Peer Neighborhood Mechanisms: A Framework for Mechanism Generalization

Adam Richardson, Boi Faltings

École Polytechnique Fédérale de Lausanne
Artificial Intelligence Laboratory

Abstract

Peer prediction incentive mechanisms for crowdsourcing are generally limited to eliciting samples from categorical distributions. Prior work on extending peer prediction to arbitrary distributions has largely relied on assumptions on the structures of the distributions or known properties of the data providers. We introduce a novel class of incentive mechanisms that extend peer prediction mechanisms to arbitrary distributions by replacing the notion of an exact match with a concept of neighborhood matching. We present conditions on the belief updates of the data providers that guarantee incentive-compatibility for rational data providers, and admit a broad class of possible reasonable updates.

1 Introduction

With the advent of machine learning in data analysis, governance, recommendation systems, and commercial products, there is an increasing need for abundant and accurate data with which to train these models. In some contexts, acquiring the necessary data can be expensive, slow, or require access to private information like medical records. For this reason, there has been ample recent research on *crowdsourcing* for data acquisition. The primary goal of such research is to design *incentive mechanisms*, or payment schemes, that incentivize independent, rational Agents to provide accurate data to a Center.

For many years, the gold standard for such incentive mechanisms has been the class of *Peer prediction* mechanisms. Peer prediction mechanisms operate in the absence of any a priori baseline evaluation of the quality of reports. Generally, a Peer prediction mechanism works by comparing an Agent's report to a randomly selected report from another Agent, called a Peer, during the same data collection period. By examining correlations between reports, such mechanisms can induce desirable equilibria. In most settings, the Center wants the Agents to truthfully report some observation about a real world phenomenon. The major disadvantage of such mechanisms that they are generally only applicable for eliciting data from categorical distributions because they rely on a notion of report matching, i.e. the Agent and Peer report a sample from the same category.

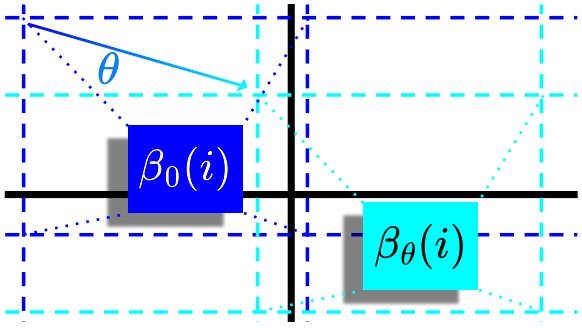

Figure 1: An example of a partition family on $\mathbb{R}^2$ with θ representing translations of the bins: $\beta_0(i)$ transforms into $\beta_\theta(i)$. Partition families are used to construct Peer Neighborhoods.

We refer to such mechanisms as *Peer Consistency* mechanisms. Because they are restricted to categorical distributions, they eschew any notion of locality among the categories. This disadvantage presents a clear theoretical roadblock for applying such mechanisms to arbitrary distributions, since continuous random variables are only understandable through measuring local neighborhoods.

Our work presents a novel framework for extending Peer Consistency mechanisms to arbitrary distributions. We call such extensions *Peer Neighborhood* mechanisms. To our knowledge, this is the only work that does so without assuming that the Center possesses a priori knowledge of properties of the Agents or of the underlying distribution. We only assume that Agents are rational and that they follow some reasonable belief update conditions, which we will show admit a broad class of updates. By analyzing an extension of the Peer Truth Serum, we prove that it can admit truthful Bayes-Nash Equilibria on the ex-ante game produced by the mechanism.

1.1 Related Work

Our work builds on a long line of advancements in incentive mechanism design based on Peer Consistency. The most basic such mechanism is Output Agreement, which simply pays Agents a constant reward when their reports are identical (Von Ahn and Dabbish 2004; Waggoner and Chen

2014). The class of Peer Consistency mechanisms was later broadened with the concept of Proper Scoring Rules (Winkler et al. 1996; Miller, Resnick, and Zeckhauser 2005). Other notable examples are multi-task peer prediction mechanisms, such as the Correlated Agreement mechanism (Dasgupta and Ghosh 2013; Shnayder et al. 2016). There is also significant literature on Bayesian mechanisms, starting in economics (d'Aspremont and Gérard-Varet 1979; McAfee and Reny 1992) with trying to elicit private, correlated information. Subsequently, the Bayesian Truth Serum was developed for eliciting subjective private information (Prelec 2004). The distinguishing feature of these Bayesian mechanisms compared to classical Peer Consistency is the requirement that the reports include not merely a data sample, but also the Agent's estimate of the probability of the sample. While the additional information provided to the Center has allowed for extensions such as Robust BTS (Witkowski and Parkes 2012; Radanovic and Faltings 2014), but it comes at great additional cost in practical terms and in the complexity of the mechanism. For this reason, we focus on Peer Consistency mechanisms which are *minimal*, meaning that they do not require Agents to report anything other than the sample.

