The Query Complexity of Scoring Rules

Pablo Azar

MIT, CSAIL

Cambridge, MA 02139, USA

azar@csail.mit.edu

Silvio Micali

MIT, CSAIL

Cambridge, MA 02139, USA

silvio@csail.mit.edu

December 19, 2012

Abstract

Proper scoring rules are crucial tools to elicit truthful information from experts. A scoring rule maps $\mathcal{X}$, an expert-provided distribution over the set of all possible states of the world, and ω, a realized state of the world, to a real number representing the expert's reward for his provided information. To compute this reward, a scoring rule queries the distribution $\mathcal{X}$ at various states. The number of these queries is thus a natural measure of the *complexity* of the scoring rule.

We prove that any bounded and strictly proper scoring rule that is deterministic must make a number of queries to its input distribution that is a quarter of the number of states of the world. When the state space is very large, this makes the computation of such scoring rules impractical. We also show a new stochastic scoring rule that is bounded, strictly proper, and which only makes two queries to its input distribution. Thus, using randomness allows us to have significant savings when computing scoring rules.

1   Introduction

Motivating examples   Very often we have uncertainty about the world. This uncertainty is typically modeled via a distribution $\mathcal{D}$, over a set Ω of possible states. A canonical example is predicting the weather, where $\Omega = \{$Sunny, Rainy$\}$ and $\mathcal{D}$ can be specified via two numbers: the probability $\mathcal{D}(Rainy)$ and its complementary probability $\mathcal{D}(Sunny)$. A more complex example is an election with an electoral college, such as the US system. In such an election, two parties compete to win majorities in each of n provinces.[1] In this example, a state of the world is a string in $\{0,1\}^n$, with the ith bit representing which party won the ith province. A probability distribution is now a much larger object, having to encode $2^n - 1$ real numbers. Another example with a potentially larger outcome space is when we want

---

[1]We use the term province and not state so as to not cause confusion with the term "state of the world".

a prediction about the outcome of a race with n contestants. In this setting, a state of the world is an ordering of the n contestants, and thus the cardinality of Ω is $n!$.

Buying a Distribution from an Expert   In many settings of interest, a decision maker must acquire information about a distribution $\mathcal{D}$ from a knowledgable expert. (For example, a farmer may hire a weather forecaster in order to decide which crops to plant for next year: e.g., wheat if its going to be a dry year and rice if its going to be a rainy year.) However, the expert is not trusted, and could report an arbitrary distribution $\mathcal{X}$ instead of the true distribution $\mathcal{D}$. To guarantee that the expert is truthful, the decision maker needs to give her some type of reward that depends not only on the expert's prediction $\mathcal{X}$, but also on his eventual observation of the real state of the world ω.[2] More formally, after agreeing on a payment rule—technically a *scoring rule*—S, the expert and decision maker participate in the following interaction:

1. the expert specifies a distribution $\mathcal{X}$; then
2. the decision maker observes a sample ω drawn according to $\mathcal{D}$; and finally
3. the decision maker pays to the expert the reward $S(\mathcal{X}, \omega)$.

To make the above process concrete, we need to specify how to communicate a distribution $\mathcal{X}$ and how to choose a scoring rule S.

1.1   Specifying a Distribution

We identify three possible ways for an expert to pin down a distribution $\mathcal{X}$.

Exhaustive List   Fixing an order of the possible states of the world, one way to specify $\mathcal{X}$ consists of the list $\mathcal{X}(\omega_1), \mathcal{X}(\omega_2), \ldots, \mathcal{X}(\omega_{|\Omega|})$.

Such a way, however is feasible only when $|\Omega|$, the cardinality of Ω, is sufficiently small.

Algorithmic Description   A potentially compact way of specifying $\mathcal{X}$ consists of an algorithm $\mathcal{A}_\mathcal{X}$ that, on input (a string encoding) a state $\omega \in \Omega$, outputs the probability $\mathcal{X}(\omega)$.

For example, when $\mathcal{X}$ is given by a bayesian network[3] the expert needs to encode and send relatively few parameters. From these the decision maker can easily compute $\mathcal{X}(\omega)$ for any state ω he cares about. Another clear example is the geometric distribution.[4]

---

[2]It is easy to see that a reward depending solely on the reported distribution $\mathcal{X}$ could not ensure the expert's truthfulness. Consider our weather example, where $\Omega = \{sunny, rainy\}$ and the true probability of $sunny$ is $\mathcal{D}$, where —say— $D(sunny) = 35\%$. Then, if a farmer asks a forecaster to reveal the true distribution over Ω, and the reward he promises to pay is maximized by another distribution $\mathcal{X}$, where— say—$\mathcal{X}(sunny) = 50\%$, then a rational forecaster will announce $\mathcal{X}$ instead of $\mathcal{D}$.

