Fooled by Robustness

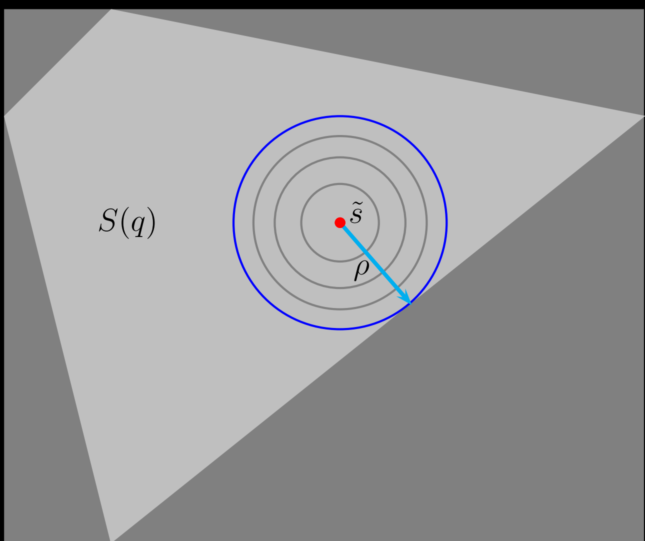
A Perspective from Down Under – the Land of the Black Swan

including a comprehensive critique of info-gap decision theory

Phascolarctos cinereus

By Ben Twist

Moshe Sniedovich



$$\max\{\rho \geq 0 : s \in S(q), \forall s \in B(\rho, \tilde{s})\}$$

Fooled by Robustness:
A Perspective from Down Under the Land of the Black Swan

Including a Comprehensive Critique of Info-Gap Decision Theory

---

Moshe Sniedovich
The University of Melbourne
Melbourne, Australia
www.moshe-online.com

Read Me First

Welcome to **Fooled By Robustness!**

This is a preview draft of the book that can be regarded at this stage as a β version. The chapters included in this version should give you a pretty good idea of what the final versions of the chapters will be like.

Please respect the copyright and refrain from copying/quoting from the material.

I plan to complete the project in the middle of the year (2011) and publish it in the second part of 2011.

Your comments on this project, especially constructive criticism, will be greatly appreciated!

Best wishes

Moshe

Preface

In his two best-selling books, *Fooled by Randomness* and *The Black Swan: The Impact of the Highly Improbable*, Nassim Taleb makes a passionate argument for the proposition that we have a propensity for being fooled by randomness. This, he argues, is manifested *inter alia* in our inclination to set great store by, indeed accept almost uncritically, the models and theories that were developed for risk management under uncertainty.

The latest edition of *The Black Swan* features a new section entitled *On Robustness and Fragility* and in the new preface Taleb (2010, p. xxiv) writes:

> It is much easier to deal with the Black Swan problem if we focus on robustness to errors rather than improving predictions.

This may well be so, however ...

As I show in this book, it is (almost) as easy to be fooled by *robustness* as it is to be fooled by *randomness!*

I demonstrate this point through a comprehensive examination of a theory that — according to its proponents — was developed expressly to furnish a reliable method for robust decision-making under severe uncertainty. This theory is known as: *info-gap decision theory*.

The objective of this book is thus twofold:

- To provide a gentle introduction to the extremely difficult problem of decision-making in the face of severe uncertainty.
- To provide a critical, but at the same time edifying and constructive analysis, of *info-gap decision theory.*

The book, however, is self-contained, meaning that no prior knowledge of *info-gap decision theory* is assumed.

As for my treatment of *severe uncertainty.*

On the whole, I follow the conventions that have been adopted in classical *decision theory* and *robust optimization,* which means that I approach severe uncertainty through the medium of non-probabilistic, likelihood-free models. Some of my Bayesian colleagues may no doubt protest the exclusion of probabilistic models from the discussion. To this I reply that, reformed Bayesian though I am, I accept that Bayesian models may offer a viable alternative to some of the approaches that I discuss in this book.

Indeed, I note this fact in the discussion on the modeling of severe uncertainty. However, I do not go into the technical issues that are involved in a probabilistic modeling of severe uncertainty. Rather, I leave it to my Bayesian colleagues to argue the case for the use of probabilistic models to model severe uncertainty.

Be that as it may, one of the immediate consequences of this non-probabilistic approach to uncertainty is that it is rendered accessible to readers/users who are not conversant with probability theory and statistics.

Still, to all those who may wonder at the proposition to "deal" with severe uncertainty by means of non-probabilistic models, indeed may even find it untenable, I want to point out the following.

It has been a long-standing tradition in *decision theory* to model robustness against severe uncertainty by using the paradigm of a game between *Nature* — assigned the role of *Uncertainty* — and the decision maker (DM). In this setting, *Nature* functions as an *adversarial opponent* of the decision maker so that the decision models yielded by this approach are *worst-case* type models.

The upshot of this is that the robustness models that I discuss in this book are *worst-case,* rather than probabilistic, type models. That is, the realization of the parameter whose (unknown) true value is subject to severe uncertainty is not determined by a probabilistic model, but rather is assumed to be equal to the *worst value* of this parameter with respect to the stipulated goals and requirements of the problem under consideration.

As for the *info-gap decision theory* content of the book.

Since the end of 2003 I have been embroiled in a protracted discussion on the obvious — and not so obvious — fundamental flaws afflicting *info-gap decision theory*. As a participant in this discussion, I have written a number of articles and have given many lectures and presentations on this topic.

I want to make it clear, though, that by referring to my participation in this discussion in these terms, I do not mean to suggest that this discussion has been nothing but a source of grief to me. To the contrary, I have enjoyed it immensely because of the unrelenting challenge to continually improve my explanations of the ills afflicting *info-gap decision theory*, and by association, my elucidations of the basic issues bearing on the treatment of *severe uncertainty*.

The trouble is, however, that to some extent this discussion has proved pointless because those *info-gap aficionados* who have taken part in this discussion have preferred to maintain their commitment to *info-gap decision theory* regardless, and in spite of the facts facing them. That is, rather than buckle down to a serious study of the problem of decision under severe uncertainty, and by implication to a careful analysis of the facts about *info-gap decision theory,* most of my disputants have preferred to uphold their continuing commitment to *info-gap decision theory* by means of . . . rhetoric.

This, by the way, was one of the triggers for the Campaign that I had launched at the end of 2006 to contain the spread of *info-gap decision theory* in Australia — the Land of the Black Swan.

More than anything else, this discussion has brought home to me the importance of spelling out in concentrated form the basic flaws of *info-gap decision theory* and of bringing to the attention of the widest readership possible how these flaws are explained away, or misrepresented, by rhetoric.

However, considering the volume of rhetoric that forms part of the books and articles on *info-gap decision theory* and its applications, I judged that combining these two aspects in one book may yield a far too lengthy volume.

I therefore decided to divide this project between two books.

In this short book I address the *technical* aspects of *info-gap decision theory*. That is, I explain what this theory is all about, what renders it a classic example of a *voodoo* decision theory, and I clarify how its proponents misconstrue its role and place in *decision theory*. I also explain the errors in the arguments advanced by *info-gap scholars* in their response to my criticism of *info-gap decision theory*.

In a second book that I hope to complete in the not too distant future, a book provisionally entitled

The Rise and Rise of Voodoo Decision Theories

I plan to raise the more general question of:

> *How is it that senior scientists/analysts can (so easily) fall for flawed theories such as info-gap decision theory?*

In this framework I intend to use *info-gap decision theory* as a case in point, a sort of a *case study*. Thus, the questions that I intend to take up in the envisioned book are questions that people often ask me about *info-gap decision theory*:

> Given that the flaws in *info-gap decision theory* are so fundamental and so obvious, who actually uses this theory?
>
> Moreover, how can its followers justify the use of such a flawed theory?

Although the present book does not set out to formulate direct answers to these intriguing questions, I believe that my discussions here will give the reader a pretty good idea of the answers that I aim to work out more fully in the envisioned book *The Rise and Rise of Voodoo Decision Theories.*

At present I am unclear as to the envisioned book's date of completion. But if you are interested in this subject, you can visit the book's website

`rise-and-rise.moshe-online.com`

where you will find additional information about it.

As for the style of presentation that I adopted in this book.

I have tried to make the general discussion as non-technical and as *math-free* as possible. However, it must be appreciated that because the key elements in *info-gap decision theory* are mathematical models, any serious discussion or assessment of the theory requires a formal, technical, mathematical treatment thereof.

Finally, a few words about the envisioned readership of this book.

This book was written primarily to benefit *info-gap scholars/analysts* in that one of its main aims is to set aright the many misconception, errors, and misguided ideas that circulate in the *info-gap literature.* However, given that this discussion is conducted in the general framework of the treatment of *robustness to severe uncertainty*, rather than in the more limited framework of *info-gap decision theory,* the book should be of interest to scholars/analysts who are interested in the *modeling aspects* of decision-making under conditions of severe uncertainty.

Indeed, I regard this book as a general "soft" introduction to the conceptual and modeling aspects of robust-decision-making in the face of severe uncertainty.

And as a final note:

> I dedicate this book to all the *info-gap scholars* with whom I have been in contact during the last seven years, hoping that they will enjoy reading it as much as I enjoyed writing it.

Moshe Sniedovich

February 15, 2011
Melbourne
The Land of the Black Swan

Contents

III Epilogue 325

Part I

Preliminaries

Chapter 1

Introduction

The main topic of discussion in this book is *robustness against severe uncertainty*. At this preliminary stage I need only indicate very broadly that, in general, this property is understood to connote resilience to the impact of unforeseen contingencies. So, in this vein, I shall understand this property to designate the ability of a system to perform effectively under conditions that are subject to severe uncertainty.

My investigation of this property will, however, be more narrowly targeted in that I shall discuss it here as a crucial element of *sound decision-making*. Hence, my discussion of this topic will revolve around these two central questions:

· How do we measure the robustness of a decision to severe uncertainty in a decision-making environment?

· How do we incorporate such measures of robustness in formal decision-making models?

I plunge head on into the study of these two intricate questions in the next chapter. However, before I do this, I want to explain a number of things about the book's character and scope and to suggest to the reader some general guidelines on how to read it.

1.1 Scope

It is important to appreciate that any serious discussion of the main issues pertaining to the problem of *robust decision-making in the face of severe uncertainty* requires a formal clarification of the concepts *robustness* and *severe uncertainty*. This is so because although on the face of it these concepts may have a familiar ring to them, they can have different interpretations in different contexts. My formal clarification of the notions of *robustness* and *severe uncertainty* will be conducted in the context of a detailed analysis of the topics of:

· *Worst-case analysis.*

· *Wald's Maximin model.*

· *Radius of Stability model.*

As the reader will progressively see, one of the interesting points that will emerge from this analysis is that the *Radius of Stability Model* (circa 1960) — a staple model of *local stability/robustness* — is in fact a simple instance of *Wald's Maximin model* (circa 1940).

But what is more, I shall show that a robustness model that in recent years has created quite a stir in a number of areas of expertise (for instance, conservation biology and applied ecology) namely, *info-gap* *decision theory's robustness model*, is in fact a simple instance of these two well known models.

In other words, *info-gap's robustness model* will be used in this discussion as a *case* illustrating, among other things, the relation between *Wald's Maximin model* and the *Radius of Stability model.*

As a matter of fact, *info-gap's robustness model*, or more generally *info-gap decision theory*, will be the only *Case Study* that will be featured in this book. So, the question obviously arising is this:

> What is it about *info-gap decision theory* that renders it so suitable a medium to serve as the sole *Case Study* in this book?

The answer to this question is very simple:

> It is not on account of its great merit as a decision theory that *info-gap decision theory* was chosen to serve as the sole case study in this book.
>
> To the contrary!
>
> It was chosen to function in this role because it provides an extremely edifying medium for demonstrating, to the novice and the expert alike, what a theory/method for robust decision in the face of uncertainty **ought not be**.
>
> So, one of the by-products of my investigation here is setting the record straight on the utterly misleading portrait painted of this theory in the *info-gap literature.* I make it abundantly clear that readers of this literature are practically lead astray with respect to the very character of *info-gap decision theory,* the role and place that it has in *decision theory,* the mode of operation, scope and capabilities of its robustness model, and many other associated matters as well.
>
> In a word, I demonstrate in this discussion what makes *info-gap decision theory* **the wrong theory** for robust decision-making in the face of uncertainty.

I go into these matters in the second part of the book. In the first part I concentrate on the conceptual and technical analysis of the notion *robustness against severe uncertainty.*

Now, such an analysis can of course be conducted from a variety of viewpoints, including the following:

· Models

· Algorithms

· Applications

My discussion in this book is devoted almost exclusively to the question of the *modeling* of *robustness*, notably the modeling of *robustness to severe uncertainty.* This means that my main thrust here is on the formulation of *robustness* models. My discussion of algorithms and of applications, is terse. However, I make it my business to refer the reader to the relevant literatures on these important topics.

Also, although I discuss in detail the mathematical aspects of robustness, this discussion should not be too demanding technically. Moreover, I shall endeavor, wherever possible, to illustrate the formal mathematical arguments and constructs graphically. I therefore believe that my mathematical treatment of this notion should not cause readers with limited mathematical backgrounds too much grief.

As for my characterization of *info-gap decision theory* in this book.

Readers who are familiar with the *info-gap literature,* in particular the three primary texts on *info-gap decision theory* (Ben-Haim 2001, 2006, 2010), will immediately see the stark contradiction between my characterization of *info-gap decision theory* and its depiction in these texts. Thus:

> While in Ben-Haim (2001, 2006, 2010) *info-gap decision theory* is presented as a distinct, new theory that is radically different from all current theories for decision under uncertainty, I show it for what it is. That is, I demonstrate that its robustness model is in fact a reinvention of a well-established concept/model — known universally as *Radius of Stability model* — that has been use extensively in numerous fields since the 1960s.

The other important difference is that:

> While in Ben-Haim (2001, 2006, 2010) *info-gap decision theory* is portrayed as a theory that is singularly suitable for the treatment of **severe** uncertainty, I show it is for what it is. That is, I demonstrate that it is in fact a theory prescribing a *local* robustness analysis of the *Radius of Stability* type which therefore renders it **utterly unsuitable** for the treatment of severe uncertainty.

To round out my explanation of the book's character and scope, I note very briefly an issue which I do not take up here but which I suspect some reader would have expected me to take on in this book.

I imagine that some readers may be puzzled by the stark incongruity between my portrayal of *info-gap decision theory,* as a manifestly fundamentally flawed theory, and the fact that a significant number of publications dealing with *info-gap decision theory* and its applications have appeared in a wide range of **peer-reviewed journals.** Thus, one may well ask:

> How is it that such a wide cross-section of peer-reviewed journals have published articles advocating the use of a theory that is so patently flawed?

This, of course, is a thoroughly legitimate question, but as indicated above, I do *not* address it in this book.

I plan to deal with it, in due course, in the sequel to this book which, as I indicated already, is tentatively entitled *The Rise and Rise of Voodoo Decision Theories*. Readers who are eager to read my analyses of publications dealing with *info-gap decision theory* are referred to my compilation of reviews of such publications on my website[1]. In Chapter 12 I discuss a number of such peer-reviewed articles.

1.2 Structure

The book is divided into three parts. In the first part, prior to turning my attention to our main topic of interest: *robustness against severe uncertainty*, I expound the background material that is required for a formal treatment of this topic. The discussion in this part of the book is thus devoted to the elucidation of these subjects:

- Illustrative example
- Radius of stability
- Severe uncertainty
- Worst-case analysis
- Wald's Maximin model
- Robustness against severe uncertainty (main topic)

[1] See http://info-gap.moshe-online.com/reviews.html

The second part of the book is devoted to a formal investigation of *info-gap decision theory*. This consists of a formal examination of *info-gap's uncertainty model*, *info-gap's robustness model*, *info-gap's opportuneness model*, *info-gap's decision models* and other related topics.

Following this I explain in detail the flaws in this theory, and I go into the technical arguments that are often adduced to presumably rebut the criticism leveled at various aspects of the theory. The program for this part is then as follows:

· First encounter
· Critique
· FAQs about info-gap decision theory
· The Maximin saga
· The power of the written (peer-reviewed) word!
· Robustness against severe uncertainty (main topic)

In the third part of the book, consisting of one short chapter, I reflect on the title on this book.

1.3 How to read this book

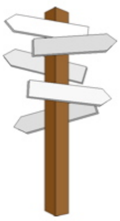
To give readers an idea of how they might approach this book, I should perhaps explain how I approached the task of writing it.

My guiding idea was that the book should be accessible to readers who are completely new to the area of *robust decision-making under sever uncertainty,* but at the same time, should prove interesting to those readers who are at home in this area of expertise. As might be expected, therefore, the book is tilted more towards the novice than the expert.

For one thing, my exposition of the material — which draws on the articles that I have written on this subject over the past five years — develops in a specific order. This order of presentation is intended to ease the novice into the complex world of *robust decision-making in the face of severe uncertainty* in a manner that will facilitate a good understanding of the issues involved.

Hence, my general advise is as follows:

· If you are new to the subject, read the first part of the book in sequence.
· If you are conversant with the subject, you can skip chapters/sections covering topics that you are familiar with.
· Consult the comprehensive *Index* for key words.



I posted road-signs, such as the one depicted here, at various points in the margins to help readers navigate their way through the book with ease. It goes without saying that following these signs is not essential/mandatory, so that they can be ignored altogether.

1.3.1 About the readers

I now address the various types of envisioned readers, commenting on what they may or may not find in this book.

Students of *decision-making in the face of severe uncertainty*

Having been engaged in the study of decision-making in the face of severe uncertainty for at least three decades, I am familiar with a good deal of the literature written on this subject.

Based on my acquaintance with this literature, I believe that my treatment of this topic in this book should be of interest to a wide cross-section of readers who would be involved in the research/application of this topic as operations research experts, statisticians, decision analysts, economists, and so on.

I submit that my analysis of the questions that are central to this topic is most instructive, because in large part, it approaches these questions from the standpoint of the errors that one might fall into due to the complexity, or if you will, the abstruseness of this subject.

I therefore urge all readers, especially those who specialize in *decision-making under severe uncertainty* to read my critique of *info-gap decision theory* in the second part of the book, as my investigation of this theory illuminates the main issues that one needs to wrestle with when dealing with the difficult problem of *decision-making under severe uncertainty*.

Titbits seekers

Some readers would have a pretty good idea of my activities over the past seven years in connection with *info-gap decision theory*, notably my encounters with directors of research centers, editors of journals, and ordinary info-gap users. Yet, there is no trace of these matters in this book. In other words, there is no account of what one might call very broadly the "politics behind the science".

So, instead of looking for savory morsels about the goings on behind the scenes, I suggest that you read this book seeking an answer to the following no less interesting questions:

- Given that the basic criticism documented in this book has been in the public domain since the end of 2006, and given that senior *info-gap scholars* have been aware of this criticism for a number of years now, how is it that the main stream *info-gap literature* is completely oblivious to this criticism?
- Indeed, how is it that the same errors, unfounded claims, misrepresentations etc. are continued to be promulgated in the *info-gap literature*, for instance in Ben-Haim (2010)?

I submit that there is enough intriguing material in these *facts* about *info-gap decision theory* that there is no need to look for more.

Scholars sitting on the fence

Those scholars who have declared themselves "undecided" as to whether my criticism of *info-gap decision theory* does indeed pull the rug out from under this theory, have an opportunity to read my criticism in concentrated form in this book and to decide which way to jump ...

As I make it crystal clear in this book, *info-gap decision theory* is in fact a re-invention of a very simple, well established idea, which alas, amounts to a *misapplication* of this simple idea.

The formal proofs demonstrating these facts are short and straightforward. Therefore, any stated position on *info-gap decision theory* must be based on a direct engagement with these formal proofs. Those scholars who are unable to do so, in effect declare themselves unqualified to express expert opinions on the role and place of *info-gap decision theory* in *decision theory* and *robust optimization.*

Robust optimization experts

There may not be a great deal of new (technical) material in this book for experts in *robust optimization*.

However!

I submit that these experts should find my exposition of standard topics such as *worst-case analysis* and *Wald's maximin model* innovative and refreshing. Also, my discussion on the *Radius of Stability model* may convince them that this topic belongs in *robust optimization.*

I also take this opportunity to call their attention to the fact that there is no point in dismissing *info-gap decision theory* out of hand simply because it does not concern itself with *algorithms* for the robust optimization problems defined by its robustness model. Nor is there any point in dismissing it on grounds that it is used primarily by scholars and analysts with limited background in classical *decision theory* and *robust optimization.*

Rather, *robust optimization* experts should look at their own camp and ask themselves whether they have done enough to penetrate those areas of expertise (for instance applied ecology, conservation biology) where the need for a robust optimization analysis is of the essence.

Bayesians

To readers of the Bayesian persuasion I say that there is no point in rejecting out of hand approaches to (severe) uncertainty solely on the grounds that their uncertainty models are non-probabilistic. In this book, non-probabilistic approaches to uncertainty are taken for what they are, namely as . . . approaches. The question then is whether these approaches make sense, are valid, realistic, and so on, within the perimeters of the *assumed* formal paradigm.

The point I want to stress in this book is then that a non-probabilistic approach to uncertainty ought to be judged on its own merits. That is, one ought to ascertain whether its uncertainty model properly or adequately (or, what have you), deals with the (severity) of the uncertainty postulated by the approach.

Thus, as we shall see, this is precisely were the trouble lies in the case of *info-gap decision theory*. The trouble with it is not in its non-probabilistic approach to (severe) uncertainty. Rather, the trouble with this theory is that its uncertainty model and its robustness model do not even begin to address the *severity* of the uncertainty postulated by it.

This fundamental flaw in *info-gap decision theory* renders it *self-contradictory,* or as I say, a *voodoo decision theory par excellence.* Because, this flaw forces the theory to violate the following two universally accepted maxims:

- Garbage In — Garbage Out (GIGO)
- Results are only as good as the estimates on which they are based.

Indeed, the more severe the uncertainty, the greater *info-gap decision theory's* violation of these maxims. It is therefore ironic in the extreme that *info-gap decision theory* is being promoted (Ben-Haim 2001, 2006, 2010) as a theory providing a reliable, realistic method for the management of an *extremely severe* type of uncertainty.

I elaborate these matters in the second part of the book.

Info-gap users

I realize that it is no easy matter to accept that a theory that one has become strongly committed to is in fact fundamentally flawed. I also realize that it is far easier to extend an error than admit to a mistake.

Yet, as I show in this book, the basic facts about *info-gap decision theory* are crystal clear:

- *Info-gap's robustness* model is a simple *Radius of Stability model* (circa 1960).

· *Info-gap's robustness model* is a simple instance of *Wald's Maximin model* (circa 1940).
· *Info-gap's robustness model* does not (indeed is in principle unable to) seek decisions that are robust to severe uncertainty.
· All that *info-gap's robustness model* is capable of is, to seek decisions that are robust against small perturbations in the nominal value of the parameter of interest.

All this is spelled out in great detail in this book. So, simply read it!

Referees of articles on *info-gap decision theory*

Consult this book to find out how erroneous/groundless/misleading are the statements describing the mode of operation, capabilities, and scope of *info-gap's robustness model,* in the papers that you had refereed. You should keep this in mind when next you are asked to referee other papers on *info-gap decision theory*.

I also urge you to read my compilation of reviews of publications on *info-gap decision theory* that is available on my website[2].

Editors of journals that published articles on *info-gap decision theory*

Refer your reviewers to this book in preparation for their assessment of articles on *info-gap decision theory*.

You may also want to refer them to my compilation of reviews of publications on *info-gap decision theory* that is available on my website[3].

Reviewers of applications for research grants based on *info-gap decision theory*.

See my advice to Referees and Editors, above.

PhD students

I remind PhD students, and MSc students as well, that it is generally expected that a PhD dissertation ought to examine the capabilities and limitations of the methods/theories adopted in the research work. So should you decide to commit to *info-gap decision theory,* you are expected to include a discussion on the limitations of this theory in your thesis.

Furthermore, you are also expected to demonstrate that you understand how *info-gap decision theory* is related to other theories for decision under severe uncertainty.

Since the "official" literature on *info-gap decision theory* does not even begin to address these issues, I am confident that you'll find this book an important and useful resource.

Indeed, I should add that one of the main reasons that prompted me, at the end of 2006, to launch my campaign to contain the spread of info-gap decision theory in Australia, was . . . to alert PhD students in Australia and elsewhere, and their supervisors, to the obvious limitations of this theory and to the fundamental flaws afflicting it.

All this is based on a number of MSc and PhD theses on *info-gap decision theory* that I have read in the last few years.

[2] See http://info-gap.moshe-online.com/reviews.html
[3] See http://info-gap.moshe-online.com/reviews.html

Info-gap aficionados

I do not expect you to severe your ties with this theory. Still, I strongly recommend that you read this book carefully so as to at least familiarize yourself with the *facts* about it.

1.4 Style

While I deliberately set out to create a relaxed atmosphere in this book, I have gone to great lengths to give my treatment of the central concepts and the main results a strictly formal presentation.

This is in line with my conviction that a "relaxed" narrative or pictorial exposition of an issue can have great pedagogical merit, as they can be greatly illuminating. But, if the basis is a mathematical model, then the "relaxed exposition" must be meticulously consistent with the model underlying it.

In our case, my discussion of the various elements that make up the concepts of *robustness, uncertainty* etc. is in parts "relaxed" [4]. However, as this discussion is grounded in formal mathematical models, I have taken every precaution to assure that the narrative analyses and the conclusions are fully consistent with the formal mathematical models that gave rise to them.

I should add that the imperative to maintain a strict correspondence between the narrative and the "mathematics" is one of the main lessons learned from what I call the *Info-Gap Experience*. In the case of *info-gap decision theory*, the (total) incongruity between the narrative (rhetoric) about the mathematical model and the model itself has resulted in a thoroughly misleading picture of what the theory actually is and does.

The method of exposition that I adopted in this book is similar to the method that I had used for many years in my lecture notes on introductory topics in applied mathematics.

As a rule, all the key concepts are explained in three different ways:

- Formally, that is mathematically.
- In plain language.
- Graphically.

I suspect that some readers may find this type of exposition belabored, because in our case the mathematical models under consideration are truly very simple. I imagine, therefore, that these readers may think it unnecessary to engage in so much explanation of what (to some) are clearly elementary concepts.

I put it to these readers, therefore, that my decision to adopt this style of presentation is based on my experience of the last seven years. That is, my experience has shown that users of *info-gap decision theory* — highly qualified mathematicians included — have serious misconceptions not only about what this theory is and does, but also about some simple concepts and ideas that come into play in this theory.

I therefore feel thoroughly justified in adopting a style of presentation where seemingly elementary concepts or constructs are explained from three different perspectives.

About **quotes.**

As you will see, I tend to quote liberally from the *info-gap literature.* My objective is twofold. First, to let this literature "speak for itself". That is, to let this literature itself attest to the plethora of errors and misconceptions that are promulgated by *info-gap decision theory*.

[4] After all, I have been living in the *Land of No Worries!* for more than twenty years now!

Second, my experience over the past seven years has shown that without backing up a reference to a position attributed to *info-gap proponents* by quotes, there is a danger of being blamed of misconstruing/misrepresenting statements made by them. It is therefore best to let *info-gap scholars* do the talking.

About the **ONIAAT and BME Principles**.

I had spent most of my compulsory military service as an instructor in a training program for operators of — what at that time — were sophisticated weapon systems.

The program's commanding officer had a number of definitive ideas on the method that ought to be used to this end. First, every effort had to be made to keep the trainees ... wide awake during the lectures.

Second, the delivery of the material had to be guided by two principles:

- The "One New Idea at a Time" Principle (ONIAAT).
- The "Beginning, Middle, End" Principle (BME).

The first principle assumes that people in general, and trainees in particular, find it difficult to cope with more than one new idea at a time.

The second principle assumes that one can never be sure in what part of a lecture a typical trainee is fully awake. Hence, to play it safe, the main (new) idea must be delivered three times during a lecture: at the Beginning, in the Middle, and at the End of the lecture.

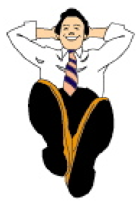
I discovered, during my academic career, that these two Principles — subject to some improvisation — work in academic institutions as well. I imagine that I have learned to apply them subconsciously.

So, when reading this book, some readers would perhaps find it helpful to ask themselves at various stages of the discussion:

- How does this idea, concept etc. sit in the chapter/section that I am reading now?
- In what context have I already encountered this idea, concept in the book?

This will enable readers, especially those who are new to this subject, to keep track of how specific ideas are being developed at various contexts of the discussion. Obviously, you can also consult the extensive *Index* to this end.

1.5 About Fred



Throughout the book I shall use the services of a (fictitious) maven on robust decision-making under severe uncertainty. Among his numerous happy clients around the world he is known, fondly, as *Mr. Robustness*. But I call him *Fred*.

The arrangement I have with Fred is that, from time to time, I report on his opinions and advice on conceptual and technical questions associated with the discussion in the book. For the most part, I share his views on robustness and related topics.

I do like his style!

I apologize in advance for some of his spontaneous interruptions. He takes the *Robustness* business very seriously, and there are occasions where he can barely contain himself. It is important that you realize that he means well.

1.6 fooled-by-robustness.moshe-online.com

I would like to draw the reader's attention to the fact that over the past four years I have posted a considerable amount of material on my website

`www.moshe-online.com`

pertaining to the topics discussed in this book. I shall refer to specific directories of this site in subsequent chapters. At this stage it suffices to mention that the directory

`decision-making.moshe-online.com`

is a good starting/meeting point.

The URL of the directory dedicated to this book is:

`fooled-by-robustness.moshe-online.com`

I plan to maintain it in support of the book.

Chapter 2

A Naive Illustrative Example

2.1 Introduction

I begin the discussion on *robust decision-making in the face of severe uncertainty* with an examination of an extremely simple, indeed naive problem, through which I want to illustrate how decision-makers would, or might, approach this task. That is, I describe in very broad strokes the positions that decision-makers might take when confronted with situations where knowledge/information/data/understanding about a critically important element (or elements) of a problem they are facing, is/are shrouded in uncertainty.

My main objective is to clarify and give content to the concept *robustness*, because, as I indicated at the outset, *robustness* is regarded a crucial factor in *sound* decision-making under (severe) uncertainty. The idea here is that decision-makers aim to identify not only a decision or set of decisions for a situation considered, but to seek out those decisions that are *robust*, namely effective vis-a-vis the uncertainty surrounding key components of the situation in question.

On the face of it, the goal of achieving *robustness* does not require a great deal of justification because the concept seems to speak for itself. That is, the merit of striving for *robustness* seems to be implied by the presumably self-evident meaning of the concept itself. But the truth is that for all its seemingly self-explanatory purport, the property of ROBUSTNESS can be defined in a variety of ways. And what is more, as a measure of "effectiveness against uncertainty" it can designate different types, or levels, or degree of *robustness*: for instance *global* as opposed to *local* robustness.

I expound these issues in the context of a problem that I call *The Treasure Hunt Problem* which runs as follows:

> **The Treasure Hunt Problem**
> Rumors are rife among the inhabitants of a big island that an ancient treasure is hidden somewhere on this island (see Figure 2.1[1]). There is much speculation about the treasure's worth and its contents, but there is general agreement that it is
>
> E N O R M O U S!
>
> So far so good.
>
> The truly serious disagreements among the inhabitants arise about the *location* of the treasure. More specifically, the inhabitants fall into two groups of approximately the same size, known as

[1]The island featured in the Treasure Hunt Problem is fictitious. I use the map of Australia to illustrate it as a (small) contribution to the ongoing campaign to promote tourism to this beautiful country/continent/island.

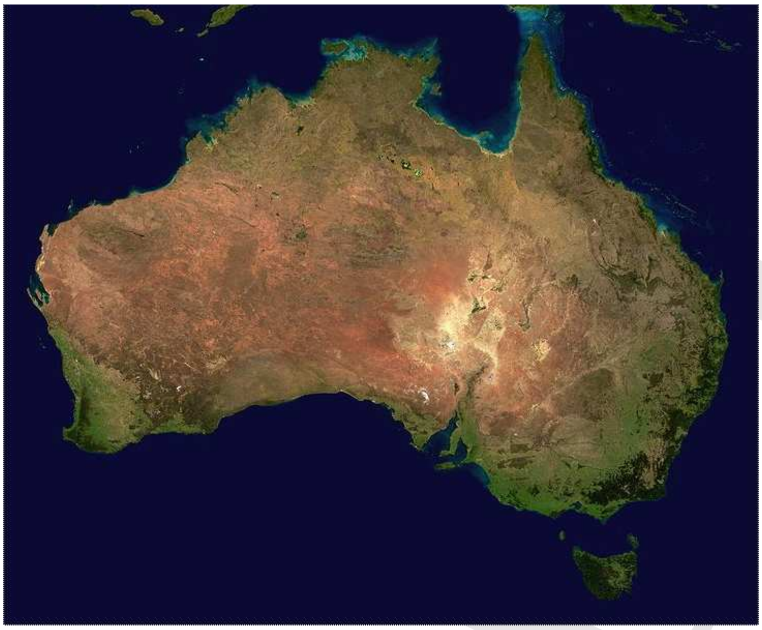


Figure 2.1: A Treasure island

Group Local and *Group Global*.

- Group **Global**:
  Members of this group hold very strongly that the uncertainty regarding the location of the treasure is *severe.* That is, these members hold that the treasure's location is shrouded in complete mystery.
- Group **Local**:
  Members of this group hold very strongly that the treasure is located in the proximity of a *Lake* (the whitish area in the center of the island). They cannot vouch for this position as they cannot be sure about it. Still, they persist in upholding it!

Five types of technologies for locating and extracting the treasure are available to the inhabitants (for a fee). These technologies are known on the island as TA, TB, TC, TD, and TE. Their costs are also known.

Now, a frequent visitor to their shores, the internationally renowned maven on robust decision-making, Fred, was approached by each group to advise them on what technology to chose for this task. Fred's advice to both groups was that considering the uncertainty in the true location of the treasure, their task was to establish the *robustness* of each technology, namely to determine how each technology fares in the face of this uncertainty.

Furthermore, for each group Fred had more specific advice. To Group Global he indicated that, in view of the fact that they take the uncertainty to be severe, each technology had to be put to a **global robustness analysis**. That is, the analysis had to evaluate the performance of each technology over the **entire island.**

To Group Local he indicated that, given their basic premiss, the uncertainty in their case is relatively *mild.* Hence, each technology had to undergo a **local robustness analysis**. That is, the analysis had to evaluate the performance of each technology in the locale of the *Lake* and its **immediate vicinity.**

Two weeks later Fred submitted a surprisingly terse report. Its main content consisted of a table, a copy of which is shown in Table 2.1, displaying the ranking of the search/excavation technologies according to their local[2] and global robustness.

Global Robustness	Local Robustness
TC	TE
TA	TB
TB	TC
TD	TD
TE	TA

Table 2.1: Ranking of the technologies according to their local and global robustness

Post script:

The inhabitants of the island were taken aback by the invoice they received from Fred for his consulting job.

And if this were not enough, the island's WIT was overheard remarking in the local bar that evening that had he been consulted on this matter, he would have given the same advice for free!

There is more to this story. So, stay tuned.

It is important to appreciate that in view of the fundamentally different positions taken by the two groups with regard to the uncertainty in the true location of the treasure, each group in effect sets itself the task of solving a different version of the *Treasure Hunt Problem*. This means that we are dealing in fact with two different *Treasure Hunt Problems*. Before I proceed to explain the differences between the two problems, let us first consider what they have in common.

In both cases:

- The objective is to rank alternative systems (search/excavation technologies) according to their robustness to uncertainty.
- The *space* containing all the possible/plausible values of the *uncertain parameter* (location of the treasure) is known (the island).
- The robustness of an alternative is assessed on the basis of its ability to perform the task (finding and excavating the treasure) given the uncertainty in the location of the treasure.
- The term *robustness* is formally undefined.

One of our main concerns in this chapter is the formulation of a suitable definition of the property ROBUSTNESS. However, prior to this, I want to point out the differences between the two versions of the *Treasure Hunt Problems*, listed in Table 2.2.

Having pointed out the similarities and the differences between the two versions of the problem, the reader should now have a clearer picture of the tasks set by each version of the problem.

And yet, to be able to explain how one would go about solving them, the problems require a more careful formulation, which means of course that key concepts in the problems would have to be carefully defined.

The concepts that I have in mind are these:

[2] Around the *Lake*.

Local	Global
· The uncertainty is very mild. · An estimate of the true location is taken to be available. · There is considerably strong confidence in the estimate.	· The uncertainty is severe. · An estimate of the true location is not considered. Indeed, it is not even considered to be available.

Table 2.2: Differences between the two versions of the *Treasure Hunt Problem*

- Severe uncertainty
- Very mild uncertainty
- Estimate
- Confidence in the estimate

As we shall see, however, giving these concepts a precise definition can prove a considerable challenge indeed, in certain respects it is a case of "easier said than done".

2.2 Mathematical abstraction

I want to preface my formal formulations of the *Treasure Hunt Problem* with a passage from *Idle Thoughts of an Idle Man* (Jerome 1889), which makes vivid the great merit of mathematical abstraction (modeling) in putting across ideas precisely and unambiguously. I realize that certain readers may perhaps find it "politically incorrect", still its ability to make this point is worth the risk involved.

> Questions of taste were soon decided in those days. When a twelfth-century youth fell in love, he did not take three paces backward, gaze into her eyes and tell her she was beautiful to live. He said he would step outside and see about it. And if, when he got out, he met a man and broke his head — the other's man's head, I mean — then that proved that his — the first fellow's girl — was a pretty girl. But if the other fellow's — the other fellow to the second fellow, that is because of course the other fellow would only be the other fellow to him, not the first fellow, who — well, if he broke his head, then his girl — not the other fellow's, but the fellow who was the — Look here, if A broke B's head, then A's girl was a pretty girl, but if B broke A's head, then A's girl wasn't pretty girl, but B's girl was. That was their method of conducting art criticism.
>
> Now-a-days we light a pipe, and let the girls fight it out amongst themselves . . .
>
> Jerome K. Jerome
> *Idle Thoughts of an Idle Fellow*
> *Being Idle,* pp. 58-59, 1889.

So let us look at the A's and B's of the *Treasure Hunt Problem*.

We have a collection of alternative *systems*, $q \in Q$. The *performance* of each system is based on the value of a certain *parameter* $u \in \mathscr{U}$. So, for each system, we partition the space $\mathscr{U}$ into two *disjoint* sets, call them $A(q)$ and $\overline{A}(q)$, such that $A(q)$ consists of the values of u for which the performance of system q is "acceptable" and $\overline{A}(q)$ consists of the values of u for which the performance of system q is "unacceptable". Formally, $\overline{A}(q)$ is the *complement* of $A(q)$ with respect to $\mathscr{U}$, that is $\overline{A}(q) = \mathscr{U} \setminus A(q)$.

The *uncertainty* is depicted in the model by the following terms:

u = *parameter of interest* whose true value is unknown as it is subject to uncertainty.

$\mathscr{U}$ = set of all "possible/plausible" values of the true value of u. This is the *uncertainty space* of u.

The *severity* of the uncertainty is expressed by virtue of the following working assumption.

Working Assumption:
The uncertainty in the true value of u is LIKELIHOOD-FREE.

What is meant by this assumption is that no likelihood whatsoever is attributed to the uncertainty space $\mathscr{U}$. The uncertainty space $\mathscr{U}$ is thus taken to be devoid of any explicit and/or implicit likelihood structure. This implies in turn that there is no ground to assume that the true value of u is more/less likely to be in the neighborhood of any given $u \in \mathscr{U}$.

I want to make it clear, though, that by postulating this working assumption one does not imply that the concept of uncertainty categorically prohibits all attribution of likelihood to the uncertainty space $\mathscr{U}$. Of course not. There is no ban on imposing likelihood structures on the uncertainty space $\mathscr{U}$, based say, on subjective experience or even for mathematical convenience.

The point is, however, that should a likelihood structure be ascribed to the uncertainty space $\mathscr{U}$, this fact must be made explicit in the formulation, so that it is crystal clear that such a property is postulated by the model in question.

So much then for the *Global* version of the *Treasure Hunt Problem*.

As we saw above, the factor giving rise to the *Local* version of the problem is that members of the so called *Group-Local,* from the start, proceed on the assumption that an *estimate* of the true value of u is available. Or more accurately, this basic approach to the *Treasure Hunt Problem* is what earns this group its appellation. This means that to be able to discuss the *Local* version of the problem formally, not only do we have to designate symbolically the *estimate* of the true value of u, we also need to designate symbolically areas in the vicinity, namely *neighborhoods,* of the estimate.

So let,

$\tilde{u}$ = *point estimate* of the true value of u.

$\mathcal{N}(\alpha, \tilde{u})$ = *neighborhood* of radius α around $\tilde{u}$.

It is important to be clear on the exact meaning of "point" in "point estimate". All that is meant by it is that the estimate of u is an element of the same space, namely $\mathscr{U}$, as the true value of u. Thus, if the true value of u is a matrix, then all the elements of $\mathscr{U}$ are matrices and so is the estimate $\tilde{u}$. And if the true value of u is a function, then all the elements of $\mathscr{U}$ are functions, and so is the estimate $\tilde{u}$. And if the true value of u is a 4-story building in Melbourne, then so are all the elements of $\mathscr{U}$, and so is the estimate $\tilde{u}$.

As for the role played by the construct NEIGHBORHOOD in this setting.

In cases where the basic approach to the problem — as in the case of *Group Local*'s approach to the *Treasure Hunt Problem* — is that a "reasonable" estimate of u is available, the robustness analysis is conducted in the neighborhood of this estimate rather than over the entire uncertainty space $\mathscr{U}$. This, needless to say, is a *local* robustness analysis. In this framework one needs to have a mechanism for marking out neighborhoods of various sizes around the estimate $\tilde{u}$. We shall refer to these as *neighborhoods* of the point estimate $\tilde{u}$.

For our purposes here, there is no need to go into the technical minutia that are associated with a formal definition of the mathematical concept NEIGHBORHOOD. We can simply proceed to define this concept as follows:

> $\mathcal{N}(\alpha,\tilde{u})$ is the subset of the uncertainty space $\mathscr{U}$ consisting of all the points in $\mathscr{U}$ that are within a DISTANCE α (or less) from $\tilde{u}$.

That is,

$$\mathcal{N}(\alpha,\tilde{u}) := \{u \in \mathscr{U} : dist(u,\tilde{u}) \leq \alpha)\}\ ,\ \alpha \geq 0 \tag{2.1}$$

where $dist(u,\tilde{u})$ denotes the distance between u and $\tilde{u}$.

Neither is it necessary to dwell on the properties that $dist$ is required to satisfy. For our purposes it is sufficient to assume that the neighborhoods considered in this discussion must have the following two simple properties:

$$\mathcal{N}(0,\tilde{u}) = \{\tilde{u}\} \text{ (contraction)} \tag{2.2}$$

$$\mathcal{N}(\alpha,\tilde{u}) \subseteq \mathcal{N}(\alpha',\tilde{u})\ ,\ \forall\, \alpha \in [0,\alpha') \text{ (nesting)} \tag{2.3}$$

For instance, the neighborhood structure specified by

$$\mathcal{N}(\alpha,\tilde{u}) = \left\{u \in \mathscr{U} : (u_1-\tilde{u}_1)^2 + (u_2-\tilde{u}_2)^2 \leq \alpha^2\right\}\ ,\ \alpha \geq 0 \tag{2.4}$$

consists of concentric circles centered at $\tilde{u}$, observing that here $\mathcal{N}(\alpha,\tilde{u})$ is a circle of radius α centered at $\tilde{u}$. In contrast, the neighborhood structure specified by

$$\mathcal{N}(\alpha,\tilde{u}) = \{u \in \mathscr{U} : |u_1-\tilde{u}_1| \leq \alpha, |u_2-\tilde{u}_2| \leq \alpha\}\ ,\ \alpha \geq 0 \tag{2.5}$$

consists of squares centered at $\tilde{u}$, observing that here $\mathcal{N}(\alpha,\tilde{u})$ is a square of side 2α centered at $\tilde{u}$.

These two most prevalent neighborhood types are displayed in Figure 2.2.

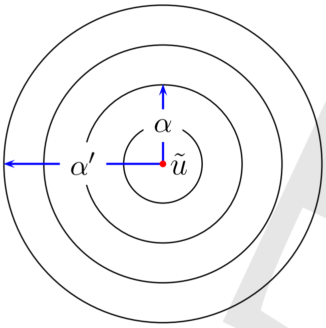
Post Script:

Fred (the renowned expert on robust decision in the face of severe uncertainty) reminds me to advise the readers that in some disciplines the neighborhood $\mathcal{N}(\alpha,\tilde{u})$ is called a "ball".

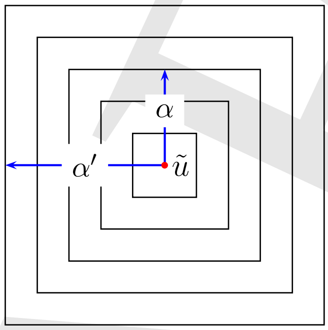
Indeed, in this book the two terms are taken to be equivalent. The same is true about the notations $\mathcal{N}(\alpha,\tilde{u})$ and $B(\alpha,\tilde{u})$ designating "neighborhood" and "ball" respectively. Observe that these "balls" need not necessarily be circular (see for instance Figure 2.2).

In certain cases it is more convenient to refer to the *radius* of a neighborhood (α) as "size", whereas in other it is more convenient to denote the size/radius by ρ rather than by α.

Having clarified the basic ingredients of the *Treasure Hunt Problem*, let us proceed to give its two versions a more formal treatment.



$$\mathcal{N}(\alpha,\tilde{u}) = \{u \in \mathscr{U} : (u_1 - \tilde{u}_1)^2 + (u_2 - \tilde{u}_2)^2 \leq \alpha^2\} \ , \ \alpha \geq 0$$



$$\mathcal{N}(\alpha,\tilde{u}) = \{u \in \mathscr{U} : |u_1 - \tilde{u}_1| \leq \alpha, |u_2 - \tilde{u}_2| \leq \alpha\} \ , \ \alpha \geq 0$$

Figure 2.2: Two common neighborhood structures on $\mathscr{U} = \mathbb{R}^2$

2.3 Treasure Hunting subject to SEVERE uncertainty

Consider the version of the problem faced by *Group Global*. Recall that members of this group take themselves to be in the dark about the (true) location of the treasure on the island.

In this case the severity of the uncertainty is expressed by dint of the objects of the uncertainty model being defined as follows:

$u =$ possible location of the treasure on the island.

$\mathscr{U} =$ set of all possible locations on the island.

$Q =$ set of alternative search/excavation technologies available.

$A(q) =$ set of locations on the island that are amenable to technology q.

$\overline{A}(q) =$ set of locations on the island that are not amenable to technology q.

The task is to rank the technologies according to their robustness against the uncertainty in the true location of the treasure. To accomplish this task, *Group Global* must first determine the robustness of each technology to severe uncertainty (which is no easy matter) and then order the technologies according to their respective measures of robustness (which is elementary). So the robustness problem in this case can be stated as follows:

Global robustness problem:
Given $\mathscr{U}$ and $A(q), q \in Q$, determine the robustness of each system $q \in Q$ against the severe uncertainty in the true value of $u \in \mathscr{U}$.

Take note that this formulation does not posit an ESTIMATE of the true value of u. Indeed, there is no assumption that an estimate is available. All that is posited is that the true value of u is an element of $\mathscr{U}$.

2.4 Treasure Hunting subject to A VERY MILD uncertainty

This version of the problem is considered by *Group Local* whose members strongly believe that the true location of the treasure is somewhere in the neighborhood of the *Lake*.

The uncertainty model consists of the same components that make up the preceding version, except for the additional, centrally important element:

$\tilde{u} =$ *point estimate* of the true value of u.

You will recall that:

> *Group Local* have considerable confidence in $\tilde{u}$, meaning that they strongly hold that the true location is somewhere in the vicinity of $\tilde{u}$.

This means of course that the estimate is viewed by *Group Local* as being "reasonably good".

Having consulted Fred, *Group Local* concluded that the way to go was to assume that the true location of the treasure was in the neighborhood $\mathcal{N}(\alpha^*, \tilde{u})$. The value of α^* was to be worked out according to a formula provided by Fred, by using values of α proposed by senior members of the Group.

The value of α^* was computed and is now kept in a specially built safe, at a secret location on the island near the main entrance to the *Lake*. Regrettably, therefore, I cannot provide the value of α^* here. Rumor has it that the secret value of α^* is 2987, but this is just a rumor.

Be that as it may, the robustness problem in this case can be stated as follows:

> **Local robustness problem:**
> Given $\mathscr{U}$, $A(q), q \in Q$ and $\tilde{u} \in \mathscr{U}$, determine the robustness of each system $q \in Q$ against the mild uncertainty in the true value of u, assuming that the true location of the treasure is in $\mathcal{N}(\alpha^*, \tilde{u})$, where α^* is the secret value of α, which is a closely guarded secret on the island.

Figure 2.3 illustrates the situation.

> **Post Script:**
> According to local rumors, it was no easy matter for members of *Group Local* to agree on a proper value of α^*. Some members felt that the value determined by Fred's formula was way too high, while others felt that it was way too low.
>
> Fred indicated that there was no cause for alarm.
>
> He pointed out that a proper value for the size of the neighborhood can be worked out by means of a SENSITIVITY ANALYSIS. Namely, by varying the value of α and observing how the ranking of the technologies changes with α.
>
> He also indicated that he will give the Group a special discount on such an analysis, should the Group decide to go ahead with it.

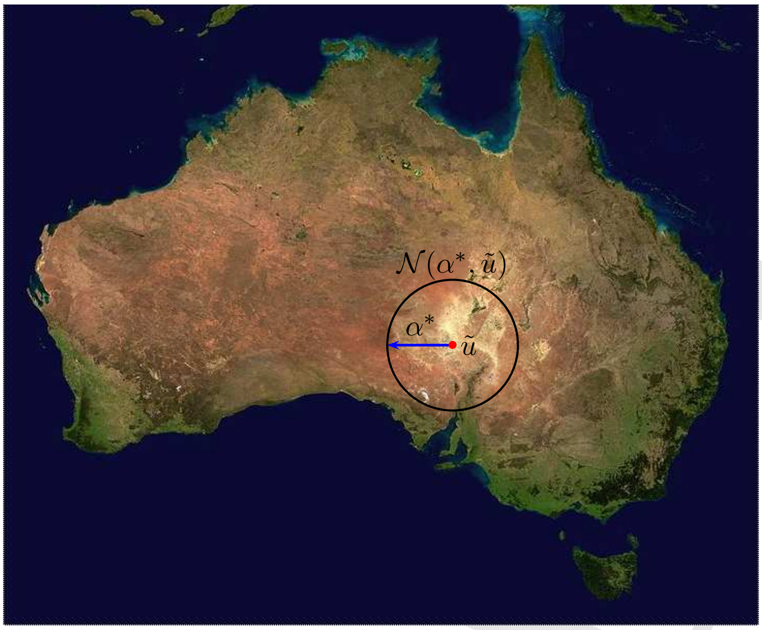


Figure 2.3: Local robustness analysis at $\tilde{u}$, $\alpha^* =$ top secret value

2.5 Where is the beef ?

The reader has no doubt noticed that, thus far, my discussion of this book's main interest: ROBUSTNESS AGAINST SEVERE SEVERE UNCERTAINTY, had not addressed directly the real issue on our agenda:

> How exactly would we define ROBUSTNESS TO SEVERE UNCERTAINTY, and how would we actually gauge it?

In the preceding sections of this chapter I emphasized the crucial role that robustness has in *decision under uncertainty*. I illustrated this point through the advice given by Fred, the renowned expert on robustness, to the island's inhabitants, to rank the alternative technologies available to them according to their respective measures of robustness. I also noted that Fred used a formula to determine the robustness of an alternative. However, I did not indicate how this formula worked nor indeed what this formula was.

I shall address the above question at various stages of the discussion in an ongoing attempt to elucidate it as best I can. To set the scene, though, I want to call attention to the following:

> **Fact:**
> It is important to appreciate that there is no one general, universally applicable definition of *robustness* that, as it were, "fits all" problems. Nor is there such a formula yielding a universally applicable measure of robustness against severe uncertainty. There are a number of approaches outlining how to deal with this difficult problem, so in subsequent chapters I shall examine how each approach proposes to take on this task.

The point here is that achieving *robustness to severe uncertainty* is an enormously complex, hence demanding, task. One need not be an expert in *decision theory* to appreciate the enormity of this task. After all, the driving force behind this undertaking is to identify decisions that are sound/effective/resilient against unforeseen contingencies. The implication therefore is that it is hard to envisage a truly useful, general-purpose definition or formula that would be applicable generally to any situation of severe uncertainty.

As we shall progressively see then, definitions of robustness and so-called *recipes* for measuring robustness, are very much *problem-oriented* meaning that these should be expected to be worked out in relation to PROBLEM TYPES.

With this as background, let us now examine what definitions or criteria for determining robustness can be contemplated in the case of the *Treasure Hunt Problem*.

2.5.1 Size criterion

The criterion that I am about to outline offers what seems to be an "intuitively obvious" means for assessing the *global robustness* of systems to variability in general and to severe uncertainty in particular. As we shall see, however, although eminently compelling, this criterion is not very practical because its implementation is fraught with difficulties.

Let us see then how this criterion would work in the context of our *Treasure Hunt Problem* when the problem is taken to be subject to severe uncertainty.

Recall that $\mathscr{U}$ is the set of all the possible/plausible locations on the island and $A(q)$ is the subset of $\mathscr{U}$ consisting of the "acceptable" elements of $\mathscr{U}$, that is, elements that are amenable to technology q. Then clearly, the ratio

$$\gamma(q) := \frac{size(A(q))}{size(\mathscr{U})} \cdot 100 \ , \ q \in Q \tag{2.6}$$

provides the most "logical" means to measure the robustness of technology q, where $size(A)$ denotes the "size" of set $A \subseteq \mathscr{U}$.

For example, if $\mathscr{U}$ is a finite set, then we can let $size(A) := |A|$, where $|A|$ denotes the cardinality[3] of set A. Thus, if $\gamma(q) = 100$, then q is super-robust: all the locations on the island are amenable to technology q. In contrast, if $\gamma(q) = 0$, then q is super-fragile: none of the locations on the island is amenable to technology q. Similarly, if $\gamma(q) = 78$, then 78% of the locations on the island are amenable to technology q.

I shall not go into the question of what formal definitions of $size$ would be better-suited for more complicated situations except to indicate the following: as pointed out in the bibliographic notes at the end of this chapter, despite its appeal, this criterion is rarely used in practice.

That said, I am going to adopt the following convention:

Moshe's working convention:
In this book, the *Size Criterion* is viewed as the DEFAULT criterion for assessing (global) robustness associated with *constraints* (requirements). That is, unless stated otherwise, (global) robustness with respect to constraints is measured according to (2.6), or according to its *local* variant

$$\lambda(q,\alpha) := \frac{size(A(q) \cap \mathcal{N}(\alpha,\tilde{u}))}{size(\mathcal{N}(\alpha,\tilde{u}))} \cdot 100 \ , \ q \in Q \tag{2.7}$$

[3] The *cardinality* of a finite set is the integer specifying the number of elements comprising the set.

where

$$A(q,\alpha) := A(q) \cap \mathcal{N}(\alpha,\tilde{u}) \ \ , \ q \in Q, \alpha \geq 0 \tag{2.8}$$

Note that, by definition, $A(q,\alpha)$ is the set of all the values of u in $\mathcal{N}(\alpha,\tilde{u})$ that are acceptable with respect to system q. Hence, by definition, $\lambda(q,\alpha)$ is the "percentage"[4] of elements of $\mathcal{N}(\alpha,\tilde{u})$ that are acceptable with respect to system q.

2.5.2 Example

An extensive investigation discovered that there are exactly 108,993,567 possible location on the island, so set $size(\mathscr{U}) = |\mathscr{U}| = 108,993,567$. It was also found that, out of these, 93,734,468 locations are amenable to the given technology under consideration (top secret), call it q^*. Hence, set $size(A(q^*)) = 93,734,468$. Therefore,

$$\gamma(q^*) = \frac{size(A(q^*))}{size(\mathscr{U})} \cdot 100 = \frac{93734468}{108993567} \cdot 100 = 86\% \tag{2.9}$$

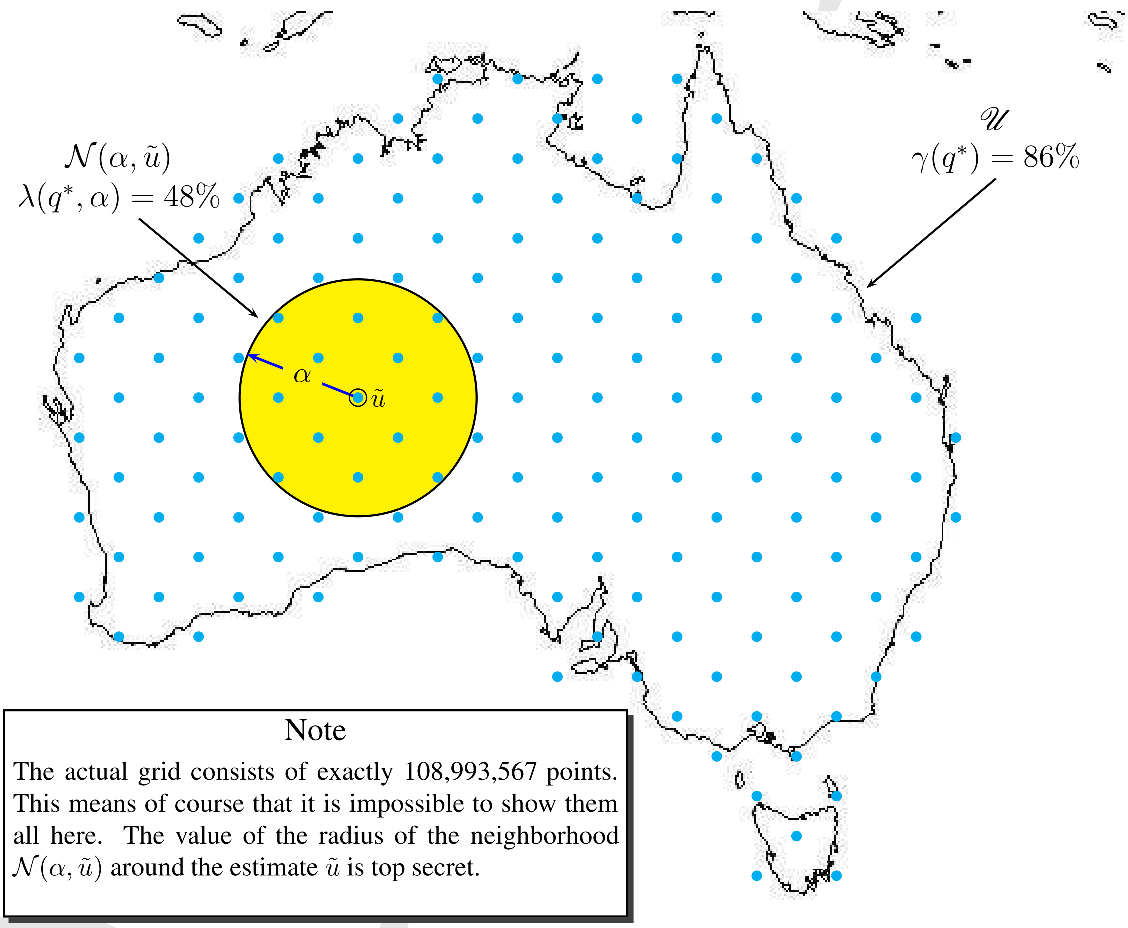


Figure 2.4: Illustration of the Size Criterion in action

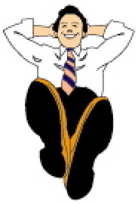
To assess local robustness, it was found that for the value of α^* under consideration (top secret), the

[4] To be precise, this is so if $size(A)$ is equal to the cardinality of A.

neighborhood $\mathcal{N}(\alpha^*, \tilde{u})$ contains 20,777,957 locations, thus we set $size(\mathcal{N}(\alpha^*, \tilde{u})) = |\mathcal{N}(\alpha^*, \tilde{u})| = 20,777,957$. Out of these, 9,973,419 were found to be amenable to technology q, hence set $size(A(q^*, \alpha^*)) = |A(q^*, \alpha^*)| = 9,973,419$. Therefore,

$$\lambda(q^*, \alpha^*) = \frac{size(A(q^*, \alpha^*))}{size(\mathcal{N}(\alpha^*, \tilde{u}))} \cdot 100 = \frac{9973419}{20777957} \cdot 100 = 48\% \tag{2.10}$$

The conclusion therefore is that, according to the *Size Criterion*, technology q^* is far more robust globally on $\mathscr{U}$ than locally on $\mathcal{N}(\alpha^*, \tilde{u})$.



Post Script:

Some members of *Group Local* expressed strong reservations about the *Size Criterion*. They contended that when computing local robustness over $\mathcal{N}(\alpha^*, \tilde{u})$, more "weight" ought to be given to locations that are closer to the estimate $\tilde{u}$ — than to those that are farther from $\tilde{u}$.

Fred, the renowned robustness expert, advised the Group that incorporating such "weights" in the definition of the local robustness is a straightforward affair.

He also indicated that, for a relatively small fee, he would gladly recompute the local robustness of the technologies considered to take these "weights" into account.

A sub-committee was thus established to determine the "weights" in question. However, if the rumors are anything to go by, members of this sub-committee found it very hard to agree on appropriate weights.

2.6 About the top secret value of α

In this section I want to explain my characterization of the value of α^* as a "closely guarded secret". My objective in using this metaphor was to alert the reader to the intricacies involved in selecting an appropriate value of α in the context of the local robustness analysis discussed above.

As we saw above, this type of analysis requires that the user (somehow!) determine (by whatever means!) the radius (α) of the neighborhood around the estimate, that is $\mathcal{N}(\alpha, \tilde{u})$, on which the performance of the alternative systems is assessed. My point is then that readers should not get the impression that because this type of analysis simply proceed to require this move from the user, that executing it is necessarily a simple, straightforward affair.

Theoretically, the value of α selected for this purpose should reflect the user's confidence in the estimate $\tilde{u}$: the greater one's confidence in $\tilde{u}$, the smaller would α be. Indeed, in the extreme case where the user is absolutely certain that $\tilde{u}$ is equal to the true value of u, she can set the value of α to zero, so that the robustness analysis would be conducted on a singleton $\mathcal{N}(0, \tilde{u}) = \{\tilde{u}\}$.

If the uncertainty is *mild* (whatever this means!), the choice of a proper value for α would involve a compromise between two countervailing considerations. A low value of α may exclude from the robustness analysis values of u that can be equal to the true value of u, that presumably would be distant from the estimate $\tilde{u}$. The analysis may thus risk being too heavily biased towards those values of u that are close to the estimate $\tilde{u}$. On the other hand, a high value of α may result in the neighborhood $\mathcal{N}(\alpha, \tilde{u})$ comprising values of u that are significantly different from the true value of u, to thus "distort" the results of the analysis.

The truth of the matter is that, as Fred, the renowned expert on robustness analysis, would tell you, in practice, the value of α would be determined on a case by case basis, rather than by a general purpose recipe.

My own view on this matter is that — should it be decided to adopt the local analysis outlined above — the ranking of alternative systems should be based on a range of values of α, because it would be well-nigh impossible to “pin down" this highly elusive single “top secret" value of α.

Having said all that, I want to examine very briefly — for the benefit of *info-gap aficionados* — the simple recipe for determining the value of α that Fred proposed in a seminar he gave during one of his frequent visits to the island.

2.7 An info-gap decision theory perspective

Suppose that the user specifies a (subjective) critical value of $\lambda(q, \alpha)$, call it λ^*. This value represents the minimum acceptable level (%) of acceptable values of u in $\mathcal{N}(\alpha, \tilde{u})$. For instance, if we set $\lambda^* = 80$, then at least 80% of the elements of $\mathcal{N}(\alpha, \tilde{u})$ should be “acceptable” (belong in $A(q)$).

Once the desired value of λ^* had been specified, the value of α selected for the local robustness analysis should satisfy the requirement

$$\lambda(q, \alpha) \geq \lambda^* \tag{2.11}$$

keeping in mind, though, that more than one value of α may satisfy this requirement.

Consider then the following recipe for determining the value of α to be used in the local robustness analysis outlined above:

$$\begin{aligned}
\alpha^*(q, \tilde{u}) :&= \max_{\alpha \geq 0} \{\alpha : \lambda(q, \alpha) \geq \lambda^*\} \ , \ q \in Q && (2.12)\\
&= \max_{\alpha \geq 0} \left\{\alpha : \frac{size(A(q,\alpha))}{size(\mathcal{N}(\alpha, \tilde{u}))} \cdot 100 \geq \lambda^*\right\} && (2.13)\\
&= \max_{\alpha \geq 0} \ \{\alpha : size(A(q,\alpha)) \geq 0.01\lambda^* size(\mathcal{N}(\alpha, \tilde{u}))\} && (2.14)
\end{aligned}$$

In words,

> $\alpha^*(q, \tilde{u})$ is the largest value of α such that the size of the subset of $\mathcal{N}(\alpha, \tilde{u})$ consisting of acceptable values of u with respect to system q is not smaller than $\lambda^*(\%)$ of the size of $\mathcal{N}(\alpha, \tilde{u})$.

According to this recipe, the larger the value of $\alpha^*(q, \tilde{u})$, the more robust q is (in the neighborhood of $\tilde{u}$). Hence, $\alpha^*(q, \tilde{u})$ can be viewed as a measure of the *local* robustness of q at $\tilde{u}$.

What then is the connection to *info-gap decision theory*?

If you are very *conservative* (expecting the worst), you’ll no doubt set $\lambda^* = 100(\%)$, in which case the recipe yields:

$$\begin{aligned}
\alpha^*(q, \tilde{u}) :&= \max_{\alpha \geq 0} \ \{\alpha : size(A(q,\alpha)) \geq 0.01\lambda^* size(\mathcal{N}(\alpha, \tilde{u}))\} && (2.15)\\
&= \max_{\alpha \geq 0} \ \{\alpha : size(A(q,\alpha)) = size(\mathcal{N}(\alpha, \tilde{u}))\} && (2.16)\\
&= \max_{\alpha \geq 0} \ \{\alpha : A(q,\alpha) = \mathcal{N}(\alpha, \tilde{u})\} && (2.17)\\
&= \max_{\alpha \geq 0} \ \{\alpha : A(q) \cap \mathcal{N}(\alpha, \tilde{u}) = \mathcal{N}(\alpha, \tilde{u})\} && (2.18)
\end{aligned}$$

recalling that $A(q, \alpha)$ is a subset of $\mathcal{N}(\alpha, \tilde{u})$.

In words,

> In the extreme case where $\lambda^* = 100(\%)$, it follows that $\alpha^*(q, \tilde{u})$ is the largest value of α such that ALL the elements of the neighborhood $\mathcal{N}(\alpha, \tilde{u})$ are *acceptable* with respect to system q.

Conceptually, to determine the value of $\alpha^*(q, \tilde{u})$ in this case, one would start with $\alpha = 0$, and keep increasing the value of α until the neighborhood $\mathcal{N}(\alpha, \tilde{u})$ would encounter a value of u that is unacceptable with respect to system q. This is illustrated in Figure 2.5.

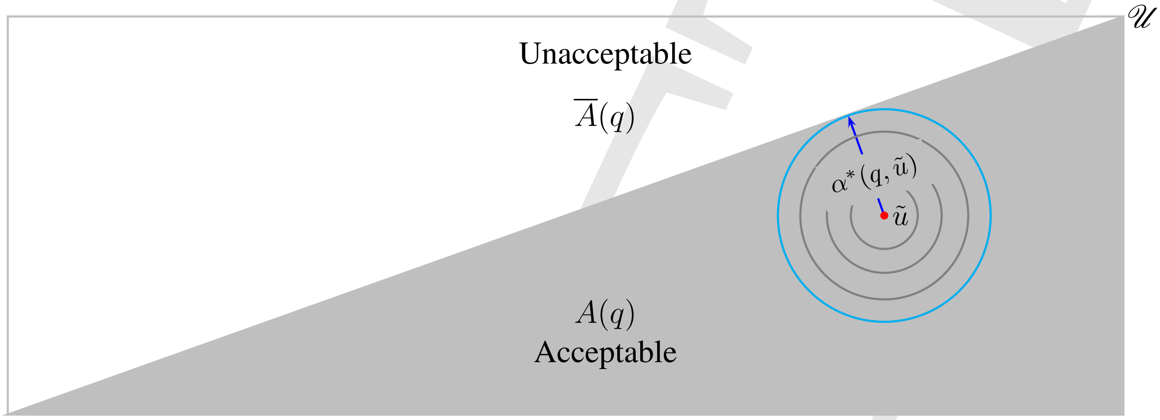


Figure 2.5: Illustration of a conservative local robustness analysis of system q

As illustrated by this picture, in cases where $\lambda^* = 100(\%)$, the value of $\alpha^*(q, \tilde{u})$ denotes the shortest distance from the estimate $\tilde{u}$ to the *boundary of the set of acceptable values of u with respect to q*. Differently put, it is the shortest distance from the estimate to the region of unacceptable values of u with respect to system q.

In this case $\alpha^*(q, \tilde{u})$ is a perfectly reasonable measure of the *local* robustness of q to the *mild* uncertainty in the true value of u. However, it cannot be expected to be capable of measuring properly the global robustness of the systems against the variation of the value of u over the uncertainty space $\mathscr{U}$.

Post Script:

This is a message from Fred, the renowned robustness expert, to info-gap aficionados.

Fred indicates that the recipe given in (2.18) is precisely the recipe stipulated by *info-gap decision theory* to determine the robustness of system q.

He further points out that, in view of this, *info-gap's robustness model* is unsuitable for the treatment of **severe** uncertainty.

In my opinion, Fred makes an excellent point here. I shall discuss this issue in greater detail in subsequent chapters. At this stage I want to point out that it follows from (2.18) that for $\lambda^* = 100(\%)$ we have

$$\alpha^*(q, \tilde{u}) := \max_{\alpha \geq 0} \{\alpha : A(q) \cap \mathcal{N}(\alpha, \tilde{u}) = \mathcal{N}(\alpha, \tilde{u})\} \tag{2.19}$$

$$= \max \{\alpha \geq 0 : u \in A(q), \forall u \in \mathcal{N}(\alpha, \tilde{u})\} \tag{2.20}$$

In words,

> If $\lambda^* = 100(\%)$, then according to (2.18), the (local) robustness of decision q at $\tilde{u}$ is the size (α) of the largest neighborhood $\mathcal{N}(\alpha, \tilde{u})$ around $\tilde{u}$ all of whose elements are acceptable with respect to decision q.

