Eliciting Properties of Probability Distributions

[Extended Abstract]

Nicolas Lambert*
Computer Science Department
Stanford University
Stanford, CA 94305

David M. Pennock
Yahoo! Research
111 West 40th Street, 17th Floor
New York, NY 10018

Yoav Shoham
Computer Science Department
Stanford University
Stanford, CA 94305

ABSTRACT

We investigate the problem of truthfully eliciting an expert's assessment of a property of a probability distribution, where a property is any real-valued function of the distribution such as mean or variance. We show that not all properties are elicitable; for example, the mean is elicitable and the variance is not. For those that are elicitable, we provide a representation theorem characterizing all payment (or "score") functions that induce truthful revelation. We also consider the elicitation of *sets* of properties. We then observe that properties can always be inferred from sets of elicitable properties. This naturally suggests the concept of *elicitation complexity*; the elicitation complexity of property is the minimal size of such a set implying the property. Finally we discuss applications to prediction markets.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Economics

General Terms

Economics, Theory

Keywords

Scoring rule, prediction market, elicitation, forecasting

1. INTRODUCTION

Asking an expert for a probability assessment, or paying a fixed amount, gives the expert no reward for being accurate or truthful. A *scoring rule* is a prescription for paying an expert that depends on both the expert's report and the actual outcome and rewards accuracy. A scoring rule is *proper* if it maximally rewards truthfulness, meaning that the expert maximizes expected score (payment) by reporting truthfully. Introduced by Brier, Good, and Savage [1, 8, 14], scoring rules have been extensively studied over the past five decades, validated experimentally [12] and applied in a variety of domains [11, 17, 13]. Winkler [19] gives a partial summary of the literature.

Scoring rules elicit probabilities directly. That is, the expert's report takes the form of a full probability distribution. This poses practical difficulties if the distribution is large or complex. First, the expert may not know the full distribution, or its estimation may be infeasible from a computational or data acquisition perspective. Even if he does, the full distribution may be too long to represent and communicate in practice. Or the expert may wish to keep certain aspects of the distribution confidential. At the same time, the person paying for the information may not care about the entire distribution, but only about certain properties of it, such as the mean, variance, or the probability it accords to some specific event.

In this paper, we characterize when a property of a probability distribution is *elicitable*, meaning when it is possible to induce a truthful estimate of the property with a given dimensionality of reports.

We are aware of two prior publications on truthful elicitation of specific properties. Savage's [14] original article proposes an adaptation of scoring rules to elicit expectations. Cervera, Munoz, Gneiting, and Raftery [7, 3] suggest scores that can elicit the quantiles for continuous outcomes. Our results are considerably more general, and the highlights are as follows.

A property is any real-valued function of a probability distribution, and the expert is asked to report a single real number estimate of the property. The property is *directly elicitable* if the expert can be incented to report the true value of the property. Unless explicitly stated, by elicitability we mean direct elicitability.

In Section 3, we ask which properties are elicitable. We show that the elicitable properties are functions whose level sets are convex, or equivalently, functions which can be expressed as linear constraints on the probability distribution. For example, the mean is elicitable but not the variance.

We next ask what payment rules induce truthful reports. We propose a simple characterization of the contracts that are incentive compatible: we will show that any property is uniquely associated with a function, called a *signature*, and that incentive compatible contracts are all positively weighted mixtures of those functions.

We then turn our attention to properties that are not directly elicitable. In Section 4, we generalize the notion of elicitability to sets of properties. For example, the expert

*Part of this work was done while the author was at Yahoo! Research, New York.

EC'08, July 8–12, 2008, Chicago, Illinois, USA.

may be asked to declare a vector of three estimates, and be incented based on her declaration and the realized value of the outcome. It turns out that sets of properties can be elicited even though the individual properties cannot. For example, while the variance is not directly elicitable, the set {mean,variance} is directly elicitable, and so the variance is *indirectly* elicitable. In fact, *any* property is indirectly elicitable, because the entire distribution is directly elicitable, from which any specific property follows. This naturally gives rise to the notion of *elicitation complexity*, discussed in Section 6. A property is of complexity $\mathcal{C}^k$ when the smallest set of elicitable properties implying that property is of size k.

Finally we discuss applications of our results to prediction markets. We show that prediction markets reveal exactly the same properties as in the scoring rule context. Properties of complexity $\mathcal{C}^k$ need at least k parallel markets to be elicited. We will then address the more specific cases of markets operated by continuous double auctions and those operated by automatic market makers. In both cases, we describe the form of the securities that can generate market estimates for a given elicitable property.

2. MODEL

2.1 Setting and notations

We consider a setting with an *experimenter* and a *forecaster*. The experimenter wishes to learn information—modeled as *distribution properties*—on a given *random experiment*, whose outcome is drawn according to some probability. The forecaster has beliefs over that probability, which is unknown to the experimenter. The experimenter can offer contracts to the forecaster, which specify payments to the forecaster as a function of the forecaster's prediction and the subsequent realized outcome.

We assume a finite set of outcomes $\Omega = \{\omega_1, \ldots, \omega_n\}$, and denote by $\mathbb{R}^\Omega$ the set of random variables over Ω, i.e., functions mapping elements of Ω to real values. We denote by $\Delta(\Omega) \subset \mathbb{R}^\Omega$ the set of density functions over Ω and we identify a probability P with its density function, with the abuse of notation $P(\{\omega_i\}) = P(\omega_i)$.

We will use $\mathbb{R}^\Omega$ as a linear space with the scalar product $\langle X, Y\rangle = \sum_{\omega\in\Omega} X(\omega)Y(\omega)$ and the distance $d(X,Y) = \|X - Y\| = \sqrt{\langle X-Y, X-Y\rangle}$. For a random variable X, and a distribution P, we write $E_P[X]$ the expected value of X under P. Our proofs will make extensive use of the fact that

$$E_P[X] = \langle P, X\rangle \ .$$

For any set $\mathcal{S}$ of $\mathbb{R}^\Omega$, there exists a unique smallest linear subspace $\mathcal{E}$ of $\mathbb{R}^\Omega$ containing $\mathcal{S}$, i.e., such that for any linear subspace $\mathcal{E}'$ containing $\mathcal{S}$, $\mathcal{E} \subseteq \mathcal{E}'$. We say that $\mathcal{E}$ is the *linear extension* of $\mathcal{S}$, and we define the dimension of $\mathcal{S}$ by $\dim \mathcal{S} = \dim \mathcal{E}$.

We recall that a set $\mathcal{S}$ is *open* in a metric space $\mathcal{E}$ when any of its points $X \in \mathcal{S}$ may be surrounded by a ball $\{Y \in \mathcal{E} / d(X,Y) < \epsilon\}$. The *interior* of $\mathcal{S}$ is the largest open set contained in $\mathcal{S}$ and is denoted $\mathrm{Int}(\mathcal{S})$.

2.2 Distribution properties

Definition 1. Given a set $\mathcal{D} \subseteq \Delta(\Omega)$, a distribution property $\Gamma : \mathcal{D} \mapsto \mathbb{R}$ is a function that assigns a real value to any probability in $\mathcal{D}$.

Common distribution properties include the probability of some event $A \subset \Omega$, $\Gamma_A(P) = P(A)$, the expectation of a given random variable X, $\Gamma_\mu(P) = E_P[X]$, or its standard deviation $\Gamma_\sigma(P) = \sqrt{E_P[(X - E_P[X])^2]}$, the kth centered moment $\Gamma_{\mu_k}(P) = E_P[(X - E_P[X])^k]$, the skewness $\Gamma_s(P) = \Gamma_{\mu_3}(P)/\Gamma_\sigma(P)^3$, the kurtosis $\Gamma_k(P) = \Gamma_{\mu_4}(P)/\Gamma_\sigma(P)^4 - 3$.

$\mathcal{D}$ is a set of probabilities which are considered "possible", it must include the true probability of the random experiment. The experimenter usually considers the set of all distributions $\Delta(\Omega)$ or those assigning positive probability to each outcome.

