Working Paper SM-13-1

Risk Analysis 101 Rhetoric in risk analysis, Part III: Much Ado About Proxy Probability of Survival*

Moshe Sniedovich
Department of Mathematics and Statistics
University of Melbourne
Melbourne Vic 3010, Australia
moshe@ms.unimelb.edu.au

February 10, 2013
Last update: July 6, 2013

Preface

In the article *"Robust satisficing and the probability of survival"* that was published recently in the peer-reviewed journal *International Journal of Systems Science,* we find this statement which attributes info-gap decision theory the ability to maximize the probability of success without using probabilistic information:

> We present three propositions, based on info-gap decision theory (Ben-Haim 2006), that identify conditions under which the probability of success is maximised by an agent who robustly satisfices the outcome without using probabilistic information.
>
> Ben-Haim (2013, p. 1)

This statement echoes earlier pronouncements in the literature on info-gap decision theory that not only pit info-gap decision theory against *Probability Theory,* but go so far as to assert that info-gap decision theory provides no less than a 'replacement theory' for *Probability Theory* itself. For instance:

> Probability and info-gap modeling each emerged as a struggle between rival intellectual schools. Some philosophers of science tended to evaluate the info-gap approach in terms of how it would serve physical science in place of probability. This is like asking how probability would have served scholastic demonstrative reasoning in the place of Aristotelian logic; the answer: not at all. But then, probability arose from challenges different from those faced the scholastics, just as the info-gap decision theory which we will develop in this book aims to meet new challenges.
>
> Ben-Haim (2001 and 2006, p. 12)

> Info-gap decision theory clearly presents a replacement theory with which we can more fully understand the relation between classical theories of uncertainty and uncertain phenomena themselves.
>
> Ben-Haim (2001 p. 305; 2006, p. 343)

*This article was written for the **Risk Analysis 101 Project** to provide a Second Opinion on pronouncements made in the info-gap literature on the use of info-gap robustness as a proxy for a probability of survival.

If you wonder on what grounds are these audacious statements made—after all, it is not very often that one comes across statements such as these in the scientific literature—let me assure you straightaway, that it does not take much to discover that the hard facts about info-gap decision theory tell a different story altogether!

My objective is then to demonstrate the yawning gap between the hard facts about info-gap decision theory, and the rhetoric attributing it the ability to maximize the probability of success without using probabilistic information, and to thereby show that this rhetoric is nothing short of farcical.

To do this, I begin by reminding the reader of the long established fact that certain highly specialized probabilistic problems have **equivalent** non-probabilistic counterparts. In fields such as *operations research, stochastic optimization* and *robust optimization,* these equivalent non-probabilistic problems are often called **deterministic equivalents** (e.g. Charnes and Cooper 1963, Charnes et al. 1964, Prékopek 1995, Birge and Louveaux 1997, Gassmann and Ziemba 2013). The idea of obtaining equivalent non-probabilistic problems from probabilistic problems in the framework of decision-making models goes back to at least the 1960s, thus predating info-gap decision theory by some forty years or so. For instance, consider the following (emphasis added):

> Chance constrained programming admits random data variations and permits constraint violations up to specified probability limits. Different kinds of decision rules and optimizing objectives may be used so that, under certain conditions, a programming problem (not necessarily linear) can be achieved that is deterministic—in that all random elements have been eliminated. Existence of such '**deterministic equivalents**' in the form of specified convex programming problems is here established for a general class of linear decision rules under the following 3 classes of objectives (1) maximum expected value ('E model'), (2) minimum variance ('V model'), and (3) **maximum probability** ('P model'). Various explanations and interpretations of these results are supplied along with other aspects of chance constrained programming For example, the 'P model' is interpreted so that H A SIMON'S suggestions for 'satisficing' can be studied relative to more traditional optimizing objectives associated with 'E' and 'V model' variants.
>
> Charnes and Cooper (1963, p. 18)

For our discussion, the 'P model' is of particular interest as it will enable us to bring to light the rationale behind the proposition to **maximize the probability of success** by means of a **non-probabilistic model.** It is important to emphasize, though, that in essence, this rationale is so rudimentary that generally it is taken for granted. Namely, one generally does not bother to elaborated it explicitly, as it is taken to be implied directly by the foundational axioms of *probability theory* (Feller 1968, Rozanov 1977). Still, for our purposes, it is a good idea to state it formally:

> **A Well Known Fact:**
> Let Y be a random variable, c be a numeric scalar, and let $P(Y \leq c)$ denote the probability of the event "$Y \leq c$". Then $P(Y \leq c)$ is non-decreasing with c, namely
>
> $$a < b \longrightarrow P(Y \leq a) \leq P(Y \leq b).$$

Note how the seemingly "non-probabilistic" information, namely "$a < b$", translates to the critically important "probabilistic information", "$P(Y \leq a) \leq P(Y \leq b)$".

And to see how this is put to work, consider the following probabilistic problem and the deterministic equivalent counterpart that clearly derives from it, on grounds of the above well known fact:

Probabilistic problem	Deterministic equivalent
$p^* := \max_{x \in X} \ P(Y \leq h(x))$	$h^* := \max_{x \in X} \ h(x)$

(1)

where h is a real-valued function on X.

For the record, then

THEOREM 1 *Any optimal solution to the deterministic equivalent in (1) is also an optimal solution to the probabilistic problem in (1). And if $P(Y \leq c)$ is strictly increasing with c, then the two problems have the same optimal solutions: $x' \in X$ is optimal with respect to the probabilistic problem iff it is optimal with respect to the deterministic equivalent.*

This basic results can be refined by allowing the random variable Y to be a function of a random n-vector W and x. So assume that $Y = t(x, W)$ where W is a random n-vector and t is a real-valued function on $X \times \mathbb{R}^n$, where $\mathbb{R}$ denotes the real line, such that the probability distribution function of $t(x, W)$ is **independent** of x. Thus, the performance constraint is of the form $t(x, w) \leq h(x)$ where $w \in \mathbb{R}^n$ denotes the realization of W. It follows then that in this case the probabilistic problem and its deterministic equivalent counterpart are as follows:

Probabilistic problem	Deterministic equivalent
$p^* := \max_{x \in X} P(t(x, W) \leq h(x))$	$h^* := \max_{x \in X} h(x)$

(2)

To illustrate, consider the constraint

$$a(x) + d(x)g(w) \geq r \tag{3}$$

where a and d are real-valued function on X, r is a given numeric scalar, g is a real-valued function on $\mathbb{R}^n$, and w is a realization of a random n-vector W. Assume that $d(x) > 0, \forall x \in X$.

By inspection, this constraint can be rewritten as follows:

$$-g(w) \leq \frac{a(x) - r}{d(x)} \tag{4}$$

hence,

Probabilistic problem	Deterministic equivalent
$p^* := \max_{x \in X} P(a(x) + d(x)g(W) \geq r)$	$h^* := \max_{x \in X} \dfrac{a(x) - r}{d(x)}$

(5)

Observe that, in this case, the random variable that yields the "standard" form $P(Y \leq h(x))$ is $Y = t(x, W) = -g(W)$, and $h(x) = \dfrac{a(x) - r}{d(x)}$.

The point to note here, however, is that the format $t(x, w) \leq h(x)$ and the requirement that the probability distribution function of $Y = t(x, W)$ be independent of x are **extremely restrictive.**

But most of all, as we just saw, in cases such as this the deterministic equivalent can be derived by inspection **directly from the performance constraint under consideration**. Which means of course that the derivation of the above deterministic equivalent **has got nothing whatsoever to do with info-gap decision theory!**

About the title of the article

Indeed, once you examine how the derivation/formulation of deterministic equivalents plays out in the case of info-gap decision theory, specifically in the case where the three propositions referred to in the statement quoted above from Ben-Haim (2013) hold, it becomes clear that all this hullabaloo about info-gap decision theory enabling you to "maximize the probability of success without using probabilistic information" turns out to be a classic case of "much ado about nothing".

To appreciate this fact, you need to keep in mind that the performance constraint employed by info-gap decision theory is of the form $f(x, w) \leq b$, where b is a given numeric scalar and

f is a real-valued function on $X \times \mathbb{W}$, where $\mathbb{W}$ denotes the *uncertainty set,* namely the set of possible/plausible values of the uncertainty parameter w. Moreover, you need to take note that info-gap decision theory's formulation of deterministic equivalents, as made clear, for instance in Ben-Haim (2013), **is predicated on** the performance constraint $f(x,w) \leq b$ having the following properties:

- The uncertainty parameter w is a numeric **scalar**, namely $\mathbb{W} \subseteq \mathbb{R}$.
- The performance function $f(x,w)$ is **monotone** with w.

Once this is clear, you realize that in situations such as this, the probabilistic robustness problem **has to begin with a trivially obvious deterministic equivalent.** Namely, the deterministic equivalent can, **without any appeal to info-gap decision theory,** or any other theory for that matter, be deduced from the probabilistic robustness problem itself, **by inspection**, to thereby render the heavy guns that info-gap decision theory brings into play for this purpose **comically redundant.**

To see that this is so, I reiterate that without any loss of generality, the assumption is that w is a **scalar** and for each $x \in X$ the value of $f(x,w)$ is **non-decreasing** with w. Clearly then, under these conditions, to maximize the value of $P(f(x,W) \leq b)$ over $x \in X$, we simply select an $x \in X$ that admits the largest value of $w \in \mathbb{W}$, subject to $f(x,w) \leq b$. In other words, in this case, to obtain the equivalent $t(x,w) \leq h(x)$ format we simply set:

$$t(x,w) \equiv w \tag{6}$$

$$h(x) = \max_{w \in \mathbb{W}} \{w : f(x,w) \leq b\} \ , \ x \in X. \tag{7}$$

Therefore, **by inspection,** we have:

Probabilistic problem	Deterministic equivalent
$p^* := \max\limits_{x \in X} \ P(f(x,W) \leq b)$	$h^* := \max\limits_{x \in X} \ h(x)$ $= \max\limits_{x \in X, w \in \mathbb{W}} \ \{w : f(x,w) \leq b\}$

(8)

Assumption: w is a numeric scalar and $f(x,w)$ is non-decreasing with w.