[3]A bayesian network is given by a directed acyclic graph $G = (V, E)$, a collection of random variables $\{X_v\}_v$ indexed by the vertices of G, and conditional probability tables $Pr[X_v|X_{pa(v)}]$ that give the conditional probability of X_v taking different values given the values of its parents in the network.

[4]Consider a coin that lands "heads" with probability p, and let $\mathcal{X}$ be the random variable representing the number of coin flips that we need until observing one head. Even though there is an infinite number of

Merkle Commitment A *Merkle tree* [11] is a universal way for a party A to "commit" to any finite number of values, $v_1, \ldots, v_n$, to another party B by sending B a single k-bit value c (e.g., $k = 500$), that is very quickly computed from the values $v_1, \ldots, v_n$. *De facto*, this short string c pins down all original values and allows the expert to subsequently reveal to B what the original i^{th} value was without cheating. He does so by sending v_i to B together with a $(k \cdot \log n)$-bit authenticating string a_i. Party B then runs a very fast algorithm on inputs c, v_i and a_i, to output either YES or NO. If the output is YES, then B knows that v_i is the original i^{th} value. Indeed, although an "authenticating" string a_i' for a different value v_i' exists, the number of computational steps that B would have to perform in order to find such a_i' is greater than —say— $2^{k/3}$. (Merkle trees have been constructed from the intractability of several problems, such as integer factorization.)

In theoretical computer science, Merkle trees have been used in order to generate more efficient zero-knowledge proofs [7] and computationally sound proofs [12].

We can also use Merkle trees to communicate very large distributions. Consider our example of an election with an electoral college. In this example, the set of states of the world is $\Omega = \{0, 1\}^{50}$, and a distribution $\mathcal{X}$ consists of a list of $2^{50} - 1$ values. Although sending this list to the decision maker would be (barely) possible for the expert, sending just a 500-bit Merkle commitment would be much easier. The decision maker could then ask the expert for $\mathcal{X}(\omega)$ for any 50-bit state ω whenever he wants and without fear of being cheated. The authenticating string a_ω would just consist of 25,000 bits. (Since the expert has committed to the distribution in advance, the decision maker can even use the expert herself in order to query $\mathcal{X}$ for the purpose of determining her reward!)

1.2 Crucial Properties of Scoring Rules

We identify three properties of a good scoring rule.

Truthfulness Since our goal is to incentivize the expert to reveal correctly the true distribution D over the states of the world, a good scoring rule must guarantee that the expert maximizes her reward when she is truthful. More formally, a scoring rule S is *strictly proper* if for all distinct distributions $\mathcal{D}$ and $\mathcal{X}$:

$$\sum_{\omega \in \Omega} \mathcal{D}(\omega) S(\mathcal{D}, \omega) > \sum_{\omega \in \Omega} \mathcal{D}(\omega) S(\mathcal{X}, \omega).$$

Two popular strictly proper scoring rules are the logarithmic scoring rule [6] $S(\mathcal{X}, \omega) = \log(\mathcal{X}(\omega))$ and the quadratic scoring rule [2] $S(\mathcal{X}, \omega) = 2\mathcal{X}(\omega) - \sum_{\omega \in \Omega} \mathcal{X}(\omega)^2 + 1$.

Examples of other scoring rules, together with a comprehensive survey and characterization, can be found in a paper by Gneiting and Raftery [5].

For brevity and when the context is clear, we will often refer to strictly proper scoring rules simply as *proper scoring rules* or *scoring rules*.

---

states of the world, the algorithm for computing $\mathcal{X}(k) = Pr[\mathcal{X} = k] = (1 - p)^{k-1} \cdot p$ is very efficient, and very easy to communicate.

Ex-Post Individual Rationality Even when the chosen scoring rule S is strictly proper, it may be the case that the expert may be reluctant to sell her knowledge of the true distribution $\mathcal{D}$ in the above envisaged process, because S gives her negative rewards in some states of the world. This is very clear with the logarithmic scoring rule, where the expert may have to make arbitrarily large payments to the decision maker instead of being paid herself. One may think that this problem can be avoided by shifting the scoring rule by some positive amount, but this is not the case.[5] To really avoid this problem, we need scoring rules that are *ex-post individually rational*. That is, $S(\mathcal{X}, \omega) > 0$ for every possible input distribution $\mathcal{X}$ and every possible realized state of the world ω. The quadratic scoring rule is a good example of an individually rational scoring rule, because its reward is bounded between $[0, 2]$ regardless of $\mathcal{X}$ and ω.

