Reliability Engineering and System Safety, 85: 249–266 (2004).

Uncertainty, Probability and Information-gaps

Yakov Ben-Haim
Itzhak Moda'i Chair in Technology and Economics
Faculty of Mechanical Engineering
Technion — Israel Institute of Technology
Haifa 32000 Israel
yakov@technion.ac.il
http://info-gap.com http://www.technion.ac.il/yakov

Contents

papers euw euw2e.tex 6.3.2002

Keywords: Epistemic uncertainty, information-gaps, info-gap models of uncertainty, robustness, opportunity, decision-making.

Abstract

This paper discusses two main ideas. First we focus on info-gap uncertainty, as distinct from probability. Info-gap theory is especially suited for modelling and managing uncertainty in system models: we invest all our knowledge in formulating the best possible model; this leaves the modeller with very faulty and fragmentary information about the variation of reality around that optimal model.

Second, we examine the interdependence between uncertainty-modelling and decision-making. Good uncertainty modelling requires contact with the end-use, namely, with the decision-making application of the uncertainty model. The most important avenue of uncertainty-propagation is from initial data- and model-uncertainties into uncertainty in the decision-domain. Two questions arise. Is the de cision robust to the initial uncertainties? Is the decision prone to opportune windfall success?

We apply info-gap robustness and opportunity functions to the analysis of representation and propagation of uncertainty in several of the Sandia challenge problems.

1 Introduction

This paper discusses two main ideas. First of all, the paper focusses on the intuition and quantification of information-gaps, as distinct from probabilistic conceptions. It will be claimed that info-gap theory provides a tool for modelling and managing uncertainty which is quite suitable for many situations, especially those characterized by highly deficient information. The prime example of an info-gap arises in system modelling: we invest all our knowledge in formulating the best possible model; this leaves the modeller in the position of having very faulty and fragmentary information about the variation of reality around that optimal model.

Second, we examine the intimate interdependence between the process of constructing a model of uncertainty, and the process of making decisions. We will assert that good uncertainty modelling requires, at critical stages, contact with the end-use, namely, with the decision-making application of the uncertainty model. This interaction between modelling and deciding has a profound impact on how one should represent the properties and propagation of an uncertain phenomenon. Stated differently, the most important avenue of propagation of uncertainty is from initial data- and model-uncertainties into uncertainty in the decision-domain. This revolves around the question of whether the decision is robust or immune to the initial uncertainties, that is, whether the decision is stable or dithers erratically due to model- and data-variation. We also consider whether the decision enjoys the opportunity for windfall success.

One of the motivations for info-gap theory is an awareness of the limitations of probability. Since the pitfalls in the application of probability are generally well known, we have relegated their discussion to the appendix, section 7. The reader who is familiar with this material can skip it without prejudicing the comprehension of the rest of the paper. However, every methodology has its limitations (even info-gap theory!) and one is well-advised to keep these limitations in mind.

In section 3.1 we outline very briefly the idea of an info-gap model of uncertainty. We then discuss the two decision functions of info-gap theory in section 3.2. The robustness function assesses the immunity to failure due to pernicious variations. The opportunity function evaluates the immunity to windfall due to propitious fluctuations. These immunity functions — robustness and opportunity — underlie distinct decision strategies. The robustness function induces a robust satisficing design strategy: maximizing the feasibility of satisfying at least minimal survival requirements. The opportunity function induces opportune windfalling: enhancing the ability to exploit favorable contingencies in order to achieve highly desirable results. In section 3.3 we examine the complementarity between information and uncertainty, and define a concept of the informativeness of an info-gap model. This leads to a method for assessing the value of an increment of information.

We apply these concepts to several of the Sandia Challenge Problems [27]. The Challenge Problems are accompanied by information which must be transformed into uncertainty models. We begin by formulating our version of descartian skepticism in section 2. Without being solipsistic (or just plain ornery), our brand of skepticism provides a method for guiding this transformation. In particular, we distinguish between apodictic and contingent interpretation of data.

In section 4 we outline our strategies for propagation and representation of uncertainty. In section 4.1 we study the propagation, into the decision domain, of info-gaps in the initial data. This involves the formulation of robustness and opportunity functions which, in our present discussion, are geared to the problem of model-based decision-making with uncertain data. In section 4.2 we study the formulation of an info-gap model of uncertainty from fragmentary initial data. This entails the use of an empirical robustness function which is used in choosing the structure and parameters of an info-gap model.

In section 5 we discuss three examples drawn from the Sandia Algebraic Challenge Problem. Section 5.1 deals with the propagation of interval uncertainty in the initial information. Section 5.2 demonstrates the info-gap representation of initial information. Section 5.3 again considers uncertainty propagation, this time of initial data which is presented probabilistically.

We do not discuss the Sandia Dynamic Challenge Problem. However, the info-gap analysis of vibrating systems may be found in a number of sources [2, 3, 4, 6].

2 Skepticism

We precede our discussion of the Challenge Problems with a brief analysis of the nature of the information they contain. This will lead us to formulate a concept of skepticism which will guide our development.

Each Challenge Problem is provided with assertions which are described as "information" about the associated parameters, such as that they belong to a particular set, or vary according to a specific probability density, or according to one of a class of probability densities.

The information is presented as propositions which can be interpreted in one of two ways: (1) as categorical, unequivocal and apodictic truths, or (2) as contingent, empirical interpretations or summaries of observations.

People tend to shy away from apodictic assertions, especially in our age of relativism. However, a fairly uncontroversial instance of an apodictic assertion concerns the net displacement in a diffusive random walk. The Einstein-Smoluchowski analysis shows this to be a normal distribution. This is of course contingent upon certain assumptions, but given the theory on which the analysis rests, the adoption of a normal distribution is very strongly supported. It is a theorem; if you accept the axioms, then you must live with the results. That's what 'apodictic' means.

A good example of a contingent empirical assertion is not hard to find. The distribution of the time to failure in low-level fatigue can be modelled as any of a number of distributions, and nobody has a really good fundamental theory of the underlying process by which we can distinguish between these distributions. (Some people believe they have a good theory, but I'm skeptical.)

We really do not know The Bard's intention in formulating the packets of information which accompany each Challenge Problem: whether apodictic or contingent. There are numerous risks entailed in elevating assertions from the status of contingent to apodictic. Some of these pitfalls are discussed in the appendix, to which the reader has already been referred. I will adopt M. Descartes' skeptical attitude which treats assertions as contingent until forced by evidence to adopt them absolutely.

What this means, to recapitulate, is that each parcel of information is viewed as a collation or extrapolation of observations. So, stating that the parameters fall in a given set is no more than a summary of data, and is not meant to imply that occurrences outside this set are impossible. Such occurrences have not yet been seen, and we do not know whether or not they can or will ever be observed. Likewise, specifying a probability density is an assertion of “best fit” to a finite collection of observations. Consequently, the far tails of the distribution may be wrong, since they predict occurrences which have never been witnessed.

The theory of info-gap uncertainty is apposite to our skeptical attitude towards contingent evidence.

3 Info-gap Decision Theory: A Précis

In this section we discuss some of the key ideas of info-gap decision theory. Section 3.1 defines the info-gap models of uncertainty upon which the theory is based. Section 3.2 discusses the two decision functions which induce the info-gap decision-strategies of robust-satisficing and opportune-windfalling. Section 3.3 touches on an intimate interdependence between information and decision which is elucidated by info-gap theory, showing that competent formulation of models of uncertainty requires cognizance of the end-use to which the models will be put: making decisions.

3.1 Info-gap Models of Uncertainty

Our quantification of knowledge-deficiency is based on non-probabilistic information-gap models [6]. An info-gap is a disparity between what the decision maker knows and what could be known. The range of possibilities expands as the info-gap grows. For instance, we may base our decisions upon the best-informed, most sophisticated system-models available. However, further study would improve our knowledge and understanding and would lead us to even better models of reality upon which to base our decisions.

Note that we distinguish between models of systems, and models of uncertainty. The latter may in fact refer to the knowledge-deficiency of our system-models. While the distinction between system-model and uncertainty-model is real, it is not iron-clad. The barrier is especially permeable when the uncertainty relates to uncertainty in the system-model itself. We will return, in section 3.3, to the interdependence between system-modelling and decision-making, on the one hand, and uncertainty-modelling on the other.

An info-gap model of uncertainty is a family of nested sets. Each set corresponds to a particular degree of knowledge-deficiency, according to its level of nesting. Each element in a set represents a possible event; a possible model of a physical or social system for instance. There are no measure functions in an info-gap model of uncertainty.

Info-gap theory provides a quantitative representation of Knight’s concept of “true uncertainty” for which “there is no objective measure of the probability”, as opposed to risk which is probabilistically measurable [19, pp.46, 120, 231–232]. Further discussion of the relation between Knight’s conception and info-gap theory is found in [6, section 12.5]. Similarly, Shackle’s “non-distributional uncertainty variable” bears some similarity to info-gap analysis [28, p.23]. Likewise, Kyburg recognized the possibility of a “decision theory that is based on some non-probabilistic measure of uncertainty.” [21, p.1094].

Events are represented as scalar or vectors, f, which may in fact be functions. Thus, for example, f specifies a model upon which our decisions are based. Knowledge-deficiency is expressed at two levels by info-gap models. For fixed α the set $\mathcal{F}(\alpha, \widetilde{f})$ represents a degree of variability of f around the centerpoint $\widetilde{f}$ (where for instance $\widetilde{f}$ represents the best known model). The greater the value of α, the greater the range of possible variation, so α is called the *uncertainty parameter* and expresses the information gap between what is known ($\widetilde{f}$ and the structure of the sets) and what needs to be known for an ideal solution (the exact value of f). The value of α is usually unknown, which constitutes the second level of imperfection of knowledge: the horizon of variation is unbounded.

Let $\Re_+$ denote the non-negative real numbers and let Ω be a Banach space in which the uncertain quantities f are defined. An info-gap model $\mathcal{F}(\alpha, \widetilde{f})$ is a map from $\Re_+ \times \Omega$ into the power set of Ω. All info-gap models obey the following two axioms. *Nesting:* $\mathcal{F}(\alpha, \widetilde{f}) \subseteq \mathcal{F}(\alpha', \widetilde{f})$ if $\alpha \le \alpha'$. *Contraction:* $\mathcal{F}(0, 0)$ is the singleton set $\{0\}$. Nesting is the most characteristic of the info-gap axioms. It expresses the intuition that possibilities expand as the info-gap grows, and imbues α with its meaning as an 'horizon of uncertainty'. The contraction axiom states that the 'centerpoint' event, $\widetilde{f}$, is the only possibility in the absence of uncertainty.

Many info-gap models encountered in practice obey further axioms which specify more specific structure. We will not need these additional axioms. For more discussion of info-gap axioms see [5].

3.2 Robustness and Opportunity

We now present two info-gap decision strategies: robust satisficing and opportune windfalling. These strategies lie at the heart of info-gap decision theory.

The decision maker will make a choice by selecting a value for q, which is a vector, or vector function of real and/or linguistic variables. The decision vector q belongs to a set of available decisions Q. The outcome of this choice is influenced by an unknown vector (or vector function) $f \in \Omega$ whose range of possible variation is represented by an info-gap model $\mathcal{F}(\alpha, \widetilde{f})$, $\alpha \ge 0$, in a Banach space Ω. In particular, we are interested in situations in which the decision is based on an incomplete or inaccurate model, which usually will be the centerpoint $\widetilde{f}$ of the info-gap model. The info-gap model $\mathcal{F}(\alpha, \widetilde{f})$ represents the decision maker's info-gaps regarding this design-base model. Since reality prefers some model f rather than $\widetilde{f}$, the outcome of decision q will be impacted by the gap between $\widetilde{f}$ and f.

