rule (Eq. 4.2) allows one to combine the prior distribution with the data $\mathbf{x}$ to get the posterior distribution $f_1(\theta|\mathbf{x})$, representing the belief in the state of θ after having observed $\mathbf{x}$:

$$f_1(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\, f_0(\theta) \tag{4.2}$$

where $f(\mathbf{x}|\theta)$ is the likelihood of the parameters θ given the data $\mathbf{x}$. A typical database used in such circumstances contains the number of positive/negative outcomes in a number of trials. If the outcome of each event is binary (e.g. success or failure), and if it can be assumed that the trials are independent of each other, the data can be described by a Bernoulli process. The parameter to estimate is the probability of failure θ and the observation is n_e, the number of failures in a total of n observed/simulated scenarios. In this case, the likelihood function $f(n_e|\theta)$ is the binomial probability mass function (PMF) with parameter θ:

$$f(n_e|\theta) = \binom{n}{n_e} \cdot \theta^{n_e} \cdot (1-\theta)^{n-n_e} \tag{4.3}$$

[8]The description of Bayesian updating for binary nodes and elicitation of beta priors is partly adapted from (Zwirglmaier et al., 2017).

The beta distribution is a conjugate prior for the binomial likelihood function (Raiffa and Schlaifer, 1961). As it greatly simplifies the mathematics of Bayesian updating, the use of conjugate priors is in general recommendable. The beta prior PDF with parameters a_0 and b_0 is:

$$f_0(\theta) = \frac{1}{B(a_0, b_0)} \theta^{a_0 - 1} (1 - \theta)^{b_0 - 1} \tag{4.4}$$

If the beta distribution is used to model $f_0(\theta)$ and the likelihood function is the binomial PMF of Eq. 4.3, the posterior $f_1(\theta|n_e)$ is beta distributed as well. In this case, the parameters of the posterior beta distribution can be calculated analytically as:

$$a_1 = a_0 + n_e \tag{4.5}$$

$$b_1 = b_0 + (n - n_e) \tag{4.6}$$

An example of a beta prior PDF, f_0 that is updated with two different datasets is shown in Fig. 4.1. In both cases the uncertainty is reduced through the Bayesian updating. Furthermore in the first case the posterior distribution of θ is shifted to the right compared to the prior.

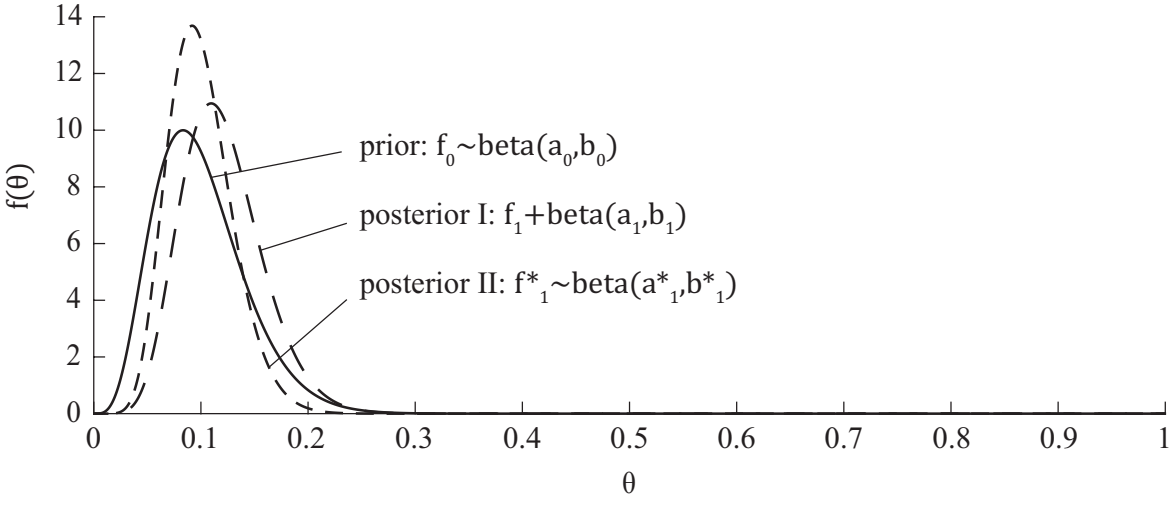


Figure 4.1: Updating the beta prior distribution using two different data sets. The parameters of the prior , $f_0(\theta)$, are $a_0 = 5$ and $b_0 = 45$. Posterior I is obtained by updating with a dataset of length $n = 25$ in which $n_e = 4$ failures are observed. Posterior II is obtained by updating with a dataset of length $n = 50$ with $n_e = 5$ failure occurrences.

Multistate nodes

The CPT of a multistate node X_i with m states requires elicitation of $m - 1$ parameters $\boldsymbol{\theta}_j = \left[\theta^1, \theta^2, \ldots, \theta^{m-1}\right]$ for each state of X_i's parents. An example of such a CPT of a multistate node X_i with binary parents $pa(X_i) = \{X_1, X_2, X_3\}$ is shown in Tab. 4.2.

Table 4.2: CPT of a multistate node X_i with discrete parents $[X_1, X_2, X_3, ..., X_n]$

X_1		x_1^1				x_1^2		
X_2	x_2^1		x_2^2		x_2^1		x_2^2	
X_3	x_3^1	x_3^2	x_3^1	x_3^2	x_3^1	x_3^2	x_3^1	x_3^2
$X_i = 1$	θ_1^1	θ_2^1	θ_3^1	θ_4^1	θ_5^1	θ_6^1	θ_7^1	θ_8^1
$X_i = 2$	θ_1^2	θ_2^2	θ_3^2	θ_4^2	θ_5^2	θ_6^2	θ_7^2	θ_8^2
$X_i = 3$	θ_1^3	θ_2^3	θ_3^3	θ_4^3	θ_5^3	θ_6^3	θ_7^3	θ_8^3
...								
$X_i = m$	$1 - \sum_{l=1}^{m-1} \theta_1^l$	$1 - \sum_{l=1}^{m-1} \theta_2^l$	$1 - \sum_{l=1}^{m-1} \theta_3^l$	...	...	...	...	...

The natural choice to express ones prior belief about $\boldsymbol{\theta}$ (the index j is suppressed for simplicity) is the Dirichlet distribution. The Dirichlet distribution can be seen as an extension of the beta distribution to multistate nodes. Furthermore, it is a conjugate prior to the multinomial distribution, a generalization of the binomial distribution. I.e. analogous to the binary case, if the prior follows a Dirichlet distribution and the likelihood is multinomial, the posterior will also follow a Dirichlet distribution. The PDF of the Dirichlet distribution reads:

$$f(\boldsymbol{\theta}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{m} \theta_i^{\alpha^i - 1} \tag{4.7}$$

where $\boldsymbol{\alpha} = \left[\alpha^1, \alpha^2, ..., \alpha^m\right]$ are its parameters and $B(\boldsymbol{\alpha})$ is a normalizing constant:

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{m} \Gamma\left(\alpha^i\right)}{\Gamma\left(\sum_{i_1}^{m} \alpha^i\right)} \tag{4.8}$$

Let $\boldsymbol{\alpha}_0$ denote the parameters of a Dirichlet prior. A dataset $\mathbf{x}$, containing observations of X_i conditional on the respective combined state of its parents can be used to update the prior beliefs about $\boldsymbol{\theta}$ (c.f. Eq. 4.2). In particular, if $n_{X_i=l}$ is the number of cases in $\mathbf{x}$ for which X_i is in state l and the parents are in their respective states, the posterior parameter α_1^l reads:

$$\alpha_1^l = \alpha_0^l + n_{X_i=l} \tag{4.9}$$

Continuous nodes

Other distribution types may be used for continuous nodes. In that case the principle from Eq. 4.2 is still valid. The computation of the posterior distribution will however differ. For further details the reader is referred to the standard textbooks (e.g. Raiffa and Schlaifer, 1961; Kottegoda and Rosso, 1997; Gelman et al., 2013).

4.3.2 Elicitation of prior beliefs for binary nodes

So far in this chapter the parameters of the beta prior a_0 and b_0 were assumed to be given. In practice these parameters typically need to be elicited from experts. Many techniques to elicit

beta parameters from experts have been proposed (e.g. Bedford and Cooke, 2001; O'Hagan et al., 2006). A review of these approaches is out of the scope of this thesis.
A straightforward approach to estimate the parameters of $f_0(\theta)$ is to first estimate the expected value $E[\theta]$. The first constraint on the distribution parameters is given through the definition of the expected value of a beta distribution:

$$E[\theta] = \frac{a_0}{a_0 + b_0} \tag{4.10}$$

A second constraint is given through the shape of the PDF. To this end the parameters and thus also the shape of the distribution are altered, in a way that the expected value $E[\theta]$ still corresponds to the elicited one, until the shape of the PDF represents the expert's uncertainty about θ. In Fig. 4.2 this is exemplarily shown for an expected value $E[\theta] = 0.1$ and varying uncertainty.

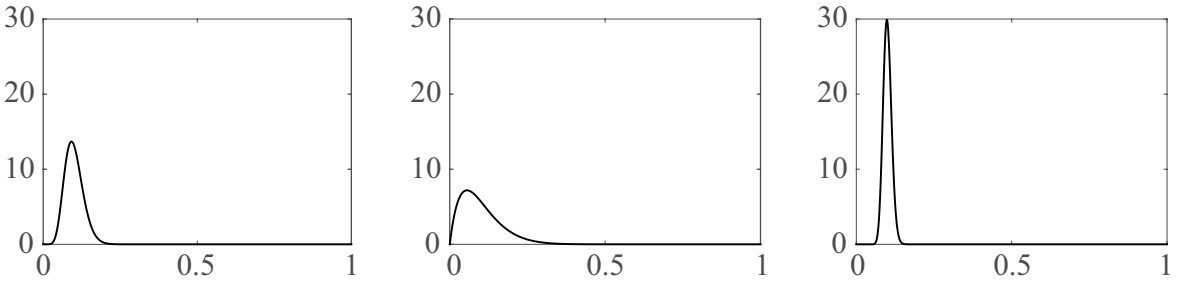


Figure 4.2: PDFs of beta distributions with mean 0.1 and varying standard deviations. The standard deviations of the distributions are a) 0.03, b) 0.07 and c) 0.01.

The elicitation of Dirichlet priors is due to its multivariate character slightly more challenging. For a detailed description of approches for the elicitation of Dirichlet priors the interested reader is referred to the relevant literature (e.g. van Dorp and Mazzuchi, 2004; Zapata-Vázquez et al., 2014).

4.4 Nodes representing engineering models

In reliability, failure events are often defined in terms of engineering models. These models can be physical or empirical. Also they may be defined in an analytical or numerical manner. If such models are to be included in discrete BNs, they need to be represented through a CPT. For example, in the application of chapter 8, the node representing runway overrun (RWO) is defined through a physically based numerical model.
Representation of engineering models in a CPT requires the estimation of probabilities of the engineering model being in a certain discrete state conditional on its basic random variables (parents in the BN) being in certain states. In the following, the engineering model is denoted as $g(\mathbf{x})$. This model is dependent on the basic random variables $\mathbf{X} = [X_1, \ldots, X_n]$ with joint

PDF $f_{\mathbf{X}}(\mathbf{x})$. Typically these basic random variables are continuous, and are discretized only for the purpose of representing them in discrete BNs. We thus distinguish between the continuous basic random variable X_i and its corresponding discretized version Y_i, whose j-th state is denoted as y_j^i. The discretized random variable is obtained by the introduction of interval boundaries $\left[d_0^i, d_1^i, \ldots, d_m^i\right]$, where d_0^i and d_m^i correspond to the domain boundaries of X_i. The j-th interval is bounded by d_{j-1}^i and d_j^i.
The probability of $g(\mathbf{x})$ being in state E conditional on its discretized parents $\{Y_1, \ldots Y_n\}$ being in states $\left[y_j^1, \ldots, y_k^n\right]$, can then be written as:

$$\Pr\left(g(\mathbf{x}) \in E | y_j^1, \ldots, y_k^n\right) = \int_{g(\mathbf{x})\in E} f\left(\mathbf{x} | d_{j-1}^1 < x_1 \leq d_j^1, \ldots, d_{k-1}^n < x_n \leq d_k^n\right) \mathrm{d}\mathbf{x} \quad (4.11)$$

where $f\left(\mathbf{x} | d_{j-1}^1 < x_1 \leq d_j^1, \ldots, d_{k-1}^n < x_n \leq d_k^n\right)$ represents the PDF $f_{\mathbf{X}}(\mathbf{x})$, truncated at the interval boundaries corresponding to the states $\left[y_j^1, \ldots, y_k^n\right]$. This problem is essentially the same as the structural reliability problem in Eq. 2.4. Like in the case of the structural reliability problem, in general also the integral in Eq. 4.11 cannot be solved analytically. Approximate results can however be obtained through standard structural reliability methods (Straub and Der Kiureghian, 2010a). For example one can apply standard MCS for this purpose, by generating n_S samples $\mathbf{x}^i$ from the joint truncated distribution $f_{\mathbf{X}}\left(\mathbf{x} | d_{j-1}^1 < x_1 \leq d_j^1, \ldots, d_{k-1}^n < x_n \leq d_k^n\right)$. In this case the probability of interest can be calculated approximately as:

$$\Pr\left(g(\mathbf{x}) \in E | y_j^1, \ldots, y_k^n\right) \approx \frac{1}{n_S} \sum_{i=1}^{n_S} I_{g(\mathbf{x})\in E}\left(\mathbf{x}^i\right) \quad (4.12)$$

where $I_{g(\mathbf{x})\in E}\left(\mathbf{x}^i\right)$ is an indicator function, which is 1 if $g\left(\mathbf{x}^i\right) \in E$ and 0 otherwise. In cases where samples cannot readily be generated from the joint truncated PDF of the basic random variables $\mathbf{X}$, samples can alternatively be generated from another density $h_{\mathbf{X}}(\mathbf{x})$ and weighted accordingly. In this case the probability of interest reads:

$$\Pr\left(g(\mathbf{x}) \in E | y_j^1, \ldots, y_j^n\right) \approx \frac{1}{n_S} \sum_{i=1}^{n_S} I_{g(\mathbf{x}^i)\in E} \frac{f_{\mathbf{X}}\left(\mathbf{x}^i | d_{j-1}^1 < x_1 \leq d_j^1, \ldots, d_{k-1}^n < x_n \leq d_k^n\right)}{h_{\mathbf{X}}\left(\mathbf{x}^i\right)} \quad (4.13)$$

A possible choice for the sampling density $h_{X_i}(x_i)$ corresponding to component X_i, is the exponential distribution, if the interval corresponds to a boundary interval, and the uniform distribution otherwise (Straub, 2009).

Part II

INFERENCE

Chapter 5

Introduction to inference

A BN is an efficient way of representing the joint probability distribution of a number of random variables $\mathbf{X} = [X_1, \ldots, X_n]$. Based on this representation, BN inference algorithms can be used to answer queries about this joint probability distribution. Typically these are queries for the distribution of a variable X_i conditional on having observed the outcome of a set of other variables i.e. $\mathbf{X}^e = \mathbf{x}^e$, with $\mathbf{X}^e \subseteq \mathbf{X} \setminus X_i$. For the case where all random variables $\mathbf{X}$ are discrete, the conditional PMF of X_i can be computed as:

$$p\left(x_i|\mathbf{x}^e\right) = \frac{\sum_{\mathbf{x}^{-(i,\mathbf{e})}} p\left(\mathbf{x}\right)}{\sum_{\mathbf{x}^{-\mathbf{e}}} p\left(\mathbf{x}\right)} \tag{5.1}$$

where $\mathbf{x}^{-\mathbf{e}} = \mathbf{x} \setminus \mathbf{x}^e$ and $\mathbf{x}^{-(i,\mathbf{e})} = \mathbf{x} \setminus \{x_i, \mathbf{x}^e\}$. For continuous distributed random variables the sums in Eq. 5.1 are replaced by integrals and one can write:

$$f\left(x_i|\mathbf{x}^e\right) = \frac{\int_{\mathbf{x}^{-(i,\mathbf{e})}} f\left(\mathbf{x}\right) \mathrm{d}\mathbf{x}^{-(i,\mathbf{e})}}{\int_{\mathbf{x}^{-\mathbf{e}}} f\left(\mathbf{x}\right) \mathrm{d}\mathbf{x}^{-\mathbf{e}}} \tag{5.2}$$

In the case of a hybrid BN, the respective equation contains both, sums and integrals.
For a discrete random variables, calculating the marginal probabilities in the numerator and in the denominator of Eq. 5.1 is theoretically possible. For BNs containing continuous random variables, the required integrals can typically not be calculated analytically (cf. Eq. 5.2). Continuous/hybrid exceptions, for which exact inference is possible, are:

- Conditional linear Gaussian (CLG) BNs, where the marginal distributions of the continuous nodes are Gaussian and their means depend linearly on the states of their parents. A severe restriction for such CLGs is that discrete nodes must not have continuous parents (Shachter and Kenley, 1989; Lauritzen, 1992).
- BNs whose continuous distributions can be represented by Mixtures of truncated basis functionss (MoTBFs), for which exact inference is possible (Langseth et al., 2009, 2012).

These two special cases will not be further considered in the scope of this thesis. Instead two exact inference algorithms for discrete BNs are introduced in the following. By discretizing

continuous random variables these algorithms can also be applied for approximate inference in hybrid or continuous BNs (c.f. chapter 6). Sampling-based inference algorithms represent another type of approximate inference algorithms for general BNs. The most important sampling based inference algorithms are briefly introduced in chapter 7.

5.1 Exact inference algorithms for discrete BNs

While there exist analytical solutions for the sums in Eq. 5.1, the exponential increase of the CPT hinders inference directly through Eq. 5.1 for BNs with a medium to large number of random variables. Exact inference algorithms aim at exploiting the independence properties in a BN to make the required calculations feasible. For this purpose, the chain rule for BNs (Eq. 2.2) is introduced into Eq. 5.1:

$$p\left(x_i|\mathbf{x}^e\right) = \frac{\sum_{\mathbf{x}^{-(i,\mathbf{e})}} \prod_{x_j \in \mathbf{x}} p\left(x_i|pa\left(x_i\right)\right)}{\sum_{\mathbf{x}^{-\mathbf{e}}} \prod_{x_j \in \mathbf{x}} p\left(x_i|pa\left(x_i\right)\right)} \tag{5.3}$$

Two of the most prominent representatives of exact BN inference algorithms are the variable elimination algorithm (Zhang and Poole, 1994) and the junction tree algorithm (Lauritzen and Spiegelhalter, 1988; Jensen et al., 1990), which is typically implemented in standard BN software tools. Both of them are discussed briefly in the following for a detailed description the reader is referred to the standard textbooks (e.g. Jensen and Nielsen, 2007; Koller and Friedman, 2009; Kjærulff and Madsen, 2013).

5.1.1 Variable elimination algorithm

Consider the BN in Fig. 2.1 and assume one is interested in the state of X_3 conditional on $X_4 = x_4^e$. The posterior probability distribution $p\left(x_3|x_4^e\right)$ can be calculated through Eq. 5.3 as:

$$p\left(x_3|x_4^e\right) = \frac{\sum_{x_1,x_2,x_5} p\left(x_1\right) p\left(x_2\right) p\left(x_3|x_1\right) p\left(x_4^e|x_1\right) p\left(x_5|x_2,x_3,x_4^e\right)}{\sum_{x_1,x_2,x_3,x_5} p\left(x_1\right) p\left(x_2\right) p\left(x_3|x_1\right) p\left(x_4^e|x_1\right) p\left(x_5|x_2,x_3,x_4^e\right)} \tag{5.4}$$

Assuming each variable X_i has m states, one needs to perform summations in a table of size m^5. More efficiently Eq. 5.4 can be rewritten as:

$$p\left(x_3|x_4^e\right) = \frac{\sum_{x_1} p\left(x_1\right) p\left(x_3|x_1\right) p\left(x_4^e|x_1\right) \sum_{x_2} p\left(x_2\right) \sum_{x_5} p\left(x_5|x_2,x_3,x_4^e\right)}{\sum_{x_1} p\left(x_1\right) p\left(x_4^e|x_1\right) \sum_{x_3} p\left(x_3|x_1\right) \sum_{x_2} p\left(x_2\right) \sum_{x_5} p\left(x_5|x_2,x_3,x_4^e\right)} \tag{5.5}$$

Taking into account that the sum of a (conditional) PMF over its outcome space is 1 – here $\sum_{x_5} p\left(x_5|x_2,x_3,x_4^e\right) = \sum_{x_2} p\left(x_2\right) = \sum_{x_3} p\left(x_3|x_1\right) = 1$ – one can write:

$$p\left(x_3|x_4^e\right) = \frac{\sum_{x_1} p\left(x_1\right) p\left(x_3|x_1\right) p\left(x_4^e|x_1\right)}{\sum_{x_1} p\left(x_1\right) p\left(x_4^e|x_1\right)} \tag{5.6}$$

Eq. 5.5 is computationally more efficient, since the largest table one needs to perform summations on, is of size m^3. In general, it can be stated that the computational complexity of variable elimination is determined by the size of the tables generated. (Note: This is closely related to the so called clique size, which will be introduced in the next subsection.)
The size of the tables generated depends on the order, according to which the variables are eliminated respectively according to which the summations are executed. For the purpose of computational efficiency it is thus desirable to find an optimal elimination order. This is however computationally expensive (in fact NP-hard) (Arnborg et al., 1987). Fortunately for most practical problems it is possible to find an efficient variable elimination order through heuristic approaches (Kjærulff, 1990).
In a nutshell the variable elimination algorithm makes inference feasible by executing the required operations in an efficient order. This order depends on the variable of interest. Therefore, if the marginal distributions of n nodes in the network are to be computed the variable elimination algorithm needs to be run n times. In such situations it is easy to imagine that the efficiency of inference can be enhanced by restructuring the problem, such that parts of the computations that have already been performed can be reused. This is the basic idea behind the junction tree algorithm.

5.1.2 Junction tree algorithm

In the junction tree algorithm, a BN is transformed into a junction tree. In the junction tree, nodes represent cliques C_i. These cliques represent subsets of nodes in the BN that form a complete subgraph in a corresponding chordal graph. A chordal graph is a undirected graph, where there is no cycle of length larger than 3 that does not have chords i.e. links between two nodes in the cycle that are not part of the cycle (Koller and Friedman, 2009).
A chordal graph for a BN can be obtained based on its DAG. To this end, first the so-called moralized graph is generated by neglecting the directions of the links in the DAG and additionally introducing undirected links between nodes that have common children. To obtain a chordal graph from this moralized graph, additional chords are introduced until the graph does not contain undirected cycles of length larger than 3.
The cliques are connected through undirected links in a tree structure, such that if X belongs to two cliques C_i and C_l it also has to belong to all cliques that lie on the path between these cliques (running intersection property). To each link connecting two cliques C_i and C_j, a separator set is assigned. This separator set $S_{i,j}$ consists of the X_i's that both cliques have in common, i.e. $S_{ij} = C_i \cap C_j$. A potential, ψ_C, is assigned to each clique. The potential of a clique C_i is determined as the product of the potentials (CPTs) of the X_i's $\in C_i$. Furthermore two messages are associated to each link, connecting two cliques C_i and C_j, i.e one for each direction. The message sent from C_i to C_j is determined by summing the potential of the clique C_i over all X_i's $\notin S_{ij}$. The message sent from C_j to C_i is computed analogously. Based on this the marginal distribution of a node X_k can be calculated by:

- selecting a clique C_i, which contains X_k
- multiplying C_i's potential with the messages, directed towards it
- summing over all X_i's $\in C_i$ except X_k

In the following, the junction tree algorithm is applied to the simple BN example from Fig. 2.1. The corresponding moral graph is shown in Fig. 5.1. In this case the moral graph is chordal. The cliques can be derived by eliminating nodes from the moral graph. From the moral graph in Fig. 5.1 one can start by eliminating X_1. The associated clique consists of the node X_1 itself and all of its neighbors i.e. X_3 and X_4. After eliminating X_1 all other nodes X_2, X_3, X_4 and X_5 form a clique. Therefore the constructed junction tree (Fig. 5.2) consists of only two cliques.

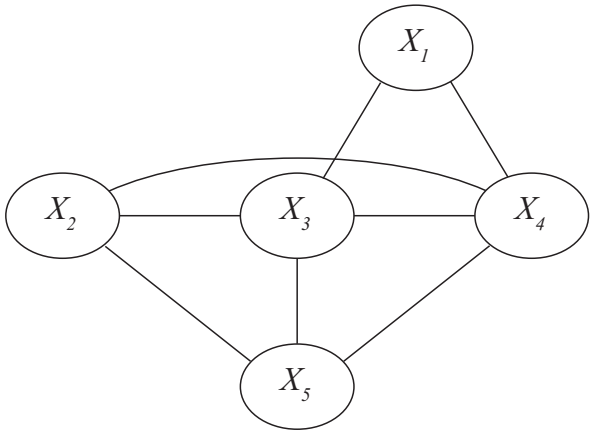


Figure 5.1: Moral graph for the BN in Fig. 2.1.

The potentials of the cliques can be calculated from the potentials of the nodes in the BN. These are:

- $\phi_1 = p(X_1)$ where the domain of ϕ_1, $\mathrm{dom}(\phi_1)$ is $\{X_1\}$
- $\phi_2 = p(X_2)$ where $\mathrm{dom}(\phi_2)$ is $\{X_2\}$
- $\phi_3 = p(X_3|X_1)$ where $\mathrm{dom}(\phi_3)$ is $\{X_1, X_3\}$
- $\phi_4 = p(X_4|X_1)$ where $\mathrm{dom}(\phi_4)$ is $\{X_1, X_4\}$
- $\phi_5 = p(X_5|X_2, X_3, X_4)$ where $\mathrm{dom}(\phi_5)$ is $\{X_2, X_3, X_4, X_5\}$

Thus the potential of clique C_1 reads:

$$\psi_{C_1} = \phi_1 \cdot \phi_3 \cdot \phi_4 \tag{5.7}$$

And the potential of clique C_2 reads:

$$\psi_{C_2} = \phi_2 \cdot \phi_5 \tag{5.8}$$

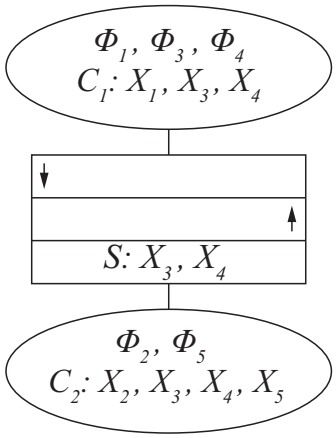


Figure 5.2: Junction tree for the BN in Fig. 2.1.

There are two messages attached to the link between C_1 and C_2. The message $\psi^{\downarrow}$ is sent from C_1 to C_2 and $\psi^{\uparrow}$ the reverse direction. $\psi^{\uparrow}$ is calculated from the BN nodes' potentials as:

$$\psi^{\uparrow} = \sum_{x_j \in C_2 \setminus S} \phi_2 \cdot \phi_5 = \sum_{x_2, x_5} \phi_2 \cdot \phi_5 \tag{5.9}$$

And $\psi^{\downarrow}$ analogously as:

$$\psi^{\downarrow} = \sum_{x_j \in C_1 \setminus S} \phi_1 \cdot \phi_3 \cdot \phi_4 = \sum_{x_1} \phi_1 \cdot \phi_3 \cdot \phi_4 \tag{5.10}$$

From this the marginal distribution of e.g. X_2 can be calculated as:

$$p(X_2) = \sum_{x_3, x_4, x_5} \psi^{\downarrow} \cdot \phi_2 \cdot \phi_5 \tag{5.11}$$

Within one evidence scenario the messages can be used for calculating the marginal distribution of any node $X_i \in \mathbf{X}$.

5.2 Computational limitations of exact inference

Exact inference is limited to discrete BNs and a couple of hybrid special cases, for which there exists an exact solution for the required integrals. Besides that, computer hardware and in particular computation time and memory space, can be a limiting factor.
Already in the phase of model elicitation memory can become an issue. This is especially the case, when CPTs become too large, e.g. because a node has many parents. In the context of inference, memory issues may furthermore arise if the size of the tables that need to be handled during variable elimination respectively the size of the cliques' factors that need to be handled within the junction tree algorithm become too large. These are dependent on the sizes of the CPTs of the nodes and also on the BN structure and its corresponding elimination order\junction tree. Further discussions on the computational complexity of exact inference algorithms are provided in (e.g. Jensen et al., 1990; Koller and Friedman, 2009).
In some situations, in which computational costs hinder the application of exact inference algorithms to discrete BNs, approximate inference (e.g. through sampling based methods) can represent a worthwhile alternative.

