Prediction Markets: Theory and Applications

A dissertation presented

by

Michael Edward Ruberry

to

The School of Engineering and Applied Sciences
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Computer Science

Harvard University
Cambridge, Massachusetts
November 2013

Dissertation Advisor: Professor Yiling Chen   Michael Edward Ruberry

Prediction Markets: Theory and Applications

Abstract

In this thesis I offer new results on how we can acquire, reward, and use accurate predictions of future events. Some of these results are entirely theoretical, improving our understanding of strictly proper scoring rules (Chapter 3), and expanding strict properness to include cost functions (Chapter 4). Others are more practical, like developing a practical cost function for the $[0, 1]$ interval (Chapter 5), exploring how to design simple and informative prediction markets (Chapter 6), and using predictions to make decisions (Chapter 7).

Strict properness is the essential property of interest when acquiring and rewarding predictions. It ensures more accurate predictions are assigned higher scores than less accurate ones, and incentivizes self-interested experts to be as accurate as possible. It is a property of associations between predictions and the scoring functions used to score them, and Chapters 3 and 4 are developed using convex analysis and a focus on these associations; the relevant mathematical background appears in Chapter 2, which offers a relevant synthesis of measure theory, functional analysis, and convex analysis.

Chapters 5–7 discuss prediction markets that are more than strictly proper. Chapter 5 develops a market for the $[0, 1]$ interval that provides a natural interface, is computable, and has bounded worst-case loss. Chapter 6 offers a framework to understand how we can design markets that are as simple as possible while still providing

an accurate prediction. Chapter 7 extends the classical prediction elicitation setting to describe decision markets, where predictions are used to advise a decision maker on the best course of action.

Contents

Citations to Previously Published Work

Portions of the mathematical background and discussion of scoring relations in Chapter 2, as well as the entirety of Chapter 3, was developed from or appears in

> "Cost Function Market Makers for Measurable Spaces", Yiling Chen, Mike Ruberry, and Jenn Wortman Vaughan, *Proceedings of the 14th ACM Conference on Electronic Commerce (EC)*, Philadelphia, PA, June 2013.

Most of Chapter 4 previously appeared in

> "Designing Informative Securities", Yiling Chen, Mike Ruberry, and Jenn Wortman Vaughan, *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence (UAI)*, Catalina Island, CA, August 2012.

Most of Chapter 5 previously appeared in the following journal paper, which precedes the published conference paper listed below

> "Eliciting Predictions and Recommendations for Decision Making", Yiling Chen, Ian A. Kash, Mike Ruberry, Victor Shnayder, *Revise and resubmit to ACM Transactions on Economics and Computation (TEAC)*, February 2013.

> "Decision Markets with Good Incentives", Yiling Chen, Ian A. Kash, Mike Ruberry, and Victor Shnayder, *Proceedings of the 7th Workshop on Internet and Network Economics (WINE)*, Singapore, December 2011.

Acknowledgments

Thank you to my big-hearted adviser Yiling Chen and the other members of my thesis committee, Jenn Wortman Vaughan and David Parkes.

Thank you to Joan Feigenbaum, who introduced me to economic computation.

Thank you to my parents, Ed and Sarah Ruberry.

Thank you to my co-authors, Yiling Chen, Jenn Wortman Vaughan, Sven Seuken, Ian Kash, Victor Shnayder, Jon Ullman and Scott Kominers.

Thank you to those who introduced me to research, Jay Budzik, Sara Owsley and Ayman Shamma.

Thank you Shadi, for pushing me on.

We choose to go to the moon in this decade and to do these other things not because they are easy, but because they are hard, because that goal will serve to organize and measure the best of our energies and skills, because that challenge is one that we are willing to accept, one we are unwilling to postpone, and one which we intend to win.

—President John F. Kennedy

Come, my friends,
'Tis not too late to seek a newer world.
Push off, and sitting well in order smite
The sounding furrows; for my purpose holds
To sail beyond the sunset, and the baths
Of all the western stars, until I die.
It may be that the gulfs will wash us down;
It may be we shall touch the Happy Isles,
And see the great Achilles, whom we knew.
Though much is taken, much abides; and though
We are not now that strength which in old days
Moved earth and heaven, that which we are, we are,
One equal temper of heroic hearts,
Made weak by time and fate, but strong in will
To strive, to seek, to find, and not to yield.

—Lord Alfred Tennyson's Ulysses

This thesis is dedicated to my father, Edward Ruberry, who resolutely seeks and accepts the greatest challenges.

From his son, Mike.

1

Introduction

All appearances being the same, the higher the barometer is, the greater the probability of fair weather.

– John Dalton, 1793[1]

...there has been vague demand for [probabilistic weather] forecasts for several years, as the usual inquiry made by the farmers of this district has always been, "What are the chances of rain?"

– Cleve Hallenbeck, 1920[2]

Verification of weather forecasts has been a controversial subject for more than half a century. There are a number of reasons why this problem has been so perplexing to meteorologists and others but one of the most important difficulties seems to be in reaching an agreement on the specification of a scale of goodness for weather forecasts. Numerous systems have been proposed but one of the greatest arguments raised against forecast verification is that forecasts which may be the "best" according to the accepted system of arbitrary scores may not be the most useful forecasts.

– Glenn W. Brier, 1950[3]

One major purpose of statistical analysis is to make forecasts for the future and provide suitable measures for the uncertainty associated with them.

– Gneiting & Raftery, 2007[4]

---

[1] From [27], see also [61] for a discussion of the history of probabilistic weather forecasts.

[2] From [46].

[3] All of Brier's quotes are from [16].

[4] From [43].

This thesis studies the now classical problem of how we elicit and score predictions about the future and some of its practical extensions. This problem is motivated by a natural desire to *acquire an accurate prediction about the likelihood of future events from one or more self-interested (risk-neutral and expected score maximizing) experts*, or – equivalently – a desire to *devise a system for scoring experts that rewards accurate predictions.* The formal study of this problem first came from meteorology, with its interest in predicting tomorrow's weather, and is now often studied independently. Systems designed to elicit accurate predictions of the future have been used to predict everything from presidential elections to technology trends, and it appears they produce better predictions than some common alternatives.[17, 22]

After about sixty years of study there are still significant challenges to our understanding of how we score predictions. Some of these challenges are theoretical – we lack a complete understanding of how to relate our problem to various mathematical objects – and many others are practical—some systems for eliciting accurate predictions are too complicated to be used in practice, and actually making use of a prediction can be surprisingly difficult. This thesis addresses some of these challenges, offering a new theoretical perspective on how we score predictions and examining several practical problems: (1) the creation of a practical securities market for events occurring in the [0,1] interval, (2) the construction of simple and informative markets, and (3) the use of predictions for decision making. Chapters 2–4 are more theoretical, presenting some mathematical background and then characterizing strictly proper scoring rules and cost functions, and Chapters 5–7 present the three more practical investigations. The rest of this introduction provides an overview of these chapters.

1.1 Convex Functions and Relations

Chapter 2 synthesizes concepts from measure theory, functional analysis, and convex analysis, to provide the mathematical tools and perspective needed in Chapters 3–5. It formalizes our discussion of predictions and scores, and shows how we can study associations between them using convex analysis. These associations describe how we score predictions, and will be the fundamental objects of study in Chapters 3 and 4.

Chapter 2 also develops some specialized new tools that let us succinctly describe strictly proper associations between predictions and scoring functions, associations that are the subject of Chapter 3.

1.2 Scoring Rules

Chapter 3 describes *strictly proper scoring rules*. Scoring rules are a popular method for acquiring predictions about the future, and strict properness is the *essential* property that guarantees they elicit and reward *accurate* predictions. These rules define an association between predictions and a means of scoring them, known as *scoring functions*, and strict properness is a property of the structure of these relations. Using the tools developed in Chapter 2, Chapter 3 identifies this strictly proper structure as always being a subset of a relation described by convex functions.

When using a scoring rule ask an "expert," like a meteorologist, to offer a prediction of the likelihood of future events, like whether or not it will rain tomorrow. A scoring rule assigns this prediction a scoring function that maps each possible out-

comes to a score. When a meteorologist is predicting the likelihood of rain there are two outcomes, RAIN and NO RAIN, and a prediction is a probability distribution over these possible outcomes. A scoring rule assigns the meteorologist's prediction a scoring function b, and if it RAINS the expert is scored $b(RAIN)$ and otherwise $b(NO\ RAIN)$.

If a scoring rule is strictly proper, then an expert expects to maximize its score only when it offers the most accurate prediction possible. Alternatively, a strictly proper scoring rule rewards accurate predictions more in expectation. If our meteorologist thinks the likelihood of rain is 70% then a strictly proper scoring rule provides a strict incentive for it to also predict a 70% likelihood. If a scoring rule is not strictly proper then our meteorologist may expect to maximize its score by predicting 50% instead, and this less accurate prediction might be rewarded just as much as or more than the more accurate one! Simply put, scoring rules that are not strictly proper fail our goal of eliciting and rewarding accurate predictions. This is why strict properness is the *essential* property we need when eliciting and scoring predictions, and this point cannot be emphasized enough.[5]

Strictly proper scoring rules have been studied heavily, ever since Brier proposed a scoring system for weather predictions he thought would encourage and reward accurate predictions [16]. Savage later characterized strictly proper scoring rules that could handle countable outcome spaces [83], and Gneiting and Raftery described them for arbitrary measurable spaces [43]. Both Savage's and Gneiting and Raftery's characterizations identify strictly proper scoring rules with strictly convex functions,

---

[5]I think methods of acquiring a prediction that are not strictly proper have some serious explaining to do.

essentially showing that a strictly proper scoring rule's association between predictions and scoring functions is described by a strictly convex function's association between points and their "subtangents."[6] This characterization is not as elegantly stated as I have paraphrased it, and, from my perspective, it has real deficits:

1. It provides little insight into *why* strictly proper scoring rules and strictly convex functions are related.

2. It uses subtangents, atypical mathematical objects that are not part of convex analysis

3. It requires the class of predictions considered is convex. Equivalently, it only allows strictly proper scoring rules with convex instead of arbitrary domains.

4. It allows scoring rules to assign scores of negative infinity to some experts, and these scores cannot be assigned in a prediction market.[7]

5. It does not suggest a way of expanding strict properness to *cost functions*, another popular method of scoring predictions. (Discussed in the next chapter.)

Gneiting and Raftery were not attempting to address these perceived deficits; they were certainly not trying to create a perspective on strict properness that would also cover cost functions! My point here is that there is room to improve our fundamental characterization of strictly proper scoring rules.

---

[6]See Chapter 3 for a more detailed analysis of Gneiting and Raftery's characterization.

[7]In a prediction market it is necessary to take the difference of two scores. The difference of negative infinity and negative infinity is undefined, and so scoring functions that assign negative infinity would result in an ill-defined market.

By approaching strict properness from the perspective provided in Chapter 2, Chapter 3 quickly arrives at a distinct characterization that shows a strictly proper scoring rule's mapping from predictions to scoring functions is always a subset of a convex function's mapping from its points to their *unique supporting subgradients*. This is very similar to Gneiting and Raftery's characterization, and it has the following advantages:

1. It clarifies the relationship between scoring rules, a type of relation, and convex functions, which are a useful tool for understanding the structure of relations.

2. It uses the idea of "supporting subgradients" instead of "subtangents," and the former is part of convex analysis.

3. It lets strict properness apply to any class of predictions, not just convex ones.

4. It restricts scores to be real-valued, letting these scores always usable by a prediction market.

5. It offers a framework for extending strict properness to cost functions.

This second-to-last point may also be seen as a negative, since Gneiting and Raftery's characterization is more general by allowing more scores. Written negatively, the last item might read: *scoring functions can no longer assign a value of negative infinity.* This is a consequence of using supporting subgradients and the tools of convex analysis instead of subtangents. Not being a complete generalization, I think both characterizations are still of interest, and I hope my own offers the reader some new insights for their own work.

1.3 Cost Functions

Cost functions are another popular means of acquiring a prediction. These functions are especially interesting since they can emulate futures markets where (one or more) traders buy and sell securities whose values anticipate future events. While scoring rules can also be used to run markets with many experts "trading" predictions, trading securities using a cost function has two significant advantages over using a scoring rule: (1) it presents a familiar interface to traders, and (2) it lets traders focus exclusively on their areas of expertise.[8] Instead of having to predict the likelihood of every future event, a cost function lets traders focus on determining whether a few securities are priced correctly. The cost function, acting as a market maker, assumes the role of translating trading behavior into a complete prediction of future events.

Futures markets have been implicitly acquiring and rewarding predictions of the future since they were first opened. The better a trader can predict the price of corn the more it expects to make trading corn futures. These markets naturally provide the same incentives a strictly proper scoring rule does for traders to acquire information and produce predictions that are as accurate as possible. Well-designed cost functions can let us act as market makers who emulate these futures markets.

Prior work on cost functions has usually developed them to have desirable economic properties, to be efficiently computable,[9] or to make some theoretical connections

---

[8]Generalizations of the scoring rules considered in Chapter 3, like those discussed in Chapter 6, can also allow traders to focus in this way. Classically, however, we think of scoring rules request an entire probability measure.

[9]When running a market with billions (or more) possibilities, accounting for the effects of one trade on the system can be very difficult. For instance, if running a market to determine the next U.S. president, it can be hard to understand how to increase the likelihood of a Democratic win if traders begin purchasing the security that says they will win in Iowa. Some excellent work on this

with other fields, especially machine learning [1]. This work has also often revealed connections between cost functions and scoring rules [1, 3], yet the idea of a strictly proper cost function was never formally developed.[10] It has also proven difficult to adapt cost functions to measurable spaces, and most work on them considers discrete spaces.

Chapter 4 characterizes strictly proper cost functions on arbitrary measurable spaces for the first time. This characterization puts our understanding of cost functions in parity with our understanding of scoring rules, and completely reveals the relationship between the two. It does this by developing the perspective on scoring rules in Chapter 3 into a more general object that I call a "scoring relation." These scoring relations are the root object in the study of strict properness, and both scoring rules and cost functions are derived from them.

Perhaps surprisingly, given our discussion so far, a cost function must be more than strictly proper to emulate a futures market. Chapter 5 discusses the additional structure required while developing a new cost function for continuous random variables.

---

problem is [53].

[10]The authors of [1] effectively show the cost functions they consider are strictly proper when they demonstrate the mathematical connections these cost functions have to strictly proper scoring rules. The concept of strict properness has been so alien to cost functions, though, that the authors do not elaborate on the incentive implications of this result.

1.4   A Cost Function for Continuous Random Variables

Chapter 5 continuous our discussion of cost functions. In Chapter 4 strictly proper cost functions were described, and Chapter 5 begins by characterizing when these functions actually emulate a futures market. In addition to being strictly proper, emulating a futures market requires cost functions reliably offer traders a consistent set of securities to buy and sell, and that they can quote meaningful prices for any bundle of securities. These are natural properties we expect any market to have.

The second part of Chapter 5 uses the techniques developed to produce a practical cost function for continuous random variables. Cost functions for non-discrete spaces have, historically, proven elusive. In [38] a set of economic properties was proposed, as we expect from work on cost functions, and it was shown that cost functions satisfying these properties must experience unbounded *worst-case loss* when working with continuous random variables. Unbounded worst-case loss means that our market maker can lose any amount of money, and this is an undesirable property to have in practice. In [67] a cost function for continuous random variables with bounded loss was incorrectly claimed, a claim withdrawn in the author's thesis [66]. These difficulties have caused prior work to discretize the outcome space of continuous random variables, or offer alternative interfaces other than a traditional cost function [68, 37].

Chapter 5 uses my characterization of strictly proper cost functions for arbitrary measurable spaces to create a market for the outcome of a (bounded) continuous random variable that (1) is strictly proper, (2) acts like a futures market, (3) has

bounded worst-case loss, and (4) can be computed using a convex program. This cost function is not perfect. It does not let traders buy and sell any security, and it is incapable of representing every possible prediction. Still, it is an interesting first step in our development of cost functions for continuous random variables, and may even be considered suitable for real use.

Chapter 5 concludes my discussion of strict properness in measurable spaces. Chapters 6 and 7 continue, like Chapter 5, to discuss markets that are more than strictly proper. The first of these chapters, Chapter 6, asks how we can design prediction markets that are simple and informative, and the second, Chapter 7, investigates how we can use expert advice to make decisions.

1.5   Simple and Informative Markets

Chapter 6, like Chapter 5, focuses on a prediction market that is more than strictly proper. In this chapter I assume a finite outcome space and Bayesian traders, with a common prior and known information structure. Our prediction market offers a set of securities, and Chapter 6 is interested in designing markets that are both *simple* and *informative*.

A market is informative if (1) traders are able to converge on security prices that reflect all their private information, and (2) we are able to uniquely infer from these prices the likelihood of some events of interest. This first property has been studied by [65], which showed a *separability* condition was necessary. In brief, this condition related the available securities to the structure of traders' information. Securities are the medium through which traders exchange ideas and debate in markets, and if

they are cleverly structured then traders are able to accurate deduce the likelihood of future events. Sometimes, however, this is not possible.

Consider, for example attempting to determine the future price of corn. Corn prices are determined by a variety of factors, like the weather and future demand, and if we understood these variables we could offer securities to determine how much it would rain, and how much demand there would be. The prices of these securities would then let traders better determine future corn prices. If we just offer a security for the future price of corn, traders would be unable to express their information about the weather, future demand for corn, etc., and the result is a less accurate prediction of future corn prices.

The second property of informativeness is straightforward: the security prices must actually be usable. This prevents us from mistakes like running a trivial market with, for instance, a constant-valued security. Traders are always able to price this security perfectly and it always tells us nothing. Thus informativeness is composed of two properties.

Returning to our future corn price example, we might think one solution to best determining the future price of corn is offering as many securities as possible, one for every possible event. This would allow traders to express a great deal of information, and the market would be very difficult, in practice, to run. Broadly speaking, the more securities a market offers the more computationally complex it becomes to run, and too many securities is computationally prohibitive. Some excellent work on making tractable markets that can handle large outcome spaces is [31, 53].

Because too many securities is computationally prohibitive, then, when designing

a market we think of both informativeness and simplicity. These are markets that informative and that use as few natural securities as possible, securities that either pay \$0 or \$1. This prevents us from offering superfluous securities, as well as especially strange securities real traders are unlikely to want to work with.

How we consider designing a market that is both simple and informative depends on our knowledge of traders' signal structure, and Chapter 6 has two significant results. The first shows that without any knowledge of how traders' information is structured a potentially huge number of securities is necessary to best identify the likelihood of a future event, as many securities as outcomes that comprise the event or its complement. The second shows that when we know traders' signal structure, designing a simple and informative market is NP-hard. Thus, designing a simple and informative market is either trivial and does not help us reduce the computational complexity of a market, or we actually have the chance of reducing a market's complexity but doing so perfectly is NP-hard.

In the end, these results that simple prediction markets likely work because information is being exchanged outside the market, or traders' information is already very simple. In our corn example traders might be receiving weather reports instead of relying on weather securities. Given the hardness of usefully designing a market that is both simple and informative, and how unlikely it is that we perfectly know traders' information structure, this chapter likely raises more questions about designing prediction markets than it answers.

1.6 Making Decisions with Expert Advice

Chapter 7 concludes my new results with an investigation of how we can use expert advice to help make decisions. Acquiring predictions of the future is, after all, only useful if it might change how we act today—if it can influence some decision we are making. The idea of a "decision market" where prediction markets would influence policy decisions was first proposed in [47], and formally studied for the first time in [69]. This latter paper revealed a tension between acting on decisions and ensuring their accuracy, and they discussed a solution for a special case of the problem.

In the first part of Chapter 7 I will fully characterize strictly proper *decision markets*, which incentivize accurate predictions just like strictly proper prediction markets. These markets consider a decision maker trying to choose between several available actions. Experts are then asked to predict what *would* happen if each action were taken. For example, a prediction of the likelihood of future events conditional on action A being taken, and another prediction of the likelihood of future events conditional on action B being taken. The decision maker can then review these predictions to assist in picking what it thinks is the best possible action it can take.

Chapter 7 shows that strictly proper decision markets exist, and can be readily built from traditional strictly proper scoring rules. Unfortunately, they also require the decision maker risk taking *any* action with *some* (arbitrary small) probability. Since this probability can be made as small as desired, this limitation still means a decision maker can use a decision market to *improve* the chances it makes a good decision.

The second part of Chapter 7 talks about decision making using the advice of

a single expert. Here it is possible to simply take a recommended option, and *recommendation rules* can be constructed to incentivize the expert to reveal the option the decision maker would most prefer. These recommendation rules are an interesting departure from scoring rules since they are not necessarily designed to reward more accurate predictions. Instead, they might give the expert a share of the decision maker's utility for the actual outcome, aligning the expert's incentives with the decision maker's. Recommendation rules are mathematically similar to scoring rules, even if conceptually different, and they suggest there may be other uses for the techniques developed in Chapters 2–4.

2

Mathematical Background

This chapter offers a relevant synthesis of some concepts from measure theory, functional analysis, and convex analysis needed in Chapters 3–5. An excellent introductory measure theory book is [7], an excellent introductory functional analysis text is [52], and a very interesting book on convex analysis is [9].

This chapter begins in Section 2.1 by showing how measure theory is an appropriate language for scoring predictions. The events we would like to predict are represented by a measurable space, predictions are probability measures, and scoring functions are bounded measurable functions. Section 2.2 shows how predictions (probability measures) and scoring functions (bounded measurable functions) can be placed in duality, and how each is actually a continuous linear function of the other. Section 2.3 shows how convex functions can be used to study relations between objects in duality, like predictions and scoring functions, and develops some refinements particular to our work. In particular, it concludes with a description of a the relation between a convex function's points and their unique supporting subgradients, a

relation that we will see describes all of strict properness.

2.1 Measures, Measurable Spaces, Sets and Functions

When using a scoring rule we start with *something we would like to predict,* then we *acquire a prediction* and *assign it a scoring function* that describes how it will be scored. Afterwards we *observe the actual outcome* and use the scoring function to *assign the prediction a score.* In this section I will formalize each of these steps using concepts from measure theory, assisted by two running examples. The first will be of a meteorologist predicting the likelihood of rain tomorrow, and this will allow us to use and compare our intuition from discrete probability theory with the measure theory; the second example will be of a statistician predicting the outcome of a continuous random variable on the [0, 1] interval, a more abstract instance that requires measure theory understand.

2.1.1 Measurable Spaces and Sets

We will represent the possible outcomes of what we would like to predict as an arbitrary *measurable space*, a tuple $(\Omega, \mathcal{F})$. This tuple consists of an *outcome space* Ω, a set that describes what may happen, and a $\sigma-$algebra $\mathcal{F}$, a set that describes the *measurable sets* of Ω. These measurable sets are the sets we can use a measure to assign a value ("size," "length," "mass") to, and are referred to as *measurable.* In our context a measurable set is also described as an *event.* A measurable space always

has at least one event, ensures the complement of any event is also an event, and requires that a countable union of events is also an event (and thus so are countable intersections of events).

Discrete probability theory does not explicitly define a $\sigma-$algebra. When Ω is a countable set, like {RAIN, NO RAIN}, it is natural to think of every subset being an event. Explicitly, such an outcome space Ω can be interpreted as belonging to the measurable space $(\Omega, 2^\Omega)$, and these spaces are the purview of discrete probability theory.

Measure theory was developed to work with countable and uncountable outcome spaces, like the [0, 1] interval, where assuming every subset is an event is mathematically problematic. The details of why this assumption is problematic is not important for our purposes, and we need only accept that $\sigma-$algebras are a mathematical necessity and that much of our intuition from discrete probability theory no longer applies in this setting. We will not encounter any subsets of interest that are not also events in this thesis, and we will never be interested directly in the structure of a $\sigma-$algebra; they are mostly carried around as notation.

A common way of quickly defining and forgetting a $\sigma-$algebra for familiar sets Ω is to generate one from a familiar or usual topology on Ω. A topology is a collection of open sets, just like a $\sigma-$algebra is a collection of measurable sets, that satisfies some similar properties we will not be concerned with. We are intuitively familiar with the "usual" Euclidean topology on the reals, where a basis of open sets are the open intervals, the empty set, and $\mathbb{R}$ itself, and the uncountable unions of these sets define the open sets that compose the topology. A Borel $\sigma-$algebra generated from

this topology is the smallest $\sigma-$algebra that contains every open set.

On the [0, 1] interval a more common $\sigma-$algebra is the Lebesgue measurable sets, which also contains every open set and so is a super set of the Borel $\sigma-$algebra. These sets are described in the next subsection along with Lebesgue measure.

2.1.2   Measures

In the previous subsection we represented the outcome space of what we would like to predict as a measurable space. This measurable space provided a structure of measurable sets or events that will let us describe how likely an event is, and a prediction will be a complete description of how likely each event is. More formally, a prediction will be a *probability measure*, a special type of measure.

Given a measurable space $(\Omega, \mathcal{F})$, a *measure* is any function that maps from the $\sigma-$algebra to the reals, $\mu : \mathcal{F} \to \mathbb{R}$. The probability measures are a special closed and convex subset of all measures that are non-negative, countably additive, and that assign a likelihood of one to Ω itself.[1] Countable additivity means that the sum of the likelihoods of countably many disjoint events is equal to the likelihood of the union of the disjoint events, $\sum_i \mu(F_i) = \mu(\cup_i F_i)$, $\forall i, j, F_i \cap F_j = \emptyset$. The set of probability measures is denoted $\mathcal{P}$.

With a discrete space, like $\Omega = \{\text{RAIN, NO RAIN}\}$, probability measures are also called probability distributions, and handling them is well understood. With an arbitrary measurable space it is not so clear what a probability measure looks

---

[1]Sometimes probability measures are allowed to be finitely additive, too. This may be an interesting extension for future work to consider. We are usually economically interested in the countably additive probability measures.

like. Luckily, in the case of the [0, 1] interval the probability measures have a very special and familiar structure. Understanding this first requires knowing a little about Lebesgue measure.

Lebesgue measure is a measure defined on the reals that acts as one might expect, assigning intervals a measure equal to their length. In fact, Lebesgue measure is "strictly positive," which means it assigns *every* open set of the interval a positive value. Lebesgue measure is usually denoted λ, and the Lebesgue measurable sets are denoted $\mathcal{L}$. We will not go into detail about this measure, suffice to say that they are a superset of the Borel measurable sets, and so contain all points, subintervals, and all their countable unions and finite intersections—every subset of interest on the [0, 1] interval. Thus we have statements like $\lambda([0,.5]) = \lambda(0,.5) = .5$, and $\lambda(.7) = 0$.

Lebesgue measure and the Lebesgue measurable sets are so important that we will always think of [0, 1] as part of the measurable space $([0,1], \mathcal{L})$. One nice thing about probability measures whose domain is the Lebesgue measurable sets is that these probability measures are identified with cumulative distribution functions (CDFs).[2] Every cumulative distribution function is such a probability measure, and every such probability measure is a cumulative distribution function. Lebesgue measure itself is the uniform "straight-line" 45 degree angle CDF.

Returning to our context of acquiring a prediction, we start with a measurable space $(\Omega, \mathcal{F})$ that represents the possible outcomes of what we would like to predict. We normally think of an expert have some *beliefs* $p \in \mathcal{P}$ of what they think most likely to occur, and they make a prediction $p' \in \mathcal{P}$. If $\Omega = \{\text{RAIN, NO RAIN}\}$

---

[2]Right-continuous functions of the [0, 1] interval that begin at zero and go to one.

then this prediction is a probability distribution, and if $\Omega = [0, 1]$ this prediction is a CDF. Strict properness is the property that attempts to make $p' = p$. That is, strict properness is about getting experts to tell us what they actually believe, or about scoring them higher (in expectation) when the expert is most accurate and does so (we take the expert's belief as the pinnacle of accuracy).

That beliefs and predictions over the $[0, 1]$ interval are equivalently CDFs will offer a great deal of useful structure that we will exploit in Chapter 5. Describing more of this structure will require understanding measurable functions, and conveniently these functions are also what we will use as *scoring functions* that determine what score to assign a prediction.

2.1.3 Measurable Functions

So far we have discussed measurable spaces, sets, and measures, especially probability measures. When acquiring a prediction, we think of a measurable space $(\Omega, \mathcal{F})$ describing the possible outcomes, and providing the structure necessary to define measures, like the probability measures, that represent an expert's beliefs and the predictions they can make. In this subsection we describe *measurable functions*, a subset of which we will use as our *scoring functions* that describe how we assign predictions a score.

Let $(\Omega_0, \mathcal{F}_0)$ and $(\Omega_1, \mathcal{F}_1)$ be two measurable spaces. A function $f : \Omega_0 \to \Omega_1$ is $\mathcal{F}_0/\mathcal{F}_1-$measurable when the inverse image of every measurable set is also a measurable set. When the measurable spaces are understood, such functions will be described simply as "measurable." This is analogous to the topological notion of

continuity, where a function is continuous when the inverse of each open set is open. If a function is continuous then it is measurable, and if a function is measurable it is "almost continuous," having at most a countable number of discontinuities.

Our scoring functions will be *bounded and measurable* functions from $(\Omega, \mathcal{F})$ to the reals with their Borel $\sigma-$algebra. The set of such functions is denoted $\mathcal{B}$, and (again) a member of this set is a function $b : \Omega \to \mathbb{R}$ that is measurable and bounded. Boundedness will be important in the next section, where we will need the supremum norm of our scoring functions $\sup_{\omega \in \Omega} |b(\omega)|$ to be well-defined. Note that, while any function $b \in \mathcal{B}$ is bounded above and below by some real k, the set itself is unbounded.

It is important our scoring functions be measurable, because this will allow us to take their expectation. If an expert has beliefs $p \in \mathcal{P}$, the expectation of a bounded measurable function $b \in \mathcal{B}$ is defined by the Lebesgue integral

$$p(b) \;=\; \int_\Omega b \; \mathrm{d}p \qquad \qquad \text{(Expectation / Lebesgue integral)}$$

This integral is a means of turning a countably additive measure, like p, into a function of measurable functions. The precise definition of the integral is too detailed for this overview; in the discrete setting we have a natural intuition about expectations, and the integral is best understood as such. In the continuous setting the integral is like the limit of a discrete expectation, and can be thought of as the values of the function b times the measure that p assigns to them.

When predicting the likelihood of rain, $\Omega = \{\text{RAIN, NO RAIN}\}$ and we interpret this outcome space as part of a discrete space. Our meteorologist has some belief about how likely rain is, and this is simply a probability distribution. Let's assume the meteorologist believes there is a 70% chance of rain, and let this measure be p.

We ask the meteorologist for a prediction $p' \in \mathcal{P}$, also a probability distribution, and assign it a scoring function $b \in \mathcal{B}$. The expert expects to score $p(b)$, its expectation for the scoring function. If $b(\text{RAIN}) = 1$ and $b(\text{NO RAIN}) = 0$, then this would be $p(b) = .7(1) + .3(0) = .7$. If RAIN occurs then the expert is scored $b(\text{RAIN}) = 1$.

When an expert is predicting the outcome of a continuous random variable its beliefs are a probability measure or CDF p, and it offers as a prediction another CDF p'. It receives a scoring function $b : [0, 1] \rightarrow \mathbb{R}$, and it expects to score $p(b)$. If b is one on $[0, .5]$ and zero on $(.5, 1]$, then $p(b) = p([0, 5])(1) + p((.5, 1])(0)$. If the outcome $.2$ occurs then the expert is scored $b(.2) = 1$.

This concludes most of the measure theory we will need in the following chapters. We have a way to represent what may happen, an understanding of beliefs and predictions as probability measures, and a knowledge of scoring functions as bounded and measurable functions. This is a formal representation of how using a scoring rule works, and in the next chapter I will focus on how we determine what scoring function b to pair with each prediction p. Before moving on to discuss Banach spaces and duality, however, it is convenient to now return to Lebesgue measure and how it relates to probability measures on the $[0, 1]$ interval. This structure will be needed in Chapter 5.

2.1.4 Lebesgue Measure as a Perspective

The measurable space $([0, 1], \mathcal{L})$ is so often of interest that we have a great deal of specialized tools available for analyzing it, and we will need these tools in Chapter 5 when we focus on acquiring predictions over the interval. As described earlier

in this section, probability measures on this interval are identified with cumulative distributions functions CDFs). Lebesgue measure is the CDF corresponding to the uniform distribution, and is a natural reference point for mathematical investigations. In this subsection we will discuss how other probability measures on relate to it.

A probability density function (PDF) is another way of describing *some* probability measures on the $[0, 1]$ interval. In the language of classical (not discrete) or calculus-based probability theory, a PDF is usually defined as a function $f : [0, 1] \to \mathbb{R}$ that is Riemann integral. The likelihood of an event is then the Riemann integral of this function over that event. For instance, the likelihood of the event $[.2, 4]$ would be

$$\int_{.2}^{.4} f \ \mathrm{d}x \qquad \text{(specifying likelihoods with a PDF)}$$

Probability measures that can be described with a PDF are called "absolutely continuous" in classical probability theory.

From a measure theory perspective, one measure μ is absolutely continuous with respect to another measure ψ when there exists a measurable function, usually written $\frac{\mathrm{d}\mu}{\mathrm{d}\psi} : [0, 1] \to [0, \infty)$, such that

$$\mu(L) \ = \ \int_L \frac{\mathrm{d}\mu}{\mathrm{d}\psi} \ \mathrm{d}\psi \qquad \text{(Radon-Nikodym derivative)}$$

and this function is known as the Radon-Nikodym derivative of μ with respect to ψ.[3] If a probability measure p is absolutely continuous with respect to Lebesgue measure, then its Radon-Nikodym derivative with respect to Lebesgue measure is then called

---

[3]I am misrepresenting the math a little here in a simplification that avoids notions like σ−finiteness. It would be more accurate here to say "any measure we might consider is absolutely continuous with respect to another one we might ever consider when...."

its PDF. To avoid a proliferation of "with respect to Lebesgue measure"s from appearing, I will adopt the classical probability theory perspective that assumes Lebesgue measure as a reference point. That is, I will also start referring to measures simply as "absolutely continuous," and we will understand it is with respect to Lebesgue measure.

