Elicitation and Evaluation of Statistical Forecasts *

Nicolas S. Lambert†‡

April 2012 (current version)

Abstract

This paper studies mechanisms for eliciting and evaluating statistical forecasts. Nature draws a state at random from a given state space, according to some distribution p. Prior to Nature's move, a forecaster, who knows p, provides a prediction for a given statistic of p. The mechanism defines the forecaster's payoff as a function of the prediction and the subsequently realized state. When the statistic is continuous with a continuum of values, the payoffs that provide strict incentives to the forecaster exist if and only if the statistic partitions the set of distributions into convex subsets. When the underlying state space is finite, and the statistic takes values in a finite set, these payoffs exist if and only if the partition forms a linear cross-section of a Voronoi diagram—that is, if the partition forms a power diagram—a stronger condition than convexity. In both cases, the payoffs can be fully characterized essentially as weighted averages of base functions.

*Preliminary versions appear in the proceedings of the 9^{th} and 10^{th} ACM Conference on Electronic Commerce, respectively "Eliciting Properties of Probability Distributions" (with David Pennock and Yoav Shoham) and "Eliciting Truthful Answers to Multiple-Choice Questions" (with Yoav Shoham). The latter paper was awarded an ACM Conference of Electronic Commerce Best Paper Award and the Best Student Paper Award.

†Stanford University Graduate School of Business. Email: nlambert@stanford.edu.

‡I am deeply indebted to David Pennock and Yoav Shoham for their guidance and invaluable advice. Part of this work was developed while I was visiting Yahoo! Research, New York. I am grateful to Jacob Abernethy, Jeremy Bulow, Christopher Chambers, Peter Cramton, Yossi Feinberg, Rafael Frongillo, Drew Fudenberg, David Kreps, Sebastien Lahaie, John Langford, Mohammad Mahdian, Roger Myerson, Michael Ostrovsky, Eduardo Perez, Philip Reny, Alvin Roth, Hugo Sonnenschein, Andrzej Skrzypacz, Adam Szeidl, Balázs Szentes and Robert Wilson for insightful comments and discussions. I thank seminar participants at various institutions. Part of this work is supported by Google Research and the National Science Foundation under grant No. 1101209. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author only and do not necessarily reflect the views of the funding agencies.

1 Introduction

A decision maker often has less information relevant for her decision than does some other agent (a forecaster). In this paper, I examine mechanisms for eliciting and evaluating information when such information consists of statistical forecasts. Given a state space Ω, Nature draws a state ω^* at random according to some probability distribution p. Before Nature chooses a state, a forecaster, who knows p, announces a prediction for a "statistic" of interest to the decision maker. In this paper a "statistic" has a fairly general meaning. It encompasses most uses of the term and captures all features a distribution can possess. In particular, a "statistic" can be a real-valued statistic such as the mean, median or variance of a random variable defined on Ω. It can also be multidimensional: for instance, it could be a confidence interval, or a variance-covariance matrix. Finally, it can be a discrete statistic that takes value in a more general set; for example, a ranking of events by likelihood, or a pair of most correlated components of a random vector.

The mechanisms I consider in this paper are scoring rules. A scoring rule assigns to the forecaster a payoff as a function of his prediction and the one realization of the state of nature, ω^*, that is observed ex-post. I focus on *strictly proper scoring rules*, scoring rules that give the forecaster (who seeks to maximize his expected payoff) a strict incentive to report truthfully: the forecaster maximizes his expected payoff if, and only if, he makes a truthful prediction.

I address two central questions. First, for which statistics does a strictly proper scoring rule exist? Second, for a given statistic, how can we construct such a strictly proper scoring rule? I show that for some statistics, a strictly proper scoring rule does not exist. There are many cases for which the declared predictions do not procure enough information to grant existence of strict incentives.[1] But there are also many cases for which it is enough.

I distinguish between two classes of statistics: the statistics that are continuous and take a continuum of values (for example, the mean or variance of a random variable), and the statistics that are discrete and take a finite number of values (for example, a ranking of events in order of likelihood, or the most correlated pair of components in a random vector). For continuous statistics, strict properness can be

[1]Note that, in this setting, the forecaster reports only a value for the statistic of interest, he does not report the full probability distribution.

achieved if and only if the statistic partitions the space of probability distributions into convex sets. That is, each level set of the statistic—those sets of distributions assigned to the same statistic value—must be convex. So, for example, a strictly proper scoring rule exists for the mean, but not for the variance. For discrete statistics, strictly proper scoring rules exist if and only if they partition the distributions into pieces of a Voronoi diagram, a geometric object more constraining than convex partitions. I will show that this implies, for example, that a strictly proper scoring rule can be constructed for the ranking of events, but not for the most correlated components. In both cases, the proper and strictly proper scoring rules are generated by the (continuous or finite) mixtures of some given functions that are entirely—and often uniquely—determined by the statistic itself.

For statistics that take values in a set endowed with a natural ordering of its elements (for example, the mode and the median of a random variable, that take real values), I introduce an alternative class of scoring rules, the *(strictly) order-sensitive scoring rules*. Order sensitivity means that the closer the prediction is to the true value of the statistic, in terms of its rank, the larger the expected payoff. For continuous statistics, strict order sensitivity is a property that all strictly proper scoring rules possess. The result carries over to discrete statistics, albeit in a limited sense: when they exist, the strictly order-sensitive scoring rules are exactly those that are strictly proper. However, the discrete statistics that admit a strictly order-sensitive scoring rule form only a small subset of all those that admit a strictly proper scoring rule. These statistics partition the distributions into "slices" separated by hyperplanes, a much stronger constraint. For example, a strictly order-sensitive scoring rule can be constructed for the median, but not for the mode.

The literature on forecast evaluation and elicitation goes back to Brier (1950). Brier envisioned a scheme to measure the accuracy of probability assessments for a set of events, in the context of weather forecasting. This scheme, the Brier score, was later recognized as part of a much larger family of functions that possess similar properties, the proper and strictly proper probability scoring rules. These scoring rules and their properties have been extensively studied (see for example McCarthy, 1956; De Finetti, 1962; Winkler and Murphy, 1968; Winkler, 1969; Hendrickson and Buehler, 1971; Savage, 1971; Schervish, 1989; Good, 1997; Selten, 1998; Dawid, 2007; Gneiting and Raftery, 2007). A vast majority of the literature has concentrated on eliciting the full information about p. [2] In contrast, this paper deals with predictions regarding partial information in a general sense.

As probability scoring rules elicit the full distribution, any statistic can be theoretically elicited in some indirect fashion. As a practical matter though, eliciting the entire distribution becomes rapidly infeasible as the underlying state space grows large. This can be observed on two grounds. First, describing densities in full entails an amount of communication that, if benign for spaces that include only few states, is prohibitive in terms of effort or technological cost for spaces whose states vary along several dimensions. Second, if, to gain sufficient knowledge, the forecaster must spend a substantial number of hours in study and investigation, it is conceivable that the capability to estimate accurately the entire distribution comes at a much greater cost than learning some of its specific features. This paper does not attempt to model these phenomena, but they provide justification for studying ways to elicit a particular statistic of interest rather than the whole distribution.

This paper embraces the setting of Savage's seminal work (Savage, 1971). Like Savage, I concentrate on the basic model and abstract away from common problems, such as unknown risk-attitudes (Karni, 2009; Offerman et al., 2009), costly access to information (Osband, 1989), or forecasters endowed with different levels of information or skills (Olszewski and Sandroni, 2007, 2011). Although the current paper focuses on a simple setting, the results are also relevant to these more complex settings as I explain in Section 5.

Another branch of the literature is concerned with the testing of forecasts. The standard setting assumes an elicitor who repeatedly interacts with a self-proclaimed informed forecaster (Foster and Vohra, 1998). At each time period, the forecaster provides a probability assessment over outcomes that materialize at the following period. For instance, a weather forecaster is asked, every day, to predict the probability of rain for the following day. The main question of interest is whether, by looking at the sequence of predictions along the sequence of realizations, the elicitor can make the distinction between a truly knowledgeable forecaster and an impostor. Naturally, if the elicitor were able to assess the exactitude of the forecasts, she would be able to

[2] The exceptions I am aware of concern Fan (1975), Bonin (1976), and Thomson (1979) who design compensation schemes to elicit, from local branch managers, the production output that can be attained with some given probability. Also, in the context of government contracting, Reichelstein and Osband (1984) and Osband and Reichelstein (1985) establish incentive contracts that induce a contracting firm to reveal truthfully moments of its prior on the costs associated to the project.

enforce appropriate incentives to tell the truth, at least, in a repeated setting. However such tests do not exist: a major and surprising result of the literature essentially asserts that for every test that the informed forecaster passes, there exists a scheme that, when employed by a completely uninformed forecaster, makes him successfully pass with arbitrarily high probability (Olszewski and Sandroni, 2008; Shmaya, 2008). More positive results exist under variants of the basic setting, notably with multiple competing forecasters (see Feinberg and Stewart, 2008; Al-Najjar and Weinstein, 2008), when the forecaster's computational abilities are sufficiently limited (Fortnow and Vohra, 2009), by imposing strong enough restrictions on the class of distributions Nature can choose amongst (Olszewski and Sandroni, 2009b; Al-Najjar et al., 2010; Feinberg and Lambert, 2011; Stewart, 2011), or when the forecaster supplies all of his probability assessments upfront and the elicitor is allowed to make use of counterfactual predictions (Dekel and Feinberg, 2006; Olszewski and Sandroni, 2009a).

In this paper there is a single forecaster, who is known to be informed, and known to maximize his expected payoff. On the other hand, I impose no restriction on computational abilities nor on the distributions of Nature, there is a single data point (the one realization ω^* of the state) and I require strict incentives for truth-telling.

The paper proceeds as follows. Section 2 details the model. Sections 3 and 4 present the results concerning, respectively, the statistics that take a finite number of values, and the statistics that take values in a continuum. Section 5 concludes. Full proofs are relegated to the Appendix.

2 Model

Denote by Ω the set of states of Nature, and by $\Delta\Omega$ be a set of probability distributions over Ω.

Statistics. This paper is concerned with the elicitation and evaluation of statistical predictions. I give to statistics a broad meaning: they can describe any feature that a distribution may possess. Formally, a *distribution statistic*, or simply *statistic*, is defined as a pair (Θ, F). Θ is the *value set* of the statistic. It is the set in which the statistic takes values. F is the *level-set function*, it is a correspondence[3] $F : \Theta \rightrightarrows \Delta\Omega$. It assigns, to every θ, the collection $F(\theta)$ of all the distributions for which θ is a correct

[3]As usual, I assume without loss of generality that for every $\theta \in \Theta$, $F(\theta)$ is not empty.

statistic value. For instance, the mean of a random variable X could be written as a pair $(\mathbb{R}, F)$. Here a distribution p belongs to $F(m)$ if, and only if, the mean of X under p, $\int X \mathrm{d}p$, equals m.

Many statistics assign a unique value to a distribution. If so, the level sets $F(\theta), \theta \in \Theta$, are pairwise disjoint. These statistics are said to *exhibit no redundancy*. The mean, variance, or entropy share this property. Other statistics may associate several values to the same distribution. If so, the level sets overlap, and the statistic contains some amount of redundancy. For instance, the median exhibits some redundancy: in some cases, several median values may be associated with the same distribution. A *statistic function* is defined as a function Γ that associates a true statistic value $\Gamma(p)$ to every distribution p (formally, $\Gamma^{-1}(\theta) \subseteq F(\theta)$ for each θ). When the statistic has no redundancy, it is associated with only one statistic function. This function then suffices to conveniently represent the statistic. In general however, we need at least two statistic functions to fully describe a statistic.

I restrict the analysis to the statistics (Θ, F) that satisfy two conditions: *(a)* F satisfies $\bigcup_\theta F(\theta) = \Delta\Omega$ (the statistic is well defined for every distribution of $\Delta\Omega$); and *(b)* for all $\theta_1 \neq \theta_2$, $F(\theta_1) \nsubseteq F(\theta_2)$ (no statistic value is fully redundant). These assumptions are without loss of generality. All statistics can be redefined on the entire set $\Delta\Omega$ by assigning a dummy value to the distributions for which it is not originally properly defined. Moreover, as the scoring rules we seek to construct offer the same expected score for all correct predictions, removing statistic values that are fully redundant does not impact the analysis.

Scoring rules. The payoffs to the forecaster are specified by scoring rules, whose original definition is expanded to account for general statistical predictions. Given a statistic with value set Θ, a *scoring rule* is a function $S : \Theta \times \Omega \mapsto \mathbb{R}$ that assigns to every prediction θ and every state ω a real-valued score $S(\theta, \omega)$. Payoffs given by scoring rules may be interpreted and used in various ways, as long as the forecaster complies with the general principle that higher expected payoffs are systematically preferred (Winkler et al., 1996). To make the discussion concrete, in this paper a scoring rule specifies the payment the forecaster receives in exchange for his prediction.

The mechanisms considered throughout this paper are entirely specified by the statistic (F, Θ) we want to learn and the scoring rule S that assigns payoff values.

They operate in three stages.

t=1 Nature selects a distribution $p \in \Delta\Omega$.
The forecaster learns p.

t=2 The forecaster reports a prediction $\theta \in \Theta$.

t=3 Nature draws a state ω^* at random according to p.
The forecaster receives an amount $S(\theta, \omega^*)$.

Properness. If the forecaster is risk-neutral, his optimal response maximizes the expected payment. To induce the forecaster to answer honestly, we must construct scoring rules that are proper or strictly proper, in the terminology of De Finetti (1962) and Savage (1971). Proper scoring rules ensure that all true predictions get the maximum expected payment. With strictly proper scoring rules, the maximum is attained if and only if the prediction is correct. The definition is easily adapted to the general case of statistical predictions:

Definition 1. A scoring rule S for statistic (Θ, F) is *proper* if, for every prediction θ and every distribution p, whenever θ is true under p, i.e., whenever $p \in F(\theta)$,

$$\theta \in \arg\max_{\hat{\theta} \in \Theta} \mathop{\mathsf{E}}_{\omega \sim p} [S(\hat{\theta}, \omega)] \ .$$

S is *strictly proper* if it is proper and if, for every prediction θ and every distribution p, whenever θ is false under p, i.e., whenever $p \notin F(\theta)$,

$$\theta \notin \arg\max_{\hat{\theta} \in \Theta} \mathop{\mathsf{E}}_{\omega \sim p} [S(\hat{\theta}, \omega)] \ .$$

Because the mechanism controls precisely how much utility the forecaster gets for each draw of Nature, risk neutrality is not a binding assumption. [4] In the same fashion, forecasters need not be endowed with complete knowledge—but for simplicity

[4]Given any forecaster with strictly increasing utility for money u, any (strictly) proper scoring rule S can be transformed into another one, $u^{-1} \circ S$, that (strictly) induces honest reports from such forecasters, and conversely. Variations in risk attitude do not affect our ability to elicit a given statistical information. It merely implies a transformation of the compensation structure. Even if we were to ignore the forecaster's utilities, enforcing strict incentives remains possible. For example, we can use a two-stage mechanism that starts by estimating the agent's preferences, as in Offerman et al. (2009), or reward forecasters with lottery tickets to neutralize risk aversion, as in Karni (2009).

of exposition it is convenient to think they are. As long as the forecaster knows the true value of the statistic, (strictly) proper scoring rules guarantee a (strictly) maximal expected payment, regardless of Nature's distribution. Had the forecaster the chance to learn the full distribution, he would find himself unable to take advantage of that extra piece of information.

Order sensitivity. Properness is concerned with how the expected scores of correct predictions compare with those of incorrect predictions. However, properness does not look at how the expected scores of incorrect predictions compare with one another. Yet, a number of situations arise in which such comparisons are desirable. A prominent example is that of eliciting the mean of a random variable. Say, the true mean is 100. Proper scoring rules dictate that forecasting 100 will maximize expected payments. But what about forecasting 99 versus 10? Properness *a priori* does not preclude that the latter prediction yields a higher expected payment than the former.

More generally, suppose the statistic takes value in a set that is naturally ordered. Given a true prediction θ and two incorrect predictions θ_1, θ_2, it seems pretty uncontroversial that whenever θ_1 is "in-between" θ and θ_2, it constitutes a "more accurate" prediction than θ_2 does. In such situations, it makes sense to offer a contract that, on average, rewards prediction θ_1 with a higher amount than it does for θ_2. This requirement is captured via the notion of order sensitivity. [5]

Definition 2. A scoring rule S for a statistic (Θ, F) is *order sensitive* with respect to total order $\prec$ on the value set Θ if, for all distributions p, all predictions θ true under p, i.e., such that $p \in F(\theta)$, and all predictions θ_1, θ_2 such that either $\theta \preceq \theta_1 \prec \theta_2$ or $\theta_2 \prec \theta_1 \preceq \theta$,

$$\mathop{\mathsf{E}}_{\omega \sim p}[S(\theta_1, \omega)] \geq \mathop{\mathsf{E}}_{\omega \sim p}[S(\theta_2, \omega)] \ .$$

S is *strictly order sensitive* when the inequality is strict whenever $p \notin F(\theta_2)$.

To reduce notation $S(\theta, p)$ denotes the expected score under p, $\mathsf{E}_{\omega \sim p}[S(\theta, \omega)]$.

[5] Order sensitivity relates to the notion of scoring rule efficiency (Friedman, 1983; Nau, 1985). Scoring rule efficiency compares probabilistic predictions according to their distance to the true distribution, with respect to some metric. In contrast, order sensitivity compares statistical predictions according to their rank relative to the true statistic value.

3 Statistics with a Finite Number of Values

In this section the set Ω includes a finite number of states, and statistics take values in a finite set. I call such statistics *finite statistics*. $\Delta\Omega$ denotes the set of all the probability distributions over Ω. To reduce notation distributions are identified with their density functions and the terms are used interchangeably.

Denote by $\mathbb{R}^\Omega$ the space of functions that map states of Nature to real values, that is, $\mathbb{R}^\Omega$ is the space of random variables. Note that every distribution—or density function—over states, p, is an element of $\mathbb{R}^\Omega$. The set $\mathbb{R}^\Omega$ is naturally endowed with the scalar product

$$\langle X, Y \rangle = \sum_\omega X(\omega)Y(\omega) \ .$$

This enables to write the expected payoff to the forecaster who announces θ,

$$\mathop{\mathsf{E}}_{\omega \sim p}[S(\theta, \omega)] \ ,$$

as the scalar product between the random payment $S(\theta, \cdot)$ and the density function p,

$$\langle S(\theta, \cdot), p \rangle \ .$$

The results of this section have a natural geometric interpretation, and viewing the set of distributions $\Delta\Omega$ as a simplex in the Euclidian space $\mathbb{R}^\Omega$ plays a key role in the arguments that follow. Besides, every finite statistic has a simple graphical representation as a finite covering of the simplex, covering that in most cases of interest is a partition except at boundary points. Several examples are included below.