We will define two real-valued reward functions defined on (q, A) where $q \in Q$ and $A \subset \Omega$. The **lower reward function** $\mathcal{R}_*(q, A)$ is monotonically decreasing on the power set of Ω, at fixed choice q:

$$A \subset B \quad \text{implies} \quad \mathcal{R}_*(q, A) \ge \mathcal{R}_*(q, B) \tag{1}$$

The **upper reward function** $\mathcal{R}^*(q, A)$ is monotonically increasing on the power set of Ω, at fixed choice x:

$$A \subset B \quad \text{implies} \quad \mathcal{R}^*(q, A) \le \mathcal{R}^*(q, B) \tag{2}$$

The monotonicity of these reward functions does not relate to the decision vector q, or to the decision maker's preferences regarding these choices. The reward functions are monotonic in the space Ω of unknown auxiliary events f, and expresses the impact of this ambient variation on the available outcomes of a choice q. Properties (1) and (2) do not establish preference relations on q.

The most common realizations of the lower and upper reward functions are in terms of least and greatest available rewards as f varies within a set $A \subset \Omega$. Let $r(q, f)$ be the reward actually realized when the choice is $q \in Q$ and the unknown quantity takes the value $f \in \Omega$. We construe the term 'reward' broadly, so that $r(q, f)$ may reflect economic profit, or technological performance such as structural stability, and so on. Employing a reward function $r(q, f)$, common choices of lower and upper reward functions $\mathcal{R}_*$ and $\mathcal{R}^*$ become:

$$\mathcal{R}_*(q, A) = \min_{f \in A} r(q, f) \tag{3}$$

$$\mathcal{R}^*(q, A) = \max_{f \in A} r(q, f) \tag{4}$$

Thus $\mathcal{R}_*(q, A)$ is the least available reward, while $\mathcal{R}^*(q, A)$ is the greatest available reward, in the uncertain environment A. In this realization, $\mathcal{R}_*$ and $\mathcal{R}^*$ still have not determined a preference relation on the decision q, both because the set A is undetermined and because the lower and upper rewards may behave differently.

More specifically, we will subsequently choose the set A as a set $\mathcal{F}(\alpha, \widetilde{f})$ in an info-gap family of nested sets. Then, in eqs.(3) and (4), $\mathcal{R}_*[q, \mathcal{F}(\alpha, \widetilde{f})]$ is the least accessible reward up to info-gap α, while $\mathcal{R}^*[q, \mathcal{F}(\alpha, \widetilde{f})]$ is the greatest accessible reward up to α. Recall that the info-gap model may represent knowledge-deficiency of the system-models upon which our decisions are based. Thus $\mathcal{R}_*[q, \mathcal{F}(\alpha, \widetilde{f})]$ and $\mathcal{R}^*[q, \mathcal{F}(\alpha, \widetilde{f})]$ are the least and greatest accessible rewards up to info-gap α in our system-models.

The reward functions $\mathcal{R}_*(x, A)$ and $\mathcal{R}^*(x, A)$ are real-valued and represent desirable reward. The decision maker prefers more rather than less reward. However, neither $\mathcal{R}_*(q, A)$ nor $\mathcal{R}^*(q, A)$ need be continuous or convex, (though we will sometimes assume continuity), nor do they derive from preference relations. Moreover, the decision maker can't know the values of $\mathcal{R}_*(q, A)$ and $\mathcal{R}^*(q, A)$ which will be realized in practice because the set A (of possible design-base models, for instance) is unknown. Finally, while $\mathcal{R}_*$ and $\mathcal{R}^*$ entail knowledge-deficiency, they are not probabilistic but instead depend on an info-gap model.

We will now use $\mathcal{R}_*(q, A)$ and $\mathcal{R}^*(q, A)$ to define two decision functions [6] which generate preferences on values of q. These preferences are not unique, nor are the same preferences necessarily derived from each of the two decision functions. These preferences will vary with aspiration level, the details of the decision problem, and possibly other exogenous factors.

Let $r_{\rm c}$ be a value of reward which the decision maker strives to achieve; more reward would be better, but less than $r_{\rm c}$ would be unacceptable. The decision maker wishes to satisfice at reward level $r_{\rm c}$. The **robustness** of choice q is the greatest level of knowledge-deficiency at which reward no less than $r_{\rm c}$ is guaranteed:

$$\widehat{\alpha}(q, r_{\rm c}) = \max \left\{ \alpha : \ \mathcal{R}_*[q, \mathcal{F}(\alpha, \widetilde{f})] \geq r_{\rm c} \right\} \tag{5}$$

$\widehat{\alpha}(q, r_{\rm c})$ is a **robust satisficing decision function.** One can readily prove that as the decision maker's aspirations rise, as expressed by increasing the critical reward $r_{\rm c}$, the robustness $\widehat{\alpha}(q, r_{\rm c})$ decreases. This is a fundamental trade-off between aspiration and feasibility. High aspirations (large $r_{\rm c}$) may not be feasible (because of low robustness, $\widehat{\alpha}(q, r_{\rm c})$), leading the decision maker to lower the aspiration. Conversely, one's aspirations may turn out, upon analysis, to be overly modest (having very large robustness), which enables one to aim at a higher target (larger $r_{\rm c}$).

Let $r_{\rm w}$ be a large value of reward (much greater than $r_{\rm c}$) which the decision maker would be delighted to achieve; lower reward would be acceptable, but reward as large as $r_{\rm w}$ is a windfall success. The **opportunity** inherent in choice q is the least level of knowledge-deficiency at which windfall can occur:

$$\widehat{\beta}(q, r_{\rm w}) = \min \left\{ \alpha : \ \mathcal{R}^*[x, \mathcal{F}(\alpha, \widetilde{f})] \geq r_{\rm w} \right\} \tag{6}$$

$\widehat{\beta}(q, r_{\rm w})$ is an **opportune windfalling decision function.** One can prove that $\widehat{\beta}(q, r_{\rm w})$ increases as $r_{\rm w}$ increases, which means that greater ambient uncertainty is needed in order to enable the possibility of greater windfall. This is a trade-off between certainty and opportunity: great certainty (small $\widehat{\beta}$) entails little opportunity for wondrous surprise (large $r_{\rm w}$).

$\widehat{\alpha}(q, r_{\rm c})$ and $\widehat{\beta}(q, r_{\rm w})$ are **immunity functions**. $\widehat{\alpha}(q, r_{\rm c})$ is the immunity against failure (reward less than $r_{\rm c}$). When $\widehat{\alpha}(q, r_{\rm c})$ is large, failure can occur only at great ambient uncertainty; the decision maker is not vulnerable to pernicious uncertainty. Similarly, $\widehat{\beta}(q, r_{\rm w})$ is the immunity against windfall (reward no less than $r_{\rm w}$). When $\widehat{\beta}(q, r_{\rm w})$ is small, windfall can occur even under mundane circumstances; the decision maker is not immune to propitious uncertainty.

These considerations lead to preference rankings on the choice vector q. While "bigger is better" for robustness $\widehat{\alpha}(q, r_{\rm c})$, "big is bad" for opportunity $\widehat{\beta}(q, r_{\rm w})$. The preferences induced by the robust-satisficing strategy are:

$$q \succeq_{\rm r} q' \quad \text{if} \quad \widehat{\alpha}(q, r_{\rm c}) \geq \widehat{\alpha}(q', r_{\rm c}) \tag{7}$$

Likewise, the preferences induced by the opportune-windfalling strategy are:

$$q \succeq_{\rm o} q' \quad \text{if} \quad \widehat{\beta}(q, r_{\rm w}) \leq \widehat{\beta}(q', r_{\rm w}) \tag{8}$$

We note that neither $\succeq_{\rm r}$ nor $\succeq_{\rm o}$ is necessarily single-valued for any pair of choices q and q': the preferences may change with $r_{\rm c}$ and $r_{\rm w}$, respectively. Moreover, $\succeq_{\rm r}$ and $\succeq_{\rm o}$ may rank the same pair of options differently. Since the aspiration parameter $r_{\rm c}$ and $r_{\rm w}$ are not fixed, the preference relations $\succeq_{\rm r}$ and $\succeq_{\rm o}$, either alone or together, do not satisfy the rationality conditions of c ompleteness and transitivity. They do not establish unique preferences for all pairs of available choices, and they hence do not entail transitivity of preference.

3.3 Decision, Uncertainty and the Value of Information

In this section we briefly examine two ideas, which are developed more fully elsewhere [6, chap. 7; 7]. First, information is the complement of uncertainty, which is particularly self-evident when uncertainty is conceived of as an info-gap. Second, the value of information can be assessed in terms of how it enhances the robustness-to-uncertainty of a decision.

The first idea, that information is the complement of uncertainty, is unproblematic. We should be able to exploit new information by improving our models of uncertainty. As our knowledge and understanding improves we should be able to discern and differentiate more finely between alternative possibilities. For instance, we should be able to select more discriminatingly between different system models. This motivates the following definition of the relative informativeness of two info-gap models.

Given two info-gap models, $\mathcal{F}_1(\alpha, \widetilde{f})$ and $\mathcal{F}_2(\alpha, \widetilde{f})$, we will say that the first is **more informative than** the second if:

$$\mathcal{F}_1(\alpha, \widetilde{f}) \subseteq \mathcal{F}_2(\alpha, \widetilde{f}) \tag{9}$$

$\mathcal{F}_1(\alpha, \widetilde{f})$ is more informative in the sense that it circumscribes the range of possibilities more restrictively than $\mathcal{F}_2(\alpha, \widetilde{f})$, at the horizon of uncertainty α. (For a wide range of structural axioms, if inclusion (9) holds at some particular strictly positive value of α, then it holds at all values of α.)

We must observe two reservations to inclusion (9) as a test for the relative informativeness of two info-gap models. First, 'relative informativeness' is not a complete ranking of all info-gap models. It may well happen that two info-gap models intersect but that neither is included in the other. They are then simply not commensurate by this criterion of informativeness. Their information is different.

Second, $\mathcal{F}_1$ is more informative than $\mathcal{F}_2$ in the sense that, if $\mathcal{F}_1$ is verified as a model of uncertain variation, then $\mathcal{F}_1$ provides a more precise picture of reality than is available if $\mathcal{F}_2$ is verified. Or, if our information is no better than $\mathcal{F}_2$, then we are less well informed than if our information as good as $\mathcal{F}_1$.

Now consider the second idea: **the value of information.** Suppose we have two info-gap models which are ranked by informativeness as in inclusion (9). $\mathcal{F}_1$ is more informative than $\mathcal{F}_2$, but is the difference substantial or insignificant? Or, thinking of these two models as objects of study whose verification or falsification we are contemplating: is $\mathcal{F}_1$ worth the effort to verify even if much greater effort is required than for the verification of $\mathcal{F}_2$?

There can be no absolute answer to these questions. Rather, the answer depends on what one wishes to do with the models. We here formulate an answer in terms of robustness to uncertainty. Each info-gap model generates a robustness function, which we denote $\widehat{\alpha}_1(q, r_{\rm c})$ and $\widehat{\alpha}_2(q, r_{\rm c})$. It is readily proven that these robustnesses are ranked numerically if and only if the corresponding info-gap models are ranked by inclusion. That is, inclusion (9) is equivalent to:

$$\widehat{\alpha}_1(q, r_{\rm c}) \geq \widehat{\alpha}_2(q, r_{\rm c}) \tag{10}$$

The more informative info-gap model induces more robust decisions, at any level of aspiration $r_{\rm c}$. This provides a means of assessing the relative value of the info-gap models. The increment of information needed to verify $\mathcal{F}_1$ over $\mathcal{F}_2$ is valuable to the extent that decision q is more robust to uncertainty given $\mathcal{F}_1$ rather than $\mathcal{F}_2$. If we can identify significantly more feasible options with $\mathcal{F}_1$ than we can with $\mathcal{F}_2$, then the former is significantly more valuable. It is clear that subjective assessment of increments of robustness in linguistic terms such as 'large', 'significant', etc., is no mean task (see [4; 6, chap. 4]). However, the point is that choosing to develop a particular info-gap model, rather than some other possible info-gap model, should depend upon the decisions one will base upon the model. By comparing the robustness functions for these info-gap models we can assess their relative utility. The optimistic and progressive maxim that one should 'do the best one can' in modelling the uncertainty is unrealistic and unnecessary, not to say unscientific.

This is not to say that one cannot compartmentalize. Having chosen a particular info-gap model-structure, one then proceeds to realize that model, for instance by estimating its parameters, without worrying day and night if some simpler and only slightly less valuable info-gap model is lurking around somewhere.

