Contracting under Heterogeneous Beliefs

by

Mario Ghossoub

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2011

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The main motivation behind this thesis is the lack of *belief subjectivity* in problems of contracting, and especially in problems of demand for insurance. The idea that an underlying uncertainty in contracting problems (e.g. an insurable loss in problems of insurance demand) is a given random variable X on some exogenously determined probability space $(\Omega, \mathcal{F}, P)$ is so engrained in the literature that one can easily forget that the notion of an *objective uncertainty* is only one possible approach to the formulation of uncertainty in economic theory.

On the other hand, the *subjectivist school*, led by De Finetti [93] and Ramsey [234], challenged the idea that uncertainty is totally objective, and advocated a *personal view of probability* (subjective probability). This ultimately led to Savage's [266] approach to the theory of choice under uncertainty, where uncertainty is entirely subjective and it is only one's preferences that determine one's probabilistic assessment.

It is the purpose of this thesis to revisit the "classical" insurance demand problem from a purely subjectivist perspective on uncertainty. To do so, we will first examine a general problem of contracting under heterogeneous subjective beliefs and provide conditions under which we can show the existence of a solution and then characterize that solution. One such condition will be called *vigilance*. We will then specialize the study to the insurance framework, and characterize the solution in terms of what we will call a *generalized deductible* contract. Subsequently, we will study some mathematical properties of collections of *vigilant* beliefs, in preparation for future work on the idea of *vigilance*. This and other envisaged future work will be discussed in the concluding chapter of this thesis.

In the chapter preceding the concluding chapter, we will examine a model of *contracting for innovation* under *heterogeneity* and *ambiguity*, simply to demonstrate how the ideas and techniques developed in the first chapter can be used beyond problems of insurance demand.

Acknowledgements

Financial support from the following sources is gratefully acknowledged:

- The Natural Sciences and Engineering Research Council of Canada (for the NSERC Doctoral Scholarship);
- The Society of Actuaries (for the James C. Hickman Scholarship, formerly known as the SOA Doctoral Stipend);
- Tata Consulting Services (for RA funding in the Fall of 2008);
- The Waterloo Research institute in Insurance, Securities and Quantitative finance (for the WatRISQ Scholarships);
- The former Waterloo Institute of Quantitative Finance and Insurance (for the IQFI Scholarship);
- The University of Waterloo (for the President's Scholarships); and,
- The Department of Statistics and Actuarial Science at the University of Waterloo (for the Graduate Scholarship, the Chair's Award, and the Mathematics Graduate Experience Award).

Thanks are due to my supervisors, Prof. Andrew Heunis and Prof. Carole Bernard, for trying their best to accommodate my choice of these research topics, which might unfortunately be deemed unorthodox in a department such as ours. I thank Prof. Andrew Heunis for the numerous mathematical references that have been very helpful in developing my knowledge of functional analysis and topological measure theory. I have enjoyed my meetings with Prof. Heunis and all the conversations we have had about the history of Mathematics. I thank Prof. Carole Bernard for her time and investment.

I warmly thank Prof. Massimiliano Amarante, Prof. Itzhak Gilboa, Prof. Fabio Maccheroni, Prof. Massimo Marinacci, and Prof. Peter Wakker for their advice. I am especially grateful to Prof. Massimiliano Amarante, Prof. Fabio Maccheroni, and Prof. Massimo Marinacci for their encouragement. The kind hospitality of the Department of Decision Sciences at Bocconi University in Milan during the month of July 2010 and the Department of Economics of the University of Montréal during the month of September 2010 is much appreciated. Part of this research was done while I was there.

I thank Prof. Pierpaolo Battigalli, Prof. Guillaume Carlier, and Prof. Virginia Young for suggesting some helpful references, and Prof. Laurent W. Marcoux for introducing me to the theory of operator algebras.

I thank my PhD thesis committee for their time and effort: Prof. Carole Bernard, Prof. Marc Henry, Prof. Andrew Heunis, Prof. Adam Kolkiewicz, Prof. Paul Marriott, Prof. Selcuk Onay, and Prof. Tony Wirjanto.

I thank Mary-Lou Dufton, the Graduate Studies Co-ordinator, for her invaluable help all throughout my PhD program.

Finally, but most importantly, I thank my fiancée Katalin for her patience, understanding, and encouragement, and for always believing in me.

Dedication

This work is dedicated to the memory of Prof. K.H. Borch (1919 – 1986) in recognition of the incommensurable significance of his work and of its tremendous impact on the field of actuarial mathematics.

The work of Prof. Borch is a testimony of the fundamental role that microeconomic theory has in the development of actuarial science. He was responsible for the introduction of *uncertainty* into actuarial thought, not as a footnote to an otherwise flourishing field, but rather as a building block of a field whose development owes much to Prof. Borch's proliferous academic output. It is hard to imagine what actuarial science would have been today without Prof. Borch's contributions.

Contents

Preface

As soon as there is life there is danger
Ralph Waldo Emerson

Uncertainty in Economic Theory

Life and *danger* are metaphysical concepts which cannot be directly observed. Just as wind is not perceptible, but its consequences and effects are, so are the concepts of *life* and *danger*. As far as economic theory is concerned, *life* is discernible through the *ability to carry on a decision-making process*, and *danger* is manifest in the presence of *uncertainty*[1], which is itself observable in mathematical objects[2] such as a random variable, a capacity, etc. This presumably posits *decision theory*, also known as the *theory of choice under uncertainty*, at the center of every musing on the nature and scope of the actions of the human being as a social being endowed with free will, and even more so if this reflection is of a mathematical nature, e.g. in economic theory[3]. Indeed, the coming of age of economics as a distinctive science of human action is inseparable from the historical development of Expected-Utility Theory, although, from a purely epistemological positioning, this needs not be the case: as long as market phenomena are explained by a general theory taking the individual's freedom of choice as a starting point, any model of the economic agent's choice under uncertainty is normatively acceptable. We shall only require its mathematical consistency.

At the center of any theory of choice lie a few simple, yet deep questions:

[1]The term "uncertainty" is used here as a generic term that encompasses both the notion of *risk* and the concept of *Knightian uncertainty*. When the need for distinction arises the proper term will be used in each situation.

[2]Note that the validity of this argument relies on a prior ontological study of mathematical objects.

[3]The Austrian School in economics went as far as to consider the concept of marginal utility, which is a concept emerging from decision theory, as the cornerstone of economic theory.

1. How can we mathematically model the notion of an *uncertain prospect*?

2. How do people *act* when facing uncertainty?

3. Do people have *free will*?

4. If people have free will, how can their behavior be *predicted*?

The first two questions will be considered in Appendix A, which is a review of the classical theory of choice under risk and uncertainty, and of some of its paradoxes. The third and fourth questions are not discussed. Here, we will briefly discuss the three major ways in which *uncertainty* is defined in microeconomic theory, and in particular in the theory of choice under uncertainty.

Formulations of Uncertainty

Objective vs. Subjective Probability

It is customary in decision theory[4] to make the following distinction:

- An *objective probability* measure is any probability measure that is not derived from the individual's preferences over objects of choice, but one which is given exogenously; and,

- A *subjective probability* measure is a probability measure that is derived from the individual's preferences over objects of choice, and is hence derived endogenously.

Objective Uncertainty

The von Neumann-Morgenstern approach to the theory of choice ([297]) defines uncertainty as totally objective, in the sense that uncertainty is simply a random variable on an objectively (exogenously) given probability space. Equivalently, objective uncertainty can

[4] The reader is referred to any introductory textbook on the subject, such as Gilboa [149] or Kreps [187].

be identified with a situation in which the objects of choice are probability measures[5]. The von Neumann-Morgenstern setting will be reviewed in Appendix A.1.

In the vast majority of the financial literature and the actuarial literature, uncertainty (be it a random loss, a stock price, etc.) is seen as random variable (or a stochastic process, on an objectively given and appropriately filtered probability space). However, the idea that uncertainty can be totally objective has been severely criticized in economic theory, starting from the advocates of a *personal view of probability* (subjective probability), such as De Finetti [93], Ramsey [234], and especially Savage [266].

Subjective Uncertainty

Contrary to the von Neumann-Morgenstern approach to uncertainty – and building upon the work of De Finetti [93], Ramsey [234], and von Neumann and Morgenstern [297] – Savage [266] formulates uncertainty in the objects of choice *without* exogenous objective probabilities. Likelihoods are then derived from preferences, and probabilities are hence *subjective*. The individual's preferences over acts of choice induce a unique *subjective probability measure*. The Savage model will be exposed in Appendix A.1.

The Anscombe-Aumann Approach

The von Neumann-Morgenstern and the Savage approaches to uncertainty are opposite extremes, and the elegance of Savage's theory comes at a high price: the derivation of subjective expected utility is not an easy task. Admitting that there might exist objective randomizing devices (such as the toss of a coin), Anscombe and Aumann [19] introduced a setting where the uncertainty is partly objective and partly subjective. This setup then tremendously simplifies the derivation of a subjective probability measure from a decision maker's preferences over appropriately defined objects of choice. The Anscombe-Aumann model will be discussed in Appendix A.1.

Risk vs. Uncertainty

It is customary in decision theory[6] to make the following distinction:

[5] If X is a random variable on a probability space $(\Omega, \mathcal{F}, P)$, then the image measure of P under X, namely the quantity $P \circ X^{-1}$ is a Borel probability measure on $\mathbb{R}$.

[6] The reader is referred to any introductory textbook on the subject, such as Gilboa [149] or Kreps [187].

- We refer to a situation of decision under objective uncertainty as a situation of *decision under risk*; and,

- We refer to a situation of decision under subjective uncertainty as a situation of *decision under uncertainty*.

Ambiguity (Knightian Uncertainty)

Ellsberg [118] proposed a series of thought experiments to show that, contrary to the Bayesian paradigm (e.g. Savage's view), there ought to exist a meaningful distinction between subjective uncertainties that are *perfectly known* and subjective uncertainties that are *imperfectly known*. This distinction was argued for by Knight [184]. Consequently, we use the term *Knightian Uncertainty* to refer to a situation where the subjective uncertainties are not perfectly known, in the sense that the decision maker is not able to formulate a unique subjective probability measure that will represent his or her beliefs. Such a situation is also called a situation of *Ambiguity*, or of *Ambiguous Beliefs*, and it was the main motivation behind what is called today *neo-Bayesian theories of choice* (e.g. [14], [121], [123], [134], [139], [148], [151], [182], [195], [274], [275]). This is discussed in more detail in Appendix A.2.3 and Appendix A.3.2.

Chapter 1

Introduction

1.1 Some Historical Notes

Reflections on the nature of the insurance problem can be seen in Bernouilli [35], Smith [283], and Walras [298]; and then in Böhm-Bawerk [44], Lindenbaum [190], Tauber [291], and Willett [300]. For instance, Adam Smith [283] writes:

> "*That the chance of loss is frequently under-valued, and scarce ever valued more than it is worth, we may learn from the very moderate profit of insurers. In order to make insurance, either from fire or sea-risk, a trade at all, the common premium must be sufficient to compensate the common losses, to pay the expence of management, and to afford such a profit as might have been drawn from an equal capital employed in any common trade. The person who pays no more than this, evidently pays no more than the real value of the risk, or the lowest price at which he can reasonably expect to insure it. But though many people have made a little money by insurance, very few have made a great fortune; and from this consideration alone, it seems evident enough, that the ordinary balance of profit and loss is not more advantageous in this, than in other common trades by which so many people make fortunes. Moderate, however, as the premium of insurance commonly is, many people despise the risk too much to care to pay it. Taking the whole kingdom at an average, nineteen houses in twenty, or rather, perhaps, ninety-nine in a hundred, are not insured from fire. Sea risk is more alarming to the greater part of people, and the proportion of ships insured to those not insured is much greater. Many sail, however, at all seasons, and even in time of war, without any insurance. This may sometimes perhaps be done without any imprudence. When a great company, or even a great merchant,*

has twenty or thirty ships at sea, they may, as it were, insure one another. The premium saved upon them all, may more than compensate such losses as they are likely to meet with in the common course of chances. The neglect of insurance upon shipping, however, in the same manner as upon houses, is, in most cases, the effect of no such nice calculation, but of mere thoughtless rashness and presumptuous contempt of the risk." – ([283], Book I, Chap. 10)

Or,

"*The only trades which it seems possible for a joint stock company to carry on successfully without an exclusive privilege are those of which all the operations are capable of being reduced to what is called a routine, or to such a uniformity of method as admits of little or no variation. Of this kind is, first, the banking trade; secondly, the trade of insurance from fire, and from sea risk and capture in time of war; thirdly, the trade of making and maintaining a navigable cut or canal; and, fourthly, the similar trade of bringing water for the supply of a great city. [...] The value of the risk, either from fire, or from loss by sea, or by capture, though it cannot, perhaps, be calculated very exactly, admits, however, of such a gross estimation as renders it, in some degree, reducible to strict rule and method. The trade of insurance, therefore, may be carried on successfully by a joint stock company without any exclusive privilege. Neither the London Assurance, nor the Royal Exchange Assurance companies, have any such privilege.*" – ([283], Book V, Chap. 1)

It was not until the revolution initiated by von Neumann and Morgenstern in their seminal work entitled *Theory of Games and Economic Behavior* ([297]) that the theory of insurance found a proper foundation in the early work of Kenneth J. Arrow on this subject (e.g. [23]).

What is now called the "classical" theory of insurance, or the post-von-Neumann-Morgenstern theory, can be traced back to Arrow ([23], [24], [25], [26]), Borch ([46], [47], [48], [49], [50], [52]), and Raviv [239]. As Borch himself writes in [51],

"It is probably fair to say that the present interest in the economics of insurance springs from the theory of the economics of uncertainty which has been developed during the last twenty years. The pioneering work in this field is certainly Arrow's paper [from 1953]."

Or, as Gollier writes in [153],

> *"It is fair to state that the blossoming of research in this area in the last thirty years springs from the work of these two authors [(Arrow and Borch)]."*

Nevertheless, and as we will argue below, it seems that the extent to which the *theory of choice under uncertainty* was involved in laying down the foundations of the "classical" theory of insurance was limited to the approach of von Neumann and Moregenstern. In other words, the underlying uncertainty was seen as purely objective. Needless to say, this disregards the tremendous contributions of the *subjectivist school*, that is, De Finetti [93], Ramsey [234], and Savage [266]. More importantly, this leaves out of consideration some very interesting problems, one of which will be considered in this thesis, namely the problem of demand for insurance under heterogeneous subjective beliefs, and some others that will be discussed in Chapter 6.

1.2 Towards a *Subjectivist* Theory of insurance

Objective Uncertainty in the "Classical" Model

In the "classical" theory of insurance, the idea that the underlying uncertainty is a random variable X on some objectively given probability space $(\Omega, \mathcal{F}, P)$, where X represents the insurable loss, is almost never questioned. What this setting implicitly assumes is the existence of the *objective* probability measure P. This is not surprising if one adopts the von Neumann-Moregenstern view (see Appendix A.1.2), where the elements of choice are the random variables on $(\Omega, \mathcal{F})$, or equivalently the collection $\left\{P \circ X^{-1} : X \text{ is a random variable on } (\Omega, \mathcal{F})\right\}$ of exogenously given objective (Borel) probability measures on $\mathbb{R}$. With two exceptions that we are aware of (and that will be discussed in section 3.1), this is precisely how the insurance demand problem was approached in the past 60 years or so.

Subjective Uncertainty in this Thesis

As mentioned in the Preface, the approach to the notion of *uncertainty* in economic theory at large, and especially in decision theory, has been in constant modification since the work of von Neumann and Moregenstern. Savage [266] proposed a purely subjective approach, while Anscombe and Aumann [19] suggested a midway position that simplified the derivation of subjective probability. Finally, Knight's [184] intuition and Ellsberg's

[118] thought experiments led to the blossoming of *neo-Bayesian decision theory*, starting with the seminal contributions of Schmeidler (e.g. [275] and [151]).

It is then only natural to examine an alternative foundation of the notion of a *random loss* in problems of insurance demand, and more generally the notion of an *underlying uncertainty* in problems of contracting. This is precisely what the object of this thesis is. All throughout this thesis we will adopt a *purely subjectivist* approach to uncertainty, by assuming that the only primitives of our model are the primitives in Savage's [266] approach, namely, a collection of *states of the world*, a collection of *events*, a collection of *acts*, or objects of choice, and an individual's preference over *acts*. We will *not* assume the existence of an objective, exogenously given probability measure. Rather, as in Savage's [266] approach, we will start from the basic philosophical stance that it is solely the individual's preference over the elements of choice that determines the individual's beliefs. These beliefs are hence entirely subjective, and are represented by a *subjective probability* measure on the state space.

At points (in Chapter 5 of this thesis), we will also consider situations of *Knightian uncertainty*, i.e. *ambiguity*, where an individual's preference determines the individual's beliefs, but the latter are not represented by an *additive* probability measure. Rather, as in Schmeidler's [275] approach, these beliefs are represented by a *non-additive* measure, also known as a *capacity*.

1.3 Outline of the Thesis

The rest of the thesis is organized as follows:

1. In Chapter 2, we examine a general model of *contracting under heterogeneous subjective beliefs*. We consider an abstract problem of demand for claims that are contingent on some given underlying uncertainty, and nondecreasing functions of that uncertainty. A decision makers' (DM) wealth depends on both the uncertainty and the contingent claim that is issued to her by a claim issuer (CI). The DM seeks to maximize her expected utility of wealth with respect to her subjective probability measure, whereas claims are evaluated by the CI according to his expected utility of wealth with respect to his subjective probability measure. We show that under a consistency requirement on the subjective probabilities that we call *vigilance*, we can drop the monotonicity constraint – hence simplifying the problem considerably – and we can show the existence of a monotone solutions. We then provide a technique for

dealing with the heterogeneity of beliefs. We also show that in most relevant situations, the assumption of *vigilance* is implied by the more or less usual assumption of a *monotone likelihood ratio*, when the latter can be defined.

2. In Chapter 3, we examine an important special case of the abstract model of Chapter 2, namely a problem of *demand for insurance with heterogeneous beliefs*. Unlike the "classical" approach to this old problem, we consider a situation where the insurer and the risk-averse DM have different subjective beliefs, and hence assign different probabilities to the realizations of a given insurable random loss. The decision maker seeks to maximize her expected utility of terminal wealth with respect to her subjective probability measure, whereas the insurer sets premiums on the basis of his subjective probability measure. We show that if *vigilance* holds then there exists an event to which the DM assigns full subjective probability and on which an optimal solution has a *generalized deductible* form.

3. In Chapter 4, we study the mathematical structures underlying some collections of *vigilant beliefs*, and we give a crisp characterization of *second-order vigilant beliefs*, i.e. beliefs about vigilant beliefs. The results of Chapter 4 will be the basic mathematical tools used in future work about the notion of *vigilance*, as discussed in Chapter 6.

4. In Chapter 5 we examine a model of *contracting for innovation* under heterogeneity and ambiguity, in order to demonstrate how the ideas and techniques developed in Chapter 2 can be used beyond problems of insurance demand. A DM is an *innovator* who wishes to sell her *innovation* to an interested CI. The latter will pay a fixed fee $H > 0$ as a lump-sum upon entering into the contract, in return of which he will receive the innovation (quantified monetarily) and pay an amount to the DM contingent on the value of that innovation. We will consider a situation where the CI's preference has a *Choquet-Expected Utility* (CEU) representation (as in Schmeidler [275] and Gilboa [148]), hence reflecting some level of *ambiguity* in the CI's beliefs. We then show that there exists a monotone solution to the DM's demand problem, using similar techniques to those of Chapter 2. We also consider a special case and characterize the solution in that case. Finally, we consider a situation where both the DM and the CI have preferences admitting a representation that exhibits some ambiguity in their beliefs, whereby that the CI is a CEU-maximizer and the DM's preferences are represented by a *symmetric Choquet integral* (defined in Appendix 5.7), hence reflecting some *gain-loss* separability, in the spirit of the *Cumulative Prospect Theory* of Kahneman and Tversky ([177] and [293]). We show that even in this setting, the notion of *vigilance* is fruitful: if *vigilance* holds, then for any payoff function Y_1 which is feasible for the DM's problem there is another feasible payoff function Y_2 which is (i) a nondecreasing function of the underlying uncertainty, and

(ii) such that for a given fixed fee $H >$ the CI prefers a contract in which he pays Y_2 to a contract in which he pays Y_1. However, to prove this result we had to extend the notion of an equimeasurable rearrangement to the case where we have a non-additive measure instead of a measure (this is done in Appendix 5.8).

5. In Chapter 6 we outline our future research agenda as it relates to the work done in this thesis. In particular, we discuss how the results of Chapter 4 will be useful in future investigations of the notion of *vigilance*. We also discuss possible extensions of the models of Chapter 2, Chapter 3, and Chapter 5.

The Appendices at the end of some chapters contain specific material directly related to some ideas developed in these chapters, whereas the Appendices at the end of this thesis provide all necessary background (both mathematical and decision-theoretic).

Finally, in Appendix A.3.1 we briefly discuss *Cumulative Prospect Theory* and some of the work done in Bernard and Ghossoub [32] and in Ghossoub [141].

Chapter 2

Contracting under Heterogeneous Beliefs

2.1 Introduction

In this chapter we examine an abstract contracting problem under heterogeneous subjective beliefs. The main motivation behind this study is its application to the insurance setting, which will be considered in Chapter 3.

The DM faces a fundamental uncertainty that she wishes to transfer to the CI for a fixed fee, and in return of a claim contingent on the initial uncertainty (e.g. a risk sharing rule). The model will be formally introduced in section 2.2. For instance, in problems of insurance demand, the CI can be interpreted as the insurer, the fee paid by the DM can be seen as the premium, and the claim issued by the CI is the indemnity paid by the insurer.

The DM's problem is that of finding the contingent claim that will maximize her expected utility of terminal wealth, subject to some constraints. These constraints typically include the CI's *individual rationality constraint*, or *participation constraint*, and a monotonicity constraint stipulating that the contingent claim should be a monotone function of the underlying uncertainty (typically, a nondecreasing function). In the insurance setting, for example, the optimal indemnity is desired to be nondecreasing in the loss, so as to prevent some moral hazard issues that might result from a downward misrepresentation of the loss by the DM (see Huberman, Mayers and Smith [169], for instance).

This contracting problem can be reformulated as a problem of demand for claims that are contingent on some given underlying uncertainty. A decision makers' (DM) wealth depends on both the uncertainty and the contingent claim that is issued to her by a claim issuer (CI). The DM seeks to maximize her expected utility of wealth with respect to her subjective probability measure, whereas claims are evaluated by the CI according to his expected utility of wealth with respect to his subjective probability measure.

Problems of Demand for Contingent Claims

Many problems of contracting can be formulated as problems of demand for contingent claims, as follows:

$$\sup_I \left\{ \int u\left(a - X + I \circ X\right) \, dP : I \in \Theta \right\} \tag{2.1}$$

for some $a \in \mathbb{R}$, where X is a representation of some underlying uncertainty, $I \circ X$ is a contingent claim that depends on the uncertainty X, $u\left(a - X + I \circ X\right)$ is the DM's utility of wealth (see below), and P is a probability measure on the state space. The collection Θ is a given problem-specific nonempty feasibility set that often restricts I to be nondecreasing, and accounts for the CI's *individual-rationality constraint* (a.k.a. *participation constraint*), as will be explained below.

For instance, "classical" problems of demand for insurance coverage against a random loss[1] can be formulated as in (2.1), with X being the underlying insurable loss against which a DM seeks an insurance coverage $Y = I \circ X$. If the DM is a risk-averse Expected-Utility (EU) maximizer, with utility function u and initial wealth W_0, and if the insurance premium is a given $\Pi > 0$ and P is a probability measure on the state space, then the DM's problem of optimal demand for insurance can be written as:

Problem 2.1. *For a given loading factor $\rho \geq 0$,*

$$\sup_I \left\{ \int u\Big(W_0 - \Pi - X + I \circ X\Big) \, dP \right\} :$$
$$\begin{cases} 0 \leq I \circ X \leq X \\ \Pi \geq (1+\rho) \int Y \, dP \\ I \text{ is a nondecreasing function} \end{cases}$$

[1]See Appendix 3.8 for a review of the "classical" insurance demand problem.

Belief Heterogeneity

The abstract problem that we will examine in this chapter takes the following form:

$$\sup_I \left\{ \int u\left(a - X + I \circ X\right)\ d\mu :\ 0 \leq I \circ X \leq X,\ I \text{ is nondecreasing}, \right.$$
$$\left. \text{and } \int v\left(b - I \circ X\right)\ d\nu \geq R \right\} \tag{2.2}$$

where X is the underlying uncertainty (defined below), $I \circ X$ is a contingent claim that depends on the uncertainty X, $a, b \in \mathbb{R}$, $u\left(a - X + I \circ X\right)$ is the DM's utility of wealth, μ is the DM's *subjective* probability measure, $v\left(b - I \circ X\right)$ is the CI's utility of wealth, ν is the CI's *subjective* probability measure, and $R \in \mathbb{R}$ is given. The functions u and v and the probability measure μ and ν are discussed and defined below.

The most important observation to be made here is that the probability measures μ and ν are not identical, *a priori*. This poses some important mathematical complications, and makes the problem radically different from what has been done in the related literature on contracting, and especially in problems of insurance demand.

This Chapter's Contribution

In this chapter, we propose the notion of *vigilant beliefs* as a consistency requirement between the subjective beliefs of both parties, and we show how *vigilance* leads to the existence of a solution to the DM's problem which is a nondecreasing function of the underlying uncertainty.

We will also give a general technique for solving the DM's demand problem, within the abstract framework of this chapter. In Chapter 3, we will apply the techniques developed here to the specific problem of optimal insurance design.

Outline

In section 2.2 we introduce some notation and definitions, as well as the general setup for our model. In section 2.3 we discuss the notion of an equimeasurable rearrangement of a measurable function with respect to another measurable function. In section 2.4 we state the DM's problem and give some general techniques for showing the existence and monotonicity of a solution, and for characterizing this solution. In section 2.5 we compare our *vigilance* requirement with an alternative "monotonicity" requirements that we might have otherwise imposed, namely a *Monotone Likelihood Ratio* (MLR). Finally, section 2.6 concludes. Appendices 2.7 and 2.8 contain some useful related results that are used in this chapter.

2.2 Preliminaries

Let S denote the set of states of the world, and suppose that $\mathcal{G}$ is a σ-algebra of subsets of S, called events. We denote by $B(\mathcal{G})$ the supnorm-normed Banach space of all bounded, $\mathbb{R}$-valued and $\mathcal{G}$-measurable functions on $(S, \mathcal{G})$ (see Proposition D.23 on p. 229), and we denote by $B^+(\mathcal{G})$ the collection of all $\mathbb{R}^+$-valued elements of $B(\mathcal{G})$. For any $f \in B(\mathcal{G})$, the supnorm of f is given by $\|f\|_s := \sup\{|f(s)| : s \in S\} < +\infty$. For $C \subseteq S$, we will denote by $\mathbf{1}_C$ the indicator (characteristic) function of C. For any $A \subseteq S$ and for any $B \subseteq A$, we will denote by $A \setminus B$ the complement of B in A.

For any $f \in B(\mathcal{G})$, we shall denote by $\sigma\{f\}$ the σ-algebra of subsets of S generated by f, and we shall denote by $B(\sigma\{f\})$ the linear space of all bounded, $\mathbb{R}$-valued and $\sigma\{f\}$-measurable functions on $(S, \mathcal{G})$. Then by Doob's measurability theorem (Theorem D.4 on p. 224), for any $g \in B(\sigma\{f\})$ there exists a Borel-measurable map $\zeta : \mathbb{R} \to \mathbb{R}$ such that $g = \zeta \circ f$. We will denote by $B^+(\sigma\{f\})$ the cone of nonnegative elements of $B(\sigma\{f\})$.

2.2.1 Uncertainty and Preferences

We assume that the DM faces a fundamental uncertainty that affects her wealth and consumption. This uncertainty will be modeled to be a (henceforth fixed) element X of $B^+(\mathcal{G})$ with a closed range $[0, M] := X(S)$, where $M := \|X\|_s < +\infty$. In other words, the uncertainty is a mapping of S onto the closed interval $[0, M]$. For instance, in problems of demand for insurance the uncertainty X can be seen as the underlying insurable loss against which the DM seeks an insurance coverage $I \circ X$. In problems of optimal debt contracting, the uncertainty X can be seen as the interest on a loan, and $I \circ X$ as the repayment scheme. Hereafter, we shall denote by Σ the σ-algebra $\sigma\{X\}$ of subsets of S generated by X.

If $\mathcal{A}$ is any sub-σ-algebra of $\mathcal{G}$ such that $\Sigma \subseteq \mathcal{A}$ and if P is any probability measure on the measurable space $(S, \mathcal{A})$, we will say that X is a continuous random variable for P when the law $P \circ X^{-1}$ of X is a nonatomic Borel probability measure[2].

The DM's decision process is assumed to consist in choosing a certain *act* among a collection of given *acts* whose realization, in each state of the world s, depends on the value $X(s)$ of the uncertainty X is the state s. Specifically, we will assume that these *acts* are the elements of $B^+(\Sigma)$.

Formally, the DM and the CI have preferences over acts in a framework à la Savage (see Appendix A.1.3). Here, the set of consequences (or prizes) is taken to be $\mathbb{R}$. Let $\mathcal{F}$ denote the collection of all $\mathcal{G}$-measurable functions $f : S \to \mathbb{R}$. The elements of choice (or *acts*) are taken to be the elements of $B^+(\Sigma) \subset \mathcal{F}$. The nature of the problem makes this a natural assumption. Indeed, what we are interested in is determining the optimal function of the uncertainty, that is, the optimal claim $Y := I \circ X \in B^+(\Sigma)$, for some Borel-measurable map $I : X(S) \to \mathbb{R}^+$, that will satisfy a certain set of requirements (constraints).

The DM's preferences $\succcurlyeq_{DM}$ over $B^+(\Sigma)$ and the CI's preferences $\succcurlyeq_{CI}$ over $B^+(\Sigma)$ admit a Subjective Expected-Utility (SEU) representation, as in Theorem A.20 in Appendix A.1.3. Formally, we assume the following:

Assumption 2.2. *There are bounded, nondecreasing, and continuous utility functions* $u, v : \mathbb{R} \to \mathbb{R}$, *unique up to a positive affine transformation, and unique countably additive*[3] *subjective probability measures* μ *and* ν *on the measurable space* (S, Σ), *such that for each* $Y_1, Y_2 \in B^+(\Sigma)$,

$$Y_1 \succcurlyeq_{DM} Y_2 \iff \int u(Y_1)\ d\mu \geq \int u(Y_2)\ d\mu \tag{2.3}$$

and

$$Y_1 \succcurlyeq_{CI} Y_2 \iff \int v(Y_1)\ d\nu \geq \int v(Y_2)\ d\nu \tag{2.4}$$

In sum, the DM's problem here is choosing the optimal *act* $Y^* \in B^+(\Sigma)$ that will

[2]A finite measure η on a measurable space $(\Omega, \mathcal{A})$ is said to be *nonatomic* if for any $A \in \mathcal{A}$ with $\eta(A) > 0$, there is some $B \in \mathcal{A}$ such that $B \subsetneq A$ and $0 < \eta(B) < \eta(A)$.

[3]Countable additivity of the subjective probability measure representing preferences can be obtained by assuming that preferences satisfy the Arrow-Villegas *Monotone Continuity* axiom (see p. 178). See Arrow [25], Chateauneuf et al. [77], and Villegas [296]. See also Corollary A.19 on p. 179.

maximize her expected utility of wealth, with respect to her subjective probability measure μ.

We also make the assumption that the uncertainty X (with closed range $[0, M]$) has a nonatomic law[4] induced by the probability measure μ, and that the CI and the DM are both aware of the fact that μ represents the DM's beliefs and ν represents that CI's beliefs. Specifically:

Assumption 2.3. *We assume that:*

1. $\mu \circ X^{-1}$ *is nonatomic;*

2. μ *is known by the CI; and,*

3. ν *is known by the DM.*

The assumption of nonatomicity of $\mu \circ X^{-1}$ is simply a technical requirement that is needed for defining the notion of an *equimeasurable monotone rearrangement*, as will be seen in section 2.3.

2.2.2 Uncertainty and Wealth

The contract between the DM and CI is a pair $(\Pi, Y) \subset \mathbb{R}^+ \setminus \{0\} \times B^+(\Sigma)$, whereby upon entering into the contract with the CI, the DM pays a fixed fee $\Pi > 0$ to the CI, in return of which she receives the amount $Y(s)$ from the CI in state $s \in S$.

The DM has initial wealth W_0, and after entering into the contract (Π, Y) with the CI, her total wealth is the Σ-measurable, $\mathbb{R}$-valued and bounded function on S defined by

$$W^{DM}(\Pi, Y)(s) := W_0 - \Pi - X(s) + Y(s), \ \forall s \in S \tag{2.5}$$

The CI has initial wealth W_0^{CI}, and after entering into the contract (Π, Y) with the DM, his total wealth is the Σ-measurable, $\mathbb{R}$-valued and bounded function on S defined by

[4] A finite measure η on a measurable space $(\Omega, \mathcal{A})$ is said to be *nonatomic* if for any $A \in \mathcal{A}$ with $\eta(A) > 0$, there is some $B \in \mathcal{A}$ such that $B \subsetneq A$ and $0 < \eta(B) < \eta(A)$.

$$W^{CI}(\Pi, Y)(s) := W_0^{CI} + \Pi - Y(s)\,,\ \forall s \in S \tag{2.6}$$

Letting $R := v\left(W_0^{CI}\right)$ be the CI's *reservation utility*, the CI's *individual rationality constraint*, or *participation constraint*, for entering into a contract (Π, Y) with the CI is then given by

$$\int v\Big(W_0^{CI} + \Pi - Y\Big)\, d\nu \geq v\left(W_0^{CI}\right) \tag{2.7}$$

2.2.3 Vigilant Beliefs and Probabilistic Consistency

Definition 2.4. *The probability measure ν is said to be (μ, X)-vigilant if for any $Y_1, Y_2 \in B^+(\Sigma)$ such that*

(i) Y_1 and Y_2 have the same distribution under μ, and,

(ii) Y_2 is a nondecreasing function of X,

the following holds:

$$W^{CI}(\Pi, Y_2) \succcurlyeq_{CI} W^{CI}(\Pi, Y_1) \tag{2.8}$$

Remark 2.5. *An equivalent definition of vigilance in this context is the following: the probability measure ν is (μ, X)-vigilant if for any $Y_1, Y_2 \in B^+(\Sigma)$ such that*

(i) Y_1 and Y_2 have the same distribution under μ,

(ii) Y_2 is a nondecreasing function of X,

the following holds:

$$\int v\Big(W_0^{CI} + \Pi - Y_2\Big)\, d\nu \geq \int v\Big(W_0^{CI} + \Pi - Y_1\Big)\, d\nu \tag{2.9}$$

Remark 2.6. *In this context, vigilance can simply be seen as a technical condition. It is a feasibility condition that will allow us to show that whenever the feasibility set of the DM's demand problem (stated below) is nonempty, there are feasible functions that are nondecreasing in X. Most importantly, vigilance will show that for any feasible claim for the the DM's demand problem, there is another feasible claim which is not only nondecreasing in X but also Pareto-improving (as defined later – Definition 2.27 on p. 38).*

However, depending on the specific application of the general contracting problem outlined here, vigilance becomes more than just a technical condition. For instance,

1. *the interpretation of (μ, X)-vigilance in the specific problem of insurance design that will be examined in Chapter 3 is more or less natural, and in that context vigilance is more than simply a feasibility condition. Indeed, the intuition behind this terminology comes from the insurance framework.*

2. *Also, in Amarante, Ghossoub, and Phelps [16], vigilance has a very natural interpretation that is very well suited for the specific problem of entrepreneurship that the authors discuss. We refer the interested reader to that paper for more about that specific setting.*

In section 2.5 we will show that in the specific setting where the DM and the CI assign different probability density functions to the uncertainty X with range $[0, M]$, the assumption of vigilance is implied by the (more or less usual) assumption of a monotone likelihood ratio. This result shows yet another useful property of the notion of vigilance. Indeed, whenever densities cannot be defined or simply do not exist, vigilance can be used.

2.3 Equimeasurable Monotone Rearrangements and Supermodularity

In this section we will discuss the notion of an *equimeasurable monotone rearrangement* of a measurable function with respect to another measurable function. This notion will be the basic tool that we will use to show the existence of a monotone solution to the DM's problem (introduced in section 2.4).

The concept of an equimeasurable rearrangement of a Borel-measurable function on $\mathbb{R}$ with respect to a finite Borel measure, and the notion of an equimeasurable rearrangement of a measurable function f from a measurable space into $\mathbb{R}$ with respect to a finite Borel measure on the range of f is by now part of the classical literature[5].

Here, we are interested in the idea of an equimeasurable rearrangement of a random variable with respect to another random variable. The nomenclature used here has been chosen with the present context in mind, whereby the same measurable space may be endowed with different measures.

In this section we introduce two specific formulations of the nondecreasing rearrangement of any element Y of $B^{+}(\Sigma)$ with respect to the fixed underlying uncertainty X. Although some of the results presented here are not new, the approach is novel, to the best of our knowledge.

2.3.1 The Nondecreasing Rearrangement

Let $(S, \mathcal{G}, P)$ be a probability space, and let $X \in B^{+}(\mathcal{G})$ be a continuous random variable[6] with range $[0, M] := X(S)$, where $M := \sup\{X(s) : s \in S\} < +\infty$, i.e. X is a mapping of S onto the closed interval $[0, M]$. Denote by Σ the σ-algebra $\sigma\{X\}$, and denote by ϕ the law of X defined by

$$\phi(B) := P\Big(\{s \in S : X(s) \in B\}\Big) = P \circ X^{-1}(B) \tag{2.10}$$

for any Borel subset B of $\mathbb{R}$.

If $I, I_n : [0, M] \to [0, M]$, for each $n \geq 1$, we will write $I_n \downarrow I$, ϕ-a.s., to signify that the sequence $\{I_n\}_n$ is a nonincreasing sequence of functions and that $\lim\limits_{n \to +\infty} I_n(t) = I(t)$, for ϕ-a.a. $t \in [0, M]$. Similarly, we will write $I_n \uparrow I$, ϕ-a.s., to signify that the sequence $\{I_n\}_n$ is a nondecreasing sequence of functions and that $\lim\limits_{n \to +\infty} I_n(t) = I(t)$, for ϕ-a.a. $t \in [0, M]$.

[5] See, e.g. Bennett and Sharpley [31], Carlier and Dana [72], [73] and [74], Chong [78], [79] and [80], Chong and Rice [81], Dana and Scarsini [88], Epperson [120], Hardy, Littlewood and Pólya [163], Luxemburg [192], or Rakotoson [233]. The results of Epperson [120] are the most relevant in this case, since [120] considers rearrangements of Borel measurable functions with respect to arbitrary nonatomic measures, that is, not necessarily the Lebesgue measure. Nevertheless, all the results for the case of a rearrangement with respect to the Lebesgue measure can be easily generalized to the case of any nonatomic finite Borel measure.

[6] Recall that this means that $P \circ X^{-1}$ is a nonatomic Borel probability measure.

Definition 2.7. *For any Borel-measurable map $I : [0, M] \to \mathbb{R}$, define the distribution function of I as the map $\phi_I : \mathbb{R} \to [0, 1]$ defined by*

$$\phi_I(t) := \phi\big(\{x \in [0, M] : I(x) \leq t\}\big) \tag{2.11}$$

Then ϕ_I is a nondecreasing right-continuous function, and the function $t \mapsto 1 - \phi_I(t)$ is called the survival function of I.

Definition 2.8. *Let $I : [0, M] \to [0, M]$ be any Borel-measurable map, and define the function $\widetilde{I} : [0, M] \to \mathbb{R}$ by:*

$$\widetilde{I}(t) := \inf\Big\{z \in \mathbb{R}^+ : \phi_I(z) \geq \phi\big([0, t]\big)\Big\} \tag{2.12}$$

The following proposition gives some useful properties of the map $\widetilde{I}$ defined above.

Proposition 2.9. *Let $I : [0, M] \to [0, M]$ be any Borel-measurable map and let $\widetilde{I} : [0, M] \to \mathbb{R}$ be defined as in (2.12). Then the following hold:*

1. *$\widetilde{I}$ is left-continuous, nondecreasing, and Borel-measurable;*
2. *For each $t \in [0, M]$, $\phi_I\left(\widetilde{I}(t)\right) \geq \phi([0, t])$;*
3. *$\widetilde{I}(t) \geq 0$, for each $t \in [0, M]$, $\widetilde{I}(0) = 0$, and $\widetilde{I}(M) \leq M$;*
4. *If $I_1, I_2 : [0, M] \to [0, M]$ are such that $I_1 \leq I_2$, ϕ-a.s., then $\widetilde{I}_1 \leq \widetilde{I}_2$;*
5. *If $Id : [0, M] \to [0, M]$ denotes the identity function, then $\widetilde{Id} \leq Id$;*
6. *$\widetilde{I}$ is ϕ-equimeasurable with I, in the sense that for any Borel set B,*

$$\phi\Big(\{t \in [0, M] : I(t) \in B\Big) = \phi\Big(\{t \in [0, M] : \widetilde{I}(t) \in B\}\Big) \tag{2.13}$$

7. *If* $\overline{I} : [0, M] \to \mathbb{R}^+$ *is another nondecreasing, Borel-measurable map which is* ϕ-*equimeasurable with* I*, then* $\overline{I} = \widetilde{I}$*,* ϕ-*a.s.;*

8. *If* $I, I_n : [0, M] \to [0, M]$*, for each* $n \geq 1$*, and* $I_n \downarrow I$*,* ϕ-*a.s., then* $\widetilde{I}_n \downarrow \widetilde{I}$*,* ϕ-*a.s.*

Proof.

1. The monotonicity of $\widetilde{I}$, and hence its Borel-measurability, is clear. Left-continuity of $\widetilde{I}$ is an immediate consequence of the monotonicity, the nonatomicity, and the continuity of the measure ϕ for monotone sequences, as well as the left-continuity of the left-inverse[7] of the distribution function of $\widetilde{I}$, namely the function $\phi_I^*(t) := \inf\Big\{z \in \mathbb{R}^+ : \phi_I(z) \geq t\Big\}$, for $t \in [0, 1]$;

2. This is an immediate consequence of the right-continuity of the distribution function ϕ_I of I;

3. By the very definition of $\widetilde{I}$ given in (2.12), we have $\widetilde{I}(t) \geq 0$ for each $t \in [0, M]$. Now, $\phi([0, 0]) = \phi(\{0\}) = 0$, by nonatomicity of ϕ. Therefore, for each $x \geq 0$, $\phi\big(\{t \in [0, M] : I(t) \leq x\}\big) \geq \phi([0, 0])$. In particular,

$$\phi\big(\{t \in [0, M] : I(t) \leq 0\}\big) = \phi\big(\{t \in [0, M] : I(t) = 0\}\big) \geq \phi([0, 0])$$

Hence, by (2.12), $\widetilde{I}(0) \leq 0$, and so $\widetilde{I}(0) = 0$. Moreover, for each $x \in [0, M]$,

$$1 = \phi([0, M]) \geq \phi\big(\{t \in [0, M] : I(t) \leq x\}\big)$$

Therefore, $\Big\{z \in \mathbb{R}^+ : \phi\big(\{x \in [0, M] : I(x) \leq z\}\big) \geq \phi\big([0, M]\big)\Big\} = \Big\{z \in \mathbb{R}^+ : \phi\big(\{x \in [0, M] : I(x) \leq z\}\big) = 1\Big\}$. Since $I(t) \leq M$ for each $t \in [0, M]$, it follows that $M \in \Big\{z \in \mathbb{R}^+ : \phi\big(\{x \in [0, M] : I(x) \leq z\}\big) = 1\Big\}$, and so from (2.12) we have that

$$\widetilde{I}(M) = \inf\Big\{z \in \mathbb{R}^+ : \phi\big(\{x \in [0, M] : I(x) \leq z\}\big) = 1\Big\} \leq M;$$

[7] See also Embrechts and Hofert [119] for more about the left-inverse (a.k.a. the left-continuous inverse) of a nondecreasing function.

4. Let $I_1, I_2 : [0, M] \to [0, M]$ be such that $I_1 \le I_2$, ϕ-a.s. Then, for each $x \ge 0$, $\phi\Big(\{t \in [0, M] : I_1(t) \le x\}\Big) \ge \phi\Big(\{t \in [0, M] : I_2(t) \le x\}\Big)$. Therefore, for each $t \in [0, M]$,

$$\Big\{z \in \mathbb{R}^+ : \phi\big(\{x \in [0, M] : I_2(x) \le z\}\big) \ge \phi\big([0, t]\big)\Big\}$$
$$\subseteq \Big\{z \in \mathbb{R}^+ : \phi\big(\{x \in [0, M] : I_1(x) \le z\}\big) \ge \phi\big([0, t]\big)\Big\}$$

It then follows from (2.12) that $\widetilde{I}_1 \le \widetilde{I}_2$;

5. By (2.12), for each $t \in [0, M]$,

$$\begin{aligned}\widetilde{Id}(t) &= \inf\Big\{z \in \mathbb{R}^+ : \phi\big(\{x \in [0, M] : Id(x) \le z\}\big) \ge \phi\big([0, t]\big)\Big\} \\ &= \inf\Big\{z \in \mathbb{R}^+ : \phi\big([0, z]\big) \ge \phi\big([0, t]\big)\Big\}\end{aligned}$$

Therefore, $\widetilde{Id}(t) \le t = Id(t)$, for each $t \in [0, M]$;

6. To show that $\widetilde{I}$ is ϕ-equimeasurable with I, we need to show that for any Borel set B,

$$\phi\Big(\{t \in [0, M] : I(t) \in B\Big) = \phi\Big(\{t \in [0, M] : \widetilde{I}(t) \in B\}\Big) \tag{2.14}$$

We first show that for each $\alpha \in [0, M]$,

$$\phi\Big(\{t \in [0, M] : I(t) \le \alpha\}\Big) = \phi\Big(\{t \in [0, M] : \widetilde{I}(t) \le \alpha\}\Big)$$

Since $\widetilde{I}$ is nondecreasing, we have that for each $\alpha \in [0, M]$ there is some $x_0 \in [0, M]$ such that the set $\Big\{x \in [0, M] : \widetilde{I}(x) \le \alpha\Big\}$ has the form $[0, x_0)$ or $[0, x_0]$, with $\widetilde{I}(x) > \alpha$ for each $x \in (x_0, M]$. But by nonatomicity of ϕ, we have

$$\phi\big([0, x_0)\big) = \phi\big([0, x_0]\big) = \phi\big(\{x \in [0, M] : \widetilde{I}(x) \le \alpha\}\big)$$

Moreover, by left-continuity of $\widetilde{I}$, it follows that $\widetilde{I}(x_0) \le \alpha$. Therefore, since ϕ_I is nondecreasing, we have

$$\phi\Big(\{t \in [0,M] : I(t) \le \widetilde{I}(x_0)\}\Big) \le \phi\Big(\{t \in [0,M] : I(t) \le \alpha\}\Big)$$

Now, from (2) above, we have $\phi_I\left(\widetilde{I}(x_0)\right) \ge \phi([0,x_0]) = \phi([0,x_0))$. Therefore,

$$\phi([0,x_0]) = \phi([0,x_0)) \le \phi_I\left(\widetilde{I}(x_0)\right) \le \phi_I(\alpha)$$

Suppose that $\phi([0,x_0]) < \phi_I(\alpha)$. Then there is some $z_0 \in (x_0, M]$ such that

$$\phi([0,z_0]) = \phi([0,z_0)) = \phi_I(\alpha)$$

Thus,

$$\begin{aligned}\widetilde{I}(z_0) &= \inf\left\{z \in \mathbb{R}^+ : \phi_I(z) \ge \phi\big([0,z_0]\big)\right\} \\ &= \inf\left\{z \in \mathbb{R}^+ : \phi_I(z) \ge \phi_I(\alpha)\,\right\} \\ &\le \alpha\end{aligned}$$

contradicting the fact that $\widetilde{I}(x) > \alpha$ for each $x \in (x_0, M]$. Therefore,

$$\phi([0,x_0]) = \phi([0,x_0)) = \phi_I(\alpha) = \phi_{\widetilde{I}}(\alpha);$$

Thus, for each $\alpha \in [0,M]$, we have

$$\phi\Big(\{t \in [0,M] : I(t) \le \alpha\}\Big) = \phi\Big(\{t \in [0,M] : \widetilde{I}(t) \le \alpha\}\Big)$$

Now, the collection $\left\{\,[0,\alpha] : \alpha \in [0,M]\,\right\}$ is a π-system (see Appendix 2.8) that generates the Borel σ-algebra on $[0,M]$ (see Resnick [240], pp. 18-19). Moreover, the collection of all Borel subsets B of $\mathbb{R}$ such that $\phi \circ I^{-1}(B) = \phi \circ \widetilde{I}^{-1}(B)$ is easily seen to be a λ-system (see Appendix 2.8 and Resnick [240], Proposition 2.2.3 on p. 37). Therefore, by Dynkin's π-λ theorem (Theorem 2.53 on p. 51), $\phi \circ I^{-1}(C) = \phi \circ \widetilde{I}^{-1}(C)$, for each Borel subset C of $[0,M]$. That is, for any Borel set B,

$$\phi\Big(\{t \in [0,M] : I(t) \in B\Big) = \phi\Big(\{t \in [0,M] : \widetilde{I}(t) \in B\}\Big) \tag{2.15}$$

7. Let $\overline{I} : [0, M] \to \mathbb{R}^+$ be another nondecreasing, Borel-measurable map which is ϕ-equimeasurable with I. To show that $\overline{I} = \widetilde{I}$, ϕ-a.s., it is enough to show that $\phi\big(\{x \in [0, M] : \widetilde{I}(x) > \overline{I}(x)\}\big) = \phi\big(\{x \in [0, M] : \overline{I}(x) > \widetilde{I}(x)\}\big) = 0$. Let $\mathbb{Q}$ denote the set of all rational numbers. Then

$$\Big\{x \in \ [0, M] : \overline{I}(x) < \widetilde{I}(x)\Big\}$$
$$= \bigcup_{q \in \mathbb{Q}} \Big(\{x \in [0, M] : \overline{I}(x) < q\} \ \cap \ \{x \in [0, M] : q \le \widetilde{I}(x)\}\Big)$$

Fix $q \in \mathbb{Q}$ arbitrarily. Since both $\widetilde{I}$ and $\overline{I}$ are nondecreasing functions, there are numbers $t_1, t_2 \in [0, M]$ such that

$$\{x \in [0, M] : \overline{I}(x) < q\} = [0, t_1) \text{ or } [0, t_1]$$

and

$$\{x \in [0, M] : q \le \widetilde{I}(x)\} = (t_2, M] \text{ or } [t_2, M]$$

By nonatomicity of ϕ, $\phi([0, t_1)) = \phi([0, t_1])$ and $\phi((t_2, M]) = \phi([t_2, M])$. Thus, since $\overline{I}$ and $\widetilde{I}$ are both ϕ-equimeasurable with I, we have

$$\phi([0, t_1)) = \phi([0, t_1]) = \phi\big(\{x \in [0, M] : I(x) < q\}\big)$$

and

$$\phi((t_2, M]) = \phi([t_2, M]) = \phi\big(\{x \in [0, M] : q \le I(x)\}\big)$$

Thus

$$\phi([0, t_1)) = \phi([0, t_1]) = 1 - \phi([t_2, M]) = \phi([0, t_2)) = \phi([0, t_2])$$

If $t_1 = t_2$, then

$$[0, t_1) \cap (t_2, M] = [0, t_1) \cap [t_2, M] = [0, t_1] \cap (t_2, M] = \varnothing$$

and

$$[0, t_1] \cap [t_2, M] = \{t_1\}$$

Thus,

$$\phi\big(\,[0, t_1) \cap (t_2, M]\,\big) = \phi\big(\,[0, t_1) \cap [t_2, M]\,\big) = \phi\big(\,[0, t_1] \cap (t_2, M]\,\big) = 0$$

and, by nonatomicity of ϕ,

$$\phi\Big(\,[0,t_1]\cap[t_2,M]\,\Big)=0$$

Therefore, $\phi\Big(\{x\in[0,M]:\overline{I}\,(x)<q\}\ \cap\ \{x\in[0,M]:q\leq\widetilde{I}\,(x)\}\Big)=0.$

If $t_1>t_2$, then

$$[0,t_1)=[0,t_2)\cup[t_2,t_1)=[0,t_2]\cup(t_2,t_1)$$

and

$$[0,t_1]=[0,t_2)\cup[t_2,t_1]=[0,t_2]\cup(t_2,t_1]$$

Since $\phi\,([0,t_1))=\phi\,([0,t_1])=\phi\,([0,t_2))=\phi\,([0,t_2])$, it then follows that

$$\phi\,((t_2,t_1))=\phi\,((t_2,t_1])=\phi\,([t_2,t_1))=\phi\,([t_2,t_1])=0$$

Therefore,

$$\begin{cases}\phi\Big(\,[0,t_1)\cap(t_2,M]\,\Big)=\phi\Big(\,(t_2,t_1)\,\Big)=0\\ \phi\Big(\,[0,t_1)\cap[t_2,M]\,\Big)=\phi\Big(\,[t_2,t_1)\,\Big)=0\\ \phi\Big(\,[0,t_1]\cap(t_2,M]\,\Big)=\phi\Big(\,(t_2,t_1]\,\Big)=0\\ \phi\Big(\,[0,t_1]\cap[t_2,M]\,\Big)=\phi\Big(\,[t_2,t_1]\,\Big)=0\end{cases}$$

Thus, $\phi\Big(\{x\in[0,M]:\overline{I}\,(x)<q\}\ \cap\ \{x\in[0,M]:q\leq\widetilde{I}\,(x)\}\Big)=0.$

Finally, if $t_2>t_1$, then

$$[0,t_2)=[0,t_1)\cup[t_1,t_2)=[0,t_1]\cup(t_1,t_2)$$

and

$$[0,t_2]=[0,t_1)\cup[t_1,t_2]=[0,t_1]\cup(t_1,t_2]$$

Since $\phi\,([0,t_2))=\phi\,([0,t_2])=\phi\,([0,t_1))=\phi\,([0,t_1])$, it then follows that

$$\phi\,((t_1,t_2))=\phi\,((t_1,t_2])=\phi\,([t_1,t_2))=\phi\,([t_1,t_2])=0$$

Therefore,

$$\begin{cases} \phi\big([0,t_2)\cap(t_1,M]\big)=\phi\big((t_1,t_2)\big)=0\\ \phi\big([0,t_2)\cap[t_1,M]\big)=\phi\big([t_1,t_2)\big)=0\\ \phi\big([0,t_2]\cap(t_1,M]\big)=\phi\big((t_1,t_2]\big)=0\\ \phi\big([0,t_2]\cap[t_1,M]\big)=\phi\big([t_1,t_2]\big)=0 \end{cases}$$

Thus, $\phi\Big(\{x\in[0,M]:\overline{I}(x)<q\}\ \cap\ \{x\in[0,M]:q\le\widetilde{I}(x)\}\Big)=0$. Since $q\in\mathbb{Q}$ was chosen arbitrarily, it then follows that

$$\phi\Big(\big\{x\in[0,M]:\overline{I}(x)<\widetilde{I}(x)\big\}\Big)=0$$

Similarly, we can show that $\phi\Big(\big\{x\in[0,M]:\widetilde{I}(x)<\overline{I}(x)\big\}\Big)=0$. Thus, $\widetilde{I}=\overline{I}$, ϕ-a.s.;

8. For each Borel-measurable and finite function $\psi:\mathbb{R}\to\mathbb{R}$ define the mapping $\delta(\psi):\mathbb{R}\to\mathbb{R}$ by

$$\delta(\psi)(t):=\inf\Big\{z\in\mathbb{R}:\phi\big(\{x\in\mathbb{R}:\psi(x)>z\}\big)\le\phi\big((-\infty,t)\big)\Big\},\ \forall t\in\mathbb{R}$$

Then, as in Epperson [120] (proposition 2 on p. 225), $\delta(\psi)$ is nonincreasing and ϕ-equimeasurable with ψ. Moreover, by Epperson [120] (proposition 2 on p. 225), if $\{f_n\}_n$ is a sequence of Borel-meaurable finite real-valued functions on $\mathbb{R}$ such that $f_n\uparrow f$, ϕ-a.s., where f is some Borel-meaurable finite real-valued functions on $\mathbb{R}$, then $\delta(f_n)\uparrow\delta(f)$.

Now, for each Borel-measurable and finite function $\psi:\mathbb{R}\to\mathbb{R}$ define the mapping

$$\iota(\psi):=-\delta(-\psi)$$

Then $\iota(\psi)$ is nondecreasing and ϕ-equimeasurable with ψ. Thus, by (7) above, for each Borel-measruable function $I:[0,M]\to[0,M]$, we have $\iota(I)=\widetilde{I}$, ϕ-a.s., that is, $\phi\Big(\big\{t\in[0,M]:\widetilde{I}(t)=\iota(I)(t)\big\}\Big)=1$. The rest then follows trivially (see also Proposition 3.60 on p. 98).

□

$\widetilde{I}$ will be called the nondecreasing ϕ-rearrangement of I (see also Epperson [120] pp. 224-225, Carlier and Dana [72] p. 876, Carlier and Dana [73] Proposition 1, p. 830, or

Carlier and Dana [74] Proposition 1 on p. 486). Now, define $Y := I \circ X$ and $\widetilde{Y} := \widetilde{I} \circ X$. Since both I and $\widetilde{I}$ are Borel-measurable mappings of $[0, M]$ into itself, it follows that $Y, \widetilde{Y} \in B^{+}(\Sigma)$. Note also that $\widetilde{Y}$ is nondecreasing in X, in the sense that if $s_1, s_2 \in S$ are such that $X(s_1) \leq X(s_2)$ then $\widetilde{Y}(s_1) \leq \widetilde{Y}(s_2)$, and that Y and $\widetilde{Y}$ are P-equimeasurable, that is, for any $\alpha \in [0, M]$, $P(\{s \in S : Y(s) \leq \alpha\}) = P\Big(\{s \in S : \widetilde{Y}(s) \leq \alpha\}\Big)$. Indeed,

$$\begin{aligned}
P\Big(s \in S : \widetilde{Y}(s) \leq \alpha\Big) &= P\Big(\{s \in S : X(s) \in \{t \in [0, M] : \widetilde{I}(t) \leq \alpha\}\}\Big) \\
&= \phi\Big(\{t \in [0, M] : \widetilde{I}(t) \leq \alpha\}\Big) \\
&= \phi\Big(\{t \in [0, M] : I(t) \leq \alpha\}\Big) \\
&= P\Big(\{s \in S : X(s) \in \{t \in [0, M] : I(t) \leq \alpha\}\}\Big) \\
&= P\Big(s \in S : Y(s) \leq \alpha\Big)
\end{aligned}$$

We will then call $\widetilde{Y}$ a **nondecreasing P-rearrangement of Y with respect to X**, and we shall denote it by $\widetilde{Y}_P$ to avoid confusion in case a different measure on $(S, \mathcal{G})$ is also considered. For example, in case both P_1 and P_2 are probability measures on the measurable space $(S, \mathcal{G})$, we shall denote by $\widetilde{Y}_{P_1}$ (resp. $\widetilde{Y}_{P_2}$) a nondecreasing P_1-rearrangement (resp. P_2-rearrangement) of Y with respect to X. In the general case, nothing can be said *a priori* about the relationship between $\widetilde{Y}_{P_1}$ and $\widetilde{Y}_{P_2}$. What can be asserted, however, is that:

1. Both $\widetilde{Y}_{P_1}$ and $\widetilde{Y}_{P_2}$ are nondecreasing in X, and hence $\widetilde{Y}_{P_1}$ and $\widetilde{Y}_{P_2}$ are comonotonic, i.e. $\Big[\widetilde{Y}_{P_2}(s) - \widetilde{Y}_{P_2}(s')\Big]\Big[\widetilde{Y}_{P_1}(s) - \widetilde{Y}_{P_1}(s')\Big] \geq 0$, for all $s, s' \in S$;

2. Y and $\widetilde{Y}_{P_1}$ are P_1-equimeasurable; and,

3. Y and $\widetilde{Y}_{P_2}$ are P_2-equimeasurable.

Note that $\widetilde{Y}_P$ is P-a.s. unique. Note also that if Y_1 and Y_2 are P-equimeasurable and if $Y_1 \in L_1(S, \mathcal{G}, P)$, then $Y_2 \in L_1(S, \mathcal{G}, P)$ and $\int \psi(Y_1)\ dP = \int \psi(Y_2)\ dP$, for any measurable function ψ such that the integrals exist.

Remark 2.10. *The previous construction of the nondecreasing rearrangement guarantees that the collection of all those functions $Y_1, Y_2 \in B^{+}(\Sigma)$ considered in Definition 2.4 is rich enough.*

Similarly to the previous construction, for a given a Borel-measurable $B \subseteq [0, M]$ with $\phi(B) > 0$, there exists a ϕ-a.s. unique (on B) nondecreasing, Borel-measurable mapping $\widetilde{I}_B : B \to [0, M]$ which is ϕ-equimeasurable with I on B, in the sense that for any $\alpha \in [0, M]$,

$$\phi\Big(\{t \in B : I(t) \leq \alpha\}\Big) = \phi\Big(\{t \in B : \widetilde{I}_B(t) \leq \alpha\}\Big) \tag{2.16}$$

$\widetilde{I}_B$ is called the nondecreasing ϕ-rearrangement of I on B[8]. Since X is $\mathcal{G}$-measurable, there exists $A \in \mathcal{G}$ such that $A = X^{-1}(B)$, and hence $P(A) > 0$. Now, define $\widetilde{Y}_A := \widetilde{I}_B \circ X$. Since both I and $\widetilde{I}_B$ are bounded Borel-measurable mappings, it follows that $Y, \widetilde{Y}_A \in B^+(\Sigma)$. Note also that $\widetilde{Y}_A$ is nondecreasing in X on A, in the sense that if $s_1, s_2 \in A$ are such that $X(s_1) \leq X(s_2)$ then $\widetilde{Y}(s_1) \leq \widetilde{Y}(s_2)$, and that Y and $\widetilde{Y}_A$ are P-equimeasurable on A, that is, for any $\alpha \in [0, M]$, $P(\{s \in S : Y(s) \leq \alpha\} \cap A) = P\Big(\{s \in S : \widetilde{Y}_A(s) \leq \alpha\} \cap A\Big)$. Indeed,

$$\begin{aligned} P\Big(s \in A : \widetilde{Y}_A(s) \leq \alpha\Big) = \phi\Big(\{t \in B : \widetilde{I}_B(t) \leq \alpha\}\Big) &= \phi\Big(\{t \in B : I(t) \leq \alpha\}\Big) \\ &= P\Big(s \in A : Y(s) \leq \alpha\Big) \end{aligned}$$

We will then call $\widetilde{Y}_A$ a **nondecreasing P-rearrangement of Y with respect to X on** A, and we shall denote it by $\widetilde{Y}_{A,P}$ to avoid confusion in case a different measure on $(S, \mathcal{G})$ is also considered. For example, in case both P_1 and P_2 are probability measures on the measurable space $(S, \mathcal{G})$, we shall denote by $\widetilde{Y}_{A,P_1}$ (resp. $\widetilde{Y}_{A,P_2}$) a nondecreasing P_1-rearrangement (resp. P_2-rearrangement) of Y with respect to X on A. In the general case, nothing can be said *a priori* about the relationship between $\widetilde{Y}_{A,P_1}$ and $\widetilde{Y}_{A,P_2}$. What can be asserted, however, is that:

1. Both $\widetilde{Y}_{A,P_1}$ and $\widetilde{Y}_{A,P_2}$ are nondecreasing in X on A, and hence $\widetilde{Y}_{P_1}$ and $\widetilde{Y}_{P_2}$ are comonotonic on A, i.e. $\Big[\widetilde{Y}_{A,P_2}(s) - \widetilde{Y}_{A,P_2}(s')\Big]\Big[\widetilde{Y}_{A,P_1}(s) - \widetilde{Y}_{A,P_1}(s')\Big] \geq 0$, for all $s, s' \in A$;

2. Y and $\widetilde{Y}_{A,P_1}$ are P_1-equimeasurable on A; and,

3. Y and $\widetilde{Y}_{A,P_2}$ are P_2-equimeasurable on A.

[8]See Carlier and Dana [72], p. 876.

Note that $\widetilde{Y}_{A,P}$ is P-a.s. unique. Note also that if $Y_{1,A}$ and $Y_{2,A}$ are P-equimeasurable on A and if $\int_A Y_{1,A}\ dP < +\infty$, then $\int_A Y_{2,A}\ dP < +\infty$ and $\int_A \psi\left(Y_{1,A}\right)\ dP = \int_A \psi\left(Y_{2,A}\right)\ dP$, for any measurable function ψ such that the integrals exist.

Lemma 2.11. *Let $Y \in B^+(\Sigma)$ and let $A \in \mathcal{G}$ be such that $P(A) = 1$ and $X(A)$ is a Borel set*[9]*. Let $\widetilde{Y}_P$ be the nondecreasing P-rearrangement of Y with respect to X, and let $\widetilde{Y}_{A,P}$ be the nondecreasing P-rearrangement of Y with respect to X on A. Then $\widetilde{Y}_P = \widetilde{Y}_{A,P}$, P-a.s.*

Proof. Since $P(A) = 1$, we have $P(S \setminus A) = 0$, and so it follows that for all $t \in \mathbb{R}$,

$$
\begin{aligned}
P\left[\{\widetilde{Y}_{A,P} \geq t\} \cap A\right] &= P\left[\{Y \geq t\} \cap A\right] \text{ by definition of } \widetilde{Y}_{A,P} \\
&= P\left[\{Y \geq t\} \cap A\right] + P\left[\{Y \geq t\} \cap (S \setminus A)\right] \text{ since } P(S \setminus A) = 0 \\
&= P\left[\{Y \geq t\}\right] = P\left[\{\widetilde{Y}_P \geq t\}\right] \text{ by definition of } \widetilde{Y}_P \\
&= P\left[\{\widetilde{Y}_P \geq t\} \cap A\right] + P\left[\{\widetilde{Y}_P \geq t\} \cap (S \setminus A)\right] \\
&= P\left[\{\widetilde{Y}_P \geq t\} \cap A\right] \text{ since } P(S \setminus A) = 0 \\
&= P\left[\{\widetilde{Y}_P \mathbf{1}_A \geq t\} \cap A\right]
\end{aligned}
$$

Furthermore, both $\widetilde{Y}_{A,P}$ and $\widetilde{Y}_P$ are nondecreasing in X on A. Hence, by the P-a.s. uniqueness of the nondecreasing P-rearrangement of Y with respect to X on A, it follows that $\widetilde{Y}_P = \widetilde{Y}_{A,P}$, P-a.s. on A, that is, P-a.s. □

2.3.2 Supermodularity and Hardy-Littlewood-Pólya Inequalities

A *partially ordered set* (*poset*) is a pair $(T, \succcurlyeq)$ where $\succcurlyeq$ is a reflexive, transitive and antisymmetric binary relation[10] on T. A point $t \in T$ is called an *upper bound* (resp. *lower bound*) for a subset S of T if $t \succcurlyeq x$ (resp. $x \succcurlyeq t$) for each $x \in S$. A point $t^* \in T$ is called

[9]Note that if $A \in \Sigma = \sigma\{X\}$ then $X(A)$ is automatically a Borel set, by definition of $\sigma\{X\}$. Indeed, for any $A \in \sigma\{X\}$, there is some Borel set B such that $A = X^{-1}(B)$. Then $X(A) = B \cap X(S)$ (see, e.g. Dieudonné [103], p. 7). Thus $X(A) = B \cap [0, M]$ is a Borel subset of $[0, M]$.

[10]See Appendix A.1.1 for more about binary relations.

a *least upper bound* (resp. *greatest lower bound*) for S if it is an upper bound (resp. lower bound) for S and for any other upper bound (resp. lower bound) t of S we have $t \succcurlyeq t^*$ (resp. $t^* \succcurlyeq t$). It is easily seen that the least upper bound and the greatest lower bound are unique.

For any $x, y \in S$ we denote by $x \vee y$ (resp. $x \wedge y$) the least upper bound (resp. greatest lower bound) of the set $\{x, y\}$. A poset $(T, \succcurlyeq)$ is called a *lattice* when $x \vee y, x \wedge y \in T$, for each $x, y \in T$.

For instance, the Euclidian space $\mathbb{R}^n$ is a lattice for the partial order $\succcurlyeq$ defined as follows: for $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, we write $x \succcurlyeq y$ when $x_i \geq y_i$, for each $i = 1, \dots, n$. It is then easy to see that $x \vee y = (\max(x_1, y_1), \dots, \max(x_n, y_n))$ and $x \wedge y = (\min(x_1, y_1), \dots, \min(x_n, y_n))$.

Definition 2.12. *Let $(T, \succcurlyeq)$ be a lattice. A function $L : T \to \mathbb{R}$ is said to be supermodular if for each $x, y \in T$,*

$$L(x \vee y) + L(x \wedge y) \geq L(x) + L(y) \tag{2.17}$$

In particular, a function $L : \mathbb{R}^2 \to \mathbb{R}$ is supermodular if for any $x_1, x_2, y_1, y_2 \in \mathbb{R}$ with $x_1 \leq x_2$ and $y_1 \leq y_2$, we have

$$L(x_2, y_2) + L(x_1, y_1) \geq L(x_1, y_2) + L(x_2, y_1) \tag{2.18}$$

Lemma 2.13. *A function $L : \mathbb{R}^2 \to \mathbb{R}$ is supermodular if and only if the function $\eta(y) := L(x + h, y) - L(x, y)$ is nondecreasing on $\mathbb{R}$, for any $x \in \mathbb{R}$ and $h \geq 0$.*

Proof. Immediate consequence of (2.18). □

Example 2.14. *The following are useful examples of supermodular functions:*

1. *If $g : \mathbb{R} \to \mathbb{R}$ is concave, and $a \in \mathbb{R}$, then the function $L_1 : \mathbb{R}^2 \to \mathbb{R}$ defined by $L_1(x, y) = g(a - x + y)$ is supermodular.*

2. *If* $f : \mathbb{R} \to \mathbb{R}$ *is concave, and* $a \in \mathbb{R}$*, then the function* $L_2 : \mathbb{R}^2 \to \mathbb{R}$ *defined by* $L_2(x,y) = f(a + x - y)$ *is supermodular.*

3. *The function* $L_3 : \mathbb{R}^2 \to \mathbb{R}$ *defined by* $L_3(x,y) = -(y-x)^+$ *is supermodular.*

4. *If* $\psi : \mathbb{R} \to \mathbb{R}$ *is nonincreasing function, then the function* $L_4 : \mathbb{R}^2 \to \mathbb{R}$ *defined by* $L_4(x,y) = -y\ \psi(x)$ *is supermodular.*

5. *If* $\zeta : \mathbb{R} \to \mathbb{R}^+$ *is a nondecreasing function,* $g : \mathbb{R} \to \mathbb{R}$ *is concave and nondecreasing, and* $a \in \mathbb{R}$*, then the function* $L_5 : \mathbb{R}^2 \to \mathbb{R}$ *defined by* $L_5(x,y) = g(a - x + y)\ \ \zeta(x)$ *is supermodular.*

6. *If* $\eta : \mathbb{R} \to \mathbb{R}^+$ *is a nonincreasing function,* $f : \mathbb{R} \to \mathbb{R}$ *is concave and nondecreasing, and* $a \in \mathbb{R}$*, then the function* $L_6 : \mathbb{R}^2 \to \mathbb{R}$ *defined by* $L_6(x,y) = f(a + x - y)\ \ \eta(x)$ *is supermodular.*

Lemma 2.15 (Hardy-Littlewood-Pólya Inequalities)**.** *Let* $Y \in B^+(\Sigma)$ *and let* $A \in \mathcal{G}$ *be such that* $P(A) > 0$ *and* $X(A)$ *is a Borel set. Let* $\widetilde{Y}_P$ *be the nondecreasing* P*-rearrangement of* Y *with respect to* X*, and let* $\widetilde{Y}_{A,P}$ *be the nondecreasing* P*-rearrangement of* Y *with respect to* X *on* A*. If* L *is supermodular then:*

1. $\int L\Big(X,Y\Big)\ dP \leq \int L\left(X,\widetilde{Y}_P\right)\ dP$*, and,*

2. $\int_A L\Big(X,Y\Big)\ dP \leq \int_A L\left(X,\widetilde{Y}_{A,P}\right)\ dP$*,*

provided the integrals exist (i.e. they are not of the form $\infty - \infty$*).*

Proof. By the P-a.s. uniqueness of the nondecreasing P-rearrangement of $Y \in B^+(\Sigma)$ with respect to X, it suffices to show that (1) holds for a pair $\left(Y,\overline{Y}\right) \in B^+(\Sigma) \times B^+(\Sigma)$ such that Y and $\overline{Y}$ are P-equimeasurable and $\overline{Y}$ is a nondecreasing function of X, and that (2) holds for a pair $\left(Y,\overline{Y}_A\right)$ such that Y and $\overline{Y}_A$ are P-equimeasurable on A and $\overline{Y}_A$ is a nondecreasing function of X on A. But these are classical results, and we refer the reader to Cambanis et al. [68], Carlier and Dana [74], and especially Dana and Scarsini [88], whose Lemma 3.4 on p. 158 is identical to our Lemma.

For a more formal treatment of this subject, within the larger framework of the Monge-Kantorovich optimal transport problems, see Rachev and Rüschendorf [230] (Chap. 3) and [231] (Chap. 7), and Villani [294] and [295]. □

Lemma 2.16. *Let $Y \in B^{+}(\Sigma)$ and let $A \in \mathcal{G}$ be such that $P(A) > 0$ and $X(A)$ is a Borel set. Let $\widetilde{Y}_P$ be the nondecreasing P-rearrangement of Y with respect to X, and let $\widetilde{Y}_{A,P}$ be the nondecreasing P-rearrangement of Y with respect to X on A. Then the following hold:*

1. *If $0 \leq Y \leq X$, P-a.s., then $0 \leq \widetilde{Y}_P \leq X$; and,*

2. *If $0 \leq Y \leq X$, P-a.s. on A, then $0 \leq \widetilde{Y}_{A,P} \leq X$, P-a.s. on A.*

Proof. Since Y is $\sigma\{X\}$-measurable, by Doob's measurability theorem there is a real-valued bounded Borel-measurable function I on $[0, M]$ such that $Y = I \circ X$. Moreover, we can write $X = Id \circ X$, where Id denotes the identity map on $[0, M]$.

If $0 \leq Y \leq X$, P-a.s. then $0 \leq I \leq Id$, ϕ-a.s. Therefore, by Proposition[11] 2.9, $0 \leq \widetilde{I} \leq \widetilde{Id} \leq Id$, where $\widetilde{I}$ denotes the nondecreasing ϕ-rearrangement of I and where $\widetilde{Id}$ denotes the nondecreasing ϕ-rearrangement of Id. Hence, $0 \leq \widetilde{Y}_P \leq X$.

