Project Reference: 608166

FP7-SEC-2013-1 Impact of extreme weather on critical infrastructure

Project Duration: 1 May 2014 – 30 April 2017

RAIN PROJECT

Security Sensitivity Committee Deliverable Evaluation

Deliverable Reference	**D 5.1**
Deliverable Name	**Risk Analysis framework for single and multiple hazards**
Contributing Partners	**TU Delft**
Date of Submission	**November 2015**

The evaluation is:

- **The content is not related to general project management**
- **The content is not related to general outcomes as dissemination and communication**
- **The content is not related to critical infrastructure vulnerability or sensitivity**

Diagram path 1-2-3. Therefor the evaluation is Public.

Decision of Evaluation	**Public**	**~~Confidential~~**
	~~Restricted~~	

Evaluator Name	**P.L. Prak, MSSM**
Evaluator Signature	**Signed by the chairman**
Date of Evaluation	**2015-11-04**

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608166

This project is funded by the European Union

RAIN – Risk Analysis of Infrastructure Networks in Response to Extreme Weather

Project Reference: 608166

FP7-SEC-2013-1 Impact of extreme weather on critical infrastructure

Project Duration: 1 May 2014 – 30 April 2017

Deliverable 5.1-Risk Analysis framework for single and multiple hazards

Authors

Noel van Erp (TU Delft)

Pieter van Gelder* (TU Delft)

***Corresponding author: p.h.a.j.m.vangelder@tudelft.nl**

Date: 30/09/2015

Dissemination level: (PU, PP, RE, CO): PU

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608166

This project is funded by the European Union

DOCUMENT HISTORY

Index	Date	Author(s)	Main modifications
E01	8-10-2015	NVE and PVG	First Draft
	25-10-2015	NVE and PVG	Second Draft
	30-10-2015	FMI, HI and ROD	Review comments
	4-11-2015	NVE and PVG	Final version

Document Name: Risk analysis framework for single and multiple hazards

Work Package: 5

Task: 5.1

Deliverable: 5.1

Deliverable scheduled date (18th Month) 30th October 2015

Responsible Partner: TU Delft

Table of Contents

Executive Summary

In this deliverable a quantitative risk analysis framework for single and multiple hazards is provided. The quantitative risk analysis framework admits two phases: (1) the inference phase, and (2) the decision phase.

It is recommended in this deliverable to use for the inference phase of our risk analysis framework all of the tools that are provided by the Bayesian probability theory. We come to this recommendation after having applied an objective ranking session (Chapter 2) and after having done research into the alternative and initially proposed Bayesian network methodology (Chapter 3). The research into the Bayesian network methodology is discussed in this deliverable in Chapter 4 and the Appendices A and B.

It is recommended in this deliverable to use the tools of the decision theory for the decision phase of our risk analysis framework. In this deliverable two contributions to the decision theory are presented. The first contribution is an alternative for the traditional criterion of choice where risk is equated with the expectation value of the possible losses; see Chapter 5. The second contribution is a consistency derivation of Bernoulli's utility function, which – for those of us who are familiar with consistency derivation – may remove the degree of freedom of which utility function to use in order to map our outcomes to utilities; see Appendix D.

In Chapter 6 the proposed alternative criterion of choice is compared with the expected outcome theory and the expected utility theory, by way of a simple toy-problem. This toy-problem is used to demonstrate the intuitive inadequacy of the criterion of choice in which one maximizes the expectation values of the probability distributions of either the monetary outcomes or their corresponding utilities.

In Chapter 7 the miscellaneous 'techniques' of (1) influence diagrams, (2) Bayesian networks, (3) event trees, and (4) GIS mappings are related to the here proposed quantitative risk analysis framework.

1. That Which Has Been Delivered

Now, in order to discuss a quantitative risk assessment framework, one first has to define both the concepts of 'risk' and 'framework'. For if these concepts are ill-defined then the concept of a 'risk assessment framework' itself becomes ambiguous. So, for clarity's sake, we will outline here what we understand in this deliverable by the terms 'risk' and 'framework'.

1.1. What Is Risk?

In the first paragraph of the deliverable 6.1 risk is defined as 'the probability of losses'. In this deliverable, however, we understand risk to be some measure on a probability distribution of (net) outcomes; where negative (net) outcomes are understood to be losses. Stated differently, the probability distribution of the losses itself is not the risk, rather risk is some scalar which is associated with that probability distribution.

For example, a traditional risk measure is the expectation value of the outcome probability distributions (Bernoulli, 1738; Jaynes, 2003):

$$E(X)=\sum_{i=1}^{n} p_i x_i \ ,$$

where X is some stochastic of net outcomes $\{x_1,\dots x_n\}$ having probabilities $\{p_1,\dots p_n\}$ of occurring. The expectation value is not necessarily a value one would 'expect', but it is a probabilistic expectation in that it gives the centre of mass of a given probability distribution.

In this deliverable we will offer up an alternative risk measure which takes into account both the probabilistic worst and best case scenarios, as well as the probabilistic expected scenario; see Section 4.2. This alternative risk measure is the primary highlight of this research into risk; see Section 4.3 for a historical overview and Chapter 5 for an application of the alternative risk measure and a demonstration of its non-trivial implications.

1.2. A Refinement of the Risk Definition

At the beginning of this section we defined risk as some measure on a probability distribution of outcomes, we now refine this definition slightly in that we define risk as a measure on a probability distribution of utilities, rather than a probability distribution of outcomes.

Utilities are a measure of distance between our initial and final asset positions. By way of Bernoulli's utility function losses may be translated to negative utilities and gains to positive utilities. By mapping the outcomes of the outcome probability distribution to their corresponding utilities, the utility probability distributions are obtained. Utilities capture the risk averseness as a function of our initial wealth and the increments in this initial wealth, or, equivalently, the outcomes.

Stated differently, where in the previous section the risk was a function of outcomes and probabilities, there we propose in this section – as a refinement – that risk is a function of outcomes, probabilities, and the initial wealth of the decision maker; 4.1. The secondary highlight of this research into risk is the consistency derivation of Bernoulli's utility function, as this derivation effectively removes the degree of freedom as to what function to use in order to map outcomes to utilities[1]; see Appendix D.

1.3. Framework vs. Implementation

In this deliverable we will make a distinction between a 'framework', on the one hand, and an 'implementation', on the other hand. A framework is that which is general and of which there is only one. An implementation is that which is specific and of which there may be legion.

1.3.1. The framework

In the quantitative risk assessment framework proposed in this deliverable we have that (1) in the inference phase outcome probability distributions are constructed, one for each possible action which is contemplated, then (2) the decision making phase that action that promises to be the most profitable is chosen, by way of a comparison of the computed risk measures on the corresponding utility probability distributions; see Figure 2.1.

In this deliverable it is argued that for the inference phase Bayesian probability theory should be used as the general inference 'toolbox'. For it may be shown that the quantitative methodologies of event trees, fault trees, and Bayesian Networks are all specific applications of Bayesian probability theory. The demonstration that Bayesian Networks are a special, more restricted instance of the more general Bayesian probability theory constitutes a tertiary highlight of this research into risk; see Section 3.3 and Appendices A and B.

Furthermore, in this deliverable it is taken as self-evident that for the making of decisions based on outcome probability distributions is the subject matter of decision theory. Note that the first two highlights of this research into risk consists of non-trivial contributions to the field of decision theory; see Sections 1.1 and 1.2.

1.3.2. An implementation

As stated above, for the inference phase of the here proposed framework it is recommended in this deliverable to use Bayesian probability theory. But this deliverable stays mute on the question which particular probability model to use in order to come to our outcome probability distributions. Stated differently, the construction of a probability model is on the implementation level rather than on the framework level.

For example, it is stated in Deliverable 6.1 that Figure 7 gives a 'generic Bayesian network framework developed for RAIN risk assessment framework'. Now, if we look at this figure, Figure 1.1, then we may see that this figure shows us not a framework, but, rather, an implementation of this framework, by way of a specific probability model. So a more apt caption would be: 'a generic

[1] Consistency derivations are a powerful class of proofs which provide us with the proof that the product and sum rules of probability theory are a necessary and sufficient set of rules by which to combine our probability distributions (Cox, 1946; Jaynes, 2003), as well as the derivation of the entropy measure of information theory (Shannon, 1948; Jaynes, 2003).

Bayesian network *model* developed *as an example of the implementation of* the RAIN risk assessment framework'. This specific probability model is generic enough, but, nonetheless, it is still only an example of an implementation.

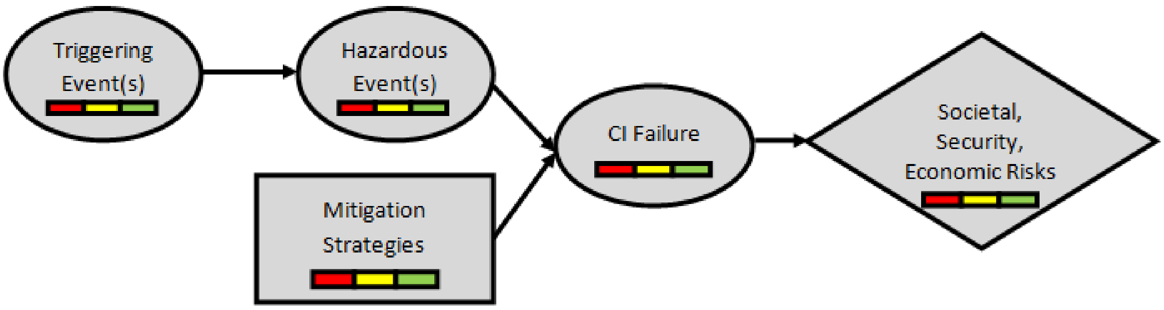


Figure 1.1 Bayesian Network Model

Moreover, this specific implementation was the product of a three-day workshop organized by the probability theorists of WP5, which had as its purpose to give the engineers in the RAIN project a sense for the proposed quantitative risk assessment framework. So more sophisticated implementations, or, equivalently, probability models, may be found by a firm grasp of probability theory and, on the other hand, extensive engineering knowledge about the inference problem at hand.

1.4. Single Hazards vs. Multiple Hazards

The quantitative risk assessment framework – on the framework level – does not distinguish between single and multiple hazards. For both types of hazards the inference phase is done by way of the Bayesian probability theory and the subsequent decision making phase by way of decision theory.

1.4.1. Future research

In the Description Of Work (DOW) of task 5.2 multiple hazards are – implicitly – equated with (inter)dependencies between failures and disruptions of the critical infrastructural systems. The modelling of interdependencies, just like the modelling of cascading effects (the subject of task 5.3), is on the implementation level of the here proposed quantitative risk assessment framework. Stated differently, interdependencies and cascading effects are to be captured by some appropriate probability model, which takes into account all the available data and the expert engineering knowledge.

Markovian processes are currently thought to be the most appropriate class of probability models by which to model interdependencies and cascading effects. For both interdependencies and cascading effects imply the presence some time variable in our probability models and Markovian processes are probability models which evolve over time (steps).

In task 5.3 probability theoretical research will be done into Markovian probability models in order to come up with a generic class of probability models by which to model both interdependencies and cascading. Stated differently, task 5.3 will be a research on the implementation level of the here proposed quantitative risk assessment framework. For on the framework level no distinction is made between risks involving interdependencies, cascading effects or both.

2. Selection process for modelling risks due to single hazards

One of the tasks of deliverable 5.1 is the selection process for selecting the optimal framework for this research project. In the Description of Work (DOW) of task 5.1 is was suggested to use the Objective Ranking Tool (ORT) for this selection process. In this chapter ORT will be introduced as decision support tool (2.1), the process for this selection process will be described (2.2) and the results of the evaluation process will be given (2.3). The conclusion will be presented in (2.4). In the Annex F more detailed outcomes of the ORT-analyses are given.

For this selection process a dedicated application within ORT is developed with the aim to evaluate existing frameworks due to model risks due to single hazards. The frameworks taken into consideration for this selection process was the outcome of the deliverable 3.2 ‘Defining critical land transport infrastructure protection methods’. This deliverable is extended with three additional frameworks: Bayesian Networks, Bayesian probabilities and Influence diagrams.

Within task 5.1 there was a preference for the use of the Bayesian probabilities approach, based on existing experience and knowledge. The question to be answered with the ORT was whether the Bayesian probabilities approach would score high in a formal selection process.

Developing an ORT-application for such a selection process resulted that ‘Influence diagrams’ was evaluated as the highest scoring framework to model risks due to single hazards. Related to a fictive reference framework (the ultimate model) the Influence Diagram scores 0,9976 out of 1,000. Bayesian networks and Bayesian probabilities gave outcomes of 0,9892 and 0,9776.

2.1. Summary of the Objective Ranking Tool (ORT) methodology

The Objective Ranking Tool (hereinafter ORT) is based on three scientific principles: Similarity Judgment, Analytic Hierarchy Processing and the use of a Delphi-panel. PSJ developed ORT as a dedicated application that can be used in any form of decision-support, decision-making and prioritisation of alternatives. ORT provides a unified process and structured support tool.

2.1.1. Introduction

The principle of 'equality' - hereinafter referred to as similarity - supposes that people make judgments and reviews about ‘phenomena’ by comparing the agreement and differences between these objects. Similarity judgment is developed within cognitive psychology for one-to-one comparisons and has applications in many areas. This principle, developed by Amos Tversky (1984), has been used by Prak (2009) to prioritize objects within the National Alert System in the Netherlands to prioritise the most vulnerable locations related to terrorism.

2.1.2. Theory

Tversky concluded that the equivalence of two phenomena is determined by the analysis of commonalities as well as by the unique characteristics of both phenomena separately. Based on this insight, Tversky developed a mathematical model: $S_{ij} = f_{ij} / [f_{i,j} + a(f_{i,\,not\,j}) + b(f_{not\,i,\,j})]$ in which: 1) features in the reference object but not in the study object, $f_{i,\,not\,j}$; 2) features in the study object, but not in the reference object, $f_{not\,I,\,j}$; 3) common features, $f_{i,\,j}$; 4) the constant 'a' and 'b' add up to '1’.

The principles behind the contrast model, on which the similarity judgment methodology is developed, may well be used to prioritize phenomena, objects or alternatives. Through the use of Delphi panels (expert judgment panels) and the use of Analytic Hierarchy Processing (AHP) a reliable identification of features is possible and weight percentages to these characteristics can be provided.

2.1.3. Objective Ranking Tool

ORT facilitates larger numbers of simultaneous judgments between different objects. ORT applies a reference object that meets all the features and compares alternatives with this reference. Features should be developed by a Delphi-panel, a dedicated group of experts, representing the subjects of interests that covers the question to be answered. By assigning to all features a weighting factor with AHP, the application calculates the degree of similarity to the reference.

All objects are assessed to the developed features. A feature can be TRUE or FALSE. Applying a value between "TRUE" and "FALSE" is possible within ORT. Features are then divided into 'substitutive' or 'additive'. When using many features it is not desirable to apply a correction factor within the contrast model. Setting a limit on the outcome of the ORT analysis, a number between '0'and '1' is not appropriate. The ORT analysis is about the relative ordering between objects. Whether a sector designates a prime location depends also on other factors. The outcome of the ORT analysis can be used to prioritize the deployment of available capacities. With the results of the ORT analysis it is possible to agree on an ordering of objects within and between sectors. This similarity is expressed as a number between 0 and 1. The closer to 1, the higher the equivalence is with the reference object. The highest scoring alternative meets the requirements as set with the developed features the best.

2.1.4. Process

Using a structured group of experts in the form of a Delphi panel is needed in order to achieve more accurate results. This has benefits for both the determination of the characteristics, determining the weighting factors as well as in assessing whether an item meets the characteristics. A Delphi panel with a number between five and 10 people each with own expertise is suggested;

For the determination of weighting factors, the Analytic Hierarchy Process (AHP) is proposed. Pairwise comparisons are carried out in a structured way in which a preference is given for one of the characteristics. A value of '1 'indicates an equivalence (no preference), a value of '9 ' gives an extreme inequality (high preference for one of the characteristics). Reliable results are obtained with up to seven comparable characteristics. Within ORT, this means an adaptation to three levels (criteria, sub-criteria, sub-sub criteria) of seven characteristics that, in theory, can lead to 343 (7^3) features to be assessed.

The combination of Delphi panel and the use of AHP will give a one-time investment in order to develop features, give these features a weight and assess the developed features;

Within ORT a differentiation is possible for assessing whether an object conforms to the described characteristics. The contrast model checks the unique and common features of objects. The ORT assigns a score of '0' when an object to be compared does not hold the characteristic (FALSE) or '1' if the object to be compared does possess the characteristic (TRUE). This is called a "substitutive" feature. In reviewing some characteristics such binary assessment is not always possible. These characteristics may be partly true or false and are qualified as 'additive' characteristics. For these 'additive' characteristics, a linear point scale is proposed with an ascending value of ‘0.1’. With this addition, a more refined result is achieved;

The constant 'a' and 'b' in the model $S_{ij} = f_{ij} / [f_{i,j} + a(f_{i, \text{not } j}) + b(f_{\text{not } i, j})]$ add up to 1. With this constant, a correction factor is suggested that one characteristic is weighed more heavily in relation to the other characteristic. The reliable determination of this constant can be achieved through a process of AHP. In a discussion in a Delphi panel on the value of this constant, it will be asked for the reason for the distinction. Arguments will be used which actually are complementary characteristics. It is therefore more valuable to invest in an investigation of these additional features then to determine the value of ' a' and ' b';

The results of the ORT process indicate a ranking in the extent to which objects has a certain degree of equivalence. The ranking of objects facilitates the decision-making process

2.1.5. Application

Nowadays the ORT-application is a dedicated web-based-tool and will be extended with additional features in the future. The web-based data can be exported by .csv-files. Screenshots of the application are given in appendix F.

2.2. Selection process

The process followed is described in in two tasks. The work of the Delphi-panel to define the aim of the application, define criteria, give those criteria weight factors, define the scope of the framework to be judged and the scoring process. After the scoring process the results of the analyses were presented by the application and a short analysis was made of the outcome. Whereas in the explanation ‘features’ are used within the ORT-application we speak about ‘criteria’.

2.2.1. Delphi-panel

The Delphi-panel was designed with the people assigned to this task. Due to the limited timeframe between the execution of deliverable 3.2 and the final deliverable 5.1 there was no possibility to extend the people involved. The process was guided by PSJ.

The first task was to define the aim of this dedicated ORT-application. The application will be the ‘selection process for modelling risks due to single hazards’.

The second task was to define the criteria that should be taken into consideration for this selection process. Also a description in more detail is made for each criteria so there is a common understanding about the meaning. As third task the weighting factors of each criterion were defined by using Analytic Hierarchy Processing. An overview of criteria and weighting factors is given in table 1.

The fourth task was to define the frameworks to be considered. The result of the deliverable 3.2 ‘Defining critical land transport infrastructure protection methods’ are used as bases. This deliverable is extended with three additional frameworks: Bayesian Network, Bayesian probabilities and Influence diagram. The reasons for this extension is that these models are not regularly used within the area of critical land transport infrastructure protection. Within the world of probabilistic science these models are used frequently.

The last task of the Delphi-panel was to score each framework on the criteria given. The result of this process is given in table 2. All the criteria were qualified as ‘additive’ which means that not only a judgment of TRUE (1,0) or FALSE (0,0) is possible but also the intermediate figures. The scoring process was executed by the following increasing list related to the fulfilment of the criteria:

- not or almost not: 0 -> 0,1
- limited 0,2 <- 0,3
- partially 0,4 <- 0,5 -> 0,6
- almost 0,7 <- 0,8 -> 0,9
- almost full or full 0,9 <- 1,0

CRITERIA	PERCENTAGE	TYPE	SUBCRITERIA	PERCENTAGE	TYPE
Knowledge and Information	18,22	additive	Overall expertise within consortium	47,37	additive
Knowledge and Information	18,22	additive	Availability of data	10,71	additive
Knowledge and Information	18,22	additive	Expert engineering knowledge	41,92	additive
Framework	26,43	additive	Completeness	6	additive
Framework	26,43	additive	Reliability	31,23	additive
Framework	26,43	additive	Validity	41,07	additive
Framework	26,43	additive	Transparancy	21,7	additive
Use of framework	8,86	additive	Attractiveness	80,82	additive
Use of framework	8,86	additive	Simplicity	6,23	additive
Use of framework	8,86	additive	Extensibility	12,95	additive
Innovative framework	5,81	additive			
Suitability for problem	40,68	additive			

Table 1: criteria and their respective weighting factors

CRITERIA	WEIGHT	CRITERIA	WEIGHT	Bayesian Network 1(50%)	BTA - Bow Tie analysis 1(50%)	CCA - Cause and Consequences Analyses 1(50%)	CL - Checklist 1(50%)	ETA - Event Tree Analysis 1(50%)	FTA - Fault Tree Analysis 1(50%)	FMEA - Fault Mode and Effect Analysis 1(50%)	HAZOP - HAZard and OPerability study 1(50%)	What If analysis 1(50%)	Point method 1(50%)	SWOT analysis 1(50%)
Knowledge and Information	0,18	Overall expertise within consortium	0,47	1	1	1	1	1	1	0,5	0,3	0,7	0,3	0,7
Knowledge and Information	0,18	Availability of data	0,11	0,9	0,5	0,5	0,5	0,5	0,5	0,3	0,3	0,3	0,5	0,5
Knowledge and Information	0,18	Expert engineering knowledge	0,42	1	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7
Framework	0,26	Completeness	0,06	0,5	0,8	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7
Framework	0,26	Reliability	0,31	1	0,8	0,8	1	0,8	0,8	0,5	0,5	0,3	0,5	0,3
Framework	0,26	Validity	0,41	1	0,5	0,5	0,7	0,5	0,5	0,5	0,5	0,5	0,5	0,3
Framework	0,26	Transparancy	0,22	1	0,5	0,5	0,7	0,5	0,5	0,7	0,5	0,5	0,7	1
Use of framework	0,09	Attractiveness	0,81	0,9	0,8	0,5	0,3	0,5	0,5	0,3	0,3	0,3	0,3	0,3
Use of framework	0,09	Simplicity	0,06	0,2	0,5	0,5	0,7	1	1	0,7	0,3	0,7	0,7	0,7
Use of framework	0,09	Extensibility	0,13	1	1	1	1	1	1	1	1	0,7	0,7	0,7
Innovative framework	0,06			1	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3
Suitability for problem	0,41			1	0,5	0,5	0,3	0,5	0,5	0,3	0,3	0,3	0,1	0,1

CRITERIA	WEIGHT	CRITERIA	WEIGHT	Decision Matrix 1(50%)	Common Safety Method 1(50%)	Critical Path Method and Program Evaluation and Review Technique 1(50%)	Method of optimal network connection 1(50%)	Method of time consuming activities 1(50%)	Gantt diagram 1(50%)	Methods and measures lowering impacts 1(50%)	Subsystem of technical security devices 1(50%)	Subsystem of organizational measures 1(50%)	Bayesian probabilities 1(50%)	Influence diagram 1(50%)
Knowledge and Information	0,18	Overall expertise within consortium	0,47	0,7	0,3	0,3	0,3	0,3	0,7	0,7	0,3	0,3	1	1
Knowledge and Information	0,18	Availability of data	0,11	0,5	0,3	0,3	0,3	0,3	0,3	0,3	0,7	0,7	0,3	0,9
Knowledge and Information	0,18	Expert engineering knowledge	0,42	0,7	0,5	0,5	0,5	0,2	0,3	0,3	0,3	0,3	1	1
Framework	0,26	Completeness	0,06	0,7	1	0,7	0,5	0,3	0,3	0,5	0,1	0,1	0,7	1
Framework	0,26	Reliability	0,31	0,3	0,7	0,7	0,7	0,5	0,7	0,7	0,1	0,1	1	1
Framework	0,26	Validity	0,41	0,3	0,7	0,7	0,5	0,7	0,7	0,7	0,7	0,7	1	1
Framework	0,26	Transparancy	0,22	1	1	1	1	1	1	0,5	0,7	0,7	1	1
Use of framework	0,09	Attractiveness	0,81	0,3	0,7	0,3	0,1	0,1	0,1	0,3	0,1	0,1	0,7	1
Use of framework	0,09	Simplicity	0,06	0,7	0,5	0,5	0,7	0,7	0,7	0,3	0,3	0,3	0,3	0,5
Use of framework	0,09	Extensibility	0,13	0,7	1	1	1	1	1	1	1	1	1	1
Innovative framework	0,06			0,3	0,7	0,3	0,3	0,3	0,3	0,3	0,3	0,3	1	1
Suitability for problem	0,41			0,1	0,5	0,1	0	0	0	0	0	0	1	1

Table 2: scoring overview of each framework on the criteria

2.3. Result and analysis

Based on the mathematical model the application presents the results. With dedicated analysis the highest scoring criteria to the result can be presented.

