Interval Estimates: How to Make Them More Adequate and How to Use Them In Economic Analysis and Decision Making

Laura Adriana Berrout-Ramos
University of Texas at El Paso

INTERVAL ESTIMATES: HOW TO MAKE THEM MORE ADEQUATE AND HOW TO USE THEM IN ECONOMIC ANALYSIS AND DECISION MAKING

LAURA A. BERROUT RAMOS

Master's Program in Computer Science

APPROVED:

Vladik Kreinovich, Ph.D., Chair

Luc Longpré, Ph.D.

Olga Kosheleva, Ph.D.

Stephen L. Crites, Ph.D.
Dean of the Graduate School

INTERVAL ESTIMATES: HOW TO MAKE THEM MORE ADEQUATE AND HOW TO USE THEM IN ECONOMIC ANALYSIS AND DECISION MAKING

by

LAURA A. BERROUT RAMOS

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Computer Science

THE UNIVERSITY OF TEXAS AT EL PASO

December 2021

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Vladik Kreinovich, for his incredible support, for his help and all his patience, and to Dr. Luc Longpré and Dr. Olga Kosheleva, members of my committee, for their advice and help.

I would also like to thank my husband Salvador Ruiz for his unending support.

Abstract

In many real-life situations, we need to make decisions in situations when we do not have full information about the consequences of different decisions. In particular, instead of the exact values of the relevant quantities, we only know lower and upper bounds on these values – i.e., we know an interval that contains the actual (unknown) value. These interval estimates often come from experts.

This fact naturally leads to the following important questions: How should we make decisions under such interval uncertainty? How to gauge the quality of the resulting decisions? And if this quality is not sufficient – because the original intervals were too wide – how can we improve the interval estimates so as to make better decisions? And if improvements are possible, why not do them from the very beginning, as a pre-processing of expert-provided intervals? In this thesis, we propose answers to these questions in several economically meaningful situations.

We start, in Chapter 1, with a general description of how rational decisions should be made – according to decision theory. To make these decisions, we need to have some information about the corresponding quantities, information that often comes in terms of expert-provided intervals. In Chapter 2, we analyze how these intervals can be improved.

In Chapter 3, we analyze how we can take interval uncertainty into account when gauging the quality of the existing decisions. Finally, in Chapter 4, we analyze how to make new decisions under interval uncertainty.

Table of Contents

Chapter 1

How Economic Decisions Are Made: Preferences and Utilities

This thesis is about making decisions. In this chapter, we provide a general overview of how decisions should be made.

To make adequate decisions – especially economic decisions – we need to know people's preferences. Preferences are a particular case of natural ordering relations that appear in all areas of human activity: causality in space-time physics, preference in decision making, and logical inference in reasoning.

In space-time physics, a 1950 theorem by A. D. Alexandrov proved that causality relation is fundamental: many other features, including numerical characteristics of time and space, can be reconstructed from this relation. In this chapter, we provide a simple proof that, similarly, the preference relation is fundamental in decision making and in logical reasoning.

As an additional result, we also prove that the corresponding logical inference relation is fundamental in reasoning.

The contents of this chapter first appeared in [30].

1.1 Preferences and Utilities: A Brief Reminder

How preferences are described in decision theory: a brief reminder. Let us recall how preferences are described in decision theory; see, e.g., [13, 14, 27, 39, 45, 46, 53].

The usual way to provide a numerical description of preferences is to select two alternatives:

- a very bad alternative A_- which is worse than anything we will actually encounter, and
- a very good alternative A_+ which is better than anything we will actually encounter.

Then, to provide a numerical value to any actual alternative A, we ask the person to compare this alternative with lotteries in which this person:

- gets A_+ with some probability p and
- gets A_- with the remaining probability $1 - p$,

for different values p. We will denote such a lottery by $L(p)$.

- When the probability p is small, the lottery is almost the same as the very bad alternative A_-. So, due to our choice of A_-, the alternative A is better than $L(p)$: $L(p) \leq A$.
- When the probability p is close to 1, the lottery is almost the same as the very good alternative A_+. So, due to our choice of A_+, the alternative A is worse than $L(p)$: $A \leq L(p)$.

As we increase the probability p of the very good outcome A_+, the lottery becomes more and more preferable. So, at some probability p_0, the decision maker switches from $L(p) < A$ to $A < L(p)$. This threshold value p_0 is known as the *utility* of the alternative A. The utility is usually denoted by $u(A)$.

Utility is not uniquely determined. The numerical value of the utility depends on our choice of the alternatives A_- and A_+. One can show that if we select a different pair A'_-, A'_+, then the corresponding numerical value $u'(A)$ is related to the original value $u(A)$ by a linear dependence

$$u'(A) = a \cdot u(A) + b, \tag{1.1}$$

for some values $a > 0$ and b (which do not depend on the alternative A).

How people actually make decisions. In the ideal world, if a person encounters n alternatives with utilities $u_1, \ldots, u_n$, this person should select the alternative with the largest utility – because, by definition of utility, this alternative is equivalent to the lottery in which the probability of winning the big prize A_+ is the highest. In reality, the person may select other alternatives as well. The probability of selecting the i-th is equal to:

$$p_i = \frac{\exp(\alpha \cdot u_i)}{\sum\limits_{j=1}^{n} \exp(\alpha \cdot u_j)}, \tag{1.2}$$

for some value $\alpha > 0$ depending on the person. This formula is known as the *discrete choice model*; see, e.g., [38, 40, 41, 42, 58]. For this formula, D. McFadden received a Nobel Prize.

The formula (1.2) confirms what we have mentioned earlier: that human preference is not a binary relation, it is characterized by the probabilities p_i of selecting different alternatives.

Comment. As shown in [10], the formula (1.2) is not just empirically valid: it can be explained by natural symmetry ideas.

1.2 Preferences Are a Particular Case of Natural Order Relations

Main objectives of science and engineering. One of the main objectives of science and engineering is to help people *select the most beneficial decisions.* To make these decisions,

- we must know people's *preferences*,
- we must have the information about different events – *possible consequences of different decisions*, and
- we must be able to use this information to come up with decisions.

Enter order relations. In each of these three categories, we have natural order relations:

- for preferences, $a \le b$ means that b is preferable to a;
- for events, $a \le b$ means that a can influence b; this relation is known as causality; and
- in reasoning, $a \le b$ means that we can infer b from a; this relation is known as *implication* and is usually denoted by $a \to b$.

Important comment: these relations are often not binary. Sometimes, we are absolutely sure that an alternative b is better than an alternative a, that an event a can influence an event b, that a statement a definitely implies the statement b. In such cases, the corresponding relation is "binary" in the sense that:

- for some pairs (a, b), the relation $a \le b$ is absolutely true, while
- for some other pairs (a, b), the relation $a \le b$ is absolutely *not* true.

However, in many cases, we are not 100% certain:

- we believe that the alternative b is most probably better, but we have some doubts; as a result, people sometimes select a;
- we believe that b can probably be inferred from a, but we are not absolutely sure; for example, in a trial by jury, there is often a reasonable doubt, and, because of this, a suspect goes free in spite of some strong indirect evidence against him.

In the following text, we will take this uncertainty into account.

1.3 Alexandrov-Zeeman Theorem: Reminder

Causality in physics: a brief reminder. To describe the Alexandrov-Zeeman theorem, let us first briefly recall how causality is defined in physics; see, e.g., [12, 57].

In Newton's physics, signals can potentially travel with an arbitrarily large speed. To describe the corresponding causality relation between events, let us denote an event occurring at the spatial location x at time t by $a = (t, x)$. In these notations, Newton's causality relation is as follows: an event $a = (t, x)$ can causally (physically) influence an event $a' = (t', x')$ if and only if $t \le t'$:

$$(t, x) \le (t', x') \Leftrightarrow t \le t'. \tag{1.3}$$

In Special Relativity, the speed of all the signals is limited by the speed of light c. As a result, $a = (t, x) \le a' = (t', x')$ if and only if $t' \ge t$ and in time $t' - t$, the speed needed to traverse the distance $d(x, x')$ does not exceed c, i.e., $\dfrac{d(x, x')}{t' - t} \le c$. The resulting causality relation has the form

$$(t, x) \le (t', x') \Leftrightarrow c \cdot (t' - t) \ge d(x, x'). \tag{1.4}$$

Alexandrov-Zeeman theorem. In 1950, A. D. Alexandrov showed that in Special Relativity, causality implied Lorenz group [1, 2]. To be more precise, he proved that every transformation of the 4-dimensional space-time that preserves the causality relation (1.4) is linear, and is a composition of:

- shifts in space and time,
- spatial rotations,
- Lorentz transformations (describing a transition to a moving reference frame), and
- re-scalings $x \to \lambda \cdot x$ (corresponding to a change of unit for measuring space and time).

This theorem was later generalized by E. Zeeman [61] and is therefore known as the *Alexandrov-Zeeman theorem*; see, e.g., [3, 4, 5, 7, 8, 16, 17, 18, 19, 20, 21, 25, 26, 29, 32, 33, 34, 35, 36, 37, 44, 52, 61].

This theorem showed that causality indeed plays a fundamental role in space-time physics: once we know this relation, we can reconstruct the linear structure of space-time, we can reconstruct (modulo a possible multiplicative constant) the values of proper time and proper space, etc.

1.4 Main Result of This Chapter: Preference Relation Is Fundamental in Decision Making

The following result shows that the preference relation – as described by the probabilities p_i – determines utilities uniquely – modulo a linear transformation (1.1). Thus, for decision making, the corresponding ordering relation is indeed also fundamental.

Proposition 1.1. *If for two sequences $u_1, \ldots, u_n$ and $u'_1, \ldots, u'_n$ and for some values $\alpha > 0$ and $\alpha' > 0$, we have $p_i = p'_i$ for all i, where*

$$p_i \stackrel{\text{def}}{=} \frac{\exp(\alpha \cdot u_i)}{\sum\limits_{j=1}^{n} \exp(\alpha \cdot u_j)} \text{ and } p'_i = \frac{\exp(\alpha' \cdot u'_i)}{\sum\limits_{j=1}^{n} \exp(\alpha' \cdot u'_j)}, \tag{1.5}$$

then there exists values $a > 0$ and b for which

$$u'_i = a \cdot u_i + b \tag{1.6}$$

for all i.

Proof. For each $i \neq 1$, if we divide the equality $p_i = p'_i$ by the equality $p_1 = p'_1$, we get

$$\frac{p_i}{p_1} = \frac{p'_i}{p'_1}. \tag{1.7}$$

Substituting the expressions (1.6) instead of p_i, p_1, p'_i, and p'_1, we conclude that

$$\frac{\exp(\alpha \cdot u_i)}{\exp(\alpha \cdot u_1)} = \frac{\exp(\alpha' \cdot u'_i)}{\exp(\alpha' \cdot u'_1)}, \tag{1.8}$$

i.e., equivalently,

$$\exp(\alpha \cdot (u_i - u_1)) = \exp(\alpha' \cdot (u'_i - u'_1)). \tag{1.9}$$

By taking logarithm of both sides and dividing both sides by α', we conclude that

$$u_i' - u_1' = a \cdot (u_i - u_1), \tag{1.10}$$

where we denoted $a \stackrel{\text{def}}{=} \alpha/\alpha'$. Hence,

$$u_i' = a \cdot u_i + (u_1' - a \cdot u_1), \tag{1.11}$$

i.e., the desired formula (1.6) for $b \stackrel{\text{def}}{=} u_1' - a \cdot u_1$.

We are almost done: we have proved the formula (1.6) for all $i \neq 1$. However, one can easily show that for $i = 1$, the right-hand side of the formula (1.11) is also equal to its left-hand side u_1'. So, the proposition is proven.

1.5 Auxiliary Result: Logical Inference Relation Is Fundamental in Reasoning

How can we describe degree of inference in logical reasoning: a brief reminder. A direction of logic that describes degrees of certainty – i.e., that analyzes statements which are imprecise ("fuzzy") is known as *fuzzy logic*; see, e.g., [6, 24, 43, 48, 51, 60].

The original – and simplest – idea is to take into account that in a computer:

- "false" is represented as 0, and
- "true" is represented as 1.

Thus, it is reasonable to describe intermediate degrees by numbers from the interval $(0, 1)$.