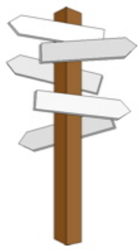
This is precisely the recipe prescribed by *info-gap decision theory* to determine the robustness of system q at $\tilde{u}$.

2.8 What next?

I hope that the relatively (technically) “soft" analysis in this chapter has made it clear that a meaningful explanation of the function of *robustness* as a crucial factor in *sound* decision under severe uncertainty requires a careful, indeed formal, definition of the notion of *robustness to uncertainty*. The point I wanted to highlight was that one cannot count on the seemingly familiar meaning of the concept *robustness* to do the work if this concept is to be used as a criterion for measuring the effectiveness of decisions against uncertainty. For one thing, a careful formulation of this concept is required to indicate clearly whether it is being applied as a measure of *local robustness* or that of *global robustness*.

In Chapter 4 I take up the concept SEVERE UNCERTAINTY, and in the ensuing chapters I examine formally models for robust decision-making under severe uncertainty.

However, prior to this, in Chapter 3 I take a close look at the concept RADIUS OF STABILITY and the local robustness models that it induces. Readers who are familiar with this important concept can proceed directly to Chapter 4 on SEVERE UNCERTAINTY. Still, I suggest that reading the next chapter can benefit even those readers who are well-versed in this topic. Not only because they stand to benefit from “my take" on this concept, but also because my discussion of this issue will enable them to better appreciate my main objectives in this book.



To explain.

At first glance, the idea of prefacing a discussion on *severe uncertainty* with an examination of *Radius of Stability* models — which as we shall see are utterly unsuitable for the treatment of severe uncertainty — does not seem to make much sense. The point, of course, is that not only does it appear odd to anticipate the treatment of *severe uncertainty* with a discussion of models that are the *wrong* models for the treatment of severe uncertainty. The real question is: why bother with such models at all if both conceptually and technically they lack the capabilities required for a sound treatment of *severe uncertainty*?

My reply to this perfectly valid question should give you some inkling into the motivation behind this book. My experience of the last seven years has shown that a detailed/careful clarification of the nature of models of *local robustness* is not only relevant to a discussion on *severe uncertainty*, it is in fact imperative. I have come to realize that such a clarification, particularly from the viewpoint of how these models stand with regard to a sound treatment of *severe uncertainty,* is badly needed.

Because, as my experience has shown, many analysts, practitioners, and even mathematicians, dealing with the problem of *robustness to uncertainty*, are “blissfully unaware" of the foundational difference between *local robustness* and *global robustness*. This means of course that they do not appreciate the vastly different ramifications that models of *local robustness* have for the management of *severe uncertainty* as opposed to those of *global robustness*.

I submit therefore that a detailed analysis of models of *local robustness* induced by the concept RADIUS OF STABILITY as a prelude to an analysis of the concept of SEVERE UNCERTAINTY should not only bring into sharp focus the mode of operation, scope, and capabilities of such models. It should also bring out more forcefully why models of *local robustness* must not even be contemplated for the treatment of *severe uncertainty*.

These remarks should also shed light on the book's title, namely on what I mean by FOOLED BY ROBUSTNESS.

One of the main objectives of this book is to warn that a lack of awareness of, or a lack of appreciation of, or inattention to the fundamental difference between *local* and *global* robustness, results in the FOOLED BY ROBUSTNESS CONDITION. What I mean by this is that a lack of appreciation that models of *local robustness* in principle lack the capabilities to conduct a robustness analysis of the kind required by the very nature of *severe uncertainty* leads to a misinterpretation of the results yielded by this analysis. This means that the results obtained from the *local robustness analysis* are misread as having global significance. In other words, the results obtained from a robustness analysis conducted by a model of *local robustness* are seen to be *robust to severe uncertainty*.

This is one of the lessons learned from what I call the *Info-gap Experience*. I discuss this issue in the second part of the book. For now, I need only point out that in the case of *info-gap scholars,* analysts, and practitioners, the FOOLED BY ROBUSTNESS CONDITION is manifested in their failure to realize that the results that they report on in their publications **are anything but** robust to severe uncertainty. This is so because they are unaware that *info-gap's robustness model* is in fact a typical *Radius of Stability model,* hence the wrong model for the analysis and solution of decision-problems subject to severe uncertainty.

Note that in *info-gap decision theory* the severity of the uncertainty is characterized by:

· A vast uncertainty space.

· A poor point estimate of the true value of the parameter of interest.

· A likelihood-free quantification uncertainty.

To conclude I want to say a few words about the place of the *estimate* in all this, given that it figures as the fulcrum of *Group-Local*'s model of uncertainty. I go into this question in the next chapter. So here I shall only alert the reader to the following:

Post Script:
Fred, the renowned expert on robustness, urged me to impress upon the reader how extremely difficult it is to make sense of a poor estimate in a framework that expresses the severity of the uncertainty in terms of a LIKELIHOOD-FREE model.

This remark refers to the countless conversations that I had with Fred about the very idea of contemplating an "estimate" in a model of severe uncertainty where the uncertainty is likelihood-free. I shall examine this issue at various stages of the discussion. At this point I need only say that in the framework of such models it is hard to see how an "estimate" can be treated (technically speaking) differently from any other point that would presumably represent other "possible" values of the parameter of interest. Moreover, if the uncertainty is indeed "severe", then surely, the estimate is assumed to be poor, indeed it can be substantially wrong, hence no more than . . . a "wild guess".

So the question is:

How would you incorporate a WILD GUESS in a likelihood-free model of uncertainty?

And once you found an answer to this vexed question, consider the obvious follow-up question:

How would you incorporate a GOOD, or EXCELLENT estimate, in a likelihood-free model of uncertainty?

Think about these questions!

2.9 Bibliographic notes

Some robustness models, for instance those based on the seemingly intuitive *Size Criterion*, were developed in the early days of robust optimization (e.g. Rosenhead et al. 1972, Rosenblat 1987, Kouvelis and Yu 1997). However, it was soon discovered that these models give rise to virtually intractable optimization problems, namely problems that can be solved in practice only if certain simplifying conditions hold (e.g. Starr 1963, 1966, Schneller and Sphicas 1983, Eiselt and Langley 1990, Eiselt et al. 1998).

In subsequent chapters I discuss more popular local and global robustness models.

As for the imperative to maintain a strict distinction between *local* and *global* robustness. I should point out that Ben-Tal et al. (2009b) argue that models of robustness, aimed for situations of severe uncertainty, that consider only the parameter's "normal range" of values — rather than the entire uncertainty space — represent a somewhat "irresponsible" decision-maker.

I shall have a great deal more to say about this matter in subsequent chapters of this book.

Chapter 3

Radius of Stability

3.1 Introduction

In this chapter I examine a concept that offers what seems to be the most intuitively obvious way to describe the idea of *local* robustness. This important concept is used extensively in many disciplines, for instance in numerical and applied mathematics, parametric optimization, operations research, economics, and control theory. It is known universally as *Radius of Stability* or *Stability Radius*.

The great merit of expressing *local* robustness in terms of the *Radius of Stability* is that this enables giving the idea of *local* robustness a simple, compelling, graphic rendition. The same is true about the terminology associated with this concept, but more on this in due course.

I describe this concept in three different ways: verbally, graphically, and mathematically. However, before I can do this I need to explain the epithet itself. That is, I need to explain why the notion of "local robustness" is referred to via the term *Radius of Stability* and not say, via "radius of local robustness".

The reason for this is quite prosaic and is due simply to the historical fact that the idea encapsulated in this concept was developed in areas of expertise whose main concern is with *Stability* (for instance numerical analysis and control theory), so that it was labeled *Radius of Stability*. But, in and of itself this idea has got nothing to do with the property of STABILITY. By this I mean of course that the concept *Radius of Stability* is based on a more rudimentary, or if you will, a more abstract notion, which due to its explanatory power, is used metaphorically to convey the concept of *stability*. This rudimentary notion simply designates the following:

> The *distance* from a *point* to a *set*.
>
> Or more accurately, the distance from a point outside a given set to the point in the set that is nearest to it.

This notion is illustrated graphically in Figure 3.1, where p denotes the point and S denotes the set under consideration.

The fact that the shortest distance from point p to set S is equal to the *radius of the smallest circle* centered at p that is tangent to S is thus "borrowed" to express a PERTURBATION in a (parameter of a) system, hence a means to determine the system's stability/robustness. In this framework, the point designates a *nominal value* of a parameter with respect to which the system is "stable", and the set designates the values of the parameter with respect to which the system is "unstable".

So, if p is an element of S, then the shortest distance from p to S is equal to zero, and so is the *Radius of Stability.*

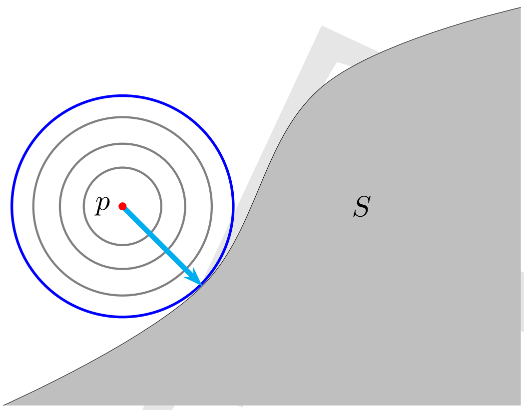


Figure 3.1: Distance from a point p to a set S

3.2 Example

Consider the situation depicted in Figure 3.2. This picture describes the emotional state of a dog called *Rex*. Note that *Rex's* emotional state is influenced by two different stimuli. The two stimuli are denoted by two key parameters. The point $\tilde{p} = (\tilde{p}_1, \tilde{p}_2)$ represents the *nominal values* of the two parameters.

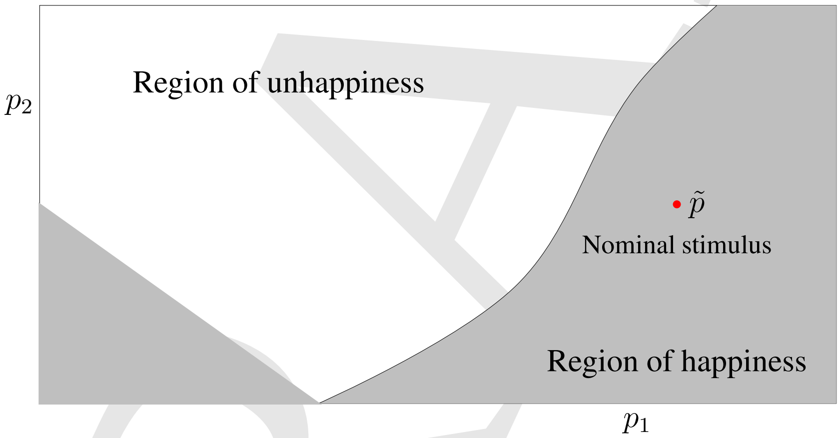


Figure 3.2: Rex's state of mind as a function of the stimulus p

Now, suppose that we want to find out to what degree do *perturbations* in the nominal stimulus $\tilde{p}$ affect Rex's happiness. The question that we would therefore ask is:

How robust is Rex's happiness to small perturbations in the nominal stimulus $\tilde{p}$?

To answer this question it is clear that we would seek to determine the distance of the nominal stimulus $\tilde{p}$ from the *Region of unhappiness*, in which case the question would be:

What is the distance from $\tilde{p}$ to the nearest point in the *Region of unhappiness*?

Or, more accurately:

What is the distance from $\tilde{p}$ to (the nearest point on) the boundary separating the *Region of happiness* and the *Region of unhappiness*?

As illustrated in Figure 3.3, this distance is equal to the RADIUS of the SMALLEST CIRCLE that is tangent with the boundary between the *Region of happiness* and the *Region of unhappiness*.

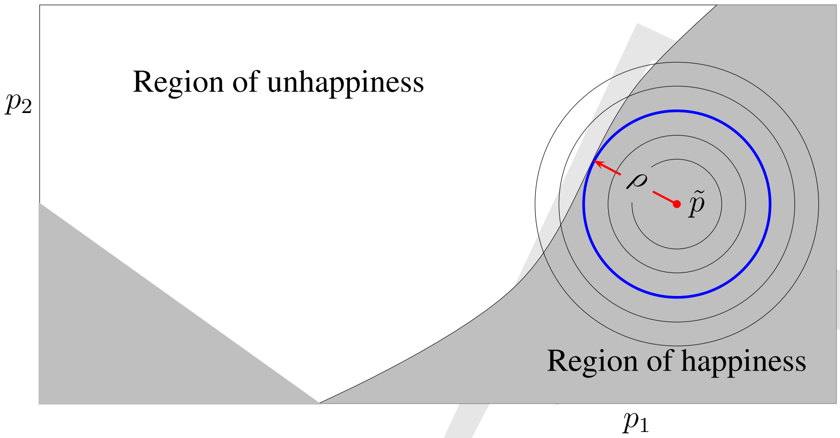


Figure 3.3: Rex's radius of happiness at $\tilde{p}$

Take note, however, that it is also the RADIUS of the LARGEST CIRCLE that is contained in the REGION OF HAPPINESS.

I call attention to this fact because my encounter with *info-gap analysts and practitioners* over the past seven years has shown that one cannot overstate the importance of a correct reading of this fact. What must be appreciated in this regard is that although a *Radius of Stability* based analysis identifies the *largest circle* in a certain *region*, the stability/robustness signified by this *largest circle* is nevertheless LOCAL, indeed it cannot be anything but LOCAL.

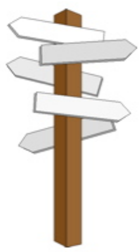
The point to note here is that the *Radius of Stability* denotes the size of the SMALLEST destabilizing perturbation in the nominal value of the parameter of interest. That the smallest destabilizing perturbation can be large does not alter the fact that, both methodologically and practically, the stability/robustness obtained is *local*. The "localness" of the robustness obtained is due to its being measured **only** with respect to a distance to the NOMINAL VALUE of the parameter of interest. So clearly, the robustness analysis in this case is in the LOCALE of the nominal value.

What is more, there are no provisions in the definition of the *Radius of Stability* to consider large perturbations in the nominal value of the parameter of interest that go beyond the smallest perturbations that effectively destabilize the system.

A correct reading of this fact is essential to a correct interpretation of the results generated by *Radius of Stability models.* Thus, methods such as *info-gap decision theory,* that claim to seek the most robust systems/decisions, and to do this deploy the *Radius of Stability* as a measure of robustness, in effect maximize the *Radius of Stability* over the set of available decisions[1]. As we shall later see, these results cannot be proclaimed (the most) robust to severe uncertainty.

3.2.1 Variations on a theme

There are various ways to describe/define the *Radius of Stability* of a system. So, before I proceed to set out the formulation that I shall adopt in this book, let us take a quick look at four (slightly) different versions of this idea. Readers who have no great interest in the minutia of such an analysis can proceed directly to section 3.3.



Consider then the following verbal formulation of the *Radius of Stability*.

Radius of Stability: Version 1.

[1]Namely, all other thing being equal, the larger the *Radius of Stability* the more robust the decision.

> The radius of stability of a system — with respect to a nominal *state* — is the size of the *smallest perturbation* in the nominal state that can destabilize the system.

This statement is immediately meaningful because it overtly relates stability to **small** perturbations in the nominal state. In other words, it gives expression to the idea that it makes ample sense to gauge stability in terms of the affect of the **smallest destabilizing perturbation** in this state. This phrasing also explains why this concept constitutes a measure of *local* robustness.

Thus, on this version, the *Radius of Stability* of a system is a measure of the distance of the nominal state from the system's region of instability. This interpretation is illustrated in Figure 3.4, where

$$\begin{aligned} \tilde{s} &= \text{nominal state of the system.} \\ S(q) &= \text{state space of system } q. \\ S_{stable}(q) &= \text{region of stability of system } q. \\ S_{unstable}(q) &= \text{region of instability of system } q. \\ \rho &= \text{radius of stability of system } q \text{ at } p = \tilde{p}. \end{aligned}$$

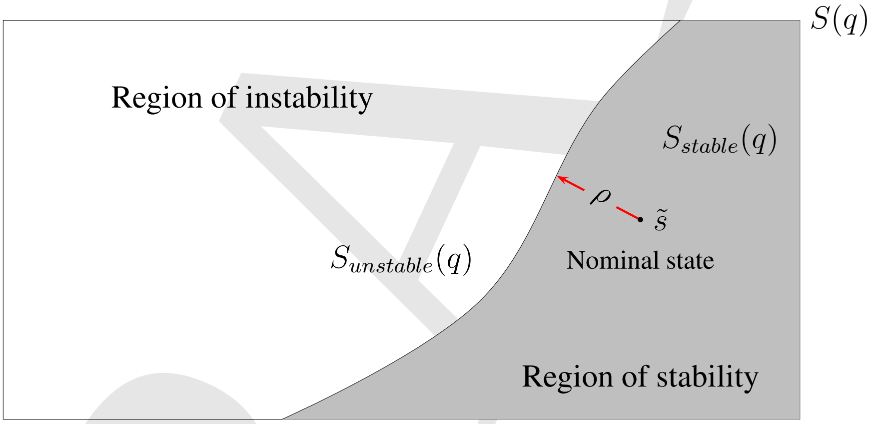


Figure 3.4: Radius of Stability of system q at $\tilde{s}$: Version 1

The red arrow connects the nominal state $\tilde{s}$ to the nearest state on the boundary of the region of instability, and the *length* of this arrow is the *Radius of Stability* of the system[2].

To illustrate, suppose that the *Radius of Stability* of the system under consideration is equal to 5. Then, given the above, the implication is that any perturbation (in any "direction") in the value of $\tilde{s}$ whose size (length) is less than 5 will yield a stable state. Differently put, ANY perturbation in the value of $\tilde{s}$ of a length less than 5, will not destabilize the system, and any perturbation of 5 or more will destabilize it. The ANY means *in any direction*.

Suppose next that the size (length) of the perturbation is equal to 6. Will this perturbation destabilize the system?

To answer this question let us draw a circle of radius 6 centered at $\tilde{s}$ and let us examine the points in this circle. Since the system's *Radius of Stability* is equal to 5, we know that some of the points in the larger circle represent unstable states meaning that, the perturbations associated with these unstable states do destabilize the system.

So, the answer to the above question is as follows:

[2] It is assumed that in this case the boundary is contained in the region of instability.

> Since the *Radius of Stability* of the system is equal to 5, we know in advance that some, but not all, the perturbations of size 6 will destabilize the system.

This is shown in Figure 3.5 where the small circle is of radius 5 and the large one is of radius 6.

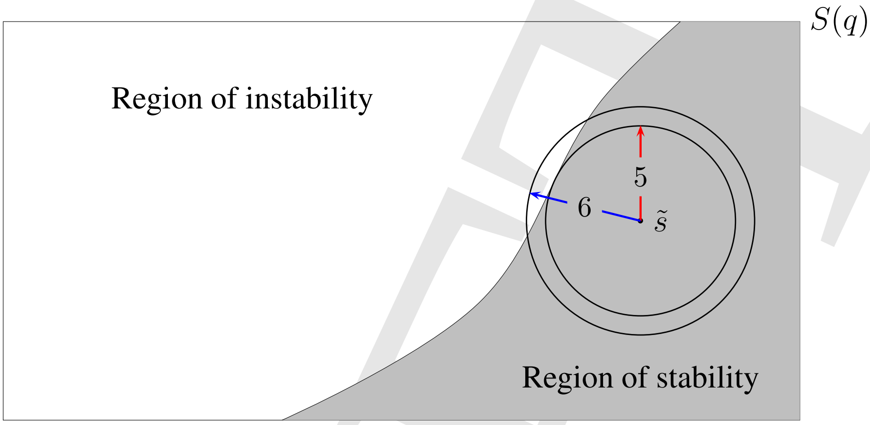


Figure 3.5: Destabilizing perturbations

To set the stage for my formulation of the second version of the *Radius of Stability*, I need to point out that although it is equivalent to Version 1, its graphic representation can be somewhat more intricate. This is so because it uses "neighborhoods" or "balls" — similarly to the circles depicted in Figure 3.5 — to represent perturbations in the value of the nominal state.

So, as in Chapter 2, let $B(\rho, \tilde{s})$ denote a ball of radius ρ centered at $\tilde{s}$. For any $\rho \geq 0$, the ball $B(\rho, \tilde{s})$ is a subset of the *state space* $\mathcal{S}$ consisting of all the points in $\mathcal{S}$ that are within a distance ρ from $\tilde{s}$. Hence, such balls have these two essential properties:

- **Contraction:** $B(0, \tilde{s})$ is the singleton $\{\tilde{s}\}$.
- **Nesting**: If $\rho' < \rho''$ then $B(\rho', \tilde{s})$ is a subset of $B(\rho'', \tilde{s})$.

This means of course, as already indicated in Chapter 2, that "balls" need not necessarily be circular. For instance, if $\mathcal{S}$ is a two dimensional Euclidean space, a ball $B(\rho, \tilde{s})$, can be say, a *rectangle* of sides ρ and 2ρ, centered at $\tilde{s}$. Each ball would thus be a rectangle whose width is twice as large as its height. There can be infinitely many such balls. Figure 3.6 displays 4 such balls.

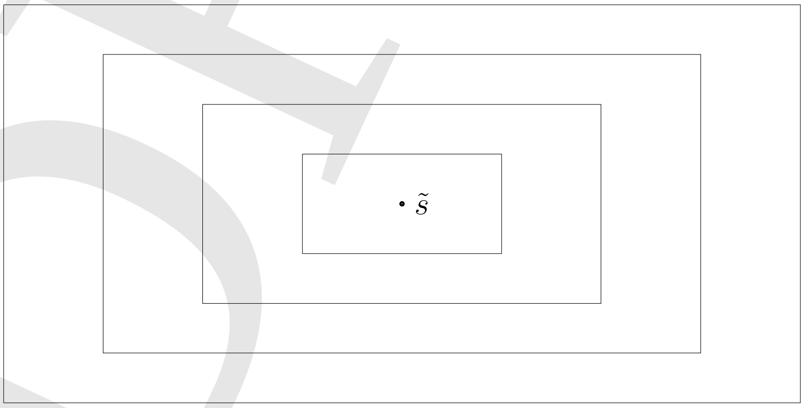


Figure 3.6: Rectangular balls centered at $\tilde{s}$

As we shall see, the definition of the "balls" is often a matter of choice, which means that it is somewhat subjective.

Consider then the following definition of the *Radius of Stability* which is based on concepts such as "ball" and "neighborhood":

> **Radius of Stability:** Version 2.
> The radius of stability of a system — with respect to its nominal state $\tilde{s}$ — is the radius of the *smallest ball* centered at $\tilde{s}$ that contains at least one *unstable* state.

This is illustrated in Figure 3.7, where the assumption is that the *boundary* between the regions is part of the region of instability. That is, the points on the boundary of the circle that are in the region of instability represent perturbations that destabilize the system.

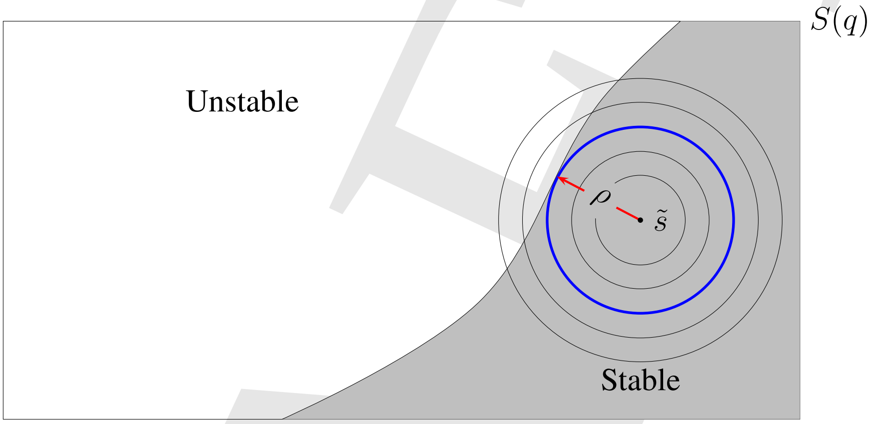


Figure 3.7: Radius of Stability of system q at $\tilde{s}$: Version 2

For the purposes of this discussion, a slightly modified Version 2 would be more appropriate. Consider then the following:

> **Radius of Stability:** Version 3.
> The radius of stability of a system — with respect to its nominal state $\tilde{s}$ — is the radius of the *largest ball* centered at $\tilde{s}$ that is contained in the region of stability of the system.

This is illustrated in Figure 3.7. Here the boundary between the two regions is part of the region of stability.

Observe that if $\tilde{s}$ is on the boundary of the region of stability, then according to Version 3, the *Radius of Stability* of the system is equal to zero: any ball centered at $\tilde{s}$ whose radius is greater than zero will contain an unstable state.

I note in passing that this discussion is based on the premiss that the nominal state of the system is stable, namely that $\tilde{s}$ is in the region of stability of system q. This means of course that we will not have to consider situations where $\tilde{s}$ is an unstable state.

To my mind, definitions of the *Radius of Stability* that are based on the immediately meaningful notion *stable ball* are particularly suitable for this discussion. Consider then the definition of a *stable ball*:

> **Stable ball**:
> A ball $B(\rho, \tilde{s})$ is said to be stable iff all its points are stable.

In other words, a ball is stable iff it is a subset of the region of stability of the system. Hence,

> **Radius of Stability**: Version 4.
> The radius of stability of a system — with respect to its nominal state $\tilde{s}$ — is the radius of the largest stable ball centered at $\tilde{s}$.

3.3 Formal definition

Let us begin by recalling the story behind the concept *Radius of Stability*. We have a set of systems, Q, whose states can be either stable or unstable. The state space of system q, denoted by $S(q)$, is partitioned into a region of stability $S_{stable}(q)$ and a region of instability $S_{unstable}(q)$. A nominal (stable) state $\tilde{s}$ is given. So, the question is:

> How robust are the systems to small perturbation in the value of the nominal state $\tilde{s}$?

The answer is spelled out by[3]:

Formal Radius of Stability Model:

$$\rho(q,\tilde{s}) := \max_{\rho\geq 0}\{\rho : s \in S_{stable}(q), \forall s \in B(\rho,\tilde{s})\}\ ,\ q\in Q \tag{3.1}$$

In words,

> The *Radius of Stability* of system q at $\tilde{s}$, denoted $\rho(q,\tilde{s})$, is the radius of the largest ball centered at $\tilde{s}$ all of whose elements are stable.

Let us consider now a specific representation of the region of stability $S_{stable}(q)$ that is particularly pertinent to this discussion, namely a representation where the system's region of stability is defined by a simple *performance requirement* of the form:

$$r^* \leq r(q,s)\ ,\ q \in Q, s \in S(q) \tag{3.2}$$

where r^* is some given numeric scalar representing the *critical level of performance* and $r(q,s)$ denotes the *performance level* of system q at state s.

Hence,

Radius of Stability Model of interest:

$$\rho(q,\tilde{s}) := \max_{\rho\geq 0}\{\rho : r^* \leq r(q,s), \forall s \in B(\rho,\tilde{s})\}\ ,\ q\in Q \tag{3.3}$$

As we shall see, this is precisely the *robustness model* deployed by *info-gap decision theory*. Its generic format is shown in Figure 3.8.

Thus, from the viewpoint of the *Radius of Stability*, the picture is this:

> Info-gap's robustness model is a *Radius of Stability* model characterized by regions of stability of the form

$$S_{stable}(q) = \{s \in S(q) : r^* \leq r(q,s)\}\ ,\ q \in Q \tag{3.4}$$

Let us now examine this model in action.

[3] This formal model is based on the assumption that the max is attained.

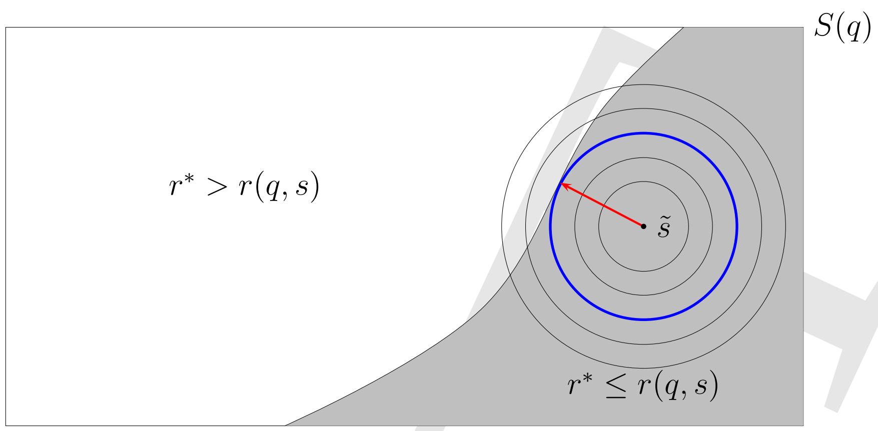


Figure 3.8: Radius of Stability of system q at $\tilde{s}$ a la info-gap decision theory

3.3.1 Example

Consider the situation where Q is some set, and

$$S(q) = \mathcal{S} = \{s \in \mathbb{R}^2 : s_1, s_2 \geq 0\} \ , \ \forall q \in Q \tag{3.5}$$

$$B(\rho, \tilde{s}) = \{s \in \mathcal{S} : (s_1 - \tilde{s}_1)^2 + (s_2 - \tilde{s}_2)^2 \leq \rho^2\} \ , \ \rho \geq 0, \tilde{s} \in \mathcal{S} \tag{3.6}$$

$$r(q, s) = a(q)s_1 + b(q)s_2 \tag{3.7}$$

where $\mathbb{R}$ denotes the real line, and $a(q)$ and $b(q)$ are some numeric scalars whose values depend on q, observing that $B(\rho, \tilde{s})$ is a circle of radius ρ centered at $\tilde{s} \in \mathcal{S}$.

So, in this case the *Radius of Stability* of q is the radius of the largest circle centered at $\tilde{s}$ such that every point s in this circle satisfies the performance requirement $a(q)s_1 + b(q)s_2 \geq r^*$.

Now, consider a specific system, say $q' \in Q$, such that $a(q') = 1$ and $b(q') = 2$ and assume that $r^* = 12$ and $\tilde{s} = (9, 4)$.

Figure 3.9 depicts the region of stability associated with system $q' \in Q$ and it illustrates how the *Radius of Stability* is determined.

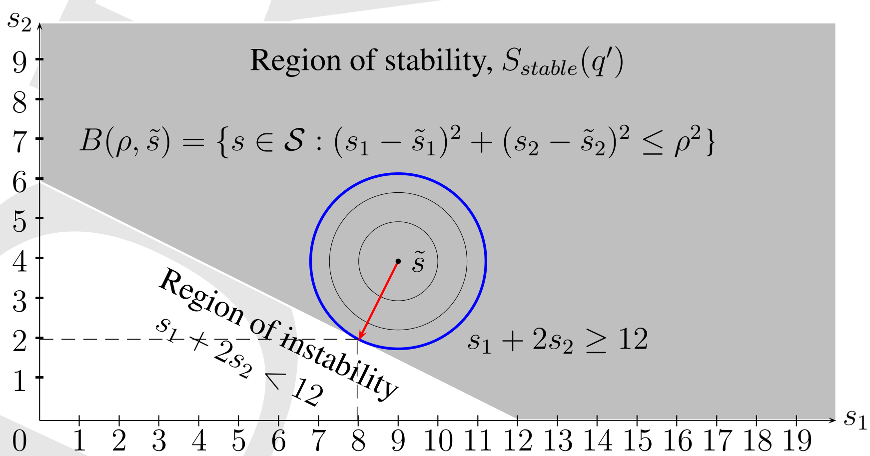


Figure 3.9: Radius of Stability of system q' at $\tilde{s}$

Note that the largest circle centered at $\tilde{s}$ that is contained in the region of stability is tangent with the boundary of this region, namely the line $s_1 + 2s_2 = 12$, at $s' = (8, 2)$. Hence, the *Radius of Stability* is equal to

$$\sqrt{(s'_1 - \tilde{s}_1)^2 + (s'_2 - \tilde{s}_2)^2} = \sqrt{(8-9)^2 + (2-4)^2} = \sqrt{5} \tag{3.8}$$

3.4 Radius of Instability

Suppose that the nominal state $\tilde{s}$ is unstable. Then, to measure the system's instability with respect to perturbations in $\tilde{s}$, we would ask: how far is $\tilde{s}$ from the system's region of stability? That is, we would seek to determine the distance of $\tilde{s}$ to the nearest stable state.

This distance is called the *Radius of Instability* of the system at $\tilde{s}$. Hence,

Formal Radius of Instability Model:

$$\sigma(q, \tilde{s}) := \min_{\rho \geq 0} \{\rho : s \in S_{stable}(q) \text{ for at least one } s \in B(\rho, \tilde{s})\} \ , \ q \in Q \tag{3.9}$$

In words,

> The *Radius of Instability* of system q at $\tilde{s}$ is the radius of the smallest ball around $\tilde{u}$ containing at least one stable state.

This is illustrate in Figure 3.10.

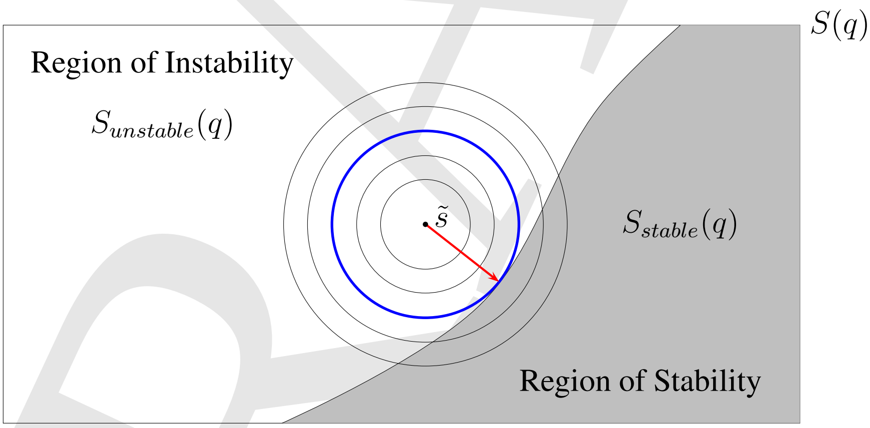


Figure 3.10: Radius of Instability of system q at $\tilde{s}$

Clearly then, technically as well as conceptually, the *Radius of Instability* can be regarded as the counterpart of the *Radius of Stability* of a system:

Issue	*Radius of Stability*	*Radius of Instability*
Nominal point	· $\tilde{s}$ is a stable state	· $\tilde{s}$ is an unstable state
Question asked	· How far is $\tilde{s}$ to instability?	· How far is $\tilde{s}$ to stability?
Preference	· The larger the better	· The smaller the better

As we shall see, the "Preference" issue is of the utmost importance.

> **G'day Moshe:**
> It is important to bring to the reader's attention that the question of *Preference* is central to the development of *conceptual models* of *Radius of Stability* and of *Radius of Instability*.

It must be made clear that the first lends itself to a *worst-case treatment*, whereas the latter lends itself to a *best-case treatment*. This means that in the framework of *classical decision theory*, the former is a simple *Maximin model* and the latter is a simple *Minimin model*.

Cheers,
Fred
Internationally Known Expert on Robust Decision-Making Under Severe Uncertainty

G'day Fred:

Point taken!

I discuss these and related topics in subsequent sections of this chapter as well as in other chapters.

Cheers,
Moshe

So, let us go back to where we left off.

A *Radius of Instability* model that is of particular interest to us here is the one deployed by *info-gap decision theory*, which in the info-gap jargon is called *opportuneness*. In this case the stability requirement is $r^\circ \leq r(q,s)$, hence,

Radius of Instability Model of Interest:

$$\sigma(q,\tilde{s}) := \min_{\rho\geq 0} \{\rho : r^\circ \leq r(q,s) \text{ for at least one } s \text{ in } B(\rho,\tilde{s})\} \tag{3.10}$$

In words, the *Radius of Instability* of the system is the radius of the smallest ball centered at $\tilde{s}$ that contains at least one stable state. The generic format is shown in Figure 3.11.

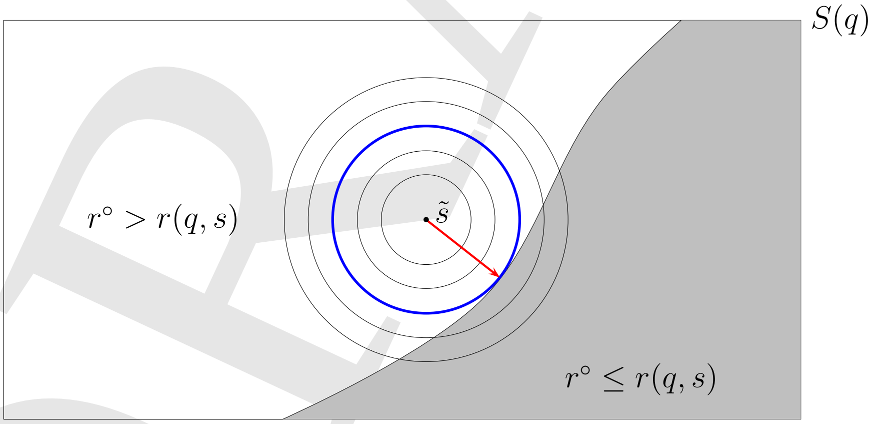


Figure 3.11: Radius of Instability of system q at $\tilde{s}$ a la info-gap decision theory

Example

As in the previous example, consider the situation where Q is some set, and

$$S(q) = \mathcal{S} = \{s \in \mathbb{R}^2 : s_1, s_2 \geq 0\}\ ,\ q \in Q \tag{3.11}$$

$$B(\rho,\tilde{s}) = \{s \in S : (s_1 - \tilde{s}_1)^2 + (s_2 - \tilde{s}_2)^2 \leq \rho^2\}\ , \rho \geq 0, \tilde{s} \in S \tag{3.12}$$

$$r(q,s) = a(q)s_1 + b(q)s_2 \tag{3.13}$$

where $\mathbb{R}$ denotes the real line, and $a(q)$ and $b(q)$ are some numeric scalars whose values depend on q, observing that $B(\rho, \tilde{s})$ is a circle of radius ρ centered at $\tilde{s} \in \mathcal{S}$.

So in this case the *Radius of Instability* of q is the radius of the smallest circle centered at $\tilde{s}$ such that at least one point s in this circle satisfies the performance requirement $a(q)s_1 + b(q)s_2 \geq r^{\circ}$.

Now, consider a specific system, say $q' \in Q$, such that $a(q') = 1$ and $b(q') = 2$ and assume that $r^{\circ} = 16$ and $\tilde{s} = (7, 3)$.

Figure 3.12 depicts the region of stability associated with system $q' \in Q$ illustrating how the *Radius of Instability* is determined.

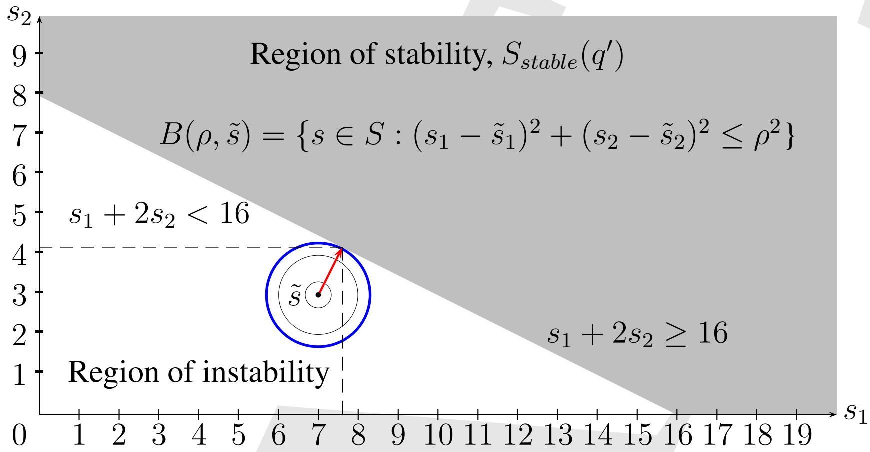
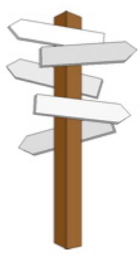


Figure 3.12: Radius of Instability of system q' at $\tilde{s}$

Note that the largest circle centered at $\tilde{s}$ that is contained in the region of stability is tangent with the boundary of this region, namely the line $s_1 + 2s_2 = 12$, at $s' = (7.6, 4.2)$. Hence, the *Radius of Instability* is equal to

$$\sqrt{(s_1' - \tilde{s}_1)^2 + (s_2' - \tilde{s}_2)^2} = \sqrt{(7.6 - 7)^2 + (4.2 - 3)^2} = \sqrt{1.8} \tag{3.14}$$

3.4.1 What next?



My exposition thus far has covered all the essentials of the concept *Radius of Stability*. So, readers with an only limited interest in this topic can proceed directly from here to the next chapter.

I submit, though, that many readers should find my discussion in the remaining sections of this chapter most instructive, particularly for the relationship that I identify between the *Radius of Stability* and other concepts/models. The relationship that I identify throws an interesting light both on the *Radius of Stability* and the concepts/models that I associate with it.

So, the first item on my agenda is to expound the *worst-case* orientation of the concept *Radius of Stability*.

3.5 Worst Case Perspective

You will recall that *worst-case analysis* entails that systems are evaluated and ranked on the basis of their *worst-case performance*. The question is then how would this reasoning be reflected in the working of the *Radius of Stability*.

It is immediately clear that to be able to bring this out, the first thing to do is to relate *stability/instablility* to *best/worst*. Based on this we can say that the *worst* state in $S(q)$ is any *unstable* state associated with system q, if $S(q)$ contains such states; and any state is *worst*, if all the states in S(q) are either stable or unstable. Similarly, the *best* state in $S(q)$ is any *stable* state, if $S(q)$ contains such states; and any state is *best*, if all the states are unstable.

We can therefore distinguish between the following three cases:

· All the elements of $S(q)$ are stable.
System q is "super-robust". Each element of $S(q)$ is both a worst case and a best case.

· All the elements of $S(q)$ are unstable.
System q is ''super-fragile". Each element of $S(q)$ is both a worst case and a best case.

· Some the elements of $S(q)$ are stable, some are unstable.
The unstable elements of $S(q)$ are worst cases, the stable elements of $S(q)$ are best cases.

In short, any state in $S(q)$ is either a *worst* state, or a *best* state, or both. The latter occurs if all the elements of $S(q)$ are of the same type (stable or unstable).

Exactly the same classification holds for the elements of the balls $B(\rho,\tilde{u}), \rho \geq 0$. That is,

· **Worst state in** $B(\rho,\tilde{s})$:
The worst state in $B(\rho,\tilde{s})$ is: any unstable state in $B(\rho,\tilde{s})$ if there are such states in $B(\rho,\tilde{s})$; and any state in $B(\rho,\tilde{s})$ if all the states in $B(\rho,\tilde{s})$ are stable.

· **Best state in** $B(\rho,\tilde{s})$:
The best state in $B(\rho,\tilde{s})$ is: any stable state in $B(\rho,\tilde{s})$ if there are such states in $B(\rho,\tilde{s})$; and any state in $B(\rho,\tilde{s})$ if all the states in $B(\rho,\tilde{s})$ are unstable.

This classification is illustrated in Figure 3.13. The small ball, $B(\rho',\tilde{s})$, contains only stable points. So, when a worst-case analysis is conducted on this ball, each point in this ball is both a worst case and a best case. The situation is different in the case of the large ball, $B(\rho'',\tilde{s})$. Here, some of the ball's elements are stable, while some are unstable. Hence, in the worst-case analysis on this ball, all the stable points are best cases and all the unstable points are worst cases.

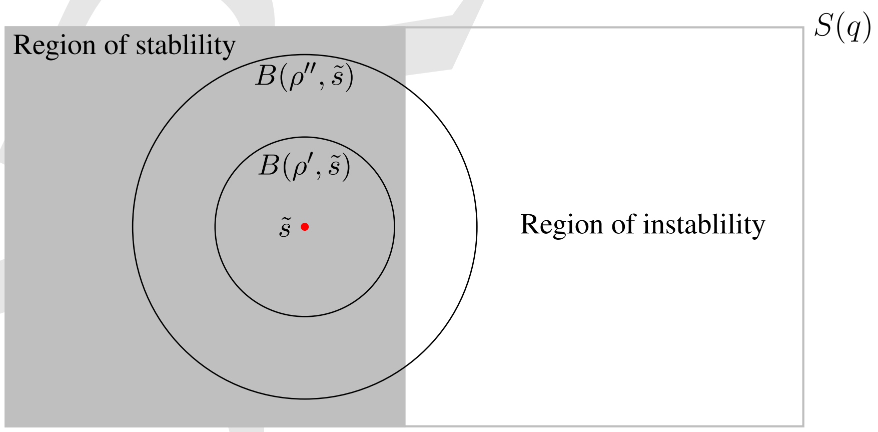


Figure 3.13: Local worst-case analysis on two balls

This paves the way for the following formulation:

Radius of Stability: Version 5.
The radius of stability of a system — with respect to its nominal state $\tilde{s}$ — is the radius of the largest ball centered at $\tilde{s}$ whose worst state is stable.

The merit of affiliating the *Radius of Stability* with worst-case analysis is that this enables formulating the *Radius of Stability* — in a most straightforward indeed, intuitive manner — as a model that places it squarely where it belongs namely in, *Classical Decision Theory*.

G'day Moshe,

I strongly suspect that some readers may fail to see how ALL the elements of a set, say $S(q)$ for some $q \in Q$, can be BOTH *worst* and *best* cases.

In particular, they may wonder how on earth can a stable state be a worst state and an unstable state a best state!

Cheers,
Fred
Internationally renowned Expert on Robust Decision Under Severe Uncertainty

I discuss this issue in detail in Chapter 5 so, here I shall explain it in very broad terms. As indicated above, insofar as *requirements* such as $s \in S_{stable}(q)$ are concerned, potentially there are only two "cases": either the requirement is satisfied (the state is stable), or it is not satisfied (the state is unstable).

Consider then a system $q \in Q$ and a ball $B(\alpha, \tilde{u})$ such that $s \in S_{stable}(q), \forall s \in B(\alpha, \tilde{u})$, namely a case where all the states in $B(\alpha, \tilde{u})$ are stable.

Clearly, since ALL the states in $B(\alpha, \tilde{u})$ are stable, ALL are also *best* cases. This makes perfect sense. What might not make sense is the claim that all these states are also WORST cases! For, how can a STABLE state possibly be a WORST case?!?!?!

Very simple!

By definition, a case, say c, in a set of cases, say C, is a worst case if all other cases in C are at least as good as c. Similarly, c is a best case if it is at least as good as all other cases in C. So if all the cases in C have the same "value", say they are all "stable", then all the elements of C are both worst cases and best cases.

To illustrate, suppose that that C consists of 4 elements, call them $c^{(1)}, c^{(2)}, c^{(3)}$, and $c^{(4)}$, where each is a positive integer indicating the number of dark chocolate bars you might find in your hotel room bar. These cases represent four possible, highly uncertain, scenarios. Now, suppose that you know for a fact that

$$c^{(1)} = 3\ ;\ c^{(2)} = 4\ ;\ c^{(3)} = 1\ ;\ c^{(4)} = 5$$

Then clearly, assuming that you like dark chocolate, the worst case in C is $c^{(3)}$ and the best case in C is $c^{(4)}$.

Suppose next that you know for a fact that

$$c^{(1)} = 2\ ;\ c^{(2)} = 2\ ;\ c^{(3)} = 2\ ;\ c^{(4)} = 2$$

Then clearly, here ALL the cases in C are **equivalent:** all are both worst cases and best cases.

Next, consider the more complicated situation where your recommended daily dose is 3 dark chocolate bars (or more). Furthermore, assume that you are in a "stable" state iff this daily dose is satisfied.

So, under these conditions, how will your stability affect the classification of the following scenarios?

$$c^{(1)} = 2\ ;\ c^{(2)} = 1\ ;\ c^{(3)} = 0\ ;\ c^{(4)} = 1$$

Since ALL the scenarios violate your daily requirement, all are **equivalent:** all are "unstable". Hence, all are both worst cases and best cases.

And what about this set of scenarios?

$$c^{(1)} = 8\ ;\ c^{(2)} = 5\ ;\ c^{(3)} = 9\ ;\ c^{(4)} = 7$$

Since ALL the scenarios satisfice your daily requirement, all are **equivalent:** all are "stable". Hence, all are both worst cases and best cases.

And how about this possibility?

$$c^{(1)} = 1\ ;\ c^{(2)} = 2\ ;\ c^{(3)} = 4\ ;\ c^{(4)} = 1$$

Here $c^{(1)}$, $c^{(2)}$ and $c^{(4)}$ are "unstable" and $c^{(3)}$ is "stable". Hence, $c^{(1)}$, $c^{(2)}$ and $c^{(4)}$ are worst cases and $c^{(3)}$ is a best case.

Some readers may argue that the preceding analysis is badly flawed, because it takes no account of "degrees" or "levels" of "satisfaction" or "stability". That is, the argument would be that for dark chocolate lovers, regardless of their daily dose requirement being met, the "more" they can have the "better", PERIOD. Hence, if for instance

$$c^{(1)} = 8\ ;\ c^{(2)} = 2\ ;\ c^{(3)} = 4\ ;\ c^{(4)} = 1$$

then the worst case is $c^{(4)}$ and the best case is $c^{(1)}$. Similarly, if

$$c^{(1)} = 8\ ;\ c^{(2)} = 6\ ;\ c^{(3)} = 4\ ;\ c^{(4)} = 5$$

then the worst case is $c^{(3)}$ and the best case is $c^{(1)}$.

Let me point out then that one of the objectives of this short digression was to make it clear that standard models of robustness, such as the *Radius of Stability* model, do **not** concern themselves with "degrees" of stability, or degrees of violation of a requirement.

As I emphasized above, a state of a system is either stable or unstable, but not both. There is no "degree" or "level" of stability. Similarly, a constraint, say $r^* \leq r(q,s)$, is either satisfied or violated. This is clearly indicated in the formal description of models such as

$$\max_{\rho \geq 0} \{\rho : s \in S(q), \forall s \in B(\alpha, \tilde{s})\} \tag{3.15}$$

and

$$\max_{\rho \geq 0} \{\rho : r^* \leq r(q,s), \forall s \in B(\alpha, \tilde{s})\} \tag{3.16}$$

Note, for example, that in (3.16) there is no distinction between a "small" violation and a "big" violation of the constraint $r^* \leq r(q,s)$. A violation is a ... violation, regardless of how small or big it is.

I should point out, though, that this does not entail that such models cannot be refined so as to accommodate situations where it is required or desirable to account for "degrees" or "levels" of stability or, "degrees" or "levels" of satisfaction/violation of requirements. This only means that standard *Radius of Stability* models are not designed for this purpose so that structurally they lack this feature.

More on this in Chapter 5. However, before I can proceed I want to call attention to the importance of a correct reading of the manner in which the worst-case analysis is performed by the *Radius of Stability* model.

3.5.1 Warning

!

My experience with *info-gap decision theory* over the past seven years has brought home to me how deeply rooted are the misconceptions about the worst-case analysis. This is manifested not only in the failure to appreciate the connection between *info-gap's robustness analysis* and worst-case analysis, but also the fundamental differences between *local* and *global* worst-case analysis. I therefore deemed it necessary to alert the readers to these facts (see Figure 3.14) so as to prevent the errors that are endemic in *info-gap decision theory* as a result of these misconceptions. I ask those readers who may find this warning superfluous, perhaps even insulting, to bear with me when I call attention to the points made in Figure 3.14.

Local vs Global worst-case analysis

It is important to be clear on how the worst-cases analysis works in the context of the *Radius of Stability* model.

- The *Radius of Stability* model provides a measure of the stability of a system. It is therefore concerned with situations of only two levels of stability. Namely, the basic assumption is that a state $s \in S(q)$ is either "stable" or "unstable" (but not both) with respect to system q.
- Furthermore, in the framework of this model, the worst-case analysis is conducted not on the state space $S(q)$ of system q but on the balls $B(\rho, \tilde{s}), \rho \geq 0$.
- Therefore, subject to conventional regularity conditions, for any $\rho \geq 0$ the ball $B(\rho, \tilde{s})$ has a worst state and a best state. Hence, the *existence* of a worst case or a best case, is not an issue here, even if the state space $S(q)$ is unbounded.
- This is the case, for example, in situations where the balls are compact sets and the worst state in a ball is determined by optimizing a continuous real-valued function on this ball.
- The point to keep in mind here is that the existence of a worst state in $B(\rho, \tilde{s})$ is not contingent on the existence of a worst state in $S(q)$. In particular, the fact that there is no worst state in $S(q)$ does not imply that for each $\rho \geq 0$ there is no worst state in $B(\rho, \tilde{s})$.

Figure 3.14: Global warning

These remarks require that I make my usage of this terminology crystal clear. Observe then that unless I expressly state otherwise:

> I use the term *local worst-case analysis* to designate a worst case analysis that is conducted on a ball $B(\rho, \tilde{s})$ whose radius ρ is finite.

In the same vein:

> I use the term *global worst-case analysis* to designate a worst-case analysis over the entire state space, namely $S(q)$, of system q.

And to play it even safer:

> Unless I specifically indicate that a worst-case analysis is *global,* the working assumption is that the analysis is local on the assumed ball $B(\rho, \tilde{s})$.

In sum, I shall make it my business to ensure that no ambiguity arises about the domain over which the worst-case analysis is conducted. I am confident that Fred, the internationally known expert on robustness, will remind me to adhere to this commitment.

3.6 Best-Case Perspective

Best-case analysis entails that systems are evaluated and ranked on the basis of their best-case performance. So, in this section I want to examine how the best-case analysis is manifested in the operation of the *Radius of Instability*.

To be on the safe side, I repeat the statements that I made in the discussion on the *worst-case perspective* of the *Radius of Stability.* Recall that in this framework the nominal state $\tilde{s}$ is *unstable*.

Relating *stability/instability* to *best/worst* implies that the *best* state in $S(q)$ is any *stable* state, if $S(q)$ contains such states; and any state is *best*, if all the states are unstable. Similarly, as indicated above, the *worst* state in $S(q)$ is any *unstable* state, if $S(q)$ contains such states; and any state is *worst*, if all the states are stable.

This means that any state in $S(q)$ is either a *worst* state, or a *best* state, or both. The same classification holds for the elements of balls:

- **Worst state in** $B(\rho, \tilde{s})$:
  The worst state in $B(\rho, \tilde{s})$ is: any unstable state in $B(\rho, \tilde{s})$ if there are such states in $B(\rho, \tilde{s})$; and any state in $B(\rho, \tilde{s})$ if all the states in $B(\rho, \tilde{s})$ are stable.
- **Best state in** $B(\rho, \tilde{s})$:
  The best state in $B(\rho, \tilde{s})$ is: any stable state in $B(\rho, \tilde{s})$ if there are such states in $B(\rho, \tilde{s})$; and any state in $B(\rho, \tilde{s})$ if all the states in $B(\rho, \tilde{s})$ are unstable.

This is illustrated in Figure 3.15.

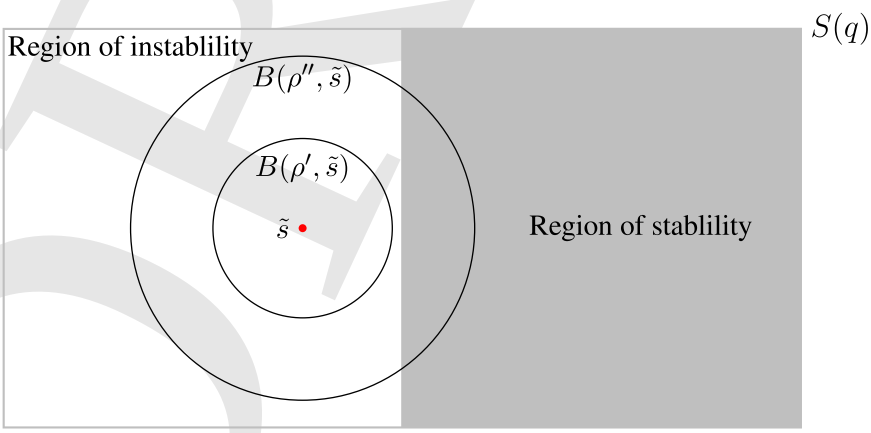


Figure 3.15: Local best-case analysis on two balls

Note that all the elements of the small ball $B(\rho', \tilde{s})$ are worst cases as well as best cases. In contrast, all the stable points in the large ball $B(\rho'', \tilde{s})$ are best cases, and all the unstable points in this ball are worst cases.

Hence,

Radius of Instability:
The radius of instability of a system — with respect to its nominal state $\tilde{s}$ — is the radius of the smallest ball centered at $\tilde{s}$ whose best state is stable.

Parallel to the affiliation of the *Radius of Stability* to the worst-case analysis, the affiliation of the *Radius of Instability* to the best-case analysis enables formulating the *Radius of Instability* directly — indeed, intuitively — as a model that places it squarely where it belongs, namely in *Classical Decision Theory*.

So let us take a quick look at these *classic* formulations of the *Radius of Stability* and the *Radius of Instability*.

3.7 Conceptual models

The formal definitions of the *Radius of Stability* and *Radius of Instability*, accompanied by their self-explanatory graphic renditions, should have given the reader a good grasp of the purport, mode of operation, scop and capabilities of the models that are based on these concepts. As we shall see, though, it is possible to enlarge on and deepen the above explanations by formulating these concepts in terms of what are staple *conceptual models* in *Classical Decision Theory*. These models are known in decision theory, operations research and other areas as "games", or more accurately as "two-person games".

I discuss such models in greater detail in the next chapter as well as in Chapter 5. Here I merely outline them in broad terms.

3.7.1 Radius of stability game

Phrasing the *Radius of Stability* in terms of the classic "two-person game" gives rise to a game involving two players: the Decision Maker (DM) and Nature. The DM plays first by selecting (determining) the radius of a ball centered at the nominal state of the system. Nature responds by selecting the worst state in the ball whose radius was determined by the DM. Based on the radius selected by the DM and the state selected by Nature, a *payoff/reward* is awarded to the DM. The assumption is that the DM seeks to maximize her payoff/reward

Here is a more concise description of the game:

The Radius of Stability Game

Step 1: DM selects a non-negative number ρ^*, representing a ball $B(\rho^*, \tilde{s})$ centered at the nominal state $\tilde{s}$ of the system.

Step 2: Nature selects the WORST state in $B(\rho^*, \tilde{s})$, call it s^*.

Step 3: A payoff, denoted $\text{payoff}(\rho^*, s^*)$, is awarded to DM.

This conceptual model is given a more graphic depiction in Figure 3.16.

It is extremely important to take full note of the defining feature of this game which is that the DM plays first. Because, the implication is that when selecting the state s^* in Step 2 of the game, Nature knows the value of ρ^* that was selected by DM. This greatly simplifies the situation depicted in the classic *zero-sum 2-person games,* where the players select their respective decisions *simultaneously.*

Now, to be able to state the concept *Radius of Stability* in terms of this game we must have at our disposal a formula for determining $\text{payoff}(\rho^*, s^*)$. This is required by the DM for selecting a value

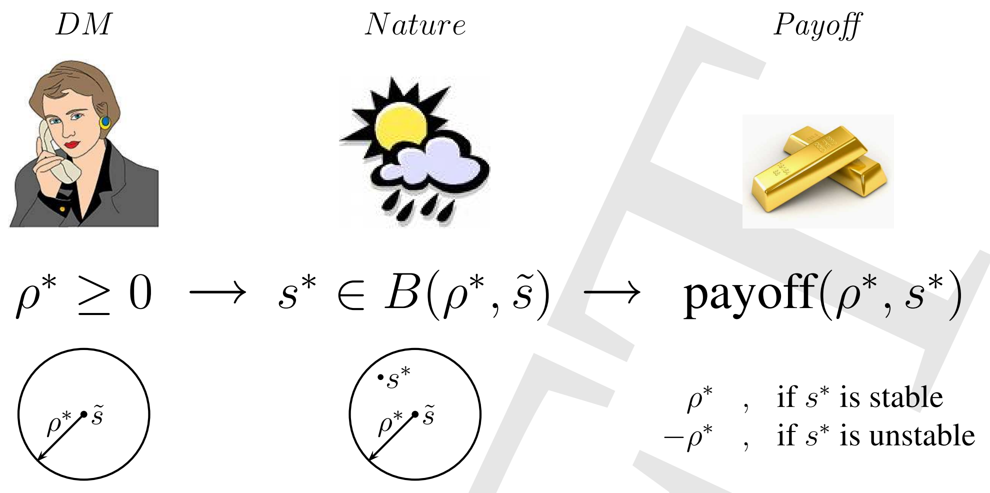


Figure 3.16: Radius of stability game

of ρ^* that is equal to the *Radius of Stability* of the system under consideration. This task presents the following modeling problem:

> How are we to define payoff(ρ^*, s^*) so as to insure that the DM selects the value of ρ that is equal to the *Radius of Stability* of the system under consideration?

As it turns out, though, the task presented by this modeling problem is rather straightforward[4]. To see that this is so, consider the following:

Recipe to determine payoff(ρ^*, s^*)

- If s^* is stable, let payoff$(\rho^*, s^*) = \rho^*$.
- Otherwise, let payoff$(\rho^*, s^*) = -\rho^*$.

Symbolically,

$$\text{payoff}(\rho^*, s^*) := \begin{cases} \rho^* & , \ s^* \in S_{stable}(q) \\ -\rho^* & , \ s^* \in S_{unstable}(q) \end{cases} \tag{3.17}$$

I should point out that the choice of $-\rho^*$ for the $s^* \in S_{unstable}(q)$ case is somewhat arbitrary (but useful!). Any negative value will do.

To ascertain that this recipe can do the job, as above let

$$\rho(q, \tilde{s}) := \text{Radius of stability of system } q \text{ with respect to the nominal state } \tilde{s} \tag{3.18}$$

and consider the following two cases:

- Case 1: The value of ρ^* selected by DM is greater than $\rho(q, \tilde{s})$.
  In this case $B(\rho^*, \tilde{s})$ contains at least one unstable state meaning that the state selected by Nature is

[4]Often, terms such as "clearly" and "straightforward" are used to describe situations where the writer/speaker find it hard to explain an argument, precisely because the simplicity of the matter involved is an impediment to a clear, concise explanation. So these terms are intended to connote the following: "listen: the argument is really and truly very simple. However, it is very difficult to explain it without complicating it. So, it would be best for you to accept that it holds water. If you can't, ask for help. In desperation, send me a note.

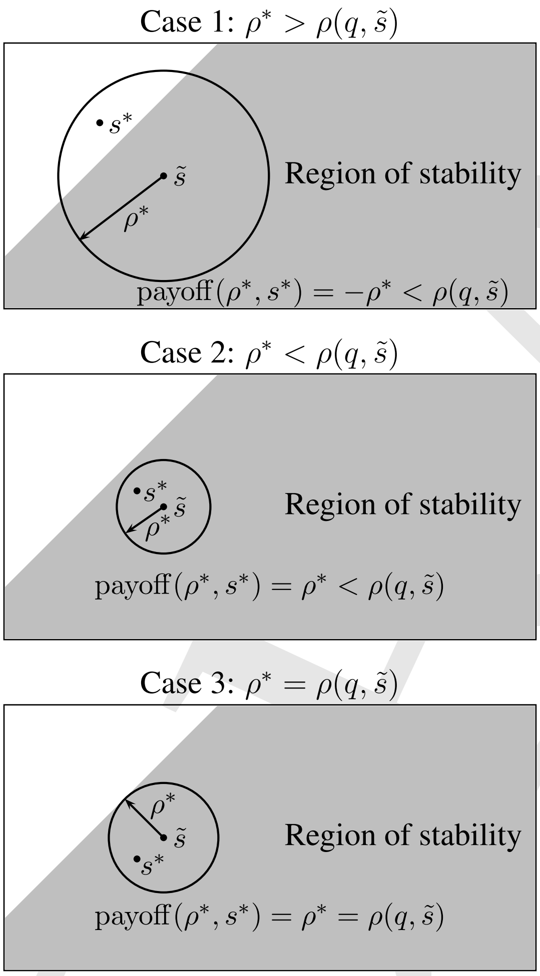
unstable. In this case the payoff determined by the above recipe is equal to $-\rho^*$. It will therefore be to the DM's advantage to select a smaller value of ρ^*.

· Case 2: The value of ρ^* selected by DM is smaller than $\rho(q,\tilde{s})$.
In this case there is a radius larger than that selected by the DM whose ball contains only stable states. Hence, given the above recipe for the payoff, it will be to the DM's advantage to select a larger value of ρ^*, thus increasing her reward.

It follows then that because both in Case 1 and Case 2 the value of ρ^* can be bettered, the only other alternative is this:

Case 3: The DM selects a value of ρ^* that is equal to $\rho(q,\tilde{s})$.

Schematic descriptions of these three cases are shown in Figure 3.17.



Explanation

In this case, the ball $B(\rho^*,\tilde{u})$ contains both stable and unstable states. Hence, the worst state in this ball is an unstable state. Nature will select such a state. Consequently, the payoff would be equal to $-\rho^*$. The DM should therefore opt for a smaller value of ρ^*.

Explanation

In this case, the ball $B(\rho^*,\tilde{u})$ contains only stable states. Hence, the worst state in this ball is a stable state. Nature will select such a state. Consequently, the payoff would be equal to ρ^*. The DM should therefore opt for a larger value of ρ^*.

Explanation

In this case, the ball $B(\rho^*,\tilde{u})$ contains only stable states. Hence, the worst state in this ball is a stable state. Nature will select such a state. Consequently the payoff would be equal to ρ^*. The DM should opt for a larger value of ρ^*, but … any larger ball will immediately trigger Case 1.

Figure 3.17: The three cases of the relationship between ρ^* and $\rho(q,\tilde{s})$

The conclusion to be drawn from all this is that the above recipe does indeed ensures that the optimal value of ρ^* selected by DM in Step 1 of the game — the value of ρ^* that maximizes her payoff — is equal to the *Radius of Stability* of the system under consideration. This means of course that the above game is a thoroughly sound conceptual model for the *Radius of Stability* model.

3.7.2 Radius of instability game

This game is similar to the *Radius of Stability game* discussed above, except that here the DM seeks to minimize her "cost" and, Nature does not oppose the DM but rather cooperates with the DM's effort to achieve her (the DM's) objective.

So, in this case as well we have two players: a decision maker (DM) and Nature. The DM plays first selecting (determining) the radius of a ball centered at the nominal state of the system. Nature responds by selecting the *best* state in this ball. Based on the radius selected by DM and the state selected by Nature, a *cost* is incurred by the DM. The assumption is that the DM seeks to minimize her cost.

The Radius of Instability Game

Step 1: DM selects a non-negative number ρ°, representing a ball $B(\rho^\circ, \tilde{s})$ centered at the nominal state $\tilde{s}$ of the system.

Step 2: Nature selects the best state in $B(\rho^\circ, \tilde{s})$, call it s°.

Step 3: A cost, denoted $cost(\rho^\circ, s^\circ)$, is incurred by DM.

This conceptual model is described pictorially in Figure 3.18.

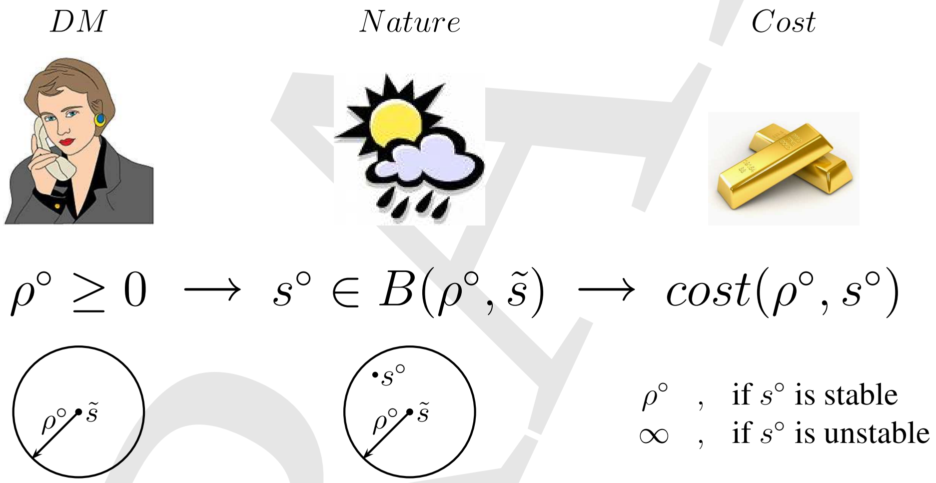


Figure 3.18: Radius of instability game

To repeat the above argument. To be able to formulate the *Radius of Instability* in terms of a two-person game model, we need to have at our disposal a recipe for determining the $cost(\rho^\circ, s^\circ)$ to insure that the DM select a value of ρ° that is equal to the Radius of Instability of the system under consideration.

So, in this case our modeling problem is this:

> Which definition of $cost(\rho^\circ, s^\circ)$ will ensure that the DM select a value of ρ that is equal to the *Radius of Instability* of the system under consideration?

As it turns out this modeling problem is also straightforward. To see that this is so, consider the following:

Recipe for $cost(\rho^\circ, s^\circ)$

- If s° is stable, let $cost(\rho^\circ, s^\circ) = \rho^\circ$.
- Otherwise, let $cost(\rho^\circ, s^\circ) = \infty$.

Symbolically,

$$cost(\rho^\circ, s^\circ) := \begin{cases} \rho^\circ & , \; s^\circ \in S_{stable}(q) \\ \infty & , \; s^\circ \in S_{unstable}(q) \end{cases} \tag{3.19}$$

Here as well the choice of ∞ for the $s^\circ \in S_{unstable}(q)$ case is somewhat arbitrary (but useful!).

To ascertain that this recipe can do the job here as well, as above let,

$$\sigma(q, \tilde{s}) := \text{Radius of instability of system } q \text{ with respect to the nominal state } \tilde{s} \tag{3.20}$$

and consider the following two cases:

· Case 1: The value of ρ° selected by DM is greater than $\sigma(q, \tilde{s})$.
In this case there exists a ball containing a stable state whose radius is smaller than ρ°. It will therefore be in the DM's advantage to select a smaller value of ρ°.