4 Representation and Propagation of Uncertainty

We have explained our interpretation of the information which accompanies each Challenge Problem as a contingent, empirical collation or summary of observations. We are now in a position to define info-gap approaches to the representation and propagation of uncertainty.

We will suggest two different aspects of uncertainty-propagation, and study one of them in section 4.1. What distinguishes the two approaches is the space into which the informational uncertainty propagates. In the first method the uncertainty is propagated directly into the space of responses of the system defined by the Challenge Problem. In the second method, upon which we will concentrate, the uncertainty propagates into the decision space: that realm where the model is used to choose an action. In this method we are concerned with the degree to which the end-use decision is instable due to uncertainty in the primary data.

In section 4.2 we explore the representation of data-deficiency with an info-gap model. Specifically, we illustrate the estimation of the parameters of an info-gap model based on data.

Let $\mathcal{X}(\alpha, \widetilde{x})$, for $\alpha \geq 0$, denote the info-gap model of the data which are summarized as the Challenge Problem's "information". $\mathcal{X}(\alpha, \widetilde{x})$ may represent info-gaps in the values of parameters or functions, or info-gaps on probability distributions of those entities. We will encounter examples later on.

The first method, upon which we will not dwell, is **direct propagation.** The Challenge Problem specifies an input/output system: uncertain input drives the system to produce uncertain output. The input is the info-gap model $\mathcal{X}(\alpha, \widetilde{x})$, which is a family of nested sets. The output is the family of sets of responses of the system. The knowledge-deficiency is propagated from the info-gap of the basic information, to an info-gap in the response space. We will not pursue this direction further, though it can be very useful and entails many interesting challenges.

4.1 Uncertainty Propagation from Data to Decision

In the second method we focus on **decision stability** and ask: how does our uncertainty about the primary data propagate to the final decision? Is the final decision robust or immune, stable or dithering, with respect to the initial informational deficiency? We will explore this with robustness functions, as discussed in section 3.2. Since uncertainty can be propitious as well as pernicious, we will also enquire if the final decision is prone to wonderful windfalls due to unanticipated favorable deviations from the nominal data. This is examined with the opportunity function.

A robustness function for evaluating decision stability can be formulated as follows. Let x be the data about which we are uncertain: for instance the parameters of the algebraic or dynamic equations in the Sandia Challenge problems. Our knowledge-deficiency regarding x is represented by the info-gap model $\mathcal{X}(\alpha, \widetilde{x})$, $\alpha \geq 0$.

The data, x, induces a decision, $D(x)$. For instance, in the algebraic Challenge Problem to be discussed in section 5, x is the parameter vector (a, b) and $D(x)$ is some function of the system response $y = (a + b)^a$. The decision algorithm $D(x)$ may be continuous and real-valued to indicate the choice of a time, a location, or some other dimension. Or $D(x)$ may be integer-valued to represent a selection between distinct options such as 'go', 'no-go', 'do this', 'do that', etc.

We are interested in modelling the uncertainty in the initial data, x, and the propagation of this uncertainty into the decision space. We wish to assess the degree of severity of this data-uncertainty for the end-use of that information. The robustness function assesses the degree of data variability which can be tolerated by the end-user. If this robustness is great, then the uncertainty in x is benign; on the other hand, low robustness implies malignantly deficient information. That is, the robustness function addresses the question: how variable is x, in terms of impact on end-use?

Our decision will be based on a best estimate x_{e} of x, which may (but need not) be the centerpoint $\widetilde{x}$ of the info-gap model. One common **robustness function** (see [8] for others) evaluates the stability of possible decisions $D(x)$ compared to the decision which will actually be made, $D(x_{\mathrm{e}})$. Let $\|D(x) - D(x_{\mathrm{e}})\|$ be a measure of how different these decisions are from one another. The robustness of decision algorithm D is the greatest info-gap in x which does not allow the decision $D(x)$ to vary unacceptably from $D(x_{\mathrm{e}})$:

$$\widehat{\alpha}(D, r_{\mathrm{c}}) = \max \left\{ \alpha : \max_{x \in \mathcal{X}(\alpha, \widetilde{x})} \|D(x) - D(x_{\mathrm{e}})\| \leq r_{\mathrm{c}} \right\} \tag{11}$$

The aspiration parameter r_{c} is the greatest acceptable fluctuation of the decision we *would have made* if we had used data x (which we don't have), compared against the decision we *actually will make* given data x_{e} (which are actually in hand). A large value of $\widehat{\alpha}(D, r_{\mathrm{c}})$ implies that the data can vary greatly without unacceptably impacting the decision. That is, a large robustness means that the data do not vary perniciously. On the other hand, a small value of $\widehat{\alpha}(D, r_{\mathrm{c}})$ implies that the data are dangerously unstable. In other words, the robustness function $\widehat{\alpha}(D, r_{\mathrm{c}})$ assesses the degree of **pernicious** variability of the initial data x, with respect to the decision maker's survival aspiration r_{c} for the end-use $D(x)$.

r_{c} expresses the decision maker's minimal requirement for stability of the algorithm. The decision maker may aspire to far greater decision-stability. Let r_{w} be a very small value of decision-fluctuation, much less than r_{c}. Achieving decision-stability as good as r_{w} is not necessary for survival, but would be viewed as a highly desirable windfall. The **opportunity function** assesses the immunity to windfall, and is the lowest level of info-gap in the data which enables, but does not guarantee, windfall:

$$\widehat{\beta}(D, r_{\mathrm{w}}) = \min \left\{ \alpha : \min_{x \in \mathcal{X}(\alpha, \widetilde{x})} \|D(x) - D(x_{\mathrm{e}})\| \leq r_{\mathrm{w}} \right\} \tag{12}$$

A small value of the opportunity function $\widehat{\beta}(D, r_{\mathrm{w}})$ means that the decision algorithm is opportune: windfall is an immediate possibility. Regarding data uncertainty, a small $\widehat{\beta}(D, r_{\mathrm{w}})$ implies that propitious fluctuations of x are imminent. The opportunity function $\widehat{\beta}(D, r_{\mathrm{w}})$ assesses the degree of **propitious** variability of the initial data x, with respect to the decision maker's windfall aspiration r_{w} for the end-use $D(x)$.

4.2 Uncertainty Representation with Info-gap Models

In section 4.1 we analyzed decision stability by employing a given info-gap model $\mathcal{X}(\alpha, \widetilde{x})$ to represent uncertainty in the initial data. We used that info-gap model to construct decision-immunity functions (robustness and opportunity functions). These immunity functions assess the propagation of the data-uncertainty into the decision domain. In this section we show how $\mathcal{X}(\alpha, \widetilde{x})$ is formulated and up-dated based upon measurements. That is, we are now concerned with **info-gap modelling.**

Let us suppose that we have a collection of data, X_i, $i \in \mathcal{I}$, where $\mathcal{I}$ is a set of indices which identify the data. We will think of each X_i as a set of measurements. This set may contain a single number, or a vector of measurements, or it may be an infinite set of data-values based on measurements. We will use these data sets X_i to choose the *properties* and *parameters* of an info-gap model $\mathcal{X}(\alpha, \widetilde{x})$. The *properties* which we choose will specify the type of info-gap model (e.g. interval, ellipsoid, integral-energy, etc.). The *parameters* to be determined can be any of the quantities which specify the specific structure of the info-gap model, such as the centerpoint. Let q be a vector specifying these properties and parameters.

The approach we take in formulating and up-dating an info-gap model involves two stages. First we use the data to evaluate an empirical robustness function, $\widehat{\alpha}_{\rm e}(q, r_{\rm c})$. This can be done in many different ways, with different interpretations, and suitable to different specific problems. We will consider one approach, in which $\widehat{\alpha}_{\rm e}(q, r_{\rm c})$ assesses the fidelity between the data and the info-gap model. Other approaches are found in [6, ch. 8; 7]. In the second stage we choose q to optimize the fidelity of the info-gap model to the data. As we will explain later, this is done by choosing q to minimize $\widehat{\alpha}_{\rm e}(q, r_{\rm c})$.

Let $\mathcal{X}_q(\alpha, \widetilde{x})$ denote the info-gap model which is specified by the up-date vector q. The **gap function** for data set X, with respect to $\mathcal{X}_q(\alpha, \widetilde{x})$, is the lowest info-gap at which $\mathcal{X}_q(\alpha, \widetilde{x})$ is consistent with X:

$$\Gamma(X, q) = \min \{\alpha : \ X \subseteq \mathcal{X}_q(\alpha, \widetilde{x})\} \tag{13}$$

If the value of the gap function $\Gamma(X, q)$ is small, then $\mathcal{X}_q(\alpha, \widetilde{x})$ is highly consistent with the data X; a large gap implies low fidelity between model and data.

Let us suppose that we have a best-estimate of the data, $X_{\rm e}$. That is, $X_{\rm e}$ is our anticipation of a typical value, or set of values, of the data. We think of $X_{\rm e}$ as a set to keep our notation consistent. Let $\rho(X, X_{\rm e})$ be a measure of the difference between the data X and the best-estimate $X_{\rm e}$. (In other approaches to the up-date task we do not need the estimate $X_{\rm e}$.)

The **empirical robustness** is the greatest info-gap, α, at which data that do not differ too greatly from $X_{\rm e}$ are also consistent with $\mathcal{X}_q(\alpha, \widetilde{x})$:

$$\widehat{\alpha}_{\rm e}(q, r_{\rm c}) = \max \{\Gamma(X_i, q), \ i \in \mathcal{I} : \ \Gamma(X_j, q) \le \Gamma(X_i, q) \ \text{ if } \ \rho(X_j, X_{\rm e}) \le r_{\rm c}, \ j \in \mathcal{I}\} \tag{14}$$

We understand the empirical robustness function $\widehat{\alpha}_{\rm e}(q, r_{\rm c})$ as follows. Data set X_j is not an outlier if it does not differ from our anticipation $X_{\rm e}$ by too much: $\rho(X_j, X_{\rm e}) \le r_{\rm c}$. We can choose $r_{\rm c}$ to be more or less lenient in our definition of acceptable deviation of data. The empirical robustness is the greatest info-gap at which the info-gap model $\mathcal{X}_q(\alpha, \widetilde{x})$ 'captures' all non-outlying data.

One may note some similarity between the optimization in eq.(14) and Tikhonov regularization [30]. The parallel is that the data is regularized (stabilized) around $X_{\rm e}$, with $r_{\rm c}$ as the regularization parameter. The optimization then reflects the comparison between distinct data sets X_i and X_j.

The info-gap model is compatible with the data if $\mathcal{X}_q(\alpha, \widetilde{x})$ captures the data at a low horizon of uncertainty. That is, a small value of $\widehat{\alpha}_{\rm e}(q, r_{\rm c})$ indicates high fidelity between data and info-gap model. Hence we up-date the info-gap model, based on the data X_i, $i \in \mathcal{I}$, by choosing q to minimize the empirical robustness. The optimal model, $\widehat{q}$, is specified by:

$$\widehat{\alpha}_{\rm e}(\widehat{q}, r_{\rm c}) = \min_q \widehat{\alpha}_{\rm e}(q, r_{\rm c}) \tag{15}$$

5 Sandia Algebraic Challenge Problem

In this section we will apply the concepts of info-gap uncertainty representation and propagation developed in section 4, to the Sandia Algebraic Challenge Problem. In section 5.1 we illustrate the propagation of uncertainty from the initial data to the decision domain, employing the first data set which contains only interval data. In section 5.2 we study the info-gap representation of uncertainty for data set 2c, which contains conflicting interval data. In section 5.3 we again study uncertainty propagation, this time with data set 4 which contains both interval and probability information.

5.1 Uncertainty Propagation from Data to Decision: I

In this section we illustrate the propagation of an info-gap from the initial data, (a, b), to the decision space. We will use the robustness and opportunity functions discussed in section 4.1.

The uncertain entity is the vector $x = (a, b)$. We will consider the first data set:

$$a \in [a_1, a_2], \quad b \in [b_1, b_2] \tag{16}$$

where $a_1 = 0.1$, $a_2 = 0.5$, $b_1 = 0.2$ and $b_2 = 0.7$. In the skeptical spirit discussed in section 2, we interpret this information to mean that all observations to date fall in these intervals. Where the next observation will fall is unknown.