There are many different *implication operations*, i.e., functions $f_{\to}(a, b)$ that transform the expert's degree of confidence a and b in some statements A and B into the estimated degree of confidence in the implication $A \to B$. Such operation should satisfy several reasonable properties. For example:

- If we accept a statement A in which we are not 100% sure, then we can get conclusions that we could not get before. In this case, the statement $A \to B$ can have larger degree of confidence than the statement B itself.

- However, if, as A, we take an absolutely true statement T, with degree $a = 1$, then adding this statement will not change what we can conclude. Thus, our degree of confidence $f_\to(1, b)$ in the implication $T \to B$ should be exactly the same as our degree of confidence b in the original statement B: $f_\to(1, b) = b$.

The values $f_\to(a, b)$ corresponding to different a and b represent the inference ordering.

Main result of this section. The following result shows that the inference relation – as described by the degrees $f_\to(a, b)$ – uniquely determines the original degrees a and b.

Thus, for logical reasoning, the corresponding ordering relation is indeed also fundamental.

Definition 1.1. *By an* implication operation*, we mean a function*

$$f_\to : [0, 1] \times [0, 1] \to [0, 1] \tag{1.12}$$

for which $f_\to(1, b) = b$ for all b.

Proposition 1.2. *Suppose that we have two sequences $a_1, \ldots, a_n$ and $a'_1, \ldots, a'_n$ such that for some i_0, we have $a_{i_0} = b_{i_0} = 1$. Suppose also that we have two implication operations $f_\to$ and $f'_\to$ for which, for all i and j, we have*

$$f_\to(a_i, a_j) = f'_\to(a'_i, a'_j). \tag{1.13}$$

Then, for all j, we have $a_j = a'_j$.

Proof. For each j from 1 to n, substituting $i = i_0$ into the formula (1.13) and taking into account that $a_{i_0} = a'_{i_0} = 1$, we conclude that $f_\to(1, a_j) = f'_\to(1, a'_j)$. Now, by definition of an implication operation, we conclude that indeed $a_j = a'_j$.

The proposition is proven.

Chapter 2

Interval Estimates and How to Make Them More Adequate

In many real-life situations, we need to make decisions in situations when we do have full information about the consequences of different decisions. In particular, instead of the exact values of the relevant quantities, we only know lower and upper bounds on these values – i.e., we know an interval that contains the actual (unknown) value. These interval estimates often come from experts.

Sometimes the intervals are unnecessarily narrow – e.g., if an expert is overconfident. Sometimes, the intervals are unnecessarily wide – if the expert is too cautious. In such situations, we need to correct these intervals, by making them, correspondingly, wider or narrower. Empirical studies show that people use specific formulas for such corrections. In this chapter, we provide a theoretical explanation for these empirical formulas.

Comment. Results from this chapter first appeared in [9].

2.1 Formulation of the Problem

Expert estimates are needed. In economics – as in many other application areas – we often rely on expert estimates.

Expert estimates are approximate. Of course, expert estimates are approximate. So, to be able to use them effectively, we need to know how accurate they are.

How do we usually gauge the accuracy of expert estimates. There are two main idea on how to gauge the accuracy of an expert estimate:

- the first is to ask the expert him/herself to gauge this accuracy, namely, to provide an *interval* of possible values instead of a single value;

- the second is to ask one or more other experts, and to consider the range formed by their values as the reasonable range of possible values for the corresponding quantity; for example, of one experts predicts the value 15, and the two others predict 10 and 20, then we take the interval $[10, 20]$.

Need for correction. When an expert provides an interval of possible values:

- sometimes, the expert is too confident, and the interval provided by the expert is too narrow; in this case, a natural idea is to widen it;

- sometimes, the expert is too cautious, and the interval provided by the expert is too wide; in this case, a natural idea is to make it narrower.

Similarly, when we have an interval range formed by numerical estimates made by different experts:

- sometimes, the experts' estimates are too close to each other, so the resulting interval is too narrow; in this case, a natural idea is to widen it;

- sometimes, the experts' estimates are too far away from each other; for example, one expert's predictions are too optimistic, and another expert's predictions are too pessimistic; in this case, the resulting interval is too wide, so a natural idea is to make it narrower.

How do people correct the corresponding intervals? Empirical data shows the corrected version $[A, B]$ of the original interval $[a, b]$ usually follows the formula

$$[A, B] = \left[a \cdot \frac{1+\alpha}{2} + b \cdot \frac{1-\alpha}{2}, a \cdot \frac{1-\alpha}{2} + b \cdot \frac{1+\alpha}{2}\right] \tag{2.1}$$

for some value $\alpha > 0$;

- the values $\alpha < 1$ correspond to shrinking, while

- the values $\alpha > 1$ correspond to stretching of the interval;

see, e.g., [56] and references therein.

Historical comment. The formula (2.1) was first proposed in [15] for expert estimates of probabilities. In [55], this formula was extended to general (not necessarily probabilistic) expert estimates.

Remaining problem. Why do people use this particular formula to correct interval-valued expert estimates?

In this chapter, we provide a possible explanation for this formula.

2.2 Analysis of the Problem

What we look for. We are looking for a formula that describes the corrected interval $[A, B]$ – i.e., that describes both endpoints A and B of this interval – in terms of the inputs interval $[a, b]$ – i.e., in terms of its endpoints a and b. In other words, we need to find algorithms $A(a, b)$ and $B(a, b)$ that, given the expert-provided values a and b, produced the corrected values A and B.

Let us analyze what are the natural properties of these algorithms.

The correction formula should not depend on the monetary unit. The same financial predictions can be described in different monetary units. For example, predictions related to Mexican economy can be made in Mexican pesos or in US dollars.

We are looking for general correction formulas, formulas that would be applicable to all possible interval-valued expert estimates. Suppose that we first applied this formula to the interval $[a, b]$ described in one monetary units. Then, in these units, the corrected interval takes the form

$$[A(a, b), B(a, b)]. \tag{2.2}$$

Another possibility is:

- to first translate into a different monetary unit,
- to make a correction there, and then
- to translate the result back into the original monetary unit.

If we select a different monetary unit which is λ times smaller than the original one, then all numerical values multiply by λ. In particular, in the new units, the original interval $[a, b]$ will take the form $[\lambda \cdot a, \lambda \cdot b]$. If we apply the same correction algorithm to this interval, we get – in the new units – the following corrected interval:

$$[A(\lambda \cdot a, \lambda \cdot b), B(\lambda \cdot a, \lambda \cdot b)]. \tag{2.3}$$

To describe the corrected interval (2.3) in the original units, we need to divide both its endpoints by λ. As a result, we get the corrected interval expressed in the original units:

$$\left[\frac{1}{\lambda} \cdot A(\lambda \cdot a, \lambda \cdot b), \frac{1}{\lambda} \cdot B(\lambda \cdot a, \lambda \cdot b)\right]. \tag{2.4}$$

As we have mentioned, it is reasonable to require that the corrected interval should be the same whether we use the original monetary units or different units, i.e., that we should have

$$[A(a, b), B(a, b)] = \left[\frac{1}{\lambda} \cdot A(\lambda \cdot a, \lambda \cdot b), \frac{1}{\lambda} \cdot B(\lambda \cdot a, \lambda \cdot b)\right]. \tag{2.5}$$

The two intervals are equal if their left endpoints are equal to each other and their right endpoints are equal to each other. So, the equality (2.5) means that:

$$A(a, b) = \frac{1}{\lambda} \cdot A(\lambda \cdot a, \lambda \cdot b) \tag{2.6}$$

and

$$B(a, b) = \frac{1}{\lambda} \cdot B(\lambda \cdot a, \lambda \cdot b). \tag{2.7}$$

Terminological comment. The properties (2.6)–(2.7) describing the fact that the transformation

$$[a, b] \mapsto [A(a, b), B(a, b)]$$

does not change if we change the measuring units (i.e., *re-scale* all the numerical values) is known as *scale-invariance*.

Shift-invariance. Suppose that the expected company's income consists of:

- the fixed amount f – e.g., determined by the current contracts – and
- some additional amount x that will depend on the relation between supply and demand.

Suppose that the expert predicts this additional amount to be somewhere in the interval $[a, b]$. This means that the overall company's income is predicted to be between $f + a$ and $f + b$, i.e., somewhere in the interval $[f + a, f + b]$.

If we believe that the expert estimate needs corrections, then we have two possible ways to perform this correction:

- we can apply the correction to the original interval $[a, b]$, resulting in the corrected interval estimate $[A(a, b), B(a, b)]$ for the additional income; in this case, the interval estimate for the overall income will be
$$[f + A(a, b), f + B(a, b)]; \tag{2.8}$$
- alternatively, we can apply the correction to the interval $[f + a, f + b]$ describing the overall income; in this case, the resulting corrected interval for the overall income will have the form
$$[A(f + a, f + b), B(f + a, f + b)]. \tag{2.9}$$

It is reasonable to require that the two methods should lead to the exact same interval estimate for the overall income:

$$[f + A(a, b), f + B(a, b)] = [A(f + a, f + b), B(f + a, f + b)], \tag{2.10}$$

i.e., equivalently,

$$f + A(a, b) = A(f + a, f + b) \tag{2.11}$$

and

$$f + B(a,b) = B(f+a, f+b). \tag{2.12}$$

Terminological comment. The properties (2.11)–(2.12) describing the fact that the transformation

$$[a,b] \mapsto [A(a,b), B(a,b)]$$

does not change if we shift all the inputs by the same amount f is known as *shift-invariance.*

Sign invariance. One of the possible expert predictions is, e.g., how much bank B_1 will owe a bank B_2 at a certain future date. This amount can be positive – meaning that the bank B_1 will owe some money to the bank B_2. This amount can also be negative – meaning that, according to the expert, at the given future date, the bank B_2 will owe money to the bank B_1.

Suppose that the expert estimates this amount by an interval $[a,b]$. This means that if we ask the same expert a different question: how much money will the bank B_2 owe to the bank B_1 – this expert will provide the interval $[-b,-a]$, i.e., the set of all the values $-x$ when $x \in [a,b]$.

In this case, we also have two possible ways to perform a correction:

- we can apply the correction to the original interval $[a,b]$, resulting in the corrected interval estimate
$$[A(a,b), B(a,b)]; \tag{2.13}$$
- alternatively, we can apply the correction to the interval $[-b,-a]$ describing how much the bank B_2 will owe to the bank B_1, and get the corrected interval $[A(-b,-a), B(-b,-a)]$ for this amount; by changing the sign, we get an interval estimate of how much the bank B_1 will owe to the bank B_2:
$$[-B(-b,-a), -A(-b,-a)]. \tag{2.14}$$

It is reasonable to require that the two methods should lead to the exact same interval estimate for the overall amount:

$$[A(a,b), B(a,b)] = [-B(-b,-a), -A(-b,-a)], \tag{2.15}$$

i.e., equivalently,

$$A(a,b) = -B(-b,-a) \tag{2.16}$$

and

$$B(a,b) = -A(-b,-a). \tag{2.17}$$

Terminological comment. We will call the properties (2.16)–(2.17) describing the fact that the transformation

$$[a,b] \mapsto [A(a,b), B(a,b)]$$

does not change if change all the signs, *sign-invariance.*

2.3 Definitions and the Main Result

Definition 2.1. *We say that a mapping* $[a,b] \mapsto [A(a,b), B(a,b)]$ *is:*

- scale-invariant *if it satisfies the properties* (2.6) *and* (2.7) *for all* $a < b$ *and for all* $s\lambda > 0$*;*

- scale-invariant *if it satisfies the properties* (2.11) *and* (2.12) *for all* $a < b$ *and* f*; and*

- sign-invariant *if it satisfies the properties* (2.16) *and* (2.17) *for all* $a < b$*.*

Comment. One can easily check that the transformation (2.1) is scale-, shift-, and sign-invariant. It turns out that every scale-, shift-, and sign-invariant transformation has the form (2.1). This explains why people use such transformations to correct interval-valued expert estimates.

Proposition 2.1. *Every scale-, shift-, and sign-invariant transformation has the form* (2.1).

