Statistical Modeling of Expert Opinions Using Imprecise Probabilities

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de Rector Magnificus, prof.dr. J.H. van Lint, voor een commissie aangewezen door het College van Dekanen in het openbaar te verdedigen op dinsdag 1 februari 1994 om 16.00 uur

door

Franciscus Petrus Antonius Coolen

geboren te Tilburg

Dit proefschrift is goedgekeurd door

de promotoren: prof.dr. P. van der Laan
                prof.dr. P.C. Sander
co-promotor: dr. M.J. Newby

VOORWOORD

Het onderzoek beschreven in dit proefschrift is uitgevoerd gedurende een vier-jarige aanstelling als Assistent In Opleiding bij de vakgroep Besliskunde & Stochastiek van de faculteit Wiskunde & Informatica aan de Technische Univer-siteit Eindhoven.

Mijn dank gaat uit naar mijn promotoren, Paul van der Laan en Peter Sander, voor het bieden van deze mogelijkheid en hun stimulerende interesse in het onderzoek. De tijd en energie die Martin Newby in het begeleiden van mijn onderzoek heeft gestoken waardeer ik ten zeerste, en is van groot belang geweest voor het tot stand komen van dit proefschrift. Voorts dank ik Emiel van Berkum, Rommert Dekker en Adriaan Smit voor hun steun bij mijn onderzoek. De heer C. van Rijn van het Koninklijke/Shell Laboratorium te Amsterdam ben ik dank verschuldigd voor de geboden mogelijkheid mijn inzicht in de problematiek van onderhoudsplanning te vergroten tijdens een stage gedurende deze periode. I am grateful to Peter Walley for his stimulating comments during this re-search, and to Tom Mazzuchi for our discussions about Bayesian statistics and the use of expert opinions. Mijn collegae te Eindhoven dank ik eveneens, Roel Wilms in het bijzonder vanwege de prettige, vriendschappelijke sfeer waarin we bijna 4 jaar lang de werkkamer hebben gedeeld.

De steun die ik heb gekregen van familie, collegae en vrienden, in het bij-zonder Pauline, is voor mij van groot belang geweest, en heeft mij doen besef-fen dat dit proefschrift tevens het begin van een nieuwe periode in mijn leven is.

Tenslotte dank ik mijn ouders voor meer dan ik kan noemen. Ondanks de dood van mijn moeder is haar steun een grote stimulans voor mij.

CONTENTS

Chapter 1

INTRODUCTION

1.1 Motivation

In every-day life one is often confronted with decision problems related to uncertain events, and almost all decisions are based on subjective informa- tion. For most decisions some rule of thumb is used, but sometimes it is ob- vious that a structural analysis of the information at hand is needed to make a good decision. For example, if one carries responsibility for maintenance planning for critical equipment in a chemical process plant, a decision sup- port system can be valuable, as well for determination of a good decision as for providing a basis for such a decision by reporting information about all aspects of interest to the problem. The information can consist of historical or experimental data or expert opinions. Information from experts plays an important role in many decision problems and the need to quantify expert opi- nions for use in decision support systems is growing.

If the decision problem relates to an uncertain event, it is generally accep- ted that quantification of expert opinions about the occurrence of this event should be in terms of probabilities or a related concept. At first sight, the theory of subjective probabilities combined with the Bayesian concept of sta- tistics and decision making (De Finetti, 1974) seems appropriate, but on ap- plying these serious problems arise as discussed by Walley (1991; chapter 5). Most of these problems relate to the interpretation of subjective probabili- ties, which is obviously very important for practical applications but unfor- tunately highly neglected by mathematicians. Definition of probabilities by using a betting behavior interpretation (De Finetti, 1974) seems appropriate. Here, a bet on an event A is such that the buyer of the bet receives 1 if A occurs and 0 otherwise, where 1 should be a linear unit of utility (Walley, 1991; section 2.2). To measure your knowledge about A, where we use 'you' to emphasize the subjective interpretation, you are supposed to be able to assess a unique value $P(A)$ such that you would buy the bet for all prices less than $P(A)$ if offered to you, and sell it for all prices higher than $P(A)$. This interpretation leads to one number $P(A)$ that is called your probability for A, and on logical grounds (De Finetti, 1974) this leads to a theory identical to the mathematical theory of probabilities following the Kolmogorov-axioms (with

perhaps some discussion left about finite additivity). Of course, once a mathematician uses this theory he can leave out the interpretation part and simply work within this well-known framework.

In practice however, problems arise from the fact that people do not naturally have a unique P(A) with the above interpretation. Of course, if one is forced to choose between buying or selling a bet against each price, this will lead to such a P(A), but a simple example illustrates that important information is deleted here. For a football match between team A and team B in which one of both will win (think about a cup-final), let A denote the event that team A wins. If you assign P(A)=1/2, this number indicates that you think it equally likely that team A or B wins, but this might as well be based on the actual knowledge that both teams are equally strong as on lack of any information about the teams and the match. Because it is important, in practical decision problems, that such a difference in knowledge is reported, some concepts have recently been proposed to overcome this problem.

The use of second order probabilities is attractive within the standard Bayesian framework, but suggests that you are able to quantify your opinions about P(A) exactly by means of a probability distribution, for example by assessing probabilities for events $P(A) \leq p$ for some $p \in [0,1]$. Walley (1991; section 5.10) discusses this idea, and there are several problems, especially for application, related to interpretations. For example, what is the meaning of your P(A) if this is uncertain, and how can the second order probability distribution be interpreted well? Analogous problems with interpretation occur in theories that use membership functions, such as fuzzy set theory (Zadeh, 1983; discussed by Walley, 1991; section 5.11) or possibility measures. There seems to be a problem with assessment of a unique P(A), and therefore one proposes solutions that are based on quantification of some form of vagueness, mostly without interpretable definitions of the concepts. It is hardly ever discussed what the difference is between membership functions and probabilities, or do these have the same interpretations? In applications, membership functions are simply assumed to have some form without a sensible justification (e.g. Chun and Ahn, 1992), which cannot be provided by consulted experts simply because nobody knows what they are talking about. Therefore, it seems to be logical to look for a solution from the opposite direction, so not starting with subjective probabilities following the Kolmogorov axioms and looking for solutions to the problems that arise, but trying to find out where the stated problem of quantification of opinions comes from, and propose as solution a theory based on quantities with a sensible interpretation together with some rational basic assumptions, without restriction to the Kolmogorov

axioms beforehand. The theory of imprecise probabilities, or more generally imprecise previsions, is such a methodology, and the mathematical foundation is provided by Walley (1991).

This theory also starts from the idea that expert knowledge is measured through betting behavior, but instead of a $P(A)$ with the above interpretation you are asked to assess two values, called lower and upper probabilities and denoted by $\underline{P}(A)$ and $\overline{P}(A)$, respectively, with the interpretation that for all prices $p<\underline{P}(A)$ you expect a profit when buying the bet, and for all prices $p>\overline{P}(A)$ you expect a profit when selling the bet. So instead of being forced to choose between buying or selling a bet for all prices, you are asked if you expect profit in case of buying or selling a bet against a price, which may leave an interval $[\underline{P}(A),\overline{P}(A)]$ of positive length containing prices against which you are indeterminate of the result that you expect on betting, in the sense that based on your current state of knowledge of A you do not expect profit either in case of buying or selling the bet for prices within this interval. Fundamental to the theory presented by Walley (1991) is coherence of betting behavior, which can be interpreted as the assumption that you, when betting on several events, do not accept combinations of bets such that you will surely loose (in units of utility). For example, it is easy to verify that a consequence of this assumption is $0\leq\underline{P}(A)\leq\overline{P}(A)\leq1$.

The length of the interval $[\underline{P}(A),\overline{P}(A)]$ is called the imprecision about A, and is denoted by $\Delta(A)=\overline{P}(A)-\underline{P}(A)$. It seems reasonable that imprecision is related to how sure you feel about your knowledge of A. To illustrate the two extreme situations possible in the brief example of the football match, if you are entirely convinced that both teams are equally strong you may assess $\underline{P}(A)=\overline{P}(A)=1/2$, but if you know nothing at all about the match and the teams, or not even about football, you may assess $\underline{P}(A)=0$ and $\overline{P}(A)=1$. The imprecision seems to be a measure that is related to the value that you add to the amount of information available at the moment you are asked to represent your knowledge through betting behavior. Only if $\underline{P}(A)=\overline{P}(A)$ for all events A the method reduces to the standard theory of probability. Indeed, this concept differs substantially from methods as fuzzy sets or higher order probabilities in that it adds no new things with vague interpretations to probabilities, but it generalizes the concept by deleting the assumption that $\underline{P}(A)=\overline{P}(A)$ for all A, that is the cause of the loss of an important part of subjective information.

Although the concept of imprecise probabilities is based on a purely subjective interpretation, the need exists for a theory of updating imprecise probabilities in the light of new information. Strictly according to the interpre-

tation, after new information your imprecise probabilities need to be measured again to present your betting behavior at the new moment in time. Mainly for practical reasons one would like to develop a theory that takes care of updating in a reasonable way, where it is not meant that you should indeed be willing to bet according to the theoretically derived imprecise probabilities, since human mind will never exactly follow updating rules.

Walley (1991; section 6.4) presents a theory of updating through a so-called Generalized Bayes Rule (GBR), that is based on the idea that updated imprecise probabilities are equal to the conditional imprecise probabilities as defined by the priors, and betting behavior should be coherent through time. Generally, coherent betting behavior means that you do not accept combinations of bets that lead to sure loss, which we accept as logical requirement for betting behavior at one moment. By coherence through time we mean the assumption that your betting behavior before and after new information has become available should be coherent together, so no combinations of bets that you are willing to accept either before or after the new information can be found that lead to sure loss. However, on discussing the GBR Walley repeatedly remarks that the conditional prior imprecise probabilities, with conditioning to the new information, may only provide bounds for the updated betting behavior, in the sense that your real betting behavior after learning the new information may be more precise than according to the GBR (Walley, 1991; sections 6.1.2, 6.11.2, 6.11.9). This leads Walley to remark that there is scope for other updating strategies than the GBR.

The main motivation for the research presented in this thesis is the question if it is possible to propose an updating theory that deletes the assumption of coherence through time as Walley's GBR and indeed may lead to less imprecise betting behavior after updating than the GBR suggests. By deleting the assumption of coherence through time we aim at explicit modeling of betting behavior based on the old and new information after updating, where the intuitive connection between imprecision and the amount of information available when betting behavior is represented is assumed to play an important role. The methodology that we propose is not meant to be as general as Walley's GBR, but restricted to the Bayesian framework of statistics, so dealing with parametric models and new information only in the form of observations of random variables of interest. This brings our concept close to the robust Bayesian methods (Berger, 1990), in which no role is played by imprecision as a measure that is related to the amount of information. The robust Bayesians assume

-4-

that probability is precise by nature, but use sets of prior distributions as a means of sensitivity analysis only.

The concept developed and discussed in this thesis shows that an updating theory exists in which the relation between imprecision and information plays a crucial role, and this updating theory could be the starting-point for new research in this area.

1.2 Overview

Before a new updating strategy is proposed in which imprecision is used as a measure directly related to the amount of information available, or better to say to the value one adds to this information, a mathematical expression for this relation is needed. In literature this aspect has not been explicitly studied before, but Walley (1991; section 5.3) suggests, on quite arbitrary grounds, an information measure related to imprecision. In section 2.2 a new argument in favor of Walley's measure of information is presented, after which this measure is used throughout the thesis. Section 2.3 presents the new updating theory, with emphasis on the relation between imprecision and information, and this theory is more deeply discussed for the case of Bernoulli experiments in section 2.4.

Next to a statistical theory for updating probabilities in the light of new information, the standard Bayesian framework provides a clear concept for decision making (Lindley, 1990). In section 3.2 a generalization of this concept is presented, related to the theory proposed in chapter 2. In section 3.3 an important problem of calculation is analyzed, leading to the conclusion that adoption of our concept of imprecise prior probabilities does not lead to an explosion of the necessary amount of numerical calculations necessary for inferences. In section 3.4 an example of application of the concept is presented, analyzing a simple replacement strategy for technical equipment.

Of course, a new theory for using expert opinions in decision problems can only be evaluated by an actual implementation, which has not been performed. Nevertheless, in chapter 4 we discuss some aspects that are important for implementation, and we illuminate possible advantages of our concept when applied to practical problems in comparison with existing theories that do not use imprecise probabilities. In section 4.2 the measurement of expert opinions, known as elicitation, is discussed. In section 4.3 the combination of imprecise probabilities provided by several experts is considered, and in section 4.4 a method is suggested to fit assumed models to information provided by experts in forms as proposed in section 4.2, where model fitting means the choice of suitable values for hyperparameters for a set of prior distributions.

Recently, reliability engineers and analysts have shown interest in new techniques to quantify knowledge about uncertain events, which has resulted in a number of papers on applications of, for example, fuzzy sets (e.g. Chun and

Ahn (1992), Dubois and Prade (1992), Rao and Dhingra (1992)). In section 4.5 implementation of our method in reliability analysis is briefly discussed, providing generalized definitions of some well-known concepts in this area.

Finally, in the epilogue some topics for future research related to this thesis are mentioned.

Notes

Some of the results presented in this thesis have already been published elsewhere, or have been accepted for publication. Parts of section 2.3 are presented in the paper 'Imprecise conjugate prior densities for the one-parameter exponential family of distributions', *Statistics & Probability Letters 16* (1993), 337-342.

Section 2.4 is the rewritten body of the paper 'On Bernoulli experiments with imprecise prior probabilities', to appear in a special issue of *The Statistician* on the Third International Conference on practical Bayesian statistics of The Institute of Statisticians (Nottingham, 1992), where this paper has been presented.

Section 3.2 is the main part of a short presentation at the second joined International Conference of the German and the Dutch Societies for Operations Research (Amsterdam, 1993), and a short paper, entitled 'Decision making with imprecise probabilities' is published in the conference proceedings.

Section 3.3 is almost similar to the paper 'Bounds for expected loss in Bayesian decision theory with imprecise prior probabilities', to appear in *The Statistician*.

The methodology proposed in this thesis has also been presented at the Symposium 'Reliability: A competitive edge', of the Society of Reliability Engineers (Arnhem, 1993), with emphasis on the example presented in section 3.4. A paper has been published in the proceedings (pp. 172-183), entitled 'Bayesian decision theory with imprecise prior probabilities applied to replacement problems'.

Section 4.5 can be regarded as a short version of the paper 'Bayesian reliability analysis with imprecise prior probabilities', with co-author M.J. Newby, that will appear in *Reliability Engineering and System Safety*.

Chapter 2

BAYESIAN STATISTICS USING IMPRECISE PROBABILITIES

2.1 Introduction

The theory of imprecise probabilities (Walley, 1991) is a generalization of the classical theory of subjective probability (De Finetti, 1974), with lower and upper probabilities related to personal betting behavior. A *bet* on event A is such that the owner receives 1 unit of utility if A occurs and 0 if not. Here unit of utility is not further specified, but Walley (1991; section 2.2) presents a method to overcome the well-known problems if utility is expressed in terms of money, which is based on the use of lottery tickets. Your *lower probability* for event A, $\underline{P}(A)$, is the supremum of all prices for which you want to buy the bet, your *upper probability* $\overline{P}(A)$ the infimum of all prices for which you want to sell the bet, assuming that you only want to buy or sell a bet if you expect profit.

In the classical theory of subjective probability you are forced to $\underline{P}(A)=\overline{P}(A)$. This creates the obvious problem that, for example, a single probability $P(head)=1/2$ when tossing a coin is used if you have no information about the coin at all as well as if you know that the coin is perfectly symmetrical. So the forced use of a single probability destroys important information, whereas the use of higher-order probabilities only pushes the problem ahead (Walley, 1991; section 5.10).

Central to the theory of imprecise probabilities is the *coherence* of betting behavior, which is the assumption that you always avoid (combinations of) bets that lead to sure loss. For example, $\underline{P}(A) \leq \overline{P}(A)$ follows from the coherence assumption, since $\underline{P}(A) > \overline{P}(A)$ would imply that prices p and q exist with $\overline{P}(A) < p < q < \underline{P}(A)$, such that you want to buy the bet on A for price q and sell it for price p, together leading to the sure loss of $q-p > 0$.

Most of Walley's (1991) monograph is about coherence, and a characterization of coherence of imprecise probabilities is presented in a list of properties (Walley, 1991; section 2.7.4). As Walley's theory is more general than the one presented in this thesis, we restrict ourselves to the presentation of the most important properties as axioms. It must be observed that coherence of the imprecise probabilities used in this thesis is ensured by the assumed models (Walley, 1991; section 4.6).

Before presenting the axioms it is useful to note that there is a second in-

terpretation for imprecise probabilities. One could assume the actual existence of a single probability for event A, say $P(A)$, following the classical theory of subjective probability (De Finetti, 1974), and regard the imprecise probabilities as bounds, such that $\underline{P}(A) \leq P(A) \leq \overline{P}(A)$. It is the assumption of existence of a single $P(A)$ that is deleted in the first interpretation which we adopt. The first interpretation also gives a sensible meaning to these bounds. If the second interpretation is preferred, one has to add assumption 2.1.1 that implies that $\underline{P}(A)$ and $\overline{P}(A)$ are chosen as sharp as possible.

Assumption 2.1.1
If you know that $p \leq P(A) \leq q$ for some $p, q \in [0,1]$, then you assess lower and upper bounds $\underline{P}(A)$ and $\overline{P}(A)$ with $\underline{P}(A) \geq p$ and $\overline{P}(A) \leq q$.

Although the first interpretation is adopted throughout this thesis, it is interesting to show how the axioms for imprecise probabilities relate to the well-known axioms of precise probabilities in this second interpretation. Discussion of the axioms in the light of the first interpretation is extensively presented by Walley (1991).

Let Ω be the set of all possible events of interest, then $\underline{P}$ and $\overline{P}$ are assumed to satisfy the following basic axioms (Wolfenson and Fine, 1982)

(IP1) For all $A \subseteq \Omega$: $\underline{P}(A) \geq 0$
(IP2) $\underline{P}(\Omega) = 1$
(IP3) For all $A, B \subseteq \Omega$, with $A \cap B = \varnothing$: $\underline{P}(A) + \underline{P}(B) \leq \underline{P}(A \cup B)$, $\overline{P}(A) + \overline{P}(B) \geq \overline{P}(A \cup B)$
(IP4) For all $A \subseteq \Omega$: $\overline{P}(A) + \underline{P}(A^c) = 1$, where $A^c = \Omega \backslash A$, the complement of A.

Basic axioms for precise probability $P(A)$ are

(PP1) For all $A \subseteq \Omega$: $P(A) \geq 0$
(PP2) $P(\Omega) = 1$
(PP3) For all $A, B \subseteq \Omega$, with $A \cap B = \varnothing$: $P(A) + P(B) = P(A \cup B)$.

We shall not discuss the problem whether or not the additivity axiom PP3 should hold for only finite numbers of events, or also for countably infinite numbers. For such discussion see De Finetti (1974), who opposes countable additivity, and Walley (1991; section 6.9), who accepts it. We follow Walley here.

Next we show how axioms IP1-IP4 follow from axioms PP1-PP3 and assumption 2.1.1, using the second interpretation with the imprecise probabilities regarded as bounds for a precise probability, $\underline{P}(A) \leq P(A) \leq \overline{P}(A)$.

Axioms IP1 and IP2 follow straightforwardly from PP1 and PP2 together with assumption 2.1.1. IP2 implies that one is absolutely sure about Ω, but this is not a serious restriction if interest is only in event A, as further specification of A^c is not needed.

Axiom IP3 follows from PP3 and assumption 2.1.1 by $P(A \cup B) = P(A) + P(B) \geq \underline{P}(A) + \underline{P}(B)$ and $P(A \cup B) = P(A) + P(B) \leq \overline{P}(A) + \overline{P}(B)$. As an example of a situation in which $\underline{P}(A \cup B) > \underline{P}(A) + \underline{P}(B)$, regard tossing a coin which is not known to be fair, and let H denote the event that the coin lands head up. Your betting behavior might, for example, be represented by $\underline{P}(H) = 0.3$ and $\underline{P}(H^c) = 0.2$. However, since $H \cup H^c = \Omega$ it follows that $\underline{P}(H \cup H^c) = 1$ while $\underline{P}(H) + \underline{P}(H^c) = 0.5$, so indeed $\underline{P}(H \cup H^c) > \underline{P}(H) + \underline{P}(H^c)$.

Axiom IP4 follows from PP2 and PP3, that lead to $P(A) + P(A^c) = 1$. We know that $P(A) \geq \underline{P}(A)$, so $1 - P(A^c) \geq \underline{P}(A)$ and $P(A^c) \leq 1 - \underline{P}(A)$, and by assumption 2.1.1 $\overline{P}(A^c) \leq 1 - \underline{P}(A)$ holds. Analogously we know that $P(A^c) \leq \overline{P}(A^c)$, so $1 - P(A) \leq \overline{P}(A^c)$ and $P(A) \geq 1 - \overline{P}(A^c)$, and assumption 2.1.1 leads to $\underline{P}(A) \geq 1 - \overline{P}(A^c)$, so $\overline{P}(A^c) \geq 1 - \underline{P}(A)$. These two results, $\overline{P}(A^c) \leq 1 - \underline{P}(A)$ and $\overline{P}(A^c) \geq 1 - \underline{P}(A)$, imply IP4.

Consequences of axioms IP1-IP4 are

(C1) $\underline{P}(\emptyset) = \overline{P}(\emptyset) = 0$
(C2) $\overline{P}(\Omega) = 1$
(C3) $\overline{P}(A) \geq \underline{P}(A)$
(C4) $A \subseteq B$ implies $\underline{P}(B) \geq \underline{P}(A)$ and $\overline{P}(B) \geq \overline{P}(A)$
(C5) $A \cap B = \emptyset$ implies $\underline{P}(A \cup B) \leq \underline{P}(A) + \overline{P}(B) \leq \overline{P}(A \cup B)$.

Consequence C1 follows from IP2, IP3 and IP4, C2 follows from IP4 and C1, and C3 follows from IP3 and IP4.
To proof C4 note that
  $\underline{P}(B) = \underline{P}(A \cup (B \backslash A)) \geq \underline{P}(A) + \underline{P}(B \backslash A) \geq \underline{P}(A)$,
since $A \cap (B \backslash A) = \emptyset$, where IP3 and IP1 have been used. Further, $A \subseteq B$ implies $B^c \subseteq A^c$, so $\underline{P}(A^c) \geq \underline{P}(B^c)$ and, using IP4, $\overline{P}(B) \geq \overline{P}(A)$.

To proof C5 use that $A\cap B=\varnothing$ implies $A\subseteq B^c$, so

$$\underline{P}(A)=\underline{P}((A\cup B)\cap B^c)=1-\overline{P}(((A\cup B)\cap B^c)^c)=1-\overline{P}((A\cup B)^c\cup B)$$
$$\geq 1-\overline{P}((A\cup B)^c)-\overline{P}(B)=\underline{P}(A\cup B)-\overline{P}(B),$$

which proves the first inequality, using some logical results of the theory of sets together with IP3 and IP4. The second inequality follows from the fact that $A\cap B=\varnothing$ implies $B\subseteq A^c$, so

$$\overline{P}(B)=\overline{P}((A\cup B)\cap A^c)=1-\underline{P}(((A\cup B)\cap A^c)^c)=1-\underline{P}((A\cup B)^c\cup A)$$
$$\leq 1-\underline{P}((A\cup B)^c)-\underline{P}(A)=\underline{P}(A\cup B)-\underline{P}(A).$$

Following Walley (1991; section 5.1), as a possible *degree of imprecision* about A we define $\Delta(A)=\overline{P}(A)-\underline{P}(A)$. According to Walley $\Delta(A)$, also called *imprecision*, is assumed to depend on the amount of information about A available to you. If new information becomes available, this may change your betting behavior and a statistical model using imprecise probabilities should explicitly use a relation between imprecision and information. The first step into such a new theory of updating imprecise probabilities is the definition of a measure of information related to imprecision. On the basis of some simple requirements, Walley (1991; section 5.3) proposes $I(A)=\Delta(A)^{-1}-1$ as a *measure of the amount of information* about A related to your imprecise probabilities. In section 2.2 the relation between imprecision and information is analyzed and a new argument in favor of $I(A)$ is presented. It would be better to call $I(A)$ the value of available information to you, as imprecision does not necessarily decrease as function of the amount of information when expressed in numbers of data. Situations are possible in which new information contradicts old information, and may lead to an increase of imprecision. A statistical theory of imprecise probabilities must be able to deal with such situations correctly, that means according to intuition since betting behavior is subjective by nature.

The reader must be made aware of a possible problem with terminology. In our concept information is related to imprecision, and therefore is different to the term as used in information theory. Walley (1991; section 5.1) discusses this, using the terminology *determinacy* and *indeterminacy* (or determinate and indeterminate uncertainty), of which one could also think in terms of *stochastic uncertainty* and *uncertainty through lack of information*. The difference is clear in the coin tossing example where a precise $P(head)=1/2$ quantifies stochastic uncertainty if the coin is perfectly fair, but fails to indicate uncertainty through lack of information. It would seem to be correct to re-

place the word information in our theory by some other term, since statisticians are used to the information theory terminology. However, we strongly believe that *information* in our concept is closer to its linguistic meaning than it is as used in information theory, and therefore we would like to pass the problem of finding a new terminology to the other side.

An extensive overview of the literature on imprecise probabilities can be found in Walley (1991; section 1.8 and 2.7). We restrict ourselves to a short survey. The idea of imprecise probabilities goes back to Boole (1854). For more than a hundred years little attention was paid to the subject until the early sixties when Smith (1961) and Good (1962) provided important contributions to the theory. The method of DeRobertis and Hartigan (1981), *'intervals of measures'*, is especially useful when imprecise probabilities are used for parameters, as in the Bayesian framework (Pericchi and Walley, 1991). The method of DeRobertis and Hartigan is adopted here. We regard imprecision as the essential concept in reporting the amount of information, in this way our method differs from the sensitivity approach (e.g. Meczarski and Zielinski, 1991) and robust Bayesian analysis (Berger, 1985, 1990; Berger and Berliner, 1986; Lavine, 1991a,b; Moreno and Cano, 1991; Zen and DasGupta, 1990).
Recently, there has been a growing interest in using the intervals of measures theory in robust Bayesian analysis (Wasserman, 1992). Dempster (1968; see also Suppes and Zanotti, 1977) proposed a theory for constructing coherent imprecise probabilities where updating has also been considered. However, Dempster does not use a relation between imprecision and information, and Walley (1991; section 5.13) shows that the theory leads to unreasonable results in some simple situations. Wasserman and Kadane (1992) remark that for further development of a theory of statistical inference based on imprecise probabilities, it is necessary to construct useful parametric models, and propose a generalization of uniform probability distributions in the theory of imprecise probabilities, without regarding updating. Walley and Fine (1982) have introduced a frequentist theory of imprecise probabilities.

As far as we know the only earlier analysis of the updating of imprecise probabilities using a relation between information and imprecision is Walley's example for Bernoulli trials (Walley, 1991; section 5.3), considered in section 2.4. The only case study using imprecise probabilities, that we are aware of, is by Walley and Campello de Souza (1990), but they do not analyze imprecision after updating.

In section 2.3 a new theory of updating imprecise probabilities is presented, with updated imprecision based on the measure of information $I(A)$. The possibility of updating imprecise probabilities is essential to the theory. Although the imprecise probabilities represent personal betting behavior, and are therefore subjective by nature, an updating theory is needed as in many practical situations it is not possible to elicit imprecise probabilities repeatedly as new data become available, for example by restrictions in the available amount of time for the consulted experts or the costs of the elicitation process. Generally, our method uses the Bayesian concept of updating, together with a second mechanism to control imprecision, where attention is paid to the imprecision in your prior imprecise probabilities and the value you add to new information compared to the information on which you have based your betting behavior. It must be remarked that new information could be contradictory to old information, which may lead to more imprecision after updating as the expert may have more doubts than before. Our goal is to look for a mathematically simple theory of updating through which the behavior of imprecise probabilities can be analyzed. Such analysis is presented in section 2.3. Imprecision enters our model through imprecise prior probabilities, while a precise parametric model is assumed for the random variable of interest, leading to precisely defined likelihood functions. Generalization to imprecise likelihood functions, that is not considered in this thesis, is briefly discussed by Walley (1991; section 8.6.9).

The theoretically updated imprecise probabilities for A will not exactly represent your betting behavior after you have learned the new data, but in the updating theory presented in section 2.3 you are asked to compare the value of new data to the value of the amount of information on which you base your prior imprecise probabilities, and the theory relies on the relation between information and imprecision expressed by the information measure $I(A)$. In section 2.4 a detailed analysis for Bernoulli experiments is given, and in section 2.5 some remarks are made about possible methods for statistical inference with imprecise probabilities, lastly some ideas are proposed for updating imprecision in case of censored data.

2.2 A measure of information related to imprecision

As far as we know no analysis of the relation between imprecision and a measure of information has yet been presented in the literature, although this is important for a theory of statistics in which imprecision is updated as well as stochastic uncertainty if new data become available, and where a measure of information is related to the value of observations with regard to betting behavior. In this section such a relation is analyzed, and this leads to a new argument in favor of Walley's measure of information $I(A)=\Delta(A)^{-1}-1$.

Writing $i(A)$ for a measure of information related to imprecision $\Delta(A)$, logical relations between $i(A)$ and $\Delta(A)$ are discussed together with other properties of $i(A)$. In a theory of imprecise probabilities an information measure $i(A)$ should be interpreted as the value added to available information on which betting behavior on the event A, presented by your imprecise probabilities, is based. It is not assumed that new data always increase the information measure (so decrease imprecision), as it may happen that new data conflicts with old information, leading to more doubts, represented in the betting behavior through more imprecision.

Walley (1991; section 5.3.7) presents four logical assumptions for $i(A)$

1. $i(A)\geq 0$
2. $i(A)=0 \Leftrightarrow \Delta(A)=1$
3. $i(A)=\infty \Leftrightarrow \Delta(A)=0$
4. $i(A)$ is strictly decreasing as function of $\Delta(A)$.

The third assumption is based on the fact that in many situations of interest no natural upper bound for the amount of information exists.

Walley suggests the use of $I(A)=\Delta(A)^{-1}-1$, that satisfies these four assumptions, without further arguments. Since many relations between a general $i(A)$ and $\Delta(A)$ can be defined that satisfy the above assumptions, an additional argument is needed to defend the use of $I(A)$.

To this end, consider a binomial model with probability θ of success. Assume that your ability to discriminate between results is determined by the amount of information. We shall relate an interval of possible values for θ to a statement about the situation of the form (n,a), with a the most likely number of successes in a binomial experiment of size n, where n is the largest in-

teger for which one feels sure about a unique value for a. For example, you may regard $a=3$ to be the most likely result if $n=4$, but not feel yourself able to give a unique most likely result for greater values of n.

According to the binomial model, it is easily proved that a successes in an experiment of size n is the most likely result for all $\theta \in [\frac{a}{n+1}, \frac{a+1}{n+1}]$. Remark that this interval contains the classical unbiased and maximum likelihood estimator $\frac{a}{n}$ as well as the estimator $\frac{a+1}{n+2}$ from Laplace's rule of succession (Laplace, 1812; see also Dale, 1991), and that the bounds are the unbiased estimators related to the two possible situations that appear if, after an experiment of size n has led to a successes, one more trial is performed.

The probability of a success, given the value of θ, is $\mathbb{P}(X=1|\theta)=\theta$, so $\theta \in [\frac{a}{n+1}, \frac{a+1}{n+1}]$ can be translated to $\underline{\mathbb{P}}(X=1)=\frac{a}{n+1}$ and $\overline{\mathbb{P}}(X=1)=\frac{a+1}{n+1}$, leading to imprecision $\Delta(X=1)=\overline{\mathbb{P}}(X=1)-\underline{\mathbb{P}}(X=1)=\frac{1}{n+1}$. Here it is used that, if one knows θ in this situation, the expected profit of buying the bet which pays 1 if $X=1$ and 0 if $X=0$ against price p is equal to $\theta-p$, so for $p<\theta$ you expect profit when buying the bet and for $p>\theta$ you expect profit when selling the bet.

