The theory of imprecise probabilities: some results for
distribution functions, densities, hazard rates and hazard functions
Coolen, F.P.A.

Published: 01/01/1991

Document Version
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):
Memorandum COSOR 91-32
The Theory of Imprecise Probabilities:
Some Results for Distribution Functions,
Densities, Hazard Rates and Hazard Functions
F.P.A. Coolen

Eindhoven University of Technology
Department of Mathematics and Computing Science
P.O. Box 513
5600 MB Eindhoven
The Netherlands

Eindhoven, November 1991
The Netherlands
The Theory of Imprecise Probabilities: Some Results for Distribution Functions, Densities, Hazard Rates and Hazard Functions

F.P.A. Coolen
Eindhoven University of Technology
Department of Mathematics and Computing Science

Abstract

The theory of imprecise probabilities is discussed, and some relations are derived for lower and upper cumulative distribution functions and densities. To introduce the concept in the theory of reliability lower and upper hazard rates and hazard functions are defined. Also some aspects of updating lower and upper prior densities in light of new information are described.
1. Introduction

The theory of imprecise probabilities (see Walley [5]) provides a suitable concept for handling uncertainty and imprecision in expert systems. The most important difference between this concept and the (classical) frequentist or Bayesian concepts is that one does not need to assign one exact (precise) value to be the probability of a certain event, but one can assign two values, indicating the lower and upper probability for that event, so creating an interval of values instead of one value. If the lower probability of an event is equal to zero, and the upper probability equal to one, this indicates total ignorance about the event, whereas the other extreme situation, with equal lower and upper probabilities, indicates precision, meaning that one is absolutely sure about the probability of the event.

In practice, the assumption of precision is very strong, and if assignment of probabilities has to be based on elicited opinions of people, imprecision is quite natural. Also if information from several sources has to be combined, the dogma of precision does not agree with reality.

Walley [6] mentions six broad criteria for evaluating measures of uncertainty. These are:

(a) Interpretation. The measure should have a clear interpretation that is sufficiently definite to be used to guide assessment, to understand the conclusions of the system (and use them as a basis for action), and to support the rules for combining and updating measures.

(b) Imprecision. The measures should be able to model partial or complete ignorance, limited or conflicting information, and imprecise assessments of uncertainty.

(c) Calculus. There should be rules for combining measures of uncertainty, updating them after receiving new information, and using them to calculate other uncertainties, to draw conclusions and to make decisions. Some justification must be given for the rules.

(d) Coherence. There should be methods for checking the consistency (coherence) of all uncertainty assessments used by the system, and the rules should ensure that the conclusions are consistent with these assessments. (That is, it should not be possible to combine the measures of uncertainty to produce irrational inferences or behaviour.)
Assessment. It should be practicable for a user of the system to make
(and feel comfortable with) all the uncertainty assessments that are needed
as input. The system should give some guidance on how to make the
assessments. It should be able to handle judgements of various types,
including expressions of uncertainty in natural language such as "if A then
probably B", and to combine qualitative judgements with quantitative
assessments of uncertainty.

Computation. It should be computationally feasible for the system to
derive inferences and conclusions from the assessments.

Walley [6] compares several measures of uncertainty (amongst others additive
probabilities, as used in the Bayesian theory), and concludes that only
lower and upper probabilities (or more general, previsions) are adequate as
general measures of uncertainty. Although, in this theory, the nature of
probability is subjective, the concept of imprecise probabilities can also
be used in a frequentist context, as described by Walley and Fine [7].

Elicitation of the opinions of experts is important, as these opinions are
often an important source of information. It is important to emphasize the
difference between elicitation of these opinions, and assessment of your
probabilities, based on these opinions and perhaps other information. For
assessment matters like combination of information from several sources
(exerts) and coherence are important. Elicitation is the process by which
beliefs are measured, through explicit judgements and choices.

The concept of imprecise probabilities allows a reasonable translation of
expressions of uncertainty in natural language into lower probabilities.
Walley [6] gives examples, such as "an event A is probably (or likely)
implies \( P(A) \approx 0.5 \), and "A has very high probability implies \( P(A) \geq 0.9 \). Here
\( P(A) \) denotes the lower probability of event A, that can be interpreted as a
lower bound of an interval of values for this probability, between which no
further distinction is made based on the available information.

The resulting imprecision seems to be natural, although the choices of the
lower probabilities are still subjective. Translation of such statements the
way we indicate here is, however, surely less doubtful than by translating
such expressions into precise probabilities, which is often done in
elicitation processes within practical Bayesian statistics.

There are many possible methods to elicit lower probabilities of events. In
reliability it is most interesting to work out practical elicitation methods for deriving lower and upper densities for lifetime distributions. These are functions that bound an area such that every function between them can, after normalisation, be regarded as a possible probability density function. Practical elicitation could, for example, be executed by using histograms. In a later report we will look at the practical possibilities for elicitation, when working with the concept of imprecise probabilities. There it will be shown that the concept of imprecise probabilities has advantages to precise probabilities when working with histograms for eliciting information, but with some parametrical family of distributions as the model (e.g. for calculations). In such case one always has many possibilities to create a certain method for choosing a member of the family that looks the most like the histogram, but this suggests precision which is not grounded by the available information.

In this report imprecise probabilities are briefly introduced in section 2, and lower and upper cumulative distribution functions and probability density functions are discussed in section 3. In section 4 we define the hazard rate and hazard function, and show that these definitions are consequent to the theory. Finally, in section 5 updating of lower and upper prior densities is discussed, and some conclusions and interesting subjects for future research are presented in section 6.
2. Imprecise probabilities

Instead of choosing one value to be the probability of occurrence of a certain event A, two values are given that bound an interval, that can be interpreted as a set of values that can all possibly be the unknown probability of A, and between which no further distinction is possible or desirable based on the available information. These bounds are called the lower and upper probability of the event A, denoted by \( \underline{P}(A) \) and \( \overline{P}(A) \) respectively, and the obvious relation \( 0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1 \) must hold (see the axioms and consequences below). A possible interpretation is to accept that the occurrence of event A has some probability \( P(A) \), that follows the usual axioms of probability (B1-B3 below), but to you the exact value of \( P(A) \) is unknown. However, based on the available information you are sure that \( \underline{P}(A) \leq P(A) \leq \overline{P}(A) \). Of course, in practice exact probabilities are hardly ever known, so the idea of choosing an interval seems logical. It is also clear that the bounds \( \underline{P}(A) \) and \( \overline{P}(A) \) are subjective, as these will depend on you and your information. Here the nature of \( P(A) \) can also be subjective, but other interpretations are also possible. The main idea is that \( P(A) \) is not exactly known to you.

It is useful, although the above interpretation is suitable for this report, to shortly discuss the subjective nature of probability, as advocated by De Finetti [2], and adopted by Walley [5], who really needs this interpretation to develop his more complete theory of imprecise previsions. Here your probabilities are related to your betting behaviour.

Suppose that a gamble is defined such that you receive 1 dollar if event A occurs, and nothing otherwise. (To overcome the fact that, to most people, the utility of money is not a linear function of its amount, Walley [5] introduces so-called probability currency.) Your lower probability of A is then interpreted as the upper bound of prices that you are willing to pay for this gamble (your supremum buying price, you actually may not want to pay \( \underline{P}(A) \), but you are willing to pay \( \underline{P}(A) - \varepsilon \) for any \( \varepsilon > 0 \)), and your upper probability of A as the amount of money \( \overline{P}(A) \) such that you are willing to sell the above gamble to someone else for any price greater than \( \overline{P}(A) \) (your infinimum selling price).
This interpretation can be related to the first one that we adopt here. There the lower and upper probabilities are interpreted as bounds, such that you know sure that \( P(A) \leq \bar{P}(A) \leq \tilde{P}(A) \). Related to the second interpretation, this would mean that indeed you would be willing to pay up to \( \tilde{P}(A) \) for the gamble described above, and in fact we need such an interpretation to make clear what we mean by 'you are sure' in the first interpretation.

