International Journal of General Systems, Vol. 34, No. 4, August 2005, 465–482

Taylor & Francis
Taylor & Francis Group

Convergence properties of an interval probabilistic approach to system reliability estimation

CLIFF JOSLYN†* and VLADIK KREINOVICH‡¶

†Knowledge and Information Systems Science Team, Modeling, Algorithms, and Informatics Group (CCS-3), Los Alamos National Laboratory, Mail Stop B265, Los Alamos, NM 87545, USA
†Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA

(Received 11 November 2002; in final form 13 October 2004)

Based on a black box model of a complex system, and on intervals and probabilities describing the known information about the inputs, we want to estimate the system's reliability. This problem is motivated by a number of problem areas, most specifically in engineering reliability analysis under conditions of poor measurement and high complexity of system models. Using the results of tests performed on the system's computer model, we can estimate the lower and upper bounds of the probability that the system is in a desirable state. This is equivalent to using Monte-Carlo sampling to estimate cumulative belief and plausibility values of functionally propagated finite random intervals. In this paper, we prove that these estimates are correct in the sense that under reasonable assumptions, these estimates converge to the actual probability bounds.

Keywords: Interval probability; Interval analysis; Reliability analysis; Dempster-Shafer evidence theory; Random sets; Random intervals

1. Introduction

In this paper, we consider the problem of modeling the risk and reliability of complex technical system. The behavior of this system is determined by the values of the corresponding parameters $x = \langle x^{(1)}, x^{(2)}, \ldots, x^{(n)} \rangle$; for example, for a nuclear reactor, these characteristics could include the thickness of the walls, the locations of the radiation absorbers, etc. For each combination, x of these parameters, the system exhibits certain characteristics $y = \langle y^{(1)}, y^{(2)}, \ldots, y^{(m)} \rangle$; e.g. for a nuclear reactor, the list of such characteristics include neutron flux, temperature, etc.

We assume that the parameters set is complete (or almost complete), so that the observed state is uniquely determined by the values x of the parameters, so that $y = f(x)$ for some function f. In this context, f acts as a model of our knowledge of the system.

*Corresponding author. Tel.: +1-505-667-9096. Email: joslyn@lanl.gov; URL: http://www.c3.lanl.gov/ ~ joslyn
¶E-mail: vladik@cs.utep.edu

International Journal of General Systems
ISSN 0308-1079 print/ISSN 1563-5104 online © 2005 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/03081070500033880

The reliability of such a system is related to the fact that some states y are desirable, while some other states are not. Thus we'll be concerned with the conditions under which the system output occurs within some desirable output set B_0.

If we knew the exact values of the parameters $x^{(i)}$, then we would be able to determine the corresponding state $y = f(x)$ and check whether it is desirable, $f(x) \in B_0$, or not. But in real life, we usually do not know the exact values of $x^{(i)}$. Instead, in some instances we may know a lower bound $\underline{x}^{(i)}$ and the upper bound $\overline{x}^{(i)}$ that are known to contain $x^{(i)}$. Often, we do not know how more or how less probable are different values within this interval, i.e. what is an actual probability distribution for $x^{(i)}$ within this interval.

In some cases, instead of a single interval, several experts provide different intervals corresponding to different possible situations. Then, in addition to these intervals, we usually also know the probability of each such situation. For example, a reactor shell can come from three different manufacturing plants, and we know the frequencies with which they come from different plants, i.e. the probabilities that a randomly selected shell is from this particular plant. For each plant, we also know the interval of possible value of thickness for shells produced by this plant.

Furthermore, systems of interest are characterized by a high complexity such that these models f are large simulation codes. These codes are sometimes so huge that each run requires days on supercomputers. As a result, we cannot typically control what inputs we feed into the code, but have to reply on the results of the testing, i.e. on some pairs $\{\langle x_1, y_1\rangle, \langle x_2, y_2\rangle, \ldots \langle x_M, y_M\rangle\}$ corresponding to these actual test runs.

Given such information, what can you know about the probability P that the resulting state is desirable? If we knew the joint probability distribution of the parameters, then we could determine the probability of different values of y and thus we could get the probability $\Pr(f(x) \in B_0)$. In reality, we only have partial information about the probability distributions. For different distributions, we may get different values of P. Our goal is, therefore, to find the interval $\mathbf{P} = \left[\underline{P}, \overline{P}\right]$ of possible values of this probability P.

In this paper, we develop the methodology of this problem formulation under conditions of only sampling information from $f(x)$, and determine convergence conditions for both the upper and lower value of this probability interval.

This class of problems is quite realistic in certain engineering modeling contexts, and was the recent focus of a major interdisciplinary research effort among the engineering modeling, risk analysis and generalized information theoretical and imprecise probability communities [2,7,11]. Most specifically, this can be understood as the problem of propagating a finite random interval [8] through the model f. Monte-Carlo sampling approaches to such random interval sampling are in development [5,6,9], and this work stands to assist that effort in providing a rigorous formulation of some of the required convergence results.

2. Probability intervals on input information

We begin by introducing our basic mathematical constructs, characterizing the uncertainty structures on the inputs, and demonstrating the resulting probability intervals.

2.1 Input information

Let $\mathbb{R}$ be the real numbers, and $\mathbb{N} := \{1, 2, \ldots\}, \mathbb{N}_n := \{1, 2, \ldots, n\}$. Assume integers $n, m \in \mathbb{N}$ where n is the number of inputs and m the number of outputs, and let $f : \mathbb{R}^n \mapsto \mathbb{R}^m$.

Let the ith input be denoted $x^{(i)}$, taking values on a set $X^{(i)} = \mathbb{R}$, for $i \in \mathbb{N}_n$. Similarly, let the lth output be denoted $y^{(l)}$, taking values on a set $Y^{(l)} = \mathbb{R}$, for $l \in \mathbb{N}_m$. Let $x := \langle x^{(1)}, x^{(2)}, \ldots, x^{(n)} \rangle \in \mathbb{R}^n$ be called input points and $y := \langle y^{(1)}, y^{(2)}, \ldots, y^{(m)} \rangle \in \mathbb{R}^m$ be called states.

Let X and Y be the sets of closed, rectangular boxes in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and generically, let $A \in X, B \in Y$. More specifically, let $B_0 \in Y$ be the box of desired states; all other states $y \in \mathbb{R}$, $y \notin B_0$ will be called undesirable.

Let $A^{(i)} = \left[\underline{x}^{(i)}, \overline{x}^{(i)}\right] \subseteq X^{(i)}$ be a closed interval of the ith input $X^{(i)}$, and $B^{(l)} = \left[\underline{y}^{(l)}, \overline{y}^{(l)}\right] \subseteq Y^{(l)}$ a closed interval of the lth output $Y^{(l)}$. We assume that for every input i from 1 to n, we have information about ith input expressed as a finite collection of $N^{(i)} \in \mathbb{N}$ weighted intervals. In particular, denote the uncertainty structure on the i'th input as

$$\mathcal{S}^{(i)} := \left\{ \left\langle A_1^{(i)}, p_1^{(i)} \right\rangle, \ldots, \left\langle A_{j^{(i)}}^{(i)}, p_{j^{(i)}}^{(i)} \right\rangle, \ldots, \left\langle A_{N^{(i)}}^{(i)}, p_{N^{(i)}}^{(i)} \right\rangle \right\},$$

where $j^{(i)} \in \mathbb{N}_{N^{(i)}}, A_{j^{(i)}}^{(i)} = \left[\overline{x}_{j^{(i)}}^{(i)}, \underline{x}_{j^{(i)}}^{(i)}\right] \subseteq X^{(i)}$ is one such interval, $p_{j^{(i)}}^{(i)} \in [0, 1]$, and we have $\forall_i \in \mathbb{N}_n$ the probabilistic normalization criterion

$$\sum_{j^{(i)}=1}^{N^{(i)}} p_{j^{(i)}}^{(i)} = 1. \tag{1}$$

So in general, across the different inputs $X^{(i)}$, we can choose a particular combination of intervals $A_{j^{(i)}}^{(i)}$, one for each input dimension $X^{(i)}$. Denote $\vec{j} := \langle j^{(1)}, j^{(2)}, \ldots, j^{(n)} \rangle$ as indicating these combinations. There are $N := \prod_{i=1}^{n} N^{(i)}$ such possible combinations. Since, there exists a bijective mapping between $\mathbb{N}_N$ and the set of all combinations $\vec{j}$, we can thereby use the $j \in \mathbb{N}_N$ to enumerate the various possible $\vec{j}$. Also, denote $j^{(i)} \in \vec{j}$ to indicate that a particular $j^{(i)}$ is one of the components of $\vec{j}$.

