Decision Making Under Interval Probabilities

Ronald R. Yager[1] and Vladik Kreinovich[2]

[1]Machine Intelligence Institute
Iona College
715 North Avenue
New Rochelle, NY 10801-18903, USA
email yager@panix.com

[2]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
email vladik@cs.utep.edu

Abstract

If we know the probabilities $p_1, \ldots, p_n$ of different situations $s_1, \ldots, s_n$, then we can choose a decision A_i for which the expected benefit $C_i = p_1 \cdot c_{i1} + \ldots + p_n \cdot c_{in}$ takes the largest possible value, where c_{ij} denotes the benefit of decision A_i in situation s_j. In many real life situations, however, we do not know the *exact* values of the probabilities p_j; we only know the *intervals* $\mathbf{p}_j = [p_j^-, p_j^+]$ of possible values of these probabilities. In order to make decisions under such interval probabilities, we would like to generalize the notion of expected benefits to interval probabilities. In this paper, we show that natural requirements lead to a unique (and easily computable) generalization. Thus, we have a natural way of decision making under interval probabilities.

1 Introduction to the Problem

Decision making: case of exactly known consequences. One of the main problems in decision making is the problem of choosing one of (finitely many) alternatives $A_1, \ldots, A_m$. For example, we may choose one of the possible locations of a new airport, one of the possible designs for a new plant; a farmer needs to choose a crop to grow, etc.

In some situations, we know the exact consequences of each choice; in particular, we know the numerical *benefits* (e.g., monetary, utilities, etc.) $C_1, \ldots, C_m$

which characterize the consequences of each choice. In such situations, the choice of the best alternative is easy: we simply choose the alternative A_i for which the value C_i is the largest

$$C_i \to \max_i .$$

Decision making: case of exactly known probabilities. Most frequently, however, for each choice A_i, the exact value of the benefit related to this choice is not known beforehand, because this value depends not only on our choice, but also on some situation which is beyond our control. For example, the farmer's benefits depend not only on his choice of a crop, but also on the weather. Usually, in such cases, we can enumerate all possible situations $s_1, \ldots, s_n$, and for each choice A_i and for each situation s_j, we know (or at least we can estimate) the value c_{ij} of the benefit that this choice will bring in the situation s_j. In such cases, in order to choose the best alternative, it helps to know how probable different situations are.

Traditional methods of decision making (see, e.g., [7, 17, 20]) are based on the assumption that we know the *probabilities* $p_1, \ldots, p_n$ of different situations s_j. In this case, we can take the average (expected) benefit $C_i = p_1 \cdot c_{i1} + \ldots + p_n \cdot c_{in}$ as a measure of quality of each alternative A_i, and select the alternative for which this expected benefit takes the largest possible value:

$$C_i = p_1 \cdot c_{i1} + \ldots + p_n \cdot c_{in} \to \max_i .$$

Decision making: a more realistic case of intervally known probabilities. In some situations, we do not know the *exact* values of the probabilities p_i. Instead, we only have the *intervals* $\mathbf{p}_i = [p_i^-, p_i^+]$ of possible values of probabilities (see, e.g., [15, 21, 31] and references therein).

Example: Cassini mission. As a recent example of the necessity of decision making under interval probabilities, we can cite the planning of a Cassini mission to Saturn; the technical discussion of the corresponding decision issues is presented, e.g., in [18]. Since this mission was sent to the far bounds of the Solar System, where the Solar light is very dim, it could not rely solely on Solar batteries (as usual planetary missions), so a plutonium energy source was added. The preference of a reasonable large amount of such highly radioactive substance as plutonium made a possible launch failure a potential serious health risk.

To make a decision, NASA followed the standard decision making paradigm and tried to estimate the probability of this failure. However, researchers soon pointed out (see [18] for more detail) that due to the large uncertainties in the database, we cannot get the exact probabilities, we can, at best, get an *interval* of possible values of these probabilities. So, instead of using the original *numerical* estimate $\widetilde{p}_1 \approx 10^{-6}$ for the probability of the disaster, the planners should have used the whole *interval* $\mathbf{p}_1 \approx [0, 10^{-3}]$ of possible values of p_1.

Although acknowledged, this idea was not formally implemented in the planning of the actual mission, mainly due to the lack of the appropriate decision making techniques. Many NASA researchers are willing to take these intervals into consideration when planning future missions.

Averaging: a natural idea. Of course, the interval probabilities $\mathbf{p}_i$ must be *consistent*, i.e., there should be values $p_i \in \mathbf{p}_i$ which form a probability distribution (i.e., for which $p_1 + \ldots + p_n = 1$). For each such distribution $p = (p_1, \ldots, p_n)$, we can compute the expected benefit $C_i(p) = p_1 \cdot c_{i1} + \ldots + p_n \cdot c_{in}$; the problem is that in the case of interval uncertainty, there are *many* (actually, *infinitely many*) possible probability distributions, and different distributions lead, in general, to different values of the expected benefit. We would like to somehow combine, "average" these values $C_i(p)$ and come up with a single numerical estimate of the quality of a given alternative. How can we do that?

In this paper, we show how this "average" can be naturally defined. Namely, we describe reasonable requirements on this "average" and then show that these conditions uniquely determine an expression for this "average". Luckily for decision making applications, this expression is easy to compute and is, thus, very practical.

2 Towards a Formalization of the Problem

The desired quality C_i of an alternative A_i should only depend on the properties of this particular alternative, and it should not depend on what other alternatives are there. So, when computing C_i, we must only take in to consideration, for each situation s_j, its interval probability $\mathbf{p}_j$ and the benefits c_{ij} which corresponds to this situation s_j (and we will not need the values c_{kj} for $j \neq i$). In view of this comment, we can simplify our notations by dropping the index i (which characterizes the alternative), and denote the benefit corresponding to the situation s_j by c_j instead of c_{ij}.

In these simplified notations, we can re-formulate our problem as follows:

- we have a finite sequence of pairs $\langle \mathbf{p}_j, c_j \rangle$, $1 \leq j \leq n$, (with consistent probability intervals $\mathbf{p}_j$); and
- we need to transform this sequence into a single number C.

In other words, we must design a *function* C which takes, as input, an arbitrary consistent finite sequence of pairs $\langle \mathbf{p}_j, c_j \rangle$ and which returns a desired estimate

$$C(\langle \mathbf{p}_1, c_1 \rangle, \ldots, \langle \mathbf{p}_n, c_n \rangle).$$

There are some natural properties that we expect from this function.

