



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Sniedovich, M

Title:

Wald's mighty maximin: A tutorial

Date:

2016-07-01

Citation:

Sniedovich, M. (2016). Wald's mighty maximin: A tutorial. *International Transactions in Operational Research*, 23 (4), pp.625-653. <https://doi.org/10.1111/itor.12248>.

Persistent Link:

<https://hdl.handle.net/11343/290816>

# Wald's Mighty Maximin: A Tutorial

Moshe Sniedovich

Department of Mathematics and Statistics, University of Melbourne, Melbourne, Victoria 3010, Email: moshe@ms.unimelb.edu.au

Ever since its appearance on the scene in the 1940s, in the then nascent field of *statistical decision theory*, Wald's maximin paradigm has played a vital role in many disciplines in the treatment of non-probabilistic uncertainty, both as a tool of thought and as a practical instrument. In fact, in some fields, such as modern *robust optimization*, Wald's paradigm dominates the scene. It is important to note that this paradigm's preeminence in many fields continues unabated, despite its obvious limitations and the criticism that had been and continues to be leveled at it. So, in this tutorial we examine the methodological aspects of this stalwart of decision theory from the viewpoint of *robust decision-making*, paying special attention to its obvious and not so obvious limitations and to its relation to other maximin paradigms.

*Keywords:* Wald's Maximin paradigm, worst-case analysis, uncertainty, decision theory, robust optimization.

## 1 Introduction

To set the stage for our discussion on Wald's maximin paradigm, let us begin by considering the simple decision problem that is represented by the table below. The entries in the table, denoted  $p(x, s)$ , designate *payoffs*, the rows designate the *alternatives* available to the decision maker, and the columns designate possible (uncertain)

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/itor.12248

This article is protected by copyright. All rights reserved.

scenarios. Hence, there are three alternatives and five possible scenarios in this case.

$p(x, s)$	$s^{(1)}$	$s^{(2)}$	$s^{(3)}$	$s^{(4)}$	$s^{(5)}$
$x^{(1)}$	5	4	4	6	9
$x^{(2)}$	8	9	7	1	2
$x^{(3)}$	8	8	4	7	5

(1)

For simplicity, assume that a unit of payoff is 1\$ (USA), and that the larger the payoff the better.

The task set by this decision problem is this: determine the best (optimal) alternative.

It is interesting to note that, to date, *decision theory* has not produced a uniform, universally accepted recipe, method, or technique to handle decision problems of this type. Indeed, to this day, experts remain divided on how best to deal with such problems.

At first glance this might appear surprising because, on the face of it, the decision problem in question seems to be simplicity itself, so much so that the last thing one would have expected is a lack of unanimity on how best to deal with it. Indeed, one would have expected decision theory to furnish a simple recipe to match the apparent simplicity of this problem. On closer scrutiny, however, it becomes clear that the main reason that this seemingly simple problem is anything but simple is due to the *uncertainty* in the five scenarios. To be precise, the difficulty here is in our ignorance as to which of the five possible scenarios will be realized.

For, had we possessed this knowledge, the decision problem would in fact be trivial. And to illustrate, had we known (for sure) that scenario  $s^{(4)}$  would be realized, the best alternative would have been  $x^{(3)}$ , yielding the payoff  $p(x^{(3)}, s^{(4)})=7$ . And had we known for sure that scenario  $s^{(2)}$  would be realized, the best alternative would have been  $x^{(2)}$ , yielding the payoff  $p(x^{(2)}, s^{(2)})=9$ . But, in the absence of this knowledge, determining the best alternative (optimal) in this case is no easy matter.

The question is then: how does one approach problems of this type?

In this tutorial we examine one of the paradigms, indeed the foremost paradigm, offered by decision theory for this purpose, namely *Wald's maximin paradigm* (circa 1940). The discussion is thus organized according to these topics:

- Section 2: The basics
- Section 3: Formal generic maximin models
- Section 4: What makes a maximin model a Wald-type model

- Section 5: Playing against Nature
- Section 6: Robustness
- Section 7: Rationality and conservatism
- Section 8: Variations on a theme
- Section 9: Robust optimization
- Section 10: Can Wald's maximin paradigm save the world?

## 2 The basics

The identifying characteristic of Wald's maximin paradigm is in its approach to uncertainty which can be stated informally as follows:

- *In the face of uncertainty, assume the worst!*

Thus, taking this approach to the above fundamental decision problem, each alternative available to the decision maker is evaluated on the basis of the worst-case scenario pertaining to the alternative considered. Then, the best alternative is judged to be that whose performance under the worst-case scenario is at least as good as the worst-case performance of all other alternatives. A recipe of this type is called a *maximin decision rule* and it can be stated in various ways (e.g. Luce and Raiffa 1957, Resnik 1987, French 1988, Peterson 2009). For our purposes, Rawls's (1971, pp. 152-3) formulation of the rule is particularly instructive:

The maximin rule tells us to rank alternatives by their worst possible outcomes: we are to adopt the alternative the worst outcome of which is superior to the worst outcome of the others.

To formulate rules of this type more formally, let  $X$  denote the set of *alternatives* available to the decision maker, let  $S$  denote the set of possible *scenarios* under consideration, and let  $O(x, s)$  denote the *outcome* generated by alternative  $x \in X$  under scenario  $s \in S$ . Formally  $O$  can be regarded as a function defined on  $X \times S$ . This being so, we can give the maximin rule the following abstract mathematical transliteration:

$$z^* := \underset{x \in X}{\text{best}} \left\{ \underset{s \in S}{\text{worst}} O(x, s) \right\} \quad (2)$$

$$= \underset{x \in X}{\text{best}} \underset{s \in S}{\text{worst}} O(x, s) \quad (3)$$

where the  $\underset{x \in X}{\text{worst}}$  operation selects the worst outcome over all the scenarios, and the  $\underset{s \in S}{\text{best}}$  operation selects the best outcome over all the alternatives. These operations represent the decision maker's preferences.

For simplicity, we assume that the outcomes are numeric scalars, namely that  $O$  is a real-valued function, and that the larger the outcome the better. In this case  $\text{best} = \max$  and  $\text{worst} = \min$ , whereupon we have what to many readers no doubt epitomizes the maximin rule, namely:

$$z^* := \max_{x \in X} \min_{s \in S} O(x, s). \quad (4)$$

And in cases where the smaller the outcome the better, we have:

$$z^* := \min_{x \in X} \max_{s \in S} O(x, s) \quad (5)$$

which is a minimax, rather than a maximin, model.

We call attention to the concept *pessimization* (IECRC 1966, Corbett 1988, Bell 1992, Mutapcic and Boyd 2009, Oskooi et al. 2012, Ben-Tal et al. 2015) as an antonym of *optimization*. In this case the abbreviated *pes* would function as an antonym of *opt*, whereupon the maximin model and the minimax model would be represented by the following single abstract model:

$$z^* := \underset{x \in X}{\text{opt}} \underset{s \in S}{\text{pes}} O(x, s). \quad (6)$$

We note though that the discussion in this tutorial focuses almost exclusively on maximin versions of this abstract model.

It should also be pointed out that although the outcomes  $O(x, s)$ ,  $x \in X$ ,  $s \in S$ , are assumed to be numeric scalars, the above simple models can easily accommodate *performance requirements*, namely *constraints*, as well as *payoffs*. Thus, methodologically speaking, it is instructive to consider three types of outcomes.

## 2.1 Outcomes representing payoffs

In this case

$$O(x, s) = p(x, s), \quad x \in X, s \in S \quad (7)$$

where  $p$  is a real-valued function on  $X \times S$  such that  $p(x, s)$  represents the *payoff* generated by alternative  $x$  under scenario  $s$ . Assuming that the larger the payoff the better, the maximin decision rule translates to the following maximin model:

$$z^* := \max_{x \in X} \min_{s \in S} p(x, s). \quad (8)$$

**Example**

Consider the following payoff table:

$p(x, s)$	$s^{(1)}$	$s^{(2)}$	$s^{(3)}$	$s^{(4)}$	$s^{(5)}$	
$x^{(1)}$	5	4	4	6	9	(9)
$x^{(2)}$	8	9	7	1	2	
$x^{(3)}$	8	8	3	7	5	

To find the best alternative, we would first append to the table a column whose entries are the worst payoffs for the respective alternatives:

$p(x, s)$	$s^{(1)}$	$s^{(2)}$	$s^{(3)}$	$s^{(4)}$	$s^{(5)}$	$\min_{s \in S} p(x, s)$	
$x^{(1)}$	5	4	4	6	9	4	(10)
$x^{(2)}$	8	9	7	1	2	1	
$x^{(3)}$	8	8	3	7	5	3	

These values are termed in the idiom of decision theory the *security levels* of the respective alternatives.

Next we would identify that row (alternative) whose security level is the largest:

$p(x, s)$	$s^{(1)}$	$s^{(2)}$	$s^{(3)}$	$s^{(4)}$	$s^{(5)}$	$\min_{s \in S} p(x, s)$	$\max_{x \in X} \min_{s \in S} p(x, s)$	
$x^{(1)}$	5	4	4	6	9	4	4	(11)
$x^{(2)}$	8	9	7	1	2	1		
$x^{(3)}$	8	8	3	7	5	3		

We would therefore conclude that, according to the maximin rule, alternative  $x^{(1)}$  is the best, seeing that its security level which is equal to 4 is larger than the security levels of the other two alternatives.