· Case 2: The value of ρ° selected by the DM is smaller than $\sigma(q, \tilde{s})$.
In this case all the states in $B(\rho^\circ, \tilde{s})$ are unstable. Nature will therefore have to select an unstable state. This will result in a huge cost (∞) to DM. Since the DM seeks to minimize her cost, it would be in her advantage to increase the value of ρ° in her search for a stable state.

It follows then that because both in Case 1 and Case 2 the value of ρ° can be bettered, the only other alternative is this:

Case 3: DM selects the value of ρ° that is equal to $\sigma(q, \tilde{s})$.

Schematic descriptions of these three cases are shown in Figure 3.19.

The conclusion to be drawn in this case as well is that the above recipe ensures that the optimal value of ρ° selected by DM in Step 1 of the game — the value of ρ° that minimizes her cost — is equal to the *Radius of Instability* of the system under consideration. And the implication is that, the above game provides a sound conceptual model for the *Radius of Instability*.

3.8 Local vs global stability

I have already pointed out that the concept *Radius of Stability* stipulates a stability that is inherently *local* because it is defined only with regard to $\tilde{s}$ — the nominal state of the system. Put another way, the *Radius of Stability* is a measure of *local stability.*

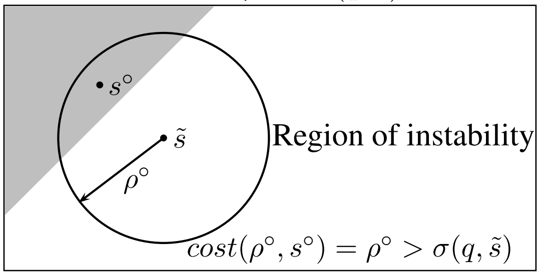
This is illustrated in Figure 3.20 where the *Radius of Stability* of a system is shown for two *nominal* states, s' and s''.

What this figure tells us is that, according to the precepts of the *Radius of Stability* model, system q is much more stable at state $\tilde{s}'$ than at state $\tilde{s}''$. So, if our objective is to maximize the stability of the system and if we have a choice in this matter, we would prefer state $\tilde{s}'$ to state $\tilde{s}''$, all other things being equal.

There are, however, situations where our concern is the *global* stability of the system. In other words, our goal is to determine how stable the system is over its entire state space $S(q)$.

For example, consider the situation shown in Figure 3.21 where the regions of stability two systems, A and B, are shown, observing that both have the same state space $S(A) = S(B) = S$.

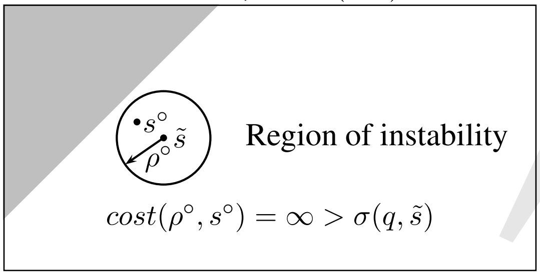
Case 1: $\rho^\circ > \sigma(q, \tilde{s})$



Explanation

In this case, the ball $B(\rho^\circ, \tilde{u})$ contains both stable and unstable states. Hence, the best state in this ball is a stable state. Nature will select such a state. Consequently, the cost would be equal to ρ°. The DM should therefore opt for a smaller value of ρ^*.

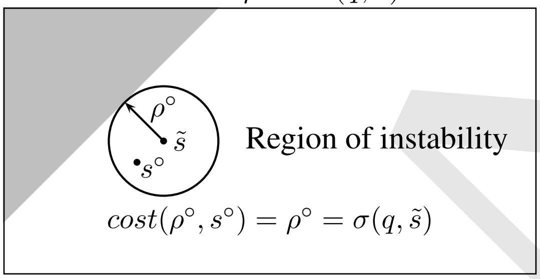
Case 2: $\rho^\circ < \sigma(q, \tilde{s})$



Explanation

In this case, the ball $B(\rho^\circ, \tilde{u})$ contains only unstable states. Therefore, the best state in this ball is an unstable state whereupon Nature will select such a state. The resulting cost will therefore be equal to ∞.

Case 3: $\rho^\circ = \sigma(q, \tilde{s})$



Explanation

In this case, the ball $B(\rho^\circ, \tilde{u})$ contains both unstable and stable states (perhaps only one stable state). Therefore, the best state in this ball is a stable state whereupon Nature will select such a state. The resulting cost is therefore equal to ρ°. The DM cannot decrease the radius of this ball because this will immediately trigger Case 2.

Figure 3.19: The three cases of the relationship between ρ° and $\sigma(q, \tilde{s})$

Suppose that the objective is to select the most stable system, where stability is sought with respect to the entire state space $S = S(A) = S(B)$.

So, which system is more stable globally over $S = S(A) = S(B)$? Is it System A or System B?

It goes without saying that it is impossible to give a meaningful answer to this without a criterion to measure the respective *global* robustness of the systems considered. We can use the *Size Criterion* to this end. But, as we shall see, there are other measures of global robustness.

The important point is that it is vital to take full note of the fundamental difference between *local* and *global* stability/robustness. Meaning that a system that is globally stable/robust is not necessarily locally stable/robust, and vice versa.

3.9 Invariance property of radius of stability models

In subsequent chapters of this book I shall make frequent reference to a property of *Radius of Stability* models that I shall call the *Invariance Property*. However much this property may appear to certain readers as trivially obvious to warrant any comment, I can vouch from my own experience that it most certainly requires a thorough discussion.

The reason that I deemed it necessary to call attention to this property nay, give it a formal statement, was to point out to *info-gap scholars* the profound error of using a *Radius of Stability* type robustness

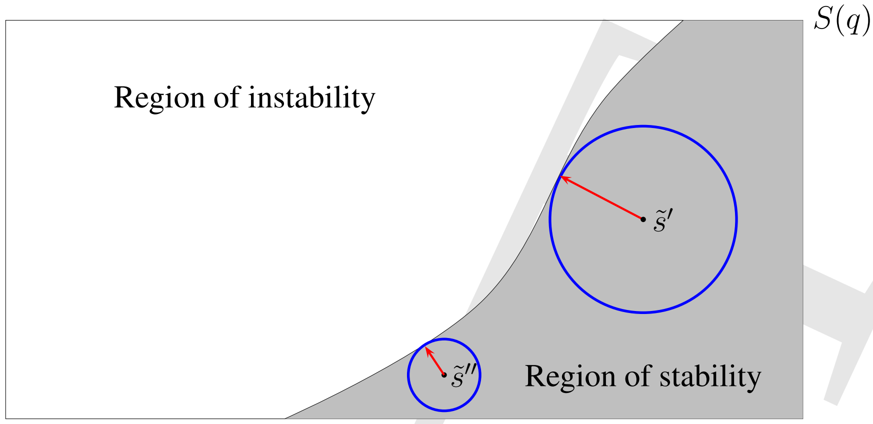


Figure 3.20: Radius of Stability of two nominal states

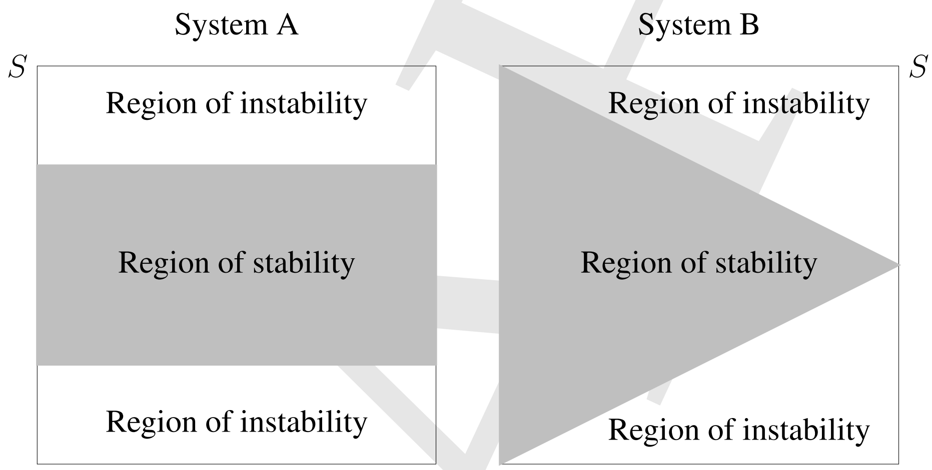


Figure 3.21: Regions of stability of two systems

analysis (read: an *info-gap robustness analysis*) in the pursuit of robustness to severe uncertainty which requires a *global analysis*. The point of the *Invariance Property* is to make vivid the *localness* of this analysis. Namely, to show that regardless of how vast the state space (which is a manifestation of the severity of the uncertainty), the robustness analysis prescribed by a *Radius of Stability* type model is indifferent to indeed, takes no account of this fact. Thus, the *robustness* identified by a *Radius of Stability* model is determined in total disregard to the vast space **outside** the area marked out by the *Radius of Stability*, and is thus surprise surprise! *local*.

However, before I proceed to state the *Invariance Property* formally and prove its validity, I want to illustrate it informally by example. The question that we are going to address then is the following:

> What is the impact of the sheer *size* of the state space $S(q)$ on the *Radius of Stability* of system q?

For example, suppose that you were put in charge of system $q' \in Q$ and were requested to compute its *Radius of Stability*. You determined that this was equal to $\rho(q', \tilde{s}) = 27.89$ and you sent off the results to *Head Office*. Two weeks later you received the following note from Head Office:

> G'day:
>
> We regret that due to circumstances beyond our control, the state space $(S(q'))$ of system q'

actually turned out to be four times larger than anticipated. That is, $S(q')$ is a square of side 10, not 5. The nominal state ($\tilde{s}$) remains unchanged.

Kindly recompute the system's *Radius of Stability* based on this, much larger, state space, and send us the results at your earliest convenience.

We apologize for any inconvenience caused!

Cheers,

Chris

In the best Aussie tradition, you reply

No worries, Chris!

Cheers,

Y.

To organize your thoughts, you call a meeting of the Team to discuss Chris's request. Two hours later the Team meets in the Tea Room. John unfolds on the table a large schematic picture of the situation which he drew on his way to the meeting. It shows the old *Radius of Stability* analysis and the new expanded state space $S'(q)$ (see Figure 3.22).

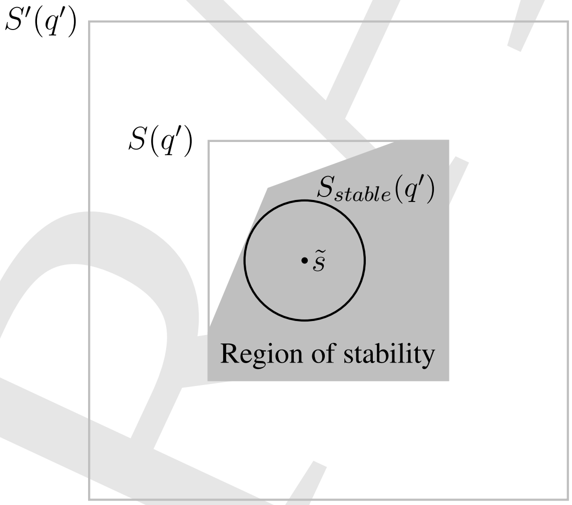


Figure 3.22: John's schematic picture of the situation

John explains that no data is available to determine the extension of the region of stability associated with the expanded state space $S'(q)$. Consequently, it is impossible to determine the new region of stability $S'_{stable}(q)$. In other words, it is impossible to determine the system's new *Radius of Stability* because the data required to determine which new states (that is states in $S'(q)$ that are not in $S(q)$) are stable, is lacking.

It is agreed that a request for the relevant data should be sent off to Head Office (Att. Chris) immediately.

As the meeting is about to break up, Felix suddenly suggests to contact Fred, the internationally known expert on robustness, to seek his advice on this matter. Felix indicates that in one of Fred's recent seminars on *Radius of Stability* models, he mentioned something called "Invariance Property". Felix intimates that he was not fully awake during that seminar, so he cannot remember the details, but . . . he "has a sense" that the "Invariance Property" is very much relevant to the situation in question.

John then indicates that although he did not attend Fred's recent seminar, he downloaded from Fred's website the slides that Fred had used at the seminar, so that he can quickly check the slides to see what this "Invariance Property" is all about.

John dashes off to his office and ten minutes later returns with a neat copy outlining the "Invariance Property" of *Radius of Stability* models. Here is a copy thereof:

> **Invariance Property**:
> Consider a system $q \in Q$ and a stable nominal state $\tilde{s} \in S_{stable}(q)$. Assume that a ball centered at $\tilde{s}$ whose radius is slightly larger than the *Radius of Stability* of system q is a subset of the state space $S(q)$. Then the *Radius of Stability* of the system will not be affected by an increase/decrease in the size of $S(q)$ so long as the modified state space $S'(q)$ contains this slightly larger ball.

John explains that this property implies that in the case of Figure 3.22, there is no need to recompute the *Radius of Stability* to account for the new (larger) state space. Because, the increase in the size of the state space (from $S(q')$ to $S'(q')$) has no impact on the system's *Radius of Stability*.

Felix then asks John to give the "Invariance Property" a more formal exposition and to prove its validity.

John explains.

Suppose that $B(\rho', \tilde{s})$ is a subset of $S(q)$ for some $\rho' > \rho(q, \tilde{s})$, recalling that $\rho(q, \tilde{s})$ denotes the *Radius of Stability* of q at $\tilde{s}$. Then, the *Invariance Property* asserts that the *Radius of Stability* of system q will not be affected by changes in $S(q)$ if these changes are such that the modified state space contains the ball $B(\rho', \tilde{s})$.

John's proof is as follows:

John's formal proof of the Invariance Property:

Assume that for some system $q \in Q$ we have a $\rho' > \rho(q, \tilde{s})$ such that $B(\rho', \tilde{s}) \subseteq S(q)$. By definition of the *Radius of Stability* of system q, we have

$$\rho(q, \tilde{u}) := \max\left\{\rho \geq 0 : s \in S_{stable}(q), \forall s \in B(\rho, \tilde{s})\right\} \tag{3.21}$$

$$= \max\left\{\rho \geq 0 : s \in B(\alpha', \tilde{s}), \forall s \in B(\rho, \tilde{s})\right\},\ \forall \alpha' > \rho(q, \tilde{s}) \tag{3.22}$$

observing that because the balls are nested, for any $\rho' > \rho(q, \tilde{s})$, the ball $B(\alpha', \tilde{s})$ contains at least one unstable state. Hence,

$$\rho(q, \tilde{u}) = \max\left\{\rho \geq 0 : s \in S'(q), \forall s \in B(\rho, \tilde{s})\right\} \tag{3.23}$$

for any state space $S'(q)$ containing $B(\alpha', \tilde{s})$. QED

Back to Chris and Figure 3.22.

The above observations imply that in the case of Figure 3.22 there is no need to recompute the *Radius of Stability* to account for the new (larger) state space. The increase in the size of the state space has no impact on the *Radius of Stability* of the system. So, G'day Chris:

The *Radius of Stability* of system q' is unaffected by the increase in the size of the state space that you specified in your letter.

Cheers,

Y.

To sum up then. The *Invariance Property* brings out that the "size" of the state space is immaterial to determining the value of the *Radius of Stability* of a system. So, if you were to wonder:

> What exactly is the relation between the state space $S(q)$, particularly its "size", and the *Radius of Stability*?

Take note that the state space $S(q)$ is not even a factor in the formal definition of the *Radius of Stability*:

$$\rho(q, \tilde{s}) := \max_{\rho \geq 0} \{\rho : s \in S_{stable}(q), \forall s \in B(\rho, \tilde{s})\} \tag{3.24}$$

Hence, the question: "where is $S(q)$ in all this, and in what way does its 'size' affects the value of $\rho(q, \tilde{s})$?" is given by the following fact:

> Our basic assumption regarding the balls $B(\rho, \tilde{s}), \rho \geq 0$, is that they are all *subsets* of $S(q)$. This means that the definition of these balls must be consistent with this assumption.

The "Invariance Property" thus indicates the obvious: if $B(\rho', \tilde{s}) \subseteq S(q)$ for some $\rho' > \rho(q, \tilde{s})$, then clearly the value of $\rho(q, \tilde{s})$ is independent of $S(q)$ provided that $B(\rho', \tilde{s}) \subseteq S(q)$.

3.10 The No Man's Land Syndrome

The property that is captured by this colorful phrase is intimately related to the "Invariance Property" of *Radius of Stability* models. I decided to state it in these terms when I realized that *info-gap scholars/practitioners* seemed not to appreciate the implications of the "Invariance Property".

My aim was to bring into sharp focus the fact that if $B(\rho', \tilde{s}) \subseteq S(q)$ for some $\rho' > \rho(q, \tilde{s})$, then the value of the *Radius of Stability* of system q is utterly independent of the performance of system q outside the ball $B(\rho', \tilde{s})$. And what better way to do this than to refer to

$$NML(q) := S(q) \setminus B(\rho', \tilde{s}) \tag{3.25}$$

as the *No Man's Land* associated with system q?

In this case as well, the point of the metaphor is to bring out the inherently *local* nature of the robustness defined/mesasured by the *Radius of Stability* model. It also aims to make vivid that as models of local robustness, *Radius of Stability* models, as a matter of principle, lack the capabilities to explore the uncertainty space outside the area determined by the *Radius of Stability*. They are therefore **utterly unsuitable** models for the treatment of severe uncertainty, especially when the severity of the uncertainty is manifested in a vast (indeed, even unbounded) uncertainty space. This is shown in Figure 3.23.

The driving force behind all this was my objective to make it abundantly clear that:

Figure 3.23: No Man's Land Syndrome of *Radius of Stability* models

> Due to their inherently local orientation, *Radius of Stability* models **ought not even be contemplated** as models for robustness against severe uncertainty.

To fully appreciate this argument it is necessary to appreciate the challenges posed by *severe uncertainty*. I discuss this topic in Chapter 4.

3.11 Robustness and Stability

As I have been arguing in this chapter, the *Radius of Stability* of a system relative to its nominal state, can also be understood as a means for determining the *robustness* of the system to changes in the value of this state. In other words, the *Radius of Stability* provides a measure of the robustness of a system in the neighborhood of the nominal state.

In fact, in this book I do not distinguish between *stability* and *robustness*. I am well aware that *stability* and *robustness* can be, indeed are described by means of different verbal definitions. My position is, however, that what counts in this matter is the formal definition of these concepts, not their ordinary language interpretations. For a quick comparison of the various meanings commonly attributed to the terms "robust" and "stable", have a look at Figure 3.24. The lists are taken from the online Encarta Dictionary[5].

This means of course that all my warnings concerning the difference between *local* and *global* stability apply in full to the difference between *local* and *global* robustness. So, as I indicate in this chapter, a failure to recognize this difference (as evidenced in *info-gap decision theory*) can lead to a misapplication of the *Radius of Stability* model, as a means for identifying/obtaining global robustness.

[5] http://encarta.msn.com/encnet/features/dictionary/dictionaryhome.aspx

Robust

1. strong and healthy: strong, healthy, and hardy in constitution
2. strongly constructed: built, constructed, or designed to be sturdy, durable, or hardwearing
3. needing physical strength: involving or requiring great physical strength and stamina Football is a robust sport.
4. determined: characterized by firmness and determination and a refusal to make concessions a robust defense
5. straightforward: showing clear thought and common sense
6. blunt or crude: rough and direct or crude
7. full-flavored: rich, strong-tasting, and full-bodied
8. system that is able to recover from unexpected conditions during operation

[Mid-16th century. Latin robustus "made of oak, hard, strong" robur "oak tree, hardness, strength"]

Stable

1. not changing: steady and not liable to change Prices have remained stable.
2. not likely to move: steady or firm and not liable to move
3. not excitable: having a calm and steady temperament, rather than being excitable or given to apparently irrational behavior
4. chemistry physics not readily undergoing change: not subject to changes in chemical or physical properties
5. physics not naturally radioactive: incapable of becoming a different isotope or element by radioactive decay

[13th century. Via Anglo-Norman and Old French< Latin stabilis]

Figure 3.24: Definitions of “Robust” and “Stable”. Slightly modified versions of the definitions given by the online Encarta Dictionary

3.12 A Storm in a Tea Cup?

To round out this discussion, I want to take up again my decision to devote a whole chapter to the *Radius of Stability*. I want to give a cogent explanation to my decision to discuss at such great length and at this level of detail a model prescribing a local stability/robustness analysis when my concern in this book is:

> *Determine the robustness of a system against severe uncertainty in the true value of the parameter of interest.*

After all, the local analysis prescribed by *Radius of Stability* models is conducted in the neighborhood of a given *nominal* value of s. So, the fundamental question that such a prescription for the treatment of uncertainty from the start fails to address, worse misrepresents, is:

> *What is the relationship between the given nominal value of s, that is the value of $\tilde{s}$, and the (unknown) true value of s?*

Because, to suggest that the local measure of stability prescribed by the *Radius of Stability* is a suitable measure of stability against severe uncertainty amounts to suggesting that the results yielded for a *nominal* value apply to the *true* value of s. And, this is akin to contending that *local optimization* and *global optimization* are one and the same thing to further imply that *a local optimum* is *a global optimum*.

I need not pursue this analogy any further to bring out the error in applying a local stability analysis of the type discussed above to determine the robustness of a system against severe uncertainty in the true value of the parameter of interest.

Of course, this does not mean that such a local analysis is without its merit. There are clearly many situations where our interest is precisely in local stability with respect to perturbations in a *nominal* value of the parameter of interest. Indeed, this is a central concern in fields ranging from numerical analysis, applied mathematics, parametric optimization, economics, to control theory, etc.

But the point remains that however useful a tool it is, local stability analysis is the wrong tool for the modeling, analysis and management of severe uncertainty.

G'day Moshe:

In my humble opinion this argument does not make a strong enough case for the inclusion in this book of a whole chapter on the *Radius of Stability*.

I realize that you plan to come back to this issue in subsequent chapters of the book. Still it is important to make your case stronger right from the start.

Cheers,
Fred
Internationally Renowned Expert on Robust Decision-Making in the Face of Severe Uncertainty

Point taken, Fred.

The point is this.

As I indicate at the end of Chapter 2, in my explanation of the phrase "fooled by robustness" in this book's title, my motivation for writing this book, was the fact that *Radius of Stability* models are advocated in a number of disciplines for the management of severe uncertainty. Proponents of this approach understand the uncertainty to be characterized by a vast uncertainty space, a poor estimate and a likelihood-free quantification of uncertainty.

Not only are proponents of this approach certain that such models are the right models for this purpose, that the results that they obtain with these models are robust to (severe) uncertainty, but they are convinced that this type of approach ought to be accepted as the standard in certain fields (for instance, *applied ecology)*.

So, one of the objectives of this lengthy, detailed examination of the *Radius of Stability* was to prepare the ground for my discussions in subsequent chapters where I intend to examine in greater detail the errors of this approach.

After all, an important element in the recognition of how misguided this approach is, namely how unsuitable for the treatment of severe uncertainty it is, is a good appreciation of the operation and role of *Radius of Stability* models, and of course the challenges posed by severe uncertainty.

G’day Reader:

I thought it necessary to intervene at this point because I suspect that some readers may find Moshe’s explanation to be somewhat contrived.

It is indeed difficult to comprehend that a *Radius of Stability* model should be proposed as the core ingredient of a methodology for decision under **severe** uncertainty where the uncertainty is characterized by a poor estimate and a vast likelihood-free uncertainty space.

I suggest, though, that you take Moshe at his word, because as the record shows, this is precisely the methodology that is advocated in numerous publications, in a number of disciplines, as a sound reliable methodology for the treatment of severe uncertainty. To lend support to his argument I note that this methodology prescribes a local robustness analysis in the neighborhood of a poor estimate, the neighborhood being infinitesimally small compared to the uncertainty space under consideration.

I am sure that Moshe will address these points in due course. For my part, I thought that all this had to be made clear from the start.

Cheers,
Fred
Internationally Renowned Expert on Robust Decision-Making in the Face of Severe uncertainty

So let us go back to where we left off.

It is no doubt true that the *Radius of Stability* model is widely recognized for what is it: a model of *local robustness*. That is, this model is generally used to do what it was designed to do, namely, to model and analyze the robustness of systems to **small perturbations** in a given nominal value of the parameter of interest.

This being the case, I suspect that some readers may accuse me of creating a storm in a tea cup.

Indeed, when I launched my Campaign[6] at the end of 2006, some of my colleagues argued as much. “Isn’t it well known" they contended, “that models of local robustness of this type are utterly unsuitable for the treatment of severe uncertainty? So, what if this is advocated by a small group of scholars who do not see the profound error in using a model of local robustness as a tool for the treatment of severe uncertainty that is characterized by a poor estimate, a vast uncertainty space and a likelihood-free quantification of uncertainty?”

[6]My campaign to contain the spread of info-gap decision theory in Australia.

"After all," the argument continued, "only one theory, namely *info-gap decision theory*, advances this untenable proposition, and what is more, this theory is used mostly by scholars who are not well-versed in classical decision theory and robust optimization."

I shall not elaborate in this book on the reasons that at that time prompted me to launch my Campaign except to point out that four years later I am satisfied that my decision was correct. I plan to discuss this issue in detail in my next book on this topic[7].

More to the point, however, I want to reiterate the reasoning behind my decision to devote a whole chapter to a detailed analysis of the *Radius of Stability* as a precursor to later discussions on its misapplication as a tool for the treatment of severe uncertainty (read its application by *info-gap decision theory*). To put across the full dimensions of the error of applying the *Radius of Stability* models to the management of severe uncertainty requires a thorough clarification of both the notion of the *Radius of Stability* and the notion of *severe uncertainty*.

So, having completed the clarification of the first, I can now turn to an examination of the notion of *severe uncertainty*.

3.13 Bibliographic Notes

There seems to be a general consensus in the area of *control theory* that the concept *Radius of Stability* was invented in the 1980s by Hinrichsen and Pritchard (1986a, 1986b).

A quick search of the literature suggests otherwise.

The earliest reference to the concept *Radius of Stability* that I managed to find (so far) dates back to the early 1960s. For example,

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

- 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.

In Milne and Reynolds (1962, p. 62) we read:

> It is convenient to use the term "radius of stability of a formula" for the radius of the largest circle with center at the origin in the s-plane inside which the formula remains stable.

The term was apparently coined independently by Hinrichsen and Pritchard (1986a, 1986b) in the field of *control theory.* So, according to Paice and Wirth (1998, p. 289):

> Robustness analysis has played a prominent role in the theory of linear systems. In particular the state-state approach via stability radii has received considerable attention, see [HP2], [HP3], and references therein. In this approach a perturbation structure is defined for a realization of the system, and the robustness of the system is identified with the norm of the smallest destabilizing perturbation. In recent years there has been a great deal of work done on extending these results to more general perturbation classes, see, for example, the survey paper [PD], and for recent results on stability radii with respect to real perturbations...

[7] See the website of the book: `rise-and-rise.moshe-online.com`

where HP2 = Hinrichsen and Pritchard (1990), HP3 = Hinrichsen and Pritchard (1992) and PD= Packard and Doyle (1993).

As for the role of *Radius of Stability* in *optimization* and *mathematical programming*, consider Zlobec's (1988, p. 129) statement:

> An important concept in the study of regions of stability is the "radius of stability" (e.g., [65, 89]). This is the radius r of the largest open sphere $S(\theta^*, r)$, centered at θ^*, with the property that the model is stable, at every point θ in $S(\theta^*, r)$. Knoweldge of this radius is important, because it tells us how far one can uniformly strain the system before it begins to "break down". (In an electrical power system, the latter may manifest in a sudden loss of power, or a short circuit, due to a too high consumer demand for energy. Our approach to optimality, via regions of stability, may also help understand the puzzling phenomenon of voltage collapse in electrical networks described, e.g., in [11].)

where [65] = Petric and Zlobec (1983), [89]= Zlobec (1987), and [11] = Carpentier et al (1984).

In the first edition of the *Encyclopedia of Optimization,* Zlobec (2001) describes the *Radius of stability* as follows:

> The radius of the largest ball centered at θ^*, with the property that the model is stable at its every interior point θ, is the radius of stability at θ^*, e.g, [69]. It is a measure of how much the system can be uniformly strained from θ^* before it starts breaking down.

where [69] = Zlobec (1988).

In accounting (Raab and Feroz, 2007, p. 400):

> Charnes, Haag, Jaska, and Semple (1992), Charnes, Rousseau, and Semple (1996) and Seiford and Zhu (1998) developed a sensitivity analysis technique based on the infinitynorm measure of a vector. This technique defines the necessary simultaneous perturbations to the component vector of a given NG that cause it to move to a state of "virtual" efficiency. Virtual efficiency is defined as a point on the efficient frontier where any miniscule detrimental perturbation (increase in inputs and/or decrease in outputs) will cause an efficient NG to become inefficient, or any miniscule favorable perturbation (decrease in inputs and/or increase in outputs) will cause an inefficient NG to become efficient.
>
> For an efficient NG, the infinity-norm measure, or the radius of stability (herein termed stability index), defines the largest "cell" in which all simultaneous detrimental perturbations to the input and output components will not cause a change in the efficiency status from technically efficient to inefficient. As such, the larger the stability index, the more robustly efficient the NG is said to be. Those efficient NGs with small stability indices will thus become technically inefficient, with smaller detrimental perturbations than those efficient NGs with larger stability indices.

NG = national government.

And here are the abstracts of two recent recent papers where the *Radius of Stability* plays a central role. First the paper "Finite cooperative games: parametrisation of the concept of equilibrium (from Pareto to Nash) and stability of the efficient situation in the Hölder metric" by Emelichev and Karlkina (2009, p. 229):

> We consider a finite cooperative game of several players with parametric principle of optimality such that the relations between players in a coalition are based on the Pareto maximum. The introduction of this principle allows us to find a link between such classical concepts as the Pareto

> optimality and the Nash equilibrium. We carry out a quantitative analysis of the stability of the game situation which is optimal for the given partition method with respect to perturbations of parameters of the payoff functions in the space with the Hölder l_p-metric, $1 \leq p \leq \infty$. We obtain a formula for the radius of stability for such situation, so we are able to point out the limiting level for perturbations of the game parameters such that the optimality of the situation is preserved.

Next, the paper entitled "Sensitivity and Stability Analysis in DEA on Interval Data by Using MOLP Methods" by Beigi et al. (2009, p. 891):

> In this paper, we suppose a method for analyzing sensitivity and stability of all the decision making units, while inputs and outputs are interval data. Therefore, for estimating radius of stability of a DMU; firstly, we classify the decision making units then we obtain the radius of stability for each classification. For analyzing the sensitivity and estimating the radius of stability analogous of each DMU, a MOLP is defined. Therefore, the interactive methods are used for finding the efficient solution in which the comment of Decision Maker is important. At the end numerical example has been solved by using the weighted-sums of the target function and also the interactive method (STEM) in MOLP problems.

Regarding the *Radius of Instability*.

For obvious reasons, the literature on this concept is much less extensive. The earliest reference to this concept that I am aware of is found in articles by Salvadori (1991) and Salvadori and Visentin (1992). The same definitions are discussed in Anderson and Bernfeld (2001).

In Ediev (2002, p. 818) we read:

> Then the circle $\Lambda = \{\lambda : |\lambda| \leq \lambda^{min}\}$ exists, such that the reproduction model possesses the ergodic property if and only if the reproduction coefficient doesn't belong to that circle. The circle mentioned will be called as an *instability circle,* and its radius will be called as an *instability radius.*

Here λ represents a numeric scalar coefficient of a demographic model. Hence, λ can be regarded as a perturbation of the nominal value $\tilde{\lambda} = 0$.

The definition given by Kraev and Fursov (2005, p. 1585) is more in line with the abstract definition used in this chapter:

> The main results of these papers are estimates of the distance from a given unstable polynomial to the nearest stable polynomial in some metric on the parameter space (the instability radius).

In symbols:

$$\iota(a) := \inf_{b \in S} ||a - b|| \tag{3.26}$$

where S denotes the set of stable polynomials, a is a given unstable polynomial and ||a-b|| denotes the distance between a and b. This is indeed the formal definition given in Kraev and Fursov (2007, p. 380) which is a translation of Kraev and Fursov (2004).

Postscript

It is interesting to note that Deines et al. (2007) have missed the connection between *info-gap robustness* and *Radius of stability.* Their discussion refers to a number of *info-gap publications* and, separately, to the concept *stability radius,* but it does not make the connection between the two. What a pity!

3.13.1 Variations on a theme

I call attention to the fact that the concept *Radius of Stability* is so basic that it is invoked in many disciplines. Its formal definition can therefore vary slightly depending on the mathematical/physical framework in which it is used. So, the definitions given above, for instance

$$\rho(q,\tilde{s}) := \max\left\{\rho \geq 0 : s \in S_{stable}(q), \forall s \in B(\rho,\tilde{s})\right\} \tag{3.27}$$

are just a few of the many variations on the same theme.

In applied mathematics the *Radius of Stability* is often defined explicitly in terms of the perturbation in the nominal value $\tilde{s}$. For instance,

$$\rho(\tilde{s},S,\Delta) := \sup_{\delta\in\Delta}\left\{||\delta|| : \tilde{s}+\delta \in S\right\} \tag{3.28}$$

where

$$\begin{aligned} \Delta = &\ \textit{set of admissible perturbations}\\ ||\delta|| = &\ \textit{magnitude of perturbation } \delta \in \Delta\\ S = &\ \textit{region of stability under consideration} \end{aligned}$$

According to this definition, the *Radius of Stability* at $\tilde{s}$ is the magnitude of the largest perturbation in $\tilde{s}$ that will keep the state in the region of stability of S. The incorporation of Δ in the definition of the stability region enables the modeler to control the “size” and “direction” of the perturbations.

There are cases where it is more convenient to define the radius of stability as follows:

$$\rho'(\tilde{s},S,\Delta) := \inf_{\delta\in\Delta}\left\{||\delta|| : \tilde{s}+\delta \notin S\right\} \tag{3.29}$$

So according to this definition, the *Radius of Stability* $\tilde{s}$ is the magnitude of the “smallest” perturbation in $\tilde{s}$ that will destabilize the system.

3.13.2 Variations on a title

It should be stressed that because the concept *Radius of Stability* is so intuitive, yet abstract, it appears in the literature under other titles, that is, titles that describe the application under consideration.

For instance, in Floréen and Orponen (1993, p. 813) it is called *Attraction Radius,* in Quirk et al. (2005) and Barmish et al. (2009. p. 974) it is called *Radius of Robustness* and in and Shcherbakov and Topunov (2008, p. 11386) it is called *Radius of Sign-Definiteness.*

And in the first example in this chapter, I called it *Radius of Happiness.* Perhaps I should have called it *Rex's Radius of Happiness*.

Remark:

One would be hard pressed to find so much as a single reference to the fundamental concept *Radius of Stability* in the entire body of literature on *info-gap decision theory*.

This of course is most surprising given that *info-gap's robustness model* is nothing less than a simple *Radius of stability* model.

More on this in the second part of the book.

Chapter 4

Severe Uncertainty

4.1 Introduction

Severe uncertainty has, of course, been a subject of reflection from time immemorial. Thus, the intellectual conundrums presented by this concept have for millennia exercised minds in many cultures. In modern times, the study of severe uncertainty has to a great extent been driven by the aspiration to identify ways of coping with uncertainty. Namely, the objective has been to identify (mathematical) methods capable of implementation in our socio/political/economic systems as a means of bolstering these systems against the unexpected, the unforeseen, the unpredictable.

To set the stage, consider this illuminating portrayal of the challenges posed by severe uncertainty from *Don Quixote*. The text in the square brackets in mine:

> [Inn keeper:] "He also has with him a monkey with the rarest talent ever seen among monkeys or imagined among men, because if he's asked something, he pays attention to what he's asked then jumps onto his master's shoulders and goes up to his ear and tells him the answer to the question, and then Master Pedro says what it is; he has much to say about past things than about future ones, and even though he isn't right all the time, he is not wrong most of the time, so he makes us think he has the devil in his body."
>
> ⋮
>
> [Don Quixote:] "Senior Soothsayer, can your grace tell me *che pesce pigliamo?* What will become of us? ..."
>
> [Soothsayer:] "Senior, this animal does not respond or give information about things to come; about past things he knows a little, and about present ones, a little more."
>
> "By god", said Sancho, "I wouldn't pay anything to have somebody tell me what's already happened to me! Who knows that better than me? And it would be foolish to pay anybody to tell me what I already know; but since he knows about present things, here's my two *reales* so His Monkeyness can tell me what my wife, Teresa Panza, is doing now, and how she's spending her time."
>
> Cervantes (2003, p. 624)

And to see how close to the mark this portrayal of severe uncertainty indeed is, consider the following quote, taken from a comment by Bernard Dixon that was published in the *New Scientist and Science*

Journal on September 16, 1971. The reference is to a lecture given by Prof. Paul Ehrlich (Stanford University) at the Institute of Biology (UK) in 1969:

> The audience loved it and gasped for more. For an hour they listened, enraptured, to dire warnings about death of all sea-food, a drastic fall in imports, international tensions, and "rocketing" death rates. "If I were a gambler", Professor Ehrlich concluded before boarding his plane, "I would take even money that England will not exist in the year 2000." He then flew off into the night, leaving behind yards of exciting headlines, a clutch of memorable phrases ("siren song of the myopic optimists", "tidal wave sweeping the country", "political turmoil will rule" ...), and the gratitude of editors who had used Ehrlich to keep the pot boiling between the Liberal and Labour party conferences.

So, before I begin my discussion on this topic, I want to make it abundantly clear that I am going to tread very carefully in this chapter.

I am not going to join the fray by adding my two pennies worth on the question of what is the meaning and nature of severe uncertainty. Nor am I going to venture into the veritable minefield of the philosophical debate on whether formal mathematical models of the type currently used by "experts in the field" are **indeed** capable of furnishing practical tools for the treatment and management of severe uncertainty.

For an idea of two diametrically opposed positions on our ability to handle severe uncertainty, I refer the reader to Nassim Taleb's two books on uncertainty *Fooled by Randomness: the Hidden Role of Chance in Life and in the Markets* and *Black Swan: the Impact of the Highly Improbable* and to Bruce de Mesquita's book *The Predictioner's game: Using the Logic of Brazen Self-Interest to See and Shape the Future*.

My objectives in this book are immeasurably more modest.

Thus, my main objective in this chapter is to examine a number of conceptual, methodological and technical issues that a modeler/analyst would face when confronted with a problem whose key parameter (parameters) is (are) subject to severe uncertainty. I am going to take it as a given that it has been a long-standing practice in many areas of expertise (operations research, economics, optimization etc.) to formulate such problems in terms of abstract models of uncertainty. Furthermore, that the models generally used for this purpose have their roots in *classical decision theory*.

The first point that I must therefore make clear in this regard is that the defining feature of the models that I am going to discuss in this book is that they are *non-probabilistic* and likelihood-free. So, strange though it may sound, I am going to discuss here only models that are *likelihood-free,* also, *fuzzy-free,* and so on.

This fact must not be (mis) construed as suggesting that probability theory has no place in the modeling, analysis and management of severe uncertainty.

Of course not!

Rather, it is a reflection of the general orientation of the discussion in this book. Namely, its being rooted firmly in the traditions of *classical decision theory* and in the state of the art in the field of *robust optimization.*

Second, it is obviously extremely difficult to avoid all discussion on the concept of *severe uncertainty* in a chapter whose central concern is the question of how to deal with problems that are subject to *severe uncertainty*. Meaning that I cannot justifiably refrain from devoting at least some attention to the interpretations of this concept in the literatures on decision-making and related areas. However, given the proliferation of positions on this concept in these literatures, and given that many come laden with babble and excessive rhetoric, I shall have to be extremely selective in my choice. I shall therefore consider only those positions on this concept that have come to be regarded as mainstays in these fields.

For the most part, I shall endeavor to stir the discussion in a direction that would seek to be as rhetoric-free, hence as technical as possible. And to get started, I begin with a clear statement of a model that I shall treat as a framework giving expression to a problem subject to *severe uncertainty.*

4.2 Generic model of severe uncertainty

As I indicated in the preceding chapters, a modeler/analyst presented with a problem where a key element (namely a parameter of interest) is unknown, indeed is subject to severe uncertainty, might or would describe the problem in question by means of an uncertainty model consisting of these two basic ingredients:

$$\begin{aligned} \mathscr{U} &= \text{ set of possible/plausible values of the parameter of interest, call it } u. \\ \tilde{u} &= \text{ point estimate of the true value of } u. \end{aligned}$$

Before I can proceed, the following comments are in order. First, the model that I have just outlined is not the only model that one would be able to formulate for an uncertainty problem. Hence, it is only one of a number of possible uncertainty models that I shall discuss in this book.

Second, this formulation puts forward an extremely simple, indeed basic, model of uncertainty. All that this model states is that $\mathscr{U}$ signifies the full range of possible/plausible values of the parameter of interest whose true value is subject to severe uncertainty. And that $\tilde{u}$ signifies our point estimate of the true value of this parameter.

The point is then that there is nothing in its present formulation to indicate that this model does indeed represent a situation where the uncertainty is *severe* rather than say, VERY MILD. This means that to enable it to designate a situation of SEVERE uncertainty, $\mathscr{U}$ and $\tilde{u}$ must have certain attributes that will aptly express characteristics that we would associate with states of affairs where the uncertainty is SEVERE, rather than say, VERY MILD.

4.2.1 Working assumptions

One way to characterize the uncertainty of the uncertainty model $(\mathscr{U}, \tilde{u})$ as SEVERE is to require that this pair satisfy the following working assumptions:

- WA-1: The uncertainty space $\mathscr{U}$ is (relatively) vast (e.g. can be unbounded).
- WA-2: The estimate $\tilde{u}$ is a poor indication of the true value of u, indeed it is likely to be substantially wrong.
- WA-3: The uncertainty model is non-probabilistic and likelihood-free.

I imagine that the reader appreciates that it is pointless to attempt to work out precise formulations of a “(relatively) vast" uncertainty space, or a “poor, likely to be substantially wrong" estimate. I trust, though, that my discussion in the next few sections will inject these notions with sufficient content to render them meaningful.

So, let us now examine the key ingredients of the uncertainty model in view of these working assumptions.

4.2.2 Parameter of interest, u

This element of the uncertainty model represents a key parameter in the system under consideration. Key in the sense that the solution to the problem faced by the decision-maker depends on the value of this parameter. That is, the output of the model depicting this system is contingent on the value of this parameter.

Methodologically, u is an abstract entity. It can be anything one wants or needs it to be. In practice, it would typically be a mathematical object such as a numeric scalar, a vector, a matrix, a function, and so on.

For example, u can be a numeric vector $u = (u_1, u_2, u_3)$ denoting the prices of three commodities, say gold, oil, and dark chocolate, on March 4, 2016. Or it can be say, a numeric scalar representing the number of koalas in the wild in 2020.

The whole point about this parameter is that its "true" value is unknown as it is subject to severe uncertainty. But, as indicated in the next sub-section, this does not mean that our ignorance about the true value of u is such that we are stumped by it. So, for instance, we shall not consider situations where

$$u = \text{ winner of the Wimbledon Men's Single Title in 2041}.$$

4.2.3 Uncertainty space, $\mathscr{U}$

To continue this thought. Although we proceed from the fact that the true value of u is subject to *severe* uncertainty, our basic premiss about it is that the set of possible/plausible values of this true value is KNOWN. We call this set the *uncertainty space* and we denote it by $\mathscr{U}$.

The immediate implication of this is that we take the uncertainty model $(\mathscr{U}, \tilde{u})$ to be an instrument for dealing with severe uncertainty. But, there is no presumption here that this model is capable of handling "unknown unknowns", namely those "unknowable entities", or "unknowable contingencies", or what have you, made famous in 2002, by the then US Defense Secretary, Donald Rumsfeld:

> There are known knowns; there are things we know that we know. There are known unknowns; that is to say, there are things that we now know we don't know. But there are also unknown unknowns; there are things we do not know we don't know.

I shall briefly discuss the difficulties posed by these elusive entities in the second part of the book in connection with my critique of *info-gap decision theory.*

In sum, the basic idea put forward by this utterly conventional definition of the uncertainty space $\mathscr{U}$ is that it is a mathematical object taken to comprise all the possible/plausible values of the parameter u, including the true value, where the **big unknown** is which element of $\mathscr{U}$ is the true value of u.

And to illustrate the relation between the unknown parameter u and the uncertainty space $\mathscr{U}$, suppose that

$$u = \text{number of koalas in the wild on March 4, 2016.}$$

The question is then: what would be the corresponding uncertainty space $\mathscr{U}$?

For an idea of how we might pin down a "proper" value for $\mathscr{U}$, consider the following exercise, the aim of which is to determine the value of $\mathscr{U}$ within say 15 minutes.

A good starting point could be the website of the *Australian Koala Foundation* (AKF)[1] where a quick search (November 21, 2010) would yield this information (color added):

[1] https://www.savethekoala.com

1. How endangered is the koala? Is it at risk of extinction? How many koalas remain? The Australian Koala Foundation's (AKF) research indicates that the koala is in trouble and that extinctions of local populations have already occurred. In contrast to the millions of koalas which were thought to be present at the time of European settlement, the AKF believes that there could be less than 80,000 remaining today, possibly as few as 43,000. If this rate of decline continues then yes, the koala is at risk of extinction.

It is clear from this statement that the uncertainty regarding the size of the *current population* of koalas is indeed *severe*. So, how are we going to decide on a value for the uncertainty space such that it will "adequately" give expression to the size of the population in 2016?

To do this we would have to take into account what are manifestly two conflicting considerations:

- On the one hand, it is clear that we would seek to ensure that $\mathscr{U}$ contains the (unknown) true value of u. This suggests that we would attempt to play it safe by assigning $\mathscr{U}$ a value large "enough" to include this element.
- But, on the other hand, it is equally clear that we would strive to exclude from consideration implausible/impossible values of u. This suggests in turn that we would desire that $\mathscr{U}$ be as small as "possible".

So, how are we going to determine $\mathscr{U}$ given these considerations?

Obviously, there is nothing to stop us from setting $\mathscr{U} = \{0, 1, 2, \dots, 200000\}$, in which case we would most assuredly be "on the safe side". But is this a good specification of the uncertainty space under consideration?

And how about $\mathscr{U} = \{20000, \dots, 100000\}$?

Clearly, we can go on in this fashion ad infinitum, which means of course that an answer to this question can be given only in *context*. By this I mean that we must never lose sight of the fact that the specification of $\mathscr{U}$ is not an academic exercise. The value of $\mathscr{U}$ would normally be determined on grounds of the objectives and requirements of the application under consideration. In some applications a rough approximation will be more than adequate, in others it can be very useful to have a good approximation, and in some applications it could be vital to have a very good approximation.

For example, in the case of the koalas problem, in some applications the very rough $\mathscr{U} = \{0, 1, 2, \dots, 200000\}$ may well be adequate, perhaps even $\mathscr{U} = \{0, 1, 2, \dots, \infty\}$.

An important factor that would come into play here is: to what extent do the results of the analysis depend on the value assigned to $\mathscr{U}$? I call attention to this point because, as we saw in Chapter 3, there are cases (models of robustness) where the results of the analysis are independent of the size of $\mathscr{U}$ (within reason).

And talking about **size**.

As we saw above, the severity of the uncertainty in the true value of u is manifested not only in the poor quality of the estimate, but also in vastness of the uncertainty space. This means that in theory — all other things being equal — the vaster the uncertainty space, the more *severe* the uncertainty. For instance, consider the situation in Figure 4.1 where two uncertainty spaces are shown for the same system. Note that the two uncertainty spaces have the same estimates, $\tilde{u}' = \tilde{u}'' = \tilde{u}$, and that these are of the same quality, both are say, *wild guesses*. The difference is in the uncertainty spaces: $\mathscr{U}''$ is manifestly much larger than $\mathscr{U}'$.

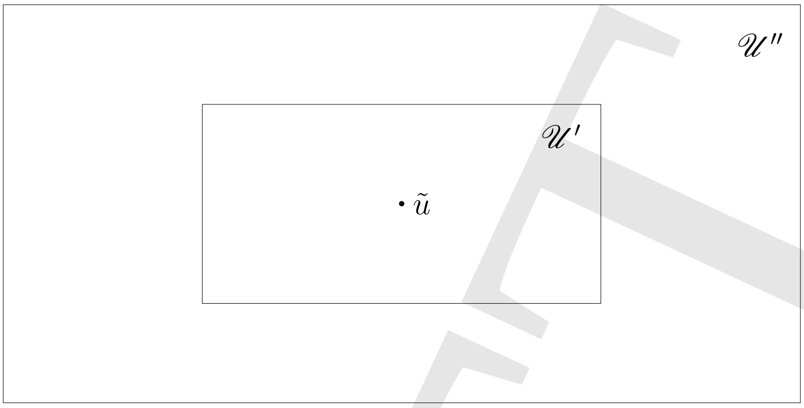


Figure 4.1: Two uncertainty spaces for the same system

This, by the way, may represent a situation where $\mathscr{U}'$ is an initial approximation of the uncertainty space that figures in a preliminary analysis of the problem, whereas $\mathscr{U}''$ is the uncertainty space in the final analysis. The change occurred as additional information about the problem became available.

Clearly, the uncertainty associated with $\mathscr{U}''$ is *more severe* than the uncertainty associated with $\mathscr{U}'$.

So far so good.

But what inferences should we draw from this depiction of the notion of *severe uncertainty*? Are we to conclude that because it is thoroughly sensible to propose that a vast, indeed unbounded, uncertainty space gives apt expression to the notion of *severe uncertainty* that this actually provides us with the wherewithal to handle any problem imaginable? Consider for instance the case where the task is:

u = prime minister of Australia in 2060.

What specification are we going to give the uncertainty space $\mathscr{U}$ if we are to give a sensible representation to this parameter? That is, how are we going to determine what uncertainty space would reasonably cover all the possible/plausible value of u?

And if this were not enough, consider an even more challenging problem:

u = president of France in 2080.

What specification should we give this uncertainty space $\mathscr{U}$ to properly reflect this parameter? How would we determine what space would cover all the possible/plausible value of u in this case?

The point in bringing up these examples is to make it clear that although the three working assumptions listed above regarding underlying $(\mathscr{U}, \tilde{u})$ methodologically remove any limiting restrictions from the parameter u, one does not have *carte blanche* to handle with this model just about any “wild" severe uncertainty that may strike one’s fancy. A sound application of this model implies that the severity of the uncertainty capable of being handled by the model $(\mathscr{U}, \tilde{u})$ is clearly limited to our ability to spell out the possible/plausible values of the parameter of interest.

4.2.4 Estimate, $\tilde{u}$

There is clearly no point in embarking on a linguistic analysis aimed at flushing out the meaning of “poor” and “substantially wrong” in the statement of WA-2, as it is hard to imagine a consensus on this matter. It therefore seems that the most obvious way to give meaning to the idea that the estimate $\tilde{u}$ is taken to be “poor", “doubtful", “highly questionable”, etc. is to characterize it as follows.

> We posit that there are no grounds to assume that the true value of u is more/less likely to be in the neighborhood of the estimate $\tilde{u}$ than in the neighborhood of any other element of $\mathscr{U}$.

As a matter of fact, there is hardly any choice in this matter.

Because, this characterization is in fact dictated by WA-3, which essentially implies that we have not the slightest inkling as to the "location" of the true value of u in the uncertainty space $\mathscr{U}$. This means, among other things, that the estimate $\tilde{u}$ is, as a matter of principle, rendered indistinguishable from any other element of $\mathscr{U}$.

To further clarify this point, consider the situation depicted in Figure 4.2 where the rectangle represents the uncertainty space $\mathscr{U}$ and u' and u'' are two arbitrary values of $u \in \mathscr{U}$.

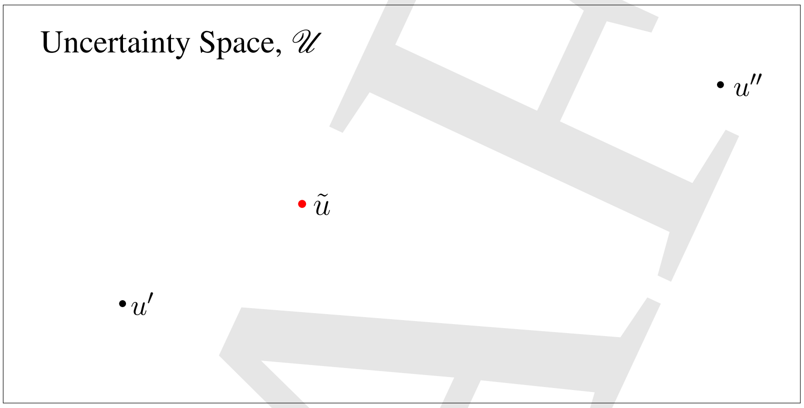


Figure 4.2: Severe uncertainty

As the uncertainty model is *likelihood-free,* there are no grounds to assume that the true value of u is more/less likely to be in any one particular neighborhood of $\mathscr{U}$. Hence, there are no grounds to assume that the true value of u is more/less likely to be in the neighborhood of $u = u'$ rather than in the neighborhood of $u = u''$. Therefore, there are no grounds to assume that the true value of u is more/less likely to be in the neighborhood of the estimate $\tilde{u}$ rather than the neighborhood of any other $u \in \mathscr{U}$.

But having said all that, it is clearly incumbent on me to explain the role of the estimate $\tilde{u}$ in this setup.

Because, it may well be argued that rather than allaying the misgivings about the ability to sensibly incorporate an estimate in this type of model, this characterization of the estimate $\tilde{u}$ in effect raises an even more vexing question:

> Doesn't this imply then that a methodological incongruity exists between WA-2 and WA-3a, in which case the function of the estimate $\tilde{u}$ in this framework is rendered ambiguous?

For according to WA-3, there is nothing in the model to tell the estimate $\tilde{u}$ apart from other elements of $\mathscr{U}$, other than its being given the epithet "estimate" and its being denoted symbolically by $\tilde{}$.

What is more, there is nothing in this setup to provide any clue as to how to "quantify" or even to interpret this "entity" called "estimate".

In particular, in the framework of the uncertainty model under consideration, the "estimate" is **not** even intended to represent the *most likely* true value of u, indeed it must not be (mis) interpreted as such.

And yet, for all that, an element of $\mathscr{U}$ which is not ascribed any properties to distinguish it from other elements of $\mathscr{U}$, is clearly being singled out in the model by virtue of its being treated as an "estimate".

I therefore want to make it clear that this is precisely how I shall treat the estimate in this book. I shall not attribute it any special properties, functions, or responsibilities. That is, other than assigning it its appellation and designating it by $\tilde{\ }$, it will be indistinguishable from all other elements of $\mathscr{U}$.

G'day Reader:

I find this explanation/excuse for incorporating an estimate in the uncertainty model, that in the end turns out to be ambiguous even self-contradictory, totally unsatisfactory.

I prefer to call a spade a spade. I maintain that as the uncertainty — as depicted above — is severe, we simply cannot claim to have an estimate of the true value of u.

This position is in line with the treatment of "severe uncertainty" in *classical decision theory* and *robust optimization*.

Cheers,
Fred
Internationally Known Expert on Robust Decision-Making Under Severe Uncertainty

Although Fred brings up an extremely important point, I beg to differ with him on this issue. Let me explain why.

I have yet to come across a "real-world" situation where it is impossible to come up with an "estimate" for the true value of the parameter of interest. In fact, picking an estimate is the least of our trouble as invariably this is no more than a trivial task. Incidentally, should you need advice on this matter consider this formula[2]

Wet your index finger and put it in the air. Think of a number and double it.

What is problematic here is identifying a "sensible", or "reliable", estimate. In other words, the issue here is not the *existence* of an estimate, the issue is the *quality* of the estimate. This is where our trouble lies.

One might argue, therefore, that WA-2 should be able to take care of this difficulty, the logic behind it being that if the uncertainty is indeed "severe", then it would be inconsistent to assume that the estimate is "good" and "reliable". Conversely, if the estimate is "not poor" and is unlikely to be "substantially wrong", then there is no sense in assuming that the uncertainty "severe".

However, the counter argument to this proposition might be that it is not simply a question of the *quality* of the estimate that is problematic here, but rather the *status* of an estimate, any estimate, in a *likelihood-free* model of uncertainty.

In other words, any estimate of any quality would seem to postulate some sort of "likelihood" structure about the uncertainty space $\mathscr{U}$. So, in one way or another, the quality of the estimate would have to be reflected in an assumed likelihood structure (or a substitute thereof), for the sake of consistency.

For example, consider a situation where the estimate is "excellent", yet the uncertainty space is vast and the uncertainty model is likelihood-free. Are we to assume that (according to WA-3) there are no grounds to hold that the true value of u is more/less likely to be in the neighborhood of the estimate than in the neighborhood of any other $u \in \mathscr{U}$? If so, in what sense is the estimate "excellent"?

All this goes to show that although there is no doubt that "one can always come up with an estimate", in the context of severe uncertainty the whole issue is fraught with difficulties. Still, whether one accepts

[2]See it online at http://wiki.answers.com/Q/What_is_best_estimate_and_how_do_i_calculat_it

my explanation, or Fred's explanation regarding the ability to incorporate an estimate in a model of *severe uncertainty*, the bottom line is that no special status can be attributed to the estimate $\tilde{u}$ in the uncertainty model. WA-2 provides the rationale for this fact.

On the whole, the above working assumptions encapsulate an understanding of "severe uncertainty" that is very much in line with the manner in which this concept is understood in *classical decision theory* and *robust optimization*.

And as a final note (as Fred would no doubt urge me to do), I want to point out that uncertainty models are not objects that stand on their own. One formulates an uncertainty model to serve as a tool of thought, which together with other tools of thought, is used in the formulation of *decision-making models*.

So, the uncertainty model described above should be viewed as the "raw" model that would be refined and adjusted to give expression to the special requirements and objectives of the decision problem that the analyst would formulate in terms of a *decision-making model.*

For this reason, judgment on this simple model should be reserved to a later stage where I examine how this model would be incorporated in models for decision in the face of severe uncertainty.

4.3 The $1 \diamond 2 \diamond 3$ Recipe

To round out my discussion on how a modeler/analyst might or would approach a problem that is subject to severe uncertainty, I want to examine very quickly a prescription for the management of severe uncertainty that I call "the $1 \diamond 2 \diamond 3$ Recipe".

Such a recipe — as readers who will make it to the second part of this book will learn — is advanced with great fanfare by *info-gap decision theory* as being particularly well-suited for this purpose.

So, although it may appear that I am anticipating things, I am going to say a few words about this recipe now, because I want to make the following point clear to those readers who will not make it to the second part of this book.

> This recipe is considered in this book **not** as an alternative to the others that will be outlined in this book. It is considered here so as to illustrate a recipe that ought not even be contemplated for the management of SEVERE uncertainty.

For, as I explain in greater detail in the second part of the book, the fundamental flaw in this recipe is that it misapplies the formula for obtaining *local robustness* to the pursuit of *global robustness*. It prescribes doing the following:

> If you have a (decision) problem where one or more than one of its key parameters is/are subject to severe uncertainty, pick an estimate(s) (which in all likelihood can be substantially wrong), conduct a robustness (read: worst-case) analysis (only) around this estimate and call it a day!
>
> The decision(s) you identified is/are robust to severe uncertainty.

Or more succinctly:

The $1 \diamond 2 \diamond 3$ Recipe for the management of severe uncertainty

1. Pick an estimate for the true value of the parameter of interest. Any estimate.
2. Ignore the difficulties arising in view of the severity of the uncertainty.
3. Conduct your analysis in the neighborhood of the estimate.

That is, in blatant disregard for the problematic issues associated with "picking" an estimate for a problem that is subject to severe uncertainty, this recipe unhesitatingly prescribes an analysis that focuses only on this estimate and its neighborhood. So, as this estimate effectively amounts to a "wild guess", by making the results of the analysis contingent so critically on this estimate, it effectively flies in the face of these two maxims:

- Garbage In — Garbage Out (GIGO).
- Results are only as good as the estimates on which they are based.

The conclusion must therefore be that because — as illustrated in Figure 4.3 — science dictates that the quality of the output is by necessity on a par with the quality of the input. That is, short of solid proofs to the contrary, the results obtained by means of this recipe are on a par with the quality of the estimate(s) on which these results are based. Namely, they can be no better than "wild guesses".

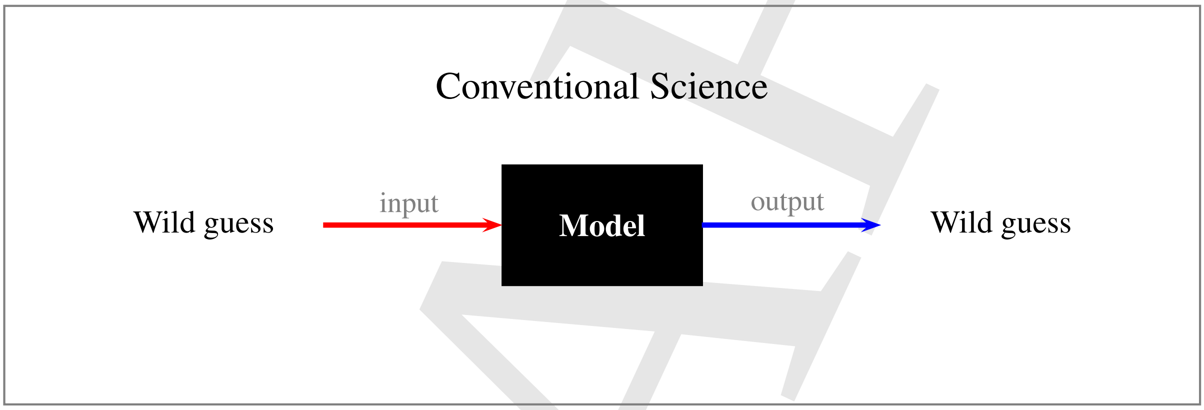


Figure 4.3: Conventional scientific models

The trouble is, however, that the rhetoric accompanying the $1 \diamond 2 \diamond 3$ Recipe portrays it as capable of generating reliable results out of this "wild guess". This rhetoric is illustrated in Figure 4.4.

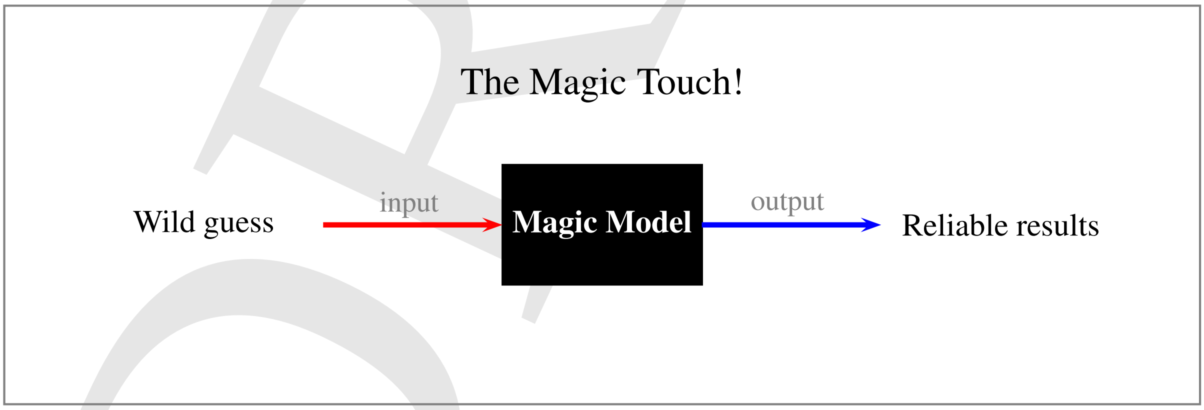


Figure 4.4: Voodoo models

As can be gathered from this picture, I regard recipes of this type as VOODOO RECIPES, or more generally as exponents of a *Voodoo theory.* This definition is in line with the accepted practice to refer to questionable pursuits by means of the sobriquet "voodoo": "voodoo economics", "voodoo science", "voodoo mathematics", "voodoo statistics", and so on.

As I argue in (Sniedovich 2011, p. 10):

> The point about a Voodoo theory is then that it is unconstrained by universally accepted scientific conventions such as: claims must be supported by facts, evidence, proofs, demonstrations and so on. This means that in stark contrast to conventional scientific theories, Voodoo theories basically have a free hand to claim just about anything. In other words, in the framework of Voodoo theories "anything goes".

In short, in this book, the $1 \diamond 2 \diamond 3$ Recipe is treated as a VOODOO RECIPE par excellence, hence as utterly unsuitable for the treatment of severe uncertainty.

But more on this in the second part of the book.

To conclude this chapter, I want to examine very briefly a taxonomy of the term "uncertainty" so as to give the preceding discussion a proper context.

4.4 Classification of uncertainty

It has been accepted practice in *classical decision theory* to distinguish between three states of affairs:

· Certainty
· Risk
· Uncertainty

The first represents situations where the "true" values of all the parameters of a model are known. The second represents situations where the true values of the model's parameters are unknown, but the uncertainty pertaining to the true values of the model's parameters can be quantified probabilistically. The third represents situations where the uncertainty in the unknown true values of the parameters cannot be quantified by probabilistic models.

The distinction between "Risk" and "Uncertainty" is attributed to the economist Frank Knight (1885 – 1972). Hence the popular phrase *Knightian uncertainty,* often used to designate an uncertainty that cannot be quantified by probabilistic models.

Since the phrase *Knightian uncertainty* is used with abandon in the *info-gap literature* to convey the type of uncertainty that *info-gap decision theory* is claimed to model, analyze and manage, it is important to point out the following.

Knight's (1921) own elaboration of the difference between "risk" and "uncertainty" suggests that he might have had in mind a broader, more potent, conception of uncertainty — one that also encompasses what we call today "unknown unknowns". Here is his phrasing of this point (emphasis added):

> To preserve the distinction which has been drawn in the last chapter between the measurable uncertainty and an unmeasurable one we may use the term "risk" to designate the former and the term "uncertainty" for the latter. The word "risk" is ordinarily used in a loose way to refer to any sort of uncertainty viewed from the standpoint of the unfavorable contingency, and the term "uncertainty" similarly with reference to the favorable outcome; we speak of the "risk" of a loss, the "uncertainty" of a gain. But if our reasoning so far is at all correct, there is a fatal ambiguity in these terms, which must be gotten rid of, and the use of the term "risk" in connection with the measurable uncertainties or probabilities of insurance gives some justification for specializing the terms as just indicated. We can also employ the terms "objective" and "subjective" probability to designate the risk and uncertainty respectively, as these expressions are already in general use with a signification akin to that proposed. **The practical difference between the**

> **two categories, risk and uncertainty, is that in the former the distribution of the outcome in a group of instances is known (either through calculation a priori or from statistics of past experience), while in the case of uncertainty this is not true, the reason being in general that it is impossible to form a group of instances, because the situation dealt with is in a high degree unique.** The best example of uncertainty is in connection with the exercise of judgment or the formation of those opinions as to the future course of events, which opinions (and not scientific knowledge) actually guide most of our conduct.
>
> Knight (1921, III.VIII.1-2)

The assertion of interest (bold faced) seems to suggest that in Knight's understanding, "uncertainty" also refers to states of affairs where our lack of knowledge impedes not only the ability to specify the probabilities of events of interest, but also the *uncertainty space* of the problem. So it is not only that the probabilities are difficult to quantify, the events themselves are unknowable.

John Maynard Keynes' (1883–1946) phrasing of the distinction between "Risk" and "Uncertainty" also seems to point in this direction:

> By "uncertain" knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty; nor is the prospect of a Victory bond being drawn. Or, again, the expectation of life is only slightly uncertain. Even the weather is only moderately uncertain. The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention, or the position of private wealth owners in the social system in 1970. About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know. Nevertheless, the necessity for action and for decision compels us as practical men to do our best to overlook this awkward fact and to behave exactly as we should if we had behind us a good Benthamite calculation of a series of prospective advantages and disadvantages, each multiplied by its appropriate probability, waiting to be summed.
>
> John Maynard Keynes (1937, pp. 213-214)
> The General Theory of Employment
> Quarterly Journal of Economics, 209-223, 1937.

Over the years, a host of descriptors have been added to the term "uncertainty" in an ongoing effort to capture the "true nature" of uncertainty and to distinguish perhaps between various degrees or levels of magnitude of uncertainty. To mention just a few:

> Strict uncertainty, Severe uncertainty, Extreme uncertainty, Deep uncertainty, Substantial uncertainty, Essential uncertainty, Hard uncertainty, High uncertainty, True uncertainty, Fundamental uncertainty, Wild uncertainty, Radical uncertainty, Profound uncertainty, Knightian uncertainty, True Knightian uncertainty.

In a recent paper, Lo and Mueller (2010) put forward the idea that it makes better sense to speak of a spectrum ranging from certainty to various "levels" of uncertainty. They propose a taxonomy of uncertainty that is far more refined: "...one capable of explaining the differences across the entire spectrum of intellectual pursuits from physics to biology to economics to philosophy and religion ...". Thus, the spectrum that they propose consists of the following levels:

· Level 1: *complete certainty.*
  An idealized deterministic world.

· Level 2: *risk without uncertainty.*
  The randomness under consideration is governed by a probabilistic model and the set of possible outcomes is completely known.
· Level 3: *fully reducible uncertainty.*
  This is a weaker version of risk in that it " ... can be rendered arbitrarily close to Level-2 uncertainty with sufficiently large amounts of data using the tools of statistical analysis. "
· Level 4: *partially reducible uncertainty.*
  Here "... there is a limit to what we can deduce about the underlying phenomenon generating the data." Consequently, in this environment "... classical statistics may not be as useful as a Bayesian perspective, in which probabilities are no longer tied to relative frequencies of repeated trials, but now represent degree of belief".
· Level 5: *irreducible uncertainty.*
  This represents *total ignorance*, that is " ... ignorance that cannot be remedied by collecting more data, using more sophisticated methods of statistical inference or more powerful computers, or thinking harder and smarter."
· Level ∞: *Zen uncertainty.*
  "Attempts to understand uncertainty are mere illusions; there is only suffering.

Considering that the model of interest in this chapter is a likelihood-free model, one would have thought that it would fall under *Level 5: Irreducible uncertainty.* Let us therefore examine what "Irreducible uncertainty" entails according to Lo and Mueller (2010, p. 13):

> Irreducible uncertainty is the polite term for a state of total ignorance; ignorance that cannot be remedied by collecting more data, using more sophisticated methods of statistical inference or more powerful computers, or thinking harder and smarter. Such uncertainty is beyond the reach of probabilistic reasoning, statistical inference, and any meaningful quantification. This type of uncertainty is the domain of philosophers and religious leaders, who focus on not only the unknown, but the unknowable.
>
> Stated in such stark terms, irreducible uncertainty seems more likely to be the exception rather than the rule. After all, what kinds of phenomena are completely impervious to quantitative analysis, other than the deepest theological conundrums? The usefulness of this concept is precisely in its extremity. By defining a category of uncertainty that cannot be reduced to any quantifiable risk — essentially an admission of intellectual defeat — we force ourselves to stretch our imaginations to their absolute limits before relegating any phenomenon to this level.