3.1 Strictly Proper Scoring Rules

I begin with a description of the statistics that accept strictly proper scoring rules. We first observe that strictly proper scoring rules may exist, but do not always exist. For example, consider the problem of predicting which one of a finite number of events $E_1, \ldots, E_m \subset \Omega$ is most likely. Such a statistic can be elicited via the scoring rule defined by $S(E_i, \omega) = \mathbb{1}\{\omega \in E_i\}$ which, obviously, is strictly proper. Now consider a simple example in a three-state world, with a random variable X taking values 1, 2 or 3. Let us look at the statistic that indicates whether X has "high"

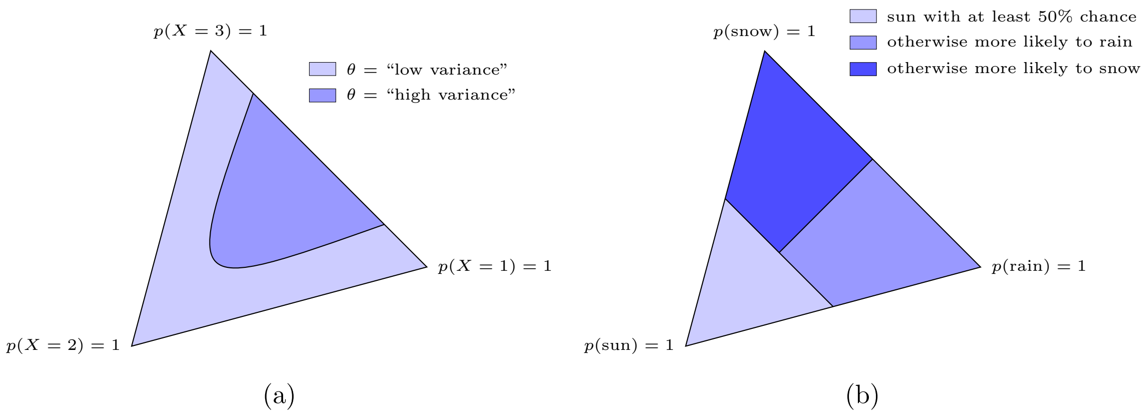


Figure 1: Statistics for the variance and for weather predictions.

or "low" variance, where the levels of variance are determined with respect to some arbitrary threshold. This statistic is depicted in Figure 1(a). In this case a strictly proper scoring rule does not exist. If a scoring rule S strictly incites the forecaster to make a truthful report when p has low variance, then $S(\text{“low variance”}, x) > S(\text{“high variance”}, x)$, as a distribution with an almost-sure state $X = x$ has a zero variance. But expected payments $S(\theta, p)$ are linear in the true distribution p. So for such a scoring rule, $S(\text{“low variance”}, p) > \text{“high variance”}, p)$ for every distributions p: the forecaster incited to make truthful predictions when the variance is low is always best off reporting low variance levels even when the true variance is high.

A necessary condition for existence of strictly proper scoring rules is that the level sets of the statistic be convex; that is, *the distributions that share the same statistic value must form a convex shape.* Consider two distributions over states, p and q. The argument relies on the simple observation that the expected payment to the forecaster when predicting θ under any mixture of p and q,

$$\mathop{\mathsf{E}}_{\omega \sim \lambda p + (1-\lambda) q} [S(\theta, \omega)] \ ,$$

equals the mixture of the expected payments when predicting θ separately on p and q,

$$\lambda \mathop{\mathsf{E}}_{\omega \sim p} [S(\theta, \omega)] + (1 - \lambda) \mathop{\mathsf{E}}_{\omega \sim q} [S(\theta, \omega)] \ .$$

Suppose S is a strictly proper scoring rule and θ is a prediction that is correct for both p and q. By reporting θ, the forecaster maximizes the expected payment under both distributions. Per the above equality, the payment remains optimal under any mixture of p and q. Since S is strictly proper, it must be the case that θ is a correct prediction for all mixtures of p and q. Hence all level sets must have a convex shape. Clearly, in the case of the variance depicted in Figure 1(a), the statistic does not partition the distributions into convex subsets.

However, in the case of finite statistics, convexity is not sufficient. [6] The exact characterization makes use of a well-known geometric structure called a *Voronoi diagram*. Voronoi diagrams specify, for a set of points called *sites*, the regions of the space that comprise the points closest to each site. Specifically, consider a metric space $\mathcal{E}$ with distance d, together with vectors $x_1, \ldots, x_n \in \mathcal{E}$ that are the sites. The *Voronoi cell* for site x_i includes all the vectors whose distance to x_i is less than or equal to the distance to any other site x_j. The collection of all the Voronoi cells is called the *Voronoi diagram* for the sites $x_1, \ldots, x_n$. Observe that the set of distributions, when viewed as a simplex in $\mathbb{R}^\Omega$ inherits its Euclidian metric. In this context it makes sense to talk about *Voronoi diagrams of distributions*, as well as *Voronoi diagrams of random variables*, since random variables are the elements of $\mathbb{R}^\Omega$. Voronoi diagrams are used in variety of fields, including mathematics, computer science, econometrics, and economics. See Aurenhammer (1991) or De Berg et al. (2008) for a literature review.

To understand the role Voronoi diagrams play in the characterization, it is helpful to start off with a simple sufficient condition: *if the level sets of a finite statistic form a Voronoi diagram of distributions, then there exists a strictly proper scoring rule for the statistic.* The argument is as follows. Let (Θ, F) be a finite statistic. Let each level set $F(\theta)$ be the Voronoi cell of some distribution $q_\theta \in \Delta\Omega$. Suppose that the forecaster is allowed to announce a full distribution q, and is rewarded according to

[6] This can be seen with the statistic pictured in Figure 1(b). Consider three possible states of the weather tomorrow: sunny, rainy, or snowy. We want to know if there will be sun with at least 50% chance (θ_A), or, if not, whether it is more likely to rain (θ_B) or to snow (θ_C). The statistic partitions the distributions in convex subsets. Yet there does not exist a strictly proper scoring rule. To see this, let us use the notation $p = (p(\text{sunny}), \text{p(rain)}, \text{p(snow)})$. Let $p_0 = (1,0,0)$, $p_1 = \left(\frac{1}{2}, \frac{1}{2}, 0\right)$, $p_2 = \left(\frac{1}{2}, 0, \frac{1}{2}\right)$, $p_3 = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$. Consider a proper scoring rule S. Both predictions θ_A and θ_B are true under p_1, so $S(\theta_A, p_1) = S(\theta_B, p_1)$. Similarly, $S(\theta_A, p_2) = S(\theta_C, p_2)$, $S(\theta_A, p_3) = S(\theta_B, p_3) = S(\theta_C, p_3)$, $S(\theta_B, p_0) = S(\theta_B, p_0)$. By linearity of the expected score, $2S(\theta_A, p_3) = S(\theta_A, p_1) + S(\theta_A, p_2)$, so $2S(\theta_C, p_3) = S(\theta_B, p_1) + S(\theta_C, p_2)$ implying $S(\theta_B, p_1) = S(\theta_C, p_1)$. Also, since the vectors p_0, p_1, p_2 are independent, $S(\theta_B, \cdot)$ is entirely specified by $S(\theta_B, p_0), S(\theta_B, p_1), S(\theta_B, p_3)$, and $S(\theta_C, \cdot)$ is entirely specified by $S(\theta_C, p_0), S(\theta_C, p_1), S(\theta_C, p_3)$. However, $S(\theta_B, p_0) = S(\theta_C, p_0)$, $S(\theta_B, p_3) = S(\theta_B, p_3)$, and $S(\theta_B, p_1) = S(\theta_C, p_1)$. Hence $S(\theta_A, \cdot) = S(\theta_B, \cdot)$ and S cannot be strictly proper.

the Brier score $S(q,\omega) = 2q(\omega) - \|q\|^2$. Aside from being strictly proper, the Brier score has the property that the closer the announced distribution is to that of Nature (in the Euclidian distance), the larger the expected payments (Friedman, 1983). In consequence if we were to force the forecaster to choose his report among the set of Voronoi sites $\{q_\theta, \theta \in \Theta\}$, his best response would be to produce the q_θ that is the closest to Nature's distribution. By forcing the forecaster to report one of these distributions, the forecaster reports the Voronoi cell that contains the distribution of Nature, thereby revealing a true value for the statistic. Because there is a one-to-one mapping between statistic values θ and sites q_θ, the reward scheme corresponds to asking a value θ for the statistic and paying the forecaster according to the strictly proper scoring rule $S(\theta,\omega) = 2q_\theta(\omega) - \|q_\theta\|^2$.

That the statistic partition $\Delta\Omega$ into a Voronoi diagram of distributions is not necessary, because the logic of the above argument applies to other probability scoring rules and other distances. But it leads the way to the exact characterization, which turns out to be a generalization of this result. Instead of focusing on a Voronoi diagram in the space of distributions, we look at a Voronoi diagram in the entire space $\mathbb{R}^\Omega$. Specifically, *the statistics that admit strictly proper scoring rules are precisely those whose level sets are included in a Voronoi diagram of random variables.*

Theorem 1. *Let (Θ, F) be a finite statistic. There exists a strictly proper scoring rule if and only if there exists a Voronoi diagram $\{\mathcal{C}_\theta\}_{\theta\in\Theta}$ in the space of random variables such that for every $\theta \in \Theta$, $F(\theta) = \mathcal{C}_\theta \cap \Delta\Omega$.*

Intersections of Voronoi diagrams with linear subsets are otherwise known as *power diagrams* in these subsets (Imai et al., 1985; Aurenhammer, 1987). Power diagrams are often interpreted as extensions of Voronoi diagrams in which a weight factor on the sites shifts the distances between vectors and sites. With that identification in mind, the theorem can be reformulated as follows: *there exists a strictly proper scoring rule if and only if the level sets of the statistic form a power diagram of distributions.*

Proof of Theorem 1. Let (Θ, F) be a finite statistic and let $\{X_\theta\}_{\theta\in\Theta}$ be a family of random variables indexed by statistic values. Consider the Voronoi diagram of this family in the space $\mathbb{R}^\Omega$. Denote by $\mathcal{C}_\theta$ the Voronoi cell for X_θ, that is, the set of all the functions $X : \Theta \to \mathbb{R}$ that are at least as close to X_θ as to any other site $X_{\theta'}$, with respect to the Euclidian distance. Suppose $F(\theta)$ is the part of $\mathcal{C}_\theta$ that is located on the simplex.

Consider the scoring rule $S(\theta, \omega) = 2X_\theta(\omega) - \|X_\theta\|^2$. The expected payment for prediction θ, under distribution p, is

$$\mathop{\mathsf{E}}_{\omega \sim p}[S(\theta, \omega)] = 2\langle X_\theta, p \rangle - \|X_\theta\|^2 = \|p\|^2 - \|p - X_\theta\|^2 \ .$$

This means that the expected payment of prediction θ under p is maximized across all possible predictions if and only if

$$\|p - X_\theta\| \leq \|p - X_{\hat{\theta}}\| \qquad \forall \hat{\theta} \in \Theta \ ,$$

which is to say that p belongs to the Voronoi cell $\mathcal{C}_\theta$ of X_θ. Since $F(\theta) = \mathcal{C}_\theta \cap \Delta\Omega$, the expected payment of prediction θ under p is maximized if and only if $p \in F(\theta)$, thereby establishing the strict properness of S.

To get the converse, assume there exists a strictly proper scoring rule S for a finite statistic (Θ, F). We need to construct random variables $\{X_\theta\}_{\theta \in \Theta}$ such that the associated Voronoi diagram in $\mathbb{R}^\Omega$ partitions the simplex $\Delta\Omega$ into the level sets of the statistics. To do so, we will use $X_\theta(\omega) = S(\theta, \omega) + k_\theta$, where k_θ is a constant to be specified later.

Saying that distribution p is in the Voronoi cell of X_θ is saying that

$$\|p - X_\theta\|^2 \leq \|p - X_{\hat{\theta}}\|^2 \qquad \forall \hat{\theta} \in \Theta \ ,$$

or equivalently, after expanding the terms,

$$-\|S(\theta, \cdot) + k_\theta\|^2 + 2k_\theta + 2\langle S(\theta, \cdot), p\rangle \geq -\|S(\hat{\theta}, \cdot) + k_{\hat{\theta}}\|^2 + 2k_{\hat{\theta}} + 2\langle S(\hat{\theta}, \cdot), p\rangle \quad \forall \hat{\theta} \in \Theta \ .$$

If the choice in k_θ is such that $\|S(\theta, \cdot) + k_\theta\|^2 - 2k_\theta$ equals a constant C independent of θ, we can cancel these terms and the last inequality becomes

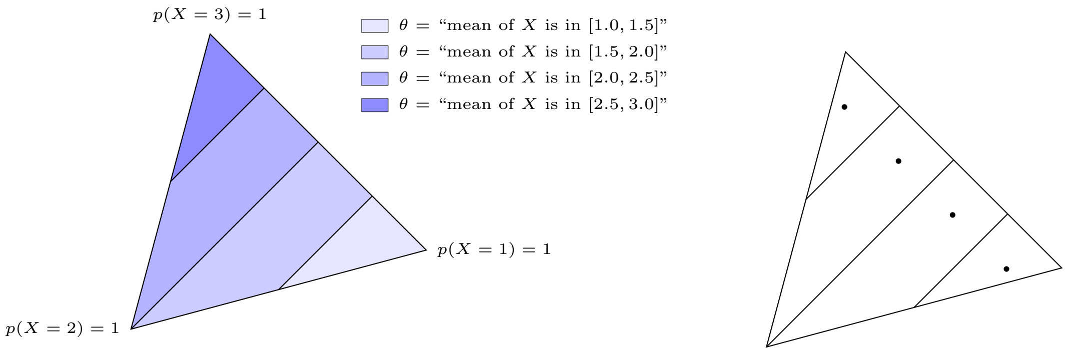
$$\mathop{\mathsf{E}}_{\omega \sim p}[S(\theta, \omega)] \geq \mathop{\mathsf{E}}_{\omega \sim p}[S(\hat{\theta}, \omega)] \qquad \forall \hat{\theta} \in \Theta \ .$$

Note that, for every θ, $\|S(\theta, \cdot) + k_\theta\|^2 - 2k_\theta$ is a parabola as a function of k_θ. As long as C is chosen to be greater that $\|S(\theta, \cdot)\|^2$ uniformly across statistic values—so that it intersects all the parabolas—it is always possible to select constants k_θ that satisfy this requirement. For such a choice of k_θ and X_θ, we have that for every distribution of Nature p, announcing prediction θ maximizes the expected payment if and only if p is located in the Voronoi cell of X_θ. As S is strictly proper, a prediction θ maximizes the expected payment if and only if it is true, that is, if $p \in F(\theta)$. Combining the two statements, we find that every level set $F(\theta)$ is the part of the Voronoi cell of X_θ located on the the simplex $\Delta\Omega$. □

The theorem essentially asserts that, as scoring rules vary, their associated value functions project onto power diagrams – or equivalently onto linear cross sections of Voronoi diagrams. Indeed, given a scoring rule S, the forecaster gets as expected payment $\max_\theta S(\theta, p)$. The expected payment, as a function of the true distribution of Nature p, is the value function. Saying that S is strictly proper is equivalent to saying that the projection of associated value function on the domain of distributions partitions $\Delta\Omega$ *exactly* as the level sets of the statistic $F(\theta)$ do. The statistics we can elicit via strictly proper scoring rules therefore correspond to the projections of all the value functions. In the case of a finite statistic, the value function describes the upper envelope of a finite number of non-vertical hyperplanes. Moreover, by an appropriate choice of S, any such envelope can be obtained. Therefore the level sets of statistics we can elicit via strictly proper scoring rules correspond exactly to the projections of hyperplane envelopes, which turn out to be the power diagrams.

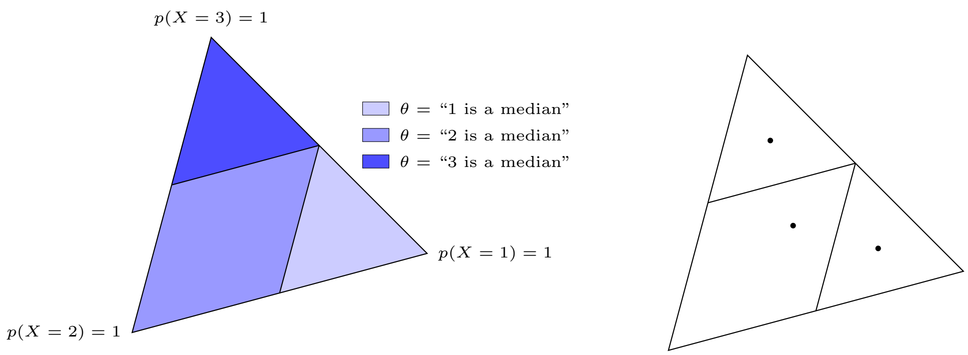
The geometric characterization of the Voronoi test is appealing. As long as the dimension of the simplex of distributions is small, a quick visual check gives a good sense of whether the statistic satisfies the condition of Theorem 1. Figures 2 to 5 depict the simplex of distributions partitioned into level sets for, respectively, the rounded mean, the median, the most likely state and the ranking of states according to their probabilities, all of which are classic exemplars of finite statistics. The right side of each figure maps a Voronoi diagram (along with the sites, all located on the simplex) that matches exactly the partition of level sets. We conclude that all these statistics admit a strictly proper scoring rule. Naturally the number of states must be kept artificially low to enable a 2-dimensional rendering of the simplex. However the Voronoi construction typically extends directly to higher dimensional simplices. And in most cases, the 2-dimensional visual test is sufficient to get convinced of its existence.

It is not difficult to exhibit statistics that fail the Voronoi test. We already saw the variance fails the convexity test in Figure 1(a), a condition weaker than the Voronoi



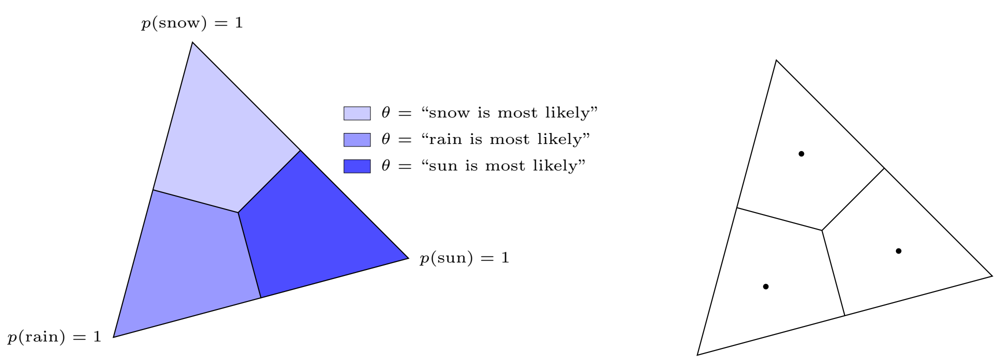
(a) (b)

Figure 2: The mean.



(a) (b)

Figure 3: The median.



(a) (b)

Figure 4: The most likely state of Nature.

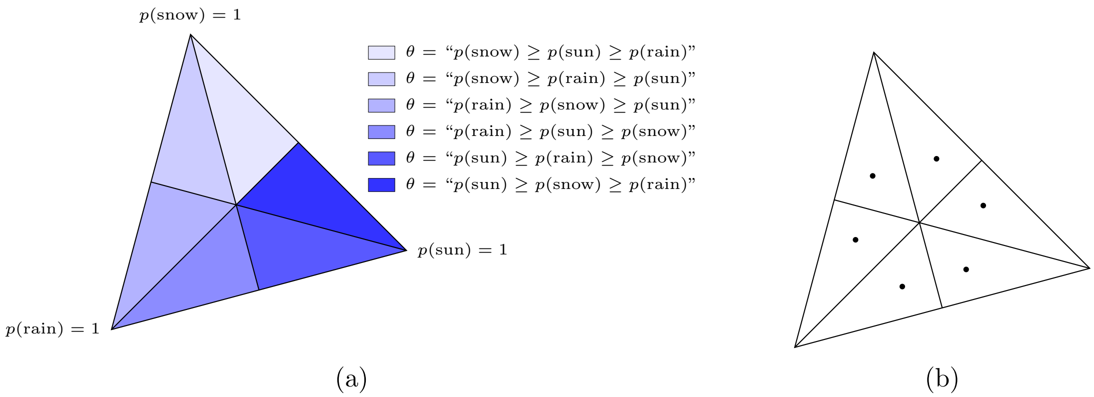


Figure 5: The ranking of states from most to least likely.

