

Optimal Scoring Rule Design under Partial Knowledge

Yiling Chen¹[0000-0002-0166-594X] and Fang-Yi Yu²[0000-0002-3697-8807]

¹ Harvard University, Cambridge, MA 02138

yiling@seas.harvard.edu

² George Mason University, Fairfax, VA 22030

fangyiyu@gmu.edu

Abstract. This paper studies the design of optimal proper scoring rules when a principal has partial knowledge of an agent’s signal distribution. Recent work [23] characterizes the proper scoring rules that maximize the increase of an agent’s payoff when the agent chooses to access a costly signal to refine a posterior belief from his prior prediction, under the assumption that the agent’s signal distribution is fully known to the principal. In our setting, the principal only knows about a set of distributions where the agent’s signal distribution belongs. We formulate the scoring rule design problem as a max-min optimization that maximizes the worst-case increase in payoff across the set of distributions.

We propose an efficient algorithm to compute an optimal scoring rule when the set of distributions is finite, and devise a fully polynomial-time approximation scheme that accommodates various infinite sets of distributions. We further remark that widely used scoring rules, such as the quadratic and log rules, as well as previously identified optimal scoring rules under full knowledge [23], can be far from optimal in our partial knowledge settings.

Keywords: Proper scoring rule · Crowdsourcing · Optimization.

1 Introduction

Proper scoring rules are scoring functions that incentivize truthful information elicitation: an agent with a subjective belief about an uncertain event maximizes his expected score by making a prediction according to his belief. If the agent can acquire a costly signal to refine his belief, which proper scoring rules maximally incentivize the agent’s information acquisition? Recent work [23] explores this question under the assumption that the principal has full knowledge about the agent’s information structure. In contrast, we investigate this optimal scoring rule design question when the principal has only partial knowledge about the agent’s information structure, only knowing the set of information structures that the agent’s belongs to.

Incentivizing information acquisition is crucial in many real-world applications. For example, in conference reviewing, a reviewer could spend little to

no effort and form his prior assessment about a paper primarily based on the length of the paper, the amount of grammatical errors in the introduction, or the references cited. But a conference chair would aspire for a high-effort review where the reviewer carefully reads the paper and then forms his, more informed, posterior assessment. As another example, in crowdsourced prediction for the reproducibility of scientific studies [27, 24], an expert could form a prediction on the replication outcome of a study based on his general knowledge about the study’s topic, the reputation of the publication venue, and the authors’ affiliations. But a more valuable and accurate prediction requires the expert to examine the study’s methodology and evaluation procedures carefully. In these applications, the principal ideally hopes to devise a proper scoring rule that maximizes the expected increase in score if the agent acquires the information, to maximally incentivize information acquisition. The principal however only has limited knowledge about the agent’s information structure.

More formally, an agent’s information structure consists of two parts: the agent’s prior belief about the random variable of interest (the prior) and the distribution of the agent’s signal conditioned on every realization of the random variable (an experiment). We formulate the optimal scoring rule design problem as a max-min optimization that maximizes the agent’s worst-case increase in score across the set of possible information structures. We explore the problem for four settings of principal’s knowledge, ranging from the special full-knowledge case to varying degree of partial knowledge:

1. The principal knows both the prior and the experiment and hence the set of information structures is a singleton. We reprove [23]’s results and show that the optimal scoring rules are v -shaped in proposition 1. In addition, proposition 1 also provides a close-form expression for an optimal scoring rule.
2. The principal knows the prior but is uncertain about the experiment. The set of information structures shares the same prior, as introduced in example 2. Interestingly, we show in theorem 2 that the same v -shaped scoring rule in proposition 1 is also optimal.
3. The principal knows the experiment but is uncertain about the agent’s prior. The set of information structures shares the same experiment, as introduced in example 1. We present a fully polynomial time approximation scheme (FPTAS) that outputs approximately optimal piece-wise linear scoring rules in theorem 4.
4. The principal is uncertain about both the prior and the experiment. In theorem 3, we propose an efficient algorithm to compute an optimal scoring rule when the set of distributions is finite. Additionally, for ρ -correlated information structures defined in example 3, we develop an FPTAS to find approximately optimal scoring rules in theorem 4. ρ -correlated information structures have interesting connections to Beta-Bernoulli model [28] and noise operator in Boolean function analysis [31].

We then run simulations to evaluate the performance of two frequently used proper scoring rules (quadratic scoring rule and log scoring rule), the v -shaped

scoring rule (which is optimal in the first two settings) and the piece-wise linear scoring rules obtained by our FPTAS for the ρ -correlated information structures. The simulations show that our algorithm's piecewise linear scoring rules perform well, and provide the most uniform incentive when the signal and the state have a large correlation. However, when the correlation is small, log scoring rule outperforms our piecewise linear scoring rules and other scoring rules. In particular, as the correlation decreases, we observe that the piecewise scoring rule approaches the log scoring rule which may suggest an interesting connection between ρ -correlated information structure and the log scoring rule. The v -shaped scoring rule empirically performs worst in our partial knowledge setting.

Organization and contribution of technical results Using Savage's representation of proper scoring rules [25, 36, 18], we convert variable space of our max-min optimization problem from scoring rules to convex functions where the increase in payoff becomes the Jensen's gap of the associated convex function as problem 1.

In section 3.1, we derive a geometric interpretation of the optimization problem. Lemma 1 shows that the information gain amounts to how *curved* the associated convex function is at the prior of the information structure. This interpretation is critical as it leads to theorem 2 that shows v -shaped scoring rule is optimal for known prior setting (example 2).

In sections 3.2 and 3.3, we delve into the scenario of unknown prior settings. We first use linear program to find an optimal piecewise linear scoring rule when the collection of information structures is finite in section 3.2. Then we venture into the domain of infinite information, and provide an FPTAS (theorem 4) for various infinite sets (examples 1 and 3). Informally, our FPTAS runs the linear program on a finite subset of the infinite set of information structures, and provides approximation guarantees as long as the finite subset is an ϵ -covering of the original set under earth mover's distance. To this end, we relate the earth mover's distance to our optimization problem in lemma 3. Additionally, we design a novel coupling that can bound the earth mover's distance of two posterior predictions by the total variation distance of their joint distributions on signal and state in lemma 4. This coupling argument may be of independent interest for designing approximation algorithms for information aggregation and elicitation.

Related work Our problem can be seen as purchasing prediction from strategic people. The work on this topic can be roughly divided into two categories according to whether agents can misreport their signal or prediction. Below we focus on the relationship of our work to the most relevant technical scholarship.