So for each such combination $\vec{j}$, we can define the box $A_{\vec{j}}$ as the Cartesian product of the corresponding intervals:

$$A_{\vec{j}} := \bigtimes_{i=1}^{n} A_{j^{(i)}}^{(i)} \in X,$$

also denoted $A_j \in X$ as appropriate. Furthermore, assume that the information corresponding to different parameters are independent. Then for each combination $\vec{j}$, we have the overall probability "mass" $p_{\vec{j}} := \prod_{i=1}^{n} p_{j^{(i)}}^{(i)}$, also denoted P_j. In this way, from the individual uncertainty structures $\mathcal{S}^{(i)}$, we can construct the overall input uncertainty structure

$$\mathcal{S} := \left\{ \left\langle A_{\vec{j}}, p_{\vec{j}} \right\rangle \right\}_{\vec{j}} = \{ \langle A_1, p_1 \rangle, \ldots, \langle A_j, p_j \rangle, \ldots, \langle A_N, p_N \rangle \}.$$

An example is shown in figure 1 for $n = 2$ input parameters and $N^{(1)} = N^{(2)} = 2$ intervals on each input parameter. The input intervals $A_1^{(1)}, A_2^{(1)}$ on $X^{(1)}$ and $A_1^{(2)}, A_2^{(2)}$ on $X^{(2)}$ are shown, with probabilities

$$p_1^{(1)} = 0.4, \quad p_2^{(1)} = 0.6, \quad p_1^{(2)} = 0.2, \quad p_2^{(2)} = 0.8,$$

which assignment satisfies equation (1). The boxes $A_{\vec{j}}$ and masses $p_{\vec{j}}$ are shown, along with their enumerated forms A_j, p_j. The bijective mapping $\mathbb{N}_N \leftrightarrow \{\vec{j}\}$ is

$$\langle 1, \langle 1, 1 \rangle \rangle, \quad \langle 2, \langle 1, 2 \rangle \rangle, \quad \langle 3, \langle 2, 1 \rangle \rangle, \quad \langle 4, \langle 2, 2 \rangle \rangle$$

shown as tuples of the form $\langle j, \vec{j} = \langle j^{(1)}, j^{(2)} \rangle \rangle$.

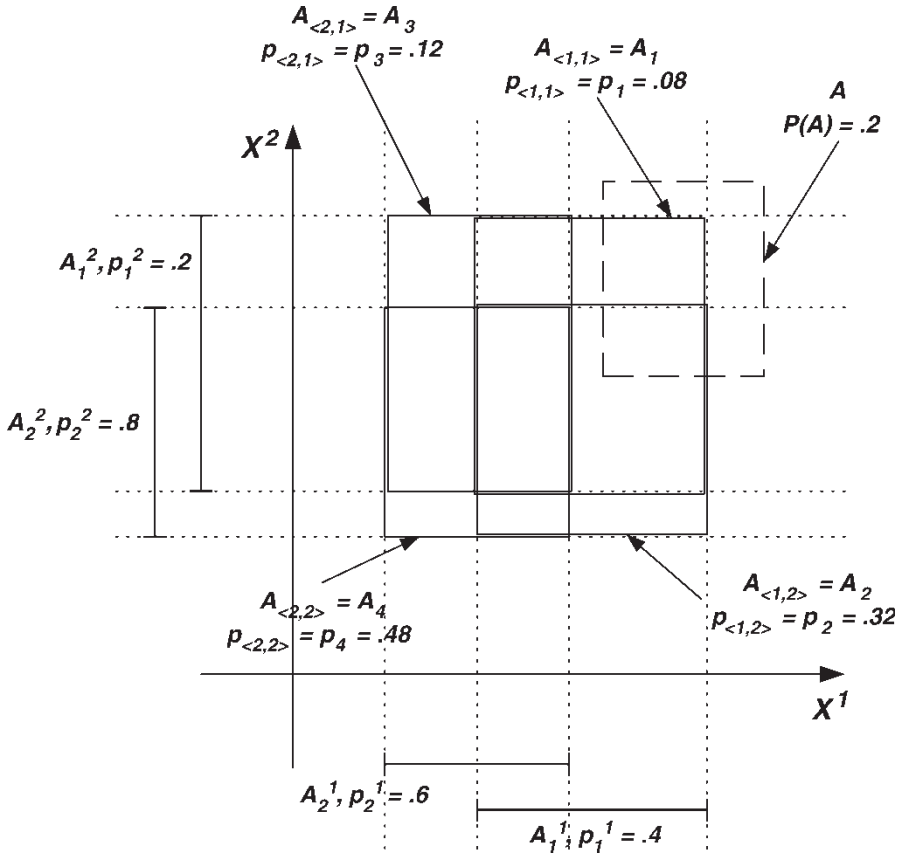

Figure 1. Example input uncertainty structure $\mathcal{S}$.

Note that for illustrative purposes, in figure 1, we show the boxes $A^{(i)}_{j^{(i)}}$ slightly offset from each other, in order to clearly distinguish them. In fact, their borders overlap where they are shown very close to each other.

2.2 Consistent probability measures

Consider a probability measure P on $\mathbb{R}^n$. For an arbitrary box in the input space $A \in X$, we say that the total probability $P(A)$ is consistent with the input uncertainty structure $\mathcal{S}$ if there exists a collection of total probabilities $\langle P_j(A)\rangle$, which are concentrated on the corresponding boxes A_j, such that

$$P(A) = \sum_{j=1}^{N} p_j P_j(A).$$

An example is also shown in figure 1, for $A \in X$ as illustrated. It can be demonstrated that $P(A) = 0.2$ is consistent with $\mathcal{S}$, because for the distribution of total probabilities $\langle P_j(A)\rangle = \langle 0.5, 0.5, 0, 0\rangle$ we have

$$\sum_{j=1}^{N} p_j P_j(A) = 0.08 \times 0.5 + 0.32 \times 0.5 + 0.48 \times 0 + 0.12 \times 0 = 0.2 = P(A).$$

2.3 Probability intervals

Our goal is to describe the smallest interval $\mathbf{P} := \left[\underline{P}, \overline{P}\right]$ that contains all possible values of $\Pr(f(x) \in B_0) = P(\{x : f(x) \in B_0\})$ for all consistent probability measures $P(A)$.

THEOREM 1

$$\underline{P} = \sum_{j:f(A_j)\subseteq B_0} p_j, \quad \overline{P} = \sum_{j:f(A_j)\cap B_0 \neq \emptyset} p_j. \tag{2}$$

2.4 Comments

- Proofs of all theorems, corollaries and propositions can be found in the Appendix.
- The problem formulation, including a different version of Theorem 1, was originally expressed by Joslyn and Helton [9].
- Many readers will recognize equation (2) as formulae from the Dempster–Shafer theory [4,13]: $\underline{P}$ is the formula for belief, and $\overline{P}$ is the formula for plausibility, with boxes A_j as focal elements, and with p_j being the mass of the "basic probability assignment" of the corresponding box A_j. The formula for the mass of the box is also familiar: it corresponds to the Dempster–Shafer combination of the corresponding "knowledge bases" $\mathcal{S}^{(i)}$. This similarity is no accident: Dempster–Shafer formalism was originally designed to describe exactly such situations—when we have only partial information about probabilities.
- In particular, we assume a random set interpretation of a Dempster–Shafer structure [1,12], so that the weights $p^{(i)}_{j^{(i)}}$ are interpreted as values of a discrete probability density over the atomic events which are actually the intervals $A^{(i)}_{j^{(i)}}$, and thus which may be overlapping, included within each other, or disjoint.
- Similarly, our formulation of the input uncertainty structure, both the dimensional form $\mathcal{S}^{(i)}$ and the overall form $\mathcal{S}$, is isomorphic to a formulation as a finite random set [3,9], which is itself isomorphic to Dempster–Shafer evidence theory.