Now, suppose that $0 \leq Y \leq X$, P-a.s. on A. To show that $0 \leq \widetilde{Y}_{A,P} \leq X$, P-a.s. on A, let $L(X,Y) := -(Y - X)^{+}$ and let $\Psi := L(X,Y)$. Then the function L is supermodular, and the functions $L(t, I(t))$ and $L\left(t, \widetilde{I}(t)\right)$ are bounded, and hence ϕ-integrable functions on $[0, M]$. Thus, Ψ is P-integrable[12]. Let $\widetilde{\Psi} := L\left(X, \widetilde{Y}_{A,P}\right)$. Then similarly, $\widetilde{\Psi}$ is P-integrable. Since $0 \leq Y \leq X$, P-a.s. on A, it follows that $\int_A \Psi \, dP = 0$. Moreover, since L is supermodular, Lemma 2.15 yields:

$$0 = -\int_A \Psi \, dP \geq -\int_A \widetilde{\Psi} \, dP = \int_A \left(\widetilde{Y}_{A,P} - X\right)^{+} \, dP \geq 0$$

Therefore, $\int_A \left(\widetilde{Y}_{A,P} - X\right)^{+} \, dP = 0$. Since $\left(\widetilde{Y}_{A,P} - X\right)^{+} \geq 0$ and $P(A) > 0$, it then follows form Lemma 2.49 that $\widetilde{Y}_{A,P} \leq X$, P-a.s. on A. The fact that $\widetilde{Y}_{A,P} \geq 0$, P-a.s. on A follows from the definition of $\widetilde{Y}_{A,P}$ and the fact that $Y \geq 0$, P-a.s. on A. □

[11] Chong and Rice [81] (Proposition 4.3 (ii) on p. 30) provide a similar result for the case where ϕ is the Lebesgue measure.

[12] See Theorem E.11 on p. 241.

2.3.3 Some "Convergence" Results

Lemma 2.17. *If f and f_n are $[0, +\infty)$-valued, Σ-measurable functions on S such that the sequence $\{f_n\}_n$ converges pointwise P-a.s. to f monotonically downwards, then the sequence $\{\widetilde{f}_{n,P}\}_n$ converges pointwise P-a.s. to $\widetilde{f}_P$ monotonically downwards, where $\widetilde{f}_P$ is the nondecreasing P-rearrangement of f with respect to X, and $\widetilde{f}_{n,P}$ is the nondecreasing P-rearrangement of f_n with respect to X, for each $n \in \mathbb{N}$.*

Proof. Follows from Proposition 2.9 (8) and Doob's measurability theorem. Similar results can be found in Epperson [120] (Proposition 2 (ii) on p. 225) and Chong [80] (p. 142), for the specific formulation of the monotone equimeasurable rearrangement that the authors use. □

Lemma 2.18. *Let f and f_n be $[0, +\infty)$-valued, Σ-measurable functions on S. If $f_n \in B^+(\Sigma)$, for each $n \geq 1$, and if the sequence $\{f_n\}_n$ converges uniformly to $f \in B^+(\Sigma)$, then*

1. *The functions $\widetilde{f}_P$ and $\widetilde{f}_{n,P}$ are in L_∞, for each $n \geq 1$, where $\widetilde{f}_P$ is the nondecreasing P-rearrangement of f with respect to X, and $\widetilde{f}_{n,P}$ is the nondecreasing P-rearrangement of f_n with respect to X, for each $n \in \mathbb{N}$; and,*

2. *The sequence $\{\widetilde{f}_{n,P}\}_n$ converges to $\widetilde{f}_P$ in the L_∞ norm.*

Proof. Since f, f_n are uniformly bounded, they are essentially bounded (i.e. they belong to L_∞) and so, by Lemma 2.16, the functions $\widetilde{f}_P$ and $\widetilde{f}_{n,P}$ are essentially bounded, for each $n \geq 1$.

Since $f, f_n \in B^+(\Sigma)$ for each $n \geq 1$, by Doob's measurability theorem there are real-valued Borel-measurable functions I and I_n on $[0, M]$ such that $f = I \circ X$ and $f_n = I_n \circ X$, for each $n \in \mathbb{N}$. Moreover, the functions I, I_n are uniformly bounded since the functions f, f_n are uniformly bounded (by hypothesis), and since X is uniformly bounded (by hypothesis). Also, the sequence $\{I_n\}_n$ converges uniformly to I, on the range of X, i.e. on the interval $[0, M]$, and hence convergence is also in the sense of the L_∞ norm. Let $\widetilde{I}$ be the nondecreasing ϕ-rearrangement of I, and let $\widetilde{I}_n$ be the nondecreasing ϕ-rearrangement of I_n, for each $n \in \mathbb{N}$. Then the sequence $\{\widetilde{I}_n\}_n$ converges to $\widetilde{I}$ in the L_∞ norm (see, e.g. Rakotoson [233], Corollary 1.3.3 on p. 17, which shows that the monotone rearrangement is a norm-continuous operator on L_p, for each $p \in [1, +\infty]$, and hence in particular for the space L_∞. Horsley and Wrobel [168] also discuss various ways in which the nondecreasing rearrangement operator is continuous). The rest then follows trivially. □

2.3.4 Another Characterization of the Nondecrasing Rearrangement

Consider the nonatomic Borel probability measure ϕ on the interval $[0, M]$ introduced in section 2.3.1. The next proposition gives another characterization of the nondecreasing ϕ-rearrangement of any Borel-measurable function $I : [0, M] \to [0, M]$.

Proposition 2.19. *Let $I : [0, M] \to [0, M]$ be any Borel-measurable map, and let $\widetilde{I} : [0, M] \to [0, M]$ denote the nondecreasing ϕ-rearrangement of I. Then for ϕ-a.e. $t \in [0, M]$,*

$$\widetilde{I}(t) = \eta(t) \tag{2.19}$$

where,

$$\eta(t) := \sup \Big\{ u \in \mathbb{R} : \xi(u) \leq t \Big\} \tag{2.20}$$

and

$$\xi(u) := \inf \left\{ \beta \in [0, M] : \phi\Big([\beta, M] \Big) = \phi\Big(\{ s \in [0, M] : I(s) \geq u \} \Big) \right\} \tag{2.21}$$

Proof. By the ϕ-a.s. uniqueness of the nondecreasing ϕ-rearrangement of I (Proposition 2.9 (7)), it suffices to show that the map η is nondecreasing and ϕ-equimeasurable with I.

First note that ξ is nondecreasing by monotonicity of the measure ϕ, and so η is also nondecreasing. To show that η and I are ϕ-equimeasurable, note that for any u and t,

$$u > \eta(t) \iff \xi(u) > t$$

Therefore, for any $u \in \mathbb{R}$,

$$\phi\Big(\{ s \in [0, M] : \eta(s) < u \} \Big) = \phi\Big(\{ s \in [0, M] : \xi(u) > s \} \Big) = \phi\Big([0, \xi(u)) \Big)$$

Now, since $\phi\Big([0, M]\Big) = 1$, it follows that

$$\xi(u) = \inf\left\{\beta \in [0, M] : \phi\Big([0, \beta)\Big) = \phi\Big(\{s \in [0, M] : I(s) < u\}\Big)\right\}$$

Therefore, $\phi\Big([0, \xi(u))\Big) = \phi\Big(\{s \in [0, M] : I(s) < u\}\Big)$, which completes the proof. $\square$

Simple functions

Any Σ-simple function[13] $Y \in B^{+}(\Sigma)$ can be written as $Y = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{C_i}$, for some $\{\alpha_i\}_{i=1}^{n} \subset \mathbb{R}^{+}$ and a partition $\{C_i\}_{i=1}^{n}$ of S, where $C_i \in \Sigma$, for each $i \in \{1, \ldots, n\}$. Since $C_i \in \Sigma$, for each $i \in \{1, \ldots, n\}$, and since $\Sigma = \sigma\{X\}$, it follows that

$$Y(s) = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{B_i}\big(X(s)\big), \ \forall s \in S \tag{2.22}$$

where B_i is a Borel subset of $X(S) = [0, M]$, for each $i \in \{1, \ldots, n\}$, and $\{B_i\}_{i}^{n}$ is a partition of $[0, M]$. in other words, $Y = I \circ X$, where the function I is a simple function on $[0, M]$ of the form

$$I = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{B_i} \tag{2.23}$$

Since the nondecreasing rearrangement $\widetilde{Y}$ of Y with respect to X is simply $\widetilde{I} \circ X$, where $\widetilde{I}$ is the nondecreasing ϕ-rearrangement of I (recall that $\phi = P \circ X^{-1}$), it suffices to characterize $\widetilde{I}$. This is done in the following proposition.

Proposition 2.20. *Let $I = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{B_i}$ be any Borel-measurable simple function on $[0, M]$, and suppose, without loss of generality, that $\alpha_1 < \alpha_2 < \ldots < \alpha_n$. If $\widetilde{I}$ denotes the*

[13]Note that the collection of all Σ-simple functions on (S, Σ) (i.e. the collection of all finite linear combinations of indicator functions of sets in Σ) is supnorm dense in the supnorm-normed Banach space $B(\Sigma)$ of all bounded, $\mathbb{R}$-valued and Σ-measurable functions on S (Hewitt and Stromberg [166], Theorem 11.35 on p. 159), where as before $\Sigma = \sigma\{X\}$.

nondecreasing ϕ*-rearrangement of* I*, then* $\widetilde{I} = \eta,\ \phi$*-a.s., where* η *is given by*

$$\eta\left(t\right) := \sum_{i=1}^{n-1} \alpha_i \mathbf{1}_{[\beta_{i-1},\beta_i)}\left(t\right) + \alpha_n \mathbf{1}_{[\beta_{n-1},M]} \tag{2.24}$$

where:

1. $\beta_0 := 0$

2. $\beta_i := \inf\left\{\beta \in \left[0,M\right] : \phi\Big(\left[\beta,M\right]\Big) = \phi\Big(\left\{t \in \left[0,M\right] : I\left(t\right) \geq \alpha_{i+1}\right\}\Big)\right\}$*, for each* $i \in \{1,\ldots,n-1\}$.

Proof. To show that $\eta = \widetilde{I}$ ϕ-a.s., it suffices to show that η is nondecreasing and ϕ-equimeasurable with I. The fact that η is nondecreasing follows directly from (2.24). To show that I and η are ϕ-equimeasurable, fix any $i_0 \in \{1,\ldots,n\}$. Then

$$\phi\Big(\left\{t \in \left[0,M\right] : \eta\left(t\right) \geq \alpha_{i_0}\right\}\Big) = \phi\Big(\left[\beta_{i_0-1},M\right]\Big)$$

However, by definition of β_{i_0-1}, $\phi\Big(\left[\beta_{i_0-1},M\right]\Big) = \phi\Big(\left\{t \in \left[0,M\right] : I\left(t\right) \geq \alpha_{i_0}\right\}\Big)$. Since i_0 was chosen arbitrarily in $\{1,\ldots,n\}$, this completes the proof. $\square$

2.4 The DM's Demand for Contingent Claims

2.4.1 The DM's problem

The problem of designing the optimal contract can be seen as that of finding the claim that will maximize the expected utility of the DM's wealth, under her subjective probability measure, subject to the CI's participation constraint and to some constraints on the claim. Specifically, the DM's problem is the following:

Problem 2.21.

$$\sup_{Y \in B^{+}(\Sigma)} \left\{ \int u\Big(W_0 - \Pi - X + Y\Big)\ d\mu \right\} :$$
$$\begin{cases} 0 \leq Y \leq X \\ \int v\Big(W_0^{CI} + \Pi - Y\Big)\ d\nu \geq v\left(W_0^{CI}\right) \end{cases}$$

In an insurance framework, the first constraint is standard (see Arrow [25] and Raviv [239]), and says that an indemnity is nonnegative and cannot exceed the loss itself. The second constraint is simply the CI's *participation constraint*, or *individual rationality constraint*. We will discuss these constraint further in the next chapter.

Remark 2.22. *Assuming Problem 2.21 has a nonempty feasibility set, the supremum value of Problem 2.21 is finite since the utility function u is bounded. That is, there exists some $R < +\infty$ such that $u\Big(W_0 - \Pi - X(s) + Y(s)\Big) \leq R$, for each $s \in S$ and for each $Y \in B^{+}(\Sigma)$. Consequently, $\int_D u\Big(W_0 - \Pi - X + Y\Big)\ d\mu \leq R\mu(D)$, for each $D \in \Sigma$ and for each $Y \in B^{+}(\Sigma)$.*

2.4.2 Existence of a Monotone Solution and Pareto-Improving Claims

Here we will give a sufficient condition for Problem 2.21 to admit a solution which is a nondecreasing function of X.

Definition 2.23. *Let $\mathcal{F}_{SB}$ be defined by*

$$\mathcal{F}_{SB} := \left\{ Y \in B^{+}(\Sigma) : 0 \leq Y \leq X \ \textit{and} \ \int v\Big(W_0^{CI} + \Pi - Y\Big)\ d\nu \geq v\left(W_0^{CI}\right) \right\}$$

That is, $\mathcal{F}_{SB}$ is the feasibility set for Problem 2.21. In the following, we will assume that this feasibility set is nonempty:

Assumption 2.24. $\mathcal{F}_{SB} \neq \varnothing$.

Let $\mathcal{F}_{SB}^{\uparrow}$ denote the collection of all feasible $Y \in B^{+}(\Sigma)$ for Problem 2.21 which are also nondecreasing in X, i.e. of the form $Y = I \circ X$ where $I : [0, M] \to [0, M]$ is nondecreasing:

Definition 2.25. *Let* $\mathcal{F}_{SB}^{\uparrow} := \Big\{ Y = I \circ X \in \mathcal{F}_{SB} : I \text{ is nondecreasing} \Big\}$.

Lemma 2.26. *If* ν *is* (μ, X)*-vigilant, then* $\mathcal{F}_{SB}^{\uparrow} \neq \varnothing$.

Proof. By Assumption 2.24, $\mathcal{F}_{SB} \neq \varnothing$. Choose any $Y = I \circ X \in \mathcal{F}_{SB}$, and let $\widetilde{Y}_{\mu}$ denote the nondecreasing μ-rearrangement of Y with respect to X. Then (i) $\widetilde{Y}_{\mu} = \widetilde{I} \circ X$ where $\widetilde{I}$ is nondecreasing, and (ii) $0 \leq \widetilde{Y}_{\mu} \leq X$, by Lemma 2.16.

Furthermore, since ν is (μ, X)-vigilant, it follows from the definition of vigilance that

$$\int v\Big(W_0^{CI} + \Pi - \widetilde{Y}_{\mu}\Big)\, d\nu \geq \int v\Big(W_0^{CI} + \Pi - Y\Big)\, d\nu$$

However, $\int v\Big(W_0^{CI} + \Pi - Y\Big)\, d\nu \geq v\left(W_0^{CI}\right)$ since $Y \in \mathcal{F}_{SB}$. Therefore, $\int v\Big(W_0^{CI} + \Pi - \widetilde{Y}_{\mu}\Big)\, d\nu \geq v\left(W_0^{CI}\right)$. Thus, $\widetilde{Y}_{\mu} \in \mathcal{F}_{SB}^{\uparrow}$, and so $\mathcal{F}_{SB}^{\uparrow} \neq \varnothing$. □

Definition 2.27. *If* $Y_1, Y_2 \in \mathcal{F}_{SB}$, *we will say that* Y_2 *is a Pareto improvement of* Y_1 *(or is Pareto-improving) when the following hold:*

1. $\int u\Big(W_0 - \Pi - X + Y_2\Big)\, d\mu \geq \int u\Big(W_0 - \Pi - X + Y_1\Big)\, d\mu$*; and,*

2. $\int v\Big(W_0^{CI} + \Pi - Y_2\Big)\, d\nu \geq \int v\Big(W_0^{CI} + \Pi - Y_1\Big)\, d\nu$.

Lemma 2.28. *Suppose that ν is (μ, X)-vigilant and that $\mathcal{U}(X,Y) := u\Big(W_0 - \Pi - X + Y\Big)$ is supermodular*[14]. *If $Y \in \mathcal{F}_{SB}$, then there is some $Y^* \in \mathcal{F}_{SB}^{\uparrow}$ which is Pareto-improving.*

Proof. First note that by Lemma 2.26 $\mathcal{F}_{SB}^{\uparrow} \neq \varnothing$. Choose any $Y \in \mathcal{F}_{SB}$, and let $Y^* := \widetilde{Y}_\mu$, where $\widetilde{Y}_\mu$ denotes the nondecreasing μ-rearrangement of Y with respect to X. Then $Y^* \in \mathcal{F}_{SB}^{\uparrow}$, as in the proof of Lemma 2.26. Moreover, since $\mathcal{U}(X,Y)$ is supermodular, it follows from Lemma 2.15 that

$$\int u\Big(W_0 - \Pi - X + Y^*\Big)\, d\mu \geq \int u\Big(W_0 - \Pi - X + Y\Big)\, d\mu$$

Finally, since ν is (μ, X)-vigilant, it follows from the definition of (μ, X)-vigilance that

$$\int v\Big(W_0^{CI} + \Pi - Y^*\Big)\, d\nu \geq \int v\Big(W_0^{CI} + \Pi - Y\Big)\, d\nu$$

Therefore, $Y^* \in \mathcal{F}_{SB}^{\uparrow}$ is a Pareto improvement of $Y \in \mathcal{F}_{SB}$. □

Proposition 2.29. *If ν is (μ, X)-vigilant and $\mathcal{U}(X,Y) := u\Big(W_0 - \Pi - X + Y\Big)$ is supermodular (e.g. u is concave), then Problem 2.21 admits a solution which is a nondecreasing function of X.*

Proof. The main idea of this proof (the use of *maximizing sequences*) is rather standard. For instance, Dana and Scarsini [88] use a similar idea based on *maximizing sequences* to prove their Theorem 3.8 and Theorem 8.4. Also, Cardaliaguet and Tahraoui [70] use a similar idea based on *minimizing sequences* to show existence of solutions to some classes of Calculus of Variations problems (their Theorem 3.3).

By Lemma 2.28, we can choose a maximizing sequence $\{Y_n\}_n$ in $\mathcal{F}_{SB}^{\uparrow}$ for Problem 2.21. That is,

$$\lim_{n \to +\infty} \int u\Big(W_0 - \Pi - X + Y_n\Big)\, d\mu = N$$

[14]This happens for instance when the utility function u is concave, i.e. when the DM is risk-averse. See Example 2.14 (1) and (2).

where $N < +\infty$ is the supremum value of Problem 2.21. Since $0 \leq Y_n \leq X \leq M := \|X\|_s$, the sequence $\{Y_n\}_n$ is uniformly bounded. Moreover, for each $n \geq 1$ we have $Y_n = I_n \circ X$, with $I_n : [0, M] \rightarrow [0, M]$. Consequently, the sequence $\{I_n\}_n$ is a uniformly bounded sequence of nondecreasing Borel-measurable functions. Thus, by Lemma 2.50, there is a nondecreasing function $I^* : [0, M] \rightarrow [0, M]$ and a subsequence $\{I_m\}_m$ of $\{I_n\}_n$ such that $\{I_m\}_m$ converges pointwise on $[0, M]$ to I^*. Hence, I^* is also Borel-measurable, and so $Y^* := I^* \circ X \in B^+(\Sigma)$ is such that $0 \leq Y^* \leq X$. Moreover, the sequence $\{Y_m\}_m$, defined by $Y_m = I_m \circ X$, converges pointwise to Y^*. Thus, by continuity of the utility function v (Assumption 2.2), the sequence $\left\{v\Big(W_0^{CI} + \Pi - Y_m\Big)\right\}_m$ converges pointwise to $v\Big(W_0^{CI} + \Pi - Y^*\Big)$. Boundedness of v and Lebesgue's Dominated Convergence Theorem (Theorem E.9 on p. 241), hence give that

$$\lim_{m \rightarrow +\infty} \int v\Big(W_0^{CI} + \Pi - Y_m\Big)\ d\nu = \int v\Big(W_0^{CI} + \Pi - Y^*\Big)\ d\nu$$

Now, since $Y_m \in \mathcal{F}_{SB}$, for each $m \geq 1$, it follows that:

$$\int v\Big(W_0^{CI} + \Pi - Y_m\Big)\ d\nu \geq v\left(W_0^{CI}\right),\ \forall m \geq 1$$

Therefore, $Y^* \in \mathcal{F}_{SB}^{\uparrow}$.

Similarly, by continuity and boundedness of the utility function u (Assumption 2.2) and by Lebesgue's Dominated Convergence Theorem, we have

$$\begin{aligned}\int u\Big(W_0 - \Pi - X + Y^*\Big)\ d\mu &= \lim_{m \rightarrow +\infty} \int u\Big(W_0 - \Pi - X + Y_m\Big)\ d\mu \\ &= \lim_{n \rightarrow +\infty} \int u\Big(W_0 - \Pi - X + Y_n\Big)\ d\mu = N\end{aligned}$$

Hence Y^* solves Problem 2.21. □

2.4.3 Characterization of the Solution

By Lebesgue's decomposition theorem (Theorem D.10 on p. 226) there exists a unique pair (ν_{ac}, ν_s) of (nonnegative) finite measures on (S, Σ) such that $\nu = \nu_{ac} + \nu_s$, $\nu_{ac} << \mu$, and $\nu_s \perp \mu$. That is, for all $B \in \Sigma$ with $\mu(B) = 0$, we have $\nu_{ac}(B) = 0$, and there is some $A \in \Sigma$ such that $\mu(S \setminus A) = \nu_s(A) = 0$. It then also follows that $\nu_{ac}(S \setminus A) = 0$ and $\mu(A) = 1$. Hence, for all $Z \in B^+(\Sigma)$, $\int Z \, d\nu = \int_A Z \, d\nu_{ac} + \int_{S\setminus A} Z \, d\nu_s$. Moreover, by the Radon-Nikodým theorem (Theorem E.22 on p. 244) there exists a μ-a.s. unique Σ-measurable and μ-integrable function $h : S \to [0, +\infty)$ such that $\nu_{ac}(C) = \int_C h \, d\mu$, for all $C \in \Sigma$. Consequently, for all $Z \in B^+(\Sigma)$, $\int Z \, d\nu = \int_A Zh \, d\mu + \int_{S\setminus A} Z \, d\nu_s$. Furthermore, since $\nu_{ac}(S \setminus A) = 0$, it follows that $\int_{S\setminus A} Z \, d\nu_s = \int_{S\setminus A} Z \, d\nu$. Thus, for all $Z \in B^+(\Sigma)$, $\int Z \, d\nu = \int_A Zh \, d\mu + \int_{S\setminus A} Z \, d\nu$. In particular, $\int Y \, d\nu = \int_A Yh \, d\mu + \int_{S\setminus A} Y \, d\nu$.

In the following, the Σ-measurable set A on which μ is concentrated (and $\nu_s(A) = 0$) is assumed to be fixed all throughout. Moreover, since $A \in \Sigma$ and since $X(S) = [0, M]$, $X(A)$ is a Borel subset[15] of $[0, M]$.

Lemma 2.30. *Let Y^* be an optimal solution for Problem 2.21, and suppose that ν is (μ, X)-vigilant and that $\mathcal{U}(X, Y) := u\Big(W_0 - \Pi - X + Y\Big)$ is supermodular. Let $\widetilde{Y}^*_\mu$ be the nondecreasing μ-rearrangement of Y^* with respect to X. Then:*

1. *$\widetilde{Y}^*_\mu$ is optimal for Problem 2.21; and,*

2. *$\widetilde{Y}^*_\mu = \widetilde{Y}^*_{\mu,A}$, μ-a.s., where $\widetilde{Y}^*_{\mu,A}$ is the nondecreasing μ-rearrangement of Y^* with respect to X on A.*

Proof. Optimality of $\widetilde{Y}^*_\mu$ for Problem 2.21 is an immediate consequence of Lemma 2.28 and its proof.

Let $\widetilde{Y}^*_{\mu,A}$ be the nondecreasing μ-rearrangement of Y^* with respect to X on A. Since $\mu(A) = 1$, then by Lemma 2.11 we have that $\widetilde{Y}^*_\mu = \widetilde{Y}^*_{\mu,A}$, μ-a.s. □

[15] for any $A \in \Sigma = \sigma\{X\}$, there is some Borel set B such that $A = X^{-1}(B)$. Then $X(A) = B \cap X(S)$ (see, e.g. Dieudonné [103], p. 7). Thus $X(A) = B \cap [0, M]$ is a Borel subset of $[0, M]$.

Lemma 2.31. *Let an optimal solution for Problem 2.21 be given by:*

$$Y^* = Y_1^* \mathbf{1}_A + Y_2^* \mathbf{1}_{S \setminus A} \tag{2.25}$$

for some $Y_1^*, Y_2^* \in B^+(\Sigma)$. *Let* $\widetilde{Y}_\mu^*$ *be the nondecreasing* μ*-rearrangement of* Y^* *with respect to* X*, and let* $Y_{1,\mu}^*$ *be the nondecreasing* μ*-rearrangement of* Y_1^* *with respect to* X*. Then* $\widetilde{Y}_\mu^* = \widetilde{Y}_{1,\mu}^*$*,* μ*-a.s.*

Proof. Let $\widetilde{Y}_{\mu,A}^*$ be the nondecreasing μ-rearrangement of Y^* with respect to X on A. Since $\mu(A) = 1$, then by Lemma 2.11 we have $\widetilde{Y}_\mu^* = \widetilde{Y}_{\mu,A}^*$, μ-a.s.

Similarly, let $\widetilde{Y}_{1,\mu,A}^*$ be the nondecreasing μ-rearrangement of Y_1^* with respect to X on A. Then $\widetilde{Y}_{1,\mu}^* = \widetilde{Y}_{1,\mu,A}^*$, μ-a.s.

Therefore, it suffices to show that $\widetilde{Y}_{\mu,A}^* = \widetilde{Y}_{1,\mu,A}^*$, μ-a.s. Since both $\widetilde{Y}_{\mu,A}^*$ and $\widetilde{Y}_{1,\mu,A}^*$ are nondecreasing functions of X on A, then by the μ-a.s. uniqueness of the nondecreasing rearrangement, it remains to show that they are μ-equimeasurable with Y^* on A. Now, for each $t \in [0, M]$,

$$\begin{aligned} \mu\Big(\{s \in A : \widetilde{Y}_{\mu,A}^*(s) \le t\}\Big) &= \mu\Big(\{s \in A : Y^*(s) \le t\}\Big) = \mu\Big(\{s \in A : Y_1^*(s) \le t\}\Big) \\ &= \mu\Big(\{s \in A : \widetilde{Y}_{1,\mu,A}^*(s) \le t\}\Big) \end{aligned}$$

where the first equality follows from the definition of $\widetilde{Y}_{\mu,A}^*$ (equimeasurabiltiy), the second equality follows from equation (2.25), and the third equality follows from the definition of $\widetilde{Y}_{1,\mu,A}^*$ (equimeasurabiltiy).

□

2.4.4 Some Sufficient Problems

Consider now the following two problems[16]:

[16]The general "splitting" procedure that we use hereafter is inspired by a similar technique used by Jin and Zhou [175], albeit in a different setting and for different reasons.

Problem 2.32. *For a given $\beta \in \mathbb{R}$,*

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int_A u\Big(W_0 - \Pi - X + Y\Big)\, d\mu \right\} :$$
$$\begin{cases} 0 \leq Y\mathbf{1}_A \leq X\mathbf{1}_A \\ \int_A v\Big(W_0^{CI} + \Pi - Y\Big)\, d\nu = \beta \end{cases}$$

Problem 2.33.

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int_{S\backslash A} u\Big(W_0 - \Pi - X + Y\Big)\, d\mu \right\} :$$
$$\begin{cases} 0 \leq Y\mathbf{1}_{S\backslash A} \leq X\mathbf{1}_{S\backslash A} \\ \int_{S\backslash A} v\Big(W_0^{CI} + \Pi - Y\Big)\, d\nu \geq v\left(W_0^{CI}\right) - \beta, \text{ for the same } \beta \text{ as in Problem 2.32} \end{cases}$$

Remark 2.34. *By Remark 2.22, the supremum value of each of the above two problems is finite when their feasibility sets are nonempty.*

Definition 2.35. *For a given $\beta \in \mathbb{R}$, let:*

1. *$\Theta_{A,\beta}$ be the feasibility set of Problem 2.32 with parameter β. That is,*

$$\Theta_{A,\beta} := \Bigg\{ Y \in B^+(\Sigma) : 0 \leq Y\ \mathbf{1}_A \leq X\mathbf{1}_A, \text{ and } \int_A v\Big(W_0^{CI} + \Pi - Y\Big)\, d\nu = \beta \Bigg\}$$

2. *$\Theta_{S\backslash A,\beta}$ be the feasibility set of Problem 2.33 with parameter β. That is,*

$$\Theta_{S\backslash A,\beta} := \Bigg\{ Y \in B^+(\Sigma) : 0 \leq Y\mathbf{1}_{S\backslash A} \leq X\mathbf{1}_{S\backslash A}, \text{ and } \int_{S\backslash A} v\Big(W_0^{CI} + \Pi - Y\Big)\, d\nu \geq v\left(W_0^{CI}\right) - \beta \Bigg\}$$

Denote by Γ the collection of all β for which the feasibility sets $\Theta_{A,\beta}$ and $\Theta_{S\setminus A,\beta}$ are nonempty:

Definition 2.36. *Let* $\Gamma := \left\{ \beta \in \mathbb{R} : \Theta_{A,\beta} \neq \varnothing, \;\Theta_{S\setminus A,\beta} \neq \varnothing \right\}$

Lemma 2.37. $\Gamma \neq \varnothing$.

Proof. By Assumption 2.24, there is some $Y \in B^+(\Sigma)$ such that $0 \leq Y \leq X$, and $\int v\Big(W_0^{CI} + \Pi - Y\Big)\, d\nu \geq v\left(W_0^{CI}\right)$. Let $\beta_Y := \int_A v\Big(W_0^{CI} + \Pi - Y\Big)\, d\nu$. Then, by definition of β_Y, and since $0 \leq Y \leq X$, we have $Y \in \Theta_{A,\beta_Y} \cap \Theta_{S\setminus A,\beta_Y}$, and so $\Theta_{A,\beta_Y} \neq \varnothing$ and $\Theta_{S\setminus A,\beta_Y} \neq \varnothing$. Consequently, $\beta_Y \in \Gamma$, and so $\Gamma \neq \varnothing$. □

Now, consider the following problem:

Problem 2.38.

$$\sup_{\beta} \left\{ F_A^*(\beta) + F_A^*\left(v\left(W_0^{CI}\right) - \beta\right) : \beta \in \Gamma \right\} :$$

$$\begin{cases} F_A^*(\beta) \text{ is the supremum value of Problem 2.32, for a fixed } \beta \in \Gamma \\ F_A^*\left(v\left(W_0^{CI}\right) - \beta\right) \text{ is the supremum value of Problem 2.33, for the same fixed } \beta \in \Gamma \end{cases}$$

Lemma 2.39. *If β^* is optimal for Problem 2.38, Y_3^* is optimal for Problem 2.32 with parameter β^*, and Y_4^* is optimal for Problem 2.33 with parameter β^*, then $Y_2^* := Y_3^*\mathbf{1}_A + Y_4^*\mathbf{1}_{S\setminus A}$ is optimal for Problem 2.21.*

Proof. Feasibility of Y_2^* for Problem 2.21 is immediate. To show optimality of Y_2^* for Problem 2.21, let $\widetilde{Y}$ be any other feasible function for Problem 2.21, and define $\alpha := \int_A v\Big(W_0^{CI} + \Pi - \widetilde{Y}\Big)\, d\nu$. Then α is feasible for Problem 2.38, and $\widetilde{Y}\mathbf{1}_A$ (resp. $\widetilde{Y}\mathbf{1}_{S\setminus A}$) is feasible for Problem 2.32 (resp. Problem 2.33) with parameter α. Hence

$$\begin{cases} F_A^*(\alpha) \geq \int_A u\Big(W_0 - \Pi - X + \widetilde{Y}\Big)\, d\mu \\ F_A^*\left(v\left(W_0^{CI}\right) - \alpha\right) \geq \int_{S\setminus A} u\Big(W_0 - \Pi - X + \widetilde{Y}\Big)\, d\mu \end{cases}$$

Now, since β^* is optimal for Problem 2.38, it follows that

$$F_A^*(\beta^*) + F_A^*\left(v\left(W_0^{CI}\right) - \beta^*\right) \geq F_A^*(\alpha) + F_A^*\left(v\left(W_0^{CI}\right) - \alpha\right) \tag{2.26}$$

However,

$$\begin{cases} F_A^*(\beta^*) = \int_A u\Big(W_0 - \Pi - X + Y_3^*\Big)\, d\mu \\ F_A^*\left(v\left(W_0^{CI}\right) - \beta^*\right) = \int_{S\setminus A} u\Big(W_0 - \Pi - X + Y_4^*\Big)\, d\mu \end{cases}$$

Therefore,

$$\int u\Big(W_0 - \Pi - X + Y_2^*\Big)\, d\mu \geq \int u\Big(W_0 - \Pi - X + \widetilde{Y}\Big)\, d\mu \tag{2.27}$$

Hence, Y_2^* is optimal for Problem 2.21. □

Remark 2.40. *By Lemma 2.39, we can restrict ourselves to solving Problems 2.32 and 2.33 with a parameter $\beta \in \Gamma$.*

Remark 2.41. *By Lemmata 2.30, 2.31, and 2.39, if ν is (μ, X)-vigilant, $\mathcal{U}(X,Y) := u\Big(W_0 - \Pi - X + Y\Big)$ is supermodular, β^* is optimal for Problem 2.38, Y_1^* is optimal for Problem 2.32 with parameter β^*, and Y_2^* is optimal for Problem 2.33 with parameter β^*, then $\widetilde{Y}_\mu^*$ is optimal for Problem 2.21, and $\widetilde{Y}_\mu^* = \widetilde{Y}_{1,\mu}^*$, μ-a.s., where $\widetilde{Y}_\mu^*$ (resp. $\widetilde{Y}_{1,\mu}^*$) is the μ-a.s. unique nondecreasing μ-rearrangement of $Y^* := Y_1^* \mathbf{1}_A + Y_2^* \mathbf{1}_{S\setminus A}$ (resp. of Y_1^*) with respect to X.*

2.4.5 Solving Problems 2.32 and 2.33

Since $\mu(S \setminus A) = 0$, it follows that, for all $Y \in B^+(\Sigma)$, we have

$$\int_{S \setminus A} u\Big(W_0 - \Pi - X + Y\Big)\, d\mu = 0$$

Consequently, any Y which is feasible for Problem 2.33 with paramter β is also optimal for Problem 2.33 with parameter β.

Now, for a fixed parameter $\beta \in \Gamma$, we will solve Problem 2.32 "statewise", as follows:

Lemma 2.42. *If $Y^* \in B^+(\Sigma)$ satisfies the following:*

1. $0 \leq Y^*(s) \leq X(s)$, *for all* $s \in A$;
2. $\int_A v\Big(W_0^{CI} + \Pi - Y^*\Big) h\, d\mu = \beta$; *and,*
3. *There exists some* $\lambda \geq 0$ *such that for all* $s \in A$,

$$Y^*(s) = \underset{0 \leq y \leq X(s)}{\arg\max} \left[u\Big(W_0 - \Pi - X(s) + y\Big) - \lambda\, v\Big(W_0^{CI} + \Pi - y\Big) h(s) \right] \tag{2.28}$$

Then the function Y^* solves Problem 2.32 with parameter β.

Proof. Suppose that $Y^* \in B^+(\Sigma)$ satisfies (1), (2), and (3) above. Then Y^* is clearly feasible for Problem 2.32 with parameter β. To show optimality of Y^* for Problem 2.32 note that for any other $Y \in B^+(\Sigma)$ which is feasible for Problem 2.32 with parameter β, we have, for all $s \in A$,

$$\begin{aligned} & u\Big(W_0 - \Pi - X(s) + Y^*(s)\Big) - u\Big(W_0 - \Pi - X(s) + Y(s)\Big) \\ \geq\ & \lambda \left[h(s)\ v\Big(W_0^{CI} + \Pi - Y^*(s)\Big) - h(s)\ v\Big(W_0^{CI} + \Pi - Y(s)\Big) \right] \end{aligned}$$

Consequently,

$$\int_A u\Big(W_0 - \Pi - X + Y^*\Big)\, d\mu - \int_A u\Big(W_0 - \Pi - X + Y\Big)\, d\mu \;\geq\; \lambda\Big[\beta - \beta\Big] = 0$$

which completes the proof. □

Remark 2.43. *In Chapter 3 we will apply these ideas to a problem of demand for insurance under heterogeneous subjective beliefs, and we will characterize the solution.*

2.5 Monotone Likelihood Ratios and Vigilance of Beliefs

2.5.1 The Monotone Likelihood Ratio Assumption

The purpose of this subsection is to show that our assumption of *vigilance of beliefs* is implied by the assumption of a *monotone likelihood ratio* in a setting where the DM and the insurer assign a different probability density function (pdf) to the random loss on its range. Needless to say, this presupposes the existence of such pdf-s. We will model this situation as follows: let $(S, \mathcal{G})$ be a measurable space, and let $X \in B^+(\mathcal{G})$ be a random variable with range $X(S) := [0, M]$ on the real line, where $M := \sup\{X(s) : s \in S\} < +\infty$. Let $\Sigma := \sigma\{X\}$. The DM's subjective probability measure μ on (S, Σ) is such that the law $\mu \circ X^{-1}$ is absolutely continuous with respect to the Lebesgue measure, with a Radon-Nikodým derivative f, where $f(t)$ is interpreted as the pdf that the DM assigns to the loss X. Similarly, the insurer's subjective probability measure ν on (S, Σ) is such that the law $\nu \circ X^{-1}$ is absolutely continuous with respect to the Lebesgue measure, with a Radon-Nikodým derivative g, where $g(t)$ is interpreted as the pdf that the insurer assigns to the loss X. Both f and g have support $[0, M]$.

Definition 2.44. *The likelihood ratio is the function $LR : [0, M] \to \mathbb{R}^+$ defined by*

$$LR(t) := g(t)/f(t) \tag{2.29}$$

for all $t \in [0, M]$ such that $f(t) \neq 0$.

Now, define the map $Z : S \to \mathbb{R}^+$ by $Z := LR \circ X$. Then Z is nonnegative and Σ-measurable, and LR is a nondecreasing (resp. nonincreasing) function on its domain if and only if Z is a nondecreasing function (resp. a nonincreasing function) of X. Consider the following two conditions that one might impose.

Condition 2.45 (Monotone Likelihood Ratio)**.** *LR is a nonincreasing function on its domain.*

Condition 2.46 (Vigilance)**.** *ν is (μ, X)-vigilant.*

2.5.2 MLR vs. Vigilance

The following proposition shows that the *vigilance* condition is implied by the *monotone likelihood ratio* condition in this particular setting.

Proposition 2.47. *If Condition 2.45 (Monotone Likelihood Ratio) holds, and if the map $v\left(I \circ X\right) LR\left(X\right) : S \to \mathbb{R}$ is μ-integrable for each $I \circ X \in B^+\left(\Sigma\right)$, then condition 2.46 (Vigilance) holds.*

Proof. First note that since the utility function v is a nondecreasing function (by Assumption 2.2), it follows from Condition 2.45 and Lemma 2.13 that the map $L : [0, M] \times [0, M] \to \mathbb{R}$ defined by

$$L\left(x, y\right) := v\Big(W_0^{CI} + \Pi - y\Big) LR\left(x\right) \tag{2.30}$$

is supermodular (see also Example 2.14 (6)).

Suppose that Condition 2.45 holds. To show that Condition 2.46 is implied, choose $Y_1, Y_2 \in B^+\left(\Sigma\right)$ such that Y_1 and Y_2 have the same distribution under μ, and Y_2 is a nondecreasing function of X. Then by the μ-a.s. uniqueness of the nondecreasing μ-rearrangement, Y_2 is μ-a.s. equal to $\widetilde{Y}_{1,\mu}$, where $\widetilde{Y}_{1,\mu}$ is the nondecreasing μ-rearrangement of Y_1 with respect to X, that is, $Y_2 = \widetilde{Y}_{1,\mu}$, μ-a.s.

Since the function $L\left(x, y\right)$ is supermodular, as observed above, then part (1) of Lemma 2.15 yields

$$\int L\left(X, \widetilde{Y}_{1,\mu}\right)\ d\mu \geq \int L\Big(X, Y_1\Big)\ d\mu$$

that is,

$$\int v\Big(W_0^{CI} + \Pi - \widetilde{Y}_{1,\mu}\Big)\ Z\ d\mu \geq \int v\Big(W_0^{CI} + \Pi - Y_1\Big)\ Z\ d\mu$$

where Z is as defined above. Since $Y_2 = \widetilde{Y}_{1,\mu}$, μ-a.s., we then have

$$\int v\Big(W_0^{CI} + \Pi - Y_2\Big)\ Z\ d\mu \geq \int v\Big(W_0^{CI} + \Pi - Y_1\Big)\ Z\ d\mu$$

which yields (by two "changes of variable"[17], and using the definition of f and g as Radon-Nikodým derivatives of $\mu \circ X^{-1}$ and $\nu \circ X^{-1}$, respectively, with respect to the Lebesgue measure) the following:

$$\int v\Big(W_0^{CI} + \Pi - Y_2\Big)\ d\nu \geq \int v\Big(W_0^{CI} + \Pi - Y_1\Big)\ d\nu$$

as required. Condition 2.46 hence follows from Condition 2.45. This completes the proof of Proposition 2.47. □

Remark 2.48. *Needless to say, the Likelihood Ratio is only defined in situations where densities exist. In other words, when the DM and the CI assign distributions to the underlying uncertainty, with probability density functions, then the Likelihood Ratio can be defined. What Proposition 2.47 asserts is that in more abstract situations where densities do not necessarily exist and hence likelihood ratios cannot be defined, the notion of vigilance might serve as a substitute.*

[17]As in Theorem E.11 on p. 241, and since the map $v(I \circ X)\, LR(X) : S \to \mathbb{R}$ is μ-integrable for each $I \circ X \in B^{+}(\Sigma)$.

2.6 Conclusion

In this chapter we considered an abstract problem of contracting under heterogeneous beliefs, restated as a problem of demand for contingent claims under belief heterogeneity. This problem is an abstraction of many contracting problems where belief heterogeneity is allowed for, such as the problem of optimal insurance design under heterogeneous beliefs that we will examine in Chapter 3.

We showed that under a specific *probabilistic consistency* assumption on the subjective beliefs of the decision maker (DM) and the Claim Issuer (CI) that we called *Vigilance*, there exists a solution which is a nondecreasing function of the underlying uncertainty. We then provided a general method for solving the problem, based on a *splitting* procedure suggested by Lebesgue's Decomposition Theorem.

Technically, the assumption of *Vigilance* is essential to show existence of optimal claims which are nondecreasing functions of the underlying uncertainty. *Vigilance of beliefs* is implied by the assumption of a *monotone likelihood ratio*, as discussed in section 2.5.

In Chapter 4 we will study some mathematical properties of collections of *vigilant beliefs*.

2.7 Appendix: Related Analysis

Lemma 2.49. *Let $(\Omega, \mathcal{F})$ be a given measurable space, and suppose that η is a finite non-negative measure on $(\Omega, \mathcal{F})$. Let Z be any $\mathbb{R}^+$-valued, bounded, and $\mathcal{F}$-measurable function on Ω. If $A \in \mathcal{F}$ is such that $\eta(A) > 0$, then the following are equivalent:*

(i) $\int_A Z \, d\eta = 0$

(ii) $Z = 0$, *η-a.s. on A.*

Proof. The proof is elementary and will be skipped. See, for instance, Aliprantis and Border [3], Theorem 11.16–(3) on p. 412. □

Lemma 2.50. *If $(f_n)_n$ is a uniformly bounded sequence of nondecreasing real-valued functions on some closed interval $\mathcal{I}$ in $\mathbb{R}$, with bound N (i.e. $|f_n(x)| \leq N,\ \forall x \in \mathcal{I},\ \forall n \geq 1$), then there exists a nondecreasing real-valued bounded function f^* on $\mathcal{I}$, also with bound N, and a subsequence of $(f_n)_n$ that converges pointwise to f^* on $\mathcal{I}$.*

Proof. See Carothers [75], Lemma 13.15 on p. 211. Lemma 2.50 is a special case of Helly's First Theorem, sometimes also called Helly's Selection Theorem or Helly's Compactness Theorem (see Carothers [75], Theorem 13.16 on p. 212, or Doob [106], Theorem X.9 on pp. 165-166). □

2.8 Appendix: Dynkin's π-λ Theorem

Definition 2.51 (π-system)**.** *Let S be a nonempty set. A nonempty collection $\mathcal{P}$ of subsets of S is said to be a π-system if for each $A, B \in \mathcal{P}$, $A \cap B \in \mathcal{P}$.*

Hence, a π-system is a nonempty collection of subsets of a set, which is closed under finite intersections.

Definition 2.52 (λ-system, or Dynkin class)**.** *Let S be a nonempty set. A nonempty collection $\mathcal{L}$ of subsets of S is said to be a λ-system if*

1. *$S \in \mathcal{L}$;*
2. *If $A, B \in \mathcal{L}$ are such that $A \subset B$, then $B \setminus A \in \mathcal{L}$; and,*
3. *If $\{A_n\}_n$ is a nondecreasing sequence of elements of $\mathcal{L}$ such that $A_n \uparrow A := \bigcup_{n=1}^{+\infty} A_n$, then $A \in \mathcal{L}$.*

Theorem 2.53 (Dynkin's π-λ Theorem)**.** *Let S be a nonempty set, $\mathcal{P}$ a π-system in S, and $\mathcal{L}$ a λ-system in S. If $\mathcal{P} \subset \mathcal{L}$ then $\sigma\{\mathcal{P}\} \subset \mathcal{L}$, where $\sigma\{\mathcal{P}\}$ is the σ-algebra of subsets of S generated by $\mathcal{P}$.*

Proof. See Aliprantis and Border [3] (pp. 135-136), or Resnick [240] (p. 37). □

Chapter 3

The Demand for Insurance under Heterogeneous Subjective Beliefs

3.1 Introduction

The problem of optimal design of an insurance contract has become part of the folklore of the theory of insurance, as it were. From the outset, the problem was studied within the framework of Expected-Utility Theory (EUT), as in the seminal work of Arrow [25], Borch [52], and Raviv [239], where it was shown that full insurance above a deductible is optimal when the premium principle depends on the actuarial value of the indemnity, when the decision maker (DM) is a risk-averse Expected-Utility Maximizer, and when both the DM and the insurer share the same (additive) probabilistic beliefs about the realization of a given insurable loss. These basic results were then extended in many different directions, while maintaining the assumption of homogeneity of beliefs. For instance,

1. analysis of the optimal deductible level was done in Drèze [107], Eeckhoudt et al. [116], Gould [156], Meyer and Ormiston [211], Moffet [213], Mossin [214], Pashigian et al. [223], and Schlesinger [272];

2. Cummins and Mahul [86] impose an additional upper limit on coverage, and show that the optimal indemnity is full insurance above a deductible up to a cap;

3. effects of changes in the distribution of the loss on the level of the deductible were studied by Demers and Demers [95], Eeckhoudt et al. [116], and Schlee [271];

4. effects of changes in the DM's risk-aversion (curvature of the concave utility) on the level of the deductible were studied by Schlesinger [272];

5. optimal insurance with more general premium principles was investigated in Carlier and Dana [72], Deprez and Gerber [97], Promislow and Young [228], and Young [304].

In this "classical" approach, however, the uncertainty inherent in the insurable loss is assumed to be totally objective, *a priori*. Indeed, the insurable loss is assumed to be a random variable X on an *objective* probability space $(\Omega, \mathcal{F}, P)$. That is, the probability measure P is a totally objective object, the only role of which is to induce a law for X on the real line. Hence, P cannot be a reflection of the *subjective beliefs* of the insurer and the DM, which have no reason whatsoever to be identical *a priori*. The main motivation behind the present study is precisely this lack of subjectivity in the "classical" insurance model.

Even in insurance models with Non-Expected Utility representation of preferences, the insurable loss has always been taken to be an exogenously given random variable X on an *objective* probability space $(\Omega, \mathcal{F}, P)$. This was done, for instance, in Carlier and Dana [72] and [74], Doherty and Eeckhoudt [104], Gollier and Schlesinger [155], Karni [179], Machina [198], Safra and Zilcha [264], Schlesinger [273], and Zilcha and Chew [307].

Information Asymmetry

In a related stream of the literature, problems of information asymmetry in insurance markets were usually studied in the context of adverse selection or moral hazard. Adverse selection was introduced into the theory of insurance starting from the ground-breaking work of Akerlof [2], Rothschild and Stiglitz [247], Stiglitz [287], and Wilson [301]. The classical setup considers a risk-neutral EU-maximizing insurer and two types of risk-averse EU-maximizing insureds: a high-risk type (h) and a low risk-type (l). There are only two states of the world: an *accident* state and a *no-accident* state. An insurable loss X then takes the value 0 with a probability $1 - p_i$, for $i = h, l$, and a fixed value $L > 0$ with probability p_i. However, for each risk type i, both the insurer and the i-type insured have perfectly homogeneous beliefs, in that they both assign the distribution $(L, p_i; 0, 1 - p_i)$ to the loss. In this framework, "information asymmetry" refers to the fact that the type of the insured is private information, not the insured's perception of the loss distribution.

Moral hazard issues, initially considered within the more general principal-agent setting, were introduced into the theory of insurance by Arnott and Stiglitz [20], [21], and [22], Shavell [278], and Stiglitz [288], for instance. In these models, there is risk-neutral EU-maximizing insurer and a risk-averse EU-maximizing insured. There is a two-state world (accident and no-accident) where the loss X can take a value $L > 0$ (in case of accident), with a known probability p_e that depends on the insured's effort (e) in preventing the loss. In the no-accident state, the loss takes the value 0 with probability $1 - p_e$. However, both the insurer and the insured have perfectly homogeneous beliefs at each effort level. That is, for a given effort level e, both the insurer and the insured assign the distribution $(L, p_e; 0, 1 - p_e)$ to the loss. In this framework, "information asymmetry" refers to the fact that the effort level of the insured is private information, not the insured's perception of the loss distribution for a given effort level.

Recently, Jeleva and Villeneuve [173] extended the two-state adverse selection model of Stiglitz [287] to account for belief heterogeneity. In their framework, for each type of insured, the insurer and the insured have different beliefs about the probability of the loss taking the positive value L. Although the authors give an interesting analysis, one might argue that the two-state framework is of limited interest to actuaries, for at least two reasons: (i) typically, financial and insurance risks are not binary risks as in the two-state model; and (ii) assuming the classical constraint that an indemnity $I(X)$ be nonnegative and not larger than the loss itself, a two-state model where the loss X has a distribution $(L, p; 0, 1 - p)$ cannot determine the *shape* of the optimal indemnity schedule. For instance, the indemnity can be a *deductible* contract of the form $I(X) = (X - d)^+$, for some $d \geq 0$. Indeed, in the no-accident state, the loss is 0, and so the indemnity is 0. In the accident state, the loss is $L > 0$, and so the indemnity is N, for some $N \in (0, L]$. Letting $d = L - N \geq 0$, we can then write the indemnity as $I(X) = (X - d)^+$. However, the indemnity can also be of the *coinsurance* type, i.e. of the form $I(X) = \alpha X$, for some $\alpha \in (0, 1]$. Indeed, in the no-accident state, the loss is 0, and so the indemnity is 0. In the accident state, the loss is $L > 0$, and so the indemnity is N, for some $N \in (0, L]$. Letting $\alpha = N/L$, we can then write the indemnity as $I(X) = \alpha X$. From an actuarial viewpoint, the interest in such a framework is limited.

Heterogeneity of Subjective Beliefs

In this chapter, we examine a general problem of *belief asymmetry* in an insurance model with one insurer and one DM. We adopt a decision-theoretic approach to belief formation and rely on the heterogeneity of subjective beliefs as a proxy for belief asymmetry. Moreover, we do not assume that the DM's actions influence the realization of the random loss under consideration. We then use the results and techniques of Chapter 2 to determine the optimal form of an indemnity, within an insurance model that resembles the "classical" model. We show that if *Vigilance* holds, then there exists an event to which the DM assigns full subjective probability, and on which an optimal solution has a *generalized deductible* form (defined below).

Although the problem is a very natural one, the effect of the heterogeneity of beliefs in an insurance market on the shape of an optimal contract was only studied at a more or less suitable level of generality by Marshall [207], to the best of our knowledge. After the exposition of this chapter's main result, we will explain Marshall [207]'s model in more detail, so as to show how the model presented in this chapter is radically different in spirit and in scope.

Outline

In section 3.2 we introduce some notation and definitions, as well as the general setup for our model. In section 3.3 we state the problem and our main result. In section 3.4 we prove the main result of this chapter. In section 3.5 we discuss some special cases of our setting. In section 3.6 we discuss the work of Marshall [207] and how it differs from ours. Finally, section 3.7 concludes. Appendix 3.8 reviews the "classical" insurance design problem, and Appendix 3.9 provides a useful mathematical result that will be used in this chapter.

3.2 Preliminaries

Consider the setting of section 2.2 to be applicable all throughout this chapter. In the context of this chapter, the underlying uncertainty X is interpreted as a given random loss against which the DM seeks an insurance coverage. As in section 2.2, X is a bounded, measurable, real-valued function on the state space $(S, \mathcal{G})$, with closed range $X(S) = [0, M]$, where $M = \|X\|_s < +\infty$. In other words, the random loss X is a mapping of S onto the closed interval $[0, M]$. In particular, there are states of the world in which the loss takes a zero value, that is, $\{s \in S : X(s) = 0\} \neq \varnothing$. Henceforth, we shall denote by Σ the σ-algebra $\sigma\{X\}$ of subsets of S generated by the random loss X. Moreover, in the present context, the CI is simply the insurer and the insurance contract is a claim of the form $I \circ X$.

The insurance market gives the DM the possibility of entering into an insurance contract with the insurer. Such a contract is represented by a pair (Π, I), where $\Pi > 0$ is the premium paid by the DM in return of the indemnity I. The indemnity is a Borel-measurable map $I : [0, M] \rightarrow [0, M]$, such that $0 \leq I(X(s)) \leq X(s)$ for all $s \in S$. Then $Y := I \circ X \in B^+(\Sigma)$, the collection of all bounded, Σ-measurable $\mathbb{R}^+$-valued functions on S, where $\Sigma := \sigma\{X\}$, as in section 2.2.

Setup

As in section 2.2, the DM and the insurer have preferences over the elements of $B^+ (\Sigma)$. Both the DM's and the insurer's preferences have a Subjective Expected-Utility (SEU) representation (see Appendix A.1.3). The DM's preferences induce a utility function $u : \mathbb{R} \to \mathbb{R}$, unique up to a positive linear transformation, and the insurer's preferences induce a utility function $v : \mathbb{R} \to \mathbb{R}$, also unique up to a positive linear transformation. Both the DM's and the insurer's preferences are also assumed to satisfy the Arrow-Villegas *Monotone Continuity* axiom (Arrow [25], Chateauneuf et al. [77], and Villegas [296] – see also p. 178), hence yielding a unique countably additive subjective probability measure on the measurable space (S, Σ), for each (as in Corollary A.19 on p. 179).

The subjectivity of the beliefs of each of the DM and the insurer is reflected in the different subjective probability measure that each has over the measurable space (S, Σ):

Assumption 3.1. *The DM's beliefs are represented by the countably additive probability measure* μ *on* (S, Σ)*, and the insurer's beliefs are represented by the countably additive probability measure* ν *on* (S, Σ)*.*

Additionally, we suppose that the DM is risk averse, having a utility index u such that following holds:

Assumption 3.2. *The DM's utility is bounded and satisfies Inada's [172] conditions. Specifically,*

1. u *is bounded;*

2. $u(0) = 0$*;*

3. u *is strictly increasing and strictly concave;*

4. u *is continuously differentiable; and,*

5. $u'(0) = +\infty$ *and* $u'(+\infty) = 0$*.*

Remark 3.3. *Assumption 3.2 (1) above on the DM's utility function is, strictly speaking, redundant. Indeed, the utility function given from the DM's preferences in Savage's SEU representation is a bounded function*[1]*.*

[1] See, e.g. Fishburn [128] (Theorem 14.1 on pp. 192-193) or Gilboa [149] (Theorem 10.2 on pp. 108-109). See also Theorem A.20 in Appendix A.1.3.

Moreover, assuming that u is strictly concave and continuously differentiable implies that u' is both continuous and strictly decreasing. This then implies that $(u')^{-1}$ is continuous and strictly decreasing, by the Inverse Function Theorem.

We will also assume that the DM has initial wealth W_0 strictly larger than the premium Π.

Assumption 3.4. $W_0 > \Pi > 0$.

We also make the assumption that the random loss X (with closed range $[0, M]$) has a nonatomic law induced by the probability measure μ, and that the CI and the DM are both aware of the fact that μ represents the DM's beliefs and ν represents the insurer's beliefs. Moreover, we will assume that the subjective probability measures μ and ν are not mutually singular[2], that the DM is almost certain that the random loss she will incur is not larger than her remaining wealth after the premium has been paid. Specifically:

Assumption 3.5. *We assume that:*

1. *$\mu \circ X^{-1}$ is nonatomic;*

2. *$X \leq W_0 - \Pi$, μ-a.s. In other words, $\mu\Big(\{s \in S : X(s) > W_0 - \Pi\}\Big) = 0$;*

3. *μ is known by the insurer, and ν is known by the DM; and,*

4. *μ and ν are not mutually singular.*

Remark 3.6. *Assumption 3.5 (1) is a technical requirement that is needed for defining the equimeasurable monotone rearrangement, as in section 2.3.*

Assumption 3.5 (2) simply states that the DM is well-diversified so that the particular loss exposure X against which she is seeking an insurance coverage is sufficiently small.

Assumption 3.5 (4) means that the insurer and the DM do not have beliefs that are totally incompatible. However, this does not prevent the agents from assigning different probabilities to events, and they typically do not assign same likelihoods to the realizations of the uncertainty X.

[2]Two finite nonnegative measures m_1 and m_2 on the measurable space (S, Σ) are said to be mutually singular, denoted by $m_1 \perp m_2$, if there is some $A \in \Sigma$ such that $m_1(S \setminus A) = m_2(A) = 0$. In other words, $m_1 \perp m_2$ if there is a Σ-partition $\{A, B\}$ of the set S of states of nature such that μ_1 is concentrated on A and μ_2 is concentrated on B.

The total wealth of the DM is the Σ-measurable, $\mathbb{R}$-valued and bounded function on S defined by

$$W(s) := W_0 - \Pi - X(s) + Y(s), \ \forall s \in S \tag{3.1}$$

Finally, we assume that the insurer is risk-neutral. This assumption is common in contracting problems, principal-agent problems, and especially in the insurance framework (as in Arrow [25]). Since the insurer's utility function v is unique up to a positive linear transformation[3], we can then assume, without loss of generality, that v is simply the identity function. The total wealth of the insurer is the Σ-measurable, $\mathbb{R}$-valued and bounded function on S defined by

$$W^{ins}(s) := W_0^{ins} + \Pi - Y(s) - \rho\, Y(s), \ \forall s \in S \tag{3.2}$$

where W_0^{ins} is the insurer's initial wealth and $\rho > 0$ is a (proportional) cost associated with handling the insurance contract Y, as in the model of Arrow [25] (or section III of Raviv [239]).

Vigilance of Beliefs

Proposition 3.7. *The probability measure ν is (μ, X)-vigilant if and only if for any $Y_1, Y_2 \in B^+(\Sigma)$ such that*

(i) Y_1 and Y_2 have the same distribution under μ,

(ii) Y_2 is a nondecreasing function of X,

the following holds:

$$\int Y_2 \, d\nu \le \int Y_1 \, d\nu \tag{3.3}$$

Proof. By Definition 2.4 and by risk-neutrality of the insurer, ν is (μ, X)-vigilant if and only if $\int \left(W_0^{ins} + \Pi - Y_2 - \rho\, Y_2\right) d\nu \ge \int \left(W_0^{ins} + \Pi - Y_1 - \rho\, Y_1\right) d\nu$, that is

$$\left[W_0^{ins} + \Pi\right] - (1+\rho) \int Y_2 \, d\nu \ge \left[W_0^{ins} + \Pi\right] - (1+\rho) \int Y_1 \, d\nu,$$

[3] See Theorem A.20 in Appendix A.1.3.

which yields $\int Y_2 \, d\nu \leq \int Y_1 \, d\nu$. $\square$

One possible interpretation of (μ, X)-vigilance, when one thinks of $\int Y \, d\nu$ as the insurer's subjective measure (albeit very rudimentary) of the risk associated with $Y \in B^+(\Sigma)$, in reference to X, is as follows: if ν is (μ, X)-vigilant, then for a given risk Y_1 for the insurer that depends on X, if Y_2 is another such risk that depends on X and that the DM believes to be identically distributed as Y_1 (under her subjective probability measure μ) and nondecreasing in X, then the insurer will not assign a higher subjective risk measure to Y_2 than to Y_1. In this sense, *the insurer is vigilant in his assessment of the riskiness of* Y_2 *as a function of* X. In a sense, the insurer assigns some *credibility* to the DM's subjective assessment of Y_2 in reference to both Y_1 and X. This implies a certain *probabilistic consistency* between the DM's and the insurer's subjective beliefs for a class of risks which are seen as functions of a given risk. As we shall see later on, this consistency requirement is crucial to rule out moral hazard problems that might result from a downward misrepresentation of the loss by the DM.

Remark 3.8. *Recall that in section 2.5 we showed that in the specific setting where the DM and the insurer assign different probability density functions to a random loss* X *with range* $[0, M]$*, the assumption of vigilance is implied by the assumption of a monotone likelihood ratio.*

Generalized Deductible Contracts

Henceforth, we shall adopt the following terminology:

Definition 3.9. *The indemnity schedule* $I : [0, M] \to [0, M]$ *will be called:*

1. *A full insurance when* $I(t) = t$*, for all* $t \in [0, M]$*;*

2. *A deductible when there is some* $d \in [0, M]$ *such that*

$$I(t) = \begin{cases} 0 \text{ if } t \in [0, d) \\ t - d \text{ if } t \in [d, M] \end{cases}$$

3. *A capped deductible when there are* $d, c \in [0, M]$ *such that*

$$I\left(t\right)=\min\left[I_d\left(t\right),c\right],\ \forall\ t\in\left[0,M\right]$$

where I_d is a deductible contract with deductible d.

4. *A generalized deductible when there is some $d\in[0,M]$ such that*

$$I\left(t\right)=\begin{cases}0 \ \text{if}\ t\in[0,d)\\ f\left(t\right) \ \text{if}\ t\in[d,M]\end{cases}$$

for some nondecreasing Borel-measurable function $f:[0,M]\to[0,M]$ such that $0\le f\left(t\right)\le t$ for $t\in[0,M]$.

Figure 3.1 below illustrates the shapes of these insurance contracts.

3.3 Design of the Optimal Insurance Contract

The DM's problem

The problem of designing the optimal insurance contract can be seen as that of finding the indemnity that will maximize the expected utility of the DM's wealth, under her subjective probability measure, subject to a constraint on the premium and to some constraints on the indemnity function. Specifically, the DM's problem is the following:

Problem 3.10.

$$\sup_{Y\in B^+(\Sigma)}\left\{\int u\Big(W_0-\Pi-X+Y\Big)\,d\mu\right\}:$$
$$\begin{cases}0\le Y\le X\\ \int v\Big(W_0^{ins}+\Pi-\left(1+\rho\right)Y\Big)\,d\nu\ge v\left(W_0^{ins}\right)\end{cases}$$

The first constraint is standard (see Arrow [25] and Raviv [239]), and says that an indemnity is nonnegative and cannot exceed the loss itself. The latter requirement simply rules out situations where the DM has an incentive to create damage (see Huberman, Mayers and Smith [169]). The second constraint is simply the insurer's *participation constraint*, or *individual rationality constraint*, where $R=v\left(W_0^{ins}\right)$ is the insurer's *reservation utility.*

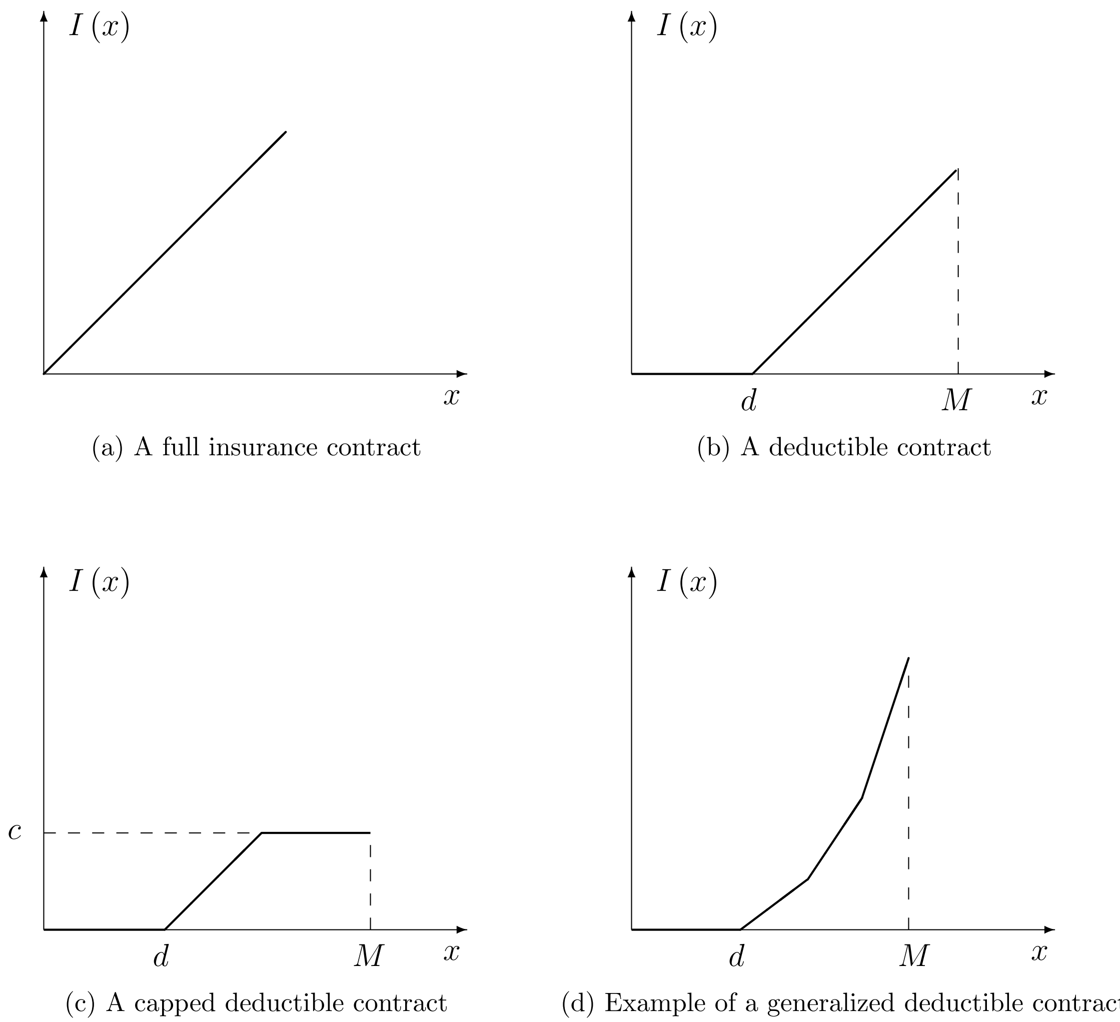


Figure 3.1: Some examples of insurance contracts.

Since v was assumed to be the identity function (by risk-neutrality of the insurer), the insurer's individual rationality constraint can be re-written as the following *premium constraint*:

$$W_0^{ins} + \Pi - (1+\rho) \int Y \, d\nu \geq W_0^{ins},$$

that is,

$$\Pi \geq (1+\rho) \int Y \, d\nu$$

Hence, we can write the DM's problem as follows:

Problem 3.11. *For a given loading factor $\rho > 0$,*

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int u\Big(W_0 - \Pi - X + Y\Big) \, d\mu \right\} :$$
$$\begin{cases} 0 \leq Y \leq X \\ \Pi \geq (1+\rho) \int Y \, d\nu \end{cases}$$

The second constraint in Problem 3.11 is commonplace in insurance problems. Raviv [239], for instance, assumes that the premium is at least equal to $(1+\alpha) \int (I \circ X) \, dP$, where P in Raviv's [239] context is the probability measure common to both the DM and the insurer, and $\alpha \geq 0$ is a loading factor.

Also typical in the classical insurance setting is to impose an additional monotonicity constraint on the desired optimal indemnity by requiring that it be nondecreasing in the loss. This constraint, first introduced by Huberman, Mayers and Smith [169], is meant to prevent moral hazard issues that might result from a downward misrepresentation of the loss by the DM. Here, we prefer not to impose this monotonicity as a constraint, but rather to achieve it as a property of an optimal indemnity. This is done using rearrangement techniques and the assumption of *vigilance*.

In the following, we will assume that the feasibility set for Problem 3.11 is nonempty:

Assumption 3.12. *$\mathcal{F}_{SB} \neq \varnothing$, where $\mathcal{F}_{SB}$ is the collection of all feasible $Y \in B^+(\Sigma)$ for Problem 3.11.*

The Main Results

Recall from section 2.4.3 that there exists a unique pair (ν_{ac}, ν_s) of (nonnegative) finite measures on (S, Σ) such that $\nu = \nu_{ac} + \nu_s$, $\nu_{ac} << \mu$, and $\nu_s \perp \mu$. As in section 2.4.3, let $h : S \to [0, +\infty)$ de the Radon-Nikodým derivative of ν_{ac} with respect to μ. That is, h

is the μ-a.s. unique Σ-measurable and μ-integrable nonnegative function on S such that $\nu_{ac}(C) = \int_C h \, d\mu$, for all $C \in \Sigma$. The following Theorem characterizes an optimal solution of Problem 3.11.

Theorem 3.13. *Suppose that the previous assumptions hold, and for each $\lambda \geq 0$ define the function $Y_\lambda^* \in B^+(\Sigma)$ by:*

$$Y_\lambda^* := \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1} (\lambda h) \right] \right) \right] \tag{3.4}$$

Let $\widetilde{Y}_{\lambda,\mu}^*$ denote the μ-a.s. unique nondecreasing μ-rearrangement of Y_λ^* with respect to X. If the insurer's subjective probability measure ν is (μ, X)-vigilant, then there exists a $\lambda^* \geq 0$ and an optimal solution $\mathcal{Y}^*$ to Problem 3.11 which is nondecreasing in the loss X, and such that $\mathcal{Y}^* = \widetilde{Y}_{\lambda^*,\mu}^*$, μ-a.s.

Corollary 3.14. *Under the previous assumptions, and provided the insurer's subjective probability measure ν is (μ, X)-vigilant, there exists an event $E^* \in \Sigma$ such that $\mu(E^*) = 1$, and an optimal solution $\mathcal{Y}^*$ to Problem 3.11 which is nondecreasing in the loss X, and such that for μ-a.a. $s \in S$,*

$$\mathcal{Y}^*(s) = \begin{cases} 0 \ \text{if } X(s) \in [0, a^*) \\ f(X(s)) \ \text{if } X(s) \in [a^*, M] \end{cases} \tag{3.5}$$

for an $a^* \geq 0$ and a nondecreasing, left-continuous, and Borel-measurable function $f : [0, M] \to [0, M]$ such that $0 \leq f(t) \leq t$ for each $t \in [a^*, M]$.

Moreover, $a^* > 0$ when $\mu(\mathcal{D}_{E^*}) \neq 0$, where:

(i) $\mathcal{D}_{E^*} := \left\{ s_0 \in E^* : X(s_0) > 0, \ h(s_0) > 0, \ \int_{E^*} \mathcal{Y}^* \ h \ d\mu < \overline{L}(s_0) \right\}$*; and,*

(ii) $\overline{L}(s_0) := \int_{E^*} \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1} \left(\frac{u'(W_0 - \Pi - X(s_0))}{h(s_0)} \ h \right) \right] \right) \right] \ h \ d\mu.$

Finally, there exists $\kappa \in \mathbb{R}^+$ such that $a^* > 0$ when $\mu(\mathcal{E}_{E^*}) \neq 0$, where:

$$\mathcal{E}_{E^*} := \Big\{ s_0 \in E^* : \ h(s_0) > 0, \ \ \kappa\, h(s_0) > u'(W_0 - \Pi), \qquad 0 < X(s_0) < W_0 - \Pi - (u')^{-1}(\kappa\, h(s_0)) \Big\} \tag{3.6}$$

Corollary 3.14 essentially says that when *vigilance* holds, there is a measurable set D to which the DM assigns full (subjective) probability, and such that an optimal indemnity schedule I^* will pay the DM, in the state of the world $s \in D$, the amount $I_{op}(s)$, where I_{op} is a *generalized deductible* contract on D.

Remark 3.15. *The most interesting implication of Corollary 3.14 is the existence of the deductible a^*, mainly because of the resemblance with the classical result of Arrow [25], Borch [52], and Raviv [239]. We do not provide an explicit characterization of the function f that appears in Corollary 3.14, although it is possible to do so using the ideas developed in section 2.3.4 (see Remark 3.48 on p. 89).*

3.4 Proof of the Main Results

3.4.1 Proof of Theorem 3.13

Existence of a Monotone Solution and its Characterization

Since the function $\mathcal{U}(x, y) := u(W_0 - \Pi - x + y)$ is supermodular (see Example 2.14 (1)) and since ν is assumed to be (μ, X)-vigilant, the results of section 2.4.2 hold here (with Y replaced by $(1 + \rho)Y$ in the CI's wealth process and participation constraint). In particular, there exists a solution to Problem 3.11 which is a nondecreasing function of the loss X.

Similarly, the results of section 2.4.3 also hold. In the following, the Σ-measurable set A on which μ is concentrated[4] (and $\nu_s(A) = 0$) is assumed to be fixed all throughout.

[4] As in the setting of section 2.4.3.

Some Sufficient Problems

Lemma 3.16. *Let Y^* be an optimal solution for Problem 3.11, and suppose that ν is (μ, X)-vigilant. Let $\widetilde{Y}^*_\mu$ be the nondecreasing μ-rearrangement of Y^* with respect to X. Then:*

1. *$\widetilde{Y}^*_\mu$ is optimal for Problem 3.11; and,*

2. *$\widetilde{Y}^*_\mu = \widetilde{Y}^*_{\mu,A}$, μ-a.s., where $\widetilde{Y}^*_{\mu,A}$ is the nondecreasing μ-rearrangement of Y^* with respect to X on A.*

Proof. See Lemma 2.30. □

Lemma 3.17. *Let an optimal solution for Problem 3.11 be given by:*

$$Y^* = Y_1^* \mathbf{1}_A + Y_2^* \mathbf{1}_{S \setminus A} \tag{3.7}$$

for some $Y_1^*, Y_2^* \in B^+(\Sigma)$. Let $\widetilde{Y}^*_\mu$ be the nondecreasing μ-rearrangement of Y^* with respect to X, and let $Y^*_{1,\mu}$ be the nondecreasing μ-rearrangement of Y_1^* with respect to X. Then $\widetilde{Y}^*_\mu = \widetilde{Y}^*_{1,\mu}$, μ-a.s.