For the remaining of this paper we will make two assumptions. The first concerns the shape of the possible probabilities: as in [10, 14] we will assume convex domains $\mathcal{D}$. The second concerns the property itself: we will restrict our attention to *nice* properties Γ satisfying the following conditions:

(1) Γ is continuous, and

(2) Γ is not locally constant.[1]

In the sequel when we mention properties we always mean nice properties.

Since $\mathcal{D}$ is convex and Γ is continuous, the set of possible property values, $\Gamma(\mathcal{D}) = \{\Gamma(P)/P \in \mathcal{D}\}$, is an interval I, and its interior $\mathrm{Int}(\Gamma(\mathcal{D})) = (a,b)$, for some $a, b \in \mathbb{R}\cup\{\pm\infty\}$, is the same interval I without boundary points. We will only accept reports of values that exclude boundary points. This is not restrictive, it allows for unbounded score functions such as the logarithmic score. When the score is bounded, elicitation usually remains valid at boundary points by continuous extension of the score. [10, 14] make a similar assumption for probability scoring rules.

When considering a property Γ, we call a report r (resp. a probability P) *admissible* when $r \in \mathrm{Int}(\Gamma(\mathcal{D}))$ (resp. $\Gamma(P) \in \mathrm{Int}(\Gamma(\mathcal{D}))$).

2.3 Score functions

The family of contracts binding the experimenter and the forecaster is expressed by *score functions*, which are probability scoring rules adapted to the case of distribution properties. A score function associates a real-valued function of the outcomes to every valid report. It represents a contract that specifies a payment for all possible outcomes.

Definition 2. A *score function* for a property Γ with domain $\mathcal{D}$ is a function $s : \mathrm{Int}(\Gamma(\mathcal{D})) \mapsto \mathbb{R}^\Omega$.

When a forecaster reports a property value r, the experimenter and the forecaster commit to the contract $c = s(r)$, meaning that the experimenter pays the forecaster $c(\omega^*)$ when the true outcome ω^* of the random experiment becomes known.

A forecaster who believes in P and seeks to maximize her expected reward makes a report r that solves

$$\max_r E_P[s(r)] \ .$$

To induce truthful report of a property Γ, the score function used by the experimenter must ensure that the solution to the above maximization problem is $\Gamma(P)$ for every admissible probability P. When a score function is such that an

[1]Which means that there does not exist an open set of $\mathcal{D}$ on which Γ is constant.

optimal report is always consistent with her subjective probability, it is *strictly proper*, by analogy to probability scoring rules. In the broader game theory community, strictly proper score functions would more naturally be called *incentive compatible*.

Definition 3. A score function $s : \text{Int}(\Gamma(\mathcal{D})) \mapsto \mathbb{R}^\Omega$ is *strictly proper* for a property Γ with domain $\mathcal{D}$ if

$$E_P[s(r)] < E_P[s(\Gamma(P))]$$

for all admissible P and $r \neq \Gamma(P)$.

Note that looking at strictly proper scores does not prevent from considering risk-averse or risk-seeking forecasters, as long as they act as expected utility maximizers.[2] In fact, the question of whether we can or cannot elicit a certain information is independent of the forecaster's risk-attitude.

It may be desirable that a forecaster whose report is closer to the true value of the property than that of another forecaster has higher expected reward. This idea originated from Staël von Holstein [18] and was later investigated by Friedman [6]. We say that a score function whose expected value increases with accuracy is *accuracy-rewarding*. Accuracy-rewarding score functions are also strictly proper.

Definition 4. A score function s for a distribution property Γ is *accuracy-rewarding* if $E_P[s(r)] < E_P[s(r')]$ when either $r < r' \leq \Gamma(P)$ or $\Gamma(P) \leq r' < r$ for all admissible r, r', P.

Another interesting feature is that of first-order. It means that the first-order condition of the optimization problem determines the maximum of the expected score. This feature is important when relating score functions to securities in prediction markets, as equilibria are characterized by first-order conditions (Section 6). First-order score functions are also accuracy-rewarding and strictly proper.

Definition 5. A score function s for a property Γ is of *first-order* if, for all admissible P, the function

$$r \mapsto E_P[s(r)]$$

is continuously differentiable and has a maximum determined by the first-order condition:

$$\left.\frac{\partial E_P[s(r)]}{\partial r}\right|_{r=r^*} = 0 \qquad \Leftrightarrow \qquad r^* = \Gamma(P) \ .$$

3. DIRECT ELICITABILITY

The purpose of this section is (a) to determine the distribution properties that can be elicited truthfully directly, that is, for which there exists contracts that induce a truthful revelation the property by the forecaster, and (b) to identify those contracts.

Definition 6. A property Γ is *(directly) elicitable* when there exists a strictly proper score function for Γ.

[2]In the general case where the forecaster has a utility for money u (with inverse u^{-1}), then the experimenter can use any score function $\tilde{s}(r) = u^{-1} \circ s(r)$, with s a strictly proper score, to obtain truthful reports. Inversely, if $\tilde{s}$ induces truthful report, then $s(r) = u \circ \tilde{s}(r)$ is strictly proper. Consequently, we can assume w.l.o.g. that the forecaster has quasi-linear utility and concentrate on strictly proper score functions.

Those properties are called directly elicitable because the forecaster provides directly the value of the property: she gives exactly the information of interest, and no more. If the forecaster gives more information than the value of the property under consideration (for example, the full distribution), a property which may not be elicitable directly could still be inferred from that information. These cases of indirect elicitability will be discussed in Sections 5 and 6. Unless mentioned otherwise, in the rest of the paper elicitability will always mean direct elicitability.

3.1 A characterization of elicitable properties

Not every distribution property Γ is elicitable. A necessary condition is that Γ have its level sets convex. We recall that the level sets of Γ are the sets of probabilities $\Gamma^{-1}(r) = \{P/\Gamma(P) = r\}$ that map to the same property value.

LEMMA 1. *If a property Γ is elicitable, then $\Gamma^{-1}(r)$ is convex for all admissible r.*

PROOF. Let s be strictly proper for Γ, r be admissible property value, and $P, Q \in \Gamma^{-1}(r)$. Then for any admissible r' with $r' \neq r$, $E_P[s(r)] > E_P[s(r')]$ and $E_Q[s(r)] > E_Q[s(r')]$. So, if $0 < \lambda < 1$,

$$\begin{aligned} E_{\lambda P+(1-\lambda)Q}[s(r)] &= \lambda E_P[s(r)] + (1-\lambda)E_Q[s(r)] \\ &> \lambda E_P[s(r')] + (1-\lambda)E_Q[s(r')] \\ &= E_{\lambda P+(1-\lambda)Q}[s(r')] \ . \end{aligned}$$

Hence $\Gamma(\lambda P + (1-\lambda)Q) = r$. $\square$

We will see that convexity is also a sufficient condition of elicitability. We will use *convex maximal* sets. Convex maximal sets are easier to manipulate than convex sets. While convex sets contain all the barycenters of points with positive weights, convex maximal sets contain barycenters of points with positive or negative weights. Intuitively, a set is convex maximal in $\mathcal{S}$ when there does not exist a larger convex set in $\mathcal{S}$ of the same dimension.

Definition 7. A set $\mathcal{S}$ is convex maximal in $\mathcal{S}'$, subset of a linear space, if for all $P_1, \dots, P_k \in \mathcal{S}$, and all $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ with $\sum_i \lambda_i = 1$,

$$\sum_i \lambda_i P_i \in \mathcal{S}' \Rightarrow \sum_i \lambda_i P_i \in \mathcal{S} \ .$$

We omit to specify the reference set $\mathcal{S}'$ when there is no ambiguity (in most of our applications, we use $\mathcal{S}' = \mathcal{D}$). The following lemma will be useful.

LEMMA 2. *If a property Γ is such that for all admissible r, $\Gamma^{-1}(r)$ is convex, then it is also convex maximal (in $\mathcal{D}$).*

Convex maximal sets of a set $\mathcal{S}$ are geometrically highly regular: they are the intersection of $\mathcal{S}$ and linear spaces. Using Lemma 2, we show that when a property has convex maximal level sets, they form hyperplanes of the domain.