3. Basic sampling results

Theorem 1 describes how we can compute the bounds $\underline{P}$ and $\overline{P}$ in the ideal situation when we know the function $f(x)$. In reality, all we know are some samples $\langle x_k, f(x_k)\rangle$ from this function.

Denote $y_k = f(x_k)$, and let $S := \{\langle x_1, y_1\rangle, \ldots, \langle x_k, y_k\rangle, \ldots\}$ be an infinite sequence of sample points for $k \in \mathbb{N}$. For some $M \in \mathbb{N}$, denote $S_M := \{\langle x_1, y_1\rangle, \ldots, \langle x_M, y_M\rangle\}$ as the initial finite subsequence of M sample points.

How can we estimate $\underline{P}$ and $\overline{P}$ based on these samples?

3.1 Lower probability

Let us start with $\underline{P}$. According to Proposition 1, the actual value $\underline{P}$ is the sum of the values p_j for all the boxes A_j for which $f(A_j) \subseteq B_0$. This set theoretical condition can be re-expressed in logical terms:

$$f(A_j) \subseteq B_0 \equiv \{f(x) : x \in A_j\} \subseteq B_0 \equiv \forall x \in A_j, f(x) \in B_0 \equiv x \in A_j \rightarrow f(x) \in B_0$$

Thus, the left side of equation (2) can be restated as:

$$\underline{P} = \sum_{j: x \in A_j \rightarrow f(x) \in B_0} p_j.$$

So, since we only know the value of $f(x)$ on M different inputs x_k, it makes sense to use as an estimator the same expression on the sample data set, that is to define:

$$\underline{P}_M := \sum_{j: \forall k \leq M, x_k \in A_j \rightarrow y_k \in B_0} p_j. \tag{3}$$

We will show below that $\underline{P}_M$ is, indeed, a good estimator of $\underline{P}$.

Before that, it is useful to consider some of the properties of $\underline{P}_M$. First, note that equation (3) can be restated, by logical expansion of the implication operator and de Morgan over universal quantification, as

$$\underline{P}_M = \sum_{\forall j: \forall k \leq M, x_k \notin A_j \vee y_k \in B_0} p_j \quad = \sum_{\forall j: \nexists k \leq M, x_k \in A_j \wedge y_k \notin B_0} p_j.$$

In other words, $\underline{P}_M$ includes the p_j values for any box A_j which is not "contradicted" by a data point $\langle x_k, y_k \rangle$ such that x_k is in the box A_j, but nonetheless $y_k \notin B_0$. Note that for $M = 0$, there can be no such data points, since S_0 is empty (there are no data points at all). Thus $\underline{P}_0 = \sum_{j=1}^{N} p_j = 1$, so that all boxes A_j are included. For $M > 0$, as data points $\langle x_k, y_k \rangle \in S_M$ are encountered for which $y_k \notin B_0$, all the boxes $A_j \ni x_k$ become excluded from $\underline{P}_M$. Hence $0 < M < M' \rightarrow \underline{P}_M \geq \underline{P}_{M'}$, so that $\underline{P}_M$ is monotonically non-increasing in M.

Consider now a particular box A_j. There are two possibilities:

1. $\forall x \in A_j, f(x) \in B_0$: By equation (2), p_j is not excluded from $\underline{P}$. Moreover, no contradictory data point will be encountered, so p_j can never be excluded from $\underline{P}_M$.
2. $\exists x \in A_j, f(x) \notin B_0$: Now p_j will be excluded from $\underline{P}$. But it might be that no such contradictory x is encountered as an x_k in S_M, so that p_j may or may not be excluded from $\underline{P}_M$.

Thus, we can see that $\forall M > 0, \underline{P} \leq \underline{P}_M$.

So $\underline{P}_M$ is a monotonically non-increasing sequence bounded below by $\underline{P}$. Thus $\underline{P}_M$ has a limit, and it's reasonable to ask if that limit is, indeed $\underline{P}$, and to ask about the convergence of $\underline{P}_M \rightarrow \underline{P}$ as $M \rightarrow \infty$. Our proof actually proves a stronger result: not only does $\underline{P}_M$ give the correct value of $\underline{P}$ in the limit as $M \rightarrow \infty$, but it also does so for a sufficiently large, but finite, number M_0 of sample points.

THEOREM 2. Let $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a continuous function, $S = \{\langle x_1, y_1 \rangle, \ldots, \langle x_k, y_k \rangle, \ldots\}$ be an infinite sequence of pairs such that the values x_k are everywhere dense in X, and $y_k = f(x_k)$. Let S_M be the finite subsequence of S for $k \leq M$, and let $\underline{P}_M$ be defined as in equation (3) on S_M. Then, there exists an integer M_0 such that $\forall M \geq M_0$, $\underline{P}_M = \underline{P}$.

COROLLARY 3. Given the conditions holding in Theorem 2, then

$$\lim_{M \rightarrow \infty} \underline{P}_M = \underline{P}.$$

PROPOSITION 4. The condition that a sequence $\{x_1, \ldots, x_k, \ldots\}$ is everywhere dense is satisfied if to select x_k, we generate independent random vectors—random relative to some probability distribution for which the probability density function $\rho(x)$ is continuous and positive everywhere on the set X.

3.2 Upper probability

By similar reasoning to $\underline{P}$, according to Theorem 1, the probability $\overline{\mathrm{P}}$ is the sum of the probabilities p_j for all boxes j for which $f(A_j) \cap B_0 \neq \emptyset$. This set theoretical condition can also be expressed in logical terms:

$$f(A_j) \cap B_0 \neq \emptyset \equiv \{f(x) : x \in A_j\} \cap B_0 \neq \emptyset \equiv \exists x \in \mathbb{R}^n, x \in A_j \wedge f(x) \in B_0.$$

Thus, the right side of equation (2) can be restated as:

$$\overline{\mathrm{P}} = \sum_{j:\exists x \in \mathbb{R}^n, x \in A_j \wedge f(x) \in B_0} p_j.$$

So in seeking an estimator for $\overline{P}$, we can similarly advance:

$$\overline{P}_M := \sum_{j:\exists k \leq M, x_k \in A_j \wedge y_k \in B_0} p_j. \tag{4}$$

Similarly, but conversely, to $\underline{P}$ we can easily conclude that $\overline{P}_M$ is a monotonically non-decreasing sequence bounded above by $\overline{P}$. Thus it also has a limit. Does this limit equals $\overline{P}$? Well, unlike $\underline{P}$, this limit may be different from the desired value $\overline{P}$.

THEOREM 5. There exists a continuous function $f(x)$ and an everywhere dense sequence $\langle x_k, y_k \rangle$ for which $\overline{P}_M \to 0$ and $\overline{P} = 1$.

However, a natural minor modification of (2) considered in Section 4.2 below will lead to the desired result.

3.3 Comments

- Equations (3) and (4) were originally proposed by Joslyn and Helton [9] as estimators of the Dempster–Shafer uncertainty measures Bel and Pl, respectively.
- As mentioned above, the Dempster–Shafer formalism inspiring this formulation is isomorphic to a random set approach [3]. From this viewpoint, our convergence result can be obtained as a particular case of convergence results for random sets [10].
- How algorithmic are equations (3) and (4)? For each rectangular box $A_{\vec{j}} = \times_{i=1}^{n} A_{j^{(i)}}^{(i)}$, checking whether a given input $x = \langle x^{(1)}, \ldots, x^{(i)}, \ldots, x^{(n)} \rangle$ belongs to this box means checking that for every $i \in \mathbb{N}_n$, the value $x^{(i)}$ belongs to the corresponding interval $A_{j^{(i)}}^{(i)}$ for each $j^{(i)} \in \vec{j}$, i.e. checking $\forall i \in \mathbb{N}_n$, $\forall j^{(i)} \in \vec{j}$, whether $\underline{x}_{j^{(i)}}^{(i)} \leq x^{(i)} \leq \overline{x}_{j^{(i)}}^{(i)}$. So since A_j and B_0 are rectangular boxes, checking whether $x_k \in A_j$ or whether $y_k \in B_0$ means checking n and m corresponding double inequalities respectively. Thus, computing the above estimates $\underline{P}_M$ and $\overline{P}_M$ requires finitely many computational steps.
- For the above algorithm, the number of steps is proportional to the total number of boxes and it can actually be quite large. In Section 5 we will show how we can decrease

the computation time when the number of boxes is large.
- In the proof of Theorem 2 and Corollary 3, we do not use the fact that B_0 is a closed *box*, only that it is a closed *set*.

