Appearing in the *Journal of the Franklin Institute*, 336: 1093–1117 (1999).

Set-Models of Information-Gap Uncertainty: Axioms and an Inference Scheme

Yakov Ben-Haim
Faculty of Mechanical Engineering
Technion — Israel Institute of Technology
Haifa 32000 Israel
e-mail: yakov@technion.ac.il

Abstract

The sparsity and complexity of information in many technological situations has led to the development of new methods for quantifying uncertain evidence, and new schemes of inference from uncertain data. This paper deals with set-models of information-gap uncertainty which employ geometrical rather than measure-theoretic tools, and which are radically different from both probability and fuzzy-logic possibility models. The first goal of this paper is the construction of an axiomatic basis for info-gap models of uncertainty. The result is completely different from Kolmogorov's axiomatization of probability. Once we establish an axiomatically distinct framework for uncertainty, we arrive at a new possibility for inference and decision from uncertain evidence. The development of an inference scheme from info-gap models of uncertainty is the second goal of this paper. This inference scheme is illustrated with two examples: a logical riddle and a mechanical engineering design decision.

Keywords. Uncertainty modelling, axioms of uncertainty, inference with uncertainty, information-gap models, convex models.

1 Introduction

The problem of inference and decision with uncertain evidence has engaged the attention of innumerable mathematicians, philosophers and scientists, and the variety of approaches is great enough to boggle the most stout-hearted professional. Within this diversity, however, an irrevocable connection between inference and probability has been made by many thinkers, and in a way which experience in technological inference leads me to regard as surprising.

Keynes asserts:

> Part of our knowledge we obtain direct; and part by argument. The Theory of Probability is concerned with that part which we obtain by argument, and it treats of the different degrees in which the results so obtained are conclusive or inconclusive. ...
>
> The method of this treatise has been to regard subjective probability as fundamental and to treat all other relevant conceptions as derivative from this. [20, pp.3, 281–282]

Among Carnap's "basic conceptions" is the contention that

> all inductive reasoning, in the wide sense of nondeductive or nondemonstrative reasoning, is reasoning in terms of probability [15, p.v].

It would be a mistake to view these brief almost slogan-styled statements as comprehensive reflections of these men's positions. However, they are characteristic of a prevalent attitude which

[0] papers axiom axiom.tex TME 462 24.5.99

views probability as the only means of handling uncertainty. This attitude fails to account for an inference procedure which is important in engineering. The aim of the present paper is to develop the foundations of a very non-probabilistic theory of uncertainty and to demonstrate, through both example and analysis, how uncertain evidence is exploited in drawing inferences and making decisions. The set-models of information-gap uncertainty whose axiomatization is presented here are in use in many areas of technology. (A non-technical discussion may be found in [3].) The purpose of this study is, first to clarify the fundamental structure of info-gap models of uncertainty and their relationship to more classical probabilistic uncertainty models and second, to develop a non-probabilistic inference scheme for uncertain evidence.

The spirit of this paper is well expressed by Kyburg's premonition that there are

> a number of uncertainty formalisms. ... They can all be reflected as special cases of the approach of taking probabilities to be determined by sets of probability functions defined on an algebra of statements. ... But, it might be the case that some novel procedure could be used in a decision theory that is based on some non-probabilistic measure of uncertainty. [24, p.189]

One main theme of a previous article, [3], was that the rigors of technological necessity have led to the emergence of new methods for quantifying and exploiting uncertain information. By far the most well known is possibility theory based on fuzzy logic [16], which is extensively used in electro-mechanical control and decision systems. Possibility theory, while axiomatically distinct from probability [14, 18], is nonetheless similar to the extent that it quantifies uncertainties with non-negative real functions (membership functions rather than probability densities).

This paper deals with models of uncertainty which employ geometrical rather than measure-theoretic tools, and which are greatly different from both probability and possibility models. None of the traditional interpretations of probability — Laplace's enumeration of basic events, subjective preferences, frequency of recurrence — provide an explanation of info-gap models. And what is more important, the axiomatic basis of info-gap models is completely different from Kolmogorov's axiomatization of probability. Constructing an axiomatic basis of info-gap models of uncertainty is the first goal of this paper.

I can agree with Carnap's assertion after one modification: all inductive reasoning, in the wide sense of nondeductive or nondemonstrative reasoning, is reasoning in terms of **uncertainty.** Once we establish an axiomatically distinct framework for uncertainty, we arrive at new possibilities for inference and decision with uncertain evidence. The development of an inference scheme from info-gap models of uncertainty is the second goal of this paper.

In section 2 we very briefly describe several info-gap models of uncertainty which are typical in engineering applications. In section 3 we develop an axiomatic basis for info-gap models of uncertainty and derive a number of immediate consequences. Proofs of the theorems are collected in the appendix. Discussion is deferred until later in the paper. In sections 4 and 5 we discuss two examples of inference and decision problems with info-gap models of uncertainty. Finally, in section 6, we discuss some general questions.

2 Info-Gap Models of Uncertainty In Engineering

Convex info-gap models of uncertainty have been described elsewhere, both technically [2, 6, 13, 17] and non-technically [3]. In this section I will be very brief, and provide just a minimal intuitive framework from which to continue. I will describe three typical engineering scenarios which are incompletely described by the available information. The type of uncertainty one often faces in technological design and analysis can be described as a gap between what is known and what is not known. The quantification of this disparity leads to the info-gap model of uncertainty.

Geometrical imperfections. A thin-walled cylindrical shell, like the outer casing of a missile, is designed to have a particular nominal shape. Let $u(z, \theta)$ represent the nominal radius of the cylinder at height z and azimuthal angle θ. The shell is manufactured to conform to this shape to within a radial tolerance, α. Any shell coming off the construction line has an uncertain shape, deviating in some form from the nominal shape. $\mathcal{U}(\alpha, u)$ is the set of all shapes $f(z, \theta)$ which are consistent with the design to within tolerance α:

$$\mathcal{U}(\alpha, u) = \{ f(z, \theta) : \ |f(z, \theta) - u(z, \theta)| \ \leq \ \alpha \} , \quad \alpha \geq 0 \tag{1}$$

$\mathcal{U}(\alpha, u)$ is an info-gap model for the uncertainty in the shape of actual shells. Since the tolerance α is variable, $\mathcal{U}(\alpha, u)$ is in fact a family of nested sets: $\alpha < \beta$ implies that $\mathcal{U}(\alpha, u) \subset \mathcal{U}(\beta, u)$.

Spectral characterization of shape. In many situations the geometrical imperfections of a solid structure are characterized in terms of their waviness [22, 25, 26]. That is, shape defects are defined in terms of the amplitude and spatial wavelength of the shape-components which make them up. Shallow depressions have low spatial-frequency components, while sharp dents have both low and high spatial-frequencies. Fourier analysis is the standard tool for this representation, which leads to specifying a shape in terms of a vector of Fourier coefficients, where each coefficient is the amplitude of contribution of a particular spatial wavelength to the total shape. The typical or nominal shape has the discrete Fourier spectrum u, while the discrete spectrum f of an actual shape deviates from u. The most common info-gap model of uncertainty describes this deviation as a 'cloud' of vectors comprising an ellipsoid [29]. The info-gap model which results is:

$$\mathcal{U}(\alpha, u) = \left\{ f : \ (f - u)^T W (f - u) \leq \alpha^2 \right\} , \quad \alpha \geq 0 \tag{2}$$

where T implies matrix transposition and W is a known, real, symmetric, positive-definite matrix specifying the shape of the ellipsoid. α determines the size of the ellipsoid, which is unknown, so again $\mathcal{U}(\alpha, u)$ is a family of nested sets.

Seismic ground motion. An earthquake applies time-varying forces to buildings, bridges, and so on. Let $u(t)$ represent the temporal profile of force variation of a typical earthquake. Actual earthquakes of similar magnitude and character will deviate from the typical profile. The following info-gap model of uncertainty characterizes the range of unknown variation of the force profile in terms of the "energy of deviation" of an actual seismic event from the typical profile:

$$\mathcal{U}(\alpha, u) = \left\{ f(t) : \ \int_0^\infty [f(t) - u(t)]^2 \, \mathrm{d}t \leq \alpha^2 \right\} , \quad \alpha \geq 0 \tag{3}$$

In all three cases, eqs.(1)–(3), α is a non-negative "uncertainty parameter" whose magnitude determines the range of uncertain variation, and u is a "center point" around which the set $\mathcal{U}(\alpha, u)$ expands and contracts like a balloon as α grows and shrinks. The uncertainty parameter α is often unknown, so the info-gap model $\mathcal{U}(\alpha, u)$ is a family of nested sets. Each set has been formulated as the collection of all elements consistent with a given body of initial data. Quite often the set turns out to be convex, as in the three examples here, even though convexity is not assumed to begin with. Because of the analytical importance and practical prevalence of the property of convexity of info-gap models of uncertainty, we refer to these models as convex models of uncertainty.

3 An Axiomatization

In this section we present an axiomatic formulation of info-gap models and derive several theorems. The most important results are the "non-representation" theorems — 5 and 6 — which show that info-gap models cannot be represented in terms of Kolmogorov probability.