Proof. See Lemma 2.31. □

Remark 3.18. *What Lemma 3.16 asserts is that when the insurer's subjective probability measure is vigilant with respect to the DM's subjective probability measure in regards to the risk X, and if there exists an indemnity schedule which is perceived by the DM as optimal for her initial problem, then there exists another indemnity schedule which is perceived by the DM as optimal for her initial problem, and which rules out any possibility of moral hazard resulting from a voluntary downward misrepresentation of losses by the DM. Indeed, as long as an indemnity schedule is nondecreasing in the insurable loss, there is no incentive for the DM to misrepresent the loss downwards.*

Consider now the following three problems:

Problem 3.19. *For a given* $\beta \in \left[0, \min\left(\Pi/\left(1+\rho\right), \int_A X\ d\nu\right)\right]$,

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int_A u\Big(W_0 - \Pi - X + Y\Big)\ d\mu \right\} :$$
$$\begin{cases} 0 \le Y\mathbf{1}_A \le X\mathbf{1}_A \\ \int_A Y\ d\nu = \beta \end{cases}$$

Problem 3.20.

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int_{S \setminus A} u\Big(W_0 - \Pi - X + Y\Big)\ d\mu \right\} :$$
$$\begin{cases} 0 \le Y\mathbf{1}_{S\setminus A} \le X\mathbf{1}_{S\setminus A} \\ \int_{S\setminus A} Y\ d\nu \le \min\left(\frac{\Pi}{1+\rho} - \beta, \int_{S\setminus A} X\ d\nu\right),\ \textit{for the same } \beta \textit{ as in Problem 3.19} \end{cases}$$

Problem 3.21.

$$\sup_{\beta} \left[F_A^*\left(\beta\right) + F_A^*\left(\frac{\Pi}{1+\rho} - \beta\right) : 0 \le \beta \le \min\left(\Pi/\left(1+\rho\right), \int_A X\ d\nu\right) \right] :$$
$$\begin{cases} F_A^*\left(\beta\right)\ \textit{is the supremum value of Problem 3.19, for a fixed } \beta \\ F_A^*\left(\frac{\Pi}{1+\rho} - \beta\right)\ \textit{is the supremum value of Problem 3.20, for the same fixed } \beta \end{cases}$$

Remark 3.22. *The feasibility sets of Problems 3.19 and 3.20 are nonempty. To see why this is true, first note that:*

1. *Since* μ *and* ν *are not mutually singular, by Assumption 3.5, and since* $\mu\left(S \setminus A\right) = 0$, *it follows that* $\nu\left(A\right) > 0$;

2. *Since* $\nu\left(A\right) > 0$, $h \ge 0$, *and* $\nu\left(A\right) = \nu_{ac}\left(A\right) + \nu_s\left(A\right) = \nu_{ac}\left(A\right) = \int_A h\ d\mu$, *it follows from Lemma 2.49 that there exists some* $B \in \Sigma$ *such that* $B \subseteq A$, $\mu\left(B\right) > 0$, *and* $h > 0$ *on* B.

If $\int_A X \, d\nu = \int_A Xh \, d\mu = 0$, then by Lemma 2.49 we have $Xh = 0$, μ-a.s. on A. However, $h > 0$ on B. Thus, $X = 0$, μ-a.s. on B. Consequently, there is some $C \in \Sigma$, with $C \subseteq B$ and $\mu(C) > 0$, such that $X = 0$ on C and $\mu(B \setminus C) = 0$. Therefore, $\mu(B) = \mu(C)$. Now, since $X(s) = 0$, for each $s \in C$, it follows that $C \subseteq \{s \in S : X(s) = 0\}$. Thus, by monotonicity of μ, $\mu(C) \leq \mu(\{s \in S : X(s) = 0\}) = \mu \circ X^{-1}(\{0\})$. But $\mu \circ X^{-1}(\{0\}) = 0$, by nonatomicity of $\mu \circ X^{-1}$ (Assumption 3.5). Therefore, $\mu(C) = 0$, a contradiction. Hence $\int_A X \, d\nu > 0$.

Now, for a given $\beta \in \left[0, \min\left(\Pi/(1+\rho), \int_A X \, d\nu\right)\right]$, the function $Y_1 := \frac{\beta X}{\int_A X \, d\nu}$ is feasible for Problem 3.19 with parameter β.

If $\int_{S\setminus A} X \, d\nu = 0$, then $Y_2 := 0$ is feasible for Problem 3.20. If $\int_{S\setminus A} X \, d\nu > 0$, then $Y_3 := \frac{\alpha X}{\int_{S\setminus A} X \, d\nu}$, with $\alpha := \min\left(\frac{\Pi}{1+\rho} - \beta, \int_{S\setminus A} X \, d\nu\right) \Big/ 2$, is feasible for Problem 3.20 with parameter β, for any given $\beta \in \left[0, \min\left(\Pi/(1+\rho), \int_A X \, d\nu\right)\right]$.

Remark 3.23. *Just as in Remark 2.22, the supremum value of each of the above three problems is finite.*

Lemma 3.24. *If β^* is optimal for Problem 3.21, Y_3^* is optimal for Problem 3.19 with parameter β^*, and Y_4^* is optimal for Problem 3.20 with parameter β^*, then $Y_2^* := Y_3^* \mathbf{1}_A + Y_4^* \mathbf{1}_{S\setminus A}$ is optimal for Problem 3.11.*

Proof. See Lemma 2.39. □

Remark 3.25. *By Lemmata 3.16, 3.17, and 3.24, if ν is (μ, X)-vigilant, β^* is optimal for Problem 3.21, Y_1^* is optimal for Problem 3.19 with parameter β^*, and Y_2^* is optimal for Problem 3.20 with parameter β^*, then $\widetilde{Y}_\mu^*$ is optimal for Problem 3.11, and $\widetilde{Y}_\mu^* = \widetilde{Y}_{1,\mu}^*$, μ-a.s., where $\widetilde{Y}_\mu^*$ (resp. $\widetilde{Y}_{1,\mu}^*$) is the μ-a.s. unique nondecreasing μ-rearrangement of $Y^* := Y_1^* \mathbf{1}_A + Y_2^* \mathbf{1}_{S\setminus A}$ (resp. of Y_1^*) with respect to X.*

Lemma 3.26. *If β^* is optimal for Problem 3.21, then $\beta^* > 0$.*

Proof. First note the following:

(i) Since $\mu(S \setminus A) = 0$, it follows that $\int_{S\setminus A} Z \, d\mu = 0$, for each $Z \in B(\Sigma)$, and so $F_A^*\left(\frac{\Pi}{1+\rho} - \beta\right) = 0$, for each $\beta \in \left[0, \min\left(\Pi/(1+\rho), \int_A X \, d\nu\right)\right]$. Consequently, $F_A^*(\beta) + F_A^*\left(\frac{\Pi}{1+\rho} - \beta\right) = F_A^*(\beta)$, for each $\beta \in \left[0, \min\left(\Pi/(1+\rho), \int_A X \, d\nu\right)\right]$. Therefore, in particular, $F_A^*(\beta^*) + F_A^*\left(\frac{\Pi}{1+\rho} - \beta^*\right) = F_A^*(\beta^*)$.

(ii) Since μ and ν are not mutually singular, by Assumption 3.5, it follows that $\nu(A) > 0$.

(iii) Since $\nu(A) > 0$, $h \geq 0$, and $\nu(A) = \nu_{ac}(A) + \nu_s(A) = \nu_{ac}(A) = \int_A h \, d\mu$, it follows from Lemma 2.49 that there exists some $B \in \Sigma$ such that $B \subseteq A$, $\mu(B) > 0$, and $h > 0$ on B.

Now, suppose, *per contra*, that $\beta^* = 0$ is optimal for Problem 3.21, and let Y_0 be optimal for Problem 3.19 with parameter 0, so that $F_A^*(0) = \int_A u(W_0 - \Pi - X + Y_0) \, d\mu$.

Since $\beta^* = 0$ is optimal for Problem 3.21, we have $F_A^*(0) \geq F_A^*(\beta)$, for each $\beta \in \left[0, \min\left(\Pi/(1+\rho), \int_A X \, d\nu\right)\right]$.

Since Y_0 is feasible for Problem 3.19 with parameter $\beta^* = 0$, we have $\int_A Y_0 \, d\nu = \int_A Y_0 h \, d\mu = \beta^* = 0$. Now, since $\mu(A) > 0$ and $Y_0 h \geq 0$, it follows from Lemma 2.49 that $Y_0 h = 0$, μ-a.s. on A. Moreover, since $h > 0$ on B and $\mu(B) > 0$, it follows that $Y_0 = 0$, μ-a.s. on B.

Define the function Z by $Z := Y_0 \mathbf{1}_{A\setminus B} + \min\Big(X, \Pi/(1+\rho)\Big)\mathbf{1}_B$, and let $K_Z := \int_A Z \, d\nu$. Then the following clearly hold:

(i) $Z \in B^+(\Sigma)$;

(ii) $0 \leq Z\mathbf{1}_A \leq X\mathbf{1}_A$;

(iii) $0 \leq K_Z \leq \min\Big(\int_A X \, d\nu, \Pi/(1+\rho)\Big)$.

Therefore, in particular, K_Z is feasible for Problem 3.21 and Z is feasible for Problem 3.19 with parameter K_Z. Moreover,

$$\begin{aligned} 0 \leq K_Z &= \int_{A\setminus B} Y_0 \, d\nu + \int_B \min\Big(X, \Pi/(1+\rho)\Big) \, d\nu \\ &= \int_B \min\Big(X, \Pi/(1+\rho)\Big) \, d\nu = \int_B \min\Big(X, \Pi/(1+\rho)\Big) h \, d\mu \end{aligned}$$

If $K_Z = 0$, then $\int_B \min\Big(X, \Pi/(1+\rho)\Big)h \ d\mu = 0$ and $\min\Big(X, \Pi/(1+\rho)\Big)h \geq 0$. Hence, by Lemma 2.49, $\min\Big(X, \Pi/(1+\rho)\Big)h = 0$, μ-a.s. on B. However, $h > 0$ on B. Thus, $\min\Big(X, \Pi/(1+\rho)\Big) = 0$, μ-a.s. on B. Since $\Pi > 0$, this yields $X = 0$, μ-a.s. on B. Consequently, there is some $C \in \Sigma$, with $C \subseteq B$ and $\mu(C) > 0$, such that $X = 0$ on C and $\mu(B \setminus C) = 0$. Therefore, $\mu(B) = \mu(C)$. Now, since $X(s) = 0$, for each $s \in C$, it follows that $C \subseteq \{s \in S : X(s) = 0\}$. Thus, by monotonicity of μ, $\mu(C) \leq \mu(\{s \in S : X(s) = 0\}) = \mu \circ X^{-1}(\{0\})$. But $\mu \circ X^{-1}(\{0\}) = 0$, by nonatomicity of $\mu \circ X^{-1}$ (Assumption 3.5). Therefore, $\mu(C) = 0$, a contradiction. Hence $K_Z > 0$.

Finally,

$$\begin{aligned}
F_A^*(K_Z) &\geq \int_A u(W_0 - \Pi - X + Z)\ d\mu \\
&= \int_{A\setminus B} u(W_0 - \Pi - X + Y_0)\ d\mu + \int_B u(W_0 - \Pi - X + \min(X, \Pi/(1+\rho)))\ d\mu \\
&\geq \int_{A\setminus B} u(W_0 - \Pi - X + Y_0)\ d\mu + \int_B u(W_0 - \Pi - X)\ d\mu \\
&= \int_A u(W_0 - \Pi - X + Y_0)\ d\mu := F_A^*(0) = F_A^*(\beta^*)
\end{aligned}$$

This contradicts the optimality of $\beta^* = 0$ for Problem 3.21. Consequently, if β^* is optimal for Problem 3.21 then $\beta^* > 0$. □

Solving Problem 3.20

Since $\mu(S \setminus A) = 0$, it follows that, for all $Y \in B^+(\Sigma)$, we have

$$\int_{S\setminus A} u\Big(W_0 - \Pi - X + Y\Big)\ d\mu = 0$$

Consequently, any Y which is feasible for Problem 3.20 with paramter β is also optimal for Problem 3.20 with parameter β. For instance, define $Y_4^* := \min\left[X, \max\left\{0, X - \overline{d}_\beta\right\}\right]$, where $\overline{d}_\beta$ is chosen such that $\int_{S\setminus A} Y_4^*\ d\nu \leq \min\Big(\frac{\Pi}{1+\rho} - \beta, \int_{S\setminus A} X\ d\nu\Big)$. Then $Y_4^* \mathbf{1}_{S\setminus A}$ is optimal for Problem 3.20.

Remark 3.27. *The choice of* $\overline{d}_\beta$ *so that* $\int_{S\backslash A} Y_4^* \, d\nu \le \min\left(\frac{\Pi}{1+\rho} - \beta, \int_{S\backslash A} X \, d\nu\right)$ *is justified by the following argument: define the function* $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ *by*

$$\phi(\alpha) = \int_{S\backslash A} Y_{4,\alpha} \, d\nu \tag{3.8}$$

where $Y_{4,\alpha} := \min[X, \max\{0, X - \alpha\}]$*, for each* $\alpha \ge 0$*. Then* ϕ *is a nonincreasing function of* α*. Moreover, by the continuity of the functions* $\max(0, .)$ *and* $\min(x, .)$*, and by Lebesgue's Dominated Convergence Theorem,* ϕ *is a continuous function of the parameter* α*.*

Now, by the continuity of the functions max *and* min*,* $\lim_{\alpha \to 0} Y_{4,\alpha} = X$ *and* $\lim_{\alpha \to +\infty} Y_{4,\alpha} = 0$*. Therefore, by continuity of the function* ϕ *in* α*,*

$$\lim_{\alpha \to 0} \phi(\alpha) = \int_{S\backslash A} X \, d\nu$$

and

$$\lim_{\alpha \to +\infty} \phi(\alpha) = 0$$

Consequently, ϕ *is a continuous nonincreasing function of* α *such that* $\lim_{\alpha \to +\infty} \phi(\alpha) = 0$ *and* $\lim_{\alpha \to 0} \phi(\alpha) = \int_{S\backslash A} X \, d\nu$*. Thus, by the Intermediate Value Theorem, one can always choose* α *such that* $\phi(\alpha) \le \min\left(\frac{\Pi}{1+\rho} - \beta, \int_{S\backslash A} X \, d\nu\right)$*, for any* $\beta \in \left[0, \min\left(\Pi/(1+\rho), \int_A X \, d\nu\right)\right]$*.*

Solving Problem 3.19

For a fixed parameter $\beta \in \left[0, \min\left(\Pi/(1+\rho), \int_A X \, d\nu\right)\right]$, we will solve Problem 3.19 "statewise" as in Lemma 2.42. Moreover, by Lemma 3.26, we can restrict the analysis to the case where $\beta \in \left(0, \min\left(\Pi/(1+\rho), \int_A X \, d\nu\right)\right]$.

Lemma 3.28. *If* $Y^* \in B^+(\Sigma)$ *satisfies the following:*

1. $0 \le Y^*(s) \le X(s)$*, for all* $s \in A$*;*

2. $\int_A Y^* h \, d\mu = \beta$*, for some* $\beta \in \big(0, \min\left(\Pi/\left(1+\rho\right), \int_A X \, d\nu\right)\big]$*; and,*

3. There exists some $\lambda \geq 0$ *such that for all* $s \in A \setminus \{s \in S : h\left(s\right) = 0\}$,

$$Y^*\left(s\right) = \underset{0 \leq y \leq X(s)}{\arg\max} \left[u\Big(W_0 - \Pi - X\left(s\right) + y\Big) - \lambda y h\left(s\right)\right]$$

Then the function $Z^* := Y^* \mathbf{1}_{A \setminus \{s \in S : h(s) = 0\}} + X \mathbf{1}_{A \cap \{s \in S : h(s) = 0\}}$ *solves Problem 3.19 with parameter* β.

Proof. Suppose that $Y^* \in B^+\left(\Sigma\right)$ satisfies (1), (2), and (3) above. Then Z^* is clearly feasible for Problem 3.19 with parameter β. To show optimality of Z^* for Problem 3.19 note that for any other $Y \in B^+\left(\Sigma\right)$ which is feasible for Problem 3.19 with parameter β, we have, for all $s \in A \setminus \{s \in S : h\left(s\right) = 0\}$,

$$\begin{aligned}
& u\Big(W_0 - \Pi - X\left(s\right) + Z^*\left(s\right)\Big) - u\Big(W_0 - \Pi - X\left(s\right) + Y\left(s\right)\Big) \\
& = u\Big(W_0 - \Pi - X\left(s\right) + Y^*\left(s\right)\Big) - u\Big(W_0 - \Pi - X\left(s\right) + Y\left(s\right)\Big) \\
& \geq \lambda\Big[h\left(s\right) Y^*\left(s\right) - h\left(s\right) Y\left(s\right)\Big] = \lambda\Big[h\left(s\right) Z^*\left(s\right) - h\left(s\right) Y\left(s\right)\Big]
\end{aligned}$$

Furthermore, since u is increasing, since $0 \leq Y \leq X$ on A, and since $Z^*\left(s\right) = X\left(s\right)$ for all $s \in \{s \in S : h\left(s\right) = 0\} \cap A$, it follows that for all $s \in \{s \in S : h\left(s\right) = 0\} \cap A$,

$$u\Big(W_0 - \Pi - X\left(s\right) + Z^*\left(s\right)\Big) = u\Big(W_0 - \Pi\Big) \geq u\Big(W_0 - \Pi - X\left(s\right) + Y\left(s\right)\Big)$$

Thus,

$$\int_{A \cap \{s \in S : h(s) = 0\}} u\Big(W_0 - \Pi - X + Z^*\Big) \, d\mu \; - \; \int_{A \cap \{s \in S : h(s) = 0\}} u\Big(W_0 - \Pi - X + Y\Big) \, d\mu \geq 0$$

Consequently,

$$
\begin{aligned}
&\int_A u\Big(W_0 - \Pi - X + Z^*\Big)\ d\mu - \int_A u\Big(W_0 - \Pi - X + Y\Big)\ d\mu \\
&\geq \int_{A\setminus\{s\in S: h(s)=0\}} u\Big(W_0 - \Pi - X + Z^*\Big)\ d\mu - \int_{A\setminus\{s\in S: h(s)=0\}} u\Big(W_0 - \Pi - X + Y\Big)\ d\mu \\
&\geq \lambda\Big[\beta - \beta\Big] = 0
\end{aligned}
$$

which completes the proof. □

Lemma 3.29. *For any $\lambda \geq 0$, the function given by*

$$
Y_\lambda^* := \min\left[X, \max\left(0, X - \left[W_0 - \Pi - \left(u'\right)^{-1}(\lambda h)\right]\right)\right] \tag{3.9}
$$

satisfies conditions (1) *and* (3) *of Lemma 3.28.*

Proof. This results directly from the fact that u is strictly increasing and concave. The proof is almost identical to the standard proof for the "classical" case, where the term h does not appear in the objective function. We refer the reader to any proof of the classical insurance problem, such as Bernard and Tian [34], pp. 75-76, for instance. The same methodology was also used in Bernard and Tian [33], pp. 722-724.

To prove this result directly, one has to solve an elementary Kuhn-Tucker problem of maximizing a concave function with two inequality constraints, which is exactly what the problem appearing in part (3) of Lemma 3.28 is. □

Lemma 3.30. *For Y_λ^* defined in equation* (3.9)*, the following holds:*

$$
Y_\lambda^* \mathbf{1}_{A\setminus\{s\in S: h(s)=0\}} + X\mathbf{1}_{A\cap\{s\in S: h(s)=0\}} = Y_\lambda^* \mathbf{1}_A \tag{3.10}
$$

Therefore,

$$
\int_A \left[Y_\lambda^* \mathbf{1}_{A\setminus\{s\in S: h(s)=0\}} + X\mathbf{1}_{A\cap\{s\in S: h(s)=0\}}\right]\ d\nu = \int_A Y_\lambda^*\ d\nu = \int_A Y_\lambda^* h\ d\mu \tag{3.11}
$$

Proof. Indeed, if $s \in \{s \in S : h(s) = 0\}$, then $(u')^{-1}(\lambda h(s)) = (u')^{-1}(0) = +\infty$, by Assumption 3.2. Thus, for each $s \in \{s \in S : h(s) = 0\}$ we have

$$Y_\lambda^*(s) = \min\left[X(s), \max\left(0, X(s) - \left[W_0 - \Pi - (u')^{-1}(0)\right]\right)\right] = X(s)$$

The rest then follows trivially. ☐

Lemma 3.31. *Define the function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ as follows: for each $\lambda \in \mathbb{R}^+$,*

$$\begin{aligned} \phi(\lambda) &:= \int_A \left[Y_\lambda^* \mathbf{1}_{A \setminus \{s \in S : h(s) = 0\}} + X \mathbf{1}_{A \cap \{s \in S : h(s) = 0\}}\right] \, d\nu \\ &= \int_A Y_\lambda^* \, d\nu = \int_A Y_\lambda^* h \, d\mu \end{aligned} \tag{3.12}$$

Then ϕ is a continuous nonincreasing function of the parameter λ.

Proof. First, recall that

$$Y_\lambda^* := \min\left[X, \max\left(0, X - \left[W_0 - \Pi - (u')^{-1}(\lambda h)\right]\right)\right] \tag{3.13}$$

Continuity of ϕ is a direct consequence of Lebesgue's Dominated Convergence Theorem and of the continuity of each of the functions $(u')^{-1}$, $\max(0, .)$, and $\min(x, .)$ (the function $(u')^{-1}$ is continuous by Remark 3.3). The fact that ϕ is nonincreasing in λ results from the concavity of u, i.e. from the fact that u' is a nonincreasing function. ☐

Lemma 3.32. *Consider the function ϕ defined above. Then:*

1. $\lim\limits_{\lambda \to 0} \phi(\lambda) = \int_A X \, d\nu$*; and,*

2. $\lim\limits_{\lambda \to +\infty} \phi(\lambda) = 0.$

Proof. By continuity of the functions $(u')^{-1}$, $\max(0,.)$, and $\min(x,.)$, we have that for each $s \in S$,

$$\lim_{\lambda\to 0} Y_\lambda^*(s) = \min\left[X(s), \max\left(0, X(s) - \left[W_0 - \Pi - (u')^{-1}(0)\right]\right)\right]$$

Moreover, as we showed above,

$$\min\left[X, \max\left(0, X - \left[W_0 - \Pi - (u')^{-1}(0)\right]\right)\right] = X$$

Therefore, $\lim_{\lambda\to 0} Y_\lambda^*(s) = X(s)$, for each $s \in S$. Hence, by continuity of the function ϕ in λ, it follows that

$$\lim_{\lambda\to 0} \phi(\lambda) = \int_A X \; d\nu$$

Similarly, by continuity of the functions $(u')^{-1}$, $\max(0,.)$, and $\min(x,.)$, we have that for each $s \in S$,

$$\lim_{\lambda\to +\infty} Y_\lambda^*(s) = \min\left[X(s), \max\left(0, X(s) - \left[W_0 - \Pi - (u')^{-1}(+\infty)\right]\right)\right]$$

However, by continuity of the function ϕ in λ, we have

$$\lim_{\lambda\to +\infty} \phi(\lambda) = \int_A \lim_{\lambda\to +\infty} Y_\lambda^* \; d\nu$$

But by Assumption 3.2, $(u')^{-1}(+\infty) = 0$, and by Assumption 3.5, $X \le W_0 - \Pi$, μ-a.s. Moreover, $\mu(A) = 1$. Therefore,

$$\int_A \lim_{\lambda\to +\infty} Y_\lambda^* \; d\nu = \int_A \lim_{\lambda\to +\infty} Y_\lambda^* \; h \; d\mu = 0$$

□

Remark 3.33. *Hence, summing up, the function ϕ defined above is a nonincreasing continuous function of the parameter λ such that $\lim_{\lambda\to 0} \phi(\lambda) = \int_A X\ d\nu$ and $\lim_{\lambda\to+\infty} \phi(\lambda) = 0$. Therefore, $\phi(\lambda) \in \left[0, \int_A X\ d\nu\right]$, and so by the Intermediate Value Theorem, for each $\beta \in \left(0, \min\left(\Pi/(1+\rho), \int_A X\ d\nu\right)\right]$ we can chose $\overline{\lambda} = \overline{\lambda}_\beta \in [0, +\infty)$ such that*

$$\beta = \phi\left(\ \overline{\lambda}\ \right) = \int \left[Y^*_{\overline{\lambda}} \mathbf{1}_{A\setminus\{s\in S: h(s)=0\}} + X\mathbf{1}_{A\cap\{s\in S: h(s)=0\}}\right] h\ d\mu \tag{3.14}$$

Therefore, by Lemmata 3.28 and 3.29, the function $Y^*_{\overline{\lambda}}$ defined above solves Problem 3.19, with parameter β. Finally, let β^* be optimal for Problem 3.21, let λ^* be chosen for β^* just as $\overline{\lambda}$ was chosen for β in Remark 3.33, and let $Y^*_{\lambda^*}$ be a corresponding optimal solution for Problem 3.20 with parameter β^*. The rest then follows from Remark 3.25. This concludes the proof of Theorem 3.13.

3.4.2 Proof of Corollary 3.14

Approximating $Y^*_{\overline{\lambda}}$

Fix $\beta \in \left(0, \min\left(\Pi/(1+\rho), \int_A X\ d\nu\right)\right]$, and let $\overline{\lambda}$ be the corresponding λ, chosen as in to Remark 3.33. Since h is nonnegative, Σ-measurable and μ-integrable, there is a sequence $\{h_n\}_n$ of nonnegative, μ-simple and μ-integrable functions on (S, Σ) that converges monotonically upwards and pointwise to h (Proposition D.6 on p. 224). Therefore, since u' is bicontinuous (so that, in particular, $(u')^{-1}$ is continuous), it follows that the sequence $\{Y_{\overline{\lambda},n}\}_n$, defined by

$$Y_{\overline{\lambda},n} := X - W_0 + \Pi + (u')^{-1}\left(\overline{\lambda} h_n\right), \tag{3.15}$$

converges pointwise to $Y_{\overline{\lambda}}$, defined by

$$Y_{\overline{\lambda}} := X - W_0 + \Pi + (u')^{-1}\left(\overline{\lambda} h\right) \tag{3.16}$$

Since the sequence $\{h_n\}_n$ converges monotonically upwards and pointwise to h, and since $(u')^{-1}$ is continuous and decreasing, it follows that the sequence $\{Y_{\overline{\lambda},n}\}_n$ converges monotonically downwards and pointwise to $Y_{\overline{\lambda}}$.

Now, for each $n \in \mathbb{N}$, there is some $m_n \in \mathbb{N}$, a Σ-partition $\{B_{i,n}\}_{i=1}^{m_n}$ of S, and some nonnegative real numbers $\alpha_{i,n} \geq 0$, for $i = 1, ..., m_n$, such that

$$h_n = \sum_{i=1}^{m_n} \alpha_{i,n} \; \mathbf{1}_{B_{i,n}} \tag{3.17}$$

Since $X - W_0 + \Pi$ can be written as $\sum_{i=1}^{m_n} (X - W_0 + \Pi) \, \mathbf{1}_{B_{i,n}}$, it is then easy to see that

$$Y_{\overline{\lambda},n} = \sum_{i=1}^{m_n} \Big((u')^{-1} \left(\overline{\lambda} \; \alpha_{i,n} \right) + X - W_0 + \Pi \Big) \, \mathbf{1}_{B_{i,n}} \tag{3.18}$$

Define $Y^*_{\overline{\lambda},n}$ by

$$Y^*_{\overline{\lambda},n} := \min \left[X, \max \left(0, Y_{\overline{\lambda},n} \right) \right] \tag{3.19}$$

By continuity of the functions $\max(0, .)$ and $\min(x, .)$, and since $\max(0, t)$ and $\min(X(s), t)$ are nondecreasing functions of t for each $s \in S$, it follows that the sequence $\{Y^*_{\overline{\lambda},n}\}_n$ converges monotonically downwards and pointwise to $Y^*_{\overline{\lambda}}$ (given by equation (3.9)).

Remark 3.34. *For each $n \geq 1$, let $\widetilde{Y}^*_{\overline{\lambda},n,\mu}$ denote the μ-a.s. unique nondecreasing μ-rearrangement of $Y^*_{\overline{\lambda},n}$ with respect to X. Then by Lemma 2.17, the sequence $\{\widetilde{Y}_{\overline{\lambda},n,\mu}\}_n$ converges monotonically downwards and pointwise μ-a.s. to $\widetilde{Y}_{\overline{\lambda},\mu}$.*

Note that, for each $n \in \mathbb{N}$, we can rewrite $Y^*_{\overline{\lambda},n}$ as

$$Y^*_{\overline{\lambda},n} = \sum_{i=1}^{m_n} I^*_{\overline{\lambda},n,i} \; \mathbf{1}_{B_{i,n}} \tag{3.20}$$

where, for $i = 1, ..., m_n$,

$$I^*_{\overline{\lambda},n,i} := \min \left[X, \max \left(0, X - d_{\overline{\lambda},n,i} \right) \right] \tag{3.21}$$

and

$$d_{\overline{\lambda},n,i} := W_0 - \Pi - (u')^{-1} \left(\overline{\lambda} \alpha_{i,n} \right) \tag{3.22}$$

Lemma 3.35. *For each $n \in \mathbb{N}$, and for each $i_0 \in \{1, 2, ..., m_n\}$, $I^*_{\overline{\lambda},n,i_0}$ is either a full insurance contract or a deductible contract (with a strictly positive deductible) on the set $B_{i_0,n}$.*

Proof. Fix $n \in \mathbb{N}$, and fix $i_0 \in \{1, 2, ..., m_n\}$. If $\alpha_{i_0,n} > 0$ and $\overline{\lambda} \le u'(W_0 - \Pi)/\alpha_{i_0,n}$, then since u' is decreasing (u is concave) it follows that

$$(u')^{-1}\left(\overline{\lambda}\alpha_{i_0,n}\right) \ge W_0 - \Pi$$

Therefore, $(u')^{-1}\left(\overline{\lambda}\alpha_{i_0,n}\right) - W_0 + \Pi + X \ge X \ge 0$, and so $I^*_{\overline{\lambda},n,i_0} = X$, a full insurance contract (on $B_{i_0,n}$).

If $\alpha_{i_0,n} = 0$, then $I^*_{\overline{\lambda},n,i_0} = \min\left[X, \max\left(0, (u')^{-1}(0) + X - W_0 + \Pi\right)\right]$. But $(u')^{-1}(0) = +\infty$, by Assumption 3.2. Therefore, $(u')^{-1}(0) - W_0 + \Pi + X \ge X \ge 0$, and so $I^*_{\overline{\lambda},n,i_0} = X$, a full insurance contract (on $B_{i_0,n}$).

If $\alpha_{i_0,n} > 0$ and $\overline{\lambda} > u'(W_0 - \Pi)/\alpha_{i_0,n}$, then since u' is strictly decreasing (u is strictly concave) it follows that $(u')^{-1}\left(\overline{\lambda}\alpha_{i_0,n}\right) < W_0 - \Pi$. Therefore, $0 < W_0 - \Pi - (u')^{-1}\left(\overline{\lambda}\alpha_{i_0,n}\right) = d_{\overline{\lambda},n,i_0}$, and so $I^*_{\overline{\lambda},n,i_0} = \left(X - d_{\overline{\lambda},n,i_0}\right)^+$, a deductible insurance contract (on $B_{i_0,n}$) with a strictly positive deductible, where for any $a, b \in \mathbb{R}$, $(a - b)^+ := \max(0, a - b)$. □

Remark 3.36. *Hence, we have constructed a sequence $\{Y^*_{\overline{\lambda},n}\}_n$ converging pointwise (on S and hence on A) to $Y^*_{\overline{\lambda}}$. Consequently, by Egoroff's theorem (Theorem D.12 on p. 227), for each $\varepsilon > 0$, there exists some $B_\varepsilon \in \Sigma$, $B_\varepsilon \subseteq A$, with $\mu(A \setminus B_\varepsilon) < \varepsilon$, such that $\{Y^*_{\overline{\lambda},n}\}_n$ converges to $Y^*_{\overline{\lambda}}$ uniformly on B_ε. In other words, for each $\varepsilon > 0$, there is some $B_\varepsilon \in \Sigma$, $B_\varepsilon \subseteq A$, with $\mu(A \setminus B_\varepsilon) < \varepsilon$, and there is some $N_\varepsilon \in \mathbb{N}$ such that for all $n \ge N_\varepsilon$, $|Y^*_{\overline{\lambda},n}(s) - Y^*_{\overline{\lambda}}(s)| < \varepsilon/2^n$, for all $s \in B_\varepsilon$.*

Rearrangement of the Approximation

The following lemma is a direct consequence of Lemmata 2.11 and 2.17, and it is hence stated without a proof.

Lemma 3.37. *If $\widetilde{Y}^*_{\overline{\lambda},n,\mu}$ (resp. $\widetilde{Y}^*_{\overline{\lambda},\mu}$) denotes the nondecreasing μ-rearrangement of $Y^*_{\overline{\lambda},n}$ (resp. $Y^*_{\overline{\lambda}}$) with respect to X, then $\{\widetilde{Y}^*_{\overline{\lambda},n,\mu}\}_n$ converges monotonically downwards and pointwise μ-a.s. to $\widetilde{Y}^*_{\overline{\lambda},\mu}$. Moreover, $\widetilde{Y}^*_{\overline{\lambda},n,\mu} = \widetilde{Y}^*_{\overline{\lambda},n,A,\mu}$, μ-a.s., where $\widetilde{Y}^*_{\overline{\lambda},n,A,\mu}$ denotes the nondecreasing μ-rearrangement of $Y^*_{\overline{\lambda},n,\mu}$ with respect to X on A.*

Let $C_{2,n} := \left\{ s \in S : Y^*_{\overline{\lambda},n}(s) = X(s) \right\}$. Then $C_{2,n}$ is of the form[5]

$$C_{2,n} = B_{k_1,n} \cup ... \cup B_{k_N,n} \tag{3.23}$$

for some $\{k_1, k_2, ..., k_N\} \subseteq \{1, 2, ..., m_n\}$. Therefore,

$$Y^*_{\overline{\lambda},n} = \sum_{j \in J} \left(X - d_{\overline{\lambda},n,j}\right)^+ \mathbf{1}_{B_{j,n}} + X\mathbf{1}_{C_{2,n}} \tag{3.24}$$

for $J = \{1, 2, ..., m_n\} \setminus \{k_1, k_2, ..., k_N\}$.

Lemma 3.38. *Fix $n \in \mathbb{N}$. If there exists some $i_0 \in \{1, 2, \ldots, m_n\}$ such that $\alpha_{i_0,n} = 0$ and $B_{i_0,n} \setminus \{s \in S : X(s) = 0\} \neq \varnothing$, then $C_{2,n} \setminus \{s \in S : X(s) = 0\} \neq \varnothing$.*

Proof. Trivial, in light of the second paragraph in the proof of Lemma 3.35. □

Lemma 3.39. *If μ is not absolutely continuous with respect to ν, then for each $n \in \mathbb{N}$ there is some $i_0 \in \{1, 2, ..., m_n\}$ such that $\alpha_{i_0,n} = 0$.*

Proof. Suppose, *per contra*, that μ is not absolutely continuous with respect to ν but that there is some $n \in \mathbb{N}$ such that $\alpha_{i_0,n} > 0$, for each $i_0 \in \{1, 2, ..., m_n\}$. Then $h_n = \sum_{i=1}^{m_n} \alpha_{i,n} \mathbf{1}_{B_{i,n}} > 0$. But the sequence $\{h_n\}_n$ converges monotonically upwards, and pointwise, to $h := d\nu_{ac}/d\mu$. Hence, since $h_n > 0$, it follows that $h(s) \geq h_k(s) > 0$, for each $s \in S$ and for each $k \geq n$. Consequently, $h > 0$. Therefore μ and ν_{ac} are mutually absolutely continuous (i.e. equivalent[6]). Furthermore, the finite measures ν, ν_{ac}, and ν_s are nonnegative, and hence $\nu_{ac} << \nu$. Thus, $\mu << \nu$, a contradiction. □

Lemma 3.40. *If $\mu = \nu$ then $C_{2,n} \setminus \{s \in S : X(s) = 0\} = \varnothing$, for each $n \geq 1$.*

[5] Note that since the random loss X is a mapping of S onto the closed interval $[0, M]$, it follows that $\{s \in S : X(s) = 0\} \neq \varnothing$, as mentioned previously (see section 3.2). Now, since $0 \leq Y_{\overline{\lambda},n} \leq X$, it follows that $\varnothing \neq \{s \in S : X(s) = 0\} \subseteq C_{2,n}$. Therefore, $C_{2,n} \neq \varnothing$.

[6] See, e.g. Bogachev [42] (p. 179).

Proof. Suppose that $\mu = \nu$. Then, in this case, $\nu_s = 0$, $\nu_{ac} = \nu = \mu$, and so $h = 1$ and $A = S$. Thus, $h_n = 1$, for all $n \in \mathbb{N}$.

We claim that $\overline{\lambda} > u'(W_0 - \Pi)$. Suppose, *per contra*, that $\overline{\lambda} \leq u'(W_0 - \Pi)$. Then by concavity of u, u' is decreasing, and so $(u')^{-1}\left(\overline{\lambda}\right) \geq W_0 - \Pi$. Therefore,

$$Y^*_{\overline{\lambda}} = \min\left[X, \max\left(0, X - \left[W_0 - \Pi - (u')^{-1}\left(\overline{\lambda}\right)\right]\right)\right] = X,$$

contradicting the classical result that a deductible insurance contract, with a positive deductible, is optimal in this case (as in Raviv [239] or Proposition 3.59). Therefore, $\overline{\lambda} > u'(W_0 - \Pi)$. But then from the proof of Lemma 3.35 it follows that $C_{2,n} \setminus \{s \in S : X(s) = 0\} = \varnothing$, for each $n \geq 1$. □

Now, let $C_{1,n} := \left\{s \in S : Y^*_{\overline{\lambda},n}(s) = 0\right\}$. Then $C_{1,n}$ is non-empty[7] and of the form

$$C_{1,n} = C^{(i)}_{1,n} \cup C^{(ii)}_{1,n} \tag{3.25}$$

Where $C^{(i)}_{1,n} \subseteq C_{2,n}$ and $C^{(ii)}_{1,n} \subseteq S \setminus C_{2,n}$. Indeed, since $\left\{s \in S : X(s) = 0\right\} \neq \varnothing$ and $0 \leq Y^*_{\overline{\lambda},n} \leq X$, it follows that for all $s \in \left\{s \in S : X(s) = 0\right\}$ we have $Y^*_{\overline{\lambda},n}(s) = X(s) = 0$. It is then easily verified that

$$C^{(i)}_{1,n} := \left\{s \in C_{2,n} : Y^*_{\overline{\lambda},n}(s) = 0\right\} = \left\{s \in S : X(s) = 0\right\} \neq \varnothing \tag{3.26}$$

Therefore, $C_{1,n} = \left\{s \in S : X(s) = 0\right\} \cup C^{(ii)}_{1,n}$. Moreover, we can write

$$C^{(ii)}_{1,n} = \bigcup_{j=k_{N+1}}^{k_Q} B_{j,n} \tag{3.27}$$

for some $\{k_{N+1}, ..., k_Q\} \subseteq J$. Letting $J' := J \setminus \{k_{N+1}, ..., k_Q\}$, it follows that

[7] Since $0 \leq Y^*_{\overline{\lambda},n} \leq X$ and $\{s \in S : X(s) = 0\} \neq \varnothing$.

$$0 < \left(X - d_{\overline{\lambda},n,j}\right)^{+} = X - d_{\overline{\lambda},n,j} < X \tag{3.28}$$

for each $j \in J'$. Therefore,

$$Y^{*}_{\overline{\lambda},n} = 0\mathbf{1}_{C_{1,n}} + \sum_{j \in J'} \left(X - d_{\overline{\lambda},n,j}\right) \mathbf{1}_{B_{j,n}} + X\mathbf{1}_{C_{2,n} \setminus \{s \in S : X(s) = 0\}} \tag{3.29}$$

We can assume, without loss of generality, that $\alpha_{j,n} < \alpha_{k,n}$, for all $j, k \in J'$ such that $j < k$. Then it is easily verified that $d_{\overline{\lambda},n,j} < d_{\overline{\lambda},n,k}$, because of the concavity of u.

Now, if $\widetilde{Y}^{*}_{\overline{\lambda},n,\mu}$ denotes the nondecreasing μ-rearrangement of $Y^{*}_{\overline{\lambda},n}$ with respect to X, we have the following result:

Lemma 3.41. *Let $\widetilde{Y}^{*}_{\overline{\lambda},n,\mu}$ denotes the nondecreasing μ-rearrangement of $Y^{*}_{\overline{\lambda},n}$ with respect to X. There exists $a_n \in [0, M]$ such that for μ-a.a. $s \in S$,*

$$\widetilde{Y}^{*}_{\overline{\lambda},n,\mu}(s) = \begin{cases} 0 \text{ if } X(s) \in [0, a_n) \\ f_n(X(s)) \text{ if } X(s) \in [a_n, M] \end{cases} \tag{3.30}$$

where $f_n : [0, M] \to [0, M]$ is a nondecreasing and Borel-measurable function such that $0 \leq f_n(t) \leq t$ for each $t \in [0, M]$, and, for $\mu \circ X^{-1}$-a.a. $t \in [0, M]$, we have $f(t) > 0$ if $t > a_n$.

Proof. First note that $0 \leq \widetilde{Y}^{*}_{\overline{\lambda},n,\mu} \leq X$, by Lemma 2.16, since $0 \leq Y^{*}_{\overline{\lambda},n} \leq X$, by definition of $Y^{*}_{\overline{\lambda},n}$. Moreover, we have $Y^{*}_{\overline{\lambda},n} = I_{\overline{\lambda},n} \circ X$, for some Borel-measurable function $I_{\overline{\lambda},n} : [0, M] \to [0, M]$. Therefore, $\widetilde{Y}^{*}_{\overline{\lambda},n,\mu} = \widetilde{I}^{*}_{\overline{\lambda},n} \circ X$, where $\widetilde{I}^{*}_{n}$ is the nondecreasing $\mu \circ X^{-1}$-rearrangement of $I_{\overline{\lambda},n}$. Let $f_n := \widetilde{I}^{*}_{\overline{\lambda},n}$. Then $0 \leq f_n(t) \leq t$, for each $t \in [0, M]$, and $f_n : [0, M] \to [0, M]$ is nondecreasing and Borel-measurable. Now, note that

$$\begin{aligned}\mu\Big(\big\{s\in S: Y^*_{\overline{\lambda},n}(s)\le 0\big\}\Big) &= \mu\Big(\big\{s\in S: Y^*_{\overline{\lambda},n}(s)= 0\big\}\Big)=\mu\left(C_{1,n}\right)\\ &=\mu\Big(\big\{s\in S: Y^*_{\overline{\lambda},n}(s)\le 0, X(s)=0\big\}\Big)\\ &\qquad+\mu\Big(\big\{s\in S: Y^*_{\overline{\lambda},n}(s)\le 0, X(s)>0\big\}\Big)\\ &=\mu\Big(\big\{s\in S: X(s)=0\big\}\Big)+\mu\left(C^{ii}_{1,n}\right)\\ &=\mu\left(C^{ii}_{1,n}\right)\end{aligned}$$

where the last equality follows form the nonatomicity of $\mu\circ X^{-1}$ (Assumption 3.5). Moreover, by equimeasurability, we have that

$$\mu\Big(\big\{s\in S: Y^*_{\overline{\lambda},n}(s)\le 0\big\}\Big)=\mu\Big(\big\{s\in S: \widetilde{Y}^*_{\overline{\lambda},n,\mu}\le 0\big\}\Big)$$

However,

$$\begin{aligned}\mu\Big(\big\{s\in S: \widetilde{Y}^*_{\overline{\lambda},n,\mu}(s)\le 0\big\}\Big) &= \mu\Big(\big\{s\in S: \widetilde{Y}^*_{\overline{\lambda},n,\mu}(s)= 0\big\}\Big)\\ &=\mu\Big(\big\{s\in S: \widetilde{Y}^*_{\overline{\lambda},n,\mu}(s)\le 0, X(s)=0\big\}\Big)\\ &\qquad+\mu\Big(\big\{s\in S: \widetilde{Y}^*_{\overline{\lambda},n,\mu}(s)\le 0, X(s)>0\big\}\Big)\\ &=\mu\Big(\big\{s\in S: X(s)=0\big\}\Big)\\ &\qquad+\mu\Big(\big\{s\in S: \widetilde{Y}^*_{\overline{\lambda},n,\mu}(s)= 0, X(s)>0\big\}\Big)\\ &=\mu\Big(\big\{s\in S: \widetilde{Y}^*_{\overline{\lambda},n,\mu}(s)= 0, X(s)>0\big\}\Big)\end{aligned}$$

where the last equality follows form the nonatomicity of $\mu\circ X^{-1}$ (Assumption 3.5). Consequently,

$$\mu\left(C_{1,n}\right)=\mu\left(C^{ii}_{1,n}\right)=\mu\Big(\big\{s\in S:\ \widetilde{Y}^*_{\overline{\lambda},n,\mu}(s)=0,\ X(s)>0\big\}\Big) \tag{3.31}$$

Thus, if $\mu\left(C^{ii}_{1,n}\right)\neq 0$, then there exists $a_n>0$ such that for μ-a.a. $s\in S$, $\widetilde{Y}^*_{\overline{\lambda},n,\mu}(s)=0$ if $X(s)$ belongs to $[0,a_n]$ or $[0,a_n)$, and $\widetilde{Y}^*_{\overline{\lambda},n,\mu}(s)>0$ if $X(s)>a_n$. Therefore, $f_n(t)>0$ if $t>a_n$, for $\mu\circ X^{-1}$-a.a. $t\in[0,M]$.

If $\mu\left(C^{ii}_{1,n}\right)=0$, then $\mu\Big(\big\{s\in S:\ \widetilde{Y}^*_{\overline{\lambda},n,\mu}(s)=0,\ X(s)>0\big\}\Big)=0$, and so for μ-a.a.

$s \in S$, $\widetilde{Y}^*_{\overline{\lambda},n,\mu}(s) = 0$ if $X(s) = 0$, and $\widetilde{Y}^*_{\overline{\lambda},n,\mu}(s) > 0$ if $X(s) > 0$. Thus, with $a_n = 0$, $\widetilde{Y}^*_{\overline{\lambda},n,\mu}$ is μ-a.s. of the form (3.30), with $f_n(t) > 0$ if $t > a_n = 0$, for $\mu \circ X^{-1}$-a.a. $t \in [0, M]$. $\square$

Remark 3.42. *For each $n \geq 1$, let $E_n \in \Sigma$ be the event such that $\mu(E_n) = 1$ and $\widetilde{Y}^*_{\overline{\lambda},n,\mu}$ is of the form* (3.30) *on E_n. Let $E := \bigcap_{n=1}^{+\infty} E_n$. Then $E \in \Sigma$ and, by Proposition 3.60, $\mu(E) = 1$. Moreover, for each $s \in E$, and for each $n \geq 1$, $\widetilde{Y}^*_{\overline{\lambda},n,\mu}(s)$ is given by* (3.30)*.*

Convergence to an Optimal Solution of Problem 3.11

By Lemma 3.37, the sequence $\{\widetilde{Y}^*_{\overline{\lambda},m,\mu}\}_m$ defined by equation (3.30) converges pointwise μ-a.s. to $\widetilde{Y}^*_{\overline{\lambda},\mu}$, the nondecreasing μ-rearrangement of $Y^*_{\overline{\lambda}}$ with respect to X.

Now, let $Y^*_{4,\beta}$ be an optimal solution to Problem 3.20 with parameter β, as defined previously, and for each $m \in \mathbb{N}$ let

$$\widetilde{Y}^*_{m,\beta} := \widetilde{Y}^*_{\overline{\lambda},m,\mu}\mathbf{1}_A + Y^*_{4,\beta}\mathbf{1}_{S\setminus A} \tag{3.32}$$

Finally, let β^* be optimal for Problem 3.21, let λ^* be chosen for β^* just as $\overline{\lambda}$ was chosen for β, and let Y^*_{4,β^*} be a corresponding optimal solution for Problem 3.20 with parameter β^*. For each $m \geq 1$, let

$$\widetilde{Y}^*_{m,\beta^*} := \widetilde{Y}^*_{\lambda^*,m,\mu}\mathbf{1}_A + Y^*_{4,\beta^*}\mathbf{1}_{S\setminus A} \tag{3.33}$$

Then, by Remark 3.25, the sequence $\{\widetilde{Y}^*_{m,\beta^*}\}_m$ converges pointwise μ-a.s. to an optimal solution of the initial problem (Problem 3.11), which is μ-a.s. nondecreasing in the loss X. Henceforth, we shall denote by $\mathcal{Y}^*$ that optimal solution. Then

$$\mathcal{Y}^*\mathbf{1}_A = \widetilde{Y}^*_{\lambda^*,\mu}\mathbf{1}_A \tag{3.34}$$

A Characterization of the Optimal Solution of Problem 3.11

To conclude the proof of Corollary 3.14, we now show that the optimal solution $\mathcal{Y}^*$ to Problem 3.11 obtained above has the form of a *generalized deductible* contract, μ-a.s. That is, we show that $\widetilde{Y}^*_{\lambda^*,\mu}$ has the form of a *generalized deductible* contract, μ-a.s.

Recall that $\widetilde{Y}^*_{\lambda^*,\mu}$ is the μ-a.s. unique nondecreasing μ-rearrangement of $Y^*_{\lambda^*}$ with respect to X, where

$$Y^*_{\lambda^*} := \min\left[X, \max\left(0, Y_{\lambda^*}\right)\right] \tag{3.35}$$

and

$$Y_{\lambda^*} := X - W_0 + \Pi + (u')^{-1}(\lambda^* h) \tag{3.36}$$

Moreover, the sequence $\{Y_{\lambda^*,m}\}_m$, defined by

$$Y_{\lambda^*,m} := X - W_0 + \Pi + (u')^{-1}(\lambda^* h_m), \tag{3.37}$$

converges pointwise to Y_{λ^*}. Since the sequence $\{h_m\}_m$ converges monotonically upwards and pointwise to h, and since $(u')^{-1}$ is continuous and decreasing, it follows that the sequence $\{Y_{\lambda^*,m}\}_m$ converges monotonically downwards and pointwise to Y_{λ^*}. Consequently, one can easily check that the sequence $\{Y^*_{\lambda^*,m}\}_m$ converges monotonically downwards and pointwise to $Y^*_{\lambda^*}$, where for each $m \geq 1$,

$$Y^*_{\lambda^*,m} := \min\left[X, \max\left(0, Y_{\lambda^*,m}\right)\right] \tag{3.38}$$

Remark 3.43. *For each $m \geq 1$, let $\widetilde{Y}^*_{\lambda^*,m,\mu}$ denote the μ-a.s. unique nondecreasing μ-rearrangement of $Y^*_{\lambda^*,m}$ with respect to X. Then by Lemma 2.17, the sequence $\{\widetilde{Y}_{\lambda^*,m,\mu}\}_m$ converges monotonically downwards and pointwise μ-a.s. to $\widetilde{Y}_{\lambda^*,\mu}$. That is, there is some $A^* \in \Sigma$ with $A^* \subseteq A$ and $\mu(A^*) = 1$, such that for each $s \in A^*$ the sequence $\{\widetilde{Y}_{\lambda^*,m,\mu}(s)\}_m$ converges monotonically downwards to $\widetilde{Y}_{\lambda^*,\mu}(s)$.*

Now, as in Lemma 3.41, for μ-a.a. $s \in S$,

$$\widetilde{Y}^*_{\lambda^*,n,\mu}(s) = \begin{cases} 0 \text{ if } X(s) \in [0, a_n) \\ f_n(X(s)) \text{ if } X(s) \in [a_n, M] \end{cases} \tag{3.39}$$

for a given $a_n \geq 0$, and $f_n : [0, M] \to [0, M]$, a nondecreasing and Borel-measurable function such that $0 \leq f_n(t) \leq t$ for each $t \in [0, M]$, and $f(t) > 0$ if $t > a_n$ for $\mu \circ X^{-1}$-a.a. $t \in [0, M]$.

Lemma 3.44. *The sequence $\{a_m\}_m$ is bounded and nondecreasing.*

Proof. Since $\{a_m\}_m \subset [0, M]$, boundedness of the sequence $\{a_m\}_m$ is clear. We show that it is nondecreasing. Fix $m \in \mathbb{N}$. Since the sequence $\{\widetilde{Y}_{\lambda^*,m,\mu}\}_m$ is nonincreasing pointwise on A^* (as in Remark 3.43), we have $\widetilde{Y}_{\lambda^*,m,\mu}(s) \geq \widetilde{Y}_{\lambda^*,m+1,\mu}(s)$, for each $s \in A^*$.

To show that $a_m \leq a_{m+1}$, first note that if $a_m = 0$, then $a_{m+1} \geq 0 = a_m$. If $a_m > 0$, let $E \in \Sigma$ be as in Remark 3.42, let $A^* \in \Sigma$ be as in Remark 3.43, and choose $s \in E \cap A^*$ such that $X(s) \in [0, a_m)$. Then $0 = \widetilde{Y}_{\lambda^*,m,\mu}(s) \geq \widetilde{Y}_{\lambda^*,m+1,\mu}(s) \geq 0$, and so $\widetilde{Y}_{\lambda^*,m+1,\mu}(s) = 0$. Consequently, $X(s) \in [0, a_{m+1}]$, and so $[0, a_m) \subseteq [0, a_{m+1}]$. Therefore, $0 < a_m \leq a_{m+1}$. □

Hence, the sequence $\{a_m\}_m$ is bounded and monotone. Therefore, it has a limit. Let

$$a := \lim_{m \to +\infty} a_m \tag{3.40}$$

Moreover, if there is some $n \geq 1$ such that $a_n > 0$, then for each $m \geq n$, we have $a_m \geq a_n > 0$.

Lemma 3.45. *With a as defined above, we have $0 \leq a \leq M$, and $a > 0$ if there is some $n \geq 1$ with $a_n > 0$.*

Proof. Since $0 \leq a_m \leq M$, for each $m \geq 1$, it follows that $0 \leq a \leq M$. Moreover, if there is some $n \geq 1$ such that $a_n > 0$, then for each $m \geq n$ we have $a_m \geq a_n > 0$. Therefore, $a \geq a_m > 0$, for each $m \geq n$, and so $a > 0$. □

Lemma 3.46. *There exist $a^* \geq 0$ such that for μ-a.a. $s \in S$,*

$$\mathcal{Y}^*(s) = \begin{cases} 0 \text{ if } X(s) \in [0, a^*) \\ f(X(s)) \text{ if } X(s) \in [a^*, M] \end{cases} \tag{3.41}$$

for some nondecreasing, left-continuous, and Borel-measurable function $f : [0, M] \to [0, M]$ such that $0 \leq f(t) \leq t$ for each $t \in [a^*, M]$.

Proof. Let $a^* := a$, where $a = \lim_{m \to +\infty} a_m$, as above, let $E \in \Sigma$ be as in Remark 3.42, let $A^* \in \Sigma$ be as in Remark 3.43, and let $E^* := E \cap A^*$. Suppose that there exists some $s_1 \in E^*$ such that $X(s_1) \in [0, a^*)$, but $\mathcal{Y}^*(s_1) > 0$. Then for each $m \geq 1$ we have $\widetilde{Y}_{\lambda^*,m,\mu}(s_1) > 0$, since the sequence $\{\widetilde{Y}_{\lambda^*,m,\mu}\}_m$ converges monotonically downwards and pointwise on E^* to $\widetilde{Y}_{\lambda^*,\mu}$ and $\mathcal{Y}^* \mathbf{1}_{E^*} = \widetilde{Y}^*_{\lambda^*,\mu} \mathbf{1}_{E^*}$, by definition of $\mathcal{Y}^*$. Consequently, $X(s_1) \geq a_m$, for each $m \geq 1$. Therefore, $X(s_1) \geq a^* = a = \lim_{m \to +\infty} a_m$, a contradiction. Hence, for each $s \in E^*$, $X(s) \in [0, a^*) \Rightarrow \mathcal{Y}^*(s) = 0$. Also, since $\mu(E) = \mu(A^*) = 1$, it follows form Proposition 3.60 that $\mu(E^*) = 1$.

Moreover, $\widetilde{Y}^*_{\lambda^*,\mu} = \widetilde{I} \circ X$, for some bounded, nonnegative, nondecreasing, left-continuous, and Borel-measurable function $\widetilde{I}$ on the range $[0, M]$ of X (see section 2.3.1). Let $f := \widetilde{I}$. We then have, for each $s \in E^*$, $\mathcal{Y}^*(s) = f(X(s))$ if $X(s) \in [a^*, M]$. Furthermore, since $0 \leq \widetilde{Y}^*_{\lambda^*,\mu} \leq X$, it follows that $0 \leq f(t) \leq t$, for each $t \in [0, M]$. In particular, $f(0) = 0$. □

Remark 3.47. *Note that If there is some $n \geq 1$ such that $a_n > 0$, then $a > 0$ by Lemma 3.45, and hence it follows from the definition of a^* that $a^* > 0$.*

Positivity of the Deductible Level a^*

Let $E \in \Sigma$ be as in Remark 3.42, let $A^* \in \Sigma$ be as in Remark 3.43, and let $E^* := E \cap A^*$, as above. For each $s_0 \in E^*$, define $\overline{L}(s_0)$ by:

$$\overline{L}(s_0) := \int_{E^*} \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1} \left(\frac{u'(W_0 - \Pi - X(s_0))}{h(s_0)} \; h \right) \right] \right) \right] \; h \, d\mu$$

Then

$$\overline{L}(s_0) = \int_{A} \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1} \left(\frac{u'(W_0 - \Pi - X(s_0))}{h(s_0)} \; h \right) \right] \right) \right] \; h \, d\mu$$

Now, let

$$\mathcal{D}_{E^*} := \left\{ s_0 \in E^* : X(s_0) > 0,\ h(s_0) > 0,\ \int_{E^*} Y^*\ h\ d\mu < \overline{L}(s_0) \right\}$$

Then

$$\mathcal{D}_{E^*} = \left\{ s_0 \in E^* : X(s_0) > 0,\ h(s_0) > 0,\ \int_{A} Y^*\ h\ d\mu < \overline{L}(s_0) \right\}$$

Suppose that $\mu(\mathcal{D}_{E^*}) \neq 0$. Then, in particular, $\mathcal{D}_{E^*} \neq \varnothing$. Fix some $s_0 \in \mathcal{D}_{E^*}$. Then $X(s_0) > 0$, $h(s_0) > 0$, and $\int_A \mathcal{Y}^*\ h\ d\mu < \overline{L}(s_0)$. In other words,

$$\beta^* = \phi(\lambda^*) = \int_A \mathcal{Y}^*\ h\ d\mu < \phi\Big(u'(W_0 - \Pi - X(s_0)) \Big/ h(s_0)\Big) = \overline{L}(s_0).$$

Therefore,

$$\lambda^* \geq u'(W_0 - \Pi - X(s_0)) \Big/ h(s_0),$$

since ϕ is a nonincreasing function. Consequently, $X(s_0) \leq W_0 - \Pi - (u')^{-1}(\lambda^* h(s_0))$, and so

$$Y^*_{\lambda^*}(s_0) = \min\left[X(s_0), \max\Big(0, X(s_0) - \Big[W_0 - \Pi - (u')^{-1}(\lambda^* h(s_0))\Big]\Big)\right] = 0$$

Hence, for each $s_0 \in \mathcal{D}_{E^*}$, we have $X(s_0) > 0$ and $Y^*_{\lambda^*}(s_0) = 0$. Since $\mu(\mathcal{D}_{E^*}) \neq 0$ by hypothesis, it follows that

$$\mu\Bigg(\Big\{ s \in E^* :\ X(s) > 0,\ Y^*_{\lambda^*}(s) = 0 \Big\}\Bigg) \neq 0$$

Thus, the fact that in this case we have $a^* > 0$ follows from the properties of the equimeasurable rearrangement (recall equation (3.34) and the proof of Lemma 3.41).

Now, let $\kappa = \lambda^*$, and define the set $\mathcal{E}_{E^*}$ as follows:

$$\mathcal{E}_{E^*} := \left\{ s_0 \in E^* : \ h(s_0) > 0, \ \kappa \, h(s_0) > u'(W_0 - \Pi), \right.$$
$$\left. 0 < X(s_0) < W_0 - \Pi - (u')^{-1}(\kappa \, h(s_0)) \right\}$$

Suppose that $\mu(\mathcal{E}_{E^*}) \neq 0$. Then, in particular, $\mathcal{E}_{E^*} \neq \varnothing$. Fix some $s_0 \in \mathcal{E}_{E^*}$. Then $h(s_0) > 0$, $\lambda^* > u'(W_0 - \Pi)/h(s_0)$, $X(s_0) > 0$, and $X(s_0) < W_0 - \Pi - (u')^{-1}(\lambda^* h(s_0))$. Since the sequence $\{h_n\}_n$ of nonnegative, μ-simple functions on (S, Σ) previously defined converges pointwise to h, we can choose n large enough so that $h_n(s_0)$ is close enough to $h(s_0)$ and the following hold:

1. $h_n(s_0) > 0$;
2. $\lambda^* > u'(W_0 - \Pi)/h_n(s_0)$; and,
3. $0 < X(s_0) < W_0 - \Pi - (u')^{-1}(\lambda^* h_n(s_0))$.

Therefore, from the proof of Lemma 3.35 (third paragraph), we have $X(s_0) > 0$ and $Y^*_{\lambda^*,n}(s_0) = 0$. Since $\mu(\mathcal{E}_{E^*}) \neq 0$ by hypothesis, it follows that

$$\mu\Bigg(\Big\{s \in E^* : \ X(s) > 0, \ Y^*_{\lambda^*,n}(s) = 0, \ \text{for some } n \geq 1\Big\}\Bigg) \neq 0$$

Thus, there exists $n^* \geq 1$ such that $\mu\Big(\big\{s \in E^* : \ X(s) > 0, \ Y^*_{\lambda^*,n^*}(s) = 0\big\}\Big) \neq 0$. For such n^*, we have $a_{n^*} > 0$ by properties of the equimeasurable rearrangement (as in the proof of Lemma 3.41), and by definition of the function $\widetilde{Y}^*_{\lambda^*,n^*,\mu}$ given in (3.30). This then yields $a > 0$ (by Lemma 3.45) and so $a^* > 0$. This completes the proof of Corollary 3.14.

Remark 3.48. *We have mentioned that the function f that appears in Corollary 3.14 can be characterized using the ideas developed in section 2.3.4. Indeed, the optimal solution appearing in Corollary 3.14 has been constructed as the limit of a sequence of rearrangements of nonnegative functions (each bounded by $M := \|X\|_s$). For ease of notation, let* *us refer to the initial sequence as $(k_m)_{m\geq 1}$, so that the sequence $\left(\widetilde{k}_m\right)_{m\geq 1}$ of nondecreasing rearrangements converges to the optimal solution appearing in Corollary 3.14. Now each element k_m of the initial sequence can in turn be approximated by a nondecreasing sequence $\{l_{m,i}\}_{i\geq 1}$ of Σ-simple nonnegative functions. By Proposition 2.20, we can completely characterize the nondecreasing rearrangement $\widetilde{l}_{m,i}$ of each one of these simple functions $l_{m,i}$. Each sequence $\{\widetilde{l}_{m,i}\}_{i\geq 1}$ of nondecreasing simple functions hence obtained converges to the nondecreasing rearrangement $\widetilde{k}_m$ of each element k_m of the initial sequence $\{k_m\}_{m\geq 1}$ (by Lemma 2.18).*

3.5 Some Special Cases

3.5.1 Perfect Intersubjectivity of Beliefs

When the DM's and the insurer's beliefs are perfectly intersubjective, in the sense that $\mu = \nu$, then we recover the classical setup where the random loss endured by the DM is a nonnegative, bounded and continuous random variable on an objective probability space (S, Σ, P).

In this case, $\nu_s = 0$, $\nu_{ac} = \nu = \mu$, and so $h = 1$ and $A = S$. Thus, $h_n = 1$, for all $n \in \mathbb{N}$, and the optimal indemnity is a deductible insurance contract, just as in Arrow [25] and Raviv [239][8].

3.5.2 Equivalent Subjective Beliefs

When the DM's and the insurer's beliefs are not perfectly intersubjective but only equivalent, in the sense that $\mu << \nu$ and $\nu << \mu$, then $\nu_s = 0$, $\nu_{ac} = \nu$, $A = S$, and $h > 0$, μ-a.s. (see Proposition E.23 on p. 244), and hence ν-a.s. Moreover, $d\mu/d\nu = h^{-1}$. This case is treated similarly to the general case, taking into consideration the fact that the measurable set $\{s \in S : h(s) = 0\}$ has measure 0 (for both μ and ν), and that there is no need for a "splitting" procedure with respect to the events A and $S \setminus A$ (as in Problems 3.19, 3.20, and 3.21, and Lemma 3.24) since $S \setminus A = \varnothing$.

[8]See also Bernard and Tian [34], Proposition 2.1 on p. 53 and Lemma C.1 on p. 75.

3.5.3 Absolute Continuity

Consider the case where the DM's subjective beliefs fail to capture events that are otherwise deemed possible by the insurer. This situation arises naturally in insurance markets where an insurer's past experience and expertise give him an advantage over the DM in the understanding of the insurable risk under consideration.

We will model this state of affairs as a situation where the DM's subjective probability μ is absolutely continuous with respect to the insurer's subjective probability ν, with a Radon-Nikodým derivative $r = d\mu/d\nu$. The motivation behind this assumption is that events of zero probability for the insurer are those trivial events that even the DM deems impossible, whereas there might exist events that are considered impossible by the DM but to which the insurer attaches a positive probability: the latter are precisely the DM's *Black Swan* events[9] that might not be so for the insurer. The DM's problem becomes the following:

Problem 3.49. *For a given loading factor $\rho > 0$,*

$$\sup_{Y \in B^{+}(\Sigma)} \left\{ \int u\Big(W_0 - \Pi - X + Y\Big) r \; d\nu \right\} :$$
$$\begin{cases} 0 \le Y \le X \\ \Pi \ge (1+\rho) \; \int Y \; d\nu \end{cases}$$

This case is treated similarly to the general case exposed in the previous section, albeit with a slight modification to account for the fact that the function r now appears in the integrand. Moreover, the "splitting" procedure requires an adjustment here, as will be apparent below.

Lemma 3.50. *Let Y^* be an optimal solution for Problem 3.49, and suppose that ν is (μ, X)-vigilant. Let $\widetilde{Y}^*_\mu$ be the nondecreasing μ-rearrangement of Y^* with respect to X. Then $\widetilde{Y}^*_\mu$ is optimal for Problem 3.49 and nondecreasing in the loss X.*

Proof. See Lemma 2.30. □

Let $B := \{s \in S : r(s) \neq 0\} = \{s \in S : r(s) > 0\}$, and consider the following three problems:

[9]The terminology is borrowed from Taleb [290].

Problem 3.51.

$$\sup_{Y\in B^+(\Sigma)}\left\{\int_B u\Big(W_0-\Pi-X+Y\Big)r\ d\nu\right\}:$$
$$\begin{cases} 0\le Y\le X,\ \ on\ B \\ \int_B Y\ d\nu=\beta\in\left[0,\min\left(\Pi/\left(1+\rho\right),\int_B X\ d\nu\right)\right] \end{cases}$$

Problem 3.52.

$$\sup_{Y\in B^+(\Sigma)}\left\{\int_{S\setminus B} u\Big(W_0-\Pi-X+Y\Big)r\ d\nu\right\}:$$
$$\begin{cases} 0\le Y\le X,\ \ on\ S\setminus B \\ \int_{S\setminus B} Y\ d\nu\le\min\left(\frac{\Pi}{1+\rho}-\beta,\int_{S\setminus B} X\ d\nu\right),\ \textit{for the same } \beta \textit{ as in Problem 3.51} \end{cases}$$

Problem 3.53.

$$\sup_{\beta}\left[F_B^*\left(\beta\right)+F_B^*\left(\frac{\Pi}{1+\rho}-\beta\right):0\le\beta\le\min\left(\Pi/\left(1+\rho\right),\int_B X\ d\nu\right)\right]:$$
$$\begin{cases} F_B^*\left(\beta\right)\ \textit{is the supremum value of Problem 3.51, for a fixed} \\ \qquad\qquad\qquad\qquad\beta\in\left[0,\min\left(\Pi/\left(1+\rho\right),\int_B X\ d\nu\right)\right] \\ F_B^*\left(\frac{\Pi}{1+\rho}-\beta\right)\ \textit{is the supremum value of Problem 3.52, for the same fixed } \beta \end{cases}$$

Lemma 3.54. *If β^* is optimal for Problem 3.53, Y_3^* is optimal for Problem 3.51 with parameter β^*, and Y_4^* is optimal for Problem 3.52 with parameter β^*, then $Y_2^*:=Y_3^*\mathbf{1}_B+Y_4^*\mathbf{1}_{S\setminus B}$ is optimal for Problem 3.49.*

Proof. Similar to the proof of Lemma 3.24. □

Remark 3.55. *By Lemmata 3.54 and 3.50, if ν is (μ,X)-vigilant, β^* is optimal for Problem 3.53, Y_3^* is optimal for Problem 3.51 with parameter β^*, Y_4^* is optimal for Problem 3.52 with parameter β^*, and $Y_2^*:=Y_3^*\mathbf{1}_B+Y_4^*\mathbf{1}_{S\setminus B}$, then $\widetilde{Y}_{2,\mu}^*$ is optimal for Problem 3.49 and nondecreasing in the loss X, where $\widetilde{Y}_{2,\mu}^*$ is the μ-a.s. unique nondecreasing μ-rearrangement of Y_2^* with respect to X.*

Now, in order to solve Problem 3.52, note that for any $s \in S \setminus B$ we have $r(s) = 0$, and so for any $Y \in B^+(\Sigma)$, we have $\int_{S \setminus B} u\Big(W_0 - \Pi - X + Y\Big) r \, d\nu = 0$. Therefore, for any $\beta \in \left[0, \min\left(\Pi/(1+\rho), \int_B X \, d\nu\right)\right]$, any Y which is feasible for Problem 3.52 with parameter β is also optimal for Problem 3.52 with parameter β. As was done in the general case, take for instance $Y_4^* := \min\left[X, \max\left\{0, X - \overline{d}_\beta\right\}\right]$, where $\overline{d}_\beta$ is chosen such that $\int_{S \setminus B} Y_4^* \, d\nu \leq \min\left(\frac{\Pi}{1+\rho} - \beta, \int_{S \setminus B} X \, d\nu\right)$. Then $Y_4^* \mathbf{1}_{S \setminus B}$ is optimal for Problem 3.20 with parameter β.

Note that the choice of $\overline{d}_\beta$ so that $\int_{S \setminus B} Y_4^* \, d\nu \leq \min\left(\frac{\Pi}{1+\rho} - \beta, \int_{S \setminus B} X \, d\nu\right)$ is justified by an argument similar to that used in the general case.

In order to solve Problem 3.51, we will use a "statewise" technique as in the general case.

Lemma 3.56. *If $Y^* \in B^+(\Sigma)$ satisfies the following:*

1. $0 \leq Y^*(s) \leq X(s)$ *on* B;

2. $\int_B Y^* \, d\nu = \beta$, *for some* $\beta \in \left[0, \min\left(\Pi/(1+\rho), \int_B X \, d\nu\right)\right]$; *and,*

3. *There exists some* $\lambda \geq 0$ *such that for each* $s \in B$,

$$Y^*(s) = \underset{0 \leq y \leq X(s)}{\arg\max} \left[u\Big(W_0 - \Pi - X(s) + y\Big) r(s) - \lambda y\right]$$

Then $Y^* \mathbf{1}_B$ solves Problem 3.51 with parameter β.

Proof. Fix some $\beta \in \left[0, \min\left(\Pi/(1+\rho), \int_B X \, d\nu\right)\right]$, and suppose that $Y^* \in B^+(\Sigma)$ satisfies (1), (2), and (3) above. Then Y^* is feasible for Problem 3.51 with parameter β. To show optimality of Y^* for Problem 3.51 with parameter β note that for any other $Y \in B^+(\Sigma)$ which is feasible for Problem 3.51 with parameter β, we have, for each $s \in B$,

$$\begin{aligned} &u\Big(W_0 - \Pi - X(s) + Y^*(s)\Big) r(s) - u\Big(W_0 - \Pi - X(s) + Y(s)\Big) r(s) \\ &\geq \lambda \left[Y^*(s) - Y(s)\right] \end{aligned}$$

Consequently,

$$\int_B u\Big(W_0 - \Pi - X + Y^*\Big) r \; d\nu - \int_B u\Big(W_0 - \Pi - X + Y\Big) r \; d\nu \geq \lambda \left[\beta - \beta\right] = 0$$

which completes the proof. □

Lemma 3.57. *For any $\lambda \geq 0$, the function given by:*

$$Y_\lambda^* := \min\left[X, \max\left(0, X - \left[W_0 - \Pi - \left(u'\right)^{-1}\left(\lambda/r\right)\right]\right)\right]\mathbf{1}_B \tag{3.42}$$

satisfies conditions (1) *and* (3) *of Lemma 3.56 (recall that* $B := \{s \in S : r\left(s\right) > 0\}$*).*

Proof. Trivial, since u is strictly increasing and concave. The proof is almost identical to the "classical" case. □

Now, as was done previously for the general case, by the continuity of the function $\int_B Y_\lambda^* \; d\nu$ in the parameter λ and by the Intermediate Value Theorem, for any $\beta \in \left[0, \min\left(\Pi/\left(1+\rho\right), \int_B X \; d\nu\right)\right]$ one can chose $\overline{\lambda}$ such that

$$\beta = \int_B Y_{\overline{\lambda}}^* \; d\nu \tag{3.43}$$

On B, r can be approximated by a sequence $\{r_n\}_n$ of ν-simple, ν-integrable, positive functions. The continuity of $\left(u'\right)^{-1}$ then gives an approximation of $Y_{\overline{\lambda}}^*$ on B, as was done previously. Indeed, for each $n \in \mathbb{N}$, r_n can be written as

$$r_n = \sum_{i=1}^{m_n} \alpha_{i,n} \mathbf{1}_{C_{i,n}} \tag{3.44}$$

for some $m_n \in \mathbb{N}$, a Σ-partition $\{C_{i,n}\}_{i=1}^{m_n}$ of S, and some positive real numbers $\alpha_{i,n} > 0$, for $i = 1, ..., m_n$. Then, we can write

$$\overline{\lambda}/r_n = \sum_{i=1}^{m_n} \left(\overline{\lambda}/\alpha_{i,n}\right) \mathbf{1}_{C_{i,n}} \tag{3.45}$$

By continuity of the map $x \mapsto 1/x$ on $(0, +\infty)$, it follows that $\overline{\lambda}/r_n \to \overline{\lambda}/r$, and hence that $(u')^{-1}\left(\overline{\lambda}/r_n\right) \to (u')^{-1}\left(\overline{\lambda}/r\right)$, by continuity of $(u')^{-1}$. Then, as done previously, $Y^*_{\overline{\lambda},n} \to Y^*_{\overline{\lambda}}$, where $Y^*_{\overline{\lambda},n} = \sum_{i=1}^{m_n} I^*_{\overline{\lambda},n,i}\mathbf{1}_{C_{i,n}}$; where, for $i = 1, ..., m_n$, $I^*_{\overline{\lambda},n,i} = \min\left[X, \max\left(0, X - d_{\overline{\lambda},n,i}\right)\right]$, and $d_{\overline{\lambda},n,i} = W_0 - \Pi - (u')^{-1}\left(\overline{\lambda}/\alpha_{i,n}\right)$. One can easily check, as was done previously, that for each $i_0 \in \{1, 2, ..., m_n\}$, $I^*_{\overline{\lambda},n,i_0}$ is either a full insurance contract or a deductible contract with a positive deductible. The rest then follows as was done previously.

