Published full-text article available at: http://dx.doi.org/10.1016/j.ijar.2015.12.008

Combined analysis of unique and repetitive events in quantitative risk assessment

R Flage[1,*], Dubois D[2] & T Aven[1]

[1] University of Stavanger, Norway
2 IRIT – CNRS Université de Toulouse, France
* Corresponding author. Address: Department of industrial economics, risk management and planning, Faculty of Science and Technology, University of Stavanger, 4036
Stavanger, Norway. Telephone: (+47) 51 83 19 08. E-mail: roger.flage@uis.no.

Abstract

For risk assessment to be a relevant tool in the study of any type of system or activity, it needs to be based on a framework that allows for jointly analysing both unique and repetitive events. Separately, unique events may be handled by predictive probability assignments on the events, and repetitive events with unknown/uncertain frequencies are typically handled by the probability of frequency (or Bayesian) approach. Regardless of the nature of the events involved, there may be a problem with imprecision in the probability assignments. Several uncertainty representations with the interpretation of lower and upper probability have been developed for reflecting such imprecision. In particular, several methods exist for jointly propagating precise and imprecise probabilistic input in the probability of frequency setting. In the present position paper we outline a framework for the combined analysis of unique and repetitive events in quantitative risk assessment using both precise and imprecise probability. In particular, we extend an existing method for jointly propagating probabilistic and possibilistic input by relaxing the assumption that all events involved have frequentist probabilities; instead we assume that frequentist probabilities may be introduced for some but not all events involved, i.e. some events are assumed to be unique and require predictive – possibly imprecise – probabilistic assignments, i.e. subjective probability assignments on the unique events without introducing underlying frequentist probabilities for these. A numerical example related to environmental risk assessment of the drilling of an oil well is included to illustrate the application of the resulting method.

Key words: risk assessment; unique; repetitive; imprecise probability; possibility theory

1 Introduction

The Kaplan and Garrick (1981) approach for describing risk has for several decades served as a cornerstone to the field of quantitative engineering risk assessment. According to this approach, risk is equal to and expressed by the set of triplets consisting of (accident) scenarios, the likelihoods λ of these scenarios and their consequences. Three likelihood settings are mentioned by Kaplan (1997): repetitive situation with known frequency (λ = f, where f is a frequentist probability), unique situation (λ = p, where p is a subjective probability), and repetitive situation with unknown frequency (λ = H(f), where H is a subjective probability distribution on the unit interval of frequencies). The last mentioned setting is typically dealt with using the so-called probability of frequency approach, wherein knowledge about all potentially occurring events involved are assumed to be represented by uncertain frequentist probabilities of occurrence, and the epistemic uncertainties about the true values of the frequentist probabilities are described using subjective (also referred to as judgmental or knowledge-based) probabilities. Of course, the first case above is a special case of the third.

A repetitive event is an event whose occurrence or not can be embedded into a hypothetically infinite sequence of similar situations (technically: an exchangeable sequence), whereas a unique event cannot because such a sequence cannot reasonably be envisaged. As an example of a repeatable event, consider the tossing of a die. We can envisage tossing the die over and over again under similar conditions, thus generating a limiting frequency of the event that the die shows '1' (say). On the other hand, as an example of a unique event, consider the case of a particular election for choosing the president of a country. We cannot reasonably envisage a hypothetically infinite number of repetitions of this particular election, because among other issues, candidates are not all the same from one election to the other. Hence, we cannot introduce a frequentist probability of some candidate winning, but we rather just directly assign a predictive subjective probability of this event. Determining whether to treat an event as repeatable or unique is often a judgement call by the analyst; we refer to Section 7 for a discussion of this issue.

As another example, consider an oil and gas company performing an environmental risk assessment of the activity drilling a wildcat oil well. If there is oil at the drilling location, some might be inclined to argue that a number of factors (such as the physical characteristics of the reservoir, e.g. the reservoir pressure at a particular point in space and time; and the performance of technical barrier systems, e.g. whether a particular barrier element functions on demand) come into play to generate a frequentist probability of an oil spill due to a blowout or well-leak (since the reservoir pressure at a particular point in space changes over time, and the barrier element is not perfectly reliable). On the other hand, if there is no oil present in the reservoir, the frequentist probability of a blowout of oil is zero. Oil or no oil is a fixed but unknown state of the world – it is not subject to randomness. The situation cannot be repeated such that in some cases there is oil in this particular reservoir, while in others there is not.

As another example, to the extent that Probabilistic Risk Assessment (PRA) is suitable for terrorism risk, a terrorist attack is also a unique event. It is however not a fixed but as-yet-unrevealed event, like presence of oil in the environmental risk assessment example. Yet a relative frequentist probability of a terrorist attack cannot be meaningfully defined (Aven & Renn, 2009).

The quantitative risk assessment setting can be formally summarized as follows: We are interested in a quantity Z (possibly a vector) and introduce a model g(X) with input quantities $X = (X_1, X_2, \ldots, X_n)$ to predict Z. The quantities Z and X could be the total number and a vector of the number of fatalities due to different accident scenarios, respectively. Alternatively, for a particular explosion accident scenario, Z and X could be the explosion pressure and a set of factors (quantities) affecting the explosion pressure, respectively. Or Z could be an indicator quantity for some overall event of interest, e.g. blowout in an offshore QRA setting or meltdown in a nuclear QRA/PRA setting, and X could be a set of indicator quantities for events that, through various combinations, could lead to the occurrence of the overall event, respectively.

In the present paper we focus on the last setting. Hence, for our purpose Z and X are 0-1-valued unknown quantities where Z equals 1 if some high level event A takes place and 0 otherwise, i.e. Z = I(A). Corresponding to $X = (X_1,X_2,\ldots,X_n)$ are the lower level events $B = (B_1, B_2, \ldots, B_n)$, in the sense that $X_i = I(B_i)$, $i=1,2,\ldots,n$, where I is the indicator function equal to 1 if its argument is true and 0 otherwise. This is a setting commonly dealt with using basic risk analysis models such as fault trees (comprising top events and basic events) and event trees (comprising initiating events, branching events and end events).

In the probability of frequency approach one would by default introduce a frequentist probability f of the event A, and a set of frequentist probabilities $q = (q_1,q_2,\ldots,q_n)$ of the basic events B according to the structure of an event model g. A frequentist probability expresses the fraction of times the event of interest occurs when repeating the situation considered over and over again infinitely. The setting then yields f = g(q), and by establishing a subjective probability distribution over the vector q and propagating it through the model g, a probability distribution of the frequency probability f is established. Particular attention might then be put on the expected value of f, which is assumed to be also the predictive probability of A, i.e. E[f] = P(A).