1. First, we want to make sure that when we know the probabilities *exactly*, i.e., when all the intervals are degenerate $\mathbf{p}_i = [p_i, p_i]$, we get the expected value:

$$C(\langle [p_1, p_1], c_1 \rangle, \ldots, \langle [p_n, p_n], c_n \rangle) = p_1 \cdot c_1 + \ldots + p_n \cdot c_n. \tag{1}$$

2. A similar relation must be true when there is an uncertainty, but this uncertainty is fictitious: namely, if we have only two situations, and we know the exact probability p_1 for one of them (i.e., $\mathbf{p}_1 = [p_1, p_1]$), then, although we may be given a non-degenerate interval $\mathbf{p}_2$ for the second probability, we know that, due to the equality $p_1 + p_2 = 1$, the only possible value of this second probability is $p_2 = 1 - p_1$. In this case, the width of the interval $\mathbf{p}_2$ is irrelevant and it is therefore reasonable to require that the resulting benefit will be the same whether we use a wide interval $\mathbf{p}_2$, or the degenerate interval $[1 - p_1, 1 - p_1]$:

$$C(\langle [p_1, p_1], c_1 \rangle, \langle \mathbf{p}_2, c_2 \rangle) = C(\langle [p_1, p_1], c_1 \rangle, \langle [1 - p_1, 1 - p_1], c_2 \rangle). \tag{2}$$

3. The third desired property comes from the fact that the order of the situations is usually pretty much arbitrary: what was a situation # 1 could as well be situation # 5, and vice versa. Therefore, the value of the desired function should not change if we simply swap i-th and j-th situations:

$$\begin{gathered} C(\langle \mathbf{p}_1, c_1 \rangle, \ldots, \langle \mathbf{p}_{i-1}, c_{i-1} \rangle, \langle \mathbf{p}_i, c_i \rangle, \langle \mathbf{p}_{i+1}, c_{i+1} \rangle, \ldots, \\ \langle \mathbf{p}_{j-1}, c_{j-1} \rangle, \langle \mathbf{p}_j, c_j \rangle, \langle \mathbf{p}_{j+1}, c_{j+1} \rangle, \ldots, \langle \mathbf{p}_n, c_n \rangle) = \\ C(\langle \mathbf{p}_1, c_1 \rangle, \ldots, \langle \mathbf{p}_{i-1}, c_{i-1} \rangle, \langle \mathbf{p}_j, c_j \rangle, \langle \mathbf{p}_{i+1}, c_{i+1} \rangle, \ldots, \\ \langle \mathbf{p}_{j-1}, c_{j-1} \rangle, \langle \mathbf{p}_i, c_i \rangle, \langle \mathbf{p}_{j+1}, c_{j+1} \rangle, \ldots, \langle \mathbf{p}_n, c_n \rangle). \end{gathered} \tag{3}$$

4. The fourth desired property comes from the following: whatever are the (unknown) actual probabilities $p_j \in \mathbf{p}_j$, the benefit cannot be worse than the worst of the possibilities and cannot be better than the best of the possibilities. In other words, the desired value C must always be between $\min c_j$ and $\max c_j$:

$$\min_j c_j \le C(\langle [p_1, p_1], c_1 \rangle, \ldots, \langle [p_n, p_n], c_n \rangle) \le \max_j c_j. \tag{4}$$

5. The fifth property is related to the fact that while we have so far considered a *single* decision process (choosing A_i), we may have *two* or more independent decisions one after another:

- first choosing an alternative A_i from the *first* list of alternatives $A_1, \ldots, A_m$, and then
- choosing an alternative A'_k from the *second* list of alternatives $A'_1, \ldots, A'_q$.

The fact that these choices are independent means that for each pair of choices A_i and A'_k, the resulting benefit $\widetilde{c}_j$ in situation s_j is simply equal to the sum of the two benefits: the benefit c_{ij} of choosing A_i and the benefit c'_{kj} of choosing A'_k. It is natural to require that in such a situation, the expected benefit of the situation s_j for the double choice is simply equal to the sum of expected benefits corresponding to c_j and c'_j. In other words, we require that

$$C(\langle \mathbf{p}_1, c_1 + c'_1 \rangle, \ldots, \langle \mathbf{p}_n, c_n + c'_n \rangle) =$$

$$C(\langle \mathbf{p}_1, c_1 \rangle, \ldots, \langle \mathbf{p}_n, c_n \rangle) + C(\langle \mathbf{p}_1, c'_1 \rangle, \ldots, \langle \mathbf{p}_n, c'_n \rangle). \tag{5}$$

6. The sixth property is related to the following fact: When we analyze the possible consequences of our decisions, we try to list all possible situations by imagining all possible combinations of events. Some of these events may be relevant to our decision, some may later turn out to be irrelevant. As a result, we may end up with two different situations, say s_1 and s_2, which result in the exact same benefit value $c_1 = c_2$. To simplify computations, it is desirable to combine these two situations into a single one.

If we know the exact probabilities p_1 and p_2 of each of the original situations, then the probability of the combined situation is equal to $p_1 + p_2$. If we do not know the exact probability of each situation, i.e., if we only know the *intervals* of possible values $\mathbf{p}_1 = [p_1^-, p_1^+]$ and $\mathbf{p}_2 = [p_2^-, p_2^+]$ of these probabilities, then the probability of the combined event can take any value $p_1 + p_2$ where $p_1 \in \mathbf{p}_1$ and $p_2 = \mathbf{p}_2$. This set of possible values is known to be also an interval, with the bounds $[p_1^- + p_2^-, p_1^+ + p_2^+]$. In interval computations (see, e.g., [8, 9, 11, 12, 13, 19]), this new interval is called the *sum* of the two intervals $\mathbf{p}_1$ and $\mathbf{p}_2$ and denoted by $\mathbf{p}_1 + \mathbf{p}_2$.

The benefit of the decision should not change if we simply combine the two actions with identical consequences into one. In other words, we must have:

$$C(\langle \mathbf{p}_1, c_1 \rangle, \langle \mathbf{p}_2, c_1 \rangle, \langle \mathbf{p}_3, c_3 \rangle, \ldots, \langle \mathbf{p}_n, c_n \rangle) =$$

$$C(\langle \mathbf{p}_1 + \mathbf{p}_2, c_1 \rangle, \langle \mathbf{p}_3, c_3 \rangle, \ldots, \langle \mathbf{p}_n, c_n \rangle). \tag{6}$$

7. Finally, small changes in the probabilities p_j^- or p_j^+ or small changes in benefits c_j should not drastically affect the resulting benefit function C. In other words, we want the function C to be *continuous* for any given n.

3 Definitions and the Main Result

Definition 1.

- *By an interval probability* $\mathbf{p}$, *we mean an interval* $\mathbf{p} = [p^-, p^+] \subseteq [0,1]$.
- *We say that a finite sequence of interval probabilities* $\mathbf{p}_1, \ldots, \mathbf{p}_n$ *is consistent* (or, *to be more accurate, forms an interval probability distribution*), *if there exist values* $p_1 \in \mathbf{p}_1$, $\ldots$, $p_n \in \mathbf{p}_n$ *for which* $p_1 + \ldots + p_n = 1$.

Proposition 1. *A sequence of interval probabilities* $\mathbf{p}_1 = [p_1^-, p_1^+], \ldots, \mathbf{p}_n = [p_n^-, p_n^+]$ *is consistent if and only if* $p_1^- + \ldots + p_n^- \le 1 \le p_1^+ + \ldots + p_n^+$.

Proof: By definition, a sequence of probability intervals is consistent if and only if 1 can be represented as $p_1 + \ldots + p_n$ for some $p_j \in \mathbf{p}_j$. According to the above definition of the sum of intervals, this condition is, in its turn, equivalent to $1 \in \mathbf{p}_1 + \ldots + \mathbf{p}_n$. From the above result about the sum of the intervals, we know the exact expression for the endpoints of the interval $\mathbf{p}_1 + \ldots + \mathbf{p}_n$, so the fact that 1 belongs to this intervals can expressed by the inequalities given in the formulation of the proposition. The proposition is proven.