Your lower and upper probabilities should reflect the amount of information available to you, and from this point of view the possibility that \( \tilde{P}(A) \neq \bar{P}(A) \) is quite logical. Within the standard Bayesian framework (see De Finetti [2]), always \( \bar{P}(A) = \tilde{P}(A) \) is assumed, leaving no method to take the amount of available information into account.

We define the degree of imprecision about \( A \) as \( \Delta(A) = \tilde{P}(A) - \bar{P}(A) \). If \( \Delta(A) = 0 \) the probabilities for \( A \) are called precise, whereas the probabilities in the other extreme situation, with \( \bar{P}(A) = 0 \) and \( \tilde{P}(A) = 1 \), are called vacuous. This last situation perfectly represents total absence of information about \( A \), while precise probabilities represent perfect information. It seems logical that \( \Delta(A) \) should be a decreasing function of the amount of relevant information.

A thorough discussion of the importance of imprecision is provided by Walley [5, chapter 5]. As a simple example think about tossing a coin. Suppose you are interested in the event \( H \) that a coin will fall 'head up' when tossed. The knowledge that the coin is fair and so is the way of tossing could lead you to assign precise probabilities \( \bar{P}(H) = \tilde{P}(H) = 0.5 \), meaning that you are absolutely sure that the probability \( P(H) \) is 0.5. Using the second interpretation, this would mean that your supremum buying price and infimum selling price of the gamble, such that one dollar is paid if \( H \) occurs and nothing otherwise, are equal. But even in this situation the equality is \( \bar{P}(H) = \tilde{P}(H) \) not necessary, as you still may have some doubts. However, if you have no knowledge at all about the coin, or about the way it is tossed, the assumption that you know that \( P(H) = 0.5 \) (or alternatively, that you would be willing to pay 0.5 dollar for this bet) is too strong. This complete lack of information could even lead you to choose vacuous probabilities (in the second interpretation, to reject the gamble if you have to pay any positive sum of money for it).
This example makes clear that lack of total knowledge about events and about the probability of occurrence of these events, forces you to enable the use of imprecision within the concept of probability. Furthermore, the concept of imprecise probabilities can be used to reflect the amount of information that is available. In many situations this is very important, for example, if a person who has to make a decision asks the opinions of several experts, using the concept of probabilities to express their uncertainty.

Let $\Omega$ be the set of all possible events of interest, then $P$ and $\overline{P}$ are assumed to satisfy the following basic axioms (according to Wolfenson and Fine [8]):

(A1) For all $A \in \Omega$: $P(A) \geq 0$.
(A2) $P(\Omega) = 1$.
(A3) For all $A, B \in \Omega$, with $A \cap B = \emptyset$: $P(A) + P(B) \leq P(A \cup B)$ and $\overline{P}(A) + \overline{P}(B) \geq \overline{P}(A \cup B)$.
(A4) For all $A \in \Omega$: $\overline{P}(A) + P(A^c) = 1$, where $A^c$ is the complement of $A$.

These axioms result from the interpretation of $P(A)$ and $\overline{P}(A)$, as sure bounds for the unknown precise probability $P(A)$. According to the well-known theory of precise probabilities, $P$ must satisfy the axioms:

(B1) For all $A \in \Omega$: $P(A) \geq 0$.
(B2) $P(\Omega) = 1$.
(B3) For all $A, B \in \Omega$, with $A \cap B = \emptyset$: $P(A \cup B) = P(A) + P(B)$.

(We do not discuss the problem whether or not the additivity axiom B3 should hold for only finite numbers of events, or also for countably infinite numbers. For such discussion, see for example De Finetti [2], who is against countable additivity, or Walley [5, section 6.9], who accepts countable additivity. We accept Walley's point of view.)

An important behavioural assumption is that you do not choose the lower (upper) probability smaller (greater) than a certain value that is surely known not to be greater (smaller) than $P(A)$.
This assumption, together with B1 and B2, straightforwardly lead to axioms A1 and A2.

Axiom A3 follows from B3, by \( P(A \cup B) = P(A) + P(B) \geq P(A) + P(B) \) and \( P(A \cup B) = P(A) + P(B) - P(A) + P(B) \), so the above assumption leads to A3. As a simple example in which \( P(A \cup B) > P(A) + P(B) \) again look at the example we used earlier, about tossing a coin. Suppose you have no information about the coin, but have the results of some earlier tosses, and you assign \( \bar{H} = 0.3 \) and \( \bar{H^C} = 0.2 \), where \( H^C \) means that the coin falls 'tail up'. Because \( H \cup H^C = \Omega \) and \( H \cap H^C = \emptyset \) it is obvious that, according to A2, \( P(H \cup H^C) = 1 \), while \( P(H) + P(H^C) = 0.5 \).

Finally, axiom A4 is proved by using B2 and B3 that lead to \( P(A) + P(A^C) = 1 \). We know that \( P(A) \leq P(A) \), so \( 1 - P(A^C) \leq P(A) \) and this leads to \( P(A^C) \leq 1 - P(A) \). The above assumption gives \( \bar{P(A^C)} \leq 1 - P(A) \). Analogously, we know that \( P(A^C) \leq \bar{P}(A^C) \), so \( 1 - P(A) \leq \bar{P}(A^C) \) and this leads to \( P(A) \leq 1 - \bar{P}(A^C) \). Using the above assumption we derive at \( P(A) \leq 1 - \bar{P}(A^C) \), so \( \bar{P(A^C)} \leq 1 - P(A) \). The two results \( \bar{P}(A^C) \leq 1 - P(A) \) and \( P(A^C) \leq 1 - \bar{P}(A) \) imply A4.

Elementary consequences of the axioms A1-A4 are

(C1) \( P(\emptyset) = \bar{P}(\emptyset) = 0 \) (use axioms A2, A3 and A4).

(C2) \( \bar{P}(\Omega) = 1 \) (use A4 and C1).

(C3) \( P(A) \leq \bar{P}(A) \) (A3 and A4).

(C4) \( A \subseteq B \) implies \( P(B) \leq P(A) \) and \( \bar{P}(B) \leq \bar{P}(A) \).

(C5) \( A \cap B = \emptyset \) implies \( P(A \cup B) \leq P(A) + P(B) \).

Walley [5] provides the complete theory, starting from the interpretation that your probabilities relate to your betting behaviour. The theory is more
general, and the reasoning that leads to the axioms that correspond to A1-A4 is quite long-winded. This is avoided in this report by using the first mentioned interpretation. Nevertheless, interested readers should feel themselves challenged to read the book by Walley [5], and to study the arguments within his theory. We regard this book as a very solid foundation for the theory of imprecise previsions, with imprecise probabilities as an important special case.
3. Distribution function and density

Special cases of imprecise probabilities are lower and upper cumulative distribution functions (cdf) for a real variable X. These are the lower and upper probabilities of the events \( X \leq x \) for \( x \in \mathbb{R} \), and are denoted by \( F_l(x) = P(X \leq x) \) and \( F_u(x) = P(X \leq x) \).

A suitable method to elicit a person's opinion about X is to ask him to assign two functions, say \( l(x) \) and \( u(x) \) with \( 0 \leq l(x) \leq u(x) \) for all \( x \), such that all functions between \( l \) and \( u \) can, after normalisation, be regarded as possible probability density functions (pdf) for X. This method is discussed by Walley [5, section 4.6.4] and by DeRobertis and Hartigan [3]. The functions \( l \) and \( u \) are called lower and upper density functions. We assume, in this report, that \( l \) and \( u \) have finite integrals (although theoretically this is not necessary), and that \( l \) and \( u \) are continuous functions. This second assumption could be generalised by assuming that both may have a countable number of discontinuity points.

First we show how to construct lower and upper cdf's that relate to lower and upper density functions.

**Lemma 1**

The lower and upper cdf's, defined by

\[
F_l(x) = \frac{\int_{-\infty}^{x} l(\omega) \, d\omega}{\int_{-\infty}^{\infty} l(\omega) \, d\omega + \int_{-\infty}^{\infty} u(\omega) \, d\omega} \quad \text{and} \quad F_u(x) = \frac{\int_{-\infty}^{x} u(\omega) \, d\omega}{\int_{-\infty}^{\infty} l(\omega) \, d\omega + \int_{-\infty}^{\infty} u(\omega) \, d\omega}
\]

are the bounds of all cdf's that can be constructed from densities (after normalisation) that lie between \( l \) and \( u \).