In the first category, to ensure agents reporting their signals or predictions truthfully, there are two settings according to whether money is used for incentive alignment. In the first setting, the analyst uses monetary payments to incentivize agents to reveal their data truthfully. The challenge is to ensure truth-telling gets the highest payments. Existing works verify agents' reports by using either an observable ground truth (proper scoring rules) or peers' reports (peer prediction). For the second setting, individuals' utilities directly depend on an inference or learning outcome (e.g. they want a regression line to be as close to their own data

point as possible) and hence they have incentives to manipulate their reported data to influence the outcome. [11, 26, 34, 22, 8]

Our setting generalizes [23]'s. Our max-min optimization formulation captures the principal's partial knowledge about agents' information structure, while theirs focuses more on known information structure. We consider both ex-ante and ex-post setting which allow us to compare our result to the log scoring rule which is arguably one of the most important proper scoring rules but cannot be ex-post bounded. They show the optimal scoring rule is *v*-shaped when the information structure is known. We generalize the result to known prior setting in theorem 2.

[29] also study optimal scoring rule design problem, but is more related to sequential method [17]. Instead of imposing bounded payment conditions, their objective comprises both payment and accuracy. They consider a special case of Beta-Bernoulli information structures where the prior is uninformative and design scoring rule for an agent to sequentially acquire samples from a Bernoulli distribution.

Finally, [33] study an information acquisition problem: When the information structure and the cost of information are known, the principal chooses a proper scoring rule (menu of contracts) to incentivize information acquisition as cheaply as possible subject to limited liability. Informally, their formulation can be seen as the dual of our problem. Instead of maximizing information gain subjected to bounded payment conditions, they want to minimize payment with a lower bound on the information gain. Their optimal scoring rules are also *v*-shaped and similar to our known information structure setting.

In the second category, agents cannot misreport their signal or prediction. The problem of purchasing data from people has been investigated with different focuses, e.g. privacy concerns [16, 12, 15, 30, 10, 37], effort and cost of data providers [35, 5, 1, 7, 38, 9], and reward allocation [14, 2].

More generally, our problem is also related to contract theory [19]. Recent work also studies optimal contracts where the principal does not fully know the agent's cost. [21, 3] Moreover, [4] study delegation problem that tries to optimize the efforts of others. However, the critical difference that sets ours apart from these works is the principal's preference. Most of the work in contract theory aims to maximize the principal's utility, but ours treats the principal's preference as a budget constraint and optimize the efforts of the agents. Our formulation may be more suitable for complicated problems, e.g., peer grading or conference review, that consists of multiple sub-problems, and the principal's utility can not be easily decomposed according to sub-problems. For the modelling choice, [13] studies auction design where the unknown value distribution is from a known collection, without assuming a prior over the collection.

2 Model and Preliminaries

In section 2.1, we first introduce proper scoring rules and two boundedness notions. Then we define information structures that formalize the relationship be-

tween the costly signal and the event of interest, and we specify the principal's partial knowledge and introduce three motivating examples. Finally, we define the value of information and our scoring rule design problem. We further simplify our problem by connecting scoring to convex function in section 2.2.

2.1 Optimal Scoring Rule for Costly Information

This paper studies the design of scoring rules for binary state³ which maps an agent's reported prediction x and the realized ground state $\omega \in \{0, 1\}$ to a score for the agents, $PS(\omega, x)$. As a principal designs a scoring rule for the agent, one desirable scoring rule should elicit truthful predictions.

Definition 1. A scoring rule is proper if for all predictions $x, x' \in [0, 1]$,

$$\mathbb{E}_{\omega \sim x}[PS(\omega, x)] \geq \mathbb{E}_{\omega \sim x}[PS(\omega, x')].$$

In other words, if an agent's prediction for $\omega = 1$ is x , he cannot gain a higher score by misreporting x' . Here $\omega \sim x$ denote random variable $\omega = 1$ with probability x , and 0 otherwise.

Another desirable property of a scoring rule is boundedness as the principal has a finite budget for the reward. Here we consider the following two notions.

Definition 2. A scoring rule PS is ex-post bounded by $B > 0$ if for all $\omega \in \{0, 1\}$ and $x \in [0, 1]$, $PS(\omega, x) \in [0, B]$. Alternatively, PS is ex-ante bounded by B if for all $x \in [0, 1]$ $\mathbb{E}_{\omega \sim x}[PS(\omega, x)] \in [0, B]$.

In addition to properness and boundedness, the principal often wants to design proper scoring rules that incentivize effort. Specifically, when the agent can refine his prediction by exerting a binary effort for a costly signal, how can the principal design a scoring rule that maximizes the agent's perceived gain from exerting effort without complete knowledge of the prior and posterior distributions?

Prior, posterior, and information structures The agent can access a costly signal $S \in \mathcal{S}$ with a finite support that improves his prediction of the unknown binary state of the world $W \in \{0, 1\}$. Specifically, the agent has an **information structure** P which is a joint distribution on the state and signal. An information structure P consists of a *prior* $\pi \in [0, 1]$ of the state and an *experiment* $\sigma : \{0, 1\} \rightarrow \Delta(\mathcal{S})$ which is a conditional distribution of the signal given the state, so that for all state $\omega \in \Omega$ and signal $s \in \mathcal{S}$,

$$\pi = P(W = 1) \text{ and } \sigma(s|\omega) = P(S = s \mid W = \omega).$$

We will refer P by the pair of prior and experiment (π, σ) .

Given P with (π, σ) , if the agent ignores the signal, his truthful prediction of the state is the *prior*, $P(W = 1) = \pi$. If the agent accesses the signal and sees $S = s$, his truthful prediction becomes the *posterior* $P(W = 1 \mid S = s) =$

³ We discuss the d -dimensional setting in the appendix.

$\frac{\pi\sigma(s|1)}{\pi\sigma(s|1)+(1-\pi)\sigma(s|0)}$. The posterior prediction $X(s) := P(W = 1 \mid S = s)$ is a random variable over the randomness of signal. We will omit s and write the random variable of posterior prediction as X . For instance, the expectation of posterior equals prior $\mathbb{E} X = \pi$ by Bayes' rule.

Partial knowledge of information structures In our partial knowledge setting, the principal only knows a collection of information structures \mathcal{P} that the agent's information structure falls into. Below are three examples of collections of information structures that model the principal's partial knowledge. We will use these as running examples throughout the paper.

First the principal may know the signal distribution given ground state (experiment) but does not know the agent's background (prior). Intuitively, this captures online crowdsourcing setting, e.g., image annotation, where the background of the agent is unknown, but quality of signal can be controlled. Note that the priors are bounded away from zero and one by $\delta > 0$, because if the prior is zero or one, the posterior is identical to the prior.