The inference to be drawn from this portrayal of *irreducible uncertainty* is that the severe uncertainty that is of interest to us in this book is of the *Level 4: partially reducible uncertainty* type, rather than of the *Level 5: irreducible uncertainty* type. This implies in turn that Bayesian models of uncertainty can be used to quantify the severe uncertainty under consideration here. And, reformed Bayesian though I am, I have no problem with this conclusion.

As a matter of fact, I am going to show — in the next chapter — that the so-called Bayesian "belief based" analysis is not necessarily the preserve of a Bayesian approach. That is, my investigation of how "robustness against severe uncertainty" is pursued in *classical decision theory* and in *robust optimization* will examine what kind of "beliefs" are appealed to in these fields to quantify severe uncertainty, and how these "beliefs" shape the structure of the relevant robustness models. My point is then that no one has the sole rights to the idea that "beliefs" determine the analysis ... not even the Reverend Thomas

Bayes (1702 – 1761).

To conclude this section, I want to note very briefly — given the implications of this discussion for my impending critique of *info-gap decision theory* — a classification of uncertainty put forward by *info-gap scholars.* Consider then the summary shown in Figure 4.5.

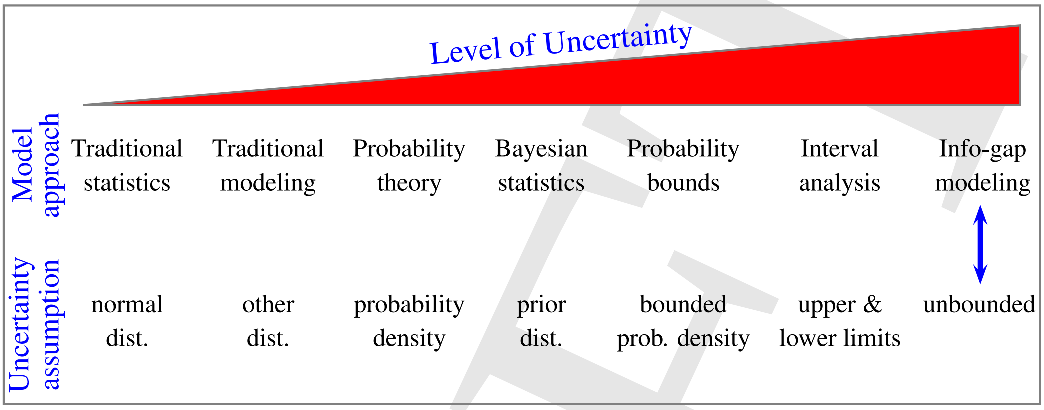


Figure 4.5: Treatment of various levels of uncertainty

This summary is a reproduction of the assessment given by Halpern et al. (2006) of the domains of applicability of various methods and approaches relative to the severity of the uncertainty that they postulate. Clearly, the uncertainty that *info-gap decision theory* is claimed to address is the severest on this scale. Because, *info-gap's likelihood-free model of uncertainty* is depicted by this classification, as a tool for the modeling, analysis and management of an uncertainty that is characterized by an UNBOUNDED uncertainty space.

4.5 Winds of change

The following query was posted recently on the website of *INFORMS Decision Analysis Society*[3] by Scott L. Mitchell, Chairman and CEO, OCEG (Open Compliance & Ethics Group):

> Proper Definitions of Risk and Uncertainty
>
> by smitchell360 Thu Jan 27, 2011 1:44 pm
>
> Fairly recently, the International Standards Organization issued ISO 31000 which is a standard for Risk Management. I wanted to get some perspectives on definitions in the standard. In particular, the definition for risk:
>
> ISO 31000 says
>
> - Risk is "the effect of uncertainty on objectives"
> - Effect is "a positive or negative deviation from the expected"
> - Uncertainty is "the state, even partial, of deficiency of information related to, understanding or knowledge of, an event, its consequence, or likelihood"
>
> So . . . (edited a bit from simple word replacement to make more sense)
> Risk is the positive or negative deviation from the expected likelihood and consequences of an event; caused by deficiency of information about an event, its consequence, or likelihood.

[3] See http://www.syncopation.com/forums/viewtopic.php?f=6&t=396

> So . . . High risk means a high positive or negative deviation from the expected likelihood and consequence of an event.
> Low risk means a low positive or negative deviation from the expected likelihood and consequence of an event.
>
> So . . . I can have high risk without the possibility of a negative deviation. In other words, loss or a less desirable outcome is not required for risk to exist.
>
> Am I reading and reconstructing this correctly?
>
> I don't think that ANYONE would disagree that the discipline of risk management is about managing BOTH the positive and negative. The question is whether one has to re-define the word risk to make this point? Also if these definitions are accepted, what does that do to 400 years of formulas and textbooks in the fields of uncertainty, risk and decisions science?

It reflects the grim reality that the established distinction between "risk" and "uncertainty" is unsatisfactory.

And was pointed out above, there are serious difficulties in dealing with non-probabilistic, likelihood-free quantification of uncertainty, the point being that as things stand, *Uncertainty* eludes *measuring*. It is simply impossible to provide a means by which we would "measure" the level, or degree, of *Uncertainty* to thereby indicate how great or daunting it is.

So, as explained above, to make up for this difficulty a tradition has developed whereby the level, or degree, of *Uncertainty* is captured descriptively, that is informally, through the use of labels such as these:

- Strict uncertainty
- Severe uncertainty
- Extreme uncertainty
- Deep uncertainty
- Substantial uncertainty
- Essential uncertainty
- Hard uncertainty
- High uncertainty
- True uncertainty
- Fundamental uncertainty
- Wild uncertainty
- Radical uncertainty
- Profound uncertainty
- Knightian uncertainty
- True Knightian uncertainty

The trouble is, however, that all too often, these terms are used as no more than buzzwords with a web of empty rhetoric spun around them. So, to guard against this, it is important to be clear on their meaning in the context of the problem under consideration.

It is heartening, therefore, that this message comes through loud and clear in the following "official" assessment:

> Another concern of the committee regarding the content of this chapter involves the use of the concept of "robustness." The committee finds that this term is insufficiently defined. A plausible argument can be made that there is no meaningful distinction from usual optimality analysis and that the concept discussed in this report is a matter of a poorly defined utility function. If indeed there is a real technical distinction to be made, the authors should consider expanding and supporting the discussion of this concept. Furthermore, the committee suggests that the authors address the concept of adaptive management in conjunction with discussions of robustness and in particular address how different sources of uncertainty affect different kinds of decisions. Finally, the committee would appreciate a further elucidation of what the author considers to

> constitute “deep uncertainty” (page 34 and other locations). The committee understands that there is overlap between this concept and the others defined in this section (e.g., “robust”), but nevertheless finds that it is not entirely clear when the author considers the situation inappropriate for use of conventional methods for characterizing uncertainty.
>
> Review of the U.S. Climate Change Science
> Program’s Synthesis and Assessment Product 5.2
> Best Practice Approaches for Characterizing, Communicating, and Incorporating
> Scientific Uncertainty in Climate Decision Making
> pp. 17-18, 2007
> http://www.nap.edu/catalog/11873.html

There are, of course, also the legal aspects of this terminology. For instance,

> The Norwest court noted that uncertainty is also required under the Code Sec. 174 regulations and under the process of experimentation test of Code Sec. 41. But since the economic risk test uses the term “substantial uncertainty,” whereas Code Secs. 174 and 41 require only “uncertainty," the court found that the economic risk test required internal-use software to take a “further step” and to have a higher threshold of technology advancement than in other fields.
>
> Rashkin (2007, p. 174)

So, I just wonder what the Court would rule, in cases where there is a distinction between “extreme uncertainty” and "substantial uncertainty". More importantly, who is to decide, and on the basis, whether a given “uncertainty” is “extreme uncertainty", or “true Knightian uncertainty”, or just “substantial uncertainty?

Suffice it to say that . . . the winds of change have reached our shores.

But until more useful universally accepted definitions of “risk” and “uncertainty” become available, if ever, it is the responsibility of scholars/analysts using this terminology to ensure that the terminology used is consistent with the formal models deployed in the modeling and analysis of . . . “risk” and “uncertainty”.

In line with this, in this book I do not evaluate models of uncertainty on the basis of the rhetoric associated with them. Rather, in this book models of severe uncertainty are evaluated on the basis of their structure, features, and the way they actually work in the framework of the associated decision model.

4.6 Bibliographic notes

TBW

Chapter 5

Worst-case Analysis

5.1 Introduction

The interest that *worst-case analysis* has for our discussion is due of course to the prominent role that this analysis has in *robust decision-making*. The type of tactic put forward by *worst-case analysis* has been recognized, at least since the 1940's, as being well-suited for *decision under uncertainty*. In more recent years, the mathematical technique prescribed by the *worst-case analysis* has been incorporated in algorithms designed to identify robust optimal solutions for decision problems subject to severe uncertainty.

But, we need not dig too deeply to realize that the "secret weapon against uncertainty" offered by *worst-case analysis* is not necessarily the preserve of the expert. Indeed, the type of approach manifested in *worst-case analysis* seems to attest to a very basic human attitude towards the unknown, as clearly reflected in a host of popular maxims. For instance:

- Hope for the best, and prepare for the worst!
- It is best to expect the worst!
- When in doubt, assume the worst!

Surely, one would be hard pressed to find a more beautifully phrased statement of the worst-case analysis than the one Rustem and Howe (2002) use as a motto for their book *Algorithms for Worst-case Design and Applications to Risk Management*:

The gods to-day stand friendly, that we may,
Lovers of peace, lead on our days to age!
But, since the affairs of men rests still incertain
Let's reason with the worst that may befall.

William Shakespeare
Julius Caesar, Act 5, Scene 1

Dedicated to those who have suffered the worst case.

Needless to say, by adopting this highly pessimistic approach towards the uncertain and the unknown, one clearly "plays it safe" — perhaps too safe — with respect to unforeseen risks. But, as French (1988, p. 37) reminds us, one must also remember that "nothing ventured, nothing gained".

My main objective in this chapter is to examine how *worst-case analysis* is incorporated in decision-making models subject to uncertainty. I hope that this exposition will also serve to enlighten *info-gap* *scholars* — especially those who are still unwilling to accept this fact — on why *info-gap's robustness analysis* is a worst-case analysis par excellence.

In section 3.5 I explained in detail how worst-case analysis is conducted in the framework of the generic *Radius of Stability* model. So, readers who find this explanation sufficient for their needs may want to proceed directly to the next chapter.

I submit, though, that my discussion in this chapter should be of interest even to readers who are conversant with this concept, because the abstraction that I give worst-case analysis in this chapter provides for a deeper understanding of this extremely important concept.

5.2 Background and motivation

As I indicate above, worst-case analysis is a well-known, well-established strategy. It has been used for decades in many fields. And yet, as demonstrated in the *info-gap literature,* this concept can be badly misunderstood even by experts specializing in risk analysis, decision under uncertainty and related areas. This is evidenced, for instance, in the string of erroneous claims in the *info-gap literature,* about the manner in which worst-case analysis operates, the range of its applicability, the purposes for which it is used, indeed the very rationale behind it.

So, my objective in this chapter is to show that worst-case analysis is used for a variety of purposes. I examine it as a means for dealing with *variability*, including variability induced by *uncertainty,* and I clarify the difference between *local* and *global* worst-case analysis.

But the point I want to emphasize above all is that when it comes to the management of *uncertainty,* worst-case analysis is not necessarily the measure of last resort. Namely, one does not turn to a worst-case analysis only in cases where probabilistic models cannot be used for lack of data and/or knowledge.

Indeed, in many cases worst-case analysis supplements, rather than supplants, probabilistic methods. That is, there are many cases where we are interested not only in say, the *expected value* of the outcome, but also in the *worst outcome.*

And to illustrate, consider this passage from Rustem and Howe (2002, p. xiii):

> The conventional approach to decision under uncertainty is based on expected value optimization. The main problem with this concept is that it neglects the worst-case effect of the uncertainty in favor of expected values. While acceptable in numerous instances, decisions based on expected value optimization may often need to be justified in view of the worst-case scenario. This is especially important if the decision to be made can be influenced by such uncertainty that, in the worst case, might have drastic consequences on the system being optimized. On the other hand, given an uncertain effect, some worst-case realizations might be so improbable that dwelling on them might result in unnecessarily pessimistic decisions. Nevertheless, even when decisions based on expected value optimization are to be implemented, the worst-case scenario does provide an appropriate benchmark indicating the risks.

And to further elaborate (Rustem and Howe, 2002, p. xiv):

> Through its inherent pessimism, the minimax strategy may lead to a serious deterioration of performance. Alternatively, the realization of the worst-case scenario may result in an unacceptable performance deterioration for the strategy based on expected value optimization. *Neither minimax nor expected value optimization provide a substitute for wisdom.* At best, they can be regarded as risk management tools for analyzing the effects of uncertain events.

There is not much that need be added to this extremely well put statement except to illustrate it by example.

Consider then the following well-known problem which can be given a probabilistic (expected value) treatment and a worst-case treatment.

5.2.1 Example: The counterfeit coin problem

You are given a collection of N coins all of which, except one, have the same weight. Your task is to identify the odd coin, using a balance scale. For simplicity assume that the odd coin is heavier than the other coins.

Figure 5.1: Counterfeit coin problem

Consider the first weighing: you place x coins on each side of the scale whereupon $N - 2x$ coins are left off the scale. Since you have no clue where the odd coin is, you cannot predict the result of the weighing. Hence, there is an uncertainty in the result of the first weighing.

The implication is then that the number of weighings required to identify the odd coin cannot be determined a priori even if the weighing strategy is stipulated in detail a priori. There is a strong element of "chance" here.

To deal with the uncertainty regarding which coin is the odd one, we can develop a probabilistic model to determine the results of the weighings. For instance, we can assume that the probability that the odd coin is on the scale is equal to the proportion of the number of coins placed on the scale. Thus, if for instance, there are 80 coins altogether and 30 are placed on each side of the scale, then the probability that the odd coin is on the scale would be equal to 60/80. This means that the weighing has two possible outcomes: the number of coins left for inspection after the weighing will be either 30 or 20 with probability 0.75 and 0.25, respectively.

Using this approach we can then set up an optimization model that will require the optimal solution to minimize the *expected value* of the number of weighings needed to identify the odd coin (e.g. Sniedovich 2003).

Alternatively, we may want to find a weighing policy that is best under the *worst-case scenario.* That is, a policy that minimizes the number of weighings required to identify the odd coin assuming that Nature (Chance) is "playing against us". Note that if the assumption is that Nature (Chance) is a hostile opponent, then in each weighing the odd coin will be taken to be hidden in the *largest* of the three batches of coins: the two on the scale and the one off the scale. This means that Nature's antagonistic attitude enables us to predict with certainty Her policy, to thereby remove the uncertainty regarding the result of a weighing altogether.

For example, if in the first weighing we place x coins on each side of the scale, then the result of the weighing will be as follows:

- If $x \geq N - 2x$ then x coins will be left for inspection.
- If $x < N - 2x$ then $N - 2x$ coins will be left for inspection.

More compactly, $\max(x, N - 2x)$ coins will be left for inspection after the first weighing.

Invoking this observation, we can set up an optimization model that requires the optimal weighing policy to minimize the number of weighings required to identify the odd coin under the worst-case scenario (e.g. Sniedovich 2003).

The following example illustrates the notion *security level* that serves as a measure of *robustness.* It is a fundamental concept in classical *game theory* and *decision theory.*

5.2.2 Example

Consider the two-player game where the *payoff table* for Player 1 (in $AUD) is shown in Table 5.1.

		Player 2				
		A_1	A_2	A_3	A_4	A_5
Player 1	a_1	3	2	5	6	2
	a_2	9	8	0	8	7
	a_3	3	5	4	4	3

Table 5.1: A simple 2-player game

Here Player 1 has 3 alternatives (rows) to choose from, namely a_1, a_2, or a_3, whereas Player 2 has 5 alternatives (columns) to chose from, namely A_1, A_2, A_3, A_4 or A_5. Thus, if for instance, Player 1 chooses alternative a_2 and Player 2 chooses alternative A_5, the payoff to Player 1 is equal to 7. Player 2 may have her own payoff table, but it is not shown here.

Now suppose that Player 1 has no inkling as to what alternative Player 2 will choose. How should she determine the *worst-case scenario* for this game?

Consider for instance alternative a_1. The worst payoff to Player 1, should she choose this alternative, is the smallest payoff shown in the first row of the table, which is equal to 2. Similarly, the worst payoff to Player 1, should she choose alternative a_2, is the smallest payoff in the second row of the payoff table, which is equal to 0. And the worst payoff to Player 1, should she choose alternative a_3 is the smallest payoff in the third row of the payoff table, which is equal to 3.

This back-of-the-envelope analysis is summarized in Table 5.2, where the worst payoffs for alternatives available to Player 1 are identified, and where a new column entitled *SL* for *Security Level* is appended to the original payoff table.

		Player 2					
		A_1	A_2	A_3	A_4	A_5	SL
Player 1	a_1	3	$\boxed{2}$	5	6	$\boxed{2}$	2
	a_2	9	8	$\boxed{0}$	8	7	0
	a_3	$\boxed{3}$	5	4	4	$\boxed{3}$	3

Table 5.2: Security levels for the alternatives for Player 1

The security level of an alternative for Player 1 is then the worst payoff to Player 1 associated with this alternative. It can be regarded as a *measure of robustness* against the uncertainty in the alternative to be chosen by Player 2.

For example, consider alternative a_3, whose security level is equal to 3: regardless of what alternative Player 2 will choose, should Player 1 choose alternative a_3, her payoff will be at least \$AUD 3.

So, according to this measure of robustness, the most robust alternative for Player 1 is the alternative whose security level is the highest, namely a_3. Similarly, the least robust alternative is the alternative whose security level is the lowest, namely a_2.

The next example is a precursor to my extensive discussion in Chapter 7 on robust optimization problems. It is somewhat lengthy, so take a deep breath!

5.2.3 Example

Consider the following abstract constrained optimization problem

$$z^* := \max_{x \in X} \ g(x) \text{ subject to } \ r^* \leq r(x,p), \forall p \in P \tag{5.1}$$

where X and P are some given sets, g is a real valued function on X, r is a real valued function on $X \times P$, and r^* is a given numeric scalar.

This problem represents a typical worst-case analysis because the clause $\forall p \in P$ in the constraint $r^* \leq r(x,p), \forall p \in P$ requires the decision variable x to satisfice the constraint $r^* \leq r(x,p)$ for the worst (smallest) value of $r(x,p)$ over $p \in P$. Conversely, if the constraint $r^* \leq r(x,p)$ is satisfied for the value of p that minimizes $r(x,p)$ over $p \in P$, then this constraint is satisfied by all $p \in P$. Indeed, if the worst (smallest) value of $r(x,p)$ over $p \in P$ is attained, we have:

$$\left\{ \max_{x \in X} \ g(x) \ \text{ s.t. } \ r^* \leq r(x,p), \forall p \in P \right\} = \left\{ \max_{x \in X} \ g(x) \ \text{ s.t. } \ r^* \leq \min_{p \in P} \ r(x,p) \right\} \tag{5.2}$$

Observe that for a "$\geq$" type of constraint, the relation is as follows:

$$\left\{ \max_{x \in X} \ g(x) \ \text{ s.t. } \ r^* \geq r(x,p), \forall p \in P \right\} = \left\{ \max_{x \in X} \ g(x) \ \text{ s.t. } \ r^* \geq \max_{p \in P} \ r(x,p) \right\} \tag{5.3}$$

And what about constrained optimization problems of the following generic form? Do they represent a worst-case analysis?

$$\max_{x \in X} \ g(x) \text{ subject to } \ r(x,p) \in H(x), \forall p \in P \tag{5.4}$$

where for each $x \in X$, set $H(x)$ is a subset of some set $\mathscr{H}$.

Here as well, the $\forall p \in P$ clause in the constraint indicates that a worst-case analysis is lurking beneath the surface, because: since $r(x,p) \in H(x)$ must be satisfied for all $p \in P$, it follows that it must be satisfied by the worst $p \in P$.

But ... what exactly is the "worst $p \in P$" in this case? Namely, what is the worst $p \in P$ with respect to the constraint $r(x,p) \in H(p)$?

The answer is crystal clear:

> For a given $s \in X$, the worst $p \in P$ with respect to the constraint $r(x,p) \in H(p)$, is a $p \in P$ that VIOLATES this constraint, if such a $p \in P$ exists. Otherwise, namely if such a $p \in P$ does

not exist, that is if $r(x,p) \in H(x), \forall p \in P$, then EVERY $p \in P$ is a worst case.

More generally, consider the following generic constrained optimization problem

$$z^* := \max_{x\in X} g(x) \text{ subject to } constraints(x;p), \forall p \in P \tag{5.5}$$

where $constraints(x;p)$ denotes a set of constraints on x that may depend on a parameter $p \in P$. For instance, suppose that $P = [1,3]$, $X = \mathbb{R}^2$ where $\mathbb{R}$ denotes the real line, and the set of constraints on $x \in X$ is as follows:

$$constraints(x;p) = \{x_1 \geq 1, x_2 \geq 1, x_1 + x_2 \leq 7, x_1 + 2px_2 \geq 9, px_1 + x_2 \leq 8\} \tag{5.6}$$

observing that the first three constraints, namely $x_1 \geq 1, x_2 \geq 1$ and $x_1 + x_2 \leq 7$, are independent of p.

To determine the worst values of $p \in P$ with respect to each $x \in X$ pertaining to these constraints, observe that

$$x_1 + 2px_2 \geq 9 \quad \longrightarrow \quad p \geq \frac{9-x_1}{2x_2} \tag{5.7}$$

$$px_1 + \quad x_2 \leq 8 \quad \longrightarrow \quad p \leq \frac{8-x_2}{x_1} \tag{5.8}$$

We therefore conclude that for a given $x \in \mathbb{R}^2$ such that $x_1 \geq 1, x_2 \geq 1, x_1 + x_2 \leq 7$, the constraints are satisfied for any $p \in P$ such that

$$\frac{9-x_1}{2x_2} \leq p \leq \frac{8-x_2}{x_1} \tag{5.9}$$

and are violated for any $p \in P$ such that

$$\text{either } \; p < \frac{9-x_1}{2x_2} \; \text{ or } \; p > \frac{8-x_2}{x_1} \tag{5.10}$$

or both. Thus, let

$$P_v(x) := \left\{ p \in P : \text{either } \; p < \frac{9-x_1}{2x_2} \; \text{ or } \; p > \frac{8-x_2}{x_1} \text{ or both} \right\} \; , \; x \in X' \tag{5.11}$$

where $X' = \{x \in X : x_1 \geq 1, x_2 \geq 1, x_1 + x_2 \leq 7\}$.

For example, for $x = (1,1)$, we have

$$P_v((1,1)) = \left\{ p \in [1,3] : \text{either } \; p < \frac{9-1}{2} \; \text{ or } \; p > \frac{8-1}{1} \text{ or both} \right\} \tag{5.12}$$

$$= \{p \in [1,3] : \text{either } \; p < 4 \; \text{ or } \; p > 8\} \tag{5.13}$$

$$= [1,3] = P \tag{5.14}$$

We therefore conclude that for $x = (1,1)$, all the values of p are worst cases. Indeed, all the values of $p \in P$ violate the constraint $x_1 + 2px_2 \geq 9$ for $x = (1,1)$.

In contrast, consider the case where $x = (4,3)$, for which we have

$$P_v((4,4)) = \left\{ p \in [1,3] : \text{either } \; p < \frac{9-4}{6} \; \text{ or } \; p > \frac{8-3}{4} \text{ or both} \right\} \tag{5.15}$$

$$= \left\{ p \in [1,3] : \text{either } \ p < \frac{5}{6} \ \text{ or } \ \ p > \frac{5}{4} \right\} \tag{5.16}$$

$$= (1.25, 3] \tag{5.17}$$

Note that for $x = (4, 3)$, any $p \in (1.25, 3]$ violates the constraint $px_1 + x_2 \leq 8$, hence all the elements of $P_v((4, 4))$ are worst cases.

5.3 Abstraction

To formalize the notion "worst-case analysis" so as to render it directly applicable to our discussion, we need three ingredients:

- A set of *cases,* call it C.
- A *value function,* call it v, to determine the "value" of the cases $c \in C$. Formally, v is a function on C with values in some set V.
- A *preference criterion* to rank elements of V.

In the framework of this simple model, the worst case, namely the worst $c \in C$, is a $c' \in C$ such that according to the preference criterion, the "value" of c', namely $v(c')$, is the worst element of $v(C) := \{v(c) : c \in C\}$. In other words, assuming that the preference criterion is *complete*[1], $c' \in C$ is a worst element of C iff $v(c)$ is at least as good as $v(c')$ for all $c \in C$. Similarly, $c' \in C$ is a best element of C iff $v(c')$ is at least as good as $v(c)$ for all $c \in C$. In general, there could be more than one worst case, or no worst case at all.

The task of the worst-case analysis is then to identify a c', such that $v(c)$ is at least as good as $v(c')$ for all $c \in C$.

The main thing to note is that in order to conduct a worst-case analysis on C, it is necessary to specify the *value function* under consideration (v), as well as the assumed *preference criterion*.

It goes without saying that if the worst-case analysis is with respect to a parameter whose "true" value is unknown and is subject to uncertainty, then C would be the *uncertainty space* of the parameter. The value function v and the preference criterion quantify the goals and requirements of the decision problem under consideration. Often, more than one representation/formulation of a worst-case analysis is possible for a problem, so it is important to be clear on which representation/formulation is used.

The objective of the next two examples is to dispose of the groundless, badly misleading assertion, often repeated in the *info-gap literature* (see second part of the book), that there is no worst case if the set of cases C is **unbounded.**

5.3.1 Example

Consider the situation where $C = (-\infty, \infty)$, $v(c) = 17 - 3c + (4.5 - c)^2$ and the preference criterion is "the larger the better". So here the worst c in C is a c in $(-\infty, \infty)$ that yields the smallest value for $v(c)$ over $c \in C$. To determine this value, we solve this simple minimization problem

$$v^* := \min_{c \in C} v(c) = \min_{-\infty < c < \infty} \{17 - 3c + (4.5 - c)^2\} \tag{5.18}$$

[1] Any pair of elements of $v(C)$ can be compared.

A bit of calculus indicates that the minimum is attained at $c^* = 6$, yielding $v^* = v(6) = 1.25$. The picture is shown in Figure 5.2.

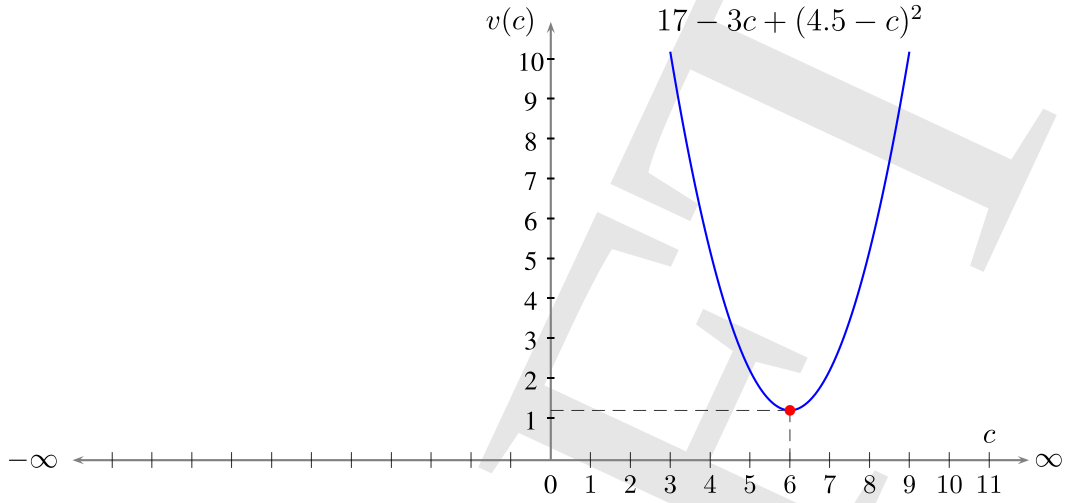


Figure 5.2: Simple worst-case analysis

In short, this is a simple illustration that there can be a worst case even if the set of cases, C, is unbounded.

The next example features a situation where there are infinitely many worst cases.

5.3.2 Example

Consider the case where $C = (-\infty, \infty)$ and the performance criterion is specified by $v(c) = 3+\sin(c)$, assuming that the preference criterion is "the smaller the better". Since $v(c)$ increases with $sin(c)$, the worst value of $c \in C$ is that which maximizes $\sin(c)$ over $c \in C$. Hence, the worst value of c is any value of $c \in C$ such that $\sin(c) = 1$, meaning that there are infinitely many (countable) such worst values (see Figure 5.3).

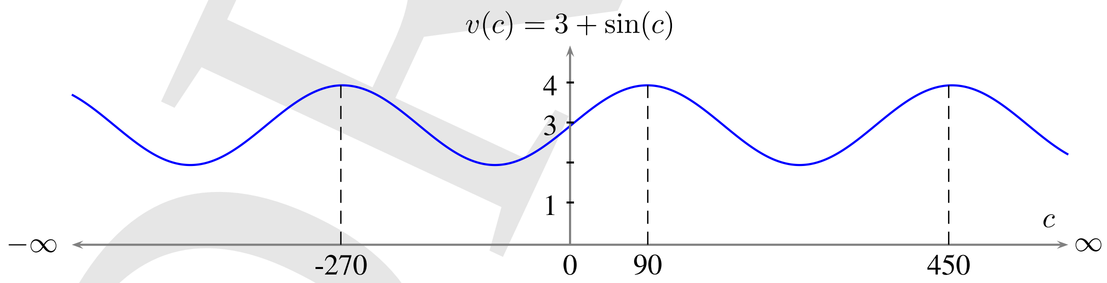


Figure 5.3: Worst cases within an unbounded parameter space, $q = 3$

It is important therefore to be clear on what exactly is at issue here.

It is no doubt true that in the absence of any assumptions about C, the value function v, and the preference criterion, **there is no assurance** that a worst case with respect to $v(c)$ exists on C. However, the fact that C is unbounded **does not imply** that a worst-case does not exist.

By the same token, as indicated by the next example, the fact that the set of cases, C, is bounded is no guarantee that there is a worst case.

5.3.3 Example

Consider the case where $C = [0, 1)$, $v(c) = c$ and the preference criterion is "the smaller the better". The worst $c \in C$ in this situation is then a $c \in C$ whose value $v(c)$ is the largest possible, namely it is the largest value in C. However, set $C = [0, 1)$ does not have a "largest" element. That is, for every $c \in C$ there is a larger element in C.

The inference is therefore clear. Although the set of cases, C, is bounded, there is no worst case here.

5.3.4 Example

The point of this example is to make it clear that worst-case analysis can be conducted indeed, is conducted as a matter of course (in various branches of Optimization), with respect to CONSTRAINTS or REQUIREMENTS and not only with respect to "payoffs" or "costs".

My objective is not only to bring this fact to the attention of readers who are not well-versed in this topic. But to reject out of hand the badly misleading statements in the *info-gap literature* which portray the *info-gap robustness analysis* (the so-called "Robust-satisficing" analysis) as being special/different perhaps even unique, because of its constraint driven robustness analysis (see second part of the book).

To illustrate this point, consider the case where, again, as above, $C = (-\infty, \infty)$ and the value function is defined by $v(c) = 17 - 3c + (4.5 - c)^2$. However, here the *preference criterion* is not "larger is better", as above, but rather is as follows:

Preference criterion:

· $v(c)$ is "acceptable" iff $4 \leq v(c) \leq 5$, otherwise it is "unacceptable".

· "Acceptable" is better than "unacceptable".

Note that this recipe induces a *complete* preference criterion on C, namely for any pair of elements of C, say (c', c''), exactly one of the following applies:

· c' is better than c''.

· c'' is better than c''

· c' and c'' are equally desirable/undesirable.

Note that the last item refers to situations where c' and c'' are either both "acceptable" or both "unacceptable".

Observe that since in this example some elements of C are "acceptable" and some are "unacceptable", it follows that the worst c in C is any value of $c \in C$ such that $v(c)$ is "unacceptable", namely any value of $c \in C$ such that $v(c) \notin [4, 5]$. This is shown in Figure 5.4.

Thus, the set of worst values of c is as follows:

$$WC := \{c \in C : v(c) \notin [4, 5]\} \tag{5.19}$$
$$= C \setminus BC \tag{5.20}$$

where BC denotes the set of "best" values of c, namely

$$BC := \{c \in C : 4 \leq v(c) \leq 5\} \tag{5.21}$$
$$= \{c \in (-\infty, \infty) : 4 \leq 17 - 3c + (4.5 - c)^2 \leq 5\} \tag{5.22}$$

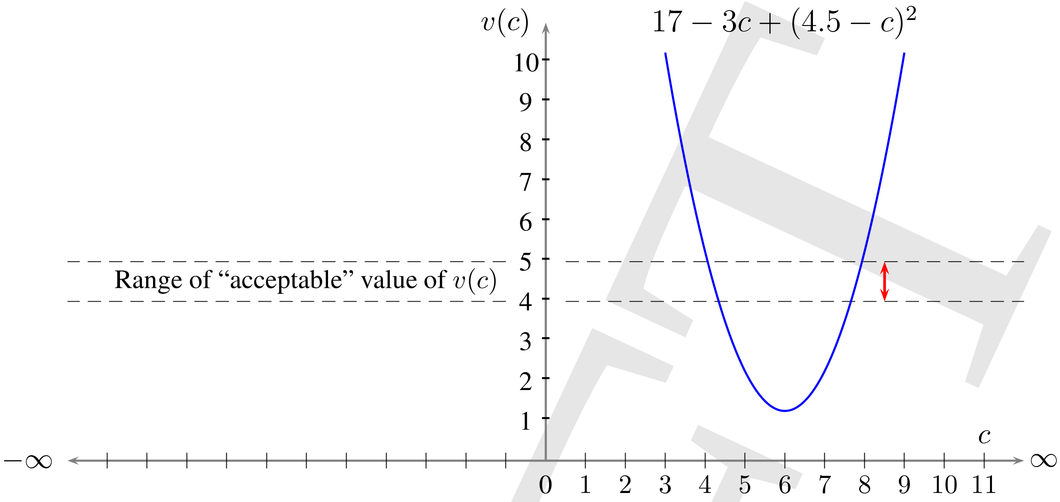


Figure 5.4: A simple worst-case analysis

Note that there are infinitely many "worst" values for c and infinitely many "best" values for c in this example. Readers may at this point try their hand at formulating the set WC explicitly in terms of intervals of values of c.

And to see how easily an alternative formulation can be given to the worst-case analysis under consideration, consider this situation:

Let $g(c) = 17 - 3c + (4.5 - c)^2$ and define the value function v as follows:

$$v(c) := \begin{cases} \text{“}acceptable\text{''} & , \ g(c) \in [4, 5] \\ \text{“}unacceptable\text{''} & , \ g(c) \notin [4, 5] \end{cases} \tag{5.23}$$

Obviously, an "acceptable" value of c is preferable to an "unacceptable" value of c. Hence, the set of worst values of c is as follows:

$$\begin{aligned} WC' : &= \{c \in C : v(c) = \text{“}unacceptable\text{''}\} && (5.24) \\ &= \{c \in C : g(c) \notin [4, 5]\} && (5.25) \\ &= C \setminus \{c \in C : g(c) \in [4, 5]\} && (5.26) \\ &= C \setminus BC' && (5.27) \end{aligned}$$

where BC' denotes the set of "best" values of c, namely

$$\begin{aligned} BC' : &= \{c \in C : v(c) = \text{“}acceptable\text{''}\} && (5.28) \\ &= \{c \in C : 4 \leq g(c) \leq 5\} && (5.29) \end{aligned}$$

Clearly, $BC' = BC$ and $WC' = WC$.

5.3.5 Example

To liven up things, this example features a situation where every $c \in C$ is both a "worst" case and a 'best" case of the parameter of interest. So, let us go back to the previous example and simply add 4 to the expression defining the value function v, namely assume that $v(c) = 21 - 3c + (4.5 - c)^2$ (see

Figure 5.5).

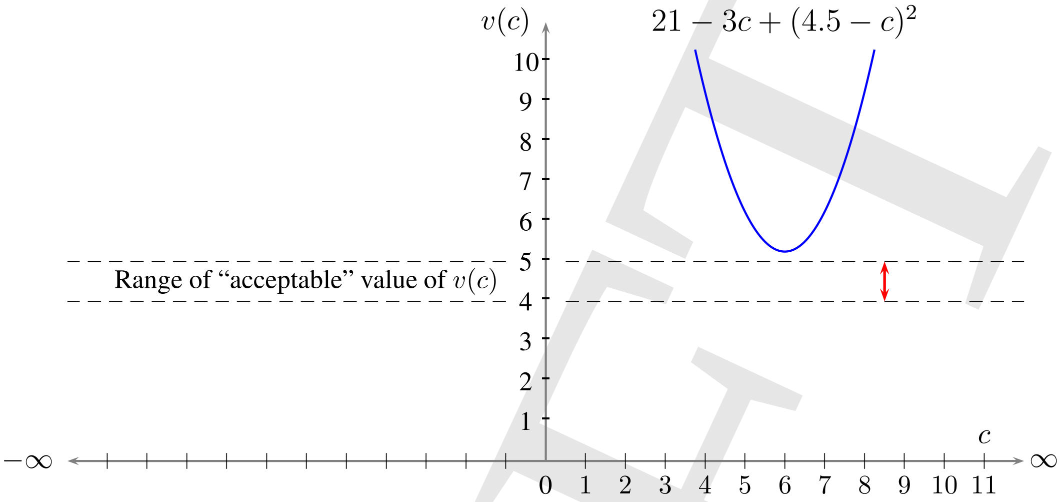


Figure 5.5: A simple worst-case analysis

Observe that this time $v(c) > 5, \forall c \in C$, namely all the values of c violate the requirement $4 \leq v(c) \leq 5$. Thus, if our *preference criterion* remains "a value of c such that $4 \leq v(c) \leq 5$ is preferable to a value of c such that $v(c) \notin [4, 5]$", then any value of $c \in C$ is as good/bad as any other value of $c' \in C$. That is, all the elements of C are equally good/bad: all are "worst" elements of C and all are "best" elements of C.

5.3.6 Example

It is important to take note that in all the previous examples, the three ingredients of the worst-case analysis model were specified in explicitly in the formulation of the problem.

But this, as the reader no doubt suspects, is not necessarily the case in practice. Problems (amenable to a worst-case analysis) do not come our way presenting business cards such as the one shown in Figure 5.6.

Worst-Case for All !

G'day:

Let me introduce myself. I am

$$\beta^* := \max \{\beta \geq 0 : h(p) \in H, \forall p \in T(\beta)\},$$

a typical worst-case analysis problem!
If you have not encountered this type of problem before, you may not be able to recognize me for what I am.

Please feel free to contact me on this matter at your convenience (see my contact details at the back of this card).

Worst wishes,
Your Friendly Worst-Case Analysis Model.

Figure 5.6: A sample business card of a worst-case analysis

The question on our agenda is therefore this:

> What makes the following model a typical worst-case analysis model?
>
> $$\beta^* := \max\,\{\beta \geq 0 : h(p) \in H, \forall p \in T(\beta)\} \tag{5.30}$$
>
> Here p is a parameter whose set of possible/plausible values is say P, and $T(\beta) \subseteq P, \forall \beta \geq 0$.

More specifically, referring to the worst-case analysis model considered in this chapter, the question is as follows:

> What is the set of cases C, what is the value function v, and what preference criterion is used in (5.30) to rank elements of V?

Before I take up these important technical questions, I want to point out that the telltale feature that should immediately reveal that the model specified by (5.30) is a worst-case analysis model, is the clause $\forall p \in T(\beta)$.

For obviously, if the requirement $h(p) \in H$ is to be satisfied by all $p \in T(\beta)$, then it must be satisfied by the worst $p \in T(\beta)$ — whatever it is, assuming that it exists. This immediately suggests that we can set $c = p$, and $C = T(\beta)$. Note that the implication is that from the standpoint of the underlying worst-case analysis, β is a *parameter* and for each value of this parameter a worst-case analysis is conducted on $C = T(\beta)$ with respect to the constraint $h(p) \in H$.

Regarding the preference criterion, the situation is straightforward because the model makes it crystal clear that the requirement $h(p) \in H$ must be satisfied. In other words, the preference is for a p that satisfices this requirement rather than for a p that does not satisfice this requirement. This suggests that h can serve as a value function. Given this informal preliminary analysis we can now argue formally as follows:

> The model specified by (5.30) is a typical worst-case analysis model in the following sense. Consider a given, fixed value of $\beta \geq 0$, let
>
> $$c = p \tag{5.31}$$
> $$C = T(\beta) \tag{5.32}$$
> $$v = h \tag{5.33}$$
>
> and stipulate that values of $c \in C$ such that $v(c) \in H$ are deemed better than values of $c \in C$ such that $v(c) \notin H$. It therefore follows that:
>
> $$\beta^* := \max\,\{\beta \geq 0 : h(p) \in H, \forall p \in T(\beta)\} \tag{5.34}$$
> $$= \max\,\{\beta \geq 0 : \text{ the worst } p \in T(\beta) \text{ satisfices the constraint } h(p) \in H\} \tag{5.35}$$

Differently put, the model specified by (5.30) seeks the largest value of β so that the WORST element of $T(\beta)$, call it $p^*(\beta)$, is "acceptable" in the sense that it satisfices the requirement $h(p^*(\beta)) \in H$.

G'day Moshe:
My extensive experience in this area has left me in no doubt that many analysts are not really at home with the modeling aspects of worst-case analysis.

I therefore urge you to outline here a more "challenging" case, so as to sheds more light on the insight and skills required from the modeler in this endeavor.

Cheers,
Fred
Expert on Robust Decision-Making in the Face of Severe Uncertainty

5.3.7 Example

Taking Fred's comment on board, consider the following model. Keep in mind that it is compared to the model specified by (5.30):

$$\beta^{\circ} := \max\{\beta \geq 0 : h(p) \in H \text{ for at least one } p \in T(\beta)\} \tag{5.36}$$

In case you have not noticed, . . . the only difference between the two models is that the clause "$\forall p \in T(\beta)$" in (5.30) is replaced by the clause *"for at least one $p \in T(\beta)$"* in (5.36).

Since the clause *"for at least one $p \in T(\beta)$"* greatly mitigates the clause "$\forall p \in T(\beta)$" — in fact it is the most mitigated (yet, not altogether disarmed) version therefore — one would suspect that (5.36) is in fact the BEST CASE counterpart of (5.30).

And, to see that this is indeed so, observe that if we use the same setup as in (5.31)-(5.33), we would have

$$\begin{aligned}\beta^{\circ} &:= \max\{\beta \geq 0 : h(p) \in H \text{ for at least one } p \in T(\beta)\} && (5.37)\\ &= \max\{\beta \geq 0 : \text{ the best } p \in T(\beta) \text{ satisfices the constraint } h(p) \in H\} && (5.38)\end{aligned}$$

The lesson to be learned from this exercise is that, just as the clause "$\forall p \in T(\beta)$" proclaims that a WORST-CASE analysis is lurking underneath, the clause "at least one $p \in T(\beta)$" proclaims that a BEST-CASE analysis is in the wings. Note that in this example this is manifested in the relation $\beta^{\circ} \geq \beta^{*}$. Figure 5.7 illustrates this point where $T(\beta)$ represents a circle of radius 1 centered at a point on the line segment $\overline{AB}$ whose distance from point A is equal to β.

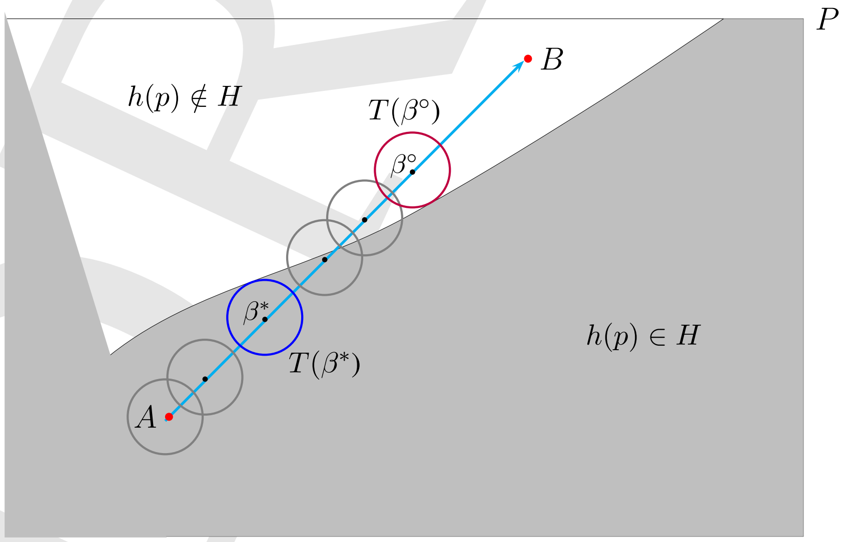


Figure 5.7: Best and Worst case analyses

Note that $T(\beta^{*})$ is contained in H, whereas $T(\beta^{\circ})$ just touches the boundary of H. As expected, $\beta^{\circ} > \beta^{*}$.

5.4 Global, partial and local worst-case analysis

In this section I distinguish between three types of worst-case analysis, namely *global, partial and local.* The importance of this distinction is in the implications that it has for *robustness.* Obviously, these three types of worst-case analysis give rise to three corresponding types of *robustness*: global, partial, and local robustness, respectively.

At first glance, this distinction seems rather obvious so as not to require special attention. However, as my *Info-Gap Experience* has shown, this is not so. Indeed, my experience has shown that robustness can be easily misinterpret so that local robustness would be taken to signify global robustness. Since in my view this is often the result of a lack of awareness of the distinction between local and global worst-case analysis, I decided to include in this chapter a short discussion on this point.

To do this I divide the discussion into two parts: Part One and Part Two (what else?!). In the first I give this issue a somewhat simplistic treatment, simplistic in the sense that I do not go into the fine points of how a blurring of this distinction results in misinterpretation and misrepresentation. In the second part I take up these "fine points" and I illustrates them.

Part One.

5.4.1 Global worst-case analysis

The worst-case analysis described in section 5.3 is *global* in the sense that the domain of the analysis, namely set C, represents the set of ALL the possible/plausible cases under consideration. To reiterate, the model consists of three ingredients:

- A set of *cases,* call it C.
- A *value function,* call it v, that determines the values of cases $c \in C$. Formally, v is a function on C with values in some set V.
- A *preference criterion* to rank elements of V.

The worst-case analysis explores the entire set C with the view to determine the least favorable element(s) of C based on the value function v and the *preference criterion* under consideration.

This description reminds us, of course, of the definition of a *global minimum*:

$$x^* \in X : \ \ g(x^*) \leq g(x) \ , \ \forall x \in X \tag{5.39}$$

where the "$\forall x \in X''$ clause indicates that $x^* \in X$ is the product of a *global* search of the set of possible/plausible values of x specified by X.

As a matter of fact, if the value function v is a real-valued function on C and the preference criterion is "larger is better", than the global worst-case analysis boils down to this:

$$c' \in C : \ \ v(c') \leq v(c) \ , \ \forall c \in C \tag{5.40}$$

That is, c' is a worst element of C iff it is a global minimum of $v(c)$ over $c \in C$.

5.4.2 Partial worst-case analysis

There are many practical reasons why it is often necessary and/or desirable to conduct the worst-case analysis not over C but rather over a SUBSET thereof, call it $\widehat{C}$. This subset can be determined a priori, that is prior to the implementation of the worst-case analysis, or it can be determined impromptu, based on the intermediate results obtained by the analysis.

The term *partial worst-case analysis* obviously indicates that the analysis is conducted only on a part of the set of possible/plausible cases $c \in C$.

In a sense, $\widehat{C}$ can be regarded as a "sample" of the elements of C and a major consideration in determining this "sample" is that it should be an "adequate" representation of the set C. In practice, this is often easier said than done.

> **Dear Reader:**
>
> Pace Moshe, my experience of the last fifty years or so has shown that, in practice, there is no such thing as "a global worst-case analysis".
>
> Without exception, all the worst-case analyses that I carried out were, what Moshe calls here, "partial worst-case analyses".
>
> As Moshe ought to mention in the bibliographic notes to this chapter, it is often utterly impractical to take into account the "real" worst case scenarios for fear that they will give the analysis a far too pessimistic tilt and would thereby call for extremely costly measures.
>
> Cheers,
> Fred
> Expert on Robust Decision-Making Under Severe Uncertainty

I do not disagree with Fred on this issue, but let me note the following.

> The stipulation that a set C designates the set of ALL the possible/plausible cases is not based on the information provided by private investigators who determined that this set is indeed ... the set of ALL the possible/plausible cases. We treat this stipulation for what it is: a proposition or a working assumption.
>
> The descriptor "partial worst-case analysis" given above does not question the merit of this proposition/assumption. It merely indicates that the set on which the worst-case analysis is conducted, namely $\widehat{C}$, is a subset of C — the set representing ALL the possible/plausible cases under consideration.
>
> I agree with Fred that in practice C itself is (often) already a subset of the all the (real) possible/plausible cases. This is precisely the reason that I take care to append the clause "under consideration" to the description of C.
>
> And if this were not enough, I should add that, things can get even more complicated because for mathematical convenience, we often include in C elements that are **not** possible/plausible cases.

A type of partial worst-case analysis that is of particular interest to us in this book, is that which is conducted on a set $\widehat{C} \subset C$ that is a *neighborhood* of some case $\tilde{c} \in C$.

5.4.3 Local worst-case analysis

Suppose that we restrict the worst-case analysis to a neighborhood, $\mathcal{N}(\tilde{\rho}, \tilde{c})$, of C, where the center of the neighborhood, $\tilde{c}$, is given and the radius of the neighborhood, $\tilde{\rho}$, is either specified a priori or is determined impromptu during the analysis.

In this situation the worst-case of c in $\mathcal{N}(\tilde{\rho}, \tilde{c})$, call it c', would have a typical *local* character. Namely, c' would be the least favorable element of $\mathcal{N}(\tilde{\rho}, \tilde{c})$ based on the value function v and the preference criterion under consideration. For instance, if v is a real valued function on C and the preference criterion is "larger is better", then

$$c' \in \mathcal{N}(\tilde{\rho},\tilde{c}): \quad v(c') \leq v(c) \ , \ \forall c \in \mathcal{N}(\tilde{\rho},\tilde{c}) \tag{5.41}$$

Not surprisingly, this reminds us of the definition of *local minimum*.

And to preempt Fred's intervention, I hasten to add that in the absence of all knowledge about C and its relationship with $\mathcal{N}(\tilde{\rho},\tilde{c})$, the local robustness analysis on $\mathcal{N}(\tilde{\rho},\tilde{c})$ can (erroneously) be taken for a global robustness analysis on $C = \mathcal{N}(\tilde{\rho},\tilde{c})$.

The moral of the story is then that it is important to take note that what renders a local robustness analysis *local*, is the discrepancy between the set of all possible/plausible cases, namely C, and the neighborhood of C on which the worst-case analysis is actually conducted, namely $\mathcal{N}(\tilde{\rho},\tilde{c})$. The smaller this neighborhood (compared to C), the more pronounced the "localness" of the analysis, hence the more complex the relationship between them is. This, in turn, raises obvious follow-up questions such as these:

- What is the exact role of the center case $\tilde{c}$ in the analysis?
- Why does it serve as the center of the analysis?
- How is the value of the radius $\tilde{\rho}$ determined?
- How is the local worst-case analysis integrated within a broader framework, if any?

My aim here is not to attempt to answer these questions, but to indicate that they must be addressed whenever a local worst-case analysis is conducted. It is important that the reader take note of the ramifications of local worst-case analysis in situations where C designates an uncertainty space whose objective is to represent *severe* uncertainty.

Par Two[2]

5.4.4 Fine points

The classification described in the preceding sections is based on the set of cases on which the worst-case analysis is actually conducted:

- *Global:* the analysis is conducted on the *entire set* of cases, C.
- *Partial:* the analysis is conducted on a *subset* of C.
- *Local:* the analysis is conducted on a *neighborhood* of an element of C.

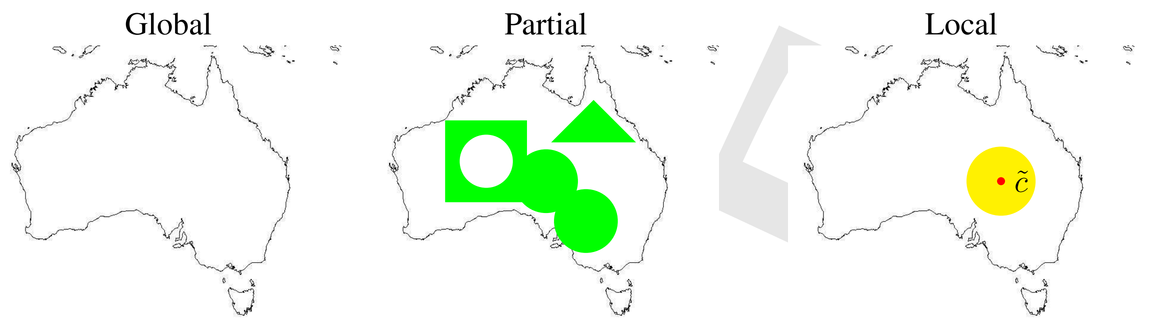
This is shown schematically in Figure 5.8 where, for illustrative purposes, the domain of the partial worst-case analysis, $\widehat{C}$, is deliberately depicted as having an irregular shape.

A more obvious depiction of the distinction between local, partial and global worst-case analysis can hardly be envisaged.

Fred[3] tells me that an exotic country that he had recently visited — call it XYZ — had outlawed local robustness analysis altogether to the effect that it is now illegal to conduct a local worst-case analysis in XYZ.

[2] Recall my decision to defer a discussion on the more subtle issue related to the distinction between global, partial and local worst-case analysis to a later stage.

[3] The internationally known expert on robust decision in the face of severe uncertainty.



C = entire island $\qquad$ $\widehat{C}$ = green area $\qquad$ $\mathcal{N}(\hat{\rho}, \tilde{c})$ = yellow neighborhood of $\tilde{c}$

Figure 5.8: Classification of worst-case analysis based on the domain of the analysis

One of his clients in XYZ, a heavy user of *Radius of Stability* models, was therefore anxious to find out whether it was possible to state (formally) *Radius of Stability* models as models of *global* worst-case analysis, rather than models of *local* worst case analysis.

More specifically, the question was this:

[4]**$A36,000 Question:**

Is it possible to formulate the following *local* worst-case analysis model

$$\rho(q, \tilde{c}) := \max \left\{ \rho \geq 0 : r^* \leq r(q, c), \forall u \in \mathcal{N}(\rho, \tilde{c}) \right\}, \; q \in Q \tag{5.42}$$

as a model of *global* worst-case analysis?

For simplicity assume that $r^* \leq r(q, \tilde{c}), \forall q \in Q$.

Note that the offending clause here is $\forall c \in \mathcal{N}(\rho, \tilde{c})$, where as usual, $\mathcal{N}(\rho, \tilde{c})$ denotes a neighborhood of radius ρ around $\tilde{c}$. It is assumed that all such neighborhoods are contained in the set of all possible/plausible cases, C.

Fred's Answer to the $A36,000 Question:

Definitely, yes! Here is a simple recipe:

$$\underset{\text{Local model}}{\max \left\{ \rho \geq 0 : r^* \leq r(q, c), \forall c \in \mathcal{N}(\rho, \tilde{c}) \right\}} \equiv \underset{\text{Global model}}{\max \left\{ \rho \geq 0 : r^* \leq R(q, c, \rho), \forall c \in C \right\}}$$

where

$$R(q, c, \rho) := \begin{cases} r(q, c) & , \; c \in \mathcal{N}(\rho, \tilde{c}) \\ \infty & , \; c \notin \mathcal{N}(\rho, \tilde{c}) \end{cases}, \; q \in Q, c \in C, \rho \geq 0 \tag{5.43}$$

You can let C be the smallest set containing all the neighborhoods $\mathcal{N}(\rho, \tilde{c}), \rho \geq 0$.

Let me explain Fred's $A36,000 solution to the problem.

The recipe specified by (5.42) calls for the following local worst-case analysis for given triplets $(q, \rho, \tilde{c})$, where $\tilde{c} \in \mathcal{N}(\rho, \tilde{c})$, $q \in Q$ and $\rho \geq 0$:

[4]Consulting fee charged by Fred.

Local worst-case analysis for $(q, \rho, \tilde{c})$**:**
Find the (local) worst $c \in \mathcal{N}(\rho, \tilde{c})$ with respect to the requirement $r^* \leq r(q, c)$ over the neighborhood $\mathcal{N}(\rho, \tilde{c})$.

This analysis is presently outlawed in XYZ, so the idea is to conduct instead a *global* worst-case analysis (on C) that will yield the same results, namely the same worst value of c. Fred's \$36,000 solution suggests the following:

Proposed global worst-case analysis for $(q, \rho, \tilde{c})$**:**
Find the (global) worst value of $c \in C$ with respect to the requirement $r^* \leq R(q, c, \rho)$ over C.

Note that the worst outcome of $r^* \leq R(q, c, \rho)$ over $c \in C$ is "failure" if this constraint is violated by some $c \in \mathcal{N}(\rho, \tilde{c})$, otherwise it is "success". Similarly, the worst outcome for the requirement $r^* \leq r(q, c)$ over the neighborhood $\mathcal{N}(\rho, \tilde{c})$ is "failure" if this requirement is violated by some $c \in \mathcal{N}(\rho, \tilde{c})$, otherwise it is "success".

The equivalence between the model of local worst-case analysis and the model of global worst-case analysis specified in Fred's Answer means that, according to Fred, these two models yields the same results. That is:

For any triplet $(q, \rho, \tilde{c})$, the worst outcome for the requirement $r^* \leq R(q, c, \rho)$ over $c \in C$ is the same as the worst outcome for the requirement $r^* \leq r(q, c)$ over c in the neighborhood $\mathcal{N}(\rho, \tilde{c})$. More specifically,

$$r^* \leq R(q, c, \rho), \forall c \in C \quad \text{iff} \quad r^* \leq r(q, c), \forall c \in \mathcal{N}(\rho, \tilde{c}) \tag{5.44}$$

$$\{c \in C : r^* > R(q, c, \rho)\} \quad = \quad \{c \in \mathcal{N}(\rho, \tilde{c}) : r^* > r(q, c)\} \tag{5.45}$$

A brief explanation regarding the function R defined by (5.43) is in order.

Functions of this type are used extensively in situations where it is required/desirable to "free" a problem from its "constraints" but at the same time allow it to retain its basic characteristics. In the case above, it is necessary to get rid of the "local" clause "$\forall c \in \mathcal{N}(\rho, \tilde{c})$" and replace it with the global clause "$\forall c \in C$", without affecting the results of the worst-case analysis.

This is accomplished by modifying the constraint. Instead of the original constraint $r^* \leq r(q, c)$, Fred invokes the modified constraint $r^* \leq R(q, c, \rho)$, where R is defined in such a way that (5.44)-(5.45) holds — as required. Note that, by construction,

$$R(q, c, \rho) = r(q, c) \ , \ \forall \rho \geq 0, c \in \mathcal{N}(\rho, \tilde{c}) \tag{5.46}$$

$$r^* \leq R(q, c, \rho) \ , \ \forall \rho \geq 0, c \in C \setminus \mathcal{N}(\rho, \tilde{c}) \tag{5.47}$$

This means that the proposed global worst-case analysis reports on a "failure" of $r^* \leq R(q, c, \rho)$ for some $c \in C$ iff the local worst-case analysis reports on a "failure" of $r^* \leq r(q, c)$ for some $c \in \mathcal{N}(\rho, \tilde{c})$, and furthermore, the values of c that cause the global "failure" are the same values that cause the local "failure".

Readers who are averse to, or are uncomfortable with, the presence of ∞ in mathematical expressions, can replace ∞ in (5.43) with any numeric scalar not smaller than r^*.

5.5 The info-gap connection

As I have already indicated in Chapter 3, *info-gap's generic robustness model* is stated as follows:

Robustness of decision q :

$$\hat{\alpha}(q,\tilde{u}) := \max\left\{\alpha \geq 0 : r^* \leq r(q,u), \forall u \in \mathcal{N}(\alpha,\tilde{u})\right\}\ ,\ q \in Q \tag{5.48}$$

where, as usual, $\mathcal{N}(\alpha,\tilde{u})$ denotes a neighborhood of radius α around $\tilde{u}$.

Thus, for each pair (q,α) considered by this model, a *local* worst-case analysis is conducted on the neighborhood $\mathcal{N}(\alpha,\tilde{u})$. In words,

> The local robustness of decision $q \in Q$ at $\tilde{u}$, denoted $\hat{\alpha}(q,\tilde{u})$, is the radius of the largest neighborhood $\mathcal{N}(\alpha,\tilde{u})$ whose worst element, call it $u^*(\alpha)$, satisfices the performance requirement $r^* \leq r(q,u^*(\alpha))$.

This fact is so obvious that one is left gasping at the repeated statements in the *info-gap literature* that ... *info-gap's robustness analysis* is not a worst-case analysis.

Note that under mild regularity conditions, a worst element of $\mathcal{N}(\alpha,\tilde{u})$ can be found by minimizing $r(q,u)$ over $u \in \mathcal{N}(\alpha,\tilde{u})$. In other words,

$$\max\left\{\alpha \geq 0 : r^* \leq r(q,u), \forall u \in \mathcal{N}(\alpha,\tilde{u})\right\} = \max\left\{\alpha \geq 0 : r^* \leq \min_{u\in\mathcal{N}(\alpha,\tilde{u})} r(q,u)\right\} \tag{5.49}$$

provided that the $\min$ is attained.

More on this in the second part of the book but also ... in the following section.

5.6 Modeling issues

As is normally the case in modeling, a host of technical and conceptual issues need to be dealt with when one sets out to formulate a worst-case analysis model. Some of this issues are workaday, others are not. In this section I discuss two related issues that come into play in the modeling of worst-case analysis.

5.6.1 Model choice

The first has to do with the fact that often it is possible to give a problem more than one worst-case analysis representation. This means of course that one's first task would be to chose the right model for the problem considered. Since a discussion of the ins and outs of "choosing the right model" would require going into intricate questions that belong in the domain of "the art of modeling", I shall not dwell on this question here, except ... to note again that it is important that the rhetoric explaining and justifying the model matches the specification of the formal model chosen.

Recall then that the three basic ingredients of a worst-case analysis model are these:

· The set of cases, C.

· The value function, v according to which the "value" $v(c)$ of case c is determined.

· The preference criterion according to which the values of the cases are ranked.

With this is mind, let us examine the following example as an illustration of the fact that a problem can be given various worst-case analysis formulations.

Example

Consider the *Radius of Stability* model which we have just discussed above. Recall that this is the model employed by *info-gap decision theory* to define/assess the robustness of decisions against the uncertainty in the true value of the parameter of interest:

Robustness of decision q :

$$\hat{\alpha}(q,\tilde{u}) := \max \left\{\alpha \geq 0 : r^* \leq r(q,u), \forall u \in \mathcal{N}(\alpha,\tilde{u})\right\}, \ q \in Q \tag{5.50}$$

where, as usual, $\mathcal{N}(\alpha,\tilde{u})$ denotes a neighborhood of radius α around $\tilde{u}$.

Before I proceed to show that this robustness model can be given different worst-case analysis formulations, I call attention to the following manifestly obvious fact:

> This robustness model does **not** prescribe a worst-case analysis of the values that $r(q,u)$ takes for a given $q \in Q$ as the value of u varies over the set
>
> $$\mathscr{U} := \text{ smallest set that contains } \mathcal{N}(\alpha,\tilde{u}) \text{ for all } \alpha \geq 0. \tag{5.51}$$

Should you wonder why I bother to make an observation that states the obvious, and what is more, seems so out of place here, take note that not only is it imperative to make this observation. It is in fact very much apropos my ensuing discussion which will illustrate that the above model has various equivalent worst-case analysis formulations.

The point I want to make then — and this is important for my impending critique of *info-gap decision theory* — is that this obvious fact is routinely used in the *info-gap literature* as (supposedly) providing the "proof" that *info-gap's robustness analysis* is **not** a worst-case analysis, PERIOD. As we shall see in the second part of the book, the argument normally goes like this:

> For the above robustness model to be a worst-case analysis model, there must be a worst case. Namely, there must exist a worst case for u. But, as the value of α is unbounded, then for each value of u in $\mathcal{N}(\alpha,\tilde{u})$ for some $\alpha \geq 0$, there is an even worse value, say u', in a larger neighborhood $\mathcal{N}(\alpha',\tilde{u}), \alpha' > \alpha$. Meaning that a worst value of u may not exist.
>
> And so . . . the argument continues, if the existence of a worst case is not guaranteed, how can it be claimed that the above robustness model is a worst-case analysis model?

I shall explain in detail the staggering errors in this argument in the second part of the book. For the moment I merely want to point out that it is obvious that the above robustness model does not prescribe a worst case analysis of $r(q,u)$ over $u \in \mathscr{U}$.

Rather, as clearly indicated in (5.51), this robustness model prescribes a worst-case analysis of the constraint $r^* \leq r(q,u)$ over $u \in \mathcal{N}(\alpha,\tilde{u})$. This worst-case analysis is *local,* namely it is confined to the neighborhood $\mathcal{N}(\alpha,\tilde{u})$. More specifically, for each value of $\alpha \geq 0$ a local worst-case analysis driven by the constraint $r^* \leq r(q,u)$ is conducted over $u \in \mathcal{N}(\alpha,\tilde{u})$. The purpose of this local analysis is to determine whether the stipulated value of α is admissible for the given q, namely to determine whether the robustness constraint $r^* \leq r(q,u), \forall u \in \mathcal{U}(\alpha,\tilde{u})$ is satisficed.

This clear, let us now consider the following two formulations for such a local worst-case analysis, viewing q and α as given.

Formulation # 1:

Let $C = \mathcal{N}(\alpha, \tilde{u})$, $v(u) = r(q, u), u \in C$ and let the preference criterion on $v(C)$ be based on the simple "larger is better" rule. Hence, in this case the worst u in $\mathcal{N}(\alpha, \tilde{u})$ is one that *minimizes* $r(q, u)$ over $\mathcal{N}(\alpha, \tilde{u})$. The worst-case analysis model is then as follows:

$$z^*(q, \alpha) := \min_{u \in \mathcal{N}(\alpha, \tilde{u})} r(q, u) \tag{5.52}$$

If $z^*(q, \alpha) \geq r^*$, then the worst-case analysis indicates that the value of α is admissible. If, on the other hand, $z^*(q, \alpha) < r^*$, then the value of α is inadmissible.

Since the neighborhoods are nested, in the latter case the implication is that the value of α is too large, meaning that it must be decreased.

Formulation # 2:

Let $C = \mathcal{N}(\alpha, \tilde{u})$,

$$v(u) = \begin{cases} \text{"acc"} & , \quad r^* \leq r(q, u) \\ \text{"unacc"} & , \quad r^* > r(q, u) \end{cases} , \ u \in C \tag{5.53}$$

and let the preference criterion on $v(C)$ be based on the simple rule that "acc" is better than "unacc'. Hence, here the worst u in $\mathcal{N}(\alpha, \tilde{u})$ is one for which $v(u) =$"unacc" if $\mathcal{N}(\alpha, \tilde{u})$ contains such a u. If $\mathcal{N}(\alpha, \tilde{u})$ does not contain such a u, then all the elements of $\mathcal{N}(\alpha, \tilde{u})$ are worst cases.

If the worst value of $v(u)$ over $\mathcal{N}(\alpha, \tilde{u})$ is equal to "acc", then the worst-case analysis indicates that the value of α is admissible. If, on the other hand, the worst value of $v(u)$ over $\mathcal{N}(\alpha, \tilde{u})$ is equal to "unacc", then the value of α is inadmissible.

Since the neighborhoods are nested, in the latter case the implication is that the value of α is too large and must therefore be decreased.

Remark

It should be noted that although the two formulations yield the same value for $\hat{\alpha}(q, \tilde{u})$ in (5.50), they are not completely equivalent. The difference is that typically, the first formulation, based as it is on the minimization of $r(q, u)$, may not yield all the "critical" or "worst" values of u associated with a given value of α. In fact, the optimization problem (5.52) is completely oblivious to the value of r^* and to the constraint $r^* \leq r(q, u)$.

Formally, let

$$WC^{(1)}(q, \alpha) : = \{u' \in \mathcal{N}(\alpha, \tilde{u}) : r(q, u) = z^*(q, \alpha)\} \tag{5.54}$$

$$WC^{(2)}(q, \alpha) : = \begin{cases} \mathcal{N}(\alpha, \tilde{u}) & , \quad \text{"unacc"} \notin V(\mathcal{N}(\alpha, \tilde{u})) \\ \{u \in \mathcal{N}(\alpha, \tilde{u}) : v(u) = \text{"unacc"}\} & , \quad \text{"unacc"} \in V(\mathcal{N}(\alpha, \tilde{u})) \end{cases} \tag{5.55}$$

Then clearly $WC^{(1)}(q, \alpha) \subseteq WC^{(2)}(q, \alpha)$.

5.6.2 Objective function vs constraints

As we saw in the preceding sections of this chapter, there are two modes of formulation of a worst-case analysis.

If the analysis involves the optimization of a real valued function and the "cases" are values of a parameter of the *objective function,* then the worst-case analysis with respect to the parameter can be introduced by means of a second-tier optimization problem whose decision variable is the parameter of interest.

The mechanics would be as follows: consider the generic optimization problem

$$\overline{z}(p) := \max_{x\in X} g(x;p) \tag{5.56}$$

where $p \in$ is some parameter and the $g(x;p)$ notation entails that x designates the "official" argument of function g whereas p is a parameter. Differently put, in this framework x is a *decision variable* whereas p is . . . "just" a parameter.