**Proof**

These relations follow from formulas presented by Walley [5, section 4.6]:

Let \( A \subseteq \mathbb{R} \), then \( P(x \in A) = \frac{\int_{A} l(\omega) \, d\omega}{\int_{-\infty}^{\infty} l(\omega) \, d\omega + \int_{A} u(\omega) \, d\omega} \) and

\[
P(x \in A^c) = \frac{\int_{A^c} l(\omega) \, d\omega + \int_{A} u(\omega) \, d\omega}{\int_{-\infty}^{\infty} l(\omega) \, d\omega + \int_{A} u(\omega) \, d\omega}
\]
Some simple relations between $l, u$ and $F, \overline{F}$ follow from this lemma:

**Corollary 1**

If for all $x$: $l(x)=0$ and $u(x)>0$, then $F(x)=0$ and $\overline{F}(x)=1$.

If for all $x$: $l(x)=u(x)=0$, then the lower and upper cdf's are not defined by the above forms, and the problem is degenerated.

If for a $x_0$: $l(x)=0$ for all $x\geq x_0$ and $u(x)=0$ for all $x\leq x_0$, then $F(x_0)$ is not defined by the above form.

If for a $x_1$: $u(x)=0$ for all $x\leq x_1$ and $l(x)=0$ for all $x\geq x_1$ (so $l(x)=0$ for all $x$), then $\overline{F}(x_1)$ is not defined by the above form.

**Proof**

These remarks follow easily from lemma 1. $\Box$

To avoid problems as indicated in corollary 1, from now on we restrict to $l$ and $u$ having positive integrals, meaning that both are positive for some $x$.

This is reasonable, as $l(x)=0$ for all $x$ leads either to incorrect definitions or to vacuous distributions ($F(x)=0$ and $\overline{F}(x)=1$ for all $x$), depending on $u$. Of course, if $l$ has positive integral, also $u$ has, and if $u$ has finite integral, also $l$ has.

According to the following corollary, the situation of precision can be expressed by equality of $l$ and $u$ as well as by equality of $F$ and $\overline{F}$.

**Corollary 2**

$l(x) = u(x)$ for all $x$ $\Leftrightarrow$ $F(x) = \overline{F}(x)$ for all $x$.

**Proof**

This result follows by rewriting

$$F(x) = \left(1 + \frac{\int_{-\infty}^{x} u(\omega) \, d\omega}{\int_{-\infty}^{x} l(\omega) \, d\omega}\right)^{-1} \quad \text{and} \quad \overline{F}(x) = \left(1 + \frac{\int_{-\infty}^{x} l(\omega) \, d\omega}{\int_{-\infty}^{x} u(\omega) \, d\omega}\right)^{-1}.$$

From this forms it is also easy to verify that $F(x) \leq \overline{F}(x)$. $\Box$
The following two lemmas can, together, also be regarded as proof of lemma 1. Lemma 2 states that every cdf that results from a pdf between 1 and \( u \), after normalisation, lies between \( F \) and \( \bar{F} \), and that these lower and upper cdf's result themselves from a density between 1 and \( u \) if and only if \( l=u \), so if there is precision. Lemma 3 shows that it is not possible to reach sharper bounds.

**Lemma 2**

(a) If a density \( g \) lies between 1 and \( u \), so \( 1(\omega) \leq g(\omega) \leq u(\omega) \) for all \( \omega \), and

\[
c = \int_{-\infty}^{\infty} g(\omega) d\omega,
\]

then the cdf that corresponds to the pdf \( g(\omega)/c \),

\[
G(x) = \int_{-\infty}^{x} \frac{g(\omega)}{c} d\omega,
\]

lies between \( F \) and \( \bar{F} \), so \( F(x) \leq G(x) \leq \bar{F}(x) \) for all \( x \).

(b) There is a density \( g \) between 1 and \( u \) such that its corresponding cdf \( G \) is equal to \( F \) (\( F(x)=G(x) \) for all \( x \)) ⇔ \( l(x)=u(x) \) for all \( x \) ⇔ \( \bar{F}(x)=G(x) \) for all \( x \).

**Proof**

(a) \( F(x) = \frac{\int_{-\infty}^{x} 1(\omega) c d\omega}{\int_{-\infty}^{\infty} 1(\omega) c d\omega} \leq \frac{\int_{-\infty}^{x} 1(\omega) c d\omega}{\int_{-\infty}^{\infty} 1(\omega) c d\omega} \leq \frac{\int_{-\infty}^{\infty} g(\omega) d\omega}{\int_{-\infty}^{\infty} g(\omega) d\omega} = G(x) \leq \frac{\int_{-\infty}^{\infty} u(\omega) c d\omega}{\int_{-\infty}^{\infty} u(\omega) c d\omega} \leq \frac{\int_{-\infty}^{\infty} u(\omega) c d\omega}{\int_{-\infty}^{\infty} u(\omega) c d\omega} = \bar{F}(x) \).

(b) Let \( F(x)=G(x) \) for all \( x \). Now it is obvious that, in the proof of (a), the first two \( \leq \) both need to be equalities for all \( x \), and this straightforwardly leads to \( l(x)=g(x)=u(x) \) for almost all \( x \) (if \( g \) does not need to be continuous, it may differ from 1 and \( u \) in a countable
number of points \( x \). The continuity assumption for \( l \) and \( u \) leads to \( l(x) = u(x) \) for all \( x \).

If \( l = u \), there is only one possible density \( g \), and the result is obvious.

The second 'if and only if'-relation is proved analogously. \( \square \)

**Lemma 3**

The lower and upper cdf's \( F \) and \( \bar{F} \), given in lemma 1, are the sharpest bounds for all cdf's that can be constructed from densities between \( l \) and \( u \).

**Proof**

It is enough to show that, for each \( x_0 \), the value \( F(x_0) \) is equal to \( G(x_0) \) for some cdf \( G \) that relates to a density \( g \) between \( l \) and \( u \).

This is simply done by defining \( g(x) = l(x) \) for \( x \leq x_0 \), and \( g(x) = u(x) \) for \( x > x_0 \).

Note that this density is not continuous. However, for every \( \epsilon > 0 \) we can also find a continuous density \( g \) between \( l \) and \( u \) such that \( G(x_0) < F(x_0) + \epsilon \). This can be done by defining \( g(x) = l(x) \) for \( x \leq x_0 - \delta \), \( g(x) = u(x) \) for \( x \geq x_0 + \delta \), and \( g(x) \) increasing linearly from \( l(x_0 - \delta) \) to \( u(x_0 + \delta) \) on the interval \( [x_0 - \delta, x_0 + \delta] \), with \( \delta > 0 \) depending on \( \epsilon \).

It is interesting to regard relationships between \( F, \bar{F} \) and the cdf's that correspond to \( l \) and \( u \), after normalisation.

**Corollary 3**

Let \( F_l(x) = \frac{-\int_{-\infty}^{x} l(\omega) \, d\omega}{\int_{-\infty}^{\infty} l(\omega) \, d\omega} \) and \( F_u(x) = \frac{-\int_{-\infty}^{x} u(\omega) \, d\omega}{\int_{-\infty}^{\infty} u(\omega) \, d\omega} \), then \( F(x) \leq F_l(x) \leq \bar{F}(x) \) and \( \underline{F}(x) \leq F_u(x) \leq \bar{F}(x) \) for all \( x \).

Furthermore, \( F = F_l \Leftrightarrow F_l = \bar{F} \Leftrightarrow F = F_u \Leftrightarrow F = \bar{F} \Leftrightarrow l = u \), where equality of functions means equality of their values for all \( x \).

**Proof**

Because the normalised \( l \) and \( u \) are allowed pdf's themselves, the first results follow from lemma 2a.

