On the use of imprecise probabilities in reliability

F.P.A. Coolen
Department of Mathematical Sciences
University of Durham, Durham, DH1 3LE, UK
Frank.Coolen@durham.ac.uk

Abstract

Theory of imprecise probability generalizes classical probability theory, by assigning to each event an interval instead of a single number. In this paper, we briefly discuss this generalization and some recently suggested applications of imprecise probabilities in reliability. We also comment on challenges for research and applications.

1 Introduction

Most reliability models which include aspects of uncertainty use probability theory to quantify information about these uncertain aspects. For example, to take a future lifetime T of a component into account in a model for system reliability, a probability distribution for T can be used, which e.g. can be assumed to belong to a certain class of parametric distributions such as Weibull, Gamma, or Lognormal. Such a choice of a single distribution is often highly subjective, even if there is a fair amount of relevant data available there are still many possible distributions which are equally well supported by the data. Several authors (e.g. [1, 5, 25]) have suggested that data are often sparse in reliability applications, if available at all, so quantification of uncertain aspects should be based on subjective information, i.e. experts' judgements. Such judgements are then to be translated into, again, a single probability distribution for the uncertain aspects of the model, with the same problem that many distributions could be used without the possibility to distinguish between these.

Central to these quantifications of uncertainty, either based on data, experts' judgements, or a combination of both, is the restrictive assumption of using precise probabilities and probability distributions. During the past few decades there has been increasing attention to generalized uncertainty quantification, including concepts such as fuzzy sets and belief functions, for further references to such concepts, and discussion from the perspective of generalized probability, see [32, 33].

One such a concept is called 'imprecise probability' [32], also known as 'interval probability' [34, 35], and in recent years this has particularly been a growing area of research, as it is realized that it provides a consistent theory for generalized uncertainty quantification (which to some extend coincides with belief functions and other concepts). Researchers with widely varying backgrounds are currently contributing to theory, and indeed applications, of imprecise probability, including mathematicians, statisticians, computer scientists, and researchers working on artificial intelligence, medicine, and a variety of engineering areas. Such researchers are brought together via the Society for Imprecise Probability Theory and Applications (SIPTA, *www.sipta.org*), which also organizes biennial conferences.

In this paper, we discuss possible use of imprecise probabilities in reliability. We explain some methods recently presented in the literature, without aiming to provide a full literature survey, and outline several topics for future research to enable applications of imprecise probability in this field. Section 2 provides a concise informal introduction to imprecise probability. In Section 3 we consider the use of imprecise probability in reliability by discussing some methods and models presented in the literature. Finally, in Section 4 we discuss some challenges for research and applications. Throughout the paper we refer to other sources for details on theory and further examples.

2 Imprecise probability

Imprecise probability generalizes classical probability in the sense that uncertainties about events are quantified via intervals, instead of single numbers. For an uncertain event A, where classical probability theory would require one value $P(A)$, we now assign an interval $[\underline{P}(A), \overline{P}(A)]$, with $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$, where $\underline{P}(A)$ is called the *lower probability* for event A, $\overline{P}(A)$ the *upper probability* for event A, and $\Delta(A) = \overline{P}(A) - \underline{P}(A)$ is called the *imprecision* for event A. When introducing such a generalization, careful attention is required to interpretation, axioms, theorems, further concepts, and development of related statistical models and corresponding inference, including decision making. In addition, insight into advantages and disadvantages, when compared to the established theory of precise probabilities, is required, in particular to convince practitioners to start using the new concepts and methods. It is amusing that quite many mathematicians and statisticians have expressed fairly strong views against the use of imprecise probability, see e.g. [23], whereas the onus of justification is clearly on the side of using the far more restrictive precise probabilities, which are a special case of imprecise probability.

