Safety, Reliability and Risk Analysis: Beyond the Horizon – Steenbergen et al. (Eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00123-7

Are interval probabilities a viable way for quantifying uncertainties?

K. Breitung
Engineering Risk Analysis Group, Technical University of Munich, Germany

ABSTRACT: In the last decade several authors propagated the use of interval probabilities as alternative to Bayesian models in reliability problems. The basic idea of this approach is to start from some lower and upper bounds for functions of random variables describing the failure probabilities or rates of the components of a system and then to derive from these then bounds for the failure probability of the system. The advantage of such bounds is that there are no classical or Bayesian confidence probabilities, one is 100% certain that the calculated probabilities lie in the derived bounds.

If one considers the basic problem in reliability of finding the failure probability, this can be seen as collecting information, one starts from total ignorance and gathering more and more information one arrives at more specific estimates of the probability. Using the mathematical definition of entropy and information, here it is shown that the method of interval probabilities requires an infinite amount of data as prerequisite. A prerequisite which in halfway realistic problems cannot be fulfilled.

1 INTRODUCTION

Modelling in reliability problems is done mainly with probabilistic models. One point that causes often uneasiness in people used to deterministic thinking is that there is almost never 100% percent certainty in probabilistic models. In both conceptions of probability, classical or Bayesian, after deriving confidence intervals, there is always a remaining risk that something is outside the bounds one has found. The larger the confidence probability is taken, the larger the corresponding confidence intervals will become.

There have been attempts to develop structural reliability methods which avoids such probability statements, as an example see (Ben-Haim 1996). Avoiding probability statements seems appealing, since so all these fine points of probability theory are not needed and don't have to be explained. On the other hand there is the danger that such schemes are used especially because of these benefits without seeing serious shortcomings in them.

One of these alternative concepts competing with the Bayesian approach is the interval probability method. Here it seems that one has suddenly no more confidence intervals with some probability content less than 100%, but absolute certainty that the parameters of the system are contained in the derived intervals. It remains the question, how this can be achieved. This problem will be examined here.

2 BASICS OF THE INTERVAL PROBABILITY METHOD

In reliability problems the usual form is that a number of random variables (rv's) $X_1,\ldots,X_n$ is given and a function g which describes the state of the system such that if $g(X_1,\ldots,X_n) \geq 0$ the system is intact and if $g(X_1,\ldots,X_n) < 0$ the system is defect. The probability of failure $P(g(X_1,\ldots,X_n) < 0)$ has to be estimated.

Here a very short outline of the interval probability method for reliability problems following the papers by Coolen (2004) and Utkin & Coolen (2007) is given. In reliability problems the usual form is that a number of random variables $X_1,\ldots,X_n$ is given and a function g which describes the state of the system such that if $g(X_1,\ldots,X_n) \geq 0$ the system is intact and if $g(X_1,\ldots,X_n) < 0$ the system is defect.

The starting point of such methods is that for random variables which characterize in some way the failure probabilities of system components some upper and lower bounds are given. As an example consider a system consisting of n components where the behavior of each component is described by a random variable X_i, $i = 1,\ldots,n$. Further there are given functions $f_{ij}(X_i)$ $i = 1,\ldots,n, j = 1,\ldots,m_i$ of the random variables $X_1,\ldots,X_n$. Here m_i is the number of functions that are related to the i-th component. In figure 1 such a configuration is shown for three components X_1, X_2, X_3 where

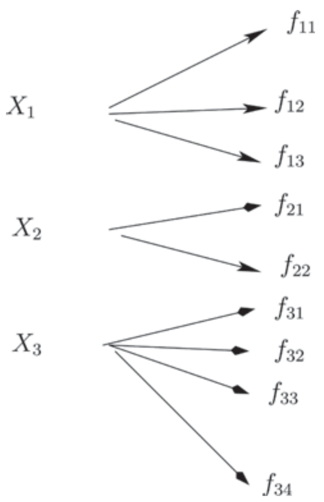

Figure 1. Example with 3 rv's and $m_1 = 3$, $m_2 = 2$, $m_3 = 4$.

there are three judgments for the first rv X_1, two for the second rv X_2 and four for X_3.