For this situation, where all available information has led to the expression (n,a), it seems reasonable to assume that the information measure should be proportional to n, the number of observations available. This is achieved by the information measure proposed by Walley, since $I(X=1)=\Delta(X=1)^{-1}-1=n$.

If the ratio a/n is a constant while n increases then it seems in agreement with intuition that the expression (n,a) is based on more information, one is more certain about the situation so to say. This idea is only used to find a relation between a measure of information and imprecision, and not proposed as a method for elicitation (discussed in section 4.2) of your opinions. Nevertheless, an example with four experts is used as an illustration. Suppose four experts have provided (n,a)

expert	(n,a)	$\underline{\mathbb{P}}(X=1)=\frac{a}{n+1}$	$\overline{\mathbb{P}}(X=1)=\frac{a+1}{n+1}$
A	(4,3)	0.60	0.80
B	(24,18)	0.72	0.76
C	(9,7)	0.70	0.80
D	(5,3)	0.50	0.67

With the related betting behavior represented by $\underline{\mathbb{P}}(X=1)$ and $\overline{\mathbb{P}}(X=1)$, only between experts B or C and D could bets be exchanged (with D selling the bet). Experts A and B have similar ideas about the ratio a/n, but B is much more certain, with the result that A does not want to buy the bet from D whereas B does.

2.3 An updating theory for imprecise probabilities

The basic concept for the theory that is developed here, is that of *intervals of measures* (DeRobertis and Hartigan, 1981). For a continuous random variable $X \in \mathbb{R}$, generalization to discrete (see section 2.4) or multi-dimensional variables is straightforward, a class of *probability density functions* (pdf) is assumed that consists of all pdf's that are proportional to unnormalized density functions that lie between specified lower and upper bounds, ℓ and u respectively, called the *lower* and *upper densities* (in our terminology a density does not integrate to one, a pdf does). It is assumed that $0 \leq \ell(x) \leq u(x)$ for all x and that ℓ and u are continuous with $\int_{-\infty}^{\infty} \ell(y)dy > 0$ and $\int_{-\infty}^{\infty} u(y)dy < \infty$. This class of pdf's for X is defined by

$$\Pi(\ell,u) = \{ \ell \mid \ell(x) = q(x)/C_q, \ \ell(x) \leq q(x) \leq u(x) \text{ for } x \in \mathbb{R}, \ C_q = \int_{-\infty}^{\infty} q(y)dy \}. \qquad (2.1)$$

The fact that the functions q in (2.1) can be normalized to give probability densities means that they are integrable. Although the notation of Riemann integrals is used throughout, if necessary generalization to Stieltjes integrals is possible.

In the literature on robust Bayesian analysis the name density ratio class is used as well as intervals of measures class (Berger, 1990). In subsection 2.3.1 results are presented for the imprecise *cumulative distribution functions* (cdf) $\underline{F}(x) = \underline{P}(X \leq x)$ and $\overline{F}(x) = \overline{P}(X \leq x)$ corresponding to $\Pi(\ell,u)$, and the according imprecision for events $X \leq x$. In subsection 2.3.2 an updating theory is proposed and analyzed.

2.3.1 On cumulative distribution functions and corresponding imprecision

Imprecise probabilities for an event $X \in S \subseteq \mathbb{R}$, corresponding to $\Pi(\ell,u)$, are (Walley, 1991; section 4.6)

$$\underline{P}(X \in S) = \frac{\int_S \ell(y)dy}{\int_S \ell(y)dy + \int_{S^c} u(y)dy} \qquad (2.2)$$

and

$$\bar{\mathbb{F}}(X \in S) = \frac{\int\limits_{S} u(y)\,dy}{\int\limits_{S} u(y)\,dy + \int\limits_{S^c} \ell(y)\,dy}. \tag{2.3}$$

Imprecise cdf's corresponding to $\Pi(\ell,u)$ are derived by setting $S=(-\infty,x]$

$$\underline{F}(x) = \frac{\int\limits_{-\infty}^{x} \ell(y)\,dy}{\int\limits_{-\infty}^{x} \ell(y)\,dy + \int\limits_{x}^{\infty} u(y)\,dy} \tag{2.4}$$

and

$$\bar{F}(x) = \frac{\int\limits_{-\infty}^{x} u(y)\,dy}{\int\limits_{-\infty}^{x} u(y)\,dy + \int\limits_{x}^{\infty} \ell(y)\,dy}, \tag{2.5}$$

These are the sharpest bounds, also called lower and upper envelopes, of all cdf's that can be constructed from pdf's that belong to $\Pi(\ell,u)$. Some simple facts about these imprecise cdf's are given in lemma 2.3.1 (the proofs are straightforward).

Lemma 2.3.1

(a) For all $x \in \mathbb{R}$: $\underline{F}(x)= \int\limits_{-\infty}^{x} \ell(y)\,dy$, with $\ell \in \Pi(\ell,u)$ $\Leftrightarrow$ For all $x \in \mathbb{R}$: $\ell(x)=u(x)$

(b) $\underline{F}(x)=0$ $\Leftrightarrow$ $\ell(y)=0$ for all $y \le x$

(c) $\bar{F}(x)=0$ $\Leftrightarrow$ $u(y)=\ell(y)=0$ for all $y \le x$

(d) $\underline{F}(x)=1$ $\Leftrightarrow$ $u(y)=\ell(y)=0$ for all $y \ge x$

(e) $\bar{F}(x)=1$ $\Leftrightarrow$ $\ell(y)=0$ for all $y \ge x$.

Without loss of generality we can write the upper density in the form

$$u(x) = \ell(x) + ca(x), \tag{2.6}$$

where $a(x) \ge 0$ for all x, and $c \ge 0$ is a constant. We introduce the notation

$$L(x) = \int\limits_{-\infty}^{x} \ell(y)\,dy, \quad L = \int\limits_{-\infty}^{\infty} l(y)\,dy, \quad A(x) = \int\limits_{-\infty}^{x} a(y)\,dy, \quad A = \int\limits_{-\infty}^{\infty} a(y)\,dy,$$

$$U(x) = \int\limits_{-\infty}^{x} u(y)\,dy = L(x)+cA(x) \quad \text{and} \quad U = \int\limits_{-\infty}^{\infty} u(y)\,dy = L+cA.$$

Unless stated otherwise, we restrict ourselves to positive L, A and c (if one of these is equal to 0 the results are obvious), $L=A=1$ can also be assumed, but this normalization of ℓ and a is not necessary. Remark that, given ℓ and u, a and c are only determined up to a constant factor. However, this choice does not affect the results that follow, and this special form will be used explicitly in subsection 2.3.2. Next, some results on imprecision and information for events $X \leq x$ are presented, using

$$\underline{F}(x) = \frac{L(x)}{L+c[A-A(x)]} \tag{2.7}$$

$$\overline{F}(x) = \frac{L(x)+cA(x)}{L+cA(x)} \tag{2.8}$$

$$\Delta(X \leq x) = \frac{c\{L(x)[A-A(x)] + A(x)[L-L(x)]\} + c^2 A(x)[A-A(x)]}{L^2 + cLA + c^2 A(x)[A-A(x)]} \tag{2.9}$$

$$I(X \leq x) = \frac{L^2 + cLA - c\{L(x)[A-A(x)] + A(x)[L-L(x)]\}}{c\{L(x)[A-A(x)] + A(x)[L-L(x)]\} + c^2 A(x)[A-A(x)]} \tag{2.10}$$

These expressions for $\Delta(X \leq x)$ and $I(X \leq x)$ are easily derived by calculation of $\Delta(X \leq x) = \overline{F}(x)-\underline{F}(x)$ and $I(X \leq x) = \Delta(X \leq x)^{-1}-1$.

Theorem 2.3.2 provides an upper bound for $\Delta(X \leq x)$.

Theorem 2.3.2

$$\Delta(X \leq x) \leq \frac{cA}{L + cA}. \tag{2.11}$$

Proof

$$\Delta(X \leq x) \leq \frac{cA}{L + cA} \quad \Leftrightarrow$$

$$\frac{L(x)[A-A(x)] + A(x)[L-L(x)] + cA(x)[A-A(x)]}{L^2 + cLA + c^2 A(x)[A-A(x)]} \leq \frac{A}{L + cA} \quad \Leftrightarrow$$

$$cK_1 + K_2 \leq 0,$$

with

$$K_1 = 2ALA(x) + A^2 L(x) - 2AA(x)L(x) - LA^2(x) - LA^2$$

and

$$K_2 = ALL(x) + L^2 A(x) - 2LA(x)L(x) - AL^2.$$

-20-

The proof is completed by showing that $K_1 \leq 0$ and $K_2 \leq 0$ for all ℓ (L), a (A) and x. Using the relation

$$AL(\theta) + LA(\theta) = AL + A(\theta)L(\theta) + [A-A(\theta)][L(\theta)-L]$$

twice, it follows that

$$K_1 = ALA(x) + A[A-A(x)][L(x)-L] - AA(x)L(x) - LA^2(x)$$

$$= ALA(x) + A[A-A(x)][L(x)-L] - A(x)\Big[AL + A(x)L(x) + [A-A(x)][L(x)-L]\Big]$$

$$= -A^2(x)L(x) + [A-A(x)]^2[L(x)-L] \leq 0,$$

since $0 \leq A(x) \leq A$ and $0 \leq L(x) \leq L$, and

$$K_2 = L\Big[AL + A(x)L(x) + [A-A(x)][L(x)-L]\Big] - 2LA(x)L(x) - AL^2$$

$$= L\Big[-A(x)L(x) + [A-A(x)][L(x)-L]\Big] \leq 0. \qquad \square$$

Next, the unknown maximum value of $\Delta(X \leq x)$, denoted by Δ_{max} is considered. A value x for which $\Delta(X \leq x) = \Delta_{max}$ is denoted by x_m, so $\Delta(X \leq x_m) = \Delta_{max}$. The corresponding minimum information $I(X \leq x_m)$ is denoted by I_{min}. Let m_a be the median of the distribution with cdf $A(x)/A$, so $A(m_a) = A/2$, and m_ℓ the median of the distribution with cdf $L(x)/L$, so $L(m_\ell) = L/2$. It is assumed that m_a and m_ℓ are unique, although the theory can be generalized to allow situations in which $A(x) = A/2$ or $L(x) = L/2$ for more than one value of x.

Corollary 2.3.3 is a simple result on Δ_{max}.

Corollary 2.3.3

$$\Delta_{max} = \frac{cA}{L + cA} \quad \Leftrightarrow \quad \exists x: \; [A(x)=A \text{ and } L(x)=0] \text{ or } [A(x)=0 \text{ and } L(x)=L]. \quad (2.12)$$

Proof

See the proof of theorem 2.3.2: $\Delta_{max} = \dfrac{cA}{L + cA} \quad \Leftrightarrow \quad K_1 = K_2 = 0.$ $\qquad \square$

According to corollary 2.3.3, for given A and L the maximum imprecision can be derived in two situations. If $A(x)=A$ and $L(x)=0$ for some x, so $a(y)=0$ for all $y > x$ and $\ell(y)=0$ for all $y < x$, then $\underline{F}(x)=0$ and $\overline{F}(x)=\dfrac{cA}{L+cA}$, while if $A(x)=0$ and $L(x)=L$ for some x, so $a(y)=0$ for all $y < x$ and $\ell(y)=0$ for all $y > x$, then

$$\underline{F}(x)=\frac{L}{L+cA} \quad \text{and} \quad \overline{F}(x)=1.$$

Lemma 2.3.4 provides a lower bound for Δ_{max}.

Lemma 2.3.4

$$\Delta_{max} \geq \frac{cA}{2L + cA}. \tag{2.13}$$

Proof

$$\Delta(X \leq m_\alpha) = \frac{cLA/2 + c^2A^2/4}{L^2 + cLA + c^2A^2/4} = \frac{cA(2L+cA)}{(2L+cA)^2} = \frac{cA}{2L + cA}. \qquad \square$$

From the bounds of Δ_{max}, following from theorem 2.3.2 and lemma 2.3.4, the statements of corollary 2.3.5 easily follow, where one should remark that the assumption that L, A and c are positive implies that $0<\Delta_{max}<1$.

Corollary 2.3.5

(a) $\quad \lim_{c \downarrow 0} \Delta_{max} = 0$ \hfill (2.14)

(b) $\quad \dfrac{\Delta_{max}L}{(1-\Delta_{max})A} \leq c \leq \dfrac{2\Delta_{max}L}{(1-\Delta_{max})A}$ \hfill (2.15)

(c) $\quad \dfrac{L}{cA} \leq I_{min} \leq \dfrac{2L}{cA}.$ \hfill (2.16)

Theorem 2.3.6 tells us more about the location of the unknown value x_m.

Theorem 2.3.6

$$\min(m_\alpha, m_\ell) \leq x_m \leq \max(m_\alpha, m_\ell). \tag{2.17}$$

Proof

$$\frac{d\Delta(X \leq x)}{dx} = \frac{c\{[A-2A(x)]K_3 + [L-2L(x)]K_4\}}{\{L^2 + cLA + c^2A(x)[A-A(x)]\}^2},$$

with

$$K_3 = L^2\ell(x) + cL^2a(x) + cLA\ell(x) + c^2L(x)A(x)a(x)$$
$$\quad + c^2[L-L(x)][A-A(x)]a(x) + c^2A(x)[A-A(x)]\ell(x)$$

-22-

and

$$K_4 = \left[L^2 + cLA + c^2A(x)[A-A(x)] \right] a(x).$$

Under the assumption that L, A and c are positive (restricted to $\ell(x)>0$ and $a(x)>0$, both K_3 and K_4 are positive)

$$x<\min(m_a,m_\ell), \text{ so } A>2A(x) \text{ and } L>2L(x) \quad \Rightarrow \quad \frac{d\Delta(X\le x)}{dx}>0, \text{ so } x<x_m$$

and

$$x>\max(m_a,m_\ell), \text{ so } A<2A(x) \text{ and } L<2L(x) \quad \Rightarrow \quad \frac{d\Delta(X\le x)}{dx}<0, \text{ so } x>x_m.$$

The restriction to positive $\ell(x)$ and $a(x)$ is allowed, since if for some $x<\min(m_a,m_\ell)$ or $x>\max(m_a,m_\ell)$ we get $\frac{d\Delta(X\le x)}{dx} = 0$ because $\ell(x)=a(x)=0$, then x is an inflection (the sign of $\frac{d\Delta(X\le x)}{dx}$ can only change for values of x between the medians). $\qquad\qquad\square$

In our model, the location of the maximum imprecision in imprecise cdf's is, by theorem 2.3.6, restricted to the area between the medians related to ℓ and a. Situations can be thought of where one wishes to model lower and upper cdf's with imprecision not maximal in this area, leaving an interesting topic for future research. For example, if X is the time to failure of a technical unit and historical information does not provide failure times but only whether or not a unit has failed before some time x_0, then your imprecise probabilities for events $X\le x$ may be highly imprecise for x less than x_0 as well as for x greater than x_0, while imprecision may be small for the event $X\le x_0$. This would lead to at least two local maxima for imprecision in the imprecise cdf's that would not necessarily be in the area between the medians related to ℓ and a.

Corollary 2.3.7 states two simple results immediately following from theorem 2.3.6 and $\Delta(X\le m_a)$, as calculated in the proof of lemma 2.3.4.

Corollary 2.3.7

(a) $m_a=m_\ell \quad \Rightarrow \quad x_m=m_a=m_\ell$ and $\Delta_{max}=\dfrac{cA}{2L + cA}$ $\qquad\qquad$ (2.18)

(b) $a(x)=\ell(x)$ for all $x\in\mathbb{R} \quad \Rightarrow \quad \Delta_{max}=\dfrac{c}{c+2}.$ $\qquad\qquad$ (2.19)

The situation $\alpha(x)=\ell(x)$ for all x, so $u(x)=(c+1)\ell(x)$, occurs in the model proposed by Coolen (1993), that is also discussed in subsection 2.3.2. Note that, for this situation, corollary 2.3.7 states that $\Delta_{max}=\frac{c}{c+2}$, whereas theorem 2.3.2 states that $\Delta(X\leq x)\leq\frac{c}{c+1}$ for all x.

2.3.2 *An updating theory for imprecise probabilities*

To develop an updating theory for imprecise probabilities for the random variable $X\in\mathbb{R}$, a parametric model is assumed with pdf $f_X(x|\theta)$, where the parameter θ is assumed to be real-valued and one-dimensional, so $\theta\in\mathbb{R}$, for ease of notation. Generalization to a parameter space $\Theta\subseteq\mathbb{R}^k$ for $k\in\mathbb{N}$ is straightforward. In most situations of interest X is observable, and we assume that the parameter θ is also random in order to derive a Bayesian scheme to model learning from experience. As in the standard Bayesian framework the introduction of a random θ is needed to take all historical information with regard to stochastic uncertainty into account, within the model assumed. The form of the model is assumed to be known exactly so that the form of the pdf $f_X(x|\theta)$ is precise. Imprecision enters through the parameter θ and is expressed as an imprecise prior for θ, modeled by a class of prior densities according to (2.1). Let ℓ_θ and u_θ be lower and upper densities for θ, then lower and upper *predictive* densities for X are defined by

$$\ell_X(x) = \int_{-\infty}^{\infty} f_X(x|\omega)\ell_\theta(\omega)d\omega \qquad (2.20)$$

and

$$u_X(x) = \int_{-\infty}^{\infty} f_X(x|\omega)u_\theta(\omega)d\omega. \qquad (2.21)$$

The class of all predictive pdf's for X related to pdf's for θ that belong to the class $\Pi_\theta(\ell_\theta,u_\theta)$ is equal to $\Pi_X(\ell_X,u_X)$, and $u_\theta(t)=\ell_\theta(t)+c\alpha_\theta(t)$ for all t implies

$$u_X(x)=\ell_X(x)+c\alpha_X(x) \qquad (2.22)$$

for all x, with

$$\alpha_X(x) = \int_{-\infty}^{\infty} f_X(x|\omega)\alpha_\theta(\omega)d\omega. \qquad (2.23)$$

Suppose that a model is assumed consisting of pdf $f_X(x|\theta)$ and *imprecise prior densities* $\ell_{\theta,0}$ and $u_{\theta,0}$, where the additional index 0 is used to indicate a

-24-

prior situation in the Bayesian context. Through $\ell_{\theta,0}$ and $u_{\theta,0}$ a class of prior pdf's $\Pi_{\theta,0}(\ell_{\theta,0}, u_{\theta,0})$ according to (2.1) is determined. This assumed model should be such that your betting behavior for events $X \leq x$ is represented well by the imprecise cdf's that relate to the corresponding *imprecise prior predictive densities* for X, denoted by $\ell_{X,0}$ and $u_{X,0}$ and derived from $\ell_{\theta,0}$ and $u_{\theta,0}$ through relations (2.20) and (2.21). The imprecise prior predictive cdf's are determined by $\ell_{X,0}$ and $u_{X,0}$ through relations (2.4) and (2.5).

A theory is needed for updating these prior densities in the case that n independent observations $X=x_i$ $(i=1...,n)$, denoted by

$$x(n)=\{x_1,...x_n\} \tag{2.24}$$

become available. Updating should lead to *imprecise posterior densities* $\ell_{\theta,n}(t \,|\, x(n))$ and $u_{\theta,n}(t \,|\, x(n))$, and corresponding *imprecise posterior predictive densities* $\ell_{X,n}(x \,|\, x(n))$ and $u_{X,n}(x \,|\, x(n))$, related to each other as in relations (2.20) and (2.21), such that the imprecise posterior cdf's for $X \leq x$ represent a logical betting behavior in the light of the new information. Now the additional index n is used to indicate a posterior situation in the Bayesian context, and our purpose is to find suitable definitions for these imprecise posterior densities in which the relation between imprecision and information plays a central role.

Walley (1991) proposes a complete theory of coherent imprecise probabilities, including updating. The basic idea of Walley's coherent updating is that your future betting behavior is entirely determined by your current betting behavior, leading to a so-called Generalized Bayes Rule (GBR). However, on developing a theory of imprecise probabilities with imprecision related to information, coherence of betting behavior through time can be illogical, since it does not seem fair to use conditional probabilities based on little information to describe betting behavior when new data have become available. Walley (1991; section 6.11.2) also recognizes this point, and states that there is a scope for other updating strategies, where it is reasonable to require only that posterior lower probabilities dominate the conditional prior lower probabilities according to the GBR. Such a theory is developed here.

We accept coherence as fundamental to the theory of imprecise probabilities, but only for sets of imprecise probabilities describing betting behavior at the same moment in time. If new information becomes available, updated betting behavior should not be based only on former imprecise probabilities, as

this would lead to the situation that betting behavior on a certain event is determined forever by the first stated imprecise probabilities on this event, that are often based on little information.

Next, the impact of updating by Walley's GBR is briefly discussed for a set $\Pi(\ell,u)$ of pdf's of the form (2.1), and a new updating theory is proposed where imprecision is controlled by additional parameters related to the information measure. The method is introduced through a simple model, and thereafter a more general model is proposed and behavior of imprecision in case of updating is discussed. The theory shows that a central role for the relation between information and imprecision in case of updating is possible, without claiming that it can only be done by a model of this form. The theory presented here is only a first step in a new direction, where the philosophy of the role of information and imprecision is original.

The likelihood function for parameter θ, with data $x(n)$, is denoted by $\mathcal{L}(\theta\,|\,x(n))$, and is equal to (based on the assumption that the random variables X_i are conditionally independent)

$$\mathcal{L}(\theta\,|\,x(n)) = \prod_{i=1}^{n} f_X(x_i\,|\,\theta). \tag{2.25}$$

According to Walley's concept of coherent updating through the GBR, the class of posterior pdf's for θ, related to the prior class $\Pi_{\theta,0}(\ell_{\theta,0},u_{\theta,0})$, would be the intervals of measures class of the form (2.1) with lower and upper densities equal to $\mathcal{L}(t\,|\,x(n))\ell_{\theta,0}(t)$ and $\mathcal{L}(t\,|\,x(n))u_{\theta,0}(t)$, respectively. It is easily seen that this posterior class is equal to the set of all standard Bayesian posterior pdf's derived from the prior pdf's in $\Pi_{\theta,0}(\ell_{\theta,0},u_{\theta,0})$. In our theory, if new data become available this class is reduced, and to explain this let us write the prior upper density for θ in the form

$$u_{\theta,0}(t)=\ell_{\theta,0}(t)+c_0 a_{\theta,0}(t). \tag{2.26}$$

Assume that $a_{\theta,0}(t)=\ell_{\theta,0}(t)$ for all $t\in\mathbb{R}$ (Coolen, 1993), so

$$u_{\theta,0}(t)=(c_0+1)\ell_{\theta,0}(t). \tag{2.27}$$

This model is also called the constant odds-ratio model (Walley, 1991; section 2.9.4), and has been studied before in the literature on robust Bayesian analysis (Berger, 1990) without the relation between imprecision and information

that is fundamental to our theory. The fundamental difference between our theory and Walley's coherent updating through the GBR is that we replace c_0 by a c_n on updating after data $x(n)$, where we will define c_n such that the imprecision in the posterior model is logically related to the amount of information. Our updating rule leads to imprecise posterior densities for θ

$$\ell_{\theta,n}(t \mid x(n)) = \mathcal{L}(t \mid x(n)) \ell_{\theta,0}(t) \tag{2.28}$$

and

$$u_{\theta,n}(t \mid x(n)) = (c_n + 1) \mathcal{L}(t \mid x(n)) \ell_{\theta,0}(t) = (c_n + 1) \ell_{\theta,n}(t \mid x(n)). \tag{2.29}$$

The corresponding imprecise posterior predictive densities are

$$\ell_{X,n}(x \mid x(n)) = \int_{-\infty}^{\infty} f_X(x \mid \omega) \ell_{\theta,n}(\omega \mid x(n)) d\omega \tag{2.30}$$

and

$$u_{X,n}(x \mid x(n)) = \int_{-\infty}^{\infty} f_X(x \mid \omega) u_{\theta,n}(\omega \mid x(n)) d\omega = (c_n + 1) \ell_{X,n}(x \mid x(n)). \tag{2.31}$$

The factor $c_n + 1$ plays the same role in the densities for θ as for X. In the following analysis the random variable X is considered, as this is the variable of interest and in many situations X is observable while the random parameter θ is not (Geisser, 1993).

The above argument shows that each update yields a new class of posterior predictive pdf's $\Pi_{X,n}(\ell_{X,n},(c_n + 1) \ell_{X,n})$. At each stage corollary 2.3.7 implies that imprecision for events $X \leq x$ is a maximum for $x = m_{\ell,n}$, the median of the probability distribution with pdf proportional to $\ell_{X,n}$. This maximum imprecision is

$$\Delta_{max,X,n} = \Lambda_{X,n}(X \leq m_{\ell,n}) = \frac{c_n}{c_n + 2}. \tag{2.32}$$

Analogously, the maximum imprecision corresponding to the class of prior predictive pdf's $\Pi_{X,0}(\ell_{X,0},(c_0 + 1) \ell_{X,0})$ is equal to

$$\Delta_{max,X,0} = \Lambda_{X,0}(X \leq m_{\ell,0}) = \frac{c_0}{c_0 + 2}. \tag{2.33}$$

If c_n remains constant, $c_n = c_0$, as in Walley's coherent updating through the GBR, then the maximum imprecision and the minimum of the information measure

do not change in the light of new information,

$$I_{min,X,n} = I_{X,n}(X \leq m_{\ell,n}) = \frac{2}{c_n} \tag{2.34}$$

and

$$I_{min,X,0} = I_{X,0}(X \leq m_{\ell,0}) = \frac{2}{c_0} \tag{2.35}$$

are equal. This leads to suggest a form of c_n that is an increasing function of n, of course with $c_n=c_0$ for $n=0$. To derive at a logical relation between the information measure $I_{X,n}$ and the number of observations n, in our model we define

$$c_n = \frac{c_0}{1 + n/\xi}, \tag{2.36}$$

with ξ a positive real number. This leads to

$$I_{min,X,n} = \frac{2(1+n/\xi)}{c_0} = (1+n/\xi)I_{min,X,0} \tag{2.37}$$

$$I_{min,X,k\xi} = (1+k)I_{min,X,0} \tag{2.38}$$

$$I_{min,X,\xi} = 2I_{min,X,0}, \tag{2.39}$$

so ξ can be interpreted as the number of new data that provide an equal amount of information as that on which the imprecise prior probabilities were based. In our theory the role of ξ is important as it urges one to compare the value of current information with that of future data, which seems to be necessary to model future betting behavior when coherence through time is not assumed.

Definition (2.36) is a simple form for c_n that brings only two additional parameters into the model to control imprecision, namely c_0 related to the imprecision in the priors, and ξ to model the value of new data compared to the value of prior information. We suggest that updating of imprecision in general parametric models cannot be controlled with less additional parameters in the model.

The assumption that $a_{\theta,0}(t)=\ell_{0,0}(t)$ for all t implies that $\ell_{0,0}$ and $a_{\theta,0}$ are equally changed by multiplication by the likelihood, which also means that new data do not support either $\ell_{X,0}$ or $a_{X,0}$ more than the other. As the assumption that $u_{X,0}$ is proportional to $\ell_{X,0}$ is rather restrictive, the behavior of

imprecision on updating in the general model is of interest. Let

$$a_{\theta,n}(t|x(n)) = \mathcal{L}(t|x(n))a_{\theta,n}(t) \tag{2.40}$$

$$a_{X,n}(x|x(n)) = \int_{-\infty}^{\infty} f_X(x|\omega)a_{\theta,n}(\omega|x(n))d\omega \tag{2.41}$$

$$u_{X,n}(x|x(n)) = \ell_{X,n}(x|x(n)) + c_n a_{X,n}(x|x(n)). \tag{2.42}$$

We introduce the notation

$$L_{X,0} = \int_{-\infty}^{\infty} \ell_{X,0}(y)dy, \qquad L_{X,n} = \int_{-\infty}^{\infty} \ell_{X,n}(y)dy,$$

$$A_{X,0} = \int_{-\infty}^{\infty} a_{X,0}(y)dy, \qquad A_{X,n} = \int_{-\infty}^{\infty} a_{X,n}(y)dy.$$

According to theorem 2.3.2 and lemma 2.3.4 the following bounds hold for the maximum prior and posterior predictive imprecision

$$\frac{c_0}{2L_{X,0}/A_{X,0} + c_0} \leq \Delta_{max,X,0} \leq \frac{c_0}{L_{X,0}/A_{X,0} + c_0} \tag{2.43}$$

and

$$\frac{c_n}{2L_{X,n}/A_{X,n} + c_n} \leq \Delta_{max,X,n} \leq \frac{c_n}{L_{X,n}/A_{X,n} + c_n}. \tag{2.44}$$

Theorem 2.3.6 gives an indication of the location of the unknown values $x_{m,0}$ and $x_{m,n}$ for which $\Delta_{max,X,0} = \Delta_{X,0}(X \leq x_{m,0})$ and $\Delta_{max,X,n} = \Delta_{X,n}(X \leq x_{m,n})$. Corollary 2.3.5 leads to the following bounds for the minimum prior and posterior predictive information

$$\frac{L_{X,0}}{c_0 A_{X,0}} \leq I_{min,X,0} \leq \frac{2L_{X,0}}{c_0 A_{X,0}} \tag{2.45}$$

and

$$\frac{L_{X,n}}{c_n A_{X,n}} \leq I_{min,X,n} \leq \frac{2L_{X,n}}{c_n A_{X,n}}. \tag{2.46}$$

These bounds depend not only on c_n, but also on the ratio $L_{X,n}/A_{X,n}$. The pro-

posed $c_n = \dfrac{c_0}{1 + n/\xi}$ implies that if this ratio is a constant,

$$L_{X,n}/A_{X,n} = L_{X,0}/A_{X,0} = B,\tag{2.47}$$

which can be interpreted as if the data support, by means of the likelihood, $\ell_{X,0}$ and $a_{X,0}$ equally, then

$$\frac{B}{c_0} \le I_{min,X,0} \le \frac{2B}{c_0}\tag{2.48}$$

and

$$\frac{B}{c_0}(1+n/\xi) \le I_{min,X,n} \le \frac{2B}{c_0}(1+n/\xi).\tag{2.49}$$

The increments in these lower and upper bounds are proportional to n, with the bounds of $I_{min,X,k\xi}$ equal to $(1+k)$ times the bounds of $I_{min,X,0}$, for $k \ge 0$ and $k\xi \in \mathbb{N}$.

If $L_{X,n}/A_{X,n} = L_{X,0}/A_{X,0}$ for all n, the first result in corollary 2.3.5 implies (since $n \to \infty$ implies $c_n \downarrow 0$)

$$\lim_{n \to \infty} \Delta_{max,X,n} = 0\tag{2.50}$$

and $I_{min,X,n}$ tends to infinity.

If n is large, the likelihood will be concentrated in a very small area of the parameter space, and $L_{X,n}/A_{X,n}$ will remain almost constant as more data become available. For small n the minimum information $I_{min,X,n}$ is still influenced by the prior information, and whether the data conflict with this information or confirms it. For large n, $I_{min,X,n}$ tends to infinity proportionally to n.

If data support $\ell_{X,0}$ and $a_{X,0}$ about equally, $L_{X,n}/A_{X,n} \approx L_{X,0}/A_{X,0}$, then the bounds of $I_{min,X,n}$ increase by the influence of c_n.

If data support $\ell_{X,0}$ more than $a_{X,0}$, so $L_{X,n}/A_{X,n} > L_{X,0}/A_{X,0}$, then the bounds of $I_{min,X,n}$ increase as well by the influence of the data as by re-placement of c_0 by c_n, so new data confirm the prior information, with the effect that you become quite sure and this is expressed in your betting behavior.