Note that the change of integral from the Riemann to the Lebesgue here is a minor issue, since while more functions are Lebesgue-integrable than Riemann-integrable, the Riemann integral is equivalent to Lebesgue integration with respect to Lebesgue measure wherever the former is defined.

Measures that are absolutely continuous do not have unique PDFs, and as mentioned not every probability measure has a PDF. In particular, probability measures with point masses do not have PDFs. (These are measures that assign positive mass to a single real number.[4] There are also *singular continuous* measures, which do not have point masses and are still not absolutely continuous. These measures are difficult to work with (an example of a singular continuous measure is the probability measure that has uniform support on the Cantor set[5]) and we will, in fact, exclude them from our consideration in Chapter 5 when designing a practical system for acquiring predictions on the [0, 1] interval.

While not every measure has a PDF, every measure on the interval can be partitioned from the perspective of Lebesgue measure in what is known as a *Lebesgue decomposition.* This partition results in three measures, one consisting only of point

---

[4]This is why probability measures are *countably* additive, and not simply additive. Many probability measures on [0, 1], like Lebesgue measure, assign a likelihood of zero to *every* point.

[5]Good luck trying to draw that CDF.

masses known as a *pure point* part, an absolutely continuous part, and a singular continuous part. Further, the pure point part has a countable number of point masses, and this fact and this decomposition will be used in Chapter 5. In fact, we can immediately derive the fact that the pure point part has a countable number of point masses because every point mass is a discontinuity in a CDF, and since CDFs are right-continuous they have at most a countable number of discontinuities. Results like this demonstrate the utility of working with probability measures on the interval, where we can leverage the structure of CDFs.

Before concluding our discussion of measure theory and moving on to functional analysis, I will prove that we can create a strictly convex function of the absolutely continuous probability measures over the interval by using a strictly convex function of the reals $f : \mathbb{R} \to \mathbb{R}$. Formally:

Lemma 1 (Strictly Convex Functions of Absolutely Continuous Measures)**.** *Letting $f : \mathbb{R} \to \mathbb{R}$ be a strictly convex function, the function*

$$\psi(\mu) \;=\; \int_{[0,1]} f(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}) \, \mathrm{d}\lambda$$

is a strictly convex function of measures μ over $([0,1], \mathcal{L})$ that are absolutely continuous, where $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ is the Radon-Nikodym derivative of μ with respect to Lebesgue measure.

To prove this I will use the following lemma.

Lemma 2 (CDF Distinguishability)**.** *Any two CDFs F and G on $[0,1]$ such that $\exists x \in [0,1]$ such that $F(x) \neq G(x)$ must differ on a non-empty open set.*

Proof. We begin by showing distinct right-continuous functions differ on a non-empty open set, then applying this results to CDFs.

Let f and g be two right-continuous functions defined on $[a, b) \in \mathbb{R}$. Assume there exists $x \in [a, b)$ such that $f(x) \neq g(x)$. Let $c = f(x) - g(x)$, then by right-continuity there exists $\delta_f, \delta_g > 0$ such that $f(x) - f(x') < c/2$ for all $x' \in (x, x + \delta_f)$, and symmetrically for g. Let $\delta = \min(\delta_f, \delta_g)$, then on the interval $[x, x + \delta)$ f and g are nowhere equal since f is always within $c/2$ of $f(x)$ on that interval and g is always within $c/2$ of $g(x)$, and $f(x)$ and $g(x)$ differ by c, so no number is within $c/2$ of both of them.

Since any two right-continuous functions differ on a non-empty open subset and CDFs are right-continuous if two CDFs F and G differ on $[0, 1)$ the result is immediate. If the functions do not differ on $[0, 1)$ they do not differ anywhere since the extension of a CDF to $[0, 1]$ is unique. $\qquad\square$

Which we now apply.

Proof. Let F and G be the CDFs of two probability measures absolutely continuous with respect to the Lebesgue measure. A Radon-Nikodym derivative (density function) of the measure $\alpha F + (1 - \alpha)G$ is then $\alpha \frac{\mathrm{d}F}{\mathrm{d}\lambda} + (1 - \alpha)\frac{\mathrm{d}F}{\mathrm{d}\lambda}$. Using the strict convexity of f, we have the inequality

$$f\left(\alpha\frac{\mathrm{d}F}{\mathrm{d}\lambda} + (1 - \alpha)\frac{\mathrm{d}F}{\mathrm{d}\lambda}\right) < \alpha\psi(\frac{\mathrm{d}F}{\mathrm{d}\lambda}) + (1 - \alpha)f(\frac{\mathrm{d}F}{\mathrm{d}\lambda})$$

And the same inequality holds for the integrals

$$\int_0^1 f\left(\alpha\frac{\mathrm{d}F}{\mathrm{d}\lambda} + (1 - \alpha)\frac{\mathrm{d}F}{\mathrm{d}\lambda}\right)\,\mathrm{d}x < \int_0^1 \alpha f\left(\frac{\mathrm{d}F}{\mathrm{d}\lambda}\right) + (1 - \alpha)f\left(\frac{\mathrm{d}F}{\mathrm{d}\lambda}\right)\,\mathrm{d}x$$

since it holds pointwise and applying Lemma 2 we have that the CDFs differ on an open set and this implies their densities do, too, so the inequality is strict.

Finally, we note that any other Radon-Nikodym derivative differs from the one we constructed only on a Lebesgue-negligible set so the value of any such integral is equivalent and the choice of density function is immaterial to the inequality. □

This result will be used in Chapter 5. It is interesting because it lets us take an easy to understand strictly convex function from the reals, and create a strictly convex function of the absolutely continuous probability measures, a much more difficult class of objects to work with.

2.2 Banach Spaces and Duality

Strict properness is a property of a relation, the association between predictions and scoring functions or, as we saw in the last section, the association between probability measures and bounded measurable functions. Convex analysis will allow us to study the structure of these associations because it lets us understand relationships between the elements of a Banach space and its dual. This brief section describes what those are, and how they apply to our interests.

2.2.1 Banach Spaces

A Banach space is a complete metric space. That is, it is a set X coupled with a metric d where every Cauchy sequence converges to a limit in X. Elements of a Banach space are vectors, and like all vector spaces these vectors may be added together or multiplied by a scalar, and there always exists a zero vector.

Letting $(\Omega, \mathcal{F})$ be a measurable space, there are two Banach spaces we will be

interested in. The first is the ca space of (bounded, signed and) countably additive measures, since this space contains the probability measures $\mathcal{P}$ as a closed convex subset, and these represent beliefs and predictions. The metric we will use on the probability measures is the total variation distance, defined as

$$||p_0 - p_1|| \;\; = \;\; \sup_{F \in \mathcal{F}} |p_0(F) - p_1(F)| \qquad \text{(total variation distance)}$$

Intuitively, the total variation distance of two probability measures is the greatest difference in likelihood they assign any event.[6]

It is important to realize that the probability measures are not, themselves, naturally a Banach space: multiplying by any scalar other than one does not give us a probability measure, nor does adding two probability measures together; plus, there is no zero vector. Hence why we situate the probability measures in the ca space. While we will not explicitly reference the ca space after this section, it will continue to be important to think of the probability measures as a thin slice of a larger space, as this geometric thinking offers valuable intuition in the next section.

The second Banach space we will be interested in is the bounded measurable functions $\mathcal{B}$, which will become our scoring functions. These are part of the dual space of the probability measures (described below), and convex analysis will let us study pairings between them. A norm on $\mathcal{B}$ is the supremum norm

$$||b|| \;\; = \;\; \sup_{\omega \in \Omega} |b(\omega)| \qquad \text{(supremum norm)}$$

and we use this to define a metric that is simply the greatest difference two functions assign any point. Our need for Banach spaces is why we must restrict attention to

---

[6]This metric is derived from the total variation norm on the ca space: $||\mu|| = \mu_+(\Omega) + |\mu_-(\Omega)|$, where μ_+ is the positive part of the measure μ, and μ_- is the negative part.

the *bounded* measurable functions. Boundedness lets us define our norm (and thus our metric), and if the functions were allowed to have infinite values we could not add them together and would not have a vector space.

Gneiting and Raftery did not require their scoring functions to be bounded (they did require them to be measurable), and this distinction is complicated because it is, on the one hand, less general, and yet it lets us apply the powerful tools convex analysis has to study strict properness. I think the key to understanding this trade-off is that allowing unbounded scoring functions is, quite simply, uninteresting, and well worth trading off for the rich theoretical framework we gain. First, infinite scores are impractical, and scoring functions that actually attain infinite values cannot be used in prediction markets where the difference of two scoring functions must be taken. Second, in the discrete setting and on the $[0, 1]$ interval the only unbounded functions must actually attain infinite values, and the interest of functions that are real-valued and unbounded is then, at best, specific to domains not yet considered. Finally, in addition to being impractical it is theoretically limiting, a special case that requires ad hoc tools and regularity conditions. I am happy to leave unboundedness behind, at least for now, to leverage the standard tools of convex analysis.[7]

As mentioned, these sets $\mathcal{P}$ and $\mathcal{B}$, are of interest because they can be placed in duality and studied using convex analysis. This section concludes with a discussion of this duality.

---

[7]For those familiar with strictly proper scoring rules, the logarithmic scoring rule is commonly used as an example strictly proper scoring rule. This rule is unsuitable to use in a prediction market, for the reasons mentioned, even though it often appears in that context. Further, my framework still includes the logarithmic scoring rule, it just does not allow its domain to be *any* possible prediction. When its domain is restricted the logarithmic scoring rule can associate every prediction with a bounded scoring function, and this is the only version suitable for use in a prediction market.

2.2.2 Duality

Two compatible Banach spaces can be paired, or placed in duality, and relations between them studied using convex analysis. In particular, we can pair the ca space, which includes the probability measures $\mathcal{P}$, with the bounded measurable functions $\mathcal{B}$. We will be interested in this pairing because it associated beliefs and predictions with scoring functions, and these associations will be fundamental to our study of strict properness.

The continuous dual space of a Banach space X is the set of continuous linear functions from X to the reals. The continuous linear functions from X to the reals, denoted X^*, is also a Banach space, and its continuous dual space contains X. Two Banach spaces that are part of the others' continuous dual spaces are considered paired or *placed in duality*, and they have a natural bilinear form between them, a function from $X \times X^*$ to the reals that is linear in both arguments.

The ca space and the bounded measurable functions can be placed in duality, and the bilinear form between them is simply the Lebesgue integral. This is conventionally written:

$$\langle \mu, b \rangle \;=\; \mu(b) \;=\; \int_\Omega b \, \mathrm{d}\mu \qquad\qquad \text{(bilinear form)}$$

for a countably additive measure μ and bounded measurable function b.

As mentioned, convex analysis lets us study relations between spaces in duality. Since the probability measures are not the entire continuous dual space of the bounded measurable functions, and the bounded measurable functions are not the entire continuous dual space of the probability measures, we will exercise caution in the next section to be sure we are only dealing with these objects of interest. This

will become more apparent shortly.

2.3 Convex Functions and their Subdifferentials

In the previous two sections we represented the possible outcomes we are trying to predict as a measurable space $(\Omega, \mathcal{F})$. The probability measures $\mathcal{P}$ on this measurable space are an expert's beliefs and the possible predictions, and the bounded measurable functions $\mathcal{B}$ are the the possible scoring functions. We discussed how $\mathcal{P}$ and $\mathcal{B}$ were part of each others' dual spaces, and I said this meant we could study relations between them using convex analysis. In this section we will see what convex analysis offers us. This section, unlike the other two in this chapter, actually contains some specialized results of my own motivated by our focus on $\mathcal{P}$ and $\mathcal{B}$, and we will need these results in Chapters 3 and 4.

2.3.1 Functions and Relations

In this section we will be discussing many functions and relations, and we will need some general notation for them.

A relation between two sets X and Y is a non-empty set of ordered pairs consisting of an element from X and an element from Y. The domain of a relation is the elements of X in it, and its range is the elements of Y in it. A relation between X and Y is usually introduced as $R \subseteq X \times Y$, and I write R^T for the transpose of R, where $(y, x) \in R^T$ when $(x, y) \in R$. The notation $R|_C$ is the restriction of R to $C \subseteq X$, the set of pairs from R such that $(x, y) \in R$ and $x \in C$. Then notation $R(C)$ is the image of C under R, or all y such that $(x, y) \in R$ for some $x \in C$.

A function $f : X \to Y$ also defines a special type of many-to-one relation, and we equivalently write $f(x) = y$ and $(x, y) \in f$. Functions, unlike relations, can be described as lower semicontinuous (l.s.c.), an extremely useful property when studying convex analysis, and continuous. Whenever we discuss the continuity or lower semicontinuity of a function it will be a function between two normed spaces, and continuity will be with respect to the norm topologies on X and Y.

2.3.2 Convex Functions

A convex functions maps a Banach space X to the extended reals $\bar{\mathbb{R}}$ such that[8]

$$f : X \to \bar{\mathbb{R}} \qquad\qquad\qquad \text{(convex function)}$$

$$\alpha f(x_0) + (1 - \alpha)f(x_1) \; \geq \; f(\alpha x_0 + (1 - \alpha)x_1), \; \forall x_0, x_1 \in X, \alpha \in [0, 1]$$

if the inequality is strict for all $x_0, x_1 \in X$ and $\alpha \in (0, 1)$ then we say f is strictly convex. If the inequality is strict whenever tested on a subset $W \subseteq X$ then I will describe f as strictly convex on W. That is, a function is strictly convex on a set W if the inequality is strict whenever x_0, x_1 and $\alpha x_0 + (1 - \alpha)x_1$ are in W. This is my own generalization of strict convexity, and we will use it when characterizing the structure of strictly proper scoring rules.

Convex functions have some special notation. The effective domain of a convex function f is where it is real-valued and is denoted $\text{dom} f \subseteq X$. If $W \subseteq X$ and I write $f : W \to \bar{\mathbb{R}}$ then I mean the effective domain of f is a subset of W and it is $+\infty$ elsewhere. If a function is real-valued somewhere and nowhere negative infinity

---

[8]The mapping is also often described as from a convex subset of X. This distinction is uninteresting since the domain of such a function can be extended to all of X by defining it as $+\infty$ outside its original domain. This extension preserves convexity, properness and lower semicontinuity.

then it is called proper. If a function is both l.s.c. and proper I will call it closed. This language will be especially useful as it will avoid a profusion of "propers" in our discussion. This language is also appropriate since a proper convex function is l.s.c. if and only if its epigraph[9] is closed, which is the case when the effective domain of the function is a closed convex set. One incredibly useful fact is that a function is closed and convex if and only if it is the supremum of a family of continuous affine functions.[9, p. 80] I will actually use a family of continuous linear functions in Chapter 3, a special case of this result.

Two useful facts about l.s.c. convex functions that I will use later are that (1) l.s.c. convex functions are bounded below on bounded sets [13, p. 144] and (2) l.s.c. and real-valued convex functions of Banach spaces are, in fact, continuous [9, p. 74].

2.3.3 The Subdifferential

Convex functions admit a generalization of the classical derivative known as the subdifferential. Subgradients, elements of the subdifferential, are points from the dual space of the convex function's domain, and a convex function's association between points and subgradients describes a class of relations between two spaces placed in duality. In our case, a convex function $f : \mathcal{P} \to \mathbb{R}$ will have subgradients that are elements of $\mathcal{B}$, and a convex function $f : \mathcal{B} \to \mathbb{R}$ will have subgradients that are elements of $\mathcal{P}$. Of course, as mentioned previously $\mathcal{P}$ and $\mathcal{B}$ are not each others' entire dual space, and so the subdifferential of these functions may contain elements

---

[9]We will not need to know what the epigraph is. Very roughly, we can get some intuition into what the epigraph is by saying that for a convex function $f : \mathbb{R} \to \bar{\mathbb{R}}$ the epigraph is the set in $\mathbb{R}^2$ defined as the space above the function.

from outside our sets of interest. I will create a refinement of the subdifferential that lets us restrict attention to just these sets of interest. To reiterate our goal, we are interested in relations between predictions from $\mathcal{P}$ and scoring functions from $\mathcal{B}$, and these relations will be encoded or embodied or identified with an association between points and subgradients of a convex function.

Let X^* be the continuous dual space of X. The subdifferential of a (proper)[10] convex function f is the function ∂f defined as[11]

$$\partial f : \mathrm{dom} f \to 2^{X^*} \quad \text{(subdifferential)}$$

$$\partial f(x_0) \;=\; \{x_0^* | x_0^* \in X^*, \; f(x) - f(x_0) \;\geq\; \langle x - x_0, x_0^* \rangle, \; \forall x \in X\}$$

Following convention I let $\mathrm{dom} \partial f$ be the subset of X where the subdifferential of f is non-empty. An element of $\partial f(x)$ is referred to as a subgradient of f at x, and if $\mathrm{dom}\ \partial f = \mathrm{dom} f$ I will simply describe the function as subdifferentiable. A useful fact is that the subgradients of a convex function always form a closed convex set in the continuous dual space.

It is important to remember that a subgradient is a continuous linear function of the domain of a convex function. When studying convex functions in Euclidean space, $f : \mathbb{R}^n \to \mathbb{R}$, these functions can be identified with vectors from $\mathbb{R}^n$. This is because $n-$dimensional Euclidean space *is its own continuous dual space*. Every linear function on $\mathbb{R}^n$ can be represented as a vector from $\mathbb{R}^n$, and the bilinear form between these two spaces is the dot product. When working in a discrete setting, like our meteorologist predicting rain tomorrow, we have a finite number of outcomes

---

[10]For the subdifferential of a convex function to be nonempty it must be proper.

[11]Following convention that 2^X is the collection of all subsets of X.

and a probability measure is also a vector in $\mathbb{R}^n$. In that example it is actually an element of $\mathbb{R}^2$. We also saw that its scoring functions had two values, and so could be identified with elements of $\mathbb{R}^2$ as well. This is to be expected because a scoring function comes from the continuous dual space of the probability measures. Note that it is easy to become confused, and think of this dual space as always having the same structure, and this example reveals how this is not the case. The continuous dual space of the probability measures depends greatly on the measurable space we are considering. When we consider a convex function $f : \mathbb{R}^2 \to \mathbb{R}$, its subgradients will also be members of $\mathbb{R}^2$, and we will use the association between points on the function and its subgradients to associate predictions with scoring functions.

On the $[0, 1]$ interval a probability measure is a CDF, and so our convex function will map CDFs to the extended reals, $f : \mathcal{P} \to \bar{\mathbb{R}}$. On its effective domain it may be subdifferentiable, and where subdifferentiable it creates an association between CDFs and elements of their continuous dual space. Unlike $\mathbb{R}^2$, this may or may not be a bounded measurable function. Assuming it is, the convex function will describe an association between CDFs and bounded measurable functions of the interval $b : [0, 1] \to \mathbb{R}$. I will next introduce a refinement of the subdifferential that ensures we do not accidentally describe an association between probability measures and other mathematical objects.[12]

---

[12]There does not seem to be a good description of what, exactly, the continuous dual space of the probability measures on an arbitrary measurable space is. Although we do know that the continuous dual space of the bounded measurable functions is the ba space of all bounded and finitely additive signed measures, which includes the ca space as a closed subspace.

2.3.4 Refining the Subdifferential

As mentioned, we will need to refine the subdifferential so that we can restrict it only to the objects we are interested in, like the probability measures and bounded measurable functions, so we can focus on relations only between them.[13]

Letting $Y \subseteq X^*$, the $Y-$subdifferential of a convex function is

$$\partial_Y f : \operatorname{dom} f \to 2^Y \qquad\qquad (Y-\text{subdifferential})$$

$$\partial_Y f(x_0) \;=\; \partial f(x_0) \cap Y$$

In particular, the $\mathcal{B}-$subdifferential of a convex function will only include the bounded measurable functions, and the $\mathcal{P}-$subdifferential will only include probability measures.

Another, further refinement will be to focus on a convex function's association between points and their *unique subgradients.* This is because it will be useful later to be sure that only one probability measure p is associated with each bounded measurable function b, and this association can be identified with a convex function $f : \mathcal{P} \to \bar{\mathbb{R}}$ where b is a subgradient of f at p and only at p. In fact, it is this unique subdifferential relation that is necessary and sufficient for there to be a strictly proper relationship between the probability measures and bounded measurable functions, although elaborating on this will have to wait until Chapter 3.

---

[13]Previously I mentioned that Gneiting and Raftery did not require boundedness. Maybe future work will not even require measurability, and allow any element of the continuous dual space to somehow be used as a scoring function. How, exactly, this would work is beyond my understanding, as the continuous dual space of the ca space is an unknown menagerie with objects so exotic they are unlikely to be both functions from the ca space and from our state space Ω. Our ability to interpret the bounded measurable functions as both is essential, since as functions from the ca space they define an expectation, and as functions from the state space they define a score.

Formally, the unique $(Y-)$subdifferential of a convex function f is

$$\partial_Y f : \text{dom} f \times 2^{\text{dom} f} \to 2^Y \qquad \text{(unique subdifferential)}$$

$$\partial_Y f(x_0; W) = \{x_0^* \mid x_0^* \in \partial_Y f(x_0), x_0^* \notin \partial_Y f(x), \ \forall x \in W, x \neq x_0\}$$

This says that the unique $Y-$subdifferential at a point x_0 with respect to a set W is the set of $Y-$subgradients of x_0 that are not also subgradients at other points in W. So if $f : \mathcal{P} \to \bar{\mathbb{R}}$ and $b \in \partial_{\mathcal{B}} f(p; \mathcal{P})$ then the bounded measurable function b is in the $\mathcal{B}-$subdifferential of f at p, and nowhere else.

These refinements are my own, and maybe in the future they will be standardized better. They are needed for the particular analysis we will be doing, as will the following little lemma that connects unique subgradients with the subgradient inequality holding strictly. This lemma will be used in my characterization of scoring rules, and appears to be known (used in [43] without proof) but not formalized elsewhere.

Lemma 3 (Uniqueness and Strict Subgradient Inequality). *Let X be a Banach space and X^* its continuous dual space; let $f : X \to \bar{\mathbb{R}}$ be a (proper) convex function with $x_0^* \in \partial f(x_0)$. If $W \subseteq X$, then $x_0^* \in \partial f(x_0, W)$ if and only if $f(x) - f(x_0) > \langle x - x_0, x_0^* \rangle$ for all $x \in W, x \neq x_0$.*

Proof. The subgradient inequality implies

$$f(x) - f(x_0) \geq \langle x - x_0, x_0^* \rangle, \ \forall x \in X \qquad \text{(subgradient inequality)}$$

$$f(x) - \langle x, x_0^* \rangle \geq f(x_0) - \langle x_0, x_0^* \rangle$$

So if there exists $x' \in W$ such that

$$f(x') - f(x_0) = \langle x' - x_0, x_0^* \rangle \qquad \text{(Case 1)}$$

then

$$f(x) - \langle x, x_0^* \rangle \;\geq\; f(x') - \langle x', x_0^* \rangle, \;\; \forall x \in X$$

and so Case 1 implies x_0^* is also a subgradient of f at x', and so not in the unique subdifferential of f at x_0 with respect to W.

Alternatively

$$f(x) - f(x_0) \;>\; \langle x - x_0, x_0^* \rangle, \;\; \forall x \in W, x \neq x_0 \qquad \text{(Case 2)}$$

$$f(x_0) - f(x) \;<\; \langle x_0 - x, x_0^* \rangle$$

yet if $x_0^* \in \partial f(x)$ then

$$f(x_0) - f(x) \;\geq\; \langle x_0 - x, x_0^* \rangle \qquad \text{(subgradient inequality)}$$

a contradiction, and so this case implies $x_0^* \notin \partial f(x), \;\; \forall x \in W, x \neq x_0$. Thus, since we assumed $x_0^* \in \partial f(x_0)$, it is in the unique subdifferential of f at x_0 with respect to W. $\qquad\square$

Before moving on, the subdifferential, being a function, is naturally a relation between a set X and 2^{X^*}. It is incredibly convenient and conventional to pretend it is instead a relation between X and X^* itself, with $(x, x^*) \in \partial f$ when $x^* \in \partial f(x)$. I will use the same convention for similar functions through this chapter and Chapters 3–5.

2.3.5   Gâteaux differential

There are many notions of differentiability suitable for working in Banach spaces, one closely related with the notion of subdifferentiability is the Gâteaux differential.

Understanding this particular differential and how it relates to strict properness is useful because it is a familiar mathematical property, and in Chapter 5 it will offer us a natural notion of prices for securities as well as a means of associating many probability measures with a bounded measurable function. The details of these last two advantages must, of course, be left for Chapter 5 since they require a great deal of new context to be understood.

Let X be a Banach space, and $f : X \to \bar{\mathbb{R}}$ a function. Assume the limit

$$\lim_{\tau \to 0} \frac{f(x + \tau h) - f(x)}{\tau}$$

exists for all $h \in X$ at a point $x \in X$, the Gâteaux variation of f at x is the function

$$\nabla f(x; \cdot) : X \to \bar{\mathbb{R}} \qquad \text{(Gâteaux variation)}$$

$$\nabla f(x; h) \;=\; \lim_{\tau \to 0} \frac{f(x + \tau h) - f(x)}{\tau}$$

And f is Gâteaux differentiable at x if the variation is a continuous linear function of h, in which case we refer to it as the Gâteaux differential. That is, f is Gâteaux differentiable at x if the Gâteaux variation exists and is an element of the continuous dual space of X. For a function $f : \mathbb{R} \to \mathbb{R}$ this means the differential is simply a real number and is, in fact, the derivative, and for a function $f : \mathbb{R}^n \to \mathbb{R}$ the Gâteaux differential is the gradient.

The subdifferential and Gâteaux differential are sometimes related. If a convex function has a single subgradient at a point where it is finite and continuous then the function is Gâteaux differentiable there and its subgradient is the differential. Conversely, if a convex function is Gâteaux differentiable at a point it has a single subgradient at that point equal to the differential, and if a convex function is l.s.c.

and Gâteaux differentiable at a point it is continuous there, too [9, p. 87][13, p. 159].

2.3.6   Cyclic Monotonicity and the Subdifferential

So far we have defined the subdifferential and a few refinements, and mentioned that the relationship between points and subgradients of a convex function can let us study relations between spaces in duality. This subsection describes how convex functions lets us study cyclically monotonic relations, and importantly how any such relation is always part of the subdifferential of some closed convex function. This last fact will let us focus exclusively on this class of convex functions without loss, letting us use the great additional structure we get with closed functions to study our relation of interest, that between the probability measures and bounded measurable functions.

I just mentioned how we interpret ∂f as a relation between a Banach space X and its continuous dual X^*, and it turns out these relations are exactly the cyclically monotone ones between these spaces [82]. A relation $R \subseteq X \times X^*$ is cyclically monotone when

$$\sum_{i \in I} \langle x_i, x_i^* \rangle \; \geq \; \sum_{i \in I} \langle x_{\sigma(i)}, x_i^* \rangle \qquad \text{(cyclic monotonicity)}$$

for every finite set of points I, $(x_i, x_i^*) \in R$ and where σ is any permutation of I. A relation is a subset of the subdifferential relation of a convex function if and only if the relation is cyclically monotone.[14], and is the subdifferential of a closed convex function

---

[14]As a concrete example, the Rockafellar function of a relation R always encodes the relation. It

if and only if it is maximal cyclically monotone.[15] Every cyclically monotone relation can be extended to a maximal cyclically monotone one, and a maximal cyclically monotone relation interpreted as a subdifferential ∂f uniquely defines f up to an additive constant [82].[16] Importantly, this means that any subdifferential relation is part of the subdifferential relation of some closed convex function, and this allows us to restrict attention to this class, which sometimes offers valuable structure.

2.3.7   Conjugate Functions

A useful tool when studying the subdifferential relation of a convex function, especially closed convex functions, is a function's conjugate. Intuitively, the conjugate of a closed convex function is also a closed convex function where the subdifferential relationship is flipped. Conjugates will be used in Chapter 4 where I describe cost functions, and Chapter 5 as a means of identifying the subdifferential of a particular convex function.

is defined as

$$f_R : X \to \bar{\mathbb{R}} \qquad \text{(Rockafellar function)}$$
$$f(x) = \sup\{\langle x - x_n, x_n^* \rangle + \cdots + \langle x_1 - x_0, x_0^* \rangle\}$$

where the supremum is taken over all finite sets pairs $(x_i, x_i^*) \in R$. If R is cyclically monotone then f_R is a closed convex function, and if also $(x, x^*) \in R$ then $x^* \in \partial f_R(x)$.

[15]This is true when X is a Banach space, as I have assumed. A relation is maximal cyclically monotone if no additional pairs can be added to it without violating the cyclic monotonicity condition. Equivalently, a relation is maximal cyclically monotone if it is not a subset of another cyclically monotone relation. See [12] for a good survey of and introduction to monotonic functions.

[16]Recall that we are treating the subdifferential as a subset of $X \times X^*$.

Formally, the conjugate of a convex function $f : X \to \bar{\mathbb{R}}$ is defined as

$$f^* : X^* \to \bar{\mathbb{R}} \qquad\qquad \text{(conjugate function)}$$

$$f^*(x^*) \;=\; \sup_{x \in X} \; \langle x, x^* \rangle - f(x)$$

Conjugates have many interesting properties. The conjugate of a proper convex function is always a closed convex function. The biconjugate of f is the conjugate of its conjugate and is written f^{**}; if f is a closed convex function then $f(x) = f^{**}(x)$ for all $x \in X$.[17] In the future I will write $f(x) \overset{X}{=} f^{**}(x)$ when two functions agree on a set X.

As an example, a closed convex function $f : \mathcal{P} \to \bar{\mathbb{R}}$ has a conjugate function f^* that can be restricted to $\mathcal{B}$, $f^*|_{\mathcal{B}} : \mathcal{B} \to \bar{\mathbb{R}}$. We will see that if f describes the expected score function of a scoring rule, then $f^*|_{\mathcal{B}}$ describes a cost function. Alternatively, if $f : \mathcal{B} \to \bar{\mathbb{R}}$ is a closed convex function describing a cost function, then its restricted conjugate $f^*|_{\mathcal{P}} : \mathcal{P} \to \bar{\mathbb{R}}$ will describe the expected score function of a scoring rule. These facts are elaborated on in Chapter 4.

What makes the conjugate so useful for the study of the subdifferential of a closed convex function is the conjugate-subgradient theorem, adapted here from [9].

Theorem 1 (Conjugate-Subgradient Theorem)**.** *Let X be a Banach space and X^**

---

[17]I am intentionally careful not to say the two functions are identical. The biconjugate may be well-defined outside the domain of the original function since it maps not from the original space X but X^{**}, the bidual of X. The bidual of the ca space is not itself, for example. When a space is its own bidual it is called reflexive and admits many special properties. Finite dimensional spaces like $\mathbb{R}^n$ are always reflexive and convex analysis in these spaces admits a great deal of additional structure. In fact, Euclidean n-space is also its own continuous (and algebraic) dual space. This is why we often see subgradients defined as a dot product between two n-dimensional vectors: the dot product is the bilinear form between n-dimensional Euclidean space and its continuous dual (itself). Structural differences like this can make convex analysis in general Banach spaces difficult to understand since we are likely to bring our Euclidean assumptions along with us to places they no longer belong.

its continuous dual space; also let $f : X \to \bar{\mathbb{R}}$ be a proper and convex function. Then the following three properties are equivalent:

1. $x^* \in \partial f(x)$

2. $f(x) + f^*(x^*) = \langle x, x^* \rangle$

If the function is also l.s.c. (closed) then the above properties are also equivalent to

3. $x \in \partial f^*(x^*)$

The first and third equivalences state that the conjugate of a closed convex function "flips" its subdifferential relation. The second equivalence describes where the conjugate function expression attains its supremum. That is, if $x^* \in \partial f(x)$, we have that

$$f^*(x^*) = \langle x, x^* \rangle - f(x)$$
$$x \in \arg\max_{x \in X} \langle x, x^* \rangle - f(x) \qquad \text{(conjugate attainment)}$$

This equivalence will be useful for understanding a convex function's supporting subgradients, and how they relate to the subdifferential. These supporting subgradients are essential to our understanding of strict properness, and are described in the next subsection.

2.3.8   Supporting Subgradients

A critical relationship for strict properness is that between points on a convex function f and its *supporting subgradients*. These are subgradients that agree with

the function when they are in the subdifferential or, equivalently, subgradients for which f^* is zero. It is precisely these supporting subgradients that we will identify with scoring functions, and relations between points and *unique* supporting subgradients will describe every strictly proper association between predictions and scoring functions.

This subsection, unlike the others in this section, specifically addresses convex functions of the probability measures $f : \mathcal{P} \to \mathbb{R}$. These functions have supporting subgradients wherever they are subdifferentiable, and this is not generally the case. Other functions always have *supporting hyperplanes*, but these are *affine* and not *linear* functions. Working with them is difficult since they do not fit in our standard duality framework. Luckily, when we focus exclusively on the probability measures we can stick with more familiar linear functions for our analysis.

The subdifferential of (proper) convex functions $f : \mathcal{P} \to \mathbb{R}$ actually has several interesting properties we will use. First, it may contain elements from the continuous dual space of the ca space, and this may mean functions outside of $\mathcal{B}$. Thus we will restrict our attention to the $\mathcal{B}-$subdifferential. A notable special case where this is not required is when our measurable space $(\Omega, \mathcal{F})$ has a finite set Ω. In this case a convex function of the probability measures is a function $f : \mathbb{R}^{|\Omega|} \to \mathbb{R}$,[18] and the continuous dual space is also represented by $\mathbb{R}^{|\Omega|}$. Our scoring functions are vectors describing a score for each outcome, there are $|\Omega|$ outcomes, and so the scoring functions are identified with vectors in $\mathbb{R}^{|\Omega|}$. The bilinear form between the probability measures

---

[18]More precisely the function maps from the probability simplex of Ω, not any vector in the Euclidean plane.

and bounded measurable functions in this case is the dot product.[19]

A second fact about the subdifferential of a convex function $f : \mathcal{P} \to \mathbb{R}$, is that if it contains $b_0 \in \mathcal{B}$ at a point $p_0 \in \mathcal{P}$, then it also contains $b_0 + k$ for all $k \in \mathbb{R}$ at that point.[20] We can check this using the subgradient inequality:

$$f(p) - f(p_0) \geq \langle p - p_0, b_0 + k \rangle \qquad \text{(subgradient inequality)}$$

$$f(p) - f(p_0) \geq \langle p - p_0, b_0 \rangle + \langle p - p_0, k \rangle$$

$$f(p) - f(p_0) \geq \langle p - p_0, b_0 \rangle + k - k$$

$$f(p) - f(p_0) \geq \langle p - p_0, b_0 \rangle$$

the last line being true because we assumed $b_0 \in \partial f(p_0)$. Intuitively, the expected value of a constant function k is simply k, and using the separability of the bilinear form shows that if one function is in the subdifferential of such a function f, then every translation of that function is, too. Further, if b_0 is in the unique subdifferential, then all its translations are, too.