Now partial information about the reliability of components is represented as a set of lower and upper expectations $\underline{E} f_{ij}$ and $\overline{E} f_{ij}, i = 1,\ldots,n, j = 1,\ldots,m_i$ of these functions f_{ij}. The simplest case is that $f(X_i) = IE(X_i)$ and the bounds are for this mean value, i.e.

$$\underline{E}(X_i) \leq IE(X_i) \leq \overline{E}(X_i) \tag{1}$$

In the same way bounds for other moments can be defined taking $f(X_i) = X_i^k$; then

$$\underline{E}(X_i^k) \leq IE(X_i^k) \leq \overline{E}(X_i^k) \tag{2}$$

Further to get bounds for fractiles, if X_i is the time to failure for the i-th component, the probability that a failure of this component is in the time interval $[a, b]$ can be represented as expectation of the indicator function $I_{[a,b]}(X_i)$ such that

$$I_{[a,b]}(X_i) = 1, \text{ if } \; X_i \in [a,b] \; \text{ and } \; 0 \; \text{ elsewhere} \tag{3}$$

The partial information here is a lower and upper bound for this expectation

$$\begin{aligned} &\underline{E}(I_{[a,b]}) \leq IE(I_{[a,b]}(X_i)) \leq \overline{E}(I_{[a,b]}) \\ &\underline{E}(I_{[a,b]}) \leq \mathrm{P}(a \leq X_i \leq b) \leq \overline{E}(I_{[a,b]}). \end{aligned} \tag{4}$$

Said in plain words the probability that the value of lies between a and b is larger equal than $\underline{E}(I_{[a,b]})$ and less equal than $\overline{E}(I_{[a,b]})$. Using these bounds interval probability methods derive then bounds for the failure probability of the system given by $\mathrm{P}(g(X_1,\ldots,X_n) < 0)$. In general, such bounds can be derived easily if the limit state function $g(X_1,\ldots,X_n)$ is a linear or polynomial function of the random variables $X_1,\ldots,X_n$.

Utkin (2004a) applied such concepts to structural reliability, taking as first example the basic load-resistance model with L the load and R the resistance; where failure occurs if $L > R$. Then it is assumed that for the load L and the resistance R bounds for their respective CDF's are known.

$$\begin{aligned} &\underline{p}_i \leq \mathrm{P}(L \leq \alpha_i) \leq \overline{p}_i \\ &\underline{q}_j \leq \mathrm{P}(L \leq \beta_j) \leq \overline{q}_j \end{aligned} \tag{5}$$

For $i = 1,\ldots,n$ and $j = 1,\ldots,m$. So here the $\underline{p}_i(\overline{p}_i)$ are lower (upper) bounds for the CDF of the load L at the points α_i and respectively the $\underline{q}_j(\overline{q}_j)$ for the CDF of the resistance R at the points β_j. This can be used to obtain bounds for the failure probability $\mathrm{P}(L > R)$.

Further reliability problems, especially system reliability are studied in Utkin (2004b) and Utkin (2005). Generalizations of the interval probability method are second-order reliability models where the problems of contradicting judgements are discussed, see for example (Kozine & Utkin 2001).

Looking now at the starting point of the method, one question arises. How one does get such information about the CDF's and is it a realistic assumption that such information is available? To answer this question first in the next paragraph the concept of entropy will be introduced as a tool for solving this problem.

3 THE CONCEPT OF ENTROPY

The probability of an event A can be seen as uncertainty about the occurrence of this event. Considering now the question, how to model the uncertainty about which of a number of possible but incompatible events will occur, leads to the concepts of entropy and information. The outline here follows the presentation given in Papoulis (1991). A more detailed presentation can be found for example in Gray (1991). For a given probability space S a partition is a collection of mutually exclusive subsets such that the union of these sets equals S. Let a partition of S be denoted by $\mathcal{A}$. The concept of entropy assigns a measure of uncertainty not to a single event but to a partition of the probability space. This measure $H(\mathcal{A})$ is called the entropy of the partition. It models our uncertainty which of the possible events will occur.

A partition consisting only of two sets A_1 and A_2 is called a *binary* partition. So if for a binary partition we would be quite sure that A_1 will occur, the entropy is low; on the other hand if we give A_1 and A_2 the same chance, we would be more uncertain, the entropy is large.