In the present paper, following Kaplan and Garrick (1981), we relax the assumption that frequentist probabilities of occurrence can be defined for all events involved. Instead we consider a setting where frequentist probabilities can be defined for some but not all events involved – some events are assumed to be unique and their uncertainty can only be assessed using predictive subjective probability assignments, or more generally, as will be argued, as imprecise probabilities.

The main contribution of the present paper is the formulation of a set-up for the combined analysis of unique and repetitive events in quantitative risk assessment that extends the Kaplan and Garrick (1981), methodology to a setting allowing for imprecise probabilities, that either represent epistemic uncertainty on frequentist probabilities, or represent expert uncertainty concerning unique events. The imprecise probability approach has been argued to be a more faithful representation of partial information than unique distributions. As a particular case, the paper describes the extension, to the joint presence of unique and repetitive events, of an existing method for jointly propagating probabilistic and possibilistic inputs to be applicable in such a setting. So this paper contributes to a better understanding of how to articulate the joint presence of aleatory and epistemic uncertainty in risk assessment.

Table 1 Frameworks and methods for the study of different types of events using different types of probability.

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="3">Events</th></tr>
<tr><th>Repeatable</th><th>Unique</th><th>Repeatable and unique</th></tr>
<tr><td rowspan="2">Subjective probabilities</td><td>Precise</td><td colspan="3">Bayesian framework</td></tr>
<tr><td>Imprecise</td><td>Hybrid methods</td><td>Interval analysis</td><td>Present paper method</td></tr>
</table>

Table 1 positions the present paper and its suggested methods in relation to existing frameworks and methods. The Bayesian framework (e.g. Bernardo & Smith, 1994) with its precise probabilities is straightforward to apply to settings involving respectively both repeatable and unique events, as well as to their combination. Imprecise probability methods can also handle these settings: Unique event probabilities are handled by interval analysis (e.g. Moore, 1966) when (some of) the probabilities are imprecise. Efforts have been made to develop hybrid methods for handling repeatable events using precise and imprecise probabilities jointly (e.g. Cooper et al., 1996; Guyonnet et al., 2003; Baudrit et al. 2006; Montgomery, 2009). The present paper extends these efforts by addressing the setting of both repeatable and unique events.

The remainder of this paper is organized as follows. In Section 2, we introduce and discuss the notion of subjective probability. In Section 3, we formalize the set-up for combined analysis of unique and repetitive events using the Bayesian framework, and in Section 4, we extend the set-up to the case of imprecise probabilities. In Section 5 and Section 6, a hybrid probabilistic and possibilistic method is presented and applied, respectively. It can be viewed as a special case of the former. In Section 7, we discuss the findings of the paper, and Section 8 gives a summary and some final remarks.

2 Subjective probabilities and the Kaplan-Garrick setting revisited

There exist different interpretations of a subjective probability, see e.g. Aven & Reniers (2013) for a recent overview of probability interpretations in a risk and safety setting. Here we consider two common interpretations: A subjective probability P(A) interpreted with reference to a standard is the number such that the uncertainty about the occurrence of A is considered equivalent with the uncertainty about the occurrence of some standard event, e.g. drawing, at random, a red ball from an urn that contains P(A) x 100 % red balls (Kaplan and Garrick, 1981; Lindley, 2000; 2006; see also Bernardo & Smith, 1994). Alternatively the subjective probability P(A) can be interpreted as the amount of money that the assigner would be willing to bet if he/she would receive a single unit of payment if the event A were to occur, and nothing otherwise. The agent must also accept the opposite bet, exchanging roles between buyer and seller. Hence the probability of an event by this interpretation is the (fair) price at which the assigner is neutral between buying and selling a ticket that is worth one unit of payment if the event occurs, and worth nothing if not (Singpurwalla 2006). This is the view advocated already in the 1930’s by De Finetti (1936).

Which interpretation to use is a debated topic; see Kaplan (1988, 1997), SEP (2011), Aven (2010, 2013) and Dubois (2010). This discussion is however beyond the scope of the present paper and will not be addressed in the following. What is important here is that both the “random drawing from an urn” analogy and the betting view force the assigner/analyst to provide unique probabilities for describing his/her state of beliefs. Subjective probabilities are based on an axiom assuming that precise measurements of uncertainties can be made; see e.g. Bernardo & Smith (1994 p. 31). However, many researchers have questioned this assumption of and requirement for precise measurements of uncertainties (e.g. Dempster, 1968; Shafer, 1976; Walley, 1991; see also Dubois et al. 1996; Dubois and Prade, 2009). In many situations, the basis for assigning the probabilities could be weak. A specific number can be assigned, but the rationale for it can be questioned – strong assumptions may be needed to justify a concrete number. The assigned probability is precise, but such a precision may be considered rather arbitrary. One may, for example, subjectively assess that two different situations have probabilities equal to 0.7 say, but in one case the assignment is supported by a substantial amount of relevant data, whereas, in the other, by effectively no data at all. The strength of knowledge, the probabilities are based on, is not at all reflected by the assigned number (Aven & Zio 2011).

These considerations have led to the use of alternative approaches and methods for representing the uncertainties (Dubois and Prade, 2009; Dubois 2010). Here our focus is on interval probabilities or, more generally, imprecise probabilities. Imprecise probability theory generalises probability by using an interval [$\underline{P}$(A), $\overline{P}(A)$] to represent uncertainty about an event A, with lower probability $\underline{P}$(A) and upper probability $\overline{P}(A) = 1 - \underline{P}(\overline{A})$, where $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$ and $\overline{A}$ is the complement of set *A*.

In line with the uncertainty standard interpretation of a subjective probability (Lindley, 2006) we interpret an imprecision interval, say [0.2, 0.5] for the event A, as follows: The analyst states that his/her assigned degree of belief is greater than the urn chance of 0.20 (the degree of belief of drawing one particular ball out of an urn comprising 5 balls) and less than the urn chance of 0.5. The analyst is not willing to make any further judgments. Then the interval [0.2, 0.5] can be considered an imprecision interval for the hypothetical urn probability P(A). In contrast, following Walley's betting interpretation (Walley, 1991), the lower probability is interpreted as the maximum price for which one would be willing to buy a bet which pays 1 if A occurs, and 0 if not and the upper probability as the minimum price for which one would be willing to sell the same bet. If the upper and lower probabilities are equal, we are led to precise probabilities à la De Finetti. In this case, the lower bound 0.2 is directly interpreted as the degree of belief in *A*, the degree of belief of its complement being 0.5.

