Structure and Design of Informational Substitutes

by

Jerry Anunrojwong

A thesis submitted in partial fulfillment for the
degree of Bachelors of Arts in
Applied Mathematics

Harvard University
Cambridge, Massachusetts

March 2018

Abstract

We analyze structure and design of informational substitutes and complements, as proposed by Chen and Waggonner (2016). First, we characterize "universal" complements, or information structures such that signals are complements for every decision problem, as precisely variants of the exclusive-or (XOR) of binary signals. This characterization is important because equilibria in the corresponding prediction market games are always the "worst-possible" regardless of design. Second, we show that the problem of designing the market for substitutability is equivalent to solving a linear program, and that for many common information structures, such a linear program can be solved in polynomial time. Third, we extend informational substitutes to predicting continuous distributions and distribution properties, such as mean and median, and show that they sometimes behave unintuitively. In particular, conditionally independent gaussian signals are complements under a wide range of standard decision problems.

Acknowledgements

I acknowledge and appreciate guidance and mentorship from Yiling Chen, Gordon McKay Professor of Computer Science at Harvard University, who is my primary advisor for this thesis. I also want to thank Bo Waggoner, a postdoctoral researcher at University of Pennsylvania at the time of this writing, for his valuable ideas. This thesis is related to a research project that the three of us are currently collaborating on.

I am indebted to the community of researchers and advisors at Harvard University who contributed to my undergraduate experience. I would like to thank David Parkes, a second reader on this thesis, who introduced me to the fascinating area of economics and computation. He continues to be a source of advice as I start doing research in the field.

I especially want to thank the following individuals for guiding me through the dynamic and sometimes turbulent world of research: Cynthia Dwork, Benjamin Golub, Scott Kominers, Nicole Immorlica, Brendan Lucier, Yaron Singer, and Robert Townsend.

Lastly, I would like to thank my father, Watchara Anunrojwong, and my mother, Jirapa Anunrojwong, for their constant support and encouragement.

Contents

Chapter 1

Introduction

Agents need to make decisions in an uncertain world. Such decisions are based on pieces of information that they need to collect and interpret; the more they collect information, the higher the expected payoff will be. This naturally leads to the notion of value of information. We want to understand and guide how agents acquire and use information.

Consider the analogous situation of value of *items*, where agents have a *valuation function* over subsets of items. This valuation function is a set function. A set of items are *substitutes* if this valuation function is *submodular*, meaning that the marginal value of an additional item is decreasing in the subset of items. The more items they have, the lower value they assign to an additional item. There is a rich body of theories relating substitutability of items to structural, algorithmic, and game-theoretic properties. For example, substitutability implies existence of market equilibria, and vice versa for complementarity (Alexander S. Kelso and Crawford, 1982; Hatfield and Milgrom, 2005; Ostrovsky, 2008), and bundles of substitute items are "easy" to sell in auctions (Lehmann et al., 2006). More importantly, these connections are made through submodularity of the valuation function.

Chen and Waggonner (2016) proposes an analogous definition of substitutability for information based on submodularity of the *value of information* function, just like substitutability of items is based on submodularity of the valuation function of items. The value of information function is exactly what it sounds like: the value of information A is the expected reward that an agent will get by using A in making a decision in a decision problem of interest. This definition of informational substitutes has a crucial difference from that of substitutes of items: while an agent can have a valuation function over items independently of context, the value of information to the agent depends on context, the *decision problem* for which these pieces of information will be used. Moreover, each piece of information can have its internal structures and probabilistic relationships that do not exist for items. Chen and Waggonner (2016) accounts for this structure by modeling signals as *lattices*. In a lattice, information can be partially ordered (a piece of information might be more informative than another one, but some pairs are incomparable) and combined (if an agent knows two pieces of information, that agent knows a "combined piece" of information). These two properties make lattices a good model for information. Importantly, the diminishing marginal value definition of submodularity still holds on lattices, and they reduce to the standard submodular set function definition when the lattice treats each signal as an "item" without internal structure. This correspondence between the definition of informational substitutes with the more established definition of substitutability of items suggests that informational substitutes is a natural notion in the study of decision making under uncertainty.

Beyond the aforementioned correspondence, Chen and Waggonner (2016) shows that informational substitutes gives "best-possible" information aggregation and informational complements gives "worst-possible" information aggregation in prediction markets. This result unifies several previous results on information aggregation in prediction markets with strategic agents. To fully appreciate the significance of this result, we will first describe the function and significance of markets in general, and prediction markets in particular.

The market mechanism is a common way people reveal and aggregate information and learn from others. If the relevant information is captured in some form of tradable assets, such as stocks (representing value of a company) or bonds (representing a borrower's creditworthiness), the market can be implemented by letting agents buy and sell the corresponding tradable assets. The market prices at a given time can be interpreted as representing the "consensus" valuation, a result of aggregating information from participating agents.

We can also use the market to aggregate information about the likelihood of future uncertain outcomes such as the outcome of an upcoming election, or the chance of a successful product launch. These outcomes are not tradable assets, but the market designer can create synthetic assets or financial contracts *whose values depend on uncertain outcomes of interest*. Such markets are called *prediction markets*. For example, if we are interested in whether a binary event X will occur, we can create a contract C that pays 1 if X occurs and 0 otherwise. Then, we can open the market and attract traders to speculate on the value of the contract. At any given time, the market price of C is between 0 and 1, and can be interpreted as the probability of X occurring.

Prediction markets have several advantages over alternative methods of aggregating information, such as polls and expert surveys: they encourage broad participation; they directly gauge confidence of agents based on their stakes; they align the incentives of agents with that of the information aggregator. In fact, prediction markets and "wisdom of the crowds" are often more accurate than experts (Mollick and Nanda, 2015; Tetlock and Gardner, 2016).

Despite the success of prediction markets in practice, progress on theoretical analysis of their information aggregation properties is quite slow. Savage (1971) suggests using *(strictly) proper scoring rules*, which (strictly) maximizes the immediate reward under truthful report. However, proper scoring rules only guarantee good information aggregation if agents participate only once, or are myopic. A non-myopic trader can conceivably employ non-truthful strategies to bring market probabilities away from her subjective probabilities and mislead other traders so she can reap greater profit later when she brings the probabilities back to her believed values.

This concern is not a mere theoretical curiosity: every seasoned speculator in financial markets is familiar with the "pump and dump" market manipulation scheme whereby a trader repeatedly buys a worthless security at a gradually increasing price to create a perception of activity and value. When enough other investors follow suit and drive up the price, the trader sells off her position at a profit and the price crashes. In our framework, the trader is not truthful at the beginning because she buys the asset even when the price of the asset is higher than her value for it, but she does so to convince other traders that the asset is valuable and drive up the price for her future profit.

We can see that the problem of analyzing strategic behaviors in prediction markets is very challenging. It therefore comes as no surprise that previous works (Chen et al., 2007; Dimitrov and Sami, 2008; Chen et al., 2010; Gao et al., 2013) analyze equilibria of prediction market games only in very special cases. They tell us, for instance, that independent signals are "bad" and conditionally independent signals are "good" for information aggregation. Our notion of informational substitutes unifies all these special cases into one natural framework. Specifically, we view the prediction market game as an extensive form game. Signals are substitutes if and only if, in any equilibrium and for any arrival order of traders, all traders rush to truthfully reveal their information at their first opportunity: the "all-rush" equilibrium under substitutes is the best-possible. Signals are complements if and only if, in any equilibrium and for any arrival order of traders, all traders delay and truthfully reveal their information at their last opportunity: the "all-delay" equilibrium under complements is the worst-possible.

The natural correspondence via submodular functions and the game-theoretic results in prediction markets, taken together, suggest that informational substitutes should play a central role in the study of strategic information revelation and aggregation. However, informational substitutes is complex, making it hard for practical use. The value of information function is a complicated mathematical object that depends jointly on the information structure and the decision problem. Checking submodularity of this value of information function, which informational substitutes demands, is even more complicated. To make informational substitutes an operational definition, we need to better understand its *structure*. Moreover, since the market designer has control over the decision problem (equivalently, the reward function) but not the information structure (which is a structural property of the situation at hand), we are interested in the *design* problem. Given an information structure, can we design a decision problem such that signals are substitutes? In light of prediction market results of informational substitutes, we can restate the design question as follows: can we design market scoring rules such that it is incentive compatible for self-interested forward-looking traders to report their true beliefs? The study of *structure and design* of informational substitutes is the focus of this work.

1.1 Overview and Contribution

We model the information structure with the standard Bayesian model of information with Aumann partition, following Aumann (1976). At the beginning of the game, each trader receives a signal. The joint distribution between all signals and the outcome is common knowledge, but the value of each signal is only known to the trader who receives it. In other words, every trader knows exactly what type of information other traders get but not specific values; she knows only the value of her own signal.

Chen and Waggonner (2016) identify a condition called *informational substitutes*, generalizing the substitutability condition of Börgers et al. (2013). We already explain and motivate their definition of informational substitutes in the introduction. We defer statements of precise definitions and formal model setups to Chapter 2.

The informational substitutes condition takes into account the information structure and the decision problem jointly. If the information structure is such that signals are substitutes (complements) for every decision problem, we call that information structure *universal substitutes (complements)*. Universal substitutes and complements are interesting because they provide the strongest possible guarantee independent of the decision problem, and because they remove the complex interaction between scoring rules and information structures. Chen and Waggonner (2016) have shown that universal weak substitutes must be "almost trivial" and universal moderate and strong substitutes must be "trivial" in a certain well-defined sense. The main contribution of Chapter 3 is characterizing universal complements. We show that universal weak complements are precisely variants of an exclusive-or (XOR) of binary signals, and universal moderate and strong complements must be trivial.

The characterization of universal complements naturally leads to our next investigation in Chapter 4, designing for substitutability. Given an information structure, find a decision problem (equivalently, a scoring rule) such that the signals are substitutes. This is the content of Section 4.2. Unless our problem falls into a rather small class of universal complements, we know that such a task is not *a priori* hopeless. However, there are many problems that are neither substitutes nor complements, so information structures that are not universal complements might still not have associated substitutability-inducing scoring rules. We show that given an information structure, the problem of deciding existence of a substitutability-inducing scoring rule, and explicitly computing one if it exists, can be formulated as a linear programming problem. Therefore, it can always be solved in time polynomial in the number of variables and constraints in the linear program. We show that important classes of information structures, those with independent signals or with conditionally independent signals, can be solved efficiently with this method. We also described geometric requirements of substitutability that suggest how market designers can find a scoring rule tailored to an information structure under consideration without explicitly invoking linear programs.

The linear programming approach can handle discrete signals, but not continuous signals. Fortunately, standard models of continuous signals, such as gaussian distributed signals, are well-structured and we can combine this with a parametrized class of scoring rules for substitutability. This is the content of Section 4.3. We also consider substitutability under elicitation of properties of distributions in Section 4.4. The two sections 4.3 and 4.4 also carry a broader message that substitutability of signals sometimes behaves quite unintuitively, and complete characterization is likely to be challenging, if not impossible. For example, conditionally independent gaussian signals can be complements for a wide range of scoring rules.

Lastly, Chapter 5 recaps the contribution of this work, and discusses possible future directions.

Chapter 3 and 4 are original to this thesis.

1.2 Related Work

The closest paper to our work is Chen and Waggonner (2016). They propose a definition of substitutes and complements in information markets via a connection to submodularity of the value of information function. Their definition is analogous to substitutability for items and is well-motivated. Thus, we adopt their definition in our work. Their work builds on Börgers et al. (2013) which proposes a similar but less expressive definition of informational substitutes.

The main technical tools in this work are inequalities of convex functions, convex optimization, and linear programming. The main references are Boyd and Vandenberghe (2004) and Bertsekas et al. (2003).

The idea that information should be "quickly" aggregated in markets is called "efficient market hypothesis." Fama (1970) discusses different versions of this hypothesis. However, few theoretical works directly address information aggregation. Works in this line of research are often concerned with market microstructure and asset price dynamics. Hasbrouck (2007) gives an overview of this literature. Kyle (1985) proposes a model of financial markets with informed and noise traders. Kyle's model becomes a benchmark model, but not much is known about game-theoretic equilibria of this model until Ostrovsky (2012) shows that information is eventually aggregated under certain natural conditions. However, it is challenging in their framework to find how fast information is aggregated. The notion of information substitutes ensures that information is aggregated as fast as it possibly can.

Our model of the market is based on *prediction markets*. The notion of cost-based market maker in prediction markets, which is crucial to our market model, is due to Hanson (2003). Such models are based on proper scoring rules (Savage, 1971; Gneiting and Raftery, 2007). There are some previous works regarding equilibria of prediction market games (Chen et al., 2007; Dimitrov and Sami, 2008; Chen et al., 2010; Gao et al., 2013). As noted by Chen and Waggonner (2016), these results are subsumed by our results on informational substitutes.

Chapter 2

Definitions and Background

In this chapter, we formally present definitions and model settings. We also present and discuss theorems and results that are related to current work. We also discuss classical results in probability theory and statistics that we use extensively in this work.

2.1 Settings: Information Structures, Decision Problems, Signal Lattices

We model the information structure with the standard Bayesian model of probabilistic information. This modeling choice is standard and we largely follow the notations of Chen and Waggonner (2016).

Definition 2.1 (Information Structure)**.** There is a random event E of interest and n "base signals" $A_1, \ldots, A_n$ modeled as random events. An *information structure* is a prior joint distribution of $(E, A_1, \ldots, A_n)$. This joint distribution is common knowledge.

We will use p to denote a probability distribution on E, an event of interest, so $p(e)$ refers to the probability that $E = e$, and $p(e|a_i)$ refers to the probability that $E = e$ conditional on $A_i = a_i$, which can be computed via Bayesian updating $p(e|a_i) = p(e, a_i)/p(a_i)$. We will use the following shorthand notation p_a to refer to the posterior distribution of e conditional on $A = a$, and $p_{a,b}$ to refer to the posterior distribution of e conditional on $A = a$ and $B = b$, and so on. We overload the notation and also write E for the set of outcomes, so we can write, for example, $e \in E$. The notation $\mathbb{E}_{a \sim A}$ is the expectation over $a \in A$, which we sometimes shorten to $\mathbb{E}_a$. A set of signals is denoted $\mathcal{L}$. A probability distribution on E is equivalent to a simplex on E, and is denoted Δ_E,.

Definition 2.2 (Decision Problem)**.** A *decision problem* consists of a set of event outcomes E, a decision space $\mathcal{D}$, and a utility function $u : \mathcal{D} \times E \to \mathbb{R}$ where $u(d, e)$ is the utility for taking action d when the event outcome is $E = e$.

The agent derives her value from this decision problem in the context of an information structure. Specifically, an agent has prior p and after observing signal A, she updates her information on E to the posterior p_a and then choose a decision to maximize her expected utility given this posterior belief. The value of signal A to her is the expectation over $a \sim A$ of her expected utility from such a decision.

Definition 2.3 (Value of Information)**.** Given a decision problem with a set of event outcomes E, a decision space $\mathcal{D}$, and a utility function $u : \mathcal{D}\times E \to \mathbb{R}$, and an information structure, the prior P over E and the signals, the *value of information function* is

$$\mathcal{V}^{u,P}(A) = \mathop{\mathbb{E}}_{a}\left[\max_{d\in\mathcal{D}} \mathop{\mathbb{E}}_{e}[u(d, e)|A = a]\right]$$

This notation follows Chen and Waggoner (2016). However, this definition has been proposed earlier in the literature, for example, in Howard (1966); Athey and Levin (2001) in the context of decision theory.

We now give definitions of signal lattices, which are our models of information. Following Chen and Waggoner (2016), we consider three types of lattices, the subsets signal lattice, the discrete signal lattice, and the continuous signal lattice. These three types correspond to different levels of fineness in the Aumann information partition (Aumann, 1976), with the subsets signal lattice being the coarsest and the continuous signal lattice the finest.

Definition 2.4 (Lattice)**.** A lattice $(U, \preceq)$ is a set U together with a partial order $\preceq$ on it such that for all $A, B \in U$, there are a *meet* $A \wedge B$ and a *join* $A \vee B$ in U satisfying

1. $A \wedge B \preceq A \preceq A \vee B$ and $A \wedge B \preceq B \preceq A \vee B$
2. the meet and join are the "highest" and "lowest" (respectively) elements in the order satisfying these inequalities.

In a lattice, $\perp$ denotes the "bottom" element and $\top$ the "top" element, *i.e.* $\perp \preceq A \preceq \top$ for all $A \in U$.

Definition 2.5 (Subsets Signal Lattice)**.** The subsets signal lattice generated by $A_1, \ldots, A_n$ consists of an element A_S for each subset S of $\{A_1, \ldots, A_n\}$, where A_S is the signal conveying all realizations $\{A_i = a_i : i \in S\}$. Its partial order is $A_S \preceq A_{S'}$ if an only if $S \subseteq S'$. Hence, its meet operation is given by set intersection and join by set union.

The other two definitions depend on Aumann's classical model of *information partition*. Let $\Gamma \subseteq A_1 \times \cdots A_n$ consists of all signal realizations $(a_1, \ldots, a_n)$ in the support of the prior distribution. A partition is a collection of subsets of Γ such that each $\gamma \in \Gamma$ is in exactly one subset. Each signal A_i corresponds to a partition of Γ with one subset for each outcome a_i, namely, the set of realizations $\gamma = (\cdots, a_i, \cdots)$.

The partitions of Γ form a lattice and the partial ordering is that $A \preceq B$ if the partition of A is coarser than that of B, that is, each element of A is partitioned by elements of B. If $A \preceq B$, we also say that B is finer than A. The join of two partitions is then the coarsest common refinement, and the meet, the finest common coarsening.

Note that $A \preceq B$ or that A is coarser than B means that A is "less informative" than B. This will be true for all three definitions of signal lattices. (In fact, the Aumann partition is implicit in the definition of subsets signal lattice, where we consider the partition that treats each signal independently.) The bottom element $\perp$ is a null signal, corresponding to no information beyond the prior, and the top element $\top$ corresponds to the maximum amount of information, observing all signals.

Definition 2.6 (Discrete Signal Lattice)**.** The *discrete signal lattice* generated by $A_1, \ldots, A_n$ consists of all signals corresponding to partitions of Γ, where Γ is the subset of $A_1 \times \cdots A_n$ with positive probability. Its partial order has $A \preceq B$ if the partition of A is coarser than that of B.

Definition 2.7 (Continuous Signal Lattice)**.** For each partition Π of Γ, let R_Π be drawn independently from the uniform distribution. Let $\Gamma' = \Gamma \times \mathbf{R}$ where $\mathbf{R} = \times_\Pi R_\Pi$. The *continuous signal lattice* consists of a signal corresponding to each partition of Γ'. Its partial order has $A \preceq B$ if the partition of A is coarser than that of B.

2.2 Substitutes and Complements

The main goal of this work is to understand the structure and design of substitutes and complement of information signals. Therefore, the following definition of informational substitutes and complements (S&C), taken from Chen and Waggonner (2016), is central to this work.

Definition 2.8. A function f from a lattice to the reals is *submodular* if it exhibits diminishing marginal value: for all A', A, B on the lattice with $A \wedge B \preceq A' \preceq A$,

$$f(B \wedge A') - f(A') \geq f(B \wedge A) - f(A).$$

It is *supermodular* if it exhibits increasing marginal value: for all A', A, B on the lattice with $A \wedge B \preceq A' \preceq A$, the above inequality is reversed. The sub- or super-modularity is *strict* if, whenever A and B are incomparable on the lattice's ordering and $A' \neq A$, the inequality is strict.[1]

Definition 2.9 (Informational S&C, Chen and Waggonner (2016))**.** In the context of a decision problem u and prior P, the signals $A_1, \ldots, A_n$ are *(weak, moderate, strong) substitutes* if the value of information function $\mathcal{V}^{u,P}$ is submodular on the (subsets signal lattice, discrete signal lattice, continuous signal lattice) respectively. The signals are *strict* substitutes if $\mathcal{V}^{u,P}$ is strictly submodular.

The signals $A_1, \ldots, A_n$ are *(weak, moderate, strong) complements* if $\mathcal{V}^{u,P}$ is supermodular on the (subsets signal lattice, discrete signal lattice, continuous signal lattice) respectively. The signals are *strict* substitutes if $\mathcal{V}^{u,P}$ is strictly supermodular.

We note that strong substitutes imply moderate substitutes, which imply weak substitutes.

As we will show in this work, moderate and strong substitutes are very strong notions of informational substitutes that are unlikely to be satisfied except in mostly trivial cases. Moreover, weak substitutes are easier to work with. Therefore, in this work, *when we refer to substitutes without qualification, we mean weak substitutes.* Similarly, when we refer to complements without qualification, we mean weak complements.

To demonstrate the intuition of this definition, we consider the case of independent signals. Independent signals should be complements because knowing the value of one signal does not give any information about any other signal by virtue of independence. Knowing more signals should therefore not decrease the marginal value of an additional signal. In other words, the value of information function is supermodular. This can be formalized in the following proposition, which is Proposition 5.2.1 from Chen and Waggonner (2016).

Proposition 2.10 (Chen and Waggonner (2016))**.** *Independent signals are strong complements in any decision problem where G has a jointly convex Bregman divergence $D_G(p, q)$.*

[1]The reader might note that the standard submodularity condition on a lattice is $f(x) + f(y) \geq f(x \vee y) + f(x \wedge y)$ for every x and y on the lattice. We can convert the standard submodularity condition to the diminishing marginal return definition as follows. We start with $f(x) + f(y) \geq f(x \vee y) + f(x \wedge y)$ or $f(x) - f(x \wedge y) \geq f(x \vee y) - f(y)$. Comparing this with the marginal value inequality gives $x = B \vee A', y = A, x \wedge y = A', x \vee y = B \wedge A$, so we insist that by joining and meeting expressions for x and y we do get the last two. The join is immediate: $x \vee y = (B \vee A') \vee A = B \vee (A' \vee A) = B \vee A$. The meet gives us the condition we want $x \wedge y = (B \vee A') \wedge A = (B \wedge A) \vee (A' \wedge A) = (B \wedge A) \vee A'$, so if this is equal A', then $A \wedge B \preceq A'$. The converse can be checked similarly, so the two definitions are equivalent.