There are several possible interpretations of classical probabilities, and the same holds for imprecise probabilities. A relatively straightforward interpretation of such an interval is in terms of 'all possible values' for a precise probability, so all values one has not yet been able to exclude. This implies that one thinks that A does have a precise probability $P(A)$, which happens not to be precisely known, but it is known (or at least one strongly believes) that it is within $[\underline{P}(A), \overline{P}(A)]$. Many researchers in fields related to uncertainty prefer a subjective interpretation of probability, which also underlies Bayesian statistics, and where a precise probability $P(A)$ is defined in terms of a personal 'fair price' for a bet on event A, loosely speaking such that you consider £$P(A)$ the fair price for a bet that pays £1 if A occurs, and nothing otherwise (formally, some unit of linear utility is required instead of pounds [32] to avoid influence of personal attitudes towards risk). Generalizing this to imprecise probability, one distinguishes between the maximum price for which one would wish to buy this bet, which is $\underline{P}(A)$, and the minimum price for which one would wish to sell this bet, which is $\overline{P}(A)$. The idea is that, if one has a lot of information relevant to the uncertainty of A, one might be able to determine upper and lower probabilities which are close to each other, whereas if one hardly has any such information, one may wish to assess lower and upper probabilities close to 0 and 1, respectively. Clearly, this allows different levels of expertise underlying subjective information to be taken into account. For example, an experienced engineer might confidently assess lower and upper probability of 0.7 and 0.8, respectively, for the event that a particular unit will not fail during its first year in use, whereas a less experienced engineer, for example knowing less relevant facts about the process and environment in which the unit is used, might only feel confident to assess the values 0.5 and 0.9 for the same event. If two experts would assign 'conflicting values', in the sense that one expert's upper probability for an event is less than the other expert's lower probability for the same event, then this is indeed very useful to know (note that, when restricting to precise probabilities, it is very likely that they would assess different values, so it is less clear that there may be 'conflict'). In such a situation, these values would, from a subjective interpretation point of view, imply that they could exchange a bet such that both would find this favourable. More practically, however, noticing such difference in experts' imprecise probabilities might point towards them having different information or knowledge available on which to base their quantifications, in such cases discussion of the reasons for their assessments might be very useful in reliability applications, we return to this issue in Subsection 3.1.

During the last two decades, a variety of generalizations of classical probability theory have been presented, including axiom systems and important further concepts and theorems, for example enabling corresponding statistical inference and decision making. It appears that, after several concepts were developed more or less in parallel early on, many researchers now agree that the most complete framework is offered by Walley's [32] theory of imprecise probability, and, closely related to this, Weichselberger's [34, 35] theory of interval probability, where the latter theory is more directly formulated as a generalization of Kolmogorov's classical probability axioms, and puts less emphasis on development via coherent subjective betting behaviour than Walley's theory. Without wishing to go into details on axioms and further concepts, it should be remarked that these are far more complex than for precise probabilities. For example, the important concept of conditional probability, which e.g. plays a crucial role in Bayesian statistics, does not have a unique generalization to imprecise probability, where conditioning can be done via several differing concepts [2, 34].

Statistical inference based on imprecise probability offers a wide range of modelling opportunities, among these are well-known models from robust (Bayesian) statistics [3, 18], which however have a different interpretation from imprecise probability perspective, with differences between upper and lower bounds for inferences fundamentally quantifying indeterminacy in the information used for the quantification of the uncertain events of interest, which goes beyond the classical sensitivity interpretation (which, of course, can still be of use). In Section 3 we will briefly show some possibilities for inference based on imprecise probabilities, including attention to decision making (Subsection 3.4). There, we also comment on advantages and disadvantages of the use of imprecise probabilities for quantification of uncertainties in reliability, including attention to elicitation (Subsection 3.1).

3 Imprecise probability in reliability

There is a wide variety of reasons why imprecise probability might be of particular relevance in the area of reliability. A main reason is the fact that, in many reliability applications, there may be few, if any, statistical data available, implying stronger dependence on subjective information in the form of expert judgements [1, 5, 25]. Apeland, *et al.* [1] suggest a fully subjective approach, without even considering statistical data, precisely because of this reason. However, they still rely on precise probabilities, where it is not clear how imperfections in the experts' uncertainty quantifications can be dealt with. Such imperfections may, for example, occur due to constraints on time and resources to elicit all probabilities of interest, inconsistency of elicited subjective probabilities on related events, or the use of information from a group of possibly disagreeing experts. This last issue also raises interesting questions on combination of subjective probabilities from a group of experts, which we will address in Subsection 3.1.

