Optimization of Scoring Rules*

YINGKAI LI, Department of Computer Science, Northwestern University.
JASON D. HARTLINE, Department of Computer Science, Northwestern University.
LIREN SHAN, Department of Computer Science, Northwestern University.
YIFAN WU, Department of Computer Science, Northwestern University.

CCS Concepts: • **Theory of computation** → **Algorithmic mechanism design**; • **Applied computing** → **Economics**.

Additional Key Words and Phrases: scoring rules, mechanism design, optimization

ACM Reference Format:
Yingkai Li, Jason D. Hartline, Liren Shan, and Yifan Wu. 2022. Optimization of Scoring Rules. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC '22), July 11–15, 2022, Boulder, CO, USA*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3490486.3538338

This paper provides a framework for a principal to optimize over proper scoring rules. Proper scoring rules are mechanisms that incentivize a forecaster to reveal her true beliefs about a probabilistic state. Proper scoring rules are well studied in theory and widely used in practice. The optimization framework of the paper is relevant for applications that include peer grading, peer prediction, and exam scoring.

Proper scoring rules incentivize a forecaster to reveal her true belief about an unknown and probabilistic state. The principal publishes a scoring rule that maps the reported belief and the realized state to a reward for the forecaster. The forecaster reports her belief about the state. The state is realized and the principal rewards the forecaster according to the scoring rule. A scoring rule is proper if the forecaster's optimal strategy, under any belief she may possess, is to report that belief. Proper scoring rules are also designed for directly eliciting a statistic of the distribution such as its expectation.

Not all proper scoring rules work well in any a given scenario. This paper considers a mathematical program for optimization of scoring rules where (a) the objective captures the incentive for the forecaster to exert effort and (b) the boundedness constraints prevent the principal from scaling the scores arbitrarily. For (a), we focus on a simple binary model of effort where the forecaster does or does not exert effort and with this effort the forecaster obtains a refined posterior distribution from the prior distribution on the unknown state (e.g., by obtaining a signal that is correlated with the state). We adopt the objective that takes the perspective of the forecaster at the point of the decision with knowledge of both the prior and the distributions of posteriors that is obtained by

*This work was supported by NSF CCF-1733860, CCF-1955351, and CCF-1934931. The full version of this paper is available at https://arxiv.org/abs/2007.02905. The order of the authors are certified random. The records are available in https://www.aeaweb.org/journals/policies/random-author-order/search?RandomAuthorsSearch%5Bsearch%5D=4FJdnUr4sE80. Part of the research was done when the fourth author was an undergraduate student at Peking University. Email address: {yingkai.li, hartline, lirenshan2023, yifan.wu}@u.northwestern.edu.

EC '22, July 11–15, 2022, Boulder, CO, USA.

ACM ISBN 978-1-4503-9150-4/22/07.
https://doi.org/10.1145/3490486.3538338

exerting effort. We want a scoring rule that maximizes the difference in expected scores for the posterior distribution and prior distribution. For (b), we impose the ex post constraint that the score is in a bounded range, i.e., without loss, between zero and one.

Results. We solve for the optimal scoring rule for reporting the expectation in single-dimensional space. As we expect for single-dimensional mechanism design problems for an agent with linear utility, the optimal scoring rule fixing any realized state is a step function. To implement this scoring rule, it is sufficient for the designer to know the prior mean instead of the details on the distribution over posteriors. The optimal scoring rule is the same as asking a binary choice question: whether the belief is higher or lower than the mean of the prior. We also demonstrate a first result for prior-independent analysis of scoring rules. Among prior-independent scoring rules for reporting the expectation, the quadratic scoring rule is within a constant factor of optimal.

For multi-dimensional forecasting when the distribution over posterior means and the state space are given explicitly, we provide a polynomial time algorithm that computes the optimal scoring rule. For multi-dimensional forecasting with symmetric distributions, we give an analytical characterization of the optimal scoring rule. For multi-dimensional forecasting without a symmetry assumption, we identify a scoring rule that gives an 8-approximation. This scoring rule can be interpreted as scoring the dimension for which the agent's posterior in the optimal single-dimensional scoring rule gives the highest utility. Equivalently, it can be implemented by letting the agent select which dimension to score and only scoring that dimension (after exerting effort to learn the posterior mean of all dimensions). While optimal mechanisms generally depend on the distribution over posteriors, our approximation bounds are proved for simple mechanismsthat depend only on the prior mean, and do not require detailed knowledge of the distribution over posteriors. For the peer grading example, e.g., it is sufficient to know that the mean grade is $0.8 \in [0, 1]$. In addition, due to the simple form of the approximately optimal scoring rule, even when the designer is ignorant of the prior mean, the designer can estimate it using samples and the expected incentive loss for using the sample estimate is negligible. Finally, we show that the ad-hoc approach of averaging the score of each dimension may have an multiplicative loss in incentives for effort that is linear in the size of the dimension.

Application to Peer Grading. Optimization of scoring rules is a mechanism design problem. A significant challenge for algorithmic methods in mechanism design is a lack of applications to which researchers can readily apply mechanism design results. Optimization of scoring rules, however, has application to peer grading and can be deployed in classrooms where algorithms researchers teach. The questions of this paper were in fact motivated by the failure of classical approaches to scoring rules in this context.

While peer grading may be employed to reduce effort of course staff, a primary concern is in improving learning outcomes. For peers to learn from peer reviewing they must be incentivized to put in effort, i.e., the peer reviews themselves must be graded. One way to algorithmically grade peer reviews is to compare the peer's marks to ground truth marks provided by the teaching staff. Specifically, a peer can be asked to review the submission and forecast the true marks.

If the grading rubric has multiple elements (denoted by n), the natural approach from the literature would be to score each dimension separately and then take the sum. In contrast, the optimal multi-dimensional rule is not the sum over separate rules but the maximum over separate scoring rules. For a prior such that independently for each dimension, the signal reveals the state with probability $\frac{1}{n}$, these two are significantly different. Specifically, the incentives for effort for the separate scoring rule is $O(\frac{1}{n})$ while the incentives for effort for optimal scoring rule is $O(1)$. Thus optimal scoring rule can be unboundedly better than separate scoring rule.