The actual form of $H(\mathcal{A})$ is derived from some postulates formalizing the intuitive understanding of uncertainty. The usual set of postulates is the following (Shannon & Weaver 1949):

1. $H(\mathcal{A})$ is a continuous function of $p_i = \mathrm{P}(A_i)$,
2. If $p_1 = p_2 = \cdots = p_n = 1/n$, then $H(A)$ is an increasing function of n,
3. If a new partition $\mathcal{B}$ is formed by subdividing one of the sets of $\mathcal{A}$, then $H(\mathcal{B}) \geq H(\mathcal{A})$.

From these postulates one can derive that $H(\mathcal{A})$ (up to a multiplicative constant)

$$H(\mathcal{A}) = -p_1 \log(p_1) - \cdots - p_n \log(p_n). \tag{6}$$

Here it is assumed that $0 \log(0) = 0$, which can be justified by using L'Hospital's rule. It can be derived easily that the entropy is maximal if all probabilities p_i of the partition are equal, i.e. if $p_1 = p_2 = \cdots = p_n = 1/n$ Then

$$H(\mathcal{A}) = -n\frac{1}{n}\log(1/n) = \log(n). \tag{7}$$

For a binary partition we have $P(A_2) = 1 - P(A_1)$ and the entropy is a function of $P(A_1)$ alone (see figure 2). In the case that $P(A_1) = 0$ or $P(A_2) = 0$ the entropy is equal to zero, we are totally certain which of the two events will occur. If now we observe an event M, the entropy of the partition changes, since now the probability space is $S \cap M$ and no more S. Calculating then the entropy under the condition that M was observed, gives the conditional entropy $H(\mathcal{A} \mid M)$ defined as

$$H(\mathcal{A} \mid M) = -\sum_{i=1}^{n} \mathrm{P}(A_i \mid M)\log(\mathrm{P}(A_i \mid M)). \tag{8}$$

The difference $H(\mathcal{A}) - H(\mathcal{A} \mid M)$ is the information about $H(\mathcal{A})$ contained in M. With more and more observations and data the entropy should decrease, since we get more information. If we have a binary partition, we have for the conditional entropy

$$\begin{aligned} H(\mathcal{A} \mid M) = &-\mathrm{P}(A_1 \mid M)\log(\mathrm{P}(A_1 \mid M)) \\ &-\mathrm{P}(A_2 \mid M)\log(\mathrm{P}(A_2 \mid M)). \end{aligned} \tag{9}$$

If we have perfect information about the partition after observing M, i.e. if no uncertainty is left

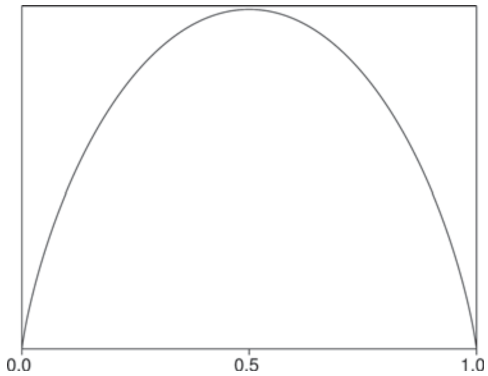

Figure 2. Entropy of a binary partition as function of $P(A_1)$.

and we know for example that A_1 is the true state, the conditional entropy would be zero.

4 A BAYESIAN ANALYSIS OF INTERVAL PROBABILITY METHODS

Here we will examine now the proposed interval probability in the light of the Bayesian paradigm. For this we will consider a totally simple example, Bernoulli experiments, and study interval probability methods for it. All these partial information about distributions described in the second paragraph can be considered as information about a Bernoulli random variable. Now, if X is a Bernoulli random variable, denoting for example $X = 1$ the failure of a component or the probability that a random variable Y takes on values in a specific interval, its expected value $\mathbb{E}(X) = p$ is the only parameter of its distribution. Now we assume that partial information about this expected value is given, a lower bound $\underline{p}$ and an upper bound $\overline{p}$ so that

$$\underline{p} \leq p \leq \overline{p}. \tag{10}$$

Interval probability followers take this as starting point for their methods. But how do we get such an information? An attempt to answer this question is now made, putting all into a Bayesian framework.

In a Bayesian framework we start from total ignorance, i.e. we assume a uniform prior distribution over the unit interval [0, 1] and then we learn from the observed data. Let now be defined $A_1 = [\underline{p}, \overline{p}]$. Then the prior probabilities for the two sets are

$$\begin{aligned} &P(A_1) = \left|\overline{p} - \underline{p}\right|, \\ &P(A_2) = 1 - P(A_1) = 1 - \left|\overline{p} - \underline{p}\right|. \end{aligned} \tag{11}$$

We exclude the trivial case that $\mathrm{P}(A_1)=1$, i.e. that $\overline{p}=1$ and $\underline{p}=0$, so we assume that $\mathrm{P}(A_2)>0$ and $P(A_1)<1$.