R is the set of the non-negative real numbers and D is a subset of a Banach space S. We will define R, D and S in this way throughout the paper. Information-gap models of uncertainty are formally represented as maps, $\mathcal{U}(\alpha, u)$, from $R \times D$ into the power set[1] of S. Thus $\mathcal{U}(\alpha, u)$ is a set-valued function, ascribing a subset of S to each point (α, u) in $R \times D$. For reasons which are evident from the examples of section 2, we refer to α as the *uncertainty parameter* and to u as the *center point* of the map. We will say that $\mathcal{U}(\alpha, u)$ is "centered" at u and has "size" or "uncertainty" α. The particular maps we are interested in are called info-gap models of uncertainty and have the following four properties:

Axiom 1 **Nesting.** *The info-gap model $\mathcal{U}(\alpha, u)$ is nested: $\alpha \leq \beta$ implies that*

$$\mathcal{U}(\alpha, u) \subseteq \mathcal{U}(\beta, u) \tag{4}$$

Axiom 2 **Contraction.** *The info-gap model $\mathcal{U}(0, 0)$ is a singleton set containing its center point:*

$$\mathcal{U}(0, 0) = \{0\} \tag{5}$$

Axiom 3 **Translation.** *Info-gap models translate linearly:*

$$\mathcal{U}(\alpha, u) = \mathcal{U}(\alpha, 0) + u \tag{6}$$

where $\mathcal{U} + u$ means that u is added to each element of $\mathcal{U}$.

Axiom 4 **Linear expansion.** *Info-gap models centered at the origin expand linearly from the origin:*

$$\mathcal{U}(\beta, 0) = \frac{\beta}{\alpha}\mathcal{U}(\alpha, 0), \quad \text{for all } \alpha,\ \beta > 0 \tag{7}$$

where $\beta\mathcal{U}$ means that β multiplies each element of $\mathcal{U}$.

As examples of info-gap models, we note that all the convex models discussed in section 2 obey the axioms of info-gap models.

The first axiom — nesting — imposes the property of 'clustering' which is characteristic of information-gap uncertainty. Axioms 2–4 specify info-gap models with particular structural properties. Some of the results we will prove depend only upon axiom 1, and we will refer to a map which obeys the first axiom as an **uncertainty map** or simply as a **U-map.**

We now develop some properties of U-maps and info-gap models. All proofs appear in the appendix.

Lemma 1 *The class of U-maps is closed under translation. That is, if $\mathcal{U}(\alpha, u)$ is a U-map and $v \in S$, then $\mathcal{U}(\alpha, u) + v$ is a U-map.*

Lemma 2 *Info-gap models of size zero are singleton sets:*

$$\mathcal{U}(0, u) = \{u\} \tag{8}$$

Lemma 3 *Info-gap models expand and translate with respect to their center points:*

$$\mathcal{U}(\beta, v) = \frac{\beta}{\alpha}\left[\mathcal{U}(\alpha, u) - u\right] + v, \quad \text{for all } \alpha,\ \beta \neq 0 \tag{9}$$

[1]We will denote the power set of a set A by $\mathcal{P}(A)$.

A single-valued transformation[2] T between two Banach spaces, S and V, is applied to a U-map $\mathcal{U}$ from $R \times D$ into $\mathcal{P}(S)$, by applying the transformation to each element in the range of $\mathcal{U}$. It is evident that the nesting property of U-maps is preserved. That is, since $\mathcal{U}(\alpha, u)$ is nested, we conclude that $\alpha \le \beta$ implies:

$$T[\mathcal{U}(\alpha, u)] \subseteq T[\mathcal{U}(\beta, u)] \tag{10}$$

This is important in the treatment of dynamic systems subject to uncertainty. If the input set, which drives the system, is a U-map, then the set of responses is also a U-map, where we identify the transformation T as the input-output relation of the system. Similarly, if the properties of the system, such as its physical parameters, are uncertain and represented by a U-map, then again the response set is a U-map, where T is the transformation between the physical properties and the response, for fixed input. It is noted that these conclusions do not require the transformation T to be linear or even continuous.

The analogous statement for info-gap models is much weaker, and refers only to linear transformations.

Theorem 1 *Let $\mathcal{U}(\alpha, u)$ be an info-gap model, mapping from $R \times D$ into the power set of S, where $D \subseteq S$. Let T be a linear transformation between Banach spaces S and V. Then $T[\mathcal{U}(\alpha, u)]$ is an info-gap model from $R \times T(D)$ to the power set of V.*

For any subset $\mathcal{U}$ of the Banach space S, the complement of $\mathcal{U}$ with respect to S is denoted $\mathcal{U}^{\rm c}$. The complement of an info-gap model is not necessarily an info-gap model. This will be important in discussing the Kolmogorov axiomatization of probability.

Theorem 2 *Let $\mathcal{U}(\alpha, u)$ be an info-gap model, mapping from $R \times D$ into $\mathcal{P}(S)$. Let $h(\cdot)$ be a function[3] from R onto R. Define the map $\mathcal{V}(\alpha, u) = \mathcal{U}^{\rm c}(h(\alpha), u)$. $\mathcal{V}(\alpha, u)$ is* not *an info-gap model.*

Lemma 4 *Let $\mathcal{U}(\alpha, u)$ be an info-gap model, mapping from $R \times D$ into $\mathcal{P}(S)$. If $\mathcal{U}(\alpha, u)$ is a convex set for some particular (α, u), $\alpha \ne 0$, then it is a convex set for all $(\beta, y) \in R \times D$.*

For any subsets $\mathcal{U}$ and $\mathcal{V}$ of S, their sum $\mathcal{U} + \mathcal{V}$ is defined as the set of all pair-wise sums of elements of $\mathcal{U}$ and $\mathcal{V}$. Let $\|\cdot\|$ denote the norm of the Banach space S. For two sets $\mathcal{U}$ and $\mathcal{V}$ in S, the Hausdorff metric is $h(\mathcal{U}, \mathcal{V}) = \max\left(\max_{u \in \mathcal{U}} \min_{v \in \mathcal{V}} \|u - v\|, \ \max_{v \in \mathcal{V}} \min_{u \in \mathcal{U}} \|u - v\|\right)$. Convergence of sets is meant in terms of the Hausdorff metric. The following result is closely related to [1].

Theorem 3 *Let $\mathcal{U}_1(\alpha, u)$ be a U-map and define the sequence of sets:*

$$U_n(\alpha, u) = \frac{1}{n!}\left[\underbrace{\mathcal{U}_1(\alpha, u) + \cdots + \mathcal{U}_1(\alpha, u)}_{n! \text{ times}}\right], \quad n = 1, 2, \ldots \tag{11}$$

$U_n(\alpha, u)$ is a U-map for each n and the sequence converges to the convex hull of $\mathcal{U}_1(\alpha, u)$.

Theorem 3 is important in suggesting a connection between convexity and uncertainty. We will return to this in section 6. Theorem 3 holds for info-gap models as well, as shown by the following result.

Theorem 4 *Let $\mathcal{U}_1(\alpha, u)$ be an info-gap model and define the sequence of sets in eq.(11). $U_n(\alpha, u)$ is an info-gap model for each n and the sequence converges to the convex hull of $\mathcal{U}_1(\alpha, u)$.*

---

[2]That is, for each $u \in S$, $T(u)$ takes a single value in V.

[3]That is, $h(\alpha)$ takes a single value for each $\alpha \in R$.

We now consider some properties of combinations of several distinct U-maps, $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$, ..., all defined on the same set of real numbers R and Banach space S. For instance these different U-maps may be differently-shaped ellipsoidal convex models, or they may be various types of energy-bound convex models. The following elementary result is presented without proof.

Lemma 5 *For any two U-maps, $\mathcal{U}(\alpha, u)$ and $\mathcal{V}(\alpha, v)$ from $R \times D$ into the power set of S, their union and intersection are each maps from $R \times D \times D$ into $\mathcal{P}(S)$ defined as:*

$$\mathcal{W}(\alpha, u, v) = \mathcal{U}(\alpha, u) \cup \mathcal{V}(\alpha, v) \tag{12}$$
$$\mathcal{X}(\alpha, u, v) = \mathcal{U}(\alpha, u) \cap \mathcal{V}(\alpha, v) \tag{13}$$

$\mathcal{W}(\alpha, u, v)$ and $\mathcal{X}(\alpha, u, v)$ obey axiom 1, that is, they are nested: $\alpha \le \beta$ implies that $\mathcal{W}(\alpha, u, v) \subseteq \mathcal{W}(\beta, u, v)$ and $\mathcal{X}(\alpha, u, v) \subseteq \mathcal{X}(\beta, u, v)$.

This can obviously be extended to the case of multiple unions and intersections.

Lemma 5 states that the class of U-maps and their unions and intersections all have the property of nesting, axiom 1. The unions and intersections of U-maps are not themselves actually U-maps because their domains of definition are more complicated.

Lemma 5 can be specialized to info-gap models in the following way.