If data support $a_{X,0}$ more than $\ell_{X,0}$, so $L_{X,n}/A_{X,n} < L_{X,0}/A_{X,0}$, then the ef-

fect of the data on $I_{min,X,n}$ consists of two more or less contradictory parts. First, as the data support the difference between $u_{X,0}$ and $\ell_{X,0}$ more than $\ell_{X,0}$, the value one adds to the total amount of information decreases. Secondly, if new data become available it seems to be logical that the measure of information increases, which is modeled by c_n replacing c_0. So if the data support $a_{X,0}$ more than $\ell_{X,0}$, there are arguments for decreasing (conflict of new and old information) or increasing (new information, e.g. more observations) the bounds on $I_{min,X,n}$.

The fact that the change of the value one adds to the total amount of information depends on both the amount of new data and the way these data support $\ell_{X,0}$ and $a_{X,0}$ is a strong argument in favor of the chosen model.

The special form of $u_{\theta,0}$ leads to analytically simple updating if both $\ell_{\theta,0}$ and $a_{\theta,0}$ are members of a conjugate family (Raiffa and Schlaifer, 1961) with regard to the chosen model density $f_X(x|\theta)$. For example, if the distribution of X belongs to the one-parameter exponential family (Coolen, 1993; Lee, 1989; for the examples in this thesis models are assumed that are members of this family), where the pdf and the likelihood can, in general forms, be written as

$$f_X(x|\theta=t) = g_X(x)h_\theta(t)\exp\{s_X(x)\psi_\theta(t)\} \tag{2.51}$$

and

$$\mathcal{L}_\theta(t|x(n)) = \left[\prod_{i=1}^{n}g_X(x_i)\right]h_\theta^n(t)\exp\left\{\sum_{i=1}^{n}s_X(x_i)\psi_\theta(t)\right\}, \tag{2.52}$$

then for $\ell_{\theta,0}$ and $a_{\theta,0}$ members of the family of conjugate priors can be chosen. Generally (Lee, 1989) these conjugate priors are described by two hyperparameters v and τ, and are proportional to $h_\theta^v(t)\exp\{\tau\psi_\theta(t)\}$. Remark that the hyperparameters v and τ play similar roles as n and $\sum_{i=1}^{n}s_X(x_i)$, respectively, in the likelihoodfunction, so v and τ can be interpreted as sufficient statistics of an imaginary set of data (the prior information), since n and $\sum_{i=1}^{n}s_X(x_i)$ are sufficient statistics for data $x(n)=\{x_1,...,x_n\}$. Within our theory $\ell_{\theta,0}$ and $a_{\theta,0}$ can be chosen from this family of conjugate priors, with the hyperparameters denoted by (v_ℓ,τ_ℓ) and (v_a,τ_a) respectively. If data $x(n)=\{x_1,...,x_n\}$ become available $\ell_{\theta,0}$ and $u_{\theta,0}$ are updated by replacing c_0 by

c_n, together with replacement of (ν_ℓ, τ_ℓ) by $(\nu_\ell + n, \tau_\ell + \sum_{i=1}^{n} s_X(x_i))$, and (ν_a, τ_a) by $(\nu_a + n, \tau_a + \sum_{i=1}^{n} s_X(x_i))$.

To derive $\ell_{\theta,n}(t|x(n))$ and $a_{\theta,n}(t|x(n))$, it is not necessary to calculate the factor $\prod_{i=1}^{n} g_X(x_i)$, but, of course, these posterior densities cannot be normalized because $L_{X,n}$ and $A_{X,n}$ play an important role. Generalization to the multi-parameter exponential family (Lee, 1989) is straightforward, where inferences for normal distributions with two parameters are of special interest.

2.4 On Bernoulli experiments using imprecise probabilities

For simplicity, the theory in section 2.3 was presented for $X \in \mathbb{R}$ and $\theta \in \mathbb{R}$. In this section a standard Bernoulli situation, with discrete X and $\theta \in [0,1]$, is considered to analyze the proposed theory in more detail, especially the behavior of imprecision and information concerning X in case of updating.

Independent random variables X_i $(i=1,2,..)$ with identical binomial probability mass functions

$$\mathbb{P}(X_i = x \mid \theta) = \theta^x (1-\theta)^{1-x} \qquad (x=0 \text{ or } 1; \quad 0 \leq \theta \leq 1) \tag{2.53}$$

are sampled. We refer to the observation of a number, say k, of the independent random variables X_i as a Bernoulli experiment (if $k=1$, the experiment is also called a trial). An observation of X_i with value 1 is called a success and an observation with value 0 a failure. Let X be the number of successes in an experiment of size k ($k \in \mathbb{N}$ predetermined), then the distribution of X is the binomial, $X \sim \text{Bin}(k,\theta)$

$$\mathbb{P}_X(X=x \mid \theta) = \binom{k}{x} \theta^x (1-\theta)^{k-x} \tag{2.54}$$

for $x \in \{0,1,...,k\}$ and $0 \leq \theta \leq 1$. We are interested in the probability of a success in a future trial, or, more generally, in the probability of x successes in a future experiment of size k.

A conjugate prior distribution for θ is a beta (Colombo and Constantini, 1980), $\theta \sim \text{Be}(\alpha,\beta)$, with pdf

$$f_\theta(t) \propto t^{\alpha-1}(1-t)^{\beta-1}. \tag{2.55}$$

for $0 \leq t \leq 1$ ($\alpha,\beta > 0$). If an experiment of size n leads to a successes this prior can be updated giving a posterior $\theta \sim \text{Be}(\alpha+a,\beta+n-a)$. We denote the data by (n,a). Lee (1989; chapter 3) shows that many reasonably smooth unimodal distributions on $[0,1]$ can be approximated by some beta distribution.

The predictive distribution for X, corresponding to the conjugate prior, is a Polya distribution, $X \sim \text{Pol}(k,\alpha,\beta)$, also called a beta-binomial distribution (Raiffa and Schlaifer, 1961), with probability mass function

$$\mathbb{P}_X(X=x) = \binom{k}{x} \text{B}(\alpha+x,\beta+k-x) / \text{B}(\alpha,\beta) \tag{2.56}$$

for $x \in \{0,1,...,k\}$, with beta function

$$\mathbf{B}(\alpha,\beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}dy \quad (\alpha,\beta>0).$$ (2.57)

On updating (α,β) is replaced by $(\alpha+\alpha,\beta+n-\alpha)$.

The problem of statistical inference for Bernoulli trials, using imprecise probabilities with imprecision related to information, has also been discussed by Walley (1991; chapter 5), using a set of prior densities consisting only of beta densities. However, although his model and examples are of interest, it is not as general as the one proposed in this thesis that can be used for all parametric models, since the structure of the binomial model and beta priors is essential in Walley's model.

In subsection 2.4.1 the model is presented and imprecision is analyzed for $k=1$. In subsection 2.4.2 results for general k are given together with an example, and attention is paid to the problem of modeling lack of prior knowledge.

2.4.1 Imprecision for Bernoulli trials

Using the method presented in section 2.3, imprecise prior densities are chosen such that a conjugacy property holds, so

$$\ell_{\theta,0}(t) = t^{\alpha-1}(1-t)^{\beta-1}/\mathbf{B}(\alpha,\beta) \quad \text{for } 0\leq t\leq 1 \ (\alpha,\beta>0)$$ (2.58)

and

$$u_{\theta,0}(t) = \ell_{\theta,0}(t) + c_0 a_{\theta,0}(t)$$ (2.59)

with $c_0 \geq 0$ and

$$a_{\theta,0}(t) = t^{\lambda-1}(1-t)^{\mu-1}/\mathbf{B}(\lambda,\mu) \quad \text{for } 0\leq t\leq 1 \ (\lambda,\mu>0).$$ (2.60)

Lower and upper prior predictive probability mass functions are derived by discretized versions of (2.20)-(2.23), for $x=0,1,...k$

$$\rho_{X,0}(x) = \binom{k}{x}\mathbf{B}(\alpha+x,\beta+k-x)/\mathbf{B}(\alpha,\beta)$$ (2.61)

and

$$\bar{\rho}_{X,0}(x) = \binom{k}{x}\mathbf{B}(\alpha+x,\beta+k-x)/\mathbf{B}(\alpha,\beta) + c_0\binom{k}{x}\mathbf{B}(\lambda+x,\mu+k-x)/\mathbf{B}(\lambda,\mu). \tag{2.62}$$

Corresponding imprecise prior probabilities for the event $X=x$ are

$$\underline{P}_{X,0}(X=x) = \frac{\underline{\rho}_{X,0}(x)}{\underline{\rho}_{X,0}(x)+\sum_{i\neq x}\bar{\rho}_{X,0}(i)} \tag{2.63}$$

and

$$\bar{P}_{X,0}(X=x) = \frac{\bar{\rho}_{X,0}(x)}{\bar{\rho}_{X,0}(x)+\sum_{i\neq x}\underline{\rho}_{X,0}(i)}. \tag{2.64}$$

If data $(n,\mathfrak{a})$ become available, the imprecise prior densities are updated to imprecise posterior densities

$$\ell_{\theta,n}(t|n,\mathfrak{a}) = t^{\alpha+\mathfrak{a}-1}(1-t)^{\beta+n-\mathfrak{a}-1}/\mathbf{B}(\alpha,\beta) \tag{2.65}$$

and

$$u_{\theta,n}(t|n,\mathfrak{a}) = \ell_{\theta,n}(t|n,\mathfrak{a}) + c_n a_{0,n}(t|n,\mathfrak{a}) \tag{2.66}$$

where

$$a_{\theta,n}(t|n,\mathfrak{a}) = t^{\lambda+\mathfrak{a}-1}(1-t)^{\mu+n-\mathfrak{a}-1}/\mathbf{B}(\lambda,\mu). \tag{2.67}$$

Note that $\ell_{\theta,n}$ and $a_{0,n}$ are not normalized. The corresponding imprecise posterior predictive probabilities for $X=x$ are derived analogously to (2.61) and (2.62), now using the imprecise posterior predictive probability mass functions

$$\underline{\rho}_{X,n}(x|n,\mathfrak{a}) = \binom{k}{x}\mathbf{B}(\alpha+\mathfrak{a}+x,\beta+n-\mathfrak{a}+k-x)/\mathbf{B}(\alpha,\beta) \tag{2.68}$$

and

$$\bar{\rho}_{X,n}(x|n,\mathfrak{a}) = \binom{k}{x}\mathbf{B}(\alpha+\mathfrak{a}+x,\beta+n-\mathfrak{a}+k-x)/\mathbf{B}(\alpha,\beta)$$
$$+ c_n\binom{k}{x}\mathbf{B}(\lambda+\mathfrak{a}+x,\mu+n-\mathfrak{a}+k-x)/\mathbf{B}(\lambda,\mu). \tag{2.69}$$

The model for $k=1$ is used to analyze imprecision and to provide more concrete arguments for the role of c_n. To this end $\ell_{\theta,0}$ and $a_{\theta,0}$ are assumed to be equal, with $\lambda=\alpha>0$ and $\mu=\beta>0$, and we take a constant $c_n=c>0$ for all n instead of (2.36). Given data $(n,\mathfrak{a})$, the posterior imprecision for $X=0$ becomes (re-

mark that $\Delta_X(X=0) = \Delta_X(X=1)$, as $\overline{\mathbb{P}}_X(X=0) = 1 - \underline{\mathbb{P}}_X(X=1)$ and $\underline{\mathbb{P}}_X(X=0) = 1 - \overline{\mathbb{P}}_X(X=1)$)

$$\Delta_{X,n}(X=0 \mid n, \mathfrak{o}) = \frac{(c^2+2c)(\alpha+\mathfrak{o})(\beta+n-\mathfrak{o})}{c^2(\alpha+\mathfrak{o})(\beta+n-\mathfrak{o}) + (c+1)(\beta+\alpha+n)^2}. \qquad (2.70)$$

The prior imprecision for $X=0$ is

$$\Delta_{X,0}(X=0) = \frac{(c^2+2c)\alpha\beta}{c^2\alpha\beta + (c+1)(\beta+\alpha)^2}, \qquad (2.71)$$

As a function of α, for a constant β, this prior imprecision is a maximum if $\alpha = \beta$, strictly increasing if $\alpha < \beta$ and strictly decreasing if $\alpha > \beta$. The following analysis is based on this simple form for Δ_X.

The effect of data on the imprecision for $X=0$ is of interest, so we analyze the sign of

$$D_X(n, \mathfrak{o}) = \Delta_{X,0}(X=0 \mid 0,0) - \Delta_{X,n}(X=0 \mid n, \mathfrak{o}) \qquad (2.72)$$

where it is assumed that $n > 0$. For simplicity a continuous variable $\mathfrak{o} \in [0, n]$ is used in place of the discrete $\mathfrak{o} \in \{0, 1, ..., n\}$, thus making D_X a continuous function

$$D_X(n, \mathfrak{o}) = \frac{(c^2+2c)(c+1)q(n, \mathfrak{o})}{[c^2\alpha\beta+(c+1)(\beta+\alpha)^2][c^2(\alpha+\mathfrak{o})(\beta+n-\mathfrak{o})+(c+1)(\beta+\alpha+n)^2]} \qquad (2.73)$$

with

$$q(n, \mathfrak{o}) = (\alpha+\beta)^2(\mathfrak{o}-n)\mathfrak{o} + (\alpha^2-\beta^2)(\alpha+\beta)\mathfrak{o} + \alpha n(\beta n+\beta^2-\alpha^2). \qquad (2.74)$$

The denominator of (2.73) is positive, so the sign of $D_X(n, \mathfrak{o})$ is equal to the sign of $q(n, \mathfrak{o})$. The results are given in Table 2.4.1. Moreover $q(n, \mathfrak{o}) > 0$ means that imprecision decreases and therefore that information according to Walley's measure increases, so a '+' in table 2.4.1 indicates that information increases. In table 2.4.1 the values $\mathfrak{o}_1$ and $\mathfrak{o}_2$ are derived by

$$q(n, \mathfrak{o}) = 0 \iff \mathfrak{o} = \left(\frac{\alpha}{\alpha+\beta}\right)n =: \mathfrak{o}_1 \quad \text{or} \quad \mathfrak{o} = \beta-\alpha+\left(\frac{\beta}{\alpha+\beta}\right)n =: \mathfrak{o}_2 \qquad (2.75)$$

(note that $\mathfrak{o}_2 \in [0, n]$ does not always hold).

$\alpha,\ \beta,\ n$	sign of q	$a_1,\ a_2$
I. $\quad\alpha = \beta$	q: $\quad+\quad 0\quad +$ a: $0\qquad n/2\qquad n$	$a_1 = n/2$ $a_2 = n/2$
II. $\quad\alpha < \beta$		$0 < a_1 < n/2 < a_2$
$n < \left(\dfrac{\alpha+\beta}{\alpha}\right)(\beta-\alpha)$	q: $\quad+\quad 0\quad -$ a: $0\quad a_1\qquad n$	$a_1 < \beta-\alpha$ $a_2 > n$
$n = \left(\dfrac{\alpha+\beta}{\alpha}\right)(\beta-\alpha)$	q: $\quad+\quad 0\quad -\quad 0$ a: $0\quad a_1\qquad n=a_2$	$a_1 = \beta-\alpha$ $a_2 = n$
$n > \left(\dfrac{\alpha+\beta}{\alpha}\right)(\beta-\alpha)$	q: $\quad+\quad 0\quad -\quad 0\quad +$ a: $0\qquad a_1\quad a_2\quad n$	$a_1 > \beta-\alpha$ $a_2 < n$
III. $\quad\alpha > \beta$		$a_2 < n/2 < a_1 < n$
$n < \left(\dfrac{\alpha+\beta}{\beta}\right)(\alpha-\beta)$	q: $\quad-\qquad 0\quad +$ a: $0\qquad a_1\quad n$	$a_1 < (\alpha-\beta)\alpha/\beta$ $a_2 < 0$
$n = \left(\dfrac{\alpha+\beta}{\beta}\right)(\alpha-\beta)$	q: $0\qquad -\qquad 0\quad +$ a: $0=a_2\qquad a_1\quad n$	$a_1 = (\alpha-\beta)\alpha/\beta$ $a_2 = 0$
$n > \left(\dfrac{\alpha+\beta}{\beta}\right)(\alpha-\beta)$	q: $\quad+\quad 0\quad -\quad 0\quad +$ a: $0\quad a_2\qquad a_1\quad n$	$a_1 > (\alpha-\beta)\alpha/\beta$ $a_2 > 0$

Table 2.4.1 *Analysis of imprecision for Bernoulli trials*

This analysis shows some interesting facts about imprecision in this simpli-
fied model. First of all, if data become available with $a=a_1$, so $\dfrac{a}{n} = \dfrac{\alpha}{\alpha+\beta}$
(the relative number of successes agrees with your prior ideas), then
$D_X(n,a)=0$ for all n, and neither the imprecision nor Walley's measure of in-
formation changes. This situation suggests an increasing information measure
(decreasing imprecision), and is indeed a strong argument in favor of choosing

c_n as a decreasing function of n instead of a constant (by an analogous argument, it seems logical that $c_n \to 0$ as $n \to \infty$).

Secondly, it is worth remarking that data contradictory to the prior ideas decrease the information measure if n is small. However, a large n can also make the information measure increase (as shown by the role of a_2). Such behavior seems to be natural, and is called prior-data conflict (Walley, 1991; section 5.4).

The behavior of imprecision around a_1 is also remarkable, but easily explained within the model by considering the behavior of Δ_X as function of α and β. It is more interesting to discuss this behavior from an intuitive point of view. Here the idea is that, if the updated densities are further away from $\theta=1/2$ than the priors are, the information measure increases. If you feel, after getting new data, that θ will probably be closer to the nearest extreme value (0 or 1) than you thought before, imprecision for the probability of success at the next trial decreases.

The effect of new data can be divided into two parts. First, some aspects of the data, e.g. mean value (these aspects are often summarized by sufficient statistics), influence your thoughts. Secondly, the amount of data is important, and this aspect was not taken into consideration in the simple model for $k=1$ discussed above, but this is solved by choosing a c_n according to (2.36). Using that c_n, imprecision according to the complete model for Bernoulli experiments is discussed in subsection 2.4.2.

2.4.2 Imprecision for the binomial model

For experiments of size k, the above model leads to imprecise prior predictive cdf's $(x=0,1,...,k)$

$$\underline{F}_{X,0}(x) = \underline{P}_{X,0}(X \le x) = \frac{\sum_{i=0}^{x} \underline{p}_{X,0}(i)}{\sum_{i=0}^{x} \underline{p}_{X,0}(i) + \sum_{i=x+1}^{k} \overline{p}_{X,0}(i)} \tag{2.76}$$

and

$$\overline{F}_{X,0}(x) = \overline{P}_{X,0}(X \le x) = \frac{\sum_{i=0}^{x} \overline{p}_{X,0}(i)}{\sum_{i=0}^{x} \overline{p}_{X,0}(i) + \sum_{i=x+1}^{k} \underline{p}_{X,0}(i)} \tag{2.77}$$

-38-

with $\underline{p}_{X,0}(x)$ and $\bar{p}_{X,0}(x)$ as (2.61) and (2.62), depending on hyperparameters α, β, λ, μ and c_0. The corresponding prior imprecision for $X \leq x$ is

$$\Delta_{X,0}(X \leq x) = \bar{F}_{X,0}(x) - \underline{F}_{X,0}(x) \quad (x=0,1,..,k). \tag{2.78}$$

After updating, the hyperparameters are changed as before, and the imprecise posterior predictive cdf's and posterior imprecision are again easily derived. However, the resulting posterior imprecision and information are difficult to handle, except for their behavior related to ξ. It is easy to verify that the posterior imprecision for the event $X \leq x$ is strictly increasing as function of c_n (for $n,c_0 > 0$ and $x=0,1,..,k$)

$$\frac{d\Delta_{X,n}(X \leq x \mid n, \diamond)}{dc_n} > 0. \tag{2.79}$$

The chain rule, $\dfrac{d\Delta_{X,n}(X \leq x \mid n, \diamond)}{d\xi} = \dfrac{d\Delta_{X,n}(X \leq x \mid n, \diamond)}{dc_n} * \dfrac{dc_n}{d\xi}$, and the fact that $\dfrac{dc_n}{d\xi} > 0$, which is easy to verify from (2.36), lead to

$$\frac{d\Delta_{X,n}(X \leq x \mid n, \diamond)}{d\xi} > 0 \tag{2.80}$$

and therefore, by the definition of Walley's measure of information

$$\frac{dI_{X,n}(X \leq x \mid n, \diamond)}{d\xi} < 0. \tag{2.81}$$

Logically, $\Delta_{X,n}(X=x \mid n, \diamond)$ and $I_{X,n}(X=x \mid n, \diamond)$ depend analogously on ξ.

In the examples in this chapter the values of the hyperparameters are arbitrary, as our main goal is to analyze imprecision, information and the possibilities of the suggested model. In chapter 4 some attention will be paid to the problem of eliciting expert opinions and assessing values for the hyperparameters to fit the model.

Example 2.4.1

In table 2.4.2 results are given for the following model: $\alpha=2$, $\beta=8$, $\lambda=4$, $\mu=6$ and $c_0=1$. First the lower and upper probabilities for the events $X\leq x$ (cdf's) and $X=x$ are given for the prior situation, and thereafter posteriors for five different cases. The figures 2.4.1 represent the values of the imprecise cdf's, according to table 2.4.2, for the prior and two posterior cases.

It should be remarked that the figures 2.4.1 do not show the actual cdf's, as these are stepfunctions. The correct cdf values at the integers are connected by lines. This presentation is regarded to be more suitable to show the behavior of imprecision, especially when more cases are presented in one plot.

In this discussion we consider imprecision and information for $X=x$, but it is also interesting to study imprecision for $X\leq x$ from table 2.4.2. For the prior, the maximum imprecision is $\Delta_{max,X,0}(X=2) = 0.1934$ and minimum information $I_{min,X,0}(X=2) = 4.17$. First, with $\xi=10$, results for data (10,8) and (5,4) are compared

$$\Delta_{max,X,10}(X=6\,|\,10,8) = 0.5731 \qquad I_{min,X,10}(X=6\,|\,10,8) = 0.74$$
$$\Lambda_{max,X,5}(X=5\,|\,5,4) = 0.4521 \qquad I_{min,X,5}(X=5\,|\,5,4) = 1.21.$$

For these examples, the data obviously conflict with the prior thoughts, and this leads to a decreasing value for the total amount of information for most $X=x$, except for small values of x, for which the probabilities become very small. The other three cases represented in table 2.4.2 all have data (10,3), that do not conflict with prior opinion, and from the numbers in the table the behavior of the imprecise probabilities depending on the value of ξ can be studied. For $\xi=5$ the information of the data has more value, compared to prior thoughts, than for $\xi=10$ and $\xi=20$, and leads to less posterior imprecision

$$\Delta_{max,X,10}(X=3\,|\,10,3;\xi=5) = 0.1072 \qquad I_{min,X,10}(X=3\,|\,10,3;\xi=5) = 8.33$$
$$\Delta_{max,X,10}(X=3\,|\,10,3;\xi=10) = 0.1502 \qquad I_{min,X,10}(X=3\,|\,10,3;\xi=10) = 5.66$$
$$\Lambda_{max,X,10}(X=3\,|\,10,3;\xi=20) = 0.1881 \qquad I_{min,X,10}(X=3\,|\,10,3;\xi=20) = 4.32.$$
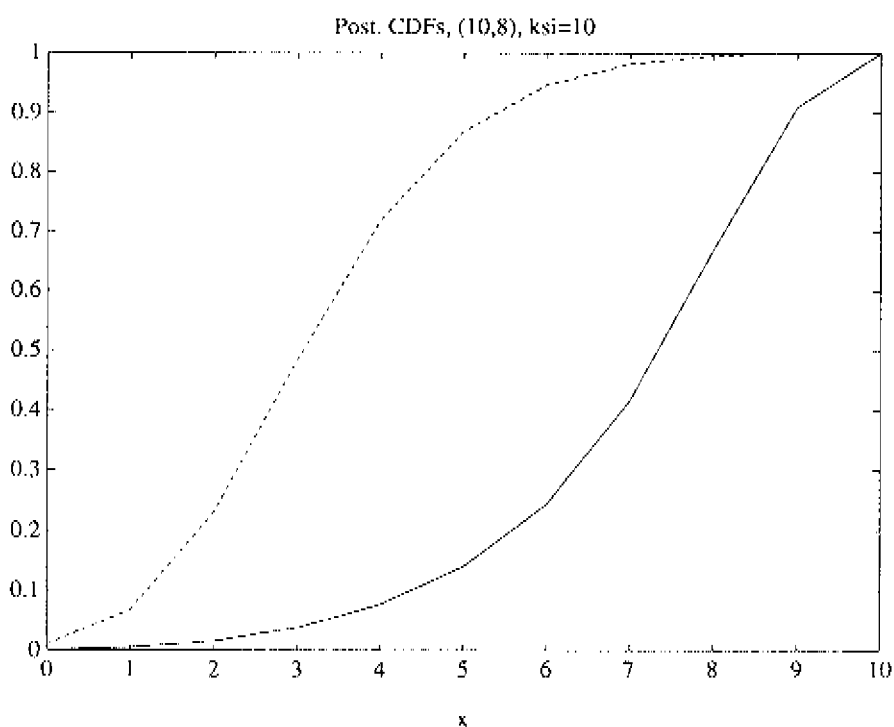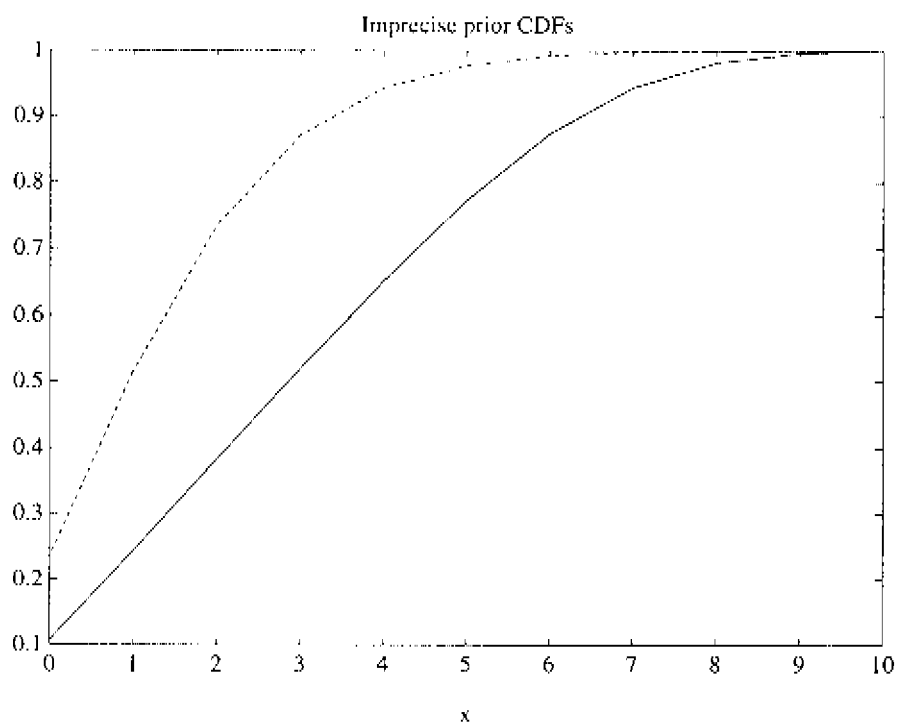
$\square$

Prior   Post. (10,8) ξ=10

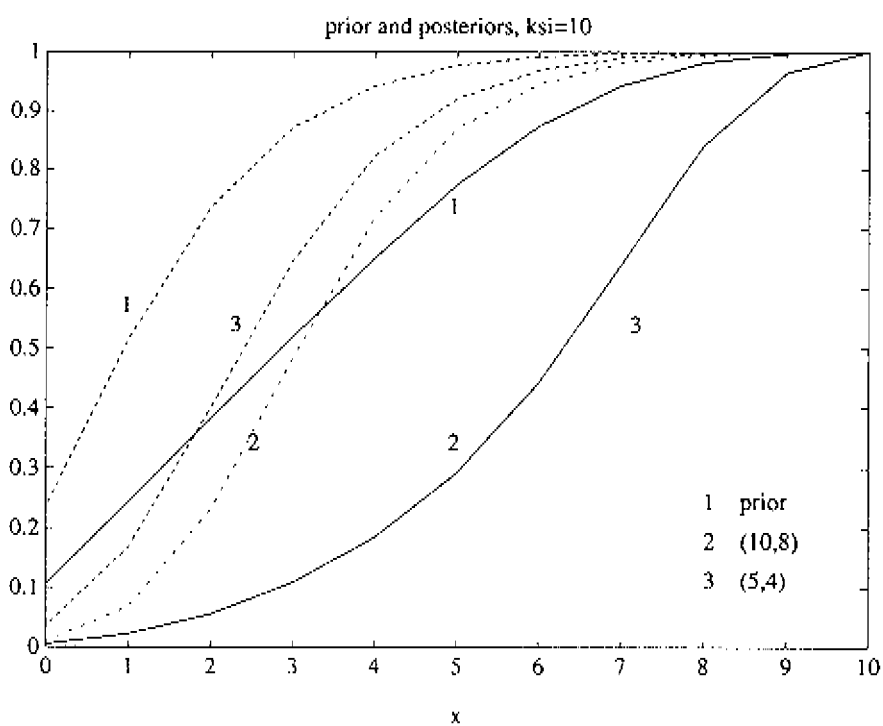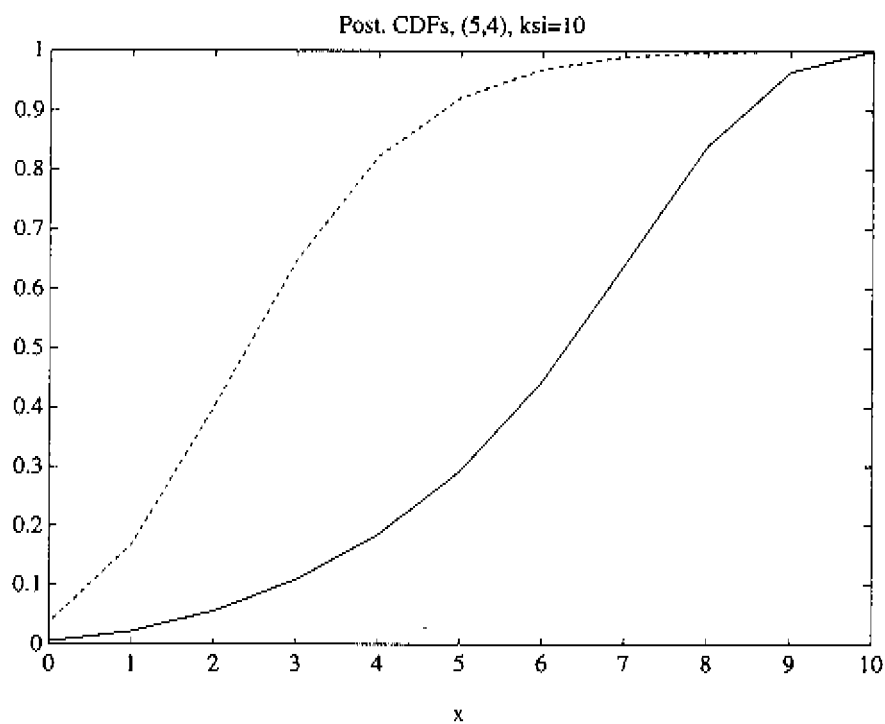| x	$\underline{F}_{X,0}$	$\overline{F}_{X,0}$	$\underline{P}_{X,0}$	$\overline{P}_{X,0}$	$\underline{F}_{X,10}$	$\overline{F}_{X,10}$	$\underline{P}_{X,10}$	$\overline{P}_{X,10}$
0	.1070	.2354	.1070	.2354	.0007	.0098	.0007	.0098
1	.2436	.5159	.1294	.3077	.0046	.0679	.0038	.0588
2	.3831	.7355	.1123	.3057	.0155	.2309	.0107	.1770
3	.5196	.8712	.0813	.2732	.0379	.4841	.0209	.3443
4	.6514	.9431	.0509	.2280	.0767	.7189	.0312	.4947
5	.7729	.9773	.0278	.1771	.1399	.8692	.0369	.5842
6	.8736	.9921	.0133	.1252	.2434	.9467	.0340	.6071
7	.9437	.9977	.0054	.0775	.4150	.9815	.0236	.5670
8	.9819	.9995	.0018	.0395	.6684	.9951	.0119	.4594
9	.9968	.9999	.0004	.0149	.9097	.9993	.0041	.2819
10	1	1	.0001	.0032	1	1	.0007	.0903

Post. (5,4) ξ=10   Post. (10,3) ξ=10

x	$\underline{F}_{X,5}$	$\overline{F}_{X,5}$	$\underline{P}_{X,5}$	$\overline{P}_{X,5}$	$\underline{F}_{X,10}$	$\overline{F}_{X,10}$	$\underline{P}_{X,10}$	$\overline{P}_{X,10}$
0	.0050	.0361	.0050	.0361	.0628	.1144	.0628	.1144
1	.0219	.1680	.0168	.1361	.2014	.3524	.1344	.2485
2	.0555	.4013	.0320	.2772	.3874	.6112	.1623	.3096
3	.1085	.6466	.0445	.4031	.5837	.8037	.1396	.2898
4	.1847	.8225	.0488	.4773	.7565	.9151	.0925	.2227
5	.2923	.9214	.0430	.4951	.8818	.9687	.0489	.1437
6	.4435	.9693	.0302	.4624	.9543	.9904	.0210	.0768
7	.6402	.9899	.0167	.3837	.9866	.9977	.0072	.0330
8	.8401	.9975	.0071	.2657	.9973	.9996	.0019	.0108
9	.9650	.9997	.0021	.1331	.9997	1.0000	.0003	.0024
10	1	1	.0003	.0350	1	1	.0000	.0003

Post. (10,3) ξ=5   Post. (10,3) ξ=20

x	$\underline{F}_{X,10}$	$\overline{F}_{X,10}$	$\underline{P}_{X,10}$	$\overline{P}_{X,10}$	$\underline{F}_{X,10}$	$\overline{F}_{X,10}$	$\underline{P}_{X,10}$	$\overline{P}_{X,10}$
0	.0714	.1090	.0714	.1090	.0561	.1198	.0561	.1198
1	.2265	.3364	.1517	.2342	.1813	.3676	.1207	.2623
2	.4280	.5904	.1818	.2877	.3539	.6299	.1466	.3302
3	.6290	.7873	.1559	.2631	.5445	.8178	.1265	.3146
4	.7936	.9060	.1034	.1957	.7227	.9227	.0836	.2480
5	.9044	.9648	.0550	.1214	.8604	.9718	.0441	.1648
6	.9644	.9892	.0237	.0624	.9444	.9915	.0188	.0908
7	.9899	.9974	.0082	.0258	.9834	.9980	.0064	.0400
8	.9980	.9996	.0021	.0082	.9966	.9997	.0017	.0133
9	.9998	1.0000	.0004	.0018	.9996	1.0000	.0003	.0030
10	1	1	.0000	.0002	1	1	.0000	.0004

Table 2.4.2   *Imprecise probabilities, example 2.4.1*

Figures 2.4.1 *Imprecise cdf's, example 2.4.1*

Imprecise prior CDFs

Post. CDFs, (10,8), ksi=10

Post. CDFs, (5,4), ksi=10

prior and posteriors, ksi=10

1   prior
2   (10,8)
3   (5,4)

The situation of no information on event A can theoretically be covered by the so-called vacuous probabilities $\underline{P}(A)=0$ and $\overline{P}(A)=1$, implying $\Delta(A)=1$ and $I(A)=0$. In our model this cannot be achieved for all events $X \leq x$ through a class of the form (2.1), unless the lower prior density is equal to zero everywhere (see lemma 2.3.1), in which case also the lower posterior density is zero everywhere which would give rise to vacuous posterior probabilities. The same happens in case of updating according to Walley's GBR. However, it is unlikely that one has no information at all, and the problem of vacuous posterior probabilities disappears for all non-vacuous prior probabilities, so total lack of prior information can be approximated by choosing the hyperparameters such that the lower probabilities for all events are arbitrarily close to zero and the upper probabilities close to one.