Proof. Indeed, due to shift-invariance for $f = a$, we have

$$B(a, b) = a + B(0, b - a). \tag{2.18}$$

Due to scale-invariance for $\lambda = b - a$, we have

$$B(0, 1) = \frac{1}{b - a} \cdot B(0, b - a),$$

hence

$$B(0, b - a) = (b - a) \cdot B(0, 1). \tag{2.19}$$

Let us denote

$$\alpha \stackrel{\text{def}}{=} 2B(0, 1) - 1,$$

then

$$B(0, 1) = \frac{1 + \alpha}{2},$$

and the formula (2.19) takes the form

$$B(0, b - a) = (b - a) \cdot \frac{1 + \alpha}{2} = b \cdot \frac{1 + \alpha}{2} - a \cdot \frac{1 + \alpha}{2}. \tag{2.20}$$

Substituting the expression (2.20) into the formula (2.18), we get

$$B(a, b) = a + b \cdot \frac{1 + \alpha}{2} - a \cdot \frac{1 + \alpha}{2} = a \cdot \frac{1 - \alpha}{2} + b \cdot \frac{1 + \alpha}{2}.$$

This is exactly the expression for $B(a, b)$ corresponding to the formula (2.1).

Now, by using sign-invariance, namely, the formula (2.16), we conclude that

$$A(a, b) = -B(-b, -a) = -\left((-b) \cdot \frac{1 - \alpha}{2} + (-a) \cdot \frac{1 + \alpha}{2}\right) =$$

$$a \cdot \frac{1 + \alpha}{2} + b \cdot \frac{1 - \alpha}{2}.$$

This is also exactly the expression for $A(a, b)$ corresponding to the formula (2.1). Thus, the proposition is proven.

Comment. In addition to empirically confirming the formula (2.1), the paper [56] also mentions an additional empirical fact that it finds difficult to explain: that if two experts provide interval estimates, people put trust into pairs of intervals that have comparable width.

From our viewpoint, this is easy to explain: if two experts, based on the same data, provide completely different estimates of how accurately we can make a prediction based on this data, then we do not trust any of these experts. This is one of the cases when the two supposed experts are inconsistent. A similar phenomenon happens when the two experts provide drastically different numerical estimates: in this case, we do not trust either of them.

Chapter 3

Economic Analysis under Interval Uncertainty: How Effective Are We?

In order to make new decisions, we need to gauge the quality of the existing decisions. In particular, in many economic applications, it is desirable to estimate how effective is a given country, a given plant, a given farm. In some cases, such an estimate is easy: for example, if we have a similar farm which is much more productive, then it is clear than the original farm is not effective. However, in many situations, such a direct comparison is not possible: for example, most countries are different, so while their productivity is different, the difference in productivity may be due to difference in climate, etc., and not necessarily caused by insufficient effectiveness.

To estimate effectiveness in such situations, special Stochastic Frontier techniques have been invented. The problem with the currently use Stochastic Frontier techniques is that these techniques are based on reasonably arbitrary assumptions about the probability distribution of effectiveness – assumptions which are motivated mostly by computational efficiency and which do not have any convincing economic motivations. Because of this arbitrariness, the conclusions of Stochastic Frontier analysis are often not convincing to users. To make these conclusions more convincing, we propose to use different families of distributions – families which *are* economically motivated. Interestingly, while for the new families, the corresponding computational complexity may increase slightly, the corresponding estimation algorithms are still very efficient – especially since they are based on such actively-used-in-economics techniques as least squares and linear programming.

Comment. Results from this chapter first appeared in [50].

3.1 Current Stochastic Frontier Techniques and Their Limitations

Main problem: how effective are we? The success of a country's economy depends on its geographical location, on its natural resources, on the skill of its people, on the amount of available capital – in short, on many characteristics. But it also depends on how efficient are the government policies that intend to boost the economic growth. There are many examples when a change in economic policies provided a boost or, vice versa, a slowdown.

Similarly, a success of a company depends on many objective characteristics – but also on the effectiveness of its management. The success of a farmer depends on what crops he/she plants, what techniques he/she uses, etc. – but also on the farmer's effectiveness.

From this viewpoint, it is important to know how effective is a country, a company, a farm – if it has reached its effectiveness limit, then the only way to further growth is by improving the corresponding objective characteristics – borrowing money, hiring more people, etc. On the other hand, if insufficient effectiveness is the problem, then we can boost productivity simply by making production more effective, often without the need to make large investments. So how do we know how effective we are?

Sometimes, this is easy, but not always. In some cases, it is easy to see that we are not as effective as we would like to be: if there is another country, another company, another farm whose objective characteristics are similar but which is much more productive, then clearly our effectiveness can be improved. However, in many other cases, such a direct comparison is not possible: all other countries, companies, and farm have somewhat different objective characteristics.

For example, a farm in a less fertile area will be usually less productive, but is this "less" due to a less fertile soil or also to a less effective management?

How can we describe this: ideal case. To understand how this question is answered now, let us first consider the ideal case of absolutely effective management.

There are certain quantities $x_1, \ldots, x_n$ that affect the productivity. For example, for a farm, it is important to take into account the climate (e.g., how many warm days per year), quality of soil, how many workers the farm has, etc. In the ideal case, when all the resources are used effectively, we have some ideal productivity y – that depends on the values of all these quantities

$$y = f(x_1, \ldots, x_n), \tag{3.1}$$

for some function $f(x_1, \ldots, x_n)$.

Of course, even when all the resources are used effectively, there is a large number of factors on which the productivity depends. We can take many major factors into account, but there are so many different factors that it is not possible to take all of them into account. For example, for a farm, an unnecessary dry season can decrease the productivity – and, vice versa, an unusually warm season can increase it. As a result, for the same values of the major quantities $x_1, \ldots, x_n$, in different situations, we may have different productivity y. The actual productivity differs from its ideal value $f(x_1, \ldots, x_n)$ – it can be slightly larger, it can be slightly smaller. In other words, the actual productivity is equal to

$$y = r \cdot f(x_1, \ldots, x_n), \tag{3.2}$$

where the factor r is *random* – in the sense that we cannot predict it.

Realistic case. In reality, management is rarely 100% effective. As a result, for the same resources, the productivity is smaller than its ideal value $r \cdot f(x_1, \ldots, x_n)$ by some factor $f \leq 1$ that describes the effectiveness level:

$$y = f \cdot r \cdot f(x_1, \ldots, x_n). \tag{3.3}$$

Towards formulating the problem in precise terms. In several situations $k = 1, \ldots, K$, we observe the values $x_{k,1}, \ldots, x_{k.n}$ of the corresponding quantities, and the value y_k of the resulting productivity. According to the formula (3.3), for each situation k, we have

$$y_k = f_k \cdot r_k \cdot f(x_{k,1}, \ldots, x_{k,n}). \tag{3.4}$$

Our goal is to estimate the values of the factors f_k.

How we can formulate and solve this problem: general discussion. To specify the dependence described by the formula (3.4), we need to know the function $f(x_1, \ldots, x_n)$, and we need to have some information about the factors r and f. All this has to be determined from data.

At any given moment of time, we only have finitely many data points, so we can only determine the values of finitely many parameters. Thus, instead of general functions $f(x_1, \ldots, x_n)$, we only consider functions from an appropriate finite-parametric family

$$f(x_1, \ldots, x_n, c_1, \ldots, c_m) \tag{3.5}$$

depending on parameters $c_1, \ldots, c_m$.

What about the factors r_k and f_k? Since the factors r_k cannot be predicted, it is reasonable to consider them random, with some appropriate distribution. This distribution also needs to be determined based on the empirical data. Thus, similarly to the case of selecting a function, it makes sense to restrict ourselves to some finite-parametric family of distributions, with parameters $a_1, \ldots, a_\ell$.

The factors f_k also cannot be predicted based on the known values of the quantities $x_1, \ldots, x_n$, so it also makes sense to consider them random variables. To describe their distribution, we can select some other finite-parametric family of distributions, with parameters $b_1, \ldots, b_d$. When selecting this family, we need to make sure that we always have $f_k \le 1$.

Now, we have the following exact formulation of the problem:

- we know the values $x_{k,1}, \ldots, x_{k,n}$ and y_k corresponding to different situations k;
- we know the family $f(x_1, \ldots, x_n, c_1, \ldots, c_m)$ of possible functions, and we know two

families of probability distributions characterized, correspondingly, by the parameters $a_1, \ldots, a_\ell$ and $b_1, \ldots, b_d$.

First thing we do is, based on this information, we find the values of all the parameters c_i, a_i, and b_i that leads to the best fit in the formula (3.4). A natural way to find the values of all these parameters is to use the maximum likelihood method, i.e., to find the values of the parameters for which the equality (3.4) is the most probable.

Then, we use the determined values of all these parameters to estimate, for each situation k, the desired effectiveness f_k as

$$f_k = \frac{y_k}{r \cdot f(x_1, \ldots, x_n, c_1, \ldots, c_m)}, \tag{3.6}$$

where r is a random variable distributed according to the distribution with the estimates values of the parameters b_i.

This procedure of using statistical methods to analyze how close we are to the productivity frontier – corresponding to maximal effectiveness – is known as the *stochastic frontier analysis*; see, e.g., [31].

How this problem is usually simplified. The general problem of fitting the formula (3.4) is highly nonlinear and thus, difficult to solve. To make this problem easier to solve, it is usually simplified – in the following three steps.

The need for the first simplification step comes from the fact that already the formula (3.4) itself is nonlinear: it contains the product of three different terms. From the computational viewpoints, a sum is much easier to compute than a product – since the usual computation of a product of two numbers means, in effect, adding several values. In general, it is easier to deal with sums than with the products. It is well known that we can reduce products to sums: the use of logarithms reduces multiplication to addition, since the logarithm of the product is equal to the sum of the logarithms. This is not some unexpected magic property: this property is exactly why logarithms were invented in the first place – to make computing products easier. So, we take the logarithms $Y_k \stackrel{\text{def}}{=} \ln(y_k)$,

$R_k \stackrel{\rm def}{=} \ln(r_k)$, $F_k \stackrel{\rm def}{=} \ln(f_k)$, and $F(x_1,\ldots,x_n) \stackrel{\rm def}{=} \ln(f(x_1,\ldots,x_n))$, for which the formula (3.4) takes the simplified form

$$Y_k = F_k + R_k + F(x_{k,1},\ldots,x_{k,n}) =$$

$$F_k + R_k + F(x_{k,1},\ldots,x_{k,n},c_1,\ldots,c_m). \tag{3.7}$$

The need for the second simplification step comes from the fact that the dependence of the function F on the parameters c_i may be non-linear. It is therefore much more convenient to consider the case when the dependence of the function F on the parameters c_i is linear. This does not limit possible non-linearity of the resulting functions – e.g., every continuous function (and economic dependencies are usually continuous) can be approximated by a polynomial, and every polynomial can be represented as a linear combination of monomials. Thus, we usually consider families of the type

$$F(x_1,\ldots,x_n) = G_0(x_1,\ldots,x_n) +$$

$$\sum_{i=1}^m c_i \cdot G_i(x_1,\ldots,x_n), \tag{3.8}$$

for which the formula (3.7) takes a simplified form

$$Y_k = F_k + R_k + G_0(x_{k,1},\ldots,x_{k,n}) +$$

$$\sum_{i=1}^m c_i \cdot G_i(x_{k,1},\ldots,x_{k,n}), \tag{3.9}$$

i.e.,

$$Y_k = F_k + R_k + G_{k,0} + \sum_{i=1}^m c_i \cdot G_{k,i}, \tag{3.9a}$$

where we denoted

$$G_{k,i} \stackrel{\rm def}{=} G_i(x_{k,1},\ldots,x_{k,n}). \tag{3.9b}$$

The need for the third simplification step comes from the fact that, as we have mentioned earlier, the more quantities x_i we take into account, the more accurate is the description $f(x_1,\ldots,x_n)$ of productivity corresponding to maximal effectiveness – and thus, the smaller the role of the random factor r_k that take all other quantities into account. Thus, if we take sufficiently many quantities x_i into account, we have $r_k \approx 1$ and thus, $R_k \approx 0$. For $R_k = 0$, the formula (3.9a) has an even simpler form

$$Y_k = F_k + G_{k,0} + \sum_{i=1}^{m} c_i \cdot G_{k,i}. \tag{3.10}$$

In this case, once we found the corresponding values c_i, we can then determine the desired values F_k as

$$F_k = Y_k - G_{k,0} - \sum_{i=1}^{m} c_i \cdot G_{k,i} \tag{3.11}$$

and

$$f_k = \exp(F_k). \tag{3.12}$$

How do we describe the distribution of f_k. In most applications, there is not enough data to determine this distribution based on the empirical data – and the problem is that different distributions lead, in general, to different effectiveness estimates.