The use of imprecision intervals is a debated topic also in the applied probability literature; see for example Lindley (2006) and Aven (2011). In their paper, Kaplan and Garrick (1981 p. 18) claim that "Statistics, as a subject, is the study of frequency type information. That is, it is the science of handling data. On the other hand probability, as a subject, we might say is the science of handling the lack of data. […] When one has insufficient data, there is nothing else one can do *but* use probability." The above discussion emphasizing the emergence of generalisations of subjective probability and the use of intervals for describing incomplete information make Kaplan and Garrick's emphatic statement on the exclusive position of probability theory in belief representation indeed debatable. However, we leave this discussion here.

For the purpose of the present work we simply assume that the analysts either accept the precise subjective stance (the Bayesian approach), or adopt an approach based on imprecise probabilities. The former approach is a special case of the latter, but we choose to treat these approaches separately as they are based on different ideas and it is easier to understand the general case by first casting the analysis within the simpler Bayesian framework. For an extension of the Kaplan and Garrick risk perspective, which allows for alternative ways of representing uncertainties, including imprecise probabilities, see Aven (2011).

3 Combined analysis of unique and repetitive events using the Bayesian framework

In this section we present a form of the Kaplan-Garrick set-up for the combined analysis of unique and repetitive events according to a Bayesian approach. We do so first for a setting involving two events, before introducing the more general setting involving a complex combination of events. The set-up is based on precise probabilities in this section. Imprecise probabilities are considered in Section 4.

3.1 Two-event case

Let A denote an event whose occurrence depends on the outcome of a unique event U and of a repeatable event R in such a way that A occurs if both U and R occur. Furthermore, define $X = I(U)$ and $Y = I(R)$, where I is the indicator function which equals 1 if its argument is true and 0 otherwise. Then we have

$$I(A) = XY.$$

Suppose that an expert supplies a subjective probability p for U, i.e. $p = P(U) = P(X = 1)$ and knows the precise frequentist probability f of R, i.e. $f = F(R) = F(Y = 1)$; here P is a subjective probability measure and F a frequentist one. To proceed we rely on the following two principles:

1. A complex event made up of a logical combination of repeatable and non-repeatable events is non-repeatable (as we are interested in the particular occurrence of R taking place along with the unique occurrence of U).
2. The subjective probability of the occurrence of a repeatable event should be measured by its

real frequency if the latter is known (the Hacking frequency principle (Hacking, 1965)).

Based on the above principles and assuming independence between U and R, we can consider pf as the subjective probability of A. To see this, note that

$$P(A) = E[XY] = P(XY = 1) = P(U \cap R) = P(U|R)P(R) = P(U)F(R) = pf.$$

In this calculation, the subjective probability $P(R)$ is measured by means of the frequency of occurrence of R, $F(R)$, during the unique experiment where U is observed or not. This interpretation is implicit when we compute the product pf.

Actually the probability $P(A)$ should be written $P(A|f)$ as f is assumed known. If this assumption is dropped, we obtain the unconditional probability of the event A:

$$P(A) = \int P(A|f)dH(f) = \int pfdH(f) = p \cdot E(f), \tag{1}$$

where H is a subjective probability distribution of f and $E(f)$ is its mean value.

3.2 Complex combination of events

More generally, let A denote an event whose occurrence depends on the outcome of n unique events $U = (U_1, U_2, \dots, U_n)$ and m repetitive events $R = (R_1, R_2, \dots, R_m)$ through a function g. For example, A could be the top event and (U, R) the basic events of a fault tree. Or A could be an end event and (U, R) the initiating event and barrier failure events of an event tree. Define a vector of *unknown* quantities $X = (X_1, X_2, \dots, X_n)$ and a vector of *random* quantities $Y = (Y_1, Y_2, \dots, Y_m)$, and let $X_i = I(U_i)$, $i = 1,2, \dots, n$, and $Y_i = I(R_i)$, $j = 1,2, \dots, m$. Then we have

$$I(A) = g(X, Y).$$

Let $p = (p_1, p_2, \dots, p_n)$ be the subjective probabilities of the unique events X, and let $H(f)$ denote the subjective probability distribution of f, where we now extend the definition of f from now on to be a vector $f = (f_1, f_2, \dots, f_m)$ and $f_i = F(R_i)$ is the frequentist probability of the event R_i. We assume that the analysts' uncertainties about f are not dependent on the outcomes of X. Denoting by $\Omega = \{0, 1\}^m$ the domain of f, $p(x) = P(X = x)$ and $g(x, f) = P(A|X = x, f)$ and taking expectations, yields:

$$P(A) = E\left[E[g(X,Y)|f]\right] = \int_\Omega E[g(X,f)]dH(f) = \int_\Omega \sum_x g(x,f)p(x)\,dH(f)$$
$$= \sum_x p(x)\int_\Omega g(x,f)dH(f) = \sum_x p(x)E[g(x,f)].$$

Moreover the subjective probability $p(x)$ is of the form $\prod_{i:\,x_i=1} p_i \cdot \prod_{i:\,x_i=o}(1-p_i)$, which since x_i is Boolean can be written $\prod_{i=1,\dots,n}\left(x_i p_i + (1-x_i)(1-p_i)\right)$.

4 Combined analysis of unique and repetitive events using imprecise probabilities

We now extend the set-up presented in Section 3 from a precise probability framework to an imprecise probability framework. The objective is to establish an interval $\left[\underline{P}(A), \overline{P}(A)\right]$, with $\underline{P}(A) \leq \overline{P}(A)$. This interval represents, in the presence of unique events, the degree of belief in A for the expert, and, by duality, the degree of belief in its complement $\left(\overline{A}\right)$. Total ignorance is obtained if both degrees of belief are 0.

4.1 Two-event case

Consider the setting introduced in Section 2.1 involving two events, and first assuming that the precise frequentist probability f is known and that $\underline{p} = \underline{P}(U)$ and $\overline{p} = \overline{P}(U)$ are lower and upper probabilities of the unique event U, respectively. Note that, under Walley's gamble-based approach, the subjective interval $[\underline{p}\,, \overline{p}]$ represents the analyst's state of belief about U and does not contain a "true subjective probability" that would represent the analyst's opinion in an accurate way (Walley, 1991). The same holds for the other interpretation described in Section 2, where the analyst's uncertainty about U is compared to that related to a standard event.