Chapter 6

Efficient discretization for rare events

In the previous chapter, algorithms for exact inference in discrete BNs were presented. The same algorithms can be used for approximate inference in hybrid or continuous BNs if the continuous random variables are discretized. Static discretization (i.e. prior to inference) in combination with exact inference algorithms is of special interest for applications that require the near real-time performance of exact inference algorithms. Standard static discretization methods like equal-width discretization or equal-frequency discretization are not tailored for reliability problems with small failure probabilities. In this chapter, the development of an efficient discretization procedure for reliability problems, where the performance of a system is defined through a LSF and where the failure probabilities are potentially small is described.[9]

6.1 Methodology

6.1.1 Treatment of a reliability problem in a BN

Discrete BNs and structural reliability concepts are combined to facilitate updating of rare event (failure) probabilities under new observations. The general problem setting is illustrated in the BN of Fig. 6.1. The presentation here is limited to component reliability problems; system problems are considered later. The binary random variable *Component performance* is described by the LSF $g(\mathbf{X})$.

The basic random variables $\mathbf{X}$ of the model are included in the BN as parents of *Component performance*. The nodes M_i represent measurements of individual random variables X_i, and nodes I_j represent factors influencing the basic random variables. Dependence between the variables in $\mathbf{X}$ is modeled either directly by links among them (here $X_2 \rightarrow X_1$ and $X_4 \rightarrow X_5$) or through common influencing factors (here $I_2 \rightarrow X_3$ and $I_2 \rightarrow X_4$) the component performance node can have (multiple) child nodes. However that does not have an impact on the discretization of the reliability problem. Ultimately, the goal is to predict the component performance, i.e. $\Pr(F)$, conditional on observations of other variables, typically of the

[9] This chapter is adapted from (Zwirglmaier and Straub, 2016a).

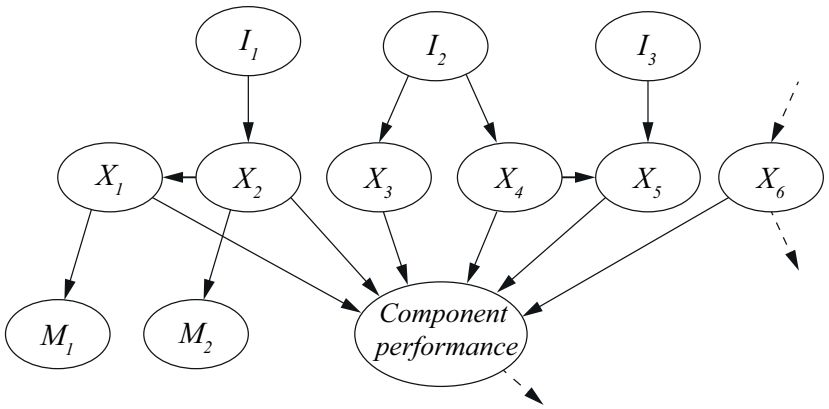


Figure 6.1: General structure of a BN model of a component reliability problem.

measurement variables M_i, but possibly of any other random variable in the BN, such as the influencing variables I_j. Whenever new evidence on these variables is available, the BN should be evaluated (in near real-time) utilizing exact BN inference algorithms. To enable exact inference algorithms, all continuous random variables are discretized. These include the $\mathbf{X}$, and possibly the M_i and I_j. In the general case, the computational effort for solving the BN is a direct function of the CPT size of *Component performance*. The size of this CPT is $2\prod_{i=1}^{n} n_i$ where n is the number of random variables in $\mathbf{X}$, and n_i is the number of states used for discretizing X_i. Discretization of random variables M_i and I_j is not described here, since it is typically straightforward and does not contribute significantly to computational performance. The key parameter for computational efficiency and accuracy is the discretization scheme for $\mathbf{X}$, which will be described in this section. Since in many cases it is computationally not feasible to represent all basic random variables $\mathbf{X}$ as nodes in a discrete BN, prior to the discretization procedure an approach for simplifying a BN through removal of some of the X_i's is presented in chapter 3.

6.1.2 Discretization of basic random variables

For ease of presentation, first discretization of statistically independent basic random variables $\mathbf{X}$ is considered. This is shown in Fig. 6.2, which represents a special case of the BN from Fig. 6.1, in which the X_i's have no parents. In general in a BN the basic random variables $\mathbf{X}$ are independent if they are not connected through links, they have no common (unknown) ancestors and no evidence is available on any of their descendants. The proposed procedure is extended to the general case of dependent basic random variables thereafter.

Independent basic random variables

The situation is illustrated in Fig. 6.2. The performance of the component depends on n statistically independent random variables and is described by a LSF $g(\mathbf{X}) = g(X_1, \ldots, X_n)$. For all basic random variables X_i, corresponding measurements M_i can be performed. To

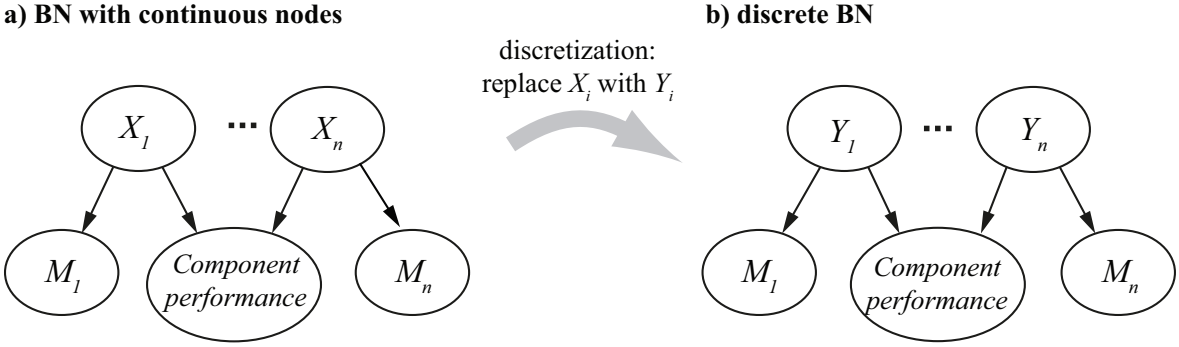


Figure 6.2: Representation of a basic reliability problem with n independent basic random variables in a BN. Left: original problem with continuous basic random variables X_i, right: discrete BN, in which X_i are substituted with discrete nodes Y_i.

obtain an equivalent discrete BN, the continuous X_i are replaced by the discrete random variables Y_i, and the LSF is replaced by the CPT of component performance conditional on $Y = [Y_1; \ldots; Y_n]$. For each discrete random variable Y_i with n_i states $1, 2, \ldots, n_i$, we define a discretization scheme $D_i = \left[d_0, d_1, \ldots, d_{n_{i-1}}, d_{n_i}\right]$ consisting of n_{i+1} interval boundaries. The first and the last interval boundaries are given by the boundaries of X_i's outcome space.

Since here the X_i, and thus the Y_i, have no parents, the PMF of Y_i is defined as:

$$p_{Y_i}(j) = F_{X_i}(d_j) - F_{X_i}(d_{j-1}); \text{with } j \in [1, \ldots, n_i] \tag{6.1}$$

where F_{X_i} denotes the CDF of X_i. The probability of failure corresponding to the discrete BN in Fig. 6.2 can be calculated as:

$$\Pr(F) = \sum_{y_1=1}^{n_1} \ldots \sum_{y_n=1}^{n_n} p_{Y_1}(y_1) \cdot \ldots \cdot p_{Y_n}(y_n) \cdot \Pr(F | Y_1 = y_1 \cap \ldots \cap Y_n = y_n) \tag{6.2}$$

Note that the discretization does not introduce any approximation error here, as long as the conditional probability $\Pr(F|Y_1 = y_1 \cap \ldots \cap Y_n = y_n)$ is computed exactly.

Once measurements from the nodes $\mathbf{M} = [M_1; \ldots; M_n]$ are available, the conditional failure probability can be calculated as:

$$\begin{aligned}\Pr(F|\mathbf{M} = \mathbf{m}) \approx \frac{1}{p_{\mathbf{M}}(\mathbf{m})} \sum_{y_1=1}^{n_1} \ldots \sum_{y_n=1}^{n_n} & p_{Y_1}(y_1) \cdot p_{M_1|Y_1}(m_1|y_1) \cdot \ldots \cdot p_{Y_n}(y_n) \cdot \\ & p_{M_n|Y_n}(m_n|y_n) \cdot \Pr(F|Y_1 = y_1 \cap \ldots \cap Y_n = y_n)\end{aligned} \tag{6.3}$$

where $\Pr(F|Y_1 = y_1, \ldots, Y_n = y_n)$ is the conditional probability of component failure given $y_1, \ldots, y_n$. If no measurements are available for some of the basic random variables, the corresponding likelihood terms $p_{M_j|Y_j}(m_j|y_j)$ are simply omitted in Eq. 6.3. While the computation of the unconditional failure probability following Eq. 6.2 is exact, the computation of the conditional failure probability through Eq. 6.3 is only an approximation. The reason is

that the dependence between the measurement variable M_i and the *Component performance* variable is not fully captured in the discrete BN (see also Straub and Der Kiureghian, 2010a). In Fig. 6.3, this is illustrated for a reliability problem with one basic random variable X_i. Both the continuous distribution (Fig. 6.3a) and the corresponding discretized distribution (Fig. 6.3b) are updated correctly after observing M_1. However, for Eq. 6.3 to be exact, also the conditional failure probabilities $\Pr(F|Y_1 = y_1)$ would need to be updated.

This can be observed in Fig. 6.3a: in interval $Y_1 = 3$, which is the one cut by the limit state surface, the ratio of the probability mass in the failure domain to that in the safe domain changes from the prior to the posterior case, i.e. $\Pr(F|Y_1 = 3) \neq \Pr(F|Y_1 = 3, M_1 = m_1)$. Since the computation of the conditional failure probability following Eq. 6.3 is based on the prior probability $\Pr(F|Y_1 = 3)$, the discretization introduces an approximation in this case. The error occurs only in the intervals that are cut by the limit state surface. In the simple one-dimensional case of Fig. 6.3, an optimal discretization approach would be to discretize the whole outcome space in two intervals, one capturing the survival and one the failure domain. This discretization would have zero approximation error. However, already in a two-dimensional case, such a solution is not possible. This is illustrated in Fig. 6.4, where the cells cut by the limit state surface are indicated in grey. The failure probability conditional on measurements calculated according to Eq. 6.3 will necessarily be an approximation. The approximation error will be small, if the contribution of the cells cut by the limit state surface (the grey cells in Fig. 6.4) to the total failure probability is small. An efficient discretization will thus limit this contribution with as few intervals as possible.

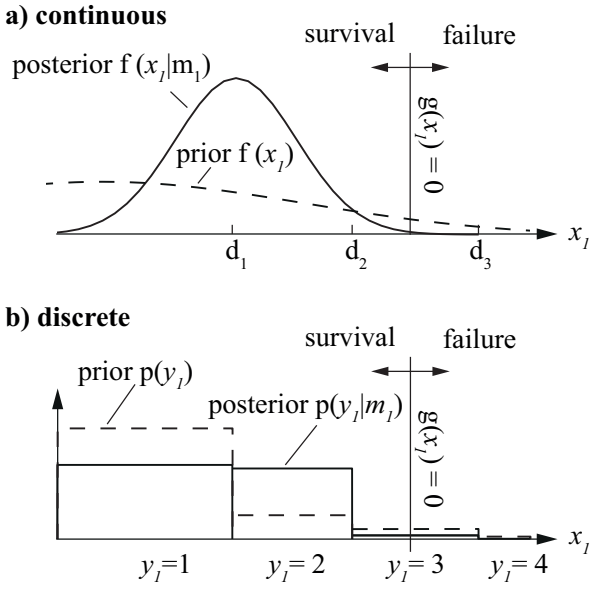


Figure 6.3: Discretization error in 1D.

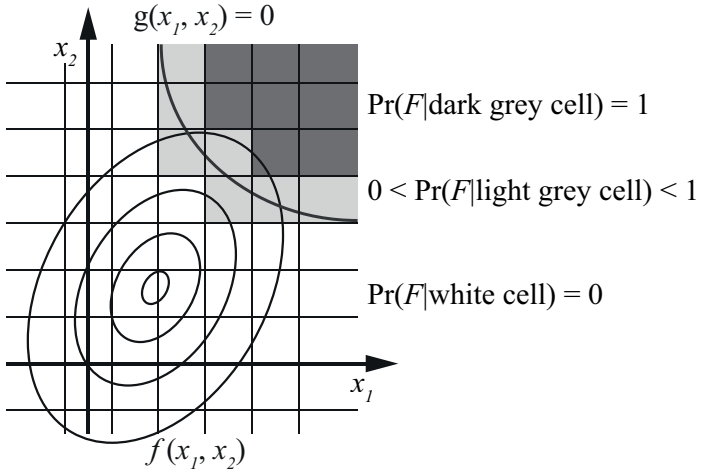


Figure 6.4: General structure of a BN model of a component reliability problem.

Dependent basic random variables

Eqs. (6.1-6.3) must be adjusted when dependence among the X_i's is present, in accordance with the case-specific BN structure. However, the principles outlined above for independent $X_1, \ldots, X_n$ hold equally for dependent basic random variables: The discretization error is a function of the cells cut by the LSF. When determining an optimal discretization, it is proposed in the following to find the FORM approximation of the reliability problem, which can readily account for the dependence among the random variables. Hence, there is no need to distinguish between the cases with independent or dependent random variables.

6.1.3 Efficient discretization

Optimal discretization of linear problems in standard normal space

To find an efficient discretization of $\mathbf{X}$, the FORM solution to the reliability problem is considered. Evaluating the linearized FORM LSF $G_L(\mathbf{U})$ is computationally inexpensive once the design point $\mathbf{u}^*$ is known. Therefore, it is feasible to find a discretization of $\mathbf{U}$ that is optimal for the event $\{G_L(U) \leq 0\}$ through optimization. If $G(\mathbf{U})$ is not strongly non-linear, this solution will be an efficient discretization for $\{G(\mathbf{U}) \leq 0\}$ and, after a transformation to the original space, also for $\{g(\mathbf{X}) \leq 0\}$. As discussed in subsection 6.1.2, the approximation error of the discretization is associated with the change from the prior to the posterior distribution of the basic random variables. A measure of optimality must thus consider possible measurements of $\mathbf{X}$ or $\mathbf{U}$. Considered are hypothetical measurements $\tilde{M}_i$ (this notation is used to distinguish hypothetical measurements from actual measurements $M_i = m_i$) of all standard normal random variables U_i with additive measurement error $\epsilon_i \sim N(0, \sigma_{\epsilon_i})$. Since both the prior distribution and the measurement error are normal distributed, the likelihood is also normal distributed:

$$\tilde{M}_i | \{U_i = u_i\} \sim N(u_i, \sigma_{\epsilon_i}) \tag{6.4}$$

The posterior, i.e. the conditional distribution of U_i given a measurement outcome $\tilde{M}_i = \tilde{m}_i$, is the normal distribution with mean $\frac{1}{1+\sigma_{\epsilon_i}^2}\tilde{m}_i$ and standard deviation $\sqrt{\left(1 - \frac{1}{1+\sigma_{\epsilon_i}^2}\right)}$. An error measure is defined that is based on comparing the true posterior probability of failure $P_{F|\tilde{M}}(\tilde{m})$ with the posterior probability of failure calculated from the discretized $\mathbf{U}$, denoted by $\hat{P}_{F|\tilde{M}}(d;\tilde{m})$. Here, $\mathbf{d}$ are the parameters defining the discretization. The proposed error measure is:

$$e(\mathbf{d},\tilde{\mathbf{m}}) = \left| \frac{\log_{10}\hat{P}_{F|\mathbf{M}}(\mathbf{d};\tilde{\mathbf{m}}) - \log_{10} P_{F|\tilde{\mathbf{M}}}(\tilde{\mathbf{m}})}{\log_{10} P_{F|\tilde{\mathbf{M}}}(\tilde{\mathbf{m}})} \right| \tag{6.5}$$

The error measure of Eq. 6.5 represents a tradeoff between the absolute and the relative error. It weights the (logarithmic) relative error by the magnitude of the posterior failure probability. This ensures that the same relative error is considered worse at a higher probability level compared to an error at a lower probability level. A-priori, the measurement outcomes are not known. Hence we define the optimal discretization as the one that minimizes the expected preposterior error $E_{\tilde{\mathbf{M}}}\left[e\left(\mathbf{d},\tilde{\mathbf{M}}\right)\right]$:

$$\mathbf{d}^{opt} = \arg\min_{\mathbf{d}} \int_{\tilde{\mathbf{M}}} e(\mathbf{d},\tilde{\mathbf{m}})\, f_{\tilde{\mathbf{M}}}(\tilde{\mathbf{m}})\, \mathrm{d}\tilde{\mathbf{m}} \tag{6.6}$$

The optimization is thus based on the computation of an expected value with respect to the possible measurements outcomes $\tilde{\mathbf{M}} = \tilde{\mathbf{m}}$ before having taken any measurements. This is analogous to a preposterior analysis (Raiffa and Schlaifer, 1961; Straub, 2014a). However, unlike in traditional preposterior analysis, the objective is not to identify an optimal action under future available information, but to find the optimal discretization parameters $\mathbf{d}^{opt}$. The integral in Eq. 6.6 is evaluated through a simple Monte Carlo approach. All $\tilde{M}_i$ have the normal distribution with zero mean and variance $1 + \sigma_\epsilon^2$. The parameters in $\mathbf{d}$ describing the discretization scheme are:

- n_i: number of intervals used to discretize each random variable U_i,
- w_i: width of the discretization frame in the dimension of U_i, and
- v_i: position of the midpoint of the discretization frame relative to the design point

These parameters are illustrated in Fig. 6.5. For a problem with n basic random variables, the full set of optimization parameters is $\mathbf{d} = [w_1,\ldots,w_n,n_1,\ldots,n_{n-1},v_1,\ldots,v_n]$. Clearly, the discretization error reduces with increasing n_i. Because the computational efficiency of the final BN is a direct function of the size of the CPT associated with component performance, which is $\prod_{i=1}^{n} n_i$, its size is constrained. To this end, c_{up} is defined as the maximum allowed number of parameters of the CPT of the *component performance node*. This puts a constraint on the optimization of Eq. 6.6:

$$\prod_{i=1}^{n} n_i \leq c_{up} \tag{6.7}$$

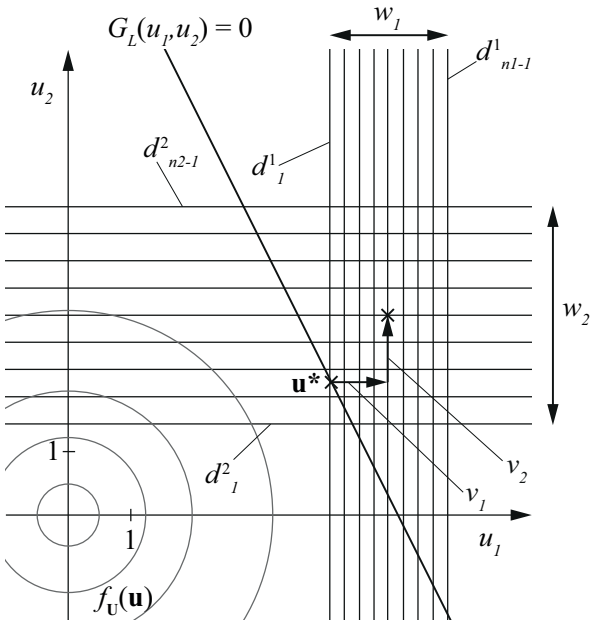


Figure 6.5: Schematic representation of a discretization of a linear 2D reliability problem. w_i is the distance between interval boundaries d^i_1 and $d^i_{n_i-1}$. All intervals between these boundaries are equi-spaced. v_i is the position of the midpoint of the discretization frame relative to the design point $\mathbf{u}^*$ in dimension i.

The optimization is implemented through a two-level approach. The optimization of the continuous parameters width w_i and position of the discretization frame v_i for all $i = 1, \ldots, n$ is carried out using unconstrained nonlinear optimization for fixed values of n_i. The optimization of the discrete n_i is performed through a local search algorithm. Note that the optimization is performed offline, i.e. prior to running the BN, hence it does not affect the goal of near real-time performance of the BN. Furthermore, in section 6.2 a heuristic is derived that can replace the time-consuming solution of the optimization problem.

Efficient discretization of the original random variables X

Since the nodes in the BN represent random variables $\mathbf{X}$ in their original outcome space, the discretization schemes, which are derived for the corresponding standard normal random variables $\mathbf{U}$, need to be transformed to the X-space. In the case of mutually independent random variables X_i, any point on the i-th interval boundary in U-space – if transformed – will result in the same corresponding i-th interval boundary in X-space. This is not the case for dependent random variables X_i, where a mapping of the interval boundaries in U-space to X-space will not lead to an orthogonal discretization scheme in X-space. To preserve orthogonality throughout the transformation, we propose to represent each interval boundary through a characteristic point and determine the boundary in X-space through a transformation of this point. For transforming the interval boundary of X_i, the characteristic point is selected as

the design point $\mathbf{u}^*$, where the i-th element is substituted by the coordinate of the interval boundary. In Fig. 6.6 this is shown for an example with $n = 2$ random variables.

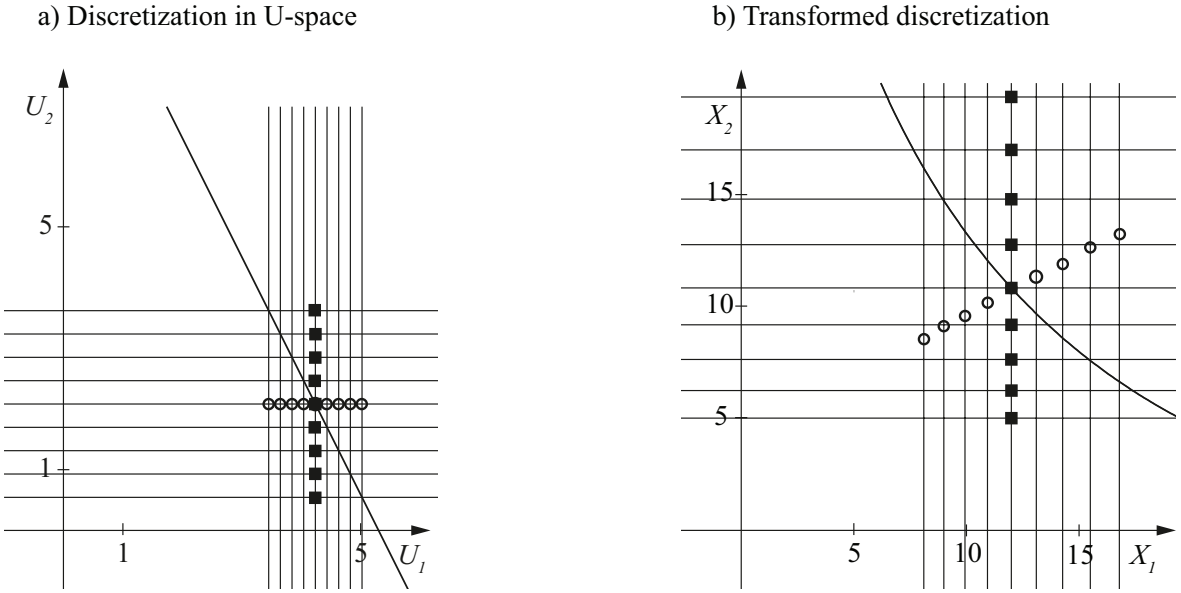


Figure 6.6: Transformation of a discretization scheme from U-space to X-space. To preserve orthogonality each interval boundary in U-space is represented by a characteristic point. The random variables X_1 and X_2 are Weibull distributed with scale and shape parameter 1 and their correlation is 0.5.

6.2 Development of an efficient discretization procedure

6.2.1 Optimization of the FORM approximation

We present the optimal discretization for the FORM approximation $G_L(\mathbf{U})$ for $n = 2$ and $n = 3$ dimensions. Extension to higher numbers of random variables is discussed. Because the linear LSF employed in FORM is described only by the reliability index β_{FORM} and the vector $\boldsymbol{\alpha}$ of FORM sensitives (Eq. 2.7), it facilitates parametric studies. Initially, we consider a reliability index $\beta_{FORM} = 4.26$, corresponding to a probability of failure of 10^{-5}. The standard deviation of the additive measurement error is set to either $\sigma_\epsilon = 0.5$ or $\sigma_\epsilon = 1.0$. Different combinations of FORM sensitivity values α_i are selected, to investigate their effect on the optimal discretization. In all investigated cases, we find that the position of the midpoint of the optimal discretization frame coincides with the design point, i.e. $v_i^{opt} = 0$. The optimal discretization widths w_i^{opt} vary significantly with the importance measures α_i, as shown in section 6.2.1. However, the optimal number of intervals n_i^{opt} is approximately the same for all random variables in all investigated cases, independent of the α_i values, i.e. $n_i^{opt} = c_{up}^{1/n}$.

On why the optimal number of intervals n_i^{opt} is independent of α_i

To better understand why the number of intervals does not depend on $|\alpha_i|$ (for $|\alpha_i|$ that are significantly larger than 0), recall that an efficient discretization scheme should focus on the area around the limit state surface. More precisely, the discretization error in the posterior case is induced by the cells that are cut by the limit state surface. Exemplarily, Fig. 6.7 shows a linear problem in standard normal space with two basic random variables U_1 and U_2, where $\alpha_1 = 0.45$ and $\alpha_2 = 0.89$. While Fig. 6.7a shows an optimal discretization with 5 intervals per dimension, Fig. 6.7b shows a discretization scheme, where the more important random variable U_2 is discretized with 6 intervals and U_1 with 4 intervals. In both cases, the discretization frame is centered at the design point. It is observed that the probability mass of the (grey) cells, associated with the discretization error in the posterior case, is higher in Fig. 6.7b than for the optimal discretion scheme in 6.7a. In the example shown here it is $1.4 \cdot 10^{-3}$ compared to $2.5 \cdot 10^{-4}$.

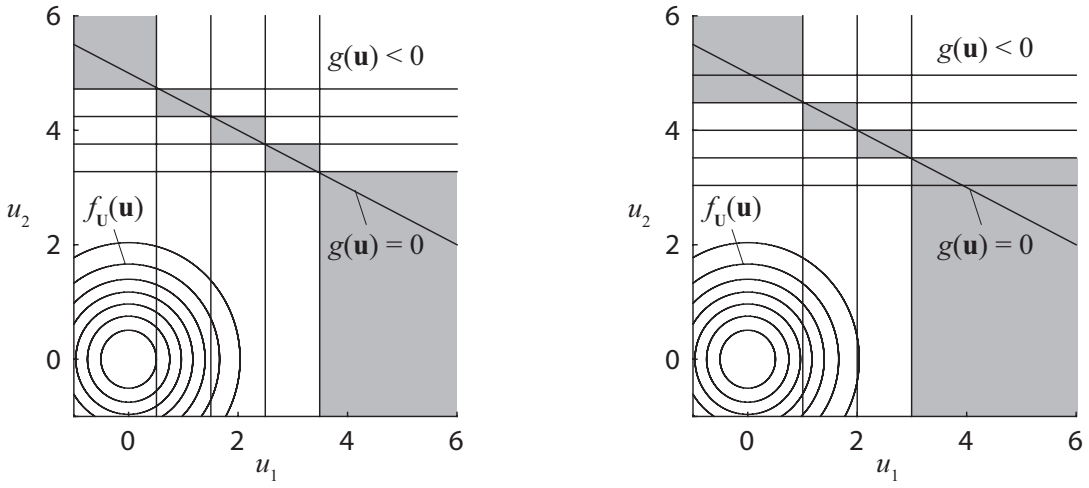


Figure 6.7: Linear problem in standard normal space, with two random variables U_1 and U_2, where $\alpha_1 = 0.45$ and $\alpha_1 = 0.89$. The intervals cut by the limit state surface, i.e. those which potentially lead to a posterior discretization error are marked in grey.

For input random variables X_i with a value of $|\alpha_i|$ close to zero, the above observations do not hold. Following section 3.2.2, these variables should be removed from the BN prior to discretization.

Dependence of the optimal discretization width on α_i

In the optimization, it is found that the optimal discretization width w_i^{opt} varies strongly with the random variable's importance, expressed through α_i. It is reminded that the width w_i describes the domain in which a fine discretization mesh is applied (Fig. 6.5). In general, w_i^{opt} decreases with increasing α_i. This effect can be observed in Fig. 6.7a, where U_2 is the

more important input random variable and it is $w_2^{opt} < w_1^{opt}$. A clear relation between w_i^{opt} and α_i can be observed by plotting the (logarithm of the) probability mass enclosed by w_i^{opt} against α_i, as shown in Fig. 6.8. The results of Fig. 6.8 indicate that the probability mass contained within this interval should be a direct function of α_i. The more important the variable, the finer the discretization around the design point should become. The observed relationship between this probability mass and α_i follows a clear trend, and a function can be fitted (Fig. 6.8). Neither the dimensionality of the problem nor the standard deviation of the measurement error appear to have an influence on this relation. However, as shown in the following section, it is found that the relation does depend on the prior failure probability of the problem (i.e. on β_{FORM}) and on the number of intervals n_i used to discretize the domain.