In practice, usually, the distribution of F_k is selected either as exponential or as half-Gaussian – i.e., Gaussian with mean 0 limited to negative values of F_k. Why? The only reason is that this leads to efficient algorithms.

The problem with this selection is that it lacks good *economic* motivations – and thus, the results of using these distributions are not very convincing – especially since these two distributions lead, in general, to different effectiveness results.

What we do in this chapter. In this chapter, we provide two alternative distributions, distributions which do have economic motivations: Gaussian and uniform. Good news is that replacing distributions selected solely for their computational efficiency with these economically motivated distributions does not make the problem too complex: for both distributions, we provide reasonably efficient estimation algorithms.

What is known and what is new. In contrast to practice, where mostly half-normal and exponential distributions are used, in research papers, many different distributions have been proposed. In particular, uniform distributions were considered, among others, in [49]. However, in all these cases, the motivations for considering these distributions were mostly computational or mathematical – to make the class of distributions as general as possible.

Our novelty is that we provide an economy-based motivations for the use of the selected distributions, and that we explain how, when using these distributions, we can estimate effectiveness – by using only algorithms which are actively used in economic analysis, such as least squares and linear programming.

3.2 First Alternative: Normal Distributions

Main idea. What causes insufficient effectiveness? It is very rare that there is only one reason for decreased productivity. In most real-life situations, there are many different reasons leading to the decrease in productivity: from not-very-effective upper level management to not-very-effective middle level management all the way to not well trained (thus very effective) workers (not to mention a not very effective rewards system).

Let us describe this idea in precise terms. For each situation k, each of the many above-mentioned reasons decreases productivity by a corresponding factor: the first factor decrease productivity by some factor $f_{k,1}$, the second reason decreases productivity by a factor $f_{k,2}$, etc., all the way to the last reason that decreases the productivity by the factor $f_{k,N}$. If the ideal productivity is y_0, then the first factor decreases it to $f_{k,1} \cdot y_0$, the second decrease the resulting productivity even further, to the value

$$f_{k,2} \cdot (f_{k,1} \cdot y_0) = (f_{k,1} \cdot f_{k,2}) \cdot y_0, \tag{3.13}$$

and similarly, the joint effect f_k of all the reasons can be described as the product of the corresponding factors:

$$f_k = f_{k,1} \cdot f_{k,2} \cdot \ldots \cdot f_{k,N}. \tag{3.14}$$

How can this help? At first glance, the above idea only made the problem more complicated and difficult to solve: instead of a single variable f_k whose distribution we do not know, we now have a large number of variables $f_{k,i}$ with unknown distributions. How can this help?

Actually, it does help. Let us see how it helps. The above formulas use the logarithm $F_k = \ln(f_k)$ of the factor f_k. If we take the logarithm of both sides of the formula (3.14), we conclude that

$$F_k = F_{k,1} + F_{k,2} + \ldots + F_{k,N}, \tag{3.15}$$

where we denoted $F_{k,i} \stackrel{\text{def}}{=} \ln(f_{k,i})$. Now, the desired random variable F_k is equal to the sum of a large number of random variables, each of which is relatively small and all of which are reasonably independent. This is exactly a situation to which we can apply the Central Limit Theorem (see, e.g., [54]), according to which, for large N, the distribution of the sum of many small independent random variables is close to Gaussian.

Resulting conclusion. Thus, with reasonable accuracy, we can conclude that the distribution of F_k is Gaussian, with some mean μ and standard deviation σ.

How to estimate effectiveness for this distribution: towards an algorithm. To uniquely determine the model, we need to find the parameters c_i of the productivity function and the parameters μ and σ of the corresponding normal distribution for f. We need to find the values c_i, μ, and σ, it is reasonable to use the maximum likelihood approach. For the normal distribution, the maximum likelihood method reduced to the usual Least Squares. So, we find the values c_i and μ by minimizing the expression

$$\sum_{k=1}^{K} \left(Y_k - \mu - G_{k,0} - \sum_{i=1}^{m} c_i \cdot G_{k,i} \right)^2 . \tag{3.16}$$

Once we have found the values c_i, we can then estimate F_k and f_k by using the formulas (3.11) and (3.12).

Comment. Measurements are noisy, models are approximate – as a result, instead of values F_k close to 0 – which corresponds to $f_k = 1$ – we can have values F_k slightly larger than

0, i.e., to $f_k > 1$, which makes no sense: you cannot be more productive that the most effective plant. So, if we get such a result, we will simply replace it with 0. In other words, instead of using the formula (3.11), we should use the correspondingly modified formula

$$F_k = \min\left(Y_k - G_{k,0} - \sum_{i=1}^{m} c_i \cdot G_{k,i}, 0\right). \tag{3.17}$$

What if we take noise into account. In the above discussions, we ignored the random "noise" r_k. As we explained, this ignoring makes sense, since the noise is usually relatively small. To get a more accurate estimate, a natural idea is to take this noise into account.

In this case, arguments similar to the above arguments about the distribution of F_k enable us to conclude that it is reasonable to assume that the distribution of the values R_k is also (close to) Gaussian. The main difference between the distributions for F_k and R_k is that, while the value F_k is always negative – and thus, its mean μ is negative, noise can lead both to increase and decrease in productivity – so R_k can be equally probably positive and negative and its mean is 0.

Interestingly, in this case, the same formula (3.16) is still applicable – and thus, the same least-squares-based algorithm that we described earlier works for this more accurate model as well.

3.3 Second Alternative: Uniform Distributions

Idea. We do not know anything about the values F_k. All we know is that these values should be smaller than or equal to 0, and maybe that they should be larger than or equal to some negative value F_0 – which means that companies, farms, etc. cannot be too ineffective. A farm can produce 2, even maybe 10 times less crops than under effective measurement, but not 100 or 1000 times less.

In other words, all the know is that the values F_k are all located on an interval $[F_0, 0]$. We have no reasons to believe that some values from this interval are more probable and some are less probable. Thus, it makes sense to conclude that all the values from this interval are equally probable, i.e., that we have a uniform distribution on this interval. A similar argument is often used in statistics and data processing. It is known as *Laplace Indeterminacy Principle*, and it is the basis of a highly successful Maximum Entropy approach; see, e.g., [23].

Resulting conclusion. Thus, with reasonable accuracy, we can conclude that the values F_k are uniformly distributed on the interval $[F_0, 0]$, for some value $F_0 < 0$.

How to estimate effectiveness for this distribution: towards an algorithm. To uniquely determine the model, we need to find the parameters c_i of the productivity function and the parameter F_0 of the corresponding uniform distribution for f. To find the values c_i and F_0, it is reasonable to use the maximum likelihood approach. For the uniform distribution, the probability density is equal to $1/|F_0|$. The overall probability density is equal to the product of K such terms, i.e., to $1/|F_0|^K$. The largest value of this likelihood corresponds to the smallest possible value of $|F_0|$ – i.e., equivalently, to the largest possible value of $F_0 = -|F_0|$.

The only restriction on F_0 is that all the differences F_k – as determined by the formula (3.11) – should be between F_0 and 0. Thus, we arrive at the same formulation of the problem: to find the values c_i and F_0, we need to find the largest possible values of F_0 under the following constraints:

$$\begin{gathered} F_0 \le Y_1 - G_{1,0} - \sum_{i=1}^{m} c_i \cdot G_{1,i} \le 0; \\ \ldots \\ F_0 \le Y_K - G_{K,0} - \sum_{i=1}^{m} c_i \cdot G_{K,i} \le 0. \end{gathered} \tag{3.18}$$

In terms of the unknowns c_i and F_0, we thus get a problem of maximizing a linear function under linear constraints. Such optimization problems are known as problems of *linear programming*; see, e.g., [59].

There exist efficient algorithms for solving linear programming problems. By applying these algorithms, we can efficiently find the coefficients c_i. Once we have found these coefficients, we can then find the desired values F_k by using the formulas (3.11) and (3.12).

What if we take noise into account. Similarly to the previous section, we can ask a question what happens if we take noise r_k into account. Similarly to our arguments about F_k, we can argue that the only thing we know about the value R_k – which can be both positive and negative – is that its absolute value cannot exceed some value R_0. So, we expect only values R_k from the interval $[-R_0, R_0]$. Similarly to the above, we can then conclude that it is reasonable to assume that the actual value R_k is uniformly distributed on this interval. For this distribution, the probability density is equal to $1/(2R_0)$, so the overall probability density is equal to

$$\left(\frac{1}{|F_0|}\right)^K \cdot \left(\frac{1}{2R_0}\right)^K = \frac{1}{2^K \cdot (|F_0| \cdot R_0)^K}. \tag{3.19}$$

Thus, the largest possible likelihood value is attained when the product $|F_0| \cdot R_0$ takes the smallest possible value.

So, in this case, we need to find the smallest possible value of this product under the following constraints:

$$\begin{gathered} F_0 - R_0 \le Y_1 - G_{1,0} - \sum_{i=1}^m c_i \cdot G_{1,i} \le R_0; \\ \ldots \\ F_0 - R_0 \le Y_K - G_{K,0} - \sum_{i=1}^m c_i \cdot G_{K,i} \le R_0. \end{gathered} \tag{3.20}$$

This is no longer a linear programming problem – constraints are still linear, but the objective function $|F_0| \cdot R_0$ is now non-linear.

3.4 In General, How Reliable Are the Results?

Natural question: can we make a certain conclusion. A natural question is: if by using some of these methods – whether it is stochastic frontier analysis with traditional distributions or with new distributions – we conclude that $f_k < 1$, does this mean, with certainty, that the current production is not maximally effective?

First answer: no, we cannot be absolutely certain. The correct answer is: not really. What we are doing is representing the difference $Y_k - F(x_{k,1}, \ldots, x_{k,n})$ – between the actual productivity Y_k (measured logarithmically) and the productivity $F(x_{k,1}, \ldots, x_{k,n})$ predicted by the model – as the sum $R_k + F_k$ of the truly random term R_k (which can be positive and negative) and the ineffectiveness term F_k which can only be negative. But of course, there is nothing unique about this representation: from the mathematical viewpoint, we can take one of the components of the term F_k and add it to the random term.

Second answer: but this is still useful. This mathematical non-uniqueness does not mean, of course, that stochastic frontier analysis is not practically helpful. As usual with statistical methods, all the conclusions are valid only with some certainty.

So, if the estimate for F_k is negative, then it is worth looking into why the productivity of this particular country, company, or farm is lower than the model predicts. Maybe there are some factors that we did not take into account when designing a model – in this case, we need to adjust our model by taking these factors into consideration.

If after this adjustment, we still get negative F_k, then it is a good idea to analyze everything and to see what can be improved.

Chapter 4

How to Make Decisions under Interval Uncertainty

Once we have the information about the situation – which often comes in terms of intervals – and we know that the currently made decisions are not perfect, we need to come up with new decisions.

In the ideal world, we know the exact consequences of each action. In this case, it is relatively straightforward to compare different possible actions and, as a result of this comparison, to select the best action. In real life, we only know the consequences with some uncertainty. A typical example, as we have mentioned, is interval uncertainty, when we only know the lower and upper bounds on the expected gain. How can we compare such interval-valued alternatives?

A usual way to compare such alternatives is to use the optimism-pessimism criterion developed by Nobelist Leo Hurwicz. In this approach, we maximize a weighted combination of the worst-case and the best-case gains, with the weights reflecting the decision maker's degree of optimism.

There exist several justifications for this criterion; however, some of the assumptions behind these justifications are not 100% convincing. In this chapter, we propose new, hopefully more convincing justifications for Hurwicz's approach.

Comment. Results from this chapter first appeared in [11].

4.1 Formulation of the Problem

Need to make decisions under interval uncertainty. In many real-life situations, we need to make a decision, i.e., we need to select one of the possible alternatives. For example, we want to select the best investment strategy, we need to decide whether to accept a new job offer, etc.

In the ideal world, we should know the exact consequence of each possible alternative. In such an ideal case, we select an alternative which is the best for us. For example, if the goal of the investment is to save for retirement, then we should select the investment strategy that will bring us the larger amount of money by the expected retirement date.