**2.2 Outcomes representing performance requirements**

Suppose that there are no payoffs, but the  $(x, s)$  pairs are subject to certain performance constraints. In this case the outcomes would be defined as follows:

$$O(x, s) = \begin{cases} 1 & \text{, the pair } (x, s) \text{ satisfies the constraints} \\ 0 & \text{, the pair } (x, s) \text{ violates the constraints} \end{cases} \quad , \quad x \in X, s \in S. \quad (12)$$

This means that alternative  $x \in X$  is *admissible* iff the pair  $(x, s)$  satisfies the constraints for all  $s \in S$ . If none of the alternatives is admissible, the conclusion is that the maximin problem has no solution: there is no  $x \in X$  such that  $(x, s)$  satisfies the constraints for all  $s \in S$ .

**Example**

Consider the following outcome table, representing performance requirements imposed on  $(x, s)$  pairs:

$O(x, s)$	$s^{(1)}$	$s^{(2)}$	$s^{(3)}$	$s^{(4)}$	$s^{(5)}$	(13)
$x^{(1)}$	1	0	1	0	1	
$x^{(2)}$	1	1	1	1	1	
$x^{(3)}$	1	1	1	1	0	

By inspection, there is only one admissible alternative, namely  $x^{(2)}$ . Hence, according to the maximin rule, this alternative is the best.

**2.3 Outcomes representing payoffs and performance requirements**

Suppose that we have both payoffs and performance requirements (constraints). In this case, as in optimization theory, constraint satisfaction, namely *feasibility*, has priority over payoffs. Hence, outcomes would be defined as follows:

$$O(x, s) = \begin{cases} p(x, s) & , \text{ the pair } (x, s) \text{ satisfies the constraints} \\ -\infty & , \text{ the pair } (x, s) \text{ violates the constraints} \end{cases} , \quad x \in X, s \in S. \tag{14}$$

This yields the following maximin model:

$$z^* := \max_{x \in X} \min_{s \in S} O(x, s) \tag{15}$$

$$= \max_{x \in X} \min_{s \in S} \{p(x, s) : (x, s) \text{ satisfies the constraints } \forall s \in S\}. \tag{16}$$

**Example**

Consider the situation described by the following table where the entries represent payoffs and constraint satisfaction. That is, each entry is a pair whose first element represents a payoff and its second element is a boolean scalar indicating whether the constraints are satisfied by the respective  $(x, s)$  pair.

	$s^{(1)}$	$s^{(2)}$	$s^{(3)}$	$s^{(4)}$	(17)
$x^{(1)}$	(9, 1)	(9, 0)	(8, 1)	(9, 0)	
$x^{(2)}$	(3, 1)	(9, 1)	(6, 1)	(8, 1)	
$x^{(3)}$	(4, 1)	(4, 1)	(5, 1)	(6, 1)	

For example, the entry (3, 1) for  $(x^{(2)}, s^{(1)})$  indicates that the payoff is equal to 3 and the constraints are satisfied. The entry (9, 0) for  $(x^{(1)}, s^{(4)})$  indicates that the payoff is equal to 9 and the constraints are violated.

We thus obtain the following outcome table

$O(x, s)$	$s^{(1)}$	$s^{(2)}$	$s^{(3)}$	$s^{(4)}$	$\min_{s \in S} O(x, s)$	$\max_{x \in X} \min_{s \in S} O(x, s)$
$x^{(1)}$	9	$-\infty$	8	$-\infty$	$-\infty$	
$x^{(2)}$	3	9	6	8	3	
$x^{(3)}$	4	4	5	6	4	4

(18)

The conclusion therefore is that, according to the maximin rule, the best decision is  $x^{(3)}$ , whose security level is equal to 4.

### 3 Formal generic maximin models

From what we have seen so far, it is clear that maximin models give formal mathematical expression to the maximin decision rule. For the purposes of this tutorial it is instructive to consider three generic maximin models, generic in the sense that they embody the basic ingredients of Wald’s maximin paradigm. These basic ingredients are as follows:

- A *decision space*,  $X$ , namely a set consisting of all the decisions (alternatives) available to the decision maker.
- A *state space*,  $S$ , namely a set consisting of all the possible states (scenarios) under consideration.
- A real-valued function  $p$  defined on  $X \times S$  specifying the *payoffs* generated by the decisions and states under consideration.
- A list of *performance requirements*, namely a list of *constraints*, imposed on the (decision, state) pairs. Let  $constraints(x, s)$  denote this list.

The first generic model we consider is characterized by the property that it is constraints-free, meaning that there are no explicit constraints on (decision, state) pairs (the list  $constraints(x, s)$  is empty). This model thus takes this simple form:

**Unconstrained model:**

$$z^* := \max_{x \in X} \min_{s \in S} p(x, s). \tag{19}$$

The second generic model is characterized by the property that the list of constraints is not empty. This model thus takes this form:

**Constrained model:**

$$z^* := \max_{x \in X} \min_{s \in S} \{p(x, s) : \text{constraints}(x, s), \forall s \in S\}. \quad (20)$$

The third model is characterized by the property that the payoffs are independent of the state variable. Note that in this case the iconic  $\min_{s \in S}$  operation is superfluous. Models of this type therefore take this form:

**State-free-payoff model:**

$$z^* := \max_{x \in X} \{p(x) : \text{constraints}(x, s), \forall s \in S\}. \quad (21)$$

We call attention to the fact that these three generic models are interchangeable. Namely, each one of these models can be rewritten so as to assume the form of any one of the other two models. For example, the *Unconstrained model* can be written as follows:

$$\max_{x \in X} \min_{s \in S} p(x, s) = \max_{x \in X} \max_{v \in \mathbb{R}} \{v : v \leq p(x, s), \forall s \in S\} \quad (22)$$

$$= \max_{x \in X, v \in \mathbb{R}} \{v : v \leq p(x, s), \forall s \in S\} \quad (23)$$

where  $\mathbb{R}$  denotes the real line. The model on the right-hand side is clearly a *State-free-payoff model*: its payoff  $v$  is independent of the state variable  $s$ .

Note that this transformation exploits the fact that a minimization problem, say  $\min_{y \in Y} g(y)$ , can be rewritten as a maximization problem as follows

$$\min_{y \in Y} g(y) = \max_{v \in \mathbb{R}} \{v : v \leq g(y), \forall y \in Y\}. \quad (24)$$

And to express the *Constrained* model as an *Unconstrained* model, we let

$$f(x, s) = \begin{cases} p(x, s) & , \text{ pair } (x, s) \text{ satisfies the constraints} \\ -\infty & , \text{ pair } (x, s) \text{ does not satisfy the constraints} \end{cases}, \quad x \in X, s \in S \quad (25)$$

so that

$$\max_{x \in X} \min_{s \in S} \{p(x, s) : \text{constraints}(x, s), \forall s \in S\} = \max_{x \in X} \min_{s \in S} f(x, s). \quad (26)$$

In practice, the choice between these three generic maximin models is a matter of style and convenience.

What emerges then is that from a maximin point of view we can distinguish between two types of decisions, namely *admissible* and *inadmissible* decisions, where a decision is said to be admissible iff it satisfies the performance constraints for **all** the states (scenarios) under consideration. More formally, the set of admissible decisions is defined as follows:

$$X_a := \{x \in X : \text{constraints}(x, s), \forall s \in S\}. \quad (27)$$

Observe that in the context of the *Unconstrained* model,  $X_a = X$ .

For simplicity we assume that  $X_a$  is not empty, namely that there is at least one admissible decision.

### Remark

The maximin paradigm can be incorporated in *sequential* decision processes to generate maximin-type *sequential* decision models. For example, the following is a dynamic programming functional equation for a minimax version of the famous *counterfeit coin problem*:

$$f(s) = 1 + \min_{x \in D(s)} f(\max(x, s - 2x)) , \quad s = 2, 3, \dots, n \quad (28)$$

with  $f(1) = 0$ ,  $D(s) := \{1, 2, \dots, I(s/2)\}$  and  $I(s)$  denotes the integer part of  $s$ .

Here  $s$  denotes the number of coins left for inspection (state variable),  $x$  denotes the number of coins placed on each side of the scale (decision variable), and  $f(s)$  denotes the minimum number of weighings required to identify the counterfeit coin under the worst-case scenario, given that  $s$  coins are left for inspection (See Sniedovich (2003) for details about this model).

## 4 What makes a maximin model a Wald-type model?

As we pointed out at the outset, our main concern in this tutorial is with Wald-type maximin models which, as we also noted, were introduced into *statistical decision theory* in the 1940s. To this we should now add that this type of model was introduced into this field by the mathematician Abraham Wald (1902-1950). By the early 1950s models of this type became the mainstay of *decision theory* (Luce and Raiffa 1957, Resnik 1987, French 1988, Peterson 2009) and from there they started entering

many other disciplines, for instance engineering, economics, management, operation research, and so on.

The question is then this: what is it that distinguishes between a Wald-type maximin model and other types of maximin models?

As we shall see, this distinction is not due necessarily to the formal mathematical format of the respective models, but rather to the meaning attaching to the entities that the mathematical objects of these models represent.