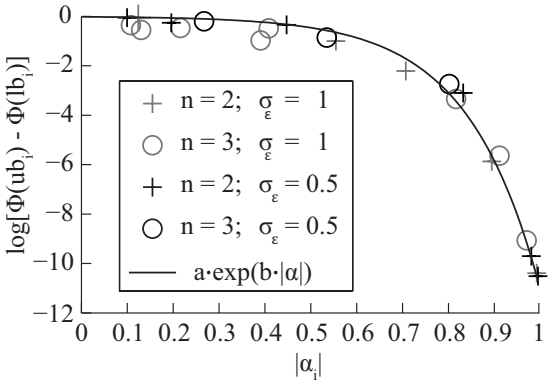


Figure 6.8: Logarithm of the probability mass enclosed by the discretization frame plotted against $|\alpha_i|$. Φ denotes the standard normal CDF and ub_i respectively lb_i the last (upper) and the first (lower) interval bound in dimension i.

To facilitate the application in practice and extending the results to larger numbers of random variables, in section 6.2.2 parametric functions are fitted to the optimization results to capture the dependency between the optimal discretization width w_i^{opt} and the FORM importance measures α_i.

Dependence of the optimal discretization on the reliability index β and the number of discretization cells c_{up}

The influence of the prior failure probability and the maximum size of the CPT, c_{up}, on the optimal discretization is investigated through 10 problems with $n = 2$ random variables in standard normal space. The FORM importance measures of the random variables are selected between 0.1 to 0.995 and the standard deviation of the measurement error is fixed to $\sigma_{epsilon} = 1.0$. We find that the optimal discretization frame is generally centered at the design point, i.e. $v_i^{opt} = 0$, and that the intervals are distributed uniformly among the dimensions. Firstly, the maximum CPT size c_{up}, i.e. the total number of discretization cells, is varied. The reliability index is $\beta_{FORM} = 5.2$. Fig. 6.9 shows the influence of c_{up} on the resulting width of the discretization frame w_i. Three cases are considered: $c_{up} = 25$, $c_{up} = 100$ and

Table 6.1: Parameters a and b of Eq. 6.8 for $\beta = 3.1$, $\beta = 4.3$ and $\beta = 5.2$ as well as 5, 10 and 20 intervals per dimension.

(a, b)	$n_i = 5$	$n_i = 10$	$n_i = 20$
$\beta = 3.1$	$(-0.28, 2.9)$	$\left(-1.6 \cdot 10^{-2}, 5.8\right)$	$\left(-9.8 \cdot 10^{-4}, 8.7\right)$
$\beta = 4.3$	$(-0.15, 4.3)$	$\left(-2.4 \cdot 10^{-2}, 6.1\right)$	$\left(-2.1 \cdot 10^{-2}, 6.2\right)$
$\beta = 5.2$	$(-0.36, 3.7)$	$(-0.11, 5.0)$	$\left(-3.7 \cdot 10^{-2}, 6.0\right)$

$c_{up} = 400$. These choices correspond to 5, 10 and 20 intervals for each random variable. The left side of Fig. 6.9 shows the relation between the optimal w_i and α_i^2. The right side of Fig. 6.9 shows the same relation, where the w_i's are scaled as in Fig. 6.8, i.e. the logarithm of the probability mass enclosed by the outer interval boundaries is depicted. As in Fig. 6.8, there is a clear dependence between the scaled w_i values and the α_i^2's. The interval frames increase with increasing number of random variables. Secondly, the prior failure probability is varied from 10^{-3} ($\beta = 3.1$) to 10^{-7} ($\beta = 5.2$). The results are shown in Fig. 6.10. Again, a distinct dependence between the scaled w_i values and the α_i^2's is found. The interval frames decrease with increasing reliability index (with decreasing failure probability).

6.2.2 Parametric function of optimal discretion frame

As evident from Fig. 6.9 and Fig. 6.10, there is a clear dependence of the probability mass enclosed by the optimal discretization frame (with width w_i) on the FORM sensitivity values α_i^2. The following parametric function captures this dependence:

$$\log\left(\Phi\left(ub_i\right) - \Phi\left(lb_i\right)\right) = a \cdot \exp\left(b \cdot |\alpha_i|\right) \tag{6.8}$$

ub_i is the upper and lb_i the lower interval boundary in dimension i, such that $w_i = ub_i - lb_i$. a and b are the parameters of the exponential function. This function is depicted in Figs. 6.9 and 6.10. Tab. 6.1 shows the parameter values a and b for the different combinations of the prior reliability index β and number of intervals per dimension

From the left sides of Fig. 6.9 and Fig. 6.10, it can be observed that the relation between α_i^2 and the optimal w_i is fairly diffuse for random variables with $|\alpha_i| < 0.6$. Here, the parametric relationship of Eq. 6.8 is less accurate. However, these random variables by definition have lower importance on the reliability estimate. Hence, the inaccuracy of Eq. 6.8 for random variables with $|\alpha_i| < 0.6$ is not critical, as is confirmed by the numerical investigations performed in the remainder of the paper. The parameter values of Tab. 6.1 are derived from two-dimensional problems. In Fig. 6.8 it is shown that there are no notable differences between two and three dimensions. On this basis, it is hypothesized that the heuristics are applicable also to problems with higher dimensions. This assumption is furthermore supported by the verification examples presented in section 6.3, where the heuristics are applied also to four-dimensional problems without any notable deterioration in the results.

6.2.3 Summary of the proposed procedure

The steps of the proposed procedure are:

1. Formulate the reliability problem
2. Set up the corresponding BN
3. Perform a FORM analysis for the reliability problem
4. Simplify the BN by removing nodes based on:
   (a) their importance for prediction
   (b) their observability
   (c) whether or not a node simplifies modeling of dependencies
   (d) whether or not it is desired to explicitly show a node in the BN for communication purposes
5. Find the discretization scheme in U-space based on the proposed heuristics i.e.:
   (a) the discretization scheme is centered at the design point from the FORM analysis
   (b) the same number of intervals is used for each random variable
   (c) the width of the discretization frame follows Eq. 6.8
6. Transform the discretization scheme to X-space
7. Compute the CPTs of the *component performance* node and the basic random variables using Monte Carlo simulation or Latin hypercube sampling

A MATLAB-based software tool performing these steps is available for download under www.era.bgu.tum.de/software.

6.3 Verification examples

6.3.1 Verification example I

For verification purposes, the proposed methodology is applied to the discretization of a general limit state with non-normal dependent random variables. The approximation error made by this discretization is investigated for different measurement outcomes. Failure is defined through the LSF $g(\mathbf{x})$:

$$g(\mathbf{x}) = a - \prod_{i=0}^{n} X_i \tag{6.9}$$

i.e., failure corresponds to the event $\{\prod_{i=1}^{n} X_i \geq a\}$. The basic random variables are distributed as $X_1 \sim LN(0, 0.5)$ and $X_2, \ldots, X_n \sim LN(1, 0.3)$ (values in parenthesis are the

parameters of the lognormal distribution). The statistical dependence among the X_i is described through a Gaussian copula model, with pairwise correlation coefficients ρ_{ij}. The parameters a and ρ_{ij} determine the prior failure probability P_F. Measurements $M_i = m_i$ are available for all basic random variables; they are associated with multiplicative measurement errors $\epsilon_i \sim LN(0, 0.71)$. In Tabs. 6.2 and 6.3, different cases with 3 and 4 random variables are shown. These cases differ with respect to the prior failure probability P_F, the correlation between the random variables ρ_{ij} and the observed measurements $\mathbf{m}$. For each case, a reference solution $P_{F|\mathbf{M}}$ is calculated analytically.

Table 6.2: Evaluation of the discretization error for different measurement outcomes m, for problems with $n = 3$ random variables. a is the constant in the LSF, Eq. 6.9; ρ_{ij} is the correlation coefficient between X_i and X_j for all $i \neq j$; P_F and $P_{F|\mathbf{M}}$ denote the analytically calculated prior and posterior failure probabilities; $\hat{P}_{F|\mathbf{M}}$ is the conditional failure probability calculated with the discrete BN.

a	c_{up}	ρ_{ij}	P_F	$\mathbf{m}$	$P_{F\|\mathbf{M}}$	$\hat{P}_{F\|\mathbf{M}}$	Abs. error	Rel. error [%]
100	10^3	0	3.6e-5	$[3.0, 2.9, 2, 9]$	4.3e-5	4.5e-5	3e-6	6
100	10^3	0	3.6e-5	$[2.3, 1.1, 2, 1]$	4.6e-6	5.3e-6	7e-7	14
100	10^3	0	3.6e-5	$[0.9, 2.4, 0.9]$	2.8e-7	3.5e-7	7e-8	25
200	15^3	0.5	1.6e-4	$[1.6, 2.0, 1.2]$	1.4e-6	1.4e-6	1e-7	4
400	8^3	0.5	6.4e-6	$[2.6, 3.0, 3.2]$	8.2e-7	8.9e-7	7e-8	9
400	12^3	0.5	6.4e-6	$[3.6, 3.3, 4, 3]$	4.9e-6	5.0e-6	1e-9	3

The results in Tabs. 6.2 and 6.3 show that the proposed methodology for discretization leads to errors in the posterior probability estimate that are acceptably small for most engineering applications. (It is reminded that discretization does not lead to a discretization error in the prior case.) As expected, the relative error is larger when the posterior probability is low, and the absolute error is larger when the posterior probability is high. This follows from the error measure defined in Eq. 6.5, which balances the relative with the absolute error. In addition,

Table 6.3: Evaluation of the discretization error for different measurement outcomes m. The number of random variables $n = 4$; a is the constant in the LSF, Eq. 6.9; ρ_{ij} is the correlation coefficient between X_i and X_j for all $i \neq j$; P_F and $P_{F|M}$ denote the analytically calculated prior and posterior failure probabilities; $\hat{P}_{F|\mathbf{M}}$ is the conditional failure probability calculated with the discrete BN.

a	c_{up}	ρ_{ij}	P_F	$\mathbf{m}$	$P_{F\|\mathbf{M}}$	$\hat{P}_{F\|\mathbf{M}}$	Abs. error	Rel. error [%]
400	10^4	0	1.7e-5	$[2.2, 3.2, 2.4, 3.4]$	9.5e-6	1.0e-5	9e-7	9
400	10^4	0	1.7e-5	$[1.6, 1.6, 1.6, 2.0]$	6.5e-7	7.9e-7	1e-7	21
400	10^4	0	1.7e-5	$[1.1, 2.3, 1.9, 1.2]$	2.4e-7	3.0e-7	6e-8	26
600	10^4	0.5	1.3e-3	$[3.3, 1.7, 2.8, 2.6]$	4.2e-4	4.3e-4	2e-5	4
800	8^4	0.5	5.3e-4	$[1.9, 2.0, 1.9, 2.4]$	1.8e-5	1.9e-5	1e-6	8

the results do not display any apparent effect of correlation on the accuracy.
To assess the effect of the choice of the number of discretization intervals, the failure probability $\hat{P}_{F|\mathbf{M}}$ was calculated for a discretization scheme with up to 20 intervals per RV for the fourth measurement case in Tab. 6.2. The estimated failure probabilities $\hat{P}_{F|\mathbf{M}}$ are plotted together with the exact solution in Fig. 6.11.

6.3.2 Verification example II

The failure criterion applied in verification example I (Eq. 6.9) leads to a linear LSF in U-space. To verify the accuracy of the proposed method for problems with non-linear LSFs in U-space, we additionally investigate the following LSF:

$$g(\mathbf{x}) = a - \sum_{i=1}^{n} X_i \tag{6.10}$$

Again the basic random variables X_1 to X_n are distributed as $X_1 \sim LN(0, 0.5)$ and $X_2, \ldots, X_n \sim LN(1, 0.3)$. Different cases with $n = 2$, 3 and 4 random variables are investigated. Measurements $M_i = m_i$ are available for all basic random variables; associated to these measurement are multiplicative measurement errors $\epsilon_i \sim LN(0, 0.71)$. For independent random variables X_i it is possible to determine posterior distributions $f_{X_i|M_i}(x_i|m_i)$ analytically. The posterior failure probabilities $P_{F|M}$, which are used as reference solutions, are calculated through importance sampling with 10^7 samples. The results are presented in Tab. 6.4.
The results in Tab. 6.4 do not differ substantially from Tabs. 6.2 and 6.3. This indicates that the (weak) non-linearity of the LSF function describing failure does not affect the accuracy significantly.

Table 6.4: Evaluation of the discretization error for different measurement outcomes m. The problems have $n = 2$,3 or 4 random variables; a is the constant in the LSF, Eq. 6.10; $\rho_i j$ is the correlation coefficient between X_i and X_j for all $i \neq j$; P_F and $P_{F|\mathbf{M}}$ denote the prior respectively posterior failure probabilities, which are calculated through importance sampling with 10^7 samples; $\hat{P}_{F|\mathbf{M}}$ is the conditional failure probability calculated with the discrete BN. Since there is no analytical solution, the updating of the basic random variables was performed through rejection sampling with more than 5e7 accepted samples.

a	c_{up}	ρ_{ij}	P_F	$\mathbf{m}$	$P_{F\|\mathbf{M}}$	$\hat{P}_{F\|\mathbf{M}}$	Abs. error	Rel. error [%]
$n = 2$:								
12	10^2	0	1.3e-5	$[2.8, 4.5]$	1.4e-5	1.2e-5	2e-6	15
12	10^2	0	1.3e-5	$[2.3, 2.4]$	3.3e-6	3.5e-6	2e-7	6
10	12^2	0	1.7e-4	$[4.0, 3.2]$	4.0e-4	3.7e-4	3e-5	7
12	10^2	0.5	1.7e-4	$[2.3, 2.4]$	4.8e-5	5.0e-5	2e-6	4
$n = 3$:								
15	10^3	0	3.7e-5	$[2.1, 5.6, 5.0]$	4.8e-5	4.5e-5	4e-6	7
15	10^3	0	3.7e-5	$[1.1, 3.7, 3.4]$	1.5e-5	1.8e-5	3e-6	20
13	12^3	0	5.0e-4	$[3.0, 3.0, 3.0]$	5.4e-4	5.4e-5	3e-6	1
16	10^3	0.5	9.1e-4	$[3.0, 6.0, 5.0]$	1.8e-3	1.9e-3	3e-5	2
$n = 4$:								
20	8^4	0	7.4e-6	$[2.0, 4.0, 3.4, 3.0]$	4.9e-6	5.6e-6	6e-7	13
17	8^4	0	3.1e-4	$[1.0, 1.4, 1.2, 2.0]$	4.5e-5	5.2e-5	7e-6	15
17	12^4	0	3.1e-4	$[3.1, 2.0, 3.3, 2.4]$	2.5e-4	2.5e-4	7e-6	3
24	8^4	0.5	3.7e-4	$[1.0, 1.4, 1.2, 2.0]$	1.1e-5	1.1e-5	9e-7	8

6.4 Discussion

When modeling with BNs, it is often necessary or beneficial to discretize continuous random variables. When the BN includes rare events that are a function of such random variables, the choice of the discretization scheme is non-trivial. In this chapter, discretization based on FORM concepts is investigated, and a heuristic procedure for an efficient discretization in these cases is proposed. This is based on importance measures α_i obtained through a FORM analysis, which represent the influence of the uncertainty associated with a random variable X_i. The most important finding is that discretization should focus on the area around the most likely failure point (design point), identified by a FORM analysis. Furthermore, is is found that optimally all random variables should be discretized with approximately equal numbers of intervals, independent of their importance, as long as $|\alpha_i|$ is not close to zero. The widths of the intervals should be selected based on the FORM importance α_i of the random variables. With increasing importance, the interval width should be reduced, leading to finer discretization for larger $|\alpha_i|$. This relation is particularly evident for $|\alpha_i| \geq 0.8$.

It is shown that it is possible to fit a parametric function to approximate the relation between $|\alpha_i|$ and the optimal width of the region on which the discretization should focus. This parametric function is used to derive a heuristic procedure for finding an efficient discretization. This allows the extrapolation of the optimization results to problems with more random variables. As demonstrated by the verification examples, the heuristic procedure leads to accurate results.

This paper is restricted to static discretization. Application of the proposed procedure within dynamic discretization (e.g. Neil et al., 2008) should be investigated. The results of the procedure can serve as an initial discretization scheme, which is iteratively adjusted within dynamic discretization. This might strongly enhance the convergence performance of these algorithms. The gain in computational efficiency resulting from the proposed procedure over alternative static discretization approaches is problem specific. Some insights can be gained from Fig. 6.7. A discretization with intervals of equal width centered at the origin would require between approximately 2 to 5 times more intervals per random variable to achieve the same accuracy. Because of the exponential increase of computational effort with number of random variables, this leads to a considerable increase in efficiency. The gain compared to discretization with equal-frequency intervals is expected to be even higher, since equal frequency intervals focuses the fine intervals on the region of high probability density rather than on the tails of the distribution.

The number of basic random variables in a single LSF that can be modeled explicitly in a BN is limited to around 5 to 8. This is due to the exponential growth of the target nodes CPT with increasing number of parents and is independent of the discretization method. Despite this limitation, BNs are applicable to many practical problems – particularly if one considers that usually not all basic random variables need to be modeled explicitly as nodes, as demonstrated in the presented example. While in this chapter the focus was on the discretization of the basic

random variables, it is straightforward to incorporate the BNs discussed into larger models. Exemplarily in chapter 8 the procedure proposed here will be applied to the development of a prototype of a warning system for RWO.

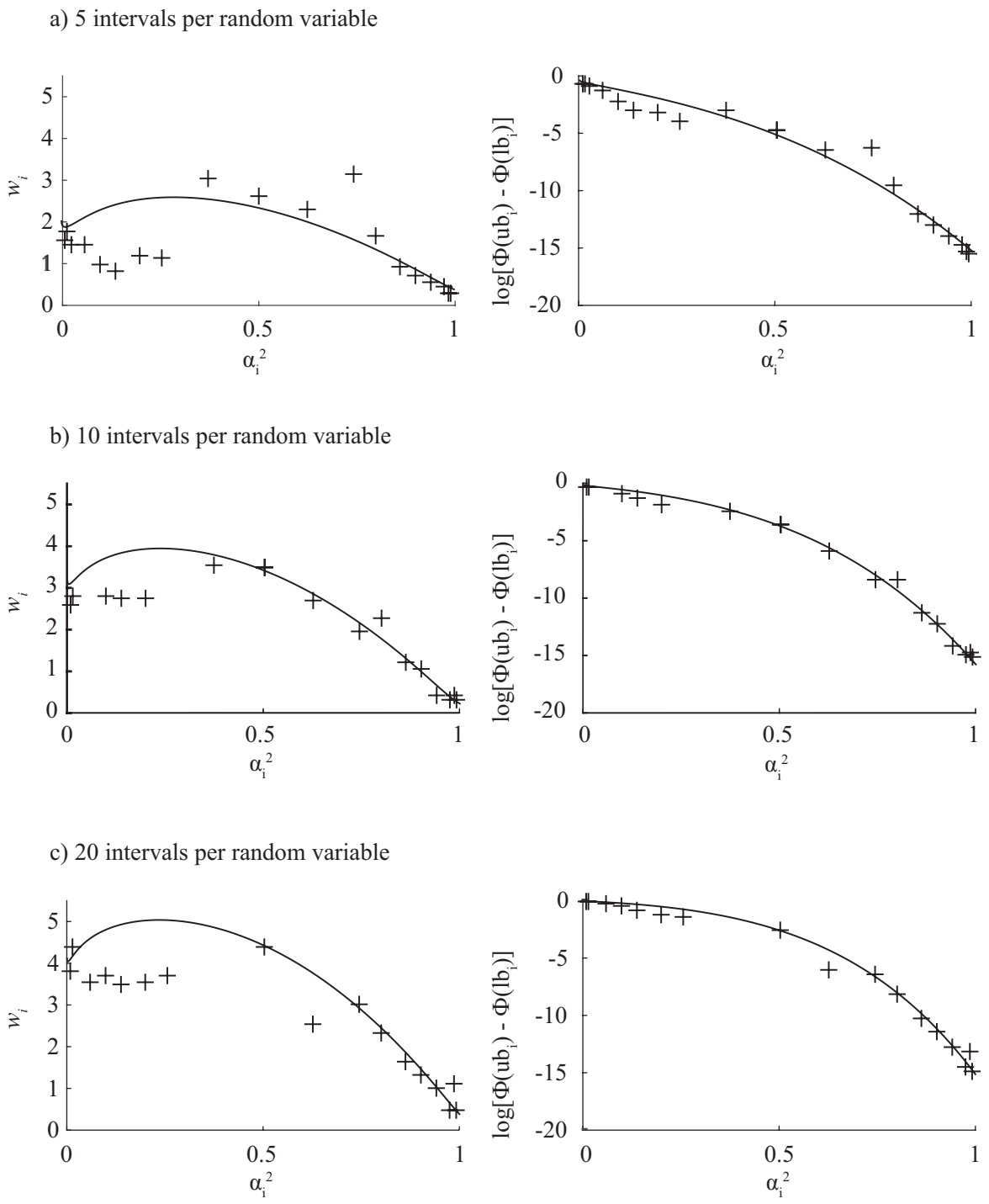


Figure 6.9: Optimization results for 10 two-dimensional, linear problems in standard normal space, which are discretized with 5, 10 and 20 intervals per dimension. In all cases the prior failure probability is 10^{-7} ($\beta = 5.2$). The crosses represent the optimization results. The solid lines are the fitted parametric functions (Eq. 6.8). The left-hand side shows the relation between the width of a discretization frame w_i and α_i^2 and the right-hand side shows the relation between the probability mass enclosed by the discretization frame with width w_i and α_i^2.

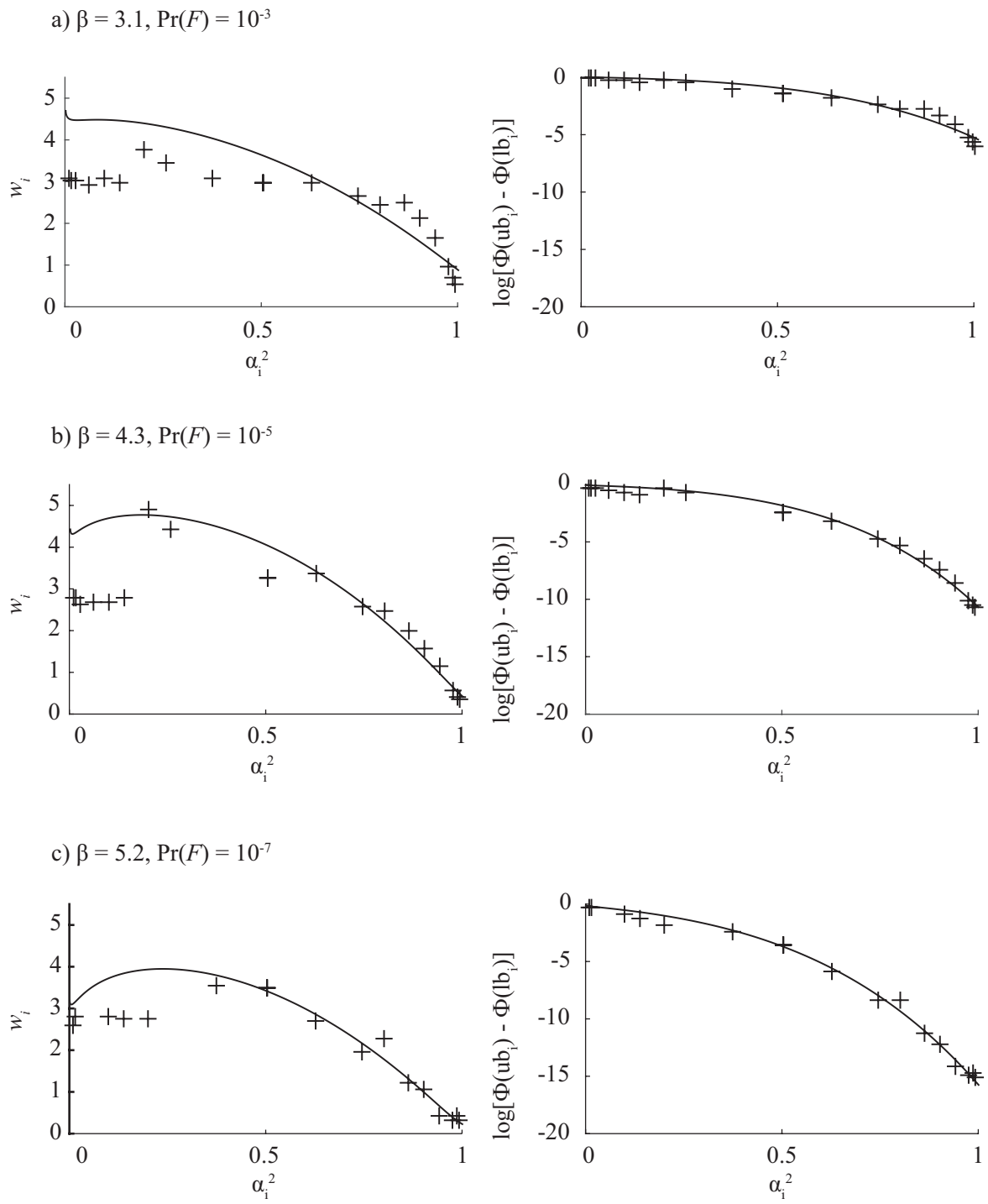


Figure 6.10: Optimization results for 10 two-dimensional, linear problems in standard normal space, which are discretized with 10 intervals per dimension. The prior failure probabilities are 10^{-3} ($\beta = 3.1$), 10^{-5} ($\beta = 4.3$) and 10^{-7} ($\beta = 5.1$). The crosses represent the optimization results. The solid lines are the fitted parametric functions (Eq. 6.8). The left-hand side shows the relation between the width of a discretization w_i and α_i^2 and the right-hand side shows the relation between the probability mass enclosed by the discretization frame with width w_i and α_i^2.

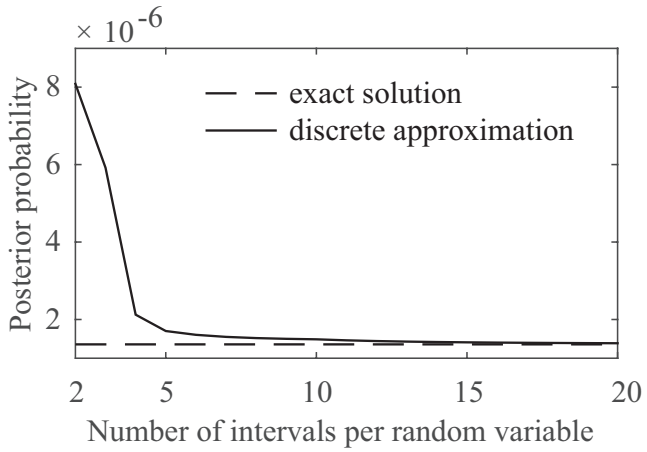


Figure 6.11: Posterior probability $\hat{P}_{F|\mathbf{M}}$ as a function of the number of intervals per random variable together with the exact (analytical) solution $P_{F|\mathbf{M}}$ for the fourth measurement case $[1.6, 2.0, 1.2]$ in Tab. 6.2.

Chapter 7

Sampling based inference

As discussed in chapter 5, exact inference is not possible for general hybrid BNs. But also for large discrete BNs, the computations required for exact inference can become unfeasible. In these cases, approximate inference through sampling is a worthwhile alternative. However, depending on the problem sampling based inference may be much slower than exact inference. Several approaches to BN sampling exist. In the following, forward sampling, likelihood weighting and inference through MCMC sampling are briefly discussed. For a in-depth description the interested reader is referred to the standard textbooks (e.g. Jensen and Nielsen, 2007; Koller and Friedman, 2009).

We assume a joint distribution $f_{\mathbf{X}}(\mathbf{x})$ with random variables $\mathbf{X} = [X_1, \ldots, X_n]$ that is represented through a BN. For the prior case – where no observations on any of the nodes X_i in the network is available, generating samples from $f_{\mathbf{X}}(\mathbf{x})$ is straightforward. More challenging is the task of sampling from the posterior distribution $f_{\mathbf{X}^{-\mathbf{e}}|\mathbf{X}^{\mathbf{e}}}(\mathbf{x}^{-\mathbf{e}}|\mathbf{x}^{\mathbf{e}})$ i.e. the distribution of $\mathbf{X}$ if information on variables $\mathbf{X}^{\mathbf{e}} \subseteq \mathbf{X}$ is available. $\mathbf{X}^{-\mathbf{e}}$ denotes the variables that have not received evidence i.e. $\mathbf{X}^{-\mathbf{e}} = \mathbf{X} \backslash \mathbf{X}^{\mathbf{e}}$. For the posterior PDF one can write:

$$f_{\mathbf{X}^{-\mathbf{e}}|\mathbf{X}^{\mathbf{e}}}\left(\mathbf{x}^{-\mathbf{e}}|\mathbf{x}^{\mathbf{e}}\right) \propto f_{\mathbf{X}^{-\mathbf{e}}}\left(\mathbf{x}^{-\mathbf{e}}\right) \cdot f_{\mathbf{X}^{\mathbf{e}}|\mathbf{X}^{-\mathbf{e}}}\left(\mathbf{x}^{\mathbf{e}}|\mathbf{x}^{-\mathbf{e}}\right) \tag{7.1}$$

7.1 Forward sampling, rejection sampling and likelihood weighting

From a prior distribution $f_{\mathbf{X}}(\mathbf{x})$ that is represented through a BN, samples can readily be generated through forward sampling (Henrion, 1988). The generated samples are i.i.d. (independent and identically distributed). If the BN is used to estimate the probability of an event E, $\Pr(E)$, one can determine the number of i.i.d. samples required to get an estimate $\hat{\Pr}(E)$ with a coefficient of variation smaller than ρ as:

$$n_s \geq \frac{1 - \Pr(E)}{\rho^2 \cdot \Pr(E)} \tag{7.2}$$

Forward sampling in a BN is performed one random variable at a time. The random variables $\mathbf{X} = [X_1, \ldots, X_n]$ in the BN should follow a topological ordering, meaning that for any two nodes X_i and X_j, where X_i is a parent of X_j, X_i should occur in the ordering before X_j. Following the topological ordering, for each node X_i a sample x_i is generated from $f_{X_i|pa(X_i)}\left(x_i|pa\left(x_i\right)\right)$, its distribution conditional on its parents. The procedure for generating n_S samples from $f_{\mathbf{X}}\left(\mathbf{x}\right)$ is summarized in algorithm 1.