In real world, there is uncertainty. We can rarely predict the exact consequences of each action. In the simplest case, instead of knowing the exact amount of money m resulting from each alternative, we only know that this amount will be somewhere between the values $\underline{m}$ and $\overline{m}$. In other words, we do not know the exact value m; instead, we only know the interval $[\underline{m}, \overline{m}]$ that contains the actual (not yet known) value m. Such a situation is known as the situation of *interval uncertainty.* If we know intervals corresponding to different alternatives, which alternative should we select?

In other cases, in addition to the bounds $\underline{m}$ and $\overline{m}$, we also have some information about which values from the corresponding interval are more probable and which are less probable. In other words, we have some information – usually partial – about the actual probability distribution on the interval $[\underline{m}, \overline{m}]$. Sometimes, we know the exact probability distribution. In this case, we can, e.g., select the alternative for which the expected gain is the largest – of, if we want to be cautious, e.g., the alternative for which the gain guaranteed with a certain probability (e.g., 80%) is the largest.

In practice, we rarely know the exact probability distribution. Even if we know that the distribution is, e.g., Gaussian, we still do not know the exact values of the corresponding parameters – from the observations, we can only determine parameters with some uncertainty. For different possible combinations of these parameters, the expected gain – or whatever else characteristic we use – may take different values. Thus, for each alternative, instead of the *exact* value m of the corresponding objective function (such as expected gain), we have a whole *interval* $[\underline{m}, \overline{m}]$ of possible values of this objective function. So, we face the exact same problem as in the simplest possible case – we need to select an alternative in a situation when for each alternative, we only know the interval of possible values of the objective function.

How decisions under interval uncertainty are currently made. As we have mentioned earlier, decision making under interval uncertainty is an important practical problem. Not surprisingly, methods for solving this problem have been known for many decades. Usually, practitioners use a solution proposed in the early 1950s by the future Nobelist Leo Hurwicz; see, e.g., [22, 27, 39]. According to this solution, a decision maker should:

- first, select a parameter α from the interval $[0, 1]$, and then
- select an alternative for which the following combination attains the largest possible value:

$$\alpha \cdot \overline{u} + (1 - \alpha) \cdot \underline{u}.$$

This idea is known as the *optimism-pessimism* criterion, and the selected value α is known as the *optimism parameter*. The reason for these terms is straightforward:

- If $\alpha = 1$, this means that the decision maker simply selects the alternative with the largest possible value of $\overline{m}$. In other words, the decision maker completely ignores the possibility that the outcome of each alternative can be smaller than in the best possible case, and bases his/her decision exclusively on comparing these best possible consequences of different actions. This is clearly an extreme case of an optimist.
- Vice versa, if $\alpha = 0$, this means that the decision maker simply selects the alternative with the largest possible value of $\underline{m}$. In other words, the decision maker completely ignores the possibility that the outcome of each alternative can be better that in the worst possible case, and bases his/her decision exclusively on comparing these

worst possible consequences of different actions. This is clearly an extreme case of a pessimist.

Both these situations are extreme. In real life, most people take into account both good and bad possibilities, i.e., in Hurwicz terms, they make decisions based on some intermediate value α – which is larger than the pessimist's 0 but smaller than the optimist's 1.

How can we explain the current approach to decision making under uncertainty. There exist reasonable explanations for Hurwicz criteria, both:

- for the case when the outcome of each alternative is simply monetary and
- for the case when the outcome is not monetary – in this case, decision theory helps us describe the user's preferences in terms of special values known as *utilities*; see, e.g., [13, 27, 39, 45, 53] and references therein.

Remaining problem and what we do in this chapter. In both monetary and utility cases, to derive Hurwicz's formula, we need to make certain assumptions;

- some of these assumptions are more reasonable,
- some of these assumptions are slightly less convincing.

Natural questions are:

- Do we need these somewhat less convincing assumptions?
- Can we avoid them altogether – and, if not, can we replace them with somewhat more convincing assumptions?

These are the question that we will analyze – and answer – in this chapter.

Structure of this chapter. We will start the chapter with the easier-to-describe and easier-to-analyze case of monetary alternatives. First, in Section 2, we describe the usual assumptions leading to the Hurwicz criterion, explain how the Hurwicz criterion can be derived from these assumptions (in this, we largely follow [28]), and why some of these assumptions may not sound fully convincing. Then, in Section 3, we present new – hopefully more convincing – assumptions, and show how Hurwicz criterion can be derived from the new assumptions.

Then, we deal with the utility case. In Section 4, we briefly remind the readers who are not familiar with all the technical details of decision theory, what is utility and what are the properties of utility. In Section 5, we describe the usual assumptions leading to Hurwicz criterion for the utility case (they are somewhat different from the monetary case), explain how Hurwicz criterion can be derived from these assumptions (in this, we also largely follow [28]), and why some of these assumptions may not sound fully convincing. Finally, in Section 6, we show that the Hurwicz criterion can be derived from the (hopefully) more convincing assumptions in the utility case as well.

4.2 Monetary Case: Usual Derivation of the Hurwicz Criterion and the Limitations of This Derivation

To make a decision, we need to have an exact numerical equivalent for each interval. We want to be able to compare different alternatives with interval uncertainty. In particular, for each interval-valued alternative $[\underline{m}, \overline{m}]$ and for each alternative with a known exact monetary value m, we need to be able to decide:

- whether the exact-valued alternative is better or

- whether the interval-valued alternative is better.

Of course, if $m < \underline{m}$, then no matter what is the actual value from the interval $[\underline{m}, \overline{m}]$, this value will be larger than m. Thus, in this case, the interval alternative is clearly better. We will denote this by $m < [\underline{m}, \overline{m}]$.

Similarly, if $m > \overline{m}$, then no matter what is the actual value from the interval $[\underline{m}, \overline{m}]$, this value will be smaller than m. Thus, in this case, the interval alternative is clearly worse: $[\underline{m},\overline{m}] < m$.

If $m < [\underline{m},\overline{m}]$ and $m' < m$, the clearly $m' < [\underline{m},\overline{m}]$. Similar, if $[\underline{m},\overline{m}] < m$ and $m < m'$, then $[\underline{m},\overline{m}] < m'$.

One can show that because of this, there is a threshold value separating the two cases, namely, the value

$$\sup\{m : m < [\underline{m},\overline{m}]\} = \inf\{m : [\underline{m},\overline{m}] < m\}.$$

Let us denote this threshold value – depending on $\underline{m}$ and $\overline{m}$ – by $f(\underline{m},\overline{m})$.

By definition, for every $\varepsilon > 0$, we have

$$f(\underline{m},\overline{m}) - \varepsilon < [\underline{m},\overline{m}] < f(\underline{m},\overline{m}) + \varepsilon.$$

In particular, this property holds for an arbitrarily small ε, including such small ε that no one will notice the difference between the value m and the values $m-\varepsilon$ and $m+\varepsilon$. So, from the practical viewpoint, we can say that the interval $[\underline{m},\overline{m}]$ is *equivalent* to the monetary value $f(\underline{m},\overline{m})$. We will denote this equivalence by

$$[\underline{m},\overline{m}] \equiv f(\underline{m},\overline{m}).$$

From this viewpoint, all we need to do to describe decision making under interval uncertainty is to describe the corresponding function $f(\underline{m},\overline{m})$.

The numerical value $f(\underline{m},\overline{m})$ should always be between $\underline{m}$ and $\overline{m}$. Since, as we have mentioned earlier, for every value $m < \underline{m}$, we have $m < [\underline{m},\overline{m}]$, the set $\{m : m < [\underline{m},\overline{m}]\}$ contains all the values from the set $(-\infty,\underline{m})$. Thus, its supremum $f(\underline{m},\overline{m})$ has to be greater than or equal to all the values $m < \underline{m}$, in particular, than all the values $m = \underline{m} - 1/n$. So, we must have

$$\underline{m} - 1/n < f(\underline{m},\overline{m})$$

for all n. In the limit $n \to \infty$, we conclude that $\underline{m} \le f(\underline{m},\overline{m})$.

Similarly, since for every value $m > \overline{m}$, we have $[\underline{m},\overline{m}] > m$, the set

$$\{m : [\underline{m},\overline{m}] < m\}$$

contains all the values from the set $(\overline{m}, \infty)$. Thus, its infimum $f(\underline{m}, \overline{m})$ has to be smaller than or equal to all the values $m > \overline{m}$ – in particular, than all the values $m = \overline{m} + 1/n$. So, we must have $f(\underline{m}, \overline{m}) < \overline{m} + 1/n$ for all m. In the limit $n \to \infty$, we conclude that $f(\underline{m}, \overline{m}) \le \overline{m}$.

Based on the two above examples, we should always have $\underline{m} \le f(\underline{m}, \overline{m}) \le \overline{m}$.

Let us prepare for the usual derivation of Hurwicz criterion. In order to explain the usual derivation of Hurwicz criterion from several assumptions, let us first provide the usual motivation for these assumptions.

Monotonicity. Let us assume that we start with an interval $[\underline{m}, \overline{m}]$, and then we:

- delete all the lowest-value options – i.e., options for which $m \le \underline{m}'$ for some $\underline{m}' > \underline{m}$, and/or:
- add several higher-value options, with $m > \overline{m}$, e.g., all the values from $\overline{m}$ to some larger value $\overline{m}' > \overline{m}$.

After this, we get a clearly better interval $[\underline{m}', \overline{m}']$. Thus, we conclude that the function $f(\underline{m}, \overline{m})$ should be *monotonic*: if $\underline{m} \le \underline{m}'$ and $\overline{m} \le \overline{m}'$, then $f(\underline{m}, \overline{m}) \le f(\underline{m}', \overline{m}')$.

Additivity. Suppose that we have two situations:

- in the first situation, we can get any value from $\underline{a}$ to $\overline{a}$, and
- in the second situation, we can get any value from $\underline{b}$ to $\overline{b}$.

By definition of the function $f(\underline{m}, \overline{m})$, we are willing to pay the value $f(\underline{a}, \overline{a})$ to participate in the first situation, and the value $f\left(\underline{b}, \overline{b}\right)$ to participate in the second situation.

What if we consider these two choices as a single situation? In this case, the smallest possible value that we get overall – in both situations – is when we get the smallest possible value $\underline{a}$ in the first situation and the smallest possible value $\underline{b}$ in the second situation. In this case, the overall value is $\underline{a} + \underline{b}$.

Similarly, the largest possible value that we get overall – in both situations – is when we get the largest possible value $\overline{a}$ in the first situation and the largest possible value $\overline{b}$ in the second situation. In this case, the overall value is $\overline{a} + \overline{b}$.

Thus, when we consider these two choices as a single situation, the interval of possible monetary gains has the form $\left[\underline{a} + \underline{b}, \overline{a} + \overline{b}\right]$. So, the equivalent monetary value of the two choices treated as a single situation is $f\left(\underline{a} + \underline{b}, \overline{a} + \overline{b}\right)$.

It is reasonable to require that the price that we pay for two choices sold together should be equal to the sum of the prices that we pay for two choices taken separately, i.e., that $f\left(\underline{a} + \underline{b}, \overline{a} + \overline{b}\right) = f(\underline{a}, \overline{a}) + f\left(\underline{b}, \overline{b}\right)$. This property is known as *additivity.*

The usual derivation of Hurwicz criterion. Now, we are ready to describe the usual derivation of Hurwicz criterion.

Definition 4.1.

- *By a* value function*, we mean a function $f(\underline{m}, \overline{m})$ that assigns, to each pair $(\underline{m}, \overline{m})$ of real numbers for which $\underline{m} \le \overline{m}$, a real number $f(\underline{m}, \overline{m})$ for which $\underline{m} \le f(\underline{m}, \overline{m}) \le \overline{m}$.*

- *We say that a value function $f(\underline{m}, \overline{m})$ is* monotonic *if whenever $\underline{m} \le \underline{m}'$ and $\overline{m} \le \overline{m}'$, then $f(\underline{m}, \overline{m}) \le f(\underline{m}', \overline{m}')$.*

- *We say that a value function $f(\underline{m}, \overline{m})$ is* additive *if for all possible values $\underline{a} \le \overline{a}$ and $\underline{b} \le \overline{b}$, we have $f\left(\underline{a} + \underline{b}, \overline{a} + \overline{b}\right) = f(\underline{a}, \overline{a}) + f\left(\underline{b}, \overline{b}\right)$.*

- *We say that a value function $f(\underline{m}, \overline{m})$ has a* Hurwicz *form if it has the form $f(\underline{m}, \overline{m}) = \alpha \cdot \overline{m} + (1 - \alpha) \cdot \underline{m}$ for some $\alpha \in [0, 1]$.*

Proposition 4.1. *For a value function $f(\underline{m}, \overline{m})$, the following two conditions are equivalent to each other:*

- *the value function is monotonic and additive;*

- *the value function has the Hurwicz form.*

Proof. It is easy to prove that a Hurwicz-form value function is monotonic and additive.