Query Efficiency Both the quadratic and logarithmic scoring rules are evaluated by querying the distribution $\mathcal{X}$ at different states of the world $\omega_1, ..., \omega_k \in \Omega$, and then returning as a reward some function of $\mathcal{X}(\omega_1), ..., \mathcal{X}(\omega_k)$. If the cardinality of Ω is small, then querying $\mathcal{X}$ is not a problem. In this case in fact the expert may send the decision maker the distribution $\mathcal{X}$ encoded as an exhaustive list, and querying $\mathcal{X}$ on *all* states of the world would only increase complexity by a constant factor. However, if, as in our election and ranking examples, Ω is very large, communicating the distribution may be easy, but computing a scoring rule making many queries to $\mathcal{X}$ becomes prohibitively expensive. Thus, a third and crucial property of scoring rules is that they query $\mathcal{X}$ only at very few inputs. The ideal case would be to have a *local* scoring rule, that is, a scoring rule that only queries $\mathcal{X}$ on the realized state of the world ω. The logarithmic scoring rule is an example of a local scoring rule. In contrast, the quadratic scoring rule depends on the norm $\sum_{\omega \in \Omega} \mathcal{X}(\omega)^2$, and thus needs to make $|\Omega|$ queries to the distribution. That is, it must query $\mathcal{X}$ on each possible state of the world.

1.3   Our Results

Note that the logarithmic scoring rules is truthful and makes only one query, but is not ex-post individually rational. The quadratic scoring rule is truthful and ex-post individually rational, but is not query efficient. The ideal situation would be having a scoring rule that is truthful, individually rational and local. Unfortunately, the following well known lemma implies that this ideal situation is impossible.[6]

---

[5]That is, because the linear transformation $a \cdot \ln(\mathcal{X}, \omega) + b$ is also truthful and local, one may attempt to ensure the individual rationality of the logarithmic scoring rule by choosing b very large, so as to offsett the negative reward that the expert would obtain in most states of the world. However, when $\mathcal{X}$ can assign probability zero to some states of the world, this approach cannot yield an individually rational scoring rule. Even restricting ourselves to full-support distributions, for any linear transformation $a \cdot \ln(\mathcal{X}, \omega) + b$ of the logarithmic scoring rule, there exists a distribution $\mathcal{X}$ such that this transformed scoring rule still gives a negative reward.

[6]For completeness, we give a proof sketch of the lemma in Appendix A.

Lemma 1 (McCarthy - Savage [10] [14]). *If $|\Omega| \geq 3$ then the only local strictly proper scoring rules are of the form $S(\mathcal{X}, \omega) = a + b \ln \mathcal{X}(\omega)$ where $b > 0$.*

Since local scoring rules only make a single query, the McCarthy-Savage lemma does not exclude the existence of truthful and individually rational scoring rules that make a constant number of queries. However, we prove that is far from begin the case for any *deterministic* scoring rule.

More precisely, our first theorem shows that any truthful and individually rational scoring rule must make at least $\frac{|\Omega|}{4}$ queries to its input distribution.

On the positive side, our second theorem shows that, if we allow *stochastic* scoring rules, then we can construct new scoring rules that are truthful, ex-post individually rational, and query efficient.

2 Prior Work

2.1 Generalizations of Locality

Parry, Dawid and Lauritzen [13] study proper scoring rules for the real line. Thus, they focus on continuous distributions given by probability mass functions $f : \mathbb{R} \to [0, 1]$. They study a generalization of locality which they call k-locality. In their definition, a scoring rule $S(f, x)$ is k-local if and only if we can write $S(f, x) = g(x, f'(x), f''(x), ..., f^{(k)}(x))$, where $f^{(k)}(x)$ is the kth derivative of f evaluated at x. They present k-local scoring rules for all even k, and show that one does not need the distribution to be normalized ($\int f(x) dx = 1$) in order for these scoring rules to be effective. Ehm and Gneiting [4] characterize the 2-local scoring rules which depend only on the density function at x and its first two derivatives.