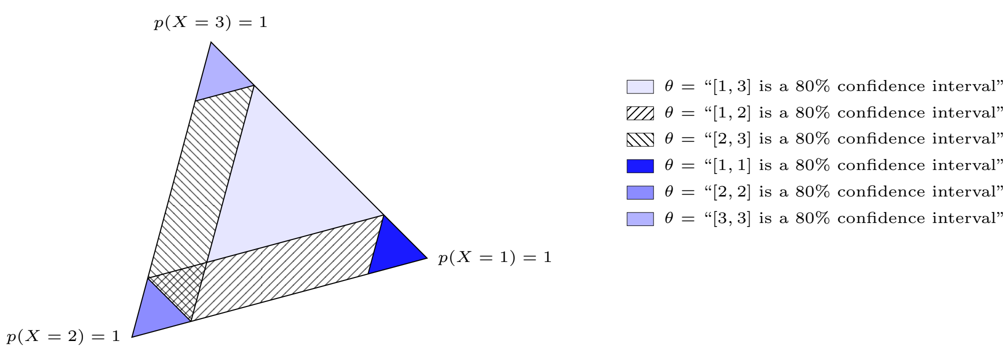


Figure 6: Confidence intervals.

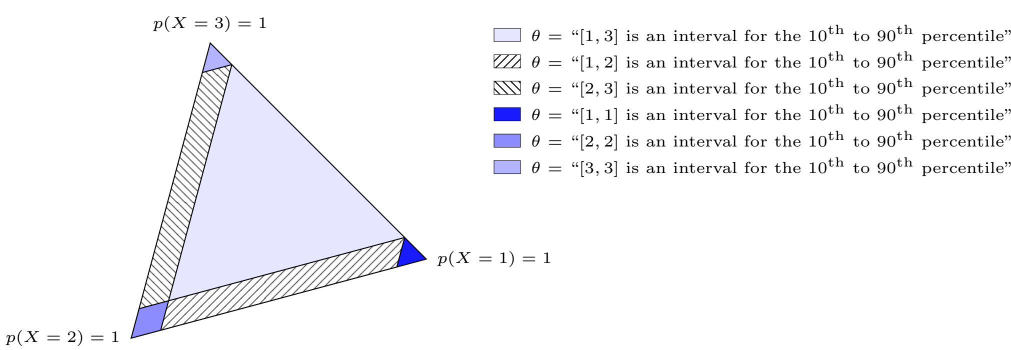


Figure 7: Intervals for the 10^{th} and 90^{th} percentiles.

test. In Figure 6, we are interested in a 80% confidence interval for a random variable.[7] The statistic fails the Voronoi test because of the large overlap for two of its level sets, corresponding to intervals $[1, 2]$ and $[2, 3]$. In this region, the densities cannot be equidistant to two distinct random variables. Whatever the scoring rule being used, there will be cases where a forecaster who reports one of the two intervals will not maximize his expected payoff, even when both intervals are correct. In general, to be able to elicit the predictions of a discrete statistic, there must exist situations for which two or more predictions are simultaneously correct. But those situations should almost never happen, in the sense that the level sets should be a proper partition of the space of distributions except for a measure zero set of points which belong to two or more level sets. To properly elicit confidence intervals, we must reduce the overlap. For example we can require that predictions take the form of symmetric intervals as in Figure 7, that are the ranges between the 10^{th} and the 90^{th} percentiles. It is easily seen that the Voronoi test is then satisfied, and so a strictly proper scoring rule exists.

Although Voronoi diagrams and convex partitions look alike, a Voronoi test can be much stronger than a convexity test. This is especially true in high dimensions. Nonetheless, most statistics that exhibit a high level of symmetry are naturally shaped as Voronoi diagrams. Non-symmetric cases can arise as well. For example, the statistic that gives the most likely of a list of (possibly overlapping) events. In such cases, the Voronoi sites are typically off the density simplex, precisely to shape the asymmetric structures. These cases are somewhat harder to visualize.

Now let us focus on a statistic that passes the Voronoi test of Theorem 1. How can we construct strictly proper scoring rules? The next result asserts that the (strictly) proper scoring rules are essentially the mixtures of a finite number of carefully chosen proper scoring rules. These proper scoring rules form a base. Fixed once and for all, the base is entirely determined by the statistic being elicited.

Theorem 2. *Let (Θ, F) be a finite statistic that satisfies the Voronoi test in Theorem 1. There exist $\ell \geq 1$ proper scoring rules $S_1, \dots, S_\ell$, called a base, such that a*

[7] For a discrete random variable X, $[a, b]$ is a 80% confidence interval if the probability that $X \in [a, b]$ is at least 80% and if there is no interval $[c, d] \subsetneq [a, b]$ for which $X \in [c, d]$ with at least 80% probability.

scoring rule S is proper (resp. strictly proper) if, and only if,

$$S(\theta, \omega) = \kappa(\omega) + \sum_{i=1}^{\ell} \lambda_i S_i(\theta, \omega) \ , \qquad \forall \theta \in \Theta, \omega \in \Omega \ ,$$

for some function $\kappa : \Omega \mapsto \mathbb{R}$, and nonnegative (resp. strictly positive) reals λ_i, $i = 1, \ldots, \ell$.

Proof intuition for Theorem 2. The proof is based on the following idea. Let S be a proper scoring rule. The properness property is captured by the following constraints:

$$S(\theta, p) \geq S(\hat{\theta}, p) \qquad \forall \theta, \hat{\theta} \in \Theta, \forall p \in F(\theta) \ .$$

There are uncountably many inequalities. However, whenever the statistic satisfies the criterion of Theorem 1 the sets $F(\theta)$ are polyhedra. Observing that the inequalities are linear in p, they need only be satisfied at the extreme vertices of these polyhedra. Thus the properness condition boils down to a finite system of homogeneous inequalities. By standard arguments (see, for example, Eremin, 2002), the solutions form a polytope that consists of a cone in the space of scoring rules, which is being copied and translated infinitely many times along some linear subspace. The directrices of the cone generate the "base" scoring rules. The kernel of the system, which gives rise to the translations, produces the complementary state-contingent payments. Adding strict properness substitutes some weak inequalities for strict ones in the above system, which complicates matters. Nonetheless the outcome remains intuitive: the strict inequalities only slightly perturb the solution space by excluding the boundary of the translated cone. In effect, this exclusion is responsible for the strictly positive weights to all the scoring rules of the base. □

For instance:

- For the statistic that gives the most likely of n arbitrary events $E_1, \ldots, E_n \subset \Omega$, there is only one base scoring rule, and the family of all the (strictly) proper scoring rules are written

$$S(E_k, \omega) = \kappa(\omega) + \lambda \mathbb{1}\{\omega \in E_k\} \ .$$

  for arbitrary functions κ and nonnegative (strictly positive) constants λ.

- For the median of a random variable X, the (strictly) proper scoring rules take the form

$$S(m,\omega) = \kappa(\omega) + \sum_{i=1}^{n} \lambda_i \cdot \left\{ \begin{array}{ll} -1 & \text{if } m < x_i, X(\omega) \leq x_i \\ 0 & \text{if } m \geq x_i \\ +1 & \text{if } m < x_i, X(\omega) > x_i \end{array} \right\} .$$

in which $x_1, \ldots, x_n$ are the values taken by X. After simplification, it can be seen that the scoring rules that are (strictly) proper for the median take the form

$$S(m,\omega) = \kappa(\omega) - |g(m) - g(X(\omega))| \ ,$$

for arbitrary functions κ and g, where g is (strictly) increasing.

- Divide the range $[0,1]$ into n intervals of equal size, $\left[\frac{k-1}{n}, \frac{k}{n}\right]$, $k = 1, \ldots, n$. For the statistic that gives an interval that contains the probability for some given event $E \subset \Omega$, the (strictly) proper scoring rules are

$$S\left(\left[\tfrac{k-1}{n}, \tfrac{k}{n}\right], \omega\right) = \kappa(\omega) + \sum_{i=1}^{n-1} \lambda_i \cdot \left\{ \begin{array}{ll} n-i & \text{if } i < k, \omega \in E \\ 0 & \text{if } i \geq k \\ -i & \text{if } i < k, \omega \notin E \end{array} \right\} .$$

After simplification, we find that the (strictly) proper scoring rules for probability intervals take the form

$$S\left(\left[\tfrac{k-1}{n}, \tfrac{k}{n}\right], \omega\right) = \kappa(\omega) + \mathbb{1}\{\omega \in E\}(g(k) - g(1)) + \frac{1}{n}\sum_{i=1}^{k-1}(g(k) - g(i)) \ ,$$

for arbitrary functions κ and g, where g is (strictly) increasing.

Some of these examples are detailed below.

3.2 Strictly Order-Sensitive Scoring Rules

The remainder of this section discusses order sensitivity and its interplay with properness. Consider a scoring rule that takes value in a set attached with a natural ordering of its elements.

The result below is a test for the existence of strictly order-sensitive scoring rules.

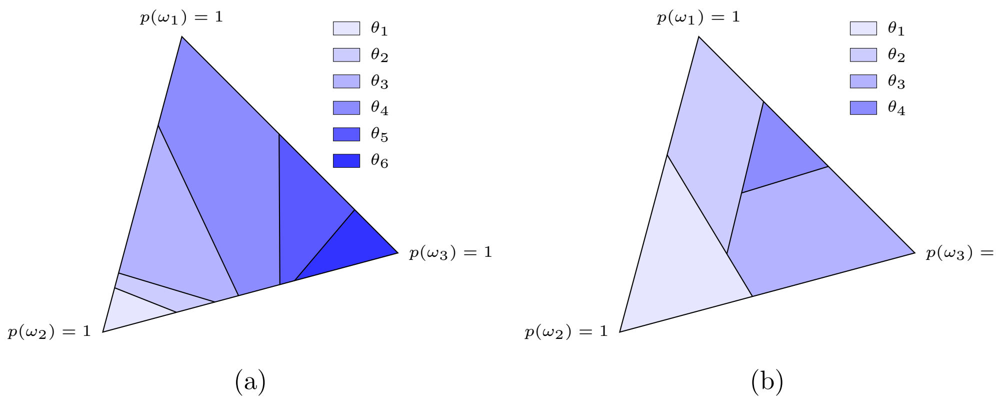


Figure 8: Strict order sensitivity can be enforced on the left statistic only.

As expected, the test is stronger than the Voronoi test of Theorem 1. But it is also easier to carry out. A statistic passes the test if and only if it partitions the distributions into "slices", as in Figure 8(a), and as opposed to Figure 8(b).

Theorem 3. *Let* $(\Theta = \{\theta_1, \ldots, \theta_n\}, F)$ *be a finite statistic, with* $\theta_1 \prec \cdots \prec \theta_n$*. There exists a scoring rule that is strictly order sensitive with respect to the order relation* $\prec$ *if, and only if, for all* $i = 1, \ldots, n-1$*,* $F(\theta_i) \cap F(\theta_{i+1})$ *is a hyperplane of* $\Delta\Omega$*.*[8]

Proof intuition for Theorem 3. The proof idea is best conveyed through an example. Consider a strictly order-sensitive scoring rule for a statistic whose value set Θ contains three elements, θ_1, θ_2 and θ_3. If both θ_1 and θ_3 are correct predictions under some distribution p, but θ_2 is not, then the expected payment, under p, is maximized only when responding θ_1 or θ_3. Adding a small perturbation to p, we can pull out a distribution $\tilde{p}$ for which the only true prediction is θ_1, while announcing θ_3 yields an expected payment that is nearly maximized and larger than that derived from announcing θ_2. This contradicts strict order sensitivity. This means that, whenever we choose some $p \in F(\theta_1)$ and $q \in F(\theta_3)$, the segment of distributions must go through $F(\theta_2)$. More generally, suppose the statistic takes more than three values. For any two distributions $p \in F(\theta_i)$ and $q \in F(\theta_j)$, $i < j$, the segment of distributions starting from p and ending at q must pass by, in order, through $F(\theta_i), F(\theta_{i+1}), \ldots, F(\theta_j)$, by which the hyperplane separation holds. The converse can be made clear through an explicit construction of the strictly order-sensitive scoring rules, which is the object of Theorem 4. $\square$

[8] Hyperplanes of distributions can be viewed as hyperplanes in the Euclidian space $\mathbb{R}^\Omega$ that intersect the simplex of density functions.

For example, we can apply Theorem 3 to the case of the median and the mode of a random variable X. For the median, Figure 3 suggests that the statistic passes the slice test of Theorem 3.[9] In contrast, consider the mode of X. This statistic gives the most likely value of X. Figure 4, for which the mode is a special case, clearly indicates that the statistic fails the test of Theorem 3.[10]

Strictly order-sensitive scoring rules are also strictly proper, and the form of the strictly proper contracts follows the rule given in Theorem 2. One benefit of statistics that admit strictly order-sensitive scoring rules is that the base scoring rules are easily derived; they are $0-1$ factors of the normals to the boundaries of consecutive level sets. Together Theorem 2 and Theorem 4 can be used to obtain the strictly proper scoring rules for any statistic that satisfies the condition of Theorem 3.

Theorem 4. *Let $(\Theta = \{\theta_1, \dots, \theta_n\}, F)$ be a finite statistic with $\theta_1 \prec \cdots \prec \theta_n$. Assume there exists a strictly order-sensitive scoring rule S with respect to the order relation $\prec$. The scoring rules $S_1, \dots, S_{n-1}$, defined by*

$$S_i(\theta_j, \omega) = \begin{cases} 0 & \text{if } j \le i \ , \\ \mathbf{n}_i(\omega) & \text{if } j > i \ , \end{cases}$$

form a base, with $\mathbf{n}_i$ being a positively oriented normal (i.e., oriented towards $F(\theta_{i+1})$) to the hyperplane of random variables in $\mathbb{R}^\Omega$ generated by $F(\theta_i) \cap F(\theta_{i+1})$.

The proof is immediate and relegated to the Appendix.

It remains to characterize the (strictly) order-sensitive scoring rules. As it turns out, as long as a strictly order-sensitive scoring rule exists, all the (strictly) proper scoring rules are also (strictly) order sensitive, so that the characterization of Theorem 4 still applies. The result implies that when a statistic admits a strictly order-sensitive scoring rule, it does so for exactly two order relations, one being the reverse

[9]Indeed, choosing two consecutive values for X, x and y, we easily verify that $F(x) \cap F(y)$ is a hyperplane. If both are possible median values under a distribution p, then $p(X \le x) \ge \frac{1}{2}$, $p(X \ge x) \ge \frac{1}{2}$, and $p(X \le y) \ge \frac{1}{2}$, $p(X \ge y) \ge \frac{1}{2}$. Hence $p(X > x) = p(X \ge y) \ge \frac{1}{2}$, and, as $p(X \le x) + p(X > x) = 1$, $p(X \le x) = \frac{1}{2}$. The converse is immediate. This means that the set $F(x) \cap F(y)$ is the hyperplane defined by $\sum_{z \le x} p(X = z) = \frac{1}{2}$. Hence the criterion of Theorem 3 is satisfied.

[10]To be convinced of this assertion, choose two consecutive values of X, x and y. The set $F(x) \cap F(y)$ contains all distributions p such that $p(X = x) = p(X = y)$, equality that indeed defines a hyperplane. However it is only part of a hyperplane, because there are distributions that assign the same probability to both x and y, and yet whose most likely values are attained elsewhere. As $F(x) \cap F(y)$ does not cover an entire hyperplane of distributions, it fails the above criterion.

of the other. As demonstrated in the sketch proof of Theorem 3, the result breaks down without the existence requirement. It breaks down even when restricted to weak order sensitivity, which, obviously, exists for all statistics.

Proposition 1. *Let $(\Theta = \{\theta_1, \ldots, \theta_n\}, F)$ be a finite statistic with $\theta_1 \prec \cdots \prec \theta_n$. Assume there exists a strictly order-sensitive scoring rule with respect to the order relation $\prec$. A scoring rule is (strictly) proper if and only if it is (strictly) order sensitive, with respect to $\prec$.*

The proof is direct and relegated to the Appendix. Take, for example, the median statistic of random variable X. Let x_i be the i-th smallest value taken by X. The hyperplane that separates two consecutive level sets of the median, for respective values x_k and x_{k+1}, is, as established previously, specified by equation $\sum_{i \le k} p(X = x_i) = \frac{1}{2}$. And so, the following

$$\mathbf{n}_k(\omega) = \begin{cases} -1 & \text{if } X(\omega) \le x_k \ , \\ +1 & \text{if } X(\omega) > x_k \ , \end{cases}$$

defines a positively oriented normal for each k. The normals generate the $n-1$ base scoring rules used at the end of the preceding subsection

$$S_k(m, \omega) = \begin{cases} -1 & \text{if } m < x_i, X(\omega) \le x_i \ , \\ 0 & \text{if } m \ge x_i \ , \\ +1 & \text{if } m < x_i, X(\omega) > x_i \ . \end{cases}$$

Now let us return to the statistic that gives a probability interval $\mathcal{I}_k = \left[\frac{k-1}{n}, \frac{k}{n}\right]$ for an event E. These intervals are naturally ordered by $\mathcal{I}_1 \prec \cdots \prec \mathcal{I}_n$. The hyperplane that separates two consecutive level sets is specified by the set of distributions p such that both $p(E) \in \mathcal{I}_k$ and $p(E) \in \mathcal{I}_{k+1}$, that is, $p(E) = \sum_{\omega \in E} p(\omega) = k/n$. Hence the following

$$\mathbf{n}_k(\omega) = \begin{cases} 1 - k/n & \text{if } \omega \in E \ , \\ -k/n & \text{if } \omega \notin E \ , \end{cases}$$

defines a positively oriented normal for every k, from which we derive the base scoring rules presented at the end of the preceding subsection.

4 Statistics with a Continuum of Values

This section focuses on statistics that take values in a one-dimensional continuum. The state space Ω is a compact metric space, and $\mathcal{G}$ its Borel σ-algebra. Let μ be an arbitrary finite reference measure on $(\Omega, \mathcal{G})$. The set of distributions of Nature, $\Delta\Omega$, is the set of all probability distributions over $(\Omega, \mathcal{G})$ that are absolutely continuous with respect to μ, and whose density functions are continuous and strictly positive over Ω.[11] Denote by $\mathcal{C}(\Omega)$ the set of continuous functions over Ω endowed with the L^1 norm $\|f\| = \int |f| \mathrm{d}\mu$. For notational convenience, probability distributions are identified with their density functions. $\Delta\Omega$ is viewed as a subset of $\mathcal{C}(\Omega)$ that inherits the same topology.

For technical tractability, in the remainder of this section I restrict attention to statistics (Θ, F) that satisfy three conditions:

Real Valued Θ is a subset of the real line.

No Redundancy The sets $F(\theta)$ are pairwise disjoint.

Continuity Their (unique) statistic function is continuous and nowhere locally constant[12].

These statistics are called *regular real-valued continuous statistics*. As long as we are interested in statistics that vary along a single dimension, these assumptions are not very restrictive. Common statistics such as the mean and moments, median and quantiles, variance, entropy, skewness, kurtosis (when $\Omega = [a, b]$) and covariance (when $\Omega = [a, b] \times [c, d]$) all satisfy the three conditions. To guarantee existence of expectations, I focus on the scoring rules S such that the function $S(\theta, \cdot)$ is μ-integrable for all θ.

Finally, I focus the discussion on the (strict) properness property. Indeed, for the statistics that satisfy the three conditions, any scoring rule that is (strictly) proper is also (strictly) order sensitive, for the usual ordering on the real line.

Proposition 2. *Consider a statistic that satisfies conditions* (Real Valued) *and* (No Redundancy)*, and whose statistic function is continuous in some topology on* $\Delta\Omega$.[13]

[11] Positiveness is not required in the technical arguments but makes the results applicable to a broader class of statistics.

[12] The statistic function is not constant on any open set of distributions.

[13] By topology on $\Delta\Omega$ I mean topology inherited from the linear topological space $\mathcal{C}(\Omega)$, *i.e.*, one that makes addition and scalar multiplications continuous operations.

A scoring rule is (strictly) proper if and only if it is (strictly) order sensitive.

Convexity of the level sets remains a necessary condition as for finite statistics. However, unlike the case of finite statistics, the continuity condition imposed on the statistics examined in this section makes this condition sufficient.