LEMMA 3. *If a property Γ is such that $\Gamma^{-1}(r)$ is convex for all admissible r, then $\dim(\Gamma^{-1}(r)) = \dim \mathcal{D} - 1$.*

Putting together Lemma 2 and Lemma 3, we can build an appropriate score function for any property with convex level sets, which permits a complete characterization of elicitable properties.

THEOREM 1. *A distribution property Γ is elicitable* if and only if $\Gamma^{-1}(r)$ *is convex for all admissible property values* r.

PROOF SKETCH. ($\Rightarrow$) We proved this side in Lemma 1.
($\Leftarrow$)
Let (a, b) the set of admissible property values, interior of $\Gamma(\mathcal{D})$. Let $\mathcal{E}$ be the linear extension of $\mathcal{D}$. By applying Lemma 3, the linear extension of $\Gamma^{-1}(r)$ is an hyperplane $\mathcal{H}_r$ of $\mathcal{E}$, for any $r \in (a, b)$.

Lemma 2 and the intermediate value theorem imply that $\mathcal{H}_r$ divides $\mathcal{D}$ in two parts, a "negative half space", the set $\mathcal{P}_r^-$ composed of the probabilities P with $\Gamma(P) < r$, and a "positive half space", the set $\mathcal{P}_r^+$ composed of the probabilities P with $\Gamma(P) > r$.

Consequently we can generate a unique vector $v(r)$ of $\mathcal{E}$ orthogonal to $\mathcal{H}_r$, normalized (such that $\|v(r)\| = 1$), and oriented towards the positive half-space, meaning that $\langle v(r), P\rangle > 0$ when $P \in \mathcal{P}_r^+$, and $\langle v(r), P\rangle < 0$ when $P \in \mathcal{P}_r^-$.

We can show that $v(r)$ is continuous in r by using the continuity of Γ. Intuitively, a Γ continuous implies that the hyperplanes $\mathcal{H}_r$ vary "smoothly" in r, and so do the direction of their normal vectors.

Continuity makes it possible to consider the function $s(r) = \int_{r_0}^{r} v(t)\mathrm{d}t$ on (a, b), for any r_0 on (a, b). Take any admissible r and P such that $r < \Gamma(P)$, then

$$\frac{\mathrm{d}E_P[s(r)]}{\mathrm{d}r} = \langle v(r), P\rangle > 0$$

as P is in the positive half-space $\mathcal{P}_r^+$, which implies that $E_P[s(\cdot)]$ is strictly increasing on the interval $(a, \Gamma(P))$. By a symmetric argument, $E_P[s(\cdot)]$ is strictly decreasing on the interval $(\Gamma(P), b)$. Hence s is strictly proper for Γ, which concludes the proof. $\square$

As the convexity of a property's level sets is equivalent to their maximal convexity, we also have:

COROLLARY 1. *A property* $\Gamma : \mathcal{D} \mapsto \mathbb{R}$ *is elicitable* if and only if $\Gamma^{-1}(r)$ *is convex maximal for all admissible property values* r.

This is useful to obtain an alternative view of elicitable properties as defined by linear constraints. Indeed, elicitable properties Γ are exactly those for which the equation $\Gamma(P) = r$ corresponds to a linear constraint on P. Applied to common distribution properties, we find that probabilities, expectations, and moments are elicitable, while variance, centered moments, kurtosis and skewness are not whenever $|\Omega| > 2$.

We say that a function $L : \mathcal{S} \mapsto \mathbb{R}$, for $\mathcal{S} \subseteq \mathbb{R}^\Omega$ is *linear in* $\mathcal{S}$ if for all $x, y \in \mathcal{S}$, and all $\alpha, \beta \in \mathbb{R}$ such that $\alpha x + \beta y \in \mathcal{S}$, $L(\alpha x + \beta y) = \alpha L(x) + \beta L(y)$.

THEOREM 2. *A property Γ with domain $\mathcal{D}$ is elicitable* if and only if*, for all admissible property value r, there exists a function $L^r : \mathcal{D} \mapsto \mathbb{R}$ linear in $\mathcal{D}$ such that, for all admissible r, P,*

$$L^r(P) = 0 \qquad \Leftrightarrow \qquad \Gamma(P) = r\ .$$

PROOF. ($\Leftarrow$) This side follows directly from the linearity of L^r. Indeed, let r be an admissible report, the characterization of Γ as linear constraints gives $\Gamma^{-1}(r) = \{P \in \mathcal{D}/L^r(P) = 0\}$. For any $P_1, \ldots, P_k \in \Gamma^{-1}(r)$, $\lambda_1, \ldots, \lambda_k$ with $\sum_i \lambda_i = 1$, $L^r(\sum_i \lambda_i P_i) = \sum_i \lambda_i L^r(P_i) = 0$ so $\sum_i \lambda_i P_i \in \Gamma^{-1}(r)$. Hence $\Gamma^{-1}(r)$ is convex maximal and by Corollary 1, Γ is elicitable.

($\Rightarrow$) Assume Γ is elicitable. Let r be an admissible report. Let $\mathcal{E}$ be the linear extension $\mathcal{D}$, and $\mathcal{H}$ the linear extension $\Gamma^{-1}(r)$. By Lemma 3, $\mathcal{H}$ is an hyperplane of $\mathcal{E}$. Since hyperplanes are kernels of non-null linear forms, there exist a linear form L^r on $\mathcal{E}$ such that $L^r(P) = 0$ if and only if $P \in \mathcal{H}$. As Γ is elicitable, $\Gamma^{-1}(r)$ must be convex maximal by Corollary 1, so $\Gamma^{-1}(r) = \mathcal{H} \cap \mathcal{D}$, and the restriction of L^r to $\mathcal{D}$ is null exactly on $\Gamma^{-1}(r)$, which concludes the proof. $\square$

3.2 Score function representations

Strictly proper score functions are not arbitrary but well structured. One of the first characterizations goes back to Shuford et al. [16], who considers eliciting a distribution on binary events. The idea is pushed further by Schervish in [15], who shows that there exists a one-to-one correspondence between proper scoring rules (for probability of a binary event) and non-negative measures on $[0, 1]$. Recently, Buja et al. [2] propose a taxonomy of scoring rules based on their Schervish measure, and use it in a statistical learning setting.

We obtain a generalization the results of Schervish [15]. We show that, for an elicitable property, there exists a one-to-one association between continuously differentiable score functions that are strictly proper and nonnegative absolutely continuous measures on property values that are not locally null. Indeed, each elicitable property is associated to a unique function, that we call *signature*. Strictly proper scores for a property are mixtures of the signature of that property, weighted positively almost everywhere. This means that an incentive compatible payment scheme is fully specified by two components: its signature, uniquely associated to the property that it elicits, and its *weight function*, which assigns non-negative weights to property values.

Two incentive-compatible payment schemes differ only by their weights. These functions indicate the amplitude of variation of the payment with respect to the reports. As we can choose them freely, more amplitude should be given in regions of the property values that are of higher interest [14].

Once we know the signature of a property, it becomes possible to generate a wealth of score functions – in fact, all of them – very easily, by varying the weights, whose sole constraint is to be positive almost everywhere. Besides, the signature of a property Γ is easily determined by differentiating any strictly proper score for Γ. Consequently we can generate all the strictly possible incentive compatible payment functions from one sample.

THEOREM 3. *If a property Γ with domain $\mathcal{D}$ is elicitable, denoting by (a, b) the interval of admissible reports for Γ, and $\mathcal{E}$ the linear extension of $\mathcal{D}$, then there exists a unique function $v : (a, b) \mapsto \mathbb{R}^\Omega$, taking values in $\mathcal{E}$, with $\|v(\cdot)\| = 1$ and such that a continuously differentiable score function s is strictly proper (resp. of first-order) for Γ if and only if*

$$s(r) = s_0 + \int_{r_0}^{r} \lambda(t)v(t)\mathrm{d}t \tag{1}$$

for $s_0 \in \mathbb{R}^\Omega$ and a continuous weight function $\lambda : (a, b) \mapsto [0, +\infty)$ that is not locally null (resp. never null).