Our work differs from the above papers in two ways. First, our notion of k-query scoring rules is a different generalization of locality than Parry, Dawid and Lauritzen's. Instead of studying scoring rules which depend on $\mathcal{X}(\omega)$ and its derivatives, we study scoring rules that depend on $\mathcal{X}$ when evaluated at ω and at $k - 1$ other points in the state space. Second, we focus on discrete distributions, whereas they focus on continuous ones.

A more closely related paper, by Dawid, Lauritzen and Parry [3], studies scoring rules for discrete spaces and interprets the sample space Ω as a graph $\mathcal{G}$. They define a scoring rule to be $\mathcal{G}$-local if $S(\mathcal{X}, \omega)$ is a function only of the neighbors of ω in the graph $\mathcal{G}$. Our proof of our theorem is motivated by their interpretation of Ω as a graph. These results are useful for developing good heuristics for special cases of the distribution $\mathcal{D}$. For instance, example 4.4 in their paper gives a deterministic scoring rule over $\{0, 1\}^n$ which is individually rational and makes only $2n$ queries to its input distribution. However, this rule is only strictly proper when $\mathcal{D}$ gives positive weight to every state of the world. One interesting question is whether there exists a comparable rule that only makes $2n$ queries to the input distribution but is strictly proper *for all inputs*.

Our first theorem shows that a search for such a scoring rule would be fruitless. For arbitrary distributions over $\Omega = \{0, 1\}^n$, a deterministic, individually rational and strictly proper scoring rule always needs to make $\frac{|\Omega|}{4} = \frac{2^n}{4}$ queries to its input distribution. Even for

moderately sized n, this is a very large number of queries required to evaluate the scoring rule.

2.2 Cumulative Scoring Rules

Our impossibility theorem only applies to deterministic scoring rules and raises the question of whether using randomization can help us obtain new rules satisfying our three desired properties. By changing the definition of a scoring rule to accept a cumulative distribution function (CDF) instead of a probability mass function (PMF), Matheson and Winkler [9] provided a positive answer. However, the CDF may not be easily accessible. For example, for bayesian models, the expert may easily learn and communicate the PMF of the true distribution, but not necessarily its CDF. Furthermore, the decision maker and expert need to agree on which order of the state space to use to compute the CDF. This could be a problem. Consider the case when the expert (before being hired) already determined the cumulative distribution function of the true distribution for some arbitrary order of states of the world. Then, either communicating this order to the decision maker, or recomputing the cumulative distribution function for a new agreed upon order, may be very expensive.

In contrast, our second theorem gives a new truthful, individually rational and stochastic scoring rule with low query complexity that, as usual, takes the PMF of the distribution as input. This illustrates that what helps us reduce the query complexity of scoring rules is not necessarily the encoding of the distribution but whether we use deterministic or stochastic scoring rules.

3 An Impossibility Theorem

Before stating our theorems, we first give a formal definition of what it means for a scoring rule to have query complexity k.

Definition 1. *A scoring rule S has* query complexity k *if*

   1. *there exists a function $g : [0, 1]^k \to \mathbb{R}$ and*

   2. *for all states $\omega \in \Omega$, there exist $q_1(\omega), ..., q_k(\omega) \in \Omega$ such that,*

for all probability mass functions $\mathcal{X}(\cdot)$,

$$S(\mathcal{X}, \omega) = g(\mathcal{X}(q_1(\omega)), ..., \mathcal{X}(q_k(\omega))).$$

If the function g is deterministic we say that the scoring rule is deterministic, and if g is randomized, we say that the scoring rule is stochastic.

With the above definition of query complexity, we can state the main result of this paper. Generalizing what was previously known for local scoring rules, we prove that, when $\Omega = \{0, 1\}^n$, the number of queries required for a deterministic scoring rule to be individual rational is exponential in n, that is, exponential in the number of bits needed to describe a state in Ω.

Theorem 1. *For all $k \leq \frac{|\Omega|}{4}$, no k-query deterministic scoring rule is individually rational.*

Before proving this theorem, we need to give some definitions and results from extremal graph theory, which will be useful in the proof.