The relation \( F = F_l \Leftrightarrow l = u \) can be verified easily by comparing the above form of \( F_l \) to \( F \) as presented in lemma 1, and the other three equivalent
statements are verified analogously (continuity of 1 and u is needed). In fact, these relations also follow from lemma 2b.

The lower and upper cdf's, constructed from the lower and upper densities as above, are said to have pdf's \( f \) and \( \overline{f} \) respectively. So

\[
\begin{align*}
\underline{f}(x) &= \frac{d\underline{F}(x)}{dx} = \frac{1(x) \int_{-\infty}^{x} u(\omega)d\omega + u(x) \int_{x}^{\infty} l(\omega)d\omega}{\left(\int_{-\infty}^{x} l(\omega)d\omega + \int_{x}^{\infty} u(\omega)d\omega\right)^2}, \\
\overline{f}(x) &= \frac{d\overline{F}(x)}{dx} = \frac{u(x) \int_{-\infty}^{x} l(\omega)d\omega + l(x) \int_{x}^{\infty} u(\omega)d\omega}{\left(\int_{-\infty}^{x} u(\omega)d\omega + \int_{x}^{\infty} l(\omega)d\omega\right)^2}.
\end{align*}
\]

In the following corollary we derive a useful relation between 1, \( f \) and \( u \).

**Corollary 4**

Let \( c(x) = \int_{-\infty}^{x} l(\omega)d\omega + \int_{x}^{\infty} u(\omega)d\omega \), then \( 1(x) \leq c(x) \underline{f}(x) \leq u(x) \) for all \( x \).

Furthermore, let \( x_0 = \inf\{x | l(x) > 0\} \) (\( x_0 \) may be \( -\infty \)), then there are two possible situations, namely \( l(x_0) = 0 \) or \( l(x_0) > 0 \).

If \( l(x_0) = 0 \), then: \( l(x) = c(x) \underline{f}(x) \) for all \( x \Rightarrow l(x) = u(x) \) for all \( x \geq x_0 \).

If \( l(x_0) > 0 \), then: \( l(x) = c(x) \underline{f}(x) \) for all \( x \Rightarrow l(x) = u(x) \) for all \( x \geq x_0 \).

For the second inequality we have:

\( c(x) \underline{f}(x) = u(x) \) for all \( x \Rightarrow l(x) = u(x) \) for all \( x \).

Analogous relations hold for \( \overline{f}(x) \).

**Proof**

The first relation is clear from the fact that

\[ c(x) \underline{f}(x) = w_1(x) l(x) + w_2(x) u(x), \]

with \( w_1(x) \) and \( w_2(x) \) both non-negative, and \( w_1(x) + w_2(x) = 1 \) for all \( x \).

The other relations follow straightforwardly by comparing \( c(x) \underline{f}(x) \) to \( l(x) \) and \( u(x) \) respectively, remembering the assumption of continuity of \( l \) and \( u \).
Obviously, continuity of \( l \) and \( u \) leads to continuity of \( F, \bar{F}, F \) and \( \bar{F} \).

Because both \( f \) and \( \bar{f} \) are pdf's, the relation \( f(x) \leq \bar{f}(x) \) for all \( x \) would imply \( f(x) = \bar{f}(x) \) for all \( x \), representing precision.

The following corollary is an important result of lemma 2b.

**Corollary 5**

There is a density \( g \) between \( l \) and \( u \), such that \( f \) is equal to \( g \) after normalisation \( \Rightarrow l(x) = u(x) \) for all \( x \).

A same relation holds for \( \bar{f} \).

**Proof**

This follows immediately from lemma 2b, and the fact both \( F \) and \( f \) are continuous.

This simply means that, unless there is precision, the pdf \( f \) is not a normalised version of a density that lies between \( l \) and \( u \).

The above expressions for \( f(x) \) and \( \bar{f}(x) \) lead to

\[
\begin{align*}
\bar{f}(x) &= \frac{l(x) [1 - F(x)] + u(x) F(x)}{\int_{-\infty}^{x} l(\omega) d\omega + \int_{x}^{\infty} u(\omega) d\omega} \quad \text{and} \quad \bar{f}(x) = \frac{u(x) [1 - \bar{F}(x)] + l(x) \bar{F}(x)}{\int_{-\infty}^{x} u(\omega) d\omega + \int_{x}^{\infty} l(\omega) d\omega}.
\end{align*}
\]

If \( u(x) = l(x) \), then \( \bar{f}(x) = \bar{f}(x) = \frac{l(x)}{\int_{-\infty}^{x} l(\omega) d\omega} \).

If \( u(x) > l(x) \), then \( \bar{f}(x) = \frac{f(x) \left( \int_{-\infty}^{x} l(\omega) d\omega + \int_{x}^{\infty} u(\omega) d\omega \right) - l(x)}{u(x) - l(x)} \quad \text{and} \quad \bar{f}(x) = \frac{u(x) - \bar{f}(x) \left( \int_{-\infty}^{x} u(\omega) d\omega + \int_{x}^{\infty} l(\omega) d\omega \right)}{u(x) - l(x)}.
\]

From these forms, but also directly from lemma 1, the relations in the following corollary can be proved easily (remember the assumption of continuity of \( l \) and \( u \)).
Corollary 6

\[ F(x_0) = 0 \iff 1(x) = 0 \text{ for all } x \leq x_0; \]

\[ F(x_1) = 0 \iff u(x) = 0 \text{ for all } x \leq x_1 (\text{also } 1(x) = 0 \text{ for all } x \leq x_1); \]

\[ F(x_2) = 1 \iff u(x) = 0 \text{ for all } x \geq x_2 (\text{also } 1(x) = 0 \text{ for all } x \geq x_2); \]

\[ F(x_3) = 1 \iff 1(x) = 0 \text{ for all } x \geq x_3. \]

Proof

These relations follow from lemma 1, and the assumption that 1 and \( u \) both have positive, finite integrals. \( \square \)
4. Hazard rate and hazard function

Within reliability theory, the reliability function, hazard rate and hazard function are important characteristics of probability distributions, where the hazard rate provides a natural description of a failure process. Here we propose definitions of lower and upper hazard rates and hazard functions, that are consequent to the above theory of imprecise probabilities. We also derive some simple relations. We restrict to non-negative real random variables.

Definitions

Suppose continuous lower and upper densities $l$ and $u$ are specified, and the according functions $F$, $\bar{F}$, $f$ and $\bar{f}$ are derived. The lower and upper Reliability Functions are $R(x) := 1 - \bar{F}(x)$ and $\bar{R}(x) := 1 - F(x)$.

The lower and upper hazard rates are $lhr(x) := \frac{l(x)}{1 - F(x)}$ and $uhr(x) := \frac{u(x)}{1 - F(x)}$.

The lower and upper hazard functions are $H(x) := -\ln(R(x)) = -\ln(1 - F(x))$ and $\bar{H}(x) := -\ln(\bar{R}(x)) = -\ln(1 - \bar{F}(x))$.

It must be remarked that $lhr$ and $uhr$ are not real hazard rates, that means they are not hazard rates according to some probability distribution. This is equivalent to the fact that $l$ and $u$ are not pdf's. This might be a problem for interpretation of $lhr$ and $uhr$.

Also the first derivatives of the hazard functions play a role. These are $h(x) = \frac{dH(x)}{dx} = \frac{f(x)}{1 - F(x)}$ and $\bar{h}(x) = \frac{d\bar{H}(x)}{dx} = \frac{\bar{f}(x)}{1 - \bar{F}(x)}$.

The functions $h$ and $\bar{h}$ relate obviously to $f$ and $\bar{f}$, and are real hazard rates. It is important to notice that $h(x) \leq \bar{h}(x)$ does not necessarily hold.

It is clear that $l(x) = u(x)$ for all $x$, which implies precision, is equivalent to each one of the following equalities (compare corollary 2):

$lhr(x) = uhr(x)$, $h(x) = \bar{h}(x)$ and $H(x) = \bar{H}(x)$ for all $x$. In case of precision the following relation holds: $\frac{lhr(x)}{h(x)} = \int_{-\infty}^{\infty} l(\omega) d\omega$. 
Continuity of \( l \) and \( u \) also implies continuity of \( \ln lhr \), \( \ln uhr \), \( \ln H \), \( \ln h \) and \( \ln \bar{H} \).