2.3.1. Results

The mathematical model designed by Tversky ($S_{ij} = f_{ij} / [f_{i,j} + a(f_{i,\text{ not } j}) + b(f_{\text{not } i,\, j})]$) is computed within the application. ‘A’ and ‘b’ are set equal (each 0,5) because there is not a limitation in the number of features so a correction between the frameworks is not necessary. Each criteria is counted on the relative weight as a result of the AHP-analysis and in the computational model the scoring input is used. All frameworks are compared to a reference which scores a maximum of 1.0 on each criteria.

Related to a fictive reference framework (the ultimate model) the Influence Diagram scores 0,9976 out of 1,000. Bayesian networks and Bayesian probabilities gave outcomes of 0,9892 and 0,9776.

	REFERENCE	Influence diagram	Bayesian Network	Bayesian probabilities	BTA - Bow Tie analysis	Common Safety Method	FTA - Fault Tree Analysis	ETA - Event Tree Analysis	CCA - Cause and Consequences Analyses	CL - Checklist	FMEA - Fault Mode and Effect Analysis	What If analysis
REFERENCE	1	0,9976	0,9892	0,9776	0,7528	0,7384	0,7368	0,7368	0,7346	0,6976	0,597	0,583

	REFERENCE	HAZOP - HAZard and OPerability study	Critical Path Method and Program Evaluation and Review Technique	SWOT analysis	Decision Matrix	Point method	Gantt diagram	Methods and measures lowering impacts	Method of optimal network connection	Method of time consuming activities	Subsystem of organization al measures	Subsystem of technical security devices
REFERENCE	1	0,5659	0,5358	0,5085	0,5085	0,4933	0,4905	0,4751	0,4462	0,4206	0,3689	0,3689

Table 3: sorted outcome of the level of similarity judgment

2.3.2. Analysis

Within the application it is possible to evaluate on those criteria that contributes the most to the outcome. Selections can be made for the ownership of any criteria (not used), if any criteria can be influenced by the defined owner (not used) and the number of criteria. In this example this number was set to five.

Full analyses of the differences between the Influence diagram, Bayesian probabilities and the Bayesian Network, the three highest scoring frameworks shows that the differences respectively based on (Influence diagram / Bayesian network / Bayesian probabilities)

- Availability of data 0,9 / 0,9 / 0,3
- Completeness 1,0 / 0,5 / 0,7
- Attractiveness 1,0 / 0,9 / 0,7
- Simplicity 0,5 / 0,2 / 0,3

Although this outcome the choice has been made for Bayesian probabilities. The reason for this is in Appendix A of this deliverable (A Non-Technical Introduction into Bayesian Networks).

An overview of the analysis on all frameworks is given in table 4.

CRITERIA	CRITERIA	WEIGHT FACTOR TOTAL	Bayesian Network 1(50%)	BTA - Bow Tie analysis 1(50%)	CCA - Cause and Consequences Analyses 1(50%)	CL - Checklist 1(50%)	ETA - Event Tree Analysis 1(50%)	FTA - Fault Tree Analysis 1(50%)	FMEA - Fault Mode and Effect Analysis 1(50%)	HAZOP - HAZard and OPerability study 1(50%)	What If analysis 1(50%)	Point method 1(50%)	SWOT analysis 1(50%)
Knowledge and Information	Overall expertise within consortium	8,63%	1.0	1.0	1.0	1.0	1.0	1.0	0.5		0.7		0.7
Knowledge and Information	Availability of data	1,95%											
Knowledge and Information	Expert engineering knowledge	7,64%	1.0		0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Framework	Completeness	1,59%											
Framework	Reliability	8,25%	1.0	0.8	0.8	1.0	0.8	0.8	0.5	0.5		0.5	
Framework	Validity	10,85%	1.0	0.5	0.5	0.7	0.5	0.5	0.5	0.5	0.5	0.5	0.3
Framework	Transparancy	5,74%								0.5	0.5	0.7	1.0
Use of framework	Attractiveness	7,16%		0.8									
Use of framework	Simplicity	0,55%											
Use of framework	Extensibility	1,15%											
Innovative framework		0,06%											
Suitability for problem		0,41%	1.0	0.5	0.5	0.3	0.5	0.5	0.3	0.3	0.3	0.1	0.1

CRITERIA	CRITERIA	WEIGHT FACTOR TOTAL	Decision Matrix 1(50%)	Common Safety Method 1(50%)	Critical Path Method and Program Evaluation and Review Technique 1(50%)	Method of optimal network connection 1(50%)	Method of time consuming activities 1(50%)	Gantt diagram 1(50%)	Methods and measures lowering impacts 1(50%)	Subsystem of technical security devices 1(50%)	Subsystem of organizational measures 1(50%)	Bayesian probabilities 1(50%)	Influence diagram 1(50%)
Knowledge and Information	Overall expertise within consortium	8,63%	0.7			0.3	0.3	0.7	0.7	0.3	0.3	1.0	1.0
Knowledge and Information	Availability of data	1,95%											
Knowledge and Information	Expert engineering knowledge	7,64%	0.7		0.5	0.5		0.3	0.3	0.3	0.3	1.0	1.0
Framework	Completeness	1,59%											
Framework	Reliability	8,25%		0.7	0.7	0.7	0.5	0.7	0.7			1.0	1.0
Framework	Validity	10,85%	0.3	0.7	0.7	0.5	0.7	0.7	0.7	0.7	0.7	1.0	1.0
Framework	Transparancy	5,74%	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.7	0.7		
Use of framework	Attractiveness	7,16%		0.7									
Use of framework	Simplicity	0,55%											
Use of framework	Extensibility	1,15%											
Innovative framework		0,06%					0.3			0.3	0.3		
Suitability for problem		0,41%	0.1	0.5	0.1							1.0	1.0

Table 4: overview of the analysis of the five most contributing criteria to the outcome of the similarity judgment.

2.4. Conclusion

A total of twelve criteria have been developed for the selection process for modelling risks due to single hazards. With the use of the principles of Analytic Hierarchy Processing those criteria have been given a relative weight for the computation. A number of 22 frameworks are defined for this selection process. Each framework has been scored on the criteria with an outcome between 0 and 1. The computation calculated the similarity between a fictive reference framework which fulfils all criteria on the maximum of 1. The outcome of the similarity judgment process was that the Influence diagram is evaluated as highest scoring framework (0,9976) however the Bayesian probabilities (0,9776) was used within task 5.1 as outlined in Appendix A.

3. A General Outline of the Proposed Risk Analysis Framework

In order to quantify both single and multi-mode risks and the impacts of extreme weather events on interconnected critical (infrastructure) systems, a risk analysis framework is proposed that – in the inference phase – quantifies the probabilities of the possible outcomes as a function of the actions that a risk manager might be contemplating to take and then – in the decision phase – helps to choose that action that promises to minimize the pertinent risks.

The necessity of reasoning as best we can in situations where our information is incomplete is faced by all of us, every waking hour of our lives. Should I wear a raincoat today, eat that egg, cross that street, tote that bale, buy that book? For risk managers this is no different. They have to ask themselves, for example: do we improve the flood surge defences, and if so how much are we willing to spend on these improvements, or do we leave the flood surge defences as they are?

As a rule, we must decide what to do next, even though we cannot be certain what the consequences will be. Introspection would suggest that before deciding what action to take our intuition organizes the preliminary reasoning in the following stages (Jaynes, 1985):

1. Try to foresee all the possibilities that might arise.
2. Judge how likely each is, based on everything you can see and all your past experience.
3. In the light of this, judge what the probable consequences of various actions would be.
4. Now make your decision.

From the earliest times this process of decision making has been recognized. Herodotus, in about 500 BC, discusses the policy decisions of the Persian kings. He notes that a decision is wise if the evidence at hand indicated it as the best one to make (Jaynes, 1985).

So this kind of reasoning has been around for a long time, and has been well understood for a long time. Furthermore, it is so well organized in our minds in qualitative form that is seems obvious that the above stages of reasoning can be reproduced in quantitative form by a mathematical model, and that such a mathematical model would be very useful in such areas as engineering and economics, where we are obliged constantly to make decisions as best we can in spite of incomplete information, but the number of possibilities and amount of data are far too great for intuition to keep track of (Jaynes, 1985).

In Figure 2.1. an outline of such a mathematical model – the here proposed quantitative risk analysis framework – is given; where the inference phase corresponds with the first three rows and the decision making phase corresponds with the last two rows.

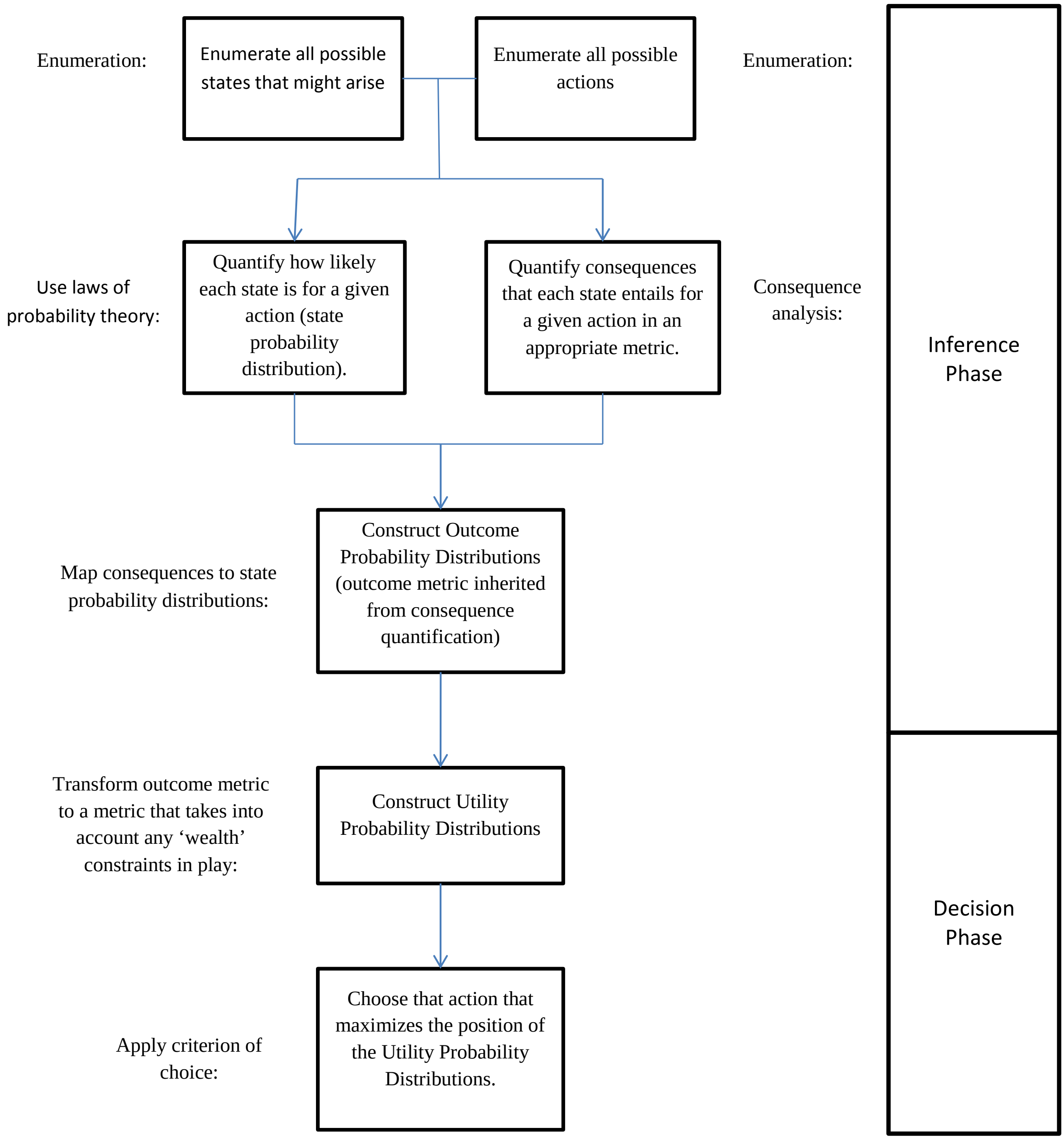


Figure 2.1 Quantitative Risk Analysis Framework, consisting of 5 rows

Fig. 2.1 served as a basis for Dx` in which the framework is preceded by 'Establishing the context' and the main blocks are denoted as 'Risk Identification', 'Mitigation measures', 'Risk inference', and 'Risk evaluation'.

3.1. The Inference Phase

The inference phase of the risk analysis framework in Figure 2.1 consists of a enumeration phase, a quantification phase, and a construction sub-phase. Note that all of the inferential effort of any risk analysis is concentrated in the inference phase.

In the enumeration sub-phase we make an enumeration of the possible states of the system under consideration, as well as an enumeration of the possible actions that may reduce risk. The enumeration of the possible states of a system is an enumeration of the hypothesis space of the (engineering) stakeholders. So, possible states that are not entertained as possibilities by the stakeholders will not enter the risk analysis. Likewise, the enumeration of the possible actions is an enumeration of the action space of the risk managers. Again, possible actions that are not entertained as alternatives by the risk managers will not enter the risk analysis.

In the quantification sub-phase we do the probability and the consequence analyses. In the probability analysis the qualitative common-sense considerations of the (engineering) stakeholders and the available data on the system under consideration are translated to state probability distributions which are logically connected[2]. By way of the product and sum rules[3] (Jaynes, 2003), these connected state probability distributions are then combined into the state probability distributions of interest. In the here proposed risk analysis framework the state probability distributions are the marginal conditional probability distributions of all the outcome states of the system given some action[4]. In the consequence analysis the outcome states of the system are assigned some numerical consequence in the metric of interest[5]. The numerical consequences that are assigned to the possible states of the system are - like the state probabilities – based on the qualitative common-sense considerations of the stakeholders and the available data on like outcomes, past or present.

In the construction sub-phase of the inference phase the state probability distributions and the numerical consequences of the states of the states of these probability distributions are combined. This results in the outcome probability distributions which admit cumulants (i.e., mean, standard deviation, skewness, etc.) in the consequence metric.

3.2. The Decision Phase

---

[2] An example of a logical connection is that a landslide at location x is related to the state of the road at location x. For this simple example we will in all likelihood first assign a probability distribution to the occurrence of a landslide and then a conditional probability distribution for the state of the road, given the occurrence or non-occurrence of a landslide. The logical connection between landslides and road states then finds it quantitative expression in the conditional dependence of the probability of the road state given a landslide occurrence.

[3] The strict adherence to the laws of probability theory, that is, the product and sum rules, makes the here proposed risk framework 'Bayesian' on the inference level.

[4] An example of a possible action is the investment in some preventive or mitigating measure that will, respectively, lower the probability of extraordinary events or lower the negative consequences of these extraordinary events. But actions also admit a finer resolution, as we define actions as investments of size x we make in order to improve a system's safety; where x is variable, lying in the interval $a < x < b$; see Chapter 5.

[5] Some examples of consequence metrics are monetary losses/gains, measures of losses/gains in citizen confidence, and loss-of-life increments.

The decision phase of the risk analysis framework in Figure 2.1 consists of a construction phase and a maximization sub-phase.

In the construction sub-phase of the decision phase the numerical consequences of the outcome probability distributions are transformed to their corresponding utilities, by way of an appropriate utility function. By so doing, we transform our outcome probability distributions from the inference phase to their corresponding utility probability distributions. These utility probability distributions admit cumulants (i.e., mean, standard deviation, skewness, etc.) in a properly transformed metric, that takes into account either risk averseness or debt relief in the consequence metric, whenever pertinent[6]. This concludes the discussion of the construction phase.

In the maximization sub-phase we take that action which promises to be most advantageous. If the metric axis is oriented so that a disutility is given a negative metric value and an utility a positive metric value. Then we have that the action that puts the position of the utility probability distribution the most to the right is the most advantageous decision in terms of the adopted metric[7]. Positions of probability distributions on the same metric axis may be ordered transitively by way of the sum of (a) the lower confidence bound, (b) the expectation value, and (c) the upper confidence bound (van Erp *et al.*, 2015). The larger the sum of the lower and upper confidence bound and expectation value of a given utility probability distribution[8], the more advantageous the action which gives rise to this utility probability distribution.

[6] For infinite wealth, or, equivalently, for relatively infinitesimal increments or decrements in the current wealth, the utilities become proportional to the untransformed numerical consequences. As the difference in proportionality cancels out automatically in the decision theoretical (in)equalities, we have that for infinite wealth outcome probability distributions, third row Figure 2.1, are equivalent to utility probability distributions, fourth row Figure 2.1.

[7] The here proposed maximization of the position of the outcome probability distribution, as opposed to only its means, results in a non-trivial refinement of the criterion of choice of the decision theoretical Expected Utility Theory. This refinement together with the derivation of a consistency proof of the Bernoulli utility function, also known as the Weber-Fechner law of psycho-physics, makes the here proposed risk framework 'Bayesian' on the decision theoretical level, as explained in (van Erp *et al.*, 2014).

[8] Risk is traditionally defined as probability times consequence, or, equivalently, the means of the outcome distributions. But this definition neglects the risk exposure as captured in the standard deviation and higher order cumulants of the outcome/utility probability distributions.

4. The Inference Phase

We now will proceed to give a more in-depth discussion of the separate boxes in the schema of Figure 2.1 which belong to the inference phase. In the discussion of these boxes we will work from the left to right, where we start from the top and work our way downwards.

4.1. Enumeration of possible states

The enumeration of the possible states is an exercise in qualitative analysis. In this analysis we have to define the hypothesis space regarding the system states. The resulting hypothesis space will be a reflection of the expert knowledge of the involved risk managers and (structural) engineers. Seeing that those events we do not foresee are the ones that are also apt to blind-side us, care should be taken to make an exhaustive inventory of all to possible system states.

4.2. Enumeration of possible actions

In the enumeration of the possible actions care should be taken that all relevant alternatives are considered. This is so the risk manager can come to a well-balanced choice, where all feasible options have been examined and weighed.

4.3. Quantification of likelihoods

The quantification of the likeliness of the possible states of the system under consideration results gives us the state probability distributions.

4.3.1. The laws of probability theory

Denoting various propositions by A, B, C, *etc.*, let AB stand for the proposition "both A and B are true", $\overline{A}$ = " A is false", and let the symbol $p(A|B)$ stand for "the probability that A is true, given that B is true". Then the basic laws of probability theory are the product and sum rules (Jaynes, 2003):

$$p(AB|C) = p(A|BC)\, p(B|C) \tag{1}$$

$$p(A|C) + p(\overline{A}|C) = 1 \tag{2}$$

But AB and BA are the same proposition, so consistency requires that we may interchange A and B in the right-hand side of (1). If $p(B|C) > 0$, we thus have what is called "Bayes' Theorem" today, although Bayes never wrote it[9] (Jaynes, 1985):

$$p(A|BC) = p(A|C)\frac{p(B|AC)}{p(B|C)} \tag{3}$$

---

[9] What is commonly known as 'Bayes' Theorem', in accordance with the law of epynomy, was first given by Laplace (Jaynes, 1978).

So, Bayes' Theorem is nothing more than the statement that the product rule (2) is consistent; why is such a seeming triviality important?

In (3) we have a mathematical representation of the process of learning. $p(A|C)$ is our "prior probability" for A, when we know only C. $p(A|BC)$ is its "posterior probability", updated as a result of acquiring new information B. Typically, A represents some hypothesis, or theory, whose truth we wish to ascertain, B represents new data from some observation, and the "prior information" C represents the totality of what we knew about A before getting the data B.

For example – as in a famous analysis done by Laplace – proposition A might be the statement that the unknown mass M_S of Saturn lies in a specified interval, B the data from observatories about the mutual perturbations of Jupiter and Saturn, C the common sense observation that M_S cannot be so small that Saturn would lose it rings; or so large that Saturn would disrupt the solar system.

Laplace found, by way of (3), astronomical perturbation data up to the end of the 18th Century, and common sense celestial mechanics considerations, that M_S may be estimated to be (1/3512) of the solar mass. Furthermore, Laplace gave a probability of 0.99991, or odds of 11,000:1, that the actual value of M_S lay within 1% of this estimate. The advent of the space age, 150 years hence, and the subsequent further accumulation of new data has raised this 18th Century estimate by 0.63 percent; well within the bounds as originally predicted by Laplace's Bayesian probability analysis (Jaynes, 1973). So, we see that Bayesian probability theory, or, equivalently, the consequent application of the product and sum rules, respectively, (1) and (2), in the right hands, is a powerful tool of quantitative inference.

4.3.2. Bayesian probability theory vs. Bayesian networks

Bayesian probability theory is the inference instrument of choice for the physics community (Knuth, 2014). In the engineering community the situation is more diffuse. Some engineers come to Bayesian probability theory by way of the work of formal Bayesians – be they ISBA or MaxEnt Bayesians – and others come to Bayesian probability by way of Bayesian Networks, which were proposed by the artificial intelligence community.

Bayesian probability theory, as explained in the previous section, was first discovered as the quantification of our common sense by Laplace; an accomplished physicist who was called by his contemporaries the French Newton of his time. But at the end of the 19th Century, as relativity theory burst upon the physics community, X-rays and radioactivity were discovered, and quantum theory started to develop, physicists had all but completely retired from the field they had created, as a host of new experimental facts needed unravelling and new revolutions of thought needed digesting (Jaynes, 1973).