In this theorem, we call the function v the *signature* of the property, and λ the *weight function*. For common domains, e.g., $\mathcal{D} = \Delta(\Omega)$, $\mathcal{E}$ is $\mathbb{R}^\Omega$ the set of random variables over Ω.

Proof (Theorem 3). We start by constructing the signature of Γ. Let $r \in (a,b)$, $\mathcal{E}$ be the linear extension of $\mathcal{D}$. By Lemma 3 the linear extension $\mathcal{H}_r$ of $\Gamma^{-1}(r)$ is an hyperplane of $\mathcal{E}$. In the proof of Theorem 1, we identified a vector $v(r) \in \mathcal{E}$ that is (1) normal to $\mathcal{H}_r$, (2) continuous in r, and (3) such that $\langle v(r), P\rangle > 0$ when $r < \Gamma(P)$ and such that $\langle v(r), P\rangle < 0$ when $\Gamma(P) < r$. We refer to v as Γ's signature.

We now proceed to the main proof.

($\Rightarrow$) Assume s elicits Γ and s is continuously differentiable. Then by the first-order condition

$$\left.\frac{\mathrm{d}E_P[s(r)]}{\mathrm{d}r}\right|_{r=\Gamma(P)} = 0$$

for P admissible. Therefore

$$\left\langle \frac{\mathrm{d}s(r)}{\mathrm{d}r}, P \right\rangle = 0$$

for any $r \in (a,b)$ and all $P \in \Gamma^{-1}(r)$.

Since vectors of $\mathcal{H}_r$ are linear combinations of $\Gamma^{-1}(r)$, if a vector n is such that $\langle n, P\rangle = 0$ for all $P \in \Gamma^{-1}(r)$, then $\langle n, X\rangle = 0$ for all $X \in \mathcal{H}_r$. So $\mathrm{d}s(r)/\mathrm{d}r$ is orthogonal to all vectors of $\mathcal{H}_r$, and, as $\mathcal{H}_r$ is an hyperplane of $\mathcal{E}$, is colinear to any normal vector of $\mathcal{H}_r$ in $\mathcal{E}$. As $v(r)$ is normal to $\mathcal{H}_r$, for all $r \in (a,b)$,

$$\frac{\mathrm{d}s(r)}{\mathrm{d}r} = \lambda(r)v(r) \tag{2}$$

for some scalar $\lambda(r)$. Note that $\mathrm{d}s(r)/\mathrm{d}r$ is continuous, $\|v(\cdot)\| = 1$, and $v(\cdot)$ is continuous, thus $\lambda(\cdot)$ is continuous. Therefore, by integration, we obtain Equation (1). It remains to prove that λ is nonnegative and not locally null, and that the function v is unique.

If $\lambda(r^*) < 0$ for some r^*, by continuity λ is negative in a neighborhood of r^*, on some interval $(r^*-\epsilon, r^*+\epsilon)$ for some $\epsilon > 0$. Then, considering any P with $\Gamma(P) = r^*$, we have, for all $r \in (r^*, r^*+\epsilon)$,

$$\frac{\mathrm{d}E_P[s(r)]}{\mathrm{d}r} = E_P[\lambda(r)v(r)] = \lambda(r)\langle v(r), P\rangle \ .$$

But within that range, $\lambda(r) < 0$ and $\langle v(r), P\rangle < 0$ since $r > \Gamma(P)$. So $E_P[s(r)] > E_P[s(\Gamma(P))]$ and s is not strictly proper. Hence $\lambda \geq 0$.

If λ is locally null, then by Equation (2), s is locally constant, and so $E_P[s(\cdot)]$ is locally constant for all admissible P, implying that s is not strictly proper. Therefore λ is not locally null.

We now turn to uniqueness. Suppose there exists $w(r) \in \mathcal{E}$, continuous in r, with $\|w(r)\| = 1$, and verifying Equation (1). Then $\mathrm{d}s(r)/\mathrm{d}r$ is positively colinear with $w(r)$, and as it is orthogonal to $\mathcal{H}_r$ hyperplane of $\mathcal{E}$, $w(r) = \mu(r)v(r)$ for some scalar $\mu(r)$. But $\|v(r)\| = 1$, so, for any r, either $\mu(r) = 1$, or $\mu(r) = -1$. Since μ must be continuous, $\mu = 1$ or $\mu = -1$. Yet the only possible μ that induces a nonnegative weight function is 1, so $w(r) = v(r)$.

Of course, if s is of first-order, λ is never null, otherwise the derivative of $E_P[s(r)]$ would always be null at r for which $\lambda(r) = 0$, for all admissible P, thus violating conditions of first-order.

($\Leftarrow$) Let s be written as in Equation (1), with v being Γ's signature. Then s is differentiable and $\mathrm{d}s(r)/\mathrm{d}r = \lambda(r)v(r)$. We verify that s is strictly proper by expressing the derivative of the expected score in terms of $v(r)$. For any admissible P,

$$\begin{aligned}\frac{\mathrm{d}E_P[s(r)]}{\mathrm{d}r} &= E_P\left[\frac{\mathrm{d}s(r)}{\mathrm{d}r}\right] \\ &= \left\langle \frac{\mathrm{d}s(r)}{\mathrm{d}r}, P \right\rangle \\ &= \lambda(r)\langle v(r), P\rangle \ .\end{aligned}$$

v is such that, if $r < \Gamma(P)$, $\langle v(r), P\rangle > 0$, and since $\lambda \geq 0$, $\mathrm{d}E_P[s(r)]/\mathrm{d}r \geq 0$, so the expected score $r \mapsto E_P[s(r)]$ is non-decreasing on $(a, \Gamma(P))$. By a symmetric argument, $r \mapsto E_P[s(r)]$ is non-increasing on $(\Gamma(P), b)$. Moreover, if $E_P[s(r)]$ was constant on some open interval of values r, $\lambda(r)$ would be null on that interval, impossible by assumption. Hence for all admissible P, $r \mapsto E_P[s(r)]$ is strictly increasing for $r < \Gamma(P)$ and strictly decreasing for $r > \Gamma(P)$, thus s is strictly proper for Γ.

If, in addition, λ is never null, then according to the last equation, $\mathrm{d}E_P[s(r)]/\mathrm{d}r$ is null exactly when $\langle v(r), P\rangle$ is null, and so is null only when $r = \Gamma(P)$, which means that s is of first-order. $\square$

The Schervish theorem [15] is a special case when using our result applied to a two-outcome space and Γ the probability of one outcome. Our result may be interpreted as follows: the direction and orientation of the gradients $\mathrm{d}s(r)/\mathrm{d}r$ is determined by the property that s elicits, while the gradient norms $\lambda(r) = \|\mathrm{d}s(r)/\mathrm{d}r\|$ are freely decided independently of the property.

The theorem shows how to build a wide variety of score functions that are strictly proper or of first-order. Further, it implies that any score function that is continuously differentiable and strictly proper *is also* accuracy-rewarding.

Corollary 2. *A continuously differentiable score function is strictly proper* if and only if *it is accuracy-rewarding.*

For example, if Ω contains only two outcomes 0 and 1, then the signature of $\Gamma(P) = P(0)$ is

$$v(t) = \frac{1}{\sqrt{t^2 + (1-t)^2}} \begin{bmatrix} 1-t \\ -t \end{bmatrix} \ .$$

The weight function of the logarithmic scoring rule,

$$s(r)(\omega) = b[\omega \log r + (1-\omega)\log(1-r)] \ ,$$

is

$$\lambda(t) = b\sqrt{\frac{1}{t^2} + \frac{1}{(1-t)^2}} \ ,$$

and the weight function of the quadratic score

$$s(r)(\omega) = -b(\omega - r)^2$$

corresponds to

$$\lambda(t) = b\sqrt{t^2 + (1-t)^2} \ .$$

4. VECTORS OF PROPERTIES

We now consider vectors of properties, for two major reasons. In the first place, the experimenter may be interested in more than a single feature of the distribution, or in a feature that is only expressible as a multi-dimensional value.