We primarily build on the work of Radanovic et al. on the Peer Truth Serum (PTS), which expands on the Output Agreement mechanism by considering different update conditions for Agent beliefs (Radanovic, Faltings, and Jurca 2016). The Output Agreement mechanism can be characterized as incentive-compatible with respect to the *self-dominating* update condition. The PTS is incentive-compatible with respect to a much broader class of updates, characterized by the *self-predicting* condition. Like the Peer Consistency mechanisms before it, the PTS is only applicable to categorical distributions. It has long been a goal of incentive mechanism design to uncover general concepts that allow for application over arbitrary distributions. One attempt is the Logarithmic Peer Truth Serum (LPTS), which does so by assuming a locality structure among the Agents so that more localized Peers have more similar statistical properties (Radanovic and Faltings 2015), and the Personalized Peer Truth Serum, which extends the LPTS for subjective private data (Goel and Faltings 2020). Finally, the work of (Chen, Shen, and Zheng 2020) and (Kong and Schoenebeck 2019) consider mechanisms which reward Agents according to the mutual information between reports, but assume that the distribution can be parameterized as an element of a known family of distributions. We see that in all cases, the ability to extend Peer Consistency concepts to arbitrary distributions relies on a priori assumptions about the structure of the distributions or the Agent properties. Our work eliminates these assumptions by considering a novel mechanism extention concept, which we call Peer Neighborhoods, and identifying a class of Agent updates that make such mechanisms incentive-compatible.

2 Model

In a crowdsourcing setting there is a Center that wishes to learn an arbitrary distribution Φ, which we call the *true distribution*, but the Center can't probe this distribution in a meaningful way. The Center tries to learn the distribution by collecting *reports* from a set of independent, self-interested Agents who can sample Φ to produce an *observation*. Because the Agents are self-interested, they must be *incentivized* to produce reports that help the Center learn Φ. The incentive an Agent experiences is a personal utility function that depends on the Agent's reporting strategy and a set of *beliefs* the Agent has about the setting, such as the distribution Φ and the reporting strategies of other Agents. Agents always act rationally, so they will adopt the reporting strategy which maximizes their expected utility under their current beliefs. The Center's goal is to choose a payment function which dispenses utility to the Agents in exchange for reports, such that Agents will be incentivized to adopt "good" reporting strategies. In our case, we seek *truthful* reporting, meaning Agents report their observations.

In the setting we consider, the Agent does not have a static set of beliefs. We refer to the belief of the Agent about the true distribution before making the observation as π, the *prior belief*. After making an observation o, the prior is updated to π_o, the *posterior belief*. Prior to the data collection period, the Center also has its own belief about the true distribution, R, which it makes public to the Agents. We refer to R as the *public prior*. It is assumed that these probability measures are on a shared measurable space (Ω, Σ). When discussing arbitrary distributions, we will assume that $\Omega = \mathbb{R}^d$ for some d, and $\Sigma = \mathcal{B}(\Omega)$, the Borel sets of Ω.

The game sequence is as follows: a data collection period begins with a set of Agents possessing prior beliefs. Each Agent then makes an observation and updates their prior to a posterior belief. Each Agent then makes a report to maximize its expected payment according to the mechanism, which is public knowledge. At the end of the data collection period, the Center will have received a set of reports $\{r\}$. For each report, it randomly picks a Peer report and performs some comparison between the two reports. This informs the payment for the report.

3 Peer Neighborhood Mechanisms

3.1 Peer Consistency

In the original setting for Peer Consistency mechanisms, Φ is a categorical distribution, and the mechanisms pay an Agent when its report matches with a randomly selected Peer report. We formalize this concept:

Definition 3.1. A *Peer Consistency* mechanism is a mechanism which assumes some public prior R with categorical distribution, takes a report r from an Agent and a report rr from a randomly chosen Peer, and pays the Agent $\tau_R(r, rr) = f(rr) + s_R(r) * \mathbb{1}_{r=rr}$ where f depends only on rr, and s_R is a non-negative *scoring function* which depends on the distribution R.

When discussing the incentives of a Peer Consistency mechanism, this is typically in regards to some *update condition*:

Definition 3.2. Given a prior probability measure π, an observation o, and a posterior probability measure π_o, an *update condition* is a predicate $S(\pi, \pi_o)$. We call $S^*(\pi, \pi_o)$ the *natural update condition* for some scoring function s_π as $\forall x \neq o : \pi_o(o) * s_\pi(o) > \pi_o(x) * s_\pi(x)$.

A notable example that we will use is the *self-predicting* update condition for the Peer Truth Serum (Radanovic, Faltings, and Jurca 2016). This is the natural update condition and is given by $s_\pi(r) = \frac{1}{\pi(r)}$, so the condition is $\forall x \neq o : \frac{\pi_o(o)}{\pi(o)} > \frac{\pi_o(x)}{\pi(x)}$. Assumptions about Agent behaviors can be embedded in update conditions in a way that reflects an informal notion of probabilistic reasonableness. For example, the self-predicting condition satisfies Bayes's Rule.

Definition 3.3. Given a prior probability measure π and an observation $o \in \Omega$, an *update process* is a function $\mathcal{U}(\pi, o) = \pi_o$. We say an update process satisfies an update condition S if $\forall \omega \in \Omega : S(\pi, \mathcal{U}(\pi, \omega))$ is true.

Definition 3.4. A Peer Consistency mechanism with public prior R is *incentive-compatible* with respect to an update condition S if an Agent, with prior $\pi = R$ and an update process which satisfies S, believes that for any observation o, their expected payment, with the Peer reports distributed according to the posterior, is maximized by truthfully reporting o.