A second reason is the relaxation of dependence on precise statistical models justified by physical arguments. Although there is some overlap with the issues relating to expert judgements, some probability distributions can be assumed based on physical properties, e.g. the shape of the Weibull distribution [22] for the lifetime of a unit might be assumed a known constant based on knowledge of the ageing process of a unit, e.g. shape parameter 1, giving the exponential distribution, relates to no ageing effects, and shape parameter 2 leads to linearly increasing hazard rate (a Rayleigh distribution), where roughly speaking, if unit A is twice as old as unit B, it is twice as likely to fail within the next short time period. In practice, there may indeed be relevant ageing knowledge backed up by physical reasons, yet this knowledge may not be strong enough to justify restriction to a single statistical model. We will address this issue a bit further in Subsection 3.2.

A third reason may be encountered in study of system reliability. An assumption underlying most mathematical work in this field is that the exact system structure and dependence relations between components are known, which may well be unrealistic in many applications for all but the simplest systems. We address this issue in Subsection 3.3.

There are more reasons suggesting benefits of using imprecise probability in reliability, e.g. the nature of data collection such as grouped data [7, 13], but we reckon that the ones we discuss explicitly in this paper are strong enough to emphasize the possible benefits of using imprecise probabilities in reliability. In Subsection 3.4 we briefly address the important aspect of actually making decisions when using imprecise probabilities, which is often the ultimate goal of analyses of uncertainty in reliability.

3.1 Expert judgements: elicitation and combination

Many authors have presented strong cases for the use of expert judgements in reliability applications which involve uncertainty (e.g. [1, 5]). Crucial to inclusion of such information in probabilistic or statistical models is proper quantification of such judgements, a process often called 'elicitation'. When interest is in the uncertain time to failure of a technical unit, an often used approach is to assume a particular convenient family of probability distributions for this random quantity, e.g. Weibull distributions with two unknown parameters, and ask an expert just enough (i.e. two in this case) characteristics to choose a particular Weibull distribution. For example, one might ask the expert's judgement on the expected value and variance of this distribution, or two different quantiles. At a more sophisticated level, one might try to use a Bayesian approach with probability distributions for the parameters, but effectively this just adds one more layer of model assumptions, as typically expert's judgements are still elicited via a few summaries [5, 25]. Whether or not the full resulting probability distribution carefully reflects the expert's judgements is often not clear, if one would elicit more characteristics than the number of parameters to be chosen then there would almost certainly not be a precisely fitting probability distribution in the chosen family of distributions.

Imprecise probabilistic methods allow as many characteristics of probability distributions to be elicited as the expert, or the statistician working on the problem, feels appropriate to carefully quantify the expert's judgements. Of course, if different such quantifications are implicitly contradictory this should be pointed out, but otherwise a set of probability distributions can be used such that all the elicited values correspond to some of the probability distributions in this set. In addition, and perhaps the most important advantage, such methods do not require the expert to quantify his judgements via single numbers for characteristics of the probability distribution, explicitly allowing indeterminacy which could possibly reflect the confidence that the expert has in his knowledge with regard to the particular random quantity of interest.

For example, suppose that for a reliability study it is important to model the uncertainty about the lifetime X (in years) of a particular unit, and that two experts are consulted. Let $\underline{P}_i(A)$ and $\overline{P}_i(A)$ be the lower and upper probabilities for event A, respectively, as assessed by expert $i = 1, 2$. Suppose that the following values have been elicited: $\underline{P}_1(X > 1) = 0.9$, $\overline{P}_1(X > 1) = 1.0$, $\underline{P}_1(X > 2) = 0.8$, $\overline{P}_1(X > 2) = 0.9$, $\underline{P}_2(X > 1) = 0.8$, $\overline{P}_2(X > 1) = 0.95$, $\underline{P}_2(X > 2) = 0.3$, $\overline{P}_2(X > 2) = 0.7$. With such elicited values, it is possible to use sets of lifetime distributions per expert, such that these sets correspond to these elicited values. For example, for expert 1 all Weibull distributions with probability of surviving 1 year between 0.9 and 1.0, and of surviving 2 years between 0.8 and 0.9, could be found, and used for inference.