3.1 Facts from Graph Theory

Definition 2. *A* graph *$G = (V, E)$ consists of a set V of vertices and a set $E \subset V \times V$ of edges. Unless declared otherwise, all edges of G are* directed*, and an edge from u to v is denoted by (u, v).*

If $(u, v) \in E \iff (v, u) \in E$ for all vertices u, v, then we say that the graph is undirected*, and denote the edge between u, v by the set $\{u, v\}$.*

Definition 3. *An* independent set *in a graph G is a subset of vertices $S \subset V$ such that, for all $u, v \in S$, there is no edge between u and v.*

Definition 4. *A* triangle *in an undirected graph G is a set of vertices α, β, γ such that $\{\alpha, \beta\}, \{\beta, \gamma\}, \{\gamma, \alpha\}$ are all edges of G.*

Lemma 2 (Mantel [8], Turan [15])**.** *Let $G = (V, E)$ be an undirected graph on n vertices with at least $\frac{n^2}{4}$ edges. Then G has a triangle.*

3.2 Proof of Theorem 1

Let S be a k-query scoring rule. Let $g(\mathcal{X}(q_1(\omega)), ..., \mathcal{X}(q_k(\omega)))$ be a function that computes $S(\mathcal{X}, \omega)$ as in definition 1. Recall that that the set of points $(q_1, ..., q_k)$ on which g queries $\mathcal{X}$ is completely determined by ω. We call this set the *neighborhood of ω* and denote it $N(\omega)$. We show that, when $|N(\omega)| \leq \frac{|\Omega|}{4}$, the scoring rule S cannot be individually rational.

This neighborhood relation defines a directed graph $\vec{G}$ over the elements of Ω. There is an edge (α, β) if $\beta \in N(\alpha)$. That is, there is an edge from α to β if the computation of $S(\mathcal{X}, \alpha)$ depends on the value of $\mathcal{X}(\beta)$.

Suppose that $\{\alpha, \beta, \gamma\} \subset \Omega$ is an independent set in $\vec{G}$. That is, there are no edges between any of α, β, γ. This means that the value of $S(\mathcal{X}, \alpha)$ does not depend on the values of $\mathcal{X}(\beta), \mathcal{X}(\gamma)$. Similarly, the values of $S(\mathcal{X}, \beta), S(\mathcal{X}, \gamma)$ do not depend on $\mathcal{X}$ evaluated on the two other points.

Now we have three points α, β, γ such that computing $S(\mathcal{X}, \cdot)$ for any one event does not require consulting the distribution on the other two events. Consider now only inputs $\mathcal{X}$ supported on $\{\alpha, \beta, \gamma\}$. On such distributions, the scoring rule $S(\mathcal{X}, \omega)$ makes at most one query to its input distribution, and when it does this query is to $\mathcal{X}(\omega)$. Thus, it is a local scoring rule: there is a function $h(\cdot)$ such that $S(\mathcal{X}, x) = h(\mathcal{X}(x))$. Since this holds true for all $x \in \{\alpha, \beta, \gamma\}$ and all distributions $\mathcal{X}$ supported on this set, the McCarthy-Savage lemma implies that $h(\mathcal{X}(x)) = a + b \ln(\mathcal{X}(x))$ at least on such inputs. Since the probabilities assigned by $\mathcal{X}$ on α, β, γ can be arbitrarily low, this function is not individually rational.

So, if we can find an independent set $\{\alpha, \beta, \gamma\}$ in $\vec{G}$, we will have proved the theorem. This is where Turan's result comes in.

Consider the *undirected graph* G induced by $\vec{G}$. The vertices are the same as those of $\vec{G}$ and there's an edge $\{\alpha, \beta\}$ if either (α, β) or (β, α) are present in $\vec{G}$. Clearly, if α, β, γ is an independent set of G, it is also an independent set of $\vec{G}$. Thus, it suffices to show that G has an independent set with three elements.

Each vertex $\alpha \in G$ has as many edges coming out of it as there are elements in $N(\alpha)$. By assumption $|N(\alpha)| \leq \frac{|\Omega|}{4}$. Thus, there are at most $\frac{|\Omega|^2}{4}$ edges in $\vec{G}$. The number of edges in G is less than the number of edges in $\vec{G}$, so there are at most $\frac{|\Omega|^2}{4}$ edges in G.

Now let H be the complement graph of G. That is, the vertices of H are the same as those of G, and $\{u, v\}$ is an edge of H if and only if $\{u, v\}$ is *not* an edge of G. Since G has at most $\frac{|\Omega|^2}{4}$ edges, and the complete graph has $\frac{|\Omega|^2}{2}$ edges, we have that the number of edges in H is greater or equal to

$$\frac{|\Omega|^2}{2} - \frac{|\Omega|^2}{4} = \frac{|\Omega|^2}{4}.$$

By Turan's Theorem, H has so many edges that it must contain a triangle. Let α, β, γ be the vertices of this triangle. This means that H contains all the edges $\{\alpha, \beta\}, \{\beta, \gamma\}, \{\gamma, \alpha\}$. But since H is the complement graph of G, it means that *none* of these edges are in G. This means that α, β, γ is an independent set of G.