3.6 Belief Subjectivity in the Work of Marshall [207]

The motivation behind Marshall [207] stems partly from his observation that heterogeneity of beliefs is an inherent trait of insurance markets. The way such heterogeneity in beliefs is introduced into the insurance model in Marshall [207] is very specific, and we quote the author for the sake of completeness (see Marshall [207], p. 261):

> "*The client's beliefs deviate from those of the insurer when she (the client) holds a probability density function $g(t)$ different from $f(t)$*"

where it is understood that $f(t)$ in that context is the probability density function that the insurer attributes to the (bounded) insurable loss X, and $g(t)$ is the probability density function that the DM attributes to X.

One immediate observation to make here is that Marshall [207] hence implicitly assumes that we are given a measurable space (S, Σ), two probability measures P_1 and P_2 representing the DM and the insurer's subjective beliefs, respectively, and a bounded random variable X, such that the laws $P_1 \circ X^{-1}$ and $P_2 \circ X^{-1}$ are both absolutely continuous with respect to the Lebesgue measure on the range of X. This is a first limitation of the scope of the problem.

A second limitation is the author's requirement that the DM be *more optimistic than* the insurer, which is essentially sort of a probabilistic consistency requirement. The author defines the notion of "*more optimistic than*" as follows (see Marshall [207], p. 262): If $G(t)$ denotes the distribution function associated with $f(t)$ and $G(t)$ that associated with $g(t)$, then we say that the DM is more optimistic than the insurer when for all s,

$$\int_0^s G(t)\ dt > \int_0^s F(t)\ dt \tag{3.46}$$

This occurs, for instance, when $g(t)$ is obtained from $f(t)$ by shifting probability mass from areas of large loss to areas of smaller loss, as the author notes. However, even under such a limitation, the author provides an example showing that optimum contracts might not be "insurance contracts", where the author calls a contract an "insurance contract" when it is a deductible contract.

The author then imposes a further restriction on the way in which the beliefs can differ: he assumes that, conditional on the event that the loss is nonzero, $f(t)$ and $g(t)$ agree, but that the probability of a zero loss is higher for the DM than for the insurer. In this case, the DM is of course *more optimistic than* the insurer.

The author then shows that under these limitations, an optimal contract is an "insurance contract". Needless to say, this is considerably more restrictive than the approach considered in this chapter.

3.7 Conclusion

The subjectivity of beliefs in problems of optimal insurance design was largely overlooked in the relevant actuarial literature. The classical actuarial approach to insurance design has traditionally assumed that the insurer and the insured share the same probabilistic beliefs about the realization of a given insurable loss. While Aumann's Agreement Theorem (Aumann [28]) might give a justification for such an assumption, it might be argued that in real-life situations, instances of divergence of beliefs are the most interesting ones to examine.

In this chapter we have shown that under a specific *probabilistic consistency* assumption on the subjective beliefs of the decision maker (DM) and the insurer, an optimal insurance contract has the form of a *generalized deductible contract* on an event to which the decision maker assigns full subjective probability. This *probabilistic consistency* requirement can be intuitively interpreted in terms of the credibility that the insurer gives to the decision maker's subjective assessment of the risk inherent in the insurance contract. It essentially stipulates that the insurer will not assign a higher *subjective risk assessment* to the insurance contract on an event which is certain for the decision maker and on which the decision maker's information about the aforementioned risk is "better" in a specific sense.

Technically, this assumption of *probabilistic consistency* between the DM and the insurer is essential to show existence of optimal indemnities which are comonotonic with the random loss X (a.s. for the DM), and hence to avoid probems of moral hazard arising from a downward misrepresentation of losses by the DM.

Vigilance of beliefs is implied by the assumption of a *monotone likelihood ratio*, as discussed in section 2.5. In the absence of this assumption of *probabilistic consistency* there might exist solutions to the optimal insurance design problem that are not nondecreasing in the loss, and hence moral hazard issues might arise.

The work done in this chapter implicitly assumed that there is no risk associated with the insurer's default on payments. Future research will incorporate such default risk in the present framework in order to generalize, for instance, the work done by Cummins and Mahul [85]. In Chapter 6 we will consider some other possible extensions of this chapter's setting that are left for future work.

3.8 Appendix: The "Classical" Insurance Demand Problem

The classical problem of demand for insurance considers two agents: an insurer and a DM. The DM seeks an insurance coverage against a random loss she is facing. The market gives the DM the opportunity to purchase a coverage I from an insurer, for a premium Π set by the latter.

The random loss is modeled as an essentially bounded nonnegative random variable X on some exogenously given probability space $(\Omega, \mathcal{G}, P)$. Both the DM and the insurer are assumed to know what the distribution of X is and to agree on that distribution, in the sense that both assign to X the law $P \circ X^{-1}$ on the range $\mathcal{D} \subseteq \mathbb{R}^+$ of X. The insurance coverage is modeled as a Borel-measurable mapping $I : \mathcal{D} \to \mathbb{R}^+$.

Moreover, the insurer is assumed to be a risk-neutral Expected-Utility (EU) maximizer, and the DM is assumed to be a strictly risk-averse EU maximizer, with a strictly increasing, concave, and twice continuously differentiable utility function u.

The DM has an initial wealth of amount W_0, and her wealth in state $\omega \in \Omega$ is given by

$$W(\omega) := W_0 - \Pi - X(\omega) + I(X(\omega)) \tag{3.47}$$

Since the DM's preference has an EU representation, the problem from the perspective of the DM is to find the indemnity I^* that will maximize her expected utility of wealth, subject to some feasibility and participation constraints. Formally, the problem is the following:

Problem 3.58. *For a given loading factor $\rho > 0$,*

$$\sup_I \left\{ \int u\Big(W_0 - \Pi - X + I \circ X\Big)\, dP \right\} :$$
$$\begin{cases} 0 \leq I \circ X \leq X \\ \Pi \geq (1+\rho) \int I \circ X \, dP \end{cases}$$

These constraints were previously discussed.

Optimality of the Deductible Contract

The classical literature has studied Problem 3.58 extensively. Starting with Arrow [25], Borch [52], and Raviv [239], it was shown that the solution to Problem 3.58 is given by a deductible contract, when the premium principle used by the insurer depends on the actuarial value of the indemnity, and when the DM is a risk-averse EU maximizer. Cummins and Mahul [86] show that if an additional upper limit constraint on the indemnity is imposed, then the optimal insurance contract is a capped deductible. The following proposition summarizes the classical theory:

Proposition 3.59. *There exists a deductible contract I_d, for some $d > 0$, such that $(I_d \circ X)$ is optimal for Problem 3.58.*

Proof. See Raviv [239], Corollary 1 on p. 90 (and the setting of his model on p. 86), for instance. See also Mossin [214] (pp. 561-562), Arrow [25] (p. 212), Gollier [154] (p. 372), Moffet [213] (p. 674), or Bernard and Tian [34] (p. 53). □

3.9 Appendix: Related Analysis

Proposition 3.60. *Let* (S, Σ, μ) *be a finite nonnegative measure space. If* $\{A_n\}_n \subset \Sigma$ *is such that* $\mu(A_n) = \mu(S)$*, for each* $n \geq 1$*, then* $\mu\left(\bigcap_{n=1}^{+\infty} A_n\right) = \mu(S)$*.*

Proof. Since for each $n \geq 1$ we have $\mu(A_n) = \mu(S)$, it follows that $\mu(S \setminus A_n) = 0$, for each $n \geq 1$. Therefore, since μ is nonnegative, and by countable subadditivity of countably additive measures[10], it follows that

$$0 \leq \mu\left(\bigcup_{n=1}^{+\infty} S \setminus A_n\right) \leq \sum_{n=1}^{+\infty} \mu(S \setminus A_n) = 0$$

Therefore, $\mu\left(\bigcap_{n=1}^{+\infty} A_n\right) = \mu(S) - \mu\left(\bigcup_{n=1}^{+\infty} S \setminus A_n\right) = \mu(S)$. □

[10]See, e.g. Cohn [82], Proposition 1.2.2 on p. 10.

Chapter 4

More on Vigilant Beliefs

In this chapter we provide a characterization of the mathematical structure underpinning collections of what we previously called *vigilant beliefs*. In particular, several *convergence* results are provided, as well as a characterization of *second-order vigilant beliefs* (i.e. beliefs about vigilant beliefs) on classes of vigilant beliefs.

The results presented here will serve as the basic mathematical tools for future work related to the notion of *vigilance*. This will be discussed in more detail in Chapter 6. For instance, in section 6.4 we outline the main ideas of an ongoing work aimed at characterizing a DM's preference among insurers that are *vigilant* with respect to the DM.

4.1 Preliminaries

The notation adopted here is standard and is consistent with the Appendices, where all necessary background material is given. Relevant references from which necessary background material is drawn include Bartle, Dunford, and Schwartz [30], Diestel [100] (chap. VII), Dunford and Schwartz [109] (sections IV.1, IV.2, IV.5, and IV.9), Gänssler [132], Maccheroni and Marinacci [194], Marinacci and Montrucchio [204], Megginson [210] (chap. 2), or Rao and Rao [238] (sections 2.2, 2.3, 2.4, 4.7, 9.1, and 9.2).

Let S denote the set of states of the world, and suppose that Σ is a σ-algebra of subsets of S, called events. Let $B(S)$ denote the Banach space of all bounded, $\mathbb{R}$-valued functions on S, normed by the supnorm $\|.\|_s$ (see Dunford and Schwartz [109], IV.2.13 on p. 240, and the first paragraph of p. 258). Let $B_0(\Sigma)$ denote the normed linear space of $\mathbb{R}$-valued, Σ-measurable simple functions on S (i.e. finite linear combinations of indicator functions of sets in Σ), endowed with the supnorm. Let $B(\Sigma)$ denote the collection of all bounded Σ-measurable $\mathbb{R}$-valued functions on S. Then $B(\Sigma)$ is a Banach space for $\|.\|_s$, by Proposition D.23 (p. 229). Let $B^+(\Sigma)$ denote the cone of nonnegative elements of $B(\Sigma)$.

Let $ba(\Sigma)$ denote the linear space of all bounded finitely additive set functions on Σ, endowed with the usual mixing operations. Let $ba^+(\Sigma)$ denote the cone of nonnegative elements of $ba(\Sigma)$, and let $ba_1^+(\Sigma)$ denote the collection of those elements μ of $ba^+(\Sigma)$ for which $\mu(S) = 1$. Elements of $ba_1^+(\Sigma)$ are the finitely additive probability charges on Σ. When endowed with the variation norm $\|.\|_v$, $ba(\Sigma)$ is a Banach space, by Proposition D.20 (p. 228).

By Theorem D.24 (p. 229), $(ba(\Sigma), \|.\|_v)$ is isometrically isomorphic to the norm-dual of $B(\Sigma)$ (respectively, $B_0(\Sigma)$), via the duality $<\phi, \mu>= \int \phi \, d\mu,\ \forall \mu \in ba(\Sigma),\ \forall \phi \in B(\Sigma)$ (respectively, $B_0(\Sigma)$). Consequently, we can endow $ba(\Sigma)$ with the weak* topologies $\sigma(ba(\Sigma), B(\Sigma))$ or $\sigma(ba(\Sigma), B_0(\Sigma))$. These two topologies coincide[1] on $ba_1^+(\Sigma)$. We can also endow $ba(\Sigma)$ with the weak topology $\sigma(ba(\Sigma), ba^*(\Sigma))$.

Let $ca(\Sigma)$ denote the linear subspace of $ba(\Sigma)$ consisting of countably additive set functions. When endowed with the variation norm $\|.\|_v$, $ca(\Sigma)$ is a Banach space, by Proposition D.20 (p. 228). In particular, $ca(\Sigma)$ is a $\|.\|_v$-closed linear subspace of $ba(\Sigma)$. Let $ca^+(\Sigma)$ denote the cone of nonnegative elements of $ca(\Sigma)$, and let $ca_1^+(\Sigma)$ denote the collection of those elements ν of $ca^+(\Sigma)$ for which $\nu(S) = 1$. Elements of $ca_1^+(\Sigma)$ are the countably additive probability measures on Σ. We can endow $ca(\Sigma)$ with the weak topology $\sigma(ca(\Sigma), ca^*(\Sigma))$, which coincides with the weak topology that $ca(\Sigma)$ inherits from the weak topology of $ba(\Sigma)$ (see Proposition C.16 on p. 219). In particular, weak compactness in $ca(\Sigma)$ is equivalent to weak compactness in $ba(\Sigma)$.

4.2 Some Definitions

Recall that two elements X and Y of $B(\Sigma)$ are called *comonotonic* if for all $s, t \in S$, $\Big[X(s) - X(t)\Big]\Big[Y(s) - Y(t)\Big] \geq 0$.

Definition 4.1. *Let $X \in B^+(\Sigma)$ be given, and let $M := \sup\{s \in S : X(s)\} < +\infty$. Let $\mu, \nu \in ca_1^+(\Sigma)$. For a given $B \in \Sigma$, let $\mathcal{A}_{B,\mu}$ be the subset of $B^+(\sigma\{X\}) \times B^+(\sigma\{X\})$*

[1]See Maccheroni, Marinacci and Rustichini [195], p. 1475, for instance.

consisting of all those pairs (W, Z) such that W and Z are identically distributed under μ on B, and Z is comonotonic with X on B. That is,

$$\mathcal{A}_{B,\mu} = \Bigg\{ (W, Z) \in B^+ (\sigma \{X\}) \times B^+ (\sigma \{X\}) : \\ \mu\Big(\{s \in B : W(s) \in \Gamma\} \Big) = \mu\Big(\{s \in B : Z(s) \in \Gamma\} \Big), \text{ for any Borel set } \Gamma, \\ \text{and } \Big[X(s) - X(t)\Big]\Big[Z(s) - Z(t)\Big] \geq 0, \text{ for all } s, t \in B \Bigg\} \tag{4.1}$$

The following is a slight generalization of the definition of *vigilance* in the context of the insurance problem of Chapter 3.

Definition 4.2. *Let $X \in B^+(\Sigma)$. For $\mu, \nu \in ca_1^+(\Sigma)$, we say that ν is (μ, X)-vigilant if for any $A \in \Sigma$ with $\mu(A) = 1$ and for any $(Y_1, Y_2) \in \mathcal{A}_{A,\mu}$, the following holds:*

$$\int_A Y_2 \; d\nu \leq \int_A Y_1 \; d\nu \tag{4.2}$$

Definition 4.3. *Let $\mu \in ca_1^+(\Sigma)$ and $X \in B^+(\Sigma)$ be given. We define the subsets $\mathcal{C}_\mu$ and $\mathcal{C}_\mu^*$ of $ca_1^+(\Sigma)$ as follows:*

$$\mathcal{C}_\mu = \Big\{ \nu \in ca_1^+(\Sigma) : \nu \text{ is } (\mu, X)\text{-vigilant (in the sense of Definition 4.2)} \Big\} \tag{4.3}$$

and

$$\mathcal{C}_\mu^* = \Big\{ \nu \in ca_1^+(\Sigma) : \int Y_2 \; d\nu \leq \int Y_1 \; d\nu, \; \forall (Y_1, Y_2) \in \mathcal{A}_{S,\mu} \Big\} \tag{4.4}$$

Remark 4.4. *In light of Definition 4.2, $\mathcal{C}_\mu$ can be written as:*

$$\mathcal{C}_\mu = \Bigg\{ \nu \in ca_1^+(\Sigma) : \int_A Y_2 \; d\nu \leq \int_A Y_1 \; d\nu, \\ \forall A \in \Sigma \text{ with } \mu(A) = 1, \; \forall (Y_1, Y_2) \in \mathcal{A}_{A,\mu} \Bigg\} \tag{4.5}$$

Furthermore, both $\mathcal{C}_\mu$ and $\mathcal{C}_\mu^*$ are non-empty since $\mu \in \mathcal{C}_\mu \cap \mathcal{C}_\mu^*$.

4.2.1 Some "Convergence" Properties of $\mathcal{C}_\mu$ and $\mathcal{C}_\mu^*$

Proposition 4.5. *The sets $\mathcal{C}_\mu$ and $\mathcal{C}_\mu^*$ are convex, norm-bounded, and hence weakly bounded and weak* bounded.*

Proof. The convexity of $\mathcal{C}_\mu$ and $\mathcal{C}_\mu^*$ is clear, and their norm-boundedness follows from the norm-boundedness of $ca_1^+(\Sigma)$. Consequently, $\mathcal{C}_\mu$ and $\mathcal{C}_\mu^*$ are also weakly bounded (by Proposition C.21 on p. 221). Finally, since $B(\Sigma)$ is a Banach space, $\mathcal{C}_\mu$ and $\mathcal{C}_\mu^*$ are also weak* bounded (also by Proposition C.21 on p. 221). □

Proposition 4.6. *The set $\mathcal{C}_\mu$ has the following properties:*

(i) *$\mathcal{C}_\mu$ is convex, norm-bounded and weak* bounded;*

(ii) *$\mathcal{C}_\mu$ is weak* compact and weakly compact;*

(iii) *$\mathcal{C}_\mu$ is weak* closed, weakly closed, and norm-closed;*

(iv) *$\mathcal{C}_\mu$ is weakly complete.*

Proof. Convexity, norm-boundedness, weak boundedness and weak* boundedness of $\mathcal{C}_\mu$ follow from Proposition 4.5. Let $\{\nu_\alpha\}_{\alpha\in\Gamma}$ be a net in $\mathcal{C}_\mu$ that converges to some $\nu \in ca(\Sigma)$ in the weak* topology. Then by the standard duality it follows that the net $\left\{\int \phi \; d\nu_\alpha\right\}_{\alpha\in\Gamma}$ converges to $\int \phi \; d\nu$, for all $\phi \in B(\Sigma)$.

Now, fix any $A \in \Sigma$ such that $\mu(A) = 1$ and choose any $(Y_1, Y_2) \in \mathcal{A}_{A,\mu}$. Then $Y_1, Y_2, Y_1\mathbf{1}_A, Y_2\mathbf{1}_A \in B(\Sigma)$. Hence, the net $\left\{\int Y_1\mathbf{1}_A \, d\nu_\alpha\right\}_{\alpha\in\Gamma}$ converges to $\int Y_1\mathbf{1}_A \, d\nu$, and the net $\left\{\int Y_2\mathbf{1}_A \, d\nu_\alpha\right\}_{\alpha\in\Gamma}$ converges to $\int Y_2\mathbf{1}_A \, d\nu$. Moreover, for each $\alpha \in \Gamma$ we have

$$0 \le \int_A Y_2 \, d\nu_\alpha = \int Y_2\mathbf{1}_A \, d\nu_\alpha \le \int_A Y_1 \, d\nu_\alpha = \int Y_1\mathbf{1}_A \, d\nu_\alpha$$

since $\nu_\alpha \in \mathcal{C}_\nu$ for each $\alpha \in \Gamma$. Therefore,

$$0 \le \int_A Y_2 \, d\nu = \lim_{\alpha\in\Gamma} \int_A Y_2 \, d\nu_\alpha \le \lim_{\alpha\in\Gamma} \int_A Y_1 \, d\nu_\alpha = \int_A Y_1 \, d\nu$$

and so $\nu \in \mathcal{C}_\mu$. Hence, $\mathcal{C}_\mu$ is weak* closed.

Since $\mathcal{C}_\mu$ is also norm bounded, it follows from the Banach-Alaoglu theorem (Theorem C.24 on p. 222) that $\mathcal{C}_\mu$ is weak* compact. Now from Theorem D.31 (p. 232) it follows that $\mathcal{C}_\mu$ is weakly compact. Therefore $\mathcal{C}_\mu$ is weakly closed (since the weak topology is Hausdorff and using Proposition B.16 on p. 206). Norm-closure of $\mathcal{C}_\mu$ follows from its weak closure. Finally, since $ca(\Sigma)$ is weakly complete (by Proposition D.28 on p. 231), and $\mathcal{C}_\mu$ is weakly closed, it follows that $\mathcal{C}_\mu$ is weakly complete. □

Proposition 4.7. *The set $\mathcal{C}_\mu^*$ has the following properties:*

(i) $\mathcal{C}_\mu^*$ is convex, norm-bounded and weak* bounded;

(ii) $\mathcal{C}_\mu^*$ is weak* compact and weakly compact;

(iii) $\mathcal{C}_\mu^*$ is weak* closed, weakly closed, and norm-closed;

(iv) $\mathcal{C}_\mu^*$ is weakly complete.

Proof. Immediate consequence of Proposition 4.6. □

4.3 Vigilance and Absolute Continuity

We now characterize the collection of all probability measures ν on (S, Σ) that are both absolutely continuous with respect to some $\mu \in ca_1^+(\Sigma)$ and (μ, X)-vigilant, for a given $X \in B^+(\Sigma)$.

Definition 4.8. *For a given $\mu \in ca_1^+(\Sigma)$, let*

$$ca(\Sigma, \mu) := \left\{\nu \in ca(\Sigma) : \nu << \mu\right\} \tag{4.6}$$

and

$$\mathcal{AC}_\mu := \left\{\nu \in \mathcal{C}_\mu : \nu << \mu\right\} \tag{4.7}$$

and

$$\mathcal{H} := \left\{h \in L_1(S, \Sigma, \mu) : h = d\nu/d\mu, \text{ for some } \nu \in \mathcal{AC}_\mu\right\} \tag{4.8}$$

Remark 4.9. *As in Remark 4.4, one can easily check that in light of Definition 4.2, $\mathcal{H}$ can be written as:*

$$\begin{aligned}\mathcal{H} = \Bigg\{ & h \in L_1(S, \Sigma, \mu) : h \geq 0, \ \mu\text{-a.s., and for any } A \in \Sigma \text{ with } \mu(A) = 1, \\ & \int_A Y_2 h \, d\mu \leq \int_A Y_1 h \, d\mu, \ \forall (Y_1, Y_2) \in \mathcal{A}_{A,\mu} \Bigg\}\end{aligned} \tag{4.9}$$

Note that for any $A \in \Sigma$ with $\mu(A) = 1$, and for any $(Y_1, Y_2) \in \mathcal{A}_{A,\mu}$, we have $\int_A Y_2 h \, d\mu = \int Y_2 h \, d\mu$ and $\int_A Y_1 h \, d\mu = \int Y_1 h \, d\mu$.

By the Radon-Nikodým theorem (Theorem E.22 on p. 244), there is an isometric isomorphism between the space $ca(S, \Sigma, \mu)$ and the space $L_1(S, \Sigma, \mu)$, given by $\nu(E) = \int_E f \, d\mu$, $\forall E \in \Sigma$ (see Corollary E.25 on p. 245). Hence, we can identify $\mathcal{AC}_\mu$ with $\mathcal{H}$.

Proposition 4.10. *If $\mathcal{AC}_\mu$ is uniformly absolutely continuous with respect to μ, then the set $\mathcal{AC}_\mu$ has the following properties:*

(i) $\mathcal{AC}_\mu$ is convex, norm-bounded and weak bounded;*

(ii) $\mathcal{AC}_\mu$ *is weakly sequentially compact;*

(iii) $\mathcal{AC}_\mu$ *is weak** *compact and weakly compact;*

(iv) $\mathcal{AC}_\mu$ *is weak** *closed, weakly closed, and norm-closed;*

(v) $\mathcal{AC}_\mu$ *is weakly complete.*

Proof. Convexity of $\mathcal{AC}_\mu$ is clear, and norm-boundedness (resp. weak boundedness, weak* boundedness) of $\mathcal{AC}_\mu$ follows from Proposition 4.5. For all $\nu \in \mathcal{AC}_\mu$, we have $\nu << \mu \in ca_1^+(\Sigma)$ and, given any $E \in \Sigma$, the limit $\lim_{\mu(E)\to 0} \nu(E) = 0$ is uniform with respect to $\nu \in \mathcal{AC}_\mu$, by hypothesis. Consequently, since $\mathcal{AC}_\mu$ is norm-bounded, it follows that $\mathcal{AC}_\mu$ is weakly sequentially compact (by Theorem D.27 on p. 231), and hence weakly compact by the Eberlein-Šmulian theorem (Theorem C.25 on p. 222). Now from Theorem D.31 (p. 232) it follows that $\mathcal{AC}_\mu$ is weak* compact. Therefore $\mathcal{AC}_\mu$ is weak* closed and weakly closed (since both of these topologies are Hausdorff, and using Proposition B.16 on p. 206). Norm-closure of $\mathcal{AC}_\mu$ follows from its weak closure. Finally, since $ca(\Sigma)$ is weakly complete (by Proposition D.28 on p. 231), and $\mathcal{AC}_\mu$ is weakly closed, it follows that $\mathcal{AC}_\mu$ is weakly complete. □

Remark 4.11. *In Proposition 4.10, if* $\mathcal{AC}_\mu$ *is countable, that is,* $\mathcal{AC}_\mu$ *is of the form* $\{\nu_n, n \geq 1\}$*, and if* $\lim_n \nu_n(A)$ *exists for each* $A \in \Sigma$*, then the requirement that* $\mathcal{AC}_\mu$ *be also uniformly absolutely continuous with respect to* μ *is superfluous by the Vitali-Hahn-Saks theorem (Theorem D.26 on p. 231).*

Proposition 4.12. $\mathcal{H}$ *is a convex, norm-closed and weakly closed subset of the Banach space* $L_1(S, \Sigma, \mu)$*.*

Proof. The convexity of $\mathcal{H}$ is clear. To show that $\mathcal{H}$ is closed in the L_1-norm simply note that by standard duality results, there is an isometric isomorphism between the topological dual of the space $L_1(S, \Sigma, \mu)$ and the space $L_\infty(S, \Sigma, \mu) \supseteq B(\Sigma)$ given by $T(f) = \int fg \, d\mu$, for every continuous linear functional T on $L_1(S, \Sigma, \mu)$ and for every $f \in L_1(S, \Sigma, \mu)$, where g is some element of $L_\infty(S, \Sigma, \mu)$ (see Theorem E.29 on p. 246).

Therefore, it follows that for any $Y_1, Y_2 \in \mathcal{A}_{A,\mu}$ the maps $T_1 : \mathcal{H} \to \mathbb{R}^+$ and $T_2 : \mathcal{H} \to \mathbb{R}^+$ defined by $T_1(h) = \int Y_1 h \, d\mu$ and $T_2(h) = \int Y_2 h \, d\mu$ are continuous in the L_1-norm topology, and hence sequentially continuous (by Proposition B.47 on p. 212). Now, choose a sequence $\{h_n\}_n$ of elements of $\mathcal{H}$ that converges in the L_1-norm to some $h \in L_1(S, \Sigma, \mu)$. Then by continuity of the maps T_1 and T_2, it follows that

$$\lim_{n\to+\infty} \int Y_1 h_n \, d\mu = \int Y_1 h \, d\mu \tag{4.10}$$

and

$$\lim_{n\to+\infty} \int Y_2 h_n \, d\mu = \int Y_2 h \, d\mu \tag{4.11}$$

Moreover, since $h_n \in \mathcal{H}$, for each $n \in \mathbb{N}$, it follows that for each $n \geq 1$

$$0 \leq \int Y_2 h_n \, d\mu \leq \int Y_1 h_n \, d\mu \tag{4.12}$$

and therefore, taking limits as $n \to +\infty$ yields

$$0 \leq \int Y_2 h \, d\mu \leq \int Y_1 h \, d\mu \tag{4.13}$$

Consequently, $h \in \mathcal{H}$, and hence $\mathcal{H}$ is closed in the norm-topology. Since $\mathcal{H}$ is also convex, it follows by Mazur's theorem (Theorem C.23 on p. 221) that it is also closed in the weak topology. □

Corollary 4.13. *$\mathcal{H}$ is both norm-complete and weakly complete.*

Proof. Being a norm-closed subset of the Banach space $L_1(S, \Sigma, \mu)$ (as shown in the previous proof), $\mathcal{H}$ is complete in the L_1-norm. Furthermore, the space $L_1(S, \Sigma, \mu)$ is weakly complete (see Theorem E.30 on p. 246), and hence $\mathcal{H}$ is also weakly complete, being weakly closed. □

4.4 Geometric Properties of Some Collections of Vigilant Beliefs

Both the weak and weak* topologies on $ca(\Sigma)$ are locally convex and Hausdorff linear topologies[2]. Consequently, in light of Propositions 4.6, 4.7 and 4.10, both the Krein-Milman

[2]See, e.g. Schaefer [270], p. 52, and note that the topological dual of $ca(\Sigma)$ for the strong (variation norm) topology on $ca(\Sigma)$ separates points of $ca(\Sigma)$ since any norm topology is a locally convex topology.

Theorem (Theorem F.17 on p. 251, and Corollary F.25 on p. 253) and the Choquet-Bishop-De Leeuw Theorem (Theorem F.27 on p. 253) can be used to characterize the collections $\mathcal{C}_\mu$, $\mathcal{C}_\mu^*$ and $\mathcal{AC}_\mu$ in terms of (i) their extreme points, and of (ii) second-order beliefs on them, i.e. beliefs about beliefs.

4.4.1 Extreme Points

Proposition 4.14. *For any $\nu \in \mathcal{C}_\mu$ there exists a net $\{\lambda_\alpha\}_{\alpha\in\Gamma}$ in $\mathcal{C}_\mu$ and a net $\{\phi_\beta\}_{\beta\in\Gamma}$ in $\mathcal{C}_\mu$ such that:*

1. *For each $\alpha \in \Gamma$, λ_α is a convex combination of extreme points of $\mathcal{C}_\mu$;*
2. *For each $\beta \in \Gamma$, ϕ_β is a convex combination of extreme points of $\mathcal{C}_\mu$;*
3. *The net $\{\lambda_\alpha\}_{\alpha\in\Gamma}$ converges to ν in the weak topology; and,*
4. *The net $\{\phi_\beta\}_{\beta\in\Gamma}$ converges to ν in the weak* topology.*

Proof. By Proposition 4.6, the set $\mathcal{C}_\mu$ is a nonempty, convex and weakly compact subset of $ca(\Sigma)$, which is a locally convex and Hausdorff topological vector space when equipped with its weak topology. Therefore, by the Krein-Milman Theorem, $\mathcal{C}_\mu$ is the weakly closed convex hull of the set of its extreme points. Therefore, there exists a net in $\mathcal{C}_\mu$ which converges to ν in the weak topology, and each element of which is a convex combination of extreme points of $\mathcal{C}_\mu$.

Similarly, by Proposition 4.6, the set $\mathcal{C}_\mu$ is a nonempty, convex and weak* compact subset of $ca(\Sigma)$, which is a locally convex and Hausdorff topological vector space when equipped with its weak* topology. Therefore, by the Krein-Milman Theorem, $\mathcal{C}_\mu$ is the weak* closed convex hull of the set of its extreme points. Therefore, there exists a net in $\mathcal{C}_\mu$ which converges to ν in the weak* topology, and each element of which is a convex combination of extreme points of $\mathcal{C}_\mu$. □

See also Rudin [248] (Theorem 3.10 on p. 62 and Section 3.14 on p. 66) or Aliprantis and Border [3] (pp. 211-212).

Remark 4.15. *By Proposition 4.6 and the Krein-Milman Theorem, the collection $\mathcal{C}_\mu$ is the weakly closed convex hull of the set of its extreme points. However, the weak and strong closure of a convex subset of a locally convex Hausdorff topological vector space coincide (see, e.g. Dunford and Schwartz [109], Theorem V.3.13 on p. 422, or Rudin [248], Theorem 3.12 on p. 64). Therefore, $\mathcal{C}_\mu$ is also the norm-closed convex hull of the set of its extreme points, and so for every $\nu \in \mathcal{C}_\mu$ there exists a net $\{\eta_\alpha\}_{\alpha\in\Gamma}$ in $\mathcal{C}_\mu$, each element of which is a convex combination of extreme points of $\mathcal{C}_\mu$, and which converges to ν in the strong (variation norm) topology.*

A net converges to some limit if and only if ever subnet of that net converges to the same limit (see, e.g. Willard [299], 11.4.e on p. 75). In particular, if a net converges to some limit, then every subsequence of that net converges to the same limit. We then have the following corollary:

Corollary 4.16. *For any $\nu \in \mathcal{C}_\mu$ there exist three sequences $\{\lambda_n\}_{n\geq 1}$, $\{\phi_n\}_{n\geq 1}$, and $\{\eta_n\}_{n\geq 1}$ in $\mathcal{C}_\mu$ such that:*

1. *For each $n \geq 1$, λ_n, ϕ_n, and η_n are convex combinations of extreme points of $\mathcal{C}_\mu$;*
2. *The sequence $\{\eta_n\}_{n\geq 1}$ converges to ν in the strong topology;*
3. *The sequence $\{\lambda_n\}_{n\geq 1}$ converges to ν in the weak topology; and,*
4. *The sequence $\{\phi_n\}_{n\geq 1}$ converges to ν in the weak* topology, that is, for any $\Psi \in B(\Sigma)$, $\lim_{n\to+\infty} \int \Psi \, d\phi_n = \int \Psi \, d\nu$.*

Proof. Immediate □

Remark 4.17. *By Proposition 4.7, the previous results also apply to the collection $\mathcal{C}^*_\mu$.*

Similarly, in light of Proposition 4.10 and the Krein-Milman Theorem, the following two results hold for the collection $\mathcal{AC}_\mu$:

Proposition 4.18. *For any $\nu \in \mathcal{AC}_\mu$ there exists three nets $\{\lambda_\alpha\}_{\alpha\in\Gamma}$, $\{\phi_\beta\}_{\beta\in\Gamma}$, and $\{\eta_\delta\}_{\delta\in\Gamma}$ in $\mathcal{AC}_\mu$ such that:*

1. *For each* $\alpha \in \Gamma$, λ_α *is a convex combination of extreme points of* $\mathcal{AC}_\mu$*;*

2. *For each* $\beta \in \Gamma$, ϕ_β *is a convex combination of extreme points of* $\mathcal{AC}_\mu$*;*

3. *For each* $\delta \in \Gamma$, η_δ *is a convex combination of extreme points of* $\mathcal{AC}_\mu$*;*

4. *The net* $\{\lambda_\alpha\}_{\alpha\in\Gamma}$ *converges to* ν *in the weak topology;*

5. *The net* $\{\phi_\beta\}_{\beta\in\Gamma}$ *converges to* ν *in the weak* topology; and,*

6. *The net* $\{\eta_\delta\}_{\delta\in\Gamma}$ *converges to* ν *in the strong topology.*

Proof. Immediate □

Corollary 4.19. *For any* $\nu \in \mathcal{AC}_\mu$ *there exist three sequences* $\{\lambda_n\}_{n\geq 1}$, $\{\phi_n\}_{n\geq 1}$, *and* $\{\eta_n\}_{n\geq 1}$ *in* $\mathcal{AC}_\mu$ *such that:*

1. *For each* $n \geq 1$, λ_n, ϕ_n, *and* η_n *are convex combinations of extreme points of* $\mathcal{AC}_\mu$*;*

2. *The sequence* $\{\eta_n\}_{n\geq 1}$ *converges to* ν *in the strong topology;*

3. *The sequence* $\{\lambda_n\}_{n\geq 1}$ *converges to* ν *in the weak topology; and,*

4. *The sequence* $\{\phi_n\}_{n\geq 1}$ *converges to* ν *in the weak* topology, that is, for any* $\Psi \in B(\Sigma)$, $\lim_{n\to+\infty} \int \Psi \, d\phi_n = \int \Psi \, d\nu$.

Proof. Immediate □

Remark 4.20. *It is well-known that the extreme points of* $ca_1^+(\Sigma)$ *are those probability measures on* (S, Σ) *taking only the values* 0 *and* 1*. An example of an extreme point for* $ca_1^+(\Sigma)$ *is a Dirac measure, i.e. a set function of the form*

$$\begin{aligned} \delta_x : \Sigma &\to [0,1] \\ A &\mapsto \delta_x(A) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \end{aligned} \tag{4.14}$$

for some $x \in S$. In general, not all extreme points of $ca_1^+(\Sigma)$ are Dirac measures[3]*. However, in many common situations the extreme points of the collection of probability measures are precisely the Dirac measures. These situations occur frequently in practice. For instance, the extreme points of the collection of all Borel probability measures on a Polish space (e.g. $\mathbb{R}$ with the usual metric) are the Dirac measures. More generally, the extreme points of the collection of all Borel probability measures on any Souslin space (Definition D.40 on p. 234, and Proposition D.41 on p. 234) are the Dirac measures*[4]*.*

4.4.2 Second-Order Beliefs

Corollary F.25 (p. 253) and Theorem F.27 (p. 253) can be used to characterize second-order beliefs on some collections of vigilant beliefs, as follows:

Proposition 4.21. *Let $\nu \in \mathcal{C}_\mu$, and let $\mathcal{E}(\mathcal{C}_\mu)$ denote the set of extreme points of $\mathcal{C}_\mu$. Denote by Σ the Borel σ-algebra on $\mathcal{C}_\mu$ generated by the weak topology, and denote by Σ^* the Borel σ-algebra on $\mathcal{C}_\mu$ generated by the weak* topology. Then there exist a regular probability measure λ_1 on Σ and a regular probability measure λ_2 on Σ^* such that:*

1. *$\phi(\nu) = \int \phi \, d\lambda_1$, for any linear and weakly continuous function ϕ on $\mathcal{C}_\mu$;*
2. *$\psi(\nu) = \int \psi \, d\lambda_2$, for any linear and weak* continuous function ψ on $\mathcal{C}_\mu$;*
3. *λ_1 is supported by the weak closure of $\mathcal{E}(\mathcal{C}_\mu)$; and,*
4. *λ_2 is suported by the weak* closure of $\mathcal{E}(\mathcal{C}_\mu)$.*

Proof. Immediate consequence of Corollary F.25 (p. 253) and Proposition 4.6. □

[3]Suppose, for instance, that S is uncountable and that Σ is the collection of all subsets of S that are either countable or whose complement is countable. Then Σ is a σ-algebra on S. Define the set function μ on Σ as follows: For each $A \in \Sigma$, let $\mu(A)$ be equal to 0 when A is countable and let $\mu(A)$ be equal to 1 when A is uncountable. Then μ is an extreme point of $ca_1^+(\Sigma)$ but is not a Dirac measure.

[4]See, e.g. Bugajski et al. [67] (pp. 3-4), Topsøe [292] (Theorem 11.1 on p. 48), Winkler [302] (p. 585) and Winkler [303] (p. 11). Adamski [1] studies certain types of measurable spaces in which the Dirac measures are precisely the zero-one measures.

By Proposition 4.7, a similar result holds for the collection $\mathcal{C}_\mu^*$.

Proposition 4.22. *Let $\nu \in \mathcal{C}_\mu$, and let $\mathcal{E}(\mathcal{C}_\mu)$ denote the set of extreme points of $\mathcal{C}_\mu$. Denote by Σ the Borel σ-algebra on $\mathcal{C}_\mu$ generated by the weak topology, and denote by Σ^* the Borel σ-algebra on $\mathcal{C}_\mu$ generated by the weak* topology. Denote by $\mathcal{B}$ the Baire σ-algebra on $\mathcal{C}_\mu$ generated by the weakly compact G_δ-s. Similarly, denote by $\mathcal{B}^*$ the Baire σ-algebra on $\mathcal{C}_\mu$ generated by the weak* compact G_δ-s (note that $\mathcal{B} \subseteq \Sigma$ and $\mathcal{B}^* \subseteq \Sigma^*$).*

Then there exist a regular probability measure λ_3 on Σ and a regular probability measure λ_4 on Σ^* such that:

1. *$\phi(\nu) = \int \phi \, d\lambda_3$, for any linear and weakly continuous function ϕ on $\mathcal{C}_\mu$;*
2. *$\psi(\nu) = \int \psi \, d\lambda_4$, for any linear and weak* continuous function ψ on $\mathcal{C}_\mu$;*
3. *λ_3 vanishes on every subset of $\mathcal{B}$ which is disjoint from $\mathcal{E}(\mathcal{C}_\mu)$; and,*
4. *λ_4 vanishes on every subset of $\mathcal{B}^*$ which is disjoint from $\mathcal{E}(\mathcal{C}_\mu)$.*

Proof. Immediate consequence of Theorem F.27 (p. 253) and Proposition 4.6. □

By Proposition 4.7, a similar result holds for the collection $\mathcal{C}_\mu^*$. Similarly, using Proposition 4.10, Corollary F.25 (p. 253), and Theorem F.27 (p. 253) we obtain the following:

Proposition 4.23. *Let $\nu \in \mathcal{AC}_\mu$, and let $\mathcal{E}(\mathcal{AC}_\mu)$ denote the set of extreme points of $\mathcal{AC}_\mu$. Denote by Σ the Borel σ-algebra on $\mathcal{AC}_\mu$ generated by the weak topology, and denote by Σ^* the Borel σ-algebra on $\mathcal{AC}_\mu$ generated by the weak* topology. Then there exist a regular probability measure λ_1 on Σ and a regular probability measure λ_2 on Σ^* such that:*

1. *$\phi(\nu) = \int \phi \, d\lambda_1$, for any linear and weakly continuous function ϕ on $\mathcal{AC}_\mu$;*
2. *$\psi(\nu) = \int \psi \, d\lambda_2$, for any linear and weak* continuous function ψ on $\mathcal{AC}_\mu$;*
3. *λ_1 is supported by the weak closure of $\mathcal{E}(\mathcal{AC}_\mu)$; and,*
4. *λ_2 is suported by the weak* closure of $\mathcal{E}(\mathcal{AC}_\mu)$.*

Proof. Immediate ∎

Proposition 4.24. *Let $\nu \in \mathcal{AC}_\mu$, and let $\mathcal{E}\left(\mathcal{AC}_\mu\right)$ denote the set of extreme points of $\mathcal{AC}_\mu$. Denote by Σ the Borel σ-algebra on $\mathcal{AC}_\mu$ generated by the weak topology, and denote by Σ^* the Borel σ-algebra on $\mathcal{AC}_\mu$ generated by the weak* topology. Denote by $\mathcal{B}$ the Baire σ-algebra on $\mathcal{AC}_\mu$ generated by the weakly compact G_δ-s. Similarly, denote by $\mathcal{B}^*$ the Baire σ-algebra on $\mathcal{AC}_\mu$ generated by the weak* compact G_δ-s (note that $\mathcal{B} \subseteq \Sigma$ and $\mathcal{B}^* \subseteq \Sigma^*$).*

Then there exist a regular probability measure λ_3 on Σ and a regular probability measure λ_4 on Σ^* such that:

1. *$\phi\left(\nu\right) = \int \phi \, d\lambda_3$, for any linear and weakly continuous function ϕ on $\mathcal{AC}_\mu$;*
2. *$\psi\left(\nu\right) = \int \psi \, d\lambda_4$, for any linear and weak* continuous function ψ on $\mathcal{AC}_\mu$;*
3. *λ_3 vanishes on every subset of $\mathcal{B}$ which is disjoint from $\mathcal{E}\left(\mathcal{AC}_\mu\right)$; and,*
4. *λ_4 vanishes on every subset of $\mathcal{B}^*$ which is disjoint from $\mathcal{E}\left(\mathcal{AC}_\mu\right)$.*

Proof. Immediate ∎

Chapter 5

Contracting for Innovation under Heterogeneity and Ambiguity

5.1 Introduction

The purpose of this chapter is to show how the techniques of Chapter 2 can be used beyond problems of insurance demand. Specifically, we will consider a problem of *contracting for innovation* under heterogeneous and ambiguous beliefs, as will be explained below.

In Chapter 2 both the CI and the DM had beliefs that were unambiguous, in the sense that these beliefs were represented by an *additive* measure. In reality, however, it is often the case that one party's preferences over the acts of choice (i.e. the contracts) induce a representation of beliefs that reflects some level of *ambiguity*, while the other party's beliefs are *unambiguous*. Nevertheless, this is not identical to a situation of information asymmetry. It simply is a state of affairs whereby (i) the two parties have divergent beliefs, and (ii) one party's beliefs are *ambiguous*.

The situation that we will examine here can be summarized as follows: a DM is an *innovator* who improves upon a certain commodity for which there is a market. This improvement may be due to technological or other advances. A CI wishes to purchase that innovation, or the rights for using that innovation. The *innovation* will be modeled as a random variable X (see below). A contract between the DM and the CI is a pair (H, Y), where $H > 0$ is a fixed fee that the CI pays as a lump-sum upon entering into the contract with the DM. The DM then promises to transfer to the CI the amount $X(s)$, in the state of the world $s \in S$, in return of which the CI promises to pay the contingent amount $Y(s) = I(X(s))$, for some predetermined payment scheme I. The DM's problem is that of determining the optimal payment scheme I, or equivalently the optimal contingent claim $Y = I \circ X$.

Additionally, the CI's preferences admit a representation in terms of a *Choquet-Expected Utility* (CEU), as in Schmeidler [275] and Gilboa [148] (see Appendix A.3.2), hence exhibiting a certain level of *ambiguity*.

We will then illustrate how we can still use some of the techniques of Chapter 2 in this setting so as to show the existence of a monotone solution to the DM's problem. Additionally, we will consider a special case and characterize the solution to the DM's problem in that case using a similar "splitting" method to the one used in Chapter 2 and Chapter 3.

Finally, we also consider a situation where both the DM and the CI have preferences over the elements of choice that admit a representation which exhibits some ambiguity in their beliefs. We will extend the definition of *vigilance* to such a situation, and we will then show how in this case for each feasible solution to the DM's problem there corresponds at least one other feasible solution which is a nondecreasing function of the underlying uncertainty. However, to prove this result we will need to extend the notion of an equimeasurable rearrangement to the case where we have a non-additive measure instead of a measure (this is done in Appendix 5.8).

Outline

In section 5.2 we introduce definitions as well as the general setup. In section 5.3 we state the DM's problem and use the techniques of Chapter 2 to show the existence and monotonicity of a solution. In section 5.4, we consider a special case, and characterize the solution in that case using a "splitting" procedure as was done in Chapter 2. In section 5.5 we briefly consider yet another extension, whereby both the DM and the CI have ambiguous beliefs, and we show the existence of a monotone solution to the DM's problem. Finally, section 5.6 concludes. Appendix 5.7 contains a useful theorem that will be used in this chapter. Appendix 5.8 extends the idea of an equimeasurable rearrangement to a situation where we have a capacity (defined below) instead of a measure. This will be used in the proof of the main Theorem of section 5.5.

5.2 Preliminaries and Setup

In this section we recall the definition of a *capacity* and of the *Choquet integral*, which are the building blocks the CEU model and of the contracting model we will consider in this chapter.

5.2.1 Capacities and the Choquet Integral

Definition 5.1. *A capacity on* $(S, \mathcal{G})$ *is a set function* $\nu : \mathcal{G} \to [0, 1]$ *such that*

1. $\nu(\varnothing) = 0$;

2. $\nu(S) = 1$; *and,*

3. ν *is monotone: for any* $A, B \in \mathcal{G}$, $A \subseteq B \Rightarrow \nu(A) \leq \nu(B)$.

Remark 5.2. *An example of a capacity on a measurable space* $(S, \mathcal{G})$ *is a set function* $\nu := T \circ P$, *where* P *is a probability measure on* $(S, \mathcal{G})$ *and* $T : [0, 1] \to [0, 1]$ *is increasing with* $T(0) = 0$ *and* $T(1) = 1$.

Definition 5.3. *A capacity* ν *on* $(S, \mathcal{G})$ *is said to be continuous from above if for any sequence* $\{A_n\}_n$ *in* $\mathcal{G}$ *such that* $A_{n+1} \subseteq A_n$ *for each* $n \geq 1$, *we have:*

$$\lim_{n \to +\infty} \nu(A_n) = \nu\left(\bigcap_{n=1}^{+\infty} A_n\right) \tag{5.1}$$

A capacity ν *on* $(S, \mathcal{G})$ *is said to be continuous from below if for any sequence* $\{A_n\}_n$ *in* $\mathcal{G}$ *such that* $A_n \subseteq A_{n+1}$ *for each* $n \geq 1$, *we have:*

$$\lim_{n \to +\infty} \nu(A_n) = \nu\left(\bigcup_{n=1}^{+\infty} A_n\right) \tag{5.2}$$

Finally, a capacity ν *on* $(S, \mathcal{G})$ *is said to be continuous if it is both continuous from above and continuous from below.*

Remark 5.4. *If P is a probability measure on $(S, \mathcal{G})$ and $T : [0,1] \to [0,1]$ is increasing and continuous, with $T(0) = 0$ and $T(1) = 1$, then the set function $\nu := T \circ P$ is a capacity on $(S, \mathcal{G})$ which is continuous. This is an immediate consequence of the continuity of the measure P for monotome sequences*[1] *and the continuity of T.*

Definition 5.5. *A capacity ν on $(S, \mathcal{G})$ is said to be submodular if for each $A, B \in \mathcal{G}$,*

$$\nu (A \cup B) + \nu (A \cap B) \leq \nu (A) + \nu (B) \tag{5.3}$$

Remark 5.6. *If P is a probability measure on $(S, \mathcal{G})$ and $T : [0,1] \to [0,1]$ is increasing and concave, with $T(0) = 0$ and $T(1) = 1$, then the set function $\nu := T \circ P$ is a capacity on $(S, \mathcal{G})$ which is submodular*[2]*.*

Definition 5.7. *For a given capacity ν and a given $\psi \in B^+ (\mathcal{G})$, the Choquet integral $\oint \psi \; d\nu$ of ψ with respect to ν is defined by*

$$\oint \psi \; d\nu := \int_0^{+\infty} \nu \left(\{ s \in S : \psi (s) \geq t \} \right) \; dt \tag{5.4}$$

Moreover, for any $A \in \mathcal{G}$, $\oint_A \psi \; d\nu$ will be defined by:

$$\oint_A \psi \; d\nu := \oint \psi \; \mathbf{1}_A \; d\nu \tag{5.5}$$

If $\phi \in B (\mathcal{G})$, then the Choquet integral $\oint \phi \; d\nu$ of ϕ with respect to ν is defined by

$$\oint \phi \; d\nu := \int_0^{+\infty} \nu \left(\{ s \in S : \phi (s) \geq t \} \right) \; dt + \int_{-\infty}^{0} \left[\nu \left(\{ s \in S : \phi (s) \geq t \} \right) - 1 \right] \; dt \tag{5.6}$$

Moreover, for any $A \in \mathcal{G}$, $\oint_A \phi \; d\nu$ will be defined by:

$$\oint_A \phi \; d\nu := \oint \phi \; \mathbf{1}_A \; d\nu \tag{5.7}$$

[1] See, e.g. Cohn [82], Proposition 1.2.3 on p. 11.
[2] See, e.g. Denneberg [96], Example 2.1 on pp. 16-17.

Remark 5.8. *The Choquet integral with respect to a measure is simply the usual Lebesgue integral with respect to that measure*[3]*. The Choquet integral is a well-defined Riemann integral*[4]*. Moreover, for any capacity ν on $(S, \mathcal{G})$ and for any $\psi \in B^+(\mathcal{G})$, the following holds*[5]*:*

$$\oint \psi \, d\nu = \int_0^{+\infty} \nu\left(\{s \in S : \psi(s) \geq t\}\right) \, dt = \int_0^{+\infty} \nu\left(\{s \in S : \psi(s) > t\}\right) \, dt \tag{5.8}$$

Furthermore, for any capacity ν on $(S, \mathcal{G})$ and for any $\phi \in B(\mathcal{G})$, the following holds[6]*:*

$$\oint \psi \, d\nu = \int_0^{+\infty} \nu\left(\{s \in S : \phi(s) > t\}\right) \, dt + \int_{-\infty}^{0} \left[\nu\left(\{s \in S : \phi(s) > t\}\right) - 1\right] \, dt \tag{5.9}$$

Finally, as a functional on $B(\mathcal{G})$, the Choquet integral (with respect to some given capacity) is supnorm-continuous being Lipschitz continuous[7]*.*

Definition 5.9. *Two functions $Y_1, Y_2 \in B(\mathcal{G})$ are said to be comonotonic if*

$$\Big[Y_1(s) - Y_1(s')\Big]\Big[Y_2(s) - Y_2(s')\Big] \geq 0, \text{ for all } s, s' \in S \tag{5.10}$$

Similarly, two functions $Y_1, Y_2 \in B(\mathcal{G})$ are said to be anti-comonotonic if

$$\Big[Y_1(s) - Y_1(s')\Big]\Big[Y_2(s) - Y_2(s')\Big] \leq 0, \text{ for all } s, s' \in S \tag{5.11}$$

For instance any $Y \in B(\mathcal{G})$ is comonotonic with any $c \in \mathbb{R}$. Moreover, if $Y_1, Y_2 \in B(\mathcal{G})$, and if Y_2 is of the form $Y_2 = I \circ Y_1$, for some Borel-measurable function I, then Y_2 is comonotonic (resp. anti-comonotonic) with Y_1 if and only if the function I is nondecreasing (resp. nonincreasing).

[3] See, e.g. Marinacci and Montrucchio [204], p. 59.
[4] See, e.g. Marinacci and Montrucchio [204], p. 61.
[5] See, e.g. Marinacci and Montrucchio [204], Proposition 4.8 on p. 60.
[6] See, e.g. Marinacci and Montrucchio [204], p. 60.
[7] See, e.g. Marinacci and Montrucchio [204], Proposition 4.11–(iv) on p. 64.

Proposition 5.10. *Let ν be a capacity on $(S, \mathcal{G})$. If $\phi_1, \phi_2 \in B(\mathcal{G})$ are comonotonic, then the following holds:*

$$\oint (\phi_1 + \phi_2) \, d\nu = \oint \phi_1 \, d\nu + \oint \phi_2 \, d\nu \tag{5.12}$$

Proof. See Marinacci and Montrucchio [204], Theorem 4.3 on p. 66. □

Proposition 5.11. *Let ν be a capacity on $(S, \mathcal{G})$. If $\phi \in B(\mathcal{G})$ and $c \in \mathbb{R}$, then the following holds:*

$$\oint (\phi + c) \, d\nu = \oint \phi \, d\nu + c \tag{5.13}$$

Proof. See Denneberg [96], Proposition 5.1–(v) on p. 65. This is an immediate consequence of the definition of the Choquet integral and of Proposition 5.10. □

Proposition 5.12. *Let ν be a capacity on $(S, \mathcal{G})$. Then the following hold:*

1. *If $A \in \mathcal{G}$ then*

$$\oint \mathbf{1}_A \, d\nu = \nu(A) \tag{5.14}$$

2. *If $\phi \in B(\mathcal{G})$ and $a \geq 0$, then*

$$\oint a \, \phi \, d\nu = a \oint \phi \, d\nu \tag{5.15}$$

3. *If $\phi_1, \phi_2 \in B(\mathcal{G})$ are such that $\phi_1 \leq \phi_2$, then*

$$\oint \phi_1 \, d\nu \leq \oint \phi_2 \, d\nu \tag{5.16}$$

Proof. See Denneberg [96] (Proposition 5.1–(i), (ii), and (iv) on pp. 64-65) or Marinacci and Montrucchio [204] (Proposition 4.11–(i), (ii), and (iii) on p. 64). □

Proposition 5.13. *Let ν be a submodular capacity on $(S, \mathcal{G})$. If $\phi_1, \phi_2 \in B(\mathcal{G})$, then the following holds:*

$$\oint (\phi_1 + \phi_2)\ d\nu \leq \oint \phi_1\ d\nu + \oint \phi_2\ d\nu \tag{5.17}$$

Proof. See Denneberg [96], Proposition 6.3 on p. 75, or Marinacci and Montrucchio [204], Theorem 4.6 on p. 73. □

5.2.2 Uncertainty, Preferences, and Beliefs

Representation of preferences

As in Chapter 2, S is the set of states of the world and $\mathcal{G}$ is a σ-algebra of events on S. $B^+(\mathcal{G})$ denotes the collection of bounded, $\mathbb{R}^+$-valued, and $\mathcal{G}$-measurable functions of S. The *innovation* will be quantified in monetary terms and will be taken to be a henceforth fixed $X \in B^+(\mathcal{G})$ with a closed range $[0, M] := X(S)$, where $M := \|X\|_s < +\infty$, and Σ denotes the σ-algebra $\sigma\{X\}$ of subsets of S generated by X.

The DM has preference $\succcurlyeq_{DM}$ over $B^+(\Sigma)$, and the CI has preference $\succcurlyeq_{CI}$ over $B^+(\Sigma)$. We will assume that the DM is an Expected-Utility maximizer, whereas the CI's preferences $\succcurlyeq_{CI}$ over $B^+(\Sigma)$ admit a Choquet Expected-Utility (CEU) representation (as in Schmeidler [275] or Gilboa [148])[8]. Specifically, we will assume the following about the representation of preferences:

Assumption 5.14. *There is a bounded, nondecreasing, and continuous utility function $u : \mathbb{R} \to \mathbb{R}$, unique up to a positive affine transformation, and a unique countably additive subjective probability measure μ on the measurable space (S, Σ), such that for each $Y_1, Y_2 \in B^+(\Sigma)$,*

$$Y_1 \succcurlyeq_{DM} Y_2 \iff \int u(Y_1)\ d\mu \geq \int u(Y_2)\ d\mu \tag{5.18}$$

Moreover, there is a nondecreasing and continuous utility function $v : \mathbb{R} \to \mathbb{R}$, which is bounded on bounded subsets of $\mathbb{R}$ and unique up to a positive affine transformation,

[8] See also Appendix A.3.2.

and a unique subjective capacity ν on the measurable space (S, Σ), such that for each $Y_1, Y_2 \in B^+ (\Sigma)$,

$$Y_1 \succcurlyeq_{CI} Y_2 \iff \oint v(Y_1) \, d\nu \geq \oint v(Y_2) \, d\nu \tag{5.19}$$

Additionally, as usual in contracting problems and principal-agent problems, we suppose that the DM is risk-averse, having a concave utility function, and that the CI has a linear utility index v. In the Expected-Utility framework this is equivalent to assuming risk-neutrality of the CI. Since the utility index v is given up to a positive affine transformation, we might then assume the following:

Assumption 5.15. *The CI's utility function v is the identity function, that is, $v(t) = t$ for each $t \in \mathbb{R}$.*

As in Chapter 2, we also make the assumption that the innovation X (with closed range $[0, M]$) has a nonatomic law induced by the probability measure μ, and that the CI and the DM are both aware of the fact that μ represents the DM's beliefs and ν represents the CI's beliefs. We also assume that the capacity ν that represents that CI's beliefs is continuous. Specifically:

Assumption 5.16. *We assume that:*

1. *$\mu \circ X^{-1}$ is nonatomic;*

2. *ν is continuous;*

3. *μ is known by the CI; and,*

4. *ν is known by the DM.*

The assumption of continuity of ν is a technical assumption that is needed for the proof of the existence of a monotone solution to the DM's demand problem, which will be stated below. Moreover, as in the previous chapters, the assumption of nonatomicity of $\mu \circ X^{-1}$ is simply a technical requirement that is needed for defining the *equimeasurable monotone rearrangement* of some $Y \in B^+ (\Sigma)$ with respect to X, as in section 2.3.

Uncertainty and Wealth

The contract between the DM and CI is a pair $(H, Y) \subset \mathbb{R}^+ \setminus \{0\} \times B^+(\Sigma)$, whereby upon entering into the contract with the CI, the DM receives a fixed fee $H > 0$ from the CI, in return of which she accepts to transfer to the CI the amount $X(s)$ of the innovation X in the state of the world $s \in S$, and to receive the amount $Y(s)$ from the CI.

The DM has initial wealth W_0, and after entering into the contract (H, Y) with the CI, her total wealth is the Σ-measurable, $\mathbb{R}$-valued and bounded function on S defined by

$$W^{DM}(H, Y)(s) := W_0 + H - X(s) + Y(s), \ \forall s \in S \tag{5.20}$$

The CI has initial wealth W_0^{CI}, and after entering into the contract (H, Y) with the DM, his total wealth is the Σ-measurable, $\mathbb{R}$-valued and bounded function on S defined by

$$W^{CI}(H, Y)(s) := W_0^{CI} - H + X(s) - Y(s), \ \forall s \in S \tag{5.21}$$

Letting $R := v\left(W_0^{CI}\right) = W_0^{CI}$ be the CI's *reservation utility*, the CI's *individual rationality constraint*, or *participation constraint*, for entering into a contract (H, Y) with the CI is then given by $\oint v\Big(W_0^{CI} - H + X - Y\Big)\, d\nu \geq v\left(W_0^{CI}\right)$. However, since we have assumed v to be the identity function, the CI's *participation constraint* becomes $\oint \Big(W_0^{CI} - H + X - Y\Big)\, d\nu \geq W_0^{CI}$. Now, since $W_0^{CI} - H \in \mathbb{R}$ and $X - Y \in B(\Sigma)$, it follows from Proposition 5.11 then yields that the CI's *participation constraint* is given by:

$$\oint \big(X - Y\big)\, d\nu \geq H \tag{5.22}$$

Vigilant Beliefs and Probabilistic Consistency

As in Chapter 2, we will adopt the following definition of *vigilance*.

Definition 5.17. *The capacity ν is said to be (μ, X)-vigilant if for any $Y_1, Y_2 \in B^+(\Sigma)$ such that* *(i)* Y_1 *and* Y_2 *have the same distribution under* μ*, and,*

(ii) Y_2 *is a nondecreasing function of* X*, i.e.* Y_2 *and* X *are comonotonic,*

the following holds:

$$W^{CI}(H, Y_2) \succcurlyeq_{CI} W^{CI}(H, Y_1) \tag{5.23}$$

Remark 5.18. *Equivalently, we can define vigilance in this setting as follows: the capacity* ν *is* (μ, X)*-vigilant if and only if for any* $Y_1, Y_2 \in B^+(\Sigma)$ *such that*

(i) Y_1 *and* Y_2 *have the same distribution under* μ*, and,*

(ii) Y_2 *is a nondecreasing function of* X*, i.e.* Y_2 *and* X *are comonotonic,*

the following holds:

$$\oint (X - Y_2)\ d\nu \geq \oint (X - Y_1)\ d\nu \tag{5.24}$$

5.3 The DM's Demand for Contingent Claims

5.3.1 The DM's problem

The problem of designing the optimal contract can be seen as that of finding the claim that will maximize the expected utility of the DM's wealth, under her subjective probability measure, subject to the CI's participation constraint and to some constraints on the claim. Specifically, the DM's problem is the following:

Problem 5.19.

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int u(W_0 + H - X + Y)\ d\mu \right\} :$$
$$\begin{cases} 0 \leq Y \leq X \\ \oint (X - Y)\ d\nu \geq H \end{cases}$$

Remark 5.20. *Assuming Problem 5.19 has a nonempty feasibility set, the supremum in Problem 5.19 is finite since the utility function u is bounded.*

5.3.2 Existence of a Solution and Pareto-Improving Claims

Here we will give a sufficient condition for Problem 5.19 to admit a solution which is comonotonic with X.

Definition 5.21. *Let $\mathcal{F}_{SB}$ be given by*

$$\mathcal{F}_{SB} := \left\{ Y \in B(\Sigma) : 0 \le Y \le X \text{ and } \oint (X - Y) \, d\nu \ge H \right\}$$

That is, $\mathcal{F}_{SB}$ is the feasibility set for Problem 5.19. In the following, we will assume that this feasibility set is nonempty:

Assumption 5.22. $\mathcal{F}_{SB} \neq \varnothing$.

Let $\mathcal{F}_{SB}^{\uparrow}$ denote the collection of all feasible $Y \in B^{+}(\Sigma)$ for Problem 5.19 which are also comonotonic with X, i.e. of the form $Y = I \circ X$ where $I : [0, M] \to [0, M]$ is nondecreasing:

Definition 5.23. *Let* $\mathcal{F}_{SB}^{\uparrow} := \left\{ Y = I \circ X \in \mathcal{F}_{SB} : I \text{ is nondecreasing} \right\}$.

Lemma 5.24. *If ν is (μ, X)-vigilant, then $\mathcal{F}_{SB}^{\uparrow} \neq \varnothing$.*

Proof. Similar to the prof of Lemma 2.26. □

Definition 5.25. *If $Y_1, Y_2 \in \mathcal{F}_{SB}$, we will say that Y_2 is a Pareto improvement of Y_1 (or is Pareto-improving) when the following hold:*

1. $\int u\Big(W^{DM}(H, Y_2)\Big)\ d\mu \geq \int u\Big(W^{DM}(H, Y_1)\Big)\ d\mu$*; and,*

2. $\oint W^{CI}(H, Y_2)\ d\nu \geq \oint W^{CI}(H, Y_1)\ d\nu$.

Lemma 5.26. *Suppose that ν is (μ, X)-vigilant and that $\mathcal{U}(X, Y) := u\Big(W_0 + H - X + Y\Big)$ is supermodular (e.g. u is concave). If $Y \in \mathcal{F}_{SB}$, then there is some $Y^* \in \mathcal{F}_{SB}^{\uparrow}$ which is Pareto-improving.*

Proof. Similar to the proof of Lemma 2.28. □

Proposition 5.27. *If ν is (μ, X)-vigilant and $\mathcal{U}(X, Y) := u\Big(W_0 + H - X + Y\Big)$ is supermodular (e.g. u is concave), then Problem 5.19 admits a solution which is comonotonic with X.*

Proof. The main idea of this proof is to extend the methods used for the proof of Proposition 2.29 and to adapt them to the present context.

By Lemma 5.26, we can choose a maximizing sequence $\{Y_n\}_n$ in $\mathcal{F}_{SB}^{\uparrow}$ for Problem 5.19. That is,

$$\lim_{n \to +\infty} \int u\big(W_0 + H - X + Y_n\big)\ d\mu = N$$

where $N < +\infty$ is the supremum value of Problem 5.19. Since $0 \leq Y_n \leq X \leq M := \|X\|_s$, the sequence $\{Y_n\}_n$ is uniformly bounded. Moreover, for each $n \geq 1$ we have $Y_n = I_n \circ X$, with $I_n : [0, M] \to [0, M]$. Consequently, the sequence $\{I_n\}_n$ is a uniformly bounded sequence of nondecreasing Borel-measurable functions. Thus, by Lemma 2.50, there is a nondecreasing function $I^* : [0, M] \to [0, M]$ and a subsequence $\{I_m\}_m$ of $\{I_n\}_n$ such that $\{I_m\}_m$ converges pointwise on $[0, M]$ to I^*. Hence, I^* is also Borel-measurable, and so $Y^* := I^* \circ X \in B^+(\Sigma)$ is such that $0 \leq Y^* \leq X$. Moreover, the sequence $\{Y_m\}_m$, defined by $Y_m = I_m \circ X$, converges pointwise to Y^*. Thus, the sequence $\{X - Y_m\}_m$ is nonnegative, uniformly bounded, and converges pointwise to $X - Y^*$. Therefore, since ν is continuous by Assumption 5.16, it follows from Corollary 5.70 on p. 147 that

$$H \leq \lim_{m \to +\infty} \oint \big(X - Y_m\big)\, d\nu = \oint \big(X - Y^*\big)\, d\nu$$

and so $Y^* \in \mathcal{F}_{SB}^{\uparrow}$.

Now, by continuity and boundedness of the function u (Assumption 5.14), and by Lebesgue's Dominated Convergence Theorem, we have

$$\begin{aligned}\int u\big(W_0 + H - X + Y^*\big)\, d\mu &= \lim_{m \to +\infty} \int u\big(W_0 + H - X + Y_m\big)\, d\mu \\ &= \lim_{n \to +\infty} \int u\big(W_0 + H - X + Y_n\big)\, d\mu = N\end{aligned}$$

Hence Y^* solves Problem 5.19. □

5.4 An Application: The Case of a Concave Distortion of a Measure

Here we consider a special case of the previous general model. Namely, we suppose that $\nu = T \circ P$, for some probability measure P on (S, Σ) and some function $T : [0,1] \to [0,1]$, increasing, concave and continuous, with $T(0) = 0$ and $T(1) = 1$. Then $T \circ P$ is a continuous submodular capacity on (S, Σ). Based on Gilboa [147], we may assume that both the *distortion function* T and the probability measure P are *subjective*, i.e. they are determined entirely from the CI's preferences, since ν is[9]. We will also assume that $P \circ X^{-1}$ is a nonatomic Borel probability measure. Specifically:

Assumption 5.28. *We assume that $\nu = T \circ P$, where:*

1. *P is a probability measure on (S, Σ) such that $P \circ X^{-1}$ is nonatomic;*

[9]Theorem 3.1 of Gilboa [147] also yields that both T and P are unique.

2. $T : [0,1] \to [0,1]$ *is increasing, concave and continuously differentiable; and,*

3. $T(0) = 0$ *and* $T(1) = 1$.

Remark 5.29. *Since the seminal work of Schmeidler [275], it is customary in the CEU-model to equate the submodularity of the capacity with an ambiguity seeking attitude. Specifically, if* ν *is a submodular capacity on the measurable space* (S, Σ)*, then there exists a (weak*-)compact and convex set* $\mathcal{C}_0$ *of probability measures on* (S, Σ) *such that for any* $\phi \in B^+(\Sigma)$,

$$\oint \phi \, d\nu = \max_{\lambda \in \mathcal{C}_0} \int \phi \, d\lambda \tag{5.25}$$

It is not surprising in the present context to assume that the CI is ambiguity seeking. Indeed, (5.25) *can be interpreted as indicating that the CI is optimistic about the possible realizations of the innovation* X*. Arguably, this is in practice a necessary condition for the existence of such markets for innovation in the first place, and – more generally – for venture capitalism.*

We will also make the assumption that the DM is risk-averse, i.e. that u is a concave utility function.

Assumption 5.30. *The DM's utility function* u *is concave*[10] *and nondecreasing.*

For each $Z \in B^+(\Sigma)$, let $F_Z(t) := P\big(\{s \in S : Z(s) \le t\}\big)$ denote the distribution function of Z with respect to the probability measure P, and let $F_X(t) := P\big(\{s \in S : X(s) \le t\}\big)$ denote the distribution function of X with respect to the probability measure P. Let $F_Z^{-1}(t)$ be the left-continuous inverse of the distribution function F_Z (that is, the *quantile function* of Z), defined by

$$F_Z^{-1}(t) := \inf\Big\{z \in \mathbb{R}^+ : F_Z(z) \ge t\Big\}, \ \forall t \in [0,1] \tag{5.26}$$

Definition 5.31. *Denote by* $\mathcal{AQ}$ *the collection of all quantile functions* f *of the form* F^{-1}*, where* F *is the distribution function of some* $Z \in B^+(\Sigma)$ *such that* $0 \le Z \le X$.

[10] This then implies that the function $\mathcal{U}(x, y) = u(W_0 + H - x + y)$ is supermodular.

By Lebesgue's decomposition theorem, there exists a unique pair (μ_{ac}, μ_s) of (nonnegative) finite measures on (S, Σ) such that $\mu = \mu_{ac} + \mu_s$, $\mu_{ac} << P$, and $\mu_s \perp P$. That is, for all $B \in \Sigma$ with $P(B) = 0$, we have $\mu_{ac}(B) = 0$, and there is some $A \in \Sigma$ such that $P(S \setminus A) = \mu_s(A) = 0$. It then also follows that $\mu_{ac}(S \setminus A) = 0$ and $P(A) = 1$. Note also that for all $Z \in B(\Sigma)$, $\int Z \, d\mu = \int_A Z \, d\mu_{ac} + \int_{S \setminus A} Z \, d\mu_s$. Furthermore, by the Radon-Nikodým theorem, there exists a P-a.s. unique Σ-measurable and P-integrable function $h : S \to [0, +\infty)$ such that $\mu_{ac}(C) = \int_C h \, dP$, for all $C \in \Sigma$. Consequently, for all $Z \in B(\Sigma)$, $\int Z \, d\mu = \int_A Zh \, dP + \int_{S \setminus A} Z \, d\mu_s$. Moreover, since $\mu_{ac}(S \setminus A) = 0$, it follows that $\int_{S \setminus A} Z \, d\mu_s = \int_{S \setminus A} Z \, d\mu$. Thus, for all $Z \in B(\Sigma)$, $\int Z \, d\mu = \int_A Zh \, dP + \int_{S \setminus A} Z \, d\mu$.

Moreover, since $h : S \to [0, +\infty)$ is Σ-measurable and P-integrable, there exists a Borel-measurable and $P \circ X^{-1}$-integrable map $\phi : X(S) \to [0, +\infty)$ such that $h = d\mu_{ac}/dP = \phi \circ X$. We will also make the following assumption:

Assumption 5.32. *The Σ-measurable function $h = \phi \circ X = d\mu_{ac}/dP$ is anti-comonotonic with X, i.e. ϕ is nonincreasing.*

Since $P \circ X^{-1}$ is nonatomic (by Assumption 5.28), it follows that $F_X(X)$ has a uniform distribution over $(0, 1)$ (see Föllmer and Schied [129], Lemma A.1 on p. 409), that is, $P(\{s \in S : F_X(X)(s) \leq t\}) = t$ for each $t \in (0, 1)$. Letting $U := F_X(X)$, it follows that U is a random variable on the probability space (S, Σ, P) with a uniform distribution on $(0, 1)$. Consider the following quantile problem:

Problem 5.33. *For a given $\beta \geq H$,*

$$\sup_f \left\{ V(f) := \int u\big(W_0 + H - f(U)\big) \, \phi\left(F_X^{-1}(U)\right) \, dP \right\} :$$
$$\begin{cases} f \in \mathcal{AQ} \\ \int T'(1-U) \, f(U) \, dP = \beta \end{cases}$$

The following theorem characterizes the solution of Problem 5.19 in terms of the solution of the relatively easier quantile problem given in Problem 5.33, if the previous assumptions hold.

Theorem 5.34. *Under the previous assumptions, there exists a parameter $\beta^* \geq H$ such that if f^* is optimal for Problem 5.33 with parameter β^*, then the function*

$$Y^* := \big(X - f^*(U)\big)\mathbf{1}_A + X\mathbf{1}_{S\setminus A}$$

is optimal for Problem 5.19.

5.4.1 Proof of Theorem 5.34

"Splitting"

In the following, the Σ-measurable set A on which P is concentrated (and $\mu_s(A) = 0$) is assumed to be fixed all throughout.

Lemma 5.35. *Let Y^* be an optimal solution for Problem 5.19, and suppose that ν is (μ, X)-vigilant. Let $\widetilde{Y}^*_\mu$ be the nondecreasing μ-rearrangement of Y^* with respect to X. Then $\widetilde{Y}^*_\mu$ is optimal for Problem 5.19 and comonotonic with X.*

Proof. This is an immediate consequence of Assumption 5.30 and of Lemma 5.26 and its proof. □

Consider now the following two problems:

Problem 5.36. *For a given $\beta \geq H$,*

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int_A u\big(W_0 + H - X + Y\big)\, d\mu \right\} :$$
$$\begin{cases} 0 \leq Y \leq X \\ \oint \big(X - Y\big)\, dT \circ P = \beta \end{cases}$$

Problem 5.37.

$$\sup_{Y \in B^{+}(\Sigma)} \left\{ \int_{S\setminus A} u\big(W_0 + H - X + Y\big)\ d\mu \right\} :$$
$$\begin{cases} 0 \leq Y\mathbf{1}_{S\setminus A} \leq X\mathbf{1}_{S\setminus A} \\ \oint_{S\setminus A} \big(X - Y\big)\ dT \circ P = 0 \end{cases}$$

Remark 5.38. *By Remark 5.20, the supremum value of each of the above two problems is finite when their feasibility sets are nonempty. Now, the function X is feasible for Problem 5.37, and so Problem 5.37 has a nonempty feasibility set.*

Definition 5.39. *For a given $\beta \geq H$, let $\Theta_{A,\beta}$ be the feasibility set of Problem 5.36 with parameter β. That is,*

$$\Theta_{A,\beta} := \left\{ Y \in B^{+}(\Sigma) : 0 \leq Y \leq X,\ \oint (X - Y)\ \ d\nu = \beta \right\}$$

Denote by Γ the collection of all β for which the feasibility set $\Theta_{A,\beta}$ is nonempty:

Definition 5.40. *Let* $\Gamma := \left\{ \beta \geq H : \Theta_{A,\beta} \neq \varnothing \right\}$

Lemma 5.41. $\Gamma \neq \varnothing$.