Definition 2. *By an averaging operation for interval probabilities, we mean a function* C *that transforms every finite sequence of pairs*

$$(\langle \mathbf{p}_1, c_1 \rangle, \ldots, \langle \mathbf{p}_n, c_n \rangle)$$

with consistent interval probabilities into a real number

$$C(\langle \mathbf{p}_1, c_1 \rangle, \ldots, \langle \mathbf{p}_n, c_n \rangle),$$

which is continuous for any n, *and which satisfies the conditions* (1)–(6).

Theorem. *There exists exactly one averaging operation with interval probabilities, and this averaging operation has the form*

$$C(\langle [p_1^-, p_1^+], c_1 \rangle, \ldots, \langle [p_n^-, p_n^+], c_n \rangle) = \widetilde{p}_1 \cdot c_1 + \ldots + \widetilde{p}_n \cdot c_n, \tag{7}$$

where

$$\widetilde{p}_j = \frac{\Sigma^+ - 1}{\Sigma^+ - \Sigma^-} \cdot p_j^- + \frac{1 - \Sigma^-}{\Sigma^+ - \Sigma^-} \cdot p_j^+, \tag{8}$$

$$\Sigma^- = p_1^- + \ldots + p_n^-, \tag{9}$$

and

$$\Sigma^+ = p_1^+ + \ldots + p_n^+. \tag{10}$$

Comments.

- So, if we have several alternatives A_i, and we know:
  - the benefits c_{ij} of each alternative under each situation s_j, and
  - the interval probability $\mathbf{p}_j = [p_j^-, p_j^+]$ of each situation,

  we recommend to select a decision A_i for which

  $$C_i = \widetilde{p}_1 \cdot c_{i1} + \ldots + \widetilde{p}_n \cdot c_{in} \to \max_i,$$

  where $\widetilde{p}_j$ are determined by the formulas (7) and (8).

- Formula (8) can be re-written in the following equivalent form:

  $$\widetilde{p}_j = p_j^- + \frac{\Delta p_j}{\Delta p} \cdot (1 - p_1^- - \ldots - p_n^-), \tag{8a}$$

  where $\Delta p_j = p_j^+ - p_j^-$, and $\Delta p = \Delta p_1 + \ldots + \Delta p_n$. Since $\Sigma^- = p_1^- + \ldots + p_n^- \le 1$, what we are doing is essentially adding to the lower probability p_j^- an amount proportional to the width $\Delta p_j = p_j^+ - p_j^-$ of the corresponding probability interval $[p_j^-, p_j^+]$. This width is a natural measure of uncertainty with which we know the probabilities.

- Alternatively, we can represent formula (8) in another equivalent form:

  $$\widetilde{p}_j = p_j^+ - \frac{\Delta p_j}{\Delta p} \cdot (p_1^+ - \ldots + p_n^+ - 1), \tag{8b}$$

  Since $\Sigma^+ = p_1^+ + \ldots + p_n^+ \ge 1$, what we are doing is essentially subtracting to the upper probability p_j^+ an amount proportional to the width $\Delta p_j = p_j^+ - p_j^-$ of the corresponding probability interval $[p_j^-, p_j^+]$.

- The proof of the Theorem is given in Appendix 1.

Examples:

- If all the interval probabilities coincide, we get $\widetilde{p}_j = 1/n$ for all j, so we must choose an alternative A_i for which

  $$C_i = \frac{c_{i1} + \ldots + c_{in}}{n} \to \max_i.$$

- If we only know the *upper* bounds p_i^+ for the probabilities, i.e., if $p_i^- = 0$ for all i, then

  $$\widetilde{p}_j = \frac{p_j^+}{p_1^+ + \ldots + p_n^+}.$$

In this example, we must choose an alternative A_i for which

$$C = \frac{p_1^+ \cdot c_1 + \ldots + p_n^+ \cdot c_n}{p_1^+ + \ldots + p_n^+} \to \max.$$

- If we only know the *lower* bounds p_i^- for the probabilities, i.e., if $p_i^+ = 1$ for all i, then

$$\widetilde{p}_j = \frac{n-1}{n - \sum p_i^-} \cdot p_j^- + \frac{1 - \sum p_i^-}{n - \sum p_i^-}.$$

Comment. We have already mentioned that sometimes, the interval uncertainty is *fictitious*: e.g., if we have only two situations, and we know the exact probability p_1 for one of them (i.e., $\mathbf{p}_1 = [p_1, p_1]$), then, although we may be given a non-degenerate interval $\mathbf{p}_2$ for the second probability, we know that, due to the equality $p_1 + p_2 = 1$, the only possible value of this second probability is $p_2 = 1 - p_1$. In general, if we have n interval probabilities $[p_i^-, p_i^+]$, $1 \le i \le n$, then, due to the condition $p_1 + \ldots + p_n = 1$, we have $p_i = 1 - (p_1 + \ldots + p_{i-1} + p_{i+1} + \ldots + p_n)$; therefore, the actual value of p_i must lie between $1 - (p_1^+ + \ldots + p_{i-1}^+ + p_{i+1}^+ + \ldots + p_n^+)$ and $1 - (p_1^- + \ldots + p_{i-1}^- + p_{i+1}^- + \ldots + p_n^-)$. As a result, for each i from 1 to n, only the values from the "reduced" interval $\mathbf{p}'_i = [p'^-_i, p'^+_i]$ are possible, where:

$$p'^-_i = \max\left(p_i^-, 1 - \sum_{j \ne i} p_i^+\right) \text{ and } p'^+_i = \min\left(p_i^+, 1 - \sum_{j \ne i} p_i^-\right). \tag{11}$$

For example, if we start with a sequence $\mathbf{p}_1 = [0, 1]$, $\mathbf{p}_2 = [0, 0.5]$, we get new intervals $\mathbf{p}'_1 = [0.5, 1]$ and $\mathbf{p}'_2 = [0, 0.5]$. Here, the interval $\mathbf{p}'_1$ is narrower than the original interval $\mathbf{p}_1 = [0, 1]$. In such situations, when one of these new intervals is narrower than the original one, this means that a part of the original uncertainty is "fictitious".

There are two possible approaches to such "fictitious" uncertainty:

- In the above text, we assumed that we can have an arbitrary sequence of interval probabilities which is consistent in the sense of Definition 1. In particular, we may have a sequence $\mathbf{p}_1 = [p_1, p_1]$, $\mathbf{p}_2 = [1 - p_1, 1]$, which includes "fictitious" probabilities. For this case, Theorem 1 justifies the use of formulas (7)–(10).
- Alternatively, we can first *reduce* the original sequence of probability distributions to a new sequence (11), and then apply the formulas (7)–(10) to the resulting sequence $\mathbf{p}'_1, \ldots, \mathbf{p}'_n$. The justification of this second approach is provided by the following: if, in Definition 2, we restrict ourselves only to *reduced* sequences of interval probabilities, then our proof of Theorem 1 shows, in effect, that thus restricted mapping is described by the same formulas (7)–(10).