For each $x \in X$, the worst $p \in P$ is the value of $p \in P$ that MINIMIZES $g(x;p)$ over $p \in P$. Hence, the worst-case analysis for a given $x \in X$ consists of solving the following problem:

$$\overline{\overline{z}}(x) := \min_{p\in P}\ g(x;p)\ ,\ x \in X \tag{5.57}$$

Similarly, in the framework of

$$\underline{z} := \min_{x\in X} g(x;p) \tag{5.58}$$

for each $x \in X$, the worst $p \in P$ is the value of $p \in P$ that MAXIMIZES $g(x;p)$ over $p \in P$. The respective second-tier optimization problem is then as follows:

$$\underline{\underline{z}}(x) := \max_{p\in P}\ g(x;p)\ ,\ x \in X \tag{5.59}$$

The second mode of formulation of a worst-case analysis is used in situations where the "cases" are values of a parameter of the *constraints* under consideration.

The mechanics are as follows: let $x \in X$ denote the decision variable, let $p \in P$ denote the parameter under consideration and let $constraints(x;p)$ denote the set of constraints imposed on x.

As we saw above, the worst-case analysis with respect to these constraints can be formulated as follows:

$$constraints(x;p)\ ,\ \forall p \in P \tag{5.60}$$

In other words, in the framework of a the worst-case analysis of this type, for each $x \in X$ the constraints stipulated by $constraints(x;p)$ must be satisfied for all $p \in P$.

It should be pointed out, though, that the two modes of formulation are related. For instance, note that subject to minor regularity conditions[5],

$$\min_{p \in P} g(x;p) \equiv \max\{v \in \mathbb{R} : v \leq g(x;p), \forall p \in P\} \tag{5.61}$$

$$\max_{p \in P} g(x;p) \equiv \min\{v \in \mathbb{R} : v \geq g(x;p), \forall p \in P\} \tag{5.62}$$

It should also be observed that *constraint satisfaction problems* can be represented by (equivalent) *optimization problems.* For example,

Satisficing problem:

Find an $x \in X$ that satisfices the constraints stipulated by $constraints(x)$. (5.63)

is "equivalent" to

Optimization problem:

$$\max_{x \in X} f(x) \tag{5.64}$$

where

$$f(x) := \begin{cases} 1 & , \ x \text{ satisfices the constraints stipulated by } constraints(x) \\ 0 & , \ x \text{ violates the constraints stipulated by } constraints(x) \end{cases} \tag{5.65}$$

To convince yourself that this is indeed so, simply try to establish that any optimal solution to the *Optimization problem* is a feasible solution to the *Satisficing problem*, and vice versa.

I note these facts in anticipation of the discussions in subsequent chapters of the book. The point is that although from the standpoint of *robustness* it is in principle possible to "unify" the treatment of *constraints* and *objective functions* of optimizations problems, in this book I prefer to keep the distinction alive.

5.7 Trivial worst-case/robustness problems

Dear Reader,

In this section I want to call your attention to the fact that however central the worst case analysis is in *robust decision*, there is a whole category of problems that are so "obviuosly easy" to solve — a category of problems that for simplicity I shall call "trivial robustness problems" — that a "formal" worst case analysis need not even be contemplated in their case.

If you wonder why on earth do I bother to call attention to this fact, if it is indeed so obvious, allow me to point out that, not only is it imperative to make this point clear, it is in fact necessary to discuss it in broader terms. This — as you might have guessed — is another lesson learned from the *Info-Gap Experience*. It turns out that it is indeed important to raise the question of "trivial worst-case/robustness problems" with the view to clarify the role and place that such problems have in research, in education and most of all, in articles destined for ... peer reviewed journals.

[5]Existence of the min and max values.

To give you an immediate insight into the position that I am coming from, take note that my basic position is that "trivial worst-case/robustness problems" do have an important role in the study of *robust decision-making*. I therefore accept that they can feature in discussions in peer reviewed publications. I for one, use such problems extensively in my investigations of the topic of robustness, in my lectures and in my writings — including this book.

But the point is that I expect authors, analysts, lecturers, to be forthright on this matter so as to make it clear that, if a robustness problem under consideration is "trivial", then this problem ought to be presented for what it is . . . a "trivial" robustness problem.

Because I tend to express my views on this issue somewhat "too bluntly", I shall let Fred the renowned expert on robust decision in the face of severe uncertainty, to do the talking in this section.

I am confident that he will do an excellent job!

Cheers,
The Author

Thanks for your kind words, Moshe!

I'll do my best!

Cheers,
Fred
Expert on Robust Decision-Making Under Severe Uncertainty

Having read a significant number of *info-gap publications*, I fully support Moshe's position on the need to bring this matter to the attention of readers of these publications. Particularly because not all readers of these publications, say in applied ecology journals, would be well versed in decision theory, optimization, and related fields, and would therefore be unable to recognize these problems for what they are.

Of course, it is impossible to discuss the role of a"trivial robustness problem" without placing it in the broader context of the role of "trivial problems" in general. So, let me note that I accept that it is often beneficial/expedient/illuminating to simplify a problem under consideration to a degree that it is rendered "trivial", specifically, it submits to solution "by inspection."

I also accept that not all "trivial problems" are necessarily contrived or academic. In fact, there are many extremely important practical problems that are indeed "trivial". Trivial in this very sense that they are easily solved "almost by inspection".

But, having said all that, the point is that it is vital that "trivial" problems not be treated as though they were "non-trivial problems", indeed as though they were challenging problems for the purpose of say, advancing a specific methodology.

I know for a fact that Moshe does not object to the use of "trivial" problems to illustrate the working of this or that solution method. What, he does object to most vehemently is that scholars do not bother to make it clear (as evidenced in *info-gap publications*), that a problem considered is trivial, that it should not perhaps be judged as a real test for the methodology in question, that the problem can be solved by inspection, and so on. Making these facts clear he considers a *sine qua non*!

Since I realize that Moshe is going to examine such problems in the second part of the book, I shall not dwell on this any further. Instead, I shall consider a particular class of "trivial wordt-case/robustness problems", in order to bring out why these problems are "trivial" and why it is important to recognize that they are "trivial". In particular, I shall explain the role of "worst-case analysis models" in the modeling, analysis and solution of such problems.

So, let us begin the discussion with a very simple illustrative example that Moshe will no doubt come back to, in the second part of the book.

Example

Consider the case where we need to assess the robustness of a small number of systems $q \in Q$ with respect to the following simple constraint:

$$A(q) + u \leq r(q) \tag{5.66}$$

where A and r are real valued function on Q such that $A(q) > 0, \forall q \in Q$ and u is a parameter whose range of possible/plausible values is $\mathscr{U} = [\underline{u}, \overline{u}]$, where $\underline{u} < \overline{u}$ are given positive numeric scalars.

Now consider the following robustness problem associated with this constraint:

Robustness Problem:
Determine the robustness of each $q \in Q$ (with respect to the above constraint) against possible/plausible variations in the value of u over $\mathscr{U}$.

Before we formally examine the "global" and "local" versions of this robustness problem in detail, let us examine why Moshe considers this problem a "trivial worst-case/robustness problem".

Note then that the robustness question here is in essence this:

Essential Robustness Question:
What is the range of values of u that satisfice the constraint $A(q) + u \leq r(q)$?

That is, regardless of how we shall ultimately "solve" the worst-case/robustness problem, somewhere along the line we shall have to deal with this question, either implicitly or explicitly. Once we answered this question, the robustness problem is as good as (97.456%) solved.

Hence, Moshe would argue as follows:

The Robustness Problem under consideration is **trivial** because the answer to the Essential Robustness Question is **trivial** in that it can be obtained **by inspection** directly from the constraint itself. In other words:

Answer to the Essential Robustness Question:
By inspection, the range of values of u that satisfice the constraint $A(q) + u \leq r(q)$ is

$$\mathbf{F}(q) = (-\infty, u^*(q)] \ , \ \text{ where } \ u^*(q) := r(q) - A(q) \ , \ q \in Q \tag{5.67}$$

For obvious reasons, let us call $u^*(q)$ the **critical** value of u for $q \in Q$. This value is CRITICAL because

- The constraint $A(q) + u \leq r(q)$ is SATISFICED for all $u \in \mathbb{R}$ such that $u \leq u^*(q)$.
- The constraint $A(q) + u \leq r(q)$ is VIOLATED for all $u \in \mathbb{R}$ such that $u > u^*(q)$.

So, Moshe would argue that to all intents and purposes the robustness problem has been solved once we determined the value of $u^*(q)$. Since determining the value of $u^*(q)$ is a trivial task in this case, the specific robustness problem under consideration is "trivial" PERIOD. This is illustrated in Figure 5.9 for a case where $\mathscr{U} = [0, 7]$, $A(q') = 3$ and $r(q') = 7$ for some $q' \in Q$. This yields $u^*(q') = 4$.

Let us now examine two versions of the problem.

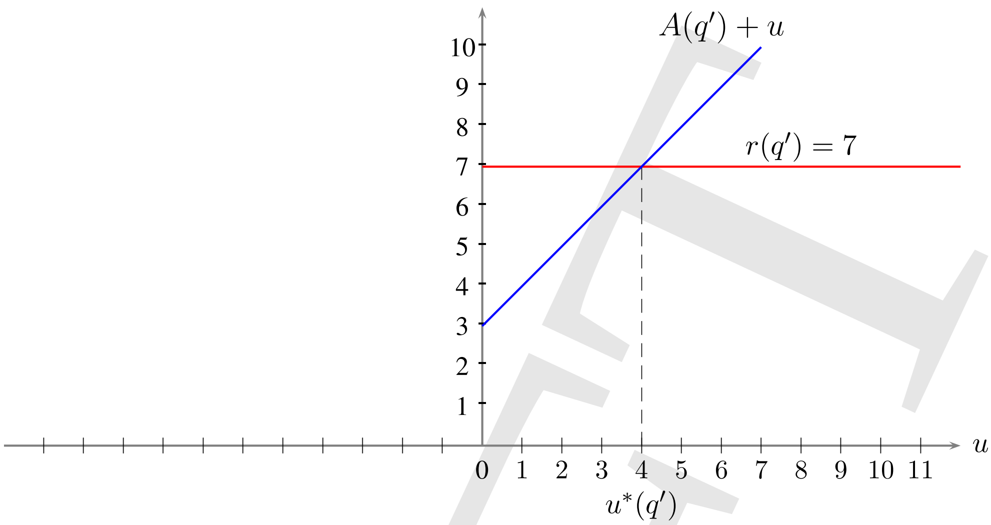


Figure 5.9: A trivial linear robustness problem

Global Robustness

Suppose that we seek global robustness against the variation in the value of u over the interval $\mathscr{U} = [\underline{u}, \overline{u}]$ with respect to the constraint $A(q) + u \leq r(q)$ and that the *Size Criterion* is used as a (global) robustness measure. Then the robustness of system q would be defined as the "size" of the subinterval of $\mathscr{U} = [\underline{u}, \overline{u}]$ over which the constraint is satisfied by system $q \in Q$.

Note that the subinterval of $\mathscr{U} = [\underline{u}, \overline{u}]$ over which the constraint is satisfied by system q is specified as follows:

$$I(q) := \begin{cases} \{\,\} & , \ u^*(q) < \underline{u} \\ \mathscr{U} & , \ u^*(q) > \overline{u} \\ [\underline{u}, u^*(q)] & , \ \underline{u} \leq u^*(q) \leq \overline{u} \end{cases} \quad , \ q \in Q \tag{5.68}$$

where $\{\,\}$ denotes the empty set.

And so, we can define the global robustness of system q to variation in the value of u over the interval $\mathscr{U} = [\underline{u}, \overline{u}]$, with respect to the constraint $A(q) + u \leq r(q)$ as follows:

$$G(q) := \begin{cases} -\infty & , \ u^*(q) < \underline{u} \\ \overline{u} - \underline{u} & , \ u^*(q) > \overline{u} \\ u^*(q) - \underline{u} & , \ \underline{u} \leq u^*(q) \leq \overline{u} \end{cases} \quad , \ q \in Q \tag{5.69}$$

Note that

- q is *super-fragile* if $G(q) < 0$.
- q is *super-robust* if $G(q) = \overline{u} - \underline{u}$.

Also note that the robustness of $q \in Q$ is non-decreasing with $u^*(q)$, and strictly increasing with $u^*(q)$ for $u^*(q) \in \mathscr{U}$. This means that basically, the value of $u^*(q)$ itself can be used as a measure of robustness: "the larger the better" (over $\mathscr{U}$).

Of course, you can establish a Committee to consider other (similar) definitions of global robustness[6].

[6] For a small fee, I'd be delighted to advise this Committee.

Local robustness

Suppose that we have a pretty good (reliable) estimate of the "true" value of $u \in \mathscr{U}$, call it $\tilde{u}$, and assume that $\tilde{u} \in \mathscr{U}$. Furthermore, assume that we are interested in a *Radius of Stability*-like local robustness of the system in the neighborhood of the estimate $\tilde{u}$. We can define the local robustness of system $q \in Q$ in the neighborhood of the estimate $\tilde{u}$ as follows:

$$L(q,\tilde{u}) := \begin{cases} -\infty & , \ u^*(q) < \underline{u} \\ u^*(q) - \tilde{u} & , \ u^*(q) < \tilde{u} \\ \overline{u} - \tilde{u} & , \ u^*(q) > \overline{u} \\ u^*(q) - \tilde{u} & , \ \tilde{u} \leq u^*(q) \leq \overline{u} \end{cases} \quad , \ q \in Q \tag{5.70}$$

Note that a negative local robustness indicates that system $q \in Q$ violates the performance constraint at the estimate $\tilde{u}$ and at values of u that are slightly larger than $\tilde{u}$.

So now a second Committee can be quickly established to deal with possible variations on this theme[7].

5.7.1 So what is the moral of this story for worst-case analysis?

The reader must have noticed that no reference to "worst-case analysis" was made in the analysis of the problem featured in the above example. Furthermore, the "solution" was obtained by inspection. The question is then

> What exactly is it in the constraint $A(q) + u \leq r(q)$ that renders it so easy to handle in the robustness analysis?
>
> Indeed, are there other such "trivial" constraints?

To answer these questions, recall that typically, worst-case analyses associated with robustness problems involve robustness constraints of the generic form

$$constraints(q;u) \ , \ \forall u \in \mathscr{U}(q) \tag{5.71}$$

where $constraints(q;u)$ denotes a list of constraints on decision $q \in Q$ that depend on parameter $u \in \mathscr{U}$ and $\mathscr{U}(q)$ denotes the set of possible/plausible values of u associated with decision q. So consider the two cases:

Generic form	Trivial example
$constraints(q;u) \ , \ \forall u \in \mathscr{U}(q)$	$A(q) + u \leq r(q), \forall u \in \mathscr{U} \subset \mathbb{R}$

Clearly, three factors render this example trivial — relative to the generic form of the robustness constraints.

· In the trivial example, the parameter of interest, u, is a "one-dimensional" object, more specifically, u is a NUMERIC SCALAR (implied by $\mathscr{U} \subset \mathbb{R}$).

· In the trivial example, there is only ONE robustness constraint.

· The range of "acceptable" values of u can be determined by INSPECTION.

[7] My offer to this Committee is the similar.

Regarding the third observation, note that it is not always implied by the first two. For instance, consider the simple case where the single robustness constraint is as follows:

$$B(q,u) \leq r(q) \ , \ \forall u \in \mathscr{U} \subset \mathbb{R} \tag{5.72}$$

where r is a real valued function on Q and B is a real valued function on $Q \times \mathbb{R}$.

Although u is a numeric scalar and there is only one robustness constraint, it might not be easy to determine the set of values of u that violate/satisfice this constraint. This task would be relatively simple if $B(q,u)$ is UNIMODAL[8] with u, in which case — subject to simple technical regularity conditions — if the set of values of u that satisfice the constraint is not empty, then it is an interval that is bounded by two CRITICAL values of u, call them $\underline{u}^*(q)$ and $\overline{u}^*(q)$, assuming that $\underline{u}^*(q) \leq \overline{u}^*(q)$. These two CRITICAL values are found by solving the equation $B(q,u) = r(q)$ for u. If this equation does not have a solution then q is either super-fragile: there are no values of u that satisfice the constraint $B(q,u) \leq r(q)$ for the given value of q; or super-robust: there are no values of u that violate the constraint $B(q,u) \leq r(q)$ for the given value of q.

Note that in the case of the problem featured in the above example, $B(q,u) = A(q) + u$, hence the two bounds are $\underline{u}^*(q) = -\infty$ and $\overline{u}^*(q) = r(q) - A(q)$.

To illustrate, consider the case where the robustness constraint is as follows:

$$16.75 - qu + (4.5 - u)^2 \leq 2q - 1 \ , \ \forall u \in \mathscr{U} = \mathbb{R} \tag{5.73}$$

where $q \in Q = [0, 10]$.

So we can let $r(q) = 2q - 1$ and $B(q,u) = 16.75 - qu + (4.5 - u)^2$, observing that $B(q,u)$ is *convex*, hence unimodal, with u for each $q \in Q$. Thus, for each $q \in Q$ it is straightforward to determine the critical values of u. For instance, for $q = 3$, we solve $16.75 - 3u + (4.5 - u)^2 = 5$ and obtain the two critical values $\underline{u}^*(3) = 4$ and $\overline{u}^*(3) = 8$. This is shown in Figure 5.10

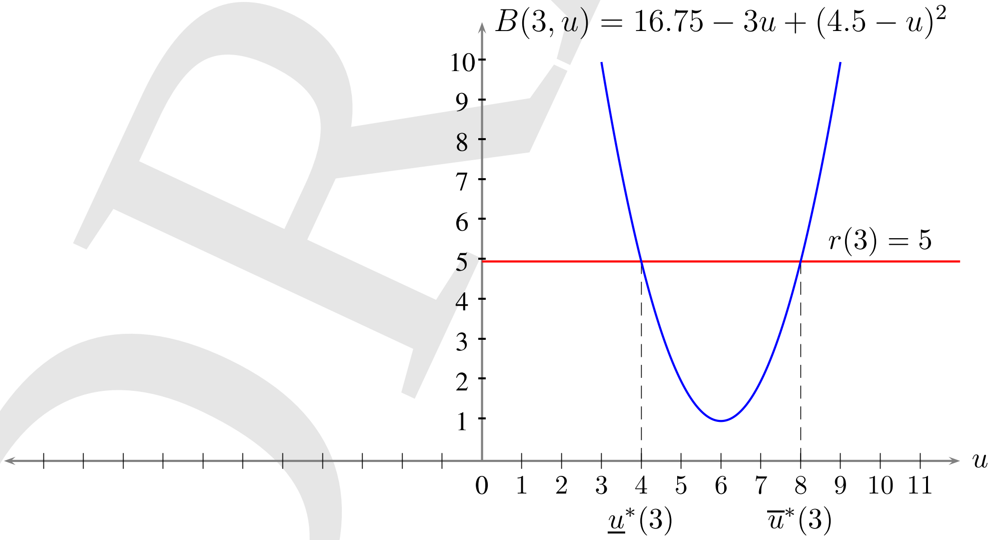


Figure 5.10: A trivial nonlinear robustness problem

The point is then that in the case of trivial robustness problems of this kind, there is no need for a formal worst-case analysis to determine the worst values of the parameter of interest. More than this, in

[8]Roughly, a function f of a scalar x is *unimodal* on the interval $[a,b]$ if for some $c \in [a,b]$ the function is monotonically increasing on $[a,c]$ and is monotonically decreasing on $[c,b]$, or vice versa.

the case of such trivial robustness problems, the distinction between *local* and *global* robustness can be seriously blurred because the set of acceptable values of the parameter u is determined by at most two critical points that can be easily determined by solving the equality form of the constraint.

To sum up this argument I reiterate that there is nothing objectionable in a discussion featuring trivial robustness models. But there is a great deal to object to if the discussion in question does not expressly point out that the problems considered are trivial.

So, it is quite acceptable if an article (in a peer-reviewed journal) stated in the introduction the following:

> In the third section we consider a very important, practical, robustness problem. Our consultant, Fred[9], put it to us that in the field of *robust optimization*, this problem is regarded as trivial because it can be solved by inspection. The trivial solution obtained by inspection is described in detail in the appendix.
>
> Nevertheless, we formulate, analyze and solved this problem with a generic *Radius of Stability* model, in spite of the fact the we recognize that this unduly complicates the formulation of the problem, its analysis and its solution. The reason we do this is that our objective in this section is to illustrate how problems of this type — trivial though they are — as well as problems that are much more complicated, can be treated with this well established and widely used methodology.

Of course, authors may not be in a position to recognize that a problem considered is trivial. I accept this as “a fact of life".

The problem I have is with authors who are determined to use — by hook or by crook — their favorite solution methodology, even in cases where the problem itself is trivial and can be solved by inspection, or by more direct, or much “easier’, methods.

Remark:

I know for a fact that Moshe is particularly concerned that readers of such publications (for instance, *info-gap publications*) are being fed a diet of “trivial robustness problems" as presumably illustrative of real-world problems, encountered in certain areas of expertise (for instance, applied ecology). The trouble with such publications is that they tend to give the wrong impression that *local* robustness and *global* robustness are essentially the same. Or more specifically, that global robustness can be deduced from local robustness.

The purpose of this remark is to make the point that, while this is definitely true in the case of the many trivial robustness problems featured in the *info-gap literature,* this is not the case in general.

So, for the record, let me stress again that one has to take great care to avoid giving the impression that the conclusions (about robustness to uncertainty) deriving from an analysis of trivial robustness problems are as a rule applicable to . . . non-trivial robustness problems.

Cheers,
Fred
Internationally Known Expert on Robust Decision-Making Under Severe Uncertainty

Dear Reader,

[9]Expert on robust decision-making in the face of severe uncertainty.

As promised, Fred did an excellent job in this section so that all that is left to do is to point out the following:

You must have noticed that in the above examples featuring the trivial linear and nonlinear robustness problems, the critical values of u are totally independent of the *estimate* of the true value of u. As a matter of fact, Fred's analysis did not so much as refer to an "estimate" in the global robustness case. In the local robustness case (consistent with our definition of the local case) the *estimate* involves a slight adjustment of the critical value of u.

Think about this point when next you read an *info-gap publication*.

Cheers
The Author

5.8 What next?

Worst-case analysis is central to robustness analysis. It was therefore important to clarify its mode of operation, its capabilities and its scope.

This done, my next goal will be to examine how worst-case analysis is integrated in robustness models of the type used to analyze and manage severe uncertainty.

To this end I shall examine in the next chapter *Wald's Maximin model* which is based on the following extremely simple (to state) idea:

Maximin Rule:
Assess the performance of alternatives, systems, courses of action, decisions, and so on, on the basis of their WORST *performance or outcome!*

Hence, select the alternative, system, course of action, decision, and so on, whose worst-case performance/outcome is BEST.

In the framework of our discussion, the WORST performance/outcome refers to the worst possible/plausible values of a parameter whose true value is unknown, indeed is subject to severe uncertainty.

5.9 Bibliographic notes

TBW

Chapter 6

The Mighty Maximin!

6.1 Introduction

Ever since its debut, in *classical decision theory*, in the early 1950s, *Wald's Maximin model* has figured as the predominant paradigm for the modeling and management of severe uncertainty. In recent years it has been integrated into models developed in the field of *robust optimization* where the objective is to identify robust optimal solutions to decision problems subject to severe uncertainty.

The main attraction of the *Maximin model* is in the extremely simple rule that it puts forth for ranking alternative systems, decisions, policies and so on. Stated verbally the *Maximin* instructs the following:

Maximin Rule
Select the alternative with the best worst outcome!

Clearly then, the *worst-case analysis* is a basic ingredient in this prescription. However, the Maximin goes a step further to prescribe the selection of a *best-worst outcome*. So, conceptually we can talk about a selection process comprising two steps:

Maximin Recipe

- **Step 1:** Determine the worst outcome for each alternative.
- **Step 2:** Select the alternative with the best worst outcome.

In the previous chapter I focused on the modeling of the notion *worst outcome*. In this chapter I develop a mathematical model that governs the selection of the best-worst outcome. The *Maximin Recipe* is thus stated in terms of a mathematical model. Or, in other words, the result is a formal *Maximin model*.

However, before I turn to the discussion itself, I want to clarify a number of points that bear on my investigation of this stalwart of *classical decision theory* and *robust optimization*.

- The term *Maximin*.
  The combination of (Max) and (min) obviously says it all. It signifies that the objective here is to identify the best (Max) worst (min) outcome out of a set of outcomes. The term *Minimax* conveys the same idea: identify the best (min) worst (max) outcome out of the set of outcomes. The difference between the two phrasings is that — from the decision maker's point of view — in the former "larger is better" whereas in the latter "smaller is better". As we shall see, these two paradigms

are equivalent subject to a simple transformation. The implication is therefore that, with no loss of generality, we can concentrate exclusively on the *Maximin model*.

I recognize that some readers may be more accustomed to the *Minimax* phraseology. This however, should present no impediment whatsoever to following my argument throughout this chapter.

· The *Maximin model* featured in this book is **not** the classic maximin model of the *2-person, zero-sum game*. The model I discuss here is far simpler both conceptually and technically.

· However much the *Maximin model* has come to be identified with the modeling and management of *uncertainty,* it is important to keep in mind that, in and of itself, the *Maximin model* has got nothing to do with uncertainty. Essentially, it is a rule dictating the selection of a best-worst alternative. This means of course that it is applicable not only to situations of uncertainty, but to various states of affairs.

· Over the years, the *Maximin model* has gained the reputation (some would say, notoriety) of being far too "conservative". This characterization has developed in view of the fact that it gives rise to "ultra cautious" results . I want to make it clear, though, that as we shall see in this book, this is true about certain but not all of the Maximin's applications.

· The *Maximin model* can be discussed and studied at various levels of technical, namely mathematical, sophistication. The level at which I discuss it in this book is quite elementary.

· Finally, I want to call attention to the badly misleading treatment given to the *Maximin model* in the *info-gap literature*. The reason that it is imperative to point this out is that, as my *Info-Gap Experience* has shown, the erroneous, groundless, misinforming statements about the *Maximin model* in this literature have lead many a scholar and practitioner astray.

But more about all this as we go along.

6.1.1 Example

Consider again the *payoff table* that I constructed in the previous chapter, for a simple 2-player game. Here it is shown in Table 6.1. Recall that, the *SL* column displays the *security levels* of the alternatives available to Player 1 (row player).

		Player 2					
		A_1	A_2	A_3	A_4	A_5	SL
Player 1	a_1	3	$\boxed{2}$	5	6	$\boxed{2}$	2
	a_2	9	8	$\boxed{0}$	8	7	0
	a_3	$\boxed{3}$	5	4	4	$\boxed{3}$	3

Table 6.1: Security levels for Player 1's alternatives

Recall also that the security level of an alternative for Player 1 is the *worst* payoff for that alternative. Hence, following the precepts of the *Maximin rule,* the alternatives are ranked according to their respective security levels: *the larger the better.* The implication is then that the *Maximin rule* prescribes that the best alternative for Player 1 is a_3. Because, the value of the security level of a_3 is higher than those of the other two alternatives.

Expressed symbolically, let:

$$P(i,j) = \text{ payoff to Player 1 if she selects alternative } a_i \text{ and Player 2 selects alternative } A_j.$$
$$SL(i) = \text{ security level of alternative } a_i.$$

Then,

$$SL(i) := \min_{1\le j\le 5} P(i,j)\ ,\ i = 1,2,3 \tag{6.1}$$

and the best alternative (according to the Maximin rule) for Player 1 is determined by solving the following simple *Maximin problem:*

$$z^* := \max_{1\le i\le 3} SL(i) \tag{6.2}$$
$$= \max_{1\le i\le 3} \min_{1\le j\le 5} P(i,j) \tag{6.3}$$

Note that in the absence of a specification of the payoffs to Player 2, it is impossible to rank the alternatives available to her. The payoffs specified by the above table are only for Player 1.

Remark:
This note immediately suggests that it would be extremely difficult to justify this analysis without certain assumptions being made about Player 2. That is, for this analysis to make sense we need to know what type of payoffs does this player stand to obtain, what her objectives are in this game, and what information about Player 1 is available to her.

So, let us pause here for a moment and let us give this player a determinate definition. This definition will apply to all the forthcoming games that we will encounter from now on. And while we are at it, let us do the same with regard to Player 1.

Assumptions:

- All the 2-player games featured in this books are in fact 1-player games, that are played by a *Decision Maker* (DM).
- The so-called "second" player (e.g. Player 2, *Nature,* and so on) is in fact a reification of the DM's *attitude* to *uncertainty/variability.*
- In the framework of the *Maximin paradigm,* the second player represents the most pessimistic attitude to uncertainty/variability imaginable: *expect the worst!*

So from now on, these are going to be the ruling assumptions. In the case of *Maximax* and *Minimin* models that I shall discuss shortly, the ultra-pessimistic attitude is replaced by an ultra-optimistic attitude to uncertainty/variability. More on this in due course.

The next two examples are designed to highlight that although the *Maximin model* is, for the most part, used to model situations where decisions need to be made under conditions of *uncertainty,* as a modeling paradigm it does not obtain its meaning only from an association with uncertainty. For, as indicated at the outset, as a modeling paradigm the *Maximin model* is applicable to various states of affairs — uncertainty being one of them.

6.2 Example

Consider the following simple "bottleneck" problem. The task is to find the best path from node 1 to node 7 in the network shown in Figure 6.1 where the labels on the arcs designate *capacity.* For instance, the capacity of the arc from node 2 to node 5 is equal to 2.

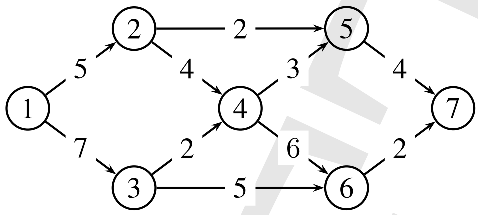


Figure 6.1: Bottleneck problem

The problem is called "bottleneck" because the idea here is that the capacity of a *path* is equal to that of the arc with the *lowest capacity* on this path. For example, the capacity of the path represented by the nodes (1,3,6,7) is equal to 2, which is the capacity of the arc from node 6 to node 7, which is the lowest capacity on this path.

By inspection, the optimal solution is the path going through the nodes (1,2,4,5,7), whose capacity is equal to 3.

This is a typical *Maximin problem* because our goal is to maximize (Max) over all the paths, the smallest (min) capacity on a path. Expressed symbolically, let:

· $P(1, 7)$ = set of paths from node 1 to node 7.

· $A(p)$ = collection of the arcs on path $p \in P(1, 7)$, organized as a set.

· $K(a)$ = capacity of arc a.

Then the bottleneck problem under consideration is as follows:

$$k^* := \max_{p \in P(1,7)} \min_{a \in A(p)} K(a) \tag{6.4}$$

This is a typical, simple, *Maximin problem.*

6.2.1 Example

Consider the collection of points shown in Figure 6.2 and assume that they represent the locations of farm houses in a rural area.

The task is to determine the best location on the map for a *garbage dump* to serve the farms represented by the points on the map.

I consider two versions of this problem, associated with two distinct *odors* emanating from the dump:

· Garlic

· Roses

The Garlic case represents situations where each farmer prefers to be as *far* away from the dump as possible. The Roses case represents situations where each farmer prefers to be as *close* to the dump as possible[1].

[1] I am well aware that to some people the smell of garlic is all roses. So, feel free to change the odors to suit

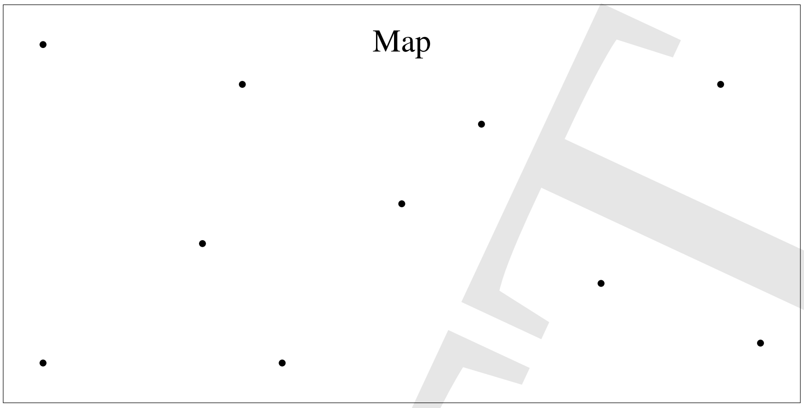


Figure 6.2: A garbage dump location problem

In both cases the term "worst performance" refers to the *least desirable* distance of a farm house from the dump. Thus, in the Garlic case, the worst farm house is the one *nearest* the dump, whereas in the Roses case, the worst farm house is the one *farthest* from the dump.

So the two decision-making situations are as follows:

· *Garlic case*:
Find a point on the map whose distance to the *nearest* farm house is the *largest* possible.

· *Roses case*:
Find a point on the map whose distance to the *farthest* farm house is the *smallest* possible.

Expressed symbolically, let:

$$
\begin{aligned}
F &= \text{ set of farm houses.}\\
X &= \text{ set of admissible locations for the garbage dump.}\\
dist(f,x) &= \text{ distance of farm house } f \in F \text{ to location } x \in X.
\end{aligned}
$$

Then the two versions of the problem are as follows:

Garlic Problem

$$\max_{x\in X}\min_{f\in F} dist(f,x)$$

Roses Problem

$$\min_{x\in X}\max_{f\in F} dist(f,x)$$

However, for the benefit of readers who are not familiar with a *Maximin worst-case analysis,* it should be more instructive to consider a significantly simpler version of this problem. Consider then a case where all the farms houses are located on a *straight line,* as shown in Figure 6.3, and where the garbage dump must also be located on this straight line. The dots represent the farm houses.
So the two decision-making situations are these:

· *Garlic case*:
Find a point on the line whose distance from the *nearest* farm house is the *largest* possible.

· *Roses case*:
Find a point on the line whose distance from *farthest* farm house is the *smallest* possible.

your olfactory nerve

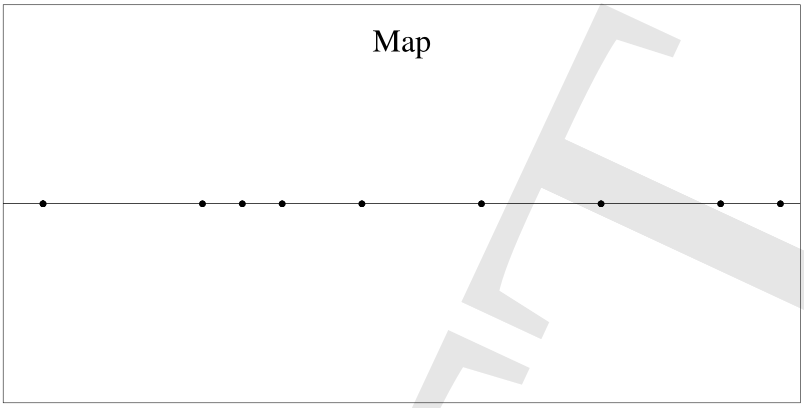


Figure 6.3: A simple garbage dump location problem

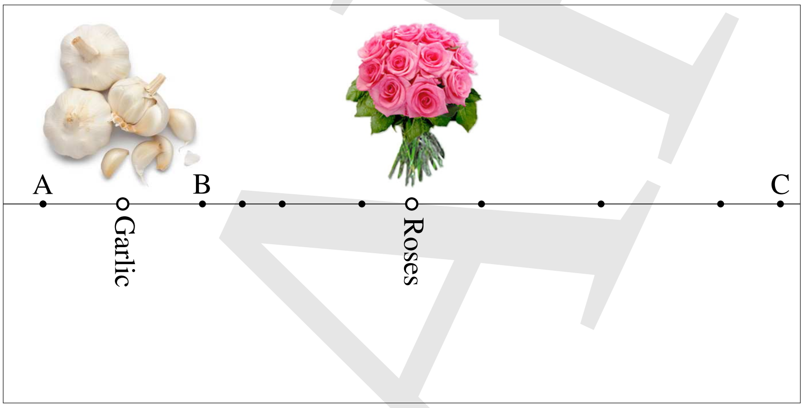


Figure 6.4: Solution to the simple garbage dump location problem

The solutions to these problems are shown in Figure 6.4 and can be explained as follows:

- *Garlic Case:*
  The best location of the dump is either: one of the end points of the line or the mid-point of a pair of adjacent farm houses. In the latter case, it would be the mid-point of the pair of adjacent farm houses that are the farthest away from one another. These are points A and B in Figure 6.3. This midpoint is better than the two end points, hence it must be the optimal location of the Garlic dump.
- *Roses Case:*
  Here we can ignore all the points on the line except point A and point C, as regardless where the dump is placed, either point A or point C will be the worst point. So, the optimal location of the dump is the midpoint of this pair of points.

6.3 Maximin games

At the risk of appearing to contradict myself[2], I am going to maintain that it is instructive to describe the *Maximin model* as a game between two players: the *Decision Maker* (DM) and *Nature*. The former controls the *decision variable*, the latter the *state variable*. The DM plays first, her aim is to maximize her payoff. Nature responds to this move by attempting to minimize the payoff awarded to DM.

Here is a more formal description of the game:

Maximin Game

- **Step 1:** DM selects a decision $x \in X$.
- **Step 2:** Nature selects the worst state in $S(x)$, call it $s(x)$.
- **Step 3:** A payoff $f(x, s(x))$ is awarded to DM.

We refer to X as the *decision space*, to $S(x)$ as the *state space* associated with decision x and to f as the *payoff function*. Note that when Nature selects her state in $S(x)$, she knows what decision was selected by DM, thus Nature's choice would typically be contingent on the decision selected by the DM. This conceptual model is shown in Figure 6.5.

$$\overset{DM}{x \in X} \;\longrightarrow\; \overset{Nature}{s(x) \in S(x)} \;\longrightarrow\; \overset{Payoff}{f(x, s(x))}$$

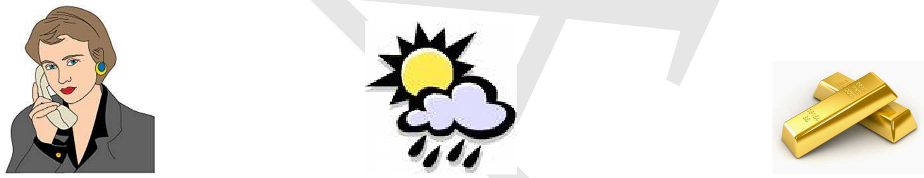


Figure 6.5: Maximin game

The mathematical formulation of this game is as follows[3]:

$$p^* := \max_{x \in X} \min_{s \in S(x)} f(x, s) \tag{6.5}$$

observing that the outer $\max$ represents DM and the inner $\min$ represents Nature.

6.3.1 Example

Let us go back to the Garlic case in Example 6.2.1. Here the DM selects the location of the garbage dump, so:

$x =$ location of the garbage dump.

For simplicity assume that this location is expressed (in the usual manner) as a pair of numbers, representing the two coordinates of the point's location on the map. Thus, in this case

$X =$ set of all the points of the region on the map.

[2]Recall my remark above that all the games discussed in this book are one-player games. I am going to refer to the Maximin game as a two-player game to avoid constant repetition that the second player is the reification of the only player's attitude to uncertainty.

[3]The assumption is that the problem is nice and smooth, namely that the $\max$ and $\min$ values are attained.

Nature selects the "worst" farm house associated with the location determined by DM. Hence, in this case we can set

$S(x) =$ set of the locations of the farm houses.

Note that here $S(x)$ is independent of x.

Last but not least, the payoff $f(x,s)$. This is the distance between the location of the dump (x) determined by DM and the location of the farm house selected by Nature (s). Thus,

$f(x,s) =$ distance between farm house s and location x.

Observe that the distance $f(x,s)$ could be, but does not have to be, the Euclidean distance between the two points.

The Roses case is similar, except that here the DM is seeking to minimize her payoff while Nature is attempting to maximize it. Hence, the model is as follows:

$$p^* := \min_{x\in X} \max_{s\in S(x)} f(x,s) \tag{6.6}$$

This is a Minimax, rather than a Maximin, model.

As I indicated above, I am not going to discuss *Minimax models* in this chapter. I should therefore note that any *Minimax model* can be reformulated as an equivalent *Maximin model* by simply multiplying the payoff function f by -1, invoking the following fact:

$$\min_{y\in Y} g(y) = -\max_{y\in Y} -g(y) \tag{6.7}$$

where g is a real-valued function of some set Y. The implication is then that

$$\min_{x\in X} \max_{s\in S(x)} f(x,s) = -\max_{x\in X} \min_{x\in S(x)} -f(x,s) \tag{6.8}$$

Note that this simple trick does not affect the optimal values of x and s. Namely, (x,s) is optimal with respect to the *Minimax problem* on the LHS of (6.8) iff it is optimal with respect to the *Maximin problem* on the RHS of (6.8).

6.4 Mathematical programming format

Often, it is more convenient to state the *Maximin model* as a model of a "conventional" maximization problem, that is a problem of the form:

$$\max_{y\in Y} g(y) \tag{6.9}$$

$$s.t. \quad \text{list of constraints on } y. \tag{6.10}$$

Or, expressed more succinctly:

$$\max_{y\in Y} \{g(y) : \text{ some constraints on y}\} \tag{6.11}$$

An appeal to the following fact can be most helpful in such cases, as it enables eliminating the inner

$\min$ operation from the *Maximin model:*

$$\begin{array}{ccc} \text{Classic format} & & \text{MP format} \\ \max_{x\in X} \min_{s\in S(x)} f(x,s) & \equiv & \max_{x\in X, v\in\mathbb{R}} \{v : v \le f(x,s), \forall s \in S(x)\} \end{array} \tag{6.12}$$

The equivalence means that these two models yield the same results. If (x^*, s^*) is an optimal solution with respect to the classic format then (x^*, s^*, v^*) is an optimal solution with respect to the MP format, with $v^* = f(x^*, s^*)$. Conversely, if (x^*, v^*) is an optimal solution with respect to the MP format, then (x^*, s^*) is an optimal solution with respect to the classic format, with $v^* = f(x^*, s^*)$ for some $s^* \in S(x^*)$. More specifically, s^* is a value of $s \in S(x^*)$ that minimizes $f(x^*, s)$ over $s \in S(x^*)$.

Note that if (x^*, v^*) is an optimal solution to the problem specified by the RHS of (6.12), then

$$v^* = f(x^*, s^*) = \min_{s\in S(x^*)} f(x^*, s) = \max_{x\in X} \min_{s\in S(x)} f(x,s) \tag{6.13}$$

Similarly, for the *Minimax model* we have:

$$\begin{array}{ccc} \text{Classic format} & & \text{MP format} \\ \min_{x\in X} \max_{s\in S(x)} f(x,s) & \equiv & \min_{x\in X, v\in\mathbb{R}} \{v : v \ge f(x,s), \forall s \in S(x)\} \end{array} \tag{6.14}$$

Note that if (x^*, v^*) is an optimal solution to the problem specified by the RHS of (6.14), then

$$v^* = f(x^*, s^*) = \max_{s\in S(x^*)} f(x^*, s) = \min_{x\in X} \max_{s\in S(x)} f(x,s) \tag{6.15}$$

If you are unfamiliar with the MP format of *Maximin model* so that incorporating v in the model may strike you as “bizarre” , you may want to take a breather at this point, to contemplate this move. I can assure you that this measure is indispensable in the modeling of the *Maximin paradigm* so that you should become adept at performing it.

6.5 Constrained Maximin models

One of the great merits of the MP format is that it enables an easy integration of *constraints* in the formulation of a *Maximin model.* For example, suppose that we want to incorporate the constraint:

$$g(x,s) \in G(x) \ , \ \forall s \in S(x) \tag{6.16}$$

in the *Maximin model* specified by (6.12).

In the framework of the MP format, this task is straightforward as the new (adjusted) *Maximin model* would be stated thus:

$$\max_{x\in X, v\in\mathbb{R}} \{v : v \le f(x,s), g(x,s) \in G(x), \forall s \in S(x)\} \tag{6.17}$$

On the other hand, the equivalent Classic format would be:

$$\max_{x\in X} \min_{s\in S(x)} h(x,s) \tag{6.18}$$

where

$$h(x,s) = \begin{cases} f(x,s) & , \ g(x,s) \in G(x) \\ -\infty & , \ g(x,s) \notin G(x) \end{cases} \ , \ x \in X, s \in S(x) \tag{6.19}$$

The stiff penalty $(-\infty)$ stipulated in (6.19) for violating the constraint $g(x,s) \in G(x)$ ensures that DM would not select an $x \in X$ that violates this constraint for some $s \in S(x)$. In response, Nature will not select a state $s \in S(x)$ that violates this constraint for the decision selected by DM. Should the optimal solution (x^*, s^*) to (6.18) yield $h(x^*, s^*) = -\infty$, the implication would be that the problem does not have a feasible solution.

It goes without saying that this type of constraint can be easily added (explicitly) to the Classic format as follows:

$$\max_{x \in X} \min_{s \in S(x)} \{f(s,x) : g(x,s) \in G(x), \forall s \in S(x)\} \tag{6.20}$$

This is a perfectly kosher *Maximin model* and so is

$$\max_{x \in X} \min_{s \in S(x)} \{f(s,x) : r^* \leq r(x,s), \forall s \in S(x)\} \tag{6.21}$$

where r^* is a given numerical scalar and r is a real-valued function of the decision and state variables.

And for greater effect you may want to consider this model:

$$\max_{\substack{x \in X \\ g(x,s) \in G(x) \\ \forall s \in S(x)}} \min_{s \in S(x)} f(s,x) \tag{6.22}$$

observing that it is equivalent to the one specified in (6.20). However, I suggest that you resist such temptation.

After considerable deliberations, I decided to use in this book the following two equivalent generic formats for the constrained Maximin model:

$$\underbrace{\max_{x \in X} \min_{s \in S(x)} \{f(x,s) : con(x,s), \forall s \in S(x)\}}^{\text{Classic format}} \equiv \overbrace{\max_{x \in X, v \in \mathbb{R}} \{v : v \leq f(x,s), con(x,s), \forall s \in S(x)\}}^{\text{MP format}} \tag{6.23}$$

where $con(x,s)$ denotes the set of constraints (on x) under consideration. Note that in the MP format the clause $\forall s \in S(x)$ to both $con(x,s)$ and $v \leq f(x,s)$.

The notation $con(x,s)$ indicates that some, but not necessarily all, the constraints are dependent on s. For instance, $con(x,s)$ may represent the following two constraints:

$$x_1 + x_2 + x_3 = 100 \tag{6.24}$$

$$s_1 x_1 + s_2 x_2 + s_3 x_2 \geq 200 \tag{6.25}$$

observing that the first constraint is independent of s.

6.6 Modeling issues

As I indicated above, by incorporating the decision variable v in the MP format of the *Maximin model,* one rids this format of the inner $\min$ that is part of the combined $\max\min$ operation in the Classic format. At first glance this trick yields a somewhat unattractive curly brackets expression $\{v : v \leq f(x,s), \dots\}$ that may perhaps puzzle the uninitiated. Observe then that in practice, this generic expression can be simplified by exploiting the structure of function f.

Indeed, often the decision variable v need not be specified at all in the formulation of the MP format, as its role can be taken over by the "genuine" decision variable x. In particular, note that in the framework of the constrained Classic format, if $f(x,s)$ is independent of s, then the inner $\min$ of the combined $\max\min$ operation is redundant so that it can be dropped.

For example, consider this:

$$\max_{x\in X}\min_{s\in S(x)} \{f(x,s) : r^* \leq r(x,s), h(x,s) \in H(x), \forall s \in S(x)\} \tag{6.26}$$

where $f(x,s) = g(x)$ for some real valued function g on X.

We would then have,

$$\max_{x\in X}\min_{s\in S(x)} \{f(x,s) : r^* \leq r(x,s), h(x,s) \in H(x), \forall s \in S(x)\} \tag{6.27}$$

$$\equiv \max_{x\in X}\min_{s\in S(x)} \{g(x) : r^* \leq r(x,s), h(x,s) \in H(x), \forall s \in S(x)\} \tag{6.28}$$

$$\equiv \max_{x\in X} \{g(x) : r^* \leq r(x,s), h(x,s) \in H(x), \forall s \in S(x)\} \tag{6.29}$$

This occurs in cases where robustness (hence worst-case analysis) is sought with respect to the *constraints* of an optimization problem but not with regard to its *objective function.*

So, from the viewpoint of modeling the *Maximin paradigm*, it is instructive to consider three classes of *Maximin models:*

· The objective function f is independent of the state s, but some of the constraints in $cons(x,s)$ depend on s.

· The objective function f depends on s, but all the constraints in $con(x,s)$ are independent of s.

· Both the objective function and some of the constraints depend on s.

For obvious reasons, I do not consider the degenerate case where both the objective function and the constraints are independent of s.

I should also point out that the set of constraints denoted by $con(x,s)$ in (6.23) that is imposed on the *Maximin model* can be classified, in view of their dependence on s, as follows:

· All the constraints are independent of s.

· At least one of the constraints depends on s.

The first case designates situations where the worst-case analysis represented by the *Maximin model* is conducted only with respect to the payoff $f(x,s)$.

So for these and other reasons it is instructive to distinguish between three classes of *Maximin models,* the distinction arising from the types of worst-case analysis that they perform.

6.6.1 A payoff driven worst-case analysis

In situations where the worst-case analysis is performed only with respect to the payoff (not the constraints), the two formats of the constrained *Maximin model* would be formulated as follows:

$$\max_{x\in X}\min_{s\in S(x)}\{f(x,s): con(x)\} \equiv \max_{x\in X, v\in\mathbb{R}}\{v: v\leq f(x,s), \forall s\in S(x), con(x)\} \tag{6.30}$$

where $con(x)$ represents a set of constraints on x, observing that these constraints are independent of s, hence the $\forall s \in S(x)$ clause does not apply to them. For this reason, it is perhaps more convenient to incorporate these constraints in X and use the unconstrained formats of the *Maximin model.*

The important point to note about the MP format in this case is that it clearly exhibits the two hallmarks of the generic MP format, namely the $v \leq f(x,s)$ expression and the $\forall s \in S(x)$ clause.

6.6.2 A constraints driven worst-case analysis

In situations where the worst-case analysis is conducted only with respect to the constraints (not the payoff), the two formats of the constrained *Maximin model* would be formulated as follows:

$$\max_{x\in X}\min_{s\in S(x)}\{f(x): con(x,s), \forall s\in S(x)\} \equiv \max_{x\in X, v\in\mathbb{R}}\{v: v\leq f(x), con(x,s), \forall s\in S(x)\} \tag{6.31}$$

where $con(x,s)$ represents a set of constraints on x and s, jointly.

Observe that the payoff $f(x)$ is independent of s, therefore the $\forall s \in S(x)$ clause does not apply to it. For this reason, the two formats can be simplified to yield a common format, namely:

$$\max_{x\in X}\{f(x): con(x,s), \forall s\in S(x)\} \tag{6.32}$$

Note that this format retains the hallmark clause $\forall s \in S(x)$.

6.6.3 A payoff and constraints driven worst-case analysis

In this case, the MP format is the "lock stock and barrel" MP formulation specified by (6.23) which exhibits both the hallmark $\forall s \in S(x)$ clause and the intriguing $v \leq f(x,s)$ expression.

6.7 Relation to the Radius of Stability model

Consider the generic *Radius of Stability* model that we discussed in Chapter 3, namely

$$\rho(q,\tilde{s}) := \max_{\rho\geq 0}\{\rho: s\in P_{stable}(q), \forall s\in B(\rho,\tilde{s})\}\ ,\ q\in Q \tag{6.33}$$

The clause "$s \in P_{stable}(q), \forall s \in B(\rho,\tilde{s})$" is a clear indication that the model represents a worst-case analysis of some sort. This much is immediately clear.

But, the question is: is this a *Maximin model?*

The first thing to note then about this model is that, the payoff here is independent of s. Considering however that the worst-case analysis is carried out with respect to s, it follows that, this worst-case analysis is driven only by the constraints, not the payoff. The implication is then that the generic *Maximin model* in this case is of the form specified by (6.32), namely

$$\max_{x \in X} \{f(x) : con(x,s), \forall s \in S(x)\} \tag{6.34}$$

A quick visual comparison of this model with the model specified by (6.33) reveals that the latter is indeed a Maximin model. The correspondence between the two models is displayed in Figure 6.6.

Radius of Stability model	Maximin model
$\max_{\rho \geq 0} \{\rho : s \in P_{stable}(q), \forall s \in B(\rho, \tilde{s})\}$	$\max_{x \in X} \{f(x) : con(x,s), \forall s \in S(x)\}$
ρ	x
ρ	$f(x)$
$s \in P_{stable}(q)$	$con(x,s)$
$B(\rho, \tilde{s})$	$S(x)$

Figure 6.6: Correspondence between two models

The conclusion is therefore clear: the *Maximin model* specified in Figure (6.6) is equivalent to the *Radius of Stability* model specified in (6.33). In other words, the *Radius of Stability* model specified in (6.33) is an *instance* (special case)[4] of the generic *Maximin model*.

This fact is brought out more forcefully by the following typical *Maximin Game* that is associated with the Classic unconstrained Maximin format, specified in Figure 6.6, namely:

$$\max_{\rho \geq 0} \min_{s \in B(\rho, \tilde{s})} f(x,s) \tag{6.35}$$

where

$$f(\rho, s) := \begin{cases} \rho & , \ s \in P_{stable}(q) \\ -\rho & , \ s \notin P_{stable}(q) \end{cases} , \ \rho \geq 0, s \in S(q) \tag{6.36}$$

The Maximin game would run as follows:

The Radius of Stability Game

Step 1: DM selects a non-negative number ρ, representing a ball of radius ρ centered at the nominal state $\tilde{s}$ of the system, namely $B(\rho, \tilde{s})$.

Step 2: Nature selects the worst state in $B(\rho, \tilde{s})$, namely the state in $B(\rho, \tilde{s})$ that minimizes $f(\rho, s)$ over $s \in B(\rho, \tilde{s})$. Call this state $s(\rho)$.

Step 3: A payoff $f(\rho, s(\rho))$ is awarded to DM.

The inference is that to maximize her payoff, the DM must select the value of ρ that is equal to the *Radius of Stability* of the system. The formal proof is short:

Formal proof.
If the DM selects a ρ that is greater than the system's *Radius of Stability*, then the worst state in $B(\rho, \tilde{s})$ will violate the requirement $r^* \leq r(q,s)$ whereupon the payoff to DM will be equal to $-\rho$. Clearly, the DM will be ill-advised to select such a value.

[4]See section 6.17

If the DM selects a value of ρ that is smaller than the system's *Radius of Stability*, there would be a larger value of ρ that satisfices the requirement $r^* \leq r(q,s)$ for all $s \in B(\rho,\tilde{s})$. The DM will therefore prefer to select such a larger value of ρ.

The conclusion is therefore that the optimal value of ρ is EQUAL to the *Radius of Stability* of the system. □

At this point I want to call the attention of all those who may not make it to the second part of the book, to take note of the direct implications that the above discussion has for *info-gap's robustness model*. Observe then that as I have already established in Chapter 3 that *info-gap's robustness model* is a *Radius of Stability* model, it follows that *info-gap's robustness model* is a *Maximin model.* The two equivalent formats (Classic and MP) of this model are as follows:

$$\begin{array}{ccc} & \text{Info-gap's robustness model} & \\ \text{Classic Maximin format} & & \text{MP Maximin format} \\ \max_{\rho\geq 0}\ \min_{s\in B(\rho,\tilde{s})} h(\rho,s) & \equiv & \max_{\rho\geq 0}\{\rho : r^* \leq r(q,s), \forall s\in B(\rho,\tilde{s})\} \end{array} \tag{6.37}$$

where

$$h(\rho,s) := \begin{cases} \rho & , \ r^* \leq r(q,s) \\ -\rho & , \ r^* > r(q,s) \end{cases} \tag{6.38}$$

Feel free to replace the $-\rho$ in (6.38) with any negative number.

And just for the fun of it, let us repeat the formulation, this time using the Classic format in (6.37) as the starting point. This yields the following short progression:

$$\overbrace{\max_{\rho\geq 0}\ \min_{s\in B(\rho,\tilde{s})} h(\rho,s)}^{\text{Maximin: Classic format}} \equiv \overbrace{\max_{\rho\geq 0, v\in\mathbb{R}}\{v : v \leq h(\rho,s), \forall s\in B(\rho,\tilde{s})\}}^{\text{Maximin: MP format}} \tag{6.39}$$

$$(\text{From the structure of } h) \equiv \max_{\rho\geq 0, v\in\mathbb{R}}\{v : v\leq \rho, r^* \leq r(q,s), \forall s\in B(\rho,\tilde{s})\} \tag{6.40}$$

$$(\text{optimal } v = \text{optimal } \rho) \equiv \underbrace{\max_{\rho\geq 0}\{\rho : r^* \leq r(q,s), \forall s\in B(\rho,\tilde{s})\}}_{\text{info-gap's robustness model}} \tag{6.41}$$

This means of course that *info-gap's robustness model* is an instance[5] of the Maximin model specified by (6.37). That is, it is the instance obtained from the function h defined in (6.38).

6.8 Role in robust decision-making

As I noted at the outset, *Wald's Maximin model* is the most important paradigm for modeling and managing robust decision to have come out of *classical decision theory*. A superficial examination of the robust optimization literature immediately reveals how prevalent this model, and its many variants are in this area of expertise. But not only there.

Consider for example the following three quotes. The first is the abstract of the entry *Robust Control* by Noah Williams in the *New Palgrave Dictionary of Economics,* Second Edition, 2008[6]:

[5] See section 6.17

[6] See http://www.dictionaryofeconomics.com/article?id=pde2008_R000250&q=robust%20control&topicid=&result_number=1

> Robust control is an approach for confronting model uncertainty in decision making, aiming at finding decision rules which perform well across a range of alternative models. This typically leads to a minimax approach, where the robust decision rule minimizes the worst-case outcome from the possible set. This article discusses the rationale for robust decisions, the background literature in control theory, and different approaches which have been used in economics, including the most prominent approach due to Hansen and Sargent.

The second is from the book *Robust Statistics* by Huber and Ronchetti (2009, p. 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 ϵ-deviation from the assumptions. The optimally robust procedure minimizes this degradation and hence will be a minimax procedure of some kind.

The third is from the paper *Solution of macromodels with Hansen-Sargent robust policies: some extensions* by Paolo Giordani and Paul Söderlind (2004, p. 2370):

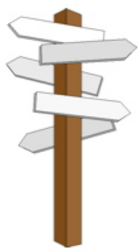
> From a technical point of view, robustness involves a switch from a minimization problem (minimizing a loss function) to an appropriately specified min-max problem. In order to set up and solve a min-max problem, it is convenient to work with a two-agent representation: the policy function selected by the planner is the equilibrium outcome of a two person game in which a fictitious evil agent, whose only goal is to maximize the planner's loss, chooses a model from the available set, and the planner chooses a policy function.

So, my suggestion to you is that if you get this strange "Eureka" feeling, which leads you to believe that you have just come up with a new theory for robust decision-making that is radically different from all current theories for robust decision-making, you'd better take a deep breath and make sure that you have not reinvented *Wald's Maximin model,* or one of its many variates.

Because, if you plan to make it to the second part of this book, you will see that this is precisely what happened in the case of *info-gap decision theory*. This theory is an example of what can happen when *Wald's Maximin model* is misunderstood, or ignored, or misinterpreted, or what have you.

6.9 What next?

In the ensuing sections of this chapter I look at certain modeling aspects of the *Maximin model.* These are immediately relevant to *info-gap decision theory* especially to *info-gap's robustness model*.



Readers with no great interest in the modeling aspects the *Maximin model* can proceed directly to the next chapter where I take up the main topic of this book, namely *robustness against severe uncertainty*.

6.10 Conservatism

It is hard to imagine what individuals' lives would have been like, or what kind of socio/political/ economic systems we would have constructed, had we been guided by policies based on "pure" worst-case analysis.

It seems that adopting such an approach would have resulted in unimaginably costly systems perhaps even in a complete paralysis of our lives and the infrastructure that we depend on.

It is important to note, therefore, that the extreme conservatism inherent in the *worst-case analysis*, hence the unimaginable extreme consequences of basing decision-making on the *Maximin* were recognized right from the start:

> It should be mentioned that Wald advocated the minimax principle in a tentative way and because of certain formal advantages. I am informed that he was still interested in finding a less conservative and more satisfactory principle for statistical inference.
>
> To my mind, it is somewhat doubtful if principles of this kind are really applicable in the social sciences [26]. They are without any doubt applicable in industrial applications (quality control, etc.)
>
> Tintner (1952, p. 24)

and

> It should also be remarked that the minimax principle even if it is applicable leads to an extremely conservative policy.
>
> Tintner (1952, p. 25)

Tintner's reference to the Maximin's applicability to the"social sciences" is interesting in view of the role that *Wald's Maximin Principle* is taken to play in Rawls' argument in the (1971) *Theory of Justice* and the ensuing criticism of Rawls' argument by Harsanyi's (1975, p. 595). Harsanyi's critique (which does not undermine Rawls' argument) is interesting for its colorful description of the ramifications of the Maximin:

> If you took the maximin principle seriously then you could not ever cross a street (after all, you might be hit by a car); you could never drive over a bridge (after all, it might collapse); you could never get married (after all, it might end in a disaster), etc. If anybody really acted this way he would soon end up in a mental institution.
>
> Conceptually, the basic trouble with the maximin principle is that it violates an important continuity requirement: It is extremely irrational to make your behavior wholly dependent on some highly unlikely unfavorable contingencies regardless of how little probability you are willing to assign to them.
>
> Harsanyi (1975, p. 595)

To illustrate this last point, consider the following simple payoff table and the associated security levels of Player 1's alternatives:

		Player 2									
		A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	SL
Player 1	a_1	3	$\boxed{2}$	5	6	$\boxed{2}$	3	4	3	4	2
	a_2	39	58	$\boxed{0}$	78	97	30	40	50	60	0

Table 6.2: Security levels for Player 1's alternatives

All the payoffs associated with alternative a_2 are much larger than the corresponding payoffs associated with alternative a_1, except the one in column A_3. This single payoff renders the security level of a_2 less attractive (smaller) than the security level of a_1, meaning that the Maximin rules that, a_1 is more robust than a_2.

A similar position is attributed by French (1988) to those who question the merit of appealing to a performance measure that is based on so grim (pessimistic) a view of uncertainty:

> It is, perhaps, a telling argument against Wald's criterion that, although there are many advocates of this approach, there are few, if any, of the maximax return criterion. Why is it more rational to be pessimistic than optimistic? An old proverb may tell us that 'it is better to be safe than sorry', and it is true that Wald's criterion is as cautious as possible: but one must remember also that 'nothing ventured, nothing gained'.
>
> French (1988, p. 37)

But, it is important to realize that the Maximin analysis, as a tool for the modeling/analysis of situations of uncertainty, must not be judged in isolation from the context in which it is used. In other words, the fact that radical consequences are inherent in a (pure) Maximin analysis must be juxtaposed against the fact that the Maximin provides some sort of "weapon" against the unforseen risks inherent in uncertainty. This point is explained extremely eloquently in the following passages that I quoted already in Chapter 5:

> The conventional approach to decision under uncertainty is based on expected value optimization. The main problem with this concept is that it neglects the worst-case effect of the uncertainty in favor of expected values. While acceptable in numerous instances, decisions based on expected value optimization may often need to be justified in view of the worst-case scenario. This is especially important if the decision to be made can be influenced by such uncertainty that, in the worst case, might have drastic consequences on the system being optimized. On the other hand, given an uncertain effect, some worst-case realizations might be so improbable that dwelling on them might result in unnecessarily pessimistic decisions. Nevertheless, even when decisions based on expected value optimization are to be implemented, the worst-case scenario does provide an appropriate benchmark indicating the risks.
>
> Rustem and Howe (2002, p. xiii)

> Through its inherent pessimism, the minimax strategy may lead to a serious deterioration of performance. Alternatively, the realization of the worst-case scenario may result in an unacceptable performance deterioration for the strategy based on expected value optimization. *Neither minima nor expected value optimization provide a substitute for wisdom*. At best, they can be regarded as risk management tools for analyzing the effects of uncertain events.
>
> Rustem and Howe (2002, p. xiv)

Of course, if the performance measure based on the Maximin's grim view of uncertainty/variability is at odds with your view on what robustness against uncertainty ought to be based on, then . . . don't use this paradigm as a framework for your robustness model.

But at the same time, it is important to appreciate that the Maximin's conservatism is not set in stone. The *Maximin model* does not dictate to the analyst/modeler how the model's constituent elements, should be defined in a given application. These are determined by the analyst, which means that the analyst can have some input into determining how "conservative" the *Maximin model* would turn out to be. For instance, the analyst may decide that in a particular application positing a *Noah's flood* as the worst scenario is uncalled for and that positing a 100-year flood should do.

As a matter of fact, as I show below, the very intuitive *Size Criterion* can be regarded as a mitigated, more genial *Maximin model* (in disguise), with no trace of "conservatism" about it!

6.11 Maximax and Minimin

The optimists among us would surly be pleased that *classical decision theory* offers a model where Nature is a *cooperating,* rather than an *antagonistic* player — as is the case in the *Maximin model.*

This model is stated formally as follows:

$$\begin{array}{ll} \text{Classic format} & \text{MP format} \\ z^* := \max_{x\in X} \max_{s\in S(x)} f(x,s) \equiv & \max_{s\in X, s\in S(x)} f(x,s) \end{array} \tag{6.42}$$

and it represents the following game:

Maximax Game

- **Step 1:** DM selects a decision $x \in X$.
- **Step 2:** Nature selects the best state in $S(x)$, call it $s(x)$.
- **Step 3:** A reward $f(x, s(x))$ is awarded to DM.

where it is assumed that DM tries to maximize her payoff $f(x, s)$.

And in case the DM's objective is to minimize the cost, we have a *Minimin model:*

$$\begin{array}{ll} \text{Classic format} & \text{MP format} \\ z^* := \max_{x\in X} \min_{s\in S(x)} f(x,s) \equiv & \min_{s\in X, s\in S(x)} f(x,s) \end{array} \tag{6.43}$$

It represents the following game:

Minimin Game

- **Step 1:** DM selects a decision $x \in X$.
- **Step 2:** Nature selects the best state in $S(x)$, call it $s(x)$.
- **Step 3:** A cost $f(x, s(x))$ is incurred to DM.

where it is assumed that DM tries to minimize her cost $f(x, s)$.

Note that in both cases the game can be phrased as a "conventional" optimization problem, namely a game involving a *single player*:

$$\max_{x\in X} \max_{s\in S(x)} f(x,s) \;\equiv\; \max_{x\in X, s\in S(x)} f(x,s) \tag{6.44}$$

and

$$\min_{x\in X} \min_{s\in S(x)} f(x,s) \;\equiv\; \min_{x\in X, s\in S(x)} f(x,s) \tag{6.45}$$

The constrained versions of these models would be obtained as demonstrated above, namely in the same manner as the constrained *Maximin model* was formulated above. Note however, that here the hallmark clause signaling that the best-case analysis is constraints driven in not $\forall s \in S(x)$ — as is the case in the *Maximin model* — but rather "for at least one s in $S(x)$.

For instance, the constrained Minimin model where the best-case analysis is conducted only with respect to the constraints, is as follows:

$$\max_{x\in X} \{f(x) : \; con(x,s) \text{ is satisfied for at least one } s \in S(x)\} \tag{6.46}$$

To illustrate, consider the test model

$$\beta(q,\tilde{u}) := \min \left\{ \alpha \geq 0 : r^* \leq \max_{u \in U(\alpha,\tilde{u})} r(q,u) \right\} \ , \ q \in Q \tag{6.47}$$

Observe that from the definition of $\max$ it follows that this model is equivalent to the following model

$$\min \ \{\alpha \geq 0 : r^* \leq r(q,u) \text{ for at least one } u \in U(\alpha,\tilde{u})\} \tag{6.48}$$

Hence, (6.47) is a *Minimin model.*

Readers who are familiar with *info-gap decision theory* should have noticed that (6.47) is *info-gap's opportuneness model*. The above analysis shows then that *info-gap's opportuneness model* is an instance[7] of the generic Minimin model. Expressed in terms of the classic format, this model is as follows:

$$\min_{\alpha \geq 0} \ \min_{u \in U(\alpha,\tilde{u})} f(\alpha,u) \ \ \text{where} \ \ f(\alpha,u) := \begin{cases} \alpha & , \ r^* \leq r(q,u) \\ -\alpha & , \ r^* > r(q,u) \end{cases} , \ \alpha \geq 0, u \in U(\alpha,\tilde{u}) \tag{6.49}$$

More on this model in the second part of the book.

6.12 Variations on a theme

The concern over the Maximin's inherent "conservatism" has prompted the formulation of modified versions of the generic Maximin. The best known adaptations are *Savage's Minimax regret* paradigm and *Hurwicz's optimism-pessimism* paradigm.

· Savage's Minimax regret paradigm is:

$$z^* := \min_{x \in X} \max_{s \in S(x)} g(x,s) \tag{6.50}$$

where

$$g(x,s) := \left\{ \max_{x \in X} f(x,s) \right\} - f(x,s) \tag{6.51}$$

Note that $g(x,s)$ is non-negative. It stipulates the amount difference between the best payoff (over all decisions) for state s and the payoff $f(x,s)$. The *security level* of decision x is then

$$SL(x) := \max_{s \in S(x)} g(x,s) \tag{6.52}$$

and it can be regarded as a robustness measure: the smaller the better.

To illustrate, consider the payoff table shown in Figure 6.3 and the associated security levels for Player 1. Note that according to the Maximin rule, a_1 is more robust than a_2.

To see which alternative is more robust according to the Minimax rule we compute, for each $A_j, j = 1,2,\ldots,10$, the corresponding maximum payoff. The results are as follows:

[7] See section 6.17

		Player 2										
		A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	SL
Player 1	a_1	3	2	5	6	2	3	4	3	4	4	2
	a_2	39	58	0	78	97	30	40	50	60	80	0

Table 6.3: Payoff table for Player 1

j	1	2	3	4	5	6	7	8	9	10
$\max\limits_{i=1,2} f(i,j)$	39	58	5	78	97	30	40	50	60	80

We obtain the regrets by subtracting the payoffs from these maximum values, column by column. The end result is the regret table for Player 1, including the security levels, shown in 6.4.

		Player 2										
		A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	SL
Player 1	a_1	36	56	0	72	95	27	36	47	57	76	95
	a_2	0	0	5	0	0	0	0	0	0	0	5

Table 6.4: Regret table for Player 1

Note that according to the *Minimax rule,* the security level of an alternative for Player 1 is the largest regret pertaining to this alternative, and that the smaller the security level the better, namely the more robust the alternative is. In this case, therefore, the *Minimax Regret rule* deems that the most robust alternative is alternative a_2. Its security level is equal to 5. The implication is then that, regardless of what alternative will be selected by Player 2, if this alternative is selected by Player 1, the regret for Player 1 will not exceed 5.