Within the model proposed this can be achieved by choosing c_0 large. For c_0 increasing to infinity the limiting values of the prior predictive probabilities for X, $\underline{P}_{X,0}(X=x)$ and $\overline{P}_{X,0}(X=x)$, are zero and one respectively. For the model of section 2.4.1 (where $c=c_0$) this easily follows from

$$\Delta_{X,0}(X=0) = \frac{(c^2+2c)\alpha\beta}{c^2\alpha\beta + (c+1)(\beta+\alpha)^2} \quad\quad (2.82)$$

so

$$\lim_{c \to \infty} \Delta_{X,0}(X=0) = 1. \quad\quad (2.83)$$

Example 2.4.2 shows how this idea of 'almost no prior information' can be used. It is important to remark that, as example 2.4.2 also shows, although this method seems to be promising for modeling lack of prior information, further research is necessary, for example on the form of c_n. The c_n defined by (2.36) performs well if prior knowledge is regarded as informative, but its relation to c_0 is doubtful if c_0 is very large, as is necessary when approximating prior lack of information. With the proposed form of c_n, small changes in the prior imprecision, if this is close to one, can lead to large changes in the posterior imprecision. However, the relative change (by updating) of Walley's measure of information for this discrete predictive distribution is almost equal for different c_0, and therefore insensitive to the choice of c_0.

Example 2.4.2

Let $k=10$ and $\alpha=\lambda=\beta=\mu=1$. Then

$$\underline{p}_{X,0}(x) = 1/11 \quad \text{and} \quad \overline{p}_{X,0}(x) = (c_0+1)/11 \quad \text{for all } x=0,1,...10,$$

so

$\underline{\mathbb{P}}_{X,0}(X{=}x) = 1/(11+10c_0)$ and $\bar{\mathbb{P}}_{X,0}(X{=}x) = (c_0{+}1)/(c_0{+}11)$ for all x.

The corresponding imprecise prior predictive cdf's are

$$\underline{\mathbb{F}}_{X,0}(x) = \frac{x+1}{11 + c_0(10{-}x)}$$

and

$$\bar{\mathbb{F}}_{X,0}(x) = \frac{(c_0{+}1)(x{+}1)}{11 + c_0(x{+}1)}.$$

It is easy to see that c_0 can be chosen such that $\underline{\mathbb{P}}_{X,0}(X{=}x)$ and $\bar{\mathbb{P}}_{X,0}(X{=}x)$ are arbitrarily close to 0 or 1, although 0 and 1 cannot be reached for finite c_0. This can also be done for all $\underline{\mathbb{F}}_{X,0}(x)$ and $\bar{\mathbb{F}}_{X,0}(x)$, where it should be remarked that $\underline{\mathbb{F}}_{X,0}(10) = \bar{\mathbb{F}}_{X,0}(10) = 1$.

As an example of possible application, suppose that there are 10 identical machines and the variable of interest is the number X of these machines that will operate during one month without breakdown, while nothing at all is known about these machines. One possible description of this prior lack of information is to define c_0 such that $\underline{\mathbb{P}}_{X,0}(X{=}x){\leq}0.01$ and $\bar{\mathbb{P}}_{X,0}(X{=}x){\geq}0.99$ for all x. The value $c_0{=}1000$ satisfies these constraints.

Suppose that in the first month 8 of these 10 machines operated successfully (and there is no wear out). Then your prior opinions can be updated. The parameters ξ and c_0 should already be chosen, to indicate the value you assign to new data, compared to your prior knowledge. Remember that ξ can be interpreted as the number of data that gives you an equal amount of information as you had before, so it is logical to choose ξ very small. In table 2.4.3 the results of updating are given for $\xi{=}1$ and $\xi{=}0.01$. This last value means that you regard the amount of information in a single observation as 100 times your prior information. Also results are given if the data are (60,48) instead of (10,8), indicating for example a situation wherein after six months you know that on 48 occasions a machine has operated for a month successfully, again with 10 machines in the sample. Finally, the results are given for a situation with $c_0{=}10000$, which implies $\underline{\mathbb{P}}_{X,0}(X{=}x){\leq}0.001$ and $\bar{\mathbb{P}}_{X,0}(X{=}x){\geq}0.999$. Note that the results seem to be very sensitive, but remember the above remark that Walley's measure of information is not (for example, if prior imprecision $\Delta_{X,0}(X{=}x)$ increases from 0.98 to 0.998, then information $I_{X,0}(X{=}x)$ decreases from 2/98 to 2/998). The numbers for the cdf's are also presented in the figures 2.4.2, where again not the entire cdf's, which are stepfunctions, are given, but the correct values at the integers are connected to give a clear presentation of imprecision.

The prior for $c_0=10000$ is not given in table 2.4.3. For this c_0 we have $\underline{P}_{X,0}(X=x)=0.0000$ and $\overline{P}_{X,0}(X=x)=0.9990$ for all x. So if we examine the prior imprecision for $X=x$, then

$$\Delta_{X,0}(X=x\,|\,c_0=1000) = 0.9901-0.0001 = 0.9900$$

$$\Delta_{X,0}(X=x\,|\,c_0=10000) = 0.9990-0.0000 = 0.9990$$

$$I_{X,0}(X=x\,|\,c_0=1000) = 0.010$$

$$I_{X,0}(X=x\,|\,c_0=10000) = 0.001.$$

For the corresponding imprecise posterior predictive probabilities, after data (60,48) with $\xi=0.01$, the maximum values for $\Delta_{X,60}(X=x)$ are

$$\Delta_{max,X,60}(X=8\,|\,60,48;c_0=1000) = 0.0620$$

and

$$\Delta_{max,X,60}(X=8\,|\,60,48;c_0=10000) = 0.3817.$$

The corresponding minimum values of Walley's information measure are

$$I_{min,X,60}(X=8\,|\,60,48;c_0=1000) = 15.1$$

and

$$I_{min,X,60}(X=8\,|\,60,48;c_0=10000) = 1.62.$$

The relative difference in the amount of information, that plays a central role in our theory, has changed very little after updating.

The model in this example has the form of (2.28)-(2.29), which implies that maximum imprecision for events $X\leq x$ would be given by (2.32) if X was continuous. It is easy to see that (2.32) provides an upper bound for maximum imprecision in case of discrete X, which is confirmed by the results in table 2.4.3, as shown by the following numbers (with the right-hand sides the upper bounds provided by (2.32), with c_n according to (2.36))

$$\Delta_{max,X,0}(X\leq 5\,|\,\xi=1,c_0=1000) = 0.9980 \leq 0.9980$$

$$\Delta_{max,X,10}(X\leq 8\,|\,10,8;\xi=1,c_0=1000) = 0.9781 \leq 0.9785$$

$$\Delta_{max,X,60}(X\leq 8\,|\,60,48;\xi=1,c_0=1000) = 0.8773 \leq 0.8913$$

$$\Delta_{max,X,10}(X\leq 7\,|\,10,8;\xi=0.01,c_0=1000) = 0.3302 \leq 0.3331$$

$$\Delta_{max,X,60}(X\leq 8\,|\,60,48;\xi=0.01,c_0=1000) = 0.0715 \leq 0.0769$$

$$\Delta_{max,X,60}(X\leq 8\,|\,60,48;\xi=0.01,c_0=10000) = 0.4283 \leq 0.4544. \qquad \square$$

x	Prior $\quad c_0=1000$				Post. $(10,8)$ $\xi=1$ $c_0=1000$			
	$\underline{F}_{X,0}$	$\overline{\overline{F}}_{X,0}$	$\underline{P}_{X,0}$	$\overline{P}_{X,0}$	$\underline{F}_{X,10}$	$\overline{\overline{F}}_{X,10}$	$\underline{P}_{X,10}$	$\overline{P}_{X,10}$
0	.0001	.9901	.0001	.9901	.0000	.0169	.0000	.0169
1	.0002	.9955	.0001	.9901	.0000	.1277	.0000	.1144
2	.0004	.9973	.0001	.9901	.0001	.4043	.0001	.3467
3	.0006	.9983	.0001	.9901	.0003	.6948	.0002	.6115
4	.0008	.9988	.0001	.9901	.0007	.8617	.0004	.7899
5	.0012	.9992	.0001	.9901	.0018	.9374	.0009	.8841
6	.0017	.9994	.0001	.9901	.0040	.9711	.0016	.9308
7	.0027	.9996	.0001	.9901	.0088	.9869	.0024	.9535
8	.0045	.9998	.0001	.9901	.0215	.9946	.0030	.9626
9	.0099	.9999	.0001	.9901	.0713	.9985	.0028	.9599
10	1	1	.0001	.9901	1	1	.0015	.9287

x	Post. $(60,48)$ $\xi=1$ $c_0=1000$				Post. $(10,8)$ $\xi=.01$ $c_0=1000$			
	$\underline{F}_{X,60}$	$\overline{\overline{F}}_{X,60}$	$\underline{P}_{X,60}$	$\overline{P}_{X,60}$	$\underline{F}_{X,10}$	$\overline{\overline{F}}_{X,10}$	$\underline{P}_{X,10}$	$\overline{P}_{X,10}$
0	.0000	.0000	.0000	.0000	.0001	.0004	.0001	.0004
1	.0000	.0006	.0000	.0005	.0008	.0032	.0007	.0028
2	.0000	.0063	.0000	.0058	.0037	.0145	.0029	.0114
3	.0002	.0440	.0001	.0381	.0122	.0472	.0085	.0331
4	.0008	.1925	.0006	.1607	.0328	.1193	.0201	.0756
5	.0031	.4882	.0023	.4103	.0754	.2457	.0399	.1423
6	.0103	.7597	.0065	.6635	.1547	.4223	.0682	.2264
7	.0305	.9050	.0142	.8130	.2906	.6208	.1004	.3085
8	.0905	.9678	.0219	.8712	.5029	.8017	.1230	.3591
9	.3109	.9927	.0191	.8551	.7794	.9338	.1154	.3426
10	1	1	.0073	.6891	1	1	.0662	.2206

x	Post. $(60,48)$ $\xi=.01$ $c_0=1000$				Post. $(60,48)$ $\xi=.01$ $c_0=10000$			
	$\underline{F}_{X,60}$	$\overline{\overline{F}}_{X,60}$	$\underline{P}_{X,60}$	$\overline{P}_{X,60}$	$\underline{F}_{X,60}$	$\overline{\overline{F}}_{X,60}$	$\underline{P}_{X,60}$	$\overline{P}_{X,60}$
0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0001
2	.0003	.0004	.0003	.0004	.0001	.0010	.0001	.0009
3	.0023	.0031	.0019	.0027	.0010	.0070	.0009	.0060
4	.0116	.0157	.0093	.0127	.0051	.0353	.0041	.0285
5	.0449	.0601	.0331	.0446	.0201	.1276	.0148	.0964
6	.1348	.1749	.0885	.1168	.0638	.3264	.0408	.2321
7	.3194	.3897	.1764	.2258	.1703	.5934	.0857	.3999
8	.5972	.6687	.2500	.3120	.3935	.8218	.1273	.5090
9	.8706	.9015	.2253	.2835	.7464	.9544	.1129	.4749
10	1	1	.0985	.1294	1	1	.0456	.2536

Table 2.4.3 *Imprecise probabilities, example 2.4.2*

Figures 2.4.2 *Imprecise cdf's, example 2.4.2*

prior CDFs

post. CDFs, ksi=1, c=1000

1. (10,8)
2. (60,48)
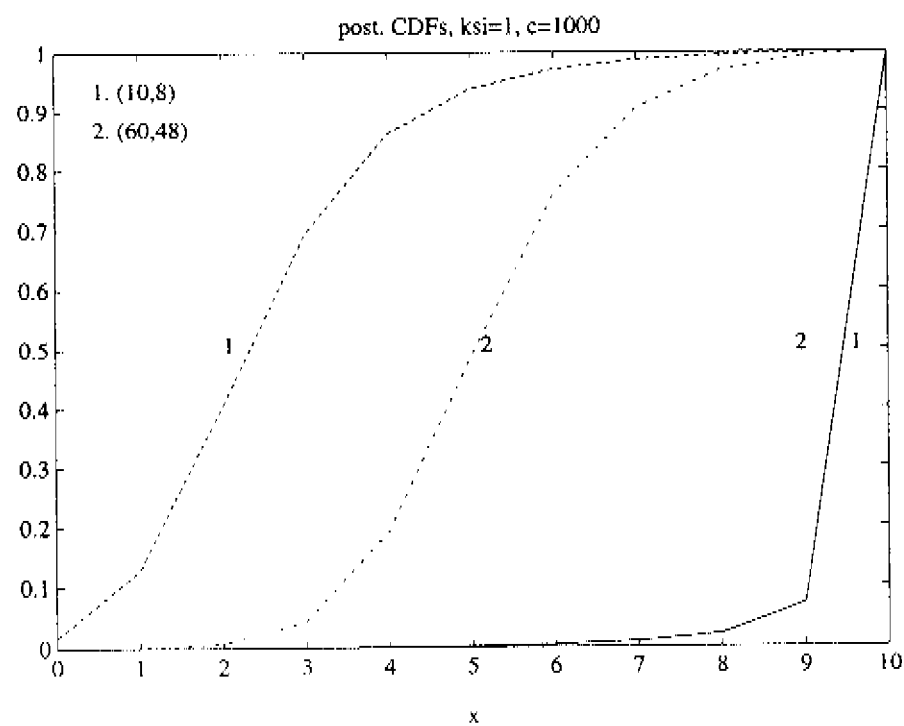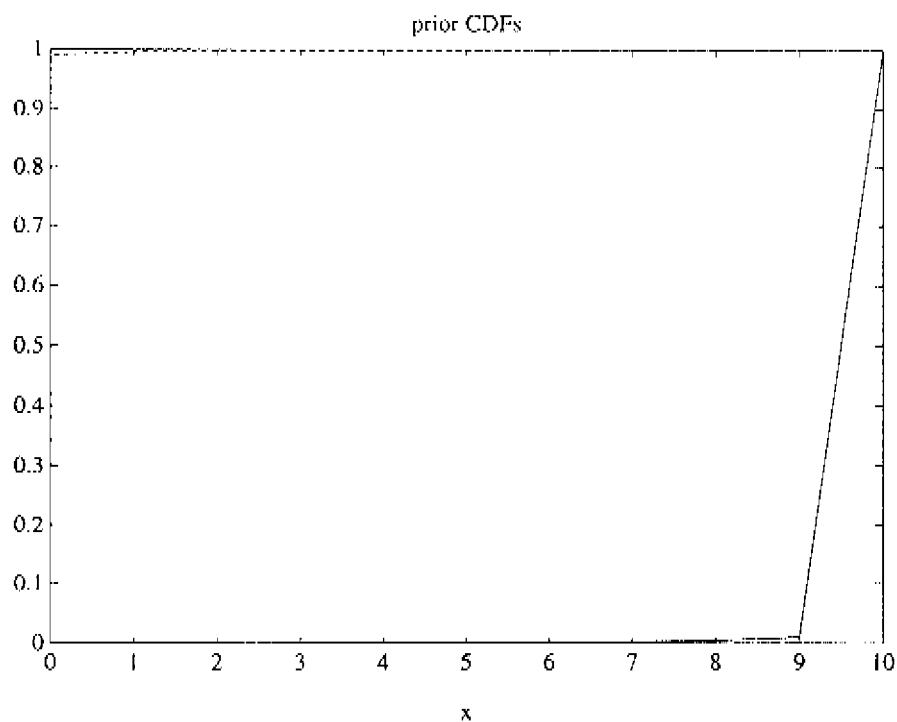
post. CDFs (60,48), ksi=.01
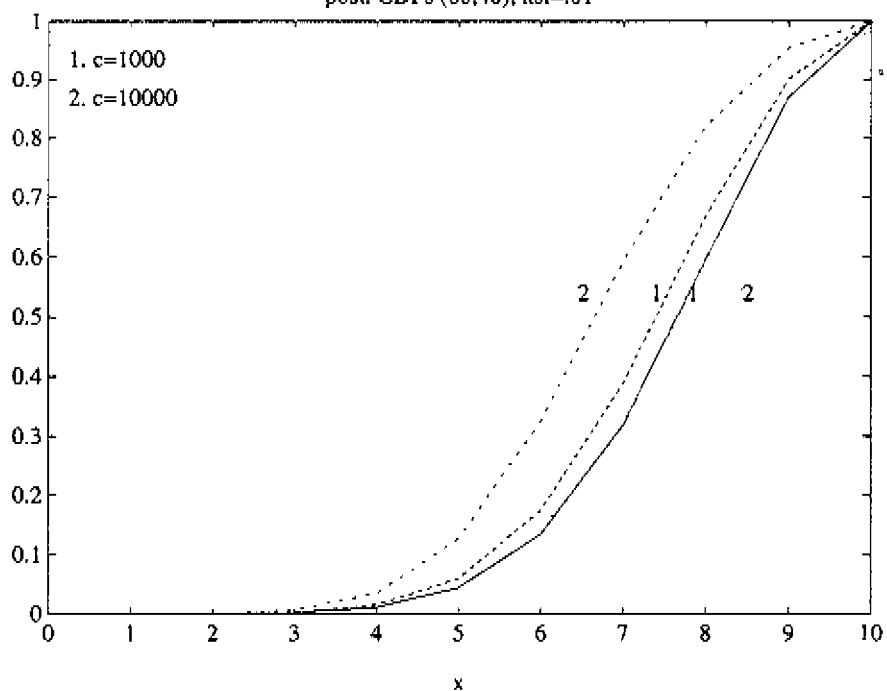
1. c=1000
2. c=10000

post. CDFs (10,8), c=1000

1. ksi=1
2. ksi=.01
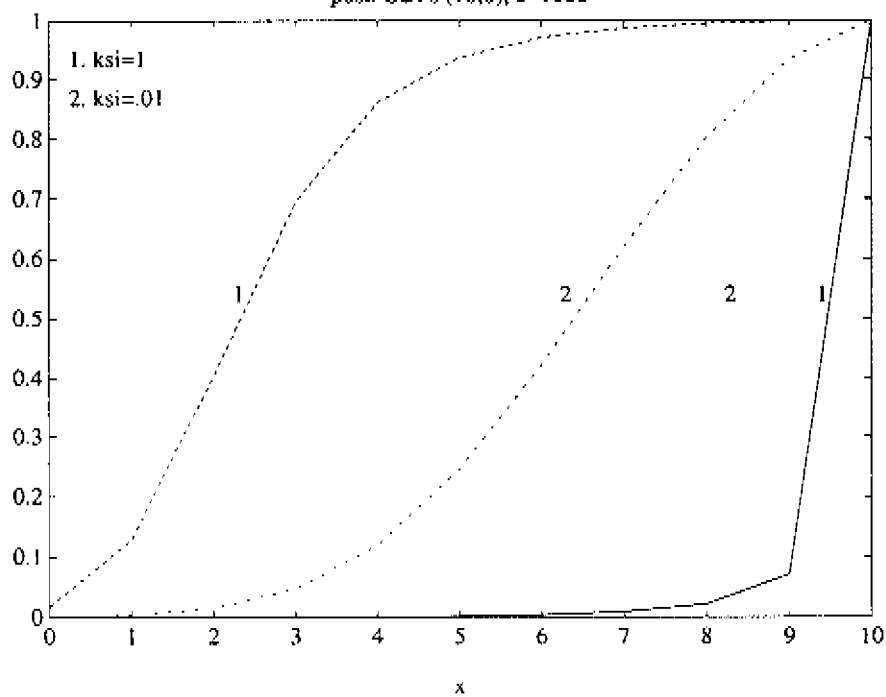
2.5 Additional remarks

At the end of this chapter we briefly mention some aspects of methods of statistical inference based on imprecise probabilities, and some attention is paid to updating in case of censored data.

In the classical Bayesian framework of statistical inference (Lindley, 1990) all statistical inferences are based on the combination of a pdf for the parameters of the assumed parametric model and a loss function. Generalization of this approach to our concept is presented in chapter 3.

A different approach is the use of highest density regions for either the distribution of the parameters or for the predictive distribution of X. Many Bayesian statisticians restrict themselves to inferences about parameters, but to our opinion it is more useful to emphasize predictive distributions for X, the random variable of interest itself (Geisser, 1993).

Let q be a pdf of X, then a $(1-\alpha)$-*Highest Density Region* (HDR) for X, denoted by $\mathcal{R}_q(\alpha)$, is such that $(0 \leq \alpha \leq 1)$

$$\mathbb{P}(X \in \mathcal{R}_q(\alpha)) = 1-\alpha \qquad (2.84)$$

and

$$\int_{\mathcal{R}_q(\alpha)} dy \text{ is a minimum.}$$

In case of a unimodal continuous probability distribution for a real random variable $\mathcal{R}_q(\alpha)$ is the shortest interval with probability $1-\alpha$.

As a generalization of highest density regions to the theory of imprecise probabilities, using a class of pdf's $\Pi(\ell,u)$ according to (2.1), we define a $(1-\alpha)$-*Imprecise Highest Density Region* (IHDR) $\mathcal{R}_{\ell,u}(\alpha) \subseteq \mathbb{R}$ for random variable X such that $(0 \leq \alpha \leq 1)$

$$\underline{\mathbb{P}}(X \in \mathcal{R}_{\ell,u}(\alpha)) = 1-\alpha \qquad (2.85)$$

and

$$\int_{\mathcal{R}_{\ell,u}(\alpha)} dy \text{ is a minimum.}$$

Another interesting region is found by adding the condition that $\mathcal{R}_{\ell,u}(\alpha)$ should be an interval. Both kinds of regions are special cases of sets studied by DasGupta (1991). Theorem 2.5.1 is a simplified form of DasGupta's theorem 2.1, written in terms of the class $\Pi(\ell,u)$, which says that $\mathcal{R}_{\ell,u}(\alpha)$ is equal to a precise highest density region for one particular pdf in the class $\Pi(\ell,u)$. Again we write $u(x)=\ell(x)+ca(x)$ with $\int_{-\infty}^{\infty} \ell(y)dy = \int_{-\infty}^{\infty} a(y)dy = 1$, which can be assumed without loss of generality.

Theorem 2.5.1

$$\mathcal{R}_{\ell,u}(\alpha) = \mathcal{R}_{q}(\alpha_c) \qquad (2.86)$$

with

$$q(x) = \frac{\ell(x) + (1-\alpha)ca(x)}{1 + (1-\alpha)c} \qquad (2.87)$$

so $q \in \Pi(\ell,u)$, and

$$\alpha_c = \frac{\alpha}{1 + (1-\alpha)c}. \qquad (2.88)$$

Proof
See theorem 2.1 of DasGupta (1991). □

Note that $\mathcal{R}_{\ell,u}(\alpha)$ is not the union of all $(1-\alpha)$-HDR's related to pdf's in $\Pi(\ell,u)$, that $\alpha_c < \alpha$ if $c>0$ and $0<\alpha<1$, and that α_c is decreasing as a function of c.
If a is identical to ℓ also the pdf q in theorem 2.5.1 is equal to ℓ. A simple example is added with a equal to ℓ.

Example 2.5.2
Let $\ell(x)$ be the pdf of a $N(\mu,\sigma^2)$ distribution and $u(x)=(1+c)\ell(x)$. The $(1-\alpha)$-IHDR is equal to the $(1-\alpha_c)$-HDR according to ℓ,

$$[\mu - \sigma\Phi^{-1}(1-\alpha_c/2) \, , \, \mu + \sigma\Phi^{-1}(1-\alpha_c/2)],$$

with Φ the standard normal cdf.

The upper probability of the event $X \in \mathcal{R}_{\ell,u}(\alpha)$ is calculated according to (2.2). For the special case that a and ℓ are identical this leads to

$$\overline{\mathbb{P}}(X \in \mathcal{R}_{\ell,(1+c)\ell}(\alpha)) = 1 - \frac{\alpha}{(1-\alpha)(1+c)^2 + \alpha} \tag{2.89}$$

and

$$\Lambda(X \in \mathcal{R}_{\ell,(1+c)\ell}(\alpha)) = \frac{\alpha(1-\alpha)c(2+c)}{(1-\alpha)(1+c)^2 + \alpha}. \tag{2.90}$$

It is easily verified that $\overline{\mathbb{P}}(X \in \mathcal{R}_{\ell,(1+c)\ell}(\alpha)) > 1-\alpha$ for $0<\alpha<1$ and $c>0$.

In terms of upper probabilities it may be interesting to regard a region $Q_{\ell,u}(\alpha) \subseteq \mathbb{R}$ such that $(0 \leq \alpha \leq 1)$

$$\overline{\mathbb{P}}(X \in Q_{\ell,u}(\alpha)) = 1-\alpha \tag{2.91}$$

and

$$\int_{Q_{\ell,u}(\alpha)} dy \quad \text{is a maximum.}$$

However, axiom IP4 of section 2.1, $\underline{\mathbb{P}}(A)+\overline{\mathbb{P}}(A^c)=1$, implies

$$Q_{\ell,u}(\alpha)=\mathcal{R}^c_{\ell,u}(1-\alpha) \tag{2.92}$$

so this adds nothing to the theory.

Next, we briefly propose a generalization of our concept to deal with censored data, which is an interesting topic for future research.

We restrict the discussion of censoring to a simple mechanism that is of interest if X represents the lifetime of technical equipment. Assume that, as before, data $x(n)$ can become available, together with data $z(m)=\{z_1,...,z_m\}$ representing independent observations $X>z_j$.

In the case of censored data there are no problems with the likelihood that is used in updating imprecise probabilities. If one assumes that one censored observation z_j provides an equal amount of information as one observation x_i, then our method for updating imprecise probabilities can be used, with c_{n+m} according to (2.36). However, it seems to be logical that one would like to take the difference of the data into account; x_i may be more informative than z_j. We suggest that this situation could be modeled by defining

$$c_{n,m} = \frac{c_0}{1 + n/\xi + m/\zeta} \tag{2.93}$$

-52-

with ξ as before, and ζ playing an analogous role in comparing the information provided by censored data to the prior information. A straightforward generalization of (2.37), for the model with a and ℓ identical, would be

$$I_{min,X,n,m} = (1+n/\xi+m/\zeta)I_{min,X,0}$$ (2.94)

Other results of section 2.3 are also easily generalized.

In a theory of imprecise probabilities that gives a central role to the relation between imprecision and information, one is urged to express the value, in terms of information, that one assigns to possible observations, compared to the prior amount of information. This may be rather complicated in case of several censoring mechanisms, especially when they are unknown at the moment the prior imprecise probabilities are elicited. Although this leaves many questions open, it seems to be a natural feature when modeling expert opinions by a theory related to betting behavior. If data become available in a form that was not considered while assessing the prior imprecise probabilities, it is quite logical that problems with updating arise, and probably the only reasonable solution is to visit the expert again and to repeat the elicitation process, with more information available.

Chapter 3

BAYESIAN DECISION THEORY USING IMPRECISE PROBABILITIES

3.1 Introduction

Without suggesting that every decision in life should be based on a quantitative analysis of the situation, it is believed that quantitative methods can be valuable tools for making good decisions. Citing Gärdenfors and Sahlin (1983), 'decision making would be easier if we knew all the relevant facts about the decision alternatives open to us', which suggests that generally a distinction could be made between problems where all relevant facts are known and problems where some facts are not known, bringing uncertainty to the decision problem. However, it is necessary to distinguish between two fundamentally different kinds of uncertainty, stochastic uncertainty and uncertainty by lack of information. This is discussed by Walley (1991; section 5.1.2), who uses the terms determinacy and indeterminacy instead of stochastic uncertainty and uncertainty by lack of information, respectively.