From the definitions it follows that both \( \ln H \) and \( \ln \bar{H} \) are non-decreasing.

The following relations hold:
\[
\begin{align*}
\ln H(x) &= \ln(lhr(x)) - \ln(1(x)), \quad \ln \bar{H}(x) = \ln(uhr(x)) - \ln(u(x)); \\
\ln F(x) &= 1 - \exp(-\ln H(x)), \quad \ln \bar{F}(x) = 1 - \exp(-\ln \bar{H}(x)); \\
\frac{\ln f(x)}{\ln h(x)} &= \frac{l(x)}{lhr(x)}, \quad \frac{\ln \bar{f}(x)}{\ln \bar{h}(x)} = \frac{u(x)}{uhr(x)}.
\end{align*}
\]

All functions of interest can be written as functions of \( l \) and \( u \). \( F \) and \( f \) are already given (and also the according upper cdf and pdf). Furthermore we have (the proofs are straightforward):
\[
\begin{align*}
lhr(x) &= \frac{1}{x} \left( \int_0^x l(\omega) \, d\omega + \int_x^\infty u(\omega) \, d\omega \right), \\
R(x) &= \frac{x}{x} \left( \int_0^x l(\omega) \, d\omega \right), \\
h(x) &= \frac{1}{x} \left( \int_0^x u(\omega) \, d\omega + \int_x^\infty l(\omega) \, d\omega \right) \quad \text{provided} \ u(x) > l(x), \quad \text{and} \\
\ln H(x) &= \ln \left( \int_0^x l(\omega) \, d\omega + \int_x^\infty u(\omega) \, d\omega \right) - \ln \left( \int_0^x u(\omega) \, d\omega \right).
\end{align*}
\]

Analogous relations for the upper functions are easily derived.

It is important to show that, given \( l \) and \( u \), the definitions of the lower hazard rate and hazard function are correct, in the way that the hazard rates and hazard functions that relate to densities between \( l \) and \( u \), indeed lie between \( lhr \) and \( uhr \), or \( \ln H \) and \( \ln \bar{H} \), respectively.

**Corollary 7**

Let \( g \) be a density that lies between \( l \) and \( u \), \( c = \int_{-\infty}^\infty g(\omega) \, d\omega \), and let the cdf that corresponds to the pdf \( g(\omega)/c \) be \( G(x) = \int_{-\infty}^x \frac{g(\omega)}{c} \, d\omega \). The according (precise) hazard rate is defined by \( \ln hr_g(x) = \frac{g(x)}{1-G(x)} \), and the (precise) hazard
function by \( H_g(x) = -\ln(1-G(x)) \).

Then \( l_h(x) \leq h_r(x) \leq u_h(x) \), and \( H(x) \leq H_g(x) \leq \overline{H}(x) \) for all \( x \).

**Proof**

These inequalities follow immediately from \( l(\omega) \leq g(\omega) \leq u(\omega) \) for all \( \omega \), and lemma 2. \( \square \)

Another simple, but important result is given in corollary 8.

**Corollary 8**

\( l_h(x) \leq c(x)h(x) \leq u_h(x) \), with \( c(x) \) as in corollary 4.

An analogous relation holds for \( \overline{h}(x) \).

**Proof**

This follows immediately from \( l(x) \leq c(x)\_f(x) \leq u(x) \) and \( F(x) \leq \overline{F}(x) \). \( \square \)

In the following lemma a useful property of the hazard function, that is well-known in the precise case, is generalised (we assume that the inverse functions of \( H \) and \( \overline{H} \) exist).

**Lemma 4**

If \( X \) has a standard exponential distribution, then \( H^{-1}(X) \) has distribution with cdf \( F \), and \( \overline{H}^{-1}(X) \) with cdf \( \overline{F} \).

**Proof**

\[ P(H^{-1}(X) \leq x) = P(X \leq H(x)) = 1 - \exp(-H(x)) = F(x), \] where it is used that \( H \) is non-decreasing. The proof for \( \overline{H}^{-1}(X) \) is analogous. \( \square \)
5. Updating

In the introduction to [5] Walley writes: "This book is concerned with probabilistic reasoning, which involves various methods for assessing imprecise probabilities, modifying the assessments where necessary to achieve coherence, updating them to take account of new information, and combining them to calculate other probabilities and to draw conclusions. We are especially concerned with the application of probabilistic reasoning in statistical problems which we call statistical reasoning, where the goal is to draw conclusions about statistical parameters or future occurrences."

The approach is to assess imprecise prior probabilities and combine these with the statistical data through a (generalised) Bayes' rule. The resulting posterior probabilities are imprecise, and their precision should increase with the amount of statistical information.

The conclusions of a statistical analysis (or decision analysis) are robust when a realistically wide class of probability (and utility) models all lead to essentially the same conclusions. Conclusions based on an imprecise model are automatically robust, because these do not rely on arbitrary or doubtful assumptions. Of course, these conclusions may be highly indeterminate, but in that case the conclusions obtained from any precise model will be unreliable.

When working with probability distributions for continuous random variables, for example lifetime distributions in reliability analyses, two methods of working with lower densities, and updating these in light of new data, seem to be most promising.

First it is possible to assume some class of distributions, for example one kind of parametrical family, where the parameter is within a certain interval. Then the lower density can be defined as the lower envelope of all these distributions (so for each x we get l(x) is the minimum value of all allowed densities in the point x), and the upper density as the upper envelope. Updating can take place by updating all the allowed densities by the usual Bayes' rule, and then again taking the lower and upper densities. The difficulty in this method is the determination of the lower and upper envelopes.

A second method is to assign l and u directly, so without reference to some class of probability densities, and update these through a generalised
Bayes’ rule, which is a simple routine in case of density functions. Here the difficulty is that, in literature, there are not yet models presented such that updating of \( l \) and \( u \) is simple, while at the same time the relating lower and upper pdf’s \( f \) and \( \tilde{f} \) have simple forms. Some models that have these properties will be presented in a following report.

In our approach we assume imprecise prior densities for a parameter, while the model is precise. However, it is also possible to assume an imprecise model, resulting in lower and upper likelihood functions. For example, instead of using a precise Weibull distribution with shape parameter equal to 2, with an imprecise prior for the shape parameter, we could also assume that the scale parameter is between 1.5 and 2.5, while not defining a prior distribution for this parameter. This subject is not further discussed in this report.

Another possibility, that is attractive because it leads to simple calculations, is to restrict to a class of conjugate densities only. In the general approach, all densities between \( l \) and \( u \) are regarded as possible, while here only conjugate priors are allowed. This can be seen as sensitivity analyses towards some parameter of the conjugate prior (see for example Walley [5, section 5.3] and Meczarski and Zielinski [4]). However, our theory is different to sensitivity analysis, with regard to the prior, in the standard Bayesian framework. Our approach treats lack of perfect knowledge within the concept, while the use of sensitivity analysis leads to the contradictory fact that a concept is used that needs complete knowledge of probabilities, whereas the absence of this knowledge is the reason for the sensitivity analysis. Another important difference is that in sensitivity analysis only a finite number of distributions can be compared, while in our approach a set of an infinite number of prior distributions is used (except in case of precision).

It is also important to regard the problem of reaching a (statistical) decision in a practical problem, but here we refer to Walley [5, section 3.9], who gives an outline of the approach of this problem, and to Wolfenson and Fine [8].
Now we come to the important issue of updating prior densities, in the light of new data.

Let the random variable of interest be a parameter $\theta \in \Theta$ (where $\Theta$ is the parameter space, the set of all possible values for $\theta$), so that the pdf’s and cdf’s in the following part can be seen as prior information on the parameter, or, after updating, posteriors.

The following lemmas show that it is necessary to use $l$ and $u$ for statistical reasoning, when new data, say $x$, come available. Here we conclude that restriction to statistical inferences based only on $F$ and $F$, after updating, is not allowed, because the updated versions of $F$ and $F$ will generally not be the lower and upper cdf’s corresponding to the updated $l$ and $u$.