In the absence of the physicists a new school in statistics arose that rejected the notion of probability as describing a state of knowledge, as too subjective a basis for inference, and insisted, instead, that the only objective definition of a probability was probability as the long-term frequency of an imaginary infinity of replications in an imaginary random experiment[10]. This new school of thought was so extremely aggressive that its methods, now known as 'orthodox statistics', for a time dominated the field so completely that those who were students in the period 1930-1960 were hardly aware that any other conception of probability had ever existed (Jaynes, 1978).

This state of affairs changed when the physicists re-entered the field of probability theory. Notably, Lee Lusted, Arnold Zellner, and Edwin Jaynes, respectively, a bio-statistician, an econometrist, and a physicist by trade, though all three physicists by training, reintroduced Laplace's ideas, as transmitted to them by the work of Jeffreys (Jeffreys, 1939) – the only physicist still working in the field of probability theory in the orthodox interim – in their respective fields of work[11] (Lusted 1968; Zellner, 1971; Jaynes, 2003). As a result the polemical trench-warfare commenced that was the Bayesian vs. Orthodox Statistics schism[12].

In parallel with the Great Schism between Bayesian and Orthodox statisticians, during the late 1970s and early 1980s computing power became readily available and, as a consequence, the Artificial Intelligence (AI) community, consisting of computer scientists, saw its inception. The AI-community, unsatisfied with the then dominant frequentist probability theory and rejecting the Bayesian alternative out of hand – as was the fashionable thing to do in those days – started to introduce their own suite of methods to quantify inference and knowledge representation. Examples of these attempts were Default Logic, Non-Monotonic Logic, Certainty Factors, Schafer-Dempster Theory, Confirmation Theory, Fuzzy Logic, and Endorsements. However, all these attempts were short-lived[13], and even Fuzzy Logic (Zadeh, 1973), which for a time enjoyed some popularity, has now fallen by the scientific way-side (Skilling, 2005).

The AI-community, by way of (Pearl, 1988), eventually came to recognize the power of Bayesian probability theory as a quantification of our common sense. But even though Pearl on a conceptual level saw the Bayesian probability calculus as the ideal, he, nonetheless, felt that on an implementation level there was an 'obvious' computational infeasibility in the exponential explosion in the evaluation of the joint probability distribution of the propositions under consideration. This

---

[10] For example, in a 'frequentist' analysis of the mass-of-Saturn estimation problem one would need to invoke an infinity of parallel universes, where the mass of Saturn is a long-term frequency of the masses of all the respective Saturn planets in some imaginary random experiment where we sample masses M_S from this imaginary ensemble of universes.

[11] Zellner founded the ISBA Bayesian community; a community that consists of econometrists, bio-statisticians, psychologists, and sociologists. Jaynes founded the MaxEnt community; a community that consists mainly of physicists. Jaynes described (Zellner, 1971) as "In spite of the word 'econometrics' in the title, this work contains universal principles and will be highly valuable to all scientists and engineers.". But Zellner seemed loathe to cite the work Jaynes. So, ISBA Bayesians seem to be relatively unfamiliar with the work done by their MaxEnt collegues, and as a consequence have a slightly less fundamental outlook on what it is that constitutes a Bayesian analysis (Skilling, 2008).

[12] A polemical highlight in the Great Debate is the paper *Confidence Intervals vs. Bayesian Intervals* (Jaynes, 1973), with rejoinders of Maxfield and Kempthorne, and counter rejoinders of Jaynes; to be found on the dedicatory website to E.T. Jaynes, where a collection of both his published and unpublished papers may be found.

[13] For a critical discussion for these AI-attempts to capture and quantify the laws of inference, we refer the interested reader to (Cheeseman, 1985; Cheeseman, 1988; Jaynes 1990).

led him to formulate an axiomatic system, which finds its intuitive interpretation in graph theory, for the formal definition of the notion informational dependency (Pearl, 1988). So, in order to implement Bayesian probability, Pearl thought it necessity to embellish Laplace's original theory with graph theoretical extensions. The graph theoretical embellished version of Bayesian probability models have gained considerable popularity in the engineering and medical communities, where these models fly under the flag of Bayesian Networks.

In order to demonstrate that the graph theoretical embellishments are just that, that is, embellishments, WP5 has developed an algorithm that has the same functionality as the commercial HUGIN software package which implements Bayesian Networks, but which bypasses both informational dependency algebra, as given in (Pearl, 1988), as well as the use of the graph theoretical Junction Tree algorithm that lies at the heart of HUGIN's algorithmic core (Andreassen *et al.*, 1991).

The WP5 algorithm, henceforth called the Bayesian BN algorithm, implements the product and sum rules, respectively, (1) and (2), by way of a handful of heuristics, which automate a Bayesian's use of these laws of probability theory. It has been found that WP5's Bayesian BNs successfully reproduces HUGIN's results – up to small rounding errors – for the (Gulvanessian *et al.*, 1999) constellation of (conditional) probability distributions.

This successful reproduction holds two implications. The first implication is that HUGIN does what it promises to do; that is, it implements – up to a small rounding error – the product and sum rules of Bayesian probability theory. However, the second, more important implication is that we may dispense with Pearl's Bayesian Networks. As it is now shown that a judicious use of Laplace's Bayesian probability theory (Laplace, 1819), without the graph theoretical additions, may suffice to evaluate the kind of large-scale inference problems which according to the AI community could only be done by way of Bayesian Networks.

4.3.3. How to use the laws of probability theory in the RAIN project

For the quantification of the likelihoods, the first box in the second row of Figure 2.1, of RAIN there was initially proposed in Helsinki a Bayesian network approach. But after having done our research and after having managed to construct a fully Bayesian alternative for the AI Bayesian Networks – as outlined in the Appendices A and B – we now, instead, propose to use the inference framework already in use by the physics community; that is, Bayesian probability theory, or, equivalently, a strict adherence to the product and sum rules, without any unnecessary graph theoretical embellishments (Sivia, 1996; Jaynes, 2003; Skilling 2005).

In Bayesian probability theory chains of inference are connected by way of the product rule (1) and collapsed over the variables which are not directly of interest by way of the sum rule (2. If the unconditional 'prior distribution' is a discrete probability distribution, then the collapsing over the 'nuisance' variables will be done by way of summation); see Section 3.3.1 and Appendix B. If the unconditional 'prior distribution' is continuous, then this collapsing will be done by way of integration; integration being a limit case of summation, as the number of propositions that enumerate the possible states of a variable tend to infinity.

4.4. Consequence analysis

We now discuss the implementation of the second box of the second row of Figure 2.1, the consequence quantification of the possible states of the system under consideration. The numerical consequences that we assign to the outcome states of the system should, just like the state probability distributions, based on the qualitative common-sense considerations of the stakeholders and the available data on like outcomes, past or present. Note that consequences may be assigned either deterministically or probabilistically.

The most obvious and straightforward metric is the monetary metric. If each damage state of the system has some cost associated with it and, likewise, if the actions that are contemplated, in the second box of the first row of Figure 2.1, have some monetary cost associated with them, then the cost of a given (damage) state for a given action is just the sum of the cost of that damage state plus the cost incurred by the action taken. But where large sums of money are involved there will also be the cost of money that will have to be taken into account. If the costs of either the damage states or the actions contemplated exceed the current budget of decision maker, then money will have to be borrowed at the capital markets. This borrowing will, typically, come with some additional interest costs. In the same vein, money spend on either the reparation of damage states or invested in preventive or mitigating actions cannot be invested elsewhere, thus, incurring some additional discounting costs.

Another straightforward metric is the loss of life metric. Whereas monetary costs are relatively easy to estimate deterministically, we have that loss of life may have to be estimated probabilistically; seeing that more uncertainties are typically involved (Jonkman, 2007). So, for a given damage state a conditional probability distribution of loss life may have to be constructed which is conditional on the specific damage state of the system. Each loss of life probability distribution may then be connected in the third row of Figure 2.1, to its corresponding damage state, by way of the product rule (1), after which we may summate over the damage state variable, by way of the sum rule (2), thus leaving us with the outcome probability distribution which has as its metric a loss of life metric.

A more challenging metric is the citizen confidence metric. The challenge in the construction for this metric lies in the fact that it asks for a cross-disciplinary research effort, where the social sciences and the structural engineering sciences, together, link the possible damage states of the system, via the possible preventive and mitigating actions taken during the natural hazard, to increments and decrements in citizen confidence.

4.5. Constructing outcome probability distributions

By mapping numerical consequences to the states of state probability distributions we obtain the outcome probability distributions. This construction, once the boxes in the second row in Figure 2.1 are completed, is a purely a question of relabelling if the numerical consequences are deterministic. In the following sub-section we will give the formal mathematical tools for this relabelling which is applicable to both deterministic and probabilistic consequences.

4.5.1. A formal mathematical notation

Let $p(S_i \mid A_j)$ be the state probability distribution for the possible (damage) states S_i under the action A_j, and let C_i be the numerical consequence associated with the state S_i on the numerical consequence metric c. Then by relabelling of the propositions S_i with the numerical values C_i gives us the outcome probability distributions:

$$p(S_i \mid A_j) \rightarrow p(C_i \mid A_j), \qquad \text{by mapping } S_i \rightarrow C_i \tag{4}$$

where the outcome probability distributions $p(C_i \mid A_j)$ have inherited the metric of the numerical consequences C_i.

If the numerical consequences are probabilistic, then we let $p(c \mid \boldsymbol{\theta}_{S_i}, S_i)$ be the consequence probability distribution for the numerical consequences c where $\boldsymbol{\theta}_{S_i}$ are the parameters that define this consequence probability distribution for the state S_i. Then we may combine the state and consequence probability distributions, by way of the product rule (2):

$$p(S_i \mid A_j)\, p(c \mid \boldsymbol{\theta}_{S_i}\, S_i) = p(c, S_i \mid \boldsymbol{\theta}_{S_i}\, A_j) \tag{5}$$

By way of the sum rule (2), we then may summate out the unwanted state variable, leaving us with the outcome probability distribution of interest:

$$\sum_{S_i} p(c, S_i \mid \boldsymbol{\theta}_{S_i}\, A_j) = p(c \mid \boldsymbol{\theta}_{S_i}\, A_j) \tag{6}$$

Note that the straightforward relabelling in (4) is a limit case of (6). This can be shown as follows.

If the numerical consequences c are deterministically set to C_{S_i} for damage state S_i, then we may express this determinism by way of the Dirac delta distribution, which has its sufficient parameter $\boldsymbol{\theta}_{S_i}$ the deterministic value C_{S_i}, so that $\boldsymbol{\theta}_{S_i} = C_{S_i}$:

$$p(c \mid \boldsymbol{\theta}_{S_i}, S_i)\, dc = p(c \mid C_{S_i}, S_i)\, dc = \delta(c - C_{S_i})\, dc = \begin{cases} 1, & c = C_{S_i} \\ 0, & c \neq C_{S_i} \end{cases} \tag{7}$$

Substituting (7) into (5) and (6), we then obtain the alternative expression for the outcome probability distribution (4):

$$\sum_{S_i} p(S_i \mid A_j)\, \delta(c - C_{S_i}) = p(c \mid C_{S_i}\, A_j) \tag{8}$$

The reason that we initially forewent the formal mathematical treatment, (8), of the straightforward process of relabelling, (4), is that the formal Dirac delta notation, (7), may obfuscate the simplicity of the relabelling procedure for the deterministic case.

And the reason that we point to the fact that the construction of outcome probability distributions for deterministic consequences admit the same treatment as for probabilistic consequences is in order to point the reader to the consistency of the construction procedure of outcome probability distributions; where only the product and sum rules are needed to go from state probability distributions to outcome probability distributions.

5. The Decision Phase

We now will proceed to give a more in-depth discussion of the separate boxes in the schema of Figure 2.1 which belong to the decision phase. In the discussion of these boxes we will work from the left to right, where we start from the top and work our way downwards.

5.1. Constructing utility probability distributions

We now discuss the implementation of the fourth row in Figure 2.1, the construction of the utility probability distributions.

5.1.1. The mathematical modelling of wealth constraints

Governmental decision makers, as a rule, will view a 50/50 chance of losing or gaining $1 billion with a lot more trepidation than a 50/50 chance of losing $5 million, even though the expectation value of the outcome is the same in both cases:

$$E(c) = -0.5 \times 1{,}000{,}000{,}000 + 0.5 \times 1{,}000{,}000{,}000 = -0.5 \times 5{,}000{,}000 + 0.5 \times 5{,}000{,}000 = 0$$

A government has a large portfolio of projects and programs and therefore can act fully rational with confidence that, on average, things will turn out well if its portfolio consists of only $5 million projects and programs. However, in the case of single large or politically sensitive project, involving $1 billion, the question of a wealth constraints might arise (Treasury Board of Canada Secretariat, 1998).

The wealth constraint comes in play if the increments or decrements in wealth are a sizable percentage of the total initial wealth which is at the disposal of the decision maker (Bernoulli, 1738; Laplace, 1819; Jaynes, 2003). If a governmental department has a yearly budget of, say, $10 billion, then an increment or decrement of $5 million in this initial wealth will be negligible. However, if an increment or decrement of $1 billion is involved, then the outcome of the project or program will be felt more acutely.

Likewise, going from a monetary metric to a fatality metric, if in a country there are, say, 1000 fatal traffic accidents per year, then an increment in the number of traffic related fatalities by 100 will be notable, but not that much of a calamity; as the initial 'wealth' in fatalities is relatively large compared with the increment. However, if in another country we only have 200 fatal traffic accidents per year, then the same increment of 100 traffic related fatalities will be a calamity; as the initial 'wealth' in fatalities is now not that far removed from the increment.

The operating mechanism for 'wealth' constraints is that if increments or decrements are relatively small relative to the initial 'wealth' will be just noticeable, if at all, whereas increments and decrements that are relatively large relative to the initial 'wealth' will be highly noticeable. The mathematical function that captures this mechanism for positive metrics is Bernoulli's utility function (Bernoulli, 1738) and the mathematical function for negative metrics is the negative Bernoulli utility function (van Erp *et al.*, 2014).

Let M be the initial wealth in the positive monetary metric – the more money, the better – and let ΔM be some increment or decrement in wealth. Then the (dis)utility, say, u, of ΔM is given as (Bernoulli, 1738):

$$u(\Delta M \mid M) = q \log \frac{M + \Delta M}{M}, \qquad (1a)$$

where q is some scaling constant which scales utilities to Just Noticeable Differences[14] (JNDs); these scaling constants will typically differ per metric. Alternatively, let the initial 'wealth' M be in the negative fatality metric[15] – the more fatalities, the worse – and let ΔM be some increment or decrement in 'wealth'. Then the (dis)utility, say, u, of ΔM is given by the negative Bernoulli utility function (van Erp *et al.*, 2015):

$$u(\Delta M \mid M) = -\, q \log \frac{M + \Delta M}{M} \qquad (1b)$$

Note that as the ratio $\Delta M / M$ tends to zero, by way of the properties of the logarithm (9) will tend to:

$$u(\Delta M \mid M) \to \pm \frac{q}{M} \cdot \Delta M, \qquad \text{as} \qquad \frac{\Delta M}{M} \to 0, \qquad (2)$$

and we see that for relatively small increments or decrements ΔM in the initial wealth M, utilities become proportional, with a factor q/M, to the outcomes, or, equivalently, the losses/gains ΔM. So, if we return to the first example given at the beginning of this section, for a governmental department having a budget of $M = \$10{,}000{,}000{,}000$ no wealth constraints will be at play for $\Delta M = \pm\,\$5{,}000{,}000$, as

$$\pm \frac{\Delta M}{M} = \pm \frac{5{,}000{,}000}{10{,}000{,}000{,}000} = \pm\, 0.0005$$

and we return, up to scaling constant q/M, to the untransformed outcome scale (2). But for the large project with $\Delta M = \pm\,\$1{,}000{,}000{,}000$, the wealth constraints will be still in play, as the ratio

$$\pm \frac{\Delta M}{M} = \pm \frac{1{,}000{,}000{,}000}{10{,}000{,}000{,}000} = \pm\, 0.1,$$

forces us to forgo of the approximation (2) in favour of utility function that takes into the wealth constraint (1).

[14] The scaling constant q may be determined experimentally, or, alternatively, may be left defined, as it will fall away in the decision theoretical (in)equalities, which are set up in the box of the fifth row of Figure 2.1.

[15] Another possible negative metric for which the negative Bernoulli utility function is applicable is the monetary debt metric – the more debt, the worse. For an in-depth discussion of debt utilities we refer the interested reader to (van Erp *et al.*, 2015).

Now, Bernoulli's utility function (1) is also the function that guides our sense perception (Fechner, 1860). A sensory stimulus is always perceived relative to some background stimulus level. For example, in the background noise is zero, then we can 'hear a pin drop'. But if we are at a rock-concert, then we will have to shout to make ourselves heard. So, as the background stimulus level, or, equivalent, the initial 'wealth', is increased, the just noticeable increments and decrements will also increase. Psychological experimentation has found that for all sensory stimuli the JNDs are a function of both the initial background stimulus level M and the increment or decrement M, as given in (1), or in its power-law form (Stevens, 1961; Fancher, 1990), where the utility function (1) is exponentiated, and we go from an utility to a exponentiated utility scale (van Erp *et al.*, 2015).

But where Bernoulli derived (1) by way of a monetary variance argument (Bernoulli, 1738), Fechner derived that self-same (1) by way of the Weber law (Fechner, 1860). Because of this alternative derivation Fechner called his derivation of Bernoulli's utility the Weber's law of sense perception, which later became known as the Fechner-Weber law (Masin, *et al*. 2009). Yet another alternative consistency derivation of (1) is given in (van Erp *et al.*, 2015). This derivation[16] makes use of the desideratum of consistency, that the route taken of an increment or decrement in initial wealth should be of no consequence for the final utility of that increment or decrement[17], and the desideratum of invariance for a rescaling in the stimulus units[18]; see Appendix D.

The consistency derivation in (van Erp *et al.*, 2015) shows why it is that increments and decrements in monetary stimuli follow the same law as sensory stimuli. Any other law will violate the consistency and rescaling invariance desiderata. The implication then is that we, in our stimulus perception – money may also be seen as a stimulus that moves us to action – are inherently consistent. This phenomenon is not limited to sense perception, for the product and sum rules of probability theory are also derived by consistency desiderata (Cox, 1946; Jaynes, 2003). And MaxEnt physicists are now in the process of deriving the laws of physics, by way of consistency constraints on lattices of events (Goyal *et al.*, 2010; Knuth, 2010; Knuth, 2014).

5.1.2. Mapping utilities to outcomes

In order to construct the utility probability distributions, we have to transform the numerical consequences of the outcome probability distributions to their corresponding (dis)utilities, by way of Bernoulli's utility function (1).

Let M be the initial wealth constraint, then we have that for the continuous outcome probability distribution (6) in Chapter 3:

$$\sum_{S_i} p\left(c, S_i \mid \boldsymbol{\theta}_{S_i}\ A_j\right) = p\left(c \mid \boldsymbol{\theta}_{S_i}\ A_j\right) \tag{3}$$

---

[16] As an aside, the derivation of Bernoulli's utility function, or, equivalently, the Weber-Fechner law, makes use of the assumption of differentiability. An alternative, more primitive proof, which drops this weak assumption, has been completed and will be published soon (van Erp and Knuth, *under construction*).

[17] A loss of $1000 should invoke the same disutility as two losses of $500 combined.

[18] A loss $1000 should invoke the same disutility as the loss of 100,000 dollar-cents.

we must make the change of variable:

$$u = q \log \frac{M + c}{M}, \qquad \frac{du}{dc} = \frac{q}{M + c}, \tag{4}$$

from which it follows that:

$$c = M \left[\exp\left(\frac{u}{q}\right) - 1 \right], \qquad \frac{M + c}{q} du = dc \tag{5}$$

Substituting (5) into (3), we then obtain the continuous utility probability distribution:

$$p\left(c \mid \boldsymbol{\theta}_{S_i}\, A_j\right) dc \rightarrow \frac{M \exp(u/q)}{q} p\left(M \left[\exp\left(\frac{u}{q}\right) - 1 \right] \mid \boldsymbol{\theta}_{S_i}\, A_j \right) du\,. \tag{6}$$

Note that if we let C be the numerical consequence stochastic (3):

$$C \sim p\left(c \mid \boldsymbol{\theta}_{S_i}\, A_j\right), \tag{7}$$

and let X be the function of the stochastic (7):

$$X = \log \frac{M + C}{M}, \tag{8}$$

where M is a constant. Then the first two cumulants of the alternative stochastic Y, (4):

$$Y = q \log \frac{M + C}{M}, \tag{9}$$

where q is also a constant, are related to the first two cumulants of the stochastic (8), by way of the moment identities (Lindgren, 1993):

$$E(Y) = q\,E(X), \qquad std(Y) = q\,std(X), \tag{10}$$

It is because of this linearity in the scaling constant q of the first two cumulants of (9), that we may set q arbitrarily to one, as this scaling constant automatically falls away in the decision theoretical (in)equalities of the fifth row of Figure 2.1.

The case for discrete outcome distributions is mathematically less involved, as a simple relabelling will suffice. Let $p(C_i \mid A_j)$ be discrete outcome distributions for the different actions A_j. Then the relabelling of the propositions C_i with the transformation

$$U_i = q \log \frac{M + C_i}{M} \tag{11}$$

gives us the utility probability distributions[19]:

$$p(C_i \mid A_j) \rightarrow p(U_i \mid A_j), \quad \text{by transforming} \quad C_i \rightarrow U_i = q \log \frac{M + C_i}{M}. \tag{12}$$

where the identities (10), by necessity, also hold.

5.2. Maximizing the positions of the utility probability distributions

We now discuss the implementation of the fifth row in Figure 2.1, the maximization of the position of the utility probability distributions.

5.2.1. The criterion of choice as a degree of freedom

Let A_1 and A_2 be two actions we have to choose from. Let o_i, for $i = 1, \ldots, n$, and o_j, for $i = 1, \ldots, m$, be the monetary outcomes associated with, respectively, actions A_1 and A_2. Then in the Bayesian decision theory we first construct the two outcome distributions that correspond with these actions:

$$p(o_i \mid A_1), \quad \text{and} \quad p(o_j \mid A_2) \tag{13}$$

We then proceed, by way of the Bernoulli utility function (9*a*), or, equivalently, the Weber-Fechner law, to map utilities to the monetary outcomes o_i and o_j in (20). This leaves us with the utility probability distributions:

$$p(u_i \mid A_1), \quad \text{and} \quad p(u_j \mid A_2) \tag{14}$$

Now, our most primitive intuition regarding the utility probability distributions (21) is that the action which corresponds with the utility probability distribution which lies more to the right will also be the action that promises to be the most advantageous. So, when making a decision we ought to compare the positions of the utility probability distributions on the utility axis and then choose that action which maximizes the position of these utility probability distributions.