The trouble is, of course, that an appeal to info-gap decision theory's heavy guns for this purpose in fact conceals from view how **trivially** simple this derivation is. And the bigger trouble is that in the literature on info-gap decision theory, for example in Ben-Haim (2013), all this lies buried under piles of rhetoric.

The end result is that, not only do readers get a totally distorted picture of info gap decision theory's claimed ability to "maximize the probability of success without using probabilistic information". Readers are misled to believe that this is an inherent capability of this theory when the facts are that the existence and deduction of deterministic equivalents have got nothing whatsoever to do with info-gap decision theory.

Hence, the title of this article.

Melbourne
Australia, *The Land of the Black Swan*
July 6, 2013

Moshe

1 Introduction

Over the past nine years, I had many occasions to discuss with risk analysts the proposition to use a **non-probabilistic** decision theory, such as info-gap decision theory, as a tool for *maximizing the* **probability** *of survival,* or *success.* I also kept abreast with publications advocating this idea in a number of peer-reviewed journals, e.g. *Medical Hypotheses*, *International Journal of Systems Science,* and *Risk Analysis: An International Journal.*

So, in this article I share with the reader some of my conclusions about this proposition, notably the manner in which it is examined/discussed in the literature on info-gap decision theory. To set the stage, I suggest that you ask yourself the following question:

> How can a **non-probabilistic** decision theory, such as info-gap decision theory, possibly enable the **maximization of the probability** of survival/success?

I imagine that your immediate hunch would probably be that such a proposition

- may well be too good to be true; or that
- it must have a touch of alchemy about it; or that
- the probabilistic problems under consideration are **trivial.**

To illustrate the last point, consider the following performance constraint

$$f(x,w) \leq b \tag{9}$$

where $x \in X$ represents a *decision variable*, $w \in \mathbb{R}^n$ represents an *uncertainty parameter,* b is a given numeric scalar, f is a real-valued function on $X \times \mathbb{R}^n$, and $\mathbb{R}$ denotes the real line.

Now, suppose that the uncertainty parameter w is in fact a realization of some *random n-vector* W, and the objective is to maximize the probability that this performance constraint is satisfied. The probabilistic problem that we would have on our hands is then this:

$$\textbf{Probabilistic Problem:} \quad p^* := \max_{x \in X} \; P(f(x,W) \leq b). \tag{10}$$

To complicate matters, assume, **as info-gap decision theory indeed does**, that the probability measure P is **completely unknown.**

The question is then this: how does one tackle this probabilistic problem considering the uncongenial, indeed forbidding, conditions that it is subject to?

One avenue of approach is to attempt to derive for it a **deterministic equivalent**: a "deterministic" optimization problem whose optimal solutions are also optimal for the probabilistic problem (e.g. Charnes and Cooper 1963, Charnes et al. 1964, Prékopek 1995, Birge and Louveaux 1997, Gassmann and Ziemba 2013).

And yet, however naturally obvious this approach appears at first, it must be realized, given our vast experience of the last fifty years or so, that this approach is applicable only in very simple, namely highly specialized, situations. Practically speaking, this means that this approach cannot be applied to the probabilistic problem as it is stated above, unless **simplifying conditions** are impose on the problem. Indeed, a considerable simplification in the above probabilistic problem would be required for a **deterministic equivalent** to exist.

The discussion on this idea in the *Preface* points in the direction of a simplification that would be obtained by reasoning along the lines outlined in the next subsection.

1.1 A class of trivial probabilistic problems

Suppose that there is a real-valued function t on $\mathbb{R}^n$ and a real-valued function h on X such that the performance constraint $f(x,w) \leq b$ is equivalent to the constraint $t(w) \leq h(x)$, namely

$$\forall (x,w) \in X \times \mathbb{R}^n : \; t(w) \leq h(x) \longleftrightarrow f(x,w) \leq b \tag{11}$$

meaning that a pair $(x,w) \in X \times \mathbb{R}^n$ satisfies the constraint $f(x,w) \leq b$ iff it satisfies the constraint $t(w) \leq h(x)$.

Note that if this assumption holds, then the probabilistic problem under consideration, namely (10), can be written as follows:

$$p^* := \max_{x\in X} \ P(f(x,W) \leq b) \tag{12}$$

$$= \max_{x\in X} \ P(t(W) \leq h(x)) \tag{13}$$

$$= \max_{x\in X} \ P(Y \leq h(x)) \ , \ Y = t(W). \tag{14}$$

And since $P(Y \leq c)$ is non-decreasing with c, it follows that

$$h^* := \max_{x\in X} \ h(x) \tag{15}$$

is a deterministic equivalent of (10). In short, here we have:

Probabilistic problem	Deterministic equivalent
$p^* := \max_{x\in X} \ P(f(x,W) \leq b)$	$h^* := \max_{x\in X} \ h(x)$

(16)

Simplicity itself.

Example

Consider the rather exotic performance constraint

$$\alpha(x) + [\beta(x)]^{\gamma(w)} \geq b \tag{17}$$

where α and β are real-valued function on X and γ is a real-valued function on $\mathbb{R}^n$. Assume that $\beta(x) > 1$ and $b > \alpha(x)$ for all $x \in X$.

Note that under these conditions, this performance constraint can be re-written as follows:

$$-\gamma(w) \leq -\frac{\ln(b - \alpha(x))}{\ln(\beta(x))}. \tag{18}$$

Hence, we can set $t(w) = -\gamma(w)$ and $h(x) = \dfrac{-\ln(b - \alpha(x))}{\ln(\beta(x))}$. So, in this case we have:

Probabilistic problem	Deterministic equivalent
$p^* := \max_{x\in X} \ P\left(\alpha(x) + \beta(x)^{\gamma(W)} \geq b\right)$	$h^* := \max_{x\in X} \ \dfrac{-\ln(b - \alpha(x))}{\ln(\beta(x))}$

(19)

Again, simplicity itself.

The next example represents a raft of typical probabilistic problems that have been known in the fields of *operations research* and *stochastic optimization* ever since the concept of formulating "deterministic equivalents" was introduced in the 1960's (e.g. Charnes and Cooper 1963, Charnes et al. 1964, Prékopek 1995, Birge and Louveaux 1997, Gassmann and Ziemba 2013). This example also serves to show that "deterministic equivalents" of probabilistic problems can also be used to good effect in situations where the probabilities of the events of interests are known.

To illustrate, let us generalize somewhat the representation $t(w) \leq h(x)$ by allowing t to depend on x. Thus, consider the case where the constraint $f(x,w) \leq b$ can be written as $t(x,w) \leq h(x)$ where t is a real valued function on $X \times \mathbb{R}^n$.

However, in this case, $P(t(x,W) \leq c)$ is required to be **independent** of x. In other words, the cumulative probability distribution function of the random variable $Y = t(x,W)$ is required to be independent of x. This condition ensures that $P(t(x,W) \leq h(x))$ is non-decreasing with $h(x)$, thereby ensuring that the deterministic equivalent would remain $\max_{x\in X} h(x)$.

Example

Suppose that the decision variable $x \in X$ and the random n-vector W interact[1] in such a way that $f(x,w) = g(x) + v(x,w)$, where g is a real-valued function on X and $v(x,w)$ is the realization of a *normal random variable* with mean $\mu(x)$ and variance $\sigma^2(x) > 0$, namely $v(x,W) \sim \mathcal{N}(\mu(x), \sigma^2(x))$.

Thus, the performance constraint $f(x,w) \leq b$ can be written as follows:

$$v(x,w) \leq b - g(x). \tag{20}$$

Now, this expression can be normalized/standardized, in the usual manner, by means of the mean $\mu(x)$ and the variance $\sigma^2(x)$, to obtain the following equivalent normalized/standardized constraint:

$$\frac{v(x,w) - \mu(x)}{\sigma(x)} \leq \frac{b - g(x) - \mu(x)}{\sigma(x)}. \tag{21}$$

Hence, we can set

$$t(x,w) = \frac{v(x,w) - \mu(x)}{\sigma(x)} \tag{22}$$

$$h(x) = \frac{b - g(x) - \mu(x)}{\sigma(x)} \tag{23}$$

observing that since $v(x,W)$ is a *normal random variable,* it follows that $t(x,W)$ is a *standard normal random variable,* namely $t(x,W) \sim \mathcal{N}(0,1)$, hence its distribution function is independent of x. Consequently:

Probabilistic problem	Deterministic equivalent
$p^* := \max_{x \in X} P(g(x) + v(x,W) \leq b) \mid_{v(x,W) \sim \mathcal{N}(\mu(x),\sigma^2(x))}$	$h^* := \max_{x \in X} \frac{b - g(x) - \mu(x)}{\sigma(x)}$

(24)

Again, simplicity itself.

I shall not go into the well familiar procedure of using *location* and/or *scaling* parameters of probability distribution functions to derive "standard" random variables. Suffice it to say that in such cases, one can sometimes pre-process the performance constraint $f(x,W) \leq b$ so as to obtain an equivalent constraint of the form $\xi \leq h(x)$ where ξ denotes the "standard" random variable of the family of random variables under consideration. In such cases the deterministic equivalent would be $\max_{x \in X} h(x)$.

DEFINITION 1 *Consider the performance constraint $f(x,w) \leq b$ discussed above. We say that this constraint is* TRIVIAL *if it submits to the representation $t(w) \leq h(x)$, as discussed above, or the representation $t(x,w) \leq h(x)$, as discussed above, such that $P(t(x,W) \leq c)$ is independent of x. In either case, we refer to the probabilistic problem* $\max_{x \in X} P(f(x,w)) \leq b)$ *as* TRIVIAL.