Proof. By Assumption 5.22, there is some $Y \in B^{+}(\Sigma)$ such that $0 \leq Y \leq X$, and $\oint (X - Y)\ \ d\nu \geq H$. Let $\beta_Y := \oint (X - Y)\ \ d\nu$. Then, by definition of β_Y, and since $0 \leq Y \leq X$, we have $Y \in \Theta_{A,\beta_Y}$, and so $\Theta_{A,\beta_Y} \neq \varnothing$. Consequently, $\beta_Y \in \Gamma$, and so $\Gamma \neq \varnothing$. □

Lemma 5.42. *X is optimal for Problem 5.37*

Proof. The feasibility of X for Problem 5.37 is clear. To show optimality, let Y be any feasible function for Problem 5.37. Then for each $s \in S \setminus A$, $Y(s) \leq X(s)$. Therefore, since u is increasing, we have $u\big(W_0 + H - X(s) + Y(s)\big) \leq u\big(W_0 + H - X(s) + X(s)\big) = u\big(W_0 + H\big)$, for each $s \in S \setminus A$. Thus,

$$\int_{S \setminus A} u\big(W_0 + H - X + Y\big)\, d\mu \leq \int_{S \setminus A} u\big(W_0 + H - X + X\big)\, d\mu = u\big(W_0 + H\big)\mu(S \setminus A)$$

□

Remark 5.43. *Since $P(S \setminus A) = 0$ and $T(0) = 0$, it follows that $T \circ P(S \setminus A) = 0$, and so $\oint \mathbf{1}_{S \setminus A}\, dT \circ P = T \circ P(S \setminus A) = 0$, by Proposition 5.12 (1). Therefore, for any $Z \in B^+(\Sigma)$, it follows form the monotonicity and positive homogeneity of the Choquet integral (Proposition 5.12 (2) and (3)) that*

$$\begin{aligned} 0 &\leq \oint_{S \setminus A} Z\, dT \circ P = \oint Z \mathbf{1}_{S \setminus A}\, dT \circ P \\ &\leq \oint \|Z\|_s \mathbf{1}_{S \setminus A}\, dT \circ P = \|Z\|_s \oint \mathbf{1}_{S \setminus A}\, dT \circ P = 0 \end{aligned}$$

and so $\oint_{S \setminus A} Z\, dT \circ P = 0$. Consequently, it follows form Proposition 5.13 that for any $Z \in B^+(\Sigma)$,

$$\oint Z\, dT \circ P \leq \oint Z \mathbf{1}_A\, dT \circ P = \oint_A Z\, dT \circ P$$

Now, consider the following problem:

Problem 5.44.

$$\sup_{\beta \in \Gamma} \left\{ F_A^*(\beta) : F_A^*(\beta) \text{ is the supremum value of Problem 5.36, for a fixed } \beta \in \Gamma \right\}$$

Lemma 5.45. *Under Assumption 5.28, if β^* is optimal for Problem 5.44, and if Y_1^* is optimal for Problem 5.36 with parameter β^*, then $Y^* := Y_1^* \mathbf{1}_A + X \mathbf{1}_{S \setminus A}$ is optimal for Problem 5.19.*

Proof. By the feasibility of Y_1^* for Problem 5.36 with parameter β^*, we have $0 \leq Y_1^* \leq X$ and $\oint (X - Y_1^*) \ dT \circ P = \beta^*$. Therefore, $0 \leq Y^* \leq X$, and

$$\begin{aligned} \oint (X - Y^*) \ dT \circ P &= \oint \left[(X - Y_1^*) \mathbf{1}_A + (X - X) \mathbf{1}_{S \setminus A} \right] \ dT \circ P \\ &= \oint_A (X - Y_1^*) \ dT \circ P \geq \oint (X - Y_1^*) \ dT \circ P \\ &= \beta^* \geq H \end{aligned}$$

where the inequality $\oint_A (X - Y_1^*) \ dT \circ P \geq \oint (X - Y_1^*) \ dT \circ P$ follows from the same argument as in Remark 5.43. Hence, Y^* is feasible for Problem 5.19. To show optimality of Y^* for Problem 5.19, let $\overline{Y}$ be any other feasible function for Problem 5.19, and define $\alpha := \oint \left(X - \overline{Y} \right) \ dT \circ P$. Then $\alpha \geq H$, and so $\overline{Y}$ is feasible for Problem 5.36 with parameter α, and α is feasible for Problem 5.44. Hence

$$F_A^* (\alpha) \geq \int_A u\big(W_0 + H - X + \overline{Y}\big) \ d\mu$$

Now, since β^* is optimal for Problem 5.44, it follows that

$$F_A^* (\beta^*) \geq F_A^* (\alpha)$$

Moreover, $\overline{Y}$ is feasible for Problem 5.37 (since $0 \leq \overline{Y} \leq X$ and so $\oint_{S \setminus A} \left(X - \overline{Y} \right) \ dT \circ P = 0$ by Remark 5.43). Thus,

$$\begin{aligned} F_A^* (\beta^*) + u\big(W_0 + H\big) \mu (S \setminus A) &\geq F_A^* (\alpha) + u\big(W_0 + H\big) \mu (S \setminus A) \\ &\geq \int_A u\big(W_0 + H - X + \overline{Y}\big) \ d\mu \\ &\qquad + u\big(W_0 + H\big) \mu (S \setminus A) \\ &\geq \int_A u\big(W_0 + H - X + \overline{Y}\big) \ d\mu \\ &\qquad + \int_{S \setminus A} u\big(W_0 + H - X + \overline{Y}\big) \ d\mu \\ &= \int u\big(W_0 + H - X + \overline{Y}\big) \ d\mu \end{aligned}$$

However,

$$F_A^*(\beta^*) = \int_A u\big(W_0 + H - X + Y_1^*\big)\ d\mu$$

Therefore,

$$\begin{aligned}\int u\big(W_0 + H - X + Y^*\big)\ d\mu &= F_A^*(\beta^*) + u\big(W_0 + H\big)\mu(S \setminus A) \\ &\geq \int u\big(W_0 + H - X + \overline{Y}\big)\ d\mu\end{aligned}$$

Hence, Y^* is optimal for Problem 5.19. □

Remark 5.46. *By Lemma 5.45, we can restrict ourselves to solving Problem 5.36 with a parameter* $\beta \in \Gamma$*.*

Solving Problems 5.36

Recall that for all $Z \in B(\Sigma)$, $\int Z\ d\mu = \int_A Zh\ dP + \int_{S \setminus A} Z\ d\mu$, where $h = d\mu_{ac}/dP$ is the Radon-Nikodým derivative of μ_{ac} with respect to P. Moreover, by definition of the set $A \in \Sigma$, we have $P(S \setminus A) = \mu_s(A) = 0$. Therefore, $\int_A Zh\ dP = \int Zh\ dP$, for each $Z \in B(\Sigma)$. Hence, we can rewrite Problem 5.36 (restricting ourselves to parameters $\beta \in \Gamma$ and recalling that $h = \phi \circ X$) as the following problem:

Problem 5.47. *For a given* $\beta \in \Gamma$*,*

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int u\big(W_0 + H - X + Y\big)\ \phi(X)\ dP \right\} :$$

$$\begin{cases} 0 \leq Y \leq X \\ \oint (X - Y)\ dT \circ P = \beta \end{cases}$$

Now, consider the following problem:

Problem 5.48. *For a given $\beta \in \Gamma$,*

$$\sup_{Z \in B^+(\Sigma)} \left\{ \int u\big(W_0 + H - Z\big)\ \phi(X)\ dP \right\} :$$
$$\begin{cases} 0 \leq Z \leq X \\ \oint Z\ dT \circ P = \beta = \int_0^{+\infty} T\Big(P\big(\{s \in S : Z(s) \geq t\}\big)\Big)\ dt \end{cases}$$

Lemma 5.49. *If Z^* is optimal for Problem 5.48 with parameter β, then $Y^* := X - Z^*$ is optimal for Problem 5.47 with parameter β.*

Proof. Let $\beta \in \Gamma$ be given, and suppose that Z^* is optimal for Problem 5.48 with parameter β. Define $Y^* := X - Z^*$. Then $Y^* \in B(\Sigma)$. Moreover, since $0 \leq Z^* \leq X$, it follows that $0 \leq Y^* \leq X$. Now,

$$\oint (X - Y^*)\ dT \circ P = \oint \Big(X - (X - Z^*)\Big)\ dT \circ P = \oint Z^*\ dT \circ P = \beta$$

and so Y^* is feasible for Problem 5.47 with parameter β. To show optimality of Y^* for Problem 5.47 with parameter β, suppose *per contra* that $\overline{Y} \neq Y^*$ is feasible for Problem 5.47 with parameter β and

$$\int u\big(W_0 + H - X + \overline{Y}\big)\ h\ dP > \int u\big(W_0 + H - X + Y^*\big)\ h\ dP$$

that is, with $\overline{Z} := X - \overline{Y}$, we have

$$\int u\big(W_0 + H - \overline{Z}\big)\ h\ dP > \int u\big(W_0 + H - Z^*\big)\ h\ dP$$

Now, since $0 \leq \overline{Y} \leq X$ and $\oint \big(X - \overline{Y}\big)\ dT \circ P = \beta$, we have that $\overline{Z}$ is feasible for Problem 5.48 with parameter β, hence contradicting the optimality of Z^* for Problem 5.48 with parameter β. Thus, $Y^* := X - Z^*$ is optimal for Problem 5.47 with parameter β. $\square$

The next result shows that for any feasible claim, there is a another feasible claim which is comonotonic with X and Pareto-improving.

Lemma 5.50. *Fix a parameter $\beta \in \Gamma$. If Z is feasible for Problem 5.48 with parameter β, then $\widetilde{Z}$ is feasible for Probem 5.48 with parameter β, comonotonic with X, and Pareto-improving, where $\widetilde{Z}$ is the nondecreasing P-rearrangement of Z with respect to X.*

Proof. Let Z be feasible for Problem 5.48 with parameter β, and note that by Assumptions 5.30 and 5.32 and by Lemma 2.13, the map $L(X, Z) := u\big(W_0 + H - Z\big)\ \phi(X)$ is supermodular (see Example 2.14 (6)). Let $\widetilde{Z}$ denote the nondecreasing P-rearrangement of Z with respect to X. Then by Lemma 2.16 (1) and by equimeasurability of Z and $\widetilde{Z}$, the function $\widetilde{Z}$ is feasible for Problem 5.48 with parameter β. Also, by Lemma 2.15 (1) and by supermodularity of $L(X, Z)$, it follows that $\widetilde{Z}$ is Pareto-improving. □

Quantile reformulation

Fix a parameter $\beta \in \Gamma$, let $Z \in B^+(\Sigma)$ be feasible for Problem 5.48 with parameter β, and let $\widetilde{Z}$ denote the nondecreasing P-rearrangement of Z with respect to X. Since $Z \in B^+(\Sigma)$, it can be written as $\psi \circ X$ for some nonnegative Borel-measurable and bounded map ψ on $X(S) = [0, M]$. Moreover, since $0 \leq Z \leq X$, ψ is a mapping of $[0, M]$ into $[0, M]$. Let $\phi := P \circ X^{-1}$ be the image measure of P under X. By Assumption 5.28, ϕ is nonatomic. We can then define the mapping $\widetilde{\psi} : [0, M] \to [0, M]$ as in Section 2.3 (see equation (2.12) on p. 20) to be the nondecreasing ϕ-rearrangement of ψ, that is,

$$\widetilde{\psi}(t) := \inf \Big\{ z \in \mathbb{R}^+ : \phi\big(\{x \in [0, M] : \psi(x) \leq z\}\big) \geq \phi\big(\,[0, t]\,\big) \Big\} \tag{5.27}$$

Then, as in Section 2.3, $\widetilde{Z} = \widetilde{\psi} \circ X$. Therefore, for each $s_0 \in S$,

$$\begin{aligned} \widetilde{Z}(s_0) &= \widetilde{\psi}(X(s_0)) \\ &= \inf \Big\{ z \in \mathbb{R}^+ : \phi\big(\{x \in [0, M] : \psi(x) \leq z\}\big) \geq \phi\big(\,[0, X(s_0)]\,\big) \Big\} \end{aligned} \tag{5.28}$$

However, for each $s_0 \in S$,

$$\phi\big([0,X(s_0)]\big) = P\circ X^{-1}\big([0,X(s_0)]\big) = F_X(X(s_0)) := F_X(X)(s_0) \tag{5.29}$$

Moreover,

$$\begin{aligned}\phi\big(\{x\in[0,M]:\psi(x)\le z\}\big) &= P\circ X^{-1}\big(\{x\in[0,M]:\psi(x)\le z\}\big)\\ &= P\big(\{s\in S:\psi(X(s))\le z\}\big) = F_Z(z)\end{aligned} \tag{5.30}$$

Consequently, for each $s_0 \in S$,

$$\begin{aligned}\widetilde{Z}(s_0) &= \inf\Big\{z\in\mathbb{R}^+ : F_Z(z)\ge F_X(X)(s_0)\Big\}\\ &= F_Z^{-1}(F_X(X(s_0))) := F_Z^{-1}(F_X(X))(s_0)\end{aligned} \tag{5.31}$$

That is,

$$\widetilde{Z} = F_Z^{-1}(F_X(X)) \tag{5.32}$$

where F_z^{-1} is the left-continuous inverse of F_Z, as defined in (5.26).

Hence, by Lemma 5.50 and equation (5.32), we can restrict ourselves to finding a solution to Problem 5.48 of the form $F^{-1}(F_X(X))$, where F is the distribution function of a function $Z \in B^+(\Sigma)$ such that $0 \le Z \le X$ and $\oint Z\, dT\circ P = \beta$. Moreover, since X is a nondecreasing function of X and P-equimeasurable with X, it follows from the P-a.s. uniqueness of the equimeasurable nondecreasing P-rearrangement (see Section 2.3) that $X = F_X^{-1}(F_X(X))$, P-a.s. (see also Föllmer and Schied [129], Lemma A.1 on p. 409). Thus, for any $Z \in B^+(\Sigma)$,

$$\begin{aligned}\int u\big(W_0 + H - F_Z^{-1}(F_X(X))\big)\ \phi\big(F_X^{-1}(F_X(X))\big)\ dP &= \int u\big(W_0+H-\widetilde{Z}\big)\ \phi(X)\ dP\\ &\ge \int u\big(W_0+H-Z\big)\ \phi(X)\ dP\end{aligned}$$

where the inequality follows from the proof of Lemma 5.50.

Moreover, since $P \circ X^{-1}$ is nonatomic (by Assumption 5.28), it follows that $F_X(X)$ has a uniform distribution over $(0,1)$ (see Föllmer and Schied [129], Lemma A.1 on p. 409), that is, $P\big(\{s \in S : F_X(X)(s) \leq t\}\big) = t$ for each $t \in (0,1)$. Finally, letting $U := F_X(X)$,

$$\begin{aligned}
\oint F^{-1}(U)\ dT \circ P &= \int_0^{+\infty} T\Big[P\big(\{s \in S : F^{-1}(U)(s) \geq t\}\big)\Big]\ dt \\
&= \int_0^{+\infty} T\Big[P\big(\{s \in S : F^{-1}(U)(s) > t\}\big)\Big]\ dt \\
&= \int_0^{+\infty} T\Big[1 - F(t)\Big]\ dt \\
&= \int_0^1 T'(1-t)\, F^{-1}(t)\ dt = \int T'(1-U)\, F^{-1}(U)\ dP
\end{aligned}$$

where the third and last equalities above follow from the fact that U has a uniform distribution over $(0,1)$, and where the second-to-last equality follows from a standard argument (see, e.g. Denneberg [96], Proposition 1.4 on p. 8 and the discussion on pp. 61-62. See also Jin and Zhou [175] p. 418, He and Zhou [164] p. 210 and p. 213, or the remark of Carlier and Dana [71] on p. 207).

Now, recall from Definition 5.31 that $\mathcal{AQ}$ is the collection of all *admissible quantile functions*, that is the collection of all functions f of the form F^{-1}, where F is the distribution function of a function $Z \in B^+(\Sigma)$ such that $0 \leq Z \leq X$, and consider the following problem:

Problem 5.51. *For a given $\beta \in \Gamma$,*

$$\sup_f \left\{ V(f) := \int u\big(W_0 + H - f(U)\big)\ \phi\left(F_X^{-1}(U)\right)\ dP \right\} :$$
$$\begin{cases} f \in \mathcal{AQ} \\ \int T'(1-U)\, f(U)\ dP = \beta \end{cases}$$

Lemma 5.52. *If f^* is optimal for Problem 5.51 with parameter $\beta \in \Gamma$, then the function*

$f^*(U)$ is optimal for Problem 5.48 with parameter β, where $U := F_X(X)$. Moreover, $X - f^*(U)$ is optimal for Problem 5.47 with parameter β.

Proof. Fix $\beta \in \Gamma$, suppose that $f^* \in \mathcal{AQ}$ is optimal for Problem 5.51 with parameter β, and let $Z^* \in B^+(\Sigma)$ be the corresponding function. That is, f^* is the quantile function of Z^* and $0 \leq Z^* \leq X$. Let $\widetilde{Z}^* := f^*(U)$. Then $\widetilde{Z}^*$ is the equimeasurable nondecreasing P-rearrangement of Z^* with respect to X, and so $0 \leq \widetilde{Z}^* \leq X$ by Lemma 2.16 (1). Moreover,

$$\begin{aligned}
\beta &= \int T'(1-U)\, f^*(U)\ dP = \oint f^*(U)\ dT \circ P \\
&= \oint \widetilde{Z}^*\ dT \circ P = \int_0^{+\infty} T\Big[P\big(\{s \in S : \widetilde{Z}^*(s) \geq t\}\big)\Big]\ dt \\
&= \int_0^{+\infty} T\Big[P\big(\{s \in S : Z^*(s) \geq t\}\big)\Big]\ dt = \oint Z^*\ dT \circ P
\end{aligned}$$

where the second-to-last equality follows from the P-equimeasurability of Z^* and $\widetilde{Z}^*$. Therefore, $\widetilde{Z}^* = f^*(U)$ is feasible for Problem 5.48 with parameter β. To show optimality, let Z be any feasible function for Problem 5.48 with parameter β, and let F be the distribution function for Z. Then, by Lemma 5.50, the function $\widetilde{Z} := F^{-1}(U)$ is feasible for Probem 5.48 with parameter β, comonotonic with X, and Pareto-improving. Moreover, $\widetilde{Z}$ has also F as a distribution function. To show optimality of $\widetilde{Z}^* = f^*(U)$ for Problem 5.48 with parameter β, it remains to show that

$$\int u\big(W_0 + H - \widetilde{Z}^*\big)\ \phi(X)\ dP \geq \int u\big(W_0 + H - \widetilde{Z}\big)\ \phi(X)\ dP$$

Now, let $f := F^{-1}$, so that $\widetilde{Z} = f(U)$. Since $\widetilde{Z}$ is feasible for Probem 5.48 with parameter β, we have

$$\begin{aligned}
\beta &= \oint \widetilde{Z}\ dT \circ P = \oint F^{-1}(U)\ dT \circ P \\
&= \int_0^1 T'(1-t)\, F^{-1}(t)\ dt = \int T'(1-U)\, f(U)\ dP
\end{aligned}$$

Hence, f is feasible for Problem 5.51 with parameter β. Since f^* is optimal for Problem 5.51 with parameter β we have

$$\int u\big(W_0 + H - f^*(U)\big)\ \phi\left(F_X^{-1}(U)\right)\ dP \geq \int u\big(W_0 + H - f(U)\big)\ \phi\left(F_X^{-1}(U)\right)\ dP$$

Finally, since $X = F_X^{-1}(U)$, P-a.s., we have

$$\int u\big(W_0 + H - \widetilde{Z}^*\big)\ \phi(X)\ dP \geq \int u\big(W_0 + H - \widetilde{Z}\big)\ \phi(X)\ dP$$

Therefore, $\widetilde{Z}^* = f^*(U)$ is optimal for Problem 5.48 with parameter β. Finally, by Lemma 5.49, $Y^* := X - \widetilde{Z}^* = X - f^*(U)$ is optimal for Problem 5.47 with parameter β. □

By Lemmata 5.45 and 5.52, this completes the proof of Theorem 5.34.

5.5 Contracting under Bilateral Ambiguity

5.5.1 Uncertainty, Preferences, and Beliefs

In this section we consider a situation where both the DM and the CI have preferences over $B^+(\Sigma)$ that admit a representation that reflects some *ambiguity* in their beliefs. We also suppose that the DM's preferences are represented by a *symmetric Choquet integral* (see Appendix 5.7), hence reflecting some *gain-loss* separability, in the spirit of the *Cumulative Prospect Theory* (CPT) of Kahneman and Tversky ([177] and [293]). We refer the reader to Appendix A.3.1, p. 189, for a description of CPT. For more about the separability properties of a CPT-functional, including several equivalent characterizations of the CPT-functional, as well as a discussion and definition of the notion of *Loss Aversion*, see Ghossoub [141]. Bernard and Ghossoub [32] also provide some equivalent formulations of a CPT-functional.

Specifically, we will assume the following about the representation of preferences:

Assumption 5.53. *There is a bounded, nondecreasing, and continuous utility function $u : \mathbb{R} \to \mathbb{R}$, unique up to a positive affine transformation, and a unique subjective capacity η on the measurable space (S, Σ), such that for each $Y_1, Y_2 \in B^+(\Sigma)$,*

$Y_1 \succcurlyeq_{DM} Y_2 \iff$

$$\left[\oint (u \circ Y_1)^+ \ d\eta - \oint (u \circ Y_1)^- \ d\eta\right] \geq \left[\oint (u \circ Y_2)^+ \ d\eta - \oint (u \circ Y_2)^- \ d\eta\right] \tag{5.33}$$

Moreover, there is a nondecreasing and continuous utility function $v : \mathbb{R} \to \mathbb{R}$, which is bounded on bounded subsets of $\mathbb{R}$ and unique up to a positive affine transformation, and a unique subjective capacity ν on the measurable space (S, Σ), such that for each $Y_1, Y_2 \in B^+(\Sigma)$,

$$Y_1 \succcurlyeq_{CI} Y_2 \iff \oint v(Y_1) \ d\nu \geq \oint v(Y_2) \ d\nu \tag{5.34}$$

As in Appendix 5.8, we define "continuity" of the random variable X with respect to the capacity η as follows:

Definition 5.54. *We say that $\psi := \eta \circ X^{-1}$ is nonatomic if for any Borel set A with $\psi(A) > 0$, there is some Borel set $B \subsetneq A$ such that $0 < \psi(B) < \psi(A)$.*

We will say that X is a "continuous" random variable for η if ψ is nonatomic.

We will make the following assumptions:

Assumption 5.55. *We assume that:*

1. *The CI's utility function v is the identity function, that is, $v(t) = t$ for each $t \in \mathbb{R}$;*
2. *η is continuous and submodular;*
3. *$\eta \circ X^{-1}$ is nonatomic;*

4. ν is continuous;

5. η is known by the CI; and,

6. ν is known by the DM.

As in the previous setting (section 5.2.2), the contract between the DM and CI is a pair $(H, Y) \subset \mathbb{R}^+ \setminus \{0\} \times B^+(\Sigma)$. Also, for a given contract (H, Y), the wealth of the DM (resp. of the CI) is the element $W^{DM}(H, Y)$ (resp. $W^{CI}(H, Y)$) of $B(\Sigma)$ given in equation (5.20) (resp. equation (5.21)). Moreover, the CI's *participation constraint* is given by:

$$\oint \big(X - Y\big)\, d\nu \geq H \tag{5.35}$$

Definition 5.56. *We say that $Y, Z \in B^+(\Sigma)$ have the same distribution under η if for each $\alpha \geq 0$,*

$$\eta\Big(s \in S : Y(s) \leq \alpha\Big) = \eta\Big(s \in S : Z(s) \leq \alpha\Big)$$

Remark 5.57. *In Definition 5.56, if η were a bona fide (additive) measure then the fact that for each $\alpha \geq 0$,*

$$\eta\Big(s \in S : Y(s) \leq \alpha\Big) = \eta\Big(s \in S : Z(s) \leq \alpha\Big)$$

is enough to imply that for any Borel set B,

$$\eta\Big(\{s \in S : Y(s) \in B\Big) = \eta\Big(\{s \in S : Z(s) \in B\}\Big),$$

as was shown in the proof of Proposition 2.9 (6).

However, in more general situations where η is not a measure, the fact that

$$\eta\Big(s \in S : Y(s) \leq \alpha\Big) = \eta\Big(s \in S : Z(s) \leq \alpha\Big)$$

for each $\alpha \geq 0$ need not imply that

$$\eta\Big(\{s \in S : Y(s) \in B\Big) = \eta\Big(\{s \in S : Z(s) \in B\}\Big),$$

for each Borel set B. Consequently, there are several ways in which one can define the notion of " η-equimeasurability" of two elements of $B^+(\Sigma)$, when η is not additive. For instance:

1. *Denneberg [96] (p. 46) defines the "distribution function" of $Y \in B^+(\Sigma)$ with respect to a capacity η on a measurable space (S, Σ) as the function*

$$G_{\eta,Y}(t) := \eta\Big(\{s \in S : Y(s) > t\}\Big)$$

   One would then say that $Y, Z \in B^+(\Sigma)$ have the same distribution under η if for each $\alpha \geq 0$,

$$G_{\eta,Y}(\alpha) = G_{\eta,Z}(\alpha)$$

2. *One could also say that $Y, Z \in B^+(\Sigma)$ have the same distribution under η if for each Borel set B,*

$$\eta\Big(\{s \in S : Y(s) \in B\Big) = \eta\Big(\{s \in S : Z(s) \in B\}\Big)$$

Here, we will adopt the following definition of *vigilance*:

Definition 5.58. *The capacity ν is said to be (η, X)-vigilant if for any $Y_1, Y_2 \in B^+(\Sigma)$ such that*

(i) Y_1 and Y_2 have the same distribution under η (in the sense of Definition 5.56) and,

(ii) Y_2 is a nondecreasing function of X, i.e. Y_2 and X are comonotonic,

the following holds:

$$\oint \big(X - Y_2\big)\, d\nu \geq \oint \big(X - Y_1\big)\, d\nu \tag{5.36}$$

5.5.2 The DM's Problem

The DM's problem is the following:

Problem 5.59.

$$\sup_{Y \in B^+(\Sigma)} \left\{ \oint \Big(u\big(W_0 + H - X + Y\big)\Big)^+ \, d\eta - \oint \Big(u\big(W_0 + H - X + Y\big)\Big)^- \, d\eta \right\} :$$

$$\begin{cases} 0 \leq Y \leq X \\ \oint (X - Y) \ d\nu \geq H \end{cases}$$

Remark 5.60. *By boundedness of the utility function u (Assumption 5.53) and by Proposition 5.12, the supremum value of Problem 5.59 is finite whenever its feasibility set is nonempty.*

Definition 5.61. *Let $\mathcal{F}_{SB}$ be given by*

$$\mathcal{F}_{SB} := \left\{ Y \in B(\Sigma) : 0 \leq Y \leq X \ \text{and} \ \oint (X - Y) \ d\nu \geq H \right\}$$

That is, $\mathcal{F}_{SB}$ is the feasibility set for Problem 5.59. In the following, we will assume that this feasibility set is nonempty:

Assumption 5.62. $\mathcal{F}_{SB} \neq \varnothing$.

Let $\mathcal{F}_{SB}^{\uparrow}$ denote the collection of all feasible $Y \in B^+(\Sigma)$ for Problem 5.59 which are also comonotonic with X, i.e. of the form $Y = I \circ X$ where $I : [0, M] \to [0, M]$ is nondecreasing:

Definition 5.63. *Let* $\mathcal{F}_{SB}^{\uparrow} := \left\{ Y = I \circ X \in \mathcal{F}_{SB} : I \ \text{is nondecreasing} \right\}$.

Theorem 5.64. *If ν is (η, X)-vigilant, then $\mathcal{F}_{SB}^{\uparrow} \neq \varnothing$ and for each $Y \in \mathcal{F}_{SB}$ there exists some $\widetilde{Y} \in \mathcal{F}_{SB}^{\uparrow}$ such that*

$$W^{CI}\left(H, \widetilde{Y}\right) \succcurlyeq_{CI} W^{CI}(H, Y)$$

Proof. By Assumption 5.62, $\mathcal{F}_{SB} \neq \varnothing$. Choose any $Y = I \circ X \in \mathcal{F}_{SB}$, and let $\widetilde{Y}$ denote a nondecreasing η-rearrangement of Y with respect to X, as defined in Appendix 5.8. Then (i) Y and $\widetilde{Y}$ have the same distribution under η; (ii) $\widetilde{Y} = \widetilde{I} \circ X$ where $\widetilde{I}$ is nondecreasing, and (iii) $0 \leq \widetilde{Y} \leq X$, by Proposition 5.80 on p. 153.

Furthermore, since ν is (η, X)-vigilant, it follows from the definition of vigilance that

$$\oint \left(X - \widetilde{Y}\right) \, d\nu \geq \oint \left(X - Y\right) \, d\nu$$

However, $\oint \left(X - Y\right) \, d\nu \geq H$ since $Y \in \mathcal{F}_{SB}$. Therefore, $\oint \left(X - \widetilde{Y}\right) \, d\nu \geq H$. Thus, $\widetilde{Y} \in \mathcal{F}_{SB}^{\uparrow}$, and so $\mathcal{F}_{SB}^{\uparrow} \neq \varnothing$.

Moreover, by Proposition 5.11,

$$\begin{aligned}
\oint \left(X - \widetilde{Y}\right) \, d\nu \ &\geq \ \oint \left(X - Y\right) \, d\nu \\
&\iff \ \oint \left(W_0^{CI} - H + X - \widetilde{Y}\right) \, d\nu \geq \oint \left(W_0^{CI} - H + X - Y\right) \, d\nu \\
&\iff \ W^{CI}\left(H, \widetilde{Y}\right) \succcurlyeq_{CI} W^{CI}(H, Y)
\end{aligned}$$

where the last equivalence holds because we have assumed that the CI's utility function v is the identity function (Assumption 5.55 (1)). □

Remark 5.65. *What Theorem 5.64 asserts is that if vigilance holds, then for any function $Y_1 \in B^+(\Sigma)$ which is feasible for the DM's problem there is another feasible function $Y_2 \in B^+(\Sigma)$ which is (i) comonotonic with the underlying uncertainty X, and (ii) is such the contract (H, Y_2) is preferred by the CI to the contract (H, Y_1).*

5.6 Conclusion

In this chapter we examined a problem of *contracting for innovation* under heterogeneous and ambiguous beliefs in order to demonstrate how the techniques of Chapter 2 can be used beyond problems of insurance demand. We showed that even in such a framework, and even if preferences admit a representation that reflects a certain level of ambiguity in beliefs, the notion of *vigilance* yields the existence a solution to the DM's problem which is a nondecreasing function of the underlying *innovation* (what we called the underlying uncertainty in Chapter 2).

We then considered the special case where the CI's beliefs are represented by a *distorted probability* measure, with a concave distortion function. We provided a general technique for characterizing the solution of the DM's problem in terms of the solution of a relatively easier *quantile problem*, using a "splitting" procedure resembling what was done in Chapter 2.

Finally, we considered a situation where both the DM and the CI have preferences over the elements of choice that admit a representation that exhibits some ambiguity in their beliefs. We assumed that the CI is a CEU-maximizer and the DM's preferences are represented by a *symmetric* Choquet integral, hence reflecting some *gain-loss* separability, in the spirit of the *Cumulative Prospect Theory* of Kahneman and Tversky ([177] and [293]). We showed that in this setting, if *vigilance* holds, then for any payoff function Y_1 which is feasible for the DM's problem there is another feasible payoff function Y_2 which is (i) comonotonic with the underlying uncertainty, and (ii) such that for a given fixed fee $H > 0$ the CI prefers a contract in which he pays Y_2 to a contract in which he pays Y_1.

5.7 Appendix: A Dominated Convergence Theorem for the Choquet Integral

5.7.1 The Šipoš Integral

Given a capacity ν on a measurable space (S, Σ), recall from Definition 5.7 and Remark 5.8 that the *Choquet integral* of a given $\phi \in B(\Sigma)$ with respect to ν is defined as

$$\oint \phi \, d\nu := \int_0^{+\infty} \nu\left(\{s \in S : \phi(s) \geq t\}\right) \, dt + \int_{-\infty}^0 \left[\nu\left(\{s \in S : \phi(s) \geq t\}\right) - 1\right] \, dt$$
$$= \int_0^{+\infty} \nu\left(\{s \in S : \phi(s) > t\}\right) \, dt + \int_{-\infty}^0 \left[\nu\left(\{s \in S : \phi(s) > t\}\right) - 1\right] \, dt \tag{5.37}$$

If ϕ^+ (resp. ϕ^-) denotes the nonnegative (resp. nonpositive) part of ϕ, then

1. $\phi = \phi^+ + (-\phi^-)$; and,

2. ϕ^+ and $-\phi^-$ are comonotonic.

Therefore, by Proposition 5.10, we have:

$$\oint \phi \, d\nu = \oint \phi^+ \, d\nu + \oint \left(-\phi^-\right) \, d\nu \tag{5.38}$$

However, unless ν is a *bona fide* measure or $\phi \geq 0$, it does not hold that $\oint \left(-\phi^-\right) \, d\nu = -\oint \phi^- \, d\nu$. Indeed, by (5.37),

$$\oint \left(-\phi^-\right) \, d\nu = \int_0^{+\infty} \nu\left(\{s \in S : -\phi^-(s) > t\}\right) \, dt + \int_{-\infty}^0 \left[\nu\left(\{s \in S : -\phi^-(s) > t\}\right) - 1\right] \, dt$$
$$= 0 + \int_{-\infty}^0 \left[\nu\left(\{s \in S : -\phi^-(s) > t\}\right) - 1\right] \, dt$$
$$= \int_{-\infty}^0 \left[\nu\left(\{s \in S : \phi^-(s) < -t\}\right) - 1\right] \, dt$$

Letting $u = -t$, we then have

$$\oint \left(-\phi^-\right) \, d\nu = -\int_{+\infty}^0 \left[\nu\left(\{s \in S : \phi^-(s) < u\}\right) - 1\right] \, du$$
$$= \int_0^{+\infty} \left[\nu\left(\{s \in S : \phi^-(s) < u\}\right) - 1\right] \, du$$
$$= -\int_0^{+\infty} \left[1 - \nu\left(\{s \in S : \phi^-(s) < u\}\right)\right] \, du$$

On the other hand,

$$
\begin{aligned}
-\oint \phi^- \, d\nu &= -\int_0^{+\infty} \nu\left(\{s \in S : \phi^-(s) \geq u\}\right) \, du - \int_{-\infty}^0 \left[\nu\left(\{s \in S : \phi^-(s) \geq u\}\right) - 1\right] \, du \\
&= -\int_0^{+\infty} \nu\left(\{s \in S : \phi^-(s) \geq u\}\right) \, du
\end{aligned}
$$

However, since ν is not additive,

$$
\nu\left(\{s \in S : \phi^-(s) \geq u\}\right) \;\neq\; 1 - \nu\left(\{s \in S : \phi^-(s) < u\}\right)
$$

Therefore, if $\phi^- \neq 0$, then $\oint(-\phi^-) \, d\nu \neq -\oint \phi^- \, d\nu$, and so

$$
\oint \phi^+ \, d\nu + \oint \left(-\phi^-\right) \, d\nu \;\neq\; \oint \phi^+ \, d\nu - \oint \phi^- \, d\nu \tag{5.39}
$$

This is the motivation behind the following definition.

Definition 5.66 (The Šipoš Integral)**.** *Let ν be a given capacity on a measurable space (S, Σ), and let $\phi \in B(\Sigma)$. The Šipoš integral (a.k.a. the symmetric Choquet integral) of ϕ with respect to ν is defined as:*

$$
\square\!\!\!\!\int \phi \, d\nu := \oint \phi^+ \, d\nu - \oint \phi^- \, d\nu \tag{5.40}
$$

Proposition 5.67. *If ν is a given capacity on a measurable space (S, Σ), and if $\phi \in B^+(\Sigma)$, then*

$$
\oint \phi \, d\nu = \square\!\!\!\!\int \phi \, d\nu
$$

Proof. Trivial. □

Unlike the Choquet integral which is merely positively homogeneous (Proposition 5.12), the Šipoš integral is homogeneous, and therefore it is also symmetric:

Proposition 5.68. *If ν is a given capacity on a measurable space (S, Σ), and if $\phi \in B(\Sigma)$. Then for any $a \in \mathbb{R}$, we have*

$$\int\!\!\!\!\!\square\; a \ \phi \ d\nu = a \ \int\!\!\!\!\!\square\; \phi \ d\nu$$

In particular, the Šipoš integral is symmetric, in the sense that

$$\int\!\!\!\!\!\square\; (-\phi) \ d\nu = -\int\!\!\!\!\!\square\; \phi \ d\nu$$

Proof. See Denneberg [96] (Propostion 7.1 on p. 88) or Pap [221] (Theorem 7.10 on p. 155). □

5.7.2 A Dominated Convergence Theorem

Theorem 5.69 (Dominated Convergence)**.** *Let ν be a continuous capacity (Definition 5.3) on a measurable space (S, Σ). If $\{\phi_n\}_n$ is a sequence in $B(\Sigma)$ that converges pointwise to some $\phi \in B(\Sigma)$, then*

$$\lim_{n \to +\infty} \int\!\!\!\!\!\square\; \phi_n \ d\nu = \int\!\!\!\!\!\square\; \phi \ d\nu \tag{5.41}$$

Proof. This is a special case of a more general theorem given in Pap [221] (Theorem 7.16 on p. 166). □

Since the Choquet integral and the Šipoš integral of a nonnegative (bounded and measurable) function coincide, we then have the following corollary immediately:

Corollary 5.70. *Let ν be a continuous capacity on a measurable space (S, Σ). If $\{\phi_n\}_n$ is a sequence in $B^+(\Sigma)$ that converges pointwise to some $\phi \in B^+(\Sigma)$, then*

$$\lim_{n \to +\infty} \oint \phi_n \ d\nu = \oint \phi \ d\nu \tag{5.42}$$

There are several other convergence theorems for the Choquet integral in the literature, but not for pointwise convergent sequences. For instance:

1. By Marinacci and Montrucchio [204] (Proposition 4.11–(iv) on p. 64), the Choquet integral is a supnorm-continous operator on $B(\Sigma)$, and so it preserves uniform convergence, i.e. convergence in the supnorm on $B(\Sigma)$;

2. Greco [157] also gives a Dominated Convergence Theorem for the Choquet integral, but for uniformly convergent sequences of functions;

3. Denneberg [96] (Theorem 8.9 on p. 101) provides a Dominated Convergence Theorem for the Choquet integral (with respect to a given capacity ν), but for sequences of functions that converge *"in ν-distribution"*. Denneberg's definition of "convergence in ν-distribution" is similar to the usual definition for measures, with the exception that here ν is not additive.

5.8 Appendix: Equimeasurable Monotone Rearrangements with Respect to a Continuous Submodular Capacity

In section 2.3, we introduced a specific formulation of the *equimeasurable monotone P-rearrangement* of an element Y of $B^+(\Sigma)$ with respect to X, where $\Sigma = \sigma\{X\}$ and P is a given probability measure on the space (S, Σ). Any such Y can be written as $Y = I \circ X$, for some Borel-measurable nonnegative function on the range of X. The construction of the equimeasurable rearrangement of $\widetilde{Y}_P$ of Y with respect to X was defined as $\widetilde{I} \circ X$, where $\widetilde{I}$ was the equimeasurable ϕ-rearrangement of I, with $\phi = P \circ X^{-1}$.

In this Appendix we use a similar construction to define an *equimeasurable monotone η-rearrangement* of an element Y of $B^+(\Sigma)$ with respect to X, where η is a given continuous and submodular capacity on the space (S, Σ). The results presented here are new, to the best of our knowledge.

5.8.1 Distribution Functions

Let $(S, \mathcal{G})$ be a measurable space, and let $X \in B^+(\mathcal{G})$ with $[0, M] := X(S)$, where $M = \|X\|_s < +\infty$. Denote by Σ the σ-algebra $\sigma\{X\}$ of subsets of S generated by X.

Let η be a continuous submodular capacity on (S, Σ), and recall from section 5.2 that this means the following:

1. $\eta(\varnothing) = 0$;

2. $\eta(S) = 1$;

3. η is monotone: for any $A, B \in \Sigma,\ A \subseteq B \Rightarrow \eta(A) \leq \eta(B)$;

4. η is submodular: for each $A, B \in \Sigma$,
$$\eta(A \cup B) + \eta(A \cap B) \leq \eta(A) + \eta(B) \tag{5.43}$$

5. η continuous from above: for any sequence $\{A_n\}_n$ in Σ such that $A_{n+1} \subseteq A_n$ for each $n \geq 1$, we have:
$$\lim_{n \to +\infty} \eta(A_n) = \eta\left(\bigcap_{n=1}^{+\infty} A_n\right) \tag{5.44}$$

6. η is continuous from below: for any sequence $\{A_n\}_n$ in Σ such that $A_n \subseteq A_{n+1}$ for each $n \geq 1$, we have:
$$\lim_{n \to +\infty} \eta(A_n) = \eta\left(\bigcup_{n=1}^{+\infty} A_n\right) \tag{5.45}$$

Definition 5.71. *We define the "image capacity" of η under X as the quantity $\psi := \eta \circ X^{-1}$, where for each Borel set B, $\psi(B) = \eta\big(X^{-1}(B)\big)$.*

Proposition 5.72. *The set function ψ is a continuous submodular capacity on the Borel σ-algebra of the range of X.*

Proof. Immediate. □

Definition 5.73. *For any Borel-measurable map $I : [0, M] \to \mathbb{R}$, define the distribution function of I as the map $\psi_I : \mathbb{R} \to [0, 1]$ given by*
$$\psi_I(t) := \psi\big(\{x \in [0, M] : I(x) \leq t\}\big) \tag{5.46}$$

Proposition 5.74. *For any Borel-measurable map $I : [0, M] \to \mathbb{R}$, the function $\psi_I : \mathbb{R} \to [0, 1]$ is nondecreasing and right-continuous.*

Proof. The monotonicity (resp. right-continuity) of ψ_I is a consequence of the monotonicity (resp. continuity) of ψ. □

5.8.2 Equimeasurable Rearrangements

Here, we will construct a ψ-equimeasurable rearrangement of any Borel-measurable map $I : [0, M] \to [0, M]$, assuming that η satisfies a "nonatomicity" condition.

Definition 5.75. *We say that $\psi = \eta \circ X^{-1}$ is nonatomic if for any Borel set A with $\psi(A) > 0$, there is some Borel set $B \subsetneq A$ such that $0 < \psi(B) < \psi(A)$.*

When $\eta \circ X^{-1}$ is nonatomic, we will say that X is a continuous random variable for η.

Assumption 5.76. *$\psi = \eta \circ X^{-1}$ is nonatomic.*

Remark 5.77. *Since ψ is nonatomic, $\psi(\{t\}) = 0$, for each $t \in [0, M]$. Therefore, it follows that $\psi\big([0, t]\big) = \psi\big([0, t)\big)$, for each $t \in [0, M]$. Indeed, we have $\psi\big([0, t]\big) \geq \psi\big([0, t)\big)$, by monotonicity of ψ. Moreover, since $[0, t] = [0, t) \cup \{t\}$, it follows from the submodularity of ψ that*

$$\psi\big([0, t]\big) = \psi\big([0, t) \cup \{t\}\big) + \psi\big([0, t) \cap \{t\}\big) \leq \psi\big([0, t)\big) + \psi\big(\{t\}\big) = \psi\big([0, t)\big)$$

Similarly, $\psi\big((t, M]\big) = \psi\big([t, M]\big)$, for each $t \in [0, M]$.

Definition 5.78. *Let $I : [0, M] \to [0, M]$ be any Borel-measurable map, and define the function $\widetilde{I} : [0, M] \to \mathbb{R}$ by:*

$$\widetilde{I}(t) := \inf\Big\{z \in \mathbb{R}^+ : \psi_I(z) \geq \psi\big([0, t]\big)\Big\} \tag{5.47}$$

The following proposition gives some useful properties of the map $\widetilde{I}$ defined above.

Proposition 5.79. *Let $I : [0, M] \to [0, M]$ be any Borel-measurable map and let $\widetilde{I} : [0, M] \to \mathbb{R}$ be defined as in* (5.47)*. If Assumption 5.76 holds, then:*

1. *$\widetilde{I}$ is left-continuous, nondecreasing, and Borel-measurable;*

2. *For each $t \in [0, M]$, $\psi_I\left(\widetilde{I}(t)\right) \geq \psi([0, t])$;*

3. *$\widetilde{I}(t) \geq 0$, for each $t \in [0, M]$, $\widetilde{I}(0) = 0$, and $\widetilde{I}(M) \leq M$;*

4. *If $I_1, I_2 : [0, M] \to [0, M]$ are such that $I_1 \leq I_2$ except on a Borel set C such that $\psi(C) = 0$, then $\widetilde{I}_1 \leq \widetilde{I}_2$;*

5. *If $Id : [0, M] \to [0, M]$ denotes the identity function, then $\widetilde{Id} \leq Id$;*

6. *$\widetilde{I}$ is ψ-equimeasurable with I, in the sense that for any $\alpha \in [0, M]$,*

$$\psi\Big(\{t \in [0, M] : I(t) \leq \alpha\}\Big) = \psi\Big(\{t \in [0, M] : \widetilde{I}(t) \leq \alpha\}\Big) \tag{5.48}$$

Proof.

1. The monotonicity of $\widetilde{I}$, and hence its Borel-measurability, follows from the monotonicity of ψ. By Remark 5.77 and by the monotonicity and the continuity of ψ, left-continuity of $\widetilde{I}$ is an immediate consequence of the left-continuity of the function $\psi_I^*(t) := \inf\left\{z \in \mathbb{R}^+ : \psi_I(z) \geq t\right\}$, for $t \in [0, 1]$ (i.e. the left-continuous inverse[11] of ψ_I);

2. This is an immediate consequence of the right-continuity of the distribution function ψ_I of I;

3. Similar to the proof of Proposition 2.9 (3);

[11]See also Embrechts and Hofert [119] for more about the left-inverse (a.k.a. the left-continuous inverse) of a nondecreasing function.

4. Let $I_1, I_2 : [0, M] \to [0, M]$ be such that $I_1 \le I_2$, except on a Borel subset C of $[0, M]$ with $\psi(C) = 0$. Let $D := [0, M] \setminus C$, then for each $x \ge 0$,

$$\psi\Big(\{t \in [0, M] : I_1(t) \le x\}\Big) \ge \psi\Big(\{t \in D : I_1(t) \le x\}\Big),$$

by monotonicity of ψ. On the other hand, since ψ is submodular, we have

$$\begin{aligned}
\psi\Big(\{t \in [0, M] : I_1(t) \le x\}\Big) &= \psi\Big(\{t \in D : I_1(t) \le x\} \cup \{t \in C : I_1(t) \le x\}\Big) \\
&= \psi\Big(\{t \in D : I_1(t) \le x\} \cup \{t \in C : I_1(t) \le x\}\Big) \\
&\quad + \psi\Big(\{t \in D : I_1(t) \le x\} \cap \{t \in C : I_1(t) \le x\}\Big) \\
&\le \psi\Big(\{t \in D : I_1(t) \le x\}\Big) + \psi\Big(\{t \in C : I_1(t) \le x\}\Big) \\
&= \psi\Big(\{t \in D : I_1(t) \le x\}\Big)
\end{aligned}$$

where the last equality follows form monotonicity of ψ and from the fat that $\psi(C) = 0$. Therefore, for each $x \ge 0$, we have

$$\psi\Big(\{t \in [0, M] : I_1(t) \le x\}\Big) = \psi\Big(\{t \in D : I_1(t) \le x\}\Big)$$

Similarly, or each $x \ge 0$, we have

$$\psi\Big(\{t \in [0, M] : I_2(t) \le x\}\Big) = \psi\Big(\{t \in D : I_2(t) \le x\}\Big)$$

Since $I_1 \le I_2$ on D, it follows from the monotonicity of ψ that for each $x \ge 0$,

$$\begin{aligned}
\psi\Big(\{t \in [0, M] : I_1(t) \le x\}\Big) &= \psi\Big(\{t \in D : I_1(t) \le x\}\Big) \\
&\ge \psi\Big(\{t \in D : I_2(t) \le x\}\Big) \\
&= \psi\Big(\{t \in [0, M] : I_2(t) \le x\}\Big)
\end{aligned}$$

Therefore, for each $t \in [0, M]$,

$$\begin{aligned}
&\Big\{z \in \mathbb{R}^+ : \psi\big(\{x \in [0, M] : I_2(x) \le z\}\big) \ge \psi\big([0, t]\big)\Big\} \\
&\qquad \subseteq \Big\{z \in \mathbb{R}^+ : \psi\big(\{x \in [0, M] : I_1(x) \le z\}\big) \ge \psi\big([0, t]\big)\Big\}
\end{aligned}$$

It then follows from (2.12) that $\widetilde{I}_1 \leq \widetilde{I}_2$;

5. Similar to the proof of Proposition 2.9 (5);

6. To show that for each $\alpha \in [0, M]$,

$$\psi\Big(\{t \in [0, M] : I(t) \leq \alpha\}\Big) = \psi\Big(\{t \in [0, M] : \widetilde{I}(t) \leq \alpha\}\Big)$$

the proof is similar to the one in the proof of Proposition 2.9 (6), bearing in mind Remark 5.77.

□

$\widetilde{I}$ will be called a nondecreasing ψ-rearrangement of I. Now, define $Y := I \circ X$ and $\widetilde{Y} := \widetilde{I} \circ X$. Since both I and $\widetilde{I}$ are Borel-measurable mappings of $[0, M]$ into itself, it follows that $Y, \widetilde{Y} \in B^+(\Sigma)$. Note also that $\widetilde{Y}$ is nondecreasing in X, in the sense that if $s_1, s_2 \in S$ are such that $X(s_1) \leq X(s_2)$ then $\widetilde{Y}(s_1) \leq \widetilde{Y}(s_2)$, and that Y and $\widetilde{Y}$ are η-equimeasurable, that is, for any $\alpha \in [0, M]$, $\eta(\{s \in S : Y(s) \leq \alpha\}) = \eta\left(\{s \in S : \widetilde{Y}(s) \leq \alpha\}\right)$. Indeed,

$$\begin{aligned}
\eta\Big(s \in S : \widetilde{Y}(s) \leq \alpha\Big) &= \eta\Big(\{s \in S : X(s) \in \{t \in [0, M] : \widetilde{I}(t) \leq \alpha\}\}\Big) \\
&= \psi\Big(\{t \in [0, M] : \widetilde{I}(t) \leq \alpha\}\Big) \\
&= \psi\Big(\{t \in [0, M] : I(t) \leq \alpha\}\Big) \\
&= \eta\Big(\{s \in S : X(s) \in \{t \in [0, M] : I(t) \leq \alpha\}\}\Big) \\
&= \eta\Big(s \in S : Y(s) \leq \alpha\Big)
\end{aligned}$$

We will then call $\widetilde{Y}$ a **nondecreasing η-rearrangement of Y with respect to X**, and we shall denote it by $\widetilde{Y}_\eta$ to avoid confusion in case a different continuous submodular capacity on (S, Σ) is also considered.

Finally, in light of Proposition 5.79, we have the following:

Proposition 5.80. *Under Assumption 5.76, if $Y \in B^+(\Sigma)$ is such that $0 \leq Y \leq X$, then $0 \leq \widetilde{Y}_\eta \leq X$.*

Proof. With $M = \|X\|_s$, as before, since Y is Σ-measurable, by Doob's measurability theorem there is a real-valued bounded Borel-measurable mapping I on $[0, M]$ such that $Y = I \circ X$. Moreover, we can write $X = Id \circ X$, where Id denotes the identity map on $[0, M]$.

If $0 \leq Y \leq X$ then $0 \leq I \leq Id$, since $X(S) = [0, M]$. Therefore, by Proposition 5.79 (4) and (5), $0 \leq \widetilde{I} \leq \widetilde{Id} \leq Id$, where $\widetilde{I}$ denotes the nondecreasing ψ-rearrangement of I and where $\widetilde{Id}$ denotes the nondecreasing ψ-rearrangement of Id. Hence, $0 \leq \widetilde{Y}_\eta \leq X$. □

Chapter 6

Conclusion and Future Work

6.1 Some Concluding Remarks

In this thesis we have considered the effect of the *subjectivity of beliefs* in contracting problems, and especially in problems of insurance demand, which were the motivation behind this thesis. As we mentioned in Chapter 1, the idea that the random loss in insurance models is a given random variable on an objectively and exogenously given probability space is inherited from the von Neumann-Morgenstern approach, stipulating that uncertainty is totally objective. In contrast, the work done in this thesis is based on Savage's [266] approach to uncertainty, and is entirely in the *subjectivist tradition* of De Finetti [93] and Ramsey [234].

This subjectivist foundation then gave us a proper framework in which heterogeneity of beliefs would arise naturally as a consequence of the heterogeneity of preferences. We then examined an insurance model in which the beliefs of both parties are subjective and heterogeneous. In other words, there are two probability measures on an underlying measurable space, and the insurable loss is a random variable on that underlying space. Each party then evaluates the likelihood of the different realizations of this loss according to their own probability measure. This is radically different from the "classical" insurance model, both in scope and philosophy. Nevertheless, we showed that under a consistency requirement on these different probability measures that we called *vigilance*, we can show the existence of an optimal contract which is a nondecreasing function of the insurable loss. Moreover, we provided a general technique for characterizing the solution, and we did so in terms of a *generalized deductible contract*.

Finally, we examined a problem of *contracting for innovation* under heterogeneous and ambiguous beliefs, to illustrate the techniques of Chapter 2 outside of the insurance framework. We showed that in that case, and even if preferences admit a representation that reflects a certain level of ambiguity in beliefs, the notion of *vigilance* yields the existence a solution to the DM's problem which is a nondecreasing function of the underlying *innovation* (what was called the underlying uncertainty in Chapter 2). We then characterized the solution is a special case, using a "slitting" procedure as in Chapter 2.

In this chapter we will discuss some possible extensions of the results presented in this thesis, as well as some more properties of the notion of *vigilance*.

6.2 The Effect of the "Distance" between Subjective Beliefs

The work done in Chapter 3 did not consider the effect of a change in the insurer's subjective probability measure on the shape of the optimal contract, or on the existence of a monotone solution to the DM's demand problem. Future research will examine how the shape of the optimal contract changes with the "distance" between the subjective beliefs of the DM and the insurer, for an appropriately defined notion of "distance".

There are several ways in which one can define a "distance" between two probability measures. Let P and Q be probability measures on a measurable space $(\Omega, \mathcal{F})$. We can then define a "distance" between P and Q in many different ways, such as:

1. If $Q << P$, let $d_{KL}(P,Q) = D_{\mathrm{KL}}(P \| Q) = -\int_\Omega \log \frac{\mathrm{d}Q}{\mathrm{d}P}\, \mathrm{d}P$, where for any probability measures $\mu_2 << \mu_1$, $D_{\mathrm{KL}}(\mu_1 \| \mu_2)$ is *Kullback-Leibler divergence* from μ_1 to μ_2.

2. $d_H(P,Q) = H^2(P,Q) = \frac{1}{2}\int_\Omega \left(\sqrt{\frac{dP}{d\lambda}} - \sqrt{\frac{dQ}{d\lambda}}\right)^2 d\lambda$, where $\lambda = (P+Q)/2$, and where for any probability measures μ_1, μ_2, and μ_3 such that $\mu_1 << \mu_3$ and $\mu_2 << \mu_3$, $H^2(\mu_1, \mu_2)$ is called the *Hellinger distance*, and it does not depend on the specific choice of the probability measure μ_3. For instance, $\mu_3 = (\mu_1 + \mu_2)/2$ is acceptable.

3. $d_{JS}(P,Q) = \frac{1}{2}D_{\mathrm{KL}}(P \| \lambda) + \frac{1}{2}D_{\mathrm{KL}}(Q \| \lambda)$, where $\lambda = (P+Q)/2$. The quantity $d_{JS}(P,Q)$ is called the *Jensen-Shannon divergence*.

4. For a given convex function f with $f(1) = 0$, if $Q << P$, let $d_f(P, Q) = D_f(Q \| P) = \int_\Omega f\left(\frac{dQ}{dP}\right) dP$, where for any probability measures $\mu_2 << \mu_1$, $D_f(\mu_2 \| \mu_1)$ is *f-divergence* of μ_2 from μ_1.

5. $d_{TV}(P, Q) = \sup\{|P(A) - Q(A)| : A \in \mathcal{F}\}$ is the *total variation distance* between the probability measures P and Q.

6. $d_W(P, Q) = \sup\left\{\left|\int_\Omega f\ dP - \int_\Omega f\ dQ\right| : f \text{ is 1-Lipschitz}\right\}$ is the *Wassertein distance*, or the *Kantorovich-Monge distance* between the probability measures P and Q.

6.3 The Demand for Insurance in the Presence of Moral Hazard and Belief Heterogeneity

6.3.1 *Ex-ante* Moral Hazard in Insurance Models

The setting of Chapter 3 did not incorporate the possibility of moral hazard in the insurance model considered. Future work will consider such a possibility, while maintaining heterogeneity of subjective beliefs. Specifically, we will consider a situation where the DM can exercise an *effort* in preventing the loss X or reducing its severity. This effort level is *unknown* to the insurer, and hence *ex ante* moral hazard exists.

Before we can state the mathematical model that suits this situation, we need the following definition:

Definition 6.1. *A probability kernel from a measurable space $(S_1, \mathcal{F}_1)$ to a measurable space $(S_2, \mathcal{F}_2)$ is a mapping $Q : S_1 \times \mathcal{F}_2 :\to [0, 1]$ such that:*

1. *For each $t \in S_1$, the mapping*

$$\begin{aligned} Q(t, .) : \mathcal{F}_2 &\to [0, 1] \\ B &\mapsto Q(t, B) \end{aligned} \tag{6.1}$$

is a (countably additive) probability measure on $(S_2, \mathcal{F}_2)$;

2. *For each $B \in \mathcal{F}_2$, the mapping*

$$\begin{aligned} Q(.,B) : S_1 &\to [0,1] \\ t &\mapsto Q(t,B) \end{aligned} \tag{6.2}$$

is $\mathcal{F}_1$-measurable.

6.3.2 A Model of Moral Hazard in Insurance

Mathematically, we will model this insurance market with moral hazard as follows:

1. As in section 3.2, S denotes the set of all *states of the world*, and $\mathcal{G}$ is a σ-algebra of *events* on S.

2. The insurable loss is a fixed element X of $B^+(\mathcal{G})$ with closed range $X(S) = [0, M]$, where $M = \|X\|_s < +\infty$. Let Σ be the σ-algebra $\sigma\{X\}$ of subsets of S generated by the random loss X.

3. The insurance market gives the DM the possibility of entering into an insurance contract with the insurer. Such a contract is represented by a pair (Π, I), where $\Pi > 0$ is the premium paid by the DM in return of the indemnity I. The indemnity is a Borel-measurable map $I : [0, M] \to [0, M]$, such that $0 \leq I(X(s)) \leq X(s)$ for all $s \in S$. Then $Y := I \circ X \in B^+(\Sigma)$.

4. Both the DM and the insurer have preferences over the elements of $B^+(\Sigma)$. The insurer's preferences $\succcurlyeq_{CI}$ have a Subjective Expected-Utility (SEU) representation of the form
$$Y_1 \succcurlyeq_{CI} Y_2 \iff \int_S v(Y_1)\, d\nu \geq \int_S v(Y_2)\, d\nu$$
where $v : \mathbb{R} \to \mathbb{R}$ is a continuous utility, bounded on bounded sets, and ν is the insurer's subjective probability measure on (S, Σ).

5. $(E, \mathcal{E})$ is a measurable space of possible *effort levels* that the DM can exercise, and $\succcurlyeq_E$ is a total order on E. The cost of effort is modeled by a cost function $c : E \to \mathbb{R}^+$, where c is nondecreasing, i.e. for each $e_1, e_2 \in E$,
$$e_1 \succcurlyeq_E e_2 \implies c(e_1) \geq c(e_2)$$

Both the effort level e and its cost $c(e)$ are unknown to the insurer, and unobservable by him.

6. The DM's preferences $\succcurlyeq_{DM}$ over $B^{+}(\Sigma)$ yield the existence of a probability kernel Q form $(E, \mathcal{E})$ to (S, Σ) and a continuous, bounded, concave, and increasing utility function $u : \mathbb{R} \to \mathbb{R}$ such that for each effort level $e \in E$ and for each $Y_1, Y_2 \in B^{+}(\Sigma)$,

$$Y_1 \succcurlyeq_{DM} Y_2 \iff \int_S u(Y_1)\ dQ(e,.) \geq \int_S u(Y_2)\ dQ(e,.)$$

7. The DM has initial wealth $W_0 > \Pi > 0$, and for each effort level e the total wealth of the DM is the Σ-measurable, $\mathbb{R}$-valued and bounded function on S defined by

$$W_e(s) := W_0 - \Pi - c(e) - X(s) + Y(s),\ \forall s \in S \tag{6.3}$$

where $c(e)$ is the cost associated with the effort level e.

8. The insurer has initial wealth W_0^{CI} and final wealth $W^{CI} \in B(\Sigma)$.

6.3.3 The DM's Demand Problem

The DM's problem can then be formulated as the following problem:

Problem 6.2.

$$\sup_{Y \in B^{+}(\Sigma)} \left\{ \int_S u\Big(W_0 - \Pi - c(e) - X + Y\Big)\ dQ(e,.) \right\} :$$

$$\begin{cases} 0 \leq Y \leq X \\ \int v\left(W^{CI}\right)\ d\nu \geq v\left(W_0^{CI}\right) \\ e \in \arg\max\limits_{z \in E} \Big[\int_S u\big(W_0 - \Pi - c(z) - X + Y\big)\ dQ(z,.) \Big] \end{cases}$$

The first constraint has been discussed previously. The second constraint is simply the insurer's *participation constraint*, or *individual rationality constraint*, where $v\left(W_0^{CI}\right)$

is the insurer's *reservation utility*. The third constraint is commonly called an *incentive-compatibility constraint* in the literature on principal-agent problems. However, in that context, the beliefs of the principal and the agent are identical, for a given effort level.

For more about moral hazard in principal-agent problems, including an interpretation of the *incentive-compatibility constraint*, we refer the reader to Arnott and Stiglitz [20], [21], and [22], Conlon [84], Grossman and Hart [160], Jewitt [174], Kadan and Swinkels [176], Mirrlees [212], Page [219] and [220], Rogerson [245], Shavell [278], Sinclair-Desgagné [279] and [280], or Stiglitz [288], for instance.

6.4 The DM's Preference for "Compatibility"

6.4.1 A Collection of Insurers

We saw in Chapter 3 that one possible interpretation of the notion of *vigilance* is that the insurer assigns a certain *credibility* to the DM's assessment of a given risk Y that is a function of the underlying random loss X.

However, in the setting of Chapter 3, both the DM and the insurer are given *a priori*. The DM is not assumed to have a choice over which insurer to purchase the insurance coverage from. Future work on the notion of *vigilance* will also examine a situation where the DM not only seeks an insurance contract that maximizes her expected utility of terminal wealth with respect to her subjective measure, but she also has a choice between which insurers to contract with, and will seek the "best" such insurer according to some representation of some preference over a suitable set of available insurers.

It is hence natural to start by defining a suitable collection of available insurers with which the DM can contract, and then define a proper notion of preference among these insurers. For this, we need the following terminology:

Definition 6.3 (Compatibility). *For a given DM whose beliefs are represented by a subjective probability measure ν, we will say that the DM is compatible with an insurer CIN if the insurer's beliefs are represented by a subjective probability measure ν which is (μ, X)-vigilant. Alternatively, we will also say that the DM and the insurer CIN are compatible.*

6.4.2 A Ranking of Insurers

We will start with a DM and a collection of *compatible insurers*. Given this collection of insurers, we assume that the DM has an ability to rank these *compatible insurers* via a "*compatibility preference*". This preference is to be interpreted as follows: a *compatible insurer* CIN_1 is preferred by the DM to another *compatible insurer* CIN_2 if the DM believes that CIN_1 is "*more compatible*" than CIN_2.

The DM then chooses which *compatible insurer* to contract with, solely on the basis of her *compatibility preference*. Once her choice of insurer is made, the DM then seeks the optimal form of insurance coverage as in Chapter 3.

We will take as primitives of our model the DM's preference $\succcurlyeq_{DM}$ over indemnity schedules (i.e. elements of $B^+(\Sigma)$), and her preference over *compatible insurers*. Moreover, for a given *compatible insurer*, the insurer's preference over $B^+(\Sigma)$ is also taken as a primitive.

Formally, the DM's preference $\succcurlyeq_{DM}$ over $B^+(\Sigma)$ determine her utility u and a subjective probability measure μ representing her beliefs. As in Chapter 4, we let $\mathcal{C}_\mu \subset ca_1^+(\Sigma)$ denote the collection of all (μ, X)-vigilant beliefs.

Definition 6.4. *We denote by $\mathcal{CIN}$ the collection of all compatible insurers, indexed by $\mathcal{C}_\mu$. That is, $CIN \in \mathcal{CIN}$ if and only if $\nu \in \mathcal{C}_\mu$, where $\nu \in ca_1^+(\Sigma)$ represents CIN's beliefs.*

Then the DM is assumed to have a preference over the collection $\mathcal{CIN}$ of all *compatible insurers*:

Assumption 6.5. *The DM has a preference $\succcurlyeq_{comp}$ over $\mathcal{CIN}$.*

The problem is then to give a certain axiomatization of this preference $\succcurlyeq_{comp}$ in order to obtain a representation, in the spirit of the various representation theorems given in Appendix A. To do so, we need to

6.4.3 A Representation Theorem for the "Compatibility Preference"

Here we will outline two possible ways in which one can obtain an integral representation of the preference relation $\succcurlyeq_{comp}$. The first method is a straightforward application of the von Neumann-Morgenstern Representation Theorem (Theorem A.11 on p. 174), while the second relies heavily on the mathematical structure of $\mathcal{C}_\mu$, as discussed in Chapter 4.

Method I

By definition of $\mathcal{CIN}$, we can identify $\mathcal{CIN}$ with $\mathcal{C}_\mu$, and hence assume that $\succcurlyeq_{comp}$ is a preference over elements of $\mathcal{C}_\mu$. By Proposition 4.5, the collection of all *vigilant* beliefs is a convex subset of $ca_1^+(\Sigma)$. Then following the results of section A.1.2, or Fishburn [128] (pp. 137-143), we can give axioms that would imply an integral representation for $\succcurlyeq_{comp}$ of the form

$$CIN_1 \succcurlyeq_{comp} CIN_2 \iff \int_S \xi \, d\nu_1 \geq \int_S \xi \, d\nu_2, \ \forall CIN_1, CIN_2 \in \mathcal{CIN} \tag{6.4}$$

where $\xi : S \to \mathbb{R}$ is a bounded "utility" function.

Method II

The approach outlined here relies on the properties of collections of *vigilant* beliefs given in Propositions 4.6 and 4.7, as well as on the results of Propositions 4.21 and 4.22. Indeed, by identifying $\mathcal{CIN}$ with $\mathcal{C}_\mu$, the results of either Proposition 4.21 or Proposition 4.22 can be used to give an integral representations for $\succcurlyeq_{comp}$, as long as we can find a linear functional $L : \mathcal{C}_\mu \to \mathbb{R}$ which is continuous in either the weak or weak* topologies on $\mathcal{C}_\mu$.

A possible starting point would be the classical *Mixture Space Theorem* of Herstein and Milnor [165].

Definition 6.6. *A mixture space is a nonempty set M and a collection $\{r_\alpha\}_{\alpha\in[01]}$ of functions $r_\alpha : M \times M \to M$, such that for all $s, t \in M$ and for all $\alpha, \beta \in [0,1]$,*

1. $r_1(s,t) = s$;
2. $r_\alpha(s,t) = r_{1-\alpha}(t,s)$;
3. $r_\alpha(r_\beta(s,t), t) = r_{\alpha\beta}(s,t)$.

For instance, any convex subset of a real vector space is a mixture space, with the operation of vector addition being the mixing operation. Therefore, by Proposition 4.5, the collection of all *vigilant* beliefs is a *mixture space*, for the usual vector operations on the real vector space $ba\,(\Sigma)$ of all finitely additive measures of bounded variation on (S, Σ). That is, $r_\alpha\,(\nu_1, \nu_2) = \alpha\nu_1 + (1-\alpha)\,\nu_2$.

Theorem 6.7 (Mixture Space Theorem ([165]))**.** *Let $\succcurlyeq$ be a binary relation on a mixture space $\Big(M, \{r_\alpha\}_{\alpha\in[01]}\Big)$. Then the following are equivalent:*

1. *There exists a function $L : M \to \mathbb{R}$ such that for all $s, t \in M$ and for all $\alpha \in [0,1]$,*
   (a) *$s \succcurlyeq t \;\Leftrightarrow\; L\,(s) \geq L\,(t)$;*
   (b) *$L\big(r_\alpha\,(s,t)\big) = \alpha L\,(s) + (1-\alpha)\,L\,(t)$.*
2. *The binary relation $\succcurlyeq$ satisfies the following three axioms:*
   (a) *Axiom 1 (weak order): $\succcurlyeq$ is complete and transitive, that is, (i) for all $s, t \in M$, either $s \succcurlyeq t$, or $t \succcurlyeq s$, or both; and (ii) for all $s, t, z \in M$, if $s \succcurlyeq t$ and $t \succcurlyeq z$, then $s \succcurlyeq z$;*
   (b) *Axiom 2 (Independence): for all $\alpha \in (0,1]$ and for all $s, t, z \in M$, if $s \succcurlyeq t$ then $r_\alpha\,(s,z) \succcurlyeq r_a\,(t,z)$;*
   (c) *Axiom 3 (continuity): for all $s, t, z \in M$, the sets $\{\alpha \in [0,1] : r_\alpha\,(s,t) \succcurlyeq z\}$ and $\{\alpha \in [0,1] : z \succcurlyeq r_\alpha\,(s,t)\}$ are closed.*

Moreover, if L represents $\succcurlyeq$ in the sense of (1)(a), then L is unique up to a positive affine transformation.

By identifying $\mathcal{CIN}$ with $\mathcal{C}_\mu$, we can think of $\mathcal{CIN}$ as a mixture space with the mixing operation defined as follows: fix $CIN_1, CIN_2 \in \mathcal{CIN}$. Then there are $\nu_1, \nu_2 \in \mathcal{C}_\mu$ such that ν_1 represents CIN_1's beliefs and ν_2 represents CIN_2's beliefs. We then define, for each $\alpha \in [0,1]$,

$$r_\alpha\,(CIN_1, CIN_2) := \alpha\nu_1 + (1-\alpha)\,\nu_2 \tag{6.5}$$

Then r hence defined is a mixing operation on $\mathcal{CIN}$ since $\mathcal{C}_\mu$ is a convex set (Proposition 4.5). Hence we can assume that $\succcurlyeq_{comp}$ is a preference over elements of the mixture space

$\left(\mathcal{CIN}, \{r_\alpha\}_{\alpha\in[01]}\right)$ hence defined. The *Mixture Space Theorem* can then be applied to $\left(\mathcal{CIN}, \{r_\alpha\}_{\alpha\in[01]}\right)$ to obtain a representation $L : \mathcal{CIN} \to \mathbb{R}$ of the preference relation $\succcurlyeq_{comp}$.

APPENDICES

Appendix A

Some Elements of the Theory of Choice under Risk and Uncertainty

A.1 The Classical Theory

In this section we will review the classical theory of choice under risk and uncertainty, as introduced by von Neumann and Morgenstern [297], Savage [266], and Anscombe and Aumann [19]. First, however, we start from the very basics: binary relations on arbitrary sets and numerical representations of order relations. This will then lead us to von Neumann and Morgenstern's Expeted-Utility Representation Theorem, to Savage's Subjective Expected-Utility Representation Theorem, and to Anscombe and Aumann's Subjective Expected-Utility Representation Theorem.

A.1.1 Preliminaries

Preferences

Binary Relations Throughout this chapter, let S denote an arbitrary nonempty set, and let $S \times S$ denote the usual Cartesian product of S with itself. Then $S \times S$ is the collection of all ordered pairs of elements of S. That is,

$$S \times S := \{(x, y) : x, y \in S\} \tag{A.1}$$

Definition A.1. *A binary relation on S is a subset $\mathcal{R}$ of $S \times S$, that is, a collection of ordered pairs of elements of S. We write $x\mathcal{R}y$ to mean that $(x, y) \in \mathcal{R}$, and we write $\neg(x\mathcal{R}y)$ to mean that $(x, y) \notin \mathcal{R}$.*

The following proposition shows that the concept of a binary relation is "*well-defined*", in a sense.

Proposition A.2. *If $\mathcal{R}$ is a binary relation on an arbitrary nonempty set S then for any $x, y \in S$, only one of the following holds:*

1. *$x\mathcal{R}y$; or*

2. *$\neg(x\mathcal{R}y)$.*

Moreover, only one of the following holds:

1. *$x\mathcal{R}y$ and $y\mathcal{R}x$; or,*

2. *$x\mathcal{R}y$ and $\neg(y\mathcal{R}x)$; or,*

3. *$\neg(x\mathcal{R}y)$ and $y\mathcal{R}x$; or,*

4. *$\neg(x\mathcal{R}y)$ and $\neg(y\mathcal{R}x)$.*

Proof. See Fishburn [128], p. 10. □

Thus far, we have introduced the notion of a binary relation and shown that this concept is *philosophically consistent*. We now turn to some properties that a binary relation might have.

Definition A.3. *Let $\mathcal{R}$ be a binary relation on an arbitrary nonempty set S. Then $\mathcal{R}$ is said to be:*

1. *Reflexive if $x\mathcal{R}x$, for each $x \in S$;*

2. *Irreflexive if $\neg(x\mathcal{R}x)$, for each $x \in S$;*

3. *Symmetric if* $x\mathcal{R}y \Rightarrow y\mathcal{R}x$, *for each* $x, y \in S$;

4. *Asymmetric if* $x\mathcal{R}y \Rightarrow \neg(y\mathcal{R}x)$, *for each* $x, y \in S$;

5. *Antisymmetric if* $(x\mathcal{R}y \text{ and } y\mathcal{R}x) \Rightarrow (x = y)$, *for each* $x, y \in S$;

6. *Transitive if* $(x\mathcal{R}y \text{ and } y\mathcal{R}z) \Rightarrow x\mathcal{R}z$, *for each* $x, y, z \in S$;

7. *Negatively transitive if*

$$(\neg(x\mathcal{R}y) \text{ and } \neg(y\mathcal{R}z)) \Rightarrow \neg(x\mathcal{R}z), \text{ for each } x, y, z \in S;$$

8. *Weakly complete if* $(x \neq y) \Rightarrow (x\mathcal{R}y \text{ or } y\mathcal{R}x)$, *for each* $x, y \in S$;

9. *Complete if* $x\mathcal{R}y$ *or* $y\mathcal{R}x$ *or both, for each* $x, y \in S$;

To illustrate the previous definition, consider the usual strict order $>$ on $\mathbb{R}$. Then $>$ is irreflexive, asymmetric, transitive, negatively transitive, and weakly complete. Now, just as $>$ on $\mathbb{R}$ possesses more than just one of the above property, a binary relation $\mathcal{R}$ on an arbitrary nonempty set S might have more than just one of the above properties. It is customary to give special names to special combinations of the above properties, as in the following definition.