This all sounds intuitive enough. But how do we define the position of a probability distribution? Ideally we would have some consistency derivation of what constitutes a position measure of a probability distribution, say,

$$H_n(p_1, \ldots, p_n, x_1, \ldots, x_n) \tag{15}$$

where p_i are the probabilities of the values x_i, for $i = 1, \ldots, n$. But in the absence of such a consistency derivation we have to take our recourse to *ad hoc* common sense considerations. Stated differently, the criterion of choice in our decision theory constitutes a degree of freedom.

[19] The simple relabelling procedure given here also admits the Dirac delta procedure given in Section 3.5, (van Erp *et al.*, 2015).

5.2.1.1 *The expectation value as a position measure*

From the introduction of expected outcome theory in the 17th century and expected utility theory in the 18th century the implicit assumption has been that the measure of a position of a probability distribution is given by its expectation value (Jaynes, 2003; Bernoulli, 1738):

$$E(X) = \sum_{i=1}^{n} p_i x_i = H_n(p_1, \ldots, p_n, x_1, \ldots, x_n) \tag{16}$$

But this criterion of choice has proven to be so unsatisfactory that it has given rise to the paradigm of behavioural economics which holds as its central tenet that human decision making does not adhere to the maximization of expectation values (Kahneman, 2011). So, this is why we set out to search for a more appropriate criterion of choice.

5.2.1.2 *The confidence bounds as a position measure*

Now we may imagine a decision problem in which we are only interested in the positions of the probabilistic worst or best case scenarios.

The absolute worst case scenario is:

$$a = \min(x_1, \ldots, x_n) \tag{17}$$

The criterion of choice (17) is also known as the minimax criterion of choice (Lindgren, 1993).

The *k*-sigma lower bound of a given probability distribution is a given as

$$lb(X) = E(X) - k \ \mathrm{std}(\mathrm{X}) \tag{18}$$

where k is the sigma level of the lower bound and where, (16),

$$\mathrm{std}(\mathrm{X}) = \sqrt{\sum_{i=1}^{n} p_i x_i^2 - [E(X)]^2} \tag{19}$$

is the standard deviation. The probabilistic worst case scenario then is given as, (17) and (18):

$$LB(X) = \begin{cases} E(X) - k \ \mathrm{std}(\mathrm{X}), & lb(X) \geq a, \\ a, & lb(X) < a. \end{cases} \tag{20}$$

So, we have that the probabilistic worst case scenario (20) holds the minimax criterion of choice (17) as a special case.

The absolute best case scenario is:

$$b = \max(x_1, \ldots, x_n) \tag{21}$$

The criterion of choice (21) is also known as the maximax criterion of choice.

The k-sigma upper bound of a given probability distribution is a given as:

$$ub(X) = E(X) + k \ \mathrm{std}(\mathrm{X}) \tag{22}$$

where k is the sigma level of the upper bound. The probabilistic best case scenario then is given as:

$$UB(X)=\begin{cases} b, & ub(X)>b, \\ E(X)+k\ \mathrm{std}(X), & ub(X)\leq b. \end{cases} \tag{23}$$

So, we have that the probabilistic best case scenario (30) holds the maximax criterion of choice (21) as a special case.

If we take as our criterion of choice (20) then we only endeavour to minimize our 'losses' and if we take as our criterion of choice (23) then we only endeavour to maximize our 'gains'. A more rational, that is, balanced, criterion of choice would be to make a trade-off between the losses/gains in the probabilistic worst case scenarios (20) and the corresponding gains/losses in the probabilistic best case scenarios (23).

5.2.1.3 *The sum of confidence bounds as a position measure*

If we take as our criterion of choice, (20) and (23),

$$\frac{LB(X)+UB(X)}{2}=\begin{cases} E(X), & lb(X)\geq a,\ ub(X)\leq b, \\ \dfrac{a+E(X)+k\ \mathrm{std}(X)}{2}, & lb(X)<a,\ ub(X)\leq b, \\ \dfrac{E(X)-k\ \mathrm{std}(X)+b}{2}, & lb(X)\geq a,\ ub(X)>b, \\ \dfrac{a+b}{2}, & lb(X)<a,\ ub(X)>b. \end{cases} \tag{24}$$

then it can be shown that we have a position measure which makes a trade-off between the losses/gains in the probabilistic worst case scenarios (20) and the corresponding gains/losses in the probabilistic best case scenarios (23); see Appendix E.

This alternative position measure, as an added benefit, also holds the traditional criterion of choice (16) as a special case, when no undershoot and overshoot of, respectively, the lower and upper sigma confidence bounds occur, as well as Hurwitz's criterion of choice with a balanced pessimism factor of $\alpha=1/2$, when both an undershoot and an overshoot occur. Nonetheless, it may be found that the criterion of choice (24) is vulnerable to a simple counter-example.

Imagine two utility probability distributions having equal lower and upper bounds, but one being right-skewed and the other being left-skewed. Then the criterion of choice (24) will leave us undecided between the two, whereas our intuition would give preference to the decision corresponding with the left-skewed distribution, as the bulk of the probability distribution of the left-skewed distribution will be more to the right than that of the right-skewed distribution.

5.2.1.4 *The sum of confidence bounds and the expectation value as a position measure*

What we seek to maximize in our decision theory is the position of the utility probability distributions; as we have that the decision that puts our utility probability distribution most to the right promises to be the most profitable decision. In this there is little room for manoeuvring. But in our choice of the measure that captures the position of a given probability distribution there is all the more.

Taking a cue from the behavioural economists we have derived as an alternative to (23) the criterion of choice (31) that also takes into account the standard deviation of a given probability distributions, by way of the positions of the under and overshoot corrected lower and upper bounds. But only to find its universality compromised by the simple counter example of a right-skewed and a left-skewed distribution which have the same lower and upper bounds.

Now, also taking a cue from the intuitive results which flow forth from (24) (van Erp *et al.*, 2015), we may 'repair' our criterion of choice (24), albeit in an *ad hoc* fashion, by taking as our position measure for a probability distribution the weighted sum:

$$\frac{LB(X)+E(X)+UB(X)}{3}=\begin{cases} E(X), & lb(X)\geq a,\ ub(X)\leq b, \\ \dfrac{a+2E(X)+k\ \mathrm{std}(X)}{3}, & lb(X)< a,\ ub(X)\leq b, \\ \dfrac{2E(X)-k\ \mathrm{std}(X)+b}{3}, & lb(X)\geq a,\ ub(X)> b, \\ \dfrac{a+E(X)+b}{3}, & lb(X)< a,\ ub(X)> b. \end{cases} \tag{25}$$

For in this criterion of choice we not only take into account the trade-off between the probabilistic worst and best case scenarios, but also the location of the bulk of the probability density in a uni-model probability distribution; thus, accommodating the intuitive preference for the left-skewed distribution of the above counter example.

The position measure (25) is the weighted sum of the positions of, respectively, the probabilistic worst, expected, and best case. The uncorrected lower and upper bounds, (18) and (22), have been traditionally used as simplifying proxies for their generating probability distributions, by way of confidence intervals:

$$\text{proxy} = [lb(X), ub(X)] = \text{uncorrected CI} \tag{26}$$

We, alternatively, take as our simplifying proxy the corrected lower and upper bounds, (19) and (23), and the expectation value (16):

$$\text{proxy} = [LB(X), E(X), UB(X)] \tag{27}$$

Because of the corrections for lower bound undershoot and upper bound overshoot in (17), by way of (19) and (23), we have that for skewed distributions the distance between $E(X)$ and $LB(X)$ may differ from the distance between $UB(X)$ and $E(X)$; thus, reflecting the asymmetry present in these distributions. The position of the generating probability distribution then is taken to be the weighted sum of the positions of the elements of our simple proxy distribution. This then is the rationale behind the criterion of choice (25).

5.2.1.5 *Discussion of the criterion of choice (25)*

In any problem of choice we will endeavour to choose that action which has a corresponding utility probability distribution that is lying most the right on the utility axis; that is, we will choose to maximize our utility probability distributions. In this there is little freedom. But we are free, in principle, to choose the measures of the positions of our utility probability distributions any way we see fit. Nonetheless, we believe that it is always a good policy to take into account all the pertinent information we have.

If we only maximize the expectation values of the utility probability distributions, then we will, by definition, neglect the information that the standard deviations of the utility probability distributions have to bear on our problem of choice, by way of the symmetry breaking in the case of an overshoot of one of the bounds.

Likewise, we are free to only maximize one of the confidence bounds of our utility probability distributions, while neglecting the other. But in doing so, we will be performing probabilistic minimax or maximax analyses, and, consequently, neglect the possibility of either the (catastrophic) losses in the lower bound or the (astronomical) gains in the upper bound.

However, if we only maximize the sum of the lower and upper bound, or a scalar multiple thereof, then we will make a trade-off between the probabilistic worst and best case scenarios. But in the process, we will, for unimodal distributions, be neglecting the location of the bulk of our probability distributions. This is why, in our minds, the scalar multiple the sum of the lower bound, expectation value, and upper bound currently is the most all-round position measure for a given probability distribution, as it reflects the position of the probabilistic worst and best case scenarios, as well as the position of the expected outcome.

5.2.2. Deciding upon the optimal action

We take that action which promises to be most advantageous. If the metric axis is oriented so that a disutility is given a negative metric value and an utility a positive metric value. Then we have that the action that puts the position of the outcome probability distribution the most to the right is the most advantageous decision in terms of the adopted metric.

Positions of probability distributions on the same metric axis may be ordered transitively by way of the sum of their lower and upper bounds (van Erp *et al.*, 2015). The larger the sum of the lower and upper bound of a given utility probability distribution, the more advantageous the action which gives rise to this utility probability distribution. It follows that the cumulants of utility probability distributions allow us to identify that action which will maximize the position of the utility probability distribution; seeing that we construct our lower and upper bounds by way of the cumulants, and seeing that these cumulants, by construction, are conditional on the action under consideration.

In closing, it may very well be that that decision which maximizes the position of the utility probability distribution[20] in, say, a utility transformed monetary metric will be the same decision that

[20] The fundamental maximization principle is uncontested in the decision theoretical field. We quote (Edwards, 1954), a broad and authoritative overview of the decision theoretical field until 1954: "In the literature on

will minimize the position of the utility probability distribution in a utility transformed citizen confidence metric, and vice versa. It will then be up to the decision maker's discretion which metric weighs more heavily in his considerations.

5.3. A short historical digression

We close our discussion of the decision phase of the here proposed risk analysis framework with a short historical overview, that paints the development of this framework. We do this for the benefit of the interested reader, whose expertise does not necessarily lie in the decision theoretical field.

The historical record shows that the notion of 'expectation of profit' was very intuitive to the first workers in probability theory; even than that of probability (Jaynes, 2003). Consider each possibility, $i = 1, \ldots, n$, assign probabilities to them p_i to them, and also assign numbers M_i which represent the 'profit' we would obtain if the *i*th possibility should in fact turn out to be true. Then the expectation of profit[21] is:

$$E(M) = \sum_{i=1}^{n} p_i M_i \tag{28}$$

The prosperous merchants in the 17th century Amsterdam bought and sold expectations as if they were tangible goods. It seemed obvious to many that a person acting in pure self-interest should always behave in such a way as to maximize his expected profit. This, however, led to some paradoxes which led Bernoulli to recognize that the maximization of simple expectation of profit is not always a sensible criterion of action.

For example, suppose that your information leads you to assign probability 0.51 to heads in a slightly biased coin. Now you are given the choice of two actions: (1) your whole wealth M_0 you have on heads for the next coin toss; (2) not to bet at all. According to the criterion of expectation of profit, you should always choose to gamble when faced with this choice. Your expectation of profit, if you do not gamble, is zero; but if you gamble it is

$$E(M) = 0.51 M_0 + 0.49(-M_0) = 0.02 M_0 > 0 \tag{29}$$

Nevertheless it is obvious that nobody would really choose the first alternative. This means that our common sense, in some cases, rejects the criterion of maximized expected profit (Jaynes, 2003).

Bernoulli proposed to resolve these paradoxes by recognition that the true value to a person, of receiving a certain amount of money, is not measured simply by the amount received; it depends also upon how much he already has, by way of the utility function (1):

statistical decision making and the theory of games, various other fundamental principles are considered, but they are all maximization principles of one sort or another.". In the Bayesian decision theory, that what is maximized is the position of the utility, or, if no wealth constraints are in play, outcome probability distributions (van Erp *et al.*, 2015).

[21] Note that in the expectation profit we may recognize the concept of risk as probability times consequence.

$$u(M_i) = q \log \frac{M_0 + M_i}{M_0} . \tag{30}$$

A modern economist is expressing the same idea when he speaks of the 'diminishing marginal utility of money'. So, Bernoulli said that we should recognize that the mathematical expectation of profit (28) is not the same thing as its 'moral expectation', or, equivalently, its expected utility (Bernoulli, 1738):

$$E[u(M)] = \sum_{i=1}^{n} p_i \, u(M_i) = q \sum_{i=1}^{n} p_i \log \frac{M_0 + M_i}{M_0} \tag{31}$$

Bernoulli's expected utility theory drifted in the centuries that followed into obscurity. Even though Bernoulli's utility function (30) was found by Fechner to hold for all sensory perceptions (Fechner, 1860), and, as a consequence, gave rise to psychology as a legitimate exact science, as opposed to the metaphysical discipline it had been until then (Fancher, 1990).

It was only when von Neumann and Morgenstern published their Theory of Games and Economic Behaviour (1944), that expected utility theory again stepped into the scientific limelight, as a theory of choice. However, von Neumann and Morgenstern's version of expected utility left out Bernoulli's utility function (30), which left their theory vulnerable for paradoxes like the ones postulated by Ellsberg and Allais, (Allais, 1953; Ellsberg, 1961).

When it was realised that von Neumann and Morgenstern's expected utility theory failed to adequately model human decision making in certain instances, leading to such paradoxes as the Ellsberg and Allais paradox, the psychologists stepped in. It was then found by these psychologists, who later on called themselves behavioural economists, because their field of research combined behavioural psychology and economics, that, apart from the Ellsberg and Allais paradoxes, expected utility theory was also unable to reproduce the specific convex-down and concave-up shape of the fair probability curves in certainty bets (Kahneman and Tversky, 1992).

Currently, the behavioural economics paradigm of the heuristics and biases approach and cumulative prospect theory is firmly entrenched in the decision theoretic field. Heuristics are mental shortcuts or, equivalently, 'rules of thumb'. It is said that as we don't always have the time or resources to compare all the information at hand we use heuristics to do inference quickly and efficiently. It is postulated by the behavioural economists that most of the times these mental short-cuts will be helpful, but in other cases they lead to systematic errors or cognitive biases, (Kahneman and Tversky, 1972; Kahneman and Tversky, 1973).

In 1979 Kahneman and Tversky complemented their inductive heuristics and biases approach with a decision theoretic prospect theory, (Kahneman and Tversky, 1979). Prospect theory aimed at solving the paradoxes that plagued von Neumann and Morgenstern's expected utility theory. However, this theory was soon found to have its own particular set of serious paradoxes (Mongin, 1997). So, Kahneman and Tversky revised their initial prospect theory in a new cumulative prospect theory, (Tversky and Kahneman, 1992).

Cumulative prospect theory is in structure, though not in implementation, much akin to expected utility theory. In the latter weighted sums, obtained by adding the utility values of outcomes multiplied by their respective probabilities, are compared. In the former weighted sums are also compared, but these weighted sums are obtained by multiplying the utility values, obtained by applying a two part power function to both monetary gains and losses, (Tversky and Kahneman, 1992), by their respective decision weights. These decision weights are transformed probabilities. If the transformation function is the unity function, then the decision weights equal the probabilities themselves, (Fennema and Wakker, 1997), and cumulative prospect theory collapses to von Neumann and Morgernsten's expected utility theory.

However, it has been found that cumulative prospect theory, which is not build from first principles, but, rather, is built from the outset to accommodate the Ellsberg and Allais paradox, as well as the specific convex-down and concave-up shape of the fair probability curves in certainty bets, may be replaced by Bernoulli's original proposal[22] (Bernoulli, 1738), with the adjusted criterion of choice that the confidence bound overshoot corrected position measure of the utility probability distribution (25), should be maximized; rather than the expected utility value (31). This, mathematically trivial adjustment of Bernoulli's expected utility theory[23] – henceforth called Bayesian decision theory – accommodates the experimental results which were in contradiction with von Neumann and Morgenstern's expected utility proposal and, moreover, is built from first principles (van Erp *et al.*, 2014).

We summarize, the first stage of the development of the here proposed risk framework starts with its inception, as expected utility theory, in 1738 by Bernoulli. In the second stage the framework is reintroduced, in a fundamentally altered form, but under the same name, by von Neumann and Morgenstern (1944). In the third stage the von Neumann and Morgenstern version of expected utility theory is challenged by behavioural economists, because of its ability to accommodate experimentally found betting preferences, and the alternative decision theoretical framework of cumulative prospect theory is proposed by Kahneman and Tversky (1992). In the fourth stage it is found that Bernoulli's original 1738 proposal, with an slight adjustment of its criterion of choice accommodates the experimentally found betting preferences, from first principles, thus, offering a viable, because mathematically straightforward[24], alternative for cumulative prospect theory (van Erp *et al.*, 2015).

[22] Bernoulli's original 1738 article was translated, from its original Latin in English, only as late as 1954.

[23] Note that the Ellsberg and some of the Allais paradoxes were already accommodated by Bernoulli's original proposal. Whereas it is the adjusted criterion of choice, where (34) is maximized rather than (42), which allows us to reproduce, from first principles, the specific convex-down and concave-up shape of the fair probability curves, as found in the experimental certainty bets, as well as the remaining Allais paradoxes not accounted for by Bernoulli's expected utility theory.

[24] The here proposed Bayesian risk framework may be mathematical condensed into four mathematical identities: the product and sum rules, equations (1) and (2) of Chapter 3, for the inference phase, with the addition of Bernoulli's utility function (1) and the adjusted criterion of choice (25) which is to be maximized in the decision phase.

6. A Toy-Problem

6.1. Introduction

In this chapter a comparison is made between expected outcome theory, expected utility theory, and Bayesian decision theory. A simple toy-problem is given where the investment willingness to avert a simple type II risk is modelled for the three mathematical decision theories; type II risks are typically High Impact Low Probability (HILP) scenarios.

6.2. An overview of the three decision theories

In what follows we give a short overview of the three decision theories of expected outcome theory, expected utility theory, and the Bayesian decision theory.

Expected Outcome Theory

The algorithmic steps of expected outcome theory are as follows:

(1) For each possible decision construct an outcome probability distributions; i.e. for each possible decision, assign to every conceivable contingency both an estimated net-monetary-consequence and a probability.

(2) Choose that decision which maximizes the expectation values (i.e. means) of the outcome probability distributions.

Expected Utility Theory

The algorithmic steps of expected utility theory are as follows:

(1) For each possible decision construct an outcome probability distributions; i.e. for each possible decision, assign to every conceivable contingency both an estimated net-monetary-consequence and a probability.

(2) Transform outcome probability distributions to their corresponding utility probability distributions; i.e. convert the outcomes of the outcome probability distributions to their corresponding utilities, using Bernoulli's utility function.

(3) Choose that decision which maximizes the expectation values (i.e. means) of the utility probability distributions.

Bayesian Decision Theory

The algorithmic steps of Bayesian decision theory are as follows:

(1) For each possible decision construct an outcome probability distributions; i.e. for each possible decision, assign to every conceivable contingency both an estimated net-monetary-consequence and a probability.

(2) Transform outcome probability distributions to their corresponding utility probability distributions; i.e. convert the outcomes of the outcome probability distributions to their corresponding utilities, using Bernoulli's utility function.

(3) Maximize a scalar multiple of the sum of the lower bound, the expectation value, and the upper bound of the utility probability distributions.

We summarize, expected utility theory differs from expected outcome theory in that the former transforms monetary outcomes to their subjective utilities. However, for an infinite initial wealth we have that expected utility theory reverts back to expected outcome theory; or, equivalently, expected outcome theory is a special instance of the more general expected utility theory.

The concept of utilities was initially derived in order to accommodate the St. Petersburg paradox (Bernoulli, 1738), but we note here that the concept of utilities explains why the insurance company provides insurance policies and why the policy holders buy these policies.

If a risk spells out (near) financial ruin, then policy holders will be willing to pay in excess to the actual objective risks involved, if this transfers this risk to the insurance company. The insurance company takes both the actual objective risk, as well as the excess sum that was paid by the policy holder; the latter being the margin of profit for the insurance company.

Stated in the language of expected utility theory, would the negative event of the negative risk event(s) materialize, then the initial asset position of the policy holder will be proportionally worse affected than the initial asset position of the insurance company; hence, the negative risk event is more risky for the individual policy holder than for the insurance company, as the consequences are proportionally worse for him.

Bayesian decision theory differs from expected utility theory, in that a different criterion of choice is proposed, which is to be maximized. However, for symmetric probability distributions we have that Bayesian decision theory reverts back to expected utility theory; or, equivalently, expected utility theory is a special instance of the more general Bayesian decision theory.

6.3. A Simple Scenario

The Bayesian framework is now applied to a scenario in which a decision maker must decide on much he is willing to invest in order to reduce the probability of a type II risk event occurring. The two decisions under consideration in this simple scenario are:

$D_1 =$ keep the status quo,

$D_2 =$ improve barrier for type II event.

The possible outcomes in the risk scenario remain the same under either decision, and therefore are not dependent upon the particular decision taken. These outcomes are

$O_1 =$ catastrophic type II event occurs,

$O_2 =$ no type II event.

The hypothetical damages associated with these outcomes are,

$O_1 = -x$ euros,

$O_2 = 0$ euros, (1)

respectively, and the investment costs associated with the additional improvement of the type II event barriers are expressed by the parameter

$I =$ investment costs. (2)

The decision whether to improve the type II event barriers or not is of influence on the probabilities of the respective outcomes. Under the decision to make no additional investments in the type II event barriers and keep the status quo, D_1, the probabilities of the outcomes will be, say,

$$P(O_1 \mid D_1) = \theta,$$

$$P(O_2 \mid D_1) = 1 - \theta. \tag{3}$$

Under the decision to improve the type II event barriers, D_2, the probability of the catastrophic type II event will be decreased, say,

$$P(O_1 \mid D_2) = \phi,$$

$$P(O_2 \mid D_2) = 1 - \phi. \tag{4}$$

where $\phi < \theta$. Stated differently, the proposed barrier improvements will decrease the chances of the catastrophic type II event by a factor of $c = \theta/\phi$.

In what follows, the solution of this problem of choice will be given for expected outcome theory, expected utility theory, and Bayesian decision theory. These solutions will be given in terms of variable x, θ, and ϕ, respectively, (1), (3), and (4).

6.3.1. Expected outcome theory solution

The prosperous merchants in the 17th century Amsterdam bought and sold expectations as if they were tangible goods. It seemed obvious to many that a person acting in pure self-interest should always behave so as to maximize his expected profit (Jaynes, 2003).