Lemma 6 *For any two info-gap models with the same center point, $\mathcal{U}(\alpha, u)$ and $\mathcal{V}(\alpha, u)$ from $R \times D$ into the power set of S, their union and intersection are each info-gap models from $R \times D$ into $\mathcal{P}(S)$ defined as:*

$$\mathcal{W}(\alpha, u) = \mathcal{U}(\alpha, u) \cup \mathcal{V}(\alpha, u) \tag{14}$$
$$\mathcal{X}(\alpha, u) = \mathcal{U}(\alpha, u) \cap \mathcal{V}(\alpha, u) \tag{15}$$

The following "non-representation" theorem will be important in discussing the Kolmogorov axiomatization of probability in section 6. We first require a definition.

Definition 1 *A U-map from $R \times D$ into $\mathcal{P}(S)$ is* dispersed *if, for any $\alpha \in R$ and any positive integer n, there exist points $u_1, \ldots, u_n$ contained in D such that:*

$$\mathcal{U}(\alpha, u_i) \cap \mathcal{U}(\alpha, u_j) = \emptyset, \quad \text{for all} \quad i \ne j \tag{16}$$

We note that all the convex models of section 2 are dispersed, as well as many others. However, non-dispersed convex models do arise in some problems of robust stability and control of dynamical systems.[4]

Theorem 5 *Let $\mathcal{U}(\alpha, u)$ be a dispersed U-map from $R \times D$ into $\mathcal{P}(S)$. Let E be a field of sets belonging to S and containing all the sets in the range of $\mathcal{U}(\alpha, u)$. Let there be a probability function $P(\cdot)$ (consistent with the Kolmogorov axioms) defined on E such that, for each set $\mathcal{U}(\alpha, u)$, its probability is a function only of α: $P[\mathcal{U}(\alpha, u)] = h(\alpha)$. Then $h(\alpha) = 0$ for all $\alpha \in R$.*

Since the uncertainty parameter α of a U-map represents the "size" of the sets it is sometimes asserted that α is related to a probability. What theorem 5 is saying, roughly, is that if a dispersed U-map is embedded into a probability field, then neither α nor any non-negative function of α can, alone, "meaningfully" represent the probability of the U-map sets. We will discuss this further in section 6.

We can obtain an even stronger result with a slightly stronger restriction on the U-map. We first need definition 2. A *ball* of radius r and centered at point u in the Banach space S is the set $B(r, u) = \{f : \|f - u\| \le r\}$, where $\|\cdot\|$ is the norm in the space S.

[4] I am indebted to Prof. Vladimir Kharitonov for bringing examples of non-dispersed convex models to my attention [21].

Definition 2 *A U-map from $R \times D$ into $\mathcal{P}(S)$ is* unbounded *if, for any ball $B(r,u)$, there is a finite value of the uncertainty parameter α such that $B(r,u) \subset \mathcal{U}(\alpha,u)$.*

Let us note that the property of unboundedness implies, in addition to the unlimited expansiveness of the U-map, that it also has "interior" points and is not only a surface for example. The convex models of section 2 are all unbounded as well as dispersed.

Theorem 6 *Let $\mathcal{U}(\alpha,u)$ be a dispersed and unbounded U-map from $R \times D$ into $\mathcal{P}(S)$. Let E be a field of sets belonging to S and containing all the sets in the range of $\mathcal{U}(\alpha,u)$. There is* **no** *probability function $P(\cdot)$ (consistent with the Kolmogorov axioms) defined on E such that, for each set $\mathcal{U}(\alpha,u)$, its probability is a function only of α: $P[\mathcal{U}(\alpha,u)] = h(\alpha)$.*

4 Robust Inference: The 3-Box Riddle

In this section we show how an info-gap model is used to generate a decision rule based on severely deficient information. We will show that the structure of the decision rule is different from probabilistic inference.

In the 3-box riddle we will be presented with uncertain evidence and then confronted with a choice: either 'do nothing' or 'take an action'. The inference scheme by which we choose the correct option based on the evidence will be formulated as a severe test of the hypothesis, H_0, that a particular probability exceeds 1/2, against the alternative hypothesis H_1 that H_0 is false.

We will have incomplete probabilistic information, so that we are unable to use a maximum likelihood or a bayesian argument. This imperfect information will be presented to us as an info-gap model. We will develop the idea of *robust inference* [9], which is closely related to the idea of robust reliability [5, 6, 7].

4.1 3-Box Riddle

One traditional probabilistic formulation of the three-box riddle is the following. We know that a prize has been placed in one of three closed boxes. We know the probabilities, p_n, that the nth box holds the prize,[5] for $n = 1, 2, 3$. We are asked to choose a box, and if our choice is correct, we win the prize. For convenience, let us call the box we choose C. At least one of the two remaining boxes is, of course, empty. This remaining empty box we will call E. We will refer to the third box as T. Now the Master of Ceremonies (MC), who knows both our choice and the correct box, opens E, shows us that it is empty, and gives us the option of changing our choice from C to T. The question is: *do we have any rational basis for revising our choice?*

We will consider a modification of this traditional formulation, in which the probabilistic information is uncertain. We are initially given the probability model $\overline{p} = (\overline{p}_1, \overline{p}_2, \overline{p}_3)$ with $\overline{p}_1 \geq \overline{p}_2 \geq \overline{p}_3$, so our initial choice is box number 1, $C = 1$. After the MC opens box E he informs us that the initial probability model may be incorrect, and that the correct probability distribution is constrained to the set:

$$\mathcal{U}(\alpha,\overline{p}) = \left\{ (p_1,\, p_2) : \quad (p_1 - \overline{p}_1)^2 + (p_2 - \overline{p}_2)^2 \leq \alpha^2 \right\} \tag{17}$$

but he does not tell us the value of α. (The value of p_3 is determined from p_1 and p_2 by normalization.) If α is very small then the initial probability model is nearly correct. The uncertainty in the initial model rises as α increases. α is the uncertainty parameter of this info-gap model which is centered at $(\overline{p}_1, \overline{p}_2)$. We do not know the value of α, so $\mathcal{U}(\alpha,\overline{p})$ should be thought of as a family of sets: $\mathcal{U}(\alpha,\overline{p})$ for $\alpha \geq 0$. A specific value of α generates a specific set, which is another way of saying that $\mathcal{U}(\alpha,\overline{p})$ is a set-valued function of α, like all info-gap models.

[5] These probabilities are usually assumed to be equal, based on a 'principle of indifference'. We consider a more general formulation in which the probabilities may take any values.

There is uncertainty as to the correct probability model, both because we lack the correct value of α and because, unless $\alpha = 0$, there is an infinity of probability models in the set $\mathcal{U}(\alpha, \overline{p})$. The bayesians would proceed to construct a prior probability of α and of the elements of each set $\mathcal{U}(\alpha, \overline{p})$, and thus reduce the problem to one involving ordinary random variables. Another approach would be to construct lower and upper probabilities for possible events. We will take a different approach, based on what we will call the robustness of the decision, and motivated by the structure of info-gap models of uncertainty.

4.2 Robustness of the Decision

The initial probabilities are ranked as $\overline{p}_1 \geq \overline{p}_2 \geq \overline{p}_3$, so the initial decision is $C = 1$. In the classical probabilistic analysis of the 3-box riddle (with known prior probabilities) one finds that it is rational, in a maximum-likelihood sense, to change our bet from C to T if and only if $\overline{p}_1 < 1/2$. How robust is this decision? By how much can the probabilistic model vary and still keep the nominal decision correct? We will refer to the *robustness* or the *robust reliability* of the decision as: the greatest[6] value of α for which the decision would be the same if it were based on *any* probability model in the set $\mathcal{U}(\alpha, \overline{p})$. Formally, we can express the robustness of the decision as the supremum of the set of α-values for which the decision is constant:

$$\widehat{\alpha} = \sup \{\alpha : \text{ Decision} = \text{constant}, \quad \text{for all} \quad p \in \mathcal{U}(\alpha, \overline{p})\} \tag{18}$$

The decision whether or not to switch the bet from C to T will *not* be the same for all models in $\mathcal{U}(\alpha, \overline{p})$ if $\mathcal{U}(\alpha, \overline{p})$ contains a value of p_1 equal to 1/2. Eq.(17) implies that the range of p_1 values in $\mathcal{U}(\alpha, \overline{p})$ is:

$$\overline{p}_1 - \alpha \leq p_1 \leq \overline{p}_1 + \alpha \tag{19}$$

So, the least value of the uncertainty parameter α for which the decision is *not* the same for all models in $\mathcal{U}(\alpha, \overline{p})$ satisfies either $\overline{p}_1 - \alpha = 1/2$ or $\overline{p}_1 + \alpha = 1/2$. This is the robustness of the decision:

$$\widehat{\alpha} = \left| \overline{p}_1 - \frac{1}{2} \right| \tag{20}$$

For any lesser value of α the decision based on $\overline{p}_1$ is the same as the decision based on any other value of p_1 in $\mathcal{U}(\alpha, \overline{p})$. If $\alpha \leq \widehat{\alpha}$ then the decision based on $\overline{p}$ is the same regardless of which probability model in $\mathcal{U}(\alpha, \overline{p})$ actually occurs. If $\widehat{\alpha}$ if large, then the decision based on $\overline{p}$ will be correct even if the correct distribution deviates greatly from $\overline{p}$.

The robustness of the decision is small or zero when $\overline{p}_1$ is near or equal to 1/2: even small deviation of the nominal from the actual model could result in an erroneous decision. On the other hand, when $\overline{p}_1$ is very different from 1/2 then the actual model could vary by quite a bit without altering the validity of the decision concerning revision of the bet.