And to illustrate, consider the following simple *Unconstrained* maximin model:

$$z^* := \max_{v \in \mathbb{R}} \min_{w \in \mathbb{R}} \{w^2 + 2vw - v^2\}. \quad (29)$$

Is this a Wald-type model?

To answer this question we need to know what entities do the variables  $v$  and  $w$  represent. For this model to be a Wald-type maximin model  $w$  must represent a decision variable whose value is controlled by the decision maker and  $v$  must represent *uncertainty*, namely it must represent an entity whose “true” value is unknown. What is more, the uncertainty in  $w$  must be non-probabilistic.

The following example features a minimax model that is devoid of any uncertainty. Hence, this model is clearly not a Wald-type minimax model.

### Example: location problem

Deterministic maximin and minimax models are used extensively in applied mathematics. To illustrate this point, consider the following “thoroughly deterministic” generic *geometric* problem that was proposed by the English mathematician James Joseph Sylvester (1814–1897) in the article *A Question in the Geometry of Situation* (Sylvester 1857):

- **Smallest (bounding) sphere problem:** Find the smallest  $n$ -sphere containing all the points in a given subset of  $\mathbb{R}^n$ .

Figure 1 illustrates a solution to a simple smallest *circle* problem ( $n = 2$ ). The given set of points consists of the points represented by the small black squares,  $c^*$  denotes the center point of the optimal (smallest) circle, and  $\rho^*$  denotes the radius of this optimal (smallest) circle.

Figure 1: Smallest circle problem

Let  $S$  denote the given set of points in  $\mathbb{R}^n$ , let  $x \in \mathbb{R}^n$  denote the center point of an  $n$ -sphere, and let  $\rho(x, s)$  denote the distance from point  $x$  to point  $s \in S$ . Then

the smallest n-sphere problem can be stated as follows:

$$\rho^* = \min_{x \in \mathbb{R}} \max_{s \in S} \rho(x, s) \quad (30)$$

$$= \min_{x, r \in \mathbb{R}} \{r : \rho(x, s) \leq r, \forall s \in S\} \quad (31)$$

observing that, by definition,

$$r^*(x) := \max_{s \in S} \rho(x, s) \quad (32)$$

is the distance from  $x$  to the most distant point in  $S$  from  $x$ , hence it is the radius of the smallest n-sphere centered at  $x$  that contains **all** the points in  $S$ . By definition,  $\rho^*$  is equal to the radius of the smallest such n-sphere.

In the above minimax model, the optimizing  $x$  specifies the center of the smallest n-sphere and  $\rho^*$  is equal to the radius of this n-sphere. Clearly, in this *minimax* problem no uncertainty is associated with the elements of set  $S$ . In other words, this is a “thoroughly deterministic” minimax problem.

Generic problems of this type have many practical applications. For instance, they are used extensively in *location theory*, where set  $S$  represents the location of a given collection of *demand* points and  $x$  represents the location of a *supply* point. In this framework, the objective is to determine the location of the supply point that minimizes the distance from this point to the most distant demand point. These problems are known as *minimax location problems* (e.g. Elzinga and Hearn 1972).  $\square$

Another point of interest brought out by the above example is the distant roots of maximin models. For as this example indicates, the use of deterministic maximin models predates not only the use of Wald-type maximin models in *statistical decision theory* (Wald 1939, 1945, 1950), but also that of maximin models of the type deployed in game theory (circa 1920s). In fact, the use of maximin models can be traced back at least to *error analysis*, namely, at least to Leonhard Euler (1707-1783). And to illustrate, consider for instance the article *Origin of the theory of errors*, by Sheynin (1966), which argues that Johann Heinrich Lambert (1728-1777) “. . . should be given precedence over Gauss as the originator of the theory of errors . . .” because, as Sheynin (1966, p. 1004) points out:

- (f) An enunciation of the “minimax” principle (minimization of the maximum residual error in geodetic adjustments—the minimum being sought among all possible solutions . . .) But Lambert confessed that he did not know how to use this principle “in a general manner, without many devious ways” (*auf cine allgemeine Art, und ohne mlele Umwege*). The use

of this principle in a rudimentary form for solving a redundant system of linear algebraic equations should be credited to Euler. . . .

And to go back to the features that render a maximin model a Wald-type model, we call attention to the fact that the non-probabilistic nature of the uncertainty that characterizes Wald-type models, is manifested in the assumption that **all** that is known about the “true” value of the state variable  $s$  is that it can be any element of the state space  $S$ . Thus, no assumption whatsoever is made as to whether certain elements of  $S$  are more/less likely to be the true value of  $s$  than others. We simply have no clue.

### Example: an uncertain linear programming problem

Consider the following standard linear programming problem

$$v^* := \max_x \{c^T x : Ax \leq b, x \geq 0\}. \quad (33)$$

There are situations where some of the parameters of this problem are uncertain. For example, consider the case where vector  $b$  is uncertain: it can take values in some ball  $B(\varepsilon, \tilde{b})$  of radius  $\varepsilon$  in  $\mathbb{R}^m$  around some given point  $\tilde{b} \in \mathbb{R}^m$ . But, there is no inkling as to which point in  $B(\varepsilon, \tilde{b})$  is the true value of  $b$ . To repeat, other than knowing that this true value is an element of  $B(\varepsilon, \tilde{b})$  nothing else is known.

In this case, the maximin counterpart of the above standard linear programming problem would require the decision variable  $x$  to satisfy the constraint  $Ax \leq b$  for all  $b \in B(\varepsilon, \tilde{b})$ . This model would then be as follows:

$$z^* := \max_x \{c^T x : Ax \leq b, x \geq 0, \forall b \in B(\varepsilon, \tilde{b})\}. \quad (34)$$

This is a *State-free-payoff* maximin model where  $b$  designates the state variable and  $B(\varepsilon, \tilde{b})$  the state space. Note that this is not a “run of the mill” linear programming model, but rather a *semi-infinite* linear programming problem. Namely, it is a linear programming problem with infinitely many constraints (Ben-Tal et al. 2009).  $\square$

And to sum it all up. From what we have seen thus far, it emerges that Wald-type maximin/minimax models are distinguished by the following four basic properties:

- The state variable  $s \in S$  represents uncertainty, namely the true value of  $s$  is unknown.

- The state space  $S$  consists of all the possible/plausible values of  $s$  under consideration, such that only one of these values is the true value of  $s$ .
- The decision  $x \in X$  is made before the true value of the state variable  $s \in S$  is revealed (realized).
- The uncertainty in the true value of the state variable  $s \in S$  is probability-free, likelihood-free, plausibility-free, chance-free, belief-free, etc. All that is known about this true value is that it is an element of  $S$ .

It goes without saying that the last characteristic implies that the uncertainty that Wald-type models are designed to deal with is severe.

## 5 Playing against Nature

Now, the immediate implication of the fact that in the framework of Wald’s maximin paradigm the state variable  $s \in S$  embodies *uncertainty* is that the realized value of  $s$  is not fully controlled by the decision maker. This situation readily lends itself to a representation of a Wald-type maximin model as a **game** involving two players: the decision maker and *Nature*, with the latter personifying uncertainty (or rather the decision maker’s perception of uncertainty).

In this game the decision maker plays first by selecting a decision  $x \in X$ . In response, *Nature* selects the worst  $s \in S$  associated with this decision. The decision-maker’s objective is to select an  $x \in X$  that yields the best payoff, given the anticipated antagonistic response by *Nature*.

### 5.1 Equilibrium

Unlike the players in classic zero-sum two-person games (von Neumann and Morgenstern 1944, Thie 1988), here the two players do **not** seek an equilibrium. Thus, a solution to a Wald-type maximin problem is not required to be stable. For example, consider the maximin associated with this payoff table:

$$\begin{array}{c|cc||c|c}
 p(x, s) & s^{(1)} & s^{(2)} & \min_{s \in S} p(x, s) & \max_{x \in X} \min_{s \in S} p(x, s) \\
 \hline
 x^{(1)} & \boxed{2} & 3 & \boxed{2} & 2 \\
 x^{(2)} & 4 & 1 & 1 & \\
 \end{array} \tag{35}$$

As a Wald-type maximin model, this game has a unique optimal solution, namely  $(x^*, s^*) = (x^{(1)}, s^{(1)})$ . Observe, however, that this solution is not a saddle point.

Hence, as a classic zero-sum two-person game, this game does not have an optimal *pure strategy*.

Thus, insofar as optimality conditions are concerned, Wald-type maximin models are far simpler than models of the classic zero-sum two-person games ilk.

Before we proceed any further, it is important to be thoroughly clear on how *Nature* selects the worst  $s \in S$  pertaining to a decision  $x \in X$  selected by the decision maker.

## 5.2 Nature's worst-case analysis

Because in Wald-type maximin models the players do not seek an equilibrium, the worst-case analysis conducted by *Nature* is straightforward.

Thus, in the framework of the *Unconstrained* model, the worst-case analysis of decision  $x \in X$  amounts to no more than a minimization of the payoff  $p(x, s)$  over  $s \in S$ . This yields the *security level* of decision  $x$ , namely

$$SL(x) := \min_{s \in S} p(x, s) , \quad x \in X. \quad (36)$$

Therefore, if the decision maker selects decision  $x$ , the anticipated payoff will not be below  $SL(x)$ .