Secondly, property vectors provide a solution to the problem of obtaining truthful reports of a property that is not elicitable. In fact the properties that compose the vectors need not be elicitable for the vector to be elicitable. For example, the pair *(expectation, variance)* is elicitable while variance is not. Note that eliciting a vector of properties is equivalent to eliciting a set of those properties, but vectors offer a convenient way to express score functions.

We will consider vectors composed of properties that share the same domain, and that are *independent*, that is, for a vector $\Gamma = (\Gamma_1, \dots, \Gamma_k)$, the knowledge of the value of one property does not give the value of another. More formally, for $i \neq j$, and all admissible P, there exists an admissible P' such that $\Gamma_i(P) = \Gamma_i(P')$ and $\Gamma_j(P) \neq \Gamma_j(P')$.

Getting reports for vectors of properties is accomplished in a similar fashion as for single properties. The forecaster provides a report for each of the properties composing the vector. Based on these reports, the experimenter chooses a contract that specifies payments to be made for every possible outcome of the random experiment. The family of contracts is formulated as a score function, to be interpreted in same way as for single properties and whose definition is extended to the multi-dimensional case.

Definition 8. A *score function for the vector of properties* $(\Gamma_1, \dots, \Gamma_k)$ is a function $s : (a_1, b_1) \times \cdots \times (a_k, b_k) \mapsto \mathbb{R}^\Omega$, with (a_i, b_i) the interior $\Gamma_i(\mathcal{D})$ with $\mathcal{D}$ domain of the vector.

Properties that compose the vector are often related, and the shape taken by the range of possible vectors of property values may not be a (multi-dimensional) rectangle. However from a practical standpoint, it is more convenient to specify a window of acceptable reports for each individual property than a complex multi-dimensional object.

As previously, strictly proper scores maximize expected score of truthful reports, and a property is *(directly) elicitable* when it admits a strictly proper score function.

Definition 9. A score function s is *strictly proper for a vector of properties* $\Gamma = (\Gamma_1, \dots, \Gamma_k)$ if

$$E_P[s(r_1, \dots, r_k)] < E_P[s(\Gamma_1(P), \dots, \Gamma_k(P))]$$

for all admissible P, $(r_1, \dots, r_k) \neq (\Gamma_1(P), \dots, \Gamma_k(P))$.

Obviously a set composed of elicitable properties is elicitable, but an elicitable set may contain properties that are unelicitable. In fact, some elicitable sets may be entirely composed of unelicitable properties.

Convexity of the level sets is a necessary condition by the same argument as in Lemma 1.

THEOREM 4. *If a vector of distribution properties* $\Gamma = (\Gamma_1, \dots, \Gamma_k)$ *is elicitable, then* $\Gamma^{-1}(r_1, \dots, r_k)$ *is convex for admissible* $r_1, \dots, r_k$.

Unfortunately Theorem 1 cannot be adapted as such, and the question of whether convexity is a sufficient condition remains open. However proposing rewards that increase with the forecaster's accuracy is also desirable in the present situation, and instead of identifying bundles of properties that admit strictly proper scores we shall identify those that admit accuracy-rewarding scores, and give a simple characterization of those scores.

Regarding a vector of properties $(\Gamma_1, \dots, \Gamma_k)$, a score function s is accuracy-rewarding if for all admissible $(r_1, \dots, r_k)$ and $(r'_1, \dots, r'_k)$, with $(r_1, \dots, r_k) \neq (r'_1, \dots, r'_k)$, when for all i, either $r_i \leq r'_i \leq \Gamma(P)$ or $\Gamma(P) \leq r'_i \leq r_i$, then $E_P[s(r_1, \dots, r_k)] < E_P[s(r'_1, \dots, r'_k)]$.

Unlike the case of a single distribution property, there exists elicitable vectors that do not admit an accuracy-rewarding score function. For example, the theorem that follows imply that there does not exist any accuracy-rewarding score for the pair *(expectation, variance)*. When a vector is composed of elicitable properties, the experimenter can always use a sum of accuracy-rewarding score functions for each of the properties as an accuracy-rewarding score function for the whole vector. The following result shows that this is always the case. In particular, *no* score function is accuracy-rewarding for a vector of properties that contains at least one unelicitable property.

THEOREM 5. *A score function s for a vector of properties $(\Gamma_1, \dots, \Gamma_k)$ that is twice continuously differentiable is accuracy-rewarding* if and only if *there exists strictly proper score functions s_i for Γ_i twice continuously differentiable and such that*

$$s(r_1, \dots, r_k) = \sum_i s_i(r_i) \ .$$

PROOF. Let $\mathcal{D}$ be the domain of the vector of properties and (a_i, b_i) be the interior of $\Gamma_i(\mathcal{D})$.

($\Leftarrow$) Take any admissible $(r_1, \dots, r_k)$, $(r'_1, \dots, r'_k)$ and an admissible probability P, with $(r_1, \dots, r_k) \neq (r'_1, \dots, r'_k)$, and assume that, for $1 \leq i \leq k$, either $r_i \leq r'_i \leq \Gamma_i(P)$ or $\Gamma_i(P) \leq r'_i \leq r_i$. If the s_i are strictly proper they are also accuracy-rewarding by Corollary 2. Hence, for all i, $E_P[s(r_i)] \leq E_P[s(r'_i)]$ (with a strict inequality if $r_i \neq r'_i$). Therefore,

$$\begin{aligned} E_P[s(r_1, \dots, r_k)] &= \sum_i E_P[s_i(r_i)] \\ &< \sum_i E_P[s_i(r'_i)] \\ &= E_P[s(r'_1, \dots, r'_k)] \ . \end{aligned}$$

Thus s is accuracy-rewarding.

($\Rightarrow$) We will use the following lemma:

LEMMA 4. *If $f : (a_1, b_1) \times \cdots \times (a_k, b_k) \mapsto \mathbb{R}$ function twice continuously differentiable verifies $\partial^2 f(x_1, \dots, x_k)/\partial x_i \partial x_j = 0$ for $i \neq j$, then there exists $f_i : (a_i, b_i) \mapsto \mathbb{R}$ such that*

$$f(x_1, \dots, x_k) = \sum_i f_i(x_i) \ . \tag{3}$$

Consider any admissible $(r_1, \dots, r_k)$, and any admissible probability P. Let $1 \leq \ell \leq k$, and construct $(r'_1, \dots, r'_k)$ with $r'_i = r_i$ for all $i \neq \ell$, and $r'_\ell = \Gamma_\ell(P)$. Then as s is accuracy-rewarding, if $r_\ell \neq \Gamma_\ell(P)$, $E_P[s(r_1, \dots, r_k)] < E_P[(r'_1, \dots, r'_k)]$. This means that $r_\ell \mapsto s(r_1, \dots, r_k)$ is strictly proper for all ℓ, and also that Γ_ℓ is elicitable.

Take arbitrary i, j with $i < j$. For convenience, we use the notation $s(r_i, r_j) = s(r_1, \dots, r_k)$.