2.3 Scoring Rules and Convex Functions

In this section, we show that every decision problem can be reduced to scoring rules, which in turn can be reduced to an associated convex function. Therefore, propositions about decision problems can be rewritten with convex functions. This reduction allows us to use many tools from the geometry of convex functions to analyze and design "good" decision problems/scoring rules.

In the following definition (McCarthy, 1956; Savage, 1971; Gneiting and Raftery, 2007), the forecaster predicts a probability distribution $\hat{q}$ over events of interest E. Once the event is realized $E = e$, the forecaster receives a score (or utility) $S(\hat{q}, e)$. Before event realization, at the time of forecast, the forecaster has some belief over the event, expressed as a probability distribution q. The forecaster wants to maximize her subjective expected score, with the expectation taken according to her subjective belief $\mathbb{E}_{e\sim q} S(\hat{q}, e)$. A proper scoring rule is such that $\hat{q} = q$ is a maximizer of the expected score, and a strictly proper scoring rule is such that the maximizer is unique. Therefore, (strictly) proper scoring rules encourage one-stage truthful predictions.

Definition 2.11 (Scoring Rule for Probability Distribution Prediction)**.** A *scoring rule* for an event E is a function $S : \Delta_E \times E \to \mathbb{R}$ so that $S(\hat{q}, e)$ is the score assigned to a prediction $\hat{q}$ when the true realized outcome is $E = e$. We use the standard notation $S(\hat{q}; q) = \mathbb{E}_{e\sim q} S(\hat{q}, e)$. The scoring rule is *(strictly) proper* if for all E, q, setting $\hat{q} = q$ (uniquely) maximizes the expected score $S(\hat{q}; q)$.

We then have the following "revelation principle" reductions from decision problems to scoring rules and convex functions, respectively. For proofs, see Theorem 2.3.1 and Corollary 2.3.1 of Chen and Waggonner (2016).

Proposition 2.12 (Chen and Waggonner (2016))**.** *For any decision problem u, there exists a proper scoring rule $S : \Delta_E \times E \to \mathbb{R}$ that is equivalent to the original decision problem in that for all information structures P and signals A, $\mathcal{V}^{S,P}(A) = \mathcal{V}^{u,P}(A)$*

Proposition 2.13 (Chen and Waggonner (2016))**.** *For any decision problem u there exists a corresponding convex function $G : \Delta_E \to \mathbb{R}$, and for every such G there exists a decision problem u such that $G(q)$ is the expected utility for acting optimally when the agent's posterior belief on E is q.*

Hence, for instance, $\mathcal{V}(A) = \mathbb{E}_a G(p_a)$ where p_a is the posterior on E given $A = a$.

Proposition 2.13 is our main computational tool. For example, we can recast the substitutes condition in terms of convex functions as follows.

Definition 2.14. For any decision problem, let G be the associated expected score function (which is convex). Signals are respectively *(weak, moderate, strong) substitutes* for that decision problem if and only if for all A', A, B on the (subsets, discrete, continuous) lattice with $A \wedge B \preceq A' \preceq A$,

$$\mathop{\mathbb{E}}_{a',b} G(p_{a'b}) - \mathop{\mathbb{E}}_{a'} G(p_{a'}) \geq \mathop{\mathbb{E}}_{a,b} G(p_{a,b}) - \mathop{\mathbb{E}}_{a} G(p_a). \tag{2.1}$$

The above inequality is reversed for complements.

Weak substitutes and complements correspond to set functions being submodular and supermodular. The following characterization for submodular set functions is more convenient. Let Ω be a finite set, called a ground set, a submodular function is a set function $f : 2^\Omega \to \mathbb{R}$, where 2^Ω is the power set of Ω, satisfying the following condition. For every $S \subseteq \Omega$ and $a, b \in \Omega \setminus S$, we have $f(S \cup \{a\}) + f(S \cup \{b\}) \geq f(S \cup \{a, b\}) + f(S)$. Therefore, we can recast the weak substitutes condition as follows.

Definition 2.15. For any decision problem, let G be the associated expected score function (which is convex). Signals $\mathcal{L} = \{A_1, \ldots, A_n\}$ are weak substitutes for that decision problem if and only if for all $S \subseteq \mathcal{L}$ and $A, B \in \mathcal{L} \setminus S$,

$$\mathop{\mathbb{E}}_{S,a} G(p_{Sa}) + \mathop{\mathbb{E}}_{S,b} G(p_{Sb}) \geq \mathop{\mathbb{E}}_{S,a,b} G(p_{Sab}) + \mathop{\mathbb{E}}_{S} G(p_S) \tag{2.2}$$

The above inequality is reversed for complements.

This definition is especially convenient when there are $n = 2$ signals, where the condition reduces to

$$\mathop{\mathbb{E}}_{a_1} G(p_{a_1}) + \mathop{\mathbb{E}}_{a_2} G(p_{a_2}) \geq \mathop{\mathbb{E}}_{a_1,a_2} G(p_{a_1 a_2}) + G(p) \tag{2.3}$$

for substitutes, and the inequality is reversed for complements.

2.4 Scoring Rules for Distribution Properties

A scoring rule for probability distribution prediction is often just called a scoring rule, and indeed this is what we will mean when we say a scoring rule without qualification. However, we make this distinction because we are also interested in a scoring rule when the forecaster provides not the entire distribution over outcomes, but a summary statistic (such as the mean) over events. This is useful when the aggregator is only interested in some property, such as mean or median, of the target event distribution, and not necessarily the entire distribution. This is also useful when the probability distribution over outcomes is large and it is unreasonable to expect forecasters to predict or even communicate the probability of every event, as is often the case when the event is combinatorial or continuous.

Lambert et al. (2008) first formalizes the above problem as follows. Given an outcome space Ω and an arbitrary map $\Gamma : \Delta_\Omega \to \mathbb{R}$, under what circumstances can we construct a proper scoring rule $s : \mathbb{R} \times \Omega \to \mathbb{R}$ for Γ, i.e. where,

$$\Gamma(P) \in \arg\min_{r \in \mathbb{R}} \mathop{\mathbb{E}}_{e \sim P} s[r](e) \tag{2.4}$$

for every $P \in \Delta_\Omega$?

In our prediction market context, we can ask a corresponding question. The forecaster now predicts the summary of the distribution of outcome of interest (which can be a scalar, or a vector of scalars), and not the entire probability distribution. The question becomes: given an information structure and a distribution statistics (or a *property*) to be predicted, can we design a market scoring rule such that signals are substitutes?

Abernethy and Frongillo (2012) characterizes all possible scoring rules for *linear properties*, defined as follows.

Definition 2.16. A *linear property* is given by the map $\mu \mapsto \int_\Omega \rho d\mu$ for some $\rho : \Omega \to U$. We call this property a ρ-linear property.

In other words, a linear property is a "weighted average" of all the points, where each point in the domain has a weight. The expectation of a distribution is a linear property with ρ an identity function $\rho(e) = e$.

Definition 2.17. The Bregman divergence of a function f with subderivative df is the function

$$D_{f,df}(x, y) = f(x) - f(y) - df_y \cdot (x - y) \tag{2.5}$$

When f is differentiable, and hence df is unique, we can simply write $D_f(x, y)$.

Definition 2.18. Given $\rho : \Omega \to U$, we can associate a Bregman score s to the triple (f, df, ρ) defined by

$$s[r](e) = -D_{f,df}(\rho(e), r) + f(\rho(e)) = f(r) + df_r \cdot (\rho(e) - r) \tag{2.6}$$

The main result of Abernethy and Frongillo (2012) is that every "nice" scoring rule that elicit a ρ-linear property is "equivalent" to some (f, df, ρ) Bregman score, where two scoring rules are equivalent if they differ by the amount that depends only on realized event values and not on the property of the distribution. By "nice" we mean some technical restrictions, such as the scoring rule being differentiable in some special space, or that the condition holds only in the relative interior. Such technicalities do not concern us here.

Proposition 2.19 (Abernethy and Frongillo (2012))**.** *A "nice" proper scoring rule s elicits a ρ-linear property if and only if it is equivalent to a Bregman score (f, df, ρ) for some convex f.*

This result reduces the problem on scoring rules (in the case of linear properties) to a problem over convex functions which we are equipped to solve. We can view this result as analogous to the classical result that reduces the search over scoring rules over probability distributions to convex functions given by the expected score.

This result is a corollary of Proposition 2.12 for the case of eliciting the expectation. We cannot find the exact statement of this result anywhere, but the result is not hard. Later, we will study information structures and scoring rules that are substitutes for eliciting expectation using this result.

Corollary 2.20. *A scoring rule elicits expectation r if and only if the expected score can be written in the form $f(r)$ where f is a convex function.*

Proof. Proposition 2.19 implies that the scoring rule can be written in the form $s[r](e) = f(r) + df_r(\rho(e) - r) = f(r) + df_r(e - r)$ because the expectation property has $\rho(e) = e$. Moreover, since r is the expectation, we must have $\mathbb{E}_e(e - r) = 0$, so $\mathbb{E}_e\, s[r](e) = f(r)$. Conversely, given a convex f, the scoring rule $s[r](e) = f(r) + df_r(e - r)$ elicits the expectation r by the converse of Proposition 2.19. □

Even though Corollary 2.20 says any convex function of the reported expectation can be a scoring rule, in practice there is a special scoring rule that is often used, which comes from minimizing weighted sum of squares. In one dimension, if the reported expectation is r and the realized outcome is e, the forecaster gets $s[r](e) = u(r, e) := -(r - e)^2$

To have $\mathcal{V}^{S,P}(A) = \mathcal{V}^{u,P}(A)$ we have $S(\hat{q}, e) = G(\hat{q}) + \langle G'(\hat{q}), \delta_e - \hat{q}\rangle$ with $S(\hat{q}, e) = u(d^*_{\hat{q}}, e)$ and $d^*_{\hat{q}} = \text{argmax}_{d \in \mathcal{D}} \mathbb{E}_{e\sim q}\, u(d, e)$. Since the property is just a decision, we have $S(\hat{q}, e) = u(r^*_{\hat{q}}, e)$ and $r^*_{\hat{q}} = \text{argmax}_r \mathbb{E}_{e\sim q}\, u(r, e)$. Also note that $G(q) = S(q; q) = \mathbb{E}_{e\sim q}\, S(q; e)$. It must be true that $G(q)$ is convex in q – and remember that q is a distribution not a number.

We can compute $G(q)$ when we want to elicit mean $u(r, e) = -(r - e)^2$. We have $r^*_q = \mathbb{E}\, q$ so $S(q, e) = u(\mathbb{E}\, q, e)$ and $G(q) = S(q; q) = \mathbb{E}_{e\sim q}\, S(q, e) = \mathbb{E}_{e\sim q} -(\mathbb{E}\, q - e)^2 = -\text{Var}(q)$

So G is the negative variance of the distribution, an elegant result.

We call this mean-eliciting scoring rule *canonical*. Note that this canonical scoring rule we just derived is a special case of Corollary 2.20 when $f(r) = r^2$, because the scoring rule then is $s[r](e) = f(r) + f'(r)(e-r) = r^2 + 2r(e-r) = e^2 - (r-e)^2$, which is equivalent to the negative square because it differs from $-(r-e)^2$ by e^2, independent of r.

More generally, we can write $s[r](e) = f(r) + df_r(e-r) = -D_f(e,r) + f(e)$, which is equivalent to $s[r](e) = -D_f(e,r)$. If a proper scoring rule elicits the mean, $r = \mathbb{E}\, q$, where q is the belief of the forecaster, so the expected score function is then $G(q) = -\mathbb{E}_{e\sim q} D_f(\mathbb{E}\, q, e)$.

A classical result tells us that the expected score function G must be convex. We can directly verify that a canonical score function is indeed convex as follows.

$$G\left(\frac{q_1+q_2}{2}\right) - \frac{G(q_1)+G(q_2)}{2} = -\mathrm{Var}\left(\frac{q_1+q_2}{2}\right) + \frac{\mathrm{Var}(q_1)+\mathrm{Var}(q_2)}{2} = \mathrm{Var}\left(\frac{q_1-q_2}{2}\right) \geq 0$$

We summarize our discussion in the following definition.

Definition 2.21. A scoring rule that elicits expectation is *canonical* if the expected score is $-\mathrm{Var}(q)$, where q is the belief of the forecaster. This is equivalent to the expected score of $(\mathbb{E}\, q)^2 = f(\mathbb{E}\, q)$, where $f(r) = r^2$ in the spirit of Corollary 2.20.

A canonical scoring rule also has the nice property that the Bregman divergence of two beliefs is canonical. To the best of our knowledge, this result is new. However, there is a well-known analogue: if $G(q) = ||q||_2^2$, then $D_G(p,q) = ||p-q||_2^2$.

Proposition 2.22. *If $G(q) = -\mathit{Var}(q)$ then $D_G(p,q) = -\mathit{Var}(p-q)$.*

Proof. Prove the statement on discrete distributions and pass to the limit. The rest is calculation. The details are relegated to Appendix A.1.1. □

In particular, Proposition 2.22 implies that with the canonical mean-eliciting scoring rule, the Bregman divergence $D_G(p,q)$ is jointly convex in p and q. Proposition 2.10 then implies that independent signals are strong complements.

Corollary 2.23. *Independent signals are strong complements under a canonical mean-eliciting scoring rule.*

Abernethy and Frongillo (2012) only handles linear properties, but there are other properties of distributions that we are interested in, such as the median. Note that the median only makes sense if the target distribution is one-dimensional. As is standard in statistics, the mean is elicited by the scoring rule $s[r](e) = -(r-e)^2$, which gives rise to the canonical mean-eliciting scoring rule. An analogue for median elicitation is $s[r](e) = -|r-e|$, and we also call this scoring rule canonical among the median-eliciting scoring rules. A median-eliciting canonical scoring rule as the expected score on $q = \mathcal{N}(\mu, \sigma^2)$ as $G(q) = \mathbb{E}_{e\sim\mathcal{N}(\mu,\sigma^2)} -|\mu - e| = -c\sigma$ for constant $c = \sqrt{2/\pi} > 0$. We summarize our discussion in the following definition.

Definition 2.24. A scoring rule that elicits the median is *canonical* if the scoring rule is $s[r](e) = -|r-e|$. If $q = \mathcal{N}(\mu, \sigma^2)$, then the expected score is $-c\sigma$, where $c = \sqrt{2/\pi} > 0$.

Chapter 3

Universal Substitutes and Complements

The informational substitutes condition takes into account the information structure and the scoring rule jointly. If the information structure is such that signals are substitutes (complements) for every scoring rule, we call that information structure *universal substitutes (complements)*. In this chapter, we characterize universal substitutes and complements.

Universal substitutes and complements are especially of interest because independence of signal substitutability from market settings removes much of the complexity incurred in thinking about substitutability of pieces of information compared to substitutability of items of goods; whether items are substitutes for an agent depends only on her valuation function over items and not on which market she is in. Universal substitutes also provide a very strong guarantee that the market can function as intended when the market designer has no control over the choice of scoring rule, or is interested in the robustness of such a choice. Unfortunately, Chen and Waggonner (2016) has shown that all universal substitutes are, in a precisely defined sense, "almost trivial." One interpretation of this result is as follows: in information markets, the information structure and scoring rules cannot be decoupled and must necessarily be considered jointly.

We are also interested in the opposite problem, that of *universal complements*. Characterizing universal complements is the main contribution of this chapter. Identifying universal complements is useful as an exercise in impossibility results. If we know that an information structure we are dealing with is in a class of universal complements, we know that information aggregation is worst-possible regardless of design. Under universal complements, market designers need not try to find a substitutability-inducing scoring rule because it does not exist. We show that if signals and the outcome are binary such that the outcome is the exclusive or (XOR) of signals, then the signals are universal complements. We also show that, for a binary outcome, if binary signals are universal complements, then the outcome must be the XOR of signals. Moreover, we show that for multi-valued signals to be universal complements "generically," they must collapse into binary signals. Thus, we characterize the class of universal complements.

We argue that universal moderate (and strong) substitutes and complements are too strong in the sense that only trivial information structures satisfy them. We prove this fact in the last section, Section 3.4. Henceforth, when we refer to universal substitutes or complements, we mean the *weak* notion. Section 3.1 develops technical machineries on linear convex function inequalities that we will use throughout this chapter. Section 3.2 discusses the universal substitutes result of Chen and Waggonner (2016). Section 3.3 characterizes universal complements, starting from the simpler case when the outcome is a deterministic function of signals (Subsection 3.3.1), then extending it to the general case (Subsection 3.3.2).

Definition 3.1 (Universal Substitutes and Complements)**.** Given an information structure $E, A_1, \ldots, A_n$ with prior P, the signals $A_1, \ldots, A_n$ are *universal weak substitutes* if they are weak substitutes for every decision problem. To make the dependence on E explicit, we sometimes say that $(A_1, \ldots, A_n; E)$ are universal weak substitutes. Universal (weak/moderate/strong) (substitutes/complements) are defined analogously.

Note that the universal substitutes property is a property of the *information structure* alone, that is, it is a property of the joint prior distribution $(A_1, \ldots, A_n; E)$.

We have already seen that any decision problem can be reduced to a convex function G and the inequalities that determine the substitutes and complements are linear in the function G. The universal substitutes and complements condition means the inequality holds for every convex function G. We will work with such inequalities extensively in this section, so before we proceed, we will develop some results on how to deal with them.

3.1 Linear Convex Function Inequalities

We first show that G can be scaled by a positive factor and by a linear combination of its coordinate inputs. This will allow us to normalize the values of G at the corners of the domain to 0, which aid computation.

Let E takes d values, then the function $G(q)$ has as its argument q a distribution over d values, that is, q is in a $(d-1)$-simplex $\Delta_{d-1} = \{(q_1, \ldots, q_d) \in \mathbb{R}^d_{\geq 0} : q_1 + \cdots + q_d = 1\}$.

For $i = 1, \ldots, d$, let $e_i \in \Delta_{d-1}$ be such that the ith coordinate is 1, and the rest are 0. In the following proposition, we show that we can scale G such that we can set $G(e_i) = 0$ for $i = 1, \ldots, d$ without loss of generality.

Proposition 3.2. *Let $\lambda > 0$ and $\alpha \in \mathbb{R}^d$ be constants. Let $G, \tilde{G} : \Delta_{d-1} \to \mathbb{R}$, be such that $\tilde{G}(q) = \lambda G(q) + \langle q, \alpha \rangle$ for every $q \in \Delta_{d-1}$, where $\langle \cdot, \cdot \rangle$ is a dot product, i.e., $\langle q, \alpha \rangle = \sum_{i=1}^{d} q_i \alpha_i$. then G induces substitutability if and only if $\tilde{G}$ does. This implies that we can set $G(e_i) = 0$ for $i = 1, \ldots, d$ without loss of generality.*

Proof. G induces substitutability if and only if for $A \wedge B \preceq A' \preceq A$,

$$\mathop{\mathbb{E}}_{a',b} G(p_{a',b}) - \mathop{\mathbb{E}}_{a'} G(p_{a'}) \geq \mathop{\mathbb{E}}_{a,b} G(p_{a,b}) - \mathop{\mathbb{E}}_{a} G(p_a)$$

and same for $\tilde{G}$. We show that $\mathbb{E}_{a,b}\, p_{a,b} = p$. This is true because for each outcome e, $\sum_{a,b} p(a,b)p(e|a,b) = \sum_{a,b} p(e,a,b) = p(e)$. Similarly, we can show that $\mathbb{E}_{a,b}\, p_{a,b} = \mathbb{E}_{a',b}\, p_{a',b} = \mathbb{E}_a\, p_a = \mathbb{E}_{a'}\, p_{a'} = p$. Therefore,

$$\begin{aligned}
\mathop{\mathbb{E}}_{a',b} \tilde{G}(p_{a',b}) - \mathop{\mathbb{E}}_{a'} \tilde{G}(p_{a'}) &= \lambda \left(\mathop{\mathbb{E}}_{a',b} G(p_{a',b} - \mathop{\mathbb{E}}_{a'} G(p_{a'}) \right) + \langle \mathop{\mathbb{E}}_{a',b} p_{a',b}, \alpha \rangle - \langle \mathop{\mathbb{E}}_{a'} p_{a'}, \alpha \rangle \\
&= \lambda \left(\mathop{\mathbb{E}}_{a',b} G(p_{a',b}) - \mathop{\mathbb{E}}_{a'} G(p_{a'}) \right)
\end{aligned}$$

Similarly,

$$\mathop{\mathbb{E}}_{a,b} \tilde{G}(p_{a,b}) - \mathop{\mathbb{E}}_{a} \tilde{G}(p_a) = \lambda \left(\mathop{\mathbb{E}}_{a,b} G(p_{a,b}) - \mathop{\mathbb{E}}_{a} G(p_a) \right)$$

Since $\lambda > 0$, substitutability for G is equivalent to substitutability for $\tilde{G}$.

Now, let G satisfies substitutability and for $i = 1, \ldots, d$, let $\alpha_i = -G(e_i)$. If $\tilde{G}(q) = G(q) + \langle \alpha, q \rangle$, then we just proved that $\tilde{G}$ must also satisfy substitutability and $\tilde{G}(e_i) = G(e_i) + \langle \alpha, e_i \rangle = G(e_i) + \alpha_i = 0$. Therefore, it is not without loss of generality to assume that $G(e_i) = 0$ for $i = 1, \ldots, d$.

□

The following result is classical.

Proposition 3.3. *If $S \subseteq \mathbb{R}^d$ is an open set, and $G : S \to \mathbb{R}$ is convex, then G is continuous on S.*

Often we will be working with a convex G on a domain that is a $(d-1)$-simplex $\Delta_{d-1} := \{(x_1, \ldots, x_d) : x_i \geq 0, x_1 + \cdots + x_d = 1\}$. G can only be discontinuous at the boundaries, but such functions are pathological and we want to rule them out. Therefore, from this point on we assume that G is continuous in the entire domain.