Algorithm 1 Forward sampling

$\mathbf{X} = [X_1, \ldots, X_n]$ follows a topological ordering
for $j = 1$ to n_S **do**
  for $i = 1$ to n **do**
    generate a sample x_i^j from $f_{X_i|pa(X_i)}\left(x_i|\mathbf{x}_{pa(X_i)}^j\right)$, where $\mathbf{x}_{pa(X_i)}^j$ is the j-th sample of $pa\left(X_i\right)$
  end for
end for

Algorithm 1 is applicable as long as non of the variables have received any evidence. A straightforward extension to this algorithm, in order to be able to sample from posterior distributions, is to add a rejection step. Again, samples are generated from $f_{\mathbf{X}}\left(\mathbf{x}\right)$. However, a sample is only accepted if the sample that has been generated for the random variables $\mathbf{X}^e$ complies with $\mathbf{x^e}$. The approach is summarized in algorithm 2 .

Algorithm 2 Rejection sampling

$\mathbf{X} = [X_1, \ldots, X_n]$ follows a topological ordering;
set $j = 1$;
while $j \leq n_S$ **do**
  for $i = 1$ to n **do**
    generate a candidate sample x_i^c from $f_{X_i|pa(X_i)}\left(x_i|\mathbf{x}_{pa(X_i)}^c\right)$, where $\mathbf{x}_{pa(X_i)}^c$ is the current candidate sample of $pa\left(X_i\right)$
  end for
  if $\mathbf{x}^j$ complies with $\mathbf{x^e}$ **then**
    $\mathbf{x}^j = \mathbf{x}^{cand}$ (accept the sample)
    $j = j + 1$
  end if
end while

The efficiency of this procedure depends on the probability of observing $\mathbf{x^e}$ in a sample, drawn from $f_{\mathbf{X}}\left(\mathbf{x}\right)$. In fact, if at least one of the continuous nodes in the BN has received hard evidence, this probability is 0 and therefore this approach will not be directly applicable. In that case approximate results can be obtained by introducing virtual intervals around the hard evidence on continuous nodes. Instead of accepting a generated sample if it complies exactly with the given evidence one can then accept a sample if it falls within the predefined virtual interval around the evidence. This approach is referred to as approximate Bayesian computation (ABC) (Robert, 2016).

In general, rejection sampling is inefficient since (possibly most of the) samples are discarded. An approach that overcomes this issue is likelihood weighting (Shachter and Peot, 1990; Fung and Chang, 1989). Again samples can be generated from the BN in a similar way except that each element of a sample that corresponds to a random variable in $\mathbf{X^e}$ is set to it's respective observed state $\mathbf{x^e}$. To compensate for the fact that these samples do not follow the target distribution $f_{\mathbf{X^{-e}}|\mathbf{X^e}}\left(\mathbf{x^{-e}}|\mathbf{x^e}\right)$ a weight $w\left(\mathbf{x^e}, \mathbf{x}^j\right)$ is given to each sample $\mathbf{x}^j$. The weight is calculated as:

$$w\left(\mathbf{x^e}, \mathbf{x}^j\right) = \prod_{X_i \in \mathbf{X^e}} p\left(x_i^e | pa\left(X_i\right) = \mathbf{x}^j_{pa(X_i))}\right) \tag{7.3}$$

While likelihood weighting is an improvement compared to rejection sampling, it is known to perform poorly when there are large fluctuations in the likelihood weights $w\left(\mathbf{e}, \mathbf{x}^i\right)$ (Fung and Chang, 1989), which is often the case when the evidence is unlikely (Shachter and Peot, 1990).

7.2 MCMC sampling

In likelihood weighting, samples are not generated directly from the distribution of interest. This is accounted for by weighting the samples. Differently, Markov chain Monte Carlo (MCMC) methods aim at generating samples directly from the target distribution. MCMC sampling represents a popular approach for sampling from the posterior distribution $f_{\mathbf{X^{-e}}|\mathbf{X^e}}\left(\mathbf{x^{-e}}|\mathbf{x^e}\right)$ also in the context of BNs (Gilks et al., 1996). MCMC has already been briefly described in this thesis in the context of SuS (section 2.2.1). In this section MCMC is introduced in greater detail with a focus on inference in BNs.
A Markov Chain is generated from a random process that is discrete in the time domain, $\left[\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^t, \ldots, \mathbf{X}^n\right]$. This random process has the Markov property i.e. each new state, $\mathbf{X}^t = \mathbf{x}^t$, of the chain can be generated from a transition distribution that is conditional only on the previous state, $\mathbf{X}^{t-1} = \mathbf{x}^{t-1}$:

$$p\left(\mathbf{x}^t | \mathbf{x}^{t-1}, \ldots, \mathbf{x}^1\right) = p\left(\mathbf{x}^t | \mathbf{x}^{t-1}\right) \tag{7.4}$$

The transition distribution is stationary if the transition probability $p\left(\mathbf{x}^{t+k} | \mathbf{x}^{t+k-1}\right)$ does not depend on k.
A MCMC method is required to have the distribution of interest (i.e. the posterior $f_{\mathbf{X^{-e}}|\mathbf{X^e}}\left(\mathbf{x^{-e}}|\mathbf{x^e}\right)$) as its stationary distributions. In order to converge to a stationary distribution, a homogeneous chain needs to fulfill two properties. These are (Gilks et al., 1996):

- Irreducibility, i.e. any region of the outcome space, for which $f_{\mathbf{X^{-e}}|\mathbf{X^e}}\left(\mathbf{x^{-e}}|\mathbf{x^e}\right) > 0$ can be reached by a chain, with probability larger than zero.
- All states of the Markov chain are positiv-recurrence, i.e. the probability of a chain returning to a state in a finite number of steps is 1.

7.2.1 The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm generates a candidate, $\mathbf{x}^c$, for the next state of the Markov chain, $\mathbf{x}^t$, from a proposal distribution $q\left(\mathbf{x}^c|\mathbf{x}^{t-1}\right)$. The candidate state is accepted as the new state with probability:

$$\alpha\left(\mathbf{x}^c, \mathbf{x}^{t-1}\right) = \min\left(1, \frac{f_{\mathbf{X}^{-\mathbf{e}}|\mathbf{X}^{\mathbf{e}}}\left(\mathbf{x}^c|\mathbf{x}^{\mathbf{e}}\right) \cdot q\left(\mathbf{x}^{t-1}|\mathbf{x}^c\right)}{f_{\mathbf{X}^{-\mathbf{e}}|\mathbf{X}^{\mathbf{e}}}\left(\mathbf{x}^{t-1}|\mathbf{x}^{\mathbf{e}}\right) \cdot q\left(\mathbf{x}^c|\mathbf{x}^{t-1}\right)}\right) \tag{7.5}$$

If $\mathbf{x}^c$ is not accepted as the next state, the next state $\mathbf{x}^t$ is equal to the current state $\mathbf{x}^{t-1}$. This algorithm was proposed by (Metropolis et al., 1953) for symmetric proposal distributions. In that case $q\left(\mathbf{x}^c|\mathbf{x}^{t-1}\right) = q\left(\mathbf{x}^{t-1}|\mathbf{x}^c\right)$ and thus the proposal distribution cancels out in Eq. 7.5. The algorithm was later generalized for non-symmetric proposal distributions by (Hastings, 1970). Markov chains generated through the Metropolis-Hastings algorithm fulfill the reversibility condition from Eq. 7.6 (e.g. Green and Hastie, 2009; Brooks et al., 2011; Papaioannou et al., 2015) and thus they have $f_{\mathbf{X}^{-\mathbf{e}}|\mathbf{X}^{\mathbf{e}}}\left(\mathbf{x}^{-\mathbf{e}}|\mathbf{x}^{\mathbf{e}}\right)$ as their stationary distributions. The reversibility condition reads:

$$p\left(\mathbf{x}^t|\mathbf{x}^{t-1}\right) f_{\mathbf{X}^{-\mathbf{e}}|\mathbf{X}^{\mathbf{e}}}\left(\mathbf{x}^{t-1}|\mathbf{x}^{\mathbf{e}}\right) = p\left(\mathbf{x}^{t-1}|\mathbf{x}^t\right) f_{\mathbf{X}^{-\mathbf{e}}}\left(\mathbf{x}^t|\mathbf{x}^{\mathbf{e}}\right) \tag{7.6}$$

where $p\left(\mathbf{x}^t|\mathbf{x}^{t-1}\right)$ is the transition PDF that describes the transition from $\mathbf{x}^{t-1}$ to $\mathbf{x}^t$. This reversibility condition ensures that $f_{\mathbf{X}^{-\mathbf{e}}|\mathbf{X}^{\mathbf{e}}}\left(\mathbf{x}^{-\mathbf{e}}|\mathbf{x}^{\mathbf{e}}\right)$ is the stationary distribution for an irreducible Markov chain.

7.2.2 Gibbs sampling

Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990) is a MCMC method that is particularly suited to sampling from a posterior $f_{\mathbf{X}^{-\mathbf{e}}|\mathbf{X}^{\mathbf{e}}}\left(\mathbf{x}^{-\mathbf{e}}|\mathbf{x}^{\mathbf{e}}\right)$ if the joint distribution of $\mathbf{X}$ is defined through a BN. In the context of Gibbs sampling, a new state of a component X_i is generated directly from the distribution of X_i conditional on the current state of all other random variables. The Gibbs sampler can be considered a special case of the Metropolis-Hastings algorithm, where each new sample is accepted with probability 1 (Gilks et al., 1994). Gibbs sampling for BNs is readily available through the well-known open source software OpenBUGS (Bayesian inference Using Gibbs Sampling) as well as a number of other software tools (Lunn et al., 2009). In algorithm 3 an approach for generating a Markov chain of length n_C by Gibbs sampling is described.

Where x_i^e denotes the hard evidence corresponding to the random variable X_i. A way to initiate $\mathbf{x}^1 = \left[x_1^1, \ldots, x_n^1\right]$ is by forward sampling. If the initial sample does not follow the posterior distribution burn-in needs to be considered. The Gibbs sampler, as it is described here does not fullfil the reversibility condition from Eq. 7.6.

The independence assumptions of the BN simplify Gibbs sampling due to the fact that a random variable X_i is d-separated from the rest of the network, given it's Markov blanket i.e.

Algorithm 3 Gibbs sampling.

initiate $\mathbf{x}^1 = \left[x_1^1, \ldots, x_n^1\right]$
for $j = 2$ to n_c **do**
  for $i = 1$ to n **do**
    if $X_i \in \mathbf{X^e}$ **then**
      $x_i^j = x_i^e$
    else
      sample x_i^j from $f_{X_i|X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n}\left(x_i^j|x_1^j,\ldots,x_{i-1}^j,x_{i+1}^{j-1},\ldots,x_n^{j-1}\right)$
    end if
  end for
end for

$f_{X_i|X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_m}\left(x_i^j|x_1^j,\ldots,x_{i-1}^j,x_{i+1}^{j-1},\ldots,x_m^{j-1}\right) = f_{X_i|MB(X_i)}\left(x_i^j|MB\left(x_i\right)\right)$. Thus one can sample a new state of X_i from $f_{X_i|MB(X_i)}\left(x_i^j|MB\left(x_i\right)\right)$. For discrete BNs, deriving an analytical expression for this distribution and sampling from it is straightforward. For continuous nodes, the distribution can in general not be derived analytically, however, efficient methods for sampling from this univariate conditional distribution exist (e.g. Gilks and Wild, 1992; Gilks et al., 1995; Martino et al., 2015).

7.2.3 Challenges of MCMC

Some aspects regarding convergence

The main challenge in the context of MCMC is to ensure convergence. This includes two aspects: (1) Does a Markov chain (theoretically) converge to the desired stationary distribution (2) How long does convergence take, i.e. at what point in time can one assume that a sample is generated from a distribution, which is sufficiently close to the desired stationary distribution?

Properties that need to be fulfilled in order for Markov chains to converge to the desired posterior are given in section 7.2. As an example of a BN for which Gibbs sampling does not converge consider the network structure in Fig. 7.1 with the CPTs from Tabs. 7.1 – 7.3. The BN could represent a series system with components X_1, X_2 and the random variable X_3 representing the system state. Starting from an initial state $\left[x_1^1, x_2^1, x_3^1\right]$ the states of the random variable X_3 chain will not change within the Markov chain. The chain would fulfill the convergence properties, if all entries in the CPTs were larger than 0. In fact it can be shown that the Gibbs sampler fulfills the properties required for convergence for any discrete BN, if all entries of its CPTs are larger than 0 (Koller and Friedman, 2009).

However, if all zeros were replaced by small probabilities, the Gibbs sampler would still take a large number of steps to converge. In general it is not possible to determine if a chain has converged. Nevertheless, a number of diagnostic heuristics to check whether a chain has converged have been proposed to this end (Cowles and Carlin, 1996).

Typically the starting point of a Markov chain is chosen arbitrarily and is thus not distributed

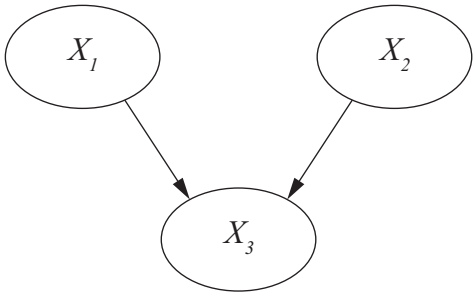


Figure 7.1: A simple BN structure.

Table 7.1: CPT of the node X_1 in Fig. 7.1.

$\Pr(X_1)$	
$X_1 = 0$	0.1
$X_1 = 1$	0.9

according to the desired distribution. To reduce the influence of an arbitrary starting point, it is common to discard the first samples of a chain. This part of a Markov chain is known as its burn-in phase. After the burn-in phase, the samples of the Markov chain should correspond to a distribution, which is sufficiently close to the desired distribution. Unfortunately, it is in general not possible to determine the length of the burn-in phase.

Computational cost

Unlike samples from traditional MCS, samples generated through MCMC are dependent. Therefore, the number of MCMC samples required to get an estimate of some statistics of the desired distribution is higher than for an equally good estimate through standard MCS. In fact, the higher the correlation between samples, the more samples are required.
Furthermore, the computational cost for generating one sample should be taken into account, when evaluating the efficiency of a MCMC method. This is especially the case if a node is defined through a computationally costly numerical model, $g(\mathbf{x})$. For example, the M-H algorithm as described in subsection 7.2.1 requires one evaluation of $g(\mathbf{x})$ per sample. In contrast the Gibbs sampler from algorithm 3, where each component is sampled individually, requires at least one evaluation of $g(\mathbf{x})$ for each component that has the $g(\mathbf{x})$-node in its Markov blanket. However, depending on the method used for sampling from the univariate conditional distributions, the number of evaluations of $g(\mathbf{x})$ per component sampled might be significantly higher.

Table 7.2: CPT of the node X_2 in Fig. 7.1.

$\Pr(X_2)$	
$X_2 = 0$	0.2
$X_2 = 1$	0.8

Table 7.3: CPT of the node X_3 in Fig. 7.1.

$\Pr(X_3)$	$X_1 = 0$		$X_1 = 1$	
	$X_2 = 0$	$X_2 = 1$	$X_2 = 0$	$X_2 = 1$
$X_3 = 0$	1	1	1	0
$X_3 = 1$	0	0	0	1

7.3 BN sampling for rare events

Motivated by the need for estimating probabilities of rare (failure) events based on numerically costly models, efforts have been made to reduce the number of samples (respectively model evaluations) required for getting a good estimate of the probabilities of these rare events. This resulted in a number of structural reliability methods, some of which are described in chapter 2 .

One representative of these methods is SuS (Au and Beck, 2001), described in subsection 2.2.1. The idea behind SuS is to decompose the probability of a rare event $\Pr(F)$ into a product of m larger (conditional) probabilities, $\Pr(F_i|F_{i-1})$, as (c.f. Eq. 2.11):

$$\Pr(F) = \Pr(F_1) \prod_{i=2}^{m} \Pr(F_i|F_{i-1}) \tag{7.7}$$

where the intermediate events F_i are defined as $F_i = \{g(\mathbf{x}) \leq a_i\}$, with a_i being a positive constant. Furthermore $F_n = F = \{g(\mathbf{x}) \leq 0\}$. Typically in SuS these probabilities are estimated through standard MCS (in the case of $Pr(F_1)$) or through MCMC (in the case of $\Pr(F_i|F_{i-1})$). Furthermore the intermediate failure events can be chosen, such that $Pr(F_1) = \Pr(F_i|F_{i-1}) = \ldots = \Pr(F_{n-1}|F_{n-2}) = p_0$. In that case the probability of failure can be rewritten as in Eq. 2.12.

7.3.1 SuS in the context of BNs

We consider BNs for reliability problems, as shown e.g. in Fig. 6.1 from chapter 6. Such a BN might also be included in larger networks. In this BN, the performance of an engineering system is described through a physical model $g(\mathbf{x})$ that is represented through the binary node

Component performance, with states failure (F: $\{g(\mathbf{x}) \leq 0\}$) or no failure ($\overline{F}$:$\{g(\mathbf{x}) > 0\}$). Fig. 7.2 shows an extended version of the BN in Fig. 6.1, which directly includes $g(\mathbf{x})$ as a node.

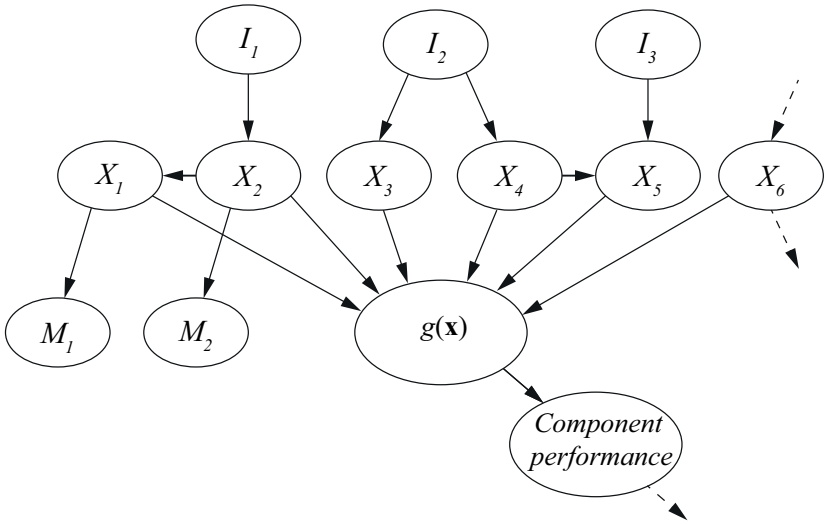


Figure 7.2: The general reliability BN, extended by explicitly including the model $g(\mathbf{x})$ as a node.

The probability of the node *Component performance* being in state failure (F), conditional on having evidence, $[\mathbf{X}^e = \mathbf{x}^e, \mathbf{I}^e = \mathbf{i}^e, \mathbf{M}^e = \mathbf{m}^e]$ on some of the nodes, can be computed efficiently by applying SuS to the BN. The procedure is described in algorithm 4.
The algorithm will not only output $\Pr(F)$ but also a number of samples from the nodes $[\mathbf{X}, \mathbf{M}, \mathbf{I}, g(\mathbf{x})]$ that are conditional on both $[\mathbf{X}^e = \mathbf{x}^e, \mathbf{I}^e = \mathbf{i}^e, \mathbf{M}^e = \mathbf{m}^e]$ and $g(\mathbf{x}) \leq 0$. From these samples, samples within the failure domain (F) can be generated by Gibbs sampling.

7.3.2 A simple verification example in OpenBUGS

For verification, the approach is applied to a simple example defined through the LSF:

$$g(\mathbf{X}) = a - \sum_{i=1}^{n} X_i \tag{7.8}$$

where the n random variables X_i are assumed to be independent (thus in the corresponding BN there are no links or common parents between the X_i's) and standard normal distributed and the constant a is chosen corresponding to different probabilities of failure, $\Pr(F)$, from $1.5 \cdot 10^{-1}$ to $1.5 \cdot 10^{-5}$ as:

$$a = -F^{-1}(\Pr(F)) \tag{7.9}$$

where F^{-1} is the inverse normal CDF with mean $\mu = 0$ and standard deviation $\sigma = \sqrt{n}$. For this verification example, algorithm 4 has been implemented in MATLAB, the MCMC

Algorithm 4 SuS in the context of BNs

p_0; n_S; $n_{seeds} = {}^1/_{p_0}$; $i_{Level} = 1$;

- Gnereate n_S samples, $[\mathbf{x}, \mathbf{i}, \mathbf{m}, g(\mathbf{x})]$, from the BN (consider burn-in if necessary)
- Select a_1 as the $(p_0 \cdot 100\%)$-th percentile of samples the node $g(\mathbf{x})$
- Select the n_{seeds} samples of $[\mathbf{x}, \mathbf{i}, \mathbf{m}, g(\mathbf{x})]$, for which $g(\mathbf{x}) \leq a_1$ as seeds

while $a_{i_{Level}} > 0$ **do**

- $i_{Level} = i_{Level} + 1$
- set $\{g(\mathbf{x}) \leq a_{i_{Level}-1}\}$ as (inequality) evidence in the BN
- draw n_S new samples, $[\mathbf{x}, \mathbf{i}, \mathbf{m}, g(\mathbf{x})]$, from the BN – to this end generate n_{seeds} Markov chains of length ${}^{n_S}/_{n_{seeds}}$, each starting from a different seed
- Select $a_{i_{Level}}$ as the $(p_0 \cdot 100\%)$-th percentile of samples of the node $g(\mathbf{x})$
- Select the n_{seeds} samples of $[\mathbf{x}, \mathbf{i}, \mathbf{m}, g(\mathbf{x})]$, for which $g(\mathbf{x}) \leq a_{i_{Level}}$ as seeds

end while

- $\hat{\Pr}(F) \approx p_0^{i_{Level}-1} \cdot \frac{1}{n_S} \cdot \sum_{\mathbf{x}_i \in \mathbf{x}^{i_{Level}}} I_{g(\mathbf{x}_i) \leq 0}(\mathbf{x}_i)$

sampling within the algorithm is performed through Gibbs sampling in OpenBUGS. The MATLAB based software Matbugs (Murphy and Mahdaviani, 2005) is used as an interface between MATLAB and OpenBUGS. For each SuS-level, $n_S = 7000$ samples are used and the intermediate thresholds a_i are chosen, such that $p_0 = 0.1$. In Fig. 7.3 the relative error for the example (averaged over 14 simulation runs) is shown for $n = 2$ and $n = 30$ random variables X_i. The corresponding estimate of the coefficient of variation is shown in Fig. 7.4.

7.3.3 Discussion of the sampling approach for rare events

The results in Figs. 7.3 and 7.4 show a relative error and a coefficient of variation, which is reasonable for most practical applications. Ultimately, the approach should be incorporated into inference for larger BNs that include a reliability problem similar to the one discussed. In this context, convergence of the applied MCMC method should be carefully investigated. Even in the simple example presented here, the coefficient of variation is relatively high for small failure probabilities. This may be an indication for slow convergence. Both the coefficient of variation and also the relative error are (at least for small $\Pr(F)$) higher for the case with only two random variables. A possible explanation for that, is that in the two-dimensional case the random variables are more correlated than in the $n = 30$ case. And it is known that convergence of the Gibbs sampler is typically slower for models, whose random variables are highly correlated (Koller and Friedman, 2009).

An estimate of $\Pr(F) = 1.5 \cdot 10^{-5}$ was here obtained based on 35000 samples. Estimating $\Pr(F)$ through MCS requires $2 \cdot 10^6$ (respectively 10^7) to get results that are comparable (in terms of the coefficient of variation) to the $n = 2$-case (respectively the $n = 30$-case). When

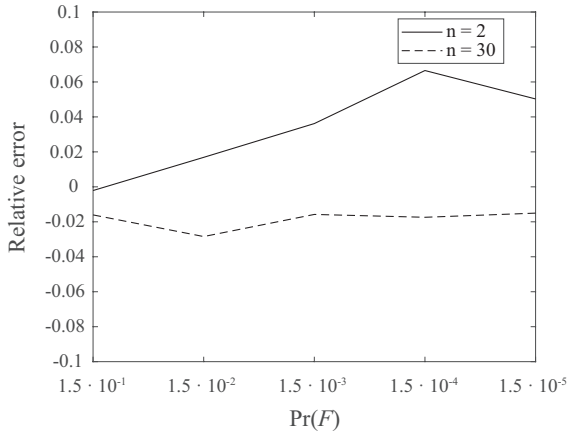


Figure 7.3: Relative error for the LSF in Eq. 7.8 averaged over 14 simulation runs (Note: The number of simulation runs is limited because of the applied software tools). The results are shown for $n = 2$ and $n = 30$ random variables X_i.

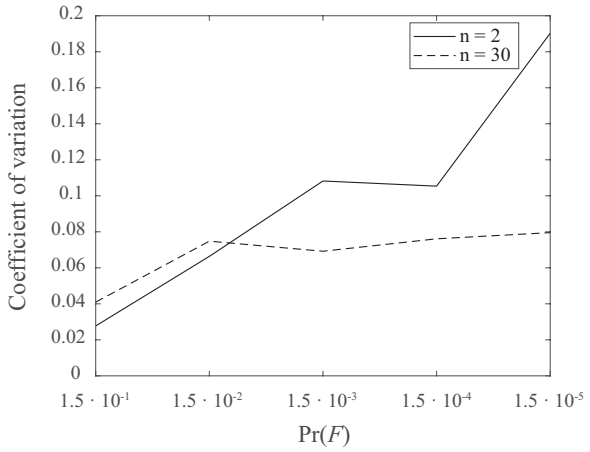


Figure 7.4: Coefficient of variation for the LSF in Eq. 7.8 estimated from 14 simulation runs . The results are shown for $n = 2$ and $n = 30$ random variables X_i.

estimating the probabilities by means of standard MCMC significantly more samples would be required, since the samples within a Markov chain are not independent.
However, if an evaluation of the LSF, $g(\mathbf{x})$, is computationally costly, more important than the number of samples is the number of LSF-evaluations. In standard MCS, the number of LSF-evaluations is equal to the number of samples. The same is true for the Metropolis-Hastings algorithm as described in subsection 7.2.1. In the case of Gibbs sampling (subsection 7.2.2) at least one LSF-evaluation is required for resampling each node that has $g(\mathbf{x})$ in its Markov blanket. Future research should thus on the one hand look deeper into convergence of the method and on the other hand into its computational efficiency.

Part III

APPLICATIONS

Chapter 8

Reliability analysis for runway overrun

[10]The airlines that are organized in the international air transport association (IATA) carried around 3.6 billion passengers in 2012. Among these 414 were killed in an aviation accident (IATA, 2013). Conserving such high safety standards necessitates quantitative risk analysis. The small accident probabilities hinder an estimation of the accident-probabilities purely based on statistical methods.

The most common accident type in the field of civil aviation is runway excursion. Several subtypes of runway excursion exist, among them RWO of a landing aircraft (Fig. 8.1) is the most critical one (in the following simply referred to as RWO). This chapter describes the development of a BN that is capable of providing RWO probabilities, conditional on the state of the aircraft and the environment, in near real-time. Continuously providing the pilots with these RWO probability estimates the BN is supposed to help pilots in their decision, wether it is save to land or wether they should attempt a second approach, either at the same airport or at an alternate airport.

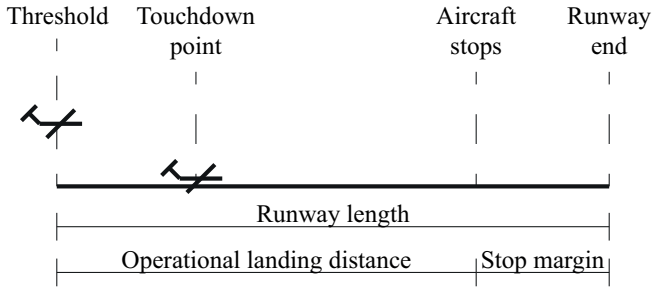


Figure 8.1: Runway overrun definitions

[10]The description of this appliction is based on (Zwirglmaier et al., 2014; Zwirglmaier and Straub, 2015, 2016a).