The first order condition, applied to the strictly proper score functions $s(\cdot, r_j)$ and $s(r_i, \cdot)$, gives

$$\left. \frac{\partial E_P[s(r_i, r_j)]}{\partial r_i} \right|_{r_i = \Gamma_i(P)} = 0 \ ,$$

$$\left. \frac{\partial E_P[s(r_i, r_j)]}{\partial r_j} \right|_{r_j = \Gamma_j(P)} = 0 \ .$$

As Ω is finite we can swap the differential operator and the expectation operator. By a second differentiation, and noting that $\partial^2 s/\partial r_i \partial r_j = \partial^2 s/\partial r_j \partial r_i$,

$$E_P\left[\frac{\partial^2 s(\Gamma_i(P), r_j)}{\partial r_i \partial r_j}\right] = 0 \;, \tag{4}$$

$$E_P\left[\frac{\partial^2 s(r_i, \Gamma_j(P))}{\partial r_i \partial r_j}\right] = 0 \;. \tag{5}$$

Consider any $r_i^* \in (a_i, b_i)$, $r_j^* \in (a_j, b_j)$. Let's assume by contradiction that

$$v = \frac{\partial^2 s(r_i^*, r_j^*)}{\partial r_i \partial r_j} \neq 0 \;.$$

Then for all $P_i \in \Gamma_i^{-1}(r_i^*)$, $P_j \in \Gamma_j^{-1}(r_j^*)$, by Equation (4) and (5),

$$E_{P_i}[v] = E_{P_j}[v] = 0 \;, \tag{6}$$

Besides, since Γ_i and Γ_j are elicitable, by Lemma 3, the linear extensions $\mathcal{H}_i$ and $\mathcal{H}_j$ of respectively $\Gamma_i^{-1}(r_i^*)$ and $\Gamma_j^{-1}(r_j^*)$ are hyperplanes of the linear extension $\mathcal{E}$ of $\mathcal{D}$. Note that we can write $E_Q[v] = \langle v, Q\rangle$ for $Q = P_i, P_j$ and any vector of $\mathcal{H}_i$ is a linear combination of vectors of $\Gamma_i^{-1}(r_i^*)$ (and similarly for j). By linearity of $\langle v, \cdot\rangle$, Equation (6) implies that $\langle v, X\rangle = 0$ for all $X \in \mathcal{H}_i$ and all $X \in \mathcal{H}_j$. As $\mathcal{H}_i$ and $\mathcal{H}_j$ are both hyperplanes of $\mathcal{E}$, they must be equal.

In addition, by Corollary 1, $\Gamma_i^{-1}(r_i^*) = \mathcal{H}_i \cap \mathcal{D}$ and $\Gamma_j^{-1}(r_j^*) = \mathcal{H}_j \cap \mathcal{D}$, so $\Gamma_i^{-1}(r_i^*) = \Gamma_j^{-1}(r_j^*)$. Hence, if Q is such that $\Gamma_i(Q) = r_i^*$, then $\Gamma_j(Q) = r_j^*$, and inversely. However this is impossible by the independence assumption.

Therefore,

$$\frac{\partial^2 s(r_i^*, r_j^*)}{\partial r_i \partial r_j} = 0$$

for all r_i^*, r_j^*. By applying Lemma 4, s is additively separable. The components are all strictly proper since we showed that $r_\ell \mapsto s(r_1, \ldots, r_k)$ is strictly proper. □

5. THE COMPLEXITY OF DISTRIBUTION PROPERTIES

Our analysis so far has focused on properties that can be directly elicited. In this section, we look at properties that are not directly elicitable.

To elicit properties that are not directly elicitable, the experimenter must ask for more information than necessary, and infer the property value from that information. For example, she may call for reports of the full distribution, in that sense every property is *implicitly* elicitable. Fortunately, in most situations, less information is needed. In general, the experimenter may infer a property (for example, the variance of X) either from an elicitable set (for example, $\{E[X], \mathrm{Var}(X)\}$), or from a set of elicitable properties (from example, $\{E[X], E[X^2]\}$). In both cases, it would be interesting to know what is the minimal size of the set from which a property may be inferred. Obviously the minimal size in the first case is no bigger than in the second case, but the reverse is not obvious. Here we focus on the second case: indirect elicitation from sets of directly elicitable properties. We do so for two reasons: as we argued in the previous section, accuracy-rewarding scores exist only for sets of elicitable properties, and as we will see in the next section, the minimal size of such set has a natural interpretation for prediction markets.

If a property can be inferred from k elicitable properties, it is k-*elicitable.* The notion of k-elicitability completes that of direct elicitability.

Definition 10. A property Γ is said to be k-elicitable if there exists elicitable properties $\Gamma_1, \ldots, \Gamma_k$ sharing the same domain such that $\Gamma = f(\Gamma_1, \ldots, \Gamma_k)$ for some function f.

While directly elicitable properties are 1-elicitable, the reverse statement is not true. For example, $|E[X]|$ is 1-elicitable but is not directly elicitable.

Naturally, an experimenter typically wishes to use as few properties as possible. The more reports needed, the harder it is to evaluate a distribution property. In that sense, the minimal number of elicitable properties needed to infer a distribution property measures the difficulty to evaluate that property. This leads to the introduction the *complexity* of distribution properties.

Definition 11. A property is said to be of complexity $\mathcal{C}^k$, $k \geq 1$ if it is k-elicitable, but is not $(k-1)$-elicitable, when $k > 1$.

For example, the probability of a binary event and the expectation of a random variables are properties of complexity $\mathcal{C}^1$, the variance is of complexity $\mathcal{C}^2$. It can be shown that the skewness is of complexity $\mathcal{C}^3$. The kurtosis is of complexity $\mathcal{C}^3$ or $\mathcal{C}^4$ (we conjecture $\mathcal{C}^4$).

The notion of complexity extends our view beyond the dichotomy of elicitability, and creates a rich taxonomy of properties. Intuitively, properties of complexity $\mathcal{C}^a$ are more difficult to elicit than those of complexity $\mathcal{C}^b$ when $a > b$, and two properties of the same complexity are equally hard to elicit. In particular, obtaining a truthful reports of a property of complexity $\mathcal{C}^k$ is exactly as difficult as obtaining truthful report of a full distribution with $k+1$ possible outcomes. For example, eliciting variance is as hard as eliciting a distribution in a 3-outcome worlds.

This raises a natural question: how hard can it be to elicit a distribution property? Surprisingly, there are properties which are as hard to obtain as the full distribution, as stated in the result below. For example, the property $\max_{\omega\in\Omega} P(\{\omega\})$ is of complexity $\mathcal{C}^{|\Omega|-1}$: it is as difficult to elicit as the $|\Omega|$ marginal probabilities that specify the distribution. Further, distributions properties can be more or less difficult to evaluate, and form a continuum between the properties that are directly elicitable and those which require the full distribution.

THEOREM 6. *If $|\Omega| = n$, any property belongs to one of the classes $\mathcal{C}^1, \ldots, \mathcal{C}^{n-1}$. In addition, none of the complexity classes is empty.*

6. APPLICATION TO PREDICTION MARKETS

Prediction markets are markets designed for the purpose of estimating probabilistic information. Empirical evidence suggests that they are powerful prediction tools [5, 20], and they are now used regularly in industry. In prediction markets, participants trade securities whose payoffs are contingent upon the realization of uncertain events. Relying on the efficient market hypothesis, these markets are designed in such a way that the market state at equilibrium reveals certain information of interest.

So far our analysis involved a single agent. In contrast, a prediction market is a multi-agent setting. However, under standard assumption, market equilibria lead to a consensus of posterior beliefs, so that the crowd behaves as single agent. While we cannot query this hypothetical agent, our analysis may be applied to show that (a) with one type of securities, markets reveal exactly the same information as single-agent reward schemes, and (b) in the two main market mechanisms, there exists a one-to-one correspondence between securities and score functions.

As in the single-agent case, we assume that a random experiment will produce an outcome of a finite set Ω according to some probability a priori unknown, but on which market participants form beliefs. We assume traders are risk-neutral and seek to maximize their expected profit according to their belief.

In most prediction markets, participants trade on securities indexed by a real-valued parameter r [20].[3] Each security is described by its price p_r and its payoff $\mathbf{q}_r$. The payoff gives, for each outcome ω, the monetary value $\mathbf{q}_r(\omega)$ of one unit of security r. We will assume that prices and payoffs vary continuously with the index. In the simplest situation, securities are indexed by price and the payoff is 1 if some event A occurs and 0 otherwise: for a security r, $p_r = r$ and $\mathbf{q}(\omega) = 1$ if $\omega \in A$, $\mathbf{q}(\omega) = 0$ otherwise. Following the classification of Wolfers and Zitzewitz [20], this is a *winner-takes-all* security, in use in a majority of markets, such as TradeSports[4] and the Iowa Electronic Market [5].