It may also be of interest to combine the judgements from both experts. First, however, note that, from a betting interpretation perspective, these experts disagree about survival past 2 years to the extent that they would be happy to exchange a bet, which follows from the fact that expert 2's upper probability is less than expert 1's lower probability for the event $X > 2$. For example, a bet that would pay £1, say, if the unit actually survives 2 years, and nothing else, could be exchanged for a price of £0.75, such that expert 1 pays expert 2 £0.75 now, getting the bet in return, and both experts are eager to do so as they expect a profit. Such differences in opinion might be important to notice, whereas when restricting to precise probabilities it would be a mere coincidence if the two experts would exactly agree.

If one wishes to combine the expert judgements, there is a variety of possibilities. Most obviously, we could define a combined lower (upper) probability for A as the minimum (maximum) of the individual lower (upper) probabilities for A. This could be interpreted as a cautious approach, as all experts would agree with the resulting value being a lower (upper) bound of the probability, or, in terms of the betting interpretation, all experts would indeed support buying the bet for this combined lower probability. Obviously, this might lead to fairly large differences between corresponding upper and lower probabilities. An alternative would be to take the maximum (minimum) individual lower (upper) probability as the combined lower (upper) probability, but this might lead to incoherent probabilities. For example, with the two experts above, this would lead to lower probability 0.8 of surviving past 2 years, but upper probability 0.7 for the same event, which is not acceptable for obvious reasons. Such combination rules are discussed in more detail by, for example, Walley [32] and Kozine and Filimonov [20]. Coolen [6] also discusses the possibility of taking weighted averages of individual lower (upper) probabilities as the combined lower (upper) probability, where weights could reflect the expertise of the individuals, generalizing weighted averaging of precise subjective probabilities [5], but theory into appropriate definitions for such weights has not yet been developed for imprecise probabilities. There are good opportunities to use calibration of the experts [5] related to such weights, and one route towards sensible weights for experts for combining imprecise probabilities might be to actually relate weights to the betting interpretations of lower and upper probabilities, as budgets of 'good experts' would tend to increase, whereas those of 'bad experts' would tend to decrease over time if they were actually buying and selling bets. Finally, we need to warn about any simple rule on combining quantified experts' judgements, as resulting combined imprecise probabilities may not take into account that the information of some of the experts may not be independent. For precise probabilities some progress has been made on dealing with this aspect [27], it may well be that imprecise probabilities are again better suited to deal with this but, as far as we are aware, this has not yet been reported in the literature.

Imprecise probabilities also naturally occur when looking for statistical inferential methods with a minimum of expert judgements included, to the point that one also wishes to reduce structural modelling assumptions. Of course, it remains necessary to assume some mathematical structure, to link random quantities on future observations to past observations, which can be achieved by a post-data assumption related to exchangeability, leading to so-called nonparametric predictive inference [2, 9, 14], brief examples of such inference are presented in Subsections 3.2 and 3.4.

Finally, with ever growing computational powers, statisticians have successfully developed models allowing very many random quantities, with careful representation of dependence structures between these quantities. For example, Bayesian graphical models [15, 19] have proven successful for large scale applications [26, 36]. Typically, the elicitation task for such models is enormous, as it is exponential in the number of quantities included in the model and these models normally add unobservable random quantities ('parameters') to the observable random quantities. For application of such models, elicitation tends to suffer from serious time constraints, so methods are needed that can leave many probabilities unspecified, or at best only partially specified. Such methods have been suggested during the past few years, see e.g. Cozman [16], and are being developed further, where main problems involve development of fast optimisation algorithms combined with the algorithms for learning in such models. Studying imprecision in such models is also likely to be of use for indicating on which parts of the models it is best to focus the elicitation effort, but methods for this have not yet been fully developed.

3.2 Lifetime models and inference

For statistical inference in reliability, with imprecise probabilities, statistical models for lifetimes are required. Several models have been suggested explicitly based on imprecise probabilities, we briefly discuss a few below. In addition, there is a rich literature on robust statistics, both Bayesian [3] and frequentist [18], suitable models presented there can also be used within imprecise probability theory [10]. This is generally discussed by Walley [32].