In this chapter Bayesian decision theory with imprecise probabilities is discussed as a tool for dealing with both kinds of uncertainty. In section 3.2 some aspects of mathematical decision theory are illustrated and a philosophical background for the methodology is discussed, with emphasis on imprecise probabilities as the concept to be used to quantify opinions. Within the concept of Bayesian decision theory with imprecise probabilities, imprecision enters the model through a prior distribution for parameters.
A method for calculating bounds on the expected value of a loss function is needed and some results for this are presented in section 3.3. Section 3.4 presents an application to a simple problem of replacement planning for technical equipment, illustrating the attractive features of the concept compared with a precise Bayesian approach. The approach is especially attractive to non-mathematicians, which is important for convincing decision makers of the power of quantitative methods as a tool to help them to make good decisions. In section 3.5 some additional remarks are made.

3.2 Decision making using imprecise probabilities

In many decision problems where uncertainty plays a role, there is one person who has to make a decision, and therefore has the responsibility for the decision. This person is called the *decision maker* (DM), and we assume that the DM wants to make the best decision (the definition of 'best' is not always trivial). To make a decision, the DM needs information about the situation. Possible sources of information are historical data, experimental data and expert opinions, and of course the DM's own opinions and knowledge can be included.

Historical data may contain useful information, especially if it is reasonable to assume that these data come from a similar situation to the one of interest. When exchangeability can be assumed (De Finetti, 1974; see also Cooke, 1991; chapter 7) the value of historical data as source of information is beyond doubt. However, even when the current situation is not identical to the past (exchangeability cannot be assumed), historical data about related situations can contain useful information. This is the main body of what is called experience, which is strongly related to expertise. Most decisions in life are based on human knowledge, ideas or intuition.

We call people who are consulted by a DM *experts*, and *opinion* is a collective noun for the possible states of mind. There is obviously a need for a sound and general methodology do deal with expert opinion, and a useful recent monograph is written by Cooke (1991). However, Cooke restricts his discussion to precise probabilities as the mathematical concept for handling uncertainty, also when used to transfer information about uncertain events from experts to a DM. We believe that this restriction is too narrow, as important information is not quantified. For example, consider a DM who has to decide about betting on the outcome of two different tennis matches (this is a version of an example by Gärdenfors and Sahlin, 1982). The DM does not know anything about tennis at all, and consults Miss Julie since he knows that she has some knowledge of the game, and perhaps also of the matches of interest. We further assume that the DM and Miss Julie can only communicate by means of probabilities, either precise or imprecise. This may seem strange here, but a methodology is needed such that in decision problems the information can be stored, for which means numbers are suited, where these numbers must have an interpretation known to everybody.

Gärdenfors and Sahlin (1982) give the following story. As regards match A, Miss Julie is very well-informed about the two players - she knows everything about the results of their earlier matches, she has watched them play several

times, she is familiar with their present physical condition and the setting of the match, etc. Given all this information, she predicts that it will be a very even match and that a mere chance will determine the winner. In match B, she knows nothing whatsoever about the relative strength of the contestants (she has not even heard their names before) and she has no other information that is relevant for predicting the winner of the match.

If *pressed* to evaluate precise probabilities of the various possible outcomes of the matches, Miss Julie would say that in both matches, given the information she has, each of the players has a 50% chance of winning. In this situation a strict Bayesian would say that Miss Julie should be willing to bet at equal odds on one of the players winning in one of the matches if and only if she is willing to place a similar bet in the other match. It seems, however, perfectly rational if Miss Julie decides to bet on match A, but not on B.

This difference between her opinions about both matches should also be represented by the quantified expression of her thoughts, and is obviously of interest to the DM. When using the concept of imprecise probabilities to describe her betting behavior, and thereby quantifying her opinions, Miss Julie could represent the difference between both matches in a logical way, and the idea behind it is that she would be free to bet if she was offered the possibility, and was not pressed. By means of the different imprecise probabilities Miss Julie gives the DM essential information that is lost through the use of precise probabilities.

The combination of opinions of several experts is also important. In chapter 4 combination is more fully considered, but at this point it is important to stress the fact that a sensible interpretation is available for some forms of combined imprecise probabilities for all experts, which is not the case when using precise probabilities. For example, the minimum of the individual lower probabilities for a certain event is the supremum of the prices for which all experts want to buy the bet (see section 4.3 for more details). One could also say that there is consensus if the imprecise probabilities per expert reflect that no experts want to exchange bets. This offers a DM a useful tool for reporting important information on the consensus among the experts. With the interpretation of precise probabilities in the standard Bayesian framework two experts always seem to be willing to exchange a bet on an event, except when their probabilities are equal. If the opinion of a single expert is elicited by different methods using imprecise probabilities (as discussed in section 4.2), combination of these imprecise probabilities does not lead to problems of incoherence of the same kind as confronted with when using the

classical Bayesian methodology.

Many decision problems of practical interest are too complex to ask an expert to state directly the best decision, and this is the point where mathematical modeling enters the framework. Lindley (1990) gives an overview of the Bayesian ideology of decision making that can be generalized to allow imprecise prior probabilities in decision making. An overview of important contributions and interesting ideas on decision making with imprecise probabilities is presented by Walley (1991; section 1.8).

Decision problems (Lindley, 1990) can, with a good deal of generality, be described as follows. Let $\mathcal{X}$ be the sample space of points x and Θ the parameter space of points θ. These are connected by a probability density for the random variable X, for a given θ. Let D be the decision space of points d, and $L(d,\theta)$ the *loss* in selecting d when θ obtains (we assume $L(d,\theta) \in \mathbb{R}$ for all d and θ). The Bayesian approach uses a prior probability density $\pi(\theta)$ over the parameter space Θ and chooses as the optimum decision that d which minimizes the *expected loss* (if such d exists)

$$EL(d \mid x) = \int_\Theta L(d,\theta)\pi(\theta \mid x)d\theta \qquad (3.1)$$

where $\pi(\theta \mid x)$ is the posterior probability density of θ, given x, obtained by Bayes' theorem. If no data x are available, the decision can be based on the prior, in which case the optimum decision is that d which minimizes the expected loss

$$EL(d) = \int_\Theta L(d,\theta)\pi(\theta)d\theta. \qquad (3.2)$$

Although in literature notation in terms of posteriors is often preferred (e.g. Moreno and Pericchi, 1992), without loss of generality we use the form (3.2), since only the fact that there is some pdf for θ is of interest here, and the posterior at one stage is just the prior for the next.

In the context of this thesis, the assumption of a single pdf $\pi(\theta)$ is generalized by assuming a class Π of pdf's for θ. Throughout we assume that the loss function $L(d,\theta)$ is known precisely. The expected loss for $\pi \in \Pi$ and $d \in D$ is

$$EL(d,\pi) = \int_{\Theta} L(d,\theta)\pi(\theta)d\theta. \qquad (3.3)$$

For each decision $d \in D$ this leads to a set values of expected loss, denoted by

$$\mathcal{EL}(d,\Pi) = \{EL(d,\pi) \mid \pi \in \Pi\}. \qquad (3.4)$$

To choose an optimal decision an additional criterion is needed, by which these sets $\mathcal{EL}(d,\Pi)$ can be compared for all $d \in D$. To this end, two useful characteristics of $\mathcal{EL}(d,\Pi)$ are the lower expected loss $\underline{\mathcal{EL}}(d,\Pi)$ and the upper expected loss $\overline{\mathcal{EL}}(d,\Pi)$

$$\underline{\mathcal{EL}}(d,\Pi) = \inf \mathcal{EL}(d,\Pi) \qquad (3.5)$$

and

$$\overline{\mathcal{EL}}(d,\Pi) = \sup \mathcal{EL}(d,\Pi). \qquad (3.6)$$

According to De Finetti (1974; chapter 3), without the generalization that allows imprecision, your prevision for the random variable X should be equal to the mean value according to the probability distribution for X if this distribution is known, so $\underline{\mathcal{EL}}(d,\Pi)$ and $\overline{\mathcal{EL}}(d,\Pi)$ can also be interpreted as the lower and upper previsions of the random loss. For the generalized theory of previsions we refer to Walley (1991).

Since the main goal is to minimize expected loss, $d_1 \in D$ is obviously preferred to $d_2 \in D$ if $\overline{\mathcal{EL}}(d_1,\Pi) < \underline{\mathcal{EL}}(d_2,\Pi)$. For other situations the preference is less clear, although it seems reasonable that d_1 is also preferred to d_2 if both $\underline{\mathcal{EL}}(d_1,\Pi) < \underline{\mathcal{EL}}(d_2,\Pi)$ and $\overline{\mathcal{EL}}(d_1,\Pi) < \overline{\mathcal{EL}}(d_2,\Pi)$. Walley (1991; section 3.9.7) briefly proposes a minimax rule as possible additional criterion, leading to the choice of a decision d_0 with $\overline{\mathcal{EL}}(d_0,\Pi) \leq \overline{\mathcal{EL}}(d,\Pi)$ for all $d \in D$, but Walley immediately remarks that there does not seem to be any good reason to prefer this rule to other possibilities. In fact, it appears to be attractive to confront the DM with the consequences of the lack of perfect information by presenting a set of possible values of expected loss corresponding to each possible decision, based on the current amount of information with no further distinction possible. For practical acceptance of a quantitative decision theory it is better that lack of perfect information is reported by means of imprecise expected loss than that an optimal decision is presented, based on precise loss as the result of a classical Bayesian analysis, especially if one has little information.

In the framework of the theory proposed in chapter 2, we restrict Π to the form given by (2.1). In section 3.3 we focus on the calculation of $\underline{\mathcal{E}\mathcal{L}}(d,\Pi)$ and $\overline{\mathcal{E}\mathcal{L}}(d,\Pi)$.

As remarked before, in this thesis the loss function $L(d,\theta)$ is assumed to be precisely known. If this cannot be assumed, e.g. because some cost figures are not precisely known, but it is known that $\underline{L}(d,\theta) \leq L(d,\theta) \leq \overline{L}(d,\theta)$, then generalization of a decision theory based on comparison of possible decisions by $\underline{\mathcal{E}\mathcal{L}}(d,\Pi)$ and $\overline{\mathcal{E}\mathcal{L}}(d,\Pi)$ is straightforward by determining $\underline{\mathcal{E}\mathcal{L}}(d,\Pi)$ corresponding to loss function $\underline{L}(d,\theta)$ and $\overline{\mathcal{E}\mathcal{L}}(d,\Pi)$ corresponding to $\overline{L}(d,\theta)$.

Estimation of a parameter plays an important role in frequentist statistics, but is essentially non-Bayesian, since all knowledge about the parameter is represented by a distribution. Nevertheless, one may be interested in some estimator, and in the standard Bayesian theory such is derived through the choice of an appropriate loss function (where squared-error loss is often used since this leads, in many cases, to estimators that have some nice frequentist properties, an argument that should never be used within the Bayesian framework but is sometimes reasonable to convince non-Bayesian statisticians). This has led to a Bayesian estimation theory which is obviously nothing more than Bayesian decision theory, and it is therefore not necessary to pay more attention to a generalization of Bayesian estimation theory in a concept of imprecise probabilities.

3.3 Bounds for expected loss

To choose an optimal decision the sets $\mathcal{EL}(d,\Pi)$ for all $d \in D$ must be compared, this requires an additional criterion. An attractive simplification is derived by replacing this problem by one of comparing intervals $[\underline{\mathcal{EL}}(d,\Pi),\overline{\mathcal{EL}}(d,\Pi)]$. This reduces, for each $d \in D$, the calculation of $EL(d,\pi)$ for all $\pi \in \Pi$ to the calculation of only two values, and seems to be reasonable as $\mathcal{EL}(d,\Pi) \subset [\underline{\mathcal{EL}}(d,\Pi),\overline{\mathcal{EL}}(d,\Pi)]$, whereas no smaller closed interval exists that contains $\mathcal{EL}(d,\Pi)$. For the calculation of $\underline{\mathcal{EL}}(d,\Pi)$ and $\overline{\mathcal{EL}}(d,\Pi)$ theorem 3.3.3 is an important result, if Π has the form (2.1), as it shows that $\underline{\mathcal{EL}}(d,\Pi)$ and $\overline{\mathcal{EL}}(d,\Pi)$ are determined by considering only a subclass of Π consisting of probability densities that depend on $L(d,\theta)$, $\ell(\theta)$ and $u(\theta)$, and that can be characterized by one additional parameter belonging to a bounded parameter space if the loss function is bounded. To derive theorem 3.3.3 some additional notation and two lemmas are needed.

For each decision $d \in D$ we introduce

$$L_{l,d} = \inf\{L(d,\theta)\,|\,\theta \in \Theta\} \tag{3.7}$$

and

$$L_{u,d} = \sup\{L(d,\theta)\,|\,\theta \in \Theta\} \tag{3.8}$$

which may be equal to $-\infty$ or $+\infty$, and a partition of the parameter space Θ, $\omega \in [L_{l,d}, L_{u,d}]$

$$\Theta_u(d,\omega) = \{\theta \in \Theta\,|\,L(d,\theta) > \omega\} \tag{3.9}$$

$$\Theta_\omega(d,\omega) = \{\theta \in \Theta\,|\,L(d,\theta) = \omega\} \tag{3.10}$$

$$\Theta_l(d,\omega) = \{\theta \in \Theta\,|\,L(d,\theta) < \omega\}. \tag{3.11}$$

We restrict the discussion to loss functions $L(d,\theta)$ such that

$$\int_{\Theta_\omega(d,\omega)} d\omega = 0 \tag{3.12}$$

for all $\omega \in [L_{l,d}, L_{u,d}]$.

If $\Theta = \mathbb{R}$ this restriction implies that there is no interval in Θ on which $L(d,\theta)$

is constant. At the end of this section we briefly discuss the situation if restriction (3.12) does not hold, in which case analogous results can be derived.

For $d \in D$ we define $\quad (\omega \in [L_{l,d}, L_{u,d}])$

$$C_{u,d}(\omega) = \int_{\Theta_u(d,\omega)} u(\omega)d\omega + \int_{\Theta_l(d,\omega)} \ell(\omega)d\omega \qquad (3.13)$$

and

$$C_{l,d}(\omega) = \int_{\Theta_u(d,\omega)} \ell(\omega)d\omega + \int_{\Theta_l(d,\omega)} u(\omega)d\omega. \qquad (3.14)$$

We define two probability densities in Π that depend through ω on the above partition of Θ

$$\pi^u_{d,\omega}(\theta) = \begin{cases} u(\theta)/C_{u,d}(\omega) & \text{for} \quad \theta \in \Theta_u(d,\omega) \cup \Theta_\omega(d,\omega) \\ \ell(\theta)/C_{u,d}(\omega) & \text{for} \quad \theta \in \Theta_l(d,\omega) \end{cases} \qquad (3.15)$$

and

$$\pi^l_{d,\omega}(\theta) = \begin{cases} \ell(\theta)/C_{l,d}(\omega) & \text{for} \quad \theta \in \Theta_u(d,\omega) \cup \Theta_\omega(d,\omega) \\ u(\theta)/C_{l,d}(\omega) & \text{for} \quad \theta \in \Theta_l(d,\omega) . \end{cases} \qquad (3.16)$$

Next we define subclasses of Π by considering the normalizing constants of original densities between ℓ and u that determine the probability densities in Π. We introduce normalizing constants

$$C_\ell = \int_\Theta \ell(\omega)d\omega \qquad (3.17)$$

$$C_u = \int_\Theta u(\omega)d\omega \qquad (3.18)$$

and

$$C_q = \int_\Theta q(\omega)d\omega. \qquad (3.19)$$

We define a subclass of Π, denoted by $\Pi(C)$, as the class of pdf's obtained form density functions q with $\ell(\theta) \leq q(\theta) \leq u(\theta)$ for $\theta \in \Theta$, which have C as normali-

zing constant $(C_\ell \leq C \leq C_u)$

$$\Pi(C) = \{ \ \pi \in \Pi \ | \ \ell(\theta) \leq q(\theta) \leq u(\theta) \text{ for } \theta \in \Theta, \ \pi(\theta) = q(\theta)/C_q \text{ and } C_q = C \ \}. \quad (3.20)$$

Lemma 3.3.1

$$\Pi = \bigcup_{C \in [C_\ell, C_u]} \Pi(C) = \bigcup_{\omega \in [L_{l,d}, L_{u,d}]} \Pi(C_{u,d}(\omega)) = \bigcup_{\omega \in [L_{l,d}, L_{u,d}]} \Pi(C_{l,d}(\omega)) \quad (3.21)$$

Proof

The first equality is an obvious consequence of (3.20).

The second equality of (3.21) holds because restriction (3.12) and continuity of $\ell(\theta)$ and $u(\theta)$ imply that $C_{u,d}(\omega)$ is a continuous non-increasing function of ω, with $C_{u,d}(L_{l,d}) = C_u$ and $C_{u,d}(L_{u,d}) = C_\ell$.

Analogously, the third equality of (3.21) holds because $C_{l,d}(\omega)$ is a continuous non-decreasing function of ω, $C_{l,d}(L_{l,d}) = C_\ell$ and $C_{l,d}(L_{u,d}) = C_u$.

$\square$

Lemma 3.3.2

Let $d \in D$, and let $\omega \in [L_{l,d}, L_{u,d}]$. Then relation (3.22) holds for all probability densities $\pi \in \Pi(C_{u,d}(\omega))$

$$EL(d,\pi) \leq EL(d, \pi^u_{d,\omega}). \quad (3.22)$$

And for all probability densities $\pi \in \Pi(C_{l,d}(\omega))$ relation (3.23) holds

$$EL(d,\pi) \geq EL(d, \pi^l_{d,\omega}). \quad (3.23)$$

Proof

We prove (3.22). Let $\pi \in \Pi(C_{u,d}(\omega))$ and let $q_\pi(\theta)$ be a corresponding function between ℓ and u, so $\pi(\theta) = q_\pi(\theta)/C_{u,d}(\omega)$ with $\ell(\theta) \leq q_\pi(\theta) \leq u(\theta)$ for all $\theta \in \Theta$. Then

$$EL(d,\pi) = \int\limits_{\Theta} L(d,\omega)\pi(\omega)d\omega = \int\limits_{\Theta} L(d,\omega)q_{\pi}(\omega)d\omega/C_{u,d}(\omega)$$

$$= \int\limits_{\Theta_u(d,\omega)} L(d,\omega)q_{\pi}(\omega)d\omega/C_{u,d}(\omega) + \int\limits_{\Theta_l(d,\omega)} L(d,\omega)q_{\pi}(\omega)d\omega/C_{u,d}(\omega)$$

$$\leq \int\limits_{\Theta_u(d,\omega)} L(d,\omega)u(\omega)d\omega/C_{u,d}(\omega) + \int\limits_{\Theta_l(d,\omega)} L(d,\omega)\ell(\omega)d\omega/C_{u,d}(\omega)$$

$$= \int\limits_{\Theta} L(d,\omega)\pi^{u}_{d,\omega}(\omega)d\omega = EL(d,\pi^{u}_{d,\omega}).$$

To prove the inequality herein we use the fact that

$$C_{u,d}(\omega) = \int\limits_{\Theta_u(d,\omega)} u(\omega)d\omega + \int\limits_{\Theta_l(d,\omega)} \ell(\omega)d\omega = \int\limits_{\Theta_u(d,\omega)} q_{\pi}(\omega)d\omega + \int\limits_{\Theta_l(d,\omega)} q_{\pi}(\omega)d\omega,$$

so

$$\int\limits_{\Theta_u(d,\omega)} [u(\omega)-q_{\pi}(\omega)]d\omega = \int\limits_{\Theta_l(d,\omega)} [q_{\pi}(\omega)-\ell(\omega)]d\omega,$$

which, in combination with (3.9) and (3.11), leads to

$$\int\limits_{\Theta_u(d,\omega)} L(d,\omega)u(\omega)d\omega + \int\limits_{\Theta_l(d,\omega)} L(d,\omega)\ell(\omega)d\omega - \int\limits_{\Theta_u(d,\omega)} L(d,\omega)q_{\pi}(\omega)d\omega \cdot \int\limits_{\Theta_l(d,\omega)} L(d,\omega)q_{\pi}(\omega)d\omega$$

$$= \int\limits_{\Theta_u(d,\omega)} L(d,\omega)[u(\omega)-q_{\pi}(\omega)]d\omega + \int\limits_{\Theta_l(d,\omega)} L(d,\omega)[\ell(\omega)-q_{\pi}(\omega)]d\omega$$

$$\geq \int\limits_{\Theta_u(d,\omega)} \omega[u(\omega)-q_{\pi}(\omega)]d\omega + \int\limits_{\Theta_l(d,\omega)} L(d,\omega)[\ell(\omega)-q_{\pi}(\omega)]d\omega$$

$$= \int\limits_{\Theta_l(d,\omega)} \omega[q_{\pi}(\omega)\cdot\ell(\omega)]d\omega + \int\limits_{\Theta_l(d,\omega)} L(d,\omega)[\ell(\omega)-q_{\pi}(\omega)]d\omega$$

$$= \int\limits_{\Theta_l(d,\omega)} [\omega-L(d,\omega)][q_{\pi}(\omega)-\ell(\omega)]d\omega \geq 0.$$

The proof of (3.23) is analogous.   □

Theorem 3.3.3 shows that for a given decision $d \in D$, one needs to consider only probability densities of the form $\pi^u_{d,\omega}(\theta)$ to determine $\overline{\mathcal{EL}}(d,\Pi)$, and probability densities of the form $\pi^l_{d,\omega}(\theta)$ to determine $\underline{\mathcal{EL}}(d,\Pi)$.

We introduce the notations

$$\Pi^u_d = \{ \pi^u_{d,\omega} \in \Pi \mid \omega \in [L_{l,d}, L_{u,d}] \} \tag{3.24}$$

$$\mathcal{EL}(d,\Pi^u_d) = \{ EL(d,\pi) \mid \pi \in \Pi^u_d \} \tag{3.25}$$

$$\overline{\mathcal{EL}}(d,\Pi^u_d) = \sup \mathcal{EL}(d,\Pi^u_d) \tag{3.26}$$

$$\Pi^l_d = \{ \pi^l_{d,\omega} \in \Pi \mid \omega \in [L_{l,d}, L_{u,d}] \} \tag{3.27}$$

$$\mathcal{EL}(d,\Pi^l_d) = \{ EL(d,\pi) \mid \pi \in \Pi^l_d \} \tag{3.28}$$

$$\underline{\mathcal{EL}}(d,\Pi^l_d) = \inf \mathcal{EL}(d,\Pi^l_d). \tag{3.29}$$

Theorem 3.3.3

Let $d \in D$ and let the loss function $L(d,\theta)$ satisfy restriction (3.12), then

$$\overline{\mathcal{EL}}(d,\Pi) = \overline{\mathcal{EL}}(d,\Pi^u_d) \tag{3.30}$$

and

$$\underline{\mathcal{EL}}(d,\Pi) = \underline{\mathcal{EL}}(d,\Pi^l_d). \tag{3.31}$$

Proof

We prove (3.30).

$$\overline{\mathcal{EL}}(d,\Pi) = \sup \{EL(d,\pi) \mid \pi \in \Pi\} = \sup \{\sup \{EL(d,\pi) \mid \pi \in \Pi(C)\} \mid C \in [C_l, C_u]\}$$

$$= \sup \{\sup \{EL(d,\pi) \mid \pi \in \Pi(C_{u,d}(\omega))\} \mid \omega \in [L_{l,d}, L_{u,d}]\}$$

$$= \sup \{EL(d,\pi^u_{d,\omega}) \mid \omega \in [L_{l,d}, L_{u,d}]\} = \sup \mathcal{EL}(d,\Pi^u_d) = \overline{\mathcal{EL}}(d,\Pi^u_d).$$

The second and third equality are based on lemma 3.3.1, the fourth on lemma 3.3.2. The other equalities in the proof follow from (3.4), (3.5), (3.6), (3.24), (3.25) and (3.26).

The proof of (3.31) is analogous. □

For each $d \in D$, $\overline{\mathcal{EL}}(d,\Pi)$ and $\underline{\mathcal{EL}}(d,\Pi)$ can, as a result of theorem 3.3.3, be determined through the calculation of $\overline{\mathcal{EL}}(d,\Pi_d^u)$ and $\underline{\mathcal{EL}}(d,\Pi_d^l)$. Hereto, we must consider $EL(d,\pi_{d,\omega}^u)$ and $EL(d,\pi_{d,\omega}^l)$ for all $\omega \in [L_{l,d}, L_{u,d}]$. We next show that calculation of $\overline{\mathcal{EL}}(d,\Pi_d^u)$ and $\underline{\mathcal{EL}}(d,\Pi_d^l)$ is simpler than it seems to be, because of the form of $EL(d,\pi_{d,\omega}^u)$ and $EL(d,\pi_{d,\omega}^l)$ as functions of ω.

Theorem 3.3.4

Let $d \in D$ and $\omega \in [L_{l,d}, L_{u,d}]$. For $EL(d,\pi_{d,\omega}^u)$ the following relations hold

If $EL(d,\pi_{d,\omega}^u) \geq \omega$, then for all $\omega \in [L_{l,d}, \omega)$

$$EL(d,\pi_{d,\omega}^u) \leq EL(d,\pi_{d,\omega}^u). \tag{3.32}$$

If $EL(d,\pi_{d,\omega}^u) \leq \omega$, then for all $\omega \in (\omega, L_{u,d}]$

$$EL(d,\pi_{d,\omega}^u) \leq EL(d,\pi_{d,\omega}^u). \tag{3.33}$$

For $EL(d,\pi_{d,\omega}^l)$ analogous relations are

If $EL(d,\pi_{d,\omega}^l) \geq \omega$, then for all $\omega \in [L_{l,d}, \omega)$

$$EL(d,\pi_{d,\omega}^l) \geq EL(d,\pi_{d,\omega}^l). \tag{3.34}$$

If $EL(d,\pi_{d,\omega}^l) \leq \omega$, then for all $\omega \in (\omega, L_{u,d}]$

$$EL(d,\pi_{d,\omega}^l) \geq EL(d,\pi_{d,\omega}^l). \tag{3.35}$$

Proof

We prove (3.32). Let $\omega, \omega \in [L_{l,d}, L_{u,d}]$ with $\omega < \omega$. The fact that the following relations hold

$$\Theta_u(d,\omega) \subseteq \Theta_u(d,\omega)$$

$$\Theta_l(d,\omega) \subseteq \Theta_l(d,\omega)$$

and

-66-

$$\Theta_u(d,\omega)\backslash\Theta_u(d,\omega) = \{\theta\in\Theta\,|\,\omega<L(d,\theta)\leq\omega\} = [\Theta_l(d,\omega)\cup\Theta_\omega(d,\omega)]\backslash\Theta_l(d,\omega).$$

leads, in combination with restriction (3.12), to

$$EL(d,\pi^u_{d,\omega}) \leq EL(d,\pi^u_{d,\omega}) \quad\Leftrightarrow$$

$$\frac{\displaystyle\int_{\Theta_u(d,\omega)} L(d,\theta)\,u(\theta)d\theta + \int_{\Theta_l(d,\omega)} L(d,\theta)\,\ell(\theta)d\theta}{\displaystyle\int_{\Theta_u(d,\omega)} u(\theta)d\theta + \int_{\Theta_l(d,\omega)} \ell(\theta)d\theta} \leq EL(d,\pi^u_{d,\omega}) \quad\Leftrightarrow$$

$$\frac{\displaystyle\int_{\Theta_u(d,\omega)} L(d,\theta)u(\theta)d\theta + \int_{\Theta_l(d,\omega)} L(d,\theta)\ell(\theta)d\theta + \int_{\Theta_u(d,\omega)\backslash\Theta_u(d,\omega)} L(d,\theta)[u(\theta)-\ell(\theta)]d\theta}{\displaystyle\int_{\Theta_u(d,\omega)} u(\theta)d\theta + \int_{\Theta_l(d,\omega)} \ell(\theta)d\theta + \int_{\Theta_u(d,\omega)\backslash\Theta_u(d,\omega)} [u(\theta)-\ell(\theta)]d\theta}$$

$$\leq \frac{\displaystyle\int_{\Theta_u(d,\omega)} L(d,\theta)u(\theta)d\theta + \int_{\Theta_l(d,\omega)} L(d,\theta)\ell(\theta)d\theta}{\displaystyle\int_{\Theta_u(d,\omega)} u(\theta)d\theta + \int_{\Theta_l(d,\omega)} \ell(\theta)d\theta} \quad\Leftrightarrow$$

$$\frac{\displaystyle\int_{\Theta_u(d,\omega)\backslash\Theta_u(d,\omega)} L(d,\theta)[u(\theta)-\ell(\theta)]d\theta}{\displaystyle\int_{\Theta_u(d,\omega)\backslash\Theta_u(d,\omega)} [u(\theta)-\ell(\theta)]d\theta} \leq EL(d,\pi^u_{d,\omega}). \tag{3.36}$$

From the fact that $L(d,\theta) \leq \omega$ for all $\theta \in \Theta_u(d,\omega)\backslash\Theta_u(d,\omega)$, we conclude that (3.36) holds if $EL(d,\pi^u_{d,\omega}) \geq \omega$.

The proofs of (3.33), (3.34) and (3.35) are analogous. □

Theorem 3.3.4 is a useful result for the optimization process necessary to calculate $\overline{\mathcal{E}\mathcal{L}}(d,\Pi)$ and $\underline{\mathcal{E}\mathcal{L}}(d,\Pi)$, as it provides a way of deleting a part of Θ while searching for these optima (after each calculation) and a sufficient condition for a calculated value to be an optimum.

Corollary 3.3.5 is a useful result of theorem 3.3.4.

Corollary 3.3.5
Let $d \in D$ and let $L(d,\theta)$ be continuous, bounded (so $L_{l,d} > -\infty$ and $L_{u,d} < \infty$) and such that restriction (3.12) holds (with, as before, ℓ and u continuous).
Then there exist $\omega_0, \omega_1 \in [L_{l,d}, L_{u,d}]$ such that

$$\overline{\mathcal{E}\mathcal{L}}(d,\Pi) = EL(d, \pi^u_{d,\omega_0}) = \omega_0 \tag{3.37}$$

and

$$\underline{\mathcal{E}\mathcal{L}}(d,\Pi) = EL(d, \pi^l_{d,\omega_1}) = \omega_1. \tag{3.38}$$

Proof
We prove (3.37). Let $\omega_0 \in [L_{l,d}, L_{u,d}]$, then, as (3.12) holds, theorem 3.3.4 states that a sufficient condition for $\overline{\mathcal{E}\mathcal{L}}(d,\Pi) = \omega_0$ is $EL(d, \pi^u_{d,\omega_0}) = \omega_0$.

For $\omega \in [L_{l,d}, L_{u,d}]$ we derive

$$EL(d, \pi^u_{d,\omega}) - \omega = 0 \quad \Leftrightarrow$$

$$k_1(\omega) := \int_{\Theta_u(d,\omega)} [L(d,\theta) - \omega] u(\theta) d\theta = \int_{\Theta_l(d,\omega)} [\omega - L(d,\theta)] \ell(\theta) d\theta =: k_2(\omega).$$

Here $k_1(\omega)$ is a continuous non-increasing function of ω with

$$k_1(L_{l,d}) = \int_{\Theta} [L(d,\theta) - L_{l,d}] u(\theta) d\theta \geq 0 \quad \text{and} \quad k_1(L_{u,d}) = 0,$$

and $k_2(\omega)$ is a continuous non-decreasing function of ω with

$$k_2(L_{l,d}) = 0 \quad \text{and} \quad k_2(L_{u,d}) = \int_{\Theta} [L_{u,d} - L(d,\theta)] \ell(\theta) d\theta \geq 0.$$

This implies that there is at least one $\omega_0 \in [L_{l,d}, L_{u,d}]$ for which $k_1(\omega_0) = k_2(\omega_0)$, which proves (3.37).