4.3 The Inference as a Severe Test

The decision whether or not to change the bet from box C to box T can be viewed as a choice between two hypotheses:

$$H_0 : \qquad p_1 > 1/2 \tag{21}$$

$$H_1 : \qquad p_1 \leq 1/2 \tag{22}$$

H_0 asserts that box 1 (which is chosen initially as box C) contains the prize with a probability in excess of 1/2, while H_1 asserts the opposite. If H_0 is true then the maximum likelihood argument

[6] Actually, by "greatest" we mean "least upper bound".

leads to keeping the bet on C, while if H_1 is true one changes the bet to T in order to maximize the chances of winning. These are mutually exclusive and jointly comprehensive: only one can be true; at least one must be true. Consequently, it is sufficient to test (accept or reject) only one of them.

We are subjecting H_0 to a *severe test* if an erroneous inference concerning the truth of H_0 can result only under extraordinary circumstances. Mayo has developed an account of scientific inference based on learning from error and severe tests of hypotheses which, while employing probabilistic thinking, is quite distinct from the classical approach of assigning a posterior probability to an hypothesis. Mayo suggests the following "underlying thesis": [28, p.445]:

> It is learned that an error is absent to the extent that a procedure of inquiry with a high probability of detecting the error if and only if it is present nevertheless detects no error.

If we have sufficient probabilistic information we would severely test H_0 by requiring the probability of failing to detect an error in H_0 to be exceedingly small. This is precisely the approach underlying classical statistical testing of hypotheses. If we do not have the requisite probabilistic information we can nonetheless proceed with a small modification of the probabilistic severe-test idea. Changing only a few words in Mayo's formulation we have:

> It is learned that an error is absent to the extent that a procedure of inquiry with high *robustness* for detecting the error if and only if it is present nevertheless detects no error.

The robust implementation of the severe test is:

> Accept H_0 if $\overline{p}_1 > 1/2$ and if the robustness $\widehat{\alpha}$ is large.

We accept H_0 if p_1 (the unknown true value) could deviate by a large margin from $\overline{p}_1$ (the nominal value given as evidence) without altering the validity of H_0. This test would fail to detect an error in H_0 only in the 'extraordinary' circumstance that $\overline{p}_1$ is 'extraordinarily far' from correct, since a large value of $\widehat{\alpha}$ means that even large error in $\overline{p}_1$ does not jeopardize the truth of the decision based on H_0.

The question now is, what does 'large' mean? How large a value of $\widehat{\alpha}$ is large enough to vindicate H_0 in the severe test? We somehow know what 'high probability' means in Mayo's formulation, or at least we believe that in any given application the protagonists will be able to decide to their own satisfaction how probable is probable enough: 0.9, 0.999, 0.99999, etc. They will use their judgement tempered by experience to decide what magnitude of probability is acceptable and what is not.

Similarly, one can calibrate $\widehat{\alpha}$ by relating its values to quantities whose acceptability or unacceptability is known from experience. This is not necessarily a simple judgement to make, nor can it be done in only one unique fashion with one unique answer. In chap. 9 of [6] I discuss two alternative approaches, one based on evaluating the severity of the decision consequences, the other on comparing differing degrees of initial uncertainty.

In the present example one can calibrate $\widehat{\alpha}$ in the following simple manner. First let us define the quantity:

$$\alpha_{\max} = \max\{\overline{p}_1,\, 1 - \overline{p}_1\} \tag{23}$$

Referring to eq.(17) or (19) one sees that, if $\alpha \geq \alpha_{\max}$ then $\mathcal{U}(\alpha, \overline{p})$ in fact does not constrain the value of p_1 at all since p_1, being a probability, must lie in the interval from zero to one. $\alpha_{\max}$ is in this sense an upper limit of meaningful values of the uncertainty parameter.

Now let us consider two extreme situations: (1) $\widehat{\alpha} = 0$ and (2) $\widehat{\alpha} > \alpha_{\max}$. The first case corresponds to *complete fragility* of the decision: the slightest deviation of p_1 from $\overline{p}_1$ entails the possibility of an error in the decision based on $\overline{p}_1$. The second case represents *complete robustness:* no deviation of p_1 from $\overline{p}_1$ can jeopardize our conclusion based on $\overline{p}_1$. The range of α-values from 0 to $\alpha_{\max}$ provides an intuitive or qualitative scale for calibration of the robustness. In any particular application we are given the value of $\overline{p}_1$ and we calculate the robustness $\widehat{\alpha}$ from eq.(20). If $\widehat{\alpha}$ is much nearer to zero than to $\alpha_{\max}$ we consider the decision as 'fragile' rather than 'robust', while if $\widehat{\alpha}$ is much closer to $\alpha_{\max}$ than to zero our judgement is the reverse. Our answer to the question 'How large an $\widehat{\alpha}$ is large enough' is: $\widehat{\alpha}$ near $\alpha_{\max}$ is 'large'; $\widehat{\alpha}$ near zero is 'small'. We are now in a position which is fairly analogous to where the probabilistic argument puts us when we say: a probability of no-failure near unity is 'large', while a probability of no-failure near zero is 'small'.

4.4 Discussion

The 'evidence' upon which our decision is based is the nominal or initial probability distribution, $\overline{p} = (\overline{p}_1, \overline{p}_2, \overline{p}_3)$. In addition, we know that these probabilities are uncertain, and that their available variations are constrained by the info-gap model $\mathcal{U}(\alpha, \overline{p})$, eq.(17). This uncertainty model represents the uncertainty in the probabilistic evidence in the sense of an *information gap*. $\mathcal{U}(\alpha, \overline{p})$ expresses the gap between the explicit evidence which we have, $(\overline{p}_1, \overline{p}_2, \overline{p}_3)$, and the other possibilities which may in fact occur. Alternatively, we can view the non-probabilistic uncertainty expressed by $\mathcal{U}(\alpha, \overline{p})$ as a geometrical expression of how the available alternatives *cluster* around the evidence. Our evidence is $(\overline{p}_1, \overline{p}_2, \overline{p}_3)$, and $\mathcal{U}(\alpha, \overline{p})$ tells us how the actual probabilities can arrange themselves around this nominal value. Both of these interpretations are non-probabilistic. Both hinge on α as an uncertainty parameter which orders the nesting of the sets in the info-gap model $\mathcal{U}(\alpha, \overline{p})$, $\alpha \geq 0$. Both interpretations are consistent with the axioms of section 3.

Now we consider the relation between the robust inference scheme and bayesian decision theory as well as statistical hypothesis testing.

In the bayesian approach to multi-hypothesis decision-making one combines the evidence with prior probabilities of the hypotheses and with probabilities of the evidence conditioned on the hypotheses. This data base is used either to minimize a cost function or to calculate conditional probabilities of the hypotheses based on the evidence.

Non-bayesian statistical hypothesis testing avoids the use of prior probabilities and uses the probabilities of the evidence conditioned on each hypothesis. When testing binary hypotheses, one uses the level of significance — the probability, conditioned on the null hypothesis, of obtaining a result more extreme than the evidence — to severely test the null hypothesis in Mayo's probabilistic sense. This is extended in more complex hypothesis tests by means of the Neyman-Pearson lemma and related results concerning the power of a test.

In the robust inference scheme for binary hypotheses, the robustness is a function of the evidence which indicates how large a change in the evidence is needed to cause the alternative hypothesis to be preferred over the null hypothesis. More precisely, the robustness is the greatest value of the uncertainty parameter for which the decision is the same for all possible evidence in the info-gap model.

While non-bayesian statistical hypothesis testing jettisons the prior probabilities which the bayesians rely upon, the robust inference scheme abandons the conditional probabilities as well. However, the robust inference scheme remains structurally somewhat similar to statistical testing, once we replace the level of significance by the robustness, $\widehat{\alpha}$. In statistical testing of binary hypotheses, H_0 is accepted if the level of significance is large, indicating that the evidence is neither extreme nor unusual with respect to H_0 and that a vast amount of more extreme evidence could have been obtained which would also be compatible with H_0. In robust inference, H_0 is accepted if the robustness, $\widehat{\alpha}$, is large, indicating that a vast amount of other evidence might have been obtained which is also consistent with H_0 and that only an extraordinarily large change in the evidence would alter the decision.

Further discussion of the relation between robust inference and statistical hypothesis testing, specifically in the context of sequential tests, is found in [11].

There is an additional parallel between Mayo's severe test procedure and the robust severe test. Mayo emphasizes inference as a process of learning from error, as a sequence of tests of the experimental procedures, of the working hypotheses, of the underlying models and assumptions. Mayo emphasizes evaluating the details of the method, rather than obtaining a posterior probability that a given final conclusion is true. This is characteristic of robust severe testing as well, as illustrated in detail by the seismic example discussed in section 5.

The "non-representation" theorems, 5 and 6, would seem to preclude an attempt to formulate the robust decision, based on the uncertainty parameter of an info-gap model, as a special case of statistical inference.