A particularly interesting class of lifetime distributions, leading to imprecise reliability inferences, has recently been proposed by Utkin and Gurov [30]. For a non-negative random lifetime T, with cumulative hazard function $H(t) = \int_0^t h(x)dx$, where $h(\cdot)$ is the hazard rate, they define the class of distributions $\mathcal{H}(r, s)$, for $0 \leq r \leq s \leq \infty$, as those distributions for which $H(t)/t^r$ increases and $H(t)/t^s$ decreases for all t. Such classes can contain a wide variety of distributions [30], e.g. $\mathcal{H}(1, \infty)$ is the class of all distributions with increasing hazard rate, $\mathcal{H}(0, 1)$ those with decreasing hazard rate, while bath-tub shaped hazard rates might be embedded in such classes by choosing $r < 1 < s$. Furthermore, a Weibull distribution with shape parameter β belongs to all such classes with $r \leq \beta \leq s$, and a Gamma distribution with shape parameter k belongs to $\mathcal{H}(1, k)$ for $k \geq 1$ and to $\mathcal{H}(k, 1)$ for $k \leq 1$. Utkin and Gurov discuss in great detail the optimisation problems related to calculation of upper and lower bounds for reliability characteristics corresponding to such classes of distributions, where they particularly focus on system reliability. For example, if n independent components, where component i has lifetime distribution belonging to $\mathcal{H}(r_i, s_i)$, form a series system, then the system lifetime distribution belongs to $\mathcal{H}(\min r_i, \max r_i)$, while if they form a parallel system it belongs to $\mathcal{H}(\min r_i, \sum s_i)$.

Such classes of distributions allow partial information to be taken into account without requiring many additional assumptions. To make practical use of such models more attractive, statistical theory for such models must be developed, e.g. on fitting such classes to available data. Furthermore, generalizations may be useful, for example by partitioning the time axis and defining separate such classes on each interval.

Nonparametric predictive methods [9] enable lifetime inferences with a minimum of subjective assumptions added to data. For example, Coolen and Yan [12] present predictive upper and lower survival functions for a future lifetime T_{n+1}, based on observations of n such previous lifetimes, which may include right-censored observations. This gives an alternative to the well-known product-limit estimate by Kaplan and Meier (see e.g. [22]), and to the predictive method by Berliner and Hill [4], which is based on similar foundations as the Coolen-Yan method, but cannot deal with exact censoring information. For example, the following data are part of an example discussed by Coolen and Yan [12], who provide details on context and the original source. The 16 observations are (in days, t^* denotes right-censoring):

$$90, 142, 150, 269, 291, 468^*, 680, 837, 890^*, 1037, 1090^*, 1113^*, 1153, 1297, 1429, 1577^*$$

Figure 1 gives the lower and upper survival functions for X_{17} according to the method developed by Coolen and Yan [12], together with the Kaplan-Meier and Berliner-Hill alternatives. The use of imprecision enables clear indication of the moments where right-censoring takes place, at which the lower survival function decreases. The two alternative methods do not clearly indicate the immediate effects of right-censoring. Also, it seems natural to interpret right-censoring as a loss of information, which is reflected by greater imprecision.

Applications of such nonparametric predictive inference to replacement problems have been presented [8, 11], results for age replacement are presented at this conference [14]. Coolen and Yan [13] present such inference for grouped lifetime data, e.g. life tables, where only numbers of observed events and right-censorings per interval for a partition of the time-axis are given. Here, imprecision results from three different sources, namely the few assumptions added to the data, right-censoring, and the fact that the data are not exactly observed. Although these are quite different reasons for imprecision, indeed it appears that all cause some reduced level of information which is rightly reflected by increased imprecision. An alternative method for such grouped lifetime data, also using imprecise probability but closer in nature to a robust Bayesian approach, was presented by Coolen [7].