THEOREM 2 *If the constraint $f(x,w) \leq b$ is* TRIVIAL, *then any decision that maximizes $h(x)$ over $x \in X$ also maximizes $P(f(x,W) \leq b)$ over $x \in X$, that is,*

Probabilistic problem	*Deterministic equivalent*
$p^* := \max_{x \in X} P(f(x,W) \leq b)$	$h^* := \max_{x \in X} h(x)$

(25)

Pure and simple!

[1]For example, consider the case where x is also an n-vector and $f(x,w) = x^T w$. If W is a vector of n stochastically independent normal random variables, then $f(x,w)$ is a normal random variable.

Remark

That said, I want to clarify my usage of the term TRIVIAL in this context. Take note then that this term does not suggest that in cases where the representation $t(w) \le h(x)$ exists, it is always trivially easy to derive it from $f(x, w) \le b$. Rather, its function is to indicate that, however complicated and unappealing the constraint $f(x, w) \le b$ might appear to be, once you are motivated by the proposition to tackle the problem considered via a"deterministic equivalent", you realize that this constraint can essentially be boiled down to a generic constraint of the simple form $y \le h(x)$ where y represents the realization of a (real-valued) random variable Y. Once this is done, the task of deriving a deterministic equivalent for the probabilistic problem $\max_{x \in X} P(Y \le h(x))$, hence of $\max_{x \in X} P(f(x, W) \le b)$, turns out to be a veritably trivial exercise.

In short, the deterministic equivalents under consideration here are derived from the probabilistic problems **by inspection,** by capitalizing on the well known fact that $P(Y \le c)$ is non-decreasing with c. Specifically, these deterministic equivalents can be derived, by inspection, directly from the performance constraints themselves.

I need hardly point out, though, that the label TRIVIAL by no means implies that the optimization problem defined by the deterministic equivalent $\max_{x \in X} h(x)$ is trivial in the sense that it is easy to solve. Indeed, for all we know, it may well turn out to be extremely difficult to solve.

In the next section I take a quick look at a number of specific performance constraints. Each and every one of these constraints is featured in the literature on info-gap decision theory. And as you might have guessed, each and every one of these constraints is indeed TRIVIAL in the above sense.

2 A gallery of trivial probabilistic problems

The objective of this section is to demonstrate that the probabilistic robustness problems for which info-gap decision theory provides deterministic equivalents are manifestly trivial. The main point here is that it is patently obvious in these cases, that it is straightforward to deduce the $t(w) \le h(x)$ representation or the $t(x, w) \le h(x)$ representation, directly from the constraint $f(x, w) \le b$. The question therefore arising is this: what is the point, the merit, indeed the justification, for using info-gap decision theory's sledgehammer approach for this purpose?

To allow the reader to easily relate the analysis here to the analysis in the info-gap literature, I adopt in each case the notation used in the info-gap literature. The corresponding $f(x, w) \le b$ and $t(w) \le h(x)$ formats used in this article should be obvious.

2.1 Example: foraging behavior problem

This is a recent version of an info-gap model that made its debut appearance in an article by Carmel and Ben-Haim (2005). Here I consider the version discussed in Ben-Haim (2013), where the performance constraint is formulated thus:

$$tg_0 + (T - t)g_1 \ge G_c \tag{26}$$

where $t \in [0, T)$ represent the decision variable, G_c and g_0 are given numeric scalars, and the numeric scalar $g_1 \ge 0$ represents the uncertainty parameter.

By inspection, the performance constraint can be written as follows:

$$g_1 \le \frac{tg_0 - G_c}{T - t}. \tag{27}$$

So if we let G_1 denote the random variable governing the values of g_1, then, by inspection,

Probabilistic problem	Deterministic equivalent
$p^* := \max\limits_{0 \le t < T} P(tg_0 + (T - t)G_1 \ge G_c)$	$h^* := \max\limits_{0 \le t < T} \dfrac{tg_0 - G_c}{T - t}$

(28)

Nothing to it!

2.2 Example: Uncertain non-linear oscillator problem

Ben-Haim and Cogan (2011) consider the following performance constraint:

$$|X_c(q)| - |X_r(c)| \geq \delta \tag{29}$$

where $q \in Q$ represents the decision variable, c represents the uncertainty parameter and $|X_c(q)|$ and $|X_r(c)|$ are positive numeric scalars that depend on q and c, respectively.

By inspection, this constraint can be written as follows:

$$|X_r(c)| \leq |X_c(q)| - \delta. \tag{30}$$

Let C denote the random variable governing the value of c. Then, by inspection

Probabilistic problem	Deterministic equivalent
$p^* := \max_{q\in Q} \ P(\lvert X_c(q)\rvert - \lvert X_r(C)\rvert \geq \delta)$	$h^* := \max_{q\in Q} \lvert X_c(q)\rvert$

(31)

Again, nothing to it!

2.3 Example: Risky assets

Ben-Haim (2013, Appendix C) examines the performance constraint:

$$u(w - p^T r) + u(r^T q) \geq G_c \tag{32}$$

where $r \in \mathbb{R}^n$ represents the decision variable, $q \in \mathbb{R}^n$ represents the uncertainty parameter, w and G_c are given numeric scalars, p is a given n-vector, and u is a strictly increasing real-valued function on the real line. The admissible values of r are constrained by $0 \leq p^T r \leq w$ and $r_j \geq 0, j = 1, 2, \dots, n$. Let u^{-1} denote the inverse of u.

By inspection, we can rewrite this constraint as follows:

$$r^T q \geq u^{-1}(G_c - u(w - p^T r)). \tag{33}$$

According to the assumptions made in Ben-Haim (2013, p. 17), $r^T q$ is a realization of a *normal random* variable with mean $m(r) = r^T \mu$ and variance $s^2(r) = r^T \Sigma r > 0$. We can therefore standardize the constraint as follows:

$$-\frac{r^T q - r^T \mu}{r^T \Sigma r} \leq \frac{r^T \mu - u^{-1}(G_c - u(w - p^T r))}{r^T \Sigma r} \tag{34}$$

observing that according to the above assumptions,

$$t(r, q) := \frac{r^T q - r^T \mu}{r^T \Sigma r} \tag{35}$$

is the realization of a *standard normal random variable,* hence the probability function governing the value of $t(r, q)$ is independent of r. Therefore, by inspection

Probabilistic problem $r^T q \sim \mathcal{N}(r^T\mu, r^T\Sigma r)$	Deterministic equivalent
$p^* := \max_{r\in X} \ P\left(u(w - p^T r) + u(r^T q) \geq G_c\right)$	$h^* := \max_{r\in X} \frac{r^T \mu - u^{-1}(G_c - u(w - p^T r))}{r^T \Sigma r}$

(36)

where $X = \{r \in \mathbb{R}^n : 0 \leq p^T r \leq w, r_j \geq 0, j = 1, 2, \dots, n\}$.

Again, nothing to it!

What all this goes to show is that to make a case for the employment of info-gap's robustness model as a tool for *maximizing the* **probability** *of survival,* or *success,* one must answer the following questions.

2.4 Questions

- What justification could there possibly be for complicating the derivation of the deterministic equivalents of probabilistic problems, such as those featured above, by foisting on these probabilistic problems the additional constructs that are required by info-gap decision theory, in order to cast them in terms of an info-gap robustness model (see section 4.1)?
- Indeed, what justification could there possibly be for turning to info-gap decision theory in the first place, given that the deterministic equivalents are obtained immediately by inspection from the performance constraint?
- Isn't it clear as daylight that the simplicity of these examples sends a clear message that the transformation of the constraint $f(x, w) \leq b$ into $t(w) \leq h(x)$ is possible only in highly specialized cases?
  And to illustrate, consider the seemingly simple constraint

  $$(w - x)^3 \leq 100 \tag{37}$$

  where x and w are **numeric scalars,** representing the decision variable and the uncertainty parameter, respectively. However simple this representation seems to be, there is no obvious way to deduce from this constraint the $t(w) \leq h(x)$ format. Indeed, it is doubtful that such a format exists for this case.

In the next section I examine a slightly different scheme for deriving deterministic equivalents for probabilistic problems. This scheme provides further evidence that deterministic equivalents can be derived only in highly specialized cases of the type discussed above.

3 A slightly different perspective

As I noted at the outset, the preceding analysis sought to demonstrate that the ability to deduce deterministic equivalents hangs on the premise (well known fact) that $P(Y \leq c)$ is non-decreasing with c. But to give a fuller picture of why the application of deterministic equivalents is extremely limited in cases where P is completely unknown, it is instructive to explain this fact in terms of another inherent property of probability measures, implied by the axioms of probability theory (Feller 1968, Rozanov, 1977).

Let C and D denote two (non-empty) events (sets) associated with some probability space. Then, as we all know,

$$C \subseteq D \longrightarrow P(C) \leq P(D) \tag{38}$$

where $P(E)$ denotes the probability of event E.

The point to note then is that to be able to count on this inherent property of probability measures, the two sets (events) we compare must be **nested**, namely at least one of these two set must be a subset of the other. For, suppose that C and D are not nested. Then it would follow that $C \setminus D$ and $D \setminus C$ are not empty. Therefore, if P is completely unknown, we cannot discount the possibility that $P(C \setminus D) = 1$, $P(D) = 0$ and therefore that $P(C) > P(D)$. Nor can we discount the possibility that $P(D \setminus C) = 1$, $P(C) = 0$ and therefore that $P(D) > P(C)$. In short, in this case we are in no position to rule whether $P(C) \leq P(D)$ or $P(D) \leq P(C)$.