4. Advanced results

We now consider some other results which hold for this problem.

4.1 Continuity and density requirements

Theorem 2 was proven using two assumptions: that the function $f(x)$ is continuous, and that the sequence x_k is everywhere dense. The following propositions show that both the conditions are necessary.

THEOREM 6. There exists a discontinuous function $f(x)$ and an everywhere dense sequence x_k for which $\underline{P}_M \nrightarrow \underline{P}$.

The counter-example used in the proof is quite natural. Moreover, for discontinuous functions, not only is it the case that our method cannot extract the correct value $\underline{P}$, but moreover, it is impossible to do so by any method.

THEOREM 7. There exists a discontinuous function $f(x)$ and a continuous function $\tilde{f}(x)$, for which $\underline{P} \neq \underline{\tilde{P}}$, but for which, for some everywhere dense sequence x_k, we have $f(x_k) = \tilde{f}(x_k)$ for all k.

Thus, we have the same set of pairs $\langle x_k, y_k \rangle$ to start with, so no matter what method we use, we cannot end up with two different values for $\underline{P}$.

Similarly, if the sequence is not everywhere dense, we cannot reconstruct $\underline{P}$, no matter what method we use.

THEOREM 8. Let the boxes and probabilities be given, and let $B_0 \subset \mathbb{R}^m$ be a closed set. Then, for every sequence x_k which is not everywhere dense in X, there exist different continuous functions $f(x)$ and $\tilde{f}(x)$ for which $\underline{P} \neq \underline{\tilde{P}}$, but for which we have $f(x_k) = \tilde{f}(x_k)$ for all k.

4.2 Modification of upper probability conditions

Similar to the counter-example from the proof of Theorem 6, the counter-example used in the proof of Theorem 5 is also very natural. So, we have to modify the original estimator $\overline{P}_M$ for the upper probability. Fortunately, such a modification is relatively easy and straightforward, introducing only modest considerations for the semantics of the kinds of risk and reliability problems concerning us.

First, we know that the box of desirable states B_0 can be characterized by intervals on the dimensions of the output space as

$$B_0 = \bigtimes_{l=1}^{m} B_0^{(l)} = \bigtimes_{l=1}^{m} \left[\underline{y}_0^{(l)}, \overline{y}_0^{(l)}\right].$$

So now consider a small positive real number $\alpha > 0$, and define the extended desirable box as:

$$B_\alpha := \left[\underline{y}_0^{(1)} - \alpha, \overline{y}_0^{(1)} + \alpha\right] \times \left[\underline{y}_0^{(2)} - \alpha, \overline{y}_0^{(2)} + \alpha\right] \times \ldots \times \left[\underline{y}_0^{(m)} - \alpha, \overline{y}_0^{(m)} + \alpha\right].$$

Elements of this extended box B_α do not necessarily satisfy all $2m$ desired inequalities generated by the output intervals $B_0^{(l)}$, but their deviation from each of these inequalities does not exceed α.

Now, our reliability requirement becomes that $f(x)$ belongs to this extended box, i.e. $f(x) \in B_\alpha$. The upper bound for this probability is then

$$\overline{P}_\alpha = \sum_{j: f(A_j) \cap B_\alpha \neq \emptyset} p_j. \tag{5}$$

Since the modified condition $f(x) \in B_\alpha$ is less demanding than the original condition $f(x) \in B_0$, it is now easier for a state to be desirable, so the probability for a state to be desirable is higher: $\overline{P} \leq \overline{P}_\alpha$. The actual upper probability can be anywhere between $\underline{P}$ and $\underline{P}_\alpha$. From this viewpoint, when we compute an estimate for $\underline{P}$, it is also reasonable, instead of sticking to the original set B_0, to depart from (4) and instead use our slightly enlarged set B_α :

$$\overline{P}_{\alpha,M} := \sum_{j: \exists k \leq M, x_k \in A_j \wedge y_k \in B_\alpha} p_j.$$

THEOREM 9. Let $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a continuous function, $S = \langle x_k, y_k \rangle$, $k > 0$ be an infinite sequence of pairs such that the values x_k are everywhere dense in X, and $y_k = f(x_k)$. Let S_M be the finite subsequence of S for $k \leq M$, and for a small number $\alpha > 0$, let $\overline{P}_{\alpha,M}$ be defined as in equation (3) on S_M. Then there exists an integer M_0 such that $\forall M \geq M_0$, $\overline{P}_{\alpha,M} \in [\overline{P}, \overline{P}_\alpha]$.

The main idea of the above result is that the required bounds on the state variables y_k are not exact, they can be exceeded a little bit—by some small value α—without any harm to the system. In the above result, we used the same value $\alpha > 0$ for every M. Intuitively, the more pairs we have, the more accurately we can describe the requirements. Therefore, it seems reasonable, instead of selecting a single α for all M, to make α decrease to 0 when $M \to \infty$: $\alpha_M \to 0$. Then, if we can still prove the inequality $\overline{P} \leq \overline{P}_{\alpha_M,M} \leq \overline{P}_{\alpha_M}$, we will be able to conclude, in the limit $M \to \infty$, that $\overline{P}_{\alpha_M,M} \to \overline{P}$. But, as shown below, this is not possible—and in this sense, Theorem 9 is the best we can get.

THEOREM 10. Let $\alpha_M \to 0$ be a sequence of positive real numbers. Then, there exists a continuous function $f(x)$ and an infinite sequence of pairs $\langle x_k, y_k \rangle$ for which the values x_k are everywhere dense in X, and $\overline{P}_{\alpha_M,M} \to \overline{P}$.

5. Estimator calculations for large numbers of inputs

Equations (3)–(5) requires every single box j to be analyzed. If for each variable $x^{(i)}$, we have $N^{(i)}$ possible intervals, then have $N = \prod_{i=1}^{N} N^{(i)}$ boxes. In a nuclear facility example, we could have about $n = 100$ variables, and at least two boxes $N^{(i)} \geq 2$ for each variable.

Thus, the total number of boxes is 2^{100}—which is approximately 10^{30}. Testing reveals that many boxes are well beyond the capacity of modern computers. So what do we do?

A natural idea is to use Monte-Carlo simulation to estimate, e.g. $\underline{P}_M$. Indeed, equation (3) can be interpreted as follows. For each box j, let

$$\chi_M(j) = \begin{cases} 1, & \forall k \le M, x_k \in A_j \rightarrow y_k \in B_0 \\ 0, & \text{otherwise} \end{cases}$$

Then equation (3) takes the form $\underline{P} = \sum_{j=1}^{N} p_j \chi_M(j)$. In other words, $\underline{P}$ is a mathematical expectation of $\chi_M(j)$ under the probability distribution in which each box A_j appears with probability p_j.

Recall that we constructed a bijective mapping between the integers $\mathbb{N}_N$ and the vectors $\vec{j} = \langle j^{(1)}, \ldots, j^{(n)} \rangle$. Since the probability p_j is defined as the product $\prod_{i=1}^{N} p_{j^{(i)}}^{(i)}$ of the corresponding probabilities $p_{j^{(i)}}^{(i)}$, to get this probability distribution, it is sufficient to independently select each situation $j^{(i)} \in \vec{j}$ with probability $p_{j^{(i)}}^{(i)}$. This can be done, for example, as follows: we subdivide the interval [0,1] into $N^{(i)}$ subintervals of lengths $p_1^{(i)}, p_2^{(i)}$, etc.—i.e. into the intervals $[0, p_1^{(i)}], [p_1^{(i)}, p_1^{(i)} + p_2^{(i)}]$, etc. and then run a random number generator corresponding to the uniform distribution on [0,1] to select a situation depending on the interval into which the resulting random number falls.