Thus, since G has an independent set of size 3, the scoring rule that induces G cannot be individually rational. ∎

A remark on tightness Since any scoring rule can make at most $|\Omega|$ queries to its input distribution, our bound is tight up to a constant factor. However, it would be interesting to see if our lower bound of $\frac{|\Omega|}{4}$ queries for an individually rational scoring rule can be improved, or whether there exists a truthful and individually rational scoring rule that makes $\frac{|\Omega|}{4} + 1$ queries to its input distribution. If such a scoring rule existed, its induced graph would not have any independent sets of size 3, and it would be highly regular. We do not see an immediate way in which a scoring rule inducing such a structured graph would guarantee good incentives. It is thus plausible that our lower bound can be improved.

Let us discuss two deterministic attempts at bypassing our lower bound.

4 Trying to Bypass Our Impossibility Theorem

Reduction to distributions with full support As already mentioned, Dawid, Lauritzen and Parry [3] give a deterministic scoring rule over $\Omega = \{0, 1\}^n$ that is individually rational, truthful, and only queries $O(\log |\Omega|)$ states of the world. However, their scoring rule is truthful only when restricted to distributions which have full support, that is, which do not assign zero probability to any state of the world. Note that the notion of a scoring rule should apply to *arbitrary distributions*. Indeed, the true distribution may assign zero

probability to many states (especially when Ω is very large) and if we do not allow the expert to communicate non-full support distributions then the expert cannot possibly be truthful.

Can this problem be avoided is by smoothing distributions that do not have full support? That is, can the decision maker add a small probability ϵ to every state of the world, and then renormalize to obtain a valid probability distribution? We show that, even though this is syntactically possible, this smoothing approach can severely distort the input distribution, to the point where it no longer conveys useful information.

As a concrete example, let $\Omega = \{1, 2, 3, ..., 2^n\}$ and let $\mathcal{X}$ be the distribution assigning all of its mass to the state of the world $\omega = 1$. The expectation of $\mathcal{X}$ is therefore 1. It is important for the expert to be able to convey such a distribution, because it implies certainty about a given state of the world.

If one smooths $\mathcal{X}$ by adding ϵ probability to every state of the world, the expectation of the smoothed distribution becomes significantly skewed. The new expectation will be greater than $\frac{1+\epsilon}{1+2^n \cdot \epsilon} 1 + \frac{2^n \cdot \epsilon}{1+2^n \cdot \epsilon} 2^{n-1}$. As n becomes very large, the expectation of the smoothed distribution is higher than 2^{n-1}, while the expectation of the original $\mathcal{X}$ is always one. Thus, smoothing can severely distort the important properties of the distribution conveyed by the expert.

Amortizing Computation in Market Scoring Rules   Another way to bypass our impossibility result in practice would be through amortizing the computation of the scoring rule reward over time. A possible application of this technique is in designing market scoring rules, which allow the aggregation of multiple experts' knowledge. In this setting, the decision maker initializes the market by posting her initial beliefs $\mathcal{X}_0$. At time t, any expert participating in the market can suggest an updated belief distribution $\mathcal{X}_t$. After all these predictions are made, the state of the world ω is revealed, and the experts are paid. The expert who made the prediction $\mathcal{X}_t$ at time t receives a payment $S(\mathcal{X}_t, \omega) - S(\mathcal{X}_{t-1}, \omega)$, reflecting how much she improved the prediction at time $t-1$. If the expert decreased the quality of the prediction, then her reward can be negative and she needs to pay the decision maker. In a setting where $\mathcal{X}_t$ is not much different from $\mathcal{X}_{t-1}$, the number of queries made to the predicted distributions may be amortized over time.

