International Forum on Engineering Decision Making
6th IFED Forum, January 26-29, 2012, Lake Louise, Canada

Interval Probabilities for Reliability: Panacea or Pipe Dream?

Karl W. Breitung
Technical University of Munich, 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 information. A prerequisite which in halfway realistic problems cannot be fulfilled.

1. Introduction

Structural reliability started to develop in the 50ies of the last century, usually the paper by Freudenthal (1956) is seen as marking the begin of this research field. Using in the beginning the ideas of classical probability, the favorite approach soon shifted to Bayesian methods. A seminal book pioneering this approach was Benjamin and Cornell (1970). Since then the overwhelming majority of research in this field is based on this paradigm.
One point that causes often uneasiness in people entrenched in deterministic thinking is that there is no 100% percent certainty. 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.
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 shortcomings in them, which might be serious.
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 this paragraph a sketch of the interval probability method following the paper by Uktin and Coolen (2007) is given.
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 $X_i,\ i=1,\ldots,n$. Suppose that partial information about reliability of components is represented as a set of lower and upper expectations $\underline{E} f_{i\,j}$ and $\overline{E} f_{i\,j}$, $i=1,\ldots,n,\ j=1,\ldots,m_i$ (i.e. numbers) of functions $f_{i,j}$. Here m_i is a number of judgments that are related to the *i*-th component; $f_{i\,j}(X_i)$ is a function of the random variable X_i of the *i*-th component or some different random variable, describing the *i*-th component reliability and corresponding to the *j*-th judgment about this component.
The simplest cases are bounds for the moments of X_i . Here this means for the mean

$$\underline{E}(X_i) \le E(X_i) \le \overline{E}(X_i) \quad . \tag{1}$$

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

$I_{[a,b]}(X_i)=1$ if $X_i \in [a,b]$ and
$I_{[a,b]}(X_i)=0$ if $X_i \notin [a,b]$.

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

$$\underline{E}(I_{[a,b]}) \ \le\ E(I_{[a,b]}(X_i)) \ \le\ \overline{E}(I_{[a,b]})$$

$$\underline{E}(I_{[a,b]}) \ \le P(a \ \le\ X_i \ \le\ b) \ \le\ \overline{E}(I_{[a,b]}). \tag{2}$$

Said in plain words the probability that the value of X_i lies between *a* and *b* is larger equal than $\underline{E}(I_{[a,b]})$ and less equal than $\overline{E}(I_{[a,b]})$.
Utkin (2004) 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

$$\underline{p_i} < P(L < \alpha_i) < \overline{p_i},\ \ \underline{q_j} < P(R < \beta_j) < \overline{q_{j.}} \tag{3}$$

for i=1,...,n and j=1,..,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 *P(L>R)*. Using now this information, further bounds for the reliability are derived.
Further reliability problems, especially system reliability are studied in Utkin (2004b, 2005). Generalizations of the interval probability method are second-order reliability models where the problems of contradicting judgements are discussed, see for example Kozin and 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 $A_{1,}\ldots,A_n$ such that the union of these sets equals S.

Let the partition of n sets $A_{1,}\ldots,A_n$ be denoted by $\mathbf{A}$. The concept of entropy assigns a measure of uncertainty not to a single event but to a partition of the probability space. This measure $H(\mathbf{A})$ is called the *entropy* of the partition. It models our uncertainty which of the possible events will occur.

So if there are only two events A_1 and A_2 in the partition and $P(A_1)=0.999$ and $P(A_2)=0.001$, we would be quite sure that A_1 will occur, the entropy is low; on the other hand if $P(A_1)=0.5,\ P(A_2)=0.5$, we would be more uncertain, the entropy is large.

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

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

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

$$H(\mathbf{A}) = -p_1\log(p_1)-\ldots-p_N\log(p_N). \tag{4}$$

Here it is assumed that $0\cdot\log(0)=0$, 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 $p_1=\ldots=p_n=1/n$. Then

$$H(\mathbf{A}) = n\frac{1}{n}\log(1/n) = \log(n). \tag{5}$$

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* $H(\mathbf{A}|M)$ defined as

$$H(\mathbf{A}|M) = -\sum_{i=1}^{N} P(A_i|M)\log P(A_i|M). \tag{6}$$

The difference $H(\mathbf{A})$ - $H(\mathbf{A}|M)$ is the *information* about $H(\mathbf{A})$ 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, i.e. the probability space S is divided in only two sets A_1 and A_2 , we have for the conditional entropy

$$H(\mathbf{A}|M) \;=\; -P(A_1|M)\log P(A_1|M) - P(A_2|M)\log P(A_2|M). \tag{7}$$

If we have perfect information about the partition after observing M, i.e. if no uncertainty is left 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 $E(X)=p$ 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

$$\underline{p} \;\leq\; p \;\leq\; \overline{p}. \tag{8}$$

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

$$P(A_1) \;=\; |\,p - \underline{p}\,| \;\;,\;\; P(A_2) \;=\; 1 - |\,\overline{p} - \underline{p}\,| \tag{9}$$