For this simple model the only observations we make, are the successes or failures of a sequence of Bernoulli experiments. How can we arrive at the conclusion

$$\mathrm{P}_{post}(A_1)=1, \quad \mathrm{P}_{post}(A_2)=1-\mathrm{P}_{post}(A_1)=0? \tag{12}$$

Here P_{post} denotes a posterior probability distribution of which it is known only that $P(A_1)=0$ and $P(A_2)=1$.

If we can derive the result in eq. (12) in a Bayesian setting, it must be so that after a finite number of observations the then achieved posterior distribution of p gives us this result. This assumption is the basis of interval probability methods in this elementary example. After a *finite* amount of data being collected one arrives at the conclusion in eq. (12), which is then the starting point for interval probability methods.

Now, if we observe the results of more and more Bernoulli experiments, the entropy should diminish, until we reach the state where we can deduce that $\mathrm{P}_{post}(A_1)=1$. But this would mean that the entropy $H(\mathcal{A})$ of the binary partition consisting of A_1 and A_2 with this posterior probability distribution is zero, since

$$\begin{aligned} H(\mathcal{A}\mid M) &= -\mathrm{P}_{post}(A_1)\log(\mathrm{P}_{post}(A_1)) \\ &\quad -\mathrm{P}_{post}(A_2)\log(\mathrm{P}_{post}(A_2)) \\ &= -1\log(1)-0\log(0)=0. \end{aligned} \tag{13}$$

If we can derive the result in eq. (12) in the Bayesian framework, the conditions must be

1. A finite number k of Bernoulli experiments is observed,
2. After these experiments, if the observed outcomes lie in a set M, the result $\mathrm{P}_{post}(A_1)=1$ is obtained.

So the procedure should look somehow as shown in figure 3. But as will be shown in the following such a result—i.e. a posterior with $P(A_1)=1$—cannot be obtained. If only data are used to derive the statement in eq. (12), after a finite number k of experiments has been observed this conclusion must have been reached. Here nothing is said about which set M would bring this conclusion, but if only a finite amount of information is used, such a conclusion must be based on the observation of some set M and nothing else. Since this is the essential part of the argument, to repeat it, if the deduction is made in a rational way derived from the observed data, it must be done in such a way. Other ways to follow that are not justifiable in a rational way in this context.

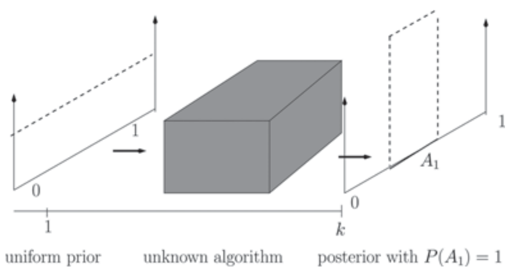

Figure 3. Prior and posterior distribution.

In a finite sequence of k Bernoulli experiments, all elementary events E are binary strings of length k. For example, for $k=3$ the elementary events are

$$\begin{aligned} &\{000\},\{100\},\{010\},\{001\}, \\ &\{110\},\{101\},\{011\},\{111\}. \end{aligned} \tag{14}$$

Each event M observed in a sequence of k Bernoulli trials can be written as a union of the elementary events contained in M. Due to their definition these elementary events are disjoint.

For each such elementary event E the posterior probability of A_i given E is (Press (1989), p. 40)

$$\mathrm{P}(A_i\mid E)\propto \mathrm{P}(A_i)\binom{k}{n_E}\int_{A_i} p^{n_E}(1-p)^{k-n_E}\,dp \tag{15}$$

Here n_E is the number of ones in the string E. For an arbitrary event M which is composed of some elementary events E, the posterior probabilities are then

$$\mathrm{P}(A_i\mid M)\propto \mathrm{P}(A_i)\sum_{E\subset M}\binom{k}{n_E}\int_{A_i} p^{n_E}(1-p)^{k-n_E}\,dp \tag{16}$$