To illustrate why this is the case, imagine that $\mathcal{X}_t$ and $\mathcal{X}_{t-1}$ are almost the same, except that they differ in the probabilities that they assign to two outcomes α and β. As usual, let ω be the revealed state of the world, and for the purposes of this example, assume that $\omega \neq \alpha, \beta$. Assume that the decision maker uses the quadratic scoring rule and has already computed $S(\mathcal{X}_{t-1}, \omega) = 2\mathcal{X}_{t-1}(\omega) - \sum_{x \in \Omega} \mathcal{X}_{t-1}(x)^2 + 1$. Now she needs to compute $S(\mathcal{X}_t, \omega) = 2\mathcal{X}_t(\omega) - \sum_{x \in \Omega} \mathcal{X}_t(x)^2 + 1$. Since $\omega \neq \alpha, \beta$, we have that the difference $S(\mathcal{X}_t, \omega) - S(\mathcal{X}_{t-1}, \omega)$ is equal to the difference between the square norms

$$\sum_{x \in \Omega} \mathcal{X}_t^2(x) - \sum_{x \in \Omega} \mathcal{X}_{t-1}^2(x) = \mathcal{X}_t^2(\alpha) + \mathcal{X}_t^2(\beta) - \mathcal{X}_{t-1}^2(\alpha) - \mathcal{X}_{t-1}^2(\beta).$$

Thus, given that the decision maker has already computed $\sum_{x \in \Omega} \mathcal{X}_{t-1}^2(x)$, they can compute $\sum_{x \in \Omega} \mathcal{X}_t^2(x)$ using only 4 queries to these input distributions. This means that each

new computation of a reward $S(\mathcal{X}_t, \omega)$ requires a constant number of queries to the input distribution, allowing us to amortize the cost of computing the scoring rule over time. This technique would apply when $S(\mathcal{X}_0, \omega)$ can be pre-computed, and when each new expert only changes the existing predicted distribution at a few states of the world. In the general case where each new expert is allowed to give an arbitrary new distribution, the computation of the market scoring rule rewards cannot be amortized in this way.

5   A new Scoring Rule

Theorem 1 applies only to scoring rules that are deterministic. We now show how to construct a *stochastic* scoring rule which is truthful, ex-post individually rational, and makes only two queries to its input distribution.

In contrast with cited prior work [9], we emphasize that we want the input distribution to be given by its probability mass function rather than its cumulative distribution function.

Theorem 2. *There exists a stochastic scoring rule that is truthful, individually rational and makes only two queries to its input distribution.*

Proof. We prove this theorem by constructing the required scoring rule.

Stochastic Scoring Rule $\mathcal{R}$

On inputs a distribution $\mathcal{X}$ and a state $\omega \in \Omega$:

1. Sample a uniformly random state α from Ω.

2. Query $\mathcal{X}(\alpha)$ and compute $\mathcal{X}(\alpha)^2$.

3. Output a reward $S(\mathcal{X}, \omega, \alpha) = \frac{2}{|\Omega|}\mathcal{X}(\omega) - \mathcal{X}(\alpha)^2 + 1$.

Note that the above scoring rule only makes two queries to $\mathcal{X}$: at the realized state of the world ω, and at a uniformly random state of the world α. In addition, it is very easy to compute once the queries have been made. We now show that this scoring rule is truthful and ex-post individually rational.

Given $\mathcal{X}$ and ω, let $S(\mathcal{X}, \omega) = \mathbb{E}_\alpha[S(\mathcal{X}, \omega, \alpha)]$ be the expected reward where the expectation is taken over the scoring rule's own coin tosses. Since α is a uniformly random state from Ω, we have that this expected reward is $\frac{2}{|\Omega|}\mathcal{X}(\omega) - \frac{1}{|\Omega|}\sum_{\alpha \in \Omega}\mathcal{X}(\alpha)^2 + 1$. Note that this expected reward is *exactly* a linear transformation of the reward given by the quadratic scoring rule. Since the quadratic scoring rule remains truthful under positive linear transformations, our new scoring rule is also truthful.

To argue ex-post individual rationality, we need to show that our new scoring rule always gives a non-negative reward, for every possible value of $\mathcal{X}, \omega$ and α. When these inputs are fixed, the new scoring rule gives a reward equal to $\frac{2}{|\Omega|}\mathcal{X}(\omega) - \mathcal{X}(\alpha)^2 + 1$. Since $\mathcal{X}(\alpha)^2 \leq 1$ and $\frac{2}{|\Omega|}\mathcal{X}(\omega)$ is always non-negative, we obtain that the scoring rule always gives a non-negative reward. ∎

6 Final Remarks

We showed that stochastic scoring rules provide significant advantages. Indeed, we proved that there is a very significant tradeoff between individual rationality and query efficiency in deterministic scoring rules. In contrast, we also proved that there are stochastic scoring rules that are individually rational and make only two queries to their input distribution.