5 Robust Reliability: A Seismic Example

In this section we will consider a mechanical engineering design decision based on evidence, part of which is uncertain and represented by an info-gap model of uncertainty. The purpose is to illustrate the design procedure in which a decision based on uncertain data is made by a variation of the robust inference scheme described in section 4. The example will be tremendously simplified in order to succinctly exhibit the main lines of reasoning. The reader is asked to bear with me in my approximations, and to accept that the gist of the analysis is realistic.

One of the grand and hoary problems of structural engineering is to construct a building which won't collapse, come what may. Hammurabi imposed terrible penalties upon a builder whose handiwork falls down [6, p.215] and the attitude has remained that structural collapse, even under severe conditions, is unacceptable. Nevertheless, the 1995 earthquake in Los Angeles demonstrated that even four millenia after Hammurabi we are far from proficient. We will illustrate a design analysis for seismic reliability both because this is an important area of research in engineering and because the consequences of the decision are quite significant. We will show that the design is an inference based on uncertain evidence.

We perform the reliability analysis in section 5.1 and then describe the design procedure in section 5.2.

5.1 Reliability Analysis

Buildings fail during earthquakes due to the swaying induced by the ground motion. The stiffer the building, the lower is the amplitude of the sway, so in our simple analysis we will assume that 'stiffer is safer'.[7] The question is: how stiff should we make the building? That depends on the magnitude and nature of the seismic event, which is uncertain.

The engineering analysis of reliability entails three components: (1) a model of the system, (2) a model of the uncertainties, and (3) a criterion for failure. After outlining these components we will define exactly what is meant, here, by reliability, and then consider the design analysis as an inference problem in the next subsection.

System model. The building will be modelled as a one-dimensional undamped linear harmonic oscillator. That is, the displacement of the building with respect to the ground is represented by the following differential equation:

$$m\ddot{x}(t) + kx(t) = f(t) \tag{24}$$

where the dots imply differentiation with respect to time t, m is the mass of the structure and k is its stiffness. The force exerted on the structure by the seismic motion of the earth is $f(t)$. The initial conditions are $x(0) = \dot{x}(0) = 0$, meaning that the building is at equilibrium and motionless when the earthquake begins. The solution of eq.(24) is:

$$x_f(t) = \frac{1}{m\omega}\int_0^t f(\tau)\sin\omega(t-\tau)\,\mathrm{d}\tau \tag{25}$$

[7] This is the most glaring technical simplification which I will make, since I am ignoring the possibilities of passively damping and actively dissipating the seismic energy [30]. However, even if these considerations were included, the structure of the argument would remain formally the same, though it would become substantially more complex.

where $\omega = \sqrt{k/m}$ is the 'natural' or 'resonant' frequency of vibration of the building.

Uncertainty model. The uncertainty in the seismic load, $f(t)$, is represented by the info-gap model of uncertainty $\mathcal{U}(\alpha, u)$ in eq.(3). $\mathcal{U}(\alpha, u)$ is the set of seismic load functions, $f(t)$, whose "energy of deviation" from the function $u(t)$ does not exceed α^2. The function $u(t)$ is the known 'nominal' or 'anticipated' or 'design-base' seismic load. α is the uncertainty parameter, whose magnitude determines the disparity between $u(t)$ and what can really occur. When α is small, all seismic events are essentially the same as $u(t)$; as α grows, the gap between what is known, the typical result $u(t)$, and the possible unknown future outcome, $f(t)$, grows as well. We assume that we know the nominal load function, $u(t)$. However we definitely do not know a specific value for the uncertainty parameter, α, which can take any non-negative value. In other words, the uncertainty in the seismic load is not represented by a specific set, but rather by the family of nested sets $\mathcal{U}(\alpha, u)$ where u is known but $\alpha \geq 0$. As we know from section 2, $\mathcal{U}(\alpha, u)$ is a convex info-gap model with uncertainty parameter α and center point $u(t)$.

Failure criterion. The building will fail if and only if its displacement with respect to the ground exceeds a known critical value:

$$|x(t)| \geq x_{\text{cr}} \tag{26}$$

The stiffness of the structure is expressed by the coefficient k in eq.(24), and this is the parameter whose value must be chosen by the designer to assure that the building is reliable. Given a proposed design (a value of k) we cannot determine whether or not the structure will fail because we do not know $f(t)$ (the earthquake motion) and hence we cannot calculate $x_f(t)$. The uncertainty in $f(t)$ prevents us from directly applying failure criterion (26) to test whether or not k is acceptable. We cannot even perform a "worst-case" analysis, seeking the most damaging load function, since the uncertainty parameter α is also unknown, and the family of sets, $\mathcal{U}(\alpha, u)$ for $\alpha \geq 0$, contains loads of unbounded severity. What are we to do in order to design a reliable structure?

To rely on something means "to have confidence based on experience" [32]. Traditionally (over the past forty years or so) engineers have quantified this intuitive lexical definition of reliability in terms of probability: the reliability of something is the probability that it will not fail. Classical reliability theory is thus based on the mathematical theory of probability, and depends on knowledge of probability density functions of the uncertain quantities. However, in the present situation we cannot apply this quantification of reliability because our information is much too scanty to verify a probabilistic model. The info-gap model tells us how the unknown seismic loads cluster and expand with increasing uncertainty, but it tells us nothing about their likelihoods. (For comparison of probabilistic and info-gap models of uncertainty in seismic applications see [12]; in mechanical shell buckling see [25, 26].)

The idea of robustness or immunity to uncertainty provides us with an alternative quantification of reliability which, on the one hand is satisfactorily close to the intuitive idea of reliability and, on the other hand can be implemented with the available information. We will say the building is reliable if it will not fail even in the presence of great uncertainty. Conversely, an unreliable structure can fail even with small uncertainty. In other words, a reliable system is robust or immune to uncertainty, while unreliability means fragility or vulnerability to deviation from nominal conditions. This concept of reliability is (not unexpectedly) well suited to implementation with info-gap models. This non-probabilistic theory of reliability is developed in [4, 5, 6, 7].

For fixed uncertainty parameter α, the 'response set' is the collection of all values taken by $x_f(t)$ as $f(t)$ varies over the input set $\mathcal{U}(\alpha, u)$:

$$\mathcal{X}(\alpha, x_u) = \{x(t): \ x(t) = x_f(t), \quad \text{for all} \quad f(t) \in \mathcal{U}(\alpha, u)\} \tag{27}$$

The input/output relation for the system of eq.(24) is linear so, in light of theorem 1, the response set $\mathcal{X}(\alpha, x_u)$ is an info-gap model. Consequently, $\mathcal{X}(\alpha, x_u)$ is nested with respect to α. The robust reliability is the greatest value of α for which $\mathcal{X}(\alpha, x_u)$ does not include the critical value for failure,

$x_{\rm cr}$. Formally, the robustness is:

$$\widehat{\alpha} = \sup \{ \alpha : \ \mathcal{X}(\alpha, x_u) \ \cap \ \{x_{\rm cr}\} = \emptyset \} \tag{28}$$

(We are assuming that the structure will not fail with the nominal load, so that $|x_u| < x_{\rm cr}$.)

With a modest effort (which the reader will be spared) one finds the following expression for the robustness:

$$\widehat{\alpha}(t) = \frac{m\omega}{\sigma(t)} \left[x_{\rm cr} - |x_u(t)| \right] \tag{29}$$

where we have defined $\sigma(t) = \sqrt{\int_0^t \sin^2 \omega t \, {\rm d}t}$. The robust reliability depends on the design parameter k which appears in ω, x_u and σ, it depends on the nominal response x_u, on the critical value for failure $x_{\rm cr}$, and on the time t. The robust reliability $\widehat{\alpha}$, eq.(28), exploits the nested structure of the info-gap model $\mathcal{X}(\alpha, x_u)$, which is centered at the nominal response x_u, in order to determine the unique least upper bound of uncertainty-parameter values for which the structure satisfies the no-failure condition. The critical features of $\mathcal{X}(\alpha, x_u)$ are nesting and centering; that $\mathcal{X}(\alpha, x_u)$ is also convex and expands linearly makes the analytical evaluation of $\widehat{\alpha}$ fairly easy (also in more complex and realistic formulations [10]).

5.2 Design Decision as a Robust Inference

The designer wants to choose the stiffness coefficient k so that the structure is very robust with respect to the earthquake load. He wants a stiffness which will enable the building even to resist earthquakes which are widely different from the nominal design-base load $u(t)$. That is, k should be chosen so that $\widehat{\alpha}$ is large.[8] We can understand this design procedure as a robust inference in the sense of section 4.3, as we now explain.

Formally what the designer is doing is making a binary decision between two hypotheses:

$$H_0 : \quad \text{The structure will not fail.} \tag{30}$$

$$H_1 : \quad \text{The structure will fail.} \tag{31}$$

The designer must use the reliability analysis to severely test H_0. The evidence upon which the reliability analysis rests is the numerical data and physical understanding of the system and its environment: the dynamic model eq.(24), the uncertainty model (3), and the failure criterion (26). An error is absent in H_0 to the extent that the reliability analysis has determined that the structure is robust to uncertainty in the seismic load. It may be that some extraordinary seismic event could cause collapse of the building. However, if $\widehat{\alpha}$ is large then eq.(28) assures us that the given structure will survive every earthquake within a large set centered at the design-base earthquake. In other words, if $\widehat{\alpha}$ is large, then we can claim that we have sought in vain for a fatal earthquake by a procedure which will detect this fatality if and only if it is present unless it is extraordinarily different from the nominal.