In the case of *Constrained* models, selecting the worst  $s \in S$  pertaining to decision  $x$  is much more involved. This is due to the presence of performance constraints on  $(x, s)$  pairs, and to the fact that constraint satisfaction takes precedence over payoffs. Thus, in response to the selection of a decision  $x \in X$  by the decision maker, *Nature* selects a state  $s \in S$  as follows:

- If there is a state  $s \in S$  such that  $(x, s)$  violates the performance constraints, then *Nature* select such a state.
- If there is no such  $s \in S$ , namely if for every  $s \in S$  the pair  $(x, s)$  satisfies the performance constraints, then *Nature* select a state  $s \in S$  that minimizes  $p(x, s)$  over  $s \in S$ .

Hence, in the context of *Constrained* models, the security level of an admissible decision  $x$  is as follows:

$$SL(x) := \min_{s \in S} p(x, s) , \quad x \in X_a. \quad (37)$$

The same procedure is followed by *Nature* in the case of *State-free-payoff* models, except that here there is no need to minimize the payoff with respect to the state variable, because the payoff is independent of the state variable. So, the procedure runs as follows:

- If there is an  $s \in S$  such that  $(x, s)$  violates the performance constraints, then *Nature* selects such a state.
- If there is no such  $s \in S$ , then *Nature* selects any  $s \in S$ .

Thus, in the context of *State-free-payoff* models, the security level of an admissible decision  $x$  is as follows:

$$SL(x) := p(x) , x \in X_a. \quad (38)$$

In all maximin models, the iconic  $\max_{x \in X}$  operation gives expression to the decision maker's selection of the best  $x \in X$  subject to *Nature's* antagonist stance. The idea therefore is that the decision maker would avoid selecting an inadmissible decision. Hence, if there are no admissible decisions, namely if  $X_a$  is empty, then the problem under consideration is *infeasible* from a maximin perspective.

### 5.3 The importance of playing first

Recall that in the framework of Wald's maximin paradigm, the decision maker "plays" first and *Nature* responds to the decision maker's moves. Therefore, methodologically, it is perfectly kosher, indeed beneficial, to generalize somewhat the formulation of the generic maximin models discussed above by assuming that each decision  $x \in X$  may have its own state space. In other words, let

$$S(x) := \text{set of states (scenarios) pertaining to decision } x, x \in X. \quad (39)$$

Under this assumption, the *Unconstrained* model will take the form

$$z^* := \max_{x \in X} \min_{s \in S(x)} p(x, s). \quad (40)$$

Similarly, the *Constrained* model will be as follows

$$z^* := \max_{x \in X} \min_{s \in S(x)} \{p(x, s) : \text{constraints}(x, s), \forall s \in S(x)\} \quad (41)$$

and the *State-free-payoff* model will have this format:

$$z^* := \max_{x \in X} \{p(x) : \text{constraints}(x, s), \forall s \in S(x)\}. \quad (42)$$

### Example

Consider the following *State-free-payoff* maximin model

$$x^* := \max_{x \geq 0} \{x : g(s) \geq 0, \forall s \in S(x)\} \quad (43)$$

where  $S(x)$  denotes a *neighborhood* of size  $x$  around some point  $\tilde{s} \in \mathbb{R}^n$ , and  $g$  is a real-valued function on  $\mathbb{R}^n$ . By definition then,  $x^*$  is the size of the largest neighborhood around  $\tilde{s}$  all whose points satisfy the constraint  $g(s) \geq 0$ .

This is illustrated in Figure 2 where  $n = 2$  and the neighborhoods are circles around  $\tilde{s}$ . Note that in this case  $x^*$  is equal to the radius of the largest circle around  $\tilde{s}$  that is contained in the region of  $\mathbb{R}^2$  where  $g(s) \geq 0$ .

Figure 2: Local stability/robustness at  $\tilde{s}$

Models of this type are used extensively in many fields to measure the *local* stability/robustness of systems against perturbations in a nominal value ( $\tilde{s}$ ) of a parameter. We discuss such models, called *radius of stability* models, in the next section.

## 6 Robustness

The fact that Wald’s maximin paradigm seeks to determine decisions that perform well (relative to other decisions) under the worst-case scenario, has made Wald-type maximin models ready mediums for the purpose of identifying decisions that are *robust* against uncertainty/variability. And to be sure, these models are used extensively in many fields to this end.

That said, it is important to point out that the phrase “robustness”, as it is currently used in decision theory and optimization theory, is of a relatively recent vintage (circa 1970s) which means that it had not been part of the original idiom of Wald’s paradigm. Still, the idea that is articulated by the modern term “robustness” has been an essential element of the maximin paradigm from the start, except that in the framework of this paradigm it is referred to as the *security level* of decisions. Thus, in the context of a Wald-type maximin model the robustness of decisions is measured in by their *security levels*: the larger the security level the more robust the decision.

The two examples below illustrate the use of maximin models to define robustness of systems/decisions in situations devoid of explicit payoffs, so that robustness is sought only with respect to performance requirements (constraints).

**Example: “size robustness”**

As above, let  $constraints(x, s)$  denote the list of constraints imposed on (decision, state) pairs, and assume that we seek a decision that is robust against uncertainty with respect to these constraints. Obviously, if a decision  $x \in X$  such that

$$constraints(x, s), \forall s \in S(x) \quad (44)$$

exists, then such a decision would be deemed “super-robust”, hence the best (optimal).

Now, suppose that there is no such decision. Rather, for each  $x$  there is some  $s \in S(x)$  such that the pair  $(x, s)$  violates the constraints.

In this case we might opt to seek a decision  $x \in X$  that satisfies the constraints over a large *subset* of  $S(x)$  such that the larger this subset, relative to  $S(x)$ , the better. With this in mind, let

$$A(x) := \text{subset of } S(x) \text{ whose elements satisfy } constraints(x, s), x \in X \quad (45)$$

and define the robustness of decision  $x$  as follows:

$$robust(x) := \frac{\#A(x)}{\#S(x)}, x \in X \quad (46)$$

where  $\#B$  denotes the “size” of set  $B$ . For example, if  $B$  is a discrete set we can let  $\#B$  denote the *cardinality* of set  $B$ . If set  $B$  is say a ball in  $\mathbb{R}^3$ , then we can let  $\#B$  be the radius of this ball, or its volume.

In words, the robustness of decision  $x$ , denoted  $robust(x)$ , is equal to the size of the largest subset of  $S(x)$  over which  $x$  satisfies the constraints, divided by the size of  $S(x)$ . Since  $\#V \leq \#S(x), \forall V \subseteq S(x)$ , it follows that  $0 \leq robust(x) \leq 1$ . In any case, to find the most robust decision, we would solve this problem

$$r^* := \max_{x \in X} robust(x) \quad (47)$$

$$= \max_{x \in X} \frac{\#A(x)}{\#S(x)}. \quad (48)$$

To see the link of these optimization problems to the maximin paradigm, observe that it follows from the definition of  $A(x)$  that

$$\#A(x) = \max_{V \subseteq S(x)} \{\#V : constraints(x, s), \forall s \in V\}, x \in X. \quad (49)$$

Clearly, this is a *State-free-payoff* maximin model where  $V$  plays the role of a decision variable. Therefore:

$$robust(x) = \frac{1}{\#S(x)} \max_{V \subseteq S(x)} \{\#V : constraints(x, s), \forall s \in V\} \quad (50)$$

and

$$r^* = \max_{x \in X, V \subseteq S(x)} \left\{ \frac{\#V}{\#S(x)} : constraints(x, s), \forall s \in V \right\}. \quad (51)$$

These two are also *State-free-payoff* maximin models. In the latter  $(x, V)$  plays the role of a decision variable.

Figure 3 illustrates the implementation of this robustness measure. Here we let  $\#A(x)$  denote the *area* in  $\mathbb{R}^2$  taken up by set  $A(x)$ . Clearly, according to this measure, decision  $x'$  is, by inspection, more robust than decision  $x''$ . Decision  $x'$  satisfies the constraints over 80% of  $S(x')$ , whereas  $x''$  satisfies the constraints only over 25% of  $S(x'')$ .

Figure 3: Size robustness of two decisions

Models of this type date back to the 1960s (e.g. Starr 1963, 1966, Gupta and Rosenhead 1968, Rosenhead et al. 1972, Rosenblat 1987, Eiselt and Laporte 1992, Eiselt et al. 1998, Herman et al. 2014, Whateley et al. 2014).

### Example: radius of stability

There are many situations where we seek decisions that are robust against *perturbations* in a *nominal value* of the state variable, call it  $\tilde{s}$ . In this case we would define the robustness of decision  $x$  as the largest perturbation in  $\tilde{s}$  that does not cause  $x$  to violate the performance constraints in the neighborhood around  $\tilde{s}$  which is demarcated by this perturbation. Formally then, we would let

$$ros(x|\tilde{s}) := \max_{\alpha \geq 0} \{\alpha : constraints(x, s), \forall s \in \mathcal{N}(\alpha, \tilde{s})\}, \quad x \in X \quad (52)$$

where  $\mathcal{N}(\alpha, \tilde{s})$  denotes a neighborhood of radius  $\alpha$  around  $\tilde{s}$ . Note that this is a *State-free-payoff* maximin model where  $\alpha$  plays the role of the decision variable.