Example 1. Given an experiment σ and $0 < \delta \leq 1/2$, a collection of information structures with homogeneous experiment is $\mathcal{P}_\sigma = \{(\pi, \sigma) : \pi \in [\delta, 1 - \delta]\}$.

To another extreme, the principal may know agent's prior but does not know the agent's experiment. This can model peer grading in classroom where the student's prior is known but the quality of his work is uncertain.

Example 2. Given a prior $\pi \in (0, 1)$ and a set of experiments σ_i where $i \in \mathcal{I}$, a collection of information structures with homogeneous prior is $\{(\pi, \sigma_i) : i \in \mathcal{I}\}$.

The first example has the same experiment for all information structures, and the second has the same prior. We provide an example where the prior and experiment both vary.

Example 3. Given $\rho \in [0, 1]$, $0 < \delta \leq 1/2$ and a binary signal space $\mathcal{S} = \{0, 1\}$, a ρ -correlated experiment produces signal that equals the ground state with probability ρ and samples from prior independently otherwise. A collection of ρ -correlated information structures \mathcal{P}_ρ is

$$\{(\pi, \sigma) : \pi \in [\delta, 1 - \delta], \sigma(1|1) = \rho + (1 - \rho)\pi, \sigma(1|0) = (1 - \rho)\pi\}.$$

Note that the value ρ controls the correlation between the signal and the ground state. In particular, if $\rho = 0$, the signal is independent of the ground state, and if $\rho = 1$, the signal perfectly agrees with the ground state. Similar to example 1, the priors are bounded away from zero and one.

There are several interesting interpretations of ρ -correlated experiments. First, the information structure is the posterior predictive distribution of Beta-Bernoulli model [28] given an additional sample, and ρ captures the strength of prior. Second, the distribution of the signal and the state is also known as ρ -correlated pair in Boolean function analysis [31]. We formalize this connection in the appendix.

Information gain and max-min optimization problem Now we formalize the agent's perceived gain from exerting effort under a proper scoring rule PS . Let $PS(x) := \mathbb{E}_{\omega \sim x} PS(\omega, x)$ be the expected score of truthful reporting a prediction x . Given PS and P with (π, σ) , the expected score of the agent truthfully reporting his initial prediction is $PS(\pi)$. Alternatively, if the agent accesses the costly signal $S = s$, he gets score $PS(X(s))$. Thus, his expected score of accessing the costly signal before seeing it is $\mathbb{E} PS(X)$ over the randomness of X . Since $\mathbb{E} X = \pi$, the agent's gain from exerting effort is

$$\mathbb{E} PS(X) - PS(\pi) = \mathbb{E} PS(X) - PS(\mathbb{E} X).$$

We call the above **information gain** on P under the proper scoring rule PS . Because the information gain is fully determined by the posterior distribution random variable, we can use P and X interchangeably. Additionally, we let \mathcal{X} be the collection of random variables of prediction induced from information structures P in \mathcal{P} .

To maximize the agent's gain from exerting effort for any possible information structure in \mathcal{X} , the principal finds a bounded proper scoring rule $PS : \{0, 1\} \times [0, 1] \rightarrow \mathbb{R}$ which maximizes the worst-case information gain

$$\max_{PS} \min_{X \in \mathcal{X}} \mathbb{E} PS(X) - PS(\mathbb{E} X).$$

We will simplify our optimization in section 2.2 as problem 1.

2.2 Savage's Representation of Proper Scoring Rules

We characterize the space of bounded proper scoring rules. First, Savage's representation of proper scoring rules connects proper scoring rules and convex functions.

Theorem 1 ([25, 36]). *When the state space $\Omega = \{0, 1\}$ is binary, for every proper scoring rule PS , there exists a convex function $H : [0, 1] \rightarrow \mathbb{R}$ so that for all $x \in [0, 1]$ and $\omega \in \{0, 1\}$*

$$PS(\omega, x) = \begin{cases} H(x) + \partial H(x) \cdot (1 - x) & \text{when } \omega = 1, \\ H(x) - \partial H(x) \cdot x & \text{when } \omega = 0 \end{cases}$$

where $\partial H(x)$ is a subgradient of H at x .

Conversely, for every convex function $H : [0, 1] \rightarrow \mathbb{R}$, there exists a proper scoring rule such that the above condition hold.

With above characterization, we can specify proper scoring rules by their associated convex functions. For instance, a *quadratic scoring rule* has a convex function $2(x^2 + (1-x)^2) - 1$, and a *log scoring rule* has a convex function $x \log_2 x + (1-x) \log_2(1-x) + 1$. We will study **piecewise-linear scoring rules** whose associated convex function is piecewise linear $\max\{f_1(x), \dots, f_m(x)\}$ with affine functions f_i , $i = 1, \dots, m$. Finally, **v-shaped scoring rules** in [23] are special

cases of piecewise linear scoring rules, whose corresponding convex function is *v*-shaped with parameter (a, b, c, x_0) so that $H_{(a,b,c,x_0)}(x) = \max\{a(x - x_0) + c, b(x - x_0) + c\}$ with the *vertex* at x_0 . fig. 3a shows examples of quadratic and log scoring rules, fig. 1 is for *v*-shaped scoring rules, and fig. 2 is for piecewise scoring rules.

Now we reformulate our optimization problem in terms of convex functions with the following lemmas by simple applications of theorem 1. The first lemma converts *the information gain as the gap of Jensen's inequality of the associated convex function*. The second lemma shows the bounded conditions can be checked by the value and sub-gradient of the convex function. The proofs are in the appendix.

Lemma 1. *For any proper scoring rule PS with convex function H , the expected score of truthfully reporting x is $PS(x) = H(x)$, and the information gain of information structure X under the proper scoring rule PS is $\mathbb{E} PS(X) - PS(\mathbb{E} X) = \mathbb{E}[H(X)] - H(\mathbb{E} X)$.*

We will define the information gain of information structure X under the proper scoring rule with H as

$$\mathfrak{J}_X(H) := \mathbb{E}[H(X)] - H(\mathbb{E} X). \quad (1)$$

Lemma 2. *Given $B > 0$, for any proper scoring rule PS with convex function H , PS is ex-post bounded by B if and only if $H(x) + \partial H(x)(1 - x)$ and $H(x) - \partial H(x)x$ are in $[0, B]$ for all $x \in [0, 1]$.*

PS is ex-ante bounded by B if and only if $H(x) \in [0, B]$ for all $x \in [0, 1]$.

Let $\mathcal{B}_{\text{expost}}$ denote the set of ex-post bounded convex functions and $\mathcal{B}_{\text{exante}}$ for ex-ante bounded convex function as lemma 2. By lemma 2, $\mathcal{B}_{\text{expost}} \subseteq \mathcal{B}_{\text{exante}}$. We now derive the simplified program for our max min optimization problem over the space of convex functions.