Combining (1), (2), (3) , and (4), one may construct the outcome probability distributions under the decisions D_1 and D_2:

$$p(O_i \mid D_1) = \begin{cases} \theta, & O_1 = -x, \\ 1-\theta, & O_2 = 0, \end{cases} \tag{5}$$

and

$$p(O_i \mid I, D_2) = \begin{cases} \phi, & O_1 = -x - I, \\ 1-\phi, & O_2 = -I, \end{cases} \tag{6}$$

where in (6) one may explicitly conditionalize on the investment parameter I, which is to be to estimated.

The expected outcomes of these probability distributions are, respectively (Lindgren, 1993):

$$E(O \mid D_1) = -\theta\, x\,, \tag{7}$$

and

$$E(O \mid I, D_2) = -\phi\, x - I \tag{8}$$

The decision theoretical equality

$$E(O \mid D_1) = E(O \mid I, D_2) \tag{9}$$

represents the equilibrium situation, where it will be undecided to choose between the decision to keep the status quo D_1 and the decision to invest in additional barrier improvements D_2. Now, if one solves for I in (9), by way of (7) and (8):

$$I = (\theta - \phi)x \tag{10}$$

then we find that investment quantity where one will be undecided between either decision.

Stated differently, any investment cost smaller than (10) will turn (9) into an inequality, where D_2 becomes more attractive than D_1. It follows that the equilibrium investment (10) is also the maximal investment one will be willing to make in order to improve the type II event barriers.

6.3.2. Expected utility theory solution

For a rich man hundred euros is an insignificant amount of money. So, the prospect of gaining or losing a hundred euros will fail to move the rich man, that is, an increment of hundred euros for him has a utility which tends to zero. For the poor man a hundred euros will be a significant amount of money. So, the prospect of gaining or losing hundred euros will most likely move the poor man to action. It follows that for him an increment of a hundred euros has a utility significantly greater than zero.

In 1738 Daniel Bernoulli derived the utility function for the subjective value of objective moneys by way of a variance argument, in which he considered the subjective effect of a given fixed monetary increment c for two persons holding different initial wealths. Based on this variance argument he derived the utility function of going from an initial asset position x to the asset position $x + c$:

$$u(x, x + c) = q \log \frac{x + c}{x} \tag{11}$$

where q is some scaling constant greater than zero (Bernoulli, 1738; van Erp et al., 2015); in Appendix A an alternative consistency argument for the derivation of Bernoulli's utility function is given.

In expected utility theory the expected values of the utility probability distributions are maximized. Assuming that the decision maker has a total wealth, that is, an actual income and asset portfolio, of

$$M = m \text{ euros,} \tag{12}$$

then, using (11), or, equivalently,

$$U_i = q \log \frac{M + O_i}{M}\,, \tag{13}$$

one may construct from (5) and (6) the corresponding utility probability distributions as:

$$p(U_i \mid D_1) = \begin{cases} \theta, & U_1 = q \log \frac{m-x}{m}, \\ 1-\theta, & U_2 = 0, \end{cases} \tag{14}$$

and

$$p(U_i \mid I, D_2) = \begin{cases} \phi, & U_1 = q \log \frac{m-x-I}{m}, \\ 1-\phi, & U_2 = q \log \frac{m-I}{m}. \end{cases} \tag{15}$$

The expected outcomes of the utility probability distributions are, respectively (Lindgren, 1993):

$$E(U \mid D_1) = q\left(\theta \log \frac{m-x}{m}\right) \tag{16}$$

and

$$E(U \mid I, D_2) = q\left(\phi \log \frac{m-x-I}{m-I} + \log \frac{m-I}{m}\right). \tag{17}$$

The decision theoretical equality

$$E(U \mid D_1) = E(U \mid I, D_2) \tag{18}$$

represents the equilibrium situation, between the decision to keep the status quo D_1 and the decision to invest in additional barriers D_2. Now, if one substitutes (16) and (17) into (18), then one obtains the closed expression for that investment value where one is indifferent between either decision:

$$\log \frac{m-I}{m} = \theta \log \frac{m-x}{m} - \phi \log \frac{m-x-I}{m-I}. \tag{19}$$

Any investment cost smaller than the numerical solution of I in (19) will turn (19) into an inequality, where D_2 becomes more attractive than D_1. It follows that the equilibrium investment (19) is also the maximal investment one will be willing to make to improve the type II event barriers.

6.3.3. Bayesian decision theory solution

In Bayesian decision theory the scaled sum of the confidence bounds and the expectation value of the utility probability distributions is maximized as the risk measure that captures the position of the underlying utility probability distribution (see Section 1.1.7.):

$$R(U \mid D_j) = \frac{LB(U \mid D_j) + E(U \mid D_j) + UB(U \mid D_j)}{3}, \tag{20}$$

where the lower confidence bound is corrected for undershooting the worst possible outcome, say, A:

$$LB(U \mid D_j) = \begin{cases} A, & E(U \mid D_j) - k \ \mathrm{std}(U \mid D_j) < A, \\ E(U \mid D_j) - k \ \mathrm{std}(U \mid D_j), & E(U \mid D_j) - k \ \mathrm{std}(U \mid D_j) \geq A, \end{cases} \tag{21}$$

and the upper confidence bound is corrected for overshooting the best possible outcome, say, B:

$$UB(U \mid D_j) = \begin{cases} E(U \mid D_j) + k \ \mathrm{std}(U \mid D_j), & E(U \mid D_j) + k \ \mathrm{std}(U \mid D_j) \leq B, \\ B, & E(U \mid D_j) - k \ \mathrm{std}(U \mid D_j) > B. \end{cases} \tag{22}$$

Substituting (21) and (22) into (20), one obtains the risk index:

$$R(U \mid D_j) = \begin{cases} E(U \mid D_j), & \text{Neither overshoot nor undershoot,} \\ \dfrac{A + 2E(U \mid D_j) + k \ \mathrm{std}(U \mid D_j)}{3}, & \text{Undershoot and no overshoot,} \\ \dfrac{2E(U \mid D_j) - k \ \mathrm{std}(U \mid D_j) + B}{3}, & \text{Overshoot and no undershoot,} \\ \dfrac{A + E(U \mid D_j) + B}{3}, & \text{Both overshoot and undershoot.} \end{cases} \tag{23}$$

where it is noted that the first row of (23) corresponds with the expected utility theory criterion of choice (Jaynes, 2003); whereas the second and third rows of (23) are the symmetry breaking conditions which arise from skewness in outcome probability distributions and which are responsible for the inverted *S*-shape of certainty bets involving positive and negative prospects, respectively, (van Erp *et al.*, 2015), and the fourth row is a kind of adjusted Hurwitz criterion of choice, which takes into account that two probability distributions may have the same minimal and maximal values while at the same time having an opposite skewness.

In the toy-problem under consideration a simple type II risk scenario is modelled, which is typically a high impact low probability scenario; that is, both large monetary costs and small probabilities for the high-impact event, or, equivalently, on the impact side (1), $x >> 0$ and, on the probability side (3) and (4), $\theta, \phi << 0.5$. Stated differently, the utility probability distributions (14) and (15) under consideration will both be highly skewed to the left and, as a consequence, will lead to the third condition in (23):

$$R(U \mid D_j) = \frac{2E(U \mid D_j) - k \ \mathrm{std}(U \mid D_j) + B}{3}. \tag{24}$$

The best possible outcome under decision D_1 is (14):

$$B = 0, \tag{25}$$

and the standard deviation of (14) is (Lindgren, 1993):

$$\mathrm{std}(U \mid D_1) = -q\sqrt{\theta(1-\theta)} \log \frac{m-x}{m}. \tag{26}$$

So, the risk index under the decision to keep the status quo is, substituting (16), (25), and (26), into (24):

$$R(U \mid D_1) = \frac{q}{3}\left[2\theta + k\sqrt{\theta(1-\theta)}\right] \log \frac{m-x}{m} \tag{27}$$

The best possible outcome under decision D_2 is (15):

$$B = q \log \frac{m-I}{m}, \tag{28}$$

and the standard deviation of (15) is (Lindgren,, 1993):

$$\mathrm{std}(U \mid I, D_2) = -q\sqrt{\phi(1-\phi)}\, \log \frac{m-x-I}{m-I}. \tag{29}$$

So, the risk index under the decision invest in additional barriers is, substituting (17), (28), and (29), into (24):

$$R(U \mid I, D_2) = \frac{q}{3}\left[2\phi + k\,\sqrt{\phi(1-\phi)}\right] \log \frac{m-x-I}{m-I} + q \log \frac{m-I}{m}. \tag{30}$$

The decision theoretical equality

$$R(U \mid D_1) = R(U \mid I, D_2) \tag{31}$$

represents the equilibrium situation, between the decision to keep the status quo D_1 and the decision to invest in additional risk barriers D_2. Now, if one substitutes (27) and (30) into (31), then one obtains the closed expression for that investment value which will leave one undecided:

$$\log \frac{m-I}{m} = \frac{1}{3}\left\{\left[2\theta + k\,\sqrt{\theta(1-\theta)}\right] \log \frac{m-x}{m} - \left[2\phi + k\,\sqrt{\phi(1-\phi)}\right] \log \frac{m-x-I}{m-I}\right\}. \tag{32}$$

Any investment smaller than the numerical solution of I in (32) will turn (32) into an inequality, where D_2 becomes more attractive than D_1. It follows that the equilibrium investment (32) is also the maximal investment one will be willing to make to improve the type II event barriers.

Note that the 'Weber-constant' q has fallen away in both the decision theoretical equalities (19) and (32). This will hold in general, as both the expectation values and standard deviations in (14) and (15) are both linear in the unknown constant q. It follows that one may set, without any loss of generality, q equal to one.

6.4. Some numerical results

In our simple toy-problem we have a decision maker who must decide on how much he is willing to invest in further improvements of his type II risk barriers.

6.4.1. Removing unsafety

After the great Dutch flooding in 1953 the 'Oosterschelde Waterkering' was built. This was a movable dike that allowed for an improved safety from $1/100$ to $1/4000$, while keeping the Oosterschelde connected to the North Sea. This open connection to the North Sea was decided upon in order to keep the salt-sea ecological system of the Oosterschelde river intact.

The total costs of the Oosterschelde Waterkering where about 2.5 billion euros. The bulk of these costs where due to the movable character of this dike. Had the Dutch government decided to build an unmovable dike, then the costs would only have been about 175 million euros.

The total value of the assets at risk at the time where about $1/20$th of the GDP at that time,

$$x = 3.75 \times 10^{9} \text{ euros.} \tag{33}$$

The wealth of the decision maker, that is, the Dutch government, was about 40% of the Dutch GDP at that time, aggregated over a period of five years; five years being the total construction time of the movable Oosterschelde dyke:

$$m = 1.5 \times 10^{11} \text{ euros.} \tag{34}$$

The status quo probability of a catastrophic flooding had right after the great flood been estimated to be, (3):

$$\theta = \frac{1}{100} \tag{35}$$

whereas the probability of the catastrophic flooding under the improved flood defences had been estimated as, (4):

$$\phi = \frac{1}{4000} \tag{36}$$

Substituting the values (33) through (36) into (10), (19), and (32), one obtains the following solutions for the maximal investments I :

- Expected outcome theory:
  - Any sigma level: $I = 36.6 \times 10^{6}$ euros

and

- Expected utility theory:
  - Any sigma level: $I = 37.0 \times 10^{6}$ euros

and

- Bayesian decision theory:
  - 1-sigma level: $I = 129.8 \times 10^{6}$ euros
  - 2-sigma level: $I = 234.9 \times 10^{6}$ euros
  - 3-sigma level: $I = 340.1 \times 10^{6}$ euros

It is noted here that after the great Dutch flood the discussion was not whether to build additional flood defenses or not, but, rather, whether or not to choose for the expensive solution over the `cheap' solution, which would keep the Oosterschelde salt-sea ecosystem intact. Under expected utility theory the cheap solution of an unmovable dyke would have been too expensive by a factor of three, whereas under Bayesian decision theory the cheap solution was well within the 3-sigma bounds.

6.4.2. Maintaining safety

The current total value of the assets at risk in the Oosterschelde region are about $1/20$th of the current GDP, (1):

$$x = 30 \times 10^{9} \text{ euros.} \tag{37}$$

The wealth of the decision maker, that is, the Dutch government, is about 20% of the current Dutch GDP:

$$m = 1.2 \times 10^{11} \text{ euros.} \tag{38}$$

If one assumes the current probability of a catastrophic flooding to be $1/4000$, and if one assumes that in the absence of any maintenance the flood defences will have deteriorated such that the probability of a catastrophic flooding will have doubled to $1/2000$ five years from now. Then $\sqrt[5]{2}$ is the implied 'doubling' one year away from the latest maintenance round. Using this doubling factor of $\sqrt[5]{2}$, the probability of a catastrophic flooding becomes, (3):

$$\theta = \frac{\sqrt[5]{2}}{4000}. \tag{39}$$

If one assumes that the probability of a catastrophic flooding under the flood defence maintenance is our current probability of a catastrophic flooding, (4):

$$\phi = \frac{1}{4000}. \tag{40}$$

Then one has a scenario in which one wishes to prevent a current situation, which is very safe (40), from sliding into a somewhat less safe situation (39).

Substituting the values (37) through (40) into into (10), (19), and (32), one obtains the following solutions for the maximal investments I:

- Expected outcome theory:
  - Any sigma level: $I = 1.1 \times 10^{6}$ euros

and

- Expected utility theory:
  - Any sigma level: $I = 1.3 \times 10^{6}$ euros

and

- Bayesian decision theory:
  - 1-sigma level: $I = 13.9 \times 10^{6}$ euros
  - 2-sigma level: $I = 26.9 \times 10^{6}$ euros
  - 3-sigma level: $I = 39.8 \times 10^{6}$ euros

It is noted here that in order to obtain the very real safety benefit of preventing the probability of a catastrophic flooding of $\phi = 1/4000$ from sliding to $\theta = \sqrt[5]{2}/4000$, expected utility theory is not willing to invest more than 1.3 million euros, whereas Bayesian decision theory with utility transformation, under a 2-sigma safety level, is willing to invest 26.9 million euros for the safety maintenance of the Oosterschelde Waterkering.

So it would seem that Bayesian decision theory solution is more commensurate with observed safety management practices, seeing that the Dutch government yearly spends about 20 million euros to keep the Oosterschelde Waterkering maintained.

6.5. Discussion

In this chapter we have compared expected outcome theory, expected utility theory, and Bayesian decision theory, for a simple toy-problem in which we look at the investment willingness to avert a high impact low probability event.

We have demonstrated that the adjusted criterion of choice, in which scalar multiples of the sum of the lower confidence bound, expectation value, and upper confidence bound of the utility probability distributions are maximized, though mathematical trivial (van Erp et al., 2015), has non-trivial practical implications for the modelled investment willingness.

7. A Discussion of Some Miscellaneous Techniques

In the Description Of Work (DOW) it is stated that influence diagrams, Bayesian networks, event trees, GIS mapping are to be related to the here proposed quantitative risk analysis framework.

7.1. Influence Diagrams

Influence Diagrams admit the same structure as the here proposed risk analysis framework. Influence Diagrams, which find their inception in the AI community, use for their inference phase the Bayesian network methodology – rather than Bayesian probability theory proper – and for their decision phase the expectation value as the criterion of choice which is to be maximized (Gulvanessian *et al.*, 1999) – rather than the criterion of choice (4.25) as recommended in this deliverable.

7.2. Bayesian Networks

It has been found that the Bayesian network methodology does not provide us with any functionality which was not already there in Bayesian probability theory. Moreover, important functionality which is present in the more general Bayesian probability theory (i.e. the handling of arbitrary continuous pdfs, order statistics, and beta-like distributions) is lacking in the more restricted Bayesian network methodology; see Chapter 3 and Appendices A and B.

7.3. Event Trees

Event trees just like Bayesian networks are about the decomposition of some joint distribution of interest in condition probability distributions; see Appendix C. However, Bayesian networks are more flexible, as they allow for the reversal of the flow of inference; see Appendix B. Stated differently, event trees are special instances of the more general Bayesian network methodology.

But seeing that Bayesian networks themselves are only special instances of the more general Bayesian probability theory (see Chapter 3 and Appendices A and B), it then follows – a fortiori – that event trees are special instances of the more general Bayesian probability theory.

7.4. GIS Mapping

GIS maps provide us with data inputs which in the inference phase of the risk analysis framework may be used to come to our outcome probability distributions. For a non-trivial example of the use of GIS data in a probability model we refer to (Gret-Regamey and Straub, 2006).

8. References

Allais M., (1953). “Le Comportement de l'Homme Rationel devant le Risque, Critique des Postulates et Axiomes de l'Ecole Americaine”, Econometrica 21,503-546.

Andreassen S., Jensen F.V., and Olesen K.G., (1991). “Medical Expert Systems Based on Causal Probabilistic Networks”, Int. J. Biomed. Comput. 1-30, Elsevier Scientific ?Publishers Ireland Ltd.

Bernoulli D., (1738). “Exposition of a New Theory on the Measurement of Risk”. Translated from Latin into English by Dr Louise Sommer from `Specimen Theoriae Novae de Mensura Sortis', Commentarii Academiae Scientiarum Imperialis Petropolitanas, Tomus V, 175-192, (1954).

Cheeseman P. (1985). “In Defence of Probability”, In Proc. 9th Int. Joint. Conf. on Artificial Intelligence, Joshi A.K., ed., Morgan Kaufmann, San Francisco, CA.

Cheeseman P. (1988). “An Inquiry Into Computer Understanding”, Comput. Intell. 4. 58-66. See also the following 76 pages of discussion.

Cox R.T., (1946). “Probability, Frequency and Reasonable Expectation”, American Journal of Physics, 14, 1-13.

Edwards W., (1954). “The Theory of Decision Making”, Psychological Bulletin, Vol 51., No. 4.

Ellsberg D., (1961). “Risk, Ambiguity, and the Savage Axioms”, Quarterly Journal of Economics 75, 643-699, (1961).

Fancher R.E., (1990). “Pioneers of Psychology”, W.W. Norton and Company, London.

Fechner G.J., (1860). “Elemente der Psychophysik”, 2 vols.; Vol. 1 translated as Elements of Psychophysics, Boring, E.G. and Howes, D.H., eds. Holt, Rinehart and Winston, New York, (1966).

Fennema H. and Wakker P., (1997). “Original Cumulative Prospect Theory: A Discussion of Empirical Differences”, Journal of Behavioral Decision Making, Vol. 10, 53-64.

Goyal P., Knuth K.H., and Skilling J., (2010). “Origin of Complex Quantum Amplitudes and Feynman's Rules”, Phys Rev. A 81, 022109, arXiv:0907.090 [quant-ph].

Gret-Regamey A. and Straub D., (2006). “Spatially explicit avalanche risk assessment linking Bayesian networks to a GIS”, Nat. Hazards Earth Syst. Sci., 6, 911–926.

Gulvanessian H., Holicky M., Cajot L.G., and Schleich J.B., (1999). “Probabilistic Analysis of Fire Safety Using Bayesian Causal Network”, Safety and Reliability, Schueller and Kalka, eds., Balkema, Rotterdam.

Jaynes E.T., (1973). “Confidence Intervals vs Bayesian Intervals”, Harper W.L. and Hooker C.A., eds. Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Reidel Publishing Co., Dordrecht, Holland, (1976).

Jaynes E.T., (1978). “Where Do We Stand on Maximum Entropy?”, in Levine R.D. and Tribus M., eds., The Maximum Entropy Formalism, M.I.T. Press, Cambridge MA.

Jaynes E.T., (1985). “Bayesian Methods: General Background”, In the Proceedings Volume, “Maximum Entropy and Bayesian Methods in Applied Statistics”, Justice J.H., Ed., Cambridge University Press.

Jaynes E.T., (1990). “Probability Theory as Logic”, Maximum Entropy and Bayesian Methods, Fougere P.F., ed., Kluwer Academic Publishers, Dordrecht, Holland.

Jaynes E.T., (2003). “Probability Theory; the Logic of Science”, Cambridge University Press.

Jeffreys H., (1939). “Theory of Probability Theory”, Clarendon Press, Oxford.

Jonkman S.D., (2007). “Loss of Life Estimation in Flood Risk Assessment”, Delft University of Technology, Delft, Holland.

Kahneman D. and Tversky A., (1972). “Subjective Probability: a Judgment of Representativeness”, Cognitive Psychology 3, 430-454.

Kahneman D. and Tversky A., (1973). “On the Psychology of Prediction”, Psychological Review, Vol. 80, No. 4.

Kahneman D. and Tversky A., (1979). “Prospect Theory: an Analysis of Decision Under Risk”, Econometrica, 47(2), 263-291.

Kahneman D. (2011). “Thinking, Fast, and Slow”, Penguin, Random House, UK.

Knuth K.H., (2010). “A Derivation of Special Relativity from Causal Sets”, arXiv: 1005.4172v2 [math-ph].

Knuth K.H. and Skilling J. , (2010). “Foundations of Inference”, arXiv: 1008.4831v1 [math.PR].

Knuth K.H., (2014). “Information-Based Physics: An Observer-Centric Foundation”. Contemporary Physics, (Invited Submission). doi:10.1080/00107514.2013.853426. arXiv:1310.1667 [quant-ph]

Knuth, K.H. (2014). “Bayesian Evidence and Model Selection”, arXiv:1411.3013v1 [stat.ME].

Lindgren B.W., (1993). “Statistical Theory”, Chapman \& Hall, Inc., New York.

Lusted, L. (1968). “Introduction to Medical Decision Making”, C.C. Thomas, Springfield, IL.

Masin S.C., Zudini V., and Antonelli M., (2009). “Early Alternative Derivations of Fechner's Law”, Journal of Behavioral Sciences, 45, 56-65.

Mongin P, (1997). “Expected Utility Theory”, Prepared for the Handbook of Economic Methodology (slightly longer version than the published one), Davis J., Hands W., and Maki U., eds., London, Edward Elgar.

Pearl J.G., (1988). “Probabilistic Reasoning: Networks of plausible Inference”, Morgan Kaufmann, San Francisco, CA.

Prak, P. (2009). ‘Onderzoek naar de toepassing van similarity judgment bij het vaststellen van alerteringslocaties binnen de spoorsector’, SSM thesis, TU Delft.

Shannon C.E., (1948). “A Mathematical Theory of Communication”, Bell Sys. Tech. J., 27, 379-423, 623-56.

Sivia D.S., (1996). “Data Analysis: A Bayesian Tutorial”, Clarendon Press, Oxford.

Skilling J., (2005). “Bayesics”, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, eds. Knuth, K.H., Abbas A.E., Morris R.D., and Castle J.P., American Institute of Physics.

Skilling J., (2008). “This Physicist’s View of Gelman’s Bayes”, Bayesian Analysis, (also) to be found on the internet.

Stevens S.S., (1961). “To Honour Fechner and Repeal His Law”, Science, New Series, Vol. 133, No. 3446, 80-86.