When the domain of G is 1-simplex, we can restrict it without loss of generality to a function of the first coordinate which is in $[0, 1]$. In other words, for $x \in [0, 1]$, we write $G(x, 1-x)$ as $G(x)$, and G is still convex in x.

We often have to deal with inequalities of the form

$$\sum_{i=1}^{n} \alpha_i G(x_i) \geq 0$$

where $\alpha_1, \ldots, \alpha_n, x_1, \ldots, x_n$ are known, and we want to know whether the above inequality holds for every convex function G. Since every constant function is convex, using the constant function 1 and -1 tells us that $\sum_{i=1}^{n} \alpha_i = 0$ is a necessary condition. This condition is satisfied in our application in universal substitutes and complements.

Proposition 3.4. *Given $\alpha_1, \ldots, \alpha_n, x_1, \ldots, x_n \in \mathbb{R}$ such that $\sum_{i=1}^{n} \alpha_i = 0$ and $0 \leq x_1, \ldots, x_n \leq 1$. We can check in time linear in n whether the inequality*

$$\sum_{i=1}^{n} \alpha_i G(x_i) \geq 0 \tag{3.1}$$

holds for every convex G such using Algorithm 1.

Proof. We can set $G(0) = G(1) = 0$ by Proposition 3.2.

Let $x_0 = 0, x_{n+1} = 1$, and $y_i = G(x_i)$ for $0 \leq i \leq n+1$; in particular, $y_0 = G(0) = 0, y_{n+1} = G(1) = 0$.

We will show by induction on i that the inequality holds for every convex G if it holds for a special G that is linear from $x_0 = 0$ to x_i and piecewise linear on $[x_j, x_{j+1}]$, $j \geq i$, following the transformations in Algorithm 1.

We first note that it is sufficient to consider a G that is piecewise linear on $[x_j, x_{j+1}]$ for every j, because only the values of G are present in the inequality in question. The base case $i = 1$ is evident because G is linear on $[x_0, x_1]$. Now, assume that G is linear on $[x_0, x_i]$. We will show that y_i can take any value within the upper and lower bound derived later. Given $(x_0, y_0), (x_{i+1}, y_{i+1}), (x_{i+2}, y_{i+2}), \ldots, (x_{n+1}, y_{n+1})$, the necessary and sufficient conditions for y_i such that G is convex are

$$\frac{y_i}{x_i} \leq \frac{y_{i+1} - x_{i+1}}{x_{i+1} - x_i} \text{ and } \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \leq \frac{y_{i+2} - y_{i+1}}{x_{i+2} - x_{i+1}}$$

which is equivalent to

$$\left(\frac{x_{i+2}-x_i}{x_{i+2}-x_{i+1}}\right)y_{i+1}-\left(\frac{x_{i+1}-x_i}{x_{i+2}-x_{i+1}}\right)y_{i+2}\leq\left(\frac{x_i}{x_{i+1}}\right)y_{i+1}$$

If $\alpha_i \geq 0$, then the expression $\sum_{j\geq i}\alpha_j y_j$ (the sum is over $j \geq i$ because the inductive assumption that G is linear on $[x_0, x_i]$ means that any y_j, $j \leq i$, can be written as a linear combination of y_0 and y_i and cleared away) is lower bounded by the lower bound of y_i, so we can replace the y_i with the lower bound expression. If this replacement makes the inequality holds, the fact that it is a lower bound means that the inequality always holds. Conversely, this lower bound can be achieved so the inequality must hold at the lower bound. Replacing y_i with the lower bound means adding $\frac{x_{i+2}-x_i}{x_{i+2}-x_{i+1}}$ to the coefficient α_{i+1} of x_{i+1} and subtracting $\frac{x_{i+1}-x_i}{x_{i+2}-x_{i+1}}$ from the coefficient α_{i+2} of x_{i+2}.

If $i = n-1$, then $x_{i+2} = 0$, not a variable, so the coefficient change has no effect. This corresponds to the update within the $\alpha_i > 0$ loop in Algorithm 1. If $\alpha_i < 0$, then analogously we can replace y_i with the upper bound, which corresponds to adding $\frac{x_i}{x_{i+1}}$ to the coefficient α_{i+1} of x_{i+1}. This also corresponds to the update within the $\alpha_i < 0$ loop in the algorithm.

Once we exhaust all of $i = 1, \ldots, n-1$, we reduce the inequality to $\alpha_n y_n \leq 0$, where y_n can take any non-positive value. This inequality holds if and only if $\alpha_n \leq 0$. This corresponds to the last part of the algorithm. □

Proposition 3.5. *Given $\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s \geq 0$ and $0 = x_0 < x_1 < \cdots < x_r < x_{r+1} = 1$ and $0 \leq x'_1, \ldots, x'_s \leq 1$. The inequality*

$$\sum_{i=1}^{r}\alpha_i G(x_i) \geq \sum_{j=1}^{s}\beta_j G(x'_j) \tag{3.2}$$

holds for every convex G if and only if it holds for a special G that is piecewise linear on each of the intervals $[x_i, x_{i+1}]$, $i = 0, \ldots, r$.

Proof. A line segment connecting two points on the graph of a convex function lies weakly above the graph of the function between that two points. □

Proposition 3.5 implies that we can replace each of the $G(x'_j)$ with a linear interpolation between $G(x_{i_j})$ and $G(x_{i_j+1})$ for $x'_j \in [x_{i_j}, x_{i_j+1}]$ to reduce the inequality to the ones that only have the terms $G(x_i)$. We can then appeal to Algorithm 1 to finish the check. If all the coefficients and function arguments are given, this does not save computation time. The computation time is still $O(r+s)$. However, it is a very powerful conceptual

Algorithm 1 Checking whether a linear inequality 3.1 holds for every convex G.

Precondition: $\sum_{i=1}^{n} \alpha_i = 0$ and $0 = x_0 < x_1 < \cdots < x_n < x_{n+1} = 1$

function LinearConvexIneq$(\alpha_1, \ldots, \alpha_n, x_1, \ldots, x_n)$
  for $i \leftarrow 1$ to $n-1$ **do**
    if $\alpha_i > 0$ **then**
      $\alpha_{i+1} \leftarrow \alpha_{i+1} + \frac{x_{i+2}-x_i}{x_{i+2}-x_{i+1}} \alpha_i$
      if $i < n-1$ **then**
        $\alpha_{i+2} \leftarrow \alpha_{i+2} - \frac{x_{i+1}-x_i}{x_{i+2}-x_{i+1}} \alpha_i$
      end if
    end if
    if $\alpha_i < 0$ **then**
      $\alpha_{i+1} \leftarrow \alpha_{i+1} + \frac{x_i}{x_{i+1}} \alpha_i$
    end if
    $\alpha_i \leftarrow 0$
  end for
  if $\alpha_n \leq 0$ **then**
    return true
  else
    return false
  end if
end function

tool when we want to prove that a family of information structures satisfy universal substitutes or complements.

Proposition 3.5 has the following very useful corollary.

Corollary 3.6. *If* $\alpha_1, \ldots, \alpha_n \geq 0$ *and* $0 \leq x, x_1, x_2, \ldots, x_k \leq 1$, *the inequality*

$$G(x) \geq \sum_{i=1}^{k} \alpha_i G(x_i) \tag{3.3}$$

holds for every convex G *if and only if*

$$\sum_{i:x_i \leq x} \alpha_i \frac{x_i}{x} + \sum_{i:x_i \geq x} \alpha_i \frac{1-x_i}{1-x} \geq 1 \tag{3.4}$$

3.2 Universal Substitutes

Chen and Waggonner (2016) essentially already solved the universal substitutes problem in Section 5 of their paper.

Definition 3.7. Given an information structure $E, A_1, \ldots, A_n$ with prior P, the signals are *trivial substitutes* if for every realization $a_1, \ldots, a_n$ of $A_1, \ldots, A_n$ in the prior's support, $p_{a_i} = p_{a_1,\ldots,a_n}$ for all i. The signals are *trivial complements* if every realization

$a_1, \ldots, a_n$ of $A_1, \ldots, A_n$ in the prior's support, $p_{\{a_j : j \neq i\}} = p$ for all i. We term them *somewhat trivial* if the prior is a mixture distribution that is equal to a trivial structure with some probability, and some other arbitrary other structure with the remaining probability.

The following proposition is Proposition 5.1.1 of Chen and Waggonner (2016).

Proposition 3.8 (Chen and Waggonner (2016))**.** *If $(A_1, \ldots, A_n; E)$ are universal weak substitutes, then they are somewhat trivial. Furthermore, their "trivial" component is more informative than the nontrivial component, in the following sense. Let $X_i \subseteq \Delta_E$ be the convex hull of $\{p_{a_i} : a_i \in A_i\}$, and let $Y \subseteq \Delta_E$ be the convex hull of $\{p_{a_1,\ldots,a_n} : a_1 \in A_1, \ldots, a_n \in A_n\}$. If $(A_1, \ldots, A_n; E)$ are universal substitutes, then $X_i = Y$ for all i.*

3.3 Universal Complements

We characterize universal complements in this section. An inspection of the inequalities that determine complements (Definition 2.15) shows that the key variables are the posterior probabilities of the outcome E conditional on subsets of signals. Therefore, the math is simpler in the case where the outcome E is a deterministic function of signals $A_1, \ldots, A_n$; in other words, signals completely determine the outcome. This is an important special case not only because it is mathematically simple, but also because it represents an ideal case where someone who knows all the signals can predict the outcome with certainty.

We will study this setup in Subsection 3.3.1. We will obtain a clean characterization of universal complements when signals determine the outcome: all signals must "collapse" into binary signals, and the outcome is an exclusive-or (XOR) of a subset of signals. We then extend this characterization to general information structures in Subsection 3.3.2.

To rule out degenerate cases, we assume that every signal is *nontrivial.* Note that this is slightly weaker than the distinguishability criterion common in previous works (Chen et al., 2010; Ostrovsky, 2012; Gao et al., 2013; Chen and Waggonner, 2016).

Definition 3.9. Given an information structure $(A_1, \ldots, A_n; E)$ with prior P, a signal A_i is *nontrivial* if A_i takes at least two values, and for any two different realizations a_i and a'_i of A_i, there exists a realization a_{-i} of $A_{-i} = \{A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_n\}$ such that the two posterior distributions $E|A_i = a_i, A_{-i} = a_{-i}$ and $E|A_i = a'_i, A_{-i} = a_{-i}$ are different.

We will assume in this section that all signals are nontrivial. This is without loss of generality because if two realizations a_i and a_i' of A_i give the same posterior distribution of the outcome every any realization a_{-i} of A_{-i}, then for the purpose of predicting outcome, agent i can treat a_i and a_i' as the same signal. If the signal A_i takes only one value, then the value of this signal is not informative about the outcome, so the signal is trivial and agent i only knows the common prior.

3.3.1 When Signals Completely Determine The Outcome

The main theorem in this section is the following characterization of universal complements as XOR of boolean signals.

Theorem 3.10. *If $(A_1, \ldots, A_n; E)$ are universal complements, all signals are nontrivial and binary, and all prior probabilities are positive, then either $E = A_1 \oplus A_2 \oplus \cdots \oplus A_n$ or $E = \neg(A_1 \oplus A_2 \oplus \cdots \oplus A_n)$. If we assume that $(A_1, \ldots, A_n; E)$ are universal complements for every set of positive prior probabilities, then nontrivial signals must be binary, so the binary signals assumption can be dropped.*

Proof. We give a proof sketch here. See Appendix A.2.1 for the full proof.

The key technical tools of the proof are linear convex function inequalities that we develop in Section 3.1.

The core of this result is when $n = 2$ and E is binary. In such a case, the substitutability condition in Definition 2.15 has only one inequality which is linear in $G(\cdot)$ at various conditional probabilities that can be computed. For each $(A_1, A_2) \in \{0, 1\} \times \{0, 1\}$, there are two possible values of E, as E is a function of A_1 and A_2. Each function gives rise to a different inequality. Then, Corollary 3.6 allows us to convert this inequality that holds for every convex G to an inequality in scalars, which can be readily checked case by case. By exploiting the symmetry of the problem, there are only very few cases to consider. Only the XOR passes the test.

We can use the $n = 2$ result to extend to E binary and general n by induction on n, considering the $n - 1$ signals as a block that has to be XOR by inductive hypothesis. By $n = 2$ result, that block also has to XOR with the nth signal.

Now we relax the binary signals assumption but assume that signals are complements for every set of positive prior probabilities. By setting prior probabilities of all but two signals to near zero carefully, we can invoke continuity to derive a contradiction or reduce back to the case that E is binary. To show that each signal has to be binary, assume that one signal A_1 takes at least 3 values. Fixing all signals but A_1 and A_2 and consider the value of E for each value of $(A_1, A_2) \in \{0, 1, 2\} \times \{0, 1\}$. Case analysis reveals a contradiction. $\square$

The assumption that all prior probabilities are positive is essential. Let $n = 2$, A_1, A_2 are binary with $\mathbb{P}(A_1 = a_1, A_2 = a_2) = \pi_{a_1 a_2}$ for $a_1, a_2 \in \{0, 1\}$. If $\pi_{00} = 0$ then the signals under $E = \mathbf{1}[A_1 = 1, A_2 = 1] = A_1 \wedge A_2$ are in fact universal substitutes. This can be proved by bounding the right side with the help of Corollary 3.6 to get the coefficient

$$\geq (\pi_{10} + \pi_{11}) \frac{1 - \frac{\pi_{11}}{\pi_{10}+\pi_{11}}}{1 - \pi_{11}} + (\pi_{01} + \pi_{11}) \frac{1 - \frac{\pi_{11}}{\pi_{01}+\pi_{11}}}{1 - \pi_{11}} = \frac{\pi_{10}}{1 - \pi_{11}} + \frac{\pi_{01}}{1 - \pi_{11}} = 1$$

and inequality (3.4) is satisfied. Similarly, if $\pi_{01} = 0$, then $E = \mathbf{1}[A_1 = 1, A_2 = 0]$ is a universal complement, and so on. Case analysis shows that these are the only other possibilities apart from the XOR. These approaches can be extended to general n as well, depending on which prior probabilities are zero, by carefully conditioning on subsets of signals just like the proof.

The theorem immediately implies the following corollary by letting S be a subset of nontrivial signals.

Corollary 3.11. *If $(A_1, \ldots, A_n; E)$ are universal complements, all signals are binary, and all prior probabilities are positive, then there exists $S \subseteq \{1, 2, \ldots, n\}$ such that $E = \oplus_{i \in S} A_i$ or $E = \neg \oplus_{i \in S} A_i$. If we assume that $(A_1, \ldots, A_n; E)$ are universal complements for every set of positive prior probabilities, then nontrivial signals must be binary, so the binary signals assumption can be dropped.*

3.3.2 When Signals Do Not Completely Determine The Outcome

In Subsection 3.3.1 we show that in universal complements, if the outcome is a deterministic function of signals, then signals must be binary and the outcome must be the XOR of signals. We now show that if the randomization in the outcome E is arbitrary, then the outcome must either be the XOR of all signals or its negation.

Theorem 3.12. *Let $A = (A_1, \ldots, A_n)$ be nontrivial signals. Let $f_1(A), \ldots, f_k(A)$ be deterministic functions of the signals, and the outcome E conditional on A is defined by $E = f_i(A)$ with probability $\alpha_i > 0$, for $1 \leq i \leq k$, where $\alpha = (\alpha_1, \ldots, \alpha_k)$ is such that $\sum_{i=1}^{k} \alpha_i = 1$. Then, $(A; E)$ are universal complements for every α if and only if all signals and the outcome are binary and either $k = 1$ or 2. Moreover, if we denote $A_1 \oplus \cdots \oplus A_n$ by $\oplus A$, then f_i must be of the following form. If $k = 1$, then $E = f_1(A)$ is $\oplus A$ or $\neg(\oplus A)$. If $k = 2$, then $\{f_1(A), f_2(A)\} = \{\oplus A, \neg(\oplus A)\}$.*

Proof. For each i, consider the α such that $\alpha_j \to 0$ for $j \neq i$, and $\alpha_i \to 1$, then by continuity of G, $(A; f_i(A))$ are universal complements. By Theorem 3.10, $f_i(A) \in \{\oplus A, \neg(\oplus A)\}$. Since this holds for every i, and the f_i are distinct, k must be 1 or 2, and f_i must be of the described form. If $k = 1$, then Theorem 3.10 implies the converse for $k = 1$ that $\oplus A$ and $\neg(\oplus A)$ are universal complements. Now we need to show the converse for $k = 2$ that if $E|A$ is $\oplus A$ with probability α and $\neg(\oplus A)$ with probability $1 - \alpha$, then $(A; E)$ are universal complements. It suffices to check the case $n = 2$. The general n follows by induction analogously to the last part of Theorem 3.10.

By symmetry, assume $0 \leq \alpha \leq \frac{1}{2}$. The inequality to be proved is

$$
\begin{aligned}
&G\left((\pi_{01} + \pi_{10})(1-\alpha) + (\pi_{00} + \pi_{11})\alpha\right) + (\pi_{00} + \pi_{11})G(\alpha) + (\pi_{01} + \pi_{10})G(1-\alpha) \\
&\geq (\pi_{00} + \pi_{01})G\left(\frac{\pi_{01}(1-\alpha) + \pi_{00}\alpha}{\pi_{00} + \pi_{01}}\right) + (\pi_{10} + \pi_{11})G\left(\frac{\pi_{10}(1-\alpha) + \pi_{11}\alpha}{\pi_{10} + \pi_{11}}\right) \\
&+ (\pi_{00} + \pi_{10})G\left(\frac{\pi_{10}(1-\alpha) + \pi_{00}\alpha}{\pi_{00} + \pi_{10}}\right) + (\pi_{01} + \pi_{11})G\left(\frac{\pi_{01}(1-\alpha) + \pi_{11}\alpha}{\pi_{01} + \pi_{11}}\right)
\end{aligned}
$$

Let $\tilde{G}(x) = G((1-\alpha)x + \alpha(1-x))$ which is convex since $1 - 2\alpha \geq 0$. The inequality to be proved is equivalent to

$$
\begin{aligned}
&\tilde{G}(\pi_{01} + \pi_{10}) + (\pi_{00} + \pi_{11})\tilde{G}(0) + (\pi_{01} + \pi_{10})\tilde{G}(1) \\
&\geq (\pi_{00} + \pi_{01})\tilde{G}\left(\frac{\pi_{01}}{\pi_{00} + \pi_{01}}\right) + (\pi_{10} + \pi_{11})\tilde{G}\left(\frac{\pi_{10}}{\pi_{10} + \pi_{11}}\right) \\
&+ (\pi_{00} + \pi_{10})\tilde{G}\left(\frac{\pi_{10}}{\pi_{00} + \pi_{10}}\right) + (\pi_{01} + \pi_{11})\tilde{G}\left(\frac{\pi_{01}}{\pi_{01} + \pi_{11}}\right)
\end{aligned}
$$

This is the exact same inequality that we proved for deterministic XOR in Theorem 3.10 and we are done. □

3.4 Universal Moderate Complements Are Trivial

Throughout this work we deal almost exclusively with weak substitutes and complements. In the introduction, we argue that the moderate and strong notions of substitutes and complements are too strong to be satisfied except in the trivial cases.

Chen and Waggonner (2016) concludes that all universal moderate substitutes are trivial in their Theorem 5.1.1.

Proposition 3.13 (Chen and Waggonner (2016)). *All universal moderate substitutes are trivial. (Hence, the same holds for universal strong substitutes.)*

We also claim that all universal moderate complements are also trivial. Hence, the same holds for universal strong complements. The key is that under moderate complements, the inequalities have to be satisfied for any admissible information partition, and the set of all admissible partitions is too rich for one information structure to satisfy them all.

Theorem 3.14. *If the signals $A_1, \ldots, A_n$ are signals that each take a finite number of possible values and $E = f(A_1, \ldots, A_n)$ is a deterministic function of $A_1, \ldots, A_n$, then if $(A_1, \ldots, A_n; E)$ are universal moderate complements, then f is a constant function.*

Proof. We give a proof sketch here. See Appendix A.2.2 for the full proof.

We first prove the case $n = 2$ and f is boolean. Each function f corresponds to one of the 16 possible values of $(a_{00}, a_{01}, a_{10}, a_{11})$. For notational convenience list the value of E when $(A_1, A_2) = (0,0), (0,1), (1,0), (1,1)$ consecutively; for example, 0001 represents the information structure where $E = 1$ if $A_1 = A_2 = 1$ and 0 otherwise. Write $\pi_{a_1 a_2}$ the prior probability that $(A_1, A_2) = (a_1, a_2)$ and $a_{a_1 a_2} := f(A_1 = a_1, A_2 = a_2)$.

Consider the following partition

$$\begin{aligned} A' &= \{\{(0,0), (0,1), (1,0), (1,1)\}\} \\ A &= \{\{(0,0)\}, \{(0,1), (1,0), (1,1)\}\} \\ B &= \{\{(0,1)\}, \{(0,0), (1,0), (1,1)\}\} \end{aligned}$$

We can show that 0100, 0101, 0110, 0111 are not permissible by choosing an appropriate counterexample G that violates the inequality. We will assume throughout that $G(0) = G(1) = 0$. For each of 0100 and 0111, the inequality is Jensen's inequality with the wrong sign, so it is false. For 0101 and 0110, there are two Gs on both the left and the right. We can find a contradiction by first selecting a G such that the two Gs on the left are equal. The left side reduces to one G, so we can apply Corollary 3.6 to reduce it to an inequality of scalars, which can be checked to be false. A more methodical approach, which requires more case analysis, is to apply Algorithm 1 directly to reduce a linear inequality of Gs to an inequality of scalars.

By symmetry, we can rule out all other f's except 0000 and 0001, or their symmetric equivalents. 0001 can also be ruled out by another information structure, so only 0000, which corresponds to a constant function, remains.

Extending the case $n = 2$ to general n, f boolean, is straightforward. Assume that it is true up to $n - 1$, then a function $f(A_1, \ldots, A_n)$ of n variables must be constant when you fix the A_n, so $f(A_1, \ldots, A_n)$ depends only on A_n, but similarly it depends only on A_{n-1} so it depends on neither, that is, a constant function.