Problem 1. Given a set of information structures \mathcal{X} and a set of convex functions \mathcal{B} , find a convex function $H : [0, 1] \rightarrow \mathbb{R}$ which maximizes the worst-case information gain

$$\max_{H \in \mathcal{B}} \min_{X \in \mathcal{X}} \mathfrak{J}_X(H).$$

We will focus on \mathcal{B} being $\mathcal{B}_{\text{expost}}$ in ex-post bounded setting and $\mathcal{B}_{\text{exante}}$ in ex-ante setting.

3 Main Results

In section 3.1, we explore the setting of known information structure and show *v*-shaped scoring rules are optimal in proposition 1. We further show that the same *v*-shaped scoring rule is also optimal in the setting of known prior but uncertain about the experiment in theorem 2. We shift our focus to the unknown prior setting. We consider finite collections of information structures in section 3.2,

and devise an efficient algorithm that solves for the optimal scoring rules in theorem 3. Finally, we design an FPTAS in theorem 4 for infinite collections of information structures, examples 1 and 3.

We provide outlines of proofs and intuitions, while complete proofs are deferred to the appendix.

3.1 Singleton and Homogeneous Prior Information Structures

As a warm-up, let's consider the principal exactly knows the agent's information structure so that $\mathcal{X} = \{X\}$ is a singleton. We show that the optimal H can be *v*-shaped (defined in section 2.2) with the vertex at the prior in both ex-ante and ex-post bounded settings.

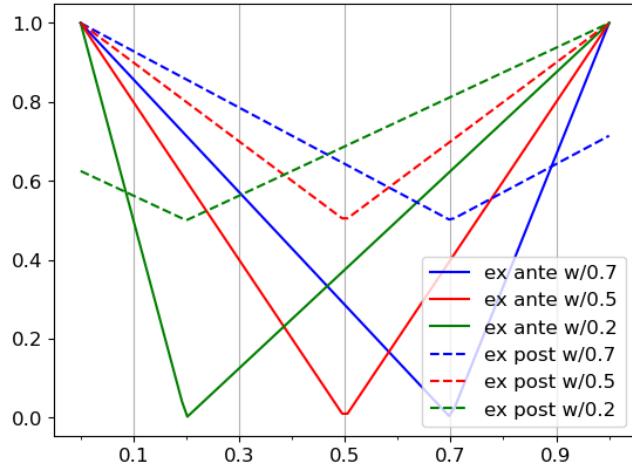


Fig. 1: The optimal scoring rules for the homogeneous prior setting. With theorem 2, the dashed lines are the optimal *v*-shaped convex functions for three different priors (0.7, 0.5, and 0.2) in the ex-post setting, and the solid ones are in the ex-ante setting.

Proposition 1 (singleton). *If $\mathcal{X} = \{X\}$ is singleton in ex-post or ex-ante bounded settings, there exists an optimal scoring rule associated with a *v*-shaped convex function for problem 1.*

1. A *v*-shaped convex function with $x_0 = \pi$, $a = \frac{-B}{\pi}$, $b = \frac{B}{1-\pi}$ and $c = 0$ is optimal in ex-ante bounded setting with B .
2. A *v*-shaped convex function with $x_0 = \pi$, $a = \frac{-B}{2 \max\{\pi, 1-\pi\}}$, $b = \frac{B}{2 \max\{\pi, 1-\pi\}}$, and $c = \frac{B}{2}$ is optimal in ex-post bounded setting with B .

While the above result is already proved in [23], we provide an explicit closed-form expression of the optimal solution and present an alternative proof in the appendix. We generalize the proof and show that v -shaped scoring rule is optimal for any collection of information structure when they share the same prior (defined in example 2).

Theorem 2 (homogeneous prior). *Given any collection of information structures with homogeneous prior π (example 2), there exists an optimal scoring rule associated with a v -shaped convex function with the vertex at π in both ex-ante and ex-post bounded settings respectively.*

The proof follows from the fact that the optimal v -shaped scoring rules only depend on the prior. When multiple information structures have the same prior, the same v -shaped scoring rule is still optimal.

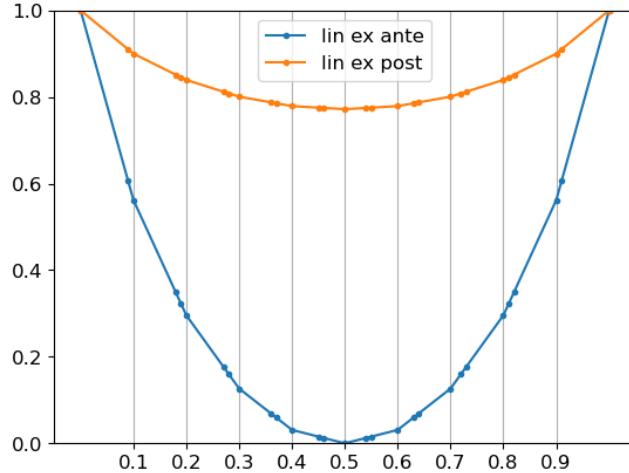


Fig. 2: The optimal piecewise linear functions for a finite collection of information structures $\mathcal{P}_{\rho, N} = \{(\pi, \sigma) \in \mathcal{P}_\rho : N\pi \in \mathbb{N}\}$ with $\rho = 0.1$ and $N = 10$ (section 4). In both ex-post and ex-ante setting, the vertex of the function is at the prior of the collection of information structures which is aligned with our intuition in section 3.3 which suggests maximizing curvature at the prior.

These results suggest that the principal should choose H that is “curved” at the prior in order to incentivize the agent to derive the signal and move away from the prior as fig. 1.

3.2 Finite Collections of Information Structures

If the collection of information structures is *finite* so that $|\mathcal{X}| < \infty$, we give an efficient algorithm to compute an optimal piecewise linear scoring rule. Finite

collections of information structures are natural when there are finite types of agents, and is useful to approximate some infinite collections of information structures as shown in the next section.

Theorem 3. *If \mathcal{X} is finite, for both bounded settings, there exists an optimal proper scoring rule that is piecewise linear and can be derived by solving a linear program in time polynomial in $|\mathcal{X}| \cdot |\mathcal{S}|$.*

The main idea is that when \mathcal{X} is finite $\mathfrak{J}_X(H)$ in eq. (1) only depends on the evaluations of H on the support of \mathcal{X} $\overline{\text{supp}}(\mathcal{X}) := \cup_{X \in \mathcal{X}}(\text{supp}(X) \cup \{\mathbb{E} X\})$. Thus, instead of searching all possible bounded convex functions, we can reduce the dimension of problem 1 and use a linear program whose variables contain the evaluations of H in $\overline{\text{supp}}(\mathcal{X})$ and linear constraints ensure that the evaluations can be extended to a convex piecewise linear function. Figure 2 presents an example of piece-wise linear convex function outputted by our linear program. This observation allows us to solve the problem in weakly polynomial time in $|\overline{\text{supp}}(\mathcal{X})| \leq |\mathcal{X}| \cdot |\mathcal{S}|$. We present the formal proof in the appendix.