Treasury Board of Canada Secretariat (1998). “Benefit-Cost Analysis Guide”, draft July 1988, Ottawa.

van Erp H.R.N., Linger R.O., and van Gelder P.H.A.J.M., (2015). “Fact Research on the Bayesian Decision Theory”, arxiv.

von Neumann J. and Morgenstern O., (1944). “Theory of Games and Economic Behavior”, Princeton, Princeton Univer. Press.

Zadeh L.A., (1973). “Outline of a New Approach to the Analysis of Complex Systems”, Trans. On Systems, Man and Cybernetics SMC-3, 28-44.

Zellner, A. (1971). “An Introduction to Bayesian Inference in Econometrics”, J. Wiley & Sons, Inc., New York.

9. Appendix A: A Non-Technical Introduction into Bayesian Networks

9.1. Introduction

We give a short non-technical overview of our research into Bayesian networks paradigm (not to be confused with Bayesian probability theory).

9.2. AI Bayesian Networks

While researching Bayesian Networks, also called Causal Networks, we have found that these networks, at best, are nothing but a suboptimal implementation of Bayesian probability theory. Pearl, the founder of the Bayesian Network paradigm, recognized the conceptual strength of Bayesian probability theory, we quote (Pearl, 1988, p.20):

> "We take for granted that probability calculus is unique in the way it handles context-dependent information and that no competing calculus exists that closely covers so many qualitative aspects of plausible reasoning."

But seeing that Pearl then states:

> "So, the calculus is worthy of exploitation, emulation, or at the very least, serious exploitation. We therefore take probability calculus as an initial model for human reasoning from which more refined models may originate, if needed. By exploring the limits of using probability calculus in machine implementations of plausible inference, we hope to identify conditions under which extensions, refinements, and simplifications are warranted."

where 'probability calculus' points to the Bayesian probability theory, we feel trouble brewing. And indeed we then read:

> "Obviously, there are applications where strict adherence to the dictates of probability theory would be computationally infeasible, and there compromises will have to be made. Still, we find it more comfortable to compromise an ideal theory that is well understood than to search for a new surrogate theory, with only gut feeling for guidance."

So, we may summarize as follows. Pearl, on a conceptual level, saw the Bayesian probability calculus as the ideal, but on an implementation level he felt there to be a problem, seeing that he mentions an obvious computationally infeasibility. Now, what constitutes this 'obvious computationally infeasibility'?

Say, we have 26 variables $A_a, B_b, C_c, \ldots, Z_z$, all admitting ten values, that is, having sub-indices

$$a = 1, \ldots, 10, \qquad b = 1, \ldots, 10, \qquad \ldots \qquad z = 1, \ldots, 10.$$

If we connect these variables by the conditional inference chain:

$$P(A_a, B_b, C_c, \ldots, Z_z) = P(A_a) P(B_b \mid A_a) P(C_c \mid B_b) \cdots P(Z_z \mid Y_y). \tag{1}$$

Then the probability distribution $P(A_a, B_b, C_c, ..., Z_z)$ has an 26-dimensional parameter space admitting $26^{10} = 1.4 \times 10^{14}$ distinct values.

Now, if we want to find the conditional probability of $P(S_s \mid Q_q)$, then we first must marginalize the left-hand side of (1) over all variables, save Q_q and S_s, that is:

$$P(Q_q, S_s) = \sum_{\notin\{q,s\}} P(A_a, ..., Q_q, \ldots, S_s, \ldots, Z_z), \tag{2}$$

We then marginalize the left-hand-side of (2) over the variable S_s, that is:

$$P(Q_q) = \sum_{s} P(Q_q, S_s) \tag{3}$$

Bayes theorem then states that

$$P(S_s \mid Q_q) = \frac{P(Q_q, S_s)}{P(Q_q)} \tag{4}$$

Now in going from (1) to (2), we have to make

$$(26 - 2)^{10} = 24^{10} = 6.3 \times 10^{13} \tag{5}$$

summations, which is a computationally prohibitive value. This is the 'obvious computationally infeasibility' mentioned by Pearl (1988, p.20).

In order to accommodate this infeasibility an ideal theory is comprised, as may be read in the last quote, The compromising of the ideal of Bayesian probability theory then constitutes the AI class of models, generally known as Bayesian networks. But any Bayesian could have told Pearl that one may evaluate (3) sequentially as:

$$
\begin{aligned}
P(Q_q, S_s) &= \sum_{\notin\{q,s\}} P(A_a) P(B_b \mid A_a) P(C_c \mid B_b) \cdots P(Z_z \mid Y_y) \\
&= \sum_{\notin\{q,s\}} \sum_{a} [P(A_a) P(B_b \mid A_a)] P(C_c \mid B_b) \cdots P(Z_z \mid Y_y) \\
&= \sum_{\notin\{q,s\}} \sum_{a} P(A_a, B_b) P(C_c \mid B_b) \cdots P(Z_z \mid Y_y) \\
&= \sum_{\notin\{q,s\}} P(B_b) P(C_c \mid B_b) P(D_d \mid C_c) \cdots P(Z_z \mid Y_y) \\
&= \sum_{\notin\{q,s\}} \sum_{b} [P(B_b) P(C_c \mid B_b)] P(D_d \mid C_c) \cdots P(Z_z \mid Y_y) \\
&= \sum_{\notin\{q,s\}} \sum_{b} P(B_b, C_c) P(D_d \mid C_c) \cdots P(Z_z \mid Y_y) \\
&= \sum_{\notin\{q,s\}} P(C_c) P(D_d \mid C_c) \cdots P(Z_z \mid Y_y) \\
&= etc...
\end{aligned}
\tag{6}
$$

which brings us in

$$24 \times 10 \times 10 = 2400 \tag{7}$$

summations to the needed probability distribution (2).

So, marginalizing the left-hand side of (1) over all variables, save Q_q and S_s, if done sequential, (6), may be done in 2400 summations, (7), as opposed to the bulk evaluation, (2), which would require an unfeasible 6.3×10^{13} summations, (5). So, there is no need to compromise the ideal[25] and, in the process, introduce an added layer of complexity by way of graph theory and all kinds of exotic parameter estimation methods.

9.3. Bayesian Bayesian Networks

Making use of sequential updating (6), we have developed for this work package a Bayesian Bayesian Network as opposed to an AI Bayesian Network. This is a trivial thing to do. One only has to capture

[25] As an aside, in their deviating from the ideal the AI community have allowed for unspecified nodes. These unspecified nodes are purely structural in that they are empty of content but restrict the flow of probability propagation, thus, leading to a numerical value of the unspecified nodes that are not marginalized, that is, summated away. This latent variable procedure is both suspect and not pertinent for the RAIN project.

in an algorithm the computational steps that a Bayesian performs routinely whenever he does an inference problem.

It has been found that are no more than ten guiding heuristics in a Bayesian analysis. So, in RAIN we will introduce Bayesian BNs. We would like to emphasize here that these Bayesian BNs are only an automation of the application of Bayesian probability theory. In other words, the use of Bayesian BNs are equivalent to a straight forward application of Bayesian probability theory.

Stated differently, the Bayesian network methodology does not provide us with anything which is not already available to us in the more general Bayesian probability theory. Moreover, Bayesian probability theory may provide us with 'tools' which are not available in the more restricted Bayesian network methodology; examples of such tools are the use of continuous probability distributions, order statistics, principles to construct probability distributions, beta-like distributions, etc….

This then is why we recommend to use the Bayesian probability theory as the inferential 'methodology' for the RAIN project, rather than Bayesian networks.

10. Appendix B: A More Technical Introduction into Bayesian Networks

10.1. Bayesian Probability Theory

Bayesian probability theory combines probabilities by way of the Product Rule and the Sum Rule, respectively:

$$p(A \mid C)\ p(B \mid A) = p(AB \mid C) = p(B \mid C)\ p(A \mid B) \tag{1}$$

and

$$p(A \mid C) + p(\overline{A} \mid C) = 1 \tag{2}$$

The strict adherence to these two rules, as the necessary and sufficient operators for probabilities are what defines a Bayesian[26].

The Product and Sum rules may be generalized to apply to distributions of probabilities. Let $p(A_i \mid C_k)$ be a collection of probability distributions of the set of propositions $\{A_i\}$, conditional on the propositions $C_1, \ldots, C_K$, then it may be found that (1) and (2) generalize to, respectively,

$$p(A_i \mid C_k)\ p(B_j \mid A_i) = p(A_i B_j \mid C_k) = p(B_j \mid C_k)\ p(A_i \mid B_j), \tag{3}$$

where $p(A_i B_j \mid C_k)$ is a collection of bivariate probability distributions of the set of proposition $\{A_i B_j\}$, conditional on the propositions $C_1, \ldots, C_K$, and

$$\sum_{i=1}^{I} \sum_{j=1}^{J} p(A_i B_j \mid C_k) = \sum_{i=1}^{I} p(A_i \mid C_k) = 1 \ , \tag{4a}$$

or, equivalently,

$$\sum_{i=1}^{I} \sum_{j=1}^{J} p(A_i B_j \mid C_k) = \sum_{j=1}^{J} p(B_j \mid C_k) = 1 . \tag{4b}$$

The product rule (3) allows us to put sets of propositions, or, equivalently, variables, like $\{A_i\}$ and $\{B_j\}$ that are placed behind the conditionality sign before the conditionality sign. The sum rule (4) then allows to summate any set of proposition which is placed before the conditionality sign away as 'nuisance' parameters.

10.2. Graphical Representation of Chains of Inference

[26] The product and sum rules tell us how to combine probabilities once we have assigned them, but it tells us not how to assign them. Nonetheless, Bayesians also concern themselves how to translate our state of information to probabilities, by way of such principles as the transformation group and maximum entropy (MaxEnt) principles. But this probability assignment part of Bayesian practice is not pertinent to the subject at hand, that of Bayesian networks.

If the probability distribution of, say, the set of propositions $\{A_i\}$ is conditional dependent upon the propositions $\{B_j\}$, whose probability distributions in their turn are again dependent upon the propositions $\{C_k\}$, and if the probability distribution of the $\{C_k\}$ is not dependent on any other proposition, then we have an inference chain consisting of the elements: $p(A_i \mid B_j)$, $p(B_j \mid C_k)$, and $p(C_k)$, where $p(C_k)$ connects with $p(B_j \mid C_k)$, and $p(B_j \mid C_k)$ connects with $p(A_i \mid B_j)$.

We may represent, if we so wish, this chain of inference, as a graphical heuristic, by way of nodes and arrows:

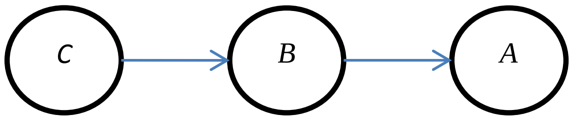


Figure B.1: a chain of inference

Figure 1 tells us at a glance that the (trivariate) joint probability distribution of the proposition elements $\{A_i B_j C_k\}$ is obtained by way the product, (3):

$$p(A_i B_j C_k) = p(C_k)\ p(B_j \mid C_k)\ p(A_i \mid B_j). \tag{5}$$

Figure 1 then becomes:

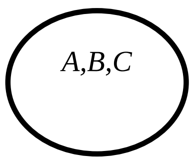


Figure B.2: a joint node

It follows that the product rule (3) allows us to fold the separate nodes of Figure 1 into one joint node.

10.3. Reversing the Flow of Inference

In what follows we will derive the nodes which reverse the flow of inference in Figure 1, by way of the product and sum rules, (3) and (4). The reversal of the flow of inference is one of the defining qualities of Bayesian probability theory, as it allows one to go from the likelihood (i.e. the probability distribution of the observed data given the value of some parameter of interest) to the posterior (i.e. the probability distribution of the value of some parameter of interest given the observed data).

We apply the sum rule (4) to (5) find the marginal distribution of $\{A_i\}$:

$$p(A_i) = \sum_{k=1}^{K}\sum_{j=1}^{J} p(A_i B_j C_k), \tag{6}$$

the marginal distribution of $\{B_j\}$:

$$p(B_j) = \sum_{k=1}^{K}\sum_{i=1}^{I} p(A_i B_j C_k), \tag{7}$$

the marginal distribution of the conjunction of $\{A_i\}$ and $\{B_j\}$:

$$p(A_i B_j) = \sum_{k=1}^{K} p(A_i B_j C_k), \tag{8}$$

and the marginal distribution of the conjunction of $\{B_j\}$ and $\{C_k\}$:

$$p(B_j C_k) = \sum_{i=1}^{I} p(A_i B_j C_k). \tag{9}$$

By way of the product rule (3), or, equivalently, Bayes' Theorem, we then combine (6) and (8) into the conditional distribution:

$$p(B_j \mid A_i) = \frac{p(A_i B_j)}{p(A_i)}, \tag{10}$$

and (7) and (9) into the conditional distribution:

$$p(C_k \mid B_j) = \frac{p(B_j C_k)}{p(B_j)} . \tag{11}$$

In (6), (10), and (11), we now have the nodes which reverse the flow of inference in Figure 1:

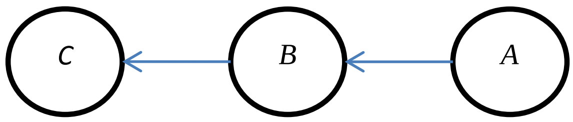


Figure B.3: a reversed chain of inference relative to Figure 1

As an aside Figure 3 tells us again at a glance that the (trivariate) joint probability distribution of the proposition elements $\{A_i B_j C_k\}$ is obtained by way the product, (3):

$$p(A_i B_j C_k) = p(A_i) p(B_j \mid A_i) p(C_k \mid B_j), \tag{12}$$

10.4. Removing Nuisance Variables

In what follows we will demonstrate how to remove nodes from our chain of inference which are not directly of interest. The removal of ‘nuisance’ nodes is done by way of the sum rule (4), and is another defining quality of interest of the Bayesian probability theory, as it allows one to remove in a principled way variables of which the exact parameter values are unknown.

Say, we want remove the intermediate node *B* in Figure 1, in order to obtain the alternative chain of inference in Figure 4:

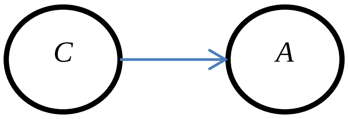


Figure B.4: a collapsed chain of inference relative to Figure 1

There are two ways of going from the chain of inference in Figure 1 to the chain of inference in Figure 4. One of these ways will become computationally infeasible as the number of intermediate nodes between *C* and *A* goes from just the one node in Figure 1 to *N* nodes :

Figure B.5: an expanded chain of inference relative to Figure 1

Whereas the other way will remain computational feasible as the number of intermediate nodes *N* grows.

10.4.1. The Computational Inefficient Method

In the computational inefficient method we go to the chain of inference in Figure 4, by way of the joint distribution node in Figure 2, or, equivalently, (5):

$$p(A_i B_j C_k) = p(C_k)\, p(B_j \mid C_k)\, p(A_i \mid B_j). \tag{13}$$

Using the sum rule we may obtain from (13):

$$p(A_i C_k) = \sum_{j=1}^{J} p(A_i B_j C_k) \tag{14}$$

and from (14):

$$p(C_k) = \sum_{i=1}^{I} p(A_i C_k) \tag{15}$$

By way of the product rule (3), or, equivalently, Bayes' Theorem, we then may combine the nodes (14) and (15) in order to obtain the new marginalized conditional node:

$$p(A_i \mid C_k) = \frac{p(A_i C_k)}{p(C_k)} \tag{16}$$

The nodes (15) and (16) then are the nodes of the desired chain of inference in Figure 4.

Now would we have that the joint distribution node would be constructed from the partitioning in Figure 5:

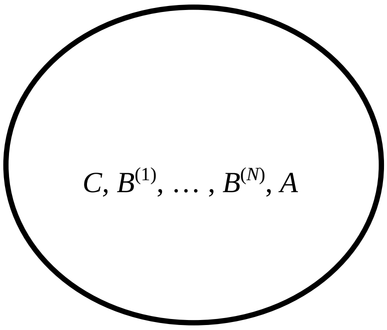


Figure B.6: the joint node of the chain of inference in Figure 5

then (13) would go to

$$p\left(A_i B_{j_1}^{(1)} \cdots B_{j_N}^{(N)} C_k\right) = p(C_k)\, p\left(B_{j_1}^{(1)} \mid C_k\right) \cdots p\left(B_{j_k}^{(k)} \mid B_{j_{k-1}}^{(k-1)}\right) \cdots p\left(A_i \mid B_{j_N}^{(N)}\right), \tag{17}$$

whereas (14) would go to

$$p(A_i C_k) = \sum_{j_N=1}^{J_N} \cdots \sum_{j_1=1}^{J_1} p\left(A_i B_{j_1}^{(1)} \cdots B_{j_N}^{(N)} C_k\right) \tag{18}$$

As the evaluation of (18) requires $J_1 \times \cdots \times J_N$ summations, this approach becomes quickly computationally infeasible as the number of nuisance nodes, *N*, increases.

10.4.2. The Computational Efficient Method

In the computational efficient method we first lose the second multiplication 'sign' on the right-hand side of (5):

$$p(C_k)\left[p(B_j \mid C_k)\, p(A_i \mid B_j)\right] = p(C_k)\, p(A_i B_j \mid C_k) \tag{19}$$

The probability distributions in the right-hand side of (19) correspond with the chain of inference:

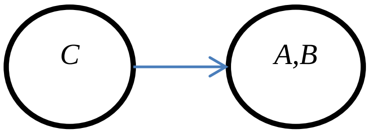


Figure B.7: a graphical representation of the right-hand side of (19)

In (19) we now may summate out variable *B*, by way of the sum rule (4):

$$p(C_k)\ p(A_i \mid C_k) = p(C_k) \sum_{j=1}^{J} p(A_i B_j \mid C_k) \tag{20}$$

In (20) we then have the probability distributions which correspond with the nodes in Figure 5.

If we have a large joint node, like in Figure 6, then in the computational efficient method to remove nuisance variables, all the nuisance nodes are removed sequentially. This is done as follows, first we combine the nodes which are dependent on a given nuisance node, using the product rule, as was done in (19), then we summate out this nuisance node, using the sum rule, as was done in (20).

For example, we have that (17), by way of the product rule (3), is equivalent to:

$$\begin{aligned} p\left(A_i B_{j_1}^{(1)} \cdots B_{j_N}^{(N)} C_k\right) &= p(C_k)\ p\left(B_{j_1}^{(1)} \mid C_k\right) p\left(B_{j_2}^{(2)} \mid B_{j_1}^{(1)}\right) \cdots p\left(A_i \mid B_{j_N}^{(N)}\right) \\ &= p(C_k)\ p\left(B_{j_1}^{(1)} B_{j_2}^{(2)} \mid C_k\right) p\left(B_{j_3}^{(3)} \mid B_{j_2}^{(2)}\right) \cdots p\left(A_i \mid B_{j_N}^{(N)}\right) \end{aligned} \tag{21}$$

From (21) we may summate out the node $B^{(1)}$, by way of the sum rule (4):

$$\begin{aligned} p\left(A_i B_{j_2}^{(2)} \cdots B_{j_N}^{(N)} C_k\right) &= \sum_{j_1=1}^{J_1} p\left(A_i B_{j_1}^{(1)} B_{j_2}^{(2)} \cdots B_{j_N}^{(N)} C_k\right) \\ &= p(C_k) \left[\sum_{j_1=1}^{J_1} p\left(B_{j_1}^{(1)} B_{j_2}^{(2)} \mid C_k\right) \right] p\left(B_{j_3}^{(3)} \mid B_{j_2}^{(2)}\right) \cdots p\left(A_i \mid B_{j_N}^{(N)}\right) \\ &= p(C_k)\ p\left(B_{j_2}^{(2)} \mid C_k\right) p\left(B_{j_3}^{(3)} \mid B_{j_2}^{(2)}\right) \cdots p\left(A_i \mid B_{j_N}^{(N)}\right) \end{aligned} \tag{22}$$

We have that (22), by way of the product rule (3), is equivalent to:

$$\begin{aligned} p\left(A_i B_{j_2}^{(2)} \cdots B_{j_N}^{(N)} C_k\right) &= p(C_k)\ p\left(B_{j_2}^{(2)} \mid C_k\right) p\left(B_{j_3}^{(3)} \mid B_{j_2}^{(2)}\right) \cdots p\left(A_i \mid B_{j_N}^{(N)}\right) \\ &= p(C_k)\ p\left(B_{j_2}^{(2)} B_{j_3}^{(3)} \mid C_k\right) p\left(B_{j_4}^{(4)} \mid B_{j_3}^{(3)}\right) \cdots p\left(A_i \mid B_{j_N}^{(N)}\right) \end{aligned} \tag{23}$$

From (23) we may summate out the node $B^{(2)}$, by way of the sum rule (4):

$$p\left(A_i B_{j_3}^{(3)} \cdots B_{j_N}^{(N)} C_k\right) = \sum_{j_2=1}^{J_2} p\left(A_i B_{j_2}^{(2)} B_{j_3}^{(3)} \cdots B_{j_N}^{(N)} C_k\right)$$

$$= p(C_k)\left[\sum_{j_2=1}^{J_2} p\left(B_{j_2}^{(2)} B_{j_3}^{(3)} \mid C_k\right)\right] p\left(B_{j_4}^{(4)} \mid B_{j_3}^{(3)}\right) \cdots p\left(A_i \mid B_{j_N}^{(N)}\right) \tag{24}$$

$$= p(C_k)\, p\left(B_{j_3}^{(3)} \mid C_k\right) p\left(B_{j_4}^{(4)} \mid B_{j_3}^{(3)}\right) \cdots p\left(A_i \mid B_{j_N}^{(N)}\right)$$

etcetera.

Note that in the computational efficient method of nuisance node removal the evaluation of the bulk joint node in Figure 6 is replaced by a sequence of evaluations of sub-joint nodes (e.g. the right-hand joint node *AB* of Figure 4). The evaluation of this sequence of sub-joint nodes will be computational feasible as this evaluation is linear in the number of nuisance nodes *N*; in contrast, the evaluation of (18) will be exponential in the number of nuisance nodes *N*.

10.4.3. The MATLAB algorithm made for RAIN

The computational infeasibility of the computational inefficient method of removing nuisance nodes is why the AI community set their Bayesian Networks apart from Bayesian probability theory proper and introduce their own independence language (Pearl, 1988).

Bayesian Networks from the AI community use junction tree algorithms for the removal of nuisance nodes (Andreassen *et al.*, 1991), which are computational efficient in their evaluation of (18). However, junction trees need not be invoked in order to come to a computational efficient evaluation of (18).

In order to demonstrate this a Bayesian network algorithm along purely lines was developed in MATLAB. This algorithm performs the following four algorithmic steps:

(a) Connect separate nodes into joint nodes by way the product rule (3),
(b) Remove nodes by way of the sum rule (4)
(c) Sequentially eliminate the nuisance nodes in a given collection of nodes in such an order that a minimal number of summations in step (b),
(d) Apply Bayes' Theorem to the marginalized joint distributions obtained in (c).

11. Appendix C: Event Trees as an Application of Bayesian networks

In Bayesian networks, which is just Artificial Intelligence community new-speak (Pearl, 1988) for Bayesian probability theory (Laplace, 1784), the entity of interest is the joint probability distribution.