The above also tells us if f is (uniquely) $\mathcal{B}-$subdifferentiable at a point p_0, then it has a (unique) supporting $\mathcal{B}-$subgradient at that point, too. This requires formally defining a supporting subgradient. Letting $P \subseteq \mathcal{P}$ and $B \subseteq \mathcal{B}$, we can define this mapping as

$$\partial_B^\star f : \mathcal{P} \times 2^{\mathcal{P}} \to 2^B \qquad \text{(unique supporting } B-\text{subgradients)}$$

$$\partial_B^\star f(p; P) = \{x^\star \mid x^\star \in \partial_B f(p; P), x^\star(p) = f(p)\}$$

---

[19] The dot product equals the Lebesgue integral in this case.

[20] That is, it contains the function defined by adding b_0 with the constant function k, which can be represented by a real value.

and as a shorthand I will let $\partial_B^\star f(p) = \partial_B^\star f(p; p)$ be the not necessarily unique supporting $B-$subgradients of f at p. So letting $b_0 \in \partial f(p_0)$ as before, then the function $b_0 - f^*(b_0)$ is a supporting subgradient of f at p_0, and this construction shows that (unique) subdifferentiability implies the existence of a (unique) supporting subgradient.

The next chapter demonstrates these unique supporting $\mathcal{B}-$subgradients describe every scoring rule, and Chapter 4 shows they describe all of strict properness.

3

Scoring Rules

The old-established way of measuring a person's belief is to propose a bet, and see what are the lowest odds which he will accept. This method I regard as fundamentally sound; but it suffers from being insufficiently general, and from being necessarily inexact.

. . .

Suppost next that the subject is capable of doubt; then we could test his degree of belief in different propositions by making him offers of the following kind. Would you rather have world α in any event; or world β if p is true, and world γ if p is false?

. . .

This is, of course, a very schematic version of the situation in real life, but it is, I think, easier to consider it in this form.

– Frank P. Ramsey, 1926[1]

. . . one essential criterion for satisfactory verification is that the verification scheme should influence the forecaster in no undesirable way. Unfortunately, the criterion is difficult, if not impossible to satisfy, although some schemes will be much worse than others in this respect.

– Glenn W. Brier, 1950

---

[1]All Ramsey's quotations are taken from his essay *Truth and Probability* [81].

Strictly proper scoring rules are a popular means of acquiring and rewarding accurate predictions of the future. Our study of these rules is motivated by a desire to predict the likelihood of future events, events we assume can be represented by a measurable space $(\Omega, \mathcal{F})$. We ask an expert for a prediction from $\mathcal{P}$, and use a scoring rule to associate it with a scoring function from $\mathcal{B}$. If the rule is strictly proper then this association is designed to incentivize self-interested experts to offer the most accurate prediction possible, or, equivalently, this association is designed to reward, in expectation, more accurate predictions more than less accurate ones.

Strict properness is the *essential* property for eliciting and valuing predictions. If we ask experts for predictions and attempt to score them in a non-strictly proper fashion then we arrive back at the problem Brier was trying to solve in 1950: our "verification scheme" influences our forecaster in undesirable ways. It may let the expert be lazy and offer a less accurate prediction without penalty, or it may actively encourage the expert be inaccurate. Neither of these cases is desirable, and from a narrative perspective we will remain exclusively interested in strict properness. Traditionally, however, a weaker property known simply as "properness" is also characterized alongside strict properness, and this is mathematically so easy to do that I will follow convention and my formal statements will describe proper and strictly proper scoring rules.

In this chapter I characterize strictly proper scoring rules. I begin with a formal definition in Section 3.1, then discuss how these rules can also be used to run a prediction market in Section 3.2. Section 3.3 offers the formal characterization, and Section 3.4 concludes by comparing my characterization to that offered in [43], which

also characterized strictly proper scoring rules for measurable spaces. My result stresses the relationship between predictions and scoring functions, and how it is this relationship that a class of convex functions describes. Situating my discussion within the realm of convex analysis offers many advantages, and one minor and arguable disadvantage

3.1   Scoring Rules, Formally

Scoring rules have been defined and characterized many times, notably for discrete spaces in [83] and for measurable spaces in [43]. The typical definition of a scoring rule differs from the one I will offer, although the two are essentially the same barring notational differences. My definition will fit much better in our narrative, however. The classical definition appears at the end of this chapter when discussing Gneiting and Raftery's characterization of strictly proper scoring rules. My own definition is:

Definition 1 ((Strictly Proper) Scoring Rule)**.** *A scoring rule is any function $S :$ $\mathcal{P} \to \mathcal{B}$.[2] If S is a scoring rule with domain P and image or codomain B, it is called $(P/B-)$proper when*

$$p(Sp) \ \geq \ p(Sp') \qquad\qquad \text{(properness)}$$

for all $p, p' \in P$. The rule is strictly $(P/B-)$proper when the inequality holds strictly unless $p' = p$.

We can relate this definition back to our motivating story. When using a scoring

---

[2]A scoring rule is an operator, and it will often be convenient to follow convention and write Sp instead of $S(p)$.

rule we begin with a desire to predict the likelihood of future events, represented by $(\Omega, \mathcal{F})$. An expert has some beliefs $p \in \mathcal{P}$ about how likely these future events, and we ask this expert for a prediction $p' \in \mathcal{P}$. Our scoring rule maps this prediction to a scoring function, $b = Sp'$, and after we wait and observe the actual outcome $\omega \in \Omega$ the expert is scored $b(\omega)$. Strict properness says the expert's expected score, $p(Sp')$, is uniquely maximized when the expert reveals its beliefs. This is a strict incentive for self-interested (and risk neutral) experts to accurately report what they think likely to occur. Further, a strictly proper scoring rule rewards, in expectation, an expert reporting what it believes to be the most accurate prediction of the future.

A strictly proper scoring rule S defines a one-to-one relation between $\mathcal{P}$ and $\mathcal{B}$. Each prediction (probability measure) can be associated with only one scoring function (bounded measurable function), and each scoring function can only be associated with one prediction. This is best understood as thinking of a scoring rule as offering its expert a menu of scoring functions B. The expert picks a function from this menu, and we then infer from its choice what its beliefs are. If one scoring function was associated with two predictions then we would be unable to infer the expert's belief. A scoring rule also must associate only one scoring function with each prediction because it is a function of the predictions.

Thinking of scoring rules as offering a menu will be an especially useful intuition in the next chapter on cost functions. For now, the key point is to realize that the expert's prediction p' is simply the language it uses to select a scoring function. A scoring rule is strictly proper, then, when the menu of scoring functions it offers separates the possible predictions, with each prediction expecting its own scoring

function to score highest. Note that I did not say "with each scoring function being maximized by a different prediction." This is a distinct property and it does not relate to strict properness. Strict properness is about each prediction or belief "preferring" a different function from the menu offered.

3.2   Prediction Markets and Scoring Rules

A prediction market is a popular mechanism for acquiring a consensus prediction from a group of experts, not just one expert as we have been discussing. Prediction markets using scoring rules were first proposed in [48], and have been heavily studied since. Predictions markets often use a scoring rule or a cost function, and in this section I will describe them using scoring rules.

A prediction market that uses a scoring rule, or more concisely a scoring rule market, lets one or more "traders" interact with a scoring rule a countable number of times. It begins, like always, with a desire to predict the likelihood of future events, represented by $(\Omega, \mathcal{F})$. The market starts or "opens" with an initial prediction p_0, then accepts traders' public and countable predictions as $p_1, p_2, \ldots$. Eventually the market closes with a final prediction p, either because traders stop offering predictions or because the market stops accepting them, and afterwards we observe the actual outcome $\omega \in \Omega$ and score each trader the sum of the scores of their predictions minus the sum of the scores of the immediately preceding predictions, as determined by our scoring rule.

How traders are scored in a market and how we interpret these scores requires elaboration. Let our scoring rule market be using a scoring rule $S : \mathcal{P} \to \mathcal{B}$, then the

trader who made prediction p_i receives a score of

$$(Sp_i - Sp_{i-1})(\omega) \qquad\qquad \text{(market's score for a prediction)}$$

for that prediction. Prediction markets reward traders for improving over the last prediction. If the scoring rule S is strictly proper then I will also describe the market as strictly proper. A strictly proper market has the important property that if the last prediction made disagrees with a trader's beliefs, then that trader has an incentive to adjust the market. In particular, if the market closes with a prediction p that differs from a trader's belief, then that trader had an opportunity to score higher than it did. Thus, if trading stops organically in a market we are assured of a consensus among the traders.

Our alternative motivation for strict properness was rewarding accurate predictions, and a strictly proper market can be thought of as rewarding the improvement a more accurate prediction makes upon a less accurate one. It also penalizes predictions that are less accurate than the ones immediately preceding them. These markets are especially clever because they reward the marginal information of a prediction. In practice, they are also often considered cost effective because the total payout to traders is only $(Sp - Sp_0)(\omega)$.

I said that if trading stops organically in a strictly proper market then it implies a consensus among traders. It may be, however, that trading would never stop of traders' own accord, or that the consensus reached is a poor one. How markets do or do not reach consensus and aggregate information is discussed more in Chapter 6. Still, with these classical prediction markets we think of accepting the last prediction made as the best we will receive. Remember, though, our discussion of predictions

markets here is just motivational and contextual. We need nor make no assumptions about them or their use until Chapter 5 and beyond, where we will discuss particular types of markets.

3.3   Characterizing Strictly Proper Scoring Rules

A scoring rule is a relation between $\mathcal{P}$ and $\mathcal{B}$, two paired spaces. Chapter 2 discussed how some relations between such pairs could be studied using convex analysis, and the following characterization of strict properness shows that strictly proper scoring rules are represented by a subset of the structure of a particular class of convex function.

Theorem 2 (Strictly Proper Scoring Rules and Convex Functions with Unique Supporting Subgradients)**.** *Let $P \subseteq \mathcal{P}$ and $B \subseteq \mathcal{B}$. A scoring rule $S : P \to B$ is strictly $P-$proper if and only if there exists a convex function $f_S : \mathcal{P} \to \bar{\mathbb{R}}$ such that $S \subseteq \partial^\star_\mathcal{B} f_S(P; P)$. The rule is $P-$proper if and only if $S \subseteq \partial^\star_\mathcal{B} f_S(P)$. (In plain English, a scoring rule is strictly proper if and only if it is a subset of the unique supporting subgradient relation of a convex function of the probability measures.)*

Proof. Let $S : P \to B$ be a $P/B-$proper scoring rule (the assumption that S is onto is without loss of generality), and let f_S be the closed convex function defined as the pointwise supremum of B, $f_S(p) = \sup_{b \in B} p(b)$. Letting $(p_0, b_0) \in S$, properness says

$$\langle p_0, b_0 \rangle \ \geq \ \langle p_0, b \rangle, \ \forall b \in B \qquad\qquad \text{(properness)}$$

and this implies $f_S(p_0) = p_0(b_0)$. We now verify b_0 is a supporting subgradient of f_S

at p_0 by checking the subgradient inequality:

$$f_S(p) - f_S(p_0) \geq \langle p - p_0, b_0 \rangle \qquad \text{(subgradient inequality)}$$

$$\sup_{b \in B} p(b) - \langle p_0, b_0 \rangle \geq \langle p, b_0 \rangle - \langle p_0, b_0 \rangle$$

$$\sup_{b \in B} p(b) \geq \langle p, b_0 \rangle$$

which is always true, and so this convex function is such that $S \subseteq \partial_{\mathcal{B}}^{\star} f_S(P)$, as desired.

When S is strictly $P/B-$proper we have

$$\langle p_0, b_0 \rangle > \langle p_0, b \rangle, \ \forall b \in B \qquad \text{(strict properness)}$$

still implying $f_S(p_0) = p_0(b_0)$. Following the above b_0 is a supporting subgradient of f_S at p_0, and we verify it is unique with respect to P by checking the strict subgradient inequality on P (see Lemma 3):

$$f_S(p) - f_S(p_0) > \langle p - p_0, b_0 \rangle, p \in P \qquad \text{(strict subgradient inequality)}$$

$$\sup_{b \in B} p(b) - \langle p_0, b_0 \rangle > \langle p, b_0 \rangle - \langle p_0, b_0 \rangle$$

$$p(b) > \langle p, b_0 \rangle, (p, b) \in P$$

which follows from strict properness. These two arguments show that every proper scoring rule implies the existence of a (closed) convex function with a corresponding supporting $B-$subgradient relation, and that every strictly proper scoring rule implies the existence of a (closed) convex function with a corresponding unique supporting subgradient relation.

Now let f_S be a (proper) convex function $f_S : \mathcal{P} \to \mathbb{R}$, and let $S \subseteq \partial_{\mathcal{B}}^{\star} f_S(P)$ be a non-empty one-to-one relation, and let the range of this relation be B. The

subgradient inequality and the structure of this relation tell us

$$f_S(p) - f_S(p_0) \geq \langle p - p_0, b_0 \rangle, \ \forall (p_0, b_0) \in S$$

which we substitute into using $f_S(p_0) = p_0(b_0)$, to obtain

$$f_S(p) \geq \langle p, b_0 \rangle$$

and if $(p, b) \in S$ this gives

$$\langle p, b \rangle \geq \langle p, b_0 \rangle, \forall (p, b) \in S, b_0 \in B$$

which is properness. If $S \subseteq \partial_{\mathcal{P}}^{\star} f_S(P; P)$ then the subgradient and following inequalities are strict whenever $b \neq b_0$, and we have strict properness. This concludes our proof by showing that selecting any of the supporting subgradients determines a proper scoring rule, and selecting any subset of the unique supporting subgradients determines a strictly proper scoring rule.

□

In plain English, the above says that we can identify a strictly proper scoring rule with a one-to-one subset of the unique supporting subgradient relation of a convex function. It is important to notice that we do not identify a strictly proper scoring rule with the unique supporting subgradient relation in its entirety, because it is possible that the latter relation may associate one prediction with multiple scoring functions. That is, there may be multiple unique supporting subgradients at a point. Since a scoring rule is a function it can only map each point to one of these subgradients. Cost functions, discussed in the next chapter, avoid this limitation.

The proof also discusses the convex function f_S defined as the pointwise supremum of the range of a scoring rule S. This function is often called the expected score function of S, since if an expert has beliefs $p \in \mathcal{P}$ then their expected score for making an accurate prediction is $f_S(p)$. The expected score function of a scoring rule is uniquely defined, and there is a many-to-one association between scoring rules and expected score functions. That is, multiple scoring rules, even multiple strictly proper scoring rules, may share the same expected score function. Chapter 4 will elaborate on expected score functions further.

Now that we understand there is a structural relationship between strictly proper scoring rules and the unique supporting subgradient relation of a convex function, a natural follow-up is better understanding when and where a convex function has unique supporting subgradients. As discussed in Chapter 2, this is equivalent to a convex function $f : \mathcal{P} \to \bar{\mathbb{R}}$ having unique subgradients.

Theorem 3 (Unique Subdifferential Relations). *Let X be a Banach space and X^* its continuous dual space. Let $Y \subseteq X^*$ and $f : X \to \bar{\mathbb{R}}$ be a closed convex function; the following are equivalent:*

1. *Y is a subset of the range of the unique subdifferential relation of f*

2. *f is strictly convex on dom $\partial_Y f$*

3. *the subgradient inequality, $f(x) - f(x_0) \leq \langle x - x_0, x_0^* \rangle$, holds strictly whenever $x \neq x_0$ for all $x \in X$ and $x_0^* \in \partial_Y f(x_0)$*

4. *the $X-$subdifferentials of f^* on Y are singleton sets*

Proof. I first prove (1) implies (2). Assume, for a contradiction, that (2) is false and (1) is true. Then there exists $x, x_0, x_1 \in \text{dom } \partial_Y f$ and an $\alpha \in (0, 1)$ such that $\alpha x_0 + (1 - \alpha)x_1 = x$ and

$$\alpha f(x_0) + (1 - \alpha)f(x_1) \;=\; f(x) \qquad\qquad \text{(convex equality)}$$

Let $x^* \in Y$ be a subgradient of f at x. The conjugate-subgradient theorem says

$$f^*(x^*) \;=\; \langle x, x^* \rangle - f(x)$$

$$f^*(x^*) \;=\; \langle \alpha x_0 + (1 - \alpha)x_1, x^* \rangle - \alpha f(x_0) - (1 - \alpha)f(x_1)$$

$$\langle x, x^* \rangle - f(x) \;=\; \alpha\big(\langle x_0, x^* \rangle - f(x_0)\big) + (1 - \alpha)\big(\langle x_1, x^* \rangle - f(x_1)\big)$$

and since the conjugate-subgradient theorem says $x \in \arg\max_{\bar{x}} \langle \bar{x}, x^* \rangle - f(\bar{x})$ and the above is a convex combination we have that both terms must be equal, and so we conclude $x_0, x_1 \in \arg\max_{\bar{x}} \langle \bar{x}, x^* \rangle - f(\bar{x})$, too, and thus x^* is in the $Y-$subdifferential of x_0 and x_1. This contradicts our assumption of (1) that Y is a subset of the range of the unique subdifferential relation of f, so (1) implies (2).

Now I show (2) implies (1). Assume, for a contradiction, that (1) is false and (2) is true. Let $x, x_0, x_1 \in \text{dom } \partial_Y f$ such that $x^* \in \partial_Y f(x_0) \cap \partial_Y f(x_1)$ and there exists $\alpha \in (0, 1)$ such that $x = \alpha x_0 + (1 - \alpha)x_1$. Since (2) is true

$$\alpha f(x_0) + (1 - \alpha)f(x_1) \;>\; f(x)$$

The two subgradient inequalities

$$f(x_0) - f(x) \;\leq\; \langle x_0 - x, x^* \rangle$$

$$f(x_1) - f(x) \;\leq\; \langle x_1 - x, x^* \rangle$$

together with our assumptions imply

$$\alpha f(x_0) + (1 - \alpha)f(x_1) \ \leq \ f(x)$$

a contradiction of (2). Since we assume (2) true, however, this shows (1) and (2) are equivalent.

The last two equivalences following immediately from Lemma 3 and the conjugate-subgradient theorem. □

In particular, the above shows that a strictly convex $f : \mathcal{P} \to \bar{\mathbb{R}}$ has unique supporting subgradients at every point, although these subgradients may not be bounded measurable functions. Historically, characterizations of strict properness have focused entirely on strict convexity, and this was possible because they also assume the probability measures of interest, $P \subseteq \mathcal{P}$, were a convex set. In this case the above theorem shows strict convexity is equivalent every subgradient being unique. The next section of this chapter will discuss one of these characterizations.

This theorem on unique subdifferentials only applies to closed convex functions, and we need to recall that every scoring rule is a subset of the unique supporting subgradient relation of a closed convex function, an example of which is its expected score function. Because the subdifferential of every convex function is a subset of the subdifferential of a closed convex function, we can focus on the closed convex functions without loss of generality.

Lastly, before moving on, observe that although the prior theorem tells us the convex function is uniquely subdifferential on a portion of its domain, it does not tell us how to find the actual relation matching points from X and subgradients from X^*. This is a real challenge in need of some active work, and I will not return to

the problem until Chapter 5, when I need to better describe the subdifferential of a particular convex function. Doing so requires some special tricks that may be hard to generalize.

3.4   Gneiting and Raftery's characterization

In [43] Gneiting and Raftery presented a characterization of scoring rules for arbitrary measurable spaces. They use a slightly different definition of a scoring rule, and some unique concepts that are outside standard convex analysis. I will go through this definition and these adjuncts, then present and compare their characterization with my own.

Gneiting and Raftery start with an arbitrary measurable space $(\Omega, \mathcal{F})$ and define a scoring rule as any function $S : \mathcal{P} \times \Omega \to \bar{\mathbb{R}}$, such that the partial functions $S(p, \cdot) : \Omega \to \bar{\mathbb{R}}$ are $\mathcal{P}-$quasiintegrable. This is the usual way to define such scoring rules, with an equivalent definition appearing in [83] for discrete spaces. It is simply easier for my narrative to have a scoring rule map predictions to scoring functions.

From our perspective, we can interpret this as saying that Gneiting and Raftery allow a broader class of scoring functions than the bounded measurable functions, since they allow any $\mathcal{P}-$quasi-integrable function of Ω. Quasi-integrability is like regular integrability, which applies to measurable functions that have real-valued Lebesgue integrals, except it allows the integral to have infinite values.

Since these scoring functions are not part of the dual space of the probability measures (ca space) they require two specialized concepts to handle. The authors define a "subtangent" of a convex function $f : \mathcal{P} \to \bar{\mathbb{R}}$ at a point p as function

61

$t : \Omega \to \bar{\mathbb{R}}$ that is (1) $p-$integrable, (2) $\mathcal{P}-$quasi-integrable and (3) satisfies the subgradient inequality. They also describe a scoring rule as "regular" if the expected score of any scoring function is real-valued unless it differs from the experts beliefs, in which case the expected score can be negative infinity.[3]

We can now state their characterization, which I present a little differently than they do for clarity.

Theorem 4 (Gneiting and Raftery). *Let $P \subseteq \mathcal{P}$ be a convex set. A regular scoring rule $S : P \times \Omega \to \bar{\mathbb{R}}$ is $P-$proper if and only if there exists a convex function $f : P \to \mathbb{R}$ such that*

$$S(p, \omega) \;=\; f(p) - \int_\Omega t \; \mathrm{d}p + t(\omega)$$

where t is a subtangent of f at p. The rule is strictly $P-$proper when f is strictly convex.

The above statement is obtuse. If we pretend that t is a bounded measurable function and subgradient, then we can apply the conjugate-subgradient theorem, though, which says $f(x) + f^*(x^*) \;=\; \langle x, x^* \rangle$, and lets us interpret the above expression as defining $S(p, \cdot) : \Omega \to \mathbb{R}$ as $t - f^*(t)$, which is a bounded measurable function and a supporting subgradient of f. So in this case the statement becomes similar to my own, and reveals some of its intuition.

While Gneiting and Raftery's characterization was a significant accomplishment in our development of scoring rules, my own characterization has the following benefits:

---

[3]This can happen when a score of negative infinity is associated with events a prediction assigns zero likelihood to.

1. It shows that strictly proper scoring rules are a one-to-one subset of a convex function's unique supporting subgradients relation.

2. It does not use subtangents or require regularity conditions.

3. It allows P to be any subset of the probability measures, not just a convex one.

4. It will suggest a ready extension of strict properness to cost functions (see the next chapter).

5. It ensures that any strictly proper scoring rule can be used in a prediction market.

Of course, this last point may also be characterized a flaw in my characterization. Gneiting and Raftery do allow a more general class of scoring function than I do. This generalization is, however, unlikely to be interesting. Practically it is hard to think of how we would assign or enforce a score of negative infinity, much less when we might want to do so. Plus, we are usually interested in strictly proper scoring rules to use in prediction markets, and scoring functions that assign scores of negative infinity cannot be used in that setting since we must take the difference of scores, and the difference of negative infinity and negative infinity is undefined.[4]

---

[4]The logarithmic scoring rule is a popular scoring rule in discrete settings, and using it in a market requires restricting the domain of predictions it will accept. If allowed to accept any prediction it will produce scores of negative infinity. Any restriction that requires its scores always be real-valued is also a scoring rule in my framework, since each scoring function is then a bounded measurable function. When there are a finite number of outcomes any regular scoring rule, in the sense of Gneiting and Raftery, can have its domain restricted to produce a scoring rule that always assigns real-valued scores. Regular scoring rules for arbitrary measurable spaces, however, may be such that any prediction can produce a score of negative infinity. Restricting the domain of these scoring rules will not let them be used in a market.

That is not to say that my own characterization is perfect. It may still be possible to generalize beyond $\mathcal{P}$, possibly including some finitely additive measures, or go beyond $\mathcal{B}$ and include other members of the dual space of the ca space. Or there may be an alternative mathematical object that more succinctly and powerfully expresses the structure of strictly proper scoring rules. Most likely, there might be a fascinating notion of $\epsilon-$strict properness that can be developed by studying the $\epsilon-$subdifferential of a convex function. Noting that there is still a great deal to do, I think one major advantage of my characterization is that it places our thinking about scoring rules firmly in the realm of convex analysis, where we have many tools available to do it with. We will use these tools again in the next chapter to define and understand strictly proper cost functions.

4

Cost Functions

The previous chapter discussed strictly proper scoring rules and reiterated that strict properness is the *essential* property for acquiring and rewarding accurate predictions of the future. Strictly proper scoring rules have been the focus of a great deal of work, including [43] which characterized them for arbitrary measurable spaces. Our understanding of cost functions, another popular method for eliciting predictions, has lagged behind our knowledge of scoring rules. Until now, cost functions for measurable spaces were not characterized, and there was no notion of a strictly proper cost function.

A cost function, like a scoring rule, is a means of acquiring and rewarding predictions about the likelihood of future events. These functions are especially interesting because they can allow us to create prediction markets that emulate existing futures markets where traders can buy and sell securities. Trading securities may have two significant advantages over using a scoring rule: (1) the interface is likely to be considered more familiar in many settings and (2) it lets traders focus on their area of

expertise, instead of requiring their prediction specify the likelihood of every event.

As an example, we can consider running two prediction markets for the NCAA tournament. One will use a scoring rule, and it requires traders submit predictions specifying the likelihood of every possible event. There are more than a quintillion of these events, and so traders using this scoring rule must submit more than a quintillion numbers to specify their predictions.[1] This market is unlikely to see much participation.

In our second market we will act as a market maker and offer securities for each event that pay \$1 when that event occurs. We will quote prices for these securities, and if a trader's expected value for a security differs from our quoted price then it will have an incentive to buy or sell it. If, say, our security for Gonzaga beating Louisville has a price of 30 cents, and a trader thinks the likelihood of this event is 80%, then that trader will purchase this security and increase its price. This market is far easier to interact with than the first one, and it allows traders to focus on their areas of expertise (like the chances Gonzaga beats Louisville).

Cost functions enable us to run this second kind of market with its simple and natural interface. In these markets prices represent an implicit prediction, and the price of a single security is one aspect or part of this representation. Traders no longer need to deal with the entire prediction, only these parts. Offering this simplicity to traders shifts some complexity from them to the market, however. When a trader

---

[1]It is possible that we could design a system that interprets succinct representations into full predictions, and I will return to this idea in the conclusion. Currently, a significant challenge of working in discrete spaces is their complete lack of natural structure. In the next chapter we will heavily exploit the structure of probability measures on the [0, 1] interval in our development of a cost function for it.

purchases shares in Gonzaga beating Louisville they are telling us this event is more likely than we thought, and we will want to adjust the prices of other related events – like the chances Gonzaga wins the tournament – accordingly. Intuitively, we, as market maker, are responsible for translating a trader's myopic adjustment into a set of feasible prices that reflect some reasonable prediction about the future. Instead of traders dealing with quintillions of numbers, now the market has to.

This chapter focuses exclusively on characterizing strictly proper cost functions, and Chapter 5 will discuss how they can emulate futures markets. It begins in Section 4.1 with the development of scoring relations, the root objects in our study of strict properness. Section 4.2 defines strictly proper cost functions, and Section 4.3 concludes with a brief discussion of their relation to scoring rules, scoring relations, and expected score functions, that better reveals the structure of strict properness..

4.1   Scoring Relations

A scoring relation is a generalization of a scoring rule that allows many scoring functions to be associated with each prediction. Scoring relations are truly the fundamental object in our study of strict properness, and happily it is straightforward to generalize from scoring rules to them.

Definition 2 ((Strictly Proper) Scoring Relation)**.** *Any non-empty one-to-many relation $R \subseteq \mathcal{P} \times \mathcal{B}$ is a scoring relation. Letting P be the domain and B the range of R, such a relation is $(P/B-)$proper when*

$$p(b) \ \geq \ p(b') \qquad\qquad \text{(strict properness)}$$

for all $(p, b) \in S$ and $b' \in B$. The relation is strictly $(P/B-)$proper when the inequality is strict whenever $(p, b') \notin R$.

A scoring relation is a lot like a scoring rule (scoring rules are simply one-to-one scoring relations) and its definition of strict properness is analogous. A scoring relation describes an association between predictions and scoring functions, just like a scoring rule, and we can also think of it as offering a menu of scoring functions. If a scoring relation is strictly proper, then the clutch of scoring functions it associates with a prediction has two properties: (1) an expert who thinks that prediction is most likely prefers choosing one of these scoring functions over all others and (2) the expert is indifferent among the scoring functions in this set. Strict properness also requires that (3) no two predictions are associated with the same (set) of scoring functions. Thus, if we offer the range of a strictly proper scoring relation as a menu of scoring functions we can still uniquely infer an expert's beliefs from their choice of scoring function: only that prediction thinks that scoring function maximizes its expected score.

As we saw in the previous chapter, strictly proper scoring rules represent subsets of the relation between points and unique supporting subgradients (that are also bounded measurable functions) of a convex function. Strictly proper scoring relations have the same characterization.

Theorem 5 (Strictly Proper Scoring Relations and Convex Functions with Unique Supporting Subgradients). *A scoring relation R is strictly $P/B-$proper if and only if there exists a convex function $f_R : \mathcal{P} \to \mathbb{R}$ such that $R = \partial_B^\star f_R(P; P)$. The relation is $P/B-$proper if and only if $R = \partial_B^\star f_R(P)$.*

The proof is immediate from the characterization of strictly proper scoring rules.

Scoring relations also admit expected score functions, which we will now need to formalize.

Definition 3 ((Strictly Proper) Expected Score Functions)**.** *A closed convex function* $f_R : \mathcal{P} \to \bar{\mathbb{R}}$ *is an expected score function for a scoring relation* R *when* $(p, b) \in R$ *implies* $b \in \partial^\star f_R(p)$. *If* f_R *is an expected score function then it is called (strictly) proper exactly when* R *is (strictly) proper.*

Given a strictly $P/B-$proper scoring relation R, the pointwise supremum of B is an expected score function for R. This simple construction lets us consider only closed functions as expected score functions without loss of generality (and see Chapter 2 for a more general perspective on why restricting attention to this class is without loss). The name "expected score function" is derived from the property that $f(p) = p(b)$ for all $(p, b) \in R$, and we can interpret this as saying that the function's value at a prediction $p \in P$ is the highest expected score an expert holding that belief can obtain (when choosing a function from the scoring menu B).

There may be multiple expected score functions for a scoring relation, and we knew this since scoring rules are scoring relations and they may have multiple expected score functions. The requirement for a function f_R to be an expected score function for our scoring relation R is that its points and unique supporting subgradients contain the association between predictions and scoring functions described by the relation. Importantly, any expected score function of a scoring relation has the same value on the domain of the scoring relation. Thus, if the domain of a scoring relation is all of $\mathcal{P}$, it will uniquely define an expected score function.

Strictly proper scoring relations are a way of capturing the complete structure offered by the points and unique supporting subgradients of a convex function. Scoring rules may not be able to accomplish this since they are one-to-one mappings, and a convex function may have multiple unique supporting subgradients at a point. Cost functions, described in the next section, can be thought of as extending strictly proper scoring relations to capture the structure of a convex function's points and their unique subgradients.

4.2   Strictly Proper Cost Functions

Our goal with a cost function is to let traders buy and sell securities like they do in existing futures markets, and to use this trading behavior to infer a prediction. This goal will not be fully realized until the next chapter, and it is far from our current understanding. It can also be difficult to see how the mathematical definitions in this and the previous section relate to this more practical goal. I ask the reader to bear with me as we first continue to abstract even farther away from futures markets before returning to them.

One way to think of a cost function is as a language exposing a scoring menu. Throughout this thesis I have described scoring rules, and now scoring relations, as offering a menu of bounded measurable functions. Strictly proper scoring relations actually describe all of (classical) strict properness in that they can describe every scoring menu and how to infer a probability measure from them. Not all scoring relations, however, immediately offer us a means of *exposing* this menu. Scoring rules are useful because they provide a natural language, the language of predictions, as a

means to access this menu.

One natural alternative idea to using a scoring rule is to let traders select a scoring function directly from the menu. This is almost what a cost function does, and is an excellent intuition for how they work. Of course, we could also design other languages, possibly even ones that are many-to-one mappings into the scoring functions. For example, we could have two scoring functions b_0 and b_1, and design a silly new means of acquiring and rewarding accurate predictions that required traders give us some object from an arbitrary set $\mathcal{L}$ that then mapped to one of these two scoring functions. This is, of course, absurd, but it illustrates the point that our systems for inferring predictions can be thought of as (1) a menu of scoring functions and (2) a language that lets traders select a scoring function. Again, using scoring rules as an example, their language for selecting a scoring function was the set of predictions.

I will describe the language of cost function as *securities*. Mathematically these securities are translations of scoring functions that the cost function exposes. For example, when a cost function exposes a scoring menu B, a security will be any real-valued translation of an element of B. So if $b \in B$, then $b + k$ for any real k is a security. Intuitively the language of cost functions lets traders not only request a scoring function directly (they can ask for b itself), but also allows them to ask for them in the apparently (for now) mathematically circuitous fashion of specifying a translation of the scoring function instead. This roundabout language is what will let us emulate a futures market, and this context is described in the next chapter. For now, we can think of a trader "purchasing" the security $b + k$ at a "cost" of k to arrive

back at the scoring function b. This thinking gives us the name "cost function:" the function determines how much a trader must pay for a security.