Wolfers and Zitzewitz describe two other types of securities commonly used in practical contexts: *index* securities, with a variable price $p_r = r$ and a payoff that varies linearly with a numeric outcome $\mathbf{q}_r(\omega) = \omega$ (e.g., the number of points a sport team wins by), and *spread* securities, with a fixed price $p_r = 1$ and a payoff that depends on whether a numeric outcome is below or above the index: $\mathbf{q}_r(\omega) = 2$ if $\omega \geq r$ and $\mathbf{q}_r(\omega) = 0$ otherwise.

In many market environments, participants trade securities among themselves. In that case, there is *equilibrium* when participants are no longer willing to trade with each other. Formally, for any two traders with subjective probability P_1 and P_2, there does not exist a security r with

$$E_{P_1}[\mathbf{q}_r] - p_r > 0 \; , \tag{7}$$

$$p_r - E_{P_2}[\mathbf{q}_r] > 0 \; . \tag{8}$$

To reveal information, the market index, or the index of the security of the most recent exchange, must have a particular meaning. For example, the price of a winner-take-all security gives the consensus probability of an event, the price of an index security provides a market consensus on the expected value of a numeric outcome, and spread securities give the median of the distribution of a numeric outcome.

More generally, we say that a market *reveals* a property Γ when, at equilibrium, the market index equals $\Gamma(P)$ for P the subjective probability of any trader. In particular, there is a unique equilibrium corresponding to a given $\Gamma(P)$, and so, as prices and payoffs are continuous, by (7) and (8), the equilibrium market index is characterized by $E_P[\mathbf{q}_r] - p_r = 0$ for all P subjective probability of any trader.

[3] Although we consider only one family of securities, markets may contain multiple families, i.e., different types of securities, typically one for each outcome.

[4] `www.tradesports.com`

Since $E_P[\mathbf{q}_r - p_r] = 0$ if and only if $r = \Gamma(P)$, by Theorem 2, Γ is elicitable. Inversely, if Γ is elicitable, then by choosing any p_r, $\mathbf{q}_r$ such that $\mathbf{q}_r - p_r$ equals the signature of Γ evaluated at r, the market reveals Γ. In other words, subject to the continuity assumptions above, and considering a single family of securities,

THEOREM 7. *Prediction markets reveal exactly the same distribution properties as one-dimensional score functions.*

This sheds new light on the interpretation of the complexity of a property: Γ is of complexity $\mathcal{C}^k$ when at least k prediction markets executed in parallel (or k types of securities) are needed to compute a market estimate of Γ.

Thus only properties that correspond to linear constraints may be revealed by the market. It is worth mentioning the manifest connection between our findings and those of Feigenbaum et al. [4]. Feigenbaum et al. model prediction markets as Shapley-Shubik games in which traders receive private binary signals. They show that, under Aumann's information partition model with common prior, prediction markets may only reveal a statistic of signals when the statistic is a linear function of the input signals.

In principle, these markets need not be organized: participants can negotiate securities with each other directly. However, when many traders are present, it is more efficient to use a centralized mechanism. Most existing prediction markets are of two sorts: markets using Continuous Double Auctions (CDA) – used for example in TradeSports, Betfair, the Iowa Electronic Markets, or the Google internal prediction market – and markets operated by automatic market makers, such as Hanson's logarithmic market scoring rule market maker [9] – used in YooNew[5], Inkling Markets[6], and in Microsoft's internal prediction markets.

6.1 Markets with continuous double auction

A Continuous Double Auction (CDA) attempts to match, at any time, orders to buy a security with orders to sell. We describe a general model of CDA, that can be used with any family of securities, including winner-take-all, index, and spread. Common CDAs, like those used in TradeSports, Betfair, Iowa Eletronic Market, and many financial places, are a special case.

A CDA allows traders to buy and sell securities as follows. At any time, a trader may submit an order to buy (resp. sell) a *given quantity of securities* whose index is below (resp. above) some *given index value*. When submitting a buying order (resp. selling order), the order is placed in a *bid queue* (resp. *ask queue*). At any time, an attempt is made to match orders of the bid queue with orders of the ask queue. If r_{bid} is the bid index, i.e., the highest index in the bid queue and r_{ask} is the ask index, i.e., the lowest index in the ask queue, and if $r_{bid} > r_{ask}$, then a trade occurs between the corresponding market participants on the security indexed by r^* with, for example, $r^* = (r_{bid} + r_{ask})/2$.

The functioning of the CDA relies on the assumption that a trader willing to buy (resp. sell) a security r_{bid} (resp. r_{ask}) is also willing to buy a security $r < r_{bid}$ (resp. sell a security $r > r_{ask}$). Therefore, the family of securities must conform to the following condition for any probability P, for

[5] `www.yoonew.com`

[6] `www.inklingmarkets.com`

all $r' < r$,

$$E_P[\mathbf{q}_r] - p_r \geq 0 \Rightarrow E_P[\mathbf{q}_{r'}] - p_{r'} \geq 0 \ , \quad (9)$$

$$E_P[\mathbf{q}_{r'}] - p_{r'} \leq 0 \Rightarrow E_P[\mathbf{q}_r] - p_r \leq 0 \ . \quad (10)$$

The equilibrium condition imposes that the CDA reveals Γ only when $\mathbf{q}_r - p_r$ is colinear to the signature of the property defined in Theorem 2, and Equation (9) and (10) are verified only when the colinearity is positive. Therefore, from Theorem 3,

THEOREM 8. *A CDA reveals a property Γ if and only if each market security r has a price p_r and a contingent payoff $\mathbf{q}_r$ such that $\mathbf{q}_r - p_r = \mathrm{d}s(r)/\mathrm{d}r$ for a first-order score function s for Γ.*

Note that this may used in both ways: to design securities for prediction markets from score functions, and to find strictly proper score functions from prediction market securities.

6.2 Markets with automated market maker

In prediction markets with automated market maker (MwAMM), participants do not trade directly with each other, but with a market-maker, who decides of the offer. Currently, many prediction markets use Hanson's logarithmic market scoring rule market maker [9]. Traders may sell or buy at any time the security present on the market, which is fixed by the market maker. This type of markets would correspond to a financial market where all traders are price takers. When traders buy or sell securities, the market-maker raises or lowers the index of the security being traded, effectively reacting to the demand by changing the offer. Contrary to a CDA, the patron of a MwAMM is subject to potential losses that occur almost certainly when traders are better informed than the patron, a natural assumption for prediction markets. Losses depend on how the offer vary with the demand, as the offer is decided by the market-maker, it is generally possible to keep a control over the worst-case loss.

In addition to a family of securities, the market also specifies a function g that determines, at any time, the security available for trade. The market-maker keeps track of the net quantity of securities she has sold, Q, and sets $r = g(Q)$ as the index of the security available. To be consistent, the market maker should keep moving the index in the same direction when buying securities, so we can impose w.l.o.g. that g be increasing.

Traders interact with the market by specifying a quantity of securities to buy ΔQ (to sell if $\Delta Q < 0$). As securities present on the market change continuously with trading activity, prices and payoffs of a market order must be calculated by summing over infinitesimal trades. When a trader buys a infinitely small quantity of securities $\mathrm{d}Q$, he pays the market maker $p_r \mathrm{d}Q$ and gets a payoff $\mathbf{q}_r(\omega)\mathrm{d}Q$ for an outcome ω, while the market maker raises the market index by $g'(Q)\mathrm{d}Q$. By integration, when a quantity ΔQ of securities is purchased at a time when market maker has sold a net quantity Q_1 of securities, the price is

$$\int_{Q_1}^{Q_1+\Delta Q} p_{g(Q)} \mathrm{d}Q = \int_{g(Q_1)}^{g(Q_1+\Delta Q))-g(Q_1)} \frac{p_r}{g'(g^{-1}(r))} \mathrm{d}r \ ,$$

and the payoff for outcome ω is

$$\int_{Q_1}^{Q_1+\Delta Q} \mathbf{q}_{g(Q)}(\omega) \mathrm{d}Q = \int_{g(Q_1)}^{g(Q_1+\Delta Q)} \frac{\mathbf{q}_r(\omega)}{g'(g^{-1}(r))} \mathrm{d}r \ .$$

Noting $n(t) = \mathbf{q}_t - p_t$, the net profit at outcome ω is $(s(g(Q_1 + \Delta Q)) - s(g(Q_1)))(\omega)$ with the function $s(r) = \int_{r_0}^{r} \lambda(t)v(t)\mathrm{d}t$ and $v(t) = n(t)/\|n(t)\|$ and $\lambda(t) = \|n(t)\|/g'(g^{-1}(t))$. Note that s is a score function.