Should it be desired/required, it would be possible to *scale* the regret, say by dividing it by $\max\limits_{x\in X} f(x,s)$ (assuming that it is strictly positive). In this case the regret would be phrased as follows:

$$g(x,s) := \frac{\left\{\max\limits_{x\in X} f(x,s)\right\} - f(x,s)}{\max\limits_{x\in X} f(x,s)} \tag{6.53}$$

In cases where the maximization required by $\max\limits_{x\in X} f(x,s)$ proves difficult, one may seek to measure the regret with respect to some bench-mark function of s rather than $\max\limits_{x\in X} f(x,s)$. In this case the *Minimax regret model* would be phrased as follows:

$$\min_{x\in X}\ \max_{s\in S(x)}\ \{b(s) - f(x,s)\} \tag{6.54}$$

where $b(s)$ represents a bench-mark payoff pertaining to state s, say an approximation of $\max\limits_{x\in X} f(x,s)$.

· Hurwicz's optimism-pessimism paradigm:

$$z^*(\alpha) := \max_{x\in X}\left\{\alpha \min_{s\in S(x)} f(x,s) + (1-\alpha)\max_{s\in S(x)} f(x,s)\right\}\ ,\ \alpha\in[0,1] \tag{6.55}$$

where α, the famous optimism-pessimism index, measures the decision maker's optimism/pessimism with regard to the uncertain state of nature.

In case you are unsure as to how optimistic-pessimistic you are, here is the recipe for calculating this. Consider the following simple payoff table:

	A_1	A_2
a_1	1	0
a_2	v	v

and ask yourself:

> If I am the row-player, what value of v will render me *indifferent* to whether alternatives a_1 or a_2 is the case?

Once you determined the value of v according to your preference, you can compute your optimisim-pessimism index with this recipe:

$$\alpha = 1 - v \tag{6.56}$$

Note that if you are "rational", the value you will assign v will be in the range $[0, 1]$, and so will be your magic value of α.

Remark

One must make sure, when using Savage's Minimax regret model and *Hurwicz's optimism-pessimism model,* that the operations of addition and scalar multiplication are "sound". The point is that both models require that these operations be performed on the payoffs. For instance, in the case of *Savage's Minimax regret model,* to compute regrets from payoff we *subtract* payoffs from payoffs. It is important therefore to insure that this operation is meaningful, in the sense that the results of the subtraction operation make sense in the context of the problem under consideration.

Note that this problem does not arise in the context of *Wald's Maximin model* because no arithmetic operations are called for by the *Maximin Recipe,* only comparisons. The comparison operations are unproblematic because the assumption is that the payoffs yielded are real numbers so that *the larger the better* preference criterion is straightforward.

6.13 Conservatism revisited

The fact that the Maximin's conservatism has prompted the formulation of adapted versions thereof — such as those discussed in the preceding section — is often (mis)taken to imply that its conservatism is immutable.

But this is not entirely correct. It is important to appreciate that there are definitely ways available to control the Maximin's conservatism, or as I say above, the Maximin's conservatism is not set in stone. This is particularly pertinent to *robustness.* It is up to the modeler/analyst to decide how conservative/liberal a particular *robustness Maximin model* should be with respect to the objective function and the constraints. The point to note here is that standard modeling measures can be used to *control* the degree, namely the severity, of the robustness that one would expect/desire the *Maximin model* to provide. After all, it is the modeler/analyst who decides what the objective function and constraints of the *Maximin model* would be. These are not dictated by the *Maximin model* itself!

So, in a nutshell, by virtue of prescribing the choice of the best worst outcome the *Maximin paradigm* by definition tilts the analysis in a conservative direction. But, at the same time, because it allows the modeler/analyst to decide what the "best case" and "worst case" ought to be, it provides for a significant measure of flexibility.

6.13.1 Example

Consider the following simple constrained Maximin model, given in its MP format:

$$z^* := \max_{x \in X} \{g(x) : r^* \leq r(x,s), \forall s \in S(x)\} \tag{6.57}$$

where g is a real valued function on X and r is a real valued function on $X \times \mathcal{S}$ for some $\mathcal{S}$ that contains $\cup_{x \in X} S(x)$.

Now, what would you do in a situation where the problem specified by this model has no feasible, let alone optimal, solutions? That is, what would you do if there is no $x \in X$ such that $r^* \leq r(x,s), \forall s \in S(x)$?

As Fred[8] would no doubt bear me out, such situations are commonplace in practice. This means of course that, this issue is important not only for its methodological import, it is important for its obvious practical implications.

There are two obvious ways to deal with this difficulty:

- We can advise whoever it may concern that, as things stand, the problem has no solution.
- We can ease the constraint imposed on the problem, say by decreasing somewhat the value of r^*, to the point where a solution for the problem is eventually found.

I shall not argue here the pro's and con's of these obvious options. My objective in bringing them up is to illustrate two things:

- First, that requirements such as $r^* \leq r(x,s), \forall s \in S(x)$ are often too exacting.
- Second, that there are ways to "relax" worst-case requirements such as $r^* \leq r(x,s), \forall s \in S(x)$.

For instance, consider the following "mitigation scheme".

Suppose that in view of the difficulty that the decisions $x \in X$ have in meeting the worst-case requirement $r(x,s) \leq r^*, \forall s \in S(x)$, we introduce "wavers". The sort of thing (exceptions, subsidies, and so on) that governments often introduce to help business/industry cope with government regulations, including payment of taxes, meeting pollution standards, etc.

That is, recognizing that some values of the state $s \in S$ might be too exacting for the decisions $x \in X$ to handle on their own, we propose that rather than decrease the value of r^* by a *constant*, say $\delta > 0$, we decrease the value of r^* by varying the value that is contingent on s. So, consider the following scheme:

$$r^* \leq r(x,s) + \Delta(s)\ ,\ s \in S(x) \tag{6.58}$$

[8] Internationally Known Expert on Robust Decision-Making in the Face of Severe Uncertainty

where Δ is a real valued function on S such that $\Delta(s) \geq 0, \forall s \in \mathcal{S}$. The more "troublesome" s is, the bigger the waver granted to $\Delta(s)$.

The *Maximin model* corresponding to this scheme is then as follows:

$$z^{\circ} := \max_{x \in X} \{g(x) : r^* \leq r(x,s) + \Delta(s), \forall s \in S(x)\} \tag{6.59}$$

Furthermore, we may even face situations where it would make sense to allow the waver depend on the decision $x \in X$, in which case the Maximin model would be as follows:

$$z' := \max_{x \in X} \{g(x) : r^* \leq r(x,s) + \Delta(x,s), \forall s \in S(x)\} \tag{6.60}$$

where now formally Δ is a real valued function on $X \times \mathcal{S}$ such that $\Delta(x,s) \geq 0, \forall s \in X, s \in S(x)$.

G'day Moshe

I recall that in one of your recent lectures you indicated that there might be cases where it would make a lot of sense to allow $\Delta(x,s)$ t be *negative*. I do not remember the details, but your arguments were very persuasive.

I therefore suggest that you mention this option here.

Cheers,
Fred
Internationally renowned Expert on Robust Decision-Making Under Severe Uncertainty

The reason that Fred raised this point is that just as there could be some states $s \in \mathcal{S}$ that would be "troublesome", there could be other states $s \in \mathcal{S}$ that not only would not be "troublesome" but would actually be "auspicious".

The implications that this point has for the question of the *Maximin model*'s conservatism is illustrated by these two models:

Original Maximin model	Modified Maximin model	
$\max_{x \in X} \{g(x) : r^* \leq r(x,s), \forall s \in S(x)\}$	$\max_{x \in X} \{g(x) : r^* \leq \hat{r}(x,s), \forall s \in S(x)\}$	(6.61)

where $\hat{r}$ is a modified ("relaxed" ?) version of r designed to deal with the conservative nature of the Maximin model. In the above scheme $\hat{r}(x,s) = r(x,s) + \Delta(x,s)$.

Needless to say, the same argument applies to a possible modification of the *objective function* (payoff) of the *Maximin model* where the objective is similar. Dealing with *Maximin model*'s innate conservatism.

I shall return to this point in the next chapter in my a discussion of robustness models that are based on this type of schemes.

6.14 Maximin models in disguise

I was surprised to learn over the past seven years that many analysts who are familiar with Wald's Maximin model are not familiar with its MP format. To put it somewhat bluntly, some senior risk analysts proceed on the assumption (hence, argue/write along these lines) that only *Maximin models* "dressed" in the official classic "dress" are *Maximin models,* namely:

$$\max_{x\in X}\ \min_{s\in S(x)}\ f(x,s) \tag{6.62}$$

Indeed, over the past five years I have had several exchanges with some senior risk analysts on whether the following model is a *Maximin model:*

$$\max\ \left\{\alpha\geq 0 : r^* \geq \max_{u\in U(\alpha,\tilde{u})} r(q,u)\right\}\ ,\ q\in Q \tag{6.63}$$

where $U(\alpha,\tilde{u})$ denotes a ball of radius α around $\tilde{u}$.

The argument put to me was the following: how could (6.63) be a *Maximin model* when it is so obviously different from (6.62)?

It was even suggested to me by senior risk analysts that (6.63) is a *Maximax model,* not a *Maximin model*!!!

This led me to coin the phrase MAXIMIN MODELS IN DISGUISE so as to designate models that do not "look" like *Maximin models* but are in fact *Maximin models.*

This raises the following important question:

How do we determine whether a given model is a Maximin model?

In the following two subsections I take up this question, addressing it from two angles that can be described broadly as "formal" and "informal".

In preparation I note that it will be convenient to conduct the discussion in the framework of the constrained *Maximin model* whose two (equivalent) formats are as follows:

$$\text{Classic format:}\quad \max_{x\in X}\ \min_{s\in S(x)} \{f(x,s) : con(x,s), \forall s\in S(x)\} \tag{6.64}$$

$$\text{MP format:}\quad \max_{x\in X, v\in\mathbb{R}} \{v : v\leq f(x,s), con(x,s), \forall s\in S(x)\} \tag{6.65}$$

where $con(x,s)$ represents a set of constraints on x and s.

So with regard to the model specified by (6.63), the question is: is this model an instance[9] of the generic constrained Maximin model specified by (6.64)-(6.65)?

6.14.1 A formal approach

Maximin models in disguise invariably appear on the scene as (constrained) optimization models of the generic form

$$\max_{y\in Y}\ \{g(y) : con(y)\} \tag{6.66}$$

where $con(y)$ represents . . . a set of constraints on the decision variable y.

The point to note here is that as:

$$\max_{y\in Y}\ \{g(y) : con(y)\} \equiv \max_{y\in Y, w\in\mathbb{R}}\ \{w : w\leq g(y), con(y)\} \tag{6.67}$$

[9] See section6.17

it is more convenient in such cases to relate such optimization models to the MP format of the Maximin model, rather than the Classic format.

So, let us place the two models one next to the other, so as to be able to compare them first of all "visually":

$$\text{Maximin (MP format):} \ \max_{x\in X, v\in\mathbb{R}} \{v : v \leq f(x,s), con(x,s), \forall s \in S(x)\} \tag{6.68}$$

$$\text{Test model:} \ \max_{y\in Y, w\in\mathbb{R}} \{w : w \leq g(y), con(y)\} \tag{6.69}$$

Clearly, the main difference between these two models is in the clause $\forall s \in S(x)$ in the MP format. The significance of this clause is twofold:

· It introduces an *exogenous* variable, s, and its range of feasible values, $S(x)$, into the model.

· It incorporates a *worst-case analysis* into the model via the $\forall s \in S(x)$ requirement.

Note that the value of the exogenous variable s is not under the decision maker's control (the $\max$ operator). Only its range of feasible values, namely $S(x)$, is controlled by the DM.

In short, for the Test Model (6.69) to be a Maximin model, it must exhibit these two properties. In other words,

· It must include an *exogenous parameter.*

· It must include a $\forall$ requirement with regard to some of the constraints and/or the $w \leq g(y)$ requirement.

For example, consider this Test Model:

$$\max \left\{\alpha \geq 0 : r^* \geq \max_{u\in U(\alpha,\tilde{u})} r(q,u)\right\} \tag{6.70}$$

Is this a *Maximin model?*

The key point here is that if the $\max$ inside the curly brackets exists, then

$$r^* \geq \max_{u\in U(\alpha,\tilde{u})} r(q,u) \equiv r^* \geq r(q,u) \ , \ \forall u \in U(\alpha,\tilde{u}) \tag{6.71}$$

Hence,

$$\max \left\{\alpha \geq 0 : r^* \geq \max_{u\in U(\alpha,\tilde{u})} r(q,u)\right\} \equiv \max\left\{\alpha \geq 0 : r^* \geq r(q,u), \forall u \in U(\alpha,\tilde{u})\right\} \tag{6.72}$$

We therefore conclude that this is indeed a *Maximin model.* Its classic and MP formats are as follows:

Constrained Classic format:

$$\max_{\alpha\geq 0} \min_{u\in U(\alpha,\tilde{u})} \{\alpha : r^* \geq r(q,u), \forall u \in U(\alpha,\tilde{u})\} \tag{6.73}$$

Constrained MP format:

$$\max_{\alpha\geq 0, v\geq 0} \{v : v \leq \alpha : r^* \geq r(q,u), \forall u \in U(\alpha,\tilde{u})\} \tag{6.74}$$

Note that because the objective function is independent of u, both models can be immediately simplified, to yield the model specified by (6.75).

Let us now consider the following similar, but more subtle, Test Model:

$$\max \left\{\alpha \geq 0 : r^* \geq \min_{u \in U(\alpha,\tilde{u})} r(q,u)\right\} \tag{6.75}$$

Is this a *Maximin model?*

The key point here is that if the $\min$ inside the curly brackets exists, then

$$r^* \geq \min_{u \in U(\alpha,\tilde{u})} r(q,u) \equiv r^* \geq r(q,u) \ , \ \text{for at least one } u \in U(\alpha,\tilde{u}) \tag{6.76}$$

Hence,

$$\max \left\{\alpha \geq 0 : r^* \geq \min_{u \in U(\alpha,\tilde{u})} r(q,u)\right\} \equiv \max \{\alpha \geq 0 : r^* \geq r(q,u), \\ \text{for at least one } u \in U(\alpha,\tilde{u})\} \tag{6.77}$$

We therefore conclude that this is not a *Maximin model,* it is a *Maximax model!* More on this shortly.

And how about this Test Model:

$$\max \left\{\alpha \geq 0 : r^* \leq \min_{u \in U(\alpha,\tilde{u})} r(q,u)\right\} \tag{6.78}$$

Is this a *Maximin model?*

The key point here is that if the $\min$ inside the curly brackets exists, then

$$r^* \leq \min_{u \in U(\alpha,\tilde{u})} r(q,u) \equiv r^* \leq r(q,u) \ , \forall u \in U(\alpha,\tilde{u}) \tag{6.79}$$

Hence,

$$\max \left\{\alpha \geq 0 : r^* \leq \min_{u \in U(\alpha,\tilde{u})} r(q,u)\right\} \equiv \max \{\alpha \geq 0 : r^* \leq r(q,u), \forall u \in U(\alpha,\tilde{u})\} \tag{6.80}$$

We therefore conclude that this is indeed a *Maximin model.* Its (constrained) classic and MP formats are as follows:

Constrained Classic format:

$$\max_{\alpha \geq 0} \min_{u \in U(\alpha,\tilde{u})} \{\alpha : r^* \leq r(q,u), \forall u \in U(\alpha,\tilde{u})\} \tag{6.81}$$

Constrained MP format:

$$\max_{\alpha \geq 0, v \geq 0} \{v : v \leq \alpha : r^* \leq r(q,u), \forall u \in U(\alpha,\tilde{u})\} \tag{6.82}$$

Note that because the objective function is independent of u, both models can be immediately simplified, to yield (6.78).

And how about this Test Model:

$$
\begin{aligned}
&\max_{w,x,y,z} z && (6.83)\\
&s.t.\ \ 2w^2+3x^2+2y^2 \geq z+20 && (6.84)\\
&\qquad\ 4w\ \ +5x\ \ +2y \geq z-1 && (6.85)\\
&\qquad\ \ w\ \ +\ \ z\ \ +\ \ y = 8 && (6.86)\\
&\qquad\ w,x,y \geq 0 && (6.87)
\end{aligned}
$$

To test your modeling skills, demonstrate that this is indeed a *Maximin model* whose classic format is as follows:

$$
\begin{aligned}
&\max_{w,x,y} \min \{2w^2+3x^2+2y^2-20\ ,\ 4w+5x+2y+1\} && (6.88)\\
&\qquad s.t.\ \ w+x+y=8 && (6.89)\\
&\qquad\qquad w,x,y \geq 0 && (6.90)
\end{aligned}
$$

If you are a purist insisting on indicating the details of $\min$ below, you can rewrite this model as follows:

$$
\max_{w,x,y} \min_{s\in\{1,2\}} \{g_s(w,x,y) : w+x+y=8, w,x,y\geq 0\} \tag{6.91}
$$

where

$$
g_s(w,x,y) := \begin{cases} 2w^2+3x^2+2y^2-20 & ,\ \ s=1\\ 4w+5x+2y+1 & ,\ \ s=2 \end{cases} \tag{6.92}
$$

In the next subsection I consider a more easygoing approach to the modeling issue under consideration.

6.14.2 An informal approach

The structure of the generic *Maximin model* is not that intricate that one has to engage in a formal mathematical argument in order to establish whether a given model is a *Maximin model.* So, when you are presented with a Test Model and asked to determine whether this model is a *Maximin model,* simply ask ourself the following question:

Does the Test Model select the best decision prescribed by the *Maximin paradigm?*

If the answer is in the affirmative, then the model is a *Maximin model* and it should not be too difficult to give it a rigorous formal formulation in terms of the classic and/or MP formats. As for complying with the *Maximin Rule* prescription, the test model must do the following:

Maximin Rule:

· Rank decisions on the basis of their worst outcomes.

· Select the decision whose worst outcome is the best.

To show you how I would proceed to work this out consider the following.

For me, a *Maximin model* is a model giving expression to a GAME between a decision maker (DM) and an adversary, call her Nature. The arrangement is such that the DM selects her decision first whereupon, Nature responds by selecting an option that is worst for the DM. Figure 6.7 depicts the case where the DM seeks to maximize her payoff.

$$\begin{array}{ccccc} DM & & Nature & & \\ \max & & \min & & f(x,s) \\ x\in X & \rightarrow & s\in S(x) & \rightarrow & \text{payoff} \end{array}$$

Figure 6.7: Maximin game

Thus, when I examine a model with the view to ascertain whether it is a *Maximin model,* the first thing I try to establish is which construct in the model represents the DM's decision (x) and which construct represents Nature's options (s).

This done, all I have to do is to verify that the model indeed selects the option available to Nature that is worst for the DM. This requires specifying a proper payoff (objective) function (f).

Let me illustrate this approach in action by considering (again) the following simple model which, according to some senior risk analysts, does not look like a Maximin model.

$$\max \left\{ \alpha \geq 0 : r^* \geq \max_{u\in U(\alpha,\tilde{u})} r(q,u) \right\} \ , \ q \in Q \tag{6.93}$$

As I explained above, the first thing I would do is identify the two players involved in this game and their respective decision variables.

In this particular case, this task is extremely simple because each player is represented in the model by its very own $\max$ operation.

Clearly, the outer $\max$ represent the DM, so it follows that the inner $\max$ represents Nature. The respective decision variables are ρ for DM and u for nature.

Clearly, this was an extremely simple exercise.

Next, I have to determine whether this is a *Maximin game.* That is, I have to determine whether the inner $\max$ selects the worst $u \in U(\alpha,\tilde{u})$ with respect to the DM's payoff α.

To do this I observe that the requirement under consideration in this model is as follows:

$$r^* \geq r(q,x) \tag{6.94}$$

So the question is:

> Is the $u \in U(\alpha,\tilde{u})$ that maximizes $r(q,u)$ indeed the worst $u \in U(\alpha,\tilde{u})$ insofar as the DM's attempt to maximize the value of α is concerned?

The key point here is that the constraint under consideration is of the $r^* \geq r(q,x)$ type. The $\geq$ sign in this constraint means that the "smaller" $r(q,u)$ is the "better" it is. Or equivalently, the "larger" $r(q,u)$ is the "worse" it is. The conclusion is therefore this:

> By selecting the $u \in U(\alpha,\tilde{u})$ that maximizes the value of $r(q,u)$, the inner $\max$ indeed selects the worst $u \in U(\alpha,\tilde{u})$ with respect to the DM's payoff α.

My conclusion is therefore that this is indeed a Maximin model. So I confidently write:

$$\max \left\{ \alpha \geq 0 : r^* \geq \max_{u \in U(\alpha,\tilde{u})} r(q,u) \right\} \equiv \max_{\alpha \geq 0} \min_{u \in U(\alpha,\tilde{u})} f(\alpha,u) \tag{6.95}$$

where $f(\alpha,u) = \alpha$ if $r^* \geq r(q,u)$.

To complete the task, I need to determine the value of the payoff $f(\alpha,\tilde{u})$ for $u \in U(\alpha,\tilde{u})$ such that $r^* < r(q,u)$. The value to be specified for such a $f(\alpha,\tilde{u})$ should discourage the DM from selecting such a value of α.

One way to do this is to set up a Committee, so that after due deliberations, it will report on what is the most appropriate value to be assigned to the payoff $f(\alpha,\tilde{u})$.

That's fine with me.

However, until we hear from the Committee, let us simply use the penalty $-\rho$ for this purpose, namely let us formally define

$$f(\alpha,u) := \begin{cases} \alpha & , \ r^* \geq r(q,u) \\ -\alpha & , \ r^* < r(q,u) \end{cases} , \ \alpha \geq 0, u \in U(\alpha,\tilde{u}) \tag{6.96}$$

Let us now repeat the exercise with the following Test Model:

$$\min \left\{ \alpha \geq 0 : r^* \leq \max_{u \in U(\alpha,\tilde{u})} r(q,u) \right\} , \ q \in Q \tag{6.97}$$

What kind of model is this?

Based on the preceding analysis, we can view this model as a game between the the DM and Nature. As above, the DM selects a value for α whereupon Nature selects a value for u. Since the DM is minimizing the value of its payoff α, the game is definitely not a *Maximin game.*

To determine what kind of a game this is, we examine whether the selection of $u \in U(\alpha,\tilde{u})$ by the inner $\max$ player is the worst or best value in $u \in U(\alpha,\tilde{u})$ with respect to the DM's payoff.

Observe then that the constraint under consideration here is of the form

$$r^* \leq r(q,u) \tag{6.98}$$

Hence, in this context, the "larger" $r(q,u)$ is the "better" it is. Thus, by selecting the largest value of $r(q,u)$, the inner $\max$ player selects the BEST value of $u \in U(\alpha,\tilde{u})$ with respect to the DM's payoff. In other words, in this case Nature is "cooperating" with the DM.

I thus conclude that this is a *Minimin game,* so that I confidently write:

$$\min \left\{ \alpha \geq 0 : r^* \leq \max_{u \in U(\alpha,\tilde{u})} r(q,u) \right\} \equiv \min_{\alpha \geq 0} \min_{u \in U(\alpha,\tilde{u})} f(\alpha,u) \tag{6.99}$$

where $f(\alpha,u) = \alpha$ if $r^* \leq r(q,u)$.

To discourage the DM from selecting an α such that $r^* > r(q,u), \forall u \in U(\alpha,\tilde{u})$, we can set $f(\alpha,u) = \infty$ for such a choice of α. In short, we can formally define

$$f(\alpha,u) := \begin{cases} \alpha & , \ r^* \leq r(q,u) \\ \infty & , \ r^* > r(q,u) \end{cases} \ , \ \alpha \geq 0, u \in U(\alpha,\tilde{u}) \tag{6.100}$$

The Committee that is currently deliberating on the choice for f for the preceding *Maximin model* may decide to modify the choice of ∞ as a penalty in (6.100).

Remark:
Note that if $U(\alpha,\tilde{u})$ denotes a ball of radius α around $\tilde{u}$, then $r^* \leq r(q,\tilde{u})$ implies that the optimal (minimal) payoff to DM is equal to 0. More generally, in this case the optimal payoff to DM is equal to the radius of the smallest ball centered at $\tilde{u}$ that contains a u such that $r^* \leq r(q,u)$.

I discuss such models in the second part of the book in connection with *info-gap's opportuneness model.*

6.15 Local vs global models

The distinction that we drew in Chapter 5 between *local* and *global* worst-case analysis suggests an analogous distinction, between *local* and *global* Maximin models.

Since the analogy is so transparent, I shall not dwell on the distinction between *local* and *global* Maximin models, except to illustrate it.

Let us begin by observing that in the context of the generic Maximin model

$$\max_{x\in X} \min_{s\in S(x)} f(x,s) \tag{6.101}$$

the worst-case analysis is represented by the inner $\min\limits_{s\in S(x)}$ operation, whose domain is $S(x)$.

That is, in this context the range of variability of the state s associated with decision $x \in X$ is the set $S(x)$.

The situation is similar in the context of other Maximin models. For instance, consider the constrained Maximin model whose MP format is as follows:

$$\max_{x\in X} \{g(x) : r^* \leq r(x,s), \forall S(x)\} \tag{6.102}$$

Here the worst-case analysis is represented by the clause $\forall s \in S(x)$, whose domain is also $S(x)$.

In both cases the distinction between *local* and *global Maximin models* is due to the structure of the decision variable $x \in X$ and the structure of the sets $S(x), x \in X$. It should be more instructive to examine first the *local* model. But to put things in context consider this obvious *global* Maximin model:

$$\max_{x\in X} \min_{s\in S} f(x,s) \tag{6.103}$$

and this one:

$$\max_{x\in X} \{g(x) : r^* \leq r(x,s), \forall s \in S\} \tag{6.104}$$

Take note that in both cases the state space over which the worst-case analysis is conducted, namely S, is "fixed". That is, it is independent of the decision $x \in X$.

6.15.1 Local models

Consider the case where $S(x)$ is a neighborhood of some set $\mathcal{S}$ around some state $\tilde{s}(x) \in \mathcal{S}$.

In this case, we can rewrite the above two models as follows:

$$\max_{x \in X} \min_{s \in \mathcal{N}(\rho(x), \tilde{s}(x))} f(x, s) \tag{6.105}$$

and

$$\max_{x \in X} \{g(x) : r^* \leq r(x, s), \forall s \in \mathcal{N}(\rho(x), \tilde{s}(x))\} \tag{6.106}$$

where $\rho(x)$ denotes the radius of the neighborhood around the state $\tilde{s}(x)$ associated with decision $x \in X$.

These Maximin models are *local* in the sense that for each value of the decision, $x \in X$, the worst-case analysis is conducted locally in the neighborhood of a given $s \in \mathcal{S}$, namely $\tilde{s}(x)$.

For example, the generic *Radius of Stability* model

$$\rho(q, \tilde{s}) := \max_{\rho \geq 0} \{\rho : s \in P_{stable}(q), \forall B(\rho, \tilde{s})\} \ , \ q \in Q \tag{6.107}$$

is a typical (degenerate) *local Maximin model* — degenerate in the sense that $x = \rho, g(\rho) = \rho$ and that $\tilde{s}(\rho)$ is independent of ρ.

Observe that for each decision $\rho \geq 0$, the worst case analysis with respect to $s \in \mathcal{S}$ is conducted on a neighborhood of radius ρ around $\tilde{s}$.

In words:

> The *Radius of Stability* of system q, denoted by $\rho(q, \tilde{s})$, is equal to the radius of the largest ball centered at $\tilde{s}$ all of whose states are stable.

Although the graphic illustration of this model has already been given in Chapter 3, to emphasize this point I repeat it here as well (see Figure 6.8).

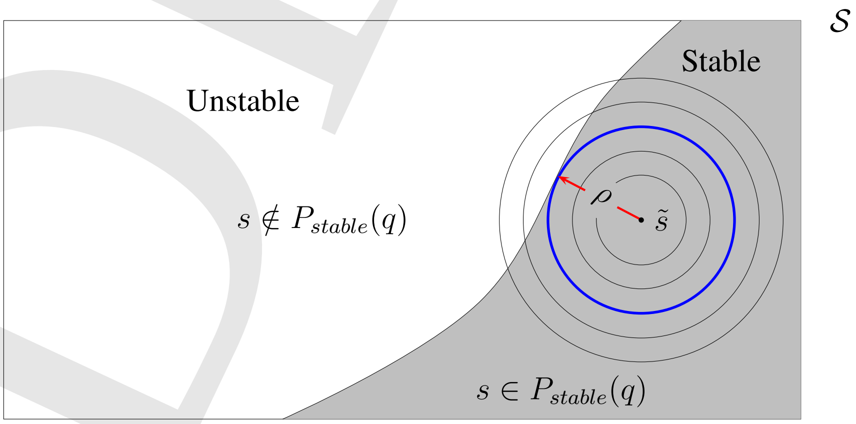


Figure 6.8: Radius of Stability of system q at $\tilde{s}$

6.15.2 Global models

Global Maximin models are models where the worst-case analysis is conducted over the entire set of possible/plausible values of the parameter of interest. As indicated above, the *Maximin models*

$$\max_{x\in X}\min_{s\in S} f(x,s) \tag{6.108}$$

and

$$\max_{x\in X}\{g(x): r^* \leq r(x,s), \forall s\in S\} \tag{6.109}$$

are manifestly global models, because in both cases the space over which the worst-case analysis is conducted, namely S, is "fixed". It is independent of the decision $x\in X$.

But the point to note is that *Maximin models* of the form

$$\max_{x\in X}\min_{s\in S(x)} f(x,s) \tag{6.110}$$

and

$$\max_{x\in X}\{g(x): r^* \leq r(x,s), \forall s\in S(x)\} \tag{6.111}$$

can also be global, despite the fact that the worst-case analysis for a given decision $x\in X$ is conducted on a set $S(x)$ whose structure and size may depend on x.

The determining factor here is whether the $S(x)$ represents the set of all the possible/plausible values of the parameter of interest pertaining to decision x. For example, consider again the "bottleneck problem" depicted in Figure 6.9 (see section 6.2).

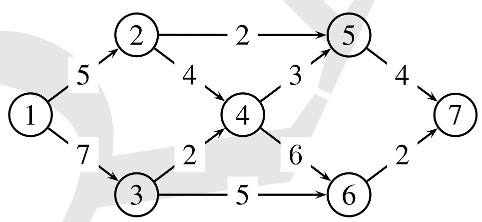


Figure 6.9: Bottleneck problem

The task is to find the best path from node 1 to node 7 in the network, where the labels on the arcs designate their capacity. For instance, the capacity of the arc from node 2 to node 5 is equal to 2. The "bottleneck" version of the problem dictates that the capacity of a *path* is equal to the arc with the lowest capacity on that path. Hence, the capacity of the path (1,2,5,7) is equal to the capacity of arc (2,5) which is equal to 2.

The implication is that the best path from node 1 to node 7 is a path from node 1 to node 7 whose arc has the largest (over all paths) smallest (over all arcs on a path) capacity. This gives rise to the following *Maximin model:*

$$k^* := \max_{p\in P(1,7)}\min_{a\in A(p)} K(a) \tag{6.112}$$

where

· $P(1,7)$ = set of all feasible paths from node 1 to node 7.
· $A(p)$ = collection of arcs on path $p \in P(1,7)$, organized as a set.
· $K(a)$ = capacity of arc a.

Note that in this framework the worst-case analysis for each decision $p \in P(1,7)$ is conducted on a subset $A(p)$ of all possible arcs on the network. Yet, the worst-case analysis with respect to each $p \in P(1,7)$ is *global:* it is conducted over all the elements of $A(p)$, which are the only possible arcs associated with path p.

Final example.

Consider the *Maximin model*

$$t^* := \max_{C \in W} \min_{c \in Cities(C)} t(c, C) \tag{6.113}$$

where

· W = a (finite) set of countries.
· $Cities(C)$ = a small set of cities in country $C \in W$.
· $t(c, C)$ = average daily temperature on January 1 in city c in country C (see footnote[10]).

The story behind this Maximin mode is as follows:

· You drew first prize in a lottery winning a visit to an exotic city, all expenses paid!
· You choose the country (C) whereupon the Prize Committee will decide the on the city (c) in country C that you will visit.
· The visit will take place in the first week of January.
· Your trusted travel agent advised you that based on the Prize Committee's past decisions, and your aversion to cold weather, it would be best for you to select the country whose coldest city is the warmest (in January).

Once you selected a country C, the only relevant cities for the analysis are those in the set $Cities(C)$. This means that the worst-case analysis is conducted on the entire $Cities(C)$ set, or in other words the worst-case analysis is *global*. The implication is therefore that the *Maximin model* is global even though the worst-case analysis is not conducted over the set of all the cities associated with the Prize, namely $\bigcup_{C \in W} Cities(C)$.

6.16 Problems vs Models

I hope that the reader has noticed that in the preceding sections I consistently (and I might add deliberately) used the term "model" rather than "problem" in reference to specific "expressions" such as this:

$$\min \left\{ \alpha \geq 0 : r^* \leq \max_{u \in U(\alpha, \tilde{u})} r(q, u) \right\} \ , \ q \in Q \tag{6.114}$$

So the question is this. Is this a "problem" or a "model"?

[10]The $t(c, C)$ notation reflect the fact that different cities may have identical names (e.g. Melbourne, Australia, vs Melbourne, Florida, USA).

The reason that I am raising this question is that often the same “problem” or “task” can be represented by different models, sometimes models that are significantly different from one another.

For example, the famous *traveling salesman problem* has various considerably different formulations.

So, my point is that, (sometimes) it is important to distinguish between the problem itself, say the “traveling salesman problem”, and a particular model used to describe it.

Some scholars may no doubt argue that this distinction is of no consequence. Indeed, that there is no reason to make a fuss about this whole issue, because the fact is that what we have to deal with are . . . problems or tasks.

Other scholars may strongly disagree, arguing that there are no such “entities” as “problems”. For, what we call “problems” are products of our intellect, so that to be able to refer to a “problem” we need a “model” to represent it.

The reader may no doubt wonder what prompted me to open this can of worms here. What relevance does the fundamental question of “problem” vs “model” possibly have for our discussion?

So, let me point out that I am raising this issue here to make it clear that this distinction is indeed very relevant to our discussion here. Because, insofar as our discussion goes, it is indeed important to appreciate that the same “problem” may have various formulations (models) — *Maximin* and non-Maximin formulations. Furthermore that sometimes even the *Maximin* formulations (models) can be considerably different.

G’day Moshe

I suggest that instead of philosophizing on this matter, simply give an example to illustrate it.

Fred

Internationally renowned Expert on Robust Decision Under Severe Uncertainty

Good idea, Fred!

6.16.1 Illustrative examples

The first example illustrates some of the issues that come into play when we seek to model the *Size Criterion* — discussed in Chapter 3 in connection with the distinction between *local* and *global* robustness.

Example

Consider the following very simple problem:

> For a given system $q \in Q$, determine the size of the largest subset of set $W(q)$ all of whose elements satisfice the constraints $h(q,w) \le H$ and $g(q,w) \le G$, where h and g are real valued functions on $Q \times W(q)$.

Let $size(V)$ denote the size of set $V \subseteq W(q)$. Then the above problem can be stated as follows:

$$z^* := \max_{V \subseteq W(q)} \{Size(V) : h(q,v) \le H, g(q,v) \le G, \forall v \in V\} \tag{6.115}$$

This is a typical Maximin model where the worst-case analysis is driven only by the constraints. The Classic format of this *Maximin model* is as follows:

$$z^* := \max_{V \subseteq W(q)} \min_{v \in V} f(V,v) \tag{6.116}$$

where

$$f(V,v) := \begin{cases} size(V) & , \ h(q,v) \leq H, g(q,v) \leq G \\ -size(V) & , \ \text{otherwise} \end{cases} \tag{6.117}$$

If $W(q)$ is a *discrete* set, then we can let $size(V) = |V|$, in which case (6.115) can be rewritten as follows:

$$z^* := \max_{V \subseteq W(q)} \{|V| : h(q,v) \leq H, g(q,v) \leq G, \forall v \in V\} \tag{6.118}$$

where $|V|$ denotes the cardinality of set V.

The equivalent classic format of this Maximin model is as follows:

$$z^* := \max_{V \subseteq W} \min_{v \in V} f(V,v) \tag{6.119}$$

where

$$f(V,v) := \begin{cases} |V| & , \ h(q,v) \leq H, g(q,v) \leq G \\ -|V| & , \ \text{otherwise} \end{cases} \tag{6.120}$$

The point I want to make then is that, the fact that the above problem can be formulated as a *Maximin problem* does not imply that it cannot be formulated otherwise. To be precise, when I say that this problem is a *Maximin problem* I am not proposing that it can be formulated only in terms of a *Maximin model.* There may well be other models by means of which the same problem can be formulated.

For example, the discrete version of the above problem can be stated as follows:

How many elements of $W(q)$ satisfice the constraints $h(q,w) \leq H$ and $g(q,w) \leq G$?

To see that this problem can be formulated by models other than Maximin, let $W(q) = \{w_1, w_2, \dots, w_n\}$, and consider the following extremely simple model:

$$z^\circ := \sum_{j=1}^{n} I(j) \tag{6.121}$$

where

$$I(j) := \begin{cases} 1 & , \ h(q,w_j) \leq H, g(q,w_j) \leq G \\ 0 & , \ \text{otherwise} \end{cases} \ , \ j = 1,2,\dots,n \tag{6.122}$$

Convince yourself that $z^\circ = z^*$.

It is possible to formulated the same problem by other, far more intricate models. For example, consider the following model where M is a sufficiently large positive number:

$$z' := \max_{y_1,\dots,y_n} \sum_{j=1}^{n} y_j \tag{6.123}$$

$$s.t. \quad h(q,w_j) \leq H + M(1-y_j)\ ,\ j = 1,2,\dots,n \tag{6.124}$$

$$h(q,w_j) \leq G + M(1-y_j)\ ,\ j = 1,2,\dots,n \tag{6.125}$$

$$y_j \in \{0,1\}\ ,\ j = 1,2,\dots,n \tag{6.126}$$

Note that if $y_j = 0$, then the constraint $h(q,w_j) \leq H + M(1-y_j)$ reduces to $h(q,w_j) \leq H + M$, and since M is sufficiently large, this constraint is satisfied by w_j. If, on the other hand, $y_j = 1$, the constraint $h(q,w_j) \leq H + M(1-y_j)$ reduces to the original constraint $h(q,w_j) \leq H$. This means that the model allows y_j to be equal to 1 only if the original constraints $h(q,w) \leq H$ and $g(q,w) \leq G$ are satisfied for $w = w_j$. Since the model seeks to maximize $\sum_j^n y_{j=1}$, the optimal solution will set $y_j = 1$ for every j such that the original constraints are satisfied by w_j.

Convince yourself that $z' = z^* = z^\circ$.

Example

Next, let us consider a case where a problem can be formulated in terms of a *Minimin model* as well as a *Maximin model.* Consider then the following simple *Radius of Stability* model. You will recall that this is in fact *info-gap's generic robustness model*:

$$\alpha(q,\tilde{u}) := \max \left\{ \alpha \geq 0 : r^* \geq \min_{u \in U(\alpha,\tilde{u})} r(q,u) \right\}\ ,\ q \in Q \tag{6.127}$$

where $U(\alpha,\tilde{u})$ denotes a ball of radius α around $\tilde{u}$.

As we saw on a number of occasions in this discussion, this is a constrained *Maximin model* where the worst-case analysis is driven only by the constraint $r^* \leq r(q,u)$.

We also saw in this chapter that the following is a *Minimin model:*

$$\beta(q,\tilde{u}) := \min \left\{ \alpha \geq 0 : r^* \leq \max_{u \in U(\alpha,\tilde{u})} r(q,u) \right\}\ ,\ q \in Q \tag{6.128}$$

The kinship between these two models is manifested in the fact that they yield the same results. This is born out by the fact that subject to minor regularity conditions[11] $\beta(q,\tilde{u}) = \alpha(q,\tilde{u})$, and by the fact that the optimal values of u are the same in both cases.

Figure 6.10 illustrates the two models, where the rectangle $\mathcal{U}(q)$ represents the set of all the possible/plausible value of u pertaining to system q and the black curve separating the white region and the shaded region represents the constraint $r^* = r(q,u)$.

To form a clear idea of the relationship between the *Minimin* and *Maximin models,* think of the *Maximin model* as representing a process where the value of α (the circles' radius) is gradually being increased, until the constraint $r^* \geq r(q,u)$ is violated for some $u \in U(\alpha,\tilde{u})$. In contrast, in the case

[11]The optimal values of u in both cases yield $r^* = r(q,u)$.

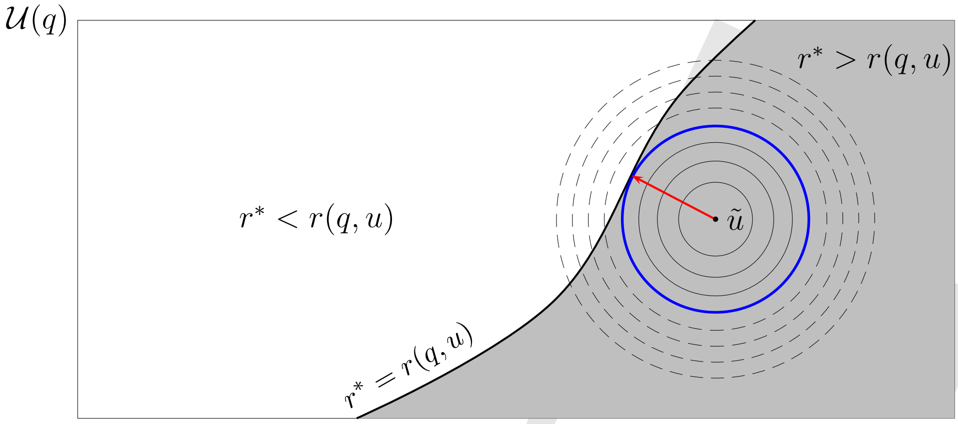


Figure 6.10: Maximin and Minimin representation of the same problem

of the *Minimin model,* think of it as representing a process where the value of α (the circles' radius) is gradually being decreased, until the constraint $r^* \leq r(q,u)$ is violated for all $u \in U(\alpha, \tilde{u})$.

So in the picture, the solid, bold, blue circle represents two things: the smallest dashed gray circle containing at least one u such that $r^* \leq r(q,u)$, and the largest solid gray circle all of whose elements satisfice the constraint $r^* \geq r(q,u)$.

My point is then that the same problem can be formulated both as a *Minimin model* and as a *Maximin model.*

The main reason that I am raising this issue is to point out that it is important to appreciate the distinction between the problem that a given model represents, and the model itself, because in certain cases this fact may have implications for the very solution of the problem concerned. Indeed, it may even decide how one would approach its solution.

And the relevance of all this for our discussion is that: while a given robustness problem expressed in terms of one model may prove intractable, when expressed in terms of a different model, the same problem may well submit to solution.

6.16.2 Discussion

Having said all that, a puzzling question seems to arise. The question is this: considering that a *Maximin model* represents a "pessimistic" approach to uncertainty/variability and a *Minimin model* represents an "optimistic" approach to uncertainty/variability, does this imply that one and the same problem can give expression to both an optimistic and a pessimistic view of the world? Surely, at least at first glance, the above example seems to be *paradoxical.*

Or, one might also ask: granting that one and the same problem can be given both a *Maximin* and a *Minimin* formulation, which of the two models captures more faithfully or more accurately or what have you, the problem in question? Surely one would assume that one formulation does this more "correctly", or properly, than the other?

This is precisely where the distinction between the problem and the model enables to illuminate why it is possible to give a problem both a *Maximin* and a *Minimin* formulation.

What needs to be appreciated in this regard is that **all** that the two models do is:

> To instruct how to compute the values of $\alpha(q, \tilde{u})$ and $\beta(q, \tilde{u})$, respectively, for a given $q \in Q$ and a given value of $\tilde{u}$.

Therefore, to decide whether the **problem itself** represents a "pessimistic" view of the world or an "optimistic" view of the world, we need to ask ourselves: what is the DM's preference regarding the distance of $\tilde{u}$ to the boundary specified by $r^* = r(q,u)$? Does the DM want $\tilde{u}$ to be distant from or close to the boundary?

As you will recall, the *Minimin model*'s "optimism" gives expression to the preference criterion "the smaller α is the better", whereas the *Maximin model*'s "pessimism" gives expression to the preference criterion "the larger α is the better".

Hence, if we assume that the DM prefers that $\tilde{u}$ be distant from the boundary, then the problem represented by these two models takes a "pessimistic" view of uncertainty/variability. Because, it is based on the notion that the distance of $\tilde{u}$ from the boundary is equal to the distance of $\tilde{u}$ to the NEAREST point on the boundary.

That the problem can be formulated by a *Minimin model* does not alter this fact.

I go into this issue in greater detail in the next chapter where I draw a distinction between *robustness models* and *decision models*. It is the latter that provide the framework for specifying the decision maker's preferences.

6.17 What is an "instance"?

To round out my discussion on the *Maximin model* I want to set the record straight on a fallacy that is circulating in the *info-gap literature* regarding the *Maximin paradigm* and its relation to *info-gap's robustness model* — the so-called *Robust-satisficing model.*

The reason that this is important is that, as the record shows, the errors, misconceptions, and misleading statements on this matter, that in the *info-gap literature* have lead many a scholar astray.

So, what I want to clarify here is a point that seems to be lost (wittingly, or unwittingly?) on *info-gap scholars*. This is the question of: what is an "instance", or more generally, where does an "instance" fit in the hierarchy of mathematical objects, concepts and so on.

My objective is to point out that in *info-gap circles* every effort is made to beat about the bush on the question of the kinship between the *Maximin model* and *info-gap's robustness model* so as to avoid accepting that *info-gap's robustness model* is in fact an instance of *Wald's Maximin model.*

For our purposes, we can take the term *instance* to signify a *specific case*, or a *special case*. For example, the polynomial

$$p(x) = x^2 - 3x + 1 \tag{6.129}$$

is an instance of the polynomial

$$q(x) = A + Bx + Cx^2 + Dx^3 \tag{6.130}$$

and that

$$\text{Model A:}\ \ \max\left\{\alpha \geq 0 : r^* \leq r(q,u), \forall u \in B(\alpha,\tilde{u})\right\},\ q \in Q \tag{6.131}$$

is an instance of

$$\text{Model B:}\ \ \max_{x\in X}\left\{g(x) : r^* \leq r(q,u), \forall u \in B(\rho(x),\tilde{v}(x))\right\},\ q \in Q \tag{6.132}$$

An *instance* is obtained from a more general "object", call it a "prototype", by specifying certain features (e.g. parameters) of the "prototype". Thus, for example, to show that $p(x) = x^2 - 3x + 1$ is an instance of $q(x) = A + Bx + Cx^2 + Dx^3$ we have to show that there are values for A, B, C and D for which $q(x) = A + Bx + Cx^2 + Dx^3$ is identical/equivalent to $q(x) = A + Bx + Cx^2 + Dx^3$.

This is straightforward as by inspection,

$$A + Bx + Cx^2 + Dx^3 \,|_{A=1,B=-3,C=1,D=0} \equiv 1 - 3x + x^2 \equiv p(x) \tag{6.133}$$

That is, $p(x) = x^2 - 3x + 1$ is the instance of $q(x) = A + Bx + Cx^2 + Dx^3$ specified by $A = 1, B = -3, C = 1, D = 0$.

As for the relation between Model A and model B, observe that Model A is the instance of Model B specified by $X = [0, \infty)$, $g(x) \equiv x$, $\rho(x) \equiv x$ and $\tilde{v}(x) \equiv \tilde{u}$.

One reason that it is important to be aware that one model is an instance of another model is that this hierarchy implies that the instance inherits all the essential properties of the prototype. This means that the body of knowledge available about the prototype is immediately relevant to its instances.

In this book I also use the term *simple instance* to indicate that the kinship between the prototype and the instance is simple in the sense that it is clearly identifiable, easily understood and readily verifiable.

So, in the preceding sections of this chapter I indicated, for instance (no pun intended!), that the *Radius of Stability* model and *info-gap's robustness model* are simple instances of *Wald's Maximin model.* I also established earlier that *info-gap's robustness model* is a simple instance of the Radius of Stability model.

Some readers may no doubt wonder:

> Why it is so important to be apprised of the fact that *info-gap's robustness model* is a simple instance of the *Radius of Stability* model and of *Wald's Maximin model.*

The principal reasons that this is important are these:

- It is important that readers of the *info-gap literature* be made aware of this fact because of its systematic misrepresentation in *info-gap publications*. Right from its introduction, *info-gap decision theory* has been hailed as a distinct, novel, revolutionary theory that is radically different from all current theories for decision under uncertainty.
  So, it is important that readers of this literature take note that the mathematical facts prove this position to be grossly in error. The simple mathematical truth is that — as I state above — *info-gap's robustness model* is a simple instance of the Radius of Stability model and both are simple instances of *Wald's Maximin model.*
- The reason that it is important to be aware of this fact is not only academic. It is important that readers of the *info-gap literature* realize that *Wald's Maximin model* and the *Radius of Stability* model have been the subjects of research and applications for many years. There is therefore a huge body of knowledge about these models and a great deal of experience with their application.
  In a word, it is important that readers of the *info-gap literature* be in a position to relate *info-gap decision theory* to the state of the art in the area of decision under (severe) uncertainty.

More on this in the second part of the book.

G'day Moshe,

Don't forget to point out the difference between: something being an *instance of* something, and something being *related to* something. This is important for the point you are discussing.

Cheers,
Fred
Internationally Renowned Expert on Robust Decision Under Severe Uncertainty

Thanks, Fred!

Fred's reminder of the importance to discuss the difference between an *instance of* and *related to* is due to the constant attempts in the *info-gap literature* to play down the kinship between *info-gap's robustness model* and *Wald's maximin model* from being an *instance of* to being *related to*.

So it is important to remind readers of this literature that the difference between something being an *instance of*, to its being *related to* something, is fundamental not only in mathematics or science more generally, but in ordinary parlance as well.

That is, the statement "Model C is an instance of Model D" is not only more informative than the statement "Model C is related Model D". It is more to the point, more accurate, more rigorous and so on. Observe that the first implies the second, but not vice versa.

By analogy, the purport of the two statements

· Fred is Sue's youngest son

· Fred is related to Sue

is **not** the same. The first is more informative, more to the point, more accurate etc. than the second.

The same is true for the following pair:

· *Info-gap's robustness model* is a simple instance of Wald's Maximin model and the *Radius of Stability* model.

· *Info-gap's robustness model* is related to *Wald's Maximin model* and the *Radius of Stability* model.

the first is far more informative, more to the point, far more accurate, more rigorous and the list goes on . . .

It is important that readers of the *info-gap literature* take full note of this point to avoid falling into the errors manifested, for instance, in the following publication (emphasis added):

> While there is a **superficial similarity with minimax decision making,** no fixed bounds are imposed on the set of possibilities, leading to a comprehensive search of the set of possibilities and construction of functions that describe the results of that search.
>
> Hine and Hall (2010, p. 16)

Do Hine and Hall (2010) imply that *Maximin models* necessarily impose fixed bounds on the "set of possibilities"? Are we then to conclude that

$$\max_{-\infty<x<\infty} \min_{-\infty<y<\infty} \{y^2 + 2xy - x^2\} \tag{6.134}$$

is not a *Maximin model*?

More on this in the second part of the book.

6.18 Bibliographic notes

TBW

Chapter 7

Robustness Against Severe Uncertainty

7.1 Introduction

In this chapter I take up the topic that is my main concern in this book: *Robustness Against Severe Uncertainty*. To this end, I pull together those elements that bear on a clarification of this issue, which I have discussed already in the preceding chapters. These are: *severe uncertainty* discussed in Chapter 4, *worst-case analysis* discussed in Chapter 5, and *Wald's Maximin model* discussed in Chapter 7.

The principal contribution of this chapter is not so much in the new material that it brings to light, as it is in its fitting together all these elements into a systematic explanation of the idea of *Robustness to Severe Uncertainty*.

For the benefit of readers who landed directly in this chapter, I briefly revisit these main elements beginning with the thinking that came out of *classical decision theory* in the early 1950s. I then proceed to discuss the ideas that developed in the field of *robust optimization* which in the early 1970s, emerged from *classical decision theory*, as a distinct branch of optimization theory specializing in robust decision in the face of severe uncertainty.

The reader is reminded that as our focus here is on *severe uncertainty,* namely an uncertainty that is quantified by likelihood-free models, I shall not examine methods such as *scenario generation* and *stochastic optimization* that are based on *probabilistic* models of uncertainty.

7.2 Classical decision theory

As indicated in Chapter 4, classical decision theory distinguishes between three states of affairs:

- *Certainty:* no uncertainty.
- *Risk:* probabilistic structures (e.g. probability distribution function) are used to quantify uncertain contingencies.
- *Uncertainty:* non-probabilistic, likelihood-free structures are used to quantify uncertain contingencies.

Obviously, *severe uncertainty* falls in the third category, which means of course that it is quantified by non-probabilistic likelihood-free models.

To model decision-making under severe uncertainty, *classical decision theory* proposes the use of a *payoff table* where the rows represent the decisions available to the decision-maker (DM) and the columns represent the "states of the world", namely the events that are governed by *Uncertainty* or, *Nature*. The entries in the payoff table represent the awards allotted to the DM, which reflect the decision selected by the DM and the state selected by *Nature*.

For example, consider the payoff table shown in Figure 7.1. The DM has 3 options (decisions) to choose from, and Nature has 5 states to choose from.

		Nature				
		s_1	s_2	s_3	s_4	s_5
DM	a_1	3	2	5	6	2
	a_2	9	8	0	6	7
	a_3	0	5	4	3	0

Table 7.1: Payoff table

If the DM selects say, decision a_1, the severity of the uncertainty is manifested in her total lack of knowledge as to which of the five payoffs associated with this decision, namely 9, 8, 0, 6, or 7, will be realized.

Before I commence the discussion on how *classical decision theory* addresses severe uncertainty of this type, let us examine, very quickly, the other two cases.

7.2.1 Certainty

In this case the payoff table consists of only *one column*. Namely, there is only one possible state of nature to contend with. Therefore, if the DM wants to maximize her payoff, she will select the decision whose payoff is the largest. And, if she wants to minimize her payoff, she will select the decision whose payoff is the smallest.

The ruling convention in this book is that — unless expressly stated to the contrary — the assumption is that the DM seeks to *maximize* her payoff.

Conceptually then, the situation is quite simple, so that the decision rule for the DM is straightforward:

Decision Rule:
Select the decision whose payoff is the largest.

But, I need hardly point out that this does not imply that this instruction can be easily implemented *in practice*. Determining which decision yields the largest payoff can be an extremely difficult task. I shall elaborate on this matter in my discussion on the *robust optimization* paradigm. For now it suffices to say that the implementation of this rule may require the solution of an extremely complex/difficult *optimization problem.*

G'day Moshe,
For the benefit of *info-gap users/scholars* it might be a good idea to point out that the above decision rule can also handle "satisficing" problems — as supposedly distinguishable from "optimizing" problem.

Cheers,
Fred
Internationally renowned Expert on Robust Decision Under Severe Uncertainty

It is indeed important to point this out not only for the benefit of *info-gap scholars and users,* but also for all those who make a big fuss about the supposedly fundamental difference between "optimizing" and "satisficing".

In the *info-gap literature* the fuss is about the purportedly unique treatment given by the *info-gap approach* to situations where decisions are made under severe uncertainty. The *info-gap rhetoric* argues that the great merit of *info-gap's robustness model*, the so-called *info-gap robust-satisficing model* is that, contrary to what it calls "direct optimization models", which according to this rhetoric, aim only to optimize utility, *info-gap's robustness model* seeks decisions that are *robust-satisficing*.

I shall take up this issue in the second part of the book where I outline my criticism of *info-gap decision theory*. For now, it suffices to say that this argument, which is based on an alleged superiority of "satisficing" over "optimizing", is a straw man argument.

The fact is that there is no fundamental difference between "optimizing" and "satisficing". Because, any "satisficing" problem can be reformulated as an equivalent "optimizing" problem so that any optimal solution to the optimizing problem is a "satisfactory" solution to the "satisficing" problem. Conversely, any "satisfactory" solution to the "satisficing" problem is an optimal solution to the "optimizing" problem. This fact is illustrated in the next subsection.

So, the moral of this story is that the above decision rule is applicable to so-called "optimizing problems" and to so-called "satisficing problems'.

7.2.2 Risk

In situations classified as "decision under risk", the uncertainty in the state variable s is quantified by a probability mass function that specifies the probability of each possible state. Let $p(j)$ denote the probability that Nature will select state s_j.

Classical decision theory proposes that in such cases decisions should be ranked on the basis of the *expected value* of their respective payoffs. Hence,

Decision Rule:
Select a decision that maximizes the expected value of the payoff.

This is illustrated in Table 7.2 where the assumed probabilities are specified in the bottom row. The entries in the E column are the expected values of the payoffs[1].

		Nature					
		s_1	s_2	s_3	s_4	s_5	E
DM	a_1	3	2	5	6	2	4.8
	a_2	9	8	0	6	7	5.6
	a_3	0	5	4	3	0	3.4
	$p(s_j)$	0.1	0.2	0.3	0.4	0.1	

Table 7.2: Payoff table

Thus, in this case the best decision is a_2. Its expected payoff is equal to 5.6.

A note for info-gap scholars and all those who go on about the shortcomings of the *expected value* paradigm.

The fact that the decision rule prescribed by the *expected value* paradigm calls for the maximization of the *expected payoff* in no way implies that this rule is not applicable to situations where the objective is

[1]To compute the expected value of a payoff, multiply the probabilities by their respective payoffs, and add up the products. Thus, the expected payoff for a_2 is equal to $(0.1 \cdot 9 + 0.2 \cdot 8 + 0.3 \cdot 0 + 0.4 \cdot 6 + 0.1 \cdot 7) = 5.6$.

to maximize the *probability of a satisfactory outcome*. This only means that, if one's objective is to meet this requirement, then one ought to formulate the problem so as to give full expression to this objective in the formulation.

And to illustrate. Suppose that in the context of the payoff table given in Table 7.2, we seek to satisfice the requirement that "the payoff should be strictly positive". And more than that, suppose that we want to rank the decisions on this basis. Indeed, we want to select the decision with the highest probability that the payoff will be strictly positive. By inspection then, the optimal decision in this case is a_2, as the result that the payoff is strictly positive has a probability of 0.8.

The point is then that the simple ''maximization of expected payoff" model proposed by *classical decision theory* is fully capable of handling this requirement. This is so because *classical decision theory* **does not** dictated to the analyst how to define "payoff". This is the analyst's business. Meaning that the onus is on the analyst to make sure that the model representing the problem in question takes full account of this consideration.

In our particular example, this modeling task amounts to no more than a modification of the original payoffs so as to allow them to properly represent the objective in question. This is shown in Table 7.3, where the payoffs are *boolean variables* representing the requirement "the payoff is strictly positive". That is, the modified payoff is equal to either 1 or 0, depending on whether the original payoff is strictly positive.

		Nature					
		s_1	s_2	s_3	s_4	s_5	E
DM	a_1	1	1	1	1	1	1
	a_2	1	1	0	1	1	0.8
	a_3	0	1	1	1	0	0.8
	$p(s_j)$	0.1	0.2	0.3	0.4	0.1	

Table 7.3: Modified Payoff table

The E column in this table represents the expected values of the modified payoffs, which are equal to the respective probabilities that the original payoffs are strictly positive. Note that here the optimal decision is a_1. It yields a strictly positive (modified) payoff with probability 1.

This of course is not an accident.

It is a simple illustration of the fact that the probability of an event can be expressed as the expected value of the event's *indicator function*. For example, let $\tilde{x}$ be a discrete[2] random variable, let x denotes its realization, let $\widehat{X}$ be the range of values that it can take, and let X be a subset of $\widehat{X}$. Then,

$$Prob(x \in X) = E(I_X(x)) \tag{7.1}$$

where $Prob(x \in X)$ denotes the probability that x takes a value in X and $E(I_X(x))$ denotes the expected value of $I_X(x)$, where

$$I_X(x) := \begin{cases} 1 & , \ x \in X \\ 0 & , \ x \notin X \end{cases} , \ x \in \widehat{X} \tag{7.2}$$

[2]This is just a technicality.

More on this in the second part of the book.

So what is the moral of this short story?

The moral of the story is that the *info-gap literature* and all those who carry on about the deficiencies of the ‘expected utility maximization” paradigm vis-a-vis the "satisficing paradigm" give an utterly distorted picture of *Classical decision theory's* well-established “expected utility maximization” paradigm.

The point is that as this paradigm **does not dictate** to us what “utility” is or ought to be, and as “expectation” can represent “probability of events”, furthermore, as “maximization” can represent “satisfaction” the implication is clear. The paradigm is fully capable (insofar as modeling it goes) to accommodate a variety of objectives and requirements one seeks to meet or satisfice, including “satisficing” probabilistic requirements.

I return to this matter at a later stage in the chapter.

7.2.3 Severe Uncertainty

Classical decision theory proposes two basic approaches to deal with situations where the severity of the uncertainty hinders the quantification of the uncertainty by means of a probabilistic structure.

One approach proposes to attribute a likelihood structure to the “state of Nature”. This proposition in effect converts the given situation from subject to *severe uncertainty* to one falling under *risk*. The other is even more extreme: it postulates a rule which effectively transforms the given “severe uncertainty” into “certainty”.

So, let us examine these two approaches with the view to clarify how they deal with the following two important issues:

· Quantification of severe uncertainty.

· Quantification of robustness.

It will be convenient to conduct this discussion in the framework of the “classical” *payoff table*.

7.2.4 Laplace's Principle of Insufficient Reason

This principle, named after the famous French mathematician and astronomer Pierre-Simon, marquis de Laplace (1749–1827), is also known as the *Principle of Indifference.*

Roughly, the principle argues that if you face $n > 1$ events that are mutually exclusive and collectively exhaustive, then it makes sense to assume that these events are *equally likely,* hence that each occurs with probability $1/n$. The most famous applications of this principle are in the exciting area of *gambling*, where all sort of games of luck are played with coins, dice, cards, and wheels.

To illustrate, in the case of the payoff table given in Table 7.1, the principle argues that each of the five states occurs with probability $1/5 = 0.2$.

Once this choice of a probabilistic structure is made, the decision-making situation is treated as “decision-making under risk”. That is, the decisions are ranked on the basis of the *expected payoff* that they generate.

As a matter of fact, computing the expected values is not really necessary because the sum of the payoffs will yield the same ranking. The expected values are equal to the sums divided by 5 (see Table 7.4).

According to this, the best decision is a_2. Its expected payoff is equal to 6.

		Nature						
		s_1	s_2	s_3	s_4	s_5	SUM	E
DM	a_1	3	2	5	6	2	18	3.6
	a_2	9	8	0	6	7	30	6.0
	a_3	0	5	4	3	0	12	2.4
	$p(s_j)$	0.2	0.2	0.2	0.2	0.2		

Table 7.4: Expected Payoffs, $E = SUM/5$

An obvious limitation of the principle is that it cannot be applied in situations where the range of feasible values of the state variable does not submit to a *uniform* probability distribution. For example, if the set of possible values of s is the non-negative section of the real line $\mathbb{R}_+ = [0, \infty)$, then it is impossible to formulate a probabilistic structure such that all the feasible values of the state are equally likely.

One must also be careful with regard to cases where the state is *multivariate* and its components are not "independent". Because, if the assumption in such cases is that a given component is uniformly distributed, the distribution functions of other components may not be uniform. The question is then, how to determine which component should be uniformly distributed. The point is that the answer is not always straightforward.

In fairness to *Laplace's Principle,* it should be noted that this difficulty afflicts distribution functions in general.

Needless to say, from a Bayesian point of view, the employment of a *uniform distribution* to quantify severe uncertainty is rather extreme (see the discussion in Chapter 4).

7.2.5 Wald's Maximin model

As I indicate in Chapter 6, the recipe prescribed by *Wald's Maximin model* to handle *severe uncertainty* is based on a *worst-case analysis.* That is, the idea is that Nature is a *hostile adversary,* her hostility being manifested in her selecting the worst state pertaining to the decision selected by the DM.

I also pointed out that this paradigm of *severe uncertainty* effectively renders uncertainty *predictable*. Because, for all her hostility towards the DM's selections, given that Nature is consistent in this hostility, the implication is that the DM no longer deals with a situation of uncertainty.

Thus, *Wald's Maximin model* trades "uncertain outcomes" for "certain bleak outcomes". Or in other words, it trades the convenience of a certain, but "grim world", for the inconvenience of a severely uncertain but "mixed world". Needless to say, the hope is that the worst-case will not be realized: *hope for the best but plan for the worst!*

This is illustrated in Table 7.5, where the "certain" worst-case states are shown for each alternative.

Insofar as "robustness" is concerned, as in classical *game theory*, *classical decision theory* employs the notion SECURITY LEVEL to measure the performance of an alternative to severe uncertainty. Formally, the *security level of an alternative* is the payoff yielded by the worst-case state(s) pertaining to this alternative. Thus, for an $m \times n$ payoff table,

$$SL(i) := \min_{1 \leq j \leq n} \textit{payoff}(i, j) \ , \ i = 1, 2, \ldots, m \tag{7.3}$$

The decision rule is therefore as follows:

		Nature				
		s_1	s_2	s_3	s_4	s_5	S^*
DM	a_1	3	$\boxed{2}$	5	6	$\boxed{2}$	$\{s_2, s_5\}$
	a_2	9	8	$\boxed{1}$	6	7	$\{s_3\}$
	a_3	$\boxed{0}$	5	4	3	$\boxed{0}$	$\{s_1, s_5\}$

Table 7.5: Expected Payoffs, S^* = set of worst-case states

Decision rule:
Rank alternatives according to their security levels. Hence, select that alternative whose security level is the highest.

Equivalently:
Rank alternatives according to their worst-case payoff. Hence, select the alternative whose worst-case payoff is the largest.

Symbolically,

$$z^* := \max_{1\leq i\leq m} SL(i) \tag{7.4}$$

$$= \max_{1\leq i\leq m} \min_{1\leq j\leq n} payoff(i,j) \tag{7.5}$$

Table 7.7 illustrates this recipe in action.

		Nature						
		s_1	s_2	s_3	s_4	s_5	SL	S^*
DM	a_1	3	$\boxed{2}$	5	6	$\boxed{2}$	$\boxed{2}$	$\{s_2, s_5\}$
	a_2	9	8	$\boxed{1}$	6	7	1	$\{s_3\}$
	a_3	$\boxed{0}$	5	4	3	$\boxed{0}$	0	$\{s_1, s_5\}$

Table 7.6: Wald's Maximin analysis

The highest security level — equal to 2 — is associated with a_1. Hence, according to the *Maximin rule,* a_1 is the best (most robust) alternative available to the DM. If the DM selects this alternative, then regardless of what state will be realized, her payoff will not be less than 2. There are two worst-case states for this alternatives, namely $S^*(1) = \{s_2, s_5\}$.

As indicated in Chapter 6, this model's main drawback is that it can yield extremely "conservative" results. For example, consider the payoff table shown in Table 7.8.

According to the *Maximin rule,* alternative a_2 is the least attractive, even though four out of its five possible payoffs are much higher than those associated with the other two alternatives. In a word, one bad apple tells against the performance of an otherwise highly attractive alternative.

Another feature that, to the minds of some experts, renders this paradigm "problematic" is that:

> An addition of a **constant** to the payoffs associated with a given state (column of the payoff table) may change the ranking of the alternatives.

This is illustrated in Table 7.8 which is obtained by adding 3 to the payoffs associated with state s_3.

		Nature						
		s_1	s_2	s_3	s_4	s_5	SL	S^*
DM	a_1	3	$\boxed{2}$	5	6	$\boxed{2}$	2	$\{s_2, s_5\}$
	a_2	900	800	$\boxed{1}$	600	700	1	$\{s_3\}$
	a_3	$\boxed{3}$	5	4	6	$\boxed{3}$	$\boxed{\boxed{3}}$	$\{s_1, s_5\}$

Table 7.7: Wald's Maximin analysis

		Nature						
		s_1	s_2	s_3	s_4	s_5	SL	S^*
DM	a_1	3	$\boxed{2}$	8	6	$\boxed{2}$	2	$\{s_2, s_5\}$
	a_2	900	800	$\boxed{4}$	600	700	$\boxed{\boxed{4}}$	$\{s_3\}$
	a_3	$\boxed{3}$	5	7	6	$\boxed{3}$	3	$\{s_1, s_5\}$

Table 7.8: Wald's Maximin analysis

Note that in the case of the payoffs listed in Table 7.7 alternative a_2 is the least attractive, whereas according to those listed in Table 7.8 this alternative is the most attractive.

So, the fault that some experts find with this paradigm is that an addition of the same constant to all the payoffs associated with a given state (column) should not have the drastic effect of altering the alternatives' ranking.

It should be pointed out that this problem does not affect the decision rule based on *Laplace's Principle of Insufficient Reason.* This decision rule has, however, a fault of its own which is that its ranking of alternatives may change if a column of the payoff table is **duplicated**. This difficulty does not affect the rule stipulated by *Wald's Maximin paradigm.*

7.2.6 Variations on a theme

I refer the reader to Chapter 6 for a discussion on the variants of *Wald's Maximin model.* Their common denominator is that they end converting the *severe uncertainty* into *certainty* by considering the worst/best case payoffs associated with the states.

For the benefit of *info-gap scholars*, it is instructive to illustrate this point in the context of the *Minimin model* because, as we have seen already, *info-gap's opportuneness model* is in fact a simple instance of the classic *Minimin model.*

So, recall that in the framework of this model, Nature cooperates with the DM, meaning that She selects the best-case state pertaining to the alternative selected by the DM. The formal model is then as follows:

$$z^{\circ} := \min_{1 \leq i \leq m} \; \min_{1 \leq j \leq n} \; payoff(i, j) \tag{7.6}$$

where the outer $\min$ represents the DM and the inner $\min$ represents Nature.

Because Nature cooperates with the DM, the decision-making situation represented by this model can be viewed as a simple optimization problem, involving only the decision maker, who selects both the alternative (a_i) and the state (s_j) with the view to *minimize her payoff.* Hence, symbolically,

$$z^\circ := \min_{\substack{1\le i\le m\\ 1\le j\le n}} \textit{payoff}(i,j) \tag{7.7}$$

This ultra-optimistic paradigm forms part of *Hurwicz's model,* namely the *Optimism Pessimism Index,* where its function is to moderate the extreme pessimism of *Wald's Maximin model* (see section 6.12).