Definition A.4. *Let* $\mathcal{R}$ *be a binary relation on an arbitrary nonempty set* S. *Then* $\mathcal{R}$ *is said to be:*

1. *A weak order if it is asymmetric and negatively transitive;*

2. *A strict order if it is a weakly complete weak order;*

3. *An equivalence if it is reflexive, transitive, and symmetric.*

As an example, the strict order $>$ on $\mathbb{R}$ is indeed a *strict order* in the understanding of the above definition. The usual equality $=$ on $\mathbb{R}$ is easily seen to be an *equivalence* relation.

Preference Relations Theories of choice often take a preference relation over some collection S of elements of choice as a primitive notion. The Decision Maker (DM) is assumed to express his preferences over those elements of choice via statements of the form "I strictly prefer x to y". We will write this as $x \succ y$, for $x, y \in S$, and where $\succ \subseteq S \times S$ is a binary relation that will express the DM's strict preference over elements of S.

A moment's reflection will show that two natural properties to require of $\succ$ are asymmetry and negative transitivity. Henceforth, we will thus assume that a DM's preference over elements of S are expressed by a *weak order* $\succ$ on S that we will refer to as a *strict preference*. Hence, the DM's strict preference $\succ$ is a primitive of choice. We assume it to be given prior to any theoretical investigation. Form this primitive notion $\succ$, we can define two additional binary relations on S to express the notions of *weak preference* and *indifference*, as in the following definition.

Definition A.5. *Let $\succ$ be a strict preference (asymmetric and negatively transitive) over elements of an arbitrary nonempty set S. From $\succ$ we define two additional binary relations $\succcurlyeq$ and $\sim$ on S, respectively called weak preference and indifference, as follows: For each $x, y \in S$,*

1. $x \succcurlyeq y \Leftrightarrow \neg (y \succ x)$;

2. $x \sim y \Leftrightarrow (\neg (x \succ y)$ *and* $\neg (y \succ x)) \Leftrightarrow (y \succcurlyeq x$ *and* $x \succcurlyeq y)$.

Throughout this thesis, we will often say that $\succcurlyeq$ and $\sim$ are defined from $\succ$ *"in the usual manner"* whenever $\succcurlyeq$ and $\sim$ are defined from $\succ$ according to Definition A.5. Now, given $\succ$ as a primitive of choice, and suppose that we have defined weak preference and indifference from strict preference in the usual manner, what can we say about the properties of the latter two notions? The following Theorem gives an answer.

Theorem A.6. *If $\succ$ is a strict preference (asymmetric and negatively transitive) over elements of an arbitrary nonempty set S then*

1. $\succ$ *is transitive;*

2. $\sim$ *is an equivalence relation;*

3. $\succcurlyeq$ *is transitive and complete;*

4. Exactly one of the following holds, for each $x, y \in S$:

(a) $x \succ y$*; or,*

(b) $y \succ x$*; or,*

(c) $x \sim y$*.*

Proof. See Fishburn [128], Theorem 2.1 on p. 13. □

Cardinal Utility

A fundamental mathematical issue that arises naturally whenever one is dealing with some sort of an ordering of elements of some set S is whether it is possible to represent this ordering by an ordering on $\mathbb{R}$ (in the usual ordering on $\mathbb{R}$) via some mapping of S into $\mathbb{R}$ that preserves the ordering. For instance, if $\succ$ is a strict preference on S, does there exist a mapping $u : S \to \mathbb{R}$ such that for all $x, y \in S$, $x \succ y \Leftrightarrow u(x) > u(y)$, where $>$ is the usual strict order on $\mathbb{R}$? Needless to say, this is a deep and serious mathematical question, and it has indeed occupied such mathematicians as the great Cantor himself (Cantor [69]).

Definition A.7. *Let $\succ$ be a strict preference over elements of an arbitrary nonempty set S, and define $\succcurlyeq$ from $\succ$ in the usual manner. A subset D of S is called $\succ$-order dense if for each $x, y \in S$ such that $x \succ y$, there exists $z \in D$ such that $x \succcurlyeq z \succcurlyeq y$.*

Theorem A.8 (Cantor)**.** *Let $\succ$ be a binary relation on an arbitrary nonempty set S. Then there exists a real-valued function u on S such that:*

$$\forall (x, y) \in X \times X, \;\; x \succ y \Leftrightarrow u(x) > u(y) \tag{A.2}$$

if and only if $\succ$ on S is a strict preference relation (asymmetric and negatively transitive) and there exists a countable $\succ$-order dense subset of S. Furthermore, when such a function u exists, it is unique up to a strictly increasing transformation. We will refer to u as a cardinal utility function, or simply a utility function.

Proof. See Kreps [187], Theorem 3.5 on p. 25 and Theorem 3.6 on p. 26. □

Cantor's theorem guarantees the existence of a *utility* function that *represents* the DM's strict preference $\succ$ over elements S in terms of a strict ordering on $\mathbb{R}$. However, the theorem does not mention continuity of the utility function u, and in fact such a question is meaningless in the setting of the previous theorem, for there is no topology on S. Moreover, even if there were a topology $\mathcal{T}$ on S, Cantor's theorem does not guarantee that u would be $\mathcal{T}$-continuous. This motivates the following theorem:

Theorem A.9 (Debreu-Fishburn)**.** *Let* $(S, \mathcal{T})$ *be a topological space, and let* $\succ$ *be a binary relation on* S*. Suppose that there is a real-valued function* u_1 *on* S *representing* $\succ$ *in the sense of* (A.2)*. Then there is another real-valued function* u_2 *on* S*, satisfying* (A.2) *and continuous in the topology* $\mathcal{T}$*, if and only if either one of the following two conditions holds:*

1. *For each* $y \in S$*,* $\{x \in S : x \succ y\} \in \mathcal{T}$ *and* $\{x \in S : y \succ x\} \in \mathcal{T}$*; or,*

2. *If* $x, y \in S$*, then there are sets* $T_x, T_y \in \mathcal{T}$ *such that:*

   (a) $x \in T_x$*,* $y \in T_y$*;*

   (b) $x \succ y'$ *for all* $y' \in T_y$*; and,*

   (c) $x' \succ y$ *for all* $x' \in T_x$*.*

Proof. See Fishburn [128], Theorem 3.5 on p. 36. Also, see Debreu [92], Corollary on p. 289. □

Theorem A.10 (Debreu-Eilenberg)**.** *Let* $(S, \mathcal{T})$ *be a connected and separable topological space, and let* $\succ$ *be a strict preference (asymmetric and negatively transitive) on* S*. If, for each* $y \in S$*, the sets* $\{x \in S : x \succ y\}$ *and* $\{x \in S : y \succ x\}$ *are open (for* $\mathcal{T}$*), then there is a real-valued function* u *on* S *representing* $\succ$ *in the sense of* (A.2)*, and continuous in the topology* $\mathcal{T}$*.*

Proof. See Debreu [92], Proposition 4 on p. 291, or Fishburn [128], Lemma 5.1 on p. 62. The work of Eilenberg [117] can be seen as the foundation of continuous cardinal utility, and we refer the reader to [117] for many interesting results. □

For more about continuity properties of cardinal utility functions we refer the reader to Debreu [90], [91] and [92], Newman and Read [218], Peleg [224], pr Rader [232], for instance.

A.1.2 The von Neumann-Morgenstern Approach

Setup

In the setting of von Neumann and Morgenstern [297], a decision maker (DM) expresses preference between exogenously given objective probability measures on some measurable space. We will model this situation as follows: Let (S, Σ) be a given measurable space and let $\mathcal{P}$ denote the collection of all probability measures on (S, Σ). We suppose that the DM has a *strict preference* $\succ$ over elements of $\mathcal{P}$. Define *weak preference* $\succcurlyeq$ and *indifference* $\sim$ from $\succ$ in the usual manner, and note that the collection $\mathcal{P}$ is closed under countable convex combinations. For each $x \in S$, define the element δ_x of $\mathcal{P}$ as the degenerate probability distribution that assigns value 1 to the (Σ-measurable) singleton $\{x\}$. By a customary slight abuse of notation, we will write $x \succ y$, for $x, y \in S$ to mean $\delta_x \succ \delta_y$.

We will make the following structural assumptions on Σ:

1. For each $x \in S$, $\{x\} \in \Sigma$; and,

2. For each $y \in S$, $\{x \in S : x \succ y\} \in \Sigma$ and $\{x \in S : y \succ x\} \in \Sigma$.

Axioms on Preferences

Consider the following axioms for the DM's preference $\succ$ over elements of $\mathcal{P}$:

Axiom A.1 (Weak Order). $\succ$ *is asymmetric and negatively transitive.*

Axiom A.2 (Independence Axiom). *for all* $\mu, \nu, \eta \in \mathcal{P}$*, and for all* $\alpha \in (0,1)$*, we have:*

$$\mu \succ \nu \Longleftrightarrow \alpha\mu + (1-\alpha)\,\eta \succ \alpha\nu + (1-\alpha)\,\eta \tag{A.3}$$

Axiom A.3 (Archimedean Axiom). *for all* $\mu, \nu, \eta \in \mathcal{P}$*, we have:*

$$\mu \succ \nu \succ \eta \Rightarrow \exists\ \alpha, \beta \in (0,1)\,, \alpha\mu + (1-\alpha)\,\eta \succ \nu \succ \beta\mu + (1-\beta)\,\eta \tag{A.4}$$

Axiom A.4 (Monotonicity)**.** *For any $\mu \in \mathcal{P}$ and for any $A \in \Sigma$, the following holds:*

1. *If $\mu(A) = 1$, $y \in S$, and $\delta_x \succcurlyeq \delta_y$, for each $x \in A$, then $\mu \succcurlyeq \delta_y$; and,*

2. *If $\mu(A) = 1$, $z \in S$, and $\delta_z \succcurlyeq \delta_x$, for each $x \in A$, then $\delta_z \succcurlyeq \mu$.*

Expected Utility Representation of Preferences

Theorem A.11 (Expected Utility Representation)**.** *If $\succ$ on $\mathcal{P}$ satisfies axioms A.1, A.2, A.3 and A.4, then there exists some bounded function $u : S \to \mathbb{R}$ such that for each $\mu, \nu \in \mathcal{P}$:*

$$\mu \succ \nu \Longleftrightarrow \int u \, d\mu > \int u \, d\nu \tag{A.5}$$

Moreover, such a function u is unique up to a positive linear transformation.

Proof. See Fishburn [128], Theorem 10.3 on p. 139 and Lemma 10.5 on p. 138. □

Theorem A.12 (Expected Utility Representation with Continuous, Bounded Utility)**.** *Let (S, d) be a separable metric space, and denote by Σ the Borel σ-algebra on S. Let $\succ$ be a binary relation on the collection $\mathcal{P}_B$ of all (Borel) probability measures on the measurable space (S, Σ). If $\succ$ is such that:*

1. *$\succ$ satisfies axioms A.1 and A.2; and,*

2. *The projection of $\succ$ onto $\mathcal{P}_B$ is continuous in the weak* topology on $\mathcal{P}_B$, that is, for any $\nu \in \mathcal{P}_B$, the sets $\{\mu \in \mathcal{P}_B : \mu \succcurlyeq \nu\}$ and $\{\mu \in \mathcal{P}_B : \nu \succcurlyeq \mu\}$ are weak* closed*[1]*,*

[1] The collection $\mathcal{P}_B$ of Borel probability measures on a metrizable topological space S can be endowed with the weak* topology $\sigma(\mathcal{P}_B, C_b(S))$, where $C_b(S)$ is the collection of all bounded continuous $\mathbb{R}$-valued functions on S. This topology is characterized by the fact that a net $\{\mu_\alpha\}_{\alpha \in \Gamma}$ of Borel probability measures on S converges in the weak* topology to some Borel probability measure μ on S if and only if the net $\{\int \phi \, d\mu_\alpha\}_{\alpha \in \Gamma}$ converges to $\int \phi \, d\mu$, for each continuous bounded real function ϕ on S. We refer the reader to Chap. 15 of Aliprantis and Border [3] for more about the weak* topology of $\mathcal{P}_B$. All required background material is given in the mathematical appendix, and notably in appendix D.4.

then there exists some continuous and bounded function $u : S \to \mathbb{R}$ such that for all $\mu, \nu \in \mathcal{P}_B$,

$$\mu \succ \nu \Longleftrightarrow \int_S u \; d\mu > \int_S u \; d\nu \tag{A.6}$$

Moreover, such a function u is unique up to a positive linear transformation.

Proof. See Kreps [187], pp. 65-66. For more about the weak* topology on $\mathcal{P}_B$, we refer the reader to Aliprantis and Border [3], Chap. 14 (Riesz Representation Theorem) and Chap. 15 (Duality theory for $\mathcal{P}_B$), Billingsley [38], Theorem 2.1 on p. 16 (Portmanteau Theorem), Dunford and Schwartz [109], Theorem IV.6.2 on p. 262 (Riesz Representation Theorem), or Hewitt and Stromberg [166], Theorem 20.48 on p. 364 (Riesz Representation Theorem). □

Remark A.13. *The results of this section can be generalized to any convex subcollection $\mathcal{P}^*$ of $\mathcal{P}$, provided $\mathcal{P}^*$ satisfies some structural requirements. The interested reader is referred to Fishburn [128] (pp. 137-143).*

A.1.3 The Savage Approach

Setup and Preliminaries

The Savage Setting Let S be an arbitrary nonempty set. We interpret S as the set of all possible *states of the world*, and 2^S as the collection of all possible *events*. Let X be an arbitrary nonempty set that we interpret as the set of all possible *outcomes* (e.g. X might be taken to be $\mathbb{R}$ if we are considering monetary outcomes). We denote by $\mathcal{F}$ the collection of all mappings of S into X, and we refer to elements of $\mathcal{F}$ as *acts*.

A decision maker (DM) expresses preferences over acts, i.e. $\mathcal{F}$ is the choice set. Let $\succ$ be a binary relation on $\mathcal{F}$ that will be taken to be the DM's preference over acts, and define $\succcurlyeq$ and $\sim$ from $\succ$ in the usual manner. Note that, contrary to the von Neumann-Morgenstern setting, the only primitive in this context is the agent's preference over acts. There are no exogenous objective probabilities given *a priori* here. Probability and utility will be derived from $\succ$, and are hence entirely *subjective*. We seek a *Subjective Expected Utility Representation* of $\succ$ of the form:

$$\forall f, g \in \mathcal{F},\; f \succ g \Longleftrightarrow \int_S u\left(f(s)\right)\; d\mu\left(s\right) > \int_S u\left(g(s)\right)\; d\mu(s) \tag{A.7}$$

for some *utility* function $u : X \to \mathbb{R}$ and some *subjective probability measure* μ on (S, Σ).

Definitions

Definition A.14 (Conditional Preference)**.** *If $A \subseteq S$, and $f, g \in \mathcal{F}$, we say that $f \succ g$* given A *if there exist $f', g' \in \mathcal{F}$, with $f' \succ g'$ and:*

1. *$f = f'$ and $g = g'$ on A;*

2. *$f' = g'$ on $S \setminus A$.*

Definition A.15 (Null Event)**.** *$A \subseteq S$ is called* null *if for all acts $f, g \in \mathcal{F}$, we have $f \sim g$ given A.*

Definition A.16 (Constant Act)**.** *For each $x \in X$, we define the constant act $\overline{x}$ as that element of $\mathcal{F}$ that yields the outcome x in each states of the world.*

Hence, we can embed X into $\mathcal{F}$ by identifying $\overline{x}$ with x, for any $x \in X$.

Axioms on Preferences

Consider the following seven axioms on the binary relation $\succ \subseteq \mathcal{F} \times \mathcal{F}$:

Axiom P.1 (Weak Order)**.** *$\succ$ is a strict preference (asymmetric and negatively transitive) on $\mathcal{F}$.*

Axiom P.1 states that $\succ$ is a weak order.

Axiom P.2 (Sure-Thing Principle)**.** *If $A \subseteq S$, $f, g, f', g' \in \mathcal{F}$, $f = f'$ on A, $g = g'$ on A, $f = g$ on $S \setminus A$, and $f' = g'$ on $S \setminus A$, then $f \succ g \Leftrightarrow f' \succ g'$.*

Axiom P.2 says that the preference between two acts depends only on the states where those acts have different consequences.

Axiom P.3 (Eventwise Monotonicity)**.** *If A is not null, $x, y \in X$, $f = \overline{x}$ on A, and $g = \overline{y}$ on A, then $f \succ g$ given $A \Leftrightarrow \overline{x} \succ \overline{y}$.*

Axiom $P.3$ says that if you consider an act that yields the outcome x for each state of the world in an event A and you modify it only on A so as to get an another act guarantying the outcome y on A, then the preference between these two acts should depend only on the preference between the constant acts $\overline{x}$ and $\overline{y}$.

Axiom P.4 (Weak Comparative Probability)**.** *Suppose $f, g, f', g' \in \mathcal{F}$, $x, y, x', y' \in X$, $A \subseteq S$, $B \subseteq S$, and:*

1. *$\overline{x} \succ \overline{y}$, and $\overline{x'} \succ \overline{y'}$;*
2. *$f = \overline{x}$ on A and $f' = \overline{x'}$ on A;*
3. *$f = \overline{y}$ on $S \setminus A$ and $f' = \overline{y'}$ on $S \setminus A$;*
4. *$g = \overline{x}$ on B and $g' = \overline{x'}$ on B;*
5. *$g = \overline{y}$ on $S \setminus B$ and $g' = \overline{y'}$ on $S \setminus B$;*

Then $f \succ g \Leftrightarrow f' \succ g'$

Axiom $P.4$ is the counterpart of axiom $P.3$ for ranking events, and it will be an essential axiom for inferring likelihood judgments on elements of 2^S from preference ranking of elements of $\mathcal{F}$. Axiom $P.4$ essentially says that likelihood of events are not affected by acts.

Axiom P.5 (Nondegeneracy))**.** *There exist some $x, y \in X$ such that $\overline{x} \succ \overline{y}$.*

Axiom P.6 (Small Event Continuity)**.** *For each $x \in X$ and for each $f, g \in F$ such that $f \succ g$, there exists a finite partition of S such that for every event A in that partition:*

1. *If $f' = \overline{x}$ on A and $f' = f$ on $S \setminus A$, then $f' \succ g$;*
2. *If $g' = \overline{x}$ on A and $g' = g$ on $S \setminus A$, then $f \succ g'$.*

Axiom $P.6$ is a continuity axiom that has a clear Archimedean flavor, and it is essential for the derivation of a subjective probability measure on $\left(S, 2^S\right)$.

Axiom P.7 (Uniform Monotonicity). *For all $A \subseteq S$,*

1. *If $f \succ g(s)$ given A, for all $s \in A$, then $f \succcurlyeq g$ given A;*

2. *If $g(s) \succ f$ given A, for all $s \in A$, then $g \succcurlyeq f$ given A.*

Axiom P.8 (Monotone Continuity). *Let $f, g \in \mathcal{F}$ and $x \in X$. Let $\{A_n\}_{n\geq 1}$ be a sequence of events such that $A_{n+1} \subseteq A_n$ and $\bigcap_{n\geq 1} A_n = \varnothing$. For each $n \geq 1$, define the acts f'_n and g'_n as follows:*

1. *$f'_n = x$ on A_n and $f'_n = f$ on $S \setminus A_n$; and,*

2. *$g'_n = x$ on A_n and $g'_n = g$ on $S \setminus A_n$.*

If $f \succ g$ then there exists some $N \geq 1$ such that:

1. *$f'_N \succ g$; and,*

2. *$f \succ g'_n$.*

Likelihood Judgements from Preferences

Definition A.17. *We define a binary relation $\dot{\succ}$ on 2^S as follows: let A and B be any two events. Let $x, y \in X$ be such that $\overline{x} \succ \overline{y}$, let $f \in \mathcal{F}$ be such that $f = \overline{x}$ on A and $f = \overline{y}$ on $S \setminus A$, and let $g \in \mathcal{F}$ be such that $g = \overline{x}$ on B and $g = \overline{y}$ on $S \setminus B$. We say that the DM judges A to be more likely than B, written $A \dot{\succ} B$, when $f \succ g$.*

Definition A.17 gives us a way to determine the DM's likelihood judgements over events from his preference ranking of acts in $\mathcal{F}$. The derivation of subjective probability from preferences should be consistent with the likelihood ranking of events as defined above. Specifically, we seek a unique probability measure μ on $\left(S, 2^S\right)$ such that for any events A and B,

$$A \dot{\succ} B \Longleftrightarrow \mu(A) > \mu(B) \tag{A.8}$$

Theorem A.18. *If $\succ$ on $\mathcal{F}$ satisfies axioms P.1 to P.7 above, then there exists a unique nonatomic*[2] *and finitely additive probability measure μ on $\left(S, 2^S\right)$ such that:*

1. *For all events A and B, we have $A \mathrel{\dot{\succ}} B \Leftrightarrow \mu(A) > \mu(B)$; and,*

2. *For any event A and for any $r \in [0, 1]$, there exists some $B \subseteq A$ and $\mu(B) = r\mu(A)$.*

Proof. See Kreps [187], Proposition 9.2 on p. 133, Proposition 9.3 on p. 133, and Theorem 8.10 on p. 125. □

Corollary A.19. *If $\succ$ on $\mathcal{F}$ satisfies axioms P.1 to P.8 above, then there exists a unique nonatomic and countably additive probability measure μ on $\left(S, 2^S\right)$ such that:*

1. *For all events A and B, we have $A \mathrel{\dot{\succ}} B \Leftrightarrow \mu(A) > \mu(B)$; and,*

2. *For any event A and for any $r \in [0, 1]$, there exists some $B \subseteq A$ and $\mu(B) = r\mu(A)$.*

Proof. See Theorem A.18, Arrow [25] p. 48 and p. 76, Chateauneuf et al. [77] p. 974, and Villegas [296] Theorem 2 on p. 1794. □

Subjective Expected Utility Representation

Theorem A.20 (Subjective Expected Utility Representation)**.** *If $\succ$ on $\mathcal{F}$ satisfies axioms P1 to P7, and if the binary relation $\dot{\succ}$ on S is defined as in Definition A.17, then:*

1. *There exists a unique finitely additive nonatomic probability measure μ on $\left(S, 2^S\right)$ such that*

   (a) *For all events A and B,*

   $$A \mathrel{\dot{\succ}} B \Longleftrightarrow \mu(A) > \mu(B) \tag{A.9}$$

   (b) *For any event A, and for any $r \in [0, 1]$, there exists an event $B \subseteq A$ such that*

   $$\mu(B) = r\ \mu(A) \tag{A.10}$$

[2]That is, for any $A \subseteq S$ with $\mu(A) > 0$, there is some $B \subset A$ and $0 < \mu(B) < \mu(A)$.

2. *For μ given above, there exists a bounded utility function $u : X \to \mathbb{R}$ such that for all $f, g \in \mathcal{F}$,*

$$f \succ g \Longleftrightarrow \int_S u\left(f\left(s\right)\right)\ d\mu\left(s\right) > \int_S u\left(g\left(s\right)\right)\ d\mu\left(s\right) \tag{A.11}$$

Moreover, u is unique up to a positive linear transformation. If $\succ$ also satisfies axiom P.8, then μ is countably additive.

Proof. See Fishburn [128] (Theorem 14.1 on p. 192), Gilboa [149] (Theorem 10.2 on p. 109), Kreps [187] (Theorem 9.16 on p. 136), and Corollary A.19 above. □

Remark A.21. *The previous theorem can be generalized to a measurable space (S, Σ) instead of $\left(S, 2^S\right)$. Events will then have to be taken to be only those Σ-measurable subsets of S, and acts will be taken to be only those Σ-measurable elements of $\mathcal{F}$.*

A.1.4 The Anscombe-Aumann Approach

Setup and Preliminaries

The Anscombe-Aumann Setting Let S be an arbitrary nonempty set. We interpret S as the set of all possible *states of the world*, and 2^S as the collection of all possible *events*. Let X be an arbitrary nonempty set that we interpret as the set of all possible *outcomes* (e.g. X might be taken to be $\mathbb{R}$ if we are considering monetary outcomes), and endow X with a σ-algebra Σ of events. Assume that $\{x\} \in \Sigma$, for each $x \in X$. We denote by $\Delta\left(X\right)$ the collection of all simple countably additive probability measures on (X, Σ)[3]. When endowed with the usual mixing operations[4] for set functions, $\Delta\left(X\right)$ is a convex subset of the real vector space of all countably additive finite measures on (X, Σ).

We denote by $\mathcal{H}$ the collection of all mappings of S into $\Delta\left(X\right)$, and we refer to elements of $\mathcal{H}$ as *Anscombe-Aumann (AA) acts*. As in the previous chapter, we also denote by $\mathcal{F}$ the collection of all mapping of S into X, and we refer to elements of $\mathcal{F}$ as *Savage acts*.

[3] A countably additive probability measure P on Σ is called simple if there is a finite Σ-measurable set A such that $P\left(A\right) = 1$. In other words, P is simple if there is a finite collection $\{x_1, \ldots, x_n\}$ of distinct points in X such that $\{x_i\} \in \Sigma$, for each $i \in \{1, 2, \ldots, n\}$, and $\sum_{i=1}^{n} P\left(\{x_i\}\right) = 1$.

[4] The usual mixing operations for set functions are the set-wise mixing operations. That is, for any set functions μ and ν, define the set function $\mu + \nu$ by $\left(\mu + \nu\right)\left(A\right) = \mu\left(A\right) + \nu\left(A\right)$, for any set A. Moreover, for any set function μ and for any scalar α, define the set function $\alpha.\mu$ by $\left(\alpha.\mu\right)\left(A\right) = \alpha.\mu\left(A\right)$, for any set A.

For each $x \in X$ let $\delta_x \in \Delta(X)$ denote the degenerate simple probability distribution assigning value 1 to the event $\{x\}$. We can then embed $\mathcal{F}$ into $\mathcal{H}$ by identifying x with δ_x, for each $x \in X$.

For each $s \in S$, let $X(s) := \{f(s) : f \in \mathcal{F}\}$ be the set of all possible consequences in the state of the world s. Then $X = \bigcup_{s \in S} X(s)$. In the following, we will assume (as Fishburn [128] does in Section 13.3) that $X(s) = X$, for each $s \in S$.

Preferences A decision maker (DM) expresses preferences over elements of $\mathcal{H}$, via a binary relation $\succ \subseteq \mathcal{H} \times \mathcal{H}$. We define the binary relations $\succcurlyeq$ and $\sim$ from $\succ$ in the usual manner.

The DM's preference $\succ \subseteq \mathcal{H} \times \mathcal{H}$ then implies a preference over Savage acts as the restriction of $\succ$ to $\mathcal{F}$, when the latter is seen as a subset of $\mathcal{H}$. We will sometimes make the usual abuse of notation and write:

1. $\mu \succ \nu$, for $\mu, \nu \in \Delta(X)$, to mean that $f \succ g$, where $f, g \in \mathcal{H}$ are such that $f(s) = \mu$ and $g(s) = \nu$, for each $s \in S$;

2. $x \succ y$, for $x, y \in X$, to mean that $\delta_x \succ \delta_y$;

3. $h \succ \mu$, for $h \in \mathcal{H}$ and $\mu \in \Delta(X)$, to mean that $h \succ f$, where $f \in \mathcal{H}$ is such that $h(s) = \mu$, for each $s \in S$; or,

4. $h \succ x$, for $h \in \mathcal{H}$ and $x \in X$, to mean that $h \succ \delta_x$.

Definitions

Definition A.22. *An AA act $h \in \mathcal{H}$ is said to be constant on an event $A \subseteq S$ when $h(s) = h(t)$, for each $s, t \in A$.*

Definition A.23. *Let $h \in \mathcal{H}$, $A \subseteq S$, and $\mu \in \Delta(X)$. We will say that $h = \mu$ on A when $h(s) = \mu$, for each $s \in A$.*

Definition A.24. *Let $f, h \in \mathcal{H}$ and $A \subseteq S$. We will say that $f = h$ on A when $f(s) = h(s)$, for each $s \in A$.*

Definition A.25. *If $A \subseteq S$, and $f, g \in \mathcal{H}$, we say that $f \succ g$ given A if there exist $f', g' \in \mathcal{H}$, with $f' \succ g'$ and:*

1. *$f = f'$ and $g = g'$ on A;*

2. *$f' = g'$ on $S \setminus A$.*

Definition A.26. *$A \subseteq S$ is called* null *if for all AA acts $f, g \in \mathcal{H}$, we have $f \sim g$ given A.*

Axioms on Preferences

Consider the following axioms on the DM's preference $\succ$ over elements of $\mathcal{H}$:

Axiom AA.1 (Weak Order)**.** *$\succ$ is asymmetric and negatively transitive.*

Axiom AA.2 (Independence Axiom)**.** *for all $f, g, h \in \mathcal{H}$, and for all $\alpha \in (0,1)$, we have:*

$$f \succ g \Longleftrightarrow \alpha f + (1-\alpha) h \succ \alpha g + (1-\alpha) h \tag{A.12}$$

Axiom AA.3 (Archimedean Axiom)**.** *for all $f, g, h \in \mathcal{H}$, we have:*

$$f \succ g \succ h \Rightarrow \exists\ \alpha, \beta \in (0,1), \alpha f + (1-\alpha) h \succ g \succ \beta f + (1-\beta) h \tag{A.13}$$

Axiom AA.4 (Nondegeneracy)**.** *There exist some $\mu, \nu \in \Delta(X)$ such that $\mu \succ \nu$.*

Axiom AA.5 (State-Independence)**.** *Let $A \subseteq S$ be a non-null event. Let $h, g \in \mathcal{H}$, and let $\mu, \nu \in \Delta(X)$. Suppose that $h = \mu$ on A, $g = \nu$ on A, and $h = g$ on $S \setminus A$. Then:*

$$h \succ g \Leftrightarrow \mu \succ \nu \tag{A.14}$$

Axiom AA.6 (Monotonicity). *For each $f, g, h \in \mathcal{H}$:*

1. $(f(s) \succ h,\ \forall s \in S) \Rightarrow f \not\prec h$; *and,*

2. $(h \succ g(s),\ \forall s \in S) \Rightarrow h \not\prec g$.

Anscombe-Aumann Representation Theorem

Theorem A.27 (Anscombe-Aumann Expected Utility Representation). *If $\succ$ on $\mathcal{H}$ satisfies axioms AA.1 to AA.6 above then there exists a utility function $u : X \to \mathbb{R}$ and a unique probability measure μ on $\left(S, 2^S\right)$ such that for all $h, g \in \mathcal{H}$,*

$$h \succ g \Longleftrightarrow \int_S \left[\int_X u\ dh(s)\right]\ d\mu(s) > \int_S \left[\int_X u\ dg(s)\right]\ d\mu(s) \tag{A.15}$$

Moreover, u is unique up to a positive linear transformation, and for all $A \subseteq S$, $\mu(A) = 0$ if and only if A is a null event.

Proof. See Fishburn [128] (Theorem 13.3 on p. 179) and Gilboa [149] (Theorem 14.1 on p. 144). □

Remark A.28. *The previous theorem can be generalized to a measurable space $(S, \mathcal{B})$ instead of $\left(S, 2^S\right)$. Events will then have to be taken to be only those $\mathcal{B}$-measurable subsets of S, and acts will be taken to be only those $\mathcal{B}$-measurable elements of $\mathcal{F}$.*

Moreover, the AA setting can be generalized so that acts are mappings of S into the collection $ca_1^+(\Sigma)$ of all countably additive probability measures on the measurable space (X, Σ) of consequences. This generalization is based on the general von Neumann-Morgenstern representation theorem (Theorem A.11) and hence requires, among other things, an additional axiom on $\succ$, in the spirit of axiom A.4 of the von Neumann-Morgenstern setting.

A.2 The Paradoxes of Expected Utility Theory

A.2.1 Introduction

The three versions of *Expected-Utility Theory* (EUT) discussed thus far (namely, the von Neumann-Morgenstern version, the Savage version, and the Anscombe-Aumann version) provide a *normative* approach to decision making. In other words, if a decision maker's (DM) preference is assumed to satisfy a set of axioms then EUT essentially asserts that the DM's behavior can be predicted according to a certain representation theorem. In practice, however, there has been considerable documentation of the inadequacy of EUT from a purely *descriptive* standpoint. We refer the reader to Machina [197] for a historical overview.

In this chapter we briefly discuss two major challenges to EUT that experimental studies have put forward: The Allais pradox ([4] and [5]) and the Ellsberg paradox ([118]). The former can be seen as a descriptive argument against *linearity in probabilities*, whereas the latter can be seen as a descriptive argument against the Bayesian paradigm in Subjective Expected Utility (SEU), whereby the DM's beliefs are represented by a *unique additive prior*, that is, a *unique subjective probability measure*.

A.2.2 The Allais Paradox

Independence of Irrelevant Alternatives

Recall that in the setting of von Neumann and Morgenstern of decision under risk the DM's preference $\succ$ over elements of $\mathcal{P}$ (the collection of all probability measures on a given measurable space) was assumed to satisfy the following *independence axiom*:

Axiom A.2 (Independence Axiom)**.** *for all $\mu, \nu, \eta \in \mathcal{P}$, and for all $\alpha \in (0,1)$, we have:*

$$\mu \succ \nu \Rightarrow \alpha\mu + (1-\alpha)\,\eta \succ \alpha\nu + (1-\alpha)\,\eta \tag{A.16}$$

The *independence axiom* states that when comparing two alternative elements of choice, the DM should only focus on the differences between these elements. In Savage's setting, where a DM has preference $\succ$ over elements of $\mathcal{F}$ (the collection of all mappings of the states space S into a space X of consequences), a similar axiom says that the preference between two acts depends only on the states where those acts have different consequences. This is Savage's *sure-thing principle* that we recall below:

Axiom P.2 (Sure-Thing Principle)**.** *If* $A \subseteq S$*,* $f, g, f', g' \in \mathcal{F}$*,* $f = f'$ *on* A*,* $g = g'$ *on* A*,* $f = g$ *on* $S \setminus A$*, and* $f' = g'$ *on* $S \setminus A$*, then*

$$f \succ g \Leftrightarrow f' \succ g' \tag{A.17}$$

The main implication of the *independence axiom* on the form of the functional that represents the DM's preference (namely, the Expected-Utility functional) is *linearity in the probabilities* (see Machina [197], p. 127).

Allais' Experiment

Allais [5]'s famous experimental study showed that this *independence axiom* is typically violated in practice. Here we recall Allais' experimental setting.

Consider the following four gambles:

1. Gamble A:

Win \$100	with probability 1.

2. Gamble B:

Win \$500	with probability 0.10
Win \$100	with probability 0.89
Win \$0	with probability 0.01

3. Gamble C:

Win \$100	with probability 0.11
Win \$0	with probability 0.89

4. Gamble D:

Win \$500	with probability 0.10
Win \$0	with probability 0.90

Lemma A.29. *If the DM's preference* $\succ$ *over elements of* $\mathcal{P}$ *admits an Expected-Utility representation in the spirit of Theorem A.11, then:*

$$(\textit{Gamble } A \succ \textit{Gamble } B) \Rightarrow (\textit{Gamble } C \succ \textit{Gamble } D) \tag{A.18}$$

Proof. Suppose that the DM has preference $\succ$ over elements of $\mathcal{P}$ that admits an Expected-Utility representation as in Theorem A.11. Then there exists some real-valued (utility) function u on the space S such that for any $\mu, \nu \in \mathcal{P}$, $\mu \succ \nu \Leftrightarrow \int_S u \; d\mu > \int_S u \; d\nu$.

Now suppose that Gamble $A \succ$ Gamble B and assume, without loss of generality, that $u(0) = 0$. It then follows that $u(100) > 0.1u(500) + 0.89u(100)$. Thus, $u(500) < \left(\frac{1-0.89}{01}\right) u(100) = 1.1u(100)$, that is, $0.1u(500) < 0.11u(100)$. This means that Gamble $C \succ$ Gamble D. □

However, as noted by Allais [5] (and Slovic and Tversky [282]), most individuals typically strictly prefer Gamble A to Gamble B and Gamble D to Gamble C, hence violating Lemma A.29, and therefore violating the *independence axiom* (that is, the *linearity in probabilities*).

A.2.3 The Ellsberg Paradox

Ellsberg's One-Urn Problem

The Ellsberg Paradox (Ellsberg [118]) is descriptively important for it demonstrates a violation of Savage's axioms underlying the existence of a unique subjective probability measure. It is also historically important in a normative sense, for to a large extent, the Ellsberg Paradox can be seen as the starting point of theoretical studies of the concept of *ambiguity* (a.k.a. *Knightian uncertanity*).

Ellsberg's ([118]) problem is as follows: Consider an urn containing 30 red balls and 60 other balls that are either black or yellow. All balls are very well mixed. Consider the following four gambles:

1. Gamble A:

   Win \$100 — if you draw a red ball
   Win \$0 — otherwise

2. Gamble B:

   Win \$100 — if you draw a black ball
   Win \$0 — otherwise

3. Gamble C:

   Win \$100 — if you draw a red or yellow ball
   Win \$0 — otherwise

4. Gamble D:

Win \$100	if you draw a black or yellow ball
Win \$0	otherwise

Lemma A.30. *Suppose that a DM has preference $\succ$ over Savage acts (i.e. elements of $\mathcal{F}$), and that his preference admits a subjective expected-utility representation in the spirit of Theorem A.20. Then the following necessarily holds:*

$$(Gamble\ A \succ Gamble\ B) \Rightarrow (Gamble\ C \succ Gamble\ D) \tag{A.19}$$

Proof. According to Subjective Expected-Utility Theory (SEU), the DM will choose gamble A over gamble B if and only if he believes that the probability of drawing a red ball is higher than that of drawing black ball. Similarly, according to SEU, he will prefer gamble C to gamble D if and only if he believes that drawing a red or yellow ball is more likely than drawing a black or yellow ball. Hence, if you prefer gamble A to gamble B, you will also prefer gamble C to gamble D, assuming that your beliefs are represented by a unique subjective probability measure (in particular, an *additive* set function). □

In Ellsberg's problem, Individuals are asked to rank their preferences between gambles A and B on the one hand, and gambles C and D on the other hand. Ellsberg predicted (and his prediction was supported by empirical evidence, as in Slovic and Tversky [282]) that most individuals tend to strictly prefer gamble A to gamble B and gamble D to gamble C, violating the prediction of SEU. In essence, people prefer *known uncertainties* to *unknown uncertainties*: the probability of winning \$100 in gamble A is exactly $1/3$, whereas the probability of winning \$100 in gamble B is unknown. Similarly, the probability of winning \$100 in gamble D is exactly $2/3$, whereas the probability of winning \$100 in gamble C is unknown.

Aversion to Knightian Uncertainty (Ambiguity)

Recall that we referred to a situation of decision under objectively known uncertainty as a situation of decision under *risk*, an example of which is the von Neumann-Morgenstern setting. We also referred to a situation of decision under subjective uncertainty as a situation of decision under *uncertainty*, an example of which is the setting of Savage. What the Ellsberg Paradox demonstrates is that, contrary to the Bayesian paradigm (e.g. Savage's view), there ought to exist a meaningful distinction between subjective uncertainties that are *perfectly known* and subjective uncertainties that are *imperfectly known*. This distinction was argued for by Knight [184]. Consequently, we use the term *Knightian Uncertainty* to refer to a situation where the subjective uncertainties are not perfectly known, in the sense that the DM is not able to formulate a unique subjective probability measure that will represent his beliefs. Such a situation is also called a situation of *Ambiguity*, or of *Ambiguous Beliefs*, and it was the main motivation behind the work of Epstein et al. ([121], [122], [123]), Gilboa ([148], [151]), Marinacci et al. ([77], [134], [139], [182], [195], [201], [202], [203], [205]), and Schmeidler ([274], [275]).

A.3 Alternative Theories of Choice

The challenges to Expected-Utility Theory (EUT) created the need for alternative theories of choice that would give a normative foundation for decision making which would be consistent with the paradoxes of EUT. Some of these models include models that deal with *Knightian uncertainty*, such as the ***Choquet Expected-Utility Model (CEU)*** (Gilboa [148], Schmeidler [275]), the ***Multiple Priors Model, or Maxmin Expected-Utility model (MEU)*** (Amarante [13], Amarante and Filiz [15], Amarante and Maccheroni [17], Chateauneuf, Maccheroni, Marinacci, and Tallon [77], Epstein and Schneider [122], Ghirardato and Siniscalchi [140], Gilboa, Maccheroni, Marinacci, and Schmeidler [150], Gilboa and Schmeidler [151], Maccheroni [193], and Marinacci [201] and [203]), ***Bewley's Model*** (Bewley [36] and [37], and Gilboa, Maccheroni, Marinacci, and Schmeidler [150]), the ***Smooth Ambiguity Model*** (Klibanoff, Marinacci, and Mukerji [182] and [183]), and the ***Variational Preference Model*** (Maccheroni, Marinacci, and Rustichini [195] and [196]).

A unifying model of decision under ambiguity, called the ***Invariant Biseparable Preference Model (IBP)***, encompasses many of the most popular ambiguity models. It was axiomatized and studied by Ghirardato, Maccheroni, and Marinacci ([134] and [135])[5], and it includes as special cases the CEU and the MEU, among other models. Recently, Amarante ([12] and [14])[6] has extended Ghirardato-Maccheroni-Marinacci's IBP model, hence providing a very general, and quite elegant, model of decision under ambiguity[7].

[5]See also Ghirardato, Maccheroni, Marinacci, and Siniscalchi [136], and Ghirardato and Marinacci [138] and [139].

[6]See also Amarante [8], [9], [10], and [11].

[7]Scholarly work aimed at defining the notions of ***ambiguity*** and ***ambiguity aversion*** and examining some properties of these definitions include Amarante [7] and [13], Amarante and Filiz [15], Epstein [121], Epstein and Zhang [123], Ghirardato [133], Ghirardato, Maccheroni, and Marinacci [134] and [135], Ghirardato and Marinacci [137] and [139], Nehring [216] and [217], and Zhang [306].

Other models, such as *Prospect Theory* ([177]) and *Cumulative Prospect Theory* ([293]), were primarily developed to tackle empirical findings that are mainly of a psychological nature, such as *loss aversion* and *framing effects*. In this chapter we will only discuss *(Cumulative) Prospect Theory* (CPT), the *Choquet Expected-Utility* Model (CEU) (Gilboa [148], Schmeidler [275]), and the *Multiple Priors Model*, or *Maxmin Expected-Utility Model* (MMEU) (Gilboa and Schmeidler [151]) as alternatives to EUT and Subjective Expected-Utility Theory (SEU).

A.3.1 Cumulative Prospect Theory (CPT)

Loss Aversion

The Markowitz-Kahneman-Tversky Heritage An important empirical finding that partly motivated the development of *Prospect Theory* is called *loss aversion* and refers to the fact that people typically experience a gain of a certain amount as less psychologically severe than a loss of the same amount. Markowitz [206] writes (p. 154):

> *"Generally people avoid symmetric bets. This suggests that the curve falls faster to the left of the origin than it rises to the right of the origin. (I.e., $U(X) > |U(-X)|, X > 0$)"*

Even though the term *loss aversion* was not explicitly used by Markowitz [206], the idea behind the predominant view that loss aversion is a property of the utility that manifests itself in the fact that the utility of a given gain is lower than the absolute value of the utility of a loss of the same magnitude was noted by Markowitz [206], in 1952; so was the idea that *people dislike symmetric bets*, which was the definition of loss aversion given by Kahneman and Tversky [177], a mere 27 years later. Kahneman and Tversky [177] write (p. 279):

> *"A salient characteristic of attitudes to changes in welfare is that losses loom larger than gains. The aggravation that one experiences in losing a sum of money appears to be greater than the pleasure associated with gaining the same amount [...] Indeed, most people find symmetric [50:50] bets [...] distinctively unattractive. Moreover, the aversiveness of symmetric fair bets generally increases with the size of the stake"*

Experimental Setting The following example is adapted from Gilboa [149] (p. 155). Consider the following four gambles:

1. Gamble A:

Win \$500	with probability 1

2. Gamble B:

Win \$1000	with probability 0.50
Win \$0	with probability 0.50

3. Gamble C:

Lose \$500	with probability 1

4. Gamble D:

Lose \$1000	with probability 0.50
Lose \$0	with probability 0.50

When asked to rank their preference between Gamble A and Gamble B on the one hand, and Gamble C and Gamble D on the other hand, people typically prefer A to B (that is, prefer a sure gain to a risky gain – hence exhibiting *risk aversion over gains*) and D to C (that is, prefer a risky loss to a sure loss – hence exhibiting *risk seeking over losses*). This phenomenon was coined *the reflection effect* by Kahneman and Tversky [177] and is a direct consequence of what Tversky and Kahneman [293] later called *loss aversion*, that is, the fact the *losses loom larger than gains*.

A More Robust Definition of Loss Aversion (Ghossoub [141])

Few theoretical investigations of loss aversion in CPT have dealt with the probability weighting process as an inherent constituent of loss aversion, and the ones that were carried out were done in a context where the objects of choice are lotteries, that is, discrete probability distributions (see, e.g. Schmidt and Zank [276] and Zank [305]). Accordingly, the definitions proposed are very specific to this particular case. However, in most applications of CPT to finance and insurance, for instance, one deals with an underlying (financial or actuarial) risk which has a continuous distribution on the real line or on an interval thereof (see, e.g. Barberis and Huang [29], Bernard and Ghossoub [32], He and Zhou [164], or Jin and Zhou [175]). In such circumstances, a proper definition of loss aversion simply does not exist as yet, to the best of my knowledge. For instance, the original definition of loss aversion given by Kahneman and Tversky [177] is aversion to *symmetric 50:50 bets*; but what exactly is a *symmetric 50:50 bet* in the continuous case?

One should note that Blavatskyy [41] recently explored the notion of loss aversion outside of CPT, and in a general framework where outcomes are not necessarily monetary, but with a finite state space and where the elements of choice are lotteries, i.e. simple probability distributions. Blavatskyy's [41] definition of loss aversion is behavioral, based on the properties of a preference over the set of all lotteries. However, his definition is essentially comparative, and an "absolute" notion of loss aversion is defined as "more loss aversion than a loss-neutral" preference. The major complication, as the author remarks, is that it is not immediately clear how to define loss-neutrality in that context.

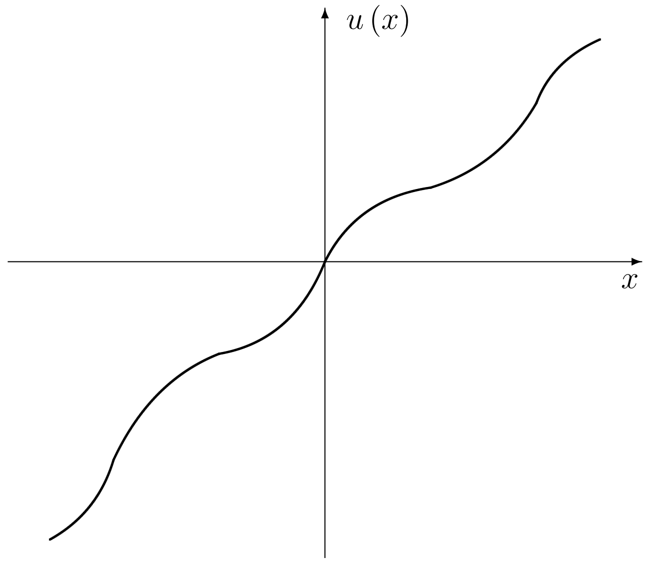
It was the purpose of Ghossoub [141] to provide a proper definition of loss aversion in the continuous case, and to examine its relation to the existing definitions. In Ghossoub [141], we give a general behavioral definition of loss aversion for preferences that exhibit probabilistic sophistication in the sense of Machina and Schmeidler [199] and *gain-loss separability* in a specific sense. The class of such preferences is large enough, and includes, *inter alia*, preferences that admit a *Subjective Expected-Utility* (SEU) representation and preferences that have a CPT representation. We show that, under CPT, our proposed definition generalizes many existing ones and that our analysis extends that of Schmidt and Zank [276] and Zank [305].

We also define an index of loss aversion for preferences that are probabilistically sophisticated, gain-loss separable, and adequately continuous. We show that CPT preferences verify these requirements, and that our proposed index of loss aversion under CPT generalizes that of Köbberling and Wakker [185]: the latter is a special case of the index of loss aversion that we introduce, when the probability weighting functions are identical over losses and gains, that is, when beliefs are not affected by the sign of the outcome.

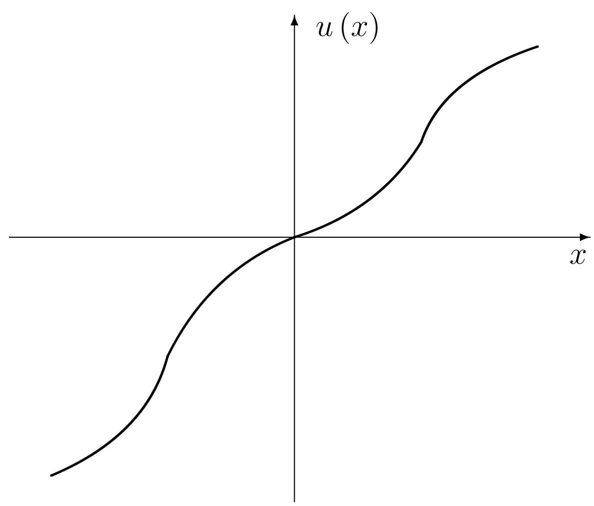
Finally, we examine loss aversion in SEU, and we show that in SEU loss aversion is not equivalent to the utility function having an S shape. We show that loss aversion in SEU holds for a class of utility functions that includes S-shaped functions, but which is strictly larger than the collection of these functions. This class also includes utility functions that are of the Friedman-Savage type ([131]) over both gains and losses (as illustrated in Figure A.1a below), and utility functions such as the one postulated by Markowitz [206] (and illustrated in Figure A.1b below), for instance. In sum, loss aversion might exist even for utility functions that are not S-shaped.

Probability Weighting

The idea that individuals typically distort probabilities during their decision making process is not exclusively a property of prospect theory. It was noted by Dale [87], Edwards ([111],[112], [113], [114], [115]), Fellner [125], Griffith [159], Handa [162], Karmarkar [178],



(a) An example of a utility function where the positive part is of the Friedman-Savage type.



(b) An example of a utility function of the type postulated by Markowitz [206].

Figure A.1: Two examples of a utility function for which loss aversion holds in SEU.

Mosteller and Nogee [215], Preston and Baratta [227], Slovic and Lichtenstein [281], and Sprowls [284], for instance, and all of these results appeared before Kahneman and Tversky [177].

Typically, a *probability distortion function*, or a *probability weighting function*, is a function $T : [0,1] \to [0,1]$ such that $T(0) = 0$ and $T(1) = 1$, and a *probability weighting* *model* of choice is any model of choice where probabilities are distorted. Two examples of probability weighting models are the *Original Prospect Theory* (OPT) and the *Cumulative Prospect Theory* (CPT).

Original Prospect Theory

In the literature, the term *Original Prospect Theory* (OPT) refers to the theory developed in Kahneman and Tversky [177], whereas *Cumulative Prospect Theory* (CPT) refers to the refinement of OPT as presented in Tversky and Kahneman [293]. Here, we will discuss OPT, and the next subsection deal with CPT.

Suppose that a DM has preference $\succ$ over *lotteries*, that is, discrete probability distributions with finitely many outcomes (also called simple probability distributions). We can represent a lottery L as follows: Suppose that the lottery takes on n distinct values $x_1, x_2, \ldots, x_n$, for some $n \geq 1$, with probabilities $p_1, p_2, \ldots, p_n$, respectively, so that $\sum_{i=1}^{n} p_i = 1$. Without loss of generality, we may assume that outcomes are ranked as follows $x_1 \geq x_2 \geq \ldots \geq x_k \geq 0 > x_{k+1} \geq \ldots \geq x_{n-1} \geq x_n$, for some $1 \leq k \leq n$. We then write

$$L = (x_1, p_1; x_2, p_2; \ldots; x_n, p_n) \tag{A.20}$$

The positive outcomes $(x_1, \ldots, x_k)$ are called *gains* and the negative outcomes $(x_{k+1}, \ldots, x_n)$ are called *losses*.

We say that the DM's preference has an OPT representation when there is an increasing utility function $u : \mathbb{R} \to \mathbb{R}$ and a distortion function $T : [0, 1] \to [0, 1]$ such that for any two lotteries L_1 and L_2,

$$L_1 \succ L_2 \Leftrightarrow V_{opt}(L_1) > V_{opt}(L_2) \tag{A.21}$$

where for any lottery $L = (x_1, p_1; x_2, p_2; \ldots; x_n, p_n)$,

$$V_{opt}(L) := \sum_{i=1}^{n} u(x_i)\, T(p_i) \tag{A.22}$$

In Kahneman and Tversky [177] experimental findings suggested that the utility function (value function) u has an S shape, i.e. it is convex over losses and concave over gains, and is steeper for losses than for gains, to account for the fact that *losses loom larger than gains*. A typical value function in OPT is illustrated below.

Kahneman and Tversky's [177] experimental findings also suggested that people typically overweight low probabilities and underweight high ones. A typical probability weighting in OPT is illustrated below.

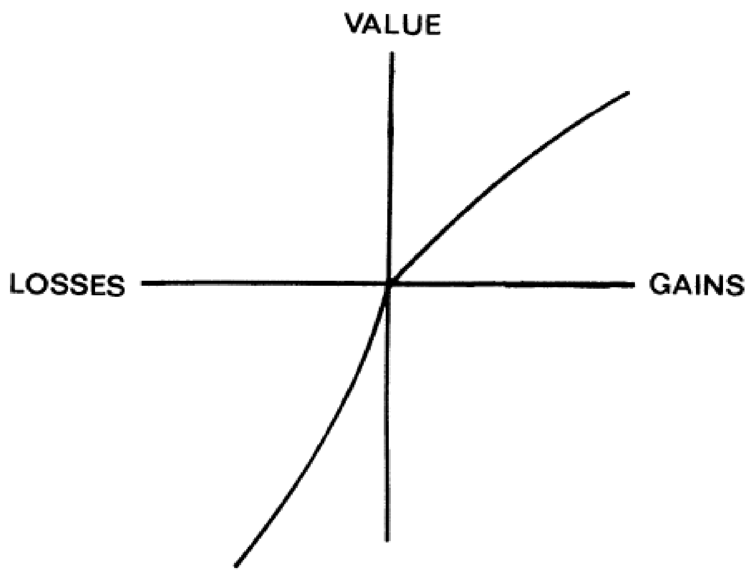


Figure A.2: Source: Kahneman and Tversky [177]

Cumulative Prospect Theory

How OPT violates first-order stochastic dominance Although quite plausible from a descriptive point of view, OPT has a major flaw from a normative perspective: it violates first-order stochastic dominance. Before we show how this might happen, a definition is needed:

Definition A.31. *Let X and Y be two random variables on a probability space (S, Σ, P). We say that X dominates Y in the sense of first-order stochastic dominance, if for each $t \in \mathbb{R}$, we have*

$$P\left(\{s \in S : X\left(s\right) \leq t\}\right) \leq P\left(\{s \in S : Y\left(s\right) \leq t\}\right) \tag{A.23}$$

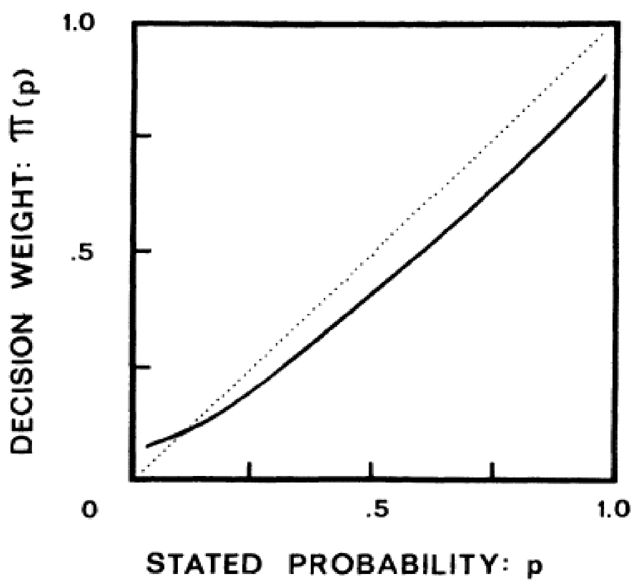


Figure A.3: Source: Kahneman and Tversky [177]

From a purely normative perspective, it seems natural to require of a functional representing preference to preserve first-order stochastic dominance. In other words, if a DM has preference $\succ$ over lotteries (or, more generally, over probability distributions) such that for any two lotteries L_1 and L_2, $L_1 \succ L_2 \Leftrightarrow V(L_1) > V(L_2)$, for some functional V over the DM's choice set, then it seems natural to ask of V to be such that $V(L_a) \geq V(L_b)$ for any two lotteries such that L_a dominates L_b in the sense of *first-order stochastic dominance.* For instance, the expected-utility functional preserves first-order stochastic dominance. However, this needs not hold for the function V_{opt} introduced above. To see why this is the case, consider the following four gambles:

1. Gamble A:

Win \$240	with probability 1

2. Gamble B:

Win \$1000	with probability 0.25
Win \$0	with probability 0.75

3. Gamble C:

Lose \$750	with probability 1

4. Gamble D:

Lose \$1000	with probability 0.75
Lose \$0	with probability 0.25

As we saw previously, in such a situation most individuals would prefer Gamble A to Gamble B and Gamble D to Gamble C. Now consider grouping gambles A and D together into a gamble that we will call Gamble AD and suppose that Gambles B and C are grouped into a gamble that we will call Gamble CD. Then:

1. Gamble AD:

Win \$240	with probability 0.25
Lose \$760	with probability 0.75

2. Gamble BC:

Win \$250	with probability 0.25
Lose \$750	with probability 0.75

It is easy to verify that Gamble BC dominates Gamble AD in the sense of first-order stochastic dominance, yet most individuals would prefer Gamble A to Gamble B and Gamble D to Gamble C.

CPT and the distortion of cumulative probabilities To overcome this major difficulty with OPT, Tversky and Kahneman [293] proposed to adopt the ideas developed by Quiggin [229] where cumulative probabilities of ranked outcomes are distorted instead of individual probabilities of ranked outcomes. This led to what is now called *Cumulative Prospect Theory* (CPT), which we describe below.

Suppose that a DM has preference $\succ$ over *lotteries* of the form

$$L = (x_1, p_1; x_2, p_2; \ldots; x_n, p_n) \tag{A.24}$$

Without loss of generality, we may assume that outcomes are ranked as follows $x_1 \geq x_2 \geq \ldots \geq x_k \geq 0 > x_{k+1} \geq \ldots \geq x_{n-1} \geq x_n$, for some $1 \leq k \leq n$. The positive outcomes $(x_1, \ldots, x_k)$ are called *gains* and the negative outcomes $(x_{k+1}, \ldots, x_n)$ are called *losses*.

We say that the DM's preference has a CPT representation when there is an increasing utility function $u : \mathbb{R} \to \mathbb{R}$ and two distortion functions $T^+ : [0,1] \to [0,1]$ and $T^- : [0,1] \to [0,1]$ such that for any two lotteries L_1 and L_2,

$$L_1 \succ L_2 \Leftrightarrow V_{cpt}(L_1) > V_{cpt}(L_2) \tag{A.25}$$

where for any lottery $L = (x_1, p_1; x_2, p_2; \ldots; x_n, p_n)$,

$$\begin{aligned} V_{cpt}(L) :=& \sum_{i=1}^{k} u(x_i) \left[T^{+}(p_1 + \ldots + p_i) - T^{+}(p_1 + \ldots + p_{i-1})\right] \\ &+ \sum_{i=k+1}^{n} u(x_i) \left[T^{-}(p_i + \ldots + p_n) - T^{-}(p_{i+1} + \ldots + p_n)\right] \end{aligned} \tag{A.26}$$

with the convention that $\sum_{j=1}^{0} p_j = 0$ and $\sum_{j=n+1}^{n} p_j = 0$.

Now, going back to the previous example, consider the gambles:

1. Gamble AD:

Win \$240	with probability 0.25
Lose \$760	with probability 0.75

2. Gamble BC:

Win \$250	with probability 0.25
Lose \$750	with probability 0.75

Recall that Gamble BC dominates Gamble AD in the sense of first-order stochastic dominance. Moreover,

$$V_{cpt}(AD) = u(240)\, T^{+}(0.25) + u(-760)\, T^{-}(0.75) \tag{A.27}$$

and

$$V_{cpt}(BC) = u(250)\, T^{+}(0.25) + u(-750)\, T^{-}(0.75) \tag{A.28}$$

Since the utility function is increasing, $u(250) > u(240)$ and $u(-750) > u(-760)$. Hence

$$V_{cpt}(BC) > V_{cpt}(AD) \tag{A.29}$$

and so V_{cpt} is consistent with the fact that Gamble BC dominates Gamble AD in the sense of first-order stochastic dominance.

Tversky and Kahneman [293] proposed a specific parameterization of their CPT model, by giving specific forms for the function u and for the distortion functions T^+ and T^-. Bernard and Ghossoub [32] and Ghossoub [141] extended the parameterization of Tversky and Kahneman [293] to continuous distributions and gave some equivalent alternative formulations of the CPT-functional. We refer the interested reader to [32] and [141] for more details.

A.3.2 The Choquet Expected-Utility Model and the Multiple Priors Model

The *Choquet Expected-Utility Model* (CEU) and the *Multiple Priors Model* (or *Maxmin Expected Utility* (MMEU) model) were introduced, respectively, by Schmeidler [275] and by Gilboa and Schmeidler [151] as a response to the inadequacy of Bayesianism in capturing situations of *Ambiguity*, or *Knightian uncertainty*, as shown in the Ellsberg paradox. As opposed to the *Bayesian* view of the Savage-Anscombe-Aumann tradition, the Gilboa-Schmeidler approach stipulates that in situations of Knightian uncertainty a decision maker (DM) is unable to formulate a unique additive prior (a unique subjective probability measure) to represent his beliefs. Instead, the DM can formulate either (i) a *unique non-additive prior* (a capacity), as in Shcmeidler [275] (or Gilboa [148]); or, (ii) a *unique collection of additive priors* as in Gilboa and Schmeidler [151]. In the latter case, the DM then bases his decision making process on the worst-case scenario: the minimum expected utility over his set of priors.

In this chapter, we recall the setting and axiomatization of (i) Schmeidler's representation theorem (schmeidler89); and, (ii) the Gilboa-Schmeidler representation theorem that appeared in [151]. Both used an Anscombe-Aumann framework. We will also discuss (i) the axiomatization of the CEU model in a purely subjective Savage framework, as given by Gilboa [148]; and, (ii) the axiomatization of the MMEU model in a purely subjective Savage setting as given by Alon and Schmeidler [6], or Casadesus-Masanell, Klibanoff and Ozdenoren [76].

Setup

Setting Let S be the set of all possible *states of the world*, and let Σ be a σ-algebra of *events*. Let X be the set of all possible *outcomes* and endow X with a σ-algebra Σ^*. Assume that $\{x\} \in \Sigma^*$, for each $x \in X$, and denote by $\Delta(X)$ the collection of all simple countably additive probability measures on (X, Σ^*). When endowed with the usual mixing operations for set functions, $\Delta(X)$ is a convex subset of the real vector space of all bounded countably additive measures on (S, Σ^*). Moreover, for each $s \in S$, let $X(s) := \{f(s) : f \in \mathcal{F}\}$ be the set of all possible consequences in the state of the world s. Then $X = \bigcup_{s \in S} X(s)$. In the following, we will assume (as Fishburn [128] does in Section 13.3) that $X(s) = X$, for each $s \in S$.

Denote by $\mathcal{H}$ the collection of all mappings of S into $\Delta(X)$, and the elements of $\mathcal{H}$ *Anscombe-Aumann (AA) acts.* Denote by $\mathcal{F}$ the collection of all mapping of S into X, and call the elements of $\mathcal{F}$ *Savage acts.* For each $x \in X$ let $\delta_x \in \Delta(X)$ denote the degenerate simple probability distribution assigning value 1 to the event $\{x\}$. We can then embed $\mathcal{F}$ into $\mathcal{H}$ by identifying x with δ_x, for each $x \in X$. Let $\mathcal{H}_0$ be the set of all those elements of $\mathcal{H}$ that take on finitely many values, that is, the finite step function in $\mathcal{H}$. Let $\mathcal{H}_c$ denote the collection of constant elements of $\mathcal{H}_0$.

Definition A.32. *Given a strict preference* $\succ$ *on* $\mathcal{H}_c$*, define weak preference* $\succcurlyeq$ *in the usual way. An act* $h \in \mathcal{H}$ *is called:*

1. Σ*-measurable when* $\{s \in S : h(s) \succ \mu\} \in \Sigma$ *and* $\{s \in S : h(s) \succcurlyeq \mu\} \in \Sigma$*, for each* $\mu \in \Delta(X)$*; and,*

2. $\succ$*-bounded if there are* $\mu_1, \mu_2 \in \Delta(X)$ *such that* $\mu_1 \succcurlyeq h(s) \succcurlyeq \mu_2$*, for each* $s \in S$*.*

Definition A.33 (Comonotonic Acts)**.** *Two acts* $f, g \in \mathcal{H}$ *are said to be comonotonic if there are no states of the world* $s, t \in S$ *such that*

$$f(s) \succ f(t) \quad \textit{but} \quad g(t) \succ g(s)$$

Preferences A decision maker (DM) expresses preferences over elements of $\mathcal{H}_0$, via a binary relation $\succ \subseteq \mathcal{H}_0 \times \mathcal{H}_0$. We define the binary relations $\succcurlyeq$ and $\sim$ from $\succ$ in the usual manner. We will sometimes make the usual abuse of notation as explained in section A.1.4 where the AA approach is discussed. Also, all definitions of section A.1.4 apply here.

Let $\mathcal{H}_{bm}$ denote the collection of all acts that are Σ-measurable and $\succ$-bounded. Then $\mathcal{H}_{bm}$ and contains $\mathcal{H}_0$, and hence $\mathcal{H}_c$.

Axioms on Preferences

Consider the following axioms on the DM's preference $\succ$ over elements of $\mathcal{H}_0$:

Axiom AA.1 (Weak Order)**.** *$\succ$ is asymmetric and negatively transitive.*

Axiom AA.2* (Certainty-Independence Axiom)**.** *for all $f, g \in \mathcal{H}_0$, for all $h \in \mathcal{H}_c$, and for all $\alpha \in (0,1)$, we have:*

$$f \succ g \Longleftrightarrow \alpha f + (1-\alpha) h \succ \alpha g + (1-\alpha) h \tag{A.30}$$

Axiom AA.2** (Comonotonic-Independence Axiom)**.** *for every pairwise comonotonic acts $f, g, h \in \mathcal{H}_0$, and for all $\alpha \in (0,1)$, we have:*

$$f \succ g \Longleftrightarrow \alpha f + (1-\alpha) h \succ \alpha g + (1-\alpha) h \tag{A.31}$$

Axiom AA.3 (Archimedean Axiom)**.** *for all $f, g, h \in \mathcal{H}_0$, we have:*

$$f \succ g \succ h \Rightarrow \exists\ \alpha, \beta \in (0,1), \alpha f + (1-\alpha) h \succ g \succ \beta f + (1-\beta) h \tag{A.32}$$

Axiom AA.4 (Nondegeneracy)**.** *There exist some $h, g \in \mathcal{H}_0$ such that $h \succ g$.*

Axiom AA.5 (Monotonicity)**.** *For each $g, h \in \mathcal{H}_0$, if $g(s) \succcurlyeq h(s)$, for all $s \in S$, then $g \succcurlyeq h$.*

Axiom AA.6 (Uncertainty Aversion)**.** *For each $g, h \in \mathcal{H}$, and for all $\alpha \in (0,1)$,*

$$h \sim g \Rightarrow \alpha h + (1-\alpha) g \succcurlyeq h \tag{A.33}$$

Schmeidler's Representation Theorem

Theorem A.34. *Let $\succ$ be a strict preference over elements of $\mathcal{H}_0$. If $\succ$ satisfies axioms AA.1, AA.2**, AA.3, AA.4, and AA.5, then there exists a nonconstant utility function $u : X \to \mathbb{R}$, unique up to a positive affine transformation, and a unique capacity ν on (S, Σ) such that for each $g, f \in \mathcal{H}_0$,*

$$h \succ g \Longleftrightarrow \oint_S \left[\int_X u \; dh\,(s)\right] \; d\nu\,(s) > \oint_S \left[\int_X u \; dg\,(s)\right] \; d\nu\,(s) \tag{A.34}$$

where the sign $\oint$ refers to integration in the sense of the Choquet integral (see Definition 5.7). This theorem can be extended to $\mathcal{H}_{bm}$ (see Schmeidler [275]).

The Gilboa-Schmeidler Representation Theorem

Theorem A.35. *Let $\succ$ be a strict preference over elements of $\mathcal{H}_0$. If $\succ$ satisfies axioms AA.1, AA.2*, AA.3, AA.4, AA.5, and AA.6, then there exists a nonconstant utility function $u : X \to \mathbb{R}$, unique up to a positive affine transformation, and a closed*[8] *and convex set $\mathcal{C}$ of probabilites on (S, Σ) such that for each $g, f \in \mathcal{H}_0$,*

$$h \succ g \Longleftrightarrow \min_{\mu \in \mathcal{C}} \int_S \left[\int_X u \; dh\,(s)\right] \; d\mu\,(s) > \min_{\mu \in \mathcal{C}} \int_S \left[\int_X u \; dg\,(s)\right] \; d\mu\,(s) \tag{A.35}$$

This theorem can be extended to $\mathcal{H}_{bm}$ (see Gilboa and Schmeidler [151]).

Purely Subjective CEU and MMEU

The setting of Schmeidler [275] and of Gilboa and Schmeidler [151] is simply the Anscome-Aumann setup where uncertainty is neither totally objective nor totally subjective. Is it possible to obtain Choquet expected utility and / or a multiple priors representation of preferences in a purely subjective framework à la Savage? This has been done by Gilboa [148] for the case of CEU and by Alon and Schmeidler [6], and Casadesus-Masanell, Klibanoff and Ozdenoren [76] for the case of MMEU.

[8] In the weak* topology on $\Delta(X)$. For more about *weak** topologies on collections of probability measures see Appendix D.4.

We will not review the axiomatization here; we will simply state the results. Alon and Schmeidler [6] consider a finite state space, and a connected topological space of outcomes. We will only review the setup of Casadesus-Masanell, Klibanoff and Ozdenoren [76] since they deal with an arbitrary state space.

Let S be a set of states of nature and Σ a σ-algebra of events on S. let X be a set of prizes, taken to be some closed interval on the real line.

Definition A.36. *We will use the following terminology:*

1. *An act is a mapping $f : S \to X$;*
2. *An act is called simple if takes only finitely many values (i.e. it is a finite step function);*
3. *An act f is called Σ-measurable when $f^{-1}(B) \in \Sigma$ for any Borel subset B of X.*

Let $\mathcal{F}$ denote the collection of all Σ-measurable acts. Then Casadesus-Masanell, Klibanoff and Ozdenoren [76] give axioms for a DM's preference $\succ$ over elements of $\mathcal{F}$ to have a representation of the form: For all $f, g \in \mathcal{F}$,

$$f \succ g \Longleftrightarrow \min_{P \in \mathcal{C}} \int_S u \circ f \; dP > \min_{P \in \mathcal{C}} \int_S u \circ g \; dP \tag{A.36}$$

for some continuous and strictly increasing utility function $u : X \to \mathbb{R}$, unique up to a positive affine transformation, and a (weak*) compact and convex set $\mathcal{C}$ of probabilities on (S, Σ).

Gilboa [148] gives axioms for the DM's preference $\succ$ over elements of $\mathcal{F}$ to have a representation of the form: For all $f, g \in \mathcal{F}$,

$$f \succ g \Longleftrightarrow \oint_S u \circ f \; d\nu > \oint_S u \circ g \; d\nu \tag{A.37}$$

for some bounded utility function $u : X \to \mathbb{R}$, unique up to a positive affine transformation, and a unique capacity on (S, Σ).

Appendix B

Topology

All of the definitions and results of this section are fairly elementary. Three classical references on this subject are Bourbaki [58], Kelley [180] and Willard [299]. All proofs will be omitted and the reader is referred to either one of the aforementioned references. Other references include Aliprantis and Border [3] (chap. 2 and 3), Dudley [108] (chap. 2), Dunford and Schwartz [109] (chap. I), Hewitt and Stromberg [166] (chap. 2), and Kolmogorov and Fomin [186] (chap. 2 and 3).

B.1 Topological Spaces

B.1.1 Definitions

Definition B.1. *A topology $\mathcal{T}$ on a nonempty set X is a subset of the collection $\mathcal{P}(X)$ of all subsets of X, such that:*

1. *$\varnothing \in \mathcal{T}$ and $X \in \mathcal{T}$;*

2. *$\mathcal{T}$ is closed under arbitrary unions; and,*

3. *$\mathcal{T}$ is closed under finite intersections*

If $\mathcal{T}$ is a topology on a set X, then we say that the space $(X, \mathcal{T})$ is a topological space.

Definition B.2. *An element of $\mathcal{T}$ is called an open set. Moreover, a subset F of X is called closed if and only if its complement in X is open.*

From the previous two definitions, it follows that arbitrary intersections of closed sets are closed and finite unions of closed sets are closed.

Definition B.3. *Let $(X, \mathcal{T})$ be a topological space and $H \subseteq X$. The closure of H (in X for the topology $\mathcal{T}$) is the smallest closed set containing H. We denote the closure of H by $\overline{H}$. It is easily verified that*

$$\overline{H} = \bigcap \{F : H \subseteq F, X \setminus F \in \mathcal{T}\}$$

Definition B.4. *Let $(X, \mathcal{T})$ be a topological space and $H \subseteq X$. The interior of H (in X for the topology $\mathcal{T}$) is the largest open set contained in H. We denote the interior of H by H°. It is easily verified that*

$$H^{\circ} = \bigcup \{U : U \subseteq H, U \in \mathcal{T}\}$$

Definition B.5. *Let $(X, \mathcal{T})$ be a topological space and $H \subseteq X$. H is said to be dense in X if $\overline{H} = X$.*

Definition B.6. *A topological space $(X, \mathcal{T})$ is called separable if it has a countable dense subset.*

For instance, $\mathbb{R}$ with its usual topology is separable, since the rationals are dense.