With this in mind, consider again the performance constraint $f(x, w) \leq b$, and let $A(x)$ denote the set of acceptable values of w associated with decision $x \in X$, namely define

$$A(x) = \{w \in \mathbb{R}^n : f(x, w) \leq b\} \ , \ x \in X. \tag{39}$$

This representation enables writing the constraint imposed on w by decision x as follows:

$$w \in A(x). \tag{40}$$

As usual, to simplify the notation, we shall write $P(A(x))$ instead of $P(W \in A(x))$, observing that $P(A(x)) = P(f(x,W) \le b)$. So here is an obvious deterministic equivalent:

$$\forall x', x'' \in X: \quad A(x') \subseteq A(x'') \longrightarrow P(A(x')) \le P(A(x'')). \tag{41}$$

This suggests the following.

Definition 2 *Two sets, say C and D, are said to be* **nested** *iff either $C \subset D$, or $D \subset C$, or $C = D$. A collection of sets is said to be nested iff every two sets in this collection are nested.*

Definition 3 *Let $SIZE$ be a real-valued function on $\{A(x) : x \in X\}$. We say that $SIZE$ is a* **size function** *for $\{A(x) : x \in X\}$ iff*

$$\forall x', x'' \in X: \quad SIZE(A(x')) < SIZE(A(x'')) \longrightarrow A(x') \subset A(x''). \tag{42}$$

Intuitively, the real number $SIZE(A(x))$ is viewed as the "size" of set $A(x)$. For instance, if the sets $A(x), x \in X$ consist of finitely many elements, we can let $SIZE(A(x)) = |A(x)|$, where $|A(x)|$ denotes the cardinality of $A(x)$. It follows then that

Theorem 3 *Assume that the collection of sets $A(x), x \in X$ is nested and that $SIZE$ is a size function for this collection. Then*

$$\forall x', x'' \in X: \quad SIZE(A(x')) < SIZE(A(x'')) \longrightarrow P(A(x')) \le P(A(x'')). \tag{43}$$

What we have then under these conditions is the following:

Probabilistic problem	Deterministic equivalent
$p^* := \max_{x \in X} P(A(x))$	$h^* := \max_{x \in X} SIZE(A(x))$

(44)

Corollary 1 *Assume that the collection of sets $A(x), x \in X$ is nested and that $SIZE$ is a size function for this collection. Then, any optimal solution to the deterministic equivalent in (44) is also an optimal solution to the probabilistic problem in (44).*

It should be noted that if P is completely unknown and for two decisions, say $x', x'' \in X$ the corresponding sets $A(x')$ and $A(x'')$ are not nested, then it would be impossible to determine categorically whether $P(A(x')) < P(A(x'')$ or $P(A(x')) > P(A(x'')$, or $P(A(x')) = P(A(x'')$. In fact, if $A(x')$ and $A(x'')$ are not nested, it would be possible to construct two probability measures, say P' and P'' such that $P'(A(x')) < P'(A(x'')$ and $P''(A(x')) > P''(A(x'')$. For instance, set $P'(A(x') \setminus A(x'')) = 1$ and $P''\,(A(x'') \setminus A(x')) = 1$.

Thus, for the above scheme for deducing deterministic equivalents to work, the sets $A(x), x \in X$ must be nested and there must be at least one decision $x^* \in X$ such that

$$A(x) \subseteq A(x^*)\ ,\ \forall x \in X. \tag{45}$$

These, needless to say, are extremely restrictive conditions.

To illustrate, consider the very simple, one might argue naive, case where w and x are numeric scalars, $f(x,w) = (w-x)^2$, and $b = 100$. The constraint $f(x,w) \le b$ can therefore be written as $(w-x)^2 \le 100$, or equivalently as $|w-x| \le 10$. Thus,

$$A(x) = \{w \in \mathbb{R} : |w - x| \le 10\}\ ,\ x \in \mathbb{R} \tag{46}$$
$$= [x - 10, x + 10]. \tag{47}$$

Clearly, the sets $A(x), x \in X$ are **not** nested.

3.1 Some veritably trivial problems

Consider the case where $w \in \mathbb{R}$, and that for every $x \in X$ the value of $f(x, w)$ is non-decreasing with w. Also, assume that the values of

$$\overline{w}(x) := \max_{w \in \mathbb{R}} \{w : f(x, w) \leq b\} \ , \ x \in X \tag{48}$$

are attained.

Then, in this case we have

$$A(x) = (-\infty, \overline{w}(x)] \ , \ x \in X \tag{49}$$

and therefore the sets $A(x), x \in X$ are nested.

Furthermore, by inspection, the function $SIZE$ defined by $SIZE(A(x)] := \overline{w}(x)$ is a size function for the sets $A(x), x \in X$. So, for this simple case we have:

$$\begin{array}{c|l} \text{Probabilistic problem} & \text{Deterministic equivalent} \\ \hline p^* := \max\limits_{x \in X} \ P(A(x)) & h^* := \max\limits_{x \in X} \ \overline{w}(x) \\ & \quad\ = \max\limits_{x \in X, w \in \mathbb{R}} \{w : f(x, w) \leq b\} \end{array} \tag{50}$$

If on the other hand, $f(x, w)$ is non-increasing with w, we would have

$$A(x) = [\underline{w}(x), \infty) \ , \ x \in X \tag{51}$$

where

$$\underline{w}(x) := \min_{w \in \mathbb{R}} \{w : f(x, w) \leq b\} \ , \ x \in X \tag{52}$$

assuming that these values are attained.

Note that the sets $A(x), x \in X$ are nested and that $SIZE(A(x)) = -\underline{w}(x)$ is a size function. Hence:

$$\begin{array}{c|l} \text{Probabilistic problem} & \text{Deterministic equivalent} \\ \hline p^* := \max\limits_{x \in X} \ P(A(x)) & h^* := \min\limits_{x \in X} \ \underline{w}(x) \\ & \quad\ = \min\limits_{x \in X, w \in \mathbb{R}} \{w : f(x, w) \leq b\} \end{array} \tag{53}$$

And along the same lines, assume that $w \in \mathbb{R}^n$ and that $f(x, w) = F(x, g(w))$ where g is a real-valued function on $\mathbb{R}^n$ and F is a real-valued function on $X \times \mathbb{R}$. We can let $v = g(w)$ and write the constraint $f(x, w) \leq b$ as $F(x, v) \leq b$, observing that v is a numeric scalar. Thus, subject to the transformation $v = g(w)$, we are back to the two simple cases examined above, namely the case where the uncertainty parameter is a numeric scalar and the constraint function is monotone with this parameter.

3.2 Examples of veritably trivial problems

Recall that the performance constraint associated with the foraging behavior problem discussed in section 2.1 is as follows:

$$tg_0 + (T - t)g_1 \geq G_c \tag{54}$$

where the numeric scalar g_1 is the uncertainty parameter, and $t \in [0, T)$ is the decision variable.

Clearly, for $t \in [0, T)$ the performance level $tg_0 + (T - t)g_1$ is strictly increasing with q_1. Hence, this problem falls squarely in the class of the veritably trivial cases discussed in the preceding section.

The same goes for the uncertain non-linear oscillator problem featured in section 2.2, where the performance constraint is as follows:

$$|X_c(q)| - |X_r(c)| \geq \delta \tag{55}$$

recalling that c represents the uncertainty parameter, and q denotes the decision variable.

By reformulating the uncertain parameter in terms of the scalar $w = |X_r(c)|$, we can write the performance constraint as follows: $|X_c(q)| - w \geq \delta$. Clearly, the performance level $|X_c(q)| - w$ is strictly decreasing with w.

And the same is true for the risky assets problem in section 2.3, where the performance constraints is as follows:

$$u(w - p^T r) + u(r^T q) \geq G_c \tag{56}$$

recalling that $r \in \mathbb{R}^n$ represents the decision variable, $q \in \mathbb{R}^n$ represents the uncertainty parameter, w and G_c are given numeric scalars, p is a given n-vector, and u is a strictly increasing real-valued function on the real line.

By reformulating the uncertainty parameter in terms of the scalar $v = r^T q$, we can write the performance constraint as follows: $u(w - p^T r) + u(v) \geq G_c$. Clearly, the performance level $u(w - p^T r) + u(v)$ is strictly increasing with $u(v)$, and since $u(v)$ is strictly increasing with v, it follows that the performance level $u(w - p^T r) + u(v)$ is also strictly increasing with v.

3.3 Taking stock

The preceding analysis sought to demonstrate that, as can best be ascertained, it is hard to see how deterministic equivalents can be deduced for $\max_{x \in X} P(f(x, W) \leq b)$ in cases where the uncertainty parameter w **is not** a numeric scalar, or can be transformed into a numeric scalar. But more than this, even in this simple case, it seems that the derivation of deterministic equivalents is contingent on the performance level $f(x, w)$ being *monotone* with w, in which case **the derivation would have been trivially obvious to begin with.**

That said, the question is what would happen in the two dimensional case, namely where $w \in \mathbb{R}^2$? Shouldn't it be possible to construct non-trivial two dimensional cases?

Consider then the manifestly simple two dimensional case $w \in \mathbb{R}^2$, where

$$X = \{x \in \mathbb{R}^2 : x_1 + x_2 = 1, x_1, x_2 \geq 0\} \tag{57}$$

and the performance constraint is as follows:

$$x_1 w_1 + w_2 x_2 \leq 20. \tag{58}$$

In this case,

$$A(x) = \{w \in \mathbb{R}^2 : x_1 w_1 + w_2 x_2 \leq 20\} \ , \ x \in X. \tag{59}$$

By inspection, these sets are **not nested.** For instance, consider these three sets

$$A(1, 0) = \{w \in \mathbb{R}^2 : w_1 \leq 20\} = \mathbb{R} \times (-\infty, 20] \tag{60}$$
$$A(0, 1) = \{w \in \mathbb{R}^2 : w_2 \leq 20\} = (-\infty, 20] \times \mathbb{R} \tag{61}$$
$$A(0.5, 0.5) = \{w \in \mathbb{R}^2 : w_1 + w_2 \leq 10\} \tag{62}$$

Clearly, these sets are not nested.