Proposition 3.5. *A Peer Consistency mechanism is incentive-compatible with respect to the natural update condition.*

The proof is in the Appendix of the long version of this paper (Richardson and Faltings 2023).

3.2 Partition Spaces

Peer Neighborhood mechanisms place a layer of abstraction on top of Peer Consistency mechanisms to introduce a notion of locality. They do so by considering a *family of partitions* of the space of reports. A standard approach to applying a Peer Consistency mechanism to a continuous distribution, such as a Gaussian distribution, is to pick some fixed discretization, or partition. Each *bin* of the partition corresponds to a category for the mechanism. A mechanism with a truthful equilibrium for some update condition over a discrete distribution would then have a bin-truthful equilibrium over this continuous distribution if the Agent's belief update satisfies the update condition with respect to the bin-categories. An Agent would have an equal incentive to report any value inside the bin containing the truthful report.

By considering a family of partitions rather than a single one, the incentives can be refined. Consider a Gaussian distribution partitioned into integer length bins $(n, n+1]$. An incentive-compatible Peer Consistency mechanism would then incentivize any report with the correct integer value. If a second partition is introduced with bins $(n + \frac{1}{2}, n + \frac{3}{2}]$, and the Agent satisfies the update condition for both partitions, the report is now incentivized to be within an interval of length $\frac{1}{2}$, corresponding to the intersection of the truthful bins from each partition. We will see that if the partition family is constructed correctly, this intersection can be refined to contain only the truthful report.

Definition 3.6. A *partition family* T is a function which maps a parameter $\theta \in \Theta$ to a partition, which is a countable set of measurable bins β that are disjoint and cover Ω: $T(\theta) = \{\beta_\theta(i)\}_{i \in \mathbb{Z}^*_\theta}$ where $\mathbb{Z}^*_\theta \subseteq \mathbb{Z}$ and $\beta_\theta(i) \in \mathcal{B}(\Omega)$ such that $\forall \theta$, $\bigcup_{i \in \mathbb{Z}^*_\theta} \beta_\theta(i) = \Omega$ and $\forall i \neq j$, $\beta_\theta(i) \cap \beta_\theta(j) = \emptyset$

A simple example of a partition family over $\mathbb{R}^2$ is shown in Figure 1. The Center must have some way of selecting a partition from the family. We have the Center pick the partition randomly according to some distribution, which we call the *partition selection distribution*.

Definition 3.7. The *partition selection distribution* is given by a probability measure Ψ over some measurable space (Θ, Σ), where Θ is the set of parameters for the partition family and Σ is some σ-field over Θ. Without loss of generality, let Ψ be supported on Θ. We call the pair (T, Ψ) the *partition space*.

For ease of reading we will often use the *bin selection function* to identify the bin that contains a particular point:

Definition 3.8. The *bin selection function* with respect to a partition family $T(\theta)$ is a function $\mathbb{X}_\theta : \Omega \to \mathbb{Z}^*_\theta$ such that $\mathbb{X}_\theta(z) = i$ if and only if $z \in \beta_\theta(i)$.

The Bin Selection Function is well-defined as a result of the bins of each partition being disjoint and covering Ω. It is important for the Center's implementation that this function be computable.

If an Agent is to have a strictly truthful incentive, there must be a non-zero probability of any other report failing to match with the truthful report under the partition family. We call this *point-isolating*:

Definition 3.9. A partition space (T, Ψ) is *point-isolating* over R if: $\omega_1 \neq \omega_2$ in the support of $R \Rightarrow \Psi(\{\theta : \mathbb{X}_\theta(\omega_1) \neq \mathbb{X}_\theta(\omega_2)\}) > 0$.

Reports with a matching probability of 0 in R can result in infinite payments under some Peer Consistency mechanisms, such as the Peer Truth Serum. To avoid degenerate payments, we impose the following condition on the partition family:

Definition 3.10. A partition space (T, Ψ) is *bin-supported* over R if $\forall \omega \in \Omega : \Psi(\{\theta : R(\beta_\theta(\mathbb{X}_\theta(\omega))) = 0\}) = 0$.

In simpler terms, for any possible report, the probability of selecting a partition with a 0 R-probability bin containing the report is 0 in Ψ.

Proposition 3.11. *$\forall R$, $\exists (T, \Psi)$ such that (T, Ψ) is point-isolating and bin-supported over R*

The proof is in the long version.

3.3 The Mechanism Extension

Now, we can introduce the Peer Neighborhood mechanism extension. First we must modify the probability measures:

Definition 3.12. Let π be a probability measure on $(\Omega, \mathcal{B}(\Omega))$. Let $T(\theta)$ be a partition of Ω. Then for $i \in \mathbb{Z}^*_\theta$, let the *partitioned probability measure* $\pi^\theta(i) = \pi(\beta_\theta(i))$.