Definition B.7. *Let $(X, \mathcal{T})$ be a topological space, and fix $x \in X$. A neighborhood of x is a subset U of X such that $x \in G \subseteq U$, for some $G \in \mathcal{T}$. We denote by $\mathcal{U}_x$ the collection of all neighborhoods of x, and we call it the neighborhood system at x.*

Definition B.8. *Let $(X, \mathcal{T})$ be a topological space. A neighborhood base $\mathcal{B}_x$ at a point $x \in X$ is a collection $\mathcal{B}_x \subseteq \mathcal{U}_x$ such that*

$$U \in \mathcal{U}_x \Rightarrow \ \exists \ B \in \mathcal{B}_x, B \subseteq U \tag{B.1}$$

We refer to the elements of $\mathcal{B}_x$ as basic neighborhoods of the point x.

Definition B.9. *Let $(X, \mathcal{T})$ be a topological space. A base for the topology is a collection $\mathcal{B} \subseteq \mathcal{T}$ such that for every $G \in \mathcal{T}$, there exists $\mathcal{C} \subseteq \mathcal{B}$ such that that $G = \bigcup\{B : B \in \mathcal{C}\}$. That is, every open set is a union of elements of $\mathcal{B}$. Note that if $\mathcal{C}$ is empty, then $\bigcup\{B : B \in \mathcal{C}\}$ is also empty, so we do not need to include the empty set in our base.*

A subbase for the topology is a collection $\mathcal{S} \subseteq \mathcal{T}$ such that the collection $\mathcal{B}$ of all finite intersections of elements of $\mathcal{S}$ forms a base for the topology $\mathcal{T}$.

B.1.2 Compactness

Definition B.10. *Let $(X, \mathcal{T})$ be a topological space and let $E \subseteq X$. A cover (resp. open cover) of E is a collection $\{H_\alpha\}_{\alpha\in\Lambda}$ of sets (resp. open sets) in X such that $E \subseteq \bigcup_{\alpha\in\Lambda} H_\alpha$.*

A subcover (resp. open subcover) for E of the cover (resp. open cover) $\{H_\alpha\}_{\alpha\in\Lambda}$ is a subcollection of $\{H_\alpha\}_{\alpha\in\Lambda}$ which is again a cover (resp. open cover) of E.

Definition B.11. *Let $(X, \mathcal{T})$ be a topological space. A subset K of X is called compact if every open cover of K has a finite subcover.*

Definition B.12. *Let $(X, \mathcal{T})$ be a topological space and $Y \subseteq X$. The relative topology on Y is the collection $\mathcal{T}_Y := \{G \cap Y : G \in \mathcal{T}\}$.*

One can easily verify that the relative topology on Y is itself a topology on Y. More importantly, compactness is not a relative notion. specifically:

Proposition B.13. *Let $(X, \mathcal{T})$ be a topological space, and let $K \subseteq Y \subseteq X$. Then the following are equivalent:*

1. *K is a compact subset of the topological space $(X, \mathcal{T})$; and,*
2. *K is a compact subset of the topological space $(Y, \mathcal{T}_Y)$.*

Proposition B.14. *Closed subsets of compact sets are compact.*

B.1.3 Hausdorff Spaces and Normal Spaces

Definition B.15. *A topological space $(X, \mathcal{T})$ is called a Hausdorff space if for any two points $x, y \in X$ such that $x \neq y$, there is some $U \in \mathcal{U}_x$ and some $V \in \mathcal{U}_y$ such that $U \cap V = \varnothing$.*

Proposition B.16. *Let $(X, \mathcal{T})$ be a Hausdorff topological space, and let $K \subseteq X$. If K is compact then K is closed.*

In particular, singletons are closed in a Hausdorff space.

Definition B.17. *A topological space $(X, \mathcal{T})$ is called a normal space if for every disjoint closed sets A and B, there are disjoint open sets U and V such that $A \subset U$ and $B \subset V$.*

B.1.4 Convergence in Topological Spaces

Definition B.18. *A set Λ is called a directed set if there is a binary relation $\leq$ on Λ such that:*

1. *$\leq$ is reflexive;*
2. *$\leq$ is transitive; and,*
3. *For all $\lambda_1, \lambda_2 \in \Lambda$, there is some $\lambda_3 \in \Lambda$ such that $\lambda_1 \leq \lambda_3$ and $\lambda_2 \leq \lambda_3$.*

For instance, the set $\mathbb{N}$ with its usual order is a directed set.

Definition B.19. *A net on a topological space $(X, \mathcal{T})$ is a function $p : \Lambda \to X$, where Λ is a directed set. It is customary to write x_λ instead of $p(\lambda)$, for any $\lambda \in \Lambda$. We denote the net by $(x_\lambda)_{\lambda \in \Lambda}$*

For instance, any sequence on $(X, \mathcal{T})$ is a net on $(X, \mathcal{T})$.

Definition B.20. *Let $(M, \leq_1)$ and $(\Lambda, \leq_2)$ be nonempty directed sets. A function $q : M \to \Lambda$ is called:*

1. *Increasing if for all $m_1, m_2 \in M$,*

$$m_1 \leq_1 m_2 \Rightarrow q(m_1) \leq_2 q(m_2)$$

2. *Cofinal if for any $\lambda_0 \in \Lambda$, there is some $m_0 \in M$ with $\lambda_0 \leq_2 q(m_0)$*

Definition B.21. *Let $(X, \mathcal{T})$ be a topological space, and let $(M, \leq_1)$ and $(\Lambda, \leq_2)$ be directed sets. Let $(x_\lambda)_{\lambda\in\Lambda}$ be a net in X. A subnet of $(x_\lambda)_{\lambda\in\Lambda}$ is a function of the form $p \circ q : M \to X$, where $q : M \to \Lambda$ is increasing and cofinal. It is customary to write $\left(x_{\lambda_\mu}\right)_{\mu\in M}$ for the subnet, i.e. $x_{\lambda_\mu} = p(q(\mu))$.*

For instance, any subsequence of a sequence is a subnet of that sequence.

Definition B.22. *A net $(x_\lambda)_{\lambda\in\Lambda}$ in a topological space $(X, \mathcal{T})$ is said to converge to a point $y \in X$, and we write $\lim_{\lambda\in\Lambda} x_\lambda = y$ if for any $U \in \mathcal{U}_y$ there is some $\lambda_0 \in \Lambda$ such that for all $\lambda \in \Lambda$,*

$$\lambda_0 \leq \lambda \Rightarrow x_\lambda \in U$$

The previous definition is a natural generalization of the notion of convergence for sequences. In a metric space, sequences can be used to characterize continuity of functions, closure of sets, and compactness. In more genral topological spaces, however, this can only be done using nets. As a matter of fact, this is the main motivation behind the notion of a net, as shown in the following propositions (which are natural generalization of classical results for metric spaces).

Definition B.23. *Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. A mapping $f : X \to Y$ is called continuous if $f^{-1}(G) \in \mathcal{T}_X$, for each $G \in \mathcal{T}_Y$.*

Let $(x_\lambda)_{\lambda\in\Lambda}$ be a net in a topological space $(X, \mathcal{T}_X)$. Let $(Y, \mathcal{T}_Y)$ be another topological space, and let $f : X \to Y$ be any mapping of X into Y. One can then easily observe that $(f(x_\lambda))_{\lambda\in\Lambda}$ is a net in the topological space $(Y, \mathcal{T}_Y)$.

Proposition B.24. *Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. A mapping $f : X \to Y$ is continuous if and only if for any net $(x_\lambda)_{\lambda\in\Lambda}$ in X such that $\lim_{\lambda\in\Lambda} x_\lambda = x$, for some $x \in X$, we have $\lim_{\lambda\in\Lambda} f(x_\lambda) = f(x) \in Y$.*

Proposition B.25. *Let $(X, \mathcal{T})$ be a topological space and $H \subseteq X$. Denote by $\overline{H}$ the closure of H. Then a point $x \in X$ belongs to $\overline{H}$ if and only if there is a net $(x_\lambda)_{\lambda \in \Lambda}$ of points in H such that $\lim_{\lambda \in \Lambda} x_\lambda = x$.*

Corollary B.26. *Let $(X, \mathcal{T})$ be a topological space. A subset F of X is closed if and only if F contains all the limit points of all nets in F.*

Limits of nets need not be unique. However, when the space is Hausdorff, all limits are unique:

Proposition B.27. *Let $(X, \mathcal{T})$ be a topological space. Then the following are equivalent:*

1. *$(X, \mathcal{T})$ is a Hausdorff space; and,*

2. *Every convergent net in X has a unique limit.*

Proposition B.28. *Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be a topological spaces. Let K be a compact subset of X. If $f : K \to Y$ is a continuous mapping then $f(K)$ is a compact subset of Y.*

Proposition B.29. *Let $(X, \mathcal{T})$ be a topological space, and let K be a compact subset of X. If $f : K \to \mathbb{R}$ is continuous then f is bounded f attains its supremum and its infimum.*

B.1.5 Nets and Compactness

Proposition B.30. *Let $(X, \mathcal{T})$ be a topological space, and let $(x_\lambda)_{\lambda \in \Lambda}$ be a net in X. Suppose that $\lim_{\lambda \in \Lambda} x_\lambda = x \in X$. If $\left(x_{\lambda_\mu}\right)_{\mu \in M}$ is a subnet of $(x_\lambda)_{\lambda \in \Lambda}$, then $\lim_{\mu \in M} x_{\lambda_\mu} = x$.*

Proposition B.31. *Let $(X, \mathcal{T})$ be a topological space. Then the following are equivalent:*

1. *X is compact; and,*

2. *Every net in X admits a convergent subnet.*

B.1.6 Connectedness

Definition B.32. *A topological space $(X, \mathcal{T})$ is called disconnected if there are some nonempty sets $G, H \in \mathcal{T}$ such that:*

1. $G \cap H = \varnothing$*; and,*

2. $G \cup H = X$.

We say that X is connected when it is not disconnected.

Proposition B.33. *A topological space $(X, \mathcal{T})$ is connected if and only if the only subsets of X that are both open and closed are $\varnothing$ and X.*

Proposition B.34. *Let $(X, \mathcal{T}_X)$ be a connected topological space, let $(Y, \mathcal{T}_Y)$ be an arbitrary topological space, and let $f : X \to Y$ be any mapping of X into Y. If f is continuous then $(Y, \mathcal{T}_Y)$ is connected.*

B.1.7 Product Topologies

The following definitions are taken form lecture notes for the course PMATH 702, as given by Prof. Laurent W. Marcoux. The first is a natural abstraction of the notion of a Cartesian product of finitely many spaces.

Definition B.35. *Let $\{(X_\alpha, \mathcal{T}_\alpha)\}_{\alpha\in\Gamma}$ be an arbitrary collection of topological spaces. We define the Cartesian product of the sets $\{X_\alpha\}_{\alpha\in\Gamma}$ as the collection*

$$\prod_{\alpha\in\Gamma} X_\alpha := \left\{ x : \Gamma \to \bigcup_{\alpha\in\Gamma} X_\alpha \mid x(\alpha) \in X_\alpha, \forall\ \alpha \in \Gamma \right\}$$

It is customary to denote $(x_\alpha)_{\alpha\in\Gamma}$ by x.

Definition B.36. *Let $\{(X_\alpha, \mathcal{T}_\alpha)\}_{\alpha\in\Gamma}$ be an arbitrary collection of topological spaces, and let $\prod_{\alpha\in\Gamma} X_\alpha$ denote their Cartesian product. For each $\beta \in \Gamma$, the map $\pi_\beta : \prod_{\alpha\in\Gamma} X_\alpha \to X_\beta$ defined by $\pi_\beta(x) = x_\beta$ is called the βth projection map.*

Definition B.37. *The product topology on $\prod_{\alpha\in\Gamma} X_\alpha$ is the smallest collection for which the collection $\mathcal{B} = \left\{\prod_{\alpha\in\Gamma} U_\alpha\right\}$ is a base, where:*

1. *$U_\alpha \in \mathcal{T}_\alpha$, for all $\alpha \in \Gamma$; and,*
2. *$U_\alpha = X_\alpha$, for all but finitely many $\alpha \in \Gamma$.*

In the definition above, it suffices to ask that we take $U_\alpha \in \mathcal{B}_\alpha$, where $\mathcal{B}_\alpha$ is a fixed base for $\mathcal{T}_\alpha$, for each $\alpha \in \Gamma$. Moreover, note that if $U_\alpha = X_\alpha$ for all $\alpha \in \Gamma$ except for $\alpha_1, \alpha_2, \ldots, \alpha_n$, then

$$\prod_{\alpha\in\Gamma} U_\alpha = \pi_{\alpha_1}^{-1}\left(U_{\alpha_1}\right) \cap \pi_{\alpha_2}^{-1}\left(U_{\alpha_2}\right) \cap \ldots \cap \pi_{\alpha_n}^{-1}\left(U_{\alpha_n}\right)$$

Therefore, the collection $\{\pi_\alpha^{-1}(U_\alpha) : U_\alpha \in \mathcal{B}_\alpha, \alpha \in \Gamma\}$ is a subbase for the product topology on $\prod_{\alpha\in\Gamma} X_\alpha$, where $\mathcal{B}_\alpha$ is a base (or even a subbase) for the topology on X_α. Consequently, the product topology is the smallest topology that makes all the projection maps continuous.

B.2 Metric Spaces

B.2.1 Metrization

Definition B.38. *A metric on an arbitrary nonempty space X is a mapping $d : X\times X \to \mathbb{R}$ such that:*

1. *For all $x, y \in X$, $d(x, y) \geq 0$;*
2. *For all $x, y \in X$, $d(x, y) = d(y, x)$;*
3. *For all $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$; and,*
4. *$d(x, x) = 0$ and for all $x, y \in X$, $d(x, y) = 0$ if and only if $x = y$.*

When d is a metric on a set X, we call the pair (X, d) a metric space.

Definition B.39. *Let (X, d) be a metric space, $x \in X$, and $r > 0$. We define the open ball of radius r with center at x as the subset $B_r(x)$ of X defined by:*

$$B_r(x) := \{y \in X : d(x, y) < r\}$$

Definition B.40. *Let (X, d) be a metric space, and fix $x \in X$. The collection $\mathcal{B}_x := \{B_r(x) : r > 0\}$ of all open balls centered at x forms a neighborhood base at x for a topology $\mathcal{T}$ on X. We refer to $\mathcal{T}$ as a topology induced by the metric d, and we often say that the metric d is compatible with the topology $\mathcal{T}$, or that the topology $\mathcal{T}$ is metrizable by the metric d.*

Definition B.41. *Let $(X, \mathcal{T})$ be a topological space. We say that $\mathcal{T}$ is metrizable if there exists a metric d on X such that $\mathcal{T}$ is induced by the metric d.*

Definition B.42. *Let (X, d) be a metric space, and let $(x_n)_{n\geq 1}$ be a sequence in X. We say that:*

1. *$(x_n)_{n\geq 1}$ converges to some $x \in X$ if for any $\varepsilon > 0$ there is some $N \in \mathbb{N}$ such that for any $n \geq 1$,*
$$n \geq N \Rightarrow d(x_n, x) < \varepsilon$$

2. *$(x_n)_{n\geq 1}$ is a Cauchy sequence if for any $\varepsilon > 0$ there is some $N \in \mathbb{N}$ such that for any $m, n \geq 1$,*
$$m, n \geq N \Rightarrow d(x_m, x_n) < \varepsilon$$

Any convergent sequence is hence a Cauchy sequence.

Proposition B.43. *Any metric space is a Hausdorff space. Therefore, limits are unique in metric spaces.*

Proposition B.44. *Any metric space is a normal space.*

Definition B.45. *A metric space (X, d) is called:*

1. *Complete if every Cauchy sequence in (X, d) converges in (X, d);*

2. *Separable if the topology induced by the metric is separable; and,*

3. *A Polish space if it is both complete and separable.*

Proposition B.46. *A closed subset of a complete metric space is complete (for its relative metric topology). Moreover, any complete subset of a metric space is closed.*

Proposition B.47. *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f : X \to Y$ be any mapping of X into Y. Then the following are equivalent:*

1. *f is continuous; and,*

2. *For any $x \in X$, if $(x_n)_{n\geq 1}$ is a sequence in X that converges to x, then the sequence $(f(x_n))_{n\geq 1}$ in Y converges to $f(x)$ (in this case, we say that f is sequentially continuous).*

Proposition B.48. *A compact subset of a metric space is closed*

B.2.2 Total Boundedness and Sequential Compactness

Definition B.49. *Let (X, d) be a metric space, and fix $\varepsilon > 0$. An ε-net for X is a finite subset of X of the form $\{x_1, x_2, \ldots, x_N\}$, for some $N \in \mathbb{N}$, such that $X \subseteq \bigcup_{i=1}^{N} B_\varepsilon(x_i)$.*

Definition B.50. *A metric space (X, d) is called totally bounded if it admits an ε-net, for every $\varepsilon > 0$.*

Proposition B.51. *Every totally bounded metric space is bounded. The converse generally fails to hold.*

Definition B.52. *A topological space $(X, \mathcal{T})$ is called sequentially compact if every sequence $(x_n)_{n\geq 1}$ in X admits a convergent subsequence.*

Proposition B.53. *Let (X, d) be a metric space. Then the following are equivalent:*

1. *X is compact;*
2. *X is sequentially compact; and,*
3. *X is complete and totally bounded.*

Appendix C

Duality in Normed Linear Spaces

Although extremely rich and deep, the theory of duality in topological vector spaces has fairly basic foundations. In this Appendix we focus on the special case of normed vector spaces. All material presented here is classical, and can be found in Aliprantis and Border [3] (chap. 5 and 6), Dudley [108] (chap. 6), Hewitt and Stromberg [166] (chap. 4), Kolmogorov and Fomin [186] (chap. 4 and 5), Megginson [210] (chap. 1 and 2), or Rudin [248] (chap. 4). For a very thorough treatment we refer the reader to Diestel [100], Dunford and Schwartz [109], Kelley and Namioka [181], Schaefer [270]. A nice unpublished introductory book is Marcoux [200]. The latter is based on lecture notes of Prof. Laurent W. Marcoux for the course PMATH 653.

C.1 Normed Linear Spaces

Definition C.1. *Let X be any vector space over the field $\mathbb{R}$, and let 0 denote the zero vector. A norm on X is a mapping $\|.\| : X \to \mathbb{R}$ such that:*

1. *For all $x \in X$, $\|x\| \geq 0$;*
2. *For all $x \in X$ and for all $c \in \mathbb{R}$, $\|c.x\| = |c|.\|x\|$;*
3. *For all $x, y \in X$, $\|x + y\| \leq \|x\| + \|y\|$; and,*
4. *$\|x\| = 0$ if and only $x = 0$.*

If $\|.\|$ is a norm on a real vector space X then we refer to the pair $(X, \|.\|)$ as a normed linear space.

Let $(X, \|.\|)$ be a normed linear space, and define the mapping $d : X \times X \to \mathbb{R}$ by $d(x, y) := \|x - y\|$, for all $x, y \in X$. Then one can easily verify that d is a metric on X. We refer to d as the *norm metric*. Let $\mathcal{T}$ be the topology on X induced by the norm metric d. We then call $\mathcal{T}$ the *norm topology*, or *strong topology*, on X. Hence, any normed linear space is a metrizable topological space.

Clearly, for any $x \in X$ a neighborhood base at x for the strong topology on the normed linear space $(X, \|.\|)$ is given by the collection $\mathcal{B}_x := \{B_r(x) : r > 0\}$, where $B_r(x) := \{y \in X : \|x - y\| < r\}$, for any $r > 0$. Moreover, whenever we refer to *convergence*, *continuity*, *closure*, *compactness*, etc. *in norm*, we mean *convergence*, *continuity*, *closure*, *compactness*, etc. for the *norm topology* defined above.

Definition C.2. *A Banach space is a normed linear space which is complete for the norm metric defined above.*

C.2 The Topological Dual

Definition C.3. *Let $(X, \|.\|)$ be a normed linear space over the field $\mathbb{R}$, and let $T : X \to \mathbb{R}$ be a functional on X. T is called linear if for any $x, y \in X$ and for any $\alpha, \beta \in \mathbb{R}$,*

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$$

Definition C.4. *Let $(X, \|.\|)$ be a normed linear space, and let $T : X \to \mathbb{R}$ be a linear functional on X. We say that T is bounded if there is some $c \geq 0$ such that, for all $x \in X$, $|T(x)| \leq c\|x\|$.*

Definition C.5. *Let $(X, \|.\|)$ be a normed linear space, and let $T : X \to \mathbb{R}$ be a bounded linear functional on X. We define*

$$\|T\|^* := \inf \{c \geq 0 : |T(x)| \leq c\|x\|, \ \forall\, x \in X\}$$

Then, by boundedness of T, $\|T\|^* < +\infty$

Proposition C.6. *Let $(X, \|.\|)$ be a normed linear space, and let $T : X \to \mathbb{R}$ be a bounded linear functional on X. Then the following quantities are equal:*

1. $\|T\|^*$;

2. $\sup\{|T(x)| : x \in X, \|x\| = 1\}$; and,

3. $\sup\{|T(x)|/\|x\| : x \in X, x \neq 0\}$.

Proposition C.7. *Let $(X, \|.\|)$ be a normed linear space, let 0 denote the zero vector in X, and let $T : X \to \mathbb{R}$ be a linear functional on X. Then the following are equivalent:*

1. *T is continuous (on X);*

2. *T is continuous at 0;*

3. *T is bounded; and,*

4. $\sup\{|T(x)| : x \in X, \|x\| = 1\} < +\infty$.

Definition C.8. *Let $(X, \|.\|)$ be a normed linear space. We denote by X^* the collection of all bounded linear (and hence continuous) functionals on X, and we call X^* the (topological) dual (or norm-dual, or simply dual) of X.*

Proposition C.9. *Let $(X, \|.\|)$ be a normed linear space, and let X^* denote the dual of X. Then $\|.\|^*$ is a norm on X^* and the normed linear space $(X^*, \|.\|^*)$ is a Banach space.*

Since $(X^*, \|.\|^*)$ is itself a normed linear space we can, in turn, construct the dual space of $(X^*, \|.\|^*)$. We call this space the double dual, or second dual of X.

Definition C.10. *Let $(X, \|.\|)$ be a normed linear space, and let $(X^*, \|.\|^*)$ be its dual space. We define the double dual (or second dual) of X as the dual of X^*, and we denote it by X^{**}.*

Definition C.11. *Let $(X, \|.\|)$ be a normed linear space, let $(X^*, \|.\|^*)$ be its dual space, and let X^{**} be its second dual. We can embed X into X^{**} via the canonical embedding $\Gamma : X \to X^{**}$ defined by $\Gamma(x) = \hat{x}$, for any $x \in X$, where for any $x \in X$, $\hat{x} : X^* \to \mathbb{R}$ is defined by $\hat{x}(\phi) = \phi(x)$, for each $\phi \in X^*$.*

One can easily verify that the canonical embedding is a linear map.

Definition C.12. *Let $(X, \|.\|)$ be a normed linear space with second dual X^{**}, and let Γ denote the canonical embedding of X into X^{**}. If $\Gamma(X) = X^{**}$ then X is called reflexive.*

C.3 Weak and Weak* Topologies

C.3.1 Definitions and Properties

Let $(X, \|.\|)$ be a normed linear space, let $(X^*, \|.\|^*)$ be its dual space, and let X^{**} be its second dual. The topology on X^* induced by the metric norm associated with the norm $\|.\|^*$ is typically a very strong topology, in the sense that it has too many open sets. In linear analysis it is customary to define two weaker topologies on X^*, namely the topology on X^* induced by X^{**} (denoted by $\sigma(X^*, X^{**})$) and the topology on X^* induced by X (denoted by $\sigma(X^*, X)$). Before we define and discuss these topologies, we first define the notion of a *topology induced* on a set by some other set.

Definition C.13. *Let X be an arbitrary nonempty set, and for each $\alpha \in \Lambda$ let $f_\alpha : X \to X_\alpha$ be a mapping of X into some topological space $(X_\alpha, \mathcal{T}_\alpha)$. Let Y denote the collection $\{f_\alpha\}_{\alpha\in\Lambda}$. The topology on X induced by Y is the smallest topology on X that makes f_α continuous, for each $\alpha \in \Lambda$. A subbase for this topology is the collection $\{f_\alpha^{-1}(U_\alpha) : \alpha \in \Lambda\}$, where $U_\alpha \in \mathcal{T}_\alpha$, for each $\alpha \in \Lambda$.*

In the definition above, it suffices to ask that we take $U_\alpha \in \mathcal{B}_\alpha$, where $\mathcal{B}_\alpha$ is a fixed base for $\mathcal{T}_\alpha$, for each $\alpha \in \Lambda$. Moreover, recall the definition of the product topology on a Cartesian product of topological space, and not that the product topology is simply the topology induced by the collection of all projection mappings.

Now, if in the previous definition $(X_\alpha, \mathcal{T}_\alpha)$ is simply the real line with the Borel σ-algebra, then a subbase for the topology on X induced by Y is the collection $\{f_\alpha^{-1}(U_\alpha) : \alpha \in \Lambda\}$, U_α is a Borel set, for each $\alpha \in \Lambda$.

Definition C.14. *Let $(X, \|.\|)$ be a normed linear space, let $(X^*, \|.\|^*)$ be its dual space, and let X^{**} be its second dual. Let Γ be the canonical embedding of X into X^{**}. Then:*

1. *The weak topology on X, denoted by $\sigma(X, X^*)$ is the topology on X induced by X^*. That is, $\sigma(X, X^*)$ is the topology on X that makes all $\|.\|$-bounded and linear functionals on X continuous;*

2. *The weak topology on X^*, denoted by $\sigma(X^*, X^{**})$ is the topology on X^* induced by X^{**}. That is, $\sigma(X^*, X^{**})$ is the topology on X^* that makes all $\|.\|^*$-bounded and linear functionals on X^* continuous; and,*

3. *The weak* topology on X^*, denoted by $\sigma(X^*, X)$ is the topology on X^* induced by $\Gamma(X)$.*

Proposition C.15. *Let $(X, \|.\|)$ be a normed linear space, let $(X^*, \|.\|^*)$ be its dual space, and let X^{**} be its second dual. If $\mathcal{T}^*$ denotes the strong topology on X^*, that is the topology induced by the metric generated by the operator norm $\|.\|^*$, then:*

$$\sigma(X^*, X) \subseteq \sigma(X^*, X^{**}) \subseteq \mathcal{T}^*$$

This means that the weak* topology on the dual space is weaker than the weak topology, which is in turn weaker than the strong topology.

Proposition C.16. *Let $(X, \|.\|)$ be a normed linear space, and let M be a linear subspace of X. Then the weak topology of the normed linear space $(M, \|.\|)$ coincides with the relative weak topology on M.*

Proof. See Megginson [210], proposition 2.5.22 on p. 218. □

In what follows we refer to compactness (resp. closure, convergence, continuity, etc.) with respect to the weak topology on X^* as *weak compactness* (resp. *weak closure*, *weak convergence*, *weak continuity*, etc.). Similarly, we refer to compactness (resp. closure, convergence, continuity, etc.) with respect to the weak* topology on X^* as *weak* compactness* (resp. *weak* closure*, *weak* convergence*, *weak* continuity*, etc.).

Definition C.17. *Let $(X, \|.\|)$ be a normed linear space, let $(X^*, \|.\|^*)$ be its dual space, and let X^{**} be its second dual. Then:*

1. *A net $(x_\alpha)_{\alpha\in\Gamma}$ in X converges weakly to some $x \in X$ if and only if the net $(T(x_\alpha))_{\alpha\in\Gamma}$ converges to $T(x)$, for each $T \in X^*$;*

2. *A net $(\phi_\alpha)_{\alpha\in\Gamma}$ in X^* converges weakly to some $\phi \in X^*$ if and only if the net $(\Psi(\phi_\alpha))_{\alpha\in\Gamma}$ converges to $\Psi(\phi)$, for each $\Psi \in X^{**}$; and,*

3. *A net $(\phi_\alpha)_{\alpha\in\Gamma}$ in X^* weak*-converges to some $\phi \in X^*$ if and only if the net $(\phi_\alpha(x))_{\alpha\in\Gamma}$ converges to $\phi(x)$, for each $x \in X$.*

Definition C.18. *For any normed linear space $(X, \|.\|)$, define the closed unit ball of X as*

$$B_1[X] := \{x \in X : \|x\| \leq 1\}$$

Theorem C.19 (Goldstine)**.** *Let $(X, \|.\|)$ be a normed linear space, let $(X^*, \|.\|^*)$ be its dual space, and let X^{**} be its second dual. Denote by Γ the canonical embedding of X into X^{**}. Then $\Gamma(B_1[X])$ is weak*-dense in $B_1[X^{**}]$. Consequently, X is weak*-dense in X^{**}.*

C.3.2 Boundedness

Definition C.20. *Let $(X, \|.\|)$ be a normed linear space, let $(X^*, \|.\|^*)$ be its dual space, and let X^{**} be its second dual. Let 0 denote the zero vector of X, and let 0^* denote the zero vector of X^*. Denote by:*

1. *$\mathcal{T}$ the norm topology on X;*

2. *$\sigma(X, X^*)$ the weak topology on X;*

3. *$\mathcal{T}^*$ the norm topology on X^*;*

4. *$\sigma(X^*, X^{**})$ the weak topology on X^*; and,*

5. *$\sigma(X^*, X)$ the weak* topology on X^*.*

For any subset A of X, and for any $t > 0$, denote by $t.A$ the set $\{t.x : x \in A\}$. Let H be any subset of X and H^* any subset of X^*. We say that:

1. *H is bounded (or $\mathcal{T}$-bounded) if for any $\mathcal{T}$-neighborhood U of 0 there is some $t_U > 0$ such that $H \subseteq s.U$, for all $s > t_U$;*

2. *H is weakly bounded (or $\sigma(X, X^*)$-bounded) if for any $\sigma(X, X^*)$-neighborhood V of 0 there is some $t_V > 0$ such that $H \subseteq s.V$, for all $s > t_V$;*

3. *H^* is bounded (or $\mathcal{T}^*$-bounded) if for any $\mathcal{T}^*$-neighborhood U of 0^* there is some $t_U > 0$ such that $H^* \subseteq s.U$, for all $s > t_U$;*

4. *H^* is weakly bounded (or $\sigma(X^*, X^{**})$-bounded) if for any $\sigma(X^*, X^{**})$-neighborhood V of 0^* there is some $t_V > 0$ such that $H^* \subseteq s.V$, for all $s > t_V$; and,*

5. *H^* is weak* bounded (or $\sigma(X^*, X)$-bounded) if for any $\sigma(X^*, X)$-neighborhood W of 0^* there is some $t_W > 0$ such that $H^* \subseteq s.W$, for all $s > t_W$.*

Proposition C.21. *Let $(X, \|.\|)$ be a normed linear space, and let $(X^*, \|.\|^*)$ be its dual space. Let H be any subset of X and H^* any subset of X^*. Then:*

1. *H is bounded if and only if H is weakly bounded; and,*

2. *If X is a Banach space, then H^* is bounded if and only if H^* is weak* bounded.*

Proof. See Megginson [210], Theorem 2.5.5 on p. 213, and Theorem 2.6.7 on p. 225. □

C.3.3 Convexity

A subset of a linear space is called *convex* when it is closed under convex combinations, that is:

Definition C.22. *Let X be a real vector space and let H be any subset of X. Then H is called convex when $\alpha.x + (1-\alpha).y \in H$, for all $x, y \in H$ and for all $\alpha \in (0,1)$.*

Clearly, any linear subspace of X is convex. The interesting properties of convex subsets of normed linear spaces stem form the following theorem:

Theorem C.23 (Mazur)**.** *Let $(X, \|.\|)$ be a normed linear space, let $(X^*, \|.\|^*)$ be its dual space, and let C be a convex subset of X. Then the norm-closure and weak-closure of C in X coincide. In particular, the following are equivalent:*

1. *C is norm-closed; and,*

2. *C is weakly closed.*

Proof. See Diestel [100], Theorem 1 on p. 11 and Corollary 4 on p. 12. □

For instance, the norm-closure and weak-closure of a linear subspace of a normed linear space coincide.

C.3.4 Compactness and Sequential Compactness

The main results pertaining to notions of compactness in normed linear spaces are undoubtedly the Banach-Alaoglu Theorem and the Eberlein-Šmulian Theorem, which we state below.

Theorem C.24 (Banach-Alaoglu)**.** *Let $(X, \|.\|)$ be a normed linear space and let $(X^*, \|.\|^*)$ be its dual space. Then $B_1[X^*]$ is weak*-compact. Thus, norm-bounded and weak*-closed subsets of X^* are weak*-compact.*

Proof. See Aliprantis and Border [3], Theorem 6.21 on p. 235. □

Theorem C.25 (Eberlein-Šmulian)**.** *A subset of a normed linear space is weakly compact if and only if it is weakly sequentially compact.*

Proof. See Diestel [100] p. 18. □

We then have the following result:

Corollary C.26. *Let $(X, \|.\|)$ be a linear space and let $(X^*, \|.\|^*)$ be its dual space. Let H be a subset of X and let H^* be a subset of X^*. Then:*

1. *H is weakly compact if and only if it is weakly sequentially compact; and,*

2. *H^* is weakly compact if and only if it is weakly sequentially compact.*

Appendix D

Duality in Spaces of Measures

The material presented in this Appendix can be found in references such as Aliprantis and Border [3], Billingsley [38], Cohn [82], Doob [106], Dunford and Schwartz [109], Halmos [161], Hewitt and Stromberg [166], Rao and Rao [238], or Saks [265]. We will also state a remarkable theorem by Maccheroni and Marinacci [194] (which is essentially due to Bartle, Dunford, and Schwartz [30], and Gänssler [132]).

D.1 Measures and Measurability

D.1.1 Measurable Functions

Definition D.1. *Let S an arbitrary nonempty set and Σ any σ-algebra of subsets of S. We call the pair (S, Σ) a measurable space.*

Definition D.2. *Let (S, Σ) be a measurable space and f a real-valued function on S. We say that the function $f : S \to \mathbb{R}$ is Σ-measurable, or simply measurable, if for any Borel set B in $\mathbb{R}$, $f^{-1}(B) \in \Sigma$, where $f^{-1}(B) := \{x \in S : f(x) \in B\}$ is called the preimage of B under f.*

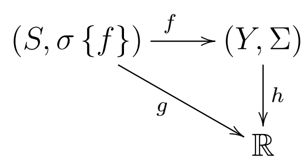
It is customary to refer to a real-valued Σ-measurable function on a measurable space (S, Σ) as a *random variable* on (S, Σ).

Definition D.3. *Let S be an arbitrary non-empty set and let (Y, Σ) be a measurable space, and let f be a mapping of S into Y. We define a collection $\sigma\{f\}$ of subsets of S as follows:*

$$\sigma\{f\} := \left\{f^{-1}(A) : A \in \Sigma\right\} \tag{D.1}$$

Then $\sigma\{f\}$ is a σ-algebra on S, and we refer to it as the σ-algebra on S generated by f.

Theorem D.4 (Doob)**.** *Let S be an arbitrary non-empty set and let (Y, Σ) be a measurable space. Let f be a mapping of S into Y, and denote by $\sigma\{f\}$ the σ-algebra on S generated by f. Let $g : S \to \mathbb{R}$. Then the function g is $\sigma\{f\}$-measurable if and only if there exists a Σ-measurable function $h : Y \to \mathbb{R}$ such that $g = h \circ f$.*



Proof. See Aliprantis and Border [3] (Theorem 4.41 on p. 147), Dellacherie and Meyer [94] (Theorem 18 on p. 12), or Doob [105] (p. 603). □

For a subset A of some set S, let $\mathbf{1}_A$ denote the indicator (characteristic) function of A.

Definition D.5. *A Σ-simple function on a measurable space (S, Σ) is a function $f : S \to \mathbb{R}$ of the form $f = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}$, for some $n \geq 1$, $\{\alpha_1, \ldots, \alpha_n\} \subset \mathbb{R}$ with $\alpha_i \neq \alpha_j$ if $i \neq j$, and a partition $\{A_i\}_{i=1}^{n}$ of S into elements of Σ.*

Hence, a Σ-simple function is in particular measurable.

Proposition D.6. *Let (S, Σ) be a measurable space, and let $f : S \to [0, +\infty]$ be a Σ-measurable function. Then there exists a sequence $\{h_n\}_{n \geq 1}$ of $[0, +\infty)$-valued Σ-simple function such that:*

1. *$h_n(x) \leq h_{n+1}(x)$, for all $n \geq 1$ and for all $x \in S$; and,*

2. *$\lim_n h_n(x) = f(x)$, for all $x \in S$.*

Proof. See Cohn [82] (Proposition 2.1.7 on p. 54), or Hewitt and Stromberg [166] (Theorem 11.35 on p. 159). □

D.1.2 Measures

Definition D.7. *Let (S, Σ) be a measurable space. Then:*

1. *A finitely additive measure on (S, Σ) is a set function $\phi : \Sigma \to [-\infty, +\infty]$ such that:*

   (a) *$\phi(\varnothing) = 0$; and*

   (b) *$\phi(A \cup B) = \phi(A) + \phi(B)$, for all $A, B \in \Sigma$ such that $A \cap B = \varnothing$.*

2. *A countably additive measure, or simply a measure, on (S, Σ) is a set function $\mu : \Sigma \to [-\infty, +\infty]$ such that:*

   (a) *$\mu(\varnothing) = 0$; and*

   (b) *$\mu\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mu(A_n)$, for any sequence $\{A_n\}_{n\geq 1}$ of pairwise disjoint elements of Σ.*

3. *A nonnegative finitely additive measure if it is both nonnegative and finitely additive; and,*

4. *A nonnegative countably additive measure if it is both nonnegative and countably additive.*

Note that if ϕ is a finitely additive measure then, by induction, for any collection $\{A_i\}_{i=1}^{n}$ of pairwise disjoint elements of Σ, $\phi(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \phi(A_i)$. Henceforth, we will refer to a countably additive measure as simply a measure.

Definition D.8. *Let (S, Σ) be a measurable space. Then:*

1. *A finitely additive measure λ on Σ is called bounded when $\sup_{A\in\Sigma} |\lambda(A)| < +\infty$;*

2. *A nonnegative finitely additive measure λ on Σ is called finite when $|\lambda(S)| < +\infty$;*

3. *A nonnegative finitely additive measure λ on Σ is called a finitely additive probability measure when $\lambda(S) = 1$; and,*

4. *A nonnegative countably additive measure λ on Σ is called a countably additive probability measure, or simply a probability measure, when $\lambda(S) = 1$.*

Definition D.9. *Let (S, Σ) be a measurable space, and let μ and ν be two finite nonnegative measures on (S, Σ). We say that:*

1. *μ is absolutely continuous with respect to ν, and we write $\mu << \nu$, if for all $A \in \Sigma$, $\nu(A) = 0 \Rightarrow \mu(A) = 0$;*

2. *μ and ν are equivalent or mutually absolutely continuous, and we write $\mu \sim \nu$, when $\mu << \nu$ and $\nu << \mu$; and,*

3. *μ and ν are mutually singular, and we write $\mu \perp \nu$ (or $\nu \perp \mu$), when there is some $B \in \Sigma$ such that $\mu(B) = \nu(S \setminus B) = 0$.*

Theorem D.10 (Lebesgue Decomposition)**.** *Let (S, Σ) be a measurable space and let μ and ν be two finite nonnegative measures on (S, Σ). Then there is a unique pair (ν_{ac}, ν_s) of finite nonnegative measures on (S, Σ) such that:*

1. *$\nu = \nu_{ac} + \nu_s$;*

2. *$\nu_s \perp \mu$; and,*

3. *$\nu_{ac} << \mu$.*

Proof. See Aliprantis and Border [3] (Theorem 10.61 on p. 401), Cohn [82] (Theorem 4.3.1 on p. 141), Hewitt and Stromberg [166] (Theorem 19.42 on p. 326), or Saks [265] (Theorem 14.6 on pp. 33-34). □

Definition D.11. *Let $(h_n)_{n\geq1}$ be a sequence of functional on an arbitrary nonempty space S, and let $A \subseteq S$. We say that the sequence $(h_n)_{n\geq1}$ converges uniformly on A to some functional ϕ on S if, for each $\varepsilon > 0$, there is some $N \in \mathbb{N}$ such that for all $n \geq 1$,*

$$n \geq N \Rightarrow |h_n(x) - \phi(x)| < \varepsilon, \ \forall\, x \in A$$

Theorem D.12 (Egoroff)**.** *Let (S, Σ) be a measurable space, let μ be a finite nonnegative measure on (S, Σ), and let $E \in \Sigma$ be such that $\mu(E) > 0$. Let $(f_n)_{n \geq 1}$ be a sequence of $\mathbb{R}$-valued Σ-measurable functions on S such that the sequence $(f_n(x))_{n \geq 1}$ converges to some $f(x)$, for all $x \in A$, for some $A \in \Sigma$, $A \subseteq E$, and $\mu(E \setminus A) = 0$. Then for any $\varepsilon > 0$, there is some $F \in \Sigma$, with $F \subseteq E$ and $\mu(E \setminus F) < \varepsilon$, such that the sequence $(f_n)_n$ converges to f uniformly on F.*

Proof. See Aliprantis and Border [3] (Theorem 10.38 on p. 389), Dunford and Schwartz [109] (Definition III.6.1 on p. 145 and Theorem III.6.12 on p. 149), Hewitt and Stromberg [166] (Theorem 11.32 on p. 158), or Saks [265] (Theorem 9.6 on p. 18). □

D.2 The Dual of the Space $B(\Sigma)$

D.2.1 The spaces $ba(\Sigma)$ and $ca(\Sigma)$

Definition D.13. *Let (S, Σ) be a measurable space and μ a finitely additive measure on (S, Σ). Define the nonnegative set function $|\mu|$ on Σ as follows: For each $E \in \Sigma$,*

$$|\mu|(E) := \sup \left\{ \sum_{i=1}^{n} |\mu(A_i)| : \{A_i\}_{i=1}^{n} \text{ is a partition of } E, A_i \in \Sigma \text{ for } i = 1, \ldots, n \right\}$$

We call $|\mu|$ the total variation of μ, and for each $E \in \Sigma$, we call $|\mu|(E)$ the total variation of μ on E.

Proposition D.14. *If (S, Σ) be a measurable space and μ is a finitely additive measure on (S, Σ), then the total variation $|\mu|$ of μ is a nonnegative finitely additive measure on (S, Σ).*

Note that when μ is also nonnegative $|\mu|(E) = \mu(E)$, for each $E \in \Sigma$.

Definition D.15. *Let (S, Σ) be a measurable space, and let μ and ν be two finitely additive measures on (S, Σ). We say that μ is absolutely continuous with respect to ν, and we write $\mu << \nu$, if*

$$\lim_{|\nu|(E) \to 0} \mu(E) = 0$$

Definition D.16. *Let (S, Σ) be a measurable space, and let μ and ν be two finite measures on (S, Σ). Then $\mu << \nu$ if and only if for all $A \in \Sigma$,*

$$|\nu|(A) = 0 \Rightarrow \mu(A) = 0$$

Definition D.17. *Let (S, Σ) be a measurable space and μ a finitely additive measure on (S, Σ). We say that μ is of bounded variation if $|\mu|(S) < +\infty$*

Proposition D.18. *Let (S, Σ) be a measurable space and μ a finitely additive measure on (S, Σ). Then*

$$\sup_{E \in \Sigma} |\mu(E)| \leq |\mu|(S) \leq 4 \sup_{E \in \Sigma} |\mu(E)|$$

Consequently, a finitely additive measure has bounded variation if and only if it is bounded.

Definition D.19. *For a measurable space (S, Σ), we denote by:*

1. *$ba(\Sigma)$ the space of all finitely additive measures on (S, Σ) having bounded variation;*
2. *$ba^{+}(\Sigma)$ the collection of nonnegative elements of $ba(\Sigma)$;*
3. *$ca(\Sigma)$ the collection of all countably additive elements of $ba(\Sigma)$;*
4. *$ca^{+}(\Sigma)$ the collection of nonnegative elements of $ca(\Sigma)$; and,*
5. *$ca_1^{+}(\Sigma)$ the collection of all probability measures on (S, Σ).*

Under the usual setwise operations of addition and scalar multiplication, the spaces $ba(\Sigma)$ and $ca(\Sigma)$ are real vector spaces, and $ca(\Sigma)$ is a linear subspace of $ba(\Sigma)$. They turn out to be much more than that:

Proposition D.20. *Let (S, Σ) be a measurable space and for each $\mu \in ba(\Sigma)$, let $\|\mu\|_v := |\mu|(S)$. Then:*

1. *$\|.\|_v$ is a norm on $ba(\Sigma)$;*
2. *The space $(ba(\Sigma), \|.\|_v)$ is a Banach space; and,*
3. *$ca(\Sigma)$ is a $\|.\|_v$-closed linear subspace of $ba(\Sigma)$. Hence, $ca(\Sigma)$ is $\|.\|_v$-complete, i.e. $(ca(\Sigma), \|.\|_v)$ is a Banach space.*

D.2.2 The Space $B(\Sigma)$ and its Dual

Definition D.21. *For a measurable space (S,Σ), denote by:*

1. *$B_0(\Sigma)$ the collection of $\mathbb{R}$-valued, Σ-measurable Σ-simple functions on S;*
2. *$B(\Sigma)$ the collection of all bounded Σ-measurable $\mathbb{R}$-valued functions on S; and,*
3. *$B^+(\Sigma)$ the collection of all nonnegative elements of $B(\Sigma)$.*

When endowed with the usual setwise operations of addition and scalar multiplication, both $B_0(\Sigma)$ and $B(\Sigma)$ are real vector spaces. Moreover, there is a natural and useful norm one can define on these spaces:

Definition D.22. *Let $f : S \to \mathbb{R}$ be a bounded function, and define*

$$\|f\|_s := \sup\{|f(s)| : s \in S\}$$

Proposition D.23. *The quantity $\|.\|_s$ is a norm on the space $B(\Sigma)$, called the supnorm, or the uniform norm. Moreover, the space $(B(\Sigma), \|.\|_s)$ is a Banach space.*

Proof. See Dunford and Schwartz [109], IV.2.12 on p. 240, and the first paragraph of p. 258. □

A natural issue to examine at this point is the characterization of the topological dual of the Banach space $(B(\Sigma), \|.\|_s)$. Fortunately, this is a rather classical problem.

Theorem D.24. *Let $B^*(\Sigma)$ denote the topological dual of the Banach space $(B(\Sigma), \|.\|_s)$. Then $B^*(\Sigma)$ is isometrically isomorphic to the space $ba(\Sigma)$, via the identity*

$$\Psi(f) = \int_S f\ d\mu \tag{D.2}$$

Therefore, for each $\Psi \in B^*(\Sigma)$ there is a unique $\mu \in ba(\Sigma)$ such that $\Psi(f) = \int_S f\ d\mu$, for each $f \in B(\Sigma)$. Moreover, for each $\mu \in ba(\Sigma)$ there is a unique $\Psi \in B^*(\Sigma)$ such that $\Psi(f) = \int_S f\ d\mu$, for each $f \in B(\Sigma)$.

Proof. See Diestel [100] (Theorem 7 on p. 77), Dunford and Schwartz [109] (Theorem IV.5.1 on p. 258), Fichtenholz and Kantorovich [126], or Hildebrandt [167]. □

This is a deep and powerful result which will allow us to endow $ba(\Sigma)$ with the weak* topology $\sigma(ba(\Sigma), B(\Sigma))$. We can also endow $ba(\Sigma)$ with the weak topology $\sigma(ba(\Sigma), ba^*(\Sigma))$. Finally, we can endow $ca(\Sigma)$ with the weak topology $\sigma(ca(\Sigma), ca^*(\Sigma))$, which coincides with the weak topology that $ca(\Sigma)$ inherits from the weak topology of $ba(\Sigma)$, by Proposition C.16. In particular, weak compactness in $ca(\Sigma)$ is equivalent to weak compactness in $ba(\Sigma)$.

D.3 Topological Properties of $ba(\Sigma)$ and $ca(\Sigma)$

Definition D.25. *Let (S,Σ) be a measurable space, let $\lambda \in ba(\Sigma)$, and let $M \subseteq ba(\Sigma)$. We say that M is uniformly absolutely continuous with respect to λ if, for every $\varepsilon > 0$, there is some $\delta > 0$ such that*

$$(A \in \Sigma \text{ and } |\lambda|(A) < \delta) \Rightarrow (|\nu(A)| < \epsilon, \text{ for each } \nu \in M)$$

In other words, the limit

$$\lim_{|\lambda|(A)\to 0} \nu(A) = 0$$

is uniform with respect to $\nu \in M$.

Theorem D.26 (Vitali-Hahn-Saks)**.** *Let (S,Σ) be a measurable space and $(\mu_n)_{n\geq 1}$ a sequence in $ca(\Sigma)$ such that, for each $A \in \Sigma$, the sequence $(\mu_n(A))_{n\geq 1}$ converges to some real number. Define the set function μ on Σ by*

$$\mu(A) := \lim_n \mu_n(A), \ \forall\ A \in \Sigma \tag{D.3}$$

Then:

1. *μ is a bounded measure on (S,Σ); and,*
2. *The collection $M := \{\mu_n, n \geq 1\}$ is uniformly absolutely continuous with respect to λ, for each $\lambda \in ca(\Sigma)$ that satisfies $\mu_n << \lambda$, $\forall\ n \in \mathbb{N}$.*

Proof. See Rao and Rao [238], Theorem 8.1.4 on pp. 204-205. □

Theorem D.27. *Let (S, Σ) be a measurable space, $M \subseteq ca(\Sigma)$, and $N \subseteq ba(\Sigma)$. Then:*

1. *M is weakly sequentially compact if and only if M is norm-bounded and there is some $\lambda \in ca^+(\Sigma)$ such that M is uniformly absolutely continuous with respect to λ; and,*

2. *N is weakly sequentially compact if and only if there is some $\phi \in ba^+(\Sigma)$ such that N is uniformly absolutely continuous with respect to ϕ;*

Proof. See Dunford and Schwartz [109], Theorem IV.9.2 on p. 306 and Theorem IV.9.12 on p. 314. □

Proposition D.28. *Let (S, Σ) be a measurable space, then:*

1. *The space $ca(\Sigma)$ is weakly complete; and,*

2. *The space $ba(\Sigma)$ is weakly complete.*

Proof. See Dunford and Schwartz [109], Theorem IV.9.4 on p. 308 and Theorem IV.9.9 on p. 311. □

Proposition D.29. *Let (S, Σ) be a measurable space. A sequence $(\mu_n)_{n\geq 1}$ in $ca(\Sigma)$ converges weakly to some $\mu \in ca(\Sigma)$ if and only if it is norm-bounded and the limit $\lim_n \mu_n(E)$ exists and equals $\mu(E)$, for each $E \in \Sigma$.*

Proposition D.30. *Let (S, Σ) be a measurable space. A sequence $(\mu_n)_{n\geq 1}$ in $ba(\Sigma)$ converges weakly to some $\mu \in ba(\Sigma)$ if and only if the sequence $(\mu_n)_{n\geq 1}$ converges to μ in the weak* topology.*

Although quite surprising at first sight, the following theorem by Maccheroni and Marinacci [194] (which is essentially due to Bartle, Dunford, and Schwartz [30], and Gänssler [132]) is remarkable for its extreme importance in practice. It can be seen as a complement to the Eberlein-Šmulian theorem (Theorem C.25), or as a Heine-Borel theorem for the space $ba(\Sigma)$.

Theorem D.31 (Maccheroni-Marinacci)**.** *Let* (S, Σ) *be a measurable space and* $M \subseteq ca(\Sigma)$*. Then the following are equivalent:*

1. *M is weak* closed and norm-bounded;*
2. *M is weak* compact;*
3. *M is weakly compact;*
4. *M is sequentially weak* compact; and,*
5. *M is sequentially weakly compact.*

Proof. See Maccheroni and Marinacci [194], Theorem 1 on p. 355. □

D.4 Duality in Spaces of Borel Probability Measures

In this section we will discus a special case of a much larger, richer and deeper theory of duality on topological measure spaces. We refer the reader to Bourbaki [59] and Kuratowski [188] for many definitions and further results. The reader may also want to consult Aliprantis and Border [3] (chap. 12, 14, and 15), Bogachev [43] (chap. 6, 7, and 8), and Cohn [82] (chap. 7 and 8). The basic duality result that we will discuss here is based on the *Riesz representation theorem*. We then apply it to the collection of probability measures on some appropriately defined σ-algebra of a topological space to characterize a notion of convergence for probability measures. First, however, we need some preliminary definitions and results.

D.4.1 Preliminaries and the Riesz Representation Theorem

Definition D.32. *Let* $(S, \mathcal{T})$ *be an arbitrary nonempty topological space. The Borel* σ*-algebra of* S *is the* σ*-algebra* $\mathcal{B}$ *generated by* $\mathcal{T}$*. Elements of* $\mathcal{B}$ *are called the Borel subsets of* S*.*

Proposition D.33. *Let $(S, \mathcal{T})$ be an arbitrary nonempty topological space and $\mathcal{B}$ its Borel σ-algebra. If a mapping $f : S \to \mathbb{R}$ is continuous, then it is Borel-measurable, i.e. $\mathcal{B}$-measurable.*

Definition D.34. *Let $(S, \mathcal{T})$ be an arbitrary nonempty topological space and $\mathcal{B}$ its Borel σ-algebra. We refer to any countably additive (resp. finitely additive) measure on the measurable space $(S, \mathcal{B})$ as a countably additive (resp. finitely additive) Borel measure on S. As usual, the term measure is reserved for countably additive measures.*

Definition D.35. *A nonnegative Borel measure μ (either finitely additive or countably additive) on a topological space $(S, \mathcal{T})$ with Borel σ-algebra $\mathcal{B}$ is called:*

1. *Outer regular if for each $A \in \mathcal{B}$, $\mu(A) = \inf\{\mu(U) : U \in \mathcal{T}, A \subseteq U\}$;*
2. *Inner regular if for each $A \in \mathcal{B}$, $\mu(A) = \sup\{\mu(F) : F \text{ is closed, } F \subseteq A\}$;*
3. *Normal if it is both inner and outer regular;*
4. *Tight if for each $A \in \mathcal{B}$, $\mu(U) = \sup\{\mu(K) : K \text{ is compact, } K \in \mathcal{B}, K \subseteq A\}$; and,*
5. *Regular if $\mu(K) < +\infty$, for each compact subset K of S in $\mathcal{B}$, and it is both outer regular and tight.*

Moreover, a bounded Borel measure μ (either finitely additive or countably additive) on the topological space $(S, \mathcal{T})$ is said to possess any of the above properties when its total variation $|\mu|$ does.

Note that if in the definition above the space $(S, \mathcal{T})$ is Hausdorff, so that every compact subset of S is closed, then the requirement that K be in $\mathcal{B}$ in the definition of tightness and regularity is redundant.

Proposition D.36. *A finite finitely additive nonnegative Borel measure is outer regular if and only if it is inner regular (and hence if and only if it is normal).*

Proposition D.37. *On a Hausdorff space, any tight finite finitely additive nonnegative Borel measure is also regular.*

Proposition D.38. *On a metrizable space, any finite finitely additive nonnegative Borel measure is normal.*

Recall the definition of a Polish space: it is a metric space which is complete and separable.

Proposition D.39. *Every bounded Borel measure on a Polish space is regular. In particular, any Borel probability measure on a Polish space is regular.*

This result can be generalized to more abstract spaces such as *Lusin spaces* and *Souslin spaces* which we define below.

Definition D.40. *A Lusin space is a Hausdorff space which is the image of a Polish space under a continuous bijection (one-to-one and onto). A Souslin space is a Hausdorff space which is the image of a Polish space under a continuous surjection (onto).*

Hence, any Polish space is a Lusin space, and any Lusin space is a Souslin space. Moreover, one can think of a Lusin space as a transformation of a Polish space, whereby the metric topology is replaced by some weaker Hausdorff topology.

Proposition D.41. *Every bounded Borel measure on a Souslin space is regular. In particular, any Borel probability measure on a Souslin space is regular.*

Definition D.42. *For any Hausdorff topological space $(S, \mathcal{T})$ with Borel σ-algebra $\mathcal{B}$, we define the following collections of set functions:*

1. $ba_r(S)$ *is the collection of all regular bounded finitely additive Borel measures on* $(S, \mathcal{B})$*;*
2. $ca_r(S)$ *is the collection of all countably additive elements of* $ba_r(S)$*;*
3. $ba_n(S)$ *is the collection of all normal bounded finitely additive Borel measures on* $(S, \mathcal{B})$*; and,*
4. $ca_n(S)$ *is the collection of all countably additive elements of* $ba_n(S)$*;*

All of the above collections of set functions are real vector spaces when endowed with usual mixing operations. Furthermore,

Proposition D.43. *If* $\|.\|_v$ *denotes the total variation norm, then the spaces* $(ba_r(S), \|.\|_v)$, $(ca_r(S), \|.\|_v)$, $(ba_n(S), \|.\|_v)$, *and* $(ca_n(S), \|.\|_v)$ *are Banach spaces.*

Definition D.44. *Let* $(S, \mathcal{T})$ *be a topological space. A function* $f : S \to \mathbb{R}$ *is said to be bounded if* $\sup\{|f(s)| : s \in S\} < +\infty$.

Definition D.45. *Let* $(S, \mathcal{T})$ *be a topological space. We denote by* $C_b(S)$ *the collection of all continuous and bounded* $\mathbb{R}$*-valued functions on* $(S, \mathcal{T})$.

With the usual mixing operations for functions, $C_b(S)$ is a real vector space. Moreover:

Proposition D.46. *Let* $(S, \mathcal{T})$ *be a topological space, and for each* $f \in C_b(S)$ *let* $\|f\|_s := \sup\{|f(s)| : s \in S\}$. *Then* $\|.\|_s : C_b(S) \to [0, +\infty)$ *is a norm on* $C_b(S)$, *and the space* $(C_b(S), \|.\|_s)$ *is a Banach space.*

We now come to this section's two main results:

Theorem D.47 (Dual of $C_b(S)$)**.** *Let* $(S, \mathcal{T})$ *be a Hausdorff normal topological space, and let* $C_b^*(S)$ *denote the topological dual of the space* $C_b(S)$. *Define the map* $T : ba_n(S) \to C_b^*(S)$ *as follows: For each* $\mu \in ba_n(S)$, *we have*

$$T(\mu)(f) := \int_S f \; d\mu, \;\; \forall \; f \in C_b(S) \tag{D.4}$$

Then the map T *is a surjective isometry. Hence, we can identify the Banach space* $C_b^*(S)$ *with the Banach space* $ba_n(S)$.

Theorem D.48 (Positive Functionals on $C_b(S)$)**.** *Let* $(S, \mathcal{T})$ *be a Hausdorff normal topological space, and let* T *be a positive linear functional on* $C_b(S)$. *Then there exists a unique finite nonnegative and finitely additive normal measure* μ *on the Borel* σ*-algebra* $\mathcal{B}$ *of* $(S, \mathcal{T})$ *such that:*

1. $T(f) := \int_S f \; d\mu$, *for each* $f \in C_b(S)$; *and,*

2. $\|\mu\|_v = \mu(S) = \|T\| = T(1)$, *where* 1 *denotes the constant function yielding* 1.

D.4.2 "Weak Convergence" of Borel Probability Measures

Any Borel probability measure on a Polish space (e.g. $\mathbb{R}$ with its usual metric) is regular and normal. Therefore, the collection $\mathcal{P}(\mathbb{R})$ of Borel probability measures on (the Borel σ-algebra of) $\mathbb{R}$ is a subset of the collection $ba_n(\mathbb{R})$ of all bounded normal finitely additive Borel measures on $\mathbb{R}$.

Moreover, as per Theorem D.47, the topological dual of the space of all bounded continuous $\mathbb{R}$-valued functions on a Hausdorff normal space S (e.g. $\mathbb{R}$ with its usual metric) can be identified with the Banach space $ba_n(S)$. Therefore, we can identify the topological dual $C_b^*(\mathbb{R})$ of $C_b(\mathbb{R})$ with the Banach space space $ba_n(\mathbb{R})$. Hence, we can endow the space $ba_n(\mathbb{R})$ with the weak* topology $\sigma(ba_n(\mathbb{R}), C_b(\mathbb{R}))$. Consequently, there is a weak* topology on $\mathcal{P}(\mathbb{R})$.

Definition D.49. *"Weak convergence" in $\mathcal{P}(\mathbb{R})$ refers to convergence in the weak* topology $\sigma(ba_n(\mathbb{R}), C_b(\mathbb{R}))$. That is, a net $\{\mu_\alpha\}_{\alpha\in\Lambda}$ of Borel probability measures on $\mathbb{R}$ converges "weakly" to some Borel probability measure μ on $\mathbb{R}$ if and only if the net $\left\{\int \phi \, d\mu_\alpha\right\}_{\alpha\in\Lambda}$ converges to $\int \phi \, d\mu$, for each continuous bounded real-valued function ϕ on $\mathbb{R}$.*

This terminology is unfortunate, for "weak convergence" in the language of probability theory is nothing but weak* convergence in the language of functional analysis. However, weak convergence in the language of functional analysis refers to convergence in the topology $\sigma(ba_n(\mathbb{R}), ba_n^*(\mathbb{R}))$ on $ba_n(\mathbb{R})$. Nevertheless, this rather bad terminology is so engrained in the literature that we see no point in challenging it here. However we will always write "weak convergence" (between quotation marks) to mean weak* convergence.

Definition D.50. *Let $\{X_\alpha\}_{\alpha\in\Lambda}$ be a net of random variables on a measurable space (S, Σ), and let μ be a probability measure on (S, Σ). We say that the net $\{X_\alpha\}_{\alpha\in\Lambda}$ converges in distribution to some random variable Y on (S, Σ) when the net $\{\mu \circ X_\alpha^{-1}\}_{\alpha\in\Lambda}$ of Borel probability measures on $\mathbb{R}$ converges "weakly" to the Borel probability measure $\mu \circ Y^{-1}$ on $\mathbb{R}$.*

Definition D.51. *For each random variable X on a probability space (S, Σ, P), we define the distribution function F_X of X as the function $F_X : \mathbb{R} \to [0, 1]$ given by*

$$F_X(t) := P(\{s \in S : X(s) \le t\}) = P \circ X^{-1}((-\infty, t]) \tag{D.5}$$

Similarly, for each Borel probability measure μ on $\mathbb{R}$, we define the distribution function F_μ of μ as the function $F_\mu : \mathbb{R} \to [0,1]$ given by

$$F_\mu(t) := \mu((-\infty, t]) \tag{D.6}$$

Proposition D.52. *Let $\{\mu_\alpha\}_{\alpha\in\Lambda}$ be a net of Borel probability measures on $\mathbb{R}$. Then $\{\mu_\alpha\}_{\alpha\in\Lambda}$ converges "weakly" to some Borel probability measure ν on $\mathbb{R}$ when the net $\{F_{\mu_\alpha}\}_{\alpha\in\Lambda}$ converges to F_ν at the points of continuity of F_ν.*

We can even strengthen this result further, but we first need some definitions:

Definition D.53. *Let $(S, \mathcal{T})$ be a topological space and A an arbitrary subset of S. Let $\overline{A}$ denote the closure of A and let A° denote the interior of A. We define the boundary of A, denoted by ∂A, as the set $\partial A = \overline{A} \setminus A^\circ$.*

Note that the boundary of any set is closed. In particular, it is a Borel set.

Theorem D.54 (Portmanteau)**.** *Let $(S, \mathcal{T})$ be a metrizable topological space, let $\mathcal{P}(S)$ denote the collection of all Borel probability measures on S, and let $C_b(S)$ denote the collection of all continuous bounded $\mathbb{R}$-valued functions on S. For a net $\{\mu_\alpha\}_{\alpha\in\Lambda}$ in $\mathcal{P}(S)$ and for some $\mu \in \mathcal{P}(S)$, the following are equivalent:*

1. *The net $\{\mu_\alpha\}_{\alpha\in\Lambda}$ converges to μ "weakly";*
2. *The net $\left\{\int_S f \, d\mu_\alpha\right\}_{\alpha\in\Lambda}$ converges to $\int_S f \, d\mu$, for each $f \in C_b(S)$; and,*
3. *The net $\{\mu_\alpha(B)\}_{\alpha\in\Lambda}$ converges to $\mu(B)$, for each Borel set B such that $\mu(\partial B) = 0$.*

Proof. See e.g. Billingsley [38], Theorem 2.1 on p. 16. □

Appendix E

Duality in L_p Spaces

All the material presented in this Appendix is classical. We refer the reader to any textbook on measure theory for a treatment of integration, the duality between L_p spaces, and the Radon-Nikodým theorem[1]. Many of the topics discussed below can be found in Prof. Andrew Heunis' lecture notes for the course STAT902. Classical references include Aliprantis and Border [3], Ash [27], Billingsley [38], Cohn [82], Dudley [108], Doob [106], Halmos [161], Hewitt and Stromberg [166], or Saks [265]. For the specific problem of characterizing the dual of an L_p space, we recommend the beautiful monograph by Gretsky [158].

E.1 Integration

In the following, let (S, Σ) be a measurable space and μ a countably additive nonnegative and finite measure on (S, Σ). We call the triple (S, Σ, μ) a *finite measure space*. If in addition $\mu(S) = 1$, then (S, Σ, μ) will be called a *probability space*.

Definition E.1. *Let $f : S \to \mathbb{R}$ be a Σ-simple function f on S of the form $f = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}$, for some $n \geq 1$, $\{\alpha_1, \ldots, \alpha_n\} \subset \mathbb{R}$ with $\alpha_i \neq \alpha_j$ if $i \neq j$, and a partition $\{A_i\}_{i=1}^{n}$ of S into elements of Σ. We define the integral $\int_S f \; d\mu$ of a f with respect to μ as*

$$\int_S f \; d\mu := \sum_{i=1}^{n} \alpha_i \mu(A_i) \tag{E.1}$$

[1] For a different approach altogether to the theory of measure and integration, we refer the reader to Bourbaki [53], [54], [55], [56], and [57], or Schwartz [277], for instance.

Definition E.2. *Let $\mathcal{S}_+$ denote the collection of all nonnegative, Σ-simple functions $f : S \to [0, +\infty)$.*

Definition E.3. *For any Σ-measurable function $f : S \to [0, +\infty]$, we define the integral $\int_S f \; d\mu$ of f with respect to μ as*

$$\int_S f \; d\mu := \sup \left\{ \int_S g \; d\nu : g \in \mathcal{S}_+ \text{ and } g(x) \leq f(x), \; \forall \; x \in S \right\} \tag{E.2}$$

Definition E.4. *For any function $f : S \to [-\infty, +\infty]$, define the functions $f^+ : S \to [0, +\infty]$ and $f^- : S \to [0, +\infty]$ by $f^+ = \max\{0, f\}$ and $f^- = (-f)^+$. Then f^+ (resp. f^-) is called the nonnegative part of f (resp. the nonpositive part of f), and $f = f^+ - f^-$.*

Definition E.5. *A Σ-measurable mapping $f : S \to [-\infty, +\infty]$ is called μ-integrable if $\int_S f^+ \; d\mu < +\infty$ and $\int_S f^- \; d\mu < +\infty$.*

Proposition E.6. *Let $f : S \to [-\infty, +\infty]$ be a Σ-measurable mapping, and define $|f| := f^+ + f^-$. Then f is μ-integrable if and only if $\int_S |f| \; d\mu < +\infty$.*

Definition E.7. *For any Σ-measurable and μ-integrable function $f : S \to [-\infty, +\infty]$, we define the integral $\int_S f \; d\mu$ of f with respect to μ as*

$$\int_S f \; d\mu := \int_S f^+ \; d\mu - \int_S f^- \; d\mu \tag{E.3}$$

Theorem E.8 (Monotone Convergence)**.** *Let $f, f_n : S \to [0, +\infty]$ be Σ-measurable functions on S, for each $n \geq 1$. Suppose that:*

1. *$f_n(x) \leq f_{n+1}(x)$, for each $n \geq 1$ and for μ-a.a. $x \in S$; and,*
2. *$\lim_n f_n(x) = f(x)$, for μ-a.a. $x \in S$.*

Then $\lim_n \int_S f_n \; d\mu = \int_S f \; d\mu$.

Theorem E.9 (Dominated Convergence)**.** *Let $g : S \to [0, +\infty]$ be a Σ-measurable and μ-integrable function on S, and let $f, f_n : S \to [-\infty, +\infty]$ be Σ-measurable functions on S, for each $n \geq 1$. Suppose that:*

1. $\lim_n f_n(x) = f(x)$, *for μ-a.a. $x \in S$; and,*

2. $|f_n(x)| \leq g(x)$, *for each $n \geq 1$ and for μ-a.a. $x \in S$.*

Then f and f_n are μ-integrable, for each $n \geq 1$, and $\lim_n \int_S f_n \, d\mu = \int_S f \, d\mu$.

Proof. See Aliprantis and Border [3] (Theorem 11.21 on p. 415) or Cohn [82] (Theorem 2.4.4. on p. 72). □

Definition E.10. *A measure μ on a measurable space (S, Σ) is called σ-finite if S is a countable union of Σ-measurable sets each having finite measure under μ.*

Theorem E.11 (Change of Variable)**.** *Let (X, Σ_X) and (Y, Σ_Y) be two measurable spaces, and let $T : (X, \Sigma_X) \to (Y, \Sigma_Y)$ be measurable, in the sense that $T^{-1}(A) \in \Sigma_X$ for each $A \in \Sigma_Y$. Let μ be a nonnegative (countably additive) measure on (X, Σ_X) and let $\nu := \mu \circ T^{-1}$. Then ν is a measure on (Y, Σ_Y), called the image measure of ν under T.*

Moreover, for each function $f : Y \to \mathbb{R}$, we have:

1. *If f is ν-integrable, then $f \circ T$ is μ-integrable and $\int_Y f \, d\nu = \int_X f \circ T \, d\mu$;*

2. *If ν is σ-finite, f is ν-measurable, and $f \circ T$ is μ-integrable, then f is ν-integrable and $\int_Y f \, d\nu = \int_X f \circ T \, d\mu$.*

Proof. See Aliprantis and Border [3] (Theorem 13.46 on p. 484), Dunford and Schwartz [109] (Lemma III.10.8 on p. 182), or Resnick [240] (Theorem 5.5.1 on pp. 135-136). □

E.2 $\mathcal{L}_p$ and L_p Spaces

In the following, let (S, Σ) be a measurable space and μ a countably additive nonnegative and finite measure on (S, Σ). We call the triple (S, Σ, μ) a *finite measure space.* If in addition $\mu(S) = 1$, then (S, Σ, μ) will be called a *probability space.*

Definition E.12. *For each $p \in [1, +\infty)$ we denote by $\mathcal{L}_p(S, \Sigma, \mu)$ the collection of all Σ-measurable functions f on S such that $|f|^p$ is μ-integrable. In particular, $\mathcal{L}_1(S, \Sigma, \mu)$ is the collection of all μ-integrable functionals on S.*

Definition E.13. *For each $p \in [1, +\infty)$, define the functional $\|.\|_p : \mathcal{L}_p(S, \Sigma, \mu) \to [0, +\infty)$ as follows: For each $f \in \mathcal{L}_p(S, \Sigma, \mu)$, let*

$$\|f\|_p := \left[\int_S |f|^p \; d\mu\right]^{1/p} \tag{E.4}$$

Proposition E.14. *For each $p \in [1, +\infty)$, define the binary relation $\sim$ on $\mathcal{L}_p(S, \Sigma, \mu)$ as follows: For each $f, g \in \mathcal{L}_p(S, \Sigma, \mu)$,*

$$f \sim g \Leftrightarrow f = g, \mu\text{-a.s.} \tag{E.5}$$

Then $\sim$ is an equivalence relation on $\mathcal{L}_p(S, \Sigma, \mu)$.

Definition E.15. *For each $p \in [1, +\infty)$ and for each $f \in \mathcal{L}_p(S, \Sigma, \mu)$, let $[f]$ denote the equivalence class of the function f for the equivalence relation $\sim$ of μ-a.s. equality defined above. We define $L_p(S, \Sigma, \mu)$ as the collection of all equivalence classes of functions in $\mathcal{L}_p(S, \Sigma, \mu)$:*

$$L_p(S, \Sigma, \mu) := \{[f] : f \in \mathcal{L}_p(S, \Sigma, \mu)\} \tag{E.6}$$

Moreover, we define the functional $\|.\|_p : L_p(S, \Sigma, \mu) \to [0, +\infty)$ by

$$\|\,[f]\,\|_p := \|f\|_p, \;\; \forall \; f \in \mathcal{L}_p(S, \Sigma, \mu) \tag{E.7}$$

Note that for each $\alpha, \beta \in \mathbb{R}$, for each $f, g \in \mathcal{L}_p(S, \Sigma, \mu)$, for each $f' \in [f]$, and for each $g' \in [g]$, we have $[\alpha.f + \beta.g] = [\alpha.f' + \beta.g']$. Consequently, we can define vector addition and scalar multiplication in $L_p(S, \Sigma, \mu)$ as follows:

$$\alpha.[f] + \beta.[g] := [\alpha.f' + \beta.g']$$

This makes $L_p(S, \Sigma, \mu)$ a real vector space with zero vector $[0]$.

Proposition E.16. *For each $p \in [1, +\infty)$, the functional $\|.\|_p$ hence defined is a norm on $L_p(S, \Sigma, \mu)$, called the L_p-norm, and $(L_p(S, \Sigma, \mu), \|.\|_p)$ is Banach space.*

It is customary to abuse the notation and identify $L_p(S, \Sigma, \mu)$ with $\mathcal{L}_p(S, \Sigma, \mu)$, and we shall follow the same abuse of notation. Now, in order to define the space $\mathcal{L}_\infty(S, \Sigma, \mu)$, and hence the space $L_\infty(S, \Sigma, \mu)$, we need the following definition:

Definition E.17. *A functional $f : S \to \mathbb{R}$ is called μ-essentially bounded if there is some $M \in [0, +\infty)$ such that the set $\{x \in S : |f(x)| > M\}$ is μ-null, that is, $|f| \leq M, \mu$-a.s.*

We now turn to the definition of the spaces $\mathcal{L}_\infty(S, \Sigma, \mu)$ and $L_\infty(S, \Sigma, \mu)$.

Definition E.18. *We define the space $\mathcal{L}_\infty(S, \Sigma, \mu)$ as the collection of all μ-essentially bounded functions on S, and we define the space $L_\infty(S, \Sigma, \mu)$ as the collection of all equivalence classes of functions in $\mathcal{L}_\infty(S, \Sigma, \mu)$ under the equivalence relation of μ-a.s. equality.*

Proposition E.19. *Define the functional $\|.\|_\infty : \mathcal{L}_\infty(S, \Sigma, \mu) \to [0, +\infty)$ as follows: For each $f \in \mathcal{L}_\infty(S, \Sigma, \mu)$,*

$$\|f\|_\infty := \inf\{M > 0 : |f(x)| \leq M \text{ for } \mu\text{-a.a. } x \in S\} \tag{E.8}$$

Moreover, define the functional $\|.\|_\infty : L_\infty(S, \Sigma, \mu) \to [0, +\infty)$ as follows: For each $f \in \mathcal{L}_\infty(S, \Sigma, \mu)$,

$$\| [f] \|_\infty := \|f\|_\infty \tag{E.9}$$

Then $\|.\|_\infty$ hence defined is a norm on $L_\infty(S, \Sigma, \mu)$, called the L_∞-norm or the essential supnorm, and $(L_\infty(S, \Sigma, \mu), \|.\|_\infty)$ is Banach space.