And if this were not enough, then there is also the issue of formulating a size function. The question is how to measure the size of the sets $A(x), x \in X$ in this very simple case, considering that these sets are *half-spaces* of $\mathbb{R}^2$ generated by the hyperplane $x_1 w_1 + w_2 x_2 = 20$ for given values of x_1 and x_2.

3.4 Conclusion

Practically based considerations, as well as methodologically based considerations, seem to suggest that in cases where P is completely unknown, the scope of both the $SIZE(A(x)))$ approach and the $P(t(w) \leq h(x))$ approach to deducing deterministic equivalents for the probabilistic problem considered is limited in the extreme. Indeed, the requirement that the sets of acceptable values of w, namely sets $A(x), x \in X$, be **nested** dictates drastic simplifications in the performance constraint. So much so, that even if we assume that the uncertainty parameter w is a **scalar,** we must assume on top of this that the performance levels $f(x, w)$ are **monotone** with w.

And the upshot of all this is that probabilistic problems of the "veritably trivial type" (such as those examined in this section) seem to be the only type of problem that the $SIZE(A(x)))$ approach and/or the $P(t(w) \leq h(x))$ approach can deal with.

4 A radius of stability perspective

The concept *radius of stability* (circa 1960) is used extensively in many fields to measure the local robustness/stability of systems against **small perturbations** in the nominal value of a parameter (e.g. Wilf 1960, Milne and Reynolds 1962, Hindrichsen and Pritchard 1986, 1986a, Zlobec 1987, 1988, , Paice and Wirth 1998, Anderson et al. 2001, Cooper et al. 2004).

To explain this intuitive concept, let $S(x)$ denote the set of **acceptable** states associated with system/decision $x \in X$ and let $\tilde{s} \in S(x)$ denote a nominal value of the state variable s. Then, the radius of stability of set $S(x)$ at $\tilde{s}$ is the radius (size) of the largest neighborhood around the nominal state that is contained in the set. More formally then, the radius of stability of system/decision $x \in X$ at $\tilde{s}$ is as follows:

$$\widehat{\rho}(x, \tilde{s}) := \max_{\rho \geq 0} \{\rho : s \in S(x), \forall s \in B(\rho, \tilde{s})\} \ , \ x \in X \tag{63}$$

where $B(\rho, \tilde{s})$ denotes a *neighborhood,* or *ball,* of radius (size) ρ around $\tilde{s}$.

For example, in the case of the constraint $f(x, w) \leq b$, we can view w as a state variable and $S(x)$ as the set of acceptable states associated with decision x. The radius of stability of decision x at $\tilde{w}$ would therefore be as follows:

$$\widehat{\rho}(x, \tilde{w}) : = \max_{\rho \geq 0} \{\rho : w \in S(x), \forall w \in B(\rho, \tilde{w})\} \ , \ x \in X \tag{64}$$

$$= \max_{\rho \geq 0} \{\rho : f(x, w) \leq b, \forall w \in B(\rho, \tilde{w})\}. \tag{65}$$

In words, the radius of stability of decision x at $\tilde{w}$ is the radius ρ of the largest neighbourhood $B(\rho, \tilde{w})$ around $\tilde{w}$ all of whose elements satisfy the performance constraint $f(x, w) \leq b$.

This means that in order to use this model to determine the local robustness of decision x at $\tilde{w}$ with respect to the performance constraint $f(x, w) \leq b$, we have to specify two additional constructs:

- The nominal value of w, namely $\tilde{w}$.
- A neighborhood structure $B(\rho, \tilde{w}), \rho \geq 0$ around $\tilde{w}$.

The implication is then that the value of the radius of stability of decision x at $\tilde{w}$, namely the value of $\widehat{\rho}(x, \tilde{w})$, is contingent on the value $\tilde{w}$ and the definition/specification of the neighbourhoods $B(\rho, \tilde{w}), \rho \geq 0$.

4.1 Info-gap proxy theorems

Info-gap decision theory (Ben-Haim 2001, 2006, 2010) uses radius of stability models of the above type to define the robustness of decisions, and its robust-satisficing approach ranks decisions according to the criterion that the larger the robustness the better. Hence, the optimal robust-satisficing decision is that whose robustness (radius of stability) $\widehat{\rho}(x, \tilde{w})$ is the largest.

Now, the contention that info-gap's robust-satisficing approach enables maximizing the probability of success without using probabilistic information is explained in the info-gap literature as being vouchsafed by what, in this literature, are called "proxy theorems" (Ben-Haim 2013). These are theorems that specify conditions under which the ranking of decisions according to their radius of stability (info-gap robustness), is equivalent to their ranking according to their probability to satisfy the performance constraint $f(x,w) \leq b$.

In other words, these theorems stipulate conditions under which $\widehat{\rho}(x,\tilde{w})$ can be used as a proxy for $P(f(x,W) \leq b)$ for the purposes of ranking decisions according to their probability of satisfying the performance constraint.

But the rub is that, as the probabilistic robustness $P(f(x,W) \leq b)$ is inherently a measure of **global** robustness, whereas radius of stability robustness $\widehat{\rho}(x,\tilde{w})$ is inherently a measure of **local** robustness, it is crucial for info-gap's proxy theorems to ensure that these two radically different measures of robustness are "coherent" with each other.

To this end, info-gap's proxy theorems posit some truly stringent conditions so that the non-trivial robustness problems satisfying them are likely to be **rare**.

One of the most comprehensive examination of this topic can be found in Davidovitch (2009) where a distinction is made between *strong proxy theorems* and *weak proxy theorems*. Members of the *strong* class do not impose any requirements on the probability/likelihood structure of the uncertainty space under consideration. So the task of "strong proxy theorems" is to stipulate conditions guaranteeing that the decision selected by info-gap's robustness model maximizes the "probability of success", regardless of the underlying (completely unknown) probability/likelihood structure.

In contrast, the *weak* class imposes some "coherency" conditions on the probability/likelihood structure. These conditions are designed to ensure that the probability/likelihood structure associated with the uncertainty space "mimic" the implicit "distance" function that is used in info-gap's model of uncertainty to create the neighborhoods $B(\rho,\tilde{w}), \rho \geq 0$ around $\tilde{w}$.

It is important to take note that Davidovitch's (2009) overall conclusion is that proxy theorems are expected to be "very rare" (emphasis added):

> We have shown that the definition of strong proxy theorems discussed by Ben-Haim (2007), is **very restrictive,** and that when the uncertainty is multi-dimensional, strong proxy theorems are expected to be **very rare.** Then we shall prove that even this weaker definition does **not** hold for a **wide family of common problems.**
>
> Davidovitch (2009, p. 137)

Since the technical issues are discussed in detail in Davidovitch (2009), I shall not elaborate on them here.

However!

What I must point out here is that an important issue that is central to this entire enterprise is not even broached in Davidovitch (2009), nor in Ben-Haim (2013). It is important therefore to discuss it here in some detail.

If you keep in mind the fundamental difference between **global** and **local** stability/robustness, you would no doubt expect that the conditions stipulated by info-gap's proxy theorems (which essentially aim to establish a tie between the two) would have to be exacting to a degree that the structure of the performance constraint $f(x,w) \leq b$ is bound to become greatly simplified. The question is therefore this:

- Given that the stringent requirements imposed by info-gap proxy theorems have the effect of rendering the performance constrain $f(x,w) \leq b$ **trivially simple,** what is the merit, the point, the advantage, of using info-gap's robustness model as a deterministic equivalent for the probabilistic robustness problem? More generally, what is the point, the merit, the advantage of such proxy theorems say, vis-á-vis say theorems such as Theorem 2, that stipulate a transformation of $f(x,w) \leq b$ into $t(w) \leq h(x)$?

More fundamentally, what is the point of foisting on the probabilistic robustness problem $\max_{x\in X} P(f(x,W)\leq b)$ an arbitrary ad hoc neighborhood structure $B(\rho,\tilde{w}),\rho\geq 0$ when all that this manages to do is to complicate the derivation of deterministic equivalents?

To illustrate this important point, let us go back to the main example in Ben-Haim (2013), namely the Foraging Behavior problem (section 2.1) featuring the following performance constraint

$$tg_0 + (T-t)g_1 \geq G_c \tag{66}$$

where $t\in[0,T)$ represent the decision variable, G_c and g_0 are given numeric scalars, and $g_1\geq 0$ represents the uncertainty parameter.

By inspection, the performance constraint can be written as follows:

$$g_1 \leq \frac{tg_0 - G_c}{T-t}. \tag{67}$$

Let G_1 denote the random variable governing the values of g_1. Then, by inspection

Probabilistic problem	Deterministic equivalent
$p^* := \max\limits_{0\leq t<T} P(tg_0 + (T-t)G_1 \geq G_c)$	$h^* := \max\limits_{0\leq t<T} \dfrac{tg_0 - G_c}{T-t}$

(68)

The same deterministic equivalent is obtained by the $SIZE(P(A(x)))$ approach, observing that in this example the sets

$$\begin{aligned} A(t) &= \left\{ g_1\in\mathbb{R} : g_1 \leq \frac{tg_0-G_c}{T-t}\right\} \ , \ 0\leq t<T & (69)\\ &= \left(-\infty, \frac{tg_0-G_c}{T-t}\right] & (70)\end{aligned}$$

are nested, so that we can set

$$SIZE(A(t)) := \frac{tg_0-G_c}{T-t} \ , \ 0\leq t<T. \tag{71}$$

In sharp contrast, the deterministic equivalent proposed in Ben-Haim (2013) according to the precepts of info-gap decision theory is the following:

$$\begin{aligned} \widehat{\rho}(\tilde{g}_1) &: = \max_{0\leq t<T} \ \widehat{\rho}(x,\tilde{g}_1) & (72)\\ &= \max_{0\leq t<T}\max_{\rho\geq 0} \ \{\rho : tg_0 + (T-t)g_1 \geq G_c, \forall g_1\in B(\rho,\tilde{g}_1)\} & (73)\end{aligned}$$

where the nominal point $\tilde{g}_1$ and the neighborhoods $B(\rho,\tilde{g}_1),\rho\geq 0$ are yet to be determined.