3.3 Infinite Collections of Information Structures

For a non-finite collection of information structures, solving problem 1 exactly is generally infeasible, because even evaluating the objective value $\min_{X \in \mathcal{X}} \mathfrak{J}_X(H)$ may depend on evaluating all $X \in \mathcal{X}$, which could require unbounded time. In particular, this limitation also applies to examples 1 and 3. Below, we extend our algorithms for finite collections of information structures to more general collections. We demonstrate the idea by approximately solving the optimal scoring rule for examples 1 and 3 in theorem 4. Remark 1 discusses the potential and limitation of our method for abstract collections of information structures.

Theorem 4. *Given any $\epsilon > 0, \delta > 0$ and $B > 0$, in ex-post bounded setting $\mathcal{B} = \mathcal{B}_{\text{expost}}$ (or ex-ante setting $\mathcal{B} = \mathcal{B}_{\text{exante}}$), there exists an efficient algorithm that outputs an ϵ -optimal scoring rule H on information structures with a homogeneous experiment (in example 1) so that,*

$$\min_{X \in \mathcal{X}} \mathfrak{J}_X(H) \geq \max_{H' \in \mathcal{B}} \min_{X \in \mathcal{X}} \mathfrak{J}_X(H') - \epsilon,$$

with running time polynomial in B and $1/\epsilon$ (or B , $1/\epsilon$ and $1/\delta$ respectively). The same results hold for ρ -correlated information structures in example 3.

Our FPTAS computes a finite collection of information structures, and runs our linear program in theorem 3. Recall that \mathcal{P}_σ is the collection of information structures with homogeneous experiment (in example 1), our algorithm picks a finite set information structures

$$\mathcal{P}_{\sigma, N} = \{(\pi, \sigma) \in \mathcal{P}_\sigma : N\pi \in \mathbb{N}\}$$

with $N = \lfloor B/(2\epsilon) \rfloor$ and outputs an optimal piecewise linear scoring rule $H_{\sigma, N}$ for $\mathcal{P}_{\sigma, N}$ using theorem 3. Similarly, for ρ -correlated information structures, let

$\mathcal{P}_{\rho,N} = \{(\pi, \sigma) \in \mathcal{P}_\rho : N\pi \in \mathbb{N}\}$. Our algorithm outputs an optimal piecewise linear scoring rule $H_{\rho,N}$ for \mathcal{P}_ρ .

The main challenge is to show the approximation guarantees. We observe that given a pair of information structures, the difference between information gains should be small if their posterior distribution is close. Formally, given two posteriors X and X' induced by two information structures P and P' respectively, the **earth mover's distance** (EMD) [6] between posteriors is $d_{EM}(X, X') = \sup_{f: \|f\|_{Lip} \leq 1} \mathbb{E}[f(X)] - \mathbb{E}[f(X')]$ where $\|f\|_{Lip}$ is the Lipschitz constant of f . The following lemma relates the earth mover's distance to our optimization problem.

Lemma 3. *Consider two posteriors X and X' , and $B > 0$. For any $H \in \mathcal{B}_{\text{expost}}$, $|\mathfrak{J}_X(H) - \mathfrak{J}_{X'}(H)| \leq 2B \cdot d_{EM}(X, X')$. Similarly, if $\text{supp}(\mathcal{X})$ and $\text{supp}(\mathcal{X}')$ are contained in $[1/L, 1 - 1/L]$ for some $L > 0$, for any $H \in \mathcal{B}_{\text{exante}}$, $|\mathfrak{J}_X(H) - \mathfrak{J}_{X'}(H)| \leq 2LB \cdot d_{EM}(X, X')$.*

To use the above results, we need to show $\mathcal{P}_{\sigma,N}$ is a good *covering* for \mathcal{P}_σ so that for all information structures in \mathcal{P}_σ there exists one in $\mathcal{P}_{\sigma,N}$ that has a small earth mover's distance on the posteriors. Here, we design a novel coupling that can bound the earth mover's distance of two posterior predictions by the total variation distance on signal and state space. Given two information structures P and P' on $\Omega \times \mathcal{S}$, the **total variation distance** (TVD) [6] between these information structures is $d_{TV}(P, P') = \frac{1}{2} \sum_{w,s} |P(w, s) - P'(w, s)|$.

Lemma 4. *Given two information structures P and P' with induced posterior predictions X and X' , $d_{EM}(X, X') \leq d_{TV}(P, P')$.*

Lemma 4 upper bound the EMD of posterior predictions by the TVD of information structures. The proof use the maximal coupling of P and P' to couple X and X' using the maximal coupling of P and P' . Then we show the expected difference of X and X' in our coupling can be converted to the total variation distance. Note that the EMD between two random variables is always smaller than the TVD between them, but lemma 4 uses the TVD of the information structures, which can be much smaller than the TVD of posterior.⁴

Lemma 5. *Given $\delta > 0$, N , and σ , for all $P \in \mathcal{P}_\sigma$, there exists $P' \in \mathcal{P}_{\sigma,N}$ so that $d_{TV}(P, P') \leq \frac{1}{N}$. Similarly, given $\rho > 0$, for all $P \in \mathcal{P}_\rho$, there exists $P' \in \mathcal{P}_{\rho,N}$ so that $d_{TV}(P, P') \leq \frac{2-\rho}{N} \cdot \overline{\text{supp}}(\mathcal{P}_{\sigma,N}) = O(N|\mathcal{S}|)$, and $\overline{\text{supp}}(\mathcal{P}_{\rho,N}) = O(N)$.*

Lemma 5 shows $\mathcal{P}_{\sigma,N}$ and $\mathcal{P}_{\rho,N}$ are good coverings for \mathcal{P}_σ and \mathcal{P}_ρ in TVD respectively, by direct computation.