The proof of (3.38) is analogous. □

No analytic forms for ω_0 and ω_1 of corollary 3.3.5 have been found, but good numerical approximations can be derived quite easily as a result of theorem 3.3.4.

We give a simple example of the above theory.

Example 3.3.6

For a given decision $d \in D$, let the loss function be equal to

$$L(d,\theta) = \theta(1-\theta), \quad \text{with } \theta \in [0,1].$$

We want to determine, for this decision d, $\underline{\mathscr{E}\mathscr{L}}(d,\Pi)$ and $\overline{\mathscr{E}\mathscr{L}}(d,\Pi)$, with Π of the form (2.1), with for all $\theta \in [0,1]$

$$\ell(\theta) = 1$$

and

$$u(\theta) = 2.$$

As $L(d,\theta)$ is not constant on an interval of positive length, this loss function satisfies restriction (3.12) and theorem 3.3.3 holds. Therefore, we can restrict the discussion to probability densities of the form (3.15) to determine $\overline{\mathscr{E}\mathscr{L}}(d,\Pi)$ and to probability densities of the form (3.16) to determine $\underline{\mathscr{E}\mathscr{L}}(d,\Pi)$.

For this example, we have $L_{l,d} = 0$, $L_{u,d} = 1/4$ and, for $\omega \in [0,1/4]$

$$\Theta_u(d,\omega) = ((1-\sqrt{1-4\omega})/2,(1+\sqrt{1-4\omega})/2)$$

$$\Theta_l(d,\omega) = [0,(1-\sqrt{1-4\omega}]/2))\cup((1+\sqrt{1-4\omega})/2,1]$$

and

$$\Theta_\omega(d,\omega) = \{(1-\sqrt{1-4\omega})/2,(1+\sqrt{1-4\omega})/2\},$$

so it is clear that indeed restriction (3.12) holds. Further

$$C_{u,d}(\omega) = 1+\sqrt{1-4\omega} \quad \text{and} \quad C_{l,d}(\omega) = 2-\sqrt{1-4\omega}.$$

Using theorem 3.3.3 we determine $\overline{\mathscr{E}\mathscr{L}}(d,\Pi)$ by maximization of

$$EL(d,\pi^u_{d,\omega}) = \frac{1 + (1+2\omega)\sqrt{1-4\omega}}{6(1+\sqrt{1-4\omega})}$$

leading to $\overline{\mathcal{EL}}(d,\Pi) = 0.1875$ (for $\omega=0.1875$).

Analogously, we determine $\underline{\mathcal{EL}}(d,\Pi)$ by minimization of

$$EL(d,\pi^l_{d,\omega}) = \frac{2 - (1+2\omega)\sqrt{1-4\omega}}{6(2-\sqrt{1-4\omega})}$$

leading to $\underline{\mathcal{EL}}(d,\Pi) = 0.1435$ (for $\omega=0.1435$).

The fact that the optima are equal to the values of ω for which these optima are adopted is in agreement with theorem 3.3.4.

<div align="right">□</div>

The analysis becomes more complicated if restriction (3.12) does not hold, in which case lemma 3.3.1 does not hold anymore. For this analysis we refer to a technical report (Coolen, 1992) where it is shown that the bounds for expected loss are derived by probability densities that correspond to (3.15) and (3.16), which are proportional to either $l(\theta)$ or $u(\theta)$ for all $\theta \in \Theta_\omega(d,\omega)$.

Finally, remark that $L(d,\theta)=0$ leads to bounds for the mean of the parameter θ, and so the above theory can also be used to calculate these bounds, as well as for the calculation of bounds for higher order moments.

In the example presented in section 3.4 the results from this section are used for the calculation of the bounds on expected loss.

3.4 An age-replacement example

In this section application of the Bayesian decision theory with imprecise prior probabilities to a simple replacement problem is presented, showing output in the form of lower and upper bounds for expected loss.

The problem considered is that of finding an optimal age-replacement rule (e.g. Tijms, 1986; section 1.2), which prescribes the replacement of a unit (system, component) upon failure or upon reaching the age $\mathcal{T}$, whichever occurs first, where $\mathcal{T}$ is a control parameter. To show the use of Bayesian decision theory with imprecision in this section an example is presented in which $\mathcal{T}$ is discrete, $\mathcal{T} \in \{3,6,9,..\}$ months. The random variable $X>0$, the lifetime of the technical unit of interest, is assumed to have a probability distribution that is known up to a scale parameter $\theta \geq 0$ (this restriction on the parameter space is not essential to the theory but for ease of exposition), with cdf $F_X(x|\theta)$. A loss function is the expected cost per time-unit over an infinite time horizon for decision $\mathcal{T}$

$$L(\mathcal{T}|\theta) = \frac{1 + (c-1)F_X(\mathcal{T}|\theta)}{\int_0^{\mathcal{T}} (1 - F_X(y|\theta))\,dy}. \tag{3.39}$$

The cost of preventive replacement is assumed without loss of generality to be 1 and $c>1$ is the relative cost of corrective replacement. The assumption that c is precisely known may not be realistic, but generalization of the theory to $c \in [c_\ell, c_u]$ is quite easy because the comparison of the decisions $\mathcal{T}$ is based on the bounds of the expected value of $L(\mathcal{T}|\theta)$ over the possible distributions for θ, and in case of $c_\ell \neq c_u$ the lower bound is accepted for costs c_ℓ and the upper bound for costs c_u. Under the above assumptions the remaining problem is that θ is unknown, and we assume that the only information available is expert opinion about the lifetime X.

Since the parameter θ is unobservable, opinions of the experts should be elicited by asking questions about X (this is discussed in chapter 4), and we assume, to start the example, the results of such an elicitation process. In chapter 4 we discuss methods for combining these opinions and for translating the information about X into sets of prior distributions for θ. It is important for the application of this concept of decision making that elicitation is studied in real-life cases by groups of researchers from several disciplines. So far we know of only one case-study of a decision problem using imprecision (Walley and Campello de Souza, 1990) where the concept is used

more as sensitivity analysis than to relate imprecision to the amount of information available. This last interpretation of imprecision is interesting if new information becomes available, which was not discussed in that case-study.

Example 3.4.1

For this example gamma distributions are used, as they are mathematically attractive and describe the randomness of lifetimes of deteriorating units reasonably well. We further assume that the shape parameter of the gamma distribution of X, the lifetime of the technical unit, is equal to 3, leaving a one-dimensional non-negative scale parameter θ in the model. The cumulative distribution function of the model is

$$F_X(x|\theta) = \Gamma_{\theta x}(3)/2, \quad x \geq 0 \text{ and } \theta > 0$$

with incomplete gamma function

$$\Gamma_w(z) = \int_0^w y^{z-1}e^{-y}dy, \quad w \geq 0 \text{ and } z > 0$$

and the probability density function is

$$f_X(x|\theta) = \theta^3 x^2 \exp(-\theta x)/2.$$

Completing the model in the context of Bayesian theory with imprecise prior probabilities defined through a class of the form (2.1), the assumed lower and upper prior densities are

$$u(\theta) = \ell(\theta) + c_0 a(\theta)$$

with $c_0 \geq 0$,

$$\ell(\theta) = \tau_\ell^{10} \theta^9 \exp(-\tau_\ell \theta)/\Gamma(10)$$

and

$$a(\theta) = \tau_a^{10} \theta^9 \exp(-\tau_a \theta)/\Gamma(10).$$

-72-

These densities ℓ and a are pdf's of gamma distributions that are conjugate to the gamma family with random scale parameter, leaving hyperparameters τ_ℓ, τ_a and c_0 in the model to be chosen to make it fit well with expert opinions. To this end we compare lower and upper cumulative distribution functions resulting from elicitation with the ones resulting from the model. The imprecise cdf's for X have the forms (2.4) and (2.5), with ℓ and u the predictive densities based on the priors, denoted by $\ell_{X,0}$ and $u_{X,0}$.

It is also possible to drop the assumption that the shape parameters of these gamma priors are known, but this leads to more hyperparameters and more calculations when fitting the model to subjective data. In practical applications it would be sensible to perform sensitivity analyses with regard to these assumptions.

For this example, we assume that there is one DM who wants to know the opinions of three experts. Walley and Campello de Souza (1990) suggest using imprecise cdf's for X in the elicitation process, and we continue with the example assuming that we obtained the following results ($\underline{F}_X(0)=\overline{F}_X(0)=0$ and $\underline{F}_X(\infty)=\overline{F}_X(\infty)=1$)

x:	6	12	18	24
Expert A: $\underline{F}_X^A(x)$:	.04	.22	.46	.66
$\overline{F}_X^A(x)$:	.15	.42	.68	.83
Expert B: $\underline{F}_X^B(x)$:	.05	.20	.41	.60
$\overline{F}_X^B(x)$:	.17	.42	.63	.78
Expert C: $\underline{F}_X^C(x)$:	.11	.32	.55	.80
$\overline{F}_X^C(x)$:	.47	.76	.89	.96

Experts A and B have similar ideas about the lifetime of the unit, whereas expert C is much more pessimistic and also less sure, which can be seen from the imprecision in the above numbers. If these numbers are interpreted as betting behavior, both experts B and C would be pleased by a bet on the event $T \leq 24$ for price 0.79 (where B sells the bet to C).

To fit the model to subjective data of this kind (per expert), by determination of suitable values for the hyperparameters, c_0 is set equal to

$2\Delta_{max}/(1-\Delta_{max})$, with Δ_{max} the maximum of the imprecision for an event $T \le t$ according to the subjective data (this is an *ad hoc* methodology based on (2.15) that reduces the necessary amount of calculations to fit the model, see section 4.4), and thereafter values for τ_ℓ and τ_a are determined such that the imprecise cdf's fit well to the expert's cdf's in the points where these are given. Here the distance between the lower cdf's (and for the upper cdf's) is defined as the expected squared distance of the discretized cdf's (over the intervals used in elicitation), where the expectation is with regard to the subjective lower distribution (see section 4.4 for details). Then τ_ℓ and τ_a are determined by minimization of the sum of these expected squared distances of the upper and of the lower cdf's (we could have determined an optimal value for c_0 as well by this method, but that would lead to much more calculations). The hyperparameters for the models that fit to the above expert's cdf's are

	c_0	τ_ℓ	τ_a
Expert A:	0.56	57.6	52.2
Expert B:	0.59	66.9	43.2
Expert C:	1.57	48.7	27.2

To calculate the bounds on the expected loss according to the above theory, the cost of corrective replacement is set to $c=10$. The results are (for decisions $\mathcal{T}$; $\mathcal{T}=300$ months is effectively no preventive replacement)

Expert	$\mathcal{T}$:	3	6	9	12	15	18	21	24	27	30	36	300
A:	$\underline{\mathcal{EL}}(\mathcal{T},\Pi)$:	.34	.18	.13	.11	.10	.10	.10	.10	.10	.10	.11	.16
	$\overline{\mathcal{EL}}(\mathcal{T},\Pi)$:	.34	.18	.14	.12	.12	.12	.13	.13	.13	.14	.15	.20
B:	$\underline{\mathcal{EL}}(\mathcal{T},\Pi)$:	.34	.17	.13	.10	.09	.09	.09	.09	.09	.09	.10	.14
	$\overline{\mathcal{EL}}(\mathcal{T},\Pi)$:	.34	.18	.14	.12	.12	.12	.12	.13	.13	.14	.14	.19
C:	$\underline{\mathcal{EL}}(\mathcal{T},\Pi)$:	.34	.18	.14	.12	.11	.11	.12	.12	.13	.13	.14	.19
	$\overline{\mathcal{EL}}(\mathcal{T},\Pi)$:	.36	.23	.22	.23	.24	.26	.27	.28	.29	.30	.31	.35

From such a scheme it is obvious that an additional criterion is needed to reach a final decision. There are some good arguments in this example in favor of decision $\mathcal{T}=15$, but the final choice is to be made by the DM.

Another approach is to combine the information from the experts first (see section 4.3 for more details). One possibility is to define the combined

lower probability as the minimum of the lower probabilities per expert, and the combined upper probability as the maximum of the upper probabilities per expert. This lower probability can be interpreted as the supremum of the prices for which all the experts are willing to buy the bet. A second possibility of combination defines the new lower probability as the maximum of the lower probabilities per expert, and the new upper probability as the minimum of the upper probabilities per expert, with interpretation that the DM, when adopting these combined imprecise probabilities, wants to buy or sell a bet if at least one expert wants this. Note that this second method of combination can lead to incoherent betting behavior of the DM since there may be a price at which one expert would be willing to buy a bet while another expert wants to sell the same bet (in our example this would be caused by the disagreement between experts B and C on the event $T{\leq}24$). The fact that this method actually indicates such disagreement between experts is useful in practice, and is not provided by the classical Bayesian theory (where experts always disagree, except when they assess exactly the same precise probabilities).

An important consideration for decision theory is the possibility of incorporating other information. The updating methodology proposed in chapter 2 is now illustrated, assuming highly imprecise prior information of an expert D. In case of additional data consisting of n observed independent failures of the technical unit, with failure times x_i $(i=1,..,n)$ and total time on test

$$u = \sum_{i=1}^{n} x_i$$

the priors are updated by replacement of the hyperparameters of the densities ℓ and α according to the classical Bayesian theory (here the choice of conjugate densities leads to simple calculations) together with replacing c_0 by

$$c_n = \frac{c_0}{1+n/\xi}$$

according to (2.36). The additional parameter ξ is to be chosen by the DM, and can be interpreted as the amount of additional data that provides as much information as the prior (subjective) information does.
The assumed prior information is

x:	6	12	18	24
Expert D: $\underline{F}_X^D(x)$:	.03	.17	.34	.54
$\overline{F}_X^D(x)$:	.51	.76	.89	.96

The hyperparameters to fit the model to these data are $c_0=2.92$, $\tau_\ell=69.8$, $\tau_a=31.5$, and the value of new information compared to the prior information is indicated by $\xi=5$.

The decisions after updating are analyzed for two possible cases

(I) $n=10$, $tt=150$

and

(II) $n=20$, $tt=180$.

The bounds of expected loss for the prior and the two posterior situations are

Expert 𝒥:	3	6	9	12	15	18	21	24	27	30	36	300
D: $\underline{\mathcal{EL}}(\mathcal{J},\Pi)$:	.34	.17	.12	.10	.09	.08	.08	.08	.08	.08	.09	.13
$\overline{\mathcal{EL}}(\mathcal{J},\Pi)$:	.36	.23	.22	.22	.23	.25	.26	.27	.28	.29	.30	.33
(I) $\underline{\mathcal{EL}}(\mathcal{J},\Pi)$:	.34	.18	.13	.11	.11	.11	.11	.11	.12	.12	.13	.19
$\overline{\mathcal{EL}}(\mathcal{J},\Pi)$:	.34	.18	.14	.13	.13	.13	.14	.14	.15	.16	.17	.23
(II) $\underline{\mathcal{EL}}(\mathcal{J},\Pi)$:	.34	.20	.17	.16	.17	.19	.20	.21	.22	.23	.25	.30
$\overline{\mathcal{EL}}(\mathcal{J},\Pi)$:	.35	.21	.18	.19	.20	.21	.23	.24	.26	.27	.29	.33

If, for example, the DM wants to use the minimax criterion, then $\mathcal{J}=9$ or $\mathcal{J}=12$ are good decisions in the prior situation, where $\mathcal{J}=12$ seems to be more logical when also regarding the lower bound for the expected loss. For the first posterior situation the decisions $\mathcal{J}=12$, $\mathcal{J}=15$ or $\mathcal{J}=18$ seem to be reasonable (no differences in both the upper and lower bounds, up to two decimals), and for the second posterior $\mathcal{J}=9$ would be a good decision. For this last situation notice that $\underline{\mathcal{EL}}(12,\Pi) < \underline{\mathcal{EL}}(9,\Pi)$ and $\overline{\mathcal{EL}}(12,\Pi) > \overline{\mathcal{EL}}(9,\Pi)$, so based on the information available here the expected loss according to decision $\mathcal{J}=9$ is believed to be between 0.17 and 0.18, while according to decision $\mathcal{J}=12$ these bounds are 0.16 and 0.19, implying that one is less sure about the consequences of decision $\mathcal{J}=12$ than about the consequences of decision $\mathcal{J}=9$.

□

3.5 Additional remarks

The value of any theory for decision making based on subjective information can only be analyzed by application to real world problems, which will in turn bring to light new aspects not encountered in the theoretical approach. In chapter 4 more attention is paid to implementation, but with the aim of developing the concepts, thus without any real application so far.

This also limits the discussion of future research, although already some questions come to mind. For example, how does the form of the loss function affect the comparison of decisions only in terms of the bounds on expected loss? If it is necessary to distinguish between possible decisions by comparing the entire sets of the form (3.4), this may lead to problems in calculation. Another interesting area of research is calculation of the bounds of expected loss according to section 3.3, especially when the parameter space is multi-dimensional with restrictions. It may be possible to address the problems with tools such as Gibbs sampling (Gelfand and Smith, 1990), that have proved to be useful in the standard Bayesian framework.

Also of interest are problems arising from the possibility that the loss function is not precisely known. Should we use imprecise loss functions, which might perhaps lead to very wide bounds of expected loss without the possibility to distinguish whether this results from the imprecise loss function or the imprecision that is modeled through the priors, or should we use precise loss functions and perform sensitivity analyses, for example with regard to cost parameters?

The main challenge is application of the concept to real world problems, and comparing not only results but also the acceptability of the concept for decision makers and experts.

Chapter 4

IMPLEMENTATION

4.1 Introduction

For the practical application of a theory of decision making based on expert opinions, elicitation and combination of opinions is of the utmost importance. One problem for research in this area is that the theories can only be tested in actual applications and this requires a multidisciplinary approach since psychological and communicational aspects are as important as mathematics. We have not performed such a test, but as a first step towards implementation of a decision theory using imprecise probabilities we discuss elicitation and combination of opinions, emphasizing arguments that intuitively favor this concept when compared to theories using only precise probabilities. We believe that a decision theory using imprecise probabilities has more appeal to decision makers and consulted experts, and can therefore play an important role in convincing managers of the power of quantitative methods as tools to assist them in making decisions under uncertainty.

Elicitation of expert opinions has to be preceded by a mathematical analysis of the situation, leading to a model based on clear assumptions to give insight into the aspects on which more information is needed. Assumptions made in this stage may also be based on subjective ideas, and insight can be gained, for example, by performing sensitivity analysis (e.g. Coolen and Dekker, 1994). These aspects are not considered any further in this thesis, but it is important to emphasize that one should always pay attention to model building and the corresponding assumptions before starting an elicitation process. In decision problems almost all information has a price, which certainly holds for expert opinions ('time is money'), so first the analyst should learn which information is needed in the decision problem at hand, and also be able to explain to other people (decision makers, experts) why this information is needed.

In section 4.2 elicitation of expert opinions is discussed, however without conclusions for elicitation within the concept of imprecise probabilities since no case-studies have been performed. In section 4.3 combination of imprecise probabilities is discussed, in which case the concept offers more ele-

gant solutions than when we are restricted to precise probabilities. It is emphasized that combination can have two different goals, firstly summarization of the opinions of a group of experts and secondly good decision making. It seems unreasonable to expect that one combination method is suitable for both goals. In section 4.4 the resulting problem of model fitting, or choosing values for hyperparameters such that the model fits well to the expert opinions, is discussed briefly.

Section 4.5 provides a brief introduction of imprecise probabilities in reliability theory, an area of application in which the use of expert opinion is often unavoidable, and finally in section 4.6 some concluding remarks are made.

4.2 Elicitation of expert opinions

Elicitation is the process of quantification of expert opinions, resulting in imprecise probabilities. The success of a quantitative concept for decision making based on expert opinions depends heavily on the possibility of eliciting the information by a method that is attractive to the experts consulted, is not mathematically demanding for them, asks them to express their opinions about phenomena that are familiar to them, and is simple to learn. The obvious requirement that the results of the elicitation process can indeed be used as information in the mathematical model, is discussed in section 4.4. In this section elicitation is discussed, however without strong general conclusions since such can only be derived after real-world applications, which takes much effort by researchers from several disciplines.

In the literature on Bayesian statistics elicitation methods frequently assume experts are able to state opinions about a parameter in a mathematical model. We restrict the elicitation to questions about observable random variables, thus giving predictive distributions a key role in fitting the model (Geisser, 1993; see section 4.4 for more details). A histogram method for eliciting expert opinions about lifetimes of technical equipment is discussed. Although every practical decision problem asks for a detailed analysis of the elicitation methods to be used, some general topics for future research are mentioned.

In the literature on imprecise probabilities elicitation is discussed by Walley (1982 and 1991; chapter 4), and by Walley and Campello de Souza (1990) in their case-study of the economic viability of solar energy systems. Much more attention has been paid to the subject under the restriction to precise probabilities, for example by Cooke (1991), Van Lenthe (1993) and Vreugdenhil (1993), who also present useful reviews of the literature.
Attention is often restricted to the elicitation of proportions, so in the case of Bernoulli experiments this implies elicitation of the model parameter of interest (Van Lenthe, 1993). We are also concerned with elicitation of opinions about a random variable $X \in \mathbb{R}$, for example lifetimes of technical equipment.
The restriction to precise probabilities is the cause of many problems, and in the literature many *ad hoc* solutions have been proposed to solve these problems. This does not take away the fact that literature on elicitation of precise probabilities teaches us some interesting things that need to be kept

in mind when an elicitation method for imprecise probabilities is required.

Vreugdenhil (1993) discusses the calibration of subjective probabilities. A subject, who may or may not be an expert in the field of interest, is asked to answer a question by choosing between two alternatives, and thereafter is asked to state his subjective probability p that his answer is correct. When p is used a number of times, one expects that the proportion of correct answers (also called predictions if verification is not yet possible when the question is asked) is nearly equal to p. If this is the case, the subject is said to be well calibrated. Usually in calibration research the result is overconfidence, meaning that people tend to overestimate their knowledge, especially when the questions are about intellectual knowledge. People tend to overestimate their intellectual capacities. When the questions are about future events there seems to be a difference in calibration to general knowledge questions. In the literature underconfidence is found in predicting future events (Vreugdenhil, 1993; section 1.4.4). This feature is especially important in areas like risk analysis.

Vreugdenhil also discusses calibration of experts, deriving the important conclusion that, although calibration research in which the subjects are experts in the area of interest does involve a lot of practical problems, it can be very informative from the experts point of view. It is stated as an important problem that calibration research with real applications is to be preferred, but many observations are necessary to obtain sensible calibration measures. Since calibration is an indication of the correctness of subjective probabilities when interpreted as relative frequencies of observations, it is said to measure the external validity of subjective probabilities. Although calibration studies in general teach us interesting facts about human opinions, for many decision problems in which expert opinions are to be used calibration is impossible precisely because of a lack of other information, moreover, in future a few observations may become available. In elicitation of expert opinions however, the results of Vreugdenhil make us aware of features like over and underconfidence and the possibility of confronting experts with calibration results for their probabilities about some events, which might lead to subjective probabilities with more external validity for related events. This creates the possibility of training before real elicitation is started.

Lourens (1984) and Terlouw (1989) have contributed to improvement of external validity of subjective probability density functions by developing an interactive computer package for elicitation, in which the subjects have to estimate proportions using different methods, with the idea that the results

should converge. After some practice, subjects were able to give stable and consistent density functions, but the functions were too peaked which means that subjects still were overconfident.

Another aspect of subjective probabilities is internal validity, whether or not they accurately reflect the uncertain knowledge of a person. It seems impossible to measure this internal validity, and the question remains what the real thoughts are and how to elicit them, one may try to create an elicitation technique that stimulates experts to give internally valid subjective probabilities.

Van Lenthe (1993) has developed an elicitation tool, the computer package ELI, that stimulates subjects to give internally valid probability distributions for proportions by using proper scoring rules. The results of application studies are promising, and show that user-friendly computer packages are helpful elicitation tools and that an idea of scoring can be attractive in communication with the subject, possibly because a terminology of scores is more in accordance with every-day life than the use of probability distributions. Score functions offer *ad hoc* solutions to elicitation problems but are restricted in their use by the lack of a sensible interpretation. When score functions are used similar problems to those encountered in fuzzy set theory arise. The concept of precise probabilities seems to be too tight and little or no attention is paid to interpretation of subjective probabilities, therefore the solution proposed is adding some mathematical concept (scoring rules or, in fuzzy sets theory, membership functions), to the probability density function without a clear interpretation. The proper scoring rules, as used by Van Lenthe, are quite arbitrarily chosen functions of probability densities, and to assess values for such functions or to learn from values related to possible observations of the random variable of interest, not only the interpretation of probability densities must be clear, but also the mathematical function that generates the scores must be understood. This last prerequisite is both as important as and as impossible as the sensible use of a membership function in fuzzy sets theory. The fact that Van Lenthe obtains results that seem to be satisfying should not lead to the conclusion that human opinions should be represented by scoring rules.

What can be learned from the application of scoring rules in the elicitation process is that people may be more motivated to try to reach a high score, so that it brings more or less some competitive element into the elicitation process. This feature, here derived by artificial means, is inherent to elicita-

tion of imprecise probabilities for events with an interpretation in terms of betting behavior, where scores should logically be based on winning bets. It would be of great interest to study the possible use of scoring ideas directly related to betting behavior in the elicitation of imprecise probabilities, for which means an easy to handle computer package like ELI should be developed. For training sessions the scoring ideas, based on betting behavior and known results, would be useful, since nobody can be expected to be able to assess imprecise probabilities without some training, and the studies of Van Lenthe indicate that people can learn to quantify their opinions better by training, with feedback in some form of scores, where probably the logical rewards in betting can be understood quite well.

When discussing the literature on elicitation, Van Lenthe (1993) remarks that reviews of evaluation and comparison studies reveal that different techniques elicit different distributions, this has been used explicitly by Lourens (1984) and Terlouw (1989). These differences are an additional argument for accepting imprecision of probability statements, since the theory of imprecise probabilities allows combination of several assessments by a single person, for example by the unanimity rule for combination as discussed in section 4.3 (see also Walley, 1991; chapter 4).

Cooke (1991; chapter 8) discusses elicitation and scoring of precise subjective probabilities, and concludes with the following list of practical guidelines that are also useful for the development of elicitation techniques for imprecise probabilities.

- The questions must be clear.
- Prepare an attractive format for the questions and graphic format for the answers.
- Perform a dry run.
- An analyst must be present during the elicitation.
- Prepare a brief explanation of the elicitation format, and of the model for processing the responses.
- Avoid coaching.
- An elicitation session should not exceed 1 hour.

As further recommendations for the entire process, Cooke mentions bringing the experts together, defining and decomposing the issues, explaining the methods

and training in probability assessment (at least 1 day), followed by the actual elicitation per expert, and finally resolving disagreements, where we think that one need not try to get all experts to agree, but this may be used to gain insight in the causes of serious disagreements. Although in our concept agreement is not necessary (see section 4.3 for details on combination), if one believes that the different experts disagree seriously this could be caused by different information which may be discussed before the actual elicitation takes place, when all experts are gathered. For the training sessions we believe that the conclusions of Vreugdenhil (1993) and Van Lenthe (1993) are valuable, so calibration measures and feedback by some form of scoring should be used in the training stage.

Cooke's first guideline, that the questions must be clear, seems to be obvious. Nevertheless, in the literature on subjective probabilities, especially on the Bayesian concept of statistics and decision making, it is often assumed that subjects are able to assess quantities that are hard to interpret. Such disadvantages of existing methods also hold within our concept, and we emphasize the importance of eliciting only opinions about observable random events, and not about purely mathematical devices such as parameters in an assumed model, except for some special situations in which the interpretation of the parameter is straightforward.

In the Bayesian framework a parametric model is assumed for a random variable X of interest, for example a probability density function for X that is completely specified up to parameter θ, $\ell(x|\theta)$. A second part of the assumed model is a parametric model for the parameter (this is also regarded as a random variable) with pdf $q(\theta|\tau)$, τ is called a hyperparameter. The choice of q is often based on mathematical simplicity. In most textbooks on Bayesian statistics (e.g. Lee, 1989; Press, 1989) it is simply assumed that a consulted expert can state his opinions about the model parameter θ, for example by giving some properties of its probability distribution, after which a value for τ is chosen such that $q(\theta|\tau)$ fits well to the information provided. We believe that a major problem is the fact that θ is often not observable, which implies that an expert in the field of interest, who is rarely an expert in statistics and mathematical modeling too, will not reach opinions about θ naturally, and is forced to say something about an unknown phenomenon that may not have a clear interpretation for him. In many situations the random variable of interest, X, is observable or has a clear definition and interpretation, for example in many problems about maintenance planning X is the lifetime of technical equipment, that can be defined correctly if some effort is

paid to it. This is also necessary for the discussion with the expert, who needs to know exactly which phenomenon his opinion is needed about.

With restriction to real random variables, let us say non-negative lifetimes of technical equipment, we propose an elicitation methodology that seems to be simple, and whose results are suitable for combination and model fitting, as discussed in sections 4.3 and 4.4. Besides these restrictions to observables in the elicitation process, it might be useful to ask experts in relation to technical equipment of interest to state their opinions about the event that the lifetime will be within an interval. To do this the time-axis can be partitioned into intervals of a length that is familiar and logical to the experts. Here one could think of weeks or months, or of some fixed length of the period between two production stops, for example for inspections, that has been used in the past. A simple mathematical formulation can be given, and for practical application it seems to be attractive to use a graphical form with both numerical and histogram representation of the imprecise probabilities.

Let $I_1,...,I_m$ be a partition of the time-axis, $I_i=[x_{i-1},x_i)$ with $0=x_0<x_1<...<x_m=\infty$. Suppose that the random variable of interest, X, is the time to failure of a technical system when functioning under specified conditions. An expert with knowledge of the system will have ideas about an event $X\in I_i$ (especially if the x_i are familiar to him), and a bet on this event can easily be defined and understood, so with some effort the expert's supremum buying price and infimum selling price can be elicited. Let these lower and upper probabilities be denoted by $\underline{p}_i=\underline{P}(X\in I_i)$ and $\overline{p}_i=\overline{P}(X\in I_i)$, $i=1,..,m$. These imprecise probabilities are coherent if and only if the following conditions hold for all i (Walley, 1991; section 4.6.1)

(COH 1) $\quad 0 \leq \underline{p}_i \leq \overline{p}_i \leq 1$ \hfill (4.1)

and

(COH 2) $\quad \overline{p}_i+\sum_{j\neq i}\underline{p}_j \leq 1 \leq \underline{p}_i+\sum_{j\neq i}\overline{p}_j.$ \hfill (4.2)

A third condition stated by Walley, $\sum_{j=1}^{m}\underline{p}_j \leq 1 \leq \sum_{j=1}^{m}\overline{p}_j$, is implied by COH 1 and COH 2.

In the elicitation process the validity of these conditions must be checked. If COH 1 is violated, this is obviously a serious kind of incoherence since the expert would be willing to bet against himself, and probably has not un-

derstood clearly what is asked, or has made some mistake in writing his answers. Incoherence by violation of the second condition can be corrected by replacing $\bar{p}_i$ by $1-\sum_{j\neq i}\underline{p}_j$ if the first inequality does not hold, or $\underline{p}_i$ by $1-\sum_{j\neq i}\bar{p}_j$ if the second inequality does not hold, where, of course, condition COH 1 must still be satisfied. Nevertheless, this violation should also be presented to the expert, with the suggestion of these replacements as explained by example 4.2.1.