To repeat then: what merit can there possibly be in using info-gap's sledgehammer approach to deal with this problem?

Indeed, what merit can there possibly be in turning to info-gap decision theory to begin with, considering that implementing it requires the specification of ad hoc values for $\tilde{g}_1$ and the neighborhoods $B(\rho,\tilde{g}_1),\rho\geq 0$, that in the end have no impact whatsoever on the results generated by this deterministic equivalent?

The point I want to emphasise is that not only is this alternative deterministic equivalent more complicated than those based on the $P(Y\leq h(x))$ approach and the $SIZE(P(A(x)))$ approach. The truly important point here is that the radius of stability approach (i.e info-gap decision theory's approach) covers up the fact that the probabilistic robustness problem under consideration is a problem of **global** robustness that has an **obvious** deterministic equivalent, derived by inspection on grounds of a basic property of **probability measures.** This point is completely obscured by the radius of stability approach adopted by info-gap decision theory.

5 So what is the bottom line?

The bottom line is that in the framework of the $SIZE(P(A(x)))$ approach and the $P(t(w) \leq h(x))$ approach, the global probabilistic robustness problem is treated from the start as a ... global robustness problem and the derivation of the deterministic equivalents is transparently based on fundamental properties of probability measures. In sharp contrast, the radius of stability approach (i.e. the approach adopted by info-gap decision theory) conceals from view the role that fundamental properties of probability functions play in the derivation of deterministic equivalents.

The point here is that these properties in fact are **very much appealed to** in the framework of the radius of stability approach. The trouble is that this fact gets lost in the mass of technical considerations that need to be reckoned with in order to reconcile the inherently local orientation of the radius of stability model, with the inherently global orientation of the probabilistic robustness problem under consideration.

This means that the statement (emphasis added):

> We present three propositions, based on info-gap decision theory (Ben-Haim 2006), that identify conditions under which the probability of success is maximised by an agent who robustly satisfices the outcome **without using probabilistic information.**
>
> Ben-Haim (2013, p. 1)

is grossly misleading.

Because, to repeat, the deduction of the deterministic equivalents by means of info-gap decision theory's radius of stability approach, most definitely makes use of "probabilistic information" through its implicit appeal to the fundamental properties of probability measures.

Let me clarify this important point.

Suppose that I assert that there is a decision $x^* \in X$ such that $A(x) \subseteq A(x^*)$ for all $x \in X$. Does this proposition count as "probabilistic information"? Similarly, suppose that I assert that there is no $x^* \in X$ such that $A(x) \subseteq A(x^*), \forall x \in X$. Does this proposition count as "probabilistic information"?

Obviously, what I am driving at is that in an analysis of the values of $P(A(x)), x \in X$, the proposition asserting that"there exists an $x^* \in X$ such that $A(x) \subseteq A(x^*), \forall x \in X$" imparts a crucially important piece of "probabilistic information" as it immediately implies that $P(A(x)) \leq P(A(x^*)), \forall x \in X$. Similarly, in the context of such an analysis, the proposition asserting that "there exists no $x^* \in X$ such that $A(x) \subseteq A(x^*), \forall x \in X$" equally imparts a crucially important piece of "probabilistic information" as it immediately implies that there is no assurance that a deterministic equivalent exists for the probabilistic robustness problem under consideration if P is indeed completely unknown.

And to amplify this point, in the context of an analysis of values of $P(Y \leq h(x)), x \in X$, the assertion that "there exists an $x^* \in X$ such that $h(x) \leq h(x^*), \forall x \in X$" is vitally important "probabilistic information". Indeed, the concept of "deterministic equivalents" itself is based on "probabilistic information" such as this, and on "probabilistic information" pertaining to fundamental properties of probability functions, such as:

$$a < b \longrightarrow P(Y \leq a) \leq P(Y \leq b) \tag{74}$$

$$C \subset D \longrightarrow P(C) \leq P(D). \tag{75}$$

The inference to be drawn then is that info-gap scholars would be well advised to reflect on the implications of the above "probabilistic information" for the role of info-gap's proxy theorems in a global probabilistic robustness analysis. They would also do well to pinpoint where such "probabilistic information" is appealed to, in the proofs of info-gap's proxy theorems.

6 Summary and conclusions

The idea of deriving **deterministic equivalents** for probabilistic problems on the strength of fundamental properties of **probability measures** is long standing. Still, this idea is applicable only in highly specialized cases, and in the case of the probability measures being completely unknown, to **trivial** problems only.

In the case of probabilistic global robustness problems such as those considered in this discussion, namely $\max_{x \in X} P(f(x, W) \leq b)$, where P is completely unknown, the derivation of deterministic equivalents calls for significant simplifications in the performance constraint $f(x, w) \leq b$. Such simplifications effectively render the derivation process itself trivially simple. This means that the proposition to use radius of stability models to obtain deterministic equivalents in such cases is nothing short of a "sledgehammer approach".

It is akin to the proposition to solve the quadratic equation

$$(x - a)(x - b) = 0 \tag{76}$$

by rewriting it as

$$x^2 - (a + b)x + ab = 0 \tag{77}$$

to then solve this equation by means of the"quadratic formula"

$$x = \frac{(a + b) \pm \sqrt{(a + b)^2 - 4ab}}{2} \tag{78}$$

to obtain the two solutions $x_1 = a, x_2 = b$, that can be obtained, by inspection, directly from (76).

The main point one needs to appreciate about info-gap's proxy theorems is the fundamental difficulty that they have to grapple with. This is the incongruity between the fact that the probabilistic robustness problem $\max_{x \in X} P(f(x, W) \leq b)$ aims for **global** robustness, but as a radius of stability model, info-gap's robustness model seeks **local** robustness. So, the conditions that the proxy theorems must therefore stipulate end simplifying the info-gap robustness problem to such an extent that its deterministic equivalent is trivially obvious. Namely, its local robustness is a priori guaranteed to be global. This is so because these conditions either require the sets $A(x), x \in X$ to be nested, or they effectively require the probabilities $P(A(x)), x \in X$ to be monotone with the radii of stability of the decisions $x \in X$.

7 Epilogue

It is important to take note that each and every proposition in Ben-Haim (2013) figuring in the illustrations of info-gap's claimed ability to "maximize the probability of success without using probabilistic information" imposes two essential conditions on the probabilistic problem considered, namely on

$$\textbf{Probabilistic Problem:} \quad p^* := \max_{x \in X} \ P(f(x, W) \leq b). \tag{79}$$

These two conditions are as follows:

- The uncertainty parameter w in the constraint $f(x, w) \leq b$ is a numeric **scalar** representing the realisation of the random variable W.
- The performance level $f(x, w)$ is **monotone** with this parameter.

As we saw in this discussion (see section 3.1), these conditions simplify the problem considered to such an extent that the derivation of deterministic equivalents is immediately obvious.

For, if w is a numeric scalar and for each $x \in X$ the value of $f(x,w)$ is say non-decreasing with w, then by inspection, the constraint $f(x,w) \leq b$ is equivalent to the constraint $w \leq \overline{w}(x)$ where

$$\overline{w}(x) := \max_{w \in \mathbb{R}} \{w : f(x,w) \leq b\} \ , \ x \in X \tag{80}$$

assuming that the max is attained. If not, replace the max by sup, and if the level set $\{w : f(x,w) \leq b\}$ is unbounded above, set $\overline{w}(x) = \infty$.

The implication is then that $f(x,w) \leq b$ is equivalent to $w \leq \overline{w}(x)$, hence that $P(f(x,w) \leq b) = P(W \leq \overline{w}(x))$. In short, we have:

THEOREM 4 *Assume that w is a numeric scalar and that for each $x \in X$ the value of $f(x,w)$ in non-decreasing with w. Then,*

Probabilistic problem	*Deterministic equivalent*
$p^* := \max_{x \in X} P(f(x,W) \leq b)$	$h^* := \max_{x \in X} \overline{w}(x) = \max_{x \in X, w \in \mathbb{R}} \{w : f(x,w) \leq b\}$

(81)

Namely, any optimal solution to the deterministic equivalent is also optimal with respect to the probabilistic problem. Note that if $\overline{w}(x) = \infty$ for some $x \in X$, then $P(f(x,W) \leq b) = 1$ and x is an optimal solution to both problems.

The implication is therefore that using the heavy guns of radius of stability models deployed by info-gap decision theory for this purpose is totally unnecessary, indeed, farcical.

The trouble is that, as in previous discussions on info-gap's claimed ability to maximize the probability of success without using probabilistic information (e.g. Ben-Haim et al. 2009, Ben-Haim and Cogan 2011, Ben-Haim (2012a), the hard facts about this claim lie submerged in Ben-Haim (2013) under layers of rhetoric.

Hence the title of this article.