A strictly proper cost function, then, must map securities to scoring functions. To define these functions we need a formal definition of B^+, the securities generated from the scoring menu B. We formally define these translations as:

$$B^+ \;=\; \{b^+|b^+ = b + k, b \in B, k \in \mathbb{R}\} \qquad \text{(translations of elements of } B)$$

and it will be convenient to describe a cost function as strictly P/B^+-proper even though it exposes the scoring menu B, leading to the following definition of a strictly proper cost function.[2]

Definition 4 (Strictly Proper Cost Function)**.** *Any function $C : \mathcal{B} \to \mathcal{B}$ is a cost function.*

Letting R be a (strictly) $P/B-$proper scoring relation, a cost function $C : B^+ \to B$ may also be described as (strictly) P/B^+-proper when $C(b + k) = b$, for all $b \in B$, $k \in \mathbb{R}$. I will also call such cost functions "cost functions for R."

Sometimes it will be useful to restrict the domain of C to a subset of $B° \subseteq B^+$, and in these cases I will describe C as (strictly) $P/B°-$proper.

One subtle and very important thing that is easy to miss in this definition is that simply exposing the scoring menu of a strictly $P/B-$proper scoring relation R makes the cost function strictly P/B^+-proper in the most meaningful way: if a trader has beliefs $p \in P$ then they will strictly prefer choosing securities $b^+ \in B^+$ such that

---

[2]I am abusing the idea that k is a real number and a constant-valued function represented by a real number, and will continue to do this for the immense convenience it offers. There should be no confusion for the reader.

$(p, C(b^+)) \in R$. The strict properness part of this definition only demands the menu exposed be the same as some such scoring relation. It is the "cost function" part of the definition that defines how the language of securities is mapped to the scoring functions.

This definition follows closely on our discussion. A cost function exposes a scoring menu B, and it does so by mapping translations of scoring functions, called "securities," back to the original functions. We can also derive this definition geometrically from expected score functions. Letting f_R be a strictly $P/B-$proper expected score function, and $b \in B, b \in \partial^* f_R(p)$, the function $b + k$ is also a subgradient of f_R at p. It is not a supporting subgradient unless k is zero, and we can think of a cost function as offering these subgradients as its language and associating them with their corresponding supporting hyperplanes.

Yet another way of thinking about a cost function involves the conjugate of f_R, f_R^*. Applying the conjugate-subgradient theorem, we see that f_R^* maps translations of supporting subgradients into the real values k that they are translated by. In other words, if $b \in B, b \in \partial^* f_R(p)$, then $f_R^*(b+k) = k$ for all $k \in \mathbb{R}$. The classical definition of a cost function fits this idea closely.

Definition 5 (Strictly Proper Classical Cost Function)**.** *Any function $\dot{C} : \mathcal{B} \to \mathbb{R}$ is a classical cost function.*

Letting R be a (strictly) $P/B-$proper scoring relation, a classical cost function $\dot{C} : B^+ \to \mathbb{R}$ may also be described as (strictly) P/B^+-proper when $\dot{C}(b+k) = k$, for all $b \in B$, $k \in \mathbb{R}$. I will also call such classical cost functions "classical cost functions for R."

Letting f_R be an expected score function for R, this can be equivalently stated as $\dot{C}(b+k) \overset{B^+}{=} f_R^*(b+k)$. *That is,* $\dot{C}$ *agrees with the conjugate of* f_R *on* B^+.

Essentially, I like to think of cost functions mapping translations of scoring functions (securities) back to scoring functions, and the classical way of thinking of a cost function actually had them map to a "cost" for each security. Translating between these two definitions is trivial, and I will switch between the representations as is convenient. The next section elaborates on some properties that can make the classical cost function representation easier to work.

4.3 Cost Functions in Duality

Scoring relations are the fundamental object of study when understanding strict properness, and particular methods of acquiring and rewarding accurate predictions simply offer different languages to access their menus of scoring functions. Scoring rules use the language of predictions, and cost functions offer translations of the scoring functions called "securities." Both of these languages are intuitive and natural, the first because we are, after all, attempting to acquire a prediction, and the second since we can think of a market maker offering securities to trade, just like in a futures market (a context we will soon return to).

Having now described scoring rules, scoring relations, expected score functions and cost functions, it can all become too confusing how these objects relate to one another, or what separates one from another, or how one might actually work with these objects. Before formalizing some of these connections, let me offer a (relatively) simple example.

The function $f(x, y) = x^2 + y^2$ is a strictly convex and continuous two-dimensional function. This function is about the simplest we can hope to work with, since each dimension represents an outcome. This function has a well-defined Gâteaux differential everywhere, its gradient, and this implies it has a unique subgradient at every point (see Chapter 2). For instance, at the point (.5, .5) it has the subgradient (1, 1).[3] Since f is strictly convex, each of these subgradients are unique (see the discussion in the previous chapter). Note that (1, 1) is not a supporting subgradient of f at (.5, .5), and that there is no supporting subgradient at that point.

The line segment defined by $x + y = 1, x > 0, y > 0$ represents the probability measures over two discrete outcomes, and so this set will be $\mathcal{P}$. Restricting f to this set we obtain $f|_{\mathcal{P}}$, which is no longer continuous or Gâteaux differentiable since it has an empty interior. Further, it now has an infinite number of subgradients at every point. At (.5, .5) it still has the unique subgradient (1, 1), and it also has unique subgradients $(1, 1) + k$ for any real-valued k. Because this function is restricted to the probability measures it now has unique supporting subgradients everywhere, and the unique supporting subgradient at $(.5, .5)$ is $(.5, .5)$. The function remains lower semi-continuous, proper and strictly convex.

We can use $f|_{\mathcal{P}}$ as an expected score function, and identifying points with the supporting subgradients that are translations of the original function's f gradients defines a strictly $\mathcal{P}-$proper scoring relation.[4] This scoring relation is one-to-one by construction, and so is actually a scoring rule. If an expert offers a prediction of

---

[3] The continuous dual space and, in fact, just the dual space of $\mathbb{R}^2$ is itself. The bilinear form is the dot product.

[4] We know it is proper for all of $\mathcal{P}$ by our construction, and the Gâteaux differentiability of f. We do not immediately know what scoring menu B this scoring rule exposes.

$(.5, .5)$ this scoring rule returns the scoring function $(.5, .5)$, and if the first outcome occurs the expert is scored $.5$ and if the second outcome occurs the expert is scored $.5$. If the expert offers the prediction $(.75, .25)$, we compute their scoring function by finding the gradient of f at $(.75, .25)$, which is $(1.5, .5)$, then translating that to be a supporting subgradient of $f|_{\mathcal{P}}$, to receive $(.875, -.125)$. So an expert predicting $(.75, .25)$ receives a score of $.875$ if the first outcome occurs, and $-.125$ if the second outcome occurs. This is to be expected: predicting one outcome is more likely than another will result in a higher score when that outcome occurs, and a lower score when the other does.

In general it can be very hard to understand the relation between a convex function's points and supporting subgradients. For finite outcomes we are lucky to have the above technique, where we can use a "nice" strictly convex function, take its gradient, and then determine a supporting subgradient of the restricted function from it.

Now we can think of using $f|_{\mathcal{P}}$ to define a cost function, which accepts securities, or subgradients of $f|_{\mathcal{P}}$, and returns scoring functions (supporting subgradients). So a trader might ask for the security $(1.5, .5)$ and receive the scoring function $(.875, -.125)$. This emulates the trader purchasing the security $(1.5, .5)$ at a cost of $.625$. We can compute this necessary translation using the conjugate of $f|_{\mathcal{P}}$, which maps subgradients into the translations needed to make them supporting subgradients. So $f|_{\mathcal{P}}^{*}(1.5, .5) = .625$.

There are still some gaps in our analysis. First, it is not clear what the domain of the cost function should be, since we do not know what scoring menu we are exposing.

Second, in general, computing the conjugate may often be difficult. In this case we can address both these gaps by using some results about convex programs:

$$f|_{\mathcal{P}}^*(x, y) = \sup_{p \in \mathcal{P}} \; \langle p, (x, y) \rangle - f|_{\mathcal{P}}(p)$$

$$= \sup_{p \in \mathbb{R}^2} \; p \cdot (x, y) - f|_{\mathcal{P}}(p)$$

which is a convex program in two variables. Further, since $f|_{\mathcal{P}}$ is strictly convex and $\mathcal{P}$ is closed this convex program will always have a unique solution, which implies a cost function derived from $f|_{\mathcal{P}}$ is actually strictly $\mathcal{P}/\mathcal{B}-$proper. This technique does not tell us what scoring menu $B \subseteq \mathcal{B}$ is actually exposed (that would require identifying all the supporting subgradients), but it does let us compute everything needed to run a scoring rule or cost function derived from $f|_{\mathcal{P}}$. In the next chapter I will also rely on our understanding of convex programs to describe a more complicated cost function.

Formally, the above example can be thought of as an instantiation of this next theorem:

Theorem 6 (The Structure of Strict Properness). *Letting R be a scoring relation with domain P and range B, the following statements are equivalent:*

1. *R is a strictly $P/B-$proper scoring relation.*

2. *There exists a l.s.c. convex function $f_R : P \to \mathbb{R}$ that is a strictly $P/B-$proper scoring function for R. The convex conjugate of f_R restricted to $\mathcal{B}$ is a classical cost function for R.*

3. *R is a subset of the unique supporting subgradient relation of a convex function $f_R : \mathcal{P} \to \bar{\mathbb{R}}$.*

4. *There exists a continuous convex function $\dot{C}_R : \mathcal{B} \to \mathbb{R}$ that is a strictly P/B^+- proper classical cost function for R and whose convex conjugate restricted to $\mathcal{P}$ is an expected score function for R.*

The statements follows readily from my characterization of strictly proper scoring rules and the conjugate-subgradient theorem, as well as l.s.c. convex functions on Banach spaces being continuous.

Some prior work has investigated the connections between scoring rules and cost functions, most notably [1], and also [3]. Both investigated the discrete case, and their statements are not simple. Essentially, they showed that there could exist an equivalence between strictly proper scoring rules and some cost functions, in the sense that both could offer the same scoring functions to traders. They also showed that "prices" in a cost function market can correspond exactly to predictions in a scoring rule market, although this thinking will have to wait until we develop prices in the next chapter. My treatment is more complete since it describes all of strict properness through the structure of strictly proper scoring relations, which are analogous to the structures between points and unique supporting subgradients of convex functions of the probability measures. Everything else is derived from this fundamental structure. This interest in association scoring rules and cost functions leads me to formalize the following simple duality between them:

Theorem 7 (Scoring Rules and Cost Functions)**.** *For any strictly $P/B-$proper scoring rule, there exists a strictly P/B^+-proper cost function with the same range (scoring functions).*

This fact is incredibly simple when presented from the perspective developed in

this thesis, and I think it highlights the utility of our approach. It says that for every scoring rule there is a cost function with the same set of scoring functions, and thus the only difference is the language used to select a scoring function. Remember, though, that since not every scoring relation is a scoring rule, this duality is imperfect. For every scoring rule there is a cost function, but there is not a scoring rule for every cost function.

There are many advantages to focusing on strict properness directly. It is an *essential* property for eliciting and rewarding accurate predictions, and so we would like to understand what is a strictly proper scoring rule, what is a strictly proper cost function, and what are strictly proper scoring relations. The greatest advantage of this approach is understanding strictly proper mechanisms as presenting languages that expose a scoring menu. The above theorem is a testament to the clarity of understanding this offers us, in contrast to some convoluted historical results.[5] The next chapter further demonstrates the utility of this approach. I have also chosen to define strict properness in such a way that we can identify strictly proper scoring relations with the points and unique supporting subgradients of convex functions of the probability measures. This lets us readily leverage the tools of convex analysis, like we saw in the above example. Prior work, by not conceptualizing scoring relations, had a more difficult time applying these tools.

Now that we understand what a strictly proper cost function is, we will use them in the next chapter to create prediction markets that emulate futures markets. These cost functions have more structure than we are assuming here, and we will find this

---

[5]I have written several of these convoluted historical results myself, and remain proud of them.

structure to be incredibly useful and natural.

5

Practical Cost Functions

This chapter is the culmination of our work in Chapter 2–4. So far we have been focused exclusively on strict properness, the *essential* property for eliciting and rewarding accurate predictions of the likelihood of future events. This chapter expands this focus, showing how cost functions can closely emulate futures markets in Section 5.1, and then demonstrating the utility of all this work by creating a cost function for bounded continuous random variables in Section 5.2.

5.1 Cost Functions as Futures Markets

In the previous chapter I described cost functions and strictly proper cost functions. Strictly proper cost functions mapped securities into scoring functions, and they did so in a way that mimicked buying and selling securities. Letting C be a strictly $P/B^{+}-$proper cost function, the security $b + k$ being mapped to b can be thought of as the cost function selling the security $b + k$ at a price of k. A classical

cost function offered another interpretation of this, simply mapping $b + k$ to k and describing its price directly.

This interpretation is especially nice because it provides traders with the familiar interface of trading securities with a market. Not all strictly proper cost functions, however, have all the nice properties we associate with a market maker. In particular, a strictly proper cost function (1) may not reliably offer every security for trade, and (2) it may not be able to quote meaningful prices. Consider, for example, a trivial strictly proper cost function with a single scoring function b. Then that cost function can sell the security once, and likely not twice, unless $b + k = 2b$. That is, the cost function is not defined on $2b$, and this means it cannot always trade this security. This is likely to be considered a real deficit in practice, and we can readily construct non-trivial examples that demonstrate this lack.

The second challenge I mentioned is that strictly proper cost functions may not always quote meaningful prices. By "meaningful," I mean that the market should be able to quote a price for a security such that if a trader's expected value for that security differs from the price, then that trader expects to profit by trading the security. Either buying *some amount* of the security if its price is lower than the trader's expectation, or selling *some amount* if the price is higher. Note that this *price* is likely not the *cost* of a security, and that we are not considering securities purchased in discrete units. Not every strictly proper cost function can quote a price like this.

It turns out these two problems are related, and Section 5.1.2 describes how requiring B^+ be a vector space solves both. When a strictly proper cost function has a

vector space for its domain it guarantees every security is reliably offered for trade, and it can quote a meaningful price for each. These results are almost immediate from the assumption. Two other desirable properties for a market maker to have are that it cannot lose more than a bounded amount of money, and that it offers traders no arbitrage. These problems are discussed in Section 5.1.3. Before either of those, however, it is time we described cost function prediction markets. This will provide the context needed for us to continue.

5.1.1 Cost Function Prediction Markets

So far we have discussed a great deal about cost functions, and in Chapter 3 I described a scoring rules market, but we have not yet looked at prediction markets that use cost functions, or cost function markets. This section will help us understand exactly how these markets work, and what specialized notation we need when working with them.

Let $B \subseteq \mathcal{B}$ and $C : B^+ \to B$ be a cost function. I have described cost functions as offering *securities*, which are elements of B^+ and mapping them to scoring functions in B. This is an excellent intuition when working with a single trader, as it mimics a market maker who charges the trader a price of k to purchase the security $b + k$, for some scoring function $b \in B$ and real k.

When a cost function is used in a market, however, it is best to think of it as mapping *liabilities* to scoring functions. Mathematically nothing changes. A liability is any element of B^+, and a scoring function is still an element of B. The reason for this change is that the term "liability" will better capture our intuition about how

cost functions emulate existing futures markets. The market can still be thought of as offering securities, and in Section 5.1.2 we will see how, under a very reasonable condition, every element of B^+ is also a security. In general, however, this may not be the case.

This is almost certainly confusing, so let's begin by adding some context. A cost function $C : B^+ \to B$ prediction market is thought of as opening with an initial liability $\ell \in B^+$. Usually $\ell = 0$, the constant zero function. Like a scoring rule prediction market, this market then accepts a countable and public series of trades or updates, and these updates are functions from B^+ that we will call *liabilities*. So, traders offer a countable and public series of liabilities $\ell_0, \ell_1, \ldots$.

Offering liabilities does not connect well with our idea of trading in a futures market. Fortunately, we can interpret these offers as traders buying and selling *securities*. When a trader requests a liability $\ell_i \in B^+$, they receive the scoring function

$$C(\ell_i) - C(\ell_{i-1}) \hspace{3cm} \text{(cost function market score)}$$

and we can think of this as the trader purchasing the security $\ell_i - \ell_{i-1}$. So securities are now the differences between two liabilities. When we think of traders purchasing securities, the latest liability offered is the *sum* of these securities, and thus represents the market maker's potential net payouts to the traders. Hence why I refer to them as "liabilities."

Eventually the market closes with a final liability $\ell \in B^+$, and if our cost function is a cost function for a strictly $P/B-$proper scoring relation R, then we extract the prediction p such that $(p, C(B^+) \in R$. Again, this is like a prediction market that uses a scoring rule, where the last prediction made is extracted from the market. This

implies the market maker's worst-case loss is

$$\sup_{\ell \in B^+, \omega \in \Omega} \big(C(\ell) - C(\ell_0)\big)(\omega) \qquad \text{(worst-case loss)}$$

This description of a cost function is entirely mathematically correct, and, I think, still very unsatisfying. It is odd that traders are offering liabilities, and it is especially strange that securities are only defined implicitly as the difference of these liabilities. This can make for some very strange behavior. In particular, a "security" may or may not be available depending on the market's current liability. Again, this is because we consider a security as a difference between two liabilities. So if the market has two scoring functions, b_0 and b_1, if the market has liabilities b_0 then the only meaningful security that can be purchased is $b_1 - b_0$, and if the market has liabilities b_1 the only meaningful security that can be purchased is $b_0 - b_1$. I think of these markets as being "unreliable," in that they do not consistently offer traders a set of securities to buy and sell. This failure also does not let our cost function offer meaningful prices for its securities. The next (sub)section addresses these oddities with a simple structural assumption.

5.1.2 Prices and the Reliable Market Maker

It is desirable, in practice, to let traders reliably buy and sell securities when interacting with a market maker. This means that, regardless of the market's current state, a trader can buy or sell any of a set of securities.

We might attempt to create such a market by starting with a set of *basis securities* $X \subseteq \mathcal{B}$. The market could then let traders purchase bundles of these securities, bundles that are elements of the vector space B^+ that is generated using X as a basis.

This market admits an interesting interpretation. Its liabilities and its securities are elements of B^+, and since B^+ is a vector space any security can be purchased or sold regardless of the market's current liabilities.[1] In other words, this market reliably offers a set of securities.

With our interest in strictly proper cost functions the above naturally suggests attempting to use our scoring functions as a set of basis securities. This may not do what is intended, however, since the vector space created from this basis may contain elements that are not translations of the scoring functions. Consider a two outcome state space. Then we might have two scoring functions (1, 0) and (0, 1). If we use these scoring functions as a basis they will generate a vector space that contains (.5, .5), which is not a translation of either scoring function. Our cost function has no means of interpreting a request for this vector.

Instead, we can require that B^+ simply be a vector space. This ensures that every element of B^+ can always be purchased as a security, since the sum of two elements in B^+ is also in B^+. Thus if the market has liability ℓ, we can take any element $b^+ \in B^+$ and add it to ℓ to achieve $\ell + b^+ \in B^+$. So an expert can always move the market from ℓ to $\ell + b^+$, effectively purchasing the security b^+. Again, in this case the space of liabilities and securities is the same, and the different words are only for intuitive clarity. The market's liability is the sum of all securities purchased, and we can think of traders buying and selling securities, with the set B^+ defining these securities.

It is this thinking that lets us return to why we defined cost functions as mapping

---

[1]Liabilities and securities are both elements of B^+ because the difference of any two liabilities is a security, and any such difference is also an element of the vector space B^+ by construction, so the space of liabilities and securities coincides. Any security can always be purchased because B^+ is a vector space, and the sum of any two elements must then also be in the space.

securities back to scoring functions. At first there seemed to be no reason to let translations of scoring functions be mapped into scoring functions. It would have been much simpler if traders requested scoring functions directly. The scoring functions alone, however, cannot describe a vector space, and without a vector space we cannot reliably offer securities. Considering the same market as before, if a trader purchases the security (1, 0), and another the security (0, 1), the market's liabilities are (1, 1). The next trader may then also want to purchase the security (1, 0), moving the market's liabilities to (2, 1). These liabilities arise naturally through trading securities, and they are not a scoring function. Thus we let cost functions map translations of scoring functions back to scoring functions. We also can only let a cost function map translations back to scoring functions because this mimics paying a cost for each security. If the security mapped back to a scoring function were not a simple translation then we could not create a cost for it that would perfectly emulate the scoring function.

When a cost function is strictly P/B^+-proper, and B^+ is a vector space, then we also obtain a natural notion of prices. This is easiest to see, as are many results on cost functions, using the classical interpretation of a cost function. Let R be a strictly $P/B-$proper scoring relation, and f_R an expected score function for it. The conjugate of f_R restricted to $\mathcal{B}$ is a classical cost function $\dot{C}_R$ for R, and it is a continuous convex function of the bounded measurable functions.

Now let $b^+ \in B^+$. Since $\dot{C}_R$ is strictly proper, its Gâteaux variation at b^+ agrees with the probability measure R associates with $C(b^+)$ on B^+. Let p be this probability measure, this means the Gâteaux variation of $\dot{C}_R$ at b^+, $\nabla \dot{C}_R(b^+; \cdot) \overset{B^+}{=} p$. Let's walk

through this equivalence. The variation of $\dot{C}_R$ at $\ell \in B^+$ in the direction of $b^+ \in B^+$ is

$$\nabla \dot{C}_R(\ell; b^+) \;=\; \lim_{\tau \to 0} \frac{\dot{C}_R(\ell + \tau b^+) - \dot{C}_R(\ell)}{\tau} \qquad \text{(price / Gâteaux variation)}$$

then if an expert has beliefs p such that $p(b^+) \neq \nabla \dot{C}_R(\ell; b^+)$, the above expression implies there exists $\tau \neq 0$ such that

$$\tau p(b^+) \;>\; \dot{C}_R(\ell + \tau b^+) - \dot{C}_R(\ell)$$

demonstrating that if a trader's expected value for a security differs from the Gâteaux variation in the direction of that security, the trader expects to profit by trading it. Thus, if the Gâteaux variation at $\ell \in B$ does not agree with the probability measure R associates with $C(\ell)$, a trader with beliefs p has a trade it expects to be profitable, and this implies the market is not strictly proper as I assumed.

Assuming B^+ is a vector space is vital for this result because it means the security τb^+ can actually be purchased. Without a vector space we can quote these prices for the securities just fine, but they are not meaningful since we cannot guarantee traders can *act* on them.

More formally, using the conjugate-subgradient theorem we can restate this argument as follows:

Theorem 8 (Strict Properness and the Gâteaux Variation)**.** *Let f be a strictly $P/B-$proper expected score function, with $B \subseteq \mathcal{B}$ a convex set. Then the conjugate of f restricted to $\mathcal{B}$, $f^*|_{\mathcal{B}}$ is a classical cost function $\dot{C}$ such that for all $(p_0, b_0) \in R$,*
$$\nabla \dot{C}(b_0; \cdot) \;\overset{B^+}{=}\; p_0.$$

Recall the points of a classical cost function are bounded measurable functions (securities) and its subgradients are probability measures (predictions). If a probability measure is a subgradient of the classical cost function at a point, then it is the probability measure our implicit scoring relation associated with that point (a bounded measurable functions). This is an easy consequence of the conjugate-subgradient theorem, and is a mathematical way of stating that the Gâteaux variation agrees with the probability measure we associate with each security. Finally, note the theorem only requires B be convex, and if B^+ is a vector space then this implies B is convex.

Note also that the Gâteaux variation satisfies our notion of *meaningful* prices. This implies that the prediction we associate with a liability is also how we determine prices. The price of a security is its expected value with respect to this prediction.

Letting the domain of a cost function, B^+ be a vector space adds a great deal of structure to a cost function. It lets us reliably offer securities and quote meaningful prices. This seem like highly desirable properties in practice, and the cost function we will develop shortly will offer a vector space of securities for this reason.

5.1.3   Bounded Loss and Arbitrage

The last two properties I will consider before describing a particular cost function for bounded continuous random variables are bounded loss and a lack of arbitrage. Let's begin with bounded loss:

Definition 6 (Bounded Loss)**.** *A $P/B-$proper scoring relation R has bounded loss if B is bounded above. A cost function for R has bounded loss if and only if R does.*

Note that only B must be bounded, and not the difference describing the worst-

89

case loss. This is because any particular scoring function is, itself, bounded, and so our only concern is that the set of all scoring functions is unbounded above. We now have the following simple result.

Theorem 9 (Boundedness and Bounded Loss). *Let R be a strictly $P/B-$proper scoring relation, and $f_R : \mathcal{P} \to \bar{\mathbb{R}}$ the pointwise supremum of the range of R. f_R is an expected score function for R, and its conjugate restricted to $\mathcal{B}$ is a strictly P/B^+-proper classical cost function $\dot{C}$. $\dot{C}$ has bounded worst-case loss if f_R is bounded above.*

Proof. If f_R is bounded above then the expectation of every scoring function is bounded above, by construction. If a scoring function were unbounded above, then it would be unbounded above on $\mathcal{P}$, too. (There must be a series of measurable sets with increasing real values, and for every measurable set there exists a probability measure that assigns all its mass to that set, so for any value of a scoring function there exist probability measures whose expectation for that function is that value.) Thus, if f_R is bounded above, every scoring function is also bounded above and we have bounded worst-case loss. □

This theorem statement and proof are a nice demonstration of the utility of thinking about strict properness in terms of scoring relations or menus of scoring functions. It lets us easily understand that the only possible payoffs a trader can get are described by scoring functions, and that if these payouts are bounded above so is our loss.

The second property, no arbitrage, is immediately satisfied by any strictly proper cost function. We first define it as:

Definition 7 (No Arbitrage). *A strictly P/B^+-proper cost function C allows no arbitrage when*

$$\inf_{\omega \in \Omega} \big(C(\ell_1) - C(\ell_0)\big)(\omega) \;\leq\; 0, \; \forall \ell_0, \ell_1 \in B^+ \qquad \text{(no arbitrage)}$$

In English, no arbitrage means that every trade has some possibility of being weakly unprofitable. And we have the following formal result:

Theorem 10 (Strictly Proper Cost Functions Permit No Arbitrage). *A P/B^+-proper cost function $C : B^+ \to B$ permits no arbitrage.*

Proof. Assume for a contradiction that there exists liabilities ℓ_0 and ℓ_1 in B^+ such that

$$\inf_{\omega \in \Omega} \big(C(\ell_1) - C(\ell_0)\big)(\omega) \;>\; 0$$

Then every probability measure expects to profit by purchasing the security $C(\ell_1) - C(\ell_0)$, but by properness (and the above argument about prices) the probability measure associated with $C(\ell_0)$ does not expect to profit by purchasing any security when the market has liabilities ℓ_0, so the existence of arbitrage opportunities contradicts properness. $\square$

In fact, when our cost function is strictly proper we can strengthen the above result so the no arbitrage inequality becomes strict. Some papers, like [1] and [31] relax the no arbitrage property as it can be computationally difficult to enforce. These possibilities are discussed further in the conclusion.

Now that we have discussed a market reliably offering securities, quoting prices, having bounded worst-case loss and admitting no arbitrage, we can discuss an actual

cost function for bounded continuous random variables. Cost functions for this space with properties like bounded loss have proven elusive historically.

5.2   A Cost Function for Bounded Continuous Random Variables

Cost functions with good properties for continuous random variables have been discussed for some time, yet no one has yet produced a cost function with bounded loss for this space. In [38] the authors showed markets for continuous random variables could not have bounded loss when satisfying some other economically motivated properties. In [68] a continuous outcome space was discretized, which is a common approach to the problem, and in [67] a cost function for continuous random variables with bounded loss was mistakenly claimed, a claim corrected in the author's thesis [66]. Most work on cost functions has been for discrete spaces, like the best characterization of cost functions for discrete spaces [1], and work especially focused on making cost functions for large discrete spaces tractable [31, 53].

In the previous chapter, however, we saw that describing a cost function for a measurable space is as simple as describing a scoring rule for one, and such scoring rules have been known for some time [58] and were characterized in [43]. In fact, we can begin with any strictly convex function of the probability measures on the $[0, 1]$ interval, like

$$S(p) \;=\; \int_{0,1} F_p^2 \; \mathrm{d}\lambda \qquad\qquad \text{(expected score function)}$$

where F_p is the cumulative distribution function (CDF) identified with the probability

measure $p \in \mathcal{P}$. The conjugate of this function restricted to $\mathcal{B}$ is a classical cost function for the interval

$$\dot{C}(b) \;=\; \sup_{p \in \mathcal{P}} \; \langle p, b \rangle - \int_{[0,1]} F_p^2 \; \mathrm{d}\lambda \qquad \text{(classical cost function)}$$

and since S is bounded above this classical cost function has bounded worst-case loss. It is not so clear what *other* properties this market has, however. It is not even clear, for instance, where the original expected score function is subdifferentiable, or what its subgradients are, and this means we do not understand what sets $P \subseteq \mathcal{P}$ and $B \subseteq \mathcal{B}$ it is strictly P/B^+-proper for.

In this section I will develop a more practical cost function for bounded continuous random variables whose properties we can understand. On a broad, natural class of securities this cost function can be solved for using a convex program with a finite number of variables, too. While this cost function is imperfect, I think it represents an important first step in our understanding of cost functions for continuous outcome spaces.

5.2.1   Unbiased Cost Functions

The classical cost function

$$\dot{C}(b) \;=\; \sup_{p \in \mathcal{P}} \; \langle p, b \rangle - \int_{[0,1]} F_p^2 \; \mathrm{d}\lambda \qquad \text{(classical cost function)}$$

also has an odd property we have not previously discussed. If the market opens with the constant zero function, as we usually expect, then the probability measure the market will initially assume assigns probability one to the event one occurring. That is, it assigns the single point one a probability of one, and the price of securities like

93

[0, .99] is zero! I will call markets like these *biased*, since despite uniform liabilities they assume non-uniform beliefs. Unbiased markets may be considered more natural to work with, and admit a nice computational interpretation, as we will see shortly. Formally, we define bias as:

Definition 8 (Unbiased Market). *Let $\mathcal{P}$ be the set of probability measures, and $\mathcal{B}$ the set of bounded measurable functions, on $([0, 1], \mathcal{L})$. A scoring relation $R \subseteq \mathcal{P} \times \mathcal{B}$ is unbiased when $(p, b) \in R$ implies that for any two measurable sets $L_0, L_1 \in \mathcal{L}$ where b is constant-valued and $b(L_0) = b(L_1)$, we have*

1. *if $\lambda(L_0) = \lambda(L_1) = 0$, then $p(L_0) = p(L_1)$,*

2. *and if $\lambda(L_0), \lambda(L_1) > 0$, then $p(L_0)/\lambda(L_0) = p(L_1)/\lambda(L_1)$.*

A cost function is unbiased when it is the cost function for an unbiased scoring relation.

This formal definition of unbiasedness is a mouthful. It says that if the market's liabilities are the same on two measurable sets with positive Lebesgue measure, then the price of these sets is in proportion to their "size," as determined by Lebesgue measure. For example, if the market's liabilities are 3 on [0, .1] and [.2, .6], being unbiased implies the market assigns four times the probability to the interval [.2, .6] since it is four times as large.

When two sets are Lebesgue-negligible, unbiasedness requires their price be the same if their liabilities are the same. This has some interesting implications about the prices of these Lebesgue-negligible sets, like that every Lebesgue-negligible set with a positive price contains a countable number of point masses whose prices sum to the

price of the set. This in turn means that no market that assumes prices consistent with a singular continuous measure can be unbiased. The market we construct will exclude these measures from consideration.

Before moving on, note that this definition of unbiasedness is particular to markets for the [0, 1] interval. The idea of something being unbiased is always a matter of perspective, and in an arbitrary measurable space it is not clear what perspective is natural, or even if there is always a notion of uniformity. Our familiarity with the [0, 1] interval, Lebesgue measure, and identifying probability measures with CDFs are all reasons why constructing a market for it may be much easier than constructing cost functions in other measurable spaces.

5.2.2   A New Cost Function

The cost function I will soon describe requires some new notation. Instead of accepting securities from all of $\mathcal{B}$, it only accepts securities that are composed of a finite number of constant real-valued segments. I will call these "interval functions" and denote the set of them by $\mathcal{B}_{interval}$. These functions can be described as a collection of tuples associating non-overlapping subintervals of [0, 1] with real numbers, like $([0,.1], 5), ((.1,.3), -3), ([.3,.3], 0) \ldots .$[2] Crucially, this set of of securities is a vector

---

[2]Note that this class is not the set of piecewise-constant functions, since these have no requirement that they consist of a finite number of segments. Further, it is also not the class of simple functions. This latter class is any function (usually circumscribed to the bounded and measurable functions) that only attains a finite number of values. The difficulty with simple functions is that these values can be obtained in unusual ways. For example, the Dirichlet function is simple because it only attains the values zero or one, and it is one on rational numbers and zero on irrational numbers. The practical market I will propose cannot handle this kind of erratic behavior. Unfortunately, my conference paper [23] that this chapter draws heavily from has an error where it uses the simple functions instead of these interval functions.

space.

Correspondingly, I will let $\mathcal{P}_{interval}$ be the set of "interval" probability measures. These are probability measures p that can also be described by tuples of non-overlapping subintervals of $[0, 1]$ with real numbers, and the real value associated with a subinterval specifies the measure p assigns uniformly to that interval. For instance, Lebesgue measure would be $\big(([0,1],1)\big)$, since it is uniform on the entire interval. Importantly, these measures admit only pure point and absolutely continuous parts in Lebesgue decomposition, and they are a convex set.

Now I can define my new cost function for the $[0, 1]$ interval. Since it is useful to have a name for it, I will call it a *dynamic discretization market,* since it effectively allows traders to arbitrarily discretize the interval with their securities. I will define this market as the classical cost function:

$$\dot{C} : \mathcal{B}_{interval} \to \mathbb{R} \qquad\qquad \text{(dynamic discretization market)}$$

$$\dot{C}(b) \;=\; \max_{p \in \mathcal{P}_{interval}} \langle p, b\rangle - \left(\sum_{\omega \in p_{pp}} p^2(\omega) - \int_\Omega \arctan(\frac{\mathrm{d}p_{cont}}{\mathrm{d}\lambda}) \, \mathrm{d}\lambda \right)$$

where p_{pp} is the pure point part of p, which is also treated as a countable set of points in the interval in a minute abuse of notation,[3] and p_{cont} is the absolutely continuous part of p. Again, measures in $\mathcal{P}_{interval}$ have no singular continuous parts.