Lemma 5 shows that, if $F(\theta) \leq G(\theta) \leq F(\theta)$ for all $\theta$, then the relation $F(\theta|x) \leq G(\theta|x) \leq F(\theta|x)$ does not necessarily hold for all $\theta$ and $x$. This is shown by an example in which $G_1(\theta) \leq G_2(\theta)$ does not imply $G_1(\theta|x) \leq G_2(\theta|x)$, caused by the form of the likelihood that plays a role in the updating process.

Lemma 6 shows how $l$ and $u$ must be updated, when data $x$ come available, by using the likelihood $L(\theta|x)$.

**Lemma 5**

Given two prior cdf’s, $G_1$ and $G_2$, such that $G_1(\theta) \leq G_2(\theta)$ for all $\theta \in \Theta$. After updating, if data $x$ have come available, the relation $G_1(\theta|x) \leq G_2(\theta|x)$ does not necessarily hold.

**Proof**

We give a counter-example. Of course the fact that in this lemma the likelihood needed in the updating part is totally free enables such a counter-example.

Suppose $\Theta=[0,1]$. Let the pdf of the model be $f(x|\theta)$, with corresponding likelihood function $L(\theta|x)$. Let $g_i$ be the pdf corresponding to $G_i$ (i=1,2), then the standard Bayes’ rule for updating leads to:

$$g_i(\theta|x) = \frac{L(\theta|x)g_i(\theta)}{\int_0^1 L(\xi|x)g_i(\xi)\,d\xi},$$

where $i=1,2$. 

22
\[
G_1(\theta|x) = \frac{\int_0^\theta L(\xi|x) g_1(\xi) d\xi}{\int_0^1 L(\xi|x) g_1(\xi) d\xi} = \left[ 1 + \frac{\int_0^1 L(\xi|x) g_1(\xi) d\xi}{\int_0^\theta L(\xi|x) g_1(\xi) d\xi} \right]^{-1}.
\]

Therefore, \( G_1(\theta|x) \leq G_2(\theta|x) \) \( \Rightarrow \) \( \int_0^\theta L(\xi|x) g_1(\xi) d\xi \geq \int_0^\theta L(\xi|x) g_2(\xi) d\xi \). (\( \ast \))

Let \( g_1(\theta) = \begin{cases} 
3 - 4\theta & \text{if } \theta \in \left[ 0, \frac{3}{8} \right] \\
-2\theta^2 + 3\theta - \frac{9}{16} & \text{if } \theta \in \left[ \frac{3}{8}, \frac{5}{8} \right], \text{ and } g_2(\theta) = 1 \text{ for } \theta \in [0,1]. \\
2\theta^2 - 2\theta + 1 & \text{if } \theta \in \left[ \frac{5}{8}, 1 \right]
\end{cases} \)

So \( g_1 \) is continuous, and consists of three lines, first increasing linearly from 0 for \( \theta=0 \) to 1.5 for \( \theta=\frac{3}{8} \), then decreasing linearly to 0.5 for \( \theta=\frac{5}{8} \), and finally increasing linearly to 2 for \( \theta=1 \).

For these densities we have \( g_1(\theta)=g_2(\theta) \) if \( \theta \in \left( \frac{1}{4}, \frac{1}{2}, \frac{3}{4} \right) \), \( g_1(\theta)<g_2(\theta) \) if \( \theta \in \left( 0, \frac{1}{4} \right) \cup \left( \frac{1}{2}, \frac{3}{4} \right) \) and \( g_1(\theta)>g_2(\theta) \) if \( \theta \in \left( \frac{1}{4}, \frac{1}{2} \right) \cup \left( \frac{3}{4}, 1 \right) \).

The according cdf’s are:

\[
G_1(\theta) = \begin{cases} 
2\theta^2 & \text{if } \theta \in \left[ 0, \frac{3}{8} \right] \\
-2\theta^2 + 3\theta - \frac{9}{16} & \text{if } \theta \in \left[ \frac{3}{8}, \frac{5}{8} \right], \text{ and } G_2(\theta) = \theta \text{ for } \theta \in [0,1]. \\
2\theta^2 - 2\theta + 1 & \text{if } \theta \in \left[ \frac{5}{8}, 1 \right]
\end{cases}
\]

The relation \( G_1(\theta) \leq G_2(\theta) \) for all \( \theta \) is easily proved.

Now suppose that the likelihood function, needed for updating when data \( x \) come available, is equal to \( L(\theta|x) = \begin{cases} 
1 & \text{if } \theta \in \left[ \frac{1}{4}, \frac{3}{4} \right] \\
0 & \text{else}
\end{cases} \).

23
Then straightforward calculation shows that
\[
\frac{1}{2} \int_0^1 L(\xi \mid x) g_1(\xi) d\xi = \frac{3}{5} \quad \text{and} \quad \frac{1}{2} \int_0^1 L(\xi \mid x) g_2(\xi) d\xi = 1.
\]

Now relation (*) shows us that \( G_1(\frac{1}{2} \mid x) > G_2(\frac{1}{2} \mid x) \), which completes the counter-example.

\[\square\]

**Lemma 6**

After observing data \( x \), the updated versions of \( l(\theta) \) and \( u(\theta) \) are

\[
l(\theta \mid x) = L(\theta \mid x) l(\theta) \quad \text{and} \quad u(\theta \mid x) = L(\theta \mid x) u(\theta)
\]

respectively, where \( L(\theta \mid x) \) is the likelihood function defined by the chosen model (in many situations \( L(\theta \mid x) = f(x \mid \theta) \) is a suitable likelihood function, where \( f(x \mid \theta) \) is interpreted as function of \( \theta \), for known \( x \)). We assume that the likelihood is finite for all \( \theta \).

**Proof**

Let \( l(\theta) \leq h(\theta) \leq u(\theta) \), then after updating we want \( l(\theta \mid x) \leq h(\theta \mid x) \leq u(\theta \mid x) \). Let \( c \) be the normalising constant for the density \( h(\theta) \), then \( h(\theta) / c \) is a pdf, so by updating through the standard Bayes' rule, we get

\[
h_1(\theta \mid x) = \frac{L(\theta \mid x) h(\theta) / c}{\int L(\theta \mid x) h(\theta) / c d\theta}
\]

with \( c_1 \) another normalising constant. However, \( c_1 \) plays no role at all when working with the lower and upper densities, so defining \( l(\theta \mid x) = L(\theta \mid x) l(\theta) \), \( u(\theta \mid x) = L(\theta \mid x) u(\theta) \) and \( h(\theta \mid x) = c_1 h_1(\theta \mid x) \) leads to \( l(\theta \mid x) \leq h(\theta \mid x) \leq u(\theta \mid x) \), because \( L(\theta \mid x) \geq 0 \). This is possible because the functions between \( l \) and \( u \) are interpreted as unnormalised densities, so these are always defined up to a constant factor.

\[\square\]

It must be remarked that the choice \( L(\theta \mid x) = f(x \mid \theta) \) may not always be suitable. For example, if \( \theta \) is a location parameter for a truncated lifetime distribution, than this choice of the likelihood function may lead to serious problems (see Coolen and Newby [1]).
Lemma 7 shows that updating of \( l(\theta) \) and \( u(\theta) \) according to lemma 6 indeed results in \( l(\theta|x) \) and \( u(\theta|x) \) that are correct lower and upper posterior densities. Here we mean that every prior cdf that lies between the lower and upper cdf's resulting from \( l(\theta) \) and \( u(\theta) \), leads, after updating, to a posterior cdf that lies between the lower and upper cdf's resulting from \( l(\theta|x) \) and \( u(\theta|x) \).