Vice versa, let us assume that a value function $f(\underline{m}, \overline{m})$ is monotonic and additive. Let us denote $\alpha \stackrel{\text{def}}{=} f(0,1)$.

Due to additivity, for every natural number n, we have

$$[0, 1/n] + \ldots + [0, 1/n] \ (n \text{ times}) = [0, 1],$$

thus

$$f(0, 1/n) + \ldots + f(0, 1/n) \ (n \text{ times}) = n \cdot f(0, 1/n) = f(0, 1) = \alpha,$$

hence $f(0, 1/n) = \alpha \cdot (1/n)$.

Similarly, for every m and n, we have

$$f(0, m/n) = f(0, 1/n) + \ldots + f(0, 1/n) \ (m \text{ times}) = m \cdot f(0, 1/n) = \alpha \cdot (m/n).$$

For every real number r, we have $m/n \le r \le (m+1)/n$, where $m \stackrel{\text{def}}{=} \lfloor r \cdot n \rfloor$. Thus, due to monotonicity, we have $f(0, m/n) \le f(0, r) \le f(0, (m+1)/n)$, i.e., $\alpha \cdot (m/n) \le f(0, r) \le \alpha \cdot (m+1)/n$. Here, $0 \le r - m/n \le 1/n$, so in the limit $n \to \infty$, we have $m/n \to r$ and $(m+1)/n \to r$. Thus, the above inequality leads to $f(0, r) = \alpha \cdot r$.

In particular, for every $\underline{m} \le \overline{m}$, we have $f(0, \overline{m} - \underline{m}) = \alpha \cdot (\overline{m} - \underline{m})$. By the property of a value function, we have $\underline{m} \le f(\underline{m}, \underline{m}) \le \underline{m}$, i.e., $f(\underline{m}, \underline{m}) = \underline{m}$. Thus, due to additivity,

$$f(\underline{m}, \overline{m}) = f(\underline{m} + 0, \underline{m} + (\overline{m} - \underline{m})) = f(\underline{m}, \underline{m}) + f(0, \overline{m} - \underline{m}) = \underline{m} + \alpha \cdot (\overline{m} - \underline{m}).$$

One can easily check that this is indeed the Hurwicz expression.

Limitations. The above motivations are reasonably reasonable, but they may not be 100% convincing.

Indeed, we argued that if the worst-case scenario is possible for each of the two situations, then it is possible that we have the worst-case scenario in both situations. This may sound reasonable, but it is not in full agreement with common sense. Indeed, e.g., when we fly from point A to point B, we understand:

- that there may an unexpected delay at the airport A,
- that a plane may have a problem in flight and we will have to get back,
- that there may a problem at the airport B and we will get stuck on the plane, etc.,

but we honestly do *not* believe that all these low-probable disasters will happen at the same – this only happens in comedies describing lovable losers who always get into trouble.

We can raise another issues about the additivity requirement: that additivity assumes that for the combination of two items, we always pay the same price as for the two items separately. Sometimes, this is true, but often, this is not true: there *are* discounts if you buy several items (or several objects of the same type) at the same time.

What should we do? Since the arguments that we used above to justify the assumptions are not 100% convincing, maybe we can find somewhat more convincing arguments in favor of Hurwicz formula – or, alternatively, maybe these more convincing arguments can lead us to a different formula?

This is what we will analyze in the next section.

4.3 Monetary Case: New, Hopefully More Convincing, Derivation of the Hurwicz Criterion

Shift-invariance. Suppose that we offer a user a package deal in which he/she gets m dollars cash *and* an alternative in which he/she gets between $\underline{m}$ and $\overline{m}$. The equivalent value for the interval-value alternative is $f(\underline{m}, \overline{m})$, so the overall value for this package is $m + f(\underline{m}, \overline{m})$.

On the other hand, if we consider this a package deal, then in this deal, we get any amount between $m + \underline{m}$ and $m + \overline{m}$. Thus, the value of this package deal should be equal to $f(m + \underline{m}, m + \overline{m})$. It is reasonable to require that these two valuations should lead to the same result, i.e., that we should have $m + f(\underline{m}, \overline{m}) = f(m + \underline{m}, m + \overline{m})$. In mathematical terms, this property is known as *shift-invariance.*

Discussion. At first glance shift-invariant is very similar to additivity. Indeed, it can be viewed as a particular case of additivity, in which the first interval is simply the interval $[m, m]$ consisting of a single number m.

But good news is that both above objections to general additivity do not apply here. Indeed, we are not talking about a combination of rare events, so the first objection is not applicable. The second objection is also not applicable, since while we may expect a discount if we buy two big bottles of milk, no one expects a discount if we buy a bottle of milk and a fixed amount of money (e.g., when we ask to change a big banknote when paying).

Need for additional assumptions. If we limit ourselves only to shift-invariance, we will get too many possibilities in addition to Hurwicz formula: specifically, one can see that we can have a more general expression

$$f(\underline{m}, \overline{m}) = \underline{m} + F(\overline{m} - \underline{m}),$$

where $F(z)$ is a monotonic function defined for all $z \geq 0$ for which $F(z) \leq z$ for all z – e.g., $F(z) = z/(1 + z)$. (By the way, it is possible to show that the above expression is the most general form of a monotonic shift-invariant value function.)

To narrow down the class of possible value functions, we need to make additional reasonable assumptions. We will describe one such assumption right away.

A new assumption: transitivity. Let us start with the same interval $[0, 1]$ with which we started the proof of Proposition 4.1. Similarly to this proof, let us denote the value $f(0, 1)$ corresponding to this interval by α.

What can we conclude that from the fact that $f(0, 1) = \alpha$? Well, due to shift invariance, we can conclude that for every x, we have $f(x, 1 + x) = \alpha + x$. From the mathematical viewpoint, this is all that we can conclude. However, from the common sense viewpoint, we can make yet another conclusion.

Indeed, e.g., for each x from the interval $[0,1]$, the alternative corresponding to the interval $[x, 1+x]$ is equivalent to getting a monetary amount $\alpha + x$: $[x, 1+x] \equiv \alpha + x$. If we do not know which of these intervals the alternative corresponds to – but we know that it corresponds to one of these alternatives, this means that the actual gain can take any value from the *union* of these intervals. Each of these intervals is equivalent to the value $\alpha + x$, thus, the union of these intervals is equivalent to the set of all possible values $\alpha + x$ when $x \in [0,1]$:

$$\bigcup_{x \in [0,1]} [x, 1+x] \equiv \{\alpha + x : x \in [0,1]\}.$$

Let us estimate the left-hand side and the right-hand side of this equality.

- The smallest possible value in the left-hand side is when we take the smallest value from the interval $[x, 1+x]$ – i.e., the value x – for the smallest possible value x from the interval $[0,1]$ (i.e., for the value $x = 0$). Thus, the smallest possible value in the left-hand side is equal to 0.

- The largest possible value in the left-hand side is when we take the largest value from the interval $[x, 1+x]$ – i.e., the value $1 + x$ – for the largest possible value x from the interval $[0,1]$ (i.e., for the value $x = 1$). Thus, the largest possible value in the left-hand side is equal to $1 + 1 = 2$.

So, the left-hand side of the above equality is the interval $[0,2]$.

Similarly:

- The smallest possible value in the right-hand side is when we take the smallest possible value x from the interval $[0,1]$, i.e., the value $x = 0$. Thus, the smallest possible value in the right-hand side is equal to $\alpha + 0 = \alpha$.

- The largest possible value in the right-hand side is when we take the largest possible value x from the interval $[0,1]$, i.e., the value $x = 1$. Thus, the smallest possible value in the right-hand side is equal to $\alpha + 1$.

So, the left-hand side of the above equality is the interval $[\alpha, 1+\alpha]$.

Thus, the above equivalent takes the form $[0,2] \equiv [\alpha, 1+\alpha]$. Good news is that we already known – as a particular case of shift-invariance – that the interval $[\alpha, 1+\alpha]$ is equivalent to the value $\alpha + \alpha = 2\alpha$. Thus, by transitivity of equivalence, we conclude that the interval $[0,2]$ is equivalent to 2α, i.e., that $f(0,2) = 2\alpha$. Then, by shift-invariance, we will get $f(x, 2+x) = 2\alpha + x$ for each x.

By similarly combining intervals $[x, 1+x]$ corresponding to $x \in [0,2]$, we conclude that $[0,3] \equiv [\alpha, 2+\alpha]$, and since we already know that $[\alpha, 2+\alpha] \equiv 2\alpha + \alpha$, by transitivity, we will have $f(0,3) = 3\alpha$.

Instead of stacking intervals of width 1, we could similarly stack intervals of a different width w.

New derivation of Hurwicz formula. It turns out that this way, we can indeed get a new derivation of Hurwicz formula. Let us describe all this in precise terms.

Definition 4.2.

- *We say that a value function $f(\underline{m}, \overline{m})$ is* shift-invariant *if for every m and for all $\underline{m} \le \overline{m}$, we have $m + f(\underline{m}, \overline{m}) = f(m + \underline{m}, m + \overline{m})$.*

- *We say that a value function is* transitive *if for each w and for all $\underline{m} \le \overline{m}$, we have $f\left(\underline{\ell}, \overline{\ell}\right) = f(\underline{r}, \overline{r})$, where*

$$\left[\underline{\ell}, \overline{\ell}\,\right] \stackrel{\text{def}}{=} \bigcup_{m \in [\underline{m}, \overline{m}]} [m, w+m]$$

and

$$[\underline{r}, \overline{r}] \stackrel{\text{def}}{=} \{f(m, m+w) : m \in [\underline{m}, \overline{m}]\}.$$

Comment. In this definition, we only described transitivity for the case when all combined intervals have the exact same width. Our main motivation for this restriction is that, as we will show, only such transitivity is needed – and in derivations, it is always desirable to avoid unnecessarily general assumptions and to limit ourselves only to weakest possible assumptions – weakest possible among those that will lead to the desired derivation.

There is another reason for this limitation: as we show later in this section, if we generalize this property too much, then there will be no realistic value function at all that would satisfy thus generalized property.

Proposition 4.2. *For a value function $f(\underline{m}, \overline{m})$, the following two conditions are equivalent to each other:*

- *the value function is monotonic, shift-invariant, and transitive;*
- *the value function has the Hurwicz form.*

Proof. Similarly to our above arguments, we can see that $\left[\underline{\ell}, \overline{\ell}\,\right] = [\underline{m}, \overline{m}+w]$, so $f\left(\underline{\ell}, \overline{\ell}\,\right) = f(\underline{m}, \overline{m} + w)$.

Due to the monotonicity of the value function, we have

$$[\underline{r}, \overline{r}] = [f(\underline{m}, \overline{m}), f(\underline{m} + w, \overline{m} + w)].$$

Due to shift-invariance, we have

$$[\underline{r}, \overline{r}] = [f(\underline{m}, \overline{m}), w + f(\underline{m} + w, \overline{m} + w)],$$

so, again due to shift-invariance – this time in relation to a shift by $f(\underline{m}, \overline{m})$ – we get $[\underline{r}, \overline{r}] = f(\underline{m}, \overline{m}) + [0, w]$.

Thus, again due to shift-invariance, $f(\underline{r}, \overline{r}) = f(\underline{m}, \overline{m}) + f(0, w)$. Therefore, transitivity means that

$$f(\underline{m}, \overline{m} + w) = f(\underline{m}, \overline{m}) + f(0, w).$$

One can easily see that the Hurwicz formula is shift-invariant and satisfies the above property for all w and for all $\underline{m} \le \overline{m}$.