Intuitively, this market lets traders define their own discretization of the interval. Importantly, unlike an *ex ante* discretization, traders can define this discretization multiple times *ex interim.* This means traders can create the precise discretization they like to best express their beliefs (as long as those beliefs are in $\mathcal{P}_{interval}$). An

---

[3]Recall that in Chapter 2 we saw this set was always a countable collection of point masses.

ex ante discretization may not accomplish this. How valuable, exactly, this is, is a matter for debate. Developing a framework for describing the value of offering a broader set of predictions would be an interesting area for future work.

We are now, of course, interested in what properties this cost function has. Its securities are a vector space, and it is strictly proper, but it's not clear if it exhibits bounded loss, or what set $P \subseteq \mathcal{P}$ it is strictly proper for, or how we might readily compute it. The next two subsubsections do precisely this.

Bounded Loss

For a market to have bounded loss its scoring functions must be bounded above. In this case, bounded loss is equivalent to showing that there exists some $k \in \mathbb{R}$ such that

$$k > \sup_{b \in \mathcal{B}_{interval}, \omega \in [0,1]} b(\omega) - \sup_{p \in \mathcal{P}_{interval}} \langle p, b \rangle - \left(\sum_{\omega \in p_{pp}} p^2(\omega) - \int_{\Omega} \arctan(\frac{\mathrm{d}p_{cont}}{\mathrm{d}\lambda}) \, \mathrm{d}\lambda \right)$$

which follows quickly since the negative arctan function is bounded below, the summation is bounded above, and so the entire term in large parentheses is bounded for all probability measures and can be removed. This leaves the difference:

$$k > \sup_{b \in \mathcal{B}_{interval}, \omega \in [0,1]} b(\omega) - \sup_{p \in \mathcal{P}_{interval}} \langle p, b \rangle$$

which is always less than or equal to zero, since $\sup_{p \in \mathcal{P}_{interval}} \langle p, b \rangle = \sup_{\omega \in [0,1]} b(\omega)$. Thus the market has bounded loss, which we state formally:

Theorem 11 (The Dynamic Discretization Market has Bounded Loss). *The dynamic discretization market has bounded worst-case loss.*

Using the negative arctan function may have seemed like an odd choice. Here we see that it is useful since it is bounded below. In the next subsubsection we will also see how important it is that it is a negative strictly convex function.

Strictly $\mathcal{P}_{interval}/\mathcal{B}_{interval}-$Proper

This subsubsection is complicated, complicated enough to be confusing. Before leaping into our narrative, it will help to describe what we will be doing:

1. Proving the conjugate of the dynamic discretization market is strictly convex on $\mathcal{P}_{interval}$.

2. Showing the market is unbiased.

3. Demonstrating the market is strictly $\mathcal{P}_{interval}/\mathcal{B}_{interval}-$proper.

As a side effect, I will show that the market can be computed using a convex program with a finite number of variables, one variable per every interval described by the market's current liabilities.

First, we need to show the market's conjugate is strictly convex on $\mathcal{P}_{interval}$. This result will be needed for the next two.

Lemma 4. *The function*

$$S : \mathcal{P}_{interval} \to \mathbb{R}$$

$$S(p) \;=\; \sum_{\omega \in p_{pp}} p^2(\omega) - \int_{\Omega} \arctan(\frac{\mathrm{d}p_{cont}}{\mathrm{d}\lambda}) \; \mathrm{d}\lambda$$

is a strictly convex function of the interval probability measures.

Proof. Let p_0, p_1 be two distinct probability measures in $\mathcal{P}_{interval}$. We need to show that

$$\alpha \left(\sum_{\omega \in (p_0)_{pp}} p_0^2(\omega) - \int_\Omega \arctan(\frac{\mathrm{d}(p_0)_{cont}}{\mathrm{d}\lambda}) \, \mathrm{d}\lambda \right) + (1-\alpha) \left(\sum_{\omega \in (p_1)_{pp}} p_1^2(\omega) - \int_\Omega \arctan(\frac{\mathrm{d}(p_1)_{cont}}{\mathrm{d}\lambda}) \, \mathrm{d}\lambda \right)$$
$$> \sum_{\omega \in (p_0+p_1)_{pp}} (\alpha p_0 + (1-\alpha)p_1)^2(\omega) - \int_\Omega \arctan(\frac{\mathrm{d}(\alpha p_0 + (1-\alpha)p_1)_{cont}}{\mathrm{d}\lambda}) \, \mathrm{d}\lambda$$

and proceed by analyzing two cases. First, assume that $(p_0 + p_1)$ has a pure point part where p_0 and p_1 differ, then

$$\alpha \sum_{\omega \in (p_0)_{pp}} p_0^2(\omega) + (1-\alpha) \sum_{\omega \in (p_1)_{pp}} p_1^2(\omega) > \sum_{\omega \in (p_0+p_1)_{pp}} (\alpha p_0 + (1-\alpha)p_1)^2(\omega)$$

since $f(x) = x^2$ is strictly convex, so $\alpha p_0^2(\omega) + (1-\alpha)p_1^2(\omega) \geq (\alpha p_0 + (1-\alpha)p_1)^2(\omega)$ for all $\omega \in \Omega$ and the inequality holds strictly where the measures are distinct, and thus the inequality holds for the sum, too (since the sum is finite and bounded).

Alternatively if $(p_0 + p_1)$ has an absolutely continuous part where p_0 and p_1 differ, then

$$-\alpha \int_\Omega \arctan(\frac{\mathrm{d}(p_0)_{cont}}{\mathrm{d}\lambda}) \, \mathrm{d}\lambda - (1-\alpha) \int_\Omega \arctan(\frac{\mathrm{d}(p_1)_{cont}}{\mathrm{d}\lambda}) \, \mathrm{d}\lambda$$
$$> - \int_\Omega \arctan(\frac{\mathrm{d}(\alpha p_0 + (1-\alpha)p_1)_{cont}}{\mathrm{d}\lambda}) \, \mathrm{d}\lambda$$

because $-\arctan$ is a strictly convex function, allowing us to apply Lemma 1.

Since we assumed $p_0 \neq p_1$ they must differ on their pure point or absolutely continuous parts, so one of the last two inequalities above must hold strictly and their summation proves the desired original inequality. $\square$

A helpful way to think of this conjugate is as a "regularization" or "penalty" function for our cost function. It prevents the cost function from assuming extremal

beliefs by penalizing them. The strict properness of this penalty function will be crucial to the next two proofs.

Next on our list was showing the market is unbiased.

Theorem 12 (Unbiasedness of the Dynamic Discretization Market). *The dynamic discretization market is unbiased.*

Proof. The proof proceeds by contradiction in two cases. First assume that there exists $b \in \mathcal{B}_{interval}$ such that the corresponding probability measure p_b is biased. One possibility is that there exist two Lebesgue-negligible measurable sets L_0 and L_1 where b is constant and equally valued on both sets, yet p_b is such that that $p(L_0) \neq p(L_1)$. I will show we can improve on this supremum, since

$$\langle p, b \rangle - p^2(L_0) - p^2(L_1) \; < \; \langle p, b \rangle - 2(\alpha p_0 + (1 - \alpha)p_1)^2(L_0 + L_1)$$

for all $\alpha \in (0, 1)$ by the strict convexity of $f(x) = x^2$. Thus we can improve on this probability measure by equalizing the probability assigned to the sets L_0 and L_1, contradicting our assumption that the market would create such an unbiased measure.

Alternatively, there are two Lebesgue-measurable sets L_0 and L_1 such that $\lambda(L_0)$ and $\lambda(L_1)$ are both greater than zero, b is constant and equally valued on both sets, yet p_b is such that $\frac{p(L_0)}{\lambda(L_0)} \neq \frac{p(L_1)}{\lambda(L_1)}$. Again, we can improve on this supremum, since

$$\langle p, b \rangle - \int_{L_0} \arctan(\frac{dp}{d\lambda}) \, d\lambda - \int_{L_1} \arctan(\frac{dp}{d\lambda}) \, d\lambda \; < \; \langle p, b \rangle - \arctan(\frac{p(L_0 + L_1)}{\lambda(L_0 + L_1)})$$

for all $\alpha \in (0, 1)$, following again by the strict convexity of the negative arctan function.

So any unbiased probability measure can be improved upon, and thus we conclude that the market cannot assume such a measure as the solution to its supremum expression. Hence, the market is unbiased. □

Unbiasedness is a natural property for a market to have, and it will also let us solve the market using a convex program. The next proof is interesting because it first shows we can use a convex program to find a solution for the market, then takes some results from convex analysis in Euclidean spaces to prove the market is strictly $\mathcal{P}_{interval}/\mathcal{B}_{interval}-$proper.

Theorem 13 (Strict $\mathcal{P}_{interval}/\mathcal{B}_{interval}-$properness of the Dynamic Discretization Market)**.** *The dynamic discretization market is strictly $\mathcal{P}_{interval}/\mathcal{B}_{interval}-$proper.*

Proof. We begin by showing we can solve the dynamic discretization market using a convex program, which requires showing that the solution to the supremum

$$\sup_{p \in \mathcal{P}} \langle p, b \rangle - \left(\sum_{\omega \in p_{pp}} p^2(\omega) - \int_{\Omega} \arctan(\frac{\mathrm{d}p_{cont}}{\mathrm{d}\lambda}) \, \mathrm{d}\lambda \right)$$

is always in $\mathcal{P}_{interval}$.

Every function in $\mathcal{B}_{interval}$ can be described as a finite subset of non-overlapping subintervals I of $[0, 1]$ associated with real values $(I_0, r_0), (I_1, r_1), \ldots)$. Any possible solution, then, is distinguished by how much probability it assigns to each interval and how it does so. If an interval is degenerate, the only mass that can be assigned to it is through a pure point part. If the interval has positive length, on the other hand, mass can be assigned to it through a combination of pure point, singular continuous or absolutely continuous parts. The first lowers the value of the supremum, the second

leaves it unchanged, and the last actually *increases* it.[4]

This thinking implies any mass assigned to a non-degenerate interval must be in an absolutely continuous part, and now we need to see this absolutely continuous part is uniform on each interval. Luckily, this follows immediately from the definition of unbiasedness, which we just proved. This argument shows that if the supremum attains a solution it does so in $\mathcal{P}_{interval}$.

We can now use this fact to solve for $\dot{C}$ using a convex program:

$$\dot{C}(b) \;=\; \max_{p \in \Delta(\mathbb{R}^{|N|})} \sum_{i \in N} r_i p_i - \left(\sum_{i \in M} p_i^2 - \sum_{i \in O} \arctan \frac{p_i}{\lambda(L_i)} \right)$$

where N indexes all the pairings of intervals and reals defining b, and M indexes the degenerate intervals and O indexes the non-degenerate intervals. This program can alternatively be expressed as minimizing a strictly convex function minus a linear function over a closed convex space, and this implies it always attains a solution. Thus we conclude that the solutions to the practical security market are in $\mathcal{P}_{interval}$, and the market attains such solutions for any $b \in \mathcal{B}_{interval}$.

We can now use this result to investigate the strict properness of our dynamic discretization market. The market attaining a unique maximum in $\mathcal{P}_{interval}$ at each point in $\mathcal{B}_{interval}$ implies it is strictly $?/\mathcal{B}_{interval}-$proper, where ? is some (still unknown to us) subset of $\mathcal{P}_{interval}$. We will use another result about convex programs to identify the space of beliefs.

Let Π be a partition of the $[0, 1]$ interval, and $\mathcal{B}_\Pi$ the functions that are constant-valued on each element of this partition, representable as functions $b : \Pi \to \mathbb{R}$. A

---

[4]These features are vital for my argument. I need the pure point part be penalized with a positive strictly convex function, and the absolutely continuous part to be "rewarded" with another strictly convex function.

convex program with $|\Pi|$ variables can solve for $\dot{C}$ over $\mathcal{B}_\Pi$, and can be written as

$$\dot{C}(b) \;=\; \max_{p \in \Delta(\Pi)} \; \sum_{\pi \in \Pi} b(\pi)p(\pi) - \left(\sum_{\pi \in \Pi, \lambda(\pi)=0} p^2(\pi) - \sum_{\pi \in \Pi, \lambda(\pi)>0} \arctan \frac{p(\pi)}{\lambda(\pi)} \right)$$

Importantly, any probability measure in $\Delta(\Pi)$ can be obtained as a solution to this expression, because any Dirac measure is obtainable and the subdifferential of a convex function is a closed and convex set. This implies that any measure in $\mathcal{P}_{interval}$ is obtainable as a solution, and our cost function is strictly $\mathcal{P}_{interval}/\mathcal{B}_{interval}-$proper, as desired. $\qquad\square$

I think the above proof is incredibly interesting. It begins by showing that a convex program can solve for the classical cost function at any point, then using facts about convex programs to prove properties of the function as a whole. It is truly a fascinating technique that I have not seen elsewhere.

5.3 Practical Cost Functions in Review

The dynamic discretization market just discussed is a strictly $\mathcal{P}_{interval}/\mathcal{B}_{interval}-$proper cost function for the $[0, 1]$ interval. It has bounded loss and is unbiased. Being strictly proper it also admits no arbitrage. Since $\mathcal{B}_{interval}$ is a vector space, it can be thought of as reliably offering every security for trade, and it can quote meaningful prices, too. Finally, at any point we can use a convex program to solve for both the classical cost function and cost function version of the market, letting us extract a prediction and obtain a cost readily.

Before this, markets for $[0, 1]$ that experienced bounded loss were unknown. Some prior work, notably [2, 3] and the unpublished [36], has begun to approach my char-

acterization of strict properness, and could likely be developed into a similar result. It seems the community has been getting ready for this moment, and that we are all collectively on its threshold. This thesis is simply one of the first past the post, and I am sure there will soon be many more.

The idea of a market being biased or unbiased is new. Unbiased markets can occur even in discrete spaces, and these markets were likely not noticed because in those settings it is very natural to work with a probability distributions PDF over its CDF. Working with a PDF avoids many of the issues of bias, and working with a CDF makes developing unbiased markets far trickier. Not only do unbiased markets make more intuitive sense, they also offer practical computational speedups by letting us solve them using convex programs.

In [1] (the best paper characterizing classical cost functions for discrete spaces) the authors also discuss how they can solve for their classical cost functions of interest using a convex program. They go farther than I do in this analysis, describing it for the entire class of classical cost functions on discrete spaces. They also suggest a fascinating way of relaxing the convex program so that it is easier to compute. This sacrifices no arbitrage, although it maintains other desirable properties. For discrete spaces this is a superb practical guide, and it suggests a great opportunity for further work.

The dynamic discretization market, despite all its nice new properties, is still far from ideal. It only offers securities from $\mathcal{B}_{interval}$, and can only express beliefs from $\mathcal{P}_{interval}$. This may or may not be an issue in practice, depending on the complexity of traders' beliefs. In settings where beliefs are especially complicated it may be

preferable to use a scoring rule over a cost function. The value of a cost function is implicitly predicated on traders not holding complete beliefs, and when they are so complex it suggests it is easier for traders to specify their beliefs directly than trade securities. Matheson's scoring rule [58] is an extremely easy to compute and strictly proper scoring rule for this setting.

Clearly there is still a great deal of work to do on developing cost functions, and I will return to this discussion in the conclusion.

6

Designing Informative and Simple Prediction Markets

Chapters 3–5 have discussed strict properness. As we saw in Chapter 5, however, strict properness is not the *only* property of interest for prediction markets. This chapter investigates the *design* of markets that are both *simple* and *informative*. These are markets where we learn the likelihood of our events of interest as if we knew all the traders' private information, and we do so using as few securities as possible. Running a market with fewer securities is computationally easier than running one with many securities, and so removing superfluous securities is a natural design goal.

This chapter's setting is distinct from the previous chapters'. I will assume only a finite number of states of the world Ω,[1] and that traders are Bayesian agents with a common prior and knowledge of how their private information is structured. Further,

---

[1]Implicitly interpreted as part of the measurable space $(\Omega, 2^\Omega)$ since we are working in a discrete setting.

I will use the same mathematical abstraction as [65], which has traders predicting the expected future prices of the securities. This prediction market does not at first resemble the scoring rule or cost function markets we have described so far, although we can think of it as a cost function market where the market maker does not enforce feasible prices. This setting offers the necessary structure to formalize our discussion of information aggregation in prediction markets and let us use results from [65] without a great deal of work translating them to another abstraction.

Semi-formally, in this chapter I consider there are some *events of interest* we are interested in learning the likelihood of. We offer a bundle of securities to the traders, and they begin exchanging information about their private information / signals by offering public predictions in the market. The more securities available the more expressive the traders can be, and the more likely they will reach a consensus that reveals the likelihood of the events of interest as if we knew all the traders' private information. Of course, having too many securities is undesirable, so our goal is to find the fewest securities we can offer such that traders' will most accurately reveal what we would like to know.

As mentioned in the introduction, a good example of this design challenge is creating a market for corn futures. There are many variables that impact the future price of corn. The amount of sunshine and rain, future demand for corn, the introduction of new varieties of corn and growing techniques, etc.. Traders may have varying levels of knowledge about each of these variables. Some may be experts at predicting the weather (meteorologists), others are experts at understanding future demand (cereal producers). To pool all their information, then, we must offer them enough securities

so they can express this knowledge. At the same time, however, we do not want to offer superfluous securities, and there may be clever ways to offer one security instead of two. Finding just the right, minimal set of securities to let the market best reveal the future price of corn is our challenge.

The above informal example provides a helpful intuition, and I think the mathematical complexities also suggest a more precise and formal example that demonstrates the tension between informativeness and simplicity. This example is derived from [41].

Example 1. *Consider a market offering a single security worth \$1 if a particular candidate wins the U.S. presidential election and \$0 otherwise. The market has two participants: a political analyst in Washington and an Iowa caucus-goer who is well-informed on local politics. The analyst understands the importance of Iowa on the campaign and knows whether a win or loss there will mean the candidate is elected. The caucus-goer, on the other hand, knows whether the candidate will win or lose the caucus, but not its broader effect.*

This situation can be described by defining four states of the world, ω_1, ω_2, ω_3, and ω_4: The analyst knows if the true state of the world is on the diagonal or not

		Iowa	
		Wins	Loses
General Election	Wins	ω_1	ω_2
	Loses	ω_3	ω_4

(the effect of the caucus) and the caucus-goer knows which column the true state is in (the results of the caucus). If they could reveal their private information they would learn the true state of the world, ω^*. But with a uniform prior over the state

space, both think the likelihood of election is 1/2 and value the security at $0.50 no matter what their private information is, since every signal contains a state where the candidate wins the election and another where the candidate loses. Thus the market closes without accomplishing anything useful, with the final security price still being $0.50.

This market is clearly very simple, and it is not informative. If, instead, it offered one security for each state of the world, however, then the traders could perfectly express their private beliefs and combine their knowledge to better predict the likely results of the presidential race.

Semi-formally, I will call a market informative if it does two things:

1. It offers a set of securities so that, in perfect Bayesian equilibrium, as traders continue trading the value of these securities converges to their expectation conditional on all traders' private information / signals. Markets with this property are said to aggregate their information.

2. The prices of these securities always reveals the likelihood of each event of interest.

This first property was studied by [65], who showed it was dependent on the securities being *separable,* a technical condition we will review shortly. One problem with prices being accurate alone, however, is they may not reveal what we need to know, hence the second property. It may sound odd that a market can aggregate information ins a useless manner. Trivially, we might only decide to offer securities related to the weather when trying to predict the price of corn. The market may perfectly aggregate all traders' knowledge about these securities and accurately determine their prices,

but we can infer nothing from them about the future price of corn. More generally, we are concerned that while prices may be correct, they can be consistent with many possible interpretations of the likelihood of our events of interest. Informativeness requires both that (1) prices always aggregate all traders' information and (2) we can use these prices to uniquely infer a single interpretation of the likelihood of our events of interest.

This chapter also focuses on simplicity, and a market is both *simple* and *informative* when

1. it is informative,

2. it offers as few securities as possible,

3. and these securities are associated with events, offering $1 if an event occurs and $0 otherwise.

I already mentioned the first two properties; the third is a naturalness condition on the type of securities we may offer. These securities appear common in practice, and I will show that with more exotic securities odd results are possible. We might, for example, be able to use a single carefully and strangely designed security to represent a lot of complex information, and it is unlikely real traders would be able to work with such a security.[2]

The rest of this chapter is organized as follows. A discussion of related work appears in Section 6.1, followed by a formal description of our model in Section 6.2. Section 6.3 discusses information aggregation in prediction markets, building on prior work to show the importance of offering securities that are *separable*. Section 6.4

---

[2]We rarely see a single security acting as a summary statistic for an entire market.

begins our discussion of designing simple and informative markets, and Section 6.5 concludes.

How we are able to design markets that are simple and informative depends on our understanding of the structure of trader's knowledge. Returning to our corn example, it depends on how well we understand what variables are relevant to future corn prices. Our mathematical example also shows it requires understanding how traders information is related. This leads to the following breakdown of Section 6.4 on design:

- In Section 6.4.2, I assume we know nothing of how traders' information is structured. In this setting determining the likelihood of an event requires as many securities as there are outcomes in that event or its complement (minus one), possibly a prohibitive number in practice. This section also discusses complete markets, and how they are the only markets that always reveal the likelihood of *every* event.

- Section 6.4.3 shows that with perfect knowledge of traders' signal structure a *single* security can create an informative market. This security is likely too strange to use in practice, however, and this market should not be considered simple. This section motivates our restriction to only use securities associated with an events, paying \$1 if that event occurs and \$0 otherwise.

- Finally, Section 6.4.4 considers designing simple and informative markets given perfect knowledge of the trader's signal structure. In this case, designing a simple and informative market is NP-hard.

These results are a little disheartening. Designing simple and informative markets

is either trivial when we know nothing about how traders' knowledge is structured, but requires a prohibitive number of securities. On the other hand, when we know everything about how traders' knowledge is structured their design is NP-hard. Still, I think this chapter is interesting as it introduces this natural problem of how we can create markets that are informative and simple. Perhaps these results will inspire future work to look harder at information aggregation occurring *outside the market*, or develop a formal theory of *partial* aggregation that is more tractable, or provide further motivation for developing markets that can handle very large outcome spaces, like the work of [53] and [31].

This chapter references some material in an appendix that appears on Yiling Chen's website[3]. I have decided not to include this material for the sake of a streamlined narrative.

6.1 Related Work

Information aggregation and the design of prediction markets have been discussed in many other papers. A series of papers have shown that prediction markets are empirically effective in settings like politics [10], business [86, 26], disease surveillance [79], and entertainment [73]. Experiments with predictions markets have also shown them effective [76, 77, 78], and substantial work has analyzed the theory of how markets aggregate information, including at rational expectations [80, 6, 45], competitive [87, 74], and game theory equilibria [65, 51, 18].

The early foundations for the study of information aggregation come from [8],

---

[3]http://yiling.seas.harvard.edu/wp-content/uploads/informativesecurities.pdf

whose abstract results are refined and extended by [41, 62, 60] and [63]. These papers show that a set of Bayesian agents regularly announcing their posterior probability distributions over a set of outcomes will eventually converge in their beliefs. Critically, these papers do not discuss whether this convergence is *informative*, only that the agents reach some impassable consensus.

In a prediction market the above is analogous to traders agreeing on the price of each security. As mentioned in the preface to this chapter, this agreement may not reveal anything useful. Some prior work on prediction markets has investigated what I call "informativeness," where traders reach a revealing consensus. In [32] informative Shapley-Shubik markets (see [84]) were characterized, assuming straightforward, non-strategic traders. And [65] characterized information aggregation in prediction markets with strategic, risk-neutral traders at perfect Bayesian equilibrium. This latter paper demonstrated the importance of a market being sufficiently expressive to let this aggregation occur, and the separability property it develops is essential to my work. An extended discussion of this paper and its results appears in Section 6.3. [51] generalized this model to risk-averse agents.

The work of [32], [65], and [51] focuses on understanding the aggregation of information relevant to the value of a given, fixed security. My work differs because it considers *design*: how we can simplify markets for large outcome spaces like the 9.2. quintillion outcomes of the NCAA tournament [88], the over 2^{50} ways for states to vote in the U.S. Presidential election, and the $n!$ rankings for a competition with n candidates, while keeping these markets informative. Offering a security for each state would be theoretically informative and practically unmanageable. Prior work

on simplifying markets, like [19], has not considered whether the result is informative.

6.2   Formal Model

In this section, I describe my model of traders' information and the market mechanism. This model closely follows [65], but is generalized to handle a *vector* of securities (often simply referred to as a set of securities) instead of a single security.

6.2.1   Modeling Traders' Information

We will consider n traders, $1, \cdots, n$, and a finite set Ω of mutually exclusive and exhaustive states of the world. Traders share a common knowledge prior distribution P_0 over Ω. Before the market opens Nature draws a state ω^* from Ω according to P_0 and traders learn some information about ω^* that, following [8], is based on partitions of Ω. A *partition* of a set Ω is a set of nonempty subsets of Ω such that every element of Ω is contained in exactly one subset. For example, $\{\{A, B\}, \{C\}, \{D\}\}$ and $\{\{A, D\}, \{B, C\}\}$ are both partitions of $\{A, B, C, D\}$. I assume that every trader i receives $\Pi_i(\omega^*)$ as their private signal, where $\Pi_i(\omega)$ denotes the element of the partition Π_i that contains ω. In other words, trader i learns that the true state of the world lies in the set $\Pi_i(\omega^*)$.

I refer to the vector $\Pi = (\Pi_1, \cdots, \Pi_n)$ as the traders' *signal structure*, which is assumed to be common knowledge for all traders. The *join* of the signal structure, denoted join(Π), is the coarsest common refinement of Π, that is, the partition with the smallest number of elements satisfying the property that for any ω_1 and ω_2 in the same element of the partition, $\Pi_i(\omega_1) = \Pi_i(\omega_2)$ for all i. For example, the join of the

partitions $\{\{A, D\}, \{B, C\}\}$ and $\{\{A, C, D\}, \{B\}\}$ is $\{\{A, D\}, \{B\}, \{C\}\}$. The join is unique. I use $\Pi(\omega)$ to denote the element of the join containing ω. Note that if two states appear in the same element of the join, no trader can distinguish between these states.

6.2.2 Market Scoring Rules

The market mechanism that we will consider is a market scoring rule [48, 49]. In this chapter, I will describe a market scoring rule as a mechanism that allows traders to sequentially report their probability distributions or expectations. While focusing on market scoring rules may seem restrictive, market scoring rules are surprisingly general. In particular, any market scoring rule that allows traders to report probability distributions over Ω has an equivalent implementation as a cost-function-based market where the mechanism acts as an automated market maker who sets prices for $|\Omega|$ Arrow-Debreu securities, one for each state and taking value 1 in that state and 0 otherwise, and is willing to buy and sell securities at the set prices [48, 25].

This result can easily be extended to general scoring rules by applying the results of [2, 3]. In particular, their results imply that any market scoring rule that allows traders to report their expectations has an equivalent implementation as a cost-function-based market that allows traders to trade securities with the market maker. Thus, without loss of generality, my model and analysis are presented for market scoring rules.

Before describing the market scoring rule mechanism, let's first review the idea of a strictly proper scoring rule. *Scoring rules* are most frequently used to evaluate and

incentivize probabilistic forecasts [39, 43], but can also be used to elicit the mean or other statistics of a random variable [54]. The scoring rules that we consider will be used to elicit the mean of a *vector* of random variables [83]. Let $\mathbf{X} = (x_1, \cdots, x_m)$ be a vector of bounded real-valued random variables. A scoring rule s maps a forecast $\vec{y}$ in some convex region $\mathcal{K} \subseteq \mathbb{R}^m$ (e.g., the probability simplex in the case of probabilistic forecasts) and a realization of $\mathbf{X}$ to a score $s(\vec{y}, \mathbf{X}(\omega))$ in $\mathbb{R}$.[4] A scoring rule for eliciting an expectation is said to be *proper* if a risk neutral forecaster who believes that the true distribution over states Ω is P maximizes his expected score by reporting $\vec{y} = \mathrm{E}_P[\mathbf{X}]$, that is, if $\mathrm{E}_P[\mathbf{X}] \in \arg\max_{\vec{y} \in \mathcal{K}} \sum_{\omega \in \Omega} P(\omega) s(\vec{y}, \mathbf{X}(\omega))$. (For random vectors $\mathbf{X}$, I use $\mathrm{E}_P[\mathbf{X}]$ to denote the expected value $\Sigma_{\omega \in \Omega} P(\omega) \mathbf{X}(\omega)$.) A scoring rule is *strictly proper* if $\mathrm{E}_P[\mathbf{X}]$ is the unique maximizer.[5]

One common example of a strictly proper scoring rule is the Brier scoring rule [16], which is based on Euclidean distance and can be written, for any $b > 0$, as $s(\vec{y}, \mathbf{X}(\omega)) = -b\sum_{j=1}^{m}(y_j - x_j(\omega))^2 = -b||\vec{y} - \mathbf{X}(\omega)||^2$.

Strictly proper scoring rules incentivize myopic traders to report truthfully, but do not provide a mechanism for aggregating predictions from multiple traders. Hanson [48, 49] introduced *market scoring rules* to address this problem. A market scoring rule is a sequentially shared strictly proper scoring rule.[6]

---

[4]Technically, the region $\mathcal{K}$ should include the convex hull of the possible realizations of $\mathbf{X}$, a set equivalent to the possible expected values of $\mathbf{X}$. A full discussion of this and other properties of scoring rules is beyond the scope of this thesis, but interested readers can see [83].

[5]This discussion is mathematically redundant with my discussion earlier in the thesis. I think it is valuable to contextualize the prior work into the setting of discrete probability theory, however. Also note that earlier I did not discuss eliciting a statistic, only eliciting the complete belief. Eliciting a statistic is a common goal for a scoring rule, and the earlier analysis can be easily extended to this case.

[6]The descriptor "sequentially shared ..." is meaningless. It has been adopted as a kind of cant

Formally, let $\mathbf{X}$ be a vector of random variables.[7] The market operator specifies a strictly proper scoring rule s and chooses an initial prediction $\vec{y}_0$ for the expected value of $\mathbf{X}$; when there is a known common prior P_0, it is most natural to set $\vec{y}_0 = \mathrm{E}_{P_0}[\mathbf{X}]$. The market opens with initial prediction $\vec{y}_0$, and traders take turns submitting predictions. The order in which traders make predictions is common knowledge. Without loss of generality, I assume that traders $1, 2, \cdots, n$ take turns, in order, submitting predictions $\vec{y}_1, \vec{y}_2, \cdots, \vec{y}_n$, then the process repeats and the traders, in the same order, submit predictions $\vec{y}_{n+1}, \vec{y}_{n+2}, \cdots, \vec{y}_{2n}$. Traders repeat this process an infinite number of times before the market closes and Nature reveals ω^*. Each trader then receives a score $s(\vec{y}_t, \mathbf{X}(\omega^*))$ for each prediction made at some time t, but must pay $s(\vec{y}_{t-1}, \mathbf{X}(\omega^*))$, the score of the previous trader. The total payment to trader i (which may be negative) is then $\sum_{t=0}^{\infty} s(\vec{y}_{tn+i}, \mathbf{X}(\omega^*)) - s(\vec{y}_{tn+i-1}, \mathbf{X}(\omega^*))$.

6.2.3 Modeling Traders' Behavior

Together, the traders, state space, signal structure, security vector, and market scoring rule mechanism define an extensive form game with incomplete information. I consider Bayesian traders either acting in perfect Bayesian equilibrium or behaving myopically in this game. A perfect Bayesian equilibrium is a subgame perfect Bayesian Nash equilibrium. Loosely speaking, at a perfect Bayesian equilibrium, it must be the case that each player's strategy is optimal (i.e., maximizes expected

---

to describe how prediction markets work. I think this is simply because the literature has done such a poor job formalizing these mechanisms it has turned to magical insider language as its best recourse.

[7]Typically market scoring rules are used for probabilistic forecasts in which case $\mathbf{X}$ would be a vector of indicator random variables, but this need not be the case.

utility) given the player's beliefs and the strategies of other players at any stage of the game, and that players' beliefs are derived from strategies using Bayes' rule whenever possible. See Gonzálex-Díaz and Meléndez-Jiménez [44] for a more formal description.[8]

Perfect Bayesian equilibria can be difficult to compute and it is an open question whether they always exist in prediction markets, although in some special cases they do [18]. An alternative is to consider *myopic* Bayesian traders who simply maximize their expected payoff for the current round. Since strictly proper scoring rules myopically incentivize honest reports, these traders report their current posteriors each time they make a prediction.

6.3   Information Aggregation

Separability is used to characterize the conditions under which securities aggregate information about their own values. Building on ideas from DeMarzo and Skiadas [29, 30], [65] characterized separability for a single security. [65] showed that in every perfect Bayesian equilibrium market prices will, in the limit, reflect the value of the security as if traders had revealed their private signals if and only if the security is separable. If a security is not separable, then there always exist priors and equilibrium strategies where no information aggregation occurs.

In this section, I generalize these prior definitions to multiple securities and arbitrary signal structures. Ostrovsky assumed a restricted class of signal structures without loss of generality, and my generalizations are uninteresting when only consid-

---

[8]Finding a good, formal description of PBE is very, very, very hard.

ering aggregation. They will be necessary to discuss informativeness, however, as the results of the next section demonstrate. I will then restate Ostrovsky's equilibrium aggregation result in this setting. As previously discussed, perfect Bayesian equilibrium may or may not exist in prediction markets, and I also adapt and formalize prior work on information aggregation to show separability is also the necessary and sufficient condition for myopic traders to always aggregate their information.

Informative markets require separable securities. If a market uses separable securities then both Bayesian traders acting myopically and in perfect Bayesian equilibrium will, in the limit, value the security as if their private signals were revealed, and this allows a market designer to directly infer the likelihood of his events of interest from the securities' value. If a set of non-separable securities were used then the market designer could be required instead to perform additional inference and know the prior and traders' strategies.