An interval-bound info-gap model is a family of nested intervals of a and b values. Since the sizes of these intervals are unknown and unbounded, this info-gap model allows the variables to take any real values, and entails no assumptions about the likelihood of different values. We will consider shortly the implications of using symmetric, or asymmetric intervals.

We will adopt a symmetric interval-uncertainty info-gap model, where the intervals are centered on the data, and the sizes of the intervals are unknown:

$$\mathcal{X}(\alpha, \widetilde{x}) = \left\{ x = (a, b) : \ \frac{|a - \widetilde{a}|}{(a_2 - a_1)/2} \le \alpha, \quad \frac{|b - \widetilde{b}|}{(b_2 - b_1)/2} \le \alpha \right\}, \quad \alpha \ge 0 \tag{17}$$

$\widetilde{x}_1 = \widetilde{a} = (a_2 + a_1)/2$ and $\widetilde{x}_2 = \widetilde{b} = (b_2 + b_1)/2$. This info-gap model is a family of nested sets of (a, b) vectors. Each value of α, the horizon of uncertainty, determines a set of vectors, $\mathcal{X}(\alpha, \widetilde{x})$. Each set $\mathcal{X}(\alpha, \widetilde{x})$ specifies the range of variation of (a, b) at info-gap α. The value of α, of course, is unknown.

Other info-gap models are available. For instance, one could employ asymmetric intervals or different normalizations of α (different denominators in eq.(17)). The effect of renormalization is to change the calibration of the robustness. Some insight can be obtained, and even some specific decisions can be made, without confronting the difficult value judgments entailed in calibrating the robustness (answering the question: how much robustness is enough?). Other aspects of the decision problem require calibration (see [6, chap. 4]). After the robustness is calibrated in value-terms, the specific normalization is no longer important (though it cannot be changed without modifying the calibration). Related, though more difficult, considerations apply to asymmetric intervals.

We are exploring the propagation of the data-uncertainty into the decision space, so we need to know how the decision varies with x. That is, we wish to evaluate the data-uncertainty in terms of its impact on the stability of the decision. Our measure of decision-fluctuation will be the fractional error in the response of the algebraic system:

$$\|D(x) - D(x_{\mathrm{e}})\|^2 = \left(\frac{y(x) - y(x_{\mathrm{e}})}{y(x_{\mathrm{e}})} \right)^2 \tag{18}$$

where $y(x) = (a + b)^a$ and $x_{\mathrm{e}} = (a_{\mathrm{e}}, b_{\mathrm{e}})$ is a best-estimate of (a, b). The final decision will be based on x_{e}. The error of this decision, eq.(18), is the fractional deviation of the system response with x_{e}, from the system response with the possibly correct but unknown x. The robustness function, $\widehat{\alpha}(D, r_{\mathrm{c}})$ in eq.(11), is the greatest horizon of uncertainty α at which the error of the decision can never be greater than the critical stability parameter r_{c}. The opportunity function, $\widehat{\beta}(D, r_{\mathrm{w}})$ in eq.(12), is the least horizon of uncertainty α at which the error of the decision may be as small as the windfall stability parameter r_{w}. "Bigger is better" for $\widehat{\alpha}(D, r_{\mathrm{c}})$ while "big is bad" for $\widehat{\beta}(D, r_{\mathrm{w}})$.

Fig. 1 shows robustness and opportunity functions versus aspirations r_{c} and r_{w}. Consider first the robustness function $\widehat{\alpha}(D, r_{\mathrm{c}})$, which is the immunity to failure, so a bigger value of $\widehat{\alpha}$ is better than a smaller value. A large value of the critical level of decision-error, r_{c}, is less demanding than a small value, so the robustness $\widehat{\alpha}(D, r_{\mathrm{c}})$ increases with increasing r_{c}. For $r_{\mathrm{c}} < 0.1$ the robustness is zero, indicating that even the nominal data, $\widetilde{x}$, fail to guarantee decision-stability as good as r_{c}.

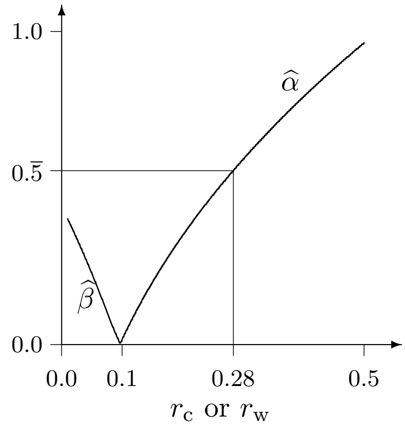

Figure 1: Robustness $\widehat{\alpha}(D, r_c)$ (increasing curve), and opportunity $\widehat{\beta}(D, r_w)$ (decreasing curve), versus aspiration r_c or r_w. a_e= 0.35, b_e= 0.25.

This monotonic increase of $\widehat{\alpha}(D, r_c)$ expresses the trade-off between great aspiration (small r_c) and great immunity to uncertainty (large $\widehat{\alpha}$).

The opportunity function is the decreasing curve in fig. 1, which is interpreted as follows. The opportunity $\widehat{\beta}(D, r_w)$ is the least level of info-gap at which decision-stability as good as r_w is possible, though not guaranteed. As we have explained, $\widehat{\beta}(D, r_w)$ is the immunity to windfall, so a small value of $\widehat{\beta}(D, r_w)$ is better than a large value. A large value of the windfall decision-stability, r_w, is less demanding than a small value, so the opportunity function $\widehat{\beta}(D, r_w)$ decreases (improves) with increasing r_w. For $r_w > 0.1$ the opportunity function is zero, indicating that even the nominal data, $\widetilde{x}$, enable decision-stability as good as r_w. This monotonic decrease of $\widehat{\beta}(D, r_w)$ expresses the trade-off between great aspiration (small r_w) and great certainty (small $\widehat{\beta}$).

The numerical data a_i and b_i upon which the info-gap model of eq.(17) is based are the ranges of the observed values of a and b. These extreme values, possibly single outlying observations, are the least reliable aspect of the data, as explained in our skeptical discussion in section 2. To understand the import of our data-uncertainty, it is important to recognize when these limiting values are relevant, and when they are not. The parameter-intervals in the uncertainty-set $\mathcal{X}(\alpha, \widetilde{x})$ are all contained within the observed ranges if and only if $\alpha \leq 0.\bar{5}$. On the other hand, when $\alpha > 0.\bar{5}$, the interval of b-values in $\mathcal{X}(\alpha, \widetilde{x})$ extends beyond the observations; likewise, when $\alpha > 0.\bar{6}$ the interval of a-values also extends beyond the observations.

Now, referring to fig. 1, we note that for fractional error r_c less 0.28, the robustness $\widehat{\alpha}$ is less than $0.\bar{5}$. If our prediction requirements are much less demanding that 28% (that is, if $r_c \gg 0.28$ is acceptable), then our data-fluctuation is not problematical because the robustness far exceeds the observed variation of the data. On the other hand, if we aspire to much lower prediction error ($r_c \ll 0.28$ is demanded), then the data-uncertainty is a major impediment because $\widehat{\alpha}$ is less than the observed data-fluctuation.

Now we apply similar reasoning to the opportunity function $\widehat{\beta}(D, r_w)$ in fig. 1. Windfall performance is possible, though not guaranteed, in the r_w-range evaluated, for the data-variability which has been observed. For instance, the point (0.05, 0.22) lies on the $\widehat{\beta}$-curve, indicating that fractional decision-error as small as 0.05 is possible with an info-gap of 0.22. The data support variability as large, or larger, than $\alpha = 0.22$.

Fig. 2 shows how the robustness and opportunity of the end-use decision varies with the best-estimate a_e of parameter a, for fixed b_e, r_c and r_w. This provides a different assessment of the propagation of the initial data-uncertainty into the decision-space. We see that the robustness is lowest (for this range of a_e-values) when the estimate a_e equals the info-gap-centerpoint $\widetilde{a} = 0.3$. Similarly, the opportunity is worst ($\widehat{\beta}$ is greatest) when $a_e = \widetilde{a}$. This results (rather tortuously) from

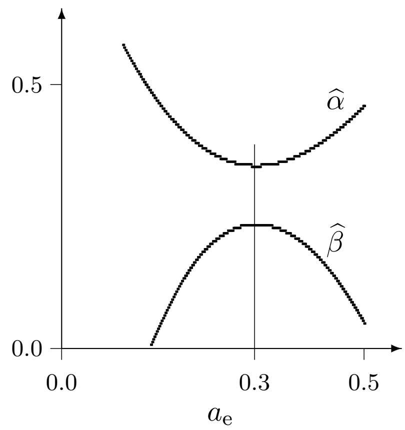

Figure 2: Robustness $\widehat{\alpha}(D, r_{\rm c})$ (upper curve), and opportunity $\widehat{\beta}(D, r_{\rm w})$ (lower curve), versus best-estimate $a_{\rm e}$. $b_{\rm e}$= 0.25, $r_{\rm c}$= 0.20, $r_{\rm w}$= 0.05.

the fact that $y(a, b)$ is a decreasing function of a in this range of values ($a < 1$).

We note in fig. 2 that the immunity functions, $\widehat{\alpha}(D, r_{\rm c})$ and $\widehat{\beta}(D, r_{\rm w})$, are fairly insensitive to the estimate $a_{\rm e}$ when $a_{\rm e}$ is near the centerpoint, $\widetilde{a}$. For instance, the immunity curves are rather flat as $a_{\rm e}$ varies in the interval 0.3 ± 0.06. Data-fluctuations which cause the best estimate of a to vary in that range will not significantly impact the robustness or opportunity of the final decision. On the other hand, if the initial data cause a to vary substantially outside this range then we anticipate important impact on the performance of our end-use decision algorithm.

Finally, we see in fig. 2 that robustness and opportunity vary sympathetically as $a_{\rm e}$ changes. Any change in $a_{\rm e}$ which improves (increases) $\widehat{\alpha}(D, r_{\rm c})$ also improves (decreases) $\widehat{\beta}(D, r_{\rm w})$. Robustness and opportunity are not always sympathetic in this sense.

5.2 Info-gap Representation of Uncertainty

We will now illustrate the method for estimating the parameters of an info-gap model which is described in section 4.2. Our example will employ the Sandia algebraic challenge problem 2c: the information about a is a single interval and the information about b is a collection of intervals which are not mutually consistent.

The information about a is the interval $A = [a_1, a_2]$ and the information about b is the 4 intervals $B_i = [b_1^i, b_2^i]$, $i = 1, \ldots, 4$. Specifically, $A = [0.1, 1.0]$, $B_1 = [0.6, 0.8]$, $B_2 = [0.5, 0.7]$, $B_3 = [0.1, 0.4]$, $B_4 = [0.0, 1.0]$. In the notation of section 4.2, our data sets are $X_i = (A, B_i)$, $i = 1, \ldots, 4$.

We will consider the following class of fractional-error info-gap models:

$$\mathcal{X}(\alpha, \widetilde{x}) = \left\{ (a, b) : \ \frac{|a - \widetilde{a}|}{\widetilde{a}} \le \alpha, \quad \frac{|b - \widetilde{b}|}{\widetilde{b}} \le \alpha \right\}, \quad \alpha \ge 0 \tag{19}$$

The decision vector q will be the centerpoint vector $\widetilde{x} = (\widetilde{a}, \widetilde{b})$. We will choose a positive $(\widetilde{a}, \widetilde{b})$ to minimize the empirical robustness function, eq.(14).