Before showing the algorithm, we introduce some auxiliary computations:

- For each $i \in \mathbb{N}_n$, we compute the values $q_0^{(i)} := 0$ and $q_{j^{(i)}+1}^{(i)} := q_{j^{(i)}}^{(i)} + p_{j^{(i)}+1}^{(i)}$, $0 \le j^{(i)} \le N^{(i)} - 1$.
- For each $k \in \mathbb{N}_M$, we check whether $y_k \in B_0$, i.e. whether the corresponding inequalities $\underline{y}^{(l)} \le y_k^{(l)} \le \bar{y}^{(l)}$ are satisfied for all $l \in \mathbb{N}_m$.

Then, to estimate $\underline{P}$, we select the number of runs R; the larger R is, the better is the estimate.

Now for the algorithm itself, for each run $r \in \mathbb{N}_R$, we do the following:

- For i from 1 to n:
  - Run a random number generator (RNG) corresponding to the uniform distribution on the interval [0,1] and store the result in $r^{(i)}$.
  - By comparing the result $r^{(i)}$ of this RNG with the values $q_{j^{(i)}}^{(i)}$, we find and retain the value $j^{(i)}$ for which $r^{(i)} \in [q_{j^{(i)}-1}^{(i)}, q_{j^{(i)}}^{(i)}]$.
- For every $k \in \mathbb{N}_M$ for which $y_k \notin B_0$, we check whether $x_k \in A_j$, i.e. whether the inequalities $\underline{x}_{j^{(i)}}^{(i)} \le x_k^{(i)} \le \bar{x}_{j^{(i)}}^{(i)}$ hold for all $i \in \mathbb{N}_n$. After that:
  - If for some k for which $y_k \notin B_0$, we have $x_k \in A_j$, this means that

$$\neg\, (\forall k \le M, x_k \in A_j \rightarrow y_k \in B_0),$$

    so we set $\chi_r = 0$;
  - Otherwise, we set $\chi_r = 1$.

Finally, we take the average $\sum_{r=1}^{R} \chi_r / R$ as the desired estimate for $\underline{P}_M$.

The above algorithm is a standard Monte-Carlo algorithm, so when $R \rightarrow \infty$, its result converges to $\underline{P}_M$. Due Theorem 2, for sufficiently large M, we have $\underline{P}_M = \underline{P}$. Therefore, we can conclude that for sufficiently large M, the results of the above algorithm converge to $\underline{P}_M$.

An algorithm for computing $\overline{P}_{\alpha,M}$ is similar to the algorithm for computing $\underline{P}_M$, with the only difference that here:

- $\chi_M(j) = 1$ if $\exists k \le M$, $x_k \in A_j \wedge y_k \in B_\alpha$.
- $\chi_M(j) = 0$ otherwise

THEOREM 1. We need to show three things:

1. For every consistent probability distribution P, $\underline{P} \le \text{Prob}(f(x) \in B_0) \le \overline{P}$.

By definition, the fact that the probability measure $P(A)$ is consistent means that

$$P(A) = \sum_{j=1}^{N} p_j P_j(A)$$

for some probability measures $P_j(A)$ which are concentrated within the corresponding boxes A_j. The probability $\Pr(f(x) \in B_0)$ is equal to the probability that x belongs to the set $f^{-1}(B_0)$ of all values x for which $f(x) \in B_0$, so that $\Pr(f(x) \in B_0) = P(f^{-1}(B_0))$. In particular, for $A = f^{-1}(B_0)$, we have $P := P(f^{-1}(B_0)) = \sum_{j=1}^{N} p_j P_j(f^{-1}(B_0))$.

For each combined situation j, the probability measure $P_j(A)$ is located on the box A_j. Thus:

- If a box A_j has no elements x with $f(x) \in B_0$—i.e. if $f(A_j)$ has no common elements with B_0—we have $P_j(f^{-1}(B_0)) = 0$.
- For those boxes that do contain elements x with $f(x) \in B_0$—i.e. for which $f(A_j)$ has a non-empty intersection with B_0, the conditional probability $P_j(f^{-1}(B_0))$—just like any other probability—cannot exceed 1.

Thus, replacing $P_j(f^{-1}(B_0))$ with 0 for boxes for which $f(A_j)$ that do not intersect B_0 and with 1 for boxes that do, we get an upper bound for P—the upper bound which is exactly the expression $\overline{P}$ from equation (2).

Similarly, since the probability measure $P_j(A)$ is located on the box A_j, we have $P_j(A_j) = 1$. Therefore:

- $P_j(f^{-1}(B_0)) = 1$ for all boxes A_j for which $f(B_0) \subseteq A_j$.
- For all other boxes, the conditional probability is a non-negative number $P_j(f^{-1}(B_0)) \ge 0$—just like any other probability.

Thus, replacing $P_j(f^{-1}(B_0))$ with 0 for boxes for which $f(A_j)$ that do not contain B_0 and with 1 for boxes that do, we get an upper bound for P—the upper bound which is exactly the expression $\underline{P}$ from equation (2).

2. There exists a consistent probability distribution $P(A)$ for which $\Pr(f(x) \in B_0) = \underline{P}$.

We select the probability $P_j(A)$ in one of two ways:

$f(A_j) \subseteq B_0$: Select an arbitrary point $x_j \in A_j$, and select $P_j(A)$ such that this point occurs with probability 1, i.e. for which $P_j(A) = 1$ if $x_j \in A$ and $P_j(A) = 0$, otherwise.

$f(A_j) \subseteq B_0$: By definition of the subset relation, there exists a point $y_j \in f(A_j)$ with $y_j \notin B_0$. Since $y_j \in f(A_j)$, there exists a value $x_j \in A_j$ for which $y_j = f(x_j)$ and for which, therefore, $f(x_j) \notin B_0$. For each such box, as $P_j(A)$, we take a probability distribution in which this point occurs with probability 1, i.e. for which $P_j(A) = 1$ if $x_j \in A$ and $P_j(A) = 0$, otherwise.

For these selected distributions, $P_j(f^{-1}(B_0)) = 1$ for boxes for which $f(A_j) \subseteq B_0$, and $P_j(f^{-1}(B_0)) = 0$ for all other boxes. Thus, for the resulting distribution $P(A)$, the probability $P = P(f^{-1}(B_0))$ is equal to $\underline{P}$.

3. There exists a consistent probability distribution $P(A)$ for which $\Pr(f(x) \in B_0) = \overline{P}$.

- For each box A_j for which $f(A_j) \cap B_0 = \emptyset$, we select an arbitrary point x_j within this box, and, as $P_j(A)$, take a probability distribution in which this point occurs with probability 1, i.e. for which $P_j(A) = 1$ if $x_j \in A$ and $P_j(A) = 0$, otherwise.
- For each box A_j for which $f(A_j) \cap B_0 \neq \emptyset$, by definition of a non-empty set, there exists a point $y_j \in f(A_j) \cap B_0$. Since $y_j \in f(A_j)$, there exists a value $x_j \in A_j$ for which $y_j = f(x_j)$ and for which, therefore, $y_j = f(x_j) \in B_0$. For each such box, as $P_j(A)$, we take a probability distribution in which this point occurs with probability 1, i.e. for which $P_j(A) = 1$ if $x_j \in E$ and $P_j(A) = 0$, otherwise.

Thus, for selected distributions, $P_j(f^{-1}(B_0)) = 0$ for boxes for which $f(A_j) \cap B_0 = \emptyset$, and $P_j(f^{-1}(B_0)) = 1$ for all other boxes. Thus, for the resulting distribution $P(A)$, the probability $P = P(f^{-1}(B_0))$ is equal to $\overline{P}$.