3.3 Systems

In the previous subsection we briefly mentioned reliability results for systems corresponding to the distribution classes by Utkin and Gurov [30]. Utkin and Gurov [28, 29], Kozine and Filimonov [20], and Utkin and Kozine [31] have presented a variety of useful results on system reliability with imprecise probability, mostly by

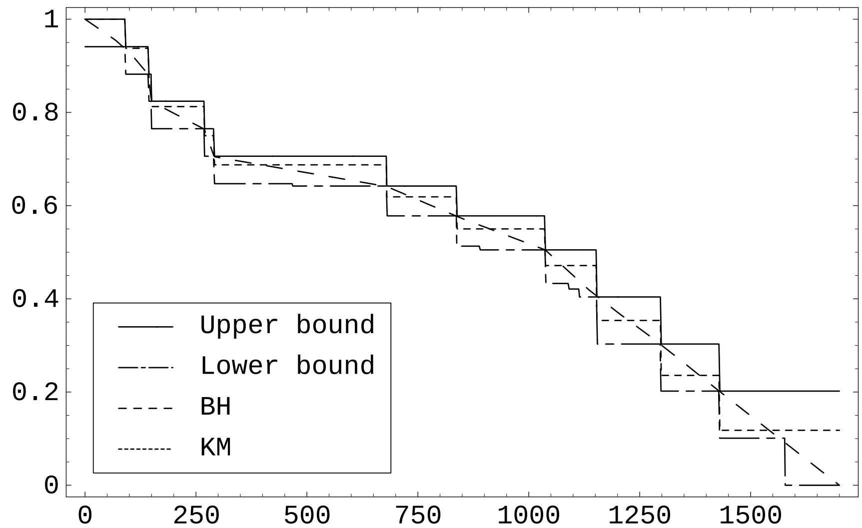

Figure 1: Survival functions; upper and lower, Berliner-Hill, and Kaplan-Meier.

solving the often complex optimisation problems to find optimal bounds for the reliability characteristics corresponding to classes of reliability probabilities or lifetime distributions per component, where the classes are often only defined by rather few specified characteristics, and dependence structures of components in the systems may not be fully known. Often, however, bounds for system reliability characteristics become very wide if no knowledge at all is assumed about the dependence of the components' lifetimes, making such methods of less practical relevance for larger systems. It is important that the mathematical methodology, provided by the work of these authors, shows that larger scale applications are possible for as far as computation is concerned, since computational complexity is a main disadvantage of imprecise probability, as often complex constrained optimisation problems need to be solved. In general for such methods, if bounds become very wide it is due to the restricted information and assumptions put into the models. An interesting direction of extending such work is the study of possibilities to model, with imprecision, varying forms of partial knowledge about components' dependencies, and possibly also of system structure for large systems. In practice, it is often not likely that experts have no useful information on dependence of components' lifetimes, yet indeed they may not have complete knowledge or time to fully reflect on this and quantify all their knowledge.

3.4 Reliability decisions

Imprecise probabilities, or corresponding statistical inferences, are often not the final goal of reliability analyses. In many practical situations, such analyses are performed because decisions are required, e.g. on replacement of a unit or release of a system. If imprecise probabilities reflect indeterminacy, then there is often a natural justification to use either the lower or upper probability from a cautious perspective. Of course, if all probabilities between the lower and upper probability would lead to (about) the same decision, such a decision is strongly supported and quite robust with regard to the aspects about which we did not have perfect information. On the other hand, if decisions corresponding to such probabilities would vary widely, then this would indicate that the information at hand may not be sufficient to suggest a clearly optimal decision. This could be regarded as a disadvantage, but seems a realistic reflection of the fact that on some occassions more information, or more modelling assumptions, may be required. Decision theory is well developed for models in robust (Bayesian) statistics, which can also be employed for several imprecise probability models.

There are situations in which it is natural to focus only on either a lower or upper probability. For safety critical systems, one may for example only wish to focus on lower probabilities of zero failures. For example, nonparametric predictive inference [9] gives imprecise probabilities for r failures in m future observations (e.g. not functioning of a safety critical system when required), based on n tests without failures, as (using straightforward notation):

$$\overline{P}((m,r)|n) \;=\; \frac{\binom{m}{r}}{\binom{n+m}{r}},$$

and

$$\underline{P}((m,r)|n) = \begin{cases} \frac{n}{n+m} & \text{for } r = 0, \\ 0 & \text{for } r > 0. \end{cases}$$