Proof (Proof of theorem 4). Given $\delta, \epsilon > 0$, and $\mathcal{P}_\sigma = \{(\pi, \sigma) : \pi \in [\delta, 1 - \delta]\}$ in example 1, we run our algorithm in theorem 3 on a finite collection of information

⁴ For instance, the TVD of posterior distribution between two ρ -correlated information structures with prior π and π' respectively is 1, but the TVD of information structures is less than $2|\pi - \pi'|$ as shown in lemma 5.

structures $\mathcal{P}_{\sigma,N} = \{(\pi, \sigma) : \pi \in [\delta, 1 - \delta], N\pi \in \mathbb{N}\}$ with an integer $N = \lceil 2B/\epsilon \rceil$. Because $|\mathcal{P}_{\sigma,N}| = O(B/\epsilon)$, the running time is polynomial in B/ϵ and $|\mathcal{S}|$. For the approximation guarantee, given any ex-post bounded $H' \in \mathcal{H}_{\text{expost}}$ and $X \in \mathcal{P}_{\sigma}$, there exists $X' \in \mathcal{P}_{\sigma,N}$ so that

$$d_{EM}(X, X') \leq \frac{\epsilon}{2B}$$

by lemmas 4 and 5. Moreover, by lemma 3 and $N \geq 2B/\epsilon$

$$|\mathfrak{J}_X(H') - \mathfrak{J}_{X'}(H')| \leq 2B \cdot d_{EM}(X, X') \leq \epsilon.$$

Therefore, the optimal H for $\mathcal{P}_{\sigma,N}$ satisfies that

$$\min_{X \in \mathcal{X}} \mathfrak{J}_X(H') \leq \min_{X \in \mathcal{X}'} \mathfrak{J}_X(H') + \epsilon \leq \min_{X \in \mathcal{X}'} \mathfrak{J}_X(H) + \epsilon$$

which completes the proof for the ex-post bounded case. The proof for ex-ante bounded case is similar to the above.

Remark 1. Our FPTAS runs the linear program on a finite subset of an infinite set of information structures and provides approximation guarantees as long as the finite subset is an ϵ -covering of the original set under earth mover's distance by lemma 3. The approach easily extends to any set with efficiently computable small coverings.⁵ Furthermore, the optimization problem 1 can be seen as zero-sum games between a designer choosing a scoring rule and the nature picking an information structure. [20] Hence, if a best response oracle on Θ or an ϵ -covering is available, various no-regret and best response algorithms can be employed to solve it.

However, the main challenge of solving problem 1 lies in understanding the set Θ . Unlike standard optimization problems, Θ is often non-convex, including examples 1 and 3. Consequently, a general efficient algorithm assuming the existence of efficiently computable small coverings or best response oracle may appear vacuous. Instead, we believe our method of finding and proving the small covering property could be a valuable tool for future research in analyzing their own Θ .

4 Simulations

We compare the performance of common scoring rules (section 2.2) within the context of problem 1. We will focus on ρ -correlated information structures (example 3) in the ex-ante setting. This collection of information structures offers a platform to evaluate these scoring rules under the uncertainty of prior (and the experiment to less extent). The other combinations (homogeneous experiment, or ex-post setting) yield comparable outcomes and are deferred to the appendix.

We consider the correlation of signal and the state is either small $\rho = 0.025$ or large $\rho = 0.25$, and the prior π is in $[0.01, 0.99]$. For proper scoring rules, we

⁵ By [32, Theorem 2.2.11], there always exists an ϵ -covering with size $O(1/\epsilon)$.

use 1) a quadratic scoring rule, 2) a log scoring rule, 3) a v -shaped scoring rule with $(a, b, c, x_0) = (-2, 2, 0, 0.5)$ which is optimal when prior is 0.5 as theorem 2, and 4) the piecewise linear scoring rule $H_{\rho, N}$ for $\mathcal{P}_{\rho, N}$ derived by our algorithm in theorem 4 with $N = 50$. Figure 3a shows the associated convex functions for these five proper scoring rules. We include these scoring rules to investigate three questions: How good are classic scoring rules in our setting (the quadratic and log scoring rules)? How does misspecified prior harm the optimal scoring rule in the known prior setting (v -shaped scoring rule)? How does our FPTAS generalizes (piecewise linear scoring rules)?

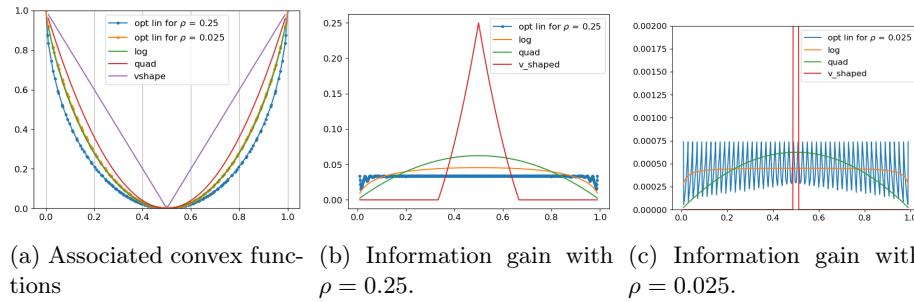


Fig. 3: Figure 3a shows the five associated convex functions in in section 4. By lemma 2 all scoring rules are ex-ante bounded by $B = 1$. Figures 3b and 3c present the information gain on each ρ -correlated information structures. x axis is the prior and y axis is the information gain.

To begin with, we evaluate all our scoring rules with the information gain on each information structure in $\mathcal{P}_{\rho, N'}$ with $N' = 1000 > N$ that serves as a proxy of the infinite collection \mathcal{P}_{ρ} and is a superset of $\mathcal{P}_{\rho, N}$. Figures 3b and 3c present the outcomes where the y -axis represents the information gain, while the x -axis denotes the position of the prior. Note that our piecewise linear scoring rules are optimal for $\mathcal{P}_{\rho, N}$, but only ϵ -optimal for $\mathcal{P}_{\rho, N'}$ with $\epsilon \leq \frac{2-\rho}{N} \leq 0.04$ by theorem 4. The difference of information gains between $\mathcal{P}_{\rho, N}$ and $\mathcal{P}_{\rho, N'}$ measure how well our method generalizes.

Figure 3b shows the information gain on ρ -correlated information structures with large correlation $\rho = 0.25$. The piecewise linear scoring rule for $\mathcal{P}_{\rho, N}$ has the worst-case information gain 0.0341 on the original set $\mathcal{P}_{\rho, N}$, and generalizes well on the superset $\mathcal{P}_{\rho, N'}$ with the worst-case information gain 0.0149. Comparatively, log scoring rule gets 0.0094, the quadratic scoring rule gets 0.0024, and the v -shaped scoring rule gets 0 worst-case information gain on $\mathcal{P}_{\rho, N'}$. The v -shaped scoring rule performs varies significantly: It achieves the highest information gain on information structures with prior centered around 0.5 aligning with the implications of theorem 2, but also gets zero information gain when prior is above 0.7 or below 0.3. This is because when the prior is far away from the vertex (0.5), information structure can have a support contained in a flat

area $(0, 0.5)$ or $(0.5, 1)$ and leading to zero information gain because the Jensen's inequality is tight on affine functions.