Bibliograpy

1. Anderson, P. and Bernfeld, S.R. (2001) Properties of the radii of stability and instability. Pp. 1-9 in *Differential Equations and Nonlinear Mechanics,* K. Vajravelu (editor), Kluwver, Norwell.
2. Bai, D., Carpenter, T., and Mulvey, J. (1997) Making a case for robust optimization models. *Management Science,* 43, 895-907.
3. Beigi, G., Hosseinzadeh, L.F., Noora, A.A., Mozaffari M.R., Gholami K. and Dehghan, F. (2009) Sensitivity and stability analysis in DEA on interval data by using MOLP methods. *Applied Mathematical Sciences,* 3(8), 891-908.
4. Ben-Haim, Y. (2001) *Information-gap decision theory.* Academic Press.
5. Ben-Haim, Y. (2005) Value-at-risk with info-gap uncertainty. *The Journal of Risk Finance,* 6(5), 388-403.
6. Ben-Haim, Y. (2006) *Info-gap decision theory.* Elsevier.
7. Ben-Haim, Y. (2007) Faqs about info-gap. http://www.technion.ac.il/yakov/IGT/faqs01.pdf
8. Ben-Haim, Y. (2010) *Info-gap Economics: an Operational Introduction.* Palgrave.
9. Ben-Haim, Y. (2012) Doing Our Best: Optimization and the Management of Risk. *Risk Analysis,* 32(8), 1326-1332.
10. Ben-Haim, Y. (2012a) Why Risk Analysis is Difficult, and Some Thoughts on How to Proceed. *Risk Analysis,* 32(10), 1638-1646.

11. Ben-Haim, Y. (2012b) Modeling and design of a Hertzian contact: an info-gap approach. *The Journal of Strain Analysis for Engineering Design,* 47(2), 153-162.

12. Ben-Haim, Y. (2012c) Doing Our Best: Optimization and the Management of Risk. *Risk Analysis* (Early View)

13. Ben-Haim, Y (2013) Robust satisfying and probability of survival. *International Journal of Systems Science,* in press.

14. Ben-Haim, Y., Dacso, CC., Carrasco, J., and Rajan, N. (2009) Heterogeneous uncertainties in cholesterol management. *International Journal of Approximate Reasoning,* 50, 1046-1065.

15. Ben-Haim, Y. and Cogan, S. (2011) Linear Bounds on an Uncertain Non-Linear Oscillator: an Info-Gap Approach. UTAM Symposium on the Vibration Analysis of Structures with Uncertainties. In Belyaev, A.K. and Langley, RZ.S., Editors, IUTAM Bookseries Volume 27, pp 3-14.

16. Y. Ben-Haim, Y., Zacksenhouse, M., Keren, C., C.C. Dacso, C.C. (2009) Do we know how to set decision thresholds for diabetes? *Medical Hypotheses,* 73(2),189-193.

17. Ben-Tal, A., Bertsimas, D., and Brown, D.B. (2010) A soft robust model for optimization under ambiguity. *Operations Research,* 58(4), 1220-1234.

18. Ben-Tal, A., den Hertog, D., and Vial, J-P. (2012) Deriving robust counterparts of nonlinear uncertain inequalities. Working paper No. 2012-0053, Centrum Wiskunde & Informatica, Tilburg University.
http://www.optimization-online.org/DB_HTML/2012/07/3520.html

19. Ben-Tal A., El Ghaoui, L., and Nemirovski, A. (Editors, 2006)*Mathematical Programming.* Special issue on Robust Optimization, Volume 107(1-2).

20. Ben-Tal, A., El Gaoui, L., and Nemirovski, A. (2009) *Robust optimization.* Princeton University Press.

21. Ben-Tal, A., Golany, B., and Shtern, S. (2009a) Robust multi- echelon multi-period inventory control. *European Journal of Operational Research,* 199(3), 922-935.

22. Ben-Tal, A., and Nemirovski, A. (1999) Robust solutions to uncertain programs. *Operations Research Letters,* 25, 1-13.

23. Ben-Tal, A., and Nemirovski, A. (2002) Robust optimization—methodology and applications. *Mathematical Programming,* Series B. 92,453-480.

24. Bertsimas, D, and Sim, M. (2003) Robust discrete optimization and network flows. *Mathematical Programming,* Series B, 98, 47-91.

25. Bertsimas, D., and Sim, M. (2004) The price of robustness. *Operations Research,* 52(1), 35-53.

26. Bertsimas, D., and D.B. Brown. D.B. (2009) Constructing uncertainty sets for robust linear optimization. *Operations Research,* 57(6),1483-1495.

27. Bertsimas, D., and D.B. Brown. D.B., and Caramanis, C. (2011) Theory and applications of robust optimization. *SIAM Review,* 53(3), 464-501.

28. Birge, J.R., and Louveaux, F. (1997) *Introduction to Stochastic Programming.* Springer.

29. Brock, W.A., Durlauf, S.N. (2005). Local robustness analysis: theory and application. *Journal of Economic Dynamics & Control,* 29, 2067-2092.

30. Byron, M. (1998) Satisficing and Optimality. *Ethics,* 109, 67-93.

31. Byron, M. (Editor, 2004) *Satisficing and Maximizing: Moral Theorists on Practical Reason.* Cambridge University Press.

32. Carmel, Y., and Ben-Haim, Y. (2005) Info-Gap Robust-Satisficing Model of Foraging Behavior: Do Foragers Optimize or Satisfice? *The American Naturalist,* 166(5), 663-641.

33. Charnes, A., and Cooper, W.W. (1963) Deterministic equivalents for optimizing and satisficing under chance constraints. *Operations Research,* 11(1), 18-39.

34. Charnes, A., Cooper, W.W., and Thompson, G.L. (1964) Critical path analysis via chance constrained and stochastic programming. *Operations Research,* 12(3), 460-470.

35. Cooper, W.W., Seiford, L.M., and Zhu, J. (2004) *Handbook on Data Envelopment Analysis.* Kluwer, Norwell.

36. Chisholm, R. A. (2010) Trade-offs between ecosystem services: Water and carbon in a biodiversity hotspot. *Ecological Economics,* 69, 1973-1987.

37. Ecker J.G., and Kupferschmid, M. (1988) *Introduction to Operations Research.* Wiley.

38. Eiselt, H.A. and Langley A. (1990) Some extensions of domain criteria in decision making under uncertainty. *Decision Sciences,* 21, 138-153.

39. Eiselt, H. A., Sandblom, C.L., and Jain, N. (1998) A spatial criterion as decision aid for capital projects: locating a sewage treatment plant in Halifax, Nova Scotia. *Journal of the Operational Research Society,* 49(1), 23-27.

40. Feller, W. (1968) *An Introduction to Probability Theory and Its Applications,* Vol. 1. Wiley.

41. French, S.D. (1988) *Decision Theory.* Ellis Horwood, NY.

42. Gabrel, V., Muraty, C., and Thielez, A. (2012) Recent Advances in Robust Optimization and Robustness: An Overview. Technical report, LAMSADE, Universite Paris-Dauphine, Paris, France.
http://www.optimization-online.org/DB_HTML/2012/07/3537.html.

43. Gassmann, H.I., and Ziemba, W.T. (Editors, 2013) *Stochastic Programming: Applications in Finance, Energy, Planning and Logistics.* World Scientific.

44. Gilboa, I. (2009) *Theory of Decision Under Uncertainty.* Cambridge University Press.

45. Gilboa, I. (2010) *Rational Choice.* The MIT Press.

46. Gilboa, I., and Schmeidler D. (1989) Maximin expected utility with non-unique prior. *Journal of Mathematical Economics,* 18. 141-153.

47. Gilboa, I., Postlewaite, A., and Schmeidler D. (2009) Is it always rational to satisfy Savage's axioms? *Economics and Philosophy,* 25, 285-296.

48. Giordani, P. and Paul Śoderlind, P. (2004) Solution of macromodels with Hansen–Sargent robust policies: some extensions *Journal of Economic Dynamics and Control,* 28, 2367-2397.

49. Goh, J., and Sim, M. (2010) Distributional robust optimization and tractable approximations. *Operations Research,* 58(4), 902-917.

50. Goh, J., and Sim, M. (2011) Robust optimization made easy with ROME. *Operations Research,* 59(4), 973-985.

51. Good, I.J. (1950) *Probability and the Weighing of Evidence.*

52. Good, I.J. (1952) Rational Decisions. Journal of the Royal Statistical Society, Series B (methodological), 14(1), 107-14.

53. Good, I.J. (1983) *Good Thinking: The Foundations of Probability and its Applications.* University of Minnesota Press.

54. Good, I.J. (1995). Reliability always depends on probability of Course . *Journal of Statistical Computation and Simulation,* 52, 192-193.

55. Good, I.J (1996) Reply to Prof. Y. Ben-Haim.*Journal of Statistical Computation and Simulation,* 55, 265-266.

56. Gupta, S.K., and Rosenhead, J.V. (1968) Robustness in sequential investment decisions. *Management Science,* 15(2) Application Series, pp. B18-B29.

57. Hafner, M., Koeppl, H., Hasler, M., and Wagner, A. (2009) "Glocal" robustness analysis and model discrimination for circadian oscillators. *PLoS Computational Biology,* 5(10): e1000534.

58. Hall, J., and Ben-Haim, Y. (2007) Making Responsible Decisions (When it Seems that You Can't) Engineering Design and Strategic Planning Under Severe Uncertainty. http://www.floodrisknet.org.uk/a/2007/11/hall-benhaim.pdf/view

59. Hall, J.W. (2003) Handling uncertainty in the hydroinformatic process. *Journal of Hydroinformatics,* 5(4), 215-232.

60. Hall, J., and Dimitri Solomatine, D. (2008) A framework for uncertainty analysis in flood risk management decisions. *International Journal of River Basin Management,* 6(2), 85-98.

61. Hall, J., and Harvey, H. (2009) Decision making under severe uncertainty for flood risk management: a case study of info-gap robustness analysis. *Eighth International Conference on Hydroinformatics,* January 12-16, 2009, Concepcion, Chile. http://www.floodsite.net/html/partner_area/project_docs/T20-08-05-hic-infogap.pdf

62. Hall, J., Lempert, R.J., Keller, K., Hackbarth, C., Mijere, A., and McInerney, D.J. (2012) Robust climate policies under uncertainty: A comparison of Info-Gap and RDM methods. *Risk Analysis* (Early View).