I need hardly point out, though, that *classical decision theory* does not propose that the ultra-optimistic *Minimin paradigm* be used on it own, as one would be hard pressed to justify the logic behind the employment of such a model in practice.

7.3 Robust optimization models

The great attraction of the formal models offered by *classical decision theory* for *decision under severe uncertainty* – discussed in the preceding sections – is their *extreme simplicity*. The decision problem as a whole is described in terms of a *payoff table*.

This characteristic also renders these models highly convenient frameworks for expounding the *fundamental issues* underlying decision-making obviously , as perceived by *classical decision theory*. So, it is not surprising that these models are discussed in many introductory texts on *decision theory* and related fields such as *operations research* and *management science*.

In the late 1960s, early 1970s, the basic ideas expressed in these models were incorporated into *optimization models* whose objective is the recovery of solutions (to optimizations problems) that are not only *optimal* but also *robust* to severe uncertainty in the true value of the models' parameters.

Thus emerged the field that is known today as *robust optimization*.

The main difference between the models of *classical decision theory* and those of *robust optimization theory* is in the treatment of *constraints*. That is, in *classical decision theory* the focus is on the *payoffs,* so that the models are formulated as *payoff tables.* In fact, they make no explicit reference to constraints.

In contrast, the models deployed by *robust optimization theory* are *constrained optimization models*, where *constraints* are as "important" as the *payoffs.* As for the treatment of payoffs, in line with the conventions of optimization theory, instead of being specified by a *table*, they are specified by an *objective function*.

Finally, as might have been expected, because in *robust optimization*, the element of "robustness" is central to the decision-making process, *Wald's Maximin paradigm* is a central element in *robust optimization* models.

Now, in this chapter I follow the tradition of explaining *robust optimization models* as natural counterparts of *parametric optimization models*. So my plan for this chapter is to review the essential features of *optimization models*, then discuss the related *parametric optimization models* that they induce, and finally examine the so called *robust counterparts* of these models.

I remind the reader that in this book my concern is only with non-probabilistic models of robustness. This means that these robustness models are designed to handle *variability* in general rather than only variability that is due to severe uncertainty.

In the last section of this chapter I discuss the issues that arise when these models are used in the pursuit of robustness to severe uncertainty.

7.3.1 Optimization models

The optimization model that will accompany us throughout this chapter is rather simple. This is a deliberate choice, designed to avoid digressions into peripheral technical matters that are related to questions such as the existence of optimal solutions, availability of solution methods, and so on.

So without further ado, here is the model in question:

Model P:

$$z^* := \max_{x \in X} g(x) \text{ subject to } constraints(x) \tag{7.8}$$

where X is some set, g is a real valued function on X and $constraint(x)$ denotes a list of constraints on x. In this framework x represents the *decision variable* and X is the set of decisions available to the decision maker (DM). I shall refer to X as the *decision space* and to g as the *objective function*.

Let X^* denote the set of all the optimal solutions to the problem described by this model. The assumption is that X^* is not empty. It will be useful to let X' denote the set of *feasible* solutions to this problems. Hence, let X' denote the subset of X whose elements satisfice the constraints specified by $constraints(x)$.

An *instance* of this model is obtained by specifying the three objects that comprise *Model P*, namely the *decision space* X, the *objective function* g, and the list of *constraints* that are included in $constraint(x)$. Such an instance gives expression to a *concrete optimization problem.*

Example

Consider the following simple, concrete optimization problem:

$$z := \max_{x_1, x_2} \{x_1 + x_2\} \text{ subject to } x_1^2 + x_2^2 \leq 8 \tag{7.9}$$

To show formally that this problem is an instance of the generic problem described by *Model P*, we can set $x = (x_1, x_2)$, $X = \mathbb{R}^2$, $g(x) = x_1 + x_2$ and let the list $constraints(x)$ consists of the single constraint $x_1^2 + x_2^2 \leq 8$. Then clearly, $X' = \{x \in X : x_1^2 + x_2^2 \leq 8\}$. By inspection, the (unique) optimal solution is $x^* = (2, 2)$, hence, $X^* = \{(2, 2)\}$. This is shown graphically in Figure 7.1.

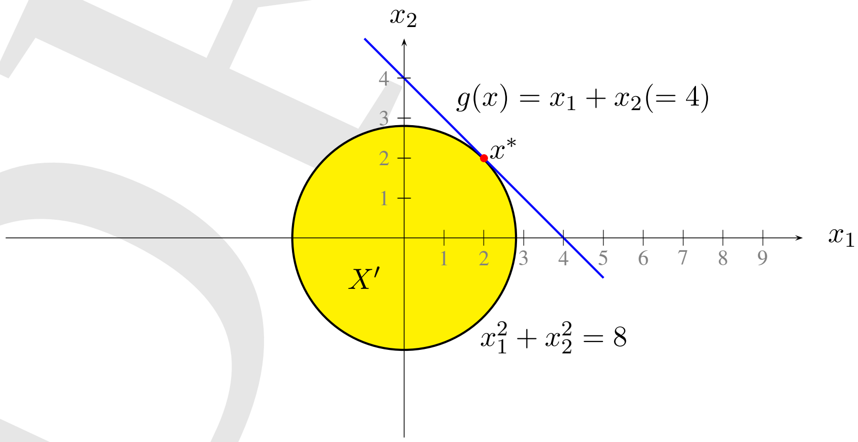


Figure 7.1: A simple optimization problem, $x^* = (2, 2)$

The straight blue line is the collection of all the points $x = (x_1, x_2)$ that satisfice the equation $g(x) = x_1 + x_2 = 4$. The circle represents the feasible domain of x. □

For the benefit of readers who are not conversant with optimization theory and its practical aspects, I should point out that the general formulation stated above does not imply that in parallel to this formulation, a general-purpose recipe is available for the solution of optimization problems of this form. Indeed, there is no such thing as a “general-purpose solution recipe for optimization problems. Indeed, often the availability of solution recipes may depend on the “size” of the problem.

Typically, the “size” of an optimization problem is measured by (a) the number of its decision variables and, (b) the number of its constraints. The first refers to the number of elements comprising the decision $x \in X$. In practice, an optimization problem can consist of thousands of decision variables and thousands of constraints.

This means that some optimization problems are very easy to solve, while others are extremely difficult to solve. Generally, the difficulties encountered in the problem’s solution are due to the volume of computations required to solve the problem, which may grow disproportionately with the problem’s size.

And last but not least, a few words about *optimization methods,* namely the methods used to solve optimization problems such as those subsumed by *Model P*.

Not unlike the experts who wield them, optimization methods are highly specialized. Specialized in the sense that they tend to be particular about the problems that “they admit for treatment". This means that one cannot take it for granted that just because a certain optimization problem can be solved by say, Method A, a “slightly” modified version of the problem would also be amenable to Method A.

This fact has immediate implications for our discussion, because the robust optimization models that are of concern to us here are “modified versions” — so called *robust counterparts* — of *Model P*. And if this were not enough, then generally, these modifications can be significant. The inference is then that a method capable of solving a given optimization problem may not be capable of solving the problem’s *robust counterpart*. Usually, the *robust counterpart* of an instance of *Model P* is expected to be more difficult — sometimes considerably more difficult — to solve than the instance in question.

And for the benefit of *info-gap scholars*, I want to point out that *Model P* enables formulating problems with **no objective functions.** I shall refer to such models as “satisficing models”. In such cases you would simply create an ad hoc objective function. So, if you want to ensure that the optimization model will yield all the solution to the “satisficing model”, let the objective function be equal to a *constant*, say let $g(x) = 0, \forall x \in X$. Obviously, in this case we would have $X^* = X'$. There could, of course, be more useful choices for the objective function.

In short, in this discussion I regard “satisficing models” as degenerate cases of “optimization models”. The important point to remember is that for any given “satisficing model” there is an equivalent “optimization model”, that is an optimization model whose optimal solutions are the feasible solutions for the “satisficing” model. I come back to this issue at a later stage in this chapter.

7.3.2 Parametric models

Suppose that both the objective function and the constraints in Model P call for values of a parameter u whose set of feasible/plausible values is $\mathscr{U}$. Then, Model P can be seen as comprising a *family* of optimization models, each affiliated with a value of $u \in \mathscr{U}$. Consider then this family of models:

Model $P(u), u \in \mathscr{U}$:

$$z^\circ(u) := \max_{x \in X} g(x; u) \text{ subject to } constraints(x; u) \tag{7.10}$$

where the u in $g(x; u)$ and in $constraints(x; u)$ simply indicates that the object in question "may depend" on the stipulated value of u.

Note that since $constraints(x; u)$ represents a *list* of constraints, for a decision to be feasible with respect to a given u, it must satisfice *all* the constraints in this list for the given value of u.

To distinguish between the *family* of these models and *instances* thereof, I shall use the following notation: the expression *Model P(u)* denotes the particular *instance* of the family that is specified by the value of u considered, whereas *Model P($\mathscr{U}$)* denotes the family (set) of all the instances comprised by $\mathscr{U}$.

For each $u \in \mathscr{U}$, let $X'(u)$ denote the set of feasible solutions to *Model P(u)* and let $X^*(u)$ denote the set of optimal solutions to this model (instance). The assumption is that for each $u \in \mathscr{U}$ the set $X^*(u)$ — hence also $X'(u)$ — are not empty.

And before I proceed to illustrate this model is action, I take a short break to explain the notation that I use in this book.

Notation break

The motivation for writing x under the $\max$ in the formulation of *Model P(u)* and for including u in $z^\circ(u)$ is to draw a clear distinction between the *decision variable* x and the *parameter* u and to highlight their respective roles in a robust optimization model. This notation brings out at a glance that the parameter of interest is denoted by u. What is more, it brings out that x is controlled by the decision maker ($\max$) whereas u is not. This convention allows employing whatever symbols one fancies for the decision variable and the parameter. For instance, in

$$c(v) := \max_{y \in \mathbb{R}^2} \{v_1 y_1 + v_2 y_2\} \text{ subject to } |y_1| + |y_2| \leq b \, , \; v \in V \tag{7.11}$$

y denotes the decision variable, v denotes the parameter of interest, V denotes the parameter space and b is some other parameter, regarded as "given".

To keep things simple, I shall do my best to stick to the (x, u, X, $\mathscr{U}$) notation throughout the remainder of the book.

Example

Consider the following parametric optimization problem:

$$z^\circ(u) := \max_{x \in \mathbb{R}^2} \{u_1 x_1 + x_2\} \text{ subject to } x_1^2 + x_2^2 \leq u_2^2 \, , \; u \in \mathscr{U} := [1, 10] \times [2, 4] \tag{7.12}$$

Note that the optimization problem featured in section 7.3.1 is the instance of this model that corresponds to $u = (1, \sqrt{8})$. Figure 7.2 displays the solution to the instance corresponding to $u = (2, 4)$, namely the problem:

$$z^\circ(2, 4) := \max_{x \in \mathbb{R}^2} \{2x_1 + x_2\} \text{ subject to } x_1^2 + x_2^2 \leq 16 \tag{7.13}$$

The unique optimal solution is $x^* = \big(8/\sqrt{5}, 4/\sqrt{5}\big)$.

It is straightforward to show that the optimal solution to the parametric problem specified by (7.12),

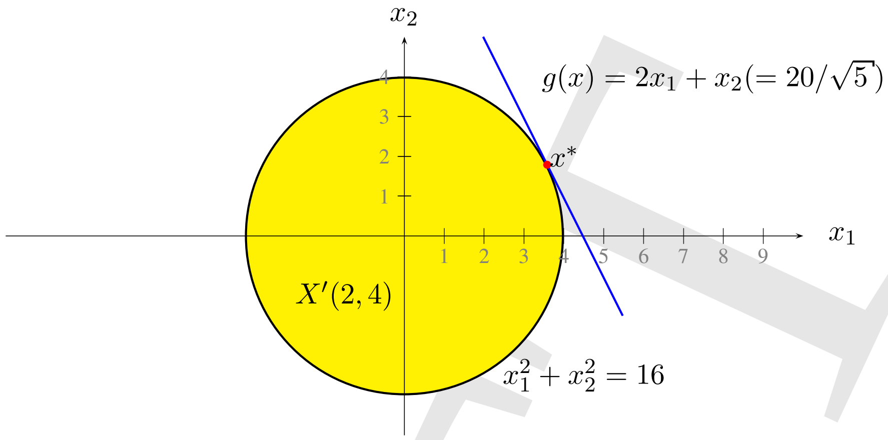


Figure 7.2: A simple optimization problem, $x^* = \left(8/\sqrt{5}, 4/\sqrt{5}\right)$

call it $x^*(u)$, can be expressed explicitly in terms of u as follows:

$$x^*(u) = \left(\frac{u_1 u_2}{\sqrt{u_1^2 + 1}} , \frac{u_2}{\sqrt{u_1^2 + 1}}\right) \ , \ u \in \mathscr{U} \tag{7.14}$$

Hence,

$$\begin{aligned} z^\circ(u) &= u_1 x_1^*(u) + x_2^* \ , \ u \in \mathscr{U} && (7.15)\\ &= u_1 \left(\frac{u_1 u_2}{\sqrt{u_1^2+1}}\right) + \frac{u_2}{\sqrt{u_1^2+1}} && (7.16)\\ &= u_2 \sqrt{u_1^2 + 1} && (7.17)\end{aligned}$$

□

Of course, eventually the parameter u will emerge as an object whose true value is subject to severe uncertainty, in which case $\mathscr{U}$ will represent the set of possible/plausible values of the (unknown) true value of u. For the time being, however, we do not ascribe uncertainty to u. We simply regard it as a *parameter* of the optimization model so that $\mathscr{U}$ is referred to as the *parameter space*.

It is important to take full note of this fact because robustness models need not necessarily be associated with uncertainty, severe or otherwise, They are also applicable to situations that have got nothing to do with uncertainty.

I should also point out that because the uncertainty model postulated by this discussion is non-probabilistic and likelihood-free, most of the robustness models that I shall examine here are suitable not only for cases where robustness is sought against severe uncertainty, but also for cases where robustness is sought against *prescribed variability*.

With this as background, I begin my discussion on how to define robustness in the context of *Model P(u)* by examining two candidate concepts that might be considered for this purpose. The idea is to show how these concepts drive the quest for a definition of robustness. This will also shed light on a point that I had made earlier, namely that there is no one formal definition of robustness that, as it were, fits all problems. Rather, the definition of robustness would depend on the objectives that one seeks to accomplish.

It stands to reason that one's objective would be to obtain the highest degree, or level, of robustness possible. Consider then the following proposition.

Definition 7.3.1 SUPER-ROBUST DECISIONS.
In the context of Model P($\mathscr{U}$), a decision $x^{\circ} \in X$ is said to be SUPER-ROBUST *if $x^{\circ} \in X^{*}(u), \forall u \in \mathscr{U}$. That is, a super-robust decision is a decision that is simultaneously optimal for all instances of Model P($\mathscr{U}$).*

That said, I hasten to add that super-robust decisions **hardly** exist.

Let me explain the exact meaning of this assertion.

There are of course practical situations where one decision *dominates* all other decisions, for all the values of the parameter under consideration, in which case one would be justified in saying that a super-robust decision exists. But the point about such practical situations is that they would be trivial and rare.

Likewise, it is no doubt possible, indeed very easy, to contrive examples where super-robust decisions exist. But here as well, the models featured in such contrived examples would be ... either trivial and/or unrepresentative of the difficulties encountered in situations calling for a robustness analysis.

After all, the whole point of resorting to the parametric model in the first place is that the optimal decision to *Model P(u)* is expected to depend on and vary with u. The preceding example vividly illustrates this point: for each $u \in \mathscr{U}$ there is a distinct optimal decision $x^{*}(u)$ specified by (7.14).

To put it in more general terms, it is precisely the difficulty that super-robust decisions are rare in the extreme, or in other words, that *Model P($\mathscr{U}$)* does not constitute a friendly paradigm for robust decision-making, that the field of *robust decision-making* is so challenging, and I might add so interesting.

So, to sum up, both from a practical point of view and methodologically, super-robust decisions can be deemed rare entities. Therefore, other than suggesting the obvious — namely, that if a super-robust decision exists, it should be treated as a rare commodity — I do not plan to pursue this matter any further.

I should add, though, that *Model P($\mathscr{U}$)* and the notion "super-robust decision" are instructive in that they bring out the difficulties associated with robust decision-making problems.

The second idea that would immediately spring to mind — considering of course the lengthy discussions in the preceding chapters on "worst-case analysis" — would be to ground robustness in a "worst instance" of *Model P($\mathscr{U}$)*. That is, to express robustness in terms of that instance of *Model P($\mathscr{U}$)* that yields the "worst" $u \in \mathscr{U}$.

So the question is ... what is the "worst" $u \in \mathscr{U}$ in the context of *Model P($\mathscr{U}$)*?

Given that the goal in *Model P(u)* is to *maximize* the objective function subject to the given constraints, it can be argued that it makes sense to rank the "undesirability" of $u \in \mathscr{U}$ in accordance with their $z^{\circ}(u)$ values: the smaller $z^{\circ}(u)$, the worse it is. Hence, the "worst" u in $\mathscr{U}$ is that which *minimizes* $z^{\circ}(u)$ over $\mathscr{U}$. This being so, the following criteria immediately suggest themselves (take a deep breath):

Definition 7.3.2 WORST-BEST INSTANCE.
In connection with Model P($\mathscr{U}$), let

$$\mathscr{U}^{(wb)} := \left\{ u' \in \mathscr{U} : z^{\circ}(u') = \min_{u \in \mathscr{U}} z^{\circ}(u) \right\} \tag{7.18}$$

We shall refer to elements of $\mathscr{U}^{(wb)}$ as the worst-best elements of $\mathscr{U}$. Accordingly, a worst-best instance of Model P($\mathscr{U}$) is an instance Model P(u) such that $u \in \mathscr{U}^{(wb)}$.

Definition 7.3.3 WORST-BEST PAYOFF.
We shall refer to

$$z^{(w)} := \min_{u\in\mathscr{U}} z^{\circ}(u) \tag{7.19}$$

$$= \min_{u\in\mathscr{U}} \max_{x\in X} \ g(x;u) \ \textit{subject to} \ \ constraints(x;u) \tag{7.20}$$

as the WORST-BEST PAYOFF *associated with Model P($\mathscr{U}$).*

Definition 7.3.4 WORST-BEST DECISION.
We shall refer to the elements of

$$X^{(wb)} := \bigcup_{u\in\mathscr{U}^{(wb)}} X^{*}(u) \tag{7.21}$$

as the WORST-BEST DECISIONS *associated with Model P($\mathscr{U}$).*

Dear Reader:

To preempt interjections from Fred[3], I want to point out that Definition 7.3.4 is thoroughly *bona fide*. I should also draw your attention to the fact that the expression on the right hand side of (7.20) is not a *Maximin* operation, but rather a *Minimax* operation. Furthermore, that the outer $\min$ operates on $u \in \mathscr{U}$ and the inner $\max$ operates on $x \in X$.

This means that in this *Minimax game,* the u player, namely Nature, plays first. This is in contrast to all the models that we examined in the preceding chapters where the x player, namely DM, makes the first move.

In the discussion that follows I explain the logic behind this definition.

Cheers,
The Author

But before I proceed any further, I want to illustrate how this definition comes into play in the context of the preceding example.

Example

In the case of (7.17), we have

$$z^{(wb)} := \min_{u\in\mathscr{U}} z^{\circ}(u) \tag{7.22}$$

$$= \min_{u} \left\{ u_2\sqrt{u_1^2+1} : u \in [1,10]\times[2,4] \right\} \tag{7.23}$$

$$= 2\sqrt{2}\,,\ u^{(wb)} = (1,2) \tag{7.24}$$

hence,

$$\mathscr{U}^{(wb)} = \{(1,2)\} \tag{7.25}$$

$$X^{(wb)} = X^{*}(1,2) = \{x^{*}(1,2)\} \tag{7.26}$$

[3] Internationally renowned expert on robust decision in the face of severe uncertainty.

where, according to (7.14),

$$x^*(1,2) = \left(\frac{2}{\sqrt{2}}\ ,\ \frac{2}{\sqrt{2}}\right) = \left(\sqrt{2}, \sqrt{2}\right) \tag{7.27}$$

So the inference is that if we solve the instance *Model P(u)* for each $u \in \mathscr{U}$, obtaining the best outcome (over all the associated feasible values of x) for each $u \in \mathscr{U}$, the worst outcome is the one given by $u^{(wb)} = (1,2)$. The best decision for this instance is then $x^*(1,2) = \left(\sqrt{2}, \sqrt{2}\right)$, yielding the worst-best payoff $z^{(wb)} = 2\sqrt{2}$.

But, what will happen if we decide to use the "best-worst case" approach instead[4]?

In this case the problem representing the "best-worst case" would be as follows:

$$\begin{aligned}
z^* :&= \max_{x \in X} \min_{u \in \mathscr{U}} \{g(x;u) : constraints(x;u), \forall u \in \mathscr{U}\} && (7.28)\\
&= \max_{x \in X} \min_{u \in \mathscr{U}} \{u_1 x_1 + x_2 : x_1^2 + x_2^2 \leq u_2^2, \forall u \in \mathscr{U}\} && (7.29)\\
&= \max_{x \in \mathbb{R}^2} \min_{\substack{1 \leq u_1 \leq 10 \\ 2 \leq u_2 \leq 4}} \{u_1 x_1 + x_2 : x_1^2 + x_2^2 \leq u_2^2, \forall u \in [1,10] \times [2,4]\} && (7.30)\\
&= \max_{x \in \mathbb{R}^2} \{x_1 + x_2 : x_1^2 + x_2^2 \leq 4\}\ ,\ u^* = (1,2) && (7.31)\\
&= 2\sqrt{2}\ ,\ x^* = \left(\sqrt{2}, \sqrt{2}\right) && (7.32)
\end{aligned}$$

On the face of it there seems to be no hindrance in doing this as the same results are obtained for the "best-worst-case" approach as in the "worst-best-case" approach.

However, as the following two examples illustrate, this is not always the case.

Example

Consider the following extremely naive case:

$$\begin{aligned}
&\text{Model P(u)}, u \in \mathscr{U} = [0,10]\\
&z^\circ(u) := \max_{x \in [5,20]} \{u + x\} \text{ subject to } 2 + u \leq x \leq 8 + u && (7.33)
\end{aligned}$$

Let us now apply the two approaches and compare the results.

Worst-best-case approach: Here the problem is as follows:

$$\begin{aligned}
z^{(wb)} :&= \min_{0 \leq u \leq 10} \max_{5 \leq x \leq 20} \{u + x : 2 + u \leq x \leq 8 + u\} && (7.34)\\
&= \min_{0 \leq u \leq 10} \{2u + 8\}\ ,\ x^{(wb)}(u) = 8 + u && (7.35)\\
&= 8\ ,\ u^{(wb)} = 0 && (7.36)
\end{aligned}$$

Hence, the optimal decision is $x = x^{(wb)}(0) = 8$.

Best-worst-case approach: Here the problem is as follows:

$$z^* := \max_{5 \leq x \leq 20} \min_{0 \leq u \leq 10} \{u + x : 2 + u \leq x \leq 8 + u, \forall u \in [0,10]\} \tag{7.37}$$

[4] Best with respect to x and worst with respect to to u.

But ... this problem is infeasible!

In other words, there is no $x \in [5, 20]$ that satisfices the constraint $2 + u \leq x \leq 8 + u$ for all $u \in [0, 10]$. Observe that

$$u = 0 \quad \longrightarrow \quad 2 \leq x \leq 8 \tag{7.38}$$

$$u = 10 \quad \longrightarrow \quad 12 \leq x \leq 18 \tag{7.39}$$

Obviously, the existence of a x such that $12 \leq x \leq 8$ can hardly be envisaged.

So, the conclusion is that it is not always the case that if the "worst-best-case" approach yields a problem with feasible solutions, then the "best-worst-case" approach will also yield such a problem.

Example

Consider the case where

$$\begin{aligned} &\text{Model P(u)}, u \in \mathscr{U} = \{1, 2, 3, 4, 5\} : \\ &z^{\circ}(u) := \max_{x \in \{1,2,3\}} g(x; u) \end{aligned} \tag{7.40}$$

where $g(x; u)$ is specified by the following payoff table:

		u				
		1	2	3	4	5
	1	3	2	9	6	4
x	2	8	9	7	8	8
	3	3	5	5	6	4

The results generated by the best "best-worst-case approach" and the "worst-best-case approach" are shown in Figure 7.3.

Best-Worst-Case

		u					
		1	2	3	4	5	SL
	1	3	2	9	6	4	2
x	2	8	9	7	8	8	$\boxed{7}$
	3	3	5	5	6	4	3

Worst-Best-Case

		u				
		1	2	3	4	5
	1	3	2	9	6	4
x	2	8	9	7	8	8
	3	3	5	5	6	4
SL		$\boxed{8}$	9	9	$\boxed{8}$	$\boxed{8}$

Figure 7.3: Results generated by the "best-worst-case" and the "worst-best-case" approach

The SL column in the table associated with the "best-worst-case" approach provides the minimum value of $g(x; u)$ over $u \in \mathscr{U}$ for each (row) value of $x \in X$. The SL row in the table associated with the "worst-best-case" approach provides the maximum value of $g(x; u)$ over $x \in X$ for each (column) value of $u \in \mathscr{U}$.

Clearly, the two approaches yield significantly different results. $\square$

As might be expected, there is a plethora of theorems (called "minimax theorems") that lay down the conditions under which these two approaches yield the same result (See for instance Sion (1958)). For our purposes it suffice to note that the "best-worst-case" approach is at least as "conservative" as the "worst-best-case" approach in the following sense.

Consider the two problems:

$$\textbf{Worst-best-case problem: } z^{(wb)} := \min_{u\in\mathscr{U}} \max_{x\in X} \{g(x;u) : constraints(x;u)\} \tag{7.41}$$

$$\textbf{Best-worst-case problem: } z^{(bw)} := \max_{x\in X} \min_{u\in\mathscr{U}} \{g(x;u) : constraints(x;u), \forall u\in\mathscr{U}\} \tag{7.42}$$

Theorem 7.3.1 BEST-WORST-CASE VS WORST-BEST-CASE

(a) Any decision $x \in X$ that is feasible with respect to the **Best-worst-case problem** *is also feasible with respect to the* **Worst-best-case problem***.*

(b) Assume that both problems have optimal solutions. Then $z^{(wb)} \geq z^{(bw)}$, namely

$$\min_{u\in\mathscr{U}} \max_{x\in X} \{g(x;u) : constraints(x;u)\} \geq \max_{x\in X} \min_{u\in\mathscr{U}} \{g(x;u) : constraints(x;u), \forall u\in\mathscr{U}\} \tag{7.43}$$

That is, the **Best-worst-case problem** *is at least as "conservative" as the* **Worst-best-case problem***.*

PROOF. The first part of the theorem is born out by the $\forall u \in \mathscr{U}$ clause in the formulation of the **Best-worst-case problem** which requires that x satisfice the constraints for all $u \in \mathscr{U}$. To prove the second part, observe that if $(x^{(wb)}, u^{(wb)})$ is an optimal solution to the **Worst-best-case problem**, then

$$\min_{u\in\mathscr{U}} \max_{x\in X} \{g(x;u) : constraints(x;u)\} = \max_{x\in X} \left\{g(x;u^{(wb)}) : constraints(x;u^{(wb)})\right\} \tag{7.44}$$

$$\geq \max_{x\in X} \min_{u\in\mathscr{U}} \{g(x;u) : constraints(x;u), \forall u\in\mathscr{U}\} \tag{7.45}$$

QED

To make sense of these results it is instructive to think of the two problems as two versions of a 2-person game. The difference is in the *order of play.* The point is that from the DM's point of view it is to her advantage to play . . . second, as this enables her to select the best decision relative to the value of u selected by Nature.

To illustrate, consider the feasibility issue.

If the DM plays first, then she must select a decision x that satisfies the constraints for every $u \in \mathscr{U}$ because she has no clue as to what value of u will be selected by Nature. On the other hand, if the DM plays second, then her choice of x is made in response to the specific value of u selected by Nature. Clearly, playing second gives the DM more leeway in her choice of decisions. This is manifested not only in $z^{(wb)} \geq z^{(bw)}$ but also in the first part of Theorem 7.3.1.

And the inference to be drawn from all this is that the conservatism associated with the "best-worst-case" approach is a manifestation of the SEVERE UNCERTAINTY in the value of u which calls upon the DM to take account of every possible/plausible value of u, including the worst value.

The objective of the next example is to illustrate that the **Worst-best-case criterion** can be utterly unsuitable for the pursuit of decisions that are required to be robust over large subsets of $\mathscr{U}$.

Example

Consider the simple case

$$z^{\circ}(u) := \max_{x \in \{a,b\}} g(x;u) \ , \ u \in \mathscr{U} = \mathbb{R} \tag{7.46}$$

where

$$g(x;u) := \begin{cases} 3 & , \ x = a \\ 2 + u^2 & , \ x = b \end{cases} \ , \ u \in \mathbb{R} \tag{7.47}$$

Figure 7.4 gives a graph of the two "branches" of this function.

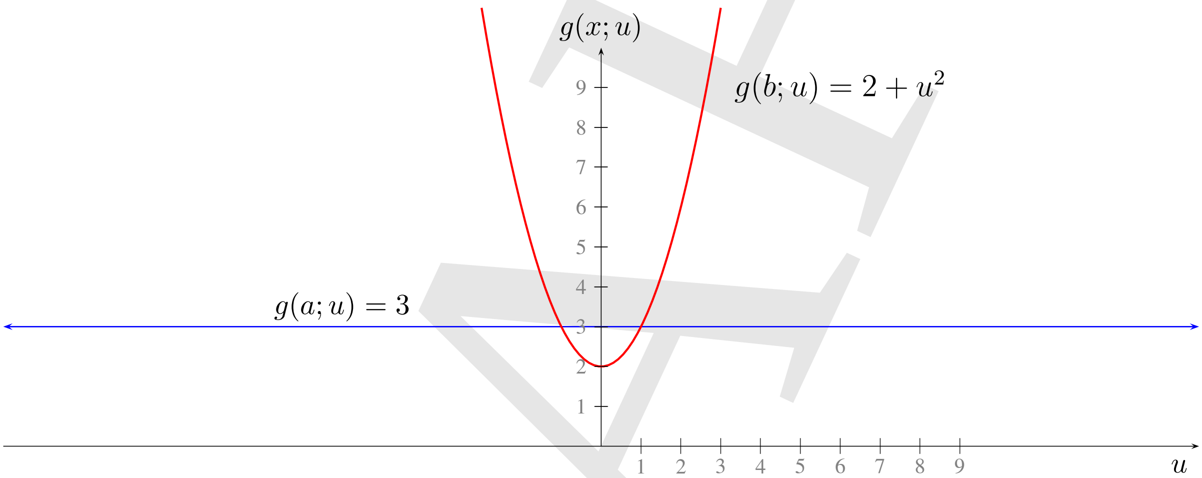


Figure 7.4: The two branches of g

By inspection, for all $u \in (-1.1)$ the best decision is $x = a$ and for all u outside $[-1,1]$ the best decision is $x = b$. The worst-best-cases occur in $(-1,1)$, hence the worst-best decision is $x = a$.

However, given that the objective is to maximize the payoff $g(x;u)$ and that $g(b;u) > g(a;u)$ for all u outside the very small interval interval $[-1,1]$, it clearly makes more sense to regard decision $x = b$ as the more robust against the variations in the value of u over $\mathscr{U}$.

This illustrates a chronic flaw of the "worst-best-case" approach which is due to the $\min\limits_{u \in \mathscr{U}}$ operation representing the "worst" in this clause. That is, this operation may be halted by a single, or a small number, of bad cases, from continuing to explore the "big picture" to thereby miss out on the identification of better outcomes.

And the trouble is that the same is true about the "best-worst-case" approach. That is, the $\min\limits_{u \in \mathscr{U}}$ operation representing the "worst" in the "best-worst- case" approach suffers from the same failing afflicting the $\min\limits_{u \in \mathscr{U}}$ operation representing the "worst" in the "worst-best-case" approach.

Furthermore, it is also important to realize that the seemingly innocuous clause $constraints(x;u)$, $\forall u \in \mathscr{U}$ can also be problematic. For in "reality" it is not always possible/advisable to satisfice ALL the requirements. Thus, it is often necessary to "coordinate" the implementation of this clause with the construction of the parameter space $\mathscr{U}$. I discuss this issue in section 7.5.2.

To illustrate how the combined effect of these two difficulties can adversely affect the formulation of decision rules based on worst-case analysis, let us consider the following example.

Suppose that $\mathscr{U}$ consists of say 987654 elements, and that some given $x' \in X$ satisfices the constraints for about 987648 elements of $u \in \mathscr{U}$, meaning that the constraints are violated for only 6 values of $u \in \mathscr{U}$ (out of 987654 values).

The question is then: should x' be regarded "admissible" or should it be dismissed for violating the constraints?

There are in fact "two horns" to this dilemma:

· Decision x' could be an excellent candidate, except for the violation mentioned above.

· The other decisions are not perfect either.

Ideally, such questions should be investigated carefully on a case by case basis. But the difficulty is that in reality the existence of such cases is masked by the sheer size of a model which impedes a "case by case" treatment of decisions.

In short, the combination of the $\min\limits_{u \in \mathscr{U}}$ operation in the problem's objective function and the stringent demands of the $constraints(x; u), \forall u \in \mathscr{U}$ clause in the problem's constraints, predesposes the "best-worst-case" approach to excessive "conservatism".

And the overall conclusion is that much as the family of parametric models, namely *Model P($\mathscr{U}$)*, and the related **Worst-best-case criterion**, are edifying in their illustration of some of the issues encountered in the formulation of robust optimization models, on their own they cannot provide a suitable framework for robust decision-making.

So, my next task is to examine how a *single* optimization model representing a problem whose optimal solution is robust — in some sense — against the variation in the value of u over $\mathscr{U}$, can be derived from the *family* of parametric problems represented by *Problem P($\mathscr{U}$)*.

As we shall see, this is a tall order!

7.3.3 Robust Counterpart models

Very broadly, the pursuit of robustness in the context of *Model P($\mathscr{U}$)* entails a search of an x that is *robust* against the variation in the value of u over $\mathscr{U}$. More precisely, a search for an x that "performs well" (both with respect to the objective function and the constraints) as the value of u varies over $\mathscr{U}$.

As I noted above, the ideal is a "super-robust" decision, namely a decision that is simultaneously optimal for all the instances of *Model P($\mathscr{U}$)*. But, as I pointed out, such decisions are a rare commodity.

So, let us examine how one would go about formulating a *robust counterpart* to *Model P($\mathscr{U}$)* aimed at recovering a robust x. As suggested by this section's opening statement, this endeavor would be guided by consideration related to the two central constructs of *Model P($\mathscr{U}$)*, namely

· The constraints

· The objective function

Thus, the factor that would decide whether a solution/decision x performs well, namely is robust, on $\mathscr{U}$ with regard to the *constraints* stipulated by *Model P($\mathscr{U}$)*, would be an x satisficing these constraints over a "large" subset of $\mathscr{U}$.

And from the standpoint of the *objective function* stipulated by *Model P(u)*, this would be the case if $g(x; u)$ would "not be much' smaller" than $z^{\circ}(u)$ over a "large" subset of $\mathscr{U}$.

Or to give it a more formal formulation,

> Solution/decision x would be regarded robust over $\mathscr{U}$ if x would satisfice the constraints over a "large" subset of $\mathscr{U}$ and $g(x;u)$ would "not be much smaller" than $z^{\circ}(u)$ over a "large" subset of $\mathscr{U}$.
>
> Reminder: $\begin{aligned} z^{\circ}(u) := & \max_{x\in X}\ g(x;u) \\ \text{subject to}\ & constraints(x;u) \end{aligned}$

It goes without saying that to be able to forge this formulation into a "working definition", I would have to be a great deal more specific about what I mean by "large" and "not much smaller" in this statement.

> **G'day Moshe:**
>
> I realize that you plan to discuss the issue that I am about to raise now, at a later stage in this chapter, still ... I believe that it is important to note it in the present context.
>
> This has to do with the formulation of a "working definition" of robustness.
>
> It is important to point out that a key consideration in formulating a "working definition" of robustness ought to be the solvability of the optimization problem induced by this definition. That is, the definition ought to give rise to an optimization problem that would be amenable to solution either by existing methods and/or by special methods developed specifically for this purpose.
>
> The brief history of *robust optimization* makes it eminently clear that otherwise a definition of robustness may end being hollow.
>
> Cheers,
> Fred
> Internationally Renowned Expert on Robust Decision Under Severe Uncertainty

While I agree with Fred's observation, I shall not allow this consideration to dictate my discussion in this chapter. By this I mean that I fully intend to discuss robustness models that may be deemed impractical because the optimization problems that they give rise to are extremely complex, hence not amenable to an "exact" solution with existing methods.

In due course I shall explain the rationale behind this deliberate decision. At this point it suffices to indicate that many problems that were deemed "too difficult to solve" say 20 years ago, are relatively easy to solve now. So it would unwise to let the current (2011) state of the art in optimization tools dictate the scope of a discussion on the modeling aspects robustness against severe uncertainty.

So what's next?

In the remainder of this chapter I shall discuss a number of robust optimization models, that is a number of so called *robust counterparts*, of *Model P(u)*. My discussion of these models will revolve around the following questions:

- The object with respect to which robustness is sought:

  · Objective function · Constraints · Both

- The scope of the robustness analysis:

  · Global · Partial · Local

In the case of the "Partial scope" and "Local scope", a further distinction can be made between two types of domains (subsets of the parameter space $\mathscr{U}$) on which the robustness analysis is conducted:

· Pre-determined · Determined impromptu

The former refers to cases where the domain of the robustness analysis is determined a priori, and the latter refers to situations where the domain is marked out by the robustness analysis itself.

So for example, we can distinguish between say a "pre-determined, local robustness of the constraints" and a "global robustness of the objective function". Note that according to these classification criteria, the robustness defined/determined by the *Radius of Stability* model — hence by *info-gap decision theory* — is an "impromptu, local robustness of the constraints".

Finally, the manner in which the quest for "robustness" is incorporated in a model.

As we shall see, the pursuit of robustness can be incorporated in an optimization model in a number of ways. In this discussion I shall distinguish between "robustness requirements" and "robustness measures".

"Robustness requirements" are stipulations laid down by the model that the solution to an optimization problem must meet in order to count robust over the parameter space $\mathscr{U}$.

A "robustness measure" is a recipe/gauge for determining the robustness of a solution to the optimization problem considered. If the robust counterpart model is a *Maximin model,* then the robustness of a decision is its *security level*. In some cases there is no explicit measure of robustness, only robustness requirements.

7.4 Global, partial and local robustness

As indicated in Chapter 5, the terms *global, partial* and *local* designate the domains of the robustness analysis, that is the subsets of the parameter space $\mathscr{U}$ over which the robustness analysis is conducted. I also pointed out that the other side of the coin is that the scope of this analysis is the main criterion by which robustness is classified. We thus have *global, partial* and *local* robustness.

It is important to appreciate though that this classification is not intended to be taken as "an academic exercise". Namely, it is not driven merely by the desire to unravel the mysteries of the parameter space $\mathscr{U}$. However gratifying such an investigation may be, the classification of robustness has immediate implication for the nature of the results obtained from a robustness analysis. This means of course that a correct reading of the type of robustness yielded by the analysis: *global, partial* or *local* robustness, is crucial for a correct interpretation of the results that one obtains for a problem considered.

A correct reading of the domain over which the robustness analysis is conducted is, of course, of the essence in the context of *severe uncertainty*, especially if the uncertainty is taken to be modeled by a non-probabilistic, likelihood-free model of uncertainty. Because as you will recall one of the main ingredients of a model of uncertainty is the parameter space $\mathscr{U}$:

$$\mathscr{U} = \text{set of all possible/plausible values of the parameter } u.$$

So, if the uncertainty under consideration is *severe*, meaning that it is not quantified by a likelihood/probabilistic model, then methodologically, any robustness model seeking robustness against the severe uncertainty in the true value of u must be conducted over $\mathscr{U}$, or over a "good approximation" thereof. In other words, a *local* robustness analysis will not do!

And to be more specific, if our parameter space $\mathscr{U}$ is say the island of *Australia,* then we would have to conduct the robustness analysis over the entire . . . island of *Australia*, as conducting the robustness analysis only over say the states of *South Australia* and *New South Wales* will not do.

Figure 7.5: Map of an island

In other words, if we seek to determine the robustness of a solution to an optimization problem, say a problem defined by *Model P(u)*, against variations in the value of u over $\mathscr{U}$, then . . . given that $\mathscr{U}$ is the set of all possible/plausible values of u, the robustness analysis must be conducted on $\mathscr{U}$.

I imagine that some readers may wonder why on earth do I go to such great lengths to explain a point that seems rather obvious.

Let me point out then that however obvious this point may appear to be, as attested by the *info-gap literature,* it is clearly not obvious to all. And this is precisely the rationale behind the title of this book, more specifically the “Fooled by Robustness” phrase in the title. For, as we shall see in the second part of the book, it is apparently quite easy to mistake a robustness analysis conducted over a subset of $\mathscr{U}$, in fact a minuscule subset of it, for an analysis over $\mathscr{U}$ itself. I address this issue below in the subsection “Local robustness".

This clear, let us resume the discussion on the differences between global, partial and local robustness.

7.4.1 Global robustness

I shall consider three types of global robustness models, or more precisely three types of global *robust counterparts* of *Model P($\mathscr{U}$).* The model types are determined by the kind of global robustness that each type aims to obtain, namely:

- Global robustness with respect to the *constraints.*
- Global robustness with respect to the *objective function.*
- Global robustness with respect to *both the constraints and the objective function.*

I begin with an instance of *Model P($\mathscr{U}$)* where the parameter u affects only the constraints. In this case, *Model P(u)* will induce to the following parametric model:

$$\text{Model } C(u), u \in \mathscr{U} :$$
$$z^c(u) := \max_{x \in X} \ g(x) \ \text{ subject to } \ constraints(x; u) \tag{7.48}$$

observing that the objective function g does not depend on u.

The following is an intuitive robust counterpart of *Model C($\mathscr{U}$)*:

$$\text{Model } RC :$$
$$z^{(rc)} := \max_{x \in X} \ g(x) \ \text{ subject to } \ constraints(x; u), \forall u \in \mathscr{U}_x \tag{7.49}$$

where for each $x \in X$ the set $\mathscr{U}_x$ is a subset of $\mathscr{U}$ containing all the possible/plausible values of u pertaining to decision $x \in X$. In many cases $\mathscr{U}_x = \mathscr{U}, \forall x \in X$. With no loss of generality assume that $\mathscr{U}$ is the smallest set containing $\mathscr{U}_x$ for all $x \in X$.

Observe that *Model RC* is a *Maximin model* expressed in terms of the *MP format.*

Example

Consider the following simple robust optimization problem, observing that robustness is sought here only with respect to the constraints:

$$z^{(rc)} := \max_{x \in \mathbb{R}^2} \{x_1 + x_2\} \tag{7.50}$$

$$\text{s.t.} \quad x_1 + 2x_2 \leq 10 + u; \ x_1 - x_2(1 + |u - 1| \leq 4; \ x_2 \leq 5 - u; \ x_1, x_2 \geq 0, \forall u \in [0, 2] \tag{7.51}$$

Set $\mathscr{U} = \mathscr{U}_x = [0, 2], \forall x \in \mathbb{R}^2$. Note that the objective function is *linear* with the two decision variables (x_1, x_2) and that for any given $u \in \mathscr{U}$, the constraints are also *linear* with these variables. The problem's extreme simplicity enables its solution by inspection.

First, let us determine the worst $u \in \mathscr{U}$ for each of the constraints, and the resulting feasible domain of the decision variables. This will enable to determine the set of feasible solutions associated with the maximization problem. By inspection:

Constraint	worst $u \in \mathscr{U}$	worst-case constraint
$x_1 + 2x_2 \leq 10 + u$	0	$x_1 + 2x_2 \leq 10$
$x_1 - x_2(1 + \|u - 1\|) \leq 4$	1	$x_1 - x_2 \leq 4$
$x_2 \leq 5 - u$	2	$x_2 \leq 3$

It therefore follows that the feasible domain of $x = (x_1, x_2)$ is determined by the constraints

$$x_1 + 2x_2 \leq 10 \ ; \ x_1 - x_2 \leq 4 \ ; \ x_2 \leq 3 \ ; \ x_1, x_2 \geq 0 \tag{7.52}$$

This is shown in Figure 7.6, where the shaded area depicts the feasible domain. The optimal solution $x^* = (6, 2)$ is obtained graphically, yielding $z^{(rc)} = 8$. The blue line represent the objective function[5].

[5] To readers who are not familiar with linear programming's "graphic method" ... this may look like magic. But, it is simple linear algebra "made real".

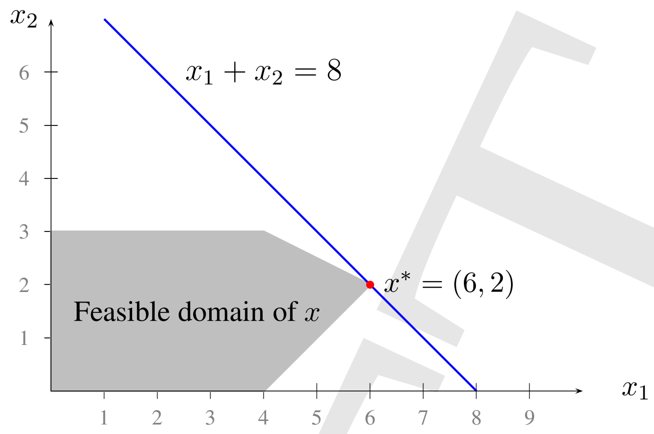


Figure 7.6: Feasible domain and optimal solution

In cases where the robustness constraints, namely $constraints(x; u), \forall u \in \mathscr{U}_x$, are too stringent, and as a consequence the optimization problem has no feasible solution, or the value yielded for z^{rc} is too small, then it may prove necessary to relax these constraints.

The objective of the next example is to illustrate the merit of the $\mathscr{U}_x$ notation, which enables indicating that the set of possible/plausible values of u may depend on x.

Example

Consider the following real-world problem where a decision has to be made as to the right location for a kosher Chinese restaurant. The decision space consists of the possible/plausible locations for instance:

$$X = \{\text{Melbourne, Tucson, Pretoria, Tel Aviv, Hong Kong}\} \tag{7.53}$$

Assume also that the plan is to open the restaurant to the public in 3 years time.

There are 28 parameter of interest in this case. One of these, call it u_k, will no doubt be the special costs inherent in the kosher character of the project (licenses, special products, etc.). Another parameter of interest is the size of the clientele, call it u_c. The variability/uncertainty in the true values of these two severely uncertain parameters — and perhaps the values of the other 26 parameters of interest — clearly hangs on the location x. It is thus convenient to formulate a separate parameter space $\mathscr{U}_x$ for each location $x \in X$.

I should add, though, that this is a mere technicality because we can just as well employ an expanded parameter space $\mathscr{U}$ where relevant subsets are invoked depending on the location x. $\square$

Next in line is a robustness model where robustness is sought only with respect to the objective function. Here *Model P(u)* induces the following parametric model:

$$\begin{aligned}&\text{Model } G(u), u \in \mathscr{U} :\\ &z^{(g)}(u) := \max_{x \in X} \ g(x; u) \ \text{ subject to } \ constraints(x)\end{aligned} \tag{7.54}$$

observing that the objective function g is independent of u.

The following is an intuitive robust counterpart:

Model RG :

$$z^{(rg)} := \max_{x \in X} \min_{u \in \mathscr{U}_x} g(x;u) \text{ subject to } constraints(x) \tag{7.55}$$

Clearly, this is a constrained *Maximin model*[6].

Example

Consider the following very simple robust optimization problem:

$$z^{(rg)} := \max_{x \in \mathbb{R}^2} \min_{u \in \mathscr{U}} \left\{ \frac{x_2}{x_1} u_1^2 + u_2 \right\} \text{ subject to } x_1 + x_2 \geq 4, x_1 - 2x_2 \leq 1, -x_1 + x_2 \leq 1 \tag{7.56}$$

where

$$\mathscr{U} = \{u \in \mathbb{R}^2 : u_1 + u_2 \geq 4\} \tag{7.57}$$

For a given pair $(x_1, x_2) \in \mathbb{R}^2$ satisficing the above constraints, the inner minimization problem over $u \in \mathscr{U}$ is as follows:

$$\min_{u \in \mathbb{R}^2} \left\{ \frac{x_2}{x_1} u_1^2 + u_2^2 \right\} \text{ subject to } u_1 + u_2 \geq 4 \tag{7.58}$$

A bit of elementary calculus yields the following optimal solution for u as a function of x:

$$u(x) = \left(\frac{4x_1}{x_1 + x_2}, \frac{4x_2}{x_1 + x_2} \right) \tag{7.59}$$

So the outer maximization problem corresponding to x is as follow:

$$\max_{x \in \mathbb{R}^2} \left\{ \frac{x_2}{x_1} u_1^2(x) + u_2^2(x) \right\} = \max_{x \in \mathbb{R}^2} \left\{ \frac{x_2}{x_1} \left(\frac{4x_1}{x_1 + x_2} \right)^2 + \left(\frac{4x_2}{x_1 + x_2} \right)^2 \right\} \tag{7.60}$$

subject to $x_1 + x_2 \geq 4, x_1 - 2x_2 \leq 1, -x_1 + x_2 \leq 1$.

But, this complex and somewhat intimidating problem can be simplified to yield a more genial problem:

$$\max_{x \in \mathbb{R}^2} \left\{ \frac{16x_2}{x_1 + x_2} \right\} \text{ subject to } x_1 + x_2 \geq 4, x_1 - 2x_2 \leq 1, -x_1 + x_2 \leq 1 \tag{7.61}$$

This is a simple *linear fractional programming problem* that can be solved semi-graphically as shown in Figure 7.7. The optimal solution is $x^* = (1.5, 2.5)$, yielding $z^{(rg)} = 10$. The optimal value of u is $u^* = (1.5, 2.5)$. $\square$

To conclude I consider a robustness model where robustness is sought with respect to both the objective function and the constraints. Namely, I consider a case of *Model P(u)* where both the objective function and the constraints depend on u.

The question is then what would be the appropriate robust counterpart in this case?

[6] Of course, some readers may prefer to use the associated Minimax *regret* model for this purpose.

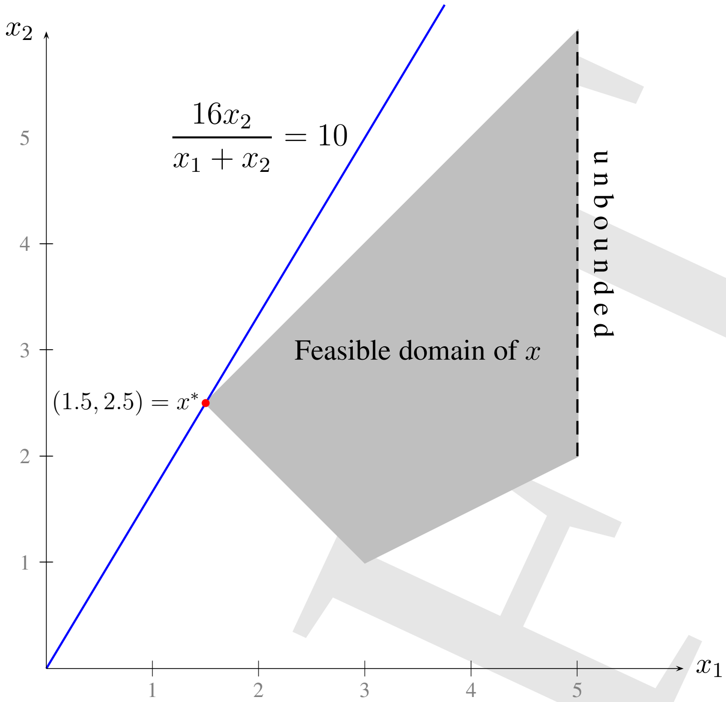


Figure 7.7: Feasible domain of x and and optimal solution

So, consider this model:

Model RGC :

$$z^{(rgc)} := \max_{x\in X} \min_{u\in \mathscr{U}_x} g(x;u) \ \text{ subject to } \ constraints(x;u), \forall u \in \mathscr{U}_x \tag{7.62}$$

Clearly, this is a constrained *Maximin model.*

Example

Consider the case where

$$z^{(rgc)} := \max_{x\in X} \min_{u\in \mathscr{U}} g(x;u) \ \text{ subject to } \ r(x;u) \leq l(x;u), \forall u \in \mathscr{U} \tag{7.63}$$

$X = \{1,2,3,4,5\}$, $\mathscr{U}_x = \mathscr{U} = \{1,2,3,4,5\}, \forall x \in X$ and the objective function and the constraints are specified by the tables given in Figure 7.9. Think of $l(x;u)$ as the performance level of decision x given u and of $r(x;u)$ as the performance requirement imposed on decision x given u.

Objective function

		u				
	$g(x;u)$	1	2	3	4	5
	1	3	2	9	6	4
	2	8	9	7	8	8
x	3	3	5	5	6	4
	4	6	7	3	4	5
	5	2	8	9	3	3

Performance levels

		u				
	$l(x;u)$	1	2	3	4	5
	1	5	8	4	8	9
	2	6	7	4	4	6
x	3	1	2	3	3	5
	4	4	5	3	2	1
	5	6	9	5	4	3

Performance requirements

		u				
	$r(x;u)$	1	2	3	4	5
	1	4	7	4	7	8
	3	6	7	5	3	5
x	3	3	2	4	2	6
	4	3	5	2	1	1
	5	5	9	4	2	3

Figure 7.8: Objective function and constraints

The results are shown in Figure 7.9, where the entries in the boolean table indicate whether the respective (x, u) values satisfice the performance constraints.

The inference is that only $x = 1$, $x = 4$, and $x = 5$ satisfice the robustness constraint. Hence, the standard Maximin payoff operation is conducted only on the rows corresponding to these decisions. The results are shown in Figure 7.9b. The conclusion is that the optimal decision is $x = 4$, yielding $z^{(rgc)} = 2$.

Constraints

		u 1	2	3	4	5	
	1	1	1	1	1	1	1
	2	1	1	0	1	1	0
x	3	0	1	0	1	0	0
	4	1	1	1	1	1	1
	5	1	1	1	1	1	1

(a)

Maximin of $g(x; u)$

	$g(x;u)$	u 1	2	3	4	5	SL
	1	3	2	9	6	4	2
	2	8	9	7	8	8	–
x	3	3	5	5	6	4	–
	4	6	7	3	4	5	3
	5	2	8	9	3	3	$\boxed{2}$

(b)

Figure 7.9: Objective function and constraints

G'day Moshe,

I submit that it is important to call the reader's attention to the fact that however formidable, *Model RGC* is not more general than *Model RC*.

Cheers,
Fred
Internationally renowned Expert on Robust Decision Under Severe Uncertainty

Fred is right.

Note that the MP format of *Model RGC* is as follows:

$$z^{(rgc)} := \max_{\substack{x \in X \\ v \in \mathbb{R}}} v \text{ subject to } v \leq g(x;u), constraints(x;u), \forall u \in \mathscr{U}_x \tag{7.64}$$

But this model can be stated as a *Model RC* associated with the decision variable $y = (y_1, y_2) = (v, x) \in Y := \mathbb{R} \times X$. That is, consider the following *Model RC* associated with this new decision variable:

$$z^{(rgcy)} := \max_{y \in Y} h(y;w) \text{ subject to } Constraints(y;w), \forall v \in W_y \tag{7.65}$$

where $h(y;w) = y_1$, $W_y = \mathscr{U}_{y_2}$, and $Constriants(y;w)$ is the set of constraints on y consisting of the constraints in $constraints(y_2;w)$ and the constraint $y_1 \leq g(y_2;w)$.

It follows then that

$$z^{(rgcy)} := \max_{y \in Y} h(y;w) \text{ subject to } Constraints(y;w), \forall w \in W_y \tag{7.66}$$

$$\equiv \max_{\substack{x \in X \\ v \in \mathbb{R}}} v \text{ subject to } v \leq g(x;u), constraints(x;u), \forall u \in \mathscr{U}_x \tag{7.67}$$

The conclusion therefore is that, although *Model RGC* has the appearance of a more impressive model, it is in fact an instance of the seemingly less impressive *Model RC*.

7.4.2 Partial robustness

Models seeking *partial robustness* are similar to models of global robustness except that here, for each $x \in X$ the robustness analysis (e.g. worst case analysis with respect to u) is conducted on a given subset of $\mathscr{U}_x$, say V_x, rather than on $\mathscr{U}_x$.

There is therefore no need elaborate on this type of model. However to illustrate its working let us consider the following case.

Consider a situation where $\mathscr{U}_x = [0, \infty), \forall x \in X$, in which case the parameter u is a non-negative number. Now suppose that while it is a fact that the true value of u can be any element of $\mathscr{U}_x$, the analyst, Jack, is quite confident that the true value of u is somewhere in the set that is the union of the following three intervals: $I(1) = [300, 800]$, $I(2) = [5000, 12000]$, and $I(3) = [200000, 500000]$.

In this case, Jack would probably seek to conduct the robustness analysis on the set $V_x = I(1) \cup I(2) \cup I(3)$ instead of $\mathscr{U}_x = [0, \infty)$.

But this give rise to a tricky question.

If Jack is indeed so confident that the true value of u is in one of these three intervals, hence in V_x, why bother with $\mathscr{U}_x$ to begin with, especially if it is so much larger than V_x?

But more than this, if Jack is indeed so confident that the true value of u is in V_x, how does this position square with the basic premiss that the relevant uncertainty model is likelihood-free?

That said, I want to make it clear that I am not arguing against the use of models of *partial robustness*. What I do argue, though, is that if you determine to use a model of *partial robustness,* then you had better comply with at least these basic “requirements”:

- Give a definite specification to the assumed parameter space $\mathscr{U}$.
- Give a definite specification to the assumed parameter space $\mathscr{U}_x$ for each $x \in X$.
- Delineate clearly the subset V_x of $\mathscr{U}_x$ in which robustness is sought for decision $x \in X$.
- Give a cogent justification to your proposition to use V_x rather than $\mathscr{U}_x$ as the domain of the robustness analysis.
- If possible, compare the results obtained from the robustness analyses conducted on the two domains. Namely, compare the results obtained from the analysis on V_x to those obtained from the analysis on $\mathscr{U}_x$.

The justification commonly given to the proposition to resort to *partial robustness* models is that a global robustness model (in the case considered) stands to yield conservative, namely extremely costly, results. In such cases one may opt to partition the parameter space $\mathscr{U}$ into two disjoint domains. A “normal” domain and an “abnormal” domain, where V represents the “normal” domain. The robustness analysis would then be confined to the “normal” domain.

The question then arising is: how is one going to handle the “abnormal” domain?

This brings us to Taleb’s *Black Swans* and to the phrase *Fooled by Robustness* in the title of this book. I shall have a great deal more to say about this matter in the second part of the book.

For now, I merely want to point out that a robustness analysis that is claimed to be able to handle rare events, catastrophes, shocks and surprises (as the *info-gap literature* indeed claims about *info-gap’s robustness analysis*), must have the capabilities to deal, in one way or another, with the “abnormal” domain of the parameter space under consideration. A failure to deliver on such claims amounts to a misrepresentation of the robustness analysis in question.

7.4.3 Local robustness

This type of robustness is a special case of partial robustness. Namely, in this case the robustness analysis is conducted on a subset V_x of $\mathscr{U}_x$ which is a *neighborhood* of a point $\tilde{u}_x \in \mathscr{U}_x$. It can be a very small neighborhood of $\mathscr{U}_x$. Let $\mathcal{N}(\rho_x, \tilde{u}_x)$ denote this neighborhood, recalling that ρ denotes the radius of the neighborhood and $\tilde{u}$ denotes its center.

That said, I imagine that careful readers would be greatly surprised that I propose to discuss *local robustness* in a chapter that is devoted to a discussion on robustness to **severe** uncertainty. I imagine that they would argue as follows:

> Are you suggesting that a *local robustness analysis* can so much as be contemplated in a framework where $\mathscr{U}_x$ represents the set of all the possible/plausible values of u pertaining to decision x and the model of uncertainty is likelihood-free?
>
> Surely, such a proposition had better be backed up by good answers to these questions:
>
> · What exactly does $\mathcal{N}(\rho_x, \tilde{u}_x)$ represent?
>
> · Who determines the location ($\tilde{u}_x$) and size (ρ_x) of this neighborhood, and how?
>
> · What guarantee is there, if any, that the robustness determined for this neighborhood adequately reflects the robustness for $\mathscr{U}_x$?

I shall address these questions at a later stage. But to justify my decision to take up *local robustness* in this context, I want to point out that my experience of the past seven years has shown that not even experienced scholars and analysts can escape being “fooled by" a *local analysis.* That is, even experts specializing in risk analysis and related areas, may not realize that the analysis that they conduct is a *local analysis,* or they may not appreciate the ramifications of a *local analysis,* and so on. It is important therefore to demonstrate how one can be “fooled" by a *local analysis*.

To illustrate what I have in mind, consider the following decidedly fictitious story. Its main protagonists are the parametric space $\mathscr{U}$ and its small subset V shown in Figure 7.10.

> Dear Kelly,
>
> As instructed, I conducted the robustness analysis on your system using $\mathscr{U}$ as the parametric space.
>
> I am pleased to report that, based on this analysis, your new system is more robust than the old one against the severe uncertainty in the true value of $u \in \mathscr{U}$.
>
> Please note that according to the robustness model that I used in the analysis — a *Radius of Stability model* — the results of the analysis, hence the robustness of the new system, are completely independent of the performance of the system outside the set V shown in Figure 7.10.
>
> Isn’t it amazing what math can do?!
>
> I shall contact you about the next consulting job when I return from the French Riviera at the end of the summer.
>
> Cheers,
> P.

So while P. is on holiday on the French Riviera, a perplexed Kelly decides to find out how on earth could a local analysis on a very small subset of $\mathscr{U}$ determine the robustness of the new system against the severe uncertainty in the true value of the parameter u.

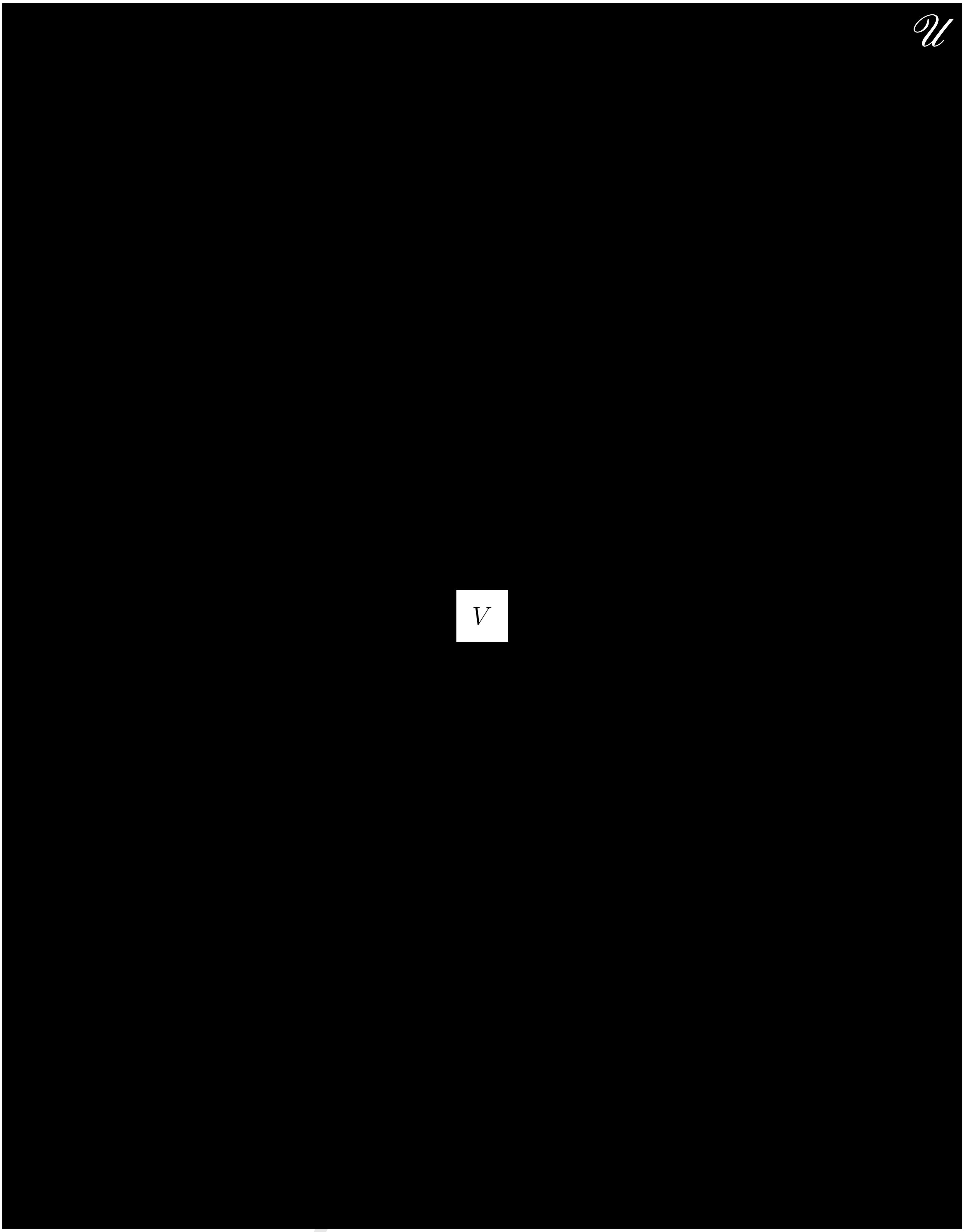
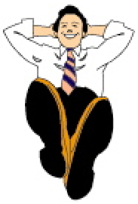


Figure 7.10: Local analysis: $\mathscr{U}$ vs V

As P. is incommunicado, Kelly very reluctantly contacts Fred[7] for a second opinion on P.'s report. A week later Kelly receives the following note from Fred.



> **G'day Kelly,**
>
> I am afraid that I have bad news.
>
> I'll tell you all about it when I see you tomorrow at the club.
>
> Cheers,
> Fred
> Internationally Renowned Expert on Robust Decision Under Severe Uncertainty

What follows is, more or less, the explanation that Fred gave Kelly at the tennis club after their mixed doubles match.

The robustness model that P. used, namely a *Radius of Stability model*, is a model of *local robustness*. The center points of models of of this type are the centers of the neighborhoods $\mathcal{N}(\rho_x, \tilde{u}_x)$, $x \in X$ on which the robustness analysis is conducted. Typically $\tilde{u}_x$ represents a *reference point* of u associated with decision x. In cases where the variability in the value of u is due to uncertainty, this center point is typically a *point estimate* of the true value of u.

Under severe uncertainty this estimate is assumed to be a poor indication of the true value of u and is likely to be substantially wrong. It is often called a "guess". Furthermore, the uncertainty model is non-probabilistic and likelihood-free, hence the choice of this "guess" cannot be justified on the basis of its "likelihood".

This is not a good starting point, is it?

Indeed, the idea of resolving the difficult task of decision-making under severe uncertainty by conducting an analysis in the neighborhood of a single "guess" of the true value of the parameter of interest is not very promising.

This difficulty is made poignant by the following relevant question:

> How does one explain the exclusion from the analysis of the *No Man's Land* part of $\mathscr{U}_x$, namely
>
> $$NML(\rho_x, \tilde{u}_x, \mathscr{U}_x) := \mathscr{U}_x \setminus \mathcal{N}(\rho_x, \tilde{u}_x) \tag{7.68}$$

In Figure 7.10 the *No Man's Land* part of $\mathscr{U}$ is represented by the black area. Needless to say, this question is particularly important in cases where the parametric space $\mathscr{U}_x$, hence by implication also $NML(\rho_x, \tilde{u}_x, \mathscr{U}_x)$, are **unbounded.**

As far as recipes for determining the size (radius) of the neighborhood $\mathcal{N}(\rho_x, \tilde{u}_x)$ on-the-fly is concerned, the one used by the *Radius of Stability* model comes immediately to mind:

> For decision $x \in X$, conduct the local robustness analysis on the set $V_x = \mathcal{N}(\rho_x, \tilde{u}_x)$, where ρ_x denotes the *Radius of Stability* of decision $x \in X$. That is, let ρ_x be the largest value of ρ such that $constraints(x; u)$ are satisficed for all $u \in \mathcal{N}(\rho, \tilde{u}_x)$.

Voilá!

Fred then went on to explain why this is in fact the wrong methodology for determining robustness to sever uncertainty. But, as I go into all this in the second part of the book, for now it suffices to note

[7] Internationally renowned expert on robust decision in the face of severe uncertainty.

that Fred was particularly scathing about the proposition to determine robustness to severe uncertainty on grounds of a local robustness analysis in the neighborhood of a “guess” of the true value of the parameter of interest.

This, he argued, amounts to a flagrant violation of the following two universally accepted maxims:

· Garbage In – Garbage Out (GIGO)

· Results are only as good as the estimate on which they are based.

It goes without saying that Fred was quick to suggest that, for a reasonable fee, he would be most willing to conducted a proper *global* robustness analysis of the system.

7.4.4 Explore but ignore

I imagine that some readers would be curious to know on what grounds P. could justify the use of model of local robustness in this case. To crystalize the issues at play here, consider a case where $\mathscr{U}$ is a square of side 100, whereas V is a square of side 1 at the center of $\mathscr{U}$, so $\mathscr{U}$ is 10000 larger than V.

Clearly, a robustness analysis that is confined to V is *local* compared to a robustness analysis that is conducted on $\mathscr{U}$.

So how in the world can one possibly determine whether the system under consideration is robust against the severe uncertainty in the true value of u on grounds of its performance over the set V?

It is indeed hard to imagine what argument can P. possibly adduce to back up her analysis. But ... based on the bits and pieces that I have come across in the past few years, I imagine that she might argue as follows:

> The fact that my robustness analysis yields results that are independent of the behavior/performance of the new system outside the set V does not mean that the analysis does not explore the entire parameter space $\mathscr{U}$. In fact, my analysis does explores the entire parameter space $\mathscr{U}$, but ... it “ignores” the performance of the system over the space outside V.