Example 4.2.1

Suppose elicitation has led to an expert's imprecise probabilities (with a partition of the time-axis into three intervals)

$$\underline{p}_1=0.1 \qquad \bar{p}_1=0.3$$
$$\underline{p}_2=0.3 \qquad \bar{p}_2=0.6$$
$$\underline{p}_3=0.2 \qquad \bar{p}_3=0.3.$$

Condition COH 1 is satisfied, but the second inequality of COH 2 is not for $i=2$, since $\underline{p}_2+\bar{p}_1+\bar{p}_3=0.9$. Given that $\bar{p}_1=\bar{p}_3=0.3$, this expert's upper probability for the event $X\in I_1\cup I_3$ should not exceed 0.6, so his lower probability for $X\in I_2$ should not be less than 0.4 (this is based on axioms IP3 and IP4 of section 2.1). If this form of incoherence is found, the expert should be asked to reconsider the assessed values, where indeed replacement of $\underline{p}_2=0.3$ by $\underline{p}_2=0.4$ is the simplest solution to make the assessments coherent, but of course the expert may prefer to change $\bar{p}_1$ or $\bar{p}_3$, in which case also the other inequalities of COH 2 need to be verified again.

□

To end this section we remark that Walley and Campello de Souza (1990) have elicited imprecise probabilities for events $X\leq x$, and thus imprecise cdf's, in their case-study, and the consulted experts felt comfortable with it. In example 3.4.1 we also assumed data of this kind. For practical elicitation of imprecise probabilities for lifetimes one could ask the experts which type of events to use, $X\in I_i$, $X\leq x_i$ or e.g. $X>x_i$. For combination and model fitting it is no problem if different experts use different events as long as the same x_i are used.

4.3 Combination of imprecise probabilities

In this section combination of the opinions about event A, expressed by imprecise probabilities, of $K \in \mathbb{N}$ experts is discussed, assuming that expert k ($k=1,..,K$) has assigned $\underline{P}_k(A) \leq \overline{P}_k(A)$. Combination of imprecise probabilities can have two different goals, summarization of the opinions of several experts, or making the best decision possible. For the second goal it seems to be logical that scoring rules are used, related to weights for the experts, so that experts who have shown to have good knowledge in the past get high weights. Calibration should be possible either by the use of old probability assessments and observations thereafter or by using seed variables (Cooke, 1991; Vreugdenhil, 1993).

Three simple methods for combining imprecise probabilities are discussed, two of these can be used for the first goal, and the last one, combination by weighted averages seems to be promising for making decisions. We focus on interpretation of the combined imprecise probabilities in terms of group betting behavior and briefly discuss some ideas for theories of weights for experts, leaving an important topic for research with perhaps possibilities to generalize the ideas used in the theories for combination of precise probabilities (Cooke, 1991).

From the interpretation of imprecise probabilities in terms of betting behavior, two rules of combination follow immediately (Walley, 1991; chapter 4), with obvious interpretations in terms of group betting behavior

$$(I) \qquad \underline{P}^I(A) = \min_{k} \underline{P}_k(A); \qquad \overline{P}^I(A) = \max_{k} \overline{P}_k(A); \qquad (4.3)$$

$$(II) \qquad \underline{P}^{II}(A) = \max_{k} \underline{P}_k(A); \qquad \overline{P}^{II}(A) = \min_{k} \overline{P}_k(A). \qquad (4.4)$$

Combination by rule (I) implies that the group only accepts to buy or sell the bet on A if each member of the group accepts it individually, so these numbers indicate consensus. Walley (1991; section 4.3.9) calls this the unanimity rule for combination, referring to the fact that a bet is only accepted by the group if all experts accept it individually.

Combination by rule (II) implies that the group agrees to buy or sell the bet on A if at least one member accepts this. Walley (1991; section 4.3.8) calls this the conjunction rule for combination, since the set of bets that are accepted by the group is the conjunction of all sets of accepted bets per ex-

pert.

It is easy to see that rule (I) leads to coherent betting behavior (Walley, 1991; section 4.3.9), but rule (II) can lead to incoherence since $\underline{P}^{II}(A) > \bar{P}^{II}(A)$ is possible. This indicates that there are two members within the group who would be willing to exchange a bet on A at prices between $\bar{P}^{II}(A)$ and $\underline{P}^{II}(A)$, and thus there is obvious conflict between the thoughts of these experts.

The results of combination by rules (I) and (II) represent facts of the opinions of all members of the group that are important for a decision maker. In the case of precise probabilities two experts are not willing to exchange bets only when their probabilities are identical. In literature much attention has been paid to combination of precise probabilities (Cooke, 1991), but no rules like (I) and (II) follow straightforwardly from the interpretations of probabilities. For decision making these rules are generally not attractive, since rule (I) can lead to great imprecision and rule (II) to very little imprecision, or even incoherent group betting behavior. Nevertheless, situations may occur in which, for example, rule (I) is logical, think about a group of experts such that each has different information about event A, or a completely different background that affects their opinion, and a DM might decide to accept a bet only if every single expert agrees.

A third possibility for combining imprecise probabilities is the weighted average rule. In the literature on combination of precise probabilities this plays an important role (Cooke, 1991), and it has also been discussed in an early work by Walley (1982), who pointed out a disadvantage that discouraged further development (it is not discussed in Walley (1991)). After presenting the general form of this rule, we examine this argument against the rule, which we believe is not a serious disadvantage, and we give an interpretation in terms of betting behavior that seems to be logical. An important topic for future research is a theory of weights, and we will make remarks in this direction.

Suppose the decision maker assigns weight w_k to expert k, with $0 < w_k < 1$ and $\sum_{k=1}^{K} w_k = 1$. The combined lower and upper probabilities by the weighted average rule are

$$\text{(III)} \qquad \underline{\mathbb{P}}^{\text{III}}(A) = \sum_{k=1}^{K} w_k \underline{\mathbb{P}}_k(A); \qquad \overline{\mathbb{P}}^{\text{III}}(A) = \sum_{k=1}^{K} w_k \overline{\mathbb{P}}_k(A). \qquad (4.5)$$

Walley (1982) starts with some criteria for combining imprecise probabilities (these are deleted in his later work). The weighted average rule violates some of these, most importantly the so-called reconciliation criterion that is regarded as crucial to the theory. This criterion states that if some member of the group of experts is willing to sell the bet on event A for price p, then the group should not commit itself to pay more than p for the bet, so the combined lower probability should not be greater than the minimum of the individual upper probabilities.

When violating this criterion, disagreement within the individual opinions may not be reported, so the weighted average rule is not suitable for the first goal, summarizing of opinions to give the decision maker an overview. For the second goal, making a decision, there are no arguments given so far in favor of the reconciliation criterion, and the possible interpretation of the combined imprecise probabilities by the weighted average rule seems acceptable and in accordance with intuition.

Suppose a decision maker has to state his lower probability for event A, that is his supremum buying price for the bet that pays 1 if A occurs, and 0 otherwise. It is obvious that this lower probability is not greater than 1, so 1 could be regarded as his budget. For this decision he wants to rely on K experts, each of them will be responsible for a part of his budget. If expert k is responsible for budget w_k, than his individual lower probability $\underline{\mathbb{P}}_k(A)$ implies that he is willing to pay up to $w_k \underline{\mathbb{P}}_k(A)$ for a bet on A that pays w_k if A occurs and 0 otherwise. The supremum buying price for the decision maker's bet on A is the sum of the $w_k \underline{\mathbb{P}}_k(A)$ of all experts, indeed giving a logical interpretation to $\underline{\mathbb{P}}^{\text{III}}(A)$ in terms of group betting behavior.

For practical use of rule (III) research is needed on derivation of sensible weights w_k, especially if new information becomes available, which has not yet been presented in literature. A first possibility would be to let the decision maker choose the weights, where $w_k = 1/k$ may be logical if no reasons exist for discriminating between the experts. Restricted to precise probabilities, weighting according to confidence in the expertise of the experts has been studied by Lindley and Singpurwalla (1986), where the decision maker is asked to take correlations between experts into account. We believe that such models are mathematically attractive, but can rarely be used in practice.

Other possibilities have been mentioned in the literature on combination of

precise probabilities by weighted averaging, e.g. the use of seed variables, using a scoring rule that indicates the performance of the experts on some questions about related issues that can be evaluated (Cooke, 1991), an inter-action model in which each expert k assigns weights to all other members of the group that are to be combined (De Groot, 1974), or the use of scoring rules related to calibration (Cooke, 1991; Vreugdenhil, 1993) that seem to be promising but can only be used when many observations have become available so that opinions can be compared to real outcomes.

Perhaps a theory of weights can be derived by use of the actual betting inter-pretations of imprecise probabilities. This idea has not been discussed before, but seems to be attractive for a theory of weights through time, since the budget of a good expert will increase by the betting. Such ideas need much consideration, since imprecise probabilities state a supremum buying price and an infimum selling price for the bet, and both should be regarded such that the weight also depends on the amount of imprecision in a logical way, where research must be started by discussing what is logical here. We believe that it would be very useful for a decision maker to get information about the performances of the experts through time, and a theory is desirable in which the better ones are rewarded by more responsibility.

Although the weighted average combination has a logical interpretation, it might also be useful to consider non-linear weighted combinations of expert opinions, that have some advantages in combining precise probabilities (Cooke, 1991), but may turn out to have no reasonable interpretation when generalized to our concept.

4.4 Model fitting

To apply the theory proposed in chapters 2 and 3, the expert opinions must be modeled by imprecise prior distributions for a parameter of an assumed model. In section 4.2 however, we suggested to restrict elicitation to statements about observable random variables. In this subsection we propose a solution for the remaining problem of choosing values for the hyperparameters of an assumed model such that the model fits well to the expert's imprecise probabilities. Also some related aspects are briefly discussed, not as a contribution but to make the reader aware of interesting topics in literature that are also important for our theory.

The first problem in practical application of a statistical theory in decision making is determination of the random variable of interest and, in our case, the choice of an appropriate parametric model. Restricting the discussion to a random variable X that represents the lifetime of a technical system, the assumption of a distribution for X described by a probability density function $f_X(x|\theta)$ on $\mathbb{R}_+$ is often quite reasonable. The choice of such a pdf, defined up to the value of θ, should be based on an analysis of the situation of interest, where the analyst should also discuss underlying assumptions with the decision maker and experts with regard to the system. This pdf is also important for its role in the likelihood function that is used in Bayesian updating. Bayarri, DeGroot and Kadane (1988) discuss the likelihood function related to a model, and show that some consideration is necessary if it is not entirely clear which random variable will be observed and which model parameter must be updated in the Bayesian framework.

Mendel (1992) and Spizzichino (1992) discuss Bayesian lifetime models based on the idea of De Finetti-type representations. This research is based on the following philosophical problem (Mendel, 1992)

'A popular method for obtaining Bayesian models is to adopt a frequentist model and simply provide the (frequentist) parameters with prior distributions. A Bayesian should justify the adoption of such models, for instance using information about the physical processes underlying the wear of the units. Moreover, the frequentist parameters are abstract quantities indexing possible 'true models' (in the frequentist sense) without an operational meaning in terms of physically observables. It is unclear how a Bayesian provides such quantities with meaningful priors.'

We have restricted ourselves to the 'popular method', and only mention the possibility of De Finetti-type representations with imprecision as an interesting subject for future research. One important argument for this lack of confidence in the so-called frequentist models is the seemingly paradoxical observation that the hazard rate (also called failure rate in the literature) of a predictive distribution behaves differently to the hazard rate of the model $\ell_X(x|\theta)$, where for example an exponential pdf $\ell_X(x|\theta)$ with constant hazard rate almost always leads to a predictive distribution with decreasing hazard rate. Although throughout this thesis much emphasis is put on the role of predictive distributions (Geisser, 1993), it is believed that Barlow (1985) is right in showing that this feature is not in disagreement with subjective ideas and follows immediately from exchangeability of trials by a full subjective Bayes description, as shown for the discrete case with geometric models by Rodrigues and Wechsler (1993). Therefore, 'frequentist' arguments for the choice of a lifetime model pdf $\ell_X(x|\theta)$, such as increasing hazard rate, are regarded valid.

It is also useful to consider the parameterization within the model with regard to the adequacy and efficiency of numerical and analytical techniques (Hills and Smith, 1992). The dependence of statistical inferences on the actual choice of the parameter is less severe than is often suggested in arguments against the Bayesian methodology. An interesting argument against the virtual problem of parameterization is provided by Singpurwalla (1992; also Abel and Singpurwalla, 1991), who shows that in the case of an exponential model with mean θ and hazard rate $\lambda=\theta^{-1}$, with conjugate priors, at any time a failure is more informative (using information theoretic arguments) about θ than a survival, but that the exact opposite is true for λ. As an additional argument we remark that the predictive distribution is the same, either arrived at through the model with θ or λ and a conjugate prior. Lindley (1992) comments that, although Bayesian ideas are often best discussed in the context of observables, there are circumstances in which parameters, though not observable, are useful, particularly in scientific studies where the parameters have well-understood meanings for the scientist. Inferences for model parameters are only of interest if these have clear meanings for the scientists, and the parameterization issue is not a serious disadvantage to Bayesian methods.

Once a pdf $\ell_X(x|\theta)$ is chosen, all further information is only used for inferences about θ, that is also regarded to be a random variable, and the second part of the model is assumption of a class $\Pi_{\theta,0}(\ell_{\theta,0},u_{\theta,0})$ of prior densities

for 0 of the form (2.1).

Computation of posterior densities is simplified by choosing $\ell_{\theta,0}(t)$ and $u_{\theta,0}(t)=\ell_{\theta,0}(t)+c_0\alpha_{\theta,0}(t)$, with $\ell_{\theta,0}$ and $\alpha_{\theta,0}$ members of a conjugate family with regard to the model, if such family exists (section 2.3). Although simple analytical calculation is often mentioned as the only reason to adopt conjugate priors (Raiffa and Schlaifer, 1961), there is a more basic foundation for the conjugacy property. A prior is conjugate if it has the same functional form as the likelihood function, implying that the posterior will also have the same form. The hyperparameters of the prior play exactly the same role as sufficient statistics of data in the likelihood, and can therefore logically be interpreted as sufficient statistics for imaginary data, so prior knowledge can be expressed by or compared with such imaginary data. Although we do not directly use this feature in elicitation, if values for the hyperparameters are chosen according to a method as will be proposed next, interpretation of the resulting values for the hyperparameters as sufficient statistics of imaginary data can be helpful for the statistician in the analysis of the subjective data. For the binomial case, the logical way in which a conjugate prior follows from the parametric model is discussed by Colombo and Constantini (1980; see also section 2.4), but the same feature holds in general for conjugate priors and is therefore an argument in favor of restriction to such priors that is stronger than simplicity of calculations on updating. Of course, from a purely subjective viewpoint the prior must represent the thoughts of the subject and should therefore not be restricted beforehand, but because of elicitation problems, and in most cases the richness and flexibility of the family of conjugate priors, restriction to such families seems quite reasonable for practical application of our theory.

Although the choice of a parametric model is an important element of our concept of statistics and decision making, once expert opinions are elicited and a final decision is to be made the statistician may perform the analysis using several models, for example by means of sensitivity analysis. If expert opinions are elicited about many different events concerning X, so the expert's imprecise cdf's are specified very accurately, then it could also be possible to distinguish between some possible models and choose that one which fits best according to some criterion as discussed in this section. The statistician should be careful not to overvalue the results of the elicitation process, since in most situations these will not be as good as possible by restrictions in time available per expert for thinking about the events of interest.

Another possibility is also to define the model imprecisely, leading to imprecise likelihood functions as discussed briefly by Walley (1991; sections 8.5 and 8.6). At this moment few results have been presented related to imprecise likelihood functions, which suggests an important area of research, with many applications since often the assumption of a precise model pdf $\ell_X(x|\theta)$ is doubtful, for example if nuisance parameters are used to create a robust sampling model (Walley, 1991; section 7.2.7).

Suppose that for $X \in \mathbb{R}_+$ the expert opinions are quantified by imprecise cdf's at $0 = x_0 < x_1 < ... < x_m = \infty$, with discussion restricted to values for one expert, but of course these can easily well be interpreted as the results of combination. Such imprecise cdf's can either be achieved directly, or by application of the histogram method proposed in section 4.2, that leads to $\underline{p}_i$ and $\overline{p}_i$ $(i=1,..,m)$ that are assumed to be coherent according to (4.1) and (4.2), and result in the following values for the imprecise cdf's (compare (2.76) and (2.77))

$$\underline{F}_X(x_i) = \frac{\sum_{j=1}^{i} \underline{p}_j}{\sum_{j=1}^{i} \underline{p}_j + \sum_{j=i+1}^{m} \overline{p}_j} \tag{4.6}$$

and

$$\overline{F}_X(x_i) = \frac{\sum_{j=1}^{i} \overline{p}_j}{\sum_{j=1}^{i} \overline{p}_j + \sum_{j=i+1}^{m} \underline{p}_j}. \tag{4.7}$$

Next we outline a general methodology for fitting the assumed model, so choosing appropriate values for the hyperparameters such that the according imprecise prior predictive cdf's at x_i are close to the values representing the expert opinions. The imprecise prior predictive cdf's for the assumed model and class of priors as in section 2.3, are denoted by $\underline{F}_X^M(x_i|\tau)$ and $\overline{F}_X^M(x_i|\tau)$, with τ the hyperparameter (say $\tau \in T$, $T \subseteq \mathbb{R}^k$ for finite k) that was suppressed in the notation used in chapter 2. Since for fitting the model to the expert opinions we need to consider only the priors, the additional index 0 used in chapter 2 is deleted here.

To solve the problem of choosing a reasonable value for τ, a suitable expression is needed for the distance between the expert's and model cdf's. Let

the interval probabilities related to the lower and upper cdf's be defined by $(i=1,..,m)$

$$\underline{f}_i = \underline{\mathbb{F}}_X(x_i) - \underline{\mathbb{F}}_X(x_{i-1}) \tag{4.8}$$

$$\overline{f}_i = \overline{\mathbb{F}}_X(x_i) - \overline{\mathbb{F}}_X(x_{i-1}) \tag{4.9}$$

$$\underline{f}_i^M(\tau) = \underline{\mathbb{F}}_X^M(x_i|\tau) - \underline{\mathbb{F}}_X^M(x_{i-1}|\tau) \tag{4.10}$$

$$\overline{f}_i^M(\tau) = \overline{\mathbb{F}}_X^M(x_i|\tau) - \overline{\mathbb{F}}_X^M(x_{i-1}|\tau). \tag{4.11}$$

As criterion to determine a suitable value for τ we propose minimization of the sum of mean distances, related to a metric $d(v,w)$, defined by

$$\text{SMD}(\tau) = \sum_{i=1}^{m} \left\{ \underline{f}_i d(\underline{f}_i, \underline{f}_i^M(\tau)) + \overline{f}_i d(\overline{f}_i, \overline{f}_i^M(\tau)) \right\}, \qquad \tau \in T. \tag{4.12}$$

So a solution of the model fitting problem is $\hat{\tau} \in T$, if it exists, such that

$$\text{SMD}(\hat{\tau}) = \inf\{\text{SMD}(\tau) \mid \tau \in T\}. \tag{4.13}$$

In many situations models corresponding to τ_ε with

$$\text{SMD}(\tau_\varepsilon) \leq \text{SMD}(\hat{\tau}) + \varepsilon, \tag{4.14}$$

where $\varepsilon > 0$ is small, also provide good approximations to the expert's cdf's, solving the theoretical problem that occurs if $\hat{\tau}$ does not exist as well as the practical problem of exact determination of $\hat{\tau}$ when it exists.

This criterion is based on the idea that the interval probabilities according to the lower and upper model cdf's should approximate the interval probabilities according to the lower and upper expert's cdf's as well as possible. This form, with the mean of the metric distances taken with regard to the expert's cdf's can be regarded as a generalization of Kullback-Leibler information (Kullback and Leibler, 1951), where instead of a general metric $d(v,w)$ the natural logarithm of the ratio of v and w is used.

In frequentist statistics, the Kullback-Leibler information provides a foundation for Maximum Likelihood estimation (Ranneby, 1984), and some situations

where the ML-method does not work properly are easily explained by the fact that the logarithm in Kullback-Leibler information is not a metric. Solutions for problems in calculation of Bayesian posteriors caused by unbounded likelihoods can also be found through the use of Kullback-Leibler information, as shown by Coolen and Newby (1994).

The SMD according to (4.12) cannot be split up easily into separate parts for the lower and upper cdf's, because different parts of τ that are related to the lower and the upper prior densities both occur in $\underline{F}_X^M(x_i|\tau)$ as well as in $\overline{F}_X^M(x_i|\tau)$. Numerical calculation of $\hat{\tau}$ or τ_ε for a small ε needs consideration since the dimension of T may be high. In the model proposed in chapter 2, τ consists of the hyperparameters of both the lower and upper densities to which also c_0 belongs, according to relation (2.26). For practical problems reasonable model approximations of the expert's imprecise cdf's can possibly also be derived at by choosing some of the parts of τ based on other criteria. In example 3.4.1 c_0 was determined by an ad hoc methodology based on relation (2.15), simplifying the numerical problem considerably.

To use the minimum SMD criterion (4.13), a metric $d(v,w)$ must be chosen by the statistician, where for example the metrics

$$d_1(v,w) = (v-w)^2 \tag{4.15}$$

$$d_2(v,w) = |v-w| \tag{4.16}$$

and

$$d_3(v,w) = |\ln(v)-\ln(w)| \tag{4.17}$$

relate to distance measures that are quite familiar in statistics, with d_3 a variation on Kullback-Leibler information. The choice should be based on what is regarded as best in relation to the metric, for example d_1 could be chosen when one prefers many small differences to a few larger differences. At this point the statistician could perhaps also communicate with the decision maker, although it may be hard to get useful information. Of course, as often in statistics we could just propose to use a certain metric, but as there are seemingly no general reasons in favor of a certain metric we will not do so. In frequentists statistics d_1 is often preferred for simplicity of calculations, which hardly plays any role in our case. In example 3.4.1 the metric d_1 has been used for model fitting.

Finally, in our model the value of ξ must be chosen by the decision maker (see relations (2.36)-(2.39)), based on the interpretation of ξ as the number of new data that provides as much information as that on which the imprecise prior probabilities are based, and therefore ξ is purely subjective and represents the relevance the DM adds to new information when compared to prior information. Only by applications it can be discovered if assessment of such ξ is directly possible.

4.5 Reliability theory using imprecise probabilities

Reliability analysis is concerned with the ability of a system to function satisfactorily under a given set of circumstances, and thus gives important information about the quality of a system, when to replace it, or when and how to maintain it. Many probabilistic models for the reliability of components and systems are available and can be used without too much difficulty whenever the parameters of the system are given or are deducible from the systems properties. The situation is more difficult when parameters must be estimated from data. The use of models is often limited by a lack of satisfactory data on which estimates of system parameters can be based. The reasons for the lack of data are numerous, sometimes there is no effective registration system, sometimes, the data are corrupted, and changes in design or operating conditions mean that historical data are no longer directly relevant.

The Bayesian approach to statistical inference provides a framework for dealing with expert opinions and prior knowledge, and therefore offers solutions to some of the difficulties in reliability analysis (Martz and Waller, 1982; Sander and Badoux, 1991). To deal correctly with both kinds of uncertainty present in expert opinions, as discussed by Walley (1991) and this thesis, generalization to imprecise probabilities is necessary, and introduction of imprecision in reliability theory asks for generalized definitions of concepts used herein. For a continuous, nonnegative real random variable X (in many reliability problems X is a random lifetime) we define lower and upper reliability functions, hazard rates and hazard functions, related to an intervals of measures class $\Pi(\ell,u)$ as given by relation (2.1), and we show the possible use of imprecise probabilities in reliability problems by an example.

In reliability theory, the reliability function, hazard rate and hazard function are important characterizations of probability distributions, with the hazard rate providing a natural description of a failure process.

Definitions

Suppose continuous lower and upper densities ℓ and u are specified for X, and the class of possible pdf's for X is $\Pi(\ell,u)$ as given by (2.1). The corresponding imprecise cdf's $\underline{F}$ and $\overline{F}$ are given by (2.4) and (2.5).

The *lower and upper Reliability functions* are $(x \geq 0)$

$$\underline{R}(x) := 1 - \overline{F}(x) \tag{4.18}$$

and

$$\overline{R}(x) := 1 - \underline{F}(x), \tag{4.19}$$

respectively.

The *lower and upper hazard rates* are

$$\underline{h}(x) := \frac{\ell(x)}{\displaystyle\int_x^\infty u(y)\,dy} \tag{4.20}$$

and

$$\overline{h}(x) := \frac{u(x)}{\displaystyle\int_x^\infty \ell(y)\,dy}, \tag{4.21}$$

respectively.

The *lower and upper Hazard functions* are

$$\underline{H}(x) := -\ln(1-\underline{F}(x)) \tag{4.22}$$

and

$$\overline{H}(x) := -\ln(1-\overline{F}(x)), \tag{4.23}$$

respectively.

The definitions (4.18) and (4.19) of imprecise reliability functions are obvious. The imprecise hazard rates (4.20) and (4.21) are the tightest lower

and upper envelopes of the set of all hazard rates of pdf's that are derived from normalization of densities between ℓ and u, which is easily proved by taking a function q with $\ell(x) \leq q(x) \leq u(x)$ for all $x \geq 0$ and $C_q = \int_0^\infty q(y)dy$ the normalizing constant. The hazard rate corresponding to pdf $q(x)/C_q$ is

$$h_q(x) = \frac{q(x)/C_q}{\int_x^\infty q(y)/C_q \, dy} = \frac{q(x)}{\int_x^\infty q(y)dy} \tag{4.24}$$

and the fact that ℓ and u are the lower and upper envelopes of all such q makes clear that $\underline{h}$ and $\overline{h}$, defined by (4.20) and (4.21), are the lower and upper envelopes of all such h_q.

The imprecise hazard functions (4.22) and (4.23) are the lower and upper envelopes of all possible hazard functions relating to possible pdf's $q(x)/C_q$, since the hazard function for such a pdf is $H_q(x) = -\ln(1-F_q(x))$, with F_q the cdf corresponding to pdf $q(x)/C_q$.

Remark that $\underline{h}$ and $\overline{h}$ are not the first derivatives of $\underline{H}$ and $\overline{H}$, in fact $\underline{h}(x) \leq \underline{H}'(x)$ and $\overline{h}(x) \geq \overline{H}'(x)$, with equalities if and only if $\ell(x) = u(x)$ for all $x \geq 0$.
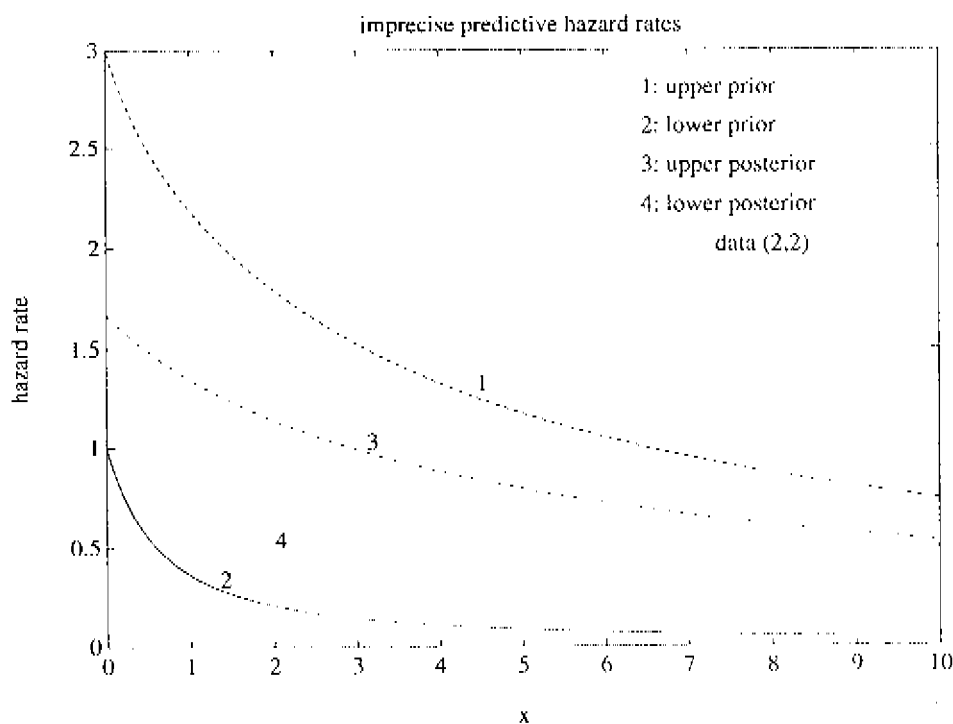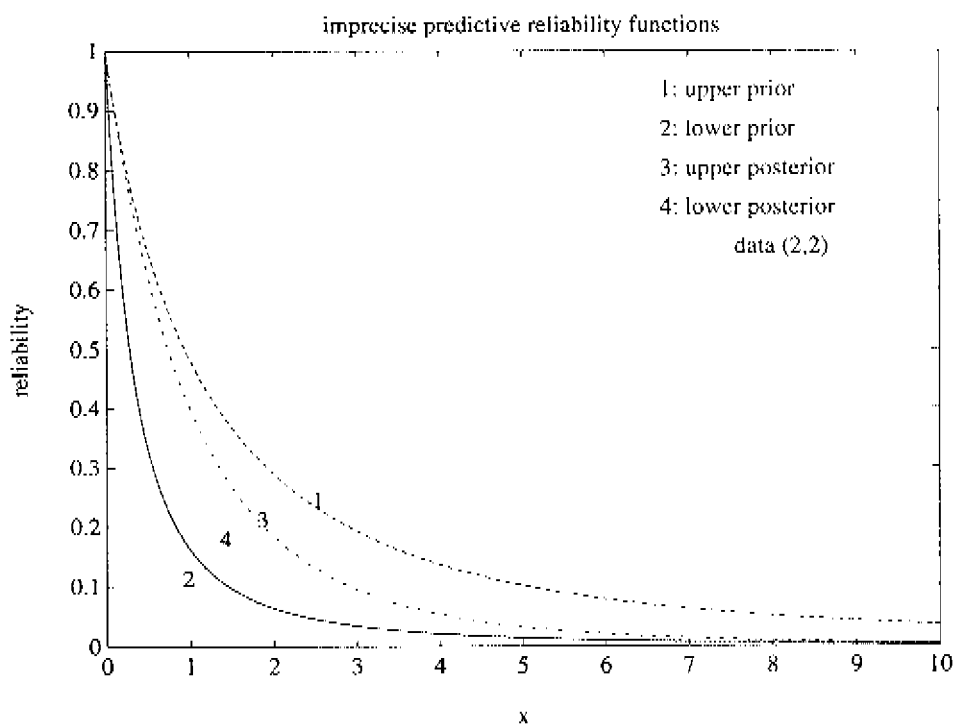
Possible application of the concept of imprecise probabilities in the theory of reliability is briefly illustrated by example 4.5.1. Of course, also the examples presented in section 2.4 on Bernoulli experiments can be of interest in reliability problems.
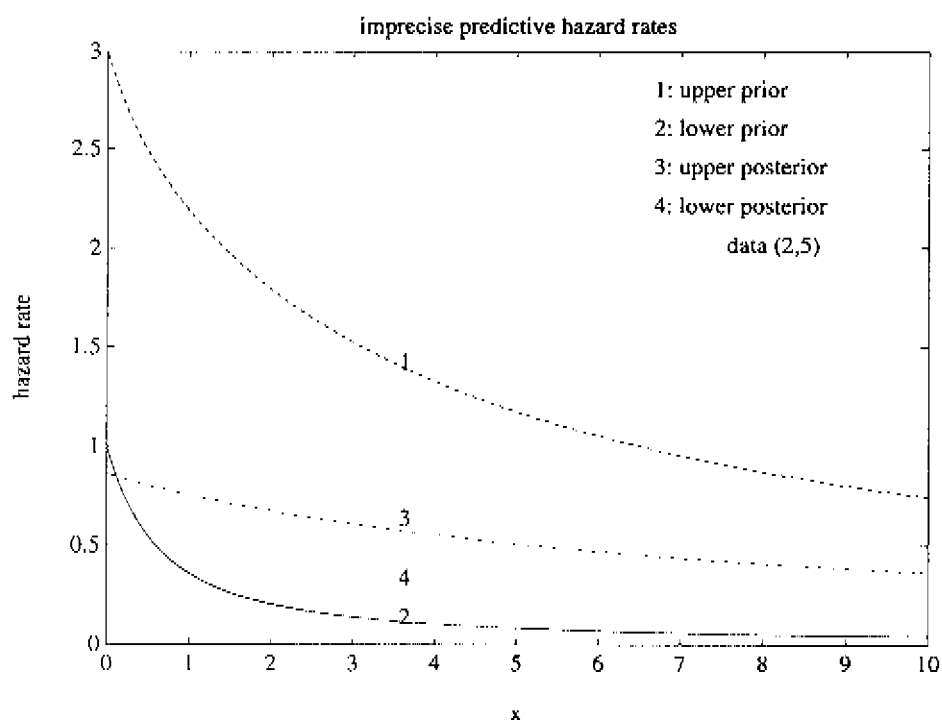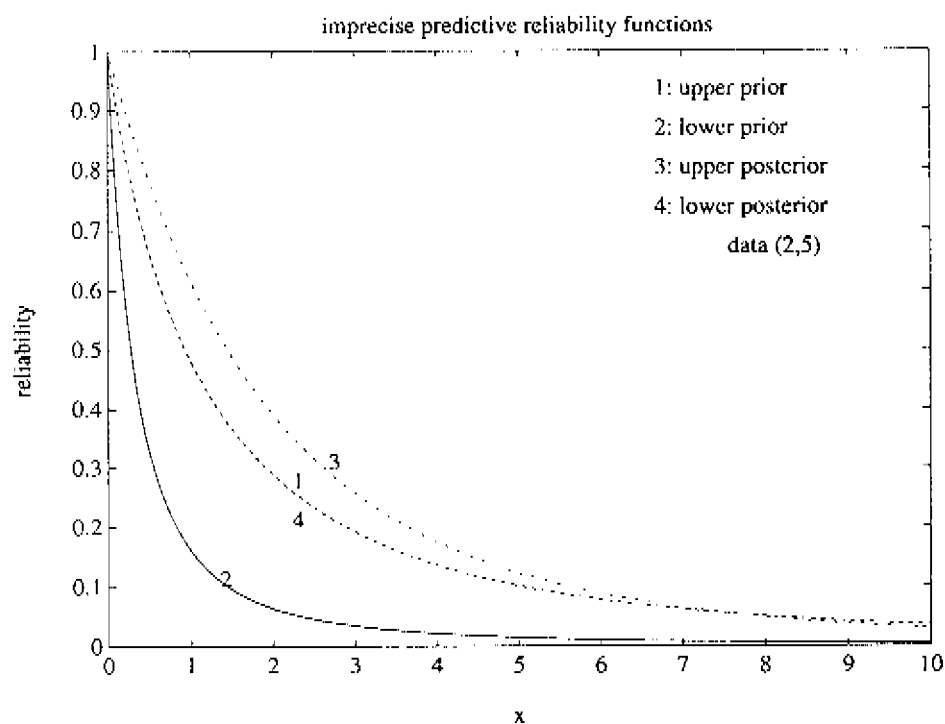

Example 4.5.1

Suppose that we are interested in the length of intervals between breakdowns of a system, which follow a Poisson process, thus the intervals are independent and identically exponentially distributed. We consider the problem of determining a distribution for the length of the intervals between consecutive failures that will occur in future. The assumed parametric model for the length X of intervals between failures is

$$\ell_X(x|\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0 \text{ and } \lambda > 0. \tag{4.25}$$

Figures 4.5.1 *Illustrations example 4.5.1*

imprecise predictive reliability functions

1: upper prior
2: lower prior
3: upper posterior
4: lower posterior
data (2,2)

imprecise predictive hazard rates

1: upper prior
2: lower prior
3: upper posterior
4: lower posterior
data (2,2)

imprecise predictive reliability functions

1: upper prior
2: lower prior
3: upper posterior
4: lower posterior
data (2,5)

imprecise predictive hazard rates

1: upper prior
2: lower prior
3: upper posterior
4: lower posterior
data (2,5)

As a conjugate class of prior densities we suggest $\Pi_{\lambda,0}(\ell_{\lambda,0},u_{\lambda,0})$ specified by (for $t\geq0$)

$$\ell_{\lambda,0}(t) = te^{-t} \tag{4.26}$$

and

$$u_{\lambda,0}(t) = \ell_{\lambda,0}(t) + c_0 * 4te^{-2t} \tag{4.27}$$

with $c_0=1$. Further, to specify c_n according to (2.36) in our model, let $\xi=1$, implying that the value of the subjective prior information is regarded to be equivalent to that of one observation.