In words, the radius of stability of decision  $x$  at  $\tilde{s}$ , denoted  $ros(x|\tilde{s})$ , is equal to the radius  $\alpha$  of the largest neighborhood  $\mathcal{N}(\alpha, \tilde{s})$  around  $\tilde{s}$  all whose elements satisfy the constraints imposed on  $x$  and  $s$ . This is illustrated in Figure 4 where the

Figure 4: Radius of stability at  $\tilde{s}$

neighborhoods  $\mathcal{N}(\alpha, \tilde{s})$  are circles around  $\tilde{s}$ . The radius of stability of the decision under consideration at  $\tilde{s}$  is then equal to the radius of the largest circle contained in the region of  $\mathbb{R}^2$  where the constraints are satisfied.

The larger the radius of stability of  $x$  at  $\tilde{s}$ , the more robust  $x$  at  $\tilde{s}$ . Thus, the most robust decision can be found by solving this *State-free-payoff* maximin model:

$$r^*(\tilde{s}) := \max_{x \in X} \text{ros}(x|\tilde{s}) \quad (53)$$

$$= \max_{x \in X, \alpha \geq 0} \{ \alpha : \text{constraints}(x, s), \forall s \in \mathcal{N}(\alpha, \tilde{s}) \} , \quad x \in X \quad (54)$$

where, formally, the decision variable is  $(x, \alpha)$  and the state space associated with  $(x, \alpha)$  is equal to  $\mathcal{N}(\alpha, \tilde{s})$ .

Models of this type are used extensively in numerous fields (e.g. von Hoerner 1957, Wilf 1960, Milne and Reynolds 1962, Leont'ev 1975, Hindrichsen and Pritchard 1986, Zlobec 1988, 2009, Charnes et al. 1992, Paice and Wirth 1998, Bingham and Ting 2013).  $\square$

The above illustrations of these two measures of robustness bring out the importance of being conversant with the fundamental distinction between *local* and *global* robustness. Broadly, *global robustness* measures the performance of decisions over their *entire* respective state spaces, whereas *local robustness* measures the performance of decisions over a *neighborhood* around a nominal value of the state variable. Needless to say, this distinction is reminiscent of the distinction between local and global optimization, and of that between a local and a global anesthetic, local and global news, local and global weather, and so on.

Thus, one must never lose sight of the fact that a decision that is locally robust is not necessarily globally robust and vice versa. The implication therefore is that one must ensure that the type of robustness analysis that one opts for is suitable for the task at hand. That is, one must ensure that the robustness analysis that one sets out to conduct is fully compatible with the uncertainty under consideration. And the further implication is that if one sets out to employ a model of local robustness for the purpose of determining robustness under a severe non-probabilistic uncertainty, then the use of such a model must be justified or at least argued for. The issue here is that the variability of the state variable over a *neighborhood* of the state space may not properly represent the variability of the state variable over the state space.

This is illustrated in Figure 5 where the large square represents the state space  $S$  and the small white circle represents a neighborhood of  $S$  over which the local

robustness analysis is conducted. All the elements of  $S$  outside this neighborhood are ignored by the local robustness analysis.

Figure 5: A local robustness analysis over a neighborhood of  $S$  around  $\tilde{s}$

This illustration makes vivid the imperative to provide a cogent explanation, indeed a justification, for using a model of local robustness to determine robustness when the uncertainty under consideration is non-probabilistic, likelihood-free, chance-free, plausibility-free and so on. In other words, one needs to justify on what grounds are all the states outside the small neighborhood shown in Figure 5 left out of the analysis. Indeed, on what grounds is the analysis conducted in the neighborhood around  $\tilde{s}$  rather than around some other state in  $S$ . And, the size of the neighborhood must also be justified.

As far as maximin models are concerned, the distinction between a local robustness analysis and a global robustness analysis is linked to the nature of the state spaces  $S(x), x \in X$ . Thus, if  $S(x)$  represents the set of all possible/plausible states pertaining to decision  $x$ , then the analysis would be global. And if  $S(x)$  represents a neighborhood of the set of all possible/plausible states pertaining to decision  $x$ , then the analysis would be local. The *size robustness* model is an example of a global maximin robustness model and the *radius of stability* model is an example of a local maximin robustness model.

Discussions on local robustness can be found in Sims (2001), Brock and Durlauf (2005), Hafner et al. (2009), Ben-Tal et al. (2009a), Sniedovich (2012, 2012a, 2014) and Hayes et al. (2013).

## 7 Rationality and conservatism

Now that the basic structures of the main types of generic maximin models have been described, and the main points associated with their functioning and implementation have been clarified, we are in a position to take up some of these models' limitations. As it will progressively become clear, methodologically, the central question that we are actually concerned with is this:

- What is the rationale behind the maximin philosophy?

To address this question in a meaningful manner we need to distinguish between two related, yet patently distinct, facts about worst-case analysis which bear on the character of the maximin paradigm. These are:

- The fact that it is often informative/useful/required to have a clear picture of the performance of decisions/systems under worst-case scenarios.
- The fact that decisions that perform well under worst-case scenarios do not necessarily perform well under “realistic” scenarios.

It goes without saying that the maximin paradigm scores well as regards the first fact. However, when it comes to the second, things become a bit more complicated so that special attention is required to deal with it. This is due to the fact that the maximin paradigm is, by virtue of its definition, concerned only with worst-case scenarios, meaning that it remains aloof to so called “realistic” scenarios. This inherent characteristic of the maximin paradigm can be easily (mis)construed as suggesting that it may lead to “irrational” decisions/behavior. And to illustrate, consider the following payoff table:

$p(x, s)$	$s^1$	$s^{(2)}$	$s^{(3)}$	$s^{(4)}$	$s^{(5)}$	$s^{(6)}$	$s^{(7)}$	$s^{(8)}$	$s^{(9)}$	$s^{(10)}$	$SL(x)$	$Avg(x)$
$x^{(1)}$	0	100	200	400	300	500	600	700	800	900	0	450
$x^{(2)}$	3	3	3	3	3	3	3	3	3	3	3	3
$x^{(3)}$	900	200	1	300	400	600	500	800	700	100	1	450.1
$x^{(4)}$	200	800	700	2	500	400	300	100	900	600	2	450.2

(55)

where the  $SL(x)$  column lists the security levels of the decisions and the  $Avg(x)$  column lists the (arithmetic) average payoffs generated by the decisions.

Clearly, the maximin paradigm decrees that the best (optimal) decision is  $x^{(2)}$ , yielding a security level of 3. However, it is equally clear that this decision performs very poorly relative to the other three decisions under *all* scenarios, except the worst ones:  $s^{(1)}$  for decision  $x^{(1)}$ ,  $s^{(3)}$  for decision  $x^{(3)}$  and  $s^{(4)}$  for decision  $x^{(4)}$ .

The question is then what conclusions should be drawn from this fact?

For one thing, it seems rather clear that if one seeks decisions that perform well with respect to worst-case scenarios as well as to “realistic” scenarios, then the maximin paradigm may not be able to meet this objective.

This takes us directly to the related matter of “conservatism”, in fact to the well-known criticism that the maximin paradigm may lead to decisions that are too “conservative”. The following example illustrates this point.

### Example

Suppose that a decision maker is presented with four alternative projects designed to protect a large public facility from natural disasters. The four alternatives differ

in the protection level that they provide, ranging from “None” to “Very High”. The alternatives, the associated costs, and the damage caused under five scenarios are as follows:

Project	level of protection	cost (M\$)	$s^{(1)}$	$s^{(2)}$	$s^{(3)}$	$s^{(4)}$	$s^{(5)}$
$x^{(1)}$	None	0	0	2	5	9	16
$x^{(2)}$	Low	1	0	0	4	7	14
$x^{(3)}$	Medium	3	0	0	0	5	11
$x^{(4)}$	High	6	0	0	0	0	7
$x^{(5)}$	Very High	12	0	0	0	0	0

(56)

Damage (M\$)

Note that the “conservative” costly High level protection ( $x^{(5)}$ ) project provides full protection (no damage) against all the scenarios.

The Total Costs (cost + damage) table pertaining to this problem, and the associated (minimax) security levels are as follows :

M\$	$s^{(1)}$	$s^{(2)}$	$s^{(3)}$	$s^{(4)}$	$s^{(5)}$	SL(x)
$x^{(1)}$	0	2	5	9	16	16
$x^{(2)}$	1	1	5	8	15	15
$x^{(3)}$	3	3	3	8	14	14
$x^{(4)}$	6	6	6	6	13	13
$x^{(5)}$	12	12	12	12	12	12

(57)

The implication is then that the best (optimal) minimax decision in this case is the “conservative”  $x^{(5)}$ . Note that this decision performs better than the other decisions only under scenario  $s^{(5)}$ . Under all other scenarios the other decisions performs much better (in terms of total-cost).  $\square$

The question is then what are we to make of this minimax verdict?

From what we have seen so far, it seems that one can view the inherent “conservatism” of the maximin paradigm from two radically different perspectives.