Or to put a more “subtle" spin on her claim, she might argue as follows:

> The fact that my robustness analysis yields results that are independent of the behavior of the system under consideration outside set V does not mean that the analysis itself does not explore the entire parameter space $\mathscr{U}$. In fact, the robustness analysis does explores the entire parameter space $\mathscr{U}$, but the results of the analysis are independent of the performance of the system outside V.

As Kelly has some mathematical savvy, she realizes that P. claims that:

> I can formulate a *local* robustness analysis on $V \subset \mathscr{U}$ as a *global* robustness analysis on $\mathscr{U}$.

But surely, Kelly would counter:

> The fact that a *local* robustness analysis on $V \subseteq \mathscr{U}$ can be formulated as a *global* robustness analysis on $\mathscr{U}$ does not alter by one iota the clear distinction between a local and a global robustness analysis. Because, nothing can alter the fact that P.’s analysis is local in the sense that the results that it generates are completely independent of the performance of the system outside V and the fact that V is much smaller than $\mathscr{U}$.

In other words, P. would have to engage in pure scholasticism to explain the difference between:

- A robustness analysis that explores $V \subset \mathscr{U}$.
- A robustness analysis that explores $\mathscr{U}$, but ... ignores the impact of values of u outside $V \subset \mathscr{U}$ on the performance of the system.

P. may well be interested in doing exactly this on her return from the French Riviera at the end of the summer. But I do not intend to open this book to such scholasticism. Nor do I intend to request the UN to install international observers on $\mathscr{U}$ to find out whether P. analysis indeed explores the entire uncertainty space.

More on this in the second part of the book.

7.5 Modeling issues

I now turn to a discussion of the modeling aspects of *robust* optimization where my objective is to illustrate how the various robust counterparts of the parametric *Model P(u)* are constructed.

And to begin, let us remind ourselves that *Model P(u)* is rooted in this optimization model:

$$\begin{aligned}&\text{Model } P:\\ &z^* := \max_{x\in X}\ g(x) \text{ subject to } constraints(x)\end{aligned} \tag{7.69}$$

That is,

$$\begin{aligned}&\text{Model } P(u), u \in \mathscr{U}:\\ &z^\circ(u) := \max_{x\in X}\ g(x;u) \text{ subject to } constraints(x;u)\end{aligned} \tag{7.70}$$

The first item on my agenda is to explain why I employ an "optimization model" as a paradigm for robust decision-making. To do this I am going to discuss this matter under the heading "optimizing vs satisficing".

Readers who are familiar with certain sectors of the economics literature, the decision-making literature, or the literatures of certain disciplines of the social sciences, are no doubt aware that a great fuss is made by certain authors about an alleged superiority of "satisficing" over "optimizing". For the benefit of readers who are not familiar with this thesis I ought to point out that proponents of the "satisficing is better than optimizing " thesis argue that an *optimizing paradigm* only seeks to maximize/minimize cost or what have you (which is bad), whereas a *satisficing paradigm* seeks to satisfy requirements (which is good).

As you might have guessed, this topic takes pride of place in the *info-gap literature* where the thesis that "satisficing" is superior to "optimizing" is claimed to be corroborated not only by human behavior but also by the behavior of (some) animals. The latter, according to this thesis, seek to satisfy their needs rather than optimize their performance, or maximize their payoff, or minimize their cost, or whatever.

7.5.1 Optimizing vs satisficing

So before I embark on this short discussion I want to make it abundantly clear that I am joining the debate not because I think that there is much merit to this entire discourse about "optimizing vs satisficing".

Indeed, I regard this discussion to be pointless, because as I show, "satisficing" is a degenerate form of "optimizing". I am joining the debate because I want to reject outright the notion that, as a methodology for decision-making, "satisficing" is superior to "optimizing".

So, let us begin with a quick look at the bare facts.

Let Y be some set and let $req(y)$ be a set of requirements (constraints) imposed on $y \in Y$. We say that $y \in Y$ is a *satisfactory* element of Y if y satisfices the requirements specified by $req(y)$. If $y \in Y$ does not satisfice these requirements, then it is said to be "unsatisfactory". Let Y^* denote the subset of Y consisting of all the satisfactory elements of Y. Then $Y \setminus Y^*$ is the set of "unsatisfactory" elements of Y.

With this in mind, consider this:

Satisficing problem:

Find a $y \in Y$ that satisfices the requirements specified by $req(y)$. (7.71)

Now, let I_Y denote the *indicator function* of Y^*, that is define

$$I_Y(y) := \begin{cases} 1 & , \ y \text{ satisfices } req(y) \\ 0 & , \ y \text{ does not satisfice } req(y) \end{cases} \ , \ y \in Y \tag{7.72}$$

Next consider this:

$$\textbf{Optimization problem:} \quad z := \max_{y \in Y} \ I_Y(y) \tag{7.73}$$

Then clearly, by inspection,

Theorem 7.5.1 THE FUNDAMENTAL THEOREM OF THE "OPTIMIZING VS SATISFICING" DEBATE:
The **Satisficing problem** *and the* **Optimization problem** *are* EQUIVALENT *in that both generate the same solutions. That is, $y \in Y$ is a solution to the satisficing problem specified by (7.71) iff it is an optimal solution to the optimization problem specified by (7.73).*

For example, consider the case where $Y = [0, 5]^2$ and the set $req(y)$ consists of the following two requirements:

$$(y_1 - 3)^2 + (y_2 - 2)^2 \leq 4 \tag{7.74}$$

$$y_1 + y_2 \geq 4 \tag{7.75}$$

This is shown in Figure 7.11, where the blue line represents the boundary of the feasible region specified by the requirement $y_1 + y_2 \geq 4$ and the red circle represents the boundary of the feasible region specified by the requirement $(y_1 - 3)^2 + (y_2 - 2)^2 \leq 4$. The set of satisfactory elements of Y is represented by the shaded area.

So, in this case the function I_Y is defined on Y as follows: $I_Y(y) = 1$ iff y is in the shaded area. Outside the shaded area $I_Y(y)$ is equal to 0. Clearly, all the points in the shaded area maximize $I_Y(y)$ over Y.

The trouble about this entire debate is that the real facts about what constitutes an optimization problem are lost in a mass of babble. The fact is that in practice, an optimization problem can consist of many constraints (requirements). This means that deciding what constitutes a "constraint" and what

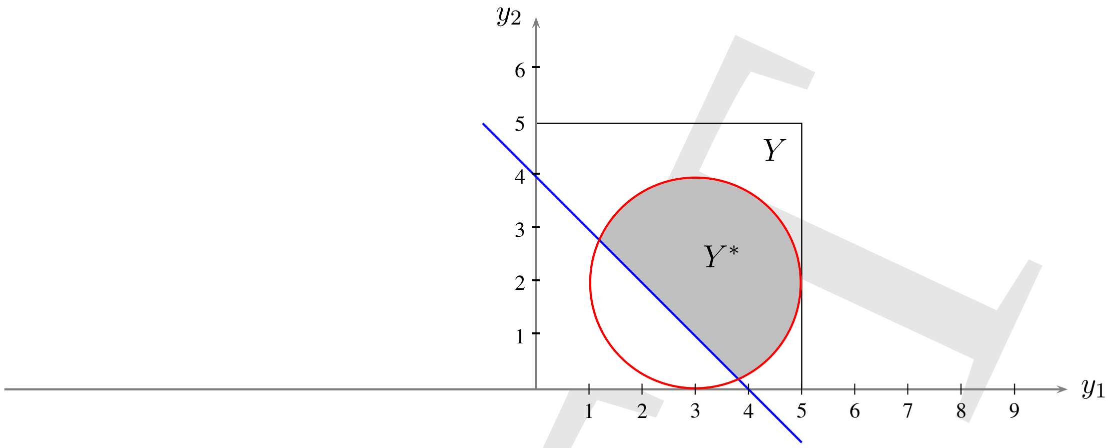


Figure 7.11: Optimizing vs Satisficing

constitutes an "objective function" is often a subjective matter. An optimization model can therefore be regarded a decision-making model where one constraint is elevated to the status of an "objective function". Which means of course that the real question that the modeler/analyst would have to deal with in such situations is not whether optimizing is better than satisficing, or vice versa. Rather, the dilemma that the analyst would have to face is WHAT to optimize and WHAT to satisfice. In other words, when presented with a problem, the analyst would have to determine what should be designated an objective function and what should be designated as constraints.

In short, the distinction between "satisficing" and "optimizing" is at best a distinction between *styles of formulation*. It does not reflect a *substantive* difference between the two forms. The trouble is, however, that, proponents of "satisficing is superior to optimizing" who hang this thesis on what Herbert Alexander Simon (1916 – 2001) called — *bounded rationality,* end trivializing some of the valid methodological and technical issues that he had raised about individuals' and group decision-making under *bounded rationality*.

And there is worse yet . . .

One of the grave difficulties that an optimization-free decision-making paradigm (namely a satisficing paradigm) faces is that it can generate **dominated** decisions, that is decisions that are dominated by (inferior to) other available decisions.

To illustrate, consider the situation depicted in Figure 7.12, where $y' = (3, 4)$ and $y'' = \left(2, 2 + \sqrt{3}\right)$ are two satisfactory values of y. The dashed blue lines are the points on the constraint $y_1 + y_2 = C$ for values of C for which the lines go through y' and y'', namely $C' = 7$ and $C'' = 4 + \sqrt{3}$, respectively.

Since

$$(y_1' - 3)^2 + (y_2' - 2)^2 = (y_1'' - 3)^2 + (y_2'' - 2)^2 = 4 \tag{7.76}$$

these two points are equivalent with respect to the first constraint, namely $(y_1 - 3)^2 + (y_2 - 2)^2 \leq 4$. However, for the second constraint, namely for $y_1 + y_2 \geq 4$, we have

$$y_1' + y_2' = 7 \ > \ y_1'' + y_2'' = 4 + \sqrt{3} = 5.73205 \tag{7.77}$$

Since this constraint is a '$\geq$' constraint, it may represent a situation where the larger the left hand side $y_1 + y_2$ is, the better.

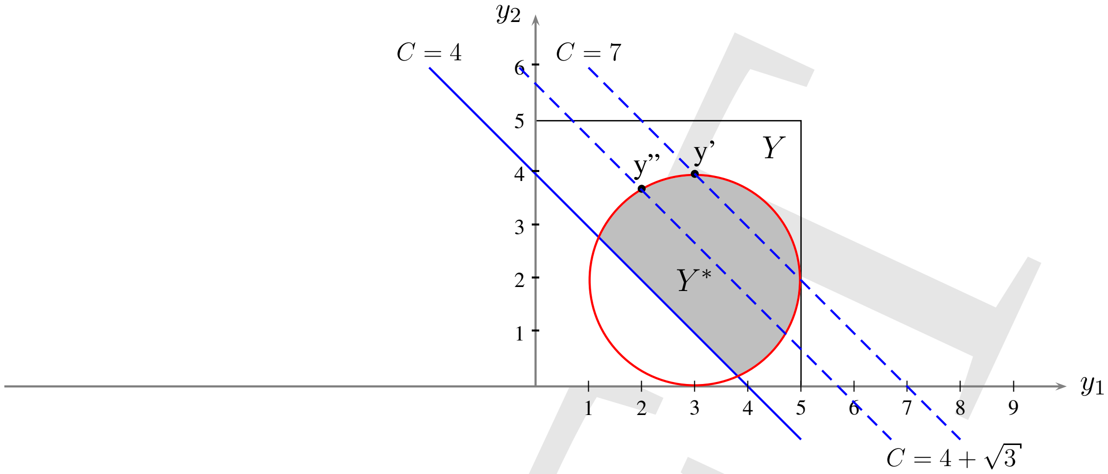


Figure 7.12: Dominated solutions

To illustrate this important point, consider the following case where

$$y_1 + y_2 \quad \equiv \quad \text{Amount (Kg) of Beluga caviar awarded to Moshe.}$$

I can assure the reader that in this case, y' is superior to y''. Yet, the purely “satisficing” model is completely oblivious to this obvious preference. Because, according to the above “satisficing” model, y'' is as “satisfactory” as y'. In fact, insofar as the “satisficing” model is concerned, all the points in the shaded area are equally satisfactory.

It is important to take note that this example does not illustrate an exception to the rule. If anything, it illustrates the rule. Constraints/requirements associated with “satisficing” models are often manifestations of preferences. But the trouble is that as preferences they are not given an appropriate representation in the “satisficing” model. This is due to the binary (0-1) structure of the indicator function I_Y. Consequently, there are no “degrees” of satisfaction, a solution $y \in Y$ is either “satisfactory” or “unsatisfactory”.

Of course, a similar complaint can be made about the optimization model specified by (7.69). Arguably, this model has only **one** objective function. Hence, additional preferences would have to be incorporated as constraints and these, would suffer from the same shortcoming described above.

For the record, I should point out though that *optimization theory* is fully capable of dealing with this issue. Indeed, the question of how to deal with preferences is the main concern in the area of *multiple-objective optimization*. However, as this topic is not within the scope of my discussion in this book, I leave it at that. The concept *Pareto Optimization* that I discuss briefly in the second part of the book is relevant to this issue.

7.5.2 The parameter space $\mathscr{U}$

At this stage I need hardly remind the reader that in the framework of likelihood-free models of uncertainty the formulation of the parameter space $\mathscr{U}$ is an extremely difficult task. For one thing, in such an environment every element of $\mathscr{U}$ counts, because in the absence of a likelihood function, or other such devices, no one element can be given more/less "weight" in the analysis than any other. This is particularly true for situations where the analysis is intended to deal not only with "normal" events, but also with rare events, surprises, catastrophes, shocks and so on (perhaps even Black Swans and "unknown unknowns"!).

I wish I could offer you a general purpose recipe for determining the parameter space of robust optimization models, but I am afraid that I cannot. The only (good) advice I can offer on this matter is to steer well clear of methods offering such recipes.

The construction of the parameter space for robust optimization models is, and in all likelihood will remain, a problem specific task that requires intimate knowledge of the problem under consideration.

In the *Bibliographic Notes* of this chapter I provide references to publications where the reader may find useful hints and advice on how to approach this formidable task. In this section I examine a number of issues related to the construction/formulation of the parameter space that I consider to be central to the discussion in this book.

It is instructive to approach this task by breaking it down into the following two related but distinct sub-tasks:

· Establish what the parameter u represents, that is specify the "real world entities" that u represents.

· Determine the range of feasible/plausible values of u.

In some cases the first sub-task is trivial. For instance, in cases where the "real world" meaning to be assigned to u is unambiguously dictated to us. But, in cases where this is not so, giving meaning to u can be an extremely complex matter. For example, in large complicated socioeconomic models there could be dozens of parameters so that it would be extremely difficult to decide which should be included in u.

For obvious reasons, in most publications dealing with practical problems, the "real world" meaning of u is "given" in the statement of the problem. A word of caution though: beware of publications incorporating "unknown unknowns" in u. More on this in the second part of the book.

I do not have a great deal more to say about the first sub-task except to repeat that it is a problem oriented task. So let us go straight to the second sub-task, namely to the question of:

How do we determine the range of feasible/plausible values of u?

Here, there are two considerations to reckon with. First, $\mathscr{U}$ should give sound expression to the variability of the "real world" counterpart(s) of u. Second, the granularity of $\mathscr{U}$ should be compatible with the requirements of the analysis under consideration.

To illustrate these issues, let us first consider a rather simple, you might say trivial case, where u is a parameter of some elaborate economic model:

$u =$ probability that the price of regular unleaded petrol in Melbourne (Australia) on January 1, 2015, will not be higher than \$2 per litre.

For the purpose of this exercise, assume that the elaborate economic model is 100% accurate — except that the true value of u is subject to severe uncertainty[8].

Obviously, since u is a *probability* of some event, the range of feasible/plausible values of u in this case is $[0, 1]$. Thus, if we want to play it safe, we can simply set $\mathscr{U} = [0, 1]$. And if we know more about

[8] As a reference point, yesterday's (December 12, 2010) price was AUD\$1.279 per litre (regular unleaded).

the global energy market, we may well be able to refine this range somewhat. In some applications it might be necessary/convenient to “discretize" the range $[0, 1]$ into say 21 equally spaced values, namely to set $\mathscr{U} = \{0.00, 0.05, 0.10, 0.15, 0.20, \dots , 0.95, 1.00\}$. In others, the discretization might have to be finer. It all depends ...

Next.

Consider the more difficult case where u is a parameter of some model dealing with the living conditions of kangaroos in the wild:

$u =$ number of kangaroos in the wild on January 1, 2015.

What should the value of $\mathscr{U}$ be in this case?

We can let $\mathscr{U} = \{\underline{u}, \dots , \overline{u}\}$, where $\underline{u}$ and $\overline{u}$ are lower and upper bounds on u, respectively. For instance, how about letting $\underline{u} = 0$ and $\underline{u} = 200$ million?[9]

While it can be argued that these bounds are indeed bounds, they are clearly not “realistic ” (tight) bounds. Thus, if the results of the robustness analysis are affected by these bounds then, it is wise to attempt to determine tighter more “realistic” bounds.

Depending on the scope of the model and its objectives, it might be sufficient to base the value of $\mathscr{U}$ on a small number of *scenarios* regarding the future population of kangaroos in Australia, generated by *population growth models.* This may yield a parameter space such as say $\mathscr{U} = \{15, 30, 60, 90, 120\}$ (in millions).

On the other hand, it might be necessary to use not only a more refined range of values of u, it may also be necessary to expand the definition of u to include other relevant details about kangaroos in Australia. For instance, their age, weight, height, length, culinary habits, and in some cases, even much more personal details such as level of aggression or lack of it, sociability etc.

And talking about scenarios.

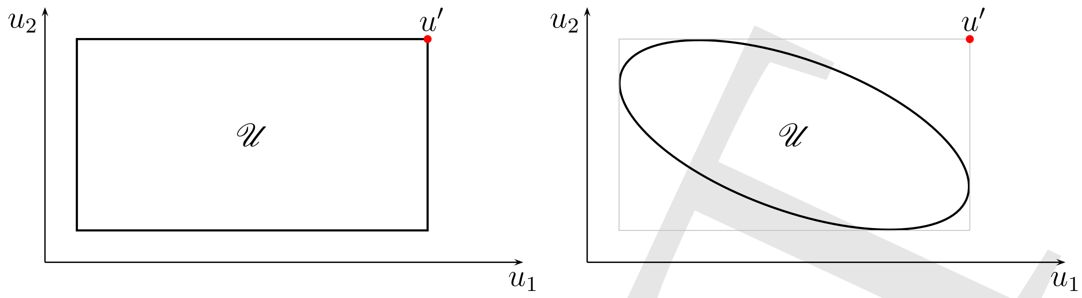
Many of the issues and difficulties associated with the formulation of parameter spaces for robust optimization models are similar to those associated with the use of scenario generation techniques to sample trajectories of complex dynamical systems/processes. But as recent debates and scandals about *climate change* indicate only too clearly, *scenario generation* involving large, complicated, severely uncertain systems is a daunting task. Witness the easy with which extremely “conservative” scenarios can be generated!

In the case of parameter spaces this can happen where u is multi-dimensional and components of u are generated independently of one another. In such situations, a rather crude, but apparently popular, way of filtering-out values of u that are too conservative, is to “anti-correlate” the values of u by properly “shaping” the parameter space.

For example, consider the case where $u = (u_1, u_2) \in \mathbb{R}^2$ and the larger these components are, the worse they are insofar as the constraints and the objective function are concerned. Consider now the parameter space $\mathscr{U}$ shown in Figure 7.13a. Its worst element is obviously u'.

One can argue, and many indeed do, that it is extremely “unlikely” that such a u will ever be realized, therefore such values should not be included in $\mathscr{U}$. This can be done by bounding the value of u_2 depending on the value of u_1 with the view to eliminate values of u that are “too conservative”. This is illustrated in Figure 7.13b.

[9]The present population is about 60 million, give or take a couple of millions.



(a) (b)

Figure 7.13: Parameter spaces

It is not surprising, therefore, that *ellipsoid-shaped* parameter spaces are very popular in robust optimization. It should be pointed out, though, that this kind of subjective "censorship" should be executed with care, especially in situations where robustness is sought not only against "normal" events, but also against "abnormal" events.

And talking about ellipsoids.

The popularity of ellipsoid-shaped parameter spaces is based on other reasons as well, mainly those that fall under the rubric of "mathematical convenience". That is, an important consideration in the formulation of the parameter space is the ease with which the robust optimization problem can be solved. In cases where u is a "continuous" parameter and the robust optimization problem is solved with conventional optimization methods, it is often convenient to "require" $\mathscr{U}$ to be a *convex* set. I shall address this point briefly in the second part of the book.

7.5.3 Auxiliary decision variables

This is a modeling issue that must be elucidated to readers who are not well-versed in optimization theory. So at the risk of boring those readers who are at home with this issue, I briefly address the role of auxiliary decision variables in the formulation of robust optimization models. I dedicate this discussion especially to *info-gap scholars*.

So what is an *auxiliary decision variable,* and how would you recognize it when it is staring at you in a robust optimization model?

To answer this question, consider the following three models. The first is the generic optimization problem under consideration in this discussion. The second is the model expressing the associated family of parametric optimization problems. The third, is the generic robustness model where robustness is sought with respect to both the constraints and the objective function:

Model P:

$$z^* := \max_{x \in X} g(x) \text{ subject to } constraints(x) \tag{7.78}$$

Model $P(u), u \in \mathscr{U}$:

$$z^{\circ}(u) := \max_{x \in X} g(x; u) \text{ subject to } constraints(x; u) \tag{7.79}$$

Model RGC :

$$z^{(rgc)} := \max_{x \in X} \min_{u \in \mathscr{U}_x} g(x; u) \text{ subject to } constraints(x; u), \forall u \in \mathscr{U}_x \tag{7.80}$$

Note that in all three cases, the decision maker (DM) — represented by the $\max$ player — controls the value of the decision variable $x \in X$. In the robustness model, Nature — represented by the $\min$ player — controls the value of the parameter u.

However, there are situations where the desired robustness is such that it is necessary/convenient to incorporate in the robustness model additional variables whose values are controlled by the decision maker. We refer to such variables as AUXILIARY decision variables.

Example

Consider the generic robustness model where robustness is sought only with respect to the constraints, namely

Model RC :

$$z^{(rc)} := \max_{x \in X} g(x) \text{ subject to } constraints(x;u), \forall u \in \mathscr{U}_x \tag{7.81}$$

Now consider a relaxation of this model as follows:

> Instead of stipulating the requirement $constraints(x;u), \forall u \in \mathscr{U}_x$, we consider the "relaxed" requirement $constraints(x;u), \forall u \in V$ where V can be any subset of $\mathscr{U}_x$ such that $size(V) \geq 0.8\, size(\mathscr{U}_x)$, recalling that $size(A)$ denotes the size of set A.

In other words, here x is required to satisfy the constraints only with respect to 80% of the elements of $\mathscr{U}_x$. Formally, the robust model would then be as follows:

$$z' := \max_{\substack{x \in X \\ V \subseteq \mathscr{U}_x}} g(x) \text{ subject to } constraints(x;u), \forall u \in V, size(V) \geq 0.8\, size(\mathscr{U}_x) \tag{7.82}$$

Observe that this is a *Maximin model.*

According to this model, the decision maker must determine the values of two objects, namely $x \in X$ and $V \subseteq \mathscr{U}_x$. The difference between these two objects is that x appears in the formulation of the generic optimization model and the formulation of the parametric optimization model, whereas V does not.

So, although the object V is manifestly a decision variable in the robustness model specified by (7.82), as it does not appear in the generic optimization model nor in the formulation of the parametric optimization model, we designate it as an "auxiliary" decision variable.

This means that to be cognizant of this fact you must understand the structure of the optimization model and the formulation of the parametric optimization model. To illustrate what I mean, suppose that you show the following model to a visitor from Mars, call her Shelly:

$$z' := \max_{x \in X} g(x) \text{ subject to } constraints(x;u), x_1 \subseteq \mathscr{U}_x, \forall u \in x_1, size(x_1) \geq 0.8\, size(\mathscr{U}_x) \tag{7.83}$$

Since in this model the decision maker controls only one object, namely x, and since on Mars the constraints associated with the parametric optimization model are of the form

$$constraints(x;u), x_1 \subseteq \mathscr{U}_x, \forall u \in x_1, size(x_1) \geq 0.8\, size(\mathscr{U}_x) \tag{7.84}$$

Shelly will not recognize x_1 for what it is, an auxiliary decision variable in disguise. To her, x — including x_1 — is a perfectly ordinary decision variable.

The same rule will apply to all of us in this book: the distinction between the ordinary decision variable (x) and an auxiliary decision variable is based on the structure of the generic optimization model and the formulation of the parametric optimization model under consideration. □

Example

Consider this familiar, one might say intuitive, robustness model based on the *Size Criterion*, namely:

$$z' := \max_{\substack{x \in X \\ V \subseteq \mathscr{U}_x}} \ size(V) \ \text{ subject to } \ constraints(x;u), \forall u \in V \tag{7.85}$$

Here V is an auxiliary decision variable.

Example

Consider the local decision model based on the *Radius of Stability model,* namely

$$z' := \max_{\substack{x \in X \\ \rho \geq 0}} \{\rho : constraints(x;u), \forall u \in B(\rho, \tilde{u})\} \tag{7.86}$$

where $B(\rho, \tilde{u})$ denotes a ball of radius ρ around $\tilde{u}$. Note that here ρ is a typical auxiliary decision variable. □

So as you can see, auxiliary decision variables have a crucial role in the formulation of robustness models. However, given my *info-gap experience* I think it imperative to make the following statement:

Declaration

The term "auxiliary" in the phrase "auxiliary variable" in this book does not in any way connote that this variable is an inferior, second class object. Not at all.

To the contrary.

In this book the vital role that these variables play in the formulation of robust decision-making models is recognized through this motto:

DON'T LEAVE HOME WITHOUT THEM!

The title "auxiliary" merely means that the auxiliary variables are not part of the generic optimization model and the associated parametric model. Therefore, their inclusion in the associated robust optimization model in fact amplifies the importance of these variables' role in the modeling and analysis of robust decision-making problems.

The Author

As indicated in the next subsection, there could be other "auxiliary" objects in a robust optimization model.

7.5.4 Levies, subsides and allowances

In the same manner that federal and local governments use various financial instruments such as levies and allowances to "guide" various sectors of the economy when the need arises, analysts can "control" robustness models by means of either tightening or easing/relaxing the models' constraints to modify their behavior. The next two examples illustrate this point.

Example

Apparently, some analysts have a much greater preference for models of robustness that are based on *regrets* rather than for models based on *payoffs*, in which case the robustness model is a *Minimax regret* model rather than a Maximin payoff model.

However, the rub here is that to use robustness models that are based on regrets one has to convert the payoffs into regrets.

This is easily done in textbook examples.

But there are many cases where computing the regrets or dealing with them on-the-fly can be extremely difficult/demanding. In such cases one is tempted to "adjust" a bit the definition of *regret* by benchmarking the performance of decisions not against the "best possible" but against some ad-hoc reference level.

Recall that in the context of the classic format

$$\max_{x\in X}\ \min_{u\in \mathscr{U}}\ g(x;u) \tag{7.87}$$

the *regrets* are defined as follows:

$$r(x;u) := \max_{y\in X}\ g(y;u) - g(x;u)\ ,\ x\in X, u\in\mathscr{U} \tag{7.88}$$

Hence the *Minimax regret* model is as follows:

$$z' := \min_{x\in X}\ \max_{u\in\mathscr{U}}\ r(x;u) \tag{7.89}$$

$$= \min_{x\in X}\ \max_{u\in\mathscr{U}} \left\{ \max_{y\in X}\ g(y;u) - g(x;u) \right\} \tag{7.90}$$

$$= \min_{x\in X}\ \max_{u\in\mathscr{U}} \{h(u) - g(x;u)\}\ ,\ h(u) := \max_{y\in X}\ g(y;u)\ ,\ u\in\mathscr{U} \tag{7.91}$$

So the idea is to replace $h(u)$ with some $\tilde{h}(u)$ that would be much easier to compute/determine. The resulting Minimax "adjusted regret" model will then be as follows:

$$\tilde{z} := \min_{x\in X}\ \max_{u\in\mathscr{U}} \left\{ \tilde{h}(u) - g(x;u) \right\} \tag{7.92}$$

In fact, $\tilde{h}(u)$ can represent the "nominal" — rather than optimal — payoff when the system is in state u. And from a methodological point of view, this nominal value can depend on x, in which case we shall have $\tilde{h}(x;u)$ rather than $\tilde{h}(u)$ in (7.91).

Example

This example illustrates the *Globalized Robustness* model proposed by Ben-Tal et al. (2006, 2009) It also illustrates the distinction between "normal" and "abnormal" values of u. This model can be used to determine robustness with regard to constraints as well as robustness with regard to the objective function. I shall illustrate the former, assuming that the model consists of one robustness constraint of the form:

$$r^* \leq r(x;u)\ ,\ \forall u\in\mathscr{U} \tag{7.93}$$

where r is a real valued function on $X \times \mathscr{U}$ and r^* is a given numeric scalar.

In practice such a constraint can be too demanding in that none of the decisions in X satisfices it.

Suppose then that we do as follows: we partition the parameter space $\mathscr{U}$ into two sets, one representing the "normal" range of values of u, call it $\mathscr{U}^+$, and the other representing the "abnormal" values of u, call it $\mathscr{U}^-$.

Next we relax the above robustness constraints as follows:

$$r^* \leq r(x;u) \ , \ \forall u \in \mathscr{U}^+ \tag{7.94}$$

$$r^* \leq r(x;u) + \beta \cdot dist(u, \mathscr{U}^+) \ , \ \forall u \in \mathscr{U}^- \tag{7.95}$$

where $\beta \geq 0$ is a given numeric scalar and $dist(u, \mathscr{U}^+)$ denotes the distance from $u \in \mathscr{U}$ to $\mathscr{U}^+$. This measure of distance is formulated so that $dist(u, \mathscr{U}^+) = 0, \forall u \in \mathscr{U}^+$, and $dist(u, \mathscr{U}^+) > 0$, $\forall u \in \mathscr{U}^+$.

Observe that for $\beta = 0$, relaxing the constraint in this fashion has the effect of reducing it to the initial robustness constraint stipulated by (7.93). Hence, $\beta = 0$ any decision that satisfices this relaxed constraints, also satisfies the constraint $r^* \leq r(x;u)$ on the normal range $\mathscr{U}^+$.

Of particular interest is, of course, the case where $\mathscr{U}^+$ is a *singleton,* namely the case where $\mathscr{U}^+ = \{\tilde{u}\}$ for some $\tilde{u} \in \mathscr{U}$. Here the robustness constraints are as follows:

$$r^* \leq r(x;\tilde{u}) \tag{7.96}$$

$$r^* \leq r(x;u) + \beta \cdot dist(u, \tilde{u}) \ , \ \forall u \in \mathscr{U} \tag{7.97}$$

where in this context $dist(u, \tilde{u})$ denotes the distance from u to $\tilde{u}$.

This being so, we can integrate this relaxed global constraint in a *Maximin model* as follows:

Problem GL:

$$\max_{x \in X} \min_{u \in \mathscr{U}} \{g(q;u) : r^* \leq r(q;u) + \beta \cdot dist(u, \mathscr{U}^+), \forall u \in \mathscr{U}\} \tag{7.98}$$

In practice, the "normal range" $\mathscr{U}^+$ can be parameterized by its "size" which can be varied by means of a *sensitivity analysis*.

Note that in this framework we can distinguish between three sets of forbidden (unacceptable) $(u, r(x,u))$ values for decision $x \in X$:

$$\mathcal{F}_{strict}(x) := \{(u,y) : u \in \mathscr{U}, y < r^*\} \tag{7.99}$$

$$\mathcal{F}_{normal}(x) := \{(u,y) : u \in \mathscr{U}^+, y < r^*\} \tag{7.100}$$

$$\mathcal{F}_{relaxed}(x) := \{(u,y) : u \in \mathscr{U}, y < r^* - \beta \cdot dist(u, \mathscr{U}^+)\} \tag{7.101}$$

The following two examples illustrate this idea and they highlight the difference between *local* and *global* robustness. For simplicity, they are presented graphically rather than algebraically. The first features a simple *Radius of Stability* analysis, and the second a global robustness analysis based on the relaxation of robustness constraints as described above.

Example

Consider a case with three decisions namely, $X = \{x', x'', x''')\}$, where the uncertainty space is the real line, that is, $\mathscr{U} = (-\infty, \infty)$. Also, assume that $\tilde{u} = 0$.

Let us consider what would be the consequences of employing a *Radius of Stability* model of the *info-gap type* in this case, namely a model where the stability regions are determined by a performance requirement $r^* \leq r(x;u)$. The critical performance level r^* is equal to 10, and the performance functions are shown in Figure 7.14. Since the uncertainty space is unbounded, only a small section of it, in the neighborhood of the estimate, is shown. Assume that the performance functions continue their trends, as shown in the figure, in both directions.

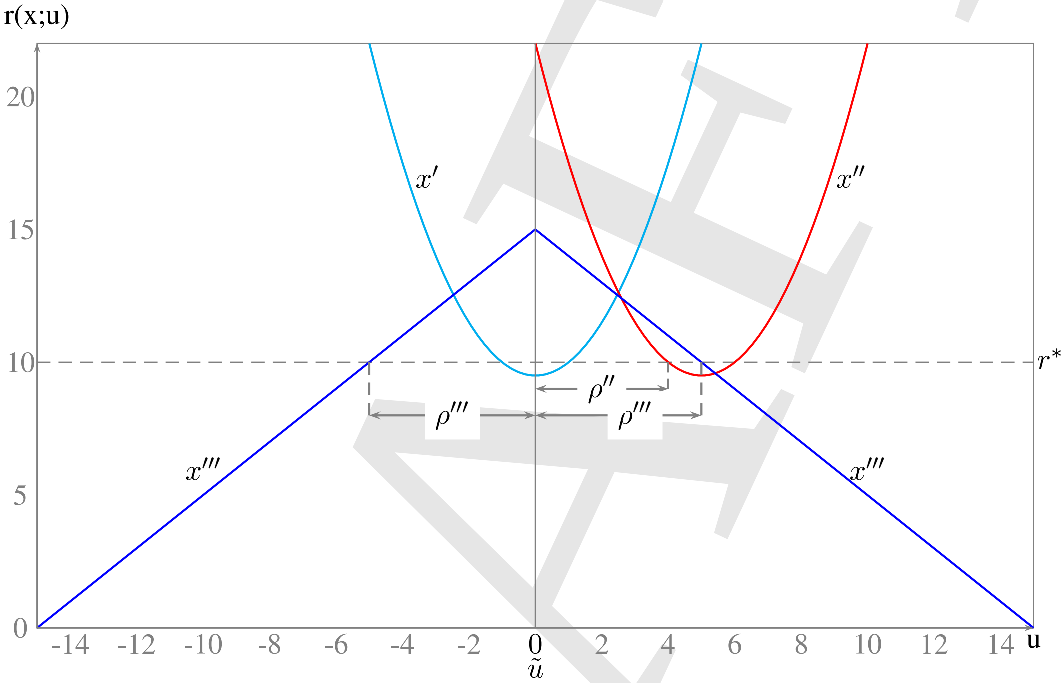


Figure 7.14: Radius of stability analysis

For simplicity, suppose that the balls (neighborhoods) used are of the form $B(\rho, \tilde{u}) = [u : |u - \tilde{u}| \leq \rho]$, in which case we have $B(\rho, 0) = [-\rho, \rho], \rho \geq 0$.

Since decision x' violates the performance requirement at the estimate $\tilde{u} = 0$, it follows that its *Radius of Stability* is equal to zero: $\rho' = 0$. By inspection, the radii of stability of x'' and x''' are $\rho'' = 4$ and $\rho''' = 5$, respectively.

Hence, according to the *Radius of Stability* approach, the most robust decision at $u = 0$ is x'''. However, this decision is decidedly not robust over its uncertainty space $\mathscr{U} = (-\infty, \infty)$.

In contrast, in spite of the fact that the *Radius of Stability* approach determines that decision x'' is not as robust as x''' at $u = 0$, this decision is far more robust than x''' globally on the uncertainty space $(-\infty, \infty)$.

But this finding is hardly surprising because, as I pointed out on a number of occasions, *Radius of Stability* models are not designed to handle global robustness.

Example

Consider the instance of Problem GL that is specified as follows:

$$X = \{x', x'', x''', x''''\}\}$$
$$\mathscr{U} = (-\infty, \infty)$$
$$\tilde{u} = 0$$
$$r^* = 10$$
$$\mathscr{U}^+ = [-2, 2]$$
$$\beta = 1$$
$$dist(s, \mathscr{U}^+) = \min_{u' \in \mathscr{U}^+} |u - u'| = \begin{cases} 0 & , \ u \in [-2, 2] \\ |u| - 2 & , \ u \notin [-2, 2] \end{cases}$$

The performance functions are shown in Figure 7.15 and the cross hatched area represents the relaxed "forbidden" region:

$$\mathcal{F}_{strict}(x) = \{(u, y) : u \in (-\infty, \infty), y < 10)\} \tag{7.102}$$
$$\mathcal{F}_{normal}(x) = \{(u, y) : u \in [-2, 2], y < 10\} \tag{7.103}$$
$$\mathcal{F}_{relaxed}(x) = \{(u, y) : u \in (-\infty, \infty), y < 10 - dist(u, \mathscr{U}^+)\} \tag{7.104}$$

Thus, the *Radius of Stability model* stipulating these performance functions determines that the most robust, hence, optimal decision is x'''.

As for the relaxed problem, by inspection, decisions x''' and x'''' are inadmissible: their graphs intrude into the forbidden region $\mathcal{F}_{relaxed}$.

To decide which decision is optimal for the relaxed problem, we consider the worst values of $g(x'; u)$ and $g(x''; u)$ over the uncertainty space $(-\infty, \infty)$, and select the best worst case. This is shown in Figure 7.16.

The worst case of $g(x'; u)$ is attained at $u = -7$ for which we have $g(x'; -7) = 5$ and the worst case of $g(x''; u)$ is attained at at $u = 0$ for which we have $g(x'; 0) = 0$. Hence, the best worst case is generated by x' and it is therefore the optimal decision in this case.

Observe that the worst case of $g(x''''; u)$ is attained at $u = 10$ and is equal to $g(x''''; 10) = 6$, which is better than the worst case of the optimal decision x'. However, as noted above, decision x'''' is inadmissible because it violates the "relaxed" global constraints (7.94)-(7.95). □

We can go a step further and regard the globalized robust optimization model

Model GL

$$\max_{x \in X} \min_{u \in \mathscr{U}} \{g(x; u) : r^* \leq r(x; u) + \beta \cdot dist(u, \mathscr{U}^+), \forall u \in \mathscr{U}\} \tag{7.105}$$

an instance of the more general model

$$\max_{x \in X} \min_{u \in \mathscr{U}} \{g(x; u) : r^* \leq r(x; u) + \Delta(x; u), \forall u \in \mathscr{U}\} \tag{7.106}$$

where $\Delta(x; u)$ represents the *allowance/levy* applicable to "abnormal" values of u. Take note that "abnormal" values of u can be either "auspicious" or "inauspicious", which in turn calls for positive or

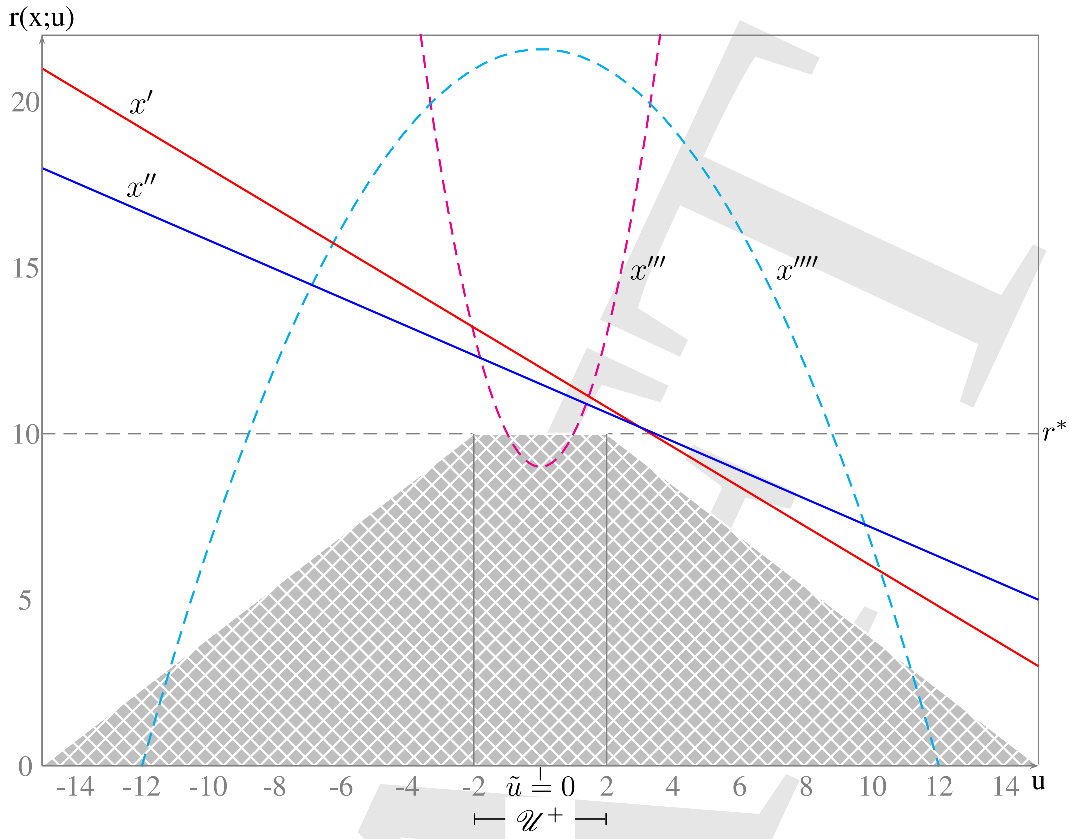


Figure 7.15: Forbidden region $\mathcal{F}_{relaxed}(x)$

negative values of $\Delta(x;u)$, respectively.

Based on the numerous conversations that I had with Fred[10] over the years on this matter, I believe that his view is far more radical than mine. Fred would argue that *Model GL* should be seen as an instance of the model:

Model Fred

$$\max_{x\in X}\ \min_{u\in\mathscr{U}} \{g(x;u) : r^* \leq \tilde{r}(x;u), \forall u \in \mathscr{U}\} \tag{7.107}$$

where the difference between the performance levels $r(x;u)$ in (7.105) and $\tilde{r}(x;u)$ in (7.107) can be taken care of by applying what is known universally as the ART OF MODELING.

In other words, Fred's point is that, in practice, it might be necessary/advisable to "massage" the "original" constraint "$r^* \leq r(x;u)$" so as to be able to take account of "abnormal" values of $u \in \mathscr{U}$.

[10]Internationally renowned expert on robust decision in the face of severe uncertainty

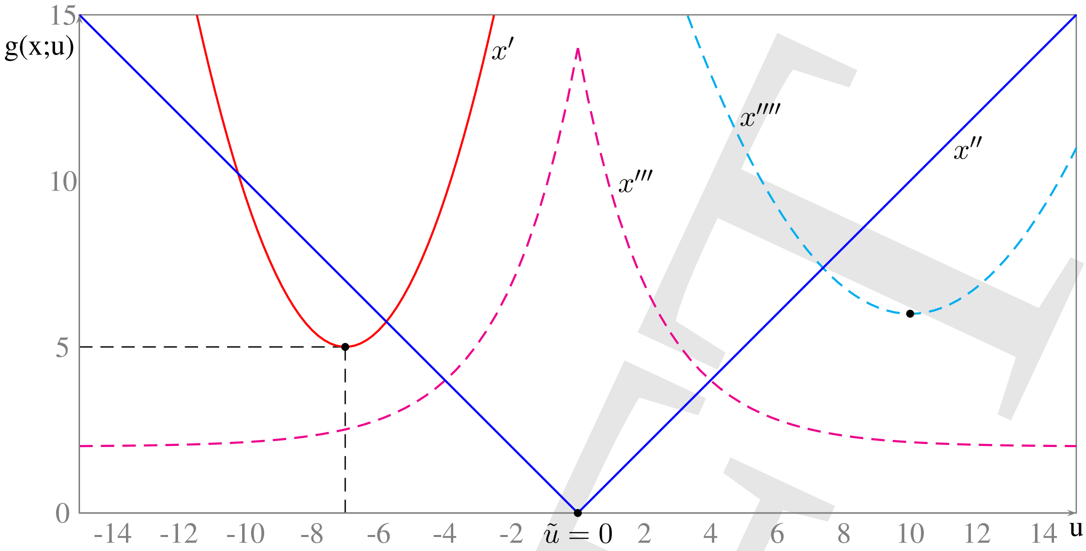


Figure 7.16: Objective functions and best worst case

But, as Fred would tell you, I take a sharply different view on this matter. I regard models such as:

Model RG

$$\max_{x\in X}\ \min_{u\in\mathscr{U}_x}\ g(x;u)\ \text{ subject to }\ constraints(x) \tag{7.108}$$

Model RC

$$\max_{x\in X}\ g(x)\ \text{ subject to }\ constraints(x;u), \forall u\in\mathscr{U}_x \tag{7.109}$$

Model RGC

$$\max_{x\in X}\ \min_{u\in\mathscr{U}_x}\ g(x;u)\ \text{ subject to }\ constraints(x;u), \forall u\in\mathscr{U} \tag{7.110}$$

as GENERIC MODELS, so my position is that from the start, the onus is on the analyst to construct the objective function, the constraints, the decision space, and the parameter space, in a manner that would give faithful representation to the problem considered.

In situations where $\mathscr{U}_x$ represent the possible/plausible values of a parameter u associated with decision $x\in X$ whose true value is subject to *severe* uncertainty, the objective function g and the constraints in $constraints(x;u)$ must be able to deal with the fact that some of the values of $u\in\mathscr{U}$ would be "abnormal" — as well as with other relevant factors affecting the analysis.

Viewed from this perspective, *Model GL* can be regarded a good illustration of the type of considerations that may affect the formulation of constraints, which under severe uncertainty, seek to deal with "abnormal" values of u.

Both Fred and I agree that taking into account that some of the values of $u\in\mathscr{U}$ are "abnormal" — as well as properly handling other relevant factors affecting the analysis — in the end boil down to . . . one's competence (perhaps even integrity) as a modeler. Fred thus makes it a point to quote Ben-Tal et al.'s (2009, p. 926) claim that models that restrict the analysis only to the "normal" range $\mathscr{U}_+$ '. . . represent a somewhat "irresponsible" decision-maker. I shall enlarge on this assertion in the second part of the book.

Having outlined the working of robust optimization models let us now examine how SEVERE UNCERTAINTY is perceived from this perspective.

7.6 Severe uncertainty revisited

You will recall that the severity of the uncertainty considered in this book is taken to be manifested in the following three properties of the uncertainty model:

- The parameter (uncertainty) *space* $\mathscr{U}$ can be *vast.*

- The *estimate* we have of the true value of u, if we have one, is a *poor* indication of the true value of u which may well turn out to be *substantially wrong.* It is often taken to be a "guess", even a "wild guess".

- The quantification of the uncertainty is *likelihood-free.*

My objective in this section is to examine how these characteristics are reckoned with in the construction of the robustness models discussed in this chapter. As we shall see, each characteristic on it own is not necessarily all that troublesome. The trouble arises, and this is what makes *severe uncertainty* so challenging, is when all three characteristics "congregate under one roof".

So, let us begin by recalling yet again that the family of parametric optimization problems under consideration is as follows:

$$
\begin{aligned}
&\text{Model } P(u), u \in \mathscr{U}:\\
&z^{\circ}(u) := \max_{x\in X}\ g(x;u) \text{ subject to } \ constraints(x;u)
\end{aligned}
\tag{7.111}
$$

The basic assumption underlying the robust counterparts of this parametric model — discussed in this chapter — is that both the objective function and the constraints are "sufficiently" sensitive (whatever that means) to variations in the value of u over $\mathscr{U}$. Of course, there can be specific instances where this is not so, but in this discussion I do not assume this to be the case.

So, in evaluating the impact of severe uncertainty on a robustness model I shall take into consideration the fact that the objective function and the constraints stipulated by the model are "sufficiently" sensitive to the value of u.

7.6.1 Size of the uncertainty space

It is important to realize that the "size" (e.g. cardinality) of the uncertainty space *per se* need not always be the determining factor in the robustness analysis. Often, the *variability* of u over this space and . . . the sensitivity of the model (objective function and constraints) to this variability overshadow the impact of the "size" of the uncertainty space on the analysis.

For example, consider the two uncertainty spaces

$$
\begin{aligned}
\mathscr{U}' &= \{0, .1, .2, .3, .4, .5, .6, .7, .8, .9, 1\}\\
\mathscr{U}'' &= \{0, .2, .4, .6, .8, 1, 1.2, 1.4, 1.6\}
\end{aligned}
$$

associated with the same optimization model.

Although $\mathscr{U}'$ is larger in size (cardinality) than $\mathscr{U}''$, its "range" is smaller than the range of the elements of $\mathscr{U}''$.

All the same, methodologically speaking, the sheer size of the uncertainty space can be a major factor in the analysis. In particular, "all other things being equal", the larger the uncertainty space, the more difficult it is to justify the use of a local robustness analysis as an approximation for a global robustness analysis.

This raises the issue of **unbounded** uncertainty spaces.

In the context of probabilistic uncertainty, it is often assumed that the uncertainty space is unbounded as this enables employing probabilistic models (e.g. normal distribution) that are "convenient" for treatment and/or yield good approximations.

But in the context of decision-making under severe uncertainty, unbounded uncertainty spaces can be extremely problematic.

For one thing, often (but certainly not always) the $constraints(x; u), \forall u \in \mathscr{U}$ requirement which characterizes constrained robust optimization models can be detrimental if $\mathscr{U}$ is unbounded. That is, in such cases the unboundedness of $\mathscr{U}$ may imply that the robustness problem has no feasible solution, let alone an optimal solution.

On the brighter side, though, I have yet to come across a real-world problem where an unbounded uncertainty space is truly called for. I shall discuss this point in detail in the second part of the book. For now I need only point out that I am raising the issue of unbounded uncertainty spaces in this context to set the stage for my critique of *info-gap decision theory* in the second part of the book. For, the claim in the *info-gap literature* is that the most commonly encountered *info-gap models* have unbounded uncertainty spaces.

7.6.2 Point Estimate

As I pointed out already, when it comes to the *point estimate,* the issue here is not whether a point estimate of the true value of u exists. For, if it does not exist, it can be easily manufactured. The real issue here is that under severe uncertainty, the estimate is expected to be poor, questionable, doubtful and the list goes on.

So, the question is why bother with an estimate at all if it is so deficient?

Well, as the reader must have noticed, this seems to be the approach taken by *robust optimization,* for without exception, none of the robustness models discussed in this chapter incorporates an estimate of the true value of u.

But on the other hand, one may well question whether this is entirely justified.

For, it can be argued that by omitting the estimate $\tilde{u}$ from the robustness model, a vital piece of information about the parameter u is being overlooked. Indeed, it can even be argued that this may well amount to NEGLIGENCE. Because, given the severity of the uncertainty, any piece of information about u should be taken into account even if the information is deficient.

I shall address this argument in the next subsection. Here I merely want to note, and this against Fred's advice, that in principle I concur with this argument — on the proviso that the **quality of the estimate is fully reckoned with in the formulation.**

My experience over the years has been that even in cases where no estimate seems to be stipulated explicitly in the robustness model, the impact of an implicit estimate is nevertheless felt in the robustness model. My point is that the specification of the uncertainty space $\mathscr{U}$ is often based (in one way or another) on an estimate. To illustrate, consider again the case where:

$u =$ number of kangaroos in the wild in Australia on January 1, 2020.

How would we determine the uncertainty space $\mathscr{U}$ for this parameter?

I do not know how you would go about it but, I would do a Google Search based on the key words "population of kangaroos in Australia". Such a search generates several links to sites providing conflicting estimates of the current population of kangaroos in Australia.

As for the quality of the estimates, I find this interesting:

> There are 48 species of macropods (kangaroos) in Australia. Of these only 4 can be commercially harvested. In addition 2 species of wallaby are harvested in Tasmania.
>
> Over 99% of the commercial kangaroo harvest occurs in the arid grazing rangelands. The populations of kangaroos in these areas are estimated every year in each State by well developed aerial survey techniques. It must be understood that these are sparsely timbered, if at all, savannah type ecosystems. Hence it is possible to fly over them and count the large animals such as kangaroos seen. Using either low flying fixed wing aircraft or helicopters, flying at heights of 2-300 meters the National Parks Authorities count the numbers of kangaroos seen over fixed transects. Thirty years of such monitoring have allowed them to develop sophisticated and accurate techniques of extrapolating out to total population numbers (Grigg and Pople 2001). Kangaroos are one of only a very few species (including humans) who have an annual census of their populations.
>
> Website of the Kangaroo Industry Association of Australia
> http://www.kangaroo-industry.asn.au/morinfo/BACKGR1.HTM
> Read on December 20, 2010

So, the situation I find myself in is as follows: I have in hand a number of conflicting estimates, but according to an authoritative source kangaroo populations estimates ought to be considered sound. How then am I going to decide on an estimate? Because my knowledge about kangaroos is minimal, and given the conflicting estimates found by my quick web search search, I conclude that currently there are about 60 millions kangaroos in Australia. Namely, my rough estimate of the true value of u is ... $\tilde{u} = 60$ million.

Now.

Regardless of how I shall ultimately work out what value to assign to the uncertainty space $\mathscr{U}$, this will very much depend on the value of $\tilde{u}$. For, suppose that I decide to let

$$\mathscr{U} = \{40, \ldots, 200\}(\text{million})$$

More than anything else, this choice is based on my very rough estimate of the 2010 population. The rather high upper bound is based on the following (color added)[11]:

> Methane from cattle and sheep is 11% of Australia's total GHG. Kangaroos, on the other hand, produce negligible amounts of methane. Farmers have few options to reduce livestock GHG emissions. An article in New York Times of 13 July 2010 described how researchers are trying to make cattle digestion more like kangaroos. It is not working however and the alternative of making greater use of kangaroos themselves to produce low emission meat is one of the objectives of the Cooperative.
>
> On the rangelands where kangaroo harvesting currently occurs, reducing cattle and sheep populations and increasing the kangaroo population to 175 million would produce the same amount

[11] GHG = greenhouse gas emissions

> of meat, and lower Australia's GHG by 16 megatonnes, or 3% of Australia's emissions. The potential carbon savings and biodiversity benefits could also be sold on in the carbon markets — voluntary or compliance, or reduce penalties when a price is placed on carbon emissions. See full paper in Conservation Letters and at download paper for details of the analysis.
>
> Using kangaroos adaptations to produce low-emission meat
> Website of Australian Wildlife Services
> http://www.awt.com.au/2010/07/15/using-kangaroos-adaptations-to-produce-low-emission-meat/
> Read on December 20, 2010

In short, although the estimate of u may not be explicitly stipulated in robustness models, one can argue that the value of $\mathscr{U}$ generally depends on some kind of an estimate of u. That is, my argument is that much as the robustness models discussed in this chapter do not incorporate an estimate as an explicit element, an estimate of the true value of u does definitely have a deeply felt impact on the value of $\mathscr{U}$.

7.6.3 Likelihood-free models of uncertainty

So, it is estimated, very roughly, that the current population of kangaroos in Australia is about 60 million and that a significant reduction in GHG can be achieved by increasing this figure to 175 million and reducing the population of cattle and sheep. Based on these and other relevant pieces of information, let us agree that for the purpose of this discussion we can set:

$$u = \text{ number of kangaroos (in million) in Australia on January 1, 2020.}$$
$$\mathscr{U} = \{40, \dots, 200\} = \text{ very rough estimate of the set of possible/plausible values of } u.$$
$$\tilde{u} = 60 = \text{ very rough estimate of the true value of } u.$$

Furthermore, we FORMALLY assume that

> The uncertainty in the true value of u is likelihood-free.

So if we are confronted by Rick, the local Bayesian, and asked:

> What is the likelihood that the population of kangaroos in Australia at the beginning of 2020 will be greater than 100 (million)?

we reply:

> Sorry, mate! This is a likelihood-free model of uncertainty!!!

And if Rick persists:

> So, is it more likely to be around the 50 million mark or the 80 million mark?

we reply:

> Sorry, mate! This is a likelihood-free model of uncertainty!!!

Of course, we may well have our own "beliefs" about the future population numbers of kangaroos in Australia, but the point is that the robustness models presented in this chapter do not provide any mechanism to incorporate such "beliefs".

For the same reason, we cannot associate any notion of likelihood or "belief" with the results generated by these robustness models.

And the implication of all this for the status of the estimate in the robustness model is crystal clear. Whatever impact the estimate has on the value of the uncertainty space $\mathscr{U}$ — and it should have significant impact — this cannot translate into assigning it special status in the uncertainty space $\mathscr{U}$ vis-a-vis other elements in this space. The simple fact is that in a likelihood-free model of uncertainty the uncertainty space $\mathscr{U}$ ... is likelihood-free. So, one cannot plead likelihood and "beliefs" in support of the estimate. The estimate is and remains indistinguishable from the other elements of $\mathscr{U}$. This fact has immediate consequences for the methods aimed at determining robustness to sever uncertainty.

More on this in the second part of the book.

7.7 Algorithms

To set the scene, I repeat the statement I made above regarding solution methods for generic optimization problems represented by *Model P*:

> Not unlike the experts who wield them, optimization methods are highly specialized. Specialized in the sense that they tend to be very particular about the problems that "they admit for treatment". This means that one cannot take it for granted that just because a certain optimization problem can be solved by say, Method A, a "slightly" modified version of the problem would also be amenable to Method A.

As we have seen, the robust counterparts of the parametric model *Model P($\mathscr{U}$)* are "complications" of *Model P*, hence as a rule, robust counterparts of an optimization problem are more difficult to solve than the optimization problem giving rise to them.

So the bad news is that not only are there no general purpose algorithms for the solution of robust optimization problems, there are also no general purpose algorithms for the solution of robust optimization problems whose generic optimization problems can be easily solved by existing methods.

In the bibliographic notes to this chapter I provide references to publications discussing the algorithmic aspects of robust optimization. In the discussion that follows I explain very briefly and I illustrate the difficulties that render robust optimization problems so hard to solve. To this end I focus on a robust optimization model where robustness is sought only with respect to the constraints, namely:

Model RC

$$z^{(rc)} := \max_{x\in X} g(x) \ \text{ subject to } \ constraints(x;u), \forall u \in \mathscr{U}_x \tag{7.112}$$

To get rid of the complication caused by the $\forall u \in \mathscr{U}_x$ clause, let $constraints(x;\mathscr{U}_x)$ denote the SET of all the constraints imposed on decision $x \in X$ by the elements of $\mathscr{U}_x$. We can then re-write this model as follows:

Model RC

$$z^{(rc)} := \max_{x\in X} g(x) \ \text{ subject to } \ constraints(x;\mathscr{U}_x) \tag{7.113}$$

This looks like an ordinary optimization model representing a problem requiring the maximization of the objective function q subject to a set of constraints on the decision variable x. So why should such problems turn out to be more difficult to solve than ordinary optimization problems?

To explain this technical point, consider again the generic optimization model that provides the basis for our discussion in this chapter, namely:

Model P

$$z^* := \max_{x\in X} g(x) \ \text{ subject to } \ constraints(x) \tag{7.114}$$

Observe that in this context $constraints(x)$ denotes a LIST of constraints imposed on decision $x \in X$. In the context of "conventional" optimization, this list is FINITE, namely the list consist of a finite number of elements.

The implication is then that as "ordinary" constrained optimization problems are optimization problems consisting of a FINITE NUMBER OF CONSTRAINTS, algorithms for "ordinary" optimization problems are designed to deal with a FINITE NUMBER OF CONSTRAINTS.

So, the complication arising from the shift from the realm of *Model P* to that of *Model RC* is in the **number of constraints** imposed on the decision variable. And here we can distinguish between two cases. The case where for each $x \in X$ set $\mathscr{U}_x$ consists of FINITELY MANY elements and the case where for some $x \in X$ set $\mathscr{U}_x$ consists of INFINITELY MANY elements.

7.7.1 A finite uncertainty space

If for each $x \in X$, the set $\mathscr{U}_x$ consists of FINITELY MANY elements, then *Model P* and *Model RC* could be of the same type, except that the latter might consist of a considerably larger number of constraints. This means that the robust optimization problem would have many more constraints than the generic optimization problem. This may, or may not make the former more difficult than the latter.

There are two kinds of possible complications in this case. The first has to do with the sheer number of constraints. In cases where the solution grows more difficult as the number of constraints increases, the robust counterpart problem can be even more difficult to solve than the generic optimization problem that gave rise to it.

The second has to do with the fact that even when the number of robustness constraints is finite, the robust counterpart may not be of the same "type" as the generic optimization problem. Consequently a method capable of solving the latter would not not necessarily be capable of solving the former.

The following two examples illustrate this point.

Example

Consider the case where *Model P* is a standard *linear programming model,* say

Model P

$$z^* := \max_{x\in X} c^T x \ \text{ subject to } \ Ax \leq b, x \geq 0 \tag{7.115}$$

where $X = \mathbb{R}^{1\times n}$, $c \in \mathbb{R}^{1\times n}$, $b \in \mathbb{R}^{1\times n}$, and $A \in \mathbb{R}^{m\times n}$.

Observe that this model has m linear constraints (other than the n non-negative constraints).

Now suppose that $\mathscr{U} = \mathscr{U}_x, \forall x \in X$, is a finite set of values of A's and b's, say $\mathscr{U} = \{u^{(1)}, \ldots, u^{(k)}\}$, where $u^{(j)} = (A^{(j)}, b^{(j)})$, $j = 1, 2, \ldots, k$.

In this case *Model RC* will be as follows:

Model RC

$$z^{(rc)} := \max_{x \in X} c^T x \text{ subject to } A^{(j)}x \leq b^{(j)}, j = 1, 2, \dots, k;\ x \geq 0 \tag{7.116}$$

$$= \max_{x \in X} c^T x \text{ subject to } \overline{A}x \leq \overline{b}, x \geq 0 \tag{7.117}$$

where

$$\overline{A} = \begin{pmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(k)} \end{pmatrix} ;\ \overline{b} = \begin{pmatrix} b^{(1)} \\ b^{(2)} \\ \vdots \\ b^{(k)} \end{pmatrix} \tag{7.118}$$

This is a standard linear programming models with $m \times k$ functional constraints. Thus, any instance of this model can be solved by a standard LP algorithm, provided that n, m and k are not too large.

Example

Consider the case where *Model P* represents the conventional *0-1 knapsack problem:*

Model P

$$z^* := \max_{x_1,\dots,x_n} \sum_{j=1}^{n} v_j x_j \text{ s.t. } \sum_{j=1}^{n} w_j x_j \leq W\ ,\ x_j \in \{0,1\}, j = 1, \dots, n \tag{7.119}$$

where the coefficients v_j, w_j and W are given positive integers.

Now suppose that $\mathscr{U} = \mathscr{U}_x, \forall x \in X$, is a finite set of values of w and W, say $\mathscr{U} = \{u^{(1)}, \dots, u^{(k)}\}$, where $u^{(i)} = (w^{(i)}, W^{(i)})$. Then the robust counterpart of this model is as follows:

Model RC

$$z^{(rc)} := \max_{x_1,\dots,x_n} \sum_{j=1}^{n} v_j x_j \text{ s.t. } \sum_{j=1}^{n} w_j^{(i)} x_j \leq W^{(i)}\ , i = 1, \dots, k\ ;\ \ x_j \in \{0,1\}, j = 1, \dots, n \tag{7.120}$$

Note that this is not a standard 0-1 knapsack model. It is a *standard 0-1 multiple constraint knapsack model.*

So, if for years, you managed to solve your standard 0-1 knapsack problems using your favorite dynamic programming (DP) algorithm for the standard 0-1 knapsack problem, you may have to consider a different algorithm for the solution of the robust counterpart of this generic optimization problem. What you would need for this problem is an algorithm that is capable of solving the *0-1 multiple constraint knapsack problem.*

But before you rush out to look for one, ... consider the next example.

Example

Suppose that in the preceding example, the parameter space $\mathscr{U}$ has the following mathematically convenient form:

$$\mathscr{U} = \mathscr{U}^{(1)} \times \mathscr{U}^{(2)} \times \cdots \times \mathscr{U}^{(k)} \times \mathbf{W} \tag{7.121}$$

for some sets $\mathscr{U}^{(1)}, \mathscr{U}^{(2)}, \ldots, \mathscr{U}^{(k)}$, and $\mathbf{W}$. Take note of the equality sign.

That is, $\mathscr{U}$ consists of all the possible combinations that can be generated by selecting w_j from some set $\mathscr{U}^{(j)}$, $j = 1, 2, \ldots, k$ and W from $\mathbf{W}$. Then, invoking simple dominance arguments, it follows that because x_j is non negative, the worst u in $\mathscr{U}^{(j)}$ is the largest element in $\mathscr{U}(j)$, and the worst value of W is the smallest value in $\mathbf{W}$.

Thus, we can replace (7.120) by

Model RC

$$z^{(rc)} = \max_{x_1,\ldots,x_n} \sum_{j=1}^{n} v_j x_j \ \text{ s.t. } \sum_{j=1}^{n} \overline{w}_j^{(i)} x_j \le \underline{W}\ ,\ x_j \in \{0,1\}, j = 1, \ldots, n \tag{7.122}$$

where

$$\overline{w}_{(j)} := \text{ largest element of } \mathscr{U}^{(j)}\ ,\ j = 1, 2, \ldots, k \tag{7.123}$$

$$\underline{W} := \text{ smallest element of } W \tag{7.124}$$

Observe that this is a standard *0-1 knapsack problem* that your favorite dynamic programming algorithm can handle.

A similar argument leads to the conclusion that under certain “independence” conditions, the size of the linear programming model specified by (7.117) can be reduced in a similar manner.

7.7.2 An infinite uncertainty space

Suppose that $\mathscr{U}$ consists of an infinite number of elements. Then the robust optimization problem under consideration, namely (7.112), will have an *infinite number of constraints.*

If the decision variables consists of a finite number of components, namely if $x = (x_1, \ldots, x_n)$ for some finite n, and the model has an infinite number of constraints, than the optimization problem is called *semi-infinite*. More precisely, a distinction is made between *semi-definite* optimization problems and *generalized semi-definite* optimization problems. the difference derives from the dependence of the constraints on the decision variable. If the set of constraints *depends on the decision variable,* then the problem is a generalized semi-infinite optimization problem, whereas if set of constraints is *independent of the decision variable* then the problem is a semi-infinite optimization problem.

In the framework of our discussion, this distinction is handled by the $\forall u \in \mathscr{U}$ vs $\forall u \in \mathscr{U}_x$ notation.

I note these facts here to draw the reader’s attention to this area of optimization theory, as it is extremely relevant to the type of optimization problems encountered in robust optimization.

7.7.3 Discussion

To conclude, as we saw in this chapter, the development of algorithms for robust optimization problems is fraught with difficulties. The greatest challenges are presented by *robustness constraints* of the generic $\forall u \in \mathscr{U}$ type.

I should point out, though, that despite the difficulties encountered in this effort, in the area of *linear programming* (and its extensions) there has been considerable success in the development of generic algorithms for a variety of robust counterparts of linear programming problems. Commercial software is also available for this purpose.

For an idea of the commercial software for *robust linear programming problems,* consider the following:

> **Robust Optimization and dealing with uncertainty**
>
> The current set-up within AIMMS allows you to solve linear and mixed integer programming (LP/MIP) models with uncertainty to optimality and create robust solutions without changing the actual structure of the models. In general, any LP/MIP solver can be used for solving RO models. However, for a specific class (i.e. when the uncertainty is defined as an ellipsoid), CPLEX or MOSEK are required as the so-called robust counterpart becomes a Second Order Cone Program (SOCP).
>
> **Read more about other ways of uncertainty modeling.**
>
> **Download a free license to try AIMMS for Robust Optimization**
>
> http://www.aimms.com/operations-research/mathematical-programming/robust-optimization
> Read on December 22, 2010

I should also add that my favorite class of optimization models, Dynamic Programming — as indicated in the bibliographic notes — also offers appropriate frameworks for the development of robust optimization models and algorithms.

Finally, because robust optimization models are typically *Maximin models,* the algorithms aimed for the solution of robust optimization problems are typically “maximin algorithms”. All the same, the specific features and capabilities of these algorithms can vary greatly. Which means of course that they are highly specialized procedures.

7.8 What next?

One of the striking facts about the main stream literature on *robust optimization* is its total oblivion of *Radius of Stability* models.

Of course, as I explained in this and in the preceding chapters, it is manifestly clear that *Radius of Stability* models are utterly unsuitable for the treatment of *severe* uncertainty, especially when the uncertainty is characterized by a *vast* uncertainty space, a *poor* estimate and a *likelihood-free* quantification of uncertainty.

The point remains, however, that robust optimization is concerned not only with *robustness to severe uncertainty*. *Robust optimization* is concerned with robustness period. It is surprising therefore that the generic *Radius of Stability model* does not even get a mention in the main stream literature on *robust optimization*.

The way I see it, this is part of an even more puzzling phenomenon, which is the absence of an in-depth discussion on the distinction between *local* and *global* robustness.

Don't get me wrong.

I am not suggesting that there is no awareness in the main stream literature on *robust optimization* of the fundamental distinction between *local* and *global* robustness. Obviously, this distinction is present, but it is present implicitly.

In other words, this distinction is not translated into a systematic treatment of the issue in the form of a formal framework similar to that used in *optimization theory* to deal with *local* and *global* optima.

To my mind, this is most unfortunate because, as the discussion in the second part of the book will amply demonstrate, such a framework is badly needed.

Which brings me to my objectives in the second part of the book.

This part of the book is devoted to a discussion of *info-gap decision theory* as a CASE STUDY. Its main objective is to illustrate how things can go wrong when the distinction between *local* and *global* robustness is not observed in the framework of robust decision-making in the face of severe uncertainty.

I describe the basic models put forth by this theory and I explain in detail the reasons that make this theory utterly unsuitable for the treatment of *severe uncertainty*.

For the benefit of *info-gap scholars*, I also explain the errors in the various attempts that have been made thus far to correct the fundamental flaws in the theory.

7.9 Bibliographic notes

TBW

Part II

info-gap decision theory

Part III

Epilogue