Theorem 5. *Let* (Θ, F) *be a regular real-valued continuous statistic. Each of the following statements imply the other two:*

1. *There exists a strictly proper scoring rule for* (Θ, F)*.*

2. *For all* $\theta \in \Theta$*,* $F(\theta)$ *is convex.*

3. *For all* θ *in the interior of* Θ*, there exists a continuous linear functional* L_θ *on* $\mathcal{C}(\Omega)$ *such that, for all distributions* $p \in \Delta\Omega$*,* $L_\theta(p) = 0$ *if and only if* $p \in F(\theta)$*.*

For instance, the mean, moments, median, quantiles of a random variable pass the convexity test (part (2) of the theorem)—and so can be elicited via a strictly proper scoring rule. However the variance, skewness and kurtosis fail the convexity test. The covariance of two random variables also fails the convexity test. Finally the entropy, which measures the level of uncertainty contained in a probability distribution, fails the convexity test as well. We therefore cannot properly incentivize forecasters to report values for these statistics.

Proof intuition for Theorem 5. Using standard notation, $\mathcal{L}$ denotes the Lebesgue measure, $L^1(\mu)$ the space of μ-integrable real functions on $(\Omega, \mathcal{G})$, and $L^\infty(\mu)$ the space of μ-measurable essentially bounded real functions on $(\Omega, \mathcal{G})$. The L^∞ norm is defined by $\|f\|_\infty = \sup\{m \mid \mu(f > m) = 0\}$. For members of $L^1(\mu)$ and $L^\infty(\mu)$ I make the usual μ-almost everywhere identification.

We have already seen how part (1) of the theorem implies part (2). Now assume part (2) is true. By continuity, Θ is an interval. Let θ be any statistic value in the interior of Θ. The distributions over states can be partitioned into three subsets, $\mathcal{D}^{<\theta}$, $\mathcal{D}^{=\theta}$, and $\mathcal{D}^{>\theta}$, that are respectively the sets of distributions whose statistic value is less than θ, equal to θ, and greater than θ. If we require that every level set of the statistic be convex, a continuity argument shows that all three sets are convex. Convexity of these three sets, along with compactness of the state space, enables the application of the Hahn-Banach separating hyperplane theorem under the L^∞-norm topology. Therefore there exists a hyperplane $\mathcal{H}_\theta$ that separates $\mathcal{D}^{<\theta}$ from $\mathcal{D}^{>\theta}$ and is closed in the L^∞ norm. As it turns out, the closeness carries over in the L^1-norm topology. The statistic being nowhere locally constant, the hyperplane ends up being the linear span of the set $\mathcal{D}^{=\theta}$, from which part (3) of the theorem follows. Hence part (2) implies part (3).

Finally, assume part (3) is true. Assume without loss of generality that $L_\theta(p)$ is strictly positive when p's statistic value is greater than θ, and strictly negative when p's statistic value is less than θ. Making use of the fact that the (topologic) dual of $L^1(\mu)$ is topologically isomorphic to $L^\infty(\mu)$, a version of the Riesz Representation theorem (for example, Theorem 1.11 of Megginson, 1998) gives a function $g_\theta \in L^\infty(\mu)$ such that, for all $f \in L^1(\mu)$,

$$L_\theta(f) = \int_\Omega f g_\theta \,\mathrm{d}\mu \ ,$$

where L_θ has been extended by continuity to the whole space $L^1(\mu)$, by the Banach extension theorem. It can be shown that the continuity of the statistic implies that $\theta \mapsto g_\theta$ is uniformly continuous on every segment of Θ, with respect to the L^1 norm. This allows existence of a bounded function $H : \Theta \times \Omega \mapsto \mathbb{R}$ that is $(\mathcal{L} \otimes \mu)$-measurable and is a good approximation of the g_θ's, in the sense that, for almost every θ, $\|H(\theta, \cdot) - g_\theta\| = 0$. Then, defining

$$S(\theta, \omega) = \int_{\theta_0}^{\theta} H(t, \omega)\mathrm{d}t \ ,$$

for an arbitrary θ_0 and in virtue of Fubini's theorem, we get, for all distributions p,

$$\begin{aligned}
\mathop{\mathsf{E}}_{\omega \sim p}[S(\theta, \omega)] &= \int_\Omega \left(\int_{\theta_0}^{\theta} H(t, \omega)\mathrm{d}t \right) p(\omega)\mathrm{d}\mu(\omega) \ , \\
&= \int_{\theta_0}^{\theta} \left(\int_\Omega H(t, \omega) p(\omega)\mathrm{d}\mu(\omega) \right) \mathrm{d}t \ .
\end{aligned}$$

Let θ^* be the unique value of the statistic under p. The condition imposed on H ensures that, for almost all $t \in (\theta, \theta^*)$, $\int_\Omega H(t, \cdot) p(\omega)\mathrm{d}\mu = L_t(p) > 0$ (with a symmetric inequality for $t \in (\theta^*, \theta)$), thereby making S strictly proper. Hence part (3) implies part (1). $\square$

Once it is established that the statistic of interest passes the convexity test of Theorem 5, it remains to design the strictly proper scoring rules. In the characterization below, I impose a smoothness condition on the scoring rules. Since the statistic varies continuously with the underlying distribution, it is reasonable to require that payments vary smoothly with the forecaster's prediction. To formalize the idea, I say that a scoring rule S is *regular* if it is uniformly Lipschitz continuous in its first variable: there exists $K > 0$ such that for all $\theta_1, \theta_2 \in \Theta$ and all $\omega \in \Omega$,

$$|S(\theta_1, \omega) - S(\theta_2, \omega)| \leq K|\theta_1 - \theta_2| \ .$$

Looking at the regular scoring rules as the only acceptable scoring rules does not limit the range of statistics to which the characterization applies. Regular strictly proper scoring rules are guaranteed to exist whenever the criteria of Theorem 5 are satisfied. On the other hand this restriction is useful in that it permits a simpler description of the scoring rules.

Assume the statistic passes the convexity test of Theorem 5. The next result asserts that there exists a particular *base scoring rule*, such that the family of strictly proper scoring rules is fully characterized (up to arbitrary state-contingent payoffs) by integrating the base scoring rule scaled by any nonnegative, nowhere locally zero weight. For proper scoring rules, the weight need only be nonnegative. In addition, the base scoring rule is unique up to a weight factor.

Theorem 6. *Let (Θ, F) be a regular real-valued continuous statistic such that, for every θ, the level set $F(\theta)$ is convex. There exists a $(\mathcal{L}\otimes\mu)$-measurable bounded scoring rule S_0 such that a regular scoring rule for the statistic is proper (resp. strictly proper) if, and only if, for all θ and μ-almost every ω,*

$$S(\theta, \omega) = \kappa(\omega) + \int_{\theta_0}^{\theta} \xi(t) S_0(t, \omega) \mathrm{d}t \ , \tag{1}$$

for some $\theta_0 \in \Theta$, $\kappa : \Omega \mapsto \mathbb{R}$, and $\xi : \Theta \mapsto \mathbb{R}_+$ a bounded Lebesgue measurable function (resp. and such that, for all $\theta_2 > \theta_1$, $\int_{\theta_1}^{\theta_2} \xi > 0$).

Proof intuition for Theorem 6. Most of the work in the proof of Theorem 5 is done for the purpose of building a function $H : \Theta \times \Omega \mapsto \mathbb{R}$ that is such that, for almost every θ and all $p \in \Delta\Omega$, the quantity

$$\int_{\Omega} H(\theta, \omega) \mathrm{d}p(\omega)$$

is respectively strictly positive when p's statistic value is greater than θ, strictly negative when it is less than θ, and zero when it equals θ. Now substitute S_0 for H in (1). By Fubini's Theorem, for all $p \in \Delta\Omega$, and all statistic values θ, θ^* where θ^* is a correct statistic value under p,

$$S(\theta^*, p) - S(\theta, p) = \int_{\theta}^{\theta^*} \xi(t) \left(\int_{\Omega} H(t, \omega) \mathrm{d}p(\omega) \right) \mathrm{d}t \ , \tag{2}$$

which makes S proper, and even strictly proper with the additional condition of positive integral on ξ.

We now get the converse. First, as $S(\cdot, \omega)$ is assumed to be Lipschitz continuous, it is in particular absolutely continuous. Hence it has an integral representation. There exists a function $G : \Theta \times \Omega \mapsto \mathbb{R}$, which can be chosen to be $(\mathcal{L} \otimes \mu)$-measurable, such that, for all θ, ω,

$$S(\theta, \omega) = \int_{\theta_0}^{\theta} G(t, \omega) \mathrm{d}t \ ,$$

where it is assumed without loss of generality that $S(\theta_0, \cdot) = 0$. Moreover, for all ω, $\theta \mapsto S(\theta, \omega)$ is differentiable for almost every θ and its derivative is given by G. Define the continuous functional Ψ_θ on $L^1(\mu)$ by

$$\Psi_\theta(f) = \int_{\Omega} G(\theta, \omega) f(\omega) \mathrm{d}\mu(\omega) \ .$$

The first step consists in showing that, outside a set of measure zero, $\theta \mapsto S(\theta, p)$ is differentiable for all distributions p, its derivative at θ being given by $\Psi_\theta(p)$. The second step makes use of the fact that when the expected payment is maximized, its derivative, when it exists, must be zero. Then, defining the continuous functional

$$\Phi_\theta(f) = \int_{\Omega} H(\theta, \omega) f(\omega) \mathrm{d}\mu(\omega) \ ,$$

an equality $\Psi_\theta = \xi(\theta)\Phi_\theta$ must hold for some $\xi(\theta)$. This is a mere consequence of the fact that the kernel of Φ_θ is the linear span of the distributions having statistic value θ. And so, in particular, the kernel of Φ_θ must be included in the kernel of Ψ_θ. ξ must remain nonnegative not to violate the order sensitivity property, which is implied by properness (Proposition 2). If, in addition, S is strictly proper, then the integral of ξ must be strictly positive on every segment, a direct consequence of (2). $\square$

The key to list all the (strictly) proper scoring rules is thus to find a possible candidate for S_0. Such candidate can typically be obtained in two ways. When we have a particular strictly proper scoring rule handy, we can differentiate it. Or we can apply part (2) of theorem 5 and write the Riesz representation of L_θ. I illustrate both ways below.

Let $\Omega = [a, b]$ be a segment of the real line, and μ be the Lebesgue measure. The distribution mean, whose statistic function is written

$$\Gamma(p) = \int_a^b \omega p(\omega)\mathrm{d}\omega \ ,$$

is (trivially) continuous, and satisfies the convexity condition of Theorem 5. It therefore admits a strictly proper scoring rule. In fact, we can easily find one: observing that the mean squared error $\mathsf{E}_{\omega\sim p}[(\theta-\omega)^2]$ is minimized precisely when θ equals the mean, paying the forecaster some positive amount minus the squared error yields a strictly proper scoring rule. Differentiating the quadratic term leads to a possible definition of S_0, $S_0(\theta, \omega) = \omega - \theta$. Altogether, Theorem 6 gives all the regular (strictly) proper scoring rules for the mean,

$$S(\theta, \omega) = \kappa(\omega) + \int_a^\theta (\omega - t)\xi(t)\mathrm{d}t \ .$$

The median, whose statistic function Γ is defined implicitly by

$$\int_a^{\Gamma(p)} p(\omega)\mathrm{d}\omega = \frac{1}{2} \ ,$$

is also continuous. This definition provides a continuous linear functional on $\mathcal{C}([a, b])$,

$$L_\theta(f) = \int_a^b \left[\mathbb{1}\{\theta \le \omega\} - \frac{1}{2}\right] f(\omega)\mathrm{d}\omega \ ,$$

leading to a candidate for the base scoring rule S_0,

$$S_0(\theta, \omega) = \mathbb{1}\{\theta \le \omega\} - \frac{1}{2} = \frac{1}{2}\left[\mathbb{1}\{\theta \le \omega\} - \mathbb{1}\{\omega \le \theta\}\right] \ .$$

Applying theorem 6 and letting $g(\theta) = \frac{1}{2}\int_a^\theta \xi(t)\mathrm{d}t$, we derive the regular (strictly) proper scoring rules for the median,

$$S(\theta,\omega) = \kappa(\omega) - |g(\theta) - g(\omega)| \ ,$$

By varying ξ under the constraints of Theorem 6, it is easily seen that g can be any weakly increasing (strictly increasing) Lipschitz function on $[a,b]$. Not surprisingly, the scoring rules for the median in the current continuous setting match those of the discrete setting of the preceding section. By the same logic, the (strictly) proper scoring rules that elicit an α-quantile take the form

$$S(\theta,\omega) = \kappa(\omega) + (2\alpha - 1)(g(\theta) - g(\omega)) - |g(\theta) - g(\omega)| \ .$$

It should be noted that continuity of the state space is not required. Dealing with finite state spaces is somewhat easier, because when equipped with the discrete metric, each of them is a compact metric space that makes continuous any function. A prevalent example is the dichotomous state space $\Omega = \{0,1\}$. The probability of occurrence of $\omega = 1$ is, obviously, a continuous statistic. Following the above example of the mean and according to Theorem 6, the form of its (strictly) proper scoring rules is given by

$$S(\theta,\omega) = \kappa(\omega) + \int_0^\theta (\omega - t)\xi(t)\mathrm{d}t \ ,$$

which is precisely the Schervish representation of probability scoring rules (Schervish, 1989). However, discrete state spaces must be used with care: statistics can be continuous with continuous states and lose continuity with discrete states. This true, in particular, of the median and quantiles.

Behind the integral representation of scoring rules lies a simple economic interpretation. From the perspective of the risk-neutral forecaster, being remunerated according to a (strictly) proper scoring rule is essentially the same as participating to an auction that sells off some carefully designed securities. Assume statistic values are bounded. Consider the auction that sells securities from a parametric family $\{R_\theta\}_{\theta\in\Theta}$; here $R_\theta(\omega)$ specifies the net payoff of the security R_θ when the realized state of Nature is ω. Net payoff is gross payoff minus initial price, which can be normalized to zero. In this auction, buyers bid on the security parameter. They are asked to bid the maximum value of θ for which they are willing to receive security R_θ. The winner is the bidder with the highest bid (ties are broken arbitrarily). Let us look at the special case in which the forecaster competes against a dummy bidder, whose bid y is distributed according to some density f. When the state ω^* materializes, the forecaster who bids x makes expected profit

$$P(y \leq x)\, \mathsf{E}[R_y(\omega^*) \mid y \leq x] = \int_{y \leq x} R_y(\omega^*) f(y) \mathrm{d}y \ . \tag{3}$$

Take any strictly proper scoring rule S of the form (1). Choosing density $f(y) = \xi(y)/\int_\Theta \xi$ and security $R_y(\omega) = \left(\int_\Theta \xi\right) S_0(y, \omega)$, (3) can be re-written

$$\int_{t \leq x} \xi(t) S_0(t, \omega^*) \mathrm{d}t \ ,$$

which is precisely the amount the forecaster would get through scoring rule S, up to a state-contingent payment. Conversely, take any bounded density function f nowhere locally zero. The forecaster's expected profit derived from participation in the auction equals the remuneration he would get with some strictly proper scoring rule. The auction interpretation is especially relevant when used on multiple forecasters, in which case dummy bidders are not needed. However, with multiple forecasters, these auctions no longer constitute the only valid incentive devices.

The family of securities that the must be auctioned off depends on the statistic of interest. When eliciting an event's probability, the goods for sale are securities that pay off the same positive amount if the event occurs and zero otherwise, minus the parameter. When eliciting a distribution's mean, they are securities that pay off a positive factor of the realized value of the underlying variable, minus the parameter. These auctions are essentially second-price auctions: for these families of securities, the gross payoff is fixed and buyers, in effect, end up bidding on the price. To elicit the median of a random variable, we should auction off securities whose net payoff $R_\theta(x) = 1$ when $x \geq \theta$ and $R_\theta(x) = -1$ if $x < \theta$ (or a positive factor of these) where θ is the parameter. Here the effect is opposite. The price of the security is fixed, say, 1, and buyers bid on a threshold parameter that only affect the gross payoff, which is then either 0 or 2, depending on whether the underlying variable realizes below or above the threshold.

The auction interpretation is also relevant to the context of prediction markets. Wolfers and Zitzewitz (2004) give examples of security markets for which bids provide aggregate information on various statistics. In each case, the securities exchanged are shaped in a particular way. The above argument demonstrates that the study of Wolfers and Zitzewitz (2004) generalizes. It is possible to use security markets to obtain aggregate estimates for *any* continuous statistic that can be elicited via a strictly proper scoring rule. Theorem 6 determines the shape of the securities to be used: these can take the form of any base scoring rule for the underlying statistic.

When we want forecasters to choose among a few alternative predictions, we can employ the finite statistics examined in the preceding section as approximations of continuous statistics. We can combine the results of the current and preceding sections to derive the (strictly) proper and (strictly) order-sensitive scoring rules. Let (Θ, F) be a regular real-valued continuous statistic. Let $\alpha_0, \alpha_n \in \mathbb{R} \cup \{+\infty, -\infty\}$ be respective lower and upper bounds of the possible values for the statistic, and $\alpha_1 < \cdots < \alpha_{n-1}$ be arbitrarily chosen in the interior of Θ (which is an interval). Consider a finite, approximate version $(\tilde{\Theta}, \tilde{F})$ of the continuous statistic. Instead of the exact value, it rounds up nearby statistical estimates: it gives an interval of the form $[\alpha_i, \alpha_{i+1}]$ that includes the exact value of the continuous statistic (Θ, F). $\tilde{F}([\alpha_i, \alpha_{i+1}])$ is therefore the set of distributions whose values, for the statistic (Θ, F), lie within $[\alpha_i, \alpha_{i+1}]$.

The collection of intervals, $\tilde{\Theta}$, is naturally equipped with the ordering $[\alpha_0, \alpha_1] \prec \cdots \prec [\alpha_{n-1}, \alpha_n]$. Suppose there exists a strictly proper scoring rule for the continuous statistic. Theorem 5 says that each $F(\theta)$ is a hyperplane of $\Delta\Omega$. As $\tilde{F}([\alpha_{i-1}, \alpha_i]) \cap \tilde{F}([\alpha_i, \alpha_{i+1}]) = F(\alpha_i)$, a direct application of Theorems 3 and 4 yields the following corollary:

Corollary 1. *If there exists a strictly proper scoring rule for statistic* (Θ, F)*, then there exist strictly proper (and strictly order-sensitive) scoring rules for the approximate statistic* $(\tilde{\Theta}, \tilde{F})$*. A scoring rule* S *is strictly proper (or strictly order sensitive) for the approximate statistic if, and only if,*

$$S([\alpha_k, \alpha_{k+1}], \omega) = \kappa(\omega) + \sum_{i<k} \lambda_i \mathbf{n}_i(\omega) \ ,$$

for any function $\kappa : \Omega \mapsto \mathbb{R}$ *and strictly positive real numbers* $\lambda_1, \ldots, \lambda_{n-1}$*,* $\mathbf{n}_i$ *being a positively oriented normal to the hyperplane generated by* $F(\alpha_i)$*.*

For example, consider again the statistic that gives, for an even E, some interval $\left[\frac{k-1}{n}, \frac{k}{n}\right]$ that includes the probability of E. Event E's probability is a regular continuous statistic. The hyperplane of distributions that give probability θ to the event has equation $\sum_{\omega\in E} p(\omega) = \theta$, and $\omega \mapsto \mathbb{1}\{\omega \in E\} - \theta$ is a positively oriented normal. This gives the strictly proper scoring rules

$$S\left(\left[\tfrac{k-1}{n}, \tfrac{k}{n}\right], \omega\right) = \kappa(\omega) + \sum_{i=1}^{n-1} \lambda_i \cdot \left\{ \begin{array}{ll} n-i & \text{if } i < k, \omega \in E \\ 0 & \text{if } i \geq k \\ -i & \text{if } i < k, \omega \notin E \end{array} \right\} .$$

As a final remark, the results of this section heavily rely on the continuity of the statistics, which can be severely constraining. While it may seem promising to look for a weaker notion of continuity so as to include a broader range of statistics, the choice of L^1-norm topology is not arbitrary. It turns out to be the strongest topology—and so the weakest continuity notion—which can be used. Indeed, whenever the statistic takes values in a bounded set, any statistic that is continuous in some topology, but not in the L^1 norm, never admits a strictly proper scoring rule.[14]

5 Concluding Remarks

This paper studies the problem of eliciting or evaluating statistical predictions. The forecaster's payoff is controlled via a generalized scoring rule, whose inputs are the announced statistical prediction and the state drawn by Nature. The focus is on two classes of statistics: the statistics that take values in a finite set, and the statistics that are continuous and take values in a one-dimensional continuum. In both cases, the paper provides a geometric characterization of the statistics that can be elicited via strictly proper scoring rules. For those statistics, it also describes the collection of proper and strictly proper scoring rules. For statistics that take values in an ordered set, additional characterizations are obtained.