The gap function for data-set X_i, eq.(13), is:

$$\Gamma(X_i, \widetilde{x}) = \min \{ \alpha : \ X_i \subseteq \mathcal{X}(\alpha, \widetilde{x}) \} \tag{20}$$

One readily finds that the gap function can be expressed as the greatest of four numbers:

$$\Gamma(X_i, \widetilde{x}) = \max \left(\frac{\widetilde{a} - a_1}{\widetilde{a}}, \ \frac{a_2 - \widetilde{a}}{\widetilde{a}}, \ \frac{\widetilde{b} - b_1^i}{\widetilde{b}}, \ \frac{b_2^i - \widetilde{b}}{\widetilde{b}} \right) \tag{21}$$

The best-estimate of $\widetilde{x}$ is the set $X_{\rm e} = (\{a_{\rm e}\}, \{b_{\rm e}\})$. The difference between a data-set X_i and the estimate-set $X_{\rm e}$ is greatest absolute difference of corresponding interval-end-points:

$$\rho(X_i, X_{\rm e}) = \max\Big(|a_1 - a_{\rm e}|,\ |a_2 - a_{\rm e}|,\ |b_1^i - b_{\rm e}|,\ |b_2^i - b_{\rm e}|\Big) \tag{22}$$

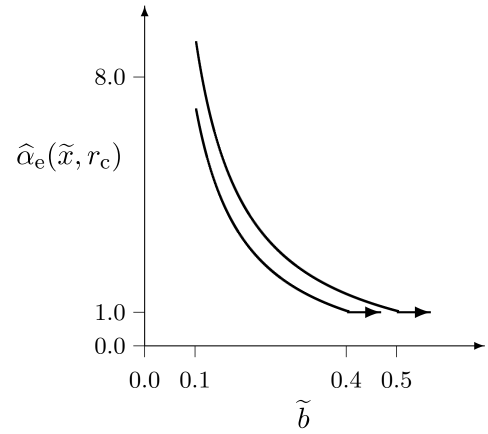

Figure 3: Empirical robustness $\widehat{\alpha}_{\rm e}(\widetilde{x}, r_{\rm c})$ vs. $\widetilde{b}$. $\widetilde{a} = 0.5$, $a_{\rm e} = 0.35$, $b_{\rm e} = 0.25$. $r_{\rm c} = 1.0$ (upper curve); $r_{\rm c} = 0.65$ (lower curve). Data set 2c.

Fig. 3 shows the empirical robustness function, $\widehat{\alpha}_{\rm e}(\widetilde{x}, r_{\rm c})$, versus the choice of the centerpoint parameter $\widetilde{b}$; the other parameters are fixed. The lower curve is evaluated with $r_{\rm c} = 0.65$, which causes the 4th data-set, $X_4 = (A, B_4)$, to be excluded by the distance criterion $\rho(X_j, X_{\rm e}) \le r_{\rm c}$ in eq.(14). The upper curve is computed with $r_{\rm c} = 1.0$, which is large enough to allow all four data-sets to participate in the estimation process.

Both $\widehat{\alpha}_{\rm e}$-curves decrease steeply as $\widetilde{b}$ increases from 0.1. As indicated by the arrows, each curve shows a kink when $\widehat{\alpha}_{\rm e}$ reaches the value of unity, and $\widehat{\alpha}_{\rm e} = 1$ for all larger values of $\widetilde{b}$. This is a result of the fractional-error structure of the info-gap model which is being estimated. The important practical conclusion from these calculations is that the fidelity between the info-gap model $\mathcal{X}(\alpha, \widetilde{c})$ and the data degenerates rapidly as $\widetilde{b}$ is reduced below 0.5 (or below 0.4 if one chooses to reject X_4). This facilitates the choice of a best estimate of $\widetilde{b}$, contingent upon the value of $\widetilde{a}$ which is used. Of course, the search must now proceed to the variation of $\widetilde{a}$, until a combination $(\widetilde{a}, \widetilde{b})$ is found which minimizes $\widehat{\alpha}_{\rm e}(\widetilde{x}, r_{\rm c})$, as expressed in eq.(15).

5.3 Uncertainty Propagation from Data to Decision: II

We now consider the 4th data set, which contains interval-uncertainty for parameter a and mixed interval/probabilistic uncertainty for b. We will concentrate on the info-gap analysis of the propagation of the probabilistic uncertainty. We should treat the interval information skeptically as well, as we did in the info-gap analysis of section 5.1. However, we will sacrifice integrity for the sake of clarity, and we will adopt the intervals apodictically. We will construct robustness and opportunity functions for the prediction of the probability of extreme responses of the system.

The information about the probability density function (pdf) of the parameter b is that it is log-normal with mean and standard deviation in specified intervals. Our skepticism regarding the probabilistic information focusses on the shape of the distribution of b. Unless we have fundamental arguments in support of the log-normal distribution (such as the Einstein-Smoluchowski theory for the normal distribution in diffusion), we must view the tail of the distribution as unknown: if the log-normal distribution is adopted on the basis of empirical evidence, there must be a point on its tail beyond which no observations have been obtained. The log-normality of this far upper tail must be viewed skeptically. Likewise, since the evidence is in any case ultimately a finite collection of data, we must acknowledge that the central "bulk" of the distribution is also imperfectly known, though we may have greater confidence in the bulk than in the tail. The knowledge-deficiency here is not simply in the value of the moments of the distribution, but in its actual shape. This sort of info-gap can be represented in various ways. We consider one possibility here: the envelope-bound info-gap model.

Define $\lambda = \ln b$, so the system response is $y = \left(a + \mathrm{e}^{\lambda}\right)^{a}$. Let $\mathcal{P}$ be the set of all mathematically possible pdfs defined on $(-\infty, \infty)$: all non-negative functions normalized to unity. Let $\widetilde{p}(\lambda|\mu,\sigma)$ be the pdf of the normal distribution with mean μ and variance σ^2. Let $M = [\mu_1,\, \mu_2]$ and $S = [\sigma_1,\, \sigma_2]$ be the intervals of means and standard deviations of λ in the Sandia data set: $\mu_1 = 0$, $\mu_2 = 1$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.5$. The nominal pdfs of λ are $\widetilde{p}(\lambda|\mu,\sigma)$ for any $\mu \in M$ and any $\sigma \in S$. Let $A = [a_1,\, a_2]$ be the Sandia data for a, where $a_1 = 0.1$ and $a_2 = 1$.

Consider the following envelope-bound info-gap model for uncertainty in a and in the pdf of λ:

$$\mathcal{X}(\alpha) = \left\{ (a, p(\lambda)) :\; a \in A;\; p \in \mathcal{P};\; \frac{|p(\lambda) - \widetilde{p}(\lambda|\mu,\sigma)|}{\widetilde{p}(\lambda|\mu,\sigma)} \le \alpha\psi(\lambda);\; \mu \in M;\; \sigma \in S \right\},\quad \alpha \ge 0 \tag{23}$$

Each pdf of λ in this info-gap model deviates, fractionally, from each nominal pdf $\widetilde{p}(\lambda|\mu,\sigma)$, $\mu \in M$, $\sigma \in S$, within an envelope of size α and shape $\psi(\lambda)$. The horizon of deviation, α, is unknown. The shape-function $\psi(\lambda)$ is dispensable, but it can be used to discriminate between different regions of the distribution. For instance, if we are quite confident of the normality of $p(\lambda)$ up to some value, λ_0, we would choose $\psi(\lambda) = 0$ for $\lambda \le \lambda_0$. Beyond this value we may choose a constant envelope. (This info-gap model does not in fact obey the contraction axiom, which entails some minor technicalities which we will skip.)

Since we are considering the propagation of data-uncertainty into the decision domain, we must specify the relevant end-use. We will evaluate the robustness and opportunity functions for evaluating the probability that extreme values of the system response, y, exceed a specified threshold.

Let η be a large value which is rarely reached by the system response y, according to the initial information. That is, suppose that the probability that y exceeds η is less than some small probability, r_{c}, for any $a \in A$ and for any nominal pdf $\widetilde{p}(\lambda|\mu,\sigma)$ for $\mu \in M$ and $\sigma \in S$. That is, '$y \ge \eta$' is very unlikely, given the initial information about a and λ. We would like to know if this conclusion is robust to the info-gaps in that information: can the actual pdf, $p(\lambda)$, deviate greatly from the nominal set of pdfs and still preserve the conclusion that '$y \ge \eta$' is rare? The answer to this question is embodied in the **robustness function:**

$$\widehat{\alpha}(\eta, r_{\mathrm{c}}) = \max \left\{ \alpha :\; \max_{(a,p)\in\mathcal{X}(\alpha)} \mathrm{Prob}(y \ge \eta) \le r_{\mathrm{c}} \right\} \tag{24}$$

If $\widehat{\alpha}(\eta, r_{\mathrm{c}})$ is large, then the actual pdf of λ may deviate greatly from the nominal normal distributions without jeopardizing the conclusion that y rarely exceeds η. On the other hand, if $\widehat{\alpha}(\eta, r_{\mathrm{c}})$ is small, then the nominal normality need only be mildly violated in order to enable y to exceed η more frequently that r_{c}.

Now suppose that the probability that y exceeds η is greater than some small probability, r_{w} (less than r_{c}), for any $a \in A$ and for any nominal pdf $\widetilde{p}(\lambda|\mu,\sigma)$ for $\mu \in M$ and $\sigma \in S$. That is, '$y \ge \eta$' is not rarer than r_{w}, given the initial information about a and λ. We would like to know if the actual pdf of λ may in fact entail lower probability that y exceeds η. The answer is contained in the **opportunity function:**

$$\widehat{\beta}(\eta, r_{\mathrm{w}}) = \min \left\{ \alpha :\; \min_{(a,p)\in\mathcal{X}(\alpha)} \mathrm{Prob}(y \ge \eta) \le r_{\mathrm{w}} \right\} \tag{25}$$

The opportunity, $\widehat{\beta}(\eta, r_{\mathrm{w}})$, is the lowest info-gap at which it is possible that the actual probability that y exceeds η is less than r_{w}. If $\widehat{\beta}(\eta, r_{\mathrm{w}})$ is small, then the favorable situation that '$y \le \eta$' is as rare as $r_{\rm w}$ is possible even at small deviations of the actual from the nominal distribution of λ. That is, small $\widehat{\beta}$ implies that opportune deviations are close at hand. On the other hand, if $\widehat{\beta}(\eta, r_{\rm w})$ is large, then only great error in our initial information entails the possibility of very favorably low probability of response as extreme as η.

We will construct the immunity functions, $\widehat{\alpha}(\eta, r_{\rm c})$ and $\widehat{\beta}(\eta, r_{\rm w})$, for η far out on the tail of the nominal distributions. The condition '$y \geq \eta$' is equivalent to '$\lambda \geq \xi(a)$' where we define $\xi(a) = \ln\left(\eta^{\frac{1}{a}} - a\right)$. Note that $\xi(a)$ decreases with increasing a, for $\eta > 1$.

The maximum probability of '$\lambda \geq \xi(a)$', up to info-gap α, occurs when $p(\lambda)$ lies on the upper envelope with respect to the nominal pdf with maximal mean and variance, and with $a = a_2$ because $\xi(a)$ decreases with increasing a. That is:

$$\max_{(a,p)\in\mathcal{X}(\alpha)} \text{Prob}(y \geq \eta) = \int_{\frac{\xi(a_2)-\mu_2}{\sigma_2}}^{\infty} \phi(z)[1+\alpha\psi(z)]\,\mathrm{d}z \tag{26}$$

$$= 1 - \Phi\left(\frac{\xi(a_2)-\mu_2}{\sigma_2}\right) + \alpha \int_{\frac{\xi(a_2)-\mu_2}{\sigma_2}}^{\infty} \phi(z)\psi(z)\,\mathrm{d}z \tag{27}$$

where $\phi(z) = \widetilde{p}(z|0,1)$ and $\Phi(z)$ is its cumulative probability function. Let us define $z_2 = \frac{\xi(a_2)-\mu_2}{\sigma_2}$. The robustness is the least upper bound of the set of α-values for which eq.(27) does not exceed the critical probability, $r_{\rm c}$. We find the robustness by equating eq.(27) to $r_{\rm c}$ and solving for α:

$$\widehat{\alpha}(\eta, r_{\rm c}) = \frac{r_{\rm c} - 1 + \Phi(z_2)}{\int_{z_2}^{\infty} \phi(z)\psi(z)\,\mathrm{d}z} \tag{28}$$

unless this is negative, in which case $\widehat{\alpha}(\eta, r_{\rm c}) = 0$, meaning that there is no robustness against failure in predicting the probability of the extreme event '$y \geq \eta$'.