We can then extend any Peer Consistency mechanism as follows:

Definition 3.13. Given some Peer Consistency mechanism with payment function τ, we define the bin-extension payment function with respect to some partition $T(\theta)$ as $\tau^\theta_R(r, rr) = \tau_{R^\theta}(\mathbb{X}_\theta(r), \mathbb{X}_\theta(rr))$. Then given some partition selection distribution Ψ such that (T, Ψ) is point-isolating and bin-supported over R, the *Peer Neighborhood*

extension mechanism pays according to:

$$\tau_R^{\Psi}(r, rr) = \mathbb{E}_{\theta\sim\Psi}[\tau_R^{\theta}(r, rr)] \quad (1)$$

3.4 Incentive-Compatibility

Given some Peer Consistency mechanism with scoring function s, we wish to discover an update condition $S^{(T,\Psi)}$ for which the associated Peer Neighborhood extension mechanism is incentive-compatible. We suggested earlier that the incentivized report region could be refined as long as the Agent is incentivized to be bin-truthful for all the partitions, so the most straightforward condition is that S is satisfied with probability 1 in Ψ.

Definition 3.14. Given a prior π, a posterior π_o, a partition space (T, Ψ), and a Peer Consistency mechanism with scoring function s, the *Partition-Invariant* (PI) update condition $S_{PI}^{(T,\Psi)}$ takes the form:

$$\Psi(\{\theta : S^*(\pi^\theta, \pi_o^\theta)\}) = 1 \quad (2)$$

Proposition 3.15. *Given a Peer Consistency mechanism with payment function τ and scoring function s, and given (T, Ψ) point-isolating over R, the Peer Neighborhood extension mechanism $\tau_R^{\Psi}(r, rr)$ is incentive-compatible with respect to the update condition $S_{PI}^{(T,\Psi)}$.*

The proof is in the long version. While this update condition clearly guarantees incentive-compatibility of the Peer Neighborhood extension mechanism, we will see that it is stronger than necessary. We present a more relaxed update condition:

Definition 3.16. Given a prior π, a posterior π_o, a partition space (T, Ψ), and a Peer Consistency mechanism with scoring function s, the *Partition-Expected* (PE) update condition $S_{PE}^{(T,\Psi)}$ takes the form:

$$\forall x \neq o : \mathbb{E}_{\theta\sim\Psi}[\pi_o^\theta(\mathbb{X}_\theta(o)) * s_{\pi^\theta}(\mathbb{X}_\theta(o))] > \mathbb{E}_{\theta\sim\Psi}[\pi_o^\theta(\mathbb{X}_\theta(x)) * s_{\pi^\theta}(\mathbb{X}_\theta(x))]$$

Proposition 3.17. *Given a Peer Consistency mechanism with payment function τ and scoring function s, and given (T, Ψ) point-isolating over R, the Peer Neighborhood extension mechanism $\tau_R^{\Psi}(r, rr)$ is incentive-compatible with respect to the update condition $S_{PE}^{(T,\Psi)}$.*

The proof is in the long version. Furthermore, we show that the PE condition is a relaxed form of the PI condition, in that any update process which satisfies PI also satisfies PE:

Lemma 3.18. *Given a partition space (T, Ψ) that is point-isolating over R, and a Peer Consistency mechanism with scoring function s, any update process $\mathcal{U}(R, o)$ which satisfies the PI extended update condition $S_{PI}^{(T,\Psi)}$ also satisfies the PE extended update condition $S_{PE}^{(T,\Psi)}$.*

The proof is in the long version.

4 Analysis of Update Processes

We have constructed a framework that extends Peer Consistency mechanisms to arbitrary distributions, but the crux of this extension is the Partition-Expected update condition, which is necessarily more restrictive than the natural update condition for the underlying discrete mechanism. We will examine what types of update processes satisfy this condition, but we must first address a practical concern which will further restrict update processes, namely whether or not an update process is consistent with convergence of the posterior to the true distribution.

4.1 Update Convergence

When an Agent makes on observation and computes a posterior according to some update process, that process should generally bring the Agent's belief closer to the true distribution. With finite observations, it is always possible that an Agent can observe a very unlikely sequence, leading to a bias in the posterior. But in the limit of infinite observations, the posterior should converge to the true distribution. We then wish to describe update processes which can be performed iteratively to converge to the true distribution.

Definition 4.1. Consider a sequence of update processes $\mathcal{U}_i$ for all $i \in \mathbb{Z}_+$. The sequence is *convergent* if, when the sequence $\{\mathcal{U}_i\}$ is applied iteratively to a sequence of i.i.d. observations $\{o_i\}$ sampled from the true distribution, the sequence of posteriors converges in distribution to the true distribution.

In order to get a better grasp on such update processes, we will restrict ourselves to a particular type of update process, which we call *additive*:

Definition 4.2. An update process $\pi_o = \mathcal{U}(\pi, o)$ is *additive* if $\pi_o = (1-\alpha)\pi + \alpha K_o$ where K_o is a probability measure which we call the *update kernel*, and $\alpha \in (0, 1)$.

Often we will refer to an update process of this form simply by referring to the update kernel. The Agent picks $(1-\alpha)$ to represent the Agent's confidence in the accuracy of its prior.