In Bayesian probability theory it is understood that a probability distribution is an information carrier of our state of knowledge (Jaynes, 2003). Humans, as a rule, have to cope with incomplete information. Stated differently, our state of knowledge is fuzzy. This fuzziness is captured by assigning numerical plausibilities to the possible outcomes in our hypothesis space.

If our outcomes, say, X and Y are related somehow, then the joint probability distribution on these outcomes will capture this dependency, by way of a mixed first moment which does not factor as the product of the respective first moments:

$$\begin{aligned} E(XY) &= \sum_i \sum_j X_i Y_j p(X_i, Y_i) \\ &\neq \left(\sum_i X_i \sum_j p(X_i, Y_i)\right)\left(\sum_j Y_j \sum_i p(X_i, Y_i)\right) \\ &= E(X)E(Y) \end{aligned} \quad (1)$$

The whole Bayesian Network methodology of the nodes (e.g., circles) and dependencies (e.g., arrows) are just a decomposition of the joint probability distribution, which, as a rule, is too complex to assign in one go, into smaller conditional probability components, which may be assigned in one go; see Appendix B.

Quantified event trees, a mainstay of engineering practice, are an example of joint probability distributions that are constructed by way of this divide-and-conquer approach. For example, the quantified event tree below – provided to us by one of our colleague work packages – translates to the following (implied) collection of conditional probability distributions.

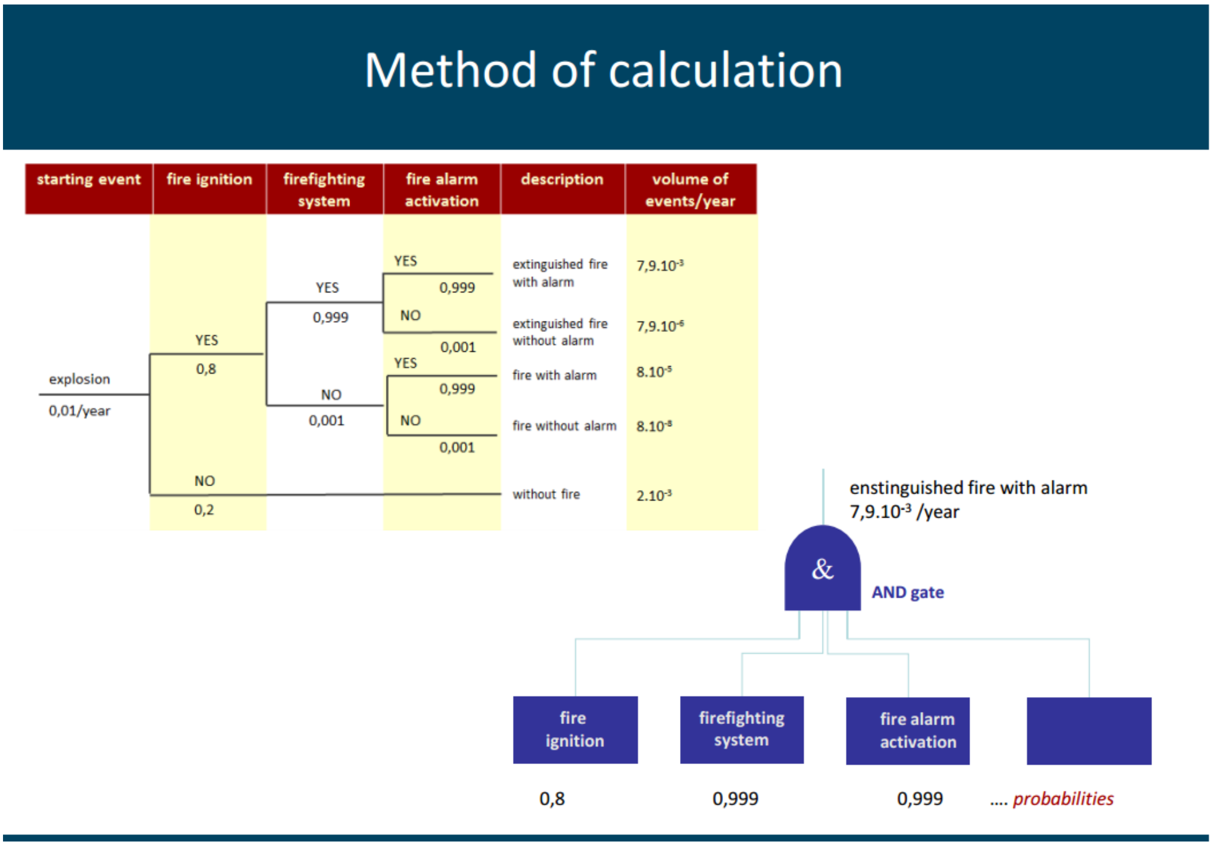


Figure C.1: Example of an event tree

Let

$$W = \text{explosion} \qquad \overline{W} = \text{no explosion}$$
$$X = \text{fire ignition} \qquad \overline{X} = \text{no fire ignition}$$
$$Y = \text{fire fighting system} \quad \text{and} \quad \overline{Y} = \text{no fire fighting system}$$
$$Z = \text{fire alarm activation} \qquad \overline{Z} = \text{no fire alarm activation}$$

Then, working from left to right:

$$\begin{cases} P(W) = 0.01 \\ P(\overline{W}) = 0.99 \end{cases}$$

$$\begin{cases} P(X \mid W) = 0.8 \\ P(\overline{X} \mid W) = 0.2 \end{cases} \qquad \begin{matrix} P(X \mid \overline{W}) = 0.0 \\ P(\overline{X} \mid \overline{W}) = 1.0 \end{matrix}$$

$$\begin{cases} P(Y \mid X) = 0.999 \\ P(\overline{Y} \mid X) = 0.001 \end{cases} \qquad \begin{matrix} P(Y \mid \overline{X}) = 0.0 \\ P(\overline{Y} \mid \overline{X}) = 1.0 \end{matrix}$$

$$\begin{cases} P(Z \mid Y, X) = 0.999 & P(Z \mid \overline{Y}, X) = 0.999 \\ P(\overline{Z} \mid Y, X) = 0.001 & P(\overline{Z} \mid \overline{Y}, X) = 0.001 \\ & \\ P(Z \mid Y, \overline{X}) = 0.0 & P(Z \mid \overline{Y}, \overline{X}) = 0.0 \\ P(\overline{Z} \mid Y, \overline{X}) = 1.0 & P(\overline{Z} \mid \overline{Y}, \overline{X}) = 1.0 \end{cases} \tag{2}$$

Now, having decomposed our joint probability distribution in more manageable conditional probability distributions we inspect the conditional building blocks and check if these quantified assumptions are commensurate with our common sense judgements.

For example, in the first two rows of the fourth block we see that the probability of a fire alarm going off is independent of the probability of the firefighting system extinguishing the fire, which seems reasonable enough. Because of this independency we may simplify the fourth block of (2):

$$\begin{cases} P(Z \mid Y, X) = 0.999 & P(Z \mid \overline{Y}, X) = 0.999 \\ P(\overline{Z} \mid Y, X) = 0.001 & P(\overline{Z} \mid \overline{Y}, X) = 0.001 \\ & \\ P(Z \mid Y, \overline{X}) = 0.0 & P(Z \mid \overline{Y}, \overline{X}) = 0.0 \\ P(\overline{Z} \mid Y, \overline{X}) = 1.0 & P(\overline{Z} \mid \overline{Y}, \overline{X}) = 1.0 \end{cases} \tag{3}$$

by dropping the reference to either Y or $\overline{Y}$:

$$\begin{cases} P(Z \mid X) = 0.999 \\ P(\overline{Z} \mid X) = 0.001 \\ \\ P(Z \mid \overline{X}) = 0.0 \\ P(\overline{Z} \mid \overline{X}) = 1.0 \end{cases} \tag{4}$$

which in a Bayesian Network mnemonic would correspond with simplifying:

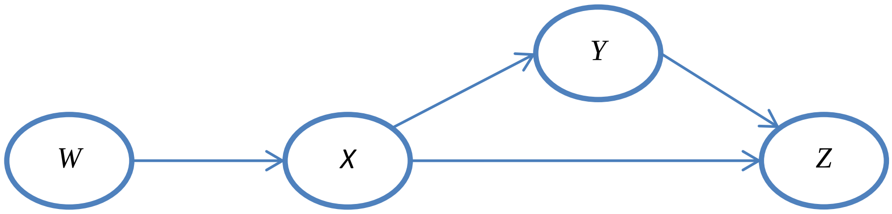


Figure C.2: Event Tree represented with nodes (2)

to

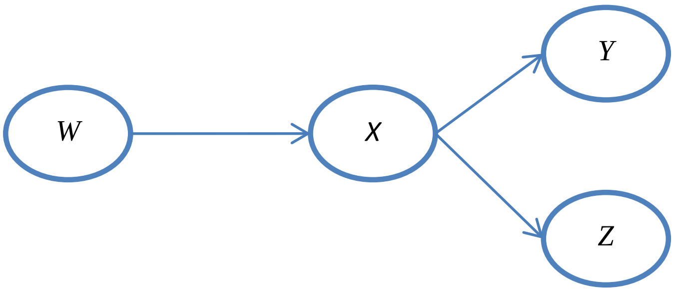


Figure C.3: Event Tree represented with simplified node (4)

The joint probability distribution the entity of interest in the Bayesian Network analysis, then is, by way of the product rule:

$$
\begin{aligned}
P(W,X,Y,Z) &= P(W)P(X\mid W)P(Y\mid X)P(Z\mid X)\\
P(\overline{W},X,Y,Z) &= P(\overline{W})P(X\mid \overline{W})P(Y\mid X)P(Z\mid X)\\
P(W,\overline{X},Y,Z) &= P(W)P(\overline{X}\mid W)P(Y\mid \overline{X})P(Z\mid \overline{X})\\
P(\overline{W},\overline{X},Y,Z) &= P(\overline{W})P(\overline{X}\mid \overline{W})P(Y\mid \overline{X})P(Z\mid \overline{X})\\
P(W,X,\overline{Y},Z) &= P(W)P(X\mid W)P(\overline{Y}\mid X)P(Z\mid X)\\
P(\overline{W},X,\overline{Y},Z) &= P(\overline{W})P(X\mid \overline{W})P(\overline{Y}\mid X)P(Z\mid X)\\
P(W,\overline{X},\overline{Y},Z) &= P(W)P(\overline{X}\mid W)P(\overline{Y}\mid \overline{X})P(Z\mid \overline{X})\\
P(\overline{W},\overline{X},\overline{Y},Z) &= P(\overline{W})P(\overline{X}\mid \overline{W})P(\overline{Y}\mid \overline{X})P(Z\mid \overline{X})\\
P(W,X,Y,\overline{Z}) &= P(W)P(X\mid W)P(Y\mid X)P(\overline{Z}\mid X)\\
P(\overline{W},X,Y,\overline{Z}) &= P(\overline{W})P(X\mid \overline{W})P(Y\mid X)P(\overline{Z}\mid X)\\
P(W,\overline{X},Y,\overline{Z}) &= P(W)P(\overline{X}\mid W)P(Y\mid \overline{X})P(\overline{Z}\mid \overline{X})\\
P(\overline{W},\overline{X},Y,\overline{Z}) &= P(\overline{W})P(\overline{X}\mid \overline{W})P(Y\mid \overline{X})P(\overline{Z}\mid \overline{X})\\
P(W,X,\overline{Y},\overline{Z}) &= P(W)P(X\mid W)P(\overline{Y}\mid X)P(\overline{Z}\mid X)\\
P(\overline{W},X,\overline{Y},\overline{Z}) &= P(\overline{W})P(X\mid \overline{W})P(\overline{Y}\mid X)P(\overline{Z}\mid X)\\
P(W,\overline{X},\overline{Y},\overline{Z}) &= P(W)P(\overline{X}\mid W)P(\overline{Y}\mid \overline{X})P(\overline{Z}\mid \overline{X})\\
P(\overline{W},\overline{X},\overline{Y},\overline{Z}) &= P(\overline{W})P(\overline{X}\mid \overline{W})P(\overline{Y}\mid \overline{X})P(\overline{Z}\mid \overline{X})
\end{aligned}
\tag{5}
$$

which evaluates, by way of (2) and (4) to:

$$
\begin{aligned}
&P(W, X, Y, Z) = (0.01)(0.8)(0.999)(0.999) = 0.00798401 \\
&P(\overline{W}, X, Y, Z) = (0.99)(0.0)(0.999)(0.999) = 0.0 \\
&P(W, \overline{X}, Y, Z) = (0.01)(0.2)(0.0)(0.0) = 0.0 \\
&P(\overline{W}, \overline{X}, Y, Z) = (0.99)(1.0)(0.0)(0.0) = 0.0 \\
&P(W, X, \overline{Y}, Z) = (0.01)(0.8)(0.001)(0.999) = 7.992 \times 10^{-6} \\
&P(\overline{W}, X, \overline{Y}, Z) = (0.99)(0.0)(0.001)(0.999) = 0.0 \\
&P(W, \overline{X}, \overline{Y}, Z) = (0.01)(0.2)(1.0)(0.0) = 0.0 \\
&P(\overline{W}, \overline{X}, \overline{Y}, Z) = (0.99)(1.0)(1.0)(0.0) = 0.0 \\
&P(W, X, Y, \overline{Z}) = (0.01)(0.8)(0.999)(0.001) = 7.992 \times 10^{-6} \\
&P(\overline{W}, X, Y, \overline{Z}) = (0.99)(0.0)(0.999)(0.001) = 0.0 \\
&P(W, \overline{X}, Y, \overline{Z}) = (0.01)(0.2)(0.0)(1.0) = 0.0 \\
&P(\overline{W}, \overline{X}, Y, \overline{Z}) = (0.99)(1.0)(0.0)(1.0) = 0.0 \\
&P(W, X, \overline{Y}, \overline{Z}) = (0.01)(0.8)(0.001)(0.001) = 8.0 \times 10^{-9} \\
&P(\overline{W}, X, \overline{Y}, \overline{Z}) = (0.99)(0.0)(0.001)(0.001) = 0.0 \\
&P(W, \overline{X}, \overline{Y}, \overline{Z}) = (0.01)(0.2)(1.0)(1.0) = 0.002 \\
&P(\overline{W}, \overline{X}, \overline{Y}, \overline{Z}) = (0.99)(1.0)(1.0)(1.0) = 0.99
\end{aligned}
\tag{6}
$$

If we gather the non-zero joint probabilities, then we see that an infidelity has crept into the third and fourth right hand leaf of the quantified event tree in Figure C.1:

$$
\begin{aligned}
&P(W, X, Y, Z) = (0.01)(0.8)(0.999)(0.999) = 0.00798401 \\
&P(W, X, \overline{Y}, Z) = (0.01)(0.8)(0.001)(0.999) = 7.992 \times 10^{-6} \\
&P(W, X, Y, \overline{Z}) = (0.01)(0.8)(0.999)(0.001) = 7.992 \times 10^{-6} \\
&P(W, X, \overline{Y}, \overline{Z}) = (0.01)(0.8)(0.001)(0.001) = 8.0 \times 10^{-9} \\
&P(W, \overline{X}, \overline{Y}, \overline{Z}) = (0.01)(0.2)(1.0)(1.0) = 0.002 \\
&P(\overline{W}, \overline{X}, \overline{Y}, \overline{Z}) = (0.99)(1.0)(1.0)(1.0) = 0.99
\end{aligned}
\tag{7}
$$

It may be checked that the above probabilities sum to 1.0.

The decomposition of our joint probability distribution (7) in the conditional probability distributions (2) and (4), allows us to partition a very complex probability assignment problem [e.g., (7)] into much simpler probability assignment sub-problems [e.g., (2) and (4)].

Because of all the zero probabilities we can recover the above quantified event tree verbatim, using the law of total probability:

$$P(W, X, Y, Z) = 0.00798401$$

$$P(W, X, \overline{Y}, Z) = 7.992 \times 10^{-6}$$

$$P(W, X, Y, \overline{Z}) = 7.992 \times 10^{-6}$$

(8)

$$P(W, X, \overline{Y}, \overline{Z}) = 8.0 \times 10^{-9}$$

$$P(W, \overline{X}) = P(W, \overline{X}, Y, Z) + P(W, \overline{X}, \overline{Y}, Z) + P(W, \overline{X}, Y, \overline{Z}) + P(W, \overline{X}, \overline{Y}, \overline{Z}) = 0.002$$

$$P(\overline{W}) = P(\overline{W}, X, Y, Z) + P(\overline{W}, \overline{X}, Y, Z) + \cdots + P(\overline{W}, \overline{X}, \overline{Y}, \overline{Z}) = 0.99$$

12. Appendix D: A Consistency Derivation of Bernoulli's Utility Function

We will now derive the Bernoulli utility function, or, equivalently, the Weber-Fechner law, or, equivalently, in content, Steven's Power law, using the desiderata of invariance and consistency. In this we follow a venerable Bayesian tradition (Cox, 1946; Jaynes, 2003; Knuth and Skilling, 2010).

Say, we have the positive quantities x, y, and z, of some stimulus or commodity of interest. Then these quantities, being numbers on the positive real, admit an ordering. So, let quantities be ordered as $x \leq y \leq z$. We now want to find the function f that quantifies the perceived decrease associated with going from, say, the quantity z to the quantity x.

The first functional equation is based on the desideratum that the unknown function f should be invariant for a change of scale in our quantities:

$$f(x,z) = f(cx,cy) \tag{1}$$

where c is positive constant.

For example, if our quantities concern sums of money, then the perceived loss of going from ten dollars to one dollar should be the same perceived loss if we reformulate this scenario in dollar cents.

The second functional equation is based on the desideratum of consistency, in which we state that the perceived decrease in going directly from z to x, ought to be the same perceived decrease in going from z to x via y:

$$f(x,z) = g[f(x,y), f(y,z)] \tag{2}$$

For example, if our quantities concern sums of money, then the perceived loss of going from ten dollars to one dollar should be the same perceived loss if we first go from ten dollars to five dollars, and then from five dollars to one dollar; seeing that in both scenarios we start out with an initial wealth of ten dollars, only to end up with a current wealth of one dollar.

The general solution to (1) is (van Erp *et al.*, 2015):

$$f(x,y) = h\left(\frac{x}{y}\right) \tag{3}$$

were h is some arbitrary function. The general solution to (2) is (Knuth and Skilling, 2010):

$$\Theta[f(x,z)] = \Theta[f(x,y)] + \Theta[f(y,z)] \tag{4}$$

where Θ is some arbitrary monotonic function. Moreover, because of this arbitrariness, we may define Θ as (Knuth and Skilling, 2010):

$$\Theta(u) = \log \Psi(u), \tag{5}$$

where Ψ itself is also arbitrary and monotonic. Using (5), we may rewrite (4), without any loss of generality, as

$$\log \Psi[f(x,z)] = \log \Psi[f(x,y)] + \log \Psi[f(y,z)] \tag{6}$$

or, equivalently, by exponentiation of both sides of (6),

$$\Psi[f(x,z)] = \Psi[f(x,y)]\,\Psi[f(y,z)] \tag{7}$$

Substituting (3) into (4) through (7), and letting, respectively,

$$\theta\left(\frac{x}{y}\right) = \Theta\left[h\left(\frac{x}{y}\right)\right] \tag{8}$$

and

$$\psi\left(\frac{x}{y}\right) = \Psi\left[h\left(\frac{x}{y}\right)\right] \tag{9}$$

we obtain the equivalent functional equations:

$$\theta\left(\frac{x}{z}\right) = \theta\left(\frac{x}{y}\right) + \theta\left(\frac{y}{z}\right) \tag{10}$$

and

$$\psi\left(\frac{x}{z}\right) = \psi\left(\frac{x}{y}\right)\psi\left(\frac{y}{z}\right) \tag{11}$$

If we assume differentiability, then (10), together with the two boundary conditions:

$$f(x,x) = \theta\left(\frac{x}{x}\right) = 0 \tag{12}$$

and

$$f(x,y) = \theta\left(\frac{x}{y}\right) < 0, \qquad \text{for } x < y, \tag{13}$$

is sufficient to find the function f that quantifies the perceived decrease associated with going from the quantity y to the quantity x.

This function f turns out to be Bernoulli's utility function, or, equivalently, the Weber-Fechner law of sense perception:

$$f(x, y) = q \log \frac{x}{y}, \qquad \text{for } q > 0, \tag{14}$$

where y is our initial asset position and x is the final asset position, and q is some arbitrary constant which has to be obtained by way psychological experimentation.

So, Bernoulli's utility function (14) is the only function that adheres to the desiderata of unit invariance and consistency, respectively, (1) and (2), and the boundary conditions that a zero change should lead to a zero perceived loss and that a perceived loss should be assigned a negative value, respectively, (12) and (13). Any other utility function will be in violation with these fundamental desiderata and specific boundary conditions.

Note that Fechner re-derived (14) in 1860 as the law that guides our sensory perception. In the years that followed (14) proved to be so successful, as it, amongst other things, gave rise to our decibel scale, that it established psychology as a legitimate experimental science (Fancher, 1990).

But as Fechner was very careful, for reasons of aesthetics, or so we hazard to guess (van Erp *et al.*, 2015}, to apply his Weber law, which later became the Fechner-Weber law, only to non-monetary stimuli, the implied universality of (14) was not recognized for the longest time. However, because of the here given consistency derivation of (14), it is now shown that the Fechner-Weber, or, equivalently, Bernoulli's utility function, is one of the consistent functions that quantifies the distance between x and y; thus, explaining the universal applicability of Bernoulli's utility function.

The other consistent distance function is Steven's power law, which may be derived as follows. If we assume differentiability, then (11), together with the two boundary conditions:

$$f(x, x) = \psi\left(\frac{x}{x}\right) = 1 \tag{15}$$

and

$$0 < f(x, y) = \psi\left(\frac{x}{y}\right) < 1, \qquad \text{for } x < y, \tag{16}$$

is sufficient to find the function f that quantifies the perceived decrease associated with going from the quantity y to the quantity x.

This function f turns out to be Steven's power law:

$$f(x, y) = \left(\frac{x}{y}\right)^q, \qquad \text{for } q > 0, \tag{17}$$

Where y is our initial asset position and x is the final asset position, and q is some arbitrary constant which has to be obtained by way psychological experimentation.

So, Steven's power law (17) is the only function that adheres to the desiderata of unit invariance and consistency, respectively, (1) and (2), and the boundary conditions that a zero change should lead to a ratio of one between the initial and final asset position and that a perceived loss should be assigned a value smaller than 1, respectively, (15) and (16). Any other utility function will be in violation with these fundamental desiderata and specific boundary conditions.

We summarize, given the desiderata (1) and (2), the Fechner-Weber law (14) results from the boundary condition that negative increments result negative utilities and a zero increment results in an utility of zero, (12) and (13); whereas Steven's power law (17) results from the boundary condition that utilities must be greater than zero and that a zero increment results an utility of one, (15) and (16). Stated differently, the Fechner-Weber law and Steven's power law are both equivalent in content, differing only in the proposed utility scale. A subtlety that seems to have been overlooked by some, seeing that the Fechner-Weber law versus the Steven's power law has been a source of controversy in psycho-physical community (Stevens, 1961).