6 Discussion

Are info-gap models probabilistic models in disguise? We will discuss a number of differences between the Kolmogorov axiomatization [23] and the properties of info-gap models of uncertainty.

(1) Both probability and info-gap models represent "events" as subsets of a universal set, E, but here the similarity ends. In Kolmogorov's axiomatization no constraint is placed on E, and the

[8]The calibration of $\widehat{\alpha}$ on a qualitative scale from 'small' to 'large' is an important technical problem which is, however, not relevant to our discussion of the designer's decision as a robust inference. We discussed this in section 4. See also [6, chap.9].

axioms begin by requiring the relevant class of subsets to be a field. In info-gap models we begin by requiring E to be a Banach space (which we have denoted by S) but the class of subsets of interest do not form a field. Indeed, the 'chronology' of the two axiom systems is different. Kolmogorov defines the underlying event space: a field of subsets which, by definition, is closed under intersection, union and difference of sets. Upon this field is then defined the probability function with its properties. In info-gap models the underlying set has the structure of a Banach space (complete, normed, linear space), but then a function is immediately postulated which, according to its properties, determines the class of relevant sets.

It is easy to see that the collection of sets defined by the range of an info-gap model need not form a field. Consider for instance the info-gap model of eq.(2), whose range is made up of ellipsoids. Unions and intersections of ellipsoids are not ellipsoids, so this class of sets is not a field.

The 'field' requirement of probability is a fundamental aspect of the structure of the event space. It is postulated that when events A and B can occur then other related events, their union, intersection and difference, also exist in the universe of possibilities and their probabilities are related to the probabilities of A and B in certain specific ways. Not so in info-gap models, which 'organize' the universe of possibilities differently. Theorem 2 shows that the family of sets defined by an info-gap model is not closed under complementation. For example, in the info-gap model of eq.(2) every element in S can 'occur', since it belongs to an infinity of info-gap models, for sufficiently large size parameters α. But the relation between one event and others does not have the 'field' logic of conjunction and disjunction, but rather the 'cluster' logic of expansion.

(2) The uncertainty parameter α of an info-gap model is a "size" parameter in the sense that the sets $\mathcal{U}(\alpha, u)$ "grow" with α. Therefore it is sometimes felt that α is just a probability measure in disguise. α itself varies over the non-negative real numbers so it does not obey the normalization requirement, but some people have suggested that this can be corrected by a suitable re-scaling of α. The fundamental question is whether α can be calibrated to have the disjoint additivity property of probability functions. We must consider whether the additivity (or nonadditivity) of α is 'essential' or 'inessential' in the sense discussed by Luce *et al* [27, pp.18–20]. α is essentially nonadditive if every re-scaling of α is nonadditive.

There is no doubt that one can define a probability function P on the range of an info-gap model or of a U-map, though to satisfy the field requirement of probability one needs to extend the collection of sets, upon which P is defined, to include unions, intersections and differences. However, theorem 5 states that any such probability function, if its restriction to $\mathcal{U}(\alpha, u)$ is only a re-scaling of α, will give zero probability to all the sets in the range of the U-map (if the U-map is dispersed). In other words, α cannot be meaningfully calibrated as the probability of a U-map; α is 'essentially' nonadditive. If the U-map is both dispersed and unbounded then theorem 6 is an even stronger result: no probability function exists at all which, on the range of the U-map, depends only on the uncertainty parameter.

From these results it is clear that α is not simply probability dressed in another garb.

Do info-gap models of uncertainty really treat uncertainty? "The concept of indeterminacy" wrote Suppes and Zanotti "is a concept for those who hold that not all sources of error, lack of certain knowlege, etc., are to be covered by a probability distribution, but may be expressed in other ways ..." [31, p.434]. Those authors proceed to study random relations and upper and lower probabilities. Info-gap models are an additional alternative.

Once we accept probability as an axiomatized mathematical theory we are recognizing that mathematical probability is not a fundamental physical or psychological phenomenon. What is fundamental, and what motivates mathematical probability, is the panoply of phenomena of uncertainty, indeterminacy, doubt and inaccuracy in all their myriad forms which confront us in our daily, and not so daily, lives.

The situation is similar to the relation between physical space and the various mathematical geometries which describe spatial attributes. The intuitive reasonableness of Euclid's postulates does not preclude alternative mathematical geometries which also have 'reasonable' interpretations. A century ago Riemann showed that modification of Euclid's 'parallel' postulate need not lead to logical inconsistency. What is more, in this century some of the non-euclidean geometries have found their place in general relativity and cosmology. A closely analogous situation is found in the various 'possibility' theories [16] in which the additivity axiom of probability is modified in various ways. Logical consistency is retained, as well as interpretations of the resulting theories in ways which correspond quite acceptably to various intuitive ideas of uncertainty.

Info-gap models of uncertainty are based on axioms which are much further from Kolmogorov's probability than are the fuzzy-logic possibility theories. Do info-gap models deal with uncertainty? The examples of sections 4 and 5, as well as many others from a range of engineering fields, seem to me to suggest an incontrovertible 'yes'. One could *define* uncertainty as only what is treated probabilistically. However, that would be a rather arbitrary and narrow definition. The experience of a large populace of men and women testifies to the utility of these logically independent methodologies for quantifying, interpreting and exploiting uncertain evidence. Much technological experience strongly indicates the need for a plurality of axiomatically distinct tools for handling uncertainty.

Robust severe inference. We have formulated inference schemes, based on robust severe tests, in the examples of sections 4 and 5. Why can one accept an inference based on a robust severe test? What is the intuition which makes such an inference plausible?

Specific examples of inference schemes tend to have distinctive personalities which can mask the underlying procedure so, before addressing this problem, I will re-iterate the main features of robust severe inference.

1. To make a *robust severe inference* about the truth of an hypothesis H_0, means to accept or reject H_0, against a competing hypothesis H_1, by subjecting it to a *robust severe test,* which we defined in section 4.3.

2. The implementation of a *robust severe test* in connection with info-gap models of uncertainty is based on the evaluation of a *robustness* $\widehat{\alpha}$ which quite often, and in both our examples, is the greatest value of the uncertainty parameter subject to some constraint involving info-gap models. The constraint is derived from the specific formulation of the problem. (See eqs.(18) and eq.(28) as well as [8]). The robustness can be succinctly expressed as:

$$\widehat{\alpha} = \sup \{\alpha : \quad \text{Constraint on } \mathcal{U}(\alpha, u)\} \tag{32}$$

3. H_0 passes the robust severe test if $\widehat{\alpha}$ is "large".

4. The calibration of $\widehat{\alpha}$ from "small" to "large" can be done in a variety of ways, which we have discussed only briefly. See section 4.3 and [6, chap.9].

One justification for accepting a robust inference is a prior belief in the *continuity of cause and effect:* distant effects come from distant causes, while small changes in the cause produce small changes in the effects. This is a pebble-splash idea: if you see a ripple, a pebble must have hit the water nearby. The calibration of what is 'nearby' is provided by the robustness. A robust severe test accepts an hypothesis (the 'cause') if it is insensitive to large variation of the evidence (the 'effect'). A robust severe test calibrates the 'distance' of evidence: even evidence quite different from the evidence in hand would arise from the same hypothesis (if it has passed a robust severe test), so that only evidence far beyond the pale of experience would arise from the competing hypothesis. If the competing hypothesis were true, we would have gotten evidence greatly different from what was observed.

A somewhat analogous intuition can be enlisted to support the plausibility of the probabilistic severe test. Mayo has analyzed at great length the foundations for probabilistic severe tests, and I do not intend to re-iterate any of her arguments. Rather, it is interesting that an apparently different justification of inference based on a probabilistic severe test can be constructed by a continuity argument which is analogous to the continuity of cause and effect. Probabilistic inference is plausible if one holds a prior belief in the *continuity of cause and frequency of effect:* events which are rare with respect to one cause can be caused frequently by distant causes, while small changes in the cause produce small changes in the likelihoods. This again is a pebble-splash idea: if you see a large ripple, the pebble must have hit the water nearby. The calibration of what is a 'large' ripple is done in terms of probability.

Neither of these continuity justifications can be viewed as incontrovertible. One could imagine discontinuities between cause and effect, in which greatly disparate causes mimic each other's effects. Similarly, one can imagine discontinuities in conditional probabilities, so that large changes in the cause (the conditioning variable) produce small changes in the probability distribution. The result would be that what is frequent with one cause becomes frequent with a greatly different cause as well. Nonetheless, both of these continuity intuitions seem "reasonable", though why one might accept one or both of them is beyond the scope of this paper.

One point to stress is that, though the intuitions which underlie robust and probabilistic severe tests are distinct, they are similar. This similarity may be responsible for the impression among some people that the tests themselves are the same, or that one can be reduced to the other. This we know to be false, since the axiomatic bases of probabilistic and info-gap models of uncertainty are quite different.

Also, it is important to realize that these two continuity intuitions are not mutually exclusive. One can hold to both simultaneously, just as one can adhere to two distinct axiomatic systems such as those of probability and info-gap models, or euclidean and lobachevskian geometries.