The opportunity function is constructed similarly. The minimum probability of '$\lambda \geq \xi(a)$', up to info-gap α, occurs when $p(\lambda)$ lies on the lower envelope with respect to the nominal pdf with minimal mean and variance, and with $a = a_1$. Define $z_1 = \frac{\xi(a_1)-\mu_1}{\sigma_1}$. The minimal probability is:

$$\min_{(a,p)\in\mathcal{X}(\alpha)} \text{Prob}(y \geq \eta) = \int_{z_1}^{\infty} \phi(z)[1-\alpha\psi(z)]\,\mathrm{d}z \tag{29}$$

The opportunity function is the least value of α at which eq.(29) does not exceed the windfall probability $r_{\rm w}$. Equating eq.(29) to $r_{\rm w}$ and solving for α yields the opportunity function:

$$\widehat{\beta}(\eta, r_{\rm w}) = \frac{1 - \Phi(z_1) - r_{\rm w}}{\int_{z_1}^{\infty} \phi(z)\psi(z)\,\mathrm{d}z} \tag{30}$$

unless this is negative, in which case $\widehat{\beta}(\eta, r_{\rm w}) = \infty$, meaning that there is no opportunity that '$y \geq \eta$' is as improbable as $r_{\rm w}$.

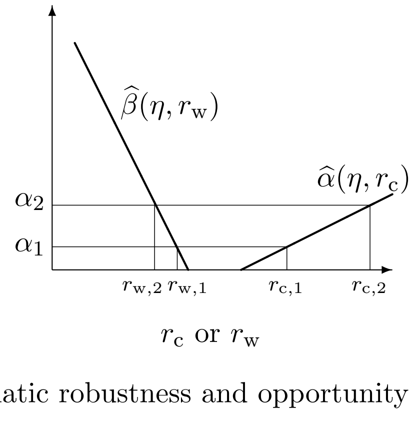

Figure 4: Schematic robustness and opportunity curves, eqs.(28) and (30).

The first thing to note in the immunity functions of eqs.(28) and (30) is their performance trade-offs: their variation against $r_{\rm c}$ and $r_{\rm w}$, shown in fig. 4. Consider the robustness, $\widehat{\alpha}(\eta, r_{\rm c})$, which increases as the probability $r_{\rm c}$ of extreme response increases. This means that the prediction of exceedence-probabilities of rare events is more vulnerable to uncertainty that the prediction of more common events. This appears in fig. 4 as the increasing $\widehat{\alpha}$-curve, which gives quantitative expression to this intuitively plausible result. Greater robustness of prediction is obtained only by predicting more common events.

The opportunity function decreases with increasing exceedence-probability, meaning that less uncertainty is needed to enable the opportunity that exceedence will be rarer than expected, as the level of exceedence probability $r_{\rm w}$ increases. Greater ambient uncertainty is needed to facilitate more favorable events. The opportunity function quantifies the adage that "nothing ventured, nothing gained".

The relative positions of the immunity curves in fig. 4, and the fact the $\widehat{\beta}$ decreases more rapidly than $\widehat{\alpha}$ increases, result from the fact that $z_1 > z_2$.

Fig. 4 illustrates a simple relationship between robustness and opportunity in the current example. The robustness at which exceedence-probability $r_{\rm c,2}$ is predicted, is greater than the robustness of predicting the rarer exceedence-probability $r_{\rm c,1}$: common events are predicted more reliably than rare events. However, notice that when the ambient uncertainty is α_2 (which is tolerable at robustness $\widehat{\alpha}(\eta, r_{\rm c,2})$), the opportunity arises for exceedence-probability as low as $r_{\rm w,2}$. In contrast, if the info-gap is only α_1 (which is the greatest acceptable uncertainty if we must predict events as rare as $r_{\rm c,1}$), then windfall exceedence probabilities will be no less than $r_{\rm w,1}$.

However, there is a much deeper relation of sympathy and antagonism between the immunity functions. To understand this relation we must first recall the preference orderings of the two immunity functions. A large value of the robustness function, $\widehat{\alpha}(\eta, r_{\rm c})$, is preferred over a small value, while for the opportunity function, $\widehat{\beta}(\eta, r_{\rm w})$, smaller values are preferred. Sloganistically: "bigger is better" for $\widehat{\alpha}$, while "big is bad" for $\widehat{\beta}$. The immunity functions vary sympathetically if a change which enhances one function also enhances the other. They vary antagonistically if improving one immunity is at the expense of the other. We will demonstrate that, in the present example, robustness and opportunity varying sympathetically as well as antagonistically, in different parameter ranges.

The robustness function in eq.(28) depends on the maximum nominal mean, μ_2, through the parameter z_2. Differential analysis shows that:

$$\frac{\partial \widehat{\alpha}(\eta, r_{\rm c})}{\partial \mu_2} < 0 \tag{31}$$

The opportunity function in eq.(30) varies with the minimum nominal mean, μ_1 which appears in z_1. Now, however, the slope of the variation may be either positive or negative, depending on the magnitude of the various quantities involved:

$$\frac{\partial \widehat{\beta}(\eta, r_{\rm w})}{\partial \mu_1} \gtrless 0 \tag{32}$$

The robustness and opportunity vary sympathetically if their slopes are opposite in sign, because then an improvement in one is accompanied by an improvement in the other. From eqs.(31) and (32) we see that any variation in our data which shifts the estimated extreme values of the mean, μ_1 and μ_2, in the same direction, may cause either antagonistic or sympathetic shifts in the immunity functions.

A similar effect is observed in connection with the extreme response threshold η, which appears in the immunity functions through z_i. We find:

$$\frac{\partial \widehat{\alpha}(\eta, r_{\rm c})}{\partial \eta} < 0, \quad \frac{\partial \widehat{\beta}(\eta, r_{\rm w})}{\partial \eta} \gtrless 0 \tag{33}$$

The immunity functions can vary either antagonistically or sympathetically as one varies the extreme response threshold η.

Finally, we wish to use the robustness function of eq.(28) to illustrate the value of information discussed in section 3.3.

The envelope function $\psi(z)$ which appears in the denominator of eq.(28) originates in the info-gap model of eq.(23). Consider two different envelope functions where one provides a 'tighter' envelope than the other:

$$\psi_1(z) \leq \psi_2(z) \tag{34}$$

For example, $\psi_1(z)$ may be zero over a range in which $\psi_2(z)$ is positive, while they are equal elsewhere. Denote the corresponding info-gap models by $\mathcal{X}_1(\alpha)$ and $\mathcal{X}_2(\alpha)$. It is clear that:

$$\mathcal{X}_1(\alpha) \subseteq \mathcal{X}_2(\alpha) \tag{35}$$

Thus $\mathcal{X}_1(\alpha)$ is **more informative than** $\mathcal{X}_2(\alpha)$, as explained in connection with eq.(9) in section 3.3.

Examining eq.(28) shows that a tighter envelope entails greater robustness. Denoting by $\widehat{\alpha}_i$ the robustness for info-gap model $\mathcal{X}_i$, one finds that relation (34) implies that $\widehat{\alpha}_1$ is **more robust than** $\widehat{\alpha}_2$:

$$\widehat{\alpha}_1(\eta, r_{\rm c}) \geq \widehat{\alpha}_2(\eta, r_{\rm c}) \tag{36}$$

This is precisely the conclusion discussed in section 3.3: greater informativeness implies greater robustness. Furthermore, the value of the more informative info-gap model is quantified in terms of the enhanced robustness it entails. This is particularly useful in identifying potential lines of research and development. Rather than adopting the presumption that more precise models are inherently more useful, one can quantify the value of additional information in terms of the enhanced robustness of the resulting decisions. Finally, if two info-gap models have envelope functions which cross, namely, which are not ranked as in eq.(34), then their robustness functions likewise are not strictly ranked: they will intersect for some value of $r_{\rm c}$.

6 Concluding Comment

It is *not* the contention of this paper that one should never use probability densities or fixed uncertainty intervals. It is *not* the contention here that models and measurements can never be accepted apodictically, without reservation. It *is* the contention that any evidence and any mathematical representation of uncertainty should be subjected to severe scrutiny before their tentative contingent status is up-graded to a less tentative level or especially to the status of full blown truth. Info-gap theory is one among the many tools available to perform this scrutiny, and to model and manage uncertainty when alternative tools are unsuitable.

7 Appendix: The Misuse of Probability

The theory of probability has found a rightful place at the heart of much engineering analysis. However, the pitfalls of probability are numerous, and the alternative tools for modelling and managing uncertainty are none too well known. In this section we touch on a range misconceptions, both common and not so common, which arise in the use of probability. We begin with some elementary errors in the application and interpretation of inverse probability, and we discuss a range of seeming paradoxes in probability. The principle of indifference (or maximum entropy or insufficient reason) is examined in an attempt to highlight where the dangers lie. Next, extrapolation from typical to rare events, which arises in calibrating a probability distribution, is studied. Then we introduce the idea that an uncertainty model imposes structure on an uncertain reality. Finally, the conflict between rationality and realism in the modelling of uncertainty is investigated.

7.1 Surprises of Inverse Probability

The mathematical theory of probability took a decidedly new tack with the publication in 1763 of Thomas Bayes' work on inverse probability [1]. Until that time probability was calculated in a 'forward' manner: from a specified model (e.g. fair dice) to probabilities of outcomes (e.g. probability of a run of 6's). Mr. Bayes proposed to invert things. In modern language, the problem he posed was: given measured outcomes of a random experiment (e.g. an observed run of 6's), estimate the parameters of the probability model. The very possibility of such a venture was in itself sufficiently innovative to cause Bayes' friend, Richard Price, to submit the articles to the Royal Society of London after Bayes' death. Since that time the power of inverse probability has attracted the attention of generations of mathematicians. But even very simple applications of inverse probability can lead to surprising results, as all experienced statisticians know.

Disraeli claimed that prevarication comes in three forms: lies, damned lies, and statistics. The statistics textbooks are full of examples of the latter category. One of DeGroot's [11, p.71] more heart-warming instances concerns a new test for cancer. If the test is applied to a person sick with cancer (S) the probability of a positive reaction ($+$) is 0.95 and the probability of a negative reaction ($-$) is 0.05. If the test is applied to a healthy individual (H) the probability of positive and negative reactions are 0.05 and 0.95 respectively. Sounds like a pretty good test. However, if only 1 in 10^5 individuals has cancer, then the probability that a test-positive individual is actually sick is so small as to be of negligible diagnostic value. This is the conditional probability of 'S' given '$+$':

$$p(S|+) = \frac{p(+|S)p(S)}{p(+|S)p(S) + p(+|H)p(H)} = 1.9 \times 10^{-4} \tag{37}$$

The moral of the story is, of course, let the customer beware and be wise: what seemed like a good test turned out, on careful analysis, to be of no value whatever. Much of the voluminous debate on topics of more or less precise technological definability would be greatly reduced if the protagonists were more statistically sophisticated.

7.2 Probability 'Paradoxes'

Allais paradox. The expected utility theory of games and economic behavior, as initiated by von Neumann and Morgenstern [25] in the 1940's, has played a central role in the study of decision under uncertainty. Expected utility theory provides a paradigm of behavior which is founded on probabilistic concepts as well as on particular axioms of 'rational' behavior. The Allais paradox is an empirical result showing that human behavior does not always conform to the axioms of the theory. While the Allais paradox does not challenge probability theory as such, it hits at the use of the linear expectation operator. The following illustration of the Allais paradox is based on [23, p.179].

A lottery has three prizes with monetary values $V_1 = \$2.5 \times 10^6$, $V_2 = \$0.5 \times 10^6$ and $V_3 = \$0$. A lottery is specified by $L = (p_1, p_2, p_3)$ where p_n is the probability of winning the prize valued at V_n.

The decision maker is offered a choice between two pairs of lotteries. The first pair is:

$$L_1 = (0,\ 1,\ 0) \quad \text{and} \quad L_1' = (0.10,\ 0.89,\ 0.01) \tag{38}$$

The second pair of lotteries between which the decision maker must choose is:

$$L_2 = (0,\ 0.11,\ 0.89) \quad \text{and} \quad L_2' = (0.10,\ 0,\ 0.90) \tag{39}$$

It is often found that people prefer L_1 over L_1' and prefer L_2' over L_2.