4 Relation with other approaches to decision making

4.1 Averaging and Hurwicz criterion

Yet another reformulation of our result. The above formula (8) can be re-formulates as follows:

$$\widetilde{p}_j = \alpha \cdot p_j^- + (1-\alpha) \cdot p_j^+, \tag{8c}$$

where $\alpha = (\Sigma^+ - 1)/(\Sigma^+ - \Sigma^-)$. One can easily check that thus defined α belongs to the interval $[0,1]$. Thus, this formula is similar to another approach to decision making, originally proposed by Hurwicz. To explain how exactly these two approaches are similar, let us first briefly describe Hurwicz's approach.

Hurwicz criterion. This approach has been proposed for the situations in which we have no information about the probabilities p_j (i.e., in our terms, when $\mathbf{p}_j = [0,1]$ for all j).

In other words:

- for decision making, we want, for each alternative A_i, to find a numerical value C_i that would characterize the utility of this alternative;
- we do not know the exact value of *the* utility of each alternative A_i; instead, we know a *set* of possible values of utility $\{c_{i1}, \ldots, c_{in}\}$ that characterize the outcome of this action A_i in different situations;
- we do not know which of the situations is more probable and which is less probable, and therefore, we do not know which elements of this set are more probable, and which are less probable.

For this situation, Hurwicz has proposed [10, 17] to choose a real number $\alpha \in [0,1]$, and then characterize each alternative A_i by the value

$$C_i = \alpha \cdot \min\{c_{i1}, \ldots, c_{in}\} + (1-\alpha) \cdot \max\{c_{i1}, \ldots, c_{in}\}.$$

The meaning of this formula depends on α:

- When $\alpha = 0$, we judge its alternative based on the its most optimistic outcome: $C_i = \max\{c_{i1}, \ldots, c_{in}\}$.
- When $\alpha = 1$, we judge each alternative based on its most pessimistic outcome: $C_i = \min\{c_{i1}, \ldots, c_{in}\}$.
- When $0 < \alpha < 1$, we use a realistic mix of pessimistic and optimistic estimates to judge its alternative A_i.

Analogy with our situation. We have a similar situation:

- for decision making, we want, for each situation s_j, to find a numerical value $\widetilde{p}_j$ that would characterize the probability of this situation;
- we do not know the exact value of *the* probability of each situation s_j; instead, we know a *set* of possible values of probability $[p_j^-, p_j^+]$;
- we do not know which elements of this set are more probable, and which are less probable.

Following Hurwicz's idea, we can fix a real number $\alpha \in [0,1]$ and characterize each situation s_j by the numerical value

$$\widetilde{p}_j = \alpha \cdot \min\{p_j \mid p_j \in [p_j^-, p_j^+]\} + (1-\alpha) \cdot \max\{p_j \mid p_j \in [p_j^-, p_j^+]\}.$$

The corresponding minimum and maximum are, of course, equal to p_j^- and p_j^+ and therefore, we get exactly the formula (8c). The value α can be uniquely determined from the condition that the values $\widetilde{p}_j$ form a probability distribution, i.e., that $\widetilde{p}_1 + \ldots + \widetilde{p}_n = 1$. So, the *averaging operation can be viewed as an analogue of Hurwicz criterion.*

Comment. For different α, the formula (8c) has been successfully used in decision making; see, e.g., [2, 3, 4, 16, 24, 25, 26, 28, 29].

4.2 Averaging and maximum entropy approach

Maximum entropy approach. Averaging over all possible distributions is not the only possible approach. Alternatively, instead of considering *all* possible probability distributions which are consistent with the given interval al probabilities, we can select *one* probability distribution which is, in some reasonable sense, the most representative, and make decisions based on this "most representative" distribution.

One natural way of selecting the "most representative" distribution is the *maximum entropy* approach (see, e.g., [6, 14], and references therein; see also [5, 22, 27, 29]), according to which we select a probability distribution p_j for which the entropy $S = -\sum p_j \cdot \log(p_j)$ take the largest possible value. This distribution is relatively easy to describe [14]: there exists a value p_0 such that for all j:

- when $p_j^+ \le p_0$, we take $p_j = p_j^+$;
- when $p_0 \le p_j^-$, we take $p_j = p_j^-$;
- when $p_j^- \le p_0 \le p_j^+$, we take $p_j = p_0$.

This value p_0 can be computed by a quadratic-time (i.e., quite feasible) algorithm [14]. In particular, if all the interval probabilities coincide, then $p_1 = \ldots = p_n = p_0 = 1/n$.

In general, these two approaches lead to different results. In the above example, our "averaging" approach leads to the same value as the maximum entropy approach. However, in general, the resulting benefit $p_1 \cdot c_1 + \ldots + p_n \cdot c_n$ is, *different* from the one produced by averaging. As an example of this difference, let us consider the case when we have two possible situations: a situation s_1 with a small interval probability $\mathbf{p}_1 = [0, p_{\text{small}}]$ ($p_{\text{small}} \ll 1$), and a situation s_2 with the interval probability $\mathbf{p}_2 = [1 - p_{\text{small}}, 1]$. In this case:

- For *averaging*, we have $\Sigma^- = 1 - p_{\text{small}}$, $\Sigma^+ = 1 + p_{\text{small}}$, so averaging leads to $\widetilde{p}_1 = p_{\text{small}}/2$ and $\widetilde{p}_2 = 1 - p_{\text{small}}/2$. This is exactly what we can intuitively expected from averaging:
  - an average of the interval $[0, p_{\text{small}}]$ is its midpoint $p_{\text{small}}/2$, and
  - an average of the interval $[1 - p_{\text{small}}, 1]$ is its midpoint $1 - p_{\text{small}}/2$.
- For maximum entropy approach, we get $\widetilde{p}_1 = p_{\text{small}}$ and $\widetilde{p}_2 = 1 - p_{\text{small}}$.

Comment. Informally, the difference between the two approaches can be explained as follows. When all the interval probabilities coincide, both approaches return the same values of equal probabilities $\widetilde{p}_j = 1/n$. In other words, informally, both approaches try to get the probabilities as close to be equal as possible. In this, both approaches agree; the difference is in how these two approaches interpret the word "close": the maximum entropy approach uses a non-linear expression (entropy) to describe this "closeness", while in the averaging approach, we only consider expressions which are linear in p_j.

Which approach is better? Which of the two approaches is better: maximum entropy or averaging? On a general *methodological* level:

- there are arguments in favor of the maximum entropy approach (see, e.g., [6, 14]),
- but there are also arguments in favor of our averaging: e.g., unlike the maximum entropy approach, our "averaging" solution takes into consideration not just a *single* distribution, but *all* probability distributions consistent with the given interval probabilities.

From the *practical* viewpoint, which of these approaches is better depends on the objective that we want to achieve in a practical problem. For example, if the first situation s_1 leads to negative consequences, then the maximum entropy approach means that we consider the *worst-case* (*pessimistic*) scenario by assuming the worst possible probability of this negative situation, while the averaging approach takes a reasonable mid-point of the interval. So:

- if our objective is to avoid the worst-case scenario at any cost, we should use maximum entropy method;
- on the other hand, if s_1 is a reasonable risk, then averaging seems to be more reasonable.