Proposition E.20. *For each $p \in [1, +\infty]$, the collection of all simple functions in $\mathcal{L}_p(S, \Sigma, \mu)$ is norm-dense in $\mathcal{L}_p(S, \Sigma, \mu)$, and hence it determines a norm-dense in $L_p(S, \Sigma, \mu)$.*

Proposition E.21. *For each $1 \leq r < s \leq +\infty$, we have that:*

1. $\mathcal{L}_s(S,\Sigma,\mu) \subset \mathcal{L}_r(S,\Sigma,\mu)$; and,

2. The identity map of $L_s(S,\Sigma,\mu)$ into $L_r(S,\Sigma,\mu)$ is norm-continuous.

E.3 Duality

The first main result that we will state here is the Radon-Nikodým theorem, which is essentially a duality result for the space L_1, and states that $L_1(S,\Sigma,\mu)$ is isometrically isomorphic to the collection of all countably additive measures on (S,Σ) of bounded variation that are absolutely continuous with respect to μ. The second main result concerns duality between L_p spaces themselves.

Theorem E.22 (Radon-Nikodým)**.** *Let μ be a σ-finite nonnegative (countably additive) measure on a measurable space (S,Σ), and let ν be a bounded countably additive measure on (S,Σ) such that $\nu << \mu$. Then the following results hold:*

1. *There exists a μ-a.s. unique function $f \in L_1(S,\Sigma,\mu)$ such that $\nu(E) = \int_E f\, d\nu$, $\forall\, E \in \Sigma$; and,*

2. $|\nu|(S) = \|\nu\|_v = \|f\|_1$.

Proof. See Aliprantis and Border [3] (Theorem 13.18 on p. 470), Cohn [82] (Theorem 4.2.2 on p. 132), Dunford and Schwartz [109] (Theorem III.10.2 on p. 176), Hewitt and Stromberg [166] (Theorem 19.23 on p. 315), or Saks [265] (Theorem 14.11 on p. 36). □

The function f given by the Radon-Nikodým theorem is called the *Radon-Nikodým derivative of ν with respect to μ* and denoted by $d\nu/d\mu$. Moreover,

Proposition E.23. *If $\mu \sim \nu$, then $d\nu/d\mu > 0$ and $d\mu/d\nu = [d\nu/d\mu]^{-1}$. Moreover, if $d\nu/d\mu > 0$ then $\mu \sim \nu$.*

Proof. See Aliprantis and Border [3] (Corollary 13.24 on p. 473) or Bogachev [42] (p. 179). □

Proposition E.24. *Let μ be a σ-finite nonnegative measure on a measurable space (S, Σ), let ν be a bounded countably additive measure on (S, Σ) such that $\nu << \mu$, and let $g = d\nu / d\mu$ be the Radon-Nikodým derivative of ν with respect to μ. If f is a ν-integrable function on S then the function fg on S is μ-integrable and $\int_S f \; d\nu = \int_S fg \; d\mu$.*

Proof. See Aliprantis and Border [3] (Theorem 13.23 on p. 472). □

In the language of duality between normed linear spaces, one can restate the Radon-Nikodým theorem as follows:

Corollary E.25. *Let (S, Σ) be a measurable space and let μ be a σ-finite nonnegative measure on (S, Σ). Denote by $ca\,(S, \Sigma, \mu)$ the collection of all countably additive measures on (S, Σ) of bounded variation that are absolutely continuous with respect to μ. Then there is an isometric isomorphism between the space $L_1\,(S, \Sigma, \mu)$ and the space $ca\,(S, \Sigma, \mu)$, via the duality $\nu\,(E) = \int_E f \; d\mu$.*

Proof. See Aliprantis and Border [3] (Theorem 13.19 on p. 470) or Dunford and Schwartz [109] (p. 306). □

The second main duality result for L_p spaces is a duality between these spaces themselves. This result is essentially due to Riesz [244] in the context of the Lebesgue measure on $[0, 1]$, and hence we will follow Aliprantis and Border [3] in attributing this result to Riesz.

Lemma E.26. *Let μ be a σ-finite nonnegative measure on a measurable space (S, Σ). Let $p \in (1, +\infty)$ and let q be such that $1/p + 1/q = 1$. If $f \in L_p\,(S, \Sigma, \mu)$ and $g \in L_q\,(S, \Sigma, \mu)$, then $fg \in L_1\,(S, \Sigma, \mu)$. Moreover, If $h \in L_1\,(S, \Sigma, \mu)$ and $m \in L_\infty\,(S, \Sigma, \mu)$, then $hm \in L_1\,(S, \Sigma, \mu)$.*

Theorem E.27 (Riesz)**.** *Let μ be a σ-finite nonnegative measure on a measurable space (S, Σ). If $p \in (1, +\infty)$ and if q is such that $1/p + 1/q = 1$, then there is an isometric isomorphism between the space $L_q\,(S, \Sigma, \mu)$ and the space $L_p^*\,(S, \Sigma, \mu)$ (the topological dual of $L_p\,(S, \Sigma, \mu)$). Moreover, the duality between these two Banach spaces is given by*

$$T\,(f) = \int_S gf \; d\mu, \;\; \forall \; f \in L_p\,(S, \Sigma, \mu) \tag{E.10}$$

Proof. See Aliprantis and Border [3] (Theorem 13.26 on p. 473) or Dunford and Schwartz [109] (Theorem IV.8.1 on p. 286). □

Consequently, the norm-dual of the space $L_p(S,\Sigma,\mu)$ can be identified with the space $L_q(S,\Sigma,\mu)$. This then allows us to endow $L_q(S,\Sigma,\mu)$ with the weak* topology

$$\sigma\left(L_q(S,\Sigma,\mu), L_p(S,\Sigma,\mu)\right)$$

Corollary E.28. *Let μ be a σ-finite nonnegative measure on a measurable space (S,Σ). If $p\in(1,+\infty)$, then:*

1. *The space $L_p(S,\Sigma,\mu)$ is weakly complete; and,*

2. *A subset of $L_p(S,\Sigma,\mu)$ is weakly sequentially compact if and only if it is norm-bounded.*

Proof. See Dunford and Schwartz [109], Corollary IV.8.3 on p. 289 and Corollary IV.8.4 on p. 289. □

Theorem E.29 (Riesz)**.** *Let μ be a σ-finite nonnegative measure on a measurable space (S,Σ). Then there is an isometric isomorphism between the space $L_\infty(S,\Sigma,\mu)$ and the space $L_1^*(S,\Sigma,\mu)$ (the topological dual of $L_1(S,\Sigma,\mu)$). Moreover, the duality between these two Banach spaces is given by*

$$T(f)=\int_S gf\ d\mu,\ \forall\ f\in L_1(S,\Sigma,\mu) \tag{E.11}$$

Proof. See Aliprantis and Border [3] (Theorem 13.28 on p. 473) or Dunford and Schwartz [109] (Theorem IV.8.5 on p. 289). □

Consequently, the norm-dual of the space $L_1(S,\Sigma,\mu)$ can be identified with the space $L_\infty(S,\Sigma,\mu)$. This then allows us to endow $L_\infty(S,\Sigma,\mu)$ with the weak* topology

$$\sigma\left(L_\infty(S,\Sigma,\mu), L_1(S,\Sigma,\mu)\right)$$

Theorem E.30. *Let μ be a σ-finite nonnegative measure on a measurable space (S,Σ). Then the space $L_1(S,\Sigma,\mu)$ is weakly complete.*

Proof. See Dunford and Schwartz [109], Theorem IV.8.36 on p. 290. □

Appendix F

Convexity and the Choquet-Bishop-De Leeuw Theorem

Of particular interest in both linear and convex analysis are convex compact sets, essentially because of the Krein-Milmam Theorem and its generalization, the Choquet-Bishop-de Leeuw Theorem, which give a "geometric" characterization of convex sets. The theory presented in this Appendix is just a small sample of a much richer theory of convex analysis on locally convex Hausdorff topological vector spaces. All material presented here may be found in Aliprantis and Border [3] (Chapter 7), Bourbaki [55] (chapter IV, section 7), Diestel [100] (Chapter IX), Diestel and Uhl [102] (Chapter VII), Dunford and Schwartz [109] (sections V.1 and V.8), Kelley and Namioka [181] (section 4.15), Megginson [210] (sections 2.10 and 2.11), Schaefer [270] (section 2.10), and, most importantly, Phelps [226].

For more "geometric properties" in linear spaces, we refer the reader, *inter alia*, to Andrews [18], Bishop and De Leeuw [39], Bishop and Phelps [40], Bourgain [60], [61], [62], and [63], Bourgin [64] and [66], Collier [83], Davis and Phelps [89], Diestel [98] and [99], Diestel, Ruess and Schachermayer [101], Edgar [110], Farmaki [124], Figiel, Ghoussoub and Johnson [127], Fonf and Lindenstrauss [130], Ghoussoub et al. [142], Ghoussoub and Maurey [143], Ghoussoub and E. Saab [144], Ghoussoub and P. Saab [145], Ghoussoub and Talagrand [146], Gilliam [152], Huff [170], Huff and Morris [171], Lin et al. 1988 [189], Matsuda [208], Maynard [209], Phelps [225], Randrianantoanina and E. Saab [235], [236], and [237], Riddle [241] and [242], Riddle, Saab and Uhl [243], Rosenthal and Wessel [246], Ruess and Stegall [249], E. Saab [250], [252], [253], [254], [251], [255], [256], and [257], E. Saab and P. Saab [258] and [259], P. Saab [260], [261], [262], and [263], Schachermayer [268] and [269], Stegall [286], and Talagrand [289].

F.1 Preliminaries and Basic Results

Definition F.1. *A subset C of a real vector space S is said to be convex if it is closed under convex combinations, that is, if for any $x, y \in C$, and for any $\alpha \in (0.1)$,*

$$\alpha x + (1 - \alpha)\, y \in C \tag{F.1}$$

Note that by a simple induction argument, $C \subseteq S$ is convex if and only if for any $\{c_1, c_2, \ldots, c_n\} \subseteq C$ and for any $\{\alpha_1, \alpha_2, \ldots, \alpha_n\} \subset \mathbb{R}^+$ such that $\sum_{i=1}^n \alpha_i = 1$, it follows that $\sum_{i=1}^n \alpha_i c_i \in C$.

Definition F.2. *Let S be a real vector space and $\{x_1, x_2, \ldots, x_n\} \subset S$. If $\{\alpha_1, \alpha_2, \ldots, \alpha_n\} \subset \mathbb{R}^+$ are such that $\sum_{i=1}^n \alpha_i = 1$, then the linear combination $\sum_{i=1}^n \alpha_i x_i$ is called a convex combination of the points $x_1, x_2, \ldots, x_n$.*

Definition F.3. *A real topological vector space, or real topological linear space, is a real vector space S with a topology $\mathcal{T}$ such that:*

1. *The mapping $(x, y) \mapsto x + y$ of $S \times S$ into S is continuous; and,*

2. *The mapping $(\alpha, x) \mapsto \alpha.x$ of $\mathbb{R} \times S$ into S is also continuous.*

We then say that the topology $\mathcal{T}$ is a linear topology on S.

Proposition F.4. *If $\{(S_\alpha, \mathcal{T}_\alpha)\}_{\alpha \in \Gamma}$ is an arbitrary family of topological vector spaces then the Cartesian product $S = \prod_{\alpha \in \Gamma} S_\alpha$ with the product topology $\mathcal{T} = \prod_{\alpha \in \Gamma} \mathcal{T}_\alpha$ is also a topological vector space.*

If $(S, \mathcal{T})$ is a topological vector space, then an immediate consequence of the linearity of the topology $\mathcal{T}$ is that $\mathcal{T}$ is *translation invariant*, in the sense that $V \in \mathcal{T}$ if and only if $x_0 + V \in \mathcal{T}$, for each $x_0 \in S$, where $x_0 + V := \{y + x_0 : y \in V\}$. Therefore, a neighborhood base at zero[1] of the topology $\mathcal{T}$ determines a neighborhood at each point of S, by translation.

[1] Zero refers to the (unique) zero vector for the operation of addition defined on the vector space S.

Definition F.5. *A topological vector space $(S, \mathcal{T})$ is said to be locally convex if it has a neighborhood base at zero consisting of convex sets, that is, if every neighborhood of zero includes a convex neighborhood of zero.*

For instance, any normed space is a locally convex and Hausdorff topological vector space.

Proposition F.6. *Let $(S, \mathcal{T})$ be a topological vector space over the field $\mathbb{R}$, and let $\{C_\alpha\}_{\alpha \in \Gamma}$ be an arbitrary collection of convex subsets of S. Then the following results hold:*

1. *For any $\alpha, \beta \in \Gamma$, the set $C_\alpha + C_\beta$ is a convex subset of S;*
2. *For any $r \in \mathbb{R}$, the set $r.C$ is a convex subset of S;*
3. *The arbitrary intersection $\bigcap_{\alpha \in \Gamma} C_\alpha$ is itself a convex subset of S; and,*
4. *For any $\alpha \in \Gamma$, the interior C° of C and the closure $\overline{C}$ of C are both convex subsets of S.*

Definition F.7. *Let A be any nonempty subset of a real vector space S. We define the convex hull of A, denoted by $co\,(A)$ as the collection of all convex combinations of elements of A. In other words,*

$$co\,(A) = \left\{ \sum_{i=1}^{n} \alpha_i x_i : n \geq 1, x_i \in A \text{ and } \alpha_i \geq 0, \forall i = 1, \ldots, n, \text{ and } \sum_{i=1}^{n} \alpha_i = 1 \right\} \tag{F.2}$$

The convex hull of A is the smallest convex set containing A. It is hence the intersection of all convex sets containing A.

Definition F.8. *Let A be any nonempty subset of a topological vector space $(S, \mathcal{T})$. We define the closed convex hull of A, denotes by $\overline{co}\,(A)$, as the smallest closed convex set containing A.*

Proposition F.9. *Let A be any nonempty subset of a topological vector space $(S, \mathcal{T})$. Then the closed convex hull of A coincides with the closure of the convex hull of A, that is,*

$$\overline{co}\,(A) = \overline{co\,(A)} \tag{F.3}$$

Proposition F.10. *Let A be any nonempty compact subset of a Banach space S. Then $\overline{co}\,(A)$ is compact.*

F.2 Support Points

Definition F.11. *Let A be an arbitrary nonempty subset of a topological vector space $(S, \mathcal{T})$ with dual space S^*, and let $\phi \in S^*$ be nonzero. We say that a point $x \in A$ is a support point of A, and that ϕ supports A at x, if ϕ attains its maximum or its minimum on A at the point x.*

Theorem F.12 (Bishop-Phelps)**.** *Let C be a norm-closed convex subset of a Banach space S. The following results hold:*

1. *The collection of support points of C is norm-dense in the boundary ∂C of C; and,*

2. *If C is also norm-bounded, then the collection of bounded linear functionals on S that support C is norm-dense in the topological dual S^* of S.*

Proof. See Aliprantis and Border [3] (Theorem 7.43 on p. 284), Bishop and Phelps [40], Diestel and Uhl [102] (Theorem 4 on p. 189), and Megginson [210] (Theorem 2.11.9 on p. 275 and Theorem 2.11.13 on p. 278). □

The Bishop-Phelps Theorem is a fundamental result that led to a very active area of research in convex analysis and in optimization theory. Notable subsequent contributions include Bollobas [45], Bourgain [61], Lindenstrauss [191], Partington [222], Schachermayer [267], and Stegall [285].

F.3 Extreme Points

Definition F.13. *Let C be an arbitrary nonempty subset of a real vector space S. We say that a nonempty subset E of C is an extreme subset of C if no element of E can be written as a convex combination of elements of $C \setminus E$. That is, if $x \in E$ can be written as $x = \alpha y + (1-\alpha)\,z$, for some $y, z \in C$ and some $\alpha \in (0,1)$, then $y, z \in E$.*

Definition F.14. *Let C be an arbitrary nonempty subset of a real vector space S. We say that a point $x \in C$ is an extreme point of C if the singleton $\{x\}$ is an extreme subset of C. That is, $x \in C$ is an extreme point of C if it cannot be written as a (strict) convex combination of distinct points in C.*

For a given subset C of a real vector space S, we will denote by $\mathcal{E}(C)$ the collection of all extreme points of C. Note that $\mathcal{E}(C)$ is itself an extreme subset of C if and only if it is nonempty.

Lemma F.15. *Let C be a nonempty convex subset of some real vector space S, and fix $x \in C$. Then $x \in \mathcal{E}(C)$ if and only if $C \setminus \{x\}$ is a convex set.*

Proposition F.16. *Let $(S, \mathcal{T})$ be a locally convex and Hausdorff topological vector space, and let C be a nonempty subset of S. Then:*

1. *Every compact extreme subset of C contains an extreme point of C; and,*
2. *If C is also compact then C has at least one extreme point.*

Theorem F.17 (Krein-Milman). *Let $(S, \mathcal{T})$ be a locally convex and Hausdorff topological vector space, and let C be a nonempty subset of S. If C is convex and compact then C is the closed convex hull of the set of its extreme points.*

Proof. See Aliprantis and Border [3] (Theorem 7.68 on p. 297), Diestel [100] (p. 148), Dunford and Schwatrz [109] (Theorem V.8.4 on p. 440), or Megginson [210] (Theorem 2.10.6 on p. 265). □

F.4 The Choquet-Bishop-De Leeuw Theorem

F.4.1 Preliminaries

Definition F.18. *Let $(S, \mathcal{T})$ be a topological space. A subset of S is called a G_δ set if it is a countable intersection of open sets in S.*

Definition F.19. *Let $(S, \mathcal{T})$ be a topological space. The Baire σ-algebra on $(S, \mathcal{T})$ is the sub-σ-algebra of the Borel σ-algebra on $(S, \mathcal{T})$ generated by the collection of all compact G_δ sets in S.*

Proposition F.20. *Let $(S, \mathcal{T})$ be a topological space. If $\mathcal{T}$ is metrizable, then the Baire and Borel σ-algebras on $(S, \mathcal{T})$ coincide.*

Definition F.21. *Let $(S, \mathcal{T})$ be a locally convex and Hausdorff topological vector space, and let S^* denote the topological dual of S. Let K be a compact subset of S, and suppose that μ is a regular Borel probability measure on K. We will say that a point $x \in S$ is represented by μ, or is the barycenter of μ, if for each $\phi \in S^*$, we have:*

$$\phi(x) = \int_K \phi \, d\mu \tag{F.4}$$

Theorem F.22. *Let $(S, \mathcal{T})$ be a locally convex and Hausdorff topological vector space, and let F be a closed subset of S. If the closed convex hull K of F is compact then each regular Borel probability measure μ on F has a unique barycenter in K.*

Proof. See Diestel [100], Theorem 1 on p. 148. □

F.4.2 Integral Representation Theorems

Choquet's celebrated theorem which we will state below, together with its generalization given by Bishop and De Leeuw, is essentially an integral representation theorem. As a matter of fact, the Krein-Milman theorem can be seen as an integral representation theorem, as we shall see below.

Definition F.23. *Let $(K, \mathcal{T})$ be a compact Hausdorff space, and let B be a Borel subset of K. We say that a Borel probability measure μ on K is supported by B, or concentrated on B, if $\mu(K \setminus B) = 0$.*

Theorem F.24. *Let $(S, \mathcal{T})$ be a locally convex and Hausdorff topological vector space, and let K be a compact subset of S. A point $x \in S$ is in the closed convex hull of K if and only if there exists some regular Borel probability measure μ on K whose barycenter exists and is x.*

Proof. See Diestel [100], Theorem 2 on p. 149. □

The previous theorem shows that the Krein-Milman theorem is an integral representation theorem. Indeed, in light of the previous theorem, we can restate the Krein-Milman theorem as follows:

Corollary F.25. *Let $(S, \mathcal{T})$ be a locally convex and Hausdorff topological vector space, and let K be a compact and convex subset of S. Then each point $x \in K$ is the barycenter of a regular Borel probability measure μ on K which is suported by the closure of the extreme points of K.*

Proof. See Phelps [226], p. 5. □

This is the starting point of Choquet's integral representation theorem and of the Bishop-De Leeuw theorem.

Theorem F.26 (Choquet)**.** *Let $(S, \mathcal{T})$ be a locally convex and Hausdorff topological vector space, and let K be a nonempty compact and convex metrizable subset of S. Then each point $x_0 \in K$ is the barycenter of a regular Borel probability measure μ that is supported by the extreme points of K.*

Proof. See Diestel [100] (p. 154) or Phelps [226] (p. 14). □

Choquet's theorem is a beautiful mathematical result, an important limitation of which is that it requires the set K to be metrizable, which is often not the case in practice (e.g. the weak topology of a linear space is metrizable if and only if that linear space is finite-dimensional). A natural generalization of Choquet's theorem to the non-metrizable case was given by Bishop and De Leeuw [39], and we state their result below.

Theorem F.27 (Bishop-De Leeuw)**.** *Let $(S, \mathcal{T})$ be a locally convex and Hausdorff topological vector space, and let K be a nonempty compact and convex subset of S. Fix some $x_0 \in K$. Then there exists a regular Borel probability measure μ on K such that:*

1. *The point x_0 is the barycenter of μ; and,*

2. *The measure μ vanishes on every Baire subset of K which is disjoint from the set of extreme points of K.*

Proof. See Phelps [226], pp. 17-22. □

The following reformulation of the Bishop-De Leeuw theorem is often more convenient for applications:

Theorem F.28 (Bishop-De Leeuw)**.** *Let $(S, \mathcal{T})$ be a locally convex and Hausdorff topological vector space, and let K be a nonempty compact and convex subset of S. Denote by $\mathcal{E}(K)$ the set of extreme points of K, and let Σ denote the σ-algebra of subsets of K which is generated by $\mathcal{E}(K)$ and the Baire sets. Then for each point $x_0 \in K$ there exists a nonnegative measure μ on Σ with $\mu(K) = 1$ such that μ represents x_0 and $\mu(\mathcal{E}(K)) = 1$.*

Proof. See Phelps [226], p. 22. □

Needless to say, the Choquet-Bishop-De Leeuw theorem is a remarkable result, for both its depth and its elegance. In practice, it is often extremely useful in characterizing compact convex sets in a "geometric" manner. Indeed, this theorem essentially says that any point in a compact convex subset of a locally convex Hausdorff topological vector space is some sort of an "average" of extreme points of that set. We should also mention that Bourgin [65] has extended the previous results to the noncompact case.

References

[1] W. Adamski. Complete Spaces and Zero-One Measures. *Manuscripta Mathematica*, 18(4):343–352, 1976. 110

[2] G.A. Akerlof. The Market for "Lemons" : Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, 1970. 54

[3] C.D. Aliprantis and K.C. Border. *Infinite Dimensional Analysis - 3^{rd} edition.* Springer-Verlag, 2006. 50, 51, 107, 174, 175, 203, 215, 222, 223, 224, 226, 227, 232, 239, 241, 244, 245, 246, 247, 250, 251

[4] M. Allais. *Fondements d'une Théorie Positive des Choix Comportant un Risque et Critique des Postulats et Axiomes de 1'École Américaine.* Mémoire présenté au Colloque International sur le risque, Centre National de la Recherche Scientifique, Paris, 1952. 184

[5] M. Allais. Le Comportement de l'Homme Rationnel Devant le Risque: Critique des Axiomes et Postulats de l'école Américaine. *Econometrica*, 21(4):503–546, 1953. 184, 185, 186

[6] S. Alon and D. Schmeidler. Purely Subjective Maxmin Expected Utility. mimeo (2008). 198, 201, 202

[7] M. Amarante. Ambiguous Events. mimeo (2003). 188

[8] M. Amarante. Analogical Reasoning in Decision Processes. mimeo (2006). 188

[9] M. Amarante. Analogy in Decision-Making. mimeo (2009). 188

[10] M. Amarante. States, Models and Information. mimeo (2010). 188

[11] M. Amarante. States, Models and Information: A Reconsideration of Ellsberg's Paradox. mimeo (2006). 188

[12] M. Amarante. Toward a Rational-Choice Foundation of Non-Additive Theories. mimeo (2009). 188

[13] M. Amarante. Ambiguity, Measurability and Multiple Priors. *Economic Theory*, 26(4):995–1006, 2005. 188

[14] M. Amarante. Foundations of Neo-Bayesian Statistics. *Journal of Economic Theory*, 144(5):2146–2173, 2009. 4, 188

[15] M. Amarante and E. Filiz. Ambiguous Events and Maxmin Expected Utility. *Journal of Economic Theory*, 134(1):1–33, 2007. 188

[16] M. Amarante, M. Ghossoub, and E. Phelps. The Entrepreneurial Economy I: Contracting under Knightian Uncertainty. mimeo (2011). 18

[17] M. Amarante and F. Maccheroni. When an Event Makes a Difference. *Theory and Decision*, 60(2):119–126, 2006. 188

[18] K.T. Andrews. Dunford-Pettis Sets in the Space of Bochner Integrable Functions. *Mathematische Annalen*, 241(1):35–41, 1979. 247

[19] F.J. Anscombe and R.J. Aumann. A Definition of Subjective Probability. *Annals of Mathematical Statistics*, 34(1):199–205, 1963. 3, 7, 167

[20] R.J. Arnott and J.E. Stiglitz. Equilibrium in Competitive Insurance Markets with Moral Hazard. NBER Working Paper No. W3588 (1991). 54, 160

[21] R.J. Arnott and J.E. Stiglitz. Price Equilibrium, Efficiency, and Decentralizability in Insurance Markets. NBER Working Paper No. W3642 (1991). 54, 160

[22] R.J. Arnott and J.E. Stiglitz. The Basic Analytics of Moral Hazard. *The Scandinavian Journal of Economics*, 90(3):383–413, 1988. 54, 160

[23] K.J. Arrow. Le Rôle des Valeurs Boursières pour la Répartition la Meilleure des Risques. *Colloques Internationaux du CNRS, XL*, pages 41–48, 1953. 6

[24] K.J. Arrow. Uncertainty and the Welfare Economics of Medical Care. *American Economic Review*, 53(5):941–973, 1963. 6

[25] K.J. Arrow. *Essays in the Theory of Risk-Bearing*. Chicago: Markham Publishing Company, 1971. 6, 15, 37, 53, 57, 59, 61, 65, 89, 97, 179

[26] K.J. Arrow. Optimal Insurance and Generalized Deductibles. *Scandinavian Actuarial Journal*, pages 1–42, 1974. 6

[27] R. B. Ash. *Real Analysis and Probability.* Academic Press, 1972. 239

[28] R.J. Aumann. Agreeing to Disagree. *The Annals of Statistics*, 4(6):1236–1239, 1976. 95

[29] N. Barberis and M. Huang. Stocks as Lotteries: The Implications of Probability Weighting for Security Prices. *American Economic Review*, 98(5):2066–2100, 2008. 190

[30] R.G. Bartle, N. Dunford, and J.T. Schwartz. Weak Compactness and Vector Measures. *Canadian Journal of Mathematics*, 7(3):289–305, 1955. 99, 223, 231

[31] C. Bennett and R. Sharpley. *Interpolation of Operators.* Academic Press, 1988. 19

[32] C. Bernard and M. Ghossoub. Static Portfolio Choice under Cumulative Prospect Theory. *Mathematics and Financial Economics*, 2(4):277–306, 2010. 10, 138, 190, 198

[33] C. Bernard and W. Tian. Optimal Reinsurance Arrangements under Tail Risk Measures. *Journal of risk and insurance*, 76(3):709–725, 2009. 73

[34] C. Bernard and W. Tian. Insurance Market Effects of Risk Management Metrics. *The Geneva Risk and Insurance Review*, 35(1):47–80, 2010. 73, 89, 97

[35] D. Bernouilli. *Specimen Theoriae Novae de Mensura Sortis.* St. Petersburg – English translation in *Econometrica*, 22, pp. 23-46 (1954), 1738. 5

[36] T.F. Bewley. Knightian Decision Theory. Part I. Cowles Foundation Discussion Papers, Paper no. 807 (1986). 188

[37] T.F. Bewley. Knightian Decision Theory, Part II: Intertemporal Problems. Cowles Foundation Discussion Papers, paper no. 835 (1988). 188

[38] P. Billingsley. *Convergence of Probability Measures – 2^{nd}edition.* Wiley, 1999. 175, 223, 237, 239

[39] E. Bishop and K. De Leeuw. The Representations of Linear Functionals by Measures on Sets of Extreme Points. *Annales de l'institut Fourier*, 9:305–331, 1959. 247, 253

[40] E. Bishop and R.R. Phelps. The Support Functionals of a Convex Set. In *Proceedings Symposium Pure Math. American Mathematical Society, Vol. 7 (Convexity).* AMS Bookstore, 1963. 247, 250

[41] P.R. Blavatskyy. Loss Aversion. *Economic Theory*, 46(1):127–148, 2011. 191

[42] V.I. Bogachev. *Measure Theory - Volume 1.* Springer-Verlag, 2007. 79, 244

[43] V.I. Bogachev. *Measure Theory - Volume 2.* Springer-Verlag, 2007. 232

[44] E. Böhm-Bawerk. *Rechte und Verhaltnisse vom Standpunkte der Volkswirtschaflichen Guterlehre.* Innsbruck, 1881. 5

[45] B. Bollobas. An Extension to the Theorem of Bishop and Phelps. *Bulletin of the London Mathematical Society*, 2(2):181, 1970. 250

[46] K. Borch. Reciprocal Reinsurance Treaties. *Astin Bulletin*, 1(4):170–190, 1960. 6

[47] K. Borch. Some Elements of a Theory of Reinsurance. *The Journal of Insurance*, 28(3):35–43, 1961. 6

[48] K. Borch. Equilibrium in a Reinsurance Market. *Econometrica*, 30(3):424–44, 1962. 6

[49] K. Borch. The Theory of Risk. *Journal of the Royal Statistical Society. Series B (Methodological)*, 29(3):432–467, 1967. 6

[50] K. Borch. The Optimal Reinsurance Treaty. *Astin Bulletin*, 5(2):293–297, 1969. 6

[51] K. Borch. Problems in the Economic Theory of Insurance. *Astin Bulletin*, 10(1):1–11, 1978. 6

[52] K.H. Borch. *The Mathematical Theory of Insurance: An Annotated Selection of Papers on Insurance Published 1960-1972.* Lexington Books, 1974. 6, 53, 65, 97

[53] N. Bourbaki. *Éléments de Mathématique – Livre VI – Intégration: Chapitre 6.* Paris: Hermannn, 1959. 239

[54] N. Bourbaki. *Éléments de Mathématique – Livre VI – Intégration: Chapitres 7 et 8.* Paris: Hermannn, 1963. 239

[55] N. Bourbaki. *Éléments de Mathématique – Livre VI – Intégration: Chapitres 1 à 4.* Paris: Hermannn, 1965. 239, 247

[56] N. Bourbaki. *Éléments de Mathématique – Livre VI – Intégration: Chapitre 5.* Paris: Hermannn, 1967. 239

[57] N. Bourbaki. *Éléments de Mathématique – Livre VI – Intégration: Chapitre 9.* Paris: Hermannn, 1969. 239

[58] N. Bourbaki. *Éléments de Mathématique – Livre III – Topologie Générale: Chapitres 1 à 4*. Paris: Hermannn, 1971. 203

[59] N. Bourbaki. *Éléments de Mathématique – Livre III – Topologie Générale: Chapitres 5 à 10*. Paris: Hermannn, 1974. 232

[60] J. Bourgain. Strongly Exposed Points in Weakly Compact Convex Sets in Banach Spaces. *Proceedings of the American Mathematical Society*, 58(1):197–200, 1976. 247

[61] J. Bourgain. On Dentability and the Bishop-Phelps Property. *Israel Journal of Mathematics*, 28(4):265–271, 1977. 247, 250

[62] J. Bourgain. A Characterization of Non-Dunford-Pettis Operators on L_1. *Israel Journal of Mathematics*, 37(1):48–53, 1980. 247

[63] J. Bourgain. Dunford-Pettis Operators on L_1 and the Radon-Nikodym Property. *Israel Journal of Mathematics*, 37(1):34–47, 1980. 247

[64] R.D. Bourgin. *Barycenters of Measures on Certain Noncompact Convex Sets*. University of Washington Press, 1969. 247

[65] R.D. Bourgin. Barycenters of Measures on Certain Noncompact Convex Sets. *Transactions of the American Mathematical Society*, 154:323–340, 1971. 254

[66] R.D. Bourgin. *Geometric Aspects of Convex Sets with the Radon-Nikodym Property – Volume 993 of Lecture Notes in Mathematics*. Springer-Verlag, 1983. 247

[67] S. Bugajski, K.-E. Hellwig, and W. Stulpe. On Fuzzy Random Variables and Statistical Maps. *Reports on Mathematical Physics*, 41(1):1–11, 1998. 110

[68] S. Cambanis, G. Simons, and W. Stout. Inequalities for $\mathcal{E}k\,(x, y)$ when the Marginals are Fixed. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 36(4):285–294, 1976. 31

[69] G. Cantor. *Contributions to the Founding Theory of Transfinite Numbers*. Chicago/London: The Open Court Publishing Company, 1915. 171

[70] P. Cardaliaguet and R. Tahraoui. Extended Hardy-Littlewood Inequalities and Applications to the Calculus of Variations. *Advances in Differential Equations*, 5(7-9):1091–1138, 2000. 39

[71] G. Carlier and R.A. Dana. Core of Convex Distortions of a Probability. *Journal of Economic Theory*, 113(2):199–222, 2003. 136

[72] G. Carlier and R.A. Dana. Pareto Efficient Insurance Contracts when the Insurer's Cost Function is Discontinuous. *Economic Theory*, 21(4):871–893, 2003. 19, 26, 28, 54

[73] G. Carlier and R.A. Dana. Existence and Monotonicity of Solutions to Moral Hazard Problems. *Journal of Mathematical Economics*, 41(7):826–843, 2005. 19, 26

[74] G. Carlier and R.A. Dana. Rearrangement Inequalities in Non-convex Insurance Models. *Journal of Mathematical Economics*, 41(4-5):483–503, 2005. 19, 27, 31, 54

[75] N.L. Carothers. *Real Analysis*. Cambridge University Press, 2000. 51

[76] R. Casadesus-Masanell, P. Klibanoff, and E. Ozdenoren. Maxmin Expected Utility over Savage Acts with a Set of Priors. *Journal of Economic Theory*, 92(1):35–65, 2000. 198, 201, 202

[77] A. Chateauneuf, F. Maccheroni, M. Marinacci, and J.M. Tallon. Monotone Continuous Multiple Priors. *Economic Theory*, 26(4):973–982, 2005. 15, 57, 179, 188

[78] K.M. Chong. Some Extensions of a Theorem of Hardy, Littlewood and Pólya and their Applications. *Canadian Journal of Mathematics*, 26(6):1321–1340, 1974. 19

[79] K.M. Chong. Spectral Inequalities Involving the Infima and Suprema of Functions. *Pacific Journal of Mathematics*, 59(1):17–20, 1975. 19

[80] K.M. Chong. Spectral Inequalities Involving the Sums and Products of Functions. *International Journal of Mathematics and Mathematical Sciences*, 5(1):141–157, 1982. 19, 33

[81] K.M. Chong and N.M. Rice. *Equimeasurable Rearrangements of Functions*. Queens Papers in Pure and Applied Mathematics, 28, 1971. 19, 32

[82] D. L. Cohn. *Measure Theory*. Birkhauser, 1980. 98, 116, 223, 224, 226, 232, 239, 241, 244

[83] J.B. Collier. The Dual of a Space with the Radon-Nikodym Property. *Pacific Journal of Mathematics*, 64(1):103–106, 1976. 247

[84] J.R. Conlon. Two New Conditions Supporting the First-Order Approach to Multisignal Principal-Agent Problems. *Econometrica*, 77(1):249–278, 2009. 160

[85] J.D. Cummins and O. Mahul. Optimal Insurance with Divergent Beliefs about Insurer Total Default Risk. *The Journal of Risk and Uncertainty*, 27(2):121–138, 2003. 96

[86] J.D. Cummins and O. Mahul. The Demand for Insurance with an Upper Limit on Coverage. *The Journal of Risk and Insurance*, 71(2):253–264, 2004. 53, 97

[87] H.C.A. Dale. A Priori Probabilities in Gambling. *Nature*, 183:842–843, 1959. 191

[88] R.A. Dana and M. Scarsini. Optimal Risk Sharing with Background Risk. *Journal of Economic Theory*, 133(1):152–176, 2007. 19, 31, 39

[89] W.J. Davis and R.R. Phelps. The Radon-Nikodym Property and Dentable Sets in Banach Spaces. *Proceedings of the American Mathematical Society*, 45(1):119–122, 1974. 247

[90] G. Debreu. Representation of a Preference Ordering by a Numerical Function. In *R.M. Thrall, C.H. Coombs, and R.L. Davis (eds.), Decision Processes*. John Wiley and Sons, New York, 1954. 172

[91] G. Debreu. Topological Methods in Cardinal Utility Theory. In *K.J. Arrow, S. Karlin and P. Suppes (eds.), Mathematical Methods in the Social Sciences*. Stanford Univ. Press, Stanford, CA, 1960. 172

[92] G. Debreu. Continuity Properties of Paretian Utility. *International Economic Review*, 5(3):285–293, 1964. 172

[93] B. De Finetti. La Prévision: Ses Lois Logiques, Ses Sources Subjectives. *Annales de l'Institut Henri Poincaré*, 7(1):1–68, 1937. v, 3, 7, 155

[94] C. Dellacherie and P. A. Meyer. *Probabilities and Potential*. North-Holland Mathematics Studies - Issue 29, 1978. 224

[95] F. Demers and M. Demers. Increases in Risk and the Optimal Deductible. *Journal of Risk and Insurance*, 58(4):670–699, 1991. 53

[96] D. Denneberg. *Non-Additive Measure and Integral*. Kluwer Academic Publishers, 1994. 116, 118, 119, 136, 141, 147, 148

[97] O. Deprez and H.U. Gerber. On Convex Principles of Premium Calculation. *Insurance: Mathematics and Economics*, 4(3):179–189, 1985. 54

[98] J. Diestel. *Geometry of Banach Spaces: Selected Topics*. Springer-Verlag, 1975. 247

[99] J. Diestel. Remarks on Weak Compactness in $L_1(\mu, X)$. *Glasgow Mathematical Journal*, 18(1):87–91, 1977. 247

[100] J. DIESTEL. *Sequences and Series in Banach Spaces.* New York: Springer, 1984. 99, 215, 221, 222, 230, 247, 251, 252, 253

[101] J. DIESTEL, W.M. RUESS, and W. SCHACHERMAYER. Weak Compactness in $L_1(\mu, X)$. *Proceedings of the American Mathematical Society*, 118(2):447–453, 1993. 247

[102] J. DIESTEL and J.J. UHL. *Vector Measures.* AMS Mathematical Surverys, 1977. 247, 250

[103] J. DIEUDONNÉ. *Foundations of Modern Analysis.* Academic press, 1969. 29, 41

[104] N.A. DOHERTY and L. EECKHOUDT. Optimal Insurance without Expected Utility: The Dual Theory and the Linearity of Insurance Contracts. *Journal of Risk and Uncertainty*, 10(2):157–179, 1995. 54

[105] J.L. DOOB. *Stochastic Processes.* John Wiley and Sons, 1953. 224

[106] J.L. DOOB. *Measure Theory.* Springer, 1994. 51, 223, 239

[107] J.H. DRÈZE. Inferring Risk Tolerance from Deductibles in Insurance Contracts. *The Geneva Papers on Risk and Insurance*, 6(1):48–52, 1981. 53

[108] R. M. DUDLEY. *Real Analysis and Probability.* Cambridge University Press, 2002. 203, 215, 239

[109] N. DUNFORD and J.T. SCHWARTZ. *Linear Operators, Part 1: General Theory.* Wiley-Interscience, 1958. 99, 108, 175, 203, 215, 223, 227, 229, 230, 231, 241, 244, 245, 246, 247, 251

[110] G.A. EDGAR. A Noncompact Choquet Theorem. *Proceedings of the American Mathematical Society*, 49(2):354–358, 1975. 247

[111] W. EDWARDS. Probability-Preferences in Gambling. *The American Journal of Psychology*, 66(3):349–364, 1953. 191

[112] W. EDWARDS. Probability-Preferences among Bets with Differing Expected Values. *The American Journal of Psychology*, 67(1):56–67, 1954. 191

[113] W. EDWARDS. The Reliability of Probability-Preferences. *The American Journal of Psychology*, 67(1):68–95, 1954. 191

[114] W. EDWARDS. The Prediction of Decisions Among Bets. *Journal of Experimental Psychology*, 50(3):201–214, 1955. 191

[115] W. Edwards. Subjective Probabilities Inferred from Decisions. *Psychological Review*, 69(2):109–135, 1962. 191

[116] L. Eeckhoudt, C. Gollier, and H. Schlesinger. Increases in Risk and Deductible Insurance. *Journal of Economic Theory*, 55(2):435–440, 1991. 53

[117] S. Eilenberg. Ordered Topological Spaces. *American Journal of Mathematics*, 63(1):39–45, 1941. 172

[118] D. Ellsberg. Risk, Ambiguity and the Savage Axioms. *Quarterly Journal of Economics*, 75(4):643–669, 1961. 4, 8, 184, 186

[119] P. Embrechts and M. Hofert. A Note on Generalized Inverses. mimeo (2010). 21, 151

[120] J.B. Epperson. A Class of Monotone Decreasing Rearrangements. *Journal of Mathematical Analysis and Applications*, 150(1):224–236, 1990. 19, 26, 33

[121] L. Epstein. A Definition of Uncertainty Aversion. *Review of Economic Studies*, 66(3):579–608, 1999. 4, 188

[122] L. Epstein and M. Schneider. Recursive Multiple-Priors. *Journal of Economic Theory*, 113(1):1–31, 2003. 188

[123] L. Epstein and J. Zhang. Subjective Probabilities on Subjectively Unambiguous Events. *Econometrica*, 69(2):265–306, 2001. 4, 188

[124] V. Farmaki. A Geometric Characterization of the Weak-Radon Nikodym Property in Dual Banach Spaces. *Rocky Mountain Journal of Mathematics*, 25(2):611–617, 1995. 247

[125] W. Fellner. Distortion of Subjective Probabilities as a Reaction to Uncertainty. *Quarterly Journal of Economics*, 75(4):670–689, 1961. 191

[126] G. Fichtenholz and L.V. Kantorovich. Sur les Opérationes Linéaires dans L'espace des Fonctions Bornées. *Studia Mathematica*, 5:69–98, 1934. 230

[127] T. Figiel, N. Ghoussoub, and W.B. Johnson. On the Structure of Non-Weakly Compact Operators on Banach Lattices. *Mathematische Annalen*, 257(3):317–334, 1981. 247

[128] P.C. Fishburn. *Utility Theory for Decision Making*. John Wiley & Sons, 1970. 57, 162, 168, 171, 172, 174, 175, 180, 181, 183, 199

[129] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time – 2^{nd} ed.* Walter de Gruyter, 2004. 127, 135, 136

[130] V.P. Fonf and J. Lindenstrauss. Some Results on Infinite-Dimensional Convexity. *Israel Journal of Mathematics*, 108(1):13–32, 1998. 247

[131] M. Friedman and L.J. Savage. The Utility Analysis of Choices Involving Risk. *The Journal of Political Economy*, 56(4):279, 1948. 191

[132] P. Gänssler. Compactness and Sequential Compactness in Spaces of Measures. *Probability Theory and Related Fields*, 17(2):124–146, 1971. 99, 223, 231

[133] P. Ghirardato. Defining Ambiguity and Ambiguity Attitude. In *I. Gilboa (ed.), Uncertainty in Economic Theory*. Routledge, London, 2004. 188

[134] P. Ghirardato, F. Maccheroni, and M. Marinacci. Differentiating Ambiguity and Ambiguity Attitude. *Journal of Economic Theory*, 118(2):133–173, 2004. 4, 188

[135] P. Ghirardato, F. Maccheroni, and M. Marinacci. Revealed Ambiguity and its Consequences: Updating. In *Advances in Decision Making Under Risk and Uncertainty*, volume 42 of *Theory and Decision Library C*, pages 3–18. Springer, 2008. 188

[136] P. Ghirardato, F. Maccheroni, M. Marinacci, and M. Siniscalchi. A Subjective Spin on Roulette Wheels. *Econometrica*, 71(6):1897–1908, 2003. 188

[137] P. Ghirardato and M. Marinacci. Range Convexity and Ambiguity Averse Preferences. *Economic Theory*, 17(3):599–617, 2001. 188

[138] P. Ghirardato and M. Marinacci. Risk, Ambiguity, and the Separation of Utility and Beliefs. *Mathematics of Operations Research*, 26(4):864–890, 2001. 188

[139] P. Ghirardato and M. Marinacci. Ambiguity Made Precise: A Comparative Foundation. *Journal of Economic Theory*, 102(2):251–289, 2002. 4, 188

[140] P. Ghirardato and M. Siniscalchi. A More Robust Definition of Multiple Priors. mimeo (2010). 188

[141] M. Ghossoub. A More Robust Definition of Loss Aversion. mimeo (2010). 10, 138, 190, 191, 198

[142] N. Ghoussoub, G. Godefroy, B. Maurey, and W. Schachermayer. *Some Topological and Geometrical Structures in Banach Spaces – Issue 378 of the Memoirs of the American Mathematical Society*. AMS Bookstore, 1987. 247

[143] N. Ghoussoub and B. Maurey. On the Radon-Nikodym Property in Function Spaces. In *K. Nigel and E. Saab (eds.), Banach Spaces. Lecture Notes in Mathematics*. Springer, 1985. 247

[144] N. Ghoussoub and E. Saab. On the Weak Radon-Nikodym Property. *Proceedings of the American Mathematical Society*, 81(1):81–84, 1981. 247

[145] N. Ghoussoub and P. Saab. Weak Compactness in Spaces of Bochner Integrable Functions and the Radon-Nikodym Property. *Pacific Journal of Mathematics*, 110(1):65–70, 1984. 247

[146] N. Ghoussoub and M. Talagrand. Order Dentability and the Radon-Nikodym Property in Banach Lattices. *Mathematische Annalen*, 243(3):217–225, 1979. 247

[147] I. Gilboa. Subjective Distortions of Probabilities and Non-Additive Probabilities. mimeo (1985). 125

[148] I. Gilboa. Expected Utility with Purely Subjective Non-additive Probabilities. *Journal of Mathematical Economics*, 16(1):65–88, 1987. 4, 9, 114, 119, 188, 189, 198, 201, 202

[149] I. Gilboa. *Theory of Decision under Uncertainty*. Cambridge University Press, 2009. 2, 3, 57, 180, 183, 190

[150] I. Gilboa, F. Maccheroni, M. Marinacci, and D. Schmeidler. Objective and Subjective Rationality in a Multiple Prior Model. *Econometrica*, 78(2):755–770, 2010. 188

[151] I. Gilboa and D. Schmeidler. Maxmin Expected Utility with a Non-Unique Prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989. 4, 8, 188, 189, 198, 201

[152] D. Gilliam. Geometry and the Radon-Nikodym Theorem in Strict Mackey Convergence Spaces. *Pacific Journal of Mathematics*, 65(2):353–364, 1976. 247

[153] C. Gollier. Economic Theory of Risk Exchanges: A Review. In *G. Dionne (ed.), Contributions to Insurance Economics*. Kluwer Academic publishers, 1992. 6

[154] C. Gollier. Optimum Insurance of Approximate Losses. *Journal of Risk and Insurance*, 63(3):369–380, 1996. 97

[155] C. Gollier and H. Schlesinger. Arrow's Theorem on the Optimality of Deductibles: A Stochastic Dominance Approach. *Economic Theory*, 7:359–363, 1996. 54

[156] J.P. Gould. The Expected Utility Hypothesis and the Selection of Optimal Deductibles for a Given Insurance Policy. *Journal of business*, 42(2):143–151, 1969. 53

[157] G. Greco. Sulla Rappresentazione di Funzionali Mediante Integrali. *Rendiconti Seminario Matematico Università di Padova*, 66:21–42, 1982. 148

[158] N.E. Gretsky. *Representation Theorems on Banach Function Spaces - Volume 84 of the Memoirs of the American Mathematical Society*. AMS Bookstore, 1968. 239

[159] R.M. Griffith. Odds Adjustments by American Horse-Race Bettors. *The American Journal of Psychology*, 62(2):290–294, 1949. 191

[160] S.J. Grossman and O.D. Hart. An Analysis of the Principal-Agent Problem. *Econometrica*, 51(1):7–45, 1983. 160

[161] P.R. Halmos. *Measure Theory*. Springer, 1974. 223, 239

[162] J. Handa. Risk, Probabilities and a New Theory of Cardinal Utility. *Journal of Political Economy*, 85(1):97–122, 1977. 191

[163] G.H. Hardy, J.E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, Cambridge. Reprint of the 1952 edition, 1988. 19

[164] X.D. He and X.Y. Zhou. Portfolio Choice via Quantiles. *Mathematical Finance*, 21(2):203–231, 2011. 136, 190

[165] I. N. Herstein and J. Milnor. An Axiomatic Approach to Measurable Utility. *Econometrica*, 21(2):291–297, 1953. 162, 163

[166] E. Hewitt and K. Stromberg. *Real and Abstract Analysis*. Springer-Verlag: New York, 1965. 35, 175, 203, 215, 223, 224, 226, 227, 239, 244

[167] T.H. Hildebrandt. On Bounded Functional Operations. *Transactions of the American Mathematical Society*, 36(4):868–875, 1934. 230

[168] A. Horsley and A.J. Wrobel. The Mackey Continuity of the Monotone Rearrangement. *Proceedings of the American Mathematical Society*, 97(4):626–628, 1986. 34

[169] G. Huberman, D. Mayers, and C.W. Smith Jr. Optimal Insurance Policy Indemnity Schedules. *The Bell Journal of Economics*, 14(2):415–426, 1983. 11, 61, 63

[170] R.E. Huff. Dentability and the Radon-Nikodym Property. *Duke Mathematical Journal*, 41(1):111–114, 1974. 247

[171] R.E. Huff and P.D. Morris. Dual Spaces with the Krein-Milman Property have the Radon-Nikodym Property. *Proceedings of the American Mathematical Society*, 49(1):104–108, 1975. 247

[172] K. Inada. On a Two-Sector Model of Economic Growth: Comments and a Generalization. *The Review of Economic Studies*, 30(2):119–127, 1963. 57

[173] M. Jeleva and B. Villeneuve. Insurance Contracts with Imprecise Probabilities and Adverse Selection. *Economic Theory*, 23(4):777–794, 2004. 55

[174] I. Jewitt. Justifying the First-Order Approach to Principal-Agent Problems. *Econometrica*, 56(5):1177–1190, 1988. 160

[175] H. Jin and X.Y. Zhou. Behavioral Portfolio Selection in Continous Time. *Mathematical Finance*, 18(3):385–426, 2008. 42, 136, 190

[176] O. Kadan and J.M. Swinkels. On the Moral Hazard Problem without the First-Order Approach. mimeo (2010). 160

[177] D. Kahneman and A. Tversky. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47(2):263–291, 1979. 9, 138, 144, 189, 190, 192, 193, 194, 195

[178] U.S. Karmarkar. Subjectively Weighted Utility: A Descriptive Extension of the Expected Utility Model. *Organizational Behavior and Human Performance*, 21(1):61–72, 1978. 191

[179] E. Karni. Optimal Insurance: A Nonexpected Utility Analysis. In *G. Dionne (ed.), Contributions to Insurance Economics*. Kluwer Academic publishers, 1992. 54

[180] J.L. Kelley. *General Topology*. Springer, 1975. 203

[181] J.L. Kelley and I. Namioka. *Linear Topological Spaces*. Springer-Verlag, 1976. 215, 247

[182] P. Klibanoff, M. Marinacci, and S. Mukerji. A Smooth Model of Decision Making under Ambiguity. *Econometrica*, 73(6):1849–1892, 2005. 4, 188

[183] P. Klibanoff, M. Marinacci, and S. Mukerji. Recursive Smooth Ambiguity Preferences. *Journal of Economic Theory*, 144(3):930–976, 2009. 188

[184] F.H. Knight. *Risk, Uncertainty, and Profit*. Boston/New York: Houghton Mifflin, 1921. 4, 7, 188

[185] V. Köbberling and P.P. Wakker. An Index of Loss Aversion. *Journal of Economic Theory*, 122(1):119–131, 2005. 191

[186] A.N. Kolmogorov and S.V. Fomin. *Introductory Real Analysis*. New York: Dover Publications – Translated by R.A. Silverman, 1975. 203, 215

[187] D.M. Kreps. *Notes on the Theory of Choice*. Westview Press, 1988. 2, 3, 171, 175, 179, 180

[188] K. Kuratowski. *Topology – Vol. 1-2*. New York : Academic Press, 1966-1968. 232

[189] B.L. Lin, P.K. Lin, and S.L. Troyanski. Characterizations of Denting Points. *Proceedings of the American Mathematical Society*, pages 526–528, 1988. 247

[190] J. Lindenbaum. Ein Vierteljahrhundert der Bedarfstheorie der Versicherung. *Zeitscrift fur Nationalokonomie*, 2:75–99, 1931. 5

[191] J. Lindenstrauss. On Operators Which Attain Their Norm. *Israel Journal of Mathematics*, 1(3):139–148, 1963. 250

[192] W.A.J. Luxemburg. Rearrangement Invariant Banach Function Spaces. In *Queens Papers in Pure and Applied Mathematics, vol. 10: Proceeding of the Symposium in Analysis*, pages 83–144. Queen's University, Kingston, Ontario, 1967. 19

[193] F. Maccheroni. Maxmin under Risk. *Economic Theory*, 19(4):823–831, 2002. 188

[194] F. Maccheroni and M. Marinacci. A Heine-Borel Theorem for $ba(\Sigma)$. *RISEC*, 48(3):353–362, 2001. 99, 223, 231, 232

[195] F. Maccheroni, M. Marinacci, and A. Rustichini. Ambiguity Aversion, Robustnsess, and the Variational Representation of Preferences. *Econometrica*, 74(6):1447–1498, 2006. 4, 100, 188

[196] F. Maccheroni, M. Marinacci, and A. Rustichini. Dynamic Variational Preferences. *Journal of Economic Theory*, 128(1):4–44, 2006. 188

[197] M.J. Machina. Choice under Uncertainty: Problems Solved and Unsolved. *The Journal of Economic Perspectives*, 1(1):121–154, 1987. 184, 185

[198] M.J. Machina. Non-Expected Utility and the Robustness of the Classical Insurance Paradigm. *The GENEVA Papers on Risk and Insurance-Theory*, 20(1):9–50, 1995. 54

[199] M.J. Machina and D. Schmeidler. A More Robust Definition of Subjective Probability. *Econometrica*, 60(4):745–780, 1992. 191

[200] L.W. Marcoux. *An Introduction to Linear Analysis*. Unpublished Draft, 2009. 215

[201] M. Marinacci. A Simple Proof of a Basic Result for Multiple Priors. mimeo (1997). 188

[202] M. Marinacci. Ambiguous Games. *Games and Economic Behavior*, 31(2):191–219, 2000. 188

[203] M. Marinacci. Probabilistic Sophistication and Multiple Priors. *Econometrica*, 70(2):755–764, 2002. 188

[204] M. Marinacci and L. Montrucchio. Introduction to the Mathematics of Ambiguity. In *I. Gilboa (ed.), Uncertainty in Economic Theory*. Routledge, London, 2004. 99, 117, 118, 119, 148

[205] M. Marinacci and M.A. Turner. Managing Extinctions and Biodiversity. mimeo (1998). 188

[206] H. Markowitz. The Utility of Wealth. *The Journal of Political Economy*, 60(2):151–158, 1952. 189, 191, 192

[207] J. Marshall. Optimum Insurance with Deviant Beliefs. *in G. Dionne (ed.), Contributions to Insurance Economics. Boston: Kluwer Academic Publishers*, pages 255–274, 1991. xii, 56, 94, 95

[208] M. Matsuda. On Saab's Characterizations of Weak Radon-Nikodym Sets. *Publ. RIMS, Kyoto University*, 21:921–941, 1985. 247

[209] H.B. Maynard. A Geometrical Characterization of Banach Spaces with the Radon-Nikodym Property. *Transactions of the American Mathematical Society*, 185:493–500, 1973. 247

[210] R.E. Megginson. *An Introduction to Banach Space Theory*. New York: Springer, 1998. 99, 215, 219, 221, 247, 250, 251

[211] J. Meyer and M.B. Ormiston. Analyzing the Demand for Deductible Insurance. *Journal of Risk and Uncertainty*, 18(3):223–230, 1999. 53

[212] J.A. Mirrlees. The Theory of Moral Hazard and Unobservable Behaviour: Part I. *Review of Economic Studies*, 66(1):3–21, 1999. 160

[213] D. Moffet. Optimal Deductible and Consumption Theory. *Journal of risk and insurance*, pages 669–682, 1977. 53, 97

[214] J. Mossin. Aspects of Rational Insurance Purchasing. *The Journal of Political Economy*, 76(4):553–568, 1968. 53, 97

[215] F. Mosteller and P. Nogee. An Experimental Measurement of Utility. *The Journal of Political Economy*, 59(5):371–404, 1951. 192

[216] K. Nehring. Imprecise Probabilistic Beliefs as a Context for Decision-Making under Ambiguity. mimeo (2002). 188

[217] K. Nehring. Capacities and Probabilistic Beliefs: A Precarious Coexistence. *Mathematical Social Sciences*, 38(2):197–214, 1999. 188

[218] P. Newman and R. Read. Representation Problems for Preference Orderings. *Journal of Economic Behavior*, 1:149–169, 1961. 172

[219] F.H. Page. The Existence of Optimal Contracts in the Principal-Agent model. *Journal of Mathematical Economics*, 16(2):157–167, 1987. 160

[220] F.H. Page. Optimal Contract Mechanisms for Principal-Agent Problems with Moral Hazard and Adverse Selection. *Economic Theory*, 1(4):323–338, 1991. 160

[221] E. Pap. *Null-Additive Set Functions.* Kluwer Academic Publishers, 1995. 147

[222] J.R. Partington. Norm Attaining Operators. *Israel Journal of Mathematics*, 43(3):273–276, 1982. 250

[223] B.P. Pashigian, L.L. Schkade, and G.H. Menefee. The Selection of an Optimal Deductible for a Given Insurance Policy. *Journal of business*, 39(1):35–44, 1966. 53

[224] B. Peleg. Utility Functions for Partially Ordered Topological Spaces. *Econometrica*, 38(1):93–96, 1970. 172

[225] R.R. Phelps. Dentability and Extreme Points in Banach Spaces. *Journal of functional analysis*, 17(1):78–90, 1974. 247

[226] R.R. Phelps. *Lectures on Choquet's Theorem – 2^{nd} edition.* Springer-Verlag, 2001. 247, 253, 254

[227] M.G. Preston and P. Baratta. An Experimental Study of the Auction-Value of an Uncertain Outcome. *The American Journal of Psychology*, 61(2):183–193, 1948. 192

[228] S.D. Promislow and V.R. Young. Unifying Framework for Optimal Insurance. *Insurance: Mathematics and Economics*, 36(3):347–364, 2005. 54

[229] J. Quiggin. A Theory of Anticipated Utility. *Journal of Economic Behavior*, 3(4):323–343, 1982. 196

[230] S.T. Rachev and L. Rüschendorf. *Mass Transportation Problems - Vol. 1: Theory*. Springer, 1998. 31

[231] S.T. Rachev and L. Rüschendorf. *Mass Transportation Problems - Vol. 2: Applications*. Springer, 1998. 31

[232] T. Rader. The Existence of a Utility Function to Represent Preferences. *The Review of Economic Studies*, 30(3):229–232, 1963. 172

[233] J.M. Rakotoson. *Réarrangement Relatif*. Springer, 2008. 19, 33

[234] F.P. Ramsey. *The Foundations of Mathematics and other Logical Essays*. The Humanities Press: New York (Ed. R.B. Braithwaite), 1950. v, 3, 7, 155

[235] N. Randrianantoanina and E. Saab. The Complete Continuity Property in Bochner Function Spaces. *Proceedings of the American Mathematical Society*, 117(4):1109–1114, 1993. 247

[236] N. Randrianantoanina and E. Saab. Stability of Some Types of Radon-Nikodym Properties. *Illinois Journal of Mathematics*, 39(3):416–430, 1995. 247

[237] N. Randrianantoanina and E. Saab. The Near Radon-Niodym Property in Lebesgue-Bochner Function Spaces. *Illinois Journal of Mathematics*, 42(1):40–57, 1998. 247

[238] B. Rao and B. Rao. *Theory of Charges*. New York: Academic Press, 1983. 99, 223, 231

[239] A. Raviv. The Design of an Optimal Insurance Policy. *The American Economic Review*, 69(1):84–96, 1979. 6, 37, 53, 59, 61, 63, 65, 80, 89, 97

[240] S.I. Resnick. *A Probability Path*. Birkhauser, 1999. 23, 51, 241

[241] L.H. Riddle. The Geometry of Weak Radon-Nikodym Sets in Dual Banach Spaces. *Proceedings of the American Mathematical Society*, 86(3):433–438, 1982. 247

[242] L.H. Riddle. Dunford-Pettis Operators and Weak Radon-Nikodym Sets. *Proceedings of the American Mathematical Society*, 91(2):254–256, 1984. 247

[243] L.H. Riddle, E. Saab, and J.J. Uhl. Sets with the Weak Radon-Nikodym Property in Dual Banach Spaces. *Indiana University Mathematics Journal*, 32:527–541, 1983. 247

[244] F. Riesz. Untersuchungen über Systeme integrierbarer Funktionen. *Mathematische Annalen*, 69(4):449–497, 1910. 245

[245] W.P. Rogerson. The First-Order Approach to Principal-Agent Problems. *Econometrica*, 53(6):1357–1367, 1985. 160

[246] H. Rosenthal and A. Wessel. The Krein-Milman Property and a Martingale Coordinatization of Certain Non-Dentable Convex Sets. *Pacific Journal of Mathematics*, 136(1):159–182, 1989. 247

[247] M. Rothschild and J.E. Stiglitz. Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *The Quarterly Journal of Economics*, 90(4):629–649, 1976. 54

[248] W Rudin. *Functional Analysis*. New York: McGraw-Hill Book Company, 1973. 107, 108, 215

[249] W.M. Ruess and C.P. Stegall. Extreme Points in Duals of Operator Spaces. *Mathematische Annalen*, 261(4):535–546, 1982. 247

[250] E. Saab. On the Radon-Nikodym Property in a Class of Locally Convex Spaces. *Pacific Journal of Mathematics*, 75(1):281–291, 1978. 247

[251] E. Saab. On Measurable Projections in Banach Spaces. *Pacific Journal of Mathematics*, 97(2):453–459, 1981. 247

[252] E. Saab. On Dunford-Pettis Operators. *Canadian Mathematical Bulletin*, 25(2):207–209, 1982. 247

[253] E. Saab. On Dunford-Pettis Operators that are Pettis-Representable. *Proceedings of the American Mathematical Society*, 85(3):363–365, 1982. 247

[254] E. Saab. Some Characterizations of Weak Radon-Nikodym Sets. *Proceedings of the American Mathematical Society*, 86(2):307–311, 1982. 247

[255] E. Saab. On the Radon-Nikodym Property in Banach Lattices. *Proceedings of the American Mathematical Society*, 94(4):736–737, 1985. 247

[256] E. Saab. Exposed Points and the Radon-Nikodym Property. In *R.G. Bartle, N.T. Peck, A.L. Peressini, and J.J. Uhl (eds.), Geometry of Normed Linear Spaces*. AMS Bookstore, 1986. 247

[257] E. Saab. On the Weak*-Radon Nikodym Property. *Bulletin of the Australian Mathematical Society*, 37(3):323–332, 1988. 247

[258] E. Saab and P. Saab. A Dual Geometric Characterization of Banach Spaces Not Containing l_1. *Pacific Journal of Mathematics*, 105(2):415–425, 1983. 247

[259] E. Saab and P. Saab. On Unconditionally Converging and Weakly Precompact Operators. *Illinois Journal of Mathematics*, 35(3):522–531, 1991. 247

[260] P. Saab. The Choquet Integral Representation in the Affine Vector-Valued Case. *Aequationes Mathematicae*, 20:252–262, 1980. 247

[261] P. Saab. Integral Representation by Boundary Vector Measures. *Canadian Mathematical Bulletin*, 25(2):164–168, 1982. 247

[262] P. Saab. Weakly Compact, Unconditionally Converging, and Dunford-Pettis Operators on Spaces of Vector-Valued Continuous Functions. *Mathematical Proceedings of the Cambridge Philosophical Society*, 95(1):101–108, 1984. 247

[263] P. Saab. Integral Operators on Spaces of Continuous Vector-Valued Functions. *Proceedings of the American Mathematical Society*, 111(4):1003–1013, 1991. 247

[264] Z. Safra and I. Zilcha. Efficient Sets With and Without the Expected Utility Hypothesis. *Journal of Mathematical Economics*, 17(4):369–384, 1988. 54

[265] S. Saks. *Theory of the Integral.* 2^{nd} ed. Monografie Matematyczne, Warsaw; English transl. L.C. Young, 1937. 223, 226, 227, 239, 244

[266] L.J. Savage. *The Foundations of Statistics (2nd revised edition) – 1^{st} ed. 1954.* New York: Dover Publications, 1972. v, 3, 7, 8, 155, 167

[267] W. Schachermayer. Norm Attaining Operators and Renormings of Banach Spaces. *Israel Journal of Mathematics*, 44(3):201–212, 1983. 250

[268] W. Schachermayer. The Sum of Two Radon-Nikodym-Sets Need Not Be a Radon-Nikodym-Set. *Proceedings of the American Mathematical Society*, 95(1):51–57, 1985. 247

[269] W. Schachermayer. The Radon-Nikodym Property and the Krein-Milman Property are Equivalent for Strongly Regular Sets. *Transactions of the American Mathematical Society*, 303(2):673–687, 1987. 247

[270] H.H. Schaefer. *Topological Vector Spaces - 2^{nd} edition.* Springer-Verlag, 1999. 106, 215, 247

[271] E.E. Schlee. The Comparative Statics of Deductible Insurance in Expected- and Non-Expected-Utility Theories. *The GENEVA Papers on Risk and Insurance Theory*, 20(1):57–72, 1995. 53

[272] H. Schlesinger. The Optimal Level of Deductibility in Insurance Contracts. *Journal of risk and insurance*, 48(3):465–481, 1981. 53

[273] H. Schlesinger. Insurance Demand without the Expected-Utility Paradigm. *Journal of Risk and Insurance*, 64:19–39, 1997. 54

[274] D. Schmeidler. Integral Representation without Additivity. *Proceedings of the American Mathematical Society*, 97(2):255–261, 1986. 4, 188

[275] D. Schmeidler. Subjective Probability and Expected Utility without Additivity. *Econometrica*, 57(3):571–587, 1989. 4, 8, 9, 114, 119, 126, 188, 189, 198, 201

[276] U. Schmidt and H. Zank. What is Loss Aversion? *The Journal of Risk and Uncertainty*, 30(2):157–167, 2005. 190, 191

[277] L. Schwartz. *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures*. Oxford University Press, 1973. 239

[278] S. Shavell. On Moral Hazard and Insurance. *The Quarterly Journal of Economics*, 93(4):541–562, 1979. 54, 160

[279] B. Sinclair-Desgagné. The First-Order Approach to Multi-Signal Principal-Agent Problems. *Econometrica*, 62(2):459–465, 1994. 160

[280] B. Sinclair-Desgagné. Ancillary Statistics in Principal-Agent Models. *Econometrica*, 77(1):279–281, 2009. 160

[281] P. Slovic and S. Lichtenstein. Relative Importance of Probabilities and Payoffs in Risk Taking. *Journal of Experimental Psychology*, 78(3):56–67, 1968. 192

[282] P. Slovic and A. Tversky. Who Accepts Savage's Axiom? *Behavioral Science*, 19(6):349–373, 1974. 186, 187

[283] A. Smith. *The Wealth of Nations*. Edinburgh, 1776. 5, 6

[284] R.C. Sprowls. Psychological-Mathematical Probability in Relationships of Lottery Gambles. *The American Journal of Psychology*, 66(1):126–130, 1953. 192

[285] C. Stegall. Optimization of Functions on Certain Subsets of Banach Spaces. *Mathematische Annalen*, 236(2):171–176, 1978. 250

[286] C. Stegall. The Radon-Nikodym Property in Conjugate Banach Spaces. II. *Transactions of the American Mathematical Society*, 264(2):507–519, 1981. 247

[287] J.E. Stiglitz. Monopoly, Non-Linear Pricing and Imperfect Information: The Insurance Market. *The Review of Economic Studies*, 44(3):407–430, 1977. 54, 55

[288] J.E. Stiglitz. Risk, Incentives and Insurance: The Pure Theory of Moral Hazard. *The Geneva Papers on Risk and Insurance*, 8(1):4–33, 1983. 54, 160

[289] M. Talagrand. Dual Banach Lattices and Banach Lattices with the Radon-Nikodym Property. *Israel Journal of Mathematics*, 38(1):46–50, 1981. 247

[290] N.N. Taleb. *The Black Swan.* Random House Inc. 2008. 90

[291] A. Tauber. Uber Risiko und Sicherheitszuschlag. *Transactions of the Sixth International Congress of Actuaries I*, pages 781–842, 1909. 5

[292] F. Topsøe. *Topology and Measure.* Springer-Verlag, 1970. 110

[293] A. Tversky and D. Kahneman. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *The Journal of Risk and Uncertainty*, 5(4):297–323, 1992. 9, 138, 144, 189, 190, 193, 196, 198

[294] C. Villani. *Topics in Optimal Transportion.* AMS Bookstore, 2003. 31

[295] C. Villani. *Optimal Transport: Old and New.* Springer, 2009. 31

[296] C. Villegas. On Qualitative Probability σ-Algebras. *The Annals of Mathematical Statistics*, 35(4):1787–1796, 1964. 15, 57, 179

[297] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, 1947. 2, 3, 6, 167, 173

[298] L. Walras. *Eléments d'Economie Politique Pure, ou Théorie de la Richesse Sociale.* Lausanne, 1874. 5

[299] S. Willard. *General Topology.* Addison-Wesley Publidhing Company, 1970. 108, 203

[300] A. Willett. *The Economic Theory of Risk and Insurance.* Columbia University Press, 1901. 5

[301] C. Wilson. A Model of Insurance Markets with Incomplete Information. *Journal of Economic Theory*, 16(2):167–207, 1977. 54

[302] G. Winkler. Extreme Points of Moment Sets. *Mathematics of Operations Research*, 13(4):581–587, 1988. 110

[303] G. Winkler. Moment Sets of Bell-Shaped Distributions: Extreme Points, Extremal Decomposition and Chebysheff Inequalities. *Mathematische Nachrichten*, 215(1):161–184, 2000. 110

[304] V.R. Young. Optimal Insurance under Wang's Premium Principle. *Insurance Mathematics and Economics*, 25(2):109–122, 1999. 54

[305] H. Zank. On Probabilities and Loss Aversion. *Theory and Decision*, 68(3):243–261, 2010. 190, 191

[306] J. Zhang. Subjective Ambiguity, Expected Utility and Choquet Expected Utility. *Economic Theory*, 20(1):159–181, 2002. 188

[307] I. Zilcha and S.H. Chew. Invariance of the Efficient Sets when the Expected Utility Hypothesis is Relaxed. *Journal of Economic Behavior and Organization*, 13(1):125–131, 1990. 54

Glossary

S	The set of all states of the world, 14.
$\mathcal{G}$	The σ-algebra of events on S, 14.
$B(\mathcal{G})$	The supnorm-normed Banach space of all bounded, $\mathbb{R}$-valued and $\mathcal{G}$-measurable functions on $(S, \mathcal{G})$, 14).
$\|f\|_s$	For a given $f \in B(\mathcal{G})$, $\|f\|_s$ denotes the supnorm of f, i.e. $\|f\|_s := \sup\{\lvert f(s)\rvert : s \in S\} < +\infty$, 14.
X	The underlying uncertainty (or innovation). X is a fixed element of $B^+(\mathcal{G})$ with a closed range $[0, M] := X(S)$, where $M := \|X\|_s < +\infty$, 14.
Σ	The σ-algebra $\sigma\{X\}$ of subsets of S generated by X, 14.
$ba(\Sigma)$	The linear space of all bounded finitely additive set functions on a measurable space (S, Σ), endowed with the usual mixing operations. $ba^+(\Sigma)$ denotes the set of nonnegative elements of $ba(\Sigma)$, and $ba_1^+(\Sigma)$ denotes the collection of those elements μ of $ba^+(\Sigma)$ for which $\mu(S) = 1$. Elements of $ba_1^+(\Sigma)$ are the finitely additive probability charges on (S, Σ). When endowed with the total variation norm $\|.\|_v$, $ba(\Sigma)$ is a Banach space, 100.
$ca(\Sigma)$	The linear subspace of $ba(\Sigma)$ consisting of countably additive set functions. When endowed with the variation norm $\|.\|_v$, $ca(\Sigma)$ is a Banach space. In particular, $ca(\Sigma)$ is a $\|.\|_v$-closed linear subspace of $ba(\Sigma)$. $ca^+(\Sigma)$ denotes the collection of nonnegative elements of $ca(\Sigma)$, and $ca_1^+(\Sigma)$ denotes the collection of those elements ν of $ca^+(\Sigma)$ for which $\nu(S) = 1$. Elements of $ca_1^+(\Sigma)$ are the countably additive probability measures on (S, Σ), 100.

$\mathcal{C}_{\mu}$	The collection of all $\nu \in ca_1^+(\Sigma)$ that are (μ, X)-vigilant beliefs, 101.
$\oint \psi \, d\nu$	The Choquet integral of $\psi \in B(\mathcal{G})$ with respect to a capacity ν on $(S, \mathcal{G})$, 116.
$\oint \phi \, d\nu$	The Šipoš integral of $\psi \in B(\mathcal{G})$ with respect to a capacity ν on $(S, \mathcal{G})$, 146.
$\mathcal{F}_{SB}$	The feasibility set of a given problem, 38.
$\mathcal{F}_{SB}^{\uparrow}$	The collection $\left\{ Y = I \circ X \in \mathcal{F}_{SB} : I \text{ is nondecreasing} \right\}$, 38.
LR	The likelihood ratio, 47.
MLR	The monotone likelihood ratio condition, 48.
$\mathcal{AQ}$	The collection of "admissible" quantile functions, 126.
$\widetilde{Y}_P$	The nondecreasing P-rearrangement of $Y \in B^+(\sigma\{X\})$ with respect to $X \in B^+(\mathcal{G})$, where P is a probability measure on $(S, \mathcal{G})$, 27.
DM	The decision maker, 11.
CI	The claim issuer, 11.
$\succcurlyeq_{DM}$	The preferences of DM over $B^+(\Sigma)$, 15.
$\succcurlyeq_{CI}$	The preferences of CI over $B^+(\Sigma)$, 15.
W_0	The DM's initial wealth, 16.
W_0^{CI}	The CI's initial wealth, 16.
(H, Y)	The contract between the DM and the CI, with $H > 0$ and $Y \in B^+(\Sigma)$, 16.
$W^{DM}(H, Y)$	The DM's wealth after entering into the contract (H, Y), 16.

$W^{CI}(H, Y)$	The CI's wealth after entering into the contract (H, Y), 17.
u	The DM's utility function, 15.
v	The CI's utility function, 15.
μ	The set function that represents the DM's subjective beliefs, 15.
ν	The set function that represents the CI's subjective beliefs, 15.