8.1 Physical model of RWO

The operational landing distance of an aircraft can be calculated from the approach speed and the forces acting on the aircraft. A physical model for that purpose was proposed by (Drees and Holzapfel, 2012). This model is briefly described in the following. For a more detailed description, the reader is referred to the original paper by (Drees and Holzapfel, 2012). From there the RWO event can be defined as the operational landing distance exceeding the runway length (c.f. Fig. 8.1). The LSF can therefore be written as:

$$g(\mathbf{x}) = \text{Stop margin}(\mathbf{x}) \tag{8.1}$$

where the stop margin is the deterministic runway length minus the operational landing distance, which depends on a number of parameters $\mathbf{x}$. Since these parameters are a priori uncertain, they are modeled through random variables. The model to determine the operational landing distance is derived from the equations of motion. Using these the acceleration of the aircraft in x-direction can be written as:

$$\dot{V} = \frac{1}{m}\left[T - D - mg \cdot \sin\gamma - \mu_F\left(mg \cdot \cos\gamma - L\right)\right] \tag{8.2}$$

where m is the mass of the aircraft, T is the propulsion force from the aircraft engines and D is the aerodynamic drag. The term $mg \cdot \sin\gamma$ – with g being the constant of gravitation and γ the flight path angle – represents the contribution of the runway slope to the acceleration. Finally, the last term describes the influence of the friction and brake forces. L denotes the aerodynamic lift and μ_F the friction coefficient, which depends on the runway condition, on the brake force and on the velocity of the aircraft.
The drag force, D can be written as:

$$D = \frac{\rho}{2}\left(V_K - V_W\right)^2 S \cdot C_D \tag{8.3}$$

and the lift force, L as:

$$L = \frac{\rho}{2}\left(V_K - V_W\right)^2 S \cdot C_L \tag{8.4}$$

where ρ is the air density, V_K is the speed of the aircraft and V_W the speed of the surrounding air, such that $V_K - V_W$ is the speed of the aircraft relative to the wind speed. Furthermore, S is the reference area of the wings and C_D respectively C_L are the drag and the lift coefficients. Integrating Eq. 8.2 twice with respect to time will yield the operational landing distance. For practical reasons the operational landing distance is split in three parts. First the touchdown distance, which is directly modeled as a random variable; second the distance from the touchdown point to the point where the pilot deactivates the auto-brake system, which, in addition to the environmental and technical factors, depends on a number of human factors, e.g. the point in time until the pilot starts braking and the point in time when the spoiler is deployed. The third part is the distance the aircraft travels from the point of auto-brake

deactivation to the final stopping position. During a normal landing, the aircraft will not stop completely on the runway but exit the runway with slow velocity to the taxiways. However, it is assumed that in the case of a critical landing, which is likely to result in a RWO, this distance is traveled with maximum deceleration efforts i.e. maximal braking forces.
The quantities in the differential Eq. 8.2 are influenced by environmental factors on the one hand, and technical as well as human factors on the other hand. The technical aspects are captured in the RWO model by allowing certain technical components like the spoiler, the auto-brake system or the engines to be either operative or inoperative with a certain probability. The brake system can be operative, inoperative or degraded. In this thesis technical failures are neglected, thus all components are considered to be fully operative. Furthermore, it is assumed that the flaps and slats are in configuration full, which is used in 96 % of all approaches (Drees and Holzapfel, 2012). The friction coefficient between the tire and the runway μ_F is a major factor influencing the deceleration of the aircraft. This friction coefficient is mainly determined by the runway condition; a wet runway in considered here. The auto-brake system automatically applies a brake pressure when the landing gear touches ground. The magnitude of the applied pressure depends on the setting of the auto-brake system. In the scope of this chapter, we compute RWO probabilities for the auto-brake settings set to medium. In operational aviation, the times after touchdown, at which the spoiler and thrust reversers are deployed and the times at which the braking starts and ends can be obtained from operational data. On the basis of this measured data, distribution models are selected. These models are summarized in Tab. 8.1. In this application we distinguish between two airports (I and II) and two aircrafts (A and B). While the airport influences the distribution of the head wind and the approach speed deviation, the aircraft influences the distribution of the landing weight.
The touchdown point in this model, is a factor, whose variability is influenced mainly by human actions. A normal distribution, with fixed standard deviation and a mean value (μ_{TDP}) that is a function of the environmental conditions is used for this. It was found from flight operation data that pilots change their touchdown behavior according to the required landing distance estimate, which they calculate during the approach. The required landing distance is thereby calculated from aircraft and runway specific characteristics as well as from the flap setting, the head wind, the temperature and the landing weight.

Table 8.1: Basic random variables of the problem. Note: The mean value of the random variable *Touchdown point* is a deterministic function of other basic random variables (Zwirglmaier et al., 2014).

Random variable	Comment	Distribution	Mean (μ)	Std. dev. (σ)
Landing weight [t]	Aircraft A	Weibull	59.3	1.69
Landing weight [t]	Aircraft B	Weibull	64.3	1.69
Head wind [kts]	Airport I	Normal	5.4	5.8
Head wind [kts]	Airport II	Normal	6.5	5.8
Approach speed deviation [kts]	Airport I	Gumbel	4.7	4.2
Approach speed deviation [kts]	Airport II	Gumbel	5.6	4.2
Temperature [°C]		GEV	9.4	8.0
Air pressure [hPa]		Normal	1016	8.1
Touchdown point [m]		Normal	μ_{TDP}	121.9
Time of spoiler deployment [s]		Gamma	5.0	1.7
Time of breaking initiation [s]		GEV	13.1	6.4
Time of reverser deployment [s]		GEV	4.3	1.3
Time of breaking end [s]		Normal	25	5

8.2 BN model

Modeling all basic random variables of the problem as parents of the node representing the LSF is computationally unfeasible in the discrete BN. For example if each of these variables were discretized with 10 intervals, the resulting CPT would consist of $2 \cdot 10^{10}$ entries, of which 10^{10} need to be computed e.g. through sampling. The BN is thus simplified by removing nodes that are not important for predicting RWO. Criteria, based on which one can decide, which nodes to remove are provided in chapter 6. The FORM importance measures α_i that express the importance of the uncertainty associated to the random variables X_i represent an important criterion for the decision whether to remove a node. In this application removed are random variables, whose $|\alpha_i| < 0.1$. Furthermore removed are random variables, for which no observation can be obtained up to the point in time, when the decision whether to land has to be made. The FORM importance measures and reasons for removing specific random variables are given in Tab. 8.2. Despite the fact that landing weight has an α_i around 0.1 it is not removed from the network. The reason for that is that while for a particular aircraft type the associated uncertainty is not important, the landing weight changes significantly from one of the considered aircraft types to the other.

The resulting BN of the RWO warning system, after node removal, is shown in Fig. 8.2. During the aircraft approach, measurements can be obtained for the three basic random variables included in the BN. The random variables are discretized separately for each aircraft–airport combination (joint states of discrete parents) with 8 intervals each, following the discretization procedure proposed in chapter 6. In a second step, the discretization schemes are merged, i.e. the regions of the outcome space, which are discretized with fine intervals for at

Table 8.2: FORM importance measures α_i for each airport (AP) – aircraft (AC) combination and every basic random variable in the RWO application.

Random variable	α_i				Annotation
	AP I/AC A	AP I/AC B	AP II/AC A	AP A/AC II	
Landing weight [t]	0.09	0.10	0.11	0.09	Modeled
Headwind [kts]	-0.65	-0.61	-0.67	-0.60	Modeled
Approach speed deviation [kts]	0.20	0.21	0.20	0.24	Not observable
Temperature [°C]	0.03	0.00	-0.03	-0.03	Not important
Air pressure [hPa]	0.01	-0.01	-0.01	0.00	Not important
Touchdown point [m]	0.20	0.16	0.18	0.20	Modeled
Time of spoiler deployment [s]	0.00	0.00	0.01	0.01	Not important
Time of breaking initiation [s]	0.70	0.74	0.68	0.73	Not observable
Time of reverser deployment [s]	0.03	0.04	0.06	0.05	Not important
Time of breaking end [s]	-0.02	-0.01	0.01	0.02	Not important

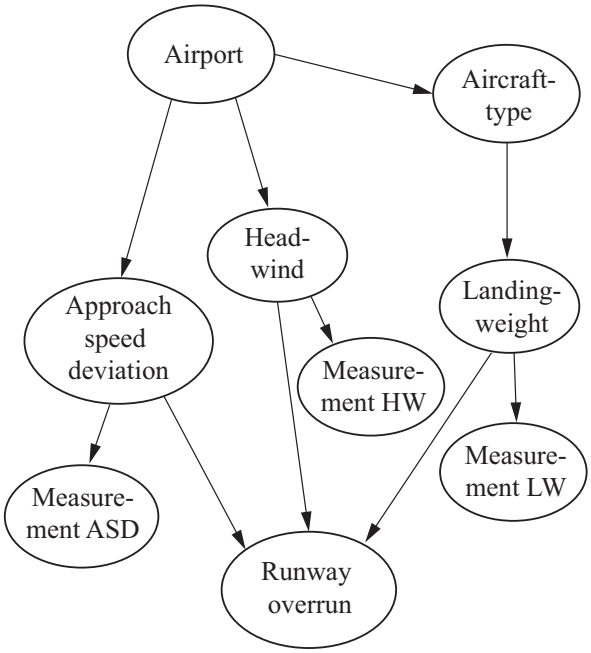


Figure 8.2: BN structure of the developed RWO warning system.

least one of the aircraft–airport combinations, are discretized with the respective fine intervals also in the merged discretization scheme. In the end 15 (landing-weight), 10 (headwind) and 9 (approach speed deviation) intervals are used to discretize the three basic random variables. For all observable quantities, the measurements m_i are modeled with an additive observation error

$$m_i = x_i + \epsilon_i \tag{8.5}$$

ϵ_i is modeled by a normal distribution with zero mean and standard deviation σ_{ϵ_i}. For the random variable landing weight (at landing time), the standard deviation of the measurement error is $\sigma_{\epsilon_{LW}} = 0.34$t. Due to turbulences governing wind speeds, the measurement of the head wind speed at the time of the measurement is only an uncertain indicator for the head wind speed at landing time; we model the measurement error with a standard deviation $\sigma_{\epsilon_{HW}} = 2.88$kts. The measurement uncertainty associated with the approach speed deviation at landing has standard deviation $\sigma_{\epsilon_{ASD}} = 4.21$kts. 49 (Measurement LW), 57 (Measurement HW) and 57 (Measurement ASD) intervals are used to discretize the measurement nodes.

8.3 Results

In Tab. 8.3, prior RWO probabilities for the different airports and aircrafts obtained with the discrete BN are compared to solutions, which were calculated by importance sampling around the design point. In Tab. 8.4, results obtained with the BN for different hypothetical cases of aircrafts approaching an airport are presented. In each of these cases, measurements associated with landing weight, headwind and the approach speed deviation are made. A threshold on the probability of RWO is used to decide, whether or not the pilot should continue landing or cancel the landing attempt. Here we assume that up to a RWO probability of 10^{-6} the pilot should continue landing.

Table 8.3: RWO probabilities, Pr(RWO), for the different airports and aircrafts calculated with the discrete BN, together with solutions calculated by importance sampling around the design point. The latter have a sampling error with coefficient of variation in the order of 10%.

Airport	Aircraft	Pr(RWO) by discrete BN	Pr(RWO) by importance sampling
I	A	$2.0 \cdot 10^{-7}$	$1.9 \cdot 10^{-7}$
I	B	$1.0 \cdot 10^{-6}$	$9.2 \cdot 10^{-7}$
II	A	$1.3 \cdot 10^{-7}$	$1.3 \cdot 10^{-7}$
II	B	$6.9 \cdot 10^{-7}$	$6.5 \cdot 10^{-7}$

Table 8.4: Probabilities of RWO and corresponding decision on landing, computed with the BN for different sets of observations.

Case	Airport	Aircraft	Measured LW [t]	M. HW [kts]	M. ASD [kts]	Pr(RWO)	Land?
a)	I	B	63	0	10.5	$2.5 \cdot 10^{-8}$	Yes
b)	I	A	61	-10	5	$4.8 \cdot 10^{-6}$	No
c)	II	B	67	3	0	$6.5 \cdot 10^{-10}$	Yes
d)	II	A	57.5	-12	3	$1.3 \cdot 10^{-6}$	No

8.4 Discussion

This application shows how a reliability problem with a physically-based performance model can be modeled through a discrete BN. The advantage of this lies in the capability of discrete BNs to rapidly update probabilities, once new information becomes available. Such a feature is especially of interest in near real-time applications. Treating continuous reliability problems in a discrete BN framework requires the discretization of the continuous outcome space of the reliability problem. This leads inevitably to a discretization error. In order to keep this error small, the heuristic developed in chapter 6 is applied. In this application the discretization procedure was done for each combination of airport and aircraft with 8 intervals per basic random variable, that is modeled explicitly as a node in the BN. In a second step these discretization schemes were merged, such that the number of free parameters of the target variable's CPT is $15 \cdot 9 \cdot 10 \cdot (2-1) = 1350$. Computing the parameters for this setting is feasible for the LSF considered in the scope of this paper. However if the number of cells, used for discretization is increased by some orders of magnitude this may lead to considerable computation cost. It is therefore necessary to decide carefully, which random variables should be modeled explicitly as nodes. For computationally more demanding LSFs it may be necessary to reduce the number of intervals per dimension and accept a larger discretization error. By applying an importance sampling approach to sample from the distributions of the implicitly modeled random variables, one can reduce the number of samples required to populate the CPT of the target variable to some extend. Furthermore while a computational demanding LSF may cause a large computational effort in the process of establishing the model, it does not have any effect on the computational effort in the application of the model. Finally it should be noted that while the final BN developed in this chapter is quite simple, such BNs can be incorporated into more complex BN models, fully exploiting the advantages of the modeling framework.

Chapter 9

Elicitation of BNs for HRA

[11]This chapter considers elicitation of BNs for HRA. A framework is proposed that derives the BN structure directly based on the relevant literature from the field of cognitive psychology. The local conditional dependencies are derived based on a combination of expert estimates and data from simulator tests. Exemplarily the CFM critical data misperceived from the IDHEAS HRA method is considered to demonstrate the proposed framework. This CFM is presented to some detail in the following. Critical data misperceived captures situations such as the one in which a parameter has to be read from a control panel or the status of some piece of equipment is to be determined from an indication on the control panel and this piece of information is critical in the sense that its misinterpretation will lead to an incorrect response (Xing et al., 2013). Three PIFs are used to describe the context: HSI/environment, workload, and training[12] , where HSI refers to Human-system interface. All the PIFs are binary with states labeled as poor and good, high and low or no and yes. In Fig. 9.1 the DT for the CFM critical data misperceived is shown. Each path through the DT represents one possible crew failure scenario. The analysts are provided with a set of two to five questions[13] for each PIF guiding them in determining the states of the PIFs. Expert elicitation was used to assign probabilities to the different crew failure scenarios.

9.1 Development of a BN structure for the CFM

As explained in chapter 3 the DAG of a BN ideally represents the causal relationships between the random variables in the model. Furthermore, the structure also defines the information (i.e., the marginal and conditional probabilities) needed to quantify the BN. In this section, the development of two BN structures for each CFM is illustrated: a first BN that contains

[11]This chapter is adapted from (Zwirglmaier et al., 2017).

[12]This DT also contains a branch for recovery potential, which is used in most IDHEAS CFM. The meaning of "recovery potential" has been defined in a generic manner as "opportunities for correction given failure". However these opportunities have not been clearly specified for the considered CFM, and thus this concept is neglected in the remaining sections of the chapter.

[13]These questions are not explicitly included in the DT. In section 9.1 a way to directly include these questions in the model is proposed.

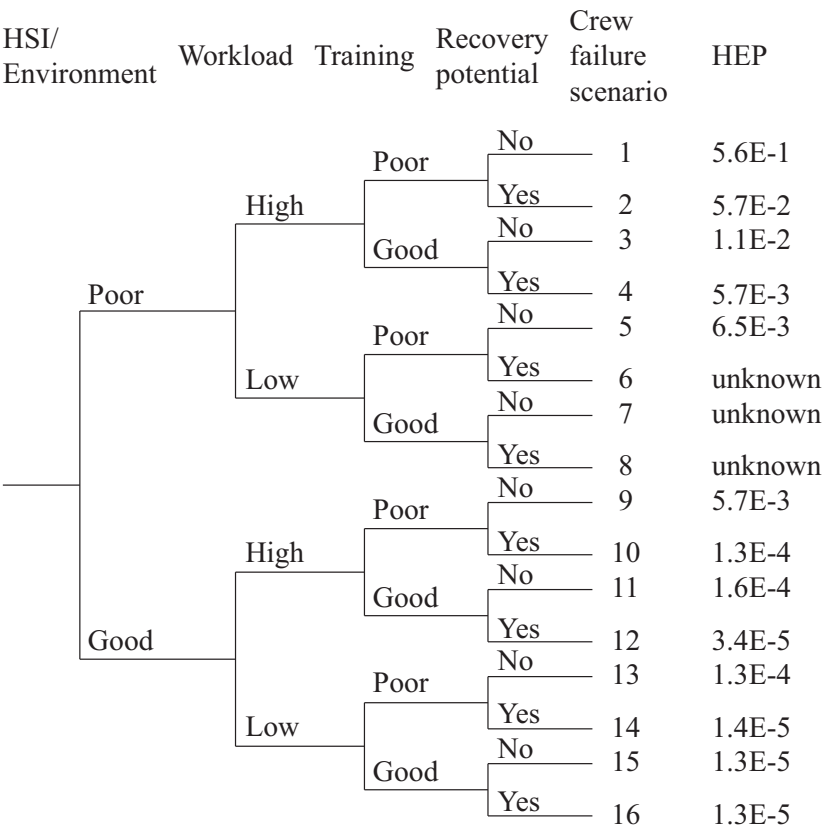


Figure 9.1: Decision tree for the crew failure mode critical data misperceived (Xing et al., 2013). The paths through the decision tree are numbered and for each path a probability was elicited from experts. E.g. the HEP for poor HSI/environment, high workload, poor training and no recovery potential is 0.56. (Note: The expert elicitation task has not been completed as of the writing of this report; some probabilities are listed as "unknown" and some may change in the final IDHEAS report.)

an expanded causal structure based on cognitive literature (Whaley et al., 2012) and PIF specification nodes corresponding to the questions in appendix B; and a second BN obtained through reduction of the first structure. Since the availability of data is the main bottleneck in HRA, the aim is to develop a BN structure, whose quantification requires roughly the same amount of information as the original DT. In section 9.1.3 the causal details in the original method are enhanced by explicitly including the PIF specifications and by adding PIFs that are essential to the interpretation of the CFM. This model can be quantified or used unquantified to help document the causal paths, on which the model is based. Section 9.1.4 demonstrates that node reduction algorithms can be used to reduce the BN with full causal details down to a structure, for which quantification is feasible with respect to data availability. In the presented example the final model is equivalent to the DT with explicit inclusion of the PIF specification nodes. In the following subsections, the general idea behind the structure development approach is discussed and the structure for the crew failure mode critical data misperceived is developed step by step. Quantification of the models is discussed

in section 9.2.

9.1.1 Summary of approach and models

An approach for developing causal (BN) models for HRA starting from the psychological basis of the models is proposed. The following steps summarize the approach:

- Review of the cognitive foundation for each CFM to identify the main causal failure paths, the PIFs and possibly other relationships.
- Development of an exhaustive causal model including all identified causal failure paths, PIFs and relationships.
- Application of node reduction algorithms, to remove nodes from the model that are not quantifiable with feasible effort.
- Elicitation of experts and initial quantification of model.
- Updating the quantification with results from human performance databases.

9.1.2 BN model of original IDHEAS DT

Each DT used to quantify the IDHEAS CFMs includes a number of PIFs. As explained in chapter 3 it is straightforward to develop a BN structure out of these PIFs and the target node, which is the node representing the CFM event. Since the PIFs influence the state of the CFM node, generally the PIFs are modeled as parents of the latter. From the structure it is clear that quantification requires conditional probability distributions for the CFM node and marginal distributions for the PIF nodes. The simple BN structure is shown in Fig. 9.2 for the CFM critical data misperceived. In this BN, the PIFs are assumed to be independent if the target node is not observed. The question if the PIFs are actually independent is not addressed within the original IDHEAS framework, since the IDHEAS decision trees provide HEPs only conditional on the states of all PIFs.

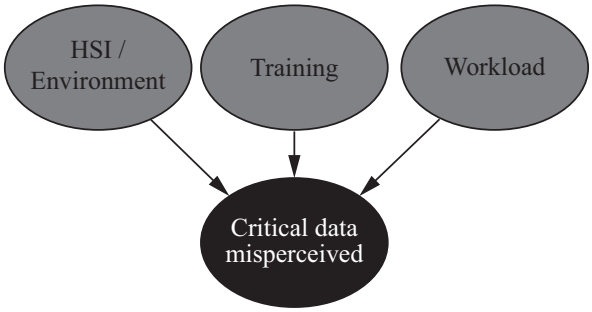


Figure 9.2: BN for the CFM critical data misperceived that corresponds to the original DT model.

9.1.3 BN model with full causal details

The BN model in Fig. 9.2, derived from the DT, reveals little about the cognitive paths leading to crew failure. This missing information is, however, essential to understanding the model with its features and limitations. The model is therefore expanded at two levels to the one BN shown in Fig. 9.3. Firstly, an additional layer of nodes is added (white in Fig. 9.3). These nodes are intended to specify the causal paths leading to error based on cognitive psychology. These nodes are often too generic or abstract for analysts to directly determine the states, but critical for correctly modeling human performance. Secondly, an additional layer of PIF specification nodes is introduced (light grey in Fig. 9.3). These are based on questions and rules for the analysts that are provided in IDHEAS to support the determination of the states of the PIFs (Appendix B).

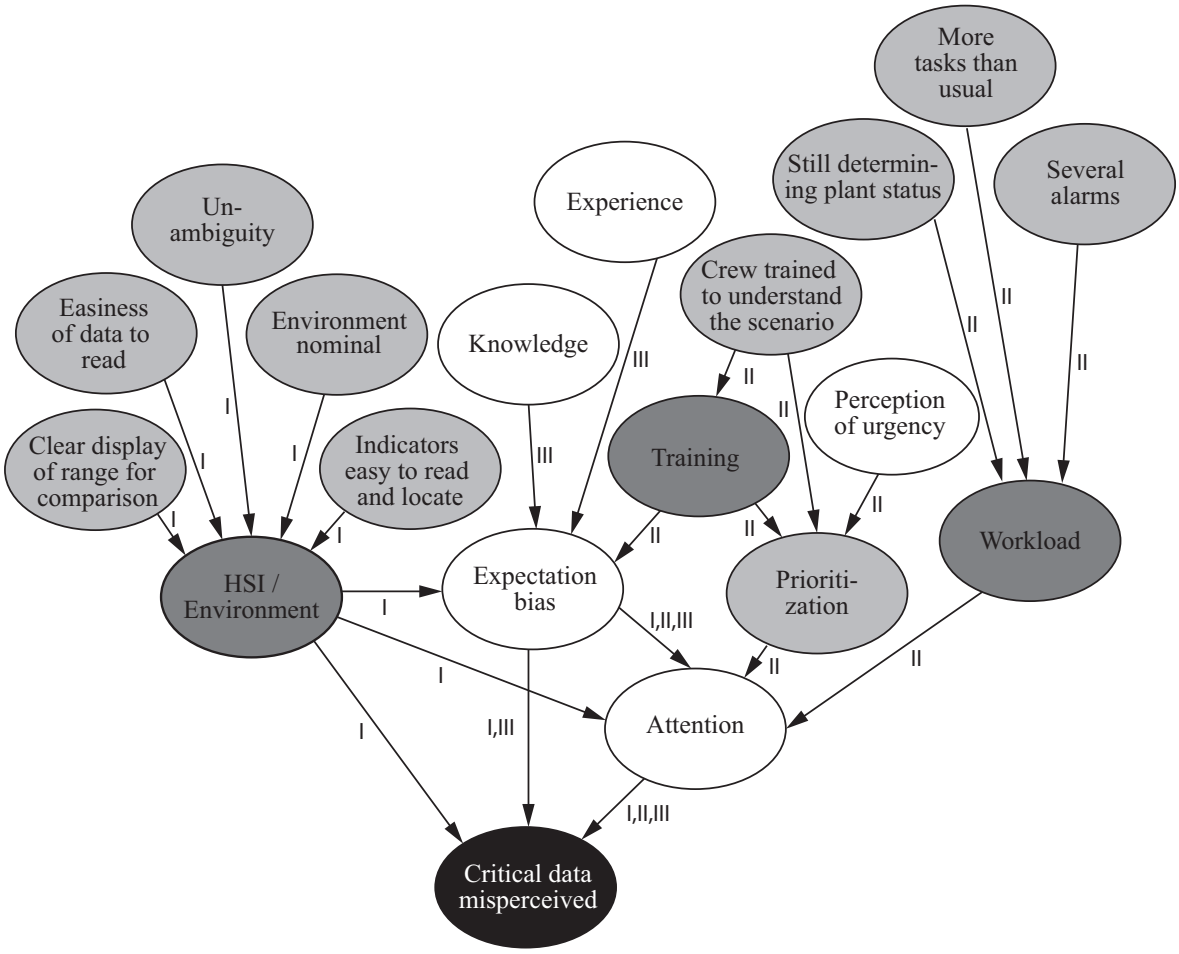


Figure 9.3: Fully expanded BN for the CFM critical data misperceived. The black node represents the target variable; dark grey nodes the PIF variables; light grey nodes the PIF specification variables and white nodes additional variables illustrating the causal paths. The causal paths I to III are indicated through roman numerals.

The literature serving as a foundation for IDHEAS (Whaley et al., 2012) summarizes the causal paths that can lead to a crew failure event, based on a comprehensive study of cognitive

psychology. These paths can be implemented directly through nodes in the model to add additional causal details extracted from scientific literature. For the example CFM (Fig. 9.3) there are three main causal paths leading to data misperceived, following cognitive literature (Köhler, 1970; Broadbent, 1958; Tversky and Kahneman, 1974; Biederman, 1987; Endsley, 1995; Klein, 1998; Warner and Letsky, 2008):

- The first causal path (path I in Fig. 9.3) corresponds to the misperception of data due to extreme HSI/environment conditions. In this case, the quality of the HSI is so poor, or certain factors in the environment are so severe, that the information is degraded in such a way that it is misperceived. For example, the operators may be inundated with many alarms so that they experience sensory overload (Broadbent, 1958) and therefore misperceive the critical data. Technically this could be seen as an instrumentation failure rather than a HFE, but this instrumentation failure would manifest as a human failure event (Endsley, 1995; Klein, 1998).

- The second causal path (path II in Fig. 9.3) is attention degradation that leads to misperception. Attention can be degraded due to a combination of factors, including characteristics of the situation and the information (e.g., the HSI and environment), high workload, multiple priorities, and through the biases introduced by training, knowledge, and experience. Training, workload and perception of urgency cause the crew to prioritize certain tasks and direct attention to these. A misdirection of attention can lead to misperception of critical data. The prioritization and the crew members' expectation biases determines the amount of attention paid to the various pieces of information, which again may lead to misperceiving the critical data (Eriksen and St. James, 1986; Endsley, 1995).

- The third causal path (path III in Fig. 9.3) stems from expectation biases related to experience and knowledge, which can cause misperception of critical data. This can occur in a direct manner, e.g., situations where a person "sees what they want to see", or indirectly through changing the person's attention to focus on other data (Einhorn and Hogarth, 1981; Endsley, 1995).

As shown in the model (Fig. 9.3), the PIFs identified in the IDHEAS model influence the occurrence of the CFM through multiple causal paths. HSI/environment influences the target CFM through one direct causal path and additionally through two indirect causal paths. Training also influences the CFM (indirectly) through two different causal paths. The third causal path, expectation bias, is only indirectly captured in the original IDHEAS model. The IDHEAS PIF specification nodes (light grey nodes) are intended to capture various aspects of the three PIFs, and are used in this model to demonstrate how observable questions can be explicitly included in the model. The node prioritization has a dual role. Firstly, it represents a PIF question specifying training, which is "Is the significance of the decision that is based on obtaining this information correctly given a high priority compared to other concurrent

tasks?". Secondly, prioritization is part of the second causal path. According to this path training influences prioritization. The link is thus directed from training to prioritization, and not like other PIF specifications the other way around. To capture the influence of prioritization according to its role as a PIF specification node correctly an additional dependence between prioritization and the node crew trained to understand the scenario needs to be introduced. Further discussions on the role of the node prioritization may be necessary, but are left for future research.