4.3 What if, in addition to interval probabilities $\mathbf{p}_i$, we also know the probabilities of different values within the intervals $\mathbf{p}_i$?

Description of the problem and the resulting formula. In the above text, we assumed that the only information that we have about the (unknown) probabilities p_j is that each of these probabilities belong to the corresponding interval $\mathbf{p}_j = [p_j^-, p_j^+]$. In some cases, however, the estimates p_j^- and p_j^+ themselves come from a statistical analysis of the existing records. In this case, in addition to *intervals* $\mathbf{p}_i$, we may also know the *probabilities* of different values within the intervals $\mathbf{p}_i$.

For example, we can simply look at all recorded situations, and count how many of them were situations s_j. If out of N total records, the situation s_j occurred in $N_j$of them, then we can take the frequency $f_j = N_j/N$ as a natural estimate for the probability p_j (for details on statistical methods, see any statistical textbook, e.g., [30]).

When the total number of records (N) is large, the error of this estimation, i.e., the difference $f_j - p_j$ between the frequency and the actual probability, is negligible small. However, in many real-life cases, N is not too large, so this difference is not negligible. It is known in statistics that the probability distribution for this difference $f_j - p_j$ is approximately Gaussian (and the larger N, the closer this distribution to Gaussian), with 0 average and known standard deviation σ_j. So, the desired probability $p_j = f_j - (f_j - p_j)$ is distributed according to the Gaussian distribution with the average f_j and standard deviation σ_j. Different estimation errors $f_j - p_j$ are independent random variables, so the random variables p_j, p_k, $j \neq k$ are independent too.

How is this information related to intervals? In practically applications of statistics, if we assume a Gaussian probability distribution with average m and standard deviation σ, and we observe a value x which is farther than $k \cdot \sigma$ from m (for some fixed k), we conclude that the distribution was wrong.

> For example, if we test a sensor with the supposed standard deviation $\sigma = 0.1$, and as a result of the testing, we get an error $x - m = 1.0$, then it's is natural to conclude (for all $k < 10$) that the sensor is malfunctioning.

Of course, for every k, there is a non-zero probability that the random variable x attains a value outside the interval $[m - k \cdot \sigma, m + k \cdot \sigma]$, but for large k, this probability is very small. In practical applications, people normally use $k = 2$ (for which the probability of error outside the interval is $\approx 5\%$), $k = 3$

(for which the probability of error outside the interval is $\approx 0.1\%$), and, in VLSI design and other important computer engineering applications, $k = 6$ (for which the probability of error outside the interval is $\approx 10^{-6}\%$). So, if we fix a value k (=2, 3, or 6), we conclude that the actual value of p_j must fit within the interval $\mathbf{p}_j = [p_j^-, p_j^+]$, where

$$p_j^- = f_j + k \cdot \sigma_j, \quad p_j^+ = f_j - k \cdot \sigma_j.$$

Vice versa, if we know this interval $[p_j^-, p_j^+]$ and the value k, we can reconstruct the parameters f_j and σ_j of the corresponding Gaussian distribution as

$$m_j = \frac{p_j^- + p_j^+}{2}, \quad \sigma_j = \frac{p_j^+ - p_j^-}{2k}.$$

If we know the distributions for p_j, then the problem of computing the average values $\widetilde{p}_j$ becomes a standard probability problem: namely, as $\widetilde{p}_j$, we take the conditional expectation of p_j under the condition that the sum of all the probabilities is 1, i.e.,

$$\widetilde{p}_j = E(p_j \,|\, p_1 + \ldots + p_n = 1).$$

We can now use the standard techniques of multi-dimensional Gaussian distributions to calculate this conditional expectation. Detailed derivation is given in Appendix 2; here we just present the result:

$$\widetilde{p}_j = f_j + \frac{(1 - \Sigma_0) \cdot \sigma_j^2}{\sum \sigma_k^2},$$

where $\Sigma_0 = f_1 + \ldots + f_n$. If we substitute, into this formula, the expressions for f_j and σ_j in terms of p_j^- and p_j^+, we get the following expression:

$$\widetilde{p}_j = \frac{p_j^- + p_j^+}{2} + \frac{(1 - \Sigma_0) \cdot (p_j^+ - p_j^-)^2}{\sum (p_k^+ - p_k^-)^2},$$

where $\Sigma_0 = (\Sigma^- + \Sigma^+)/2$.

Relation to averaging. How different is this formula from the interval-based average? If all the intervals $\mathbf{p}_j$ are of the same width, then (as one can easily see) we get the exact same averaging formula. However, if the intervals are of different width, we get different formulas: e.g., for $\mathbf{p}_1 = [0, 0.5]$ and $\mathbf{p}_2 = [0, 1]$:

- for interval-based averaging, we get:
  $\Sigma^- = 0$, $\Sigma^+ = 1.5$, so $\widetilde{p}_1 = 1/3$ and $\widetilde{p}_2 = 2/3$, while
- the statistical averaging, we get:
  $\Sigma_0 = (0 + 1.5)/2 = 0.75$, so

$$\widetilde{p}_1 = \frac{0+0.5}{2} + \frac{(1-0.75)\cdot(0.5-0)^2}{(0.5-0)^2+(1-0)^2} = 0.25 + \frac{0.25\cdot 0.25}{1.25} =$$
$$0.25 + 0.05 = 0.3 \neq \frac{1}{3},$$

and similarly $\widetilde{p}_2 = 0.7 \neq 2/3$.

Conclusion

Making decisions when there exists some uncertainty with respect to the payoff to be received as a result of one's choice of action is clearly the normal state of affairs. In many of these situations, the expected value is used as a means for comparing different courses of action. In order to calculate the expected value in a way to allow comparisons, precise information about the probability of each of the different possible payoffs resulting from an alternative is needed. Unfortunately this type of precise probabilistic information is often not available. In this work we have investigated the problem of calculating the expected value when the knowledge about the probabilities is of an *interval* type. Difficulty in attaining precise probabilities becomes particularly manifest in environments in which human beings are the source of the information about the probabilities. What is becoming quite apparent now is the human inclination for *granularization*, a fact closely related to the use of language. As a result of this quality, humans feel more comfortable providing interval probability estimates then precise exact values. This inclination is often due to the fact that *perceptions* rather then measurements are the basis of the knowledge supplied. It is our feeling that the technique developed here will play a central role in the construction of intelligent decision making systems, particularly those based upon the idea of computing with words.

Acknowledgments. This work was partially supported by NASA under cooperative agreement NCC5-209, by NSF grants No. DUE-9750858 and CDA-9522207, by the United Space Alliance grant No. NAS 9-20000 (PWO C0C67713A6), 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 number F49620-95-1-0518, and National Security Agency under Grant No. MDA904-98-1-0564.

The authors are thankful to Piero P. Bonissone and to the anonymous referees for valuable suggestions.

References

[1] J. Aczél and J. Dhombres, *Functional equations in several variables*, Cambridge University Press, Cambridge, 1989.