In the simple case of two events, assuming independence, we just have to multiply the subjective probability bounds and the frequentist probability to obtain

$$\left[\underline{P}(A), \overline{P}(A)\right] = \left[\underline{p}\cdot f, \overline{p}\cdot f\right].$$

In the case that the frequentist probability f is ill-known, still assuming independence between U and R and now also that the probability of U is not affected by knowing or not knowing the value of f, we obtain

$$[\underline{P}(A), \overline{P}(A)] = [\underline{p} \cdot \underline{f}, \overline{p} \cdot \overline{f}].$$

where $\underline{f} = \underline{E}[f]$ and $\overline{f} = \overline{E}[f]$, derive from the lower and upper cumulative distribution functions on f, respectively.

4.2 Complex combination of events

Based on an imprecise probability representation of the uncertain individual elements of the vectors f and X we want to establish bounds on $P(A)$. We assume independence everywhere, i.e. between the elements of X, between the elements of Y and between X and Y. The subjective probability $p(x)$ is then on the format $P(X_1 = x_1)\ P(X_2 = x_2) \dots P(X_n = x_n)$, $x_i \in \{0,1\}$, $i = 1,2, \dots, n$, which can be evaluated using interval analysis (Moore, 1966). It is assumed that the subjective assessment of $P(X_i = x_i)$ is specified by a probability interval $[\underline{p_i}\ \overline{p_i}]$.

In the most general case, the imprecise knowledge about the frequencies f is represented by a credal set, that is, a convex set of probabilities **F**. For specified values x of X the function $g(X, f)$ is a product of the individual elements of f and hence straightforward to evaluate. Instead of integrating $g(x, f)$ in terms of $H(f)$ we may consider directly the integral $E[g(x, f)]$ defined at the end of Section 3 (integrated over f): By obtaining bounds on $P(g(x, f) \geq \alpha)$ we can use the well-known result that for a non-negative unknown quantity W its expected value is given by (e.g. Ross, 1996)

$$E[W] = \int P(W \geq w)dw,$$

so that letting

$$\underline{P}(g(x, f) \geq \alpha) = \inf\{P(g(x, f) \geq \alpha): P \in \mathbf{F}\},$$

$$\text{and } \overline{P}(g(x, f) \geq \alpha) = \sup\{P(g(x, f) \geq \alpha): P \in \mathbf{F}\}$$

we can compute the bounds

$$\underline{E}[g(x, f)] = \int_0^1 \underline{P}(g(x, f) \geq \alpha)d\alpha, \tag{2}$$

and

$$\overline{E}[g(x,f)] = \int_0^1 \overline{P}(g(x,f) \geq \alpha)d\alpha. \quad (3)$$

In some cases (when the lower probability is 2-monotone, or is a belief function) we have equalities:

$$\underline{E}[g(x,f)] = \inf_{\mathrm{P}\in\mathbf{F}} \mathrm{E}(\mathrm{g}(\mathrm{f},\mathrm{x})); \qquad \overline{E}[g(x,f)] = \sup_{\mathrm{P}\in\mathbf{F}} \mathrm{E}(\mathrm{g}(\mathrm{f},\mathrm{x}))$$

and these quantities are Choquet integrals.

Next, we can compute upper and lower probabilities using interval optimization techniques (Moore, 1979; Lodwick et al., 2008):

$$\underline{P}(A) = inf\ \{\textstyle\sum_x \underline{E}[g(x,f)] \prod_{i=1,\dots,n}\left(x_i p_i + (1-x_i)(1-p_i)\right): p_k \in \left[\underline{p_k}\ \overline{p_k}\right],\ k = 1,\dots,n\} \quad (4)$$

and

$$\overline{P}(A) = sup\{\textstyle\sum_x \overline{E}[g(x,f)] \prod_{i=1,\dots,n}\left(x_i p_i + (1-x_i)(1-p_i)\right): p_k \in \left[\underline{p_k}\ \overline{p_k}\right],\ k = 1,\dots,n\} \quad (5)$$

where the sum is over x such that $g(x,f) > 0$. Note that these expressions are generally not easy to compute, because their monotonic behaviour in terms of probabilities p_k is not easy to guess. Generally, we know that the optimum is attained for one of the bounds of $\left[\underline{p_k}\ \overline{p_k}\right]$ but in the worst case, we must try all of them, which is of exponential complexity in the number *n* of non-repeatable events. See Jacob et al. (2011) for detailed calculations and an algorithm for the computation of such kinds of intervals. Strangely enough, there is almost no older literature on interval calculations with imprecise probabilities on Boolean expressions, stemming e.g., from fault trees.

5 Hybrid probabilistic and possibilistic approach

Like imprecise probability theory, possibility theory differs from probability theory (where a single probability measure is used) in that it uses a pair of dual set-functions, here called possibility and necessity measures, to represent uncertainty. The possibility function π is the basic building block of possibility theory and for each s in a set S, $\pi(s)$ expresses the degree of possibility of s being the true value of some ill-known quantity X. When $\pi(s) = 0$ for some s, the outcome s is considered impossible, whereas when $\pi(s) = 1$ for some s, the outcome s is possible, i.e. is just unsurprising, normal, usual (Dubois, 2006). This is a much weaker statement than a probability equal to 1. As one of the elements of S is the true value, it is assumed that $\pi(\mathrm{s}) = 1$ for at least one s. Possibility distributions on numerical spaces often take the form of fuzzy intervals (Dubois et al. 2000), namely, π is an upper semi-continuous mapping from the reals to the unit interval such that $\mathrm{V}_\alpha = \{\mathrm{v}: \pi(\mathrm{v}) \geq$

$\alpha\}$ is a closed interval for all $0 < \alpha \leq 1$. Hence, possibility theory is a suitable representation of uncertainty in situations where the available information takes the form of nested subsets with various confidence levels. Two sets are nested if one of the sets is a subset of the other and, by extension, sets in a sequence are nested if each subsequent set is contained in the next.