The paper embraces the canonical setting of Savage (1971). Since then the literature has extended and expanded Savage's model in a number of different ways. Standard settings involve one or several forecasters who announce full probability

[14] To see why, consider a bounded statistic function Γ that is continuous in some topology. Take a sequence $(p_n)_n$ that converges towards some density p_∞ in the L^1 norm. If S is strictly proper, then $\mathsf{E}_{\omega\sim p_n}[S(\Gamma(p_\infty), \omega)] \leq \mathsf{E}_{\omega\sim p_n}[S(\Gamma(p_n), \omega)]$. But simultaneously, as n grows to infinity, $\mathsf{E}_{\omega\sim p_n}[S(\Gamma(p_n), \omega)]$ gets arbitrarily close to $\mathsf{E}_{\omega\sim p_\infty}[S(\Gamma(p_n), \omega)]$ when S is regular. Since S is also strictly order sensitive (Proposition 2 holds regardless of the topology that makes the statistic continuous), the only cluster point of $(\Gamma(p_n))_n$ is $\Gamma(p_\infty)$ and so $\Gamma(p_n) \to \Gamma(p_\infty)$.

assessments. In contrast, the current paper looks at the solicitation of partial beliefs. This complementarity makes it possible to combine the results of this paper and those of the past literature, so as to obtain generalizations of known results to situations of partial information elicitation. Because many known results hold under constraints similar to strict properness, the results of the current paper apply almost directly.

For example, a large literature is devoted to the evaluation of forecasters. It is well known that comparing the quality of predictions by averaging scores or payoffs over time, as in Winkler et al. (1996), necessitate exactly strict properness – no more, and no less. Standard results consider probability scoring rules, but they apply more broadly to any generalized scoring rule. In recent research, Olszewski and Sandroni (2008) and Shmaya (2008) established several important impossibility results regarding the problem of distinguishing between an informed forecaster and an uninformed one. Their results, which concern probabilistic predictions, apply more broadly to statistical predictions when the statistic admits a strictly proper scoring rule, because the proofs of these papers continue to hold as long as the underlying statistic has convex level sets. The same observation applies to the cross-calibration method of Feinberg and Stewart (2008), which tests multiple competing forecasters simultaneously.

Another literature is concerned with issues of incentives. For example, Olszewski and Sandroni (2007, 2011) considered the problem of designing screening contracts. These contracts pay a potential expert for producing a theory, which takes the form of a series of probabilistic predictions. The problem is to design the contract so as to attract the informed forecasters and deter the uninformed ones. Olszewski and Sandroni show that, in most cases, such a contract does not exist. Their results extend directly to the statistical predictions modeled in this paper. Other research, such as Karni (2009) and Offerman et al. (2009), looked at the problem of eliciting probabilities from arbitrary (non-)expected utility maximizers. Their method requires no more than the constraint of strict properness and is independent of the statistic that is being elicited. Methods for costly information acquisition, as in Osband (1989) and Clemen (2002), also require a strictly proper scoring rule, and work essentially with every strictly proper scoring rule: they only need that the scoring rule be scaled by a sufficiently large multiplier to incite the forecaster to exert the necessary effort.

A strand of the literature uses scoring rules in a market context. Ostrovsky (2011) studies the information aggregation properties of financial markets operated either by a batch auction or by a dealer. The dealer setting uses a fixed demand/supply schedule. It is well known that a fixed schedule can be modeled via a strictly proper scoring rule (Savage, 1971; Hanson, 2003). With minor modifications, one can derive analogous results with the scoring rules of the current paper and extend the results of Ostrovsky (2011) to a broader class of securities. Solutions to betting market designs, as proposed in Johnstone (2007) and Lambert et al. (2011), also transpose directly to statistical predictions.

To finish, this paper studies strictly proper scoring rules. There are statistics for which these scoring rules do not exist. Yet, eventually, all statistics can be elicited in some indirect fashion: we can use standard methods to elicit the full probability distribution, then subsequently compute the value of any statistic of interest. Asking for the full distribution may be impractical or unnecessarily cumbersom, but it is not always required. For example, it was shown in this paper that predictions of the variance on its own cannot be elicited. But predictions of the mean *and* the variance together can be elicited. This is a mere consequence of the fact that the mean and the variance are isomorphic to the first and second moments, which both admit strictly proper scoring rules. More generally, when statistic does not convey enough information to induce truthful reports, we can rely on a finer statistic. A fascinating question for future research is what is the smallest amount of information we must ask to obtain truthful answers to what we really want to know.

Appendix

A Proofs of Section 3

Through the proofs of this section, to reduce notation I often write, for a scoring rule $S : \Theta \times \Omega \mapsto \mathbb{R}$, $S(\theta)$ to denote the random variable $S(\theta, \cdot)$. For a subset of $\mathcal{S}$ of a vector space, denote by $\dim \mathcal{V}$ the dimension of its linear span. A convex polyhedra in a convex subset $\mathcal{C}$ of a vector space is nondegenerate when it has the same dimension as $\mathcal{C}$.

The proofs make use of the following lemma.

Lemma 1. *If there exists a strictly proper scoring rule for (Θ, F), then, for all θ, $F(\theta)$ is a nondegenerate closed convex polyhedra of $\Delta\Omega$, and, when the intersection of two elements $F(\theta_1)$ and $F(\theta_2)$ is not empty, it is a degenerate closed convex polyhedron.*

Proof. The lemma is a direct consequence of Theorem 1, which asserts that when a strictly proper scoring rule exists, the $F(\theta)$'s form a power diagram of distributions. □

A.1 Proof of Theorem 2

The proof uses the following lemma.

Lemma 2. *Let $\mathcal{E}$ be an n-dimensional Hilbert space with an inner product $\langle \cdot, \cdot \rangle$. Let $y_1, \ldots, y_m$ be m vectors that generate $\mathcal{E}$. Consider the two systems of inequalities*

$$\langle y_i, x \rangle \geq 0 \ , \qquad i \in \{1, \ldots, m\} \tag{4}$$

and

$$\langle y_i, x \rangle > 0 \ , \qquad i \in \{1, \ldots, m\} \ . \tag{5}$$

If both systems admit a nonempty set of solutions, then there exist vectors $s_1, \ldots, s_\ell$ of $\mathcal{E}$ such that the set of solutions of (4) *is $\{\lambda_1 s_1 + \cdots + \lambda_\ell s_\ell, \lambda_1, \ldots, \lambda_\ell \geq 0\}$ while the set of solutions of* (5) *is $\{\lambda_1 s_1 + \cdots + \lambda_\ell s_\ell, \lambda_1, \ldots, \lambda_\ell > 0\}$.*

Proof. As (4) is a homogeneous system of weak inequalities, its set of solutions is a cone. Let $\{s_1, \ldots, s_\ell\}$ be a set of directrices of the edges of this cone. As by assumption there exists a nonzero solution, this set is not empty. The parametric form of the solutions of (4) is given by the set $\{\sum_i \lambda_i s_i, \, \lambda_1, \ldots, \lambda_\ell \geq 0\}$ (Eremin, 2002). We shall see that the cone $\mathcal{C} = \{\sum_i \lambda_i s_i, \, \lambda_1, \ldots, \lambda_\ell > 0\}$ is the set of solutions of (5).
<u>Part 1.</u> This part shows that any element of $\mathcal{C}$ is solution of (5).

Each vector s_k of $\{s_1, \ldots, s_\ell\}$ is solution of a $(n-1)$-boundary system of the form

$$\begin{cases} \langle y_i, s_k \rangle = 0 \ , & i \notin I_k \ , \\ \langle y_i, s_k \rangle > 0 \ , & i \in I_k \ , \end{cases} \tag{6}$$

for I_k a subset of $\{1, \ldots, m\}$. Let x_0 be a solution of (5). Then x_0 is also solution of (4) and so $x_0 = \sum_i \lambda_i s_i$, with $\lambda_i \geq 0$ for all i. There cannot exist j with $\langle y_j, s_k \rangle = 0$ for all k, otherwise $\langle y_j, x_0 \rangle = 0$ and x_0 would not be solution of (5). Therefore $\cup_k I_k = \{1, \ldots, m\}$.

Let $\hat{x} \in \mathcal{C}$, with $\hat{x} = \sum_i \mu_i s_i$, with $\mu_i > 0$ for all i. Since $\cup_k I_k = \{1, \ldots, m\}$, for all j there exists k such that $\mu_k \langle y_j, s_k \rangle > 0$ and $\mu_k \langle y_i, s_k \rangle \geq 0$ for all $i \neq j$. By summation, for all i, $\langle y_i, \hat{x} \rangle > 0$, and so $\hat{x}$ is solution of (5).
<u>Part 2.</u> This part shows the converse, that any solution of (5) is in $\mathcal{C}$.

Let $\hat{x}$ be a solution of (5). Let $\mathcal{B}_0$ be the open ball of diameter δ centered on $\hat{x}$, and $\mathcal{B}_1$ the open ball of diameter $\frac{3}{4}\delta$ with the same center. If δ is chosen small enough, any vector of $\mathcal{B}_0$ is solution of (5) since its inequalities define an open set of $\mathcal{E}$.

For $\epsilon > 0$, let $t = \epsilon \sum_i s_i$, and let $\mathcal{B}_1' = \mathcal{B}_1 + t$ be the translated ball by t. If ϵ is chosen small enough, the open ball $\mathcal{B}_1'$ remains contained in $\mathcal{B}_0$. In such a case, $\hat{x}$, which also belongs to $\mathcal{B}_1'$, is the image of some $x_0 \in \mathcal{B}_1$. As x_0 is solution of (4), we can write $x_0 = \sum_i \lambda_i s_i$, with $\lambda_i \geq 0$ for all i, hence $\hat{x} = \sum \mu_i s_i$, with $\mu_i = \lambda_i + \epsilon > 0$ for all i. Therefore $\hat{x} \in \mathcal{C}$. This concludes the proof of the lemma. $\square$

Let us now return to the proof of the main theorem. Denote by $\mathcal{S}$ the space of scoring rules, *i.e.*, the linear space of functions $S : \Theta \times \Omega \mapsto \mathbb{R}$, considered as a Hilbert space whose inner product is defined as $\langle S_1, S_2 \rangle = \sum_{\theta,\omega} S_1(\theta,\omega) S_2(\theta,\omega)$.

<u>Part 1.</u> Suppose that there exists a strictly proper scoring rule for the statistic (Θ, F). $S \in \mathcal{S}$ is proper if, and only if, for all $\theta, \hat{\theta} \in \Theta$,

$$\langle S(\theta), p \rangle = \langle S(\hat{\theta}), p \rangle \qquad \forall p \in F(\theta) \cap F(\hat{\theta}) \ , \tag{7}$$

$$\langle S(\theta), p \rangle \geq \langle S(\hat{\theta}), p \rangle \qquad \forall p \in F(\theta) \backslash F(\hat{\theta}) \ , \tag{8}$$

with the last inequality being strict if and only if S is strictly proper.

By Lemma 1, for all $\theta \in \Theta$, the level set $F(\theta)$ is a bounded convex polyhedron, and so is the convex hull of a set of vertices $\mathcal{V}_\theta$. We can supplement the set of vertices $\mathcal{V}_\theta$ of each polyhedron $F(\theta)$ by vertices of the other polyhedra that belong to its boundary, in such a way that, for all $\theta, \hat{\theta} \in \Theta$, and all p belonging to both $F(\theta)$ and $\mathcal{V}_{\hat{\theta}}$, p also belong to $\mathcal{V}_\theta$. Let us write $\mathcal{V}_\theta$ as $\{p_1^\theta, \ldots, p_{\ell_\theta}^\theta\}$.

Let $S \in \mathcal{S}$ be proper (resp. strictly proper). Let $\theta, \hat{\theta} \in \Theta$. If $p \in \mathcal{V}_\theta \cap \mathcal{V}_{\hat{\theta}}$, then $p \in F(\theta) \cap F(\hat{\theta})$ and so by (7), $\langle S(\theta), p \rangle = \langle S(\hat{\theta}), p \rangle$. If $p \in \mathcal{V}_\theta \backslash \mathcal{V}_{\hat{\theta}}$, then $p \in F(\theta)$ and $p \notin F(\hat{\theta})$, since by construction of $\mathcal{V}_\theta$, $p \in F(\hat{\theta})$ and $p \in \mathcal{V}_\theta$ implies $p \in \mathcal{V}_{\hat{\theta}}$. So by (8), $\langle S(\theta), p \rangle \geq \langle S(\hat{\theta}), p \rangle$ (resp. $\langle S(\theta), p \rangle > \langle S(\hat{\theta}), p \rangle$).

We shall show the sufficiency of these two conditions. Assume that if $p \in \mathcal{V}_\theta \cap \mathcal{V}_{\hat{\theta}}$, then $\langle S(\theta), p \rangle = \langle S(\hat{\theta}), p \rangle$, and if $p \in \mathcal{V}_\theta \backslash \mathcal{V}_{\hat{\theta}}$, then $\langle S(\theta), p \rangle \geq \langle S(\hat{\theta}), p \rangle$ (resp. $\langle S(\theta), p \rangle > \langle S(\hat{\theta}), p \rangle$). Let $p \in F(\theta) \cap F(\hat{\theta})$. Then p is a linear combination of vectors in $\mathcal{V}_\theta$ and $\mathcal{V}_{\hat{\theta}}$, and since the equality $\langle S(\theta), q \rangle = \langle S(\hat{\theta}), q \rangle$ holds for all vectors q that belong to these two sets, by linearity $\langle S(\theta), p \rangle = \langle S(\hat{\theta}), p \rangle$. Now let $p \in F(\theta) \backslash F(\hat{\theta})$. Then $p = \sum_i \lambda_i p_i^\theta$ for some nonnegative scalars λ_i that sum to one. Since $p \notin F(\hat{\theta})$, there exists k such that $\lambda_k > 0$ and $p_k^\theta \notin F(\hat{\theta})$. Hence $p_k^\theta \in \mathcal{V}_\theta \backslash \mathcal{V}_{\hat{\theta}}$, and $\langle S(\theta), p_k^\theta \rangle \geq \langle S(\hat{\theta}), p_k^\theta \rangle$ (resp. $\langle S(\theta), p_k^\theta \rangle > \langle S(\hat{\theta}), p_k^\theta \rangle$). For $i \neq k$, we either have $p_i^\theta \in \mathcal{V}_\theta \cap \mathcal{V}_{\hat{\theta}}$ or $p_i^\theta \in \mathcal{V}_\theta \backslash \mathcal{V}_{\hat{\theta}}$, and so in both cases $\langle S(\theta), p \rangle \geq \langle S(\hat{\theta}), p \rangle$. Hence

$$\langle S(\theta), p \rangle = \sum_i \lambda_i \langle S(\theta), p_i^\theta \rangle \geq \sum_i \lambda_i \langle S(\hat{\theta}), p_i^\theta \rangle = \langle S(\hat{\theta}), p \rangle$$

with a strict inequality when S is strictly proper. Therefore, we have shown that a scoring rule S is proper if, and only if, S is solution of the following finite linear system in the space $\mathcal{S}$,

$$\begin{cases} \langle S(\theta) - S(\hat{\theta}), p \rangle = 0 \ , \quad \theta, \hat{\theta} \in \Theta, p \in \mathcal{V}_\theta \cap \mathcal{V}_{\hat{\theta}} \ , \\ \langle S(\theta) - S(\hat{\theta}), p \rangle \geq 0 \ , \quad \theta, \hat{\theta} \in \Theta, p \in \mathcal{V}_\theta \backslash \mathcal{V}_{\hat{\theta}} \ , \end{cases} \tag{9}$$

and S is strictly proper if, and only if, S is solution of the system

$$\begin{cases} \langle S(\theta) - S(\hat{\theta}), p \rangle = 0 \ , \quad \theta, \hat{\theta} \in \Theta, p \in \mathcal{V}_\theta \cap \mathcal{V}_{\hat{\theta}} \ , \\ \langle S(\theta) - S(\hat{\theta}), p \rangle > 0 \ , \quad \theta, \hat{\theta} \in \Theta, p \in \mathcal{V}_\theta \backslash \mathcal{V}_{\hat{\theta}} \ . \end{cases} \tag{10}$$

Part 2. Let $\mathcal{S}_0$ be the space of solutions of the finite system of equalities (in $\mathcal{S}$)

$$\langle S(\theta) - S(\hat{\theta}), p \rangle = 0 \ , \quad \theta, \hat{\theta} \in \Theta, p \in \mathcal{V}_\theta \cap \mathcal{V}_{\hat{\theta}}$$

corresponding to the first part of (9) and (10).

Step 1. Let $\mathcal{S}_0^\perp$ be the orthogonal complement of $\mathcal{S}_0$ in $\mathcal{S}$. Let $S \in \mathcal{S}_0$. Then, for any vector X of $\mathcal{S}$, $\langle X, S \rangle = \langle X^{\perp\perp}, S \rangle$, with $X^{\perp\perp} \in \mathcal{S}_0$ and where $X^{\perp\perp} + X^\perp$ is the decomposition of X according to the direct sum $\mathcal{S} = \mathcal{S}_0 \oplus \mathcal{S}_0^\perp$. Therefore, there exists vectors $Y_1, \ldots, Y_m$ in $\mathcal{S}_0$ such that the solutions of (9) in $\mathcal{S}$ are exactly the solutions of the finite system of weak linear inequalities in $\mathcal{S}_0$

$$\langle Y_i, S \rangle \geq 0, \quad i = 1, \ldots, m \tag{11}$$

and the solutions of (10) are the solutions of the finite system of strict linear inequalities in $\mathcal{S}_0$

$$\langle Y_i, S \rangle > 0, \quad i = 1, \ldots, m \ . \tag{12}$$

Step 2. Let $\mathcal{K}$ be the kernel of (11) in $\mathcal{S}_0$, and $\mathcal{K}^\perp$ be its orthogonal complement in $\mathcal{S}_0$. For each Y_i, write $Y_i^{\perp\perp} + Y_i^\perp$ its decomposition according to the direct sum $\mathcal{S}_0 = \mathcal{K} \oplus \mathcal{K}^\perp$.

We can easily describe $\mathcal{K}$: $S \in \mathcal{K}$ if and only if $S \in \mathcal{S}_0$, and if, for all $\theta, \hat{\theta} \in \Theta$ and all $p \in \mathcal{V}_\theta \backslash \mathcal{V}_{\hat{\theta}}$, $\langle S(\theta) - S(\hat{\theta}), p \rangle = 0$. Since $(\mathcal{V}_\theta \cap \mathcal{V}_{\hat{\theta}}) \cup (\mathcal{V}_\theta \backslash \mathcal{V}_{\hat{\theta}}) = \mathcal{V}_\theta$, $\mathcal{K}$ is simply the solution of

$$\langle S(\theta) - S(\hat{\theta}), p \rangle = 0 \ , \quad \theta, \hat{\theta} \in \Theta, p \in \mathcal{V}_\theta \ . \tag{13}$$

Any S such that $S(\theta) = S(\hat{\theta})$ for all $\theta, \hat{\theta} \in \mathcal{S}$ is solution. By Lemma 1, $F(\theta)$ has dimension $|\Omega|$ for all θ, and so the linear span of $\mathcal{V}_\theta$ is $\mathbb{R}^\Omega$. Consequently, if S is solution of (13), then $\langle S(\theta) - S(\hat{\theta}), p \rangle = 0$ for all $\theta, \hat{\theta}$ and all $p \in \mathbb{R}^\Omega$, implying $S(\theta) = S(\hat{\theta})$. Hence $\mathcal{K} = \{S \in \mathcal{S} \mid S(\theta, \omega) = S(\hat{\theta}, \omega) \, \forall \theta \neq \hat{\theta}\}$.