[2] U. Bergsten, J. Schubert, and P. Swennson, "Beslutsstödssystemet Dezzy - En Översikt", In: *Dokumentation 7 juni av Seminarium och fackutställning om samband, sensorer och datorer för legningssystem till försvaret, MILINF 89, Försvarets materielverk, Enköping, July 5–9, 1989*, Telub AB, Växjö, 1989, 07B2:19–07B2:31, FOA Report B 20078–2.7, Department of Weapon Systems, Effects and Protection, National Defence Research Establishment, Stockholm, Sweden, 1990.

[3] U. Bergsten, J. Schubert, and P. Swennson, "Dezzy - ett Demonstrationssystem för Analys av Ubatsunderrättelser", In J. Schubert (ed.), *Delprojekt Informationssytems inom Huvudprojekt Ubatsskydd - Slutparrort*, FOA Report A 20046–2.7, Department of Weapon Systems, Effects and Protection, National Defence Research Establishment, Stockholm, Sweden, 1990.

[4] U. Bergsten and J. Schubert, "Dempster's rule for evidence ordered in a complete directed acyclic graph", *Int. J. Approx. Reasoning*, 1993, Vol. 9, No. 1, pp. 37–73.

[5] D. Dubois and H. Prade, "On several representations of an uncertain body of evidence", In: M. M. Gupta and E. Sanchez (eds.), *Fuzzy information and decision processes*, North Holland, Amsterdam, 1982, pp. 167–181.

[6] G. Erickson (ed.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 1997.

[7] P. C. Fishburn, *Utility theory for decision making*, John Wiley & Sons Inc., New York, 1969.

[8] R. Hammer, M. Hocks, U. Kulisch, D. Ratz, *Numerical toolbox for verified computing. I. Basic numerical problems*, Springer Verlag, Heidelberg, N.Y., 1993.

[9] E. R. Hansen, *Global optimization using interval analysis*, Marcel Dekker, N.Y., 1992.

[10] L. Hurwicz, *A criterion for decision-making under uncertainty*, Technical Report 355, Cowles Commission, 1952.

[11] R. B. Kearfott, *Rigorous global search: continuous problems*, Kluwer, Dordrecht, 1996.

[12] R. B. Kearfott and V. Kreinovich (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 1996.

[13] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications,* Prentice Hall, Upper Saddle River, NJ, 1995.

[14] V. Kreinovich, "Maximum entropy and interval computations", *Reliable Computing*, 1996, Vol. 2, No. 1, pp. 63–79.

[15] V. P. Kuznetsov, *Interval statistical models*, Moscow, Radio i Svyaz Publ., 1991 (in Russian).

[16] S. A. Lesh, *An evidential theory approach to judgment-based decision making*, Ph.D. Thesis, Department of Forestry and Environmental Studies, Duke University, Durham, NC, 1986.

[17] D. R. Luce and H. Raiffa, *Games and decisions, Introduction and critical survey*, J. Wiley & Sons, N.Y., 1957, reprinted by Dover, N.Y., 1989.

[18] S. J. Marcus, "Grand illusion: the risks of Cassini", *IEEE Spectrum*, March 1998, pp. 58–65 (see also discussion in April 1998, pp. 6–7).

[19] R. Moore, *Methods and Applications of Interval Analysis*, SIAM, Philadelphia, 1979.

[20] R. B. Myerson, *Game theory. Analysis of conflict*, Harvard University Press, Cambridge, MA, 1991.

[21] H. T. Nguyen, V. Kreinovich, and Q. Zuo, "Interval-valued degrees of belief: applications of interval computations to expert systems and intelligent control", *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems (IJUFKS)*, 1997, Vol. 5, No. 3, pp. 317–358.

[22] H. T. Nguyen and E. A. Walker, "On decision making using belief functions", In: R. R. Yager, J. Kacprzyk, and M. Pedrizzi (Eds.), *Advances in the Dempster-Shafer Theory of Evidence*, Wiley, N.Y., 1994, pp. 311–330.

[23] L. J. Savage, *The foundations of statistics*, Wiley, N.Y., 1954; reprinted by Dover, N.Y., 1972.

[24] J. Schubert, "On nonspecific evidence", *Int. J. Intell. Syst.*, 1993, Vol. 8, No. 6, pp. 711–725.

[25] J. Schubert, *Cluster-based specification techniques in Dempster-Shafer theory for an evidential intelligence analysis of multiple target tracks*, Ph.D. Dissertation, Royal Institute of Technology, Department of Numerical Analysis and Computer Science, Stockholm, Sweden, 1994.

[26] J. Schubert, *Specifying nonspecific evidence*, FOA Report C 20975–2.7, National Defence Research Establishment, Stockholm, 1994.

[27] P. Smets and R. Kennes, "The transferable belief model", *Artificial Intelligence*, 1994, Vol. 66, No. 2, pp. 191–234.

[28] T. M. Strat, "Decision analysis using belief model", *Int. J. Approx. Reasoning*, 1990, Vol. 4, No. 5/6, pp. 391–417.

[29] T. M. Strat, "Decision analysis using belief functions", In: R. R. Yager, J. Kacprzyk, and M. Pedrizzi (Eds.), *Advances in the Dempster-Shafer Theory of Evidence*, Wiley, N.Y., 1994, pp. 275–310.

[30] Y. Viniotis, *Probability and random processes for electrical engineers*, McGraw-Hill, Boston, 1998.

[31] P. Walley, *Statistical reasoning with imprecise probabilities*, Chapman and Hall, N.Y., 1991.

Appendix 1: Proof of the Theorem

1. Let us first fix interval probabilities $\mathbf{p}_1, \ldots, \mathbf{p}_n$, and consider C as a function of n variables $c_1, \ldots, c_n$:

$$F(c_1, \ldots, c_n) = C(\langle \mathbf{p}_1, c_1 \rangle, \ldots, \langle \mathbf{p}_n, c_n \rangle).$$

Property (5) says that this function $F(c_1, \ldots, c_n)$ is *additive*. It is known (see, e.g., [1], Section 4.1) that every continuous additive function has the form

$$F(c_1, \ldots, c_n) = \widetilde{p}_1 \cdot c_1 + \ldots + \widetilde{p}_n \cdot c_n.$$

Thus, for every sequence of n interval probabilities, there exists n real values $\widetilde{p}_1, \ldots, \widetilde{p}_n$ which depend on these interval probabilities and for which

$$C(\langle \mathbf{p}_1, c_1 \rangle, \ldots, \langle \mathbf{p}_n, c_n \rangle) = \widetilde{p}_1 \cdot c_1 + \ldots + \widetilde{p}_n \cdot c_n. \tag{12}$$

Therefore, to describe the function C, it is sufficient to describe the transformation T that maps a sequence of finitely many intervals $\mathbf{p}_j$ into a sequence of exactly as many values $\widetilde{p}_j$:

$$(\mathbf{p}_1, \ldots, \mathbf{p}_n) \rightarrow (\widetilde{p}_1, \ldots, \widetilde{p}_n).$$

2. If we take $c_1 = \ldots = c_n = 1$, then from the property (4), we conclude that

$$1 = \min_j c_j \le C(\langle \mathbf{p}_1, c_1\rangle, \ldots, \langle \mathbf{p}_n, c_n\rangle) =$$

$$\widetilde{p}_1 \cdot c_1 + \ldots + \widetilde{p}_n \cdot c_n = \widetilde{p}_1 + \ldots + \widetilde{p}_n \le \max_j c_j = 1,$$

and thus,

$$\widetilde{p}_1 + \ldots + \widetilde{p}_n = 1.$$

3. So far, we have used the properties (4) and (5). Using the equation (12), we can reformulate all other properties in terms of the transformation T.