Now we extend this to $E = f(A_1, \ldots, A_n)$ a general function. E takes a finite number of values because there are finitely many possible inputs $A_1, \ldots, A_n$. If E takes 2^k values, then we can write $E = (E_1, \ldots, E_k)$. We must have $E_j = f_j(A_1, \ldots, A_n)$ a deterministic function of A_i, and $(A_1, \ldots, A_n; E_j)$ are universal complements, so E_j is a constant for all j, so E is a constant.

We can further extend this result to signals A_i that take multiple finite number of values. We can replace any A_i with $A_{i1}, \ldots, A_{it_i}$ with A_{ij} binary, say, map any possible value of A_i into binary numbers from 0 up to a finite number less than 2^{t_i}, then A_{ij} is the jth digit of the binary representation of A_i. We are done. $\square$

Chapter 4

Designing to Create Substitutability

In this chapter, we are the market designer. The information structure is intrinsic to the market; the market designer has no control over the information structure. However, the market designer has control over the decision problem, or equivalently, the scoring rule or the associated convex function G that is the expected score. (See Chapter 2 for an extended discussion of this point.) We want to decide whether a convex G exists such that the signals are substitutes. If such a convex G exists, we want to be able to design one efficiently. The market mechanics of predicting probability distributions and predicting properties are different, so we will consider the two setups separately in this chapter.

The main technical tool to design such a G is linear programming. We first observe that the set of inequalities that determine substitutability from Definition 2.15

$$\mathop{\mathbb{E}}_{S,a} G(p_{Sa}) + \mathop{\mathbb{E}}_{S,b} G(p_{Sb}) \geq \mathop{\mathbb{E}}_{S,a,b} G(p_{Sab}) + \mathop{\mathbb{E}}_{S} G(p_S) \tag{4.1}$$

for $S \subseteq \{A_1, \ldots, A_n\} \equiv \mathcal{L}$ and $A, B \in \mathcal{L} \setminus S$, are all linear when we view each function $G(\cdot)$ as a variable. The arguments of G that are present in the above set of inequalities are precisely all possible posterior distributions of the outcome conditional on a subset of signals. Therefore, the only values of G that matter in our application are the values at all posterior distributions of outcomes, and conversely, if we can specify the values of G at each posterior distribution that satisfy the set of inequalities while maintaining the convexity of G, then we have successfully found the desired G. We need to enforce the condition that G is convex. We also need to enforce that G is nontrivial; a constant G trivially satisfies the inequality but is uninteresting. (If the trader gets the exact same score independent of the report, while it is weakly optimal to report the true belief, it is also optimal to report any belief whatsoever.) It turns out that both convexity and nontriviality conditions are also linear. We can, therefore, view each $G(\cdot)$ as a variable in the linear program, and solve the linear program. The solution of the linear program gives the value of G at desired points, and we can take those values as our G, or if the linear program has no solution, then we can conclude that no such G exists, since the two problems are equivalent.

There exist efficient algorithms that solve linear programs in polynomial time in the size of the input (Khachiyan, 1980; Karmarkar, 1984). The catch is that there might be an exponential number of posterior distribution values, i.e. the arguments to G, and thus an exponential number of variables.

Before we proceed, we need to make sure that the information structure under consideration has a *compact representation*, that is, it can be described in polynomial time. This is important because if we cannot even write down the information structure in polynomial time, then we can not give the information structure as a batch input to the algorithm in polynomial time; the input is too unwieldy for any practical application.[1] This constraint means we need to add structure or symmetry into the information structure in question. For example, n binary signals without any restriction can collectively have 2^n possible values, each with its own prior probability specified as a number, so 2^n numbers need to be written down, which is unfeasible. However, if we assume additional structure, such as "each signal is independently drawn from a specified distribution" then we have a compact representation. Some such information structures are standard in applications in computer science and economics; we enumerate and explain them in Section 4.1.

In Section 4.2, we then explain in detail how to set up a linear program to find an appropriate convex G that makes the signals substitutes, or show that such a G does not exist. The main idea is that all inequalities required for substitutability (Definition 2.15) are linear in $G(\cdot)$, evaluated at various posterior (conditional) probabilities that can be calculated. Additional assumptions on convexity and nontriviality of G are also linear in $G(\cdot)$. Therefore, given a full access to any information structure, there is an algorithm that produces a substitutability-inducing convex G or a certificate of non-existence of such G. The computational cost of the algorithm is polynomial in the size of the linear program. If there are polynomially many constraints, and $G(\cdot)$ is evaluated at polynomially many points, the problem of finding a G is equivalent to a polynomially sized linear program which can be solved in polynomial time. We then show that information structures presented in Section 4.1 fit into this regime. Lastly, we briefly consider the problem of designing interpretable G functions.

[1] We discuss other possible models for information structure query in Subsection 5.1.1.

In Section 4.3, we turn to the case where the outcome and signals have continuous probability distributions. Unlike discrete distributions considered in Section 4.2, traders can no longer report the density of their beliefs at every point because there are infinitely many such points. However, if traders' beliefs follow a known distribution (say, a gaussian), they can just report parameters of the distribution (mean and variance in the gaussian case). We focus in particular on gaussian (normal) distributions for three reasons. First, they are commonly used in modeling noisy signals. Second, posterior distributions of gaussians, which we need to check substitutability, are tractable. Third, the notion of signal correlation is most natural in the multivariate gaussian setting. We also focus on a particular parametrized class of scoring rules. This class covers most of the scoring rules used in practice.

Even within this class of scoring rules and gaussian signals, we show that substitutability of signals is quite unintuitive. Chen and Waggonner (2016) shows that conditionally independent signals are substitutes if the Bregman divergence of the expected score function G is jointly convex. This result also has an intuitive appeal; conditionally independent signals "should" be substitutes because each signal can be thought of as a noisy measurement of the true value, so knowing one signal should make the marginal signal less valuable. However, we show in Subsection 4.3.1 that conditionally independent gaussian signals are complements for a wide range of parameters in the scoring rule.

Next, in subsection 4.3.2, we analyze the effect of correlation of gaussian signals on substitutability. We show that there is a sharp threshold that the signals are substitutes if the correlation is above the threshold, and complements otherwise. This is intuitive; lower correlation makes the signals closer to independent, and higher correlation makes the signals closer to conditionally independent. We can therefore think of this result as a continuous interpolation between the two regimes. Lastly, as an aside, in subsection 4.3.3, we show how to derive new matrix inequalities from the current setup.

In Section 4.4, we consider the problem of predicting a *property* of the target distribution, such as mean or median, rather than the entire distribution. We single out commonly used scoring rules for eliciting mean and median and call them canonical. We then show that conditionally independent gaussian signals are substitutes, while independent gaussian signals are complements, under the canonical mean-eliciting and median-eliciting scoring rules. The reasons for focusing on gaussian signals are similar to those provided in Section 4.3. We show that the gaussian assumption is robust to other bell-shaped distributions; by numerical simulations, the same substitutability holds when we replace the gaussian distributions by t-distributions. However, it is not true that conditionally independent distributions are substitutes under canonical scoring rules, and we give some qualitative properties when they should be substitutes. Lastly, we consider a parametrized non-canonical scoring rules and characterize substitutability of conditionally gaussian signals. Similar to the findings in Subsection 4.3.1, we find that, surprisingly, conditionally independent gaussian signals are complements for a wide range of parameters in the scoring rule.

4.1 Common Information Structures

In this section, we list some common information structures that are used in economics and computer science and explain their motivations and variations.

4.1.1 Independent Signals

In this information structure, all signals $A_1, \dots, A_n$ are independent, and a distribution $E|A_1, \dots, A_n$ along side each distribution of A_i are specified.

This information structure is most appropriate when each person i receives a different piece of information, and all signals must be combined to get the outcome. For example, suppose E is the value of a house. The value of a house can be decomposed into the values of the concrete structures that make up house itself (A_1), the land on which the house is built (A_2) and idiosyncratic noise (such as consumer sentiment.) Agent 1, the concrete manufacturing company, gets signal A_1 and agent 2, the real estate company, gets signal A_2, each drawn from an independent distribution. The value of the outcome/house E is a randomized function of A_1 and A_2.

In the discrete case, we can assume that signal A_i can take m_i values $1, 2, \dots, m_i$ and $\mathbb{P}(A_i = a) = \alpha_{i,a}$ for $1 \le a \le m_i, 1 \le i \le n$. E takes d values $1, 2, \dots, d$ and the conditional distribution $\mathbb{P}(E = e|A_1 = a_1, \dots, A_n = a_n) = \beta_{e,a_1,\dots,a_n}$ for $1 \le e \le d$ and $1 \le a_i \le m_i$. The numbers $\alpha_{i,a}$ and $\beta_{e,a_1,\dots,a_n}$ need to be specified. This setup is unwieldy. Without additional restriction on the β's, we need to specify $dm_1 \dots m_n$ values of β's, which are exponential in n because $m_i \ge 2$. There is no canonical way to simplify this model.

We add symmetry to the above model to get *independent exchangeable signals with conditionally symmetric event*. Signals $A_1, \dots, A_n$ are independent and identically distributed, each taking s values $1, 2, \dots, s$ with prior probabilities $\alpha_1, \alpha_2, \dots, \alpha_s$. E takes d values $1, 2, \dots, d$. We are given the distribution of $E|A_1, \dots, A_n$ as follows. Assume the conditional distribution of E is symmetric in the signals, and that if there are $n_1, n_2, \dots, n_s$ of $A_1, \dots, A_n$ that take value $1, 2, \dots, s$ respectively, then the probability of $E = e$ given $A_1, \dots, A_n$ is $\beta_{e,(n_1,\dots,n_s)}$. We think of d and s as constants independent of n. To describe this model, d and s and each α_e and each $\beta_{e,(n_1,\ldots,n_s)}$. The number of β's is the number of nonnegative s-tuples $(n_1, \ldots, n_s)$ such that $n_1 + \cdots + n_s = n$, which are $\binom{n+s-1}{s} = O(n^s)$ so this model can be described in polynomial time. Moreover, this shows that s *must* not grow with n, otherwise the description of the model will no longer be polynomial in n. The assumption that d is constant corresponds to the empirical reality that in many prediction markets, the number of outcomes are small and do not grow with the size of the market. For example, in a market for election results, there are a small number d of candidates who might win.

The independent signals model is much more commonly expressed (especially in information economics) in the continuous setting. $\mu_1, \ldots, \mu_n$ and $\sigma_1^2, \ldots, \sigma_n^2 > 0$ are given such that A_i is independently drawn from a normal distribution $A_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. $\mu_\epsilon, \sigma_\epsilon^2 > 0$ are also given so that the idiosyncratic noise $\epsilon \sim \mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2)$. The outcome is modeled as a sum of components $E = A_1 + \cdots + A_n + \epsilon$.

Note that ϵ does not have to actually be idiosyncratic; it might simply contain additional information that are not in any of the signals and can thus be treated as probabilistic by every agent. All the μs can also be normalized to all be 0 because they are known constants and thus we can consider $A_1 - \mu_1, \ldots, A_n - \mu_n, \epsilon - \mu_\epsilon$ in place of $A_1, \ldots, A_n, \epsilon$ without changing the information structure. Therefore, we set $\mu_1 = \cdots = \mu_n = \mu_\epsilon = 0$

This setting is especially convenient because not only are all the signals and the outcome normally distributed, but any posterior distribution of outcome conditional on a subset of signals is also normally distributed, and its distribution can be computed as

$$E \Big| (A_i = a_i \text{ for } i \in S) \sim \mathcal{N}\left(\sum_{i \in S} a_i, \sigma_\epsilon^2 + \sum_{i \notin S} \sigma_i^2\right) \tag{4.2}$$

To describe the continuous model, we need to specify $n+1$ numbers $\sigma_1^2, \ldots, \sigma_n^2, \sigma_\epsilon^2$.

4.1.2 Conditionally Independent Signals

In this information structure, $A_1, A_2, \ldots, A_n$ are conditionally independent given E, and the distribution of $A_i|E$ is specified for each E.

This information structure is appropriate if there is a true value E and each agent i observes an independent noisy estimate A_i of E. For example, there is some amount of gold E in an unexplored gold mine. Each company i hires a private surveyor to get an independent estimate A_i of the amount of gold in the mine. (In fact, often in the economics literature, the model of this form is called the "mineral rights model.")

In the discrete case, we can assume that the outcome E takes d values $1, 2, \ldots, d$ with probabilities $\mathbb{P}(E = e) = \alpha_e$. For each i, $A_i|E$ takes m_i values $1, 2, \ldots, m_i$ with probabilities $\mathbb{P}(A_i = a|E = e) = \beta_{i,a,e}$. The numbers α_e and $\beta_{i,a,e}$ need to be specified. There are $(1 + n\sum_{i=1}^{n} m_i)\, d$ variables to specify, so the model is already polynomially described. Nevertheless, it still is rather unwieldy and asymmetric.

We can add symmetry to the above model to get the *conditionally independent and exchangeable signals*. The outcome E takes d values $1, 2, \ldots, d$ with probabilities $\mathbb{P}(E = e) = \alpha_e$. Signals $A_1, \ldots, A_n$ each take s values $1, \ldots, s$ and are independent and identically distributed conditional on any realization of $E = e$ with probabilities $\mathbb{P}(A_i = a|E = e) = \beta_{a,e}$. By Bayes' rule,

$$\mathbb{P}(E = e|A_i = a_i \text{ for } i \in S) = \frac{\alpha_e \prod_{i\in S} \beta_{a_i,e}}{\sum_{e'} \alpha_{e'} \prod_{i\in S} \beta_{a_i,e'}} \tag{4.3}$$

If $|S| = k$, and among $(A_i)_{i\in S}$, there are u_j signals with value j, $1 \leq j \leq s$, $\sum_{j=1}^{s} u_j = k$, then the above equation can be rewritten as

$$\mathbb{P}(E = e|A_i = a_i \text{ for } i \in S) = \frac{\alpha_e \prod_{j=1}^{s} \beta_{j,e}^{u_j}}{\sum_{e'} \alpha_{e'} \prod_{j=1}^{s} \beta_{j,e'}^{u_j}} \tag{4.4}$$

This model is especially important because it has a polynomial linear program solution.

The conditionally independent model is also much more commonly expressed in information economics in the continuous setting. $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2, \sigma_E^2$ are given such that $E \sim \mathcal{N}(0, \sigma_E^2)$ and $A_i|E = e \sim \mathcal{N}(e, \sigma_i^2)$ are conditionally independent.

Any posterior distribution of outcome conditional on a subset of signals is also normally distributed, and its distribution can be computed because $(A_1, \ldots, A_n, E)$ are jointly multivariate normal and a subvector of a multivariate normal conditional on another subvector is multivariate normal. To derive the analytical formulas, we use the following classical conditional result. If $y \sim \mathcal{N}(\mu, \Sigma)$ is a multivariate normal random variable that can be partitioned as

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right) \tag{4.5}$$

then $y_1|y_2 \sim \mathcal{N}(\bar{\mu}, \bar{\Sigma})$ with mean and covariance

$$\bar{\mu} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2) \tag{4.6}$$

$$\bar{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \tag{4.7}$$

In this case, we get

$$E|(A_k)_{k\in S} = \mathcal{N}\left(\frac{\sum_{k\in S} A_k\sigma_k^{-2}}{\sigma_E^{-2}+\sum_{k\in S}\sigma_k^{-2}}, \frac{1}{\sigma_E^{-2}+\sum_{k\in S}\sigma_k^{-2}}\right) \tag{4.8}$$

4.1.3 Dependent Signals

There are information structures that do not fall either in the catagories of independent or conditionally independent signals, because all we need to do is to specify the joint distribution $(A_1, \ldots, A_n, E)$. Since conditional distributions play a big role, a multivariate gaussian is often the only tractable choice. We can choose a $(n+1)\times(n+1)$ positive definite matrix Σ and let $(A_1, \ldots, A_n, E) \sim \mathcal{N}(0_{n+1}, \Sigma)$.

Avery (1998) proposes the following information structure for $n = 2$ in his model of jump bidding. Let $X \sim \mathcal{N}(0, \sigma_X^2), Y \sim \mathcal{N}(0, \sigma_Y^2), Z \sim \mathcal{N}(0, \sigma_Z^2)$ be independent and $A_1 = X + Y, A_2 = X + Z, E = A_1 + A_2 = 2X + Y + Z$.[2] We can write

$$\begin{pmatrix} E \\ A_1 \\ A_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4\sigma_X^2+\sigma_Y^2+\sigma_Z^2 & 2\sigma_X^2+\sigma_Y^2 & 2\sigma_X^2+\sigma_Z^2 \\ 2\sigma_X^2+\sigma_Y^2 & \sigma_X^2+\sigma_Y^2 & \sigma_X^2 \\ 2\sigma_X^2+\sigma_Z^2 & \sigma_X^2 & \sigma_X^2+\sigma_Z^2 \end{pmatrix}\right)$$

We can calculate conditional distributions as in Subsection 4.1.2 and get

$$E|A_1 \sim \mathcal{N}\left(\frac{2\sigma_X^2+\sigma_Y^2}{\sigma_X^2+\sigma_Y^2}A_1, \frac{\sigma_X^2\sigma_Y^2+\sigma_X^2\sigma_Z^2+\sigma_Y^2\sigma_Z^2}{\sigma_X^2+\sigma_Y^2}\right) \tag{4.9}$$

$$E|A_2 \sim \mathcal{N}\left(\frac{2\sigma_X^2+\sigma_Z^2}{\sigma_X^2+\sigma_Z^2}A_2, \frac{\sigma_X^2\sigma_Y^2+\sigma_X^2\sigma_Z^2+\sigma_Y^2\sigma_Z^2}{\sigma_X^2+\sigma_Z^2}\right) \tag{4.10}$$

$$E|A_1, A_2 = A_1 + A_2 \tag{4.11}$$

4.2 Predicting Discrete Probability Distributions

4.2.1 Convexity and Nontriviality Conditions

There are three types of inequalities that we need to consider to verify substitutability: (i) submodularity of the value of information function, (ii) convexity of G, (iii) nontriviality of G. We already discussed in the introduction that (i) can be written as a set of

[2] We scale up E by two compared to Avery's model, but in the informational context, they are equivalent. Also, in Avery's model, X, Y, Z are independent uniforms, but Avery proposed this model in the context of auctions, where uniform distributions are most tractable, but in our context, gaussian distributions are most tractable.

linear inequalities. The following proposition shows that (ii) convexity can be written as a set of linear inequalities.

Proposition 4.1. *If we are given n points (x_i, y_i) with $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$, $1 \leq i \leq n$. The necessary and sufficient condition that there exist a convex $G : \mathbb{R}^d \to \mathbb{R}$ such that $G(x_i) = y_i$ is a list of $O(\text{poly}(n^d))$ inequalities, all linear in $y_1, \ldots, y_n$.*

Proof. The necessary and sufficient condition is the following. Take any $d+1$ points (x_i, y_i). These $d+1$ points determine a hyperplane in $\mathbb{R}^d \times \mathbb{R}$ and the $d+1$ x-components x_i form a convex hull $\mathcal{H}$. For any (x_j, y_j) such that x_j lies within the convex hull $\mathcal{H}$, the value y_j must be $\leq$ the value determined by the hyperplane determined by the $d+1$ points. There are $\binom{n}{d+1} = O(\text{poly}(n^d))$ ways to choose $d+1$ points, and at most n points each to lie inside the convex hull. Note that these are necessary and sufficient, and they consist of $O(\text{poly}(n^d))$ inequalities, all linear in y_i's. □

Next, we address (iii) nontriviality. If $G : \Delta_{d-1} \to \mathbb{R}$ is convex and $G(e_i) = 0$ for all i (Proposition 3.2), then convexity of G implies that $G(q) \leq 0$ for all $q \in \Delta_{d-1}$. We can then say that G is nontrivial if $G(q) < 0$ for some q. Convexity of G implies again that if G is zero at an interior point of the simplex Δ_{d-1} then it must be identically zero. Therefore, we can put the nontriviality condition simply by picking an interior point q_0 and ask that $G(q_0)$ is strictly negative. Since Proposition 3.2 allows us to scale G by any positive factor, we can put this condition as $G(q_0) \leq -1$. We prefer this form because standard linear programming solvers only deal with $\geq$ and $\leq$ and not strict inequalities.

4.2.2 Linear Programs and Custom Design

We now develop the idea introduced in the introduction to view the values of G at specific points as variables in a linear program. The following theorem shows that in some interesting information structures, there are polynomially many points that are of concern, and hence the linear programming with look-up table for G is feasible in polynomial time. The details of information structures are repeated in the theorem statement for the reader's convenience.

Theorem 4.2. *In the following information structures, a G can be found or proved not exist in time polynomial in n using a linear program solver.*

1. *Independent Exchangeable Signals with Conditionally Symmetric Event (Subsection 4.1.1) Signals $A_1, \ldots, A_n$ are independent and identically distributed, each taking s values $1, 2, \ldots, s$ with prior probabilities $\alpha_1, \alpha_2, \ldots, \alpha_s$. E takes d values*

$1, 2, \dots, d$. The conditional distribution $E|A_1, \dots, A_n$ is symmetric in the signals, and that if there are $n_1, n_2, \dots, n_s$ of $A_1, \dots, A_n$ that take value $1, 2, \dots, s$ respectively, then the probability of $E = e$ given $A_1, \dots, A_n$ is $\beta_{e,(n_1,\dots,n_s)}$. d and s as constants independent of n.

2. *Conditionally Independent and Exchangeable Signals (Subsection 4.1.2) The outcome E takes d values $1, 2, \dots, d$ with probabilities $\mathbb{P}(E = e) = \alpha_e$. Signals $A_1, \dots, A_n$ each take s values $1, \dots, s$ and are independent and identically distributed conditional on any realization of $E = e$ with probabilities $\mathbb{P}(A_i = a|E = e) = \beta_{a,e}$.*

Proof. In both cases, signals are exchangeable and independent. Therefore, any conditional on k signals are the same.. We call this the k-expected score $ES_k = \mathbb{E}_{|S|=k}\, G(p_{a_S})$. The expression for ES_k depends on the information structure. By symmetry, the submodularity condition is equivalent to $ES_{k+2} + ES_k \geq 2ES_{k+1}$ for $0 \leq k \leq n-2$. Since we already showed in Subsection 4.2.1 that convexity and nontriviality conditions can be formulated as polynomially many linear inequalities in the number of variables. Therefore, the linear program has polynomially many inequalities. To show that it can be solved in polynomial time, we only need to show that the linear program has polynomially many variables. In other words, there are polynomially many possible arguments of G that are present in $(ES_k)_{0 \leq k \leq n-2}$.