**Lemma 7**

Let \( \Theta \subset \mathbb{R} \), let \( g \) be a density that lies between 1 and \( u \), and \( c = \int_{-\infty}^{\infty} g(\theta) \, d\theta \). Let \( l(\theta|x) \) and \( u(\theta|x) \) be defined according to lemma 6, after data \( x \) are observed, and let \( \tilde{g}(\theta|x) \) be the updated pdf resulting through Bayes' rule from \( g(\theta)/c \) and the likelihood \( L(\theta|x) \). Let \( \underline{F}(\theta|x) \) and \( \overline{F}(\theta|x) \) be the lower and upper cdf that relate to \( l(\theta|x) \) and \( u(\theta|x) \) (according to lemma 1), and let \( G(\theta|x) = \int_{-\infty}^{\theta} \tilde{g}(\omega|x) \, d\omega \). Then \( \underline{F}(\theta|x) \leq G(\theta|x) \leq \overline{F}(\theta|x) \).

**Proof**

\( l(\theta) \leq g(\theta) \leq u(\theta) \) implies \( L(\theta|x) l(\theta) \leq L(\theta|x) g(\theta) \leq L(\theta|x) u(\theta) \), and it is obvious that \( \tilde{g}(\theta|x) \) is equal, up to a constant factor, to \( L(\theta|x) g(\theta) \), so we can write \( G(\theta|x) = \frac{\int_{-\infty}^{\theta} L(\omega|x) g(\omega) \, d\omega}{\int_{-\infty}^{\infty} L(\omega|x) g(\omega) \, d\omega} \). On the other hand we have (same argument as lemma 1) \( \underline{F}(\theta|x) = \frac{\int_{-\infty}^{\theta} L(\omega|x) l(\omega) \, d\omega}{\int_{-\infty}^{\infty} L(\omega|x) l(\omega) \, d\omega + \int_{\theta}^{\infty} L(\omega|x) u(\omega) \, d\omega} \) and \( \overline{F}(\theta|x) = \frac{\int_{-\infty}^{\theta} L(\omega|x) u(\omega) \, d\omega}{\int_{-\infty}^{\infty} L(\omega|x) u(\omega) \, d\omega + \int_{\theta}^{\infty} L(\omega|x) l(\omega) \, d\omega} \). Now a same argument as used in the proof of lemma 2a completes the proof.

Lemma 5 shows that updating of \( l \) and \( u \) is necessary, because updating of \( f \) and \( \overline{f} \) does not in all situations lead to the correct lower and upper densities, that correspond to the updated \( l \) and \( u \). Now that we know how to
update 1 and $u$, we give two trivial situations for which updating of $f$ and $\bar{f}$ leads to the same results as updating of 1 and $u$ does. This is presented in lemma 8.

We have to introduce the following notation. Let $l(\theta)$ and $u(\theta)$ be the lower and upper prior densities, and $F(\theta)$ and $\bar{F}(\theta)$ the corresponding lower and upper cdf's. After data $x$ have come available, the updated (posterior) imprecise densities are $l(\theta|x)$ and $u(\theta|x)$, with corresponding posterior imprecise cdf's $F(x|\theta)$ and $\bar{F}(x|\theta)$. The updated versions of $F(\theta)$ and $\bar{F}(\theta)$ are denoted by $F_x(\theta)$ and $\bar{F}_x(\theta)$. So lemma 5 shows that it is not necessary that $F(x|\theta)=\bar{F}(x|\theta)$ and $\bar{F}(x|\theta)=\bar{F}_x(\theta)$. Lemma 8 gives two situations for which these equalities do hold.

Lemma 8

(a) If $L(\theta|x)=0$ for all $\theta$, then $F(x|\theta)=\bar{F}_x(\theta)$;

(b) If $l(\theta)=u(\theta)$ for all $\theta$ with $L(\theta|x)>0$, then $F(x|\theta)=\bar{F}_x(\theta)$.

The same holds when $F(x|\theta)=\bar{F}_x(\theta)$ is replaced by $\bar{F}(x|\theta)=\bar{F}_x(\theta)$.

Proof

(a) Let $L(\theta|x)=0$ for all $\theta$, then

$$F(x|\theta) = \frac{\int_{-\infty}^{\theta} L(\omega|x) l(\omega) d\omega}{\int_{-\infty}^{\theta} L(\omega|x) l(\omega) d\omega + \int_{\theta}^{\infty} L(\omega|x) u(\omega) d\omega} = \frac{\int_{-\infty}^{\theta} \gamma l(\omega) d\omega}{\int_{-\infty}^{\theta} \gamma l(\omega) d\omega + \int_{\theta}^{\infty} \gamma u(\omega) d\omega} = F(\theta),$$

and, by standard Bayes updating,

$$F_x(\theta) = \frac{\int_{-\infty}^{\theta} L(\omega|x) \bar{f}(\omega) d\omega}{\int_{-\infty}^{\theta} L(\omega|x) \bar{f}(\omega) d\omega + \int_{\theta}^{\infty} L(\omega|x) \bar{f}(\omega) d\omega} = \frac{\int_{-\infty}^{\theta} \gamma \bar{f}(\omega) d\omega}{\int_{-\infty}^{\theta} \gamma \bar{f}(\omega) d\omega + \int_{\theta}^{\infty} \gamma \bar{f}(\omega) d\omega} = F(\theta).$$

(b) Let $l(\theta)=u(\theta)$ for all $\theta$ with $L(\theta|x)>0$. For these $\theta$ obviously $l(\theta)=c(\theta)\bar{f}(\theta)=u(\theta)$, with $c(\theta)=\int_{-\infty}^{\infty} l(\omega) d\omega=K$, with $K>0$ constant. So for all $\theta$ we have $L(\theta|x)l(\theta)=L(\theta|x)K\bar{f}(\theta)=L(\theta|x)u(\theta)$. This leads to

$$F(x|\theta) = \frac{\int_{-\infty}^{\theta} L(\omega|x) l(\omega) d\omega}{\int_{-\infty}^{\theta} L(\omega|x) l(\omega) d\omega + \int_{\theta}^{\infty} L(\omega|x) \bar{f}(\omega) d\omega} = \frac{\int_{-\infty}^{\theta} L(\omega|x) K \bar{f}(\omega) d\omega}{\int_{-\infty}^{\theta} L(\omega|x) K \bar{f}(\omega) d\omega + \int_{\theta}^{\infty} L(\omega|x) \bar{f}(\omega) d\omega} = \frac{\int_{-\infty}^{\theta} L(\omega|x) \bar{f}(\omega) d\omega}{\int_{-\infty}^{\theta} L(\omega|x) \bar{f}(\omega) d\omega + \int_{\theta}^{\infty} L(\omega|x) \bar{f}(\omega) d\omega} = F_x(\theta).$$
The proofs for the statements with the lower cdf's replaced by the upper cdf's are analogous.

The above lemma only gives results for two situations that are of little interest, namely a constant likelihood (so we learn nothing from the data) and precision.

An important result, and justification for future research to statistical models in the concept of imprecise probabilities, would be reached if it could be proved that only for these two situations restriction to statistical inference based on $\underline{F}$ and $\overline{F}$ only, leads to the same results (posterior cdf's) as updating $\underline{I}$ and $\overline{u}$, and deriving the corresponding lower and upper cdf's, leads to.

However, we have not been able to prove this idea, nor have we found any situation, besides those of lemma 8, for which $\underline{F}(\theta|x)=\overline{F}(\theta)$ for all $\theta$. Therefore, by lemma 5 we know that there are situations for which the equality $\underline{F}(\theta|x)=\overline{F}(\theta)$ for all $\theta$ does not hold (a simple example is $\theta \in [0,1]$ with $\underline{l}(\theta)=1$, $\overline{u}(\theta)=2$ and $\underline{L}(\theta|x)=\theta$, that lead to $\underline{F}(\theta|x)=\theta^2/(2-\theta^2)$ and $\overline{F}(\theta|x)=\frac{\theta}{2-\theta}-\ln\left(\frac{2}{2-\theta}\right)/(1-\ln 2)$), but we only have an idea that the situations of lemma 8 are the only ones for which this equality does hold.

Our according conjecture, that is not proved yet, nor shown to be wrong, is (we restrict to likelihoods that are not constant, as these lead to obvious results):

**Conjecture**

If $\underline{L}(\theta|x)$ is not constant, then

$\underline{F}(\theta|x)=\overline{F}(\theta)$ for all $\theta$ $\iff$ for all $\theta$: $\underline{l}(\theta)=\overline{u}(\theta)$ or $\underline{L}(\theta|x)=0$ $\iff$ $\underline{F}(\theta|x)=\overline{F}(\theta)$ for all $\theta$.