THEOREM 2. We will show that for every box A_j, there exists an integer M_j such that for every $M \geq M_j$, the condition $f(A_j) \subseteq B_0$ is equivalent to $\forall k \leq M, x_k \in A_j \rightarrow y_k \in B_0$. Then, if we take the largest of these values M_j as M_0, we will be able to conclude that for every $M \geq M_0$, these two conditions are equivalent to each other for every box j. Thus, by comparing the formula for $\underline{P}$ from equation (2) and the definition of $\underline{P}_M$ from equation (3), we will be able to conclude that indeed $\underline{P}_M = \underline{P}$.

Let us show that the two conditions are indeed equivalent. The equivalence proof will be different for two cases, when $f(A_j) \subseteq B_0$ and when $f(A_j) \not\subseteq B_0$. Specifically, we will show that:

- When $f(A_j) \subseteq B_0$, then the finite analog of this condition is also satisfied, i.e. $\forall k \leq M$, $x_k \in A_j \rightarrow y_k \in B_0$.
- When $f(A_j) \not\subseteq B_0$, then the finite analog of this condition is also satisfied, i.e. $\neg$ ($\forall k \leq M$, $x_k \in A_j \rightarrow y_k \in B_0$, or, equivalently,

$$\exists k \leq M, x_k \in A_j \wedge y_k \notin B_0.$$

$f(A_j) \subseteq B_0$: For every $x \in A_j$, we have $f(x) \in B_0$. In particular, when $x = x_k$, from $x_k \in A_j$, we will thus be able to conclude that $y_k = f(x_k) \in B_0$, so the finite version of this condition is also satisfied.

$f(A_j) \not\subseteq B_0$: So there exists a point $x* \in A_j$ for which $y* := f(x*) \notin B_0$—or, equivalently, $f(x*) \in \overline{B}_0$, where $\bar{\cdot}$ denotes set complement. Since B_0 is a closed set, its

complement $\overline{B}_0$ is an open set. Therefore, together with a point $f(x*)$, it contains an entire open ball $B_\varepsilon(f(x*)) := \{y : \mathrm{d}(y, y*) < \varepsilon\}$ of a positive radius $\varepsilon > 0$ with a center in $f(x*)$. So, if $\mathrm{d}(y, f(x*)) < \varepsilon$, then $y \in \overline{B}_0$, i.e. $y \notin B_0$.

The function $f(x)$ is continuous, in particular, it is continuous at the point $x = x*$. By definition of continuity, this means that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that if $\mathrm{d}(x, x*) < \delta$, then $\mathrm{d}(f(x), f(x*)) < \varepsilon$. We already know that $\mathrm{d}(f(x), f(x*)) < \varepsilon$ means that $f(x) \in \overline{B}_0$, i.e. $f(x) \notin B_0$. Therefore, we can conclude that when $\mathrm{d}(x, x*) < \delta$, then $f(x) \notin B_0$.

We know that the point $x*$ belongs to the box A_j. There are two possibilities:

$x*$ is strictly inside A_j: Let δ_0 be the smallest distance from $x*$ to any of the edges. Then, $\mathrm{d}(x, x*) < \delta_0$ implies that x is also inside the box A_j. Hence, if we take $\delta_1 := \min(\delta, \delta_0)$, we can conclude the following: when $\mathrm{d}(x, x*) < \delta_1$, then $x \in A_j$ and $f(x) \notin B_0$.

We assumed that the sequence $\{x_k\}$ is everywhere dense in X. By definition, this means that for every point $x \in X$ (in particular, for $x* \in A_j$) and for every $\delta > 0$ (in particular, for $\delta = \delta_1$), there exists a point x_{k*} for which $\mathrm{d}(x_{k*}, x*) < \delta_1$. We already know that in this case, $x_{k*} \in A_j$ and $y_k^* = f(x_{k*}) \notin B_0$. Therefore, for every $M \geq k*$, there exists a $k \leq M$ (namely, $k = k*$) for which $x_k \in A_j \wedge y_k \notin B_0$. Hence, if we take this $k*$ as M_j, then the finite analog of the condition $f(A_j) \not\subseteq B_0$ is indeed satisfied for all $M \geq M_j$.

$x*$ is on the border of A_j: For each point on the border and for every δ, there exists a δ-close point inside the box. So, there exists a point $x**$ inside the box A_j for which $\mathrm{d}(x*, x**) < \delta$—and therefore, $f(x**) \notin B_0$. So, we have an internal point $x** \in A_j$ for which $f(x**) \notin B_0$. For this new point, we can repeat the same proof that we had started with $x*$, and conclude that there exists an M_j such that for every $M \geq M_j$, there exists a $k \leq M$ for which $x_k \in A_j \wedge y_k \notin B_0$—i.e. the finite analog of the condition $f(A_j) \not\subseteq B_0$ is indeed satisfied for all $M \geq M_j$.

COROLLARY 3. Follows immediately.

PROPOSITION 4. Let $x \in \cup_{j=1}^N A_j$ be a point, and $\varepsilon > 0$. Let us show that with probability 1, one of the points x_k will be ε-close to x. Indeed, let $I := X \cap B_\varepsilon(x)$, where $B_\varepsilon(x)$ is the open ball around x. For the probability distribution P_{sel} that we use to select the points x_k, the probability $P_{\mathrm{sel}}(B)$ that a randomly picked vector x is inside B is equal to $P_{\mathrm{sel}}(B) = \int_B \rho(y)\mathrm{d}y$. The intersection B has a positive volume, so, since the probability density function $\rho(y)$ is positive and continuous, this integral $P_{\mathrm{sel}}(B)$ is also positive.

Hence, for every k, the probability that $x_k \notin B$ is equal to $1 - P_{\mathrm{sel}}(B) < 1$. Since we assume that the points are independently selected on each iteration, the probability that on each of M selections, we get a point $x_k \notin B$ is equal to the product of the corresponding M probabilities, i.e. to $(1 - P_{\mathrm{sel}}(B))^M$. When $M \to \infty$, this probability tends to 0, so we conclude that the probability that none of the infinitely many points x_k is inside B is equal to 0. So, with probability 1, there is a point x_k inside B—i.e. a point $x_k \in X$ for which $d(x, x_k) \leq \varepsilon$.

Since every ball contains a smaller ball with rational center and rational radius, it is sufficient to show that we can find x_k within each ball of rational center and radius. There are countably many such balls, and for each, the probability of not having x_k inside it is 0. Thus,

the probability that one of these balls does not contain any of x_k's is also 1—so with probability 1, every ball has a point from x_k, which means, by definition, that the sequence $\{x_k\}$ is everywhere dense.

THEOREM 5. Let $n = 1$, $N^{(1)} = 1$, and $A_1^{(1)} = [0, 1]$ (with $p_1^{(1)} = 1$). Let x_k be an arbitrary everywhere dense sequence of numbers from the open interval $(0, 1)$—e.g. a sequence obtained by using a random number generator that generates numbers uniformly distributed on the interval [0,1].

Let us take $f(x) = x$ and $B_0 = [1, 2]$. Then, $f(A_j) \cap B_0 \neq \emptyset$, so $\overline{P} = 1$. On the other hand, since the values x_k are taken from the inside of the interval [0,1], none of these values is equal to 1, and therefore, none of the points $x_k \in A_j$ has the property that $y_k \in B_0$. So, for every M, we have $\overline{P}_M = 0$.

THEOREM 6. Again, let $n = 1, N^{(1)} = 1$, and $A_1^{(1)} = [0, 1]$ (with $p_1^{(1)} = 1$). Let x_k be an arbitrary everywhere dense sequence of positive numbers—e.g. a sequence obtained by using a random number generator that generates numbers uniformly distributed on the interval [0,1].