1. In the case of independent exchangeable signals with conditionally symmetric event, the k-expected score is given by

$$ES_k = \sum_{u_1+\cdots+u_s=k, 0\leq u_i \leq k} \binom{k}{u_1, \dots, u_s} \alpha_1^{u_1} \dots \alpha_s^{u_s} \times$$
$$G\left(\sum_{v_1+\cdots+v_s=n-k, 0\leq v_j \leq n-k} \binom{n-k}{v_1, \dots, v_s} \alpha_1^{v_1} \dots \alpha_s^{v_s} \beta_{(u_1+v_1,\dots,u_s+v_s)} \right)$$

where $\beta_{(n_1,\dots,n_s)} = \left(\beta_{e,(n_1,\dots,n_s)}\right)_{e\in E}$. We also note the fact that an equation $x_1 + \cdots + x_s = y$ has $\binom{y+s-1}{s-1} = O(y^{s-1})$ solutions in nonnegative integers. There are $O(k^{s-1})$ s-tuples $u = (u_1, \dots, u_s)$, and each u is associated with one variable $G(\cdot)$. The argument of that $G(\cdot)$ can be computed by summing over all $O((n-k)^{s-1}) = O(n^{s-1})$ terms of v, which can be done in polynomial time. Therefore, there are $O(k^{s-1})$ variables in ES_k, so summing over all k, there are $\sum_k O(k^{s-1}) = O(n^s)$ variables in total, which is polynomial in n.

2. In the case of conditionally independent and exchangeable signals, the k-expected score is given by

$$ES_k = \sum_{u_1+\cdots+u_s=k, 0\leq u_i \leq k} \binom{k}{u_1, \ldots, u_s} \left(\sum_{e'} \alpha_{e'} \prod_{j=1}^{s} \beta_{j,e'}^{u_j} \right) G\left(\left(\frac{\alpha_e \prod_{j=1}^{s} \beta_{j,e}^{u_j}}{\sum_{e'} \alpha_{e'} \prod_{j=1}^{s} \beta_{j,e'}^{u_j}} \right)_{e\in E} \right)$$

Every expression can be computed in polynomial time. As in 1., there are $O(k^{s-1})$ s-tuples $u = (u_1, \ldots, u_s)$, and each u is associated with one variable $G(\cdot)$, so the total number of variables is $\sum_k O(k^{s-1}) = O(n^s)$, which is polynomial in n.

□

Before we proceed, we note that Proposition 2.10 tells us that independent signals are complements if the Bregman divergence D_G of G are jointly convex. This tells us that if we can find a convex G in Theorem 4.2, then D_G is not jointly convex.

On the one hand, the meta-message of Proposition 2.10 is that independent signals are not friendly to substitutability. On the other hand, we might expect that conditionally independent signals should be quite friendly to substitutability because each person has a noisy observation of the true value (recall 4.1.2). Since they are observations of the same value, the value of the first signal should be more than the marginal signal that observes the same thing.

Broadly speaking, conditionally independent signals are more conducive to substitutability, but whether a G can be designed depends intricately on both the details (or parameters) of the information structure, and the outcome as a function of signals.

For example, consider the following problem. Let $A_1, \ldots, A_n$ be independent and identically distributed binary signals such that $\mathbb{P}(A_i = 1) = p$, and let E be a symmetric boolean function of the signals, then E can be completely characterized by the set $S \subseteq \{0, 1, \ldots, n\}$ such that

$$E = f(A_1, \ldots, A_n) = \mathbb{1}[\text{the number of 1s in } A_i \text{ is in set } S]$$

We are interested in all possible E, that is, all possible sets S, such that a substitutability-inducing convex G exists. Let's provisionally call such a set S *good* for convenience. We know a priori that E in the form of XOR or negation cannot work, that is $\{2k : 0 \leq k \leq n/2\}$ and $\{2k + 1 : 0 \leq k \leq (n - 1)/2\}$. In Proposition 4.3, we show that $\{n\}$ is good. There are also two immediate symmetries that are easily checked.

- If S is a good set, then $\{0, 1, \ldots, n\} \setminus S$ is a good set.

- If $S = \{s_1, \ldots, s_u\}$ is a good set, then $S' = \{n - s_1, \ldots, n - s_u\}$ is a good set.

However, the restriction from universal complements is far from enough. Even though some information structures are not universal complements, no substitutability-inducing G exists. They are "in between" substitutes and complements. For concreteness, let $(n, p) = (5, 1/3)$. Universal complements prohibit $\{0, 2, 4\}$ and $\{1, 3, 5\}$. There are 2^6 subsets of $\{0, 1, 2, \ldots, 5\}$, but by direct computation using linear programs as in Theorem 4.2, only 10 sets S are good:

$$\{1,2,3,4,5\}, \{0,4,5\}, \{4,5\}, \{0,1,2,3,5\}, \{0,5\}, \{1,2,3,4\}, \{4\}, \{0,1,2,3\}, \{1,2,3\}, \{0\}$$

Theorem 4.2 gives us a general recipe to use a linear program to design a G. We can, alternatively, use the "intuition" behind substitutability to design a custom G for particular information structures. In particular, we want the G to have diminishing return, so we want "high" curvature near the prior and "low" curvature near the boundary. At the same time, convexity of G enforces that the curvature near the prior cannot be too high compared to curvature near the boundary. This tension highlights the difficulty of designing a valid G. A G exists and can be designed if and only if such a resolution is possible. The drawback of this approach is that it requires a careful analysis of the information structure at hand and can be *ad hoc*. However, if we succeed, we can explicitly design G. The next proposition demonstrates this approach.

Proposition 4.3. *If $A_1, \ldots, A_n$ are independent binary signals such that each is 1 with probability $p \in (0, 1)$, then if $E = A_1 \wedge \cdots \wedge A_n$, then we can find a decision problem that these signals are substitutes.*

Proof. The k-expected score is $ES_k = p^{-k} G(p^{n-k})$. Let $G(0) = G(1) = 0$ and $G(p^{-(n-t)}) = -1 + M^t \epsilon$ with M, ϵ chosen later. Then submodularity is equivalent to $2ES_{k+1} \geq ES_{k+2} + ES_k$ for all k, which reduces to $\epsilon \leq \frac{(1-p)^2}{M^k (Mp-1)^2}$, so for any fixed $M > 1/p$ we can find ϵ small enough that this is satisfied. Convexity requires $\frac{G(p^{-(n-(k+1))}) - G(p^{-(n-k)})}{p^{-(n-(k+1))} - p^{-(n-k)}} \geq \frac{G(p^{-(n-k)}) - G(p^{-(n-(k-1))})}{p^{-(n-k)} - p^{-(n-(k-1))}}$ which reduces to $M \geq 1/p$, and $\frac{G(p^{n-1}) - G(p^n)}{p^{n-1} - p^n} \geq \frac{G(p^n) - G(0)}{p^n - 0}$ which reduces to $p(M-1)\epsilon \geq (1-p)(\epsilon - 1)$ which is true if $M > 1$ and $0 < \epsilon < 1$, and $\frac{G(1) - G(p)}{1-p} \geq \frac{G(p) - G(p^2)}{p - p^2}$ which reduces to $\epsilon \leq \frac{p}{(1+p)M^{n-1} - M^{n-2}}$ which for $M > 1$ is true for sufficiently small ϵ. So we can choose any $M > 1/p$ and ϵ sufficiently small to finish our construction. □

Note that by bit flip symmetry, we can modify the G of Proposition 4.3 to get another G that induces substitutability for $E = A_1 \vee \cdots A_n = \neg(\neg A_1 \wedge \cdots \wedge \neg A_n)$.

4.2.3 Designing Interpretable Expected Score Functions

Even if there are a polynomial number of values of G that we care about, the direct linear programming approach outlined above gives an uninterpretable lookup-table approach; the values of G are just a list of numbers that satisfy all the necessary conditions as determined by the linear program solver. We might prefer a function G from a *parametrized* family. Let $G_1, \ldots, G_m$ be known functions of a distribution q. We can let $G(q) = \sum_{i=1}^{m} \theta_i G_i(q)$, where θ_i are unknown parameters. The function G is now interpretable: it is a weighted combination of m features $G_1, \ldots, G_m$ which can be explained to market participants. Moreover, the set of inequalities are still linear in parameters $\theta_1, \ldots, \theta_m$, so we can find appropriate values for them by linear programming.

4.3 Predicting Continuous Probability Distributions

In this section, we consider the problem of predicting continuous probability distributions. Even though the forecaster cannot submit the density of her belief at every point because there is an infinite of them, if her belief is structured (say, following a parametrized distribution), she can report the type of the distribution and its parameters. The aggregator then reconstructs her belief and uses the generalization of scoring rules to continuous distributions to score her answer. We follow the notations related to continuous scoring rules from Gneiting and Raftery (2007).

Following Gneiting and Raftery (2007), there is a correspondence between a scoring rule and the convex expected score function G. We have used this fact before in the case of discrete probability distributions, but this fact holds for general measurable spaces as well. We are especially interested in scoring rules in the class $\mathcal{S}_\alpha$ that has the expected score function $G(p) = ||p||_\alpha^\alpha$, where $||p||_\alpha = \left(\int p(e)^\alpha de\right)^{1/\alpha}$ is the α-norm.

The class $\mathcal{S}_\alpha$ encompasses many strictly proper scoring rules used in practice. The log scoring rule $\text{LogScore}(p, e) = \log p(e)$ is achieved for $\alpha \to 1$. The quadratic scoring rule $\text{QS}(p, e) = 2p(e) - ||p||_2^2$ with $G(p) = ||p||_2^2$ corresponds to $\alpha = 2$. The pseudospherical score $\text{PseudoS}(p, e) = p(e)^{\alpha-1}/||p||_\alpha^{\alpha-1}$ corresponds to general α, and reduces to the spherical score when $\alpha = 2$.

We are mostly interested in this class $\mathcal{S}_\alpha$ of scoring rules in this section. In particular, for a given information structure, we might ask which values of α can induce substitutes, or whether, fixing a scoring rule in this class, varying correlations between signals can induce substitutes or complements.

First, we need to be able to compute the α-norm of standard distributions. The following proposition does so for Normal and Cauchy. We use Normal and Cauchy because we are adding random variables, and both Normal and Cauchy are stable distributions – the sum of two independent random variables in the family still belong to the family.

Proposition 4.4. *Let $\alpha > 1$.*

1. *If $p \sim \mathcal{N}(\mu, \Sigma)$ is multivariate normal, then $||p||_\alpha^\alpha = \mathrm{const}(\alpha) \times |\Sigma|^{-(\alpha-1)/2}$.*
2. *If $p \sim \mathrm{Cauchy}(x_0, \gamma)$, then $||p||_\alpha^\alpha = \mathrm{const}(\alpha) \times \gamma^{-(\alpha-1)/2}$*

where $\mathrm{const}(\alpha)$ depends only on α, and $|\Sigma|$ is the determinant of Σ.

Even though the log scoring rule can be recovered from $\mathcal{S}_\alpha$ with $\alpha \to 1$, it is often more convenient to work with it directly. With log scoring rule, the expected score is $G(p) = \int p(x) \log p(x) dx = -H(p)$, where H is the (Shannon) entropy. Entropies for Normal and Cauchy distributions are standard.

Proposition 4.5. *Let $H(p) = \int p(x) \log p(x) dx$ be the entropy of distribution p.*

- *If $p \sim \mathcal{N}(\mu, \Sigma)$, then $H(p) = \frac{1}{2} \log((2\pi e)^k |\Sigma|) = \mathrm{const} + \frac{1}{2} \log |\Sigma|$*
- *If $p \sim \mathrm{Cauchy}(x_0, \gamma)$, then $H(p) = -\log(4\pi\gamma) = \mathrm{const} - \log \gamma$*

Proposition 2.10 tells us that independent signals are complements if D_G is jointly convex. We already know that $\alpha = 2$ corresponds to a quadratic scoring rule with $D_G(p, q) = ||p - q||_2^2$ with is jointly convex. However, within the family $\mathcal{S}_\alpha$, if $\alpha > 2$ then D_G is not jointly convex.

Proposition 4.6. *Let $G(p) = ||p||_\alpha^\alpha$ for any distribution p, and $\alpha > 2$, then $D_G(p, q)$ is not convex in q, and thus not jointly convex.*

Proof. We assume that $p = (p_i)_{i=1}^n, q = (q_i)_{i=1}^n$. The infinite case follows by standard limit argument. We have

$$D_G(p, q) = G(p) - G(q) - G'(q) \cdot (p - q) = \sum_i \left(p_i^\alpha - q_i^\alpha - \alpha q_i^{\alpha-1}(p_i - q_i) \right)$$
$$= \sum_i \left(p_i^\alpha + (\alpha - 1) q_i^\alpha - \alpha q_i^{\alpha-1} p_i \right)$$

Since $\alpha > 2$, we can calculate

$$\frac{\partial^2}{\partial q_k^2} D_G(p, q) = \alpha(\alpha - 1)^2 q_k^{\alpha-2} - \alpha(\alpha - 1)(\alpha - 2) q_k^{\alpha-3} p_k$$

For a fixed q_k, we can choose p_k large enough so that $\frac{\partial^2}{\partial q_k^2} D_G(p,q) < 0$, so $D_G(p,q)$ is not convex in q. As a result, it is not jointly convex. □

4.3.1 Conditionally Independent Gaussian Signals

Before we prove the main result, we will prove a technical lemma, which is an inequality that shows up not only in this section, but in many sections after.

Lemma 4.7. *If* $0 \leq \lambda \leq 1$*, then for any* $x, y, z \geq 0$*,*

$$(x+y)^\lambda + (x+z)^\lambda \geq (x+y+z)^\lambda + x^\lambda$$

If $\lambda \leq 0$ *or* $\lambda \geq 1$*, then for any* $x, y, z \geq 0$*,*

$$(x+y)^\lambda + (x+z)^\lambda \leq (x+y+z)^\lambda + x^\lambda$$

Proof. Let $f(y) = (x+y)^\lambda + (x+z)^\lambda - (x+y+z)^\lambda - x^\lambda$, then $f'(y) = \lambda\left((x+y)^{\lambda-1} - (x+y+z)^{\lambda-1}\right)$. If $0 \leq \lambda \leq 1$, then $f'(y) \geq 0$, so $f(y) \geq f(0) = 0$. If $\lambda \leq 0$ or $\lambda \geq 1$, then $f'(y) \leq 0$, so $f(y) \leq f(0) = 0$. □

Now we use Lemma 4.7 to characterize substitutes and complements for conditionally independent gaussian signals in Proposition 4.8. This is the main result of this section. The proof is by direct calculation.

Proposition 4.8. *Let* $E \sim \mathcal{N}(0, \sigma_E^2)$ *and* $A_i|E \sim \mathcal{N}(E, \sigma_i^2)$*,* $i = 1, \ldots, n$*, are conditionally independent given* E*, then* $(A_1, \ldots, A_n; E)$ *are substitutes under the log scoring rule and* $\mathcal{S}_\alpha$ *for* $1 < \alpha \leq 3$*, and are complements under* $\mathcal{S}_\alpha$ *for* $\alpha \geq 3$*.*

Proof. We review the properties of this model in Subsection 4.1.2. In particular, since the expected score of the log or the $\mathcal{S}_\alpha$ scoring rule of a gaussian only depends on the various, we are only interested in the variance of conditional distributions, which we can write as

$$\mathrm{Var}\left(E|(A_k)_{k\in S}\right) = \left(\sigma_E^{-2} + \sum_{k\in S} \sigma_k^{-2}\right)^{-1}$$

Recall that the expected score of $\mathcal{N}(\mu, \sigma^2)$ under $\mathcal{S}_\alpha$ is $(\sigma^2)^{-\beta}$ for $\beta = (\alpha-1)/2$. To verify substitutability, let $S \subseteq \{1, 2, \ldots, n\}$ and $i, j \in \{1, 2, \ldots, n\} \setminus S$. We need to show that

$$\mathbb{E}\, G(p_{Sa_i}) + \mathbb{E}\, G(p_{Sa_j}) \geq \mathbb{E}\, G(p_{Sa_ia_j}) + \mathbb{E}\, G(p)$$

Let $x = \sigma_E^{-2} + \sum_{k \in S} \sigma_k^{-2}, y = \sigma_i^{-2}, z = \sigma_j^{-2}$, the above inequality reduces to

$$(x+y)^\beta + (x+z)^\beta \geq x^\beta + (x+y+z)^\beta$$

which is true by Lemma 4.7 since $0 < \beta \leq 1$. Note that by the same lemma, $(A_1, \ldots, A_n; E)$ are complements under $\mathcal{S}_\alpha$ for $\alpha \geq 3$.

The log scoring rule can be proved by taking the limit $\alpha \to 1$. It can also be proved directly. Let x, y, z be as above, then using the expected score expression for the log scoring rule, the inequality reduces to $(x+y)(x+z) \geq x(x+y+z)$ or $yz \geq 0$. □

4.3.2 Correlation Thresholds for Independent Gaussian Signals

Proposition 4.9. *Let $A_1 \sim \mathcal{N}(0, \sigma_1^2)$, $A_2 \sim \mathcal{N}(0, \sigma_2^2)$, $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ be such that $E = A_1 + A_2 + \epsilon$, (A_1, A_2) is independent of ϵ, and (A_1, A_2) is jointly normal with correlation ρ. Let the scoring rule be from $\mathcal{S}_\alpha$. Fix $\sigma_1^2, \sigma_2^2, \sigma_\epsilon^2, \alpha$, but let ρ vary. Either $(A_1, A_2; E)$ are always complements, or there exists a $\rho^* > 0$ such that if $\rho \geq \rho^*$, then $(A_1, A_2; E)$ are substitutes and if $\rho \leq \rho^*$, then $(A_1, A_2; E)$ are complements.*

Proof. The value ρ induces substitutes if and only if

$$(\sigma_1^2 + \sigma_\epsilon^2)^{-\beta} + (\sigma_2^2 + \sigma_\epsilon^2)^{-\beta} \geq (\sigma_\epsilon^2)^{-\beta} + (\sigma_1^2 + \sigma_2^2 + \sigma_\epsilon^2 + 2\rho\sigma_1\sigma_2)^{-\beta}$$

for $\beta = (\alpha - 1)/2 > 0$. The left hand side is fixed and the right hand side is a decreasing function of ρ. Therefore, either this inequality is always false or there exists a ρ^* such that this is true if and only if $\rho \geq \rho^*$. Moreover, this inequality is false for $\rho = 0$ by Lemma 4.7, so $\rho^* > 0$ □

The theorem quantifies an intuitive fact that signals that are more correlated tend to be closer to substitutes. Note that the result is much cleaner than the discrete distribution case, because we have tractable distributions, and we can talk sensibly about correlations in the Normal model.

For intuition, we can redo Proposition 4.9 with the log scoring rule (corresponding to $\alpha \to 1$). The value ρ induces substitutes if and only if

$$-\log(\sigma_1^2 + \sigma_\epsilon^2) - \log(\sigma_2^2 + \sigma_\epsilon^2) \geq -\log(\sigma_\epsilon^2) - \log(\sigma_1^2 + \sigma_2^2 + \sigma_\epsilon^2 + 2\rho\sigma_1\sigma_2) \tag{4.12}$$

which is equivalent to

$$\rho \geq \frac{\sigma_1 \sigma_2}{\sigma_\epsilon^2}$$

If $\sigma_\epsilon^2 < \sigma_1 \sigma_2$, then the inequality is always false. This is intuitive because if the outcome noise σ_ϵ is low compared to the noise from agent observations $\sigma_{1,2}$, then the two signals complement each other in that the two signals combined can predict the outcome precisely, but each individual signal cannot because the signal is noisy.

In general, it is *not* true that ρ^* is an increasing function of σ_1, σ_2 and a decreasing function of σ_ϵ. The reader can check that $(\sigma_2, \sigma_\epsilon, \beta) = (0.1, 0.1, 0.5)$ is a counterexample.

4.3.3 Complements In Independent Signals Prove New Inequalities

This subsection is not directly connected to the main narrative, but is interesting nonetheless and best fits here. We use the fact that independent signals are complements if D_G is jointly convex (Proposition 2.10) to prove new inequalities. The log scoring rule and the quadratic scoring rule have jointly convex Bregman divergence $D_G(p, q)$, because they are $KL(p||q)$ and $||p - q||_2^2$, so this proposition applies. Consider the information structure $E = A_1 + A_2 + \epsilon$ with $A_1 \sim \mathcal{N}(0, \Sigma_1)$, $A_2 \sim \mathcal{N}(0, \Sigma_2)$, $\epsilon \sim \mathcal{N}(0, \Sigma_\epsilon)$ are independent multivariate normals with the same dimension. The conditional distributions are easily computed

$$\begin{aligned} E &\sim \mathcal{N}(0, \Sigma_1 + \Sigma_2 + \Sigma_\epsilon) \\ E|A_1 &\sim \mathcal{N}(A_1, \Sigma_2 + \Sigma_\epsilon) \\ E|A_2 &\sim N(A_2, \Sigma_1 + \Sigma_\epsilon) \\ E|A_1, A_2 &\sim \mathcal{N}(A_1 + A_2, \Sigma_\epsilon) \end{aligned}$$

The inequality for complements is

$$\mathbb{E}\, G(p) + \mathbb{E}\, G(p_{a_1 a_2}) \geq \mathbb{E}\, G(p_{a_1}) + \mathbb{E}\, G(p_{a_2})$$

If we let $G(p) = ||p||_\alpha^\alpha$, then $\mathbb{E}\, G(p_{a_1}) = \mathbb{E}\, |\Sigma_2 + \Sigma_\epsilon|^{-(\alpha-1)/2} = |\Sigma_2 + \Sigma_\epsilon|^{-(\alpha-1)/2}$. Other terms can be computed similarly. Substituting these terms back, we get

$$|\Sigma_1 + \Sigma_2 + \Sigma_\epsilon|^{-\beta} + |\Sigma_\epsilon|^{-\beta} \geq |\Sigma_1 + \Sigma_\epsilon|^{-\beta} + |\Sigma_2 + \Sigma_\epsilon|^{-\beta}$$

for $\beta = (\alpha - 1)/2$. We know that this inequality must be true for $1 \leq \alpha \leq 2$. We can also substitute $G(p) = -H(p)$ for log scoring rule. The computation is similar; for example,

$\mathbb{E}\, G(p_{a_1}) = \mathbb{E} -\frac{1}{2}\log|\Sigma_2 + \Sigma_\epsilon| = -\frac{1}{2}\log|\Sigma_2 + \Sigma_\epsilon|$. Since $\Sigma_1, \Sigma_2, \Sigma_\epsilon$ can be any positive semidefinite matrix, we get the following

Proposition 4.10. *Let A, B, C be any positive semidefinite matrices. We have*

1. $|A + B + C| \cdot |A| \leq |A + B| \cdot |A + C|$
2. $|A + B + C|^{-\beta} + |A|^{-\beta} \leq |A + B|^{-\beta} + |A + C|^{-\beta}$ *for* $0 \leq \beta \leq 1/2$.