The property (1) says that if all intervals are degenerate, then T keeps them intact:

$$([p_1, p_1], \ldots, [p_n, p_n]) \to (p_1, \ldots, p_n). \tag{1'}$$

Similarly, (2) turns into:

$$([p_1, p_1], \mathbf{p}_2) \to (p_1, 1 - p_1). \tag{2'}$$

The property (3) turns into the following rule: If

$$(\mathbf{p}_1, \ldots, \mathbf{p}_{i-1}, \mathbf{p}_i, \mathbf{p}_{i+1}, \ldots, \mathbf{p}_{j-1}, \mathbf{p}_j, \mathbf{p}_{j+1}, \ldots, \mathbf{p}_n) \to$$

$$(\widetilde{p}_1, \ldots, \widetilde{p}_{i-1}, \widetilde{p}_i, \widetilde{p}_{i+1}, \ldots, \widetilde{p}_{j-1}, \widetilde{p}_j, \widetilde{p}_{j+1}, \ldots, \widetilde{p}_n),$$

then

$$(\mathbf{p}_1, \ldots, \mathbf{p}_{i-1}, \mathbf{p}_j, \mathbf{p}_{i+1}, \ldots, \mathbf{p}_{j-1}, \mathbf{p}_i, \mathbf{p}_{j+1}, \ldots, \mathbf{p}_n) \to$$

$$(\widetilde{p}_1, \ldots, \widetilde{p}_{i-1}, \widetilde{p}_j, \widetilde{p}_{i+1}, \ldots, \widetilde{p}_{j-1}, \widetilde{p}_i, \widetilde{p}_{j+1}, \ldots, \widetilde{p}_n). \tag{3'}$$

The property (6) means that if

$$(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n) \to (\widetilde{p}_1, \widetilde{p}_2, \ldots, \widetilde{p}_n),$$

then

$$(\mathbf{p}_1 + \mathbf{p}_2, \ldots, \mathbf{p}_n) \to (\widetilde{p}_1 + \widetilde{p}_2, \ldots, \widetilde{p}_n). \tag{6'}$$

Finally, the condition 7 means that the transformation T is continuous.

4. Let us first make a comment that will be used in the following proof. Due to symmetry (3′), if two of n intervals coincide, i.e., if $\mathbf{p}_i = \mathbf{p}_j$, then the resulting values $\widetilde{p}_i$ and $\widetilde{p}_j$ must be equal too.

5. We want to prove that the transformation T is described by the formula (8) for all intervals $\mathbf{p}_j$. To prove it, let us first start by showing that this is true for intervals $\mathbf{p}_j = [p_j^-, p_j^+]$ with *rational* endpoints.

Since all the endpoints are rational, we can reduce them to a common denominator. Let us denote this common denominator by N; then each of the endpoints p_j^- and p_j^+ has the form m/N for a non-negative integer m. Let us denote the corresponding numerators by m_j^- and m_j^+; then, we have $p_j^- = m_j^-/N$ and $p_j^+ = m_j^+/N$ (where $m_j^- = N \cdot p_j^-$ and $m_j^+ = N \cdot p_j^+$).

Each interval $\mathbf{p}_j = [m_j^-/N, m_j^+/N]$ can be represented as a sum of m_j^- degenerate intervals $[1/N, 1/N]$ and $m_j^+ - m_j^-$ non-degenerate intervals $[0, 1/N]$. Totally, we get $m_1^- + \ldots + m_n^- = N \cdot (p_1^- + \ldots + p_n^-) = N \cdot \Sigma^-$ degenerate intervals $[1/N, 1/N]$ and $N \cdot (\Sigma^+ - \Sigma^-)$ non-degenerate intervals $[0, 1/N]$. So, if we know how the transformation T transforms the resulting "long list" of $N \cdot \Sigma^- + N \cdot (\Sigma^+ - \Sigma^-) = N \cdot \Sigma^+$ intervals, we will be able to use the property (6′) and find the result of applying T to the original set of intervals.

What is the result of applying T to this long list? This long list contains intervals of two types, and intervals of each type are identical. We have already proven in part 4 of this proof that if two intervals from the list are equal, then the corresponding values of $\widetilde{p}_j$ are equal too. Thus:

- the transformation T maps all degenerate intervals $[1/N, 1/N]$ into one and the same value; we will denote this value by α;
- similarly, the transformation T maps all non-degenerate intervals $[0, 1/N]$ into one and the same value; we will denote this value by β.

So, we get the mapping

$$\left(\left[\frac{1}{N}, \frac{1}{N}\right], \ldots, \left[\frac{1}{N}, \frac{1}{N}\right], \left[0, \frac{1}{N}\right], \ldots, \left[0, \frac{1}{N}\right]\right) \rightarrow (\alpha, \ldots, \alpha, \beta, \ldots, \beta). \quad (13)$$

If we apply the property (6′) to this formula, then we can conclude that

$$(\ldots, \mathbf{p}_j, \ldots) =$$

$$\left(\ldots, \left[\frac{1}{N}, \frac{1}{N}\right] + \ldots + \left[\frac{1}{N}, \frac{1}{N}\right] (m_j^- \text{ times}) +\right.$$

$$\left.\left[0, \frac{1}{N}\right] + \ldots + \left[0, \frac{1}{N}\right] (m_j^+ - m_j^- \text{ times}), \ldots\right) \rightarrow$$

$$(\ldots, \alpha + \ldots + \alpha \ (m_j^- \text{ times}) + \beta + \ldots + \beta \ (m_j^+ - m_j^- \text{ times}), \ldots) =$$

$$(\ldots, \widetilde{p}_j, \ldots),$$

where

$$\widetilde{p}_j = m_j^- \cdot \alpha + (m_j^+ - m_j^-) \cdot \beta. \quad (14)$$

So, to find the values $\widetilde{p}_j$, it is sufficient to determine the values of the parameters α and β.