STEP 3. Let's consider the following two systems of inequalities in $\mathcal{K}^{\perp}$:

$$\langle Y_i^{\perp}, S\rangle \geq 0 \ , \quad i = 1, \ldots, m \tag{14}$$

and

$$\langle Y_i^{\perp}, S\rangle > 0 \ , \quad i = 1, \ldots, m \ . \tag{15}$$

If $S \in \mathcal{K}^{\perp}$, $\langle Y_i, S\rangle = \langle Y_i^{\perp}, S\rangle$, and the solutions of (11) (resp. (12)) are the elements of $\mathcal{K}$ added to the solutions of (14) (resp. (15)). The systems (14) and (15) have full rank in $\mathcal{K}^{\perp}$, and since by assumption there exists a strictly proper scoring rule, both admit at least one solution. By Lemma 2, there exist vectors $S_1, \ldots, S_\ell \in \mathcal{K}^{\perp}$ such that S is solution of (14) (resp. of (15)) if and only if S is a nonnegative (resp. strictly positive) linear combination of $S_1, \ldots, S_\ell$.

Therefore, S is solution of (9) (resp. of (10)) if, and only if, $S = \kappa + \sum_i \lambda_i S_i$, for $\kappa \in \mathcal{K}$ and $\lambda_1, \ldots, \lambda_\ell \geq 0$ (resp. $\lambda_1, \ldots, \lambda_\ell > 0$).

A.2 Proof of Theorem 3

If part. The construction of strictly order-sensitive scoring rules shall be done in Theorem 4 and Proposition 1.

Only if part. Let S be a strictly order-sensitive scoring rule.

STEP 1. This first step shows that for all i and $j > i+1$, if $p \in F(\theta_i)$ and $p \in F(\theta_j)$ then $p \in F(\theta_{i+1})$. Suppose by contradiction that there exist i and p with $p \in F(\theta_i)$, $p \notin F(\theta_{i+1})$, and $p \in F(\theta_j)$ for some $j > i+1$. By Lemma 1, $F(\theta_i)$ is a convex polyhedron of nonempty relative interior. Since $p \in F(\theta_i)$, there exists a sequence of vectors $\{p_k\}_{k\geq 1}$ of the relative interior of $F(\theta_i)$ that converges to p. By continuity $\lim_{k\to+\infty} S(\theta_i, p_k) \to S(\theta_i, p)$. Let $\delta_k = S(\theta_i, p_k) - S(\theta_{i+1}, p_k)$. Since p_k and p both belong to $F(\theta_i)$, but not to $F(\theta_{i+1})$, $\delta_k > 0$, and δ_k converges to $\delta = S(\theta_i, p) - S(\theta_{i+1}, p) > 0$. Therefore $\inf\{\delta_k\}_{k\geq 1} > 0$. Let $\epsilon = \inf\{\delta_k/2\}_{k\geq 1}$. By continuity, there exists K such that

$$|S(\theta_i, p) - S(\theta_i, p_K)| \leq \epsilon/2 \ ,$$

and

$$|S(\theta_j, p) - S(\theta_j, p_K)| \leq \epsilon/2 \ ,$$

so that, since θ_i and θ_j both contain p, $S(\theta_i, p) = S(\theta_j, p)$ and

$$|S(\theta_i, p_K) - S(\theta_j, p_K)| \leq \epsilon \ .$$

Hence, $S(\theta_j, p_K) > S(\theta_i, p_K) - \epsilon = S(\theta_{i+1}, p_K) + \delta_K - \epsilon > S(\theta_{i+1}, p_K)$. However, p_K is in the relative interior of $F(\theta_i)$, which means according to Lemma 1 that θ_i is the only true value of the statistic for p_K. But, since $i < i+1 < j$, and S is strictly order sensitive, we should have $S(\theta_{i+1}, p_K) > S(\theta_j, p_K)$. Contradiction.

<u>Step 2.</u> Now let $1 \leq j \leq n-1$. Let $B_j = F(\theta_1) \cup \dots \cup F(\theta_j)$, and $C_j = F(\theta_{j+1}) \cup \dots \cup F(\theta_n)$. By Lemma 1, B_j and C_j are polyhedra of dimension $|\Omega|$ and nonempty relative interior, with $B_j \cup C_j = \Delta\Omega$. Let $i \leq j < j+1 \leq k$. If $p \in F(\theta_i)$ and $p \in F(\theta_k)$, an iterative application of the claim of step 1 above yields $p \in F(\theta_i), F(\theta_{i+1}), \dots, F(\theta_k)$. In particular, $p \in F(\theta_j) \cap F(\theta_{j+1})$. Therefore $B_j \cap C_j = F(\theta_j) \cap F(\theta_{j+1})$. By Lemma 1, the dimension of $F(\theta_j) \cap F(\theta_{j+1})$ is at most $|\Omega| - 1$, so that there is a hyperplane of distributions $\mathcal{H}$ that contains $B_j \cap C_j$. Suppose that there exists a distribution p of $\mathcal{H}$ that does not belong to $B_j \cap C_j$. Since $B_j \cup C_j = \Delta\Omega$, $p \in B_j$ or $p \in C_j$. Suppose for example that $p \in B_j$. Then there exists a distribution q in the relative interior of C_j with $q \notin \mathcal{H}$. Note that the segment $]p, q]$ contains only elements of B_j or C_j. Since both sets are closed, the segment intersects $B_j \cap C_j$, which is impossible since $]p, q]$ does not intersect $\mathcal{H}$. So $B_j \cap C_j$ must be the full hyperplane of distributions $\mathcal{H}$: $\mathcal{H} = B_j \cap C_j = F(\theta_j) \cap F(\theta_{j+1})$. This concludes the proof.

A.3 Proof of Theorem 4

<u>Part 1.</u> Define

$$S(\theta_k, \omega) = \kappa(\omega) + \sum_{1 \leq i < k} \lambda_i \mathbf{n}_i(\omega) ,$$

with $\lambda_1, \dots, \lambda_{n-1} \geq 0$, and $\kappa \in \mathbb{R}^\Omega$.

As $\mathbf{n}_k$ is oriented positively, $\langle \mathbf{n}_k, p \rangle \geq 0$ for all $p \in F(\theta_{k+1}), \dots, F(\theta_n)$, and $\langle \mathbf{n}_k, p \rangle \leq 0$ for all $p \in F(\theta_1), \dots, F(\theta_k)$. The inequalities are strict if $p \notin F(\theta_k) \cap F(\theta_{k+1})$.

Let $p \in F(\theta_k)$. If $j < k$,

$$\mathop{\mathsf{E}}_{\omega \sim p}[S(\theta_k, \omega)] - \mathop{\mathsf{E}}_{\omega \sim p}[S(\theta_j, \omega)] = \sum_{j \leq i < k} \lambda_i \langle \mathbf{n}_i, p \rangle \geq 0 ,$$

and, if $j > k$,

$$\mathop{\mathsf{E}}_{\omega \sim p}[S(\theta_k, \omega)] - \mathop{\mathsf{E}}_{\omega \sim p}[S(\theta_j, \omega)] = -\sum_{k \leq i < j} \lambda_i \langle \mathbf{n}_i, p \rangle \geq 0 .$$

Therefore S is a proper scoring rule. If, in addition, $\lambda_1, \dots, \lambda_{n-1} > 0$, the inequalities become strict when $p \notin F(\theta_j)$, making S strictly proper.

<u>Part 2.</u> Now assume S is a proper scoring rule. Then, for all $p \in F(\theta_k) \cap F(\theta_{k+1})$, $1 \leq k < n$, $\langle S(\theta_k), p \rangle = \langle S(\theta_{k+1}), p \rangle$, and so $\langle S(\theta_{k+1}) - S(\theta_k), p \rangle = 0$. Theorem 3 says that $F(\theta_k) \cap F(\theta_{k+1})$ is a hyperplane of $\Delta\Omega$. Its linear span is a hyperplane $\mathcal{H}_k$ of $\mathbb{R}^\Omega$, thus

$S(\theta_{k+1}) - S(\theta_k) = \lambda_k \mathbf{n}_k$, where $\mathbf{n}_k$ is a normal to $\mathcal{H}_k$ oriented positively.

Let $p \in F(\theta_{k+1})$, $p \notin F(\theta_k)$. As S is proper, $\langle S(\theta_{k+1}), p\rangle \geq \langle S(\theta_k), p\rangle$, so $\lambda_k \langle \mathbf{n}_k, p\rangle \geq 0$. Since $p \notin \mathcal{H}_k$ and $\mathbf{n}_k$ is positively oriented, $\langle \mathbf{n}_k, p\rangle > 0$, implying $\lambda_k \geq 0$ ($\lambda_k > 0$ if S is strictly proper).

Therefore

$$S(\theta_k) = S(\theta_1) + \sum_{1 \leq i < k} (S(\theta_{i+1}) - S(\theta_i)) = \kappa + \sum_{1 \leq i < k} \lambda_i \mathbf{n}_i \ ,$$

with $\kappa = S(\theta_1)$.

A.4 Proof of Proposition 1

Assume the statistic accepts a strictly order-sensitive scoring rule with respect to the order relation $\prec$, and let $\theta_1 \prec \cdots \prec \theta_n$ be the elements of the value set of the statistic. Let S be a proper scoring rule. Theorem 4 shows that S takes the form

$$S(\theta_k, \omega) = \kappa(\omega) + \sum_{1 \leq i < k} \lambda_i \mathbf{n}_i(\omega) \ ,$$

with $\lambda_1, \ldots, \lambda_{n-1} \geq 0$. Let $p \in \Delta\Omega$. Since the normals are positively oriented, $\langle \mathbf{n}_k, p\rangle \geq 0$ if $p \in F(\theta_{k+1}), \ldots, F(\theta_n)$ and $\langle \mathbf{n}_k, p\rangle \leq 0$ if $p \in F(\theta_1), \ldots, F(\theta_k)$, the inequalities being strict if $p \notin F(\theta_k) \cap F(\theta_{k+1})$. So, for all $\theta, \theta_k, \theta_j$, if $\theta_j \prec \theta_k \prec \theta$ and $p \in F(\theta)$, then

$$\mathop{\mathsf{E}}_{\omega \sim p}[S(\theta_k, \omega)] - \mathop{\mathsf{E}}_{\omega \sim p}[S(\theta_j, \omega)] = \sum_{j \leq i < k} \lambda_i \langle \mathbf{n}_i, p\rangle \geq 0 \ .$$

Similarly, if $\theta \prec \theta_k \prec \theta_j$,

$$\mathop{\mathsf{E}}_{\omega \sim p}[S(\theta_k, \omega)] - \mathop{\mathsf{E}}_{\omega \sim p}[S(\theta_j, \omega)] = -\sum_{k \leq i < j} \lambda_i \langle \mathbf{n}_i, p\rangle \geq 0 \ .$$

Hence S is order sensitive. If S is strictly proper, the λ_i's are strictly positive, making the above inequalities strict, and S becomes strictly order sensitive.

B Proofs of Section 4

In the proofs that follow, $L^1(\mu)$ is the collection of μ-integrable real functions on $(\Omega, \mathcal{G})$. Unless mentioned otherwise, the topology used on $L^1(\mu)$, $\mathcal{C}(\Omega)$ and $\Delta\Omega$ (viewed as a subset of $\mathcal{C}(\Omega)$), will be that induced by the L^1 norm, denoted by $\|\cdot\|$ or sometimes $\|\cdot\|_1$ to avoid ambiguity. $L^\infty(\mu)$ is the collection of all μ-measurable, essentially bounded real functions on $(\Omega, \mathcal{G})$. The L^∞ norm is defined by $\|f\|_\infty = \sup\{m \mid \mu(f > m) = 0\}$. For members of $L^1(\mu)$ and $L^\infty(\mu)$ I make the usual μ-almost everywhere identification. Given a linear topological space $\mathcal{E}$, $\mathcal{E}^*$ represents the topological dual endowed with the operator norm. For an arbitrary function f, $\{f = \alpha\}$ denotes the set $\{x \mid f(x) = \alpha\}$, and $f_{|\mathcal{D}}$ the restriction of f to domain $\mathcal{D}$. Given a subset $\mathcal{S}$ of a linear (topological) space, $\mathcal{S}^\circ$ denotes the interior of $\mathcal{S}$, $\langle \mathcal{S} \rangle$ its linear span, and $\langle \mathcal{S} \rangle_a$ its affine span. For a scoring rule S, I use the short notation

$$\bar{S}_p(t) = \mathop{\mathsf{E}}_{\omega \sim p}[S(t, \omega)] \ .$$

$\mathcal{L}$ is the Lebesgue measure.

The proofs use the following elementary lemmas.

Lemma 3. *If Φ is a linear functional on $\mathcal{C}(\Omega)$ such that $\ker \Phi \cap \Delta\Omega \neq \emptyset$, then $\ker \Phi$ is the linear span of its intersection with $\Delta\Omega$.*

Proof. Let $f_0 \in \Delta\Omega$ with $\Phi(f_0) = 0$. Take any $f \in \ker \Phi$. As Ω is compact, $\inf f_0 > 0$ so that, if α is chosen large enough, $f + \alpha f_0 > 0$. So, defining $\beta = \int_\Omega (f + \alpha f_0) \mathrm{d}\mu$ and $f_1 = (f + \alpha f_0)/\beta$, we have that $\Phi(f_1) = 0$ and $f_1 \in \Delta\Omega$. Hence $f = \beta f_1 - \alpha f_0 \in \langle \ker \Phi \cap \Delta\Omega \rangle$. □

Lemma 4. *If $h : [a, b] \mapsto \mathbb{R}_+$ is a Lebesgue measurable function with $\int h > 0$, then there exists $\epsilon > 0$ such that $\{h \geq \epsilon\}$ has strictly positive measure.*

Proof. As $\{h > 0\}$ is the limit of the monotone increasing sequence of sets $\{h \geq 1/k\}$ and $\{h > 0\}$ has strictly positive measure, the sets $\{h \geq 1/k\}$ must have strictly positive measure as k grows large enough. □

Lemma 5. *If $h : [a, b] \mapsto \mathbb{R}_+$ is a Lebesgue measurable function that is strictly positive almost everywhere, and $A \subset [a, b]$ is a measurable set of strictly positive measure, then $\int_A h > 0$.*

Proof. As $A \cap \{h > 0\}$ is the limit of the monotone increasing sequence of sets $(A \cap \{h \geq 1/k\})$, for k large enough, the set $A \cap \{h \geq 1/k\}$ must have strictly positive measure, and $\int_A h \geq \lambda(A \cap \{h \geq 1/k\})/k > 0$. □

Lemma 6. *If Φ is a linear functional on $\mathcal{C}(\Omega)$ that is not continuous, then there exits a sequence $(f_n)_n$ of nonnegative functions of $\mathcal{C}(\Omega)$ converging to zero such that either ($\Phi(f_n) > 1$ for all n) or ($\Phi(f_n) < -1$ for all n).*

Proof. If, for all $\delta > 0$, there exists $\epsilon > 0$ such that for all nonnegative $f \in \mathcal{C}(\Omega)$, $\|f\| \leq \epsilon$ implies $|\Phi(f)| \leq \delta$, then Φ is continuous at 0.

Indeed, take any $\delta > 0$. The condition on Φ implies the existence of some $\epsilon > 0$, such that for all nonnegative $f \in \mathcal{C}(\Omega)$ with $\|f\| \leq \epsilon$, $|\Phi(f)| \leq \delta/2$. Take any $f \in \mathcal{C}(\Omega)$, with $\|f\| \leq \epsilon$. Then $f = f_+ - f_-$, where $f_+ = \max(f, 0)$ and $f_- = \max(-f, 0)$. Noting that $f_+, f_- \in \mathcal{C}(\Omega)$ and that $\|f_+\|, \|f_-\| \leq \|f\| \leq \epsilon$, we get that $|\Phi(f)| \leq |\Phi(f_+)| + |\Phi(f_-)| \leq \delta$. Hence f is continuous at 0, and, by linearity, continuous on $\mathcal{C}(\Omega)$.

Therefore, if Φ is not continuous, there must exist some $\delta > 0$ and a sequence $(f_n)_n$ of nonnegative elements of $\mathcal{C}(\Omega)$ converging to zero such that, for all n, $\Phi(f_n) < -\delta$ or $\Phi(f_n) > -\delta$. We get the lemma by rescaling the whole sequence and extracting a subsequence on which the same inequality holds. $\square$

Proof of Proposition 2

The proof is a simple adaptation of Proposition 3 of Nau (1985). Let Γ be the associated statistic function. Assume S is proper. Let θ_p, θ_q be two statistic values, and let p be a distribution such that $\Gamma(p) = \theta_p$. Consider the case $\theta_p < \theta_q$ and let r be a distribution such that $\theta_q \leq \Gamma(r)$. Consider the function $f : \lambda \mapsto \Gamma(\lambda r + (1-\lambda)p)$. Observe that the function is continuous. Noting that $f(0) = \theta_p < \theta_q \leq \Gamma(r) = f(1)$, there exists some $\lambda_q \in (0, 1]$ such that $f(\lambda_q) = \theta_q$. Let $q = \lambda_q r + (1-\lambda_q)p$.

As S is proper $S(\theta_p, q) \leq S(\theta_q, q)$. By linearity of the expectation operator, the inequality can be re-written as

$$\lambda_q S(\theta_p, r) + (1-\lambda_q) S(\theta_p, p) \leq \lambda_q S(\theta_q, r) + (1-\lambda_q) S(\theta_q, p)$$
$$\frac{1-\lambda_q}{\lambda_q}\left(S(\theta_p, p) - S(\theta_q, p)\right) \leq S(\theta_q, r) - S(\theta_p, r)\ .$$

The left-hand side of the inequality is nonnegative by properness of S, which makes the right-hand side of the inequality nonnegative as well. Hence S is order sensitive. The same procedure can be used to show that, if S is strictly proper, then S is also strictly order sensitive.

Proof of Theorem 5

Let (Θ, F) the a regular real-valued continuous statistic and Γ be the associated statistic function.

Part (1) $\Rightarrow$ (2) :

Let S be a strictly proper scoring rule. Take $p, q \in \Delta\Omega$, and $0 < \alpha < 1$. Suppose

$p, q \in F(\theta)$. Then, for all $\hat{\theta} \neq \theta$,

$$\mathop{\mathsf{E}}_{\omega \sim p}[S(\hat{\theta}, \omega)] \leq \mathop{\mathsf{E}}_{\omega \sim p}[S(\theta, \omega)] \ ,$$

and

$$\mathop{\mathsf{E}}_{\omega \sim q}[S(\hat{\theta}, \omega)] \leq \mathop{\mathsf{E}}_{\omega \sim q}[S(\theta, \omega)] \ ,$$

and so, by linearity of the expectation operator,

$$\begin{aligned} \mathop{\mathsf{E}}_{\omega \sim \alpha p + (1-\alpha) q}[S(\hat{\theta}, \omega)] &= \alpha \mathop{\mathsf{E}}_{\omega \sim p}[S(\hat{\theta}, \omega)] + (1-\alpha) \mathop{\mathsf{E}}_{\omega \sim q}[S(\hat{\theta}, \omega)] \\ &\leq \alpha \mathop{\mathsf{E}}_{\omega \sim p}[S(\theta, \omega)] + (1-\alpha) \mathop{\mathsf{E}}_{\omega \sim q}[S(\theta, \omega)] \\ &= \mathop{\mathsf{E}}_{\omega \sim \alpha p + (1-\alpha) q}[S(\theta, \omega)] \ , \end{aligned}$$

which, by strict properness, implies $\alpha p + (1-\alpha) q \in F(\theta)$. Hence the convexity of the sets $F(\theta)$.