We conjecture that the inequality $|A + B + C|^{-\beta} + |A|^{-\beta} \leq |A + B|^{-\beta} + |A + C|^{-\beta}$ holds for any $\beta > 0$ and positive semidefinite matrices A, B, C. This is true for one dimension because then A, B, C are positive scalars, and we can easily check that the difference between the two sides is an increasing function of B, so it reaches its minimum at $B = 0$ where the two sides are equal. This conjecture is equivalent to the information settings that if $E = A_1 + A_2 + \epsilon$ and A_1, A_2, ϵ are independently drawn from multivariate normal distributions, then $(A_1, A_2; E)$ are complements. This class of matrix inequalities seem to be closely related to the Minkowski determinant theorem and tensor product inequalities (Marcus and Minc, 2010; Berndt and Sra, 2015). In general, these types of inequalities are very hard to prove. We do not know an independent proof of Proposition 4.10.

4.4 Predicting Distribution Properties

In this section, traders are not predicting the entire probability distribution of outcomes. They only predict a *property*. Recall the extended discussions in Section 2.4. When eliciting means (expectations), the expected score function can be written as $G(\mathbb{E}\, q)$, where q is the forecaster's belief and G a convex function. We call a mean-eliciting scoring rule *canonical* if $s[r](e) = -(r - e)^2$, or equivalently, $G(r) = r^2$ where r is the reported mean. This canonical scoring rule has an equivalent representation of the expected score as $-\mathrm{Var}(q)$. We call a median-eliciting scoring rule *canonical* if $s[r](e) = -|r - e|$, which has the expected score $-c\sigma$, if $q = \mathcal{N}(\mu, \sigma^2)$.

We first show that conditionally independent normal signals are substitutes, while independent normal signals are complements, under the mean-eliciting and median-eliciting canonical scoring rules. The proofs are by direct calculation and Lemma 4.7.

Proposition 4.11. *Let $E \sim \mathcal{N}(0, \sigma_E^2)$ and $A_i|E \sim \mathcal{N}(E, \sigma_i^2)$, $i = 1, 2, \ldots, n$, are conditionally independent given E, then $(A_1, \ldots, A_n; E)$ are substitutes under the mean-eliciting and median-eliciting canonical scoring rules.*

Proof. We use the same expression as those in Proposition 4.8 that $\mathrm{Var}\left(E|(A_k)_{k\in S}\right) = \left(\sigma_E^{-2} + \sum_{k\in S}\sigma_k^{-2}\right)^{-1}$, with $G(q) = -\mathrm{Var}(q)^\beta$, with $\beta = 1$ and $1/2$ for mean-eliciting and median-eliciting scoring rules, respectively. To show that $\mathbb{E}\, G(p_{Sa_i}) + \mathbb{E}\, G(p_{Sa_j}) \geq \mathbb{E}\, G(p_{Sa_ia_j}) + \mathbb{E}\, G(p)$, let $x = \sigma_E^{-2} + \sum_{k\in S}\sigma_k^{-2}, y = \sigma_i^{-2}, z = \sigma_j^{-2}$, the inequality reduces to $(x+y)^{-\beta} + (x+z)^{-\beta} \geq x^{-\beta} + (x+y+z)^{-\beta}$, which is true by Lemma 4.7. □

Proposition 4.12. *Let $\epsilon \sim \mathcal{N}(0,\sigma_\epsilon^2)$ and $A_i \sim \mathcal{N}(0,\sigma_i^2)$, $i = 1,\dots,n$, be independent and $E = A_1+\cdots+A_n+\epsilon$, then $(A_1,\dots,A_n;E)$ are complements under the mean-eliciting and median-eliciting canonical scoring rules.*

Proof. We use the fact that $\mathrm{Var}\left(E|(A_k)_{k\in S}\right) = \sigma_E^2 + \sum_{k\notin S}\sigma_k^2$, with $G(q) = -\mathrm{Var}(q)^\beta$, with $\beta = 1$ and $1/2$ for mean-eliciting and median-eliciting scoring rules, respectively. To show that $\mathbb{E}\, G(p_{Sa_ia_j}) + \mathbb{E}\, G(p) \geq \mathbb{E}\, G(p_{Sa_i}) + \mathbb{E}\, G(p_{Sa_j})$, let $T = \{1,2,\dots,n\}\setminus(S\cup\{i,j\})$ and $x = \sigma_E^2 + \sum_{k\in T}\sigma_k^2, y = \sigma_i^2, z = \sigma_j^2$, the inequality reduces to $x^\beta + (x+y+z)^\beta \geq (x+y)^\beta + (x+z)^\beta$, which is true by Lemma 4.7. □

Proposition 4.13. *Let $X \sim \mathcal{N}(0,\sigma_X^2), Y \sim \mathcal{N}(0,\sigma_Y^2), Z \sim \mathcal{N}(0,\sigma_Z^2)$ be independent and $A_1 = X+Y, A_2 = X+Z, E = A_1 + A_2 = 2X+Y+Z$ (see Subsection 4.1.3), then $(A_1,A_2;E)$ are substitutes under the canonical mean-eliciting scoring rule, but not under the canonical median-eliciting scoring rule.*

Proof. Let $G(q) = -\mathrm{Var}(q)^\beta$, with $\beta = 1$ and $1/2$ for canonical mean-eliciting and median-eliciting scoring rules, respectively. The inequality to be proved is

$$\left(\frac{\sigma_X^2\sigma_Y^2 + \sigma_X^2\sigma_Z^2 + \sigma_Y^2\sigma_Z^2}{\sigma_X^2 + \sigma_Y^2}\right)^\beta + \left(\frac{\sigma_X^2\sigma_Y^2 + \sigma_X^2\sigma_Z^2 + \sigma_Y^2\sigma_Z^2}{\sigma_X^2 + \sigma_Z^2}\right)^\beta \leq (4\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2)^\beta$$

Let $y = \sigma_Y^2/\sigma_X^2$ and $z = \sigma_Z^2/\sigma_X^2$. For $\beta = 1$, the inequality reduces to

$$\frac{y+z+yz}{1+y+z+yz}(2+y+z) \leq 4+y+z$$

Let $y+z = 2t$. The left hand side is an increasing function in yz. Since $yz \leq t^2$, it suffices to show the inequality for $yz = t^2$, where it reduces to $\frac{2t+t^2}{(1+t)^2}(2+2t) \leq 4+2t$, which is true.

For $\beta = 1/2$, a counterexample is $(y,z) = (1,2)$. □

Proposition 4.11 shows that conditionally independent signals are substitutes under a canonical scoring rule if the signals are gaussian. This is no longer true in non-gaussian models.

The two given counterexamples are instructive. The first shows that bit-corruption (which is not additive) is an unacceptable form of noise. The second shows that even if noises are additive, signals can fail to be substitutes if the noise is monotone in the sense that the noise can only push the signal one way (in this case, make the signal value higher) rather than randomly making the outcome lower or higher with equal probability as in the case of gaussian signals.

Proposition 4.14. *It is not true that if A_1 and A_2 are conditionally independent given E, then $(A_1, A_2; E)$ are substitutes under a canonical scoring rule. The following two information structures are counterexamples in the sense that there are parameter values that make the signals not substitutes.*

1. *A_1, A_2, E are binary ($\{0,1\}$-valued) such that $\mathbb{P}(E = 1) = \pi_E$ and for each $i = 1, 2$, $A_i|E$ are conditionally independent with $\mathbb{P}(A_i = e|E = e) = \pi$ for $e \in \{0, 1\}$.*
2. *Let $E, \epsilon_1, \epsilon_2$ be independent binary signals which are 1 with probabilities π_E, π_1, π_2 respectively, and $A_1 = E + \epsilon_1, A_2 = E + \epsilon_2$.*

Proof. We check substitutability by calculation. See details in Appendix A.3.1.

In the first case, substitutability is equivalent to

$$1 + 2\pi(1-\pi) + (1-\pi)^2\pi^2 \left(\frac{1}{(1-\pi)^2\pi_E + \pi^2(1-\pi_E)} + \frac{1}{\pi_E\pi^2 + (1-\pi_E)(1-\pi)^2} \right) \geq 2\pi(1-\pi) \left(\frac{1}{(1-\pi)\pi_E + \pi(1-\pi_E)} + \frac{1}{\pi_E\pi + (1-\pi_E)(1-\pi)} \right)$$

A counterexample is $(\pi_E, \pi) = (0.99, 0.49)$.

In the second case, substitutability is equivalent to

$$1 + \frac{\pi_1\pi_2\bar{\pi}_1\bar{\pi}_2}{\pi_E\bar{\pi}_1\bar{\pi}_2 + \bar{\pi}_E\pi_1\pi_2} \geq \frac{\pi_1\bar{\pi}_1}{\pi_E\bar{\pi}_1 + \bar{\pi}_E\pi_1} + \frac{\pi_2\bar{\pi}_2}{\pi_E\bar{\pi}_2 + \bar{\pi}_E\pi_2}$$

A counterexample is $(\pi_E, \pi_1, \pi_2) = (0.95, 0.8, 0.8)$.

□

The case of gaussian signals is a special case, but we show the robustness of this result by checking substitutability for t-distributed signals in Proposition 4.15. This is especially useful since in finance and economics, sometimes we want to model the noise with something thicker-tailed than the normal, such as the t-distribution or Cauchy (which is t with one degree of freedom). The standard normal distribution is t-distributed with infinite degrees of freedom.

The t-distributions are much harder to deal with than gaussians because conditional distributions have to be computed numerically. Moreover, the fact that they are thick-tailed sometimes causes numerical instabilities in the integration procedure, and errors need to be carefully accounted for to verify submodularity inequalities within acceptable error tolerances. In fact, up to this point, we verify substitutability directly via Definition 2.15, which requires computing conditional distributions, but computing conditional distributions are challenging for most non-gaussian distributions. An interesting future research direction is coming up with a more practical way to check substitutability.

Proposition 4.15. *Let $c \cdot t_{df}$ be a distribution of cX where X is t-distributed with df degrees of freedom. Let $df_E, df_1, df_2 \geq 1$ and $c_E, c_1, c_2 > 0$. Let $E \sim c_E \cdot t_{df_E}, \epsilon_1 \sim c_1 \cdot t_{df_1}, \epsilon_2 \sim c_2 \cdot t_{df_2}$ be independent and $A_1 = E + \epsilon_1, A_2 = E + \epsilon_2$. Numerical simulations suggest that $(A_1, A_2; E)$ are substitutes under a canonical mean-eliciting scoring rule, or at least that they are substitutes under a reasonable range of parameters.*

Proposition 4.15 suggests that the gaussian signals model (that we use for most of the chapter) captures two salient aspects that are robust to perturbation. First, both the outcome and the noises are symmetric bell-shaped curves with a peak that levels off in both directions. This avoids the drawbacks of the two counterexamples above that (i) the noise does not make the outcome jump from one value to another like from 0 to 1, and (ii) the noise smears the outcome around both positively and negatively, rather than push the outcome in one direction. We do not know yet if these properties, appropriately formalized, imply substitutability; we hope to address these issues in future work.

Up until this point, we have only considered canonical scoring rules. We now consider non-canonical scoring rules. As Section 2.3 indicates, if the belief distribution is q, then the mean-eliciting scoring rule is such that the expected score function is $G(q) = f(\mathbb{E}\, q)$ for some convex function f. The canonical scoring rule corresponds to $f(x) = x^2$ which makes the conditionally independent normal signals substitutes. We will characterize functions of the form $f(t) = t^\lambda$. Since f is convex, $f''(t) = \lambda(\lambda - 1)t^{\lambda-2}$, so we need $\lambda \geq 1$. The next proposition claims that the signals are substitutes if and only if $1 \leq \lambda \leq 2$.

Proposition 4.16. *Let $E \sim \mathcal{N}(0, \sigma_E^2)$ and $A_i | E \sim \mathcal{N}(E, \sigma_i^2)$, $i = 1, \ldots, n$ be conditionally independent given E. Consider the mean-eliciting scoring rule such that the expected score function is $G(q) = f(\mathbb{E}\, q) = (\mathbb{E}\, q)^\lambda$ for $\lambda \geq 1$. Then, $(A_1, \ldots, A_n; E)$ are substitutes for every value of parameters $\sigma_E^2, \sigma_1^2, \ldots, \sigma_n^2$ if and only if $\lambda \leq 2$.*

Proof. We use the fact that if $Z \sim \mathcal{N}(0, \sigma^2)$, then $\mathbb{E}\,|Z|^p = c_p \sigma^p$ for constant c_p depending only on p. The inequality to show is $\mathbb{E}\, G(p_{Sa_i}) + \mathbb{E}\, G(p_{Sa_j}) \geq \mathbb{E}\, G(p_{Sa_ia_j}) + \mathbb{E}\, G(p_S)$. If signals are substitutes for two signals, that is, for $S = \emptyset$, then, starting from a fixed S, $E|(A_k)_{k\in S}$ is normal, so applying the two-signal results with prior $E|(A_k)_{k\in S}$ we get the inequality for a fixed S. Since this holds for every S, taking expectation over S gives the substitutes inequality for S with a_i and a_j. Conversely, if the signals are substitutes then they must be substitutes for two signals. Thus it is necessary and sufficient to deal with two signals, and by symmetry we can call these two signals A_1 and A_2. We know from (4.8) that $E|A_1$ is normal with mean $\frac{\sigma_E^2}{\sigma_E^2+\sigma_1^2}A_1$, so $\mathbb{E}_{A_1} G(p_{a_1}) = \mathbb{E}_{A_1}(\mathbb{E}\, E|A_1)^\lambda = \mathbb{E}_{A_1}\left(\frac{\sigma_E^2}{\sigma_E^2+\sigma_1^2}A_1\right)^\lambda$. Now, $A_1 \sim \mathcal{N}(0, \sigma_E^2 + \sigma_1^2)$, so $\frac{\sigma_E^2}{\sigma_E^2+\sigma_1^2}A_1 \sim \mathcal{N}\left(0, \frac{\sigma_E^4}{\sigma_E^2+\sigma_1^2}\right)$. We conclude that $E_{A_1} G(p_{a_1}) = c_\lambda \left(\frac{\sigma_E^4}{\sigma_E^2+\sigma_1^2}\right)^{\lambda/2}$. Similarly, $E_{A_2} G(p_{a_2}) = c_\lambda \left(\frac{\sigma_E^4}{\sigma_E^2+\sigma_2^2}\right)^{\lambda/2}$. We know from (4.8) that $E|A_1, A_2$ is normal with mean $\frac{\sigma_E^2\sigma_1^2 A_2+\sigma_E^2\sigma_2^2 A_1}{\sigma_E^2\sigma_1^2+\sigma_E^2\sigma_2^2+\sigma_1^2\sigma_2^2}$. To find the distribution of this expression, write it as

$$\frac{\sigma_E^2\sigma_1^2(E+\epsilon_2)+\sigma_E^2\sigma_2^2(E+\epsilon_1)}{\sigma_E^2\sigma_1^2+\sigma_E^2\sigma_2^2+\sigma_1^2\sigma_2^2} = \frac{\sigma_E^2(\sigma_1^2+\sigma_2^2)E+\sigma_E^2\sigma_2^2\epsilon_1+\sigma_E^2\sigma_1^2\epsilon_2}{\sigma_E^2\sigma_1^2+\sigma_E^2\sigma_2^2+\sigma_1^2\sigma_2^2}$$

which has mean 0 and variance

$$\frac{\sigma_E^4(\sigma_1^2+\sigma_2^2)^2\sigma_\epsilon^2+\sigma_E^4\sigma_2^4\sigma_1^2+\sigma_E^4\sigma_1^4\sigma_2^2}{(\sigma_E^2\sigma_1^2+\sigma_E^2\sigma_2^2+\sigma_1^2\sigma_2^2)^2} = \frac{\sigma_E^4(\sigma_1^2+\sigma_2^2)}{\sigma_E^2\sigma_1^2+\sigma_E^2\sigma_2^2+\sigma_1^2\sigma_2^2}$$

Therefore,

$$\mathop{\mathbb{E}}_{A_1A_2} G(p_{a_1a_2}) = c_\lambda\left(\frac{\sigma_E^4(\sigma_1^2+\sigma_2^2)}{\sigma_E^2\sigma_1^2+\sigma_E^2\sigma_2^2+\sigma_1^2\sigma_2^2}\right)^{\lambda/2}$$

Lastly, $\mathbb{E}\, G(p) = c_\lambda(\sigma_E^2)^{\lambda/2}$ Let $\theta = \lambda/2$ and $x = \sigma_E^2, y = \sigma_1^2, z = \sigma_2^2$, then it must be true that for any $x, y, z > 0$,

$$\left(\frac{x^2}{x+y}\right)^\theta + \left(\frac{x^2}{x+z}\right)^\theta \geq \left(\frac{x^2(y+z)}{xy+xz+yz}\right)^\theta$$

or

$$\frac{1}{(x+y)^\theta} + \frac{1}{(x+z)^\theta} \geq \left(\frac{y+z}{xy+xz+yz}\right)^\theta$$

Let $y = z = 1$ and $x \to 0$, we get $2 \geq 2^\theta$ or $\theta \leq 1$, or $\lambda \leq 2$. Now we show that the inequality holds for $\theta \leq 1$. After dividing both sides by $(y+z)^\theta$, this is equivalent to

$$\frac{1}{(y^2+xy+xz+yz)^\theta} + \frac{1}{(z^2+xy+xz+yz)^\theta} \geq \frac{1}{(xy+xz+yz)^\theta}$$

Normalize $xy + xz + yz$ to 1. We want to show that $(y^2+1)^{-\theta} + (z^2+1)^{-\theta} \geq 1$. We have $yz \leq xy + xz + yz = 1$, so

$$(y^2+1)^{-\theta} + (z^2+1)^{-\theta} \geq \left(\frac{1}{y^2+1}\right)^{\theta} + \left(\frac{y^2}{y^2+1}\right)^{\theta}$$

To finish to proof, we have to show that for $0 < t < 1$, $h(t) := t^{\theta} + (1-t)^{\theta} \geq 1$. We calculate $f'(t) = \theta t^{\theta-1}\left(1 - \left(\frac{t}{1-t}\right)^{1-\theta}\right)$, so $f'(t) \geq 0$ for $0 \leq t \leq \frac{1}{2}$, and $f'(t) \leq 0$ for $\frac{1}{2} \leq t \leq 1$. Since $f(0) = f(1) = 1$, we are done. □

Chapter 5

Discussion and Future Work

This work makes two main contributions. The first contribution in Chapter 3 is a characterization of universal complements. The second contribution in Chapter 4 is an investigation into designing the score function to make signals in a given information structure become substitutes. We believe that the work in Chapter 4 is a start to a fruitful research direction into structure and design of information substitutes. We list a few such directions below.

5.1 Future Work: Informational Substitutes

5.1.1 Designing G with Oracle Query on Information Structure

In Section 4.2, we initiate the study of designing the convex expected score function G for a given information structure. We assume that the entire information structure needs to be fed into the algorithm. For tractability, we therefore assume throughout that the information structure has a *compact representation*. This is not the only possible type of data access.

Since the entire information structure is a big object, a more commonly used type of data access is via an *oracle query*. The algorithm can ask for some data about the information structure in a query; precisely how depends on the modeler. There are different ways to design the oracle query, and each gives differing amounts of power to the query. We want to see whether we can design a substitute-inducing G in a polynomial number of queries, or show that it is not possible. The following queries are examples.

- Given a realization $A_1 = a_1, \dots, A_n = a_n$ of signals and $E = e$ of outcome, return $\pi_{e,a_1,\dots,a_n} := \mathbb{P}(E = e | A_1 = a_1, \dots, A_n = a_n)$. This is the weakest possible query under the assumption that the information structure is perfectly known. This is equivalent (up to a constant factor) of returning a distribution of E given $A_1 = a_1, \ldots, A_n = a_n$. Throughout Section 4.2, we only use this type of information packaged in a compact form. We believe, but do not yet have a proof, that a polynomial number of queries of this type is not sufficient to determine a G in the general case, or even to check substitutability given a fixed G.

Some proof ideas might be gleaned from Hatfield et al. (2012). This work tests substitutability in matching markets rather than prediction markets, but substitutability in both markets have connections to submodular set functions. Analogous to their work, we might be able to show that violation of substitutability can be "local" in the sense that there are very few substitutability violations relative to the number of subsets of signals.[1]

- Given a subset $S \subseteq \{1, 2, \ldots, n\}$ and a realization $(A_i = a_i)_{i \in S}$ for a subset of signals, return a distribution of E given $(A_i = a_i)_{i \in S}$. This oracle query seems quite reasonable, and most amenable to checking weak substitutability using Definition 2.15. Yet this query is quite powerful; it takes an exponential number of queries of the first bullet point type to implement this query, summing over possible realizations of signals outside of S. If, given access to this query, we still need a super-polynomial number of queries, we will be much confident that the same is true for any reasonable query on the information structure.