Vice versa, let us assume that we have a value function that satisfies this property for all w and for all $\underline{m} \le \overline{m}$. In particular, for $\underline{m} = 0$, this means that $f(0, \overline{m} + w) = f(0, \overline{m}) + f(0, w)$. This is exactly the particular case of the additivity property that we used (as well as monotonicity) in the proof of Proposition 4.1 to prove that $f(0, r) = \alpha \cdot r$

for all real numbers r. From this formula, in that proof, we used, in effect, shift-invariance to prove that the Hurwicz formula is indeed true for all $\underline{m} \le \overline{m}$. Since we still assume shift-invariance, this means that we have a derivation of the Hurwicz formula in this case as well.

The proposition is proven.

Discussion: we cannot generalize the transitivity property too much. Let us show that the transitivity assumption cannot be realistically generalized too much, to cases when united intervals have different widths.

Definition 4.3. *We say that a value function is* fully transitive *if for each family of intervals* $\{[\underline{m}(a), \overline{m}(a)]\}_{a \in A}$ *for which both sets* $\bigcup\limits_{a \in A} [\underline{m}(a), \overline{m}(a)]$ *and* $\{f(\underline{m}(a), \overline{m}(a)) : a \in A\}$ *are intervals, we have* $f\left(\underline{\ell}, \overline{\ell}\right) = f(\underline{r}, \overline{r})$*, where we denoted*

$$\left[\underline{\ell}, \overline{\ell}\,\right] = \bigcup_{a \in A} [\underline{m}(a), \overline{m}(a)]$$

and

$$[\underline{r}, \overline{r}] = \{f(\underline{m}(a), \overline{m}(a)) : a \in A\}.$$

Proposition 4.3. *For a value function* $f(\underline{m}, \overline{m})$*, the following two conditions are equivalent to each other:*

- *the value function is monotonic, shift-invariant, and fully transitive;*
- *the value function has the Hurwicz form with* $\alpha = 0$ *or* $\alpha = 1$*.*

Discussion. So, full transitivity is satisfied only in the two extreme (and unrealistic) cases:

- when $\alpha = 0$ – the case of full pessimism, and
- when $\alpha = 1$ – the case of full optimism.

Proof. One can easily check that both extreme value functions $f(\underline{m}, \overline{m}) = \underline{m}$ (that corresponds to $\alpha = 0$) and $f(\underline{m}, \overline{m}) = \overline{m}$ (that corresponds to $\alpha = 1$) are fully transitive.

Let us prove that, vice versa, every monotonic shift-invariant and fully transitive value function coincides with one of the two extreme functions. Indeed, since the general condition should be satisfied for all possible families of intervals $[\underline{m}(a), \overline{m}(a)]$, in particular, it should be satisfied for all the families from Definition 4.2. Thus, due to Proposition 4.2, the value function should have the Hurwicz form.

Now, for the family $[0, a]$, where $a \in A = [0, 1]$, the union $\left[\underline{\ell}, \overline{\ell}\right]$ is simply equal to $[0, 1]$, so $f\left(\underline{\ell}, \overline{\ell}\right) = f(0, 1) = \alpha$.

On the other hand, here, $f(0, a) = \alpha \cdot a$, so

$$[\underline{r}, \overline{r}] = \{\alpha \cdot a : a \in [0, 1]\} = [0, \alpha],$$

thus $f(\underline{r}, \overline{r}) = \alpha \cdot \alpha = \alpha^2$. Thus, the generalized transitivity is satisfied only when $\alpha = \alpha^2$, i.e., when either $\alpha = 0$ or $\alpha = 1$.

The proposition is proven.

4.4 What Is Utility and What Are the Properties of Utility: A Brief Reminder

What is utility. To apply computer-based number-oriented tools for making decisions in a non-monetary case, we need to describe the user's preferences in numerical terms. As we have mentioned in Chapter 1, in decision theory (see, e.g., [13, 27, 39, 45, 53]), this is done as follows.

Let us select the two extreme alternatives:

- a very bad alternative A_- which is worse than anything that we will actually encounter, and

- a very good alternative A_+ which is better than anything that we will actually encounter.

For each real number p from the interval $[0, 1]$, we can form a lottery – we will denote this lottery by $L(p)$ – in which:

- we get the very good alternative A_+ with probability p, and
- we get the very bad alternative A_- with the remaining probability $1 - p$.

To find how valuable is each alternative A for the decision maker, we ask him/her to compare the alternative A with lotteries $L(p)$ corresponding to different probabilities p. Here:

- when p is small, close to 0, the lottery $L(p)$ is similar to the very bad alternative A_- and is, thus, worse than A; we will denote this by $A_- < A$;
- when p is close to 1, the lottery $L(p)$ is similar to the very good alternative A_+ and is, thus, better than A: $A < L(p)$.

Also, the smaller the probability p of getting a very good alternative, the worse the lottery $L(p)$. Thus:

- if $L(p) < A$ and $p' < p$, then $L(p') < A$, and
- if $A < L(p)$ and $p < p'$, then $A < L(p')$.

Thus, similarly to the monetary case, there exists a threshold value

$$\sup\{p : L(p) < A\} = \inf\{p : A < L(p)\};$$

we will denote this threshold value by $u(A)$. This threshold value is known as the *utility* of the alternative A.

Similarly to the monetary case, for every $\varepsilon > 0$, we have $L(u(A)-\varepsilon) < A < L(u(A)+\varepsilon)$. This is true for arbitrarily small ε, in particular, for the values ε for which the difference in probabilities between $u(A)-\varepsilon$, $u(A)$, and $u(A)+\varepsilon$ is practically unnoticeable. So, we can conclude that from the practical viewpoint, the alternative A is equivalent to the lottery $L(u(A))$. We will denote this equivalence by $A \equiv L(u(A))$.

Utility is defined modulo a linear transformation. The numerical value of the utility $u(A)$ depends not only on the alternative A, it also depends on which pair (A_-, A_+) we select. What if we select a different pairs (A'_-, A'_+) – e.g., a pair for which $A_- < A'_- < A'_+ < A_+$? How will that change the numerical value of utility?

If an alternative A has utility $u'(A)$ with respect to the pair (A'_-, A'_+), this means that this alternative is equivalent to the lottery $L'(u'(A))$, in which:

- we get A'_+ with probability $u'(A)$, and
- we get A'_- with the remaining probability $1-u'(A)$.

Since $A_- < A'_- < A_+$, we can find a utility value $u(A'_-)$ for which the alternative A'_- is equivalent to the lottery $L(u(A'_-))$, in which:

- we select A_+ with probability $u(A'_-)$, and
- we select A_- with probability $1-u(A'_-)$.

Similarly, we have $A'_+ \equiv L(u(A'_+))$. Thus, the original alternative A is equivalent to a two-stage lottery, in which:

- first, we select either A'_+ (with probability $u'(A)$) or A'_- (with probability $1-u'(A)$);
- then, we select either A_+ or A_- with probabilities depending on what we selected on the first stage: if we selected A'_+ on the first stage, then we select A_+ with probability $u(A'_+)$ and A_- with probability $1-u(A'_+)$, and if we selected A'_- on the first stage, then we select A_+ with probability $u(A'_-)$ and A_- with probability $1-u(A'_-)$.

As a result of this two-stage lottery, we get either A_+ or A_-, and the probability of selecting A_+ is equal to

$$u'(A)\cdot u(A'_+) + (1-u'(A))\cdot u(A'_-).$$

By definition, this probability is the utility $u(A)$ of the alternative A with respect to the pair (A_-, A_+), thus

$$u(A) = u'(A) \cdot u(A'_+) + (1 - u'(A)) \cdot u(A'_-).$$

The right-hand side is a linear expression in terms of $u'(A)$. So, we conclude that utilities corresponding to different pairs can be obtained from each other by a linear transformation.

In other words, the numerical value of the utility is defined modulo a generic linear transformation – just like the numerical value of time and temperature, where the corresponding linear transformations mean selecting a different starting point and/or a different measuring unit.

4.5 Utility Case: Usual Derivation of the Hurwicz Criterion and the Limitations of This Derivation

Formulation of the problem. As as we have mentioned earlier, in many practical situations, we do not know the exact consequence of each action, and thus, we do not know the exact value of the corresponding utility. Instead, for such situations, we only know the interval $[\underline{u}, \overline{u}]$ of possible utility values. According to the general idea of utility, to describe the decision maker's preferences for such interval-valued situations, we must assign, to each such interval, an appropriate utility value. Similarly to the monetary case, we will denote this utility value by $f(\underline{u}, \overline{u})$, and we will call the corresponding function a value function. Clearly, we must have $\underline{u} \le f(\underline{u}, \overline{u}) \le \overline{u}$, and clearly, if $\underline{u}$ and/or $\overline{u}$ increase, the interval-valued alternative becomes better – i.e., the value function should be monotonic.

What are other natural properties of the value function?

We cannot reuse assumptions from the monetary case. We cannot simply use the same properties as in the monetary case. For example, additivity makes no sense:

- it makes perfect sense to add dollar amounts, but

- it makes no sense to add probabilities (and utilities, as we have explained, *are* probabilities).

So, we need alternative assumptions.

Assumptions used in the usual derivation of the utility-case Hurwicz formula. Since utility is defined modulo a general linear transformation, it makes sense to require that the formulas transforming the bounds $\underline{u}$ and $\overline{u}$ into an equivalent utility should remain the same if we linearly "re-scale" all utility values. In particular:

- if we have $f(\underline{u}, \overline{u}) = u$, then after shifting all the utility values by u_0 we should retain the same relation between the shifted utilities $\underline{u}' = \underline{u} + u_0$, $\overline{u}' = \overline{u} + u_0$, and $u' = u + u_0$: $f(\underline{u}', \overline{u}') = u'$;

- similarly, if we have $f(\underline{u}, \overline{u}) = u$, then after re-scaling all the utility values by a factor $c > 0$, we should retain the same relation between the shifted utilities $\underline{u}' = c \cdot \underline{u}$, $\overline{u}' = c \cdot \overline{u}$, and $u' = c \cdot u$: $f(\underline{u}', \overline{u}') = u'$.

In the shift case, if we substitute the values $\underline{u}' = \underline{u} + u_0$, $\overline{u}' = \overline{u} + u_0$, and $u' = u + u_0 = f(\underline{u}, \overline{u}) + u_0$ into the desired equality $f(\underline{u}', \overline{u}') = u'$, we get the requirement $f(\underline{u} + u_0, \overline{u} + u_0) = f(\underline{u}, \overline{u}) + u_0$. One can see that this is exactly the property that we called shift-invariance.

In the re-scaling case, if we substitute the values $\underline{u}' = c \cdot \underline{u}$, $\overline{u}' = c \cdot \overline{u}$, and $u' = c \cdot u = c \cdot f(\underline{u}, \overline{u})$ into the desired equality $f(\underline{u}', \overline{u}') = u'$, we get the requirement $f(c \cdot \underline{u}, c \cdot \overline{u}) = c \cdot f(\underline{u}, \overline{u})$. We will call this property *scale-invariance.*

Definition 4.4. *We say that a value function $f(\underline{u}, \overline{u})$ is* scale-invariant *if for every $c > 0$ and for all $\underline{u} \le \overline{u}$, we have $f(c \cdot \underline{u}, c \cdot \overline{u}) = c \cdot f(\underline{u}, \overline{u})$.*

Proposition 4.4. *For a value function $f(\underline{u}, \overline{u})$, the following two conditions are equivalent to each other:*

- *the value function is monotonic, shift-invariant, and scale-invariant;*

- *the value function has the Hurwicz form.*

Proof. It is easy to check that the Hurwicz formula is monotonic, shift-invariant, and scale-invariant. Let us show that, vice versa, every monotonic value function which is shift- and scale-invariant has the Hurwicz form.

Indeed, as in the proof of Proposition 4.1, let us denote $\alpha \stackrel{\text{def}}{=} f(0,1)$. For all $\underline{u} < \overline{u}$, due to shift-invariance with $u_0 = \underline{u}$, we have $f(\underline{u},\overline{u}) = \underline{u} + f(0,\overline{u} - \underline{u})$. Now, due to scale-invariance with $c = \overline{u} - \underline{u}$, we get $f(0,\overline{u} - \underline{u}) = (\overline{u} - \underline{u}) \cdot f(0,1) = (\overline{u} - \underline{u}) \cdot \alpha$.