What expected utility theory does with lottery-preferences such as these is to express them in terms of personal-utility values of the individual outcomes weighted by the lottery probabilities. Let u_n be the utility to the decision maker of winning the prize valued at V_n. (The u_n's are not necessarily just linear functions of the V_n's; utilities may saturate at large values for instance.) The preference of L_1 over L_1' is expressed as:

$$u_2 > 0.10u_1 + 0.89u_2 + 0.01u_3 \tag{40}$$

where the lefthand side is the expected utility of lottery L_1 and the righthand side is the expected utility of lottery L_1'. Likewise, the expected utilities of the preference of L_2' over L_2 becomes:

$$0.10u_1 + 0.90u_3 > 0.11u_2 + 0.89u_3 \tag{41}$$

These two inequalities, (40) and (41), are mutually inconsistent if one requires expected utilities to combine linearly, as implied by the probabilistic expectation operator itself. We see this by adding $0.89u_3 - 0.89u_2$ to both sides of (40), to obtain:

$$0.11u_2 + 0.89u_3 > 0.10u_1 + 0.90u_3 \tag{42}$$

which is precisely the reverse of (41). Human behavior is perverse, and the very reasonable axioms of von Neumann-Morgenstern expected utility theory, which rest intimately on concepts of probability, do not match human behavior in some circumstances, as illustrated by the pair of preferences $L_1 \succ L_1'$ and $L_2' \succ L_2$.

The economists have uncovered other vagaries of human behavior which clash with at least some probabilistic models of decision under uncertainty, such as the Ellsberg and Machina paradoxes [13; 22; 23, pp.180, 207].

Petersburg paradox. [14]. This hoary riddle has tasked the best minds since the emergence of mathematical probability. We will not review the myriad solutions [17] but only state the conundrum and leave it to the reader to learn the lesson of caution.

A fair coin is flipped until the first occurrence of 'heads' which, if it happens on the nth throw, results in awarding a prize of 2^n dollars. The question is: what is a fair stake on this game? Or, what is the maximum which a rational (that is, greedy) gambler would be willing to put up as collateral for the right of playing the game a large number of times?

A fair stake is, according to standard probability, simply the mean award implied by the rules of the game. If a prize of value v_n is won with probability p_n then the fair stake is:

$$\mathrm{E}(v) = \sum_{n=1}^{\infty} p_n v_n \tag{43}$$

In the Petersburg game we have $p_n = 1/2^n$ while $v_n = 2^n$ so the expected award is infinite, as is the fair stake.

There is nothing paradoxical (or wrong) with the mathematics. The 'paradox' is simply that a reasonable theory (probability) fails to give a reasonable result (finite stake). No gambler in his right mind (if there are any) would risk infinite wealth (or even just humongous wealth), even if he planned to break the Guinness record for non-stop gambling. The resolutions of this quandary all involve modifying the rules of the game in some way to allow probability to come up with satisfactory rules of behavior.

Transcendental Probability. The Petersburg paradox is noteworthy for illustrating a situation in which probability unexpectedly fails to produce a reasonable solution to a well defined problem. We now consider the reverse situation: an obviously insolvable problem is solved, it would seem, by an elementary probability argument.

The last riddle in Lewis Carroll's *Pillow Problems,* which he tantalizingly identifies as a riddle in "transcendental probability", reads as follows [10]:

> A bag contains 2 counters, as to which nothing is known except that each is either black or white. Ascertain their colours without taking them out of the bag.

Since this is obviously impossible the real riddle is to uncover the error in Carroll's half-page probabilistic proof that "One is black, and the other white." Not only does Mr. Carroll cheat in his algebraic manipulation of conditional probability, he slides in the assumption that the probability of white and black are equal. This assumption is not warranted by the original information about the counters, "as to which nothing is known except that each is either black or white." Any probability distribution, not just the uniform distribution, is consistent with this prior information.

It was quite common in the 19th century to interpret the phrase "as to which nothing is known" to mean "equal probabilities". However, one wonders if Carroll (to say nothing of most modern readers) was really aware that he was invoking the Principle of Indifference which has been the subject of considerable controversy. Without this enormously convenient assumption, Carroll's proof collapses, as do many modern probabilistic arguments and inferences. On this we will have more to say in section 7.3.

7.3 Total ignorance

The 'Principle of Indifference' [24] (or 'Insufficient Reason' [16] or 'Maximum Entropy' [29]) all provide solutions to the same problem: select a probability distribution when the prior information is insufficient to do so. That is, the prior information is augmented by appeal to a 'Principle' whose validity (it is claimed) is prior to the information in question.

It is patent that a distribution obtained in this manner is not a 'maximum ignorance' distribution since it depends on knowledge of the truth entailed in the Principle which was employed. For instance, knowing that an uncertain phenomenon is described by some particular distribution, e.g. the uniform distribution or a triangular distribution, is clearly more knowledge than not knowing at all what distribution describes the phenomenon, as can be illustrated with a simple probability riddle [2, chap.7].

Keynes' Counter-example. The merits of the principle in question have been debated for decades, and there is no intention here to review the arguments (see for instance [24, p.75–79]). It is illuminating however to recall one of Keynes' examples of the contradictions inherent in indiscriminate use of the Principle of Indifference [18, p.45]. I choose Keynes because of his staunch adherence to probability as the best, indeed in his day the only, means of arguing from uncertain evidence to reasonable conclusions.

Consider a material whose specific volume has the value between 1 and 3, but about which no other information is available. The Principle of Indifference allows us to assume a probability of 1/2 that the specific volume lies between 1 and 2, and 1/2 that it lies between 2 and 3. Now consider the specific density of the same material, which is the inverse of specific volume so its value must lie between 1/3 and 1. Again, nothing else is known, so (incautious) application of the Principle leads us to conclude that the probability is 1/2 that the specific volume lies between 2/3 and 1. From this we conclude therefore that the probability is 1/2 that the specific volume lies between 1 and 3/2, which contradicts the first conclusion about the specific volume.

Examples such as this abound, and the precise pitfall in applying the Principle of Indifference is not immediately evident. Keynes' ultimate limited warranty on the Principle of Indifference is an insightful defense of subjective probability based on "intuition or direct judgment" [18, p.52]. He justifies the Principle when it is used in conjunction with "our knowledge of the *form and meaning* of the alternatives" [18, p.61] (italics in the original). Keynes provides a rule — indivisibility of the alternatives into "sub-alternatives *of the same form."* [18, p.60] (italics in the original) — for testing the Principle of Indifference. Keynes' defense of the Principle of Indifference, and of the tenets of subjective probability in general [26], has been the subject of wide debate [24, p.75]. Wherever one stands in that polemic, Keynes' deep and densely reasoned argument stands as a fog-horn warning against the indiscriminate use of the Principle of Indifference.

Two Envelopes. Here's a simple way to earn some extra cash, compliments of the Principle of Indifference. Two envelopes are placed on the table, about which you know that each contains a positive amount of money and that one envelope contains twice as much money as the other. You are invited to choose one envelope and then to open and examine its contents. You do so, and find \$100 inside. Now you are given the option of exchanging this envelope for the other one, which contains either \$50 or \$200. On the principle of indifference, you assign equal probabilities to these two contingencies, and evaluate the expected reward if you make the exchange as $0.5 \times \$50 + 0.5 \times \$200 = \$125$. Worth doing!

Notice that this argument is valid regardless of the amount you found in the initial envelope (provided it is positive, as stipulated). In other words, once you had chosen an envelope, you would be willing to exchange it for the other envelope even without opening the first envelope: the expected reward is thereby augmented by 25%.

Now if this 25% expected gain works once without opening the envelope, as the principle of indifference assures that it does, then it will work again. And again. And again.

Until, I suppose, the envelopes split at the seams.

7.4 Tail Wagging the Dog

Probability distributions have parameters which depend upon moments or other measurable quantities. For instance, the exponential distribution has a single parameter which can be estimated from the mean of a random sample of the distribution; the normal distribution has two parameters, whose estimation requires two moments. Once the parameters are estimated, the probability distribution predicts the frequency distribution of events much rarer than any which were observed in the random sample. In short, distributions are calibrated from 'typical' events and give the analyst mastery of exceedingly rare events.

Two pitfalls are evident in this approach. One, the *form* of the probability distribution may be flawed. One could, for instance, calculate the mean and variance of a normal distribution from a random sample of an F distribution. The statisticians have developed many useful tools for testing the form of a proposed distribution, such as χ^2 tests and Kolmogorov-Smirnov tests, but we leave this more advanced matter for the textbooks.

The second pitfall is more elementary. Errors in parameter estimation are *magnified* when the calibrated distribution is used to predict rare events. A few simple examples will illustrate this.

But first let us recall that estimates are just that: estimates. When randomly sampling a distribution with true mean and variance μ and σ^2, the sample mean, $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$, is itself a random variable. The mean and variance of $\overline{x}$ are $\mathrm{E}(\overline{x}) = \mu$ and $\mathrm{var}(\overline{x}) = \sigma^2/N$. To get a feel for the numbers, consider the exponential distribution with parameter λ, for which $\mu = 1/\lambda$ and $\sigma^2 = 1/\lambda^2$. If this distribution is sampled N times in order to estimate its mean, the fractional error of the estimate, at one standard deviation, is:

$$\frac{\sqrt{\mathrm{var}(\overline{x})}}{\mathrm{E}(\overline{x})} = \frac{\sqrt{1/(\lambda^2 N)}}{1/\lambda} = \frac{1}{\sqrt{N}} \tag{44}$$

For instance, the fractional error in the estimation of the mean, at one standard deviation, of a sample of size $N = 1000$ is 0.032 or 3.2%.

How significant is an error of this size for estimating events far out on the tail, such as in assessing high levels of reliability or low levels of failure probability?

We will consider an example from the exponential distribution. The failure function of the exponential distribution is $F_{\mathrm{e}}(x;\lambda) = \mathrm{e}^{-\lambda x}$, which is the probability that the random variable exceeds the value x. In typical safety analyses one desires to know the value of the failure function at high quantiles. (The qth quantile is the number x_q such that x has probability q of not exceeding x_q. That is: $\mathrm{Prob}[x \le x_q] = q$.)

Let us consider the fractional change of the failure function due to small change in the parameter

λ:

$$\rho_{\rm e}(x) = \frac{F_{\rm e}(x;\lambda) - F_{\rm e}(x;\lambda_0)}{F_{\rm e}(x;\lambda_0)} = -1 + {\rm e}^{-(\lambda-\lambda_0)x} \tag{45}$$

The qth quantile of the exponential distribution is $x_q = -\frac{1}{\lambda_0}\ln(1-q)$. At the qth quantile, $x = x_q$, the fractional change in the failure function, eq.(45), becomes:

$$\rho_{\rm e}(x_q) = -1 + (1-q)^{(\lambda-\lambda_0)/\lambda_0} \tag{46}$$

Table 1 shows values of $\rho_{\rm e}(x_q)$ for various ratios of λ's and at various quantiles.

	λ/λ_0		
q	0.99	0.975	0.95
$1-10^{-3}$	0.072	0.19	0.41
$1-10^{-4}$	0.096	0.26	0.59
$1-10^{-5}$	0.122	0.33	0.78
$1-10^{-6}$	0.148	0.41	0.99

Table 1: Values of $\rho_{\rm e}(x_q)$, the fractional change in the failure function, for the exponential distribution.

From table 1 we see that shifts in λ are greatly magnified as changes in the failure function. For instance, a 1% error in λ results, at the $1-10^{-3}$th quantile, in a 7.2% shift in the probability of failure. A 2.5% error in λ results in a 26% error in the failure function at the $1-10^{-4}$th quantile. A 5% λ-error leads to a 99% failure-probability-error at the $1-10^{-6}$th quantile. It is obvious that rather modest errors in estimating the parameter λ are greatly magnified in evaluating the probability of very rare events.