This upper probability for $r = 0$ is equal to 1, for all n and m, which seems reasonable as tests without observed failures imply that there is no clear evidence against the possibility that the system might be without faults. However, the lower probability of $r = 0$ failures in m future observations, based on n tests with 0 failures, is probably most relevant from a cautious point of view. This lower probability is equal to $n/(n+m)$, which decreases with m. Suppose that one requires fairly high reliability, say lower probability of at least 0.99 to have zero failures in m future observations. Let us assume that the safety critical system is only approved if n tests have shown zero failures, else faults will be removed leading to a test-fix-test situation which could only be modelled by including further structural assumptions, which we do not consider here. This implies that the required number of tests to achieve this reliability would be $n \geq 99m$, and no test should reveal a failure. Clearly, this implies that the number of future occassions on which the system is required to function successfully, directly influences the required number of successful tests. If the safety system only needs to be used (at most) once after testing, so $m = 1$, then $n = 99$ tests without failure would be sufficient. Indeed, a large number of tests is required to achieve reasonable confidence in such high reliability, when expressing reliability in terms of the lower probability, reflecting that there are only few mathematical assumptions in this approach. We believe that both such a high number of required successful tests, and the dependence of this number of tests on the number of future observations, are fully in line with intuition of engineers, unless their engineering expertise is explicitly taken into account. A reliability requirement, such as the lower probability of at least 0.99 in this example, will typically be based on a variety of factors, such as costs, environmental factors, or risk to human life.

4 Challenges for research and application

In the sections above several research challenges have already been mentioned, including aspects with regard to large statistical models, further development of lifetime models, and combination of experts' judgements. Clearly, practical methods for elicitation of imprecise probabilities must be developed, where possible with user-friendly elicitation packages and interfaces allowing judgement quantifications at a level the experts feel comfortable with, and with a variety of feedback possibilities enabling detailed understanding of the consequences of assessed probability quantifications.

As briefly mentioned before, main difficulties for applications of statistical models with imprecise probabilities are with regard to computation. Indeed, for many inferences complex constrained optimisation problems need to be solved to calculate the optimal bounds of inferential measures relating to classes of probability distributions, where also these classes may be defined via some constraints on summaries of these distributions. The recent work by Utkin and co-authors has provided great progress on this aspect, yet much more needs to be done. For example, modern Bayesian methods often require simulation-based computational methods [17], and it is not immediately clear, from theoretical perspective, how such methods could be generalized to allow imprecise probabilities, let alone how to actually develop algorithms for such computations. Interesting theoretical results on Markov chains with imprecise probabilities have recently been presented by Kozine and Utkin [21], and these may well provide a step in the direction towards such simulation-based computational methods. In addition, for many reliability applications Markov models have been used successfully, e.g. to describe state transitions in maintenance models, and the results by Kozine and Utkin form a basis for further research on such models with imprecise probabilities.

A topic that has, as far as we know, not yet been studied at all is design of experiments with uncertainty quantified via imprecise probabilities. In such research, it is likely that well-known design optimisation criteria should be combined with the aim to reduce imprecision optimally by taking observations at points where large imprecision may prevent clear decisions to be taken. In reliability, such research could for example be relevant for design of accelerated life tests (e.g. [24]), where imprecision might realistically reflect indeterminacy with regard to modelling of the acceleration.

There is no doubt that actual applications of imprecise probability methods in reliability will bring to light many more interesting research problems, and require a wide variety of research skills to tackle varying problems. From this perspective, the wide range of backgrounds of researchers active in this field may be of great benefit for future development. We strongly hope that reliability engineers will collaborate with statisticians in the development of models and methods, to ensure applications in a field where uncertainty often plays a key role in decision making.

References

[1] Apeland, S., Aven, T. and Nilsen, T. (2002). Quantifying uncertainty under a predictive, epistemic approach to risk analysis. *Reliability Engineering and System Safety*, **75**, 93-102.

[2] Augustin, T. and Coolen, F.P.A. (2003). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, to appear.

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

[4] Berliner, B.H. and Hill, B.M. (1988). Bayesian nonparametric survival analysis. *Journal of the American Statistical Association*, **83**, 772-784.

[5] Cooke, R.M. (1991). *Experts in Uncertainty*. Oxford University Press, New York.

[6] Coolen, F.P.A. (1994). *Statistical Modeling of Expert Opinions Using Imprecise Probabilities*. PhD-Thesis, Eindhoven University of Technology (available via the author).

[7] Coolen, F.P.A. (1997). An imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations. *Reliability Engineering and System Safety*, **56**, 61-68.