The possibility function π induces a pair of necessity/possibility measures $[\mathrm{N}, \Pi]$, and the possibility of an event A, $\Pi(\mathrm{A})$, is defined by

$$\Pi(\mathrm{A}) = \sup_{\mathrm{s} \in \mathrm{A}} \pi(\mathrm{s}), \tag{6}$$

and the associated necessity measure, $\mathrm{N}(\mathrm{A})$, by

$$\mathrm{N}(\mathrm{A}) = 1 - \Pi(\overline{\mathrm{A}}) = \inf_{\mathrm{s} \notin \mathrm{A}} (1 - \pi(\mathrm{s})). \tag{7}$$

Then each cut V_α of a continuous fuzzy interval π can viewed as an interval containing the quantity of interest with confidence level $\mathrm{N}(\mathrm{I}_\alpha) \geq 1 - \alpha$.

Let $\mathbf{P}(\pi)$ be the family of probability measures $P \in \mathbf{P}(\pi)$ such that for all events A, $\mathrm{N}(\mathrm{A}) \leq \mathrm{P}(\mathrm{A}) \leq \Pi(\mathrm{A})$. It is known that (Dubois et al. 2004) that $\mathbf{P}(\pi) = \{\mathrm{P}\colon \mathrm{P}(\mathrm{I}_\alpha) \geq 1 - \alpha, 0 < \alpha \leq 1\}$. Then it can be proved that set functions N and Π are coherent lower and upper previsions in the sense of Walley (1991), namely (De Cooman and Ayels, 1999):

$$\mathrm{N}(\mathrm{A}) = \inf_{\mathrm{P} \in \mathbf{P}(\pi)} \mathrm{P}(\mathrm{A}) \quad \text{and} \quad \Pi(\mathrm{A}) = \sup_{\mathrm{P} \in \mathbf{P}(\pi)} \mathrm{P}(\mathrm{A}).$$

The possibility and necessity measures of an event can hence be interpreted as upper and lower limits, respectively, for the probability of an event, i.e. they are special cases of upper and lower probabilities. Under the subjectivist view, they model the cautious expert that sells Boolean gambles less than 1$ only if (s)he can get them for free.

When uncertainty about f is described using possibility theory, the extension principle (e.g. Zadeh, 1975) can be used to generate a possibility distribution π_g for $g(x, f)$ for specified values of x. It takes the following form (Dubois et al. 2000):

$$\pi_g(z) = \Pi(\{f\colon g(x,f) = z\}) = \sup_{f\colon g(x,f)=z} \min(\pi_1(f_1),\cdots,\pi_m(f_m))$$

This computation comes down to performing interval computations on α–cuts $V_{i\alpha}$ of possibility distributions π_i, which yield the α–cuts of π_g. From a probabilistic point of view, the choice of a single threshold α for all distributions π_i presuppose complete dependency between the possibility distributions, which corresponds to the same level of confidence for all inputs, used by a single expert. Alternatively, one may choose various thresholds for cutting the π_i, and combine interval analysis with Monte-Carlo methods, thus obtaining a random set for $g(x,f)$. This makes sense if the π_i's come from independent sources. The result of applying the extension principles are not directly comparable with the results of treating the π_i's as independent random sets. See Dubois and Prade (1991) for the relation between fuzzy and random set arithmetics, and Baudrit et al. (2006, 2007) for the theoretical foundations and implementation of a hybrid possibilistic-probabilistic propagation method using interval analysis and Monte-Carlo methods.

Then bounds of the form $\mathrm{N}(g(x,f) \geq \alpha) \leq P(g(x,f) \geq \alpha) \leq \Pi(g(x,f) \geq \alpha)$ can be generated, from which bounds on the integral $E[g(x,f)]$ can be obtained using (2) and (3) as (Baudrit et al. 2007):

$$\underline{E}[g(x,f)] = \int_0^1 \Pi(g(x,f) \geq \alpha)d\alpha = \inf_{\mathrm{P}\in\mathbf{P}(\pi_g)} \mathrm{E}(\mathrm{g}(\mathrm{f},\mathrm{x}))\,, \tag{8}$$

and

$$\overline{E}[g(x,f)] = \int_0^1 \mathrm{N}(g(x,f) \geq \alpha)d\alpha = \sup_{\mathrm{P}\in\mathbf{P}(\pi_g)} \mathrm{E}(\mathrm{g}(\mathrm{f},\mathrm{x})). \tag{9}$$

In the next section we provide a numerical example of the procedure described above.

6 Numerical example

We now return to the oil-drilling example introduced in Section 1, and define the following events:

A = Blowout

U_1 = Oil at drilling location

U_2 = Well kick barrier failure type I

R_1 = Well kick

R_2 = Well kick barrier failure type II

We assume that a blowout will occur under the combined circumstances that there is oil at the drilling location, a well kick occurs during the drilling, and either one or both of the barriers in place to avoid that a well kick develops into a blowout fails, i.e., $A = U_1 \cap R_1 \cap (U_2 \cup R_2)$. According to principle (1) described in Section 3.1, the event A is a unique event (being made up of a logical combination of repeatable and non-repeatable events). Then we have

$$I(A) = g(X, Y) = X_1Y_1\big(1 - (1 - X_2)(1 - Y_2)\big) = X_1(1 - X_2)Y_1Y_2 + X_1X_2Y_1$$

(decomposing into disjoint events) and consequently

$$g(x, f) = x_1f_1\big(1 - (1 - x_2)(1 - f_2)\big) = x_1(1 - x_2)f_1f_2 + x_1x_2f_1\,,$$

so that

$$g\big((1,1), f\big) = f_1,$$
$$g\big((0,1), f\big) = 0,$$
$$g\big((1,0), f\big) = f_1f_2,$$
$$g\big((0,0), f\big) = 0$$

Regarding dependence between the set of events (U_1,U_2,R_1,R_2), clearly $P(U_1|U_2,R_1,R_2) = P(U_1)$ and $P(R_1|R_2,U_2) = P(R_1)$, since U_1 and R_1 occur before the conditional events in these probabilities. Also, clearly $P(U_2|U_1,R_1,R_2) = P(U_2|R_1,R_2)$ and $P(R_2|U_1,U_2,R_1) = P(R_2|U_2,R_1)$, since the occurrence of R_1 makes U_1 superfluous. However, we may not always have $P(U_2|R_1,R_2) = P(U_2)$ and $P(R_2|R_1,U_2) = P(R_2)$, since the probability of a barrier failure may differ during testing and during real demand, and since knowing that one barrier has failed could induce some belief in a common cause failure having occurred. Nevertheless, we assume independence here as a simplification judged as reasonable. Finally, we have $P(R_1) = P(R_1|U_1)\,P(U_1) + P(R_1|\text{not } U_1)\,P(\text{not } U_1) = f_1\,p_1 + 0\,(1 - p_1) = f_1\,p_1$. Using precise probabilities we obtain