- Seen as a medium that enables analysis against worst-case scenarios, as well as a practical instrument that provides robustness under these conditions, the maximin paradigm can hardly be labeled “conservative”. Because, in this capacity, it does precisely what it is designed to do, which is to identify decisions that perform well under worst-case scenarios. It is hardly surprising, therefore, that decisions that perform well under worst-case scenarios should prove far

more costly than decisions that perform well under “realistic” scenarios?! One can thus view the conservatism of the maximin paradigm as the *insurance premium* it provides against the great harm/damage that can be caused by the worst-case scenario.

- On the other hand, our life experience (as individuals and as professionals) has taught us that worst-case scenarios are, or ought to be, regarded as *rare events*. Therefore, in practice, in our private lives and as professionals (say, decision/policy makers) our decisions are generally not based solely on grounds of their performance under worst-case scenarios, as advocated by the maximin paradigm. Because, pursuing this type of strategy would clearly impede a “normal” life, economic and social development, and so on.

So, the bottom line is that one must never lose sight of the special character and function of the maximin paradigm, which in Rawls’ illuminating phrasing, puts forward the following *maximin rule* (e.g. Rawls (1971, pp. 152-3):

The maximin rule tells us to rank alternatives by their worst possible outcomes: we are to adopt the alternative the worst outcome of which is superior to the worst outcome of the others.

The implication is therefore that using the maximin paradigm would be sensible and justified in situations where it makes sense to ...rank decisions according to their performance under worst-case scenarios. Using it in other situations may well lead to “irrational” decisions/behavior.

## 8 Variations on a theme

Shortly after its introduction into decision theory in the early 1950s, a number of scholars proposed various modifications to the maximin paradigm. For instance, consider this proposal by Hodges and Lehmann (1952, p. 396):

Instead of minimizing the maximum risk it is proposed to restrict attention to decision procedures whose maximum risk does not exceed the minimax risk by more than a given amount. Subject to this restriction one may wish to minimize the average risk with respect to some guessed a priori distribution suggested by previous experience. It is shown how Wald’s minimax theory can be modified to yield analogous results concerning such restricted Bayes solutions.

A particularly interesting modification was proposed by Savage (1951). It involves the concept “regret” which stipulates the deviation of the actual payoff yielded by a (decision, state) pair from the best payoff yielded by the given state. In other words, the *regrets* associated with the payoffs  $p(x, s)$ ,  $x \in X$ ,  $s \in S$  are as follows:

$$r(x, s) := \left\{ \max_{y \in X} p(y, s) \right\} - p(x, s), \quad x \in X, s \in S \quad (58)$$

observing that the smaller the regret the better.

Thus, decisions are ranked according to their worst (largest) regrets, meaning that the best decision is that whose largest regret is the smallest. This corresponds to the following minimax regret model:

$$w^* := \min_{x \in X} \max_{s \in S} r(x, s). \quad (59)$$

See Luce and Raiffa (1957), Resnik (1987), French (1988), Kouvelis and Yu (1997), Peterson (2009) for other variations on the maximin payoff theme.

## 9 Robust optimization

No discussion on the use of the maximin paradigm in *Operations Research* (OR) can possibly be considered complete without calling attention to the extensive use of this paradigm in the field of *robust optimization* (e.g. Ben-Tal and Nemirovski 1997, Bertsimas and Sim 2004, Iyengar 2005, Lim et al. 2006, Ben-Tal et al. 2006, Beyer 2007, Ben-Tal et al. 2009, Bertsimas et al. 2011, Goerigk and Schöbel 2013, Goerigk 2014, Gabrel et al. 2014).

Recall that *robust optimization* is concerned with methods, techniques, and algorithms designed to find solutions to optimizations problems that are robust against an uncertainty in the problems’ parameters.

To do this, the ruling practice in robust optimization is to transform a conventional optimization problem into a maximin problem. To see how this is done, consider the following generic constrained optimization problem:

$$\text{Problem P: } v^* := \max_{x \in X} f(x) \quad \text{s.t. constraints on } x. \quad (60)$$

where  $f$  is a real-valued function on some set  $X$ . Let  $X^*$  denote the set of optimal solutions to this problem.

Assume further that the objective function and the constraints of this problem depend on some parameter  $s \in S$ . This being so, Problem P can be linked to the following *parametric* optimization problem:

$$\text{Problem PP: } v^*(s) := \max_{x \in X} f(x, s) \text{ s.t. constraints on } (x, s) \text{ pairs, } s \in S. \quad (61)$$

Let  $X^*(s)$  denote the set of optimal solutions to this problem for the specified value of  $s \in S$ .

Ideally, there would be a decision  $x \in X$  that is optimal with respect to Problem PP for all  $s \in S$ . Should such a decision, exists, it would be deemed *super-robust*. However, as we do not operate in an ideal world, we must lower our sights and seek instead a decision that performs well with respect to both the objective function  $f$  and the constraints imposed on  $(x, s)$  pairs, over a large subset of the state space  $S$ .

Assuming that constraint satisfaction takes precedence over payoff, an application of the maximin decision rule to Problem PP yields the following maximin problem:

$$\text{Problem RC: } v^\circ := \max_{x \in X} \min_{s \in S} \{f(x, s) : \text{constraints}(x, s), \forall s \in S\} \quad (62)$$

where  $\text{constraints}(x, s)$  denotes the list of constraints imposed on  $(x, s)$  pairs. In the robust optimization literature this maximin problem is called the *robust counterpart* (RC) problem (e.g. Ronchetti and Staudte 1994, Ben-Tal et al. 2009, Bertsimas et al. 2011). From a maximin perspective, this is the *maximin counterpart* of Problem PP.

The idea of incorporating Wald's maximin rule in optimization problems with the view to obtain solutions that are robust against variations in the optimization problem's parameters dates back to at least the 1960s (e.g. Dorato and Drenick 1966). However, only in the mid 1990s did this idea begin to gain momentum so as to develop into a thriving area of expertise in the field of optimization theory (e.g. Kouvelis and Yu 1997, Ben-Tal and Nemirovski 1997).

Given the history of the maximin paradigm and its extensive use in many diverse fields, the question naturally arises: what is the difference between maximin models used in the field of robust optimization and maximin models used in other fields (statistics, machine learning, control theory, economics, engineering, and so on)?

Broadly speaking, the answer is that maximin robustness models used in *robust optimization* focus on classic mathematical programming problems, such as linear programming, quadratic programming, integer programming, and dynamic programming problems. Unsurprisingly therefore, progress in the field of robust optimization over the last twenty years has been primarily in the development of new algorithms

for the solution of robust counterpart (maximin) problems of classic mathematical programming problems.

The issue here is that a robust (maximin) counterpart of a standard mathematical programming problem, say a linear programming problem, may not necessarily turn out to be a standard mathematical programming problem. Hence the extensive effort exerted in the field of robust optimization to develop algorithms for robustness seeking optimization problems.

### Remark

That said, it is important to note that not all robust optimization models are based on worst-case analysis (e.g. Zenios 1992, Malcolm and Zenios 1994, Mulvey et al. 1995, Vladimirou and Zenios 1997) which means of course that not all robust optimization models are maximin models. The fact remains though that Wald's maximin paradigm currently holds sway in this area of expertise. So much so that according to some scholars (e.g. Bertsimas et al. 2011, Ben-Tal et al. 2015) robust optimization problems are maximin problems. And to illustrate consider the following in Ben-Tal et al. (2015, p. 628):

Robust optimization is a common framework in optimization under uncertainty when the problem parameters are not known, but it is rather known that the parameters belong to some given uncertainty set. In the robust optimization framework the problem solved is a min-max problem where a solution is judged according to its performance on the worst possible realization of the parameters.

Similar statements are made in Bertsimas et al. (2011, pp. 465-6).

To be fair though, in this respect, the field of *robust optimization* is not all that different from other fields dealing with robustness. And to illustrate, consider for instance the following statement from the book *Robust Statistics* by Huber (1981, pp. 16-17):

But as we defined robustness to mean insensitivity with regard to small deviations from assumptions, any quantitative measure of robustness must somehow be concerned with the maximum degradation of performance possible for an  $\epsilon$ -deviation from the assumptions. The optimally robust procedure minimizes this degradation and hence will be a minimax procedure of some kind.

So, to sum it all up, a quick scan of the literatures of robust optimization and of other fields concerned with robustness suggests that Wald's maximin paradigm seems to be the dominant approach in this endeavor.

## 10 Can Wald's maximin paradigm save the world?

As we pointed out at the outset, the idea behind the maximin approach to decision under uncertainty can be summed up in the dictum: *in the face of uncertainty assume the worst*. We might now add that this in fact was Wald's (1939, 1945, 1950) only stated rationale for taking this approach in the face of complete ignorance. To be precise, Wald (1939) proposed that given the severe uncertainty surrounding the problem examined in this article, it is reasonable to minimize the maximum risk.

This patently simple approach to severe uncertainty was both received enthusiastically and severely criticized by some of the prominent scholars of the time.

Thus in the article *The Theory of Statistical Decision* (Savage 1951, p. 59), which *inter alia* put forward an alternative to Wald's maximin rule, namely the *minimax regret rule* (see Section 8), we read the following:

The following general rule, called the minimax principle, is central to the theory of statistical decision functions, at least today. In fact, it is the only rule of comparable generality proposed since Bayes' was published in 1763 ...