9.1.4 BN model reduction to facilitate IDHEAS-like quantification

The full model in Fig. 9.3 can be quantified using a variety of approaches. However, a secondary objective of this work is to develop a HRA model based on IDHEAS, and thus to limit the amount of additional information that must be elicited. To achieve this goal, the model in Fig. 9.3 is reduced to a form that more closely resembles the original IDHEAS DT, but is augmented with the PIF specifications from Appendix B. As shown in chapter 3 the node removal algorithm by (Shachter, 1986) is applied to the BN of Fig. 9.3. This algorithm allows removing nodes, which have not received evidence, in a way that the independence assumptions incorporated in a BN are not altered (Straub and Der Kiureghian, 2010a). The two principles of node removal are: Firstly, a node, which has not received evidence and which does not have children can be removed from the network. Secondly, the direction of a link between two nodes Z_i and Z_j can be reversed if Z_i inherits Z_j's parents and vice versa and if this does not cause the BN structure to become cyclic. Nodes are eliminated by first reversing all links so that the nodes to be removed have no children, and then removing them. In this way, the joint probability of all remaining nodes in the BN is unaltered. Removing the white nodes from the BN in Fig. 9.3 results in the BN of Fig. 9.4.

9.1.5 Discussion/ implications of the models

Of special interest is the causal role of the node prioritization in this structure. According to causal path II discussed in section 9.1.3 high workload increases the probability of misperceiving critical data, if the crew does not set correct priorities. In Fig. 9.3, prioritization is modeled as a child of perception of urgency and training. The causal interpretation is that both training and perception of urgency influence the probability of correct prioritization. Both prioritization and workload are parents of the node attention in Fig. 9.3. The combination of ineffective prioritization together with high workload will influence the attention paid to critical data. Our derivation of the BN model from the cognitive paths proposed a direct dependency of critical data misperceived on the node prioritization, which is not considered directly as a PIF in IDHEAS. A detailed discussion on whether the inclusion of this is necessary or whether there are reasons to exclude this PIF is not within the scope of this thesis. Since multiple cognitive literature sources indicate this dependency (Eriksen and St. James, 1986; Endsley, 1995), it is considered a critical PIF for accurately representing the cognitive

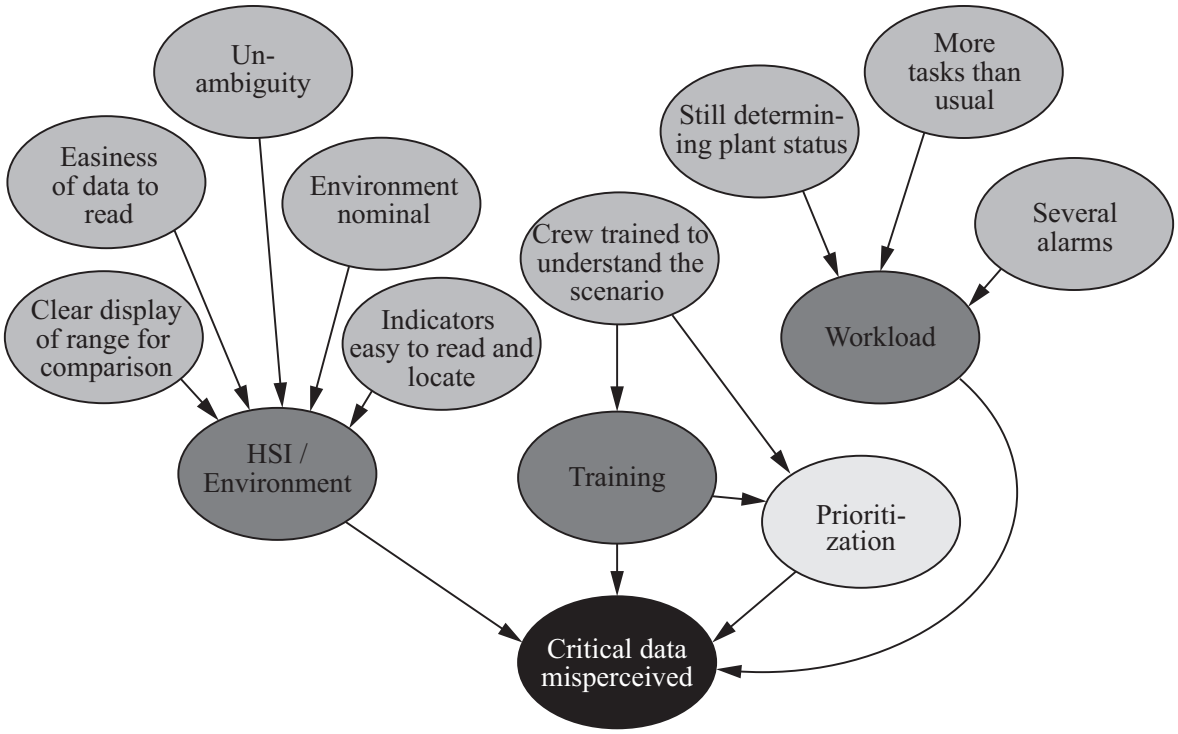


Figure 9.4: Reduced BN for the CFM critical data misperceived.

factors. The BN structure in Fig. 9.4 has advantages over the simple BN structure of Fig. 9.2. Firstly, in this model the analyst would directly answer the questions corresponding to the PIF specification nodes rather than assigning a PIF state based on implicit consideration of the questions, which is a much more abstract process . The explicit inclusion of PIF specifications in the model expands the level of documentation provided by the model, enhances the traceability from analysis input to probability estimate, and reduces variability among analysts. Secondly, if marginal probabilities are elicited for the PIF specification nodes, as was done in (Hallbert and Kolaczkowski, 2007), the BN in Fig. 9.4 can deal with missing information or uncertainty about one of the PIF specification nodes' states. For example, the HRA analyst may lack information about specific indicator design, which may make it difficult to assess the state of easiness of data to read. In situations where the analyst does not have information about one or more PIFs, the analyst can use the prior probabilities in the BN rather than guessing or making unwarranted assumptions about the system. Thirdly, the fully quantified BN can be used to reason about additional problems and gather additional insight. With identical analyst inputs, the BN structure in Fig. 9.4 will produce the same HEP assignments as the IDHEAS DT. However, the BN structure also offers the opportunity to reason about the PIFs, given knowledge of the CFM (and/or other PIFs). This provides an added benefit: the ability to identify, which PIFs (or PIF details) are likely to be present when one knows there is an HFE. This gives insight into the probabilities of the causes or HFEs, which is a critical piece of information that can be used to prevent errors (Groth and Swiler, 2013).

Table 9.1: CPT of the node critical data misperceived. The HEPs corresponding to the grey cells are marked as unknown in Fig. 9.1. For that reason, the estimates for the scenario [HSI = poor, workload = low and training = poor] were used. This corresponds to a conservative approximation, since changing the state of training from good to poor will certainly increase the HEP.

HSI/Env.	Poor				Good			
Workload	High		Low		High		Low	
Training	Poor	Good	Poor	Good	Poor	Good	Poor	Good
Recovery	No	No	No	No	No	No	No	No
Error	0.56	0.011	6.5e-3	6.5e-3	5.7e-3	1.6e-4	1.3e-4	1.3e-5
No error	0.44	0.989	0.9935	0.9935	0.9943	0.99984	0.99987	0.999987

9.2 CFM BN quantification

This section describes the quantification of the BN structures developed in the previous section. First the straightforward quantification of the BN model in Fig. 9.2 based on the IDHEAS DT is presented. It is furthermore shown how this simple model can be augmented with expert elicited data about the PIFs. Thereafter it is shown show how the BN of Fig. 9.4 can be quantified using expert estimates, and finally how information from the scenario authoring, characterization, and debriefing application (SACADA) (Chang et al., 2014) or similar databases can be used in this quantification. The BN of Fig. 9.3 can be quantified using a similar approach, but this is omitted for brevity.

9.2.1 Quantifying the BN model based on the original DT

As discussed in section 2, there is a CPT attached to each of the nodes in a BN. The CPT of the node critical data misperceived in the BN of Fig. 9.2 is identical to the conditional HEPs from the corresponding decision tree with one exception: the contribution of the recovery=yes branches is omitted because recovery is not clearly defined for this CFM. This CPT is shown in Tab. 9.1.

Quantification of the BN also requires probability distributions for each of the PIF nodes. Unlike the conditional HEPs these probability distributions are not provided by the original IDHEAS method. The marginal distributions of the PIF nodes can be quantified using dummy distributions (e.g. assigning the same probability to each state of a PIF). In that case quantifying the BNs does not require any additional probability elicitations compared to quantifying the DT models. However, if dummy distributions are used for the PIF nodes, the BN, like the DT model, is only capable of giving 'correct' HEPs if the states of all PIFs are known (i.e., the BN model will predict HEPs identical to the DT, but additional benefits of the BN cannot be realized). If the marginal distributions are actually elicited, the simple BN structure is capable of dealing with uncertainty about PIF states. In (Groth and Swiler, 2013) expert elicitations were used to quantify the CPTs of the PIFs, based on information elicited in (Hallbert and Kolaczkowski, 2007). Probability distributions for the PIFs of the

CFM critical data misperceived, which are based on (Groth and Swiler, 2013) are given in Tab. 9.2.

Table 9.2: Illustrative probabilities quantifying the CPTs of the PIFs. The probabilities are based on (Groth and Swiler, 2013).

PIF	PIF state	Pr (PIF)
HSI/Env.	good	0.16
	poor	0.84
Training	good	0.67
	poor	0.33
Workload	low	0.84
	high	0.16

9.2.2 Quantification of BN model of DT with PIF specification nodes

Quantification of the final BN model of Fig. 9.4 is illustrated. In this and similar BNs, there are three types of nodes to quantify: the CFM node (conditional on the PIFs), the PIF nodes (conditional on the PIF specification nodes), and the PIF specification nodes (marginal probabilities since these have no parents).

CFM node given PIFs

The parameters used in section 5.1 to quantify the CFM node were point estimates. Besides that (Xing et al., 2013) provides HEPs with corresponding uncertainty estimates (i.e. quantile estimates). In the following beta distributions are fitted to these quantile estimates (c.f. chapter 4). For proof-of-concept artificially generated data is used to update these fitted beta priors; the expected values of the posterior distributions are then used to quantify the final BN. Due to the updating as well as the fitting process, the parameters of the final BN do not exactly correspond to the parameters from Tab. 9.1. For quantification of the BN in Fig. 9.2 following section 9.2.1, the target node needs to be defined conditional on three parents i.e. the three PIFs. For the BN in Fig. 9.4, an additional direct dependency of the target node on the node prioritization was found. It is thus necessary to define the CPT of the target node on HSI/environment, training, workload and prioritization. Since IDHEAS does not provide probabilities for critical data misperceived conditional on the state of prioritization, simple assumptions are used here. For the purpose of illustration it is assumed that workload and prioritization interact in a way that the conditional probabilities of critical data misperceived are equal to:

- the conditional probabilities derived for low workload, if the crew is able to prioritize;
- the conditional probabilities derived for low workload, if workload is low and the crew is not able to prioritize;

- the conditional probabilities derived for high workload, if workload is high and the crew is not able to prioritize;

PIF nodes

In section 9.2.1 marginal probability distributions need to be assigned to quantify the PIF nodes of the simple BN in Fig. 9.2. For the quantification of the extended BN in Fig. 9.4, the PIF nodes are defined conditional on PIF specification nodes. The IDHEAS report (Xing et al., 2013) provides rules on defining the states of the PIFs given the states of the PIF specification nodes. To be in line with the original method these rules are used to quantify the CPTs of the PIF nodes. Presently, these rules are deterministic, which means they can be modeled as AND or OR relationships (deterministic nodes in the BN). These rules are provided for the CFM critical data misperceived as pseudo-code in Appendix C. Future research could focus on redefining these rules if it is found that these deterministic rules do not match reality, or if the relationship between the PIFs and PIF specifications is more nuanced than originally thought. It is, however, important that there is a common understanding on how the PIF specification nodes are linked to the PIF nodes, on how these nodes are included in the BN, and on how to quantitatively represent the dependence. These rules can directly be transformed to CPTs quantifying the respective nodes in the BN. For example the CPT for training given the two corresponding PIF specifications is provided in Tab. 9.3. The IDHEAS report defines training conditional on prioritization and crew trained to understand the scenario. However to represent causality, the node prioritization is considered as a child of training rather than its parent in the BNs of Fig. 9.3 and 9.4. The CPT derived from the rules provided in the IDHEAS report needs thus to be reformulated using Bayes' rule .

Table 9.3: Deterministic CPT of training given the two corresponding PIF specification nodes.

Crew trained to understand the scenario	Yes		No	
Prioritization	Yes	No	Yes	No
Good training	1	1	1	0
Poor training	0	0	0	1

PIF specification nodes

The PIF specification nodes require marginal probabilities. These probabilities are elicited from experts. Since a CFM like critical data misperceived typically does not apply in a nuclear power plant's normal state, for the elicitation process it is important that the experts understand that they are to give probabilities, which are implicitly conditional on scenarios, in which the CFM may apply. For example if the CFM critical data misperceived and the PIF specification node nominal environment are considered, the experts need to give a probability of the event environment being nominal in situations where critical data is received. As

Table 9.4: Results from the survey, carried out to elicit prior probabilities for the PIF specification nodes. Experts I and II are HRA specialists with a background in cognitive psychology and expert III is a former operator of a nuclear power plant on a submarine. The elicited numbers are probability estimates for the PIF specification nodes being in state "yes".

PIF specification node	Expert			Mean
	I	II	III	
Indications clear and unambiguous	0.9	0.8	0.9	0.87
Easiness of information to read	0.8	0.75	0.9	0.82
Clear display of range for comparison	0.8	0.8	0.95	0.85
Environment nominal	0.2	0.25	0.25	0.23
Indicators/source of data easy to read and locate	0.8	0.8	0.5	0.7
Still determining plant status	0.7	0.65	0.98	0.78
Several alarms	0.8	0.9	0.98	0.89
More tasks than usual	0.15	0.3	1.0	0.48
Crew trained to understand the scenario	0.8	0.8	0.98	0.86
Prioritization	0.95	0.9	0.9	0.92

proof-of-concept, the PIF specification nodes of the CFM critical data misperceived were quantified. The question, whether it is actually reasonable to elicit probabilities for such nodes, conditional on being in an off-normal state, from experts should be further discussed. Such a discussion is without the scope of this work.

A small survey was carried out to illustrate the process to elicit the prior probabilities of the PIF specification nodes. The survey participants were two HRA experts with a background in cognitive psychology (experts I and II) and one former operator of a nuclear power plant on a submarine (expert III). Since the final probabilities should be elicited from actual nuclear power plant operators, the numbers given in this report are intended only for illustrative purposes of the framework. The survey is shown in Appendix C and the numbers given by the experts are summarized in Tab. 9.4. There is a large spread in the experts estimates in questions 5, 6 and 8. Besides that the experts agree well on most questions. The mean values from Tab. 9.4 are directly used to quantify the marginal probabilities of the BN in Fig. 9.4, i.e. for the quantification of the PIF specification nodes parameter uncertainties are not considered.

9.3 Updating the parameters of the BN with data

In this section it is illustrated how the SACADA database (Chang et al., 2014) could be used to update the probabilities of the CFM node in the IDHEAS-BN. These HEPs in IDHEAS are conditional on the relevant PIFs. Since SACADA and IDHEAS are not completely consistent, it is not always possible to deterministically decide in which states the IDHEAS PIFs are for a given SACADA case. Nevertheless, SACADA still provides information, which can and should be used to improve the quantitative side of IDHEAS. To this end rules to probabilistically map SACADA onto IDHEAS are defined. These rules allow estimating a probability of observing a specific context $\Pr(PIF_1 = s_1, \ldots, PIF_m = s_m)$, which is defined through the states s_i of the PIFs, given a SACADA case. This concept is well known in HRA, e.g., from the quantification of SPAR-H (Gertman et al., 2005). Consider a crew failure scenario in a CFM, for which the corresponding probability $p_{CFM|PIF_1=s_1,\ldots,PIF_m=s_m}$ is to be updated. The prior distribution of $p_{CFM|PIF_1=s_1,\ldots,PIF_m=s_m}$ is beta with parameters a_0 and b_0. A database with n cases is used to update the distribution of $p_{CFM|PIF_1=s_1,\ldots,PIF_m=s_m}$. In the case where the database is not completely consistent with the PIFs, we can rewrite Eqs. 4.5 and 4.6 to:

$$a_1 = a_0 + \sum_{i=1}^{n} \Pr(\mathrm{PIF}_1 = s_1, \ldots, \mathrm{PIF}_m = s_m | \text{SACADA case } i) \cdot I_F(\text{SACADA case } i) \quad (9.1)$$

$$b_1 = b_0 + \sum_{i=1}^{n} \Pr(\mathrm{PIF}_1 = s_1, \ldots, \mathrm{PIF}_m = s_m | \text{SACADA case } i)\ (1 - I_F(\text{SACADA case } i)) \quad (9.2)$$

where I_F is an indicator function, which is 1 if a failure was recorded in the SACADA database case i and 0 otherwise. Applying Eqs. 9.1 and 9.2 requires the conditional probabilities $\Pr(\mathrm{PIF}_1 = s_1, \ldots, \mathrm{PIF}_m = s_m | \text{SACADA case } i)$. Expert estimates are used to determine the distributions of the relevant IDHEAS specification nodes for a given SACADA case i, and BN inference algorithms are used to calculate $\Pr(\mathrm{PIF}_1 = s_1, \ldots, \mathrm{PIF}_m = s_m | \text{SACADA case } i)$. Ideally the elicited experts should know both IDHEAS and SACADA well. Nevertheless if many SACADA indicators SI_i need to be mapped on PIF specification nodes PS_j from IDHEAS the expert elicitation becomes a tedious task. To simplify this process one can assign a factor $a_{\mathrm{PS}_j,\mathrm{SI}_i}$ to each SACADA indicator, which represents its effect on the PIF specification node. Qualitatively the effect of a SACADA indicator SI_i on a PIF specification node PS_j can be summarized as:

- $a_{\mathrm{PS}_j,\mathrm{SI}_i} = 0$, if SI_i being in state true causes PS_j to be in state false with certainty;
- $0 < a_{\mathrm{PS}_j,\mathrm{SI}_i} < 1$, if SI_i being in state true decreases the probability of PS_j being in state true;
- $a_{\mathrm{PS}_j,\mathrm{SI}_i} = 1$, if SI_i does not have an influence on PS_j;

- $a_{\mathrm{PS}_j,\mathrm{SI}_i} > 1$, if SI_i being in state true increases the probability of PS_j being in state true;
- $a_{\mathrm{PS}_j,\mathrm{SI}_i} = \infty$, if SI_i being in state true causes PS_j to be in state true with certainty.

Assuming that the joint effect of m SACADA indicators $\{\mathrm{SI}_1, \ldots, \mathrm{SI}_m\}$ on PS_j can be expressed as the product of the factors $a_{\mathrm{PS}_j,\mathrm{SI}_i}$ corresponding to SI_i one can write:

$$\Pr\left(\mathrm{PS}_j = \mathrm{true} | \mathrm{SI}_1, \ldots, \mathrm{SI}_m\right) = \min\left(1, \Pr\left(\mathrm{PS}_j = \mathrm{true}\right) \cdot \prod_{i=1}^{m} a_{\mathrm{PS}_j,\mathrm{SI}_i}\right) \tag{9.3}$$

For proof-of-concept these factors are estimated in Tab. 9.5 for the CFM critical data misperceived. It is important to note that the numbers in this table only serve the purpose of illustration. No factors are assigned to the SACADA indicators marked in grey in this table, since these indicators are redundant. From $\Pr\left(\mathrm{PS}_j = \mathrm{true} | \mathrm{SI}_1, \ldots, \mathrm{SI}_m\right)$ the probability $\Pr\left(\mathrm{PIF}_1 = s_1, \ldots, \mathrm{PIF}_m = s_m | \text{SACADA case } i\right)$ can be obtained through BN inference algorithms, which are implemented in any BN software.

Table 9.5: SACADA indicators, which can be related to PIF specification nodes in IDHEAS. Factors used to relate the two are given in the last column.

PIF specification node (IDHEAS) (PS_j)	SACADA indicator (SI_i)	Factor (a_{PS_j,SI_i})
	HSI/environment	
Environment nominal	Noisy background (Table A1, Miscellaneous)	0
	Overloaded (Table A2, Status of alarm board)	0.7
	Multiple alarms (Table B3)	0.7
Indicators easy to read and locate	Slight change (Table A3, Degree of change)	0.8
	Distinct change (Table A3, Degree of change)	1.5
	No mimics (Table A3, Degree of change)	0
	Small indications (Table A3, Degree of change)	0.7
	Similar displays (Table A3, Degree of change)	0.8
	Slight changes (Table B4)	
	Labeling/mimic display issues (Table B4)	
	Training	
Crew trained to understand the scenario	Standard (Table A4, Familiarity)	∞
	Novel (Table A4, Familiarity)	0.2
	Anomaly (Table A4, Familiarity)	0.2
	Unfamiliar (Table B6)	0
	Procedure-scenario mismatch (Table B6) $\rightarrow$ Novel	0.2
	Prior Experience (Table B6) $\rightarrow$ Anomaly	0.2
Prioritization	Competing priorities (Table A5, Uncertainty)	0.5
	Conflicting guidance (Table A5, Uncertainty)	0.5
	Competing priorities (Table B6) redundant	
	Conflicting guidance (Table B6) redundant	
More tasks than usual	Normal (Table A1, Workload)	0
	Concurrent demand (Table A1, Workload)	
	Multiple concurrent demands (Table A1, Workload)	2
	Multiple demands (Table A1, Miscellaneous)	
	Coordination (Table A1, Miscellaneous)	1.1
	Workload	
Several alarms	Dark (Table A2, Status of alarm board)	0
	Busy (Table A2, Status of alarm board)	∞
	Overloaded (Table A2, Status of alarm board)	∞
	Multiple alarms (Table B3, Background)	
	Not applicable (Table B3, Background) redundant	

Table 9.6: HEPs for different observations.

	Case I	Case II	Case III	Case IV	Case V
Clear display of range for comparison	-	Yes	No	-	-
Easiness of data to read	-	Yes	No	-	-
Unambiguity	-	Yes	No	Yes	-
Environment nominal	-	Yes	No	Yes	-
Indicators easy to read and locate	-	Yes	No	Yes	-
Crew trained to understand the scenario	-	Yes	No	Yes	-
Prioritization	-	Yes	No	-	No
Still determining plant status	-	No	Yes	Yes	No
More tasks than usual	-	No	Yes	-	No
Several alarms	-	No	Yes	Yes	No
HSI/environment	-	-	-	Good	
Training	-	-	-	-	
Workload	-	-	-	-	
HEP	0.01	0.00005	0.5	0.002	0.0003

9.4 Example results with the "critical data misperceived" BN

With the established BN for critical data misperceived, HEPs conditional on different observations are investigated (Tab. 9.6). Case I gives the prior HEP before having knowledge about the states of the PIFs or the PIF specification nodes. The states of the PIF specification nodes occur in that case according to the probabilities elicited from the experts. The BN gives reasonable prior HEPs if the CPTs of the PIF specification nodes are elicited (either based on data, experts or similar sources) and not populated with dummy parameters. The capability of giving such probabilities sets the BN apart from the decision trees originally used to quantify IDHEAS. Cases II and III represent the extreme cases of the CFM critical data misperceived. In Case II the states of all PIF specification nodes are observed and all of them are in a favorable state. The HEP is therefore minimal for that case. In Case III all PIF specification nodes are in an unfavorable state, hence the corresponding HEP is maximal. Both cases can also be derived from the original IDHEAS decision trees. Since evidence is here given to all PIF specification nodes, it is irrelevant if the CPTs of these nodes are elicited or populated with dummy parameters. Cases IV and V represent cases with missing information. In Case IV some of the questions corresponding to the PIF specification nodes have not been answered. The same is true for Case V, which additionally demonstrates that evidence can also be given directly at the PIF level.

Besides providing evidence at the level of PIFs or PIF specification nodes, it is possible to directly give evidence on the target node. It is for example possible to determine the distribution of the PIF nodes given a HFE as:

- Pr (HSI\Environment = poor|HFE = yes) = 0.998
- Pr (Workload = high|HFE = yes) = 0.996,
- Pr (Training = poor|HFE = yes) = 0.532
- Pr (Prioritization = no|HFE = yes) = 0.675.

9.5 Discussion

This chapter presents a comprehensive framework for the application of BNs to address shortcomings of HRA with respect to scientific basis and traceability (both causal and quantitative). A main advantage of BNs is that it allows for models that are causally traceable. As shown, to this end unobservable PIFs and concepts from psychology can be included in the BN structure and removed in a later step. Furthermore the quantification of BNs can rely on different information sources, like data and expert elicitations. Causal traceability is a major need in the field of HRA. In this chapter it was demonstrated how an expanded BN structure can qualitatively document the theoretical background of the method. Furthermore it was demonstrated how to reduce that structure to maintain causal traceability and to enable a more straightforward quantification than the full expanded structure. While both structures are quantifiable from a mathematical point of view, quantification of the expanded structure is difficult from HRA-perspective, since data or experts that are capable of estimating the specific probabilities are not available.
If the BN is implemented in a software tool, the additional nodes of the expanded structure can be marked in a separate color, to highlight that these nodes are necessary for the understanding of the causal relationships but are not quantified. While many recently developed HRA methods have a strong background in psychological research, this background becomes usually hidden for more applied users, who are presented only a reduced number of PIFs. By developing expanded BN structures and presenting them to users the theoretical background becomes more traceable even if it may not be possible to provide it in full detail in this manner. It has been found by many researchers that the results of a HRA vary strongly with the analyst (e.g. Lois et al., 2009). This is currently a major point of criticism against HRA methods. An example of how the proposed framework can help to increase causal traceability was presented in this work by the application of the framework to the CFM critical data misperceived. An additional dependency between the node prioritization and the target node was revealed through the process of building an exhaustive BN structure and reducing it. For the purpose of traceability of the HRA method it is important that the model developer is aware of additional causal details like these and communicates them to the analysts. It is then up to the model developer to decide whether quantification of these causal details is necessary or not.
Another major need in HRA, which is addressed here, is an exhaustive and rigorous quantification framework. It is generally known that HRA models are not capable or even intended

to fully capture all aspects of human behavior. In spite of this, it is necessary to model human error, using all information and knowledge available. Many HRA researchers rely on quantifying their models either through experts or through data. Our proposed quantification framework combines these two, which is in line with the Bayesian understanding of probability used throughout PRA (Kelly and Smith, 2009) and is the only method to come up with sound probability estimates in an industry with scarce data. Using Bayesian updating allows using continuously more data to update the parameters of the BN, in order to improve the quality of the model. A last point implicitly addressed in this work is the applicability of BN-based HRA methods for every-day HRA practice. While HRA researchers may be tempted to embrace BNs simply for their powerful modeling features, HRA practitioners call for models, which are applicable in their everyday practice. Not many of the BN HRA models developed up to this point satisfy this need. By developing a BN which is scalable to different sizes, we offer the potential to have the same HRA method meet the needs both practitioners and researchers.

Part IV

CONCLUSION

Chapter 10

Concluding remarks and outlook

10.1 Concluding remarks

Reliability is an important requirement for engineering systems and it is the responsibility of engineers to guarantee high reliability standards for increasingly complex systems. These high reliability standards hinder an estimation of the probability of failure of a system solely based on observed data. Instead, probabilistic modeling that is not solely based on data is required to estimate failure probabilities of such systems.

Flexible modeling tools are required in order to be able to represent complex systems. BNs represent a flexible and powerful modeling tool with large potential for engineering reliability analyses. The main reasons for that are: The ability of BNs to traceably represent the overall dependence structure of a model, the possibility for combining different sources of information for the quantification of BNs, and the existing inference algorithms allowing for straightforward Bayesian updating in BNs. A reliability analysis with BNs can be considered a three step procedure, whose steps are (I) elicitation of the qualitative model structure (the directed acyclic graph (DAG)), (II) quantification of the model and (III) inference in the model.

The DAG represents the overall dependence structure of a model. The elicitation of the DAG is crucial for two reasons. Firstly, the DAG is the part of the model that represents the overall dependence structure between the random variables in a way that is traceable even for people being non-experts in probabilistic modeling. This allows domain experts and stakeholders to challenge the assumptions made. In order to ensure the validity of models, this feature is essential. Secondly, the DAG determines the types of local conditional probability distributions that are required to quantify the model. In consequence the DAG determines also the (computational) effort required, to quantify the model and to perform inference in the model.

In practice, BN structures are often elicited in a rather ad-hoc manner. To give modelers some guidelines on how to develop a BN structure, a classification of structure elicitation approaches is proposed in this thesis. Distinguished are: (1) structure elicitation based on

other probabilistic models; (2) structure elicitation based on general models; (3) data based structure learning and (4) structure elicitation based on domain expert knowledge.

These approaches require different levels of expertise both in the problem domain as well as in BN theory. For example, it is typically straightforward to elicit a BN based on other probabilistic models like fault trees (FTs) or event trees (ETs), which are commonly used in various industries. In contrast, eliciting a BN based on domain expert knowledge typically requires a deeper understanding of BN theory and also the problem domain. In such cases, causality can be used as a guiding principle for developing valid BN models. To demonstrate how BN structures can be derived from domain expert knowledge, a framework for deriving BNs for human reliability analysis (HRA) is proposed in this work. A major shortcoming of HRA is the fact that the probabilistic models used are typically oversimplifying reality. This is mainly due to the limitations in the understanding of the cognitive processes that lead to human failure but also due to the limited flexibility of the probabilistic modeling tools that are used. It is shown in this application example, that BNs represent a well-suited modeling framework for HRA, which has advantages over many of the modeling tools that are traditionally used in this field. Future research in the area of HRA should focus on developing more realistic HRA models that use BNs as a probabilistic modeling framework.