Let $f(x)$ be the sign function

$$f(x) := \begin{cases} 0, & x = 0 \\ 1, & x > 0 \\ -1, & x < 0 \end{cases},$$

and let $B_0 = [0.5, 1.5]$. Then, for the only box $A_j = [0, 1]$, we have $f(A_j) = \{0, 1\} \not\subseteq B_0$, so $\underline{P} = 0$. However, since all the values x_k are positive, we have $y_k = f(x_k) = 1$, hence $y_k \in B_0$; thence $\underline{P}_M = 1$ for all M. Here, as $M \to \infty$, we have $\underline{P}_M \to 1$, so $\underline{P}_M \nrightarrow \underline{P}$.

THEOREM 7. Let us take $f(x)$, the boxes, and the everywhere dense sequence the same as in the proof of Theorem 6, and $\tilde{f}(x) = 1$ for all x. Then, $0 = \underline{P} \neq \underline{\tilde{P}} = 1$.

THEOREM 8. Since the sequence x_k is not everywhere dense, there exists a ball $\beta_r(x*) \subseteq \mathbb{R}^n$ that is not covered by any element from this sequence. Since $B_0 \subset \mathbb{R}^m$, there exists a point $y* \in \mathbb{R}^m$, $y* \notin B_0$. Define the function $f(x)$ as a constant equal to some point $y^{**} \in B_0$; for this function, $P(f^{-1}(B_0)) = 1$. As $\tilde{f}(x)$, we take a function

$$\tilde{f}(x) = y^{**} + \max\left(0, 1 - \frac{d(x, x*)}{r}\right) \cdot (y* - y^{**}).$$

One can easily see that $\tilde{f}(x) = f(x)$ for all $x \notin \beta_r(x*)$, so the sequences $\langle x_k, y_k \rangle$ for these two functions are indeed the same. However, since $\tilde{f}(x*) = y* \notin B_0$, there exists a box A_j—namely, any box that contains the point $x*$—for which $f(A_j) \not\subseteq B_0$ and therefore, we have $\underline{\tilde{P}} < 1$ (while $\underline{P} = 1$).

THEOREM 9. We have already proven, in the discussion of the formula for $\overline{P}_M$, that $\overline{P}_M \leq \overline{P}$. Thus, we also have $\overline{P}_{\alpha,M} \leq \overline{P}_\alpha$. So, to complete our proof, we must show that there exists an M_0 such that for every $M \geq M_0$, we have $\overline{P} \leq \overline{P}_{\alpha,M}$.

Similarly to the proof of Theorem 2, we will prove that for every box A_j, there exists a value M_j such that for every $M \geq M_j$, the condition $f(A_j) \cap B_0 \neq \emptyset$ implies that

$$\exists k \leq M, x_k \in A_j \wedge y_k \in B_\alpha.$$

If we prove this, then, for every $M \geq M_0 := \max_j M_j$, we will be able to conclude that all the terms p_j involved in the formula for $\overline{P}$ are also included in the sum that defines $\overline{P}_{\alpha,M}$ and therefore, that indeed $\overline{P} \leq \overline{P}_{\alpha,M}$.

Indeed, let $f(A_j) \cap B_0 \neq \emptyset$. This means that there exists a value $y* \in B_0$ for which $y* \in f(A_j)$, i.e. for which $y* = f(x*)$ for some $x* \in A_j$.

Since the function $f(x)$ is continuous, in particular, it is continuous at the point $x = x*$. By definition of continuity, this means that for every $\alpha > 0$, there exists a $\delta > 0$ such that if $d(x, x*) < \delta$, then $d(f(x), f(x*)) < \alpha/2$. We already know that $y* = f(x*) \in B_0$, hence $f(x)$ belongs to B_α (actually, it even belongs to $B_{\alpha/2}$). Therefore, we can conclude that when $d(x, x*) < \delta$, then $f(x) \in B_\alpha$.

We know that the point $x*$ belongs to the box A_j. There are two possibilities:

$x*$ is inside A_j: Let δ_0 be the smallest distance from $x*$ to any of the edges. Then, $d(x, x*) < \delta_0$ implies that x is also inside the box A_j. Hence, if we take $\delta_1 := \min(\delta, \delta_0)$, we can conclude that when $d(x, x*) < \delta_1$, then $x \in A_j$ and $f(x) \in B_\alpha$.

We assumed that the sequence $\{x_k\}$ is everywhere dense in X. By definition, this means that for every point $x \in X$ (in particular, for $x* \in A_j$) and for every $\delta > 0$ (in particular, for $\delta = \delta_1$), there exists a point x_{k*} for which $d(x_{k*}, x*) < \delta_1$. We already know that in this case, $x_{k*} \in A_j$ and $y_{k*} = f(x_{k*}) \in B_\alpha$. Therefore, for every $M \geq k*$, there exists a $k \leq M$ (namely, $k = k*$) for which $x_k \in A_j \wedge y_k \in B_\alpha$. Hence, if we take this $k*$ as M_j, then the finite analog of the condition $f(A_j) \cap B_\alpha \neq \emptyset$ is indeed satisfied for all $M \geq M_j$.

$x*$ is on the border of A_j: For each point on the border and for every δ, there exists a δ-close point inside the box. So, there exists a point x^{**} inside the box A_j for which $d(x*, x^{**}) < \delta$—and therefore, $f(x^{**}) \in B_{\alpha/2}$. So, we have an internal point $x^{**} \in A_j$ for which $f(x^{**}) \in B_{\alpha/2}$. For this new point, we can repeat the same proof that we had starting with $x*$, and conclude that there exists an M_j such that for every $M \geq M_j$, there exists a $k \leq M$ for which $x_k \in A_j \wedge y_k \in B_\alpha$—i.e. the finite analog of the condition $f(A_j) \cap B_\alpha \neq \emptyset$ is indeed satisfied for all $M \geq M_j$.

THEOREM 10. Without losing generality, we can assume that $\alpha_M < 1$ for all M.

In this proof, we will use the same box and the same continuous function that was used in the proof of Theorem 5 that $\overline{P}_M \not\to \overline{P}$: namely, we take $n = 1$, $N^{(1)} = 1$, $A_1^{(1)} = [0, 1]$ (with $p_1^{(1)} = 1$), $f(x) = x$, and $B_0 = [1, 2]$. For this choice, $\overline{P} = 1$.

We will show that for an appropriately chosen everywhere dense sequence $\langle x_k, f(x_k)\rangle$, we will have $\overline{P}_{\alpha_M,M} = 0$ for all M—and thus, $\overline{P}_{\alpha_M,M} \not\to \overline{P}$.

For this, we must make sure that for every $k \leq M$, we have $y_k \notin B_{\alpha_M}$. For our choice of $B_0 = [1, 2]$, we have $B_\alpha = [1 - \alpha, 2 + \alpha]$. For our choice of $f(x) = x$, we have $y_k = x_k$. Thus, the condition that we need to satisfy is $x_k < 1 - \alpha_M$ for all $M \geq k$. In the limit $M \to \infty$, $\alpha_M \to 0$, so this condition is satisfied—provided, of course, that $x_k < 1$. The requirement that

x_k is smaller than all possible values $1 - \alpha_k, 1 - \alpha_{k+1}, \ldots$, is equivalent to requiring that x_k is smaller than the smallest of the values $1 - \alpha_k, 1 - \alpha_{k+1}, \ldots$. The difference between 1 and a number is the smallest when the subtracted number is the largest, therefore, the above condition is equivalent to $x_k < 1 - \beta_k$, where $\beta_k := \max(\alpha_k, \alpha_{k+1}, \ldots)$.

Since $\alpha_M \to 0$, we can conclude that $\beta_k \to 0$, and one can easily see that β_k is a monotonic sequence: $\beta_k \geq \beta_{k+1} \geq \ldots$ So, to complete the proof, it is sufficient to find an everywhere dense sequence x_k of numbers from the interval [0,1] for which $x_k < 1 - \beta_k$ for some given monotonic sequence $\beta_k \to 0$.

Since $\alpha_M < 1$ for all M, we can conclude that $\beta_k = (\alpha_k, \alpha_{k+1}, \ldots) < 1$ for all k.