Thus, $f(\underline{u},\overline{u}) = \underline{u} + f(0,\overline{u} - \underline{u}) = \underline{u} + (\overline{u} - \underline{u}) \cdot \alpha$, which is exactly the Hurwicz formula.

The proposition is proven.

Limitations. Shift-invariance is indeed reasonable, but scale-invariance is not fully convincing. Yes, indeed, we can have different units for measuring utility – just like we can use different units for measuring money, but it is not very convincing to expect that people will make the same choices involving 100 US dollars as in situations involving 100 Pesos (which at present, in 2020, represents a 20 times smaller amount of money).

It is therefore desirable to replace scale-invariance with a more convincing assumption.

4.6 Utility Case: New, Hopefully More Convincing, Derivation of the Hurwicz Criterion

We cannot simply dismiss scale-invariance. We cannot simply dismiss scale-invariance and keep only shift-invariance: we already considered this scenario when discussing the monetary case, and we showed that in this case, there are too many values functions satisfying these requirements.

So, we need additional assumptions – assumptions which are more convincing that scale-invariance.

What we propose. What we propose is the above-described transitivity property. The arguments in favor of this property apply verbatim to the utility case. And we already know – from Proposition 4.2 – that if we require shift-invariance and transitivity, then the only value functions we get are Hurwicz ones.

Thus, indeed, we get a new, (hopefully) more convincing derivation of the Hurwicz criterion in the utility case as well.

References

[1] A. D. Alexandrov, "On Lorentz transformations", *Uspekhi Math. Nauk*, 1950, Vol. 5, No. 3, pp. 187 (in Russian).

[2] A. D. Alexandrov and V. V. Ovchinnikova, "Remarks on the foundations of special relativity", *Leningrad University Vestnik*, 1953, No. 11, pp. 94–110 (in Russian).

[3] A. D. Alexandrov, "A contribution to chronogeometry", *Canadian Journal of Mathematics*, 1967, Vol. 19, pp. 1119–1128.

[4] A. D. Alexandrov, "Mappings of spaces with families of cones and space-time transformations", *Annali di Matematica Pura ed Aplicata*, 1967, Vol. 103, pp. 229–257.

[5] A. D. Alexandrov, "On the axioms of relativity theory", *Vestnik Leningrad University, Ser. Math.*, 1967, Vol. 19, pp. 5–28.

[6] R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.

[7] W. Benz, "A characterization of plane Lorentz transformations", *Journal of Geometry*, 1977, Vol. 10, pp. 45–56.

[8] W. Benz, *Geometrische Transformationen*, BI Wissenscahftsverlag, Mannheim, 1992.

[9] L. A. Berrout Ramos, V. Kreinovich, and K. Autchariyapanitkul, *Correcting Interval-Valued Expert Estimates: Empirical Formulas Explained*, University of Texas at El Paso, Department of Computer Science, Technical Report UTEP-CS-21-72, 2021, http://www.cs.utep.edu/vladik/2021/tr21-72.pdf

[10] R. L. Cheu, H. T. Nguyen, T. Magoc, and V. Kreinovich, "Logit discrete choice

model: a new distribution-free justification", *Soft Computing*, 2009, Vol. 13, No. 2, pp. 133–137.

[11] W. Chinnakum, L. Berrout Ramos, O. Iyiola, and V. Kreinovich, "Decision making under interval uncertainty: towards (somewhat) more convincing justifications for hurwicz optimism-pessimism approach", *Asian Journal of Economics and Banking (AJEB)*, 2021, Vol. 5, No. 1, pp. 32–45.

[12] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.

[13] P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.

[14] P. C. Fishburn, *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, Maryland, 1988.

[15] T. Gajdos and J.-C. Vergnaud, "Decisions with conflicting and imprecise information", *Social Choice and Welfare*, 2013, Vol. 41, pp. 427–452,
doi: 10.1007/s00355-012-0691-1

[16] A. K. Guts, "On mappings of families of sets", *Soviet Math, Dokl.*, 1973, Vol. 14, pp. 506–508.

[17] A. K. Guts, "On mappings preserving cones in a Lobachevski space", *Math. Notes*, 1973, Vol. 13, pp. 411–415.

[18] A. K. Guts, "Mappings of an ordered Lobachevsky space", *Soviet Math. Dok.*, 1974, Vol. 15, pp. 416–419.

[19] A. K. Guts, "Invariant orders on three-dimensional Lie groups", *Siberian Math. Journal*, 1976, Vol. 17, pp. 731–736.

[20] A. K. Guts, "The chronogeometry of Gödel and de Sitter manifolds", *Siberian Math. Journal*, 1980, pp. 509–513.

[21] A. K. Guts, "Axiomatic relativity theory", *Russian Math. Surveys*, 1982, Vol. 37, No. 2, pp. 41–89.

[22] L. Hurwicz, *Optimality Criteria for Decision Making Under Ignorance*, Cowles Commission Discussion Paper, Statistics, No. 370, 1951.

[23] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.

[24] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[25] O. Kosheleva, V. Kreinovich, and P. G. Vroegindewey, "An extension of a theorem of A. D. Alexandrov to a class of partially ordered fields", *Proceedings of the Royal Academy of Science of Netherlands, Series A*, 1979, Vol. 82, No. 3, pp. 363–376.

[26] V. Kreinovich, "Approximately measured causality implies the Lorentz group: Alexandrov-Zeeman result made more realistic", *International Journal of Theoretical Physics*, 1994, Vol. 33, No. 8, pp. 1733–1747.

[27] V. Kreinovich, "Decision making under interval uncertainty (and beyond)", In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models For Social Sciences*, Springer Verlag, 2014, pp. 163–193.

[28] V. Kreinovich, "Decision making under interval (and more general) uncertainty: monetary vs. utility approaches", *Journal of Computational Technologies*, 2017, Vol. 22, No. 2, pp. 37–49.

[29] V. Kreinovich and O. Kosheleva, "Computational complexity of determining which statements about causality hold in different space-time models", *Theoretical Computer Science*, 2008, Vol. 405, No. 1–2, pp. 50–63.

[30] V. Kreinovich, O. Kosheleva, L. Bokati, and L. Berrout, *Order Relations are Ubiquitously Fundamental: Alexandrov(-Zeeman) Theorem Extended from Space-Time Physics to Logical Reasoning and Decision Making*, University of Texas at El Paso, Department of Computer Science, Technical Report UTEP-CS-21-52, 2021, http://www.cs.utep.edu/vladik/2021/tr21-52.pdf

[31] S. C. Kumbhakar and C. A. Knox Lovell, *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge, UK, 2000.

[32] H.-P. A. Künzi and V. Kreinovich, "Static space-times naturally lead to quasi-pseudometrics", *Theoretical Computer Science*, 2008, Vol. 405, No. 1–2, pp. 64–72.

[33] A. V. Kuzminykh, "Characterization of Lorentz transformations", *Soviet Math. Dokl.*, 1975, Vol. 16, No. 6, pp. 1626–1628.

[34] A. V. Kuzminykh, "Minimal condition determining the Lorentz transformations", *Siberian Mathematical Journal*, 1976, Vol. 17, No. 6, pp. 968–972.

[35] J. A. Lester, "On null cone preserving mappings", *Proceedings of Cambridge Mathematical Society*, 1977, Vol. 81, pp. 455–462.

[36] J. A. Lester, "Cone preserving mappings for quadratic cones over arbitrary fields", *Canadian Journal of Mathematics*, 1977, Vol. 29, pp. 1247–1253.

[37] J. A. Lester, "A physical characterization of conformal transformations of Minkowski spacetime", *Annals of Discrete Mathematics*, 1983, Vol. 18, pp. 567–574.

[38] D. Luce, *Individual Choice Behavior*, John Wiley and Sons, New York, 1959.

[39] R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

[40] D. Luce and P. Suppes, "Preference, utility, and subjective probability", In: D. Luce, R. Bush, and E. Galanter, *Handbook on Mathematical Psychology*, John Wiley and Sons, New York, 1965, pp. 249–410.

[41] D. McFadden, "Conditional logit analysis of qualitative choice behavior", In: P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, 1974, pp. 105–142.

[42] D. McFadden, "Economic choices", *American Economic Review*, 2001, Vol. 91, pp. 351–378.

[43] J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.

[44] G. L. Naber, *The Geometry of Minkowski Space-Time*, Springer-Verlag, N.Y., 1992.

[45] H. T. Nguyen, O. Kosheleva, and V. Kreinovich, "Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction", *International Journal of Intelligent Systems*, 2009, Vol. 24, No. 1, pp. 27–47.

[46] H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.

[47] H. T. Nguyen, V. Kreinovich, and P. Wojciechowski, "Strict Archimedean t-norms and t-conorms as universal approximators", *International Journal of Approximate Reasoning*, 1998, Vol. 18, No. 3–4, pp. 239–249.

[48] H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.

[49] N. Nguyen, *Estimation of Technical Efficiency in Stochastic Frontier Analysis*, PhD Dissertation, Bowling Green State University, 2010.

[50] Nguyen Ngoc Thach, L. Berrout, and O. Kosheleva, "How effective are we: towards a more convincing Stochastic Frontier Analysis", *Soft Computing*, 2021, Vol. 25, No. 12, pp. 7991–7996.

[51] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.

[52] V. Pambuccian, "Alexandrov-Zeeman type theorems expressed in terms of definability", *Aequationes Mathematicae*, 2007, Vol. 74, pp. 249–261.

[53] H. Raiffa, *Decision Analysis*, McGraw-Hill, Columbus, Ohio, 1997.

[54] D. J. Sheskin, *Handbook of Parametric and Non-Parametric Statistical Procedures*, Chapman & Hall/CRC, London, UK, 2011.

[55] M. Smithson, "Conflict and ambiguity: preliminary models and empirical tests", *Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications ISIPTA'2013*, Compiègne, France, July 2–5, 2013, pp. 303–310.

[56] M. Smithson, "Probability judgments under ambiguity and conflict", *Frontiers in Psychology*, 2015, Vol. 6, Paper 674, doi: 10.3389/fpsyg.2015.00674

[57] K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2017.

[58] K. Train, *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, Massachusetts, 2003.

[59] R. J. Vanderbei, *Linear Programming: Foundations and Extensions*, Springer, Cham, Switzerland, 2020.

[60] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.

[61] E. C. Zeeman, “Causality implies the Lorentz group”, *Journal of Mathematical Physics*, 1964, Vol. 5, pp. 490–493.

Curriculum Vitae

Laura A. Berrout Ramos received her Bachelor in Mechatronics Engineering degree from the Universidad Autónoma de Ciudad Juárez (UACJ) in 2012, with a thesis titled "Numerical 2D Simulation of a Thin-Film Transistor with Cadmium Sulfide as Active Layer". In 2015, she received Master's degree in Electrical Engineering from UACJ with a thesis titled "Facial Expression Recognition System with Harr-like Features and Support Vector Machines". While taking graduate classes at the University of Texas at El Paso (UTEP), she also worked as an Instructor with UTEP's Professional and Public Programs and as a Research Assistant at UTEP's Office of the Provost.

During her studies at UTEP, she has published two papers and two technical reports:

- W. Chinnakum, L. Berrout Ramos, O. Iyiola, and V. Kreinovich, "Decision making under interval uncertainty: towards (somewhat) more convincing justifications for Hurwicz optimism-pessimism approach", *Asian Journal of Economics and Banking (AJEB)*, 2021, Vol. 5, No. 1, pp. 32–45.

- Nguyen Ngoc Thach, L. Berrout, and O. Kosheleva, "How effective are we: towards a more convincing Stochastic Frontier Analysis", *Soft Computing*, 2021, Vol. 25, No. 12, pp. 7991–7996.

- L. A. Berrout Ramos, V. Kreinovich, and K. Autchariyapanitkul, *Correcting Interval-Valued Expert Estimates: Empirical Formulas Explained*, UTEP, Department of Computer Science, Technical Report UTEP-CS-21-72, 2021.

- V. Kreinovich, O. Kosheleva, L. Bokati, and L. Berrout, *Order Relations are Ubiquitously Fundamental: Alexandrov(-Zeeman) Theorem Extended from Space-Time Physics to Logical Reasoning and Decision Making*, UTEP, Department of Computer Science, Technical Report UTEP-CS-21-52, 2021.