It is obvious that these posterior probabilities remain always positive, if the prior probabilities were positive for both sets, which is the case here. Using eq. (7), we can conclude that always

$$\begin{aligned} H(\mathcal{A}\mid M) &= -\mathrm{P}(A_1\mid M)\log(\mathrm{P}(A_1\mid M)) \\ &\quad -\mathrm{P}(A_2\mid M)\log(\mathrm{P}(A_2\mid M))>0. \end{aligned} \tag{17}$$

So for any set M which can be observed after a finite number k of Bernoulli experiments always

$$H(\mathcal{A}\mid M)>0. \tag{18}$$

Therefore it is impossible to reach the conclusion in eq. (12) in any way with a finite number of Bernoulli experiments if we look at the problem in the Bayesian framework. Somehow an infinite amount of data is needed to arrive at this conclusion.

Now, defenders of the interval probability method may object, this method is no Bayesian method, we can reject the Bayesian paradigm, why should it therefore fit in here. To answer this possible objection, the henchmen of this method have to give a rational answer how to reach the conclusion in eq. (12) if only a finite amount of information is given which is quite usual in real life since this is the starting point of the whole procedure. Elsewhere if no rational explanation for this is presented, one seems to be forced to believe that somehow an information-theory king Midas transforms information from a finite set of data into information from an infinite set of data.

5 CONCLUSIONS

The results in this paper lead (at least the author) to the following conclusions:

1. The interval probability method is based on assumptions which are practically never fulfilled in halfway realistic problems.
2. How to obtain the information necessary as starting point for applying these concept, i.e. the upper and lower expectations for functions of the involved random variables, remains an enigma. The advocates of interval probability methods mention as source always expert opinions never the evaluation of data.
3. The use of classical probability theory in reliability is an anachronism. These concepts have been superseded since decades by Bayesian methods. Further this concept is static, an analysis is made and a result is obtained. There is no provision for in corporating new evidence via the theorem of Bayes.
4. The proposed methods are insofar dangerous for risk calculations as they tend to eliminate and underestimate the influence of distribution tails. This comes from the fact that only expert opinions are allowed which do not include any probability statements about that the judged quantity is not in a certain interval. Only statements in the form *"This quantity lies with probability one in the interval here"* are accepted, so forcing an expert on the bed of Procrustes.
5. Since the classical probability concept is used, only intervals are found, no probability distributions, which makes it impossible to use any optimization procedures based on expected values (risk, cost).
6. With the computing capacities available nowadays it is possible to study the influence of assumptions about the used probability distributions to avoid wrong conclusions. So there is no need to avoid any assumptions about distributions; as it might have been important decades ago, where there was no real possibility to examine the impact of these assumptions.

REFERENCES

Ben-Haim, Y. (1996). *Robust Reliability in the Mechanical Sciences*. Berlin: Springer.

Coolen, F. (2004). On the use of imprecise probabilities in reliability. *Quality and Reliability Engineering International 20*, 193–202.

Gray, R. (1991). *Entropy and Information Theory*. Springer-Verlag New York, Inc. Revised edition online: http://ee.stanford.edu/~gray/it.html.

Kozine, I. & L. Utkin (2001). Constructing coherent interval statistical models from unreliable judgements. In E. Zio et al. (Eds.), *Proc. of the European Conference on Safety and reliability ESREL 2001*, Volume 1, Torino, Italy, pp. 173–180.

Papoulis, A. (1991). *Probability, random variables, and stochastic processes* (3rd ed.). McGraw-Hill.

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

Shannon, C. & W. Weaver (1949). *The mathematical Theory of Communication*. University of Illinois Press.

Utkin, L. (2004a). Interval relibiability of typical systems with partially known probabilities. *European Journal of Operational Research 153* (3), 790–802.

Utkin, L. (2004b). Reliability models of m-out-of-n systems under incomplete information. *Computers and Operations Research 31* (10), 1681–1702.

Utkin, L. (2005). A hierarchical model of reliability by imprecise parameters of lifetime distributions. *Int. J. of Reliability, Quality and Safety Engineering 12* (2), 167–187.

Utkin, L. & F. Coolen (2007). Imprecise reliability: An introductory review. In G. Levitin (Ed.), *Computational Intelligence in Reliability Engineering, Volume 2: New Metaheuristics, Neural and Fuzzy Techniques in Reliability*, pp. 261–306. Springer-Verlag New York, Inc.