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

June 18, 2012

Abstract

Proper scoring rules are crucial tools to elicit truthful information from experts. More precisely, given some uncertain state of the world ω, and an expert opinion $\mathcal{X}$ about the probability distribution of ω, a scoring rule is a function $S(\mathcal{X}, \omega)$ mapping the expert's beliefs and the realized outcome to a real number, the expert's score or reward. To evaluate this reward, we must query the distribution $\mathcal{X}$ at different points. The number of queries that we need to make to $\mathcal{X}$ is a natural measure of the *complexity* of the scoring rule.

We prove that any bounded and strictly proper scoring rule 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.

1 Introduction

Strictly Proper Scoring Rules Let Ω be a set of states of the world. We denote by $\Delta(\Omega)$ the set of probability distributions over Ω. Given a distribution $\mathcal{X}$, we denote by $f_{\mathcal{X}}(\cdot)$ its probability density function. A strictly proper scoring rule is a function $S : \Delta(\Omega) \times \Omega \to \mathbb{R}$ such that, for all distinct distributions $\mathcal{D}$ and $\mathcal{X}$:

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

Two popular strictly proper scoring rules are the logarithmic scoring rule [8]

$$S(\mathcal{X}, \omega) = \log(f_{\mathcal{X}}(\omega))$$

and the quadratic scoring rule [3]

$$S(\mathcal{X}, \omega) = 2 f_{\mathcal{X}}(\omega) - \sum_{\omega \in \Omega} f_{\mathcal{X}}(\omega)^2 - 1.$$

Examples of other scoring rules, together with a comprehensive survey and characterization, can be found in Gneiting and Rafterty [7]. We focus on settings where Ω is a finite set. We will often consider Ω to be the set $\{0,1\}^n$ of n-bit strings, but this is not necessary for our results.

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*.

Scoring Rules as Expert-Rewarding Systems Scoring rules are crucial tools for incentivizing agents to truthfully reveal information about some future uncertain state of the world. This valuable information traditionally consists of a "true" distribution $\mathcal{D}$ over the states of the world Ω. In the classical model there are two agents, an *expert*, who knows the distribution $\mathcal{D}$, and a *principal* who wants to pay the expert for her information. The principal does not know the distribution, but does observe a sample ω drawn from the distribution $\mathcal{D}$. A scoring rule S enables them to accomplish their goals by means of the following two-stage interaction:

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

We highlight that the expert is not trusted to announce the correct distribution she knows. Thus, in principle it may be that $\mathcal{X} \neq \mathcal{D}$. However, the scoring rule ensures that the only way for her to maximize her reward is to announce exactly the distribution $\mathcal{D}$ from which the sample ω is drawn.

Applications Proper scoring rules have been applied to meteorology [12], economic forecasting [13] and medicine [4], among others. In a previous work, the authors apply scoring rules to auctions [1]. A new application of scoring rules to computational complexity has been pointed out by the authors [2]. Namely, proper scoring rules can be used to incentivize an expert with sufficient computational power to reveal the correct solution to a *hard and deterministic* computational problem.

Queries A distribution $\mathcal{X} \in \Delta(\Omega)$ may be conceptualized as an *algorithm,* which we may continue to denote by $\mathcal{X}$, that, on an input $\omega \in \Omega$, outputs the value $f_{\mathcal{X}}(\omega)$. (In the worst case, such an algorithm keeps the list of pairs $(\omega', \mathcal{X}(\omega'))$ for each state ω' in the support of $\mathcal{X}$, and, on input ω, scans the list to compute the value of $f_{\mathcal{X}}(\omega)$. As for a more interesting example, in the case of the binomial distribution $B(m,p)$, such an algorithm only keeps the binary representation of m and p, from which it can easily compute $f_{\mathcal{X}}(\omega)$ for any given ω.) Accordingly, we may envisage that a scoring rule S computes $S(\mathcal{X},\omega)$ by *querying* $\mathcal{X}$ at each desired state ω', each time obtaining in response the corresponding value of $f_{\mathcal{X}}(\omega')$.

1.1 Desirable Properties of Scoring Rules and Our Result

A desirable property for a scoring rule S is that $S(\mathcal{X}, \omega)$ always lies in some bounded and positive interval $[0, B]$. If the reward that the expert gets is negative, then it is not (ex-post) rational for her to give her knowledge about the state of the world. If the reward is too high, then the principal might not have enough money to pay the expert for his knowledge.

Another desirable property of a scoring rule is that it make few queries to its input distribution $\mathcal{X}$. Informally, S makes k queries if $S(\mathcal{X}, \omega)$ can be computed from k queries to the probability distribution function $f_{\mathcal{X}}(\omega_1), ..., f_{\mathcal{X}}(\omega_k)$. The logarithmic rule makes a single query to the input distribution, but its range is the interval $(-\infty, 0]$. By contrast, the quadratic rule (modified by adding 2 to its output) has $[0, 2]$ as its range, but requires making $|\Omega|$ queries to the input distribution. It is known that scoring rules which only make a single query to $\mathcal{X}$ *cannot be bounded*. This paper significantly extends this state of knowledge by showing that no scoring rule making a number of queries less than a quarter of all possible states can be bounded.