In closing, It may be read in (Jaynes, 2003), that to the best of Jaynes' knowledge, there are as of yet no formal principles at all for assigning numerical values to loss functions; not even when the criterion is purely economic, because the utility of money remains ill-defined. In the absence of these formal principles, Jaynes final verdict was that decision theory cannot be fundamental. The Bernoulli utility function, initially derived by Bernoulli, by way of common sense first principles (Bernoulli, 1738), has now been derived by way of a consistency argument.

This consistency argument explains why it is that Bernoulli's utility function, both in its original Fechner-Weber law and in its alternative Steven's power law form, has proven to be so ubiquitous and successful the field of sensory perception research; simply because human sense perception, like the laws of Nature (Knuth, 2014), adheres to the desideratum of consistency.

13. Appendix E: The Sum of the Lower and Upper Confidence Bound as a Measure of Position

Let A_1 and A_2 be two actions we have to choose from. Let C_i for $i = 1, \ldots, n$, and C_j for $j = 1, \ldots, m$, be, say, the monetary outcomes associated with, respectively, actions A_1 and A_2. In the Bayesian risk framework, we first construct the two outcome distributions that correspond with these actions:

$$p(C_i \mid A_1), \qquad \text{and} \qquad p(C_j \mid A_2). \tag{1}$$

We then proceed, by way of the Bernoulli utility function, or, equivalently, the Weber-Fechner law:

$$U_i = u(C_i \mid M) = q \log \frac{M + C_i}{M}, \tag{2}$$

to map utilities to the monetary outcomes C_i and C_j in (1). This leaves us with the utility probability distributions, (1) and (2):

$$p(U_i \mid A_1), \qquad \text{and} \qquad p(U_j \mid A_2). \tag{3}$$

Our most primitive intuition regarding the utility probability distributions (3) is that the decision which corresponds with the utility probability distribution which lies more to the right will also be the decision that promises to be the most advantageous. So, when making a decision we compare the positions of the utility probability distribution on the utility axis. This utility axis goes from minus infinity to plus infinity. Hence, the more-to-the-right criterion of choice. Now, the confidence bounds of (3), say:

$$\left[\, LB(U_i \mid A_1),\ UB(U_i \mid A_1) \,\right], \qquad \left[\, LB(U_j \mid A_2),\ UB(U_j \mid A_2) \,\right], \tag{4}$$

may provide us with a numerical handle on the concept of more-to-the-right.

For example, if we have that both

$$LB(U_i \mid A_1) > LB(U_j \mid A_2), \qquad \text{and} \qquad UB(U_i \mid A_1) > UB(U_j \mid A_2) \tag{5}$$

Then we will have an unambiguous preference for action A_1 over action A_2; seeing that under both the still probable worst and best case we will be better if we opt for A_1 (Treasury Board of Canada Secretariat, 1998). Likewise, if we have that either

$$LB(U_i \mid A_1) = LB(U_j \mid A_2), \qquad \text{and} \qquad UB(U_i \mid A_1) > UB(U_j \mid A_2) \tag{6}$$

or

$$LB(U_i \mid A_1) > LB(U_j \mid A_2), \qquad \text{and} \qquad UB(U_i \mid A_1) = UB(U_j \mid A_2) \tag{7}$$

Then, again, we will have an unambiguous preference for action A_1 over action A_2. In the constellation (6), we stand, all other things being equal, to be better off under the still probable best case scenario; while in the constellation (7), we stand, all other things being equal, to be less worse of under the still probable worst case scenario.

However, things become more ambiguous when, say, under action A_1, we have to make a trade-off between either a gain in the upper bound and a loss in the lower bound

$$LB(U_i \mid A_1) < LB(U_j \mid A_2), \quad \text{and} \quad UB(U_i \mid A_1) > UB(U_j \mid A_2) \tag{8}$$

or a gain in the lower bound and a loss in the upper bound

$$LB(U_i \mid A_1) > LB(U_j \mid A_2), \quad \text{and} \quad UB(U_i \mid A_1) < UB(U_j \mid A_2) \tag{9}$$

We *postulate* here that a rational criterion of choice in the respective trade-off situations (8) and (9), would be to pick that decision whose gain in either the lower or upper bound exceeds the loss in the corresponding upper or lower bound.

So, if, say, under action A_1 we stand to gain more in the still probable best case scenario than we stand to lose under the still probable worst case scenario, that is, (8):

$$LB(U_j \mid A_2) - LB(U_i \mid A_1) < UB(U_i \mid A_1) - UB(U_j \mid A_2), \tag{10}$$

then we will choose action A_1 over action A_2. Likewise, if under action A_1 we stand to gain more in the still probable worst case scenario than we stand to lose under the still probable best case scenario, that is, (9):

$$LB(U_i \mid A_1) - LB(U_j \mid A_2) > UB(U_j \mid A_2) - UB(U_i \mid A_1), \tag{11}$$

then again we will choose action A_1 over action A_2.

Note that the gains and losses in this discussion pertain to gains and losses on the utility dimension, not on the monetary outcome dimension. On the utility dimension the phenomenon of loss aversion, that is, the phenomenon that monetary losses may weigh heavier than equal monetary gains, has already been accounted for. Stated differently, the utility scale is a linear loss-aversion corrected scale for the subjective value of monies.

Now, if we look at the scenarios (10), and (11), then we see that we will choose action A_1 over action A_2 whenever we have that

$$LB(U_i \mid A_1) + UB(U_i \mid A_1) > LB(U_j \mid A_2) + UB(U_j \mid A_2). \tag{12}$$

Moreover, this single criterion of choice is also consistent with the choosing of action A_1 over action A_2 in the scenarios (5), (6), and (7). This, then, is the rationale behind the criterion of choice that we should maximize the sum of the lower and upper bounds of the utility probability distributions, in order to come to the optimal decision (van Erp *et al.*, 2015).

If the decision inequality (12) goes to an equality:

$$LB(U_i \mid A_1) + UB(U_i \mid A_1) = LB(U_j \mid A_2) + UB(U_j \mid A_2). \tag{13}$$

Then we have that we will be undecided when it comes to the deciding between the actions A_1 and A_2. Also note that for *k*-sigma bounds (4) translates to

$$E(U_i \mid A_1) \pm k \ \mathrm{std}(U_i \mid A_1), \qquad E(U_j \mid A_2) \pm k \ \mathrm{std}(U_j \mid A_2), \tag{14}$$

which, if substituted in (12), gives the inequality

$$2\ E(U_i \mid A_1) > 2\ E(U_j \mid A_2) \tag{15}$$

which brings us right back to Bernoulli's expected utility theory, as proposed in 1738, in which it is stated that the expectation value of the utility probability distribution should be maximized. As any scalar multiple of a criterion of choice is equivalent to that self-same criterion of choice; that is, (15) implies trivially

$$E(U_i \mid A_1) > E(U_j \mid A_2), \tag{16}$$

and vice versa.

Nonetheless, the criterion of choice, that a scalar multiple of the sum of the upper and lower bound should be maximized, as proposed here, will deviate from Bernoulli's initial 1738 proposal when the *k*-sigma intervals overshoots either its minimal or maximal value of the utility probability distribution. Let a and b, respectively, be the minimal and maximal values of a given utility probability distribution, and let scalar multiple be $c = 1/2$:

$$\frac{LB(U_i \mid A_k) + UB(U_i \mid A_k)}{2} = \begin{cases} E(U_j \mid A_k), & LB(U_j \mid A_2) \geq a,\ UB(U_j \mid A_2) \leq b, \\ \dfrac{a + E(U_j \mid A_k) + \mathrm{std}(U_j \mid A_k)}{2}, & LB(U_j \mid A_2) < a,\ UB(U_j \mid A_2) \leq b, \\ \dfrac{E(U_j \mid A_k) - \mathrm{std}(U_j \mid A_k) + b}{2}, & LB(U_j \mid A_2) \geq a,\ UB(U_j \mid A_2) > b, \\ \dfrac{a+b}{2}, & LB(U_j \mid A_2) < a,\ UB(U_j \mid A_2) > b \end{cases}$$

(16)

Then we may identify two additional symmetry breaking cases (van Erp *et al.*, 2015), relative to the maximization of expected utility theory (15), as well as a Hurwitz criterion of choice with a pessimism factor of $\alpha = 1/2$.

Another instance where we will deviate from Bernoulli's expected utility proposal is when we put an explicit premium on either caution or opportunity. If we take as the lower and upper bounds, whose sum is to be maximized:

$$LB(U_i \mid A_k) = E(U_i \mid A_k) - k_1 \ \mathrm{std}(U_i \mid A_k) \tag{17}$$

and

$$UB(U_i \mid A_k) = E(U_i \mid A_k) + k_2 \ \mathrm{std}(U_i \mid A_k) \tag{18}$$

Then (35) and (36) combine to:

$$\frac{LB(U_i \mid A_k) + UB(U_i \mid A_k)}{2} = E(U_i \mid A_k) + \frac{k_2 - k_1}{2} \mathrm{std}(U_i \mid A_k) \tag{19}$$

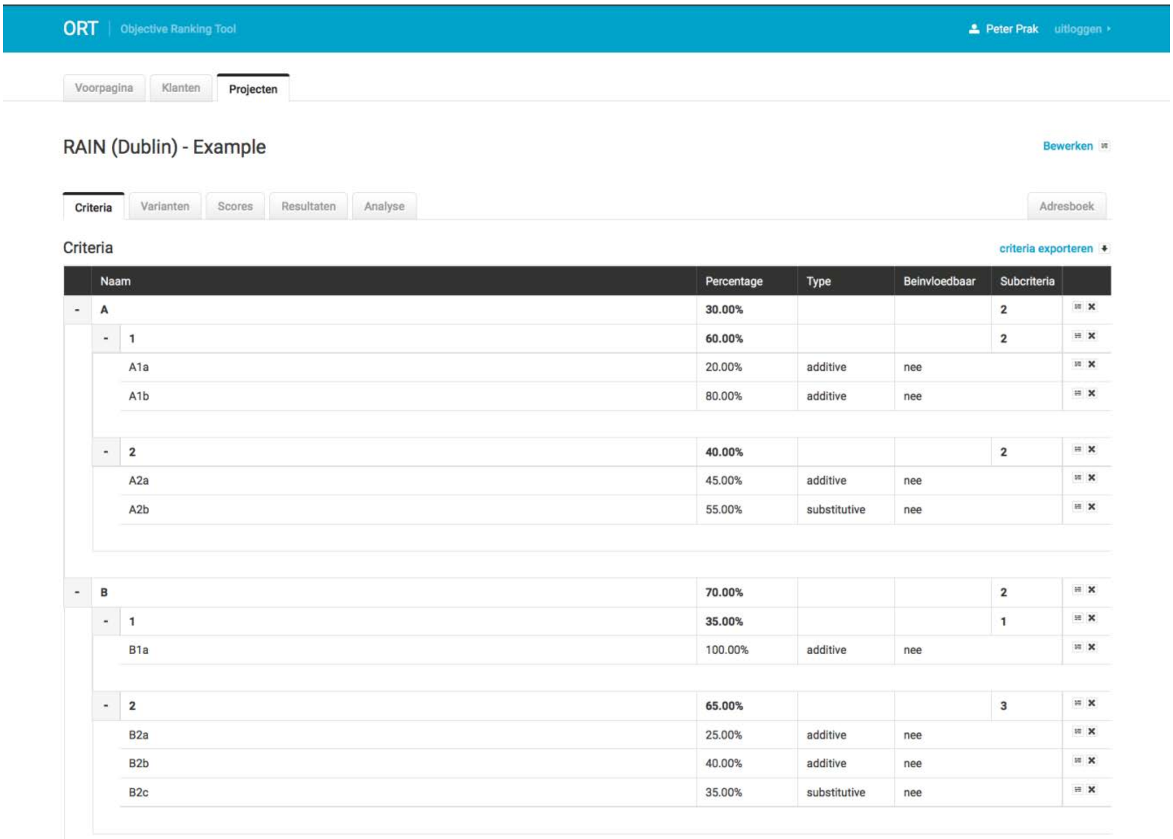
If in (37) we let $k_1 > k_2$, then we put a premium on caution; if we set $k_2 > k_1$, then we put a premium on opportunity; and if we let $k_1 = k_2$, then we have an equal trade-off between caution and opportunity taking (van Erp *et al.*, 2015).

In closing, note that the minimum and maximum utility values, respectively, a and b, just like the first two cumulants, $E(X)$ and $\mathrm{std}(X)$, are a linear function of the scaling constant q in (2), and as a consequence cancel out once we solve for the decision theoretical inequality (12) or the equality (13).

14. Appendix F: Screenshots of the web-based application ORT

The Objective Ranking Tool is developed in the Dutch language. A translation is given for each tab.

- Criteria criteria
- Naam name
- Beïnvloedbaar to be influenced

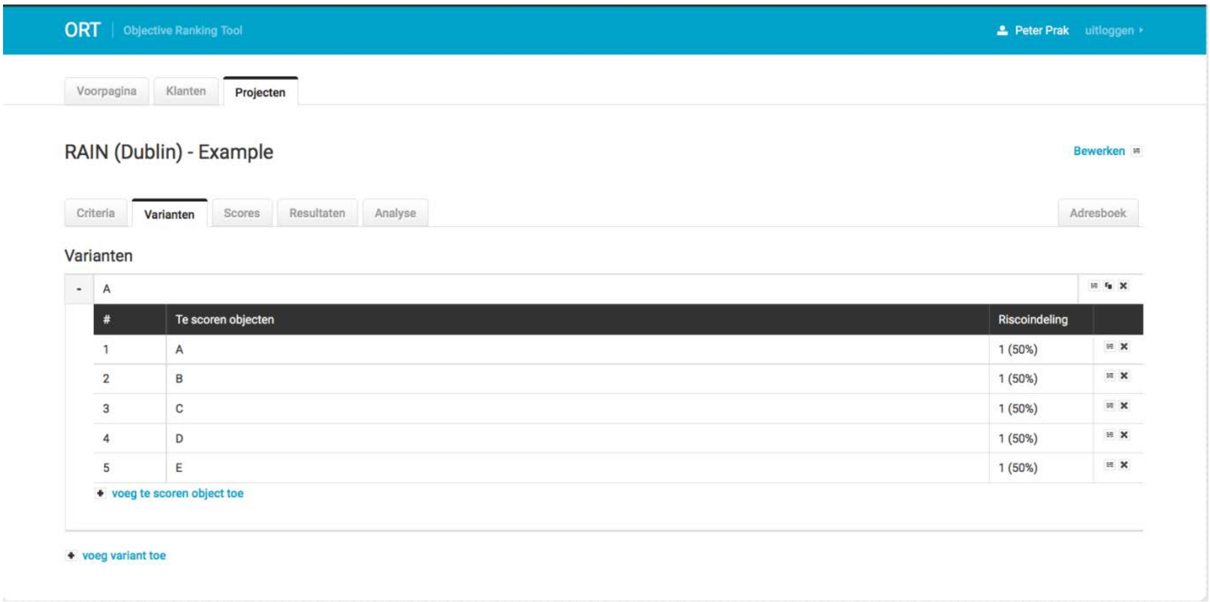


	Naam	Percentage	Type	Beinvloedbaar	Subcriteria
-	A	30.00%			2
-	1	60.00%			2
	A1a	20.00%	additive	nee	
	A1b	80.00%	additive	nee	
-	2	40.00%			2
	A2a	45.00%	additive	nee	
	A2b	55.00%	substitutive	nee	
-	B	70.00%			2
-	1	35.00%			1
	B1a	100.00%	additive	nee	
-	2	65.00%			3
	B2a	25.00%	additive	nee	
	B2b	40.00%	additive	nee	
	B2c	35.00%	substitutive	nee	

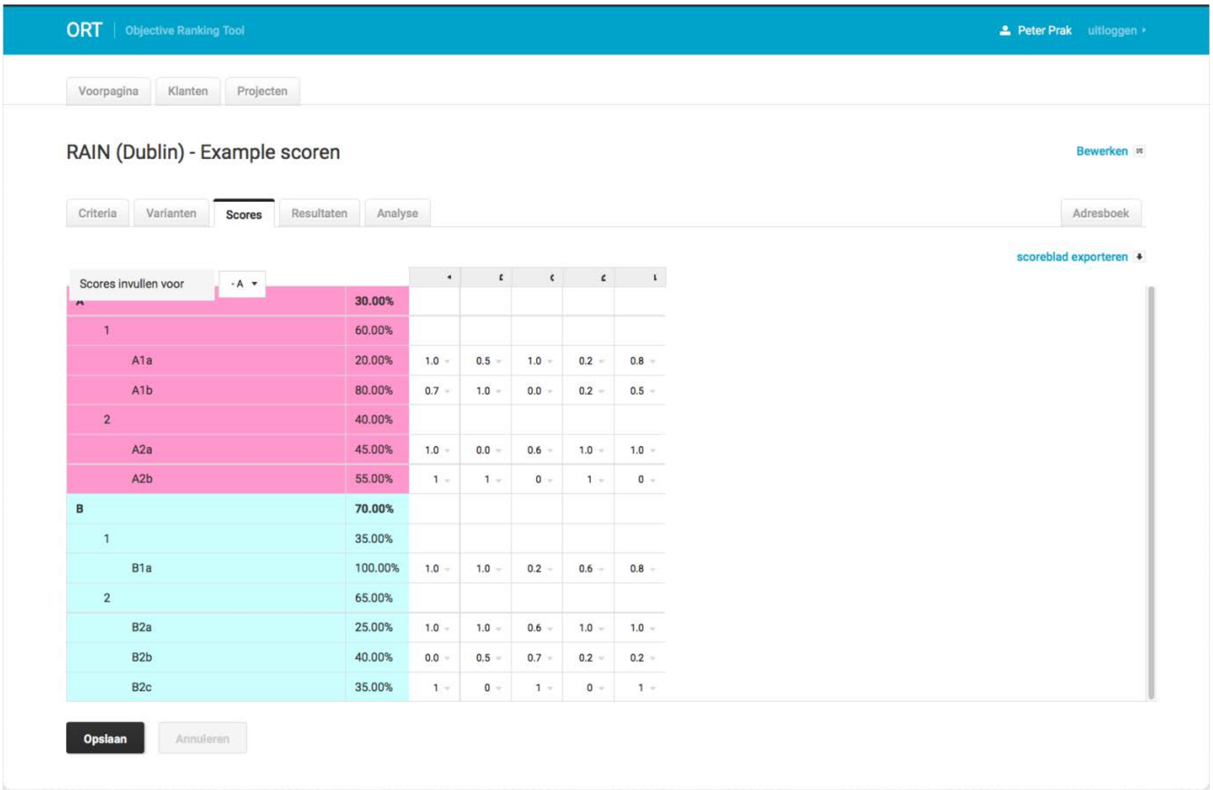
Objects to score (varianten)

- Risicoindeling risk classification

The risk classification with the presented outcome 1 (50%) is a feature within the application to set the level of 'a' and 'b' of the formula by Tversky ($S_{ij} = f_{ij} / [f_{i,j} + a(f_{i,\ not\ j}) + b(f_{not\ i,\ j})]$). 'A' and 'b' should count to '1'.



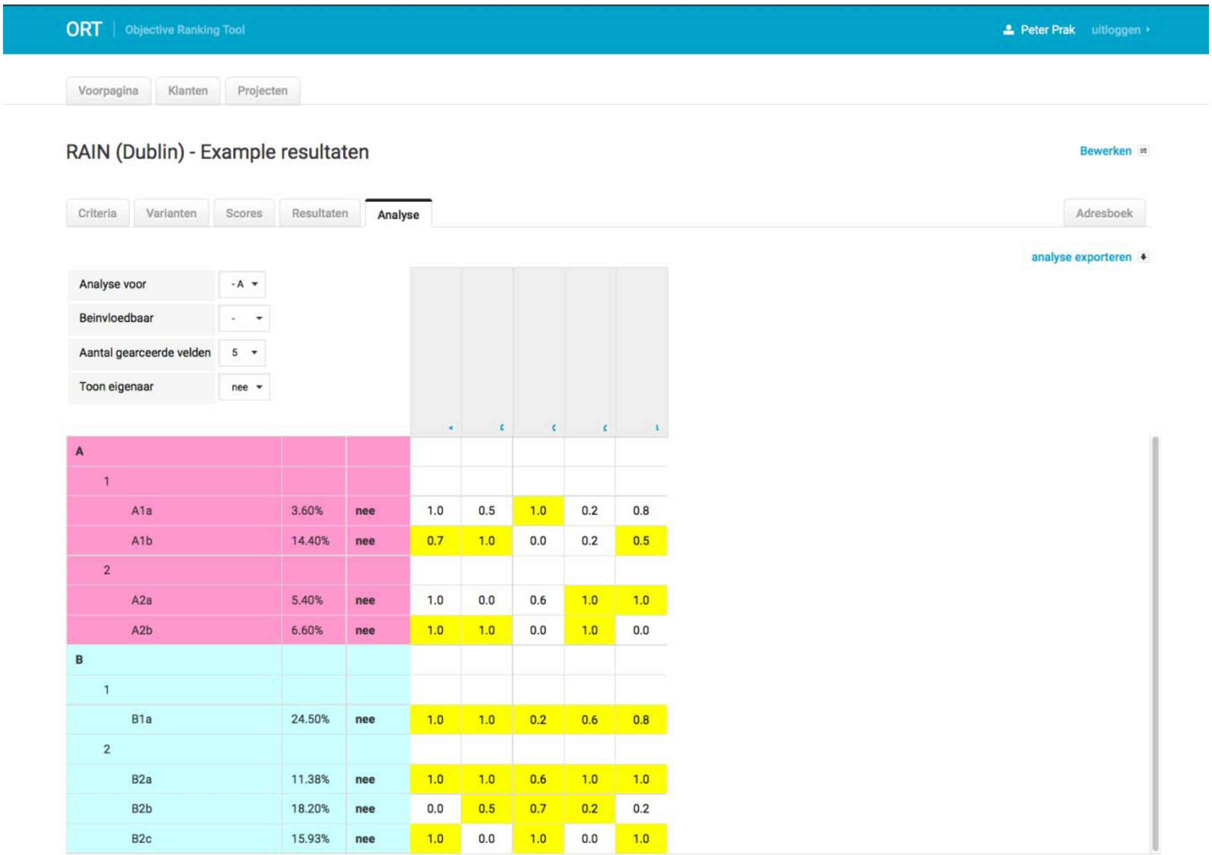
Scoring (scores)



Results (resultaten)

Analyses (analyse)

- Analyse voor — analysis for
- Beïnvloedbaar — to be influenced
- Aantal gearceerde cellen — number of hatched cells (between 1 and 10)
- Toon eigenaar — show owner



The weighting percentages of all criteria are counting to 100.01%. This is due to rounding difference. The calculation model will count with 100%.