As mentioned, I say a market aggregates information if, in the limit as time goes to infinity, the value of the securities approaches their value conditional on all the traders' private signals. Since each trader i receives the signal $\Pi_i(\omega^*)$, their pooled signal is $\bigcap_i \Pi_i(\omega^*) = \Pi(\omega^*)$.

Definition 9 (Aggregation). *Information is* aggregated *with respect to a set of securities* $\mathbf{X}$, *signal structure* Π, *and common prior* P_0, *if the sequence of predictions* $\vec{y}_0, \vec{y}_1, \vec{y}_2, \cdots$ *converges in probability to the random vector* $E_{P_0}[\mathbf{X}|\Pi(\omega^*)]$.

A set of securities is separable if and only if the traders only agree on their value when it reflects their pooled information. That is, for any prior distribution there must be at least one trader whose private information causes them to dissent from a

consensus, unless that consensus is the traders' collective best estimate.

Definition 10 (Separability). *A set of securities* **X** *is* non-separable *under partition structure* Π *if there exists a distribution P over Ω and vector $\vec{v}$ such that $P(\omega) > 0$ on at least one state $\omega \in \Omega$ in which $E_P[\mathbf{X}|\Pi(\omega)] \neq \vec{v}$, and for every trader i and state $\omega, P(\omega) > 0$,*

$$E_P[\mathbf{X}|\Pi_i(\omega)] = \frac{\sum_{\omega' \in \Pi_i(\omega)} P(\omega')\mathbf{X}(\omega')}{\sum_{\omega' \in \Pi_i(\omega)} P(\omega')} = \vec{v}. \tag{6.1}$$

If a security is not non-separable then it is separable.

Here the vector $\vec{v}$ represents a possible consensus, only agreed upon if there is no alternative when the securities are separable.

Separability is a property of the entire set of securities, as Example 2 demonstrates.

Example 2. *Let $\Omega = \{\omega_1', \omega_2^*, \omega_3, \omega_4', \omega_5^*, \omega_6\}$. Two traders have partitions as follows:*

$$\Pi_1 = \{\{\omega_1', \omega_2^*, \omega_3\}, \{\omega_4', \omega_5^*, \omega_6\}\}$$

$$\Pi_2 = \{\{\omega_1', \omega_5^*\}, \{\omega_3, \omega_4'\}, \{\omega_2^*, \omega_6\}\}$$

and there are two securities: x^* with value one when ω_2^* or ω_5^* occurs and zero otherwise, and x' with value one when ω_1' or ω_4' occurs and zero otherwise.

Both securities are individually non-separable with respect to Π. If the prior P is uniform over $\omega_1', \omega_2^*, \omega_5^*$, and ω_6, then $E_P[x^*|\Pi_i(\omega)] = 1/2$ for $i \in \{1, 2\}$ and all ω such that $P(\omega) > 0$. Similarly, if P is uniform over $\omega_1', \omega_3, \omega_4'$, and ω_5^*, then $E_P[x'|\Pi_i(\omega)] = 1/2$ for $i \in \{1, 2\}$ and all ω such that $P(\omega) > 0$. The join of traders'

partitions, however, consists of singletons. Hence, both $E_P[x^*|\Pi(\omega)]$ and $E_P[x'|\Pi(\omega)]$ have value 0 or 1, not 1/2, for all ω.

But taken together the set of securities is separable with respect to Π. Given any prior distribution P and a state ω, trader 2 either identifies ω with certain, which happens when P assigns 0 probability to the other state in its signal $\Pi_2(\omega)$, or assigns positive probability to both states in $\Pi_2(\omega)$. In the former case, $E_P[\mathbf{X}|\Pi_2(\omega)] = E_P[\mathbf{X}|\Pi(\omega)]$. In the latter case, trader 2's expected value for the securities is positive for both when $\omega \in (\omega_1', \omega_5^*)$, positive for only x' when $\omega \in (\omega_3, \omega_4')$, and positive for only x^* when $\omega \in (\omega_2', \omega_6)$. If the set of securities is non-separable there must exist a distribution $\tilde{P}$ and a vector $\vec{v}$ such that $\vec{v} \neq E_{\tilde{P}}[\mathbf{X}|\Pi(\tilde{\omega})]$ for some state $\tilde{\omega} \in \{\omega|P(\omega) > 0\}$ and $E_{\tilde{P}}[X|\Pi_2(\omega)] = \vec{v}$ for any state $\omega \in \{\omega|P(\omega) > 0\}$. This is possible only when $\tilde{P}$ assigns positive probability to the two states in $\Pi_2(\tilde{\omega})$ and 0 probability for all other states because each signal of player 2 has a distinct expectation of the securities. Given such a $\tilde{P}$, however, trader 1 always uniquely identifies the true state and has the correct expectation of the securities. Hence, the set of securities is separable with respect to Π.

6.3.1 Aggregation

Separability is a necessary and sufficient property for aggregation in two natural cases.

Theorem 14 (Equilibrium Aggregation, [65])**.** *Consider a market with securities $\mathbf{X}$ and traders with signal structure Π. Information is aggregated in every perfect Bayesian equilibrium of this market if and only if the securities $\mathbf{X}$ are separable under*

Π.

Theorem 15 (Myopic Aggregation). *Consider a market with securities* $\mathbf{X}$ *and myopic traders with signal structure* Π. *Information is aggregated in finite rounds if and only if the securities* $\mathbf{X}$ *are separable under* Π.

Ostrovsky [65] proved a special case of Theorem 14 for markets with one security. Theorem 14 stated above accommodates any finite set of securities and is proved using a simple extension of Ostrovsky's proof. Specifically, the proof shows that traders' sequences of predictions at any perfect Bayesian equilibrium are bounded martingales and must converge. Separability implies that if information is not aggregated in the limit, there exists an agent who can make an arbitrarily large profit by deviating from his equilibrium strategy, a contradiction to traders being in equilibrium.

The proof of Theorem 15 makes use of prior work on convergence to common knowledge (particularly [40]) and shows not only that myopic traders' sequences of predictions are bounded martingales but also that they must converge to the same random vector in a finite number of periods. Then, by separability, it is shown that this consensus prediction must equal $\mathrm{E}[X|\Pi(\omega^*)]$, implying aggregation. A full proof appears in the appendix mentioned at the beginning of this chapter.

If the securities are not separable then there exists a distribution P satisfying (6.1) in the definition of separability. Letting this distribution be the prior, a perfect Bayesian equilibrium is simply for traders to report the common consensus value, not allowing any meaningful Bayesian updating and preventing aggregation from occurring. Myopic traders are constrained to report this same value.

6.4   Designing Securities

In this section I discuss the design of informative markets. While separability is a sufficient and necessary condition for aggregation in two natural settings, it only implies the value of the securities reflects all the traders' private information, not that the market designer can use this value to infer that private information or the likelihood of the events of interests. I define informative securities as securities that are both separable and allow for the likelihood of the events of interest to be inferred directly from their value.

As I will show, complete markets are always informative, but deployed prediction markets are rarely complete. These markets require too many securities to be practical, and their securities present challenges for traders. A prediction market for the U.S. presidential election, for example, may need one state per outcome in the electoral college. This is over 2^{50} states and requires traders to bid on securities like "The President wins Ohio, not Florida, Illinois, not Indiana . . . " Even if alternative bidding methods were developed, traders would still be required to review the value of each security for aggregation to be formally implied. This is impractical, and so I consider good designs as those using a few natural securities. I first discuss the design requirements of markets that are always informative, and markets that are informative for a particular signal structure. The latter market allows a single security to be informative on any set of events, but arguably appears "unnatural." To describe the challenges of designing using only natural securities I then consider a constrained design process instead, where the market designer is restricted to an arbitrary subset of (possibly natural) securities.

6.4.1 Informative Markets

Informally, I would like to say that a market's securities are *informative* on a set of events with respect to a signal structure if the market organizer learns the likelihood of the events as if it knew all the traders' private signals. Assuming the values of the securities reflect traders' pooled information, if the likelihood of the events is unambiguously implied from these values then functionally all the private signals are revealed. I call this latter property *distinguishability*.

Definition 11 (Distinguishability). *Let Π be a signal structure over states Ω and $P_{\text{join}(\Pi)}$ be the set of all probability distributions over Ω that assign positive probability only to a subset of states in one element of* $\text{join}(\Pi)$ *(i.e., a trader's possible posteriors after aggregation). A set of securities* $\mathbf{X}$ *on Ω distinguishes a set of events $\mathcal{E}$ with respect to Π if and only if for any $P, P' \in P_{\text{join}(\Pi)}, E_P[\mathbf{X}] = E_{P'}[\mathbf{X}]$ implies $P(E) = P'(E), \forall E \in \mathcal{E}$.*

Equivalently a set of securities distinguishes a set of events if there exists a function from the securities' values to the likelihood of the events. When a set of securities is both separable and distinguishable I will describe it as *informative*.

Definition 12 (Informativeness). *A set of securities* $\mathbf{X}$ *is* informative *on a set of events $\mathcal{E}$ with respect to a signal structure Π if and only if $\mathbf{X}$ both distinguishes $\mathcal{E}$ and is separable with respect to Π.*

Informativeness is a strong condition. Even if securities are not informative it might be possible for a market designer to infer some information from the market, or for the market to be described as partially informative. Generalizing our framework

to account for partial aggregation would be an interesting line of future work.

6.4.2 Always Informative Markets

I first address the problem of designing a set of securities that is informative on a set of events with respect to *any* signal structure. I will call such securities *always informative.* These securities may be of practical interest if the market designer is unsure of the traders' signal structure; using a set of always informative securities implies aggregation will occur no matter what the true signal structure is.

A market is said to be *complete* if by trading securities, agents can freely transfer wealth across states [57]. Rigorously, consider the set of securities that contains a constant payoff security plus all of the securities offered by a market. The market is complete if and only if this set includes $|\Omega|$ linearly independent securities. The most common is a market with $|\Omega|$ Arrow-Debreu securities, each associated with a different state of the world, taking value 1 on that state and 0 everywhere else. For an overview of complete markets, see [33] or [57].

Complete markets are theoretically appealing because they allow traders to express any information about their beliefs. I formalize this well-known idea in our framework in the following proposition.

Proposition 1. *A market over state space Ω with securities $\mathbf{X}$ is complete if and only if for all distinct probability distributions P and P' over Ω, $E_P[\mathbf{X}] \neq E_{P'}[\mathbf{X}]$.*

Proof. Let M be a matrix containing the payoffs of $\mathbf{X}$, with one row for each outcome and one column for each security. The element at row i and column j of M takes value $\mathbf{x}_j(\omega_i)$.

125

Consider a probability distribution P represented as a row vector so, $PM = \mathrm{E}_P[\mathbf{X}]$. The system of linear equations

$$P'M = \mathrm{E}_P[\mathbf{X}] \qquad \sum_{\omega \in \Omega} P'(\omega) = 1$$

has a unique solution $P' = P$ if and only if the matrix M', which is M augmented by a column of 1s to represent the summation constraint, has rank $|\Omega|$.

If the market is complete, M' has this rank so any distinct probability distribution has distinct expectation.

Now assume $\mathrm{E}_P[\mathbf{X}] \neq \mathrm{E}_{P'}[\mathbf{X}], \forall P \neq P'$, and, for a contradiction, that the market is not complete. Then the system of equations has at least two solutions, one of which is the probability distribution P and a distinct solution Q, such that $PM = QM = \mathrm{E}_P[X]$. Let U be the uniform distribution over Ω. Then there exists $c > 0$ such that $(1 - c)U + cQ$ is a probability distribution (since Q satisfies $\sum_{\omega \in \Omega} Q(\omega) = 1$). Moreover, $(1 - c)U + cP$ is also a probability distribution and

$$\big((1 - c)U + cP\big)M = \big((1 - c)U + cQ\big)M,$$

contradicting $\mathrm{E}_P[\mathbf{X}] \neq \mathrm{E}_{P'}[\mathbf{X}], \forall P \neq P'$. Thus, the market must be complete. $\qquad \square$

This expressiveness is a necessary and sufficient condition for the likelihood of every event to be inferred, and suggests an alternative characterization of complete markets as those markets that are always informative on every event.

Theorem 16. *A market is always informative on every event E with respect to every signal structure Π if and only if it is complete.*

Proof. Distinguishing every event E is equivalent to distinguishing each state of the world $\omega \in \Omega$; the latter are also events and so must be distinguished, and if each is

distinguished then the likelihood of any event can be inferred. Proposition 1 shows that completeness is a necessary and sufficient condition for distinguishing each state of the world.

It remains to show that complete markets are also separable with respect to any signal structure Π. Assume, for a contradiction, there exists a signal structure Π and a complete market with securities $\mathbf{X}$ such that $\mathbf{X}$ is non-separable with respect to Π. Since $\mathbf{X}$ is non-separable there must exist distinct probability distributions P and P' over Ω such that $\mathrm{E}_P[\mathbf{X}] = \mathrm{E}_{P'}[\mathbf{X}]$; but by Proposition 1, in a complete market this equality only holds if $P = P'$, a contradiction. So complete markets are separable with respect to any signal structure and always distinguish every event, implying they are always informative on every event. □

Complete markets are often impractical, but rarely is every event of interest. Even if a single event is of interest, however, as many securities as almost half the states in the market may be required to create an always informative market. I let $\bar{E}$ denote the complement of E.

Theorem 17. *Any market that is always informative on an event E must have at least $\min(|E|, |\bar{E}|) - 1$ linearly independent securities.*

Proof. Let $\mathbf{X}$ be a set of securities, fewer than $\min(|E|, |\bar{E}|) - 1$ of which are linearly independent, and assume, for a contradiction, that $\mathbf{X}$ is always informative on E. Restricting attention to states in E, the argument from Proposition 1 implies this market has too few securities to distinguish every probability distribution over E and there exist probability distributions P_E and P'_E such that $E_{P_E}[\mathbf{X}] = E_{P'_E}[\mathbf{X}]$. Let the difference between these distributions be the vector $\Delta_E = P_E - P'_E$, and define vectors

Δ_E^+ and Δ_E^- such that $\Delta_E^+(\omega) = \max(0, \Delta_E(\omega))$ and $\Delta_E^-(\omega) = \min(0, \Delta_E(\omega))$. Since Δ_E is the difference of two probability distributions with the same expected value,

$$\sum_{\omega \in E} \frac{\Delta_E^+(\omega)}{||\Delta_E^+||_1} \mathbf{X}(\omega) = \sum_{\omega \in E} \frac{-\Delta_E^-(\omega)}{||\Delta_E^-||_1} \mathbf{X}(\omega). \qquad (6.2)$$

That is, $\frac{\Delta_E^+(\omega)}{||\Delta_E^+||_1}$ and $\frac{-\Delta_E^-(\omega)}{||\Delta_E^-||_1}$ are disjoint probability distributions over states in E with the same expected value, and the same argument can be made, *mutatis mutandi* for two such probability distributions over states in $\bar{E}$. Let these distributions over E be Q_E and Q_E', and the ones over $\bar{E}$ be $Q_{\bar{E}}$ and $Q_{\bar{E}}'$. Although I have been referring to these as distributions over E and $\bar{E}$ I will also consider them to be distributions over Ω that assign zero probability to all states not previously included in the distributions, and I will use these names to stand for both these distributions and the states they assign positive probability to to reduce notation.

Now suppose there are two traders with signal structure

$$\Pi_1 = \{\{Q_E, Q_{\bar{E}}\}, \{Q_E', Q_{\bar{E}}'\}\}$$

$$\Pi_2 = \{\{Q_E, Q_{\bar{E}}'\}, \{Q_E', Q_{\bar{E}}\}\}$$

and prior
$$P_0 = \frac{Q_E + Q_E' + Q_{\bar{E}} + Q_{\bar{E}}'}{4}.$$

Each trader's expectation conditional on any signal is the same since $\mathrm{E}_{Q_E}[\mathbf{X}] = \mathrm{E}_{Q_E'}[\mathbf{X}]$ and $\mathrm{E}_{Q_{\bar{E}}}[\mathbf{X}] = \mathrm{E}_{Q_{\bar{E}}'}[\mathbf{X}]$ and each signal contains one distribution over states in E and another over states in $\bar{E}$. But the join of the signal structure is $\{\{Q_E\}, \{Q_E'\}, \{Q_{\bar{E}}\}, \{Q_{\bar{E}}'\}\}$, and if $\mathbf{X}$ is separable with respect to Π the expectation conditional on any such element must also, then, be the same. This implies $E_{Q_E}[\mathbf{X}] = E_{Q_{\bar{E}}}[\mathbf{X}]$, but by construc-

tion $Q_E(E) \neq Q_{\bar{E}}(E)$, so if **X** is separable with respect to Π it does not distinguish E, contradicting our assumption that **X** is always informative on E. □

This result demonstrates the need for a market designer to allow traders to express information it finds uninteresting. It also suggests that, in practice, few markets are acquiring all of their participants' information. This is unsurprising, but I think better designs will extract more information, and that this result shows knowledge of or assumptions about the traders' signal structure may be necessary to inform those designs.

6.4.3 Fixed Signal Structures

If the join of the traders' signal structure is known and has singleton sets for its elements, then there exists a single security that is informative on every event.

Theorem 18. *For any signal structure Π such that* $\mathrm{join}(\Pi)$ *consists only of singleton sets there exists a security* **x** *that is informative on every event E with respect to Π.*

The proof uses a result from [65].

Theorem 19 ([65]). *Let Π be a signal structure such that* $\mathrm{join}(\Pi)$ *consists of singleton sets of states, and let x be a security that can be expressed as $x(\omega) = \Sigma_i f(\Pi_i(\omega))$ for an arbitrary function f mapping signals to reals. Then x is separable under Π.*

Proof of Theorem 18. To construct the security, first assign a unique identifier $s_0, s_1, s_2, \ldots$ to every signal of every trader, and define $f(s_j) = 10^j$ for all j. Let S_ω denote the set of indices of the identifiers corresponding to the signals of each trader for state ω, i.e, corresponding to $\Pi_i(\omega)$ for each trader i. The security $x(\omega) = \Sigma_{j \in S_\omega} f(s_j)$ is

separable by Theorem 19. Additionally, the sum $\Sigma_{j \in J} f(s_j)$ for any $J \subset \{0, 1, 2, \ldots\}$ is unique, and each state ω has a unique associated set of signals since I assumed the join consists of singletons. This implies the value of the security for each element of the join is unique, so the security also distinguishes every event. $\qquad\square$

The assumption that the join of traders' signal structure consists only of singleton sets is not without loss of generality. If the signal structure is known, however, the market designer can treat elements of the join as states of the world, identify the correct element of the join by running the market with a single security, then apply the prior to that element to learn the likelihood of each state as if he knew all the traders' private signals. If the prior is unknown this distribution can also be solicited from any single trader using a scoring rule.

6.4.4   Constrained Design

A single security acting as a summary statistic for an entire market is unlikely to be considered natural by any criterion. Real markets, like those on Intrade, use multiple securities. Instead of imposing our own definition of natural, in this section I consider adding a design constraint that the market's securities must be picked from a predefined set. The market designer is then challenged to find the fewest securities from this set that are informative on the events of interest with respect to the given signal structure. I call this the INFORMATIVE SET optimization problem. If the set of predefined securities is empty or has no informative subset then the problem is simply infeasible, so I assume there exists at least one such subset.

Demonstrating INFORMATIVE SET is hard would not be very interesting if exotic

and unnatural securities were required for the proof. One commonly used class of securities are *event securities* which pay \$1 if an event occurs and \$0 otherwise. The corresponding optimization problem is INFORMATIVE EVENT SET, a restriction of INFORMATIVE SET, and even solving this restricted version of the problem is NP-hard.

More formally, INFORMATIVE SET takes as input a finite outcome space Ω, a set of events of interest E, each a subset of Ω, and a set of possible securities X, each of which maps $\Omega \to \mathbb{R}$. The challenge is finding a minimal set of securities from X that is informative for the events of interest E. The INFORMATIVE EVENT SET problem is the same, except the set of possible securities X is restricted to maps $\Omega \to \{0, 1\}$. Now we can state the following results.

Theorem 20. INFORMATIVE EVENT SET *is* NP-*hard.*

This immediately implies that the more general INFORMATIVE SET problem is also hard.

Corollary 1. INFORMATIVE SET *is* NP-*hard.*

The proof appears in the appendix and demonstrates a one-to-one correspondence between set cover instances and a minimal informative set of securities for a single fully informed trader.

The complexity of these problems suggests that while knowledge of the traders' signal structure allows for better designs, a perfect design will be intractable to compute or require additional assumptions about the relationship between traders' signal structure and the set of possible securities. Practically we can only ever hope to offer

better (but not perfect) designs that extract more information from traders than current markets do. These results confirm we will always have to settle for some degree of error in our designs even if the traders' signal structure could be perfectly observed.

6.5   Designing Markets in Review

The formal framework presented here is an early step in our understanding of how to design prediction markets. In hindsight, it may be most interesting for what it does not say, and what gaps it reveals in our understanding of these markets. After all, in practice strict properness seems sufficient for markets to be accurate in practice. This suggests that either (1) a substantial amount of information aggregation and signalling occurs outside the market, and/or (2) even when the markets are not completely formally information, they are *usually partially informative.* That is, in practice, even if we are not extracting *all* the traders information, it appears we are getting a lot of it.

Two suggested areas for future work are extending the formal framework presented here to describe partial information aggregation, and empirically determining whether markets only respond to outside signals or traders aggregate information within the market itself. Reviewers for the conference version of this chapter also agreed that extending it to consider partial information aggregation would be very interesting. Empirically, this is likely what is occurring, and given the empirical success of prediction markets I have also wondered if a random set of securities is, with high probability, likely to aggregate a good deal of information.

Another possibility is that markets simply respond to outside signals. In a pres-

idential election, for example, traders can review the predictions of Nate Silver, and they might set the price in the market equal to the value he predicts. In these cases markets are more about expressing one's confidence in a belief, rather than inferring and exchanging signals. An empirical study that could reveal whether outside signals or internal signals or both were used in markets would be fascinating.

Both questions appear deeply natural and essential to our understanding of prediction markets.

7

Decision Making

The options God gives us are always conditional on our guessing whether a certain proposition is true.

         – Frank P. Ramsey, *Truth and Probability*

This chapter, like the one preceding it, is a self-contained discussion of how predictions and techniques closely related to prediction markets can be used to assist decision making. It also uses only the discrete theory, unlike Chapters 2–5, and again like Chapter 6. Written recently, and being farther from the material presented elsewhere in this thesis, this chapter stands on its own almost completely unchanged. This version references an appendix, available on Yiling Chen's website.[1] I have not included the material from this appendix since I think it bogs down our narrative.

---

[1] http://yiling.seas.harvard.edu/wp-content/uploads/DM_full_version.pdf

7.1   Introduction

Consider a company attempting to decide whether to invest in solar or wind energy. To improve its chances of making the right decision it would like to acquire some expert advice. The company needs, however, some method of incentivizing experts to be accurate. That is, it needs some means of paying experts so that they honestly reveal their private beliefs or information.

In this chapter I characterize two methods of accurately soliciting expert advice for decision making. The first elicits predictions from one or more experts about the likely effects of each available action. If these predictions are accurate then the company can use them to make an informed decision. The second simply asks a single expert to recommend an action. If the expert's incentives are aligned with the decision maker's then this action will profit them both. So with the first method our challenge is incentivizing accuracy, and in the second method it is aligning the expert's most profitable recommendation with the decision maker's most preferred action.

The first part of this Chapter (Sections 6.2–.4) focuses on eliciting predictions for decision making. One popular method of eliciting accurate forecasts of the future is a *prediction market*. In a prediction market, traders or experts produce a series of forecasts about future outcomes of interest. For instance, traders may be asked to predict whether it will rain or not on Friday next week, or which nominated film will win the best picture Oscar. These predictions are probability distributions over the outcomes and are made publicly, allowing experts to review each other's forecasts and update their own predictions accordingly. Eventually the market closes and the

future becomes the present and is observed.

A fundamental property of well-designed prediction markets is that they pay or score predictions for accuracy. If providing an accurate forecast maximizes an expert's score for that prediction, I say the market is *proper*, and if an accurate forecast uniquely maximizes the score the market is *strictly proper*. Strictly proper prediction markets are theoretically effective at aggregating expert information and providing an accurate forecast of the future under some general conditions [65, 24, 51].

Since strictly proper prediction markets are so useful for forecasting the future, we would like to provide the same incentives to experts when eliciting predictions to make a decision. Extending these incentives is not straightforward: the predictions necessary for decision making are different than those made in a classical prediction market, and making a decision changes the observed future. Put another way, in a prediction market an expert predicts *the* future, but a decision maker is interested in the *many possible futures* that can result from its choice. This implies the same techniques that make a prediction market strictly proper do not apply for the elicitation of predictions for decision making. The differences between the classical prediction and decision making settings are detailed in Section 2.

In Section 3 I introduce a model of eliciting predictions for decision making, and in Section 4 I use this model to characterize strictly proper decision making, extending the incentives of strictly proper prediction markets to decision making with both a single expert and many experts in a market (a *decision market*). Unfortunately, creating this incentive requires the decision maker use a completely mixed strategy to choose an action. Essentially, the decision maker must implement an unbiased

estimator of the future, and this requires stochastically sampling from it. This limitation suggests that eliciting predictions for decision making lets a decision maker understand its choice, but requires the decision maker not always act on this understanding. If the decision maker is initially likely to take any action, however, then eliciting predictions may *increase* the likelihood it makes the best available choice.

In Section 5 I discuss an alternative method of acquiring expert advice where the decision maker simply asks a single expert to recommend an action. In this setting I will explicitly model the decision maker's preferences, and our goal is for the expert to accurately reveal the decision maker's most preferred action. I show that we can incentivize an expert to accurately reveal this action if and only if the decision maker's preferences admit a convex weak utility representation, and that this method no longer requires the decision maker choose an action stochastically.

Related Work Decision markets were first proposed by [47] without an analysis of their incentives. [69] showed these proposed decision markets did not provide the same incentive for accuracy as a strictly proper prediction market, and I elaborate on this insight in Section 2. They also described a special case of expert recommendation that I detail and generalize in Section 5.

Other work related to eliciting predictions for decision making has considered external incentives in addition to the market's intrinsic incentives. [85] considered a prediction market where experts can affect the future by taking some actions and defined *principal-aligned* scoring rules that incentivized them to only take "helpful" actions. These rules are spiritually similar to the methods I develop in Section 5, but in this chapter's setting experts cannot take actions to affect the future except by

influencing the decision maker's action through their predictions or recommendations. More recently, Boutilier[14] has discussed decision making with an expert who has its own preferences over the decision maker's actions. Intuitively, if our experts are solar energy lobbyists they may prefer we invest in solar energy even if they must mislead our decision and receive a lower score for doing so. He (Boutilier) introduces *compensation rules* that redress the expert's loss of utility for letting other actions occur to make the expert indifferent again. He also details some realistic complexities of this setting, like the decision maker not precisely knowing the expert's utility function. Different from [14], experts in this chapter's setting do not have preferences over actions.

7.2   Prediction and Decision Markets

In this section I formally compare classical prediction elicitation and eliciting predictions for decision making. This comparison illuminates the new incentive challenges that come with making a decision. I begin by describing the classical setting.

There are many methods of eliciting predictions about the future. One popular method uses a *scoring rule* [16, 43] to evaluate a forecast, and similar rules will be the focus of this chapter. Formally, let Ω be a finite, mutually exclusive and exhaustive set of *outcomes* and $\Delta(\Omega)$ the probability simplex over Ω. A *forecast* or *prediction* is a probability distribution over Ω (an element of $\Delta(\Omega)$), and a *scoring rule* is any

function[2]

$$S : \Delta(\Omega) \times \Omega \to \mathbb{R} \qquad \text{(scoring rule)}$$

that maps a forecast and observed outcome to the reals. Intuitively, a scoring rule compares the forecast to the observed outcome and assigns a real-valued score. For example, we might be curious if it will be sunny or cloudy tomorrow[3]. In this case $\Omega = \{\text{Sunny}, \text{Cloudy}\}$ and a forecast is a binary probability distribution like $(\frac{1}{3} :$ Sunny, $\frac{2}{3}$: Cloudy). If tomorrow is sunny then this forecast's score would be $S((\frac{1}{3} :$ Sunny, $\frac{2}{3}$: Cloudy), Sunny).

Predictions from a single expert and many experts in a market are scored differently. I describe working with a single expert first. A single expert produces one prediction $p \in \Delta(\Omega)$, after which we observe the outcome $\omega \in \Omega$ and score the expert $S(p, \omega)$. If the expert believes a forecast q is the true forecast, then its expected score for a prediction p is

$$S(p, q) = \sum_{\omega \in \Omega} q(\omega) S(p, \omega) \qquad \text{(expected score)}$$

where $q(\omega)$ is the likelihood the belief q assigns to outcome ω. Not every scoring rule is useful. A desirable property is that a risk-neutral expert is incentivized to *accurately* reveal its belief. A scoring rule that provides such incentive is *proper* and satisfies

$$\arg\max_{p \in \Delta(\Omega)} S(p, q) \supseteq \{q\}, \ \forall q \in \Delta(\Omega). \qquad \text{(properness)}$$

---

[2]See Chapter 3 for why I define scoring rules to be real-valued.

[3]I assume these outcomes are mutually exclusive (no sunny cloudy days or cloudy sunny days) and exhaustive (it is either sunny or cloudy).

That is, I treat the arg max function as returning a set of maximizing arguments to the expression, and a scoring rule is proper when the expected score is maximized by accurately reporting the belief q. Even a proper scoring rule may not be useful. Always paying or scoring an expert \$5 is proper, but it provides no real incentive to be accurate. Instead, we are interested in *strictly proper* scoring rules where

$$\arg\max_{p \in \Delta(\Omega)} S(p, q) = \{q\}, \ \forall q \in \Delta(\Omega) \qquad \text{(strict properness)}$$

The expected score of a strictly proper scoring rule is uniquely maximized by accurate reporting. (Strictly) Proper scoring rules have been characterized previously in [43], [59] and [83] with convex functions. I will use the following results in later sections of this chapter.

Theorem 21 ([43]). *A scoring rule is (strictly) proper if and only if*

$$S(p, \omega) = g(p) - g_p^\star \cdot p + g_p^\star(\omega)$$

where $g : \Delta(\Omega) \to \mathbb{R}$ is a (strictly) convex function and $g_p^\star$ is a subgradient of g at the point p.[4]

Corollary 2 ([43]). *Any proper scoring rule*

$$S(p, \omega) = g(p) - g_p^\star \cdot p + g_p^\star(\omega)$$

satisfies

$$\sum_{\omega \in \Omega} p(\omega) S(p, \omega) = g(p), \ \forall p \in \Delta(\Omega)$$

---

[4]A subgradient of a convex function $g : \mathbb{R}^n \to \mathbb{R}$ at a point $p \in \mathbb{R}^n$ is a vector $g_p^\star$ such that $g(p) - g(q) \le g_p^\star \cdot (p - q)$ for all $p, q \in \mathbb{R}^n$.

As mentioned, many experts participating in a market are scored differently from a single expert. A *prediction market* [5] operated using a *market scoring rule* mechanism [48, 49] opens with an initial forecast p_0 and lets experts make a series of forecasts $p_1, p_2, \ldots$. These forecasts are public so experts can review prior predictions and update their own accordingly. Eventually the market closes and an outcome $\omega \in \Omega$ is revealed. Instead of being scored for accuracy, however, each forecast in a market is scored for how much it *improves the accuracy of the preceding forecast*; the expert who produces forecast p_i is scored or paid $S(p_i, \omega) - S(p_{i-1}, \omega)$ for the forecast.[6] An expert may make multiple forecasts in the market and its total score is the sum of the scores for its forecasts.

This method of scoring is useful since it only rewards experts for improving the accuracy of the prior prediction. Further, we can interpret the last prediction made in the market as a current market or consensus expert belief. After all, if an expert disagrees with the current prediction they have an incentive to change it. If the scoring rule S is proper then this method of scoring is also proper for experts since

$$\arg\max_{p \in \Delta(\Omega)} S(p, q) - S(p', q) \;=\; \arg\max_{p \in \Delta(\Omega)} S(p, q), \;\; \forall p, q, p' \in \Delta(\Omega)$$

Intuitively, the score of the previous forecast is fixed and so does not affect the optimization. If the scoring rule S is strictly proper then this method is strictly proper, too. I describe markets using (strictly) proper scoring rules as (strictly) proper markets.

---

[5]Prediction markets can also be operated using continuous double auctions [34, 11], automated market makers [68, 71], and other wagering mechanisms [78, 72, 56]. In this chapter, we are interested in prediction markets that use scoring rules.

[6]It is known that market scoring rules can be equivalently implemented as automated market makers [21, 25]. I restrict my discussion to the former for technical tractability.

Note that even in a (strictly) proper market it may be that an expert still expects to profit by misrepresenting its belief. An expert in a proper prediction market maximizes its score *for a forecast* by being as accurate as possible, and it does not follow that it maximizes its total score by being accurate if it can make more than one prediction in the market. In fact, an expert may find misleading other experts with false predictions to be worthwhile [18], since by leading other experts astray the expert can create an opportunity for a large correction. If experts are acting *myopically* however, then we always expect them to accurately report their beliefs in a strictly proper prediction market.