2 A formal definition of query complexity

Let us define what it means for a scoring rule to have query complexity k.

Definition 1. *A scoring rule S is a* k-query scoring rule *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 density functions $f_{\mathcal{X}}(\cdot)$,

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

We say that a scoring rule S has query complexity k if it is a k-query scoring rule.

3 Prior Work

It would be ideal to have a scoring rule that is both bounded and makes only one query to its input distribution. Such rules are usually called local. Unfortunately, scoring rules that are bounded and make only one query to their input distribution do not exist.

Lemma 1 (McCarthy - Savage [11] [15])**.** *If $|\Omega| \geq 3$ then the only local strictly proper scoring rules are of the form $S(\mathcal{X}, \omega) = a + b \ln f_{\mathcal{X}}(\omega)$ where $b > 0$.*

Parry, Dawid and Lauritzen [14] 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 k^{th} 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 [6] 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 $f_{\mathcal{X}}(\omega)$ and its derivatives, we study scoring rules that depend on $f_{\mathcal{X}}$ when evaluated at ω and at $k-1$ other points in the state space. Second, we focus on discrete distributions, whereas their work focus on continuous ones.

A more closely related paper, by Dawid, Lauritzen and Parry [5], 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.

The results in [5] are useful for developing good heuristics for special cases of the distribution $\mathcal{D}$. For instance, example 4.4 in their paper gives a scoring rule over $\{0,1\}^n$ that is bounded and which 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 new result shows that a search for such a scoring rule would be fruitless. For arbitrary distributions over $\Omega = \{0,1\}^n$, a bounded 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.

4 Our Impossibility Theorem

We show that any bounded scoring rule must make a number of queries that linear in $|\Omega|$. When $\Omega = \{0,1\}^n$, the number of queries required is exponential in n, the number of bits that we need to describe a state in Ω.

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

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

4.1 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 [10], Turan [16])**.** *Let $G = (V, E)$ be an undirected graph on n vertices with at least $\frac{n^2}{4}$ edges. Then G has a triangle.*

4.2 Proof of our theorem

We now present a proof of our theorem.

Let S be a k-query scoring rule. Let $g(f_{\mathcal{X}}(\omega), f_{\mathcal{X}}(q_1(\omega)), ..., f_{\mathcal{X}}(q_k(\omega)))$ be a function that computes $S(f_{\mathcal{X}}, \omega)$ 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 must be unbounded.

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 $f_{\mathcal{X}}(\beta)$.

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 $f_{\mathcal{X}}(\beta), f_{\mathcal{X}}(\gamma)$. 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 $S(\mathcal{X}, \cdot)$ makes only one query to its input distribution. That is, it is a local scoring rule. There is a function $h(\cdot)$ such that $S(\mathcal{X}, x) = h(f_{\mathcal{X}}(x))$. 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 $h(f_{\mathcal{X}}(x)) = a + b\ln(f_{\mathcal{X}}(x))$ at least on such inputs. Clearly this is not a bounded function. 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 must be unbounded.

5 Conclusion

We showed that there is a very significant tradeoff between boundedness and locality in scoring rules. In particular, any bounded scoring rule must make at least $\frac{|\Omega|}{4}$ queries to its input distribution. This generalizes the well known result that the only scoring rule which makes one query to its input distribution is the logarithmic scoring rule, and implies that computing bounded and strictly proper scoring rules for arbitrary distributions is computationally infeasible.

Acknowledgements

We thank Sebastien Lahaie, David Pennock, Philip Reny, Eran Shmaya and Hugo Sonnenschein for helpful comments.

This paper was partially funded by the Office of Naval Research, award number N00014-09-1-0597.

References

[1] P. Azar, J. Chen, and S. Micali. Crowdsourced bayesian auctions. In *3rd conference on Innovations in Theoretical Computer Science.*

[2] P. Azar and S. Micali. Rational proofs. In *Symposium on the Theory of Computing, 2012*. ACM, 2012.

[3] G.W. Brier. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 1950.

[4] C. Czado, T. Gneiting, and L. Held. Predictive model assessment for count data. *Biometrics*, 65(4):1254–1261, 2009.

[5] A. P. Dawid, S. Lauritzen, M. Parry. Proper local scoring rules on discrete sample spaces. *The Annals of Statistics*, Volume 40, Number 1 (2012), 593-608.

[6] W. Ehm, T. Gneiting. Local proper scoring rules of order two. *The Annals of Statistics*, Volume 40, Number 1 (2012), 609-637.

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

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

[9] N.S. Lambert, D.M. Pennock, and Y. Shoham. Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138. ACM, 2008.

[10] W. Mantel. Problem 28. *Wiskundige Opgaven*, 10:60–61, 1906.

[11] 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.

[12] A.H. Murphy and R.L. Winkler. Probability forecasting in meteorology. *Journal of the American Statistical Association*, pages 489–500, 1984.

[13] FM O'Carroll. Subjective probabilities and short-term economic forecasts: An empirical investigation. *Applied Statistics*, pages 269–278, 1977.

[14] M. Parry, A. P. Dawid, S. Lauritzen. Proper local scoring rules. *The Annals of Statistics*, Volume 40, Number 1 (2012), 561-592.

[15] L.J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, pages 783–801, 1971.

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