To determine these parameters, we will also use the additivity property (6′). Namely, from (13), we can similarly conclude that

$$\left(\left[\frac{1}{N},\frac{1}{N}\right]+\ldots+\left[\frac{1}{N},\frac{1}{N}\right] (N\cdot\Sigma^- \text{ times})+\right.$$
$$\left.\left[0,\frac{1}{N}\right]+\ldots+\left[0,\frac{1}{N}\right] (N\cdot(\Sigma^+-\Sigma^-) \text{ times})\right)\to$$
$$(\alpha+\ldots+\alpha\ (N\cdot\Sigma^- \text{ times})+\beta+\ldots+\beta\ (N\cdot(\Sigma^+-\Sigma^-) \text{ times}))=$$
$$(N\cdot\Sigma^-\cdot\alpha, N\cdot(\Sigma^+-\Sigma^-)\cdot\beta). \tag{15}$$

The sums of the intervals in the left-hand side of (15) can be explicitly calculated:

$$\left[\frac{1}{N},\frac{1}{N}\right]+\ldots+\left[\frac{1}{N},\frac{1}{N}\right] (N\cdot\Sigma^- \text{ times})=$$
$$\left[\frac{N\cdot\Sigma^-}{N},\frac{N\cdot\Sigma^-}{N}\right]=[\Sigma^-,\Sigma^-],$$

and

$$\left[0,\frac{1}{N}\right]+\ldots+\left[0,\frac{1}{N}\right] (N\cdot(\Sigma^+-\Sigma^-) \text{ times})=$$
$$\left[0,\frac{N\cdot(\Sigma^+-\Sigma^-)}{N}\right]=[0,\Sigma^+-\Sigma^-].$$

Thus, (15) takes the form

$$([\Sigma^-,\Sigma^-],[0,\Sigma^+-\Sigma^-])\to(N\cdot\Sigma^-\cdot\alpha, N\cdot(\Sigma^+-\Sigma^-)\cdot\beta). \tag{16}$$

On the other hand, from (2′), we conclude that

$$([\Sigma^-,\Sigma^-],[0,\Sigma^+-\Sigma^-])\to(\Sigma^-,1-\Sigma^-). \tag{17}$$

Comparing (16) and (17), we conclude that

$$N\cdot\Sigma^-\cdot\alpha=\Sigma^- \tag{18}$$

and

$$N\cdot(\Sigma^+-\Sigma^-)\cdot\beta=1-\Sigma^-. \tag{19}$$

From the equation (18), we conclude that

$$\alpha=\frac{1}{N}. \tag{20}$$

From the equation (19), we conclude that

$$\beta=\frac{1}{N}\cdot\frac{1-\Sigma^-}{\Sigma^+-\Sigma^-}. \tag{21}$$

Substituting the expression for α and β into the formula (14), we conclude that

$$\widetilde{p}_j = \frac{m_j^-}{N} + \frac{m_j^+ - m_j^-}{N} \cdot \frac{1 - \Sigma^-}{\Sigma^+ - \Sigma^-}. \tag{22}$$

By definition of the numbers m_j^-, we conclude that $m_j^-/N = p_j^-$ and that $(m_j^+ - m_j^-)/N = (m_j^+/N) - (m_j^-/N) = p_j^+ - p_j^-$. Therefore, (22) takes the form

$$\widetilde{p}_j = p_j^- + (p_j^+ - p_j^-) \cdot \frac{1 - \Sigma^-}{\Sigma^+ - \Sigma^-}.$$

Grouping together terms proportional to p_j^-, we conclude that

$$\widetilde{p}_j = p_j^- \cdot \left(1 - \frac{1 - \Sigma^-}{\Sigma^+ - \Sigma^-}\right) + p_j^+ \cdot \frac{1 - \Sigma^-}{\Sigma^+ - \Sigma^-}, \tag{23}$$

and finally, subtracting the two fractions in (23), we get the desired result.

6. We have shown that the formula (8) holds for all intervals with rational endpoints. Since the transformation T is continuous (property 7), and since every interval can be represented as a limit of intervals with rational endpoints, we can conclude, by tending to a limit, that this formula is true for *all* intervals. The theorem is proven.

Appendix 2: Derivation of the statistical formula for $\widetilde{p}_j$

According to mathematical statistics (see, e.g., [30], Ch. 5), if we have two Gaussian random variables X and Y, then the conditional mathematical expectation $E(X \mid Y = y)$ is equal to $a \cdot y + b$, where the coefficients a and b are determined from the condition that

$$E\left[(X - aY - b)^2\right] \to \min_{a,b}.$$

Differentiating the optimized function with respect to a and b and equating the resulting derivatives to 0, we conclude that

$$a \cdot E[Y^2] + b \cdot E[Y] = E[X \cdot Y];$$

$$a \cdot E[Y] + b \cdot E[1] = E[X].$$

In our case, $X = p_j$ and $Y = p_1 + \ldots + p_n$. Hence, $E[X] = f_j$, and $E[Y] = f_1 + \ldots + f_n$. In the following text, we will denote this sum by Σ_0.

The value $E[X \cdot Y]$ can be represented as

$$E[X \cdot Y] = E[p_j(p_1 + \ldots + p_{j-1} + p_j + p_{j+1} + \ldots + p_n)] = E[p_j^2] + \sum_{k \neq j} E[p_j \cdot p_k].$$

The first term in this sum is equal to $f_j^2 + \sigma_j^2$. Since for $k \neq j$, p_j and p_k are independent random variables, each other term is equal to $E[p_j] \cdot E[p_k] = f_j \cdot f_k$. Thus,

$$E[X \cdot Y] = f_j^2 + \sigma_j^2 + \sum_{k \neq j} f_j \cdot f_k.$$

Adding the term $f_j^2 = f_j \cdot f_j$ to the sum, we conclude that

$$E[X \cdot Y] = \sigma_j^2 + f_j \cdot \left(\sum_k f_k \right) = \sigma_j^2 + f_j \cdot \Sigma_0.$$

Similarly,

$$E[Y^2] = E\left[\left(\sum_k p_k \right) \cdot \left(\sum_k p_k \right) \right] = \sum_k E[p_k^2] + \sum_k \sum_{l \neq k} E[p_k \cdot p_l] =$$

$$\sum_k f_k^2 + \sum_k \sigma_k^2 + \sum_k \sum_{l \neq k} f_k \cdot f_l.$$

Separating the terms that correspond to $(\sum f_k)^2$, we conclude that

$$E[y^2] = \left(\sum_k f_k \right)^2 + \sum_k \sigma_k^2 = \Sigma_0^2 + \sum_k \sigma_k^2.$$

Thus, the above equations for a and b take the form:

$$a \cdot \left[\Sigma_0^2 + \sum_k \sigma_k^2 \right] + b \cdot \Sigma_0 = f_j \cdot \Sigma_0 + \sigma_j^2;$$

$$a \cdot \Sigma_0 + b = f_j.$$

If we multiply the second equation by Σ_0 and subtract the result from the first equation, we get an equation which contains only one unknown a: $a \cdot \sum \sigma_k^2 = \sigma_j^2$. Therefore,

$$a = \frac{\sigma_j^2}{\sum \sigma_k^2}.$$

Substituting this value a into the second equation, we can now calculate b as

$$b = f_j - a \cdot \Sigma_0 = f_j - \frac{\sigma_j^2}{\sum \sigma_k^2} \cdot \Sigma_0.$$

Thus, the desired conditional expectation is equal to

$$a \cdot y + b = \frac{\sigma_j^2}{\sum \sigma_k^2} \cdot 1 + f_j - \frac{\sigma_j^2}{\sum \sigma_k^2} \cdot \Sigma_0 = f_j + \frac{(1 - \Sigma_0) \cdot \sigma_j^2}{\sum \sigma_k^2}.$$

The formula is proven.