The criticism leveled at Wald's paradigm focused mainly on the underlying assumption that *Nature* is a malevolent player. Thus, perhaps not entirely fairly (considering that no reference whatsoever was made in Wald (1939) to *game theory*), Arrow (1951, p. 429) conflated Wald's maximin where the decision maker plays against Nature, with the classic zero-sum two-person game of game theory, to conclude that in the framework of Wald's maximin paradigm this assumption does not seem reasonable.

To illustrate this point, Arrow (1951, p. 429) considered the following payoff table representing two actions (decisions),  $A_1$  and  $A_2$ , and two hypothesis (scenarios),  $H_1$  and  $H_2$ :

$$\begin{array}{c|cc}
 & H_1 & H_2 \\
 \hline
 A_1 & 100 & 0 \\
 A_2 & 1 & 1
 \end{array} \tag{63}$$

Viewed as a zero-sum two-person game, and assuming that the larger the payoff the better, the solution to this game is the pure strategy specified by the *saddle point*

is  $(A_2, H_2)$ . In short, this illustration views Wald's maximin game as a zero-sum 2-person game, where the column player selects column  $H_2$ .

Arrow (1951, p. 429) therefore argued that:

Wald's theory is intuitively highly appealing in that it reflects fully the idea of complete ignorance about the hypotheses. However, the theory of the zero-sum two-person game is based on the idea of an opponent with a definite interest in reducing one's gains (see von Neumann and Morgenstern [1, pp. 98-100, 220]). No such motive can be ascribed to Nature.

and

Wald's criterion would call for choosing  $A_2$  ; yet, since it can hardly be said that Nature would choose  $H_2$  deliberately to prevent the individual from realizing the gains he would receive with  $A_1$  under  $H_1$ , it does not seem reasonable that the hope of a small gain under  $H_2$  should outweigh the possibility of a large gain under  $H_1$ , especially since it can be shown that the choice would not be altered by any accumulation of observations.

Harsher criticism yet was voiced in the article *Foundation of statistics reconsidered* (Savage 1961, pp. 578-9):

The minimax theory, too, can be viewed as an attempt to rid analysis almost completely of subjective opinions, though not of subjective value judgments. From this point of view, the minimax theory of statistics is, however, an acknowledged failure. The minimax rule has never been taken seriously as a philosophic principle for inductive behavior, and even as a rule of thumb little if any good has been found in it; the strongest apology for the rule is perhaps to be found in the latter half of my book [24], especially chapters 10, 11, and 13. Studies of the minimax rule have been stimulating for statistics, and modifications and outgrowths of the rule may prove of great value, but those of us who, twelve or thirteen years ago, hoped to find in this rule an almost universal answer to the dilemma posed by abstinence from Bayes' theorem have had to accept disappointment.

[24] = Savage (1954).

And yet, more than five decades on, Wald's maximin paradigm appears to not only have weathered this criticism but, to be going from strength to strength in many fields.

Looking at the various literatures on its applications, it seems that Wald's paradigm has been used in two capacities:

- As a decision-making paradigm.  
In line with its portrayal in *decision theory* textbooks (e.g. Luce and Raiffa 1957, Resnik 1987, French 1988, Peterson 2009), Wald's paradigm has been used primarily for the technique it provides for ranking decisions/alternatives and selecting the best (optimal) out of the latter, in situations that are subject to a non-probabilistic uncertainty. And this is precisely how it is used in *robust optimization*, where the solution of the robust-counterpart (maximin) problem identifies those decisions that are the most robust against uncertainty.
- As a decision-support tool.  
In this capacity the paradigm is not used for its prescription for selecting the best out of a set of ranked decisions. Rather, the results generated by the paradigm, together with the results generated by other tools, are treated as mere aids that the decision maker consults to select his/her choice of the "best" decision.

Needless to say, it is difficult to assess the maximin paradigm as a decision-support tool without knowing how the results generated by it are incorporated in the decision-making process and what impact do they have, if any, on the choice of the best decision by the decision maker. That said, the following quotes from the famous book *Robustness* by Hansen and Sargent (2007) give an idea of the role that the maximin paradigm can play as a decision-aid:

That is, we use a max-min decision rule. Positing a malevolent nature is just a device that the decision maker uses to perform a systematic analysis of the fragility of alternative decision rules and to construct a lower bound on the performance that can be attained by using them. A decision maker who is concerned about robustness naturally seeks to construct bounds on the performance of potential decision rules, and the malevolent agent helps the decision maker do that.

Hansen and Sargent (2007, p. 12)

### 1.9. Why max-min?

We answer this question by posing three other questions.

1. What does it mean for a decision rule to be robust? A robust decision rule performs well under the variety of probability models depicted in figure 1.7.1. How might one go about investigating the implications of

alternative models for payoffs under a given decision rule? A good way to do this is to compute a lower bound on value functions by assessing the worst performance of a given decision rule over a range of alternative models. This makes max-min a useful tool for searching for a robust decision rule.

Hansen and Sargent (2007, p. 14)

All one can say to that is that, much as computing lower bounds on the performance of decision rules could be a useful thing to do, the main question remains: given these lower bounds and other details about the decision rules under consideration, how do we determine which decision is “best”?

As for the manner in which Wald’s maximin paradigm has been used in OR/MS, consider the following *Summary and Conclusions* section from the INFORMS tutorial entitled *Model Uncertainty, Robust Optimization and Learning* (Lim et al. 2006, p. 90):

The interest in model uncertainty, robust optimization and learning in the OR/MS areas is growing rapidly. The type of model uncertainties considered in the literature can be broadly categorized into three classes: models with uncertainty sets for (1) variables, (2) parameters and (3) measures. The robust optimization approaches used to find (robust or lack thereof) solutions falls into (a) min-max and (b) min-max with bench-marking. Two common ways to bench-marking are through (1) regret and (2) competitive ratio. The main focus in OR/MS has been in the development of models with uncertainty sets for variables (deterministic models of model uncertainty) and deterministic min-max and min-max-regret robust optimization. Within this framework, the focus has been on developing efficient solution procedures for robust optimization. Only a very limited amount of work has been done on looking at stochastic models of model uncertainty and robust optimization with bench-marking. Very little is done in learning. We believe that a substantial amount of work needs to be done in the latter three topics.

To sum it all up then, what conclusions can be drawn from our discussion on Wald’s maximin paradigm?

From what we have seen here it seems fair to say that, methodologically speaking, Wald’s maximin paradigm is essentially an *ad hoc* recipe, which means that its use must be justified. This can be done either on a case by case basis, or by axiomatizing it, namely by grounding it on a set of formal axioms which guarantee its validity—

subject of course to the restrictions imposed by the axioms in question. An example of the latter case can be found in Gilboa and Schmeidler (1989).

And as a final word, let us go back to the question posed by the heading of this section, which was inspired by the title of the discussion on this topic on Prof. Larry Wasserman's website (see Wasserman 2012). Clearly, the answer to this question must be in the negative: Wald's paradigm cannot save the world! However, considering its unabated widespread use, it seems safe to say that this paradigm will continue to play a vital role in decision-making under severe uncertainty. It is important therefore that users of this paradigm make sure that they implement it properly.

As for teaching the topic "Wald's maximin paradigm" is concerned, I submit that there is room to rethink how this topic is taught in OR/MS courses and how it is presented in introductory OR/MS textbooks.

## References

- [1] Arrow, K. J., 1951. Approaches to the theory of choice in risk-taking situations. *Econometrica*, 19(4), 404-437.
- [2] Bell, K.J., 1992. Some observation on the teaching and practice of pessimization. *Heat Transfer Engineering*, 13(1), 5-6.
- [3] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A., (editors) 2006. *Mathematical Programming*, 107(1-2). Special issue on Robust Optimization.
- [4] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A., 2009. *Robust Optimization*. Princeton University Press.
- [5] Ben-Tal, A., Golany B., and Shtern, S., 2009a. Robust multi-echelon multi-period inventory control. *European Journal of Operational Research*, 199, 922-935.
- [6] Ben-Tal, A., Hazan, E., Koren, T., and Mannor, S., 2015. Oracle-Based Robust Optimization via Online Learning. *Operations Research*, 63(3), 628-638.
- [7] Ben-Tal, A., and Nemirovski, A, 1997. Robust truss topology design via semidefinite programming. *SIAM Journal of Optimization*, 7(4), 991-1016.
- [8] Bertsimas, D., Brown. D.B., and Caramanis, C., 2011. Theory and applications of robust optimization. *SIAM Review*, 53(3), 464-501.