Suppose an Agent were to observe a sequence of i.i.d. samples from the true distribution and sequentially update. It would be unreasonable for the Agent to update to a different posterior depending on the order of the sequence of samples, so the update kernels should be given equal weight. Given some additive update with α_1, the next update must then have $\alpha_2 = \frac{\alpha_1}{1+\alpha_1}$. A simple choice for α_1 would be $\frac{1}{k}$ for some positive integer k, so $\alpha_2 = \frac{1}{k+1}$. This process of decreasing α like $\frac{1}{n}$ can be applied iteratively, and we refer to this as an *additive update sequence*:

Definition 4.3. An *additive update sequence* is given by $\mathcal{U}_k(\pi, o) = \frac{k}{k+1}\pi + \frac{1}{k+1}K_o$. For a sequence of n observations $\{o_i\}$, the final posterior is given by: $\pi_{\{o_i\}} = \mathcal{U}_{k+n-1}(\mathcal{U}_{k+n-2}(\ldots\mathcal{U}_k(\pi, o_1)\ldots, o_{n-1}), o_n) = \frac{k}{k+n}\pi + \frac{1}{k+n}\sum_{i=1}^{n} K_{o_i}$.

First we show that the convergence of this update sequence depends only on the update kernels. We say this update sequence is *prior agnostic*:

Lemma 4.4. *The additive update sequence converges in distribution to the same distribution as the average of the kernels:* $\frac{1}{n}\sum_{i=1}^{n} K_{o_i} \xrightarrow{d} X \iff \frac{k}{k+n}\pi + \frac{1}{k+n}\sum_{i=1}^{n} K_{o_i} \xrightarrow{d} X$.

The proof is in the long version. We now address the structure of the update kernel K. We will consider two types of kernels, the first is a simple point mass. We call this the *Empirical Update*:

Definition 4.5. The *Empirical Update* is the additive update where $K_o(A) = \mathbb{1}_{o \in A}$.

Proposition 4.6. *The Empirical Update sequence is convergent.*

The proof is in the long version. The second type of kernel we wish to address is a kernel with a continuous CDF. We call such updates *continuous*:

Definition 4.7. An additive update is *continuous* if the CDF F_K of the update kernel K is continuous.

We show that an additive continuous update sequence is convergent if the sequence of kernels satisfies a condition on the partial sums of their concentrations around the observed samples:

Theorem 4.8. *Let* $O_n = \{o_i\}_{i \in [1,n]}$ *be a sequence of i.i.d random variables distributed with CDF* $F(x)$*. Let* K_{o_i} *be a continuous update kernel. Define* $H_n(x) = \frac{1}{n}\sum_{i=1}^{n} F_{K_{o_i}}(x)$*. Consider the random variables* $Y_i = |X_i - o_i|$ *where* X_i *is distributed according to* K_{o_i}*. Define* $C_i(\epsilon) = P(Y_i \geq \epsilon)$*. If* $\forall \epsilon > 0 : \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} C_i(\epsilon) = 0$*, then* $H_n \xrightarrow{d} F$.

The proof is in the long version. The convergence of the additive continuous update sequence with kernels satisfying this condition follows directly from Lemma 4.4.

We present a very simply condition on the kernels which will satisfy Theorem 4.8. The kernels merely need to have bounded support, with that bound converging to 0.

Corollary 4.9. *Let* $\Delta_i = \langle \delta_{i,1}, \delta_{i,2}, \ldots, \delta_{i,d} \rangle$ *with* $\delta_{i,j} > 0$ *and* $\lim_{i\to\infty} \delta_{i,j} = 0$*. Let* $A_i = [o_i - \Delta_i, o_i + \Delta_i]$ *Suppose* $K_{o_i}(A) = 1$*. Then an additive continuous update sequence with these kernels is convergent.*

The proof is in the long version.

4.2 Satisfying the Update Conditions

We now analyze update processes which satisfy our extended update conditions PI and PE. Whether or not these conditions are satisfied depends heavily on the choice of scoring function s. We choose to focus our attention to the Peer Neighborhood extension of the PTS, commonly considered the canonical example of a Peer Consistency mechanism. We will then refer to the extension as the *Peer Truth Neighborhood Extension* mechanism.

Definition 4.10. The *Peer Truth Neighborhood Extension* (PTNE) mechanism is the Peer Neighborhood extension of the Peer Consistency mechanism with scoring function $s_R(r) = \frac{c}{R(r)}$ where c is a positive constant, known as the Peer Truth Serum.

The natural update condition for the PTS is $\forall x \neq o : \frac{\pi_o(o)}{\pi(o)} > \frac{\pi_o(x)}{\pi(x)}$, known as the *self-predicting* update condition.

We first prove that the Empirical Update satisfies PI for the PTNE, and therefore also PE:

Theorem 4.11. *Given a partition space* (T, Ψ) *that is point-isolating and bin-supported over* R*, the Empirical Update process satisfies* $S_{PI}^{(T,\Psi)}$ *for the PTNE mechanism.*

The proof is in the long version. The Empirical Update is perfectly reasonable and is used quite frequently in modeling, but an Agent may want to make a continuous update as they may be unsure that their measurement of the true distribution is precise. But with additive continuous updates, whether or not they satisfy our update conditions depends on the structure of the partition space. To simplify analysis, we focus on partition spaces with a high degree of symmetry, which we call *regular*:

Definition 4.12. A *regular* partition space (T, Ψ) over a fundamental set $\Omega = \mathbb{R}^d$ is one in which each bin is a rectangular prism with side lengths $\{l_i\}_{i \in [1,d]}$, i.e. $\forall \theta \in \Theta$, $T(\theta) = \{\bigotimes_{i=1}^{d} [l_i * (n_i - \frac{1}{2}) + \theta_i, l_i * (n_i + \frac{1}{2}) + \theta_i) \forall n_i \in \mathbb{Z}\}$ and $\Theta = \bigotimes_{i=1}^{d} [0, l_i)$, with Ψ uniform over Θ.