Finally, the cause-effect continuity has no reference to likelihood or recurrence-frequency (though its similarity to the cause-frequency continuity may obscure this). The cause-effect continuity is very characteristic of info-gap modelling, and is based on a nested view of the universe of possibilities. Robust severe testing is rooted in the interpretation of uncertainty in terms of the geometrical ideas of clustering and information gap which we discussed in section 4.4.

Convexity and uncertainty. Almost all info-gap models of uncertainty used in practice employ *convex* sets. When the practitioner defines an uncertainty model as the set of all elements consistent with a given body of evidence, as we did in section 2, quite often the set turns out to be convex. Theorems 3 and 4 can be interpreted as stating that if a set of "macroscopic" events is formed as the superposition of many "microscopic" events, then it will tend to be convex. (See [13] for discussion of the analogy between theorems 3 and 4 and the central limit theorem.) Convexity of an info-gap model of uncertainty can thus arise as a by-product of the complex combination of underlying processes. A good example of this is the convexity of response sets in the assay of spatially random material, where the response to an arbitrary spatial distribution of analyte is the linear superposition of an infinity of point-source responses [2, p.25]. Convexity of info-gap models of uncertainty is the rule in practice, but convexity need not be included as an axiom by virtue of these theorems, if one is willing to allow the convexity to arise asymptotically. For more discussion see [3].

Appendix: Proofs

Proof of lemma 1. Since $\mathcal{U}(\alpha, u)$ is a U-map it is evident that $\mathcal{U}(\alpha, u) + v$ is nested. ∎

Proof of lemma 2. Direct application of contraction and translation, axioms 2 and 3. ∎

Proof of lemma 3. By axiom 4: $\mathcal{U}(\beta, 0) = \frac{\beta}{\alpha}\mathcal{U}(\alpha, 0)$. Axiom 3 implies $\mathcal{U}(\beta, v) = \mathcal{U}(\beta, 0) + v$ and $\mathcal{U}(\alpha, 0) = \mathcal{U}(\alpha, u) - u$ are U-maps. Relation (9) results by substitution. ∎

Proof of theorem 1. A linear transformation T between Banach spaces S and V has the property that $T(\beta u + \gamma v) = \beta T(u) + \gamma T(v)$ where β and γ are numbers and u, $v \in S$. When we apply T point-wise to the info-gap model $\mathcal{U}$ we can represent it as a map from $R \times T(D)$ into $\mathcal{P}(V)$

because linear transformations are single-valued functions. That is: $T[\mathcal{U}(\alpha,u)] = \mathcal{W}(\alpha,T(u)) \subseteq V$. We must show that the four axioms of info-gap models hold for the transformed map $\mathcal{W}$. (1) Nesting, axiom 1, holds by virtue of eq.(10). (2) Contraction, axiom 2: For a linear transformation, $T(0) = 0$, so $T[\mathcal{U}(0,0)] = \{0\}$. (3) Translation, axiom 3: From translation of $\mathcal{U}$ and linearity of T one finds that $T[\mathcal{U}(\alpha,u)] = T[\mathcal{U}(\alpha,0)+u] = T[\mathcal{U}(\alpha,0)] + T(u)$. But $T[\mathcal{U}(\alpha,u)] = \mathcal{W}(\alpha,T(u))$ and $T[\mathcal{U}(\alpha,0)] = \mathcal{W}(\alpha,0)$. So: $\mathcal{W}(\alpha,T(u)) = \mathcal{W}(\alpha,0) + T(u)$. (4) Expansion, axiom 4: From expansion of $\mathcal{U}$ and linearity of T one finds that $T[\mathcal{U}(\beta,0)] = T\left[\frac{\beta}{\alpha}\mathcal{U}(\alpha,0)\right] = \frac{\beta}{\alpha}T[\mathcal{U}(\alpha,0)]$. So: $\mathcal{W}(\beta,0) = \frac{\beta}{\alpha}\mathcal{W}(\alpha,0)$. ■

Proof of theorem 2. The "easiest" info-gap model axiom to violate universally is contraction. $\mathcal{V}(0,0) = \mathcal{U}^{\mathrm{c}}(h(0),0)$. By nesting (axiom 1) and contraction (axiom 2), $0 \in \mathcal{U}(h(0),0)$. Thus $0 \notin \mathcal{V}(0,0)$ which violates axiom 2. ■

Proof of lemma 4. Info-gap models of size zero are singleton sets and thus convex, so we need only consider sets of positive size. We must show that, for any $v,\, w \in \mathcal{U}(\beta,y)$, the convex combination $\gamma v + (1-\gamma)w$ also belongs to $\mathcal{U}(\beta,y)$, for any $0 \le \gamma \le 1$. Since $\mathcal{U}$ is an info-gap model, lemma 3 implies that:

$$\frac{\alpha}{\beta}(v-y)+u \;\in\; \mathcal{U}(\alpha,u), \quad \frac{\alpha}{\beta}(w-y)+u \;\in\; \mathcal{U}(\alpha,u) \tag{33}$$

The convexity of $\mathcal{U}(\alpha,u)$ implies:

$$\gamma\left[\frac{\alpha}{\beta}(v-y)+u\right] + (1-\gamma)\left[\frac{\alpha}{\beta}(w-y)+u\right] \;\in\; \mathcal{U}(\alpha,u) \tag{34}$$

After simplifying and again applying lemma 3 one finds $\gamma v + (1-\gamma)w \in \mathcal{U}(\beta,y)$. ■

Proof of theorem 3. (1) $\mathcal{U}_1$ is a U-map so it is nested, (axiom 1). It is thus elementary to show that $n!$-fold combinations of $\mathcal{U}_1$ are also nested, satisfying axiom 1. The proof is omitted.

(2) Let $\mathcal{U}(\alpha,u)$ denote the convex hull of $\mathcal{U}_1(\alpha,u)$. To prove convergence we must show that, for any $\varepsilon > 0$, there is an N_ε such that $h(\mathcal{U}_n,\mathcal{U}) < \varepsilon$ for all $n > N_\varepsilon$.

(2.1) Since $\mathcal{U}_n$ is the superposition of all $n!$ combinations of elements of $\mathcal{U}_1$, the sequence of sets is nested: $\mathcal{U}_n \subseteq \mathcal{U}_{n+1}$. The proof: An arbitrary element of $\mathcal{U}_n$ is $\frac{1}{n!}\sum_{m=1}^{n!} y_m$, for $y_1,\ldots,y_{n!} \in \mathcal{U}_1$. But this equals $\frac{1}{(n+1)!}\sum_{m=1}^{n!}(n+1)y_m$ which belongs to $\mathcal{U}_{n+1}$.

(2.2) The convex hull of a set A is the union of the convex spans of all finite subsets of A [19, p.200]. Thus any convex combination of any finite number of elements of $\mathcal{U}_1$ is an element of $\mathcal{U}$. Hence each $\mathcal{U}_n$ belongs to $\mathcal{U}$. Consequently, $\max_{x\in\mathcal{U}_n}\min_{y\in\mathcal{U}}\|x-y\| = 0$. Thus the Hausdorff metric becomes $h(\mathcal{U}_n,\mathcal{U}) = \max_{y\in\mathcal{U}}\min_{x\in\mathcal{U}_n}\|x-y\|$.

(2.3) Consequently, (2.1) implies that $h(\mathcal{U}_n,\mathcal{U}) \ge h(\mathcal{U}_{n+1},\mathcal{U})$, so to prove convergence it is sufficient to show that, for any $\varepsilon > 0$, there is an n such that $h(\mathcal{U}_n,\mathcal{U}) < \varepsilon$.

(2.4) Any $y \in \mathcal{U}$ can be represented as $y = \sum_{m=1}^{M} c_m y_m$ for some M where $y_1,\ldots,y_M \in \mathcal{U}_1$ and $c_1,\ldots,c_M$ are convex coefficients (non-negative and summing to one). Also, for any $n! \ge M$ one can choose $x \in \mathcal{U}_n$ as: $x = \sum_{m=1}^{M} r_m y_m$ where $r_1,\ldots,r_M$ are rational convex coefficients. Thus $\|x-y\| = \left\|\sum_{m=1}^{M}(r_m-c_m)y_m\right\| \le \mu\sum_{m=1}^{M}|r_m-c_m|$ where $\mu = \max_m\|y_m\|$. It is evident that, by choosing n large enough, one can make the rational coefficients r_m arbitrarily close to the real coefficients c_m, thus satisfying the convergence criterion. For example, let $\rho_m = p_m/q_m$, $m = 1,\ldots,M$, be rational convex coefficients such that, for each m, $|\rho_m - c_m| \le \varepsilon/\mu M$. Choose $n! = \left(\prod_{m=1}^{M} q_m\right)!$. It is now possible to choose each $r_m = k_m/n!$ where $k_m = n!p_m/q_m$ so $r_m = p_m/q_m$. Hence $\mu\sum_{m=1}^{M}|r_m-c_m| < \varepsilon$. ■