Forecasts for decision making are different from those in the classical prediction setting just detailed. When making a decision we have a set of actions $\mathcal{A}$ and outcomes Ω. I assume both sets are finite, mutually exclusive and exhaustive. Instead of predicting the unique future, when making a decision, experts are asked to predict *the possible futures resulting from a decision maker's choices.* This prediction can be represented by a $|\mathcal{A}| \times |\Omega|$ *action-outcome* matrix like the one in Figure 7.1, with each row representing a probability distribution over possible outcomes if the associated action is taken. The matrix in Figure 7.1 contains all the information relevant to

<div align="center">

Outcomes Ω

		Profit	Loss
Actions $\mathcal{A}$	Solar	$\frac{2}{3}$	$\frac{1}{3}$
	Wind	$\frac{2}{5}$	$\frac{3}{5}$

</div>

Figure 7.1: An example action-outcome matrix showing an expert's prediction of two possible futures: one resulting from investing in solar energy and the other from investing in wind.

making a decision. In this case, if the decision maker believes the prediction is accurate it will prefer investing in solar energy. These forecasts are elicited in a decision market just like in a classical prediction setting, except after elicitation the decision maker selects an action based on the final market prediction. Only the outcome of this action is then observed. One intuitive way to think of a decision market is as a collection of predictions markets with one prediction market per action. Instead of observing the outcome of each market, however, *we only observe it in one.*

We have not discussed how these forecasts are scored. We'd like to design a means of scoring that offers the same incentives for accuracy as strictly proper prediction markets; that is, we want to incentivize experts to accurately reveal their beliefs. [47], when introducing the idea of decision markets, suggested that forecasts in a decision market could be treated like forecasts in a set of strictly proper prediction markets, one for each action, and the markets for unchosen actions would simply be voided and unscored. This is a natural proposal, but these markets do not incentivize accuracy, as the following example describes.

Let our decision maker still be deciding whether to invest in solar or wind energy. For simplicity I'll assume the outcome space of interest is simply how likely each is to return a profit, $\Omega = \{\text{Profit, Loss}\}$. We'll be running a market, and we let the prior prediction and an expert's belief be as in Figure 7.2. In this example, I further assume that this expert is the last expert in the market and its prediction will be used by the decision maker to select an action.

We can adopt Hanson's proposed scoring scheme using the strictly proper *quadratic*

143

Prior prediction

	Profit	Loss
Solar	$\frac{2}{3}$	$\frac{1}{3}$
Wind	$\frac{2}{5}$	$\frac{3}{5}$

Expert belief

Profit	Loss
$\frac{1}{3}$	$\frac{2}{3}$
$\frac{2}{5}$	$\frac{3}{5}$

Figure 7.2: A hypothetical prior prediction (left) and expert belief (right). The expert can improve the market's prediction for what will occur if solar energy is invested in and agrees with the current prediction for wind energy.

scoring rule

$$S(p, \omega) = 2p(\omega) - \sum_{\omega \in \Omega} p^2(\omega) \qquad \text{(quadratic score)}$$

and assume the decision maker chooses the action most likely to be profitable. Unfortunately, if our expert reports accurately then its expected score is zero: the decision maker will invest in wind energy and the expert did not improve that prediction. Alternatively, the expert can lie and claim wind energy has no chance of becoming profitable. The decision maker will then invest in solar and the expert will expect to score

$$\frac{1}{3}\big(2/3 - (1/3)^2 - (2/3)^2\big) + \frac{2}{3}\big(4/3 - (1/3)^2 - (2/3)^2\big)$$

$$- \frac{1}{3}\big(4/3 - (1/3)^2 - (2/3)^2\big) - \frac{2}{3}\big(2/3 - (1/3)^2 - (2/3)^2\big)$$

$$= 2/9 > 0$$

Thus, misreporting in this decision market is preferred to reporting accurately, and we cannot claim such a market incentivizes accuracy. The intuition of this example

was first noted by [69] for working with a single expert. When an expert is not the last trader in a market, an additional complication is that the expert's prediction is not the final prediction that is used by the decision maker to select an action, although it may affect future predictions.

Experts' ability to affect which of several possible futures is observed is the salient distinguishing feature of a decision market. Eliciting predictions for decision making has the potential to improve our decisions, but without the right incentives are unlikely to be useful. In the next two sections I characterize strictly proper decision making that provides the same incentives as strictly proper prediction markets to experts. Section 7.3 starts by formalizing the decision making and scoring process.

7.3 Eliciting Predictions for Strictly Proper Decision Making

The key distinction between decision making and the classical prediction setting is that in the latter there is one possible future and in the former *experts influence which of multiple possible futures is observed.* To adapt the incentives of a strictly proper prediction market to decision making, then, requires accounting both for how the decision maker chooses an action and how accurate an expert's forecast is. In this section I first formalize my model of eliciting predictions for decision making and selecting an action, then describe how experts are scored and what it means for decision making to be strictly proper. We will see that this definition is different if we are working with a single expert or many experts in a decision market.

7.3.1 Eliciting Predictions and Decision Making

Until now I have been informal with describing how a decision maker uses the predictions it acquires to make a decision. In this subsection I formalize this process. I begin by describing what a forecast for decision making is, how these forecasts are acquired in the single expert and market settings, and then conclude with how a decision maker uses these forecasts to select an action.

Let $\mathcal{A}$ be a finite set of possible actions that a decision maker can take and Ω a finite set of mutually exclusive and exhaustive outcomes of interest to the decision maker. In our running energy investment example $\mathcal{A} = \{\text{Solar, Wind}\}$ and $\Omega = \{\text{Profit, Loss}\}$.

Experts are risk-neutral, rational agents and have private beliefs representable by an $|\mathcal{A}| \times |\Omega|$ *action-outcome* matrix associating actions with distributions over the outcomes.[7] Examples of these matrices appear in Figures 1 and 2. Each row of an expert's *action-outcome* matrix is a probability distribution over outcomes and represents the expert's subjective belief on likely outcomes when the row's action is taken. I denote the set of *action-outcome* matrices as $\mathcal{P}$. Experts are asked to produce forecasts or predictions, which are also *action-outcome* matrices in $\mathcal{P}$, but may not be the same as their beliefs.

I consider eliciting predictions from both a single expert and many experts in a market. When working with a single expert, that expert makes a single prediction $P \in \mathcal{P}$. The decision maker then applies a *decision rule* to this forecast to construct a *decision strategy*—a probability distribution over the available actions.

---

[7]Some prior work considers a setting where experts can incur cost to improve their beliefs and studies how to induce an appropriate degree of learning as well as accurate predictions [64]. I do not consider cost of obtaining additional information and assume that experts are endowed with their beliefs.

Definition 13 (Decision Rule)**.** *A decision rule is any function*

$$R : \mathcal{P} \rightarrow \Delta(\mathcal{A}),$$

mapping predictions in $\mathcal{P}$ to decision strategies from $\Delta(\mathcal{A})$. Let $R(P, a)$ be the probability the decision strategy $R(P)$ assigns to taking action a, and say a decision rule has full support if $R(P, a) > 0$ for all $P \in \mathcal{P}$ and $a \in \mathcal{A}$.

Once the decision maker has its strategy it selects an action according to it and then an outcome $\omega \in \Omega$ is observed. Intuitively, this outcome is the result of the action taken.

Multiple experts in a decision market are treated differently. A decision market opens with an initial prediction P_0 and lets experts make a series of public predictions $P_1, P_2, \ldots$. This is similar to how prediction markets operate, but with matrix forecasts instead of vectors. Eventually the market closes with a final prediction P and the decision maker applies its decision rule to this prediction to construct its decision strategy. I make no assumption on market dynamics or how this final prediction is formed.

I further assume that experts know the decision rule used by the decision maker prior to making their predictions. In Section 7.4, I will show that this assumption can be relaxed and my results hold as long as experts know that the decision maker will use a decision rule with full support.

7.3.2 Scoring Predictions

In classical prediction elicitation, forecasts are scored using a scoring rule, a function

$$S : \Delta(\Omega) \times \Omega \to \mathbb{R}. \hspace{3cm} \text{(scoring rule)}$$

In eliciting predictions for decision making, a decision maker uses the following generalization of a scoring rule instead.

Definition 14 (Decision Scoring Rule). *A decision scoring rule is a function*

$$S : \Delta(\mathcal{A}) \times \mathcal{A} \times \mathcal{P} \times \Omega \to \mathbb{R} \hspace{2cm} \text{(decision scoring rule)}$$

mapping a decision strategy, an action taken, a forecast and an observed outcome to a real number.

A decision scoring rule lets us account for how the decision maker selects its action as well as how accurate the expert's forecast is. In the next section, we'll see that this generalization is essential for strictly proper decision making. Throughout the chapter, I assume that experts know the decision scoring rule used prior to making their predictions.

When working with a single expert, the decision maker pays the expert who provides forecast P a score $S(R(P), a, P, \omega)$, when action a, drawn according to the decision strategy $R(P)$, is taken and outcome ω is observed. The expected score of an expert who believes Q and predicts P is

$$\sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a) Q(a, \omega) S(R(P), a, P, \omega). \hspace{1.5cm} \text{(expected score)}$$

Unpacking the above, each term is the likelihood an action and an outcome jointly occur, $R(P, a)Q(a, \omega)$, times the value to the expert for that outcome occurring $S(R(P), a, P, \omega)$.

In a decision market, like in a prediction market, experts receive a *net score* that is the difference of their and the previous predictions' scores. The net score for prediction P_i is $S(R(P), a, P_i, \omega) - S(R(P), a, P_{i-1}, \omega)$, when the final prediction is P, the decision maker takes action a according to decision strategy $R(P)$, and outcome ω is observed. The expected net score of an expert in a decision market who believes Q and predicts P_i, with final prediction P, is

$$\sum_{a \in \mathcal{A}, \omega \in \Omega} R(P)Q(a, \omega)\big(S(R(P), a, P_i, \omega) - S(R(P), a, P_{i-1}, \omega)\big) \quad \text{(expected net score)}$$

Note that, unlike the single expert setting, there is a separation between the prediction the decision maker creates a decision strategy from P, and an expert's prediction P_i.

7.3.3 Incentives and Strict Properness

In this subsection I define strictly proper decision making. Unlike the classical prediction setting, I will use three definitions of strict properness. One for working with a single expert, one for running a market, and one that works for both settings. Also unlike a market, strict properness is not just a function of the scoring rule or even the decision scoring rule. Instead, an expert's incentives will depend on both the decision rule and the decision scoring rule used. As a result I will describe (R, S) pairs as either strictly proper *for an expert, for a market*, or simply as strictly proper if they work for both. This is a whirlwind of specialized terms, but by going through each one, their necessity should become clear. Also, shortly after introducing pairs

that are strictly proper *for a market*, we will see that we can safely forget about them to focus on the other two. Each version of strict properness, however, brings the same expert incentives as strict properness for the classical prediction setting to the relevant decision making setting.

To begin, I say a decision rule and a decision scoring rule pair is strictly proper for an expert when a single expert uniquely maximizes its expected score for a prediction by revealing its beliefs. Thus, exactly as with a strictly proper scoring rule, truthful revelation is strictly optimal for a single expert facing such a pair.

Definition 15 (Strictly Proper for an Expert). *A decision rule and decision scoring rule pair* (R, S) *is strictly proper* for an expert *when*

$$\{Q\} = \arg\max_{P \in \mathcal{P}} \sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a) Q(a, \omega) S(R(P), a, P, \omega), \ \forall Q \in \mathcal{P}$$

Strict properness *for a market* is defined very differently, and in the next we'll see these differences are meaningful.

Definition 16 (Strictly Proper for a Market). *A decision rule and decision scoring rule pair* (R, S) *is strictly proper* for a market *when*

$$\sum_{a \in \mathcal{A}, \omega \in \Omega} R(P) Q(a, \omega) \big(S(R(P), a, Q, \omega) - S(R(P), a, P_{i-1}, \omega) \big)$$

$$\geq \sum_{a \in \mathcal{A}, \omega \in \Omega} R(P') Q(a, \omega) \big(S(R(P'), a, P_i, \omega) - S(R(P'), a, P_{i-1}, \omega) \big)$$

for all $Q, P_{i-1}, P_i, P, P' \in \mathcal{P}$, *with the inequality strict if* $P_i \neq Q$.

Understanding this definition and how it is different from the prior strict properness *for an expert* is useful. The expected score of an expert in a market is most

notably different because the decision rule may not be applied to the forecast being scored. Instead it is applied to the final forecast made, and for strict properness I require an expert always expects to strictly maximize its net score by revealing its beliefs *no matter what the decision strategy is.*

Intuitively, it is possible that an expect can *change the final prediction* to affect the decision strategy. Since I make no assumptions on the market dynamics and how the final prediction is formed, when an expert changes its prediction from Q to P_i, the final prediction may change from P to P'. What I am ruling out with the above definition is that an expert might have an incentive to change the final prediction by predicting against its belief.

I note that the focus of this chapter is not on analyzing whether and when a decision market aggregates all private information and produces a consensus prediction with rational participants. Instead, I aim to understand when a decision market provides incentives for any myopic expert to predict its belief if the expert only cares about its expected payoff of the current prediction, a property that strictly proper prediction markets have but Hanson's decision markets lack. While strict properness for a market does not allow one to immediately conclude that the final market prediction aggregates all information of rational participants in a decision market, such incentive is necessary for information aggregation — without it, as shown by the example in Section 7.2, the last participant of the market may manipulate the market prediction — and hence is fundamental to understand. In Section 7.6, I will discuss the implication of strict properness for a market on information aggregation in decision markets with forward-looking rational agents.

Carrying around two definitions of strict properness is cumbersome. We'd like to combine them into one, and we can almost accomplish this with the following.

Definition 17 (Strictly Proper Pair). *A decision rule and decision scoring rule pair* (R, S) *is strictly proper when a prediction's expected score is independent of the decision strategy*

$$\sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a) Q(a, \omega) S(R(P), a, P_i, \omega)$$

$$= \sum_{a \in \mathcal{A}, \omega \in \Omega} R(P', a) Q(a, \omega) S(R(P'), a, P_i, \omega), \ \forall Q, P_i, P, P' \in \mathcal{P} \qquad (7.1)$$

and uniquely maximized when an expert predicts its belief

$$\{Q\} = \arg\max_{P_i \in \mathcal{P}} \sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a) Q(a, \omega) S(R(P), a, P_i, \omega), \ \forall Q, P, P_i \in \mathcal{P}. \qquad (7.2)$$

Intuitively, this notion of strict properness makes decision making resemble the classical prediction setting. Like in that setting I require that future predictions cannot affect the score of prior predictions, and I demand an expert uniquely maximize its score for a prediction by revealing its beliefs. Also, this definition nearly combines the previous two, and every strictly proper pair (R, S) is strictly proper for both an expert and a market, as the following proposition formalizes.

Proposition 2. *Every strictly proper pair* (R, S) *is strictly proper for both an expert and a market.*

Proof. Let (R, S) be a strictly proper pair. For any $P \neq Q$, We have

$$\sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a)Q(a, \omega)S(R(P), a, P, \omega)$$

$$< \sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a)Q(a, \omega)S(R(P), a, Q, \omega)$$

$$= \sum_{a \in \mathcal{A}, \omega \in \Omega} R(Q, a)Q(a, \omega)S(R(Q), a, Q, \omega)$$

The inequality following from Equation 7.2 and the equality from Equation 7.1. This implies (R, S) is strictly proper *for an expert*.

Strict properness *for a market* requires

$$\sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a)Q(a, \omega)\big(S(R(P), a, Q, \omega) - S(R(P), a, P_{i-1}, \omega)\big)$$

$$\geq \sum_{a \in \mathcal{A}, \omega \in \Omega} R(P', a)Q(a, \omega)\big(S(R(P'), a, P_i, \omega) - S(R(P'), a, P_{i-1}, \omega),$$

for all $Q, P, P', P_{i-1}, P_i \in \mathcal{P}$, with the inequality strict if $P \neq Q$.

From the definition of strictly proper pairs, we have

$$\sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a)Q(a, \omega)\big(S(R(P), a, Q, \omega) - S(R(P), a, P_{i-1}, \omega)\big)$$

$$- R(P', a)Q(a, \omega)\big(S(R(P'), a, P_i, \omega) - S(R(P'), a, P_{i-1}, \omega)\big)$$

$$= \sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a)Q(a, \omega)S(R(P), a, Q, \omega) - R(P, a)Q(a, \omega)S(R(P), a, P_i, \omega) \geq 0$$

for all $Q, P, P', P_{i-1}, P_i \in \mathcal{P}$. The equality follows from Equation 7.1 and the inequality from Equation 7.2, and this inequality is strict if $P_i \neq Q$. Thus (R, S) is strictly proper *for a market*, too. $\square$

In fact, we can go further and say this definition of strict properness defines *most* of strictly proper decision making. For any pair that is *strictly proper for a market*

there is a strictly proper pair that uses the same decision rule and a decision scoring rule that provides experts the same expected net scores as before.

Proposition 3. *For every pair* $(R, \bar{S})$ *that is strictly proper* for a market, *there exists a strictly proper pair* (R, S) *such that every prediction has the same expected net score*

$$\sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a) Q(a, \omega) \big(\bar{S}(R(P), a, P_i, \omega) - \bar{S}(R(P), a, P_{i-1}, \omega) \big)$$

$$= \sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a) Q(a, \omega) \big(S(R(P), a, P_i, \omega) - S(R(P), a, P_{i-1}, \omega) \big)$$

for all $P, Q, P_{i-1}, P_i \in \mathcal{P}$.

The proof appears in the appendix.

For all practical purposes, then, we no longer need to consider pairs that are strictly proper *for a market*. A similar proposition cannot be shown for pairs that are strictly proper *for an expert*. In the next section I show that strictly proper pairs always have decision rules with full support, but some pairs that are strictly proper *for an expert* do not. These pairs do, however, create decision strategies with full support *for almost all predictions*. Hence this distinction is unlikely to be important in practice. I thus say strictly proper pairs describe *most* of strictly proper decision making.

7.4 Strictly Proper Decision Making

In this section I characterize strictly proper decision making with both many experts in a decision market and a single expert. I show that any decision rule with full support is part of a strictly proper pair, and it is easy to construct such

pairs using a strictly proper scoring rule. Unfortunately, a fundamental limitation of this approach to decision making is that it requires the decision maker always use a completely mixed strategy to select an action when running a market, and most of the time when working with a single expert. This suggests that eliciting predictions for decision making can *improve* the likelihood a decision maker takes a preferred action, but cannot *guarantee* it does so.

7.4.1   Strictly Proper Decision Markets

In this subsection I characterize strictly proper decision markets. Following our discussion in the previous section, instead of working with pairs that are strictly proper *for a market*, I restrict my attention in this subsection to pairs that are simply *strictly proper*. I start by showing that if and only if a decision rule has full support is it part of a strictly proper pair, and provide an easy means of constructing such a pair given a strictly proper scoring rule. I conclude this section with the detailed characterization of these pairs.

I begin by showing a decision rule must have full support to be part of a strictly proper pair.

Theorem 22 (Full Support is Necessary for a Strictly Proper Pair)**.** *If a pair (R, S) is strictly proper then R has full support.*

Proof. Assume, for a contradiction, that R is a decision rule without full support and S is a decision scoring rule such that (R, S) is strictly proper. Let P^* be a prediction such that $R(P^*, a') = 0$ for some action a', which must exist by my assumption that R does not have full support, and let Q and Q' be two action-outcome matrices differing

only on action a'. Then we have

$$\sum_{a \in \mathcal{A}, \omega \in \Omega} R(P^*, a) Q(a, \omega) \big(S(R(P^*), a, P_i, \omega) - S(R(P^*), a, P_{i-1}, \omega) \big)$$

$$= \sum_{a \in \mathcal{A}, \omega \in \Omega} R(P^*, a) Q'(a, \omega) \big(S(R(P^*), a, P_i, \omega) - S(R(P^*), a, P_{i-1}, \omega) \big)$$

for all $P_i, P_{i-1} \in \mathcal{P}$. This implies the same prediction maximizes the expected score of an expert who believes Q and an expert who believes Q', yet since this prediction cannot be both Q and Q' the pair (R, S) violates Equation (7.2) and so must not be strictly proper, contradicting our assumption. $\qquad \square$

Simply put, experts have no incentive to be accurate on actions that are never tested, so a decision rule without full support cannot be strictly proper. This intuition is the same one mentioned by Othman and Sandholm [69], who showed that any deterministic decision rule cannot be part of a pair that is strictly proper *for an expert*.

On the other hand, we can constructively demonstrate that any decision rule with full support is part of a strictly proper pair. Given a decision rule R with full support and any strictly proper scoring rule $\bar{S}$, we can create a decision scoring rule

$$S(R(P), a, P_i, \omega) = \frac{1}{R(P, a)} \bar{S}(P_i(a), \omega), \tag{7.3}$$

and the pair (R, S) is strictly proper since the expected score for a prediction P_i given beliefs Q and decision strategy $R(P)$ is then

$$\sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a) Q(a, \omega) \Big(\frac{1}{R(P, a)} \bar{S}(P_i(a), \omega) \Big) = \sum_{a \in \mathcal{A}, \omega \in \Omega} Q(a, \omega) \bar{S}(P_i(a), \omega)$$

the same expected score as if an expert were participating in $|\mathcal{A}|$ independent and strictly proper prediction markets, one for each action. Intuitively, dividing the scor-

ing rule's score by the likelihood that the decision maker takes an action unbiases how the score is sampled. The following complete characterization shows that all strictly proper pairs are constructed using a similar intuition.

Some additional notation is needed before stating the theorem. I use a colon between two matrices to denote their Frobenius inner product, $A : B = \sum_{i,j} A(i,j)B(i,j)$, and let $G_P^\star$ be a subgradient of the convex function $G : \mathcal{P} \to \mathbb{R}$ at P. This subgradient is also a matrix with the same dimensions as matrices in $\mathcal{P}$.

Theorem 23 (Strictly Proper Pair Characterization). *A pair (R, S) is strictly proper if and only if R has full support and there exists a subdifferentiable strictly convex function G such that*

$$S(R(P), a, P_i, \omega) = G(P_i) - G_{P_i}^\star : P_i + \frac{G_P^\star(a, \omega)}{R(P, a)} \tag{7.4}$$

Proof. I begin by showing that given a decision rule R with full support and a strictly convex G, defining a decision scoring rule S as in Equation 7.4 makes (R, S) a strictly proper pair.

An expert's expected score for predicting P with beliefs Q and decision policy $R(P)$ is

$$\sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a)Q(a, \omega)S(R(P), a, P_i, \omega)$$

$$= \sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a)Q(a, \omega)\left(G(P_i) - G_{P_i}^\star : P_i + \frac{G_{P_i}^\star(a, \omega)}{R(P, a)}\right)$$

$$= \sum_{a \in \mathcal{A}, \omega \in \Omega} \{R(P, a)Q(a, \omega)\left(G(P_i) - G_{P_i}^\star : P_i\right)\} + Q : G_{P_i}^\star$$

$$= G(P_i) - G_{P_i}^\star : P_i + Q : G_{P_i}^\star \qquad \text{(since } \sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a)Q(a, \omega) = 1\text{)}$$

$$= G(P_i) + (Q - P_i) : G_{P_i}^\star$$

which is independent of the decision strategy, and the expert's expected score for accurately predicting Q is then

$$G(Q) + (Q - Q) : G_Q^\star = G(Q)$$

and applying the subgradient inequality we have

$$G(Q) > G(P_i) + (Q - P_i) : g_{P_i}^\star, \ \forall P_i \neq Q \in \mathcal{P}$$

implying (R, S) is a strictly proper pair.

Now I show that given a strictly proper pair (R, S) it is necessary that R have full support and there exists a strictly convex G such that S is as defined in Equation 7.4. Since Theorem 22 proved the necessity of R having full support, we only need prove the latter condition.

As a shorthand, I define an expected score function

$$V(R(P), Q, P_i) = \sum_{a \in \mathcal{A}, \omega \in \Omega} R(P, a) Q(a, \omega) S(R(P), a, P_i, \omega)$$

and recall from Definition 17 that

$$V(R(P), Q, P_i) = V(R(P'), Q, P_i), \forall P, P', P_i, Q \in \mathcal{P}$$

allowing us to write simply $V(Q, P_i)$; our strictly convex function G will be $G(P_i) = V(P_i, P_i)$, which is convex (and I will verify is strictly convex shortly), and we'll use

$$G_{P_i}^\star(a, \omega) = R(P) S(R(P), a, P, \omega)$$

for any $P \in \mathcal{P}$ as our subgradient at P_i. We verify it is a subgradient by checking the

subgradient inequality:

$$G(P_i) + (Q - P_i) : G^\star_{P_i}$$

$$= V(P_i, P_i) + \sum_{a \in \mathcal{A}, \omega \in \Omega} (Q(a, \omega) - P_i(a, \omega))R(P, a)S(R(P), a, P_i, \omega)$$

$$= V(P_i, P_i) + V(Q, P_i) - V(P_i, P_i)$$

$$= V(Q, P_i)$$

$$< V(Q, Q)$$

for all $P_i \neq Q \in \mathcal{P}$. The strict inequality following since (R, S) is a strictly proper pair and this strict inequality implies G is strictly convex [50].

Before concluding, I note that since (R, S) is a strictly proper pair

$$R(P, a)S(R(P), a, P_i, \omega) = R(P')S(R(P)', a, P_i, \omega), \ \forall P, P', P_i \in \mathcal{P}, a \in \mathcal{A}, \omega \in \Omega$$

(otherwise there exist beliefs Q such that $V(R(P), Q, P_i) \neq V(R(P'), Q, P_i)$), and I use this fact to verify that G with subgradients as given is, in fact, equal to S

$$G(P_i) - G^\star_{P_i} : P_i + \frac{G^\star_P(a, \omega)}{R(P, a)}$$

$$= V(P_i, P_i) - \sum_{a \in \mathcal{A}, \omega \in \Omega} \{R(P, a)P_i(a, \omega)S(R(P), a, P_i, \omega)\} + \frac{R(P, a)S(R(P), a, P_i, \omega)}{R(P, a)}$$

$$= V(P_i, P_i) - V(P_i, P_i) + S(R(P), a, P_i, \omega)$$

$$= S(R(P), a, P, \omega)$$

So from any strictly proper pair we can construct a strictly convex G satisfying Equation 7.4. $\qquad \Box$

Theorem 23 shows that while a decision maker can take a preferred action with probability arbitrarily close to one, it must commit to a completely mixed decision

strategy. In short, the decision maker must implement an unbiased estimator of the future, and this requires stochastically sampling the actions. Note, however, that it is sufficient for experts to believe they will be scored in a strictly proper fashion, and the decision maker does not have to *ex ante* design its decision rule. Instead, it can simply review the final prediction, construct any decision strategy with full support, and then score the experts using an appropriate decision scoring rule to create a strictly proper decision market. This insight is spiritually analogous to the observation made by [14] on using compensations rules for prediction elicitation when an expert has preferences over actions. [14] noted that the expert does not need to know the decision rule to be strictly incentivized to predict its belief.

A fun analogy to the decision maker in a strictly proper decision market is to an overwhelmed teaching assistant grading a midterm. The teaching assistant does not have the time to grade every question and instead must pick one from each test. If some questions are more likely to be graded than others then students will spend more time on those and neglect the rest, biasing their scores. Only by (1) possibly grading any question and (2) weighting that question's score by the inverse likelihood that the question is graded will the teaching assistant create an unbiased estimator, where the student's expected grade is the same as if every question were reviewed. This encourages students to pay equal attention to each question and not "game the system."

7.4.2 Strictly Proper Decision Making with a Single Expert

Working with a single expert is different than running a decision market since the expert knows the decision maker will apply the decision rule to its prediction. In a decision market, on the other hand, the decision rule is applied to the final prediction. This distinction allows pairs that are strictly proper *for an expert* to use decision rules that do not have full support, although we can formally demonstrate that it is rare for these rules to create decision strategies that are not completely mixed.

Theorem 24. *For any pair (R, S) that is strictly proper* for an expert, *define a set $\mathcal{P}_0 \subset \mathcal{P}$ as the forecasts that R maps to decision strategies that are not completely mixed. The set $\mathcal{P}_0$ is nowhere dense in $\mathcal{P}$ with its natural Euclidean topology.*[8]

Intuitively, this means that for any forecast that the decision rule maps to a not completely mixed decision strategy, there is another arbitrarily close forecast that does map to a completely mixed strategy. I think it is unlikely this ability to avoid some actions will be useful in practice.

I conclude this subsection with a complete characterization of strictly proper *for an expert* pairs. The statement and its proof are similar to those of Theorem 23.

Theorem 25 (Strictly Proper for an Expert Characterization)**.** *A pair (R, S) is strictly proper* for an expert *if and only if there exists a subdifferentiable strictly convex function G and subgradients such that $G_P^\star(a) = \vec{0}$ whenever $R(P, a) = 0$ and*

$$S(R(P), a, P, \omega) = G(P) - G_P^\star : P + \frac{G_P^\star(a, \omega)}{R(P, a)}, \ \forall R(P, a) > 0. \tag{7.5}$$

---

[8]A set is nowhere dense in a topological space if the interior of its closure, with respect to the topological space, is empty.

This concludes our discussion of strictly proper decision making, where a decision maker solicits a complete mapping from actions to outcomes. In the next section I discuss an alternative where, instead of this mapping, a decision strategy or recommendation is directly solicited. This alternative allows the decision maker to deterministically take a preferred action, instead of doing so with high probability.

7.5   Recommendations for Decision Making

The previous section demonstrated that strictly proper decision making (almost always) requires the decision maker use a completely mixed strategy to select an action. Put another way, even if the decision maker learns some actions are undesirable it must risk taking them. This is certainly not ideal and possibly non-credible for the decision maker.

In this section I describe an alternative method of using expert advice to make a decision. Instead of asking experts to predict the likely outcome of each action, I instead simply ask a single expert to recommend an action. This allows the decision maker *to always take its most preferred action.*

When deciding to invest in wind or solar energy in our running example, the decision maker can run a strictly proper decision market and ask experts to predict the likely outcome of each investment. This can increase the decision maker's chances of making the right investment, but with some positive chance it must take the "wrong" or less preferred action simply to test the experts' accuracy. A simple and useful alternative is to offer a single expert a percentage of the realized profit, and ask them to suggest an action. This expert is no longer interested in making an accurate

prediction; instead its incentives are perfectly aligned with the decision maker's to produce a "good" or preferred outcome. This alignment of expert's and decision maker's incentives will let the decision maker deterministically act on the expert's recommendation. Formalizing this model of decision making is the topic of this section. I stress that if the expert recommends a single action, the *decision maker can deterministically take it,* in contrast to the previous result. This is a great benefit of asking for a recommendation. In fact, we'll see that an expert can *always recommend a single action* since a decision maker will have one action it (weakly) prefers more than the others.

This approach, like eliciting predictions for decision making, also has its limitations. It only lets us solicit a recommendation from a single expert, and eliciting an accurate recommendation is possible if and only if the decision maker's preferences admit a subdifferentiable convex weak utility representation. Still, I think it is an especially interesting option since it uses ideas from scoring rules without asking for a prediction. Instead – intuitively – a scoring rule is used to rank the actions so the expert is incentivized to choose the one the decision maker prefers most—aligning the expert's and the decision maker's preferences.

7.5.1   A Model for Recommendations

When working with expert recommendations I consider a single expert reporting a decision strategy $\sigma \in \Delta(\mathcal{A})$ and a prediction $p \in \Delta(\Omega)$ of what is likely to occur if that strategy is adopted. The decision maker then draws an action according to the strategy, observes the outcome ω, and scores the expert using a scoring rule $S(p, \omega)$.

Our goal is not to elicit any decision strategy, however, but a decision strategy that, if followed, results in the most preferred possible distribution p.

Formalizing this statement requires specifying the decision maker's preferences. I let the decision maker's preferences be a binary relation $\preceq$ on $\Delta(\Omega)$, the probability distributions or lotteries over Ω. The decision maker weakly prefers p_1 to p_0 if and only if $p_0 \preceq p_1$ and strictly prefers p_1 to p_0 if and only if $p_0 \prec p_1$. These ordinal preferences admit a *weak utility representation* [75, 5] if there exists a function $u :$ $\Delta(\Omega) \to \mathbb{R}$ such that if $p_0 \prec p_1$ then $u(p_0) < u(p_1)$ for all $p_0, p_1 \in \Delta(\Omega)$. Further, I say these preferences admit a *(strictly) convex weak utility representation* if there exists a (strictly) convex function u that is also a weak utility representation. I note that an expected value maximizing decision maker always has preferences that admit a convex weak utility representation. I continue to assume the expert is a risk-neutral expected value maximizer. While this assumption may not always hold, it is arguably reasonable for settings where the reward that the expert can receive is relatively small.

The decision maker's goal is to elicit the decision strategy that if followed results in its most preferred distribution. If the expert has belief Q, then the decision maker wants to find a decision strategy $\sigma^* \in \Delta(\mathcal{A})$ – a column vector – such that

$$Q^T \cdot \sigma^* \succeq Q^T \cdot \sigma, \ \forall \sigma \in \Delta(\mathcal{A}) \qquad \text{(preferred strategies)}$$

where Q^T is the transpose of Q and hence $Q^T \cdot \sigma$ is the lottery over outcomes created by selecting decision strategy σ. I let Σ_Q^* denote the set of such preferred strategies σ^* and

$$\Phi_Q = \arg\max_{\sigma \in \Delta(\mathcal{A})} \ \sup_{p \in \Delta(\Omega)} \ \sum_{a \in \mathcal{A}, \omega \in \Omega} \sigma(a) Q(a, \omega) S(p, \omega)$$

denote the set of decision strategies that maximize the expert's expected score. I say a scoring rule is *a recommendation rule for preferences* $\preceq$ if it always incentivizes the expert to reveal a strategy in Σ_Q^*.

Definition 18 (Recommendation Rule). *A scoring rule S is a recommendation rule for preferences $\preceq$ over $\Delta(\Omega)$ when $\Phi_Q \subseteq \Sigma_Q^*$ for all $Q \in \mathcal{P}$.*

Intuitively, a recommendation rule translates the decision maker's preferences into a payoff function (scoring rule) for the expert that incentivizes it to reveal the decision maker's most preferred strategy.

To recap, in our recommendation setting there is a decision maker and a single expert. The decision maker shows the expert a scoring rule, and the expert reports a decision strategy and makes a prediction about the outcome of this strategy. The decision maker acts according to the strategy, observes the outcome, and pays the expert based on its prediction and the observed outcome using the scoring rule. If the scoring rule is a recommendation rule then the expert has an incentive to reveal the strategy the decision maker would most prefer taking if it had the same information the expert did. I note that the expert does not need to know the decision maker's preferences.