If this conjecture would be shown to be wrong, so if $\underline{l}(\theta)$, $\overline{u}(\theta)$ and $\underline{L}(\theta|x)$ can be found such that not for all $\theta$: $\underline{l}(\theta)=\overline{u}(\theta)$ or $\underline{L}(\theta|x)=0$, while the equality of the cdf's does hold, then the next challenge would be to find all situations for which a relation, analogous to this conjecture, holds.

Within the Bayesian theory predictive distributions play an important role. When working with imprecise densities, consequent definitions of lower and upper predictive densities, based on the prior densities, are (with $\theta \in \mathbb{R}$)
\[ l_X(x) = \int_{-\infty}^{\infty} f(x | \theta) l(\theta) d\theta \quad \text{and} \quad u_X(x) = \int_{-\infty}^{\infty} f(x | \theta) u(\theta) d\theta, \]

which are again not pdf's. There are results possible that are analogous to the results in this report, by comparing lower and upper cdf's for the random variable \( X \), that can be constructed from \( l_X \) and \( u_X \) according to lemma 1. A result analogous to corollaries 3 and 7 can be proved straightforwardly. It is also easy to see that, if \( f(x | \theta) > 0 \) for all \( x \) and \( \theta \), then \( l_X(x) = u_X(x) \) for all \( x \) if and only if \( l(\theta) = u(\theta) \) for all \( \theta \).

Updating of the predictive densities is simply done by replacing \( l(\theta) \) and \( u(\theta) \) by \( l(\theta | x) \) and \( u(\theta | x) \).

Finally we remark that this theory also provides tools for statistical inferences. For example, hypotheses of the form \( \theta \in \Theta_s < \Theta \) can be tested by using \( \tilde{F}(\theta \in \Theta_s) \) and \( \tilde{F}(\theta \in \Theta_a) \), that are derived using analogous formulas as presented in the proof of lemma 1 in section 3.

Also useful are lower and upper quantiles. The lower \( \rho \)-quantile of \( X \), denoted by \( x_{\rho} \), is the infimum value of \( x \) for which \( F(x) = \tilde{F}(X \leq x) \geq \rho \). The upper \( \rho \)-quantile, denoted by \( x_{\rho} \), is the supremum value of \( x \) for which \( \tilde{F}(x) = \tilde{F}(X \leq x) \leq \rho \).
6. Conclusions and future research

The theory of imprecise probabilities can be used within reliability theory, and consequent definitions of lower and upper hazard rates and hazard functions are proposed.

When using this concept for statistical inference in non-trivial situations, it is necessary to update the imprecise prior densities, as updating of the imprecise prior cdf’s does not necessarily lead to equal results.

Much research is needed to make the concept suitable for practical use. First models must be created such that all necessary calculations in case of updating can be executed in a simple analytical way. In a following report some of such models will be proposed. Another very important aspect is practical elicitation. Here the method seems to offer a more realistic concept than standard Bayes methods do, as imprecision, for example resulting from information from several sources, which can be somehow contradictory, or from lack of information, is taken into account.

Interesting subjects for research are also decision theory in case of imprecise probabilities, and the use of imprecise models (represented by lower and upper likelihoods) together with imprecise prior densities. Only after successful application of the method in practice the concept can be regarded to be useful. This will be the most interesting part of future research.
References


List of COSOR-memoranda - 1991

<table>
<thead>
<tr>
<th>Number</th>
<th>Month</th>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>91-01</td>
<td>January</td>
<td>M.W.I. van Kraaij</td>
<td>The construction of a strategy for manpower planning problems.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W.Z. Venema</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td>91-02</td>
<td>January</td>
<td>M.W.I. van Kraaij</td>
<td>Support for problem formulation and evaluation in manpower planning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W.Z. Venema</td>
<td>problems.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td>91-03</td>
<td>January</td>
<td>M.W.P. Savelbergh</td>
<td>The vehicle routing problem with time windows: minimizing route</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>duration.</td>
</tr>
<tr>
<td>91-04</td>
<td>January</td>
<td>M.W.I. van Kraaij</td>
<td>Some considerations concerning the problem interpreter of the new</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>manpower planning system formasy.</td>
</tr>
<tr>
<td>91-05</td>
<td>February</td>
<td>G.L. Nemhauser</td>
<td>A cutting plane algorithm for the single machine scheduling problem</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M.W.P. Savelbergh</td>
<td>with release times.</td>
</tr>
<tr>
<td>91-06</td>
<td>March</td>
<td>R.J.G. Wilms</td>
<td>Properties of Fourier-Stieltjes sequences of distribution with</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>support in $[0,1)$.</td>
</tr>
<tr>
<td>91-07</td>
<td>March</td>
<td>F. Coolen</td>
<td>Analysis of a two-phase inspection model with competing risks.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R. Dekker</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A. Smit</td>
<td></td>
</tr>
<tr>
<td>91-08</td>
<td>April</td>
<td>P.J. Zwietering</td>
<td>The Design and Complexity of Exact Multi-Layered Perceptrons.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E.H.L. Aarts</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td>91-09</td>
<td>May</td>
<td>P.J. Zwietering</td>
<td>The Classification Capabilities of Exact Two-Layered Perceptrons.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E.H.L. Aarts</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td>91-10</td>
<td>May</td>
<td>P.J. Zwietering</td>
<td>Sorting With A Neural Net.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E.H.L. Aarts</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td>91-11</td>
<td>May</td>
<td>F. Coolen</td>
<td>On some misconceptions about subjective probability and Bayesian</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>inference.</td>
</tr>
<tr>
<td>Date</td>
<td>Month</td>
<td>Authors</td>
<td>Title</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>----------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>91-12</td>
<td>May</td>
<td>P. van der Laan</td>
<td>Two-stage selection procedures with attention to screening.</td>
</tr>
<tr>
<td>91-14</td>
<td>June</td>
<td>J. Korst, E. Aarts, J.K. Lenstra, J. Wessels</td>
<td>Periodic assignment and graph colouring.</td>
</tr>
<tr>
<td>91-16</td>
<td>July</td>
<td>P. Deheuvels, J.H.J. Einmahl</td>
<td>Approximations and Two-Sample Tests Based on P - P and Q - Q Plots of the Kaplan-Meier Estimators of Lifetime Distributions.</td>
</tr>
<tr>
<td>91-19</td>
<td>August</td>
<td>P. van der Laan</td>
<td>The efficiency of subset selection of an almost best treatment.</td>
</tr>
<tr>
<td>91-20</td>
<td>September</td>
<td>P. van der Laan</td>
<td>Subset selection for an -best population: efficiency results.</td>
</tr>
<tr>
<td>91-22</td>
<td>September</td>
<td>R.J.M. Vaessens, E.H.L. Aarts, J.H. van Lint</td>
<td>Genetic Algorithms in Coding Theory - A Table for ( A_2(n,d) ).</td>
</tr>
<tr>
<td>91-23</td>
<td>September</td>
<td>P. van der Laan</td>
<td>Distribution theory for selection from logistic populations.</td>
</tr>
<tr>
<td>No.</td>
<td>Date</td>
<td>Authors</td>
<td>Title</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>----------------------------------------------</td>
<td>--------------------------------------------------------------</td>
</tr>
<tr>
<td>91-26</td>
<td>October</td>
<td>E.E.M. van Berkum, P.M. Upperman</td>
<td>D-optimal designs for an incomplete quadratic model.</td>
</tr>
<tr>
<td>91-29</td>
<td>November</td>
<td>J. Wessels</td>
<td>Tools for the Interfacing Between Dynamical Problems and Models within Decision Support Systems.</td>
</tr>
<tr>
<td>91-32</td>
<td>November</td>
<td>F.P.A. Coolen</td>
<td>The Theory of Imprecise Probabilities: Some Results for Distribution Functions, Densities, Hazard Rates and Hazard Functions</td>
</tr>
</tbody>
</table>