Figure 3c shows the information gains when the correlation is small $\rho = 0.025$. The piecewise linear scoring rule has the worst-case information gain $5.77 \cdot 10^{-5}$ compared to $2.76 \cdot 10^{-4}$ under the log scoring rule, $2.48 \cdot 10^{-5}$ on quadratic scoring rule, and 0 under the v -shaped scoring rule. First, the information gains are much small in the small correlation setting, because the posterior is barely move away from prior.⁶ Second, information gains under the piecewise linear scoring rule on $\mathcal{P}_{\rho, N}$ has several periodic peaks each peak. Note that the piecewise linear can be seen as several small v -shaped scoring rules. If an information structure's prior is at a vertex, the information gain is large. However, if the support of the information structure is contained in a flat area, the information gain is near zero. Noteworthy, the number of vertices of our optimal scoring rule equal $N + 1 = 51$ by theorem 3 which is also the number of peaks in the information gains. On the other hand, the log and quadratic scoring rule are strictly convex which do not have any sharp transition between flat area and vertex, so the information gains change smoothly as the priors change. Finally, observe that

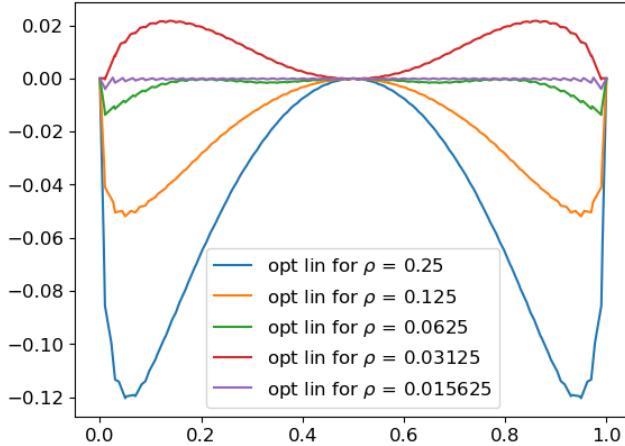


Fig. 4: The difference between log scoring rule and the piecewise optimal scoring rules for ρ -correlated information structures $\mathcal{P}_{\rho, N}$ as ρ ranging from $1/4$ to $1/64$.

the piecewise linear scoring rules are surprisingly close to the log scoring rule in fig. 3a. In fig. 4, we conduct an additional simulations and observe that difference between the piecewise scoring rule for $\mathcal{P}_{\rho, N}$ and the log scoring rule uniformly

⁶ Indeed, we can show that a correlated signal is Blackwell dominated by a correlated signal with larger correlation, and thus has a smaller information gain under any proper scoring rules.

decreases as ρ decreases. This results suggest an interesting potential connection between ρ -correlated information structure and the log scoring rule.

5 Conclusion

We propose the problem of optimal proper scoring rules design when the principal has partial knowledge of an agent's signal distribution. As shown in our simulations, this approach may serve as a benchmark for various scoring rules. We devise efficient algorithms for four principal knowledge settings, and design a novel coupling to bound the earth mover's distance of posterior by the total variation distance of the signal and state space.

Acknowledgments. We thank Grant Schoenebeck and anonymous reviewers for their valuable feedback on restructuring the paper. This work is supported by NSF under award No. IIS-2147187.

References

1. Abernethy, J., Chen, Y., Ho, C.J., Waggoner, B.: Actively purchasing data for learning. ACM EC 2015 (2015)
2. Agarwal, A., Dahleh, M.A., Sarkar, T.: A marketplace for data: An algorithmic solution pp. 701–726 (2019). <https://doi.org/10.1145/3328526.3329589>
3. Alon, T., Dütting, P., Talgam-Cohen, I.: Contracts with private cost per unit-of-effort. In: Proceedings of the 22nd ACM Conference on Economics and Computation. pp. 52–69 (2021)
4. Bechtel, C., Dughmi, S.: Delegated stochastic probing. In: Lee, J.R. (ed.) 12th Innovations in Theoretical Computer Science Conference, ITCS 2021, January 6–8, 2021, Virtual Conference. LIPIcs, vol. 185, pp. 37:1–37:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2021). <https://doi.org/10.4230/LIPICS.ITU.S.2021.37>
5. Cai, Y., Daskalakis, C., Papadimitriou, C.H.: Optimum statistical estimation with strategic data sources. In: Grünwald, P., Hazan, E., Kale, S. (eds.) Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3–6, 2015. JMLR Workshop and Conference Proceedings, vol. 40, pp. 280–296. JMLR.org (2015)
6. Chatterjee, S.: Distances between probability measures. PDF lecture notes, UC Berkeley (2008), archived from the original on July 8, 2008. Retrieved 21 June 2013
7. Chen, Y., Immorlica, N., Lucier, B., Syrgkanis, V., Ziani, J.: Optimal data acquisition for statistical estimation. In: Tardos, É., Elkind, E., Vohra, R. (eds.) Proceedings of the 2018 ACM Conference on Economics and Computation, Ithaca, NY, USA, June 18–22, 2018. pp. 27–44. ACM (2018). <https://doi.org/10.1145/3219166.3219195>
8. Chen, Y., Podimata, C., Procaccia, A.D., Shah, N.: Strategyproof linear regression in high dimensions. In: Tardos, É., Elkind, E., Vohra, R. (eds.) Proceedings of the 2018 ACM Conference on Economics and Computation, Ithaca, NY, USA, June 18–22, 2018. pp. 9–26. ACM (2018). <https://doi.org/10.1145/3219166.3219175>