$$P(A) = \sum_x p(x)E[g(x, f)] = p_1p_2E[f_1] + p_1(1 - p_2)E[f_1f_2],$$

Note that the system g is a monotone system with respect to its arguments, but the parameters *(p_1, p_2,$E[f_1]$,$E[f_1f_2]$)* that we obtain through the analysis are not directly those of elementary events, namely (p_1,p_2, $E[f_1]$, $E[f_2]$). It is, however, easy to check that in this particular case *P(A)* is increasing with $E[f_1]$ and $E[f_1f_2]$ (as patent in the expression above) as well as with p_1 and p_2. Indeed, $P(A)$ also reads $p_1(p_2(E[f_1] - E[f_1f_2]) + E[f_1f_2])$, and $E[f_1] \geq E[f_1f_2]$. So the upper and lower probabilities of *A* are obtained as

$$\underline{P}(A) = \underline{p}_1\underline{p}_2\underline{E}[f_1] + \underline{p}_1\left(1 - \underline{p}_2\right)\underline{E}[f_1f_2],$$

$$\overline{P}(A) = \overline{p}_1\overline{p}_2\overline{E}[f_1] + \overline{p}_1(1 - \overline{p}_2)\overline{E}[f_1f_2].$$

Note that the monotonicity of the function g is not always guaranteed. It would not hold if the Boolean expression of A contains an elementary event and its negation (for example an exclusive OR). In that case the computation of the upper and lower probabilities can be tricky (it may become exponential in complexity with the number of variables, see (Jacob et al. 2011). This phenomenon usually does not appear in fault-trees where the top event is just a disjunction of conjunctions and the interval analysis is straightforward.

We now assume that the lower and upper values of p_1 and p_2 are $\left[\underline{p}_1,\overline{p}_1\right] = [0.4,0.6]$ and $\left[\underline{p}_2,\overline{p}_2\right] = [0.01,0.2]$, respectively. Furthermore, we assume that the range and core values of the possibility distributions $\pi(f_1)$ and $\pi(f_2)$ are given by the triplets $(0.1,0.2,0.3)$ and $(0.2,0.5,0.7)$, respectively. These possibility distributions are shown in Figure 1, together with the possibility distribution $\pi(f_1f_2)$ resulting from the combination of $\pi(f_1)$ and $\pi(f_2)$. The distribution $\pi(f_1f_2)$ was calculated by application of the extension principle, which is equivalent to performing interval analysis on the α-cuts $V_\alpha = \left[\underline{v_\alpha}, \overline{v_\alpha}\right] = \{v: \pi(v) \geq \alpha\}$, as already explained. The interpretation of the possibility distributions in Figure 1 is as follows: The probability that the quantity in question, say f_1, belongs to V_α is $P(f_1 \in V_\alpha) \geq 1 - \alpha$. For example, looking at the top distribution of Figure 1, we see that there is a 50 % or greater probability that f_1 belongs to the interval [0.15,0.25].

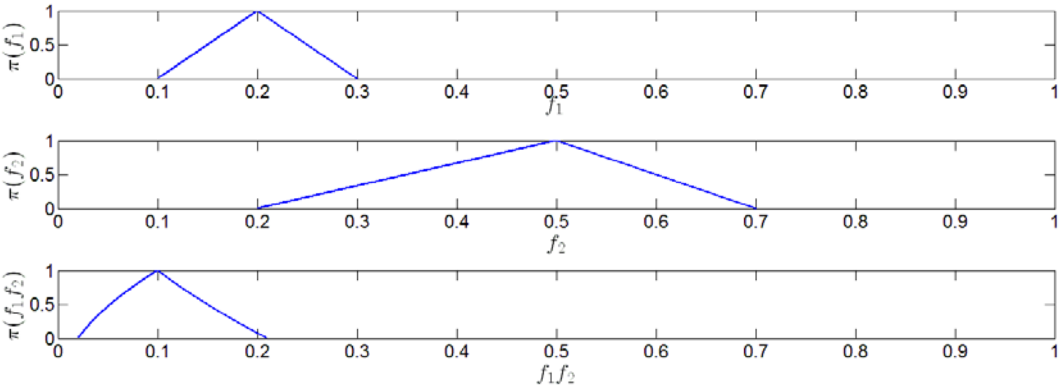

Figure 1 Possibility distributions of f_1(top), f_2 (middle) and f_1f_2 (bottom).

The cumulative and complementary cumulative Possibility and Necessity distributions of f_1f_2 induced by $\pi(f_1f_2)$ are shown in Figure 2, determined using Equations (6) and (7), respectively. The interpretation of these distributions is, for example, that considering the top set of distributions, the probability $P(f_1f_2 \leq 0.05)$ is seen to be greater than 0 and less than approximately 0.6 (cf. the left/green cumulative possibility distribution); and the probability $P(f_1f_2 \leq 0.15)$ is seen to be less than 1 and greater than approximately 0.5 (cf. the right/red cumulative necessity distribution).

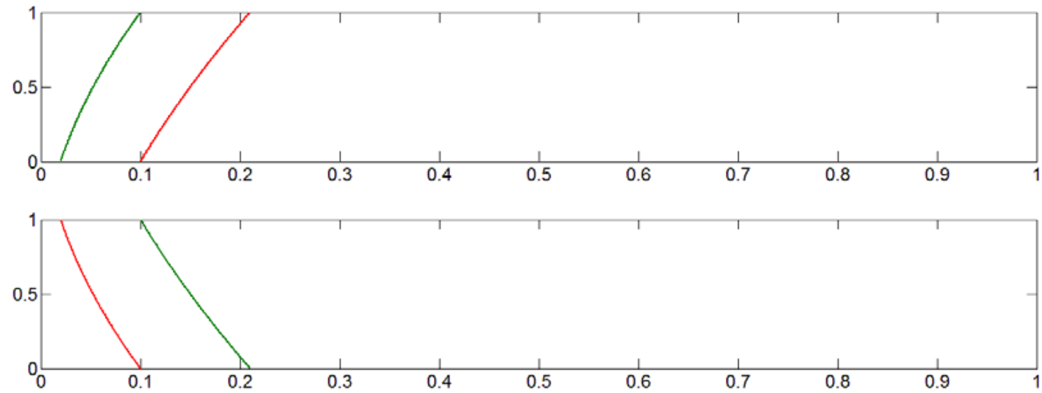

Figure 2 Cumulative (top) and complementary cumulative (bottom) Necessity and Possibility distributions of f_1f_2.