63. Hansen, L.P., and Sargent, T.J. (2007) *Robustness.* Princeton University Press.

64. Hine, D., and Hall, J.W. (2010) Information gap analysis of flood model uncertainties and regional frequency analysis. *Water Resources Research*, volume 46, W01514, doi:10.1029/2008WR007620.

65. Hindrichsen D. and Pritchard, A.J. (1986a) Stability radii of linear systems. *Systems & Control Letters,* 7, 1-10.

66. Hindrichsen D. and Pritchard, A.J. (1986b) Stability radius for structured perturbations and the algebraic Riccati equation. *Systems & Control Letters,* 8, 105-113.

67. Hinrichsen D. and Pritchard A.J. (1990) A note on some differences between real and complex stability radii. *Systems & Control Letters,*14,401-409.

68. Hinrichsen D. and Pritchard A.J. (1992) Robustness measures for linear systems with applications to stability radii of Hurwitz and Schur polynomials. *International Journal of Control,* 55, 809-844.

69. Hinrichsen D. and Pritchard A.J. (2005) *Mathematical Systems Theory I: Modelling, State Space Analysis, Stability and Robustness,* Springer.

70. Huber, P.J. (1981) *Robust Statistics.* Wiley.

71. Huber, P.J., and Ronchetti, E., (2009) *Robust Statistics,* 2nd edition, Wiley.

72. Kouvelis, P., and Yu, G. (1997) *Robust discrete optimization and its applications.* Kluwer.

73. Luce, R.D., and Raiffa, H. (1957) *Games and Decisions.* John Wiley.

74. Milne, W.E. and Reynolds, R.R. (1962) Fifth-order methods for the numerical solution of ordinary differential equations. *Journal of the ACM,* 9(1), 64-70.

75. Mongin, P. (2000) Does Optimality imply rationality? *Synthese,* 124(1-2), 73-111.

76. Mulvey, J.M., Vanderbei, R.J., and Stavros A. Zenios, S.A. (1995) Robust optimization of large-Scale systems. *Operations Research,* 43(2), 264-281.

77. Odhnoff, J. (1965) On the Techniques of Optimizing and Satisficing. *The Swedish Journal of Economics,* 67(1), 24-39.

78. Paice A.D.B. and Wirth, F.R. (1998) Analysis of the Local Robustness of Stability for Flows. *Mathematics of Control, Signals, and Systems,* 11, 289-302.

79. Prékopa, A. (1995) *Stochastic Programming.* Springer.

80. Raab, R.L., and Feroz, E.H. (2007) A productivity growth accounting approach to the ranking of developing and developed nations. *The International Journal of Accounting,* 42, 396415.

81. Rawls, J. (1971) *Theory of Justice.* Belknap Press, Cambridge, MA.

82. Resnik, M.D. (1987) *Choices: an Introduction to Decision Theory.* University of Minnesota Press, Minneapolis, MN.

83. Rosenhead, J., Elton, M., and Gupta, S.K. (1968) Robustness and Optimality as Criteria for Strategic Decisions. *Operational Research Quarterly,* 23(4), 413-431.

84. Rozanov, Y.A. (1977) *Probability Theory: A Concise Course.* Dover.

85. Savage, L.J. (1951) The theory of statistical decision. *Journal of the American Statistical Association*, 46(253), 55-67.

86. Schwartz, B., Ben-Haim, Y., and Dacso, C. (2011) What Makes a Good Decision? Robust Satisficing as a Normative Standard of Rational Decision Making. *The Journal for the Theory of Social Behaviour,* 41(2), 209-227.

87. Simon, H.A. (1961) *Administrative Behaviour.* Macmillan, New York.

88. Simon, H.A., (1976) *Administrative Behavior,* 3rd edition. Free Press, New Press.

89. Simon, H.A. (1997) *Models of Bounded Rationality: Empirically Grounded Economic Reason,* Vol. 3. Boston: MIT Press.

90. Sims, C. (2001). Pitfalls of a Minimax Approach to Model Uncertainty.*American Economic Review,* 91, 5-4.

91. Sniedovich, M. (2007) The art and science of modeling decision-making under severe uncertainty. *Decision Making in Manufacturing and Services,* 1-2, 111-136.

92. Sniedovich, M. (2008) Wald's Maximin model: a treasure in disguise! *Journal of Risk Finance,* 9(3), 287-291.

93. Sniedovich, M. (2008a) The Mighty Maximin, Working Paper No. MS-02-08, Department of Mathematics and Statistics, The University of Melbourne.
http://info-gap.moshe-online.com/mighty.pdf

94. Sniedovich, M. (2010) A bird's view of Info-Gap decision theory, *Journal of Risk Finance,* 11(3), 268-283.

95. Sniedovich, M. (2012) Black swans, new Nostradamuses, voodoo decision theories and the science of decision-making in the face of severe uncertainty. *International Transactions in Operational Research,* 19(1-2), 253-281.

96. Sniedovich, M. (2012a) Fooled by local robustness. *Risk Analysis* (Early View).

97. Sniedovich, M. (2012b) Fooled by local robustness: an applied ecology perspective. *Ecological Applications* (Early View).

98. Sotskov, Y. (2007). Recent results on stability analysis of an optimal assembly line balance. *Information Theories & Applications,* 14:174-181.

99. Sotskov, Y.N., Leontev, I.V.K., Gordeev, E.N. (1995) Some concepts of stability analysis in combinatorial optimization. *Discrete Applied Mathematics,* 58, 169-190.

100. Sotskov, Y.N., Sotskova, N.Y., and Werner, F. (1997) Stability of an optimal schedule in a job shop. *Omega,* 25(4), 397-414.

101. Sotskov, Y.N., Tanaev, V.S., and Werner, F. (1998) Stability radius of an optimal schedule: a survey and recent developments. In Yu, G. (editor) *Industrial Applications of Combinatorial Optimization,* pp. 72-108. Kluwer.

102. Sotskov, Y.N., Wagelmans, A.P.M., and Werner, F. (1998) On the calculation of the stability radius of an optimal or an approximate schedule. *Annals of Operations Research,* 83, 213-252.

103. Starr, M.K. (1963) *Product design and decision theory.* Prentice-Hall.

104. Starr, M.K. (1966) A discussion of some normative criteria for decision-making under uncertainty. *Industrial Management Review,* 8(1), 71-78.

105. Stein, J. (2010) *The Right Decision: A Mathematician Reveals How the Secrets of Decision Theory Can Help You Make the Right Decision Every Time.* McGraw-Hill.

106. Stigler, S.M. (2010) The changing history of robustness. *The American Statistician,* 64(4), 277-281.

107. Stoye, J. (2009) Minimax. In S. N. Durlauf, and L. E. Blume (Eds.), *The New Palgrave dictionary of economics.* London: Palgrave Macmillan.

108. Stoye, J. (2009a) Minimax regret. In S. N. Durlauf, and L. E. Blume (Eds.), *The New Palgrave dictionary of economics.* London: Palgrave Macmillan.

109. Stoye, J. (2011) Statistical decisions under ambiguity. *Theory and Decision,* 70(2), 129-148.

110. Taleb, N. N. 2005 *Fooled by randomness.* Random House.

111. Taleb, N. N. 2007 *The Black Swan: the impact of the highly improbable.* Random House.

112. Taleb, N,N, (2010) *The Black Swan: Second Edition: The Impact of the Highly Improbable: With a new section: "On Robustness and Fragility".* Random House.

113. Tintner, G. (1952) Abraham Wald's contributions to econometrics. *The Annals of Mathematical Statistics,* 23(1), 21-28.

114. Ullman, D.G. (2006) *Making Robust Decisions: Decision Management For Technical, Business, & Service Teams.* Trafford Publishing.

115. Untiedt, E. (2010) A parametrized model for optimization with mixed fuzzy and possibilistic uncertainty. *Studies in Fuzziness and Soft Computing,* 254, 193-208.

116. Urban, G.L. (1974) Building models for decision makers. *Interfaces,* 4(3), 1-11.

117. Van Witteloostuijn, A. (1988) Maximising and satisficing. *Journal of Economic Psychology,* 9, 289-313.

118. Wald, A. (1939) Contributions to the theory of statistical estimation and testing hypotheses. *Annals of Mathematical Statistics* 10(4), 299-326.

119. Wald, A. (1945) Statistical decision functions which minimize the maximum risk. *The Annals of Mathematics* 46(2), 265-280.

120. Wald, A. (1950). *Statistical Decision Functions.* John Wiley, NY.

121. Weisberg, M. (2007) Who is a Modeler? *British Journal for the Philosophy of Science,* 58, 207-233.

122. Weisberg, M. (2006) Robustness analysis. *Philosophy of Science,* 73, 730-742.

123. Wilf, H.S. (1960) Maximally stable numerical integration. *Journal of the Society for Industrial and Applied Mathematics,* 8(3), 537-540.

124. Williams, N. (2008). *Robust Control.* In Palgrave New Dictionary of Economics, http://www.dictionaryofe

125. Wintle, B.A., Runge, M.C., and Bekessy, S.A. (2010) Allocating monitoring effort in the face of unknown unknowns. *Ecology Letters,* 13(11), 1325-1337.

126. Zhang, Y. (2007) General robust-optimization formulation for nonlinear programming. *Journal of Optimization Theory and Applications,* 132(1), 111-124.

127. Zlobec, S. (1987) Survey of input optimization. *Optimization,* 18:309-348.

128. Zlobec, S. (1988) Characterizing Optimality in Mathematical Programming Models. *Acta Applicandae Mathematicae,* 12, 113-180.

129. Zlobec, S. (2001) Nondifferentiable optimization: Parametric programming. In *Encyclopedia of Optimization,* Floudas, C.A and Pardalos, P.M. editors, Springer.

130. Zlobec, S. (2009) Nondifferentiable optimization: Parametric programming. Pp. 2607-2615, in *Encyclopedia of Optimization,* Floudas, C.A and Pardalos, P.M. editors, Springer.