Part (2) $\Rightarrow$ (3) :

First remark that, as Γ is continuous, the set of values taken by the statistic, Θ, is an interval of the real line. This can be seen by applying the intermediate value theorem to the continuous function $\alpha \mapsto \Gamma(\alpha p + (1-\alpha) q)$, defined on $[0, 1]$ for any $p, q \in \Delta\Omega$.

STEP 1. Let us start by showing that if, for all θ, $\{\Gamma = \theta\}$ is convex, then it is also the case that $\{\Gamma \geq \theta\}$, $\{\Gamma > \theta\}$, $\{\Gamma \leq \theta\}$ and $\{\Gamma < \theta\}$ are convex. I prove the first case, the other three work in a similar fashion.

Let $\theta \in \Theta^\circ$, and $p, q \in \Delta\Omega$, with $\Gamma(p) \geq \Gamma(q) \geq \theta$. Consider the function $f(\alpha) = \Gamma(\alpha p + (1-\alpha) q)$ defined on $[0, 1]$. Note that f is continuous. To prove that $\{\Gamma \geq \theta\}$ is convex, it suffices to show that the image of f is the interval $[\Gamma(q), \Gamma(p)]$. We already know that $[\Gamma(q), \Gamma(p)] \subseteq f([0, 1])$ by continuity of f, observing that $f(0) = \Gamma(q)$ and $f(1) = \Gamma(p)$. So let

$$\begin{aligned} a &= \sup\{\alpha \in [0, 1] \mid f(\alpha) = \Gamma(q)\} \ , \\ b &= \inf\{\alpha \in [0, 1] \mid f(\alpha) = \Gamma(p)\} \ . \end{aligned}$$

By continuity of f, the above two sets are closed and nonempty, so $f(a) = f(0) = \Gamma(q)$ and $f(b) = f(1) = \Gamma(p)$. Also, by convexity of the level sets of Γ, $f([0, a]) = \{\Gamma(q)\}$ and $f([b, 1]) = \{\Gamma(p)\}$. Besides, if, for some $\alpha^* > a$, $f(\alpha^*) < f(0)$ then by continuity $f(\alpha) = f(0)$ for some $\alpha > \alpha^*$, violating a's definition. Similarly, there does not exist α^* with $f(\alpha^*) > f(1)$, and $f([0, 1]) = [\Gamma(q), \Gamma(p)]$. So $\{\Gamma \geq \theta\}$ is convex.

<u>Step 2.</u> Let $\theta \in \Theta^\circ$. Let's start by showing the existence of a nonzero linear functional Φ on $\mathcal{C}(\Omega)$, continuous in the L^∞-norm topology, such that

$$\{\Gamma < \theta\} \subset \{\Phi \leq 0\} \; ,$$
$$\{\Gamma \geq \theta\} \subset \{\Phi \geq 0\} \; .$$

First note that, as Γ is continuous in the L^1-norm topology, it is also continuous in the richer, stronger L^∞-norm topology (recall that Ω was chosen to be compact). Using the L^∞-norm topology on $\mathcal{C}(\Omega)$ will be useful, because it makes the relative interior of $\Delta\Omega$ nonempty (by compactness of Ω) so as to be able to apply the Hahn-Banach separating hyperplane theorem.

Indeed, each $p \in \{\Gamma < \theta\}$ belongs to the (L^∞) relative interior of $\{\Gamma < \theta\}$, while each $p \in \{\Gamma > \theta\}$ belongs to the (L^∞) relative interior of $\{\Gamma \geq \theta\}$. By the previous step both $\{\Gamma < \theta\}$ and $\{\Gamma \geq \theta\}$ are convex, and since they are disjoint with nonempty (L^∞) relative interior, we can apply the Hahn-Banach separating hyperplane theorem and find a nonconstant affine function Φ on the affine span of $\Delta\Omega$, continuous in the L^∞-norm topology and such that $\Phi(\{\Gamma < \theta\}) \leq 0$ and $\Phi(\{\Gamma \geq \theta\}) \geq 0$.

Let us write Φ as

$$\Phi(f) = \Phi_0 + V(f - p_0)$$

for some $p_0 \in \Delta\Omega$, where $\Phi_0 = \Phi(p_0)$ and V is a linear functional on $\{f \in \mathcal{C}(\Omega) \mid \int_\Omega f \mathrm{d}\mu = 0\}$. V is continuous with respect to the L^∞ norm. We can extend Φ to a linear functional on the whole space $\mathcal{C}(\Omega)$, also continuous with respect to the L^∞ norm, by defining

$$\Phi(f) = \Phi_0 \int_\Omega f \mathrm{d}\mu + V\left(f - \left(\int_\Omega f \mathrm{d}\mu\right) p_0\right) \; .$$

<u>Step 3.</u> Now we show that Φ is also continuous with respect to the L^1 norm, *i.e.*, $\Phi \in (\mathcal{C}(\Omega))^*$. Suppose by contradiction it is not the case. Then, by Lemma 6, there exists a sequence $(\tilde{f}_n)_n$ of nonnegative functions of $\mathcal{C}(\Omega)$ that converge to zero, and such that, for example, $\Phi(\tilde{f}_n) < -1$ for all n (the case $\Phi(\tilde{f}_n) > 1$ is treated in a similar fashion).

Let $p_0 \in \{\Gamma > \theta\}$. Then, $\Phi(p_0) \geq 0$. If α is chosen large enough, for all n, we get

$$\Phi\left(\frac{p_0 + \alpha \tilde{f}_n}{\|p_0 + \alpha \tilde{f}_n\|}\right) = \frac{1}{\|p_0 + \alpha \tilde{f}_n\|}(\Phi(p_0) + \alpha \Phi(\tilde{f}_n)) < 0 \; , \tag{16}$$

noting that $\|p_0 + \alpha \tilde{f}_n\| > 0$ as p_0 is strictly positive and $\tilde{f}_n$ is nonnegative.

By continuity of Γ, there exists $\epsilon > 0$ such that, if $p \in \Delta\Omega$ with $\|p - p_0\| \leq \epsilon$, then

$|\Gamma(p) - \Gamma(p_0)| \le \frac{\Gamma(p_0)-\theta}{2}$. In particular, we get

$$\{p \in \Delta\Omega \mid \|p - p_0\| \le \epsilon\} \subset \{\Gamma > \theta\} \subset \{\Phi \ge 0\} \ . \tag{17}$$

Note that $p_0 + \alpha \tilde{f}_n > 0$ and so $\frac{p_0+\alpha\tilde{f}_n}{\|p_0+\alpha\tilde{f}_n\|} \in \Delta\Omega$. Besides,

$$\begin{aligned} \left\| \frac{p_0 + \alpha \tilde{f}_n}{\|p_0 + \alpha \tilde{f}_n\|} - p_0 \right\| &= \frac{1}{1 + \alpha\|\tilde{f}_n\|} \left\| \alpha \tilde{f}_n - \alpha \|\tilde{f}_n\| p_0 \right\| \ , \\ &\le \|\tilde{f}_n\| \alpha (1 + \|p_0\|) \ , \end{aligned}$$

and so, as $\|\tilde{f}_n\| \to 0$, there exists N such that $\|\tilde{f}_N\|\alpha(1 + \|p_0\|) \le \epsilon$. By (17), $\Phi((p_0 + \alpha\tilde{f}_n)/\|p_0 + \alpha\tilde{f}_n\|) \ge 0$, contradicting (16). Hence $\Phi \in (\mathcal{C}(\Omega))^*$.

STEP 4. Using the same θ as in the preceding step, as $\{\Gamma < \theta\}$ and $\{\Gamma > \theta\}$ are open sets of $\Delta\Omega$, we have that $\{\Gamma < \theta\} \subset \{\Phi < 0\}$ and $\{\Gamma > \theta\} \subset \{\Phi > 0\}$. In summary, we have shown the existence of a continuous linear functional Φ on $\mathcal{C}(\Omega)$ satisfying

$$\begin{aligned} \{\Gamma < \theta\} &\subset \{\Phi < 0\} \ , \\ \{\Gamma \ge \theta\} &\subset \{\Phi \ge 0\} \ , \\ \{\Gamma > \theta\} &\subset \{\Phi > 0\} \ . \end{aligned}$$

By a symmetric argument, there exists a continuous linear functional Ψ that satisfies

$$\begin{aligned} \{\Gamma < \theta\} &\subset \{\Psi < 0\} \ , \\ \{\Gamma \le \theta\} &\subset \{\Psi \le 0\} \ , \\ \{\Gamma > \theta\} &\subset \{\Psi > 0\} \ . \end{aligned}$$

We show that Φ and Ψ are positively collinear. If they are not collinear, then $\ker\Phi \cap \Delta\Omega \ne \ker\Psi \cap \Delta\Omega$ by Lemma 3. As $\ker\Phi \cap \Delta\Omega \subseteq \{\Gamma = \theta\}$ and $\ker\Psi \cap \Delta\Omega \subseteq \{\Gamma = \theta\}$, there exist $p, q \in \{\Gamma = \theta\}$ such that $\Phi(p) = 0$ with $\Psi(p) < 0$, and $\Phi(q) > 0$ with $\Psi(q) = 0$. So,

$$\Phi\left(\frac{p+q}{2}\right) > 0 \quad \text{and} \quad \Psi\left(\frac{p+q}{2}\right) < 0 \ .$$

By continuity of Φ and Ψ, there exists an open ball $\mathcal{B}$ centered on $(p+q)/2$, in the L^1 norm, such that $\Phi(\mathcal{B})$ contains only strictly positive values and $\Psi(\mathcal{B})$ contains only strictly negative values. Since $\mathcal{B} \cap \Delta\Omega \ne \emptyset$, these assertions imply that Γ is both greater than or equal to θ and less than or equal to θ on $\mathcal{B} \cap \Delta\Omega$, and so equals θ on this open set of $\Delta\Omega$. This contradicts the regularity assumption on Γ. So Ψ and Φ are collinear, and, by their sign properties above, are positively collinear, implying $\{\Gamma = \theta\} = \ker \Phi \cap \Delta\Omega$.

In conclusion, for all $\theta \in \Theta^\circ$, there exists $L_\theta \in (\mathcal{C}(\Omega))^*$ such that $\{\Gamma = \theta\} = \ker L_\theta \cap \Delta\Omega$.

Part (3) $\Rightarrow$ (1) :

Suppose that, for all $\theta \in \Theta^\circ$, there exists $\Phi_\theta \in (\mathcal{C}(\Omega))^*$ such that, for all $p \in \Delta\Omega$, $\Phi_\theta(p) = 0$ if and only if $\Gamma(p) = \theta$. Any nonzero factor of Φ_θ possesses the same property, so we can choose without loss $\|\Phi_\theta\| = 1$, and orient Φ_θ such that $\Phi_\theta(p) > 0$ for some given $p \in \{\Gamma > \theta\}$. By continuity of Γ and convexity of $\Delta\Omega$, Φ_θ has the following properties:

$$\begin{aligned}
\{\Gamma < \theta\} &= \{\Phi < 0\} \cap \Delta\Omega \ , \\
\{\Gamma = \theta\} &= \{\Phi = 0\} \cap \Delta\Omega \ , \\
\{\Gamma > \theta\} &= \{\Phi > 0\} \cap \Delta\Omega \ .
\end{aligned}$$

By the Banach extension theorem, Φ_θ can be extended to an element of $(L^1(\mu))^*$ with $\|\Phi_\theta\| = 1$. Applying a version of the Riesz Representation theorem (Theorem 1.11 of Megginson, 1998)), there exists a function $g_\theta \in L^\infty(\Omega)$ such that, for all $f \in L^1(\mu)$,

$$\Phi_\theta(f) = \int_\Omega f g_\theta \, \mathrm{d}\mu \ .$$

Note that Ω being compact, $L^\infty(\mu) \subset L^1(\mu)$.

<u>Step 1.</u> This steps shows that the function $\theta \mapsto g_\theta$ is uniformly continuous on every segment of Θ°, with respect to the L^1 norm.

Let us begin by showing that, for all $\theta_0 \in \Theta^\circ$, $\lim_{\theta \to \theta_0} \Phi_\theta(f) = 0$ whenever $f \in \ker \Phi_{\theta_0} \cap \Delta\Omega$. To see this, let $f \in \{\Gamma = \theta_0\}$, and, for any $\epsilon > 0$, consider the open ball $\mathcal{B}_\epsilon$ in $L^1(\mu)$ of radius ϵ that is centered on f. Note that Φ_{θ_0} takes both strictly positive and strictly negative values on $\mathcal{B}_\epsilon$, meaning that Γ takes values that are both above and below θ_0. By continuity of Γ, there exists some $\delta > 0$ such that $(\theta_0 - \delta, \theta_0 + \delta) \subset \Gamma(\mathcal{B} \cap \Delta\Omega)$. In particular, for all $\theta \in (\theta_0 - \delta, \theta_0 + \delta)$, there is $g \in \mathcal{B}_\epsilon \cap \Delta\Omega$ with $\Gamma(g) = \theta$, hence $|\Phi_\theta(f)| = |\Phi_\theta(f-g) + \Phi_\theta(g)| \leq \|\Phi_\theta\|\|f-g\| \leq \epsilon$. Therefore, we have that $\lim_{\theta \to \theta_0} \Phi_\theta(f) = 0$.

Observing that $\ker \Phi_{\theta_0|\mathcal{C}(\Omega)} = \langle \ker \Phi_{\theta_0} \cap \Delta\Omega \rangle$ by Lemma 3, the above limit remains valid whenever $f \in \ker \Phi_{\theta_0|\mathcal{C}(\Omega)}$. Now we can extend the limit to all members of $\mathcal{C}(\Omega)$. To do so, take some function $v \in \mathcal{C}(\Omega)$ such that $\Phi_{\theta_0}(v) = 1$. Any $f \in \mathcal{C}(\Omega)$ can be written $f = \Phi_{\theta_0}(f)v + w$ where $w \in \ker \Phi_{\theta_0|\mathcal{C}(\Omega)}$. Then $|\Phi_\theta(f) - \Phi_{\theta_0}(f)\Phi_\theta(v)| \to 0$ as $\theta \to \theta_0$. Besides, for all θ, $\|\Phi_\theta\| = 1$, hence $|\Phi_\theta(v)| \to 1$ as $\theta \to \theta_0$ and the orientation that was decided of Φ_θ yields $\Phi_\theta(v) \to 1$. So, $\Phi_\theta(f) \to \Phi_{\theta_0}(f)$ whenever $f \in \mathcal{C}(\Omega)$.

Finally, we extend the limit on the whole space $L^1(\mu)$. Let $f \in L^1(\mu)$ and fix $\epsilon > 0$. Choose $g \in \mathcal{C}(\Omega)$ such that $\|f - g\| < \epsilon/4$. Then there exists $\delta > 0$ such that if $|\theta - \theta_0| < \delta$,

$|\Phi_\theta(g) - \Phi_{\theta_0}(g)| < \epsilon/2$. Writing $\Phi_\theta(f) = \Phi_\theta(g) + \Phi_\theta(f-g)$, and similarly for $\Phi_{\theta_0}(f)$, we get $|\Phi_\theta(f) - \Phi_{\theta_0}(f)| \leq |\Phi_\theta(g) - \Phi_{\theta_0}(g)| + \|f-g\|(\|\Phi_\theta\| + \|\Phi_{\theta_0}\|) = \epsilon$. Consequently, Φ_θ converges pointwise to Φ_{θ_0} on $L^1(\mu)$.

That $\lim_{\theta\to\theta_0}\Phi_\theta(f) = \Phi_{\theta_0}(f)$ for all $f \in L^1(\mu)$ implies that the application $\theta \mapsto g_\theta$ is continuous with respect to the L^1 norm. By the Heine-Cantor theorem the application is uniformly continuous on every segment of Θ°.

<u>STEP 2.</u> Now let $[a,b] \subset \Theta^\circ$. We shall construct a bounded function $G : [a,b]\times\Omega \mapsto \mathbb{R}$ that is $(\mathcal{L}\otimes\mu)$-measurable and that offers a nearly perfect approximation of the g_θ's, in the sense that for almost every θ, $\|G(\theta,\cdot) - g_\theta\| = 0$.

We construct the function G as a limit of functions G_n. Let $d_{k,n} = a + k\frac{b-a}{2^n}$. Let τ_n be the function defined on $[a,b]$ by $\tau_n(\theta) = d_{k,n}$ when $\theta \in [d_{k,n}; d_{k+1,n})$. Define $G_n(\theta,\omega) = g_{\tau_n(\theta)}(\omega)$. Recall that $\theta \mapsto g_\theta$ is uniformly continuous on $[a,b]$ with respect to the L^1 norm. Let $\epsilon > 0$. Choose δ such that, for $\theta_1, \theta_2 \in [a,b]$,

$$|\theta_1 - \theta_2| < \delta \Rightarrow \|g_{\theta_1} - g_{\theta_2}\| < \frac{\epsilon}{2}\ .$$

Choose N such that $1/2^N < \delta$. Take any $n, m > N$.

$$\begin{aligned}|G_n(\theta,\omega) - G_m(\theta,\omega)| &= |g_{\tau_n(\theta)}(\omega) - g_{\tau_m(\theta)}(\omega)|\ ,\\ &\leq |g_{\tau_n(\theta)}(\omega) - g_\theta(\omega)| + |g_{\tau_m(\theta)}(\omega) - g_\theta(\omega)|\ ,\end{aligned}$$

but, as $n, m > N$, $|\tau_n(\theta) - \theta|, |\tau_m(\theta) - \theta| \leq 2^{-N} < \delta$. And so, for all $\theta \in [a,b]$,

$$\begin{aligned}\|G_n(\theta,\cdot) - G_m(\theta,\cdot)\| &\leq \|g_{\tau_n(\theta)} - g_\theta\| + \|g_{\tau_m(\theta)} - g_\theta\|\ ,\\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon\ ,\end{aligned}$$

which shows that $(G_n)_n$ is a Cauchy sequence in the Banach space $L^1(\mathcal{L}\otimes\mu)$, and thus converges to some function $G \in L^1(\mathcal{L}\otimes\mu)$. Besides, as each G_n is such that $|G_n| \leq 1$, we have that $|G| \leq 1$.

Moreover, for almost every θ, $\|G(\theta,\cdot) - g_\theta\| = 0$. This can be seen by writing

$$\|G(\theta,\cdot) - g_\theta\| \leq \|G(\theta,\cdot) - G_n(\theta,\cdot)\| + \|G_n(\theta,\cdot) - g_\theta\|\ ,$$

for all n. First we observe that

$$\|G_n(\theta,\cdot) - g_\theta\| = \|g_{\tau_n(\theta)} - g_\theta\| \xrightarrow[n\infty]{} 0\ .$$

Second, as a consequence of G_n converging to G and Fubini's theorem, the function $\theta \mapsto$

$\|G(\theta, \cdot) - G_n(\theta, \cdot)\|$ converges to zero in the L^1 norm. Since convergence in L^1 implies convergence almost everywhere, for almost every θ,

$$\|G(\theta, \cdot) - G_n(\theta, \cdot)\| \underset{n\infty}{\longrightarrow} 0 \ .$$

<u>STEP 3.</u> We then extend our function G to the whole interval Θ and construct a bounded function $H : \Theta \times \Omega \mapsto \mathbb{R}$ that is $(\mathcal{L} \otimes \mu)$-measurable and such that for almost every θ, $\|H(\theta, \cdot) - g_\theta\| = 0$.

Θ is an interval and so is an (at most) countable union of closed segments of the form $[a, b]$, whose interiors are pairwise disjoint. Let $\mathcal{F}$ be such a family. For any $[a, b] \in \mathcal{F}$, there exists, by the preceding step, a function $H_{[a,b]}$ integrable on $[a, b] \times \Omega$ and such that, for almost every θ, $\|H_{[a,b]}(\theta, \cdot) - g_\theta\|$. Besides, $|H_{[a,b]}| \leq 1$.