We say that the market is at equilibrium when no trader can make a profit from an infinitesimal trade, that is, when $E_P[\mathbf{q}_r \mathrm{d}Q] - p_r \mathrm{d}Q \leq 0$ and $p_r \mathrm{d}Q - E_P[\mathbf{q}_r \mathrm{d}Q] \leq 0$ for P subjective probability of any trader. So the market reveals Γ when $\Gamma(P) = r$ if and only if $E_P[\mathbf{q}_r] - p_r = 0$, which is equivalent to saying that $\Gamma(P) = r$ if and only if $E_P[v(t)] = 0$. Thus the market reveals directly Γ if and only the function s defined above is a first-order score function for Γ.

THEOREM 9. *A MwAMM reveals a property Γ if and only if there exists a first-order score function s such that $\mathbf{q}_r - p_r = \mathrm{d}s(r)/\mathrm{d}r$.*

This means that the vector v above corresponds to the signature of the property being elicited.

In addition, the variation of the net profit is $s(r_2) - s(r_1)$ when trading activities move the index from r_1 to r_2. It follows that trading in the market corresponds *exactly* to a sequence of single-agent elicitation by the score function s, in which, at any time, a trader can move the current market estimate r_i of round i of the market property value to r_{i+1} by accepting to pay the most recent person that changed it an amount $s(r_i)$ while receiving $s(r_{i+1})$. This is not surprising: Hanson already showed the equivalence between markets with binary securities and probability scoring rules [9]. What we get is a generalization to all types of securities.

In fact, we have just shown that *any* market with automated market maker and only one family of securities is structurally equivalent to allowing agents to interact with some score function strictly proper for some property: the payoff from a traded quantity ΔQ after Q securities have been sold in the former equals a difference of scores $s \circ g(Q + \Delta Q) - s \circ g(Q)$ in the latter, where g is defined as above and determines the offer, and s is some strictly proper score function. Appropriate choices of s and g help keep control over potential losses.

For example, applied to the case of probability elicitation of a binary event, we deduct from Theorem 3 that securities must have varying prices with fixed contingent payoffs. With the logarithmic scoring rule, we get

$$\frac{1}{g'(g^{-1}(r))} = \frac{b}{r(1-r)}$$

which implies

$$g(Q) = \frac{1}{1 + e^{-Q/b}}$$

confirmed by Hanson's formula.

7. CONCLUSION

In this paper, we studied the problem of eliciting probabilistic information from an expert, when considering a random experiment with a finite set of outcomes. We made the following contributions:

1. We showed that distribution properties are elicitable *only* when, equivalently (a) their level sets are convex, (b) their level sets are convex maximal, (c) they form linear constraints.

2. We showed that properties are defined by their signature function, and provided a structural representation of incentive-compatible payments, described by two components: (a) a weight-function, independent of the property, which is any nonnegative density not locally null, and (b) the property's signature.

3. We investigated the elicitation of sets of properties, and showed that accuracy-rewarding multi-dimensional score functions are exactly those additively separable into one-dimensional score functions strictly proper for each property.

4. We introduced the complexity of distribution properties, that represent the minimal size of a set of elicitable properties from which a given property may be inferred.

5. We showed that with n outcomes, the complexity of properties vary from $\mathcal{C}^1$ to $\mathcal{C}^{n-1}$, it particular, some properties are as hard to get as the full distribution.

6. We showed that prediction markets reveal the same properties as those directly elicitable in the single-agent context.

7. We showed that, for the case of continuous double auctions or automated market-makers, securities being traded for eliciting Γ take the form of gradients of first-order score functions for Γ.

There are various avenues for future research. First, our investigation concerned continuous properties. It turns out that our results do not apply for more general properties, in particular properties having a discrete range. In that case, convexity of level sets is still a necessary condition, but is no longer sufficient. Can we characterize elicitable discrete properties? The second question refers to the outcome space, which we assumed finite. Do our results hold for more general spaces, especially continuous spaces? We conjecture that this is not the case, as our results rely indirectly on the topological completeness of a property's domain. Perhaps a workaround is to strengthen the continuity conditions, or to allow scores to be generalized functions instead of well-defined real-valued functions. Finally, many open questions remain about sets of properties. The most important one being the characterization of elicitable sets: we know that convexity of level sets is a sufficient condition, but is it necessary? We can also ask whether an integral representation *a la* Schervish is possible, as we obtained for single properties. And, naturally, it would be very interesting to establish the complexity of common distribution properties.

Acknowledgements

The authors would like to thank Mohammad Mahdian for helpful discussions and suggestions.

8. REFERENCES

[1] G. Brier. Verification of Forecasts Expressed In Terms of Probability. *Monthly Weather Review*, 78(1):1–3, 1950.

[2] A. Buja, W. Stuetzle, and Y. Shen. Loss Functions for Binary Class Probability Estimation and Classification: Structure and Applications. *Working paper*, 2005.

[3] J. Cervera and J. Munoz. Proper scoring rules for fractiles. *Bayesian Statistics*, 5:513–519, 1996.

[4] J. Feigenbaum, L. Fortnow, D. Pennock, and R. Sami. Computation in a distributed information market. *Theoretical Computer Science*, 343(1-2):114–132, 2005.

[5] R. Forsythe, F. Nelson, G. Neumann, and J. Wright. Anatomy of an Experimental Political Stock Market. *The American Economic Review*, 82(5):1142–1161, 1992.

[6] D. Friedman. Effective Scoring Rules for Probabilistic Forecasts. *Management Science*, 29(4):447–454, 1983.

[7] T. Gneiting and A. Raftery. Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[8] I. Good. *Rational Decisions*. Springer-Verlag, 1997.

[9] R. Hanson. Combinatorial Information Market Design. *Information Systems Frontiers*, 5(1):107–119, 2003.

[10] A. Hendrickson and R. Buehler. Proper Scores for Probability Forecasters. *The Annals of Mathematical Statistics*, 42(6):1916–1921, 1971.

[11] A. Murphy and R. Winkler. Probability Forecasting in Meterology. *Journal of the American Statistical Association*, 79(387):489–500, 1984.

[12] R. Nelson and D. Bessler. Subjective Probabilities and Scoring Rules: Experimental Evidence. *American Journal of Agricultural Economics*, 71(2):363–369, 1989.

[13] F. O'Carroll. Subjective Probabilities and Short-Term Economic Forecasts: An Empirical Investigation. *Applied Statistics*, 26(3):269–278, 1977.

[14] L. Savage. Elicitation of Personal Probabilities and Expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

[15] M. Schervish. A General Method for Comparing Probability Assessors. *The Annals of Statistics*, 17(4):1856–1879, 1989.

[16] E. Shuford, A. Albert, and H. Edward Massengill. Admissible probability measurement procedures. *Psychometrika*, 31(2):125–145, 1966.

[17] D. Spiegelhalter. Probabilistic prediction in patient management and clinical trials. *Stat Med*, 5(5):421–33, 1986.

[18] C. Staël von Holstein. A Family of Strictly Proper Scoring Rules Which Are Sensitive to Distance. *Journal of Applied Meteorology*, 9(3):360–364, 1970.

[19] R. Winkler. Scoring rules and the evaluation of probabilities. *TEST*, 5(1):1–60, 1996.

[20] J. Wolfers and E. Zitzewitz. Prediction Markets. *The Journal of Economic Perspectives*, 18(2):107–126, 2004.