A brief look at the normal distribution will show that it can be even more sensitive to estimation errors. Let $F_{\rm n}(x;\sigma^2)$ be the failure function of the zero-mean normal distribution with variance σ^2. As in eq.(45), the fractional error of $F_{\rm n}(x;\sigma^2)$ compared against the standard normal distribution is:

$$\rho_{\rm n}(x) = \frac{F_{\rm n}(x;\sigma^2) - F_{\rm n}(x;1)}{F_{\rm n}(x;1)} \tag{47}$$

Table 2 shows values of $\rho_{\rm n}(x_q)$ for two different quantiles and three errors in the variance.

$x_{0.9938} = 2.5$		$x_{0.9986} = 2.99$	
σ	$\rho_{\rm n}(x_q)$	σ	$\rho_{\rm n}(x_q)$
1.012	0.089	1.017	0.018
1.029	0.216	1.031	0.345
1.042	0.320	1.049	0.576

Table 2: Values of $\rho_{\rm n}(x_q)$, the fractional change in the failure function, for the normal distribution.

Table 2 shows, for instance, that a 1.2% error in the standard deviation results, at the 0.9938th quantile, in an 8.9% error in the failure probability. Or, a 4.9% error in the standard deviation leads to a 58% error in the failure probability at the 0.9986th quantile. Again, large magnification of error, and at even lower quantiles than for the exponential distribution.

7.5 Structuring

Not infrequently it is declared that "The world is probabilistic." This appealing proposition leads the declarer to the inevitable conclusion that one can do no better than to describe the world probabilistically.

However, probability is a mathematical theory, not a property of the natural or technological worlds. As Kolmogorov puts it in the preface to his axiomatization of probability: "The author set himself the task of putting in their natural place, among the general notions of modern mathematics, the basic concepts of probability theory ...." [20, p.v]. The assertion that the world is probabilistic is actually intended to assert that the mathematical theory of probability succeeds in describing some or all observations about the world.

In the same way, when we say that physical space is euclidean what we mean is that Euclid's mathematical theory of geometry describes observable properties of space, such as that the angles of a triangle sum to 180^o. But the physicists assure us that some observations, such as angle-sums of cosmic-sized triangles, are better described by non-euclidean geometries. This highlights the difference between 'space' (an attribute of nature) and 'euclidean geometry' or 'lobachevskian geometry' which are mathematical models of space. These different models capture different facets of an infinitely more subtle reality.

In the same way we can distinguish between 'uncertainty' (an attribute of nature) and 'models of uncertainty' (which are mathematical theories). There exists a range of models of uncertainty, not all probabilistic, each of which captures distinctive features of the underlying phenomena of uncertainty.

Models of uncertainty can be classified according to the principle whereby the model imposes structure upon the bubbly confusion of real events. Very briefly, *probability* organizes observations according to the concept of likelihood (in a frequentist interpretation) or according to degree of belief (for bayesians and some others). Probability theory organizes the events in the observation space into sets, to which are ascribed numerical probabilities. Events, represented by sets, become ordered (in many overlapping ways) according to their probabilities.

The most prevalent modification of probability theory to have yet emerged is *fuzzy logic,* which can be interpreted in different ways. One interpretation is in terms of the 'possibility' rather than the 'probability' of events. The distinction, which arises from a modification of Kolmogorov's axiom of additivity [9, 12, 15], hits at the comprehensiveness of mutually exclusive events. We quite reasonably think that an event A and its complement $\overline{A}$ cover the entire event space. In probability theory this motivates the result that the probability of A plus the probability of $\overline{A}$ must equal unity. However, when we consider the *possibility* of uncertain outcomes the situation is different. While A and $\overline{A}$ are indeed disjoint and comprehensive, they may both be quite possible. For instance, in a transition season in which both rain and sunshine come and go with great and unpredictable suddenness, both the proposition 'It will rain this afternoon' and the proposition 'It will not rain this afternoon' may be quite possible. That is, one's degree of belief in the *possibility* of rain may be quite high, and at the same time one's degree of belief in the possibility of no rain may also be quite high. The combined possibility of these two disjoint and comprehensive events exceeds the unit possibility of the entire event space. Fuzzy logic thus captures a distinctive feature of uncertainty which is quite absent from the theory of probability.

Both probability and fuzzy-logic models of uncertainty are quite information-intensive. They both entail real-valued measure functions defined on the space of events which express either a probability or a possibility (or some similar attribute) for each event-set in the space. For instance, this measure function applies to rare and unusual events as well as to more ordinary ones, some of the implications of which we have discussed in section 7.4. In situations of severe uncertainty it may be infeasible to verify either a probability or a fuzzy-logic uncertainty model. What characterizes severe uncertainty is the large gap between what *is known* and what *needs to be known* in order to perform an optimal analysis. *Information-gap models* of uncertainty have been developed to quantify severe uncertainty

[2, 6]. Info-gap models are, almost inevitably, informationally sparser than either probability or fuzzy-logic models. However, info-gap models provide a basis for analysis and decision without introducing the assumptions, especially about the form of a measure function, needed in order to implement richer uncertainty models. Axiomatically, info-gap models are utterly distinct from probability models [5].

The central organizing principle by which an info-gap model of uncertainty imposes structure on the event space is the idea of clustering. An info-gap model is an unbounded family of nested sets, where set-inclusion expresses the clustering of similar but uncertain events. The level of nesting expresses a level of uncertain variation. Each element of each set in the family is a possible realization of the uncertain phenomenon. Each set embodies a range of possible fluctuation at its level of uncertainty. The family of uncertainty-sets is unbounded, entailing an unknown horizon of uncertainty. No measure function exists in an info-gap model; neither likelihood nor possibility is expressed.

In choosing an uncertainty model the analyst must fit his model to his information, and not the other way around. It is not sufficient to proceed directly to the task of choosing the parameters of ones' favorite uncertainty model. The organizing principle of the model, its most fundamental structural attributes, must match the data, knowledge and understanding which the model is meant to quantify.

7.6 Rationality

The selection of an uncertainty model is sometimes made by habit, though no one would condone such a mechanism of choice without at least some qualification. To avoid choice by habit, or to augment choice by structural considerations as discussed in section 7.5, one may choose an uncertainty model to match one or more principles of rationality. This is especially attractive in the planning of experiments or projects, before actual data are available, where considerations of the structure of available knowledge don't apply since the data have yet to be obtained.

There is nothing more scientific, in the positive sense of the word, than choosing ones' model to match ones' principles of rationality. Principles abound, such as maximum-likelihood, minimum mean-square error, maximum expected utility, minimum risk, maximum entropy, minimum vulnerability, maximum robustness and so on. Each principle generates solutions which are optimal with respect to that principle. The catch is that great-sounding principles may turn sour in practice. In discussing the Allais paradox in section 7.2 we noticed that a very reasonable probabilistic postulate — that the expectations of utilities can be combined linearly — does not always describe the behavior of willful and self-interested human decision makers. Here is a case of good intentions gone wrong: a principle of rationality which leads to irrational (contradictory) predictions and recommendations.

First principles are chosen by a feeling of reasonableness, by a sense of what is right, and not by analysis, and herein lies the danger. What is reasonable is not proven, we do not even ask for proof of what seems self-evident. But nothing is more potent than a guiding principle, as we have seen in the examples discussed in sections 7.1–7.4. The line is fine indeed between dogmatism and principled decision-making.

7.7 Conclusion

We have outlined pitfalls and paradoxes, structural limitations and rationality constraints, not in order to deter the application of probability theory but rather to enhance the effectiveness of its use. But furthermore, we have emphasized the fundamental distinction between the phenomena of uncertainty which surround us in the natural and technological worlds, and the mathematics with which we model and manage that uncertainty. Since the phenomena are wondrously complex and subtle, the models must likewise be diverse in their concepts and structures. Mastery of this diversity is the only hope for competently handling the greatest of all challenges in signal processing and system analysis: managing the unknown.

Acknowledgement

The author acknowledges with pleasure the support of the Samuel Neaman Institute for Advanced Studies in Science and Technology. This article was completed while the author was a visiting professor at the Georgia Institute of Technology.

8 References

1. Bayes T. An essay toward solving a problem in the doctrine of chances, *Philosophical Transactions of the Royal Society*, London, 1763.

2. Ben-Haim Y. *Robust Reliability in the Mechanical Sciences,* Berlin: Springer-Verlag, 1996.

3. Ben-Haim Y. Reliability of vibrating structures with uncertain inputs, *Shock and Vibration Digest*, 1998, 30:106–113.

4. Ben-Haim Y. Design certification with information-gap uncertainty. *Structural Safety*, 1999, 21:269–289.

5. Ben-Haim Y. Set-models of information-gap uncertainty: Axioms and an inference scheme, *Journal of the Franklin Institute*, 1999, 336:1093–1117.

6. Ben-Haim Y. *Information-gap Decision Theory: Decisions Under Severe Uncertainty,* London: Academic Press, 2001.

7. Ben-Haim Y. Info-gap value of information in model up-dating, *Mechanical Systems and Signal Processing,* 2001, 15:457–474.

8. Ben-Haim Y. Scott Cogan and Laëtitia Sanseigne, Usability of Mathematical Models in Mechanical Decision Processes, *Mechanical Systems and Signal Processing*, 1998, 12:121–134.

9. Blockley D.I. Fuzziness and probability: A discussion of Gaines' Axioms, *Civil Engineering Systems*, 1985, 2:195–200.

10. Carroll L., *Pillow Problems,* 1895, Re-issued by Dover Press, New York, 1958.

11. DeGroot M.H. *Probability and Statistics,* 2nd ed. New York: Addison-Wesley, 1986.

12. Dubois D. and Prade H. *Possibility Theory: An Approach to Computerized Processing of Uncertainty,* New York: Plenum Press, 1988.

13. Ellsberg D. Risk, ambiguity and the Savage axioms, *Quarterly J. of Economics*, 1961, 75:643–669.

14. Feller W. *An Introduction to Probability Theory and Its Applications,* vol. 1, 3rd ed., New York: Wiley, 1968.

15. Gaines B.R. Fuzzy and probability uncertainty logics, *Information and Control*, 1978, 38:154–169.

16. Hacking I. *The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference.* U.K.: Cambridge University Press, 1975.

17. Jorland G. The Saint Petersburg Paradox: 1713–1937, In: L. Krüger, L.J. Daston and M. Heidelberger, eds., *The Probabilistic Revolution,* vol. 1: Ideas in History, U.S.: MIT Press, 1987, pp.156–190.

18. Keynes J.M. *A Treatise on Probability.* London: Macmillan, 1921.

19. Knight F.H. *Risk, Uncertainty and Profit.* Houghton Mifflin Co., 1921, Re-issued by University of Chicago Press, 1971.

20. Kolmogorov A.N. *Foundations of the Theory of Probability,* 2nd English ed., New York: Chelsea Publishing Co., 1950.

21. Kyburg H.E. jr., Getting fancy with probability, *Synthese,* 1990, 90: 189–203.

22. Machina M. Choice under uncertainty: Problems solved and unsolved, *Journal of Perspectives,* 1987, 1:121–154.

23. Mas-Colell A, Whinston M.D. and Green J.R., *Microeconomic Theory,* U.K.: Oxford University Press, 1995.

24. Mises R. von *Probability, Statistics and Truth,* 2nd English ed., New York: Dover, 1957.

25. Neumann J. von and Morgenstern O. *Theory of Games and Economic Behavior.* 3rd ed., U.S.: Princeton University Press, 1953.

26. Savage L.J. Subjective Probability and statistical practice, In: G.A. Barnard and D.R. Cox, eds., *The Foundations of Statistical Inference: A Discussion.* London: Methuen, 1962.

27. Oberkampf WL, Helton JC, Joslyn CA, Wojtkiewicz SF, and Ferson S., Challenge problems: Uncertainty in system response given uncertain parameters (this issue).

28. Shackle G.L.S. *Epistemics and Economics: A Critique of Economic Doctrines,* Cambridge University Press, 1972, re-issued by Transaction Publishers, U.S., 1992.

29. Stark H and Woods J.W. 1994, *Probability, Random Processes and Estimation Theory for Engineers,* U.S.: Prentice Hall.

30. Tikhonov A.N. and Arsenin V.Y. *Solutions of Ill-posed Problems,* translated by Fritz John, New York: Wiley. 1977.