To obtain such a sequence x_k, let us start with an arbitrary everywhere dense sequence $y_1, y_2, \ldots$ of numbers from the open interval (0,1). Let us denote $y_0 := 0$. Based on this sequence, we will design a new everywhere dense sequence x_k; this new sequence will consist of zero, one, or several repetitions of $y_0 = 0$, followed by one or several repetitions of y_1, then one or several repetitions of y_2, etc. Since all the elements from y_k are also in the sequence x_k, this new sequence is also everywhere dense in the interval [0,1].

We start by checking whether $y_1 < 1 - \beta_1$.

- If this inequality is satisfied, we start repeating y_1, i.e. take $x_1 = y_1$.
- If this inequality is not satisfied, we take $x_1 = y_0 = 0$.

In both cases, we have $x_1 < 1 - \beta_1$:

- In the first case, it is true due to our choice of x_1.
- In the second case, since $\beta_k < 1$, we have $0 < 1 - \beta_1$.

In general, if we have already selected $x_1, \ldots, x_k$, and $x_k = y_l$ for some l, then, to select x_{k+1}, we check whether $y_{l+1} < 1 - \beta_{k+1}$:

- If this inequality is satisfied, we start repeating y_{l+1}, i.e. take $x_{k+1} = y_{l+1}$.
- If this inequality is not satisfied, we continue to take $x_{k+1} = y_l$.

In both cases, we have $x_{k+1} < 1 - \beta_{k+1}$:

- In the first case, it is true due to our choice of x_{k+1}.
- In the second case, since we had $x_k = y_l < 1 - \beta_k$ and β_k is a monotonic sequence $\beta_k \geq \beta_{k+1}$, we conclude that $x_{k+1} = y_l < 1 - \beta_k \leq 1 - \beta_{k+1}$, i.e. that $x_{k+1} < 1 - \beta_{k+1}$.

To complete the proof, we must show that the process of selecting x_k will not indefinitely stumble on a value y_l and that eventually, it will move on to the next value—thus guaranteeing that all values y_l will be covered. Indeed, the value y_l is selected as x_k only until the inequality $y_{l+1} < 1 - \beta_k$ is not satisfied, i.e. until we have $y_{l+1} \geq 1 - \beta_k$. This cannot be true for arbitrarily large k because then, in the limit $k \to \infty$, we would have $y_{l+1} \geq 1$, and we assumed that all the values y_k are from the open interval (0,1). So, all the values y_l are indeed covered.

Acknowledgements

This work was partly supported by a research grant from Sandia National Laboratories as part of the Department of Energy Accelerated Strategic Computing Initiative (ASCI), by NASA under cooperative agreement NCC5-209 and grant NCC2-1232, by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-00-1-0365, and by NSF grants CDA-9522207, EAR-0112968, EAR-0225670 and 9710940 Mexico/Conacyt.

References

[1] A.P. Dempster, "Upper and lower probabilities generated by a random interval", *Annals of Mathematical Statistics*, 39(3), pp. 957–966, 1968.
[2] S. Ferson, C.A. Joslyn, J.C. Helton, W.L. Oberkampf and K. Sentz, "Summary of the epistemic uncertainty workshop: Consensus amid diversity", *Reliability Engineering and Systems Safety*, 95(1–3), pp. 355–370, 2004.
[3] J. Goutsias, R.P.S. Mahler, and H.T. Nguyen (Eds.). *Random Sets: Theory and Applications*, Springer-Verlag, 1998.
[4] J. Guan and D.A. Bell, *Evidence Theory and its Applications*, v. 1–2, North Holland, 1992.
[5] J.C. Helton, J.D. Johnson, W.L. Oberkampf and C.J. Sallaberry, "Sensitivity analysis in conjunction with evidence theory representations of epistemic uncertainty", *Reliability Engineering and Systems Safety*, 2004, in press.
[6] J.C. Helton, J.D. Johnson and W.L. Oberkampf, "An exploration of alternative approaches to the representation of uncertainty in model predictions", *Reliability Engineering and Systems Safety*, 95(1–3), pp. 39–72, 2004.
[7] J.C. Helton and W.L. Oberkampf, "Special issue: Alternative representations of epistemic uncertainty", *Reliability Engineering and Systems Safety*, 95 (1–3), 2004.
[8] C. Joslyn and S. Ferson, "Approximate representations of random intervals for hybrid uncertainty quantification", *4th Int. Conf. on Sensitivity Analysis of Model Output (SAMO 04)*, ftp://ftp.c3.lanl.gov/pub/users/joslyn/samo04.pdf 2004, p. ?.
[9] C. Joslyn and J.C. Helton, "Bounds on plausibility and belief of functionally propagated random sets", *Proc. Conf. North American Fuzzy Information Processing Society (NAFIPS 2002)*, 2002, pp. 412–417.
[10] S. Li, Y. Ogura, and V. Kreinovich (Eds.). *Limit Theorems and Applications of Set Valued and Fuzzy Valued Random Variables*, Kluwer, 2003.
[11] W.L. Oberkampf, J.C. Helton, C.A. Joslyn, S.F. Wojtkiewicz and S. Ferson, "Uncertainty in system response given uncertain parameters", *Reliability Reliability Engineering and System Safety*, 95(1–3), pp. 11–20, 2004.
[12] R.R. Yager, "Arithmetic and other operations on Dempster-Shafer structures", *Int. J. Man-Machine Studies*, 25, pp. 357–366, 1986.
[13] R.R. Yager, J. Kacprzyk, and M. Pedrizzi (Eds.). *Advances in Dempster-Shafer Theory of Evidence*, Wiley, 1994.

Cliff Joslyn (http://www.c3.lanl.gov/ ∼ joslyn) leads the Knowledge and Information Systems Science research team (http://www.c3.lanl.gov/cic3/teams/knowledge) in the Computer and Computational Science Division at the Los Alamos National Laboratory. His technical background in mathematical systems theory, computational semiotics, complex systems, and qualitative modeling has led to a research portfolio in order theoretical approaches to knowledge discovery and link analysis (ontologies and conceptual representations, hierarchy theory, lattices, partially ordered sets, etc.) and generalized information theory (possibilistic systems theory, Dempster-Shafer theory, finite random intervals, etc.). His work has been applied in computational biology, intelligence analysis, homeland defense, digital libraries, defense transformation, law enforcement, and risk and reliability analysis. He is also an active interdisciplinary systems scientist, and is a founder of the Principia Cybernetica project (http://pcp.vub.ac.be) for the distributed, collaborative development of a cybernetic philosophy. Joslyn

received his B.A. in Cognitive Science and Mathematics in 1985 from Oberlin College with High Honors in Systems Theory, and his M.S. and Ph.D. in 1989 and 1994, respectively, under George Klir in the Systems Science department of the State University of New York at Binghamton. He was a Graduate Fellow in the Software and Automation Systems Branch at NASA 's Goddard Space Flight Center from 1991–1994, and a National Academy of Sciences Postdoctoral Research Associate there from 1994–1996 before joining the Los Alamos National Laboratory as a Postdoctoral Associate. Joslyn serves on the editorial boards of the *International Journal of General Systems*, the *Journal of Biosemiotics*, and *Advances in Complex Systems.*

Vladik Kreinovich received his M.S. in Mathematics and Computer Science from St. Petersburg University, Russia, in 1974, and Ph.D. from the Institute of Mathematics, Soviet Academy of Sciences, Novosibirsk, in 1979. In 1975–80, he worked with the Soviet Academy of Sciences, in particular, in 1978–80, with the Special Astrophysical Observatory (representation and processing of uncertainty in radioastronomy). In 1982–89, he worked on error estimation and intelligent information processing for the National Institute for Electrical Measuring instruments, Russia. In 1989, he was a Visiting Scholar at Stanford University. Since 1990, he is with the Department of Computer Science, University of Texas at El Paso. Also, served as an invited professor in Paris (University of Paris VI), Hong Kong, St. Petersburg, Russia and Brazil. His main interests are: representation and processing of uncertainty, especially interval computations and intelligent control. He published six books and more than 500 papers. He is a member of the editorial board of the international journal "Reliable Computing" (formerly, "Interval Computations"), and several other journals. He is also a co-maintainer of the international website on interval computations.