This partition space is clearly point-isolating over any R as it is point isolating for all $\omega \in \Omega$. We will assume that R is such that this partition space is bin-supported over R.

4.3 Bin Edge Conditions

Let us consider some regular partition space. Let each bin have dimensions $L = \langle l_1, l_2, \ldots, l_d \rangle$. To simplify the notation, we will say the set $[-L, L) = \bigotimes_{i=1}^{d} [-l_i, l_i)$ We define the Bin Function $B : \mathbb{R}^d \to \{0, 1\}$ to be:

$$B(\omega) = \begin{cases} 1 & \omega \in [-\frac{L}{2}, \frac{L}{2}) \\ 0 & \text{otherwise} \end{cases}$$

so the Bin Function is just an indicator for a bin centered at 0. Assume that an agent with prior and posterior π and π_o respectively has PDFs f_π and f_{π_o}. It's not necessary that such PDFs exist, but we make this assumption for ease of presentation. Let us define the overhead $\sim$ to be the operator such that for a function f, $\widetilde{f}(x) = (f \circledast B)(x)$, where $\circledast$ is the convolution operator. Then $\widetilde{f_\pi}(x)$ is just the prior probability of a sample landing in a bin centered at x, and same for the posterior $\widetilde{f_{\pi_o}}(x)$. These functions can be computed only using CDFs, but it is simpler to define them this way. The quantities we are concerned with regarding the PI and PE conditions for the PTNE mechanism are the ratios $Q(x) = \frac{\widetilde{f_{\pi_o}}(x)}{\widetilde{f_\pi}(x)}$. The expected payment for reporting x is simply $\widetilde{Q}(x)$. If the update process is additive continuous, then Q and $\widetilde{Q}$ are continuous.

The PI condition gives us the following constraints: Let $N = \langle n_1, n_2, \ldots, n_d \rangle$ where $n_i \in \mathbb{Z}$ and $N \neq 0$. Then $\forall x \in (o - \frac{L}{2}, o + \frac{L}{2}] : Q(x) > Q(x + N * L)$ where $*$ is element-wise multiplication. Let $Q_o(x) = Q(o + x)$. From

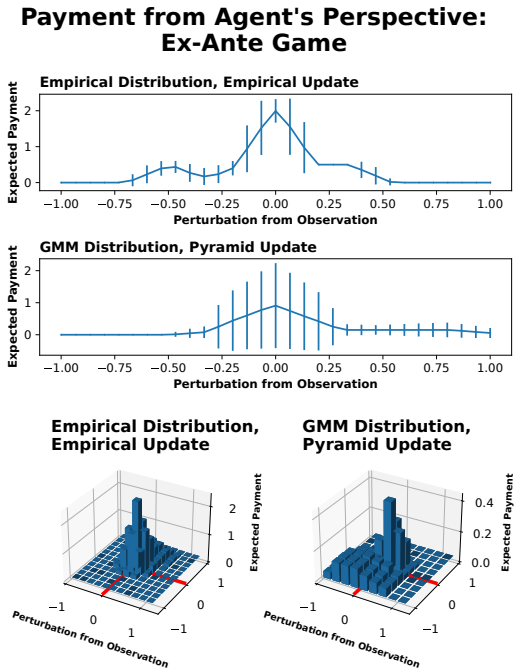

Figure 2: Expected payments for reports perturbed from the observation, computed over an Agent's posterior. Error bars are one standard deviation. In the 2D figures, red lines show the location of the maximum expected payment.

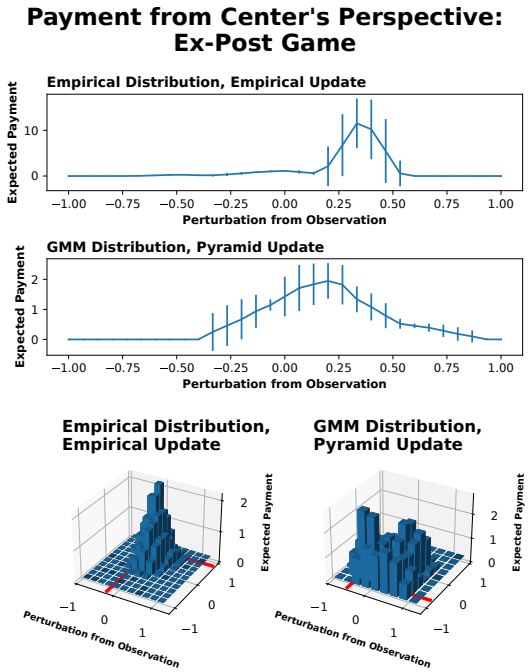

Figure 3: Expected payments for reports perturbed from the observation, computed over truthful Peer reports. Error bars are one standard deviation. In the 2D figures, red lines show the location of the maximum expected payment.

the continuity of Q, it follows that for all $i \in [1, d]$ and all $\delta_i \in [-\frac{l_i}{2}, \frac{l_i}{2}]$:

$$Q_o(\delta_1, \ldots, -\frac{l_i}{2}, \ldots, \delta_d) = Q_o(\delta_1, \ldots, \frac{l_i}{2}, \ldots, \delta_d) \qquad (3)$$

We see that these are equalities on every pair of opposing points on the boundary of the bin centered at o.