Use of Equations (8) and (9) yield $\left[\underline{E}[f_1], \overline{E}[f_1]\right] = [0.15{,}0.25]$ and $\left[\underline{E}[f_1f_2], \overline{E}[f_1f_2]\right] = [0.055{,}0.15]$, and so finally by means of Equations (4) and (5) we obtain

$$\left[\underline{P}(A), \overline{P}(A)\right] = [0.02038, 0.102].$$

Hence, the imprecise subjective probability of a blow-out is approximately [0.02, 0.1], which, under the uncertainty standard interpretation, can be interpreted to mean that the degree of belief in a blowout is greater than that of drawing one particular ball out of an urn containing 50 balls, and less than that of drawing one particular ball from an urn containing 5 balls; cf. Section 2. Alternatively, in the Walley spirit, the degree of belief in a blow-out is very small (0.02) and the degree of belief in the absence of blow-out is very large (.9).

7 Discussion

The framework described in the present paper is based on a distinction between unique and repetitive events, i.e. events are assumed to be of one kind or the other. Deciding when the situation is repetitive, respectively unique – or rather deciding which events to model as repetitive and which ones to treat as unique – is however not necessarily straightforward. The question of when to introduce probability models and frequentist probabilities (chances) instead of using predictive probability assignments is addressed by Aven (2012a) (see also Aven (2012b)), who argues that introducing frequentist probabilities (chances) should only be done when the quantities of interest are frequentist probabilities and/or when systematic information updating is important to meet the aim of the analysis.

When to use precise and imprecise probability is also not necessarily obvious. As discussed by Flage (2010) (see also Aven et al. (2014)), a recurring argument appears to be that probability is an appropriate representation of uncertainty only if a large enough amount of data is available to base it on. However, it is not obvious how to make such a prescription operational (Flage, 2010 p. 33): *'Consider the representation of uncertainty about the parameter(s) of a probability model. If a large enough amount of data exists, there would be no uncertainty about the parameter(s) and hence no need for a representation of such uncertainty. When is there enough data to justify probability, but not enough to accurately specify the true value of the parameter in question and, thus, make probability, as an epistemic concept, superfluous? […] Also, depending on the relevance of the available observations, different [amounts of data] could be judged as sufficient in different situations.'*

The latter point was also made by one of the reviewers of the present paper. He stated that reducing the issue to a question of sample size misses much of the discussion on this topic, and points out that (we allow ourselves to cite the referee comments) *'In fact, the nature of the data is also quite important. For instance, imprecise methods are particularly useful when the plus-or-minus ranges for data values are non-negligible, when there is statistical censoring or missing data, or when there is structural uncertainty about shapes of distributions, their dependencies, or the correct form of the expression or model being evaluated.'* Nevertheless, under some assumption such as unimodal distributions, it is possible to derive a possibility distribution from a single observation (assuming the latter is the mode), using a probabilistic inequality (Mauris, 2008).

As stated in Section 3, we rely on two principles related to the events and probabilities involved, that lead us to consider non-repeatable any complex events involving at least a non-repeatable one, and to admit that the frequency of a repeatable event is used as the subjective probability of any of its occurrence.

The first principle is self-evident, since the complex event occurs if and only if the non-repeatable events occur. The second principle is well-known in Bayesian statistics, where the probability of an event in an exchangeable sequence of random quantities is taken as equal to the limiting frequency (sometimes referred to as a chance (e.g. Singpurwalla & Wilson, 2009) or propensity (e.g. Lindley & Phillips, 1976) of that event in the same sequence, i.e. $P(A|p) = p$, where p is the chance of A.

Returning to the formal setting introduced in Section 1, with the model g(X) used to predict a quantity of interest Z, the methods in the present paper only consider parameter uncertainty, i.e. uncertainty about X, and neglects model (output) uncertainty, i.e. uncertainty about the difference $g(X)–Z$. In this vein, we may consider the use of set-valued models in order to cope with lack of knowledge about function g.

8 Summary and final remarks

This paper develops a framework for combined analysis of unique and repetitive events in quantitative risk assessment using both precise and imprecise probability. We start from a subjective probability (Bayesian) setting involving one event of each type, and then generalise to a setting involving a complex combination of the two types of events with their associated uncertainty represented by a combination of subjective probability and imprecise probability. One way of representing imprecise probability is through possibility theory, and we extend an existing method for jointly propagating probabilistic and possibilistic input to fit the developed framework. Finally, a numerical example related to environmental risk assessment of the drilling of an oil well is included to illustrate the application of the proposed method.

Allowing for different types of events and different types of uncertainty representations enlarges our modelling capacities. Computationally, the described methods can be handled by a combination of interval analysis and reliance on the fact that the defined complex function of unique and repetitive events is an uncertain quantity on the unit interval, i.e. a non-negative uncertain quantity, which means that we can rely on the well-known result that the expectation of a non-negative uncertain quantity is the integral over its complementary cumulative distribution function.

The described framework can be extended and applied in various ways. In the present paper we have considered the case of possibility theory as representation of uncertainty in terms of lower and upper probability. Other representations also exist for this purpose, for example the theory of belief functions (also known as evidence theory or Dempster-Shafer theory (Shafer, 1976)) as an alternative to theories of imprecise or interval probability (Walley, 1991; Weichselberger, 2000). An example where incomplete knowledge of frequencies is modeled by belief functions on the frequency range [0, 1] instead of possibility distributions is in (Limbourg, et al., 2007). While belief functions are just special cases of lower probabilities corresponding to random sets, it turns out that on Boolean frames like {0, 1}, they are mathematically equivalent. So it is tempting to represent a plain probability interval for a Boolean variable by a random set on {0, 1}. On this basis, imprecise probabilities of final events of an event tree can be efficiently obtained, and more general reliability analysis methods can be devised (Aguirre et al. 2014). The obtained results differ from those of interval analysis because the underlying independence analysis are not the same. The comparison of the two approaches is worth studying further.

Furthermore, it is conceivable to extend the framework to include finite populations, i.e. the setting between unique and repetitive events.

Acknowledgements

For R. Flage and T. Aven the work on the paper has been funded by the Research Council of Norway through the Research Centre for Arctic Petroleum Exploration (ARCEx) (grant number 228107) and the PETROMAKS2 programme (grant number grant number 228335/E30). The financial support is gratefully acknowledged.