- For moderate and strong substitutes, we need an oracle that treats each signal A_i not as a chunk but as a piece of information with internal structure. However, this task seems intractable without further assumptions on the internal structure of the signals.

Even a weaker version of this question is currently unsolved: given an information structure and a G, how can we check efficiently whether this G makes the signals substitutes. A relaxation of this idea would be that we can find an easy-to-check (and hopefully intuitive and interesting) sufficient condition that implies substitutability. The information monotonicity approach (Kong and Schoenebeck, 2018b) might be helpful. However, we have yet to make progress in this direction. Moreover, if substitutability violation is indeed "local," then substitutability itself might be quite fragile, and it is unlikely that another sufficient condition (which should be rather global because it is easy to check) can imply it.

[1] We thank Alex Wei (private communication) for this observation.

5.1.2 Designing G with Incomplete Data on Information Structure

In the previous subsection, we assume that the information structure is perfectly known, only that we access it through an oracle. This assumption is unrealistic in most settings, where we do not have a highly detailed knowledge about the information structure. Rather, we know something, but not everything about the information structure. The research question becomes: can we design a substitute-inducing G given this imperfect information? The strongest notion is that the G induces substitutability for any information structure that agrees with the part we know, while the part we do not know are completed arbitrarily. We believe that there is at least a conceptual connection between this framework and the nascent literature on *robust* optimization (see, *e.g.*, Chen et al. (2017)). A more relaxed notion might be that the designer has a prior distribution over the unknown parts of the information structure, and we want G to induce substitutes with high probability. It is probably advisable to first abstract away from computational costs and focus on the geometric and information theoretic aspects of the problem. After all, it is not even clear that designing such a G is possible.

5.1.3 Continuous Distributions and Distribution Properties

While we take the approach of designing G in Section 4.2, in Section 4.3 and 4.4 we focus on G that has specific functional forms and the signals that are gaussian rather than general. One reason is that such information structures are commonly used in economic modeling modeling and thus we see it worthwhile to investigate them. However, another reason is that in continuous regimes, standard techniques, such as linear programming, no longer work. Standard ways to describe the signal distributions also no longer work. We bypass this problem by focusing on gaussian distributed signals and outcome and do analysis on parameters of the distribution. However, this approach lacks generalities that we desire in a good theory. We believe this is a fundamental obstacle that needs to be directly addressed.

5.2 Future Work: Market Games More Broadly

5.2.1 Equilibrium Computation in Prediction Market Games

There are separate works investigating equilibria of prediction market games under different special cases, the results of which are unified under the "all-rush" and "all-delay" results of informational substitutes and complements (Chen and Waggonner, 2016). However, signals in most information structures are neither fully substitutes nor complements, and the theory developed in this work does not apply. Nevertheless, it might still be possible to use the informational substitutes idea by breaking signals into the substitutes part and complements part. Intuitively, agents should rush to reveal the substitutes part and delay the complements part until their last trades. However, formal models of this phenomenon have not yet been developed. Somewhat orthogonally, we want to develop a general algorithm that computes an equilibrium given any general information structure and trade order. As a first step in this direction, Kong and Schoenebeck (2018a) fully analyzes the equilibrium of the Alice Bob Alice (A-B-A) game, which is the main building block of our notion of informational substitutes.

5.2.2 Informational Substitutes in General Bayesian Games

In this work, we consider informational substitutes in prediction market settings where there are no interaction between agents within a round. In each round of the prediction market game, only one agent is selected to interact with the market (either a set of securities with fixed prices, or an automated market maker) which is deterministic. The interactions between agents only occur between rounds, in which the behaviors of agents in previous rounds affect the market condition of the agent in current round.

In many settings of interest, such as auctions and financial markets, several agents act simultaneously in the same round and need to take strategic behaviors of other agents into account. It is not yet clear how our notion of informational substitutes can be modified in such settings, and there are some fundamental obstacles that do not occur in our setting. For example, it is true in our setting that more information is always better; this is not generally true in multi-player settings. In multi-player settings, we need to consider both *informational* and *strategic* substitutes. Milgrom and Weber (1982) propose a definition of informational substitutes in a one-period common-value auctions, but their formulations are very different from ours. A new definition of informational substitutes in the two-sided market of Kyle (1985) will also be interesting.

5.2.3 Prediction Markets with Non-Myopic Boundedly Rational Agents

Prediction Markets are often designed using (strictly) proper scoring rules, which only guarantee that agents report the truth if they are myopic. On the other hand, as alluded to in Subsection 5.2.1, previous results that go beyond myopic agents assume that agents are fully rational. This assumption might be inappropriate if the games are long and the trade orders not fully known. To the best of our knowledge, no theoretical work has been done on analyzing prediction market games when agents follow behavioral rules such as no-regret learning. Analyzing such games using standard techniques such as the

Multiplicative Weights Update method (Arora et al., 2012) might be tractable and give us insights into dynamics of prediction markets in realistic settings.

Appendix A

Full Proofs

A.1 Full Proofs from Chapter 2

A.1.1 Proof of Proposition 2.22

Proof. We will prove this theorem when p and q each take an arbitrary finite number of values. The case where they take infinite amounts of values is achieved by the standard $\epsilon - \delta$ argument of passing to the limit.

Assume that there are n possible values that a distribution in the domain of G could take, a_i for $1 \leq i \leq n$. q takes values a_i with probability q_i and p takes values a_i with probability p_i, for each i. Then we have

$$G(p) = -\mathrm{Var}(p) = -\sum_i a_i^2 p_i^2 + \left(\sum_i a_i p_i\right)^2$$

and similarly for q. The formula for Bregman divergence is

$$D_G(p, q) = G(p) - G(q) - \langle G'(q), p - q \rangle$$

We compute the inner product in the last term

$$\begin{aligned}\langle G'(q), p-q\rangle &= \sum_i (p_i - q_i)\frac{\partial}{\partial q_i}G(q) \\ &= \sum_i (p_i - q_i)\left[-2a_i^2 q_i + 2a_i\left(\sum_k a_k q_k\right)\right] \\ &= -2\sum_i a_i^2 q_i(p_i - q_i) + 2\left(\sum_i a_i q_i\right)\left(\sum_i a_i(p_i - q_i)\right) \\ &= -2\sum_i a_i^2 p_i q_i + 2\sum_i a_i^2 q_i^2 + 2\left(\sum_i a_i p_i\right)\left(\sum_i a_i q_i\right) - 2\left(\sum_i a_i q_i\right)^2\end{aligned}$$

so

$$\begin{aligned}&D_G(p,q) \\ &= -\sum_i a_i^2 p_i^2 + \left(\sum_i a_i p_i\right)^2 + \sum_i a_i^2 q_i^2 + \left(\sum_i a_i q_i\right)^2 + 2\sum_i a_i^2 p_i q_i - 2\sum_i a_i^2 q_i^2 \\ &\quad - 2\left(\sum_i a_i p_i\right)\left(\sum_i a_i q_i\right) \\ &= -\sum_i a_i^2 p_i^2 + \left(\sum_i a_i p_i\right)^2 - \sum_i a_i^2 q_i^2 + \left(\sum_i a_i q_i\right)^2 + 2\sum_i a_i^2 p_i q_i - 2\left(\sum_i a_i p_i\right)\left(\sum_i a_i q_i\right) \\ &= -\left(\sum_i a_i p_i\right)^2 - \left(\sum_i a_i q_i\right)^2 + 2\left(\sum_i a_i p_i\right)\left(\sum_i a_i q_i\right) \\ &= \left(\sum_i a_i(p_i - q_i)\right)^2 - \sum_i a_i^2 (p_i - q_i)^2 \\ &= -\mathrm{Var}(p-q)\end{aligned}$$

□

A.2 Full Proofs from Chapter 3

A.2.1 Proof of Theorem 3.10

Proof. We first show this result for $n = 2$.

We show that E cannot take more than 2 values. Note that if we can prove this result for binary signals A_1, A_2 then we can extend the result to signals that take multiple values by letting all but two of the prior marginal probabilities go to zero, making such signals arbitrarily approximate binary signals and invoke continuity of G. Assume for the sake of contradiction that E takes $d \geq 3$ values. The arguments to the G are in the $(d-1)$-simplex. Let G be such that the value of G at the boundary of the $(d-1)$-simplex be 0, and the value of G at the prior be -1, so the graph of G looks like a pyramid. Any conditional distribution p_S for a nontrivial subset of signals S must live on the edge of the simplex, since after conditioning on a variable, there is only one free variable left, and that variable is binary, so it can take only 2 values, but the $(d-1)$-simplex has $d > 2$ arguments, so some of the arguments must be 0, and p_S is on the edge of the simplex and $G(p_S) = 0$. The substitutes inequality $G(p) + \mathbb{E}_{a_1 a_2} G(p_{a_1 a_2}) \geq \mathbb{E}_{a_1} G(p_{a_1}) + \mathbb{E}\, G(p_{a_2})$ therefore reduces to $G(p) \geq 0$ which is false.

Now that in the case $n = 2$ we know that E is binary, we will next show that E is $A_1 \oplus A_2$ or $\neg(A_1 \oplus A_2)$.

For notational convenience list the value of E when $(A_1, A_2) = (0,0), (0,1), (1,0), (1,1)$ consecutively, for example, 0001 represents the setup where $E = 1$ if $A_1 = A_2 = 1$ and 0 otherwise. The following are allowed: 0110,1001,0011,1100,0101,1010,0000,1111 (counting those with trivial signals as well). These are all the combinations with 1s appearing 0,2 or 4 times. We want to rule out the cases where 1s appear 1 or 3 times. By symmetry (flipping every 0 to 1 and vice versa) it suffices to consider the case where 1s appear 1 time. Appropriate renaming of indices reduces us to 0001. Write $\pi_{a_1 a_2}$ as the prior probability of $(A_1, A_2) = (a_1, a_2)$, the complements inequality is

$$G(\pi_{11}) \geq (\pi_{10} + \pi_{11}) G\left(\frac{\pi_{11}}{\pi_{10} + \pi_{11}}\right) + (\pi_{01} + \pi_{11}) G\left(\frac{\pi_{11}}{\pi_{01} + \pi_{11}}\right)$$

Corollary 3.6 implies that this inequality is true for every convex G if and only if the inequality 3.4 is true. The inequality reduces to $\pi_{11} - 1 \geq -\pi_{10} - \pi_{01}$ which is false under the assumption $\pi_{00} > 0$.

We therefore have proved that for $n = 2$, E must be a XOR of signals (or its negation). Now we will use this characterization for $n = 2$ to extend to all n.

We first show that if signals are nontrivial, every signal is binary. Assume for the sake of contradiction that a signal, say A_1, takes at least 3 values $0, 1, 2$. Let $E(a_1, a_2)$ be the value of E with $A_1 = a_1, A_2 = a_2$ and $A_3, \ldots, A_n$ take some fixed values. Consider the case where A_1 takes values in $\{0, 1\}$ and A_2 in $\{0, 1\}$. We can do this by considering an information structure with probabilities of other values very small and invoke continuity of G. By the $n = 2$ characterization, E on these values must be a XOR. Without loss of generality, let $E(0,0) = E(1,1) = 0$ and $E(0,1) = E(1,0) = 1$. Now consider A_1 in $\{0, 2\}$ and A_2 in $\{0, 1\}$. Since $E(0,0) = 0$ and $E(0,1) = 1$, the only possible XOR continuation is $E(2,0) = 1$ and $E(2,1) = 0$. However, when we consider A_1 in $\{1, 2\}$

and A_2 in $\{0,1\}$, the values $E(1,0) = E(2,0) = 1, E(1,1) = E(2,1) = 0$ make A_1 trivial, a contradiction. Therefore, every signal is binary.

Now we show that E is a XOR of all n signals or its negation. We show this by induction on n. The case $n = 1$ is evident and $n = 2$ is proven. Now assume that it is true for $n - 1$. Then E conditional on $A_1 = 0$ and $A_2, \ldots, A_n$ is $A_2 \oplus \cdots \oplus A_n$ or $\neg(A_2 \oplus \cdots \oplus A_n)$. Similarly, E conditional on $A_1 = 1$ and $A_2, \ldots, A_n$ is $A_2 \oplus \cdots \oplus A_n$ or $\neg(A_2 \oplus \cdots \oplus A_n)$. If both are the same, either $A_2 \oplus \cdots \oplus A_n$ or $\neg(A_2 \oplus \cdots \oplus A_n)$, then A_1 is trivial, which is not allowed. If $E|A_1 = 0, A_2, \ldots, A_n = A_2 \oplus \cdots \oplus A_n$ and $E|A_1 = 1, A_2, \ldots, A_n = \neg(A_2 \oplus \cdots \oplus A_n)$, then $E|A_1, \ldots, A_n = A_1 \oplus A_2 \oplus \cdots \oplus A_n$. If $E|A_1 = 0, A_2, \ldots, A_n = \neg(A_2 \oplus \cdots \oplus A_n)$ and $E|A_1 = 1, A_2, \ldots, A_n = (A_2 \oplus \cdots \oplus A_n)$, then $E|A_1, \ldots, A_n = \neg(A_1 \oplus A_2 \oplus \cdots \oplus A_n)$. The inductive hypothesis is proved, so we show the first part of the theorem, that E must be the XOR of signals.

We now prove the second part of the theorem, that XOR of signals are universal complements.

The case $n = 1$ is evident. We first prove the case $n = 2$. Pick any convex G and let all probability distributions on E be represented as scalars $q \in [0,1]$, the probability that $E = 1$. By Proposition 3.4, we can scale G such that $G(0) = G(1) = 0$. We must have $G(x) \leq 0$ for all $0 \leq x \leq 1$. For every $a_1, a_2 \in \{0,1\}$, let $\pi_{a_1a_2}$ be the prior probability that $A_1 = a_1$ and $A_2 = a_2$. We can compute

$$\begin{aligned}
\mathbb{E}\, G(p_{a_1a_2}) &= (\pi_{00} + \pi_{11})G(0) + (\pi_{01} + \pi_{10})G(1) = 0 \\
\mathbb{E}\, G(p_{a_1}) &= (\pi_{00} + \pi_{01})G\left(\frac{\pi_{01}}{\pi_{00} + \pi_{01}}\right) + (\pi_{10} + \pi_{11})G\left(\frac{\pi_{10}}{\pi_{10} + \pi_{11}}\right) \\
\mathbb{E}\, G(p_{a_2}) &= (\pi_{00} + \pi_{10})G\left(\frac{\pi_{10}}{\pi_{00} + \pi_{10}}\right) + (\pi_{01} + \pi_{11})G\left(\frac{\pi_{01}}{\pi_{01} + \pi_{11}}\right) \\
\mathbb{E}\, G(p) &= G(\pi_{01} + \pi_{10})
\end{aligned}$$

Therefore, we want to show that for any convex G,

$$\begin{aligned}
&\mathbb{E}\, G(p_{a_1a_2}) + \mathbb{E}\, G(p) \geq \mathbb{E}\, G(p_{a_1}) + \mathbb{E}\, G(p_{a_2}) \\
&\qquad G(\pi_{01} + \pi_{10}) \geq (\pi_{00} + \pi_{01})G\left(\frac{\pi_{01}}{\pi_{00} + \pi_{01}}\right) + (\pi_{10} + \pi_{11})G\left(\frac{\pi_{10}}{\pi_{10} + \pi_{11}}\right) \\
&\qquad\qquad + (\pi_{00} + \pi_{10})G\left(\frac{\pi_{10}}{\pi_{00} + \pi_{10}}\right) + (\pi_{01} + \pi_{11})G\left(\frac{\pi_{01}}{\pi_{01} + \pi_{11}}\right)
\end{aligned}$$

Note that if the denominator of a term is zero then the coefficient of that term is also zero so the term vanishes.

This inequality is in the form of Corollary 3.6, so we only need to check the inequality of the form (3.4) to finish the proof. We consider two cases.

Case 1 $\pi_{00}\pi_{10} \geq \pi_{01}\pi_{11}$. Observe that $(\pi_{10}+\pi_{01})(\pi_{00}+\pi_{01})-\pi_{01}(\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11}) = \pi_{00}\pi_{10} - \pi_{01}\pi_{11}$ so $\pi_{10}+\pi_{01} \geq \frac{\pi_{01}}{\pi_{00}+\pi_{01}}$ Similarly, we can show that

$$\begin{aligned}
\frac{\pi_{10}}{\pi_{10}+\pi_{01}} &\geq \pi_{10}+\pi_{01} \\
\frac{\pi_{01}}{\pi_{01}+\pi_{11}} &\geq \pi_{10}+\pi_{01} \\
\pi_{10}+\pi_{01} &\geq \frac{\pi_{01}}{\pi_{00}+\pi_{01}} \\
\pi_{10}+\pi_{01} &\geq \frac{\pi_{10}}{\pi_{00}+\pi_{10}}
\end{aligned}$$

The inequality to be proved reduces to

$$\begin{aligned}
&(\pi_{00}+\pi_{01})\frac{\pi_{01}}{(\pi_{00}+\pi_{01})(\pi_{10}+\pi_{01})} + (\pi_{10}+\pi_{11})\frac{\pi_{11}}{(\pi_{10}+\pi_{11})(\pi_{00}+\pi_{11})} \\
&+ (\pi_{00}+\pi_{10})\frac{\pi_{10}}{(\pi_{00}+\pi_{10})(\pi_{10}+\pi_{01})} + (\pi_{01}+\pi_{11})\frac{\pi_{11}}{(\pi_{01}+\pi_{11})(\pi_{00}+\pi_{11})} \\
&= 1 + \frac{2\pi_{11}}{\pi_{00}+\pi_{11}} \geq 1
\end{aligned}$$

which is immediate.

Case 2 $\pi_{00}\pi_{10} \leq \pi_{01}\pi_{11}$. Similarly, we can show that

$$\begin{aligned}
\frac{\pi_{10}}{\pi_{10}+\pi_{01}} &\leq \pi_{10}+\pi_{01} \\
\frac{\pi_{01}}{\pi_{01}+\pi_{11}} &\leq \pi_{10}+\pi_{01} \\
\pi_{10}+\pi_{01} &\leq \frac{\pi_{01}}{\pi_{00}+\pi_{01}} \\
\pi_{10}+\pi_{01} &\leq \frac{\pi_{10}}{\pi_{00}+\pi_{10}}
\end{aligned}$$

and the proof goes down as in Case 1. So the proof for $n = 2$ case is finished.

Now we prove by induction on n that the statement is true for all n. We have proved the base step $n = 1, 2$. Now assume that it is true for up to n and we will show it for $n+1$. We want to show that in the $[n+1]$ universe and $S' \subseteq S$, $T \subset [n+1] \setminus S$,

$$\mathbb{E}\, G(p_{a_i:i \in S \cup T}) - \mathbb{E}\, G(p_{a_i:i \in S}) \geq \mathbb{E}\, G(p_{a_i:i \in S' \cup T}) - \mathbb{E}\, G(p_{a_i:i \in S'})$$

We can check that $|S|, |S'|, |T| \leq n$, so the required inequality follows by inductive hypothesis, with $\oplus_{i \in S'} a_i$ prior, and the fact that $\oplus_{i \in S \setminus S'} A_i$ and $\oplus_{i \in T} A_i$ are universal weak complements, by $n = 2$ result. We are done.

□

A.2.2 Proof of Theorem 3.14

Proof. We first prove the case $n = 2$ and f is boolean. Write $\pi_{a_1 a_2}$ the prior probability that $(A_1, A_2) = (a_1, a_2)$ and $a_{a_1 a_2} := f(A_1 = a_1, A_2 = a_2)$. Consider the following partition

$$\begin{aligned} A' &= \{\{(0,0),(0,1),(1,0),(1,1)\}\} \\ A &= \{\{(0,0)\},\{(0,1),(1,0),(1,1)\}\} \\ B &= \{\{(0,1)\},\{(0,0),(1,0),(1,1)\}\} \end{aligned}$$

What this means is that A' is a null signal; A tells you whether (a_1, a_2) is in $\{(0,0)\}$ or $\{(0,1),(1,0),(1,1)\}$ with prior probabilities π_{00} and $\pi_{01} + \pi_{10} + \pi_{11}$ respectively, and similarly for B. We then have

$$\begin{aligned} B \vee A' &= B \\ B \vee A &= \{\{(0,0)\},\{(0,1)\},\{(1,0),(1,1)\}\} \end{aligned}$$

and

$$\begin{aligned} EG(p_{ba}) &= \pi_{00}G(a_{00}) + \pi_{01}G(a_{01}) + (\pi_{10} + \pi_{11})G\left(\frac{\pi_{10}a_{10} + \pi_{11}a_{11}}{\pi_{10} + \pi_{11}}\right) \\ \mathbb{E}\, G(p_a) &= \pi_{00}G(a_{00}) + (\pi_{01} + \pi_{10} + \pi_{11})G\left(\frac{\pi_{01}a_{01} + \pi_{10}a_{10} + \pi_{11}a_{11}}{\pi_{01} + \pi_{10} + \pi_{11}}\right) \\ \mathbb{E}\, G(p_{ba'}) &= \pi_{01}G(a_{01}) + (\pi_{00} + \pi_{10} + \pi_{11})G\left(\frac{\pi_{00}a_{00} + \pi_{10}a_{10} + \pi_{11}a_{11}}{\pi_{00} + \pi_{10} + \pi_{11}}\right) \\ \mathbb{E}\, G(p_{a'}) &= G(\pi_{00}a_{00} + \pi_{01}a_{01} + \pi_{10}a_{10} + \pi_{11}a_{11}) \end{aligned}$$

We also see that $A' = A \wedge B \preceq A' \preceq A$. If $(A_1, A_2; E)$ are moderate complements then we must have

$$\mathbb{E}\, G(p_{ba}) - \mathbb{E}\, G(p_a) \geq \mathbb{E}\, G(p_{ba'}) - \mathbb{E}\, G(p_{a'})$$

Substituting the expressions above and simplifying gives

$$\begin{aligned} &(\pi_{10} + \pi_{11})G\left(\frac{\pi_{10}a_{10} + \pi_{11}a_{11}}{\pi_{10} + \pi_{11}}\right) + G(\pi_{00}a_{00} + \pi_{01}a_{01} + \pi_{10}a_{10} + \pi_{11}a_{11}) \geq \\ &(\pi_{00} + \pi_{10} + \pi_{11})G\left(\frac{\pi_{00}a_{00} + \pi_{10}a_{10} + \pi_{11}a_{11}}{\pi_{00} + \pi_{10} + \pi_{11}}\right) + (\pi_{01} + \pi_{10} + \pi_{11})G\left(\frac{\pi_{01}a_{01} + \pi_{10}a_{10} + \pi_{11}a_{11}}{\pi_{01} + \pi_{10} + \pi_{11}}\right) \end{aligned}$$

Each function f corresponds to one of the 16 possible values of $(a_{00}, a_{01}, a_{10}, a_{11})$. For notational convenience list the value of E when $(A_1, A_2) = (0,0),(0,1),(1,0),(1,1)$

consecutively; for example, 0001 represents the information structure where $E = 1$ if $A_1 = A_2 = 1$ and 0 otherwise.