One caveat about using scoring rules to reward an expert is guaranteeing that the reported $\mathcal{X}$ is indeed a distribution, that is, $\sum_\omega \mathcal{X}(\omega) = 1$. For many practical ways of communicating the probability mass function, such as giving a bayesian network where each node has low degree, the decision maker can quickly verify that the expert is really providing a distribution. However, when this is not possible, the expert might be able to increase his expected reward by, for example, announcing the function $\mathcal{X}'(\cdot) = 2\mathcal{X}(\cdot)$. We note, however, that there are separate and general ways to prevent this. For instance, one may require the expert to provide a computationally sound proof [12], or a rational proof [1], that indeed $\sum_\omega \mathcal{X}(\omega) = 1$.

Acknowledgements

Appendix

A   Proof of the McCarthy-Savage Lemma

For completeness, we sketch the proof of the McCarthy-Savage Lemma.

Lemma (McCarthy - Savage [10] [14]). *If $|\Omega| \geq 3$ then the only local strictly proper scoring rules are of the form $S(\mathcal{X}, \omega) = a + b \ln \mathcal{X}(\omega)$ where $b > 0$.*

Proof. Let $S(\mathcal{X}, \omega)$ be a strictly proper local scoring rule. This means that we can write $S(\mathcal{X}, \omega) = g(\mathcal{X}(\omega))$ for some function g. For convenience of notation, we write $f(\omega)$ for $\mathcal{X}(\omega)$. In this proof sketch, we assume that g is differentiable.

Given two probability mass functions f, ϕ, let $G_f(\phi) = \sum_\omega f(\omega) \cdot g(\phi(\omega))$ be the expected reward when the expert announces ϕ and the true distribution follows the function f. Since the scoring rule is strictly proper, this function is strictly maximized by setting $\phi = f$. We can think of ϕ as a vector of real numbers, under the constraints that $\phi(\omega) > 0$ for all ω and $\sum_\omega \phi(\omega) = 1$. Thus, the Lagrange function is $\Lambda_f(\phi, \lambda) = \sum_\omega f(\omega) \cdot g(\phi(\omega)) - \lambda \cdot (\sum_\omega \phi(\omega) - 1)$. The derivative with respect to $\phi(\omega)$ is $\frac{\partial \Lambda}{\partial \phi(\omega)} = f(\omega) \cdot g'(\phi(\omega)) - \lambda$. Since the scoring rule is strictly proper, this derivative equals zero only when $\phi(\omega)$ is set to equal the true distribution $f(\omega)$. Thus, we have the first order condition

$$f(\omega) \cdot g'(f(\omega)) = \lambda.$$

Renaming $f(\omega) = x$ and dividing both sides by x we get

$$g'(x) = \frac{\lambda}{x}.$$

Solving this differential equation gives us $g(x) = \lambda \ln(x) + C$, where C is a constant.

In the above function, we assumed that g was differentiable. Note that Savage [14] shows that $g(\cdot)$ must be monotonic, and thus differentiable almost everywhere. ∎

References

[1] P.D. Azar and S. Micali. Rational proofs. In *Proceedings of the 44th symposium on Theory of Computing*, pages 1017–1028. ACM, 2012.

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

[3] A.P. Dawid, S. Lauritzen, and M. Parry. Proper local scoring rules on discrete sample spaces. *The Annals of Statistics*, 40(1):593–608, 2012.

[4] W. Ehm and T. Gneiting. Local proper scoring rules of order two. *The Annals of Statistics*, 40(1):609–637, 2012.

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

[6] I.J. Good. Rational decisions. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 107–114, 1952.

[7] J. Kilian. A note on efficient zero-knowledge proofs and arguments. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 723–732. ACM, 1992.

[8] W. Mantel. Problem 28. *Wiskundige Opgaven*, 10(60-61):320, 1907.

[9] J.E. Matheson and R.L. Winkler. Scoring rules for continuous probability distributions. *Management Science*, 22(10):1087–1096, 1976.

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

[11] R. Merkle. A digital signature based on a conventional encryption function. In *Advances in CryptologyCRYPTO87*, pages 369–378. Springer, 1987.

[12] S. Micali. Computationally sound proofs. *SIAM Journal on Computing*, 30(4):1253–1298, 2000.

[13] M. Parry, A.P. Dawid, and S. Lauritzen. Proper local scoring rules. *The Annals of Statistics*, 40(1):561–592, 2012.

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

[15] P. Turan. On an extremal problem in graph theory (in hungarian), mat. *Fiz. Lapok*, 48:436–452, 1941.