Define H as follows:

- if θ is an extremity of an interval of $\mathcal{F}$, let $H(\theta, \cdot) = 0$,
- otherwise, let $H(\theta, \omega) = H_{[a,b]}(\theta, \omega)$.

H inherits the nice properties of the $H_{[a,b]}$'s: $|H| \leq 1$, H is integrable on every segment of Θ, and, for almost every θ, $\|H(\theta, \cdot) - g_\theta\| = 0$. This last point implies that, whenever $p \in \Delta\Omega$, $\int_\Omega H(\theta, \cdot) p \mathrm{d}\mu$ equals $\Phi_\theta(p)$ for almost every θ.

<u>STEP 4.</u> At last we can construct a strictly proper scoring rule. Choose any $\theta_0 \in \Theta$ and let

$$S(\theta, \omega) = \int_{\theta_0}^{\theta} H(t, \omega) \mathrm{d}t \ .$$

As $|H|$ is bounded, for all $p \in \Delta\Omega$, a direct application of Fubini's theorem yields

$$\begin{aligned} \underset{\omega \sim p}{\mathsf{E}}[S(\theta, \omega)] &= \int_\Omega \left(\int_{\theta_0}^{\theta} H(t, \omega) \mathrm{d}t \right) p(\omega) \mathrm{d}\mu(\omega) \ , \\ &= \int_{\theta_0}^{\theta} \left(\int_\Omega H(t, \omega) p(\omega) \mathrm{d}\mu(\omega) \right) \mathrm{d}t \ . \end{aligned}$$

Suppose for example that $\Gamma(p) > \theta$, then

$$\begin{aligned} \underset{\omega \sim p}{\mathsf{E}}[S(\Gamma(p), \omega)] - \underset{\omega \sim p}{\mathsf{E}}[S(\theta, \omega)] &= \int_{\theta}^{\Gamma(p)} \left(\int_\Omega H(t, \omega) p(\omega) \mathrm{d}\mu(\omega) \right) \mathrm{d}t \ , \\ &> 0 \end{aligned}$$

since, for almost all $t < \Gamma(p)$, $\int_\Omega H(t, \cdot) p \mathrm{d}\mu = \Phi_t(p) > 0$. And similarly for $t > \Gamma(p)$. Hence S is strictly proper.

Proof of Theorem 6

If part:

In the proof of Theorem 5, we constructed a function H that is $(\mathcal{L} \otimes \mu)$-measurable, satisfies $|H| \leq 1$, and such that, for almost every θ, and all $p \in \Delta\Omega$,

$$\int_\Omega H(\theta, \omega) p(\omega) \mathrm{d}\mu(\omega)$$

is strictly positive when $\Gamma(p) > \theta$, strictly negative when $\Gamma(p) < \theta$, and zero when $\Gamma(p) = \theta$. Choose $S_0 = H$. Assume that for almost every ω and all θ, scoring rule S takes the form

$$S(\theta, \omega) = \kappa(\omega) + \int_{\theta_0}^{\theta} \xi(t) S_0(t, \omega) \mathrm{d}t \ ,$$

for some $\theta_0 \in \Theta$, $\kappa : \Omega \mapsto \mathbb{R}$, and $\xi : \mathcal{I} \mapsto \mathbb{R}_+$ a Lebesgue measurable bounded function.

For all $p \in \Delta\Omega$,

$$\bar{S}_p(\theta) = \int_\Omega \kappa \mathrm{d}\mu + \int_\Omega \int_{\theta_0}^{\theta} \xi(t) S_0(t, \omega) \mathrm{d}t \mathrm{d}\mu(\omega) \ .$$

Take for example $\theta < \Gamma(p)$. By Fubini's theorem,

$$\bar{S}_p(\Gamma(p)) - \bar{S}_p(\theta) = \int_\theta^{\Gamma(p)} \xi(t) \left(\int_\Omega S_0(t, \omega) p(\omega) \mathrm{d}\mu(\omega) \right) \mathrm{d}t \ .$$

As, for almost all $t < \Gamma(p)$, $\int_\Omega S_0(t, \omega) p(\omega) \mathrm{d}\mu(\omega) > 0$, we get $\bar{S}_p(\Gamma(p)) - \bar{S}_p(\theta) \geq 0$, implying that S is proper. If, in addition, $\int_\theta^{\Gamma(p)} \xi > 0$, then by Lemma 4, there is $\epsilon > 0$ such that $A = \{\xi \geq \epsilon\}$ is of strictly positive Lebesgue measure. Hence,

$$\bar{S}_p(\Gamma(p)) - \bar{S}_p(\theta) \geq \epsilon \int_A \left(\int_\Omega S_0(t, \omega) p(\omega) \mathrm{d}\mu(\omega) \right) \mathrm{d}t$$

which is strictly positive by Lemma 5, making S strictly proper.

Only if part:

Let S be a regular scoring rule for Γ, and $\theta_0 \in \Theta$. If S is (strictly) proper, $(\theta, \omega) \mapsto S(\theta, \omega) - S(\theta_0, \omega)$ is also (strictly) proper. Thus we can assume with loss of generality that $S(\theta_0, \cdot) = 0$.

As $S(\cdot, \omega)$ is Lipschitz continuous, it is also absolutely continuous and there is a function $G : \Theta \times \Omega \mapsto \mathbb{R}$ such that, for all θ, ω,

$$S(\theta, \omega) = \int_{\theta_0}^{\theta} G(t, \omega) \mathrm{d}t \ .$$

Moreover, for all ω, $\theta \mapsto S(\theta, \omega)$ is differentiable except possibly on a measure zero set (that may depend on ω), and

$$\frac{S(\theta, \omega)}{\partial \theta} = G(\theta, \omega) \ .$$

G can be chosen such that, if $S(\cdot, \omega)$ is not differentiable at θ, $G(\theta, \omega) = 0$. Besides, G is $(\mathcal{L} \otimes \mu)$-measurable as a limit of the measurable functions

$$(\theta, \omega) \mapsto \begin{cases} n(S(\theta + 1/n, \omega) - S(\theta, \omega)) & \text{if } S(\cdot, \omega) \text{ is differentiable at } \theta \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

Finally, as S is Lipschitz continuous, G is bounded.

For all $\theta \in \Theta$, define Ψ_θ, a continuous functional on $L^1(\mu)$, as

$$\Psi_\theta(f) = \int_\Omega G(\theta, \omega) f(\omega) \mathrm{d}\mu(\omega)$$

<u>Step 1.</u> We start by proving the existence of a set $\mathcal{Z}$ of Lebesgue measure zero such that, whenever $\theta \notin \mathcal{Z}$, $\bar{S}_p$ is differentiable at θ for all $p \in \Delta\Omega$, and

$$\bar{S}_p'(\theta) = \Psi_\theta(p) \ .$$

Observing that, for all $\theta > \theta_0$, $[\theta_0, \theta] \times \Omega$ is of finite $(\mathcal{L} \otimes \mu)$-measure, define G_f as

$$G_f(\theta) = \int_{[\theta_0, \theta] \times \Omega} G(t, \omega) f(\omega) \mathrm{d}\mathcal{L} \otimes \mu(t, \omega) \ ,$$

for every $f \in L^1(\mu)$. By application of Fubini's theorem,

$$G_f(\theta) = \int_{\theta_0}^{\theta} \Psi_t(f) \mathrm{d}t \ .$$

Consequently G_f can be written as a Lebesgue integral, implying that given some fixed $f \in L^1(\mu)$, G_f is differentiable almost everywhere, and $G_f'(\theta) = \Psi_\theta(f)$.

As $L^1(\mu)$ is separable, there exists a countable set $\mathcal{F}$ dense in $L^1(\mu)$. A countable union of measure zero sets remains of measure zero, so there exists a set $\mathcal{Z}$ of Lebesgue measure zero such that, for all $f \in \mathcal{F}$, and all $\theta \notin \mathcal{Z}$, G_f is differentiable at θ and $G_f'(\theta) = \Psi_\theta(f)$. Choose $\mathcal{Z}$ so as to include the extremities of interval Θ.

We can generalize to all $f \in L^1(\mu)$. Let $f \in L^1(\mu), T \notin \mathcal{Z}$, and $K > 0$ an upper bound of $|G|$. Let $\epsilon > 0$. We want to find a $\delta > 0$ such that, if t satisfies $|T - t| \leq \delta$, then

$$\left| \frac{G_f(T) - G_f(t)}{T - t} - \Psi_t(f) \right| \leq \epsilon \ ,$$

or equivalently,

$$\left| \frac{1}{T - t} \int_t^T \int_\Omega G(r, \omega) f(\omega) \mathrm{d}\mu(\omega) \mathrm{d}r - \int_\Omega G(T, \omega) f(\omega) \mathrm{d}\mu(\omega) \right| \leq \epsilon \ . \tag{18}$$

Let $\tilde{f} \in \mathcal{F}$ such that $\|f - \tilde{f}\| \leq \frac{\epsilon}{3K}$. Then,

$$\left| \frac{1}{T - t} \int_t^T \int_\Omega G(r, \omega) [f(\omega) - \tilde{f}(\omega)] \mathrm{d}\mu(\omega) \mathrm{d}r \right| \leq \frac{\epsilon}{3} \ .$$

Besides, $G_{\tilde{f}}$ is differentiable at T and so, for some $\delta > 0$, $|T - t| \leq \delta$ implies

$$\left| \frac{1}{T - t} \int_t^T \int_\Omega G(r, \omega) \tilde{f}(\omega) \mathrm{d}\mu(\omega) \mathrm{d}r - \int_\Omega G(t, \omega) \tilde{f}(\omega) \mathrm{d}\mu(\omega) \right| \leq \frac{\epsilon}{3} \ .$$

Finally,

$$\left| \int_\Omega G(t, \omega) [f(\omega) - \tilde{f}(\omega)] \mathrm{d}\mu(\omega) \right| \leq \frac{\epsilon}{3} \ .$$

Summing the three inequalities above yields (18). In particular, for all $\theta \notin \mathcal{Z}$, $\bar{S}_p$ is differentiable at θ for all $p \in \Delta\Omega$, and

$$\bar{S}_p'(\theta) = \Psi_\theta(p) \ .$$

<u>Step 2.</u> Assume S is proper, and let $\theta \notin \mathcal{Z}$. If $p \in \{\Gamma = \theta\}$, $\Gamma(p) \notin \mathcal{Z}$ and so $\bar{S}_p(\Gamma(p))' = 0$, which yields $\{\Gamma = \theta\} \subset \ker \Psi_\theta$. By Theorem 5, there exists a continuous linear functional Φ_θ on $\mathcal{C}(\Omega)$ such that $\{\Gamma = \theta\} = \ker \Phi_\theta \cap \Delta\Omega$. As $\{\Gamma = \theta\}$ is nonempty, applying Lemma 3 yields $\ker \Phi_\theta = \langle \{\Gamma = \theta\} \rangle$ and, as $\{\Gamma = \theta\} \subset \ker \Psi_{\theta|\mathcal{C}(\Omega)}$ we have that $\ker \Phi_\theta \subseteq \ker \Psi_{\theta|\mathcal{C}(\Omega)}$. Consequently there exists a real number $\xi(\theta)$ such that $\Psi_{\theta|\mathcal{C}(\Omega)} = \xi(\theta)\Phi_\theta$ (Lemma 3.1 of Megginson, 1998). If $\Phi_\theta = 0$ choose $\xi(\theta) = 0$. By the Banach extension theorem, Φ_θ can be extended by continuity to the whole space $L^1(\mu)$, and by density of $\mathcal{C}(\Omega)$ in $L^1(\mu)$, we have that $\Psi_\theta = \xi(\theta)\Phi_\theta$ on $L^1(\mu)$. Let $\xi(\theta) = 0$ for all $\theta \in \mathcal{Z}$.

We can choose without loss $\|\Phi_\theta\| = 1$. In the proof of Theorem 5, we showed that Φ_θ can be chosen such that $\theta \mapsto \Phi_\theta(p)$ be Lebesgue measurable for all $p \in \Delta\Omega$. Since, from its definition, $\theta \mapsto \Psi_\theta(p)$ is Lebesgue measurable, writing $\xi(\theta) = \Psi_\theta / \Phi_\theta$ leads to Lebesgue measurability of ξ. Besides, noting that $\|G(\theta, \cdot)\|_\infty = \|\Psi_\theta\| = |\xi(\theta)| \|\Phi_\theta\| = |\xi(\theta)|$, boundedness of ξ follows from boundedness of G.

Therefore, for all $p \in \Delta\Omega$, and all θ,

$$\bar{S}_p(\theta) = \int_{\theta_0}^{\theta} \Psi_t(p)\mathrm{d}t = \int_{\theta_0}^{\theta} \xi(t)\Phi_t(p)\mathrm{d}t \ .$$

By Proposition 2, S is order sensitive. This implies $\xi \geq 0$. Indeed, suppose $\xi(\theta) < 0$ for some $\theta \notin \mathcal{Z}$. Take, for example, $p \in \{\Gamma > \theta\}$. Then,

$$\bar{S}_p'(\theta) = \xi(\theta)\Phi_\theta(p) < 0 \ ,$$

and $\bar{S}_p$ is not (weakly) increasing on $\{t < \Gamma(p)\}$, contradicting order sensitivity of S. Hence $\xi \geq 0$. Assume that, in addition, S is strictly proper. Take any $\theta_1 < \theta_2$ and $p \in \{\Gamma = \theta_2\}$. Then,

$$\begin{aligned} 0 < |\bar{S}_p(\theta_2) - \bar{S}_p(\theta_1)| &= \left| \int_{\theta_1}^{\theta_2} \int_\Omega \xi(t)\Phi_t(p)\mathrm{d}t \right| \ , \\ &\leq \|f\| \int_{\theta_1}^{\theta_2} \xi \ , \end{aligned}$$

implying $\int_{\theta_1}^{\theta_2} \xi > 0$.

References

N.I. Al-Najjar and J. Weinstein. Comparative testing of experts. *Econometrica*, 76 (3):541–559, 2008.

N.I. Al-Najjar, A. Sandroni, R. Smorodinsky, and J. Weinstein. Testing theories with learnable and predictive representations. *Journal of Economic Theory*, 145 (6):2203–2217, 2010.

F. Aurenhammer. Power diagrams: Properties, algorithms and applications. *SIAM Journal on Computing*, 16(1):78, 1987.

F. Aurenhammer. Voronoi diagrams: A survey of a fundamental geometric data structure. *ACM Computing Surveys*, 23(3):345–405, 1991.

J.P. Bonin. On the design of managerial incentive structures in a decentralized planning environment. *American Economic Review*, 66(4):682–687, 1976.

G.W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.

R.T. Clemen. Incentive contracts and strictly proper scoring rules. *TEST*, 11(1): 167–189, 2002.

A.P. Dawid. The geometry of proper scoring rules. *Annals of the Institute of Statistical Mathematics*, 59(1):77–93, 2007. ISSN 0020-3157.

M. De Berg, O. Cheong, and M. van Kreveld. *Computational geometry: Algorithms and applications*. Springer, 2008.

B. De Finetti. Does it make sense to speak of "good probability appraisers"? *The Scientist Speculates: An Anthology of Partly-Baked Ideas*, pages 257–364, 1962.

E. Dekel and Y. Feinberg. Non-Bayesian testing of a stochastic prediction. *Review of Economic Studies*, 73(4):893–906, 2006.

I.I. Eremin. *Theory of linear optimization*. VSP International Science Publishers, 2002.

L.-S. Fan. On the reward system. *American Economic Review*, 65(4):226–229, 1975.

Y. Feinberg and N.S. Lambert. Mostly calibrated. *Working paper*, 2011.

Y. Feinberg and C. Stewart. Testing multiple forecasters. *Econometrica*, 76(3):561–582, 2008.

L. Fortnow and R. Vohra. The complexity of forecast testing. *Econometrica*, 77(1): 93–105, 2009.

D.P. Foster and R.V. Vohra. Asymptotic calibration. *Biometrika*, 85(2):379–390, 1998.

D. Friedman. Effective scoring rules for probabilistic forecasts. *Management Science*, 29(4):447–454, 1983.

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

IJ Good. *Rational decisions*. Springer-Verlag, 1997.

R. Hanson. Combinatorial information market design. *Information Systems Frontiers*, 5(1):107–119, 2003.

A.D. Hendrickson and R.J. Buehler. Proper scores for probability forecasters. *Annals of Mathematical Statistics*, 42(6):1916–1921, 1971.

H. Imai, M. Iri, and K. Murota. Voronoi diagram in the Laguerre geometry and its applications. *SIAM Journal on Computing*, 14(1):93, 1985.

D.J. Johnstone. The parimutuel Kelly probability scoring rule. *Decision Analysis*, 4 (2):66–75, 2007.

E. Karni. A mechanism for eliciting probabilities. *Econometrica*, 77(2):603–606, 2009.

N.S. Lambert, J. Langford, J. Wortman, Y. Chen, D. Reeves, Y. Shoham, and D. Pennock. An axiomatic characterization of wagering mechanisms. *Working paper*, 2011.

J. McCarthy. Measures of the value of information. *Proceedings of the National Academy of Sciences of the United States of America*, 42(9):654–655, 1956.

R.E. Megginson. *An introduction to Banach space theory*. Springer Verlag, 1998.

R.F. Nau. Should scoring rules be 'effective'? *Management Science*, 31(5):527–535, 1985.

T. Offerman, J. Sonnemans, G. Van De Kuilen, and P.P. Wakker. A truth serum for non-Bayesians: Correcting proper scoring rules for risk attitudes. *Review of Economic Studies*, 76(4):1461–1489, 2009. ISSN 1467-937X.

W. Olszewski and A. Sandroni. Contracts and uncertainty. *Theoretical Economics*, 2 (1):1–13, 2007.

W. Olszewski and A. Sandroni. Manipulability of future-independent tests. *Econometrica*, 76(6):1437–1466, 2008.

W. Olszewski and A. Sandroni. A non-manipulable test. *Annals of Statistics*, 37(2): 1013–1039, 2009a.

W. Olszewski and A. Sandroni. Strategic manipulation of empirical tests. *Mathematics of Operations Resesearch*, 34(1):57–70, 2009b.

W. Olszewski and A. Sandroni. Falsifiability. *American Economic Review*, 101(2): 788–818, 2011.

K. Osband. Optimal forecasting incentives. *Journal of Political Economy*, 97(5): 1091–1112, 1989.

K. Osband and S. Reichelstein. Information-eliciting compensation schemes. *Journal of Public Economics*, 27(1):107–15, 1985.

M. Ostrovsky. Information aggregation in dynamic markets with strategic traders. *Working paper*, 2011.

S. Reichelstein and K. Osband. Incentives in government contracts. *Journal of Public Economics*, 24(2):257–270, 1984.

L.J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

M.J. Schervish. A general method for comparing probability assessors. *The Annals of Statistics*, 17(4):1856–1879, 1989.

R. Selten. Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, 1(1):43–62, 1998. ISSN 1386-4157.

E. Shmaya. Many inspections are manipulable. *Theoretical Economics*, 3(3):367–382, 2008.

C. Stewart. Non-manipulable Bayesian testing. *Journal of Economic Theory*, 146(5): 2029–2041, 2011.

W. Thomson. Eliciting production possibilities from a well informed manager. *Journal of Economic Theory*, 20(3):360–380, 1979.

R.L. Winkler. Scoring rules and the evaluation of probability assessors. *Journal of the American Statistical Association*, 64(327):1073–1078, 1969.

R.L. Winkler and A.H. Murphy. “good” probability assessors. *Journal of Applied Meteorology*, 7(5):751–758, 1968.

R.L. Winkler, J. Muñoz, J.L. Cervera, J.M. Bernardo, G. Blattenberger, J.B. Kadane, D.V. Lindley, A.H. Murphy, R.M. Oliver, and D. Ríos-Insua. Scoring rules and the evaluation of probabilities. *TEST*, 5(1):1–60, 1996.

J. Wolfers and E. Zitzewitz. Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126, 2004.