Proof of theorem 4. (1) The proof of convergence is the same as for theorem 3 and will not be repeated. (2) Nesting: $\mathcal{U}_n$ is nested by theorem 3. (3) Contraction: $\mathcal{U}_1(0,0) = \{0\}$ by axiom 2, so $n!$-fold combinations of $\mathcal{U}_1(0,0)$ also equal $\{0\}$. Hence $\mathcal{U}_n(0,0)$ obeys axiom 2. (4) Translation:

$$\frac{1}{n!}\left[\underbrace{\mathcal{U}_1(\alpha,0)+\cdots+\mathcal{U}_1(\alpha,0)}_{n!\ \text{times}}\right]+u = \frac{1}{n!}\left[(\mathcal{U}_1(\alpha,0)+u)+\cdots+(\mathcal{U}_1(\alpha,0)+u)\right] \tag{35}$$

$$= \frac{1}{n!}\left[\mathcal{U}_1(\alpha,u)+\cdots+\mathcal{U}_1(\alpha,u)\right] \tag{36}$$

Hence $\mathcal{U}_n(\alpha,u)=\mathcal{U}_n(\alpha,0)+u$, obeying axiom 3. (5) Expansion:

$$\frac{\beta}{\alpha}\frac{1}{n!}\left[\underbrace{\mathcal{U}_1(\alpha,0)+\cdots+\mathcal{U}_1(\alpha,0)}_{n!\ \text{times}}\right] = \frac{1}{n!}\left[\frac{\beta}{\alpha}\mathcal{U}_1(\alpha,0)+\cdots+\frac{\beta}{\alpha}\mathcal{U}_1(\alpha,0)\right] \tag{37}$$

$$= \frac{1}{n!}\left[\mathcal{U}_1(\beta,0)+\cdots+\mathcal{U}_1(\beta,0)\right] \tag{38}$$

Hence $\mathcal{U}_n(\beta,0)=\frac{\beta}{\alpha}\mathcal{U}_n(\alpha,0)$, obeying axiom 4. ∎

Proof of lemma 6. (1) Nesting results from lemma 5. (2) Contraction is a direct result of axiom 2. (3) Translation. For unions we have $\mathcal{W}(\alpha,0)+u=[\mathcal{U}(\alpha,0)\cup\mathcal{V}(\alpha,0)]+u=\mathcal{U}(\alpha,u)\cup\mathcal{V}(\alpha,u)=\mathcal{W}(\alpha,u)$. A similar relation holds for intersections. Note that, by axiom 1, $\mathcal{U}(\alpha,0)\cap\mathcal{V}(\alpha,0)$ is not empty. (4) Expansion. For unions we have $\frac{\beta}{\alpha}\mathcal{W}(\alpha,0)=\frac{\beta}{\alpha}[\mathcal{U}(\alpha,0)\cup\mathcal{V}(\alpha,0)]=\mathcal{U}(\beta,0)\cup\mathcal{V}(\beta,0)=\mathcal{W}(\beta,0)$. A similar relation holds for intersections. ∎

Proof of theorem 5. Since $\mathcal{U}(\alpha,u)$ is dispersed, let $\mathcal{U}(\alpha,u_1)$, …, $\mathcal{U}(\alpha,u_n)$ be disjoint as in definition 1. Then:

$$P\left[\bigcup_{i=1}^{n}\mathcal{U}(\alpha,u_i)\right]=\sum_{i=1}^{n}P[\mathcal{U}(\alpha,u_i)]=nh(\alpha) \tag{39}$$

$P(\cdot)$ is normalized and additive so it nowhere exceeds unity. Thus $h(\alpha)\le 1/n$ for any α and any finite n. Hence $h(\alpha)=0$ since, suppose there is an $\epsilon>0$ such that $h(\alpha)=\epsilon$. Let $n>1/\epsilon$. Since $h(\alpha)\le 1/n<\epsilon$ we conclude that the supposition is false and that $h(\alpha)$ is equal to no positive number. ∎

Proof of theorem 6. Suppose there is such a probability function. By theorem 5 we know that $P[\mathcal{U}(\alpha,u)]=0$ for all $(\alpha,u)\in R\times S$. Since the U-map is unbounded we conclude that the probability of any ball in E is zero. Since $P(\cdot)$ is additive and E is the union some, possibly infinite, number of balls, $P(E)=0$ which violates the normalization of the probability function. Thus no such probability function exists. ∎

Acknowledgement

The author is indebted to Prof. Deborah G. Mayo for stimulating comments and suggestions.

References

1. Z. Artstein and J.C. Hansen, 1984, Convexification in limit laws of random sets in Banach spaces, *The Annals of Probability,* 13: 307–309.
2. Y. Ben-Haim, 1985, *The Assay of Spatially Random Material,* Kluwer Academic Publishers, Dordrecht, Holland.
3. Y. Ben-Haim, 1994, Convex models of uncertainty: Applications and Implications, *Erkenntnis: An International Journal of Analytic Philosophy,* 41:139–156.
4. Y. Ben-Haim, A non-probabilistic concept of reliability, *Structural Safety,* 14:227–245 (1994).
5. Y. Ben-Haim, 1995, A non-probabilistic measure of reliability of linear systems based on expansion of convex models, *Structural Safety,* 17:91–109.

6. Y. Ben-Haim, 1996, *Robust Reliability in the Mechanical Sciences,* Springer-Verlag, Berlin.
7. Y. Ben-Haim, 1997, Robust reliability of structures, *Advances in Applied Mechanics,* 33: 1–41.
8. Y. Ben-Haim, S. Cogan and L. Sanseigne, 1998, Usability of Mathematical Models in Mechanical Decision Processes, *Mechanical Systems and Signal Processing.* 12: 121–134.
9. Y. Ben-Haim, 1997, Beyond maximum entropy: Exploring the structure of uncertainty, in H.G. Natke and Y. Ben-Haim, eds., *Uncertainty: Models and Measures,* Volkswagen Foundation Workshop, 22–24 July 1996, Lambrecht, Germany. Akademie-Verlag, Berlin.
10. Y. Ben-Haim, 1998, Reliability of Vibrating Structures With Uncertain Inputs. *Shock and Vibration Digest*, 30: 106–113.
11. Y. Ben-Haim, 1998, Sequential tests based on convex models of uncertainty, *Mechanical Systems and Signal Processing,* 12: 427–448.
12. Y. Ben-Haim, G. Chen and T.T. Soong, 1996, Maximum Structural Response Using Convex Models, *ASCE J. Engineering Mechanics,* 122:325–333.
13. Y. Ben-Haim and I. Elishakoff, 1990, *Convex Models of Uncertainty in Applied Mechanics,* Elsevier Science Publishers, Amsterdam.
14. D.I. Blockley, 1985, Fuzziness and probability: a discussion of Gaines' axioms, *Civil Engineering Systems*, 2: 195–200.
15. R. Carnap, 1962, *Logical Foundations of Probability,* University of Chicago Press, 2nd ed.
16. D. Dubois and H. Prade, 1988, *Possibility Theory: An Approach to Computerized Processing of Uncertainty,* Plenum Press, New York.
17. I. Elishakoff and L.P. Zhu, 1994, *Probabilistic and Convex Modelling of Acoustically Excited Structures,* Elsevier, Amsterdam.
18. B.R. Gaines, 1978, Fuzzy and probability uncertainty logics, *Information and Control,* 38: 154–169.
19. P.J. Kelly and M.L. Weiss, 1979, *Geometry and Convexity: A Study in Mathematical Methods,* Wiley.
20. J.M. Keynes, 1929, *A Treatise on Probability,* Macmillan and Co., London.
21. D. Hinrichsen and V.L. Kharitonov, 1995 Stability of polynomials with conic uncertainty, *Mathematics of Control, Signals and Systems,* 8: 97–117.
22. S.W. Kirkpatrick and B.S. Holmes, 1989, Effect of initial imperfections on dynamic buckling of shells, *ASCE Journal of Engineering Mechanics,* 115: 1075–1093.
23. A.N. Kolmogorov, 1956, *Foundations of the Theory of Probability,* 2nd English edition, Chelsea Pub. Co. Based on the 1933 German and 1936 Russian editions.
24. H.E. Kyburg, jr., 1990, Getting fancy with probability, *Synthese,* 90: 189–203.
25. H.E. Lindberg, 1992, An evaluation of convex modelling for multimode dynamic buckling, *ASME Journal of Applied Mechanics,* 59: 929–936.
26. H.E. Lindberg, 1992, Convex models for uncertain imperfection control in multimode dynamic buckling, *ASME Journal of Applied Mechanics,* 59: 937–945.
27. R.D. Luce, D.H. Krantz, P. Suppes and A. Tversky, 1990, *Foundations of Measurement,* vol. III: *Representation, Axiomatization and Invariance,* Academic Press.
28. D.G. Mayo, 1996, *Error and the Growth of Experimental Knowledge,* University of Chicago Press, Chicago.
29. F.C. Schweppe, 1973, *Uncertain Dynamic Systems,* Prentice-Hall, Englewood Cliffs.

30. T.T. Soong, 1990, *Active Structural Control: Theory and Practice,* Wiley, New York.

31. P. Suppes and M. Zanotti, 1977, On using random relations to generate upper and lower probabilities, *Synthese,* 36: 427–440.

32. *Webster's Ninth New Collegiate Dictionary,* 1985, Merriam-Webster Inc., Springfield Massachusetts.