Despite the fact that BNs represent a powerful tool for many reliability engineering applications, there are situations in which BNs are not a suitable tool. These include problems, in which the dependence between two random variables does not have a causal direction or where the dependence is bidirectional. Examples include two points in a random field (modeled without a "hidden" common parent) or infrastructure (e.g. water) networks with bidirectional flows. In such situations other graphical probabilistic models can represent a worthwhile alternative (Koller and Friedman, 2009).

Quantitative information is typically scarce for applications in the field of reliability engineering. Because of this, all available information sources should be combined to come up with sound reliability estimates. The BN framework lends itself to such a combination of information sources, since every node is quantified independently through a local conditional probability distribution.

BN inference algorithms enable computing distributions of variables of interest (possibly) conditional on observations, obtained for other variables. Exact inference algorithms are available only for discrete BNs and a number of continuous/hybrid special cases. By discretizing continuous nodes, the same inference algorithms can also be applied to continuous/hybrid problems. This comes, however, at the cost of a discretization error. Discretization of continuous random variables in reliability applications is especially critical, since in these applications one is typically interested in rare events and thus in the tails of the distributions. An efficient discretization procedure for reliability applications that has also been implemented in a MATLAB based software tool is proposed in this thesis. This procedure aims at minimizing the discretization error, while using only a minimal number of intervals. The performance of the proposed method is investigated for a number of verification examples. For these examples

the discretization error is found to be small.
The applicability of the approach to practical problems is demonstrated through an example from the field of civil aviation. In particular a prototype of a runway overrun (RWO) warning system is developed. This warning system allows using measurements that can be obtained during the approach of a landing aircraft, to update ones believes about the aircraft's RWO probability. This probability can support the decision, whether to continue with the approach. The main reason for using exact inference algorithms in this case is their capability for performing robust and fast Bayesian updating.
While the capability for robust and fast updating makes the approach attractive for near real-time applications in general, the exponential increase of the CPT limits the number of basic random variables that can be included as nodes in the BN. In many cases it is however not necessary to directly model all basic random variables as explicitly as nodes.
Sampling based inference is applicable to both discrete and continuous\hybrid BNs, which makes discretization of continuous random variables redundant in this context. As a consequence, sampling based inference algorithms also mitigate the problem of exponentially increasing CPTs. However sampling based inference is often slower than exact inference, which is especially an issue for near real-time applications. Possible future enhancements could however make sampling based inference applicable to a larger number of near-real-time problems.
Memory issues are critical especially for large systems. In this cases the size of the CPTs or the clique sizes can hinder the application of exact (discrete) inference algorithms. Sampling based inference techniques are not sensitive to clique sizes. Furthermore, since sampling based inference algorithms do not require all nodes to be defined through CPTs, nodes having many parents can often be defined as functions of their parents or as distribution models, whose parameters depend on the parents of that nodes. Thus the memory issues typically encountered when dealing with large systems can be alleviated by applying sampling based inference. However the time required to sample from an increasingly large BN limits the size of a system, also if sampling-based inference techniques are applied. This limitation is however often less restrictive, than the limitation coming from computer memory for discrete BNs. A drawback of sampling based approaches is that they are typically not tailored towards reliability applications. I.e. they are not suited for computing small (failure-)probabilities, since the number of samples required to estimate small probabilities and thus also computation time is too high.
In order to overcome this issue, in this work the applicability of subset simulation (SuS), a sampling based method that is intended for computing small failure probabilities, to sampling based inference in BNs is investigated. A main factor, influencing the performance of SuS is the Markov chain Monte Carlo (MCMC) algorithm that is used to generate samples. Unlike in the original SuS method Gibbs sampling is applied here. The reason being that Gibbs sampling is especially efficient for BNs, since it can exploit the encoded independence properties. Just like the implementation of the original SuS algorithm, the implementation of SuS for BNs is

straightforward. In this work, OpenBUGS in combination with MATLAB were used to do so. A simple verification example has been implemented to investigate the performance of the approach. Although the results obtained are promising, convergence properties as well as computational efficiency of the method should be further investigated.
In conclusion, this thesis shows that BNs represent a powerful modeling tool for reliability engineering. Nevertheless to fully exploit their potential for this type of applications, efforts should be made to develop methods that address the needs in this field. Some points that could be addressed by future research are briefly discussed in the outlook at the end of this chapter (cf. section 10.3).

10.2 Main contributions of this thesis

The main contributions of this thesis are:

1. A classification of BN structure elicitation approaches that are commonly applied in the field of reliability engineering and engineering risk analysis.

2. A novel, efficient discretization approach that enables the use of exact BN inference algorithms for approximate inference in hybrid BNs. The procedure is applicable for reliability analyses for engineering systems, whose performance is described through an engineering model.

3. Combination of subset simulation (SuS) with sampling based inference to estimate small (failure-)probabilities from hybrid BNs that model reliability problems.

4. A framework for eliciting BN structures for HRA based on domain expert knowledge.

5. An approach for quantification of the HRA BNs that is based on expert estimates and simulator data.

6. The development of a prototype of a runway overrun warning system that is based on the proposed discretization procedure.

10.3 Outlook

Challenges for the application of BNs to reliability engineering have been addressed in this work. However there are open issues. A few points that could be approached by future research are touched upon in the following.

10.3.1 Promoting applications of BNs in industry

While there are many applications of BNs in the academic context, they are less often applied in industry. In fact, industries that perform quantitative reliability analyses are typically not familiar with BNs and thus often apply traditional modeling tools like FTs or ETs. These traditional modeling tools are typically intuitive to apply, however in many cases BNs represent a more powerful modeling tool. While modeling with BNs can be more challenging, in many cases BN elicitation can be standardized. Thus future research should focus on developing BN frameworks for specific industrial application areas.
As an example, one can consider accident modeling in the civil aviation industry. Bow ties (a combination of a FT and an ET) are typically applied in this industry. A bow tie is readily representable in a BN (cf. chapter 3). In addition BNs are capable of modeling dependencies that are not straightforward to model through bow ties. Developing a modeling framework based on BNs could promote more sophisticated probabilistic models and thus enhance reliability analyses.
Such frameworks could be accompanied by software tools that are tailored towards specific fields i.e. they could enable simple bow tie like modeling, while at the same time providing more sophisticated modeling capabilities.

10.3.2 Exact inference algorithms for large system reliability problems

System reliability problems usually define the performance of a system as a function of the performance of its components. Such problems are readily representable by means of discrete BNs (cf. chapter 3). The naïve approach for modeling such systems represents the component nodes as parents of the the system node. The size of systems that can be represented in this way is very limited. More efficient structure representations (parent divorcing, causal independence representation etc.) have been developed and applied to system reliability problems in the past (e.g. Heckerman and Breese, 1994; Bensi et al., 2013). While future research could focus on developing novel efficient structure representation techniques, this seems to be rather unpromising. It may be more promising to exploit the dependence structure of specific systems and develop efficient BNs for these systems by using the generic, efficient modeling techniques that already exist.
In order to apply exact inference algorithms to BNs representing larger systems, additionally techniques like compression of the CPTs, as described by (Tien and Der Kiureghian, 2016), can be applied.

10.3.3 Sampling based inference

Sampling based inference algorithms can be considered as an alternative to exact inference algorithms for large systems. Additionally sampling based inference algorithms are not only applicable to discrete but also to hybrid/continuous BNs. Existing sampling based BN inference algorithms are however not tailored towards reliability applications. Firstly, they are typically not suitable for estimating small (failure-)probabilities. Secondly, traditional sampling based inference algorithms perform poorly for inverse analysis i.e. problems, where the performance of a system is described through a deterministic model and the state of the system is observed.
Independent of BNs, methods have been proposed that address these issues. For instance existing structural reliability methods enable estimating small failure probabilities efficiently (Rackwitz, 2001; Au and Beck, 2001). Furthermore methods that enable efficient inverse analyses have been developed (e.g. Beck and Au, 2002; Straub, 2011; Straub and Papaioannou, 2014). A first step towards combining structural reliability methods (i.e. SuS) with inference in reliability BNs has been made in this thesis. The first results that were obtained are promising. The BUS (Bayesian updating with structural reliability methods) approach, as proposed in (Straub, 2014a) is based on SuS. Applying BUS in a BN context may be a promising approach for efficiently performing inverse analysis in BNs.
Due to that, future research should focus on applying both SuS and also BUS to BNs. Moreover the applicability of other structural reliability methods and methods for inverse analysis, to BNs should be considered by future research.

10.3.4 Combination of exact and approximate inference algorithms

In some situations that are relevant to reliability engineering, parts of a model are readily represented through a discrete BN (e.g. an infrastructure system), while other parts of the same model are more naturally described through a hybrid/continuous BN (e.g. natural hazards acting on the infrastructure system). Since exact inference algorithms exist for the former but not for the latter, combining exact and approximate inference algorithms could be a worthwhile consideration. Efforts to develop the theoretical foundation for doing so have already been made (cf. Koller and Friedman, 2009). However to the best of my knowledge this has not been applied in the context of reliability engineering. Future research could therefore focus on investigating whether these approaches are applicable and whether they perform well for reliability engineering problems.

Part V

APPENDICES

Appendix A

CPTs of the BN in Fig. 3.12a

Table A.1: CPT of the vegetation node in Fig. 3.13a

slope	$< 30°$	$30° - 45°$	$> 45°$
dense trees	0.3	0.5	0
no dense trees	0.7	0.5	1

Probabilities for the node avalanche:

- $\Pr(\text{avalanche} \mid \text{summer}) = 0$
- $\Pr(\text{avalanche} \mid \text{slope} < 30°) = 0$
- $\Pr(\text{avalanche} \mid \text{slope} > 45°) = 0$
- CPT of avalanche conditional on season, dense trees and slope between 30° and 45°:

season	spring	summer	fall	winter
avalanche	0.01	0	0.02	0.05
no avalanche	0.99	1	0.98	0.95

- CPT of avalanche conditional on season, no dense trees and slope between 30° and 45°:

season	spring	summer	fall	winter
avalanche	0.03	0	0.06	0.15
no avalanche	0.97	1	0.94	0.85

Appendix B

Questions defining the PIFs

Questions provided in IDHEAS (Xing et al., 2013) for the CFM critical data misperceived, to specify the meaning of the PIFs: The PIF human system interface (HSI)/environment is considered to be in state poor only if any of the following questions are answered with "no" (Xing et al., 2013, 93):

- "Are the indications clear and unambiguous"?
- "Is the information easy to read"?
- "Is the range (or band) with which the information is to be compared clearly identified on the display"?
- "Is the environment in the location of the indicator/source of information nominal (i.e., not challenging due to noise, heat, humidity, etc.)"?
- "Are the indicators/sources of data easy to locate and read"?

The PIF workload is considered to be in state high only if any of the following questions are answered with yes (Xing et al., 2013, p. 93f):

- "Does the need to obtain information occur at a time when the operators are still in the process of determining the plant status?"
- "Does this occur at a time when there are several alarms or indications or tasks that need attention? "
- "Is the scenario one for which the number of tasks the crew has to perform in the time available higher than would be typically addressed in training?"

The PIF training is considered to be in state poor only if both of the following questions are answered with yes (Xing et al., 2013, p. 94f):

- "Has the crew been properly trained to understand and deal with scenarios in which the information source may provide difficulties?"

- "Is the significance of the decision that is based on obtaining this information correctly given a high priority compared to other concurrent tasks?"

Appendix C

Quantification of PIF nodes

C.1 Deterministic rule quantifying the node HSI/environment:

IF

- Indications are clear and unambiguous
- AND indications are easy to read
- AND the Range (or band) with which the information is to be compared is clearly identified on the display
- AND the environment in the location of the indicator/source of information is nominal
- AND the indicators/sources of data are easy to locate and read

THEN

- HSI/environment is good

ELSE

- HSI/environment is poor

C.2 Deterministic rule quantifying the node training:

IF

- The crew has been properly trained to understand and deal with scenarios in which the information source may provide difficulties
- OR the significance of the decision that is based on obtaining this information correctly is given a high priority compared to other concurrent tasks (referred to as prioritization in the model from Figs. 9.3 and 9.4).

THEN

- Training is good

ELSE

- Training is poor

C.3 Deterministic rule quantifying the node workload:

IF

- The need to obtain information occurs at a time when the operators are still in the process of determining the plant status
- OR this occurs at a time when there are several alarms or indications or tasks that need attention
- OR the scenario is one for which the number of tasks the crew has to perform in the time available is higher than would be typically addressed in training

THEN

- Workload is high

ELSE

- Workload is low

Appendix D

Survey for eliciting probabilities of PIF specification nodes

Purpose of this survey:
The purpose of this survey is to obtain probabilities, which can be used to illustrate a framework for quantifying IDHEAS. (Note that data will be used for proof-of-concept of the mathematics, to define how IDHEAS models could be modified to include information about the probability of PIFs. IDHEAS models will not be modified based on the results of this survey. Background: IDHEAS provides models that assign the probability of a human failure event, given the state of several performance influencing factors (PIFs). IDHEAS contains 14 models and approximately 20 PIFs. Description of survey task: Consider the crew failure mode critical Data Misperceived, which is defined as: "A critical piece of information that is required to develop a plant status assessment is misperceived. A critical piece of data is one that, when misperceived in a certain way will lead to an incorrect response in that it leads to taking an incorrect or inappropriate path through the procedures or executing a response incorrectly." (From SRM Vol. 3 – Draft IDHEAS method for internal procedural events) We consider a post-initiator event i.e. the nuclear power plant (NPP) is already in an off-normal state, where the NPP-crew is confronted with critical data. For critical data misperceived, the IDHEAS model identifies three main factors PIFs, which influence human performance: Human System Interface (HSI)/environment, training, and workload. Furthermore, IDHEAS provides several questions that analysts use to assign the state of those three PIFs. In this survey, you are asked to provide a probability for each of these questions.
Note:

- A (Bayesian) probability is a degree of belief rather than an actual physically measurable quantity
- You can give your answer in one of the two forms:

  The probability of event X occurring is _____.

  Event X is _____-times more/less (more/less) likely than not X.

Name: _____
Company: _____
Position: _____

Brief description of your experience/background:
Basis for estimates in this document (e.g., "12 years of experience operating commercial NPP", "HRA database", "22 years of experience in HRA"):
Reminder: The purpose of the survey is to obtain the probability of these conditions, NOT to obtain the human error probability for these conditions.

Human System Interface (HSI)/environment:

1. Unambiguity and clearness of indications.

   The probability that the indications to this data are clear and unambiguous is _____.

   Indications are _____-times more/less likely to be clear and unambiguous than to be unclear and ambiguous.

2. Easiness of information to read.

   The probability that the information is easy to read is _____.

   Information is _____-times more/less likely to be easy to read than to be not easy to read.

3. Range (band) for comparison.

   The probability that the range (or band), with which the information is to be compared, is clearly identified on the display is _____.

   A display is _____-times more/less likely to have a clearly identified range (band) than to have a not clearly identified range (band).

4. Nominal environment.

   The probability of having a nominal environment (i.e. one that is not challenging due to noise, heat, humidity, etc.) is _____.

   During the event it is _____-times more/less likely that environment is nominal rather than non-nominal environment (challenging due to noise, heat, humidity, etc.).

5. Location and easiness to read of the indicators/sources.

   The probability that indicators/sources are easy to locate and read is _____.

   Indicators/sources are _____-times more/less likely to be easy to read and locate than to be not easy to read and locate.

Workload:

6. Crew still determining the plant status.

   The probability that the need to obtain information occurs at a time when the operators are still in the process of determining the plant status _____.

   Crew still determining the plant status is _____-times more/less likely than crew not in the process of determining the plant status.

7. Several alarms.

   The probability that the need to receive critical data occurs at a time when there are several alarms, indications or tasks that need attention is _____.

   A situation with Several alarms, indications or tasks that need attention at the same time is _____-times more/less likely than a situation without several alarms, indications or tasks.

8. More tasks in the available time than typically addressed in training.

   The probability that the number of tasks the crew has to perform in the available time is higher than it would be typically addressed in training is _____.

   A situation with a higher number of tasks than addressed in training is _____-times more/ less likely than a situation with same or a lower number of tasks than typically addressed in training.

Training:

9. Crew trained to understand the scenario.

   The probability that the crew has been properly trained to understand and deal with the scenarios, in which the information source may provide difficulties is _____.

   Crew properly trained to understand and deal with the scenario is _____-times more/less likely than crew not properly trained to understand and deal with the scenario.

10. Significance of the decision that is based on this information.

    The probability that the decision based on obtaining this information correctly is given a high priority compared to other concurrent tasks is _____.

    Giving the decision, which is based on the information, a high priority is _____-times more/less likely than not giving the decision a high priority.

List of Figures

1.1 Example of a BN modeling the damage of a house due to a storm event. . . . 5

2.1 Example BN. . . . 10
2.2 Design point and linear approximation of the limit state surface. Left side: original random variable space; right side: standard normal space (from (Straub, 2014b)). . . . 12
2.3 MCMC algorithm for SuS. . . . 15

3.1 Representation of AND/OR gate from a FT respectively a series/parallel subsystem from a RBD in a BN. The events X_i and Y are represented through binary nodes X_i^* and Y^* in the BN. . . . 23
3.2 Example FT modeling power supply. Power can be supplied, if enough power is generated and the power can be transmitted. Enough power is generated if at least two power plants are working. A power plant fails, if a tsunami occurs and no mobile flood walls are erected. . . . 24
3.3 BN corresponding to the FT from Fig. 3.2. . . . 24
3.4 Representation of an ET in a BN. It should be checked carefully that the sequence in the ET follows the causal direction. From the conditional probabilities in the ET it can be assessed, whether a link is required or not. . . . 26
3.5 Dynamic BN model of a stochastic Markov process. . . . 27
3.6 In a) a Markov network is shown, for which it is impossible to represent all its (conditional) (in-)dependence properties simultaneously in a BN. Both BN I and BN II fail at doing so. . . . 28
3.7 An ET modeling the consequences of a ship collision starting from the initiating event *ship being on a collision course* (a). The events A and B from the ET are modeled through FTs (b). The corresponding BN representing both the ET and the two FTs is shown in (c). . . . 29
3.8 Naïve BN representation of a general model $Y = g(\mathbf{X})$. To model additional dependencies among the variables $\mathbf{X}$, links between the variables X_i or common parents can be added. . . . 30
3.9 Parent divorcing. . . . 31

3.10 Causal independence representation. 32
3.11 BN derived from a physical model (Zwirglmaier and Straub, 2015). 33
3.12 NBC for runway overrun taking into account the three parents of runway overrun from Fig. 3.11 as indicators. 33
3.13 a) Original BN for predicting avalanche probability. b) BN learnt with 500 samples using a score based structure learning approach. 35
3.14 BN structure for critical data misperceived (Zwirglmaier et al., 2017). Following the node removal algorithm from (Shachter, 1988), the unquantifiable (white) nodes are removed from the network. 40
3.15 Extension of a simple causal respectively a diagnostic BN model by an additional cause. 41

4.1 Updating the beta prior distribution using two different data sets. The parameters of the prior , $f_0(\theta)$, are $a_0 = 5$ and $b_0 = 45$. Posterior I is obtained by updating with a dataset of length $n = 25$ in which $n_e = 4$ failures are observed. Posterior II is obtained by updating with a dataset of length $n = 50$ with $n_e = 5$ failure occurrences. 46
4.2 PDFs of beta distributions with mean 0.1 and varying standard deviations. The standard deviations of the distributions are a) 0.03, b) 0.07 and c) 0.01. . 48

5.1 Moral graph for the BN in Fig. 2.1. 56
5.2 Junction tree for the BN in Fig. 2.1. 57

6.1 General structure of a BN model of a component reliability problem. 60
6.2 Representation of a basic reliability problem with n independent basic random variables in a BN. Left: original problem with continuous basic random variables X_i, right: discrete BN, in which X_i are substituted with discrete nodes Y_i. 61
6.3 Discretization error in 1D. 62
6.4 General structure of a BN model of a component reliability problem. 63
6.5 Schematic representation of a discretization of a linear 2D reliability problem. w_i is the distance between interval boundaries d_1^i and $d_{n_i-1}^i$. All intervals between these boundaries are equi-spaced. v_i is the position of the midpoint of the discretization frame relative to the design point $\mathbf{u}^*$ in dimension i. 65
6.6 Transformation of a discretization scheme from U-space to X-space. To preserve orthogonality each interval boundary in U-space is represented by a characteristic point. The random variables X_1 and X_2 are Weibull distributed with scale and shape parameter 1 and their correlation is 0.5. 66

6.7 Linear problem in standard normal space, with two random variables U_1 and U_2, where $\alpha_1 = 0.45$ and $\alpha_1 = 0.89$. The intervals cut by the limit state surface, i.e. those which potentially lead to a posterior discretization error are marked in grey. 67

6.8 Logarithm of the probability mass enclosed by the discretization frame plotted against $|\alpha_i|$. Φ denotes the standard normal CDF and ub_i respectively lb_i the last (upper) and the first (lower) interval bound in dimension i. 68

6.9 Optimization results for 10 two-dimensional, linear problems in standard normal space, which are discretized with 5, 10 and 20 intervals per dimension. In all cases the prior failure probability is 10^{-7} ($\beta = 5.2$). The crosses represent the optimization results. The solid lines are the fitted parametric functions (Eq. 6.8). The left-hand side shows the relation between the width of a discretization frame w_i and α_i^2 and the right-hand side shows the relation between the probability mass enclosed by the discretization frame with width w_i and α_i^2. 76

6.10 Optimization results for 10 two-dimensional, linear problems in standard normal space, which are discretized with 10 intervals per dimension. The prior failure probabilities are 10^{-3} ($\beta = 3.1$), 10^{-5} ($\beta = 4.3$) and 10^{-7} ($\beta = 5.1$). The crosses represent the optimization results. The solid lines are the fitted parametric functions (Eq. 6.8). The left-hand side shows the relation between the width of a discretization w_i and α_i^2 and the right-hand side shows the relation between the probability mass enclosed by the discretization frame with width w_i and α_i^2. 77

6.11 Posterior probability $\hat{P}_{F|\mathbf{M}}$ as a function of the number of intervals per random variable together with the exact (analytical) solution $P_{F|\mathbf{M}}$ for the fourth measurement case $[1.6, 2.0, 1.2]$ in Tab. 6.2. 78

7.1 A simple BN structure. 84

7.2 The general reliability BN, extended by explicitly including the model $g(\mathbf{x})$ as a node. 86

7.3 Relative error for the LSF in Eq. 7.8 averaged over 14 simulation runs (Note: The number of simulation runs is limited because of the applied software tools). The results are shown for $n = 2$ and $n = 30$ random variables X_i. 88

7.4 Coefficient of variation for the LSF in Eq. 7.8 estimated from 14 simulation runs . The results are shown for $n = 2$ and $n = 30$ random variables X_i. . . . 88

8.1 Runway overrun definitions 91

8.2 BN structure of the developed RWO warning system. 95

9.1 Decision tree for the crew failure mode critical data misperceived (Xing et al., 2013). The paths through the decision tree are numbered and for each path a probability was elicited from experts. E.g. the HEP for poor HSI/environment, high workload, poor training and no recovery potential is 0.56. (Note: The expert elicitation task has not been completed as of the writing of this report; some probabilities are listed as "unknown" and some may change in the final IDHEAS report.) 100
9.2 BN for the CFM critical data misperceived that corresponds to the original DT model. 101
9.3 Fully expanded BN for the CFM critical data misperceived. The black node represents the target variable; dark grey nodes the PIF variables; light grey nodes the PIF specification variables and white nodes additional variables illustrating the causal paths. The causal paths I to III are indicated through roman numerals. 102
9.4 Reduced BN for the CFM critical data misperceived. 105

List of Tables

2.1 Crew Failure Modes used in the IDHEAS method (Xing et al., 2013) 17
2.2 CFMs used in the PHOENIX method. 18

3.1 CPT of the node *Power generation* from the BN in Fig. 3.3, which represents the 2-out-of-3 gate from the FT in 3.2. 25
3.2 Summary of the local (conditional) probabilities for the BN in Fig. 3.13a. . . 35

4.1 CPT of a binary node X_i with binary parents $\{X_1, X_2, X_3\}$ 45
4.2 CPT of a multistate node X_i with discrete parents $[X_1, X_2, X_3, ..., X_n]$ 47

6.1 Parameters a and b of Eq. 6.8 for $\beta = 3.1$, $\beta = 4.3$ and $\beta = 5.2$ as well as 5, 10 and 20 intervals per dimension. 69
6.2 Evaluation of the discretization error for different measurement outcomes m, for problems with $n = 3$ random variables. a is the constant in the LSF, Eq. 6.9; ρ_{ij} is the correlation coefficient between X_i and X_j for all $i \neq j$; P_F and $P_{F|\mathbf{M}}$ denote the analytically calculated prior and posterior failure probabilities; $\hat{P}_{F|\mathbf{M}}$ is the conditional failure probability calculated with the discrete BN. . 71
6.3 Evaluation of the discretization error for different measurement outcomes m. The number of random variables $n = 4$; a is the constant in the LSF, Eq. 6.9; ρ_{ij} is the correlation coefficient between X_i and X_j for all $i \neq j$; P_F and $P_{F|M}$ denote the analytically calculated prior and posterior failure probabilities; $\hat{P}_{F|\mathbf{M}}$ is the conditional failure probability calculated with the discrete BN. . 71
6.4 Evaluation of the discretization error for different measurement outcomes m. The problems have $n = 2$,3 or 4 random variables; a is the constant in the LSF, Eq. 6.10; $\rho_i j$ is the correlation coefficient between X_i and X_j for all $i \neq j$; P_F and $P_{F|\mathbf{M}}$ denote the prior respectively posterior failure probabilities, which are calculated through importance sampling with 10^7 samples; $\hat{P}_{F|\mathbf{M}}$ is the conditional failure probability calculated with the discrete BN. Since there is no analytical solution, the updating of the basic random variables was performed through rejection sampling with more than 5e7 accepted samples. 73

7.1 CPT of the node X_1 in Fig. 7.1. 84

7.2 CPT of the node X_2 in Fig. 7.1. 85
7.3 CPT of the node X_3 in Fig. 7.1. 85

8.1 Basic random variables of the problem. Note: The mean value of the random variable *Touchdown point* is a deterministic function of other basic random variables (Zwirglmaier et al., 2014). 94
8.2 FORM importance measures α_i for each airport (AP) – aircraft (AC) combination and every basic random variable in the RWO application. 95
8.3 RWO probabilities, Pr(RWO), for the different airports and aircrafts calculated with the discrete BN, together with solutions calculated by importance sampling around the design point. The latter have a sampling error with coefficient of variation in the order of 10%. 97
8.4 Probabilities of RWO and corresponding decision on landing, computed with the BN for different sets of observations. 97

9.1 CPT of the node critical data misperceived. The HEPs corresponding to the grey cells are marked as unknown in Fig. 9.1. For that reason, the estimates for the scenario [HSI = poor, workload = low and training = poor] were used. This corresponds to a conservative approximation, since changing the state of training from good to poor will certainly increase the HEP. 106
9.2 Illustrative probabilities quantifying the CPTs of the PIFs. The probabilities are based on (Groth and Swiler, 2013). 107
9.3 Deterministic CPT of training given the two corresponding PIF specification nodes. 108
9.4 Results from the survey, carried out to elicit prior probabilities for the PIF specification nodes. Experts I and II are HRA specialists with a background in cognitive psychology and expert III is a former operator of a nuclear power plant on a submarine. The elicited numbers are probability estimates for the PIF specification nodes being in state “yes”. 109
9.5 SACADA indicators, which can be related to PIF specification nodes in IDHEAS. Factors used to relate the two are given in the last column. 112
9.6 HEPs for different observations. 113

A.1 CPT of the vegetation node in Fig. 3.13a 127

Bibliography

P. A. Aguilera, A. Fernández, R. Fernández, R. Rumí, and A. Salmerón. Bayesian networks in environmental modelling. *Environmental Modelling and Software*, 26(12):1376–1388, 2011.

H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974.

B. Ale, L. J. Bellamy, R. Cooke, M. Duyvis, D. Kurowicka, P. H. Lin, O. Morales, A. Roelen, and J. Spouge. Causal model for air transport safety. Report, 2008.

B. J. M. Ale, L. J. Bellamy, R. van der Boom, J. Cooper, R. M. Cooke, L. H. J. Goossens, A. R. Hale, D. Kurowicka, O. Morales, A. L. C. Roelen, and J. Spouge. Further development of a causal model for air transport safety (cats): Building the mathematical heart. *Reliability Engineering & System Safety*, 94(9):1433–1441, 2009.

H. Andersen, J. Casal, A. Dandrieux, D. B, V. de Dianous, N. Duijm, C. Delvosalle, C. Fievez, L. Goossens, R. Gowland, A. Hale, D. Hourtolou, B. Mazzarotta, A. Pipart, E. Planas, F. Prats, O. Salvi, and J. Tixier. Aramis user guide. Report, European Comission Joint Research Centre Institute for the Protection and Security of the Citizen, 2004.