- [9] Bertsimas, D., and Sim, M., 2004. The price of robustness. *Operations Research*, 52(1), 35-53.
- [10] Beyer, H-G., and Sendhoff, B., 2007. Robust optimization — a comprehensive survey. *Computer Methods in Applied Mechanics and Engineering*, 196, 3190-3218.
- [11] Bingham, J.T., and Ting, L.H., 2013. Stability radius as a method for comparing the dynamics of neuromechanical systems. *IEEE transactions on Neural Systems and Rehabilitation Engineering*, 21(5), 840-848.
- [12] Brock, W.A., and Durlauf, S.N., 2005. Local robustness analysis: theory and application. *Journal of Economic Dynamics & Control*, 29, 2067-2092.
- [13] Charnes, A., Haag, S., Jaska, P., and Semple, J., 1992. Sensitivity of efficiency classifications in the additive model of Data Envelopment Analysis. *International Journal of Systems Science*, 23, 789-798.
- [14] Corbett, J.O., 1988. Uncertainty in risk analysis: an alternative approach through pessimisation. *Journal of Radiological Protection*, 8(2), 107-117.
- [15] Dorato, P., and Drenick, R.F., 1966. Optimality, sensitivity, and game theory. In *Sensitivity Methods in Control Theory*, (Radanović, L., Editor), pp. 78-102. Pergamon Press.
- [16] Eiselt, H.A., Laporte, G., 1992. The use of domains in multicriteria decision making. *European Journal of Operational Research*, 61, 292-298.
- [17] 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.
- [18] Elzinga, J., and Hearn, D.W., 1972. Geometrical solutions for some minimax Location problems. *Transportation Science*, 6(4), 379-394.
- [19] French, S.D., 1988. *Decision Theory*. Ellis Horwood, NY.
- [20] Gabrel, V., Murat, C., and Thiele, A., 2014. Recent advances in robust optimization: an overview. *European Journal of Operational Research*, 235(3), 471-483.
- [21] Gilboa, I., and Schmeidler, D., 1989. Maximin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18, 141-153.

- [22] Goerigk, M., 2014. ROPI—a robust optimization programming interface for C++. *Optimization Methods and Software*, 29(6), 1261-1280.
- [23] Goerigk, M., and Schöbel, A., 2013. Algorithm engineering in robust optimization. Technical report, Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Universität Göttingen.
- [24] Gupta, S.K., and Rosenhead, J., 1968. Robustness in sequential investment decisions. *Management Science*, 15(2), B18-B29.
- [25] Hafner, M., Koepl, H., Hasler, M., and Wagner, A., 2009. “Glocal” robustness analysis and model discrimination for circadian oscillators. *PLoS Computational Biology*, 5(10): e1000534.
- [26] Hansen, L.P., and Sargent, T.J., 2007. *Robustness*. Princeton University Press.
- [27] Hayes, K.R., Barry, S.C., Hosack, G.R., and Peters, G.W., 2013. Severe uncertainty and info-gap decision theory. *Methods in Ecology and Evolution*, 4, 601-611.
- [28] Herman, J.D., Zeff, H.B., Reed, P.M., and Characklis, G.W., 2014. Beyond optimality: multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty. *Water Resources Research*, 50, 7692-7713.
- [29] Hindrichsen, D., and Pritchard, A.J., 1986. Stability radii of linear systems. *Systems & Control Letters*, 7, 1-10.
- [30] Hodges, J.L., and Lehmann, E.L., 1952. The use of previous experience in reaching statistical decisions. *Annals of Mathematical Statistics*, 23(3), 396-407.
- [31] Huber, P.J., 1981. *Robust Statistics*. Wiley.
- [32] IECRC, 1966. Chemical reaction engineering. *Industrial and Engineering Chemistry*, 58(7), 16-17.
- [33] Iyengar, G.N., 2005. Robust dynamic programming. *Mathematics of Operations Research*, 30(2), 257-280.
- [34] Kouvelis, P., and Yu, G., 1997. *Discrete Robust Optimization and its Applications*. Springer.
- [35] Leont’ev, V.K., 1975. Stability of the travelling salesman problem. *USSR Computational Mathematics and Mathematical Physics*, 15(5), 199-213.

- [36] Lim, A.E.B., Shanthikumar, J.G., and Shen, Z.J.M., 2006. Model uncertainty, robust optimization and learning. In Johnson, M.P., Norman, B., and Secomandim, N. (editors) *Tutorials in Operations Research*, pp. 66-94. INFORMS.
- [37] Luce, R.D., and Raiffa, H., 1957. *Games and Decisions: Introduction and Critical Survey*. Wiley.
- [38] Malcolm, S., and Zenios, S.A., 1994. Robust Optimization for power capacity expansion planning. *Journal of the Operational Research Society*, 45, 1040-1049.
- [39] 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.
- [40] Mulvey, J.M., Vanderbei, R.J., and Zenios, S.A., 1995. Robust optimization of large-scale systems. *Operations Research*, 43(2), 264-281.
- [41] Mutapcic, A., and Boyd, S., 2009. Cutting-set methods for robust convex optimization with pessimizing oracles. *Optimization Methods & Software*, 24(3), 381-406.
- [42] Oskooi, A., Mutapcic, A., Noda, S., Joannopoulos, J.D., Boyd, S.P., and Johnson, S.G., 2012. Robust optimization of adiabatic tapers for coupling to slow-light photonic-crystal waveguides. *Optics Express*, 20(19), 21558-21575.
- [43] 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.
- [44] Peterson, M., 2009. *An Introduction to Decision Theory*. Cambridge University Press.
- [45] Rawls, J., 1971. *Theory of Justice*. Belknap Press, Cambridge, MA.
- [46] Resnik, M.D., 1987. *Choices: an Introduction to Decision Theory*. University of Minnesota Press, Minneapolis, MN.
- [47] Ronchetti, E., and Staudte, R.G., 1994. A robust version of Mallows's  $C_p$ . *Journal of the American Statistical Association*, 89(426), 550-559.
- [48] Rosenblat, M.J., 1987. A robust approach to facility design. *International Journal of Production Research*, 25(4), 479-486.
- [49] Rosenhead, J., Elton, M., and Gupta, S.K., 1972. Robustness and optimality as criteria for strategic decisions. *Operational Research Quarterly*, 23(4), 413-431.

- [50] Savage, L.J., 1951. The theory of statistical decision. *Journal of the American Statistical Association*, 46(253):55-67.
- [51] Savage, L.J., 1954. *The Foundations of Statistics*. Wiley, NY.
- [52] Savage, L.J., 1961. The foundation of statistics revisited. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Volume 1, pp. 575-586. University of California Press.
- [53] Sheynin, O.B., 1996. Origin of the theory of errors. *Nature*, 211, 1003-1004.
- [54] Sims, C., 2001. Pitfalls of a minimax approach to model uncertainty. *American Economic Review*, 91, 5-4.
- [55] Sniedovich, M., 2003. OR/MS Games: 3. counterfeit coin problem. *INFORMS Transactions on Education*, 3(2), 1-10.
- [56] Sniedovich, M., 2012. Black swans, new Nostradamuses, voodoo decision theories and the science of decision-making under severe uncertainty. *International Transactions in Operational Research*, 19(1-2), 253-281.
- [57] Sniedovich, M., 2012a. Fooled by local robustness. *Risk Analysis*, 32(10), 1630-1637.
- [58] Sniedovich, M., 2014. The elephant in the rhetoric on info-gap decision theory. *Ecological Applications*, 24(1), 229-233.
- [59] Starr, M.K., 1963. *Product design and decision theory*. Prentice-Hall.
- [60] Starr, M.K., 1966. A discussion of some normative criteria for decision-making under uncertainty. *Industrial Management Review*, 8(1), 71-78.
- [61] Sylvester, J.J., 1857. A question in the geometry of situation. *Quarterly Journal of Pure and Applied Mathematics*, 1, 79.
- [62] Thie, P.R., 1988. *An Introduction to Linear Programming and Game Theory*. Wiley.
- [63] Vladimirov, H., and Zenios, S.A., 1997. Stochastic Programming and Robust Optimization. In Gal, T., and Greenberg, H.J. (editors), *Advances in Sensitivity Analysis and Parametric Programming*, Chapter 12. Kluwer, Boston.

- [64] von Hoerner, S., 1957. Internal structure of globular clusters. *The Astrophysical Journal*, 125, 451-469.
- [65] von Neumann, J., Morgenstern, O., 1944. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton.
- [66] Wald, A., 1939. Contributions to the theory of statistical estimation and testing hypotheses. *Annals of Mathematical Statistics*, 10(4), 299-326.
- [67] Wald, A., 1945. Statistical decision functions which minimize the maximum risk. *The Annals of Mathematics*, 46(2), 265-280.
- [68] Wald, A., 1950. *Statistical Decision Functions*. John Wiley, NY.
- [69] Wasserman, L., 2012. Minimax Theory Saves the World. See <https://normaldeviate.wordpress.com/2012/07/17/minimax-theory-saves-the-world/>
- [70] Whateley, S., Steinschneider, S., and Brown, C., 2014. A climate change range-based method for estimating robustness for water resources supply. *Water Resources Research*, 50, 8944-8961.
- [71] Wilf, H.S., 1960. Maximally stable numerical integration. *Journal of the Society for Industrial and Applied Mathematics*, 8(3), 537-540.
- [72] Zenios, S. A., 1992. Robust Optimization for Matrix Balancing From Noisy Data. Report 92-01-02, Decision Sciences Department, University of Pennsylvania, Philadelphia.
- [73] Zlobec, S., 1988. Characterizing optimality in mathematical programming models. *Acta Applicandae Mathematicae*, 12, 113-180.
- [74] Zlobec, S., 2009. Nondifferentiable optimization: parametric programming. In Floudas C.A., and Pardalos, P.M. (editors), *Encyclopedia of Optimization*, pp. 2607-2615. Springer.