7.5.2 Characterizing Recommendation Rules

In this subsection I describe the preferences for which we can construct a recommendation rule where an expert maximizes its expected score by reporting the decision maker's most preferred decision strategy. That is, I describe the preferences for which we can strictly incentivize the expert to reveal strategies in Σ_Q^*. It turns out

this is precisely the set of of preferences admitting a subdifferentiable convex weak utility representation.

I first show that if we know a subdifferentiable convex function[9] that is a weak utility representation of the decision maker's preferences, we can use it to construct a recommendation rule.

Proposition 4 (Recommendation Rule Construction). *If a sub differentiable convex function $G : \mathbb{R}^{|\Omega|} \to \mathbb{R}$ is a weak utility representation of the decision maker's preferences $\preceq$, then the scoring rule*

$$S(p, \omega) = G(p) - G_p^\star \cdot p + G_p^\star(\omega)$$

is a recommendation rule for its preferences.

Proof. G is a weak utility representation of $\preceq$ means that $p_0 \prec p_1$ implies $G(p_0) < G(p_1)$. By Theorem 21 and Corollary 2, the scoring rule

$$S(p, \omega) = G(p) - G_p^\star \cdot p + G_p^\star(\omega)$$

is proper (in the classical sense) with expected score function

$$G(p) = \sum_{\omega \in \Omega} p(\omega) S(p, \omega).$$

An expert with belief Q maximizes its expected score by solving

$$\max_{\sigma \in \Delta(\mathcal{A}), p \in \Delta(\Omega)} \sum_{\omega \in \Omega} (Q^T \cdot \sigma)(\omega) S(p, \omega)$$

---

[9]A convex function $G : \mathbb{R}^n \to \mathbb{R}$ is subdifferentiable everywhere in its relative interior. I am requiring, for notational simplicity, it also be subdifferentiable at its relative boundary.

and since S is proper, given any σ

$$\sum_{\omega \in \Omega}(Q^T \cdot \sigma)(\omega)S(Q^T \cdot \sigma, \omega) \geq \sum_{\omega \in \Omega}(Q^T \cdot \sigma)(\omega)S(p, \omega)$$

for all Q and p. By our construction,

$$\sum_{\omega \in \Omega}(Q^T \cdot \sigma^*)(\omega)S(Q^T \cdot \sigma^*, \omega) \geq \sum_{\omega \in \Omega}(Q^T \cdot \sigma)(\omega)S(p, \omega)$$

for some $\sigma^* \in \Sigma_Q^*$, with the inequality strict if $\sigma \notin \Sigma_Q^*$, since

$$Q^T \cdot \sigma^* \succ Q^T \cdot \sigma, \ \forall \sigma \notin \Sigma_Q^*$$

and

$$G(Q^T \cdot \sigma^*) > G(Q^T \cdot \sigma), \ \forall \sigma \notin \Sigma_Q^*$$

which then, by Corollary 2, gives the desired inequality. Thus, S is a recommendation rule. $\qquad \square$

Proposition 4 indicates that the decision maker's preferences admitting a subdifferentiable convex utility representation is a sufficient condition for the existence of a recommendation rule for the preferences. In fact, it is also a necessary condition. Theorem 26 gives the complete characterization.

Theorem 26 (Recommendation Rule Characterization)**.** *If the decision maker is considering at least two actions, there exists a recommendation rule S for its preferences $\preceq$ if and only if these preferences admit a subdifferentiable convex weak utility representation.*

Proof. Proposition 4 proves that if the decision maker's preferences admit a subdifferentiable convex weak utility representation there exists a recommendation rule for them. Here I only prove the necessity of this condition.

Assume, for a contradiction, that the preferences $\preceq$ do not admit a subdifferentiable convex weak utility representation but there is a recommendation rule S for them. Let there be an expert with belief Q such that $Q(a) = q_1$ and $Q(a') = q_2$ for all $a' \neq a$. Assume the expert recommends a single action, then its expected score function given that action is

$$V(q) = \sup_{p \in \Delta(\Omega)} \sum_{\omega \in \Omega} q(\omega) S(p, \omega)$$

which is a subdifferentiable convex function of the lotteries. Since I assumed that $\preceq$ does not admit a subdifferentiable convex weak utility representation, this implies that there exists q_1 and q_2 such that

$$V(q_1) \ \geq \ V(q_2), \text{ and}$$

$$q_1 \ \prec \ q_2.$$

That is, the expert expects a (weakly) higher score by recommending a less preferred action a. Further, the expert expects to score (weakly) higher by recommending action a than any convex combination of actions because

$$V(q_1) \geq \alpha V(q_1) + (1 - \alpha) V(q_2) \geq V(\alpha q_1 + (1 - \alpha) q_2),$$

where the second inequality is due to the convexity of V. Thus, the decision strategy of taking action a with probability 1 is an element in Φ_Q but not in Σ_Q^*. This contradicts our assumption that S is a recommendation rule for preferences $\preceq$. $\qquad \square$

It is interesting that a scoring rule is used to "rank" lotteries in a way that matches the decision maker's preferences over lotteries. This lets us incentivize an expert to reveal the decision maker's most preferred decision strategy. Furthermore, because the decision maker's preferences must admit a *convex* weak utility representation, it is without loss of generality to restrict the expert to reporting *a single action* instead of *a decision strategy*. To see this, let u be the convex function representing the decision maker's preferences, and whenever $p_1 \prec p_2$, we have $u(p_1) < u(p_2)$. By convexity of u, we know that $u(\alpha p_1 + (1 - \alpha)p_2) < u(p_2)$, which implies $\alpha p_1 + (1 - \alpha)p_2 \preceq p_2$. Thus, any mixed decision strategy (which will create a convex combination of lotteries) is always (weakly) less preferred to the best single action (which leads to the most favorable lottery). The expert can simply recommend a single action for the decision maker to deterministically take.

7.5.3 Quasi-Strict Properness and Strictly Proper Recommendation Rules

Recommendation rules incentivize an expert to reveal the decision maker's best decision strategy, but not necessarily to accurately reveal their prediction on likely outcomes if that decision strategy is followed. In [69], scoring rules that the authors called *quasi-strictly proper* incentivized an expert to reveal both for a special case of decision making. In their paper, a decision maker has a finite set of actions and only two outcomes, "good" and "bad." The decision maker solicits an action-outcome matrix from a single expert, then applies a deterministic decision rule to select an action (i.e. no mixed decision strategy is allowed). The authors focus on the natural

special case of their model where the decision rule selects the action most likely to result in the "good" outcome, and show they can create a *quasi-strictly proper* rule with two nice properties: (1) the action that the expert believes will most likely result in the "good" action is always chosen by the decision maker, (2) the expert accurately reports the likely results of this action. These rules are "quasi-strictly" instead of "strictly" proper since the rest of the action-outcome matrix may not be accurate.

In this setting, we no longer request an entire action-outcome matrix when an expert makes a recommendation, and so we can simply describe recommendation rules as strictly proper when they incentivize the expert to accurately reveal its belief about the strategy's outcome.

Definition 19 (Strictly Proper Recommendation Rule). *A scoring rule S is a strictly proper recommendation rule for preferences $\preceq$ if it is a recommendation rule for $\preceq$, that is, $\Phi_Q \subseteq \Sigma_Q^*$ for all $Q \in \mathcal{P}$, and for all $\sigma^* \in \Phi_Q$,*

$$\arg\max_{p \in \Delta(\Omega)} \sum_{a \in \mathcal{A}, \omega \in \Omega} \sigma^*(a) Q(a, \omega) S(p, \omega) = \{Q^T \cdot \sigma^*\}$$

for all $Q \in \mathcal{P}$.

In practice, strictly proper recommendation rules may be interesting as they allow the decision maker to understand and plan for the likely affects of its decision. These rules can be partially characterized immediately as a corollary of my recommendation rule characterization.

Corollary 3 (Strictly Proper Recommendation Rule Characterization). *If preferences $\preceq$ admit a subdifferentiable and strictly convex weak utility representation, then*

there exists a strictly proper recommendation rule for $\preceq$.

The proof is immediate from the first part of Theorem 26, since a strictly convex function implies that the expert uniquely maximizes its expected score when the prediction p is equal to the resultant lottery $Q^T \cdot \sigma^*$. Note, however, this result is not tight, and we leave open the possibility that other types of preferences may have strictly proper recommendation rules.

7.6   Decision Making in Review

This chapter studied the elicitions of predictions and recommendations for decision making. It showed that when eliciting predictions for decision making, strict properness generally required the decision maker risk taking an action at random. This is best interpreted as telling us that a decision maker can improve the likelihood it takes a preferred or "best" action by running a decision market, even though it cannot guarantee it takes such an action. Thus, decision markets are useful, if imperfect.

When working with a single expert, on the other hand, we can acquire a recommended decision strategy and simply take that action. This suggests a trade-off between working with multiple experts, who may combine their knowledge in brilliant ways, and working with a single expert, where there is no risk of taking an action our consultant thinks will be a poor one.

Several avenues of future work are suggested by this chapter. First, the prior chapters of this thesis consider strict properness very generally (Chapters 2–4) and also the challenge of designing securities where information is aggregated. This chap-

ter ducks both challenges. It is not clear that strictly proper decision markets will aggregate information, and the extension from a discrete setting to a more general one is likely interesting.

Second, also in Chapter 4 the duality between scoring rules and cost functions was expressed. When making a decision, however, it is not clear what a cost function market would look like. Decision markets are a generalization of prediction markets, and a cost function for making decisions may also illustrate interesting properties of cost functions.

Finally, my discussion of using a single expert to make recommendations is interesting since our goal is no longer to elicit a prediction, nor even a statistic. This is far from the classical setting where we are trying to elicit a belief, and other techniques may be useful when we are asking for a recommendation. Further, our understanding of strict properness and prediction markets does not provide any method for working with multiple experts when asking for a recommendation. If this could be done it would be of great interest, not only because it would suggest a new means of aggregating information, but also because it would pool multiple experts' information and let a decision maker deterministically take an action. That is likely to be valuable in practice. I also left two immediate mathematical questions open: (1) characterizing the necessary conditions for the existence of strictly proper recommendation rules, and (2) understanding when preferences have a strictly convex weak utility representation.

In the near future it may be that prediction markets are presented through a generalization that accommodates decision making. In fact, I no longer think decision

making is necessarily the most natural lens through which to approach the mathematical discussion in this chapter. We might alternatively think of it as a discussion of how to run a prediction market *when the future is imperfectly observed.* That is, instead of the classical setting where we observe the outcome perfectly, we might consider some noise, or that the true state of the world is beyond us and we only receive some signal about it. This is essentially what is happening here, with the added challenge that we both (1) chose what observation of the true state of the world to make and (2) have preferences about the signal received from doing so.

In short, I think there is a great deal of exciting future work that can begin from this humble chapter, and I hope it will have a substantial impact on the developing field.

8

Conclusion

This thesis covered a great deal of material, from the discussion of fundamental strict properness in Chapters 3 and 4, the development of a cost function for bounded continuous random variables in Chapter 5, the design of simple and informative markets in Chapter 6, and lastly the use of predictions for decision making in Chapter 7. In this conclusion I will review the results of these chapters and add a great deal about possible extensions. I hope these extensions clarify the current work and are interesting to future researchers.

8.1　Strict Properness

This thesis offers a new perspective on strict properness, a perspective where the fundamental object of study is the scoring relation or menu of scoring functions. Both scoring rules and cost functions can be derived from these relations, and are best thought of as offering different languages to access the scoring functions. Scoring

rules use the language of predictions, and cost functions mimic futures markets where traders can buy and sell securities.

Strictly proper scoring relations can be identified with subsets of the points and unique supporting $\mathcal{B}-$subgradients of convex functions of the probability measures. Functions of the probability measures are incredibly special because they always have these supporting subgradients wherever they are subdifferentiable. We also have to be careful to use the $\mathcal{B}-$subgradients since the continuous dual space of the ca space is not well understood, and may contain other objects. This thinking led to a simpler, more geometric characterization of strict properness than that offered by [43]. One interesting distinction in their favor, though, is that they allowed some unbounded functions to be scoring functions, and I will return to this as an opportunity for future work.

This fundamental understanding of strict properness let us define strictly proper cost functions, and clarify their relationship to strictly proper scoring rules. Connections between cost functions and scoring rules had been made previously [2, 1], but none as simply or completely as we were able to offer.

Strict properness is the *essential* property for eliciting and rewarding accurate predictions of the future, but as we saw in Chapter 5 it is not the only one. Not every strictly proper cost function reliably offers securities, or is able to quote meaningful prices. If the space of securities admits a basis, however, then they do. Chapter 5 also developed a cost function for bounded continuous random variables called a *dynamic discretization market* since it effectively let traders decide on *ex interim* discretizations of the continuous outcome space. This market was strictly proper, had bounded loss

and was unbiased, a new property that is both natural and implied the market could be computed using a convex program. Prior work on developing reasonable cost functions for continuous outcome spaces suggested they all had bounded loss [?] or were mistaken [67]. Some more recent work like [2] is closer to my own, and could plausibly be developed to obtain a cost function with bounded loss for a continuous outcome space, although I believe this would be a substantial amount of work.

There is a considerable amount of work left undone and extensions suggested by these developments. In particular:

Understanding the Subdifferential

It is difficult to understand the subdifferential of convex functions for non-Euclidean spaces. This relation is at the heart of strict properness, and it is possible that progress could be made considering only functions of the probability measures. We saw this difficulty in Chapter 5, where we had use results from convex programs to obtain an *existence* result about the subdifferential of the dynamic discretization market. In particular, we need fast ways of understanding this relation so we can map from securities to scoring functions, predictions to scoring functions, and scoring functions to predictions.

The Existence of Convex Functions

Our characterization of strict properness requires a particular type of convex function exist, and it is not always clear if convex functions like it exist or not for arbitrary measurable spaces. Work here may reveal some universal techniques for constructing such functions. Unfortunately, it is not yet clear what domains, other than discrete

and continuous Euclidean space, it is interesting to apply the techniques of this thesis to. In these cases the desired convex functions certainly exist.

Extending Scoring Relations

I required scoring relations associate predictions with the bounded measurable functions. This was an elegant way to identify them with standard convex analysis concepts, but many generalizations are possible. The idea of strict properness readily supports using measurable functions, as opposed to the bounded measurable functions, for instance, although these would present many challenges and likely for little gain. More interesting would be extending scoring relations to handle statistics (also discussed below) or objects other than predictions. This may be very difficult, or even impossible in some cases, yet is likely to have improve our understanding of strict properness.

Generalizing Strict Properness

I have been saying strict properness is the *essential* property for eliciting and rewarding accurate predictions, and it is, but maybe $\epsilon-$strict properness is the *essentially* property for eliciting and rewarding $\epsilon-$accurate predictions. There are many possible generalizations of strict properness that might provide predictions *accurate enough* for our purposes. Mathematically, these might connect in a fascinating way to the notion of the $\epsilon-$subdifferential. These relaxations of strict properness might also have practical computational advantages.

Liquidity Sensitivity

As mentioned, strictly proper cost function are liquidity insensitive—they do not respond to market depth. This is an immediate consequence of strict properness and the way it translates securities to scoring functions. Both the security b and $b + k$ translate to the same scoring function, and thus the market assumes the same prices for both. With two outcomes, that means if one thousand and one securities for the first are sold, and one thousand for the second, the market has the same prices as if just one security for the first outcome were sold. Even though one out of a thousand securities seems insignificant, it has as much impact as if it were the first sold. I cannot immediately think of a solution to this, although there has been some work on liquidity sensitive markets [71, 70, 55].

Hybridizing Scoring Rules and Cost Functions

Understanding scoring relations offers us the possibility of hybridizing both scoring rules and cost functions, obtaining the benefits of both. For example, we might let traders offer predictions or trade securities, and let the market handle translations between these languages. This may even lead to scoring rules that do not require an entire probability measure but, more like the work in Chapter 6, can let traders specify only part of a probability measure, with the market supplying the rest in a clever way.

8.1.1   Valuing the Class of Elicitable Predictions

In Chapter 5 we saw a cost function that let us elicit predictions from the class $\mathcal{P}_{interval}$. It is not clear how valuable this class of predictions is, or how much *more* valuable it is versus offering a market that discretizes the interval *ex ante*. For instance, when eliciting a prediction for the outcome of a continuous random variable on $[0, 1]$, how should we value a market that elicits predictions from $\mathcal{P}_{interval}$ versus one that asks for predictions of how likely the result is to be in $[0, .5]$ or $(.5, 1]$. Intuitively, greater prediction precision seems better, and it would be nice to make a formal claim or argument that it actually *is* better.

Statistic Elicitation

One problem recently formalized is the accurate elicitation of statistics instead of entire probability measures [54]. This is a more general problem than acquiring a probability measures, and one that could likely benefit from this thesis' perspective on strict properness.

Understanding Other Markets

There is a great deal of work on prediction markets other than those described here: dynamic pari-mutuel markets [72], markets that set prices based on trader behavior [15, 28], and call markets [35, 20, 4, 42] are some examples. Understanding whether or not these markets are strictly proper, and what incentive challenges they face, could be fascinating.

8.1.2   Relaxing No Arbitrage

Some other markets [1, 31] suggest relaxing the no arbitrage property to improve how quickly the market can be computed. Computing a market quickly is a serious issue, and creating a framework for strict properness that incorporates relaxations like this would like be interesting.

8.2   Simple and Informative Markets

Chapter 6 discussed the design of simple and informative markets with Bayesian traders. Unfortunately, it showed that designing these markets was trivial and unhelpful if we knew nothing about traders' information structure, and it was NP-hard if we knew everything. Chapter 6 also allowed no opportunity for work in between. This chapter likely suggests more questions than it answers, including these two:

Empirically Understanding Internal vs. External Aggregation in Markets

We lack a good understanding of whether markets aggregate information internally, because traders are reviewing past trades and price updates, or simply responding to external signals. The observed effectiveness of simple markets seems to suggest a great deal of the latter. It may also be interesting to attempt to understand the signal structures traders have in practice, or at least how they interpret the information they are presented with, and how they think others are interpreting this information.

Notions of Partial Aggregation

As mentioned, Chapter 6's idea of informativeness is all-or-nothing. It may be useful and interesting to develop an idea of partial aggregation, where some if not all of traders' information appears in the market. Reviewers of Chapter 6 regularly suggested this extension.

I believe there is a great deal of empirical work required first before we can usefully return to aggregation in prediction markets. Theoretical work like [65] is fascinating and uses neat mathematical techniques, but it may be too far from the reality of the situation to be interesting other than as a mathematical investigation. Learning that markets mostly respond to outside sources would have dramatic implications for our field, drastically changing how we approach it. Even if only some information came from outside, this would be fascinating. Investigating the use of external sources in markets is likely the most important problem in the field today.

8.3   Expert Advice and Decision Making

Chapter 7 described how we could use techniques related to scoring predictions to make decisions. It began by describing decision markets, where a decision maker elicits predictions about the likely results of each of a set of possible actions. The decision maker can then review these predictions to help it decide on a course of action, although it must also risk taking *any* action with *some* chance. Intuitively, the market must implement an unbiased estimator of the future. This means that a decision market can *improve* the likelihood a decision maker takes a preferred action,

although the market cannot guarantee it will.

In the second part of Chapter 5 an alternative model where a single expert made a recommendation was considered. In this setting the prediction is secondary to the recommendation, and I showed that we could elicit what the expert believed to be the best available action and deterministically take it. The technique used, however, does not generalize to work with multiple experts, and so there is at trade-off between leveraging the expertise of many experts or deterministically taking the action a single expert believes optimal.

There are two particularly interesting possible extensions for this work:

Decision Markets in Practice

Running a decision market would be fascinating, both to see how well it would work, and whether a decision maker would actually be willing to risk taking any action *ex post*. Perhaps also, in practice, strict properness is not necessary because traders are altruistic, and we can safely deterministically take an action in many cases.

Alternatives to Recommendation Rules

The recommendation rules described in the second part of Chapter 5 let a single expert recommend an action to the decision maker. These rules cannot be immediately extended to handle multiple experts. What is left open as a possibility, however, is that there are alternatives to decision markets and recommendation rules that do let multiple experts pool their information and recommend an action. Such a method would be of considerable interest.

8.4   In Conclusion

Thank you for reading; I hope some parts of this thesis resonated and excite you about the many possibilities here for future work.

Bibliography

[1] Jacob Abernethy, Yiling Chen, and Jennifer Wortman Vaughan. Efficient market making via convex optimization, and a connection to online learning. *ACM Transactions on Economics and Computation (TEAC)*, 1(2):12:1–12:39, May 2013.

[2] Jacob Abernethy and Rafael Frongillo. A collaborative mechanism for crowd-sourcing prediction problems. In *Neural Information Processing Systems*, 2011.

[3] Jacob Abernethy and Rafael Frongillo. A characterization of scoring rules for linear properties. In *COLT*, 2012.

[4] Shipra Agrawal, Zizhuo Wang, and Yinyu Ye. Parimutuel betting on permutations. In *Proceedings of the 4th International Workshop on Internet and Network Economics*, pages 126–137, 2008.

[5] J.C.R Alcantud and C Rodriguez-Palmero. Characterization of the existence of semicontinuous weak utilities. *Journal of Mathematical Economics*, 32(4):503 – 509, 1999.

[6] Beth Allen. Strict rational expectation equilibria with diffuseness. *Journal of Economic Theory*, 27(1):20–46, 1982.

[7] Robert B. Ash and Catherine A. Doléans-Dade. *Probability & Measure Theory.* Academic Press, second edition, 2000.

[8] Robert J. Aumann. Agreeing to disagree. *Annals of Statistics*, 4(6):1236–1239, 1976.

[9] Viorel Barbu and Teodor Precupanu. *Convexity and Optimization in Banach Spaces.* Springer Monographs in Mathematics. Springer, 4th edition, 2012.

[10] Joyce E. Berg, Robert Forsythe, Forrest D. Nelson, and Thomas A. Rietz. Results from a dozen years of election futures markets research. In C. A. Plott and V. Smith, editors, *Handbook of Experimental Economic Results.* 2001.

[11] Joyce E. Berg and Thomas A. Rietz. Prediction markets as decision support systems. *Information Systems Frontier*, 5:79–93, 2003.

[12] Jonathan M. Borwein. Fifty years of maximal monotonicity, 2010.

[13] Jonathan M. Borwein and Jon Vanderwerff. *Convex Functions: Constructions, Characterizations and Counterexamples.* Cambridge University Press, 2010.

[14] Craig Boutilier. Eliciting forecasts from self-interested experts: Scoring rules for decision markets. *In Proceedings of the Eleventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS-12)*, 2012.

[15] Aseem Brahma, Mithun Chakraborty, Sanmay Das, Allen Lavoie, and Malik Magdon-Ismail. A Bayesian market maker. In *ACM EC*, 2012.

[16] Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, January 1950.

[17] Yiling Chen, Chao-Hsien Chu, Tracy Mullen, and David M. Pennock. Information markets vs. opinion pools: an empirical comparison. In *EC '05: Proceedings of the 6th ACM conference on Electronic commerce*, pages 58–67, New York, NY, USA, 2005. ACM.

[18] Yiling Chen, Stanko Dimitrov, Rahual Sami, Daniel M. Reeves, David M. Pennock, Robin D. Hanson, Lance Fortnow, and Rica Gonen. Gaming prediction markets: Equilibrium strategies with a market maker. *Algorithmica*, 2009.

[19] Yiling Chen, Lance Fortnow, Nicolas Lambert, David M. Pennock, and Jennifer Wortman. Complexity of combinatorial market makers. In *EC '08: Proceedings of the 9th ACM conference on Electronic commerce*, pages 190–199. ACM, 2008.

[20] Yiling Chen, Lance Fortnow, Evdokia Nikolova, and David M. Pennock. Betting on permutations. In *ACM EC*, 2007.

[21] Yiling Chen and David M. Pennock. A utility framework for bounded-loss market makers. In *UAI*, 2007.

[22] Yiling Chen, David M. Pennock, and Tejaswi Kasturi. An empirical study of dynamic pari-mutuel markets: Evidence from the Tech Buzz Game. In *WebKDD '08: Proceedings of the Workshop on Web Mining and Web Usage Analysis*, 2008.

[23] Yiling Chen, Mike Ruberry, and Jennifer Wortman Vaughan. Cost function mar-

ket makers for measurable spaces. *In Proceedings of the 14th ACM Conference on Electronic Commerce (EC)*, June 2013.

[24] Yiling Chen, Mike Ruberry, and Jennifer Wortman Vaughan. Designing informative securities. In *UAI*, 2012.

[25] Yiling Chen and Jennifer Wortman Vaughan. A new understanding of prediction markets via no-regret learning. In *ACM EC*, 2010.

[26] Bo Cowgill, Justin Wolfers, and Eric Zitzewitz. Using prediction markets to track information flows: Evidence from Google. Working paper, 2008.

[27] John Dalton. *Meteorological Observations and Essays*. 1793.

[28] S. Das and M. Magdon-Ismail. Adapting to a market shock: Optimal sequential market-making. In *NIPS*, 2008.

[29] Peter DeMarzo and Costis Skiadas. Aggregation, determinacy, and informational efficiency. *Journal of Economic Theory*, 80:123–152, 1998.

[30] Peter DeMarzo and Costis Skiadas. On the uniqueness of fully informative rational expectations equilibria. *Economic Theory*, 13:1–24, 1999.

[31] Miroslav Dudik, Sébastian Lahaie, and David M. Pennock. A tractable combinatorial market maker using constraint generation. *ACM Conference on Electronic Commerce*, pages 459–476, 2012.

[32] Joan Feigenbaum, Lance Fortnow, David M. Pennock, and Rahul Sami. Computation in a distributed information market. *Theoretical Computer Science*, 343(1-2):114–132, 2005.

[33] Mark D. Flood. An introduction to complete markets. *The Federal Reserve Bank of St. Louis Review*, 1991.

[34] Robert Forsythe, Forrest Nelson, George R. Neumann, and Jack Wright. Anatomy of an experimental political stock market. *American Economic Review*, 82(5):1142–1161, 1992.

[35] Lance Fortnow, Joe Kilian, David M. Pennock, and Michael P. Wellman. Betting boolean-style: A framework for trading in securities based on logical formulas. *Decision Support Systems*, 39(1):87–104, 2004.

[36] Rafael M. Frongillo and Ian Kash. General truthfulness characterizations via convex analysis. 2012.

[37] Xi Gao, Yiling Chen, and David M. Pennock. Betting on the real line. In *WINE*, 2009.

[38] Xi Alice Gao and Yiling Chen. An axiomatic chracterization of continuous-outcome market makers. *In Proceedings of the 6th Workshop on Internet and Network Economics (WINE)*, pages 505–514, December 2010.

[39] P. H. Garthwaite, J. B. Kadane, and A. O'Hagan. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100:680–701, 2005.

[40] John D. Geanakoplos. Common knowledge. In R. Aumann and S. Hart, editors, *Handbook of Game Theory*, volume 2, pages 1437–1496. North Holland, 1994.

[41] John D. Geanakoplos and Heracles M. Polemarchakis. We can't disagree forever. *Journal of Economic Theory*, 28(1):192–200, 1982.

[42] Mohammad Ghodsi, Hamid Mahini, Vahab S. Mirrokni, and Morteza Zadimoghaddam. Permutation betting markets: Singleton betting with extra information. In *ACM EC*, 2008.

[43] Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, March 2007.

[44] Julio González-Díaz and Miguel A. Meléndez-Jiménez. On the notion of perfect bayesian equilibrium. *TOP*, November 2011.

[45] Sanford J. Grossman. An introduction to the theory of rational expectations under asymmetric information. *Review of Economic Studies*, 48(4):541–559, 1981.

[46] Cleve Hallenbeck. Forecasting precipitation in percentages of probability. *Monthly Weather Review*, pages 645–647, November 1920.

[47] Robin Hanson. Decision markets. *IEEE Intelligent Systems*, 14(3):16–19, 1999.

[48] Robin Hanson. Combinatorial information market design. *Information Systems Frontiers*, 5(1):107–119, 2003.

[49] Robin D. Hanson. Logarithmic market scoring rules for modular combinatorial information aggregation. *Journal of Prediction Markets*, 1(1):1–15, 2007.

[50] Arlo D. Hendrickson and Robert J. Buehler. Proper scores for probability forecasters. *Annals of Mathematical Statistics*, 42(6):1916–1921, 1971.

[51] Krishnamur Iyer, Ramesh Johari, and Ciamac C. Moallemi. Information aggregation in smooth markets. In *Proceedings of the Eleventh ACM Conference on Electronic Commerce*, pages 199–205, 2010.

[52] Erwin Kreyszing. *Introductory Functional Analysis with Applications*. Wiley Classics Library. Wiley, 1989.

[53] Sebastien Lahaie, Miro Dudik, David Rothschild, and David Pennock. A combinatorial prediction market for the u.s. elections. *ACM Conference on Electronic Commerce*, 2013.

[54] Nicolas S. Lambert. Elicitation and evaluation of statistical forecasts. December 2011.

[55] Xiaolong Li and Jennifer Wortman Vaughan. An axiomatic characterization of adaptive-liquidity market makers. In *ACM EC*, 2013.

[56] Bernard Mangold, Mike Dooley, Rael Dornfest, Gary W. Flake, Havi Hoffman, Tejaswi Kasturi, and David M. Pennock. The tech buzz game. *IEEE Computer*, 38(7):94–97, 2005.

[57] A. Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomics Theory*. Oxford University Press, New York, NY, USA, 1995.

[58] James E. Matheson and Robert L. Winkler. Scoring rules for continuous probability distributions. *Management Science*, 22(10):1087–1096, June 1976.

[59] John McCarthy. Measures of the value of information. *PNAS: Proceedings of the*

National Academy of Sciences of the United States of America, 42(9):654–655, 1956.

[60] Richard D. McKelvey and Talbot Page. Common knowledge, consensus, and aggregate information. *Econometrica*, 54(1):109–127, 1986.

[61] Allan H. Murphy. The early history of probability forecasts: Some extensions and clarifications. *Weather and Forecasting*, 13:5–15, March 1998.

[62] Lars T. Nielsen. Common knowledge, communication and convergence of beliefs. *Mathematical Social Sciences*, 8:1–14, 1984.

[63] Lars T. Nielsen, Adam Brandenburger, John Geanakoplos, Richard McKelvey, and Talbot Page. Common knowledge of an aggregate of expectations. *Econometrica*, 58(5):1235–1238, 1990.

[64] Kent Osband. Optimal forecasting incentives. *Journal of Political Economy*, 97(5):pp. 1091–1112, 1989.

[65] Michael Ostrovsky. Information aggregation in dynamic markets with strategic traders. *Econometrica*, 80(6):2595–2648, November 2012.

[66] Abraham Otham. Automated market making: Theory and practice. *Thesis*, 2012.

[67] Abraham Othman and Tuomas Sandholm. Automated market makers that enable new settings: Extending constant-utility cost functions. In *AMMA*, 2010.

[68] Abraham Othman and Tuomas Sandholm. Automated market making in the large: The Gates Hillman prediction market. In *EC '10: Proceedings of the 11th*

ACM Conference on Electronic Commerce, pages 367–376, New York, NY, USA, 2010. ACM.

[69] Abraham Othman and Tuomas Sandholm. Decision rules and decision markets. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 625–632, 2010.

[70] Abraham Othman and Tuomas Sandholm. Liquidity-sensitive automated market makers via homogeneous risk measures. In *WINE*, 2011.

[71] Abraham Othman, Tuomas Sandholm, David M. Pennock, and Daniel M. Reeves. A practical liquidity-sensitive automated market maker. In *ACM EC*, 2010.

[72] David M. Pennock. A dynamic pari-mutuel market for hedging, wagering, and information aggregation. In *Proceedings of the Fifth ACM Conference on Electronic Commerce (EC'04)*, 2004.

[73] David M. Pennock, Steve Lawrence, C. Lee Giles, and Finn A. Nielsen. The real power of artificial markets. *Science*, 291:987–988, 2002.

[74] David M. Pennock and Michael P. Wellman. Representing aggregate belief through the competitive equilibrium of a securities market. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, 1997.

[75] Josep E. Peris and Begoa Subiza. A weak utility function for acyclic preferences. *Economics Letters*, 48(1):21 – 24, 1995.

[76] Charles R. Plott and Shyam Sunder. Efficiency of experimental security markets with insider information: An application of rational expectations models. *Journal of Political Economy*, 90(4):663–98, 1982.

[77] Charles R. Plott and Shyam Sunder. Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica*, 56(5):1085–1118, 1988.

[78] Charles R. Plott, Jorgen Wit, and Winston C. Yang. Parimutuel betting markets as information aggregation devices: Experimental results. Technical Report Social Science Working Paper 986, California Institute of Technology, April 1997.

[79] Philip M. Polgreen, Forrest D. Nelson, and George R. Neumann. Using prediction markets to forecast trends in infectious diseases. *Clinical Infectious Diseases*, 44(2):272–279, 2007.

[80] Roy Radner. Rational expectations equilibrium: Generic existence and the information revealed by prices. *Econometrica*, 47(3):655–678, 1979.

[81] Frank P. Ramsey. Truth and probability. In S. L. Zabell, editor, *Symmetry and its Discontents: Essays on the History of Inductive Probability*, chapter 6, pages 119–141. Cambridge University Press, 2005.

[82] R. Tyrrell Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific Journal of Mathematics*, 17(3):497–510, 1966.

[83] Leonard J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

[84] Lloyd Shapley and Martin Shubik. Trading using one commodity as a means of payment. *Journal of Political Economy*, 85:937–968, 1977.

[85] Peng Shi, Vincent Conitzer, and Mingyu Guo. Prediction mechanisms that do not incentivize undesirable actions. In *WINE '09: Proceedings of the 5th International Workshop on Internet and Network Economics*, pages 89–100, Berlin, Heidelberg, 2009. Springer-Verlag.

[86] Martin Spann and Bernd Skiera. Internet-based virtual stock markets for business forecasting. *Management Science*, 49(10):1310–1326, 2003.

[87] Justin Wolfers and Eric Zitzewitz. Interpreting prediction market prices as probabilities. Technical report, NBER Working Paper No. 10359, 2005.

[88] Lirong Xia and David M. Pennock. An efficient monte-carlo algorithm for pricing combinatorial prediction markets for tournaments. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, 2011.