How can we arrive at the conclusion

$$P_{post}(\underline{p} \;\leq\; p \;\leq \overline{p}) \;=\; P_{post}(A_1) \;=\; 1\,? \tag{10}$$

Here P_{post} denotes some unspecified posterior probability. If we can derive the result in eq. (10) in a Bayesian setting, it must be so that after a 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 basic example.
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 $P_{post}(\underline{p} \leq p \leq \overline{p}) = P_{post}(A_1) = 1$. But this would mean that the entropy $H(\mathbf{A})$ of the binary partition consisting of A_1 and A_2 with this posterior probability distribution is zero, since

$$H(\boldsymbol{A}) = P_{post}(A_1)\cdot\log(P_{post}(A_1)) + 0\cdot\log(P_{post}(A_2)) = 1\cdot\log(1) + 0\cdot\log(0) = 0. \tag{11}$$

If only data are used to derive the statement in the equation above, after a finite number K of experiments having observed an event M, 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 $P_{post}(A_1)=1$ are not justifiable in a rational way in this context.
In a finite sequence of K Bernoulli experiments, all elementary events E are binary strings of length K. For each such elementary event the posterior probability is (Press (1989), p. 40)

$$P(A_i|E) \propto P(A_i)\binom{K}{n_E}\int_{A_i} p^{n_E}(1-p)^{K-n_E}dp, \quad i=1,2. \tag{12}$$

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

$$P(A_i|M) \propto P(A_i)\sum_{E\in M}\binom{K}{n_E}\int_{A_i} p^{n_E}(1-p)^{K-n_E}dp, \quad i=1,2. \tag{13}$$

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

$$H(\boldsymbol{A}|M) > 0. \tag{14}$$

Therefore it is impossible to reach the conclusion in eq. (10) in any way with a finite number of experiments if we look at the problem in the Bayesian framework. Somehow an infinite amount of information is needed to arrive at the conclusion in eq. (10).
Now, defenders of the interval probability method may object, this method is no Bayesian method, we 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.(10) 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 an information-theorical king Midas transforms finite information from data into infinite information.

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. It was shown that the interval probability methods are quite insensitive to the influence of new data. Given the “expert opinions”, the derived intervals remain forever unimpressed by any new data.
4. 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 incorporating new evidence via the theorem of Bayes.
5. 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 statements about the probabilities that the judged quantity is in the interval. Only statements “The quantity lies with probability one here” are accepted, so forcing an expert on the bed of Procrustes.
6. Since the classical probability concept is used, only intervals are found, no probability distributions, which makes any optimization procedures based on expected values (risk, cost) impossible.
7. With the computing powers available nowadays it is possible to study the influence of assumptions about the used probability and prior distributions to avoid wrong conclusions. So there is no need to avoid 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 Verlag.

Benjamin, J. and Cornell, C.A. (1970) “Probability, Statistics, and Decisions for Civil Engineers”, New York: McGraw-Hill.

Coolen, F. P. A. (2004) “On the Use of Imprecise Probabilities in Reliability”, Quality and Reliability Engineering International, Vol. 20, pp. 193–202.

Freudenthal, A. (1956) “Safety and the probability of structural failure”, Transactions ASCE, Vol. 121, pp. 1337-1397.

Gray, R.M. (1991) “Entropy and Information Theory”, New York: Springer. Revised edition online: http://ee.stanford.edu/~gray/it.html.

Kozine I. and Utkin L.V. (2001) “Constructing coherent interval statistical models from unreliable judgements”. In: Zio E, Demichela M, Piccini N (eds.): Proc. of the European Conference on Safety and Reliability ESREL2001, Vol. 1., Torino, Italy, pp. 173-180.

Papoulis, A. (1991) “Probability, Random Variables, and Stochastic Processes”, 3rd edition, New York: McGraw-Hill.

Press, S.J. (1989) “Bayesian Statistics”, New York: John Wiley & Sons.

Shannon, C.E. and Weaver, W. (1949) “The Mathematical Theory of Communication”, University of Illinois Press.

Utkin, L.V. (2004a) “Interval reliability of typical systems with partially known probabilities”, European J. of Operational Research, Vol. 153(3), pp. 790-802.

Utkin L.V. (2004b) “Reliability models of m-out-of-n systems under incomplete information”, Computers and Operations Research, Vol. 31(10), pp. 1681-1702.

Utkin L.V. (2005) “A hierarchical model of reliability by imprecise parameters of lifetime distributions” Int. J. of Reliability, Quality and Safety Engineering, Vol.12(2), pp. 167-187.

Utkin L.V. and Coolen, F.P.A. (2007) “Imprecise reliability: An introductory overview”. Computational Intelligence in Reliability Engineering, Volume 2: New Metaheuristics, Neural and Fuzzy Techniques in Reliability, Gregory Levitin (Ed.), New York: Springer, pp. 261-306.