References

Aven, T (2010) On the need for restricting the probabilistic analysis in risk assessments to variability. *Risk Analysis*, 30(3), 354-360. With discussion 381-384.

Aguirre, F, Destercke S, Dubois D, Sallak M, Jacob C (2014) Inclusion-exclusion principle for belief functions. *International Journal of Approximate Reasoning*, 55(8): 1708-1727.

Aven, T (2011) A risk concept applicable for both probabilistic and non-probabilistic perspectives. *Safety Science*, 49: 1080-1086.

Aven, T (2012a) On when to base event trees and fault trees on probability models and frequentist probabilities in quantitative risk assessments. *International Journal of Performability Engineering*, 8(3): 311-320.

Aven, T (2012b) *Foundations of Risk Analysis*. Wiley: Chichester.

Aven, T & Reniers, G (2013) How to define and interpret a probability in a risk and safety setting. *Safety Science*, 51(1): 223–231.

Aven, T & Renn, O (2009) The role of quantitative risk assessments for characterizing risk and uncertainty and delineating appropriate risk management options, with special emphasis on terrorism risk. *Risk Analysis*, 29(4): 587–600.

Bernardo, JM & Smith, AFM (1994) *Bayesian Theory*. Chichester: Wiley.

Baudrit, C, Couso, I & Dubois, D (2007) Joint propagation of probability and possibility in risk analysis: Towards a formal framework. *International Journal of Approximate Reasoning*, 45(1): 82-105.

Baudrit, C, Dubois, D & Guyonnet, D (2006) Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. *IEEE Transactions on Fuzzy Systems*, 14(5): 593-608.

Cooper, JA, Ferson, S & Ginzburg, L (1996) Hybrid processing of stochastic and subjective uncertainty data. *Risk Analysis*, 16 (6): 785-791.

De Cooman, G & Aeyels, D (1999) Supremum preserving upper probabilities. *Information Sciences*, 118: 173-212.

Dubois, D (2006) Possibility theory and statistical reasoning. *Computational Statistics & Data Analysis*, 51, 47–69.

Dubois, D (2010) Representation, propagation and decision issues in risk analysis under incomplete probabilistic information. *Risk Analysis*, 30: 361-368.

Dubois, D, Kerre, E, Mesiar, R & Prade, H (2000) Fuzzy interval analysis. In: Dubois, D & Prade, H (eds) *Fundamentals of Fuzzy Sets*. Kluwer: Boston, Mass. The Handbooks of Fuzzy Sets Series, 483-581.

Dubois, D, Foulloy, L, Mauris, G & Prade, H (2004) Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10: 273-297

Dubois D., Prade H. (1991) Random sets and fuzzy interval analysis. Fuzzy Sets and Systems, 42(1), 87-101

Dubois D., Prade H. (2009) Formal representations of uncertainty. In: Decision-making Process- Concepts and Methods (D. Bouyssou, D. Dubois, M. Pirlot, H. Prade, Eds.), ISTE London & Wiley, Chap. 3, p. 85-156.

Dubois D., Prade H., Smets P. (1996) Representing partial ignorance. IEEE Trans. on Systems, Man and Cybernetics, 26(3), 361-377.

Guyonnet, D, Bourgine, B, Dubois, D, Fargier, H, Côme, B & Chilès, JP (2003) Hybrid approach for addressing uncertainty in risk assessments. *Journal of Environmental Engineering*, ASCE, 129: 68-78.

Hacking, I (1965) *Logic of Statistical Inference*. Cambridge University Press, Cambridge, UK.

Jacob, C, Dubois, D & Cardoso J (2011) Uncertainty handling in quantitative BDD-based fault-tree analysis by interval computation. Int. Conf. on Scalable Uncertainty Management (SUM 2011, Dayton, Ohio), (S. Benferhat, J. Grant, Eds.), Springer, Lecture Notes in Computer Science 6929, pp. 205-218.

Kaplan, S (1988) Will the real probability please stand up? *Reliability Engineering and System Safety*, 23(4): 285-292.

Kaplan, S (1997) The words of risk analysis. *Risk Analysis*, 17: 407-417.

Kaplan, S (1992) Formalism for handling phenomenological uncertainties: the concepts of probability, frequency, variability, and probability of frequency. *Nuclear Technology*, 102: 137-142.

Kaplan, S & Garrick, BJ (1981) On the quantitative definition of risk. *Risk Analysis*, 1(1): 11-27.

Limbourg P. , Savic R., Petersen J. & Kochs H.-D. (2007) Fault tree analysis in an early design stage using the Dempster-Shafer theory of evidence, *Risk, Reliability and Societal Safety* – Aven & Vinnem (eds), Taylor & Francis Group, London, 713-722.

Lindley, DV & Phillips, LD (1976) Inference for a Bernoulli process (a Bayesian view), *The American Statistician*, 30(3): 112-119.

Lodwick, W. A. (2008) Fundamentals of interval analysis and linkage to fuzzy set theory, in: W. Pedrycz, A. Skowron, V. Kreinovich (eds.), Handbook of Granular Computing, John Wiley & Sons Ltd..

Mauris, G (2008) Inferring a possibility distribution from very few measurements. In: D. Dubois et al., Eds, *Soft Methods for Handling Variability and Imprecision*, (SMPS 2008): Advances in Soft Computing , Springer 2008, 92-99.

Montgomery, V (2009) *New statistical methods in risk assessment by probability bounds*. PhD thesis, Durham University. http://www.npi-statistics.com/vm-thesis.pdf

Moore, R (1979) *Methods and Applications of Interval Analysis*. SIAM Studies in Applied Mathematics, Philadelphia.

Paté-Cornell, ME (1996) Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering and System Safety*, 54: 95-111.

Ross, SM (1996) *Stochastic Processes*. 2nd ed. Wiley.

G. Shafer (1976) A Mathematical Theory of Evidence . Princeton University Press.

Stanford Encyclopedia Philosophy (SEP) (2011) Interpretations of probability. http://plato.stanford.edu/entries/probability-interpret/.

Singpurwalla, ND (2006) *Reliability and Risk: A Bayesian Perspective*. Chichester: Wiley.

Singpurwalla, ND & Wilson, AG (2009) Probability, chance and the probability of chance. *IIE Transactions*, 41: 12-22.

Vinnem, JE (2007) *Offshore Risk Assessment: Principles, Modelling and Applications of QRA Studies*. 2nd ed. London: Springer.

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

Weichselberger, K (2000) The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24: 149-170.

Zadeh, LA (1975) The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences*, 8: 199-249.