The PE condition simply constrains o to be the global maximum of $\widetilde{Q}$. As long as the continuous update kernel has mass at o and has sufficiently bounded support, if o is a local maximum of $\widetilde{Q}$, then it will be the global maximum. We'll discuss what sufficiently bounded means later. We write the PE constraint:

$$\nabla_x \widetilde{Q}|_{x=o} = 0, \qquad \nabla_x^2 \widetilde{Q}|_{x=o} < 0.$$

From the continuity of $\widetilde{Q}$ we obtain conditions that are much less restrictive than for PI. Let L_{-i} be the vector L with entry l_i removed, and Δ_i be the vector of δ_js with entry δ_i removed. Then $\forall i \in [1, d]$:

$$\int_{-\frac{L_{-i}}{2}}^{\frac{L_i}{2}} Q_o(\delta_1, \ldots, -\frac{l_i}{2}, \ldots, \delta_d)\, d\Delta_i$$
$$= \int_{-\frac{L_{-i}}{2}}^{\frac{L_i}{2}} Q_o(\delta_1, \ldots, \frac{l_i}{2}, \ldots, \delta_d)\, d\Delta_i \qquad (4)$$

We see that rather than having an equality for every pair of opposing points on the boundary of the bin centered at o, we have a single equality for each opposing boundary surface of the bin. This is equivalent to the constraints for PI in one dimension, since the opposing boundary surfaces are just a single pair of points, but in higher dimensions it is much less constraining. We also see that a continuous update kernel has "sufficiently bounded support" if it has support within $(o - L, o + L]$. From now on we will refer to such an update kernel as *bin-bounded*.

Failing PI We will first show that it is impossible in general for a bin-bounded continuous update kernel to satisfy both PI and for the associated additive update sequence to be convergent in dimensions higher than one. We will show the proof for two dimensions, but the same argument applies to higher dimensions.

Lemma 4.13. *Given a regular partition space on $\mathbb{R}^2$, let each bin have dimensions $L = \langle l_1, l_2 \rangle$. Let $\Delta = \langle \frac{l_1}{2}, \frac{l_2}{2} \rangle$, and $A = [z - \Delta, z + \Delta]$. There is a prior π such that a continuous update kernel must have bounded probability on A: $K_o(A) < x < 1$ in order to satisfy PI for the PTNE mechanism.*

The proof is in the long version. With this we can prove that a continuous update kernel cannot allow for an additive update sequence that is convergent:

Theorem 4.14. *Given a regular partition space on $\mathbb{R}^2$, there is a prior π and true distribution Φ such that a continuous update kernel cannot satisfy PI for the PTNE mechanism and admit an additive update sequence that is convergent.*

The proof is in the long version.

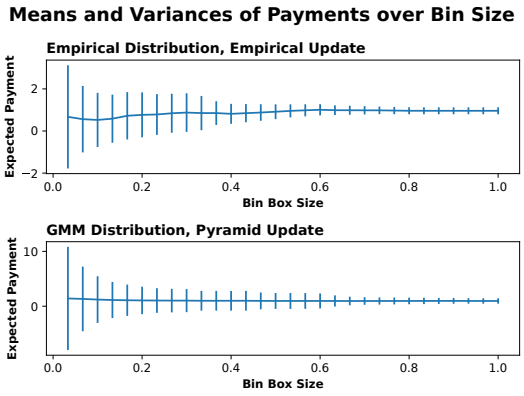

Figure 4: Smaller bins produce a larger variance in payments. Error bars are one standard deviation squared.

Satisfying PE We will now show that it is always possible to construct a sequence of continuous update kernels that satisfy both PE and are convergent. We will construct these explicitly. First we will restrict our construction so that all the probability of the kernel is within a bounded region $A = [x - \Delta, x + \Delta]$ which contains the observation point o, and where Δ can be arbitrarily small. From Corollary 4.9, we find that by allowing the sequence Δ_i to converge to 0, this update sequence will be convergent. We are left only to show that the kernel construction satisfies PE:

Theorem 4.15. *Given a regular partition space on $\mathbb{R}^d$, for any prior π, there exists a continuous update kernel that satisfies PE for the PTNE mechanism and is arbitrarily bounded around a point o.*

The proof is in the long version. It involves constructing update kernels with PDFs as hyper-pyramids with a peak at o and a base at $[x - \Delta, x + \Delta]$ for some arbitrary positive $\Delta < L$ where L is the dimensions of the bins. The proof demonstrates the existence of x such that $o \in [x-\Delta, x+\Delta]$ and the kernel satisfies the bin edge conditions in Equation 4. The proof goes further by suggesting a method for computing this x, which we take advantage of in our simulations.