It is worth to remark that Littlewood's (1991) software reliability model can be generalized in a similar way if one assumes that the mixing distributions are given imprecisely in the gamma family.

If data $x(n)$, according to (2.24), become available, sufficient statistics are (n,s) with $s=\sum_{i=1}^{n} x_i$, and the imprecise posterior densities for λ are $(t\geq0)$

$$\ell_{\lambda,n}(t|n,s) = (1+s)[(1+s)t]^{1+n} * e^{-(1+s)t} / \Gamma(2+n) \tag{4.28}$$

and

$$u_{\lambda,n}(t|n,s) = \ell_{\lambda,n}(t|n,s) + c_n * (2+s)[(2+s)t]^{1+n} * e^{-(2+s)t} / \Gamma(2+n), \tag{4.29}$$

with $c_n=(1+n)^{-1}$.

In figures 4.5.1 the imprecise predictive reliability functions for X and the corresponding imprecise hazard rates are given, together with the same functions for the imprecise posteriors derived after $n=2$ observations with sum $s=2$. Also the same functions are given with posteriors for data with $n=2$ and $s=5$, in which case the data is somehow in conflict with prior thoughts as can be concluded from the imprecise predictive reliability functions.

□

4.6 Additional remarks

The title of this chapter may have suggested more than is actually presented, since only aspects that are important for implementation of the concept of imprecise probabilities have been discussed, but without an actual implementation. It has been remarked that such implementation is necessary before conclusions can be drawn and the methodology proposed in this thesis can be evaluated. Nevertheless, by comparing the current method to standard concepts based on precise probabilities, the advantages of our concept have been illustrated, in particular the interpretation of the imprecise probabilities. This leaves many open questions for future research. For example, in section 4.2 on elicitation the idea of scoring rules could be related to results of bets, while there is still no theory for the suggested weighted average combination method in section 4.3.

A serious problem in applications is that cooperation of people from several disciplines is needed, for example because computers will play an important part in the elicitation process. Further, there seems to be an obstacle to a first application within firms, since solving of decision problems with several experts will take considerable preparation time, training in elicitation and the actual elicitation process and so on; it will demand several days from the person responsible for the final decision. Since it is not yet possible to guarantee success, many firms will not be too enthusiastic to cooperate in an experiment. It is not surprising that very few case-studies are known about decision problems using new concepts, especially when compared to the number of theoretical publications.

This chapter is intended as a possible starting point for real implementation. It is believed throughout that the interpretation of the quantities of interest is important, and that the theory of imprecise probabilities offers better possibilities from this point of view than some other recently proposed methods, such as fuzzy sets theory, that have received much attention in applications, particularly by engineers with a non-mathematical background.

EPILOGUE

The starting-point for this thesis was the need for a theory to update imprecise probabilities. The development differs from Walley's (1991) Generalized Bayes Rule. In our updating scheme, the imprecision is always less than that for Walley's GBR and the requirement of coherent betting behavior through time is dropped. The intuitive relation between imprecision and information is formalized and plays a central role.

In section 2.2 an argument is presented in favor of Walley's measure of the amount of information in terms of imprecision, leading to the use of this measure as a fundamental part of the updating theory. Nevertheless, formalization of the intuitive relation between information and imprecision still remains an important topic for research, since there is still no general theory. It is not clear if a directly measurable subjective measure of information can be defined, at the moment we are restricted to the expression of information in terms of imprecise probabilities. Also analogously to our approach other measures of information can perhaps be proposed that make sense, in which case grounds for comparison between information measures are needed.

Section 2.3 presents an updating theory for imprecise probabilities with a central role for imprecision and information, leading to less imprecise updated imprecise probabilities than those derived from Walley's GBR. The main conclusion to be drawn is that it is indeed possible to emphasize imprecision in an updating theory. The method proposed is less general than Walley's GBR, which is defined using the concept of imprecise previsions (probabilities are a special type of previsions) and deals with other sorts of information for updating. Our method is restricted to parametric models, and adopts a Bayesian framework with imprecise prior probabilities defined as intervals of measures. As such, the methodology is related to the robust Bayesian concept, in which imprecision serves only as a form of sensitivity analysis.

The main challenge for future research is an updating theory as general as Walley's GBR, but without the assumption of coherence through time and giving a central role to imprecision and information.

The use of the concept in the case of Bernoulli experiments is discussed in section 2.4, with an analysis of model behavior when the additional data conflict with prior thoughts. This is also important when considering a more general theory, and is a part of the necessary formalization of the value of information.

Directly related to the methodology described in this thesis are the topics that have been discussed briefly in section 2.5, updating in case of censored data and methods for statistical inferences.

Chapter 3 presents a generalization of the standard Bayesian framework for inference using imprecise prior distributions. Section 3.2 provides this generalization for all inferences based on loss functions (Lindley, 1990). Given the central role of loss functions, their choice remains an interesting subject for research, just as it is in the standard Bayesian framework. Generalization to allow imprecise loss functions is of interest, as well for sensitivity analysis as to cover practical situations when the loss function is not known precisely. To compare decisions we suggest the use of the bounds on expected loss, which relate naturally to the betting behavior expressed as imprecise probabilities. It may be of interest to analyze other possibilities for comparing decisions through the sets of values for expected loss.

In section 3.3 results are presented for the calculation of bounds on the expected loss that are useful in applications since the necessary numerical calculation stays within reasonable limits when compared to the standard Bayesian approach without imprecision. The calculation requirements should also be considered for more general theories, or when other models are used. In section 3.4 an example of an application to a decision problem concerning preventive replacement of technical equipment is presented. The example serves to emphasize the intuitively correct fact that lack of perfect information, modeled through imprecision, leads to intervals of values for expected loss per decision, and to stress the fact that this does not lead to indecisiveness but explicitly asks for comparison of these intervals. Hypothetical examples like this one help to convince decision makers of the possibilities of new methods like ours, which is a necessary first step in the direction of applications, that would be much more useful and could be used to evaluate the methodology.

We have now reached the issue of implementation, which is by far the most important topic for future research from a practical point of view. In chapter 4 some important aspects of implementation are discussed, but the fact that the method has not yet been applied to practical decision problems makes it impossible to draw conclusions. As a starting point for research, both practical and theoretical, some aspects of the elicitation and combination of expert opinions and model fitting have been illustrated with the theory proposed in chapter 2.

In section 4.2 the importance of using only questions about interpretable events and a user-friendly computer package for elicitation is discussed. Generally, the idea of scoring rules might be useful in the training phase, when the experts are getting used to the elicitation tool, but the use of arbitrarily chosen functions of probability densities should be avoided because of interpretation problems. The possibility of using the scoring idea inherent in betting on uncertain events can perhaps be exploited. It is obvious that practical research in this area requires the cooperation of firms, since managers and experts must have time available for all steps that are necessary for application of the theory. Moreover, the cooperation of researchers from several disciplines, for example analysts with psychological and computer science backgrounds, is needed to study the many aspects of practical elicitation.

Combination of imprecise probabilities, discussed in section 4.3, serves two important goals. Firstly, summarization of the opinions of several experts, for example to indicate serious disagreements. Secondly, good decision making based on information from several experts. The concept of imprecise probabilities offers the possibility of achieving these goals and of enabling the use of combination rules so that the combined probabilities also have a sensible interpretation. Combination by a weighted average rule has been proposed as a possible method with regard to good decision making but creates the problem of a theory for weights per expert. Other combination rules might also be suitable, or better than weighted averages, where it seems to be a logical prerequisite that the combined imprecise probabilities are interpretable in terms of betting behavior.

In section 4.4 a method is proposed for choosing values for hyperparameters so that the assumed model fits well to the elicited expert opinions. One could think of many methods for this, but it is hard to find arguments for a preferred method. Even when restricted to the proposed method, where the distance between the imprecise cumulative distribution functions of the models and the experts is minimized, the choice of a suitable measure for the distance remains open. We have suggested three metrics, but in all practical situations this needs careful consideration. It is to be remarked that this aspect is often hidden in a mathematical black box, where ease of computations is often the most important reason for selecting a distance measure. From this point of view there is no clear preference for a particular metric.

A possible area for implementation is reliability analysis, and our concept is applied in section 4.5 to provide generalized definitions of reliability functions, hazard rates and hazard functions. This area is mentioned here because

recently reliability analysts have shown interest in new quantitative concepts for decision making with relation to uncertain events, where expert opinion is often an important source of information about the problems.

In conclusion, the results presented in this thesis show that it is possible to delete the assumption of coherent updating in a theory of imprecise probabilities, and to put imprecision, related to information, in a central position. The current concept is not claimed to be the best, and the search for an updating rule that is as general as Walley's GBR remains the main research topic from a theoretical point of view. Evaluation of the current methodology will only be possible after implementation in real decision problems, where many open problems remain to be solved, some of which (especially on elicitation) explicitly need the interaction resulting from theory and practice. The concept developed in this thesis is a promising methodology for statistical modeling of expert opinions.

REFERENCES

Abel, P. and Singpurwalla, N.D. (1991). To survive or to fail?; that is the question. *Technical Report*, George Washington University, Washington.

Barlow, R.E. (1985). A Bayes explanation of an apparent failure rate paradox. *IEEE Transactions on Reliability 34*, 107-108.

Bayarri, M.J., DeGroot, M.H. and Kadane, J.B. (1988). What is the likelihood function (with discussion). In: S.S. Gupta and J.O. Berger (eds.) *Statistical Decision Theory and Related Topics IV*, 3-27 (Springer-Verlag, New York).

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York).

Berger, J.O. (1990). Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference 25*, 303-328.

Berger, J. and Berliner, L.M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *The Annals of Statistics 14*, 461-486.

Boole, G. (1854). *An Investigation of the Laws of Thought on which are founded the Mathematical Theories of Logic and Probabilities* (Macmillan, London; reprinted by Dover, 1958).

Chun, M. and Ahn, K. (1992). Assessment of the potential applicability of fuzzy set theory to accident progression event trees with phenomenological uncertainties. *Reliability Engineering and System Safety 37*, 237-252.

Colombo, A.G. and Costantini, D. (1980). Ground-hypotheses for beta distribution as Bayesian prior. *IEEE Transactions of Reliability R-29*, 17-20.

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

Coolen, F.P.A. (1992). Bounds for expected loss in Bayesian decision theory with imprecise prior probabilities. *Memorandum COSOR 92-40*, Eindhoven University of Technology.

Coolen, F.P.A. (1993). Imprecise conjugate prior densities for the one-parameter exponential family of distributions. *Statistics & Probability Letters 16*, 337-342.

Coolen, F.P.A. and Dekker, R. (1994). Analysis of a 2-phase model for optimization of condition monitoring intervals. *IEEE Transactions on Reliability*, to appear.

Coolen, F.P.A. and Newby, M.J. (1994). Bayesian estimation of location parameters in life distributions. *Reliability Engineering and System Safety*, to appear.

Dale, A.I. (1991). *A History of Inverse Probability* (Springer-Verlag, New York).

DasGupta, A. (1991). Diameter and volume minimizing confidence sets in Bayes and classical problems. *Annals of Statistics 19*, 1225-1243.

De Finetti, B. (1974). *Theory of probability* (2 volumes) (Wiley, London).

De Groot, M. (1974). Reaching a consensus. *Journal of the American Statistical Association 69*, 118-121.

Dempster, A.P. (1968). A generalization of Bayesian inference (with discussion). *Journal of the Royal Statistical Society series B 30*, 205-247.

DeRobertis, L. and Hartigan, J.A. (1981). Bayesian inference using intervals of measures. *Annals of Statistics 9*, 235-244.

Dubois, D. and Prade, H. (1992). On the relevance of non-standard theories of uncertainty in modeling and pooling expert opinions. *Reliability Engineering and System Safety 36*, 95-107.

Gärdenfors, P. and Sahlin, N.E. (1982). Unreliable probabilities, risk taking, and decision making. *Synthese 53*, 361-386.

Gärdenfors, P. and Sahlin, N.E. (1983). Decision making with unreliable probabilities. *British Journal of Mathematical and Statistical Psychology 36*, 240-251.

Geisser, S. (1993). *Predictive Inference: An Introduction* (Chapman and Hall, London).

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398-409.

Good, I.J. (1962). Subjective probability as the measure of a non-measurable set. In: E. Nagel, P. Suppes and A. Tarski (eds.) *Logic, Methodology and Philosophy of Science*, 319-329 (Stanford University Press).

Hills, S.E. and Smith, A.F.M. (1992). Parameterization issues in Bayesian inference (including discussion). In: J.M Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.) *Bayesian Statistics 4*, 227-246 (Oxford University Press).

Kullback, S. and Leibler, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics 22*, 79-86.

Laplace, P.S. de (1812). *Théorie analytique des probabilités* (Courcier, Paris).

Lavine, M. (1991a). Sensitivity in Bayesian statistics: The prior and the likelihood. *Journal of the American Statistical Association 86*, 396-399.

Lavine, M. (1991b).   An approach to robust Bayesian analysis for multidimensional parameter spaces.   *Journal of the American Statistical Association* *86*, 400–403.

Lee, P.M. (1989).   *Bayesian Statistics: An Introduction*   (Oxford University Press, New York).

Lindley, D.V. (1990).   The present position in Bayesian statistics (1988 Wald memorial lectures, with discussion).   *Statistical Science 5*, 44–89.

Lindley, D.V. (1992).   Discussion to Hills and Smith (1992).   In: J.M Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.) *Bayesian Statistics 4*, 241   (Oxford University Press).

Lindley, D.V. and Singpurwalla, N.D. (1986).   Reliability (and fault tree) analysis using expert opinions.   *Journal of the American Statistical Association 81*, 87–90.

Littlewood, B. (1991).   Forecasting software reliability.   In: P.C. Sander and R. Badoux (eds.) *Bayesian Methods in Reliability*, 135–201   (Kluwer Academic Publishers, Dordrecht).

Lourens, P.F. (1984).   The Formalization of Knowledge by Specification of Subjective Probability Distributions.   *Dissertation*, University of Groningen.

Martz, H.F. and Waller, R.A. (1982).   *Bayesian Reliability Analysis*   (Wiley, New York).

Meczarski, M. and Zielinski, R. (1991).   Stability of the Bayesian estimator of the Poisson mean under inexactly specified gamma prior.   *Statistics & Probability Letters 12*, 329–333.

Mendel, M.B. (1992).   Bayesian parametric models for lifetimes.   In: J.M Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.) *Bayesian Statistics 4*, 697–705   (Oxford University Press).

Moreno, E. and Cano, J.A. (1991).   Robust Bayesian analysis with ε-contaminations partially known.   *Journal of the Royal Statistical Society series B 53*, 143–155.

Moreno, E. and Pericchi, L.R. (1992).   Bands of probability measures: A robust Bayesian analysis.   In: J.M Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.) *Bayesian Statistics 4*, 707–713   (Oxford University Press).

Pericchi, L.R. and Walley, P. (1991).   Robust Bayesian credible intervals and prior ignorance.   *International Statistical Review 58*, 1–23.

Press, S.J. (1989).   *Bayesian statistics: Principles, Models, and Applications*   (Wiley, New York).

Raiffa, H. and Schlaifer, R. (1961).   *Applied Statistical Decision Theory*   (MIT Press, Cambridge, USA).

Ranneby, B. (1984). The maximum spacing method. An estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics 11*, 93-112.

Rao, S.S. and Dhingra, A.K. (1992). Reliability and redundancy apportionment using crisp and fuzzy multiobjective optimization approaches. *Reliability Engineering and System Safety 37*, 253-261.

Rodrigues, F.W. and Wechsler, S. (1993). A discrete Bayes explanation of a failure-rate paradox. *IEEE Transactions on Reliability 42*, 132-133.

Sander, P.C. and Badoux, R. (eds) (1991). *Bayesian Methods in Reliability* (Kluwer Academic Publishers, Dordrecht).

Singpurwalla, N.D. (1992). Discussion to Hills and Smith (1992). In: J.M Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.) *Bayesian Statistics 4*, 243 (Oxford University Press).

Smith, C.A.B. (1961). Consistency in statistical inference and decision (with discussion). *Journal of the Royal Statistical Society series B 23*, 1-25.

Spizzichino, F. (1992). Reliability decision problems under conditions of ageing. In: J.M Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.) *Bayesian Statistics 4*, 803-811 (Oxford University Press).

Suppes, P. and Zanotti, M. (1977). On using random relations to generate upper and lower probabilities. *Synthese 36*, 427-440.

Terlouw, P. (1989). Subjective Probability Distributions, a Psychometric Approach. *Dissertation*, University of Groningen.

Tijms, H.C. (1986). *Stochastic Modelling and Analysis: A computational Approach* (Wiley, New York).

Van Lenthe, J. (1993). ELI The Use of Proper Scoring Rules for Eliciting Subjective Probability Distributions. *Dissertation*, University of Groningen.

Vreugdenhil, H. (1993). Confidence in Knowledge. *Dissertation*, University of Amsterdam.

Walley, P. (1982). The elicitation and aggregation of beliefs. *Statistics Research Report 22*, University of Warwick, Coventry.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities* (Chapman and Hall, London).

Walley, P. and Campello de Souza, F.M. (1990). Uncertainty and indeterminacy in assessing the economic viability of energy options: A case study of solar heating systems in Brazil. *Energy Systems and Policy 14*, 281-304.

Walley, P. and Fine, T.L. (1982). Towards a frequentist theory of upper and lower probability. *Annals of Statistics 10*, 741-761.

Wasserman, L. (1992). Invariance properties of density ratio priors. *Annals of Statistics 20*, 2177-2182.

Wasserman, L. and Kadane, J.B. (1992). Symmetric upper probabilities. *Annals of Statistics 20*, 1720-1736.

Wolfenson, M. and Fine, T.L. (1982). Bayes-like decision making with upper and lower probabilities. *Journal of the American Statistical Association 77*, 80-88.

Zadeh, L.A. (1983). The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems 11*, 199-227.

Zen, M.M. and DasGupta, A. (1990). Estimating a binomial parameter: Is robust Bayes real Bayes? *Technical Report 90-63C*, Purdue University.

SUMMARY

The theory of imprecise probabilities, presented by Walley (1991), provides a concept for the quantification of opinions about uncertain events, to be used as source of information in decision problems. If event A is of interest, your opinion about it is measured through your desire to buy or sell a bet that pays 1 if A occurs and 0 otherwise. Your lower probability for event A, denoted by $\underline{P}(A)$, is the supremum of the prices for which you desire to buy this bet, because you expect a profit. Analogously, your upper probability $\overline{P}(A)$ for event A is the infimum of the prices for which you desire to sell this bet. Central to the theory is the assumption of coherent betting behavior, implying that you do not accept combinations of bets that lead to sure loss. One result of this assumption is $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$. The difference between your upper and lower probabilities is called (the degree of) imprecision for A, and defined by $\Delta(A) = \overline{P}(A) - \underline{P}(A)$.

The standard theory of subjective probability related to betting behavior, as presented by De Finetti (1974), is restricted to precise probabilities $P(A) = \underline{P}(A) = \overline{P}(A)$, leading to problems of the following kind. Suppose one coin is tossed, and you are totally convinced the coin is fair, so you regard both sides as equally likely to turn up. Then a second coin, which you know nothing about, is tossed. Restricted to precise probabilities this may lead you to assign $P(head) = 1/2$ in both cases, although it seems to be perfectly natural that your betting behavior is different for the second coin, with as most extreme possibility $\underline{P}(head) = 0$ and $\overline{P}(head) = 1$, implying that you do not expect profit when buying or selling the bet against any price. The restriction to precise probabilities, that are suitable to model stochastic uncertainty, leads to neglect of an important aspect of opinions, uncertainty by lack of information. In this thesis the possibility of relating the imprecision $\Delta(A)$ to the amount of information available is discussed.

The use of imprecise probabilities generalizes the standard Bayesian framework for statistics and decision making. Assuming a parametric model to describe the situation of interest, imprecision enters the model through a class of prior probability densities for the parameters. A theory for updating the priors in the light of new information is needed. Walley (1991) provides such a theory, based on a so-called Generalized Bayes Rule (GBR), but this has some serious drawbacks.

In this thesis a new updating theory for imprecise probabilities is proposed, related to the Bayesian theory and to Walley's GBR, but with control of impre-

cision based on the amount of information available. The measure of information, proposed by Walley (1991), is used after a theoretical foundation for it is provided. A general analysis of the updating theory is complemented with a more detailed study of Bernoulli experiments.

Generalization of Bayesian decision theory to allow imprecise prior probabilities is presented, with imprecision reflected by sets of values for the expected loss per decision. The calculation of the bounds for such sets takes only slightly more effort than the calculation of expected loss for precise probabilities. As an illustration a problem of a simple replacement policy for technical equipment is considered.

Finally, three important aspects for implementation, namely elicitation and combination of imprecise probabilities, and model fitting are discussed. The new theory needs to be evaluated by tests and applications to real decision problems, which has not been performed yet. The concept developed in this thesis is a promising methodology for statistical modeling of expert opinions.

SAMENVATTING

De theorie van intervalkansen, gepresenteerd door Walley (1991), draagt een concept aan voor het kwantificeren van meningen over onzekere gebeurtenissen, met als doel deze te gebruiken als informatiebron voor beslissingsproblemen. Als gebeurtenis A van belang is, wordt je mening hierover gemeten middels je verlangen een weddenschap te kopen of te verkopen, waarbij 1 uitbetaald wordt als A optreedt, en 0 als A niet optreedt. Je onderkans voor gebeurtenis A, genoteerd als $\underline{P}(A)$, is het supremum van de prijzen waarvoor je de weddenschap wilt kopen, omdat je verwacht te winnen. Analoog is je bovenkans, $\overline{P}(A)$, voor gebeurtenis A het infimum van de prijzen waarvoor je de weddenschap wilt verkopen. De theorie stoelt op de aanname dat je wedgedrag coherent is, hetgeen wil zeggen dat je geen combinaties van weddenschappen accepteert die zeker tot verlies lijden. Een gevolg van deze aanname is $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$. Het verschil tussen je boven- en onderkans wordt (de graad van) imprecisie voor A genoemd, en gedefinieerd als $\Delta(A) = \overline{P}(A) - \underline{P}(A)$.

De standaard subjectieve kanstheorie, gerelateerd aan wedgedrag als gepresenteerd door De Finetti (1974), beperkt zich tot enkelvoudige kansen $P(A) = \underline{P}(A) = \overline{P}(A)$, hetgeen tot problemen leidt van de volgende aard. Stel dat een munt wordt opgegooid en je bent er absoluut van overtuigd dat deze munt volstrekt eerlijk is, in de zin dat beide mogelijke uitkomsten, kopzijde boven of muntzijde boven, even waarschijnlijk zijn. Tevens wordt een tweede munt opgegooid waarvan je niets weet. Met de restrictie tot enkelvoudige kansen kan dit leiden tot $P(kopzijde\ boven) = 1/2$ voor beide munten, hoewel het logisch lijkt dat je wedgedrag voor de tweede munt hiervan afwijkt, met als extreme mogelijkheid $\underline{P}(kopzijde\ boven) = 0$ en $\overline{P}(kopzijde\ boven) = 1$, hetgeen impliceert dat er geen enkele prijs is waarvoor je winst verwacht bij kopen of verkopen van de weddenschap. De restrictie tot enkelvoudige kansen, die geschikt zijn voor het modelleren van stochastische onzekerheid, leidt tot verwaarlozing van een belangrijk aspect van meningen, namelijk onzekerheid door gebrek aan informatie. In dit proefschrift wordt de mogelijkheid besproken de imprecisie $\Delta(A)$ te relateren aan de hoeveelheid beschikbare informatie.

Het gebruik van intervalkansen generaliseert het standaard Bayesiaanse kader voor statistiek en besluitvorming. Als een parametrisch model wordt aangenomen ter beschrijving van een van belang zijnde situatie, wordt imprecisie in het model opgenomen middels een klasse van a priori kansdichtheden voor de parameters. Een theorie is gewenst voor het aanpassen van deze klasse aan nieuwe informatie. Walley (1991) voorziet in deze behoefte met een zogenoemde

Gegeneraliseerde Bayes Regel (GBR), deze heeft echter enkele serieuze nadelen.

In dit proefschrift wordt een nieuwe aanpassingstheorie voor intervalkansen voorgesteld, die gerelateerd is aan het Bayesiaanse concept en aan Walley's GBR, maar waarin bij aanpassing aan nieuwe gegevens de imprecisie gemanipuleerd wordt aan de hand van de hoeveelheid informatie. Hierbij wordt de informatiemaat gebruikt die door Walley (1991) gesuggereerd is, echter eerst nadat hiervoor een theoretisch argument aangedragen is. Een algemene analyse van de aanpassingstheorie wordt aangevuld met een meer gedetailleerde studie van Bernoulli experimenten.

Een generalisatie van Bayesiaanse beslissingstheorie wordt gepresenteerd die het mogelijk maakt *a priori* intervalkansen te gebruiken, waarbij de imprecisie resulteert in verzamelingen van waarden voor het verwachte verlies per beslissing. Het berekenen van de grenzen van dergelijke verzamelingen vergt slechts een geringe extra inspanning in vergelijking met het berekenen van het verwachte verlies voor enkelvoudige kansen. Ter illustratie wordt een beslissingsprobleem betreffende een eenvoudig vervangingsbeleid voor technische apparatuur besproken.

Tot slot worden drie belangrijke aspecten voor implementatie toegelicht, te weten ontlokking en combinatie van intervalkansen, en de aanpassing van een model. De nieuwe theorie kan slechts op waarde geschat worden door toepassing op echte beslissingsproblemen, hetgeen vooralsnog niet heeft plaatsgevonden. Het in dit proefschrift ontwikkelde concept is een veelbelovende methodiek voor statistische modellering van expert meningen.

Curriculum vitae

Frank Coolen werd geboren op 11 december 1965 te Tilburg. Van 1978 tot 1984 heeft hij het gymnasium β doorlopen aan het Mill-Hillcollege te Goirle. Aansluitend heeft hij wiskunde gestudeerd aan de Technische Universiteit Eindhoven, alwaar in december 1989 het doctoraal examen behaald werd bij de vakgroep Besliskunde & Stochastiek van de faculteit Wiskunde & Informatica, in de afstudeerrichting toegepaste statistiek en systeemtheorie, alsmede een aanvullend doctoraal examen in de stochastische besliskunde. Het afstudeerproject vond plaats onder leiding van prof.dr. P.C. Sander en dr. M.J. Newby van de faculteit Technische Bedrijfskunde en betrof een bedrijfszekerheidsanalyse voor warmtewisselaars, uitgevoerd bij de kunstmestfabriek Kemira bv. te Rozenburg.

Na een kort verblijf als Toegevoegd Docent Wiskunde en Statistiek bij de vakgroep Econometrie van de faculteit Economie aan de Katholieke Universiteit Brabant trad Frank Coolen in mei 1990 als Assistent In Opleiding in dienst bij de faculteit Wiskunde & Informatica aan de TU Eindhoven, en begon onder leiding van prof.dr. P. van der Laan zijn promotieonderzoek betreffende de statistische modellering van expert meningen. Tevens werd gedurende twee jaren een dag per week doorgebracht op het Koninklijke/Shell Laboratorium te Amsterdam ter verdieping van het inzicht in de problematiek van modellering en informatieverwerving bij onderhoudsplanning voor technische systemen. In oktober 1990 werd de module 'Bayesian Methods in Reliability' van het European Master's Programme in Safety and Reliability van ESRA/COMETT doorlopen.

De resultaten van het promotieonderzoek zijn vastgelegd in dit proefschrift.