[8] Coolen, F.P.A. and Coolen-Schrijner, P. (2000). Condition monitoring: a new perspective. *Journal of Operational Research Society*, **51**, 311-319.

[9] Coolen, F.P.A., Coolen-Schrijner, P. and Yan, K.J. (2002). Nonparametric predictive inference in reliability. *Reliability Engineering and System Safety*, **78**, 185-193.

[10] Coolen, F.P.A. and Newby, M.J. (1994). Bayesian reliability analysis with imprecise prior probabilities. *Reliability Engineering and System Safety*, **43**, 75-85.

[11] Coolen, F.P.A. and Newby, M.J. (1997). Guidelines for corrective replacement based on low stochastic structure assumptions. *Quality and Reliability Engineering International*, **13**, 177-182.

[12] Coolen, F.P.A. and Yan, K.J. (2002). Nonparametric predictive inference with right-censored data. *In submission*.

[13] Coolen, F.P.A. and Yan, K.J. (2003). Nonparametric predictive inference for grouped lifetime data. *Reliability Engineering and System Safety*, to appear.

[14] Coolen-Schrijner, P. and Coolen, F.P.A. (2003). Nonparametric predictive inference for age replacement with a renewal argument. *Proceedings 15th ARTS.*

[15] Cowell, R.G., Dawid, A.P., Lauritzen, S.L. and Spiegelhalter, D.J. (1999). *Probabilistic Networks and Expert Systems.* Springer, New York.

[16] Cozman, F.G. (2000). Credal networks. *Artificial Intelligence*, **120**, 199-233.

[17] Gamerman, D. (1998). *Markov Chain Monte Carlo.* Chapman and Hall, London.

[18] Huber, P.J. (1981). *Robust Statistics.* Wiley, New York.

[19] Jensen, F.V. (2001). *Bayesian Networks and Decision Graphs.* Springer, New York.

[20] Kozine, I.O. and Filimonov (2000). Imprecise reliabilities: experiences and advances. *Reliability Engineering and System Safety*, **67**, 2000, 75-83.

[21] Kozine, I.O. and Utkin, L.V. (2002). Interval-valued finite Markov chains. *Reliable Computing*, **8**, 97-113.

[22] Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data.* Wiley, New York.

[23] Lindley, D.V. (1996). Contribution to the discussion of Walley (1996). *Journal of the Royal Statistical Society, Series B*, **58**, 47-48.

[24] Meeker, W.Q. and Escobar, L.A. (1998). *Statistical Methods for Reliability Data.* Wiley, New York.

[25] Percy, D.F. (2002). Bayesian enhanced strategic decision making for reliability. *European Journal of Operational Research*, **139**, 133-145.

[26] Rees, K., Coolen, F.P.A., Goldstein, M. and Wooff, D.A. (2001). Managing the uncertainties of software testing: a Bayesian approach. *Quality and Reliability Engineering International*, **17**, 191-203.

[27] Smith, J.Q. and Faria, A.E. (2000). Bayesian Poisson models for the graphical combination of dependent expert information. *Journal of the Royal Statistical Society, Series B*, **62**, 525-544.

[28] Utkin, L.V. and Gurov, S.V. (1999). Imprecise reliability of general structures. *Knowledge and Information Systems*, **1**, 459-480.

[29] Utkin, L.V. and Gurov, S.V. (2001). New reliability models based on imprecise probabilities. In: C. Hsu (ed.), *Advanced Signal Processing Technology*, Chapter 6, 110-139. World Scientific.

[30] Utkin, L.V. and Gurov, S.V. (2002). Imprecise reliability for some new lifetime distribution classes. *Journal of Statistical Planning and Inference*, **105**, 215-232.

[31] Utkin, L.V. and Kozine, I.O. (2002). Computing system reliability given interval-valued characteristics. *In submission.*

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

[33] Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, **24**, 125-148.

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

[35] Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung: I. Intervallwahrscheinlichkeit als umfassendes Konzept.* Physika, Heidelberg (in German).

[36] Wooff, D.A., Goldstein, M. and Coolen, F.P.A. (2002). Bayesian graphical models for software testing. *IEEE Transactions on Software Engineering*, **28**, 510-525.