5 Simulations

We conduct simulations using the PTNE mechanism to demonstrate the accuracy and stability of the incentives in settings with finite data for constructing models and finite peer reports. We use artificially generated data to form the true and public distributions, which can then be used to analyze expected payments and actual payments from samples. We present two data models: 1) an Empirical distribution constructed by taking finite samples with randomized frequencies, and 2) a continuous distribution constructed as a weighted sum of Gaussian distributions, or a Gaussian Mixture Model (GMM). For the first model, Agents use the Empirical Update, while for the second they update using Pyramid kernels as described in the proof of Theorem 4.15. In all cases, the partition space is regular. We provide details of the simulation parameters in the long version (Richardson and Faltings 2023).

5.1 Report Perturbation

We simulate the expected payments for an Agent reporting a point that is a perturbation of the observation, meaning the payment for the observation itself is at 0. Figure 2 shows the expected payments computed from the perspective of the Agent over the posterior. The error bars show the standard deviation with respect to the Partition Selection distribution. We observe that the Agent believes their payment will be maximized by truthfully reporting the observation, as expected from the theory. Figure 3 shows the same expected payments, but this time computed over a set of truthful Peer reports collected by the Center. The expected payment from the Center's side is not necessarily maximized at the observation point. Since the public distribution is different from the true distribution, the observation made by the Agent might be an over-represented point in the public distribution. If this is the case, the Agent will be underpaid when compared against Peers reporting samples from the true distribution, and some perturbation of the observation might pay better. One can visually inspect the true, public, and kernel distribution figures in the long version to see how the relationships between them produce the skewed figures (Richardson and Faltings 2023). This does not matter for the incentives in the ex-ante game that the Agents play, however, as it is an ex-post calculation.

5.2 Payment Stability

We simulate the expectation and variance of payments with respect to bin size for the partition. The bin size can affect the expected payment of the Agent in complicated ways when you take into account that bin-bounded kernels must account for the bin size. From the perspective of the Center, however, the bin size should not affect the expected payment. A smaller bin means a lower probability of matching, but a proportionately higher payment when matching. Intuitively, a smaller bin size will lead to higher variance in the payments. We demonstrate this relationship in Figure 4. The stability of the payments could be a consideration for designing the mechanism to take into account either Centers or Agents who aren't risk-neutral.

6 Conclusion

We present Peer Neighborhood mechanisms, a novel framework for extending Peer Consistency from discrete distributions to arbitrary distributions with minimal additional constraints or assumptions. We formulate the explicit extension of the Peer Truth Serum to the PTNE and prove its incentive-compatibility. We show that the strengthened Agent update condition, the Partition Expected extension, still admits a broad class of reasonable update processes. Finally, we conduct simulations to demonstrate the strength of the incentives with respect to perturbations from truthfulness, and the stability of payments with respect to the bin size of the partitions chosen by the Center. Peer Neighborhood mechanisms are not only practically implementable, but present a rich concept for future research on mechanism generalization.

References

Chen, Y.; Shen, Y.; and Zheng, S. 2020. Truthful data acquisition via peer prediction. *Advances in Neural Information Processing Systems*, 33: 18194–18204.

Dasgupta, A.; and Ghosh, A. 2013. Crowdsourced judgement elicitation with endogenous proficiency. In *Proceedings of the 22nd international conference on World Wide Web*, 319–330.

d'Aspremont, C.; and Gérard-Varet, L.-A. 1979. Incentives and incomplete information. *Journal of Public economics*, 11(1): 25–45.

Goel, N.; and Faltings, B. 2020. Personalized peer truth serum for eliciting multi-attribute personal data. In *Uncertainty in Artificial Intelligence*, 18–27. PMLR.

Kong, Y.; and Schoenebeck, G. 2019. An information theoretic framework for designing information elicitation mechanisms that reward truth-telling. *ACM Transactions on Economics and Computation (TEAC)*, 7(1): 1–33.

McAfee, R. P.; and Reny, P. J. 1992. Correlated information and mecanism design. *Econometrica: Journal of the Econometric Society*, 395–421.

Miller, N.; Resnick, P.; and Zeckhauser, R. 2005. Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9): 1359–1373.

Prelec, D. 2004. A Bayesian truth serum for subjective data. *science*, 306(5695): 462–466.

Radanovic, G.; and Faltings, B. 2014. Incentives for truthful information elicitation of continuous signals. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28.

Radanovic, G.; and Faltings, B. 2015. Incentivizing truthful responses with the logarithmic peer truth serum. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, 1349–1354.

Radanovic, G.; Faltings, B.; and Jurca, R. 2016. Incentives for effort in crowdsourcing using the peer truth serum. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(4): 1–28.

Richardson, A.; and Faltings, B. 2023. Peer Neighborhood Mechanisms: A Framework for Mechanism Generalization. arXiv:2312.12303.

Shnayder, V.; Agarwal, A.; Frongillo, R.; and Parkes, D. C. 2016. Informed truthfulness in multi-task peer prediction. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, 179–196.

Von Ahn, L.; and Dabbish, L. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 319–326.

Waggoner, B.; and Chen, Y. 2014. Output agreement mechanisms and common knowledge. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 2, 220–226.

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

Witkowski, J.; and Parkes, D. 2012. A robust bayesian truth serum for small populations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26, 1492–1498.