We will show that 0100, 0101, 0110, 0111 are not permissible by choosing an appropriate counterexample G that violates the inequality. We will assume throughout that $G(0) = G(1) = 0$.

0100

$$G(\pi_{01}) \geq \pi_{00}G(0) + (\pi_{01} + \pi_{10} + \pi_{11})G\left(\frac{\pi_{01}}{\pi_{01} + \pi_{10} + \pi_{11}}\right)$$

This is false for any strictly convex G: Jensen's inequality in the wrong sign.

0101

$$\begin{aligned}(\pi_{10} + \pi_{11})G\left(\frac{\pi_{11}}{\pi_{10} + \pi_{11}}\right) + G(\pi_{01} + \pi_{11}) \geq \\ (\pi_{00} + \pi_{10} + \pi_{11})G\left(\frac{\pi_{11}}{\pi_{00} + \pi_{10} + \pi_{11}}\right) + (\pi_{01} + \pi_{10} + \pi_{11})G\left(\frac{\pi_{01} + \pi_{11}}{\pi_{01} + \pi_{10} + \pi_{11}}\right)\end{aligned}$$

We can check that $\frac{\pi_{11}}{\pi_{00}+\pi_{10}+\pi_{11}} < \frac{\pi_{01}+\pi_{11}}{\pi_{01}+\pi_{10}+\pi_{11}}$, so the two elements in G of the right hand side are not equal. So we can choose G such that the two G terms in the left hand side are equal $G\left(\frac{\pi_{11}}{\pi_{10}+\pi_{11}}\right) = G(\pi_{01} + \pi_{11})$ (such a convex G with that property always exists). Then with strictly convex G we can say that some linear combination of two Gs on the right hand side are strictly greater than the first G on the left hand side (we get strictly greater because the two things inside G of the right hand side are not equal - this is important), and also some other linear combination of the two Gs on the right hand side are greater than the second G, and if we combine the two inequalities appropriately we can get that the right hand side is greater than some linear combination of the two Gs on the left hand side - they probably have different coefficients but everything we do here is normalized so the sum of the coefficients are the same and since we assume the two Gs have equal value, we get that the right hand side is greater than the left hand side. We can also do the calculation explicitly. We have

$$\alpha G\left(\frac{\pi_{11}}{\pi_{00} + \pi_{10} + \pi_{11}}\right) + \beta G\left(\frac{\pi_{01} + \pi_{11}}{\pi_{01} + \pi_{10} + \pi_{11}}\right) > G\left(\frac{\pi_{11}}{\pi_{10} + \pi_{11}}\right)$$
$$\gamma G\left(\frac{\pi_{11}}{\pi_{00} + \pi_{10} + \pi_{11}}\right) + \delta G\left(\frac{\pi_{01} + \pi_{11}}{\pi_{01} + \pi_{10} + \pi_{11}}\right) > G(\pi_{01} + \pi_{11})$$

with $\alpha + \beta = 1, \gamma + \delta = 1$ given by

$$\alpha = \frac{\pi_{01}\pi_{10}(\pi_{00} + \pi_{10} + \pi_{11})}{(\pi_{10} + \pi_{11})(\pi_{00}\pi_{01} + \pi_{01}\pi_{10} + \pi_{00}\pi_{11})}, \quad \beta = \frac{\pi_{00}\pi_{11}(\pi_{01} + \pi_{10} + \pi_{11})}{(\pi_{10} + \pi_{11})(\pi_{00}\pi_{01} + \pi_{01}\pi_{10} + \pi_{00}\pi_{11})}$$
$$\gamma = \frac{\pi_{00}(\pi_{01} + \pi_{11})(\pi_{00} + \pi_{10} + \pi_{11})}{\pi_{00}\pi_{01} + \pi_{01}\pi_{10} + \pi_{00}\pi_{11}}, \delta = 1 - \gamma = \frac{\pi_{01}(\pi_{10} + \pi_{00}\pi_{01} + \pi_{00}\pi_{11})}{\pi_{00}\pi_{01} + \pi_{01}\pi_{10} + \pi_{00}\pi_{11}}$$

Now we multiply the first inequality by λ, the second by μ, then add, such that λ and μ satisfy

$$\alpha\lambda + \gamma\mu = \frac{(\pi_{00} + \pi_{10} + \pi_{11})}{(\pi_{00} + \pi_{10} + \pi_{11}) + (\pi_{01} + \pi_{10} + \pi_{11})}, \quad \lambda + \mu = 1$$

then rearrange to get what we want.

0101

$$(\pi_{10} + \pi_{11})G\left(\frac{\pi_{10}}{\pi_{10} + \pi_{11}}\right) + G(\pi_{01} + \pi_{10}) \geq$$
$$(\pi_{00} + \pi_{10} + \pi_{11})G\left(\frac{\pi_{10}}{\pi_{00} + \pi_{10} + \pi_{11}}\right) + (\pi_{01} + \pi_{10} + \pi_{11})G\left(\frac{\pi_{01} + \pi_{10}}{\pi_{01} + \pi_{10} + \pi_{11}}\right)$$

We can check that $\frac{\pi_{10}}{\pi_{00}+\pi_{10}+\pi_{11}} < \frac{\pi_{01}+\pi_{10}}{\pi_{01}+\pi_{10}+\pi_{11}}$. The rest proceeds exactly the same as the previous case.

0111

$$G(\pi_{01} + \pi_{10} + \pi_{11}) \geq (\pi_{00} + \pi_{10} + \pi_{11})G\left(\frac{\pi_{10} + \pi_{11}}{\pi_{00} + \pi_{10} + \pi_{11}}\right) + \pi_{01}G(1)$$

This is just Jensen with the wrong sign, same argument as 0100.

So we show that 0100, 0101, 0110, 0111 are not permissible. By symmetry, or analogously, we can also show (with A' null, A separates $(0,0)$ from the rest, and B separates $(1,0)$ from the rest) that 0010, 0011, 0110, 0111 are not permissible, so among the 8 combinations that start with 0, we are left only with 0000 and 0001. We will also show that 0001 is not permissible with the following signals: A' null, A separates $(1,1)$ from the rest, and B separates $(1,0)$ from the rest, and $G(0) = G(1) = 1$ we get

$$G(\pi_{11}) \geq (\pi_{11} + \pi_{01} + \pi_{00})G\left(\frac{\pi_{11}}{\pi_{11} + \pi_{01} + \pi_{00}}\right)$$

with $\pi_{11} < \frac{\pi_{11}}{\pi_{11}+\pi_{01}+\pi_{00}}$ this is the convex G inequality with the wrong sign.

Therefore, we show that for a sequence starting with 0, only 0000 is admissible. By symmetry, or analogously, for a sequence starting with 1, only 1111 is admissible. So we proved that in the case $n = 2$, f has to be a constant function.

Extending the case $n = 2$ to general n, f boolean, is straightforward. Assume that it is true up to $n - 1$, then a function $f(A_1, \ldots, A_n)$ of n variables must be constant when you fix the A_n, so $f(A_1, \ldots, A_n)$ depends only on A_n, but similarly it depends only on A_{n-1} so it depends on neither, that is, a constant function.

Now we extend this to $E = f(A_1, \ldots, A_n)$ a general function. E takes a finite number of values because there are finitely many possible inputs $A_1, \ldots, A_n$. If E takes 2^k values, then we can write $E = (E_1, \ldots, E_k)$. We must have $E_j = f_j(A_1, \ldots, A_n)$ a deterministic function of A_i, and $(A_1, \ldots, A_n; E_j)$ are universal complements, so E_j is a constant for all j, so E is a constant.

We can further extend this result to signals A_i that take multiple finite number of values. We can replace any A_i with $A_{i1}, \ldots, A_{it_i}$ with A_{ij} binary, say, map any possible value of A_i into binary numbers from 0 up to a finite number less than 2^{t_i}, then A_{ij} is the jth digit of the binary representation of A_i. We are done.

□

A.3 Full Proofs from Chapter 4

A.3.1 Proof of Proposition 4.14

Proof. 1. We compute

$$\mathop{\mathbb{E}}_{A_0} \text{Var}(E|A_0) = \mathop{\mathbb{E}}_{A_1} \text{Var}(E|A_1) = \frac{\pi_E(1-\pi_E)(1-\pi)\pi}{(1-\pi)\pi_E + \pi(1-\pi_E)} + \frac{\pi_E(1-\pi_E)(1-\pi)\pi}{\pi\pi_E + (1-\pi)(1-\pi_E)}$$

$$\mathbb{P}(E=1|A_0=1, A_1=1) = \frac{\pi_E(1-\pi)^2}{\pi_E(1-\pi)^2 + (1-\pi_E)\pi^2}$$

$$\mathbb{P}(E=1|A_0=1, A_1=0) = \mathbb{P}(E=1|A_0=0, A_1=1) = \frac{\pi_E(1-\pi)\pi}{\pi_E(1-\pi)\pi + (1-\pi_E)(1-\pi)\pi} = \pi_E$$

$$\mathbb{P}(E=1|A_0=0, A_1=0) = \frac{\pi_E\pi^2}{\pi_E\pi^2 + (1-\pi_E)(1-\pi)^2}$$

so

$$\begin{aligned}\mathop{\mathbb{E}}_{A_0,A_1} \text{Var}(E|A_0, A_1) = {} & \pi_E(1-\pi_E)(1-\pi)^2\pi^2 \left(\frac{1}{(1-\pi)^2\pi_E + \pi^2(1-\pi_E)} + \frac{1}{\pi_E\pi^2 + (1-\pi_E)(1-\pi)^2}\right) \\ & + 2\pi_E(1-\pi_E)(1-\pi)\pi\end{aligned}$$

The inequality reduces to

$$\begin{aligned}1 + 2\pi(1-\pi) + (1-\pi)^2\pi^2 \left(\frac{1}{(1-\pi)^2\pi_E + \pi^2(1-\pi_E)} + \frac{1}{\pi_E\pi^2 + (1-\pi_E)(1-\pi)^2}\right) \geq \\ 2\pi(1-\pi)\left(\frac{1}{(1-\pi)\pi_E + \pi(1-\pi_E)} + \frac{1}{\pi_E\pi + (1-\pi_E)(1-\pi)}\right)\end{aligned}$$

A counterexample is $(\pi_E, \pi) = (0.99, 0.49)$

2. We use the following notation $\bar{\pi} = 1 - \pi$ for $\pi \in \{\pi_E, \pi_1, \pi_2\}$.

We want to show that the following inequality has a counterexample

$$\mathrm{Var}(E) + \mathop{\mathbb{E}}_{A_1,A_2} \mathrm{Var}(E|A_1, A_2) \geq \mathop{\mathbb{E}}_{A_1} \mathrm{Var}(E|A_1) + \mathop{\mathbb{E}}_{A_2} \mathrm{Var}(E|A_2)$$

First, $\mathrm{Var}(E) = \pi_E \bar{\pi}_E$. Now, both A_1 and A_2 take value in $\{0, 1, 2\}$ but if either of them takes value 0 or 2 we know the value E for certain (if one of them takes value 0, then $E = 0$, and if one of them takes value 2, then $E = 1$). The only variability in E given A_1 and A_2 comes when $A_1 = A_2 = 1$, so

$$\begin{aligned} \mathop{\mathbb{E}}_{A_1,A_2} \mathrm{Var}(E|A_1, A_2) &= \mathbb{P}(A_1 = 1, A_2 = 1)\mathrm{Var}(E|A_1 = 1, A_2 = 1) \\ &= \mathbb{P}(A_1 = 1, A_2 = 1)\, \mathbb{P}(E = 1|A_1 = 1, A_2 = 1)(1 - \mathbb{P}(E = 1|A_1 = 1, A_2 = 1)) \end{aligned}$$

Now, $\mathbb{P}(A_1 = 1, A_2 = 1) = \pi_E \bar{\pi}_1 \bar{\pi}_2 + \bar{\pi}_E \pi_1 \pi_2$ and $\mathbb{P}(E = 1|A_1 = 1, A_2 = 1) = \bar{\pi}_E \pi_1 \pi_2 / \mathbb{P}(A_1 = 1, A_2 = 1)$ so

$$\mathop{\mathbb{E}}_{A_1,A_2} \mathrm{Var}(E|A_1, A_2) = \frac{\pi_E \bar{\pi}_1 \bar{\pi}_2 \cdot \bar{\pi}_E \pi_1 \pi_2}{\pi_E \bar{\pi}_1 \bar{\pi}_2 + \bar{\pi}_E \pi_1 \pi_2}$$

Similarly,

$$\begin{aligned} \mathop{\mathbb{E}}_{A_1} \mathrm{Var}(E|A_1) &= \mathbb{P}(A_1 = 1)\mathrm{Var}(E|A_1 = 1) \\ &= (\pi_E \bar{\pi}_1 + \bar{\pi}_E \pi_1) \frac{\pi_E \bar{\pi}_1}{\pi_E \bar{\pi}_1 + \bar{\pi}_E \pi_1} \frac{\bar{\pi}_E \pi_1}{\pi_E \bar{\pi}_1 + \bar{\pi}_E \pi_1} \\ &= \frac{\pi_E \bar{\pi}_1 \cdot \bar{\pi}_E \pi_1}{\pi_E \bar{\pi}_1 + \bar{\pi}_E \pi_1} \\ \mathop{\mathbb{E}}_{A_2} \mathrm{Var}(E|A_2) &= \frac{\pi_E \bar{\pi}_2 \cdot \bar{\pi}_E \pi_2}{\pi_E \bar{\pi}_2 + \bar{\pi}_E \pi_2} \end{aligned}$$

The inequality reduces to

$$1 + \frac{\pi_1 \pi_2 \bar{\pi}_1 \bar{\pi}_2}{\pi_E \bar{\pi}_1 \bar{\pi}_2 + \bar{\pi}_E \pi_1 \pi_2} \geq \frac{\pi_1 \bar{\pi}_1}{\pi_E \bar{\pi}_1 + \bar{\pi}_E \pi_1} + \frac{\pi_2 \bar{\pi}_2}{\pi_E \bar{\pi}_2 + \bar{\pi}_E \pi_2}$$

A counterexample to this is $(\pi_E, \pi_1, \pi_2) = (0.95, 0.8, 0.8)$.

□

Bibliography

Jacob D. Abernethy and Rafael M. Frongillo. A characterization of scoring rules for linear properties. *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*, 2012. URL http://proceedings.mlr.press/v23/abernethy12/abernethy12.pdf.

Jr. Alexander S. Kelso and Vincent P. Crawford. Job matching, coalition formation, and gross substitutes. *Econometrica*, 1982. URL https://www.jstor.org/stable/1913392.

Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing*, 2012. URL http://tocbeta.cs.uchicago.edu/articles/v008a006/v008a006.pdf.

Susan Athey and Jonathan Levin. The value of information in monotone decision problems. Technical report, 2001. URL https://web.stanford.edu/~jdlevin/Papers/VOI.pdf.

Robert J. Aumann. Agreeing to disagree. *Annals of Statistics*, 1976. URL https://projecteuclid.org/euclid.aos/1176343654.

Christopher Avery. Strategic jump bidding in english auctions. *Review of Economic Studies*, 1998. URL https://pdfs.semanticscholar.org/101d/bbab5d503d595940a85ff79c2488645bcd2f.pdf.

Wolfgang Berndt and Suvrit Sra. Hlawka-popoviciu inequalities on positive definite tensors. *Linear Algebra and its Applications*, 2015. URL https://arxiv.org/abs/1411.0065.

Dimitri P. Bertsekas, Angelia Nedic, and Asuman E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, 2003.

Tilman Börgers, Angel Hernando-Veciana, and Daniel Krhmer. When are signals complements or substitutes? *Journal of Economic Theory*, 2013. URL http://www.wiwi.uni-bonn.de/kraehmer/BHK-JET-REV-IIe.pdf.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Robert Chen, Brendan Lucier, Yaron Singer, and Vasilis Syrgkanis. Robust optimization for non-convex objectives. *Advances in Neural Information Processing Systems (NIPS)*, 2017. URL https://arxiv.org/abs/1707.01047.

Yiling Chen and Bo Waggonner. Informational substitutes. *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2016. URL https://arxiv.org/abs/1703.08636.

Yiling Chen, Daniel M. Reeves, David M. Pennock, Robin D. Hanson, Lance Fortnow, and Rica Gonen. Bluffing and strategic reticence in prediction markets. *Internet and Network Economics*, 2007. URL http://yiling.seas.harvard.edu/wp-content/uploads/bluffing-long-appendix.pdf.

Yiling Chen, Stanko Dimitrov, Rahul 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*, 2010. URL https://dash.harvard.edu/bitstream/handle/1/5027877/chen_algorithmica_author.pdf?sequence=1.

Stanko Dimitrov and Rahul Sami. http://yiling.seas.harvard.edu/wp-content/uploads/bluffing-long-appendix.pdf. *Proceedings of the 9th ACM conference on Electronic Commerce*, 2008. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.315.4411&rep=rep1&type=pdf.

Eugene F. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 1970. URL https://www.jstor.org/stable/2325486.

Xi Alice Gao, Jie Zhang, and Yiling Chen. What you jointly know determines how you act: strategic interactions in prediction markets. *Proceedings of the Fourteenth ACM Conference on Electronic Commerce (EC)*, 2013. URL https://dash.harvard.edu/bitstream/handle/1/12872200/Gao_ACM'13.pdf?sequence=1.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 2007. URL https://www.stat.washington.edu/raftery/Research/PDF/Gneiting2007jasa.pdf.

Robin Hanson. Combinatorial information market design. *Information Systems Frontiers*, 2003. URL http://mason.gmu.edu/~rhanson/combobet.pdf.

Joel Hasbrouck. *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford University Press, 2007.

John William Hatfield and Paul R. Milgrom. Matching with contracts. *American Economic Review*, 2005. URL https://web.stanford.edu/~milgrom/publishedarticles/Matching%20with%20Contracts-AER.pdf.

John William Hatfield, Nicole Immorlica, and Scott Duke Kominers. Testing substitutability. *Games and Economic Behavior*, 2012. URL http://www.immorlica.com/pubs/testingSubs.pdf.

Ronald A. Howard. Information value theory. *IEEE Transactions on Systems*, 1966.

Narendra Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 1984.

Leonid Genrikhovich Khachiyan. Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 1980.

Yuqing Kong and Grant Schoenebeck. Optimizing bayesian information revelation strategy in prediction markets: the alice bob alice case. *Proceedings of the 9th Innovations in Theoretical Computer Science (ITCS)*, 2018a. URL http://drops.dagstuhl.de/opus/volltexte/2018/8319/pdf/LIPIcs-ITCS-2018-14.pdf.

Yuqing Kong and Grant Schoenebeck. Equilibrium selection in information elicitation without verification via information monotonicity. *Proceedings of the 9th Conference on Innovations in Theoretical Computer Science (ITCS)*, 2018b. URL http://drops.dagstuhl.de/opus/volltexte/2018/8317/pdf/LIPIcs-ITCS-2018-13.pdf.

Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 1985. URL http://people.stern.nyu.edu/lpederse/courses/LAP/papers/Information,Fundamental/Kyle85.pdf.

Nicolas Lambert, David M. Pennock, and Yoav Shoham. Eliciting properties of probability distributions. *Proceedings of the 9th ACM Conference on Economics and Computation (EC)*, 2008. URL http://ai.stanford.edu/~shoham/www%20papers/LambertEtAlEC08elicitability.pdf.

Benny Lehmann, Daniel Lehmann, and Noam Nisan. Combinatorial auctions with decreasing marginal utilities. *Games and Economic Behavior*, 2006. URL https://www.sciencedirect.com/science/article/pii/S089982560500028X.

Marvin Marcus and Henryk Minc. *A Survey of Matrix Theory and Matrix Inequalities*. Dover Publications, 2010. URL https://www.amazon.com/Survey-Matrix-Theory-Inequalities-Mathematics/dp/048667102X.

John McCarthy. Measures of the value of information. *Proceedings of the National Academy of Sciences*, 1956. URL http://www.eecs.harvard.edu/cs286r/courses/fall10/papers/McCarthy56.pdf.

Paul Milgrom and Robert J. Weber. The value of information in a sealed-bid auction. *Journal of Mathematical Economics*, 1982. URL https://www.jstor.org/stable/1911865.

Ethan Mollick and Ramana Nanda. Wisdom or madness? comparing crowds with expert evaluation in funding the arts. *Management Science*, 2015. URL https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2015.2207.

Michael Ostrovsky. Stability in supply chain networks. *American Economic Review*, 2008. URL https://web.stanford.edu/~ost/papers/chains.pdf.

Michael Ostrovsky. Information aggregation in dynamic markets with strategic traders. *Econometrica*, 2012. URL https://web.stanford.edu/~ost/papers/aggregation.pdf.

Leonard J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 1971. URL https://www.jstor.org/stable/2284229.

Philip E. Tetlock and Dan Gardner. *Superforecasting: The Art and Science of Prediction*. Broadway Books, 2016.