S. Arnborg, D. G. Corneil, and A. Proskurowski. Complexity of finding embeddings in ak-tree. *SIAM Journal on Algebraic Discrete Methods*, 8(2):277–284, 1987.

S.-K. Au and J. L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277, 2001.

P. Baraldi, L. Podofillini, L. Mkrtchyan, E. Zio, and V. N. Dang. Comparing the treatment of uncertainty in bayesian networks and fuzzy expert systems used for a human reliability analysis application. *Reliability Engineering & System Safety*, 138:176–193, 2015.

D. N. Barton, S. Kuikka, O. Varis, L. Uusitalo, H. J. Henriksen, M. Borsuk, A. de la Hera, R. Farmani, S. Johnson, and J. D. Linnell. Bayesian networks in environmental and resource management. *Integrated environmental assessment and management*, 8(3):418–429, 2012.

Y. Y. Bayraktarli and M. H. Faber. Bayesian probabilistic network approach for managing earthquake risks of cities. *Georisk*, 5(1):2–24, 2011.

G. Bearfield and W. Marsh. *Generalising event trees using Bayesian networks with a case study of train derailment*, pages 52–66. Springer, 2005.

J. L. Beck and S.-K. Au. Bayesian updating of structural models and reliability using markov chain monte carlo simulation. *Journal of Engineering Mechanics*, 128(4):380–391, 2002.

T. Bedford and R. Cooke. *Probabilistic risk analysis: foundations and methods. 2001.* Cambridge: Cambridge University Press, 2001.

J. Benjamin and C. Cornell. *Probability, Statistics, and Decision for Civil Engineers*. McGraw-Hill, 1970.

M. Bensi, A. Der Kiureghian, and D. Straub. Bayesian network modeling of correlated random variables drawn from a gaussian random field. *Structural Safety*, 33(6):317–332, 2011.

M. Bensi, A. Der Kiureghian, and D. Straub. Efficient bayesian network modelling of systems. *Reliability Engineering & System Safety*, 112:200–213, 2013.

M. Bensi, A. D. Kiureghian, and D. Straub. Framework for post-earthquake risk assessment and decision making for infrastructure systems. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 1(1), 2014.

I. Biederman. Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2):115, 1987.

C. M. Bishop. *Pattern recognition and machine learning*. springer, 2006.

L. Blaser, M. Ohrnberger, C. Riggelsen, and F. Scherbaum. *Bayesian belief network for tsunami warning decision support*, pages 757–768. Springer, 2009.

A. Bobbio, L. Portinale, M. Minichino, and E. Ciancamerla. Improving the analysis of dependable systems by mapping fault trees into bayesian networks. *Reliability Engineering & System Safety*, 71(3):249–260, 2001.

R. L. Boring, C. D. Griffith, and J. C. Joe. The measure of human error: Direct and indirect performance shaping factors. In *Human Factors and Power Plants and HPRCT 13th Annual Meeting, 2007 IEEE 8th*, pages 170–176, 2007.

H. Boudali and J. B. Dugan. A discrete-time bayesian network reliability modeling and analysis framework. *Reliability Engineering & System Safety*, 87(3):337–349, 2005.

D. Broadbent. *Perception and communication*. Pergamon Press, 1958.

S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.

R. Castelo and T. Kocka. On inclusion-driven learning of bayesian networks. *The Journal of Machine Learning Research*, 4:527–574, 2003.

J. Y. Chang, D. Bley, L. Criscione, B. Kirwan, A. Mosleh, T. Madary, R. Nowell, R. Richards, E. M. Roth, S. Sieben, and A. Zoulis. The sacada database for human reliability and human performance. *Reliability Engineering & System Safety*, 125(0):117–133, 2014.

S. H. Chen and C. A. Pollino. Good practice in bayesian network modelling. *Environmental Modelling and Software*, 37(0):134–145, 2012.

D. M. Chickering. Learning equivalence classes of bayesian-network structures. *The Journal of Machine Learning Research*, 2:445–498, 2002. ISSN 1532-4435.

S. Conrady and L. Jouffe. *Bayesian Networks and BayesiaLab: A Practical Introduction for Researchers*. Bayesia USA, Franklin, TN, 2015.

R. M. Cooke. Experts in uncertainty: opinion and subjective probability in science. 1991.

S. Cooper, A. Ramey-Smith, J. Wreathall, G. Parry, D. Bley, W. Luckas, J. Taylor, and M. Barriere. A technique for human error analysis (atheana). Report NUREG/CR-6350, 1996.

M. K. Cowles and B. P. Carlin. Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434):883–904, 1996.

T. A. Cruse. *Reliability-based mechanical design*, volume 108. CRC Press, 1997.

CSNI. Simulator studies for hra purposes. Report, Committee on the Safety of Nuclear Installations, OECD Nuclear Energy Agency, 2012.

I. C. Cárdenas, S. S. H. Al-jibouri, J. I. M. Halman, and F. A. van Tol. Capturing and integrating knowledge for managing risks in tunnel works. *Risk Analysis*, 33(1):92–108, 2013.

F. De Carlo, O. Borgia, and M. Tucci. Imperfect maintenance modelling by dynamic object oriented bayesian networks. *International Journal of Engineering and Technology*, 5(5): 4282–4295, 2013.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

A. Der Kiureghian. *First- and Second-Order Reliability Methods*. CRC PressINC, 2005.

O. Doguc and J. E. Ramirez-Marquez. A generic method for estimating system reliability using bayesian networks. *Reliability Engineering & System Safety*, 94(2):542–550, 2009.

L. Drees and F. Holzapfel. *Determining and Quantifying Hazard Chains and their Contribution to Incident Probabilities in Flight Operation.* Guidance, Navigation, and Control and Co-located Conferences. American Institute of Aeronautics and Astronautics, 2012.

N. J. Duijm. Safety-barrier diagrams as a safety management tool. *Reliability Engineering & System Safety*, 94(2):332–341, 2009.

H. J. Einhorn and R. M. Hogarth. Behavioral decision theory: Processes of judgment and choice. *Journal of Accounting Research*, pages 1–31, 1981.

N. J. Ekanem and A. Mosleh. Phoenix–a model-based human reliability analysis methodology: Quantitative analysis procedure and data base. In *Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 12), Hawaii, USA*, 2014.

N. J. Ekanem, A. Mosleh, and S.-H. Shen. Phoenix – a model-based human reliability analysis methodology: Qualitative analysis procedure. *Reliability Engineering & System Safety*, 145: 301–315, 2016.

M. Endsley. Towards a theory of situation awareness in dynamic systems. *Human Factors*, 37(1):32–64, 1995.

C. Eriksen and J. St. James. Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics*, 40(4):225–240, 1986.

N. Fenton and M. Neil. The jury observation fallacy and the use of bayesian networks to present probabilistic legal arguments. *Math. Today, Southend-on-Sea*, 36(6):180–187, 2000.

N. Fenton and M. Neil. *Risk Assessment and Decision Analysis with Bayesian Networks.* CRC Press, 2012.

P. Franchin, A. Lupoi, F. Noto, and S. Tesfamariam. Seismic fragility of reinforced concrete girder bridges using bayesian belief network. *Earthquake Engineering & Structural Dynamics*, 45(1):29–44, 2016.

S. French, T. Bedford, S. J. T. Pollard, and E. Soane. Human reliability analysis: A critique and review for managers. *Safety Science*, 49(6):753–763, 2011.

N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3):131–163, 1997.

A. Friis-Hansen. *Bayesian Networks as a Decision Support Tool in Marine Applications.* Thesis, 2000.

R. Fung and K.-C. Chang. Weighing and integrating evidence for stochastic simulation in bayesian networks. In *5th Annual Conference on Uncertainty in Artificial Intelligence*, 1989.

P. Gehl and D. D'Ayala. Development of bayesian networks for the multi-hazard fragility assessment of bridge systems. *Structural Safety*, 60:37–46, 2016.

A. E. Gelfand and A. F. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian data analysis*, volume 3. Chapman & Hall/CRC Boca Raton, FL, USA, 2013.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741, 1984.

D. Gertman, H. Blackman, J. Marble, J. Byers, and C. Smith. The spar-h human reliability analysis method. Report, Idaho National Laboratory, 2005.

W. Gilks, S. Richardson, and D. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, softcover reprint of the original 1st ed. 1996 edition, 1996.

W. R. Gilks and P. Wild. Adaptive rejection sampling for gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(2):337–348, 1992.

W. R. Gilks, A. Thomas, and D. J. Spiegelhalter. A language and program for complex bayesian modelling. *The Statistician*, pages 169–177, 1994.

W. R. Gilks, N. G. Best, and K. K. C. Tan. Adaptive rejection metropolis sampling within gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44 (4):455–472, 1995.

P. J. Green and D. I. Hastie. Reversible jump mcmc. *Genetics*, 155(3):1391–1403, 2009.

K. Groth and A. Mosleh. Deriving causal bayesian networks from human reliability analysis data: A methodology and example model. *Journal of Risk and Reliability*, 226(4):361–379, 2012.

K. M. Groth and L. P. Swiler. Bridging the gap between hra research and hra practice: A bayesian network version of spar-h. *Reliability Engineering & System Safety*, 115(0):33–42, 2013.

K. M. Groth, C. L. Smith, and L. P. Swiler. A bayesian method for using simulator data to enhance human error probabilities assigned by existing hra methods. *Reliability Engineering & System Safety*, 128(0):32–40, 2014.

A. Grêt-Regamey and D. Straub. Spatially explicit avalanche risk assessment linking bayesian networks to a gis. *Natural Hazards and Earth System Science*, 6(6):911–926, 2006.

B. Hallbert and A. Kolaczkowski. The employment of empirical data and bayesian methods in human reliability analysis: A feasibility study. Report, 2007.

A. Hanea, R. Hanea, and A. Zilko. Dynamic bayesian networks as a possible alternative to the ensemble kalman filter for parameter estimation in reservoir engineering. In *11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference 2012, PSAM11*, 2012.

A. M. Hanea, D. Kurowicka, and R. M. Cooke. Hybrid method for quantifying and analyzing bayesian belief nets. *Quality and Reliability Engineering International*, 22(6):709–729, 2006.

D. Hanea and B. Ale. Risk of human fatality in building fires: A decision tool using bayesian networks. *Fire Safety Journal*, 44(5):704–710, 2009.

W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

D. Heckerman. Causal independence for knowledge acquisition and inference, 1993.

D. Heckerman and J. S. Breese. A new look at causal independence, 1994.

D. Heckerman, D. Geiger, and D. M. Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20(3):197–243, 1995.

M. Henrion. Propagation of uncertainty by probabilistic logic sampling in bayes' networks. In *Uncertainty in Artificial Intelligence*, volume 2, pages 149–164, 1988.

M. Hohenbichler and R. Rackwitz. Nonnormal dependent vectors in structural safety. *Journal of Engineering Mechanics*, 107(6):1227–1238, 1981.

E. Hollnagel. Looking for errors of omission and commission or the hunting of the snark revisited. *Reliability Engineering & System Safety*, 68(2):135–145, 2000.

IATA. Safety report 2012. Report, International Air Transport Association (IATA), 2013.

F. V. Jensen and T. D. Nielsen. *Bayesian networks and decision graphs*. Springer, New York [u.a.], 2. ed., softcover reprint edition, 2007.

F. V. Jensen, S. L. Lauritzen, and K. G. Olesen. Bayesian updating in causal probabilistic networks by local computations. *Computational statistics quarterly*, 4:269–282, 1990.

C.-G. Jong and S.-S. Leu. Bayesian-network-based hydro-power fault diagnosis system devepolment by fault tree transformation. *Journal of Marine Science and Technology*, 21(4): 367–379, 2013.

W. S. Jäger, C. K. den Heijer, A. Bolle, and A. M. Hanea. A bayesian network approach to coastal storm impact modeling. 2015.

D. L. Kelly and C. L. Smith. Bayesian inference in probabilistic risk assessment—the current state of the art. *Reliability Engineering & System Safety*, 94(2):628–643, 2009.

N. Khakzad, F. Khan, and P. Amyotte. Safety analysis in process facilities: Comparison of fault tree and bayesian network approaches. *Reliability Engineering & System Safety*, 96 (8):925–932, 2011.

N. Khakzad, F. Khan, and P. Amyotte. Quantitative risk analysis of offshore drilling operations: A bayesian approach. *Safety science*, 57:108–117, 2013.

M. C. Kim, P. H. Seong, and E. Hollnagel. A probabilistic approach for determining the control mode in cream. *Reliability Engineering & System Safety*, 91(2):191–199, 2006.

L. W. King. The code of hammurabi (translated by l.w. king), 2008. URL `http://avalon.law.yale.edu/ancient/hamframe.asp`.

B. Kirwan. *A guide to practical human reliability assessment.* CRC press, 1994.

U. Kjærulff. Triangulation of graphs – algorithms giving small total state space. Technical report, 1990.

U. B. Kjærulff and A. L. Madsen. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis.* Springer Publishing Company, Incorporated, 2nd edition, 2013.

G. A. Klein. *Sources of power: How people make decisions.* MIT press, 1998.

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques (Adaptive Computation and Machine Learning series).* The MIT Press, 2009.

D. Koller, N. Friedman, L. Getoor, and B. Taskar. 2 graphical models in a nutshell. *Statistical Relational Learning*, page 13, 2007.

N. T. Kottegoda and R. Rosso. Probability, statistics, and reliability for civil and environmental engineers. 1997.

N. T. Kottegoda and R. Rosso. *Applied statistics for civil and environmental engineers.* Blackwell Malden, MA, 2008.

N. M. Kuehn. *Empirical ground-motion models for probabilistic seismic hazard analysis: a graphical model perspective.* Phd thesis, 2010.

N. M. Kuehn, C. Riggelsen, and F. Scherbaum. Modeling the joint probability of earthquake, site, and ground-motion parameters using bayesian networks. *Bulletin of the Seismological Society of America*, 101(1):235–249, 2011.

A. Kusz, P. Maksym, and A. W. Marciniak. Bayesian networks as knowledge representation system in domain of reliability engineering. *Teka Komisji Motoryzacji i Energetyki Rolnictwa*, 11, 2011. ISSN 1641-7739.

W. Köhler. *Gestalt psychology: An introduction to new concepts in modern psychology.* WW Norton and Company, 1970.

H. Langseth and L. Portinale. Bayesian networks in reliability. *Reliability Engineering & System Safety*, 92(1):92–108, 2007.

H. Langseth, T. D. Nielsen, R. Rumí, and A. Salmerón. Inference in hybrid bayesian networks. *Reliability Engineering & System Safety*, 94(10):1499–1509, 2009.

H. Langseth, T. D. Nielsen, R. Rumı, and A. Salmerón. Mixtures of truncated basis functions. *International Journal of Approximate Reasoning*, 53(2):212–227, 2012.

S. L. Lauritzen. Propagation of probabilities, means, and variances in mixed graphical association models. *Journal of the American Statistical Association*, 87(420):1098–1108, 1992.

S. L. Lauritzen. The em algorithm for graphical association models with missing data. *Computational Statistics and Data Analysis*, 19(2):191–201, 1995.

S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and thier application to expert systems. *J. R. Statist. Soc. B*, 50(2):157–224, 1988.

E. Lois, V. N. Dang, J. Forester, H. Broberg, S. Massaiu, M. Hildebrandt, P. Braarud, G. Parry, J. Julius, R. L. Boring, I. Männistö, and A. Bye. International hra empirical study - phase 1 report: Description of overall approach and pilot phase results from comparing hra methods to simulator performance data. Report, US Nuclear Regulatory Commission, 2009.

P. Lucas. *Bayesian networks in medicine: a model-based approach to medical decision making.* na, 2001.

D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The bugs project: Evolution, critique and future directions. *Statistics in medicine*, 28(25):3049, 2009.

J. Luque and D. Straub. Reliability assessment of monitored deteriorating systems with dynamic bayesian networks. *Manuscript, submitted to Structural Safety*, 2015.

M. R. Lyu et al. *Handbook of software reliability engineering*, volume 222. IEEE computer society press CA, 1996.

A. Léger, R. Farret, C. Duval, E. Levrat, P. Weber, and B. Iung. A safety barriers-based approach for the risk analysis of socio-technical systems. In *17th IFAC World Congress*, 2008.

S. Mahadevan, R. Zhang, and N. Smith. Bayesian networks for system reliability reassessment. *Structural Safety*, 23(3):231–251, 2001.

D. Marquez, M. Neil, and N. Fenton. Improved reliability modeling using bayesian networks and dynamic discretization. *Reliability Engineering & System Safety*, 95(4):412–425, 2010.

W. Marsh and G. Bearfield. Representing parameterised fault trees using bayesian networks. In *International Conference on Computer Safety, Reliability, and Security*, pages 120–133. Springer, 2007.

L. Martino, H. Yang, D. Luengo, J. Kanniainen, and J. Corander. A fast universal self-tuned sampler within gibbs sampling. *Digital Signal Processing*, 47:68 – 83, 2015. Special Issue in Honour of William J. (Bill) Fitzgerald.

M. R. Martins, A. M. Schleder, and E. L. Droguett. A methodology for risk analysis based on hybrid bayesian networks: Application to the regasification system of liquefied natural gas onboard a floating storage and regasification unit. *Risk Analysis*, 34(12):2098–2120, 2014.

D. B. McCafferty. Successful system design through integrating engineering and human factors. *Process Safety Progress*, 14(2):147–151, 1995.

R. E. Melchers. *Structural reliability analysis and prediction*. John Wiley & Son Ltd, 1999.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6): 1087–1092, 1953.

L. Mkrtchyan, L. Podofillini, and V. N. Dang. Bayesian belief networks for human reliability analysis: a review of applications and gaps. *Reliability Engineering & System Safety*, (accepted manuscript), 2015.

S. Montani, L. Portinale, A. Bobbio, and D. Codetta-Raiteri. Radyban: A tool for reliability analysis of dynamic fault trees through conversion into dynamic bayesian networks. *Reliability Engineering & System Safety*, 93(7):922–932, 2008.

O. Morales. *Bayesian Belief Nets and Vines in Aviation Safety and other Applications*. Thesis, 2010.

A. Mosleh and Y. H. Chang. Model-based human reliability analysis: prospects and requirements. *Reliability Engineering & System Safety*, 83(2):241–253, 2004.

K. Murphy and M. Mahdaviani. Matbugs software. `https://code.google.com/archive/p/matbugs/`, 2005.

M. Musharraf, D. Bradbury-Squires, F. Khan, B. Veitch, S. MacKinnon, and S. Imtiaz. A virtual experimental technique for data collection for a bayesian network approach to human reliability analysis. *Reliability Engineering & System Safety*, 132:1–8, 2014.

S. Nadkarni and P. P. Shenoy. A causal mapping approach to constructing bayesian networks. *Decision Support Systems*, 38(2):259–281, 2004.

R. E. Neapolitan. Learning bayesian networks. 2004.

M. Neil, N. Fenton, and L. Nielson. Building large-scale bayesian networks. *The Knowledge Engineering Review*, 15(03):257–284, 2000.

M. Neil, M. Tailor, D. Marquez, N. Fenton, and P. Hearty. Modelling dependable systems using hybrid bayesian networks. *Reliability Engineering & System Safety*, 93(7):933–939, 2008.

J. J. Nielsen and J. D. Sørensen. Bayesian networks as a decision tool for o& m of offshore wind turbines. In *Proceedings of the 5th international ASRANet conference*, 2010.

U. NRC. Wash-1400: Reactor safety study. Report, 1975.

O. Nývlt and L. Ferkl. Stochastic coloured petri nets as a modelling language for complex event trees, 2013.

A. O'Hagan, C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow. *Uncertain judgements: eliciting experts' probabilities*. John Wiley and Sons, 2006.

I. Papaioannou, W. Betz, K. Zwirglmaier, and D. Straub. Mcmc algorithms for subset simulation. *Probabilistic Engineering Mechanics*, 41:89–103, 2015. ISSN 0266-8920.

P. Papakosta and D. Straub. Probabilistic prediction of fire occurrence in the mediterranean, 2013.

J. Park and W. Jung. Opera—a human performance database under simulated emergencies of nuclear power plants. *Reliability Engineering & System Safety*, 92(4):503–519, 2007.

J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Kaufmann, San Francisco, CA, 1988.

J. Pearl. *Causality*. Cambridge university press, 2009.

L. Perelman and A. Ostfeld. Bayesian networks for source intrusion detection. *Journal of Water Resources Planning and Management*, 139(4):426–432, 2012.

L. Podofillini and V. N. Dang. Aggregating expert-elicited error probabilities to build hra models, 2014.

R. Rackwitz. Reliability analysis—a review and some perspectives. *Structural Safety*, 23(4): 365–395, 2001.

H. Raiffa and R. Schlaifer. *Applied statistical decision theory*. Studies in managerial economics. Division of Research, Graduate School of Business Adminitration, Harvard University, Boston, 1961.

M. Rausand and A. Høyland. *System reliability theory: models, statistical methods, and applications*, volume 396. John Wiley & Sons, 2004. ISBN 047147133X.

C. Riggelsen. Learning bayesian networks: A map criterion for joint selection of model structure and parameter, 2008.

C. P. Robert. *Approximate Bayesian Computation: A Survey on Recent Results*, pages 185–205. Springer International Publishing, Cham, 2016.

R. W. Robinson. *Counting unlabeled acyclic digraphs*, pages 28–43. Springer, 1977.

K. Roscoe and A. Hanea. Bayesian networks in levee system reliability. 2015.

A. M. Rothblum. Human error and marine safety. In *National Safety Council Congress and Expo, Orlando, FL*, 2000.

A. Saltelli, K. Chan, and E. M. Scott. *Sensitivity analysis*, volume 1. Wiley New York, 2000.

R. Schneider. Looking across industries to improve human reliability data for quantitative risk analyses. *Process Safety Progress*, 29(2):123–126, 2010.

G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

R. Shachter. Evaluating influence diagrams. *Oper. Res.*, 34(6):871–882, 1986.

R. Shachter. Probabilistic inference and influence diagrams. *Oper. Res.*, 36(4):589–604, 1988.

R. Shachter and M. Peot. Simulation approaches to general probabilistic inference on belief networks. In *5th Annual Conference on Uncertainty in Artificial Intelligence*, pages 221–234, 1990.

R. D. Shachter and C. R. Kenley. Gaussian influence diagrams. *Management science*, 35(5): 527–550, 1989. ISSN 0025-1909.

S. A. Shappell and D. A. Wiegmann. *A human error approach to aviation accident analysis: The human factors analysis and classification system.* Ashgate Publishing, Ltd., 2012.

P. D. Spanos and R. Ghanem. Stochastic finite element expansion for random media. *Journal of engineering mechanics*, 115(5):1035–1053, 1989. ISSN 0733-9399.

O. Spačková and D. Straub. Dynamic bayesian network for probabilistic modeling of tunnel excavation processes. *Computer-Aided Civil and Infrastructure Engineering*, 28(1):1–21, 2013.

P. Spirtes, C. N. Glymour, and R. Scheines. Causality from probability. 1989.

P. Spirtes, C. N. Glymour, and R. Scheines. *Causation, prediction, and search.* MIT Press, Cambridge. Mass. [u.a.], 2. ed. edition, 2001.

H. Steck. *Constraint based structural learning in Bayesian networks using finite data sets.* PhD Thesis, TU München, 2001. München, Techn. Univ., Diss., 2001.

J. W. Stetkar. Human reliability analysis models, 2014.

M. Stewart and R. E. Melchers. *Probabilistic Risk Assessment Of Engineering Systems.* Springer, 1998.

D. Straub. Stochastic modeling of deterioration processes through dynamic bayesian networks. *Journal of Engineering Mechanics*, 135(10):1089–1099, 2009.

D. Straub. Reliability updating with equality information. *Probabilistic Engineering Mechanics*, 26(2):254–258, 2011.

D. Straub. Value of information analysis with structural reliability methods. *Structural Safety*, 49(0):75–85, 2014a.

D. Straub. Engineering risk assessment. In C. Klüppelberg, D. Straub, and I. Welpe, editors, *Risk – A Multidisciplinary Introduction*, volume 1. Springer, 2014b.

D. Straub and A. Der Kiureghian. Bayesian network enhanced with structural reliability methods: Methodology. *Journal of Engineering Mechanics*, 136(10):1248–1258, 2010a.

D. Straub and A. Der Kiureghian. Bayesian network enhanced with structural reliability methods: Application. *Journal of Engineering Mechanics*, 136(10):1259–1270, 2010b.

D. Straub and I. Papaioannou. Bayesian updating with structural reliability methods. *Journal of Engineering Mechanics*, 141(3):04014134, 2014.

O. Sträter. Considerations on the elements of quantifying human reliability. *Reliability Engineering & System Safety*, 83(2):255–264, 2004.

S. Suddle. The risk management of third parties during construction in multifunctional urban locations. *Risk Analysis*, 29(7):1024–1040, 2009.

R. Sundaramurthi and C. Smidts. Human reliability modeling for the next generation system code. *Annals of Nuclear Energy*, 52:137–156, 2013.

A. Swain and H. Guttmann. Handbook of human reliability analysis with emphasis on nuclear power plant applications. Report, U.S. Nuclear Regulatory Commission, 1983.

M. Teyssier and D. Koller. Ordering-based search: A simple and effective algorithm for learning bayesian networks. In *Proc. of the 12th Conference on Uncertainty in Artificial Intelligence (UAI-05)*, 2012.

I. Tien and A. Der Kiureghian. Algorithms for bayesian network modeling and reliability assessment of infrastructure systems. *Reliability Engineering & System Safety*, 156:134–147, 2016.

J. G. Torres-Toledano and L. E. Sucar. *Bayesian networks for reliability analysis of complex systems*, pages 195–206. Springer, 1998.

P. Trucco, E. Cagno, F. Ruggeri, and O. Grande. A bayesian belief network modelling of organisational factors in risk analysis: A case study in maritime transportation. *Reliability Engineering & System Safety*, 93(6):845–856, 2008.

A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974.

J. R. van Dorp and T. A. Mazzuchi. *Parameter specification of the beta distribution and its Dirichlet extensions utilizing quantiles*, pages 283–318. CRC press, 2004.

E. Vanmarcke. *Random fields: analysis and synthesis*. World Scientific, 2010.

K. Vogel. *Applications of Bayesian networks in natural hazard assessments*. Phd thesis, 2014.

N. Warner and M. Letsky. Empirical model of team collaboration focus on macrocognition. *Macrocognition in teams*, pages 15–33, 2008.

P. Weber and L. Jouffe. Complex system reliability modelling with dynamic object oriented bayesian networks (doobn). *Reliability Engineering & System Safety*, 91(2):149–162, 2006.

A. Whaley, J. Xing, R. L. Boring, S. Hendrickson, J. Joe, and K. Le Blanc. Building a psychological foundation for human reliability analysis (nureg-2114). Report, US Nuclear Regulatory Commission, 2012.

D. D. Woods. On taking human performance seriously in risk analysis: Comments on dougherty. *Reliability Engineering & System Safety*, 29(3):375–381, 1990.

J. Xing, M. Presley, G. Parry, J. Forester, S. Hendrickson, and V. Dang. Nrc/epri draft report for peer review: An integrated decision-tree human event analysis system (idheas) method for npp internal at-power operation. Report, NRC/EPRI, 2013.

K. Zachmann. Risk in historical perspective: Concepts, contexts, and conjunctions. In C. Klüppelberg, D. Straub, and I. M. Welpe, editors, *Risk - A Multidisciplinary Introduction*. Springer, 2014.

R. E. Zapata-Vázquez, A. O'Hagan, and L. Soares Bastos. Eliciting expert judgements about a set of proportions. *Journal of Applied Statistics*, 41(9):1919–1933, 2014.

N. L. Zhang and D. Poole. A simple approach to bayesian network computations. In *Proc. of the Tenth Canadian Conference on Artificial Intelligence*, 1994.

J. Zhu and M. Collette. A dynamic discretization method for reliability inference in dynamic bayesian networks. *Reliability Engineering & System Safety*, 138(0):242–252, 2015.

E. Zio. Reliability engineering: Old problems and new challenges. *Reliability Engineering & System Safety*, 94(2):125 – 141, 2009.

K. Zwirglmaier and D. Straub. Discretization of structural reliability problems – an application to runway overrun. In *Proc. ICASP 2015, Vancouver, Canada*, 2015.

K. Zwirglmaier and D. Straub. A discretization procedure for rare events in bayesian networks. *Reliability Engineering & System Safety*, 153:96 – 109, 2016a.

K. Zwirglmaier and D. Straub. Structure elicitation for bayesian networks. *Journal paper manuscript*, 2016b.

K. Zwirglmaier, L. Drees, F. Holzapfel, and D. Straub. Reliability analysis for runway overrun using subset simulation. In *Proc. ESREL 2014, Wrocław, Poland*, 2014.

K. Zwirglmaier, D. Straub, and K. Groth. Capturing cognitive causal paths in human reliability analysis with bayesian network models. *Reliability Engineering & System Safety*, 158: 117–129, 2017.