9. Chen, Y., Zheng, S.: Prior-free data acquisition for accurate statistical estimation. In: Karlin, A.R., Immorlica, N., Johari, R. (eds.) *Proceedings of the 2019 ACM Conference on Economics and Computation, EC 2019*, Phoenix, AZ, USA, June 24-28, 2019. pp. 659–677. ACM (2019). <https://doi.org/10.1145/3328526.3329564>
10. Cummings, R., Ioannidis, S., Ligett, K.: Truthful linear regression. In: Grünwald, P., Hazan, E., Kale, S. (eds.) *Proceedings of The 28th Conference on Learning Theory, COLT 2015*, Paris, France, July 3-6, 2015. JMLR Workshop and Conference Proceedings, vol. 40, pp. 448–483. JMLR.org (2015)
11. Dekel, O., Fischer, F.A., Procaccia, A.D.: Incentive compatible regression learning. *J. Comput. Syst. Sci.* **76**(8), 759–777 (2010). <https://doi.org/10.1016/J.JCSS.2010.03.003>
12. Fleischer, L., Lyu, Y.: Approximately optimal auctions for selling privacy when costs are correlated with data. In: Faltings, B., Leyton-Brown, K., Ipeirotis, P. (eds.) *Proceedings of the 13th ACM Conference on Electronic Commerce, EC 2012*, Valencia, Spain, June 4-8, 2012. pp. 568–585. ACM (2012). <https://doi.org/10.1145/2229012.2229054>
13. Fu, H., Haghpanah, N., Hartline, J., Kleinberg, R.: Optimal auctions for correlated buyers with sampling. In: *Proceedings of the fifteenth ACM conference on Economics and computation*. pp. 23–36 (2014)
14. Ghorbani, A., Zou, J.Y.: Data shapley: Equitable valuation of data for machine learning. In: Chaudhuri, K., Salakhutdinov, R. (eds.) *Proceedings of the 36th International Conference on Machine Learning, ICML 2019*, 9-15 June 2019, Long Beach, California, USA. *Proceedings of Machine Learning Research*, vol. 97, pp. 2242–2251. PMLR (2019)
15. Ghosh, A., Ligett, K., Roth, A., Schoenebeck, G.: Buying private data without verification. In: Babaioff, M., Conitzer, V., Easley, D.A. (eds.) *ACM Conference on Economics and Computation, EC '14*, Stanford , CA, USA, June 8-12, 2014. pp. 931–948. ACM (2014). <https://doi.org/10.1145/2600057.2602902>
16. Ghosh, A., Roth, A.: Selling privacy at auction. In: Shoham, Y., Chen, Y., Roughgarden, T. (eds.) *Proceedings 12th ACM Conference on Electronic Commerce (EC-2011)*, San Jose, CA, USA, June 5-9, 2011. pp. 199–208. ACM (2011). <https://doi.org/10.1145/1993574.1993605>
17. Ghosh, M., Mukhopadhyay, N., Sen, P.K.: *Sequential estimation*, vol. 904. John Wiley & Sons (2011)
18. Gneiting, T., Raftery, A.E.: Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**(477), 359–378 (2007)
19. Grossman, S.J., Hart, O.D.: An analysis of the principal-agent problem. In: *Foundations of insurance economics*, pp. 302–340. Springer (1992)
20. Guo, Y., Hartline, J.D., Huang, Z., Kong, Y., Shah, A., Yu, F.: Algorithmic robust forecast aggregation. *CoRR* **abs/2401.17743** (2024). <https://doi.org/10.48550/ARXIV.2401.17743>
21. Guruganesh, G., Schneider, J., Wang, J.R.: Contracts under moral hazard and adverse selection. In: Biró, P., Chawla, S., Echenique, F. (eds.) *EC '21: The 22nd ACM Conference on Economics and Computation*, Budapest, Hungary, July 18-23, 2021. pp. 563–582. ACM (2021). <https://doi.org/10.1145/3465456.3467637>
22. Hardt, M., Megiddo, N., Papadimitriou, C.H., Wootters, M.: Strategic classification. In: Sudan, M. (ed.) *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, Cambridge, MA, USA, January 14-16, 2016. pp. 111–122. ACM (2016). <https://doi.org/10.1145/2840728.2840730>

23. Li, Y., Hartline, J.D., Shan, L., Wu, Y.: Optimization of scoring rules. In: Pennock, D.M., Segal, I., Seuken, S. (eds.) EC '22: The 23rd ACM Conference on Economics and Computation, Boulder, CO, USA, July 11 - 15, 2022. pp. 988–989. ACM (2022). <https://doi.org/10.1145/3490486.3538338>
24. Liu, Y., Gordon, M., Wang, J., Bishop, M., Chen, Y., Pfeiffer, T., Twardy, C., Viganola, D.: Replication markets: Results, lessons, challenges and opportunities in AI replication. CoRR **abs/2005.04543** (2020)
25. McCarthy, J.: Measures of the value of information. Proceedings of the National Academy of Sciences of the United States of America **42**(9), 654 (1956)
26. Meir, R., Procaccia, A.D., Rosenschein, J.S.: Algorithms for strategyproof classification. Artif. Intell. **186**, 123–156 (2012). <https://doi.org/10.1016/J.ARTINT.2012.03.008>
27. Michael, G., Domenico, V., Michael, B., Yiling, C., Anna, D., Brandon, G., Felix, H., Magnus, J., Yang, L., Charles, T., Juntao, W., Thomas, P.: Are replication rates the same across academic fields? community forecasts from the darpa score programme. Royal Society Open Science **7**(200566) (2020)
28. Murphy, K.P.: Machine learning - a probabilistic perspective. Adaptive computation and machine learning series, MIT Press (2012)
29. Neyman, E., Noarov, G., Weinberg, S.M.: Binary scoring rules that incentivize precision. In: Biró, P., Chawla, S., Echenique, F. (eds.) EC '21: The 22nd ACM Conference on Economics and Computation, Budapest, Hungary, July 18-23, 2021. pp. 718–733. ACM (2021). <https://doi.org/10.1145/3465456.3467639>
30. Nissim, K., Vadhan, S.P., Xiao, D.: Redrawing the boundaries on purchasing data from privacy-sensitive individuals. CoRR **abs/1401.4092** (2014)
31. O'Donnell, R.: Analysis of boolean functions. Cambridge University Press (2014)
32. Panaretos, V.M., Zemel, Y.: An invitation to statistics in Wasserstein space. Springer Nature (2020)
33. Papireddygari, M., Waggoner, B.: Contracts with information acquisition, via scoring rules. In: Pennock, D.M., Segal, I., Seuken, S. (eds.) EC '22: The 23rd ACM Conference on Economics and Computation, Boulder, CO, USA, July 11 - 15, 2022. pp. 703–704. ACM (2022). <https://doi.org/10.1145/3490486.3538261>
34. Perote, J., Perote-Peña, J.: Strategy-proof estimators for simple regression. Math. Soc. Sci. **47**(2), 153–176 (2004). [https://doi.org/10.1016/S0165-4896\(03\)00085-4](https://doi.org/10.1016/S0165-4896(03)00085-4)
35. Roth, A., Schoenebeck, G.: Conducting truthful surveys, cheaply. CoRR **abs/1203.0353** (2012)
36. Savage, L.J.: Elicitation of personal probabilities and expectations. Journal of the American Statistical Association **66**(336), 783–801 (1971)
37. Waggoner, B., Frongillo, R.M., Abernethy, J.D.: A market framework for eliciting private data. In: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (eds.) Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada. pp. 3510–3518 (2015)
38. Zheng, S., Waggoner, B., Liu, Y., Chen, Y.: Active information acquisition for linear optimization pp. 167–176 (2018)